The relationship between Fourier and Mellin transforms, with applications to probability

Dave Collins

dcollins@unm.edu

Abstract

The use of Fourier transforms for deriving probability densities of sums and differences of random variables is well known. The use of Mellin transforms to derive densities for products and quotients of random variables is less well known. We present the relationship between the Fourier and Mellin transform, and discuss the use of these transforms in deriving densities for algebraic combinations of random variables. Results are illustrated with examples from reliability analysis.

1   Introduction

For the purposes of this paper, we may loosely define a *random variable* (RV) as a value in some domain, say $\mathbb{R}$, representing the outcome of a process based on a probability law. An example would be a real number representing the height in inches of a male chosen at random from a population in which height is distributed according to a Gaussian (normal) law with mean 71 and variance 25. Then we can say, for example, that the probability of the height of an individual from the population being between 66 and 76 inches is about .68.

For deriving such information about "nice" probability distributions (e.g., the height distribution above), we integrate the *probability density function* (pdf); in the case of the Gaussian the pdf is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]$, where μ is the mean and σ^2 is the variance.[1]

A question that frequently arises in applications is, given RVs X, Y with densities $f(x), g(y)$, what is the density of the random variable $X + Y$? (The answer is *not* $f(x) + g(y)$.) A less frequent but sometimes important question is, what is the density of the product XY? In this paper, after some brief background on probability theory, we provide specific examples of these questions and show how they can be answered with convolutions, using the Fourier and Mellin integral transforms. In fact (though we will not go into this level

---

[1] In this paper "nice" means RVs whose range is $\mathbb{R}^n$, with finite moments of all orders, and which are absolutely continuous with respect to Lebesgue measure, which implies that their pdfs are smooth almost everywhere and Riemann integrable. We will only deal with nice distributions.

of detail), using these transforms one can, in principle, compute densities for arbitrary rational functions of random variables [15].

1.1 Terminology

To avoid confusion, it is necessary to mention a few cases in which the terminology used in probability theory may be confusing:

- "Distribution" (or "law") in probability theory means a function that assigns a probability $0 \leqslant p \leqslant 1$ to every Borel subset of $\mathbb{R}$; *not* a "generalized function" as in the Schwartz theory of distributions.

- For historical reasons going back to Henri Poincaré, the term "characteristic function" in probability theory refers to an integral transform of a pdf, not to what mathematicians usually refer to as the characteristic function. For that concept, probability theory uses "indicator function", symbolized I; e.g., $I_{[0,1]}(x)$ is 1 for $x \in [0,1]$ and 0 elsewhere. In this paper we will not use the term "characteristic function" at all.

- We will be talking about pdfs being in $L^1(\mathbb{R})$, and this should be taken in the ordinary mathematical sense of a function on $\mathbb{R}$ which is absolutely integrable. More commonly, probabilists talk about *random variables* being in L^1, L^2, etc., which is quite different—in terms of a pdf f, it means that $\int |x| f(x) dx$, $\int |x|^2 f(x) dx$, etc. exist and are finite. It would require an excursion into measure theory to explain why this makes sense; suffice it to say that in the latter case we should really say something like "$L^1(\Omega, \mathcal{F}, P)$", which is not at all the same as $L^1(\mathbb{R})$.

2 Probability background

For those with no exposure to probability and statistics, we provide a brief intuitive overview of a few concepts. Feel free to skip to the end if you are already familiar with this material (but do look at the two examples at the end of the section).

Probability theory starts with the idea of the outcome of some process, which is mapped to a domain (e.g., $\mathbb{R}$) by a *random variable*, say X. We will ignore the underlying process and just think of $x \in \mathbb{R}$ as a "realization" of X, with a probability law or *distribution* which tells us how much probability is associated with any interval $[a, b] \subset \mathbb{R}$. "How much" is given by a number $0 \leqslant p \leqslant 1$.

Formally, probabilities are implicitly defined by their role in the axioms of probability theory; informally, one can think of them as degrees of belief (varying from 0, complete disbelief, to 1, complete belief), or as ratios of the number of times a certain outcome occurs to the total number of outcomes (e.g., the proportion of coin tosses that come up heads).

A probability law on $\mathbb{R}$ can be represented by its *density*, or pdf, which is a continuous function $f(x)$ with the property that the probability of finding x

in $[a, b]$ is $P(x \in [a, b]) = \int_a^b f(x)dx$. The pdf is just like a physical density—it gives the probability "mass" per unit length, which is integrated to measure the total mass in an interval. Note the defining characteristics of a probability measure on $\mathbb{R}$:

1. For any $[a, b]$, $0 \leqslant P(x \in [a, b]) \leqslant 1$.

2. $P[x \in (-\infty, \infty)] = 1$.

3. if $[a, b] \cap [c, d] = \varnothing$, then $P(x \in [a, b] \cup [c, d]) = P(x \in [a, b]) + P(x \in [c, d])$.

From these properties and general properties of the integral it follows that if f is a continuous pdf, then $f(x) \geqslant 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$.

Though we don't need them here, there are also *discrete* random variables, which take values in a countable set as opposed to a continuous domain. For example, a random variable representing the outcome of a process that counts the number of students in the classroom at any given moment takes values only in the nonnegative integers. There is much more to probability, and in particular a great deal of measure-theoretic apparatus has been ignored here, but it is not necessary for understanding the remainder of the paper.

The Gaussian or *normal* density was mentioned in section 1. We say that $X \sim N(\mu, \sigma^2)$ if it is distributed according to a normal law with *mean* or average μ and *variance* σ^2. The mean μ determines the center of the normal pdf, which is symmetric; μ is also the *median* (the point such that half the probability mass is above it, half below), and the *mode* (the unique local maximum of the pdf). If the pdf represented a physical mass distribution over a long rod, the mean μ is the point at which it would balance. The variance is a measure of the variability or "spread" of the distribution. The square root of the variance, σ, is called the *standard deviation*, and is often used because it has the same unit of measure as X.

Formally, given any RV X with pdf f, its mean is $\mu = \int_{-\infty}^{\infty} xf(x)dx$ (the average of x over the support of the distribution, weighted by the probability density). This is usually designated by $E(X)$, the expectation or *expected value* of X. The variance of X is $E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$ (the weighted average of the squared deviation of x from its mean value).

Figure 1 plots the $N(71, 25)$ density for heights mentioned in Section 1. The central vertical line marks the mean, and the two outer lines are at a distance of one standard deviation from the mean. The definite integral of the normal pdf can't be solved in closed form; an approximation is often found as follows: It is easy to show that if $X \sim N(\mu, \sigma^2)$, then $\frac{X - \mu}{\sigma} \sim N(0, 1)$; also from the properties of a probability measure, for any random variable X,

$$P(a \leqslant X \leqslant b) = P(-\infty < X \leqslant b) - P(-\infty < X \leqslant a).$$

It therefore suffices to have a table of values for $P(-\infty < X \leqslant b)$ for the $N(0, 1)$ distribution. (Viewed as a function of b, $P(b)$ is called the *cumulative distribution function*.) Such tables are found in all elementary statistics books, and give, e.g., $P(66 \leqslant X \leqslant 76) \approx .682$.
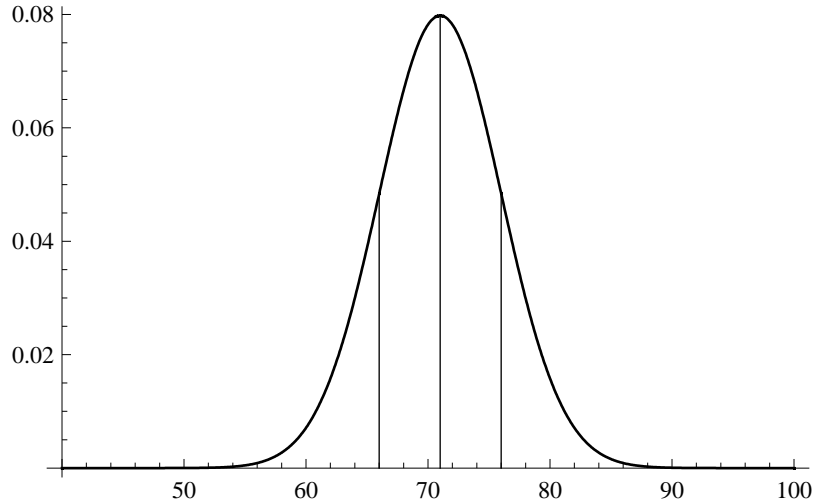
3

Figure 1: $N(71, 25)$ pdf for the distribution of heights

Many applications use random variables that take values only on $[0, \infty)$, for example to represent incomes, life expectancies, etc. A frequently used model for such RVs is the *gamma distribution* with pdf

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \text{ if } x > 0, \ 0 \text{ otherwise.}$$

(Notice that aside from the constant $\frac{1}{\Gamma(\alpha)\beta^\alpha}$, which normalizes f so it integrates to 1, and the extra parameter β, this is the kernel of the gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$, which accounts for the name.) Figure 2 shows a gamma(4, 2) pdf ($\alpha = 4$, $\beta = 2$). Because a gamma density is never symmetric, but skewed to the right, the mode, median and mean occur in that order and are not identical. For an incomes distribution this means that the typical (most likely) income is smaller than the "middle" income which is smaller than the average income (the latter is pulled up by the small number of people who have very large incomes).

Independence is profoundly important in probability theory, and is mainly what saves probability from being "merely" an application of measure theory. For the purposes of this paper, an intuitive definition suffices: two random variables X, Y are independent if the occurrence or nonoccurrence of an event $X \in [a, \ b]$ does not affect the probability of an event $Y \in [c, \ d]$, and vice versa. Computationally, the implication is that "independence means multiply". E.g., if X, Y are independent,

$$P(X \in [a, \ b] \ \& \ Y \in [c, \ d]) = P(X \in [a, \ b])P(Y \in [c, \ d]).$$

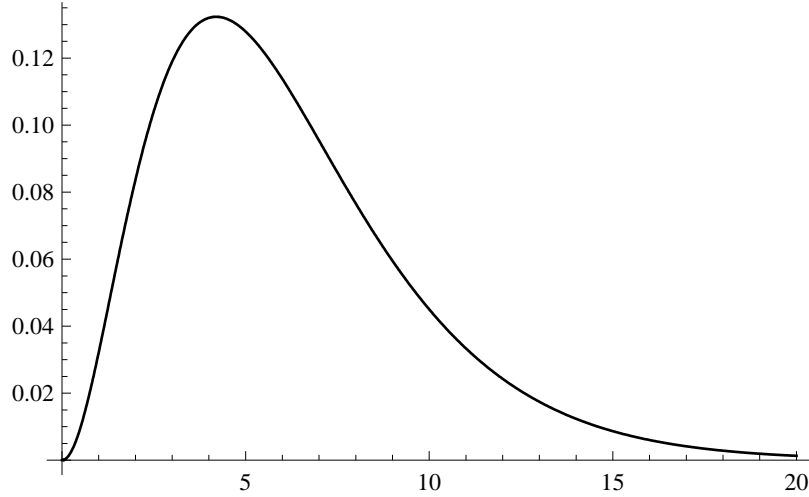In this paper, we will only consider independent random variables.

4

Figure 2: gamma(4, 2) pdf

Extending the example with which we began, suppose we consider heights of *pairs* of people from a given population, where each member of the pair is chosen at random, so we can assume that their heights are independent RVs X, Y. Then for any given pair $(x, \ y)$ we can ask, for example, about the probability that both $66 \leqslant x \leqslant 76$ and $66 \leqslant y \leqslant 76$. This requires a *joint* or *bivariate* density $f(x, y)$ of X and Y. Using the "independence means multiply" rule above and taking limits as the interval sizes go to 0, it should be fairly obvious that $f(x, y) = f_X(x)f_Y(y)$, where f_X and f_Y are the densities of X and Y. It follows that

$$P(X \in [66, \ 71] \ \& \ Y \in [66, \ 71]) = \int_{66}^{71} f_X(x)dx \int_{66}^{71} f_Y(y)dy = \int_{66}^{71} \int_{66}^{71} f(x, y)dxdy.$$

By substituting $[-\infty, \ \infty]$ for either of the intervals of integration, it is also readily seen that

$$\int_a^b f_X(x)dx \int_{-\infty}^{\infty} f_Y(y)dy = \int_a^b \int_{-\infty}^{\infty} f(x, y)dxdy = \int_a^b f_X(x)dx.$$

And it follows that by "integrating out" one of the variables from the joint density $f(x, y)$, we recover the *marginal density* of the other variable:

$$\int_{-\infty}^{\infty} f(x, y)dx = f_Y(y).$$

This is true whether or not X and Y are independent.

Figure 3 illustrates this. The 3-D plot is a bivariate standard normal density, the product of two $N(0, 1)$ densities. On the right is the marginal density of Y,

5

$f_Y(y)$,which results from aggregating the density from all points x corresponding to a given y—i.e., integrating the joint density along the line $Y = y$ parallel to the x-axis. (The marginal density $f_Y(y)$ is $N(0, 1)$, as expected.) Later, in discussing convolution, we will see that it is also useful to integrate the joint density along a line that is not parallel to one of the coordinate axes.
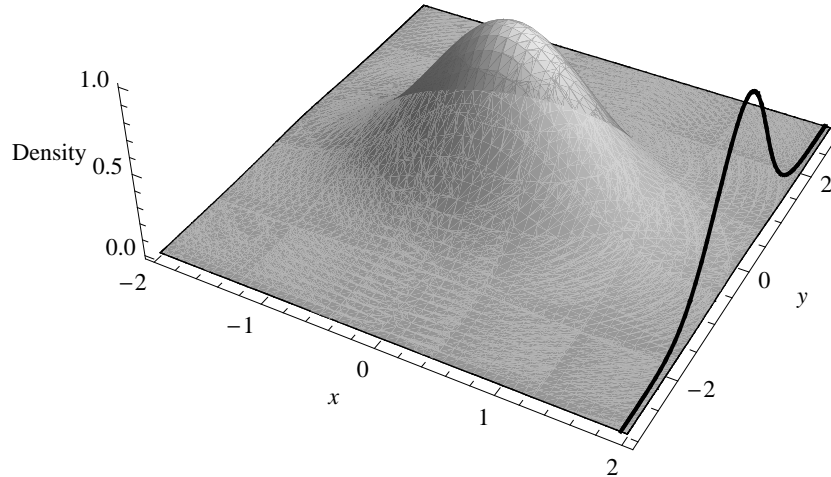

Figure 3: Bivariate normal pdf $f(x, y)$, with marginal density $f_Y(y)$

2.1   Examples

With this background, here are two examples illustrating the need to compute densities for sums and products of RVs.

Example 1 (Sum of random variables): Suppose you carry a backup battery for a cellphone. Both the backup and the battery in the phone, when fully charged, have a lifetime that is distributed according to a gamma law $f(x) \sim \text{gamma}(\alpha, \beta)$, $\alpha = 25$, $\beta = .2$, where x is in hours; $\alpha\beta = 5$ hours is the mean (average) life and $\alpha\beta^2 = 1$ is the variance. This density is shown in Figure 4; it looks similar to a bell-shaped Gaussian, but it takes on only positive values.[2] What is the probability that both batteries run down in less than 10 hours? To answer questions like this we need the distribution of the sum of the random variables representing the lifetimes of the two batteries. E.g., if the lifetimes of

---

[2]Another difference: The Gaussian, as we know, is in the Schwartz class; the gamma pdf is not, since it is not C^∞ at the origin.

the two batteries are represented by $X \sim \text{gamma}(25, .2)$, $Y \sim \text{gamma}(25, .2)$, integrating the density of $X + Y$ from 0 to 10 will give the probablity that both batteries die in 10 hours or less.
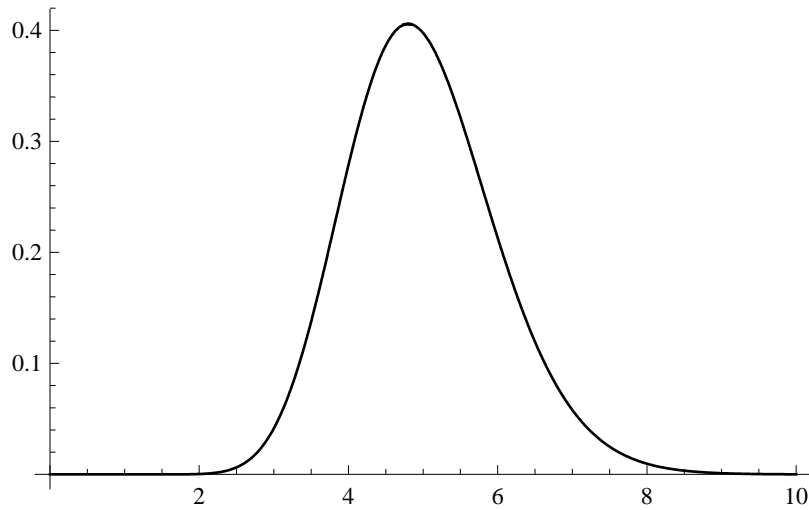

Figure 4: gamma(25, .2) pdf for Battery life

Example 2 (Product of random variables): The *drive ratio* of a pair of pulleys connected by a drive belt is (roughly) the ratio of the pulley diameters, so, e.g., if the drive ratio is 2, the speed is doubled and the torque is halved from the first pulley to the second. In practice, the drive ratio is not exact, but is a random variable which varies slightly due to errors in determining the pulley dimensions, slippage of the belt, etc.

Figure 5 shows an example: suppose a motor is turning the left-hand driveshaft at 800 rpm, which is connected to another shaft using a sequence of two belt and pulley arrangements. The nominal drive ratios are 2 and 1.5; thus the shaft connected to the pulley on the right is expected to turn at 2400 rpm ($2 \times 1.5 \times 800$).

Suppose the motor speed is taken to be constant, and we are told in the manufacturer's specifications that the first drive ratio is $2 \pm .05$ and the second is $1.5 \pm .05$. Given only this information, we might model the drive ratios as *uniform* random variables, which distribute probability evenly over a finite interval; if the interval is $[a, b]$, the uniform(a, b) pdf is $f(x) = \frac{1}{b-a} I_{[a,b]}(x)$. So the two drive ratios in this case are given by RVs $X \sim \text{uniform}(1.95, 2.05)$ and $Y \sim \text{uniform}(1.45, 1.55)$. If the reliability of the system requires that the speed of the driven shaft be within a certain tolerance, then we need to know the probability distribution describing the actual speed of the driven shaft. This will be answered by computing the probability density for the product XY.
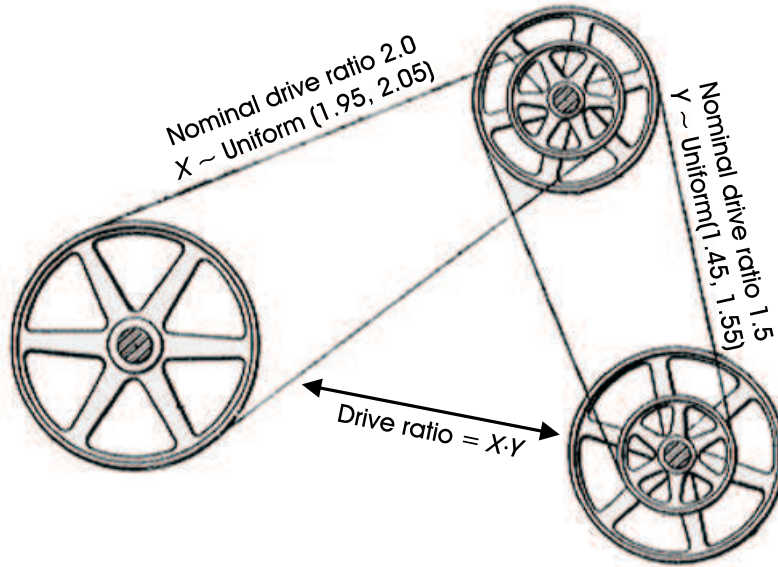
7

Figure 5: Product of random variables: Belt and pulley drive

3 Transforms for sums of random variables

Suppose that the RV X has pdf $f_X(x)$ and Y has pdf $f_Y(y)$, and X and Y are independent. What is the pdf $f_Z(z)$ of the sum $Z = X + Y$? Consider the transformation $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\psi(x, y) = (x, x + y) \equiv (x, z)$. If we can determine the joint density $f_{XZ}(x, z)$, then the marginal density $f_Z(z) = \int_{\mathbb{R}} f_{XZ}(x, z)dx$. The transformation ψ is injective with $\psi^{-1}(x, z) = (x, z - x)$ and has Jacobian identically equal to 1, so we can use the multivariate change of variable theorem to conclude that

$$
\begin{aligned}
f_Z(z) &= \int_{\mathbb{R}} f_{XZ}(x, z)dx \\
&= \int_{\mathbb{R}} f_{XY}(\psi^{-1}(x, z))dx \\
&= \int_{\mathbb{R}} f_{XY}(x, z - x)dx \\
&= \int_{\mathbb{R}} f_X(x)f_Y(z - x)dx \quad \text{by the independence of } X \text{ and } Y \\
&= f_X \star f_Y(z).
\end{aligned}
$$

The next-to-last line above is intuitive: it says that we find the density for $Z = X + Y$ by integrating the joint density of X, Y over all points where

8

$X + Y = Z$, i.e., where $Y = Z - X$. Figure 6 illustrates this for $z = 1$: $f \star g(1) = \int_{\mathbb{R}} f(1-z)g(z)dz$ is the integral of the joint density $f_{XY}(x,y) = f_X(x)f_Y(y)$ over the line $y = 1 - x$.
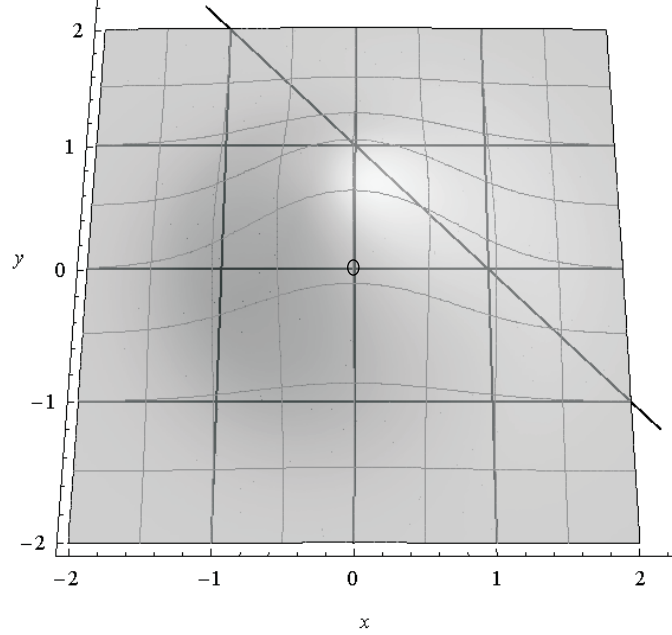

Figure 6: Integration path for $f \star g(1) = \int_{\mathbb{R}} f(1-z)g(z)dz$ along the line $y = 1-x$

In general, computation of the convolution integral is difficult, and may be intractable. It is often simplified by using transforms, e.g., the Fourier transform:

$$\widehat{f_X \star f_Y}(\xi) = \widehat{f_X}(\xi)\widehat{f_Y}(\xi)$$

The transform is then inverted to get $f_Z(z)$.

As an example, consider the gamma(α, β) pdf, whose Fourier transform is given by:

$$
\begin{aligned}
\int_{\mathbb{R}} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} e^{-2\pi i \xi x} dx &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_{\mathbb{R}} x^{\alpha-1} e^{-x\frac{1+2\pi i \xi \beta}{\beta}} dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \Gamma(\alpha) \left(\frac{\beta}{1 + 2\pi i \xi \beta} \right)^\alpha
\end{aligned}
$$

9

$$= \frac{1}{(1 + 2\pi i \xi \beta)^\alpha}$$

There is a trick here in passing from the first to the second line. Recall that the kernel of the gamma(α, β) pdf integrates to $\Gamma(\alpha)\beta^\alpha$; Then notice that the integrand is the kernel of a gamma$(\alpha, \frac{\beta}{1+2\pi i \xi \beta})$ pdf, which therefore integrates to $\Gamma(\alpha)\left(\frac{\beta}{1+2\pi i \xi \beta}\right)^\alpha$.

Thus the Fourier transform of the convolution of two independent gamma(α, β) RVs is

$$\widehat{f_X \star f_Y}(\xi) = \frac{1}{(1 + 2\pi i \xi \beta)^{2\alpha}}$$

which by inspection is the Fourier transform of a gamma$(2\alpha, \beta)$ random variable.

This answers the question posed in Example 1: If X, $Y \sim$ gamma$(25, .2)$, then $X + Y \sim$ gamma$(50, .2)$. By integrating this numerically (using Mathematica) the probability that both batteries die in 10 hours or less is found to be about .519.

In practice, we don't require the Fourier transform; we can use any integral transform T with the convolution property $T_{f\star g} = T_f T_g$. In particular, since densities representing lifetimes are supported on $[0, \infty)$, the Laplace transform $\mathcal{L}_f(s) = \int_0^\infty \exp(-st) f(t) dt$, $s \in \mathbb{R}$, is often used in reliability analysis.

Note that the convolution result is extensible; it can be shown by induction that the Fourier transform of the pdf of a sum of n independent RVs $X_1 + \cdots + X_n$ with pdfs $f_1, \ldots, f_n$ is given by[3]

$$(f_1 \star \cdots \star f_n)\widehat{}(\xi) = \hat{f}_1(\xi) \cdots \widehat{f_n}(\xi)$$

4   Transforms for products of random variables

We now motivate a convolution for products, derive the Mellin transform from the Fourier transform, and show its use to compute products of random variables. This requires a digression into algebras on spaces of functions.

4.1   Convolution algebra on $L^1(\mathbb{R})$

The general notion of an *algebra* is a collection of entities closed under operations that "look like" addition and multiplication of numbers. In the context of function spaces (in particular $L^1(\mathbb{R})$, which is where probability density functions live) functions are the entities, addition and multiplication by scalars have the obvious definitions, and we add an operation that multiplies functions.

For linear function spaces that are complete with respect to a norm (Banach spaces[4]) the most important flavor of algebra is a *Banach algebra* [2, 12], with

---

[3]An application of this result is the famous *central limit theorem*, which says that under very general conditions, the average of n independent and identically distributed random variables with any distribution whatsoever converges to a Gaussian distributed random variable as $n \to \infty$. See [8], p. 114 ff., for a proof.

[4]If the norm is given by an inner product, the Banach space is a Hilbert space.

the following properties ($\circ$ is the multiplication operator, which is undefined for the moment, λ is a scalar, and $\|$ is the norm on the space):

i) $f \circ (g \circ h) = (f \circ g) \circ h$

ii) $f \circ (g + h) = (f \circ g) + (f \circ h)$

iii) $(f + g) \circ h = (f \circ g) + (f \circ h)$

iv) $\lambda(f \circ g) = (\lambda f) \circ g = f \circ (\lambda g)$

v) $\| f \circ g \| \leqslant \| f \| \| g \|$

We can't use the obvious definition of multiplication to define an algebra over $L^1(\mathbb{R})$, because $f, g \in L^1(\mathbb{R})$ does not imply $fg \in L^1(\mathbb{R})$. For example, one can verify that $f(x) = \frac{1}{2\sqrt{\pi |x|}} e^{-|x|}$ is in $L^1(\mathbb{R})$ (in fact, it is a pdf), but $\int_{\mathbb{R}} |f(x)|^2 dx = \infty$.

Since L^1 is not closed under ordinary multiplication of functions, we need a different multiplication operation, and convolution is the most useful possibility. To verify closure, if $f, g \in L^1(\mathbb{R})$,

$$
\begin{aligned}
\| f \star g \| &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} f(y - x) g(x) dx \right| dy \\
&\leqslant \int_{\mathbb{R}} \int_{\mathbb{R}} |f(y - x)| |g(x)| dx dy \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} |f(y - x)| dy \right] |g(x)| dx \quad \text{by Fubini's theorem} \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} |f(z)| dz \right] |g(x)| dx \quad \text{by the substitution } z = y - x \\
&= \int_{\mathbb{R}} \| f \| |g(x)| dx \\
&= \| f \| \| g \| .
\end{aligned}
$$

This also verifies property (v), the norm condition, and is sometimes called *Young's inequality*.[5] The remainder of the properties are easily verified, as well as the fact that the convolution algebra is commutative: $f \star g = g \star f$.

4.2   A product convolution algebra

Consider the operator $T : f(x) \mapsto f(e^x)$ for $f \in L^1(\mathbb{R})$. Define a norm for T-transformed functions by

$$
\| f \|_T = \int_0^\infty |f(e^x)| dx = \int_0^\infty |f(y)| \frac{1}{y} dy
$$

---

[5]This is one of two different results that are called Young's inequality—see http://en.wikipedia.org/wiki/Young's_inequality.

where the last expression follows from the substitution $y = e^x$. Note that $f \in L^1(\mathbb{R})$ does not imply finiteness of the T-norm; for example, the pdf e^{-x} is in $L^1(\mathbb{R})$, but $\int_0^\infty |e^{-y}| \frac{1}{y} dx$ does not converge. This is also true for many other pdfs, including the Gaussian.

In order to salvage the T-norm for a function space that includes pdfs, we use a modified version, the $\mathcal{M}_c$-norm defined by

$$\| f \|_{\mathcal{M}_c} = \int_0^\infty |f(x)| x^{c-1} dx$$

where c is chosen to insure convergence for the class of functions we are interested in. All pdfs $f(x)$ satisfy $\int_0^\infty |f(x)| dx < \infty$, and nice ones decay rapidly at infinity so $\int_0^\infty x^p |f(x)| dx < \infty$ for $p \geqslant 1$; therefore $\| f \|_{\mathcal{M}_c} < \infty$ if $c \geqslant 1$ for f in the class of "nice" pdfs.

We can define a convolution for T-transformed functions by transforming the functions in the standard convolution $f \star g(x) = \int_{-\infty}^\infty f(x-u)g(u) du$:

$$
\begin{aligned}
f \diamond g(z) \quad &:= \quad (Tf) \star (Tg)(e^x) \quad \text{where } z = e^x \\
&= \quad \int_0^\infty f(e^{x-u}) g(e^u) du \\
&= \quad \int_0^\infty f(e^{\log z - \log w}) g(e^{\log w}) \frac{1}{w} dw \\
&= \quad \int_0^\infty f\left(\frac{z}{w}\right) g(w) \frac{1}{w} dw.
\end{aligned}
$$

(The next-to-last line follows from the substitutions $z = e^x, w = e^u$). This is called the *Mellin convolution*. It is, like the Fourier convolution, commutative: $f \diamond g = g \diamond f$.

Now for fixed $c \in \mathbb{R}$ let $\mathcal{M}_c(\mathbb{R}_+)$ be the space of functions on $(0, \infty)$ with finite $\mathcal{M}_c$-norm. Using the obvious definitions of addition and multiplication by scalars and the $\diamond$ convolution for multiplication of functions, it can be shown that $\{\mathcal{M}_c(\mathbb{R}_+), +, \diamond\}$ is a Banach algebra. Verifying closure under addition and scalar multiplication, and properties (i)–(iv), involves simple computations. The proof of property (v) and closure under $\diamond$ is lengthy, and we also need to prove that $\mathcal{M}_c(\mathbb{R}_+)$ is a Banach space relative to the $\mathcal{M}_c$-norm, i.e., that any Cauchy sequence of functions with finite $\mathcal{M}_c$-norms converges to a function with finite $\mathcal{M}_c$-norm. We omit these here; for detailed proofs, see [5].

4.3 The Mellin transform, and its relation to the Fourier transform

If $f \in \mathcal{M}_c(\mathbb{R}_+)$ for all $c \in [a, b]$, we will say that $f \in \mathcal{M}_{[a,b]}(\mathbb{R}_+)$ (our "nice" pdfs are in $\mathcal{M}_{[1,\infty)}(\mathbb{R}_+)$). Then we define the *Mellin transform* of f with argument

$s \in \mathbb{C}$ as

$$F(s) = \mathcal{M}[f](s) = \int_0^\infty f(u)u^{s-1}du$$

where $a \leqslant Re(s) \leqslant b$. (It is easy to show that if the integral converges for $s = c \in \mathbb{R}$, it converges for $s = c + it, t \in \mathbb{R}$). The subscript on $\mathcal{M}$ is usually omitted, with the assumption that the integral converges for the given s.

For $F(s) = \mathcal{M}[f](s)$, the inverse Mellin transform is

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

The condition that the inverse exists is that $F(s)x^{-s}$ is analytic in a strip $(a, b) \times (-i\infty, i\infty)$ such that $c \in (a, b)$ [5].

The Mellin transform can be derived from the Fourier transform

$$\hat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-2\pi i\xi x}dx$$

using the transformation T and the substitution $\xi = -\frac{\eta-c}{2\pi i}$ for real $c \geqslant 0$:

$$
\begin{aligned}
\widehat{Tf}(\xi) &= \int_{-\infty}^{\infty} f(e^x)e^{-2\pi i\xi x}dx \\
\widehat{Tf}\left(-\frac{\eta-c}{2\pi i}\right) &= \int_{-\infty}^{\infty} f(e^x)e^{(\eta-c)x}dx \qquad \text{for } c \geqslant 0 \\
&= \int_0^\infty f(y)e^{-(\eta-c)\log y}\frac{1}{y}dy \quad \text{with the substitution } y = e^x \\
&= \int_0^\infty f(y)y^{-c}y^\eta\frac{1}{y}dy \\
&= \int_0^\infty f^*(y)y^{\eta-1}dy \qquad\qquad \text{for } f^*(y) = f(y)y^{-c}.
\end{aligned}
$$

(An aside on the substitution $\xi = -\frac{\eta-c}{2\pi i}$: The factor of 2π is a consequence of the way we define the Fourier transform. In statistics, and in many engineering texts, the Fourier transform is defined as $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{i\xi x}dx$ (essentially measuring frequency in radians per time unit instead of cycles per time unit), which simplifies the derivation of the Mellin transform from the Fourier transform. For a summary of the different ways the Fourier transform and its inverse are represented, see [13], Appendix D.)

The same technique is used to derive the Mellin inversion formula from the Fourier inversion:

$$
\begin{aligned}
f(y) &= T^{-1}[\widetilde{\hat{f}(\cdot)}](y) \\
&= \int_{-\infty}^{\infty} \hat{f}(\xi)e^{2\pi i \log(y)\xi}d\xi
\end{aligned}
$$

13

$$
\begin{aligned}
&= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \hat{f}\left(-\frac{\eta-c}{2\pi i}\right) e^{-(\eta-c)\log(y)} d\eta \quad \text{with the substitution } \xi = -\frac{\eta-c}{2\pi i} \\
&= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f\left(-\frac{\eta}{2\pi i}\right) y^c y^{-\eta} d\eta \\
f(y)y^{-c} &= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f\left(-\frac{\eta}{2\pi i}\right) y^{-\eta} d\eta \\
&= f^*(y).
\end{aligned}
$$

In some cases the transformation T provides an easier way to invert Mellin transforms, through the use of Fourier inversion techniques.

For computing the pdf of a product of random variables, the key result will be that the Mellin transform of a Mellin convolution is the product of the Mellin transforms of the convolved functions:

$$
\begin{aligned}
\mathcal{M}[f \diamond g](s) &= \int_0^\infty \left[\int_0^\infty f(\frac{z}{w})g(w)\frac{1}{w}dw \right] z^{s-1} dz \\
&= \int_0^\infty \left[\int_0^\infty f(\frac{z}{w})z^{s-1}dz \right] g(w)\frac{1}{w}dw \qquad \text{by Fubini's theorem} \\
&= \int_0^\infty \left[\int_0^\infty f(y)y^{s-1}w^{s-1}w\, dy \right] g(w)\frac{1}{w}dw \quad \text{substituting } y = \frac{z}{w} \\
&= \int_0^\infty \left[\int_0^\infty f(y)y^{s-1}dy \right] g(w)w^{s-1}dw \\
&= \mathcal{M}[f](s)\mathcal{M}[g](s)
\end{aligned}
$$

As with the Fourier convolution, this result is extensible; it can be shown by induction that the Mellin transform of the Mellin convolution of of $f_1, \ldots, f_n$ is given by

$$
\mathcal{M}[f_1 \diamond \cdots \diamond f_n](s) = \mathcal{M}[f_1](s) \cdots \mathcal{M}[f_n](s) \tag{1}
$$

4.4   Products of random variables

Suppose we have random variables X, Y with pdfs f_X, f_Y, and the product $Z = XY$ is to be determined. Consider the transformation $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\psi(x,y) = (x, xy) \equiv (x, z)$. Except at $x = 0$,[6] ψ is injective with $(x,y) = \psi^{-1}(x,z) = (x, z/x)$ and the Jacobian of ψ^{-1} is

$$
J = \begin{vmatrix} \dfrac{\partial \psi_1^{-1}}{\partial x} & \dfrac{\partial \psi_2^{-1}}{\partial x} \\[2ex] \dfrac{\partial \psi_1^{-1}}{\partial z} & \dfrac{\partial \psi_2^{-1}}{\partial z} \end{vmatrix} = \begin{vmatrix} 1 & -\dfrac{1}{x^2} \\[2ex] 0 & \dfrac{1}{x} \end{vmatrix} = \frac{1}{x}.
$$

---

[6]This can be handled gracefully using Lebesgue integration theory, but here we ignore the problem.

Then using the multivariate change of variable theorem, the marginal density of Z is computed from the joint density of X and Z as

$$
\begin{aligned}
f_Z(z) &= \int_{\mathbb{R}} f_{XZ}(x, z)dx \\
&= \int_{\mathbb{R}} f_{XY}(\psi^{-1}(x, z))\frac{1}{x}dx \\
&= \int_{\mathbb{R}} f_{XY}\left(x, \frac{z}{x}\right)\frac{1}{x}dx \\
&= \int_{\mathbb{R}} f_X(x)f_Y\left(\frac{z}{x}\right)\frac{1}{x}dx \quad \text{by the independence of } X \text{ and } Y \\
&= f_X \diamond f_Y(z).
\end{aligned}
$$

This is precisely the Mellin convolution of f_X and f_Y. In principle, this plus the extensibility result (1) provides a way of finding product densities for arbitrary numbers of random variables.

Note that the Mellin transform is defined only for functions supported on the positive half-line $\mathbb{R}_+$, whereas many pdfs (e.g., the Gaussian) do not satisfy this requirement. For such cases, the problem can be worked around by separating the positive and negative parts of the pdf; see [15] for details.

4.5   An example

As a simple illustration of the use of the Mellin transform, we use the belt and pulley example (Example 2, p. 7). Recall that $X \sim \text{uniform}(1.95, 2.05)$, $Y \sim \text{uniform}(1.45, 1.55)$ and we seek the pdf of the product XY.

The problem can be simplified by using the fact that a $\text{uniform}(\alpha, \beta)$ random variable can be expressed as $\alpha + (\alpha - \beta)U$, where U is a $\text{uniform}(0, 1)$ random variable with pdf $I_{[0,1]}(x)$. In this case, $X = 1.95 + .1U$, $Y = 1.45 + .1U$. Then $XY = 2.8275 + .34U + .01U^2$. Since we already know how to compute sums, the problem reduces to finding the pdf for the product of two $\text{uniform}(0, 1)$ random variables.

For $Z = U^2$, the Mellin convolution evaluates to

$$
\begin{aligned}
f_Z(z) &= \int_{\mathbb{R}} f_X(x)f_Y\left(\frac{z}{x}\right)\frac{1}{x}dx \\
&= \int_z^1 \frac{1}{x}dx \\
&= \log(x)|_z^1 \\
&= -\log(z), \quad 0 < z \leqslant 1.
\end{aligned}
$$

The bounds for the integration come from $x \leqslant 1$ and $y \leqslant 1 \Rightarrow x \geqslant z$.

15

This result can also be obtained as $\mathcal{M}^{-1}\{\mathcal{M}[f_U](s)^2\}(x)$, where f_U is the pdf of U. We have

$$\mathcal{M}[f_U](s) = \int_0^1 x^{s-1}dx = \frac{1}{s}$$

so we need

$$\mathcal{M}^{-1}\left\{\frac{1}{s^2}\right\}(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{z^{-s}}{s^2}ds$$

which evaluates to the same result after an exercise in using the residue theorem of complex analysis.

In this simple case of the product of two uniform(0,1) RVs it is easier to compute the Mellin convolution directly; but the use of Mellin transforms allows computation of the pdf for a product of n uniform(0,1) RVs almost as easily, yielding $\dfrac{-\log(z)^{n-1}}{(n-1)!}$ (another exercise in residue calculus).

The difficulty of either directly integrating the Mellin convolution or inverting a product of Mellin transforms escalates quickly for less simple distributions such as the gamma or normal. In particular, whereas the transforms of Fourier convolutions of pdfs can often be evaluated by inspection (possibly using tables), this is not the case for Mellin transforms, though extensive tables do exist [3]. This seems to be a consequence of the fact that sums of RVs often have pdfs with mathematical forms similar to the individual RVs (e.g., a sum of normal RVs is normal), unlike products of RVs (e.g., the uniform example above).

The reader is referred to [15] for realistic examples, which are too lengthy to reproduce here.

5   Summary

We have presented some background on probability theory, and two examples motivating the need to compute probability density functions for sums and products of random variables. The use of the Fourier or Laplace transform to evaluate the convolution integral for the pdf of a sum is relatively straightforward. The use of the Mellin transform to evaluate the convolution integral for the pdf of a product is less well-known, but equally straightforward, at least in theory.

In practice, though the use of Fourier or Laplace transforms for sums of random variables is widely used and explained in every advanced statistics text, the Mellin transform remains obscure. Aside from Epstein's seminal paper of 1948 [9], there was a brief flurry of activity in the 1960s and 70s by Springer and Thompson (e.g., [16]) culminating in Springer's book [15]. Current texts in probability and statistics, however, do not mention the Mellin transform, and its appearance in current literature is rare.

To some extent the relative lack of interest in products of random variables is due to the lesser importance of products in applications. It probably also is

a consequence of the greater difficulty of working with the integrals involved—particularly the fact that inverting the Mellin transform requires a strong knowledge of complex variable methods, which are not part of the standard graduate curriculum in statistics. Nevertheless, it seems worthwhile for any statistician to develop at least a nodding acquaintance with Mellin transform methods. Mathematicians and engineers will also find interesting applications (see the further reading below).

5.1 Further reading

[11] is a nice summary of all the transform techniques used in probability theory. [15] is the ultimate reference on transform techniques for algebraic combinations of random variables.

For the use of integral transforms to compute sums of random variables, see any graduate textbook on probability and statistics, e.g., [4, 6].

[1], [19], [11], and [15] all cover the Mellin transform, the last two in the probability context. [3] contains an extensive table of Mellin transforms (as well as Fourier, Laplace, and other transforms). [5] contains a very complete treatment of properties of the Mellin transform, with proofs.

[1] and [19] provide considerable depth on integral transforms generally, oriented towards applied mathematics. A more abstract view is provided by [20], which includes a treatment of integral transforms of (Schwartz) distributions.

The algebraic properties of Fourier and Mellin transforms are (briefly) worked out in a series of exercises in [8] (ex. 9–15, pp. 41–43; ex. 2, p. 88; ex. 3, p. 103). For the more algebraically inclined, one can develop an abstract theory of convolution and Fourier analysis on groups. See [7], "Appendix: functions on groups" for an elementary introduction, or [14] for a full treatment.

Probability and statistics is only one application area for the Mellin transform, and it is not the most important. The Mellin transform is used in computer science for analysis of algorithms (see, for example, [17, ch. 9-10]); it has applications to analytic number theory [10]; and Mellin himself developed it in connection with his researches in the theory of functions, number theory, and partial differential equations [18].

References

[1] Andrews, L. C., and B. K. Shivamoggi. *Integral Transforms for Engineers and Applied Mathematicians*. New York: Macmillan Publishing, 1988.

[2] Arveson, William. *A Short Course on Spectral Theory*. New York: Springer, 2002.

[3] Batemen, Harry. *Tables of Integral Transforms*, Vol. I. New York: McGraw-Hill, 1954.

[4] Bauer, Heinz. *Probability Theory*. Berlin: Walter de Gruyter, 1996.

[5] Butzer, Paul L., and Stefan Jansche. "A direct approach to the Mellin transform". *The Journal of Fourier Analysis and Applications* **3** (1997), pp. 325–376.

[6] Cramér, Harald. *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press, 1946.

[7] Davis, Harry F. *Fourier Series and Orthogonal Functions.* New York: Dover Publications, 1989.

[8] Dym, H., and H. P. McKean. *Fourier Series and Integrals.* New York: Academic Press, 1972.

[9] Epstein, Benjamin. "Some applications of the Mellin transform in statistics". *The Annals of Mathematical Statistics* **19** (1948), pp. 370–379.

[10] Flajolet, Philippe, *et al.* "Mellin transforms and asymptotics: Harmonic sums". *Theoretical Computer Science* **144** (1995), pp. 2-58.

[11] Giffin, Walter C. *Transform Techniques for Probability Modeling.* New York: Academic Press, 1975.

[12] Goffman, Casper, and George Pedrick. *First Course in Functional Analysis.* Englewood Cliffs, NJ: Prentice-Hall, 1965.

[13] Hubbard, Barbara Burke. *The World According to Wavelets*, 2nd Edition. Natick, MA: A K Peters, 1998.

[14] Rudin, Walter. *Fourier Analysis on Groups.* New York: Wiley-Interscience, 1962.

[15] Springer, M. D. *The Algebra of Random Variables.* New York: John Wiley & Sons, 1979.

[16] Springer, M. D., and W. E. Thompson. "The distribution of products of independent random variables". *SIAM Journal on Applied Mathematics* **14** (1966), pp. 511–526.

[17] Szpankowski, Wojciech. *Average Case Analysis of Algorithms on Sequences.* New York: Wiley-Interscience, 2001.

[18] University of St. Andrews. "Robert Hjalmar Mellin" (biography). http://www-history.mcs.st-andrews.ac.uk/Biographies/Mellin.html.

[19] Wolf, Kurt B. *Integral Transforms in Science and Engineering.* New York: Plenum Press, 1979.

[20] Zemanian, A. H. *Generalized Integral Transforms.* New York: Interscience, 1968.