FITTING COMBINATIONS OF EXPONENTIALS TO PROBABILITY DISTRIBUTIONS

Daniel Dufresne
centre for Actuarial Studies
University of Melbourne

First version, 3 February 2005

Abstract

Techniques are described for approximating distributions on the positive half-line by combinations of exponentials, based on Jacobi polynomial expansions and on the log-beta distribution. The techniques are applied to some well-known distributions (degenerate, uniform, Pareto, lognormal, Makeham). 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.

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 is those with a distribution function F which can be written as

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

When $a_j > 0$, $j = 1, \ldots, n$, 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), 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 exponantials (ce's for short in the sequel) 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 is far form 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 developed the techniques described in this paper. So far the author has not located these techniques in the literature. Other methods rely on least-squares, moment matching, and so on. The advantage of the Jacobi 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 the phase-type distributions. However, this author has come across ce approximations 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 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 ce, and then proceed with the convolution, which is a relatively straightforward affair with ce's. This idea is explored briefly in Section 6.

Stochastic life annuities. The distribution of a stochastic annuity with an exponential lifetime is known in closed form; the mathematical result was first derived by Yor (1992), but see Dufresne (2004a) for the financial interpretation. The practically interesting case, however, is where the annuity for the duration of human life, which does not have an exponential distribution. Expressing the future lifetime distribution as an approximate ce then yields a closed form approximation for the distribution of the annuity value. The methods presented below will be applied to this problem in a separate 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 ce 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 ce's which are not precisely density functions, while method B produces true density functions. Method A works in theory for discrete distributions, but the numerical results are not very good, so a different idea is presented in Section 4. It is shown that the log-beta distribution converges to the Dirac mass, and gives a different sequence of ce's which converge to any discrete distribution with a finite number of atoms.

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

Notation and vocabulary. "Combination of exponentials is abbreviated as "ce." 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 $\delta_x(\cdot)$. Finally, $(z)_n$ is Pochhammer's symbol:

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

(*N.B.* Knuth (1992) claims that Pochhammer did not use this symbol to mean the above ascending factorial, but rather the binomial coefficient $\binom{z}{n}$.)

2. Jacobi polynomials

The Jacobi polynomials are usually defined as

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

for $\alpha, \beta > -1$. These polynomials are othogonal 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 $\alpha, \beta > -1$:

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

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, if $\alpha, \beta > -1$, the Jacobi series*

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

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

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 $N \to \infty$, 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 f 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

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

3. Fitting combinations of exponentials to continuous distribution functions

Method A: direct Jacobi approximation of the ccdf

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

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

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:

$$g(x) \;=\; \overline{F}\Big(-\frac{1}{r}\log(x)\Big), \quad 0 < x \leq 1, \quad g(0) \;=\; 0,$$

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

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

then

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

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

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

This is a combination of terms of the form

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

The advantage of 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 $\overline{F}(\cdot)$, which may be expressed in terms of the Laplace transform of the distribution of T:

$$\int_0^\infty e^{-\mu t}\overline{F}(t)\, dt = -\frac{1}{\mu}\int_0^\infty \overline{F}(t)\, de^{-\mu t} = \frac{1}{\mu}[1 - \mathsf{E}e^{-\mu T}], \quad \mu > 0. \tag{3.4}$$

The next result is an immediate consequence of Theorem 2.1.

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

$$e^{prt}\overline{F}(t) \tag{3.5}$$

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

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

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 condition (3.5) 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 there for some $p \in \mathbb{R}$ and $r > 0$*

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

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

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

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 $\overline{F}(t)$ tends to 0 as $t \to \infty$).

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

(ii) $\mathsf{E}\,(T \wedge 1)^{\alpha} \;<\; \infty$. *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

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

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 give a become density function.

Examples are given in Section 5. Observe that this procedure is not guaranteed to converge to the true function f as the number of terms used tends to infinity, because the integral to the squared truncated series may not converge to one. For instance, if one uses some $p < 0$, then the integral of the squared truncated series is infinite. 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

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

Only in the case where $\beta+1-2p=\alpha=0$ does this mean that the integral of ϕ_N tends to 1. In other cases this may or may not be true. Nevertheless, the method may still yield good (or bad) approximations.

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 to distributions with atoms; the log-beta distribution

Theorem 3.1 applies to discontinuous distributions as well, the only difference is that at points of discontinuity the series converges to

$$\frac{1}{2}[\overline{F}(x-)+\overline{F}(x+)].$$

This is not a big problem, as in practice one would use the truncated series, which is a continuous function. Numerically, however, the approximations obtained are not very good, as they oscillate a lot and are not very accurate. An alternative is given here, which is more direct, yields a true probability distribution and may be superior numerically.

Every probability distribution can be decomposed into its continuous and discrete parts, say

$$F \;=\; qF_c+(1-q)F_d.$$

The continuous part F_c may be approximated by combinations of exponentials as in the previous section. The discrete part consists of all the atoms of the distribution, say

$$dF_d(x) \;=\; \sum_j q_j\delta_{x_j}(x),$$

where δ_{x_j} is the Dirac mass at x_j. It is possible to find a combination of exponentials that approximates dF_d by adding up the combinations of exponentials that approximate each individual atom. We thus consider approximating a single Dirac mass at some point $y_0>0$.

Since

$$\lim_{c\to 0}\left(\frac{1-e^{-cx}}{c}\right)^{n-1} \;=\; x^{n-1}$$

for all x, the densities

$$Ke^{-nx}(1-e^{-cx})^{n-1}1_{\{x>0\}} \tag{4.1}$$

tend to the **Gamma**(n, n) distribution as c tends to 0 (K is a constant). The **Gamma**(n, n) distributions converge to the Dirac measure as n tends to infinity. Hence the densities in (4.1) converge (in distribution) to the Dirac measure at 1 as $c \to 0$ and $n \to \infty$.

In some circumstances, however, it may not be appropriate to use very small values of c. For this reason, we give the definition (which is not new) of the log-beta family of distributions, which includes (4.1) as a special case. It turns out that it is not required to take c very small to get accurate approximations of the Dirac measure using combinations of exponentials.

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

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

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

$$\textbf{LogBeta}(1,b,c) \;=\; \textbf{Exp}(b), \qquad \lim_{c\to 0+} \textbf{LogBeta}(a,b,c) \;=\; \textbf{Gamma}(a,b).$$

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

$$\psi(z) \;=\; -\gamma + \sum_{k=0}^{\infty}\left(\frac{1}{k+1} - \frac{1}{k+z}\right) \tag{4.2}$$

where γ is Euler's constant. The derivatives $\psi'(\cdot), \psi''(\cdot), \psi^{(3)}(\cdot), \ldots$ of the digamma function are known as the polygamma functions, and can be obtained by differentiating (4.2) term by term.

Theorem 4.1. (***Properties of the log-beta distribution***) *Suppose $X \sim$ **LogBeta**(a, b, c), $a, b, c > 0$.*

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

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

(c) The cumulants of the distribution are

$$\kappa_n \;=\; \frac{(-1)^{n-1}}{c^n}\left[\psi^{(n-1)}\left(\frac{b}{c}+a\right)-\psi^{(n-1)}\left(\frac{b}{c}\right)\right].$$

In particular,

$$\mathsf{E}\,X \;=\; \frac{1}{c}\left[\psi\left(\frac{b}{c}+a\right)-\psi\left(\frac{b}{c}\right)\right];\quad \mathsf{Var}\,X \;=\; \frac{1}{c^2}\left[\psi'\left(\frac{b}{c}\right)-\psi'\left(\frac{b}{c}+a\right)\right]$$

$$\mathsf{E}(X-\mathsf{E}\,X)^3 \;=\; \frac{1}{c^3}\left[\psi''\left(\frac{b}{c}+a\right)-\psi''\left(\frac{b}{c}\right)\right].$$

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

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

Moreover,

$$\lim_{a\to\infty}\mathsf{E}\,X^k \;=\; x_0^k,\quad k=1,2,\ldots$$

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

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

where

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

are independent, and

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

(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

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

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

$$\frac{X-\mathsf{E}X}{\sqrt{\mathsf{Var}\,X}} \xrightarrow{\mathrm{d}} N(0,1) \quad \textit{as } a\to\infty.$$

*(h) Suppose $X_1\sim$ **LogBeta**(a_1,b,c) and $X_2\sim$ **LogBeta**$(a_2,b+a_1c,c)$ are independent. Then $X_1+X_2\sim$ **LogBeta**(a_1+a_2,b,c)*

Proof. Parts (a) and (b) follow from

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

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 exlicit 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)

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

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

$$\lim_{a\to\infty}\mathsf{E}\,e^{-sX} \;=\; \lim_{a\to\infty}\frac{\Gamma(\frac{\kappa a}{c}+a)\Gamma(\frac{\kappa a+s}{c})}{\Gamma(\frac{\kappa a+s}{c}+a)\Gamma(\frac{\kappa a}{c})} \;=\; \left(\frac{\kappa}{\kappa+c}\right)^{-\frac{s}{c}}$$

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):

$$\begin{aligned}\psi(z) &\sim \log(z)-\frac{1}{2z}-\frac{1}{12z^2}+\ldots\\ \psi^{(n)}(z) &\sim (-1)^{(n-1)}\left[\frac{(n-1)!}{z^n}+\frac{n!}{2z^{n+1}}+\ldots\right]\end{aligned}$$

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

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

$$\mathsf{E}e^{-sX} \;=\; \frac{\left(\frac{b}{c}\right)_n}{\left(\frac{b+s}{c}\right)_n} \;=\; \frac{b}{b+s}\,\frac{b+c}{b+c+s}\cdots\frac{b+(n-1)c}{b+(n-1)c+s}.$$

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 (4.2) and (c) above.

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

To prove (g), first observe that, using (b),

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

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

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

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

$$\lim_{n\to\infty} \frac{1}{S_n^{2+\delta}} \sum_{j=0}^{n-1} \mathsf{E}|Y_{n,j} - \mathsf{E}Y_{n,j}|^{2+\delta} \;=\; 0 \tag{4.3}$$

for some $\delta > 0$, with

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

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

$$\mathsf{Var}\, Y \;=\; 1, \qquad \mathsf{E}(Y-1)^4 \;=\; 9.$$

This means that

$$\mathsf{Var}\, Y_{n,j} \;=\; \frac{1}{(n+jc)^2}, \qquad \mathsf{E}(Y_{n,j} - \mathsf{E}Y_{n,j})^4 \;=\; \frac{9}{(n+jc)^4}.$$

Then

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

which implies (4.3). For arbitrary a, let $n = \lfloor a \rfloor$, $r = a - n$, and note that

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

Hence,

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

where $Y_{a,j} \sim \textbf{Exp}(b + (r+j)c)$, $j = 0, \ldots, n-1$, and $Y_{a,n} \sim \textbf{LogBeta}(r, a, c)$. The Lyapounov central Limit Theorem can again be applied, the only real difference is the additional term $Y_{n,n}$. This term does not change the limit, since it can be observed (see (4.2) and (c)) that

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

This implies that

$$\lim_{a\to\infty} \frac{\mathsf{Var} Y_{a,0} + \cdots + Y_{a,n-1}}{\mathsf{Var} Y_{a,0} + \cdots + Y_{a,n}} \;=\; 1, \qquad \frac{Y_{a,n} - \mathsf{E}\, Y_{a,n}}{\sqrt{\mathsf{Var} Y_{a,0} + \cdots + Y_{a,n}}} \stackrel{\mathrm{d}}{\to} 0 \quad \text{as} \quad a \to \infty.$$

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 by performing the convolution explicitly: for $y > 0$,

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

The next two corollaries may not be as well known as the previous theorem. The first one is an immediate consequence of parts (d) and (f) of Theorem 4.1.

Corollary 4.2. *For $n \in \mathbb{N}_+$ and any fixed $\kappa, c, y_0 > 0$, the* **LogBeta**$(n, n\kappa x_0/y_0, cx_0/y_0)$ *distribution has a pdf which is a ce, and it tends to δ_{y_0} as $n \to \infty$, if $x_0 = \log(1 + c/\kappa)/c$. The same holds if x_0 is replaced with the mean of the* **LogBeta**$(n, n\kappa, c)$ *distribution.*

The classical result that ce's are dense in the set of distributions may thus be proved as follows: approximate any distribution by a discrete distribution with a finite number of atoms; then approximate each atom by a log-beta distribution as in Corollary 4.2.

From the above results, one can define a new "integrated log-beta" distribution which converges to the uniform distribution. Let $0 < y_1 < y_2$ and, for $i = 1, 2$, choose $\kappa_i, c_i > 0$, and also define $x_i = \log(1 + c_i/\kappa_i)/c_i$. We know that, for $i = 1, 2$, the distribution with density

$$\sum_{j=0}^{n-1} \alpha_j^{(i)}\lambda_j^{(i)}e^{-\lambda_j^{(i)}x}1_{\{x>0\}},$$

where

$$\alpha_j^{(i)} = c_i\left(\frac{n\kappa_i}{c_i}\right)_{n_i}\frac{(-1)^j}{j!(n-1-j)!(n\kappa_i+jc_i)}, \quad \lambda_j = \frac{x_i}{y_i}(n\kappa_i+jc_i), \quad j = 0, \ldots, n-1,$$

converges to the the Dirac mass at y_i. This implies that, for $i = 1, 2$, the corresponding ccdf converges to the indicator function $1_{\{t<y_i\}}$ at all points t except $t = y_i$. In each case, the integral of the ccdf over $[0, \infty)$ equals the mean of the distribution, namely

$$\mu_i = \frac{1}{c_i}\left[\psi\left(\frac{n\kappa_i}{c_i}+n\right)-\psi\left(\frac{n\kappa_i}{c_i}\right)\right] = \frac{y_i}{x_i}\sum_{j=0}^{n-1}\frac{1}{n\kappa_i+jc_i}, \quad i = 1, 2.$$

Hence, the ce defined as the scaled difference of the ccdf's,

$$\frac{1}{\mu_2-\mu_1}\sum_{j=0}^{n-1}(\alpha_j^{(2)}\lambda_j^{(2)}e^{-\lambda_j^{(2)}x}-\alpha_j^{(1)}\lambda_j^{(1)}e^{-\lambda_j^{(1)}x})1_{\{x>0\}} \tag{4.4}$$

is non-negative, integrates to one, and converges to the density of the $\mathbf{U}(0,1)$ distribution.

Corollary 4.3. *The ce in (4.4) converges as $n \to \infty$ to the $\mathbf{U}(y_1, y_2)$ pdf*

$$\frac{1}{y_2 - y_1} 1_{(y_1,y_2)}(t)$$

for every $t \neq y_1, y_2$. Convergence also holds in the sense of distributions, and all moments converge to those of the limit. If, in the above formulas, x_i (for $i = 1, 2$) is replaced with the mean of the **LogBeta**$(n, n\kappa_i, c_i)$ *distribution, then the same result holds if in (4.4) μ_i is replaced with y_i. To get a sequence of ce's which converges to a $\mathbf{U}(0, y_2)$ distribution, use the following as pdf:*

$$\frac{1}{\mu_2} \sum_{j=0}^{n-1} \alpha_j^{(2)} \lambda_j^{(2)} e^{-\lambda_j^{(2)} x} 1_{\{x>0\}}.$$

By successive integrations of the log-beta ccdf's one obtains sequences of ce's which converge to the **Beta**$(1, m)$ distribution on any interval $[y_1, y_2]$ ($0 \leq y_1 < y_2$), for $m = 2, 3, \ldots$ The details are omitted.

5. Examples of 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 α, β, 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 Dirac mass

One does not expect great accuracy here, but there may be practical circumstances where combinations of exponentials are sought to approximate discrete distributions. In this case the Jacobi polynomials lead to highly oscillatory approximations.

Figure 1 shows two log-beta approximations for a Dirac mass at $x = 1$. The ce's used are the **LogBeta**$(n, b, .5)$ distribution with $n = 6$ and 20, b being adjusted so that the mean of the distribution is equal to 1 in each case. However, the results are apparently not very sensitive to the parameter c.

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 a $\mathbf{U}(1,2)$ 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

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

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 $\overline{F}_3$, squaring and rescaling yields

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

There are now $2 \times 3 - 1 = 5$ distinct frequencies λ_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. For comparison, the "integrated difference of log-betas" result of Corollary 4.3 is also presented (see Figure 4). The ccdf of the distribution in that corollary is shown, with $n = 15$, which is in effect a 30-term ce. The maximum difference between the exact and approximate cdf of .208 is not very good, but the cdf has no unwanted oscillations. The pdf's are also shown in Figure 5.

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. The Pareto distribution chosen has ccdf

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

Figure 6 compares the 3 and 5-term A approximations for the ccdf, while Figure 7 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 α, β, 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 8 and 9). 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 8 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**(μ, σ^2) distribution is a constant times the pdf of a **Lognormal**$(\mu + \sigma^2, 2\sigma^2)$:

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

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 ccdf 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 9 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 10 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 intuitive explanation may be that the tail of the Makeham is much thinner than that of the exponential.

6. 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. The reality is that in applications these cases occur relatively rarely. 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 logormal distributions with first parameter 0 and second parameter σ^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 colum shows the maximum difference between the distribution functions. It is interesting to note that under Method A the error for the convolution is 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 ccdf of the distribution (see (3.3)). The latter is given by (3.4). Hence, the Jacobi expansion for the ccdf of the convolution of probability distributions may be found directly form the Laplace transforms of each of the indvidual distributions, since

$$\mathsf{E}\, e^{-\mu(T_1+\cdots+T_m)} \;=\; \prod_{k=1}^{m} \mathsf{E}\, e^{-\mu T_k}.$$

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

Table 1

Accuracy of convolutions of lognormals

σ^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 problem which arises with these methods is the magnitude of the Fourier coefficents 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 combinations of exponentials converge to the Dirac mass and also, by integrating the ccdf, to the uniform distribution; by iterating this procedure one obtains a sequence of ce's which convergence to the **Beta**$(1, m)$ distribution, $m = 2, 3 \ldots$ In the case of the Dirac mass and uniform, a good number of exponentials is required for modest accuracy, but the approximate pdf's and ccdf's so obtained are free of the oscillations that afflict the Jacobi approximations.

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. Chakravarthy 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. Chakravarthy 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.116, 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.

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 Funcions*. Cambridge University Press.

Knuth, D. (1992). Two notes on notation. *Amer. Math. Monthly* **99**: 403-422. Available at http:// www-cs-faculty.stanford.edu/∼ knuth/preprints.html.

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, Baltimaore, 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.)

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, Dirac mass at x=1

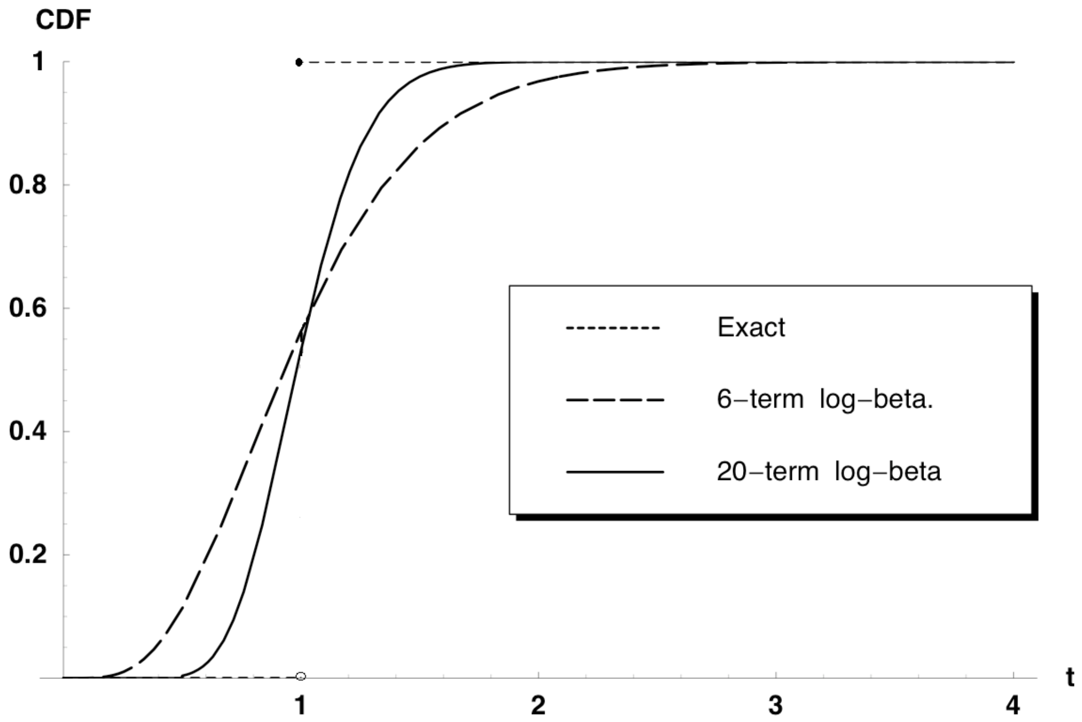

Figure 2. Example 5.2, Uniform(1,2) distribution

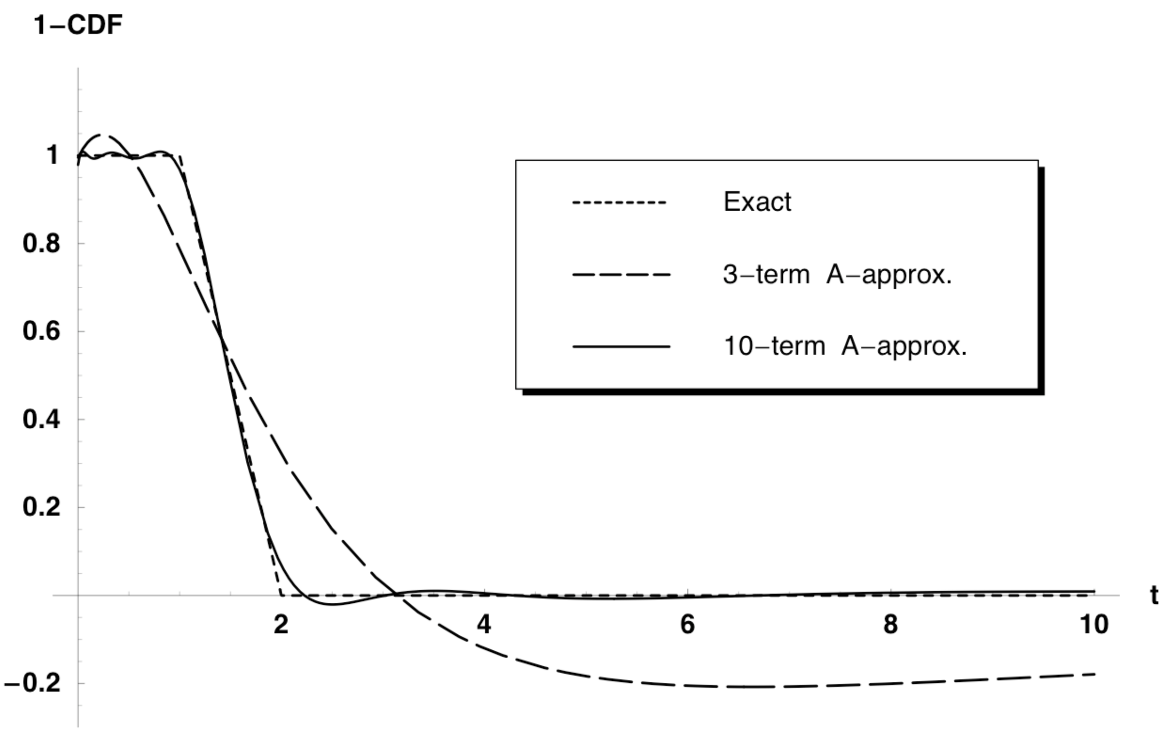

Figure 3. Example 5.2, Uniform(1,2) distribution

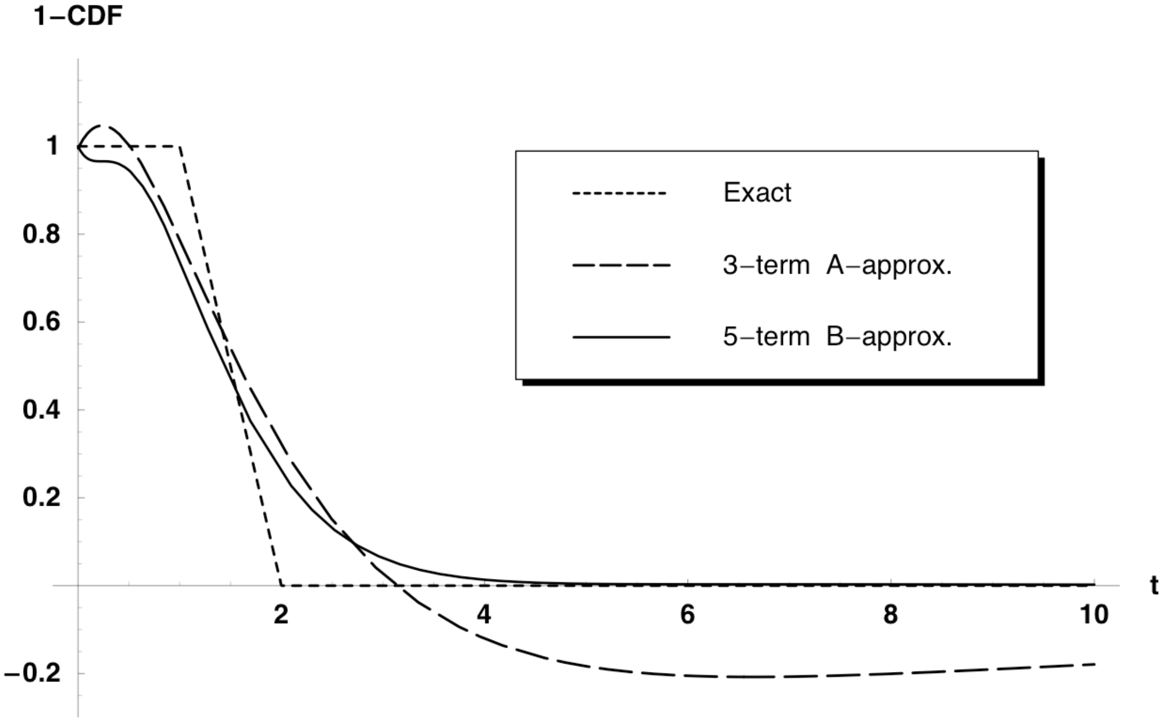

Figure 4. Example 5.2, Uniform(1,2) distribution

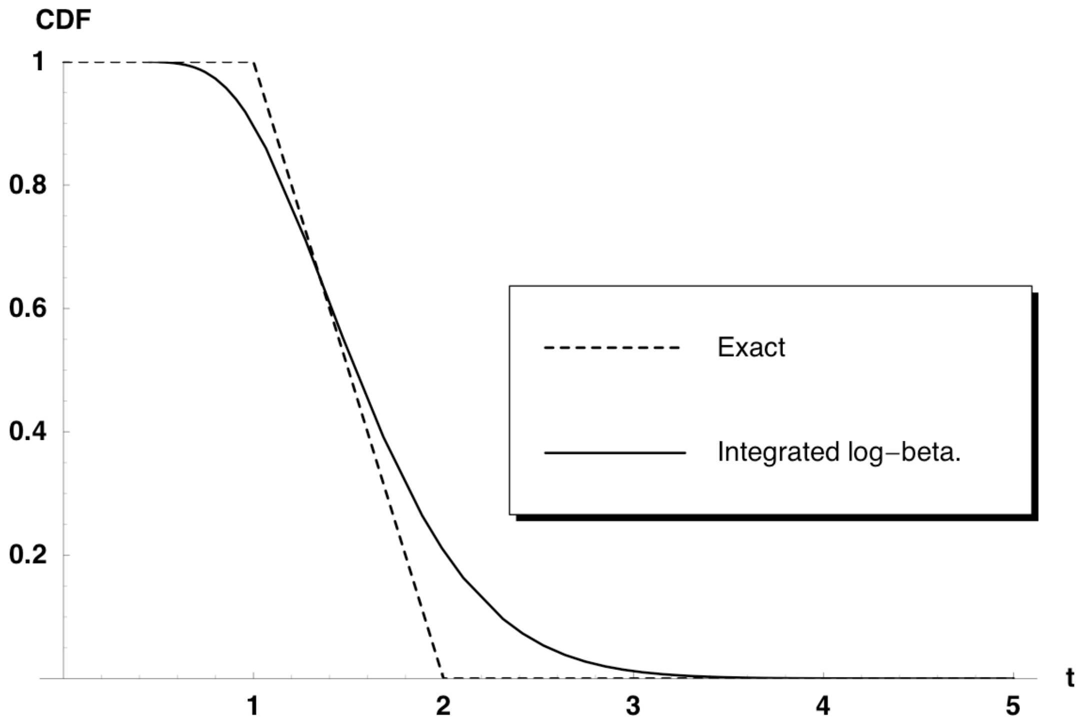

Figure 5. Example 5.2, Uniform(1,2) distribution

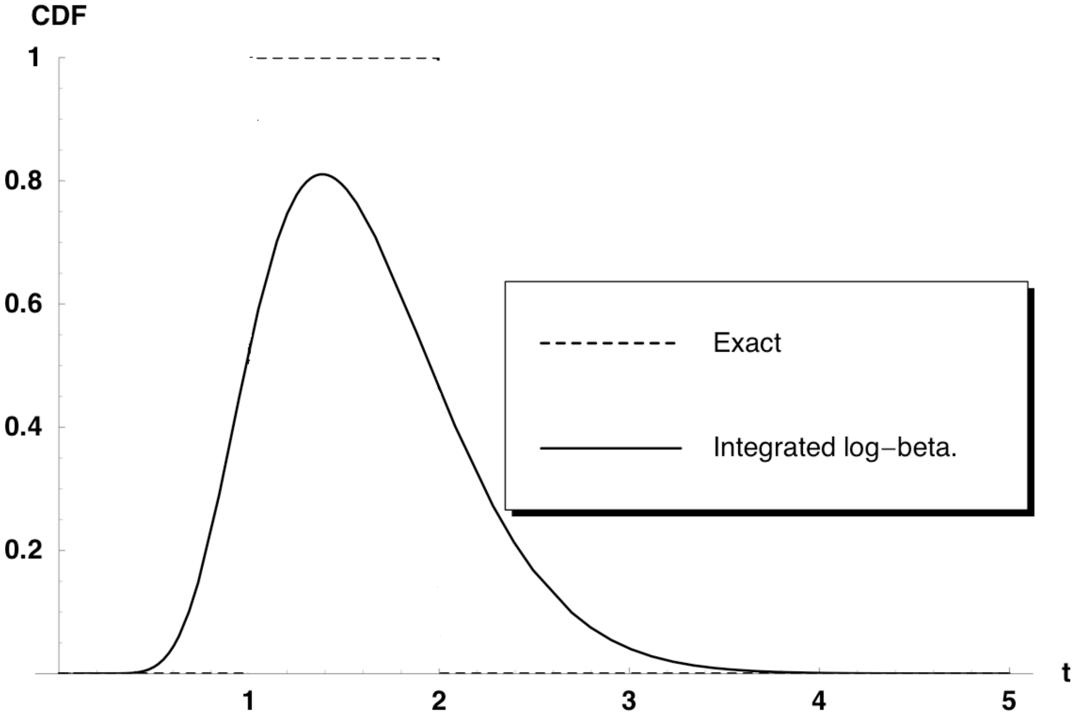

Figure 6. Example 5.3, Pareto distribution

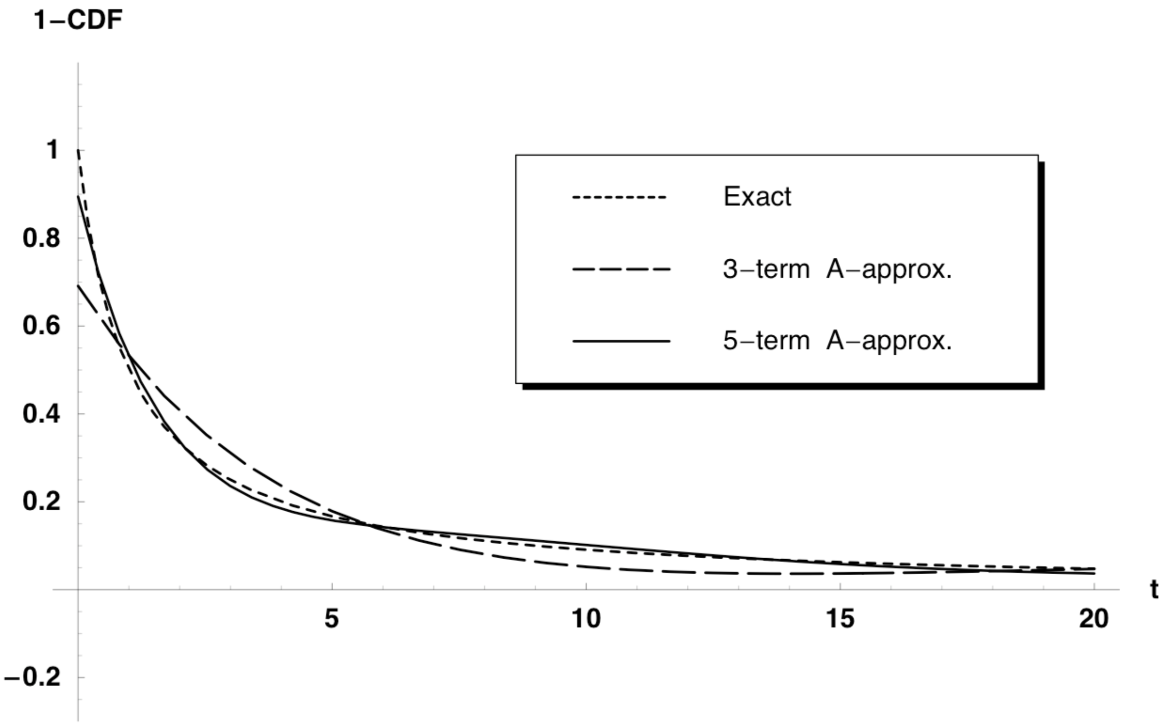

Figure 7. Example 5.3, Pareto distribution

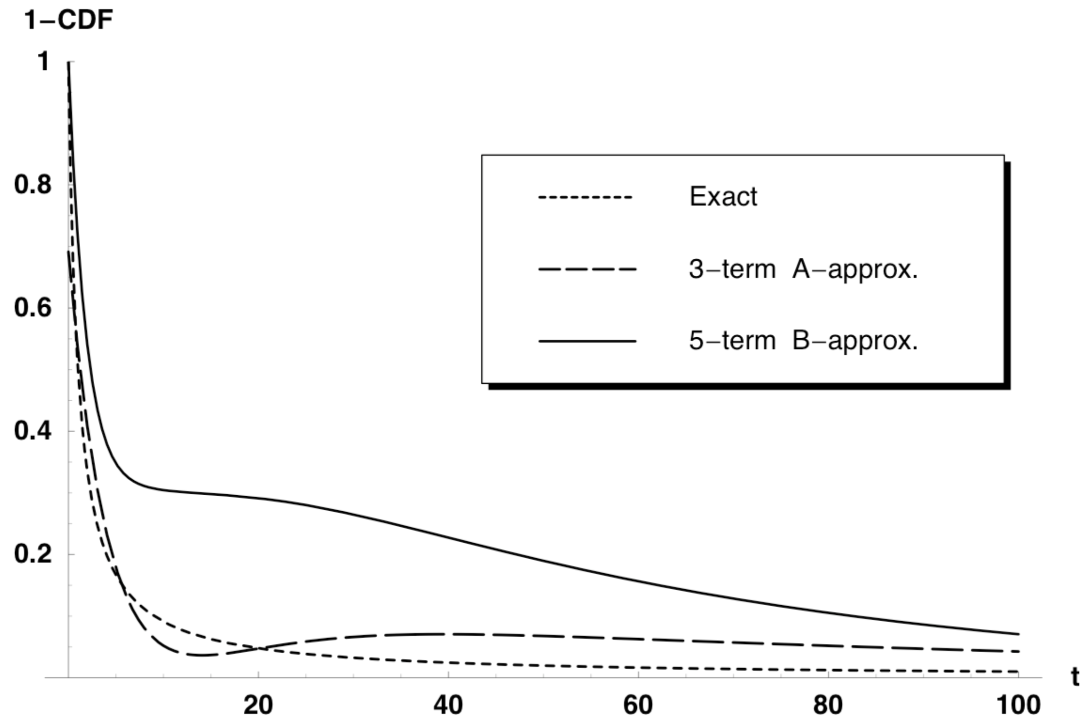

Figure 8. Example 5.4, Lognormal(0,.25) distribution

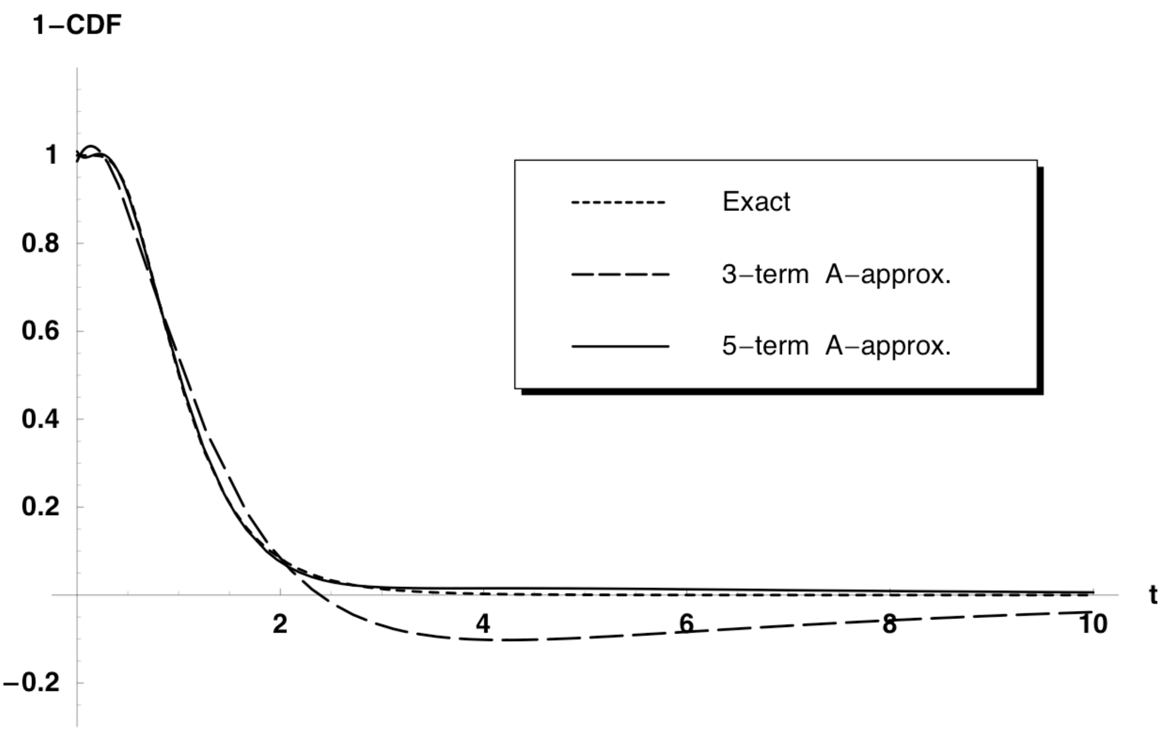

Figure 9. Example 5.4, Lognormal(0,.25) distribution

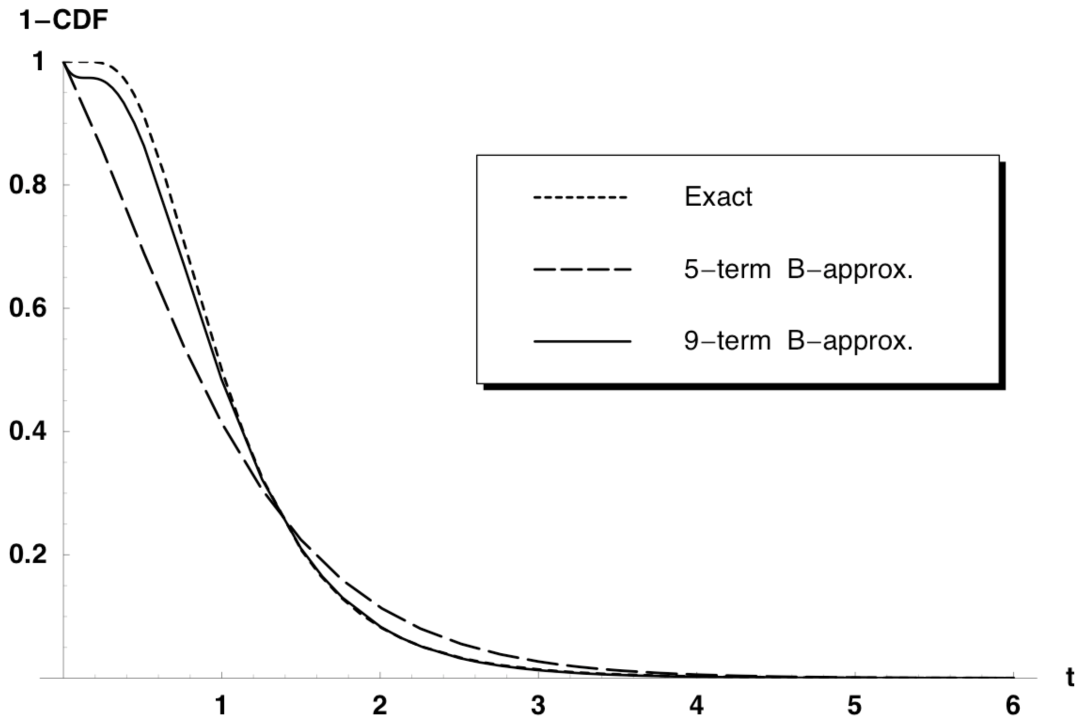

Figure 10. Example 5.5, Makeham distribution