Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 14 (16): 1889-1893, 2014
ISSN 1812-5654 / DOI: 10.3923/jas.2014.1889.1893

Estimations of the Central Tendency Measures of the Random-sum Poisson-Weibull Distribution using Saddlepoint Approximation

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

Abstract: The random-sum Poisson-Weibull variable is the sum of a random sample from a Weibull distribution with a sample size that is an independent Poisson random variable. It has a wide range of applications. This random sum is complex and difficult to analyze. Saddlepoint approximations are powerful tools for obtaining accurate expressions for closed-form distribution functions for these complex distributions. The use of saddlepoint approximations almost outperforms other methods with respect to computational costs, though not necessarily with respect to accuracy. This study introduces saddlepoint approximations to the cumulative distribution function for the Poisson-Weibull model, from which we can obtain some important statistical measures of the central tendency of a cumulative distribution. We discuss approximations of a random-sum variable using dependent components, assuming the existence of a moment-generating function. Numerical examples of Poisson-Weibull random sums are presented.

Key words: Saddlepoint approximation, random-sum, Poisson-Weibull model, central tendency measures, cumulative distribution

INTRODUCTION

Saddlepoint approximations are powerful tools for obtaining accurate terms for distribution functions that are difficult to obtain in closed form. Saddlepoint approximations almost surpass other techniques with respect to computational cost, although it does not necessarily surpass other techniques with respect to accuracy.

RANDOM-SUM DISTRIBUTIONS

Random-sum distributions have many natural applications and appear frequently in probability theory and applications. For example, these distributions have a wide range of applications in branching processes (Neyman, 1939), renewal processes, damage processes (Rao *et al*., 1980), stopped random walks (Malinovskii, 1994) and risk theory (Esscher, 1932; Jensen, 1995; Gurland, 1957). One of the most important families of random sums is the family of Poisson random sums, in which N is a Poisson (λ) random variable and the X_i's are independent and identically distributed. The random variable Y_N is said to have a random distribution if Y is of the following form (Johnson *et al*., 2005) Eq. 1:

$$Y_N = X_1 + X_2 + X_3 + \ldots + X_N \quad (1)$$

where, the number of terms N is uncertain, the random variables X_i are independent and identically distributed (with a common distribution V) and each X_i is independent of N. If N = 0, then we have Y = 0. Although, this is implicit in the definition, we want to call attention to this point for clarity. The distribution function of Y_N is given by the following Eq. 2:

$$F_{Y_N}(y) = \sum_{n=0}^{\infty} G_n(y) P[N = n] \quad (2)$$

where, for $n \geq 1$, $G_n(y)$ is the distribution function of the independent sum $X_1 + X_2 + X_3 + \ldots + X_N$. We can also express Y_N in terms of convolutions (Eq. 3):

$$f_{Y_N}(y) = \sum_{n=0}^{\infty} f^{*n}(y) P[N = n] \quad (3)$$

where, f is the common distribution function for X_i and f^{*n} is the n-fold convolution of f. If the common distribution X is discrete, then the random sum Y_N is discrete. On the other hand, if X is continuous and if P[N = n]>0, then the random sum Y has a mixed distribution. The mean of the random sum Y_N is as follows (Eq. 4):

$$E[Y_N] = E[N]\, E[X] \quad (4)$$

Corresponding Author: O. Al Mutairi Alya, Department of Applied Statistics, Faculty of Applied Science, Taibah University Al-Madinah, Kingdom of Saudi Arabia

The expected value of the mean has a natural interpretation. It is the product of the expected number of events N and the expected individual distribution X. This makes intuitive sense. The variance of the random sum is as follows (Eq. 5):

$$\text{Var}\,[Y_N] = E[N]\,\text{Var}[X]+\text{Var}[N]\,E[X]^2 \qquad (5)$$

The moment-generating function of the random sum Y_N is defined as follows (Eq. 6) (Hogg and Tanis, 1983):

$$M_{Y_N}(s) = M_N[\ln M_X(s)] \qquad (6)$$

where, the function In is the natural log function. The cumulant-generating function is defined as follows in Eq. 7:

$$K_{Y_N}(s) = \ln M_{Y_N}(s) = \ln M_N[K_X(s)] = K_N[K_X(s)] \qquad (7)$$

SADDLEPOINT APPROXIMATION TO DENSITIES AND MASS FUNCTIONS

The most basic saddlepoint approximation was introduced by Daniels (1954) and is fundamentally a formula for approximating the density and mass function from an associated moment-generating function. Saddlepoint approximations are constructed by assuming the existence of the Moment-generating Function (MGF) or equivalently, the Cumulant-generating Function (CGF), of a random variable. For improvements to the saddlepoint methodology and associated techniques, the reader is referred to Daniels (1954, 1987) for details concerning density and mass approximation, Skovgaard (1987) for a conditional version of this approximation, Reid (1988) for applications to inference, Borowiak (1999) for discussion of a tail-area approximation with a uniform relative error and Terrell (2003) for a stabilized lugannani-rice formula.

Suppose a random variable X has density function f(x) identified for all real values of x. The MGF is defined as follows (Hogg and Craig, 1978) (Eq. 8):

$$M_X(s) = E(e^{SX}) = \int_{-\infty}^{\infty} e^{SX} f(x)dx \qquad (8)$$

Wherever this expectation exists, $M_x(0)$ always exists and is equal to 1. We shall assume that $M_x(s)$ converges over the largest open neighborhood (a, b) at zero. The cumulant-generating function CGF is given by the following Eq. 9:

$$K_X(s) = \ln M_X(s),\ s\epsilon(a, b) \qquad (9)$$

For a continuous random variable X with CGF $K_x(s)$ and unknown density f(x), the saddlepoint approximation density of f(x) is given by the following Eq. 10:

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi K''_X(\hat{s})}} \exp\left(K_x(\hat{s}) - \hat{s}x\right) \qquad (10)$$

where, $\hat{s} = \hat{s}(x)$ denotes the unique solution to the saddlepoint equation $K'(\hat{s}) = x$ and $K''_x(\hat{s})$ is the second derivative (Daniels, 1954). This approximation is useful for values of x that are within the point of support $\{x: f(x)>0\} = \chi$.

The saddlepoint approximation for univariate cumulative distribution functions F(x) is given by the following Eq. 11:

$$\hat{F}(x) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})\left(\frac{1}{\hat{w}} - \frac{1}{\hat{u}}\right) & \text{if} \quad x \neq \mu \\ \frac{1}{2} + \frac{K''_X(0)}{\sqrt[6]{2\pi}K''_X(0)^{\frac{3}{2}}} & \text{if} \quad x = \mu \end{cases} \qquad (11)$$

where, the continuous random variable X has CDF F(x) and CGF CGF$K_x(s)$ with mean $\mu = E(x)$ and $\hat{w}$ and $\hat{u}$ are defined as follows (Eq. 12):

$$\begin{aligned} \hat{w} &= \text{sgn}(\hat{s})\sqrt{2\{\hat{s}x - K_X(\hat{s})\}} \\ \hat{u} &= \hat{s}\sqrt{K''_X(\hat{s})} \end{aligned} \qquad (12)$$

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

APPLICATIONS OF THE RANDOM-SUM POISSON-WEIBULL DISTRIBUTION

The Weibull distribution is a continuous probability distribution with the following probability density function (Eq. 13):

$$f(x) = \begin{cases} \frac{k}{\eta}\left(\frac{x}{\eta}\right)^{k-1} e^{-(x/\lambda)} & x \geq 0, \\ 0 & x < 0, \end{cases} \qquad (13)$$

where, k>0 is a shape parameter and η is a scale parameter of the distribution. The mean and the variance of a Weibull random variable can be expressed as follows:

$$E(x) = \eta\Gamma(1+1/k)$$
$$V(x) = \eta^2\Gamma(1+2/k)-\mu^2$$

The MGF is defined as follows:

$$M_X(s) = \sum_{n=0}^{\infty} \frac{s^n \eta^n}{n!} \Gamma\left(1 + \frac{n}{k}\right),\ k \geq 1$$

The random-sum Poisson-Weibull variable has the following form (Eq. 14):

$$Y_N = \sum_{j=1}^{N} x_j = x_1 + x_2 + x_3 + ... + x_N \tag{14}$$

where, the sample size N follows a Poisson (λ) distribution and the X_i's are i.i.d. random variables that follow a Weibull distribution. The sum Y_N is said to have a random-sum Poisson-Weibull distribution. The exact calculation of this distribution is very complex and difficult. The saddlepoint approximation method overcomes this problem. This method is based on the moment-generating function for the random sum. The cumulant-generating function for N is given by the following Eq. 15:

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

For X_i's that are i.i.d. random variables following a Weibull distribution, the cumulant-generating function is defined as follows (Eq. 16):

$$K_X(s) = \ln \sum_{n=0}^{\infty} \frac{s^n \eta^n}{n!} \Gamma\left(1 + \frac{n}{k}\right),\ k \geq 1 \tag{16}$$

This leads to the cumulant-generating function for the random-sum Poisson-Weibull distribution, which takes the following form (Eq. 17):

$$K_{Y_N}(s) = K(K_X(s)) = \lambda[\sum_{n=0}^{\infty} \frac{s^n \eta^n}{n!} \Gamma(1 + \frac{n}{k}) - 1] \tag{17}$$

The saddlepoint equation is as follows (Eq. 18):

$$K'(\hat{s}) = \lambda[\sum_{n=1}^{\infty} \frac{\hat{s}^{(n-1)} \eta^n}{(n-1)!} \Gamma(1 + \frac{n}{k})] = x \tag{18}$$

Saddlepoint computation involves first and second derivatives. The saddlepoint solution is a root of the first derivative. The unique real root $\hat{s} = \hat{s}(x)$ can be determined numerically. The second derivative of the cumulant-generating function is given by the following Eq. 19:

$$K''_{Y_N}(\hat{s}) = \lambda[\sum_{n=2}^{\infty} \frac{\hat{s}^{(n-2)} \eta^n}{(n-2)!} \Gamma(1 + \frac{n}{k})] = x, \tag{19}$$

Using Eq. 10, the saddlepoint density function for the random-sum Poisson-Weibull distribution can be expressed in the following form (Eq. 20):

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi K''_{Y_N}(\hat{s})}} \exp(K_{Y_N}(\hat{s}) - \hat{s}x) \tag{20}$$

where, $K''_{Y_N}(\hat{s})$ is given by Eq. 19 and $K_{Y_N}(\hat{s})$ is the cumulant-generating function at the value of saddlepoint $\hat{s}$.

Because of its flexible shape and its ability to model a wide range of models, the Weibull distribution has been used successfully in many applications, such as the modeling of wind speeds.

Wind speeds in most places in the world can be modeled using a Weibull distribution. This statistical tool tells us how often winds of different speeds will be observed at a location with a certain average (mean) wind speed. Knowing this helps us to choose a wind turbine with the optimal cut-in speed (the wind speed at which the turbine starts to generate usable power) and the cut-out speed (the speed at which the turbine hits the limit of its alternator and can no longer put out power with further increases in wind speed).

The shape of the Weibull distribution depends on a parameter called (helpfully) shape. In Northern Europe and most other locations around the world, the value of shape is approximately 2. The shape parameter will typically range from 1-3. For a given average wind speed.

Let N be the number of wind events that occur in one country or region during a given time period and let X be the amount of energy a wind turbine produces over a given time. Y_N gives the total amount of energy a wind turbine produces over a given time period.

Now, let $\hat{F}(x) = P(Y_N \leq X)$, as given in Eq. 11, with $\hat{w}$ and $\hat{u}$ as given in Eq. 12 and $\hat{s}$ as given in Eq. 19. Let (k = 1 shape parameter), λ = 5 and η = 4 when x = 0.02. Then, we can determine the value of the saddlepoint to be:

$$\hat{s} = -7.65569415$$

The cumulant-generating function value is:

$$K_Y(\hat{s}) = -4.841886117$$

The second derivative of the cumulant-generating function is:

$$K''_Y(\hat{s}) = 0.005059644256$$

and:

Table 1: Compares the exact probabilities with saddlepoint approximations for Poisson-Weibull distribution

X	F(X)	$\hat{F}$(X)	Relative error (%)
0.02	0.006833	0.0066384	2.847944
0.42	0.010509	0.0096275	8.388048
0.82	0.014809	0.013875	6.306975
1.22	0.019334	0.018582	3.889521
1.62	0.024341	0.023774	2.329403
2.02	0.030327	0.029458	2.865433
2.42	0.036398	0.035636	2.093522
2.82	0.042938	0.042305	1.474219
3.22	0.050316	0.049457	1.707210
3.62	0.058210	0.057083	1.936093
4.02	0.066002	0.065173	1.256023
4.42	0.074604	0.073713	1.194306
4.82	0.083911	0.082689	1.456305
5.22	0.092557	0.092085	0.509956
5.62	0.10309	0.10189	1.164031
6.02	0.11323	0.11207	1.024463
6.42	0.12341	0.12262	0.640143
6.82	0.13446	0.13353	0.691656
7.22	0.14552	0.14476	0.522265
7.62	0.15756	0.1563	0.799695
8.02	0.16886	0.16813	0.432311
8.42	0.18079	0.18023	0.309752
8.04	0.17003	0.16873	0.764571
8.44	0.18230	0.18084	0.800878
8.84	0.19418	0.1932	0.504686
9.24	0.20700	0.20578	0.589372
9.64	0.21993	0.21857	0.618379
10.04	0.23262	0.23154	0.464277
10.44	0.24570	0.24468	0.415140
10.84	0.25968	0.25797	0.658503
11.24	0.27266	0.27138	0.469449
11.64	0.28625	0.28489	0.475109
12.04	0.29943	0.29849	0.313930
12.44	0.31362	0.31216	0.465532
12.84	0.32685	0.32588	0.296772
13.24	0.34064	0.33963	0.296501
13.64	0.35480	0.35339	0.397407
14.04	0.36854	0.36716	0.374451
14.44	0.38192	0.38091	0.264453
14.84	0.39616	0.39462	0.388732
15.24	0.40910	0.40829	0.197996
15.64	0.42238	0.4219	0.113642
16.04	0.43689	0.43544	0.331891
16.44	0.45000	0.44889	0.246667
16.84	0.46347	0.46224	0.265389
17.24	0.47605	0.47548	0.119735
17.64	0.49014	0.4886	0.314196
18.04	0.50275	0.50159	0.230731
18.44	0.51553	0.51445	0.209493
18.84	0.52774	0.52715	0.111797
20.12	0.56800	0.56671	0.227113
22.12	0.62513	0.62482	0.049590
24.12	0.67771	0.67792	-0.030990
26.12	0.72677	0.72572	0.144475
28.12	0.76764	0.7682	-0.072950
30.12	0.80581	0.8055	0.038471
32.12	0.83805	0.83789	0.019092
34.12	0.86611	0.86574	0.042720
36.12	0.88950	0.88948	0.002248
38.12	0.90969	0.90954	0.016489
40.12	0.92678	0.92635	0.046397
41.82	0.93878	0.93842	0.038348
43.82	0.95055	0.95033	0.023144
45.82	0.95990	0.96012	-0.022920
47.82	0.96805	0.96813	-0.008260
49.82	0.97476	0.97464	0.012311

Table 1: Continue

X	F(X)	$\hat{F}$ (X)	Relative error (%)
51.82	0.97978	0.97990	-0.012250
53.82	0.98421	0.98413	0.008128
55.82	0.98734	0.98752	-0.018230
57.82	0.99035	0.99022	0.013127
59.82	0.99236	0.99236	0
61.82	0.99415	0.99405	0.010059
63.82	0.99541	0.99539	0.002009
65.82	0.99642	0.99643	-0.001

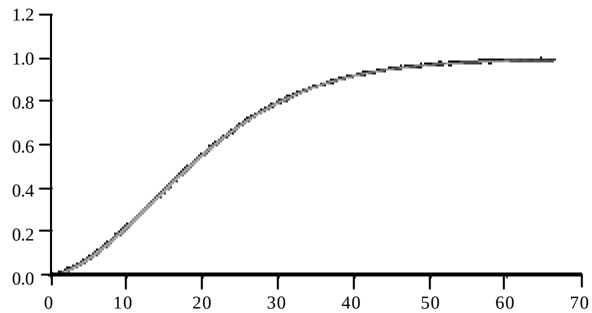

Fig. 1: Comparison of the exact CDF for a Poisson-Weibull distribution (solid line) vs. its saddlepoint approximation (dashed line)

$$\hat{w} = -3.06227766,\ \hat{u} = -0.544558531$$

Then, the saddlepoint cumulative distribution function value for the random-sum Poisson-Weibull distribution when x = 0.02 is:

$$\hat{F}(0.02) = 0.00663$$

We can use the empirical distribution function to determine the exact cumulative distribution function for the Poisson-Weibull distribution by using the MATLAB program to simulate 10^6 independent values of Y_N, where N is Poisson (5) and the X_i's are generated from a Weibull (1, 4) distribution. Table 1 compares the exact probabilities with saddlepoint approximations for Poisson-Weibull distribution, for each X, the first value of each cell of Table 1 is the exact Poisson-Weibull distribution, the second is the saddlepoint approximations and the third value is the relative error.

Figure 1 shows that the saddlepoint approximation of the CDF has the same accuracy as the exact CDF. The mean squared error of the saddlepoint approximation is MSE = $0.0^7 83707$, which indicates that the saddlepoint approximation is almost exact. From the CDF, we can drive various statistical measures of central tendency, such as the mean, the median and the mode. The median is the inverse of the CDF at 0.5. Using the MATLAB program, we find that $\Phi^{-1}(0.5) = 18$. From Eq. 4, we find that the mean = 20. Because the distribution is asymmetric, we use the formula below to drive the mode:

$$\text{Mode} = 3\text{median}-2\text{mean} = 54-40 = 14$$

CONCLUSION

This study introduced saddlepoint approximations to the cumulative distribution function for random-sum Poisson-Weibull distributions in continuous settings. We discussed approximations to random-sum random variables with dependent components assuming the presence of a moment-generating function. Measures of central tendency were derived. We used an empirical distribution function to calculate the exact CDF value by simulation of one million independent values of Y_N. A numerical example of a continuous distribution function for a Poisson-Weibull distribution was presented. We found that the saddlepoint approximation for CDF yields the same accuracy as the exact CDF and that the mean squared error of the saddlepoint approximation was close to zero.

ACKNOWLEDGMENTS

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

REFRENCES

Borowiak, D.S., 1999. A saddlepoint approximation for tail probabilities in collective risk models. J. Actuarial Pract., 7: 1-11.

Butler, R.W., 2007. Saddlepoint Approximations with Applications. Cambridge University Press, USA.

Daniels, H.E., 1954. Saddlepoint approximations in statistics. Ann. Math. Stat., 25: 631-650.

Daniels, H.E., 1987. Tail probability approximations. Int. Stat. Rev., 55: 37-48.

Esscher, F., 1932. On the probability function in the collective theory of risk. Scand. Actuarial J., 15: 175-195.

Gurland, J., 1957. Some interrelations among compound and generalized distributions. Biometrika, 44: 265-268.

Hogg, R.V. and A.T. Craig, 1978. Introduction to Mathematical Statistics. 4th Edn., Collier Macmillan Publisher, USA.

Hogg, R.V. and E.A. Tanis, 1983. Probability and Statistical Inference. 2nd Edn., Macmillan Publishing Co., New York, USA., ISBN-13: 978-0023557309.

Jensen, J.L., 1995. Saddlepoint Approximations. Clarendon Press, Oxford.

Johnson, N.L., A.W. Kemp and S. Kotz, 2005. Univariate Discrete Distributions. 3rd Edn., John Wiley and Sons, New York, ISBN-13: 9780471715801, pp: 386-388.

Malinovskii, V.K., 1994. Limit theorems for stopped random sequences, I: Rates of convergence and asymptotic expansions. Th. Prob. Appl., 38: 673-693.

Neyman, J., 1939. On a new class of contagious distributions applicable in entomology and bacteriology. Ann. Math. Statist., 10: 35-57.

Rao, C.R., R.C. Srivastava, S. Talwalker and G.A. Edgar, 1980. Characterization of probability distributions based on a generalized Rao-Rubin condition. Sankhya A, 42: 161-169.

Reid, N., 1988. Saddlepoint methods and statistical inference. Stat. Sci., 3: 213-227.

Skovgaard, I.M., 1987. Saddlepoint expansions for conditional distributions. J. Applied Probab., 24: 875-887.

Terrell, G.R., 2003. A stabilized Lugannani-Rice formula. Proceedings of the Symposium on the Interface: Computing Science and Statistics, March 12-15 2003, Salt Lake, USA.