Saddlepoint Approximation for Multivariate CDF and Probability Computations in Sampling Theory

Ronald W. Butler & Richard K. Sutton

Colorado State University

Abstract

There are four multivariate distributions that commonly arise in sampling theory: the multinomial, multivariate hypergeometric, Dirichlet, and multivariate Pólya distributions. Second-order saddlepoint approximations are given for approximating these multivariate cumulative distribution functions (CDFs) in their most general settings. Probabilities of rectangular regions associated with these CDFs are also approximated directly using second-order saddlepoint methods. All the approximations follow from characterizations of the multivariate distributions as conditional distributions.

Keywords: MULTINOMIAL, MULTIVARIATE HYPERGEOMETRIC, DIRICHLET, MULTIVARIATE PÓLYA, SADDLEPOINT APPROXIMATION, SAMPLING DISTRIBUTIONS
1. Introduction

Suppose $X = (X_1, \ldots, X_k)$ has a $(k - 1)$-dimensional mass function, such as the multinomial, multivariate hypergeometric, or multivariate Pólya distribution. Furthermore, suppose that the distribution of $X$ can be characterized as follows: there exist independent variables $Y_t = (Y_{t1}, \ldots, Y_{tk})$ such that the distribution of $X$ is the same as the conditional distribution of $Y_t$ given $\sum_{i=1}^{k} Y_{ti} = n$, i.e., $X \overset{d}{=} Y_t | \sum_{i=1}^{k} Y_{ti} = n$. (Here $t$ designates a scaling parameter for the means of $\{Y_{ti}\}$.) Such characterization holds for the three distributions noted above. Under these circumstances, probabilities of rectangular regions for $X$, defined in vector notation as

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_k \leq X_k \leq b_k) = P(a \leq X \leq b),$$

can be computed in terms of the $Y_t$ variables. Using the conditional characterization of $X$,

$$P(a \leq X \leq b) = P\left(a \leq Y_t \leq b \mid \sum_{i=1}^{k} Y_{ti} = n\right)$$

$$= P\left(\sum_{i=1}^{k} Y_{ti} = n \mid a \leq Y_t \leq b\right) \frac{\prod_{i=1}^{k} P\left(a_i \leq Y_{ti} \leq b_i\right)}{P\left(\sum_{i=1}^{k} Y_{ti} = n\right)}$$

which follows from Bayes theorem.

The computations in (1) and (2) can be easily understood using a multinomial example for which $X \sim$ multinomial($n, p_1, \ldots, p_k$), with $\sum_{i=1}^{k} p_i = 1$. The multinomial is characterized in (1) as the conditional distribution of independent Poisson variates in which $Y_{ti} \sim$ Poisson$(tp_i)$. The factors $P(a_i \leq Y_{ti} \leq b_i)$ in (2) are interval probabilities for the Poisson$(tp_i)$ mass function, which can be computed with most any statistical package. The term $P\left(\sum_{i=1}^{k} Y_{ti} = n\right)$ is a Poisson$(t)$ probability, which is explicit. The only difficult computation involves the conditional mass function of $\sum_{i=1}^{k} Y_{ti}$ given $a \leq Y_t \leq b$. This mass function is a sum of $k$ independent Poisson variates, with $Y_{ti}$ truncated at $a_i$ below and $b_i$ above. As such, it can be readily
approximated using saddlepoint methods based on the cumulant generating functions (CGFs) of the truncated $Y_i$'s.

For these distributions, CDF approximation is the special case of rectangular probability approximation in which $a_i \equiv 0$ and $b_i < \infty$ for all $i$. This is the setting in which the multinomial CDF was approximated by Levin(1981) using the relationship in (2). However, instead of using a saddlepoint approximation of the first factor in (2), Levin used an Edgeworth expansion. The relationship between our saddlepoint approximation and his Edgeworth approximation is discussed in Section three. Generally the CDF approximation that uses saddlepoint approximations is more accurate.

Saddlepoint approximations for the CDFs of $M = \max_i X_i$ and $m = \min_i X_i$ when $\mathbf{X} \sim \text{Multinomial}(n, p_1, \ldots, p_k)$ are given in Good(1957). His approach differs from ours since it is based on the inversion of certain generating functions. However, the CDFs of $M$ and $m$ involve the specific rectangular regions of the form $[a, b]$ with $a = (a, \ldots, a)$ and $b = (b, \ldots, b)$ so that their approximation is a special case of our method. We show his approximations are analytically equivalent to ours in these specific instances.

Unlike most saddlepoint applications, the leading term in the saddlepoint density expansion, the so called first-order approximation, is not sufficiently accurate for practical use here. Inclusion of the second-order correction terms in the saddlepoint approximation is necessary to achieve a close approximation for practical use.

The approximation discussed above for the multinomial can be used more generally. As long as the characterization in (1) holds, Bayes theorem can be used to convert a multivariate probability into a product of three univariate probabilities. As seen from (2), two of these are straightforward calculations, while the third is readily approximated with a second-order
saddlepoint expansion. Consider for example, the multivariate hypergeometric distribution. In this case CDF approximations are based upon its characterization as the conditional distribution of independent binomial random variables. Also, the multivariate Pólya distribution is characterized as the conditional distribution of independent negative binomial random variables. Continuous $X$ such as the Dirichlet($\alpha_1, \ldots, \alpha_k$) follows the same development, with densities and integrals replacing probabilities and sums in (1). This distribution is characterized as a conditional distribution of independent gammas which sum to one. Expressions (1) and (2) have $n = 1$ and involve gamma densities and CDFs. The saddlepoint approximation is for the density of a sum of truncated gammas evaluated at $n = 1$.

In each of the four examples, rectangular probabilities as in (2) are approximated as

$$
\hat{P}_2 \left( \sum_{i=1}^k Y_{ti} = n \mid a \leq Y_t \leq b \right) = \frac{\prod_{i=1}^k P \left( a_i \leq Y_{ti} \leq b_i \right)}{P \left( \sum_{i=1}^k Y_{ti} = n \right)}
$$

(3)

where $\hat{P}_2 (\cdot)$ is a second-order mass/density saddlepoint approximation. The true expression in (2) is invariant to the value of $t$. Remarkably, the probability approximation in (3) is also invariant to the value of $t$ for each of the four distributions. In this sense (3) provides an adequate approximation for the exact value in (2).

The paper is organized as follows. Section two describes the four distributions to be approximated, the sampling framework in which they arise, and details of their conditional characterizations. Section three provides the saddlepoint expressions for approximating mass and density functions. This section also discusses previous approximations for the multinomial CDF due to Levin(1981) and Good(1957), with numerical examples. Section four develops approximations of rectangular probabilities for the multivariate hypergeometric, Dirichlet, and Pólya distributions, along with numerical examples.
2. Multivariate Sampling Distributions

Table 1 summarizes the four distributions for approximation. The columns of the table categorize the distributions according to the size of the population sampled, whether infinite or finite. The rows categorize the distributions according to the inferential framework, whether frequentist or Bayesian. Each cell in the table lists the distribution to be approximated, along with the independent variates that characterize it through conditioning.

The multivariate hypergeometric is the distribution of frequency counts when sampling from a $k$-category finite population. Suppose $n$ items are sampled without replacement from a population of size $N = \sum_{i=1}^{k} N_i$ in which $N_1, \ldots, N_k$ are the population frequency counts. If $\mathbf{X}$ is the vector of sampled frequency counts then $\mathbf{X} \sim \text{Multivariate Hypergeometric}(n;N_1,\ldots,N_k)$ with mass function

$$P(\mathbf{X} = \mathbf{x}) = \frac{\prod_{i=1}^{k} \binom{N_i}{x_i}}{\binom{N}{n}} \left(0 \leq x_i \leq N_i, \sum_{i=1}^{k} x_i = n\right).$$  \hspace{1cm} (4)

Its conditional characterization uses independent $Y_{ti} \sim \text{binomial}(N_i,t)$, with $t \in (0,1)$, so $\mathbf{X} \overset{d}{=} \mathbf{Y}_t \mid \sum_{i=1}^{k} Y_{ti} = n$.

A Bayesian perspective takes the components of $\mathbf{X}$ as parameters for the population proportions in the $k$ categories of an infinite population. The posterior distribution on $\mathbf{X}$ resulting from a random sample of this infinite population with a conjugate prior is Dirichlet($\alpha_1,\ldots,\alpha_k$) with density

$$f(\mathbf{x}) = \Gamma(\alpha) \prod_{i=1}^{k} x_i^{\alpha_i-1} \Gamma(\alpha_i) \left(0 \leq x_i, \sum_{i=1}^{k} x_i = 1\right)$$  \hspace{1cm} (5)

where $\alpha_i > 0$ and $\alpha = \sum_{i=1}^{k} \alpha_i$. A conditional characterization is $\mathbf{X} \overset{d}{=} \mathbf{Y}_t \mid \sum_{i=1}^{k} Y_{ti} = 1$, where $Y_{ti} \sim \text{Gamma}(\alpha_i,t)$ independently with $\alpha_i > 0, t > 0$, mean $\alpha_i/t$, and variance $\alpha_i/t^2$.
A Bayesian perspective when sampling from a finite population leads to a multivariate Pólya posterior on the remaining population frequency counts, denoted as $X$. This was developed in Jeffreys (1961, III, §§3.2, 3.2.3) when $k = 2$, with a more general framework given in Hill (1968). An argument leading to the multivariate Pólya posterior is given in the Appendix. For integer-valued $x$ the mass function of the Pólya$(\alpha_1, \ldots, \alpha_k)$ distribution is

$$P(X = x) = \prod_{i=1}^{k} \left\{ \frac{\Gamma(x_i + \alpha_i)}{\Gamma(\alpha_i) x_i!} \right\} \left\{ \frac{\Gamma(n + \alpha)}{\Gamma(\alpha) n!} \right\} \quad (0 \leq x_i, \sum_{i=1}^{k} x_i = n) \quad (6)$$

for $\alpha_i > 0$, and $\alpha = \sum_{i=1}^{k} \alpha_i$. Its conditional characterization is $X \overset{d}{=} Y_t \mid \sum_{i=1}^{k} Y_{t_i} = n$, where $Y_{t_i} \sim \text{Negative Binomial} (\alpha_i, t)$ independently, with $\alpha_i > 0$, $t \in (0, 1)$, and mean $\alpha_i (1 - t)/t$. The multivariate Pólya is also known as the negative multivariate hypergeometric and the multinomial-Dirichlet. Under the latter name this distribution arises as the Bayes predictive density for a future sample from an infinite population.

3. Saddlepoint Approximation

3.1. General Methodology

Denoting $T_{t_i}$ as $Y_{t_i}$ restricted to $[a_i, \ldots, b_i]$, consider saddlepoint approximation for

$$P \left( \sum_{i=1}^{k} Y_{t_i} = n \mid a \leq Y_t \leq b \right) = P(T_t = n)$$

with $T_t = \sum_{i=1}^{k} T_{t_i}$. If $K(s, t)$ is the CGF of $T_t$ evaluated at $s$, then the first-order saddlepoint approximation to $P(T_t = n)$ is

$$\hat{P}_1(T_t = n) = \left\{ 2\pi K''(\hat{s}, t) \right\}^{-\frac{1}{2}} \exp \{ K(\hat{s}, t) - \hat{s}n \} \quad (7)$$
where saddlepoint \( \hat{s} \) is the unique solution to \( \partial K(\hat{s}, t)/\partial \hat{s} = K'(\hat{s}, t) = n \). A unique saddlepoint solution is guaranteed for \( n \in (\sum_{i=1}^{k} a_i, \sum_{i=1}^{k} b_i) \), which is the interior of the convex hull of the support for \( T_t \) as discussed in Section five.

The second-order approximation is

\[
\hat{P}_2(T_t = n) = \hat{P}_1(T_t = n) \left\{ 1 + \frac{1}{8} \frac{K^{(4)}(\hat{s}, t)}{K''(\hat{s}, t)^2} + \frac{5}{24} \frac{K^{(3)}(\hat{s}, t)^2}{K''(\hat{s}, t)^3} \right\},
\]

as given in Daniels(1954, eqn 2.6).

Consider the multinomial probability approximation, for example. The CGF of \( T_{ii} \), a Poisson restricted to \([a_i, \ldots, b_i]\), is

\[
K_i(s, t) = \log \left\{ \sum_{j=a_i}^{b_i} \exp \left( sj - tp_i \right) \left( tp_i \right)^j / j! \right\} - \log \left\{ \sum_{j=a_i}^{b_i} \exp \left( -tp_i \right) \left( tp_i \right)^j / j! \right\} \quad (-\infty < s < \infty)
\]

so \( K(s, t) = \sum_{i=1}^{k} K_i(s, t) \) is defined for \( s \in (-\infty, \infty) \).

In our probability approximations, the first-order approximation in (7) does not usually give sufficient accuracy whereas the second-order approximation in (8) does. For example, for an equiprobable, 12-dimensional multinomial, \( P(\max_i X_i > 3) = .16296 \). The first-order saddlepoint approximation yields .15308, and achieves a relative error of \(-6.1\%\). The second-order approximation gives .16291, with a relative error of \(-.027\%\).

Renormalizing the first-order saddlepoint approximation substantially improves the first-order approximation but is still not as accurate as the second-order approximation. With \( a = \sum_{i=1}^{k} a_i \) and \( b = \sum_{i=1}^{k} b_i \), a renormalized first-order approximation is computed as

\[
\hat{P}_R(T_t = n) = \hat{P}(T_t = n) / P(T_t = a) + \sum_{i=a+1}^{b} \hat{P}_1(T_t = i) + P(T_t = b).
\]

The terms \( P(T_t = a) \) and \( P(T_t = b) \) are exact values which must be used since the saddlepoint equation cannot be solved on the boundary of \( T_t \)'s support. For the 12-dimensional example
above, the renormalized, first-order approximation is .16373, with a relative error of .48%. This is not only less accurate than the second-order approximation, but also required 15 times more CPU time for its computation. Accuracy and speed favor the second-order saddlepoint approximation over both the first-order and first-order, renormalized approximations. We shall therefore consider only the second-order approximation and hereafter refer to such as simply the saddlepoint approximation.

Levin (1981) also obtained an approximation for the multinomial CDF using the characterization in (1). To approximate $P(T_i = n)$ he used a second-order Edgeworth expansion,

$$
\hat{P}_e(T_i = n) = \left\{2\pi K''(0, t)\right\}^{-\frac{1}{2}} \exp \left(-\frac{z^2}{2}\right) \left\{1 + \frac{1}{6} \frac{K^{(3)}(0, t)}{K''(0, t)^{3/2}} H_3(z) + \frac{1}{24} \frac{K^{(4)}(0, t)}{K''(0, t)^2} H_4(z) + \frac{1}{72} \frac{K^{(3)}(0, t)^2}{K''(0, t)^2} H_5(z)\right\},
$$

(10)

where $z = \left\{n - K'(0, t)\right\}/\sqrt{K''(0, t)}$ is the standardized value of $n$ and $H_3(\cdot), H_4(\cdot),$ and $H_5(\cdot)$ are Hermite polynomials. When (10) is used in place of the true value $P(T_i = n)$ in (2), an Edgeworth-based approximation results which, unlike (3), is dependent on the particular value of $t$. A choice of its value is therefore necessary. A poor choice can lead to probability estimates less than zero or greater than one. Levin set $t = n$, and justified this choice on the numerical accuracy achieved. We discuss his choice as well as other choices of $t$ below.

Approximation (10) gives a class of Edgeworth expansions indexed by $t$. A member of this class, indexed by $\hat{t}$, agrees with the saddlepoint approximation in (8) taken with $t = \hat{t}$ as we now demonstrate. The most accurate member of the class in (10) for approximating its distribution is the one in which $z = 0$, since Edgeworth expansions are known to deteriorate in accuracy as $|z|$ departs from zero (Barndorff-Nielsen and Cox (1989), p 104). Thus the $\hat{t}$ solving $z = 0$ or $n = K'(0, \hat{t})$ indexes the most accurate member. The existence of such a $\hat{t}$ is
shown in Section 6. With this the associated saddlepoint is \( \hat{s} = 0 \); the common value of (10) with \( t = \hat{t} \) and (8) with \( t = \hat{t} \) and \( \hat{s} = 0 \) is

\[
\left\{ 2\pi K''(0, \hat{t}) \right\}^{-\frac{1}{2}} \left\{ 1 + \frac{1}{8} \frac{K^{(4)}(0, \hat{t})}{K''(0, \hat{t})^2} + \frac{5}{24} \frac{K^{(3)}(0, \hat{t})^2}{K''(0, \hat{t})^3} \right\}
\]

since \( H_3(0) = 0, H_4(0) = 3 \), and \( H_6(0) = -15 \). This shows that the Edgeworth approximation indexed by \( \hat{t} \) agrees with the saddlepoint approximation in (3) with \( t = \hat{t} \). Thus the Edgeworth-based probability approximation with index \( \hat{t} \) is the same as the approximation in (3).

3.2. Numerical Results for Multinomial

Table 2 shows the accuracy of the rectangular probability approximation in (3) (2nd-order sp) when based on (8). This is compared with Levin’s suggestion (Edgeworth) which essentially substitutes (10) for (8) in the approximation (3) and uses \( t = n \). The column labeled “exact/simulated” supplies exact values of the probability when practical, or estimates based on hit-and-miss simulation. The simulations used a sufficient number of replications to yield an acceptably narrow 95% normal confidence interval halfwidth as given in parenthesis. Exact calculations and hit-and-miss simulations took minutes to days, while the saddlepoint and Edgeworth approximations took a few seconds. Superscripts are used to indicate repetition: \( (1/12)^3 = (1/12, 1/12, 1/12) \) and \( .0^{37} = .007 \). Column ”k” gives the dimension of the multinomial vector. Column ”p” defines the probabilities of the \( k \) categories. The relative errors of these probabilities are based on the lesser of the quantities \( P(a \leq X \leq b) \) and \( 1 - P(a \leq X \leq b) \). Over a wide range of dimensions, sample sizes, and values of \( p \), both the saddlepoint approximation and the Edgeworth expansion are accurate. However, more often than not, the saddlepoint approximation is more accurate, attaining relative errors on
the order of ten times smaller in four of the examples. In the fourth example, the inaccuracy of the simulated probability is caused by too few repetitions in the simulation. One billion repetitions used in the simulation were not enough to accurately estimate the probability, but more repetitions would take an impractical amount of time.

Levin's recommendation of taking \( t = n \) can be related to the saddlepoint approximation in the following manner. Suppose there is no truncation in the Poisson variates. Then the saddlepoint equation becomes \( n = E \left( \sum_{i=1}^{n} Y_i \right) = t \), which is Levin's choice. One can therefore expect that Levin's use of the Edgeworth-based approximation with \( t = n \) is reasonably accurate when the truncation of probability is not severe. Table 3 shows the effect of increasing truncation on both the saddlepoint approximation and the Edgeworth-based approximation with \( t = n \). Following the example in the second row of Table 2, we calculate \( P(0 < X < 3) \) where \( X \) is an equiprobable, 12-dimensional multinomial, and investigate the effect of truncation with increasing \( n \). For \( n = 25 \), each Poisson has a mean of 2.083, with 16% of the mass residing outside of \( \{0, 1, 2, 3\} \). For \( n = 35 \), the means are 2.917, with 33% of the mass excluded. Thus, the amount of truncation increases with \( n \). This also is reflected in a \( z \)-value which increases with \( n \) as seen in the last column. Note the saddlepoint-based approximation maintains a small relative error, while the Edgeworth-based approximation deteriorates in accuracy for larger \( n \), as one might expect.

3.3. Connections with Good(1957)

Good(1957) also used saddlepoint approximations to approximate CDFs for maxima and minima of components of the multinomial distribution. Good developed his approximations from an entirely different point of view which involves the inversion of certain generating functions.
His approximations are analytically identical to ours, taken with \( t = 1 \) as indicated below.

For \( X = (X_1, \ldots, X_k) \sim \text{Multinomial}(n, p_1, \ldots, p_k) \), let

\[
g_n = P \left( \max_i X_i \leq b \right) = P \left( 0 \leq X \leq b \right)
\]  

(11)

where \( b = (b, \ldots, b) \), so (11) is a special case of a rectangular probability. Good gives the generating function of \( \{q_n/n!\} \) as

\[
g(x) = \sum_{n=0}^{\infty} \frac{q_n}{n!} x^n = \prod_{i=1}^{k} \left\{ 1 + p_i x + \ldots + \left( p_i x \right)^b / b! \right\} \quad (-\infty < x < \infty).
\]  

(12)

Approximation of the coefficients in \( \{q_n/n!\} \) is obtained from saddlepoint inversion of (12) in the following manner. The ratio \( g(x)/g(1) \) is a probability generating function (PGF) for some random variable \( G \), so the saddlepoint approximation of its \( n \)th term is

\[
\frac{\hat{q}_n}{n! g(1)} = \hat{P}_2(G = n)
\]  

(13)

where \( \hat{P}_2(G = n) \) is a second-order saddlepoint approximation as given in (8). The expression for \( \hat{q}_n \) in (13) is essentially Good’s approximation.

To show the equivalence of \( \hat{q}_n \) in (13) and approximation (3) with \( t = 1 \), compute the CGF of \( G \) as

\[
K_G(s) = \log \left[ g \{ \exp(s) \} / g(1) \right] = \sum_{i=1}^{k} \left[ \log \left\{ \sum_{j=0}^{b_i} p_i^j \exp(j s) / j! \right\} - \log \left\{ \sum_{j=0}^{b_i} p_i^j / j! \right\} \right]
\]  

(14)

which is \( K(s, 1) \). Hence, \( \hat{P}_2(G = n) = \hat{P}_2(T_1 = n) \), the lead term in approximation (3) with \( t = 1 \). A simple calculation also shows that

\[
n! g(1) = \prod_{i=1}^{k} \frac{P \left( a_i \leq Y_{1i} \leq b_i \right)}{P \left( \sum_{i=1}^{k} Y_{1i} = n \right)}
\]

so \( \hat{q}_n \) is approximation (3) evaluated with \( t = 1 \). Recall that (3) does not depend on the actual value of \( t \).
Good also gave an approximation for

\[ r_n = P\left( \min_i X_i \geq a \right) = P(a \leq X) \]

where \( a = (a, \ldots, a) \). His approximation is analytically the same as ours. His approach is again based on saddlepoint inversion of the generating function for \( \{r_n/n!\} \), given as

\[ h(x) = \prod_{i=1}^{k} \left\{ (p_i x)^a / a! + (p_i x)^{a+1} / (a + 1)! + \ldots \right\}. \]

Using the same argument as for the maximum, we create random variable \( H \) with PGF \( h(x)/h(1) \) and obtain

\[ r_n = \hat{P}_2(H = n)n!h(1). \quad (15) \]

as Good's approximation. Simple algebra shows \( \hat{P}_2(H = n) = P_2(T_t = n) \) with \( t = 1 \) and that (15) agrees with (3) with \( t = 1 \).

4. Other Distributions

The other three distributions are approximated below. Saddlepoint-based probability approximations are compared with exact/simulated calculations, as well as Edgeworth-based approximations. For each distribution, the saddlepoint approximation in (3) is invariant to the choice of scale parameter \( t \). The arguments for showing this are similar to those used in the multinomial example in Section 6, and are therefore omitted. For each distribution, there is also a class of Edgeworth approximations for \( P(T_t = n) \) indexed by \( t \). From this class an appropriately indexed member \( \hat{t} \) (for which \( z = 0 \)) agrees with the saddlepoint approximation \( \hat{P}_2(T_t = n) \). Arguments for the existence of \( \hat{t} \) are the same as those used in the multinomial
example. Thus the Edgeworth-based probability approximation indexed by \( \hat{t} \) agrees with the saddlepoint-based approximation in (3) for each of the distributions.

4.1. Multivariate Hypergeometric \((n, N_1, \ldots, N_k)\)

This distribution, given in (4), requires approximation of the mass function of \( T_i = \sum_{i=1}^{k} T_{ni} \) when \( \{T_i\} \) are truncated binomial variables. The CGF for \( T_{ni} \) is

\[
K_i(s,t) = \log \left\{ \sum_{j=a_i}^{b_i} \binom{N_i}{j} \exp(s_j) t^j (1-t)^{N_i-j} \right\} - \log \left\{ \sum_{j=a_i}^{b_i} \binom{N_i}{j} t^j (1-t)^{N_i-j} \right\},
\]

\( K(s,t) = \sum_{i=1}^{k} K_i(s,t) \), and the computations follow from (8). These computations are displayed in Table 4, using the format of Table 2. Column "N" lists the vector of initial frequency counts \((N_1, \ldots, N_k)\). For the Edgeworth-based approximation, \( t \) was chosen to be \( n/N \), so the sum of untruncated binomials has a mean of \( n \). The saddlepoint-based and the Edgeworth-based approximations are accurate in a variety of settings, with the saddlepoint-based approximation typically more accurate. The Edgeworth-based approximation again fails when the truncation is severe. This is illustrated in the third row, where it attains a negative probability estimate. By contrast the saddlepoint-based approximation maintains extraordinarily small relative error.

4.2. Multivariate Pólya \((\alpha_1, \ldots, \alpha_k)\)

Truncated negative binomials \( \{T_{ni}\} \) are involved in approximating these probabilities. The CGF of \( T_{ni} \) is

\[
K_i(s,t) = \log \left\{ \sum_{j=a_i}^{b_i} \binom{\alpha_i + j - 1}{j} \exp(s_j)(1-t)^j \right\} - \log \left\{ \sum_{j=a_i}^{b_i} \binom{\alpha_i + j - 1}{j} t^j (1-t)^j \right\}.
\]
Table 5 investigates the accuracy of rectangular probability approximations. Column \( \alpha \) gives values of the parameters \((\alpha_1, \ldots, \alpha_k)\) in the multivariate Pólya mass function given in (6). Saddlepoint approximations usually outperform the Edgeworth-based approximations and sometimes substantially as seen in the first row. The Edgeworth-based approximation used \( t = \alpha/(\alpha + n) \) with \( \alpha = \sum_{i=1}^{k} \alpha_i \), so the mean of the sum of untruncated negative binomial variates is \( n \).

### 4.3. Dirichlet \((\alpha_1, \ldots, \alpha_k)\)

Dirichlet probability approximations are based on truncated gamma random variables \( \{T_{hi}\} \). The CGF for \( T_{hi} \) is

\[
K_i(s, t) = \log \left[ \frac{\Gamma(\alpha_i, t b_i) u^\alpha - 1 \exp \{- (t - s) u\} du}{\Gamma(\alpha, t b_i) - \Gamma(\alpha, t a_i)} \right] \tag{16}
\]

\[
= \log \left[ \frac{\Gamma \{\alpha_i, (t - s) b_i\} - \Gamma \{\alpha_i, (t - s) a_i\}}{\Gamma(\alpha_i, t b_i) - \Gamma(\alpha_i, t a_i)} \right] + \alpha_i \log \left( \frac{t}{t - s} \right) \tag{17}
\]

where \( \Gamma(\alpha_i, x) \) is the incomplete gamma function defined as \( \Gamma(\alpha, x) = \int_{0}^{x} u^\alpha - 1 \exp \{-u\} du \).

The components of \( b \) are at most one in Dirichlet probability computations. For \( 0 < b_i \leq 1 \), the CGF converges for all \( s \) as seen in (16). Expression (17) can be used for CGF computation when \( s < t \). If \( s > t \), then the integral in (16) can be evaluated by repeated integration by parts, yielding

\[
\sum_{j=0}^{\lfloor \alpha_i \rfloor - 2} (-1)^j (s - t)^{-j - 1} \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i - j)} \left[ b_i^{\alpha_i - 1 - j} \exp\{(s - t)b_i\} - a_i^{\alpha_i - 1 - j} \exp\{(s - t)a_i\} \right] + (t - s)^{-\lfloor \alpha_i \rfloor - 1} \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i - \lfloor \alpha_i \rfloor + 1)} \int_{a_i}^{b_i} u^{\alpha_i - \lfloor \alpha_i \rfloor} \exp\{(s - t)u\} du \tag{18}
\]

where \( \lfloor \alpha_i \rfloor \) is the integer portion of \( \alpha_i \). The final integral in (18) with \( u \) raised to the fractional portion of \( \alpha_i \) can be approximated by Taylor-series expansion of the exponential followed by
integration term by term. Direct Taylor-series expansion of the exponential in (16), without first reducing the power of $u$ with integration by parts, results in considerably more numerical error.

Table 6 gives probability approximations for the Dirichlet distribution. Column $\alpha$ is the vector of parameters $(\alpha_1, \ldots, \alpha_k)$. For the Edgeworth-based approximation, the scale parameter $t$ was set to $\sum_{i=1}^{k} \alpha_i$ so the untruncated sum of gammas has a mean of one. Relative errors for the saddlepoint-based approximations are considerably smaller than those of the Edgeworth-based approximations. For example, in row five the saddlepoint-based relative error is over 100 times smaller.

5. Miscellaneous Comments

All of our examples deal with the nontrivial setting of (1) which occurs when $\sum_{i=1}^{k} a_i = a < n < b = \sum_{i=1}^{k} b_i$. When this occurs the probability computation entails a summation over more than a single point in the support of $X$. When either $a = n$ or $b = n$, both sides of (1) reduce to the mass function of $X$ at $a$ or $b$ respectively since only one $k$-tuple of $Y_i$ will satisfy both the inequalities and the constraint $\sum_{i=1}^{k} Y_{ti} = n$. If either $a > n$ or $b < n$ then both sides of (1) are zero. The support of $T_i$ therefore ranges from $a$ to $b$. 

15
6. Proofs

6.1. Solution to $K'(0, \hat{\tau}) = n$

To show that a solution exists in the multinomial case we first show that $\lim_{t \to 0} K'(0, t) = a < b = \lim_{t \to \infty} K'(0, t)$. A solution must then exist by the continuity of $K'(0, t)$ in $t$. Solution to this equation is required for showing the equivalence of the $\hat{\ell}$-indexed Edgeworth and saddlepoint approximations.

Differentiate $K_i(s, t)$ with respect to $s$ to get

$$K'_i(s, t) = \frac{\sum_{j=a_i}^{b_i} j \exp(sj)(tp_i)^j / j!}{\sum_{j=a_i}^{b_i} \exp(sj)(tp_i)^j / j!}$$

(19)

Factoring out the leading terms in both the numerator and denominator and cancelling gives

$$K'_i(s, t) = a_i \left[ \frac{1 + \sum_{j=a_i+1}^{b_i} \{\exp(s)tp_i\}^{(j-a_i)} (a_i - 1)! / (j - 1)!}{1 + \sum_{j=a_i+1}^{b_i} \{\exp(s)tp_i\}^{(j-a_i)} a_i! / j!} \right]$$

(20)

$$a_i \geq 1$$

$$= \frac{0 + \sum_{j=1}^{b_i} \{\exp(s)tp_i\}^{j} / (j - 1)!}{1 + \sum_{j=1}^{b_i} \{\exp(s)tp_i\}^{j} / j!}$$

(21)

$$a_i = 0.$$

From (20) and (21), $\lim_{t \to 0} K'_i(0, t) = a_i$ so $\lim_{t \to 0} K'(0, t) = a$.

Similarly, for $t \to \infty$ and $b_i < \infty$, we obtain $b_i$ in the limit by factoring out the $b_i$ term from the sums in both the numerator and denominator, so

$$K'_i(s, t) = b_i \left[ \frac{1 + \sum_{j=a_i}^{b_i-1} \{\exp(s)p_{i\hat{\tau}}\}^{(j-b_i)} (b_i - 1)! / (j - 1)!}{1 + \sum_{j=a_i}^{b_i-1} \{\exp(s)p_{i\hat{\tau}}\}^{(j-b_i)} b_i! / j!} \right],$$

for which $\lim_{t \to \infty} K'_i(0, t) = b_i$.

For $t \to \infty$ and $b_i = \infty$, write $K'(s, t)$ in (19) as

$$K'_i(s, t) = \frac{\left\{ \sum_{j=0}^{\infty} j \exp(-tp_i) \exp(sj)(tp_i)^j / j! \right\} - \left\{ \sum_{j=0}^{a_i-1} j \exp(-tp_i) \exp(sj)(tp_i)^j / j! \right\}}{\sum_{j=a_i}^{\infty} \exp(-tp_i)(tp_i)^j / j!}$$
\[ \sum_{j=0}^{\infty} j \exp(-tp_i) \exp(sj)(tp_i)^j / j! \]
\[ = \exp\left[t p_i \left(\exp(s) - 1\right) + s\right] t p_i. \]

As \( t \to \infty, K'_i(0,t) \to \infty = b_i \). Thus \( \lim_{t \to \infty} K'(0,t) = b \), so \( K'(0,t) \) can attain any value in \((a, b)\), which includes the value \( n \).

### 6.2. Invariance of \( \hat{P}_2 (a \leq X \leq b) \) to the value of \( t \).

First, select values \( t_1 \) and \( t_2 \) and compare their respective saddlepoints \( \hat{s}_1 \) and \( \hat{s}_2 \) that solve

\[ n = K'(\hat{s}, t) = \sum_{i=1}^{b_i} \log \left( \sum_{j=a_i}^{b_i} j \exp(\hat{s}j)(tp_i)^j / j! \right) - \log \left( \sum_{j=a_i}^{b_i} \exp(\hat{s}j)(tp_i)^j / j! \right) \]

for \( t = t_1 \) and \( t = t_2 \). From the form of (22), if \( n = K'(\hat{s}_1, t_1) \) then \( n = K'(\hat{s}_2, t_2) \) when \( \hat{s}_2 = \hat{s}_1 + \log t_1 / t_2 \). Uniqueness in the solution to the saddlepoint equation guarantees that such \( \hat{s}_2 \) is the only solution to \( n = K'(\hat{s}_2, t_2) \). This therefore characterizes the relationship between \( \hat{s}_1 \) and \( \hat{s}_2 \).

With \( K'(\hat{s}_1, t_1) = K'(\hat{s}_2, t_2) \) and \( d\hat{s}_1 / d\hat{s}_2 = 1 \), all higher derivatives of \( K(s,t) \) evaluated at \((\hat{s}_i, t_i)\) will be invariant to index \( i \) so

\[ K^{(j)}(\hat{s}_1, t_1) = K^{(j)}(\hat{s}_2, t_2) \quad (j \geq 1). \]

This means that we can ignore the factor \( \{2\pi K''(\hat{s}, t)\}^{-\frac{1}{2}} \) and the second-order correction term in the saddlepoint approximation. The remaining terms in (3) include

\[ A(\hat{s}, t) = \prod_{i=1}^{k} \frac{P(a_i \leq Y_{ti} \leq b_i)}{P(\sum_{i=1}^{k} Y_{ti} = n)} \exp\left\{ \sum_{j=0}^{\infty} \frac{j \exp(-tp_i) \exp(sj)(tp_i)^j / j!}{\sum_{j=a_i}^{b_i} \exp(\hat{s}j)(tp_i)^j / j!} \right\} \]

To show (23) is invariant to \( t \) substitute

\[ K(\hat{s}_1, t_1) = K(\hat{s}_2, t_2) + \sum_{i=1}^{k} \left\{ \sum_{j=a_i}^{b_i} \frac{(t_2 p_i)^j / j!}{(t_1 p_i)^j / j!} \right\} \]

17
\[ \hat{s}_1 = \hat{s}_2 + \log \left( \frac{t_2}{t_1} \right) \]  

(24)

into the expression for \( A(\hat{s}_1, t_1) \) given in (23), so

\[
A(\hat{s}_1, t_1) = \frac{\prod_{i=1}^{k} \sum_{j=a_i}^{b_i} \frac{(t_1 p_i)^j}{j!}}{t_1^n / n!} \exp \left\{ \sum_{i=1}^{k} \left[ \frac{\sum_{j=a_i}^{b_i} \frac{(t_2 p_i)^j}{j!}}{t_1^n / n!} - n \log \frac{t_2}{t_1} \right] \right\} \exp \{ K(\hat{s}_2, t_2) - s_2 n \} 
\]

\[
= \frac{\prod_{i=1}^{k} \sum_{j=a_i}^{b_i} \frac{(t_2 p_i)^j}{j!}}{t_2^n / n!} \exp \{ K(\hat{s}_2, t_2) - \hat{s}_2 n \} 
\]

\[ = A(\hat{s}_2, t_2). \]

Similar proofs of invariance hold for the multivariate hypergeometric, Dirichlet, and multivariate Pólya CDF approximations. The technique of proof is the same in each case: re-express \( \hat{s}_1 \) and \( K(\hat{s}_1, t_1) \) in terms of \( \hat{s}_2 \) and \( K(\hat{s}_2, t_2) \). The saddlepoint relationships arising with the multivariate hypergeometric, Dirichlet, and multivariate Pólya are \( \hat{s}_1 = \log \frac{t_2(1-t_1)}{t_2(1-t_1) + \hat{s}_2}, \)

\[ \hat{s}_1 = \hat{s}_2 - t_2 + t_1, \]

and \( \hat{s}_1 = \log \frac{1-t_2}{1-t_1} + \hat{s}_2 \) respectively.

**Appendix**

The multivariate Pólya is the conjugate prior-posterior distribution associated with random finite-population sampling as we now show. Suppose \( N \) is the known population size and \( N = (N_1, \ldots, N_k) \) are population frequencies with a Pólya\( (N, \alpha_{10}, \ldots, \alpha_{k0}) \) prior, so that

\[ \pi(N) \propto \prod_{i=1}^{k} \frac{\Gamma(N_i + \alpha_{i0})}{\Gamma(N_i + 1)} \left( 0 \leq N_i, \sum_{i=1}^{k} N_i = N \right). \]

If a random sample of \( n \) items yields sample frequencies \( n = (n_1, \ldots, n_k) \) with multivariate hypergeometric likelihood

\[ P(n \mid N) \propto \prod_{i=1}^{k} \frac{\Gamma(N_i + 1)}{\Gamma(N_i + 1 - n_i)}, \]

18
then the posterior on \( M = N - \mathbf{n} \) is

\[
P(M = (m_1, \ldots, m_k) \mid \mathbf{n}) \propto \prod_{i=1}^{k} \frac{\Gamma(m_i + n_i + \alpha_{i0})}{\Gamma(m_i + 1)}
\]

which is Pólya\((N - n, n_1 + \alpha_{10}, \ldots, n_k + \alpha_{k0})\), as defined in (6).
REFERENCES


Table 1: CDF of $X$ | Distribution of $\{Y_i\}$

<table>
<thead>
<tr>
<th>Approach</th>
<th>infinite</th>
<th>finite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequentist</td>
<td>Multinomial</td>
<td>Multivariate Hypergeometric</td>
</tr>
<tr>
<td></td>
<td>Poissons</td>
<td></td>
</tr>
<tr>
<td>Bayesian</td>
<td>Dirichlet</td>
<td>Multivariate Pólya</td>
</tr>
<tr>
<td></td>
<td>Gammas</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Multinomial Examples

<table>
<thead>
<tr>
<th>k</th>
<th>p</th>
<th>n</th>
<th>a</th>
<th>b</th>
<th>exact/simulated</th>
<th>2nd-order sp</th>
<th>Edgeworth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(95% Confidence Interval)</td>
<td>% Relative Error</td>
<td>% Relative Error</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(1/8)^8</td>
<td>133</td>
<td>(0)^8</td>
<td>.94195</td>
<td>.94198</td>
<td>.94213</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(26)^8</td>
<td></td>
<td>(-.0450)</td>
<td>(-.309)</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>(1/12)^12</td>
<td>12</td>
<td>(0)^12</td>
<td>.8370</td>
<td>.8371</td>
<td>.8367</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3)^12</td>
<td></td>
<td>(.0270)</td>
<td>(.205)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(\frac{2}{420}, \frac{4}{420}, \frac{4}{420})</td>
<td>20</td>
<td>(0)^20</td>
<td>.9653</td>
<td>.9656</td>
<td>.9655</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5)^20</td>
<td>(±.03959)</td>
<td>(-.727)</td>
<td>(-.548)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(\frac{2}{420}, \frac{4}{420}, \frac{4}{420})</td>
<td>60</td>
<td>(2)^20</td>
<td>.06584</td>
<td>.06614</td>
<td>.06617</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\infty)^20</td>
<td>(±.07474)</td>
<td>(5.09)</td>
<td>(5.59)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>(1/20)^20</td>
<td>200</td>
<td>(0)^20</td>
<td>.021648</td>
<td>.021646</td>
<td>.021628</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(13)^20</td>
<td>(±.0429)</td>
<td>(-.02974)</td>
<td>(-.0910)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>(1/50)^50</td>
<td>500</td>
<td>(0)^50</td>
<td>.852919</td>
<td>.852726</td>
<td>.852732</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(19)^50</td>
<td>(±.02197)</td>
<td>(.131)</td>
<td>(.127)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>(1/50)^50</td>
<td>500</td>
<td>(3)^50</td>
<td>.87695</td>
<td>.877372</td>
<td>.877375</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(\infty)^50</td>
<td>(±.0464)</td>
<td>(-.343)</td>
<td>(-.345)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>(1/50)^50</td>
<td>300</td>
<td>(3)^50</td>
<td>.030325</td>
<td>.030323</td>
<td>.030321</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(15)^50</td>
<td>(±.0434)</td>
<td>(-.0471)</td>
<td>(-.0129)</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>(\frac{2}{40200}, \frac{4}{40200}, \frac{4}{40200})</td>
<td>300</td>
<td>(0)^200</td>
<td>.0235320</td>
<td>.0235736</td>
<td>.0235737</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5)^200</td>
<td>(±.04297)</td>
<td>(.177)</td>
<td>(.177)</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>exact</td>
<td>2nd-order sp, (% Relative Error)</td>
<td>Edgeworth, (% Relative Error)</td>
<td>z</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.050606</td>
<td>0.050609 (0.02593)</td>
<td>0.050528 (-0.154)</td>
<td>1.627</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.020912</td>
<td>0.020913 (0.02478)</td>
<td>0.021228 (1.51)</td>
<td>2.515</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>0.05289</td>
<td>0.05288 (-0.591)</td>
<td>0.05053 (-81.6)</td>
<td>3.591</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Multivariate Hypergeometric Examples

<table>
<thead>
<tr>
<th>k</th>
<th>N</th>
<th>n</th>
<th>a</th>
<th>b</th>
<th>exact/simulated (95% Confidence Interval)</th>
<th>2nd-order sp (% Relative Error)</th>
<th>Edgeworth (% Relative Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>(20,15,15)</td>
<td>10</td>
<td>(0)^3</td>
<td>(6,5,4)</td>
<td>.8086</td>
<td>.8095 (-.463)</td>
<td>.8122 (-1.86)</td>
</tr>
<tr>
<td>3</td>
<td>(20,15,15)</td>
<td>10</td>
<td>(1)^3</td>
<td>(6,5,4)</td>
<td>.794</td>
<td>.795 (-.417)</td>
<td>.796 (-.946)</td>
</tr>
<tr>
<td>8</td>
<td>(30)^8</td>
<td>133</td>
<td>(0)^8</td>
<td>(2^9,26^5)</td>
<td>.03623449</td>
<td>.03623451 (.03853)</td>
<td>-.0346 (-25,000)</td>
</tr>
<tr>
<td>10</td>
<td>(10)^10</td>
<td>35</td>
<td>(0)^10</td>
<td>(4)^10</td>
<td>.0298</td>
<td>.0299 (1.03)</td>
<td>.0100 (1.51)</td>
</tr>
<tr>
<td>10</td>
<td>(10)^10</td>
<td>35</td>
<td>(0)^10</td>
<td>(7)^10</td>
<td>.96911</td>
<td>.96906 (.143)</td>
<td>.96909 (.0653)</td>
</tr>
<tr>
<td>10</td>
<td>(10)^10</td>
<td>35</td>
<td>(3)^10</td>
<td>(7)^10</td>
<td>.010616</td>
<td>.010622 (.0565)</td>
<td>.010638 (.207)</td>
</tr>
<tr>
<td>20</td>
<td>(15)^20</td>
<td>40</td>
<td>(1)^20</td>
<td>(∞)^20</td>
<td>.06031</td>
<td>.06033 (.0362)</td>
<td>.06028 (-.0475)</td>
</tr>
<tr>
<td>50</td>
<td>(10,20, ...)</td>
<td>2000</td>
<td>(0)</td>
<td>(2,4, ..., 100)</td>
<td>.061621</td>
<td>.062036 (.674)</td>
<td>.062041 (.682)</td>
</tr>
<tr>
<td>k</td>
<td>α</td>
<td>a</td>
<td>b</td>
<td>sample size</td>
<td>exact/simulated 95% Confidence Interval</td>
<td>2nd-order (% Relative Error)</td>
<td>Edgeworth (% Relative Error)</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>-------</td>
<td>-------------</td>
<td>-----------------------------------------</td>
<td>------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>3</td>
<td>(5, 4, 4)</td>
<td>(0)^3</td>
<td>(12, 30, 3)</td>
<td>40</td>
<td>.0^2382324</td>
<td>.0^2372436</td>
<td>(-2.59)</td>
</tr>
<tr>
<td>10</td>
<td>(1, 2, ..., 10)</td>
<td>(0)^10</td>
<td>(1, 2, ..., 10)</td>
<td>50</td>
<td>.0^310139</td>
<td>.0^310136</td>
<td>(.0298)</td>
</tr>
<tr>
<td>10</td>
<td>(1, 2, ..., 10)</td>
<td>(0)^10</td>
<td>(4, 8, ..., 40)</td>
<td>50</td>
<td>.029914</td>
<td>.0300199</td>
<td>(.0300249)</td>
</tr>
<tr>
<td>20</td>
<td>(2)^20</td>
<td>(2)^20</td>
<td>(∞)^20</td>
<td>145</td>
<td>.0691</td>
<td>.0689</td>
<td>.0688</td>
</tr>
<tr>
<td>30</td>
<td>(1)^30</td>
<td>(0)^30</td>
<td>(3)^30</td>
<td>40</td>
<td>.017800</td>
<td>.017919</td>
<td>.017917</td>
</tr>
<tr>
<td>30</td>
<td>(1, 2, ..., 30)</td>
<td>(0)^30</td>
<td>(7)^30</td>
<td>40</td>
<td>.02722</td>
<td>.02735</td>
<td>.02736</td>
</tr>
<tr>
<td>100</td>
<td>(6)^100</td>
<td>(0)^100</td>
<td>(100)^100</td>
<td>3500</td>
<td>.906772</td>
<td>.906717</td>
<td>.906719</td>
</tr>
</tbody>
</table>
Table 6: Dirichlet Examples

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\alpha$</th>
<th>$a$</th>
<th>$b$</th>
<th>exact/simulated 95% Confidence Interval</th>
<th>2nd-order sp (% Relative Error)</th>
<th>Edgeworth (% Relative Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>(10, 8, 8)</td>
<td>$(0)^3$</td>
<td>$(\frac{11}{19}, \frac{10}{19}, \frac{10}{19})$</td>
<td>.955</td>
<td>.956</td>
<td>.957</td>
</tr>
<tr>
<td>3</td>
<td>(10, 8, 8)</td>
<td>$(0)^3$</td>
<td>$(.45, .32)$</td>
<td>.024</td>
<td>.0239</td>
<td>.0283</td>
</tr>
<tr>
<td>3</td>
<td>$(1)^3$</td>
<td>$(.2)^3$</td>
<td>$(.4)^3$</td>
<td>.04</td>
<td>.04001</td>
<td>.04010</td>
</tr>
<tr>
<td>3</td>
<td>$(1)^3$</td>
<td>$(.2)^3$</td>
<td>$(.4)^3$</td>
<td>.04</td>
<td>.03998</td>
<td>.03529</td>
</tr>
<tr>
<td>5</td>
<td>$(\frac{1}{7}, \frac{2}{7}, \ldots, \frac{5}{7})$</td>
<td>$(0)^5$</td>
<td>$(.4)^5$</td>
<td>.0941</td>
<td>.0936</td>
<td>.0864</td>
</tr>
<tr>
<td>50</td>
<td>$(\frac{1}{2}, 1, \ldots, 25)$</td>
<td>$(0)^{50}$</td>
<td>$(.024, .028, \ldots, .2)$</td>
<td>.968356</td>
<td>.96824568</td>
<td>.96824566</td>
</tr>
<tr>
<td>50</td>
<td>$(5)^{50}$</td>
<td>$(.01)^{50}$</td>
<td>$(1)^{50}$</td>
<td>.0215724</td>
<td>.0215745</td>
<td>.0215742</td>
</tr>
<tr>
<td>50</td>
<td>$(.5, 1, \ldots, 25)$</td>
<td>$(0)^{50}$</td>
<td>$(.029, .018, \ldots, 4.5)$</td>
<td>.036800</td>
<td>.03700630</td>
<td>.03700630</td>
</tr>
<tr>
<td>80</td>
<td>$(2)^{80}$</td>
<td>$(0)^{80}$</td>
<td>$(.032)^{80}$</td>
<td>.0270833</td>
<td>.0273532</td>
<td>.0273582</td>
</tr>
<tr>
<td>100</td>
<td>$(1.5)^{100}$</td>
<td>$(0)^{100}$</td>
<td>$(.03)^{100}$</td>
<td>.028676</td>
<td>.0285104</td>
<td>.0285159</td>
</tr>
</tbody>
</table>