Saddlepoint Approximation to Cumulative Distribution Function for Poisson–Exponential Distribution

Al Mutairi Alya O.[1,2] & Heng Chin Low[1]

[1] School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia

[2] Faculty of Applied Science, Taibah University, AlMadinah-M., Kingdom of Saudi Arabia

Correspondence: Al Mutairi Alya O., Applied Statistics Department, Faculty of Applied Science, Taibah University, AlMadinah-M., Kingdom of Saudi Arabia. E-mail: afaaq99@hotmail.com

Abstract

The random sum distribution is a key role in probability theory and its applications as well, these applications could be used in different sciences such as insurance system, biotechnology, allied health science, etc. The statistical significance in random sum distribution initiates when using the applications of probability theory in the real life, where the total quantity X can be only observed, which is included of an unknown random number X of random contributions. Saddlepoint approximation techniques overcome this problem. Saddlepoint approximations are effective tools in getting exact expressions for distribution functions that are not known in closed form. Saddlepoint approximations usually better than the other methods in which calculation costs, but not necessarily about accuracy. This paper introduces the saddlepoint approximations to the cumulative distribution function for random sum Poisson- Exponential distributions in continuous settings. We discuss approximations to random sum random variable with dependent components assuming existence of the moment generating function. A numerical example of continuous distributions from the Poisson- Exponential distribution is presented.

1. Introduction

Saddlepoint approximations are powerful tools for obtaining accurate expressions for distribution functions which are not known in closed form. Saddlepoint approximations, usually better than other methods regarding to the calculation costs, but not necessarily about accuracy. Daniels (1954) in his study presented, the primary saddlepoint approximation and that is basically a formula for approximating a density or mass function from its associated moment generating function. Saddlepoint approximations are depend on using the moment generating function (MGF) or, equivalently, the cumulant generating function (CGF), of a random variable. For discussion saddlepoint approximations methodologies and the relevant techniques see Daniels (1954) and Daniels (1987) for the details of density and mass approximation and see Skovgaard (1987) for a conditional version of saddlepoint approximation as well. In addition, Reid (1998) for applications to inference, Borowiak (1999) for discussion of a tail area approximation which has uniform relative error, and Terrell (2003) for a stabilized Lugannani-Rice formula.

We will discuss approximations to the random sum variable with dependent components assuming the moment generating function that has been exists. Suppose a continuous random variable X has density function $f(x)$ defined for all real values of x. Then the moment generating function MGF is defined as

$$M(s) = E(e^{sx}) = \int_{-\infty}^{\infty} e^{sx} f(x)dx \qquad (1)$$

Wherever this expectation exists and $M(0)$ always exists and is equal to 1. We shall assume that $M(s)$ converges over the largest open neighborhood at zero as (a,b). The cumulant generating function CGF is given by (Hogg & Craig, 1978)

$$K(s) = \ln M(s), \quad s \in (a,b) \qquad (2)$$

For continuous random variable X with CGF $K(s)$ and unknown density $f(x)$, the saddlepoint density approximation of $f(x)$ is given by (Johnson et al., 2005)

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} exp(K(\hat{s}) - \hat{s}x), \tag{3}$$

where $\hat{s} = \hat{s}(x)$ denotes the unique solution to the saddlepoint equation $K'(\hat{s}) = x$, (Daniels, 1954). The approximation is useful for values of x that is interior point of the support $\{x : f(x) > 0\} = \chi$. The normalized Saddlepoint density is defined as

$$\bar{f}(x) = \frac{\hat{f}(x)}{\int_{\chi} \hat{f}(x)dx} \tag{4}$$

As well as, it is clear to note that $\int_{\chi} \hat{f}(x)dx \neq 1$, because the interior point of the support

$$\{x : f(x) > 0\} = \chi.$$

The saddlepoint approximation for univariate cumulative distribution functions $F(x)$ is given by

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

where the continuous random variable X has CDF $F(x)$ and CGF $K(s)$ with mean $\mu = E(x)$ and $\hat{w}$ and $\hat{u}$ are defined as

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

As well as, $\hat{w}, \hat{u}$ are function of x and saddlepoint $\hat{s}$, where $\hat{s}$ is the implicitly defined function of x given by the unique solution to $K'(\hat{s}) = x$ and symbol ϕ and Φ denote the standard normal density and CDF respectively and $\text{sgn}(\hat{s})$ captures sign $(\pm)$ for $\hat{s}$ (Butler, 2007).

2. The Random Distributions

The random sum distributions have many natural applications. We motivate the notion of random distributions with an insurance application. In an individual insurance setting, we wish to model the aggregate claims during a fixed policy period for an insurance policy. In this setting, more than one claim is possible. The random variable Y is said to have a random distribution if Y is of the following form

$$Y = X_1 + X_2 + X_3 + ... + X_N \tag{7}$$

where the number of terms N is uncertain, the random variables X_i are independent and identically distributed (with common distribution X) and each X_i is independent of N. If $N = 0$ is realized, then we have $Y = 0$. Even though this is implicit in the definition, we want to call this out for clarity. In our insurance contexts, the variable N represents the number of claims generated by an individual policy or a group of independent insured over a policy period. The variable X_i represents the ith claim. Then Y represents the aggregate claims over the fixed policy period. The distribution function of Y is given by

$$f_Y(y) = \sum_{n=0}^{\infty} G_n(y)P[N = n] \tag{8}$$

Where $n \geq 1, G_n(y)$, is the distribution function of the independent sum $X_1 + X_2 + X_3 + ... + X_N$. We can also express f_Y in terms of convolutions (Johnson et al., 2005):

$$f_Y(y) = \sum_{n=0}^{\infty} f^{*n}(y)P[N = n] \tag{9}$$

where f is the common distribution function for X_i and f^{*n} is the n-fold convolution of f. If the common claim distribution X is discrete, then the aggregate claims Y is discrete. On the other hand, if X is continuous and if $P[N = n] > 0$, then the aggregate claims Y will have a mixed distribution, as is often the case in insurance applications. The mean aggregate claims $E[Y]$ is

$$E[Y] = E[N]E[X] \tag{10}$$

The expected value of the aggregate claims has a natural interpretation. It is the product of the expected number of claims and the expected individual claim amount. This makes intuitive sense. The variance of the aggregate claims $Var[Y]$ is

$$Var[Y] = E[N]Var[X] + Var[N]E[X]^2 \tag{11}$$

The variance of the overall claims also has a natural explanations. It is the sum of two elements such that, the first element stemming from the variability of the amount claimed by individual and the second element stemming from the variability of the number of claims.

The moment generating function of aggregate claims Y is defined as

$$M_Y(s) = M_N[ln\,M_X(s)] \tag{12}$$

where the function ln is the natural log function. As well as, the cumulant generating function of aggregate claims Y is defined as (Hogg & Tanis, 1983)

$$K_Y(s) = ln\,M_Y(s) = ln\,M_N[K_X(s)] = K_N[K_X(s)]\;. \tag{13}$$

The random sum Poisson distribution is a model for describing the aggregate claims arise in a group of independent insured. Let N be the number of claims generated by a portfolio of insurance policies in a fixed time period. Suppose X_1 is the amount of the first claim, X_2 is the amount of the second claim and so on. Then $Y = X_1 + X_2 + X_3 + ... + X_N$ represents the total aggregate claims generated by this portfolio of policies in the given fixed time period. In order to make this model more tractable, we make the following assumptions:

- $X_1, X_2, ...$ are independent and identically distributed.
- Each X_i is independent of the number of claims N.

The number of claims N is associated with the claim frequency in the given portfolio of policies. The common distribution of $X_1, X_2, ...$ is denoted by X. Note that X models the amount of a random claim generated in this portfolio of insurance policies.

When the claim frequency N follows a Poisson distribution with a constant parameter λ, the aggregate claims Y is said to have a random sum Poisson distribution which has the mean $\lambda = E[N]$ and the variance $Var[Y] = \lambda E[X^2]$. The moment generating function is

$$M_Y(s) = M_N[ln\,M_X(s)] = exp[\lambda(M_X(s)-1)] \tag{14}$$

3. Numerical Example

Suppose that an insurance company acquired two portfolios of insurance policies and combined them into a single block. For each portfolio the aggregate claims variable has a random Poisson distribution. For each of the portfolios, the Poisson parameter is λ and the individual claim amount has an Exponential distribution with parameter β. Let the random sum $Y = X_1 + X_2$ be the aggregate claims generated in a fixed period by an independent group. When the number of claims N follows a Poisson (λ) a distribution, $X_i's$ are i.i.d. random variables follows Exponential (β) distribution, the sum Y is said to have a Poisson-Exponential random sum distribution. The cumulant generating function for N is given by

$$K_N(s) = ln[M_N(s)] = \lambda(e^s - 1) \tag{15}$$

And for $X_i's$ which are i.i.d. random variables follows Exponential (β) distribution. The CGF for $X_i's$ is defined as

$$K_X(s) = ln[M_X(s)] = -ln(1 - \beta s) \tag{16}$$

Then, we can drive the cumulant generating function for the Poisson-Exponential random sum distribution as follows

$$K_Y(s) = K_N(K_X(s)) = \lambda \left(\frac{1}{1 - \beta S} - 1 \right), s \neq \frac{1}{\beta} \tag{17}$$

Then the saddlepoint equation is

$$K'_Y(\hat{s}) = \lambda \beta (1 - \beta \hat{s})^{-2} = x, x \in [0, \infty) \tag{18}$$

Then we can drive the saddlepoint

$$\hat{s} = \frac{1}{\beta} \left(1 - \left[\frac{x}{\lambda \beta} \right]^{-\frac{1}{2}} \right) \tag{19}$$

The second derivative of the cumulant generating function is given by

$$K''_Y(\hat{s}) = 2\lambda \beta^2 (1 - \beta \hat{s})^{-3} \tag{20}$$

Then the saddlepoint density function for the Poisson-Exponential random sum distribution by using equation (3) is in the form

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi (2\lambda \beta)^2 (1 - \beta \hat{s})^{-3}}} \exp \left[\lambda \left(\frac{1}{1 - \beta \hat{s}} - 1 \right) - \frac{1}{\beta} \left(1 - \left(\frac{x}{\lambda \beta} \right)^{-\frac{1}{2}} \right) x \right]$$

Now, let $\hat{F}(x) = P(Y \leq X)$ as given in equation (5) with $\hat{w}$, $\hat{u}$ as in Equation (6) and $\hat{s}$ as in Equation (19). In the previous insurance company example, let $\beta = 1$ and $\lambda = 5$ when $x = .01$. Then we can drive the saddlepoint as $\hat{s} = -21.36067977$. As well as, the cumulant generating function given as $K_Y(\hat{s}) = -4.776393202$. The second derivative of the cumulant generating function is $K''_Y(\hat{s}) = 0.0008944271916$ and $\hat{w} = -3.020856304$, $\hat{u} = -0.6388333294$. Then the saddlepoint cumulative distribution function for the random sum Poisson- Exponential when $x = .01$ is $\hat{F}(.01) = 0.0063977$.

On other hand, we can use the Empirical distribution function to determined exact cumulative distribution function for Poisson- Exponential by simulating 10^6 independent values of Y where N is Poisson (5) and $X_i's$ is Exponential (1) generated by using Matlab program.

Table 1 shows the comparison of the exact probabilities with saddlepoint approximations for Poisson- Exponential distribution, for each X, the first value of each cell of Table 1 is the exact Poisson- Exponential distribution, the second, is the saddlepoint approximations and the last value is the relative error .

Table 1. Compares the exact probabilities with saddlepoint approximations for Poisson- Exponential distribution

| X	F(X)	$\hat{F}(X)$	$\hat{F}(X) - F(X)$	% Relative Error
0.01	0.007239	0.0063977	-0.0008413	11.621771
0.28	0.01817	0.017361	-0.000809	4.45239406
0.55	0.032687	0.032177	-0.00051	1.56025331
0.82	0.051596	0.050571	-0.001025	1.98658811
1.09	0.073048	0.072404	-0.000644	0.88161209
1.36	0.098535	0.097426	-0.001109	1.12548841
1.63	0.12628	0.12532	-0.00096	0.76021539
1.9	0.15703	0.15572	-0.00131	0.8342355
2.17	0.18936	0.18823	-0.00113	0.59674694
2.44	0.22365	0.22244	-0.00121	0.54102392
2.71	0.25885	0.25797	-0.00088	0.33996523
2.98	0.29555	0.2944	-0.00115	0.38910506
3.25	0.33239	0.33137	-0.00102	0.30686844
3.52	0.36951	0.36853	-0.00098	0.2652161
3.79	0.406	0.40556	-0.00044	0.10837438
4.06	0.44295	0.44217	-0.00078	0.17609211
4.33	0.47932	0.47811	-0.00121	0.25244096
4.6	0.51458	0.51317	-0.00141	0.27400987
4.87	0.54816	0.54715	-0.00101	0.18425277
5.14	0.58071	0.5799	-0.00081	0.13948442
5.41	0.61208	0.61131	-0.00077	0.12580055
5.68	0.64138	0.64129	-9E-05	0.01403224
5.95	0.67008	0.66977	-0.00031	0.04626313
6.22	0.6973	0.69671	-0.00059	0.08461208
6.49	0.72243	0.7221	-0.00033	0.04567917
6.76	0.7461	0.74592	-0.00018	0.02412545
7.03	0.76873	0.7682	-0.00053	0.06894488
7.3	0.78998	0.78897	-0.00101	0.12785134
7.57	0.8085	0.80826	-0.00024	0.0296846
7.84	0.82641	0.82614	-0.00027	0.03267143
8.11	0.84245	0.84264	0.00019	0.02255327
8.38	0.85809	0.85784	-0.00025	0.02913447
8.65	0.8725	0.8718	-0.0007	0.08022923
8.92	0.88441	0.88459	0.00018	0.02035255
9.19	0.89656	0.89628	-0.00028	0.03123048
9.46	0.90673	0.90693	0.0002	0.02205728
9.73	0.91691	0.91663	-0.00028	0.03053735
10	0.9259	0.92543	-0.00047	0.05076142
10.27	0.93385	0.9334	-0.00045	0.04818761
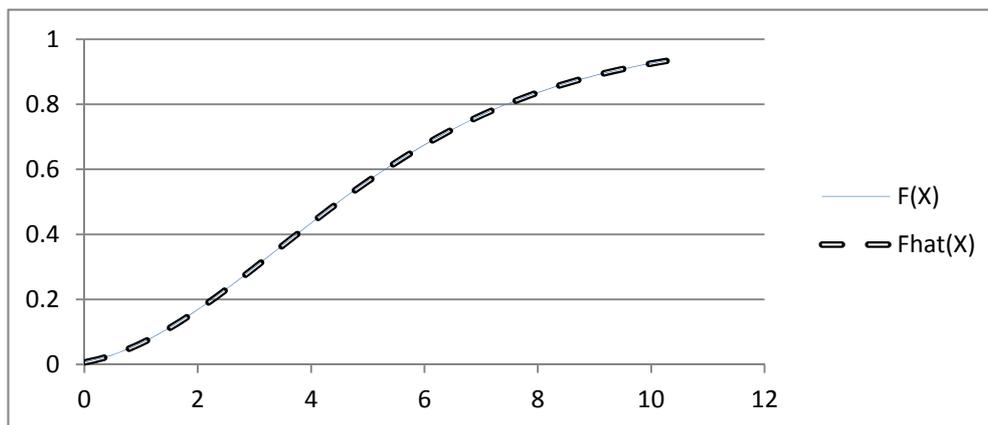
Figure 1. A comparative plot for exact probabilities with saddlepoint approximations for Poisson- Exponential distribution

F hat(x) is the saddlepoin approximation.

F(x) is the exact distribution.

Figure 1 shows a comparative plot of the true cumulative distribution function CDF, $F(x)$ (solid line) for the random sum Poisson (5) Exponential (1) with the saddlepoint approximation $\hat{F}(x)$ (dotted line).

It is clear that, saddlepoint approximation for cumulative distribution function shares the same accuracy with exact and the mean squared error of the saddlepoint approximation is MSE= $0.0^6 604708$ which shows that the saddlepoint approximation is almost exact.

4. Conclusion

This paper introduced saddlepoint approximations to the cumulative distribution function for random sum Poisson- Exponential distributions in continuous settings. We discussed approximations to random sum random variable with dependent components assuming presence of the moment generating function. We used the Empirical distribution function to calculate the exact CDF value by simulation one million independent values of Y. A numerical example of continuous distributions from the Poisson-Exponential distribution was presented. We found that, the saddlepoint approximation for CDF shares the same accuracy with exact CDF. And the mean squared error of the saddlepoint approximation is close to zero.

References

Borowiak, D. S. (1999). A saddlepoint approximation for tail probabilities in collective risk models. *Journal of Actuarial Practice, 7*, 1-11.

Butler, R. W. (2007). *Saddlepoint approximations with applications*. USA: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511619083

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics, 25*, 573-650. http://dx.doi.org/10.1214/aoms/1177728652

Daniels, H. E. (1987). Tail probability approximations. *International Statistical Review, 55*, 37-48. http://dx.doi.org/10.2307/1403269

Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). USA: Collier Macmillan Publisher. http://ar.scribd.com/doc/46835937/Hogg-Craig-Introduction-to-Mathematical-Statistics

Hogg, R. V., & Tanis, E. A. (1983). *Probability and Statistical Inference* (2nd ed.). New York: Macmillan Publishing Co. www.diss.fu-berlin.de/.../09_bibliography.pdf?

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (3rd ed.). pp. 386-388. John Wiley & Sons, Inc.

Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statist. Sci., 3*, 213-238. http://dx.doi.org/10.1214/ss/1177012906

Skovgaard, I. M. (1987). Saddlepoint expansions for conditional distributions. *Journal of Applied Probability, 24*, 875-887. http://dx.doi.org/10.2307/3214212

Terrell, G. R. (2003). A stabilized Lugannani-Rice formula aymposium on the Interface. Computing Science and Statistics. *Annual report department of statistics Virginia tech March 14*. http://www.stat.vt.edu/about/AnnualReport2003.pdf