Probabilistic Arithmetic

by

Robert Charles Williamson, B.E. (Q.I.T.), M.Eng.Sc. (Qld)

A thesis submitted for the degree of Doctor of Philosophy,
Department of Electrical Engineering, University of Queensland,
August, 1989.

Statement of Originality

To the best of the candidate's knowledge and belief, the material presented in this thesis is original, except as acknowledged in the text, and the material has not been submitted, either in whole or in part, for a degree at this or any other university.

Robert Williamson.

Abstract

This thesis develops the idea of probabilistic arithmetic. The aim is to replace arithmetic operations on numbers with arithmetic operations on random variables. Specifically, we are interested in numerical methods of calculating convolutions of probability distributions. The long-term goal is to be able to handle random problems (such as the determination of the distribution of the roots of random algebraic equations) using algorithms which have been developed for the deterministic case. To this end, in this thesis we survey a number of previously proposed methods for calculating convolutions and representing probability distributions and examine their defects. We develop some new results for some of these methods (the Laguerre transform and the histogram method), but ultimately find them unsuitable. We find that the details on how the ordinary convolution equations are calculated are secondary to the difficulties arising due to dependencies.

When random variables appear repeatedly in an expression it is not possible to determine the distribution of the overall expression by pairwise application of the convolution relations. We propose a method for partially overcoming this problem in the form of *dependency bounds.* These are bounds on the distribution of a function of random variables when only the marginal distributions of the variables are known. They are based on the Fréchet bounds for joint distribution functions.

We develop efficient numerical methods for calculating these dependency bounds and show how they can be extended in a number of ways. Furthermore we show how they are related to the "extension principle" of fuzzy set theory which allows the calculation of functions of fuzzy variables. We thus provide a probabilistic interpretation of fuzzy variables. We also study the limiting behaviour of the dependency bounds. This shows the usefulness of interval arithmetic in some situations. The limiting result also provides a general law of large numbers for fuzzy variables. Interrelationships with a number of other ideas are also discussed.

A number of potentially fruitful areas for future research are identified and the possible applications of probabilistic arithmetic, which include management of numeric uncertainty in artificial intelligence systems and the study of random systems, are discussed. Whilst the solution of random algebraic equations is still a long way off, the notion of dependency bounds developed in this thesis would appear to be of independent interest. The bounds are useful for determining robustness of independence assumptions: one can determine the range of possible results when nothing is known about the joint dependence structure of a set of random variables.

Acknowledgements

I would like to acknowledge the help and support of the following people who contributed to this thesis in various ways.

Most importantly I would like to publicly thank my wife Angharad, who has assisted me greatly in numerous ways over the years. Without her help and support this thesis would not have been possible. I am deeply grateful for all the assistance she has provided.

I would also like to thank:

My supervisor, Professor Tom Downs, for continued support, assistance and encouragement, and for being a solver of administrative problems rather than a generator of them;

Former head of department Professor Mat Darveniza, who helped me as a beginning graduate student and later assisted in my obtaining a CPRA scholarship in order to pursue my PhD studies;

The head of department, Professor Tom Parnell, for the use of departmental facilities in the preparation of this thesis;

Dr Phil Diamond of the Department of Mathematics, University of Queensland, for useful advice and discussions and for the loan of some papers;

An anonymous referee of the International Journal of Approximate Reasoning, for a question which led to section 5.2 of chapter 3;

Professor Brian Anderson of the Australian National University, whose simple question as to whether I had looked at any limiting results was part of the motivation for chapter 5;

Dr Helen McGillivray of the Department of Mathematics, University of Queensland, for a discussion on confidence intervals and other matters and the loan of a book;

Dr Guy West of the Department of Economics, University of Queensland, for discussions and the loan of some papers.

The following people assisted me by sending comments on some questions and reprints of their papers: Professor Alice Agogino (University of California at Berkeley), Dr Piero Bonissone (General Electric Research Labs, New York), Dr Joel Brenner (Palo Alto), Professors Didier Dubois and Henri Prade (University of Toulouse III), Professor Jürgen Garloff (Institut für Angewandte Mathematik Universität

Freiburg), Professor Robin Giles (Queen's University, Ontario), Professor V.L. Girko (Kiev University), Dr Ellen Hisdal (University of Oslo), Professor M. Sambandham (Atlanta University), Dr Elie Sanchez (Marseille), Professor Berthold Schweizer (University of Massachusetts, Amherst), Professor Eugene Seneta (University of Sydney), Professor Ross Shacter (Stanford University), Dr Oskar Sheynin (Moscow), Professor Ushio Sumita (University of Rochester) and Professor Klaus Weise (Physikalisch-Technische Bundesanstalt, Braunschweig, West Germany).

I would also like to thank the indulgence of the Electrical Engineering librarian Mrs Barbara Kormendy, and the Mathematics Librarian Mrs Freida Kanowski.

It is a pleasure to acknowledge a debt to Donald Knuth and Leslie Lamport for the development of TEX and LATEX under which this thesis was prepared, and to James Alexander for his Tib bibliographic preprocessor. These tools were most useful.

Financial support from a Commonwealth Postgraduate Research Award and a grant from the Australian Research Grants Scheme is gratefully acknowledged.

Finally, I would like to acknowledge an intellectual debt to Karl Popper, who, through his writings, has taught me the importance of clarity of exposition and honesty of thought.

Publications

Most of the material in this thesis has been, or will be, published elsewhere. The following is a list of papers, either published, submitted, or nearing submission, which report material on the topic of this thesis. Some of these papers report further material which is not included in this thesis.

1. Robert C. Williamson and Tom Downs, "Probabilistic Arithmetic and the Distribution of Functions of Random Variables," Proceedings of the 1st IASTED Symposium on Signal Processing and its Applications, Brisbane, August 1987, 112–119, 1987. (Parts of chapter 2, including further details and examples on the histogram method).

2. Robert C. Williamson and Tom Downs, "The Inverse and Determinant of a 2×2 Uniformly Distributed Random Matrix," *Statistics and Probability Letters,* **7,** 167–170, 1989. (Chapter 7).

3. Robert C. Williamson and Tom Downs, "Probabilistic Arithmetic: Numerical Methods for Calculating Convolutions and Dependency Bounds," accepted for publication in the *International Journal of Approximate Reasoning,* 1989. (Chapter 3).

4. Robert C. Williamson, "An Extreme Limit Theorem for Dependency Bounds of Normalised Sums of Random Variables," accepted for publication in *Information Sciences,* 1989. (Chapter 5).

5. Robert C. Williamson, "The Law of Large Numbers for Fuzzy Variables under a General Triangular Norm Extension Principle," under revision for resubmission to *Fuzzy Sets and Systems,* 1989. (Chapter 6).

6. Robert C. Williamson, "Interval Arithmetic and Probabilistic Arithmetic," to appear in the proceedings of SCAN-89 IMACS-GAMM-GI International Symposium on Computer Arithmetic and Self-Validating Numerical Methods, Basel, October 1989. (A summary of chapters 3 and 5, and a discussion of the relationship between probabilistic arithmetic and interval arithmetic).

7. Robert C. Williamson and Tom Downs, "Probabilistic Arithmetic: Relationships with Other Ideas," to be submitted to the *International Journal of Approximate Reasoning,* 1989. (Chapter 4).

8. Robert C. Williamson and Tom Downs, "Numerical Methods for Calculating Convolutions of Probability Distributions," to be submitted to *Mathematics and Computers in Simulation,* 1989. (Chapter 2).

9. Robert C. Williamson, "The Quotient of Two Normal Random Variables: An Historical Study," in preparation, 1989. (Traces the history of a simple problem related to probabilistic arithmetic).

10. Robert C. Williamson, "The Discrete T-conjugate Transform," in preparation. To be submitted to *Information Sciences,* 1989. (Extends the T-conjugate transform — see section 6.4 and chapter 5).

Contents

List of Figures

Chapter 1

Probabilistic Arithmetic — The Very Idea

Now that you have learnt about adding, minussing, multiplying and dividing you can do any sum. Even an atomic scientist only really uses these four operations.
— Mrs S. Boal (Grade 1, Oakleigh Primary School, 1968)

If one looks at the development of the measurement process during the past century, one soon observes that with increasing frequency the raw data are (probability) distribution functions or frequency functions rather than real numbers. This is so in the physical sciences; and in the biological and social sciences it is the rule rather than the exception. One may thus convincingly argue that distribution functions are the "numbers" of the future and that one should therefore study these new numbers and their arithmetic.
— Berthold Schweizer

1.1 Motivation

This thesis studies the idea of *probabilistic arithmetic,* which is the name we give to the idea of calculating the distribution of arithmetic (and perhaps other) functions of random variables. It is motivated by the hope of being able to develop procedures for solving random problems based on those already existing for the deterministic case. The fundamental observation to make is that nearly all existing numerical algorithms are based on the four operations of arithmetic: addition, subtraction, multiplication and division. For example, it may be possible to determine the distribution of the roots of polynomials with random coefficients by modification of existing algorithms for calculation of the roots in the deterministic case. Work to date on random polynomials has only obtained this for a very restricted class of polynomials [74, 356].

The first step necessary in achieving this goal is the development of an efficient and accurate method for determining the convolution of probability distribution functions. We use the word "convolution" in its general sense describing the operation on distribution functions corresponding to virtually any operation on random

variables, and not just addition. If $Z = L(X,Y)$ where X and Y are independent random variables with joint distribution function F_{XY}, then F_Z, the distribution of Z, is given by

$$F_Z(z) = \int_{L\{z\}} \mathrm{d}(F_X(u)F_Y(v)), \tag{1.1.1}$$

where $L\{z\} = \{(u,v) | \, u,v \in \Re, L(u,v) < z\}$. The ability to calculate (1.1.1) is necessary but not sufficient for the construction of a probabilistic arithmetic.

The most important difference between the deterministic and stochastic cases is the appearance of stochastic dependence. This has no counterpart in the deterministic case. We shall see that even if we restrict ourselves to independent random variables at the outset, stochastic dependencies can arise during the course of a calculation due to the occurrence of repeated terms in expressions. For example, if X, Y and Z are independent random variables, then $V = X+Y$ and $W = X \times Z$ are not necessarily independent. We give the name *dependency error* to the error incurred in calculating the distribution of some function of V and W (such as $U = V/W$) by assuming that V and W are independent.

Whilst a general solution to this problem seems impossible, we can provide a partial solution in terms of *dependency bounds.* This is the name we give to lower and upper bounds on the distribution of functions of random variables when only the marginal distributions are known. In other words the dependency bounds contain all the possible results due to all the possible joint distributions of the random variables involved. If ldb and udb denote the lower and upper dependency bounds, and $\Box$ denotes a binary operation, we write the bounds on $Z = X \Box Y$ as

$$\mathrm{ldb}(F_X, F_Y, \Box)(z) \leq F_Z(z) \leq \mathrm{udb}(F_X, F_Y, \Box)(z). \tag{1.1.2}$$

These bounds can be calculated explicitly for certain classes of binary operations.

The general approach we have just outlined has been studied previously by a number of authors. Whilst we defer a more detailed review of previous work to later chapters, let us mention the following now. The idea of an *algebra of random variables* based on the use of integral transforms for calculating the appropriate convolutions was advocated by Springer in [775] and formed the motivation for [220]. Analytical methods for determining distributions of functions of random variables are at the core of applied probability theory [632] (an early, but obscure review is given by Halina Milicer-Grużewska [575]), and of course the study of the distribution of sums (in the central limit theorem) is the basis for many of the theoretical results in probability theory [177]. Numerical methods for calculating distributions of functions of random variables have been studied by a number of authors (chapter 2 is a review of available techniques). Some of the more recent techniques are the histogram method [160] (and the related method of discrete probability distributions [423]), and the intricate H-function method [164]. The problem we study is obviously related to the propagation of errors of measurement [820]. Fuzzy arithmetic [432] derives from similar motivations and it is compared with probabilistic arithmetic in chapter 4.

There are many problems arising in the course of our study of probabilistic arithmetic, and in this thesis we present solutions to some of them. However, many

of them remain topics for future research and we give some more details on these in the final chapter of this thesis. We note here that our idea of probabilistic arithmetic can be considered to be a natural generalisation of interval arithmetic [584], which works entirely in terms of the supports of the distributions of the variables involved. We discuss this connexion in more detail later in this thesis.

The rest of this chapter gives an overview of the results contained in this thesis (section 1.2); a brief decription of the structure of the thesis and suggestions on how to read it (section 1.3); and a few standard notational conventions (section 1.4).

1.2 Outline of Results

Some of the specific results obtained in this thesis (the highlights) are now described. The general structure of the thesis is described in section 1.3.

1.2.1 Different Methods for Calculating Convolutions and the Laguerre Transform Method

In chapter 2 we study a variety of methods for calculating convolutions of probability distributions. Amongst these are the Laguerre transform methods developed by Keilson and Nunn [435]. We consider the possibility of using this method for operations other than addition and subtraction of random variables. Although we develop a number of new results, these turn out to be of little practical value because of the computational complexity of the formulae involved.

1.2.2 The Dependency Bounds and Numerical Methods of Calculating Them

The most interesting new techniques developed in this thesis are those for numerically calculating dependency bounds. We use the results of Frank, Nelsen and Schweizer [277] who show that

$$\mathrm{ldb}(F_X, F_Y, \Box)(z) = \sup_{x \Box y = z} \max(F_X(x) + F_Y(y) - 1,\, 0) \tag{1.2.1}$$

and

$$\mathrm{udb}(F_X, F_Y, \Box)(z) = \inf_{x \Box y = z} \min(F_X(x) + F_Y(y),\, 1) \tag{1.2.2}$$

for certain classes of operations $\Box$. Then by making use of a duality result to express these bounds in terms of the inverses of the distribution functions (the quantiles), we develop an efficient and accurate method for calculating the dependency bounds. A number of extensions to the dependency bounds are also considered. If *some* information is known about the joint distribution, then tighter bounds can be calculated. All this material is reported in chapter 3.

1.2.3 Precursors, Multiple Discoveries, and Relationships with Fuzzy Sets

In the course of reviewing previously published material on the topics covered in this thesis, a number of independent and multiple discoveries and precursors were found. For example George Boole discussed the analogue of dependency bounds for events in his *Investigation of the Laws of Thought* [97]. These are bounds on the probability of conjunction and disjunction of random events. They are usually associated with Fréchet [279], who rightly credits their original introduction to Boole. Boole's work on these bounds as lower and upper probabilities has recently seen a revival of interest in the area of expert systems which have to deal with uncertain information. In some cases recent authors are unaware of some of the previous work in the field [7,353,836]. We discuss this material in detail in section 4.1. There are several other instances of multiple discoveries, such as Rüschendorf [695], Makarov [536] and Frank, Nelsen and Schweizer [277] on the dependency bounds for random variables. Another example is the duality result we use for numerical calculation of the dependency bounds. This has been presented (in various degrees of generality) by Frank and Schweizer [278] Sherwood and Taylor [743], Höhle [390], Fenchel [263] (see [530]), Bellman and Karush [65], Nguyen [615] and Mizumota and Tanaka [579]. We discuss this in section 4.5. Of course these multiple discoveries should not be considered surprising, especially given the arguments of Lamb and Easton on the "pattern of scientific progress" [498].

1.2.4 The Inverse and Determinant of a Random Matrix

Chapter 7 is concerned with a new result on the inverse and determinant of a random matrix. Inverses and determinants of *interval* matrices (matrices with interval coefficients) have been studied in the literature, and so we examined the effect of interpreting an interval as a uniformly distributed random variable. Apart from some interesting conclusions in this regard, the chapter shows the disadvantage of always seeking analytical results: the formula for the density of the determinant is quite complex and it is necessary to write a computer program in order to determine its specific values. This is one of the arguments of this thesis: When specific values of distributions are required (rather than general properties), it makes more sense to accept the need for numerical calculations at the outset rather than doggedly striving for analytical results which are often practically useless.

1.2.5 A Limiting Result for Dependency Bounds

Perhaps the most interesting purely mathematical result in this thesis is that reported in chapter 5. There we show that the dependency bounds of the normalised sum $\frac{1}{N}\sum_{i=1}^{N} X_i$ converge to step functions. The position of the step functions depends solely on the support of distribution functions of the random variables $\{X_i\}$. The method used to prove this result (T-conjugate transforms) has further potential applications. These are discussed in chapters 6 and 8. The *interpretation* of

the result is that in some cases, if it is necessary to use dependency bounds (because of lack of further dependence information), then no further information will be obtained beyond that obtained by using interval arithmetic on the supports of the distributions.

1.3 Thesis Structure

Each of the chapters in this thesis is essentially self-contained and can thus be read independently of the others, although chapter 3 should be read before reading chapters 4–7. In fact each chapter was written as a paper, details of which are given in the "Publications" section at the beginning of the thesis and at the beginning of each chapter. For this reason there is some slight repetition of material in some places. We feel this will in fact be of benefit to the reader, as most of the repeated material is concerned with technical definitions which are difficult to remember for the entire length of the thesis.

The remaining chapters of the thesis are summarised below:

Chapter 2 is a survey of different methods for numerically calculating convolutions of probability distributions. As well as reviewing previous work, we present some new results on the Laguerre transform method and the histogram method. We also point out some interesting connexions with methods used in metrology.

Chapter 3 is the technical core of the thesis. Most of the other chapters are motivated by it. In it we fully develop the idea of dependency bounds and show how they can be calculated numerically. A number of examples are given. Furthermore a method for calculating ordinary convolutions (1.1.1) is presented. This uses the numerical representation developed in order to calculate dependency bounds and the result is that this method has a number of advantages over the methods described in chapter 2; not the least of which is its simplicity.

Chapter 4, which is the longest in this thesis, describes a large number of interrelationships between our dependency bound methods and other ideas. Amongst other things we discuss the Boole-Fréchet bounds, graph theoretical methods, lower and upper probabilities and fuzzy arithmetic. This latter item is of considerable interest. We show how the rules for combining fuzzy numbers are very closely related to the dependency bound formulae (1.2.1–1.2.2).

Chapter 5 is devoted to the proof and exposition of a limiting theorem for dependency bounds. We prove a generalisation of the law of large numbers under no independence assumptions. We show that convergence to a wide range of distributions is possible under the constraint that the restrictions on the supports as calculated by interval arithmetic are not violated.

Chapter 6 simply re-presents the result of chapter 5 in terms of fuzzy variables. Thus we show a law of large numbers for fuzzy variables under a general t-norm extension principle. (The meaning of this sentence is explained in the introduction to chapter 6.)

Chapter 7 presents a new result on the distribution of the inverse and determinant of a random matrix. We derive explicit formulae for perhaps the simplest case of a random matrix: a 2×2 matrix with independent elements, all having a uniform distribution on $[0, 1]$. Surprisingly enough this result does not seem to have appeared in the literature before. Our motivation for deriving it was to examine the effect of interpreting an interval in interval arithmetic as a uniformly distributed random variable.

Finally chapter 8 draws a number of conclusions from the work presented here and provides a number of suggestions for future research.

1.4 Notational Conventions

Equations and sections are numbered by chapter. Thus (2.3.5) refers to the 5th numbered equation in section 2.3, which is the third section of chapter 2. Parenthesised numbers always refer to equations. Numbers enclosed in square brackets, such as "[13]," denote references to items in the reference list in chapter 9. Closed intervals such as [0,1] ($\{x | 0 \leq x \leq 1\}$) are also denoted in this manner, but the context makes clear what is meant.

Other mathematical notations are generally standard. We list the following which may otherwise cause some confusion.

inf	infimum (greatest lower bound).
sup	supremum (least upper bound).
$\Re$	The set of real numbers.
$\Re^*$	The set of extended real numbers ($\Re^* = \Re \cup \{-\infty, \infty\}$).
$df(X)$	The distribution function of X.
$\leq$	(for functions) pointwise inequality.
supp	The *support* of a function.
Ran	The *range* of a function.
Dom	The *domain* of a function.
$[a, b)$	The half-open interval $\{x \mid a \leq x < b\}$.
$\lvert\{\ \}\rvert$	Cardinality of a set.
$\emptyset$	The null set.
$\setminus$	Set-theoretic difference.
iff	If and only if.
■	End of proof.

Other notations are introduced where needed.

Several algorithms are presented in this thesis. We use the syntax of the C programming language [453] with the exception that := denotes assignment and = equality.

Chapter 2

Numerical Methods for Calculating Convolutions of Probability Distributions

Questions on local probability and mean values are of course reducible by the employment of Cartesian or other coordinates, to multiple integrals .... The intricacy and difficulty to be encountered in dealing with such multiple integrals and their limits is so great that little success could be expected in attacking questions directly by this method.
— M.W. Crofton

2.1 Introduction, Aim and Analytical Methods

The problem of calculating convolutions of probability distribution functions arises in a wide range of applications where distributions of functions of random variables are required. In many cases analytical solutions are intractable and so numerical methods are used. This chapter will survey the numerical methods that have been presented to date. This will form a suitable background for chapter 3 where we present a new method for calculating convolutions of distribution functions. We shall see that our new method has a number of aspects in common with the methods described in the current chapter, but that it also has a number of advantages.

We will generally restrict ourselves to the problem at hand, and not discuss the applications in which convolutions arise. Generally we adopt the viewpoint of von Mises [577, p.32] who took the attitude that

> the exclusive purpose of [probability] theory is to determine, from the given probabilities in a number of initial collectives, the probabilities in a new collective derived from the initial one.

The determination of the initial probabilities is the province of statistics and we do not consider it further.

2.1.1 History, Motivation and Outline

History of the Problem

It was recognised long ago that if the inputs to some calculation are random then the final result is likely to be random also. Sheynin [747, p.117] cites a 16th century commentary by Gáńéza on the 12th century Indian writing *Lilávati* by Bháscara. On page 97 of this [75], there is a discussion of the determination of an area of a rectangle when the length and breadth are not known exactly and there is an implicit recognition that the area will be only approximate. It is suggested there that mean values should be used in order to provide a better estimate than single sample values. Of course other examples can easily be found. However the statement of the problem in terms of determining *distributions* of functions of random variables could not have occurred until last century:

> The point is that laws of distributions of functions of random quantities (even the simplest, the linear functions) could not have been considered from a general point of view at least until distribution functions themselves began to be considered *per se* [745, p.290].

Thus although "it could be hardly doubted that Laplace [in his *Théorie Analytique des Probabilités* (1812)] would have been able to transform distributions from one interval to another and from one argument [...] to another [...]" [745, p.290], one of the first *explicit* transformations, in terms of distributions, appears to have been due to Poisson in 1837 [653]. Sheynin [748, p.295] says of a problem considered there that

> elementary as it is, this [...] problem seems to be one of the first in which densities were treated as purely mathematical objects.

Another early example which arose in a completely different context concerns geometrical probability [449]. Sylvester's problem [809], whilst not couched in terms of functions of random variables (nor for that matter solved in such a fashion), is still an example of the sort of problem we are concerned with.

The Aim of this Chapter

So much for the history of the problem. We will now outline the purpose of the present chapter.

Until recently most attempts at the determination of distributions of functions of random variables have entailed the search for analytical solutions. In other words, a formula for the required distribution was sought. Quite apart from the severe difficulties encountered in this approach (see section 2.1.2 below), a new formula needs to be determined for each new problem. Thus, whilst for some restricted classes of problems (such as products and quotients of independent random variables with "standard" distributions [674]) the main results can be tabulated, in general

"it is not practicable to give a list of such occasional results, for it is clearly possible to invent further variations at will" [413, p.289]. A review of a number of exact analytical techniques for determining distributions arising in multivariate statistics is given in [547].

An alternative to the analytical approach is to use numerical methods for calculating the required distributions by using computer programs. Note that such an approach should not necessarily be considered second best compared to an exact analytical result. This is because the resulting formulae are sometimes so complex (*e.g.* infinite series of transcendental functions) that a computer program is needed to calculate the specific values and to determine the behaviour of the distribution. Thus we are really no worse off if we decide to use the computer from the outset. The study of these numerical methods is the focus of the present chapter.

Organisation of the Rest of the Chapter

The rest of this chapter is organized as follows. The remainder of this introductory section is devoted to a review of analytical methods (including integral transforms), and the possible direct numerical calculation of these. Section 2.2 summarises a number of different methods that have been proposed, including the use of moments. Sections 2.3 and 2.4 look in rather more detail at two methods which would appear to be more promising. We examine the Laguerre transform in section 2.3, and look at the Histogram (or discrete probability distribution) method in section 2.4. Section 2.5 examines interval arithmetic and methods used for the propagation of errors in metrology.

2.1.2 Exact Analytical Results

There is a well known general solution to the distribution of functions of random variables in terms of the Jacobian of transformation. We will now briefly present this result for a function of only two random variables. We restrict ourselves to this special case because we are mainly interested in functions of only two random variables (such as apply to the four arithmetic operations) and because there are notational difficulties in presenting the full general result carefully. (There is no real conceptual difficulty in extending the results to functions of N random variables.)

The General Solution in terms of the Jacobian

Let $Z = g(X,Y)$ and $W = h(X,Y)$ be two functions of the two random variables X and Y which have a joint probability density f_{XY}. We wish to determine the joint density f_{ZW} of Z and W. This is given by [632, p.201]

$$f_{ZW}(z,w) \;=\; \frac{f_{XY}(x_1,y_1)}{|J(x_1,x_2)|} + \cdots + \frac{f_{XY}(x_k,y_k)}{|J(x_k,y_k)|}, \tag{2.1.1}$$

where $(x_1, y_1), \ldots, (x_k, y_k)$ are the k real solutions to the pair of equations

$$g(x, y) = z \qquad \text{and} \qquad h(x, y) = w \tag{2.1.2}$$

in terms of z and w where x is the absolute value of x. In other words

$$g(x_i, y_i) = z \qquad \text{and} \qquad h(x_i, y_i) = w$$

for $i = 1, \ldots, k$, and $x_i \neq x_j$, $y_i \neq y_j$ for $i \neq j$. The term $J(x, y)$ is the Jacobian of transformation and is given by

$$J(x, y) \;=\; \det \begin{bmatrix} \left.\dfrac{\partial g}{\partial x}\right|_{(x,y)} & \left.\dfrac{\partial g}{\partial y}\right|_{(x,y)} \\ \left.\dfrac{\partial h}{\partial x}\right|_{(x,y)} & \left.\dfrac{\partial h}{\partial y}\right|_{(x,y)} \end{bmatrix}, \tag{2.1.3}$$

where the notation $\left.\frac{\partial g}{\partial x}\right|_{(x,y)}$ means that the partial derivative $\frac{\partial g}{\partial x}$ is evaluated at (x, y). The density f_{ZW} is equal to zero for any (z, w) such that (2.1.2) has no real solutions.

The use of Auxiliary Variables

We are often interested in m functions of n random variables with $m < n$. (In the present chapter we are only concerned with $m = 1$ and usually $n = 2$.) In order to use the above method in this situation it is necessary to proceed as follows.

Let $m = 1$ but consider general n. Write

$$Z_1 \;=\; g(X_1, \ldots, X_n)$$

for the function we are interested in. It is necessary to define $n - m$ auxiliary functions in the following manner:

$$\begin{aligned} Z_2 &= X_2 \\ &\vdots \\ Z_n &= X_n. \end{aligned}$$

The use of (2.1.1) will give $f_{Z_1 Z_2 \cdots Z_n}$. In order to determine f_{Z_1} we need to calculate

$$f_{Z_1}(z_1) \;=\; \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1} f_{Z_1 Z_2 \cdots Z_n}(z_1, \ldots, z_n) \prod_{j=2}^{n} dz_j. \tag{2.1.4}$$

The integral in (2.1.4) is almost invariably the cause of difficulties encountered in determining f_{Z_1} analytically. For practical problems n can be quite large (see for example [275] where $n = 100 \times 100 = 10000$ and thus a 9999-fold integral needs to be evaluated!). Even when the integrals are tractable they can be exceedingly tedious: see the 12 page calculation in [860, appendix IV] (which actually contains an error that simplifies the calculation). Somewhat more reasonable examples can be found in [352,632]. In section 2.1.3 we show how formulae 2.1.1 and 2.1.4 result in the standard convolution integrals for arithmetic functions of random variables. The *numerical* integration of the Jacobian of transformation has been considered by Cook and Downs [161].

Historical Remarks

The use of the Jacobian of transformation in the above manner was anticipated by Gauss in 1823 in his *Theoria Combinationis* [749, p.42], where he considered $m = 1$ and assumed $J > 0$ always and thus omitted the $|\ |$ operation in (2.1.1). The general solution was first given by Nasimov in 1889 [610]. It was subsequently studied by Poincaré and Lammel (see the footnote on page 81 of [182]).

2.1.3 Specialised Formulae for Convolutions

When the function g, of the two random variables X and Y, is one of the four arithmetic operations, we obtain the four convolution equations below [632,775].

$$Z = X + Y: \quad f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z - x, x)\, dx, \tag{2.1.5}$$

$$Z = X - Y: \quad f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z + x, x)\, dx, \tag{2.1.6}$$

$$Z = X \times Y: \quad f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{XY}(z/x, x)\, dx, \tag{2.1.7}$$

$$Z = X/Y: \quad f_Z(z) = \int_{-\infty}^{\infty} |x| f_{XY}(zx, x)\, dx. \tag{2.1.8}$$

When X and Y are independent, these reduce to

$$Z = X + Y: \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(z - x) f_Y(x)\, dx, \tag{2.1.9}$$

$$Z = X - Y: \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(z + x) f_Y(x)\, dx, \tag{2.1.10}$$

$$Z = X \times Y: \quad f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(z/x) f_Y(x)\, dx, \tag{2.1.11}$$

$$Z = X/Y: \quad f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(zx) f_Y(x)\, dx. \tag{2.1.12}$$

These latter equations (2.1.9–2.1.12) will be our main but not exclusive concern in this chapter. We shall refer to them respectively as sum, difference, product and quotient convolutions. The lesser known product and quotient convolutions would appear to have first been published in [400]. Note that equations 1.9–1.12 can be written in terms of cumulative distribution functions as the Lebesgue-Stieltjes integral (which always exists)

$$\int\limits_{L\{x\}} dF_{XY}(u, v),$$

where $L\{x\} = \{(u, v) |\, u, v \in \Re,\, L(u, v) < x\}$ [520].

The idea of "convolution" of probability distributions has been generalised in a number of ways. For example, Urbanik [834,835] has studied a different type of convolution to those presented above. His generalised convolutions are also briefly mentioned by Schweizer and Sklar [718].

2.1.4 Integral Transforms

Definitions

We now consider the relationships between (2.1.9–2.1.12) and the Fourier and Mellin transforms. The Fourier transform of a function $f(x)$ is defined by

$$\mathcal{F}_t[f(x)] \triangleq \int_{-\infty}^{\infty} f(x)e^{-itx}\,dx. \tag{2.1.13}$$

This is a specialisation of the Laplace transform

$$\mathcal{L}_s[f(x)] \triangleq \int_{-\infty}^{\infty} f(x)e^{-sx}\,dx \tag{2.1.14}$$

where $s = \sigma + it$. The Mellin transform is defined by

$$\mathcal{M}_s[f(x)] \triangleq \int_{0}^{\infty} f(x)x^{s-1}\,dx. \tag{2.1.15}$$

The Mellin transform can be derived from the Laplace transform by a logarithmic change of variables [379]. Note that the Mellin transform is only defined for functions with domain $\Re^+$. The multivariate extensions of these transforms are

$$\mathcal{F}_{\mathbf{t}}[f(x_1,\ldots,x_n)] \triangleq \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\ldots,x_n)\prod_{j=1}^{n} e^{-it_jx_j}dx_j \tag{2.1.16}$$

where $\mathbf{t} = (t_1,\ldots,t_n)$, and

$$\mathcal{M}_{\mathbf{s}}[f(x_1,\ldots,x_n)] \triangleq \int_{0}^{\infty}\cdots\int_{0}^{\infty} f(x_1,\ldots,x_n)\prod_{j=1}^{n} x_j^{s_j-1}dx_j \tag{2.1.17}$$

where $\mathbf{s} = (s_1,\ldots,s_n)$.

The Calculation of Distributions of Sums and Products

When these transforms exist, they can be used to determine the distribution of arithmetic functions of random variables as follows. If X_j $(j = 1,\ldots,n)$ are independent random variables with densities $f_{X_j}(x_j)$ respectively, and if

$$Y = \sum_{j=1}^{n} X_j,$$

then

$$\mathcal{F}_t[f_Y(y)] = \prod_{j=1}^{n} \mathcal{F}_t[f_{X_j}(x_j)].$$

If the X_j are not independent, and have a joint probability density $f_{\mathbf{X}}(\mathbf{x})$, then

$$\mathcal{F}_r[f_Y(y)] = \mathcal{F}_{\mathbf{t}}[f_{\mathbf{X}}(\mathbf{x})]|_{t_1=t_2=\cdots=t_n=r}\,.$$

Calculation of differences can be accomplished by setting $X_j' = -X_j$. If

$$Y \;=\; \prod_{j=1}^{n} X_j$$

then if X_j are all independent

$$\mathcal{M}_s[f_Y(y)] \;=\; \prod_{j=1}^{n} \mathcal{M}_s[f_{X_j}(x_j)];$$

whilst if dependent,

$$\mathcal{M}_s[f_Y(y)] \;=\; \mathcal{M}_{\mathbf{r}}[f_{\mathbf{X}}(\mathbf{x})]|_{r_1=r_2=\cdots=r_n=s}\,.$$

If $Y = X_1/X_2$, then

$$\mathcal{M}_s[f_Y(y)] \;=\; \mathcal{M}_{\mathbf{r}}[f_{\mathbf{X}}(\mathbf{x})]|_{r_1=s,\,r_2=2-s}\,;$$

and if X_1 and X_2 are independent,

$$\mathcal{M}_s[f_Y(y)] \;=\; \mathcal{M}_{r_1}[f_{X_1}(x_1)]|_{r_1=s} \;\times\; \mathcal{M}_{r_2}[f_{X_2}(x_2)]|_{r_2=2-s}\,.$$

Thus if one has a means of inverting the Fourier and Mellin transforms, one can calculate the distributions of sums, differences, products and quotients of random variables. Note that we have omitted any mention of conditions for the existence of these transforms and of the uniqueness of the inverse transform. More details can be found in [254,476,517,607,663,775,787,905].

We note that the Fourier transform has been used in statistics for determining the sums of random variables for quite some time. Zolotarev [905] attributes its introduction to Lyapunov, although it appears [721, p.79] that Gauss knew of its applicability in 1813. The widespread use by Poisson and others did not occur until some time later. Fourier transforms are usually called *characteristic functions* in probability theory, a name first used by Poincaré in 1896 [650] (not, as Cuppens [177] suggests, by Lévy [516] in 1925). The Mellin transform has been used in probability theory since at least 1938 [607].

Difficulties in using Integral Transforms

Although the use of integral transforms does seem promising, there remains the problem of inverting the transform. This is usually the most difficult part. Indeed, most of the technical difficulties encountered by Springer in his book [775] occur in the inversion of the transforms. In a number of cases the analytical inversion formulas may be used. These are

$$f(x) \;=\; \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{sx}\mathcal{L}_s[f(x)]ds$$

and

$$f(x) \;=\; \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}\mathcal{M}_s[f(x)]ds.$$

However, these formulae are of little value for the development of a general numerical method for calculating distributions of functions of random variables. In their stead, approximate numerical methods which can be readily implemented on a computer need to be considered [775, chapter 8].

2.1.5 The H-function Method

One way of avoiding the necessity of calculating the inverse Mellin transform is to consider the Mellin transform of a very general function which includes as special cases all the actual functions to be encountered. The H-function [550], which was introduced by Fox [274], contains as special cases nearly all the special functions of applied mathematics. The significance of the H-function in the present context is that multiplication by a suitable constant (to make the integral over the domain of the function equal to 1) allows one to consider H-function distributions which include many classical univariate probability distributions as special cases. These are all only defined on the positive real line and include the gamma, exponential, Chi-square, Weibull, Rayleigh, Maxwell, half-normal, uniform, half-Cauchy, half-Student, F, Beta, and Bessel distributions. Representation by H-functions is of use in the context of probabilistic arithmetic becauses the probability densities of products, quotients, and rational powers of independent H-function random variables are also H-function random variables [132,133,164,777]. Thus if one can numerically calculate the H-function inversion integral (that is the contour integral defining the H-function), one can determine the probability distribution of products, quotients and rational powers of random variables which have any of the distributions listed above.

Methods of numerically evaluating the H-function inversion integral are discussed in chapter 7 of [775] and are based on the work of Lovett [526] and Eldred [252]. Eldred's method is simplified by Cook and Barnes [164] which is, however, still rather complicated. By combining their technique of inverting the Mellin transform with a numerical method of inverting the Laplace transform described by Crump [175], Cook and Barnes produce a method for calculating the distribution of the sum of products, quotients and rational powers of H-function random variables. In [164] they present a FORTRAN program implementing the algorithm. For successful operation, this program requires a number of technical parameters to be specified, and the authors give some suggested values. They say (page 313) that the calculation of the pdf of $WX^2 + Y$, where W, X, and Y were H-function random variables, required "about 24 seconds input/output and CPU time on a Cyber 170/750".

Other methods for the numerical inversion of Laplace and Mellin transforms are given in [86,185,186,215,826] The necessity of using the Laplace transform arises because the sum or difference convolution of two (or more) H-function distributions is not an H-function distribution. Whilst an analytical expression exists for the convolution, it appears too complicated to be of any use in probabilistic arithmetic [548]. Springer suggests using Fourier transforms to perform the convolution. Although the Fourier transform of an H-function distribution is known [775, equation 6.3.2], inverting the Fourier transform may be intractable analytically. In [775, section 8.2],

Springer refers to unpublished work of Carter [132] on a numerical technique for determining the moments of the distribution of the sum of independent H-function random variables. This would allow approximations to the distribution to be obtained using standard techniques for determining distributions from moments (see section 2.2.2 below).

2.2 Miscellaneous Numerical Methods for Calculating Convolutions

The problem of numerically calculating convolutions is related to the general problem of approximating distributions which arises in many areas where the digital computer is not involved. A good survey of the general problem is given by Bowman and Shenton [104].

2.2.1 Normal Approximations and other Parameterised Distributions

Perhaps the simplest and most widespread method of calculating the distributions of certain functions of random variables is to use a normal approximation to the distributions involved. The advantage of doing this is that normal distributions are closed under sum and difference convolutions. Thus if $df(X) = N(\mu_X, \sigma_X)$ and $df(Y) = N(\mu_Y, \sigma_Y)$, then $df(X+Y) = N(\mu_X + \mu_Y, (\sigma_X^2 + \sigma_Y^2)^{1/2})$, if X and Y are independent. Therefore we could represent the distributions simply by μ and σ for the purposes of numerical calculation. This idea has been used by Pearl [642] and Sobierajski [769] and is discussed by Corsi in [168]. Note however that neither $df(XY)$ or $df(X/Y)$ are normal in general [35,267].

When using normal approximations, two different approaches can be taken in order to interpret the final results. The first approach is simply to assume that the distributions involved are normal. This is often done in error analysis. It is an unjustified assumption in many cases, not only in the theory of errors, but in other areas as well. The second approach includes the admission that the distributions are in fact non-normal but uses normal approximations for the sake of calculation. Whilst this is often acceptable, it is certainly not a suitable methodology for a general probabilistic arithmetic.

One problem in fitting a standard distribution, which arises more often than is recognised, is described by Greenberg [338] as follows:

> The analyst thus has great latitude in choosing a distribution to fit his data — naturally he will select one that is convenient to work with and easily manipulated. The only serious differences between the data and the selected distribution will probably be in the tails, where relatively few (if any) of the observation[s] will lie. However, it is in these tails that the events of interest occur: the large delays, the long queue lengths *etc.* Thus any investigation of these events either by analytic means or by simulation (especially if importance sampling is used to obtain a larger

> representation of the values in the tail) is bound to be greatly [a]ffected by the distribution chosen — and the distribution must be chosen with little or no representation in the region of most interest.

The fact that the normal distribution may be a good match to some population distribution everywhere except in the tails is the subject of Bagnold's paper [43]. (See also the reply by Rothschild [691].) The effect of the extreme tails on a final result is quite pronounced in some cases. In chapter 5 we show that the application of one method for overcoming a problem we call dependency error (see chapter 3) gives results that depend *entirely* on the extreme tail behaviour in certain circumstances.

Rather than using a normal distribution, a more flexible parameterised family of distributions could be used. Recall that we have already examined (section 2.1.5) the use of H-function distributions. A simpler alternative is to use, say, the Pearsonian curves [251,628]. Whilst these curves provide a much better fit to a wide range of distributions (compared with the normal distribution), their use as a general method for calculating distributions of functions of random variables is severely restricted by the fact that apart from certain special cases [59,482], the distribution of a function of random variables is not available in terms of the parameters of the distributions involved. Even when the distribution *is* available, it is not necessarily Pearsonian [482]. In other words, the family is not closed under convolutions. General conditions that need to be satisfied for a parameterised family to be closed under sum convolutions are given by Crow [174], and these are restrictive enough for us to discard the idea.

Thus we reject both normal and other parameterised distributions as being unsuitable for our purposes. Not only are the assumptions often invalid, but we cannot calculate the convolutions of interest in terms of the parameters involved.

2.2.2 Methods Based on Moments

Definitions and Basic Approach

Another simple method which deserves consideration is representation using moments. The *n-th central moment* of a random variable X with distribution function F_X is defined by

$$\mu_X^{(n)} = E(X^n) = \int_{\Re} x^n \, dF_X(x)$$

assuming the integral exists. The *method of moments* (introduced by Tchebyshev and Markov — see appendix II of [837]) entails calculating results in terms of moments of distributions when exact distributional results are unavailable. Given the moments of a distribution, one can either fit a distribution using the techniques described in [376], or one can use the Tchebyshev inequalities or their generalisations [212,317,318,537] to determine distributional results. Unfortunately, there are problems associated with both of these methods. Firstly, the moments do not always determine a distribution exactly [752] so that two different distributions can have all moments identical [462,635,644]; and secondly, the Tchebyshev inequalities are often quite loose.

There are also a number of ad hoc methods which use moments. For example, Broadbent [107] suggested approximating the distributions of products and quotients by lognormal distributions fitted to the moments. This is unsuitable for our needs because of restrictions on the class of distributions that can be accommodated.

Even if these problems are ignored, there are still substantial difficulties in using moments as a basis for probabilistic arithmetic. This is because although there is a simple and exact formula for the moments of a sum of two independent random variables

$$\mu_{X+Y}^{(n)} \;=\; \sum_{s=0}^{n} \binom{n}{s} \mu_X^{(s)} \mu_Y^{(n-s)}$$

(see [587, p.267]), there are no such simple results for products and quotients. Indeed, for the case of the quotient, the moments may not even exist. It is easy to prove that if $Z = X/Y$, where X and Y are random variables with densities f_X and f_Y bounded and continuous at the origin, then in order for $\mu_{X/Y}^{(n)}$ to exist it is necessary (and sufficient) for $\mu_X^{(n)}$ to exist and for f_Y to have a zero of order $n-2$ at the origin. (In some cases the odd order moments exist as Cauchy principal values.)

Approximate Formulae for Moments of Functions of Random Variables

Whilst exact formulae for moments of products and quotients do not exist, there are useful approximations which have found wide application in applied probability theory. There are two different ways of developing approximate formulae. The first method, which is based on the binomial expansion, is discussed in [208,645,827,873] and is of little use. More useful is the method based on partial differentiation of the function involved. This was studied in great detail by Tukey in [829–831]. The basic idea is to use a truncated Taylor series expansion of the function in question in order to linearise any non-linearities about expected values. Tukey says that the results obtained with only second order expansions are surprisingly accurate. Generally the approximate formulae are more accurate if the random variables have small coefficients of variation. The formulae are known in metrology as the general error propagation laws [22]. The formulae for an arbitrary function of random variables are given by Hahn and Shapiro [352, p.252] as follows. Let $Z = h(X_1, \ldots, X_n)$ be a function of n independent random variables. Then

$$\mu_Z \;\approx\; h(\mu_{X_1}, \ldots, \mu_{X_n}) \;+\; \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2 h}{\partial X_i^2} \sigma_{X_i}^2 \tag{2.2.1}$$

and

$$\sigma_Z^2 \;\approx\; \sum_{i=1}^{n} \left(\frac{\partial h}{\partial X_i} \right)^2 \sigma_{X_i}^2 \;+\; \sum_{i=1}^{n} \left(\frac{\partial h}{\partial X_i} \right) \left(\frac{\partial^2 h}{\partial X_i^2} \right) \mu_{X_i}^{(3)}, \tag{2.2.2}$$

where as usual we write σ_X^2 for $\mu_X^{(2)}$ and the partial derivatives are evaluated at their expected values. There are related but more complex formulae in terms of cross-moments for cases when the X_i are not independent. Note the following special cases. If $Z = h(X_1, X_2) = X_1 \pm X_2$,

$$\mu_Z = \mu_{X_1} \pm \mu_{X_2}, \qquad \sigma_Z^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2. \tag{2.2.3}$$

If $Z = h(X_1, X_2) = X_1 X_2$,

$$\mu_Z \approx \mu_{X_1}\mu_{X_2}, \qquad \sigma_Z^2 \approx \mu_{X_1}^2\sigma_{X_2}^2 + \mu_{X_2}^2\sigma_{X_1}^2 + \sigma_{X_1}^2\sigma_{X_2}^2. \tag{2.2.4}$$

If $Z = h(X_1, X_2) = X_1/X_2$,

$$\mu_Z \approx \frac{\mu_{X_1}}{\mu_{X_2}} + \frac{\mu_{X_1}\sigma_{X_2}^2}{4\mu_{X_2}^3}, \qquad \sigma_Z^2 \approx \frac{\sigma_{X_1}^2}{\mu_{X_2}^2} + \frac{\mu_{X_1}^2\sigma_{X_2}^2}{\mu_{X_2}^4} - \frac{\mu_{X_1}^2\mu_{X_2}^{(3)}}{2\mu_{X_2}^5}. \tag{2.2.5}$$

There are many possible variations on these formulae. For example, the derivation of an expression for the variance of products was the subject of the three papers [87, 331,332].

Uses of These Approximate Formulae

The above approximate formulae formed the basis of Metropolis's *significance arithmetic* [573,574]. The name comes from his use of the relationship

$$s_X = \left\lfloor \log_2 \left(\frac{|\mu_X|}{\sigma_X} + 0.5 \right) \right\rfloor$$

in order to determine the number of significant digits s_X in a calculated result. Metropolis stated that

> an important observation is that (2.2.3–2.2.5) do not depend on the detailed structure of the distribution function associated with each operand, apart from the natural assumption that the first and second moments exist [573, p.183].

He also noted that the formulae are only usable when the coefficients of variation are small.

Although the above approximate formulae are not entirely suitable for a completely general "probabilistic arithmetic", they are very useful in specific applications and have been used widely. See the examples in [405,808]. Moment based methods have found wide application in power systems analysis. The use of cumulants (which are directly related to moments [448]) has become popular in power systems analysis [131,542,677,783]. Representation of distributions by their cumulants has the advantage that sum convolutions can be calculated by the simple pointwise addition of cumulants. The resulting distribution is then calculated by using a Gram-Charlier expansion (see section 2.3.5 below). Whilst this method may seem elegant, there are many problems, even within the restricted application area envisaged for it by its developers. Apart from a number of specific problems caused by the area of application (which would be encountered by all methods of probabilistic arithmetic) there is the difficulty of representing distributions that are far from normal. In general there is an unknown approximation error. In some cases the fitted distribution can differ considerably from the measured distributions, especially in the tails [542, figure 7].

A moment based method has also been suggested by Petković [646], who developed an idea he calls "probable intervals." These are defined, for some random variable X with density f_X having finite support $[a, b]$, as the ordered pair $(\mu_X, q\sigma_X^2)$. The parameter $q > 0$ is chosen such that $q\sigma_X^2 < (b-a)/2$. Petković showed that if f_X is continuous and unimodal, then $\sigma_X^2 < r^2 + 2r\mu_X$, $\sigma_X^2 < 2r^3 f_X(\mu_X)/3$ and $\sigma_X^2 < r^2/3$, where $r = (b-a)/2$. He suggests that these results can be used in order to choose q. Once q is chosen, then the probable intervals can be combined using interval arithmetic (see section 2.5, especially subsection 2.5.3).

2.2.3 Non-Linear Transformations

Another method for calculating distributions of functions of random variables is to use a nonlinear transformation to convert some arbitrary distribution into, say, a normal distribution. The special properties of the normal distribution can then be used to determine the distribution of the appropriate function of the transformed random variables. The desired result is then calculated by using an inverse transformation. Transformation to normality was originally used as an aid to quadrature [46]. It has since been widely used in reliability assessment, although not in a very rigorous manner [711, p.*xxi*].

A recent application of the transformation idea is reported by Petrina *et al* in [647]. Using the probability integral transform to convert a random variable X with distribution function F_X into a random variable Y with distribution function F_Y (*i.e.* $Y = F_Y^{-1}(F_X(X))$) as their starting point, Petrina *et al.* proceed as follows. Hastings' approximation [363] to the normal distribution function is used to evaluate F_Y. They then determine an approximate overall transforming function by fitting a low degree polynomial to a discrete mapping from X to Y determined by using sample values of X. Once X has been transformed to Y (and in practice there are several Xs and several Ys), the original problem can be solved in terms of normal random variables. For the application considered by Petrina *et al*, a third degree polynomial approximation appears to give quite accurate results. The actual details of this approach are more intricate than indicated here. Whilst the technique appears quite useful in a number of circumstances, there seems little hope of developing a *general* probabilistic arithmetic in terms of it. (The idea of using polynomial transformations to normality is also discussed in [448, section 6.25].)

2.2.4 Direct Sampling of the Densities for Sum and Difference Convolutions

Allan *et al* [19], in an examination of probabilistic power systems analysis, considered the distribution of the sum of n_1 independent continuous Gaussian distributed random variables added to the sum of n_2 discrete random variables with binomial distributions. They used the results mentioned earlier for the sum of Gaussian random variables and found the distribution of the sum of the discrete random variables by numerically convolving their distributions using discrete convolution algorithms. The combination of these two results is a convolution of a Gaussian distribution

with a series of Dirac delta distributions. This is simply a superposition of Gaussian distributions.

The evaluation of the discrete convolution arising from the addition (or subtraction) of discrete random variables can also be accomplished by means of fast Fourier transform algorithms [20]. Indeed, any of the wide variety of fast discrete convolution algorithms can be used [83,625]. These fast algorithms are not applicable outside the case where the random variables can only take values on an equispaced grid. A trivial example of how this technique is used is given in [461].

The case of continuous convolutions is more difficult. A naive approach would be to approximate the continuous density $f(x)$ by the sampled discrete representation $f(kT)$, where T is the sample spacing. This is the approach used by Ackroyd [2–4]. By simply sampling a continuous density with a finite number of samples, approximations have to be made. There are two types of approximation errors. The first, which we call sampling error, is due to the density not being bandlimited, and thus the conditions under which the sampling theorem holds [375] are not satisfied [411]. The second form of approximation error is caused by the requirement of using only a finite number of samples. This makes the exact repesentation of densities with infinite (or semi-infinite) support impossible. We call this truncation error. An analysis of sampling and truncation errors for general functions (not necessarily probability densities) is given by Papoulis [633]. An analysis with particular reference to probability densities is given by Widrow [864] in his consideration of amplitude quantized sampled data systems. As well as discussing the well known Shannon sampling theorem [142,411,780], Widrow shows that if the probability density is bandlimited to W (that is if the support of the characteristic function or Fourier transform is contained within $[-W, W]$), then the set of samples spaced $1/W$ apart allows recovery of all the moments (but not the density, which requires samples only $1/(2W)$ apart).

2.2.5 The Skinner/Ackroyd Method

While it is possible, as mentioned above, to derive estimates for the error incurred in sampling the distribution function, it is preferable to provide strict upper and lower bounds on the error automatically as the calculation is performed. This has been done by Ackroyd and Kanyangarara [6]. They modified techniques presented in [2–4] by using an idea originally proposed by Skinner [758]. They sample the cumulative distribution rather than the probability density. The significance of this is that upper and lower bounds on the error caused by the sampling can be readily derived.

Given a cumulative distribution F_X of a random variable X, two discrete approximations are formed which are lower and upper bounds on F_X:

$$\underline{F}_X(x) \leq F_X(x) \leq \overline{F}_X(x) \qquad \forall x \in \Re.$$

The bounds $\underline{F}_X$ and $\overline{F}_X$ are defined by

$$\underline{F}_X(x) = F_X(kT) \qquad kT \leq x < (k+1)T$$

Figure 2.1: Illustration of Ackroyd's method of discretising a continuous probability distribution function. The solid line is $F(x)$, and the two dashed lines are $\underline{F}(x)$ and $\overline{F}(x)$ respectively.

and

$$\overline{F}_X(x) = F_X(kT) \qquad (k-1)T \leq x < kT$$

and can be understood by consideration of figure 2.1. The corresponding probability densities $\underline{f}_X$ and $\overline{f}_X$ are obtained by differencing and are related by

$$\underline{f}_X(kT) = \overline{f}_X((k-1)T).$$

Denoting n-fold convolution by a superscript (n), Ackroyd and Kanyangarara [6] showed that

$$\underline{F}_X^{(n)}(x) \leq F_X^{(n)}(x) \leq \overline{F}_X^{(n)}(x)$$

for all $x \in \Re$. For the special case of sum-convolutions it is possible to use the relationship

$$\overline{f}_X^{(n)}(kT) = \underline{f}_X^{(n)}((k+n)T)$$

in order to speed the calculation of these bounds by calculating

$$\sum_{m=-\infty}^{k} \underline{f}_X^{(n)}(mT) \;\leq\; F_X^{(n)}(x) \;\leq\; \sum_{m=-\infty}^{k+n} \underline{f}_X(mT) \qquad kT \leq x \leq (k+1)T.$$

The tightness of the bounds depends upon the size of T and the shape of F_X. There is no consideration in this scheme of the truncation error caused by representing a distribution with infinite support, although this error can be made arbitrarily small by increasing the number of samples. In chapter 3 we develop a method similar

to this where the *quantiles* F_X^{-1} are uniformly sampled and which has considerable advantages over the method described here.

Of all the techniques based on converting a continuous convolution to a discrete convolution by sampling, those which can be performed efficiently with fast Fourier transform algorithms [83,625] use uniform (equi-spaced) sampling. A different method of sampling which has been applied to the numerical calculation of characteristic functions is given by Jones and Lotwick [415]. This has been applied [416] to a method of non-parametric density estimation presented by Silverman [755, 756]. Their method reduces the error incurred by the sampling process by using a different method of assigning the sample values. We note that for some distributions non-uniform sampling [149] is better, although the practical value of this requires further investigation.

2.2.6 Spline Based Methods

Yet another approach to computing sum convolutions of probability densities was introduced by Cléroux and McConalogue [155,557]. Their idea is based on the piecewise representation of the cumulative distribution function of always positive random variables by cubic splines. Cubic splines are used as they have well behaved first derivatives (the probability density). In [557] a FORTRAN program is presented which can be used to approximate convolutions of densities that are bounded, analytic, and have support only on the positive real line. The algorithm appears to give accurate results, although estimates of the errors have not been determined: "a useful error analysis is not practicable" [155, p.1145]. The distribution function is represented by m equally spaced samples supplemented by m spline coefficients. Approximations to convolutions of distributions so represented are obtained in terms of the representing values. The details of this are quite messy and are omitted here.

One of the restrictions on the class of distributions to which the technique can be applied is removed in a generalisation presented in [558]. This generalisation overcomes problems with cubic spline approximations to infinite singularities, or where the function being approximated and all its derivatives vanish at the origin. A discussion of the original method, this generalisation, and a comparison with three other methods which are often used in solving renewal equations (the application for which this technique was originally developed) is given by Baxter in [61].

Nevertheless this generalised technique is still quite restricted, especially since it deals only with sum convolutions and because we do not know how accurate the final results are. The use of splines in conjunction with the histogram method is mentioned briefly in section 2.4.3 below. A relationship between splines and convolutions was also the subject of Sakai's paper [703].

2.3 Laguerre Transforms and Other Orthonormal Transforms

Orthonormal expansions are logical candidates for representing probability distributions so that convolutions can be calculated numerically. In this section we will look at one particular orthonormal expansion that has been widely studied in this regard and is based on Laguerre polynomials. Laguerre polynomials have been used in a number of areas such as signal processing and system identification as well as being used as a general means for representing continuous functions on a digital computer [145,152,153,272,366,401,402,460,485,507,590,591,605,624,884]. The Laguerre transform we examine was developed by Keilson and others [435,436] for calculating the sum convolutions (and other operations) of probability densities. The original motivation was to push back the "dreaded Laplacian curtain" [434, p.179]. (This is the name given to the fact that many results in queueing theory can only be expressed in terms of Laplace transforms.) The technique has been used successfully in a number of applications. In section 2.3.1 below we will briefly outline the technique and show how it can be used to calculate the sum and difference convolutions. No work has been presented to date on using Keilson's Laguerre transform to calculate product and quotient convolutions. Accordingly, in sections 2.3.2–2.3.4 we examine three possible approaches to achieve this. We obtain a number of new results but ultimately find that the methods appear intractable. In section 2.3.5 we briefly look at some other Laguerre transforms and consider their convolution relations. Our conclusions on the suitability of orthonormal expansions (particularly Laguerre transforms) as a method of representing probability densities suitable for numerically calculating convolutions are given in section 2.3.6.

Orthonormal transforms other than Laguerre transforms have been used widely in other areas. We now briefly consider some which might be useful for calculating convolutions. Hermite polynomials form an orthonormal set on $\Re$ with weight function $\exp(-x^2)$ [477]. They have been used to define Hermite transforms [190, 191,310]. Debnath [191] proved a complicated convolution formula for the Hermite transform of odd order. The convolution in question has no relation to the sum or product convolutions we are interested in and is thus of no use to us. The convolution structure of orthogonal transforms based on Jacobi polynomials (orthogonal on the interval [-1,1]) has also been investigated [39,270,292,474], but these too are of no use to us. (The convolutions are not of the form (2.1.9–2.1.12).) Other orthonormal systems such as the one recently discussed by Christov [146] are also inapplicable to our area of interest. Nonorthogonal expansions (which have some of the desirable features of orthogonal expansions without the "undesirable features") are also possible candidates for calculating convolutions. See the consideration of the Mellin transform and product convolution on pages 1281–1282 of [184].

2.3.1 The Laguerre Transform Method

Definitions

The Laguerre transform method is based on the use of the associated Laguerre functions

$$\ell_n(x) = e^{-x/2} L_n(x) \tag{2.3.1}$$

which provide an orthonormal basis in $L^2(0,\infty)$. Here $L_n(x)$ is the Laguerre polynomial of degree n which is defined by the Rodrigues' formula [810]

$$L_n(x) = \frac{1}{n!}\, e^x \left(\frac{d}{dx}\right)^n (x^n e^{-x}).$$

The polynomial $L_n(x)$ has the explicit form

$$L_n(x) = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} x^k. \tag{2.3.2}$$

An extended set of associated Laguerre functions, which allows the representation of the probability densities of random variables that are not always positive, can be defined by

$$h_n(x) = \begin{cases} \ell_n(x)U(x) & n \geq 0 \\ -\ell_n(-x)U(-x) & n < 0 \end{cases} \tag{2.3.3}$$

where $U(x) = 1$ for $x \geq 0$ and $U(x) = 0$ for $x < 0$. The set $\{h_n(x)\}_{-\infty}^{\infty}$ is an orthonormal basis of $L^2(-\infty,\infty)$. Thus for any $f \in L^2(-\infty,\infty)$ there is a unique representation

$$f(x) = \sum_{n=-\infty}^{\infty} f_n^{\dagger} h_n(x), \tag{2.3.4}$$

where the equality is in the L^2-sense. The coefficients $\{f_n^{\dagger}\}$ are known as the *Laguerre dagger coefficients* and are given by the Laguerre transform

$$f_n^{\dagger} = \int_{-\infty}^{\infty} h_n(x) f(x)\, dx. \tag{2.3.5}$$

If $\sum_n |f_n^{\dagger}| < \infty$, pointwise convergence is guaranteed for all $x \in \Re \setminus \{0\}$. The inverse transform (2.3.4) is unique if f is continuous.

Calculation of Sum and Difference Convolutions

In order to see how convolutions are calculable with Laguerre transforms we define the two generating functions

$$T_u^{\dagger}(f) = \sum_{n=-\infty}^{\infty} f_n^{\#} u^n$$

and

$$T_u^{\#}(f) = \sum_{n=-\infty}^{\infty} f_n^{\#} u^n = (1-u) T_u^{\dagger}(f).$$

The *Laguerre sharp coefficients* $\{f_n^{\#}\}_{-\infty}^{\infty}$ are related to the dagger coefficients by

$$f_n^{\#} = f_n^{\dagger} - f_{n-1}^{\dagger} = \int_{-\infty}^{\infty} f(x)\Delta h_n(x)\,dx. \tag{2.3.6}$$

where $\Delta h_n(x) = h_n(x) - h_{n-1}(x)$. The inverse relation is

$$f_n^{\dagger} = \lim_{k\to-\infty} \sum_{j=k}^{n} f_j^{\#} = -\lim_{k\to\infty} \sum_{j=n+1}^{k} f_j^{\#}.$$

Recalling our notation $\mathcal{L}_s(f) = \int_{-\infty}^{\infty} e^{-sx} f(x)\,dx$ for the Laplace transform of f, then it can be shown [791] that

$$\begin{aligned} \mathcal{L}_s(h_n) &= \int_{-\infty}^{\infty} e^{-sx} h_n(x)\,dx \\ &= \frac{1}{s+\frac{1}{2}} \left(\frac{s-\frac{1}{2}}{s+\frac{1}{2}}\right)^n \qquad -\infty < n < \infty \quad Re(s) \in [-\tfrac{1}{2}, \tfrac{1}{2}]. \end{aligned} \tag{2.3.7}$$

Equation (2.3.7) gives a relationship between the sharp generating function and the bilateral Laplace transform:

$$T_u^{\#}(f) = \mathcal{L}_{\left(\frac{1}{2}\frac{1+u}{1-u}\right)}(f).$$

Using this relationship, the sum convolution

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x-y)g(y)\,dy$$

can be calculated as follows. The convolution theorem for Laplace transforms states that $\mathcal{L}_s(f*g) = \mathcal{L}_s(f)\mathcal{L}_s(g)$. Therefore $T_u^{\#}(f*g) = T_u^{\#}(f)T_u^{\#}(g)$. This is equivalent to the discrete convolution

$$(f * g)_n^{\#} = \sum_{m=-\infty}^{\infty} f_{n-m}^{\#} g_m^{\#} \tag{2.3.8}$$

which can be readily calculated on a computer. Difference convolutions can be calculated by simply setting $g'(x) = g(-x)$ (by swapping the roles of $g_n^{\dagger}$ and $g_{-n}^{\dagger}$ for $n = 1, 2, \ldots$) and then calculating a sum convolution. In practice it is necessary to truncate the series $\{f_n^{\dagger}\}$ and $\{g_m^{\dagger}\}$ to a finite length. This will introduce errors, but they can be analysed and controlled.

Extensions and Applications

The assumption that f, the density, is in L^2 is not always satisfied. If $f \notin L^2$ then the Laguerre dagger transform defined by (2.3.5) does not exist. However one can define a *sharp transform* which does exist if $f \notin L^2$ provided that $f \in L^1$. It is equivalent to the dagger transform when $f \in L^2$ [441].

The Laguerre transform method has been successfully applied to a number of problems in applied probability that require the evaluation of sum convolutions.

Some of the applications include a study of the approach to normality in the central limit theorems [436,790], and the evaluation of renewal densities [790]. The Laguerre coefficients for a number of widely used distributions are given by Sumita in [791], although not all are given in a closed form. We note in passing that whilst it is possible to derive some formulae for the Laguerre transforms of H-function random variables by using the results in [45,549,740], these do not appear to be of any value because they are very complicated. Further details on the Laguerre transform and its applications can be found in [437–441,546,790–799].

2.3.2 Distributions of Products and Quotients using the Laguerre and Mellin Transforms

We will now consider the possibility of using the Laguerre transform for evaluating the density of the quotient or the product of two random variables in terms of the Laguerre coefficients of the operands. Because of the success with which it has been used for sum convolutions, a method of using it for calculating product and quotient convolutions would make the technique rather more complete. We should expect that the determination of the quotient or product convolutions will be rather more difficult (compared to the sum or difference convolutions) because product and quotient are non-linear operations. Also, whereas the sum convolution has the nice property that the sum convolution of any two bounded continuous distributions is bounded and continuous, the same is not true of the product convolutions. For example, the product of two standardised Gaussian random variables has a probability density $f(x) = (1/\pi)K_0(x)$, where K_0 is the Bessel function of the second kind of purely imaginary argument [35,171]. This has a singularity at the origin.

There does not appear to be any work on using the Laguerre transform to calculate product and quotient convolutions in the literature. Sumita and Kijima have considered the simpler problem of finding the Laguerre transform of the product of two functions in terms of the Laguerre coefficients of the functions [797]. That is, given the Laguerre coefficients of $f(x)$ and $g(x)$, they determine the Laguerre coefficients of $f(x)g(x)$.

Drawing on an analogy with the method of determining the relationship for the Laguerre transform coefficients for the sum (and hence difference) of two random variables by considering the Laplace transform of the associated Laguerre functions, we will now examine the (unilateral) Mellin transform of $\ell_n(x)$. If we write $\ell_n(x)$ as

$$\ell_n(x) = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} \left(e^{-x/2} x^k \right),$$

then we can determine the Mellin transform termwise. Using equations 6.3.1 and 6.1.3 of [257] we have

$$\mathcal{M}_s \left[e^{-ax} x^k \right] = a^{-(s+k)} \Gamma(s+k). \tag{2.3.9}$$

Setting $a = \frac{1}{2}$ gives

$$\mathcal{M}_s[\ell_n] = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} \left[2^{s+k}\Gamma(s+k)\right]. \tag{2.3.10}$$

Noting that $\Gamma(s+k) = \Gamma(s)\prod_{j=1}^{k}(s+j)$ we obtain

$$\mathcal{M}_s[\ell_n] = e^{cs}\Gamma(s) \sum_{k=0}^{n} \binom{n}{k} \frac{(-2)^k}{k!} \prod_{j=1}^{k}(s+j), \tag{2.3.11}$$

where $c = \ln 2$. Unfortunately this is nowhere near as simple as (2.3.7) (the Laplace transform of $h_n(x)$), and thus there is no apparent simple relation between the Laguerre coefficients for the product or quotient of two random variables. In section 2.3.5 we will briefly discuss some other relationships between Laguerre polynomials and Mellin transforms that have appeared in the literature.

2.3.3 Distribution of Products and Quotients Calculated with the Laguerre Transform Directly

Another possible approach to determining the Laguerre coefficients of product and quotient convolutions is to examine the convolutions directly. We consider the quotient here, and we let $Z = X/Y$ be the quotient of two random variables. The functions f, g, and h are the densities of X, Y and Z respectively. We assume X and Y (and hence Z) are always positive (thus $f(x) = g(x) = h(x) = 0$ for all $x < 0$). This does not sacrifice any generality as we can always determine the quotient of two random variables that are not always positive by considering the positive and negative parts separately and then combining the results in a mixing operation.

The General Approach

The convolution equation we are studying is

$$\begin{aligned} h(y) &= \int_{-\infty}^{\infty} |x| f(xy) g(x)\, dx \\ &= \int_{0}^{\infty} x f(xy) g(x)\, dx \qquad x, y \in [0, \infty). \end{aligned}$$

Let $h(y) = \sum_{k=0}^{\infty} h_k^\dagger \ell_k(y)$, $f(x) = \sum_{n=0}^{\infty} f_n^\dagger \ell_n(x)$ and $g(x) = \sum_{m=0}^{\infty} g_m^\dagger \ell_m(x)$, where $\{h_k^\dagger\}$, $\{f_n^\dagger\}$, and $\{g_m^\dagger\}$ are the respective one-sided Laguerre dagger coefficients. Then

$$h(y) = \int_0^\infty x \left[\sum_{n=0}^{\infty} f_n^\dagger e^{-xy/2} L_n(xy)\right] \left[\sum_{m=0}^{\infty} g_m^\dagger e^{-x/2} L_m(x)\right] dx. \tag{2.3.12}$$

If $\{f_n^\dagger\}$ and $\{g_m^\dagger\} \in \ell^1$, then the Laguerre expansion converges pointwise uniformly almost everywhere (see theorem 2.3.1 in the following subsection). Assuming this

we can write

$$h(y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} f_n^\dagger g_m^\dagger \int_{x=0}^{\infty} x e^{-x(y+1)/2} L_n(xy) L_m(x)\, dx$$

and we need to evaluate the integral

$$I = \int_{x=0}^{\infty} x\, e^{-x(y+1)/2} L_n(xy) L_m(x)\, dx.$$

Evaluation of the Integral

Gradshteyn and Ryzhik [336, eq. 7.4144] give

$$\begin{aligned} &\int_0^\infty e^{-bx} x^\alpha L_n^{(\alpha)}(\lambda x) L_m^{(\alpha)}(\mu x)\, dx \qquad (2.3.13)\\ &= \frac{\Gamma(m+n+\alpha+1)}{m!n!} \frac{(b-\lambda)^n (b-\mu)^m}{b^{m+n+\alpha+1}} F\left[-m, -n; -m-n-\alpha; \frac{b(b-\lambda-\mu)}{(b-\lambda)(b-\mu)}\right] \end{aligned}$$

for $Re(\alpha) > -1$, $Re(b) > 0$, where $F[\alpha, \beta; \gamma; z]$ is the hypergeometric function defined by

$$\begin{aligned} F(\alpha, \beta; \gamma; z) \;=\; & 1 + \frac{\alpha\beta}{\gamma \times 1} z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1) \times 1 \times 2} z^2 + \\ & \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{\gamma(\gamma+1)(\gamma+2) \times 1 \times 2 \times 3} z^3 + \cdots. \end{aligned}$$

The generalised Laguerre polynomial $L_n^{(\alpha)}(x)$ is defined by

$$L_n^{(\alpha)}(x) = \sum_{m=0}^{n} (-1)^m \binom{n+\alpha}{n-m} \frac{x^m}{m!}. \qquad (2.3.14)$$

Observing that $L_n^{(\alpha-1)}(x) = L_n^{(\alpha)}(x) - L_{n-1}^{(\alpha)}(x)$ [336, eq. 8.9714], we can write the integrand of I as

$$x e^{-x(y+1)/2} L_n(xy) L_m(x) = x e^{-x(y+1)/2} \left[L_n^{(1)}(xy) - L_{n-1}^{(1)}(xy)\right] \left[L_m^{(1)}(x) - L_{m-1}^{(1)}(x)\right]. \qquad (2.3.15)$$

If we set $\alpha = 1$, $b = (y+1)/2$, $\lambda = y$, and $\mu = 1$ we have (from equation 2.3.13)

$$\begin{aligned} B_{m,n}(y) \;&\triangleq\; \int_{x=0}^{\infty} x e^{-x(y+1)/2} L_n^{(1)}(xy) L_m^{(1)}(x)\, dx \\ &= \frac{(m+n+1)!}{m!n!} \frac{(-1)^n 4 (y-1)^{m+n}}{(y+1)^{m+n+2}}\; F\left[-m, -n; -m-n-1; \left(\frac{y+1}{y-1}\right)^2\right]. \end{aligned}$$

The hypergeometric function $F(\alpha, \beta; \gamma; z)$ terminates if α or β are negative integers. Since $\alpha = -m$, $\beta = -n$, and $\gamma = -m-n-1$, we can write

$$F(-m, -n; -m-n-1; z) = \sum_{i=0}^{\min(m,n)} \left[\prod_{j=0}^{i-1} \frac{(j-m)(j-n)}{(j-m-n-1)}\right] \frac{z^i}{i}. \qquad (2.3.16)$$

We also have (when $\max(i) = \min(n,m)$)

$$\begin{aligned}
&\prod_{j=0}^{i-1} \frac{(j-m)(j-n)}{(j-m-n-1)} \\
= \quad & \frac{-m(-m+1)\cdots(-m+i-1)(-n)(-n+1)\cdots(-n+i-1)}{(-m-n-1)(-m-n)(-m-n+1)\cdots(-m-n-1+i-1)} \\
= \quad & \frac{(-1)^i m(m-1)\cdots(m-i+1)(-1)^i n(n-1)\cdots(n-i+1)}{(-1)^i(m+n+1)(m+n)(m+n-1)\cdots(m+n+2-i)} \\
= \quad & \frac{(-1)^i m!\; n!\; (m+n+1-i)!}{(m-i)!\; (n-i)!\; (m+n+1)!}.
\end{aligned}$$

Therefore

$$\begin{aligned}
B_{m,n}(y) &= \frac{(m+n+1)!}{m!n!}\frac{m!n!}{(m+n+1)!}\frac{(-1)^n 4(y-1)^{m+n}}{(y+1)^{m+n+2}} \times \\
&\qquad \sum_{i=0}^{\min(m,n)} \frac{(m+n+1-i)!}{(m-i)!(n-i)!i!}\left(\frac{y+1}{y-1}\right)^{2i} \\
&= \frac{(-1)^n 4}{(y+1)^2} \sum_{i=0}^{\min(m,n)} \frac{(m+n+1-i)!}{(m-i)!(n-i)!i!}\left(\frac{y-1}{y+1}\right)^{m+n-2i}. \qquad (2.3.17)
\end{aligned}$$

Recalling equation (2.3.15) we have

$$I = B_{m,n}(y) + B_{m-1,n-1}(y) - B_{m-1,n}(y) - B_{m,n-1}(y)$$

and therefore

$$h(y) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} f_n^\dagger g_m^\dagger \left[B_{m,n}(y) + B_{m-1,n-1}(y) - B_{m-1,n}(y) - B_{m,n-1}(y)\right]. \qquad (2.3.18)$$

The Result

Equation 2.3.18 gives h in terms of $\{f_n^\dagger\}$ and $\{g_m^\dagger\}$. However what we really want is to determine $\{h_k^\dagger\}$ directly in terms of $\{f_n^\dagger\}$ and $\{g_m^\dagger\}$. We have

$$\begin{aligned}
h_k^\dagger &= \int_{y=0}^{\infty} h(y)\ell_k(y)\,dy \qquad (2.3.19) \\
&= \int_{y=0}^{\infty} \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} f_n^\dagger g_m^\dagger \left[B_{m,n}(y) + B_{m-1,n-1}(y) - B_{m-1,n}(y) - B_{m,n-1}(y)\right]\ell_k(y)\,dy,
\end{aligned}$$

where $B_{m,n}(y)$ is given by (2.3.17). Even if the inner series in (2.3.19) is uniformly convergent (a fact which is not readily apparent), this is obviously going to be a rather messy expression for $h_k^\dagger$: a complicated sum of integrals of the form $\int_0^\infty \frac{y^r e^{-y/2}}{(y+1)^p}\,dy$. So, although Gradshteyn and Ryzhik [336, eq. 3.3838] give an expression for this integral (in terms of Whittaker's function), this does not result in

a simple expression for $h_k^\dagger$. Difficulties remain even if we numerically compute the integral beforehand. This can be seen by considering

$$h_k^\dagger = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} f_n^\dagger g_m^\dagger K_{m,n,k}, \tag{2.3.20}$$

where $K_{m,n,k}$ is the precomputed integral. If we truncate the Laguerre series by using only 100 coefficients, then in order to calculate $\{h_k^\dagger\}$ for $k = 0, \ldots, 100$ in excess of 10^6 operations would be required.

A derivation of the Laguerre coefficients of the *product* of two random variables gives a similar result to that obtained above (a complicated sum of Whittaker functions) and is not considered further.

The fact that the results we have obtained are so complicated is surprising when we consider Feldheim's result [261] that the Laguerre polynomials are the only orthogonal polynomials $p_n(y)$ with a multiplicative theorem of the form

$$\lambda^n p_n(y/\lambda) = \sum_{j=0}^{n} A_{nj} (\lambda - 1)^{n-j} p_j(y) \tag{2.3.21}$$

where A_{nj} are constants. The significance of Feldheim's result can be seen by comparing (2.3.21) with the product convolution (see equation 2.3.50) where $f(x/y)$ is a term in the integrand. Nevertheless this is not as simple as [387, equation 22.13.14]

$$\int_0^x L_m(t) L_n(x-t)\, dt = \int_0^x L_{m+n}(t)\, dt = L_{m+n}(x) - L_{m+n+1}(x)$$

which was the original motiovation for Keilson's Laguerre transform [434].

Other expansions for products of Laguerre polynomials have appeared in the literature [44,255,776], but these are all in terms of Laguerre polynomials $L_n^{(\alpha)}(x)$ of different orders α. Products of other orthogonal polynomials are discussed in [36,290,291,293]. Niukkanen [622] has presented a very general result giving the product of two Laguerre polynomials in terms of a series of Laguerre polynomials. These are also presented in terms of Laguerre polynomials of different orders α. The coefficients for the Laguerre polynomials in his expansion are very complicated expressions in terms of his generalised hypergeometric series [621] and do not appear to be of any use here.

2.3.4 Distributions of Products and Quotients using the Laguerre Transform and Logarithmic/Exponential Transformation of Variables

Instead of determining the distributions of products and quotients directly, it may be possible to use logarithmic and exponential transformations. It can be shown that if X is an almost surely positive random variable with a probability density $f(x)$, then the random variable $Y = \log X$ has a density $g(y) = e^y f(e^y)$. The inverse transformation is $f(x) = (1/x) g(\log x)$. If we can determine the Laguerre coefficients

for g in terms of the coefficients for f, then we can calculate the probability density of products and quotients of random variables in terms of the Laguerre coefficients of the operands by calculating the effect of the logarithmic transformation, performing a sum or difference convolution, and then transforming back again. In this section we will derive a formula for the Laguerre coefficients of the logarithm of a random variable with a given density in terms of the Laguerre coefficients of the density. It turns out to be surprisingly difficult to do this. As a consequence, this subsection is rather longer and more intricate than the others.

The General Approach

Let $f(x) = \sum_{n=0}^{\infty} f_n^\dagger \ell_n(x)$, where $f_n^\dagger = \int_0^\infty f(x)\ell_n(x)\,dx$; and let $g(y) = e^y f(e^y) = \sum_{n=0}^{\infty} g_n^\dagger \ell_n(y)$, where $g_n^\dagger = \int_0^\infty g(y)\ell_n(y)\,dy$. We require $\{g_n^\dagger\}$ in terms of $\{f_n^\dagger\}$. Obviously

$$g_n^\dagger = \int_{y=0}^{\infty} e^y f(e^y)\ell_n(y)dy. \tag{2.3.22}$$

Let $z = e^y$, $y = \log z$, and so $\frac{dz}{dy} = e^y$ and hence $dz = e^y dy$. Then

$$\begin{aligned} g_n^\dagger &= \int_{z=1}^{\infty} f(z)\ell_n(\log z)\,dz \\ &= \int_{z=1}^{\infty} \left[\sum_{m=0}^{\infty} f_m^\dagger e^{-z/2} \sum_{p=0}^{m} \binom{m}{p} \frac{(-1)^p}{p!} z^p\right] \left[e^{-(\log z)/2} \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} (\log z)^k\right] dz \\ &= \int_{z=1}^{\infty} \sum_{k=0}^{n} \sum_{m=0}^{\infty} \sum_{p=0}^{m} \binom{n}{k} \frac{(-1)^k f_m^\dagger (-1)^p}{k!\; p!} \binom{m}{p} e^{-z/2} (\log z)^k z^{p-\frac{1}{2}}\, dz. \end{aligned} \tag{2.3.23}$$

If the infinite series in (2.3.23) is uniformly convergent to an integrable function, then we can interchange the order of integration and summation and integrate termwise. Equation (2.3.23) is equivalent to

$$g_n^\dagger = \int_{z=1}^{\infty} \sum_{m=0}^{\infty} f_m^\dagger \ell_m(z)\ell_n(\log z)\,dz. \tag{2.3.24}$$

Sumita [791, page 67] gives the following theorem regarding the uniform convergence of the extended Laguerre expansion.

Theorem 2.3.1 *Let $f(x) \in L^2(-\infty, \infty)$ have an extended Laguerre expansion*

$$f(x) \approx \sum_{n=-\infty}^{\infty} f_n^\dagger h_n(x) \qquad -\infty < x < \infty,$$

and define the partial sum $S_N(x) = \sum_{n=-N}^{N} f_n^\dagger h_n(x)$. If $\{f_n^\dagger\} \in \ell^1$, then $S_N(x)$ converges to $f(x)$ pointwise uniformly almost everywhere as $N \to \infty$.

This theorem implies that if $\{f_n^\dagger\} \in \ell^1$, then $f(x) = \sum_{n=-\infty}^{\infty} f_n^\dagger h_n(x)$ is continuous for all $x \neq 0$ because when a sequence of continuous functions (such as $S_N(x)$) converges to f uniformly, then f is continuous. The class of functions f for which

the coefficients $\{f_n^\dagger\} \notin \ell^1$ has not been determined. However we do know that if f is not continuous for all $x \in (0,\infty)$ then $\{f_n^\dagger\} \notin \ell^1$ [435, page 326]. We will assume that $\{f_n^\dagger\} \in \ell^1$ from now on.

It is required that

$$S'_N(x) = \sum_{m=0}^{N} f_m^\dagger \ell_m(z)\ell_n(\log z) \tag{2.3.25}$$

converges uniformly to an integrable function for almost all $z \in [1,\infty)$ and for $n = 0, 1, \ldots$. This is true if $\{f_m^\dagger\} \in \ell^1$ because $S'_N(x)$ converges to $f(z)\ell_n(\log z)$ (which is continuous and bounded), and $S_N(z) = \sum_{m=0}^{N} f_m^\dagger \ell_m(z)$ converges uniformly to $f(z)$. Regarding the integrability condition, we observe that

$$\begin{aligned} \ell_n(\log z) &= e^{-(\log z)/2} \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} (\log z)^k \\ &= z^{-\frac{1}{2}} O((\log z)^n) \end{aligned} \tag{2.3.26}$$

which approaches 0 as $z \to \infty$ for any given n, and therefore $f(z)\ell_n(\log z) \to 0$ as $z \to \infty$ because $f(z) \to 0$ as $z \to \infty$. Now $f(z)$ is integrable on $[1,\infty)$ (it is a probability density) and $f(z) \geq 0$ for all $z \geq 0$. Also $\ell_n(\log z)$ is bounded for all $z > 0$. Thus $f(z)\ell_n(\log z)$ is integrable on $[1,\infty)$.

Evaluation of the Integral

We can now interchange the order of integration and summation in (2.3.23) giving

$$g_n^\dagger = \sum_{k=0}^{n} \sum_{m=0}^{\infty} \sum_{p=0}^{m} \binom{n}{k} \frac{(-1)^k}{k!} f_m^\dagger \binom{m}{p} \frac{(-1)^p}{p!} \int_{z=1}^{\infty} e^{-z/2} (\log z)^k z^{p-\frac{1}{2}}\, dz. \tag{2.3.27}$$

We now need to evaluate the integral

$$I = \int_1^\infty e^{-z/2} (\log z)^k z^{p-\frac{1}{2}}\, dz. \tag{2.3.28}$$

Gradshteyn and Ryzhik [336, eq. 4.3581] give

$$\begin{aligned} \int_1^\infty x^{\nu-1} e^{-\mu x} (\log x)^m\, dx &= \frac{1}{\mu} \frac{\partial^m}{\partial \nu^m} \left(\mu^{1-\nu} \Gamma(\mu, \nu) \right) \\ & \qquad Re(\mu) > 0,\ Re(\nu) > 0,\ m = 0, 1, \ldots, \end{aligned} \tag{2.3.29}$$

where $\Gamma(\alpha, x)$ is the incomplete gamma function defined by

$$\Gamma(\alpha, x) = \int_x^\infty e^{-t} t^{\alpha-1}\, dt.$$

If we set $m = k$, $\nu = p + \frac{1}{2}$, $(p \geq 0)$, and $\mu = \frac{1}{2}$ then (2.3.29) is equivalent to (2.3.28). Therefore

$$I = \frac{\partial^k}{\partial (p+\frac{1}{2})^k} \left((\tfrac{1}{2})^{1-(p+\frac{1}{2})} \Gamma(\tfrac{1}{2}, p+\tfrac{1}{2}) \right). \tag{2.3.30}$$

Let $p' = p + \frac{1}{2}$, then

$$I = \frac{\frac{1}{2}\partial^k}{\partial p'^k}\left(2^{p'}\Gamma(\tfrac{1}{2}, p')\right). \tag{2.3.31}$$

We require an expression for

$$\frac{\partial^k}{\partial x^k}\left(e^{cx}\Gamma(a,x)\right),$$

where $x = p'$ and $c = \ln 2$. Firstly consider

$$\frac{\partial}{\partial x}\left(e^{cx}\Gamma(a,x)\right).$$

Erdelyi *et al* [256, eq. 12, p. 135] give

$$\frac{\partial}{\partial x}\left(e^{x}\Gamma(a,x)\right) = (a-1)e^{x}\Gamma(a-1,x). \tag{2.3.32}$$

If we write

$$e^{cx}\Gamma(a,x) = e^{(c-1)x} \times e^{x}\Gamma(a,x)$$

and use the product rule along with (2.3.32) we obtain

$$\frac{\partial}{\partial x}\left(e^{cx}\Gamma(a,x)\right) = (c-1)e^{cx}\Gamma(a,x) + (a-1)e^{cx}\Gamma(a-1,x). \tag{2.3.33}$$

We actually require an expression for $\frac{\partial^k}{\partial x^k}\left[e^{cx}\Gamma(a,x)\right]$. There does not appear to be one in the literature. Thus we now prove the following theorem by induction.

Theorem 2.3.2

$$\frac{\partial^n}{\partial x^n}\left[e^{cx}\Gamma(a,x)\right] \;=\; \sum_{i=0}^{n}\binom{n}{i}(c-1)^{n-i}\left[\prod_{j=1}^{i}(a-j)\right]e^{cx}\Gamma(a-i,x). \tag{2.3.34}$$

PROOF. In order to make the proof less cumbersome, we define $G(a) = e^{cx}\Gamma(a,x)$ and $D = \frac{\partial}{\partial x}$. We have already seen (equation 2.3.33) that

$$D[G(a)] \;=\; (c-1)G(a) + (a-1)G(a-1). \tag{2.3.35}$$

The theorem to be proved is

$$D^n[G(a)] \;=\; \sum_{i=0}^{n}\binom{n}{i}(c-1)^{n-i}\left[\prod_{j=1}^{i}(a-j)\right]G(a-i).$$

Thus the induction hypothesis is

$$D\left[\sum_{i=0}^{n}\binom{n}{i}(c-1)^{n-i}\left[\prod_{j=1}^{i}(a-j)\right]G(a-i)\right] \tag{2.3.36}$$

$$= \sum_{i=0}^{n+1}\binom{n+1}{i}(c-1)^{n+1-i}\left[\prod_{j=1}^{i}(a-j)\right]G(a-i). \tag{2.3.37}$$

Firstly observe that for arbitrary coefficients $\{q_i\}$,

$$D\left[\sum_i q_i G(a-i)\right] \;=\; \sum_i q_i(c-1)G(a-i) + \sum_i (a-i-1)q_i G(a-i-1).$$

This fact along with equations 2.3.35 and 2.3.36 gives

$$\begin{aligned}
&\sum_{i=0}^{n}\binom{n}{i}(c-1)^{n-i}\left[\prod_{j=1}^{i}(a-j)\right]G(a-i)(c-1) \;+\\
&\qquad\sum_{i=0}^{n}\binom{n}{i}(c-1)^{n-i}\left[\prod_{j=1}^{i}(a-j)\right](a-i-1)G(a-i-1)\\
=\;&\sum_{i=0}^{n}\binom{n}{i}(c-1)^{n+1-i}\left[\prod_{j=1}^{i}(a-j)\right]G(a-i) \;+\\
&\qquad\sum_{i=0}^{n}\binom{n}{i}(c-1)^{n-i}\left[\prod_{j=1}^{i+1}(a-j)\right]G(a-i-1)\\
=\;&\sum_{i=1}^{n}\binom{n}{i}(c-1)^{n+1-i}\left[\prod_{j=1}^{i}(a-j)\right]G(a-i) \;+\\
&\qquad\sum_{i=0}^{n-1}\binom{n}{i}(c-1)^{n-i}\left[\prod_{j=1}^{i+1}(a-j)\right]G(a-i-1)+\\
&\qquad\binom{n}{0}(c-1)^{n+1}\left[\prod_{j=1}^{0}(a-j)\right]G(a-0) \;+\\
&\qquad\binom{n}{n}(c-1)^{0}\left[\prod_{j=1}^{n+1}(a-j)\right]G(a-n-1).
\end{aligned}$$

If we set $i=k-1$ in the second sum (so $k=i+1$) this can be written as

$$\begin{aligned}
&\sum_{i=1}^{n}\binom{n}{i}(c-1)^{n+1-i}\prod_{j=1}^{i}(a-j)G(a-i) \;+\\
&\qquad\sum_{k=1}^{n}\binom{n}{k-1}(c-1)^{n-(k-1)}\prod_{j=1}^{(k-1)+1}(a-j)G(a-(k-1)-1) \;+\\
&\qquad\binom{n}{0}(c-1)^{n+1}G(a) \;+\; \binom{n}{n}\prod_{j=1}^{n+1}(a-j)G(a-n-1)\\
=\;&\sum_{i=1}^{n}\left(G(a-i)(c-1)^{n+1-i}\prod_{j=1}^{i}(a-j) \;+\right.\\
&\qquad\left.\binom{n}{i-1}(c-1)^{n+1-i}\prod_{j=1}^{i}(a-j)G(a-i)\right) \;+\\
&\qquad(c-1)^{n+1}G(a) \;+\; \prod_{j=1}^{n+1}(a-j)G(a-n-1)
\end{aligned}$$

$$= \sum_{i=1}^{n} G(a-i)(c-1)^{n+1-i} \prod_{j=1}^{i}(a-j)\left[\binom{n}{i}\binom{n}{i-1}\right] +$$
$$(c-1)^{n+1}G(a) + \prod_{j=1}^{n+1}(a-j)G(a-n-1)$$
$$= \sum_{i=1}^{n} G(a-i)(c-1)^{n+1-i} \prod_{j=1}^{i}(a-j)\binom{n+1}{i} +$$
$$(c-1)^{n+1}G(a) + \prod_{j=1}^{n+1}(a-j)G(a-n-1)$$

which equals (2.3.37). ■

The Result

The result we require (a closed form for equation 2.3.31) is obtained by substituting into (2.3.34). We find that

$$\frac{\partial^k}{\partial p'^k}\left(2^{p'}\Gamma(\tfrac{1}{2},p')\right) = \sum_{i=0}^{k}\binom{k}{i}(\ln 2-1)^{k-i}\prod_{j=1}^{i}(\tfrac{1}{2}-j)2^{p+\frac{1}{2}}\Gamma(\tfrac{1}{2}-i,p+\tfrac{1}{2}). \quad (2.3.38)$$

Substituting (2.3.38) into (2.3.31) and (2.3.27) gives

$$g_n^\dagger = \sum_{k=0}^{n}\sum_{m=0}^{\infty}\sum_{p=0}^{n}\binom{n}{k}\frac{(-1)^k}{k!}f_m^\dagger\binom{m}{p}\frac{(-1)^p}{p!} \times \quad (2.3.39)$$
$$2^{p-\frac{1}{2}}\sum_{i=0}^{k}\binom{k}{i}(\ln 2-1)^{k-i}\prod_{j=1}^{i}(\tfrac{1}{2}-j)\Gamma(\tfrac{1}{2}-i,p+\tfrac{1}{2}).$$

This can be written as

$$g_n^\dagger = \sum_{k=0}^{n}\binom{n}{k}F(k) = \sum_{k=0}^{n-1}g_{n-1}^\dagger + F(n), \quad (2.3.40)$$

where $F(k)=\sum_{m=0}^{\infty} f_m^\dagger G(k,m)$ and $G(k,m)$ does not depend on $f_m^\dagger$:

$$G(k,m) = \frac{(-1)^k}{k!}\sum_{p=0}^{m}\binom{m}{p}\frac{(-1)^p}{p!}2^{p-\frac{1}{2}}\sum_{i=0}^{k}\binom{k}{i}(\ln 2-1)^{k-i}\prod_{j=1}^{i}(\tfrac{1}{2}-j)\Gamma(\tfrac{1}{2}-i,p+\tfrac{1}{2}). \quad (2.3.41)$$

Thus $G(k,m)$ could be tabulated for $k,m=0,1,\ldots,N$, where N is the length at which we truncate the infinite series of Laguerre dagger coefficients $\{f_m^\dagger\}$ and $\{g_n^\dagger\}$. These tabulated values could then be used with equation (2.3.40) and the given Laguerre dagger coefficients $\{f_m^\dagger\}$ to calculate the set $\{g_n^\dagger\}$.

There are problems with this scheme which reduce its practicability. Firstly, if N is of the order that has been used in most of the papers on the Laguerre transform technique cited earlier (100 to 500) then the table of values of $G(k,m)$

will contain 10000 to 250000 entries. Secondly, the values of $G(k,m)$ for some m and k appear to be very large and beyond the range of most modern digital computers which can usually only represent numbers up to about 10^{38}. This second point was discovered upon programming (2.3.41) using a special package of extended floating point arithmetic subroutines [136] (which is similar to the extended-range arithmetic used for calculating Legendre polynomials in [765] and the level-index system described in [154]) and a computational algorithm for the incomplete gamma function [295]. The expression (2.3.41) is unfortunately too complicated to derive rough asymptotic values for $G(k,m)$ to verify the program's correctness. We note that Sumita has encountered problems of excessive magnitudes in the calculation of the Laguerre coefficients of Gamma densities [791, p.188].

Reducing the Extent of the Laguerre Coefficients

The problems associated with (2.3.40) and (2.3.41) would be less severe if the Laguerre coefficients were of smaller extent (*i.e.* if N, the number of coefficients was smaller). Sumita [791] has shown that the extent of the coefficients depends heavily on the concentration about zero and the extent of the tails. Scaling and exponential weighting of the density to be represented can reduce this extent considerably. Scaling involves setting $g(x) = f(cx)$ for some positive constant c, and weighting involves setting $g(x) = e^{-\theta x} f(x)$ for some positive constant θ. Sumita says that "a general procedure to find [the Laguerre coefficients of $g(x)$] from [the coefficients of $f(x)$] has yet to be developed" [791, p.168]. The following theorem solves this problem.

Theorem 2.3.3 *Let* $g(x) = e^{-\theta x} f(cx)$, $\theta \geq 0$, $c > 0$, $f(x) = \sum_{n=0}^{\infty} f_n^\dagger \ell_n(x)$, *and* $g(x) = \sum_{m=0}^{\infty} g_m^\dagger \ell_m(x)$. *Then if* $\{f_n^\dagger\} \in \ell^1$,

$$g_m^\dagger = \sum_{n=0}^{\infty} f_n^\dagger \, 2(-1)^n \sum_{i=0}^{\min(m,n)} \frac{(-1)^i (m+n-i)! \, (c-1-2\theta)^{n-i} (c-1+2\theta)^{m-i}}{(m-i)! \, (n-i)! \, i! \, (c+1+2\theta)^{m+n+1-i} (c+1-2\theta)^{-i}}. \tag{2.3.42}$$

PROOF. We have

$$g_m^\dagger = \int_0^\infty g(x) \ell_m(x) \, dx = \int_0^\infty e^{-\theta x} \left[\sum_{n=0}^{\infty} f_n^\dagger \ell_n(cx) \right] \ell_m(x) \, dx. \tag{2.3.43}$$

If $\{f_n^\dagger\} \in \ell^1$, then the Laguerre expansion in the square brackets in (2.3.43) converges uniformly (see theorem 2.3.1) and so

$$g_m^\dagger = \sum_{n=0}^{\infty} f_n^\dagger \int_0^\infty e^{-\left(\frac{c+1+2\theta}{2}\right)x} L_n(cx) L_m(x) \, dx. \tag{2.3.44}$$

If we let $b = \frac{c+1+2\theta}{2}$, $\alpha = 0$, $\lambda = c$, $\mu = 1$ and use equation 2.3.13 we obtain

$$g_m^\dagger = \sum_{n=0}^{\infty} f_n^\dagger \frac{\Gamma(m+n+1)}{m! \, n!} \frac{\left(\frac{c+1+2\theta}{2} - c\right)^n \left(\frac{c+1+2\theta}{2} - 1\right)^m}{\left(\frac{c+1+2\theta}{2}\right)^{m+n+1}} \times$$

$$F\left[-m,-n;-m-n;\frac{\left(\frac{c+1+2\theta}{2}\right)\left(\frac{c+1+2\theta}{2}-c-1\right)}{\left(\frac{c+1+2\theta}{2}-c\right)\left(\frac{c+1+2\theta}{2}-1\right)}\right]$$

$$= \sum_{n=0}^{\infty} f_n^\dagger \frac{(m+n)!\,2(1-c+2\theta)^n(c-1+2\theta)^m}{m!\,n!\,(c+1+2\theta)^{m+n+1}} \times F\left[-m,-n;-m-n;\frac{(c+1+2\theta)(c+1-2\theta)}{(c-1+2\theta)(c-1-2\theta)}\right].$$

Substituting the expression for the terminating Hypergeometric series (2.3.16) gives (2.3.42). ■

Note the following special cases.

$$(\theta=0): \quad g_m^\dagger = \sum_{n=0}^{\infty} f_n^\dagger \frac{2(-1)^n}{(c+1)} \sum_{i=0}^{\min(m,n)} \frac{(-1)^i(m+n-i)!}{(m-i)!\,(n-i)!\,i!}\left(\frac{c-1}{c+1}\right)^{m+n-2i} \tag{2.3.45}$$

$$(c=1): \quad g_m^\dagger = \sum_{n=0}^{\infty} f_n^\dagger \sum_{i=0}^{\min(m,n)} \frac{(m+n-i)!\,\theta^{m+n-2i}(1-\theta)^i}{(m-i)!\,(n-i)!\,i!\,(\theta+1)^{m+n+1-i}} \tag{2.3.46}$$

$$(\theta=0,\ c=1): \quad g_m^\dagger = \sum_{n=0}^{\infty} f_n^\dagger \delta_{mn} = f_m^\dagger. \tag{2.3.47}$$

Sumita's result for $\theta=\frac{1-c}{2}$ [791, theorem 5.5.2] is a special case of (2.3.42):

$$g_m^\dagger = \sum_{n=0}^{\infty} f_n^\dagger 2(-1)^n \sum_{i=0}^{\min(m,n)} \frac{(-1)^i(m+n-i)!\,(2(c-1))^{n-i}0^{m-i}}{(m-i)!\,(n-i)!\,i!\,2^{m+n+1-i}(2c)^{-i}}.$$

The ith term of the inner sum equals zero unless $i=m$ (because of the 0^{m-i} term). If $i=m$, then $n\geq m$ (because of the upper limit of the inner sum), and so

$$\begin{aligned} g_m^\dagger &= \sum_{n=m}^{\infty} f_n^\dagger 2(-1)^n \frac{(-1)^m(m+n-m)!\,(2(c-1))^{n-m}}{0!\,(n-m)!\,m!\,2^{m+n+1-m}(2c)^{-m}} \\ &= \sum_{n=m}^{\infty} f_n^\dagger \binom{n}{m}(-1)^{n+m}(c-1)^{n-m}c^m \\ &= \sum_{n=m}^{\infty} f_n^\dagger \binom{n}{m}(1-c)^{n-m}c^m. \end{aligned} \tag{2.3.48}$$

This is Sumita's result.

Other Uses of Logarithmic Transformation

The use of a logarithmic transformation with the Laguerre transform has been suggested elsewhere. In section 7.9 of [791], Sumita considers the problem of determining the distribution of the product of independent Beta variates. He finds the distribution of the logarithm of the variates analytically (rather than in terms of the Laguerre coefficients), and then uses the Laguerre transform to find the distribution of the sum of the transformed variates. From this, he obtains the distribution of the

product by the inverse transformation. Sumita's approach is similar to that used by Ramsey *et al* described below.

Furmanski and Petronzio [285] introduced the idea of a logarithmic transformation to solve a special class of problems encountered in Quantum Chromo-Dynamics. The method is described by Ramsey in [672]. He considers the solutions of integro-differential evolution equations of the form

$$\frac{dF(x,t)}{dt} = \int_x^1 \frac{1}{y} P(x/y) F(y,t)\, dy. \tag{2.3.49}$$

Comparing the right side of (2.3.49) with the expression for the probability density of the product of two random variables in terms of their densities suggests that it is a closely related problem. If the two random variables are independent and always positive we have from (2.1.11)

$$f_Z(z) = \int_0^\infty \frac{1}{x} f_X(z/x) f_Y(x)\, dx. \tag{2.3.50}$$

Methods that have been used to solve (2.3.49) are similar to those suitable for solving (2.3.50). Two examples are Mellin transforms [119,325,634] and brute force numerical methods (evaluating the integral) [124]. Ramsey says that the brute force numerical methods "tend to be inefficient and prone to instability". The Mellin transform requires a numerical inversion which can be difficult. The method presented by Ramsey is "more stable and more accurate than other methods in that it allows one to deal with the functions $F(x,t)$ directly rather than with their integral transforms" [672, p. 98] It involves a change of variables in (2.3.49) via a logarithmic transformation, followed by a sum convolution which is evaluated by a Laguerre transform. (Of course more than this is required in order to solve (2.3.49), but this is not our concern here.)

Ramsey discusses the effect of truncation error and round-off error, pointing out that there is an inherent tradeoff between them: making the truncation error smaller by using more coefficients to represent the functions will compound rounding errors by requiring more arithmetic operations. For his particular application, Ramsey found that the round off error became significant (greater than 1%) with double precision arithmetic when N, the number of coefficients used, was greater than 15. In general he suggests values between 8 and 15 should be used. His values of N are less than those used by Keilson and Nunn [435, p.347] who used 60, or Sumita [792, p.262] who used 500 and found that when programmed with double precision arithmetic there was "no evidence of numerical problems" (for simply calculating the convolution) [792, p.273].

Ramsey's method differs from the method we have developed above in the manner of determining the Laguerre coefficients of the logarithmically transformed function. Whereas we have attempted to determine these from the Laguerre coefficients of the function to be transformed, Ramsey assumes the Mellin transform of the original function is known. He shows on page 102 of [672] how to derive the required Laguerre coefficients in terms of these transforms. This approach is obviously no good for our application. Ramsey also suggests (page 105) the idea of expanding the function by

a power series and presents a method for determining the Laguerre coefficients in terms of the coefficients of the power series expansion. It is not clear (to the present author) how this method can work.

2.3.5 Other Types of Laguerre Transforms and Related Results

Before leaving the topic of Laguerre transforms, we briefly point out some other work on convolutions and Laguerre transforms which is not as widely known. A number of unexpected connections have been considered in [40] (see especially the editor's preface), which contains a paper on the use of Laguerre polynomials for convolutions on the Heisenberg group [62].

Hirschmann [378] develops an inversion formula for the transform defined for any real function $f(n)$ by

$$\hat{f}(x) \;=\; \sum_{m=0}^{\infty} f(m) L_m(x). \tag{2.3.51}$$

His formula is related to the Post-Widder inversion formula for the Laplace transform [379, page 65].

Debnath [189] introduced a transform in 1960 which is more closely related to Keilson's. It is a generalisation of McCully's transform [559] which was used for solving certain differential equations. Debnath's transform of a function $F(x)$ is defined by

$$f_a(n) \;=\; \int_0^{\infty} e^{-x} x^a L_n^{(a)}(x) F(x) dx \qquad a > -1. \tag{2.3.52}$$

In [192] Debnath shows that if $f_a(n)$ and $g_a(n)$ are the Debnath-Laguerre transform of the functions $F(x)$ and $G(x)$ respectively, then the Debnath-Laguerre transforms of the convolution $F(x) *_D G(x)$ exists when $Re(a) > -\frac{1}{2}$ and is equal to $f_a(n) g_a(n)$. The convolution $*_D$ is defined by

$$\begin{aligned} F(x) *_D G(x) \;=\; & \frac{\Gamma(n+a+1)}{n!\sqrt{\pi}} \int_0^{\infty} e^{-t} t^a F(t) \int_0^{\pi} e^{-\sqrt{xt}\cos\phi} \times \\ & \frac{J_{a-\frac{1}{2}}(\sqrt{xt}\sin\phi)\sin^{2a}\phi}{(\frac{1}{2}\sqrt{xt}\sin\phi)^{a-\frac{1}{2}}} \;\times\; G(x + 2t\sqrt{xt}\cos\phi)\, dt\, d\phi, \end{aligned} \tag{2.3.53}$$

where $J_n(z)$ is the Bessel function of the first kind of order n. This is much more complicated than equation 2.3.8 and is obviously of no use to us (for the purposes of probabilistic arithmetic). Debnath's transform has been further studied by Glaeske [308,309,311] and Fenyö [264]. In [308], Glaeske shows that

$$\begin{aligned} f_a(z) \;&=\; \frac{-\sin\pi z}{\pi} \int_0^{\infty} \frac{s^{-z-1}}{(s+1)^{a+1}} \int_0^{\infty} e^{-x/(s+1)} x^a F(x)\, dx\, ds \\ &=\; \frac{-\sin\pi z}{\pi} \mathcal{M}_{-z} \left\{ \frac{1}{(s+1)^{a+1}} \mathcal{L}_{(1/(s+1))} \left[x^a F(x) \right] \right\}. \end{aligned} \tag{2.3.54}$$

This relationship seems to be of even less use than equation 2.3.11. Note that the Laguerre "polynomials" in the above transform are defined for non-integral degrees

by

$$L_\nu^{(\alpha)}(x) = \frac{\Gamma(\nu+\alpha+1)}{\Gamma(\nu+1)\Gamma(\alpha+1)}\,\Phi(-\nu,\alpha+1;x), \tag{2.3.55}$$

where Φ is the degenerate hypergeometric function [336, section 9.2]. Equation 2.3.54 is equivalent to (2.3.14) when ν is an integer [410].

The Laguerre transform closest in spirit to Keilson *et al's* (which we studied above) is that discussed by Verma [840]. He obtains a real inversion formula for a general Laguerre transform of the form

$$f_n^{(\alpha,\beta)}(s) \;=\; \int_0^\infty e^{-st}(st)^\alpha L_n^{(\beta)}(st) f(t)\,dt. \tag{2.3.56}$$

His inversion formula is

$$f(t) \;=\; s(st)^{\beta-\alpha} \sum_{n=0}^{\infty} \frac{n!}{\Gamma(1+\beta+n)}\, L_n^{(\beta)}(st) f_n^{(\alpha,\beta)}(s). \tag{2.3.57}$$

He also gives the following convolution formula for this transform [841]. Defining a special case of the transform (equivalent to Debnath's)

$$V_n^{(\alpha+\beta)}[f(t)](s) \;=\; \int_0^\infty e^{-st} L_n^{(\alpha+\beta)}(st) f(t)\,dt \tag{2.3.58}$$

he shows that

$$V_n^{(\alpha+\beta)}\left[\frac{d}{dt}(f*g)\right](s) \;=\; s\sum_{r=0}^{n} V_r^{(\alpha)}[f(t)](s) V_{n-r}^{(\beta)}[g(t)](s) \tag{2.3.59}$$

where $*$ is the ordinary sum convolution of two distributions, and f and g are distributions with supports bounded on the left. This is a generalisation of the result of Genin and Calvez [300] that

$$V_n^{(0)}\left[\frac{d}{dt}(f*g)\right](s) \;=\; \sum_{r=0}^{n} V_r^{(0)}[f(t)](s) V_{n-r}^{(0)}[g(t)](s). \tag{2.3.60}$$

Noting that $L_n^{(0)} = L_n$, (2.3.59) can be seen to be equivalent to (2.3.8).

Other Laguerre transforms are studied in [37,410,421]. Various issues relating to the convergence of certain Laguerre series expansions are discussed in [38,346, 598–600,636,890]. Laguerre series expansions have also been used by Ackroyd [5] to invert a Poisson transform [91]. Laguerre polynomials have been used by Tsamasphyros *et al* for the numerical inversion of Mellin transforms [822,826] (although unfortunately this doesn't help with the problem discussed in section 2.3.2). They have also been used for the numerical calculation of Fourier transforms [855] and evaluation of certain Hankel transforms [286]. The probabilistic origin of Laguerre and other classical orthogonal polynomials is investigated by Cooper *et al* in [167, 709]. Laguerre series representations of certain special probability densities can be found in [109,303,404,814,815,823].

Laguerre polynomial expansions also appear in probability theory under the name of Gram-Charlier expansions [448]. If the successive derivatives of the Gaussian density function in the ordinary type A Gram-Charlier series are replaced by derivatives of the Gamma density then a representation of a density $f(x)$ of the following form is obtained:

$$f(x) \;=\; \left(\sum_{n=0}^{\infty} \alpha_n L_n^{(m)}(x)\right) p_m(x). \tag{2.3.61}$$

In equation 2.3.60, $p_m(x) = e^{-x}x^{m-1}/\Gamma(m)$ for $m > 0$. The coefficients $\{\alpha_n\}$ are given by

$$\alpha_n \;=\; \frac{1}{\binom{m+n+1}{n}} \int_0^{\infty} f(x) L_n(x)\, dx. \tag{2.3.62}$$

The fact that there is no exponential weighting function in (2.3.61) means that $\{\alpha_n\}_{n=0}^{\infty}$ will only represent f using (2.3.60) if all the moments of f exist. Equation 2.3.60 has the property that if only k terms of the infinite series are used, then the first k moments of the series representation of f will be correct [814, theorem 2.1]. Further details and applications can be found in [303,404,815,823].

2.3.6 Conclusions on the Laguerre Transform Method

Although we have covered a lot of detail in the analysis of the Laguerre transform method in the hope of using it for product and quotient convolutions, there are many problems which we have not mentioned. Many of these have already been examined to some extent by Sumita and others, and do not just occur in the attempt to calculate product and quotient convolutions. We have already mentioned (section 2.3.4) how the shape of the density can affect the number of Laguerre coefficients required for an accurate representation. This is clearly an undesirable property. There are other methods (which we discuss elsewhere) for representing distributions and calculating convolutions which do not suffer from this problem. Other problems associated with the Laguerre transform method include the handling of dependencies (but see [796,798]), and the manner in which one can obtain the Laguerre coefficients to start with. If the density's analytical formula is known then one can calculate the coefficients analytically. However, if the density is derived from sample values this is not possible.

To sum up then, we can say that while orthogonal series in general, and the Laguerre transform in particular, seem at first sight to be good choices for our goal, it turns out that this is not the case. The Laguerre transform can not be used to calculate product and quotient convolutions anywhere near as easily as it can be used to calculate sum and difference convolutions. Thus although a useful tool for sum and difference convolutions, it is not suitable as the basis for a more general probabilistic arithmetic.

Figure 2.2: The histogram representation: $f_X^H(x)$ approximates $f_X(x)$.

2.4 The Histogram and Related Methods

The histogram method is a way of representing probability distributions and calculating their convolutions. It was developed by Colombo and Jaarsma [159,160] based on ideas in [403] and has also been examined by Keey and Smith [433]. In this section we will examine this method in some detail, explaining how it is used and discussing the various difficulties that are encountered. We present the algorithms in question explicitly (they were only outlined by Colombo and Jaarsma in [160]). We will also compare the histogram method with two other similar methods (Kaplan's DPD method and Moore's generalisation of Interval Arithmetic).

2.4.1 The Histogram Representation

Definitions and Notation

The histogram representation is discussed with reference to figure 2.2. The probability density f_X of a random variable X is approximated by f_X^H. This is defined by the set of ordered pairs

$$H_X \;=\; \{(x_0, p_0), (x_1, p_1), \ldots, (x_{n-1}, p_{n-1}), (x_n, 0)\}. \tag{2.4.1}$$

where $\sum_i p_i = 1$ and $x_i < x_{i+1}$ for $i = 0, \ldots, n-1$. We have

$$f_X^H(x) = p_i \qquad \text{for } X \in [x_i, x_{i+1}). \tag{2.4.2}$$

In the *equiprobable* histogram (which is the focus of our attention here) we also have

$$p_i \;=\; \int_{x_i}^{x_{i+1}} f_X(x)\,dx \;=\; \int_{x_i}^{x_{i+1}} f_X^H(x)dx \;=\; \frac{1}{n}$$

for $i = 0, 1, \ldots, n-1$. The zero in the last ordered pair in (2.4.1) allows a neater definition of the histogram and it is used to indicate the end of the data structure in computer implementations.

Alternative Choices of p_i

An alternative to the equiprobable histogram is the *equispaced* histogram where $x_i - x_{i-1}$ is constant. Colombo and Jaarsma [160] discuss the relative merits of the equiprobable versus the equispaced histograms. They present an argument in favour of equiprobable intervals in terms of the choice of class intervals in χ^2 tests [343,450,543], and they compare the accuracy of the results obtained in using both techniques. Their conclusion is that equiprobable intervals are generally better.

Note that "optimal" representations of the form (2.4.2), such as those that are obtained by minimising the L^1 distance [429,430] (or "disparity" [671])

$$\int_{\Re} \left| F_X^H(x) \;-\; F_X(x) \right| dx,$$

between the distribution functions F_X^H and F_X (corresponding to the densities f_X^H and f_X) in order to choose a histogram representation, are not appropriate for our purposes. Not only is the exact distribution F_X required to construct these approximations F_X^H, but one can not easily maintain the "optimality" when histograms are combined.

Choice of the Number of Intervals and the End Points x_0 and x_n

Henceforth we shall only consider the equiprobable histogram. There thus arises the question of the choice of n, the number of intervals, and the choice of the endpoints x_0 and x_n. The number of intervals affects the accuracy of the convolutions and the time taken to compute them. We shall see below that the combination of two histograms of n bins has a computational complexity of $O(n^4)$. This restricts n to be normally less than 100.

Regarding the end points x_0 and x_n, if F_X has finite support, then they can be simply set to inf supp F_X and sup supp F_X respectively. If supp $F_X = \Re$, then there are two alternatives: one can either truncate f_X in some way, or one can use an extended number system that allows ∞ and $-\infty$ as possible values. Colombo and Jaarsma [160] argue, but not very convincingly, that a reasonable rule is simply to make the width of the end intervals some constant amount α larger than the width of the penultimate interval contiguous to it. That is,

$$x_n - x_{n-1} = \alpha(x_{n-1} - x_{n-2}) \quad \text{and} \quad x_1 - x_0 = \alpha(x_2 - x_1).$$

They suggest that α should be around 2.5 to 3 (as long as the "long tailed distributions ... are regularly shaped" [160]). The alternative method (using the extended real number system $\Re^* = \Re \cup \{-\infty, \infty\}$) has been used by several authors in interval arithmetic [358,364,418,503,727]. This alternative method was used to calculate the example results presented below. It has the advantage of not relying on (sometimes unprovable) assumptions about the distributions in question.

2.4.2 Combining Histograms to Calculate Convolutions

We now consider how the equiprobable histograms (2.4.1) can be combined in order to calculate convolutions of the probability distributions which they represent.

The Basic Combining Rule

Suppose we have two histograms

$$H_X = \{(x_0, p_0), (x_1, p_1), \ldots, (x_{n-1}, p_{n-1}), (x_n, 0)\}$$

and

$$H_Y = \{(y_0, q_0), (y_1, q_1), \ldots, (y_{m-1}, q_{m-1}), (y_m, 0)\}$$

which represent two random variables X and Y with probability densities f_X and f_Y respectively. We assume that X and Y are independent. If $\Box$ is some arithmetic operation, and $Z = X \Box Y$, then H_Z^U, is an *unsorted* histogram which approximates the density f_Z. It is given by

$$H_Z^U = \{(\underline{z}_0, \overline{z}_0, r_0), \ldots, (\underline{z}_{nm-1}, \overline{z}_{nm-1}, r_{nm-1})\}, \tag{2.4.3}$$

where

$$\begin{aligned} \underline{z}_{im+j} &= \min\{x_i \Box y_j, x_{i+1} \Box y_j, x_i \Box y_{j+1}, x_{i+1} \Box y_{j+1}\}, \\ \overline{z}_{im+j} &= \max\{x_i \Box y_j, x_{i+1} \Box y_j, x_i \Box y_{j+1}, x_{i+1} \Box y_{j+1}\}, \end{aligned} \tag{2.4.4}$$

and

$$r_{im+j} = p_i \times q_j, \tag{2.4.5}$$

for $i = 0, \ldots, n-1$ and $j = 0, \ldots, m-1$.

The rationale behind (2.4.4) and (2.4.5) is to consider all possible pairs of intervals and combine them together using the rules of interval arithmetic [17]. The upper and lower bounds of the resultant interval are the maximum and minimum of all combinations of the endpoints of the interval operands. Equation 2.4.5 says that the probability that Z will take on a value within a given interval $[\underline{z}_{im+j}, \overline{z}_{im+j}]$ is the product of the probabilities that X and Y will take on values within the intervals $[\underline{x}_i, \overline{x}_i]$ and $[\underline{y}_j, \overline{y}_j]$.

This basic combining rule is not entirely adequate because the histogram H_Z^U is unsorted, and some of its constituent intervals may overlap. This can be seen from the following example. Let

$$H_X = \{(1, 0.5), (3, 0.5), (4, 0)\}$$

and

$$H_Y \;=\; \{(2, 0.5), (4, 0.5), (6, 0)\}.$$

Then if $Z = X \times Y$,

$$H_Z^U = \{(1, 12, .25), (4, 12, .25), (6, 16, .25), (12, 24, .25)\}.$$

Observe that the four intervals in H_Z^U ([1,12), [4,12), [6,16) and [12,24)) are not disjoint.

Construction of the Disjoint Histogram and Condensation

Our final aim is to approximate f_Z by a histogram of the form of (2.4.1). The following procedure can be used in order to convert H_Z^U to this form.

1. **Sorting step.** Sort the nm 3-tuples of H_Z^U into an order specified by the relation $\leq_H$ to give H_Z^S. The relation $\leq_H$ is defined by

   $$(\underline{z}_i, \overline{z}_i, p_i) \;\leq_H\; (\underline{z}_j, \overline{z}_j, p_j)$$

   if

   $$\underline{z}_i < \underline{z}_j \text{ or } (\underline{z}_i = \underline{z}_j \text{ and } \overline{z}_i \leq \overline{z}_j)$$

   This can be perfomed using a standard sorting algorithm (such as quicksort [722]) by using an appropriate comparison function.

2. **Construction of the disjoint histogram H_Z^D.** The next step is to construct a disjoint histogram H_Z^D, where

   $$H_Z^D = \{(\underline{w}_0, \overline{w}_0, s_0),\ (\underline{w}_1, \overline{w}_1, s_1), \ldots, (\underline{w}_{nm-1}, \overline{w}_{nm-1}, 0)\}$$

   and

   $$[\underline{w}_i, \overline{w}_i] \cap [\underline{w}_j, \overline{w}_j] \;=\; \emptyset \qquad \text{for all } i \neq j.$$

   This will not necessarily be an equiprobable histogram (*i.e.* there may exist an i and j, $i \neq j$, such that $s_i \neq s_j$), although it will have the property that $\overline{w}_i = \underline{w}_{i+1}$ for $i = 0, \ldots, nm-2$. We will use the above (redundant) 3-tuple representation of the elements of H_Z^D for clarity. The procedure used to form H_Z^D from H_Z^S is as follows:

```
for (i := 0; i < nm; i++){
    w_i := z_i; w̄_i := z_{i+1}; s_i := 0;
    for (j := i; j ≥ 0; j−−){
        if (z̄_j > w_i and z_j < w̄_i){
            if (w̄_i > z̄_j and w_i ≥ z_j)
                s_i += r_j((z̄_j − w_i)/(z̄_j − z_j));
            else if (w̄_i ≤ z̄_j and w_i ≥ z_j)
                s_i += r_j((w̄_i − w_i)/(z̄_j − z_j));
        }
    }
}
```
(2.4.6)

The purpose of the inner **for** loop is to determine, for each interval $[\underline{z}_j, \overline{z}_j]$, whether it overlaps the interval $[\underline{w}_i, \overline{w}_i]$. The loop starts at $j = i$ because there cannot be any overlap of relevance at $j > i$ due to the sorting performed in step 1, and the assignments $\underline{w}_i := \underline{z}_i$ and $\overline{w}_i := \underline{z}_{i+1}$.

The first **if** statement is true when $[\underline{z}_j, \overline{z}_j] \cap [\underline{w}_i, \overline{w}_i] \neq \emptyset$. When this occurs we have to determine how much they overlap, and hence how much probability to include in the interval $[\underline{w}_i, \overline{w}_i]$ due to the probability in $[\underline{z}_j, \overline{z}_j]$. This is what the next two statements do.

Regarding the inner two **if** and **else if** statements, note that the cases ($\underline{w}_i < \underline{z}_j$ and $\overline{w}_i > \underline{z}_j$) and ($\underline{w}_i < \underline{z}_j$ and $\overline{w}_i > \overline{z}_j$) cannot occur because $\underline{w}_i = \underline{z}_i$, $\overline{w}_i = \underline{z}_{i+1}$ for $j \leq i$, and by the sorting step, $\underline{z}_i < \underline{z}_{i+p}$ or ($\underline{z}_i = \underline{z}_{i+p}$ and $\overline{z}_i \leq \overline{z}_{i+p}$) when $p > 0$.

3. **Construction of the equiprobable histogram H_Z via condensation** In order to construct an equiprobable histogram H_Z from the disjoint histogram H_Z^D we proceed in the following manner. If H_Z is to comprise ℓ intervals, *i.e.*

$$H_Z \;=\; \{(v_0, t_0), (v_1, t_1), \ldots, (v_{\ell-1}, t_{\ell-1}), (v_\ell, 0)\},$$

then $t_i = 1/(\ell-1)$ for $i = 0, \ldots, \ell-1$ and the v_i are determined by the following algorithm. The variable u keeps track of how much probability we have used up.

```
u := 0;
for(i := 0, j := 0; i < nm; i++){
    u += s_i;
    while(u ≥ 1/(ℓ − 1)){
        j++;
        v_j := w̲_i + [(1/(ℓ + 1) − (u − w̲_i))/w̲_i] ∗ (w̄_i − w̲_i);
        t_{j−1} := 1/(ℓ − 1);
        u −= 1/(ℓ − 1);
    }
}
v_ℓ := w̄_{nm};
t_ℓ := 0;
```
(2.4.7)

The histogram H_Z is the result we required. We have *condensed* the histogram H_Z^D of $nm - 1$ intervals to the histogram H_Z of $\ell - 1$ intervals. (We usually have $\ell \ll nm$.) This condensation is necessary if we are to perform a series of arithmetic operations on a number of histograms, for otherwise the final histogram would have an enormous number of intervals.

Computational Complexity and Examples

The computational complexity of the above algorithm is dominated by step 2, the construction of the disjoint histogram H_Z^D. This has time complexity $O((nm)^2)$ because for each of the nm bins in H_Z^U we have to check whether any of the other

Figure 2.3: Histogram density of $Z = X \times Y$ where $df(X) = df(Y) = U_{1,2}$ and $N = 100$.

$nm-1$ bins overlap. The sorting in step 1 ($O(nm \log nm)$) saves some time, but only by a constant factor (2). We will see below that when the histogram bins (intervals) are replaced by single points, then the complexity is dominated by the time taken for sorting.

The above combination algorithm was implemented in the form of a computer program and was used to calculate the following examples. Figure 2.3 shows the calculated histogram density of $Z = X \times Y$ where X and Y are independent random variables with uniform distributions on $[1, 2]$. All the distributions were represented by 100 histogram bins. The calculation took 20 minutes of CPU time on a Microvax II minicomputer. Figure 2.4 shows the result of the same calculation when only 50 bins were used to represent the distributions involved. In this case the CPU time was only 1.6 minutes. Figure 2.5 shows the histogram representation of Z when $Z = X + Y$. Again X and Y were independent and uniformly distributed on [1,2] and 50 bins were used in the histogram representation.

Figure 2.4: Histogram density of $Z = X \times Y$ where $df(X) = df(Y) = U_{1,2}$ and $N = 50$.

Figure 2.5: Histogram density of $Z = X + Y$ where $df(X) = df(Y) = U_{1,2}$ and $N = 50$.

Figure 2.6: The Discrete Probability Distribution representation.

2.4.3 Kaplan's Method of Discrete Probability Distributions

The DPD Representation

A method for representing and combining probability distributions similar to the histogram method was developed by Kaplan [423]. He used simple discretisations of probability distributions and called his method the DPD (Discrete Probability Distributions) method. Instead of approximating a continuous probability density by a sequence of histogram bins (as in figure 2.2), a sequence of delta functions is used (see figure 2.6). Thus a random variable X is represented by a set of ordered pairs

$$D_X = \{(x_1, p_1), \ldots, (x_n, p_n)\}$$

where p_i is the probability that the variable will take on the value x_i. Whilst this seems to be really little different to Colombo and Jaarsma's histogram method, it does in fact differ in two ways. Firstly, Kaplan considers a DPD to *be* "our state of knowledge" of the random variable in question and so the question of the accuracy of representation does not arise:

> We allow ourselves total freedom to select the x_i and p_i any way at all, save only that the set $\{(x_i, p_i)\}$ adequately represents our state of knowledge and that it be suited to the numerical procedures we have in mind [423, p.190].

Secondly, the combination rules and condensation procedure are simpler than those for the histogram method, both in a conceptual and computational complexity sense. We will briefly examine this below.

Nonuniqueness of the DPD Represenatation

Firstly let us note that under Kaplan's definition and interpretation there are an infinite number of DPDs which can describe the same random variable. This caused consternation for Nelson and Rasmussen [613]. They were unhappy with the fact that

$$\{(2, 0.02), (3, 0.1), (4, 0.08), (3, 0.08), (4, 0.40), (3, 0.32)\} \tag{2.4.8}$$

and

$$\{(2, 0.02), (3, 0.18), (4, 0.48), (5, 0.32)\} \tag{2.4.9}$$

both described the same random variable. As Kaplan does not specify any constraints such as $x_i \neq x_j$ for $i \neq j$, this is not surprising. Of course, one can represent a given random variable with the minimal number of 2-tuples if this condition is imposed. Kaplan's reply [424] points this out. The operation required to convert (2.4.8) into (2.4.9) is analogous to the condensation operation performed on the histograms described earlier, but is simpler here because we only have 2-tuples instead of 3-tuples. This condensation needs to be performed in any case, as we shall see below.

Combination and Condensation of DPDs

Two DPDs are combined in the following obvious manner. If X and Y are two random variables represented by $D_X = \{(x_i, p_i)\}_{i=1}^{n}$ and $D_Y = \{(y_j, q_j)\}_{j=1}^{m}$ respectively, then we calculate $D_Z = \{(z_k, r_k)\}_{k=1}^{nm}$, where

$$z_{i+jm} \;=\; x_i \Box y_j \tag{2.4.10}$$

and

$$r_{i+jm} \;=\; p_i \times q_j \tag{2.4.11}$$

for some operation $\Box$. This is analogous to the combination rule for the histogram method.

Since D_Z has nm points a condensation operation is necessary. We have used the following algorithm in the calculation of the examples presented below. Let $D_Z = \{(z_i, r_i)\}_{i=1}^{nm}$ be the DPD produced by the combination procedure. Assume we want to condense D_Z to ℓ points. Let us call the condensed DPD $D_Z^c = \{(z_i^c, r_i^c)\}_{i=1}^{\ell}$.

Then the following algorithm will determine D_Z^c from D_Z:

$$
\begin{array}{l}
i := j := 1; \\
spacing := (z_{nm} - z_1)/\ell; \\
v := z_1 + spacing; \\
\textbf{while}(i \leq \ell)\{ \\
\quad r_j^c := 0; \\
\quad \textbf{while}(z_i < v \text{ and } i \leq \ell)\{ \\
\quad\quad r_j^c \mathrel{+}= r_i;\ i{+}{+}; \\
\quad \} \\
\quad z_j^c := v - (spacing/2); \\
\quad v \mathrel{+}= spacing; \\
\quad j{+}{+}; \\
\}
\end{array}
\tag{2.4.12}
$$

Recently Kaplan has demonstrated an improved condensation procedure [425]. It is similar to that which we use in the method we develop in chapter 3.

Kaplan [423, p.196] suggested that upon obtaining the DPD of the desired result that it could be "smoothed" in order to make it look better. He does not make any suggestions as to how this could be carried out, although perhaps the work of Schoenberg, de Boor and others [92,100,713,847] on splines and histograms could be used. Of course there is absolutely no probabilistic justification for such a procedure. We shall see later that when working with lower and upper bounds on *distribution functions* the need for smoothing is absent.

Applications and Examples

Kaplan shows how one can define probabilistic functions based on the DPD method. His approach is analogous to the "extension principle" used for fuzzy numbers (see chapter 4). (In fact the fuzzy extension principle was originally motivated by the probabilistic counterpart.) He uses the idea of a probabilistic function as a basis for a method of seismic risk assessment [426]. Other uses are in probabilistic risk assessment [12] (see [773] and [105] for some general background), and in probabilistic fracture mechanics [487].

Using the algorithm detailed above we calculated the distribution of $Z = X \times Y$ and $X + Y$ for the same parameters as the examples for the histogram method. Figure 2.7 shows the output for $Z = X \times Y$ with n, the number of points used, equal to 100. Figure 2.8 shows the output for $Z = X + Y$ with $n = 50$. The irregularity in figure 2.7 and the dip in the middle of figure 2.8 are both due to the condensation procedure. Note that the cumulative distribution formed from these two discrete densities would be quite accurate. Whilst the simple DPD condensation algorithm has the disadvantage of producing these irregularities, it has the advantage of being much faster than the histogram method. In fact, the complexity of the DPD method is dominated by the time taken for sorting. (Sorting is only implicit in our above algorithms, but is necessary in a practical implementation.) The CPU times for the

Figure 2.7: Discrete probability distribution of $Z = X \times Y$, $df(X) = df(Y) = U_{1,2}$, and $n = 50$.

Figure 2.8: Discrete probability distribution of $Z = X + Y$, $df(X) = df(Y) = U_{1,2}$, and $n = 50$.

two examples here were 6.42 and 1.58 seconds respectively. Compare these with the 1323 and 100 seconds for the histogram method.

2.4.4 Moore's Histogram Method

The histogram method presented above has been rediscovered by R.E. Moore [585] who developed it in the context of interval analysis (see [583] and section 2.5 below). Moore [585] shows how to determine the cumulative distribution function of an arithmetic operation on random variables when the random variables are represented by the union of disjoint but contiguous intervals with a certain probability of occurence within each interval. Moore uses only a small number of subdivisions (2 or 4). He still obtains results which correspond closely with results obtained by Monte Carlo simulation (in terms of visual inspection of the graph of the cumulative distribution). This may be due in part to the fact that cumulative distributions do tend to look better than the corresponding densities because of the smoothing effect of the integration.

More important however, is Moore's idea of the *elasticity* of a function with respect to a given input variable. This is defined as the limit of the ratio of variation of the function to the variation of the input variable as the input variation becomes small. In other words, if the function is $f(x_1, \ldots, x_n)$, the elasticity of f with respect to x_i is the logarithmic derivative $\left(\frac{\partial f}{f}\right) / \left(\frac{dx_i}{x_i}\right)$. This can be approximated by considering some interval $X_i = [\underline{X}_i, \overline{X}_i]$ such that $x_i \in X_i$, and calculating

$$\left(\frac{\text{width}(f)}{\text{mid}f}\right) / \left(\frac{\text{width}(X_i)}{\text{mid}(X_i)}\right). \tag{2.4.13}$$

In equation 2.4.13 we have $\text{mid}(X_i) = (\underline{X}_i + \overline{X}_i)/2$, $\text{width}(X_i) = \overline{X}_i - \underline{X}_i$ and $\text{width}(f)$ is the range of f as x_i ranges over $[\underline{X}_i, \overline{X}_i]$. Moore's idea is to subdivide finely only those X_i with a large elasticity because increasing the subdivision of those with a small elasticity will have a negligible effect on the accuracy of the answer. This idea would decrease the computational complexity of the histogram method considerably if some of the variables have a larger elasticity than the rest and should certainly be taken into account in the development of a system for determining distributions of complex functions of random variables. The notion of subdividing only those intervals with high elasticity is equivalent to the screening procedure mentioned in [486]. Moore [585] gives an example application, an extended discussion of which appears in [582].

2.5 Interval Arithmetic and Error Analysis

We will now consider a number of techniques which determine limited information about the distribution of functions of random variables. It seems natural to group these techniques together here, more for their commonality of motivation than of method. One of the most widely used methods for determining information about

distributions of functions of random variables is *interval arithmetic*. In this section, as well as describing the basic ideas of interval arithmetic, we will consider a number of variations which have been proposed that provide some extra information about the distribution of values within an interval. We will then compare the interval arithmetic approach to the standard methods used by metrologists for the propagation of the effects of uncertainties in the results of physical measurements.

2.5.1 The Standard Interval Arithmetic

Interval arithmetic was developed by Moore [583] although the original idea was proposed by Sunaga in 1958 [800]. The original motivation was to develop an automated way of handling rounding error in numerical computation. However it can also be used to propagate uncertainties in the input variables and this explains our consideration of it here.

The basic idea of interval arithmetic is to work in terms of closed intervals of real numbers. We define

$$\mathbf{I}_{\Re} = \{[\underline{x}, \overline{x}] | \, \underline{x} \leq \overline{x}, \, \underline{x}, \overline{x} \in \Re\}$$

to be the set of real intervals. The rules of interval arithmetic (which are rules for combining elements of $\mathbf{I}_{\Re}$) follow immediately from the convolution relations (1.9–1.12). It is simply necessary to observe that an interval $X^I \in \mathbf{I}_{\Re}$ which represents some random variable X is given by

$$X^I \;=\; [\inf \operatorname{supp} f_X, \sup \operatorname{supp} f_X] \tag{2.5.1}$$

where f_X is the density of X. If $X^I, Y^I \in \mathbf{I}_{\Re}$, then for some binary operation $\Box$, $Z^I = X^I \Box Y^I$ is given by

$$Z^I = [\underline{z}, \overline{z}] = \{x \Box y | \, x \in X, \, y \in Y\} \tag{2.5.2}$$

If $\Box$ is monotonic, then it suffices to examine the end points of X and Y and we can write $\underline{z} = \min \psi$ and $\overline{z} = \max \psi$, where

$$\psi \;=\; \{\underline{x} \Box \underline{y}, \, \underline{x} \Box \overline{y}, \, \overline{x} \Box \underline{y}, \, \overline{x} \Box \overline{y}\}. \tag{2.5.3}$$

Note that according to (2.5.2), if $Z^I = X^I / Y^I$ and $0 \in Y$, then $Z^I = \Re^*$ (or is undefined if we do not allow the values $-\infty$ and ∞). When $\Box = +$ equation 2.5.2 is also known as Minkowski addition (after Hermann Minkowski who considered it in 1911) [608]. It has applications in areas other than interval arithmetic [609, p.260].

When $\Box$ is one of the four arithmetic operations, Z^I can be written explicitly. Addition and subtraction are described by

$$X^I + Y^I \;=\; [\underline{x} + \underline{y}, \overline{x} + \overline{y}]$$

and

$$X^I - Y^I \;=\; [\underline{x} - \overline{y}, \overline{x} - \underline{y}].$$

Products and quotients of intervals are more complicated because these operations are not monotonic if zero is contained in one of the intervals [584]. It turns out that there does not exist a single "coordinate-based" representation of intervals such that both sums and products can be calculated directly in terms of the coordinates (*i.e.* the whole interval has to be taken into account) [676]. There are at least three books on interval arithmetic and interval analysis [17,583,584] and a comprehensive bibliography by Garloff [289] which contains 1743 items.

We shall see below that the definition of the interval arithmetic operations is only a small part (and the easiest part) of developing interval *algorithms.* It turns out that one can not simply apply the standard interval arithmetic operations to the standard algorithms of numerical analysis. The reason for this is *dependency width* or *dependency error.* This manifests itself in the *subdistributive property:* If $X^I, Y^I, Z^I \in \mathbf{I}_{\Re}$, then $X^I(Y^I + Z^I) \subseteq X^I Y^I + X^I Z^I$. This fact was first pointed out by Young [896] in 1931 and has been considered in detail by Ratschek and Spaniol [675,770]. The idea of a general function of intervals goes back at least to Burkhill [121,122].

2.5.2 Triplex Arithmetic and Dempster's Quantile Arithmetic

We will now turn to an examination of generalisations of interval arithmetic which aim to provide some information about the distribution of the variables within the interval. Ecker and Ratschek [247] have considered intervals probabilistically in an attempt to understand the phenomena of subdistributivity and inclusion monotonicity. They also suggested a joint representation of distributions and intervals and studied some properties of Dempster's quantile arithmetic which we examine below. Ahmad [11] is supposed by Moore [584, page 17, note 4] to have looked at the arithmetic of probability distributions from the point of view of interval arithmetic. However, Ahmad's paper is solely concerned with nonparametric estimators of probability densities, and he has nothing to say about probabilistic arithmetic.

Triplex Arithmetic

Triplex arithmetic [29,618] was developed by Nickel and others to overcome a perceived inadequacy of ordinary interval arithmetic. A *triplex number* X^T is an ordered triple $[\underline{x}, \tilde{x}, \overline{x}]$ where $\underline{x}$ is the *lower bound* on X, $\tilde{x}$ is the *main value* of X, and $\overline{x}$ is the *upper bound* on X. If the $\tilde{x}$ is discarded then we are left with ordinary interval arithmetic. The purpose of the "main value" (which is treated like a single real number in Triplex arithmetic calculations) is to provide some information about the distribution of X within X^I. This is particularly important when f_X is concentrated within a small region but has very wide tails. It turns out that the provision of this "main value" can help the convergence and accuracy of iterative algorithms [617]. Problems which are unstable using ordinary interval arithmetic can be made stable when triplex arithmetic is used.

Quantile Arithmetic

Dempster [204,205,207] has developed a method called *quantile arithmetic* which can be considered as a variation on triplex arithmetic. It is perhaps worthwhile to quote Dempster's motivation for quantile arithmetic before we describe it in detail.

> [W]hile the support of the error distribution of some finite computations on interval (random) variables may become arbitrarily large, the error distribution itself may be concentrated in a small interval with probability close to one. In such a case, the error distribution of the result has low dispersion, but long *tails* in which little probability is massed. It is upon consideration of this possibility that quantile arithmetic is based [204, pp.110–111], [205, p.187], [207, pp.224–225].

Again we consider a random variable X with density f_X and distribution function F_X. The *quantile number* X^Q represents X by the approximation of f_{X^Q} to f_X:

$$f_{X^Q}(x) = \begin{cases} \alpha & \text{if } x = F_X^{-1}(\alpha), \\ 1-2\alpha & \text{if } x = F_X^{-1}(\tfrac{1}{2}), \\ \alpha & \text{if } x = F_X^{-1}(1-\alpha), \\ 0 & \text{otherwise.} \end{cases} \tag{2.5.4}$$

(Note that f_X is a density while f_{X^Q} is a discrete frequency function.)

Two quantile numbers X^Q and Y^Q are combined to give $Z^Q = X^Q \Box Y^Q$ via the rule

$$f_{Z^q}(z) = \begin{cases} f_{X^Q}(x) f_{Y^Q}(y) & \text{for } z = x \Box y, \\ 0 & \text{otherwise.} \end{cases}$$

We have assumed here that both f_{X^Q} and $f_{Y^Q} \neq 0$ for only one choice of x and y such that $z = x \Box y$ for a given z. In order to convert the nine point frequency function f_{Z^q} into a three point frequency function f_{Z^Q}, Dempster uses the following condensation algorithm: Let $z_1 \leq z_2 \cdots \leq z_9$ be the nine values of z such that $f_{Z^q}(z) \neq 0$ and let $q_i = f_{Z^q}(z_i)$. Then f_{Z^Q} is given by

$$f_{Z^Q}(z) = \begin{cases} \alpha & \text{if } z = x_1, \\ 1-2\alpha & \text{if } z = x_2, \\ \alpha & \text{if } z = x_3, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\begin{aligned} x_1 &= z_i \text{ for the largest } i \text{ such that } \sum_{j=1}^{i} q_j \leq \alpha, \\ x_2 &= z_i \text{ for the smallest } i \text{ such that } \sum_{j=1}^{i} q_j \geq \tfrac{1}{2}, \\ x_3 &= z_i \text{ for the smallest } i \text{ such that } \sum_{j=i}^{9} q_j \leq \alpha. \end{aligned}$$

The choice of the parameter α is somewhat arbitrary (like the choice of q in Petcović's probable intervals — see section 2.2.2). Dempster gives a few suggestions, but these are mainly just common sense. Dempster [207] has applied his quantile arithmetic to the approximate solution of the distribution problem in linear programming [206, 419,468,779,861].

Whilst superficially interesting, and seemingly an improvement over interval arithmetic, it turns out that quantile arithmetic has an important disadvantage compared with interval arithmetic. In quantile arithmetic, *under*estimation of the spread of a result can occur. Standard interval arithmetic often *over*estimates the spread (sometimes by a large amount), but the result is still *correct* in the sense that the "true" interval is contained within the broader one. In contrast, quantile arithmetic will sometimes give results which are *wrong,* in that the calculated interval is narrower than the true interval.

2.5.3 Interval Arithmetic and Confidence Intervals from the Metrologist's Point of View

We have seen that ordinary interval arithmetic uses only information about the support of the distribution of the random variables involved. An alternative method is to use interval arithmetic to combine *confidence intervals.* (The standard method can thus be considered to combine 100% confidence intervals.) Different confidence intervals should then give further information about the distribution of the final result.

This idea has been considered by metrologists in the context of the theory of errors and the propagation of uncertainties arising from physical measurements. We will now examine what they have had to say on this topic. Metrologists have considered the use of confidence intervals because of a shortcoming of the standard method of the statement and propagation of experimental uncertainty. (Note that the term "experimental error" is out of favour nowadays: "the uncertainty, in former times frequently called 'error' ..." [846, p.83].) The standard method is to state uncertainties as standard deviations and to use the "general law of error propagation" [820] to propagate the errors through subsequent calculations. (A good recent review can be found in [22].) This general law is simply the linearisation of nonlinear functions by truncated (first order) Taylor series expansions about expected values (see section 2.2.2). Sometimes higher order expansions [69] or the more complicated expressions taking account of the covariances [78,352,858] are used. The shortcoming of this method arises in the determination of appropriate values of uncertainty to use for subjectively estimated "systematic errors," and the difficulty in converting a final uncertainty statement in terms of standard deviations into an interval result. An interval statement is often required for calibration or legal purposes. We shall consider these two difficulties in turn.

Subjective Interval Estimates of "Systematic Error" and Standard Deviations

In recent years the distinction between "random" and "systematic" errors (first explicitly proposed by D. Bernoulli in 1780 [746, p.290] although implicitly adopted by Newton as early as 1676 [744, p.222]) has been called into question [302,604, 846,879]. The point is that there has never been an entirely satisfactory criterion for deciding which category to use in any particular instance. (Müller [604, p.375] quotes Vigoureux: "One has to remember that some errors are random for one person and systematic for another.") Similarly, "what is a systematic error for one experiment may not be for another.... Much of the skill in experimental work comes from eliminating sources of systematic error" [49, p.72]. The choice is important because the standard ("orthodox") theory propagates random and systematic errors through calculations differently. Systematic errors are added arithmetically, whereas random errors are added in quadrature. Indeed, as Müller has pointed out, this is precisely the source of the difficulty in classification: "the traditional classification of uncertainties depends upon the further use we intend to make of them, and in general this can not be known in advance" [604, p.376].

The new "randomatic" theory avoids the problem of categorization by considering *all* errors to be random. However it does distinguish between "'objective' statistical estimates" and "'subjective' guesstimates" [302, p.625]. The subjective uncertainties are often given in terms of an interval. This has to be converted into a standard deviation in order to propagate it through any subsequent calculations. A number of (admittedly somewhat *ad hoc*) methods for doing this can be found in [878]. The subjective uncertainties are usually considered to be independent [878, p.83], although there seems to be no good *a priori* reason for this to be so. It could well be argued though that any error arising from an *ad hoc* handling of subjective uncertainties should usually be negligible compared with the variability due to the "objective statistical estimates." Arguments against the randomatic theory of errors and proposals for an improved orthodox theory based on a more careful distinction between the two classes of errors can be found in the closely argued paper of Colclough [158].

Determination of a Confidence Interval for the Final Result

A more serious and older problem is the conversion of a final uncertainty statement in terms of a standard deviation into an interval statement. We have already observed (section 2.2.2) that the Chebyshev inequality or its generalisations [212, 317,318,537] can be used to determine a confidence interval in terms of means and standard deviations. However, in general these will be very loose (pessimistic) intervals. Alternative approaches which give tighter, less pessimistic, intervals require assumptions about the underlying distributions.

The oldest method is the assumption of normality. We have already discussed this in section 2.2.1. It has never found universal acceptance and had strong critics even some 60 years ago: "I reject, then, the Gaussian theory of error, without qualification and with the utmost possible emphasis; and with it go all theoretical

grounds for adopting the rules that are based on it" [129, p.162]. We need add nothing further to this.

An alternative, which is similar in its general approach to the lower and upper bounds on the distributions we consider in chapter 3, has been considered independently by Kuznetsov [488] and Weise [857]. Assuming that the true "error" distribution is symmetrical, unimodal and has finite support, Kuznetsov [488] calculated a confidence interval in terms of a variance by using the "mean distribution" derived from lower and upper distribution functions satisfying the distributional assumptions. Weise [857], somewhat more ambitiously, considered a whole class of distributions $\mathcal{D}$. He then used a distribution which was a mean of all possible distributions over $\mathcal{D}$ in order to determine a confidence interval in terms of a given variance. (Obviously the choice of $\mathcal{D}$, and the fact that the *mean* distribution is used, will affect the result; different choices will give different results.) Nevertheless this approach seems to be a promising way of investigating the effects of different distributional assumptions and of handling different degrees of optimism.

Direct and Exclusive Use of Confidence Intervals

Although they are preferable to assuming normality, the methods of Kuznetsov and Weise are still not entirely satisfying. An alternative is to avoid completely the use of standard deviations as a measure of uncertainty and to use confidence intervals only from the outset. Müller [603] and others have argued against this idea on the grounds of difficulties in combining confidence intervals. The general consensus amongst metrologists seems to be described by the DIN standard 1319 part 4 "Basic Concepts of Measurements: Treatment of Uncertainties in the Evaluation of Measurements" [858] (a summary appears in [859]. This suggests a careful but fairly straight forward application of the general law for error propagation (in terms of standard deviations). No distinction is made between "random" and "systematic" errors apart from the methods of initially estimating the numerical value to be attributed. There are still difficulties in using standard deviations when repeated measurements are not independent [18], but these seem to be manageable given some idea of the spectrum of the error sequence.

Recently Rowe [692] has presented some very interesting results which could lead to a partial combination of the two techniques (intervals and standard deviations). Using two separate approaches, Rowe has determined lower and upper bounds for μ_Y and σ_Y in terms of μ_X, σ_X, $\underline{X}$, $\overline{X}$ or limited order statistic information, for a class of transformations $Y = g(X)$. Rowe's approach is preferable to the simple first order Taylor series approximations because lower and upper bounds for $\mu_{g(X)}$ and $\sigma_{g(X)}$ are given. These allow a more rigorous propagation of uncertainty. Also, when $\underline{X}$ and $\overline{X}$ or other distributional information is available, the tighter bounds can be obtained.

Confidence Curves and Fuzzy Numbers as a Generalisation of Interval Arithmetic

Finally, to conclude this somewhat discursive exploration of the metrological uses and significance of interval arithmetic, we can mention the idea of fuzzy arithmetic [432]. This has been suggested as a natural generalisation of interval arithmetic and error propagation techniques [240] because under the standard sup-min combination rules (see chapter 4) fuzzy numbers can be combined in terms of interval arithmetic on the level sets of their membership functions. We examine fuzzy arithmetic, and point out some similarities to the idea of confidence curves introduced by Cox [169] and developed by Birnbaum and Mau [76,77,553] in chapter 4. These confidence curves are made up of nested sets of confidence intervals at different confidence levels and they may provide a useful generalisation for the purposes of the theory and propagation of errors.

2.5.4 Permutation-Perturbation and Related Methods

The original motivation for interval arithmetic was the automatic control of rounding errors in numerical calculations. There are several others approaches possible for this, and some are of interest for our goals as well.

The most widely known is Wilkinson's analysis of particular algorithms to determine the accuracy of the result that can be expected [866,867]. More explicitly stochastic methods have been adopted in recent years. See for example [50,51,176].

Another method, which has been presented several times by Vignes *et al* [1,21, 260,648,842,843], is called the Permutation-Perturbation method or CESTAC (Contrôle et Estimation Stochastique des Arrondis de Calcul). The basic idea is to consider the 2^n *perturbations* of a result of an n stage numerical computation obtained by perturbing each operation by an appropriate amount positively or negatively. This is combined with *permutations* of the order of computation (which will not always give the results that would be obtained over $\Re$ because of the failure of associativity and distributivity under floating point arithmetic). Vignes argues that only a few (two or three) of this large number of results need be considered. He *randomly* perturbs and permutes the calculation and then *estimates* the accuracy of the result. Applications of the method can be found in [21,89,90].

Stummel [784] has also considered the effect of perturbations on intermediate results in the computation of arithmetic expressions by using the computational graph concept [60]. He has obtained a number of results for the *condition number* of an algorithm and has applied his results to a careful analysis of the numerical solution of a 2×2 linear system of equations [785] and the analysis of some interval arithmetic algorithms [786].

Chapter 3

Numerical Methods for Calculating Dependency Bounds and Convolutions

In my view there is a central obligation to face squarely what we do and do not know, and to study robustness of conclusion against mistakes in a priori *assumptions of independence, conditional or not.*
— William Kruskal

In this chapter we present a new and general numerical method for calculating convolutions of a wide range of probability distributions. An important feature of the method is the manner in which the probability distributions are represented. We use lower and upper discrete approximations to the quantile function (the quasi-inverse of the distribution function). This results in any representation error being always contained within the lower and upper bounds. This method of representation has advantages over other methods previously proposed. The representation fits in well with the idea of dependency bounds.

Stochastic dependencies which arise within the course of a sequence of operations on random variables are the severest limit to the simple application of convolution algorithms to the formation of a general probabilistic arithmetic. We examine the error caused by this effect (dependency error), and show how dependency bounds are a possible means of reducing its effect. Dependency bounds are lower and upper bounds on the distribution of a function of random variables which contain the true distribution even when nothing is known of the dependence of the random variables. They are based on the Fréchet inequalities for the joint distribution of a set of random variables in terms of their marginal distributions. We show how the dependency bounds can be calculated numerically when using our numerical representation of probability distributions. Examples of the methods we develop are presented, and we briefly describe relationships with other work on numerically handling uncertainties.

The present chapter is a very slightly modified version of the paper [874]. It is essentially self contained, and thus there is some slight repetition of material presented elsewhere in this thesis.

3.1 Introduction

In order to develop automated systems for dealing with uncertainty it is necessary to be able to calculate the basic operations of an uncertainty calculus numerically. Amongst the many different uncertainty calculi now available, ordinary probability theory is the oldest. Surprisingly though there has been little detailed examination of numerical methods for calculating the distribution of arithmetic operations on random variables. Although there have been a number of schemes proposed, so far there have been none that meet the following simple criteria:

1. The method should allow the calculation of the distribution of all four arithmetic functions of random variables (and not just addition and subtraction).
2. There should be no restrictions (or only very slight restrictions) on the class of random variables that can be handled.
3. There should be a careful treatment of all the errors arising in the calculation (particularly those due to the numerical representation adopted).
4. The method should be computationally tractable and the algorithms should be described explicitly.
5. The method should be simple to understand and implement.

The present chapter's goal then is to develop a method, satisfying these criteria, for what we will call "probabilistic arithmetic".

In this introductory section we will define the problem to be studied more precisely (section 3.1.1) and introduce the notions of dependency error and dependency bounds (sections 3.1.2 and 3.1.3). Since we believe it is more worthwhile developing a method with a rigourous foundation rather than an ad hoc technique which works for some applications, we will be concentrating on the foundations rather than the applications of probabilistic arithmetic. Using the description of the necessary layering of uncertain reasoning systems due to Bonissone [93,94], we could say that we are concentrating on the representation and inference layers but ignoring the control layer.

It is worthwhile to compare our methods with other probabilistic methods as well as methods based on other uncertainty calculi (such as the theory of fuzzy sets). While we postpone a detailed examination of this to chapter 4, we do point out now that our method has some similarity to Jain's method for combining fuzzy numbers [407,408]. Jain's method was subsequently criticised by Dubois and Prade [223,224], although there own method (L-R fuzzy numbers [225]) is not without drawbacks either. More recently [238] Dubois and Prade have examined the relationship between Moore's probabilistic arithmetic [585] (itself an outgrowth of interval arithmetic [584]) and fuzzy arithmetic, by drawing on some results from the Dempster-Shafer theory of evidence [732]. In chapter 6 (see also chapter 4, section 5) we will show that the normal combination rules for fuzzy numbers are in fact equivalent to our dependency bounds and that our limiting result (chapter 5) can

be used to derive a law of large numbers for fuzzy variables under a general extension principle. Perhaps the work which is closest in spirit to that presented here is that of Grosof [341]. Grosof has taken much the same approach as we have in analysing the probabilities of events and combinations of events (rather than arithmetic operations on random variables). He has shown that a special case of his interval conditional probability logic is actually formally identical to the Dempster-Shafer theory of evidence. Several other authors [221,836] have recently adopted an approach similar to that outlined in the present chapter (*viz.* calculation of lower and upper bounds on probabilities when limited dependence information is available).

3.1.1 The Problem

Consider the following problem. Let X and Y be two random variables with distribution functions F_X and F_Y respectively. Let $Z = X \Box Y$, where $\Box$ is some arithmetic (or other) operation. Then what is F_Z, the distribution of Z? For any given $\Box$, if the joint distribution F_{XY} is known, then a solution to the problem in terms of an integral can be written down. The appropriate integral is determined from the Jacobian of transformation. In many cases, closed form solutions to the integral do not exist. Whilst series solutions can generally be obtained, the resulting formulae are often very complex and are of little value for an automated system.

If the joint distribution of X and Y is not known (*i.e.* only the marginals F_X and F_Y are known), then one can not, even in principle, calculate F_Z. One can, however, calculate lower and upper bounds on F_Z, as was recently shown by Frank, Nelsen, and Schweizer [277].

The present chapter develops numerical algorithms to solve both of these problems. Algorithms are developed for calculating lower and upper bounds on F_Z when it is known that X and Y are independent or when there is no knowledge of the dependency structure of X and Y at all. The techniques developed are quite general and can be used for almost all distributions. The method of representing the distributions and the results obtained for the convolutions are better than other numerical techniques that have been presented to date. The algorithms can be used to calculate lower and upper bounds in the manner of Frank, Nelsen, and Schweizer for a much larger class of distributions than can be managed analytically. All the algorithms are described explicitly and are computationally efficient. They have been implemented on a minicomputer and have been used to calculate some example results which are included in this chapter.

3.1.2 The Idea of Probabilistic Arithmetic and the Need for Dependency Bounds

The problems mentioned above arise naturally in the consideration of probabilistic arithmetic (the name is due to Kaplan [423]). The goal of probabilistic arithmetic is to replace the usual arithmetic operations on numbers by the appropriate operations on random variables (which are represented by their distribution functions). This is akin to several other ideas that have appeared in the literature, most notably

interval arithmetic [17,584] and fuzzy arithmetic [225,432]. The similarities and connexions with these other ideas are not considered here, but are examined in some detail in chapter 4. One of the goals of probabilistic arithmetic is to solve random algebraic equations numerically, problems for which the methods of solution available at present are still rather limited (see [8,73,74] for a review of the available methods and known results). There are numerous other possible applications if a successful probabilistic arithmetic can be developed.

Among the problems that need to be considered are the errors that can occur in a probabilistic arithmetic calculation and how they can be handled. We will show how one type of error, dependency error, is the most severe restriction on the simple application of convolution algorithms to the formation of a workable probabilistic arithmetic. The errors in probabilistic arithmetic can be classified into five types:

Representation Error. This is the error caused by the approximation of a function defined on an uncountable subset of $\Re$ by a finite number of points or coefficients.

Truncation Error. This error, which can be distinguished from general representation error for some representations, is caused by the necessity (for some distributions) of truncating the support of the distributions to a union of finite intervals.

Calculation Approximation Error. This is caused by approximations made in developing the formulae used to implement the probabilistic arithmetic. For example, if a Taylor series expansion (say for the variance of the product of two independent random variables in terms of their moments [352,829]) is truncated after a finite number of terms, then the rule itself will only be approximate. This will introduce errors into the calculated results.

Rounding Error. This is simply the error caused by performing numerical computations on machines with a finite wordlength. This will not be considered further here because in the absence of ill-conditioning, this can easily be made arbitrarily small. In any case, it is a problem associated with nearly all numerical algorithms, not only those for probabilistic arithmetic.

Dependency Error. Dependency error is the most important type of error in probabilistic arithmetic. It arises in much the same manner as spurious correlation [645]. It is explained by considering the following sequence of operations where all the quantities are random variables. Assume that V, X, and Y are independent. Then calculate

$$A = X/Y;\ B = X * V;\ C = A + B;$$

The problem is that even though the three inputs are independent, A and B are not because they both depend on X.

Although such dependencies can be handled, in principle, by techniques based on the Jacobian of transformation, it is impractical to contemplate the use of such techniques for handling sequences of computations of the type commonly

employed in the deterministic case. Thus, in order to carry out sequences of operations on random variables, one is usually obliged (for tractability) to assume independence in cases where dependencies, such as the one in the above equations, exist. A major aim of the present chapter is to investigate the question of handling the error that arises when independence is assumed (*i.e.* dependency error).

The above classification is useful for comparing the different approaches to representing and calculating with distribution functions. This is true even though no precise definition of the "error" involved has been given. There does not seem to be any one "best" measure of error between two distributions as the best measure will depend to a large extent on what the distribution being calculated will be used for.

3.1.3 Methods of Handling the Errors in Probabilistic Arithmetic

We only concern ourselves here with representation and dependency errors. (Truncation error and approximation error are discussed in chapter 2 and rounding errors are mentioned for a specific problem arising in section 3.6.3.) In section 3.4 it will be shown how a natural representation inspired by the method we develop to combat dependency error can essentially remove all the problems of representation error. The basic idea is to use lower and upper approximations to the desired distribution rather than one single approximation. Any representation error is contained within these lower and upper bounds.

Our method of handling dependency error is to use the results of Frank, Nelsen, and Schweizer [277] mentioned above. This allows the calculation of lower and upper bounds on a required distribution even if the distribution itself can not be calculated because it is not known that the variables involved are independent, or because the joint distribution is not available. There are several other possible approaches to handling dependency error. For instance, to calculate the distribution of

$$X(Y+Z)+X/V$$

it is only necessary to rewrite it as

$$X(Y+Z+1/V)$$

and the problem of dependency error introduced by repeated occurrences of the variable X disappears. This can be considered as solving the problem in Bonissone's control layer as we have changed the order in which the lower level computations are carried out. This rearrangement of expressions has been used in interval arithmetic and is discussed in more detail in chapter 4.

There are various extensions to the ideas mentioned in the previous paragraph which are worth studying. One of these, which is discussed briefly in section 3.5, is to use some measure of dependence between the random variables and to modify the combination rules to take this extra information into account. With respect to this it is of interest to note a forgotten paper of Kapteyn (it appears to have been cited

no more than three times since publication in 1912). In [427] Kapteyn considers problems such as the correlation between X and Y where $X = A + B + C + D$, $Y = A + E + F + G$ and all the $A, \ldots, G$ are independent. Another paper that deserves mention here is that of Manes [540] (see also [32,541]). Manes shows that the repeated occurrence of a variable in an expression causes problems in a wide range of fuzzy (vague or imprecise) theories (including the theory of fuzzy sets) and so the problem is not peculiar to probability theory.

3.1.4 Outline of the Rest of This Chapter

The rest of this chapter is organised as follows. Section 3.2 contains concise definitions of concepts which are needed later on. In section 3.3, the dependency bounds of Frank, Nelsen, and Schweizer are derived and explained. We extend their results by proving the pointwise best possible nature of the dependency bounds for operations other than addition and subtraction. Some examples are calculated by directly using the formulae for dependency bounds. We show that a better way to calculate dependency bounds numerically is to use the numerical representation of probability distributions developed in section 3.4 along with special discrete versions of the dependency bound formulae. Algorithms for calculating ordinary convolutions in terms of this numerical representation are also developed in section 3.4. Sections 3.5 and 3.6 contain some extensions to the basic results of sections 3.3 and 4 and include suggestions and directions for further research. Finally, section 3.7 contains a summary of the contributions of this chapter and some conclusions.

3.2 Definitions and Other Preliminaries

We now briefly present a number of definitions we need later on. Most of the material here is covered in more detail in [718].

3.2.1 Distribution Functions

Definition 3.2.1 *Let X be a random variable on $\Re$. Then its* distribution function *F_X is defined by $df(X) = F_X(x) = P\{X < x\}$ for $x \in \Re$.*

The corresponding *density*, when it exists, is denoted f_X.

Definition 3.2.2 *The* support *of F_X, denoted* supp F_X, *is the set of $x \in \Re$ such that $f_X(x) = F'_X(x)$ exists and is non zero.*

Definition 3.2.3 *The set of all distribution functions that are left continuous on $\Re$ will be denoted Δ. The subset of distribution functions in Δ such that $F(0) = 0$ will be denoted Δ^+.*

3.2.2 Binary Operations

Definition 3.2.4 *A* binary operation *on a nonempty set S is a function T from $S \times S$ into S.*

Definition 3.2.5 *Let T be a binary operation on S. An element a of S is a* left null element *of T if $T(a,x) = a$ for all $x \in S$; it is a* right null element *of T if $T(x,a) = a$ for all $x \in S$; and it is a* null element *of T if it is both a left and right null element of T.*

A binary operation can have at most one null element.

3.2.3 Quasi-Inverses

Quasi-inverses are generalisations of the inverse of a function which are defined even when the function has jump discontinuities. In this chapter we are only concerned with quasi-inverses of non-decreasing distribution functions. Let F be a non-decreasing function on a closed interval $[a,b]$. Let $y \in [a,b]$. Then $F^{-1}(y)$ is the set $\{x | F(x) = y\}$. If F has no jump discontinuities then the cardinality of $F^{-1}(y)$ is one and we simply write $F^{-1}(y) = x$. If the cardinality of $F^{-1}(y)$ is not one, then we have to somehow choose between the various elements of $F^{-1}(y)$.

Definition 3.2.6 *Let F be a non-decreasing function on a closed interval $[a,b]$. Then $Q(F)$ is the set of functions F^* (known as* quasi-inverses of F*) defined on $[F(a), F(b)]$ by*

1. *$F^*(F(a)) = a$ and $F^*(F(b)) = b$.*
2. *If $y \in \operatorname{Ran} F$ then $F^*(y) \in F^{-1}(y)$.*
3. *If $y \notin \operatorname{Ran} F$ then $F^*(y) = \sup\{x | F(x) < y\} = \inf\{x | F(x) > y\}$.*

All $F^* \in Q(F)$ are non-decreasing and coincide except on an at most a denumerable set of discontinuities. There is a unique function $F^\wedge \in Q(F)$ which is left continuous on $(F(a), F(b))$, and a unique function $F^\vee \in Q(F)$ which is right continuous on $(F(a), F(b))$. These are given by

$$F^\wedge(y) = \sup\{x | F(x) < y\}$$

and

$$F^\vee(y) = \inf\{x | F(x) > y\}.$$

For all $F^* \in Q(F)$, $F^\wedge \leq F^* \leq F^\vee$ on $(F(a), F(b))$. If F and G are non-decreasing on (a,b), then $F \geq G \Rightarrow F^\wedge \leq G^\wedge$ and $F^\vee \leq G^\vee$. If $[a,b] = \Re^* = \Re \cup \{-\infty, \infty\}$ (the "extended reals"), then $F^\wedge(0) = -\infty$. This introduces technical difficulties and so we adopt the convention that $F^\wedge(0) = F^\vee(0) = \inf \operatorname{supp} F$.

3.2.4 Triangular-norms

Definition 3.2.7 *A binary operation T on a set S is* associative *if*

$$T(T(x,y),z) = T(x,T(y,z)) \qquad \forall x,y,z \in S.$$

Definition 3.2.8 *A* triangular norm *(or* t-norm*) is an associative binary operation on $[0,1]$ that is commutative, non-decreasing in each place, and such that $T(a,1) = a$ for all $a \in [0,1]$.*

3.2.5 Copulas and Joint Distribution Functions

The most important notions for the present chapter are those of the copula and the Fréchet bounds.

Definition 3.2.9 *A two-dimensional* copula *C is a mapping $C\colon [0,1]\times[0,1] \mapsto [0,1]$ such that*

1. *$C(a,0) = C(0,a) = 0$ and $C(a,1) = C(1,a) = a$ for all $a \in [0,1]$.*
2. *$C(a_2,b_2) - C(a_1,b_2) - C(a_2,b_1) + C(a_1,b_1) \geq 0$ for all $a_1, a_2, b_1, b_2 \in [0,1]$ such that $a_1 \leq a_2$ and $b_1 \leq b_2$.*

All copulas satisfy

$$W(a,b) \leq C(a,b) \leq M(a,b) \tag{3.2.1}$$

for all $(a,b) \in [0,1]\times[0,1]$, where the t-norms W and M (which are also copulas) are given by

$$W(a,b) = \max(a+b-1,0) \tag{3.2.2}$$

and

$$M(a,b) = \min(a,b). \tag{3.2.3}$$

Copulas link joint distributions with their marginals. Let H be a two dimensional distribution function with marginals F and G. Then there exists a copula C such that

$$H(u,v) = C(F(u),G(v)) \tag{3.2.4}$$

for all $u,v \in \Re$. The inverse relation is

$$C(u,v) = H(F^*(u),G^*(v)), \tag{3.2.5}$$

where $F^* \in Q(F)$ and $G^* \in Q(G)$. The copula C contains all the dependency information of H. If $C(u,v) = \Pi(u,v) = uv$, then the random variables are independent. Equation 3.2.4 is sometimes referred to as the "uniform representation" [210,458].

Combining (3.2.1–3.2.4) gives bounds on the joint distribution in terms of the marginals:

$$\max(F(u)+G(v)-1,0) \;\leq\; H(u,v) \;\leq\; \min(F(u),G(v)). \tag{3.2.6}$$

These are known as the *Fréchet bounds*. A copula is a t-norm if and only if it is associative. A t-norm T is a copula if and only if it is 2-increasing; that is, if $T(a_2,b_2)-T(a_1,b_2)-T(a_1,b_2)-T(a_2,b_1)+T(a_1,b_1)\geq 0$ for all $a_1,a_2,b_1,b_2\in[0,1]$ such that $a_1\leq a_2$ and $b_1\leq b_2$ (see p.79ff of [718]).

Definition 3.2.10 *Let C be a copula. The* dual *of C is the function C^d defined by $C^d(x,y)=x+y-C(x,y)$ for all $x,y\in[0,1]$.*

The dual copula should not be confused with the conorm of a t-norm T given by $T^*(x,y)=1-T(1-x,1-y)$. The dual of W is $W^d(u,v)=\min(u+v,1)$. Schweizer and Sklar [718] use the notation $\overline{C}$ for a dual copula. We use the overbar notation for a different purpose below.

3.2.6 The Triangle Functions τ and ρ and σ-convolutions

The following three operations are of great importance in the sequel. The operations τ and ρ are introduced by Schweizer and Sklar [718] because of their properties as triangle functions in the theory of probabilistic metric spaces. They are known elsewhere as the supremal and infimal convolutions.

Definition 3.2.11 *Let C be a copula and let L be a binary operation from $\Re^+\times\Re^+$ onto $\Re^+$ which is non-decreasing in each place and continuous on $\Re^+\times\Re^+$ except possibly at the points $(0,\infty)$ and $(\infty,0)$ (we shall call this class of functions $\mathcal{L}$ in the sequel). Then $\tau_{C,L}$ is the function on $\Delta^+\times\Delta^+$ whose value for any $F,G\in\Delta^+$ is the function $\tau_{C,L}(F,G)$ defined on $\Re^+$ by*

$$\tau_{C,L}(F,G)(x)=\sup_{L(u,v)=x}[C(F(u),G(v))].$$

Definition 3.2.12 *Let C, L, F, and G be as in definition 3.2.11. Then $\rho_{C,L}(F,G)$ is the function defined by*

$$\rho_{C,L}(F,G)(x)=\inf_{L(u,v)=x}[C^d(F(u),G(v))].$$

Definition 3.2.13 *Let C, L, F, and G be as in definition 3.2.11. Then $\sigma_{C,L}(F,G)$ is the function defined by*

$$\sigma_{C,L}(F,G)(x)=\int_{L\{x\}}dC(F(u),G(v))\qquad x\in(0,\infty),\tag{3.2.7}$$

where $L\{x\}=\{(u,v)|\,u,v\in\Re^+,\,L(u,v)<x\}$.

In the sequel we will write $\int_{L(u,v)<x}$ for $\int_{L\{x\}}$. The function $\sigma_{C,L}$ is the distribution of $L(X,Y)$ where X and Y are random variables with joint distribution $F_{XY}(u,v)=C(F(u),G(v))$. This operation is called a σ-convolution for the operation L (the σ signifies the additive properties of the integral in contradistinction to the infimum

and supremum operations in the infimal and supremal convolutions). The well known convolution for $C = \Pi$ and $L = \mathrm{Sum}$ given by

$$\sigma_+(F,G)(x) = \int_{u+v<x} d\Pi(F(u),G(v)) = \int_{-\infty}^{x} F(x-t)\,dG(t)$$

is a special case of (3.2.6). For each of the three types of convolution (τ, ρ and σ), L is sometimes written as an infix operator $\Box$ as in $Z = X \Box Y$. Explicit formulae for the other arithmetic operations are given by (2.1.9–2.1.12).

The $\tau_{C,L}$, $\rho_{C,L}$ and $\sigma_{C,L}$ operations can actually be defined on the whole of $\Delta \times \Delta$ for appropriate L. We use these extended definitions in the sequel.

3.3 Dependency Bounds and their Properties

The dependency bounds for the four arithmetic operations of addition, subtraction, multiplication and division are now derived and their properties examined. Unless otherwise stated, the random variables considered are almost surely positive. In other words, their distribution functions are in Δ^+. Bounds on the distribution of $Z = L(X,Y)$ are derived, where $L \in \mathcal{L}$. It is then shown how to apply these bounds to subtraction and division of random variables. Following this, the pointwise best possible nature of these bounds for general $L \in \mathcal{L}$ is shown. This was not proven in [277], although our proof is a fairly straightforward generalisation of the proof there for $L = \mathrm{Sum}$. Finally a number of examples which have been numerically calculated are presented.

3.3.1 The Dependency Bounds

Let X and Y be random variables on $\Re^*$ with $df(X) = F_X$ and $df(Y) = F_Y$ such that $F_X, F_Y \in \Delta$ and let $Z = L(X,Y)$ with $df(Z) = F_Z$ where $L \in \mathcal{L}$. Let $\underline{C}_{XY}$ be a lower bound on the copula C_{XY}. Then F_Z depends on the joint distribution of X and Y and will be contained within the bounds

$$\mathrm{ldb}_{\underline{C}_{XY}}(F_X,F_Y,L)(z) \;\le\; F_Z(z) \;\le\; \mathrm{udb}_{\underline{C}_{XY}}(F_X,F_Y,L)(z) \quad \forall z \in \Re^*. \tag{3.3.1}$$

When $\underline{C}_{XY} = W$ we will simply write ldb and udb. These stand for "lower dependency bound" and "upper dependency bound" respectively.

Theorem 3.3.1 ([277]) *When $\underline{C}_{XY} = W$ the functions* ldb *and* udb *are given by*

$$\mathrm{ldb}(F_X,F_Y,L)(x) = \tau_{W,L}(F_X,F_Y)(z) = \sup_{L(u,v)=x} W(F_X(u),F_Y(v)) \tag{3.3.2}$$

and

$$\mathrm{udb}(F_X,F_Y,L)(x) = \rho_{W,L}(F_X,F_Y)(z) = \inf_{L(u,v)=x} W^d(F_X(u),F_Y(v)). \tag{3.3.3}$$

Figure 3.1: Illustration for the proof of theorem 3.3.1.

Sometimes the notation $\underline{F}_Z$ or $\overline{F}_Z$ is used for $\mathrm{ldb}(F_X, F_Y, L)$ or $\mathrm{udb}(F_X, F_Y, L)$ respectively when F_X, F_Y, and L are clear from the context.

PROOF. With reference to figure 3.1 and to section 3.2.5, it is clear that for any given copula C, and any pair of points (u_1, v_1), (u_2, v_2) on the line $L(u, v) = x$,

$$\begin{aligned} W(F_X(u_1), F_Y(v_1)) &\leq C(F_X(u_1), F_Y(v_1)) \\ &= \iint_A dC(F_X(u), F_Y(v)) \\ &\leq \sigma_{C,L}(F_X, F_Y)(x) \\ &\leq \iint_B dC(F_X(u), F_Y(v)) \\ &= F_X(u_2) + F_Y(v_2) - C(F_X(u_2), F_Y(v_2)) \\ &= C^d(F_X(u_2), F_Y(v_2)) \\ &\leq W^d(F_X(u_2), F_Y(v_2)). \end{aligned}$$

The proof is completed by observing that $\tau_{W,L}$ is simply the greatest value of $W(F_X(u_1), F_Y(v_1))$ where (u_1, v_1) is on the line $L(u, v) = x$. Similarly, $\rho_{W,L}$ is the smallest value of $W^d(F_X(u_2), F_Y(v_2))$ where (u_2, v_2) is on the line $L(u, v) = x$. ■

By a similar argument one can show that the more general bounds $\mathrm{ldb}_{\underline{C}_{XY}}$ and $\mathrm{udb}_{\underline{C}_{XY}}$ are given by (3.3.2) and (3.3.3) with W simply replaced by $\underline{C}_{XY}$. Note that since $C_{XY} \geq W$ *always*, the bounds of theorem 3.3.1 will always hold. However, as we shall see in section 3.5, the bounds for $\underline{C}_{XY} \neq W$ are tighter, and thus provide more information about F_Z. We will only concern ourselves with $\underline{C}_{XY} = W$ in this

section. Note that knowing an upper bound on C_{XY} other than M does *not* allow one to construct tighter bounds.

Theorem 3.3.1 can be used to bound the distribution of all four arithmetic operations on almost surely positive random variables. The condition of positivity is necessary because product and quotient are only monotonic on $\Re^-$ or $\Re^+$ and not over all $\Re$. All the results are collected together in the following theorem.

Theorem 3.3.2 *Let X and Y be almost surely positive random variables with distributions F_X and F_Y, and let $Z = X \Box Y$, where $\Box$ is one of the four arithmetic operations $\Box \in \{+, -, \times, \div\}$. Then the lower and upper dependency bounds for F_Z, the distribution of Z, are given by*

$$\begin{aligned} \mathrm{ldb}(F_X, F_Y, +)(x) &= \sup_{u+v=x} (\max(F_X(u) + F_Y(v) - 1, 0)) \qquad (3.3.4) \\ \mathrm{udb}(F_X, F_Y, +)(x) &= \inf_{u+v=x} (\min(F_X(u) + F_Y(v), 1)) \end{aligned}$$

$$\begin{aligned} \mathrm{ldb}(F_X, F_Y, -)(x) &= \sup_{u+v=x} (\max(F_X(u) - F_Y(-v), 0)) \qquad (3.3.5) \\ \mathrm{udb}(F_X, F_Y, -)(x) &= 1 + \inf_{u+v=x} (\min(F_X(u) - F_Y(-v), 0)) \end{aligned}$$

$$\begin{aligned} \mathrm{ldb}(F_X, F_Y, \times)(x) &= \sup_{uv=x} (\max(F_X(u) + F_Y(v) - 1, 0)) \qquad (3.3.6) \\ \mathrm{udb}(F_X, F_Y, \times)(x) &= \inf_{uv=x} (\min(F_X(u) + F_Y(v), 1)) \end{aligned}$$

$$\begin{aligned} \mathrm{ldb}(F_X, F_Y, \div)(x) &= \sup_{uv=x} (\max(F_X(u) - F_Y(1/v), 0)) \qquad (3.3.7) \\ \mathrm{udb}(F_X, F_Y, \div)(x) &= 1 + \inf_{uv=x} (\min(F_X(u) - F_Y(1/v), 0)). \end{aligned}$$

PROOF. The Sum and Product cases follow directly from theorem 3.3.1. The Difference and Quotient cases can be seen by noting that on $\Re^+$ these are both monotonic operations; they are increasing in their first argument and decreasing in their second. They can be converted to Sum and Product as follows. For Difference ($Z = X - Y$), let $Y' = -Y$. Then $F_{Y'}(y) = 1 - F_Y(-y)$. Substitution into (3.3.4) and some slight rearrangement yields (3.3.5). For Quotient ($Z = X/Y$), let $Y' = 1/Y$. Then $F_{Y'}(y) = 1 - F_Y(1/y)$. Substitution into (3.3.6) yields (3.3.7). ■

The bounds for addition and subtraction actually hold for $F_X, F_Y \in \Delta$ since addition and subtraction are monotonic over all $\Re$.

3.3.2 Pointwise Best-Possible Nature of the Bounds

The bounds in theorem 3.3.1 are the pointwise best possible. The exact meaning of this is given by the theorem below which is a generalisation of theorem 3.2 of [277]. Theorem 3.3.3 generalises theorem 3.2 of [277] in two ways: it is for general $L \in \mathcal{L}$, not just $L = \mathrm{Sum}$; and it is for general $\underline{C}_{XY}$ and not just $\underline{C}_{XY} = W$. The proof is based on that in [277] but is sufficiently modified to warrant inclusion here.

Theorem 3.3.3 *Let F_X and F_Y be distribution functions in Δ, let $x \in \Re$, let $L \in \mathcal{L}$ and let $\underline{C}_{XY}$ be a lower bound on the copula C_{XY} ($W \leq \underline{C}_{XY} \leq C_{XY}$). Then,*

1. *There exists a copula $C^{(t)}$, dependent only on the value t of $\tau_{\underline{C}_{XY},L}(F_X, F_Y)$ at x such that*
$$\sigma_{C^{(t)},L}(F_X, F_Y)(x) = \tau_{\underline{C}_{XY},L}(F_X, F_Y)(x) = t$$
if not both F_X and F_Y are discontinuous at u and v respectively such that $L(u, v) = x$.

2. *There exists a copula $C^{(r)}$ dependent only on the value r of $\rho_{\underline{C}_{XY},L}(F_X, F_Y)(x+)$ such that*
$$\sigma_{C^{(r)},L}(F_X, F_Y)(x+) = \rho_{\underline{C}_{XY},L}(F_X, F_Y)(x+) = r$$
if not both F_X and F_Y are discontinuous at u and v respectively such that $L(u, v) = x$.

This theorem says that for any dependency bounds there will always be a copula such that the true $\sigma_\square$-convolution meets the bound at a given point. In other words one can not construct bounds any tighter than (3.3.2) and (3.3.3). Thus these dependency bounds are the pointwise best possible.

PROOF. The conditions on the continuity of F_X and F_Y will be ignored for now and their effect discussed later. Consider part 1 of the theorem first. It will be shown that if $C^{(t)}(a, b)$ is the copula defined by

$$C^{(t)}(a,b) = \begin{cases} \max(t, \underline{C}_{XY}(a,b)) & (a,b) \in [t,1]^2, \\ \min(a,b) & \text{otherwise,} \end{cases} \tag{3.3.8}$$

then $\sigma_{C^{(t)},L}(F_X, F_Y)(x) = t$. Note that $C^{(t)}(a, b) \geq \underline{C}_{XY}(a, b)$ for all $(a, b) \in [0, 1]^2$.

Since L is continuous and nondecreasing in each argument, for any x, the curve $L(u, v) = x$ is continuous and nonincreasing in the $u-v$ plane (see figure 3.2). Define $\mathcal{A}_x$ and $\mathcal{B}_x$ to be the regions of the extended plane above and below the line $L(u, v) = x$:

$$\mathcal{A}_x = \{(u, v) | L(u, v) > x\}$$
$$\mathcal{B}_x = \{(u, v) | L(u, v) < x\}.$$

Firstly observe that in view of theorem 3.3.1, $\tau_{\underline{C}_{XY},L}(F_X, F_Y)(x) = 1$ implies that $\sigma_{\underline{C}_{XY},L}(F_X, F_Y)(x) = 1$. Thus it is necessary to show that for $t \in [0, 1)$,

$$\sigma_{C^{(t)},L}(F_X, F_Y)(x) = \iint_{\mathcal{B}_x} dC^{(t)}(F_X(u), F_Y(v)) \leq t. \tag{3.3.9}$$

The condition is "$\leq t$" and not "$= t$" to cope with discontinuous F_X and F_Y: if F_X and F_Y are discontinuous, then $\sigma_{C^{(t)},L}(F_X, F_Y)$ is also, and thus the situation depicted in figure 3.3 could arise. In order to show (3.3.9), note that for $(u, v) \in \overline{\mathcal{B}}_x$ (the closure of $\mathcal{B}_x$),

$$\begin{aligned} \underline{C}_{XY}(F_X(u), F_Y(v)) &\leq \tau_{\underline{C}_{XY},L}(F_X, F_Y)(L(u, v)) \leq \tau_{\underline{C}_{XY},L}(F_X, F_Y)(x) = t \\ &\Rightarrow \max(t, \underline{C}_{XY}(F_X(u), F_Y(v))) = t. \end{aligned}$$

Figure 3.2: Graph of $L(u, v) = x$.

Figure 3.3: A possible $\sigma_{C^{(t)}}(F_X, F_Y)$ when F_X and F_Y have discontinuities.

Thus using the definition of $C^{(t)}$ in (3.3.8),

$$C^{(t)}(F_X(u), F_Y(v)) = \min(F_X(u), F_Y(v), t). \tag{3.3.10}$$

From this it is obvious that $C^{(0)}(F_X(u), F_Y(v)) = 0$ for all $(u, v) \in \overline{\mathcal{B}}_x$ and so $\sigma_{C^{(0)},L}(F_X, F_Y)(x) = 0$.

It is now only necessary to consider $t \in (0, 1)$. Let

$$u_0 = \sup\{u | F_X(u) < t\}$$

for $0 < t < 1$. We show that u_0 is finite. Since $\lim_{u \to -\infty} F_X(u) = 0$, $u_0 > -\infty$. Regarding whether $u_0 < \infty$, suppose to the contrary that $u_0 = \infty$ and thus that $F_X(u) < t$ for all finite u. Now let $L^\wedge$ be the two place function defined by

$$L(u, L^\wedge(x, u)) = x.$$

That is, if $L(u, v) = x$, then $L^\wedge(x, u) = v$. $L^\wedge$ is strictly increasing in its first argument and strictly decreasing in its second and if $L = \text{Sum}$, then $L^\wedge = \text{Difference}$. For finite x and u,

$$F_Y(L^\wedge(x, u)) \leq 1.$$

Since for any copula C, $C(a, 1) = a$ (see section 3.2.2), and since C is nondecreasing in each argument, for any $\epsilon \geq 0$, $C(a, 1 - \epsilon) \leq a$. Combining these facts gives

$$\underline{C}_{XY}\left(F_X(u), F_Y(L^\wedge(x, u))\right) \leq F_X(u) < t. \tag{3.3.11}$$

But

$$\tau_{\underline{C}_{XY},L}(F_X, F_Y)(x) = \sup_u \left[\underline{C}_{XY}\left(F_X(u), F_Y(L^\wedge(x, u))\right)\right],$$

and since u in (3.3.11) is arbitrary (but finite)

$$t = \tau_{\underline{C}_{XY},L}(F_X, F_Y)(x) \leq F_X(u) < t,$$

which is a contradiction and thus u_0 is finite.

It is next shown that $F_Y(v) \geq t$ whenever $v > L^\wedge(x, u_0)$. Suppose to the contrary that there exists a $v' > L^\wedge(x, u_0)$ such that $F_Y(v') < t$. Define ${}^\wedge L$ to be the two place function such that

$$L({}^\wedge L(x, v), v) = x$$

(*i.e.* $L(u, v) = x$ implies ${}^\wedge L(x, v) = u$ and ${}^\wedge L$ is strictly increasing in its first argument and strictly decreasing in its second). Since ${}^\wedge L(x, v') < u_0$, $F_X({}^\wedge L(x, v')) < t$. Thus for $u \leq {}^\wedge L(x, v')$,

$$\underline{C}_{XY}(F_X(u), F_Y(L^\wedge(x, u))) \leq F_X(u) \leq F_X({}^\wedge L(x, v')) < t,$$

and for $u \geq {}^\wedge L(x, v')$

$$\underline{C}_{XY}(F_X(u), F_Y(L^\wedge(x, u))) \leq F_Y(L^\wedge(x, u)) \leq F_Y(v') < t,$$

and again

$$t = \tau_{\underline{C}_{XY},L}(F_X, F_Y)(x) < t.$$

Figure 3.4: $C^{(t)}(F_X(u), F_Y(v))$.

This is a contradiction and so $F_Y(v) \geq t$ whenever $v > L^\wedge(x, u_0)$.

A final fact which is needed is that $F_X(u_0) \leq t$ (because $F_X(u) < t$ for $u < u_0$ and F_X is left-continuous), and thus $F_X(u) \geq t$ for $u > u_0$. Also, $F_Y(v) \geq t$ for $v > L^\wedge(x, u_0)$.

Collecting all the facts in the above three paragraphs and substituting into (3.3.8), the definition of $C^{(t)}$, gives

$$C^{(t)}(F_X(u), F_Y(v)) = \begin{cases} \min(F_Y(v), t) & u > u_0, \\ F_X(u) & u < u_0,\ v > L^\wedge(x, u_0), \\ \min(F_X(u), F_Y(v) & u < u_0,\ v < L^\wedge(x, u_0). \end{cases}$$

These values of $C^{(t)}(F_X(u), F_Y(v))$ are indicated on figure 3.4. It now becomes fairly easy to evaluate $\sigma_{C^{(t)},L}(F_X, F_Y)$.

Recall that

$$\sigma_{C^{(t)},L}(F_X, F_Y)(x) = \iint_{L(u,v)<x} dC^{(t)}(F_X(u), F_Y(v)).$$

Following Frank, Nelsen and Schweizer [277], the (u, v) plane is divided into the five regions $R_1, \ldots, R_5$ given by

$$\begin{aligned} R_1 &= [-\infty, u_0 - \delta] \times [v_0, v_0 + \delta], \\ R_2 &= [-\infty, u_0] \times [-\infty, v_0], \end{aligned}$$

$$
\begin{aligned}
R_3 &= [u_0, u_0+\delta] \times [-\infty, v_0-\delta], \\
R_4 &= \mathcal{B}_x \cap \{(u,v) | v > v_0 + \delta\}, \\
R_5 &= \mathcal{B}_x \cap \{(u,v) | u > u_0 + \delta\},
\end{aligned}
$$

where $\delta > 0$. If

$$
I_k = \iint_{R_k} dC^{(t)}(F_X(u), F_Y(v)),
$$

then

$$
\begin{aligned}
I_1 &= F_X(u_0-\delta) - \min\left(F_X(u_0-\delta), F_Y(v_0)\right), \\
I_2 &= \min(F_X(u_0), F_Y(v_0)), \\
I_3 &= \min(F_Y(v_0-\delta), t) - \min(F_X(u_0), F_Y(v_0-\delta)), \\
I_4 &= 0, \\
I_5 &= 0.
\end{aligned}
$$

Both I_4 and I_5 are zero because both $F_X(u)$ and $\min(F_Y(v), t)$ are constant in one direction of the (u,v) plane. The value stated for I_2 is obvious, and those for I_1 and I_3 follow by consideration of the boundaries of R_1 and R_3 respectively (see figure 3.4). Since $L^\wedge$ is non-increasing in its second argument and continuous,

$$
\begin{aligned}
\sigma_{C^{(t)},L}(F_X, F_Y)(x) &= \lim_{\delta \to 0+} (I_1 + I_2 + I_3) \\
&= F_X(u_0) + \min\left(F_Y(v_0), t\right) - \min\left(F_X(u_0), F_Y(v_0)\right) \\
&= \begin{cases} t & \text{if } F_Y(v_0) > t, \\ \max(F_X(u_0), F_Y(v_0)) \leq t & \text{if } F_Y(v_0) \leq t. \end{cases} \blacksquare
\end{aligned}
$$

There are several points worth noting regarding this theorem.

1. Since $v_0 = L^\wedge(x, u_0)$, the case that $F_Y(v_0) \leq t$ in the above formula implies that $\sigma_{C^{(t)},L}(F_X, F_Y)(x)$ has a jump at $x = L(u_0, v_0)$. This is explained below. The two cases $F_Y(v_0) = t$ and $F_Y(v_0) < t$ are considered separately. If $F_Y(v_0) = t$ then $\max(F_X(u_0), F_Y(v_0)) = t$ and so the bound has been met. If $F_Y(v_0) < t$ then F_Y has a jump at v_0. This can be seen by noting that $F_Y(v) \geq t$ for $v > L^\wedge(x, u_0)$ and therefore $F_Y(v) \geq t$ for $v > v_0$. But $F_Y(v_0) \neq t$ and so $F_Y(v) > t$ for $v > v_0$. Thus $F_Y(v_0) < t$ and $F_Y(v) > t$ for $v > v_0$; *i.e.* there is a jump at v_0.

   If $F_X(u_0) = t$, $\max(F_X(u_0), F_Y(v_0)) = t$ and the bound is still met. Assume then that $F_X(u_0) < t$. Then by a similar argument, F_X has a jump at u_0. Now if F_X has a jump at u_0 and F_Y a jump at v_0, then since L is continuous, $\sigma_{C^{(t)},L}(F_X, F_Y)$ has a jump at $L(u_0, v_0)$ and this is why

$$
\sigma_{C^{(t)},L}(F_X, F_Y)(L(u_0, v_0)) < t.
$$

2. The bounds for differences and quotients constructed earlier are pointwise best possible also. This is obvious as the difference is determined in terms of a sum

and the quotient in terms of a product. If one wanted to construct an explicit copula analogous to the $C^{(t)}$ in the above proof, one could use the fact (see theorem 3.ii.2 of [719]) that if g is a non-increasing function then

$$C_{X,g(Y)}(u,v) = u - C_{XY}(u, 1-v).$$

3. The second part of theorem (3.3.3) can be proved in an entirely analogous manner (Frank *et al.* [277] give some details). The reason for having $x+$ rather than x in part 2 of the theorem is due to F_X and F_Y being left-continuous but not necessarily right-continuous.

4. Note that the $\tau_W(F_X, F_Y)$ and $\rho_W(F_X, F_Y)$ operations are *not* distributions of some function of X and Y (see [717]).

3.3.3 Examples

Some examples illustrating the lower and upper dependency bounds are now presented. All of the examples have been calculated numerically, but involve no approximation apart from rounding error (which is negligible) and an error due to the termination of the search for the infimum or supremum in calculating the τ and ρ operations. Frank, Nelsen and Schweizer [277] present a few analytical results obtained using the method of Lagrange multipliers. In general it seems rather difficult to obtain exact dependency bounds using this technique. In any case, as was mentioned in section 3.1, since it is the numerical values that are ultimately of interest, there is no real advantage in deriving analytical results if the formulae are so complicated that a computer program has to be written to calculate their specific values.

In order to calculate the dependency bounds numerically it is useful to rewrite the formulae (3.3.4–3.3.7) in the following form:

$$\mathrm{ldb}(F_X, F_Y, +)(x) = \left[\sup_u (F_X(u) + F_Y(x-u) - 1)\right]^+, \tag{3.3.12}$$

$$\mathrm{udb}(F_X, F_Y, +)(x) = 1 + \left[\inf_u (F_X(u) + F_Y(x-u))\right]^-, \tag{3.3.13}$$

$$\mathrm{ldb}(F_X, F_Y, -)(x) = \left[\sup_u (F_X(u) - F_Y(u-x))\right]^+, \tag{3.3.14}$$

$$\mathrm{udb}(F_X, F_Y, -)(x) = 1 + \left[\inf_u (F_X(u) - F_Y(u-x))\right]^-, \tag{3.3.15}$$

$$\mathrm{ldb}(F_X, F_Y, \times)(x) = \left[\sup_u (F_X(u) + F_Y(x/u) - 1)\right]^+, \tag{3.3.16}$$

$$\mathrm{udb}(F_X, F_Y, \times)(x) = 1 + \left[\inf_u (F_X(u) + F_Y(x/u))\right]^-, \tag{3.3.17}$$

$$\mathrm{ldb}(F_X, F_Y, \div)(x) = \left[\sup_u (F_X(u) - F_Y(u/x))\right]^+, \tag{3.3.18}$$

$$\mathrm{udb}(F_X, F_Y, \div)(x) = 1 + \left[\inf_u (F_X(u) - F_Y(u/x))\right]^-, \qquad (3.3.19)$$

where $[x]^+ = \max(x, 0)$ and $[x]^- = \min(x, 0)$. From these formulae it becomes straightforward to calculate lower and upper dependency bounds if one has a subroutine to calculate $F_X(x)$ and $F_Y(y)$. The calculation is simply a search for a maximum or minimum of a simple function of F_X and F_Y for a given x. This is not completely trivial to do numerically as the F_X and F_Y are defined on a continuum. The techniques developed in section 3.4 are better suited for numerical calculation.

All of the examples are for random variables with uniform distributions $U_{a,b}$ given by

$$U_{a,b}(x) = \begin{cases} 0 & x \in [-\infty, a] \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x \in [b, \infty]. \end{cases}$$

The examples are presented in figures 3.5–3.9 and details are provided in the figure captions. Some examples of dependency bounds when $\underline{C}_{XY} \neq W$ are given in section 3.5.

Some points to note about these figures are:

1. Comparison of figures 3.5 and 3.6 reveals the effect of the spread of the two random variables on the distance between the lower and upper dependency bounds. In fact, if one of the random variables has zero dispersion (a distribution function equal to a unit step), then the two bounds are identical.

2. The fact that the lower and upper bounds in figures 3.5 and 3.6 are themselves of the form $U_{a,b}$ is not really significant — it is just a peculiarity of the ldb and udb formulae for sums (and differences) of uniformly distributed random variables. Figure 3.7 shows the dependency bounds for a product. It can be seen that these are not uniform.

3. A comparison of figures 3.8 and 3.9 indicates the effect of identical distributions for numerator and denominator on the dependency bounds of their quotient. If the random variables are almost surely equal then the distribution of their quotient will be a unit step at $x = 1$. This possibility is contained within the bounds of figure 3.9.

4. The bounds for the following two cases are identical to those in figures 3.5 and 3.6 respectively:

   (a) $Z = X - Y$ with X distributed $U_{1,2}$ and Y distributed $U_{0,1}$.

   (b) $Z = X - Y$ with X distributed $U_{1,2}$ and Y distributed $U_{0,0.2}$.

Examples of dependency bounds for random variables with non-uniform distributions are given in section 3.4.

Figure 3.5: Lower and upper dependency bounds for $X + Y$ where both X and Y are uniformly distributed on $[0, 1]$.

Figure 3.6: Lower and upper dependency bounds for $X + Y$ where X is uniformly distributed on $[0, 1]$ and Y is uniformly distributed on $[0.8, 1]$.

Figure 3.7: Lower and upper dependency bounds for $X \times Y$ where both X and Y are uniformly distributed on $[0, 1]$.

Figure 3.8: Lower and upper dependency bounds for X/Y where X is uniformly distributed on $[4, 6]$ and Y is uniformly distributed on $[1, 2]$.

Figure 3.9: Lower and upper dependency bounds for X/Y where both X and Y are uniformly distributed on $[1, 2]$.

3.3.4 Exclusive Use of Lower and Upper Bounds

We will now show how it is possible to work only with the dependency bounds themselves, without worrying about the distributions within these bounds (the precise meaning of this statement is given by the following theorem).

Theorem 3.3.4 *Let* $[\underline{F}_X, \overline{F}_X]$ *and* $[\underline{F}_Y, \overline{F}_Y]$ *be two pairs of distributions functions such that* $\underline{F}_X(x) \leq \overline{F}_X(x)$ *and* $\underline{F}_Y(y) \leq \overline{F}_Y(y)$ *for all* $x, y \in \Re$*, and let* $\Box$ *be non-decreasing in each argument. Then for any* $F_X \in [\underline{F}_X, \overline{F}_X]$ *and any* $F_Y \in [\underline{F}_Y, \overline{F}_Y]$

$$\mathrm{ldb}_{\underline{C}_{XY}}(\underline{F}_X, \underline{F}_Y, \Box) \leq \mathrm{ldb}_{\underline{C}_{XY}}(F_X, F_Y, \Box) \tag{3.3.20}$$

and

$$\mathrm{udb}_{\underline{C}_{XY}}(\overline{F}_X, \overline{F}_Y, \Box) \geq \mathrm{udb}_{\underline{C}_{XY}}(F_X, F_Y, \Box). \tag{3.3.21}$$

PROOF. This follows directly from lemma 7.2.2 of [718]. ■

The only thing that remains to be determined is explicitly how to use the $\underline{F}_X$ and $\overline{F}_X$. But this is very simple. The following formulae are derived by considering the increasing or decreasing nature of the operations $\Box$ (for $\Box \in \{\times, \div\}$ we only consider random variables on $\Re^+$).

$$\begin{array}{ll}
+ & \left\{ \begin{array}{lcl} \underline{F}_Z &=& \mathrm{ldb}(\underline{F}_X, \underline{F}_Y, +) \\ \overline{F}_Z &=& \mathrm{udb}(\overline{F}_X, \overline{F}_Y, +) \end{array} \right. \\
- & \left\{ \begin{array}{lcl} \underline{F}_Z &=& \mathrm{ldb}(\underline{F}_X, \overline{F}_Y, -) \\ \overline{F}_Z &=& \mathrm{udb}(\overline{F}_X, \underline{F}_Y, -) \end{array} \right. \\
\times & \left\{ \begin{array}{lcl} \underline{F}_Z &=& \mathrm{ldb}(\underline{F}_X, \underline{F}_Y, \times) \\ \overline{F}_Z &=& \mathrm{udb}(\overline{F}_X, \overline{F}_Y, \times) \end{array} \right. \\
\div & \left\{ \begin{array}{lcl} \underline{F}_Z &=& \mathrm{ldb}(\underline{F}_X, \overline{F}_Y, \div) \\ \overline{F}_Z &=& \mathrm{udb}(\overline{F}_X, \underline{F}_Y, \div) \end{array} \right.
\end{array} \tag{3.3.22}$$

These formulae provide a consistent and neat method of handling the dependency error arising in probabilistic arithmetic. In section 3.4 the numerical implementation of these bounds will be discussed.

3.4 Numerical Representation

As the examples in section 3.3.3 show, it is possible to use the formulae (3.3.12–3.3.19) to calculate the lower and upper dependency bounds numerically. However, there are difficulties associated with this scheme. For example the search for the supremum and infimum is over a continuum of points without even a bounded range. In this section new techniques for both determining the lower and upper

dependency bounds (the $\tau_{W,L}$ and $\rho_{W,L}$ operations) and the calculation of the ordinary σ_L-convolutions will be developed. The main tool used is a duality theorem relating the operations $\tau_{W,L}$ and $\rho_{W,L}$ to functions in terms of quasi-inverses of the distribution functions involved. The σ_L-convolutions can also be calculated in terms of the quasi-inverses. As well as being computationally simpler for all these calculations, this approach provides a neat solution to the problem of representation error in the numerical calculation of σ_L-convolutions. Whilst most other techniques for the numerical calculation of σ_L-convolutions of probability distributions allow one to control the error (in the sense that it approaches zero as the number of points used to represent the distribution function approaches infinity), they do not allow the calculation of rigorous bounds on the error in any given calculation. In contrast, the method presented here provides lower and upper bounds between which the probability distribution in question *must* lie.

The present section is arranged as follows. In subsection 3.4.1 we make use of a duality theorem to express the quasi-inverses of dependency bounds in terms of quasi-inverses of the distribution functions. In subsection 3.4.2 we introduce our numerical representation and develop explicit formulae, using the representation, for calculating dependency bounds. Subsection 3.4.3 contains details on how σ_L-convolutions can be calculated using this same numerical representation. In subsection 3.4.4 it is shown how functionals of distribution functions, such as moments, can be easily calculated in terms of the numerical representation, and in subsection 3.4.5 some examples are presented which demonstrate the use of all the techniques developed here.

3.4.1 Duality

The following duality theorem has been presented in a variety of contexts and in a number of different forms. It is the basis of the level-set (or α-cut) formulae for implementing the fuzzy number convolutions (see chapter 4). Frank and Schweizer [278] have considered the duality in some detail. Their results are summarised in theorem 7.7.3 of [718] which is restated below.

Definition 3.4.1 *For any F in Δ^+ let $F^\wedge$ be the left continuous quasi-inverse of F defined in section 3.2.3. Then ∇^+ is the set $\{F^\wedge | F \in \Delta^+\}$.*

Definition 3.4.2 *For any two place function L and any copula C, $\tau^\wedge_{C,L}$ is the function $\tau^\wedge_{C,L}: \nabla^+ \times \nabla^+ \mapsto \nabla^+$ given by*

$$\tau^\wedge_{C,L}(F^\wedge, G^\wedge)(x) = \inf_{C(u,v)=x} [L(F^\wedge(u), G^\wedge(v))].$$

Definition 3.4.3 *For any two place function L and any copula C, $\rho^\wedge_{C,L}$ is the function $\rho^\wedge_{C,L}: \nabla^+ \times \nabla^+ \mapsto \nabla^+$ given by*

$$\rho^\wedge_{C,L}(F^\wedge, G^\wedge)(x) = \sup_{C^d(u,v)=x} [L(F^\wedge(u), G^\wedge(v))].$$

Theorem 3.4.4 (Theorem 7.7.3 of [718]) *Let L be a function $L: \Re^+ \times \Re^+ \mapsto \Re^+$, with ∞ as its null element, and which is continuous everywhere. Let C be a copula. Then for any $F, G \in \Delta^+$,*

$$\tau_{C,L}(F,G) = [\tau^{\wedge}_{C,L}(F^{\wedge}, G^{\wedge})]^{\wedge},$$
$$\tau^{\wedge}_{C,L}(F^{\wedge}, G^{\wedge}) = [\tau_{C,L}(F,G)]^{\wedge},$$
$$\rho_{C,L}(F,G) = [\rho^{\wedge}_{C,L}(F^{\wedge}, G^{\wedge})]^{\wedge},$$
$$\rho^{\wedge}_{C,L}(F^{\wedge}, G^{\wedge}) = [\rho_{C,L}(F,G)]^{\wedge}.$$

The advantage of using the duality theorem becomes apparent when a specific L is considered. For example, if $L = \mathrm{Sum}$, then the quasi-inverse representation of the lower and upper dependency bounds can be calculated in terms of the quasi-inverses of the distribution functions of the random variables in question by using the following formulae: Let $Z = X + Y$ and let $\mathrm{ldb}^{\wedge}_{\underline{C}_{XY}}(F_Z)$ and $\mathrm{udb}^{\wedge}_{\underline{C}_{XY}}(F_Z)$ denote the quasi-inverses of $\mathrm{ldb}_{\underline{C}_{XY}}(F_X, F_Y, +)$ and $\mathrm{udb}_{\underline{C}_{XY}}(F_X, F_Y, +)$ respectively. Then

$$\mathrm{ldb}^{\wedge}_{\underline{C}_{XY}}(F^{\wedge}_X, F^{\wedge}_Y, +)(x) = \inf_{\underline{C}_{XY}(u,v)=x} (F^{\wedge}_X(u) + F^{\wedge}_Y(v)), \tag{3.4.1}$$

$$\mathrm{udb}^{\wedge}_{\underline{C}_{XY}}(F^{\wedge}_X, F^{\wedge}_Y, +)(x) = \sup_{\underline{C}^d_{XY}(u,v)=x} (F^{\wedge}_X(u) + F^{\wedge}_Y(v)). \tag{3.4.2}$$

The standard dependency bounds (with $\underline{C}_{XY} = W$) follow with the appropriate substitution. Equations 3.4.1 and 3.4.2 are simply the maximum and minimum of the pointwise sum of quasi-inverses. The functions $\mathrm{ldb}^{\wedge}_{\underline{C}_{XY}}(F^{\wedge}_X, F^{\wedge}_Y, \Box)$ and $\mathrm{udb}^{\wedge}_{\underline{C}_{XY}}(F^{\wedge}_X, F^{\wedge}_Y, \Box)$ are quasi-inverses of the lower and upper dependency bounds and not lower and upper bounds on the quasi-inverses. That is

$$\mathrm{ldb}^{\wedge}_{\underline{C}_{XY}}(F^{\wedge}_X, F^{\wedge}_Y, \Box) \geq \mathrm{udb}^{\wedge}_{\underline{C}_{XY}}(F^{\wedge}_X, F^{\wedge}_Y, \Box) \tag{3.4.3}$$

is the consequence of

$$\mathrm{ldb}_{\underline{C}_{XY}}(F_X, F_Y, \Box) \leq \mathrm{udb}_{\underline{C}_{XY}}(F_X, F_Y, \Box).$$

Note that for $L = \mathrm{Sum}$ or Difference, theorem 3.4.4 holds for any $F, G \in \Delta$ (and not just Δ^+) because in these cases the point $x = 0$ has no special significance.

When $\underline{C}_{XY} = W$ the formulae are particularly simple, and this gives us a very simple way of calculating lower and upper dependency bounds. Recalling that $W(u,v) = \max(u + v - 1, 0)$ and $W^d(u,v) = \min(u + v, 1)$ we have

$$\mathrm{ldb}^{\wedge}\left(F^{\wedge}_X, F^{\wedge}_Y, +\right)(x) = \begin{cases} \inf\limits_{u \in [x,1]} \left(F^{\wedge}_X(u) + F^{\wedge}_Y(x - u + 1)\right) & \text{if } x \neq 0, \\ \inf\limits_{u+v-1<0} \left(F^{\wedge}_X(u) + F^{\wedge}_Y(v)\right) & \text{if } x = 0. \end{cases} \tag{3.4.4}$$

Since $F^{\wedge}_X$ and $F^{\wedge}_Y$ are non-decreasing, the case for $x = 0$ becomes

$$\mathrm{ldb}^{\wedge}(F^{\wedge}_X, F^{\wedge}_Y, +)(0) = F^{\wedge}_X(0) + F^{\wedge}_Y(0).$$

Similarly,

$$\mathrm{udb}^{\wedge}\left(F_X^{\wedge}, F_Y^{\wedge}, +\right)(x) = \begin{cases} \sup\limits_{u\in[0,x]} \left(F_X^{\wedge}(u) + F_Y^{\wedge}(x-u)\right) & \text{if } x \neq 1, \\ F_X^{\wedge}(1) + F_Y^{\wedge}(1) & \text{if } x = 1. \end{cases} \tag{3.4.5}$$

(The restrictions on the range of the supremum and infimum operations arise from the fact that $v \in [0,1]$ because $\operatorname{dom} F^{\wedge} = [0,1]$.) Likewise, we have

$$\mathrm{ldb}^{\wedge}\left(F_X^{\wedge}, F_Y^{\wedge}, \times\right)(x) = \begin{cases} \inf\limits_{u\in[x,1]} \left(F_X^{\wedge}(u) \times F_Y^{\wedge}(x-u+1)\right) & \text{if } x \neq 0, \\ F_X^{\wedge}(0) \times F_Y^{\wedge}(0) & \text{if } x = 0 \end{cases} \tag{3.4.6}$$

and

$$\mathrm{udb}^{\wedge}\left(F_X^{\wedge}, F_Y^{\wedge}, \times\right)(x) = \begin{cases} \sup\limits_{u\in[0,x]} \left(F_X^{\wedge}(u) \times F_Y^{\wedge}(x-u)\right) & \text{if } x \neq 1, \\ F_X^{\wedge}(1) \times F_Y^{\wedge}(1) & \text{if } x = 1. \end{cases} \tag{3.4.7}$$

Analogous formulae for $\mathrm{ldb}^{\wedge}$ and $\mathrm{udb}^{\wedge}$ for $\square \in \{-, \div\}$ can be determined in a manner similar to that used to derive (3.3.5) and (3.3.7). Consider the quotient first. Let $Y' = 1/Y$. If F_Y has a quasi-inverse $F_Y^{\wedge}$, what is the quasi-inverse of $F_{Y'}$? It is known that $F_Y^{\wedge}(x) = y \Rightarrow x = F_Y(y)$ and $F_{Y'}(x) = 1 - F_Y(1/x)$. If $x = (1 - F_Y(1/y))$, then $F_{Y'}^{\wedge}(x) = y$. But $1 - x = F_Y(1/y)$. Therefore $F_Y^{\wedge}(1-x) = 1/y$ and so

$$F_{Y'}^{\wedge}(x) = \frac{1}{F_Y^{\wedge}(1-x)} = y.$$

Using this and (3.4.6–3.4.7), it can be shown that

$$\mathrm{ldb}^{\wedge}\left(F_X^{\wedge}, F_Y^{\wedge}, \div\right)(x) = \begin{cases} \inf\limits_{u\in[x,1]} \left(F_X^{\wedge}(u) / F_Y^{\wedge}(u-x)\right) & \text{if } x \neq 0, \\ F_X^{\wedge}(0) / F_Y^{\wedge}(1) & \text{if } x = 0, \end{cases} \tag{3.4.8}$$

and

$$\mathrm{udb}^{\wedge}\left(F_X^{\wedge}, F_Y^{\wedge}, \div\right)(x) = \begin{cases} \sup\limits_{u\in[0,x]} \left(F_X^{\wedge}(u) / F_Y^{\wedge}(1+u-x)\right) & \text{if } x \neq 1, \\ F_X^{\wedge}(1) / F_Y^{\wedge}(0) & \text{if } x = 1. \end{cases} \tag{3.4.9}$$

Similarly, it can be shown that if $Y' = -Y$ then $F_{Y'}^{\wedge}(x) = -F_Y^{\wedge}(1-x)$, and so using (3.4.4) and (3.4.5)

$$\mathrm{ldb}^{\wedge}\left(F_X^{\wedge}, F_Y^{\wedge}, -\right)(x) = \begin{cases} \inf\limits_{u\in[x,1]} \left(F_X^{\wedge}(u) - F_Y^{\wedge}(u-x)\right) & \text{if } x \neq 0, \\ F_X^{\wedge}(0) - F_Y^{\wedge}(1) & \text{if } x = 0, \end{cases} \tag{3.4.10}$$

and

$$\mathrm{udb}^{\wedge}\left(F_X^{\wedge}, F_Y^{\wedge}, -\right)(x) = \begin{cases} \sup\limits_{u\in[0,x]} \left(F_X^{\wedge}(u) - F_Y^{\wedge}(1+u-x)\right) & \text{if } x \neq 1, \\ F_X^{\wedge}(1) - F_Y^{\wedge}(0) & \text{if } x = 1. \end{cases} \tag{3.4.11}$$

These formulae for $\mathrm{ldb}^{\wedge}$ and $\mathrm{udb}^{\wedge}$ can be seen to be slightly simpler than the corresponding formulae for ldb and udb (3.3.4–3.3.7). The one disadvantage is the

Figure 3.10: The copula W.

Figure 3.11: The numerical representation of probability distributions.

special case for $x = 0$ (for $+$ and $\times$) or for $x = 1$ (for $-$ and $\div$). The reason these special cases are necessary becomes apparent upon inspection of figure 3.10 which shows the copula W. When $W(u, v) = x \neq 0$, the values u and v can take are constrained by a linear relationship. However when $W(u, v) = 0$, u and v can take any values within the cross-hatched region. When a numerical representation of the quasi-inverses is used these special cases actually disappear.

3.4.2 Numerical Representation and the Calculation of $\mathrm{ldb}^\wedge$ and $\mathrm{udb}^\wedge$

In order to use equations (3.4.4–3.4.11) to calculate lower and upper dependency bounds numerically, it is necessary to have a discrete approximation to probability distribution functions defined on a continuum. The approach taken is illustrated in figure 3.11. The distribution F is approximated by lower and upper discrete approximations denoted $\underset{\sim}{F}$ and $\tilde{F}$ respectively. These are formed by uniformly quantising F along the vertical axis (see figure 3.11). This defines the points $\underset{\sim}{x}_i$ and $\tilde{x}_i$. For a single valued F, $\underset{\sim}{x}_i = \tilde{x}_{i+1}$ for $i = 1, \ldots, n-1$. We retain the two sets of numbers $\{\underset{\sim}{x}_i\}$ and $\{\tilde{x}_i\}$ for clarity. Interval valued distributions (*i.e.* lower and upper bounds on F) can result in $\underset{\sim}{x}_i \neq \tilde{x}_{i+1}$. The end points are

$$\begin{aligned} \inf \operatorname{supp} F &= \tilde{x}_1, \\ \sup \operatorname{supp} F &= \underset{\sim}{x}_n. \end{aligned} \tag{3.4.12}$$

The above approximation of F motivates the following numerical approximation to $F^\wedge$. Both in anticipation of the exclusive use of upper and lower bounds (to contain both dependency and representation error) and to make the ideas clearer, the discrete approximations $\underset{\sim}{F}^\wedge$ and $\tilde{F}^\wedge$ to $\underline{F}^\wedge$ and $\overline{F}^\wedge$ are presented in figure 3.12. In the following discussion it is not assumed that $\underline{F}^\wedge = \overline{F}^\wedge$, although the same results do apply to this situation which is often the case at the beginning of a calculation using probabilistic arithmetic.

Figure 3.12: The quasi-inverse representation of a distribution function.

The discrete approximations $\tilde{F}^{\wedge}$ and $\underset{\sim}{F}^{\wedge}$ are defined by

$$\tilde{F}^{\wedge}(p) = \overline{F}^{\wedge}(p_i) \qquad p \in [p_i, p_{i+1}), \tag{3.4.13}$$

$$\underset{\sim}{F}^{\wedge}(p) = \underline{F}^{\wedge}(p_{i+1}) \quad p \in [p_i, p_{i+1}), \tag{3.4.14}$$

for $i = 0, \ldots, N-1$. Whether the range in these definitions is $[p_i, p_{i+1})$ or $(p_i, p_{i+1}]$ doesn't really make any difference: as long as one choice is used consistently there are no problems. N is the number of points required to represent either $\underset{\sim}{F}^{\wedge}$ or $\tilde{F}^{\wedge}$. In figure 3.12, N=6. Note that $p_N = 1$. The quantisation is uniform and so $p_i = i/N$ for $i = 0, \ldots, N-1$. Explicit array representations of $\underset{\sim}{F}^{\wedge}$ and $\tilde{F}^{\wedge}$ suitable for computer implementation are given by

$$\tilde{F}^{[\wedge]}[i] = \overline{F}^{\wedge}(p_i) = \overline{F}^{\wedge}\left(\tfrac{i}{N}\right) \tag{3.4.15}$$

and

$$\underset{\sim}{F}^{[\wedge]}[i] = \underline{F}^{\wedge}(p_{i+1}) = \underline{F}^{\wedge}\left(\tfrac{i+1}{N}\right) \tag{3.4.16}$$

for $i = 0, \ldots, N-1$. Notice the different conventions used for defining $\underset{\sim}{F}^{[\wedge]}$ and $\tilde{F}^{[\wedge]}$ in terms of $\underline{F}^{\wedge}$ and $\overline{F}^{\wedge}$. This is necessary because of the nature of the approximation. Two special values worth noting are

$$\tilde{F}^{[\wedge]}[0] = \overline{F}^{\wedge}(p_0) = \overline{F}^{\wedge}(0)$$

and

$$\underset{\sim}{F}^{[\wedge]}[N-1] = \underline{F}^{\wedge}(p_N) = \underline{F}^{\wedge}(1).$$

These two values correspond to the points a and b in figure 3.12. Regarding $\underset{\sim}{F}^{[\wedge]}[0]$, recall the convention of redefining the value of $F^{\wedge}$ at the end points of the range (see section 3.2.3). This convention resulted in $F^{\wedge}(0) = \inf \operatorname{supp} F$, and so $\tilde{F}^{[\wedge]}[0]$ is set equal to this.

An important advantage of this representation is that any representation error (difference between $\underline{F}$ and $\underset{\sim}{F}$ or between $\overline{F}$ and $\tilde{F}$) is rigorously contained within the bounds. That is, it is perfectly correct to state that $\underset{\sim}{F} \leq F \leq \tilde{F}$ or $\underset{\sim}{F}^{\wedge} \geq F^{\wedge} \geq \tilde{F}^{\wedge}$

although this does not bound F as tightly as stating $\underline{F} \leq F \leq \overline{F}$. By always performing any necessary rounding approximations in a manner such that the width between $\underset{\sim}{F}$ and $\tilde{F}$ is made larger ("outwardly directed rounding"), this property is preserved. This is referred to below as the preservation of the representation error containment property. The representation error for any given approximation can be made arbitrarily small by increasing N. The method of approximation proposed here is better than simply using a single function F_{approx} to approximate F for which $|F_{approx}(x) - F(x)|$ is "small" for "most" x.

Now that a numerical representation has been chosen, it is necessary to derive the appropriate formulae for $\mathrm{ldb}^{[\wedge]}$ and $\mathrm{udb}^{[\wedge]}$, the lower and upper dependency bounds in terms of the numerical representation. We explicitly derive the four cases of the lower and upper dependency bounds for sums and differences.

(i) $\mathrm{ldb}^{[\wedge]}(\underset{\sim}{F}_X^{[\wedge]}, \underset{\sim}{F}_Y^{[\wedge]}, +)$

Equations 3.4.4 and 3.3.22 give

$$\mathrm{ldb}^{\wedge}(\underline{F}_X^{\wedge}, \underline{F}_X^{\wedge}, +)(x) = \inf_{u \in [x,1]} (\underline{F}_X^{\wedge}(u) + \underline{F}_Y^{\wedge}(x - u + 1)) \qquad (x \neq 0).$$

Let $x = \frac{i+1}{N}$ and $u = \frac{j+1}{N}$. Then

$$\begin{aligned} \mathrm{ldb}^{\wedge}\left(\underline{F}_X^{\wedge}, \underline{F}_Y^{\wedge}, +\right)\left(\tfrac{i+1}{N}\right) &= \inf_{\left(\frac{j+1}{N}\right) \in \left[\frac{i+1}{N}, 1\right]} \left(\underline{F}_X^{\wedge}\left(\tfrac{j+1}{N}\right) + \underline{F}_Y^{\wedge}\left(\tfrac{i+1}{N} - \tfrac{j+1}{N} + 1\right)\right) \quad \left(\tfrac{i+1}{N} \neq 0\right) \\ &= \inf_{j \in [i, N-1]} \left(\underline{F}_X^{\wedge}\left(\tfrac{j+1}{N}\right) + \underline{F}_Y^{\wedge}\left(\tfrac{i-j+N}{N}\right)\right) \quad (i \neq -1). \end{aligned}$$

Using the correspondences (3.4.15) and (3.4.16), this can be written as

$$\mathrm{ldb}^{[\wedge]}(\underset{\sim}{F}_X^{[\wedge]}, \underset{\sim}{F}_Y^{[\wedge]}, +)[i] = \inf_{j=i,\ldots,N-1} (\underset{\sim}{F}_X^{[\wedge]}[j] + \underset{\sim}{F}_Y^{[\wedge]}[i - j + N - 1]) \quad (i \neq -1). \quad (3.4.17)$$

Since it is only required to calculate $\mathrm{ldb}^{[\wedge]}[i]$ for $i = 0, \ldots, N-1$, the special case for $x = 0$ in (3.4.4) has been avoided (as we mentioned was possible at the end of section 3.4.1).

(ii) $\mathrm{udb}^{[\wedge]}(\underset{\sim}{F}_X^{[\wedge]}, \underset{\sim}{F}_Y^{[\wedge]}, +)$

Equations (3.4.5) and (3.3.22) give

$$\mathrm{udb}^{\wedge}\left(\overline{F}_X^{\wedge}, \overline{F}_Y^{\wedge}, +\right)(x) = \sup_{u \in [0,x]} \left(\overline{F}_X^{\wedge}(u) + \overline{F}_Y^{\wedge}(x - u)\right) \quad (x \neq 1)$$

Let $x = \frac{i}{N}$ and $u = \frac{j}{N}$. Then

$$\mathrm{udb}^{\wedge}\left(\overline{F}_X^{\wedge}, \overline{F}_Y^{\wedge}, +\right)\left(\tfrac{i}{N}\right) = \sup_{\left(\frac{j}{N}\right) \in \left[0, \frac{i}{N}\right]} \left(\overline{F}_X^{\wedge}\left(\tfrac{j}{N}\right) + \overline{F}_Y^{\wedge}\left(\tfrac{i}{N} - \tfrac{j}{N}\right)\right) \qquad \left(\tfrac{i}{N} \neq 1\right).$$

Using the correspondences (3.4.15) and (3.4.16) gives

$$\mathrm{udb}^{[\wedge]}\left(\tilde{F}_X^{[\wedge]},\tilde{F}_Y^{[\wedge]},+\right)[i]=\sup_{j=0,\ldots,i}\left(\underset{\sim}{F}_X^{[\wedge]}[j]+\underset{\sim}{F}_Y^{[\wedge]}[i-j]\right)\quad(i\neq N),\tag{3.4.18}$$

for $i=0,\ldots,N-1$ and the special case $(x=1)$ has been avoided.

(iii) $\mathrm{ldb}^{[\wedge]}(\underset{\sim}{F}_X^{[\wedge]},\tilde{F}_Y^{[\wedge]},-)$

From (3.4.10) and (3.3.23),

$$\mathrm{ldb}^{\wedge}\left(\underline{F}_X^{\wedge},\overline{F}_Y^{\wedge},-\right)(x)=\inf_{u\in[x,1]}\left(\underline{F}_X^{\wedge}(u)-\overline{F}_Y^{\wedge}(u-x)\right)\quad(x\neq 0).$$

Set $x=\frac{i+1}{N}$ and $u=\frac{j+1}{N}$. Then

$$\mathrm{ldb}^{\wedge}\left(\underline{F}_X^{\wedge},\overline{F}_Y^{\wedge},-\right)\left(\tfrac{i+1}{N}\right)=\inf_{\left(\frac{j+1}{N}\right)\in\left[\frac{i+1}{N},1\right]}\left(\underline{F}_X^{\wedge}\left(\tfrac{j+1}{N}\right)-\overline{F}_Y^{\wedge}\left(\tfrac{j+1}{N}-\tfrac{i+1}{N}\right)\right)\quad\left(\tfrac{i+1}{N}\neq 0\right).$$

Using the correspondences (3.4.15) and (3.4.16)

$$\mathrm{ldb}^{[\wedge]}\left(\underset{\sim}{F}_X^{[\wedge]},\tilde{F}_Y^{[\wedge]},-\right)[i]=\inf_{j=i,\ldots,N-1}\left(\underset{\sim}{F}_X^{[\wedge]}[j]-\tilde{F}_Y^{[\wedge]}[j-i]\right)\quad(i\neq -1),\tag{3.4.19}$$

which is all that is required for $i=0,\ldots,N-1$.

(iv) $\mathrm{udb}^{[\wedge]}(\tilde{F}_X^{[\wedge]},\underset{\sim}{F}_Y^{[\wedge]},-)$

Using (3.4.11) and (3.3.23),

$$\mathrm{udb}^{\wedge}\left(\overline{F}_X^{\wedge},\underline{F}_Y^{\wedge},-\right)(x)=\sup_{u\in[0,x]}\left(\overline{F}_X^{\wedge}(u)-\underline{F}_Y^{\wedge}(u-x+1)\right)\quad(x\neq 1).$$

Setting $x=\frac{i}{N}$ and $u=\frac{j}{N}$ gives

$$\mathrm{udb}^{\wedge}\left(\overline{F}_X^{\wedge},\underline{F}_Y^{\wedge},-\right)\left(\tfrac{i}{N}\right)=\sup_{\left(\frac{j}{N}\right)\in\left[0,\frac{i}{N}\right]}\left(\overline{F}_X^{\wedge}\left(\tfrac{j}{N}\right)-\underline{F}_Y^{\wedge}\left(\tfrac{j}{N}-\tfrac{i}{N}+1\right)\right)\quad\left(\tfrac{i}{N}\neq 1\right),$$

and thus

$$\mathrm{udb}^{[\wedge]}\left(\tilde{F}_X^{[\wedge]},\underset{\sim}{F}_Y^{[\wedge]},-\right)[i]=\sup_{j=0,\ldots,i}\left(\tilde{F}_X^{[\wedge]}[j]-\underset{\sim}{F}_Y^{[\wedge]}[j-i+N-1]\right)\quad(i\neq N),\tag{3.4.20}$$

which holds for $i=0,\ldots,N-1$.

The analogous formulae for product and quotient are

$$\mathrm{ldb}^{[\wedge]}(\underset{\sim}{F}_X^{[\wedge]},\underset{\sim}{F}_Y^{[\wedge]},\times)[i]=\inf_{j=i,\ldots,N-1}(\underset{\sim}{F}_X^{[\wedge]}[j]\times\underset{\sim}{F}_Y^{[\wedge]}[i-j+N-1])\quad(i\neq -1).\tag{3.4.21}$$

$$\mathrm{udb}^{[\wedge]}\left(\tilde{F}_X^{[\wedge]},\tilde{F}_Y^{[\wedge]},\times\right)[i]=\sup_{j=0,\ldots,i}\left(\tilde{F}_X^{[\wedge]}[j]\times\tilde{F}_Y^{[\wedge]}[i-j]\right)\quad(i\neq N),\tag{3.4.22}$$

$$\mathrm{ldb}^{[\wedge]}\left(\underset{\sim}{F}_X^{[\wedge]}, \tilde{F}_Y^{[\wedge]}, \div\right)[i] = \inf_{j=i,\ldots,N-1}\left(\underset{\sim}{F}_X^{[\wedge]}[j]/\tilde{F}_Y^{[\wedge]}[j-i]\right) \quad (i \neq -1), \tag{3.4.23}$$

$$\mathrm{udb}^{[\wedge]}\left(\tilde{F}_X^{[\wedge]}, \underset{\sim}{F}_Y^{[\wedge]}, \div\right)[i] = \sup_{j=0,\ldots,i}\left(\tilde{F}_X^{[\wedge]}[j]/\underset{\sim}{F}_Y^{[\wedge]}[j-i+N-1]\right) \quad (i \neq N), \tag{3.4.24}$$

All these hold for $i = 0, \ldots, N-1$. The use of lower and upper bounds alone (instead of distributions within the bounds), and whether $\underset{\sim}{F}_X^{[\wedge]}$ or $\tilde{F}_X^{[\wedge]}$ should be used in a particular instance, is discussed at the end of section 3.3.4.

Two significant points to note about the above formulae for $\mathrm{ldb}^{[\wedge]}$ and $\mathrm{udb}^{[\wedge]}$ are their low computational complexity and their lack of approximation error. To calculate a dependency bound with N points in terms of two N point discrete approximations requires only $O(N^2)$ operations. The results are free from error in the two senses of any representation error being rigorously bounded by $\underset{\sim}{F}_X^{[\wedge]}$ and $\tilde{F}_X^{[\wedge]}$ (see above), and the supremum and infimum operations are exact (compared with the numerical implementation of the formulae developed in section 3.3 for which approximation is required).

The final point which needs to be settled is how to generate $\underset{\sim}{F}_X^{[\wedge]}$ and $\tilde{F}_X^{[\wedge]}$ from a given exact distribution F_X. This is in fact quite simple. One just sets $\underline{F}_X^{\wedge}(p) = \overline{F}_X^{\wedge}(p) = F_X^{\wedge}(p)$ for $p \in [0,1)$ and uses the definitions (3.4.15) and (3.4.16). The only difficulty is finding $F_X^{\wedge}(p)$ given a formula for $F_X(x)$. While for some simple distributions (such as the uniform and triangular distributions) an exact formula for $F_X^{\wedge}(p)$ can be derived, in general one has to perform a numerical search. This will give $F_X^{\wedge}(p)$ to any desired accuracy. This method is described, along with various techniques for numerically calculating F_X for some common distributions, in chapter 5 of [451]. Some examples are given later. When the distribution F_X has unbounded support, it is necessary to curtail the distribution so that $\tilde{F}_X^{[\wedge]}[0]$ and $\underset{\sim}{F}_X^{[\wedge]}[N-1]$ are finite.

3.4.3 Numerical Calculation of σ-Convolutions

We now examine the calculation of σ-convolutions in terms of our numerical representation. First we determine whether for all $F_X \in [\underline{F}_X, \overline{F}_X]$ and $F_Y \in [\underline{F}_Y, \overline{F}_Y]$,

$$\sigma_L(\underline{F}_X, \underline{F}_Y) \leq \sigma_L(F_X, F_Y) \leq \sigma_L(\overline{F}_X, \overline{F}_Y). \tag{3.4.25}$$

It will be seen that this is indeed the case for arithmetic operations defined on $\Re^+$.

To see that (3.4.25) holds for addition and multiplication, it suffices to inspect the convolution relations which can be written

$$F_Z(x) = \int_0^x F_X(x-t)\, dF_Y(t)$$

and

$$F_Z(x) = \int_0^x F_X(x/t)\, dF_Y(t)$$

respectively [775]. If F_Y is considered as fixed, then since F_X is always greater than 0, it is apparent that

$$\underline{F}_Z(x) = \int_0^x \underline{F}_X(x-t)\, dF_Y(t)$$

is always less than or equal to F_Z. Likewise for upper bounds $\overline{F}_X$, and, by symmetry, for F_Y as well. In the same manner it can be seen that

$$\underline{F}_Z(x) = \int_0^x \underline{F}_X(x/t)\, dF_Y(t)$$

is always less than or equal to F_Z if F_Y is considered fixed. Again by symmetry this holds for upper bounds and for F_Y as well. Similar arguments could be used to show that analogous results hold for difference and quotient convolutions. In these cases it is necessary to use the lower or upper bounds on F_X and F_Y in a manner similar to equation 3.3.22. The details are omitted. As well as showing that only the lower and upper bounds on a distribution need be considered, the above result comes in useful in the "condensation" procedure which is part of the algorithm for calculating $\sigma_L^{[\wedge]}\left(F_X^{[\wedge]}, F_Y^{[\wedge]}\right)$.

The method of calculating $\sigma_L^{[\wedge]}\left(F_X^{[\wedge]}, F_Y^{[\wedge]}\right)$ is best presented in terms of the discrete frequency functions corresponding to $F_{\tilde{X}}^{[\wedge]}$ and $F_{\tilde{Y}}^{[\wedge]}$. Thus for the moment, we will ignore the use of lower and upper distributions, and we will consider the calculation of $\underline{\sigma}_+^{[\wedge]}\left(F_X^{[\wedge]}, F_Y^{[\wedge]}\right)$. Other operations L and the complications of using lower and upper distributions are considered later on in this subsection.

Given $F_{\tilde{X}}^{[\wedge]}$ and $F_{\tilde{Y}}^{[\wedge]}$ of N points each, there are corresponding discrete frequency functions f_X and f_Y given by

$$f_X(x) = \begin{cases} \frac{k}{N} & \text{if } x = F_{\tilde{X}}^{[\wedge]}[i] \text{ for } k \text{ different } i, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y(x) = \begin{cases} \frac{\ell}{N} & \text{if } x = F_{\tilde{Y}}^{[\wedge]}[i] \text{ for } \ell \text{ different } i, \\ 0 & \text{otherwise.} \end{cases}$$

The discrete frequency function of $Z = X + Y$ is obviously given by

$$f_Z(x) = \begin{cases} \frac{m}{N^2} & \text{if } x = F_{\tilde{Y}}^{[\wedge]}[i] + F_{\tilde{Y}}^{[\wedge]}[j] \text{ for } m \text{ different pairs of } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4.26}$$

This formula for F_Z follows directly from the laws of probability and has been the basis for a number of methods for calculating convolutions in terms of discrete frequency functions [423].

We thus have the following simple algorithm for calculating $F_Z^{[\wedge]}$ (expressed in the syntax of the C language [453], with the exception mentioned in section 1.4):

$$\begin{array}{l} \textbf{for}(i := 0; i < N; i{+}{+})\{ \\ \quad \textbf{for}(j := 0; j < N; j{+}{+})\{ \\ \quad\quad F_Z^{[\wedge],U}[i + j * N] := F_X^{[\wedge]}[i] + F_Y^{[\wedge]}[j]; \\ \quad \} \\ \} \\ F_Z^{[\wedge],E} := sort\left(F_Z^{[\wedge],U}\right); \end{array} \tag{3.4.27}$$

Figure 3.13: Illustration of the condensation procedure.

The array $F_Z^{[\wedge],U}$ is unsorted (hence the U) and is of size N^2. The array $F_Z^{[\wedge],E}$ is the result of sorting $F_Z^{[\wedge],U}$ into increasing order (the E stands for Exact). Observe that it is quite possible for $F_Z^{[\wedge],E}[i] = F_Z^{[\wedge],E}[j]$ for $i \neq j$. This does not matter at all: it simply results in a bigger jump in the corresponding F_Z. An important fact to notice is that while $F_X^{[\wedge]}$ and $F_Y^{[\wedge]}$ have only N points, $F_Z^{[\wedge],E}$ has N^2. This will obviously cause severe difficulties if we intend to perform a sequence of operations: the output of each operation will be an array considerably larger than its input. What is required is a method for reducing the size of $F_Z^{[\wedge],E}$ to N without introducing any error into the result. The procedure used to do this is called *condensation*.

If F_Z (or $F_Z^{\wedge}$) were simply represented by a single discrete version $F_Z^{[\wedge]}$ then it would be impossible to approximate $F_Z^{[\wedge],E}$ by an array of N points without error. However as both lower and upper approximations are available, "directed rounding" can be used to produce an approximation to $F_Z^{[\wedge],E}$ that is not in error in the sense that the quasi-inverse of the distribution is contained within the bounds. That is, it will be true that

$$\underset{\sim}{F}_Z^{[\wedge]} \geq \underset{\sim}{F}_Z^{[\wedge],E} \geq F_Z^{\wedge} \geq \tilde{F}_Z^{[\wedge],E} \geq \tilde{F}_Z^{[\wedge]}$$

where the inequality is "$\geq$" (and not "$\leq$") because we are talking about quasi-inverses. The method of determining $\underset{\sim}{F}_Z^{[\wedge]}$ and $\tilde{F}_Z^{[\wedge]}$ is indicated in figure 3.13. Only the procedure for the lower bound is shown there. The rule for condensation is simply

$$\underset{\sim}{F}_Z^{[\wedge]}[i] := \underset{\sim}{F}_Z^{[\wedge,E]}[i \times N + N - 1].$$

The analogous condensation with "upward rounding" for the upper bound is

$$\tilde{F}_Z^{[\wedge]}[i] := \tilde{F}_Z^{[\wedge,E]}[i \times N].$$

The overall algorithm (for the distribution of the sum of random variables) becomes algorithm 3.4.28 (see below). The algorithm for products is identical apart from

```
for(i := 0; i < N; i++){
    for(j := 0; j < N; j++){
        F_Z^[∧,U][i + j * N] := F_X^[∧][i] + F_Y^[∧][j];
        F̃_Z^[∧,U][i + j * N] := F̃_X^[∧][i] + F̃_Y^[∧][j];
    }
}
F_Z^[∧,E] := sort(F_Z^[∧,U]);
F̃_Z^[∧,E] := sort(F̃_Z^[∧,U]);
for(i := 0; i < N; i++){
    F_Z^[∧][i] := F_Z^[∧,E][i * N + N − 1];
    F̃_Z^[∧][i] := F̃_Z^[∧,E][i * N];
}
```

(3.4.28)

Algorithm 3.4.28

a replacement of the "+" on the right hand side of lines 3 and 4 by a "$\times$". For subtraction, lines 3 and 4 become

$$\begin{aligned} \underset{\sim}{F}_Z^{[\wedge,U]}[i + j * N] &:= \underset{\sim}{F}_X^{[\wedge]}[i] - \tilde{F}_Y^{[\wedge]}[j]; \\ \tilde{F}_Z^{[\wedge,U]}[i + j * N] &:= \tilde{F}_X^{[\wedge]}[i] - \underset{\sim}{F}_Y^{[\wedge]}[j];. \end{aligned}$$

The algorithm for quotients is obtained by replacing the subtraction in the above two lines by a division. Some examples using these algorithms are presented in section 3.4.5.

3.4.4 Calculation of Moments and other Functionals

We now consider the calculation of moments and other functionals of F_X in terms of $\underset{\sim}{F}_X^{[\wedge]}$ and $\tilde{F}_X^{[\wedge]}$. Using the change of variables result in lemma 4.4.6 of [718] we have

$$m_X^{(k)} = \int_{\Re} x^k \, dF_X(x) = \int_0^1 \left(F_X^{\wedge}(t)\right)^k \, dt \tag{3.4.29}$$

and

$$\mu_X^{(k)} = \int_{\Re} (x - \mu_X)^k \, dF_X(x) = \int_0^1 \left(F_X^{\wedge}(t) - \mu_X\right)^k \, dt, \tag{3.4.30}$$

where μ_X is the mean and $\mu_X = m_X^{(1)}$. These equations can be used to calculate moments in terms of $\underset{\sim}{F}_X^{[\wedge]}$ and $\tilde{F}_X^{[\wedge]}$ by replacing the integral by the appropriate

summation and by realising that we will only be able to calculate lower and upper bounds on the moments. Regarding the lower and upper bounds, it is easy to see that if

$$\underline{M}[i] = \min\left(\underline{F}_X^{[\wedge]}[i] - \underline{\mu}_X, \tilde{F}_X^{[\wedge]}[i] - \underline{\mu}_X, \underline{F}_X^{[\wedge]}[i] - \overline{\mu}_X, \tilde{F}_X^{[\wedge]}[i] - \overline{\mu}_X\right),$$

$$\overline{M}[i] = \max\left(\underline{F}_X^{[\wedge]}[i] - \underline{\mu}_X, \tilde{F}_X^{[\wedge]}[i] - \underline{\mu}_X, \underline{F}_X^{[\wedge]}[i] - \overline{\mu}_X, \tilde{F}_X^{[\wedge]}[i] - \overline{\mu}_X\right),$$

and $M[i] = \left[\underline{M}[i], \overline{M}[i]\right]$, then

$$\begin{aligned} \underline{\mu}_X^{(k)} &= \frac{1}{N}\sum_{i=0}^{N-1} \underline{(M[i])^k} \\ \overline{\mu}_X^{(k)} &= \frac{1}{N}\sum_{i=0}^{N-1} \overline{(M[i])^k} \end{aligned} \tag{3.4.31}$$

will be lower and upper bounds on $\mu_X^{(k)}$ for $k = 2, \ldots$. The operators $\underline{(.)^k}$ and $\overline{(.)^k}$ are the interval arithmetic lower and upper bounds for exponentiation by an integral power. If $W = [\underline{W}, \overline{W}]$, then

$$W^k = \begin{cases} [\underline{W}^k, \overline{W}^k] & \text{if } \underline{W} \geq 0 \text{ or } k \text{ is odd,} \\ [\overline{W}^k, \underline{W}^k] & \text{if } \overline{W} < 0 \text{ and } k \text{ is even,} \\ [0, |W|^k] & \text{if } 0 \in W \text{ and } k \text{ is even,} \end{cases}$$

where $|W| = \max(|\underline{W}|, |\overline{W}|)$ [584].

The use of these equations is discussed in section 3.4.5 with reference to some examples. Functionals other than moments could be calculated in an analogous manner.

3.4.5 Examples

As an illustration of the ideas developed above we now present some examples. In all of these, N (the number of sample values of the lower or upper quasi-inverses) is 50. All the results presented here are outputs of computer programs which implement the formulae derived above.

Figures 3.14 and 3.15 show the distributions arising from the numerical representation of Y and Z respectively, where Y is uniformly distributed on $[1, 2]$ and Z has a Gaussian distribution with $\mu = 4, \sigma = 1$ curtailed to $\pm 3\sigma$. These were generated by using the procedure described in section 3.4.2. Note that in each case the lower and upper distributions touch each other at the fifty sample points. Figure 3.16 shows the distribution of $X = Y/Z$ calculated using the algorithm for $\sigma_+^{[\wedge]}$ derived earlier. Notice that the lower and upper distributions no longer touch. This is due to the outwardly directed rounding in the condensation procedure. The magnitude of this effect can be seen from figures 3.17 and 3.18 where we present the distribution calculated directly from the exact quasi-inverse of a triangular distribution centred at 3 with spread of ± 1, and that of the distribution obtained using the algorithm

Figure 3.14: Numerical representation of F_Y (uniform distribution on $[1,2]$) with $N = 50$. The moments as calculated by (3.4.31) are $\mu = [1.49, 1.51]$, $\sigma^2 = [0.737, 0.937]$, $\mu^{(3)} = [-0.00501, 0.00501]$, and $\mu^{(4)} = [0.0101, 0.0152]$.

Figure 3.15: Numerical representation of F_Z, a Gaussian distribution with $\mu = 4$ and $\sigma^2 = 1$, and which is curtailed at $\mu \pm 3\sigma$. It is represented with $N = 50$. The values of the moments as calculated by (3.4.31) are $\mu = [3.94, 4.06]$, $\sigma^2 = [0.7538, 1.3058]$, $\mu^{(3)} = [-0.7243, 0.7243]$, and $\mu^{(4)} = [1.52, 5.583]$.

Figure 3.16: Numerical representation of F_X where $X = Y/Z$, and Y and Z are as in figures 3.14 and 3.15 respectively. This was calculated using the σ-convolution algorithm.

Figure 3.17: Numerical representation of an exact triangular distribution centred at 3 with a spread of ± 1.

Figure 3.18: Output of the σ_+-convolution algorithm for two random variables uniformly distributed on $[1, 2]$. This is exactly the same as figure 3.17 apart from the effect of the representation error in the inputs and the outwardly directed rounding in the condensation procedure.

Figure 3.19: Lower and upper dependency bounds for the same calculation as for figure 3.16. The values of the moments as calculated by (3.4.31) are $\mu = [0.2733, 0.5605]$, $\sigma^2 = [0.00, 0.1935]$, $\mu^{(3)} = [-0.028, 0.1383]$, and $\mu^{(4)} = [0.000, 0.2025]$.

for $\sigma_+^{[\wedge]}$ for the sum of two uniformly distributed random variables on $[1, 2]$. The difference between this representation width and the dependency width can be seen by examining figure 3.19 where we present the lower and upper dependency bounds for the same calculation as for figure 3.16.

The effect of the difference between lower and upper distributions on the tightness of the bounds for the moments can be seen by comparing the various values presented. For cases such as that presented in figure 3.19, the bounds on the moments are very loose, especially for the higher order moments. This is to be expected as one can fit a very wide range of distributions between the lower and upper bounds. Thus the bounds for moments would appear to be of little value.

3.5 Incorporating and Updating Dependency Information

3.5.1 General Idea

So far have shown how to calculate lower and upper bounds on the distribution of some arithmetic operation on random variables when it is either known that they are completely independent, or there is no knowledge of their dependency at all. However there are situations between these two extremes. These are now considered. It will be seen that the algorithms rapidly become more complicated as more information is taken into account. The methods involved can be organised according to the following classification:

1. Using the τ, ρ and σ operations.
2. Pairwise bounds on copulas or pairwise measures of dependence.
3. Pairwise joint distributions.
4. Either joint distributions of all the variables or combination of all the variables at once (rather than pairwise combinations).

So far in this paper we have only considered type 1. In this section, type 2, and in less detail, type 3, will be considered. Type 4 is generally intractable (but see the section on graph theory in chapter 4).

The fact that a lower bound on a copula C_{XY} other than W can be used to calculate dependency bounds was shown in section 3.3.2. The lower bound $\underline{C}_{XY}$ could describe bounds on the dependency of some input random variables, or it could arise in the process of a probabilistic arithmetic calculation. Dependency bounds based on $\underline{C}_{XY} \geq W$ are closer together than those based on W. In this section we will briefly examine different lower bounds on copulas and their effect on the dependency bounds. We will also consider the determination and interpretation of lower bounds other than W.

3.5.2 Interpretation of Dependencies Implied by $\underline{C}_{XY} \neq W$

The effect of a $\underline{C}_{XY}$ other than W is illustrated in figure 3.20 where lower and upper dependency bounds for the sum of X and Y with $\underline{C}_{XY} = T_p$ are presented. Here X and Y are both uniformly distributed on $[0,1]$, and T_p is the parameterised t-norm discussed on pp.72–73 of [718] and given by

$$\begin{aligned} T_p(x,y) &= (\max(x^p + y^p - 1, 0))^{1/p} \quad (p \neq 0), \\ T_0(x,y) &= xy \end{aligned} \tag{3.5.1}$$

for $p \in (-\infty, 1]$. This t-norm is related to W, Π , and M by $T_1 = W$, $T_0 = \Pi$, and $\lim_{p \to -\infty} T_p = M$.

Whilst it is possible to calculate these more general dependency bounds, it is obviously important that they have a useful interpretation and that the dependency

Figure 3.20: Lower and upper dependency bounds for the sum of two random variables each uniformly distributed on $[0, 1]$ when the lower bound on their connecting copula is given by T_p for various p.

induced by $\underline{C}_{XY} \neq W$ can be understood. The family of copulas $\{T_p |\ p \in (-\infty, 1]\}$ is in fact only one of a number of parameterised copulas that provide an infinite number of copulas between W and M [718]. All these parameterised copulas are in fact t-norms. (They have been also used in fuzzy set theory as generalised intersection operators [94,234,465].) Thus the first question we should ask is what probabilistic interpretation can be ascribed to copulas that are also t-norms; *i.e.* what does associativity of a copula imply about the dependence of random variables? Schweizer and Sklar asked this question in problem 6.7.2 of [718]. A simple argument presented in chapter 5 shows that if a copula C_{XY} is associative and satisfies $C_{XY}(a,a) < a\ \forall a \in (0,1)$ (*i.e.* it is *Archimedean*), then it is an increasing function of a joint distribution of *independent* random variables: $C_{XY}(x,y) = h(F_{UV}(x,y))$ (U and V are independent). However it is not apparent what this means probabilistically. Nevertheless such parameterised copulas have been used in practice. See for example [150,162,163,297–299,627]

We are actually interested in these parameterised copulas for their role as *lower bounds* on the actual unknown copula. We can understand their effect to an extent by making use of the following result which is given by Jogdeo [412].

Theorem 3.5.1 *Let F and G be two bivariate distribution functions such that $F(x,y) \leq G(x,y)\ \forall x,y \in \Re$. Then for every pair of nondecreasing functions f and g defined on $\Re$,*

$$\mathrm{cov}_G[f(X), g(Y)] \ \geq\ \mathrm{cov}_F[f(X), g(Y)],$$

where $\mathrm{cov}_F[f(X), g(Y)]$ is the covariance of $f(X)$ and $g(Y)$ given that X and Y have joint distribution F.

This says that stochastic ordering of F and G implies stochastic ordering of the associated covariances. For our purposes it allows us to say that

$$\mathrm{cov}_{C_{XY}}(X,Y) \ \geq\ \mathrm{cov}_{\underline{C}_{XY}}(X,Y)$$

and so $\mathrm{cov}_{\underline{C}_{XY}}(X,Y)$ is a lower bound on the covariance of X and Y. By normalising the covariance appropriately we can calculate lower bounds on the correlation coefficient of X and Y implied by $\underline{C}_{XY}$. This gives us an intuitive feel for the dependence implied by $\underline{C}_{XY}$.

The lower bounds on covariance can be calculated as follows. We have the formulae

$$\begin{aligned}\mathrm{cov}(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (F_{XY}(u,v) - F_X(u)F_Y(v))\, du\, dv \\ &= \int_0^1\int_0^1 (C_{XY}(u,v) - uv)\, dF_X^{-1}(u)\, dF_Y^{-1}(v)\end{aligned}$$

(see [388,719]). If we just consider uniform marginals $F_X = F_Y = U_{0,1}$, then the lower bound on the correlation coeeficient $r(x,y)$ is given by

$$\underline{r}(X,Y) = \frac{1}{D(X)D(Y)} \int_0^1\int_0^1 (\underline{C}_{XY}(u,v) - uv)\, du\, dv,$$

p	$\underline{r}(X,Y)$
1	-1
$\frac{1}{2}$	$\frac{-7}{15} \approx -0.4667$
0	0
-1	$4\pi^2 - 39 \approx 0.4784$
-2	$9 - 12\ln 2 \approx 0.6828$
$-\infty$	1

Table 3.1: Lower bounds on the correlation coefficient implied by lower bounds of the form T_p on the connecting copula for various p.

where $D(X)$ is the standard deviation of X. For $F_X = F_Y = U_{0,1}$, $D(X) = 1/\sqrt{12}$. This will be a lower bound on r *only* for uniform marginals. Other marginal distributions will give different results. (This is a failing of the correlation coefficent as an index of dependence: it is not invariant under transformations of the marginals [719].) If $\underline{C}_{XY} = T_p$,

$$\begin{aligned} \underline{r}(X,Y) &= 12\int_0^1\int_0^1 \left((\max(x^p + y^p - 1, 0))^{1/p} - xy\right) dx\, dy \qquad p \in (-\infty, 1] \\ &= 12(I - \tfrac{1}{4}), \end{aligned}$$

where

$$I = \int_0^1\int_0^1 (\max(x^p + y^p - 1, 0)^{1/p}\, dx\, dy.$$

The integral I can be calculated exactly for $p = 1, \frac{1}{2}, 0, -1, -2, -\infty$ with the aid of [336]. For example, with $p = -1$, we have $x^{-1} + y^{-1} > 1\ \forall (x,y) \in [0,1]^2$ and so

$$I = \int_0^1 J(y)dy.$$

Using 2.152(1) of [336] we obtain

$$\begin{aligned} J(y) &= \int_0^1 \left(\frac{1}{x} + \frac{1}{y} - 1\right)^{-1} dx \\ &= \frac{y}{1-y} + \frac{y^2}{(1-y)^2}\ln(y). \end{aligned}$$

The integral I can then be evaluated (using 4.2361 and 4.2313 of [336]) to give $I = \pi^2/3 - 3$ and thus $\underline{r}(X,Y) = 4\pi^2 - 39$. the other cases are determined similarly and the results are summarised in table 3.1. Bounds on other indices of dependence can also be calculated. For example, Genest and Mackay [298] have shown that Kendall's τ is given by $\frac{p}{2-p}$ for T_p. Note that the interpretation of $\underline{C}_{XY} = \Pi$ is particularly simple: it says that X and Y are *positively quadrant dependent* (see [246, 258,313,459,506]). Statistical tests for positive quadrant dependence are considered in [506].

3.5.3 Use of $\underline{C}_{XY} \neq W$ in Probabilistic Arithmetic Calculations

In order to use lower bounds on the connecting copula to calculate tighter dependency bounds it is necessary to calculate the dependency arising through the course of a probabilistic arithmetic calculation. That is, given random variables W, X and Y, with $Z = X \Box Y$ for some arithmetic operation $\Box$, calculate $\underline{C}_{ZX}$, $\underline{C}_{ZY}$ and $\underline{C}_{ZW}$ in terms of F_X, F_Y and $\underline{C}_{XY}$ (*cf.* [684]). If there were other variables U, V *etc.*, it would be necessary to calculate $\underline{C}_{ZU}$ and $\underline{C}_{ZV}$ *etc.* Since the calculations would be of the same form as those for $\underline{C}_{ZW}$ this is not considered further. We will consider addition ($\Box = +$) here. The details for other operations are similar.

The joint distribution of Z and X is given by

$$F_{ZX}(z,w) = \iint\limits_{D_{zw}} dF_{XY}(x,y), \tag{3.5.2}$$

where D_{zw} is the region in the (x,y) plane such that $x+y<z$ and $x<w$ (see [632, pp.201ff]). Thus

$$F_{ZX}(z,w) = \iint\limits_{D_{zw}} dC_{XY}(F_X(x), F_Y(y))$$

and so

$$C_{ZX}(u,v) = \int\int\limits_{D_{F_Z^{\wedge}(u), F_X^{\wedge}(v)}} dC_{XY}(F_X(x), F_Y(y)). \tag{3.5.3}$$

This is very similar to a σ-convolution. Using an argument along the lines of that used to prove theorem 3.3.1, it is clear that

$$\underline{C}_{ZX}(u,v) = \sup_{\substack{x+y=F_Z^{\wedge}(u) \\ y \leq F_X^{\wedge}(v)}} \left[\underline{C}_{XY}(F_X(x), F_Y(y))\right]. \tag{3.5.4}$$

It is possible to use $\underline{C}_{XY}$ rather than C_{XY} by an argument similar to that used to prove theorem 3.3.4. Equation 3.5.4 is not as complex computationally as it seems. For example, given a value of $t = \underline{C}_{ZX}(u,v)$, it is easy to calculate $\underline{C}_{ZX}(u, v+\delta)$ by

$$\underline{C}_{ZX}(u, v+\delta) = \max \left[\sup_{\substack{F_X^{\wedge}(v) < y < F_X^{\wedge}(v+\delta) \\ x+y=F_Z^{\wedge}(u)}} \underline{C}_{XY}(F_X(x), F_Y(y)), \;\; \underline{C}_{ZX}(u,v) \right]. \tag{3.5.5}$$

Obviously $\underline{C}_{ZY}$ can be calculated in a similar manner.

The calculation of $\underline{C}_{ZW}$ is rather more complicated. The joint distribution of Z and W is given by

$$\begin{aligned} F_{ZW}(z,w) &= \iiint\limits_{\substack{x+y \leq z \\ v \leq w}} dF_{XYW}(x,y,v) && (3.5.6) \\ &= \int\limits_{v \leq w} \sigma_{\Box}(F_X, F_Y; v)(z)\, dv, && (3.5.7) \end{aligned}$$

where $\sigma_\square(F_X, F_Y; v)(z)$ is simply $\sigma_\square(F_X, F_Y)(z)$ for a given v (*i.e.* for $W = v$). Thus

$$C_{ZW}(t,u) = \int\limits_{v \le F_W^\wedge(u)} \sigma_\square(F_X, F_Y; v)(F_Z^\wedge(t))\, dv$$

and

$$\underline{C}_{ZW}(t,u) = \int\limits_{v \le F_W^\wedge(u)} \mathrm{ldb}_{\underline{C}_{XY}}(F_X, F_Y, \square; v)(F_Z^\wedge(t))\, dv \tag{3.5.8}$$

where $\mathrm{ldb}_{\underline{C}_{XY}}(F_X, F_Y, \square; v)$ is the lower bound on $\sigma_\square(F_X, F_Y; v)$ and is given by

$$\begin{aligned} \mathrm{ldb}_{\underline{C}_{XY}} \quad & (F_X, F_Y, \square; q)(x) = \tau_{\underline{C}_{XY},\square}\left[\underline{F}_X(.;q), \underline{F}_Y(.;q)\right](x) \\ & = \tau_{\underline{C}_{XY},\square}\left[\underline{C}_{XW}(\underline{F}_X(.), \underline{F}_W(q)), \underline{C}_{YW}(\underline{F}_Y(.), \underline{F}_W(q))\right](x) \\ & = \sup_{u+v=x} \underline{C}_{XY}\left[\underline{C}_{XW}(\underline{F}_X(u), \underline{F}_W(q)), \underline{C}_{YW}(\underline{F}_Y(v), \underline{F}_W(q))\right]. \end{aligned} \tag{3.5.9}$$

Upon substituting (with a few changes of variables) one obtains

$$\underline{C}_{ZW}(t,u) = \int\limits_{v \le F_W^\wedge(u)} \left(\sup_{a+b=F_Z^\wedge(t)} \underline{C}_{XY}\left[\underline{C}_{XW}(\underline{F}_X(a), \underline{F}_W(v)), \underline{C}_{YW}(\underline{F}_Y(b), \underline{F}_W(v))\right] \right) dv.$$

Again this is not as computationally complex as it looks because

$$\underline{C}_{ZW}(t, u+\delta) = \underline{C}_{ZW} + \int\limits_{F_W^\wedge(u) < v \le F_W^\wedge(u+\delta)} \left(\cdots \right) dv, \tag{3.5.10}$$

where the term in the parenthesis is the same as that in the previous equation.

The numerical implementation of the above formulae will introduce further complications. For instance, a numerical representation of C_{XY} that fits in neatly with the representation already adopted for distribution functions will have to be found. Further investigation is required to determine the feasability of the approach outlined here.

3.5.4 Measures of Dependence

We have seen that different lower bounds on C_{XY} induce (or imply) different values of indices of dependence (*e.g.* the correlation coefficient). The converse is not necessarily true: different values of r do not imply unique corresponding lower bounds on the copula. Nevertheless, because of the complexity of the approach outlined in the previous subsection it seems worthwhile to examine what effect knowledge of the values of different measures or indices of dependence can have on the result of some operation combining two or more random variables. We simply point to some of the literature here.

Measures of dependence based on copulas are discussed in [719,881,882]. Other types of dependence measures are numerous. See for example [313,459,708]. Explicit

consideration of the effects of convolution on various measures of dependence can be found in [87,532,741,821,863] (this is not an exhaustive list). So far we have been unable to develop an appropriate method of dealing with operations on random variables when some dependence information is available.

3.6 Use of Mixtures for Nonmonotonic Operations

3.6.1 Introduction and General Approach

If the random variables to be combined under a division or multiplication operation are not sign definite (*i.e.* the distributions are not such that $F(0) = 0$ or 1), then the required σ-convolution can not be calculated using the above techniques or simple modifications thereof. In such a case a new approach is called for, and it is this which is the subject of the present section. Because of the intricacy of the results so far obtained, and because of some remaining problems, the present section gives less detail and is more tentative than other sections in this chapter. The general idea of handling nonmonotonic operations is to split the operation up into monotonic segments, perform the operation, and then combine the results together again. In the case under consideration here, this entails splitting the random variables involved into positive and negative parts, combining the various parts under the split operation, and then recombining them. The basic concept we use for this is a mixture.

The process is easily explained if we restrict consideration to single valued (rather than lower and upper) probability distributions which are defined on $\Re$. (The numerical approximations are considered later.) Let F_X and F_Y be the probability distributions of two random variables X and Y, neither of which is sign definite. Do the following for both X and Y (expressions are only given for X). Split X into positive and negative parts. If p_X is the probability that X is positive, then

$$p_X = 1 - F_X(0+), \tag{3.6.1}$$

and the distributions of the parts are

$$\begin{aligned} F_X^+(x) &= \tfrac{1}{p_X}(F_X(x) - F_X(0)), \\ F_X^-(x) &= \tfrac{1}{1-p_X}(F_X(0) - F_X(x)) \end{aligned} \tag{3.6.2}$$

for $x > 0$. From now on, it will be assumed that F_X and F_Y are continuous at $x = 0$ and so $F_X(0+) = F_X(0)$. There is no loss of generality in doing this as it is always possible to assume that any jump at $x = 0$ is in fact due to the random variable X being a mixture of two random variables: one without the jump, and the other consisting solely of the jump (*cf.* the Lebesgue decomposition theorem [750, p.202]). The distribution F_X can easily be reconstructed from F_X^+ and F_X^- by

$$F_X(x) = \begin{cases} p_X F_X^+(x) + (1 - p_X) & x > 0, \\ (1 - p_X) - (1 - p_X)F_X^-(-x) & x < 0. \end{cases} \tag{3.6.3}$$

In order to calculate the result of a convolution or dependency bound in terms of the positive and negative parts, it is simply necessary to observe that the positive part F_Z^+ is, for $Z = X \times Y$ or $Z = X/Y$, solely determined by $\sigma_\Box(F_X^+, F_Y^+)$ and $\sigma_\Box(F_X^-, F_Y^-)$ and that the negative part F_Z^- is solely determined by $\sigma_\Box(F_X^+, F_Y^-)$ and $\sigma_\Box(F_X^-, F_Y^+)$. Let p_Z denote the probability that Z is positive. Then

$$p_Z = p_X p_Y + (1 - p_X)(1 - p_Y). \tag{3.6.4}$$

The parts F_Z^+ and F_Z^- are given by

$$F_Z^+(z) = \frac{1}{p_Z}\left[p_X p_Y \sigma_\Box(F_X^+, F_Y^+)(z) \ + \ (1 - p_X)(1 - p_Y)\sigma_\Box(F_X^-, F_Y^-)(z)\right], \tag{3.6.5}$$

$$F_Z^-(z) = \frac{1}{1 - p_Z}\left[p_X(1 - p_Y)\sigma_\Box(F_X^+, F_Y^-)(z) \ + \ (1 - p_X)p_Y\sigma_\Box(F_X^-, F_Y^+)(z)\right], \tag{3.6.6}$$

for $z > 0$ and $\Box \in \{\times, \div\}$. Substituting Z for X in (3.6.3) gives the formula for creating F_Z from F_Z^+ and F_Z^-.

3.6.2 Complications Arising from the use of Lower and Upper Probability Distributions

The above formulae are quite straightforward. However things become more complicated when the analogous formulae for lower and upper probability distributions are considered. The corresponding formulae for (3.6.4), (3.6.2) and (3.6.3) are, respectively

$$\begin{aligned} \underline{p}_X &= 1 - \overline{F}_X(0), \\ \overline{p}_X &= 1 - \underline{F}_X(0), \end{aligned} \tag{3.6.7}$$

$$\begin{aligned} \underline{F}_X^+(x) &= \tfrac{1}{\overline{p}_X}(\underline{F}_X(x) - \underline{F}_X(0)), \\ \overline{F}_X^+(x) &= \tfrac{1}{\overline{p}_X}(\overline{F}_X(x) - \underline{F}_X(0)), \\ \underline{F}_X^-(x) &= \tfrac{1}{1-\underline{p}_X}\overline{F}_X(0) - \overline{F}_X(-x)), \\ \overline{F}_X^-(x) &= \tfrac{1}{1-\underline{p}_X}(\overline{F}_X(0) - \underline{F}_X(-x)), \end{aligned} \tag{3.6.8}$$

for $x > 0$, and

$$\begin{aligned} \underline{F}_X(x) &= \begin{cases} \overline{p}_X \underline{F}_X^+(x) + (1 - \overline{p}_X) & x > 0, \\ (1 - \underline{p}_X) - (1 - \underline{p}_X)\overline{F}_X^-(-x) & x < 0, \end{cases} \\ \overline{F}_X(x) &= \begin{cases} \overline{p}_X \overline{F}_X^+(x) + (1 - \overline{p}_X) & x > 0, \\ (1 - \underline{p}_X) - (1 - \underline{p}_X)\underline{F}_X^-(-x) & x < 0. \end{cases} \end{aligned} \tag{3.6.9}$$

Substitution of (3.6.8) into (3.6.9) gives $\underline{F}_X(x) = \underline{F}_X(x)$ and $\overline{F}_X(x) = \overline{F}_X(x)$ as one would expect.

It is apparent that it is possible to calculate the dependency bounds in terms of the positive and negative parts in the usual manner. What is not apparent however,

Figure 3.21: Illustration of (i) quantities used in splitting a distribution into positive and negative parts; and (ii) the reason why $\underline{p}_X$ is not always necessarily associated with $\overline{F}_X$ nor $\overline{p}_X$ with $\underline{F}_X$.

is whether such results can be combined together using equation (3.6.9) to give a meaningful result. Let us first examine how to calculate $\underline{p}_Z$ and $\overline{p}_Z$. Remembering the viewpoint adopted regarding the lower and upper distributions $\underline{F}_Z$ and $\overline{F}_Z$ lower and upper approximations to a single fixed (but unknown) distribution F_Z, it seems reasonable to define $\underline{p}_Z$ and $\overline{p}_Z$ by

$$\begin{array}{rcl} \underline{p}_Z & = & \min(t_1, t_2, t_3, t_4), \\ \overline{p}_Z & = & \max(t_1, t_2, t_3, t_4), \end{array} \tag{3.6.10}$$

where

$$\begin{aligned} t_1 &= \underline{p}_X \underline{p}_Y + (1 - \underline{p}_X)(1 - \underline{p}_Y), \\ t_2 &= \overline{p}_X \overline{p}_Y + (1 - \overline{p}_X)(1 - \overline{p}_Y), \\ t_3 &= \underline{p}_X \overline{p}_Y + (1 - \underline{p}_X)(1 - \overline{p}_Y), \\ t_4 &= \overline{p}_X \underline{p}_Y + (1 - \overline{p}_X)(1 - \underline{p}_Y), \end{aligned}$$

That is, not *both* $\underline{p}_X$ and $\overline{p}_X$ are used in any one possible choice: p_X is uniquely defined — it is just that its value is not known.

In order to combine the various $\sigma_\square$'s together to give $\underline{F}_Z^+$, $\overline{F}_Z^+$, $\underline{F}_Z^-$, and $\overline{F}_Z^-$, it must be realised that $\underline{p}_X$ is not necessarily associated with $\overline{F}_X$. This is because the situations depicted in figure 3.21 can occur. Thus it is necessary to calculate $\underline{F}_Z^+$, $\overline{F}_Z^+$, $\underline{F}_Z^-$, and $\overline{F}_Z^-$, by

$$\underline{F}_Z^+(z) = \underline{c}^+ \min\left[\underline{\phi}^+(\underline{p}_X, \underline{p}_Y), \underline{\phi}^+(\underline{p}_X, \overline{p}_Y), \underline{\phi}^+(\overline{p}_X, \underline{p}_Y), \underline{\phi}^+(\overline{p}_X, \overline{p}_Y)\right] \tag{3.6.11}$$

where $\underline{\phi}^+(p_X, p_Y) = p_X p_Y \underline{\sigma}_\square(F_X^+, F_Y^+)(z) + (1 - p_X)(1 - p_Y)\underline{\sigma}_\square(F_X^-, F_Y^-)(z)$,

$$\overline{F}_Z^+(z) = \overline{c}^+ \max\left[\overline{\phi}^+(\underline{p}_X, \underline{p}_Y), \overline{\phi}^+(\underline{p}_X, \overline{p}_Y), \overline{\phi}^+(\overline{p}_X, \underline{p}_Y), \overline{\phi}^+(\overline{p}_X, \overline{p}_Y)\right] \tag{3.6.12}$$

where $\overline{\phi}^{+}(p_X, p_Y) = p_X p_Y \overline{\sigma}_{\square}(F_X^{+}, F_Y^{+})(z) + (1-p_X)(1-p_Y)\overline{\sigma}_{\square}(F_X^{-}, F_Y^{-})(z)$,

$$\underline{F}_Z^{-}(z) = \underline{c}^{-} \min\left[\underline{\phi}^{-}(\underline{p}_X, \underline{p}_Y), \underline{\phi}^{-}(\underline{p}_X, \overline{p}_Y), \underline{\phi}^{-}(\overline{p}_X, \underline{p}_Y), \underline{\phi}^{-}(\overline{p}_X, \overline{p}_Y)\right] \tag{3.6.13}$$

where $\underline{\phi}^{-}(p_X, p_Y) = p_X p_Y \underline{\sigma}_{\square}(F_X^{+}, F_Y^{-})(z) + (1-p_X)(1-p_Y)\underline{\sigma}_{\square}(F_X^{-}, F_Y^{+})(z)$,

$$\overline{F}_Z^{-}(z) = \overline{c}^{-} \max\left[\overline{\phi}^{-}(\underline{p}_X, \underline{p}_Y), \overline{\phi}^{-}(\underline{p}_X, \overline{p}_Y), \overline{\phi}^{-}(\overline{p}_X, \underline{p}_Y), \overline{\phi}^{-}(\overline{p}_X, \overline{p}_Y)\right] \tag{3.6.14}$$

where $\overline{\phi}^{-}(p_X, p_Y) = p_X p_Y \overline{\sigma}_{\square}(F_X^{+}, F_Y^{-})(z) + (1-p_X)(1-p_Y)\overline{\sigma}_{\square}(F_X^{-}, F_Y^{+})(z)$, In all of these formulae c is a normalising constant such that $F_Z(\infty) = 1$. The $\underline{\sigma}_{\square}$ and $\overline{\sigma}_{\square}$ functionals are such that the appropriate (lower or upper) component of its arguments is chosen (see sections 3.3.4 and 3.4.3). Some information is lost in this procedure (because it is not known that $\underline{p}_X$ is necessarily associated with $\overline{F}_X$ *etc.*), and thus the bounds so obtained would not be pointwise best possible (even ignoring the ordinary approximation error). Nevertheless, they would appear to be the best possible given the approach that has been adopted. The question whether (3.6.9) can be used to calculate $\underline{F}_Z$ and $\overline{F}_Z$ from $\underline{F}_Z^{+}$, $\overline{F}_Z^{+}$, $\underline{F}_Z^{-}$, and $\overline{F}_Z^{-}$, will be examined below.

3.6.3 Difficulties Introduced by Using the Numerical Approximations

Unfortunately further difficulties are introduced when the above formulae are implemented in terms of the numerical approximations of the distribution functions. For instance, just the splitting of the probability distribution into positive and negative parts causes problems because of the need to maintain both the equispaced quantisation and the representation error confinement property. One could either form positive and negative parts of N points each from an N point initial distribution, or one could keep the same points (values of $F_X^{[\wedge]}[i]$) resulting in negative and positive parts of N^{-} and N^{+} points respectively, with $N \leq N^{-} + N^{+} \leq 2N$. The first choice results in approximations being necessary in the form of further outwardly directed rounding. The second choice has been adopted here, although as will be seen below, it is not without problems either. Having done this, it is necessary to generalise our algorithms for calculating lower and upper dependency bounds and σ-convolutions to handle inputs of differing sizes (*i.e.* a different number of points used in the approximations). The appropriate equations and algorithms, which are presented below, were derived in a fairly straightforward but longwinded manner and so the derivations are omitted.

Let N be the number of points with which $F_Z^{[\wedge]}$ is to be represented. Let M be the number of points used for $F_X^{[\wedge]}$ and let P be the number of points used for $F_Y^{[\wedge]}$. This is explicitly indicated in the formulae below by writing $F_X^{[\wedge,M]}$, $F_Y^{[\wedge,P]}$ and $\mathrm{ldb}^{[\wedge,N]}$ or $F_Z^{[\wedge,N]}$. The dependency bounds are given by

$$\mathrm{ldb}^{[\wedge,N]}\left(\underset{\sim}{F}_X^{[\wedge,M]}, \underset{\sim}{F}_Y^{[\wedge,P]}, \binom{+}{\times}\right)[i] =$$
$$\inf_{j=\left\lfloor \frac{(i+1)M}{N}-1 \right\rfloor,\ldots,M-1} \left(\underset{\sim}{F}_X^{[\wedge,M]}[j] \binom{+}{\times} \underset{\sim}{F}_Y^{[\wedge,P]}\left[\min\left(P-1,\ \left\lfloor \frac{P[(i+1)M-(j+1)N+NM]}{NM} - 1 \right\rfloor\right)\right]\right) \tag{3.6.15}$$

$$
\begin{array}{l}
L := M * P/N; \\
\textbf{for}(i := 0; i < M; i{+}{+})\{ \\
\quad \textbf{for}(j := 0; j < P; j{+}{+})\{ \\
\quad\quad \underset{\sim}{F}_Z^{[\wedge, M\times P, U]}[i*P+j] := \underset{\sim}{F}_X^{[\wedge,M]}[i] \binom{+}{\times} \underset{\sim}{F}_Y^{[\wedge,P]}[j]; \\
\quad\quad \tilde{F}_Z^{[\wedge, M\times P, U]}[i*P+j] := \tilde{F}_X^{[\wedge,M]}[i] \binom{+}{\times} \tilde{F}_Y^{[\wedge,P]}[j]; \\
\quad \} \\
\} \\
\underset{\sim}{F}_Z^{[\wedge, M\times P, E]} := \text{sort}\left(\underset{\sim}{F}_Z^{[\wedge, M\times P, U]}\right); \\
\tilde{F}_Z^{[\wedge, M\times P, E]} := \text{sort}\left(\tilde{F}_Z^{[\wedge, M\times P, U]}\right); \\
\textbf{for}(i := 0; i < N; i{+}{+})\{ \\
\quad \underset{\sim}{F}_Z^{[\wedge,N]}[i] := \underset{\sim}{F}_Z^{[\wedge, M\times P, E]}[i*L+L-1]; \\
\quad \tilde{F}_Z^{[\wedge,N]}[i] := \tilde{F}_Z^{[\wedge, M\times P, E]}[i*L]; \\
\}
\end{array}
\tag{3.6.19}
$$

Algorithm 3.6.19

$$
\text{udb}^{[\wedge,N]}\left(\tilde{F}_X^{[\wedge,M]}, \tilde{F}_Y^{[\wedge,P]}, \binom{+}{\times}\right)[i] = \sup_{j=0,\ldots,\left\lceil \frac{iM}{N}\right\rceil} \left(\tilde{F}_X^{[\wedge,M]}[j] \binom{+}{\times} \tilde{F}_Y^{[\wedge,P]}\left[\max\left(0, \left\lceil \frac{P(iM-jN)}{NM}\right\rceil\right)\right]\right) \tag{3.6.16}
$$

$$
\text{ldb}^{[\wedge,N]}\left(\underset{\sim}{F}_X^{[\wedge,M]}, \tilde{F}_Y^{[\wedge,P]}, \binom{-}{\div}\right)[i] = \inf_{j=\left\lfloor \frac{(i+1)M-N}{N}\right\rfloor,\ldots,M-1} \left(\underset{\sim}{F}_X^{[\wedge,M]}[j] \binom{-}{\div} \tilde{F}_Y^{[\wedge,P]}\left[\max\left(0, \left\lceil \frac{P[N(j+1)-M(i+1)]}{NM}\right\rceil\right)\right]\right) \tag{3.6.17}
$$

$$
\text{udb}^{[\wedge,N]}\left(\tilde{F}_X^{[\wedge,M]}, \underset{\sim}{F}_Y^{[\wedge,P]}, \binom{-}{\div}\right)[i] = \sup_{j=0,\ldots,\left\lceil \frac{iM}{N}\right\rceil} \left(\underset{\sim}{F}_X^{[\wedge,M]}[j] \binom{-}{\div} \underset{\sim}{F}_Y^{[\wedge,P]}\left[\min\left(P-1, \left\lfloor \frac{P(jN-iM+NM)}{NM} - 1\right\rfloor\right)\right]\right) \tag{3.6.18}
$$

for $i = 0, ..., N-1$. The symbols $\binom{+}{\times}$ and $\binom{-}{\div}$ mean that the formulae hold for either $+$ or $\times$ (or $-$ or $\div$) respectively. Equations 3.6.15–3.6.18 reduce to equations (3.4.17–3.4.24) when $M = P = N$. It is important to calculate the complex index expressions in the way they are written here. Otherwise roundoff error can be a problem (because of the floor and ceiling operations). The only floating point operation necessary if the formulae are programmed as written here is a division. As long as floating point division of an integer by one of its integer divisors gives an exact integer as a result, there should be no problems.

The algorithm for calculating $\sigma_\square$-convolutions needs to be modified also. The resulting algorithm for $\square \in \{+, \times\}$ is presented as algorithm 3.6.19 This algorithm will only work if $N|MP$ (*i.e.* if L is an integer). Otherwise the condensation

Figure 3.22: Result of the numerical calculation of a $\sigma_\times$-convolution of X and Y. The variable X is uniformly distributed on $[2, 4]$ and the variable Y has a Gaussian distribution with $\mu = 4$ and $\sigma^2 = 2$ and is curtailed to $\mu \pm 2\sigma$. In both cases $N_Y = 40$. The thin line is for $N_X = 20$, and the thick line for $N_X = 50$.

procedure will have to be considerably more complex and, because of necessary approximations, it would be less accurate (more outwardly directed rounding would be required). The condition $N|MP$ does not matter however, because, as will be shown below, it is natural to adopt the convention that $N = \min(M, P)$, and in this case the divisibility condition always holds. The modifications to the $\sigma_\square$-convolution algorithm for $\square \in \{-, \div\}$ are exactly the same.

The above formulae and algorithms give results which are not at all surprising. The two examples presented in figures 3.22 and 3.23 show that the representation error containment property has been preserved.

3.6.4 Numerical Algorithms for Splitting Distributions into Positive and Negative Parts and Recombining Them

Having dispensed with the above preliminaries, our attention can now be focussed on the details of implementing (3.6.7–3.6.14). As was the case with the original $\sigma_\square$-convolution algorithm (section 3.4.3), it will be found simpler to work directly with the algorithms, rather than attempting to determine appropriate formulae first.

Figure 3.23: The lower and upper dependency bounds for subtraction for the same variables as for figure 3.22. Again $N_Y = 40$, and the two cases presented are for $N_X = 20$ and $N_X = 50$.

```
/* First the positive part is constructed. */
for(i := 0; F_X^[∧][i] ≤ 0; i++);
p̄_X := 1 − i/N;
for(k := 0; i < N; F_X^+[∧][k++] := F_X^[∧][i++]);
N^+ := k;
for(i := 0; F̃_X^[∧][i] < 0; i++);
p_X := 1 − i/N;
for(j := 0; j < N^+ − (N − i); F̃_X^+[∧][j++] := 0);
while(j < N^+) F̃_X^+[∧][j++] := F̃_X^[∧][i++];
/* Now the negative part is constructed. */
for(i := 0; F̃_X^[∧][i] ≤ 0; i++);
N^− := i;
for(k := 0; i ≥ 0; F_X^−[∧][k++] := −F̃_X^[∧][−−i]);
for(i := 0; F_X^[∧][i] ≤ 0; i++);
for(j := 0; j < N^− − i; F̃_X^−[∧][j++] := 0);
while(j < N^−)F̃_X^−[∧][j++] := −F_X^[∧][−−i];
```

(3.6.20)

Algorithm 3.6.20

Algorithm 3.6.20 implements (3.6.7) and (3.6.8). Let N be the number of points in F_X, and let N^+ and N^- be the resulting number of points in F_X^+ and F_X^-. Then algorithm 3.6.20 splits F_X into F_X^+ and F_X^-. An example calculated with this algorithm is shown in figures 3.24–3.26. The random variable with the distribution given in figure 3.24 was split into positive and negative parts. The two parts are shown in figures 3.25 and 3.26. This algorithm results in positive and negative parts that have a number of points equal to the number of points used to represent the positive and negative values of the original distribution. That is, if i_{min} is the minimum value of i such that $\underset{\sim}{F}_X^{[\wedge]}[i] \geq 0$, then $N^+ = N - i_{min}$; and if j_{max} is the maximum value of j such that $\tilde{F}_X^{[\wedge]}[j] < 0$, then $N^- = j_{max}$. It is often the case that $N^+ + N^- > N$ because of the difference between $\underline{p}_X$ and $\overline{p}_X$.

Implementing (3.6.5) and (3.6.6) using the numerical representation is rather more complicated. Let us first consider how many points should be used to represent the results of the various $\sigma_\square$-convolutions. Because a smaller number of points used to represent an input distribution results in a greater distance between the lower and upper output distributions (see the examples presented in figures 3.22 and 3.23), it makes sense to restrict the number of points used to represent the output to at most the minimum of the number used to represent the inputs. This is henceforth adopted as a convention. It means we do not use an unnecessarily large number of points to represent the output of a $\sigma_\square$-convolution calculation.

Having done this, equations (3.6.5) and (3.6.6) can be considered to be simply

Figure 3.24: An example distribution which is to be split into positive and negative parts.

Figure 3.25: The positive part of the distribution in figure 3.24.

Figure 3.26: The negative part of the distribution in figure 3.24.

```
f := a/N; g := b/N; t := 0; k := i := j := 0;
while(k < N){
    while(t < 1/N){
        if(F^[^][i] < G^[^][j]){
            x := F^[^][i++]; t+=f;
        }
        else{
            x := G^[^][j++]; t+ = g;
        }
    }
    H^[^][k++] := x; t−=1/N;
}
```

(3.6.22)

Algorithm 3.6.22

of the form

$$H(z) = aF(z) + bG(z). \tag{3.6.21}$$

Ignoring for the moment the complications introduced by having all the above quantities interval valued (see equations 3.6.11–3.6.14), the general approach used to calculate (3.6.21) is outlined in algorithm 3.6.22. Here it is assumed that H, F, and G are all represented by N points. The idea behind the algorithm can be pictured in terms of a graph showing both $F^{[\wedge]}$ and $G^{[\wedge]}$. Moving along the x-axis from the left to the right, at each step that is encountered (be it in F or G), an appropriate amount is added to H. The variable k keeps track of how high H is and the "**while**$(t < 1/N)$" loop accumulates sufficient steps in F and G to correspond to a step in H. This algorithm is only approximate because of the implicit directed rounding associated with this **while** loop. Algorithm 3.6.22 gives the correct directed rounding for the lower approximation to H.

When all the additional complications (of a and b being interval valued, and F and G being interval valued with a different number of points used to represent them) are incorporated, the algorithm becomes rather more complicated. The details are omitted here as too much space would be required. The main idea used is that in (3.6.21), in order to obtain the minimum or maximum at any given z, it is not necessary to consider the minimum or maximum obtained at some previous z. This allows the calculation to proceed point by point.

The difficulties mentioned in the above two paragraphs are relatively straightforward to solve when compared with those encountered in attempting to calculate (3.6.9) numerically. At first sight it is surprising that this is the case as (3.6.9) seems rather simple. The difficulties arise because of the need to represent $\underline{F_Z}$ and $\overline{F_Z}$ by the same number of points, and because of the possibility of a "mismatch" either side of the $x = 0$ line, while all the time attempting to preserve the representation error containment property. No final solutions to this are offered here. It should be remarked however, that this operation of combining the positive and negative parts of each of the lower and upper distributions need only be performed if either the combination of random variables being performed is the last one in an overall

probabilistic arithmetic calculation, or if the following operation requires the whole distribution. The latter possibility is the case when the following operation is either a subtraction or an addition. If a given operation is to be followed by a multiplication (say), then there is no need to combine the positive and negative parts as the overall distribution would only have to be split up again.

3.6.5 Conclusions on the use of Mixtures for Nonmonotonic Operations

The use of mixtures for splitting random variables into positive and negative parts for calculating dependency bounds for nonmonotonic operations has been considered and many of the details worked out. Because of our desire to maintain the representation error containment property, the algorithms have been more complex than they would have been otherwise. Further work is needed to see whether the use of mixtures can solve all the difficulties associated with nonmonotonic operations.

3.7 Conclusions

The idea of probabilistic arithmetic which would allow one to work with random variables in the same way that one works with ordinary numbers was introduced in section 3.1. It was seen that the first step in developing such an arithmetic was to examine ways of numerically calculating convolutions of distribution functions. The phenomenon of dependency error and a suggested manner of handling it (dependency bounds) led to a numerical representation of distribution functions. This representation was shown to be also suitable for numerically calculating ordinary convolutions. The main difference between the representation adopted in this chapter and other methods that have been suggested is that the present method allows the representation error always to be bounded. By the representation error containment property of the lower and upper approximations, one always knows that the true distribution is contained within the lower and upper bounds.

Other methods, in comparison, not only have unknown errors (although the order of magnitude may be known), but often the combination rules are rather more complex than those used in this chapter. An example of this is the H-function based method described in [164,406,447]. This method is based on the use of rules expressing the distribution of certain convolutions of random variables whose distribution functions are H-functions in terms of other H-functions [775]. Not only is this method restricted in the type of distributions it can handle (the standard parametric distributions), but its combination rules are very complex and even when an answer is calculated (in terms of H-functions), a computer program needs to be used to calculate the point values of the distribution because its expression is so complicated. The method proposed in the present chapter is better than the methods discussed in chapter 2.

Whilst some of the groundwork for a useful probabilistic arithmetic has been laid, a good deal of further work is required. As well as the specific points mentioned earlier (measures of dependence based on copulas, the use of mixtures for non-

monotonic operations, and the development of algorithms for implementing other operations such as log and exp), it will be necessary to examine methods of using the convolution and dependency bound algorithms to calculate the sort of results we are after (solutions of random equations). The appropriate point of departure for this is the consideration of algorithms which have been used successfully in interval arithmetic [17,584]. Since interval arithmetic can be considered as a very crude form of probabilistic arithmetic where the only information known about a distribution is the smallest closed interval containing its support, it is hoped that ideas that have been useful in interval arithmetic will also be of use for more general probabilistic arithmetic.

Another aspect which deserves consideration is the acquisition of lower and upper approximations to distribution functions from sample data. A natural idea is to consider the lower and upper bounds as forming a confidence band. However there are many problems which arise with this scheme, such as whether the confidence region should be considered pointwise or overall and how one should combine the confidence levels of two distributions which are combined in the course of a probabilistic arithmetic calculation. Further investigation is required here. Also worth pursuing is the relationship between the ideas examined in this chapter and recent ideas on uncertainty such as fuzzy set theory and generalisations of probability theory (*e.g.* Dempster-Shafer theory [732,868]). This is done in the following chapter.

It seems that the phenomenon of dependency error may turn out to be the biggest obstacle to a successful probabilistic arithmetic. It may be that probabilistic arithmetic will be no better in terms of accuracy and computational complexity than Monte Carlo simulations followed by statistical estimation of the resultant probability distributions. This possible intractability of all methods other than Monte Carlo simulation has been considered by several authors in the general context of probabilistic theories of physics, especially quantum mechanics [178,265,301,883]. In [265], Feynmann argues that the only way to "simulate" some probabilistic systems is to use a probabilistic computer. This is because determining all the distribution functions and then integrating over those we are not interested in is intractable in many cases:

> For example, suppose there are variables in the system that describe the whole world (x_A, x_B) — the variables x_A you're interested in, they're "around here"; x_B are the whole result of the world. If you want to know the probability that something around here is happening, you would have to get that by integrating the total probability of all kinds of possibilities over x_B. If we had *computed* this probability, we would still have to do the integration
>
> $$P_A(x_A) = \int P(x_A, x_B)\, dx_B$$
>
> which is a hard job! But if we have *imitated* the probability, it's very simple to do it: you don't have to do anything to do the integration, you simply disregard what the values of x_B are, you just look at the region x_A [265, p.473].

What Feynmann is saying here is that if one calculated some complex joint distribution analytically, then in order to determine (not necessarily one-dimensional) marginals of this distribution it is necessary to perform the above integration. If, instead, the probabilistic processes have been simulated (imitated), one needs only to look at the quantities of interest; it is not necessary to integrate out the other variables. This is true regardless of whether the variables are independent or not.

The non-locality of physical theories can be considered to be analogous to dependency error arising in the course of a probabilistic arithmetic calculation. In such a calculation one needs to keep track explicitly of all the dependencies that arise due to common subexpressions, whereas in a Monte Carlo simulation these dependencies take care of themselves. The course for an improved probabilistic arithmetic is thus clear: it will be necessary to develop further methods of handling dependency error, making controlled approximations in order to avoid intractability. Recalling Manes' result mentioned in section 3.1.3 [540], such a course may prove useful to other uncertainty calculi as well.

In summary then, we have developed new methods for calculating convolutions and dependency bounds for the distributions of functions of random variables which are better in several respects than previously available methods. We have presented examples which show that the methods are feasible and readily implemented. We have suggested that these methods might form the basis of a "probabilistic arithmetic" suitable for calculating the distribution of functions of random variables. We have seen that the biggest obstacle to such an application is the phenomenon of dependency error, but we have also shown that the concept of dependency bounds can be used to reduce the effect of this error. The properties of these dependency bounds are explored further in chapters 4 and 5.

Chapter 4

Relationships with Other Ideas

The only essential knowledge pertains to the inter-relatedness of things.
— Jorge Luis Borges

[M]y subject does not exist because subject matters in general do not exist. There are no subject matters; no branches of learning — or, rather, of inquiry: there are only problems, and the urge to solve them.
— Karl Popper

4.1 Introduction

The techniques developed in chapter 3 have a number of interesting connexions with other ideas, some new, and others surprisingly old. This chapter is devoted to the study of these connexions. In summary, we will discuss the following issues:

- We shall see that simple cases of the analogue of our dependency bounds for random events were considered by George Boole in his *Investigation of the Laws of Thought* published some 125 years ago and we shall see how these results have been recently extended (section 4.2).
- The prospects of using graph-theoretic techniques for probabilstic arithmetic calculations will be considered, and we will review a number of ideas related to this (section 4.3).
- Different theories of interval-valued probabilities will be examined and contrasted with the probability bounds arising in chapter 3. In particular we will discuss the Dempster-Shafer theory of evidence. It will be shown that none of the methods presented to date (apart from those discussed in section 4.2) derive from the same motivation as our methods (section 4.4).
- We shall show that the dependency bounds have an intimate connexion with the methods of calculating with fuzzy numbers, and thus we will provide a probabilistic interpretation of fuzzy set theoretic operations (section 4.5).
- Finally we will discuss two different approaches to probabilistic arithmetic: the approach we have taken, and that taken in the study of probabilistic metric

spaces. We will show that the latter is a more "positivist" notion and we will argue that our approach is more meaningful, but that it engenders more severe mathematical difficulties (section 4.6).

This chapter became rather larger than the author expected because of the surprising number of interconnexions that were found. However, for ease of reading and comprehension, each of the sections can be read separately, and there is only a small amount of cross-referencing between sections.

4.2 George Boole's Results on Fréchet Bounds and Some Recent Extensions and Applications

Boole means something that no one has understood yet; the world is not ready to understand him.
— Stanley Jevons

Boole's treatment of probability baffled his contemporaries, and I do not pretend to understand him fully.
— Glenn Shafer

Never clearly understood, and considered anyhow to be wrong, Boole's ideas on probability were simply bypassed by the history of the subject, which developed along other lines.
— Theodore Hailperin

The dependency bounds we examined in chapter 3 can be traced back to Fréchet [279,280], Salvemini [705], Gini [123] (see [183]) and eventually to George Boole. Boole considered problems where no dependency information is known in his book *"An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probabilities"* [97]. In this section we will examine Boole's results on these problems, and we will describe some recent work which generalises his results. Some of these new results suggest directions for further research on our "dependency bounds" for functions of random variables. All of these results are for probabilities of *events*, and not for random variables. We shall see that Boole's work on probability is not quite as impenetrable or useless as some authors have suggested. In fact we shall find that it contains the groundwork for a potentially powerful technique of use in Artificial Intelligence systems that have to manage uncertain knowledge and inferences.

We are considerably aided in our study of Boole's work by Hailperin's admirable book [355] *"Boole's Logic and Probability."* Hailperin has presented Boole's results in modern notation and has been foremost in extending his results using the modern techniques of linear programming. His work is discussed in detail below.

4.2.1 Boole's Major and Minor Limits of Probabilities

Boole was unhappy with the need to make many assumptions in the normal application of the calculus of probability. As he puts it in a letter to De Morgan dated 4th August 1851:

> The grand difficulty in the common theory is to know what hypotheses you may lawfully make and what you cannot [764, p.51].

He was particularly concerned with problems where it was necessary to assume independence of the constituent probabilities in order to arrive at a definite answer. Interestingly (and confusingly), Boole is not consistent in his approach to such difficulties. For example, on page 256 of [97] he says

> When the probabilities of events are given, but all information respecting their dependence withheld, the mind regards them as independent.

Hailperin [355, p.221] adds that "in the face of a specific problem with material content Boole appears to back down from his position." We shall see below that it is quite the opposite view which seems appropriate for interpreting Boole's minor and major limits.

The above quoted position of Boole came under criticism by Wilbraham [865] (subsequently replied to by Boole [99]), and more recently by Jaynes [409]. Wilbraham's critcism and Boole's reply are discussed in detail by Hailperin [355, pp.266–278]. Jaynes' criticism is by far the most powerful, and, as seems invariably the case in arguments about the foundations of statistics, not a little acrimonious. He is *quite* contemptuous of Boole who he takes to task for ill-founded criticism of Laplace's use of prior distributions (in Bayesian type inference). (Boole was actually influenced quite a lot by Laplace: MacHale [534, p.204] says Laplace "played a significant part in Boole's Development.") Jaynes notes, with irony, that

> Boole, after criticizing Laplace's prior distribution based on the principle of indifference, then invokes that principle to defend his own methods against the criticisms of Wilbraham.

He goes on to say that

> [Boole's] own work on probability theory contains ludicruous errors, far worse than any committed by Laplace. ... While Laplace considered real problems and got scientifically useful answers, Boole invented artificial school-room type problems, and often gave absurd answers [409, pp.241–242].

Boole made a number of mistakes confusing probabilities of conditionals with conditional probabilities. His work on minor and major limits is essentially sound in principle though (if not in all its details). Nevertheless it seems to have been largely ignored up until very recently:

> Surprisingly, however, Boole has been given very little credit for his contributions in this area [probability theory] by present-day probability theorists and historians of mathematics; many textbooks or even history books on the subject do not even mention his name [534, p.211].

Boole first examined the type of problems we are concerned with in a paper he wrote around 1850, but which was not published until after his death, when it was communicated to the *Transactions of the Cambridge Philosophical Society* by his friend Augustus De Morgan [98]. Boole was led to examine bounds on the probabilities of certain compound events for which his "general method" for solving probability questions gave *indeterminate* answers. Boole describes these cases by saying that the data given "are insufficient to render determinant the value sought" and thus "the final expression will contain terms with arbitrary constant coefficients" [97, p.17]. Under these circumstances, one can either obtain new data (if possible) to "render the numerical solution complete," or, "by giving to their constants their limiting values 0 and 1, *determine the limits within which the probability sought must lie independently of all experience*" [97, p.17] (emphasis added).

Boole obtained "major and minor limits" for the probabilities of certain combinations of the events in terms of the unconditional event probabilities. On page 299 of [97] he gives the following bounds for the probability of a conjunction:

$$\max(P(x)+P(y)-1,0) \;\leq\; P(x\wedge y) \;\leq\; \min(P(x),P(y)), \tag{4.2.1}$$

although he does not use this notation (see also [355, p.260]). Boole also obtained analogous but rather more complex results for more complicated logical expressions. The above bounds are a special case of the more general results considered by Fréchet [279] which are given by

Theorem 4.2.1 *Let $x_1,\ldots,x_n$ be events with absolute probabilities of occurence of $p_1,\ldots,p_n$ respectively. Then the bounds*

$$\max(p_1,p_2,\ldots,p_n) \;\leq\; P(x_1\vee x_2\vee\cdots\vee x_n) \;\leq\; \min(1,p_1+p_2+\cdots+p_n) \tag{4.2.2}$$

and

$$\max(0,p_1+p_2+\cdots+p_n-(n-1)) \;\leq\; P(x_1\wedge x_2\wedge\cdots\wedge x_n) \;\leq\; \min(p_1,p_2,\ldots,p_n) \tag{4.2.3}$$

are the best possible given no further information.

PROOF. We will simply prove that the bounds for $n=2$ hold. The case for general n follows by induction. The best possible nature of the bounds is proved by constructing a set of events (for any given set of $\{p_i\}$) such that the bounds are met [279]. The bounds (4.2.2) and (4.2.3) follow from the standard rules of probability theory: For events A and B, these state that

$$\begin{aligned}
P(A) &\geq 0,\\
P(A) &= 0 \quad \text{if } A=\emptyset,\\
P(\neg A) &= 1-P(A),\\
P(A\vee B) &= P(A)+P(B)-P(A\wedge B),\\
P(A\wedge B) &= P(A|B)\times P(B).
\end{aligned}$$

First we prove the bounds for conjunction. We have $P(A\vee B)=P(A)+P(B)-P(A\wedge B) \Rightarrow P(A\wedge B)=P(A)+P(B)-P(A\vee B)$. But $P(A\vee B)\leq 1$, and

so $P(A \wedge B) \geq P(A) + P(B) - 1$. We also have $P(A \wedge B) \geq 0$, and therefore $P(A \wedge B) \geq \max(0, P(A) + P(B) - 1)$. The upper bound for $P(A \wedge B)$ follows from the fact that $P(A \wedge B) = P(A|B)P(B) = P(B|A)P(A)$. Since $P(A|B) \leq 1$ and $P(B|A) \leq 1$, we have $P(A \wedge B) \leq P(B)$ and $P(A \wedge B) \leq P(A)$. Thus $P(A \wedge B) \leq \min(P(A), P(B))$.

The bounds for disjunction are derived similarly: $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$ and $P(A \wedge B) \geq 0$ imply $P(A \vee B) \leq P(A) + P(B)$. But $P(A \vee B) \leq 1$ and so $P(A \vee B) \leq \min(1, P(A) + P(B))$. We also have $P(A \vee B) = 1 - P(\neg(A \vee B)) = 1 - P(\overline{A} \wedge \overline{B}) \geq 1 - \min(P(\overline{A}), P(\overline{B})) = \max(P(A), P(B))$. ■

It is instructive to examine Boole's interpretation of his results on bounds for the probabilities of compound events. One might expect, given the somewhat strange motivations for his logic [497], that his interpretation of probability would be peculiar. Boole has often been accused of *psychologism*[1] in his interpretation of logic. Richards [683], following the reading of Boole by Van Evra [839], has argued that whilst *motivated* by psychological considerations originally, Boole "was most certainly not a defender of logical psychologism" [683, p.31]. As already noted, Boole was not consistent in his interpretations. Perhaps the best we can do is to quote the following passage from the the introductory chapter to *The Laws of Thought*

> It will be manifest that the ulterior value of the theory of Probabilities must depend very much upon the correct formation of such mediate hypothesis, where the purely experimental data are insufficient for *definite* solution, and where that further experience indicated by the interpretation of the final logical equation is unattainable. Upon the other hand, an undue readiness to form hypotheses in subjects which from their very nature are placed beyond human ken, must re-act upon the credit of the theory of Probabilities, and tend to throw doubt in the general mind over its most legitimate conclusions [97, p.20].

4.2.2 Hailperin's Reformulation and Extension of Boole's Bounds Using Linear Programming Methods

Hailperin [353–355] has considerably extended Boole's results using the modern theory of linear programming. His bounds are for general Boolean functions of a set of events $\{A_i\}_{i=1}^n$. His results do not seem very well known and so we state them in the following two theorems [354].

Theorem 4.2.2 *Given any Boolean expression $\phi(A_1, \ldots, A_n)$:*

1. *There are numerical-valued n-ary functions L_ϕ and U_ϕ depending only on the (Boolean) structure of ϕ, such that the inequalities*

$$L_\phi(a_1, \ldots, a_n) \leq P(\phi(A_1, \ldots, A_n)) \leq U_\phi(a_1, \ldots, a_n) \tag{4.2.4}$$

hold in any probability algebra for which $P(A_i) = a_i$ $(i = 1, \ldots, n)$;

[1] Psychologism is the position that "psychology is the most fundamental branch of science, and that all other disciplines are special branches of psychology" [683, pp.19–20]

2. *The bounds in (4.2.4) are the best possible; and*

3. *The functions L_ϕ and U_ϕ are effectively determinable (by solving a linear programming problem) from the structure of ϕ.*

Hailperin extends this result to the case where only bounds are known (in place of the $a_i = P(A_i)$). He gives

Theorem 4.2.3 *Given Boolean polynomial expressions $\psi, \phi_1, \ldots, \phi_m$ in the variables $A_1, \ldots, A_n$:*

1. *There are two $2m$-ary numerical functions $L_\psi^{(\phi)}$ and $U_\psi^{(\phi)}$ depending only on the structures of $\psi, \phi_1, \ldots, \phi_m$, such that the inequalities*

$$\begin{aligned} L_\psi^{(\phi)}(a_1, \ldots, a_m, b_1, \ldots, b_m) &\leq P(\psi(A_1, \ldots, A_n)) \\ &\leq U_\psi^{(\phi)}(a_1, \ldots, a_m, b_1, \ldots, b_m) \end{aligned} \tag{4.2.5}$$

   hold in any probability algebra for which

$$a_i \leq P(\phi_i(A_1, \ldots, A_n)) \leq b_i \qquad (i = 1, \ldots, m).$$

2. *The bounds in (4.2.5) are the best possible; and*

3. *The functions $L_\psi^{(\phi)}$ and $U_\psi^{(\phi)}$ are effectively determinable (by solving a linear programming problem) from the structures of $\psi, \phi_1, \ldots, \phi_m$.*

A particularly useful special case of this theorem occurs when $\phi_i(A_1, \ldots, A_n) = A_i$ for $i = 1, \ldots, n = m$. In such a case, theorem 4.2.3 says that given bounds of the form

$$a_i \leq P(A_i) \leq b_i \qquad (i = 1, \ldots, n), \tag{4.2.6}$$

it is possible to determine bounds on $P(\psi(A_1, \ldots, A_n))$ in terms of the intervals $[a_i, b_i] \ni P(A_i)$ $(i = 1, \ldots, n)$. Theorem 4.2.3 only has content when $\{\phi_i\}$ satisfies some consistency constraints. For any given sets $\{\phi_i\}$, $\{a_i\}$ and $\{b_i\}$, the consistency is effectively decidable.

The proofs of these two theorems are essentially contained in [353]. We omit all details of the actual determination of the bounds in terms of the formulation as a linear programming problem. However we do remark that it was only through clever use of duality theorems that Hailperin obtained his results. Ursic [836] has considered the same problems as Hailperin and used similar techniques (linear algebra and linear programming) in an attempt to solve them. He studies the problem of calculating bounds which, although not the best-possible, are easier to compute. He describes how one can trade off tightness of the bounds against computational tractability. Ursic's results are far too intricate to summarise briefly here. Further work is needed to determine the practical value of his results.

Adams and Levine [7] have also considered problems similar to those studied by Hailperin in his development of probability logic (see below). They too (with Ursic)

seem to be unaware of Hailperin's work in this area. Their results are essentially a subset of Hailperin's and are not discussed further. However their *interpretation* of the results is interesting. They use their results to determine the degree to which a given uncertain inference is deductive. This allows an objective statement of the validity of an inference in terms of uncertain propositions. As Genesereth and Nilsson [296] point out, this could be of considerable value in Artificial Intelligence systems dealing with uncertain inference. A number of other authors (such as Grosof [341], Wise and Henrion [373,875,876], Ruspini *et al.* [30,581,699,700,824] and Driankov [221]) have also considered the use of lower and upper bounds on probabilities along the lines of those discussed above. We will examine their work in the following subsection.

4.2.3 Application of the Fréchet Type Bounds to Probability Logic

Hailperin uses the results in theorem 4.2.3 to develop a *probability logic* which allows inferences in terms of the lower and upper bounds for propositions. For example, generalising material implication: If $P(A) = p$, $P(A \rightarrow B) = q$ then his results imply $p(B) \in [p + q - 1, q]$ under the consistency condition $p + q \geq 1$. Hailperin's results provide a completely general and rigorous framework for uncertain inference. Not only is no new measure of uncertainty required (such as fuzzy sets or belief functions), but no interpretation problems are caused by the use of these results:

> The notion of probability is presupposed as part of the semantics, and questions as to its nature will play no more role than does the nature of truth in usual logic [354, p.201].

Hailperin's notion of probability logic seems to be the thing Popper had in mind when he discussed "probability logic" [660, p.323]:

> [T]here exists a logical interpretation of the probability calculus which makes logical derivability a special case of a relation expressible in terms of the calculus of probability. Thus I assert the existence of a probability logic which is a genuine generalization of deductive logic.[2]

The idea of a probabilistic logic can be traced back to Leibniz (see [354, p.198]). Since then several authors have examined probabilistic generalisations of logic. Let us just mention Scott and Krauss [720], Suppes [801], Rényi [682], Lewis [518] and Calabrese [126].

Implications for Dependency Bounds for Random Variables

Hailperin presented a number of examples of lower and upper bounds on the probability of Boolean functions. In many cases these bounds were obtained using his

[2] Popper was discussing this in order to refute the idea of an *inductive* probability logic that allowed (probabilistic) inductive inferences.

"full-blown linear algebra technique" [354, p.209]. However the bounds can often be calculated in a simpler manner. For instance, the results of the example mentioned above (material implication) follow easily from application of the bounds for $\vee$ and the fact that $P(\neg A) = 1 - P(A)$, given that $A \rightarrow B \equiv B \vee \neg A$. This is not always the case. As an example, Hailperin considers the determination of an upper bound for

$$P(\neg A_1 \wedge A_2 \wedge A_3 \ \vee \ \neg A_2 \wedge A_1 \wedge A_3 \ \vee \ \neg A_3 \wedge A_1 \wedge A_2) \tag{4.2.7}$$

given $P(A_i) = a_i$ for $i = 1, 2, 3$. The best upper bound is obtained (using the general linear programming technique) as

$$\min\{(a_1 + a_2 + a_3)/2,\ a_1 + a_2,\ a_1 + a_3,\ a_2 + a_3,\ 1 - a_1 + 1 - a_2 + 1 - a_3\}. \tag{4.2.8}$$

Hailperin [354, p.211] observes that this result can *not* be obtained by composition of the simple results for $\vee$, $\wedge$ and $\neg$.

The reason why the composition techniques fail for (4.2.7) is quite simple but important. Equation (4.2.7) can not be rewritten in a form with no repeated variables. This is why the probability bounds can not be determined pairwise. Observe that this is very reminsiscent of Manes' result [540] mentioned in section 3.1.3. The repeated variables cause the pairwise bounds to be not necessarily the best possible because not all possible combinations of probabilities will be allowable — there will be certain constraints which will prevent this.[3] The fact that the pairwise bounds are quite loose when repeated variables occur was the reason for Cleary's dissatisfaction with the use of (2.2) and (2.3) [151, p.146]. Even when there are not repeated variables, the bounds can rapidly become quite loose (see the example in [47, p.134]). This should not be taken as an argument against the probability bounds technique though. What it is does show is the danger, even in simple problems, of assuming independence in order to obtain a unique value at the end.

This result has implications for dependency bounds for functions of random variables. Obviously the pairwise composition of bounds will give the best possible bounds when there are no repeated variables. This will not necessarily be the case when repeated variables occur. (Whether or not, in a particular instance, the pairwise bounds are best possible is an open problem.) In such cases the prospects for determining the best possible bounds are quite daunting. Whereas Hailperin's problem is one of finding functions over $\Re^n$, the determination of dependency bounds for random variables entails finding functions over Δ^n. This will be, in general, very difficult, and it would seem that the effort would rarely be justified.[4] (One may as well calculate the actual distribution of the function of random variables in question by integrating out the Jacobian of transformation. It was due to the complexity of this that we adopted the use of the (pairwise) dependency bounds in the first place.)

[3] Repeated variables do not *imply* that the pairwise bounds will not be the best possible: the occurence of repeated variables is a necessary but not sufficient condition for this to be the case.

[4] However see our discussion of Haneveld's results on stochastic programming problems below.

Boole-Fréchet Bounds in Uncertainty in Artificial Intelligence

The use of the Boole-Fréchet bounds for "uncertain deduction" has been proposed by a number of workers in a field which has come to be known as "Uncertainty in Artificial Intelligence." Two collections of papers have been recently published [420, 511] in the field.

Quinlan [667] was perhaps the first in the Artificial Intelligence community to discuss the use of (4.2.2–4.2.3) in his "Cautious Approach to Uncertain Inference." It has most recently been considered by Grosof [341] who has developed Nilsson's ideas further [296,620]. Nilsson [620, p.76] has rederived Hailperin's result on uncertain *modus ponens*. He was aware (like Hailperin) of the need for "consistency conditions." Nilsson says

> In principle, the probabilistic entailment problem can be solved by linear programming methods, but the size of problems encountered in probabilistic reasoning is usually much too large to permit a direct solution [620, p.79].

(Nilsson was unaware of Hailperin's use of duality results which reduce the complexity considerably.) Grosof [341], who only considers the calculation of probability bounds for disjunction and conjunction, makes the following points about the use of (4.2.2–4.2.3) in a probability bounds logic:

- The use of probability bounds gives a *closed* and *consistent* uncertain inference system.
- Underdetermined information can be represented.
- The direct use of lower and upper probabilities gives a more general system than the Dempster-Shafer theory (see section 4.4 below).
- *No* unjustifiable independence assumptions are required.
- Lower and upper probabilities can be empirically acquired via standard confidence interval procedures.
- The combination rules of fuzzy set theory are a special case of the Boole-Fréchet bounds (see section 4.5 below).

Appelbaum and Ruspini [30] and Montogomery [581,699] have also proposed the use of Boole-Fréchet bounds. They too [699, p.89] say that the use of linear programming in order to find the best possible bounds "is impractical." They argue that the probability bounds approach is preferable to the use of the "principle of insufficient reason" (assumption of independence given no information to the contrary) because the probability bounds approach is rigorous — the result will always be correct. Following Martin-Clouaire and Prade [545], Appelbaum *et al.* [30, p.751, 824, p.69] have considered more general lower and upper bounds on the probability

of conjunction and disjunction based on t-norms and t-conorms. Neither Martin-Clouaire and Prade nor Appelbaum *et al.* provide a good justification for these. Their discussion is in terms of *valuations* v and they set

$$\begin{aligned}
\underline{v}(\neg a) &= 1 - \overline{v}(a) \\
\overline{v}(\neg a) &= 1 - \underline{v}(a) \\
\underline{v}(a \wedge b) &= T(\underline{v}(a), \underline{v}(b)) \\
\overline{v}(a \wedge b) &= T(\overline{v}(a), \overline{v}(b)) \\
\underline{v}(a \vee b) &= T^*(\underline{v}(a), \underline{v}(b)) \\
\overline{v}(a \vee b) &= T^*(\overline{v}(a), \overline{v}(b)).
\end{aligned}$$

Appelbaum *et al.* used $T = \Pi$ and M in their experiments with these formulae and found that the *upper* bounds were of little practical value. A more realistic (probabilistic) approach would be (in analogy with the dependency bounds for $\underline{C}_{XY} \neq W$) along the lines

$$\underline{T}(\underline{v}(a), \underline{v}(b)) \leq \underline{v}(a \wedge b) \leq \overline{v}(a \wedge b) \leq \overline{T}(\overline{v}(a), \overline{v}(b)) \tag{4.2.9}$$

(and a similar relationship for $v(a \vee b)$ in terms of T^*). The lower and upper bounds on T, when set to W and M would give the standard Boole-Fréchet bounds. If $W \leq \underline{T}$ or $\overline{T} \leq M$ then tighter bounds would be obtained. Other authors (such as Bonissone [93–95] and Goodman [330]) have also considered the use of general t-norms for the combination of uncertain evidence. Their methods are more subjective and non-probabilistic and are of little value to us.

Wise and Henrion [373,875,876] have considered the use of the Boole-Fréchet inequalities in a manner similar to that described in the previous paragraph. In [876] they consider three separate rules for conjunction and disjunction:

1. "Maximum Correlation"

$$\begin{aligned}
p(A \wedge B) &= \min(p(A), p(B)) \\
p(A \vee B) &= \max(p(A), p(B))
\end{aligned}$$

2. "Independence"

$$\begin{aligned}
p(A \wedge B) &= p(A)p(B) \\
p(A \vee B) &= p(A) + p(B) - p(A)p(B)
\end{aligned}$$

3. "Minimum Correlation"

$$\begin{aligned}
p(A \wedge B) &= \max(0, p(A) + p(B) - 1) \\
p(A \vee B) &= \min(1, p(A) + p(B)).
\end{aligned}$$

We would suggest they be used as in (2.9) thus allowing one to write (if one knew events A and B were *never* "negatively dependent")

$$\begin{aligned}
p(A)p(B) \leq \;& p(A \wedge B) \;\leq \min(p(A), p(B)) \\
p(A) + p(B) - p(A)p(B) \leq \;& p(A \vee B) \;\leq \max(p(A), p(B)).
\end{aligned}$$

Wise and Henrion [876, p.74] also consider "probabilistic *modus ponens*" in the same manner as Hailperin. They also note that the "Maximum Correlation" rules are those of fuzzy set theory (see section 4.5 below). Henrion [373, p.114] make the point (with which we agree) that the representability of some notion of dependence is a more important feature of uncertainty calculi than the fine details of representation of single events or propositions. He also makes a number of other points which we will be discussing elsewhere (use of graph models and "local event groups" (section 4.3) and Monte-Carlo methods or "logic sampling"). Cooper [165] has also used the idea of probability bounds and has combined this with the use of a graph to represent conditional dependence information. He is aware of the linear programming formulation of the problem, but is unaware of Hailperin's work on using duality results to simplify this.

Grosof [341] and many other authors (*e.g.* Hailperin and Shafer) studying lower and upper probabilities have made use of the relationship

$$\overline{p}(\neg A) = 1 - \underline{p}(A) \tag{4.2.10}$$

which comes from the fact that $p(A) + p(\neg A) = 1$. Grosof suggests that (4.2.10) allows one to work with lower probabilities only (rather than lower and upper) — if one knows the lower probability of every event, then one automatically has an upper probability for every event. We mention this because it is in fact of no use when random *variables* are considered. To see this consider

$$F_X(x) = P\{X < x\} = P(A).$$

Thus "$\neg A$" = "$X \geq x$." Assume we know $\underline{F}_X(x)$ (the lower bound on the distribution function of X). Now $P(\neg A) = P(X \geq x) = 1 - F_X(x)$. However $1 - F_X(x) = S_X(x)$ (the *survival function* of X). Thus we *can* use lower probabilities only, *if* we carry around the lower distribution function *and* the lower survival function. This is equivalent to carrying both the lower and upper distribution function and so no advantage is gained.

4.2.4 Further Results on Distributions known only by their Marginals

We will now review some further recent work on problems where only marginal distributions are known. We will briefly look at the problem of compatibility of joint distributions and then consider some problems arising in project planning. Perhaps the most surprising feature of some of these problems is that the more general (and more useful) problem where only the marginal distributions are known is *easier* to solve than the problem where independence is assumed.

Compatibility of Marginal Distributions and the Existence of Joint Distributions

One problem which has attracted a lot of attention is that of the *compatibility* of marginal distribution functions, especially higher dimensional marginals [183]. This

is related to the general problem of the existence of multivariate distributions with given marginals. Let us just mention some of the literature on the topic: [157,288, 442–445,458,483,612,697,753,781,854,863]. Schweizer and Sklar discuss the compatibility problem in terms of copulas in [718]. The extension of bivariate results to higher dimensions has proved surprisingly difficult. For example the analogue of the 2-copula W is in fact the best lower bound on multidimensional distribution functions with uniform marginals, but it is not itself a copula. A good (but somewhat dated) review of the compatibility problem is given by Dall'Aglio [183]. This problem does not concern us directly and is not discussed any further. It would be of considerable importance however if tighter bounds on compound expressions involving more than two variables were to be studied further.

Project Planning and Other Network Problems

Problems formulated in terms of sets of random variables known only by their marginals arise naturally in network problems. Let $I = \{1, \ldots, n\}$ be a set of nodes and let $I_1, \ldots, I_k$ be subsets of I such that $\cup_{j=1}^{k} I_j = I$ and so no two I_j are ordered by inclusion ($\{I_j\}$ is known as a *clutter* over I). The *blocking clutter* to $\{I_j\}$ is a clutter $J_1, \ldots, J_l$ such that $I_r \cap J_s = \emptyset \ \forall r, s$ and J_j are minimal sets with this property. Given a DAG (Directed Acyclic Graph), I is known as a *system* and $\{I_j\}_{j=1}^{k}$ and $\{J_j\}_{j=1}^{l}$ are the *paths* and *cuts* of the system. Each node i has a *weight* X_i associated with it.

The problems we will discuss are various *optimal value functions* of $\{X_i\}$. These include

$$\textit{PERT Critical Path} \quad M = \max_{1 \le j \le k} \sum_{i \in I_j} X_i \quad \text{over the clutter of paths.} \tag{4.2.11}$$

$$\textit{Maximum Flow} \quad L = \min_{1 \le j \le l} \sum_{i \in J_j} X_i \quad \text{over the clutter of cuts.} \tag{4.2.12}$$

$$\textit{Shortest Route} \quad S = \min_{1 \le j \le k} \sum_{i \in I_j} X_i \quad \text{over the clutter of paths.} \tag{4.2.13}$$

$$\textit{Reliability System Lifetime} \quad T = \max_{1 \le j \le k} \min_{i \in I_j} X_j = \min_{i \le j \le l} \max_{i \in J_j} X_i. \tag{4.2.14}$$

Two special cases are the *pure parallel* ($k = n$, $I_j = \{j\}$ $1 \le j \le n$) in which case $M = S = T = \max_{i \in I} X_i$ and $L = \sum_{i \in I} X_i$; and the *pure series* ($k = 1$) in which case $M = S = \sum_{i \in I} X_i$ and $L = T = \min_{i \in I} X_i$.

Lai and Robbins [496] have considered the determination of $E\, M_n$ where $M_n = \max(X_1, \ldots, X_n)$. They were actually interested in determining EM_n when all the X_i are *independent*. Whilst it is known that in this case

$$EM_n = n \int_{-\infty}^{\infty} x F^{n-1}(x) dF(X) = m_n^*, \tag{4.2.15}$$

where F is the common marginal distribution of X_i, (4.2.15) is in general very difficult to evaluate. In the *maximally dependent* case it can be shown that

$$EM_n = a_n + n \int_{a_n}^{\infty} (1 - F(x)) dx = m_n$$

where $a_n = \inf\{x | F(x) \geq 1 - \frac{1}{n}\}$. Lai and Robbins study how close m_n and m_n^* are. (They turn out to be surprisingly close.) We note that their constructions would be a lot simpler if the concept of a copula was used.

More interesting to us (with regard to probabilistic arithmetic) is the evaluation of the distribution of M in the pure series case. Rüschendorf [695] has used a general duality result obtained in [287] and discussed in [693] in order to obtain lower and upper bounds on $P\{M \leq x\}$. Rüschendorf's results in [695] are similar to our dependency bounds for sums. In describing these bounds he simply writes $\sup_x(F(x-) + G(t-x))$ and $\inf_x(F(x-) + G(t-x))$, obviously *implicitly* including the max and min operations required to keep the result within $[0,1]$. His results were published contemporaneously with Makarov's [536] (the precursor to Frank, Nelsen and Schweizer's paper [277]). Rüschendorf obtains one explicit result for bounds on $df(\sum_{i=1}^n X_i)$ when $df(X_i) = U_{0,1}$ (the uniform distribution on $[0,1]$). These are a special case of Alsina's results [23]:

$$\rho_{W,+}(U_{a,b}, U_{c,d}) = U_{a+c,\max(a+d,b+c)}$$

and

$$\tau_{W,+}(U_{a,b}, U_{c,d}) = U_{\min(a+d,b+c),b+d}.$$

Rüschendorf also gives some results on

$$\phi_n(t) = \sup_{F \in \mathcal{H}(F_1,\ldots,F_n)} P\{\sum_{i=1}^n X_i \leq t\}$$

where $F_i = df(X_i)$ for $i = 1,\ldots,n$ and F is the joint distribution function. If G is a subadditive, strictly isotone function with $G \circ F_i^{-1}(x) \leq x$ for all $x \in [0,1]$, then $\phi_n(t) \leq \frac{2}{n}G(t)$. If G is a superadditive isotone function with $G \circ F_i^{-1}(x) \geq x$ for all $x \in [0,1]$, then $\phi_n(t) \leq \frac{2}{n}G(t)$. Rüschendorf's results should be compared to those in chapter 5 of this thesis where we examine the dependency bounds for $df(\frac{1}{N}\sum_{i=1}^N X_i)$ as $N \to \infty$. Rüschendorf did not use copulas and seems unaware of the work of Schweizer *et al.* in this area. Nevertheless he has found the viewpoint of copulas or uniform representations useful recently [698].

Klein Haneveld's use of Duality Results

Klein Haneveld [464] made use of several duality results and rearrangement techniques [287,696] in order to determine distributions of project completion times in PERT networks. He assumes the marginal distributions of the subproject completion times are known, but not the joint distribution. He calculates the worst case results over all possible joint distributions. His paper contains a large number of detailed results and we can not do it justice in the space available here.

We note that Klein Haneveld's "inner problem" was of the following form. Determine $h(t)$, where

$$h(t) = \sup_{H \in \mathcal{H}(F_1,\ldots,F_n)} E_H f(\xi_1,\ldots,\xi_n) \qquad t \in \Re, \tag{4.2.16}$$

$f(x) = f_t(x) = [M(x) - t]^+$, $x = (\xi_1, \ldots, \xi_n) \in \Re^n$,

$$M(x) = \max_{j=1,\ldots,k} \sum_{i \in I_j} x_i,$$

$x \in \Re^n$ (the project completion time) and E_H denotes the expectation over H. This calculates the cost of a "promised completion time" of t. If independence of the random variables is assumed it is possible to obtain bounds for the expected value of $M(x)$ [284,294] or its distribution function [685,751]. The problems defined by (4.2.16) have also been studied by Cambanis *et al.* [127,128,694]. They have determined bounds for $E_H k(X,Y)$ where $H \in \mathcal{H}(F_X, F_Y)$ and k is "quasi-monotone" (*i.e.* 2-increasing like a copula). Whilst these problems are somewhat different to the type of problems we have been concerned with, they are of interest because the same basic approach has been taken (determining lower and upper bounds assuming no dependency information).

One of Klein Haneveld's most surprising results is his analysis of "worst-case marginals." Assume we only know very limited information about the marginal distributions. Then what are the worst possible marginals (in terms of project completion time) satisfying the known information? Klein Haneveld solves this problem for the situation where the support, or the support and mean or mode of the marginals are known. He notes that the case of knowing the support and mode "is the most interesting, since it contains precisely the information which is usually supposed to be known in PERT networks" [464, p.178]. Using results of Žačková [902] and Dupačová [245], Klein Haneveld manages to solve this problem in an elegant manner. We will briefly return to a discussion of this in section 4.5 below where we consider fuzzy PERT networks. Duality in marginal problems was also the subject of Kellerer's difficult paper [446].

4.3 A Forest of Graph-Theoretical Ideas

Structure *between variables has received too little attention in statistics.*
— S.L. Lauritzen and D.J. Spiegelhalter

Probability is not really about numbers, it is about the structure of reasoning.
— Glenn Shafer

The use of graphs [359] for representing *structural* relationships (conditional independence) between random quantities is a powerful technique with apparently considerable potential for probabilistic reasoning and probabilistic arithmetic. In this present section we will review the applications of graph theory to stochastic problems and show how it can be used to understand problems arising in probabilistic arithmetic. Our aim is to explore the possibility of using graph-theoretical techniques to aid probabilistic arithmetic. We have already seen (section 4.2.4) the study of probabilistic PERT, which is a problem involving random variables on a series-parallel graph.

In subsection 4.3.1 we will see how graphs have been used in studying networks of *events*. Most of this work has appeared in the artificial intelligence literature.

Subsection 4.3.2 examines analogous ideas for networks of random *variables.* These methods have appeared mainly in the statistical, biological and sociological literature. They are used in studying *causal models.* These methods can in fact be considered as a rather weak and restricted form of probabilistic arithmetic. A graph representation aids the understanding of the control layer for probabilistic arithmetic calculations. In subsection 4.3.3 we will consider how techniques used by computer scientists in the construction of optimizing compilers might be profitably employed in probabilistic arithmetic. This is a quite different use of graph-theoretical techniques. We shall see that whilst certainly providing a clear understanding of the problems, unfortunately these methods do not manage to produce solutions for the situations we are concerned with. Finally in subsection 4.3.4 we will draw some conclusions on the use of graph theoretical ideas for probabilistic arithmetic.

The one theme that is repeated through this section is the desirability of having a graphical structure which is a tree. We have already mentioned (section 3.1) Manes' general result [540] which shows this effect for a variety of uncertainty calculi.

4.3.1 Networks of Probabilities of Events: Bayesian Networks, Belief Networks and Influence Diagrams

Bayesian networks or causal networks or belief networks are a way of representing a set of interdependent probabilities by use of directed graphs. Pearl[5] [193,194,637–643] has been foremost in developing these methods (see also the recent review by Cooper [166]). The basic idea is to use a directed graph to represent the conditional independence [188] properties of the probabilities of a set of events. This is best explained by an example.

Given a joint probability $P(x_1, \ldots, x_n)$ of n events we can always write

$$P(x_1, \ldots, x_n) = P(x_n|x_{n-1}, \ldots, x_1) \cdots P(x_3|x_2, x_1)P(x_2|x_1)P(x_1). \tag{4.3.1}$$

This expansion can sometimes be simplified. For example we may be able to write

$$P(x_1, \ldots, x_n) = P(x_6|x_5)P(x_5|x_2, x_3)P(x_4|x_1, x_2)P(x_3|x_1)P(x_2|x_1)P(x_1). \tag{4.3.2}$$

Equation (4.3.2) can be expressed graphically as in figure (4.1) which describes some *conditional independence* relations. For example x_3 is *independent* of x_2 *given* x_1. But x_3 may not be independent of x_2 given x_1 and x_6 because x_6 depends on x_5 which in turn depends on x_2 and x_3. It is possible [637, p.248] to determine conditional independence relations solely in terms of the graph model.

The main advantage of this graphical representation is computational: it allows advantage to be taken of conditional independence in order to reduce the computational complexity. Whilst theoretically this is of course possible without the graphical representation, it is much harder to do so. The graphical representation is easily understood in an intuitive manner. Pearl has shown [637,643] how, given a set

[5]The review of Pearl's work below was written before the author could read [642]. This book provides a detailed examination of the material summarised below plus a lot of other ideas. It is now the preferred reference for this material.

Figure 4.1: Directed graph representation of the expansion (4.3.18).

of probabilities with a tree structure (which is unknown), it is possible to determine the actual tree structure solely from the pairwise dependencies of the leaves. This can be done in $O(n \log n)$ time, where n is the number of leaves. Whilst of interest, this is not of enormous practical use because the method is extremely sensitive to errors in the estimates of the correlation coefficients. Furthermore it does not solve the (still open) problem of determining good tree structured *approximations* to sets of probabilities that are not in fact tree structured.

Given a network which is not a tree, there are several approaches which can be taken. Brute force methods can always in principle be used to propagate probabilities, but these are intractable for realistic problems. Instead, one can attempt to *transform* the graph into a tree. Pearl [640] discusses two ways of doing this: *clustering* and *conditioning.*

Conditioning (reasoning by assumptions) entails instantiating a node (setting it to a fixed value) and then computing the probabilities. One then sets the node to a different value and recalculates. The largest probability is then chosen. The effectiveness of this method depends to a large extent on the topological properties of the network. (It is necessary to instantiate an entire cycle cutset to make the network singly connected. This results in a complexity exponential in the size of the cyclic cutsets [640, p.199].) More promising is the clustering approach. This entails forming compound events (local event groups) such that the resulting network (of clustered events) is a tree. As noticed by Barlow [501, p.202], this is similar to Chatterjee's [140] modularisation of fault trees.[6] It is also reminiscent of Downs, Cook

[6] In the same volume as Chatterjee's paper there are several other papers, the results of which may be profitably applied to belief networks. These include Mazumdar on importance sampling [555] (this would increase the efficiency of stochastic simulations of belief networks); Rosenthal [690] on the NP-completeness of finding the reliability of the top event when the fault "tree" is not in fact a tree (common cause events) — this is an analogue of the NP-completeness results known for general belief networks [166]; and Lambert [499] on measures of importance in fault trees and cutsets (this might be useful in deciding which nodes in a belief network could be discarded, in order to reduce computational complexity, without affecting the

and Rogers' *partitioning approach* for the statistical analysis of interconnected systems [220]. One method of clustering Pearl has studied in some detail [640], is based on the formation of clique-trees using an algorithm of Tarjan and Yannakakis [819]. These transformation methods are in fact relevant to a wide variety of problems on graphs (we return to this shortly below).

A method of handling graphs which are not trees which does not involve transforming the graph is to use stochastic simulation [639]. This uses random samples (Monte Carlo methods) in order to *estimate* propagated probabilities. The method has been suggested by several authors [374]. (Pearl [639] gives a brief review of early attempts.) Pearl improves the efficiency of the simple application of the method by taking some account of the structure of the graph involved. Recall that we have already discussed (section 3.7) the use of Monte Carlo simulations to perform probabilistic arithmetic type calculations.

Loops in Networks in Constraint Satisfaction Problems

As well as occurring in probabilistic problems, the presence of loops in a network (non-tree structure) causes difficulties in "any problem where globally defined solutions are produced by local computations, be it probabilistic, deterministic, logical, numerical or hybrids thereof" [640, p.199]. Examples of such problems include the class of *Constraint Satisfaction Problems* (CSP) [193,194,641]. A number of problems arising in artificial intelligence can be formulated in this manner. Constraint-Satisfaction problems are easy if they are "backtrack-free" [193, p.8]. This is equivalent to their graph being a tree. In [641] Pearl shows how the propagation of probabilities in a belief network can be considered as a CSP. In [194] he describes an algorithm for rearranging a graph in order to make it easier to handle. The algorithm uses the ideas of triangulating a graph, identification of maximal cliques and the notion of a dual graph. Different CSPs soluble by these methods are described in [193].

Constraint propagation formulations[7] of problems similar to those we have been studying in this thesis were examined by Davis [187]. The type of problems he considers are explained by the following example. Consider three nodes with the *labels* $X \in [1, 10]$, $Y \in [3, 8]$ and $Z \in [2, 7]$ combined with the constraints $X + Y = Z$ and $Y \leq X$. Davis shows how application of the Waltz algorithm [852] tightens the labels to become $X \in [3, 4]$, $Y \in [3, 4]$ and $Z \in [6, 7]$. This idea has been applied to the analysis of electrical circuits [463] and geometrical reasoning [560]. Davis encountered exactly the same sort of problems as we have been discussing in this section: When the graph representing the structure of the problem is not a tree, then the standard methods give poor results [187, p.306 and 311]. Nonlinear constraints also cause severe difficulties. In fact Davis concludes rather negatively on the whole enterprise: "We have not found any arguments that the partial results computed by

final results greatly).

[7] Constraint propagation can be considered to be simply a programming *style* (see [508]). Constraint programming languages are declarative and non-procedural and seem to be a natural and simple way of formulating many problems arising in artificial intelligence.

label inference should be adequate for the purposes of AI" [187, p.316].

Influence Diagrams

Influence diagrams are similar to Pearl's belief networks with the addition of decision information. Influence diagrams were developed some time ago as an aid to automated decision analysis [576] and have been promoted recently by Shachter [729–731] and others. The value of influence diagrams is that they allow a graphical means for manipulating and exploiting conditional independence structure. For example whilst the fact that

$$P(x,y,z) = P(x)P(y|x)P(z|y) \tag{4.3.3}$$

is equivalent to

$$P(x,y,z) = P(y)P(x|y)P(z|y) \tag{4.3.4}$$

can be readily determined using the rules of the probability calculus, similar transformations on larger sets of variables are rather more difficult. The rules for influence diagram transformations are simply a way of performing such transformations in a manner that is easier to use.

Apart from being a suitable basis for the construction of expert systems which deal with uncertain information [9,10,679], and representing "semantic modularity" for plausible reasoning [370], influence diagrams have been used for analysis purposes [53,54]. Barlow [52] has suggested them as an alternative to the decision tree for Bayesian decision analysis. He describes an application of this idea (calibration of a measuring instrument) in [53].

The mathematical basis of influence diagrams or belief networks is presented most clearly by Smith in [767] where all the relevant theorems are rigorously proved. A good recent and general review which does not go into too many details is [766]. Another recent (and very detailed) paper with considerable discussion at the end is Lauritzen and Spiegelhalter's [501]. At the end of their paper they suggest the following extensions as suitable goals for further research:

1. The incorporation of imprecision in probabilites (lower and upper bounds). These bounds could then be propagated through the system in order to provide lower and upper bounds for the final result.

2. Consideration of the "meta-level" of control. The point is that global propagation may be unnecessary because of high level restructuring.

3. Extension to nodes representing continuous measurements.

Item 1 above is obviously reminiscent of our probability bounds. Item 2 corresponds to Bonissone's [93,94] "control layer." Item 3 was solved to an extent by the recent paper [502] which integrates belief networks with covariance structure models. These are graphical models which describe interrelationships between continuous random variables. They are the topic of the next subsection. Another possible extension

(suggested by Smith [768, pp.25–26]) is the depiction of "weak" relationships — *i.e.* relationships which are "almost" conditionally independent. This is related to the use of correlation coefficients in path models (see below) and our preliminary attempts at just this problem in section 3.5.

4.3.2 Path Analysis and Covariance Structure Modelling

Path analysis is a graphical technique for statistical modelling which Moran [586, p.93] has suggested should be regarded "solely as a shorthand aid to the correct use of the regression equations such as (4.3.5) and (4.3.6)":

$$E(X_1|W_i, V_i, U_i, \ldots) = \sum_j x_{1j} W_j \ldots, \tag{4.3.5}$$

where the x_{1j} are constants and

$$\begin{aligned} E(X_1 X_2|W_i) &= E(X_1|W_i)E(X_2|W_i) \\ &= \left(\sum_j x_{1j} W_j\right)\left(\sum_k x_{2k} W_k\right) \\ &= \sum_{jk} x_{1j} x_{2k} W_j W_k. \end{aligned} \tag{4.3.6}$$

Nevertheless it is a very useful aid and one that is continuing to find wide application in the biological and social sciences.

The technique involves making the following assumptions:

1. All the variables in the system under consideration have a joint probability distribution with finite first and second moments. (Normality is not assumed though.)
2. The conditional expectations are linear functions of the other variables (4.3.5).
3. The variates are related by a conditional independence structure. Thus when conditioned on the appropriate variates, the conditional distributions are independent of all the other variables (which we have not conditioned on).

Path analysis can be explained simply in terms of a "covariance algebra" [452, chapter 2]. Writing $C(X, Y)$ for the covariance between X and Y, the rules of this are that

1. $C(X, k) = 0$,
2. $C(kX, Y) = kC(X, Y)$,
3. $C(X, X) = V(X)$ (the variance),
4. $C(X, Y + Z) = C(X, Y) + C(X, Z)$.

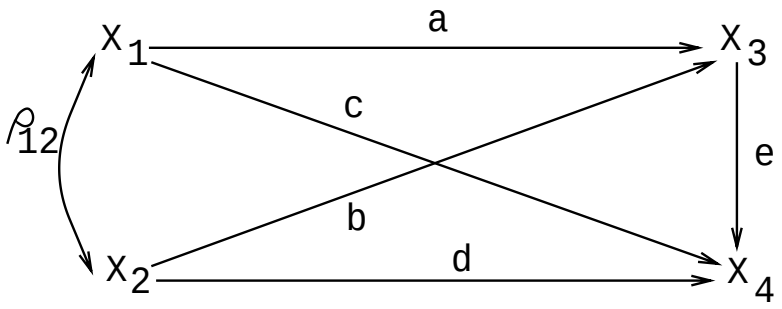

Figure 4.2: Path diagram of the linear model described by (4.3.7) and (4.3.8).

If all the variates are standardised[8] (zero mean, unit variance) by change of units, then we can represent linear models as in the following example:

$$X_3 = aX_1 + bX_2 \tag{4.3.7}$$

and

$$X_4 = cX_1 + dX_2 + eX_3. \tag{4.3.8}$$

This can be represented as in figure (4.2). The quantities a, b, c, d, e are known as *path coefficients* and ρ_{12} is the correlation between X_1 and X_2 (which equals the covariance because of the standardisation). The two main tools of path analysis are the first law and the tracing rule.

The first law says

$$\rho_{YZ} = \sum_i p_{YX_i}\rho_{X_iZ} \tag{4.3.9}$$

where p_{YX_i} is the path coefficient (or "causal parameter") from X_i to Y, ρ_{X_iZ} is the correlation between X_i and Z, and $\{X_i\}$ is the set of all variables that are "causes" of Y. This law follows from rules 2 and 4 of the covariance algebra.

The tracing rule (which only applies to hierarchical models without loops) is as follows: The correlation between X_i and X_j equals the sum of the product of all the paths obtained from each of the possible tracings beween i and j. The set of tracings includes all the possible routes from X_i to X_j given that a) the same variable is not entered twice, and b) a variable is not both entered through an arrowhead and exited through an arrowhead. For the above example this rule says that $\rho_{31} = a + b\rho_{12}$. The whole point of path analysis in practice is to help determine the causal parameters (the p's), although this is sometimes impossible [452, pp.34ff].

Path analysis is *not* intended to be a black box procedure for accomplishing the "impossible task of deducing causal relations from the values of the correlation coefficients" [887, p.193] (but see our mention of the recently developed TETRAD program below). The theory was originally developed by the geneticist Sewall Wright [885–889] and has since found wide application in the social [27,241,242,371,500] and biological [34,70,678] sciences.

[8]The advantages of using standardised variables and coefficients are explained in detail by Wright in [889].

There are, of course, many aspects we have not mentioned in this very concise overview of path analysis and covariance structure modelling. These include the different types of models (recursive *versus* nonrecursive [782]), difficulties encountered when the number of equations is far greater than the number of coefficients to be determined [887, p.193], the problem of estimating the parameters of the models [371,417], and the problem of *interpreting* the models [243,828,888]. Further details can be found in the books on the subject which include [84,244,452]. A good concise review is given in [832]. A more recent review is [321].

From our point of view (investigating probabilistic arithmetic), the most serious problem is the difficulty in handling product terms (XY). Whilst the relationships in standard path analysis do not have to be entirely linear, they have to be linear over the range of values considered [888, p.20], which of course amounts to the same thing. The topic of interaction (as it is called) has "received scant attention" [452, p.254]. The term "interaction" is used because the situation considered is $Y = (a + bZ)X$ instead of $Y = aX$. Here Z is another random variable. In his original presentation of path diagrams [886, p.563] Wright recognized "multiplying factors [as amongst] the most important." He suggests that if the coefficients of variation (σ_X/μ_X) are small, then the approximate formula

$$\sigma^2_{XZ} = \mu_X\sigma^2_Z + \mu_Z\sigma^2_X + \mu_X\mu_Z$$

could be used (see section 2.2.2).

Product terms have been considered in more detail in [88,156,312] and ratio terms are examined in [714]. Glass [312] has used Pearson's formula [645] for spurious correlation (see section 3.1) and [757]) in order to study effects of product terms in path models. Cohen [156] has generalised Glass's results and provides a discussion of the interpretation of product terms as interactions. He also considers powers of variables and compound product and polynomial terms. Nevertheless, all these efforts will only work well for small coefficients of variation because the formulae used are all based on the linearisation of a nonlinear function by a Taylor series expansion. In general, nonlinear functions are handled quite poorly. Whether the more sophisticated approach outlined in section 3.5 of will prove better is a matter for further research.

Finally let us just mention that graph-theoretical models have received increased attention in recent years with the development of sophisticated computer programs such as LISREL and TETRAD which allow more complex models to be examined. LISREL is described in [521]. It is a method of estimating the parameters of covariance structure models. The more complex task of estimating the structure has also received attention. TETRAD [315,316,712] is a program designed to do just this. Such a task is obviously at the core of scientific research and thus not surprisingly there is considerable debate about the philosophy behind this approach. There is a wide literature on the philosophy of probabilistic causality. Some of the more recent works include [118,134,217,248–250,253,357,365,394,704,802].

There is obviously room for a lot more work to be done in extending the methods of path analysis to more general operations and relationships between random variables. Perhaps eventually it will be possible to integrate the techniques discussed

here with the copula based methods discussed in section 3.5. Use of different measures of association (instead of the correlation coefficient) would be the first step in this direction.

4.3.3 Common Subexpression Elimination in Probabilistic Arithmetic by DAG Manipulations

A rather different use of graph-theoretical ideas for probabilistic arithmetic arises in the elimination of common subexpressions from arithmetic expressions. An example of this is can be found in chapter 7 where we evaluate (by hand) the distribution of the elements of the inverse of a random 2×2 matrix. The entries of the inverse matrix in terms of the entries of the original matrix can be written in the form

$$X = \frac{A}{AD - BC}. \tag{4.3.10}$$

It is not possible to evaluate the distribution of X in terms of pairwise applications of the σ_L-convolution formulae because the A term appears twice. However we can see that dividing through by A gives

$$X = \frac{1}{D - \frac{BC}{A}} \tag{4.3.11}$$

which *can* be analysed in terms of pairwise σ_L-convolutions. It is the purpose of the present subsection to study when and how such transformations can be performed, and to consider the prospects for an *algorithmic* application of these transformations.

This goal has arisen before in the field of code generation for compilers. The problem there is, given an expression such as (4.3.10), determine a sequence of machine instructions which will evaluate (4.3.10), preferably in the shortest time possible. It is common to use a DAG (Directed Acyclic Graph) representation of expressions in order to analyse this problem. We can represent (4.3.10) and (4.3.11) by the DAGs in figure 4.3. Note that the *direction* of the arcs is implicit in these diagrams by the ordering vertically down the page. We can immediately see that the property of having no common terms (or subexpressions) corresponds to the DAG being a tree. For completeness, a straight-line program[9] implementing (**) is given by

$$\begin{aligned} r_1 &:= B; \\ r_1 &:= r_1 \times C; \\ r_1 &:= r_1/A; \\ r_2 &:= D - r_1; \\ r_1 &:= 1/r_2; \end{aligned}$$

[9]Straight-line programs play an important role in computer science, not only practically but theoretically as well. Lynch [533] has suggested a number of reasons why straight-line program length is a good measure of algorithmic complexity. Straight-line programs have also been used as a method for representing and manipulating polynomials in symbolic algebra systems [282](see also [367]on the optimal evaluation of expressions in such algebra systems). We return to this aspect in the following subsection.

Figure 4.3: DAG representations for (4.3.10) and (4.3.11).

where r_1 and r_2 are *registers.*

In generating code within a compiler it is usual to *not* assume that the algebraic laws (such as distributivity) hold because of the rounding error in floating point arithmetic. Nevertheless, the problem of optimal code generation for arbitrary expressions (with common subexpressions) is NP-complete [14] (there are still difficulties in determining the evaluation order over the DAG). The problem is relatively easy for trees [13] or "collapsible graphs" [661]. Taking account of associative and commutative properties of operators is easy on trees [728] but is difficult otherwise.

The general problem (which is of interest to us) where both algebraic laws are assumed to hold and common subexpressions or repeated terms can occur "has received little attention in the literature" [324, p.633]. The only papers explicitly addressing this issue are [324] and [106]. The second reference simply reports the application of a method similar to Horner's rule [466, p.467]. Gonzalez and Ja'Ja' [324] have made a detailed and intricate study of the problem. They consider only multiplications and additions and assume that the distributive law holds. They develop an algorithm which will transform an expression DAG into a tree whenever this is possible and another which determines whether an expression calculated by the new DAG is the same as the original one. Gonzalez and Ja'Ja' prove a theorem (their theorem 5.1) which characterises when an expression can be transformed into a tree. It essentially says that this can occur only when there are no repeated terms when expressed in a standard canonical form. (This is hardly surprising.) They also present a number of complexity results including the fact that even if the arithmetic expressions are of degree 2 (degree of the expression when viewed as a polynomial), then determination of an equivalent expression with the minimal number of arithmetic operations is NP-hard [323]. Their results are not extendable to the more general case we are interested in (arbitrary expressions including subtraction and division operations).

More complex results have been obtained recently by Reif, Lewis and Tarjan

[680,681]. They consider more general program structures (not just straight-line programs), and more general arithmetic operations. Their original motivation was the "symbolic analysis" of programs. This involves constructing, for each expression in a program, a symbolic expression whose value is equal to the value of the program expression for all possible executions of the program. Whilst reasonably efficient algorithms for a restricted class of programs can be developed [681], the general problem, even over the integers, is not only difficult but it is *provably undecidable.*[10] Whilst this result only shows the impossibility of general *global* flow problems, it certainly removes some of the hope we may have had for the possibility of automatic rearrangement of even simple programs for the purposes of probabilistic arithmetic. Note that whilst there exist a number of tools and results on more general graphical structures for representing programs with control flow (such as the program dependence graph [120,397,475,725]), these are unlikely to be of value for our purposes.

Whilst the above results do not help us greatly in solving the problem of repeated variables in probabilistic arithmetic, they do provide good understanding of the difficulties involved. Manes [540] has made use of a general semantics of flowgraphs [32,541] in order to prove his powerful "meta-theorem" on the effect of tree structures in uncertainty calculi. There is still some hope for useful results in this area, but see our comments in the following subsection.

4.3.4 Conclusions: The Difficulties of Manipulating Graphs of Random Variables

Having trekked through such forests of trees what can we conclude? Unfortunately very little which can not be summarised in the motto *Trees are trivial; Nets are not.* We have seen that whether it be belief networks, covariance structure models or arithmetic expressions in compiler construction; the existence of a tree structure makes the problem feasible.[11] Whilst there are other classes of graphs which tend to be relatively easy (such as series-parallel graphs[12]), in general non-treelike structures cause severe difficulties.

Ironically, the relationship between graph structures and probability theory (from the point of view of rearranging expressions to remove common subexpressions) may well be in the opposite direction to that which we envisaged. In [282], Freeman *et al.* have considered the use of straight-line programs as a representation of polynomials in symbolic algebra systems. They often obtain large (greater than 5000 line) programs in the course of computations and have investigated ways of optimizing

[10] In order to prove this important result Reif and Lewis [680, p.308] make use of Matijasevic's [552] negative solution to Hilbert's tenth problem which says that determining whether a polynomial $Q(X_1, X_2, \ldots, X_k)$ has a root over the integers is recursively unsolvable.

[11] Recently a very wide variety of "easy" problems on graphs have been unified by the notion of tree-decomposibility [33]. These problems extend past the implementation of uncertainty calculi.

[12] Sahner and Trivedi [702] study probabilistic problems very similar to ours (they consider addition, max, min and selection of kth largest operations on random variables) on series-parallel graphs. Their method of solution, perhaps not surprisingly, entails decomposing the graph into a tree using the algorithm in [838].

these programs in the manner of [15]. A very efficient and clever method they have developed to do this is their StraightOpt3 procedure which works as follows:

> All instructions in the program are evaluated at random values for the indeterminates modulo a random prime. Then a binary search tree is built to sort their values. If a value of a new instruction is found in an existing leaf, the corresponding instructions are assumed to compute the same function and the new one is eliminated from the program. The algorithm is Monte Carlo and its running time is that of sorting, essentially $O(l \log l)$, where l is the length of the [straight-line program], since, with high probability, the search tree is well balanced [282, p.227].

Application of StraightOpt3 to a 11614 line program produced by a factorisation algorithm resulted in a 27% saving in instructions.

Whilst StraightOpt3 is Monte Carlo and thus can be wrong, the probability of this occurring could be made arbitrarily low. Thus again we see the advantage of Monte Carlo algorithms. Such algorithms have been used for multi-dimensional integration for some time now, and have recently found applications in primality testing, optimization [213] and other "hard" problems. It is our contention that ultimately it will only be Monte Carlo based methods which will be able to handle complex systems of interacting uncertainties.

4.4 Other Lower and Upper Probability Methods

Our results indicate that an objectivist, frequency- or propensity-oriented, view of probability does not necessitate an additive probability concept, and that interval-valued probability models can represent a type of indeterminancy not captured by additive probability.
— Peter Walley and Terrence Fine

My preference is for intervals because they can be based on objective knowledge of distributions, and because this compatibility is demonstrable.
— Henry Kyburg

The use of dependency bounds or Boole-Fréchet bounds leads to lower and upper probability distributions or probabilities. Since there have been a number of proposed systems utilising lower and upper (or interval-valued) probabilities in recent years, it seems worthwhile to compare these models with our use of lower and upper probabilities. We will see that interval-valued probabilities have been introduced in a wide variety of ways. Some authors have explored epistemic (subjective) probabilites and credal states in this manner. Others, notably Fine and his co-workers, have proposed an objective frequentist interpretation of interval-valued probabilities and have shown how such effects can arise in real physical systems. We will examine these and other systems below. We do not go into these in much detail. In some cases (especially Fine's work), there are a lot of very technical results which it is impossible to summarise in a couple of pages. Instead we will present a fairly non-technical overview.

4.4.1 The Dempster-Shafer Theory of Belief Functions

I do not believe in belief.
— Karl Popper

First we will examine the Dempster-Shafer theory of belief functions which has attracted a lot of attention in recent years. It is based on Shafer's [732–738] extensions of Dempster's [198,199,199–202]. The method has found a number of applications [85,322,528], and has been recommended as being suitable for uncertainty propagation in expert systems. It has recently been extended by Yager to enable representation and arithmetic operations on belief functions representing imprecise numbers [893].[13] We shall see below that the lower and upper probabilities arising in the Dempster-Shafer theory are quite different to those arising from our dependency bounds. The main difference is that we consider the bounds to be *secondary* to the primary notion of a single valued probability. Lower and upper probabilities arise in Dempster-Shafer theory because of imprecise assignment of the probabilities in the sample space: probabilities are assigned to subsets of the sample space rather than single elements of it. Thus they are more of a *primary* notion.

The Elements of the Dempster-Shafer Theory of Evidence

The following brief summary is based on that appearing in [894]. Let X be the *universe of discourse* (akin to the sample space). We assume $X = \{x_1, x_2, \ldots, x_n\}$ is finite, although most results can be extended to infinite spaces [735]. Let m be a measure on 2^X (the power set of X) such that

1. $0 \leq m(A) \leq 1 \quad \forall A \subset X$,
2. $m(\emptyset) = 0$,
3. $\sum_{A \subset X} m(A) = 1$.

The measure m is called a *basic probability assignment* (bpa) function. A *focal element* is any subset A of X such that $m(A) > 0$. If all the focal elements are singletons, then the belief structure is called a *Bayesian belief structure* and all the results reduce to the standard rules of probability theory. The union of all the focal elements of a belief function is called the *core*.

The *Belief* and *Plausibility* functions defined by

$$\mathrm{Bel}(B) = \sum_{A \subset B} m(A) \qquad A \subset B \subseteq X \tag{4.4.1}$$

[13] Yager's work can be considered to be a straight-forward analogue of the histogram or DPD methods for calculating functions of random variables in terms of discrete representations of their distribution functions (see section 2.4). Yager does not consider methods of condensation of the resultant belief structures (although this would not be difficult to do). In this sense Yager's results are trivial, and since he has not described any applications of the method, it seems to be of little interest to us here. Of course this is not to say that it does not have the *potential* to be useful. Indeed, this is why we are examining the basis of the Dempster-Shafer theory.

and

$$\mathrm{Pl}(B) = \sum_{A \cap B \neq \emptyset} m(A) \qquad A, B \subseteq X \tag{4.4.2}$$

are important in the theory and behave like lower and upper bounds on $P(A)$ (we always have $\mathrm{Bel}(A) \leq P(A) \leq \mathrm{Pl}(A)$). Writing the complement of B as $\overline{B} = X \setminus B$, we can state an important property of Bel and Pl as

$$\begin{aligned} \mathrm{Bel}(B) &= 1 - \mathrm{Pl}(\overline{B}) \\ \mathrm{Pl}(B) &= 1 - \mathrm{Bel}(\overline{B}). \end{aligned}$$

This can be rewritten as

$$\mathrm{Bel}(B) + \mathrm{Bel}(\overline{B}) \leq 1 \tag{4.4.3}$$

$$\mathrm{Pl}(B) + \mathrm{Pl}(\overline{B}) \geq 1. \tag{4.4.4}$$

If the belief structure is Bayesian then $\mathrm{Bel}(A) = P(A) = \mathrm{Pl}(A)$. Equations 4.4.3 and 4.4.4 are why the theory is called *nonadditive*. Shafer [734,736] sets considerable store by this, arguing that it is a more accurate model of human reasoning under uncertainty. Berres [71, pp.493ff] has shown how the "degree of nonadditivity" of a belief function can be measured in terms of an integral defined over the belief function.

The point of the above constructions is to allow a "more uncertain" assignment of probabilities to events. Instead of specifying an exact probability of an event, one can assign some probability to a subset $A \subset X$ without specifying how the probability mass is to be distributed within A. It is argued that this gives a more realistic representation of total ignorance than the assignment of a uniform or noninformative prior in the ordinary Bayesian probability model.

Combination of Belief Functions

When we have two pieces of evidence represented by belief functions we may wish to "combine" them in to one piece of evidence. The usual method of doing this is to use *Dempster's rule of combination* [732], although alternative rules do exist. Dempster's rule is also called the orthogonal sum ($\oplus$) of belief functions. The operator is essentially "anding" independent belief structures. Let m and m' be the bpa's of the belief functions Bel and Bel$'$ with cores $\{A_1, \ldots, A_p\}$ and $\{B_1, \ldots, B_q\}$ respectively. Then

$$(m \oplus m')(A) = \frac{\displaystyle\sum_{A_i \cap B_j = A} m(A_i) \cdot m'(B_j)}{\displaystyle\sum_{A_i \cap B_j \neq \emptyset} m(A_i) \cdot m'(B_j)} \qquad \text{if } A \neq \emptyset \tag{4.4.5}$$

and $(m \oplus m')(\emptyset) = 0$. Since p and q are of the order of $2^{|X|}$, this rule can have a very high computational complexity. Very recently approximations to (4.4.5) which reduce its complexity have been considered [844]. Dempster's combination rule can

be represented in terms of the theory of Markov chains. Norton [623] has used this representation to study the limiting effect of updating belief functions with (4.4.5). (That is, he studies $\lim_{n\to\infty} \bigoplus_{i=1}^{n} m_i$.) His results can be summarised by saying that evidential support tends to converge to single focal elements (or small groups of them). The actual details depend on the specifics of the corresponding Markov model used.

The assumption of indepedence of the separate pieces of evidence is important for (4.4.5) to be valid. This assumption has been criticised by a number of authors [235,849]. Williams [868] has argued that not only should independence not be assumed (for *epistemic* probabilities), but that the very notion of independence "is not defineable within the terms of the theory given its view of the nature of evidence" [868, p.387]. Errors resulting from repeated nodes (common cause events) in a fault "tree" when analysed using Dempster-Shafer theory have been reported by Guth [347]. His suggested solution to this amounts to the computation of a global belief function (akin to working with the complete joint distribution). However, as he confesses in a footnote, this scheme is impractical (computationally): for even a small example it was necessary to resort to Monte Carlo simulation. Dempster and Kong [203] have made some progress in applying the ideas of Lauritzen, Spiegelhalter and Pearl (see the previous section)[14] on graph theoretical techniques for handling complex probability structures to the Dempster-Shafer framework. They use the idea of trees of cliques in order to be able to independently combine subgroupings of nodes. Nevertheless, they too conclude: "We expect, however, that realistic practice will quickly drive us to Monte Carlo approximations, the study of which is just beginning" [203, p.367] Further work on more complex structures of belief functions has been reported by Kohlas [469] and Shafer and Logan [739].

Further Problems with the Dempster-Shafer Theory and its Relationship to Classical Probability Theory

As well as the difficulties of independence in combinations reported above, the Dempster-Shafer theory has been criticised in terms of the empirical acquisition of belief functions. Whilst introduced by Dempster in the context of sampling inference [198,200,202], there are many difficulties still to be resolved. Lemmer [510] has said that there is

> strong support for the conjecture that Dempster-Shafer theory cannot, *in principle,* be empirically interpreted as a calculus applicable to sample spaces, if the belief functions about these sample spaces arise from the real world by any sort of *reasonably error free* process of observation. Thus though the numbers produced by this theory are often termed probabilities, they cannot be interpreted as probabilities about sample spaces [510, p.118].

[14] Pearl [638] has compared the Dempster-Shafer theory with Bayesian probability structures on trees and has argued in favour of the latter.

Statistical aspects of Dempster-Shafer theory have recently begun to receive further attention [283,399].

It is tempting to ask whether the problems associated with Dempster-Shafer theory are simply a result of it masquerading as ordinary probability theory. In any case it is interesting to examine to what extent the procedures of the theory can be derived from the standard basis of probability theory (based on Bayesian belief structures).[15] Baron [56] has examined Dempster-Shafer theory in the light of "second-order probabilities" (probabilities of probabilities), and has shown that Dempster's rule of combination is a special case of a more general rule for combining second order probabilities. This general rule is derived using just the usual basis for probability. Baron considers belief functions to be a special case of second-order probabilities.

Falmagne [259] has studied the embedding of belief functions in higher-order probability spaces and has shown that "the values of the belief function have ... an interpretation in terms of events, in the embedding space, involving random variables" [259 p.35]. He proves a theorem (his theorem 4) which gives a representation of a belief function Bel over a finite set X in terms of a collection $\{\beta_\theta | \, \theta \in X\}$ of jointly distributed random variables. In this representation, the belief in $B \subset X$ is interpreted as the random event

$$\{\max\{\beta_\theta | \, \theta \in B\} > 0 > \max\{\beta_\theta | \, \theta \in (X \setminus B)\}\}\,. \tag{4.4.6}$$

Falmagne explains how (4.4.6) can be understood in terms of a random utility model.

Kyburg, Suppes, Zanotti and the Relationship Between Dependency Bounds and Belief Functions

Of more interest to us is the work of Kyburg [493] and Suppes and Zanotti [807]. Kyburg proves the following theorem:

Theorem 4.4.1 *Let m be a probability mass function defined over a frame of discernment θ. Let* Bel *be the corresponding belief function,* $\mathrm{Bel}(X) = \sum_{A \subset X} m(A)$. *Then there is a closed, convex set of classical probability functions S_P defined over the atoms of θ such that for every subset X of θ,*

$$\mathrm{Bel}(X) = \min_{P \in S_P} P(X).$$

This theorem says that any belief function can be represented as a lower probability. The converse is not true without a further restriction. That is, not every lower probability is a belief function. Sufficient conditions are given by Kyburg's theorem (A.2):

[15] The question of whether Dempster-Shafer theory is *better* than ordinary probability theory for some problem is a different matter altogether. We will briefly discuss this below in the context of other uncertainty calculi as well. For now just note that Smets [761] has constructed an example problem which he solves with both theories. The reader is then invited to choose which answer he likes best and thus choose the appropriate theory!

Theorem 4.4.2 *If S_P is a closed convex set of classical probability functions defined over the atoms of θ, and for every $A_1, \ldots, A_n \subset \theta$,*

$$\min P(A_1 \cup \cdots \cup A_n) \leq \sum_{I \subset \{1,\ldots,n\}} (-1)^{|I|+1} \min P\left(\bigcap_{i \in I} A_i\right), \tag{4.4.7}$$

then there is a mass function m defined over the subsets of θ such that for every X in θ, the corresponding Bel *function satisfies*

$$\text{Bel}(X) = \min_{P \in S_P} P(X).$$

This says that for a lower probability function to be a belief function it must be a *capacity of infinite order* [139,144]. (The right-hand side of equation 4.4.7 describes a capacity of order n, which is the highest order possible here.) Not all lower probabilities satisfy this. For example, Suppes and Zanotti [807, p.437] point out that if $\mathcal{P}$ is some *arbitrary* but nonempty set of probability measures, then

$$\underline{P}(A) = \inf_{P \in \mathcal{P}} P(A) \tag{4.4.8}$$

is not a capacity of infinite order, and "thus cannot be generated by a random relation on a probability space."

The result of this is that belief functions are special types of lower probabilities. Lower probabilities defined by (4.4.8) (and the analogous upper probabilites defined by $\overline{P}(A) = \sup_{P \in \mathcal{P}} P(A)$) are not belief and plausibility functions. Since we feel that it is *these* lower and upper probabilities that are of the most interest (they are how the dependency bounds arise), the theory of belief functions has little to contribute to the theory of dependency bounds and the lower and upper probabilities generated by them.

Dempster-Shafer Theory as an Alternative Uncertainty Calculus

Although Dempster-Shafer theory is of no apparent value for our dependency bound studies, it does have a number of other advantages for the purposes of a general calculus of uncertainty. We do not have the space (or the motive) to review this here. Instead we simply refer the reader to some recent reviews on this topic. Bonissone and Tong [96] and Pang *et al.* [630] have both presented brief reviews of different methods. Two better reviews giving a lot more detail are [320,372]. These papers compare Bayesian probability, Dempster-Shafer theory, Fuzzy set theory and the MYCIN certainty factors. The conclusions drawn include the fact that each method has advantages and disadvantages and thus no one method is best in all circumstances; in some applications it does not seem to matter much which is used, although in others it does; and all four methods are quite closely related (see especially [372, p.711]). Relationships between belief functions and other uncertainty calculi have also been explored in [340,738].

4.4.2 Fine's Interval-Valued Probabilities

Fine [268,339,422,484,631,850,851,880] and his co-workers have developed a theory of interval-valued probability which has some elements similar to the Dempster-Shafer theory. However Fine's motivation was rather different to Shafer's. Whereas Shafer's theory was developed from the point of view of epistemic (subjective) probability judgements, Fine has worked closer to the objective interpretations of probability. This is not to say he discards epistemological uncertainties. In fact he suggests that his model can handle both ontological and epistemological indeterminancies [851]. Elsewhere, Wolfenson and Fine [880] have considered the decision theoretic viewpoint (instead of the inferential).

From an engineering point of view one of the most interesting applications of the theory is the construction of stochastic models for empirical processes that are stationary but have fluctuating (diverging) time averages. These can not be modelled by a standard probability model (by the Ergodic theorem), but they can be modelled by interval-valued probability models [339,484]. An example of this is the modelling of flicker noise in quartz crystal oscillators. We do not have the space to provide any technical details on Fine's work here. A very good recent review article has been published by Fine [268]. In any case, the main point is that he considers lower and upper probabilities which arise in circumstances quite different to those we study (which arise from dependency bounds). However there seems to be no reason why these two approaches could not be integrated, although further research is needed in order to achieve this. What we feel is the most surprising aspect of Fine's theories is that they do not appear to have been adopted and studied further by other authors. It is not just a matter of developing a nice theory for its own sake. As Fine has said

> Overall, our objective is not the creation of yet another anemic mathematical 'theory'. Rather we wish to come to grips, through selected probability-like structures, with certain features of naturally occurring nondeterministic phenomena so that we can better understand and make use of these phenomena.

4.4.3 Other Lower/Upper Probability Theories

There have been a wide variety of other lower/upper probability theories proposed over the years. We just mention some of them briefly.

Epistemic theories of interval-valued probabilities have been studied by Koopman [472,473], Smith [762,763] and Suppes [804,805]. Higher-order probabilities (the probability of a probability) are discussed by Skyrms [760], Cyranski [179], and Good [326,327]. Models of *comparative* probability (instead of numerical probability) have also been suggested as being more suitable for modelling human judgement [269, 850]. Issues regarding making decisions based on probability bounds or lower/upper probabilities are discussed in [522,535]. (There is in fact no great difficulty in making decisions based on interval-valued probabilities.) Some good general discussions of

the idea of "intervalism" (use of lower and upper probabilities in preference to single valued ones) can be found in [492,494,513–515,523]. Levi's discussion in [514, section 9.8] is particularly relevant with a regard to our earlier remarks comparing the two approaches of taking probability intervals as primary or secondary objects. Levi takes convex sets of credal states and derives lower and upper probability envelopes. He argues (page 200) that this is preferable to beginning with an interval-valued probability measure and then developing other notions in terms of it.

Because of the amount of literature on the topic (we have not been at all exhaustive in our above citations), we do not make any further attempt to synthesise or summarise what are in many cases quite divergent opinions. We feel however that the main aim of the section has been achieved: to show that there is a wide variety of lower/upper probability theories, but in all cases the lower/upper probability arise in quite a different manner to our dependency bounds.

4.5 Fuzzy Arithmetic of Fuzzy Numbers

The concept of fuzzy numbers is expected to play important roles in solving problems describing ambiguous systems such as mathematical programming and deterministic systems as is applied to fuzzy language probability, fuzzy programs, fuzzy neurons etc.

— Masaharu Mizumoto and Kokichi Tanaka (1976)

Algebraic calculus on real fuzzy sets is devoted [sic] to play a central role in the development of the theory of fuzzy sets and its applications to systems science.

— Didier Dubois and Henri Prade (1979)

A great amount of work has already been accomplished. However, the ability to apply fuzzy concepts to practical problems requires a somewhat deeper understanding of the specificity of Zadeh's theory.

— Didier Dubois and Henri Prade (1980)

It is still too early to utter definite judgements about the actual benefit of introducing fuzzy numbers in topology or random variable theories, as well as the usefulness of this concept in engineering applications.

— Didier Dubois and Henri Prade (1987)

Fuzzy arithmetic is a topic, rather similar to probabilistic arithmetic, which has received considerable attention in recent years. In this section we will compare these two ideas in some detail. We will study the differences between the two ideas in terms of their philosophical basis and interpretations, practical computation methods and practical efficacy. We will show that the fuzzy arithmetic combination rules are in fact *almost identical* to the dependency bounds we developed in the previous chapter. This allows a probabilistic interpretation of fuzzy set operations. Much of what we have to say may be considered controversial to some, but we feel our arguments are sound and there is a only a very small amount of speculation (and this is clearly indicated as such). We begin with an outline of fuzzy sets.

Given some universe U, a *fuzzy set* F [897] on U is an imprecise subset of U. The set F is described by its membership function μ_F which takes values on $[0,1]$. Thus if $U = \Re$, then for some $x \in U$, if $\mu_F(x) = 1$, then x is definitely in F; if

$\mu_F(x) = 0$, then x is definitely not in F. The value of $\mu_X(x)$ expresses the "degree of membership" of x in F. As long as there is at least one x such that $\mu_F(x) = 1$, then F is said to be normal. A fuzzy set is usually considered to represent linguistic vagueness rather than stochastic uncertainty. However this is not always the case.

Fuzzy sets can be combined under the normal set-theoretic operations as follows:

$$\begin{array}{rrcl} \text{Intersection} & \mu_{F\cap G}(x) & = & \min(\mu_F(x), \mu_G(x)) \\ \text{Union} & \mu_{F\cup G}(x) & = & \max(\mu_F(x), \mu_G(x)) \\ \text{Complement} & \mu_{\overline{F}}(x) & = & 1 - \mu_F(x). \end{array}$$

Alternative combination rules defined in terms of t-norms and t-conorms have also been considered. We will discuss these below. Many other operations can be performed on fuzzy sets, but as we shall see there is considerable dispute about the correct definitions and interpretations for many of these. Of most interest to us is the situation where the fuzzy set represents an imprecise number. In this case one can define rules for arithmetic operations on these fuzzy numbers. This is the subject matter of fuzzy arithmetic and is the focus of the present section.

There are two main types of fuzzy numbers which we will discuss. (Whilst Dijkman *et al.* [214] talk of several different types, they can all be considered to be variations of the two types we will consider.) The first type, studied by Höhle, Klement, Lowen and others [390,391,465], considers a fuzzy number to be defined by a nondecreasing membership function (which looks like a probability distribution function). This approach is very similar to the use of distribution functions in probabilistic metric spaces [718]. We will briefly return to this idea of a fuzzy number below where we compare it with the other (more widespread) notion.

Fuzzy numbers are considered by Dubois and Prade and others [239,432,579, 898] to be a fuzzy generalisation of the notion of an interval in interval analysis: they are fuzzy sets of real numbers. We say X is a fuzzy number if its fuzzy membership function μ_X is normalised ($\sup_x \mu_X(x) = 1$) and is *pseudo-convex* or *quasi-convex* [529,833] ($\mu_X(y) \geq \min(\mu_X(x), \mu_X(z))\ \forall y \in [x, z] \subset \Re$). (Note that pseudo-convexity, otherwise known as unimodality, has sometimes erroneously been called simply "convexity" in the fuzzy set literature. Of course convexity *is* intimately related to unimodality: see [211].)

By an analogy with the method of calculating functions of random variables, the *extension principle*[16] [898, p.236] is used to calculate the membership function of $Z = f(X_1, \ldots, X_n)$ where X_i are fuzzy numbers with membership functions μ_{X_i}, $i = 1, \ldots, n$. This says that

$$\mu_Z(z) = \sup_{z=f(x_1,\ldots,x_n)} \min(\mu_{X_1}(x_1), \ldots, \mu_{X_n}(x_n)). \tag{4.5.1}$$

[16] Manes [539, p.604] has said that "Mathematicians do not usually feel that the existence of such formulas deserves to be called a 'principle' ...". He goes on to point out a number of severe deficiencies in the mathematical basis of fuzzy set theory and contrasts it with Topos theory and his general distributional set theories (see [540]). We do not have space here to describe Manes' substantial results and we simply remark that his work should be classed amongst the major contributions of the field.

Some authors *have* given derivations of the extension principle [606,853], but these are quite different to Manes's and their interpretations of fuzzy sets have a number of shortcomings.

When $n = 2$ and $f = +$, we have the "sup-min convolution"

$$\mu_Z(z) = \sup_{x_1+x_2=z} \min(\mu_{X_1}(x_1), \mu_{X_2}(x_2)).$$

The min operator is the fuzzy set intersection operator and it can be replaced [26] by a general t-norm T. We then obtain the more general extension principle (for $n = 2$) as

$$\mu_{f(X,Y)}(z) = \sup_{f(x,y)=z} T(\mu_X(x), \mu_Y(y)). \tag{4.5.2}$$

This is obviously very similar to the $\tau_{T,L}$ operations we have encountered in studying dependency bounds. The difference is mostly in interpretation (although there is the minor difference that μ_X is unimodal whereas F_X is monotonic — this leads to only slight changes in the procedures used to numerically calculate (4.5.2)). The theory of fuzzy numbers outlined above is nowadays described in terms of "possibility theory" and "possibility distributions." We shall return to this topic after we have examined the general interpretation of the modal terms "possible," "probable," and "necessary." A review of the theory of fuzzy numbers including work up to 1987 is given by Dubois and Prade in [239].

The rest of this section will cover the following material:

1. Numerical methods for calculating the fuzzy number combinations described by the extension principle. We will also consider the notion of interactivity of fuzzy numbers which is an analogue of dependence for random variables.

2. A general critical discussion of the interpretations of the modal terms "possible," "probable," and "necessary." This was undertaken because of the recent appeal by fuzzy set theorists to a "theory of possibility" on which fuzzy numbers are based. We show that the modal logic interpretations of "possible" which they use to support their development of posibility theory are not the only valid ones, and that a probabilistic semantics for modal terms is preferable for a number of reasons.

3. We also present a general discussion of the relationship between fuzzy set theory and probability. This contentious issue is as old as the theory of fuzzy sets. We aim to restrict ourselves to a concise review of the arguments on this topic and try to add something of our own.

4. The relationships between interval fuzzy numbers, especially in terms of possibility and necessity measures, and confidence curves is then examined. Confidence curves are an old but rarely used statistical technique which we show are very similar in their form and intuitive interpretation to fuzzy numbers.

5. Finally we summarise what we see as the relationships between fuzzy numbers and random variables in the light of the dependency bound operations.

This section is rather longer than most in this chapter (and thesis) solely for the reason that there is now an enormous literature (over 5000 papers) on the theory of fuzzy sets, and thus there are many different points to discuss.

In a single sentence, our conclusions could be put as follows: *The theory of fuzzy numbers, as developed to date, does essentially nothing new compared to the theory of random variables (when proper account is taken of missing independence information), and would appear to be of little value in engineering applications.*

4.5.1 Numerical Methods for Fuzzy Convolutions and the Notion of Interactive Fuzzy Variables

There are a variety of methods available for numerically calculating fuzzy arithmetic convolutions (5.1). They can be classified into three basic types:

Trapezoidal, triangular and L-R fuzzy numbers [113,224–226,228,239,432]:
These are all parametric methods. The idea is that a fuzzy number is represented by 2, 3 or 4 parameters and the arithmetic operations are implemented in terms of operations on the parameters. The implicit assumption in these methods is that these few (4) parameters "are adequate to capture the fuzzy uncertainties in human intuition" [512, p.47]. The formulae for products and quotients are necessarily approximate. The only reason the sum and difference formulae are exact is because the t-norm Min is used and so there is a "shape-preservation" effect. We discuss this in detail later on.

The basic idea of L-R fuzzy numbers is to represent the membership function μ_X of a fuzzy number by

$$\mu_X(x) = \begin{cases} L\left(\frac{\underline{m}-x}{\alpha}\right) & x \leq \underline{m} \\ 1 & x \in [\underline{m}, \overline{m}] \\ R\left(\frac{x-\overline{m}}{\beta}\right) & x \geq \overline{m}. \end{cases}$$

here $[\underline{m}, \overline{m}]$ is the *core* and $[\underline{m}-\alpha, \overline{m}+\beta]$ is the *support* of μ_X. The functions L and R from $\Re$ onto $[0,1]$ are known as *shape-functions.* Fuzzy number arithmetic operations can be determined (sometimes exactly) in terms of $\underline{m}$, $\overline{m}$, α and β for the fuzzy numbers involved.

Sampling along the x-axis (to represent $\mu_X(x)$) [407,408,544]:
This simple technique which entails only using values $\mu_X(x_i)$ for some set $\{x_i\}_{i=1}^n$ is related to Baekeland and Kerre's piecewise linear fuzzy quantities [42] and the DPD method [423,425] for calculating convolutions of probability densities, and suffers from the same sorts of problems. It should be said however that the problems can be circumvented in similar ways. For example, one of Dubois and Prade's [223] criticisms of Jain's method [407,408] can be solved by using the *condensation* procedure as used by Kaplan in [423] or by us in chapter 3 [874].

Interval Arithmetic on the Level Sets [137,138,218,431,480,669]:
This is perhaps the most interesting method. We examine it below and show that

1. It is particularly simple for $T = M$ for a good reason and it can be extended to other t-norm intersection operators.

2. The apparent success and simplicity of the method are not sufficient justification for the claim that fuzzy arithmetic is the *natural* generalisation of interval arithmetic.
3. The "fundamental result" that allows calculation of fuzzy number convolutions in terms of level sets is a special case of a duality result that has been known for some time.

Let us write F_α for the α-cut or α-level set of some fuzzy set F with membership function μ_F. That is, $F_\alpha = \{x \in \Re | \, \mu_F(x) \geq \alpha\}$. Dubois and Prade [240, p.39] state the following "fundamental result:" Let M and N be two fuzzy intervals with upper semi-continuous membership functions and assume that $M_\alpha \subset \Re$ and $N_\alpha \subset \Re$ for $\alpha > 0$. Let $f: \Re^2 \mapsto \Re$ be continuous and order-preserving (that is, $\forall u \geq u'$, $\forall v \geq v'$ $f(u,v) \geq f(u',v')$). Then, for all $\alpha > 0$,

$$[f(M,N)]_\alpha \;=\; f(M_\alpha, N_\alpha). \tag{4.5.3}$$

This means that $f(M,N)$ can be calculated in terms of interval arithmetic on the α-cuts of M and N. Dubois and Prade credit this result to Nguyen [615] who says the original idea was due to Mizumota and Tanaka [579].

Fenchel's Duality Theorem

We will now show that (4.5.3) is in fact a special case of a more general duality result [278] which we made use of in chapter 3 to calculate dependency bounds numerically in an efficient manner. Let $\Phi(p,q;r,s)$ denote the set of all non-decreasing functions from $[p,q]$ into $[r,s]$ satisfying $\Phi(p) = r$ and $\Phi(q) = s$, where $p < q$, $r < s$ and $p,q,r,s \in \Re^* = \Re \cup \{-\infty, \infty\}$. Recall the definition of the quasi-inverse $F^\wedge$ of a function $F \in \Phi(p,q;r,s)$ [278,718]:

$$F^\wedge(y) \;=\; \sup\{x | \, F(x) < y\}.$$

Define the binary operation $\tau^\wedge_{C,L}$ on the space of quasi-inverses of the elements of $\Phi(p,q;r,s)$ by

$$\tau^\wedge_{C,L}(F^\wedge, G^\wedge)(x) \;=\; \inf_{C(u,v)=x} [L(F^\wedge(u), G^\wedge(v))]. \tag{4.5.4}$$

Theorem 4.5.1 *Let $C \in \mathcal{C}$, $L \in \mathcal{L}$, $F, G \in \Phi(p,q;r,s)$. Then*

$$[\tau_{C,L}(F,G)]^\wedge \;=\; \tau^\wedge_{C,L}(F^\wedge, G^\wedge). \tag{4.5.5}$$

Upon letting $C = M$, it can be seen that (4.5.5) says the same thing as the "fundamental result" (4.5.3). This is because the infimum in (4.5.4) will always occur at $u = v = x$ since L, F and G are non-decreasing. In other words we can write

$$\tau_{M,L}(F,G)(x) \;=\; [L(F^\wedge, G^\wedge)]^\wedge(x). \tag{4.5.6}$$

When $L = \text{Sum}$, (4.5.6) reduces to

$$\tau_{M,+}(F,G)(x) \;=\; [F^\wedge + G^\wedge](x).$$

This was proved directly by Sherwood and Taylor in 1974 [743] (see their proposition 4 on page 1258) and used by Klement in [465] who was only aware of the result for $T = M$ and $L = +$ and could not see how it could be extended. Equation (4.5.5) can be used for calculating fuzzy number convolutions by decomposing membership functions into an increasing part and a decreasing part and operating on two parts separately (see chapter 6 below). In fact this duality result can be traced back to a result of Fenchel [263] in the theory of convex functions.

Fenchel's duality theorem is best explained with the aid of some diagrams. This is *not* done in [687] nor in Fenchel's original paper [263]. We shall follow Luenberger's presentation [530] which does contain diagrams. The main idea we require is that of a *dual functional* on the *dual space* of some linear vector space. We consider a convex function f over a convex domain C. The dual space C^* and the dual functional f^* are given by

$$C^* = \{x^* \in X^* | \sup_{x \in C}[\langle x, x^* \rangle - f(x)] < -\infty\},$$

and

$$f^*(x^*) = \sup_{x \in C}[\langle x, x^* \rangle - f(x)],$$

where $\langle x, x^* \rangle$ denotes the value of the functional x^* corresponding to x. (Note that x^* is a functional, even when x is a point.) The space X^* is called the dual space of a linear vector space X. It comprises the linear functionals on X and is itself a linear vector space.

This should be much clearer upon consideration of figure (4.4) which depicts a convex region C with a convex function f defined over it. The *conjugate* function f^* is a functional defined over the space X^* of dual functionals. In the example we consider, $f^*(x^*)$ is a hyperplane. In other words, for each *point* $f(x)$ there corresponds a *line* $f^*(x^*) = \langle x, x^* \rangle - r$. We refer to the set $[f, C]$ as the *epigraph* of f over C. The set $[f, C]$ is itself convex if f and C are. One can similarly define conjugate concave functionals of concave functionals of concave functions g on a concave set D: We have

$$D^* = \{x^* \in X^* | \inf_{x \in D}[\langle x, x^* \rangle - g(x)] > -\infty\},$$

and

$$g^*(x^*) = \inf_{x \in D}[\langle x, x^* \rangle - g(x)].$$

These conjugate functions are useful in solving optimization problems [72,530]. Consider determining

$$\inf_{C \cap D}[f(x) - g(x)] \tag{4.5.7}$$

where f is convex over C and g is concave over D. Examination of figure (4.5) shows this entails finding the length of the shortest dashed line in that figure — *i.e.* the smallest vertical distance between $[f, C]$ and $[g, D]$. The Fenchel duality theorem allows this problem to be solved in terms of the conjugate functionals f^* and g^*:

Figure 4.4: Illustration of a convex function f on $C \in X$ and its conjugate dual f^* on the dual space X^*.

Figure 4.5: Illustration of the optimization problem (4.5.7).

Figure 4.6: The idea behind Fenchel's duality theorem.

Theorem 4.5.2 *Assume that f and g are, respectively, convex and concave functionals on the convex sets C and D in a normed space X. Assume that $C \cap D$ contains points in the relative interior of C and D and that either $[f, C]$ or $[g, D]$ has a nonempty interior. Suppose further that $\mu = \inf_{x \in C \cap D}[f(x) - g(x)]$ is finite. Then*

$$\mu = \inf_{x \in C \cap D}[f(x) - g(x)] = \max_{x^* \in C^* \cap D^*}[g^*(x^*) - f^*(x^*)]$$

where the maximum on the right side is achieved for sone $x_0^* \in C^* \cap D^*$. If the infimum on the left is achieved by some $x_0 \in C \cap D$, then

$$\max_{x \in C}[\langle x, x_0^* \rangle - f(x)] = \langle x_0, x_0^* \rangle - f(x_0)$$

and

$$\min_{x \in D}[\langle x, x_0^* \rangle - g(x)] = \langle x_0, x_0^* \rangle - g(x_0).$$

The idea behind this theorem is shown in figure (4.6). The minimum vertical distance between f and g is the maximum vertical distance between the two parallel hyperplanes separating $[f, C]$ and $[g, D]$.

The relationship between this theorem and theorem (4.5.1) is only sketched here. Firstly consider the more general optimization problem

$$\phi(x) = \inf_{\substack{u+v=x \\ u\in C \\ v\in D}} [f(u) - g(v)]. \tag{4.5.8}$$

Setting $v = x - u$ and assuming the conditions $u \in C$ and $v \in D$ implicitly, this can be written as

$$\phi(x) = \inf_u [f(u) - g(x-u)].$$

If we now set $h(x) = -g(x)$ (so h is concave), and apply an exponential transformation (letting $F(x) = e^{f(x)}$ and $H(x) = e^{h(x)}$), we obtain

$$\Phi(x) = \inf_u [F(u)H(x-u)].$$

Finally, by use of the multiplicative generator representation of Archimedian t-norms (or t-conorms), this can be transformed into something of the form[17]

$$\Phi(x) = \inf_u T^*(F(u), H(x-u)).$$

This is the dependency bound formula and the fuzzy number convolution formula.

The Fenchel duality theorem is very closely related to the Maximum transform developed by Bellman and Karush [64–68]. They developed their results independently of Fenchel (see [65, p.551]). The maximum transform is the basis for the T-conjugate transform studied in chapter 5.

We expect that the duality theorems discussed above will prove useful in a number of different ways. We also feel there are several new results possible. One idea which we briefly examine in chapter 6 is the discrete T-conjugate transform which we think will enable an even more efficient numerical calculation of dependency bounds and fuzzy number convolutions.

The duality theorem (theorem 4.5.1) presented above is for general t-norms. Whilst not as simple for the case of $T = \min$, it is still useful. The theory of τ_T convolutions and T-conjugate transforms developed by Moynihan [592–596] appears to offer some promise of improved methods for numerically calculating fuzzy number convolutions under general t-norm intersection operators [465,872]. A different generalisation of (4.5.6) is given by Höhle [390]. Höhle's proposition 4.7 (p.101), which we will not quote here, can be viewed as "a generalisation of [(4.5.6)] to the scope of Brouwerian lattices whose duals are also Brouwerian." Interpreting his results, Höhle says

> There always exists a stochastic model in which L-fuzzy quantities admit an interpretation of abstract valued random variables, and the binary

[17] It is only "of the form" as there are a number of complications we have glossed over here. Further details can be found in chapter 5 and [596].

> operations τ_*, $\hat{\otimes}$ correspond to the usual multiplication of random variables. Thus contrary to the widespread opinion there exists a point of view from which the theory of [0,1]-fuzzy concepts can be regarded as a part of probability theory [390, p.106].

Höhle's τ_* and $\hat{\otimes}$ operations are generalisations of τ_T and ρ_T to completely distributive complete lattices. As Dubois and Prade point out [237], Höhle's approach to "fuzzy numbers" is quite different to theirs (see also [391] and [236, p.290]). We will compare Höhle's more positivist approach with ours and others in more detail in section 4.6.

Shape Invariance of Fuzzy Number Addition under Min Intersection and its Effects

A feature of the sup-min fuzzy number addition that is presented as an advantage is its shape invariance:

> Addition of fuzzy numbers is remarkably shape-invariant contrary to random variable convolution: adding triangular distributions yields triangular distributions, adding parabolic distributions yields paraboles *etc.* [233].

Dubois and Prade [233] conclude from this that fuzzy arithmetic is better than a stochastic approach for finding the shortest path of a graph [754]:

> As soon as we allow distances between vertices to be random, we are faced with many difficulties among which are the intricate dependency of paths, the necessity of performing repeated convolutions of random variables.

Viewing the sup-min convolution as the convolution of completely positively dependent random variables provides the understanding of what is happening here. The pairwise dependence problem is ignored (because all the variables are completely positively dependent upon one another), and therefore the variables can be combined simply pairwise. Similar arguments have been given by Yazenin [895].

The ironic thing is that whilst the problem of determining stochastic shortest routes when all the variables are independent is in fact difficult, the more general and more useful problem of determining bounds on the shortest routes when the joint dependence structure is unknown is computationally simpler. This is explained in detail in Klein Haneveld's excellent paper [464] (see the discussion of this in section 2.4 above). Thus the claimed advantages of using fuzzy numbers for uncertain network problems are overstated. The more difficult but more useful stochastic problem can be solved using appropriate techniques. We feel that Klein Haneveld's work in particular demonstrates that the stochastic approach to operations research problems *is* feasible and that the benefits of the fuzzy approach [113,196,431,491, 531,669] have been exaggerated (but see the review by Zimmerman [904], especially his discussion of the lack of duality results for fuzzy programming (p.56)).

Relationships with Interval Arithmetic

We turn now to a brief examination of the relationship between fuzzy arithmetic, probabilistic arithmetic and interval arithmetic. Our main point is that the following point of view (due to Dubois and Prade) is at best quite misleading, and, in our opinion, completely wrong. Dubois and Prade [228,240] contend that fuzzy arithmetic is *the* natural generalisation of interval arithmetic [17,583,584]:

> The all-or-nothing nature of interval analysis, in contrast to probability theory, which admits gradations, introduces an asymmetry between them, which one would like to remove. It is clear that the latter does not generalize the former, since a function of a uniformly distributed random variable (the probabilistic counterpart of an error interval) does not in general itself have a uniform distribution. One of the major contributions of this book is to propose a canonical generalization of interval analysis that admits of appropriate gradations [240, p.6].

Our argument runs as follows: The shape-preservation property (see above) which holds when $T = M$ should be considered as a point *against* fuzzy arithmetic. If the shapes of the initial membership functions are preserved so well, then one may as well dispense with the function all together and simply work with the support interval. The fact that adding two independent uniformly distributed random variables results in a triangular distribution shows that we have somehow included some information about the addition operation. The most important point though is the trivial fact that random variable addition *does* naturally generalise interval arithmetic if one considers the support of the distributions of the random variables involved. In fact, one can think of the probability distributions describing the distribution of values *within the support interval:* Consider two random variables X and Y with distribution functions F_X and F_Y, having support $[\ell_X, u_X]$ and $[\ell_Y, u_Y]$ respectively. If $F_Z = df(Z = X + Y)$, then

$$\text{supp}\, F_Z = [\ell_X + \ell_Y, u_X + u_Y]. \tag{4.5.9}$$

This is simply the interval arithmetic addition of supp F_X and supp F_Y. Thus it is incorrect to maintain that "a probabilistic approach to extend (4.5.9) [interval addition] would fail" [228, p.933]. See also our discussion of the relationship between probabilistic arithmetic and interval arithmetic in [871]. We return to the topic of intervals in section 4.5.4 where we examine nested sets of confidence intervals.

Noninteraction of Fuzzy Variables

One of our dissatisfactions with fuzzy arithmetic is that there has been little careful consideration of dependency. We have already mentioned the connection between the use of min intersection operators in the extension principle and complete positive dependence of random variables. Like Suppes, we feel that the concept of independence (and hence dependence) "is one of the most profound and fundamental ideas

not only of probability theory but of science in general" [802, p.109] and thus deserves careful consideration in any uncertainty calculus. Manes [540] has shown that the phenomenon of dependency error [874] will arise in a large class of uncertainty calculi. Smith [766–768] (see especially section 3 of [767]) has argued that the notion of *conditional independence* explicated with the use of directed graphs is useful for a number of different uncertainty calculi. (We have examined the use of graphs in probability theory in section 3 above.) For now we shall explore the fuzzy set analogues of probabilistic independence. We shall see that not only is the analogue of *independence* still poorly interpreted, but the analogues of *dependence* and measures of dependence have barely been considered.

The analogue of probabilistic independence in the theory of fuzzy variables is called *noninteractivity* [228,239] or *unrelatedness* [606,673]. Noninteractivity has been assumed implicitly in our discussion of extension principle in the introductory comments to this section above. Noninteractivity is defined in terms of a "joint membership function." If

$$\mu_{XY}(x,y) \;=\; \min(\mu_X(x), \mu_Y(y)) \qquad \forall (x,y) \in \operatorname{supp} \mu_X \times \operatorname{supp} \mu_Y \qquad (4.5.10)$$

(or the equivalent statement in terms of possibility distributions), then X and Y are *noninteractive*. The interpretation of noninteractivity is rather more problematic than its definition. Dubois and Prade argue as follows:

> Noninteraction plays the same role in possibility theory as independence in probability theory, but they do not convey the same meaning at all: interaction simply means a lack of functional links between the variables while, in the frequentistic view of probability, an event A is said to be independent from B if, asymptotically, A is observed as often whether it simultaneously occurs with B or $\overline{B}$ [239, p.8].

This passage would *seem* to mean that noninteraction is a much weaker notion than independence. It is certainly not an entirely satisfactory explanation of how noninteraction is to be understood. Perhaps the main difficulty is that there is no analogue of a fuzzy event (although some authors have tried to define this [606]). In probability theory the idea of a functional link is something which may *cause* stochastic dependence, but is not considered to be *equivalent* to it (see studies on probabilistic causality [802] and section 4.3 on graphical ideas). Noninteraction of fuzzy variables is also discussed in [102,380,381,670].

Statistical independence arises in the theory of fuzzy variables when membership functions (or possibility distributions) are determined from statistical experiments:

> Possibility measures usually refer to non-statistical types of imprecision, such as the one pervading natural language and subjective knowledge. Recently, however, a statistical interpretation of possibility measures has been proposed in the framework of random experiments yielding imprecise outcomes; however, the concept of statistical independence under imprecise measurements has yet to be defined [239, p.8].

Since the notion of independence is logically prior to that of a statistical experiment [660], the above deficiency is a serious problem.

Four Types of Interaction of Fuzzy Variables

There are at least four types of interaction of fuzzy variables which have been discussed in the literature to date. We will now examine each of these in turn and we will note their relationships with each other and with the statistical concepts from which they were motivated.

The first type we consider is more accurately described as a generalisation of the *non*interactivity defined by (4.5.10). Two fuzzy variables X and Y are *weakly noninteractive* [228] or T-*noninteractive* [869] or $*$-*independent* [41] if

$$\mu_{XY}(x,y) \;=\; T(\mu_X(x), \mu_Y(y)) \qquad \forall (x,y) \in \operatorname{supp}\mu_X \times \operatorname{supp}\mu_Y \tag{4.5.11}$$

where T is some t-norm other than min. Whilst it is easy to define, very little computation seems to have been done with weakly noninteractive fuzzy variables. The "extension principle" for T-noninteractive fuzzy variables is the general form (4.5.2). Dubois and Prade [228, pp.931–933] have derived some properties of additions of T-noninteractive L-R fuzzy numbers for $T = Z$, W and Π but have not made much use of them. It is in fact possible to calculate arbitrarily good approximations to the sup-T convolutions using the discretisation of the quantiles and the additive or multiplicative generator decomposition of an Archimedean t-norm. We will study T-noninteractive fuzzy additions further in chapter 6.

The idea of *strong non-fuzzy interaction* is to *restrict* the joint membership fuction as follows:

$$\mu_{XY}(x,y) \;=\; \min(\mu_X(x), \mu_Y(y)) \quad \forall (x,y) \in D \subset \operatorname{supp}\mu_X \times \operatorname{supp}\mu_Y \tag{4.5.12}$$

This can be trivially extended to higher dimensions. Dubois and Prade [228] have used this idea to determine sums of "fuzzy probabilities" (under the restriction that probabilities always must sum to one). Strong non-fuzzy interaction is equivalent to the recently introduced statistical idea of *regional dependence* [396]. Given two random variables U and V with $F_U = df(U)$ and $F_V = df(V)$, the idea of regional dependence is where the joint distribution F_{UV} is only equal to the product of the marginals on some limited region: $F_{UV}(u,v) = F_U(u)F_V(v)$ for $(u,v) \in D \subset \operatorname{supp} F_U \times \operatorname{supp} F_V$. Thus U and V are "independent on D." Regionally dependent random variables can exhibit many of the characteristics of independence such as zero correlation, constant regression functions and zero "local dependence" (see [395]). Strongly non-fuzzily interactive fuzzy variables have also been studied by Dong and Wong [219] who have very briefly considered how the "vertex method" [218] (a generalisation of simple interval arithmetic on the level sets of membership functions) can handle such interactions which arise in the solution of certain equations.

Strong fuzzy interaction is a further generalisation of strong non-fuzzy interaction where the region D is a fuzzy set. In this case the joint membership function is given

by

$$\mu_{XY}(x,y) \;=\; \min(\mu_X(x), \mu_Y(y), \mu_D(x,y)) \qquad \forall (x,y) \in \operatorname{supp}\mu_X \times \operatorname{supp}\mu_Y. \quad (4.5.13)$$

The fuzzy set μ_D is called a *fuzzy relation.* Although Dubois and Prade [239, p.23 240, p.54] discuss the possibility of calculating functions of strongly fuzzily interactive fuzzy variables using a generalisation of the extension principle such as

$$\mu_{f(X,Y;D)}(z) \;=\; \sup_{z=f(x,y)} \left[\min(\mu_X(x), \mu_Y(y), \mu_D(x,y))\right], \quad (4.5.14)$$

they say that "the study of the properties, and even more the calculation, of $f(X,Y;D)$ are in general very difficult" [240, p.54]. We are unaware of *any* examples of such calculations in the literature, although very recently Sarna [710] has considered the fast calculation of the related but simpler formula

$$\sup_{x,y} \min\left[\mu_X(x), \mu_Y(y), \mu_D(x,y)\right], \quad (4.5.15)$$

where μ_X and μ_Y are rectangular or triangular fuzzy numbers and

$$\mu_D(x,y) \;=\; \frac{\min(x,y)}{\max(x,y)}$$

with $\mu_D(0,0) = 1$. His results appear to be of little value for our purposes.

The fourth type of interaction is a further generalisation of (4.5.11) which has only really been noted in passing. Two fuzzy variables are *weakly fuzzily interactive* [228, p.929] if

$$\mu_{XY}(x,y) \;=\; T(\min(\mu_X(x), \mu_Y(y)), \mu_D(x,y)). \quad (4.5.16)$$

No examples using this have been presented.

Before we leave the topic of interactive fuzzy variables let us briefly examine the idea of *measures of association.* By analogy with measures of association in probability theory (such as the correlation coefficient), Buckley has considered the use of a single, or perhaps interval, parameter describing the "degree of association" between two fuzzy variables. Buckley and Siler have used "probabilistic analogies rather freely" [111, p.227] to develop an *ad hoc* measure of association between two fuzzy sets (not fuzzy numbers). Their measure of association R is defined "implicitly" [111, p.222] by a strongly probabilistic analogy. They consider the lower and upper Fréchet bounds for AND and OR operations and use R as a parameter taking on values within $[-1,+1]$ to give a mixture of the extreme rules. (This is very similar to the parameterised families of bivariate probability distributions).

Our criticism of Buckley and Siler's approach is that they have effectively "put the cart before the horse." They have suggested a way of using R to vary the combination rules *without*[18] *saying where the values of R are to come from.* Of course an appeal could always be made to "subjective intuitive judgement" — a common panacea of subjective methods! Buckley has used these ideas in an expert

[18] Their unsupported statement that "The proposed measures [of association] does [*sic*] seem to offer a convenient way to estimate prior associations" [111, p.227] notwithstanding.

system [112]. He has more recently turned his attention to interactive fuzzy *numbers* [114,115] where he just defines the correlation coefficient (but does not use it) and considers the effects of strong nonfuzzy interaction on sums and products of fuzzy numbers. He considers the region D to be defined by the intersection of $\text{supp}\,\mu_X \times \text{supp}\,\mu_Y$ and a quadrilateral region. His results seem to be of little value.

Our conclusion upon examining this material can be put as follows: *The idea of dependence (or interaction) is more important in uncertain reasoning than the purely "distributional" effects.* In other words, dependence (or interactivity) is something arising in uncertain reasoning which has no real counterpart in the deterministic case. Until acceptable theories of dependence are available, the theory of fuzzy variables will remain deficient. This is true even for purely subjective interpretations. Under these interpretations there are still further difficulties in considering what is meant by independence/dependence. See the discussion of this matter with regard to subjective theories of probability in Popper's *Realism and the Aim of Science* [660]: the idea of independence turns out to be contradictory under such an interpretation.

4.5.2 The Modalities "Possible" and "Probable" and their Relationship to Fuzzy Set Theory and Possibility Theory

It is natural to admit degrees of possibility and of necessity as for probability.
— Didier Dubois and Henri Prade

There can be degrees of probability, but not of possibility.
— Alan White

The use of fuzzy numbers has lately been advocated in terms of their basis in possibility theory, which is said to be quite different to probability theory. We will now present an analysis of the notions "possible" and "probable" and will suggest that it is incorrect to talk about degrees of possibility in the manner of the advocates of fuzzy set theory. The following review of interpretations of the modalities possible and probable was motivated by the apparent lack of balance in fuzzy set theorists' discussions of the topic.

Our main thesis is that the discussion to date on the interpretation of possibility has been quite confused and mostly wrong. We will argue that the subjective emphasis is misplaced or overrated and that clear and consistent objective interpretations of possibility are available. Furthermore such interpretations clear up the distinction between possibility and probability and make clear the fact that possibility does not admit degrees. That is, it is pointless to talk of "degrees of possibility." The main point to keep in mind is that the split between epistemic and physical possibilities closely mirrors that between epistemic and physical probabilities. We do not try to argue for one or the other here (such arguments in the past have not resolved the on-going debate), but rather point out some advantages and disadvantages of these two approaches. Our own preference is for the physical point of view.

All Kinds of Possibilities

Proponents of fuzzy set based possibility theories often allude to the "possibility" of modal logic [398], stressing the distinction between these two concepts. Thus Zadeh, in his well known paper *Fuzzy Sets as a Basis for a Theory of Possibility*, says "The interpretation of the concept of possibility in the theory of possibility [based on fuzzy sets] is quite different from that of modal logic." Four years later, in a paper presented (appropriately enough[19]) at the Sixth International Wittgenstein symposium, he says:

> The possibility theory which serves as a basis for test-score semantics is distinct from — but not totally unrelated to — the possibility theories related to modal logic and possible world semantics [900, p.254].

Dubois and Prade [222,240] have appealed to modal logic, and in particular Aristotle's definition of necessity in terms of possibility (see below). They do not make a clear distinction between their notion of possibility (based on fuzzy sets) and that of ordinary modal logic. However they do distinguish between epistemic and physical possibility. We examine this distinction further below.

Our purpose here is not to answer the question "What is possibility?" or "What is the true meaning or essence of possibility?" Like Popper [656–658,660], we dismiss such essentialist questions as being irrelevant. In any case, as Morgan has argued,

> To ask for the "true" meaning of necessity and possibility is parallel to asking for the true meaning of negation. Such notions have no universally constant meaning other than the minimal way they interact with our acceptance of propositions of various sorts [589, p.52].

Instead, we will examine the distinctions between a number of different concepts of possibility. Hacking [350] distinguishes between many different kinds of possibility. (A different classification is given by Lacey [495]). The main distinction, which is of medieval origin, is between *de re* modalities and *de dicto* modalities. This division, which is still controversial, is explained as follows: *De re* modalities refer to properties of *things,* whereas *de dicto* modalities refer to properties of *propositions*. Thus we could talk of *physical* modalities (*de re*) and *epistemic* (*de dicto*) modalities. Plantinga has discussed this distinction in some detail in [649].

The notion of possibility is interpreted in modal logic by means of Kripkean [478] "possible worlds semantics" [524]. This amounts to postulating an infinity of "possible worlds" and explaining modal terms with respect to identity and common properties (or accessibility [589, pp.35–36]) across all possible worlds. An obvious criticism of this, as Loux recognises, runs as follows [524]:

[19] "Appropriately enough" because Wittgenstein developed a philosophy based on the meaning of *words.* (Recall fuzzy set theory's linguistic basis.) His famous blue book [877] opens with the question: "What is the meaning of a word?" A number of philosophers, the most eminent being Karl Popper, have explicitly argued against such a conception of philosophy [57,657].

> The trouble with possible worlds, we want to say, is that they represent an exotic piece of metaphysical machinery, the armchair invention of a speculative ontologist lacking what Bertrand Russell called "a robust sense of reality."

Another difficulty is that interpreting "possibility" in terms of *possible* worlds comes very close to being a circular argument. We need not concern ourselves with possible worlds further here (but see Morgan's probabilistic semantics of modal terms below), and we simply refer the reader to [467,525] for the history of the subject. The idea of possible worlds has been *combined* with the idea of fuzzy sets by Forbes in chapter 7 of [271]. Noting what appears (to us) to be a devastating self-criticism by Forbes [271, p.96] ("The role of modal logic is more to make the theses absolutely precise than to facilitate any substantial consequences from them..."), we simply state that Forbes attempts to use the idea of degree of membership in the possible worlds setting to examine some *sōrites* type modal paradoxes associated with vague transworld identity relations. We state his conclusions in his own words as we are unable to follow his reasoning sufficiently to make them any clearer:

> Thus while every *de dicto* modal thesis about identity has the same truth value in the present framework as it has in the classical framework, a difference emerges over the *de re, not* because identity somehow becomes fuzzy, but because *de re* sentences introduce a new fuzzy relation, that of counterparthood which in turn gives rise to degrees of possibility [271, p.179].

The Relationship between Possibility and Probability

Fuzzy set theorists often talk about the dual notions of possibility and necessity. We shall see below that their idea of the interrelationship between possibility and necessity is not shared by everyone. The duality they refer to is due to Aristotle and is widely accepted in modal logic. It says that a proposition is necessary as soon as the converse proposition is not possible. We will only concern ourselves here with possibility.

The notion of possibility is hardly new in a probabilistic context. Recall the equipossible *definitions* of probability [349,351]. It is interesting to note that whilst the Greeks (*e.g.* Aristotle) did have a notion of empirical (as opposed to logical) possibility, and thus "believed in the existence of real contingency" [707, p.39], they did not develop a notion of the probable, nor did they observe the long run stability of relative frequencies [706, pp.179–181]. This is quite surprising given their penchant for gambling and given the similarities of their views otherwise. Sambursky [707] notes that the Stoics, who were, like Laplace, rigid determinists, interpreted possibility in terms of equal ignorance. Hacking [349] recalls a remark of Boudet [103] that the "perennial question about probability is whether it is *de re* or *de dicto*" [349, p.342]. Thus we can distinguish between *de dicto* (epistemic) probability and *de re* (physical) probability.

The mapping of the distinction between possibilities to probabilities and the widespread distinction between the two types of probability is considered by Hacking in [348]. Such a carrying across to probability of the distinctions between possibilities makes sense if one can in fact *define* probabilities in terms of possibilities. However, as Hacking and others have observed, this is ultimately a circular exercise: to say that some events are equipossible is simply to say that they are equally probable. All this reinforces our view that it is pointless to talk of degrees of possibility. If probability is interpreted as relative frequencies, or in terms of propensities [654, 655,660], such a circularity does not arise. The probability so defined is definitely physical (*de re*). The role of possibility in such a context is discussed below. Hacking observes that Laplace, the champion of the epistemic view of probability, does in fact make the *de re*/*de dicto* distinction:

> When he [Laplace] needs a word to refer to an unknown physical characteristic he picks on "possibility" using it in the old *de re* sense. This was the language of his early papers. When he wants to emphasise the epistemological concept which finally captivated him, he uses "possibility" in what he makes clear is the *de dicto,* epistemological sense [349, pp.353–354].

We will now consider the place of possibility within objective theories of probability. Three authors' views on the matter are examined in some detail. We will progress from the less to the more formal and rigorous, beginning with White [862].

White's Problematic and Existential Possibilities

White [862] actually distinguishes between *existential* and *problematic* probability (p.16). He distinguishes these two concepts in terms of the possibility of "can" and that of "may:"

> Existential possibility: It is possible for X to V (can).
> Problematic possibility: It is possible that X Vs (may).

He argues that it is problematic ("may") possibility which is relevant to probability. White's views on the relationship between possibility and probability are summarised by saying:

> Probability enters at this stage which is intermediate between the exclusion of the possibility of something and the exclusion of the possibility of its opposite ... There can be degrees of probability, but not of possibility: Something can be highly probable or extremely improbable, but not highly possible or extremely impossible. One thing can be more or less probable but not more or less possible than another. The probability, but not the possibility, of something can increase or decrease. Its possibility can only appear or vanish [862, pp.59–60].

Modal Language	Exact (Probabilistic) Language.
x is possible	There is a scientific theory in which $Pr(x) \geq 0$.
x is contingent	There is a scientific theory in which $0 < Pr(x) < 1$.
x is necessary	There is a scientific theory in which $Pr(x) = 1$.
x is impossible	There is *no* scientific theory in which $Pr(x) \geq 0$.
x is almost impossible	There is a scientific theory in which $Pr(x) \approx 0$.
x is almost necessary	There is a scientific theory in which $Pr(x) \approx 1$.

Table 4.1: Bunge's probabilistic interpretation of modal terms [117].

White's interpretation of probability is objective and he rejects both the epistemic degrees of belief interpretation and the Keynesian propositional interpretation [58, 454]. White says "The relation of probability to possibility is parallel with that of confidence to belief ...". Furthermore "Just as there can be degrees of probability, but not of possibility, so one can have degrees of confidence but not of belief" [862, p.60]. The notion of possibility is not, for White, the opposite of necessity, but rather it is the opposite of certainty. White also examines the *de re*/*de dicto* distinctions and argues that *de dicto* modalities do not exist:

> In the sense discussed, there is no such thing as modality *de dicto.* As we saw in detail, a wide variety of things can be qualified by different modals; but it is a variety which can all be classed as *de re.* The danger of the thesis that modality is *de dicto* is that it tempts one, particularly with such modalities as necessity, possibility, probability and certainty, to embrace subjective theories of modality according to which modality is a characteristic of thought rather than that which can be thought about [862, p.171].

Bunge's Probabilistic Degrees of Possibility and Necessity

Bunge [117] has views quite similar to those of White. Bunge argues that Aristotle's definition of necessity in terms of possibility is not applicable to physical possibility because it ignores the component of circumstance (p.19). Bunge interprets probability in terms of propensities, but differs with Popper [655] in some respects. He says that "Probability exactifies possibility but not everything possible can be assigned a probability." Ignoring a number of points Bunge makes in his discussion of the interpretation of possibility and probability, we can present his view simply by reproducing table 1. Bunge notes with respect to this table that "Whereas in modal logic there is a gap between possibility and necessity, in a probabilistic context there is a continuum between them" [117, pp.30–31]. He also remarks on what is essentially Aristotle's statistical interpretation of modality (although he does not call it this: see [377]). This is the simple fact that "What is merely possible in the short run (for a small sample or a short run of trials) may become necessary or nearly so in the long run" [117, p.31]. Finally note that "real possibility cannot be given an

'operational definition,' say in terms of frequency" because whatever is measured is actual, and not just possible. Probability as a degree of possibility was also considered by Kattsoff [147,428], who, however, used a Keynesian [454] propositional logic framework. Hart [362] considered relative frequencies across possible worlds to give degrees of possibility. Since this is based on the Kripkean possible worlds semantics (see above) it is of little interest to us.

Morgan's Probabilistic Semantics for Modal Terms

Morgan [588] has developed a probabilistic semantics for propositional modal logics. His motivation for doing this was a dissatisfaction with the standard "possible worlds" intepretation, which, apart from being circular (interpreting "possibility" in terms of *possible* worlds), does not admit a rigorous quantitative foundation. Morgan's starting point is Popper's work on conditional probability (see appendices *ii–*iv of [658] and also [342,360,504,505]). We will not attempt to present Morgan's rather technical results in the limited space we have available here. Let us just say that his probabilistic model of standard modal logic is provably sound and complete and is a strict generalisation of *any* possible worlds model. Morgan also discusses the mechanisms of belief updating and the choice of alternative logics. Although arguing that his probabilistic semantics has considerable advantages (over the possible worlds semantics), Morgan does not present his model as the only possible solution. As well as having a greater computational complexity, the probabilistic model (as currently developed) only allows the changes in the belief of the *necessity* of a proposition to occur in "rather large jumps" [588, p.115]. In any case, Morgan has said elsewhere [589, p.53] that a single unique and correct interpretation of modality is unlikely:

> Our confusion and uncertainties with regard to many modal propositions and many arguments containing the modalities is certainly strong evidence in favour of the semantically underdetermined character of necessity and possibility.

Possibility as a Degree of Effort

An alternative objective interpretation of possibility is due to von Mises. On page 67 of [577] Von Mises says how "ordinary speech recognises different degrees of possibility." He interprets possibility in terms of the varying degrees of "effort" involved in producing a particular outcome. This seems to be the sort of thing Zadeh [900] had in mind when he spoke of the possibility of squeezing a certain number of tennis balls into a box. The larger the number of balls, the lesser is the possibility of doing it: the greater the degree of effort is required. Unfortunately there does not seem to be much we can do with such an interpretation because the "degree of effort" does not seem to be adequately formalizable. Nevertheless it does seem to be the concept of possibility that has been adopted by some proponents of fuzzy set theory:

> With a probabilistic type model we are answering a question about what percentage of the number of times we perform an experiment will a given outcome occur. Whereas with possibility we are addressing questions about how easy it is for a particular outcome to occur [892, p.247].

Yager [892] does not define or explain his notion of possibility apart from saying that "In many instances the information with respect to the possibility distribution associated with a variable can be inferred from information conveyed via natural language." In other words, the degree of possibility is to be determined in terms of membership functions of fuzzy sets.

Taking all the above into account, it still seems that von Mises' interpretation of possibility as "the degree of ease" is probability in disguise. Consider how we think it is "hard" ("requires a lot of effort") to throw 3 bullseyes in a game of darts. Considered in terms of scatter properties, this is just another way of saying that with the given experimental arrangement, the probability of achieving this situation is low. Thus the degree of ease interpretation seems quite useless for any practical purposes.

Some Fuzzy Set Theorists' Views

We now examine some fuzzy set theorists' views on the modalities of possible and probable [222,229,233,662]. Dubois and Prade [233] distinguish between physical and epistemic possibility:

> The former pertains to whether it is difficult or not to perform some action, *i.e.* questions such as "is it possible to squeeze eight tennis balls in this box?" [233, p.346]

This obviously parallels the von Misian interpretation mentioned above. Dubois and Prade are really only interested in epistemic subjective interpretations, and agree with Zadeh that "epistemic possibility can be related to imprecise verbal statements." Although they are ultimately interested in the subjective possibilities, they do argue that "natural language statements are not the only reasonable source of knowledge about the possibility of occurence of events." Thus "epistemic possibility and statistical data are not completely unrelated" (although as we shall see below, their supposed relationship is by no means clear). Dubois and Prade go on to consider this relationship in more detail. Perhaps their clearest statement of how they view the two concepts of possibility and probability is the following:

> $P(A)$ denotes the probability of A, understood as how frequent A is. $\pi(A)$ denotes the possibility of A, *i.e.* a number assessing someone's knowledge about this possibility, in rely to the question 'may A occur?'.

Later they say (p. 347)

> Possibility is thus a weak notion of evidence. In [particular] what is probable must be possible but not conversely, so that we may require grades of probability to act as lower bounds on grades of possibility.

We shall discuss Dubois and Prade's possibility and necessity measures below in section 4.5.4. This idea of possibility as an upper bound for probability has been developed further by Giles. We will now examine his arguments.

Giles' Interpretation of Possibility as an Upper Bound on Probability

Giles [304–307] has made a careful study of the relationship between possibility, probability and necessity in fuzzy set theory and has developed an interpretation along the same lines as ours in that he makes use of lower and upper probabilities.

Giles' motivation for his work is that the "ordinary" approach to fuzzy sets

> gives no indication of how one is to decide what particular numerical value to assign to a grade of membership (possibility *etc.*) in a given situation, or of how one should *use* such values in (for instance) decision making. As a result, the grounds for application of the resulting theory are, to say the least, very insecure [306, p.401].

He goes on to develop a definition of grades of membership in terms of "test-procedures." A grade of membership is defined in terms of bets on the outcome of these test-procedures.

Giles' betting interpretation, which is based on the Bayesian ideas of de Finetti, Lindley and Savage, is more general than their methods because "we are not obliged to retain the assumption that every rational agent should be a 'Bayesian agent'." Thus Giles' interpretation (partially) answers the criticisms of the betting interpretation of probability put forward by Popper [656, p.79]. Giles also admits [306, p.404] that his interpretation is not the only correct or valid one.

Giles' interpretation is of particular interest to us because of his consideration of possibility in terms of lower and upper probabilities. He introduces these ideas as follows:

> Any rational agent behaves as though he believes that each proposition A has some "true" probability, but he is not himself aware of its value, knowing only that it lies in the closed interval $[p_\ell(A), p_u(A)]$.

Perhaps the best exposition of Giles' ideas is his paper [305]. In this he says how "we interpret the assignment of a degree of possibility to A as an assignment of an upper bound on the probability of A." Most interesting for us are Giles' results characterising possibility functions. Without presenting all the definitions and introductory material necessary for a complete understanding of his results, we can give an idea of their flavour [305, p.189]:

(Theorem 3) If $\{\pi_i | \, i \in I\}$ is a set of possibility functions then π, where, for every proposition A $\pi(a) = \sup\{\pi_i \, | i \in I\}$, is also a possibility function.
(Example 3) Let P be any nonempty set of probability measures on [some Boolean algebra] $\mathcal{B}$

$$\pi_P(A) = \sup\{p(A) | \, p \in P\} \quad (A \in \mathcal{B}) \tag{4.5.17}$$

then P is a possibility function. [This example] is very important, for *every* possibility function arises in this way:
(Theorem 4) If π is any possibility function then there exists a nonempty set P of probability measures on T [a totally disconnected compact Hausdorff space such that $\mathcal{B}$ is isomorphic to the Boolean algebra of all closed and open subsets of T] such that $\pi = \pi_P$, where P is defined by (4.5.17).

He interprets theorem 4 as saying that

> every rational agent *behaves as though* he believes each proposition A has some objective probability $p(A)$ of being true; the probability assignment p, however, not being known precisely, but known only to be in a subset of P of all probability assignments. For he who has such a belief will offer to bet only if he would not expect to lose no matter where p lies in P. (p. 190)

This use of lower and upper bounds on probabilities is identical to that which we advocated in chapter 3. Note that one can adopt this interpretation without necessarily talking of a betting interpretation of probability. The probability of an event (which we assume to be an objective quantity) is unknown. We know bounds on this though, and we work with these bounds in much the same manner that we use ordinary interval arithmetic for calculating with simple quantities, when we only know lower and upper bounds on the quantities.

An important departure from Giles' interpretation occurs however when we consider the combination operations. We have already seen that in fuzzy set theory possibility measures are combined using the τ_T operations. These correspond to our *lower* dependency bounds. The upper bounds are combined and calculated with the ρ_T operations. However we use these operations without introducing the idea of "possibility." Furthermore, for the oft used special case of $T = M$, we have the fact that $\rho_T = \tau_T$ [718, theorem 7.5.6]. In this case our lower and upper dependency bounds are identical.

Giles has also briefly mentioned the more general idea of belief structures (set of all acceptable bets) of rational agents corresponding to closed convex sets of probability measures [305, p.192]. This idea has been discussed in more detail in section 4.4 where we discuss the work of Kyburg and others on lower and upper probabilities.

Some Possible Conclusions, or The Necessity of Probable Possibility

What can we conclude from this general discussion of the modal terms "possible" and "probable"? Firstly it is apparent that there is no consensus as to how these terms are related and what role the various notions should play in the examination of uncertainty. Secondly, the "degrees of possibility" that fuzzy set theorists talk about only make sense if possibility and probability are considered in an epistemic sense. We have also seen that the epistemic/objective (*de dicto/de re*) distinction applies equally well to possibility and probability. Furthermore, if an objective interpretation is accepted, then probability *is* the degree or measure of possibility. The notion of degree of possibility as somehow corresponding to a degree of effort or difficulty does not seem tenable. Nor do the "possible worlds semantics" seem very valuable in this regard. We have also seen attempts (Giles) to consider possibility and probability in terms of bounds. We have examined this idea elsewhere (sections 4.2 and 4.4) where we review work along the same lines as our interpretation of the fuzzy set theoretic operations in terms of dependency bounds.

4.5.3 The Relationship Between Probability Theory and Fuzzy Set Theory

Ultimately we may think of bridging the gap between fuzzy interval arithmetic and the calculus of random variables, i.e. embedding both into a unique setting.
— Didier Dubois and Henri Prade

Artificial Intelligence is philosophical explication turned into computer programs.
— Clark Glymour

We will now attempt to briefly review various arguments on the topic of probability theory *versus*[20] fuzzy set theory, and their general interrelationship. This is, after twenty-five years, still a contentious topic. We have certainly not aimed for completeness (that would require far more space), but we have aimed to be reasonably representative. Although our preference should quickly become apparent anyway, let us explicitly state our opinions here: *We feel that fuzzy set theory is of little value in engineering applications.* Our reasons for this conclusion are several, but the main two are the philosophical basis (which we believe to be confused and wrong), and the practical efficacy (apparently close to zero, when reasonably compared with probability theory). Since we do not have space to fully develop our arguments, we do not expect to convert many people in the following text. Nevertheless we feel the issue too important to pass over in silence.

A lot of the debate between proponents of probability and fuzzy set theory is essentially philosophical. Whilst the immediate reaction of the engineer is to avoid this (just get on with building something that works), it turns out that this is neither desirable nor possible. This is especially true given that one of the main application

[20] As many people have observed, the question is really less a matter of probability *versus* fuzzy set theory than a balanced comparison of their merits. The point is that the two approaches can be, and have been, combined in a number of different ways. Nevertheless, since it is the purpose of the present section to highlight the differences, asking the question in the above form is reasonable.

areas for the methods we have been discussing is Artificial Intelligence. We agree with Glymour who says that artificial intelligence *is* philosophy. He argues that

> Since AI is philosophy, the philosophical theory a program implements should be explicit. Any claim that a program solves some well-studied problem, ... but doesn't say how, should be disbelieved [314, p.206].

We aim to show, *inter alia,* that the philosophical basis of fuzzy set theory is inadequate for engineering and artificial intelligence problems.

One of the difficulties in critically discussing fuzzy set theory was explained by Cheeseman as follows:

> Unfortunately this [the comparison between fuzzy and non-fuzzy theories] is not as easy as it sounds because the "fuzzy approach" is itself fuzzy — there are fuzzy sets, fuzzy logic, possibility theory and various higher order generalisations of these (*e.g.* fuzzy numbers within fuzzy set theory). This diversity complicates the task of critiquing the fuzzy approach [141, p.97].

Toth [825] has recently tried to clarify some of the distinctions, and to develop a more rigorous foundation for fuzzy set-theory. The reason we mention Cheeseman's complaint is that because of the variety of different views, it is difficult to know which to criticise: any criticism can be deflected by changing ground slightly. In general we will refrain from attacking the most absurd and the weakest arguments in favour of fuzzy set theory, and we will concentrate on what appears to be the most useful material.[21]

To us, the two most convincing arguments are as follows.

1. The whole enterprise of fuzzy set theory is based on the "inherent imprecision of natural language." This is supposed to be an "uncertainty" of a completely different kind to the uncertainty of probability theory. It has its roots in *sōrites* type paradoxes [81,82,701]. Arbib [31] has presented a simple argument against this. He observes that although "people can certainly draw a 'degree of tallness' curve if pushed to it, ... this does not show that our concept of tallness has such a form" [31, p.947]. He goes on to note that vague terms are normally context sensitive (a notion fuzzy set theory either ignores, or handles in a very poor manner), and that natural language is not "inherently imprecise," although it may be used imprecisely in some circumstances:

   > Perhaps the most distressing mistake of fuzzy set theorists is to believe that a natural language like English is *imprecise.* The fact that

[21] We allow ourselves one irresistable exception, namely Goguen's argument for the "social nature of truth" [319]. Goguen, upon realising the difficulties in actually determining grades of membership or degrees of truth, suggested that the notion of objective truth was not as useful as one based on social consensus: "This paper suggests we must abandon classical presuppositions about truth, and view assertions in their social context" [319, p.65]. Whilst this may appeal to totalitarian governments, it has little to recommend it otherwise!

many people use English badly is no proof of *inherent* imprecision [31, p.948].

In any case, the point at issue, for practical purposes, is the referent of a word (what the word describes), rather than the word itself. A concentration on linguistic aspects was the cause for severe difficulties in a stream of twentieth century philosophy which followed Ludwig Wittgenstein (see [57]).

2. Our second argument is more appealing to the engineer: *Fuzzy set theory based methods do not work.* More precisely, it seems generally fair to say that fuzzy set theory has not been used to develop any methods for any problems that are *demonstrably* better than probabilistic or non-fuzzy methods. Although numerous applications have been reported, the "fuzziness" of the methods is not essential to any success they may have. Furthermore there has been very little hard-headed and honest comparison with non-fuzzy techniques. We will use the example of fuzzy control to illustrate this point below.

   Regarding the general applicability of fuzzy set theory, Zeleny [903, p.302] has argued that apart from human decision making and judgement, "there are *no other* areas of application." We agree with this, but would even question the applicability to human decision making.[22]

Fuzzy Control

An example of a suggested engineering application of fuzzy set theory is *fuzzy control.* The idea of this, which seems to have been first studied by Mamdami [538], is to develop automatic controllers for dynamic plants by using linguistic information obtained by questioning human operators of the plant. That is, one asks the operator how he controls the plant, and then incorporates these "fuzzy rules" into an automatic controller. It is suggested that this approach (which does still seem to show some promise) is suitable for highly non-linear plants which it is difficult to model explicitly. A survey of fuzzy control is given by Sugeno [788]. Fuzzy control is considered to be one of the most developed and "successful application[s] of the theory of fuzzy sets" [899, p.421]: Sugeno argues that "Fuzzy control is without doubt one of the most exciting and promising fields in fuzzy engineering" [789]. Not only are the fuzzy controllers rarely compared with the classical designs, but when they are, it is only with the simplest PID (Proportional, Integral, Derivative) controllers and little advantage (if any) is claimed [181]. The main disadvantage of fuzzy control (and this is admitted by Sugeno in his survey [788, p.78]) is that there are no analytical tools to test the stability of these controllers. Furthermore, as has

[22] Some authors consider fuzzy set theory as a purely mathematical theory and suggest that it be judged on its mathematical merits. Whilst we admit that there has been some very interesting mathematical work (especially by Höhle, Lowen and others), we generally agree with MacLane's assessment (see [527] and the papers following) that *most* of fuzzy mathematics is valueless. It is believed that there are far too many mistakes and that most of the results are trivial. (A cynic might say that the same argument applies to all modern mathematics, to which we would reply "perhaps, but it applies more so to fuzzy mathematics"). See also Johnstone's open letter to Ian Graham: "Fuzzy mathematics is NOT an excuse for fuzzy thinking" [414].

been recently shown by Buckley and Ying [116], many fuzzy controllers are more closely related to linear controllers than previously thought. As the number of rules in a fuzzy linear controller is increased, its behaviour approaches that of a simple linear controller.

We argue that it is more the *structure* of the fuzzy controllers (such as those discussed in [742,817,818]) rather than the use of fuzziness *per se* that contributes to their performance. In fact, there exists a little known probabilistic analogue to these fuzzy controllers. Black [79,80] has developed a method based on *conditional expectation arrays* which would appear to be a considerable advance over the fuzzy controllers (for the application areas envisaged). The aim of Black's work is similar: control of plants which can not be readily modelled, but which seem to be controllable by a human operator. Not only do Black's methods appear to perform better, but they have a sounder basis: Rather than implementing what the operator *says* he does, the conditional expectation controller observes what he actually does, and implements control laws based on these observations. Although some fuzzy controllers which adapt according to the operator's *actions* have been reported [813] (see also [788, p.70]), these are the exception to the rule. Czogała [180] has suggested the use of combined fuzzy and probabilistic control rules. Further work is needed in order to determine the practical value of this.

To sum up: fuzzy controllers, if one of the most successful applications of fuzzy set theory, are not a good case for the practical value of fuzzy set theory for engineering applications. Whilst the methods do seem to work, they are little or no better than alternative approaches, and perhaps more importantly, there are no analytical means to test stability and robustness. (Admittedly the second point applies equally to Black's fuzzy expectation array controllers at present.) A more detailed comparison of fuzzy set based methods and probabilistic methods for a specific engineering problem can be found in [778].

Discussions of the Relationship between Fuzzy Set Theory and Probability Theory in the Literature

Let us now note some of the previous literature which compares fuzzy set theory with probability theory, and which discusses the foundations of fuzzy set theory. There does not yet appear to exist a comprehensive (and balanced) discussion of these topics.

Several authors [195,611] feel that there is value in saying that, to an extent, probability and possibility (fuzzy set theory) convey "roughly" the same information. (This is sometimes called the "possibility-probability consistency principle."[23]) Muir [601] has shown that restricting fuzzy set theory in order to save the law of excluded middle results in a *probabilistic* Boolean algebra. In [602, p.259] he gives a more discursive account of these matters. He notes (as we have done already — see section 4.5.1) that fuzzy set theory does not provide sufficient mechanisms for dealing with

[23] "This informal principle may be translated as: the degree of possibility of an event is greater than or equal to its degree of probability, which must be itself greater than or equal to its degree of necessity" [227, p.138].

relationships (dependence).

Ralsecu [668] has argued that probability and fuzzy set theory can be *combined* by considering *fuzzy random variables* [489,490] (random variables taking on inexact values). Goodman [329] has considered interconnections between fuzzy sets and random sets. This provides a bridge between fuzzy sets and probability theory. It is along quite different lines to that which we have proceeded.

An interesting episode is the discussion of Cox's theorem, by Zadeh and others [369,409,519]. Cox's theorem [170] says that under a number of reasonable assumptions (including additivity), probability is the only measure of uncertainty. (In other words, given a few reasonable axioms for a measure of uncertainty, probability is the only one which satisfies all of them.) Lindley [519] argues along similar lines and concludes that no reasonable "scoring rule" will lead to the combination laws of fuzzy sets. Zadeh's reply [519, pp.23–24] is of considerable interest. Zadeh says that he would agree with Lindley's conclusions were it not for the necessary assumption of additivity: "There would be no issue to argue about if Professor Lindley's paper were 'Scoring rules and the inevitability of probability under additivity'." (See also page 104 of [901]). Since we (and others) have shown that the (non-additive) fuzzy set theoretic combination rules arise naturally in probability theory when one makes no independence assumptions, Zadeh's comment seems to imply a tacit agreement with our conclusion: Fuzzy set theory is effectively no different to the use of probability theory when independence is not assumed.

Some Further Discussions: Hisdal, Cheeseman and Others

Hisdal is another author who has examined the relationship between fuzzy sets and probability and has studied the interpretation of grades of membership [382–386]. She has developed an alternative foundation for fuzzy set theory which she calls the TEE model (Threshold, Error, and assumption of Equivalence) [382,383]. This was developed in response to a number of perceived inadequacies of the standard theory [386] (such as the *ad hoc* modifications often necessary to the combination rules, the plethora of combination operators that have been proposed [48], the lack of consensus on the interpretation of membership functions, and questions on the place of probability in fuzzy set theory [384,386]). We do not have space to summarise Hisdal's rather intricate model. Let it suffice to say that Hisdal seems unaware of the probabilistic Boole-Fréchet bound interpretation of the standard fuzzy set combination operators. In fact she derives (to the credit of her TEE model) combination formulae *different* to the standard min-max rules.

A number of authors *have* noticed the connexion between the Boole-Fréchet bounds and the fuzzy set operators. We have already discussed this point in section 4.2.3. Let us now just make a few additional remarks. Cheeseman [141], who has noticed the connexion, has gone on to argue that one can use probabilities to do things which fuzzy set theorists maintain can only be done with fuzzy sets. The important point, he says, is that a strict frequentist view of probability is not always necessary. By admitting subjective probabilities, many "fuzzy" problems are solvable by probabilistic means. Cheeseman argues that probability theory is

a richer and more powerful framework than fuzzy set theory because it allows the rigorous representation of dependencies. He agrees with Stallings [778] that for any given problem, probability based methods seem at least as good as, or better than, fuzzy set based methods. Zadeh's reply [901] to Cheeseman is weakly argued and essentially comprises proof by example and proof by vehement assertion (see [28]).

Grosof [340,341] and Heckerman [368] have both shown how a large number of measures of uncertainty (including fuzzy sets) which have been proposed recently can be given straight-forward probabilistic intepretations. Wise, Henrion, Ruspini, Appelbaum and others have also compared fuzzy set theory with probability. See our discussion of their work in section 4.2.3. Henrion [373] has presented a closely reasoned argument in favour of probabilistic methods for handling uncertainty in artificial intelligence. As well as pointing out that probabilities can be used in a wider range of problems than sometimes stated, he notes several advantages of probabilistic schemes over non-probabilistic ones (including fuzzy set methods). These include the ability to take account of the non-independence of different sources of evidence. Whilst there are computational problems with probabilistic methods (see the concluding chapter of this thesis), these can be overcome by Monte-Carlo methods (his "logic sampling"). This points out another disadvantage of fuzzy methods: There is no way one can *simulate* the correct result to check the validity of a proposed computation method.

Concluding Remarks

Fuzzy set theory is based on the misguided premise that "People reason with words, not numbers" [361, p.89] (they do neither: they reason with ideas). Thus the linguistic basis is the wrong starting point even for purely human problems. Even if this first point is not admitted, then fuzzy sets are still of very little use for *engineering* problems. Recall the example of fuzzy control. It seems to be much better engineering to *observe* what an operator does than to ask him to *say* what he does. In any case, there is still no convincing argument that fuzzy sets are either correct or necessary in dealing with the "inherent imprecision" of natural language. When one considers the practical efficacy of fuzzy methods the story is the same: there is no evidence for superiority over probabilistic methods, neither in the breadth of the possible domains of application nor in the performance in a given domain. Finally, we have the argument which we feel clinches the matter: the fuzzy set theory combination operations can be explained simply in probabilistic terms by considering the situation where no dependency information is available. The other supposed advantages of fuzzy set theory over probability theory are also untenable. For example, it is often said that the frequentist interpretation of probability theory severely restricts its domain of application to repeatable experiments. This is incorrect. Either the subjective or the propensity interpretations of probability can assign meaning to the probability of singular events.

We conclude by speculating that one of the reasons for the appeal of fuzzy set theory, particularly in the field of artificial intelligence, is that naive introspection can lead one to believe that fuzzy sets *do* accurately model human thinking. The

goal of AI at present seems to be to mimic or implement human thinking on a computer. We suggest, following Lem [509], in fact the wrong goal, for engineering applications. Certainly intelligent machines are desired, but that does not mean they have to have anything in common with human intelligence. Lem has noted (in a retrospective review of military history of the 21st century) that artificial intelligence "became a force to be reckoned with precisely because it did *not* become the machine embodiment of the human mind" [509, pp.37–38].

4.5.4 Confidence Curves, Possibility and Necessity Measures and Interval Fuzzy Numbers

There is a little known statistical technique known as *confidence curves* which seems to be surprisingly closely related to fuzzy numbers. We will now explore this relationship. Our aim is not to show that they are *identical* (they are not), but rather to show that the "intuitive information" they capture is nearly identical.

Confidence Curves: Their Definition and Application

Confidence curves were introduced by Cox [169] in 1958. They were developed further by Birnbaum [76,77] but have received very little attention since.[24] The basic idea of confidence curves is to have a set of nested confidence intervals for some parameter at different confidence levels. Cox's motivation for this was that the ordinary confidence intervals do not give any measure of the "informativeness" of a sample:

> [W]hen we write down the confidence interval
>
> $$(\overline{x} - 1.96\sigma/\sqrt{n},\ \overline{x} + 1.96\sigma/\sqrt{n})$$
>
> for a completely unknown normal mean, there is certainly a sense in which the unknown mean θ is likely to lie near the centre of the interval, and rather unlikely to lie near the ends and in which, in this case, even if θ does lie outside the interval, it is probably not far outside. The usual theory of confidence intervals gives no direct expression of these facts [169, p.363].

Birnbaum [76] defines confidence curves as follows:

> For each c, $0 \leq c \leq 0.5$, let $\theta_L(t,c)$ and $\theta_U(t,c)$ respectively, denote lower and upper confidence limits for an unknown parameter θ at the $1-c$ level, based on the observed value of some suitable statistic t. Such a pair of estimates also represents a $1-2c$ level confidence interval. In the (θ, c) plane, for each $c < 0.5$ plot the two points $(\theta_L(t,c), c)$ and

[24] Consultation of the *Science Citation Index* revealed the only references to Birnbaum's two papers were 1. By Birnbaum, but not concerned with confidence curves; 2. Irrelevant (said nothing of value with regard to confidence curves); or 3. The papers by Kiefer [456] and Mau [553] which we discuss below.

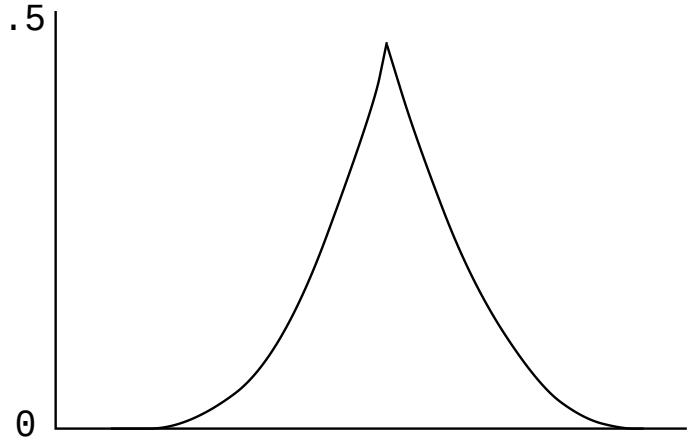

Figure 4.7: A typical confidence curve (after Birnbaum [76]).

> $(\theta_U(t,c),c)$. For $c = 0.5$, we have $\theta_L(t,c) \equiv \theta_U(t,c)$, which is a median-unbiased point estimate of θ, represented by the point $(\theta_L(t,0.5),\ 0.5)\ldots$ We denote this graph, or the function of θ which it represents, by $c(\theta,t)$. In most problems of interest such a graph is continuous and resembles that in figure 4.7 [76, p.247].

Birnbaum [77, p.118] gives the following simple example of a confidence curve. Consider an estimate of θ, the mean of a normally distributed random variable with unit variance. Then

$$c(\theta,x) = \begin{cases} \Phi(\theta - x), & -\infty \leq \theta \leq x \\ 1 - \Phi(\theta - x), & x \leq \theta \leq \infty. \end{cases} \tag{4.5.18}$$

There are many problems which arise in the use of confidence curves, particularly admissibility. These are considered by Birnbaum in some detail in [77].

Further Work on Confidence Curves for Statistical Inference

We will now examine Kiefer and Mau's comments on confidence curves. (This constitutes a comprehensive literature review of the topic!) Kiefer [456] mentioned confidence curves in his criticism of the general theory of confidence intervals. His main complaint can be best described by the following simple example (see [686]): Choose between two hypotheses H_0: $df(X) = N(0,1)$ and H_1: $df(X) = N(3,1)$ on the basis of a single observation x of X. The standard procedure would be to not reject H_0 ("accept") if $x \leq 1.5$ and accept H_1 if $x > 1.5$, the probability being 0.933 that a correct decison will follow from this procedure. Kiefer's complaint is that given two different observations $x = 1.6$ or $x = 3.6$, the same degree of conclusiveness is expressed by the test. Similar complaints have been voiced by other authors [328,554,723,724,774], and it is often said that confidence intervals give good estimates of the validity of conclusion *before* the data has been seen, but not afterwards. Kiefer's solution is to use conditional confidence intervals [110,456,457]. For the above problem, this would involve partitioning the sample space into (for example) three sets: $(-\infty,0] \cup (3,\infty)$, $(0,1] \cup (2,3]$ and $(1,2]$. The probability of the

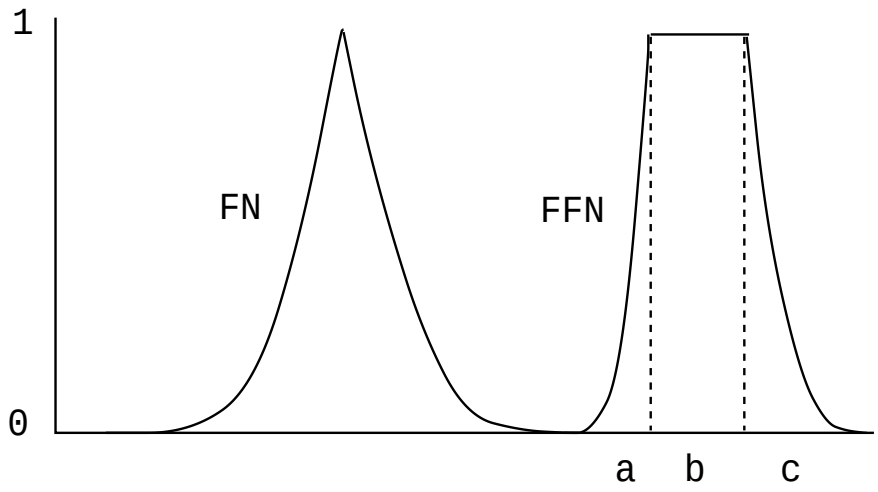

Figure 4.8: A Fuzzy Number (FN) and a Flat Fuzzy Number (FFN)

correctness of the decision to accept or reject H_0 is now *conditional* on which of these sets the sample value x falls within. The probabilities are 0.9973, 0.951 and 0.676 respectively. Kiefer says [456, p.792] that the method of nested confidence sets (confidence curves), although it "does have the frequentist interpretation of confidence," without conditioning "it suffers from the the same defect as [the example above]." This is not necessarily true (although considerable further research is required to resolve the problem completely). Birnbaum [77, p.132] has given different confidence curves for different sample values x. The intuitively "more informative sample" gives a narrower confidence curve, reflecting a more conclusive result.[25] It is surprising that these ideas have received such little attention in the literature.

Mau [553] has recently examined the use of confidence curves (he calls them confidence distributions). He carefully derives some of the properties of confidence distributions and shows how their use "quantifies the strength of evidence against a null hypothesis in the light of given data" [553, p.350] He also develops properties of central and symmetric confidence distributions. One of the most interesting points Mau makes is that confidence distribution procedures are very similar to Bayesian procedures. In fact, with one restriction, Mau's formula for *updating* a confidence distribution (in the process of accumulating data) is essentially Bayes's formula. This connexion between the two sides of statistical inference is very interesting and deserves further research.

Relationship Between Confidence Curves and Fuzzy Numbers

Our motivation for examining the relationship between confidence curves and fuzzy numbers comes from comparing figure (4.8) with figure (4.7). Apart from the factor of two difference in the vertical scaling, these two graphical representations are very similar. Furthermore the confidence curve is very directly related to the distribution function of the quantity in question. (Recall equation 4.5.18.) There are several

[25] Birnbaum's example is for a two sample test of the difference between two quantities.

Figure 4.9: Membership function constructed from histogram data

different approaches for developing a tighter analogy between the two concepts. We will examine the possibility of *conditional confidence curves* as an analogy of flat fuzzy numbers below. For now, let us consider two ideas relating fuzzy numbers and statistical data.

Dubois and Prade [233, pp.354–355] have discussed how to determine membership functions (possibility distributions) from statistical data. They begin with a nested set of intervals $\{I_i\}_{i=1}^n$ such that

$$[a,b] \subseteq I_1 \subset I_2 \subset \cdots \subset I_n = [A,B]. \tag{4.5.19}$$

They consider a series of q "imprecise" experiments for which the outcome is an interval $[a_k, b_k]$ $(k = 1, \ldots, q)$. They impose a consistency requirement that

$$\bigcap_{i=1,\ldots,q} [a_i, b_i] = [a,b] \neq \emptyset \tag{4.5.20}$$

and let $[A,B] = \bigcup_{i=1,\ldots,q}[a_i,b_i]$. Then they construct a "possibility distribution" or membership function $\mu_F(x)$ along the following lines (although their notation and presentation is rather different):

$$\mu_F(x) = \begin{cases} 0 & \text{if } x \notin I_n \\ \frac{1}{q} \sum_{j=2}^{n} |\{k|\, 1 \leq k \leq q,\ [a_k,b_k] \in I_j \setminus I_{j-1}\}| & \text{if } x \in I_n \setminus I_1 \\ 1 & \text{if } x \in I_1 \end{cases} \tag{4.5.21}$$

The result is a membership function which looks something like figure (4.9). Apart from the maximum value being 1 rather than 0.5, this is identical to the confidence

curve procedure — the vertical scale being "degree of membership" rather than "confidence." Thus at least for "statistical data" the fuzzy set method provides no more information or insight than the classical probabilistic methods. (Admittedly this is not the domain of application that proponents of fuzzy set theory have aimed for. We return to this later.)

Dubois and Prade go on to suggest the use of possibility and necessity measures based on the above acquistion procedure. McCain [556] has also considered a confidence interval interpretation of fuzzy numbers. However his hard to follow paper says very little of substance. The construction of membership functions from statistical data has also been considered by Civanlar and Trussell [148]. Their conclusions are of negligible value.

Interval Fuzzy Numbers, Possibility and Necessity Measures, Lower and Upper Probability Distributions and Higher Order Fuzzy Numbers

We will now examine Dubois and Prade's possibility and necessity measures and their relationship with lower and upper probability distributions. We will follow the presentation in [240]. More detail can be found in [231]. Further work on "interval valued fuzzy numbers" and higher order fuzzy sets (a related idea) can be found in [333–335,337,619,845].

Possibility and necessity measures are "confidence measures" that satisfy

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$$

and

$$N(A \cap B) = \min(N(A), N(B))$$

respectively for any "events" A and B. They are related by

$$\Pi(A) = 1 - N(\overline{A}).$$

Given a "possibility distribution" $\pi(\omega)$ (corresponding to a membership function $\mu(\omega)$) this means that

$$\Pi(A) = \sup\{\pi(\omega) | \, \omega \in a\} \tag{4.5.22}$$

and

$$N(A) = \inf\{1 - \pi(\omega) | \, \omega \notin A\}. \tag{4.5.23}$$

The relationship between possibility distributions and possibility measures is considered to be analogous to the relationship between probability densities and probability distributions. Possibility measures have the property that $\Pi(\bigcup_{i \in I} A_i) = \sup_{i \in I} \Pi(A_i)$. Dubois and Prade show that possibility and necessity measures are related to probability by

$$N(A) \leq P(A) \leq \Pi(A).$$

They present the theory of fuzzy numbers in terms of possibility theory and show how a fuzzy number (or, as they prefer to say, a fuzzy interval) describes lower and

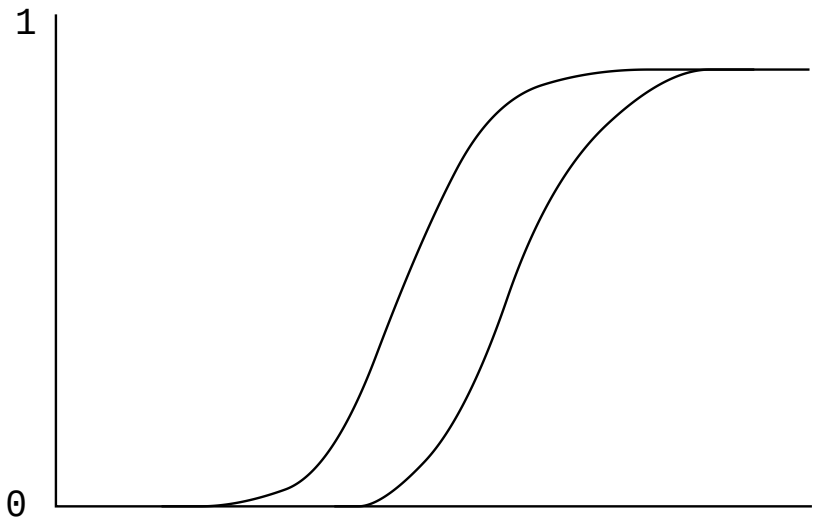

Figure 4.10: Lower and upper probability distributions for which there is no corresponding normal fuzzy number

upper probability distributions as follows:

$$\begin{aligned} F^*(u) &= \overline{P}((-\infty,u]) = \Pi((-\infty,u]) = \sup\{\mu_Q(r)|\, r \leq u\} \\ &= \begin{cases} \mu_Q(u) & u \leq q_1, \\ 1 & u > q_1, \end{cases} \end{aligned}$$

and

$$\begin{aligned} F_*(u) &= \underline{P}((-\infty,u]) = N((-\infty,u]) = \inf\{1-\mu_Q(r)|\, r > u\} \\ &= \begin{cases} 0 & u < q_2, \\ 1-\mu_Q(u+) & u \geq q_2. \end{cases} \end{aligned}$$

The fuzzy quantity Q has a membership function with a "flat" in the interval $[q_1, q_2]$ (corresponding to the region b in figure (4.8). When there is no flat ($q_1 = q_2$), there is only one value of x ($x = q_1$) such that $F_*(x) = 0$ and $F^*(x) = 1$. Since the lower and probability distributions are *derived* from normal fuzzy numbers, the situation depicted in figure (4.10) can never occur. Admittedly, by sacrificing normality, one could stretch the analogy to cover this case. We feel that taking the lower and upper probability distributions as primary (as *bounds* on some inaccurately known probability distribution) is preferable to the above approach.

Finally we note in passing that Dubois and Prade have used the connexion between possibility and necessity measures and lower and upper probability distributions in order to define and study the properties of the "mean value" of a fuzzy number [236]. They define expectations of fuzzy numbers in terms of the expectations of random variables which have the lower and upper probability distributions. They also note some connexions with the theory of random sets [551,614]. Dubois and Prade's [236, p.298] suggestions for further research include a study of the determination of $\mathcal{P}(M \oplus N)$ in terms of $\mathcal{P}(M)$ and $\mathcal{P}(N)$ where

$$\mathcal{P}(Q) = \{P|\, \forall A \text{ measurable}, \Pi(A) \geq P(A) \geq N(A)\}$$

for some fuzzy number Q, where $\oplus$ denotes the sup-min convolution and $\Pi(A)$ and $N(A)$ are defined above. Our dependency bounds in chapter 3 essentially answer this question for a more general class of probability measures. Dubois and Prade [236, p.290] also discuss the relationship between their notion of a fuzzy number and Höhle's (and others — see the beginning of section 4.5 and section 4.6). Dubois and Prade's conclusion is of some interest to us. They say that "Ultimately we may think of bridging the gap between fuzzy interval arithmetic and the calculus of random variables, *i.e.* embedding both into a unique setting" [236, p.298]. This has been our aim in the present section. We conclude by saying that there is a reasonably close correspondence between membership functions and confidence curves. This parallels the correspondence between possibility/necessity measures and lower/upper distribution functions.

4.6 General Interpretations — Positivist vs Realist

All the things in fact, that we approached by our senses reason or intellect are so different from one another that there is no precise equality between them.
— Nicolas Cusanus (1401–1464)

Elaboration of this idea leads to the concept of a space in which a distribution function rather than a definite number is associated with every pair of elements.
— Karl Menger

The object of this section is to compare two points of view for using distribution functions as a generalisation of numbers. One approach is that which we have adopted. This entails taking the view of random variables of orthodox (Kolmogorov) probability theory. The random variables are the primary entities and they are manipulated in terms of their distribution functions. The other viewpoint is that taken in the development of probabilistic metric spaces. In this case the distribution functions are considered primary and operations are performed on distribution functions without reference to any underlying random variables.

We shall see that one of the advantages of the latter viewpoint no longer holds. It was originally adopted for two reasons. The first was the severe mathematical difficulties arising in the random variable viewpoint because of intricate dependency relations. The second was that there were many functions of distribution functions which were not expressible in terms of random variables.[26] The interpretation of the $\tau_{T,L}$ and $\rho_{T,L}$ operations as dependency bounds changes this, because now these operations *can* be interpreted in terms of random variables.

4.6.1 Probabilistic Metric Spaces

We now present a very brief outline of the theory of probabilistic metric spaces. More details can be found in the excellent book by Schweizer and Sklar [718]. See

[26] Historically this statement is not true — the reasons why the development of probabilistic metric spaces went the way it did are more complex than the few sentences above would indicate. However, the main point (that the random variable viewpoint did not allow all operations) is true.

also the review of the book by Brooke [108] who mentions some of the possible physical applications of the theory. Rosen [688,689] has also considered the idea and suggested uses in physics.

The original idea of probabilistic metric spaces was proposed by Menger [562, 564] who was motivated by Poincaré's paradox [651,652] (see also [130, p.543]). Poincaré noted that for physical measurements, given that $A = B$ and $B = C$, one could not necessarily conclude that $A = C$ when the equality relation is interpreted physically. Menger suggested that a distribution function be associated with each pair of quantities, thus providing a generalisation of the notion of a metric space (first introduced by Fréchet in 1906):

> The distribution function associated with two elements of a statistical metric space might be said to give, for every x, the probability that the distance between the two points in question does not exceed x [564].

Menger subsequently developed his ideas further in [563] where he suggested that the theory could be applied to psychophysics and physics (both microscopic and macroscopic).

Menger's original proposal was for *triangle functions* T such that, *inter alia,* the triangle inequality

$$F_{pr}(x + y) \geq T(F_{pq}(x), F_{qr}(y)) \tag{4.6.1}$$

would hold (F_{pr} is the distance distribution function for the two points p and r). Wald [848] suggested that (4.6.1) could be replaced by

$$F_{pr}(x) \geq (F_{pq} * F_{qr})(x) \tag{4.6.2}$$

where $*$ denotes convolution. This has the simple probabilistic interpretation that the probability that the distance from p to r is less than x is at least as large as the probability that the sum of the distances from p to q and from q to r, regarded as independent, is less than x. These Wald spaces have not generated much interest or many results because of the severe mathematical difficulties caused by the complex dependency structure induced on the underlying random variables.

The majority of the work on probabilistic metric spaces has taken the original approach of Menger, and studied a variety of different triangle functions. One of these is the $\tau_{T,L}$ operation given by

$$\tau_{T,L}(F, G)(x) = \sup_{L(u,v)=x} T(F(u), G(v)) \tag{4.6.3}$$

for some t-norm T. An important result due to Schweizer and Sklar [717] is that apart from $L = \max$ and $T = \min$ the function $\tau_{T,L}$ is not derivable from a function on random variables. (See also Schweizer's recent remark in [716].) This means that for any F and G in Δ^+ there do not exist random variables X and Y where $F = df(X)$ and $G = df(Y)$, and a Borel measurable function $V: \Re^+ \times \Re^+ \mapsto \Re^+$ such that $df(V(X, Y)) = \tau_{T,L}(F, G)$. This implies that the viewpoint taking distribution functions as primary is in a sense more powerful. However, as we have seen (and in

fact as Schweizer *et al.* were the first to show), the $\tau_{T,L}$ operations do arise naturally as dependency bounds for functions of random variables. The impact of this on the development of probabilistic metric spaces remains to be seen. It is certainly a topic which deserves further investigation.

4.6.2 Philosophical Aspects of Probabilistic Metric Spaces

Menger developed a number of philosophical ideas from his work on probabilistic metric spaces.[27] His general view of the theory was that it was a "positivist geometry" [570,571]. Positivism, or rather logical positivism, is a philosophical doctrine which can trace its roots to Hume's empiricism. It holds that no propositions should be considered to be true unless *verified* by direct sense experience. It has been shown by a number of authors, most notably Karl Popper [658], that the whole scheme is fundamentally flawed. It is fair to say that positivism is dead. Menger's use of the term was to indicate that the objects of geometry which we perceive and can "verify" are not the ideal points or lines of Platonic geometry, but rather the blobs and fuzzy regions of a probabilistic geometry. This seems to be the appropriate viewpoint for interpreting current work on probabilisitic metric spaces.[28]

The viewpoint we adopt is that random variables are randomly varying quantities. We would prefer the quantities to be not varying, in which case we could readily calculate the functions of interest. Instead, the best we can do is to try to determine the *distribution* of the functions of interest, given the distribution of the random variables. This viewpoint seems similar to that necessary for interpreting a *random* metric space (see chapter 9 of [718]). Random metric spaces, first introduced by Špaček [771,772], have been shown by Schweizer and Sklar [718] to in fact be "a proper part of the theory of probabilistic metric spaces." Calabrese's investigations in this area seem potentially useful [125]. He shows that when the standard (distribution based) approach is examined in terms of random variabes, quite peculiar effects can arise. We feel that there is considerable scope for further research in this area, especially if the dependency bound viewpoint is adopted. We hope to pursue some of these issues ourselves at a later date.

4.7 General Conclusions

We have seen a surprisingly large number of connexions between the material we developed in chapter 3 and other ideas. Some of these connnexions have suggested areas for future research, and others have shown that some results that have been

[27] Amongst those which we do not discuss here, perhaps the most interesting and promising is his very careful analysis of the idea of a random variable from the point of view of a general theory of variables or fluents [567]. He developed new improved notation for the concepts of variables and functions [565,568, 569,572] (see also [561]), which he used in a calculus textbook [566]. This aspect of his work has received little attention in the literature since.

[28] It is no accident that the standard probabilistic metric space structure has been adopted by some fuzzy set theorists — the goals and techniques are the same: both sidestep the issue of dependence. See for example the works by Höhle, Klement, Lowen and others [135,389,391–393,465,527].

presented in the literature duplicate earlier work. The most useful connexions and directions for future research are as follows:

- The extensions by Hailperin of the Boole-Fréchet bounds using the techniques of linear programming suggest the similar application of general mathematical programming ideas to dependency bounds for functions of random variables. This would generalise the bounds to functions of more than two variables.
- Graph-theoretic techniques may be of use in probabilistic arithmetic. There are two issues: the control of calculations (using the structure of the graph) and determination and approximation of complex stochastic dependencies using graph theoretic methods; and the possibility of transforming expression DAGs in order to make determination of distribution functions easier. As we saw however, the prospects for the latter are not very good.
- Our viewpoint taken in interpreting our bounds on probability distribution functions is quite different to that taken in most contemporary theories of lower/upper probabilities. However it seems that it might be possible to integrate our approach with other theories. In particular, Fine's theories of lower probabilities deserve more attention.
- The theory and practice of fuzzy arithmetic and fuzzy numbers has a lot in common with probabilistic arithmetic which makes use of dependency bounds. The duality theory we used to develop our numerical methods encompasses the results used in calculating operations on fuzzy numbers and shows how extensions are possible. The intuitive ideas captured by fuzzy numbers can be equally well represented using probabilistic techniques such as confidence curves.
- Our viewpoint is somewhat different to that adopted in the theory of probabilistic metric spaces, although there is some scope for integrating the two approaches.

To summarise the chapter as briefly as possible: *There are numerous connexions with other results: some apparently of no use at all; some of 'interest'; some immediately useful in other areas; and some potentially useful and suitable subjects for further research.*

Chapter 5

An Extreme Limit Theorem for Dependency Bounds of Normalised Sums of Random Variables

This becomes clearer when you restrict your considerations to the maximum and minimum of quantity.
— Nicolas Cusanaus

This chapter presents a new result on the behaviour of dependency bounds of iterated normalised sums. We show that the dependency bounds converge to step functions as the number of summands increases. The step functions are positioned at points which depend only on the extremes of the supports of the summand's distribution functions. With very minor differences, this chapter will appear in *Information Sciences* (with the same title as this chapter). In order to make this chapter self-contained there is some reiteration of material covered in chapter 3. Since most of this material is not widely known, this does no harm.

5.1 Introduction

Dependency bounds are lower and upper bounds on the distribution of functions of random variables when all that is known about the random variables is their marginal distributions. They have been recently studied by Frank, Nelsen and Schweizer [277] who showed that Makarov's solution [536] to a question originally posed by Kolmogorov follows naturally from the theory of copulas [718,759]. Dependency bounds arise in the development of probabilistic arithmetic (see chapter 3), and it is in this context which we study the question of limit results for dependency bounds of normalised sums.

If we write $\mathrm{ldb}(S_N)$ and $\mathrm{udb}(S_N)$ for the lower and upper dependency bounds of $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$, and $\{X_i\}$ is a sequence of random variables with distribution functions F_i, then we show that as N approaches infinity, $\mathrm{ldb}(S_N)$ and $\mathrm{udb}(S_N)$

approach unit step functions. The position of the step functions depends only on the support of the F_i and this is why we refer to our result as an *extreme* limit theorem. (Note that our result is actually more analogous to the law of large numbers than the central limit theorem. The central limit theorem describes the behaviour of $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$.)

We prove our result in a fairly straightforward manner by making use of the properties of T-conjugate transforms. These transforms, which play a role analogous to the Laplace-Stieltjes transform in the central limit theorem for sums of random variables [520], have been developed by Richard Moynihan [595,596]. Apart from being mathematically interesting, our limit theorem has a practical interpretation with regard to the suitability of interval arithmetic for certain problems. We show that if one has to calculate the dependency bounds for the sum of a large number of random variables, then one can use interval arithmetic from the outset (using just the endpoints of the supports of the distributions) because one will lose little information in doing so.

The reason why one would want to calculate dependency bounds rather than ordinary convolutions can be explained by considering probabilistic arithmetic. The aim of this is to replace the ordinary arithmetic operations on numbers by appropriate convolutions of probability distribution functions of random variables. However, when one does this, the phenomenon of dependency error occurs. This is caused by stochastic dependencies arising between intermediate results of a calculation. In order to avoid errors in the calculations, one has to take cognizance of these dependencies when they exist. If, as seems to be the case practically with probabilistic arithmetic, one can only determine whether or not two quantities are independent (and not any measure of dependence if they are not), then one has to assume the worst and calculate dependency bounds.

The rest of this chapter is organised as follows. Section 5.2 contains all the preliminary information we require in order to prove our main result in section 5.3. We formally define the notions of dependency bounds, copulas, t-norms, the τ_T and ρ_T operations, and T-conjugate transforms as well as discussing the representation of Archimedean t-norms and the characterisation of associative copulas. Section 5.3 is devoted to the proof of our main result (theorem 5.3.1) and section 5.4 gives an explicit formula for $\mathrm{ldb}(S_N)$ for a special case. This enables us to examine the rate of convergence for theorem 5.3.1. In section 5.5 we present some examples illustrating the results of sections 5.3 and 5.4. These examples are calculated using the numerical representations and algorithms developed in chapter 3. Finally, section 5.6 draws some general conclusions from the results of this chapter.

5.2 Definitions and Other Preliminaries

We now introduce the notation and results needed to prove our result in section 5.3 in sufficient detail to make this chapter self contained. The general references for this section are [595,596,718]. We will often write inequalities between two functions F and G as $F < G$. This is to be interpreted as meaning $F(x) < G(x)$ for all x in

the common domains of F and G. A *convex* function f defined on some set A is one which satisfies

$$f(\alpha x_1 + (1-\alpha)x_2) \;\leq\; \alpha f(x_1) + (1-\alpha)f(x_2)$$

for all $x_1, x_2 \in A$ and all $\alpha \in (0,1)$. If the above inequality is reversed then f is *concave*. We write "iff" for "if and only if".

5.2.1 Distribution Functions

The *distribution function* of a random variable X is denoted $df(X)$ and is given by

$$df(X) = F(x) = P\{X < x\}$$

and is a left continuous function from $\Re$ onto I ($I = [0,1]$). The set of all distribution functions is denoted Δ. The subset Δ^+, defined by

$$\Delta^+ = \{F \in \Delta |\, F(0) = 0\},$$

is the set of distribution functions of random variables that are almost surely positive. The *support* of a distribution function is defined by

$$\operatorname{supp} F = [\ell_F, u_F]$$

where

$$\ell_F = \inf\{x |\, F(x) > 0\} \tag{5.2.1}$$

and

$$u_F = \sup\{x |\, F(x) < 1\}. \tag{5.2.2}$$

Three subsets of Δ are defined by

$$\begin{aligned} \Delta_L &= \{F \in \Delta |\, \ell_F > -\infty\}, \\ \Delta_U &= \{F \in \Delta |\, u_F < \infty\}, \end{aligned}$$

and

$$\Delta_{LU} = \Delta_L \cap \Delta_U.$$

We also have

$$\begin{aligned} \Delta_U^+ &= \{F \in \Delta^+ |\, u_F < \infty\}, \\ \Delta_L^+ &= \{F \in \Delta^+ |\, \ell_F > 0\}, \end{aligned}$$

and

$$\Delta_{LU}^+ \;=\; \Delta_L^+ \cap \Delta_U^+.$$

The step function $\epsilon_a \in \Delta$ is defined by

$$\epsilon_a(x) = \begin{cases} 0 & x \leq a, \\ 1 & x > a. \end{cases}$$

5.2.2 Triangular Norms

A *t-norm* (triangular norm) T is a two place function $T: I \times I \mapsto I$ which is symmetric, associative, non-decreasing in each place and has 1 as a unit (*i.e.* $T(a,a) = T(1,a) = a \; \forall a \in I$). An *Archimedean* t-norm is one which satisfies $T(a,a) < a \;\; \forall a \in (0,1)$. A *strict* t-norm is one which is continuous on I^2 and is strictly increasing in each place on $(0,1]^2$. All t-norms T satisfy $Z \leq T \leq M$, where

$$Z(x,y) = \begin{cases} x & x \in I, \; y = 1, \\ y & x = 1, \; y \in I, \\ 0 & x \in [0,1), \; y \in [0,1), \end{cases}$$

and

$$M(x,y) = \min(x,y) \quad x,y \in I.$$

Two other t-norms we will use are W and Π given by

$$W(x,y) = \max(x+y-1,0)$$

and

$$\Pi(x,y) = xy.$$

These four t-norms have the following properties:

Z is Archimedean, but not continuous.

W is Archimedean and continuous, but not strict.

M is continuous, but neither Archimedean nor strict.

Π is continuous, strict and Archimedean.

Any strict t-norm is Archimedean. We define $\mathcal{T}$ by

$$\mathcal{T} = \{T \,|\, T \text{ is a continuous t-norm}\},$$

and $\mathcal{T}_A$ by

$$\mathcal{T}_A = \{T \in \mathcal{T} \,|\, T \text{ is Archimedean}\}.$$

If T is a t-norm, then T^* defined by

$$T^*(x,y) = 1 - T(1-x, 1-y) \tag{5.2.3}$$

is known as a *t-conorm* (see section 5.7 of [718]).

5.2.3 Representation of Archimedean t-norms

Archimedean t-norms are of special interest because of the following representation theorems [718].

Theorem 5.2.1 *A t-norm T is continuous and Archimedean iff*

$$T(x,y) \;=\; f_T(g_T(x)+g_T(y)),$$

where

1. *g_T is a continuous strictly decreasing function from I into $\Re^+$ with $g_T(1)=0$.*
2. *f_T is a continuous function from $\Re^+$ onto I such that it is strictly decreasing on $[0,g_T(0)]$ and such that $f_T(x)=0\ \forall x \geq g_T(0)$.*
3. *f_T is a quasi-inverse of g_T (see section 3.2 of this thesis).*

The functions f_T and g_T are known as the *outer* and *inner additive generators* of T and are unique up to a multiplicative constant. If we set $h_T(x)=f_T(-\log x)$ and $k_T(x)=\exp(-g_T(x))$ we obtain a multiplicative analogue:

Theorem 5.2.2 *A t-norm T is continuous and Archimedean iff*

$$T(x,y)=h_T(k_T(x)k_T(y)),$$

where

1. *k_T is a continuous strictly increasing function from I into I with $k_T(1)=1$.*
2. *h_T is a continuous function from I onto I that is strictly increasing on $[k_T(0),1]$ and such that $h_T(x)=0\ \forall x\in[0,k_T(0)]$.*
3. *h_T is a quasi-inverse of k_T.*

The functions h_T and k_T are known as the *outer* and *inner multiplicative generators* of T and are unique up to an exponentiation. T is strict iff $h_T=k_T^{-1}$ and $k_T(0)=0$. For any $T\in\mathcal{T}_A$, h_T and k_T satisfy

$$\begin{aligned} h_Tk_T(x) &= x & \forall x\in I\\ k_Th_T(x) &= \max(k_T(0),x) & \forall x\in I. \end{aligned}$$

If $T=W$ we have $h_W(x)=\max(1+\log x,0)$ and $k_W(x)=e^{x-1}$.

5.2.4 Copulas

A *2-copula* (or just *copula*) is a mapping $C\colon I^2\mapsto I$ such that 0 is the null element, 1 the unit and C is 2-increasing:

$$C(a_2,b_2)-C(a_1,b_2)-C(a_2,b_1)+C(a_1,b_1)\geq 0$$

for all $a_1,a_2,b_1,b_2\in I$ such that $a_1\leq a_2$ and $b_1\leq b_2$. The set of all copulas is denoted $\mathcal{C}$. All copulas satisfy $W\leq C\leq M$ and are continuous.

Copulas link joint distribution functions with their marginals. If $H\colon\Re^2\mapsto I$ is a two dimensional joint distribution function corresponding to the random variables

X and Y (*i.e.* $H(x,y) = P\{X < x, Y < y\}$), and $F = df(X)$ and $G = df(Y)$ are the marginals, then

$$H(x,y) = C(F(x), G(y))$$

for some copula C, known as the *connecting copula* for X and Y. The copula contains the dependency information relating the random variables X and Y. If $C = \Pi$, then X and Y are independent.

The *dual* of a copula, written C^d, is defined by

$$C^d(x,y) = x + y - C(x,y). \tag{5.2.4}$$

For any $C \in \mathcal{C}$, the operations τ_C and ρ_C from $\Delta \times \Delta$ onto Δ are defined pointwise by

$$\tau_C(F,G)(x) = \sup_{u+v=x} C(F(u), G(v))$$

and

$$\rho_C(F,G)(x) = \inf_{u+v=x} C^d(F(u), G(v)).$$

These operations have been studied because of their properties as triangle functions when $C \in \mathcal{T}$ in the theory of probabilistic metric spaces [718]. If $C \in \mathcal{T}$, then τ_C and ρ_C are associative. They are also non-decreasing in each place (lemma 7.2.2 of [718]).

The only additional requirement on $C \in \mathcal{C}$ for C to also be in $\mathcal{T}$ is associativity. In fact, under one weak condition, C must be Archimedean:

Theorem 5.2.3 *Let $T \in \mathcal{T} \cap \mathcal{C}$ be such that $T(x,y) \neq \min(x,y)\ \forall (x,y) \in (0,1)^2$. Then $T \in \mathcal{T}_A$.*

PROOF. This follows at once from theorem 5.3.8 of [718] and the fact that $C \leq M$ for all $C \in \mathcal{C}$. ■

We define the set of T that satisfies the condition of theorem 5.2.3 as $\mathcal{T}_C \subset \mathcal{T} \cap \mathcal{C}$.

Since we will restrict ourselves to associative copulas in this chapter, it is of interest to have a probabilistic characterisation of them. This is part of problem 6.7.2 of [718]. Archimedean copulas have been studied by Genest and MacKay [299] who have looked at stochastic orderings and sequences of copulas, and have used their results to develop new parameterised families of copulas satisfying certain conditions. Schweizer and Sklar [718] present the following two theorems (their theorems 6.3.2 and 6.3.3):

Theorem 5.2.4 *A t-norm T is a copula iff*

$$T(c,b) - T(a,b) \leq c - a \qquad \forall a,b,c \in I, \text{ with } a \leq c.$$

Theorem 5.2.5 *Let $T \in \mathcal{T}_A$, then $T \in \mathcal{C}$ iff either f_T or g_T (the additive generators of T) is convex.*

Restricting C to be associative restricts the type of dependency structure two random variables X and Y can have. Noting that if C is a copula, it can be considered as the joint distribution function of X and Y, where X and Y have uniform marginal distributions on $[0,1]$, we can proceed as follows. If C is Archimedean, then

$$C(x,y) = h_C(\Pi(k_C(x), k_C(y))),$$

where Π is the joint distribution function of two independent random variables with uniform marginals on I. Therefore

$$\begin{aligned} \Pi(k_C(x), k_C(y)) &= P\{X < k_C(x),\, Y < k_C(y)\} \\ &= P\{k_C^{-1}(X) < x,\, k_C^{-1}(Y) < y\} \\ &= F_{UV}(x,y) \end{aligned}$$

where F_{UV} is the joint distribution function of U and V, $U = h_C(X)$, and $V = h_C(Y)$. We thus have

$$C(x,y) = h_C(F_{UV}(x,y)).$$

Noting that U and V are independent (because X and Y are), we have shown that if C is associative then it is an increasing function of a joint distribution of independent random variables. This demonstrates the restriction on the types of dependencies an associative copula can represent, but is not a wholly satisfying characterisation because there seems to be no natural probabilistic interpretation of such a dependence structure.

5.2.5 Dependency Bounds

We now formally define dependency bounds and show how they are related to the τ_C and ρ_C operations. We firstly consider sums of two random variables and then examine when we can extend the results to sums of N random variables.

Let X and Y be two random variables with distribution functions $F = df(X)$ and $G = df(Y)$. Let their joint distribution function be given by $H(x,y) = C(F(x), G(y))$ where C is the connecting copula for X and Y. Then the lower and upper dependency bounds $\mathrm{ldb}_{\underline{C}}$ and $\mathrm{udb}_{\underline{C}}$ on $df(X+Y)$ are such that

$$\mathrm{ldb}_{\underline{C}}(F,G) \leq df(X+Y) \leq \mathrm{udb}_{\underline{C}}(F,G)$$

for all $C \geq \underline{C} \leq W$. The function $\underline{C}$ is the lower bound on the connecting copula C. The crucial result we need is given by

Theorem 5.2.6

$$\mathrm{ldb}_{\underline{C}}(F,G) = \tau_{\underline{C}}(F,G)$$

and

$$\mathrm{udb}_{\underline{C}}(F,G) = \rho_{\underline{C}}(F,G),$$

and these bounds are the pointwise best possible.

PROOF. See chapter 3 and [277]. ■

An analogue of this theorem holds for operations other than addition on random variables. (See chapter 3.) We will be interested in the case that $\underline{C} \in \mathcal{T}_C$ in which case we will write T instead of $\underline{C}$.

These dependency bounds can be extended to sums of the form

$$S_N = \sum_{i=1}^{N} X_i,$$

where $\{X_i\}_{i=1}^N$ is a set of random variables with distribution functions $F_i = df(X_i)$ and such that for all $i \neq j$ $(i, j = 1, \ldots, N)$,

$$C_{X_i X_j} \geq T, \tag{5.2.5}$$

where $C_{X_i X_j}$ is the connecting copula for the pair of random variables (X_i, X_j). In such a case, because addition is associative and commutative, and because any $T \in \mathcal{T}_C$ is also, we find that τ_T and ρ_T are as well and we can write [277, p.211],

$$\begin{aligned} \mathrm{ldb}_T(F_1, \ldots, F_N) &= \tau_T^{(N)}(F_1, \ldots, F_N) \leq df(S_N) \\ &\leq \rho_T^{(N)}(F_1, \ldots, F_N) \\ &= \mathrm{udb}_T(F_1, \ldots, F_N). \end{aligned}$$

The N-place functions $\tau_T^{(N)}$ and $\rho_T^{(N)}$ are given by

$$\begin{aligned} &\tau_T^{(2)}(F, G) = \tau_T(F, G) \\ &\tau_T^{(N)}(F_1, \ldots, F_N) = \tau_T(F_1, \tau_T^{(N-1)}(F_2, \ldots, F_N)) = \tau_T(\tau_T^{(N-1)}(F_1, \ldots, F_{N-1}), F_N) \end{aligned} \tag{5.2.6}$$

and

$$\begin{aligned} &\rho_T^{(2)}(F, G) = \rho_T(F, G) \\ &\rho_T^{(N)}(F_1, \ldots, F_N) = \rho_T(F_1, \rho_T^{(N-1)}(F_2, \ldots, F_N)) = \rho_T(\rho_T^{(N-1)}(F_1, \ldots, F_{N-1}), F_N) \end{aligned} \tag{5.2.7}$$

We will generally drop the (N) superscript as no confusion can arise when the arguments are explicitly stated.

The condition (5.2.5) is not equivalent to

$$C_{X_1 \ldots X_N} \geq T^{(N)},$$

where $C_{X_1 \ldots X_N}$ is an N-copula and $T^{(N)}$ is an N-iterate of a t-norm:

$$T^{(N)}(x_1, \ldots, x_N) = T(T^{(N-1)}(x_1, \ldots, x_{N-1}), x_N).$$

The problem of relating high order copulas to their lower order marginals is still open. See the problems at the end of chapter 6 of [718]. It is known that all N-copulas C satisfy

$$W^{N-1} \leq C \leq M^{N-1}, \tag{5.2.8}$$

that Π^{N-1} and M^{N-1} are copulas, but that W^{N-1} is not [718, p.88]. The fact that W^N is not a copula for $N > 2$ can be understood simply as follows. For $N = 2$, if $C = W$ is a connecting copula for X and Y, then X and Y are decreasing functions of each other. However it is impossible, given three or more variables $X_1, \ldots, X_N$, for each X_i to be a decreasing function of all of the others. Nevertheless, the lower bound in (5.2.8) cannot be improved.

5.2.6 T-conjugate Transforms

Before introducing T-conjugate transforms we note how T can be represented in terms of τ_Π, h_T and k_T.

Theorem 5.2.7 ([594]) *Let $T \in \mathcal{T}_A$ and $F \in \Delta^+$. Define $k_T F \in \Delta^+$ by*

$$k_T F(x) = \begin{cases} 0 & x \leq 0, \\ k_T(F(x)) & 0 < x. \end{cases} \tag{5.2.9}$$

Then for all F and G in Δ^+, both $\tau_T(F, G) \in \Delta^+$ and $\tau_\Pi(k_T F, k_T G) \in \Delta^+$ and in fact

$$\tau_T(F, G) = h_T(\tau_\Pi(k_T F, k_T G)).$$

This follows directly from theorem 5.2.2.

Definition 5.2.8 *Let $T \in \mathcal{T}_A$ and $F \in \Delta^+$. Then the* T-conjugate transform *of F, denoted $C_T F$, is defined by*

$$C_T F(z) = \sup_{x \geq 0} e^{-xz} k_T F(x) \qquad \forall z \in \Re^+, \tag{5.2.10}$$

where $k_T F$ is given by (5.2.9).

The study of T-conjugate transforms is due to Moynihan [595,596] who extended the Prod-conjugate transform $C_\Pi F$ by using theorem 5.2.7. Note that $C_T F = C_\Pi(k_T F)$. The C_T transform has a number of properties in relation to τ_T analogous to the Laplace transform's properties with respect to ordinary convolutions. The essential one is

Theorem 5.2.9 (Theorem 3.1 of [596]) *Let $T \in \mathcal{T}_A$ and let $F, G \in \Delta^+$. Then*

$$C_T(\tau_T(F, G))(z) = \max\{k_T(0),\ C_T F(z) \times C_T G(z)\} \qquad \forall z \geq 0. \tag{5.2.11}$$

If T is strict, (5.2.11) becomes

$$C_T(\tau_T(F, G)) = C_T F \times C_T G. \tag{5.2.12}$$

We also have [595, theorem 1.1]:

Theorem 5.2.10 *Let*

$$\mathcal{A} = \{\phi: \Re^+ \mapsto (0,1] | \quad \phi \text{ is non-increasing, positive, continuous and log-convex}\} \cup \{\Theta_\infty\},$$

where $\Theta_\infty(z) = 0 \ \ \forall z \geq 0$. *Then for all* $T \in \mathcal{T}_A$, *if*

$$\mathcal{A}_T = \{\phi \in \mathcal{A} | \, \phi(z) \geq k_T(0) \ \forall z \geq 0\},$$

then $\mathcal{A}_T = \{C_T F | \, F \in \Delta^+\}$. *Thus* $\mathcal{A}_T \subseteq \mathcal{A}$ *and equality holds iff* T *is strict.*

The *inverse T-conjugate transform* C_T^* is defined by

Definition 5.2.11 *Let* $T \in \mathcal{T}_A$ *and let* $\phi \in \mathcal{A}_T$. *Then*

$$C_T^* \phi(x) = h_T(\inf_{z \geq 0} e^{xz} \phi(z)) \qquad \forall x \in \Re^+ \tag{5.2.13}$$

and $C_T^* \phi$ *is normalised to be left-continuous.*

If $F \in \Delta^+$, then we say F is *log-concave* if $\log F$ is concave on (ℓ_F, ∞), and that $F \in \Delta^+$ is *T-log-concave* if $k_T F$ is log-concave. We define

$$\Delta_T^+ = \{F \in \Delta^+ | \, F \text{ is } T\text{-log-concave}\}.$$

The *log-concave-envelope,* denoted $\overline{F}$, of $F \in \Delta^+$ is defined by

$$\begin{aligned} \overline{F}(x) &= 0 && x \leq \ell_F \\ \log \overline{F}(x) &= \sup\{p \log F(x_1) + q \log F(x_2) | \, p, q \geq 0, \\ &\qquad p + q = 1, \ x_1, x_2 > \ell_F \text{ and } x = px_1 + qx_2\} && x > \ell_F. \end{aligned}$$

In other words, the graph of $\log F$ is the upper boundary of the concave hull of the graph of $\log F$ on (ℓ_F, ∞). The *T-log-concave envelope* of F, denoted $F^{(T)}$, is in Δ^+ and is given by

$$F^{(T)} = h_T(\overline{k_T F}).$$

With this notation we state the following properties of conjugate transforms, T-log-concave envelopes and τ_T functions which we shall require later. Proofs can be found in [595,596].

Theorem 5.2.12 *For all* $T \in \mathcal{T}_A$ *and all* $F, G \in \Delta^+$,

1. $C_T^* C_T F = F^{(T)}$.
2. $F^{(T)} \geq F$ *and* $F^{(T)} = F$ *iff* $F \in \Delta_T^+$.
3. $F^{(T)} = G^{(T)}$ *iff* $C_T F = C_T G$.
4. $C_T^*(C_T F \times C_T G) = \tau_T\left(F^{(T)}, G^{(T)}\right)$.
5. $(\tau_T(F,G))^{(T)} \geq \tau_T\left(F^{(T)}, G^{(T)}\right)$ *with equality iff* T *is strict.*

6. $G(x) = F(ax)\ \forall a, x > 0 \ \Rightarrow\ C_T G(z) = C_T F(z/a)\ \forall z \geq 0.$

7. $C_T F(0) = \lim_{x\to\infty} k_T F(x).$

8. $\tau_T(F, G) \leq \tau_T\left(F^{(T)}, G^{(T)}\right).$

9. $\epsilon_\alpha^{(T)} = \epsilon_\alpha \quad \forall a \in \Re^+.$

In order to prove our limit results we also need the following theorem (theorem 3.1 of [595]) which follows by induction from theorem 5.2.9.

Theorem 5.2.13 *Let $T \in \mathcal{T}_A$ and $F_i \in \Delta^+$, $i = 1, \ldots, N$. Then for all $z \geq 0$*

$$C_T(\tau_T(F_1, \ldots, F_N))(z) = \max\left\{k_T(0),\ \prod_{i=1}^{N} C_T F_i(z)\right\}.$$

Finally we define the notion of weak convergence in order to state theorem 5.2.15 below.

Definition 5.2.14 *Let $\{F_i\}$ and F be in Δ^+. Then we say $\{F_i\}$* converges weakly *to F, and write*

$$F_i \xrightarrow{w} F$$

if $F_i(x) \to F(x)$ at each continuity point x of F.

This topology is in fact metrizable (see section 4.2 of [718]). The relationship between weak convergence of probability distributions and convergence of T-conjugate transforms is given by (theorem 2.6 of [595])

Theorem 5.2.15 *Let $T \in \mathcal{T}_A$ and let $\{F_i\}$ be a sequence of $F_i \in \Delta^+$. Then for $F \in \Delta^+$,*

$$F_i^{(T)} \xrightarrow{w} F^{(T)} \quad \textit{iff} \quad C_T F_i(z) \to C_T F(z) \qquad \forall z > 0.$$

5.3 Convergence of Dependency Bounds of Normalised Sums

This section is devoted to a proof of our main result which is

Theorem 5.3.1 *Let $\{X_i\}$ be a sequence of random variables with distribution functions $F_i = df(X_i) \in \Delta_{LU}$ and let $T \in \mathcal{T}_C$. Then*

$$\mathrm{ldb}_T\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \xrightarrow{w} \epsilon_\alpha \tag{5.3.1}$$

as $N \to \infty$, where $\alpha = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} u_{F_i}$ and u_{F_i} is as in (5.2.2).

This theorem has the following two corollaries.

Corollary 5.3.1 *Let $\{X_i\}$, $\{F_i\}$ and T be as above. Then*

$$\mathrm{udb}_T\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \xrightarrow{w} \epsilon_\beta \tag{5.3.2}$$

as $N \to \infty$, where $\beta = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \ell_{F_i}$ and ℓ_{F_i} is as in (5.2.1). Note that in the above two results if $u_{F_i} = u_F$ and $\ell_{F_i} = \ell_F$ for all i, then $\alpha = u_F$ and $\beta = \ell_F$.

Corollary 5.3.2 *Let $\{X_i\}$ be a sequence of random variables with distributions $F_i = df(X_i) \in \Delta_{LU}^+$ and let $T \in \mathcal{T}_C$. Then*

$$\mathrm{ldb}_T\left(\left(\prod_{i=1}^{N} X_i\right)^{1/N}\right) \xrightarrow{w} \epsilon_a \tag{5.3.3}$$

and

$$\mathrm{udb}_T\left(\left(\prod_{i=1}^{N} X_i\right)^{1/N}\right) \xrightarrow{w} \epsilon_b \tag{5.3.4}$$

as $N \to \infty$, where $a = \lim_{N\to\infty}\left(\prod_{i=1}^{N} u_{F_i}\right)^{1/N}$ and $b = \lim_{N\to\infty}\left(\prod_{i=1}^{N} \ell_{F_i}\right)^{1/N}$

In order to prove theorem 5.3.1 we prove the following lemmata. The proof of the theorem follows from their conjunction.

Lemma 5.3.1 *Let $F \in \Delta_U^+$ and let $T \in \mathcal{T}_A$. Then there exists a $z_u > 0$ such that for all $z \in (0, z_u)$*

$$C_T F(z) = e^{-u_F z}, \tag{5.3.5}$$

where u_F is given by (5.2.2).

PROOF. Recall that

$$C_T F(z) = \sup_{x \geq 0} \; e^{-xz} k_T F(x) \qquad z \in \Re^+. \tag{5.3.6}$$

However k_T is strictly increasing and continuous and so $k_T F$ is increasing. Also $k_T(1) = 1$ and so $k_T(x) < 1$ for $x < 1$. Therefore

$$k_T F(x) \begin{cases} = 1 & x \geq u_F, \\ < 1 & x < u_F. \end{cases}$$

Now since the slope of e^{-xz} can always be made as close to zero as desired by making z smaller, it can be seen that for any $F \in \Delta_U^+$, for small enough z the supremum in (5.3.6) must occur at $x = u_F$ (see figure 5.1). If, for a given z_u the supremum occurs at $x = u_F$, then obviously it will occur at $x = u_F$ for all $z < z_u$. ■

Figure 5.1: Illustration for the proof of lemma 5.3.1.

Lemma 5.3.2 *Let $\phi(z) = \max\{k_T(0),\ e^{-\alpha z}\}$ for $z \in \Re^+$, $\alpha \in (0,\infty)$, where k_T is the inner multiplicative generator of some $T \in \mathcal{T}_A$. Then the inverse transform of ϕ is given by*

$$C_T^*\phi(x) = \epsilon_\alpha(x) \qquad \forall x \in \Re^+.$$

PROOF. We know [718, theorem 7.8.2 (vi)] that $C_T\epsilon_\alpha(z) = \max\{k_T(0),\ e^{-\alpha z}\}$ for all $\alpha, z \in [0,\infty)$. We also know (theorem 5.2.12, part 1) that for all $T \in \mathcal{T}_A$ and for all $F \in \Delta^+$, $C_T^* C_T F = F^{(T)}$. The lemma then follows directly from the fact that $\epsilon_\alpha^{(T)} = \epsilon_\alpha$. ■

Lemma 5.3.3 *Let $T \in \mathcal{T}_C$ and let $\{X_i\}$ be a sequence of random variables with distribution functions $F_i = df(X_i)$, where $F_i \in \Delta_U^+$ for $i = 1, 2, \ldots$. Then*

$$\lim_{N\to\infty} C_T\left(\mathrm{ldb}_T\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\right)(z) \;=\; \max\{k_T(0),\ e^{-\alpha z}\} \tag{5.3.7}$$

where $\alpha = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} u_{F_i}$.

PROOF. Theorem 5.2.6 and the discussion following it tells us that $\mathrm{ldb}_T\left(\sum_{i=1}^{N} X_i'\right) = \tau_T(F_1', \ldots, F_N')$, where $F_i' = df(X_i') \in \Delta^+$. If we set $X_i' = X_i/N$, then obviously $F_i'(x) = F_i(Nx)$ for $i = 1, 2, \ldots$ and for $N > 0$. We also know (theorem 5.2.13) that

$$C_T(\tau_T(F_1', \ldots, F_N'))(z) \;=\; \max\left\{k_T(0),\ \prod_{i=1}^{N} C_T F_i'(z)\right\}$$

and that $C_T F_i'(z) = C_T F_i(z/N)$ (theorem 5.2.12, part 6). Thus

$$C_T\left(\mathrm{ldb}_T\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\right)(z) \;=\; \max\left\{k_T(0),\ \prod_{i=1}^{N} C_T F_i(z/N)\right\}.$$

However since $F_i \in \Delta_U^+$, lemma 5.3.1 tells us that

$$C_T F_i(x) = \exp\left(-u_{F_i} x\right) \qquad \text{for } x < z_u^{(i)}$$

for $i = 1, 2, \ldots, N$, where $z_u^{(i)}$ is the z_u of lemma 5.3.1 for a given F_i. If we let $z_u = \min_i\{z_u^{(i)}\}$ we can then write

$$C_T\left(\mathrm{ldb}_T\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\right)(z) \;=\; \max\left\{k_T(0),\ \prod_{i=1}^{N} \exp\left(-u_{F_i} z/N\right)\right\} \qquad \forall\left(\tfrac{z}{N}\right) \in (0, z_u). \tag{5.3.8}$$

Now for any $z \in \Re^+$, there exists an integer N_m such that for all $N > N_m$, $(z/N) < z_u$. Therefore, for all $z \in \Re^+$, (5.3.8) holds for sufficiently large N. If we now observe that

$$\prod_{i=1}^{N} \exp\left(-u_{F_i} z/N\right) \;=\; \exp\left(-\frac{1}{N}\left(\sum_{i=1}^{N} u_{F_i}\right) z\right) \;=\; e^{-\alpha z}$$

the proof is completed. ■

Lemma 5.3.4 *Let $H^{(T)} = (\mathrm{ldb}_T(S_N))^{(T)}$ be the T-log-concave envelope of $\mathrm{ldb}_T(S_N)$, where $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$, $T \in \mathcal{T}_C$, and $\{X_i\}$ is a sequence of random variables with distribution functions $F_i = df(X_i) \in \Delta_U^+$. Then*

$$H^{(T)} \stackrel{w}{\to} \epsilon_\alpha \tag{5.3.9}$$

as $N \to \infty$ and α is as in lemma 5.3.3.

PROOF. This follows directly from theorem 5.2.15, lemma 5.3.3, lemma 5.3.2 and theorem 5.2.12 (part 9). ■

Lemma 5.3.5 *Let S_N, T, and α be as in lemma 5.3.4. Then*

$$\lim_{N\to\infty} \sup \ \mathrm{supp}\ \mathrm{ldb}_T(S_N) \ = \ \alpha. \tag{5.3.10}$$

PROOF. This follows by induction from the case when $N = 2$. Thus we now prove that

$$\sup \ \mathrm{supp}\ \mathrm{ldb}_T(X_1 + X_2) \ = \ u_{F_1} \ + \ u_{F_2}.$$

Theorem 5.2.6 tells us that

$$\mathrm{ldb}_T(X_1 + X_2) \ = \ \sup_{u+v=x} \ T(F_1(u), F_2(v)).$$

Since T is an Archimedean t-norm and a copula we know that $T \leq M$ and that $T(a, b) = 1$ implies $a = b = 1$. Therefore the minimum x such that $F_1(u) = F_2(v) = 1$ is $x = u_{F_1} + u_{F_2}$. ■

Lemma 5.3.6 *Let $H \in \Delta_U^+$ be such that $H^{(T)} = (\mathrm{ldb}_T(S_N))^{(T)}$ as in lemma 5.3.4. Then*

$$H \stackrel{w}{\to} \epsilon_\alpha.$$

PROOF. Let $\mathcal{H}^{(T)} = \{G \in \Delta_U^+ |\ G^{(T)} = H^{(T)}\}$. Then

$$\forall G \in \mathcal{H}^{(T)} \quad H^{(T)} \geq G \tag{5.3.11}$$

(theorem 5.2.12, part 2). That is, $H^{(T)}$ is the maximal element in $\mathcal{H}^{(T)}$. The only H that satisfies (5.3.9) and $H \leq \epsilon_\alpha = \epsilon_\alpha^{(T)}$ is $H = \epsilon_\alpha$. ■

Lemma 5.3.7 *Let $\{F_i\}$ be a sequence of distribution functions in Δ_L and let $T \in \mathcal{T}_A$. Then if $\gamma = \min_i\{\ell_{F_i}\}$,*

$$\tau_T(F_1, \ldots, F_N)(x) \ = \ \tau_T(F_1', \ldots, F_N')(x + \gamma N),$$

where $F_i'(x) = F_i(x - \gamma)$ and $F_i' \in \Delta^+$ for $i = 1, \ldots, N$.

PROOF. For $N = 2$ the result follows directly from the definition of τ_T. The result for general N then follows by induction using the iterative construction of $\tau_T^{(N)}$ given by (5.2.6). ■

PROOF OF THEOREM 5.3.1 Immediate from lemmas 5.3.6 and 5.3.7. ■

PROOF OF COROLLARY 5.3.1 We firstly prove a restricted version of corollary 5.3.1 by a method which gives an understanding of why it works in a case of special interest ($T = W$).

Corollary 5.3.1′ *Let $\{X_i\}$ and $\{F_i\}$ be as in corollary 5.3.1 and let $T \in \mathcal{T}_A$ be such that $T^* = T^d$ (see (5.2.3) and (5.2.4)). Then*

$$\mathrm{udb}_T\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) \xrightarrow{w} \epsilon_\beta$$

as $N \to \infty$, and $\beta = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \ell_{F_i}$.

PROOF. We prove this by expressing udb_T in terms of the lower dependency bounds of some transformed random variables. Again we consider $N = 2$ and the general case follows by induction. Let $X_i' = -X_i$. Then $F_i'(x) = 1 - F_i(-x)$. Observe that

$$\begin{aligned}
\tau_T(F_1', F_2')(x) &= \sup_{u+v=x} T(F_1'(u), F_2'(v)) \\
&= \sup_{u+v=x} T(1 - F_1(-u), 1 - F_2(-v)) \\
&= \sup_{u+v=x} [1 - T^*(F_1(-u), F_2(-v))] \\
&= 1 - \inf_{u+v=x} T^*(F_1(-u), F_2(-v)) \\
&= 1 - \inf_{u+v=-x} T^d(F_1(u), F_2(v)) \\
&= 1 - \rho_T(F_1, F_2)(-x) \\
&= 1 - \mathrm{udb}_T(X_1 + X_2)(-x).
\end{aligned}$$

Writing this the other way around we have

$$\mathrm{udb}_T(X_1 + X_2)(x) = 1 - \mathrm{ldb}_T(X_1' + X_2')(-x) \tag{5.3.12}$$

and we can apply theorem 5.3.1 noting that the roles of u_{F_i} and ℓ_{F_i} will be reversed. ■

The question of when $T^* = T^d$ has in fact been solved by Frank [276] who showed that apart from $T = W$ and $T = \Pi$, the only $T \in \mathcal{T}_A$ satisfying $T^* = T^d$ are given by

$$T_s(x, y) = \log\left(1 + \frac{(s^x - 1)(s^y - 1)}{s - 1}\right) / \log s$$

for $s \in (0, \infty)$, $s \neq 1$. In fact, $T_s \in \mathcal{T}_C$.

The more general corollary 5.3.1 is proved in a different manner:

PROOF OF COROLLARY 5.3.1. Corollary 5.3.1′ establishes the result for $T = W$. We also know [597, theorem 2] that for any copulas C_1 and C_2 such that $C_1 \leq C_2$, $\rho_{C_2} \leq \rho_{C_1}$. That is

$$\forall F, G \in \Delta^+ \quad \rho_{C_2}(F, G) \;\leq\; \rho_{C_1}(F, G).$$

Since all copulas C satisfy $W \leq C$ we have

$$\rho_C \;\leq\; \rho_W. \tag{5.3.13}$$

Combining (5.3.13) with the fact (to be proved below) that $\inf\{x | \rho_C(x) = 1\} = \beta$ for all $C \in \mathcal{T}_C$ completes the proof.

In order to see that

$$\inf\{x |\, \rho_C(F_1, \ldots, F_N)(x) = 1\} = \beta \tag{5.3.14}$$

for all $F_1, \ldots, F_N \in \Delta_L$ and any $C \in \mathcal{T}_C$ we again use induction on N and start with $N = 2$. Consider then

$$\rho_C(F_1, F_2)(x) \;=\; \inf_{u+v=x} C^d(F_1(u), F_2(v)).$$

Now C^d has 1 as a null element. That is $C^d(x, 1) = C^d(1, x) = 1$ for all $x \in I$. Therefore in order for $C^d(x, y)$ to equal one it is necessary only that either x or y equal one. Thus for $C^d(F_1(u), F_2(v))$ to equal one either $F_1(u) = 1$ or $F_2(v) = 1$. The smallest x such that $u + v = x$ and either $F_1(u)$ or $F_2(v)$ equals one is $x = \ell_{F_1} + \ell_{F_2}$. The rest of the proof follows by induction noting that we have to take account of the $1/N$ term. ■

PROOF OF COROLLARY 5.3.2 Simply let $X_i' = \log X_i$, apply the results of theorem 5.3.1 and corollary 5.3.1, and exponentiate the result. ■

5.4 Rates of Convergence

Theorem 5.3.1 says that $\mathrm{ldb}_T(S_N) \xrightarrow{w} \epsilon_\alpha$. We now examine how fast the convergence is. The main tool we use is theorem 4.2 of [596] which we state below as theorem 5.4.1. Since if $T \in \mathcal{T}_A$ is not strict then Δ_T^+ is not closed under τ_T [596, page 25], we define the set $\mathcal{B}_T$ which is. Firstly define $(k_T F)_\ell \in \Delta^+$ by

$$(k_T F)_\ell(x) = \begin{cases} 0 & x \leq \ell_F \\ k_T(F(x)) & x > \ell_F \end{cases}$$

where ℓ_F is given by (5.2.1). Then define

$$\mathcal{B}_T = \{F \in \Delta^+ |\, (k_T F)_\ell \text{ is log-concave}\}.$$

If T is strict $\mathcal{B}_T = \Delta_T^+$. Otherwise $\Delta_T^+ \subset \mathcal{B}_T$ Now if for any $\mu > 0$ and $F \in \mathcal{B}_T$ with $F \neq \epsilon_\infty$ we define $F^{[\mu]} \in \Delta^+$ by

$$F^{[\mu]}(x) = h_T((k_T F)_\ell(x/\mu))^\mu$$

and let $F^{[0]} = 0$ we can then state

Theorem 5.4.1 ([596])

$$\tau_T\left(F^{[\mu]}, F^{[\nu]}\right) = F^{[\mu+\nu]} \qquad \forall \mu, \nu \geq 0.$$

Theorem 5.4.2 *Let $T \in \mathcal{T}_C$ and let $\{X_i\}$ be a sequence of random variables with identical distribution functions $df(X_i) = F \in \mathcal{B}_T$. Then*

$$\mathrm{ldb}_T\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)(x) \;=\; h_T\left[((k_T F)_\ell(x))^N\right].$$

PROOF. The result follows immediately from theorem 5.4.1 when the $\frac{1}{N}$ is taken account of in the manner of the proof of lemma 5.3.3. ■

A particularly interesting and important special case of theorem 5.4.2 occurs when $T = W$. In this case we have $k_W(x) = e^{x-1}$ and so

$$\mathcal{B}_W \;=\; \{F \in \Delta^+ \mid F \text{ is concave on } (\ell_F, \infty)\}$$

and

$$(k_W F)_\ell(x) = \begin{cases} 0 & x \leq \ell_F \\ e^{F(x)-1} & x > \ell_F. \end{cases}$$

Recalling that

$$h_w(x) = \begin{cases} 0 & 0 \leq x \leq e^{-1}, \\ 1 + \log\, x & e^{-1} \leq x < 1, \end{cases}$$

we have

$$h_W\left[((k_W F)_\ell(x))^N\right] = 0 \quad \text{for } ((k_w)_\ell(x))^N \leq e^{-1}.$$

The condition here is equivalent to

$$(e^{F(x)-1})^N \leq e^{-1}. \tag{5.4.1}$$

If we assume that F has an inverse F^{-1}, then (5.4.1) implies

$$x \leq F^{-1}\left(1 - \frac{1}{N}\right).$$

In other words, with $\{X_i\}$ as in theorem 5.4.2,

$$\mathrm{ldb}_W\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)(x) \;=\; 0 \quad \text{for } x \;\leq\; F^{-1}\left(1 - \frac{1}{N}\right).$$

If $x > F^{-1}\left(1 - \frac{1}{N}\right)$, then $h_W(x) = 1 + \log\, x$ and so

$$\mathrm{ldb}_W\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)(x) \;=\; 1 + \log\left(\left[e^{F(x)-1}\right]^N\right) \;=\; 1 + N(F(x) - 1).$$

Again assuming that F has an inverse we can write

$$\mathrm{ldb}_W^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)(y) \;=\; F^{-1}\left(\frac{y-1}{N}+1\right). \tag{5.4.2}$$

The convergence to ϵ_α is apparent from (5.4.2) upon setting $y = 1$ and $y = 0$. In these cases we have

$$\mathrm{ldb}_W^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)(1) \;=\; F^{-1}(1) \tag{5.4.3}$$

and

$$\mathrm{ldb}_W^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)(0) \;=\; F^{-1}\left(1-\frac{1}{N}\right). \tag{5.4.4}$$

Equations (5.4.3) and (5.4.4) tell us that the rate of convergence is $O(N)$.

Another interesting special case is when $T = \Pi$ in which case

$$\mathrm{ldb}_\Pi\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)(x) \;=\; [F(x)]^N,$$

which is a particularly simple result. The condition $T = \Pi$ is equivalent to Lehmann's *positive quadrant dependence* [506]. Two random variables X and Y are positively quadrant dependent if

$$F_{XY}(x,y) \;\geq\; F_X(x)F_Y(y).$$

Positive quadrant dependence has been studied and compared with other measures of dependence in [258,313,459].

If $F \notin \mathcal{B}_T$ we can use the fact (theorem 5.2.12, part 8) that $\tau_T(F,G) \leq \tau_T(F^{(T)}, G^{(T)})$ for any $T \in \mathcal{T}_A$ to bound the rate of convergence. For example, using (5.4.4) we can say that for any $F \in \Delta$,

$$\sup\left\{x \mid \mathrm{ldb}_W\left(\frac{X_1+X_2}{2}\right)(x) = 0\right\} \geq F^{(T)}\left(\tfrac{1}{2}\right).$$

This behaviour can be seen in figure 5.4 presented in the next section. Rates of convergence for udb_T are similar and follow directly using the arguments used to prove corollaries 5.3.1 and 5.3.1′.

5.5 Examples

We will now present some examples illustrating the results of sections 5.3 and 5.4. We restrict ourselves to the case $T = W$ and use the algorithms developed in chapter 3 for numerically calculating $\tau_W(F,G)$ and $\rho_W(F,G)$ when F and G are represented by discrete approximations. We check the accuracy of the results so obtained in figure 5.6 where we compare the W obtained using the numerical approximations with that obtained using (5.4.2).

Figure 5.2: Lower and upper dependency bounds for $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ with $N = 2, 3, \ldots, 8$.

Figure 5.3: Lower and upper dependency bounds for $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ with $N = 2, 4, 8, \ldots, 128$.

Figures 5.2 and 5.3 show the lower and upper dependency bounds for $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$. In this case, all the X_i are identically distributed. Their distribution is presented as the two central curves in figures 5.2 and 5.3 ($\underline{F}$ and $\overline{F}$). The reason why there are two curves is because they are the output of a confidence interval estimation procedure designed to generate the lower and upper discrete approximations to probability distribution functions developed in chapter 3. In this case we estimated the distribution of a population consisting of a mixture of samples from $N(0,1)$ and $N(8,1)$ distributions. In both cases the normal distributions were curtailed at 5. Figure 5.2 shows $\mathrm{ldb}_W(S_N)$ and $\mathrm{udb}_W(S_N)$ for $N = 2, 3, \ldots, 8$. This was generated by iteratively calculating

$$\begin{aligned} \underline{F}_i &= \mathrm{ldb}_W(\underline{F}, \underline{F}_{i-1}) \\ \overline{F}_i &= \mathrm{udb}_W(\overline{F}, \overline{F}_{i-1}). \end{aligned} \tag{5.5.1}$$

In order to speed up our view of the convergence, in figure 5.3 we used the iteration

$$\begin{aligned} \underline{F}_i &= \mathrm{ldb}_W(\underline{F}_{i-1}, \underline{F}_{i-1}) \\ \overline{F}_i &= \mathrm{udb}_W(\overline{F}_{i-1}, \overline{F}_{i-1}). \end{aligned} \tag{5.5.2}$$

Equation 5.5.2 has the effect of doubling N at each iteration. Thus figure 5.3 shows $\mathrm{ldb}_W(S_N)$ and $\mathrm{udb}_W(S_N)$ for $N = 2, 4, 8, \ldots, 128$. It can be seen that cases for $N = 2$, 4 and 8 are identical to those in figure 5.2. The convergence of $\mathrm{ldb}_W(S_N)$ and $\mathrm{udb}_W(S_N)$ to ϵ_α and ϵ_β is apparent.

Figures 5.4, 5.5 and 5.6 are all related to S_N where $df(X_i) = F$ for all i and F is an upper half Gaussian distribution. That is $F(x) = 2\Phi(x) - 0.5$, where Φ is the distribution of a $N(0,1)$ random variable. In this case F is curtailed at $\mu + 4\sigma = 4$. Again we represent F by lower and upper discrete approximations (this time touching each other) and these can be seen as the central curves in figure 5.4. Recalling the symmetry relationship between ldb_W and udb_W (see corollary 5.3.1′), we will be able to observe the effect of both a concave and convex F. (The lower and upper dependency bounds for F being a *lower* half Gaussian distribution can be seen by viewing figures 5.4–5.6 upside down and changing the axes appropriately.) Figure 5.5 shows the dependency bounds calculated using equation 5.4.2. As can be seen from figure 5.6 and by comparing figures 5.4 and 5.5, the results agree very closely with those obtained using the numerical approximations. The only difference is due to the chording effect apparent in figure 5.4. This is an artifact of the numerical representation we use in [874]. We approximate F by uniformly discretising its quantiles. Considering the effect of (5.4.2) it can be seen that the top most discrete levels are stretched down in successive iterations to result in the straight line segments which appear in figure 5.4. The bounds obtained numerically are still correct, but they are not as tight as the true dependency bounds. Once there are straight line segments, they remain present in successive iterations. This can be understood in terms of Alsina's result [23] that

$$\tau_W(U_{a,b}, U_{c,d}) = U_{\min(a+d,b+c),b+d}$$

and

$$\rho_W(U_{a,b}, U_{c,d}) = U_{a+c,\max(a+d,b+c)}$$

Figure 5.4: Dependency bounds for S_N where $df(X_i)$ is upper half Gaussian.

Figure 5.5: Dependency bounds for the same problem as figure 5.4, but calculated using (5.4.2).

Figure 5.6: Comparison between the exact and numerical calculation of the dependency bounds (figures 5.4 and 5.5).

where $U_{a,b}$ is the uniform distribution function on $[a, b]$.

Observe that udb_W converges to ϵ_β more rapidly than ldb_W converges to ϵ_α (or equivalently, ldb_W converges to ϵ_α more rapidly for lower half Gaussian F than it does for upper half Gaussian F). This is due to the concavity (or convexity) of F. Looking at figure 5.4 upside down demonstrates the effect mentioned at the end of section 5.4. Line (m) is $F^{(T)}$ for F being lower half Gaussian and it can be seen that

$$\sup\left\{x|\ \mathrm{ldb}_W\left(\frac{X_1 + X_2}{2}\right)(x) = 0\right\} = F^{(T)}(1/2).$$

5.6 Conclusions

Apart from being mathematically interesting, the results presented in this chapter have a useful practical interpretation. Theorem 5.3.1 says that in order to determine dependency bounds for the sum of a large number of random variables it is only necessary to consider the values of supp $F = [\ell_F, u_F]$. In other words, dependency bounds reduce to Minkowski sums (or products) of intervals and the methods of interval arithmetic [17,583,584,800] could be used for their calculation. The significant point is that the shapes of the F_i within supp F_i have no effect on the final result.

Observe that this situtation is quite different to the classical *central* limit theorem and law of large numbers for sums (where we are looking at ordinary convolutions rather than dependency bounds) [520]. In this case the support of the component distributions is irrelevant to the final limiting distribution apart from some non-degeneracy conditions. Note also that this result has a bearing on statistical inference. It is often considered adequate to fit a distribution to some data in such a manner that the fit is close over the central part of the distribution. When the densities are plotted in the usual manner anomalies in the tails do not show up. However, as Bagnold has shown [43], populations that appear to have normal distributions when plotted in the usual manner are quite apparently non-normal in the tails when a logarithmic vertical scale is used. Determination of the tail behaviour is of course the domain of the theory of statistics of extremes [344,345] and has significant practical application to the study of rare events.

The results presented in this chapter can also be applied to fuzzy numbers. In this case a general law of large numbers under a general t-norm extension principle is obtained (see the following chapter).

Chapter 6

A Law of Large Numbers for Fuzzy Variables

The use of fuzzy numbers for calculation with imprecisely known quantities has been advocated by a number of different authors. Fuzzy numbers are combined using extended arithmetic operations developed using the extension principle. When a general t-norm is used for the intersection operator, a general t-norm extension principle is obtained. The purpose of this chapter is to present a result for the law of large numbers for fuzzy variables when using this general t-norm extension principle. The result is that convergence to a crisp set is obtained for all Archimedean t-norm intersection operators. This generalises a previous result of Badard who conjectured that something along the lines of that presented here would be true. The result is proved by deriving it from a similar result for the law of large numbers for dependency bounds. Dependency bounds arise in probabilistic arithmetic when nothing is known of the joint distribution of two random variables apart from the marginal distributions. The bridge used to connect dependency bounds with fuzzy number operations under the t-norm extension principle can be used to give a probabilistic interpretation of fuzzy number combination. Some remarks on this interconnection are made in the final section of the chapter.

This chapter, after undergoing some revisions, will be resubmitted to *Fuzzy Sets and Systems* under the title "The Law of Large Numbers for Fuzzy Variables under a General Triangular Norm Extension Principle."

6.1 Introduction

Fuzzy numbers or variables are developed from the theory of fuzzy sets by using the extension principle [225,606,853,898]. Their properties and methods of calculating with them have been studied by a number of authors. A more general extension principle [228] makes use of a general t-norm intersection operator [24,26,230]. When this is used, the sum of two normal fuzzy variables X and Y with membership

functions μ_X and μ_Y is given by

$$\mu_Z(z) = \sup_{x+y=z} T(\mu_X(x), \mu_Y(y)), \tag{6.1.1}$$

where T is some triangular norm (or t-norm). A natural question to ask is: What is the limiting behaviour of

$$Z = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{6.1.2}$$

where $\{X_i\}$ are fuzzy variables and we use the sup-T convolution (6.1.1) to calculate the membership function of their sum? This is the fuzzy analogue of the law of large numbers for random variables [262].

In this paper we will show that for nearly all t-norm intersection operators, the membership function of Z approaches that of a crisp set or interval. This result generalises some special cases reported by Badard [41] and Rao and Rashed [673]. The result was also mentioned by Dubois and Prade in [239, p.26]. The result in the present paper is a law of large numbers for fuzzy variables (not fuzzy random variables [101,489,490,578] which are random variables that take on fuzzy values). We prove the result in three seperate ways, each of which gives a different insight into the problem. The first is a short and simple direct proof. The second uses a recently developed theorem on a related question for dependency bounds of sums of random variables [870]. We thus incidentally provide a probabilistic interpretation of the sup-T convolution (6.1.1) used for fuzzy number addition. A third proof (only sketched here) shows how T-conjugate transforms can be used in the study of operations of the form (6.1.1). These transforms play a role analogous to that of the Fourier transform in probability theory.

Some numerical examples are presented which illustrate the convergence of membership functions in calculating (6.1.2). These examples are of independent interest because they are the first "exact " calculation of operations on fuzzy numbers (not just addition) for t-norms other than min. (Dubois and Prade worked in terms of their simple L-R representation in [228] and their formulae are necessarily only approximate for operations other than addition or subtraction). We give details on the new methods used to calculate our examples.

The rest of this paper is organised as follows. Section 2 presents the main theorem and the simple direct proof. Section 3 briefly introduces the notion of dependency bounds. This is necessary for the understanding of theorem 2 which is used in the second derivation of theorem 1. Section 4 is devoted to this derivation and in section 5 we outline a different approach to proving theorem 1 by introducing the bilateral T-conjugate transform. Section 6 introduces a means of numerically calculating the sup-T convolutions, and some examples using this method are presented in section 7. These examples graphically illustrate the convergence of theorem 1. Finally, in section 8 we mention some interpretation issues raised by the bridge between probabilistic dependency bounds and fuzzy variable convolutions.

6.2 Theorem and Direct Proof

The fuzzy variables we will be concerned with are *normal* and *T-noninteractive*. A normal fuzzy number X has a membership function μ_X such that $\sup_x \mu_X(x) = 1$. If $\{X_i\}_{i=1}^N$ is a set of fuzzy variables with normalised unimodal continuous membership functions $\{\mu_{X_i}\}$ and joint membership function $\mu_{X_1 \ldots X_N}$, then $\{X_i\}$ are T-noninteractive if

$$\mu_{X_1 \ldots X_N}(x_1, \ldots, x_N) = T^{N-1}(\mu_{X_1}(x_1), \ldots, \mu_{X_N}(x_N)),$$

where T^{N-1} is the $(N-1)$th serial iterate of some t-norm T [718, page 87] defined by $T^{N-1}(x_1, \ldots, x_N) = T(T^{N-2}(x_1, \ldots, x_{N-1}), x_N)$ with $T^1 = T$. The notion of T-noninteractivity is called $*$-independence by Badard [41] and weak noninteraction by Dubois and Prade [228]. T-noninteractivity is a generalisation of Rao and Rashed's [673] min-relatedness.

Whilst noninteractivity is often assumed in discussions in the literature, the study of *interactive* variables has only begun recently [114,115,228,239,816]. Without clear semantics for the notion of interaction of fuzzy variables, there is little incentive for pursuing it further.

The t-norms we will be concerned with are Archimedean. An *Archimedean t-norm* is one which satisfies $T(a,a) < a \ \forall a \in (0,1)$. Apart from min, nearly all of the t-norms suggested in the fuzzy set literature [24,26,230,232,856] are Archimedean.

Theorem 6.2.1 *Let $\{X_i\}_{i=1}^N$ be a set of T-noninteractive fuzzy variables with membership functions μ_{X_i}, $i = 1, \ldots, N$. Assume that T is Archimedean. Let*

$$\alpha_{X_i} = \inf\{x | \, \mu_{X_i}(x) = 1\}, \tag{6.2.3}$$

$$\beta_{X_i} = \sup\{x | \, \mu_{X_i}(x) = 1\} \tag{6.2.4}$$

and let

$$Z = \frac{1}{N} \sum_{i=1}^{N} X_i$$

have a membership function μ_Z. Let α_Z and β_Z be defined analogously to α_{X_i} and β_{X_i}. Then

$$\alpha_Z = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \alpha_{X_i},$$

$$\beta_Z = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \beta_{X_i},$$

and

$$\mu_Z \stackrel{w}{\to} \mathbf{1}_{[\alpha_Z, \beta_Z]} \tag{6.2.5}$$

as $N \to \infty$. The symbol $\stackrel{w}{\to}$ in (6.2.5) means weak convergence (convergence at every continuity point), and $\mathbf{1}_{[a,b]}$ denotes the indicator function of the interval $[a,b]$ defined by

$$\mathbf{1}_{[a,b]}(x) = \begin{cases} 1 & x \in [a,b] \\ 0 & \textit{otherwise.} \end{cases}$$

This theorem says that only the values of α_{X_i} and β_{X_i} have any effect on the limiting membership function, which is that of a crisp (non-fuzzy) set. If $\alpha_{X_i} = \beta_{X_i}$ for all i, then none of the fuzzy variables have a "flat" in their membership functions, and μ_Z converges to a single delta function,

$$\mu_Z(x) = \begin{cases} 1 & x = \alpha_Z \\ 0 & x \neq \alpha_Z. \end{cases}$$

In other words the "shape preservation property" of sup-min convolutions is a special case (min is not Archimedean).

PROOF. Using the fact that T is Archimedean, and the relationship $T(b,1) = T(1,b) = b \ \forall b \in [0,1]$ (which holds for all t-norms), it can be seen that $\mu_Z^T(z) = \sup_{x+y=z} T(\mu_X(x), \mu_Y(y)) = 1$ only if there exists x and y such that $x + y = z$ and $\mu_X(x) = \mu_Y(y) = 1$. Upon noting that all Archimedean t-norms are strictly less than min, then one can use the fact that sup-min convolution has the shape preservation property to argue that $\mu_Z^T(z) < \mu_Z^M(z)$ for all z such that $\mu_Z^T(z) < 1$. Since (6.1.1) is an associative operation [718], inductive application of the above argument completes the proof. ■

6.3 Dependency Bounds

This section provides the minimum background necessary to understand the statement of theorem 2 and the discussion in section 8. Further details can be found in [277,718,874].

A *copula* is a function that links multivariate distribution functions to their marginals. It suffices to consider the bivariate case. If $H(x,y)$ is the joint distribution of two random variables X and Y defined on a common probability space, then $H(x,y) = C_{XY}(F_X(x), F_Y(y))$, where $F_X(x) = df(X) = Pr\{X < x\}$ and $F_Y(y) = df(Y) = Pr\{Y < y\}$ are the marginal distributions of X and Y respectively, and C_{XY} is their connecting copula. In other words $C_{XY}(x,y) = H(F_X^{-1}(x), F_Y^{-1}(y))$. All 2-copulas are 2-increasing ($C(a_2,b_2) - C(a_1,b_2) - C(a_2,b_1) + C(a_1,b_1) \geq 0$ for all $a_1, a_2, b_1, b_2 \in I$) and satisfy $W \leq C_{XY} \leq M$, where $W(x,y) = \max(x+y-1, 0)$ and $M(x,y) = \min(x,y)$. If $C_{XY}(x,y) = \Pi(x,y) = xy$, then X and Y are stochastically independent. Copulas are t-norms if they are associative and t-norms are copulas if they are 2-increasing.

Dependency bounds [277] are lower and upper bounds on the distribution of functions of random variables when only the marginal distributions are known. A more general form of dependency bounds considered in [874] provides tighter bounds when some lower bound $\underline{C}_{XY} > W$ on the connecting copula is known. We will assume that $\underline{C}_{XY}$ is a t-norm and use the traditional T to represent it. If $T = W$, then we obtain the bounds of [277]. We write the dependency bounds as

$$\mathrm{ldb}_T(X+Y) \ \leq \ df(X+Y) \ \leq \ \mathrm{udb}_T(X+Y)$$

where ldb_T, df and udb_T refer to lower dependency bound, distribution function

and upper dependency bound respectively. The main result we need is [277]

$$\mathrm{ldb}_T(X+Y)(z) \;=\; \tau_T(F_X,F_Y)(z) \;=\; \sup_{u+v=z} T(F_X(u),F_Y(v)) \tag{6.3.1}$$

and

$$\mathrm{udb}_T(X+Y)(z) \;=\; \rho_T(F_X,F_Y)(z) \;=\; \inf_{u+v=z} T^d(F_X(u),F_Y(v)), \tag{6.3.2}$$

where $F_X = df(X)$, $F_Y = df(Y)$, and T^d is the *dual copula* given by

$$T^d(x,y) = x + y - T(x,y).$$

This should not be confused with the *t-conorm* S defined by

$$S(x,y) = 1 - T(1-x, 1-y).$$

Since the τ_T and ρ_T operations are associative [718], we can iteratively calculate $\mathrm{ldb}_T(X_1,\ldots,X_N)$ and $\mathrm{udb}_T(X_1,\ldots,X_N)$. Theorem 2, presented below, characterises the behaviour of these quantities as $N \to \infty$.

Theorem 6.3.1 *Let $\{Y_i\}_{i=1}^N$ be a sequence of random variables with distribution functions $F_i = df(Y_i)$ such that $[\ell_{F_i}, u_{F_i}]$ is a bounded closed interval for $i = 1,\ldots,N$, where $\ell_{F_i} = \inf\{x|\, F_i(x) > 0\}$ and $u_{F_i} = \sup\{x|\, F_i(x) < 1\}$. Also let T be an Archimedean t-norm and copula. Then*

$$\mathrm{ldb}_T\left(\frac{1}{N}\sum_{i=1}^N Y_i\right) \xrightarrow{w} \varepsilon_\lambda \quad \textit{and} \quad \mathrm{udb}_T\left(\frac{1}{N}\sum_{i=1}^N Y_i\right) \xrightarrow{w} \varepsilon_\psi$$

as $N \to \infty$, where $\lambda = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N u_{F_i}$, $\psi = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \ell_{F_i}$, and ε_ϕ is the unit step distribution function at ϕ,

$$\varepsilon_\phi(x) = \begin{cases} 0 & x \le \phi, \\ 1 & x > \phi. \end{cases}$$

PROOF. See [870].

6.4 Proof of Theorem 1 via Decomposition

Theorem 1 can also be proved by representing the membership functions μ_{X_i} by two probability distributions and then applying theorem 2 (which is in terms of probability distributions).

We decompose the membership function μ_X of some fuzzy variable X as follows. Let

$$\mu_X^i(x) = \begin{cases} \mu_X(x) & x \le \alpha_X, \\ 1 & x > \alpha_X, \end{cases}$$

and

$$\mu_X^d(x) = \begin{cases} \mu_X(x) & x \geq \beta_X, \\ 1 & x < \beta_X, \end{cases}$$

where α_X and β_X are given by (6.2.3) and (6.2.4). We will work with the related pair (μ_X^I, μ_X^D), where

$$\mu_X^I(x) = \mu_X^i(x), \tag{6.4.1}$$

but

$$\mu_X^D(x) = 1 - \mu_X^d(x). \tag{6.4.2}$$

Note that μ_X^I and μ_X^D are both distribution functions (continuous, non-decreasing with range $[0,1]$).

We now need to calculate μ_Z^I and μ_Z^D in terms of the pairs (μ_X^I, μ_Y^I) and (μ_X^D, μ_Y^D) respectively. The key observation to make is that upon examining (6.1.1) and noting that $T(a,1) = T(1,a) = a$ for all $a \in [0,1]$, we find that

$$\alpha_Z = \alpha_X + \alpha_Y \quad \text{and} \quad \beta_Z = \beta_X + \beta_Y.$$

Thus we have

$$\mu_Z^I(z) = \mu_Z^i(z) = \tau_T(\mu_X^i, \mu_Y^i)(z) = \sup_{x+y=z} T(\mu_X^i(x), \mu_Y^i(y)) \tag{6.4.3}$$

and

$$\mu_Z^d(z) = \sup_{x+y=z} T(\mu_Z^d(x), \mu_Z^d(y)). \tag{6.4.4}$$

Substituting (6.4.2) into (6.4.4) gives

$$1 - \mu_Z^D(z) = \sup_{x+y=z} T(1 - \mu_X^D(x), 1 - \mu_Y^D(y)).$$

Therefore

$$\begin{aligned} \mu_Z^D(z) &= 1 - \sup_{x+y=z} T(1 - \mu_X^D(x), 1 - \mu_Y^D(y)) \\ &= \inf_{x+y=z} S(\mu_X^D(x), \mu_Y^D(y)). \end{aligned}$$

If in fact $S = T^d$, we obtain

$$\mu_Z^D(z) = \rho_T(\mu_X^D, \mu_Y^D).$$

Therefore we can use (6.3.1) and (6.3.1) and theorem 2 to prove theorem 1 for all T such that $S = T^d$. This condition can be removed by using an argument along the lines of that used in [870] to prove the udb_T part of theorem 2 for general Archimedean T. (One uses the fact that T is increasing in both places and a bound R on T such that the corresponding t-conorm (of R) equals its dual copula R^d. ■

We note in passing that one could also examine the rate of convergence in theorem 1 by using the results in section 5 of [870]. The details are omitted here. Note also that the above decomposition is of course unnecessary when the notion of a fuzzy number adopted by Höhle and others [391,465,527] is used.

Triangular norms	Generators
$W(x,y) = \max(x+y-1, 0)$	$g_W(x) = (1-x)$ $g_W^{-1}(x) = \max(1-x, 0)$
$\Pi(x,y) = xy$	$g_\Pi(x) = -\log x$ $g_\Pi^{-1}(x) = \exp(-x)$
$T_p^{(Sch)}(x,y) = \begin{cases} \sqrt[p]{\max(x^p+y^p-1,0)} & p \neq 0 \\ \Pi(x,y) & p = 0 \end{cases}$	$g_{T_p^{(Sch)}}(x) = 1-x^p$ $g^{-1}_{T_p^{(Sch)}}(x) = \sqrt[p]{1-y}$ $p \in (-\infty, 1]$
$T_p^{(Yag)}(x,y) = 1 - \min(1, \sqrt[p]{(1-x)^p + (1-y)^p})$	$g_{T_p^{(Yag)}}(x) = (1-x)^p$ $g^{-1}_{T_p^{(Yag)}}(x) = 1 - \sqrt[p]{y}$ $p \in (0, \infty)$
$T_p^{(Fra)}(x,y) = \log_p \left[1 + \frac{(p^x-1)(p^y-1)}{p-1}\right]$ $p \in (0,\infty)$ $p \neq 1$	$g_{T_p^{(Fra)}}(x) = \log_p \left[\frac{p-1}{p^x-1}\right]$ $g^{-1}_{T_p^{(Fra)}}(x) = \log_p \left[1 + \frac{(p-1)}{pe^x}\right]$

Table 6.1: Some Archimedean t-norms and their additive generators.

6.5 A Direct Proof Using Bilateral T-conjugate Transforms

Theorem 1 was proved in [870] by using the properties of T-conjugate transforms [595,596]. By extending the definition of these transforms, it is possible to construct an alternative proof of theorem 1, with no reference to dependency bounds. Most of the details would be the same as for the dependency bound case and they are omitted here. Nevertheless it seems worth mentioning the extension and its possible applications. We begin with the representation of Archimedean t-norms. These are also used in section 6.

A continuous Archimedean t-norm T can always be written [718, theorem 5.5.2]

$$T(x,y) = f_T(g_T(x) + g_T(y)), \tag{6.5.1}$$

where

- g_T (an *inner additive generator of* T) is a continuous and strictly decreasing function from $I = [0,1]$ into $\Re^+ = \{x \,|\, x \in \Re \text{ and } x \geq 0\}$ with $g(1) = 0$.
- f_T (the *outer additive generator of* T) is a continuous function from $\Re^+$ onto I that is strictly decreasing on $[0, g_T(0)]$ and such that $f_T(x) = 0$ for all $x \geq g_T(0)$.
- f_T and g_T are quasi-inverses of each other and unique up to a multiplicative constant.

If T is *strict* (continuous and strictly increasing in each place on $(0,1]^2$), then $f_T = g_T^{-1}$ and $g_T(0) = \infty$. Some examples of t-norms and their additive generators have been recently collected by Mizumoto [580] and a selection are given in table 1. Note the following special cases of the parameterised t-norms (Schweizer [718], Frank [276], Yager [891]):

$$W = T_1^{(Sch)} = T_1^{(Yag)} = \lim_{p\to\infty} T_p^{(Fra)}$$

$$\begin{aligned} \Pi &= T_0^{(Sch)} = \lim_{p\to 1} T_p^{(Fra)} \\ M &= \lim_{p\to -\infty} T_p^{(Sch)} = \lim_{p\to \infty} T_p^{(Yag)} = T_0^{(Fra)}. \end{aligned}$$

There is also a multiplicative representation: A continuous Archimedean t-norm T can always be written

$$T(x,y) = h_T(k_T(x)k_T(y)), \tag{6.5.2}$$

where

- k_T (the *inner multiplicative generator*) is a continuous strictly increasing function from I into I with $k_T(1) = 1$.
- h_T (the *outer multiplicative generator*) is a continuous function from I onto I that is strictly increasing on $[k_T(0), 1]$ and such that $h_T(x) = 0\ \forall x \in [0, k_T(0)]$.
- h_T is a quasi-inverse of k_T and they are both unique up to an exponentiation.

The *T-conjugate transform* of a probability distribution function F is defined by [595,596]

$$C_T F(z) = \sup_{x\geq 0}\ e^{-xz} k_T F(x) \quad \forall z \in \Re^+, \tag{6.5.3}$$

where

$$k_T F(x) = \begin{cases} 0 & x \leq 0, \\ k_T(F(x)) & 0 < x, \end{cases}$$

k_T is an inner multiplicative generator of the Archimedean t-norm T, and it is assumed that $F(0) = 0$. There is an inverse transform given by

$$C_T^* \phi(x) = h_T(\inf_{z\geq 0} e^{xz}\phi(z)) \quad \forall x \in \Re^+,$$

where h_T is an outer multiplicative generator of T. The significance of these transforms for the study of the τ_T-convolutions is that the following property is satisfied:

$$C_T^*(C_T F \times C_T G) = \tau_T(F^{(T)}, G^{(T)})$$

where $F^{(T)}$ is the *T-log-concave envelope* of F [596]. This means that for certain classes of F and G, it is possible to calculate τ_T-convolutions by pointwise multiplication of the T-conjugate transforms followed by an inverse T-conjugate transform. This is analogous to the use of the Fourier transform or characteristic function in probability theory [172]. The T-conjugate transform is in fact closely related to Fenchel's duality theorem [65,263,530]. In fact, the duality result of Frank and Schweizer described in the following section is yet another manifestation of Fenchel's theorem.

In order to apply T-conjugate transforms to the addition of fuzzy variables it is necessary to extend their domain of definition from probability distributions (non-decreasing) to more general unimodal functions. The obvious idea of applying the

transform to the two parts μ_X^I and μ_X^D separately turns out to be equivalent to the simple replacement of the condition $z \in \Re^+$ in (6.5.3) by the condition $z \in \Re$. We thus obtain the *bilateral T-conjugate transform:*

$$C_T\mu_X(z) = \sup_{x \geq 0} \; e^{-xz} k_T \mu_X(x) \qquad \forall z \in \Re. \tag{6.5.4}$$

This is an analogue of the Fourier transform for fuzzy variable addition.

6.6 Numerical Calculation of sup-T Convolutions

In order to compute some examples which illustrate the convergence described by theorem 1, it is necessary to be able to calculate the sup-T convolutions (6.1.1) accurately and easily. In this section we will present a new method for performing this calculation. The numerical representation is very similar to that developed in [874] for calculating convolutions and dependency bounds (when $T = W$). The method presented here is for quite general Archimedean T and determines the result in terms of the inner additive generator g_T of T.

As well as enabling an illustration of theorem 1 to be calculated, the method seems to be the first to allow the calculation of (6.1.1). Whilst some simple properties of (6.1.1) were presented by Dubois and Prade [228] in terms of L-R fuzzy numbers, these were only for the special cases of $T = W$, Π and M. In contrast, the present method allows the use of arbitrary parameterized Archimedean t-norms. Furthermore, operations other than addition are handled just as easily.

We will consider the increasing and decreasing parts of μ_Z separately.

6.6.1 Increasing Part

The main tool we use is the duality theorem of Frank and Schweizer [278]. In order to do so, consider the more general form of (6.1.1) defined by

$$\tau_{T,L}(F,G)(x) \;=\; \sup_{L(u,v)=x} T(F(u),G(v)) \tag{6.6.1}$$

where F and G are non-decreasing left continuous functions from $\Re^*$ onto I ($\Re^* = \Re \cup \{-\infty, \infty\}$ and $I = [0,1]$) and L is a binary operation. In other words F and G are distribution functions. We denote the *quasi-inverse* of F by $F^{\wedge}$ where

$$F^{\wedge}(y) = \sup\{x |\, F(x) < y\}. \tag{6.6.2}$$

Note that if F is continuous, $F^{\wedge} = F^{-1}$. The inverse distribution function is known as the *quantile distribution* in probability theory, and of course corresponds to the level-sets or α-cuts of fuzzy set theory. Frank and Schweizer have shown that

$$\tau_{T,L}^{\wedge}(F^{\wedge},G^{\wedge})(x) = \inf_{T(u,v)=x} L(F^{\wedge}(u),G^{\wedge}(v)). \tag{6.6.3}$$

When $T = M$,the infimum in (6.6.3) will occur at $u = v = x$, and (6.6.3) reduces to the well known result [218,579,615]

$$\tau^{\wedge}_{M,L}(F^{\wedge}, G^{\wedge})(x) = L(F^{\wedge}(x), G^{\wedge}(x)). \tag{6.6.4}$$

This means that interval arithmetic on the level sets can be used to calculate the sup-min convolution of fuzzy numbers. Note that Höhle [390] has given a generalistion of (6.6.3) (when $T = M$) to completely distributive complete lattices.

If we now observe that any Archimedean t-norm can be decomposed in terms of its additive generators (see the previous section), we can develop a simple way to calculate sup-T convolutions in terms of level sets. Firstly we rewrite the condition for the infimum in (6.6.3) as

$$f_T(g_T(u) + g_T(v)) = x \tag{6.6.5}$$

for a given Archimedean T. Assuming that T is strict and thus that $f_T = g_T^{-1}$, we obtain

$$\begin{aligned} & g_T(u) + g_T(v) = g_T(x) \\ \Rightarrow \qquad u \; & = g_T^{-1}(g_T(x) - g_T(v)) \end{aligned}$$

and v is constrained to range over $[x, g_T^{-1}(g_T(x) - g_T(0))] = [x, 1]$. Therefore (6.6.3) can be rewritten as

$$\tau^{\wedge}_{T,L}(F^{\wedge}, G^{\wedge})(x) \;=\; \inf_{v \in [x,1]} L(F^{\wedge}(g_T^{-1}(g_T(x) - g_T(v))), G^{\wedge}(v)). \tag{6.6.6}$$

6.6.2 Decreasing Part

Considering the decreasing part, we have

$$\mu^d_Z(z) \;=\; \sup_{L(x,y)=z} T(\mu^d_X(x), \mu^d_Y(y)) \tag{6.6.7}$$

and so

$$\begin{aligned} 1 - \mu^D_Z(z) \; & = \; \sup_{L(x,y)=z} T(1 - \mu^D_X(x), 1 - \mu^D_Y(y)) \\ \Rightarrow \qquad \mu^D_Z(z) \; & = \; 1 - \sup_{L(x,y)=z} T(1 - \mu^D_X(x), 1 - \mu^D_Y(y)). \end{aligned}$$

Let $S(x, y) = 1 - T(1 - x, 1 - y)$ be the *t-conorm* corresponding to T. We can thus write

$$\mu^D_Z(z) \;=\; \inf_{L(x,y)=z} S(\mu^D_X(x), \mu^D_Y(y)). \tag{6.6.8}$$

Frank and Schweizer [278] have shown that if

$$\tau_{S,L}(F, G)(x) = \inf_{L(u,v)=x} S(F(u), G(v)) \tag{6.6.9}$$

then

$$\tau^{\wedge}_{S,L}(F^{\wedge}, G^{\wedge})(x) = \sup_{S(u,v)=x} L(F^{\wedge}(u), G^{\wedge}(v)). \tag{6.6.10}$$

A t-conorm S derived from an Archimedean t-norm T can be decomposed by

$$S(x,y) = f^*_T(g^*_T(x) + g^*_T(y)) \tag{6.6.11}$$

where $f^*_T = 1 - f_T(x)$ and $g^*_T(x) = g_T(1-x)$. The condition $S(u,v) = x$ in (6.6.10) can be rewritten as

$$\begin{aligned} S(u,v) &= 1 - g_T^{-1}(g_T(1-u) + g_T(1-v)) = x \\ \Rightarrow \quad g_T(1-u) &+ g_T(1-v) = g_T(1-x) \\ \Rightarrow \quad u &= 1 - g_T^{-1}(g_T(1-x) - g_T(1-v)). \end{aligned}$$

Therefore

$$\tau^{\wedge}_{S,L}(F^{\wedge}, G^{\wedge})(x) = \sup_{v\in[0,z]} L(F^{\wedge}(1 - g_T^{-1}(g_T(1-x) - g_T(1-v))), G^{\wedge}(v)). \tag{6.6.12}$$

Using (6.6.6) and (6.6.12) and changing back to the fuzzy set notation we can write

$$\mu^{I\wedge}_Z(z) = \inf_{v\in[z,1]} L(\mu^{I\wedge}_X(g_T^{-1}(g_T(z) - g_T(v))), \mu^{I\wedge}_Y(v)) \tag{6.6.13}$$

and

$$\mu^{D\wedge}_Z(z) = \sup_{v\in[0,z]} L(\mu^{D\wedge}_X(1 - g_T^{-1}(g_T(1-z) - g_T(1-v))), \mu^{D\wedge}_Y(v)). \tag{6.6.14}$$

Formulae (6.6.13) and (6.6.14) provide a general and fast way of calculating sup-T convolutions for arbitrary Archimedean t-norms in terms of the level sets of the membership functions of the fuzzy variables in question. Note that for operations such as subtraction and division, the formula for $\mu^{I\wedge}_Z$ will be in terms of $\mu^{I\wedge}_X$ and $\mu^{D\wedge}_Y$, and that for $\mu^{D\wedge}_Z$ will be in terms of $\mu^{D\wedge}_X$ and $\mu^{I\wedge}_Y$. (Subtraction is thought of as negation of the right argument followed by addition.)

6.6.3 Discretisation

In order to calculate (6.6.13) and (6.6.14) numerically, we discretise μ_X, μ_Y and thus μ_Z. We will set $\mu^{I[\wedge]}_X[i] = \mu^{I\wedge}_X(\frac{i}{P})$ for $i = 0, \ldots, P$ (*i.e.* $P+1$ points). In fact we will adopt an approach similar to that used in [874] where we used lower and upper approximations. If μ is known exactly, then $\mu = \underline{\mu} = \overline{\mu}$. We thus have

$$\tilde{\mu}^{\wedge}(x) = \overline{\mu}^{\wedge}(\tfrac{i}{P}) \qquad \forall x \in \left[\tfrac{i}{P}, \tfrac{i+1}{P}\right) \tag{6.6.15}$$

$$\underset{\sim}{\mu}^{\wedge}(x) = \underline{\mu}^{\wedge}(\tfrac{i}{P}) \qquad \forall x \in \left(\tfrac{i-1}{P}, \tfrac{i}{P}\right] \tag{6.6.16}$$

for $x \in [0,1]$ and $i = 0, \ldots, P$. The advantage of this representation is that "directed rounding" can be used in order to ensure that the true result is contained within the

interval $[\underset{\sim}{\mu}_Z(z), \tilde{\mu}_Z(z)]$. This can be seen in equations (6.6.18) and (6.6.19) below. We define the array $g_T^{[]}$ by

$$g_T^{[]}[i] = g_T(i/P) \qquad i = 0, \ldots, P \tag{6.6.17}$$

and can then write the formulae

$$\mu_Z^{I[\wedge]}[i] = \inf_{j=i,\ldots,P} L(\mu_X^{I[\wedge]}[\langle P \times g_T^{-1}(g_T^{[]}[i] - g_T^{[]}[j])\rangle], \mu_Y^{I[\wedge]}[j]) \tag{6.6.18}$$

and

$$\mu_Z^{D[\wedge]}[i] = \sup_{j=0,\ldots,i} L(\mu_X^{D[\wedge]}[\langle P \times (1 - g_T^{-1}(g_T^{[]}[P-i] - g_T^{[]}[P-j]))\rangle], \mu_Y^{D[\wedge]}[j]) \tag{6.6.19}$$

for $i = 0, \ldots, P$, where

$$\begin{aligned} \langle\cdot\rangle &= \lfloor\cdot\rfloor \quad \text{for } \tilde{\mu}_Z^{I[\wedge]}, \\ \langle\cdot\rangle &= \lceil\cdot\rceil \quad \text{for } \underset{\sim}{\mu}_Z^{I[\wedge]} \end{aligned}$$

and the $\mu_X^{I[\wedge]}$, $\mu_Y^{I[\wedge]}$, $\mu_X^{D[\wedge]}$ and $\mu_Y^{D[\wedge]}$ are lower or upper approximations as appropriate (this depends on L). Note that the rounding is *down*wards for the upper approximations because we are dealing with the inverse membership function (or the level sets). Some examples calculated using (6.6.18) and (6.6.19) are presented below.

Chapter 7

The Inverse and Determinant of a 2×2 Uniformly Distributed Random Matrix

7.1 Introduction

This chapter presents two results on 2×2 matrices with iid (independent identically distributed) uniformly distributed elements. Theorem 1 gives an expression for the density of the determinant of a matrix whose elements are iid uniformly on $[0, 1]$. Theorem 2 gives an expression for the density of the elements of the inverse of this matrix. Both of these results are derived in a straightforward manner, but they do not seem to have appeared in the literature previously. The only hint of such a result which the author has found is a single sentence on page 199 of the book by Prohorov and Rozanov [665]:

> The distribution of [the determinant of a random matrix] is known only in two cases: if the [column vectors] are uniformly distributed on ℓ-dimensional unit sphere of the ℓ-dimensional space and if the [column vectors] are normally distributed with vanishing mean vector and non-degenerate correlation matrix.

We make use of the convolution relations for the difference, product and quotient of two random variables [632]. The original motivation for the results derived here was to construct an explicit example to show the misleading nature of Szulc's [811] definition of "almost everywhere non-singular matrices" (see also [812]). This chapter has been published in *Statistics and Probability Letters*, **7**, 167–170, (1989) under the same title as this chapter.

7.2 Results

Theorem 7.2.1 *Let*

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

be a 2×2 *matrix with elements* a_{ij} $(i, j = 1, 2)$*, where* a_{ij} *are iid random variables with density*

$$f_a(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & \textit{elsewhere,} \end{cases}$$

and let $D = \det A$*. If* $f_D(x)$ *is the probability density of* D *then*

$$f_D(x) = \begin{cases} (x+1)(2 - \log(x+1)) + x \left[-\log(-x) + \displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \right. & \\ \quad \left. \times \left(\displaystyle\sum_{i=1}^{k} \binom{k}{i} \frac{(x-1)^{k-i}((-x)^i - 1)}{i} + (x-1)^k \log(-x) \right) \right] & x \in [-1, 0), \\ f_D(-x) & x \in (0, 1], \\ 0 & \textit{elsewhere.} \end{cases} \tag{7.2.1}$$

PROOF. Let $p = a_{11}a_{22}$, $q = a_{21}a_{12}$ and so $D = p - q$. Obviously p and q are iid with density

$$\begin{aligned} f_q(z) = f_p(z) &= \int_{-\infty}^{\infty} \frac{1}{|x|} f_x(z/x) f_a(x)\, dx \\ &= -\log z \qquad z \in (0, 1]. \end{aligned}$$

Now

$$\begin{aligned} f_D(x) &= \int_{-\infty}^{\infty} f_p(w + x) f_q(w)\, dw \\ &= \begin{cases} \displaystyle\int_{-x}^{1} \log(w + x) \log(w)\, dw & x \in [-1, 0), \\ \displaystyle\int_{0}^{1-x} \log(w + x) \log(w)\, dw & x \in (0, 1], \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

The two cases are symmetric ($f_D(-x) = f_D(x)$) and so we will only consider $x \in [-1, 0)$. Let

$$I_x(w) = \int \log(w + x) \log(w)\, dw \qquad x \in [-1, 0).$$

Integration by parts twice gives

$$I_x(w) = w(2 - \log w) + (w + x) \log(w + x)(\log(w) - 1) + x(1 - K_x(w))$$

where

$$K_x(w) = \int \frac{\log(w + x)}{w}\, dw$$

Figure 7.1: The probability density f_D.

and cannot be expressed as a finite combination of elementary functions [336, eq. 2.7282]. We can determine a series expansion though:

$$\frac{\log(w+x)}{w} = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(w+x-1)^k}{wk} \qquad 0 < w + x \leq 2.$$

This can be integrated termwise within the region of convergence and so

$$K_x(w) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \int \frac{(w+x-1)^k}{wk}\, dw,$$

where the integral can only be over a range such that $0 < w + x \leq 2$. Using a binomial expansion of the integrand and then integrating termwise we obtain

$$K_x(w) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left[\sum_{i=1}^{k} \binom{k}{i} (x-1)^{k-i}\, \frac{w^i}{i} + (x-1)^k \log\, w \right].$$

Evaluating $I_x(w)|_{w=-x}^{1}$ gives (7.2.1). A graph of $f_D(x)$ calculated using only the first 30 terms of the infinite series is shown in figure 7.1. ■

Theorem 7.2.2 *Let A be as in theorem 7.2.1 and let*

$$A^{-1} = \begin{bmatrix} \frac{a_{22}}{a_{11}a_{22}-a_{12}a_{21}} & \frac{-a_{12}}{a_{11}a_{22}-a_{12}a_{21}} \\ \frac{-a_{21}}{a_{11}a_{22}-a_{12}a_{21}} & \frac{a_{11}}{a_{11}a_{22}-a_{12}a_{21}} \end{bmatrix} = \begin{bmatrix} p & q \\ r & s \end{bmatrix}.$$

Then p, q, r and s are identically distributed (although not independent) with common density function $f_s(x)$ given by

$$f_s(x) = \begin{cases} \dfrac{1}{x^2} + \dfrac{1}{4x(1-x)} + \dfrac{3 - 2\log(-1/x)}{4x^3} & x \in (-\infty, -1), \\ \dfrac{1}{4(1-x)} & x \in [-1, 0), \\ 0 & x \in [0, 1), \\ \dfrac{(x-1)(3 - 2\log(1 - x^{-1}))}{4x^3} & x \in [1, \infty). \end{cases} \tag{7.2.2}$$

PROOF. The fact that p, q, r and s are identically distributed is obvious due to the interchangability of the a_{ij} (they are iid). We rewrite the expression for s as

$$s = \frac{1}{a_{22} - \frac{a_{12}a_{21}}{a_{11}}} = \frac{1}{b - \frac{c}{d}} = \frac{1}{b - e} = \frac{1}{g}$$

and then successively determine the densities of c, e, g, and s. We already know (from the proof of theorem 7.2.1) that

$$f_c(x) = \begin{cases} \log\, x & x \in (0, 1], \\ 0 & \text{elsewhere.} \end{cases}$$

The densities of e and g are determined by the convolution relations for quotient and difference (see [632]). We obtain

$$f_e(x) = \begin{cases} 0 & x \in (-\infty, 0], \\ \dfrac{1}{4} - \dfrac{\log\, x}{2} & x \in (0, 1], \\ \dfrac{1}{4x^2} & x \in (1, \infty), \end{cases}$$

and

$$f_g(x) = \begin{cases} \dfrac{1}{4(x-1)} - \dfrac{1}{4x} & x \in (-\infty, -1), \\ 1 + \dfrac{1}{4(x-1)} + \dfrac{3x - 2x\log(-x)}{4} & x \in [-1, 0), \\ \dfrac{3(1-x) - 2(1-x)\log(1-x)}{4} & x \in [0, 1], \\ 0 & x \in (1, \infty). \end{cases} \tag{7.2.3}$$

The distribution of $s = 1/g$ is given by

$$f_s(x) = \frac{1}{x^2} f_g(1/x). \tag{7.2.4}$$

Substituting (7.2.3) into (7.2.4) gives (7.2.2). A graph of $f_s(x)$ is presented in figure 7.2. ■

Figure 7.2: The probability density f_s.

7.3 Discussion

Observe that $\lim_{|x|\to\infty} f_s(|x|) = 0$, and so A is almost surely non-singular in the conventional measure theoretic sense. The matrix A would not be considered to be almost everywhere nonsingular using Szulc's definition though. Szulc [811,812] is concerned with interval matrices, and defines an interval matrix A^I to be almost everywhere non-singular if there exists only a finite number of real singular matrices contained in A^I. The interpretation of an interval matrix as being a random matrix with uniformly distributed elements seems perfectly natural, and so Szulc's definition is misleading. Komlos [471] has studied the singularity of random matrices and has shown that if $\xi_{i,j}$ $(i,j = 1,2,\ldots)$ are iid with a non-degenerate distribution, then

$$\lim_{n\to\infty} P\left[\begin{vmatrix} \xi_{1,1} & \xi_{1,2} & \cdots & \xi_{1,n} \\ \xi_{2,1} & \xi_{2,2} & \cdots & \xi_{2,n} \\ \vdots & & \ddots & \vdots \\ \xi_{n,1} & \xi_{n,2} & \cdots & \xi_{n,n} \end{vmatrix} = 0\right] = 0.$$

He also settled [470] the problem of the number of singular matrices there are when the elements can be either 0 or 1.

There are very few exact results on the distribution of random determinants. A number of authors have derived results in terms of moments for a few special cases [63,273,626,664,726]. Nyquist, Rice, and Riordan [626] considered the determinants

of random matrices with independent identically normally distributed elements with zero means and derived exact densities for the 2×2, 3×3, and 4×4 cases. They also investigated the moments of the determinant in the case of nonnormally distributed elements with zero means. Nicholson [616] studied the 2×2 case for independent identically distributed elements with normal distributions having nonzero means and derived a complicated infinite series for the cumulative distribution function of the determinant. Alagar [16] has derived an expression for the density of the absolute value of the determinant of a 2×2 matrix with independent exponentially distributed elements. It is in terms of $\psi(x) = \frac{d}{dx}(\log\ \Gamma(x))$. He also presented a complex result for the 3×3 determinant in terms of G-functions.

Instead of calculating the marginal densities we could calculate the joint density. While this has a fairly simple analytical form (and can be calculated directly for the special case we are interested in or from the general result of Fox and Quirk [275]), it is of limited value. Usually one will ultimately be interested in the densities of the individual elements, and to obtain these from the joint density it is necessary to perform multiple integrations with variable limits. The joint density itself is hard to conceptualise (it is four dimensional). Note that we cannot calculate correlation coefficients to determine the dependence of the elements as none of the moments of f_s exist.

Chapter 8

Conclusions

I hate quotation. Tell me what you know.
— Ralph Waldo Emerson

8.1 Overview

The main objectives of this thesis were to study the idea of probabilistic arithmetic, to identify the major problems, and to attempt to provide solutions to some of these problems. It soon became clear that the greatest obstacle in the path to the development of probabilistic arithmetic is the phenomenon of dependency error. It is not sufficient that all the inputs to a calculation are independent since dependencies can arise in the course of a calculation because of repeated occurrences of terms. The details of how the probability distributions are represented, measured, or the σ-convolutions calculated, are of much less significance.

We have developed one approach to the handling of stochastic dependencies in the form of dependency bounds. These allow the calculation of lower and upper bounds on the distribution of a function of random variables when only the marginal distributions are known. We have seen that when some information is known (a lower bound on their copula greater than W), then tighter bounds can be determined. However it should be realised that these bounds do *not* provide a complete solution to the problem of dependency error in probabilistic arithmetic calculations. There are many other issues including the possible rearrangement of expressions in order to remove repeated terms, which have greater potential value in this regard.

Probabilistic arithmetic is similar in its motivations and problems to interval arithmetic. In fact interval arithmetic can be considered as a special case of probabilistic arithmetic. However interval arithmetic is considerably simpler because dependency error manifests itself as "dependency width," which simply causes a wider final result than could otherwise have been obtained. (Compare this with the situation where probability distributions are involved: there is no analogous simple idea of "containment" of results.) In order to achieve a similar situation with probability distributions we have to somehow incorporate the notion of *containment* (one probability distribution containing another). This is the advantage

of our lower and upper bounds on probability distributions. Whilst simple problems (for example determining the distribution of the root of a linear equation by calculating the distribution of the quotient of the two coefficients), are readily solvable using the methods presented in this thesis, more complex problems require further investigation.

Nevertheless there are a number of positive results to come out of this work. The dependency bounds and the numerical methods for calculating them seem to be of independent interest. These bounds can be used to determine the robustness of untestable independence assumptions. The connexion between these dependency bounds and the fuzzy number sup-T convolutions is also of particular interest. Not only does this provide (via our numerical method based on the duality theorem) an efficient way of calculating operations on fuzzy numbers, but it also gives a very close link between fuzzy set theory and probability theory. Whilst the connexion between the Boole-Fréchet formulae for conjunction and disjunction and the fuzzy set intersection and union operations was known before, the knowledge that this can be extended to the case of random and fuzzy variables is new, and further strengthens the argument that to a large extent the fuzzy set theory operations are better considered in terms of probability theory.

Other contributions of this thesis include

- A detailed review of methods for numerically calculating σ-convolutions of probability distribution functions.
- The new σ_L-convolution algorithm developed in chapter 3: Whilst this does not solve all the problems of probabilistic arithmetic, it is still a useful and necessary tool. Our method is computationally efficient, accurate, and simple. It is better than any of the methods discussed in chapter 2.
- The extreme limit theorem derived in chapter 5: This shows that when one has no knowledge of the dependence structure of a set of random variables, almost any distribution consistent with the constraints imposed by interval arithmetic on their supports is possible for their weighted sum.
- The result in chapter 7 on the inverse and determinant of a 2×2 random matrix, although simple, is new and demonstrates the strange results obtainable with just a few operations on random variables.

Many of the issues raised in this thesis are suitable topics for future research. In fact the author already has partial results on some of these which are not included in this thesis. The following section gives a list of possible directions for further investigation.

8.2 Directions for Future Research

There are many areas for further research suggested by the work presented in this thesis. We will very briefly mention some of these. Some of the items below (espe-

cially items 1, 4, 5 and 12) have already been studied to an extent by the author. Results will be reported elsewhere.

1. **Confidence Interval Procedures.** The numerical representation of probability distributions adopted in chapter 3 (lower and upper bounds on a distribution function) obviously suggests the use of confidence intervals for acquiring sample distributions. This would allow a consistent representation from measurement to the final result. There are a number of results available in the literature concerning confidence intervals for quantiles. Some of the questions still to be answered are: Should the confidence intervals be overall (for the entire distribution) or should they be constructed point by point?; What distributional assumptions are necessary?; Can a fast algorithm be developed for implementing the procedure chosen?; and, Given two estimated distributions with certain confidence levels, if the two distributions are combined, is there anything that can be said about the confidence of the result. (This last question lies at the heart of debates on the foundations of statistics.) We already have some answers to some of these questions and we hope to report these elsewhere.

2. **Dependency Bounds for Functions of More than 2 Arguments.** As we mentioned in section 5.2, it is possible to use linear programming techniques to calculate Boole-Fréchet bounds on the probability of complex compound events. When there are repeated terms these bounds are tighter than those obtained by repeated application of the pairwise bounds for conjunction and disjunction. Can similar techniques be applied to the determination of dependency bounds for more complex functions of random variables? This is obviously a much harder problem for random variables than it is for random events.

3. **Rearrangement Methods for Convolutions.** The σ_L-convolution algorithm developed in chapter 3 makes essential use of *sorting.* Is there a continuous analogue of this algorithm? Note that the continuous analogue of sorting a discrete function into monotonic order is an *equimeasurable rearrangement* [143]. Rearrangement techniques have been recently applied to some statistical optimization problems [696], and it seems likely that further results can be obtained using these ideas. In other words it may be possible to develop results for convolutions in terms of rearrangements of distribution functions or probability densities.

4. **Bucketing Algorithms for σ-convolutions.** The σ-convolution algorithm we developed has a computational complexity of $O(N^2 \log N)$ where N is the number of points used to represent the distribution functions. This means that it runs rather slower than the algorithm for dependency bounds, which is $O(N^2)$ (but see item 5 below). It is possible to construct an algorithm for σ-convolutions which has *average* case complexity $O(N^2)$ by using *bucketing algorithms* [209]. It would appear that this can be further improved by using adaptive recursive bucketing. This would also provide a solution to a computer

science problem, known as sorting and selection in multisets [281], which is very similar to the problem of calculating σ-convolutions.

5. **Discrete T-conjugate Transforms.** The T-conjugate transform used in chapter 5 to prove the extremal limit result can be converted into a discrete form. Furthermore there exists a fast divide and conquer algorithm for calculating this. This allows the calculation of the T-log-concave envelopes of $\tau_{T,+}$ operations in $O(N \log N)$ time. A bilateral version could be used for the very fast calculation of the sup-T convolutions of fuzzy number membership functions. This has yet to be implemented in the form of a computer program.

6. **Application of Interval Arithmetic Algorithms.** There exists a wide range of special algorithms developed for interval arithmetic [17,584]. It is possible that some of these may be of use for probabilistic arithmetic calculations. They were developed in order to reduce the amount of dependency width incurred when calculating things like the inverses of interval matrices using Gaussian elimination. In order to use these algorithms it will be necessary to develop a means whereby different lower bounds on copulas are used in the calculation of the dependency bounds so that the lower and upper bounds on the distribution functions do not diverge too far. (See the next item.)

7. **Further Study of Dependency Bounds for $\underline{C}_{XY} \neq W$.** The dependency bounds studied in section 3.5 deserve further attention. A first step would be the implementation of equation 6.4.2 as a fast means of calculating dependency bounds when $\underline{C}_{XY} \neq W$. We have had no experience yet with the updating of copulas in the manner suggested in section 3.5. The relationship between the dependencies induced by these lower bounds and other types of dependence also merits further investigation.

8. **Confidence Curve Procedures.** The confidence curves discussed in section 4.5.4 would appear to deserve further consideration. As mentioned in section 4.5.4, there is a possibility of confidence curves providing a link between Bayesian style and confidence interval style procedures in statistics. Their interpretation along the lines of fuzzy numbers also requires a closer look. Another aspect is the direct combination of confidence curves: could they be used (instead of distribution functions) as representations of random quantities?

9. **Measures of Dependence Based on Copulas and Dependency Bounds.** As we mentioned in section 3.5.4, there are a number of measures or indices of dependence which can be defined in terms of copulas. These have been investigated by Wolff and others [719]. The question is whether they can be used to provide tighter dependency bounds on distributions of functions of random variables. The effects on measures of dependence of operations on random variables also deserves investigation.

10. **Empirical Copulas.** Very recently Quesada-Molina [666] has suggested the notion of an empirical copula. Whilst this is obviously little more than a transformation of the empirical joint distribution function, its statistical properties

require investigation. Given the effect of lower bounds on copulas on the tightness of dependency bounds, it would also be worthwhile investigating one-sided lower confidence bounds on the empirical copula.

11. **Empirical Multiplicative Generators of Archimedean Copulas.** A related idea is to use the multiplicative generator representation of an Archimedean copula and thus to estimate the multiplicative generator. This is a one dimensional function rather than two dimensional. Alsina and Ger's result [25] on the convergence of t-norms in terms of convergence of their generators may be of use here.

12. **History of Quotient Normal Random Variables.** In the course of our investigations of probabilistic arithmetic we have made a study of the history of the distribution of the quotient of two normal random variables. This is obviously the sort of problem we expect our numerical methods to be able to calculate. The history of the analytical attempts on this problem, which begins with the first determination of the distribution of a quotient of random variables by Crofton in [173, p.340], and continues to the present day with a number of rediscoveries, is another interesting aspect we aim to write up one day.

13. **Application to Random Equations and Practical Problems.** As we mentioned above, the probabilistic arithmetic methods developed in this thesis have so far only been studied in terms of their foundations. It remains to be seen whether they are of value for practical problems.

14. **Use of Graph Representations.** The use of graph representations was discussed in section 4.3. There it was noted that further research is needed to determine whether these methods will be of use in probabilistic arithmetic. We expect that this will be the case if complex calculations are attempted. The first step would be to implement a simple expression rearrangement procedure.

15. **Relationship with Fine's Interval Valued Probabilities.** Although our lower and upper bounds on probability distributions can be completely understood in terms of the standard single valued notion of probability, it seems worthwhile examining whether they can be usefully integrated with Fine's interval valued probabilities (section 4.4).

16. **Relationship with Probabilistic Metric Spaces.** Nearly all of the mathematical tools we have used in this thesis were originally developed in the field of probabilistic metric spaces. We have already discussed the relationship between our methods and those of that field (section 4.6). However we suspect that there may be further results obtainable by studying this, particularly from the viewpoint of considering the $\tau_{T,L}$ and $\rho_{T,L}$ operations as dependency bounds rather than just as triangle functions.

17. **Use of Mixtures for Nonmonotonic Operations.** Further work is needed to determine the practical value of the use of mixtures advocated in section 3.6 for calculating both convolutions and dependency bounds under nonmonotonic

operations on random variables. Whilst analytically there are no difficulties, there still remain some numerical problems to be solved.

8.3 Some General Philosophical Conclusions

> *Phenomena that are statistically calculable do not become statistically incalculable suddenly, at a well-defined boundary, but, rather, by degrees. The scholar takes a position of cognitive optimism; that is, he assumes that the subjects he studies will yield to calculation. It is nicest if they do so deterministically .... It is not quite so nice if calculable probability has to substitute for certainty. But it is not nice at all when absolutely nothing can be calculated.*
>
> — Stansilaw Lem

The work presented in this thesis can be considered to be just part of a general long term trend to constructing probabilistic models of reality. General discussions on the probabilistic view of the world can be found in [803,806]. This "probabilisitic imperative" forms the background to a large portion of recent science and philosophy of science. Popper has presented a number of arguments against determinism [659], and the deterministic world view is (or at least *should* be) considered dead.

However, having adopted a probabilistic world view, problems arise that were absent in the simpler deterministic situation. For example, there is the problem of acquisition of probability values. This problem, seemingly innocuous at first, leads one through the tortuous maze of interpretations of probability. Nearly 200 years after Laplace, there is still no sign of consensus in the scientific community about this. Thus it becomes necessary for us to try and make up our own minds. Whilst we have managed to avoid the problem in this thesis, let us just mention now our preference for the propensity interpretation [479,654,655] with the rider that we are aware of many remaining problems with this.

Apart from these philosophical concerns, there is the practical problem of how one proceeds to calculate with the probabilities. We have already seen (section 4.6) the difference between our viewpoint (based on random variables) and that adopted by the majority of workers in the field of probabilistic metric spaces. Even ignoring this distinction, there remain the severe practical difficulties of solving even some of the most simply stated probabilistic problems. For example, whilst there has been some encouraging progress recently in the study of polynomials with random coefficients through the use of Kharitonov's theorem [55,455], the situation remains much the same as it did in 1956 when Hammersley confessed he was "still very far from having solved the practical problem" of determining the distribution of the roots of a random polynomial.

The point is that there is a very deep hierarchy of problems in terms of their practical difficulty. This difficulty is not the "in principle" difficulty which determinists were always sure could be removed by being clever enough. Nowadays we can state the difficulty more precisely in terms of computational complexity theory. Many problems are "insoluble" because of their exponential computational complexity. What we have seen in this thesis is a further example of this. The exact calculation of the distribution of complex functions of random variables is generally intractable.

However it *is* possible to calculate lower and upper bounds on the solution by use of the dependency bounds. This approach seems preferable to the alternative of invoking the principle of maximum entropy and assuming independence solely in order to get a single valued probability for a result. Nevertheless, randomness ultimately wins, and for the time being at least, Monte-Carlo simulation will remain the major tool for solving complex random problems.

Chapter 9

References

He has overlooked no work in his field, regardless of how barbarously or tediously written. He has expounded and criticised each work ... Do you prefer to call this madness or folly? It does not matter to me; merely confess that it is done with my assistance ....
— "Folly" (Erasmus of Rotterdam)

[1] ———, "A Variant of the CESTAC Method and its Application to Constrained Optimization," *Mathematics and Computers in Simulation* **30** (1988), .

[2] M. H. Ackroyd, "Computing the Waiting Time Distribution for the G/G/1 Queue by Signal Processing Methods," *IEEE Transactions on Communications* **28** (1980), .

[3] ———, "Iterative Computation of the M/G/1 Queue Length Distribution via the Discrete Fourier Transform," *IEEE Transactions on Communications* **28** (1980), .

[4] ———, "M/M/1 Transient State Occupancy Probabilities via the Discrete Fourier Transform," *IEEE Transactions on Communications* **30** (1982), .

[5] ———, "M/G/1 Queue Analysis via the Inverse Poisson Transform," *IEE Proceedings* **129E** (1982), .

[6] M. H. Ackroyd & R. Kanyangarara, "Skinner's Method for Computing Bounds on Distributions and the Numerical Solution of Continuous-Time Queueing Problems," *IEEE Transactions on Communications* **30** (1982), .

[7] E. W. Adams & H. P. Levine, "On the Uncertainties Transmitted from Premises to Conclusions in Deductive Inferences," *Synthese* **30** (1975), .

[8] G. Adomian, *Nonlinear Stochastic Operator Equations*, Academic Press, New York, 1986.

[9] A. M. Agogino, "AI in Computer-Aided Design: Qualitative Reasoning and Symbolic Computation," in *The Study of the Design Process*, M. B. Waldron, ed., Proceedings of the NSF Workshop in Design Theory and Methodology, Oakland, California, February 8–10, 1987.

[10] A. M. Agogino & A. Rege, "IDES: Influence Diagram Based Expert Systems," *Mathematical Modelling* **8** (1987), .

[11] R. Ahmad, "A Distribution-Free Interval Mathematics Analysis of Probability Density Functions," in *Interval Mathematics (Proceedings of the International Symposium Karlsruhe, West Germany, May 20–24, 1975)*, K. Nickel, ed., Springer-Verlag, Berlin, 1975, .

[12] S. Ahmed, R. E. Clark & D. R. Metcalf, "A Method for Propagating Uncertainty in Probabilistic Risk Assessment," *Nuclear Technology* **59** (1982), .

[13] A. V. Aho & S. C. Johnson, "Optimal Code Generation for Expression Trees," *Journal of the Association for Computing Machinery* **23** (1976), .

[14] A. V. Aho, S. C. Johnson & J. D. Ullman, "Code Generation for Expressions with Common Subexpressions," in *Conference Record of the 3rd ACM Symposium on the Principles of Programming Languages*, Association for Computing Machinery, 1976, .

[15] A. V. Aho & J. D. Ullman, "Optimization of Straight Line Programs," *SIAM Journal of Computing* **1** (1972), .

[16] V. S. Alagar, "The Distributions of Random Determinants," *The Canadian Journal of Statistics* **6** (1978), .

[17] G. Alefeld & J. Herzberger, *Introduction to Interval Computations*, Academic Press, New York, 1983.

[18] D. W. Allan, "Should the Classical Variance be Used as a Basic Measure in Standards Metrology?," *IEEE Transactions on Instrumentation and Measurement* **36** (1987), .

[19] R. N. Allan, C. H. Grigg & M. R. G. Al-Shakarchi, "Numerical Techniques in Probabilistic Load Flow Problems," *International Journal of Numerical Methods in Engineering* **10** (1976), .

[20] R. N. Allan, A. M. Leite, A. A. Abu-Nasser & R. C. Burchett, "Discrete Convolution in Power System Reliability," *IEEE Transactions on Reliability* **30** (1981), .

[21] N. Alliot, "Data Error Analysis in Unconstrained Optimization Problems with the CESTAC Method," *Mathematics and Computers in Simulation* **30** (1988), .

[22] A. Allisy, "Some Statistical Methods Used in Metrology," in *Metrology and Fundamental Constants (Proceedings of the International School of Physics "Enrico Fermi" Course LXVIII, Varenna on Lake Como, Villa Monastero, 12th-24th July 1976)*, A. F. Milone, P. Giacomo & S. Leschiutta, eds., North-Holland / Italian Physical Society, Amsterdam, 1980, .

[23] C. Alsina, "Some Functional Equations in the Space of Uniform Distribution Functions," *Aequationes Mathematicae* **22** (1981), .

[24] ———, "On A Family of Connectives for Fuzzy Sets," *Fuzzy Sets and Systems* **16** (1985), .

[25] C. Alsina & R. Ger, "Associative Operations Close to a Given One," *Comptes Rendus Mathematical Reports of the Academy of Science Canada* **7** (1985), .

[26] C. Alsina, E. Trillas & L. Valverde, "On Some Logical Connectives for Fuzzy Sets Theory," *Journal of Mathematical Analysis and Applications* **93** (1983), .

[27] D. F. Alwin & R. M. Hauser, "The Decomposition of Effects in Path Analysis," *American Sociological Review* **40** (1975), .

[28] D. Angluin, "How to Prove it," *SIGACT News (Association for Computing Machinery Special Interest Group on Automata and Computability Theory)* **15(1)** (1983), .

[29] N. Apostolatos, U. Kulisch, R. Krawczyk, B. Lortz, K. Nickel & H. -W. Wipperman, "The Algorithmic Language Triplex-Algol 60," *Numerische Mathematik* **11** (1968), .

[30] L. Appelbaum & E. H. Ruspini, "Aries: An Approximate Reasoning Inference Engine," in *Approximate Reasoning in Expert Systems*, M. M. Gupta, A. Kandel & J. B. Kiszka, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1985, .

[31] M. A. Arbib, "Review of 'Fuzzy Sets and their applications to cognitive and design processes,' 'Introduction to the theory of fuzzy subsets, volume 1,' 'Applications of fuzzy sets to systems analysis'," *Bulletin of the American Mathematical Society* **83** (1977), .

[32] M. A. Arbib & E. G. Manes, "Partially Additive Categories and Flow-Diagram Semantics," *Journal of Algebra* **62** (1980), .

[33] S. Arnborg, J. Lagergren & D. Seese, "Problems Easy for Tree-Decomposable Graphs," in *Automata, Languages and Programming (15th International Colloquium)*, Springer-Verlag, Berlin, 1988, , Lecture Notes in Computer Science 317.

[34] S. J. Arnold, "Morphology, Performance and Fitness," *American Zoologist* **23** (1983), .

[35] L. A. Aroian, "The Probability Function of the Product of Two Normally Distributed Variables," *Annals of Mathematical Statistics* **18** (1947), .

[36] R. Askey, "Linearization of the Product of Orthogonal Polynomials," in *Problems in Analysis: A Symposium in Honour of Salomon Bochner*, R. C. Gunning, ed., Princeton U.P., Princeton, N.J., 1970, .

[37] R. Askey & G. Gasper, "Convolution Structures for Laguerre Polynomials," *Journal D'Analyse Mathématique* **31** (1977), .

[38] R. Askey & S. Wainger, "Mean Convergence of Expansions in Laguerre and Hermite Series," *American Journal of Mathematics* **87** (1965), .

[39] ———, "A Convolution Structure for Jacobi Series," *American Journal of Mathematics* **91** (1969), .

[40] R. A. Askey, T. H. Koornwinder & W. Schempp, *Special Functions: Group Theoretical Aspects and Applications*, D. Riedel, Dordrecht, 1984.

[41] R. Badard, "The Law of Large Numbers for Fuzzy Processes and the Estimation Problem," *Information Sciences* **28** (1982), .

[42] R. Baekeland & E. E. Kerre, "Piecewise Linear Fuzzy Quantities: A way to Implement Fuzzy Information into Expert Systems and Fuzzy Databases," in *Uncertainty and Intelligent Systems (Lecture Notes in Computer Science 313)*, B. Bouchon, L. Saitta & R. R. Yager, eds., Springer-Verlag, Berlin, 1988, .

[43] R. A. Bagnold, "The Nature and Correlation of Random Distributions," *Proceedings of the Royal Society of London* **A338** (1983), .

[44] W. N. Bailey, "On the Product of Two Laguerre Polynomials," *Quarterly Journal of Mathematics (Oxford)* **10** (1939), .

[45] S. D. Bajpai, "An Integral Involving Fox's H-function and Whittaker Functions," *Proceedings of the Cambridge Philosophical Society* **65** (1969), .

[46] G. A. Baker, "Transformation of Non-Normal Frequency Distributions into Normal Distributions," *Annals of Mathematical Statistics* **5** (1934), .

[47] J. F. Baldwin, "Fuzzy Sets and Expert Systems," *Information Sciences* **36** (1985), .

[48] W. Bandler & L. J. Kohout, "Probabilistic versus Fuzzy Production Rules in Expert Systems," *International Journal of Man-Machine Studies* **22** (1985), .

[49] N. C. Barford, *Experimental Measurements: Precision, Error and Truth*, Addison-Wesley, London, 1967.

[50] J. L. Barlow & E. H. Bareiss, "On Roundoff Error Distributions in Floating Point and Logarithmic Arithmetic," *Computing* **34** (1985), .

[51] ———, "Probabilistic Error Analysis of Gaussian Elimination in Floating Point and Logarithmic Arithmetic," *Computing* **34** (1985), .

[52] R. E. Barlow, "Using Influence Diagrams," in *Accelerated Life Testing and Experts' Opinions in Reliability*, C. A. Clarotti & D. V. Lindley, eds., Italian Physical Society / North-Holland, Amsterdam, 1988, , Proceedings of the International School of Physics "Enrico Fermi" Course CII.

[53] R. E. Barlow, R. W. Mensing & N. G. Smiriga, "An Application of Influence Diagrams," in *Accelerated Life Testing and Experts' Opinions in Reliability*, C. A. Clarotti & D. V. Lindley, eds., Italian Physical Society / North-Holland, Amsterdam, 1988, , Proceedings of the International School of Physics "Enrico Fermi" Course CII.

[54] R. E. Barlow & X. Zhang, "Bayesian Analysis of Inspection Sampling Procedures Discussed by Deming," *Journal of Statistical Planning and Inference* **16** (1987), .

[55] B. R. Barmish, "A Generalization of Kharitonov's Four Polynomial Concept for Robust Stability Problems with Linearly Dependent Coefficient Perturbations," in *Proceedings of the 1988 American Control Conference*, 1988, .

[56] J. Baron, "Second-Order Probabilities and Belief Functions," *Theory and Decision* **23** (1987), .

[57] W. W. Bartley III, "A Popperian Harvest," in *In Pursuit of Truth: Essays on the Philosophy of Karl Popper on the Occasion of His 80th Birthday*, P. Levinson, ed., Humanities Press Inc., Atlantic Highlands, N.J., 1982, .

[58] B. W. Bateman, "Keynes's Changing Conception of Probability," *Economics and Philosophy* **3** (1987), .

[59] W. D. Baten, "Combining Two Probability Functions," *Annals of Mathematical Statistics* **5** (1934), .

[60] F. L. Bauer, "Computational Graphs and Rounding Error," *SIAM Journal of Numerical Analysis* **11** (1974), .

[61] L. A. Baxter, "Some Remarks on Numerical Convolution," *Communications in Statistics — Simulation and Computation* **B10** (1981), .

[62] R. W. Beals, P. C. Greiner & J. Vauthier, "The Laguerre Calculus on the Heisenberg Group," in *Special Functions: Group Theoretical Aspects and Applications*, R. A. Askey, T. H. Koornwinder & W. Schempp, eds., D. Riedel, Dordrecht, 1984, .

[63] R. Bellman, "A Note on the Mean Value of Random Determinants," *Quarterly of Applied Mathematics* **13** (1955), .

[64] R. Bellman & W. Karush, "On a New Functional Transform in Analysis: The Maximum Transform," *Bulletin of the American Mathematical Society* **67** (1961), .

[65] ———, "Mathematical Programming and the Maximum Transform," *Journal of the Society for Industrial and Applied Mathematics* **10** (1962), .

[66] ———, "On the Maximum Transform and Semigroups of Transformations," *Bulletin of the American Mathematical Society* **68** (1962), .

[67] ———, "On the Maximum Transform," *Journal of Mathematical Analysis and Applications* **6** (1963), .

[68] ———, "Functional Equations in the Theory of Dynamic Programming XII: An Application of the Maximum Transform," *Journal of Mathematical Analysis and Applications* **6** (1963), .

[69] L. S. Belobragina & V. I. Rudomanov, "Equivalent-Perturbation Method in Accuracy Evaluation for Indirect Measurements," *Measurement Techniques* **29** (1986), .

[70] G. L. Bennett, L. A. Swiger, R. L. Preston & V. R. Cahill, "Evaluation of Urea Space and Ultrasonic Measurement as Selectipon Criteria for Beef Animal Composition," *Journal of Animal Science* **54** (1982), .

[71] M. Berres, "On a Multiplication and a Theory of Integration for Belief and Plausibility Functions," *Journal of Mathematical Analysis and Applications* **121** (1987), .

[72] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[73] A. T. Bharucha-Reid, "Random Algebraic Equations," in *Probabilistic Methods in Applied Mathematics II*, A. T. Bharucha-Reid, ed., Academic Press, New York, 1970, .

[74] A. T. Bharucha-Reid & M. Sambandham, *Random Polynomials*, Academic Press, Orlando, 1986.

[75] Bháscara, "Lilávati," in *Algebra and Mensuration from the Sanscrit of Brahmegupta and Bháscara*, H. T. Colebrooke, ed., London, 1817, .

[76] A. Birnbaum, "Confidence Curves: An Omnibus Technique for Estimation and Testing Statistical Hypotheses," *Journal of the American Statistical Association* **56** (1961), .

[77] ———, "A Unified Theory of Estimation I," *Annals of Mathematical Statistics* **32** (1961), .

[78] A. Bjerhammar, *Theory of Errors and Generalized Matrix Inverses*, Elsevier Scientific Publishing Company, Amsterdam, 1973.

[79] J. L. Black, "Identification and Control via the Statistical Properties of Input and Output Signals," PhD thesis, University of Queensland, 1986.

[80] J. L. Black & G. F. Shannon, "A Learning and Control Strategy for Linear and Nonlinear Systems," in *Preprints of SOCOCO'88 (5th IFAC/IFIP symposium on Software for Computer Control, 26th-28th April 1988)*, 1988, separate pagination.

[81] M. Black, "Vagueness: An Exercise in Logical Analysis," *Philosophy of Science* **4** (1937), .

[82] ———, "Reasoning with Loose Concepts," in *Margins of Precision: Essays in Logic and Language*, M. Black, ed., Cornell University Press, Ithaca, 1970, .

[83] R. E. Blahut, *Fast Algorithms for Digital Signal Processing*, Addison-Wesley, Reading, Massachusetts, 1985.

[84] H. M. Blalcock Jr, ed., *Causal Models in the Social Sciences*, Aldine Publishing Company, Chicago, 1971.

[85] P. L. Bogler, "Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems," *IEEE Transactions on Systems, Man, and Cybernetics* **17** (1987), .

[86] H. Bohman, "Numerical Inversion of Characteristic Functions," *Scandnavian Acturial Journal* (1975).

[87] G. W. Bohrnstedt & A. S. Goldberger, "On the Exact Covariance of Products of Random Variables," *Journal of the American Statistical Association* **64** (1969), .

[88] G. W. Bohrnstedt & G. Marwell, "The Reliability of Products of Two Random Variables," in *Sociological Methodology 1978*, K. F. Schuessler, ed., Jossey-Bass, San Francisco, 1977, .

[89] P. Bois & J. Vignes, "A Software for Evaluating Local Accuracy in the Fourier Transform," *Mathematics and Computers in Simulation* **22** (1980), .

[90] ———, "An Algorithm for Automatic Round-Off Error Analysis in Discrete Linear Transforms," *International Journal of Computer Mathematics* **12** (1982), .

[91] L. P. Bolgiano Jr & M. J. Piovoso, "Poisson Transform Signal Analysis," *IEEE Transactions on Information Theory* **14** (1968), .

[92] L. I. Boneva, D. Kendall & I. Stefanov, "Spline Transformations: Three New Diagnostic Aids for the Statistical Data-Analyst (with discussion)," *Journal of the Royal Statistical Society Series B* **33** (1971), .

[93] P. P. Bonissone, "Theories of Plausible Reasoning," in *Reasoning with Incomplete and Uncertain Information*, General Electric Corporate Research and Development, Schenectady, New York, 1987, .

[94] ———, "Summarizing and Propagating Uncertain Information with Triangular Norms," *International Journal of Approximate Reasoning* **1** (1987), .

[95] P. P. Bonissone & K. S. Decker, "Selecting Uncertainty Calculi and Granularity: An Experiment in Trading-off Precision and Complexity," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[96] P. P. Bonissone & R. M. Tong, "Editorial: Reasoning with Uncertainty in Expert Systems," *International Journal of Man-Machine Studies* **22** (1985), .

[97] G. Boole, *An Investigation of the Laws of Thought, on which are founded The Mathematical Theories of Logic and Probabilities*, Walton and Maberley, London, 1854, Republished by Dover Publications Incorporated, New York, 1951.

[98] ———, "On Propositions Numerically Definite," *Transactions of the Cambridge Philosophical Society* **11(II)** (1868), Reprinted as pp.167–186 in R. Rhees, ed., *Studies in Logic and Probability* by George Boole, Watts and Company, London, 1952.

[99] ———, "Reply to Some Observations by Mr. Wilbraham on the Theory of Chances," *The Philosophical Magazine (series 4)* **8** (August 1854), Reprinted as pp.274–279 in R. Rhees, ed., *Studies in Logic and Probability* by George Boole, Watts and Company, London, 1952.

[100] C. de Boor, "Appendix to 'Splines and Histograms' by I.J. Schoenberg," in *Spline Functions and Approximation Theory*, A. Meir & A. Sharma, eds., Birkhäuser, Basel, 1973, .

[101] S. B. Boswell & M. S. Taylor, "A Central Limit Theorem for Fuzzy Random Variables," *Fuzzy Sets and Systems* **24** (1987), .

[102] B. Bouchon, "Fuzzy Inferences and Conditional Possibility Distributions," *Fuzzy Sets and Systems* **23** (1987), .

[103] P. M. Boudet, "Probabilité et logique de l'argumentation selon Jacques Bernoulli," *Les Études Philosophiques* **N.S. 28** (1967), .

[104] K. O. Bowman & L. R. Shenton, "Approximations to Distributions," in *Encyclopedia of Statistical Sciences, Volume 1*, S. Kotz & N. L. Johnson, eds., J. Wiley and Sons, New York, 1982, .

[105] L. Breiman, "Nail Finders, Edifices, and Oz," in *Proceedings of the Berkeley Conference in Honour of Jerzy Neyman and Jack Kiefer, Volume 1*, L. M. L. Cam & R. A. Olshen, eds., Wadsworth Advanced Books and Software, Monterey, California, 1985, .

[106] M. A. Breuer, "Generation of Optimal Code for Expressions via Factorization," *Communications of the Association for Computing Machinery* **12** (1969), .

[107] S. R. Broadbent, "Lognormal Approximation to Products and Quotients," *Biometrika* **43** (1956), .

[108] J. A. Brooke, "Review of 'Probabilistic Metric Spaces' by B. Schweizer and A. Sklar," *Foundations of Physics* **17** (1987), .

[109] R. H. Brown, "The Distribution Function of Positive Definite Quadratic Forms in Normal Random Variables," *SIAM Journal of Scientific and Statistical Computing* **7** (1986), .

[110] C. Brownie & J. Kiefer, "The Ideas of Conditional Confidence in the Simplest Setting," *Communications in Statistics — Theory and Methods* **A6** (1977), .

[111] J. Buckley & W. Siler, "Fuzzy Operators for Possibility Interval Sets," *Fuzzy Sets and Systems* **22** (1987), .

[112] J. J. Buckley, "Managing Uncertainty in a Fuzzy Expert System," *International Journal of Man-Machine Studies* **29** (1988), .

[113] ———, "Possibilistic Linear Programming with Triangular Fuzzy Numbers," *Fuzzy Sets and Systems* **26** (1988), .

[114] ———, "On the Algebra of Interactive Fuzzy Numbers," *Fuzzy Sets and Systems* (1988), To appear.

[115] ———, "On the Algebra of Interactive Fuzzy Numbers: The Continuous Case," *Fuzzy Sets and Systems* (1989), To appear.

[116] J. J. Buckley & H. Ying, "Fuzzy Controller Theory: Limit Theorems for Linear Fuzzy Controller Rules," *Automatica* **25** (1989), .

[117] M. Bunge, "Possibility and Probability," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Volume III*, W. L. Harper & C. A. Hooker, eds., D. Riedel, Dordrecht, 1976, .

[118] ———, *Causality and Modern Science (3rd Revised Edition)*, Dover Publications, New York, 1979.

[119] A. J. Buras & Gaemers, "Simple Parametrizations of Parton Distributions with Q^2 Dependence given by Asymptotic Freedom," *Nuclear Physics* **B132** (1978), .

[120] M. Burke & R. Cytron, "Interprocedural Dependence Analysis and Parallelization," *SIGPLAN Notices* **21** (1986), .

[121] J. C. Burkhill, "Functions of Intervals," *Proceedings of the London Mathematical Society (Series 2)* **22** (1923), .

[122] ———, "Derivatives of Functions of Intervals," *Fundamenta Mathematicae* **5** (1924), .

[123] C.Gini, "Di una misura della dissomiglianza tra due grappi di quantità e delle sue applicazioni allo stadio delle relazioni statistiche," *Att. R. Ist. Venoto Sc. Lettere, Art* **74** (1914), .

[124] N. Cabibbo & R. Petronzio, "Two-Stage Model of Hadron Structure: Parton Distributions and their Q^2 Dependence," *Nuclear Physics* **B137** (1978), .

[125] P. Calabrese, "The Probability of the Triangle Inequality in Probabilistic Metric Spaces," *Aequationes Mathematicae* **18** (1978), .

[126] ———, "An Algebraic Synthesis of the Foundations of Logic and Probability," *Information Sciences* **42** (1987), .

[127] S. Cambanis & G. Simons, "Probability and Expectation Inequalities," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **59** (1982), .

[128] S. Cambanis, G. Simons & W. Stout, "Inequalities for $\mathcal{E}k(X,Y)$ when the Marginals are Fixed," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **36** (1976), .

[129] N. R. Campbell, *An Account of the Principles of Measurement and Calculation*, Longmans, Green and Company Limited, London, 1928, Reprinted by University Microfilms, Ann Arbor, Michigan, 1971.

[130] ———, *Foundations of Science*, Dover, New York, 1957, (Originally *Physics: The Elements*).

[131] M. Caramanis, J. Stremel, W. Fleck & S. Daniel, "Probabilistic Production Costing: An Investigation of Alternative Algorithms," *Electrical Power and Energy Systems* **5** (1983), .

[132] B. D. Carter, "On the Probability Distribution of Rational Functions of Independent H-function Variates," PhD thesis, University of Arkansas, 1970.

[133] B. D. Carter & M. D. Springer, "The Distribution of Products, Quotients, and Powers of Independent H-function Variates," *SIAM Journal of Applied Mathematics* **33** (1977), .

[134] N. Cartwright, "Causal Laws and Effective Strategies," *Noûs* **13** (1979), .

[135] U. Cerruti & U. Höhle, "An Approach to Uncertainty Using Algebras Over a Monoidal Closed Category," *Rendiconti del Circolo Matematico di Palermo (2) — Biblioteca Matematica. Supplement Number 12* (1986), (Proceedings of the Second Topology Conference, Taormina, 1984).

[136] D. A. Chalker & D. Mackerras, "Models for Representing the Acoustic Radiation Impedance of the Mouth," *IEEE Transactions on Acoustics, Speech and Signal Processing* **33** (1985), .

[137] S. Chanas & J. Kamburowski, "The Use of Fuzzy Variables in PERT," *Fuzzy Sets and Systems* **5** (1981), .

[138] W. K. Chang, L. R. Chow & S. K. Chang, "Arithmetic Operations on Level Sets of Convex Fuzzy Numbers," in *Fuzzy Information, Knowledge Representation and Decision Analysis*, E. Sanchez, ed., IFAC (Pergamon Press), Oxford, 1984, .

[139] A. Chateauneuf & J. -Y. Jaffray, "Derivation of Some Results on Monotone Capacities by Mobius Inversion," in *Uncertainty and Intelligent Systems*, B. Bouchon, L. Saitta & R. R. Yager, eds., Springer-Verlag, Berlin, 1988, , (Lecture Notes in Computer Science 313).

[140] P. Chatterjee, "Modularization of Fault Trees: A Method to Reduce the Cost of Analysis," in *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of Systems Reliability and System Assessment*, R. E. Barlow, J. B. Fussell & N. D. Singpurwalla, eds., SIAM, Philadelphia, 1975, .

[141] P. Cheeseman, "Probabilistic versus Fuzzy Reasoning," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[142] P. M. Chirlian, C. R. Giardina & C. L. Suffel, "On the Effective Bandwidth of Sampled Signals," *IEEE Transactions on Circuits and Systems* **33** (1986), .

[143] K. M. Chong & N. M. Rice, "Equimeasurable Rearrangements of Functions," *Queen's Papers in Pure and Applied Mathematics* **28** (1971).

[144] G. Choquet, "Theory of Capacities," *Annales Institute Fourier (Grenoble)* **5** (1953), .

[145] J. -H. Chou & I. -R. Horng, "State Estimation Using Continuous Orthogonal Functions," *International Journal of Systems Science* **17** (1986), .

[146] C. I. Christov, "A Complete Orthonormal System of Functions in $L^2(-\infty, \infty)$ Space," *SIAM Journal of Applied Mathematics* **42** (1982), .

[147] A. Church, "Review of Kattsoff," *Journal of Symbolic Logic* **2** (1937), 44.

[148] M. R. Civanlar & H. J. Trussell, "Constructing Membership Functions Using Statistical Data," *Fuzzy Sets and Systems* **18** (1986), .

[149] J. J. Clark, M. R. Palmer & P. D. Lawrence, "A Transformation Method for the Reconstruction of Functions from Nonuniformly Spaced Samples," *IEEE Transactions on Acoustics, Speech and Signal Processing* **33** (1985), .

[150] D. G. Clayton, "A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika* **65** (1978), .

[151] J. G. Cleary, "Acquisition of Uncertain Rules in a Probabilistic Logic," *International Journal of Man-Machine Studies* **27** (1987), .

[152] P. R. Clement, "Laguerre Functions in Signal Analysis and Parameter Identification," *Journal of the Franklin Institute* **313** (1982), .

[153] ———, "Applications of Generalized Laguerre Functions," *Mathematics and Computers in Simulation* **27** (1985), .

[154] C. W. Clenshaw & F. W. J. Olver, "Level-Index Arithmetic Operations," *SIAM Journal of Numerical Analysis* **24** (1987), .

[155] R. Cléroux & D. J. McConalogue, "A Numerical Algorithm for Recursively-Defined Convolution Integrals Involving Distribution Functions," *Management Science* **22** (1976), .

[156] J. Cohen, "Partialed Products *are* Interactions; Partialed Powers *are* Curve Components," *Psychological Bulletin* **85** (1978), .

[157] L. Cohen, "Probability Distributions with Given Multivariate Marginals," *Journal of Mathematical Physics* **25** (1984), .

[158] A. R. Colclough, "Two Theories of Experimental Error," *Journal of Research of the National Bureau of Standards* **92** (1987), .

[159] A. G. Colombo, "Uncertainty Propagation in Fault-Tree Analysis," in *Synthesis and Analysis Methods for Safety and Reliability Studies*, G. Apostolakis, S. Garribba & G. Volta, eds., Plenum Press, New York, 1980, .

[160] A. G. Colombo & R. J. Jaarsma, "A Powerful Numerical Method to Combine Random Variables," *IEEE Transactions on Reliability* **29** (1980), .

[161] A. S. Cook & T. Downs, "Estimating Manufacturing Yield by Means of Jacobian of Transformation," *IEE Proceedings* **129G** (1982), .

[162] R. D. Cook & M. E. Johnson, "A Family of Distributions for Modelling Non-elliptically Symmetric Multivariate Data," *Journal of the Royal Statistical Society, Series B* **43** (1981), .

[163] ———, "Generalized Burr-Pareto-Logistic Distributions with Applications to a Uranium Exploration Data Set," *Technometrics* **28** (1986), .

[164] I. D. Cook Jr & J. W. Barnes, "Evaluation of the H-function Integral and of Probability Distributions of Functions of Independent Random Variables," *American Journal of Mathematical and Management Sciences* **1** (1981), .

[165] G. F. Cooper, "A Diagnostic Method that uses Causal Knowledge and Linear Programming in the Application of Bayes' Formula," *Computer Methods and Programs in Biomedicine* **22** (1986), .

[166] ———, "Current Research Directions in the Development of Expert Systems Based on Belief Networks," *Applied Stochastic and Data Analysis* **5** (1989), .

[167] R. D. Cooper, M. R. Hoare & M. Rahman, "Stochastic Processes and Special Functions: On the Probabilistic Origin of some Positive Kernels Associated with Classical Orthogonal Polynomials," *Journal of Mathematical Analysis and its Applications* **61** (1977), .

[168] F. Corsi, "Mathematical Models for Marginal Reliability Analysis," *Microelectronics and Reliability* **23** (1983), .

[169] D. R. Cox, "Some Problems Connected with Statistical Inference," *Annals of Mathematical Statistics* **29** (1958), .

[170] R. Cox, "Probability, Frequency and Reasonable Expectation," *American Journal of Physics* **14** (1946), .

[171] C. C. Craig, "On the Frequency Function of xy," *Annals of Mathematical Statistics* **7** (1936), .

[172] H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[173] M. W. Crofton, "On Mean Value and Probability," in *An Elementary Treatise on the Integral Calculus ... (3rd edition)*, Longmans, Green and Company, London, 1880, (Crofton wrote this one chapter in the treatise by Benjamin Williamson. The chapter does not appear in earlier editions of Williamson's book).

[174] E. L. Crow, "A Property of Additively Closed Families of Distributions," *Annals of Mathematical Statistics* **29** (1958), .

[175] K. S. Crump, "Numerical Inversion of Laplace Transforms using a Fourier Series Approximation," *Journal of the Association of Computing Machinery* **23** (1976), .

[176] M. Cugiani, "A Stochastic Analysis of Errors in Numerical Processes," in *Numerical Techniques for Stochastic Systems*, P. Archetti & M. Cugiani, eds., North-Holland, Amsterdam, 1980, .

[177] R. Cuppens, "The Arithmetic of Distribution Functions," in *A Collection of Papers dedicated to Eugene Lukacs*, J. Gani & V. K. Rohatgi, eds., Academic Press, New York, 1981, .

[178] R. Cuykendall, "Uncontrollable Computational Growth in Theoretical Physics," *International Journal of Theoretical Physics* **24** (1985), .

[179] J. F. Cyranski, "The Probability of a Probability," in *Maximum Entropy and Bayesian Methods in Applied Statistics — Proceedings of the 4th Workshop, Calgary, Canada, 1984*, J. H. Justice, ed., Cambridge University Press, New York, 1986, .

[180] E. Czogała, "Probabilistic Fuzzy Controller as a Generalization of the Concept of a Fuzzy Controller," *Fuzzy Sets and Systems* **26** (1988), .

[181] E. Czogała & T. Rawlik, "Modelling of a Fuzzy Controller with Application to the Control of Biological Processes," *Fuzzy Sets and Systems* **31** (1989), .

[182] E. Czuber, *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung Statistik und Lebensversicherung (Two Volumes)*, B.G. Teubner, Leipzig, 1908.

[183] G. Dall'Aglio, "Fréchet Classes and Compatibility of Distribution Functions," *Symposia Mathematica* **9** (1972), .

[184] I. Daubechies, A. Grossman & Y. Meyer, "Painless Nonorthogonal Expansions," *Journal of Mathematical Physics* **27** (1986), .

[185] B. Davies & B. Martin, "Numerical inversion of the Laplace Transform: a Survey and Comparison of Methods," *Journal of Computational Physics* **33** (1979), .

[186] R. B. Davies, "Numerical Inversion of a Characteristic Function," *Biometrika* **60** (1973), .

[187] E. Davis, "Constraint Propagation with Interval Labels," *Aritficial Intelligence* **32** (1987), .

[188] A. P. Dawid, "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B* **41** (1979), , with discussion.

[189] L. Debnath, "On Laguerre Transform," *Bulletin of the Calcutta Mathematical Society* **52** (1960), 69.

[190] ———, "On Hermite Transform," *Matematicki Vesnik (Belgrade)* **1** (1964), .

[191] ———, "Some Operational Properties of the Hermite Transform," *Matematicki Vesnik (Belgrade)* **5** (1968), .

[192] ———, "On Faltung Theorem of Laguerre Transform," *Studia Universitatis Babes-Bolyai Series Mathematica-physica* **14** (1969), .

[193] R. Dechter & J. Pearl, "Network-Based Heuristics for Constraint-Satisfaction Problems," *Artificial Intelligence* **34** (1988), .

[194] ———, "Tree Clustering for Constraint Networks," *Artificial Intelligence* **38** (1989), .

[195] M. Delgado & S. Moral, "On the Concept of Possibility-Probability Consistency," *Fuzzy Sets and Systems* **21** (1987), .

[196] M. Delgado, J. L. Verdegay & M. A. Vila, "A General Model for Fuzzy Linear Programming," *Fuzzy Sets and Systems* **29** (1989), .

[197] L. S. Dembo, "An Interview with Jorge Luis Borges," *Wisconsin Studies in Contemporary Literature* (Summer 1970).

[198] A. P. Dempster, "Upper and Lower Probability Inferences Based on a Sample from a Finite Univariate Population," *Biometrika* **54** (1967), .

[199] ———, "Upper and Lower Probabilities Induced by a Multivalued Mapping," *Annals of Mathematical Statistics* **38** (1967), .

[200] ———, "A Generalization of Bayesian Inference," *Journal of the Royal Statistical Society Series B* **30** (1968), .

[201] ———, "Upper and Lower Probabilities Generated by a Random Closed Interval," *Annals of Mathematical Statistics* **39** (1968), .

[202] ———, "Upper and Lower Probability Inferences for Families of Hypotheses with Monotone Density Ratios," *Annals of Mathematical Statistics* **40** (1969), .

[203] A. P. Dempster & A. Kong, "Uncertain Evidence and Artificial Analysis," *Journal of Statistical Planning and Inference* **20** (1988), .

[204] M. A. H. Dempster, "Distributions in Intervals and Linear Programming," in *Topics in Interval Analysis*, E. Hansen, ed., Oxford University Press, Oxford, 1969, .

[205] ———, "An Application of Quantile Arithmetic to the Distribution Problem of Stochastic Linear Programming," *Bulletin of the Institute of Mathematics and its Applications* **10** (1974), .

[206] ———, "Introduction to Stochastic Programming," in *Stochastic Programming*, M. A. H. Dempster, ed., Academic Press, London, 1980, .

[207] M. A. H. Dempster & A. Papagakipapoulias, "Computational Experience with an Approximate Method for the Distribution Problem," in *Stochastic Programming*, M. A. H. Dempster, ed., Academic Press, London, 1980, .

[208] J. B. D. Derksen, "On Some Infinite Series Introduced by Tschuprow," *Annals of Mathematical Statistics* **10** (1939), .

[209] L. Devroye, *Lecture Notes on Bucket Algorithms*, Birkhäuser, Boston, 1986.

[210] ———, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

[211] S. Dharmadhikari & K. Joag-Dev, *Unimodality, Convexity, and Applications*, Academic Press, New York, 1988.

[212] S. W. Dharmadhikari & K. Joag-Dev, "The Gauss-Tchebyshev Inequality for Unimodal Distributions," *Theory of Probability and its Applications* **30** (1986), .

[213] B. H. Dickman & M. J. Gilman, "Monte Carlo Optimization," *Journal of Optimization Theory and Applications* **60** (1989), .

[214] J. G. Dijkman, H. van Haeringen & S. J. de Lange, "Fuzzy Numbers," *Journal of Mathematical Analysis and Applications* **92** (1983), .

[215] M. Dishon & G. H. Weiss, "Numerical Inversion of Mellin and Two-Sided Laplace Transforms," *Journal of Computational Physics* **28** (1978), .

[216] J. P. Dolan, *The Essential Erasmus*, Meridian, New York, 1983.

[217] Z. Domotor, "Probabilistic and Causal Dependence Structures," *Theory and Decision* **13** (1981), .

[218] W. Dong & H. C. Shah, "Vertex Method for Computing Functions of Fuzzy Variables," *Fuzzy Sets and Systems* **24** (1987), .

[219] W. Dong & F. S. Wong, "Interactive Fuzzy Variables and Fuzzy Decisions," *Fuzzy Sets and Systems* **29** (1989), .

[220] T. Downs, A. S. Cook & G. Rogers, "A Partitioning Approach to Yield Estimation for Large Circuits and Systems," *IEEE Transactions on Circuits and Systems* **31** (1984), .

[221] D. Driankov, "Inference with Consistent Probabilities in Expert Systems," *International Journal of Intelligent Systems* **4** (1989), .

[222] D. Dubois, "Possibility Theory: Searching for Normative Foundations," in *Risk, Decision and Rationality*, B. R. Munier, ed., D. Riedel, Dordrecht, 1988, .

[223] D. Dubois & H. Prade, "Comment on 'Tolerance analysis using fuzzy sets' and 'A procedure for multiple aspect decision making'," *International Journal of Systems Science* **9** (1978), .

[224] ———, "Operations on Fuzzy Numbers," *International Journal of Systems Science* **9** (1978), .

[225] ———, "Fuzzy Real Algebra: Some Results," *Fuzzy Sets and Systems* **2** (1979), .

[226] ———, "Decision-Making Under Fuzziness," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade & R. R. Yager, eds., North-Holland, Amsterdam, 1979, .

[227] ———, *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York, 1980, (Mathematics in Science and Engineering, volume 144).

[228] ———, "Additions of Interactive Fuzzy Numbers," *IEEE Transactions on Automatic Control* **26** (1981), .

[229] ———, "On Several Representations of an Uncertain Body of Evidence," in *Fuzzy Information and Decision Processes*, M. M. Gupta & E. Sanchez, eds., North-Holland, Amsterdam, 1982, .

[230] ———, "A Class of Fuzzy Measures based on Triangular Norms," *International Journal of General Systems* **8** (1982), .

[231] ———, "Twofold Fuzzy Sets," *Fuzzy Mathematics (Huazhong, China)* **3** (1983), .

[232] ———, "A Review of Fuzzy Set Aggregation Connectives," *Information Sciences* **36** (1985), .

[233] ———, "Fuzzy Sets and Statistical Data," *European Journal of Operational Research* **25** (1986), .

[234] ———, "A Set-Theoretic View of Belief Functions: Logical Operations and Approximations by Fuzzy Sets," *International Journal of General Systems* **12** (1986), .

[235] ———, "On the Unicity of Dempster Rule of Combination," *International Journal of Intelligent Systems* **1** (1986), .

[236] ———, "The Mean Value of a Fuzzy Number," *Fuzzy Sets and Systems* **24** (1987), .

[237] ———, "Editorial (to special issue on fuzzy numbers)," *Fuzzy Sets and Systems* **24** (1987), .

[238] ———, "Evidence Theory and Interval Analysis: Toward bridging the gap between fuzzy arithmetics and Monte-Carlo simulation," in *Preprints of the Second IFSA Congress*, 1987, .

[239] ———, "Fuzzy Numbers: An Overview," in *Analysis of Fuzzy Information, Volume 1: Mathematics and Logic*, J. C. Bezdek, ed., CRC Press Inc., Boca Raton, Florida, 1987, .

[240] ———, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York and London, 1988.

[241] O. D. Duncan, "Path Analysis: Sociological Examples," *The American Journal of Sociology* **72** (1966), .

[242] ———, "Contingencies in Constructing Causal Models," in *Sociological Methodology 1969*, E. F. Borgatta, ed., Jossey-Bass, San Francisco, 1969, .

[243] ———, "Partials, Partitions, and Paths," in *Sociological Methodology 1970*, E. F. Borgatta, ed., Jossey-Bass, San Francisco, 1970, .

[244] ———, *Introduction to Structural Equation Models*, Academic Press, New York, 1975.

[245] J. Dupačová, "Stochastic and Dynamic Programming: Experience in Stochastic Programming Models," in *Survey of Mathematical Programming (Proceedings of the 9th International Mathematical Programming Symposium), Part 5*, A. Prékopa, ed., Publishing House of the Hungarian Academy of Sciences, Budapest, 1976, .

[246] N. Ebrahimi, "Bivariate Processes with Positive or Negative Dependent Structures," *Journal of Applied Probability* **24** (1987), .

[247] K. Ecker & H. Ratschek, "Intervallarithmetik mit Wahrscheinlickkeitsstruktur," *Angewandte Informatik (Elektron. Datenverarbeitung)* **14** (1972), .

[248] E. Eells, "Probabilistic Causal Interaction," *Philosophy of Science* **53** (1986), .

[249] ———, "Probabilistic Causality: Reply to John Depré," *Philosophy of Science* **54** (1987), .

[250] E. Eells & E. Sober, "Probabilistic Causality and the Question of Transitivity," *Philosophy of Science* **50** (1983), .

[251] W. P. Elderton & N. L. Johnson, *Systems of Frequency Curves*, Cambridge University Press, Cambridge, 1969.

[252] B. S. Eldred, "The Application of Integral Transform Theory to the Analysis of Probabilistic Cash Flows," Doctoral dissertation, College of Engineering, University of Texas, 1978.

[253] F. S. Ellett Jr. & D. P. Ericson, "Correlation, Partial Correlation, and Causation," *Synthese* **67** (1986), .

[254] B. Epstein, "Some Applications of the Mellin Transform in Statistics," *Annals of Mathematical Statistics* **19** (1948), .

[255] A. Erdélyi, "On Some Expansions in Laguerre Polynomials," *Journal of the London Mathematical Society* **13** (1938), .

[256] A. Erdélyi, W. Magnus, F. Oberhettinger & F. G. Tricomi, *Higher Transcendental Functions, Volume II*, McGraw-Hill, New York, 1953.

[257] ———, *Tables of Integral Transforms, Volume I*, McGraw-Hill, New York, 1954.

[258] J. D. Esary & F. Proschan, "Relationships Among Some Concepts of Bivariate Dependence," *The Annals of Mathematical Statistics* **43** (1972), .

[259] J. -C. Falmagne, "A Random Utility Model for a Belief Function," *Synthese* **57** (1983), .

[260] J. -P. Faye & J. Vignes, "Stochastic Approach of the Permutation-Perturbation Method for Round-Off Error Analysis," *Applied Numerical Analysis* **1** (1985), .

[261] P. E. Feldheim, "Une Propriété Caractéristique des polynômes de Laguerre," *Commentarii Mathematici Helvetici* **13** (1940), .

[262] W. Feller, *An Introduction to Probability Theory and its Applications (3rd Edition, Volume 1)*, John Wiley and Sons, New York, 1968.

[263] W. Fenchel, "On Conjugate Convex Functions," *Canadian Journal of Mathematics* **1** (1949), .

[264] I. Fenyö, "Some Remarks on the Theory of the Laguerre-Pinney transformation," *Aequationes Mathematicae* **25** (1982), .

[265] R. P. Feynmann, "Simulating Physics with Computers," *International Journal of Theoretical Physics* **21** (1982), .

[266] M. A. Field, "The Literary Theories of Jorge Luis Borges," PhD thesis, University of Queensland, 1978.

[267] E. C. Fieller, "The Distribution of the Index in a Normal Bivariate Population," *Biometrika* **24** (1932), .

[268] T. L. Fine, "Lower Probability Models for Uncertainty and Nondeterministic Processes," *Journal of Statistical Planning and Inference* **20** (1988), .

[269] P. C. Fishburn, "Interval Models for Comparative Probability on Finite Sets," *Journal of Mathematical Psychology* **30** (1986), .

[270] M. Flensted-Jensen & T. Koornwinder, "The Convolution Structure for Jacobi Function Expansions," *Arkiv für Matematik* **11** (1973), .

[271] G. Forbes, *The Metaphysics of Modality*, Clarendon Press, Oxford, 1985.

[272] G. E. Forsythe, "Generation and use of Orthogonal Polynomials for Data-Fitting with a Digital Computer," *Journal of the Society for Industrial and Applied Mathematics* **5** (1957), .

[273] R. Fortet, "Random Determinants," *Journal of Research of the National Bureau of Standards* **47** (1951), .

[274] C. Fox, "The G and H-functions as Symmetrical Fourier Kernels," *Transactions of the American Mathematical Society* **98** (1961), .

[275] G. Fox & J. Quirk, "Uncertainty and Input-Output Analysis," Jet Propulsion Laboratory, 1987, Unpublished.

[276] M. J. Frank, "On the Simultaneous Associativity of $F(x,y)$ and $x+y-F(x,y)$," *Aequationes Mathematicae* **19** (1979), .

[277] M. J. Frank, R. B. Nelsen & B. Schweizer, "Best-Possible Bounds for the Distribution of a Sum — A Problem of Kolmogorov," *Probability Theory and Related Fields* **74** (1987), .

[278] M. J. Frank & B. Schweizer, "On the Duality of Generalized Infimal and Supremal Convolutions," *Rendiconti di Matematica Series VI* **12** (1979), .

[279] M. Fréchet, "Généralisations du théorème des probabilités totales," *Fundamenta Mathematica* **25** (1935), .

[280] ———, "Sur les tableaux de corrélation dont les marges sont données," *Annales de l'Université de Lyon* **9A** (1951), .

[281] G. N. Frederickson & D. B. Johnson, "The Complexity of Selection and Ranking in $X+Y$ and Matrices with Sorted Columns," *Journal of Computing and Systems Science* **24** (1982), .

[282] T. S. Freeman, G. M. Imirzian, E. Kaltofen & L. Yagati, "Dagwood: A System for Manipulating Polynomials Given by Straight-Line Programs," *ACM Transactions on Mathematical Software* **14** (1988), .

[283] P. V. Fua, "Using Probability-Density Functions in the Framework of Evidential Reasoning," in *Uncertainty and Intelligent Systems*, B. Bouchon, L. Saitta & R. R. Yager, eds., Springer-Verlag, Berlin, 1988, 103, (Lecture Notes in Computer Science 313).

[284] D. R. Fulkerson, "Expected Critical Path Length in PERT type Networks," *Operations Research* **10** (1962), .

[285] W. Furmanski & R. Petronzio, "A Method of Analyzing the Scaling Violation of Inclusive Spectra in Hard Processes," *Nuclear Physics* **B195** (1982), .

[286] B. Gabutti & B. Minetti, "A New Application of the Discrete Laguerre Polynomials in the Numerical Evaluation of the Hankel Transform of a Strongly Decreasing Even Function," *Journal of Computational Physics* **42** (1981), .

[287] N. Gaffke & L. Rüschendorf, "On a Class of Extremum Problems in Statistics," *Mathematische Operationsforschung und Statistik Series Optimization* **12** (1981), .

[288] ———, "On the Existence of Probability Measures with Given Marginals," *Statistics and Decisions* **2** (1984), .

[289] J. Garloff, "Interval Mathematics: A Bibliography," *Freiburger Intervall-Berichte* **85/6** (1985), .

[290] G. Gasper, "Linearization of the Product of Jacobi Polynomials II," *Canadian Journal of Mathematics* **22** (1970), .

[291] ———, "Linearization of the Product of Jacobi Polynomials I," *Canadian Journal of Mathematics* **22** (1970), .

[292] ———, "Positivity and the Convolution Structure for Jacobi Series," *Annals of Mathematics* **93** (1971), .

[293] G. Gasper & M. Rahman, "Nonnegative Kernels in Product Formulas for q-Racah Polynomials I," *Journal of Mathematical Analysis and its Applications* **95** (1983), .

[294] W. Gaul, "Bounds for the Expected Duration of a Stochastic Project Planning Model," *Journal of Information and Optimization Sciences* **2** (1981), .

[295] W. Gautschi, "A Computational Procedure for Incomplete Gamma Functions," *ACM Transactions on Mathematical Software* **5** (1979), .

[296] M. R. Genesereth & N. J. Nilsson, *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann Publishers, Inc., 1987.

[297] C. Genest, "Frank's Family of Bivariate Distributions," *Biometrika* **74** (1987), .

[298] C. Genest & J. MacKay, "The Joy of Copulas: Bivariate Distributions with Uniform Marginals," *The American Statistician* **40** (1986), .

[299] C. Genest & R. J. MacKay, "Copules Archimédiennes et Familles de Lois Bidimensionnelles Dont les Marges Sont Données," *The Canadian Journal of Statistics* **14** (1986), .

[300] R. Genin & L. -C. Calvez, "Sur quelques propriétés de la transformation de Laguerre," *Comptes Rendus Academie des Sciences Paris Séries A-B* **2è7** (1968), .

[301] R. Geroch & J. B. Hartle, "Computability and Physical Theories," *Foundations of Physics* **16** (1986), .

[302] P. Giacomo, "On the Expression of Uncertainties," in *Quantum Metrology and Fundamental Physical Constants*, P. H. Cutler & A. A. Lucas, eds., Plenum Press, New York, 1983, .

[303] R. A. Gideon & J. Gurland, "Series Expansions for Quadratic Forms in Normal Variables," *Journal of the American Statistical Association* **71** (1976), .

[304] R. Giles, "A Logic for Subjective Belief," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Volume 1: Foundations and Philosophy of Epistemic Applications of Probability Theory*, W. L. Harper & C. A. Hooker, eds., D. Riedel, Dordrecht, 1976, .

[305] ———, "Foundations for a Theory of Possibility," in *Fuzzy Information and Decision Processes*, M. M. Gupta & E. Sanchez, eds., North-Holland, Amsterdam, 1982, .

[306] ———, "Semantics of Fuzzy Reasoning," *International Journal of Man-Machine Studies* **17** (1982), .

[307] ———, "The Concept of a Grade of Membership," *Fuzzy Sets and Systems* **25** (1988), .

[308] H. -J. Glaeske, "Die Laguerre-Pinney Transformation," *Aequationes Mathematicae* **22** (1981), .

[309] ———, "On a Class of Integral Transformations," *Wissenschaftliche Zeitschrift der Friedrich-Schiller-Universitat: Mathematisch-naturewissenschaftliche Reihe* **31** (1982), .

[310] ———, "On a Convolution Structure of a Generalized Hermite Transformation," *Serdica (Bulgaria)* **9** (1983), .

[311] H. -J. Glaeske & O. I. Maričev, "The Laguerre Transform of the Some Elementary Functions," *Zeitschrift für Analysis und ihre Anwendungen* **3** (1984), .

[312] G. V. Glass, "Correlations with Products of Variables: Statistical Formulation and Implications for Methodology," *American Educational Research Journal* **5** (1968), .

[313] J. Glaz & B. MK. Johnson, "Probability Inequalities for Multivariate Distributions with Dependence Structures," *Journal of the American Statistical Association* **79** (1984), .

[314] C. Glymour, "Artificial Intelligence is Philosophy," in *Aspects of Artificial Intelligence*, J. H. Fetzer, ed., Kluwer Academic Publishers, Dordrecht, 1988, .

[315] C. Glymour & R. Scheines, "Causal Modeling with the TETRAD Program," *Synthese* **68** (1986), .

[316] C. Glymour, R. Scheines, P. Spirtes & K. Kelly, *Discovery Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modelling*, Academic Press, Orlando, Florida, 1987.

[317] H. J. Godwin, "Generalisations of Tchebycheff's Inequality," *Journals of the American Statistical Association* **50** (1955), 923.

[318] ———, *Inequalities on Distribution Functions*, Griffin, London, 1964.

[319] J. Goguen, "Fuzzy Sets and the Social Nature of Truth," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade & R. R. Yager, eds., North-Holland, Amsterdam, 1979, .

[320] A. Goicoechea, "Expert System Models for Inference with Imperfect Knowledge: A Comparative Study," *Journal of Statistical Planning and Inference* **20** (1988), .

[321] A. S. Goldberger, "Structural Equation Models: An Overview," in *Structural Equation Models: An Overview*, A. S. Goldberger & O. D. Duncan, eds., Seminar Press, New York, 1973, .

[322] G. Gong, "Production Systems and Belief Functions," in *Computer Science and Statistics: The Interface*, D. M. Allen, ed., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[323] T. Gonzalez & J. Ja'Ja', "On the Complexity of Computing Bilinear Forms with {0.1} Constants," *Journal of Computer and Systems Sciences* **20** (1980), .

[324] ———, "Evaluation of Arithmetic Expressions with Algebraic Identites," *SIAM Journal of Computing* **11** (1982), .

[325] A. Gonzalez-Arrogo & C. Lopez, "Second-Order Contributions to the Structure Functions in Deep Inelastic Scattering (III): The Singlet Case," *Nuclear Physics* **B166** (1980), .

[326] I. J. Good, "Some History of the Hierarchical Bayesian Methodology," in *Good Thinking: The Foundations of Probability and its Applications*, University of Minnesota Press, Minneapolis, 1983, .

[327] ———, "Subjective Probability as the Measure of a Non-Measurable Set," in *Good Thinking: The Foundations of Probability and its Applications*, Univeristy of Minnesota Press, Minneapolis, 1983, .

[328] ———, "Confidence," *Journal of Statistical Computing and Simulation* **28** (1987), .

[329] I. R. Goodman, "Some New Results Concerning Random Sets and Fuzzy Sets," *Information Sciences* **34** (1984), .

[330] ———, "Some Asymptotic Results for the Combination of Evidence Problem," *Mathematical Modelling* **8** (1987), .

[331] L. A. Goodman, "On the Exact Variance of Products," *Journal of the American Statistical Association* **55** (1960), .

[332] ———, "The Variance of the Product of K Random Variables," *Journal of the American Statistical Association* **57** (1962), .

[333] S. Gottwald, "A Cumulative System of Fuzzy Sets," in *Set Theory and Hierarchy Theory (Lecture Notes in Mathematics 537)*, M. Srebrny & A. Zarach, eds., Springer-Verlag, Berlin, 1976, .

[334] ———, "Set Theory for Fuzzy Sets of Higher Level," *Fuzzy Sets and Systems* **2** (1979), .

[335] ———, "Fuzzy Uniqueness of Fuzzy Mappings," *Fuzzy Sets and Systems* **3** (1980), .

[336] I. S. Gradshteyn & I. M. Ryzhik, *Table of Integrals, Series, and Products, (4th Edition)*, Academic Press, New York, 1965.

[337] I. Grattan-Guinness, "Fuzzy Membership Mapped onto Interval and Many Valued Quantities," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **22** (1976), .

[338] I. J. Greenberg, "Distribution-Free Analysis of M/G/1 and G/M/1 Queues," *Operations Research* **21** (1973), .

[339] Y. I. Grize & T. L. Fine, "Continuous Lower Probability-Based Models for Stationary Processes with Bounded and Divergent Time Averages," *Annals of Probability* **15** (1987), .

[340] B. N. Grosof, "Evidential Confirmation as Transformed Probability: On the Duality of Priors and Updates," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[341] ———, "An Inequality Paradigm for Probabilistic Knowledge: The Logic of Conditional Probability Intervals," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[342] R. D. Gumb, "Comments on Probabilistic Semantics," in *Essays in Epistemology and Semantics*, H. Leblanc, R. Stern & R. Gumb, eds., Haven Publications, New York, 1983, .

[343] E. J. Gumbel, "On the Reliability of the Classical Chi-Square Test," *Annals of Mathematical Statistics* **14** (1943), 253.

[344] ———, *Statistics of Extremes*, Columbia University Press, New York and London, 1958.

[345] ———, "Measurement of Rare Events," in *Measurement: Definitions and Theories*, C. W. Churchman & P. Ratoosh, eds., J. Wiley and Sons, New York, 1959.

[346] D. P. Gupta, "Cesàro Summability of Laguerre Series," *Journal of Approximation Theory* **7** (1973), .

[347] M. A. S. Guth, "Uncertainty Analysis of Rule-Based Expert Systems with Dempster-Shafer Mass Assignments," *International Journal of Intelligent Systems* **3** (1988), .

[348] I. Hacking, "Possibility," *Philosophical Review* **76** (1967), .

[349] ———, "Equipossibility Theories of Probability," *British Journal for the Philosophy of Science* **22** (1971), .

[350] ———, "All Kinds of Possibility," *Philosophical Review* **84** (1975), .

[351] ———, *The Emergence of Probability*, Cambridge University Press, Cambridge, 1975.

[352] G. J. Hahn & S. S. Shapiro, *Statistical Models in Engineering*, John Wiley and Sons, New York, 1967.

[353] T. Hailperin, "Best Possible Inequalities for the Probability of a Logical Function of Events," *American Mathematical Monthly* **72** (1965), .

[354] ———, "Probability Logic," *Notre Dame Journal of Formal Logic* **25** (1984), .

[355] ———, *Boole's Logic and Probability*, North-Holland, Amsterdam, 1986, Second Edition.

[356] J. M. Hammersley, "The Zeros of a Random Polynomial," in *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, Volume II*, J. Neyman, ed., University of California Press, Berkeley and Los Angeles, 1956, .

[357] J. F. Hanna, "Probabilistic Explanation and Probabilistic Causality," in *PSA 1982 (Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association, Volume 2)*, P. D. Asquith & T. Nickles, eds., Philosophy of Science Association, East Lansing, 1983, .

[358] E. R. Hansen, "A Globally Convergent Interval Method for Computing and Bounding Real Roots," *BIT* **18** (1978), .

[359] F. Harary, R. Z. Norman & D. Cartwright, *Structural Models: An Introduction to the Theory of Directed Graphs*, John Wiley and Sons, New York, 1965.

[360] W. Harper, "A Conditional Belief Semantics for Free Quantification Logic with Identity," in *Essays in Epistemology and Semantics*, H. Leblanc, R. Stern & R. Gumb, eds., Haven Publications, New York, 1983, .

[361] A. Hart, *Knowledge Acquisition for Expert Systems*, Kogan Page, London, 1986.

[362] W. D. Hart, "Probability as Degree of Possibility," *Notre Dame Journal of Formal Logic* **13** (1972), .

[363] C. Hastings, *Approximations for Digital Computers*, Princeton University Press, Princeton, 1955.

[364] M. Hauenschild, "Extended Circular Arithmetic, Problems and Results," in *Interval Mathematics 1980*, K. Nickel, ed., Academic Press, New York, 1980, .

[365] D. Hausman, "Are there Causal Relations Among Dependent Variables?," *Philosophy of Science* **50** (1983), .

[366] J. W. Head & W. P. Wilson, "Laguerre Functions: Tables and Properties," *Proceedings of the IEE* **103C** (1956), .

[367] A. C. Hearn, "Optimal Evaluation of Algebraic Expressions," in *Algebraic Algorithms and Error-Correcting Codes*, J. Calmet, ed., Springer-Verlag, Berlin, 1986, , (Lecture Notes in Computer Science 229).

[368] D. Heckerman, "Probabilistic Interpretations for MYCIN's Certainty Factors," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[369] D. E. Heckerman, "An Axiomatic Framework for Belief Updates," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[370] D. E. Heckerman & E. J. Horvitz, "The Myth of Modularity in Rule-Based Systems for Reasoning with Uncertainty," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[371] D. R. Heise, "Problems in Path Analysis and Causal Inference," in *Sociological Methodology 1969*, E. F. Borgatta, ed., Jossey-Bass, San Francisco, 1969, .

[372] S. J. Henkind & M. C. Harrison, "An Analysis of Four Uncertainty Calculi," *IEEE Transactions on Systems, Man, and Cybernetics* **18** (1988), .

[373] M. Henrion, "Uncertainty in Artificial Intelligence: Is Probability Epistemologically and Heuristically Adequate?," in *Expert Judgement and Expert Systems*, J. L. Mumpower, L. D. Phillips, O. Renn & V. R. R. Uppuluri, eds., Springer-Verlag, Berlin, 1988, .

[374] ———, "Propagating Uncertainty in Bayesian Networks by Probabilistic Logic Sampling," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[375] J. R. Higgins, "Five Short stories about the Cardinal Series," *Bulletin (New Series) of the American Mathematical Society* **12** (1985), .

[376] T. W. Hill Jr, "On Determining a Distribution Function Known Only by its Moments," PhD thesis, Arizona State University, 1969.

[377] J. Hintakka, "Aristotle on the Realization of Possibilities in Time," in *Reforging the Great Chain of Being: Studies in the History of Modal Theories*, S. Knuuttila, ed., D. Riedel, Dordrecht, 1981, .

[378] I. I. Hirschman Jr., "Laguerre Transforms," *Duke Mathematical Journal* **30** (1963), .

[379] I. I. Hirschman Jr. & D. V. Widder, *The Convolution Transform*, Princeton University Press, Princeton, 1955.

[380] E. Hisdal, "Conditional Possibilities, Independence and Noninteraction," *Fuzzy Sets and Systems* **1** (1978), .

[381] ———, "Possibilistically Dependent Variables and a General Theory of Fuzzy Sets," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade & R. R. Yager, eds., North-Holland, Amsterdam, 1979, .

[382] ———, "Infinite-Valued Logic based on Two-Valued Logic and Probability. Part 1.1 Difficulties with the Present-day Fuzzy-Set Theory and their Resolution in the TEE Model," *International Journal of Man-Machine Studies* **25** (1986), .

[383] ———, "Infinite-Valued Logic based on Two-Valued Logic and Probability. Part 1.2 Different Sources of Fuzziness," *International Journal of Man-Machine Studies* **25** (1986), .

[384] ———, "Are Grades of Membership Probabilities?," *Fuzzy Sets and Systems* **25** (1988), .

[385] ———, "Guest Editorial," *Fuzzy Sets and Systems* **25** (1988), .

[386] ———, "The Philosophical Issues Raised by Fuzzy Set Theory," *Fuzzy Sets and Systems* **25** (1988), .

[387] U. W. Hochstrasser, "Orthogonal Polynomials," in *Handbook of Mathematical Functions*, M. Abramowitz & I. A. Stegun, eds., National Bureau of Standards, Applied Mathematics Series 55, 1965, .

[388] W. Höffding, "Masstabinvariante Korrelationstheorie," *Schriftenreihe Math. Inst. und Inst. Angew. Math. Univeristät Berlin* **5** (1941), , Mathematical Reviews 3 (1942), 5; Zbl. 24 (1941), 56..

[389] U. Höhle, "Probabilistic Uniformization of Fuzzy Topologies," *Fuzzy Sets and Systems* **1** (1978), .

[390] ———, "Representation Theorems for L-Fuzzy Quantities," *Fuzzy Sets and Systems* **5** (1981), .

[391] ———, "Fuzzy Real Numbers as Dedekind Cuts with Respect to a Multiple Valued Logic," *Fuzzy Sets and Systems* **24** (1987), .

[392] ———, "A General Theory of Fuzzy Plausibility Measures," *Journal of Mathematical Analysis and Applications* **127** (1987), .

[393] ———, "Quotients with Respect to Similarity Relations," *Fuzzy Sets and Systems* **27** (1988), .

[394] P. W. Holland, "Statistics and Causal Inference," *Journal of the American Statistical Association* **81** (1986), , With discussion.

[395] P. W. Holland & Y. J. Wang, "Dependence Function for Continuous Bivariate Densities," *Communications in Statistics — Theory and Methods* **16** (1987), .

[396] ———, "Regional Dependence for Continuous Bivariate Densities," *Communications in Statistics — Theory and Methods* **16** (1987), .

[397] S. Horwitz, T. Reps & D. Binkley, "Interprocedural Slicing Using Dependence Graphs," *SIGPLAN Notices* **23** (1988), , (Proceedings of the SIGPLAN'88 Conference on Programming Language Design and Implementation, Atlanta, Georgia, June 22–24, 1988).

[398] G. E. Hughes & M. J. Cresswell, *An Introduction to Modal Logic*, Methuen, London, 1968.

[399] R. A. Hummel & M. S. Landy, "A Statistical Viewpoint on the Theory of Evidence," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10** (1988), .

[400] E. V. Huntington, "Frequency Distribution of Product and Quotient," *Annals of Mathematical Statistics* **10** (1939), .

[401] C. Hwang & Y. -P. Shih, "Parameter Identification via Laguerre Polynomials," *International Journal of Systems Science* **13** (1982), .

[402] ———, "Laguerre Series Solution of a Functional Differential Equation," *International Journal of Systems Science* **13** (1982), .

[403] G. E. Ingram, E. L. Walker & C. R. Herrmann, "Designing for Reliability Based on Probabilistic Modelling Using Remote Access Computer Systems," in *Proceedings of the 7th Reliability and Maintainability Conference, San Francisco*, July 1968, .

[404] V. V. Ivlev, "Determination of the Parameters of the Generalized Laguerre Polynomials in Reliability Problems," *Engineering Cybernetics* **19** (1981), .

[405] P. S. Jackson, "A Second-Order Moments Method for Uncertainty Analysis," *IEEE Transactions on Reliability* **31** (1982), .

[406] H. Jacobs, J. W. Barnes & I. D. Cook Jr, "Applications of the H-Function Distribution in Classifying and Fitting Classical Probability Distributions," *American Journal of Mathematical and Management Sciences* **7** (1987), .

[407] R. Jain, "Tolerance Analysis using Fuzzy Sets," *International Journal of Systems Science* **7** (1976), .

[408] ———, "Outline of an Approach for the Analysis of Fuzzy Systems," *International Journal of Control* **23** (1976), .

[409] E. T. Jaynes, "Confidence Intervals vs Bayesian Intervals," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Volume II*, W. L. Harper & C. A. Hooker, eds., D. Riedel, Dordrecht, 1976, .

[410] A. J. Jerri, "Sampling Expansion for a Laguerre-L_ν^α Transform," *Journal of Research of the National Bureau of Standards* **80B** (1976), .

[411] ———, "The Shannon Sampling Theorem — Its Various Extensions and Applications: A Tutorial Review," *Proceedings of the IEEE* **65** (1977), , Corrections, 67 (1979), 695.

[412] K. Jogdeo, "Dependence, Concepts of," in *Encyclopedia of Statistical Sciences*, N. L. Johnson & S. Kotz, eds., John Wiley and Sons, New York, 1982, .

[413] N. L. Johnson & S. Kotz, *Continuous Univariate Distributions 2*, Houghton Mifflin Company, Boston, 1970.

[414] P. T. Johnstone, "Open Letter to Ian Graham," *Fuzzy Sets and Systems* **27** (1988), , with reply.

[415] M. C. Jones & H. W. Lotwick, "On the Errors Involved in Computing the Empirical Characteristic Function," *Journal of Statistical Computation and Simulation* **17** (1983), .

[416] ———, "A Remark on Algorithms AS176. Kernel Density Estimation Using the Fast Fourier Transform," *Applied Statistics* **33** (1984), .

[417] K. G. Jöreskog, "A General Method for Estimating a Linear Structural Equation System," in *Structural Equation Models in the Social Sciences*, A. S. Goldberger & O. D. Duncan, eds., Seminar Press, New York, 1973, .

[418] W. M. Kahane, "A More Complete Interval Arithmetic," Unpublished Lecture Notes for a Summer Course at the University of Michigan, 1968.

[419] P. Kall, *Stochastic Linear Programming*, Springer-Verlag, Berlin, 1976.

[420] L. N. Kanal & J. F. Lemmer, eds., *Uncertainty in Artificial Intelligence*, Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986.

[421] Y. Kanjin, "On Algebras with Convolution Structures for Laguerre Polynomials," *Transactions of the American Mathematical Society* **295** (1986), .

[422] M. Kaplan & T. L. Fine, "Joint Orders in Comparative Probability," *Annals of Probability* **5** (1977), .

[423] S. Kaplan, "On the Method of Discrete Probability Distributions in Risk and Reliability Calculations — Applications to Seismic Risk Assessment," *Risk Analysis* **1** (1981), .

[424] ———, "Response to Nelson and Rasmuson's Letter," *Risk Analysis* **2** (1982), 207.

[425] S. Kaplan & J. C. Lin, "An Improved Condensation Procedure in Discrete Probability Distribution Calculations," *Risk Analysis* **7** (1987), .

[426] S. Kaplan, H. F. Perla & D. C. Bley, "A Methodology for Seismic Risk Analysis of Nuclear Power Plants," *Risk Analysis* **3** (1983), .

[427] J. C. Kapteyn, "Definition of the Correlation Coefficient," *Monthly Notices of the Royal Astronomical Society* **72** (1912), .

[428] L. O. Kattsoff, "Modality and Probability," *The Philosophical Review* **46** (1937), .

[429] D. Katz, "Discrete Approximations to Continuous Density Functions that are L_1 Optimal," *Computational Statistics and Data Analysis* **1** (1983), .

[430] D. Katz & D. Z. D'Argenio, "Discrete Approximation of Multivariate Densities with Application to Bayesian Estimation," *Computational Statistics and Data Analysis* **2** (1984), .

[431] A. Kaufmann, "Fuzzy Subsets Applications in O.R. and Management," in *Fuzzy Sets Theory and Applications*, A. Jones, A. Kaufmann & H. -J. Zimmermann, eds., D. Riedel, Dordrecht, 1986, .

[432] A. Kaufmann & M. M. Gupta, *Introduction to Fuzzy Arithmetic: Theory and Applications*, Van Nostrand Reinhold, New York, 1985.

[433] R. B. Keey & C. H. Smith, "The Propagation of Uncertainties in Failure Events," *Reliability Engineering* **10** (1985), .

[434] J. Keilson, "Return of the Wanderer: A Physicist Becomes a Probabilist," in *The Craft of Probabilistic Modelling: A Collection of Personal Accounts*, J. Gani, ed., Springer-Verlag, New York, 1986, .

[435] J. Keilson & W. R. Nunn, "Laguerre Transformation as a Tool for the Numerical Solution of Integral Equations of Convolution Type," *Applied Mathematics and Computation* **5** (1979), .

[436] J. Keilson, W. R. Nunn & U. Sumita, "The Bilateral Laguerre Transform," *Applied Mathematics and Computation* **8** (1981), .

[437] J. Keilson, D. Petrondas, U. Sumita & J. Wellner, "Significance Points for Some Tests of Uniformity on the Sphere," *Journal of Statistical Computation and Simulation* **17** (1983), .

[438] J. Keilson & U. Sumita, "Waiting Time Distribution Response to Traffic Surges via the Laguerre Transform," Working Paper Series No. 8111, The Graduate School of Management, The University of Rochester, 1981.

[439] ———, "The Depletion Time for M/G/1 Systems and a Related Limit Theorem," *Advances in Applied Probability* **15** (1983), .

[440] ———, "Evaluation of the Total Time in System in a Preempt/Resume Priority Queue via a Modified Lindley Process," *Advances in Applied Probability* **15** (1983), .

[441] ———, "A General Laguerre Transform and a Related Distance between Probability Measures," *Journal of Mathematical Analysis and its Applications* **113** (1986), .

[442] H. G. Kellerer, "Funktionen auf Produkträumen mit vorgegebenen Marginal-Funktionen," *Mathematische Annalen* **144** (1961), .

[443] ———, "Maßtheoretische Marginalprobleme," *Mathematische Annalen* **153** (1964), .

[444] ———, "Marginalprobleme für Funktionen," *Mathematische Annalen* **154** (1964), .

[445] ———, "Verteilungsfunktionen mit gegebenen Marginalverteilungen," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **3** (1964), .

[446] ———, "Duality Theorems for Marginal Problems," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **67** (1984), .

[447] S. D. Kellog & J. W. Barnes, "The Distribution of Products, Quotients, and Powers of Two Dependent H-Function Variates," *Mathematics and Computers in Simulation* **29** (1987), .

[448] M. Kendall & A. Stuart, *The Advanced Theory of Statistics, Volume 1: Distribution Theory (4th Edition)*, Griffin, London, 1977.

[449] M. G. Kendall & P. A. P. Moran, *Geometrical Probability*, Griffin, London, 1963.

[450] M. G. Kendall & A. Stuart, *The Advanced Theory Of Statistics, Volume 2: Inference and Relationship (3rd Edition)*, Griffin, London, 1960.

[451] W. J. Kennedy Jr & J. E. Gentle, *Statistical Computing*, Marcel Dekker Inc., New York and Basel, 1980.

[452] D. A. Kenny, *Correlation and Causality*, John Wiley and Sons, New York, 1979.

[453] B. W. Kernighan & D. M. Ritchie, *The C Programming Language*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[454] J. M. Keynes, *A Treatise on Probability*, MacMillan and Company, London, 1921.

[455] V. L. Kharitonov, "Asymptotic Stability of an Equilibrium Position of a Family of Systems of Linear Differential Equations," *Differential'nye Uraveniya* **14** (1978), .

[456] J. Kiefer, "Conditional Confidence Statements and Confidence Estimators," *Journal of the American Statistical Association* **72** (1977), , (with discussion). Reprinted (without the discussion) in L.D. Brown, I. Olkin, J. Sacks, H.P. Wynn eds, *Jack Karl Keifer, Collected Papers II, Statistical Inference and Probability (1964–1984)*, Springer-Verlag, New York, 1985.

[457] ———, "Conditional Inference," in *Encyclopedia of Statistical Sciences*, N. L. Johnson & S. Kotz, eds., John Wiley and Sons, New York, 1982.

[458] G. Kimeldorf & A. Sampson, "Uniform Representations of Bivariate Distributions," *Communications in Statistics* **4** (1975), .

[459] G. Kimeldorf & A. R. Sampson, "Positive Dependence Orderings," *Annals of the Institute for Statistical Mathematics* **39A** (1987), .

[460] R. E. King & P. N. Paraskevopoulos, "Parametric Identification of Discrete-time SISO Systems," *International Journal of Control* **30** (1979), .

[461] N. V. Kirianaki & Y. S. Klushin, "Machine Method for Plotting the Probability Density Function of a Resultant Error with two or More Components," *Measurement Techniques* **29** (1986), .

[462] L. B. Klebanov & S. T. Mkrtchyan, "Estimation of the Closeness of Distributions in terms of Identical Moments," *Journal of Soviet Mathematics* **32** (1986), .

[463] J. de Kleer & J. S. Brown, "A Qualitative Physics Based on Confluences," *Artificial Intelligence* **24** (1984), .

[464] W. K. Klein Haneveld, "Robustness against dependence in PERT: An application of duality and distributions with known marginals," *Mathematical Programming Study* **27** (1986), .

[465] E. P. Klement, "Operations on Fuzzy Sets and Fuzzy Numbers Related to Triangular Norms," in *Proceedings of the 11th IEEE International Symposium on Multiple-Valued Logic*, 1981, .

[466] D. E. Knuth, *Seminumerical Algorithms (The Art of Computer Programming Volume 2)*, Addison-Wesley, Reading, Massachusetts, 1981, Second Edition.

[467] S. Knuuttila, *Reforging the Great Chain of Being: Studies in the History of Modal Theories*, D. Riedel, Dordrecht, 1981.

[468] V. V. Koblin, *Stochastic Programming*, D. Riedel, Dordrecht, 1977.

[469] J. Kohlas, "Conditional Belief Structures," *Probability in the Engineering and Informational Sciences* **2** (1988), .

[470] J. Komlos, "On the Determinant of (0,1) Matrices," *Studia Scientiarum Mathematicarum Hungarica* **2** (1967), .

[471] ———, "On the Determinant of Random Matrices," *Studia Scientiarum Mathematicarum Hungarica* **3** (1968), .

[472] B. O. Koopman, "The Axioms and Algebra of Intuitive Probability," *Annals of Mathematics* **41** (1940), .

[473] ———, "The Bases of Probability," *Bulletin of the American Mathematical Society* **46** (1940), .

[474] T. Koornwinder, "A New Proof of a Paley-Weiner type theorem for the Jacobi Transform," *Arkiv für Matematik* **13** (1975), .

[475] B. Korel, "The Program Dependence Graph in Static Program Testing," *Information Processing Letters* **24** (1987), .

[476] I. Kotlarski, "On the Generalized Mellin Transform of a Complex Random Variable and its Applications," *Annals of Mathematical Statistics* **36** (1965), .

[477] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley and Sons, New York, 1978.

[478] S. A. Kripke, "A Completeness Theorem in Modal Logic," *Journal of Symbolic Logic* **24** (1959), .

[479] H. Krips, "Popper, Propensities, and Quantum Theory," *British Journal for the Philosophy of Science* **35** (1984), .

[480] R. Kruse & K. D. Meyer, *Statistics with Vague Data*, D. Riedel, Dordrecht, 1987.

[481] W. Kruskal, "Miracles and Statistics: The Casual Assumption of Independence," *Journal of the American Statistical Association* **83** (1988), .

[482] S. Kullback, "The Distribution Laws of the Difference and Quotient of Variables Independently Distributed in Pearson type III laws," *Annals of Mathematical Statistics* **7** (1936), .

[483] ———, "Probability Densities with Given Marginals," *Annals of Mathematical Statistics* **39** (1968), .

[484] A. Kumar & T. L. Fine, "Stationary Lower Probabilities and Unstable Averages," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **69** (1985), .

[485] F. -C. Kung & H. Lee, "Solution of Linear State-Space Equations and Parameter Estimation in Feedback Systems using Laguerre Polynomial Expansion," *Journal of the Franklin Institute* **314** (1982), .

[486] W. Kuo & V. R. R. Uppuluri, "A Review of Error Propagation Analysis in Systems," *Microelectronics and Reliability* **23** (1983), .

[487] R. E. Kurth & D. C. Cox, "Discrete Probability Distributions for Probabilistic Fracture Mechanics," *Risk Analysis* **5** (1985), .

[488] V. P. Kuznetsov, "Choice of Coefficient for Calculating the Interval Estimate of Measurement Error," *Measurement Techniques* **26** (1983), .

[489] H. Kwakernaak, "Fuzzy Random Variables I," *Information Sciences* **15** (1978), .

[490] ———, "Fuzzy Random Variables II," *Information Sciences* **17** (1979), .

[491] M. Kwiesielewicz & L. Stolc, "Solving Linear Programming Problems with Nondeterministic Constraints using Fuzzy Numbers Arithmetic," *International Journal of Systems Science* **20** (1989), .

[492] H. E. Kyburg Jr., *The Logical Foundations of Statistical Inference*, D. Reidel, Dordrecht, 1974.

[493] H. E. Kyburg Jr, "Bayesian and Non-Bayesian Evidential Updating," *Artificial Intelligence* **31** (1987), .

[494] H. J. Kyburg Jr, "Knowledge," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[495] A. R. Lacey, "Modality," in *A Dictionary of Philosophy*, A. R. Lacey, ed., Routledge and Kegan Paul, London, 1976, .

[496] T. L. Lai & H. Robbins, "A Class of Dependent Random Variables and Their Maxima," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **42** (1978), .

[497] L. M. Laita, "Boolean Algebra and its Extra-logical Sources: The Testimony of Mary Everest Boole," *History and Philosophy of Logic* **1** (1980), .

[498] D. Lamb & S. M. Easton, *Multiple Discovery: The Pattern of Scientific Progress*, Avebury, 1984.

[499] H. E. Lambert, "Measures of Importance of Events and Cut Sets in Fault Trees," in *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of Systems Reliability and System Assessment*, R. E. Barlow, J. B. Fussell & N. D. Singpurwalla, eds., SIAM, Philadelphia, 1975, .

[500] K. C. Land, "Principles of Path Analysis," in *Sociological Methodology 1969*, E. F. Borgatta, ed., Jossey-Bass, San Francisco, 1969, .

[501] S. L. Lauritzen & D. J. Spiegelhalter, "Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems," *Journal of the Royal Statistical Society, Series B* **50** (1988), .

[502] S. L. Lauritzen & N. Wermuth, "Graphical Models for Associations Between Variables, Some of Which are Qualitative and Some Quantitative," *Annals of Statistics* **17** (1989), .

[503] S. E. Laveuve, "Definition Einer Kahan-Arithmetik Und Ihre Implementierung," in *Interval Mathematics (Lecture Notes in Computer Science 29)*, K. Nickel, ed., Springer-Verlag, Berlin, 1975, .

[504] H. Leblanc, "Probabilistic Semantics for First-Order Logic," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* **25** (1979), .

[505] ———, "Probabilistic Semantics: An Overview," in *Essays in Epistemology and Semantics*, H. Leblanc, R. Stern & R. Gumb, eds., Haven Publications, New York, 1983, .

[506] E. L. Lehmann, "Some Concepts of Dependence," *Annals of Mathematical Statistics* **37** (1966), .

[507] R. Leipnik, "Integral Equations, Biorthonormal Expansions and Noise," *Journal of the Society for Industrial and Applied Mathematics* **7** (1959), .

[508] W. Leler, *Constraint Programming Languages*, Addison-Wesley, Reading, Massachusetts, 1988.

[509] S. Lem, "The Upside-Down Evolution," in *One Human Minute*, Harcourt Brace Jovanovich, San Diego, 1986, .

[510] J. F. Lemmer, "Confidence Factors, Empiricism and the Dempster-Shafer Theory of Evidence," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[511] J. F. Lemmer & L. N. Kanal, eds., *Uncertainty in Artificial Intelligence II*, Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986.

[512] K. S. Leung & W. Lam, "Fuzzy Concepts in Expert Systems," *IEEE Computer* **21(9)** (1988), .

[513] I. Levi, "On Indeterminate Probabilities," *The Journal of Philosophy* **71** (1974), .

[514] ———, *The Enterprise of Knowledge (An Essay on Knowledge, Credal Probability and Chance)*, MIT Press, Cambridge, Massachusetts, 1980.

[515] ———, "Imprecision and Indeterminancy in Probability Judgement," *Philosophy of Science* **52** (1985), .

[516] P. Lévy, *Calcul des Probabilités*, Gauthier-Villars, Paris, 1925.

[517] J. S. Lew, "On Some Relations Between the Laplace and Mellin Transforms," *IBM Journal of Research and Development* **19** (1975), .

[518] D. Lewis, "Probabilities of Conditionals and Conditional Probabilities," *The Philosophical Review* **85** (1976), .

[519] D. V. Lindley, "Scoring Rules and the Inevitability of Probability," *International Statistical Review* **50** (1982), , With discussion.

[520] M. Loève, *Probability Theory I*, Springer-Verlag, New York, 1977.

[521] J. S. Long, "Covariance Structure Models: An Introduction to LISREL," Sage University Series on Quantitative Applications in the Social Sciences 07-034, Sage Publications, Beverly Hills, California, 1983.

[522] R. P. Loui, "Interval Based Decisions for Reasoning Systems," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[523] ———, "Computing Reference Classes," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[524] M. J. Loux, "Modality and Metaphysics," in *The Possible and the Actual: Readings in the Metaphysics of Modality*, M. J. Loux, ed., Cornell University Press, Ithaca, 1979, .

[525] A. O. Lovejoy, *The Great Chain of Being*, Harvard University Press, Cambridge, Massachusetts, 1966.

[526] J. N. Lovett, "Evaluation of the H-function Inversion Integral," PhD thesis, University of Arkansas, 1977.

[527] R. Lowen, "Editorial: Mathematics and Fuzziness, Part 1," *Fuzzy Sets and Systems* **27** (1988), .

[528] J. D. Lowrance, "Automating Argument Construction," *Journal of Statistical Planning and Inference* **20** (1988), .

[529] D. G. Luenberger, "Quasi-Convex Programming," *SIAM Journal of Applied Mathematics* **16** (1968), .

[530] ———, *Optimization by Vector Space Methods*, John Wiley and Sons, New York, 1969.

[531] M. K. Luhandjula, "On Possibilistic Linear Programming," *Fuzzy Sets and Systems* **18** (1986), .

[532] J. Lynch, G. Mimmack & F. Proschan, "Uniform Stochastic Orderings and Total Positivity," *The Canadian Journal of Statistics* **15** (1987), .

[533] N. A. Lynch, "Straight-Line Program Length as a Parameter for Complexity Analysis," *Journal of Computer and System Sciences* **21** (1980), .

[534] D. MacHale, *George Boole: His Life and Work*, Boole Press, Dublin, 1985.

[535] D. P. Madan & J. C. Owings, "Decision Theory with Complex Uncertainties," *Synthese* **75** (1988), .

[536] G. D. Makarov, "Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed," *Theory of Probability and its Applications* **26** (1981), .

[537] C. L. Mallows, "Generalization of Tchebycheff's Inequalities (with discussion)," *Journal of the Royal Statistical Society Series B* **18** (1956), .

[538] E. H. Mamdami, "Application of Fuzzy Algorithms for the Control of Simple Dynamic Plant," *Proceedings of the IEE* **121** (1974), .

[539] E. G. Manes, "Review of 'Fuzzy sets and systems: Theory and applications,' by Didier Dubois and Henri Prade," *Bulletin (New Series) of the American Mathematical Society* **7** (1982), .

[540] ———, "A Class of Fuzzy Theories," *Journal of Mathematical Analysis and its Applications* **85** (1982), .

[541] E. G. Manes & M. A. Arbib, *Algebraic Approaches to Program Semantics*, Springer-Verlag, New York, 1986.

[542] B. Manhire, R. T. Jenkins, H. R. Outhred, D. Sutanto, R. J. Kaye & F. J. Evans, "The Cumulant Method of Power System Probabilistic Simulation: Principles, Mathematics and Limitations," *Journal of Electrical and Electronics Engineering — Australia* **7** (1987), .

[543] H. B. Mann & A. Wald, "On the Choice of the Number of Class Intervals in the Application of the Chi Square Test," *Annals of Mathematical Statistics* **13** (1942), .

[544] M. Mareš & J. Horák, "Fuzzy Quantities in Networks," *Fuzzy Sets and Systems* **10** (1983), .

[545] R. Martin-Clouaire & H. Prade, "On the Problems of Representation and Propagation of Uncertainty in Expert Systems," *International Journal of Man-Machine Studies* **22** (1985), .

[546] Y. Masuda, J. G. Shanthikumar & U. Sumita, "A General Software Availability/Reliability Model: Numerical Exploration via the Matrix Laguerre Transform," Working Paper Series No. QM8425, The Graduate School of Management, The University of Rochester, 1984.

[547] A. M. Mathai, "A Review of the Different Techniques used for Deriving the Exact Distributions of Multivariate Test Criteria," *Sankhyā Series A* **35** (1973), .

[548] A. M. Mathai & R. K. Saxena, "On the Linear Combinations of Stochastic Variables," *Metrika* **20** (1973), .

[549] ———, *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences*, Springer-Verlag, Berlin, 1973, (Lecture Notes in Mathematics 348).

[550] ———, *The H-function with Applications in Statistics and Other Disciplines*, Halstead Press, New York, 1978.

[551] G. Matheron, *Random Sets and Integral Geometry*, Wiley, New York, 1975.

[552] Y. Matijasevic, "Enumerable Sets are Diophantine," *Doklady Akademii Nauk SSSR* **191** (1970), .

[553] J. Mau, "On Cox's Confidence Distribution," in *Probability and Bayesian Statistics (International Symposium on Probability and Bayesian Statistics — Insbruck, 1986)*, R. Vierti, ed., Plenum Press, New York, 1987, .

[554] D. G. Mayo, "In Defense of the Neyman-Pearson Theory of Confidence Intervals," *Philosophy of Science* **48** (1981), .

[555] M. Mazumdar, "Importance Sampling in Reliability Estimation," in *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of Systems Reliability and System Assessment*, R. E. Barlow, J. B. Fussell & N. D. Singpurwalla, eds., SIAM, Philadelphia, 1975, .

[556] R. A. McCain, "Fuzzy Confidence Intervals," *Fuzzy Sets and Systems* **10** (1983), .

[557] D. J. McConalogue, "Convolution Integrals Involving Probability Distribution Functions," *Computer Journal* **21** (1978), .

[558] ———, "Numerical Treatment of Convolution Integrals Involving Distributions with Densities having Singularities at the Origin," *Communications in Statistics — Simulation and Computation* **B10** (1981), .

[559] J. McCully, "The Laguerre Transform," *SIAM Review* **2** (1960), 185.

[560] D. V. McDermott & E. Davis, "Planning Routes Through Uncertain Territory," *Artificial Intelligence* **22** (1984), .

[561] J. L. Mehlberg, "A Classification of Mathematical Concepts," *Synthese* **14** (1962), .

[562] K. Menger, "Statistical Metrics," *Proceedings of the National Academy of Science (United States of America)* **28** (1942), .

[563] ———, "Probabilistic Geometry," *Proceedings of the National Academy of Science (United States of America)* **37** (1951), .

[564] ———, "The Theory of Relativity and Geometry," in *Albert Einstein: Philosopher-Scientist*, P. A. Schilpp, ed., Tudor Publishing Company, New York, 1951, , (See also Einstein's reply on page 686).

[565] ———, "On Variables in Mathematics and in Natural Science," *British Journal for the Philosophy of Science* **5** (1954), .

[566] ———, *Calculus: A Modern Approach*, Ginn and Company, Boston, 1955.

[567] ———, "Random Variables from the Point of View of a General Theory of Variables," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability II*, J. Neyman, ed., University of California Press, Berkeley and Los Angeles, 1956, .

[568] ———, "Mensuration and Other Mathematical Connections of Observable Material," in *Measurement: Definitions and Theories*, C. W. Churchman & P. Ratoosh, eds., John Wiley and Sons, New York, 1959, .

[569] ———, "Variables, Constants, Fluents," in *Current Issues in the Philosophy of Science*, H. Feigl & G. Maxwell, eds., New York, 1961, , (Reprinted in *Karl Menger: Selected Papers in Logic and Foundations, Dialectics, Economics*, D. Riedel, Dordrecht, 1979).

[570] ———, "Mathematical Implications of Mach's Ideas: Positivistic Geometry, The Clarification of Functional Connections," in *Ernst Mach Physicist and Philosopher*, R. S. Cohen & R. J. Seeger, eds., D. Riedel, Dordrecht, 1970, , (Boston Studies in the Philosophy of Science, VI).

[571] ———, "Geometry and Positivism: A Probabilistic Microgeometry," in *Karl Menger: Selected Papers in Logic and Foundations, Dialectics, Economics*, D. Riedel, Dordrecht, 1979, .

[572] ———, "A Counterpart of Occam's Razor," in *Karl Menger: Selected Papers in Logic and Foundations, Dialectics, Economics*, D. Riedel, Dordrecht, 1979, .

[573] N. Metropolis, "Methods of Significance Arithmetic," in *The State of the Art in Numerical Analysis*, D. Jacobs, ed., Academic Press, London, 1977, .

[574] ———, "Summation of Imprecise Numbers," *Computers and Mathematics with Applications* **6** (1980), .

[575] H. Milicer-Grużewska, "L'arithétique des variables aléatories ," *Cahiers Rhodaniens* **6** (1954), .

[576] A. C. Miller III, M. M. Merkhofer, R. A. Howard, J. E. Matheson & T. R. Rice, "Development of Automated Aids for Decision Analysis," Final technical report DO #27742, Stanford Research Institute, Menlo Park, California, 1976.

[577] R. von Mises, *Probability, Statistics and Truth (2nd revised English edition)*, George Allen and Unwin Ltd., London, 1957.

[578] M. Miyakoshi & M. Shimbo, "A Strong Law of Large Numbers for Fuzzy Random Variables," *Fuzzy Sets and Systems* **12** (1984), .

[579] M. Mizumota & K. Tanaka, "The Four Operations of Arithmetic on Fuzzy Numbers," *Systems, Computers and Controls* **7** (1976), .

[580] M. Mizumoto, "Pictorial Representations of Fuzzy Connectives, Part 1: Cases of t-norms, t-conorms and Averaging Operators," *Fuzzy Sets and Systems* **31** (1989), .

[581] C. A. Montgomery, "The Active Information System: A Data-Driven System for the Analysis of Imprecise Data," in *Proceedings of the 7th International Conference on Very Large Data Bases (Cannes, France, 1981)*, IEEE Computer Press, Los Angeles, 1981, .

[582] A. S. Moore, "Interval Risk Analysis of Real Estate Investment: A Non-Monte Carlo Approach," *Freiburger Intervall-Berichte* **85/3** (1985), .

[583] R. E. Moore, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1966.

[584] ———, *Methods and Applications of Interval Analysis*, SIAM, Philadelphia, 1979.

[585] ———, "Risk Analysis without Monte Carlo Methods," *Freiburger Intervall-Berichte* **84/1** (1984), .

[586] P. A. P. Moran, "Path Coefficients Reconsidered," *Australian Journal of Statistics* **3** (1961), .

[587] ———, *An Introduction to Probability Theory*, Clarendon Press, Oxford, 1968.

[588] C. G. Morgan, "Probabilistic Semantics for Propositional Modal Logics," in *Essays in Epistemology and Semantics*, H. Leblanc, R. Stern & R. Gumb, eds., Haven Publications, New York, 1983, .

[589] ———, "The Truebeliever, the Unbeliever, and a New Modal Semantics," in *Essays in Epistemology and Semantics*, H. Leblanc, R. Stern & R. Gumb, eds., Haven Publications, New York, 1983, .

[590] N. Morrison, "Smoothing and Extrapolation of Time Series by Means of Discrete Laguerre Polynomials," *SIAM Journal of Applied Mathematics* **15** (1967), .

[591] ———, "Smoothing and Extraploation of Continuous Time Series using Laguerre Polynomials," *SIAM Journal of Applied Mathematics* **16** (1968), .

[592] R. Moynihan, "On the Class of τ_T Semigroups of Probability Distribution Functions," *Aequationes Mathematicae* **12** (1975), .

[593] ———, "Infinite τ_T products of Distribution Functions," *Journal of the Australian Mathematical Society (Series A)* **26** (1978), .

[594] ———, "On τ_T Semigroups of Probability Distribution Functions II," *Aequationes Mathematicae* **17** (1978), .

[595] ———, "Conjugate Transforms and Limit Theorems for τ_T Semigroups," *Studia Mathematica* **69** (1980), .

[596] ———, "Conjugate Transforms for τ_T Semigroups of Probability Distribution Functions," *Journal of Mathematical Analysis and Applications* **74** (1980), .

[597] R. Moynihan, B. Schweizer & A. Sklar, "Inequalities Among Operations on Probability Distribution Functions," in *General Inequalities I, Proceedings of the 1st International Conference on General Inequalities 1976*, E. F. B. *et al*, ed., Birkäuser, Basel, 1978, .

[598] B. Muckenhoupt, "Mean Convergence of Hermite and Laguerre Series I," *Transactions of the American Mathematical Society* **147** (1970), .

[599] ———, "Conjugate Functions for Laguerre Expansions," *Transactions of the American Mathematical Society* **147** (1970), .

[600] ———, "Mean Convergence of Hermite and Laguerre Series II," *Transactions of the American Mathematical Society* **147** (1970), .

[601] A. Muir, "Fuzzy Sets and Probability," *Kybernetes (London)* **10** (1981), .

[602] ———, "Relations Between Probability and Fuzzy Theory," in *The Mathematics of Fuzzy Systems*, A. DiNola & A. G. S. Ventre, eds., Verlag Tüv Rheinland, Köln, 1986, , Interdisciplinary Systems Research ISR88.

[603] J. W. Müller, "Some Second Thoughts on Error Statements," *Nuclear Instruments and Methods* **163** (1979), .

[604] ———, "The Assignment of Uncertainties to the Results of Experimental Measurements," in *Precision Measurement and Fundamental Constants II*, B. N. Taylor & W. D. Phillips, eds., National Bureau of Standards (United States) Special Publication 617, 1984, .

[605] E. J. Muth, "The Discrete Laguerre Polynomials and their use in Exponential Smoothing," *IEE Transactions* **15** (1983), .

[606] S. Nahmias, "Fuzzy Variables," *Fuzzy Sets and Systems* **1** (1978), .

[607] U. S. Nair, "The Application of the Moment Function in the Study of Distribution Laws in Statistics," *Biometrika* **30** (1938), .

[608] I. Najfeld, "Linear Transformations of Vector Sets and their Applications," in *Interval Mathematics 1980*, K. L. E. Nickel, ed., Academic Press, New York, 1980, .

[609] I. Najfeld, R. A. Vitale & P. J. Davis, "Minkowski Iteration of Sets," *Linear Algebra and its Applications* **29** (1980), .

[610] P. S. Nasimov, "Über eine Guass'sche Abhandlung," *Nachr. Warsh. Univ.* **4** (1889), Cited as [109] by Sheynin, *Arch. Hist. Exact Sci.* 20, 21-72 (1979).

[611] B. Natvig, "Possibility Versus Probability," *Fuzzy Sets and Systems* **10** (1983), .

[612] R. B. Nelsen, "Discrete Bivariate Distributions with Given Marginals and Correlation," *Communications in Statistics — Simulation* **16** (1987), .

[613] A. J. Nelson & D. M. Rasmuson, "A Correction to S. Kaplan's Method of Discrete Probability Distributions," *Risk Analysis* **2** (1982), .

[614] H. Nguyen, "On Random Sets and Belief Functions," *Journal of Mathematical Analysis and Applications* **65** (1978), .

[615] H. T. Nguyen, "A Note on the Extension Principle for Fuzzy Sets," *Journal of Mathematical Analysis and Applications* **64** (1978), .

[616] W. L. Nicholson, "On the Distribution of 2×2 Random Normal Determinants," *Annals of Mathematical Statistics* **29** (1958), .

[617] K. Nickel, "Zeros of Polynomials and Other Topics," in *Topics in Interval Analysis*, E. Hansen, ed., Clarendon Press, Oxford, 1969, .

[618] ———, "Triplex-Algol and its Applications," in *Topics in Interval Analysis*, E. Hansen, ed., Oxford University Press, Oxford, 1969, .

[619] J. Nieminen, "On the Algebraic Structure of Fuzzy Sets of Type 2," *Kybernetika* **13** (1977), .

[620] N. J. Nilsson, "Probabilistic Logic," *Artificial Intelligence* **28** (1986), .

[621] A. W. Niukkanen, "Generalised Hypergeometric Series ${}^{N}F(x_1,\ldots,x_N)$ arising in Physical and Quantum Chemical Applications," *Journal of Physics A: Mathematical and General* **16** (1983), .

[622] ———, "Clebsch-Gordon-type Linearisation Relations for the Products of Laguerre Polynomials and Hydrogen-like Functions," *Journal of Physics A: Mathematical and General* **18** (1985), .

[623] J. Norton, "Limit Theorems for Dempster's Rule of Combination," *Theory and Decision* **25** (1988), .

[624] Y. Nurges & Y. Yaaksoo, "Laguerre State Equations for a Multivariate Discrete System," *Automation and Remote Control* **42** (1981), .

[625] H. J. Nussbaumer, *Fast Fourier Transform and Convolution Algorithms (2nd Edition)*, Springer-Verlag, Berlin, 1982.

[626] H. Nyquist, S. O. Rice & J. Riordan, "The Distribution of Random Determinants," *Quarterly of Applied Mathematics* **12** (1954), .

[627] D. Oakes, "A Model for Association in Bivariate Survival Data," *Journal of the Royal Statistical Society, Series B* **44** (1982), .

[628] J. K. Ord, *Families of Frequency Distributions*, Griffin, London, 1972.

[629] R. H. Orth & A. R. Ferguson, eds., *The Journals and Miscellaneous Notebooks of Ralph Waldo Emerson*, Belknap Press, Harvard, 1971.

[630] D. Pang, J. Bigham & E. H. Mamdani, "Reasoning with Uncertain Information," *IEE Proceedings* **134** (1987), .

[631] A. Papamarcou & T. L. Fine, "A Note on Undominated Lower Probabilities," *The Annals of Probability* **14** (1986), .

[632] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.

[633] ———, "Error Analysis in Sampling Theory," *Proceedings of the IEEE* **54** (1966), .

[634] G. Parisi & N. Sourlas, "A Simple Parametrization of the Q^2 Dependence of the Quark Distributions in QCD," *Nuclear Physics* **B151** (1979), .

[635] J. K. Patel, C. H. Kapadia & D. B. Owen, *Handbook of Statistical Distributions*, Marcel Dekker, New York and Basel, 1976.

[636] R. S. Pathak, "Summability of the Laguerre Polynomial Expansion of Generalized Functions," *Journal of the Institute of Mathematics and its Applications* **21** (1978), .

[637] J. Pearl, "Fusion, Propagation, and Structuring in Belief Networks," *Artificial Intelligence* **29** (1986), .

[638] ———, "On Evidential Reasoning in a Hierarchy of Hypotheses," *Artificial Intelligence* **28** (1986), .

[639] ———, "Evidential Reasoning Using Stochastic Simulation of Causal Models," *Artificial Intelligence* **32** (1987), .

[640] ———, "Distributed Revision of Composite Beliefs," *Artificial Intelligence* **33** (1987), .

[641] ———, "A Constraint-Propagation Approach to Probabilistic Reasoning," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[642] ———, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, Los Altos, California, 1988.

[643] J. Pearl & M. Tarsi, "Structuring Causal Trees," *Journal of Complexity* **2** (1986), .

[644] E. S. Pearson, "Some Problems arising in Approximating to Probability Distributions using Moments," *Biometrika* **50** (1963), .

[645] K. Pearson, "Contributions to the Theory of Evolution: On a Form of Spurious Correlation which may arise when Indices are used in the Measurement of Organs," *Proceedings of the Royal Society of London* **60** (1897), .

[646] M. S. Petković, "A Representation of Random Variables by the Probable Intervals," *Freiburger Intervall-Berichte* **80/10** (1980), .

[647] P. Petrina, R. Sexsmith & R. N. White, *Safety Analysis of Nuclear Concrete Containment Structures*, U.S. Nuclear Regulatory Commision, Washington, NUREG/CR-1097, RD, 1979.

[648] M. Pichat, "All Possible Computed Results in Correct Floating-Point Summation," *Mathematics and Computers in Simulation* **30** (1988), .

[649] A. Plantinga, "De Re et De Dicto," *Noûs* **3** (1969), .

[650] H. Poincaré, "Calcul des Probabilités," Paris, 1896.

[651] ———, *Science and Hypothesis*, The Science Press, Lancaster, 1905.

[652] ———, *The Foundations of Science*, The Science Press, Lancaster, 1913.

[653] S. D. Poisson, "Mémoire sur la probabilité du tir a la cible," *Mémorial de l'artillerie* **4** (1837), .

[654] K. R. Popper, "The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory," in *Observation and Interpretation: A Symposium of Philosophers and Physicists*, S. Körner, ed., Butterworths Scientific Publications, London, 1957, .

[655] ———, "The Propensity Interpretation of Probability," *British Journal for the Philosophy of Science* **10** (1959), .

[656] ———, *Objective Knowledge: An Evolutionary Approach*, Oxford University Press, Oxford, 1972.

[657] ———, *Conjectures and Refutations*, Routledge and Kegan Paul, London, 1972.

[658] ———, *The Logic of Scientific Discovery (4th Ed.)*, Hutchinson, London, 1980.

[659] ———, *The Open Universe: An Argument for Indeterminism*, Rowman and Littlefield, Totowa, New Jersey, 1982.

[660] ———, *Realism and the Aim of Science*, Rowman and Littlefield, Totawa, New Jersey, 1983.

[661] B. Prabhala & R. Sethi, "Efficient Computation of Expressions with Common Subexpressions," *Journal of the Association of Computing Machinery* **27** (1980), .

[662] H. Prade, "Modal Semantics and Fuzzy Set Theory," in *Fuzzy Set and Possibility Theory: Recent Developments*, R. R. Yager, ed., Pergamon Press, New York, 1982, .

[663] R. D. Prasad, "Probability Distributions of Algebraic Functions of Independent Random Variables," *SIAM Journal of Applied Mathematics* **18** (1970), .

[664] A. Prékopa, "On Random Determinants I," *Studia Scientiarum Mathematicarum Hungarica* **2** (1967), .

[665] Y. V. Prohorov & Y. A. Rozanov, *Probability Theory: Basic Concepts, Limit Theorems, Random Processes*, Springer-Verlag, Berlin, 1969.

[666] J. J. Quesada-Molina, "Measures of Dependence Based on the Empirical Copula," *Aequationes Mathematicae* **37** (1989), , (Report of the Twenty-sixth International Symposium on Functional Equations).

[667] J. R. Quinlan, "Inferno: A Cautious Approach to Uncertain Inference," *The Computer Journal* **26** (1983), .

[668] A. L. Ralescu & D. A. Ralescu, "Probability and Fuzziness," *Information Sciences* **34** (1984), .

[669] D. Ralescu, "A Survey of the Representation of Fuzzy Concepts and its Applications," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade & R. R. Yager, eds., North-Holland, Amsterdam, 1979, .

[670] ———, "Toward a General Theory of Fuzzy Variables," *Journal of Mathematical Analysis and Applications* **86** (1982), .

[671] A. Ramakrishnan, "A New Concept in Probability Theory," *Journal of Mathematical Analysis and Applications* **83** (1981), .

[672] G. P. Ramsey, "The Laguerre Method for Solving Integro-Differential Equations," *Journal of Computational Physics* **60** (1985), .

[673] M. B. Rao & A. Rashed, "Some Comments on Fuzzy Variables," *Fuzzy Sets and Systems* **6** (1981), .

[674] P. N. Rathie & H. Kaufman, "On the Distribution of Products and Quotients of Independent Random Variables," *Metron* **35** (1977), .

[675] H. Ratschek, "Die Subdistributivität in der Intervallarithmetik," *Zeitschrift für Angewandte Mathematik und Mechanik* **51** (1971), .

[676] ———, "Representation of Interval Operations by Coordinates," *Computing* **24** (1980), .

[677] N. S. Rau, P. Toy & K. F. Schenk, "Expected Energy Production Costs by the Method of Moments," *IEEE Transaction on Power Apparatus and Systems* **99** (1980), .

[678] R. L. Redmond, "Egg Size and Laying Date of Long-Billed Curlews Numenius Aamericanus: Implications for Female Reproductive Tactics," *Oikos* **46** (1986), .

[679] A. Rege & A. M. Agogino, "Sensor-Integrated Expert Systems for Manufacturing and PRocess Diagnostics," in *Knowledge-Based Expert Systems for Manufacturing*, S. C. -Y. Lu & R. Komanduri, eds., American Society of Mechanical Engineers, 1986, .

[680] J. H. Reif & H. R. Lewis, "Efficient Symbolic Analysis of Programs," *Journal of Computer and System Sciences* **32** (1986), .

[681] J. H. Reif & R. E. Tarjan, "Symbolic Program Analysis in Almost-Linear Time," *SIAM Journal of Computing* **11** (1982), .

[682] A. Rényi, *Foundations of Probability*, Holden-Day, San Francisco, 1970.

[683] J. Richards, "Boole and Mill: Differing Perspectives on Logical Psychologism," *History and Philosophy of Logic* **1** (1980), .

[684] H. Robbins, "A Remark on the Joint Distribution of Cumulative Sums," *Annals of Mathematical Statistics* **25** (1954), .

[685] P. Robillard & M. Traham, "The Completion Time of PERT Networks," *Operations Research* **25** (1977), .

[686] G. K. Robinson, "Confidence Intervals and Regions," in *Encyclopedia of Statistical Sciences*, N. L. Johnson & S. Kotz, eds., John Wiley and Sons, New York, 1982, .

[687] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[688] N. Rosen, "Statistical Geometry and Fundamental Particles," *Physical Review* **72** (1947), .

[689] ———, "Quantum Geometry," *Annals of Physics* **19** (1962), .

[690] A. Rosenthal, "A Computer Scientist Looks at Reliability Computations," in *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assesment*, R. E. Barlow, J. B. Fussell & N. D. Singpurwalla, eds., SIAM, Philadelphia, 1975, .

[691] L. Rothschild, "The Distribution of English Dictionary Wordlengths," *Journal of Statistical Planning and Inference* **14** (1986), .

[692] N. C. Rowe, "Absolute Bounds on the Mean and Standard Deviation of Transformed Data for Constant-Sign-Derivative Transformations," *SIAM Journal of Scientific and Statistical Computation* **9** (1988), .

[693] L. Rüschendorf, "Sharpness of Fréchet Bounds," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **42** (1978), .

[694] ———, "Inequalities for the Expectation of Δ-Monotone Functions," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **54** (1980), .

[695] ———, "Random Variables with Maximum Sums," *Advances in Applied Probability* **14** (1982), .

[696] ———, "Solution of Statistical Optimization Problem by Rearrangement Methods," *Metrika* **30** (1983), .

[697] ———, "Construction of Multivariate Distributions with Given Marginals," *Annals of the Institute of Statistical Mathematics* **37A** (1985), .

[698] ———, "Monotonicity and Unbiasedness of Tests via a.s. Constructions," *Statistics* **17** (1986), .

[699] E. H. Ruspini, "Possibility Theory Approaches for Advanced Information Systems," *IEEE Computer* **15(9)** (1982), .

[700] ———, "Approximate Inference and Interval Probabilities," in *Uncertainty in Knowledge-Based Systems*, B. Bouchon & R. R. Yager, eds., Springer-Verlag, Berlin, 1987, , (Lecture Notes in Computer Science 286).

[701] B. Russell, "Vagueness," *Australasian Journal of Philosophy* **1** (1923), .

[702] R. A. Sahner & K. S. Trivedi, "Performance and Reliability Analysis Using Directed Acyclic Graphs," *IEEE Transactions on Software Engineering* **13** (1987), .

[703] M. Sakai, "A Characterization of a Class of Polynomial Splines by Convolution," *BIT* **26** (1986), .

[704] W. Salmon, "Probabilistic Causality," *Pacific Philosophical Quarterly* **61** (1980), .

[705] T. Salvemini, "Sugli indici di omofilia," *Atti I Riunione della Societa Italiana di Statistica* (1939).

[706] S. Sambursky, *The Physical World of the Greeks*, Routledge and Kegan Paul, London, 1956.

[707] ———, "On the Possible and Probable in Ancient Greece," *Osiris* **12** (1956), .

[708] S. K. Sarkar & W. Smith, "Probability Inequalities for Ordered MTP_2 Random Variables," *Sankhyā* **48A** (1986), .

[709] O. V. Sarmanov & Z. N. Bratoeva, "Probabilistic Properties of Bilinear Expansions of Hermite Polynomials," *Theory of Probability and its Applications* **12** (1967), .

[710] M. Sarna, "Fuzzy Relations on Fuzzy and Non-fuzzy Numbers — Fast Computation Formulas," *Fuzzy Sets and Systems* **29** (1989), .

[711] S. C. Saunders, "Birnbaum's Contributions to Reliability Theory," in *Reliability and Fault Tree Analysis: Theoretical And Applied Aspects of Systems Reliability and Safety Assessment*, R. E. Barlow, J. B. Fussell & N. D. Singpurwalla, eds., Society for Industrial and Applied Mathematics, Philadelphia, 1975, .

[712] R. Scheines, "Automating Creativity," in *Aspects of Artificial Intelligence*, J. H. Fetzer, ed., Kluwer Academic Publishers, Dordrecht, 1988, .

[713] I. J. Schoenberg, "Splines and Histograms," in *Spline Functions and Approximation Theory*, A. Meir & A. Sharma, eds., Birkhäuser, Basel, 1973, .

[714] K. Schuessler, "Ratio Variables and Path Models," in *Structural Equation Models in the Social Sciences*, A. S. Goldberger & O. D. Duncan, eds., Seminar Press, New York, 1973, .

[715] B. Schweizer, "Distribution Functions: Numbers of the Future," in *Proceedings 2nd Napoli Meeting on "The Mathematics of Fuzzy Systems"*, A. Di Nola & A. G. S. Ventre, eds., Istituto di Matematica della Facoltà di Architettura Università degli Studi di Napoli, 1985.

[716] ———, "Remark," *Aequationes Mathematicae* **35** (1988), , (Report of the Twenty-fifth International Symposium on Functional Equations).

[717] B. Schweizer & A. Sklar, "Operations on Distribution Functions not Derivable from Operations on Random Variables," *Studia Mathematica* **52** (1974), .

[718] ———, *Probabilistic Metric Spaces*, North-Holland, New York, 1983.

[719] B. Schweizer & E. F. Wolff, "On Nonparametric Measures of Dependence for Random variables," *Annals of Statistics* **9** (1981), .

[720] D. Scott & P. Krauss, "Assigning Probabilities to Logical Formulas," in *Aspects of Inductive Logic*, J. Hintikka & P. Suppes, eds., North-Holland, Amsterdam, 1966, .

[721] H. L. Seal, "The Historical Use of Generating Functions in Probability Theory," *Bulletin de l'Association des Actuaires suisses* **49** (1949), , (Reprinted in M. Kendall and R.L. Plackett, eds, *Studies in the History of Statistics and Probablity*, Volume II, Charles Griffin, London, 67-86, 1977).

[722] R. Sedgewick, *Quicksort*, Garland, New York, 1978.

[723] T. Seidenfeld, *Philosophical Problems of Statistical Inference*, D. Riedel, Dordrecht, 1979.

[724] ———, "On After-Trial Properties of Best Neyman-Pearson Confidence Intervals," *Philosophy of Science* **48** (1981), .

[725] R. P. Selke, "A Rewriting Semantics for Program Dependence Graphs," in *Conference Record of the Sixteenth Annual ACM Symposium on Principles of Programming Languages*, Association for Computing Machinery, 1989, .

[726] W. Sendler, "On the Second Moments of Random Determinants," *Studia Scientiarum Mathematicarum Hungarica* **9** (1974), .

[727] B. Sendov, "Some Topics of Segment Analysis," in *Interval Mathematics 1980*, K. L. E. Nickel, ed., Academic Press, New York, 1980, .

[728] R. Sethi & J. D. Ullman, "The Generation of Optimal Code for Arithmetic Expressions," *Journal of the Association for Computing Machinery* **17** (1970), .

[729] R. D. Shachter, "Evaluating Influence Diagrams," *Operations Research* **34** (1986), .

[730] ———, "Intelligent Probabilistic Inference," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[731] ———, "Probabilistic Inference and Influence Diagrams," *Operations Research* (1987), Submitted.

[732] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.

[733] ———, "A Theory of Statistical Evidence," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Volume II*, W. L. Harper & C. A. Hooker, eds., D. Riedel, Dordrecht, 1976, .

[734] ———, "Non-Additive Probabilities in the Work of Bernoulli and Lambert," *Archive for History of Exact Sciences* **19** (1978), .

[735] ———, "Allocations of Probability," *Annals of Probability* **7** (1979), .

[736] ———, "Nonadditive Probability," in *Encyclopedia of Statistical Sciences*, N. L. Johnson & S. Kotz, eds., John Wiley and Sons, New York, 1982.

[737] ———, "The Combination of Evidence," *International Journal of Intelligent Systems* **1** (1986), .

[738] ———, "Belief Functions and Possibility Measures," in *Analysis of Fuzzy Information, Volume 1: Mathematics and Logic*, J. C. Bezdek, ed., CRC Press Incorporated, Boca Raton, Florida, 1987, .

[739] G. Shafer & R. Logan, "Implementing Dempster's Rule for Hierarchical Evidence," *Artificial Intelligence* **33** (1987), .

[740] M. Shah, "Some Results on the H-function Involving the Generalized Laguerre Polynomials," *Proceedings of the Cambridge Philosophical Society* **65** (1969), .

[741] M. Shaked, "A Family of Concepts of Dependence for Bivariate Distributions," *Journal of the American Statistical Association* **72** (1977), .

[742] S. Shao, "Fuzzy Self-Organizing Controller and its Application for Dynamic Processes," *Fuzzy Sets and Systems* **26** (1988), .

[743] H. Sherwood & M. D. Taylor, "Some PM Structures on the set of Distribution Functions," *Revue Roumaine de Mathématiques Pures et Appliquées* **19** (1974), .

[744] O. B. Sheynin, "Newton and the Classical Theory of Probability," *Archive for History of Exact Sciences* **7** (1970-71), .

[745] ———, "Finite Random Sums (A Historical Essay)," *Archive for History of Exact Sciences* **9** (1972), .

[746] ———, "D. Bernoulli's work on Probability," *RETE Strukturgeschichte der Naturwissenschaften* **1** (1972), , (Reprinted in M. Kendall and R.L. Plackett, eds, *Studies in the History of Statistics and Probablity*, Volume II, Charles Griffin, London, 1977).

[747] ———, "Mathematical Treatment of Astronomical Observations (A Historical Essay)," *Archive for History of Exact Sciences* **11** (1973), .

[748] ———, "S.D. Poisson's Work in Probability," *Archive for History of Exact Sciences* **18** (1977), .

[749] ———, "C.F. Gauss and the Theory of Errors," *Archive for History of Exact Sciences* **20** (1979), .

[750] G. E. Shilov & B. L. Gurevich, *Integral, Measure and Derivative: A Unified Approach*, Dover, New York, 1977.

[751] A. W. Shogan, "Bounding Distributions for Stochastic PERT Network," *Networks* **7** (1977), .

[752] J. A. Shohat & J. E. Tamarkin, *The Problem of Moments*, American Mathematical Society, New York, 1943.

[753] R. M. Shortt, "Strassen's Marginal Problem in Two or More Dimensions," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **64** (1983), .

[754] G. E. Sigal, A. A. B. Pritsker & J. Solberg, "The Stochastic Shortest Route Problem," *Operations Research* **28** (1980), .

[755] B. W. Silverman, "Kernel Density Estimation using the Fast Fourier Transform," *Applied Statistics* **31** (1982), .

[756] ———, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London and New York, 1986.

[757] H. A. Simon, "Spurious Correlation: A Causal Interpretation," *Journal of the American Statistical Association* **49** (1954), .

[758] C. E. Skinner, "A Numerical Inversion of the Pollaczek-Khintchine Formula," in *33rd National Meeting of the Operations Research Society of America, San Francisco, May 1968*, 1968, Brief summary in Bulletin of the Operations Research Society of America 16, Supplement 1, B-159.

[759] A. Sklar, "Random Variables, Joint Distribution Functions, and Copulas," *Kybernetika* **9** (1973), .

[760] B. Skyrms, "Higher Order Degrees of Belief," in *Prospects for Pragmatism (Essays in Honour of F.P. Ramsey)*, D. H. Mellor, ed., Cambridge University Press, Cambridge, 1980, .

[761] P. Smets, "Belief Functions versus Probability Functions," in *Uncertainty and Intelligent Systems*, B. Bouchon, L. Saitta & R. R. Yager, eds., Springer-Verlag, Berlin, 1988, , (Lecture Notes in Computer Science 313).

[762] C. A. B. Smith, "Consistency in Statistical Inference and Decision," *Journal of the Royal Statistical Society Series B* **23** (1961), , (with discussion).

[763] ———, "Personal Probability and Statistical Analysis," *Journal of the Royal Statistical Society Series A* **128** (1965), , (with discussion).

[764] G. C. Smith, *The Boole–De Morgan Correspondence 1842–1864*, Clarendon Press, Oxford, 1982.

[765] J. M. Smith, F. W. J. Olver & D. W. Lozier, "Extended-Range Arithmetic and Normalized Legendre Polynomials," *ACM Transactions on Mathematical Software* **7** (1981), .

[766] J. Q. Smith, "Belief Nets and Influence Diagrams: What they are and how they work," Report 155, Department of Statistics, University of Warwick (to appear in the Proceedings of a Conference on Influence Diagrams held at the University of California at Berkeley in May 1988), 1988.

[767] ———, "Influence Diagrams for Statistical Modelling," Report 117, Department of Statistics, University of Warwick (to appear in Annals of Statistics), 1988.

[768] ———, "Statistical Principles on Graphs," Report 138, Department of Statistics, University of Warwick (to appear in the Proceedings of a Conference on Influence Diagrams held at the University of California at Berkeley in May 1988), 1988.

[769] M. Sobierajski, "Probabilistic Analysis of Load Flow Equations," *Archiv für Electrotechnik* **69** (1986), .

[770] O. Spaniol, "Die Distributivität in der Intervallarithmetik," *Computing* **5** (1970), .

[771] A. Špaček, "Note on K. Menger's Probabilistic Geometry," *Czechoslavak Mathematics Journal* **6** (1956), .

[772] ———, "Random Metric Spaces," in *Transactions of the Second Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (June 1959)*, 1960, .

[773] T. P. Speed, "Probabilistic Risk Assessment in the Nuclear Industry: WASH-1400 and Beyond," in *Proceedings of the Berkeley Conference in Honour of Jerzy Neyman and Jack Kiefer, Volume 1*, L. M. L. Cam & R. A. Olshen, eds., Wadsworth Advanced Books and Software, Monterey, California, 1985, .

[774] S. Spielman, "A Refutation of the Neyman-Pearson Theory of Testing," *British Journal for the Philosophy of Science* **24** (1973), .

[775] M. D. Springer, *The Algebra of Random Variables*, John Wiley and Sons, New York, 1979.

[776] H. M. Srivastava, "On the Product of Two Laguerre Polynomials," *Bollettino della Unione Matematica Italiana* **(4)5** (1972), .

[777] H. M. Srivastava & R. G. Buschman, "Mellin Convolutions and H-function Transformations," *Rocky Mountain Journal of Mathematics* **6** (1976), .

[778] W. Stallings, "Fuzzy Set Theory Versus Bayesian Statistics," *IEEE Transactions on Systems, Man, and Cybernertics* **7** (1977), .

[779] I. M. Staneu-Minasian & M. J. Wets, "A Research Bibliography in Stochastic Programming 1955-1975," *Operations Research* **24** (1976), 1078.

[780] R. L. Stens, "Error Estimates for Sampling Sums Based on Convolution Integrals," *Information and Control* **45** (1980), .

[781] V. Strassen, "The Existence of Probability Measures with Given Marginals," *Annals of Mathematical Statistics* **36** (1965), .

[782] R. H. Stratz & H. O. A. Wold, "Recursive versus Nonrecursive Systems," *Econometrica* **28** (1960), .

[783] J. P. Stremel, R. T. Jenkins, R. A. Sabb & W. D. Bayless, "Production Costing Using the Cumulant Method of Representing the Equivalent Load Curve," *IEEE Transactions on Power Apparatus and Systems* **99** (1980), .

[784] F. Stummel, "Perturbation Theory for Evaluation Algorithms of Arithmetic Expressions," *Mathematics of Computation* **37** (1981), .

[785] ———, "Rounding Errors in Numerical Solutions of Two Linear Equations in Two Unknowns," *Mathematical Methods in the Applied Sciences* **4** (1982), .

[786] ———, "Rounding Error Analysis of Interval Algorithms," *Zeitschrift für Angewandte Mathematik und Mechanik* **64** (1984), .

[787] K. Subrahmaniam, "On Some Applications of Mellin Transforms to Statistics: Dependent Random Variables," *SIAM Journal of Applied Mathematics* **19** (1970), .

[788] M. Sugeno, "An Introductory Survey of Fuzzy Control," *Information Sciences* **36** (1985), .

[789] ———, "Editorial," *Fuzzy Sets and Systems* **26** (1988), 149.

[790] U. Sumita, "Numerical Evaluation of Multiple Convolutions, Survival Functions, and Renewal Functions for the One-Sided Normal Distribution and the Rayleigh Distribution via the Laguerre Transformation," Working Paper Series No. 7912, The Graduate School of Management, The University of Rochester, 1978.

[791] ———, "Development of the Laguerre Transform Method for Numerical Exploration of Applied Probability Models," PhD thesis, University of Rochester, 1981.

[792] ———, "On Sums of Independent Logistic and Folded Logistic Variants," *Journal of Statistical Computation and Simulation* **17** (1983), .

[793] ———, "The Matrix Laguerre Transform," *Applied Mathematics and Computation* **15** (1984), .

[794] ———, "The Laguerre Transform and a Family of Functions with Nonnegative Laguerre Coefficients," *Mathematics of Operation Research* **9** (1984), .

[795] U. Sumita & M. Kijima, "On Optimal Bulk Size of Single-Server Bulk-Arrival Queueing Systems with Set-up Times — Numerical Exploration via the Laguerre Transform," Working Paper Series No. QM8412, The Graduate School of Management, The University of Rochester, 1984.

[796] ———, "The Bivariate Laguerre Transform and Its Applications: Numerical Exploration of Bivariate Processes," *Advances in Applied Probability* **17** (1985), .

[797] ———, "The Laguerre Transform of a Product of Functions with Applications in Alternating Renewal Theory," Working Paper Series No. QM8517, The Graduate School of Management, The University of Rochester, 1985.

[798] ———, "Numerical Exploration of a Bivariate Lindley Process via the Bivariate Laguerre Transform," Working Paper Series No. QM8503, The Graduate School of Management, The University of Rochester, January 1985.

[799] ———, "Evaluation of Minimum and Maximum of a Correlated Pair of Random Variables via the Bivariate Laguerre Transform," Working Paper Series No. QM8522, The Graduate School of Management, The University of Rochester, June 1985.

[800] T. Sunaga, "Theory of Interval Algebra and its Application to Numerical Analysis," *RAAG Memoirs (Resarch Association of Applied Geometry)* **2** (1958), .

[801] P. Suppes, "Probabilistic Inference and the Concept of Total Evidence," in *Aspects of Inductive Logic*, J. Hintikka & P. Suppes, eds., North-Holland, Amsterdam, 1966, .

[802] ———, *A Probabilistic Theory of Causality*, (Acta Philosophica Fennica 24), North-Holland, Amsterdam, 1970.

[803] ———, "Probabilsitic Metaphysics," *Filosofiska Studier utgivna av Filosofiska Föreningen och Filosofiska Institutionen vid Uppsala Universitet* **22** (1974), (two volumes).

[804] ———, "The Measurement of Belief," *Journal of the Royal Statistical Society Series B* **36** (1974), , (with discussion).

[805] ———, "Testing Theories and the Foundations of Statistics," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Volume II*, W. L. Harper & C. A. Hooker, eds., D. Riedel, Dordrecht, 1976, .

[806] ———, "Probabilistic Metaphysics," *Basil Blackwell*, Oxford (1984).

[807] P. Suppes & M. Zanotti, "On Using Random Relations to Generate Upper and Lower Probabilities," *Synthese* **36** (1977), .

[808] A. A. Sveshnikov, *Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions*, Dover, New York, 1978.

[809] J. J. Sylvester, "On a Special Class of Questions on the Theory of Probabilities," *Birmingham British Association Report* (1865), Reprinted as pp.480–481 in volume II of his collected papers.

[810] G. Szegö, *Orthogonal Polynomials, 4th edition*, American Mathematical Society Colloquium Publication 23, Providence, Rhode Island, 1975.

[811] T. Szulc, "On Two Classes of Interval Matrices," *Freiburger Intervall-Berichte* **85/2** (1985), .

[812] ———, "Sufficient Criteria for the Nonsingularity of Interval Matrices," *Freiburger Intervall-Berichte* **85/2** (1985), .

[813] T. Takagi & M. Sugeno, "Derivation of Fuzzy Control Rules from Human Operator's Control Actions," in *Proceedings of the IFAC Symposium on Fuzzy Information, Knoeldge Representation and Decision Analysis (1983)*, E. Sanchez, ed., IFAC (Pergamon Press), Oxford, 1984.

[814] W. Y. Tan & S. P. Wong, "On the Roy-Tiku Approximation to the Distribution of Sample Variances from Nonnormal Universes," *Journal of the American Statistical Association* **72** (1977), .

[815] ———, "On Approximating the Central and Noncentral Multivariate Gamma Distributions," *Communications in Statistics — Simulation and Computation* **B7** (1978), .

[816] H. Tanaka & H. Ishibuchi, "Identification of Possibilistic Linear Systems by Quadratic Membership Functions of Fuzzy Parameters," Submitted to *Fuzzy Sets and Systems*, 1989.

[817] K. L. Tang & R. J. Mulholland, "Comparing Fuzzy Logic with Classical Controller Design," *IEEE Transactions on Systems, Man, and Cybernetics* **17** (1987), .

[818] R. Tanscheit & E. M. Scharf, "Experiments with the use of a Rule-Based Self-Organising Controller for Robotics Applications," *Fuzzy Sets and Systems* **26** (1988), .

[819] R. E. Tarjan & M. Yannakakis, "Simple Linear-time Algorithms to Test Chordality of Graphs, Test Acyclicity of Hypergraphs, and Selectively Reduce Acyclic Hypergraphs," *SIAM Journal of Computing* **13** (1984), .

[820] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, University Science Books, Mill Valley, California, 1982.

[821] A. H. Tchen, "Inequalities for Distributions with Given Marginals," *Annals of Probability* **8** (1980), .

[822] P. S. Theocaris & A. C. Chrysakis, "Numerical Inversion of the Mellin Transform," *Journal of the Institute of Mathematics and its Applications* **20** (1977), .

[823] M. L. Tiku, "Laguerre Series forms of Non-Central χ^2 and F Distributions," *Biometrika* **52** (1965), .

[824] R. M. Tong & L. A. Appelbaum, "Experiments with Interval-Valued Uncertainty," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[825] H. Toth, "From Fuzzy-Set Theory to Fuzzy Set-Theory: Some Critical Remarks on Existing Concepts," *Fuzzy Sets and Systems* **23** (1987), .

[826] G. Tsamasphyros & P. S. Theocaris, "Numerical Inversion of Mellin Transforms," *BIT* **16** (1976), .

[827] A. A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelationstheorie*, Teubner, Leipzig, 1925, *Principles of the Mathematical Theory of Correlation*, William Hodge and Company, London, 1939.

[828] J. W. Tukey, "Causation, Regression, and Path Analysis," in *Statistics and Mathematics in Biology*, O. Kempthorne, T. A. Bancroft, J. W. Gowen & J. L. Lush, eds., Hafner Publishing Company, New York, 1954, .

[829] ———, "The Propagation of Errors, Fluctuations and Tolerances. Basic Generalized Formulas," Technical Report No. 10, Statistical Techniques Research Group, Section of Mathematical Statistics, Department of Mathematics, Princeton University, 1957.

[830] ———, "The Propagation of Errors, Fluctuations and Tolerances. 3 — An Exercise in Partial Differentiation," Technical Report No. 12, Statistical Techniques Research Group, Section of Mathematical Statistics, Department of Mathematics, Princeton University, 1958.

[831] ———, "The Propagation of Errors, Fluctuations and Tolerances — Supplementary Formulas," Technical Report No. 11, Statistical Techniques Research Group, Section of Mathematical Statistics, Department of Mathematics, Princeton University, 1958.

[832] M. E. Turner & C. D. Stevens, "The Regression Analysis of Causal Paths," *Biometrics* **15** (1959), .

[833] V. A. Ubhaya, "L_p Approximation be Quasi-convex and Convex Functions," *Journal of Mathematical Analysis and Applications* **139** (1989), .

[834] K. Urbanik, "Generalized Convolutions," *Studia Mathematica* **23** (1964), .

[835] ———, "A Counterexample on Generalized Convolutions," *Colloquium Mathematicum* **54** (1987), .

[836] S. Ursic, "Generalizing Fuzzy Logic Probabilistic Inferences," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[837] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.

[838] J. Valdes, R. E. Tarjan & E. L. Lawler, "The Recognition of Series-Parallel Digraphs," *SIAM Journal of Computing* **11** (1982), .

[839] J. Van Evra, "A Reassessment of George Boole's Theory of Logic," *Notre Dame Journal of Formal Logic* **18** (1977), .

[840] T. N. Verma, "Real Inversion Formula for the Distributional Generalised Laguerre Transform," *Proceedings of the National Academy of Sciences of India* **51A** (1981), .

[841] T. N. Verma & A. B. B. Lall, "A Convolution Formula for the Distributional Laguerre Transform," *Indian Journal of Mathematics* **24** (1982), .

[842] J. Vignes, "New Methods for Evaluating the Validity of the Results of Mathematical Computations," *Mathematics and Computers in Simulation* **20** (1978), .

[843] ———, "Review of Stochastic Approach to Round-Off Error Analysis and Its Applications," *Mathematics and Computers in Simulation* **30** (1988), .

[844] F. Voorbraak, "A Computationally Efficient Approximation of Dempster-Shafer Theory," *International Journal of Man-Machine Studies* **30** (1989), .

[845] M. Wagenknecht & K. Hartmann, "Application of Fuzzy Sets of Type 2 to the Solution of Fuzzy Equations Systems," *Fuzzy Sets and Systems* **25** (1988), .

[846] S. R. Wagner, "On the Quantitative Characterization of the Uncertainty of Experimental Results in Metrology," *PTB-Mitteilungen* **89** (1979), , Die Physikalisch-Technische Bundesanstalt.

[847] G. Wahba, "Interpolating Spline Methods for Density Estimation I: Equi-spaced Knots," *Annals of Statistics* **3** (1975), .

[848] A. Wald, "On a Statistical Generalization of Metric Spaces," *Proceedings of the National Academy of Science (United States of America)* **29** (1943), .

[849] P. Walley, "Belief Function Representations of Statistical Evidence," *Annals of Statistics* **15** (1987), .

[850] P. Walley & T. L. Fine, "Varieties of Modal (Classifacatory) and Comparative Probability," *Synthese* **41** (1979), .

[851] ———, "Towards a Frequentist Theory of Upper and Lower Probability," *Annals of Statistics* **10** (1982), .

[852] D. Waltz, "Understanding Line Drawings of Scenes with Shadows," in *The Psychology of Computer Vision*, P. Winston, ed., McGraw-Hill, New York, 1975.

[853] P. -Z. Wang, "Fuzzy Contactability and Fuzzy Variables," *Fuzzy Sets and Systems* **8** (1982), .

[854] W. Warmuth, "Marginal-Fréchet-Bounds for Multidimensional Distribution Functions," *Statistics* **19** (1988), .

[855] H. Weber, "Numerical Computation of the Fourier Transform using Laguerre Functions and the Fast Fourier Transform," *Numerische Mathematik* **36** (1981), .

[856] S. Weber, "A General Concept of Fuzzy Connections, Negations and Implications Based on T-norms and T-conorms," *Fuzzy Sets and Systems* **11** (1983), .

[857] K. Weise, "Distribution-Free Confidence Intervals," *PTB-Mitteilungen* **93** (1983), , Die Physikalisch-Technische Bundesanstalt.

[858] ———, ed., *Preliminary Translation of the German Draft Standard DIN 1319 Part 4 (7th Proposal, June 1984) 'Basic Concepts of Measurements: Treatment of Uncertainties in the Evaluation of Measurements'*, Normenausschuß Einheiten und Formalgrößen, Deutsches Institut für Normung (Committee for Units and Symbols of the German Institute of Standards), Berlin, 1984, Die Physikalisch-Technische Bundesanstalt.

[859] ———, "Treatment of Uncertainties in Precision Measurements," *IEEE Transactions on Instrumentation and Measurement* **36** (1987), .

[860] G. R. West, "Sensitivity and Probabilistic Analysis of Regional Input-Output Models," PhD thesis, University of Queensland, 1983.

[861] R. J. -B. Wets, "The Distribution Problem and its Relation to Other Problems in Stochastic Programming," in *Stochastic Programming*, M. A. H. Dempster, ed., Academic Press, London, 1980, .

[862] A. R. White, *Modal Thinking*, Cornell University Press, Ithaca, 1975.

[863] W. Whitt, "Bivariate Distributions with Given Marginals," *Annals of Statistics* **4** (1976), .

[864] B. Widrow, "Statistical Analysis of Amplitude Quantized Sampled-Data Systems," *AIEE Transactions part II* **79** (1960), , = Applications and Industry 52 (1961).

[865] H. Wilbraham, "On the Theory of Chances Developed in Professor Boole's 'Laws of Thought'," *The Philosophical Magazine (series 4) (supplement)* **7** (June 1954), Reprinted as pp.473–486 in R. Rhees, ed., *Studies in Logic and Probability* by George Boole, Watts and Company, London, 1952.

[866] J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Her Majesty's Stationery Office, London, 1963.

[867] ———, "Error Analysis Revisited," *Bulletin of the Institute of Mathematics and its Applications* **22** (1986), .

[868] P. W. Williams, "On a New Theory of Epistemic Probability," *British Journal for the Philosophy of Science* **29** (1978), .

[869] R. C. Williamson, "The Law of Large Numbers for Fuzzy Variables under a General Triangular Norm Extension Principle," *Fuzzy Sets and Systems* (1989), submitted.

[870] ———, "An Extreme Limit Theorem for Dependency Bounds of Normalised Sums of Random Variables," *Information Sciences* (1989), to appear.

[871] ———, "Interval Arithmetic and Probabilistic Arithmetic," in *Proceedings of the IMACS-GAMM-GI International Symposium on Computer Arithmetic and Self-Validating Numerical Methods*, 1989, To appear, October 1989.

[872] ———, "The Discrete T-Conjugate Transform," 1989, In preparation.

[873] R. C. Williamson & T. Downs, "The History of the Problem of the Quotient of Two Normal Random Variables," 1989, In preparation.

[874] ———, "Probabilistic Arithmetic: Numerical Methods for Calculating Convolutions and Dependency Bounds," *International Journal of Approximate Reasoning* (1989), to appear.

[875] B. P. Wise, "Experimentally Comparing Uncertain Inference Systems to Probability," in *Uncertainty in Artificial Intelligence 2*, J. F. Lemmer & L. N. Kanal, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1988, .

[876] B. P. Wise & M. Henrion, "A Framework for Comparing Uncertain Inference Systems to Probability," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[877] L. Wittgenstein, *Preliminary Studies for the 'Philosophical Investigations' ('The Blue Book')*, Basil Blackwell, Oxford, 1958, Reprinted as pp.294–357 of W. Barrett and H.D. Aiken (eds): *Philosophy in the Twentieth Century, Volume Two: The Rise of the British Tradition and Contemporary Analytic Philosophy*, Harper and Row, New York 1971.

[878] W. Wöger, "Confidence Limits Resulting from Two Models for the Randomization of Systematic Uncertainties," *PTB-Mitteilungen* **93** (1983), , Die Physikalisch-Technische Bundesanstalt.

[879] ———, "Randomization of Systematic Errors and its Consequences for the Evaluation of Measurements," in *Quantum Metrology and Fundamental Physical Constants*, P. H. Cutler & A. A. Lucas, eds., Plenum Press, New York, 1983, .

[880] M. Wolfenson & T. L. Fine, "Bayes-Like Decision Making with Upper and Lower Probabilities," *Journal of the American Statistical Association* **77** (1982), .

[881] E. F. Wolff, "Measures of Dependence Derived from Copulas," PhD Thesis, University of Massachusetts, 1977.

[882] ———, "N-Dimensional Measures of Dependence," *Stochastica* **4** (1980), .

[883] S. Wolfram, "Undecidability and Intractability in Theoretical Physics," *Physical Review Letters* **54** (1985), .

[884] M. -L. Wong, K. -S. Chen & C. -L. Chou, "Solutions of Integral Equations via Modified Laguerre Polynomials," *International Journal of Systems Science* **15** (1984), .

[885] S. Wright, "The Nature of Size Factors," *Genetics* **3** (1918), .

[886] ———, "Correlation and Causation," *Journal of Agricultural Research* **20** (1921), .

[887] ———, "The Method of Path Coefficients," *Annals of Mathematical Statistics* **5** (1934), .

[888] ———, "The Interpretation of Multivariate Systems," in *Statistics and Mathematics in Biology*, O. Kempthorne, T. A. Bancroft, J. W. Gowen & J. L. Lush, eds., Hafner Publishing Company, New York, 1954, .

[889] ———, "Path Coefficients and Path Regressions: Alternative or Complementary Concepts," *Biometrics* **16** (1960), .

[890] S. P. Yadav, "Approximation of Fourier-Laguerre Expansion by its Cesàro Mean in Certain Banach Spaces," *Journal of Approximation Theory* **39** (1983), .

[891] R. R. Yager, "On a General Class of Fuzzy Connectives," *Fuzzy Sets and Systems* **4** (1980), .

[892] ———, "An Introduction to Applications of Possibility Theory," *Human Systems Management* **3** (1983), .

[893] ———, "Arithmetic and Other Operations on Dempster-Shafer Structures," *International Journal of Man-Machine Studies* **25** (1986), .

[894] ———, "On the Dempster-Shafer Framework and New Combination Rules," *Information Sciences* **41** (1987), .

[895] A. V. Yazenin, "Fuzzy and Stochastic Programming," *Fuzzy Sets and Systems* **22** (1987), .

[896] R. C. Young, "The Algebra of Many-Valued Quantities," *Mathematische Annallen* **104** (1931), .

[897] L. A. Zadeh, "Fuzzy Sets," *Information and Control* **8** (1965), .

[898] ———, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning (Part I)," *Information Sciences* **8** (1975), , Part II, *Information Sciences* **8**, 301-357; Part III, *Information Sciences* **9**, 43-80.

[899] ———, "Fuzzy Sets versus Probability," *Proceedings of the IEEE* **68** (1980), 421.

[900] ———, "Possibility Theory as a Basis for Representation of Meaning," in *Language and Ontology: Proceedings of the Sixth International Wittgenstein Symposium*, W. Leinfellner, E. Kraemer & J. Schank, eds., Hölder-Pichler-Tempsky, Vienna, 1982, .

[901] ———, "Is Probability Theory Sufficient for Dealing with Uncertainty in AI: A Negative View," in *Uncertainty in Artificial Intelligence*, L. N. Kanal & J. F. Lemmer, eds., Elsevier Science Publishers B.V. (North-Holland), Amsterdam, 1986, .

[902] J. Žačková, "On Minimax Solutions of Stochastic Linear Programming problems," *Časopis pro pěstování matematiky, roč.* **91** (1966), .

[903] M. Zeleny, "On the (Ir)relevancy of Fuzzy sets Theories," *Human Systems Managemant* **4** (1984), .

[904] H. -J. Zimmermann, "Applications of Fuzzy Set Theory to Mathematical Programming," *Information Sciences* **36** (1985), .

[905] V. M. Zolotarev, "Mellin-Stieltjes Transforms in Probability Theory," *Theory of Probability and its Applications* **2** (1957), .

I quickly found it to exercise more than my devotion: it exercised my skill (all I had): it exercised my patience, it exercised my friends too, for 'tis incomparably the hardest taske that ever I yet undertooke.
— Williams Watts, Rector of St Alban's, Wood Street

Sources of Mottoes

Page 1, Mrs S. Boal (Grade 1, Oakleigh Primary School, 1968): Class Instruction (reconstructed from memory!).

Page 1, Berthold Schweizer: [715] (see also [718, p.123]).

Page 8, M.W. Crofton: "Probability," pp.768–788 in the *Encyclopedia Brittanica,* Volume 19, Adam and Charles Black, Edinburgh, 1885

Page 66, William Kruskal: [481, pp.933–934]

Page 131, Jorge Luis Borges: [197, p.221], cited in [266, p.257]

Page 131, Karl Popper: [660, p.5]

Page 132, Stanley Jevons : Quoted by Mary Everest Boole, *Logic Taught by Love.* See footnote 128 of [497].

Page 132, Glenn Shafer: [734, p.364].

Page 132, Theodore Hailperin: [355, p.215]

Page 145, S.L. Lauritzen and D.J. Spiegelhalter: [501, p.221]

Page 145, Glenn Shafer: Quoted by Pearl in [501, p.210]

Page 156, Peter Walley and Terrence Fine: [851, p.741]

Page 156, Henry Kyburg: [494, p.265]

Page 157, Karl Popper: [660, p.21]

Page 163, Masaharu Mizumoto and Kokichi Tanaka (1976): [579, pp.80–91]

Page 163, Didier Dubois and Henri Prade (1979): [225, p.347]

Page 163, Didier Dubois and Henri Prade (1980): [227, p.372]

Page 164, Didier Dubois and Henri Prade (1987): [239, p.33]

Page 178, Didier Dubois and Henri Prade: [240, p.3]

Page 178, Alan White: [862, p.59]

Page 188, Didier Dubois and Henri Prade: [236, p.298]

Page 188, Clark Glymour: [314, p.195]

Page 200, Nicolas Cusanus (1401–1464): Nicolas Cusanus, *De Docta Ignorantia (Of Learned Ignorance)*, Book I, Chapter 4. Translated by Germain Heron,

Routledge and Kegan Paul, London, 1954

Page 200, Karl Menger: [564, pp.473–474]

Page 204, Nicolas Cusanaus: Book 1, chapter 4 of *De Docta Ignorantia (Of learned ignorance),* translated by Germain Heron, Routledge and Kegan Paul, London, 1954.

Page 249, Ralph Waldo Emerson: page 63 of his TU journal (May 1849), reprinted on page 110 of volume XI of [629].

Page 254, Stansilaw Lem: Page 87 of "The World as Cataclysm," pp.69–102 in *One Human Minute*, Harcourt, Brace and Jovanovich, San Diego, 1986.

Page 256, "Folly" (Erasmus of Rotterdam): *The Praise of Folly, c.*1510. English translation in [216, p.139].

Page 302, Williams Watts, Rector of St Alban's, Wood Street: Preface to his translation of St Augustine's *Confessions,* London, 1631.