

# FITTING COMBINATIONS OF EXPONENTIALS TO PROBABILITY DISTRIBUTIONS

Daniel Dufresne

Centre for Actuarial Studies

University of Melbourne

*First version: 3 February 2005*

*This version: 8 November 2005*

## Abstract

Two techniques are described for approximating distributions on the positive half-line by combinations of exponentials. One is based on Jacobi polynomial expansions, and the other on the log-beta distribution. The techniques are applied to some well-known distributions (degenerate, uniform, Pareto, lognormal and others). In theory the techniques yield sequences of combination of exponentials that always converge to the true distribution, but their numerical performance depends on the particular distribution being approximated. An error bound is given in the case the log-beta approximations.

**Keywords:** SUMS OF LOGNORMALS; JACOBI POLYNOMIALS; COMBINATIONS OF EXPONENTIALS; LOG-BETA DISTRIBUTION

## 1. Introduction

The exponential distribution simplifies many calculations, for instance in risk theory, queueing theory, and so on. Often, combinations of exponentials also lead to simpler calculations, and, as a result, there is an interest in approximating arbitrary distributions on the positive half-line by combinations of exponentials. The class of approximating distributions considered in this paper consists in those with a distribution function  $F$  which can be written as

$$1 - F(t) = \sum_{k=1}^n a_j e^{-\lambda_j t}, \quad t \geq 0, \quad \lambda_j > 0, \quad j = 1, \dots, n, \quad 1 \leq n < \infty.$$

When  $a_j > 0$ ,  $j = 1, \dots, n$ , this distribution is a mixture of exponentials (also called a “hyper-exponential” distribution); in this paper, however, we allow some of the  $a_j$  to be negative. The class we are considering is then a subset of the rational family of distributions (also called the matrix-exponential family, or Cox family), which comprises those distributions with a Laplace transform which is a rational function (a ratio of two polynomials). Another subset of the rational family is the class of phase-type distributions (Neuts, 1981), which includes the hyper-exponentials but not all combinations of exponentials. There is an expanding literature on these classes of distributions, see Asmussen & Bladt (1997) for more details and a list of references. For our purposes, the important property of combinations of exponentials is that they are dense in the set of probability distributions on  $[0, \infty)$ .

The problem of fitting distributions from one of the above classes to a given distribution is far from new; in particular, the fitting of phase-type distributions has attracted attention recently, see Asmussen (1997). However, this author came upon the question in particular contexts (see below)

where combinations of exponentials are the ideal tool, and where there is no reason to restrict the search to hyper-exponentials or phase-type distributions. Known methods rely on least-squares, moment matching, and so on. The advantage of the methods described in this paper is that they apply to all distributions, though they perform better for some distributions than for others. For a different fitting technique and other references, see the paper by Feldmann & Whit (1998); they fit hyper-exponentials to a certain class of distributions (a subset of the class of distributions with a tail fatter than the exponential).

The references cited above are related to queueing theory, where there is a lot of interest in phase-type distributions. However, this author has come across approximations by combinations of exponentials in relation to the following three problems.

**Risk theory.** It has been known for some time that the probability of ruin is simpler to compute if the distribution of the claims, or that of the inter-arrival times of claims, is rational, see Dufresne (2001) for details and references. The simplifications which occur in risk theory have been well-known in the literature on random walks (see, for instance, the comments in Feller (1971)) and are also related to queueing theory.

**Convolutions.** The distribution of the sum of independent random variables with a lognormal or Pareto distribution is not known in simple form. Therefore, a possibility is to approximate the distribution of each of the variable involved by a combination of exponentials, and then proceed with the convolution, which is a relatively straightforward affair with combinations of exponentials. This idea is explored briefly in Section 7.

**Financial mathematics.** Suppose an amount of money is invested in the stock market and is used to pay an annuity to a pensioner. What is the probability that the fund will last until the pensioner dies? The answer to this question is known in closed form if the rates of return are lognormal and the lifetime of the pensioner is exponential; this is due to Yor (1992), see also Dufresne (2005). In practice, however, the duration of human life is not exponentially distributed, and the distribution of the required “stochastic life annuity” can be found by expressing the future lifetime distribution as a combination of exponentials. The methods presented below are used to solve this problem in a separate paper (Dufresne, 2004a).

Here is the layout of the paper. Section 2 defines the Jacobi polynomials and gives some of their properties, while Section 3 shows how these polynomials may be used to obtain a convergent series of exponentials that represents a continuous distribution function exactly. In many applications it is not necessary that the approximating combination of exponentials be a true probability distribution, but in other cases this would be a problem. Two versions (A and B) of the method are described; method A yields a convergent sequence of combinations of exponentials which are not precisely density functions, while method B produces true density functions. A different idea is presented in Section 4. It is shown that sequences of log-beta distribution converge to the Dirac mass, and that this can be used to find a different sequence of combinations of exponentials which converges to any distribution on a finite interval.

Sections 5 and 6 apply the methods described to the approximation of the some of the usual distributions: degenerate, uniform, Pareto, lognormal and Makeham and others. The Pareto and lognormal are fat-tail distributions, while the Makeham (used by actuaries to represent future lifetime) has a thin tail. The performance of the methods is discussed in those sections and in the Conclusion. Section 7 shows how the results of Section 3 may be applied to approximating the distribution of the sum of independent random variables.

*Notation and vocabulary.* The cdf of a probability distribution is its cumulative distribution function, the ccdf is the complement of the cdf (one minus the cdf), and the pdf is the probability density function. An *atom* is a point  $x \in \mathbb{R}$  where a measure has positive mass, or, equivalently, where its distribution function has a discontinuity. A *Dirac mass* is a measure which assigns mass 1 to some point  $x$ , and no mass elsewhere; it is denoted  $\delta_x(\cdot)$ . The “big oh” notation has the usual meaning

$$a_n = \mathcal{O}(b_n) \quad \text{if} \quad \lim_{n \rightarrow \infty} \frac{a_n}{b_n} \quad \text{exists and is finite.}$$

Finally,  $(z)_n$  is Pochhammer’s symbol:

$$(z)_0 = 1, \quad (z)_n = z(z+1) \cdots (z+n-1), \quad n \geq 1.$$

## 2. Jacobi polynomials

The Jacobi polynomials are usually defined as

$$P_n^{(\alpha, \beta)}(x) = \frac{(\alpha+1)_n}{n!} {}_2F_1 \left( -n, n + \alpha + \beta + 1, \alpha + 1; \frac{1-x}{2} \right), \quad n = 0, 1, \dots,$$

for  $\alpha, \beta > -1$ . These polynomials are orthogonal over the interval  $[-1, 1]$ , for the weight function

$$(1-x)^\alpha (1+x)^\beta.$$

We will use the “shifted” Jacobi polynomials, which have the following equivalent definitions, once again for  $\alpha, \beta > -1$ :

$$\begin{aligned} R_n^{(\alpha, \beta)}(x) &= P_n^{(\alpha, \beta)}(2x-1) \\ &= \frac{(\alpha+1)_n}{n!} {}_2F_1(-n, n + \alpha + \beta + 1, \alpha + 1; 1-x) \\ &= (-1)^n \frac{(\beta+1)_n}{n!} {}_2F_1(-n, n + \alpha + \beta + 1, \beta + 1; x) \\ &= \frac{(-1)^n}{n!} (1-x)^{-\alpha} x^{-\beta} \frac{d^n}{dx^n} [(1-x)^{\alpha+n} x^{\beta+n}] \\ &= \sum_{j=0}^n \rho_{nj} x^j, \end{aligned}$$

where

$$\rho_{nj} = \frac{(-1)^n (\beta+1)_n (-n)_j (n+\lambda)_j}{(\beta+1)_j n! j!}, \quad \lambda = \alpha + \beta + 1.$$

The shifted Jacobi polynomials are orthogonal on  $[0, 1]$ , for the weight function

$$w^{(\alpha, \beta)}(x) = (1-x)^\alpha x^\beta.$$

The following result is stated in Luke (1969, p.284) (the result is given there in terms of the  $\{P_n^{(\alpha, \beta)}(x)\}$ , but we have rephrased it in terms of the  $\{R_n^{(\alpha, \beta)}(x)\}$ ).

**Theorem 2.1.** *If  $\phi(x)$  is continuous in the closed interval  $0 \leq x \leq 1$  and has a piecewise continuous derivative, then, for  $\alpha, \beta > -1$ , the Jacobi series*

$$\phi(x) = \sum_{n=0}^{\infty} c_n R_n^{(\alpha, \beta)}(x), \quad c_n = \frac{1}{h_n} \int_0^1 \phi(x) (1-x)^\alpha x^\beta R_n^{(\alpha, \beta)}(x) dx$$

$$h_n = \int_0^1 (1-x)^\alpha x^\beta R_n^{(\alpha, \beta)}(x)^2 dx = \frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{(2n+\lambda)n!\Gamma(n+\lambda)},$$

converges uniformly to  $f(x)$  in  $\epsilon < x < 1 - \epsilon$ ,  $0 < \epsilon < 1$ .

This means that the Fourier series

$$s_N(x) = \sum_{n=0}^N c_n R_n^{(\alpha, \beta)}(x)$$

converges to  $f(x)$  as  $N \rightarrow \infty$ , for all  $x$  in  $(0, 1)$ . In the case of functions with discontinuities the limit of the series is

$$\frac{1}{2}[f(x-) + f(x+)],$$

see Chapter 8 of Luke (1969) for more details. Another classical result is the following (see Higgins, 1977).

**Theorem 2.2.** *If  $\phi \in L^2((0, 1), w^{(\alpha, \beta)})$ , that is, if  $\phi$  is measurable and*

$$\int_0^1 \phi^2(x) w^{(\alpha, \beta)}(x) dx < \infty,$$

then  $s_N$  converges in  $L^2((0, 1), w^{(\alpha, \beta)})$  and

$$\lim_{N \rightarrow \infty} \|\phi - s_N\|_{L^2((0, 1), w^{(\alpha, \beta)})} = 0.$$

### 3. Fitting combinations of exponentials to using Jacobi polynomials

#### *Method A: direct Jacobi approximation of the ccdf*

The goal of this section is to approximate the complement of the distribution function by a combination of exponentials:

$$\bar{F}(t) = 1 - F(t) = P(T > t) \approx \sum_j a_j e^{-\lambda_j t}, \quad t \geq 0. \quad (3.1)$$

This can be attempted in many ways, for instance by matching moments. What we suggest is to transform the function to be approximated in the following way:

$$g(x) = \bar{F}\left(-\frac{1}{r} \log(x)\right), \quad 0 < x \leq 1, \quad g(0) = 0,$$

where  $r > 0$ . We have thus mapped the interval  $[0, \infty)$  onto  $(0, 1]$ ,  $t = 0$  corresponding to  $x = 1$ , and  $t \rightarrow \infty$  corresponds to  $x \rightarrow 0+$ ; since  $\bar{F}(\infty) = 0$ , it is justified to set  $g(0) = 0$ . If we find constants  $\alpha, \beta, p$  and  $\{b_k\}$  such that (informally)

$$g(x) = x^p \sum_k b_k R_k^{(\alpha, \beta)}(x), \quad 0 < x \leq 1,$$

then

$$\bar{F}(t) = e^{-prt} \sum_k b_k \sum_j \rho_{kj} e^{-jrt} = \sum_j \left( \sum_k b_k \rho_{kj} \right) e^{-(j+p)rt}, \quad t \geq 0. \quad (3.2)$$

This expression agrees with (3.1), if  $\lambda_j = (j+p)r, j = 0, 1, 2, \dots$  (The interchange of summation order is justified for finite sums, but not necessarily for infinite ones.) Classical Hilbert space theory says that the constants  $\{b_k\}$  can be found by

$$\begin{aligned} b_k &= \frac{1}{h_k} \int_0^1 x^{-p} g(x) R_k^{(\alpha, \beta)}(x) (1-x)^\alpha x^\beta dx \\ &= \frac{r}{h_k} \int_0^\infty e^{-(\beta-p+1)rt} (1-e^{-rt})^\alpha R_k^{(\alpha, \beta)}(e^{-rt}) \bar{F}(t) dt. \end{aligned} \quad (3.3)$$

This is a combination of terms of the form

$$\int_0^\infty e^{-(\beta-p+j+1)rt} (1-e^{-rt})^\alpha \bar{F}(t) dt, \quad j = 0, 1, \dots, k.$$

The reason for using some  $p > 0$  is that it automatically gets rid of any constant which would otherwise appear in the series. If  $\alpha = 0$ , then this integral is the Laplace transform of  $\bar{F}(\cdot)$ , which may be expressed in terms of the Laplace transform of the distribution of  $T$ :

$$\int_0^\infty e^{-\nu t} \bar{F}(t) dt = -\frac{1}{\nu} \int_0^\infty \bar{F}(t) d e^{-\nu t} = \frac{1}{\nu} [1 - \mathbb{E} e^{-\nu T}], \quad \nu > 0. \quad (3.4)$$

The next result is an immediate consequence of Theorem 2.1.

**Theorem 3.1.** *Suppose  $\alpha, \beta > -1$ , that  $\bar{F}(\cdot)$  is continuous on  $[0, \infty)$ , and that the function*

$$e^{prt} \bar{F}(t) \quad (3.5)$$

*has a finite limit as  $t$  tends to infinity for some  $p \in \mathbb{R}$  (this is always verified if  $p \leq 0$ ). Then the integral in (3.3) converges for every  $k \in \mathbb{N}$  and*

$$\bar{F}(t) = e^{-prt} \sum_{k=0}^{\infty} b_k R_k^{(\alpha, \beta)}(e^{-rt}) \quad (3.6)$$

*for every  $t$  in  $(0, \infty)$ , and the convergence is uniform over every interval  $[a, b]$ , for  $0 < a < b < \infty$ .*

Not all distributions satisfy the condition that the limit of (3.5) exists for some  $p > 0$ . The next result, a consequence of Theorem 2.2, does not need this assumption.

**Theorem 3.2.** Suppose  $\alpha, \beta > -1$  and that for some  $p \in \mathbb{R}$  and  $r > 0$

$$\int_0^\infty e^{-(\beta+1-2p)rt} (1 - e^{-rt})^\alpha \bar{F}(t)^2 dt < \infty$$

(this is always true if  $p < (\beta + 1)/2$ ). Then

$$\lim_{N \rightarrow \infty} \int_0^\infty \left[ \bar{F}(t) - e^{-prt} \sum_{k=0}^N b_k R_k^{(\alpha, \beta)}(e^{-rt}) \right]^2 e^{-(\beta+1-2p)rt} (1 - e^{-rt})^\alpha dt = 0,$$

and (3.6) holds almost everywhere, with  $\{b_k; k \geq 0\}$  given by (3.3).

For the purpose of approximating probability distributions by truncating the series, we will restrict our attention to  $0 < p < (\beta + 1)/2$  (this ensures that the truncated series for  $\bar{F}(t)$  tends to 0 as  $t \rightarrow \infty$ ).

One disadvantage of the method just described is that if only a finite number of terms of (3.2) are used, then the resulting approximation of  $F(t)$  is in general not a true distribution function. The function may be smaller than 0 or greater than 1 in places, or it might decrease over some intervals. For the cases where a true distribution function is needed, an alternative technique will now be presented.

**Method B: Modified Jacobi approximations which are true distribution functions**

This alternative method applies only to distributions which have a density  $f(\cdot)$ . It consists in (1) fitting a combination of exponentials to the square root of the density; (2) squaring it; (3) rescaling it so it integrates to 1. We thus consider an expansion of the type

$$\sqrt{f(t)} = e^{-pt} \sum_j a_j \lambda_j e^{-\lambda_j t}.$$

The square of this expression is also a combination of exponentials. The following application of Theorem 2.2 is immediate.

**Theorem 3.3.** Suppose  $\alpha, \beta > -1$ ,  $p \in \mathbb{R}$ ,  $r > 0$ , and that  $f(\cdot)$  is the density of a random variable  $T \geq 0$ . Suppose moreover that

(i)  $E e^{-(\beta+1-2p)T} < \infty$ . This is always satisfied if  $p \leq (\beta + 1)/2$ .

(ii)  $E (T \wedge 1)^\alpha < \infty$ . This is always satisfied if  $\alpha \geq 0$ .

Then

$$\sqrt{f(t)} = e^{-prt} \sum_{k=0}^{\infty} b_k R_k^{(\alpha, \beta)}(e^{-rt})$$

almost everywhere, with

$$b_k = \frac{r}{h_k} \int_0^\infty e^{-(\beta-p+1)rt} (1 - e^{-rt})^\alpha R_k^{(\alpha, \beta)}(e^{-rt}) \sqrt{f(t)} dt, \quad k = 0, 1, \dots$$

## Fitting combinations of exponentials

When truncated, the series for  $\sqrt{f}$  may be negative in places, but squaring it yields a combination of exponentials which is everywhere non-negative. It then only needs to be multiplied by the right constant (one over its integral) to become a true density function.

Examples are given in Section 5. Observe that the integral of the squared truncated series may not converge to one. The usual  $L^2$  results (e.g. Higgins, 1977), say that the norm of the truncated series converges to the norm of the limit, which in this case means that, if we let

$$\phi_N(t) = e^{-prt} \sum_{k=0}^N b_k R_k^{(\alpha, \beta)}(e^{-rt}),$$

be the  $(N + 1)$ -term approximation of  $\sqrt{f}$ , then

$$\lim_{N \rightarrow \infty} \int_0^\infty \phi_N(t)^2 e^{-(\beta+1-2p)rt} (1 - e^{-rt})^\alpha dt = \int_0^\infty f(t) e^{-(\beta+1-2p)rt} (1 - e^{-rt})^\alpha dt.$$

Only in the case where  $\beta + 1 - 2p = \alpha = 0$  does this guarantee that the integral of  $\phi_N^2$  tends to 1.

A variation on the same theme consists in finding a combination of exponentials which approximates

$$g(t) = \int_t^\infty \sqrt{f(s)} ds,$$

differentiate with respect to  $t$  and then square as before (provided of course that the integral above converges). The advantage is that the same computer code may be used as for Method A.

## 4. Fitting combinations of exponentials using the log-beta distribution

This section considers approximating distributions over a finite interval by combinations of exponentials. It will be shown that a converging sequence of approximations can be constructed by starting with any sequence of combinations of exponentials which converges in distribution to a constant. A convenient sequence of exponentials is obtained from the lobgeta distribution, as will be seen below.

Suppose the goal is to find a sequence of combination of exponentials which converge to distribution of a variable  $T$ , which takes values in  $[0, t_0]$  only. Suppose that a sequence  $\{X_n\}$  of variables with densities which are combinations of exponentials converges in distribution to  $t_0$  as  $n \rightarrow \infty$ ; suppose moreover that  $\{X_n\}$  is independent of  $T$ . Then  $T_n = T + X_n - t_0$  converges in distribution to  $T$ . Denote the density of  $X_n$  by

$$f_{X_n}(t) = \sum_j a_j^{(n)} \lambda_j^{(n)} e^{-\lambda_j^{(n)} t} 1_{\{t > 0\}}.$$

Each variable  $T_n$  thus has a density

$$\begin{aligned} f_{T_n}(t) &= \int f_{X_n}(t - u) dF_{T-t_0}(u) \\ &= \sum_{j=0}^{n-1} a_j^{(n)} \lambda_j^{(n)} e^{-\lambda_j^{(n)} t} \int e^{\lambda_j^{(n)} u} 1_{\{u < t\}} dF_{T-t_0}(u) \\ &= \sum_{j=0}^{n-1} a_j^{(n)} \lambda_j^{(n)} e^{-\lambda_j^{(n)} t} e^{-\lambda_j^{(n)} t_0} \int e^{\lambda_j^{(n)} y} 1_{\{y < t+t_0\}} dF_T(y). \end{aligned}$$

The integral in the last expression has the same value for all  $t \geq 0$  (but not for  $t < 0$ ). Hence,  $f_{T_n}(t)$  is a combination of exponentials for all  $t \geq 0$ . Since  $\mathbf{P}(T_n < 0)$  tends to 0 as  $n$  increases, the following result is obtained.

**Theorem 4.1.** *Under the previous assumptions, the distributions with pdf's equal to the combinations of exponentials*

$$f_{C_n}(t) = \sum_{j=0}^{n-1} b_j^{(n)} \lambda_j^{(n)} e^{-\lambda_j^{(n)} t} 1_{\{t \geq 0\}}, \quad b_j^{(n)} = a_j^{(n)} e^{-\lambda_j^{(n)} t_0} \mathbf{E} e^{\lambda_j^{(n)} T} / \mathbf{P}(T_n \geq 0)$$

converge to the distribution of  $T$ .

It is possible to find an upper bound for the absolute difference between the true distribution function of  $T$  and distribution function  $F_{C_n}(\cdot)$  of the combination of exponentials in Theorem 4.1, when  $T$  has a continuous distribution.

**Theorem 4.2.** *Under the previous assumptions, if it is moreover assumed that  $T$  has a pdf  $f_T(\cdot)$ ,*

$$\begin{aligned} \sup_z |F_T(z) - F_{C_n}(z)| &\leq [\sup_t f_T(t)] \mathbf{E}|X_n - t_0| + F_{T_n}(0) \\ &\leq [\sup_t f_T(t)] [\mathbf{E}(X_n - t_0)^2]^{\frac{1}{2}} + F_{T_n}(0). \end{aligned}$$

**Proof.** First,

$$\begin{aligned} |\bar{F}_T(z) - \bar{F}_{C_n}(z)| &= \left| \bar{F}_T(z) - \frac{\bar{F}_{T_n}(z)}{\bar{F}_{T_n}(0)} \right| \\ &\leq |\bar{F}_T(z) - \bar{F}_{T_n}(z)| + \bar{F}_{T_n}(z) \left( \frac{1 - \bar{F}_{T_n}(0)}{\bar{F}_{T_n}(0)} \right) \\ &\leq |\bar{F}_T(z) - \bar{F}_{T_n}(z)| + F_{T_n}(0). \end{aligned}$$

Second, let  $Y_n = X_n - t_0$ . Then

$$\begin{aligned} |F_T(z) - F_{T_n}(z)| &= |F_T(z) - \mathbf{E}F_T(z - Y_n)| \\ &\leq \int \left| \int_{z-y}^z f_T(u) du \right| dF_{Y_n}(y) \\ &\leq [\sup_t f_T(t)] \int |y| dF_{Y_n} = [\sup_t f_T(t)] \mathbf{E}|Y_n|. \quad \square \end{aligned}$$

In applications, the mean of  $X_n$  will be close to  $t_0$ , and  $F_{T_n}(0)$  will be small, so the error bound should be close to

$$[\sup_t f_T(t)] (\mathbf{Var} X_n)^{\frac{1}{2}}.$$

The log-beta family of distributions will now be introduced, because it includes explicit combinations of exponentials which converge to a constant.



**Definition.** For parameters  $a, b, c > 0$ , the **log-beta distribution** has density

$$f_{a,b,c}(x) = K_{a,b,c} e^{-bx} (1 - e^{-cx})^{a-1} \mathbf{1}_{\{x>0\}}.$$

This law is denoted **LogBeta**( $a, b, c$ ).

The name chosen is justified by the fact that

$$X \sim \mathbf{LogBeta}(a, b, c) \Leftrightarrow e^{-cX} \sim \mathbf{Beta}(b/c, a).$$

The distribution is then a transformed beta distribution that includes the gamma as a limiting case, since

$$\mathbf{LogBeta}(1, b, c) = \mathbf{Exp}(b), \quad \lim_{c \rightarrow 0^+} \mathbf{LogBeta}(a, b, c) = \mathbf{Gamma}(a, b).$$

The log-beta distribution is unimodal, just as the beta distribution.

(*N.B.* It would be possible to define the **LogBeta**( $a, b, c$ ) for  $a, b > 0$  and  $-b/a < c < 0$ , since

$$(1 - e^{-cx})^{a-1} e^{-bx} = (-1)^{a-1} (1 - e^{cx})^{a-1} e^{-(b+ac)x}.$$

is integrable over  $(0, \infty)$ . However, this law would then be identical to the **LogBeta**( $a, b + ac, -c$ ), and no greater generality would be achieved.)

The next theorem collects some properties of the log-beta distributions; the proof may be found in the Appendix. The digamma function is defined as

$$\psi(z) = \frac{1}{\Gamma(z)} \frac{d}{dz} \Gamma(z) = \frac{d}{dz} \log \Gamma(z).$$

It is known that (Lebedev, 1972, p.7)

$$\psi(z) = -\gamma + \sum_{k=0}^{\infty} \left( \frac{1}{k+1} - \frac{1}{k+z} \right), \quad (4.1)$$

where  $\gamma$  is Euler's constant. The derivatives  $\psi'(\cdot)$ ,  $\psi''(\cdot)$ ,  $\psi^{(3)}(\cdot)$ , ... of the digamma function are known as the polygamma functions, and can be obtained by differentiating (4.1) term by term.

**Theorem 4.3. (Properties of the log-beta distribution)** Suppose  $X \sim \mathbf{LogBeta}(a, b, c)$ ,  $a, b, c > 0$ .

(a)  $K_{a,b,c} = \frac{c \Gamma(\frac{b}{c} + a)}{\Gamma(\frac{b}{c}) \Gamma(a)}$ .

(b)  $E e^{-sX} = \frac{\Gamma(\frac{b}{c} + a) \Gamma(\frac{b+s}{c})}{\Gamma(\frac{b+s}{c} + a) \Gamma(\frac{b}{c})}$ ,  $s > -b$ . If  $\kappa > 0$  and  $Y = \kappa X$ , then  $Y \sim \mathbf{LogBeta}(a, b/\kappa, c/\kappa)$ .

(c) The cumulants of the distribution are

$$\kappa_n = \frac{(-1)^{n-1}}{c^n} \left[ \psi^{(n-1)} \left( \frac{b}{c} + a \right) - \psi^{(n-1)} \left( \frac{b}{c} \right) \right], \quad n = 1, 2, \dots$$

In particular,

$$\begin{aligned} \mathbf{E} X &= \frac{1}{c} \left[ \psi \left( \frac{b}{c} + a \right) - \psi \left( \frac{b}{c} \right) \right]; & \mathbf{Var} X &= \frac{1}{c^2} \left[ \psi' \left( \frac{b}{c} \right) - \psi' \left( \frac{b}{c} + a \right) \right] \\ \mathbf{E}(X - \mathbf{E} X)^3 &= \frac{1}{c^3} \left[ \psi'' \left( \frac{b}{c} + a \right) - \psi'' \left( \frac{b}{c} \right) \right]. \end{aligned}$$

(d) If  $c > 0$  is fixed,  $b = \kappa a$  for a fixed constant  $\kappa > 0$ , then

$$X \xrightarrow{d} x_0 = \frac{1}{c} \log \left( 1 + \frac{c}{\kappa} \right) \quad \text{as } a \rightarrow \infty.$$

Moreover,

$$\lim_{a \rightarrow \infty} \mathbf{E} X^k = x_0^k, \quad k = 1, 2, \dots$$

(e) If  $a = n \in \mathbb{N}_+$ , then

$$X \stackrel{d}{=} Y_0 + \dots + Y_{n-1},$$

where

$$Y_j \sim \mathbf{Exp}(b + jc), \quad j = 0, \dots, n-1,$$

are independent, and

$$\mathbf{E} X = \sum_{j=0}^{n-1} \frac{1}{b + jc}, \quad \mathbf{Var} X = \sum_{j=0}^{n-1} \frac{1}{(b + jc)^2}, \quad \mathbf{E}(X - \mathbf{E} X)^3 = 2 \sum_{j=0}^{n-1} \frac{1}{(b + jc)^3}.$$

(f) If  $a = n \in \mathbb{N}_+$ , the density of  $X$  is

$$f_{n,b,c}(x) = \sum_{j=0}^{n-1} \alpha_j \lambda_j e^{-\lambda_j x},$$

where

$$\alpha_j = c \left( \frac{b}{c} \right)_n \frac{(-1)^j}{j!(n-1-j)!(b+jc)}, \quad \lambda_j = b + jc, \quad j = 0, \dots, n-1.$$

(g) Under the same assumptions as in (d),

$$\frac{X - \mathbf{E} X}{\sqrt{\mathbf{Var} X}} \xrightarrow{d} N(0, 1) \quad \text{as } a \rightarrow \infty.$$

(h) Suppose  $X_1 \sim \mathbf{LogBeta}(a_1, b, c)$  and  $X_2 \sim \mathbf{LogBeta}(a_2, b + a_1 c, c)$  are independent. Then  $X_1 + X_2 \sim \mathbf{LogBeta}(a_1 + a_2, b, c)$

**Corollary 4.4.** For  $n \in \mathbb{N}_+$  and any fixed  $\kappa, c, y_0 > 0$ , the  $\mathbf{LogBeta}(n, n\kappa x_0/y_0, cx_0/y_0)$  distribution has a pdf which is a combination of exponentials, and it tends to  $\delta_{y_0}$  as  $n \rightarrow \infty$ , if  $x_0 = \log(1 + c/\kappa)/c$ . The same holds if  $x_0$  is replaced with the mean of the  $\mathbf{LogBeta}(n, n\kappa, c)$  distribution.

*Remark.* The classical result that combinations of exponentials are dense in the set of distributions on  $\mathbb{R}_+$  may thus be proved as follows: approximate the given distribution by a discrete distribution with a finite number of atoms; then approximate each atom by a log-beta distribution as in Corollary 4.4.

Section 6 applies Corollary 4.4 to some of the usual distributions. It is possible to be a little explicit about the maximum error made by using the combination of exponentials obtained (see Theorem 4.2). The proof of the next result is in the Appendix.

**Theorem 4.5.** Suppose  $X_n \sim \mathbf{LogBeta}(n, n\kappa x_0/t_0, cx_0/t_0)$  as in Corollary 4.4. Then

$$\text{Var}X_n \sim \frac{1}{n} \frac{t_0^2}{x_0^2} \frac{1}{\kappa(\kappa + c)} \quad \text{as } n \rightarrow \infty.$$

If, moreover, the variable  $T$  ( $0 \leq T \leq t_0$ ) has a pdf  $f_T(\cdot)$ , and  $\mathbb{E}T^{-p} < \infty$  for some  $p > 1$ , then  $F_{T_n}(0) = \mathcal{O}(n^{-\frac{p \wedge 2}{2}})$ , and the upper bound in Theorem 4.4 becomes

$$\sup_z |F_T(z) - F_{C_n}(z)| \leq [\sup_t f_T(t)] [\text{Var}X_n]^{\frac{1}{2}} + F_{T_n}(0) \sim [\sup_t f_T(t)] [\text{Var}X_n]^{\frac{1}{2}} = \mathcal{O}(1/\sqrt{n}).$$

Hence, under the assumptions of this theorem, a rule of thumb is that the maximum error is approximately

$$\frac{1}{\sqrt{n}} [\sup_t f_T(t)] \frac{t_0}{x_0} \frac{1}{\sqrt{\kappa(\kappa + c)}}.$$

## 5. Examples of Jacobi approximations

We apply the ideas above to some parametric distributions. The analysis focuses on the cdf or ccdf, rather than on the pdf.

For Methods A and B some trial and error is required to determine “good” parameters  $\alpha$ ,  $\beta$ ,  $p$  and  $r$ . All the computations were performed with Mathematica. It is not suggested that the particular approximations presented are the best in any sense. In the case of continuous distributions, a measure of the accuracy of an approximation  $\hat{F}$  of a distribution function  $F$  is the maximum of the difference between the true distribution function  $F$  and an approximation  $\hat{F}$ , which we denote

$$\|F - \hat{F}\| = \sup_{t \geq 0} |F(t) - \hat{F}(t)|.$$

Some of the low-order approximations are shown in graphs, as their larger errors make the curves distinguishable.

**Example 5.1. The degenerate distribution at  $x = 1$**

In this case the Jacobi polynomials lead to highly oscillatory approximations. This can be seen in Figure 1. The log-beta approximations do better, but at the cost of a larger number of exponentials. See next section.

**Example 5.2. The uniform distribution**

This is not known to be the easiest distribution to approximate by exponentials. Figure 2 shows the ccdf of the  $U(1, 2)$  distribution, as well as two approximations found for it using method A. The parameters used are

$$\alpha = 0, \quad \beta = 0, \quad r = .7, \quad p = .1.$$

The 3-term approximation of the ccdf is (Method A) is

$$\bar{F}_3^A(t) = -0.364027e^{-0.07t} + 3.497327e^{-0.77t} - 2.137705e^{-1.47t}$$

and is not good. The 10-term approximation does better. Both approximations are sometimes smaller than 0, sometimes greater than 1, and sometimes increasing, all things that a ccdf should not do. The accuracies do improve with the number of terms used:

$$\|F - F_3^A\| = .31, \quad \|F - F_5^A\| = .16, \quad \|F - F_{10}^A\| = .065, \quad \|F - F_{20}^A\| = .038.$$

Method B is illustrated in Figure 3. In this particular case there is a very convenient simplification, since  $\sqrt{f} = f$ . The same Jacobi approximation is thus used for  $\sqrt{f}$ . Differentiating  $\bar{F}_3$ , squaring and rescaling yields

$$\bar{F}_5^B(t) = 0.01014e^{-0.14t} - 0.357203e^{-.84t} + 10.522656e^{-1.54t} - 16.518849e^{-2.24t} + 7.343256e^{-2.94t}.$$

There are now  $2 \times 3 - 1 = 5$  distinct frequencies  $\lambda_j$ . Some of the maximum errors observed are:

$$\|F - F_3^B\| = .8, \quad \|F - F_5^B\| = .26, \quad \|F - F_{11}^B\| = .28, \quad \|F - F_{21}^B\| = .11.$$

The B approximations are all true distributions (here the 11-term B approximation happens to be a little worse than the 5-term). The 5-term B approximation does slightly better than than the 3-term A approximation, although it is really based on the same 3-term Jacobi polynomial. However, the 5-term A approximation has better precision, although it is not a true ccdf.

Overall, the Jacobi approximation is not terribly efficient for the uniform distribution, as the maximum errors above show. The graphs also reveal that the cdf (and therefore pdf) oscillates significantly. Other computations also indicate that the accuracy of the approximation varies with the interval chosen for the uniform distribution. The log-beta approximations are much smoother, see Section 6.

**Example 5.3. The Pareto distribution**

The tail of the Pareto distribution is heavier than that of a combination of exponentials, and it is not obvious *a priori* that the approximations we are considering would perform well. We consider a Pareto distribution with ccdf

$$\bar{F}(t) = \frac{1}{1+t}, \quad t \geq 0.$$

Figure 4 compares the 3 and 5-term A approximations for the cdf, while Figure 5 shows the same 3-term A approximation against the 5-term B approximation. The parameters are

$$\alpha = 0, \quad \beta = 0, \quad r = .1, \quad p = .1.$$

The accuracy increases with the number of terms used:

$$\|F - F_3^A\| = .31, \quad \|F - F_5^A\| = .12, \quad \|F - F_{10}^A\| = .006, \quad \|F - F_{20}^A\| = .0015.$$

Finding the B approximations was easy, because the square root of the pdf of this distribution is precisely its CDF, so the same Mathematica code could be used. The same parameters  $\alpha, \beta, r$  and  $p$  were used, and produced:

$$\|F - F_3^B\| = .31, \quad \|F - F_5^B\| = .24, \quad \|F - F_{11}^B\| = .018, \quad \|F - F_{21}^B\| = .0038.$$

#### Example 5.4. The lognormal distribution

The lognormal distribution has a tail which is heavier than the exponential, but lighter than the Pareto. This may explain why combinations of exponentials appear to do a little better with the lognormal than with the Pareto (see Figures 6 and 7). The parameters of the lognormal are 0 and .25 (meaning that the logarithm of a variable with this distribution has a normal distribution with mean 0 and variance .25). The first parameter of the lognormal has no importance (it only affects the scale of the distribution), but the numerical experiments performed appear to show that smaller values of the second parameter lead to worse approximations.

Figure 6 shows that the 3-term approximation does not do well at all, but that the 5-term is quite good (the exact and approximate may be difficult to distinguish). The parameters used are

$$\alpha = 0, \quad \beta = .5, \quad r = 1, \quad p = .2,$$

and some of the precisions obtained are

$$\|F - F_3^A\| = .11, \quad \|F - F_5^A\| = .014, \quad \|F - F_{10}^A\| = .0003, \quad \|F - F_{20}^A\| = 1.5 \times 10^{-6}.$$

The Method B approximations were found by first noting that the square root of the pdf of the **Lognormal**( $\mu, \sigma^2$ ) distribution is a constant times the pdf of a **Lognormal**( $\mu + \sigma^2, 2\sigma^2$ ):

$$\left[ \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\log(x)-\mu)^2/(2\sigma^2)} \right]^{1/2} = (8\pi\sigma^2)^{1/4} e^{(\sigma^2+2\mu)/4} \frac{1}{x\sigma\sqrt{2}\sqrt{2\pi}} e^{-[(\log x - (\mu + \sigma^2))]^2/4\sigma^2}.$$

It is possible to find an approximating combination of exponentials by computing a Jacobi expansion for the pdf on the right, or else by finding one for the cdf of the same distribution, and then differentiating. The two possibilities were compared, and the differences were minute. It was therefore decided to adopt the latter approach. The parameters used are

$$\alpha = 0, \quad \beta = -.3, \quad r = .7, \quad p = .1,$$

and some of the results found are

$$\|F - F_3^B\| = .46, \quad \|F - F_5^B\| = .23, \quad \|F - F_{11}^B\| = .0065, \quad \|F - F_{21}^B\| = .0003.$$

These are less accurate than the corresponding A approximations, but they do improve as the number of terms increases. Figure 7 compares the true ccdf with the 5 and 11 term B approximations. Overall, the approximations did better for the lognormal than for the Pareto in the specific cases chosen, but no general conclusion can be made based on two specific cases.

**Example 5.5. The Makeham distribution**

This distribution is used to represent the lifetime of insureds. It has failure rate

$$\mu_x = A + Bc^x.$$

We chose the parameters used in Bowers *et al.* (1997, p.78):

$$A = .0007, \quad B = 5 \times 10^{-5}, \quad c = 10^{.04}.$$

Figure 8 shows how the 3 and 7-term approximations do. The parameters chosen were

$$\alpha = 0, \quad \beta = 0, \quad r = .08, \quad p = .2,$$

and higher degree Jacobi approximations have accuracies

$$\|F - F_3^A\| = .08, \quad \|F - F_5^A\| = .04, \quad \|F - F_{10}^A\| = .0065, \quad \|F - F_{20}^A\| = .00024.$$

The accuracy is worse than with the lognormal. An explanation may be that the tail of the Makeham is much thinner than that of the exponential.

**6. Examples of log-beta approximations**

**Example 6.1. The degenerate distribution at  $x = 1$**

Figure 9 shows two log-beta approximations for a Dirac mass at  $x = 1$ . The combinations of exponentials used are the **LogBeta**( $n, n\kappa, c$ ) distribution with  $n = 20$  and  $200$ ,  $c = .1$  and  $\kappa = .1/(e^{.1} - 1)$ . The approximate ccdf's are very smooth, by comparison with the Jacobi approximations (Figure 1).

**Example 6.2. The uniform distribution**

The pdf and ccdf of some log-beta approximations for the **U**(1, 2) distribution are shown in Figures 10 and 11. The **LogBeta**( $n, n\kappa, c$ ) distribution with  $n = 20$  and  $200$ ,  $c = .1$  and  $\kappa = .1/(e^{.2} - 1)$  was used. The approximations do not oscillate at all, as the Jacobi approximations do (Figures 2 and 3). The precisions attained were

$$\|F_T - F_{C_{20}}\| = .178, \quad \|F_T - F_{C_{200}}\| = .0567.$$

The error bounds given by Theorem 4.2 are 0.451 and 0.142, respectively. In this case,  $ET^{-p} < \infty$  for any  $p$ , and  $P(T_n < 0)$  (which is a term in the error bound in Theorem 4.2) tends to 0 very

quickly:  $P(T_{20} < 0) = .000265$ , while  $P(T_{200} < 0)$  is of the order of  $10^{-9}$ . The asymptotic expression for  $\text{Var}X_n$  (Theorem 4.5) is extremely good here:

$$\begin{aligned}\text{Var}X_{20} &= 0.202698, & \frac{1}{20} \frac{1}{\kappa(\kappa + c)} &= 0.200668 \\ \text{Var}X_{200} &= 0.020087, & \frac{1}{200} \frac{1}{\kappa(\kappa + c)} &= 0.0200668.\end{aligned}$$

***Example 6.3. Squared sine distribution***

In order to test this method for a density which may not be as friendly as the previous two, consider the pdf

$$f_T(t) = 2 \sin^2(2\pi t) 1_{(0,1)}(t).$$

The **LogBeta** $(n, n\kappa, c)$  distribution with  $n = 20$  and  $200$ ,  $c = .1$  and  $\kappa = .1/(e^1 - 1)$  was used. The results are shown in Figures 12 and 13. The precisions were

$$\|F_T - F_{C_{20}}\| = .125, \quad \|F_T - F_{C_{200}}\| = .0296.$$

The error bounds given by Theorem 4.2 are .521 and .142, respectively. In this case  $ET^{-2} < \infty$ ,  $P(T_{20} < 0) = 0.0725$ , and  $P(T_{200} < 0) = .00590$ . The asymptotic expression for  $\text{Var}X_n$  (Theorem 4.5) is excellent in this case too:

$$\begin{aligned}\text{Var}X_{20} &= 0.0502929, & \frac{1}{20} \frac{1}{\kappa(\kappa + c)} &= 0.0500417 \\ \text{Var}X_{200} &= 0.00500667, & \frac{1}{200} \frac{1}{\kappa(\kappa + c)} &= 0.00500417.\end{aligned}$$

**7. Application to sums of independent variables**

Textbook examples of explicit distributions of sums of independent random variables are the normal and the gamma distributions, in the latter case only if the scale parameters are the same. In applications, however, convolutions of other distributions routinely appear, and some numerical technique is needed to approximate the distribution of the sum of two random variables. This explains why numerical convolutions (or simulations) are often required. The convolutions of combinations of exponentials being relatively easy to calculate, one might ask whether the approximations of Section 5 could be used to approximate convolutions of distributions.

Table 1 presents the results of some experiments performed with the lognormal distribution. The exact distribution function of the sum of two independent lognormal distributions with first parameter 0 and second parameter  $\sigma^2$  was computed numerically, as well as its  $N$ -term combination of exponentials approximation (the convolution of the approximation found as in Section 5). The last column shows the maximum difference between the distribution functions. It is interesting to note that under Method A the error for the convolution is very nearly double what it is for the original distribution, while under Method B the two errors are about the same. This author does not have a good explanation for this difference between the methods, and does not know whether the same holds more generally. More work is required here.

Another possible technique is the following. Section 3 showed that when  $\alpha = 0$  the Fourier coefficients are a weighted sum of the Laplace transform of the cdf of the distribution (see (3.3)). The latter is given by (3.4). Hence, the Jacobi expansion for the cdf of the convolution of probability distributions may be found directly from the Laplace transforms of each of the individual distributions, since

$$\mathbb{E} e^{-\mu(T_1+\dots+T_m)} = \prod_{k=1}^m \mathbb{E} e^{-\mu T_k}.$$

For more on sums of lognormals, the reader is referred to Dufresne (2004b).

**Table 1**  
Accuracy of convolutions of lognormals

$\sigma^2$	Approx.	N terms	$\ F - F_N\ $	$\ F^{*2} - F_N^{*2}\ $
1	A	3	.050	.102
1	A	7	.018	.037
1	A	11	.003	.006
.25	A	3	.105	.204
.25	A	7	.004	.008
.25	A	11	.00025	.0005
.25	B	5	.223	.205
.25	B	9	.050	.051
.25	B	13	.0015	.001

## Conclusion

Two advantages of the Jacobi Methods A and B presented are, first, their simplicity, and, second, the fact that all what is needed is the Laplace transform of the distribution at a finite number of points. Two disadvantages may be mentioned. First, although the sequence converges to the true value of the distribution function at all its continuity points, the truncated series under Method A is in general not a true distribution function. Second, distributions which have atoms are more difficult to approximate. A numerical problem which arises with these methods is the magnitude of the Fourier coefficients as the order of the approximation increases; for instance, some 20-term approximations have been seen to include coefficients of the order of  $10^{25}$ .

For completely monotone distributions, there is an alternative method, presented in Feldmann & Whit (1998) (but going back to de Prony (1795)) which yields a hyper-exponential approximation (this technique applies to some Weibull distributions and to the Pareto distribution, but not to the lognormal or uniform). The coefficients of a hyper-exponential being all positive, they necessarily must all be small.

The log-beta family includes explicit sequences which are combinations of exponentials and also converge to the Dirac mass at any  $y_0 \geq 0$ ; as shown in Section 4, this leads to a technique



for approximating any distribution on a finite interval by combinations of exponentials. These approximations do not have the oscillatory behaviour of the Jacobi approximations. The examples in Section 6 show that it is possible to use log-beta combinations of exponentials of fairly high order (the author has used  $n = 600$  in some cases) without incurring numerical difficulties. The precisions achieved are modest, by comparison with the Jacobi methods. I have found, however, that the log-beta method required far less trial and error than the Jacobi methods. The main disadvantage of the log-beta approximations is that they apply only to distributions over a finite interval. A positive point is, however, that there is a simple, computable error bound (Theorems 4.2 and 4.5) for the log-beta approximations. Another advantage is that the log-beta approximating pdf's are automatically positive everywhere.

## APPENDIX: PROOFS OF THEOREMS 4.3 AND 4.5

**Proof of Theorem 4.3.** Parts (a) and (b) follow from (if  $s > -b$ )

$$\int_0^\infty e^{-(b+s)x} (1 - e^{-cx})^{a-1} dx = \frac{1}{c} \int_0^1 u^{\frac{b+s}{c}-1} (1-u)^{a-1} du = \frac{\Gamma(\frac{b+s}{c})\Gamma(a)}{c\Gamma(\frac{b+s}{c} + a)}.$$

Part (c) is obtained by differentiating the logarithm of Laplace transform in the usual way.

To prove the convergence in distribution in (d), one approach is to recall that  $e^{-cX}$  has a beta distribution, which yields explicit formulas for the mean and variance of that variable. The mean equals  $\kappa/(c + \kappa)$ , and the variance tends to 0. Alternatively, use the formula (Lebedev, 1972, p.15)

$$\frac{\Gamma(z + \alpha)}{\Gamma(z + \beta)} = z^{\alpha-\beta} \left[ 1 + \frac{(\alpha - \beta)(\alpha + \beta - 1)}{2z} + \mathcal{O}(|z|^{-2}) \right] \quad \text{as } |z| \rightarrow \infty,$$

where  $\alpha$  and  $\beta$  are arbitrary constants and  $|\arg z| \leq \pi - \delta$ , for some  $\delta > 0$ . It can be seen that it implies

$$\lim_{a \rightarrow \infty} \mathbf{E} e^{-sX} = \lim_{a \rightarrow \infty} \frac{\Gamma(\frac{\kappa a}{c} + a)\Gamma(\frac{\kappa a + s}{c})}{\Gamma(\frac{\kappa a + s}{c} + a)\Gamma(\frac{\kappa a}{c})} = \left( \frac{\kappa}{\kappa + c} \right)^{\frac{s}{c}}$$

for all  $s \in \mathbb{R}$ . The convergence of moments results from the convergence of cumulants, which itself is a consequence of the expression in (c) together with the asymptotic formulas (Abramowitz & Stegun, 1972, Chapter 6):

$$\begin{aligned} \psi(z) &\sim \log(z) - \frac{1}{2z} - \frac{1}{12z^2} + \dots \\ \psi^{(n)}(z) &\sim (-1)^{(n-1)} \left[ \frac{(n-1)!}{z^n} + \frac{n!}{2z^{n+1}} + \dots \right], \quad n = 1, 2, \dots \end{aligned} \tag{A.1}$$

(valid as  $z \rightarrow \infty$  if  $|\arg z| < \pi$ ).

Next, turn to (e). If  $a = n \in \mathbb{N}_+$ , then

$$\mathbf{E} e^{-sX} = \frac{\left(\frac{b}{c}\right)_n}{\left(\frac{b+s}{c}\right)_n} = \frac{b}{b+s} \frac{b+c}{b+c+s} \dots \frac{b+(n-1)c}{b+(n-1)c+s}.$$

This proves the representation of  $X$  as the sum of independent exponential variables. The mean, variance and third central moment then follow by the additivity of cumulants, or else by using (4.1) and part (c) of this theorem.

The expression in (f) is found by expanding the density using the binomial theorem.

To prove (g), first observe that, using (b), given  $\kappa > 0$ ,

$$X \sim \mathbf{LogBeta}(a, \kappa a, c) \Leftrightarrow \kappa X \sim \mathbf{LogBeta}(a, a, c/\kappa).$$

It is then sufficient to prove the result for  $\kappa = 1$ . First let  $a = n \in \mathbb{N}$ , and express  $X$  as

$$X \stackrel{d}{=} Y_{n,0} + \cdots + Y_{n,n-1},$$

where the variables on the right are independent, and  $Y_{n,j} \sim \mathbf{Exp}(n + jc)$ . We will use the Lyapounov central Limit Theorem (Billingsley, 1987, p.371), which says that the normal limit holds if

$$\lim_{n \rightarrow \infty} \frac{1}{S_n^{2+\delta}} \sum_{j=0}^{n-1} \mathbb{E}|Y_{n,j} - \mathbb{E}Y_{n,j}|^{2+\delta} = 0 \quad (\text{A.2})$$

for some  $\delta > 0$ , with

$$S_n^2 = \sum_{j=0}^{n-1} \text{Var } Y_{n,j}.$$

Let  $\delta = 2$ . If  $Y \sim \mathbf{Exp}(1)$ , then

$$\text{Var } Y = 1, \quad \mathbb{E}(Y - 1)^4 = 9.$$

This means that

$$\text{Var } Y_{n,j} = \frac{1}{(n + jc)^2}, \quad \mathbb{E}(Y_{n,j} - \mathbb{E}Y_{n,j})^4 = \frac{9}{(n + jc)^4}.$$

Then

$$S_n^2 \geq \frac{1}{n(1+c)^2}, \quad \sum_{j=0}^{n-1} \mathbb{E}(Y_{n,j} - \mathbb{E}Y_{n,j})^4 \leq \frac{9}{n^3},$$

which implies (A.2). For arbitrary  $a$ , let  $n = \lfloor a \rfloor$ ,  $r = a - n$ , and note that

$$\mathbb{E} e^{-sX} = \left[ \prod_{j=0}^{n-1} \frac{a + (r+j)c}{a + (r+j)c + s} \right] \frac{\Gamma(\frac{a}{c} + r)\Gamma(\frac{a+s}{c})}{\Gamma(\frac{a+s}{c} + r)\Gamma(\frac{a}{c})}.$$

Hence,

$$X \stackrel{d}{=} Y_{a,0} + \cdots + Y_{a,n-1} + Y_{a,n},$$

where  $Y_{a,j} \sim \mathbf{Exp}(b + (r+j)c)$ ,  $j = 0, \dots, n-1$ , and  $Y_{a,n} \sim \mathbf{LogBeta}(r, a, c)$ . The Lyapounov central Limit Theorem can again be applied, the only real difference is the additional term  $Y_{a,n}$ . This term does not change the limit, since it can be observed (see (4.1) and part (c)) that

$$\mathbb{E} Y_{a,n} < \frac{1}{a}, \quad \text{Var } Y_{a,n} < \frac{1}{a^2}.$$

This implies that

$$\lim_{a \rightarrow \infty} \frac{\text{Var}(Y_{a,0} + \cdots + Y_{a,n-1})}{\text{Var}(Y_{a,0} + \cdots + Y_{a,n})} = 1, \quad \frac{Y_{a,n} - \mathbf{E} Y_{a,n}}{\sqrt{\text{Var}(Y_{a,0} + \cdots + Y_{a,n})}} \xrightarrow{d} 0 \quad \text{as } a \rightarrow \infty.$$

Hence, including or excluding  $Y_{n,n}$  in the normalised expression for  $X$  does not change the limit distribution. Finally, (h) may be seen to follow from writing the product of the Laplace transforms of  $X_1$  and  $X_2$ :

$$\frac{\Gamma(\frac{b}{c} + a_1)\Gamma(\frac{b+s}{c})}{\Gamma(\frac{b+s}{c} + a_1)\Gamma(\frac{b}{c})} \times \frac{\Gamma(\frac{b}{c} + a_1 + a_2)\Gamma(\frac{b+s}{c} + a_1)}{\Gamma(\frac{b+s}{c} + a_1 + a_2)\Gamma(\frac{b}{c} + a_1)} = \frac{\Gamma(\frac{b}{c} + a_1 + a_2)\Gamma(\frac{b+s}{c})}{\Gamma(\frac{b+s}{c} + a_1 + a_2)\Gamma(\frac{b}{c})},$$

or else may be obtained by performing the convolution explicitly: for  $y > 0$ ,

$$\begin{aligned} \int_0^y e^{-b(y-x)}(1 - e^{-c(y-x)})^{a_1-1} e^{-(b+a_1c)x} (1 - e^{-cx})^{a_2-1} dx \\ &= e^{-by} \int_0^y e^{-cx} (e^{-cx} - e^{-cy})^{a_1-1} (1 - e^{-cx})^{a_2-1} dx \\ &= \frac{e^{-by}}{c} \int_{e^{-cy}}^1 (u - e^{-cy})^{a_1-1} (1 - u)^{a_2-1} du \\ &= e^{-by} (1 - e^{-cy})^{a_1+a_2-1} \frac{1}{c} \int_0^1 v^{a_1-1} (1 - v)^{a_2-1} dv. \quad \square \end{aligned}$$

**Proof of Theorem 4.5.** Applying Theorem 4.3 (c),

$$\text{Var} X_n = \frac{t_0^2}{c^2 x_0^2} \left[ \psi' \left( \frac{n\kappa}{c} \right) - \psi' \left( \frac{n\kappa}{c} + n \right) \right].$$

The asymptotic behaviour of the function  $\psi'(z)$  as  $z \rightarrow \infty$  is given by (A.1), and implies

$$\text{Var} X_n \sim \frac{t_0^2}{c^2 x_0^2} \left( \frac{c}{n\kappa} - \frac{c}{n(\kappa + c)} \right) = \frac{1}{n} \frac{t_0^2}{x_0^2} \frac{1}{\kappa(\kappa + c)}.$$

Next, suppose  $\mathbf{E} T^{-p} < \infty$  for some  $p > 1$ . If  $p > 2$ , then necessarily  $\mathbf{E} T^{-2} < \infty$ , and  $p$  may be replaced with 2 in what follows.

We have

$$\begin{aligned} \mathbf{P}(T_n \leq 0) &= \mathbf{P}(T + X_n - t_0 \leq 0) \\ &= \int_0^{t_0} \mathbf{P}(X_n - t_0 \leq -t) f_T(t) dt \\ &\leq \int_0^{t_0} \mathbf{P}(|X_n - t_0| \geq t) f_T(t) dt \\ &\leq \mathbf{E} |X_n - t_0|^p \int_0^{t_0} \frac{f_T(t)}{t^p} dt \end{aligned}$$

by Markov's inequality. Now, let the  $L^r$ -norm of a random variable  $Y$  be

$$\|Y\|_r = [\mathbf{E}(|Y|^r)]^{\frac{1}{r}}.$$

By applying Hölder's inequality,

$$\mathbb{E}|X_n - t_0|^p = \|X_n - t_0\|_p^p \leq \|X_n - t_0\|_2^p = [(\mathbb{E}(X_n - t_0)^2)]^{\frac{p}{2}},$$

and, moreover,

$$\mathbb{E}(X_n - t_0)^2 = \text{Var}X_n + (\mathbb{E}X_n - t_0)^2.$$

We now need to know the asymptotic behaviour of  $\mathbb{E}X_n - t_0$  as  $n \rightarrow \infty$ . From Theorem 4.3(c) and (A.1),

$$\begin{aligned} \mathbb{E}X_n - t_0 &= \frac{t_0}{cx_0} \left[ \psi\left(\frac{n\kappa}{c} + n\right) - \psi\left(\frac{n\kappa}{c}\right) \right] - t_0 \\ &\sim \frac{t_0}{cx_0} \left[ \log\left(\frac{n\kappa}{c} + n\right) - \log\left(\frac{n\kappa}{c}\right) \right] - \frac{t_0}{2x_0} \left( \frac{1}{n\kappa + cn} - \frac{1}{n\kappa} \right) - t_0 \\ &= \frac{t_0}{cx_0} \log\left(\frac{\kappa + c}{\kappa}\right) + \frac{ct_0}{2nx_0} \frac{1}{\kappa(\kappa + c)} - t_0 \\ &= \frac{ct_0}{2nx_0} \frac{1}{\kappa(\kappa + c)}. \end{aligned}$$

Putting all this together, we get

$$\mathbb{E}|X_n - t_0|^p = \mathcal{O}(n^{-\frac{p}{2}}).$$

This implies that

$$[\sup_t f_T(t)] [\text{Var}X_n]^{\frac{1}{2}} + F_{T_n}(0) \sim [\sup_t f_T(t)] [\text{Var}X_n]^{\frac{1}{2}} \quad \square$$

## References

- Abramowitz, M., and Stegun, I.A. (1972). *Handbook of Mathematical Functions*. (Tenth printing.) Dover, New York.
- Asmussen, S. (1997). Phase-type distributions and related point processes: Fitting and recent advances. In: *Matrix-Analytic Methods in Stochastic Models*, Eds. S.R. Chakravarty and A.S. Alfa, Marcel Dekker Inc., New York.
- Asmussen, S., and Bladt, M., (1997). Renewal theory and queueing algorithms for matrix-exponential distributions. In: *Matrix-Analytic Methods in Stochastic Models*, Eds. S.R. Chakravarty and A.S. Alfa, Marcel Dekker Inc., New York.
- Billingsley, P. (1987). *Probability and Measure*. (Second Edition.) Wiley, New York.
- Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997). *Actuarial Mathematics (Second Edition)*. Society of Actuaries, Itasca, Illinois.
- Dufresne, D. (2001). On a general class of risk models. *Australian Actuarial Journal* **7**: 755-791.
- Dufresne, D. (2004a). Stochastic life annuities. Research Paper no.118, Centre for Actuarial Studies, University of Melbourne, 14 pages. Available at: <http://www.economics.unimelb.edu.au/actwww/wps2004.html>.

- Dufresne, D. (2004b). The log-normal approximation in financial and other computations. *Advances in Applied Probability* **36**: 747-773.
- Dufresne, D. (2005). Bessel Processes and a Functional of Brownian Motion. In: *Numerical Methods in Finance*, H. Ben-Ameur and M. Breton (Eds.), Kluwer Academic Publisher.
- Feldmann, A., and Whitt, W. (1998). Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Performance Evaluation* **31**: 245-279.
- Feller, W. (1971). *An Introduction to Probability Theory and its Applications, II*. (Second Edition.) Wiley, New York.
- Higgins, J.R. (1977). *Completeness and Basis Properties of Sets of Special Functions*. Cambridge University Press.
- Lebedev, N.N. (1972). *Special Functions and their Applications*. Dover, New York.
- Luke, Y.L. (1969). *The Special Functions and their Approximations*. Academic Press, New York.
- Neuts, M.F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. The Johns Hopkins University Press, Baltimore, 1981.
- de Prony, Baron Gaspard Riche (1795). Essai experimental et analytique: sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures. *Journal de l'École Polytechnique* **1**, cahier 22: 24-76. (Available at [www.essi.fr/~leroux/PRONY.pdf](http://www.essi.fr/~leroux/PRONY.pdf).)
- Yor, M. (1992). On some exponential functionals of Brownian motion. *Adv. Appl. Prob.* **24**: 509-531. (Reproduced in Yor (2001).)

Daniel Dufresne  
Centre for Actuarial Studies  
Department of Economics  
University of Melbourne  
Tel.: (0)3 8344 5324  
Fax: (0)3 8344 6899  
dufresne@unimelb.edu.au

Figure 1. Example 5.1, degenerate distribution

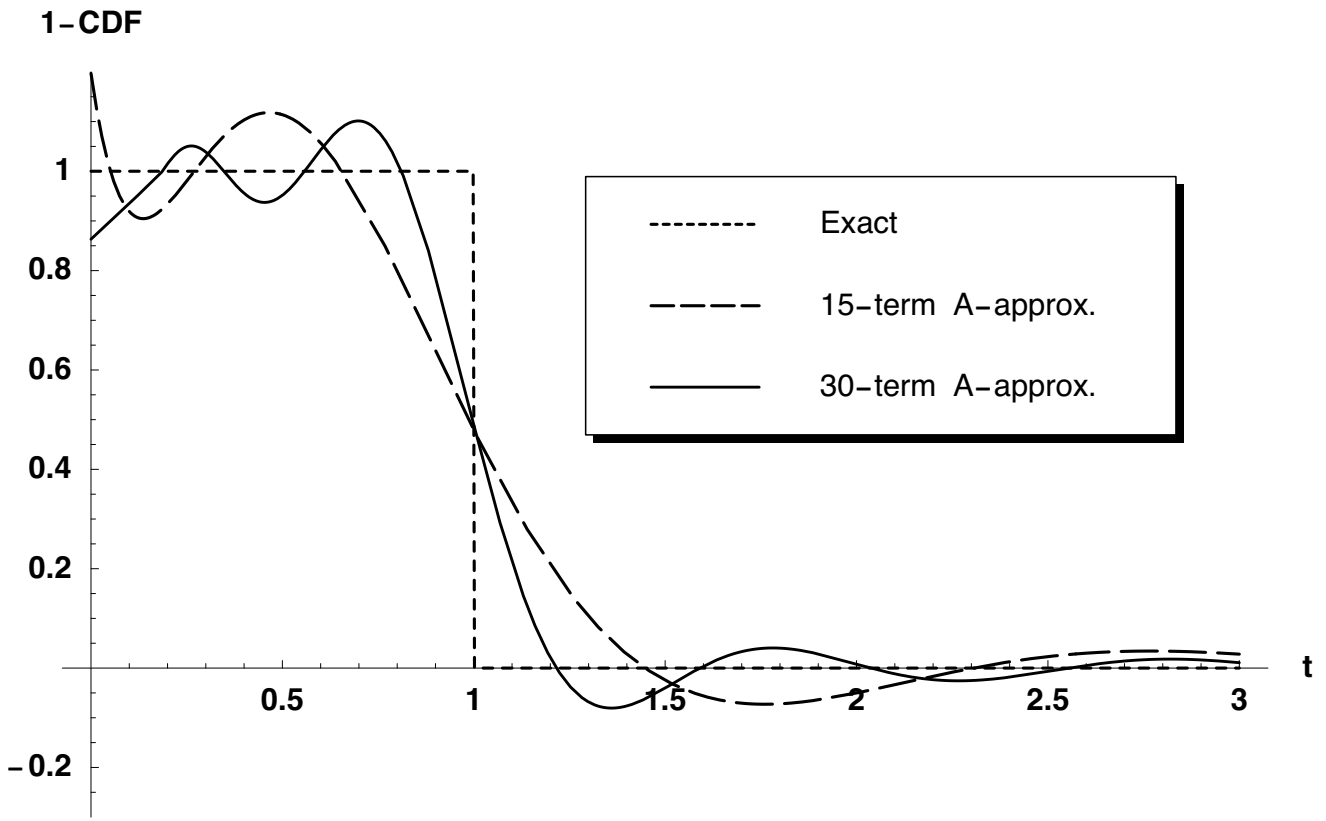


Figure 2. Example 5.2, uniform distribution

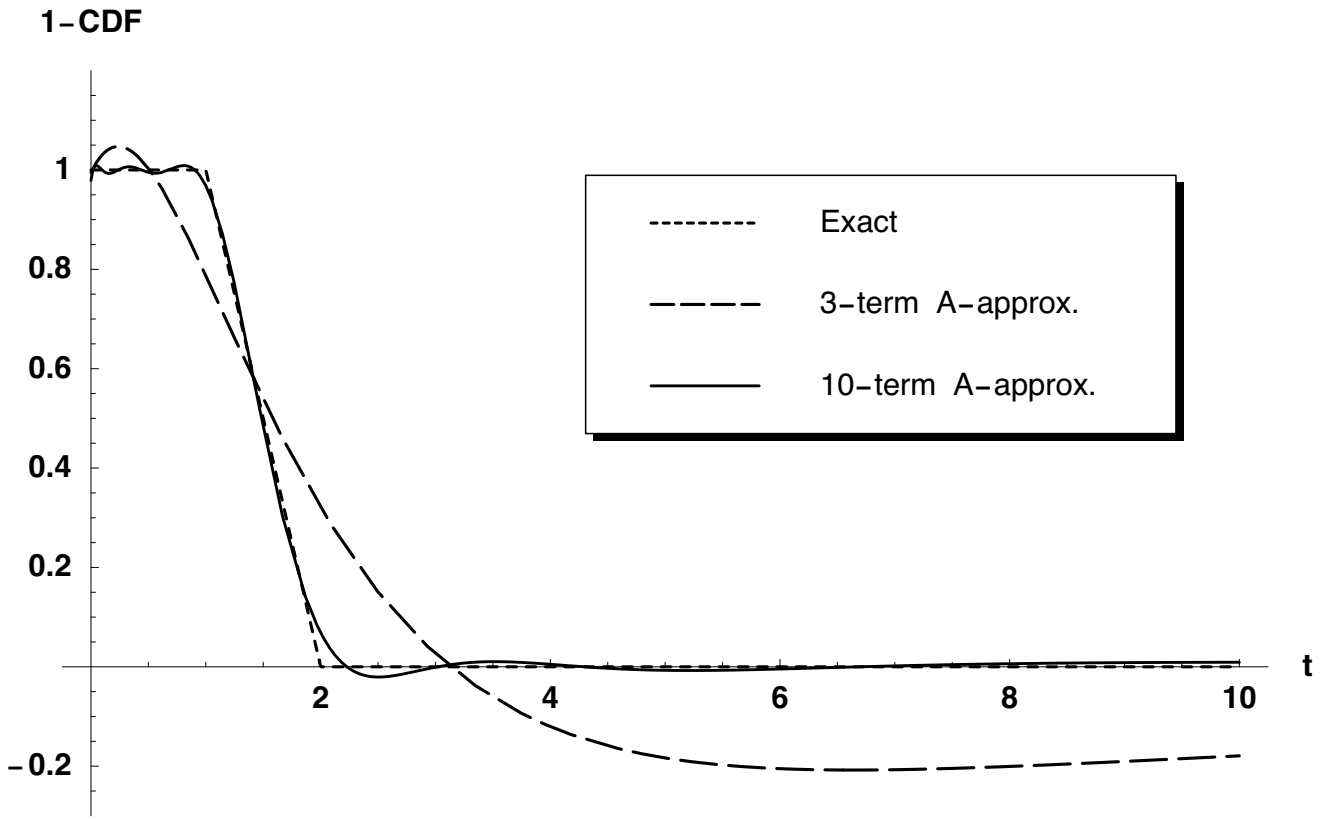


Figure 3. Example 5.2, uniform distribution

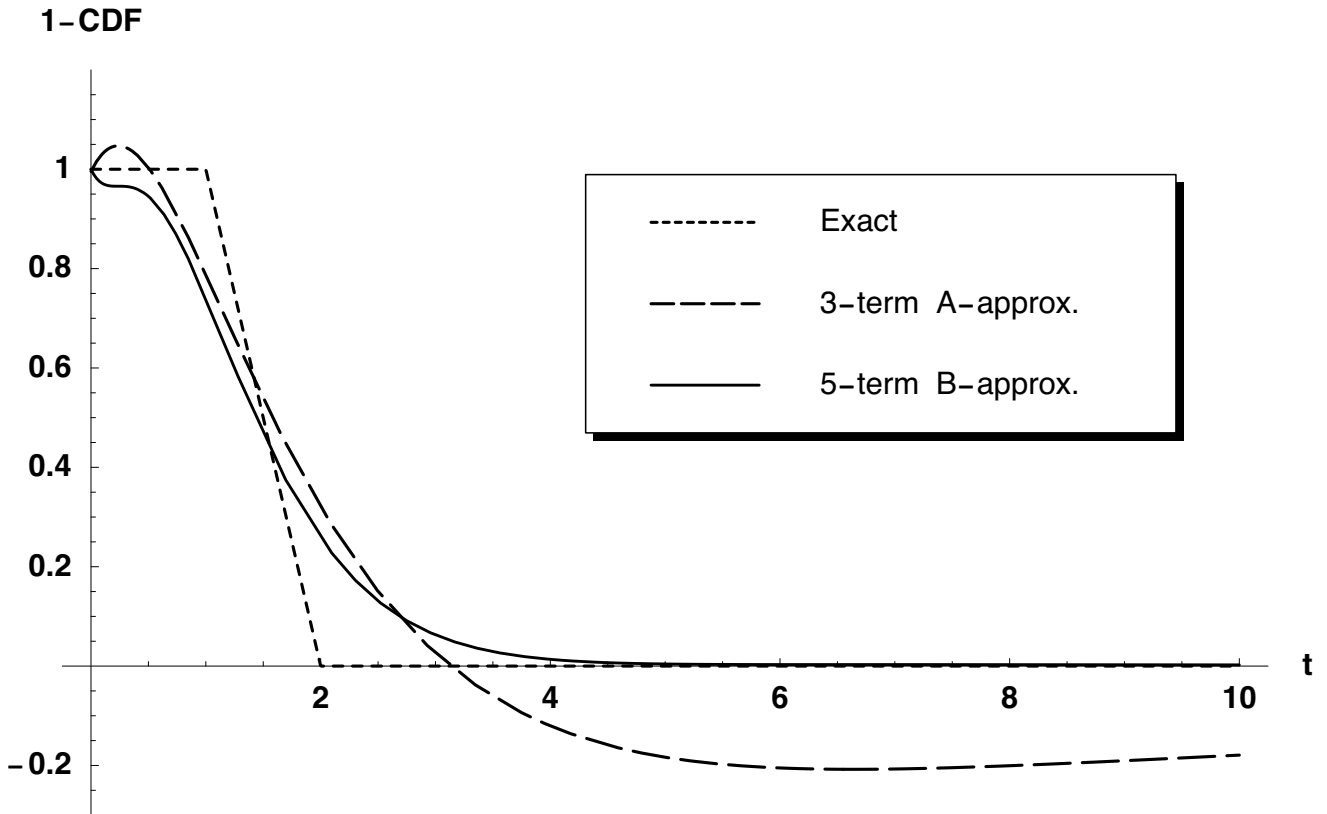


Figure 4. Example 5.3, Pareto distribution

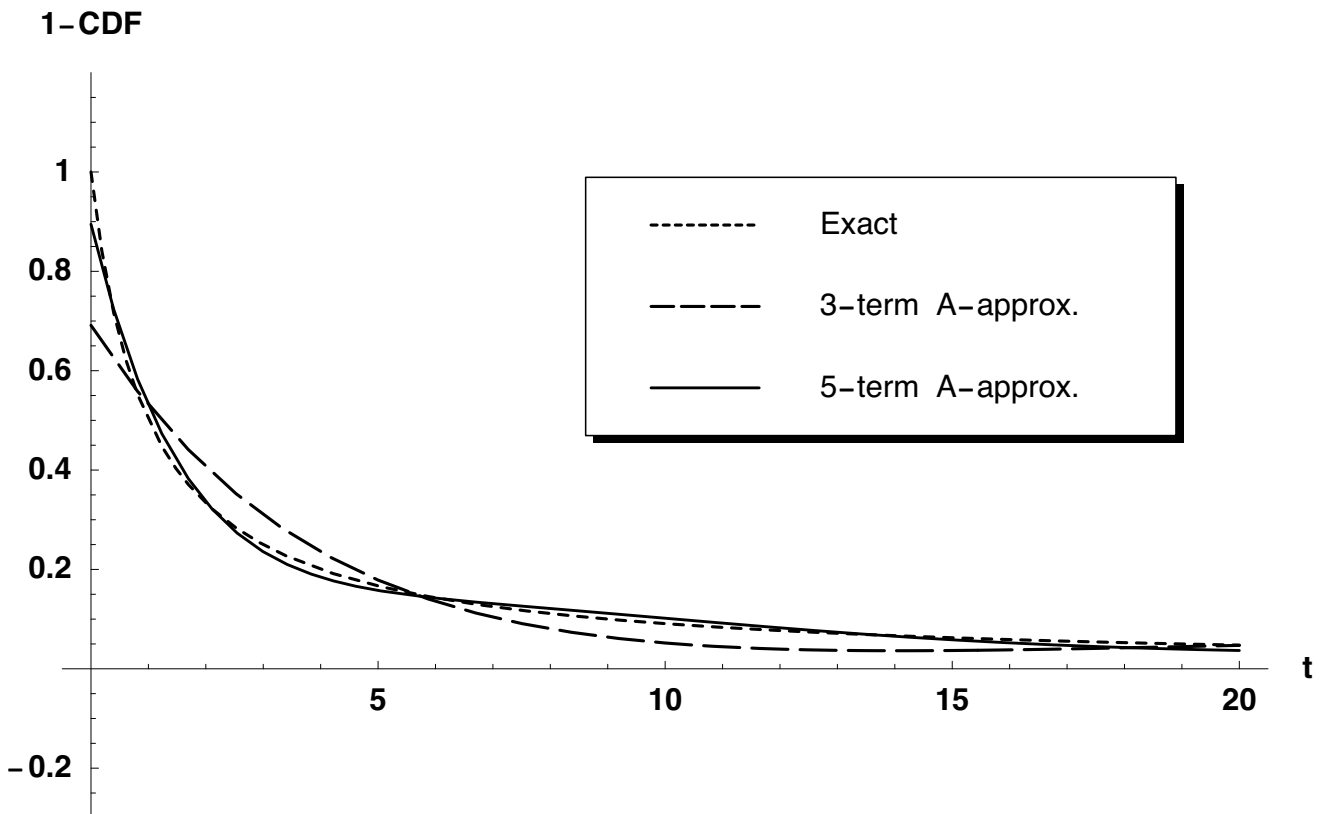


Figure 5. Example 5.3, Pareto distribution

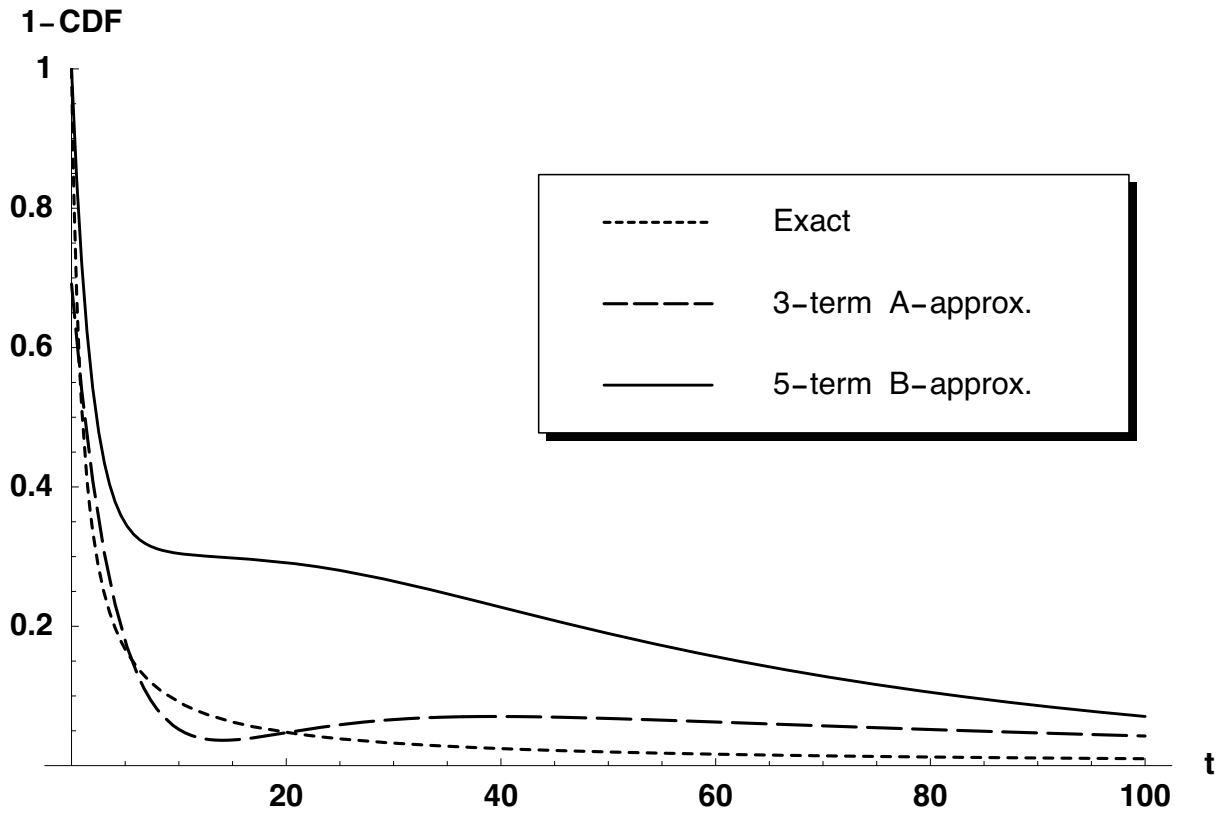


Figure 6. Example 5.4, lognormal distribution

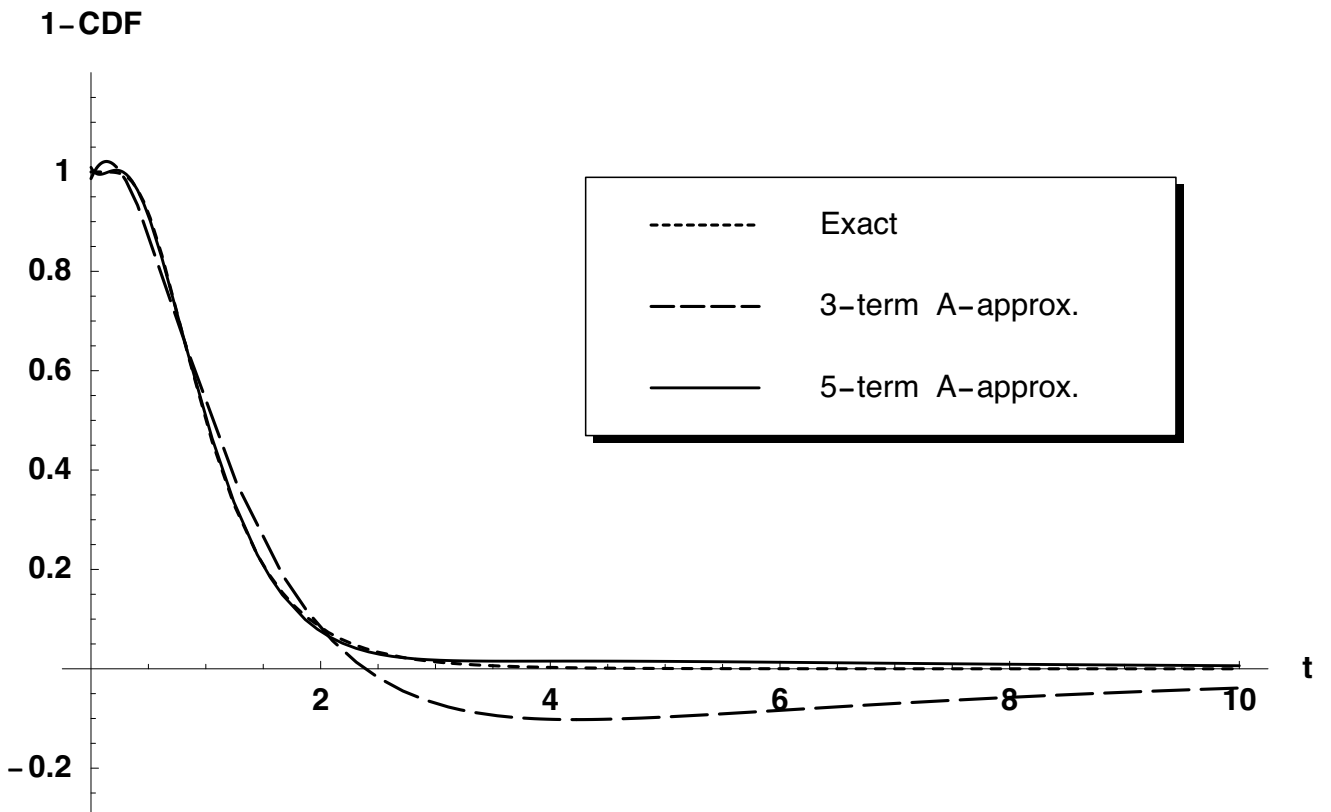




Figure 7. Example 5.4, lognormal distribution

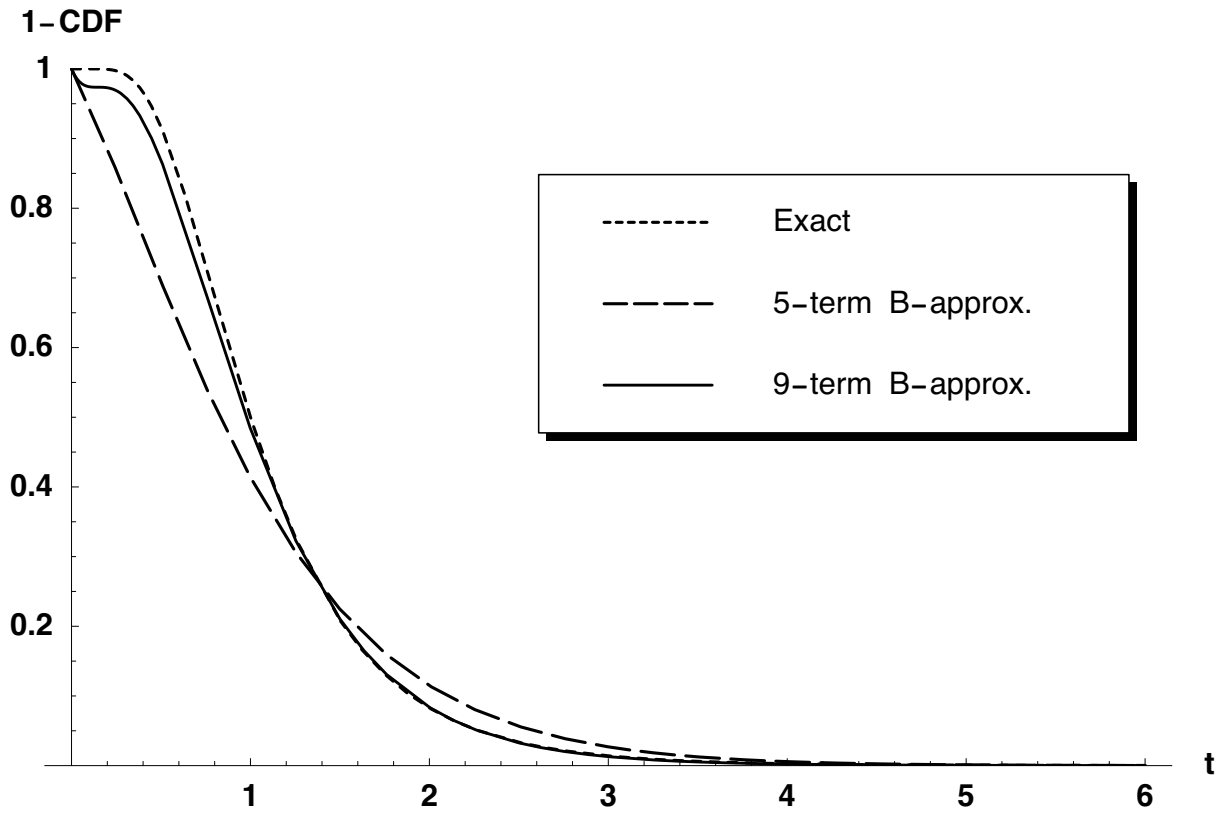


Figure 8. Example 5.5, Makeham distribution

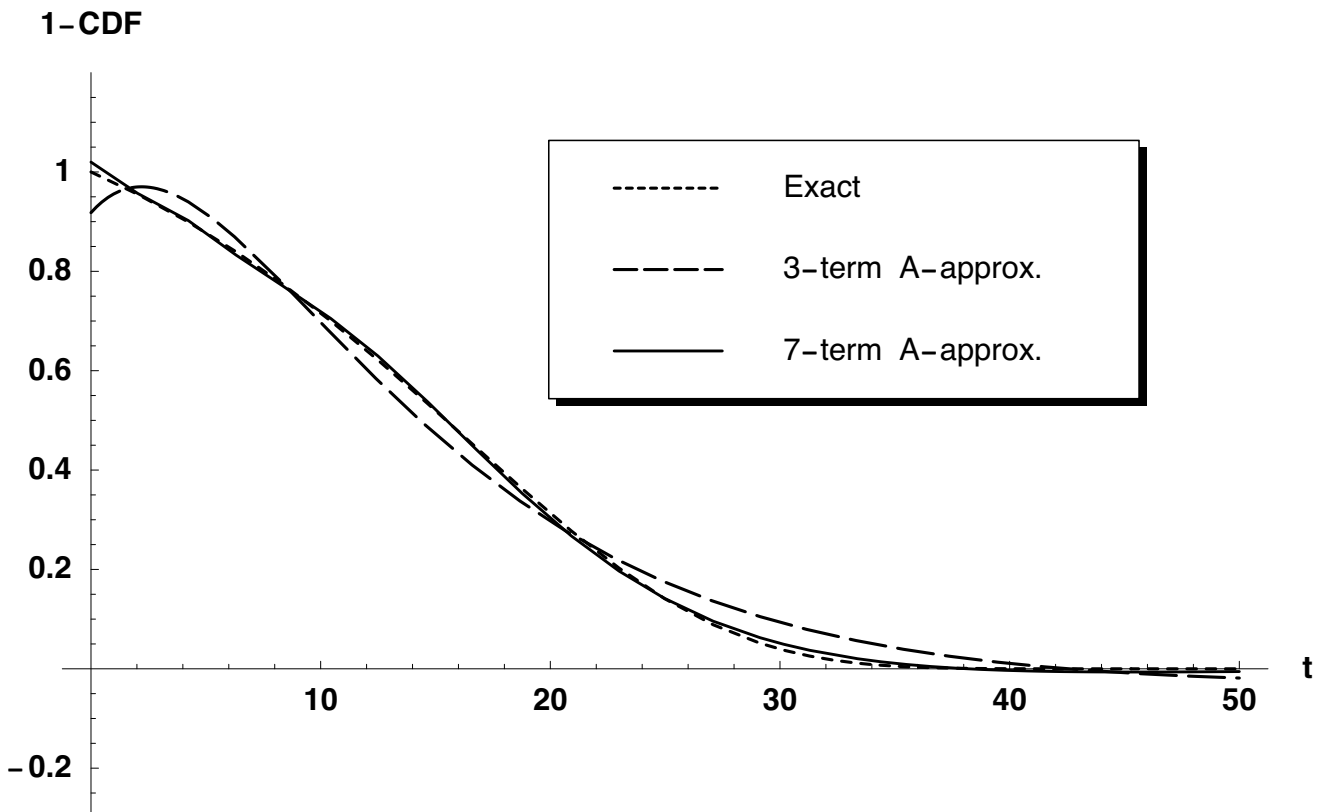


Figure 9. Example 6.1, degenerate distribution

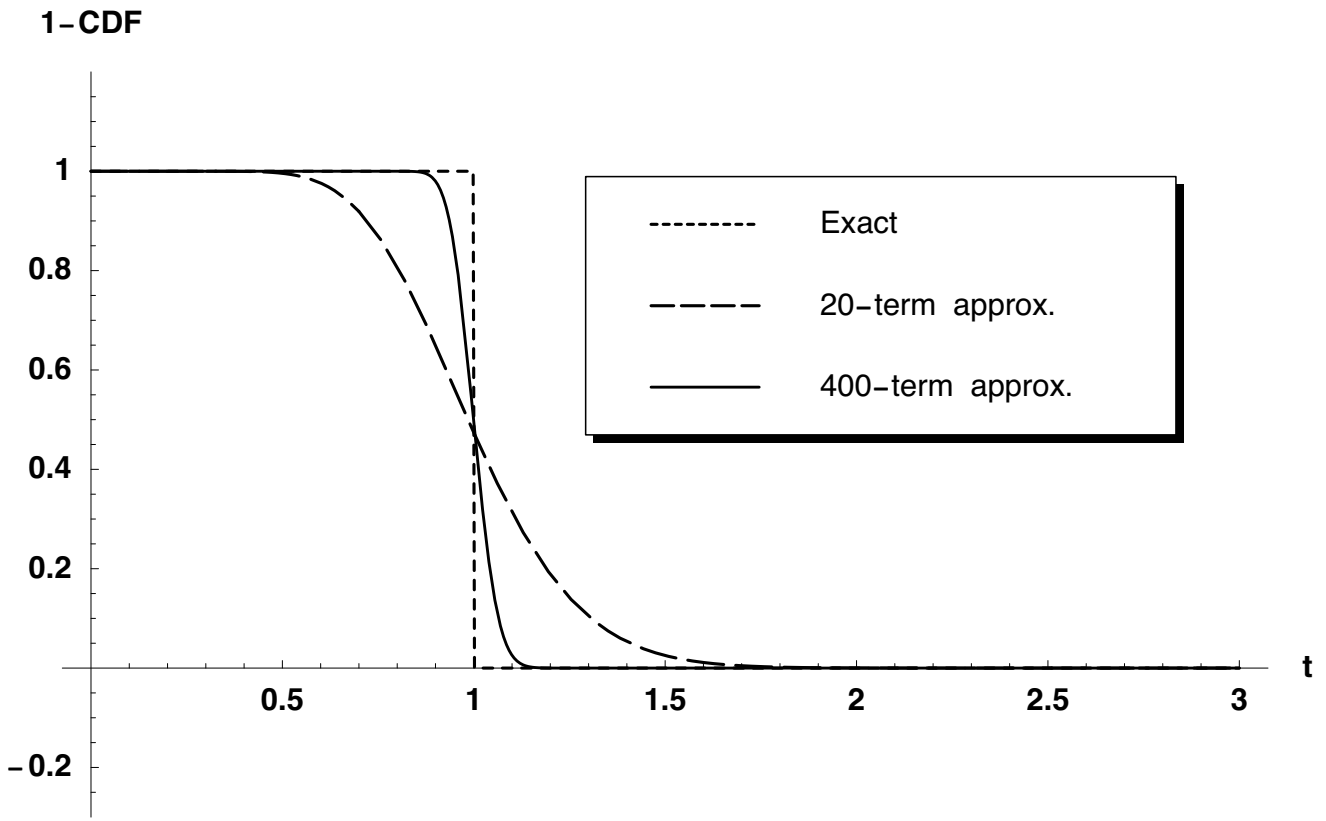


Figure 10. Example 6.2, uniform distribution

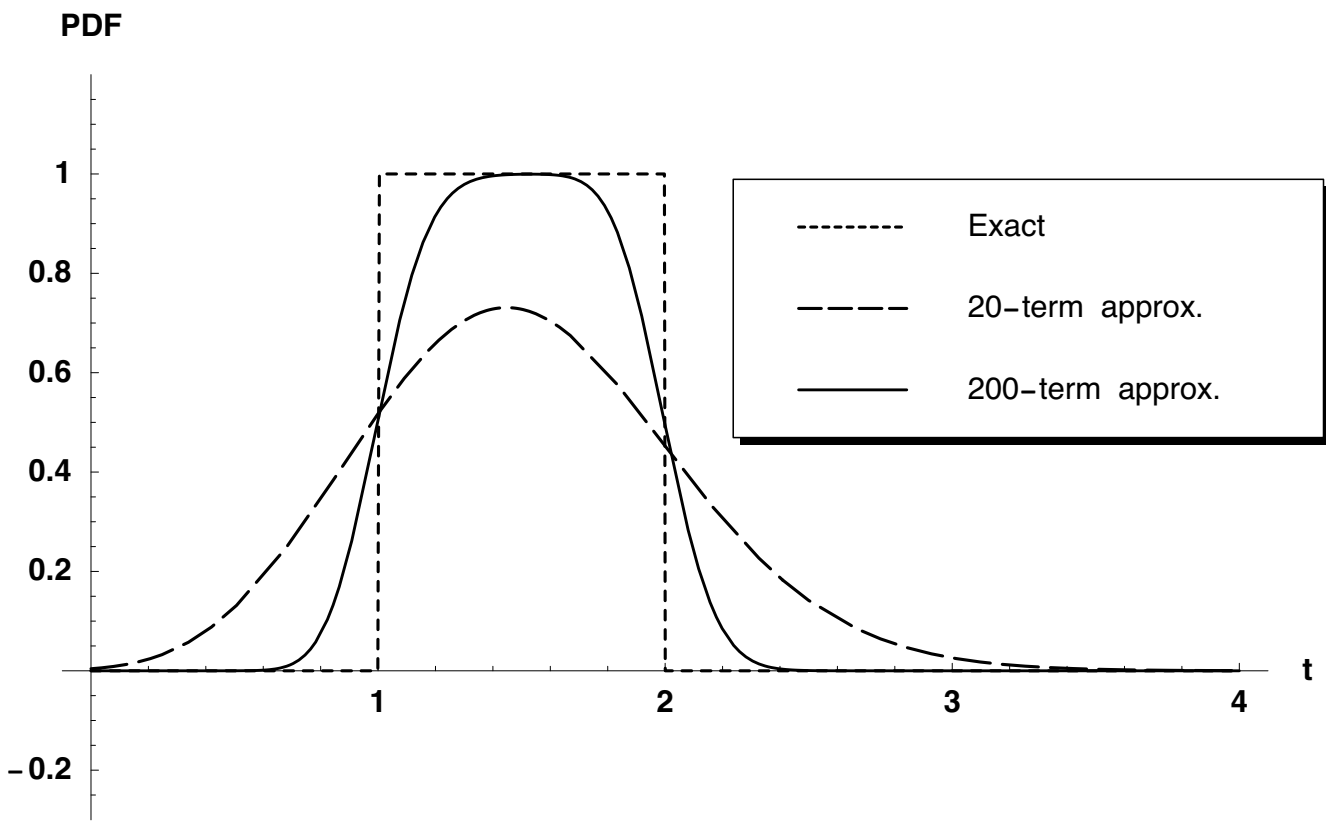


Figure 11. Example 6.2, uniform distribution

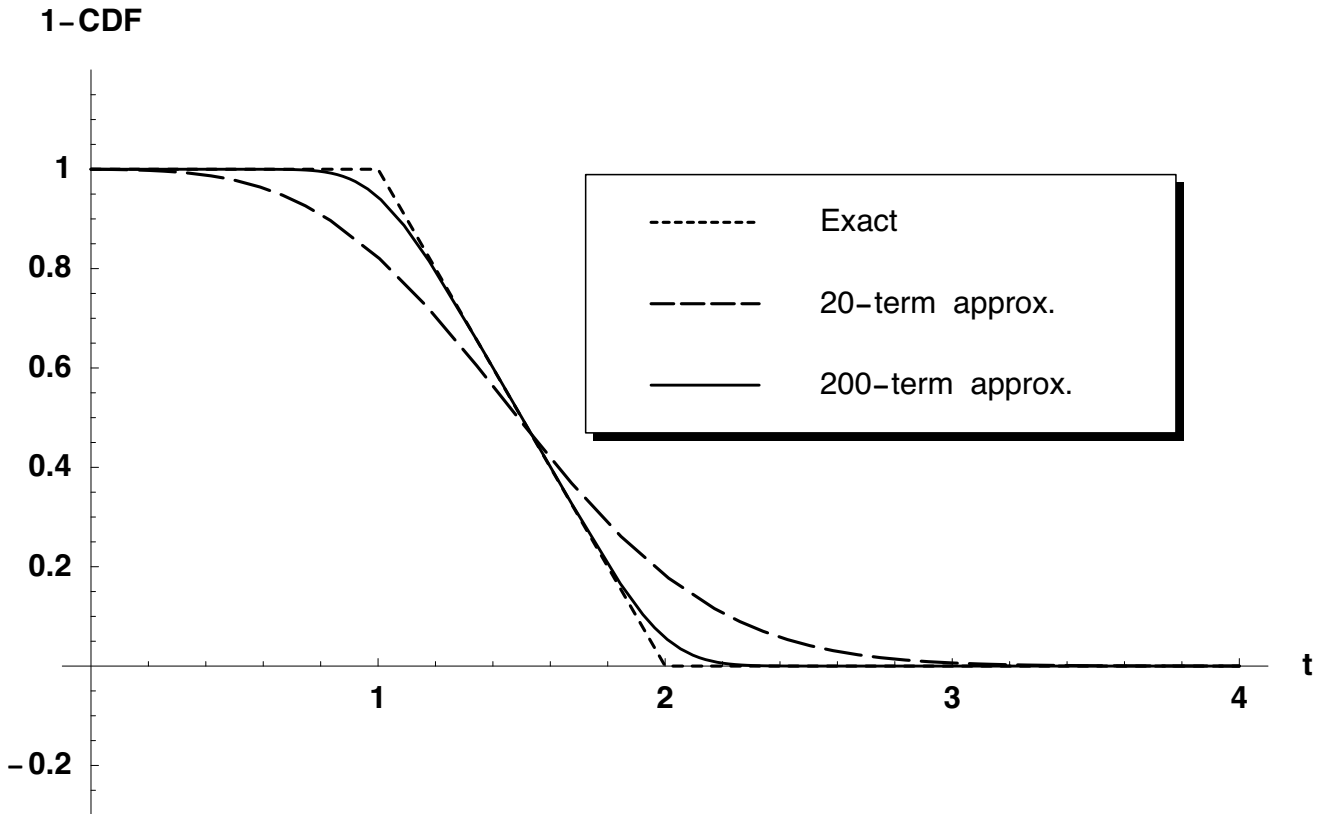


Figure 12. Example 6.3, squared sine distribution

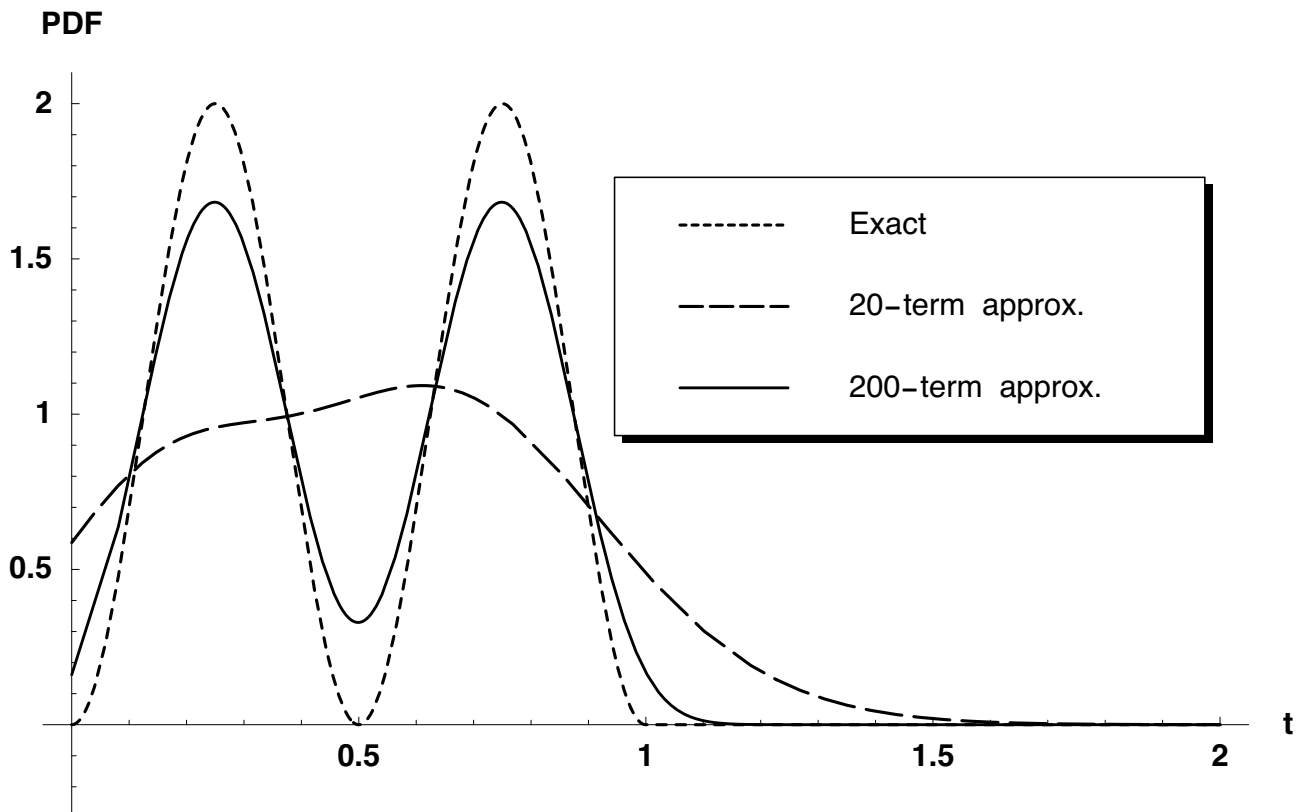


Figure 13. Example 6.3, squared sine distribution

