Saddlepoint Approximation for Moment Generating Functions of Truncated Random Variables

Ronald W. Butler and Andrew T.A. Wood
Colorado State University and University of Nottingham

September 27, 2002

Abstract

We consider the problem of approximating the moment generating function (MGF) of a truncated random variable in terms of the MGF of the underlying (i.e. untruncated) random variable. The purpose of approximating the MGF is to enable the application of saddlepoint approximations to certain distributions determined by truncated random variables. Two important statistical applications are: the approximation of certain multivariate cumulative distribution functions; and (ii) the approximation of passage time distributions in ion channel models which incorporate time interval omission. We derive two types of representation for the MGF of a truncated random variable. One of these representations is obtained by exponential tilting. The second type of representation, which has two versions, is referred to as the exponential convolution representation. Each representation motivates a different approximation. It turns out that each of the three approximations is extremely accurate in those cases "to which it is suited". Moreover, there is a simple rule of thumb for deciding which approximation to use in a given case, and if this rule is followed then our numerical and theoretical results indicate that the resulting approximation will be extremely accurate.

1 Introduction

1.1 Saddlepoint Methods

Saddlepoint methods provide approximations to densities and probabilities which are very accurate in a wide variety of settings. This accuracy is seen not only in numerical work, but also in theoretical calculations. In particular, it is often the case that relative errors of these approximations stay bounded in the extreme tails, a desirable property which is not shared by most other types of approximation used in statistics.
For development and discussion of saddlepoint methodology and related methods see Daniels (1954) for details of the density approximation; Barndorff-Nielsen and Cox (1989, 1994) for applications to inference; Lugannani and Rice (1980), Tanaka (1982) and Daniels (1987) for discussion of a tail area approximation which has uniform relative error, and Skovgaard (1987) for a conditional version of this approximation; and Reid (1988) for a review of saddlepoint techniques.

Saddlepoint approximations are constructed by performing various operations on the moment generating function (MGF) or, equivalently, the cumulant generating function (CGF), of a random variable. Let \( X \) be a continuous random variable with density \( f(x) \) with respect to Lebesgue measure, moment generating function \( M(\theta) \) and CGF \( K(\theta) = \log M(\theta) \). Then the saddlepoint density approximation to \( f(x) \) is given by

\[
\hat{f}(x) = \frac{1}{\sqrt{2\pi K''(\hat{\theta})}} \exp\{K(\hat{\theta}) - \hat{\theta}x\}
\]

where \( \theta = \hat{\theta} \) is the (unique) solution to the saddlepoint equation \( K'(\theta) = x \), and the primes denote derivatives. The Lugannani and Rice (1980) saddlepoint approximation to the CDF \( F(x) = P(X \leq x) \) is given by

\[
\hat{F}(x) = \Phi(w) + \phi(w)(w^{-1} - u^{-1})
\]

where \( \hat{\theta} \) is defined as before,

\[
w = \text{sgn}(\hat{\theta}) \sqrt{2[K'(\hat{\theta}) - \hat{\theta}x]}, \quad u = \hat{\theta} \sqrt{K''(\hat{\theta})}
\]

and \( \Phi \) and \( \phi \) are the CDF and density, respectively, of the standard normal.


Unlike much of this previous work, the current paper uses saddlepoint methods to approximate MGFs of truncated distributions with the view that these approximate MGFs may be used for further saddlepoint inversion. The work is therefore more akin to Fraser, Reid, and Wong (1991) and Butler and Wood (2002) who, for a similar purpose, use saddlepoint methods to approximate conditional MGFs from there joint MGFs.

1.2 Truncation

Suppose that \( X_i \) denotes a random variable with known MGF \( M_i(\theta) \) for \( i = 1, \ldots, n \). Now, for each \( i \), we observe \( Y_i = X_i | X_i \in (a_i, b_i) \), i.e. \( Y_i \) is \( X_i \) conditioned to lie in the set \((a_i, b_i)\). In this paper we are concerned with the following
question: is there a convenient and accurate way to approximate the CGF of \( Y_i \) using only \( K_i(\theta) \), the CGF of the untruncated variable \( X_i \)?

If we are just interested in a single random variable, \( Y_1 \) say, then the question is probably not of much interest because the density and CDF of \( Y_1 \) can be expressed simply in terms of the density and CDF of \( X_1 \) with the latter approximated using the saddlepoint approximations indicated above. However, there are situations in which approximations to the CGFs of the \( \{Y_i\} \) are potentially very useful. An important class of such examples occurs when the CGFs of the \( \{Y_i\} \) are the input for a subsequent transform or saddlepoint analysis that is based on the distribution theory of \( \{Y_i\} \). We mention two such examples.

1. We may wish to construct a saddlepoint approximation for the distribution of the sum \( \sum_{i=1}^{n} Y_i \). One such application is to the approximation of certain multivariate CDFs arising in sampling theory and extreme value theory as discussed in Butler and Sutton (1998). For these applications, the multivariate CDF is expressed in terms of the density of \( \sum_{i=1}^{n} Y_i \) where the underlying MGFs of \( \{X_i\} \) are known approximate.

2. In ion channel modeling using Markov and semi-Markov processes, a phenomenon known as time interval omission is commonly built into the model. In effect, this means that only state residences which last for longer than a given time threshold are observed (or detected), and those residences lasting for less than this threshold are not observed (or are undetected). This leads naturally to the consideration of truncating first passage time random variables \( \{X_i\} \) and calling them \( \{Y_i\} \). For the ion channel model applications, the required MGFs of \( \{Y_i\} \) are the components of the matrices \( \Phi_{ac}^D(\theta) \) and \( \Phi_{ac}^U(\theta) \) (Ball et al., 1991, §4), and these are needed when working with these models for subsequent transform analysis and also for likelihood computation and data analysis. In the language of Butler (2000, theorems 1–3), the first passage distributions for \( \{X_i\} \) are only tractable through their MGFs as expressed through the cofactor rules.

The present paper was motivated by application 2 which will be considered in Ball et al. (2002). Application 1 will be considered later in the paper.

### 1.3 Outline of the Paper

In Section 2 we consider two types of representation for the MGF of a truncated random variable, in terms of the MGF of the underlying random variable. One on these representations is obtained by exponential tilting. The second type of representation, which has two versions, is referred to as the exponential convolution representation. In section 3 we consider saddlepoint approximations to the MGF of the truncated random variable which are motivated by these representations, focussing mainly on the case of one-sided truncation. In Section 4 we consider two-sided truncation. In Section 5 results concerning the tail behavior of the various approximations are given. Computation of Dirichlet probabilities
as a consideration of application 1 above are given in Section 6. In Sections 7 and 8, MGFs of truncated lattice random variables and an extension to the multivariate case are considered briefly. A number of numerical examples are provided throughout the paper.

It turns out that each of the three approximations is extremely accurate in those cases "to which it is suited". Moreover, there is a simple rule of thumb (see subsections 3.3 and 4.1) for deciding which approximation to use in a given case, and if this rule is followed, numerical and theoretical results indicate that the resulting approximation will be extremely accurate.

2 Representations of Truncated MGFs

2.1 Preliminaries.

Let $M_0(\theta)$ denote the MGF and $K_0(\theta) = \log M_0(\theta)$ the CGF of a random variable $X$ with CDF $F_0(x) = P(X \leq x)$. Assume that $M_0(\theta)$ has a convergence strip given by $\theta \in (-\alpha, \beta)$ where $0 < \alpha, \beta \leq \infty$. Let $a < b$ denote real numbers such that $F_0(b) - F_0(a) > 0$.

Let

$$M_{(a,b)}(\theta) = \frac{1}{F_0(b) - F_0(a)} \int_a^b e^{\theta x} dF_0(x)$$

(1)

denote the MGF of $X$ truncated at $a$ and $b$, and conditioned to lie in $(a,b)$. We shall refer to $M_{(a,b)}(\theta)$ as a truncated MGF which is an abbreviation for the MGF of a truncated random variable, and similar terminology is used for other quantities such as the CGF.

In this paper we discuss how to approximate the truncated CGF $K_{(a,b)}(\theta) = \log M_{(a,b)}(\theta)$ and its derivatives in terms of the original CGF $K_0(\theta) = \log M_0(\theta)$ and its derivatives.

2.2 Tilted Representation

Let $F_\theta(x)$ denote the CDF of the $\theta$-tilted distribution of $X$, i.e. $dF_\theta(x) = e^{\theta x} dF_0(x)/M_0(\theta)$ and, if $F_0(x)$ has a density $f_0(x)$, then $f_\theta(x) = e^{\theta x} f_0(x)/M_0(\theta)$ is the density corresponding to the CDF $F_\theta$.

Then for $\theta \in (-\alpha, \beta)$, elementary manipulations show that

$$M_{(a,b)}(\theta) = M_0(\theta) \left[ \frac{F_\theta(b) - F_\theta(a)}{F_0(b) - F_0(a)} \right].$$

(2)

We shall refer to (2) as the tilted representation of $M_{(a,b)}(\theta)$.

2.3 Exponential Convolution Representations

We now focus on the continuous case and suppose that the distribution has a density $f_0$. Correspondingly, we use the notation $M_{(a,b)}$ for the right side of (2). The lattice case uses $M_{(a,b)}$ and is considered later in the paper.
We provide alternative representations of (1) which are collectively valid for all $\theta$ in the convergence interval of $M_{(a,b)}$. Define

$$\Xi_1(\theta, y) = \frac{1}{2\pi i} \int_{c_1 - i\infty}^{c_1 + i\infty} M_0(s) \frac{e^{(\theta - s)y}}{s - \theta} ds, \quad -\alpha < c_1 < \min(\beta, \theta) \quad (3)$$

and

$$\Xi_2(\theta, y) = \frac{1}{2\pi i} \int_{c_2 - i\infty}^{c_2 + i\infty} M_0(s) \frac{e^{(\theta - s)y}}{s - \theta} ds, \quad \max(-\alpha, \theta) < c_2 < \beta. \quad (4)$$

**Theorem 2.1 (Properties of $\Xi_1$ and $\Xi_2$)** Suppose that $F_0$ is absolutely continuous with density $f_0$, and assume that for each $c \in (-\alpha, \beta)$ there exists a $\nu(c) \in (0, \infty)$ such that

$$\int_{t \in \mathbb{R}} |M_0(c + it)|^{1 + \nu(c)} dt < \infty. \quad (5)$$

Then the following results hold.

(i) We have

$$\Xi_1(\theta, y) = \int_{-\infty}^{y} e^{\theta x} f_0(x) dx \quad \theta \in (-\alpha, \infty), \quad (6)$$

and

$$\Xi_2(\theta, y) = \int_{y}^{\infty} e^{\theta x} f_0(x) dx \quad \theta \in (-\infty, \beta). \quad (7)$$

Hence

$$\Xi_1(\theta, y) + \Xi_2(\theta, y) = M_0(\theta), \quad \theta \in (-\alpha, \beta). \quad (8)$$

(ii) Let $X$ denote a random variable with MGF $M_0(\theta)$ and let $E$ denote an exponential random variable with rate parameter $|\theta|$ which is independent of $X$. When $\theta > 0$

$$\Xi_1(\theta, y) = \frac{e^{\theta y}}{\theta} f_X + E(y); \quad (9)$$

and when $\theta < 0$,

$$\Xi_1(\theta, y) = M_0(\theta) - \frac{e^{\theta y}}{|\theta|} f_X - E(y). \quad (10)$$

Above and below, $f_Z$ denotes the density of a random variable $Z$.

(iii) When $\theta > 0$,

$$\Xi_2(\theta, y) = M_0(\theta) - \frac{e^{\theta y}}{\theta} f_X + E(y); \quad (11)$$

and when $\theta < 0$,

$$\Xi_2(\theta, y) = \frac{e^{\theta y}}{|\theta|} f_X - E(y). \quad (12)$$
(iv) In the respective domains of definition for $\Xi_1$ and $\Xi_2$,

$$
\mathcal{M}_{(-\infty,y)}(\theta) = \frac{\Xi_1(\theta, y)}{F_0(y)} \quad \text{and} \quad \mathcal{M}_{(y,\infty)}(\theta) = \frac{\Xi_2(\theta, y)}{1-F_0(y)}.
$$

(v) For a general interval $(a,b)$, $\mathcal{M}_{(a,b)}(\theta)$ has the alternative representations

$$
\mathcal{M}_{(a,b)}(\theta) = \frac{\Xi_1(\theta,b) - \Xi_1(\theta,a)}{F(b)-F(a)} \quad \theta \in (-\alpha, \alpha) \quad (14)
$$

and

$$
\mathcal{M}_{(a,b)}(\theta) = \frac{\Xi_2(\theta,a) - \Xi_2(\theta,b)}{F(b)-F(a)} \quad \theta \in (-\alpha, \alpha) \quad (15)
$$

We refer to (13), (14) and (15) as exponential convolution representations of the corresponding truncated MGFs.

**Proof of Theorem 2.1.** Using the convolution formula for densities (see e.g. Theorem 6.1.2 in Chung, 1974, for a precise statement) we have, for $\theta > 0$,

$$
\int_{-\infty}^{\infty} e^{\theta x} f_0(x) dx = \frac{e^{\theta y}}{\theta} \int_{-\infty}^{\infty} \theta e^{-\theta (y-u)} I(u \leq y) f_0(u) du = \frac{e^{\theta y}}{\theta} f_{X+E}(y)
$$

where $E$ is an exponential random variable with rate parameter $\theta$ which is independent of $X$. Define

$$
H_{c,\theta}(t) = \frac{M_0(c+it)}{1-(c+it)/\theta} / \frac{M_0(c)}{1-c/\theta}
$$

Note that $H_{0,\theta}(t)$ is the characteristic function (CF) of $f_{X+E}(y)$, and $H_{c,\theta}(t)$ is the CF of the $c$-tilted density $f_{X+E}(y)e^{cy}/\{M_0(c)/(1-c/\theta)\}$. Using Hölder's inequality,

$$
\int_{-\infty}^{\infty} |H_{c,\theta}(t)| dt \leq \frac{1-c/\theta}{M_0(c)} \left( \int_{-\infty}^{\infty} |M_0(c+it)|^{1+\nu(c)} dt \right)^{1/1+\nu(c)} \times \left( \int_{-\infty}^{\infty} \frac{1}{|1-(c+it)/\theta|^{1+\nu(c)}} dt \right)^{\nu(c)} < \infty
$$

for each $c \in (-\alpha, \min(\beta, \theta))$. Therefore we may apply the Fourier inversion theorem (see e.g. Chung, 1974, p.155, for a precise statement) to $H_{c,\theta}(t)$ to obtain

$$
\frac{1}{2\pi} \int_{-\infty}^{\infty} H_{c,\theta}(t) e^{-ity} dt = f_{X+E}(y)e^{cy}/\{M_0(c)/(1-c/\theta)\}.
$$

After some rearrangement, we find that (16) gives (6) for all $\theta > 0$ and $c \in (-\alpha, \min(\beta, \theta))$. This shows also that $\Xi_1(\theta, y)$ does not depend on the choice of
c_1 \) in (3). An analytic continuation argument extends (6) to \( \theta \in (-\alpha, 0] \), thus (6) is established for all \( \theta > -\alpha \).

Identical reasoning gives (7) and (12), and (8) follows immediately after adding (6) and (7). The statements (13), (14) and (15) follow directly from the definitions.

**Remark 2.1.** Note that if, for some \( c \), (5) holds with \( \nu(c) \in (0, 1] \), then absolute continuity of \( F_0 \) follows; see Theorem 11.6.1 in Kawata (1972). However, if we must take \( \nu(c) > 1 \) for all \( c \) then (5) does not guarantee that \( F_0 \) is absolutely continuous; see Theorem 13.4.2 in Kawata (1972) for a counterexample.

**Remark 2.2.** Although (8) follows immediately from the addition of (6) and (7), it is also interesting to note that (8) is a consequence of Cauchy's theorem, as we now explain. Consider the rectangle in the complex plane with vertices \( V_1 = c_2 - iT, V_2 = c_2 + iT, V_3 = c_1 + iT \) and \( V_4 = c_1 - iT \). Then the integral around this rectangle given by

\[
\frac{1}{2\pi i} \int_{V_4 \to V_3 \to \ldots \to V_1} \frac{M_0(s)}{\theta - s} ds
\]

is equal to the residue of the pole at \( s = \theta \), which is \(-M_0(\theta)\). If \( T \) is large then by standard approximation arguments, the integrals along \( V_2 \to V_3 \) and \( V_4 \to V_1 \) are small. Moreover, the integral from \( V_1 \) to \( V_2 \) is \(-\Xi_2(\theta, y)\) and the integral from \( V_3 \) to \( V_4 \) is \(-\Xi_1(\theta, y)\), so we obtain (8) as \( T \to \infty \).

## 3 Approximations

We now present approximations to the truncated CGFs

\[
K_{(-\infty, y)}(\theta) = \log M_{(-\infty, y)}(\theta)
\]

and

\[
K_{(y, \infty)}(\theta) = \log M_{(y, \infty)}(\theta)
\]

and their derivatives. An approximation to \( K_{(a, b)}(\theta) = \log M_{(a, b)}(\theta) \) of similar type is also given.

### 3.1 Lugannani and Rice Approximation

Using the tilted representation of the truncated MGF we obtain

\[
K_{(-\infty, y)}(\theta) = K_0(\theta) + \log \{F_0(y) / F_0(y)\}.
\]

We may approximate the \( \theta \)-tilted CDF \( F_0(y) \) using the Lugannani and Rice approximation applied to the tilted distribution with CGF \( K_0(s) = K_0(\theta + s) - K_0(\theta) \).

If the convergence strip of \( K_0(\theta) \) is \( \theta \in (-\alpha, \beta) \) with finite \( \beta \), then \( K_{(-\infty, y)}(\theta) \) is defined on the larger set \( (-\alpha, \infty) \) but it is not clear how to extend this
approximation to $\theta \in [\beta, \infty)$. A simple extension is discussed in §5.2, though it turns out that this extension is unsatisfactory.

The Lugannani and Rice approximation to $F_\theta(y)$ is given by

$$\hat{F}_\theta(y) = \Phi(w_\theta) + \phi(w_\theta) \left( \frac{1}{w_\theta} - \frac{1}{u_\theta} \right)$$

where $\Phi$ and $\phi$ are, respectively, the standard normal CDF and density;

$$w_\theta = \text{sgn}(t_y - \theta) \sqrt{2 \{ (t_y - \theta)y - K_0(t_y) + K_0(\theta) \}} \quad (17)$$

and

$$u_\theta = (t_y - \theta) \sqrt{K''_0(t_y)}, \quad (18)$$

where primes denote derivatives and $\text{sgn}(x) = -1, 0, 1$ depending on whether $x$ is negative, zero or positive; and $t = t_y$ is the unique solution to the saddlepoint equation $K'_0(t) = y$.

The approximation $\hat{F}_\theta(y)$ is quite simple to use since it is an explicit function of $\theta$ once $t_y$, the saddlepoint for $\theta = 0$, has been determined; thus the function $\hat{K}_{(-\infty,y)}(\theta)$ is available in explicit form once the single saddlepoint solution $t_y$ has been obtained. To see this, note that the expressions given in (17) and (18) have made use of the saddlepoint for the tilted distribution $s_\theta$ which solves

$$K'_0(s_\theta) = K'_0(s_\theta + \theta) = y = K'_0(t_y).$$

By uniqueness of the saddlepoint, $s_\theta + \theta = t_y$ so that only the computation of $t_y$ is required in order to determine $\{s_\theta : \theta \in (-\alpha, \beta)\}$. This makes CGF approximation

$$\hat{K}_{(-\infty,y)}(\theta) = K_0(\theta) + \log \left\{ \hat{F}_\theta(y) / F_\theta(y) \right\} \quad \theta \in (-\alpha, \beta) \quad (19)$$

explicit in $\theta$.

The derivatives of the approximation are given by

$$\hat{K}'_{(-\infty,y)}(\theta) = K'_0(\theta) + \frac{1}{\hat{F}_\theta(y)} \frac{\partial \hat{F}_\theta(y)}{\partial \theta}$$

and

$$\hat{K}''_{(-\infty,y)}(\theta) = K''_0(\theta) + \frac{1}{\hat{F}_\theta(y)} \frac{\partial^2 \hat{F}_\theta(y)}{\partial \theta^2} - \left\{ \frac{1}{\hat{F}_\theta(y)} \frac{\partial \hat{F}_\theta(y)}{\partial \theta} \right\}^2$$

The first partial derivative is given by

$$\frac{\partial \hat{F}_\theta(y)}{\partial \theta} = \phi(w_\theta) \left[ (y - K'_0(\theta)) \left( \frac{1}{w_\theta} - \frac{1}{u_\theta} \right) - \frac{1}{(t_y - \theta)^2 \sqrt{K''_0(t_y)}} \right]$$

and the second partial derivative $\partial^2 \hat{F}_\theta(y) / \partial \theta^2$ is most easily obtained by numerical differentiation.
In the case of \( \tilde{K}_{(y,\infty)}(\theta) \), we have the approximations

\[
\tilde{K}_{(y,\infty)} = K_0(\theta) + \log \left\{ \frac{1 - \tilde{F}_0(y)}{1 - F_0(y)} \right\} \quad \theta \in (-\alpha, \beta)
\]

\[
\tilde{K}_{(y,\infty)}' = K_0'(\theta) - \frac{1}{1 - F_0(y)} \frac{\partial \tilde{F}_0(y)}{\partial \theta},
\]

and

\[
\tilde{K}_{(y,\infty)}'' = K_0''(\theta) - \frac{1}{1 - F_0(y)} \frac{\partial^2 \tilde{F}_0(y)}{\partial \theta^2} - \left( \frac{1}{1 - F_0(y)} \frac{\partial \tilde{F}_0(y)}{\partial \theta} \right)^2,
\]

where the partial derivatives of \( \tilde{F}_0(y) \) are the same as before.

For general \( a < b \) we may approximate \( \tilde{K}_{(a,b)}(\theta) = \log M_{(a,b)} \) by

\[
\tilde{K}_{(a,b)}(\theta) = K_0(\theta) + \log \left\{ \frac{\tilde{F}_0(b) - \tilde{F}_0(a)}{F_0(b) - \tilde{F}_0(a)} \right\} \quad \theta \in (-\alpha, \beta).
\]

This is an explicit expression in \( \theta \) once two saddlepoints have been determined by solving \( K_0'(t_a) = a \) and \( K_0'(t_b) = b \). The derivatives are calculated in similar fashion.

**Example 3.1** The Exponential (1) density is truncated at 2 and \( \tilde{K}_{(2,\infty)}(\theta) \) is compared with \( K_{(2,\infty)}(\theta) \), its exact counterpart. The two graphs were indistinguishable.

*Fig. 1.* The accuracy of \( \tilde{K}_{(2,\infty)} \) in the Exponential (1) setting as measured by error (dashed) and percentage relative error (solid).
The accuracy is revealed in Figure 1 which shows the error \( \hat{K}_{(2, \infty)}(\theta) - K_{(2, \infty)}(\theta) \) (dashed) and percentage relative error (solid)

\[
r(\theta) = 100 \left\{ \frac{\hat{K}_{(2, \infty)}(\theta)}{K_{(2, \infty)}(\theta)} - 1 \right\} \quad \text{vs.} \quad \theta \in (-8, 1).
\]

**Example 3.2** The complementary truncation of the Exponential (1) concerns \( K_{(0,2)} \). Approximation (19) is quite accurate for \( \theta < 0.8 \) but not for \( \theta > 0.8 \). This may be seen in Figure 2 which plots \( K_{(0,2)} \) (dashed), its continuation (dotted) as discussed in §5.2, and \( K_{(0,2)} \) (solid) for \( \theta \in (0.5, 1.5) \). Note that the dashed and dotted portions join continuously but not smoothly and are far too inaccurate for use with further saddlepoint approximation.

![Fig. 2. Plot of \( K_{(0,2)}(\theta) \) (solid) for an Exponential (1) along with its approximation \( \hat{K}_{(0,2)}(\theta) \) (dashed) for \( \theta \leq 1 \) and its continuation (dotted) for \( \theta \geq 1 \).](image)

**Example 3.3** For the truncation of an arbitrary normal distribution, the Lugannani and Rice approximations are exact with \( \hat{K}_{(a,b)}(\theta) = K_{(a,b)}(\theta) \) for any \( (a,b) \) and \( \theta \). To see this, first note that the truncation is the same as applying the standardized truncation to the standard normal. Within the standard setting, the \( \theta \)-tilt of \( N(0,1) \) is \( N(\theta,1) \) and the Lugannani and Rice approximation is exact.

### 3.2 The Exponential Convolution Approximations

The exponential convolution approximations are obtained by applying saddlepoint approximations to the integrals defining \( \Xi_1(\theta, y) \) and \( \Xi_2(\theta, y) \). Denote
these saddlepoint approximations by \( \tilde{\xi}_1(\theta, y) \) and \( \tilde{\xi}_2(\theta, y) \). Then in this approach the CGFs \( K_{(-\infty, y)}(\theta) \) and \( K_{(y, \infty)}(\theta) \) are approximated by

\[
\tilde{K}_{(-\infty, y)}(\theta) = \log \left\{ \frac{\tilde{\xi}_1(\theta, y)}{\tilde{\xi}_1(0, y)} \right\} \quad \theta > -\alpha \tag{22}
\]

and

\[
\tilde{K}_{(y, \infty)}(\theta) = \log \left\{ \frac{\tilde{\xi}_2(\theta, y)}{\tilde{\xi}_2(0, y)} \right\} \quad \theta < \beta \tag{23}
\]

To reduce the number of formulae in this subsection, we shall use the subscripts \( 1 \) and \( 2 \) to indicate the intervals \(( -\infty, y )\) and \(( y, \infty )\), respectively.

The saddlepoint approximations to \( \tilde{\xi}_j(\theta, y) \) \( (j = 1, 2) \) are given by

\[
\tilde{\xi}_j(\theta, y) = e^{\theta y} \frac{1}{\sqrt{2\pi}} \left| K''_0(s_{j, \theta}) + \frac{1}{(\theta - s_{j, \theta})^2} \right|^{-1/2} \exp \left\{ K_0(s_{j, \theta}) - \log |\theta - s_{j, \theta}| - s_{j, \theta}y \right\} \tag{24}
\]

where \( s_{j, \theta} \) is the unique solution to

\[
K_0'(s) + \frac{1}{\theta - s} = y \tag{25}
\]

in \( (-\alpha, \beta) \) which satisfies \( s_{1, \theta} < \theta \) \( (j = 1) \) and \( s_{2, \theta} > \theta \) \( (j = 2) \).

After some simplifications we obtain

\[
\tilde{K}_j(\theta) = \theta y + D_j(\theta) - D_j(0) + K_0(s_{j, \theta}) - K_0(s_{j, \theta}) - (s_{j, \theta} - s_{j, \theta}y) \tag{26}
\]

where, using implicit differentiation, we have

\[
D_j(\theta) \equiv \frac{1}{2} \log \left( \frac{\partial s_{j, \theta}}{\partial \theta} \right) = \frac{1}{2} \log \left( 1 + (\theta - s_{j, \theta})^2 K''_0(s_{j, \theta}) \right).
\]

Note that the approximations are calibrated so that

\[
\tilde{K}_j(0) = K_j(0) = 0 \quad j = 1, 2.
\]

The first derivative of \( \tilde{K}_j(\theta) \) \( (j = 1, 2) \) is given by

\[
\tilde{K}_j'(\theta) = y + D_j'(\theta) - (y - K_0'(s_{j, \theta})) \frac{\partial s_{j, \theta}}{\partial \theta}
\]

where

\[
D_j'(\theta) = \frac{1}{2} \frac{\partial^2 s_{j, \theta}}{\partial \theta^2} \frac{\partial s_{j, \theta}}{\partial \theta}
\]

\[
\frac{\partial s_{j, \theta}}{\partial \theta} = \left\{ 1 + (\theta - s_{j, \theta})^2 K''_0(s_{j, \theta}) \right\}^{-1}
\]

and

\[
\frac{\partial^2 s_{j, \theta}}{\partial \theta^2} = -\frac{(\theta - s_{j, \theta})^2 \left[ K''_0(s_{j, \theta}) + 2(\theta - s_{j, \theta}) \{ K_0'(s_{j, \theta}) \} \right]}{\left\{ 1 + (\theta - s_{j, \theta})^2 K''_0(s_{j, \theta}) \right\}^3}.
\]

The second partial derivative \( \frac{\partial^2 \tilde{K}_j(\theta)}{\partial \theta^2} \) can be determined using numerical differentiation.
Example 3.4 For the Exponential (1), Figure 3 compares $\hat{\mathcal{R}}_1 = \hat{\mathcal{R}}_{(0,2)}$ (dashed) with $\mathcal{R}_{(0,2)}$ (solid) and Figure 4 shows $10 \times$ the error (dashed) and the percentage relative error (solid).

Fig. 3. Plot of $\hat{\mathcal{R}}_{(0,2)}(\theta)$ (dashed) and $\mathcal{R}_{(0,2)}(\theta)$ (solid) for an Exponential(1).

Fig. 4. $10 \times$ error (dashed) and percentage relative error (solid) for $\hat{\mathcal{R}}_{(0,2)}(\theta)$ in Figure 3.
Example 3.5 In the case of the standard normal, Figure 5 plots $\tilde{\mathcal{K}}_2 = \tilde{\mathcal{K}}_{(2,\infty)}(\theta)$ (dashed) and $\mathcal{K}_{(2,\infty)}(\theta)$ (solid).

Fig. 5. Plot of $\mathcal{K}_{(3,\infty)}(\theta)$ (solid) for a standard normal along with its approximation $\tilde{\mathcal{K}}_{(2,\infty)}(\theta)$ (dashed). The dotted line is (27).

Fig. 6. 10 × error (dotted) and percentage relative error (dashed) of $\tilde{\mathcal{K}}_{(2,\infty)}(\theta)$ for a standard normal.

The approximation is almost graphically indistinguishable from the true $\mathcal{K}_{(2,\infty)}(\theta)$. 
Approximation $\tilde{K}_{(2,\infty)}(\theta)$, as given in (29), is based entirely on $\tilde{M}_2(\theta, 2)$ in (24) or (26) extended over the entire range of $\theta$. The dotted line in the figure over $\theta \in (0, 4)$ refers to an approximation to $K_{(2,\infty)}(\theta)$ based on $\tilde{M}_2(\theta, 2)$ used in conjunction with identity (11). According to (11), this alternative approximation when normalized at $\theta = 0$ is

$$K_{(2,\infty)}(\theta) = \log \left\{ M_0(\theta) - \tilde{M}_2(\theta, 2) \right\} - \log \left\{ 1 - \tilde{M}_2(0, 2) \right\} \quad \theta > 0. \quad (27)$$

The plot makes it clear that switching the approximation from $\tilde{M}_2$ to $M_0(\theta) - \tilde{M}_1$ as $\theta$ passes upward through 0 leads to a much inferior approximation for $\theta > 0$.

Figure 6 shows $10 \times$ error (dashed) and the percentage relative error (solid) of $K_{(2,\infty)}(\theta)$.

Example 3.6 The Gumbel $(0, 1)$ density has CDF $F(x) = \exp(-e^{-x})$ on $x \in (-\infty, \infty)$ and CGF $K_0(s) = \log \Gamma(1 - s)$ defined for $s \in (-\infty, 1)$. The Gompertz distribution is its truncation restricted to $x \in (0, \infty)$. Figure 7 compares $K_{(0,\infty)}(\theta)$ (dashed) and $K_{(0,\infty)}(\theta)$ (dotted) with $K_{(0,\infty)}(\theta)$ (solid).

![Graph](image)

**Fig. 7.** Plot of $K_{(0,\infty)}(\theta)$ (dashed), $K_{(0,\infty)}(\theta)$ (dotted), and $K_{(0,\infty)}(\theta)$ (solid) for a Gumbel distribution.

Since $K_{(0,\infty)}(\theta)$ is the clear choice for approximation in this instance, Figure 8 plots $100 \times$ error (dashed) and percentage relative error (solid) as reflected in Figure 7.
3.3 Rule of Thumb for One-Side Truncation

1. Suppose $X$ has MGF $M_0(\theta)$ with a convergence strip of the form $(-\infty, \infty), (-\infty, \beta)$, or $(-\alpha, \infty)$ where $0 < \alpha, \beta < \infty$. If the convergence strip for truncated variable $Y$ does not need to be extended, then use the Lugannani and Rice approximation.

2. Using right truncation that extends the convergence strip $(-\infty, \beta)$ or $(-\alpha, \beta)$ to $(-\infty, \infty)$ or $(-\alpha, \infty)$ respectively, then use Lugannani and Rice for $\theta \leq 0$ and the exponentially smoothed approximation $\tilde{K}_{(-\infty, \beta)}(\theta)$ based on $\tilde{E}_1$ for $\theta \geq 0$. Note that the two pieces of the approximation join continuously at $\theta = 0$.

3. When left truncation extends the convergence strip $(-\alpha, \infty)$ or $(-\alpha, \beta)$ to $(-\infty, \infty)$ or $(-\alpha, \beta)$ respectively, then use Lugannani and Rice for $\theta \geq 0$ and the exponentially smoothed approximation $\tilde{K}_{(-\alpha, \infty)}(\theta)$ based on $\tilde{E}_2$ for $\theta < 0$.

4 Two-sided Truncation

For two-sided truncation, it is much harder to accurately approximate $K_{(\alpha, \beta)}(\theta)$. Several methods of approximation are presented and followed by a discussion of which methods are most accurate in the various settings. Generally the accuracy of a method is largely determined by the various types of convergence regions $(-\alpha, \beta)$ for $M_0(\theta)$. 

Fig. 8. Plot of $100 \times$ error (dashed) and percentage relative error (solid) and for the approximation $\tilde{K}_{(0, \infty)}(\theta)$ in Figure 5.
The methods for approximating $\mathcal{K}_{(a,b)}(\theta)$ consist of:
1. The Lugannani and Rice approximation $\tilde{\mathcal{K}}_{(a,b)}(\theta)$ in (21).
2. The \textit{joined second-order approximation} which is defined below. First let
   \[
   \tilde{\Xi}_j(\theta, y) = \xi_j(\theta, y) \left(1 + \frac{1}{8} \kappa_4(s_j, \theta) - \frac{5}{24} \kappa_3^2(s_j, \theta)\right)
   \]
   define the second-order saddlepoint approximation to $\Xi_j(\theta, y)$ where
   \[
   \kappa_3(s_j, \theta) = \frac{K_0'''(s_j, \theta) + 2(\theta - s_j, \theta)^{-3}}{\left\{K_0''(s_j, \theta) + (\theta - s_j, \theta)^{-2}\right\}^{3/2}}
   \]
   \[
   \kappa_4(s_j, \theta) = \frac{K_0''''(s_j, \theta) + 6(\theta - s_j, \theta)^{-4}}{\left\{K_0''(s_j, \theta) + (\theta - s_j, \theta)^{-2}\right\}^{3/2}}.
   \]
   There are two second order approximations to $\mathcal{K}_{(a,b)}(\theta)$ based on (14) and (15):
   \[
   \tilde{\mathcal{K}}_{1,(a,b)}(\theta) = \log \left\{ \frac{\tilde{\Xi}_1(\theta, b) - \tilde{\Xi}_1(\theta, a)}{\tilde{\Xi}_1(0, b) - \tilde{\Xi}_1(0, a)} \right\} \quad \theta \in (-\alpha, \infty) \quad (28)
   \]
   and
   \[
   \tilde{\mathcal{K}}_{2,(a,b)}(\theta) = \log \left\{ \frac{\tilde{\Xi}_2(\theta, a) - \tilde{\Xi}_2(\theta, b)}{\tilde{\Xi}_2(0, a) - \tilde{\Xi}_2(0, b)} \right\} \quad \theta \in (-\infty, \beta). \quad (29)
   \]
   The joined second-order approximation uses (28) for $\theta \in (0, \infty)$ and (29) for $\theta \in (-\infty, 0)$. The approximations join together continuously but not necessarily smoothly. The second-order term is recommended here since it tends to improve on the accuracy of first-order in this two-sided setting.
3. Use a \textit{two-step approximation} which is an iterative use of the one-sided approximations. For example, the Lugannani and Rice approximation may be used first for either truncation above $a$ or below $b$. As previously mentioned, this one-sided approximation is explicit in $\theta$ once $t_y$ has been computed. This approximation then becomes the input into the second level of approximation which accommodates the remaining truncation at $b$ or $a$ respectively. This method is quite simple to apply as just described if the Lugannani and Rice approximation is used for the first step. However, it becomes much harder if an exponential convolution approximation is used first.

When $M_0$ converges on $(-\infty, \infty)$, then generally the Lugannani and Rice approximation is most accurate but the joined approximation can also be quite good.

\textbf{Example 4.1} Consider truncating the standard normal to $(-1,2)$. The Lugannani and Rice approximation $\tilde{\mathcal{K}}_{(-1,2)}(\theta)$ is exact for this setting. For the exponentially convoluted approximation, Figure 9 compares $\tilde{\mathcal{K}}_{1,(-1,2)}(\theta)$ (dotted) and $\tilde{\mathcal{K}}_{2,(-1,2)}(\theta)$ (dashed) with $\mathcal{K}_{(-1,2)}(\theta)$ (solid) for $\theta \in (-5,5)$. The figure dramatically shows the failure of $\tilde{\mathcal{K}}_{1,(-1,2)}$ and $\tilde{\mathcal{K}}_{2,(-1,2)}$ for sufficiently negative and positive values of $\theta$ respectively.
Fig. 9. Plot of $\tilde{K}_{1,(-1,2)}(\theta)$ (dotted), $\tilde{K}_{2,(-1,2)}(\theta)$ (dashed), and $K_{(-1,2)}(\theta)$ (solid) for $\theta \in (-5,5)$.

Figure 9 also suggests that the joined approximation should provide very accurate approximation. This is confirmed in Figure 10 which plots $100 \times$ error (dashed) and percentage relative error (solid) for the joined approximation. The missing solid portion within $(-3,0)$ is outside the range of $(0,6)$.

Fig. 10. Errors of the join of $\{\tilde{K}_{1,(-1,2)}(\theta) : \theta > 0\}$ and $\{\tilde{K}_{2,(-1,2)} : \theta \leq 0\}$.
Figure 9 is representative of these exponential convolution approximations. Typically $\hat{K}_{1,(a,b)}(\theta)$ and $\hat{K}_{2,(a,b)}(\theta)$ is quite accurate for $\theta > 0$ ($\theta < 0$) but eventually quite inaccurate for $\theta << 0$ ($\theta >> 0$). The discussion on relative errors in §§5.3-5.4 explains the nature of this inaccuracy. Basically the absolute errors converge to zero as $|\theta| \to \infty$ on the accurate side but diverge on the other inaccurate side.

Example 4.2 The Gumbel distribution is truncated and restricted to $(-1,2)$. Consider a two-step approximation that first truncates by using the Lagannani and Rice approximation at $-1$. This produces $\hat{K}_{(-1,\infty)}(\theta)$ (dashed) which is barely distinguishable from $K_{(-1,\infty)}(\theta)$ (solid).

![Graph of $\hat{K}_{(-1,\infty)}(\theta)$ (dashed) and $K_{(-1,\infty)}(\theta)$ (solid) for a Gumbel.](image)

Fig. 11. Plot of $\hat{K}_{(-1,\infty)}(\theta)$ (dashed) and $K_{(-1,\infty)}(\theta)$ (solid) for a Gumbel.

A plot of $100 \times$ error (dashed) and percentage relative error (solid) is shown in Figure 12; the first step has been highly successful. The second step entails approximating $K_{(-1,2)}(\theta)$ by using the exponential convolution approximation as applied to $\hat{K}_{(-1,\infty)}(\theta)$. This leads to the two-step approximation (dashed) in Figure 13 as the final approximation to $K_{(-1,2)}(\theta)$ (solid). Note the extreme accuracy for $\theta > 0$ that is characteristic of $\hat{K}_{(-\infty,2)}(\theta)$ in (22). This accuracy is seen more clearly in Figure 14 which plots $100 \times$ error (dashed) and percentage relative error (solid).
Fig. 12. $100 \times$ error (dashed) and percentage relative error (solid) of the approximation in Figure 11.

Fig. 13. Two-step approximation $\tilde{\mathcal{K}}_{1,-1,2}(\theta)$ (dashed), the joined approximation (dotted), and $\mathcal{K}_{-1,2}(\theta)$ (solid).
Fig. 14. $100 \times$ error (dashed) and percentage relative error (solid) of the two-step approximation for $\mathcal{K}_{(-1,2)}(\theta)$.

The joined estimate is given by the dotted line in Figure 13. It is also quite accurate and highly successful in extending the approximation to $\theta > 1$. Its errors are plotted in Figure 15. Direct comparison of $100 \times$ the errors of the two-step (solid) and joined (dashed) approximations is shown in Figure 16.

Fig. 15. $100 \times$ error (dashed) and percentage relative error (solid) of the joined approximation for $\mathcal{K}_{(-1,2)}(\theta)$. 
This example has been made simple by using the Lugannani and Rice approximation to deal with the first truncation. More generally, if either $\alpha$ or $\beta$ is finite but not both, then Lugannani and Rice may always be applied first followed by an exponential convolution for the second step. In the first step of the example, the Gumbel convergence strip $(-\infty, 1)$ was preserved by $\tilde{K}_{(-1,\infty)}(\theta)$ to deal with lower truncation at $-1$. In dealing with upper truncation at 2 for the second step, the use of the $\tilde{S}_4$ approximation leading to $\tilde{K}_{2,(-1,2)}(\theta)$ accurately extends the range of convergence to $(-\infty, \infty)$.

4.1 Rule of Thumb for Two-Sided Truncation

1. If $X$ has MGF $M_0(\theta)$ with a convergence strip $(-\infty, \infty)$, then use the Lugannani and Rice approximation.

2. With convergence strips $(-\infty, \beta)$, or $(-\alpha, \infty)$ where $0 < \alpha, \beta < \infty$, then the two-step approximation or the joined approximation should be used. The former is likely to be the more accurate particularly if the first step uses the Lugannani and Rice approximation (without extending the convergence strip) followed by a second step of exponential convolution that extends the convergence strip to $(-\infty, \infty)$.

3. With finite convergence strip $(\alpha, \beta)$ for $M_0(\theta)$, then the joined approximation should be used. A more complicated approach would use the two-step approximation incorporating exponential convolution at each step. The difficulty here is that unless saddlepoints are explicit at the first step, then
two layers of saddlepoints need to be found which involves considerably more computation.

5 Theoretical Accuracy in the Tails

We now investigate the behavior of the approximations to $\mathcal{M}_{(a,b)}(\theta)$ and $K_{(a,b)}^{(r)}(\theta)$, $r = 1, 2$, as $|\theta| \to \infty$. We make the following assumptions throughout this section.

(A1): The exponential family $\{F_\theta : \theta \in (-\alpha, \beta)\}$ is steep, i.e. $|K_0^r(\theta)| \to \infty$ as $\theta \downarrow -\alpha$ and as $\theta \uparrow \beta$.

(A2): The density $f_0$ has one-sided limits at the truncation points $a$ and $b$, i.e. the limits

$$\lim_{\epsilon \downarrow 0} f_0(a + \epsilon) = f_0(a+) \quad \text{and} \quad \lim_{\epsilon \downarrow 0} f_0(b - \epsilon) = f_0(b-)$$

both exist.

Note that, under (A1) and (A2) and regardless of the value of $a \geq -\infty$, we have, as $\theta \to \infty$,

$$\mathcal{M}_{(a,b)}(\theta) \sim \begin{cases} \theta^{-1} e^{\theta b} f_0(b-) /[F_0(b) - F_0(a)] & \text{if } b < \infty \\ M_0(\theta)/[F_0(b) - F_0(a)] & \text{if } b = \infty \end{cases}$$

and, regardless of the value of $b \leq \infty$, we have, as $\theta \to -\infty$,

$$\mathcal{M}_{(a,b)}(\theta) \sim \begin{cases} \theta^{-1} e^{\theta a} f_0(a+) /[F_0(b) - F_0(a)] & \text{if } a > -\infty \\ M_0(\theta)/[F_0(b) - F_0(a)] & \text{if } a = -\infty \end{cases}$$

5.1 Accuracy of the Lugannani and Rice Approximation

We first consider the accuracy in the tails of the Lugannani and Rice (LR) approximation $\hat{\mathcal{M}}_{(a,b)}$ and its logarithmic derivatives $\hat{\mathcal{K}}_{(a,b)}$ and $\hat{\mathcal{K}}''_{(a,b)}$.

Theorem 5.1 Consider the LR approximation $\hat{\mathcal{M}}_{(a,b)}(\theta)$ specified in subsection 3.1. Assume that (A1) and (A2) both hold. Suppose also that (i) $\alpha = \infty$ in all statements concerning the left tail and $\beta = \infty$ in all results concerning the right tail; and (ii) as $|\theta| \to \infty$, $u_\theta/u_\theta^2 \to 0$, where $u_\theta$ and $u_\theta$ are given in (17) and (18) respectively, with $y = a$ or $b$ as appropriate.

(i) As $\theta \to \infty$,

$$\hat{\mathcal{M}}_{(a,b)}(\theta) \sim \begin{cases} \theta^{-1} e^{\theta b} \hat{f}_0(b)/[\hat{F}_0(b) - \hat{F}_0(a)] & \text{if } b < \infty \\ M_0(\theta)/[1 - \hat{F}_0(a)] & \text{if } b = \infty \end{cases}$$

and as $\theta \to -\infty$,

$$\hat{\mathcal{M}}_{(a,b)}(\theta) \sim \begin{cases} \theta^{-1} e^{\theta a} \hat{f}_0(a)/[\hat{F}_0(b) - \hat{F}_0(a)] & \text{if } a > -\infty \\ M_0(\theta)/\hat{F}_0(b) & \text{if } a = -\infty \end{cases}$$

22
where $\hat{f}_0$ is the saddlepoint density approximation to $f_0$ and $\hat{F}_0$ is the Lugannani and Rice approximation to the CDF $F_0$.

(ii) As $\theta \to \infty$,
\[
\tilde{\mathcal{K}}_{(a,b)}(\theta) = \begin{cases} 
 b - \theta^{-1} + o(\theta^{-1}) & \text{if } b < \infty \\
 K_0'(\theta)(1 + o(1)) & \text{if } b = \infty;
\end{cases}
\]
and as $\theta \to -\infty$,
\[
\tilde{\mathcal{K}}_{(a,b)}(\theta) = \begin{cases} 
 a - \theta^{-1} + o(\theta^{-1}) & \text{if } a > -\infty \\
 K_0'(\theta)(1 + o(1)) & \text{if } a = -\infty;
\end{cases}
\]

(iii) As $\theta \to \infty$,
\[
\tilde{\mathcal{K}}''_{(a,b)}(\theta) = \begin{cases} 
 \theta^{-2} & \text{if } b < \infty \\
 K''_0(\theta) & \text{if } b = \infty;
\end{cases}
\]
and as $\theta \to -\infty$,
\[
\tilde{\mathcal{K}}''_{(a,b)}(\theta) = \begin{cases} 
 \theta^{-2} & \text{if } a > -\infty \\
 K''_0(\theta) & \text{if } a = -\infty;
\end{cases}
\]

Proof of Theorem 5.1. The LR approximation to $\mathcal{M}_{(a,b)}(\theta)$ is given by
\[
\tilde{\mathcal{M}}_{(a,b)}(\theta) = M_0(\theta) \left[ \frac{\hat{F}_0(b) - \hat{F}_0(a)}{\hat{F}_0(b) - \hat{F}_0(a)} \right].
\]

Part (i):

Case: $\theta \to \infty$, $b = \infty$. Note that $\hat{F}_0(b) = 1$ for all $\theta$ and $\hat{F}_0(a) \to 0$ as $\theta \to \infty$, so $\tilde{\mathcal{M}}_{(a,b)}(\theta) \sim M_0(\theta)/[1 - \hat{F}_0(a)]$ as required.

Case: $\theta \to \infty$, $b < \infty$. Here, $\hat{F}_0(a)/\hat{F}_0(b) \to 0$ so
\[
\tilde{\mathcal{M}}_{(a,b)}(\theta) \sim M_0(\theta) \hat{F}_0(b)/[\hat{F}_0(b) - \hat{F}_0(a)].
\]

By assumption $u_{\theta}/u_{\theta}^3 \to 0$ as $\theta \to \infty$. Moreover, elementary calculations show that as $u_{\theta} \to -\infty$, $\Phi(u_{\theta}) \sim -\phi(u_{\theta})[u_{\theta}^{-1} + u_{\theta}^{-3}]$, and it then follows easily that
\[
\hat{F}_0(b) \sim -\phi(u_{\theta})/u_{\theta} \sim \theta^{-1} e^{b\theta} \hat{F}_0(b)/M_0(\theta) \quad \text{as } \theta \to \infty
\]
where $\hat{f}_0(b) = (2\pi)^{-1/2} |K''_0(t_b)|^{-1/2} \exp\{K_0(t_b) - t_b b\}$ is the saddlepoint approximation to $f_0(b)$. The proofs for $\theta \to -\infty$ with $a = -\infty$ and $a > -\infty$ are similar.

Part (ii): We have
\[
\tilde{\mathcal{K}}'_{(a,b)}(\theta) = K_0'(\theta) + \frac{1}{\hat{F}_0(b) - \hat{F}_0(a)} \left[ \frac{\partial \hat{F}_0(b)}{\partial \theta} - \frac{\partial \hat{F}_0(a)}{\partial \theta} \right].
\]
Case: $\theta \to \infty$, $b = \infty$. Since $\hat{F}_\theta(b) - \hat{F}_\theta(a) \to 1$, $\partial \hat{F}_\theta(b)/\partial \theta = 0$, $\partial \hat{F}_\theta(a)/\partial \theta \to 0$ and $K'_\theta(\theta) \to \infty$, the result follows.

Case: $\theta \to \infty$, $b < \infty$. Here $\hat{F}_\theta(a)/\hat{F}_\theta(b) \to 0$ and

$$\frac{\partial \hat{F}_\theta(a)}{\partial \theta} / \frac{\partial \hat{F}_\theta(b)}{\partial \theta} \to 0.$$

Therefore

$$\hat{K}'_{(a,b)}(\theta) = K'_\theta(\theta) + \frac{1}{\hat{F}_\theta(b)} \frac{\partial \hat{F}_\theta(b)}{\partial \theta} + o(\theta^{-1})$$

$$= K'_\theta(\theta) + b - K'_\theta(\theta) + \theta^{-1} + o(\theta^{-1})$$

$$= b + \theta^{-1} + o(\theta^{-1})$$

as required. The cases $\theta \to -\infty$ with $a = -\infty$ and $a > -\infty$ are proved in similar fashion.

Part (iii): The results here follow from similar but more extensive calculations.

Remark 5.1 Comparison of the results in Theorem 5.1 with the limiting results for $M_{(a,b)}(\theta)$ shows that the relative error stays bounded in all cases. With $\hat{K}'_{(a,b)}(\theta)$ and $\hat{K}''_{(a,b)}(\theta)$, the errors actually go to zero as $|\theta| \to \infty$.

5.2 Extension of the LR approximation beyond a boundary

We now discuss how to extend the Lugannani and Rice approximation beyond the boundary in a case such as Example 3.2. Suppose that the upper boundary of the convergence strip satisfies $\beta \in (0, \infty)$ and consider the Lugannani and Rice approximation to $M_{(-\infty,y)}(\theta)$ as $\theta \to \beta$ from below. For $\theta$ close to $\beta$, a fixed $y$ will be in the lower tail of the CDF $F_\theta$. Then under the condition $u_0/w_0 \to 0$, we have, as $\theta \to \beta$:

$$\hat{M}_{(-\infty,y)}(\theta) = M_0(\theta) \frac{\hat{F}_\theta(y)}{\hat{F}_0(y)} \sim -\frac{M_0(\theta)\phi(u_0)}{u_0 \hat{F}_0(y)}$$

$$\sim -e^{K_0(\theta)} \frac{1}{\sqrt{2\pi}} \frac{\exp \{K_0(t_\theta) - K_0(\theta) - (t_\theta - \theta)t_\theta \}}{(t_\theta - \theta) \sqrt{K_0''(t_\theta)}} \frac{\hat{f}_0(y)}{\hat{F}_0(y)} \sim \frac{e^{\theta t_\theta}}{\theta - t_\theta} \frac{\hat{f}_0(y)}{\hat{F}_0(y)}.$$

Since the limit is well-defined for $\theta \geq \beta$, we can take this as the continuation of the Lugannani and Rice approximation for $\theta \geq \beta$. [This extension can also be motivated more directly by another saddlepoint approximation; we omit the details.]

It is straightforward to show that if this extension is used then

$$\lim_{\theta \to -\infty} \frac{\hat{M}_{(-\infty,y)}(\theta)}{M_{(-\infty,y)}(\theta)} = \frac{\hat{f}_0(y)}{\hat{F}_0(y) / \hat{F}_0(y)}.$$
However, as shown in Example 3.2, this extension provides a continuation of the LR approximation beyond \(\beta\), but unfortunately this extension is not smooth because the derivatives blow up as \(\theta\) approaches \(\beta\) from below. More specifically,
\[
\frac{\partial w_\theta}{\partial \theta} = [K_\theta'(\theta) - y]/w_\theta \to \infty \quad \text{as} \quad \theta \uparrow \beta
\]
in the exponential case. Similar problems occur for other distributions with sufficiently heavy tails (recall that in the world of distributions which possess some finite positive and negative exponential moments, the exponential distribution is heavy tailed).

### 5.3 Accuracy of the Exponential Convolution Approximation

For \(j = 1, 2\) let \(\hat{\gamma}_j(\theta, y)\) and \(\hat{K}_j^{(r)}(\theta)\), \(r = 0, 1, 2\) be as in §3.2 and define \(\hat{M}_j(\theta) = \hat{\gamma}_j(\theta, y)/\hat{\gamma}_j(0, y)\). Also, for \(-\infty < a < b < \infty\), define
\[
\hat{M}_{1,(a,b)}(\theta) = \frac{\hat{\gamma}_1(\theta, b) - \hat{\gamma}_1(\theta, a)}{\hat{\gamma}_1(0, b) - \hat{\gamma}_1(0, a)},
\]
\[
\hat{K}_{1,(a,b)}(\theta) = \log \hat{M}_{1,(a,b)}(\theta),
\]
with corresponding definitions for \(\hat{M}_{2,(a,b)}(\theta)\) and \(\hat{K}_{2,(a,b)}(\theta)\).

**Theorem 5.2** Assume that (A1) and (A2) both hold.

(i) As \(\theta \to \infty\),
\[
\hat{M}_1(\theta) \sim \theta^{-1} e^{\theta y} \hat{f}_0(y)/\hat{\gamma}_1(0, y), \quad \hat{K}_1(\theta) = y - \theta^{-1} + o(\theta^{-1})
\]
and \(\hat{K}_1'(\theta) \sim \theta^{-2}\).

(ii) The limiting behavior of \(\hat{M}_2, \hat{K}_2^+ \) and \(\hat{K}_2^-\) in the lower tail is the same as that of \(\hat{M}_1(\theta), \hat{K}_1(\theta)\) and \(\hat{K}_1'(\theta)\) in the upper tail, as given in part (i).

(iii) If \(-\infty < a < b < \infty\) then as \(\theta \to \infty\),
\[
\hat{M}_{1,(a,b)}(\theta) \sim \frac{e^{\theta b}}{\theta} \frac{\hat{f}_0(b)}{\hat{\gamma}_1(0, b) - \hat{\gamma}_1(0, a)}, \quad \hat{K}_{1,(a,b)}(\theta) = b - \theta^{-1} + o(\theta^{-1})
\]
and \(\hat{K}_{1,(a,b)}'(\theta) \sim \theta^{-2}\).

(iv) If \(\theta \to -\infty\) then \(\hat{M}_{2,(a,b)}(\theta)\) and the derivatives of \(\hat{K}_{2,(a,b)}(\theta)\) obey results corresponding to those in part (iii), but with a replacing \(b\).

**Proof of Theorem 5.2.** In part (i), the key point to note is that \(s_1, b \to t_y\) as \(\theta \to \infty\), and then the proof follows easily. The proof is essentially the same in the other cases.

**Remark 5.2** Comparison of the results in Theorem 5.2 with the limiting results for \(M_{1,(a,b)}(\theta)\) shows that the relative error stays bounded in all cases. With \(K_1(\theta)\) and \(K_2(\theta)\), the errors actually go to zero as \(\theta \to \pm\infty\) in the cases covered by the theorem.
5.4 Behavior in the Other Tail

In Theorem 5.2 we described the limiting behavior of $\mathcal{M}_1(\theta)$ and its logarithmic derivatives as $\theta \to \infty$, and the behavior of $\mathcal{M}_2(\theta)$ and its logarithmic derivatives as $\theta \to -\infty$. In this subsection we indicate, without proof, what happens to $\mathcal{M}_1(\theta)$ and its derivatives when $\theta \to -\infty$. The results for $\mathcal{M}_2(\theta)$ are similar and are therefore omitted.

If

$$\lim_{s \to -\infty} K'^{\alpha}_0(s)/[K'_0(s)]^2 = 0$$

then

$$\mathcal{M}_1(\theta) \sim M_0(\theta)(e^{\sqrt{2\pi}}/\sqrt{\pi})/Z_1(0, y) \quad \text{as} \quad \theta \to -\infty. \quad (30)$$

Under the stronger conditions

$$\lim_{s \to -\infty} K''_0(s)/K'_0(s) \to 0 \quad \text{and} \quad \lim_{s \to -\infty} K^{(4)}_0(s)/[K'_0(s)]^3 \to 0,$$

we have

$$\tilde{K}'_1(\theta) \sim K'_0(\theta); \quad (31)$$

and still stronger conditions are needed to ensure that

$$\tilde{K}'_1(\theta) \sim K''_0(\theta). \quad (32)$$

A sufficient condition for (30), (31) and (32) to hold is the following:

$$\text{for each} \quad j \geq 2, \quad \lim_{s \to -\infty} K^{(j)}_0(s) \text{ stays bounded.} \quad (33)$$

Note that condition (33) holds for the normal distribution, gamma distribution (in the left tail), and any other distribution which has bounded support on the left. However, in the case of $-X$ where $X$ has a gamma or inverse Gaussian distribution, or if $X$ has a logistic distribution, then $\tilde{K}'_1(\theta)$ and $\tilde{K}'_1(\theta)$ do not stay bounded as $\theta \to -\infty$, and (30), (31) and (32) fail to hold.

6 Dirichlet Probability Computations

Densities for sums of independent truncated random variables are needed in order to make selected multivariate probability computations as discussed in Butler and Sutton (1998). Consider, for example, the probability that an arbitrary Dirichlet vector $D = (D_1, ..., D_n) \sim \text{Dirichlet} \{\alpha = (\alpha_1, ..., \alpha_n)\}$ lies in a general rectangular region $(a, b) = \prod_{i=1}^n (a_i, b_i) \subset (0,1)^n$. If the components of $X = (X_1, ..., X_n)$ are independent with $X_i \sim \text{Gamma} (\alpha_i, 1)$, then the Dirichlet is represented in terms of independent Gammas as $D = X/S$, where $S = \sum_{i=1}^n X_i$. Using the independence of $S$ and $X/S$, then the distribution of $D$ is also the conditional distribution of $X$ given that $S = 1$. These facts and Bayes theorem lead to

$$\Pr\{D \in (a, b)\} = \Pr\{X \in (a, b)|S = 1\} = \frac{f_S(1|X \in (a, b))}{f_S(1)} \prod_{i=1}^n \Pr\{X_i \in (a_i, b_i)\} \quad (34)$$
Here, \( f_S\{1 \mid X \in (a, b)\} \) is the density of \( Z = \sum_{i=1}^{n} Y_i \) at 1 where \( Y_i = X_i \mid X_i \in (a_i, b_i) \) which we approximate using the saddlepoint density. The other terms are explicit computations; \( \Pr\{X_i \in \{a_i, b_i\}\} \) is a gamma probability and \( f_S(1) \) is the Gamma \( \left( \sum_{i=1}^{n} \alpha_i, 1 \right) \) density of \( S \) at 1.

The CGF of \( Z \), or \( K_Z(s) = \sum_{i=1}^{n} \bar{K}_i(s) \), may be approximated as \( \bar{K}_Z(s) = \sum_{i=1}^{n} \bar{K}_i(s) \) where \( \bar{K}_i(s) \) is one of the previously discussed approximations for the CGF of truncated variable \( Y_i \). First and second-order saddlepoint density approximations using \( \bar{K}_Z(s) \) as a surrogate for the true CGF \( K_Z(s) \) provide the necessary approximations for the leading term in (34).

The most interesting question here is the extent to which saddlepoint inversion using \( K_Z(s) \) may be replicated by using \( \bar{K}_Z(s) \) and thus avoiding the difficult but exact computations of the truncated CGFs. Table 1 considers this question for some simple examples.

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \alpha )</th>
<th>( a )</th>
<th>( b )</th>
<th>Exact ( K_Z )</th>
<th>SA, Exact ( K_Z )</th>
<th>SA, ( \bar{K}_Z ) using ( \bar{E}_1 )</th>
<th>SA, ( \bar{K}_Z ) using L&amp;R</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>(10, 8, 8)</td>
<td>(0)(^2)</td>
<td>.9527</td>
<td>.8577</td>
<td>.8481</td>
<td>.9756</td>
<td>.9485</td>
</tr>
<tr>
<td>3</td>
<td>(10, 8, 8)</td>
<td>(1, 10, 11)</td>
<td>.9560</td>
<td>-23.6</td>
<td>-23.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(10, 8, 8)</td>
<td>(0)(^3)</td>
<td>.02400</td>
<td>.02479</td>
<td>.02393</td>
<td>.00141</td>
<td>-2.23</td>
</tr>
<tr>
<td>3</td>
<td>(1)(^3)</td>
<td>(.3)(^2)</td>
<td>.02390</td>
<td>36.3</td>
<td>.9667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(1)(^3)</td>
<td>(.2)(^3)</td>
<td>.04000</td>
<td>.03869</td>
<td>.04363</td>
<td>.0117</td>
<td>.00357</td>
</tr>
<tr>
<td>3</td>
<td>(1)(^3)</td>
<td>(.4)(^3)</td>
<td>.04001</td>
<td>11.50</td>
<td>10.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(1)(^3)</td>
<td>(0)(^3)</td>
<td>.04000</td>
<td>.04059</td>
<td>.03869</td>
<td>.02831</td>
<td>.002687</td>
</tr>
<tr>
<td>3</td>
<td>(1)(^3)</td>
<td>(.4)(^3)</td>
<td>.03998</td>
<td>15.59</td>
<td>.9203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(1)(^5)</td>
<td>(0)(^5)</td>
<td>.3680</td>
<td>.3540</td>
<td>.2043</td>
<td>.2535</td>
<td>.5254</td>
</tr>
<tr>
<td>5</td>
<td>(1)(^5)</td>
<td>(.4)(^5)</td>
<td>.3683</td>
<td>1.00</td>
<td>.1335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(1, 2, ..., 5)</td>
<td>(0)(^5)</td>
<td>.5526</td>
<td>.5469</td>
<td>.5445</td>
<td>.5439</td>
<td>.5373</td>
</tr>
<tr>
<td>5</td>
<td>(1, 2, ..., 5)</td>
<td>(.4)(^5)</td>
<td>.5522</td>
<td>-10.14</td>
<td>-9.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(1, 2, ..., 5)</td>
<td>(0)(^5)</td>
<td>.06(^2)2288</td>
<td>.06(^2)2336</td>
<td>.06(^2)2226</td>
<td>.075733</td>
<td>.07(^2)1532</td>
</tr>
<tr>
<td>5</td>
<td>(1, 2, ..., 5)</td>
<td>(.4)(^5)</td>
<td>.06(^2)2227</td>
<td>12.65</td>
<td>.9140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(1, 2, ..., 5)</td>
<td>(1)(^5)</td>
<td>.03220</td>
<td>.03217</td>
<td>.03090</td>
<td>.03183</td>
<td>.03142</td>
</tr>
<tr>
<td>5</td>
<td>(1, 2, ..., 5)</td>
<td>(.3)(^5)</td>
<td>.03220</td>
<td>-10.02</td>
<td>-9.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(1)(^10)</td>
<td>(1/15)(^5)</td>
<td>.06(^4)5080</td>
<td>.06(^4)3175</td>
<td>.06(^4)3150</td>
<td>.06(^4)4969</td>
<td>.06(^4)4888</td>
</tr>
<tr>
<td>10</td>
<td>(1)(^10)</td>
<td>(.3)(^5)</td>
<td>.06(^4)5080</td>
<td>-29.56</td>
<td>-28.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the various values of \( n, \alpha, \) and \( (a, b) \), the "Exact" probability as listed was computed using symbolic computation in Maple V. The ability to make such
computations has restricted the examples that are given. The entry "SA, Exact $K_Z$" uses a second-order saddlepoint density to invert the true CGF $K_Z$ and provides a baseline of comparison for the remaining approximations. Category "SA, $\tilde{K}_Z$ using $\tilde{Z}_i$" approximates the CGF of each $Y_i$ by using the appropriate (one-sided or two-sided) second-order exponential convolution approximation based on $\tilde{Z}_i$ given in (22) and (28) respectively. Upon determination of $\tilde{K}_Z$, the results of its first-order and second-order saddlepoint density inversions are listed. The final column "SA, $\tilde{K}_Z$ using L&R" shows comparable computations using the Lugannani and Rice approximation $\tilde{K}_Z$ in (21) to approximate $K_Z$ at the initial stage.

The numerical accuracy follows that predicted from theoretical considerations with one caveat. The mean of $Z$ is listed in the cell "Mean" and its value relative to value 1 determines whether the listed saddlepoints for methods $\tilde{K}_Z$ and $\tilde{K}_Z$ are negative or positive. The exponential convolution approximations show consistent accuracy with positive saddlepoints as predicted; also those using Lugannani and Rice are consistently accurate with negative saddlepoints as predicted. Inaccuracy is the result of using each method in its inappropriate tail except for two cases of $\tilde{K}_Z$. These are the first and last cases with $n = 5$ and reveal very accurate results when applied with a quite negative saddlepoint - results that are unanticipated from the theory.

7 Lattice Variables

7.1 Preliminaries

We now consider the case in which $M_0(s)$ is the moment generating function of a lattice-valued random variable. For simplicity we stick to the integer lattice, denoted $\mathbb{Z}$. Generalization of the results given below to a general lattice $L = \{u + vj : j \in \mathbb{Z}\}$ ($u, v$ in $\mathbb{R}$) is straightforward and we omit the details.

Here, 
\[ M_0(\theta) = \sum_{v \in \mathbb{Z}} p_0(v) e^{\theta v} \]

and 
\[ M_{[a,b]}(\theta) = \frac{\sum_{a < y < b} p_0(y) e^{\theta y}}{F_0(b) - F_0(a)} \]

where $p_0(y)$ is the mass function associated with the MGF $M_0(\theta)$ and $F_0(y) = \sum_{x \leq y} p_0(x)$ is its CDF. For simplicity we assume in this section that (when finite) $a$ and $b$ are integers.

7.2 Representations for Truncated MGFs

As before, we have the tilted representation 
\[ M_{[a,b]}(\theta) = M_0(\theta) \left\{ \frac{F_0(b) - F_0(a)}{F_0(b) - F_0(a)} \right\} \]
where \( F_0(y) = \sum_{x \leq y} e^{\theta y} p_0(x) / M_0(\theta) \) is the CDF of the \( \theta \)-tilted distribution. For \( y \in Z \) define

\[
\Upsilon_1(\theta, y) = e^{\theta y} \frac{1}{2\pi i} \int_{c_1-i\pi}^{c_1+i\pi} M_0(s) \frac{e^{-s y}}{1 - e^{s \theta}} ds - \alpha < c_1 < \min(\beta, \theta)
\]

and

\[
\Upsilon_2(\theta, y) = e^{\theta y} \frac{1}{2\pi i} \int_{c_2-i\pi}^{c_2+i\pi} M_0(s) \frac{e^{-s y}}{e^{s \theta} - 1} ds, \quad \max(-\alpha, \theta) < c_2 < \beta.
\]

The following result closely parallels Theorem 2.1.

**Theorem 7.1 (Properties of \( \Upsilon_1 \) and \( \Upsilon_2 \))** Below, we write \( p_2(x) \) for the probability mass function of an integer-valued random variable \( Z \).

(i) We have

\[
\Upsilon_1(\theta, y) = \sum_{x \leq y} e^{\theta x} p_0(x) \quad \theta \in (-\alpha, \infty)
\]

and

\[
\Upsilon_2(\theta, y) = \sum_{x > y} e^{\theta x} p_0(x) \quad \theta \in (-\infty, \beta).
\]

(ii) Let \( X \) denote a random variable with MGF \( M_0(\theta) \) and let \( V \) denote an independent Geometric \( (p_0 = 1 - e^{-|\theta|}) \) random variable with mass function \( \Pr(V = y) = p_0(1 - p_0)^y \). Then

\[
\Upsilon_1(\theta, y) = \begin{cases} 
  e^{\theta y} p_{X+V}(y)/(1 - e^{-\theta}) & \text{if } \theta > 0; \\
  M_0(\theta) - e^{\theta (y+1)} p_{X-V}(y+1)/(1 - e^{-|\theta|}) & \text{if } \theta < 0.
\end{cases}
\]

(iii) We have

\[
\Upsilon_2(\theta, y) = \begin{cases} 
  e^{\theta y} p_{X+V}(y)/(1 - e^{-\theta}) & \text{if } \theta > 0; \\
  M_0(\theta) - e^{\theta (y+1)} p_{X-V}(y+1)/(1 - e^{-|\theta|}) & \text{if } \theta < 0.
\end{cases}
\]

(iv) For all \( \theta \) in the respective domains of definition of \( \Upsilon_1 \) and \( \Upsilon_2 \),

\[
M_{(\alpha, \infty)}(\theta) = \frac{\Upsilon_1(\theta, y)}{F_0(y)} \quad \text{and} \quad M_{(\infty, \infty)}(\theta) = \frac{\Upsilon_2(\theta, y)}{1 - F_0(y)}.
\]

(v) For a general interval \((a, b)\) with integer end points \( a \) and \( b \), \( M_{(a,b)}(\theta) \) has the alternative representations

\[
M_{(a,b)}(\theta) = \frac{\Upsilon_1(\theta, b) - \Upsilon_1(\theta, a)}{F_0(b) - F_0(a)}
\]

and

\[
M_{(a,b)}(\theta) = \frac{\Upsilon_2(\theta, a+1) - \Upsilon_2(\theta, b+1)}{F_0(b) - F_0(a)}.
\]
7.3 Approximations

As in the continuous case, each representation motivates a different approximation. In this instance, however, there is also the issue of which continuity correction to use in connection with the Lugagnani and Rice approximation. We refer to these two corrections as “exponential” and “sinh” since these are the transformations through which the saddlepoint is expressed in the \( u(\theta) \) term as in (18). See Daniels (1987, §6) or Butler (2003, §1.2.3) for details.

For the “geometrically convoluted” approximations, the discrete formulae closely parallel those used with continuity. In the case of one-sided truncation, \( \tilde{\mathcal{K}}_j(\theta) \) as given in (26) continues to provide approximations for \( \mathcal{K}_1 = \mathcal{K}_{(-\infty,\bar{y})} \) and \( \mathcal{K}_2 = \mathcal{K}_{(y+1,\infty)} \) with the following modifications:

\[
D_j(\theta) = -\frac{1}{2} \ln |K''_0(s_j,\theta)(\tau_{j,\theta} - 1)^2 - \tau_{j,\theta}|
\]

where \( \tau_{j,\theta} = \exp(s_j,\theta - \theta) \) for \( j = 1, 2 \) and \( s_1,\theta < \theta < s_2,\theta \) are the two solutions in \((-\alpha, \beta)\) to

\[
K''_0(s_j,\theta) + \frac{1}{\exp(-|\theta - s_j,\theta|) - 1} = y, \quad j = 1, 2.
\]

The two-sided truncation formulae follow similarly with suitable modifications to the second-order correction terms.

Example 7.2 The Negative Binomial \((3, 0.2)\) distribution counts the number of tails occurring before tossing the third head with \( \Pr(\text{Heads}) = 0.2 \). Take \( y = 12 \) corresponding to its mean.

![Fig. 17. Plots of \( \tilde{\mathcal{K}}_{[0,12]}(\theta) \) (dashed) and \( \mathcal{K}_{[0,12]}(\theta) \) for a Negative Binomial \((3, 0.2)\).](image)
Figure 17 plots $\tilde{K}_{0,12}(\theta)$ (dashed) and $K_{0,12}(\theta)$ (solid) and Figure 18 considers $10\times$ error (dashed) and the percentage relative error (solid).

Fig. 18: $10\times$ error (dashed) and percentage relative error (solid) of $\tilde{K}_{0,12}(\theta)$ in Figure 17.

Fig. 19. The Negative Binomial $(3,0.2)$ truncated to $[13, \infty)$. Shown are $10\times$ error (dashed) and percentage relative error (solid) for the exponential continuity correction and the same (dotted) and (dot-dashed) for the sinh correction.
Example 7.3 The complementary truncation for the Negative Binomial \((3, 0.2)\) distribution restricts the distribution to \([13, \infty)\). Accuracy of the Lugannani and Rice approximations is shown in Figure 19. For the “exponential” continuity correction, \(10 \times \text{error (dashed)}\) and percentage relative error (solid) are shown. The same respective errors for the sinh correction are given as dotted and dot-dashed. The dotted line has been truncated but its value at \(\theta = -8\) is \(-1.15\). The exponential correction is slightly better for this particular example but which correction is better more generally is likely to be very dependent on the example.

8 Multivariate Truncated MGFs

We now consider extensions of the representations and approximations to the multivariate case.

8.1 Tilted Representation

Define \(M_0(\theta) = E[e^{\theta^T X}]\) where \(\theta \in \mathbb{R}^p\) and \(X\) is a random \(p\)-vector. Let \(\Theta = \{\theta \in \mathbb{R}^p : E[e^{\theta^T X}] < \infty\}\) and let \(\Theta^o\) denote the interior of \(\Theta\). Let \(F_0\) denote the multivariate CDF of \(X\), write \(dF_0(x) = e^{\theta^T X} dF_0(x)/M_0(\theta)\) for the \(\theta\)-tilted distribution, and for \(A \subseteq \mathbb{R}^p\) define

\[
\mu_\theta(A) = \int_{x \in A} dF_\theta(x).
\]

Then for any \(A\) such that \(\mu_\theta(A) > 0\),

\[
M_A(\theta) = M_0(\theta) \frac{\mu_\theta(A)}{\mu_\theta(A)} \tag{35}
\]

is the tilted representation of the MGF of the random vector \(X \mid X \in A\), i.e. \(X\) conditioned to lie in \(A\). Thus the tilted representation generalizes easily to the multivariate case. However, there are two significant problems which arise when using (35) as a basis for approximating \(M_A(\theta)\).

1. Generally \(\mu_\theta(A)\) will be difficult to approximate by saddlepoint methods in the multivariate case.

2. Even in cases where a convenient approximation is available for \(\mu_\theta(A)\), it may be unclear how to approximate \(M_A(\theta)\) for \(\theta\) outside \(\Theta\) in cases where \(M_A(\theta)\) has a strictly larger domain than \(M_0(\theta)\).

8.2 Convolution Representation

For simplicity we focus on the bivariate case \(p = 2\) and only consider absolutely continuous random vectors. The extension from \(p = 2\) to general \(p\) is straightforward conceptually but requires more elaborate notation. Similar results hold for multivariate lattice random vectors.
Another simplification is that we only consider the case in which the conditioning set $A$ is a semi-infinite rectangle. The extension to general rectangles is straightforward in principle but for general sets the convolution approach loses its simplicity.

Define
\[
\Xi_{jk}(\theta, y) = \frac{(-1)^{j+k}}{(2\pi l)^2} \int_{c_1-\infty}^{c_1+\infty} \int_{c_2-\infty}^{c_2+\infty} M_0(s) e^{(\theta_1-s_1) y_1} e^{(\theta_2-s_2) y_2} ds_1 ds_2
\]
where $s = (s_1, s_2)^T$, $\theta = (\theta_1, \theta_2)^T$, $y = (y_1, y_2)^T$; $\sigma_{jk} = 1$ if $j = 1$ and $k = 2$ or $j = 2$ and $k = 1$, and $\sigma_{jk} = 0$ if $j = k = 1$ or $j = k = 2$; and $\theta = (c_1, c_2)^T \in \Theta^o$ with $c_1 < \theta_1$ ($c_1 > \theta_1$) if $j = 1$ ($j = 2$), and $c_2 < \theta_2$ ($c_2 > \theta_2$) if $k = 1$ ($k = 2$). Note that $\Xi_{jk}$ is a natural generalization of $\Xi_1$ and $\Xi_2$ to the bivariate case. As in the univariate case,
\[
\Xi_{11}(\theta, y) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} e^\theta x f_0(x) dx,
\]
\[
\Xi_{21}(\theta, y) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} e^\theta x f_0(x) dx,
\]
with corresponding definitions for $\Xi_{12}(\theta, y)$ and $\Xi_{22}(\theta, y)$. Moreover, (8) extends to
\[
\Xi_{11}(\theta, y) + \Xi_{21}(\theta, y) + \Xi_{12}(\theta, y) + \Xi_{22}(\theta, y) = M_0(\theta) \quad \text{for all} \quad \theta \in \Theta^o.
\]
As in the univariate case, we can approximate the CGF of a random variable conditioned to lie in a semi-infinite rectangle by applying a saddlepoint approximation to the appropriate $\Xi_{jk}(\theta, y)$.

9 Acknowledgements

We are grateful to EPSRC for funding a Visiting Fellowship for the first author.

References


34