SENRA Academic Publishers, British Columbia
Vol. 8, No. 3, pp. 3081-3089, October 2014
Online ISSN: 1920-3853; Print ISSN: 1715-9997

Canadian Journal of pure&applied sciences

SADDLEPOINT APPROXIMATION TO CUMULATIVE DISTRIBUTION FUNCTIONS FOR SOME DIFFICULT AND UNKNOWN LINEAR COMBINATIONS OF RANDOM VARIABLES

*Al Mutairi Alya O and Heng Chin Low
School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia

ABSTRACT

Approximations are very important because it is sometimes not possible to obtain an exact representation of the probability distribution function (PDF) and the cumulative distribution function (CDF). Even when true (exact) representations are possible, approximations, in some cases, simplify the analytical treatments. In this paper, we extend the known saddlepoint tail probability approximations to univariate cases, including univariate conditional cases. Our first approximation (the weighted random sum $S_{N(t)}$) applies to unknown and very difficult statistics (we discuss the approximations within the random sum Poisson-Exponential random variables).We evaluate the performance of the saddlepoint approximation using simulations. Our second approximation (convolutions of Gamma random variables, L_N), are difficult to obtain. These computations are also compared with the exact and normal approximations. We find that the saddlepoint methods provide very accurate approximations for the CDFs probabilities that surpass other methods of approximation, such as normal approximation.The third approximation, including conditional saddlepoint approximations, uses the double saddlepoint. To demonstrate the methods of conditioning in statistical inference, we find a mid p-value using a conditionalsaddlepoint approximation for percentile modified linear rank tests. We show that in the double saddlepoint case, the saddlepoint approximations demonstrate better accuracy than the normal approximation while sharing the same accuracy.

Keywords: Saddlepoint approximation, weighted random sum, poisson-exponential random variables, convolutions of gamma random variables, percentile modified linear rank tests.

INTRODUCTION

The need to analyze distributions of linear combinations of random variables arises in many fields of research, such as biology, seismology, risk theory, insurance application and health science. A mathematical linear combination is expressed as

$$Lc_N = c_1X_1 + c_2X_2 + c_3X_3 + \cdots + c_NX_N, \quad (1)$$

where we have a set of coefficients, c_1 through c_N, that are multiplied by the corresponding variables,X_1 through X_N. During the first term, we have $c_1$times X_1, which is added to$c_2$times X_2, and so on, up to the variable X_N (Ali and Obaidullah, 1982).

This process can be expressed as the sum of the terms $c_i$times X_i, i$= 1, 2, \ldots, N$. The selection of the coefficients c_1 through $c_N$very much depends on the application of interest and the types of scientific questions that we would like to address.

The present paper is organized as follows: in section 1, we establish the basic saddlepoint approximations for the linear combination of random variables. In section 2, we discuss Saddlepoint approximation and real numerical comparisons for the continuous Random Sum Poisson-Exponential Model. In section 3, we derive results for numerical examples for a linear combination of the Gamma distribution. The performance of the saddlepoint approximation for Percentile modified linear rank test is presented in section 4.

1. The linear combination of random variables

In this paper, we discuss saddlepoint approximations to cumulative distribution functions for the linear combination of random variables (Ali and Obaidullah, 1982) in three different cases, as presented:

*Corresponding author email: afaaq99@hotmail.com
*Permanent address:Applied Statistics Department, Faculty of Applied Science, Taibah University, AlMadinah, Kingdom of Saudi Arabia

The linear combination of the sum of independent random variables when N is a random variable

The distributions considered in this study results from the combination of two independent distributions in a particular way. When all $\mathrm{c_i} = 1$, this process is termed

"generalization" by some authors (Johnson *et al*., 2005), though the term "generalized" is greatly overused in statistics. This distribution includes the sums of independent identically distributed (i.i.d.) random variables, $\{X_i\}$, with random index N, independent of

$X_i s$.

Definition 1

Let $X_1, X_2, X_3, \dots$ be a sequence of independent identically distributed (i.i.d.) random variables with a common distribution $f_X(x)$. Let Nbe a discrete random variable that takes the value $1,2,3,\dots$ and let$X_i s$be independent of N, and $\mathrm{c_i}$be non-negative real numbers. The sum

$$\mathrm{R_N} = \mathrm{c_1X_1} + \mathrm{c_2X_2} + \mathrm{c_3X_3} + \cdots + \mathrm{c_NX_N} \qquad (2)$$

is called the weighted random sum (Kasparaviciute and Leonas, 2013).

Such sums have a wide range of applications in branching processes, damage processes and risk theory. A common application of the random sum is that a total claim amount is presented to an insurance company, where N is the number of claims and theX_is are the individual claims, which are assumed to be independent.

In general, random sums are extremely difficult to investigate; therefore, approximation techniques are frequently employed. Saddlepoint methods overcome this difficulty while providing us with an influential tool for obtaining precise expressions for distribution functions that are still unknown in the closed form. In addition, these methods roughly surpass other techniques in terms of calculating expenses, but exceed no other methods in terms of accuracy.

In this paper, approximations of the unknown difficult random sum Poisson-Exponential random variables which have a continuous distribution are discussed. The saddlepoint approximation method is shown to be not only quick, dependable, stable and accurate enough for general statistical inference, but it is also applicable without deep knowledge of probability theory.

The linear combination of the sum of independent random variables when N is constant

Linear combinations of convoluted random variables occur in a wide range of fields. In most cases, the exact distribution of these linear combinations is extremely difficult to determine, and the normal approximation usually performs very badly for these complicated distributions. A better method of approximating linear combination distributions involves the additional use of saddlepoint approximation.

Saddlepoint approximation is able to provide accurate expressions for distribution functions that are unknown in their closed forms. This method not only yields an accurate approximation, near the center of the distribution but also controls the relative error in the far tail of the distribution.

Definition 2

The probability distribution of the sum of two or more independent random variables is the convolution of their individual distributions. Consider the sum of two independent random variables, $\mathrm{Z}$and Y. The distribution of their sum,$\mathrm{X} = \mathrm{Z} + \mathrm{Y}$, is the convolution of these random variables. Now, let $\mathrm{X_1, X_2, \dots, X_N}$be i.i.d. random variables and $\mathrm{c_1, c_2, \dots, c_N}$ be numbers. Thus, the random variable

$$\mathrm{L_N} = \mathrm{c_1X_1} + \mathrm{c_2X_2} + \mathrm{c_3X_3} + \cdots + \mathrm{c_NX_N} = \sum_{i=1}^{N} \mathrm{c_iX_i} \qquad (3)$$

is called the linear combination of the convolution random variable.

We derive the saddlepoint approximation of the convolution $\mathrm{aZ + bY}$, where $\mathrm{a} > 0$ and $\mathrm{b} > 0$ are real constants and Z, Y denote Gamma random variables, respectively, while being distributed independently of each other. The associated saddlepoint approximations CDFs, exact and normal approximation are derived. The plots for the CDFs arealso given.

The linear combination of sum of independent Bernoulli random variables when N is constant and $\mathrm{c_i}$ are scores

The approximation for the distribution function of a test statistic is extremely important in statistics. Many statistical procedures that are applicable to the two sample problems are based on the rank order statistics for the combined samples, and many commonly used two-sample rank tests act as a linear combination of certain indicator Bernoulli random variables for the combined ordered samples.

For the approximation presented in this paper, a saddlepoint formula proposed are given constants called weights or scores, and $\{X_i\}$ are Bernoulli distributions, $i = 1,2,\ldots,N$.

Definition 3

Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be two independent random samples drawn from populations with the continuous cumulative distribution functions, F_X and F_Y, respectively. Let $\mathrm{N} = \mathrm{m} + \mathrm{n}$; then, the statistic

$T_N = \sum_{i=1}^{N} c_i Z_i$ (4)

is called a linear rank statistic, where the $\{c_i\}$ are given constants called weights or scores, $Z_i = 1$ if the i^{th} sampled value in the combined ordered sample is X and $Z_i = 0$ if it is Y (Gibbons and Chakraborti, 2003). It is noteworthy to mention that the statistic $\mathrm{T_N}$ is a linear combination of independent indicator Bernoulli random variables $\{Z_i\}$.

This paper examines mid p-values from the null permutation simulations distributions. The permutation simulations may lead to intractable computations apart from small values for the sample size, and the normal approximation may not result in the desired accuracy, particularly when the sample size is small. Saddlepoint approximation can be used to overcome this problem. This method results in a highly accurate approximation without placing constraints or guidelines on the values of the sample sizes.

In the three cases of linear combinations involving random variables given in Equations (2), (3) and (4), we used the saddlepoint approximation formula proposed by Daniels (1954, 1987) that has the type developed by Lugannani and Rice (1980) for the cumulative distribution function of a continuous random variable X with CDF F andcumulant generating function CGF K, with mean, μ. The saddlepoint approximation for $F(x)$, as introduced by Lugannani and Rice (1980), is

$$\hat{F}(x) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\hat{u}) & \text{if } x \neq \mu \\ \dfrac{1}{2} + \dfrac{K\square(0)}{6\sqrt{2\pi K''(0)^{3/2}}} & \text{if } x = \mu \end{cases} \quad (5)$$

where Φ and ⍰⍰ denote the standard normal density and CDF, respectively, and

$$\hat{w} = sgn(\hat{s})\sqrt{2\{\hat{s}x - K(\hat{s})\}}, \qquad \hat{u} = \hat{s}\sqrt{K''(\hat{s})} \qquad (6)$$

are functions of x and saddlepoint $\hat{s}$. In this case, $\hat{s}$ is the implicitly defined function of x given as the unique solution to $K'(\hat{s}) = x$, and $sgn(\hat{s})$ captures the sign ± for $\hat{s}$.

To approximate these unknown difficult statistics based on their moment generating functions, theorems related to these unknown statistics are employed. Then, we derived the saddlepoint equations that, in some cases, can be solved using numerical methods. By performing some calculations and applying saddlepoint formulas, we can obtain the CDF for these unknown difficult statistics. Subsequently, we find the exact distributions using simulation methods and the mean square error (MSE) as well as the absolute, relative error (RE) to investigate the performance of the saddlepoint approximation.

The Skovgaard (1987) approximation when Y is a continuous variable for which $F(y|x)$ admits a densityis

$$\hat{F}(y|x) = \Phi(\hat{w}) - \phi(\hat{w})(1/\hat{w} - 1/\hat{u}), \qquad \hat{t} \neq 0 \qquad (7)$$

where

$$\hat{w} = sgn(\hat{t})\sqrt{2[\{K(\hat{s}_0, 0) - \hat{s}_0{}^T x\} - \{K(\hat{s}, \hat{t}) - \hat{s}^T x - \hat{t}y\}]} (8)$$

$$\hat{u} = \hat{t}\sqrt{\frac{|K''(\hat{s}, \hat{t})|}{|K_{ss}''(\hat{s}_0, 0)|}}$$

The components s and t are associated with X and Y, respectively.For $(x, y) \epsilon \tau_\chi$ and τ_χ is the interior of the convex hull of the support $\chi = \{(x, y) : f(x, y) > 0\}$. Here, the m-dimensional saddlepoint $(\hat{s}, \hat{t})$ solves the set of m equations $K'(\hat{s}, \hat{t}) = (x, y)$.Where $K'(\hat{s}, \hat{t})$ is the gradient with respect to both s and t. If $K''(\hat{s}, \hat{t})$ is the corresponding Hessian, the continuity corrections to CDF, as introduced by Skovgaard (1987) should be used to achieve the greatest accuracy.

First continuity correction

Suppose $(\hat{s}, \hat{t})$ is the solution to $K'(\hat{s}, \hat{t}) = (j, k)$ required for the numerator saddlepoint with $\hat{t} \neq 0$. Then,

$$\hat{P}_{r1}(Y \geq k | X = j) = 1 - \Phi(\hat{w}) - \phi(\hat{w})(1/\hat{w} - 1/\tilde{u}_1), \qquad \hat{t} \neq 0 \qquad (9)$$

where

$$\hat{w} = sgn(\hat{t})\sqrt{2[\{K(\hat{s}_0, 0) - \hat{s}_0{}^T j\} - \{K(\hat{s}, \hat{t}) - \hat{s}^T j - \hat{t}k\}]} \qquad (10)$$

$$\tilde{u}_1 = \left(1 - e^{-\hat{t}}\right)\sqrt{\frac{|K''(\hat{s}, \hat{t})|}{|K_{ss}''(\hat{s}_0, 0)|}} \qquad (11)$$

and $\hat{s}_0$ solves $K_s'(\hat{s}_0, 0) = j$; see Skovgaard (1987).

Second continuity correction

If $k^- = k - 0.5$ is the offset value of k and $(\tilde{s}, \tilde{t})$ is the offset saddlepoint solving

$K'(\tilde{s}, \tilde{t}) = (j, k - 0.5)$ (12)

with $\tilde{t} \neq 0$, then

$$\hat{P}_{r2}(Y \geq k|X = j) = 1 - \Phi(\tilde{w}_2) - \phi(\tilde{w}_2)(1/\tilde{w}_2 - 1/\tilde{u}_2), \qquad \tilde{t} \neq 0 \quad (13)$$

where

$$\tilde{w}_2 = sgn(\tilde{t})\sqrt{2[\{K(\hat{s}_0, 0) - \hat{s}_0{}^T j\} - \{K(\tilde{s}, \tilde{t}) - \tilde{s}^T j - \tilde{t}k^-\}]}(14)$$

$$\tilde{u}_2 = 2\sinh(\tilde{t}/2)\sqrt{\frac{|K''(\tilde{s}, \tilde{t})|}{|K_{ss}{}''(\hat{s}_0, 0)|}} \quad (15)$$

and the saddlepoint $\hat{s}_0$ is unchanged. Then, we find
$\hat{F}_2(k|X = j) = 1 - \hat{P}_{r2}(Y \geq k + 1|X = j),$

2. Saddlepoint approximation and real numerical comparisons for the continuous Random Sum Poisson Model

The random sum distribution plays a key role in both probability theory and its applications in biology, seismology, risk theory, meteorology and health science. The statistical significance of this distribution arises from its applicability to real-life situations, in which the researcher often observes only the total amount, say S_N, which is composed of an unknown random number N of random contributions, sayXs.

In health science, the random sum plays a very important role in many real-life applications. For example, let the number of hot spot of a contagious disease follow a Poisson distribution with a mean of ψ, and let the number of sick people within the hotspotfollow a Negative Binomial distribution. If we want to find the probability that the total number of sick people is greater than 70, then the total number of sick people within the hotspotis

$$S_{N_1} = \sum_{i=1}^{N} X_i \qquad (16)$$

where$X_i \backsim$ Negative Binomial (r, m) and $N \backsim$ Poisson(ψ) .

Another practical application of the random sum is the number of times that it rains in a given time period, say N, which has a Poisson distribution with mean λ. If the amount of rain that falls has an Exponential distribution and if the rain fallsat that time period is independent of N, then the total rainfall in the time period is

$$S_{N_2} = \sum_{i=1}^{N} Y_i \qquad (17)$$

where$Y_i \backsim$ Exponential(ϕ) and $N \backsim$ Poisson (λ) .

In fact, the total random sums S_{N_1} and S_{N_2} are composed of an unknown random number N of other random contributions, say X or Y which are very complex to analyze. In most cases, the distribution of the random sum is still unknown; in other cases, it is already known but is too complex for the computation of the distribution function, which often becomes too slow for many problems (Johnson *et al.*, 2005). The saddlepoint approximation method can help us gain knowledge of these unknown difficult statistical behavior.

In this section, we suggest that the saddlepoint approximation and efficiency analysis should be compared to the true distribution over real data compared with other methods of approximation, such as normal approximations. Suppose that the number of times it rains in a given time period, N has a Poisson distribution with mean λ. Suppose,also when it rains, the amount of rain falling has an Exponential distribution. Let the rain falling and the time period be independent of one another and of N. Then, the total rainfall in the time period is as follows:

$$S_{N(t)} = \sum_{i=1}^{N(t)} R_i \,, \qquad t > 0 \qquad (18)$$

where$\{R_i\}$are independent random variables with a distribution of R. Suppose now we observe the rainfall for Nin certain periods:$S_1, S_2, \ldots, S_N$.

Note that the probability of no rain in any such period is $P(S_{N(t)} = 0) = P_0 =$exp (λ)
Withers and Nadarajah (2011) considered the annual maximum daily rainfall data for the year 1907 to 2000 for fourteen locations in West Central Florida: Clermont, Brooksville, Orlando, Bartow, Avon Park, Arcadia, Kissimmee, Inverness, Plant City, Tarpon Springs, Tampa International Airport, St. Leo, Gainesville, and Ocala. The data were obtained from the Department of Meteorology in Tallahassee, Florida. Consider the distribution of $S_{N(t)}$, such that the unknown parameters are$\lambda$and α. The study conducted by Withers and Nadarajah (2011) found a distribution fit by these three methods, unconditional maximum likelihood estimation, conditional maximum likelihood estimation, and moments estimation. Remarkably, unconditional maximum likelihood estimation provided the best fit for each location; for example, in Orlando, the estimates were $\hat{\lambda}$= 9.565, $\hat{\alpha}$ = 2.373.

The numerical computations are plotted in figure 1 to show the ‘exact’ CDFs for a random sum Poisson (9.565)– *Exponential* (2.373)distribution $F(x)$ (the solid line), the saddlepoint $\hat{F}(x)$(the dotted line) and the normal approximation $F'(x)$ (the dashed line).

The empirical distribution function is used to determine the ‘exact’ CDF for this model by simulating 10^6 independent values of S_N, where N is Poisson(λ), and the X_i's are i.i.d random variables generated using

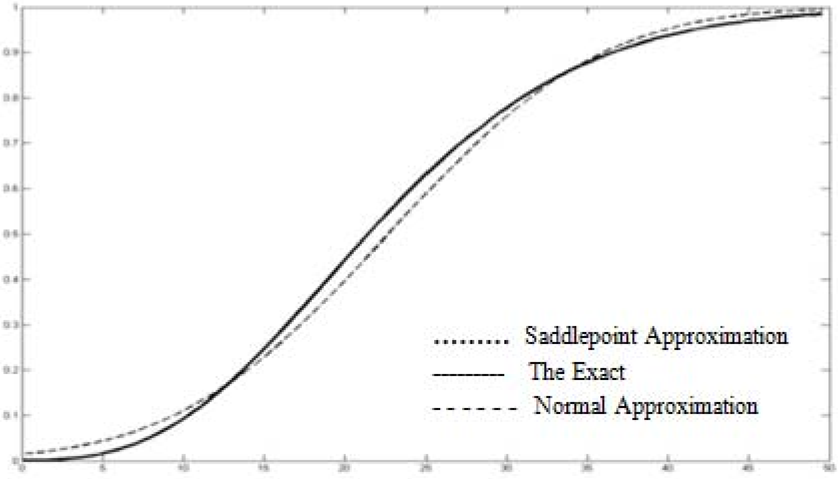

Fig. 1. The performance of the saddlepoint approximation $\hat{F}(x)$, the exact $F(x)$ and the normal approximation $F^{'}(x)$for random sum Poisson (9.565)- Exponential (2.373) model.

Table 1. Approximate left tail of the saddlepoint approximation $\hat{F}(x)$ vs. exact $F(x)$and the normal approximation $F^{'}(x)$, for random sum Poisson (9.565) Exponential (2.373) model.

x	$F(x)$	$\hat{F}(x)$	$F^{'}(x)$	%RE1	%RE2
0.1	0.000100	8.96E-05	0.014731	1.04E-01	146.3100
0.2	0.000132	0.0001255	0.015094	4.94E-02	113.3485
0.3	0.000168	0.0001662	0.015464	1.07E-02	91.04762
0.4	0.000225	0.0002122	0.015843	5.70E-02	69.41333
0.5	0.000260	0.0002639	0.016229	1.50E-02	61.41923
0.6	0.000320	0.0003217	0.016624	5.44E-03	50.95000
0.7	0.000395	0.0003862	0.017026	2.24E-02	42.10380
0.8	0.000487	0.0004575	0.017437	6.05E-02	34.80493

MATLAB program. This plot shows that the shape of $F(x)$ is the same as that of $\hat{F}(x)$(i.e., the two approximations are identical) but differs from that of $F^{'}(x)$. The plot suggests that the saddlepoint approximation for CDFs has the same accuracy as the 'exact' CDFs and is far superior to normal approximation. Table 1 shows the evaluation of the left tail probabilities for certain values of the exact $F(x)$ of the random sum in the second column, with the saddlepoint $\hat{F}(x)$ in the third column and the normal approximation in fourth column. However, based on the fifth column (the absolute relative errors), the accuracy of this method is very clear. For example, in the left tail probability, we obtained the following relative error values: 1.04E-01, 4.94E-02, 1.07E-02, 5.70E-02,1.50E-02..., and so on. All of these amounts and others suggest good approximations in the left tail. The maximum absolute relative error for the $F(x)$vs.$\hat{F}(x)$ approximation, based on our calculations for this example, appears in this tail and was 1.04E-01when $X = 0.1,\ F(x) = 0.000100$ and $\hat{F}(x) = 8.96E - 05$. However, the approximation is still good with this amount of error (acceptable).

If we refer to table 2 to examine the relative error values near the center of the distribution, we obtain 3.76E-03, 3.09E-03, 2.69E-03,4.38E-03…, and so on. Based on these values and others, the accuracy is increased compared to that in the left tail probability.

The relative error values in the right tail probability, as shown in table 3, are3.10E-04,1.07E-05,4.91E-04,2.66E-04…, and so on. At this point, the accuracy is optimal.

Throughout the entire set of results (from x=0.1 to x=40.5 with step 0.1), with its corresponding figure 1carried out usingMATLAB program, the accuracy generally appears to be increasing. In general, for this application, the mean squared error of the saddlepoint approximation is MSE(1) = 1.28625E-06, that is, very close to zero, while the means squared error of the normal approximation is MSE(2) = 000476.

Table 2. Approximate center of the distribution of the saddlepoint approximation $\hat{\mathrm{F}}(\mathrm{x})$, the exact $\mathrm{F}(\mathrm{x})$ & the normal approximation $\mathrm{F}^{'}(\mathrm{x})$ for Poisson (9.565) Exponential (2.373) model.

x	$F(x)$	$\hat{F}(x)$	$F^{'}(x)$	%RE1	%RE2
21.4	0.49764	0.49577	0.45025	3.76E-03	0.095229
21.5	0.50122	0.49967	0.45406	3.09E-03	0.09409
21.6	0.50492	0.50356	0.45788	2.69E-03	0.093163
21.7	0.50967	0.50744	0.46171	4.38E-03	0.094100
21.8	0.51366	0.51131	0.46554	4.58E-03	0.093681
21.9	0.51710	0.51518	0.46937	3.71E-03	0.092303
22.0	0.52150	0.51903	0.47320	4.74E-03	0.092617
22.1	0.52472	0.52287	0.47704	3.53E-03	0.090868

Table 3. Approximate right tail of the saddlepoint approximation $\hat{\mathrm{F}}(\mathrm{x})$, the exact $\mathrm{F}(\mathrm{x})$ & the normal approximation $\mathrm{F}^{'}(\mathrm{x})$ for random sum Poisson (9.565) Exponential (2.373) model.

x	$F(x)$	$\hat{F}(x)$	$F^{'}(x)$	%RE1	%RE2
39.8	0.93567	0.93538	0.95030	3.10E-04	0.015636
39.9	0.93630	0.93629	0.95128	1.07E-05	0.015999
40.0	0.93764	0.93718	0.95225	4.91E-04	0.015582
40.1	0.93831	0.93806	0.95320	2.66E-04	0.015869
40.2	0.93923	0.93893	0.95413	3.19E-04	0.015864
40.3	0.93974	0.93980	0.95505	6.38E-05	0.016292
40.4	0.94101	0.94065	0.95596	3.83E-04	0.015887
40.5	0.94162	0.94149	0.95685	1.38E-04	0.016174

These results indicate that the saddlepoint approximation is almost exact. Thus, we conclude that the saddlepoint approximation method provides us with an accurate approximation for this difficult statistic, the accuracy of which appears to leave no room for doubt in either of the two tails or in the center of the distribution.

3. Real numerical comparisons of the Saddlepoint approximation for linear combination of Gamma distribution

Saddlepoint approximation plays an important role in helping us gain knowledge about unknown difficult distributional behavior, such as the linear combination of random variables. In this study, we discuss the linear combination of the Gamma distribution. This convolution model is given by

$$S_N = c_1X_1 + c_2X_2, \qquad (19)$$

where, X_1 and X_2 are both independent, following an Gamma distribution with parameters (θ, ρ) and (α, β), respectively. This paper investigates the saddlepoint approximations of the convolution, where $c_1 > 0$ and $c_2 > 0$ are real constants.

In the univariate case, a general saddlepoint approximation was given for the continuous CDFs. For the linear combinations of Gamma models, this method of approximation was applied where the root was found numerically. In this setting, the efficiency of this method was explored using the empirical CDFs found by simulation methods.

Figure 2 shows a comparative plot of the ‘true’ CDFs $\mathrm{F}(\mathrm{x})$ with the saddlepoint $\hat{\mathrm{F}}(\mathrm{x})$CDFs and the normal approximation CDFs for a linear combination of Gamma distribution. It is clear from this figure that the two approximations $F(x)$ and $\hat{F}(x)$ are very close, but differs from $\mathrm{F}^{'}(\mathrm{x})$. This result means that the saddlepoint approximation for CDFs has the same accuracy as the ‘exact’ CDFs and is far superior to the normal approximation. The first value of each cell of table 4 is ‘exact’. The second and the third values are the saddlepoint approximation and normal approximation, respectively. The fourth and fifth columns show the absolute, relative errors between the saddlepoint approximation and the ‘exact’ CDFs and the relative errors between the normal approximation and the ‘exact’ CDFs, respectively.

The ‘exact’ CDFs were computed using the empirical distribution by simulating 10^6 independent values of $S_N = c_1X_1 + c_2X_2$, where $X_1 \sim Gamma(0.5,1)$, $X_2 \sim Gamma(1,2)$ and

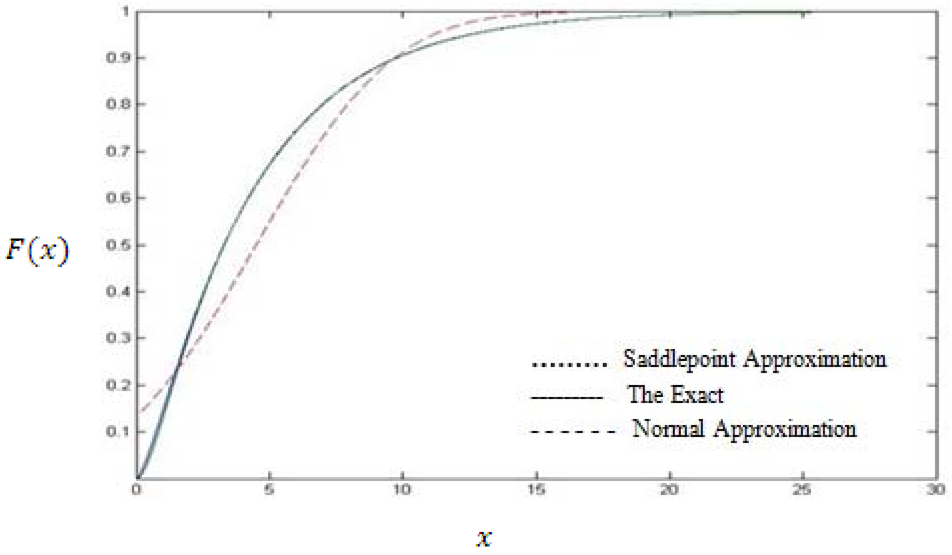

Fig. 2. Performance ofsaddlepoint approximation $\hat{F}(x)$ and the exact $F(x)$ with normal approximation $F'(x)$ for linear combination of Gamma (0.5,1,1,2).

Table 4. Comparison saddlepoint approximation $\hat{F}(x)$ and the exact $F(x)$ with normal approximation $F'(x)$ for linear combination of Gamma (0.5,1,1,2).

x	$F(x)$	$\hat{F}(x)$	$F'(x)$	%RE1	%RE2
3.5000	0.5242	0.5242	0.4028	0	0.231591
6.0000	0.7448	0.7433	0.6440	0.002014	0.135338
8.5000	0.8636	0.8621	0.8376	0.001737	0.030107
11.0000	0.9268	0.9260	0.9452	0.000863	0.019853
13.5000	0.9610	0.9603	0.9866	0.000728	0.026639
16.0000	0.9790	0.9787	0.9977	0.000306	0.019101
18.5000	0.9887	0.9886	0.9997	0.000101	0.011126
21.0000	0.9939	0.9939	1.0000	0	0.006137
23.5000	0.9968	0.9967	1.0000	0.000100	0.003210
26.0000	0.9983	0.9982	1.0000	0.000100	0.001703
28.5000	0.9991	0.9991	1.0000	0	0.231591

$c_1 = 1, c_2 = 2$ when$x = 0.01$.Generated by the MATLAB program.

Table 4 shows the relative errors for $F(x)$ vs. $\hat{F}(x)$ values of the distribution. The relative errors remain very small for the computations and the accuracy appears quite good, although the accuracy in the center is not quite as good as that in the left tail. Moreover, the numerical results indicate that the normal approximations are considerably less accurate than the saddlepoint approximations.

The values of the relative errors for $F(x)$ vs. $\hat{F}(x)$,for the right tail , the accuracy is good and very clear. However, the performance of the normal approximation in this tail appears to be much better than its performance in the center and in the left tail. Nevertheless, the saddlepoint approximation maintains its accuracy in the left, right and center of the distribution. In general, the numerical results indicate that the saddlepoint approximation is far more accurate than the normal approximation.

Table 5. Comparisonof saddlepoint approximation mid p-values and the exact with normal approximation for percentile modified linear rank tests.

Exact	Normal		Saddlepoint			
			First continuity correction		Second continuity correction	
	Mid-p-value	%RE	Mid-p-value	%RE	Mid-p-value	%RE
0.322	0.325	0.003106	0.323	0.009317	0.322	0
0.413	0.415	0	0.413	0.004843	0.413	0
0.377	0.378	0.002653	0.378	0.002653	0.377	0

Based on figure 2 with its corresponding numerical result using MATLAB program (from x = 3.5000 to x =28.5000 with step 2.5), this leads to

$$\mathrm{MSE}(1) = \frac{1}{N}\sum_{i=1}^{N}(F(x_i) - \hat{F}(x_i))^2 = 0.000077696 \quad (20)$$

Additionally, the MSE (2) for $F(x)$ vs. $F^{'}(x)$ was calculated as

$$\mathrm{MSE}(2) = \frac{1}{N}\sum_{i=1}^{N}(F(x_i) - F^{'}(x_i))^2 = 0.0019 \qquad (21)$$

We note that the value of MSE(1) is far smaller than that of MSE(2) and that MSE(1) is itself very small and close to zero (i.e. the accuracy of the estimator with a smaller mean squared error is also higher). This result indicates that the saddlepoint approximation furnishes a good fit and is superior to the normal approximation.

4. The linear combinations of the rank-order statistics for percentile modified linear rank tests

P-value is associated with a test statistic. It is the probability if the test statistic really were distributed as it would be under the null hypotheses, of observing a test statistic (as extreme as, or more extreme than) the one actually observed.

The smaller p-value, the more strongly the test rejects the null hypothesis, that is, the hypothesis being tested. A p-value of 0.05 or less rejects the null hypothesis. However, this study uses saddlepoint methods to determine mid-p-values from the linear combinations of the rank-order statistics. The two methods suggest that normal approximation and permutation simulations can be used to determine mid-p-values from the null permutation distributions. The permutation simulations lead to intractable computations apart from the small sample size. The normal approximation demands that certain conditions be applied relative to the sample size; thus, without these conditions, the results will not attain the desired accuracy, particularly when the sample size is small. The saddlepoint approximation provides a result using a highly accurate approximation without the need to place constraints or guidelines on the sample, and its accuracy is apparent even when the sample size is 1. Another advantage of these saddlepoint methods is that the required computational times are essentially negligible compared to the simulations. The real datasets used include small, intermediate and large sample sizes respectively,to show how accurate the saddlepoint method can be for all sample sizes.

For the third new estimators, table 5 shows the exact (true), normal and saddlepoint mid-p-values for linear combinations of the rank-order statistics for the two sample problems. In all examples, as a result, we determined how much the permutation simulations lead to complicated computations, apart from small values for the sample size. As indicated by the absolute relative errors, saddlepoint approximations can replace the permutation simulations and provide mid-p-values that are virtually exact for all practical purposes without the same required conditions or guidelines regarding the sample size as the normal approximation.And in most cases, the second correction is better than first corrections. Additionally, in both two continuity-corrected CDFs, the saddlepoint approximation is far more accurate than the normal approximation.All of the computations for this third new estimator were performed using FORTRAN software.

CONCLUSION

Estimating the CDFs for some linear combination of random variables is one of the problems we face in statistical inference that has many applications throughout life. The difficulty of estimating the CDFs for a given model should be detected. In this study, we used saddlepoint approximations as a better method to achieve an accurate approximation of the CDFs.Based on present study, three different versions of new saddlepoint approximations were developed. For the first approximation(the weighted random sums $S_{N(t)}$). We demonstrated the performance of the saddlepoint approximation in a wide range of applications.The proposed new estimators using the saddlepoint approximationis highly accurate. The performance of the first new estimator for the random sumwas evaluated by the relative error between the exact value and the saddlepoint approximation and between the exact value and the normal approximation for each value. In addition,the mean squared error for the saddlepoint approximation was compared with the mean squared error for the normal approximation.For the second new estimators, saddlepoint approximation to a linear combination of Gamma models was considered. These models show close agreement between the exact, and saddlepoint and far superior accuracy to the normal approximation. Moreover, for the third new estimators, saddlepoint approximations can replace the permutation simulations and provide mid-p-values that are virtually exact for all practical purposes without the same required conditions or guidelines regarding the sample size as the normal approximation.In conclusion,we confirmed the accuracy of the saddlepoint approximation in these three different settings for the linear combination model.

ACKNOWLEDGEMENT

This study was supported financially by Taibah University, AlMadinah-M., Kingdom of Saudi Arabia.

REFERENCES

Ali, MM. and Obaidullah, M. 1982. Distribution of linear combination of exponential variates. Communications in Statistics-Theory and Methods. 11:1453-1463.

Daniels, HE. 1987. Tail probability approximations. International Statistical Review. 55:37-48.

Daniels, HE. 1954. Saddlepoint approximations in statistics. Annals of Mathematical Statistics. 25:631-650.

Gibbons, JD. and Chakraborti, S. 2003. Nonparametric statistical inference. (4th edi.). Marcel Dekker, New York, USA.

Johnson, NL., Kemp, AW. and Kotz, S. 2005. Univariate discrete distributions. (3rd edi.). John Wiley and Sons, Inc., USA.

Kasparaviciute, A. and Leonas S. 2013. Large deviations for weighted random sums. Non Llinear Analysis-Modeling and Control. 18:129-142.

Lugannani, R. and Rice, SO. 1980. Saddlepoint approximations for the distribution of the sum of independent random variables. Advances in Applied Probability. 12:475-490.

Skovgaard, IM. 1987. Saddlepoint expansions for conditional distributions. Journal of Applied Probability. 24:875-887.

Withers, CS. and Nadarajah, S. 2011. On the compound Poisson-gamma distribution. Kybernetika. 47:15-37.

Received: June 22, 2014; Accepted: July 7, 2014