Advances and Applications in Statistics

Published Online: September 2016
http://dx.doi.org/10.17654/AS049030231
Volume 49, Number 3, 2016, Pages 231-243 **ISSN: 0972-3617**

SADDLE POINT APPROXIMATION FOR THE RANDOM SUM POISSON-BERNOULLI MODEL

O. Al Mutairi Alya, Ameenah Alofi, Mai Alsoubhi, Sara Alradadi and Huda Alblwi

Department of Mathematics
Faculty of Science
Taibah University
Medinah, Saudi Arabia
e-mail: afaaq99@hotmail.com

Abstract

Approximations are very important because it is sometimes not possible to precisely represent exact representation, while in some cases the exact answer is already obtained but is very difficult to apply, as well the approximations sometimes simplify the analytical treatments. Compared with other asymptotic approximations, saddle point approximations have the advantage of always generating probabilities, being very accurate in the tails of the distribution, and being accurate with small samples, sometimes even with only one observation. In this paper, saddle point approximation methods have been proven to be useful for a range of problems, such as the random sum statistics (Poisson-Bernoulli) model which is very complex model.

Received: May 13, 2016; Accepted: June 20, 2016
2010 Mathematics Subject Classification: Primary 62F15; Secondary 65.
Keywords and phrases: approximations, saddle point approximations, random sum distribution, Poisson-Bernoulli.
Communicated by K. K. Azad

1. Introduction

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, say X's. In health science, the random sum plays a very important role in many real-life applications. For example, let the number of hotbeds of a contagious disease follow a Poisson distribution with a mean of λ, and let the number of sick people within the hotbed follow a Binomial distribution. If the goal is to find the probability that the total number of sick people is greater than 70, then the total number of sick people within the hotbed is:

$$S_{N_1} = \sum_{i=1}^{N} X_i, \tag{1}$$

where $X_i \sim \text{Binomial}(n,\ p)$ and $N \sim \text{Poisson}(\lambda)$.

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 Bernoulli distribution and if the rain falling in that time period is independent of N, then the total rainfall in the time period is:

$$S_{N_2} = \sum_{i=1}^{N} Y_i, \tag{2}$$

where $Y_i \sim \text{Bernoulli}(p)$ and $N \sim \text{Poisson}(\lambda)$.

In fact, the total amounts of the 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 [1]. The saddle point approximation method can help us gain knowledge for these unknown difficult statistical behaviours.

2. Derivation of the Saddle Point CDFs for the Random Sum Poisson-Bernoulli Model

To develop new estimators CDFs using saddle point approximations for this model. Let we have the random sum $S_N = X_1 + X_2 + \cdots + X_N$, where X_i's are independent of *N*, and $N \sim$ Poisson (λ), X's $\sim$ Bernoulli(p).

The MGF for Poisson distribution is given by

$$M_N(s) = e^{\lambda(e^s - 1)}. \tag{3}$$

As well the MGF for Bernoulli distribution is given by

$$M_X(s) = (pe^s + q). \tag{4}$$

The cumulant generating function for *N* is given by

$$K_N(s) = \ln(M_N(s)),$$

$$K_N(s) = \ln e^{\lambda(e^s - 1)}$$

$$= \lambda(e^s - 1). \tag{5}$$

The cumulant generating function for *X* is given by

$$K_X(s) = \ln(M_X(s)), \tag{6}$$

$$K_X(s) = \ln(pe^s + q), \tag{7}$$

where $q + p = 1$, then we can get the cumulant generating function for the Poisson-Bernoulli random sum distribution as follows:

$$K_{S_N}(s) = K_N(K_X(s)),$$

$$K_{S_N}(s) = \lambda(e^{\ln(pe^s + q)} - 1)$$

$$= \lambda(pe^s + q - 1) \tag{8}$$

the saddle point is given as

$$K'_{S_N}(s) = \lambda p e^{\hat{s}} = x, \tag{9}$$

$$e^{\hat{s}} = \frac{x}{\lambda p},$$

$$\hat{s} = \ln \frac{x}{\lambda p}$$

and

$$K''_{S_N}(s) = \lambda p e^{\hat{s}}.$$

This leads to the saddle point mass function for Poisson-Bernoulli which is given by

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi\lambda p e^{\hat{s}}}} e^{\left\{\lambda(pe^{\hat{s}} + q - 1) - x \ln \frac{x}{\lambda p}\right\}}. \tag{10}$$

The first continuity-correction is

$$\hat{w}_1 = \operatorname{sgn}(\hat{s})\sqrt{2\left\{x\left(\ln \frac{x}{\lambda p}\right) - \lambda(pe^{\hat{s}} + q - 1)\right\}}, \tag{11}$$

$$\hat{u}_1 = \{1 - e^{-\left(\ln \frac{x}{\lambda p}\right)}\}\sqrt{\lambda p e^{\hat{s}}}.$$

The second continuity-correction is

$$\tilde{w}_2 = \operatorname{sgn}(\tilde{s})\sqrt{2\left\{\left(\ln \frac{x}{\lambda p}\right)x^{-} - \lambda(pe^{\tilde{s}} + q - 1)\right\}}, \tag{12}$$

where $x^{-} = x - 0.5,$ then

$$\tilde{u}_2 = 2\mathrm{Sinh}\left(\frac{\left(\ln \frac{x}{\lambda p}\right)}{2}\right)\sqrt{\lambda P e^{\tilde{s}}}. \tag{13}$$

And the third continuity-correction is

$$\tilde{u}_3 = \tilde{s}\sqrt{\lambda p e^{\tilde{s}}},$$

this approximation uses the same second continuity-correction while $\tilde{u}_3$ changed.

3. Saddle Point Approximation with Application to the Poisson Random Sum Distribution

The Poisson random sum distribution is the sum of a random sample from a certain distribution (continuous or discrete) with a sample size that is independent of the Poisson random variable. This sum has wide-ranging applications in fields such as insurance (e.g., the total claim size in a portfolio), meteorology, mall visit frequencies, and mortality data.

4. Saddle Point Approximation with Application to the Random Sum Poisson (λ)-Bernoulli (*p*) Model

This section provides numerical results that can be used to compare the accuracies of the various saddle point approximations in discrete cases, and considered a saddle point approximation [2] and [3] for the distribution of this random sum statistic. As an example, suppose that N is Poisson (2), and let X be the Bernoulli random variable, with $p = 0.3$. Then, the $S_{N(t)}$ is a random sum Poisson process with the following form:

$$S_{N(t)} = \sum_{j=1}^{N(t)} X_j, \quad t > 0. \tag{14}$$

As previously indicated, random sum Poisson processes are very complicated and difficult to analyze, and therefore, approximation methods are often used. The previous section identified the moment-generating function, which leads to the explicit cumulant generating function. Now let $x = 1, \lambda = 2, p = 0.3$ so, we get the saddle point

$$\begin{aligned}\hat{s} &= \ln\left(\frac{x}{\lambda p}\right)\\ &= \ln\left(\frac{1}{2(0.3)}\right)\\ &= 0.5108 \end{aligned} \tag{15}$$

and

$$\begin{aligned}K(\hat{s}) &= \lambda(pe^{\hat{s}} + q - 1)\\ &= 2(0.3e^{0.5108} + 0.7 - 1)\\ &= 0.3999. \end{aligned} \tag{16}$$

Also

$$\begin{aligned}K''(\hat{s}) &= \lambda pe^{\hat{s}}\\ &= 2(0.3)e^{0.5108}\\ &= 0.9999\end{aligned}$$

first continuity-correction

$$\begin{aligned}\hat{w}_1 &= \operatorname{sgn}(\hat{s})\sqrt{2\{\hat{s}x - k(\hat{s})\}}\\ &= +\sqrt{2\{0.5108(1) - 0.3999\}}\\ &= 0.4709\end{aligned} \tag{17}$$

and

$$\begin{aligned}\hat{u}_1 &= \{1 - e^{-\hat{s}}\}\sqrt{k''(\hat{s})}\\ &= \{1 - e^{-(0.5108)}\}\sqrt{0.9999},\end{aligned} \tag{18}$$

$$\hat{p}_{r1}(X \geq x) = 1 - \Phi(\hat{w}) - \phi(\hat{w})\left(\frac{1}{\hat{w}_1} - \frac{1}{\hat{u}_1}\right).$$

We can find the $\Phi(\hat{w}_1)$ from the normal CDF table,

$$\Phi(0.4709) = 0.6808.$$

And $\Phi(\hat{w}_1)$ by using the normal pdf distribution,

$$\phi(0.4709) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(0.4709)^2}$$

$$= 0.3570$$

Where the first approximation is

$$\hat{p}_{r1}(X \geq x) = 1 - 0.6808 - 0.3570\left(\frac{1}{0.4709} - \frac{1}{0.3999}\right)$$

$$= 0.4538 \tag{19}$$

While the second is given by $x^{-} = x - 0.5 = 1 - 0.5 = 0.5.$

Then

$$k'(\tilde{s}) = x - 0.5 = 0.5$$

$$\Rightarrow \lambda p e^{\tilde{s}} = 0.5,$$

$$\tilde{s} = \ln\left(\frac{0.5}{\lambda p}\right)$$

$$= \ln\left(\frac{0.5}{(2)(0.3)}\right)$$

$$= -0.1823$$

and

$$k(\hat{s}) = \lambda(pe^{\hat{s}} + q - 1)$$

$$= 2(0.3e^{-0.1823} + 0.7 - 1)$$

$$= -0.0999.$$

Also

$$k''(\hat{s}) = \lambda p e^{\hat{s}}$$

$$= 2(0.3)e^{-0.1823}$$

$$= 0.5000.$$

Then

$$\hat{w}_2 = \operatorname{sgn}(\hat{s})\sqrt{2\{\hat{s}x^{-} - k(\hat{s})\}}$$

$$= -\sqrt{2\{(-0.1823)(0.5) - (-0.0999)\}}$$

$$= -0.1322 \tag{20}$$

and

$$u_2 = 2\sinh\left(\frac{\tilde{s}}{2}\right)\sqrt{k''(\tilde{s})}$$

$$= 2\left(\frac{1}{2}(e^{\frac{\tilde{s}}{2}} - e^{-\frac{\tilde{s}}{2}})\right)\sqrt{k''(\tilde{s})}$$

$$= 2\left(\frac{1}{2}(e^{-0.09115} - e^{-(-0.09115)}\right)\sqrt{0.5000}$$

$$= -0.1290,$$

$$\hat{p}_{r2}(x \geq x) = 1 - \Phi(\hat{w}_2) - \phi(\hat{w}_2)\left(\frac{1}{\hat{w}_2} - \frac{1}{\hat{u}_2}\right). \tag{21}$$

We can find the $\Phi(\hat{w}_1)$ from the CDF table

$$\Phi(-0.1322) = 0.4483$$

and $\phi(\hat{w}_1)$ is given by

$$\phi(-0.1322) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(-0.1322)^2}$$

$$= 0.3954,$$

$$\hat{p}_{r2}(X \geq x) = 1 - 0.44 - 0.3954\left(\frac{1}{-0.1322} - \frac{1}{-0.1290}\right)$$

$$= 0.4775. \tag{22}$$

Third continuity-correction

$$\tilde{u}_3 = \tilde{s}\sqrt{k''(\tilde{s})}$$

$$= -0.1823\sqrt{0.5000}$$

$$= -0.1289,$$

$$\hat{p}_{r3}(X \geq x) = 1 - \Phi(\hat{w}_2) - \phi(\hat{w}_2)\left(\frac{1}{\hat{w}_2} - \frac{1}{\hat{u}_3}\right)$$

$$= 1 - 0.44 - 0.3954\left(\frac{1}{-0.1322} - \frac{1}{-0.1289}\right)$$

$$= 0.4751. \tag{23}$$

Now we use normal approximation to investigate the performance of saddle point method for the first approximation

$$E(X) = p,\ E(N) = \lambda,$$

$$E_{S_N}(s) = E(N)E(X)$$

$$= p\lambda$$

$$= (0.3)(2)$$

$$= 0.6,$$

where $V_{S_N}(s)$ is given by

$$V(x) = pq,\ V(N) = \lambda,$$

$$V_{S_N}(s) = E(N)Var(X) + (E(X))^2 Var(N)$$

$$= (2)(0.3)(0.7) + ((0.3))^2(2)$$

$$= 0.6. \tag{24}$$

From the normal approximation, we get

$$cdf\left(\frac{x-\mu}{\sigma}\right) = cdf\left(\frac{1-0.6}{\sqrt{0.6}}\right)$$

$$= cdf(0.5163),$$

$$\hat{p}_r(X \geq x) = 1 - 0.4483$$

$$= 0.5517.$$

Now we use normal approximation also to investigate the performance of saddle point method for the second approximation as

$$E(X) = p, \; E(N) = \lambda,$$

$$E_{S_N}(s) = E(N)E(X)$$

$$= p\lambda$$

$$= (0.3)(2)$$

$$= 0.6,$$

where $V_{S_N}(s)$ is found as:

$$V(x) = pq, V(N) = \lambda,$$

$$V(S) = E(N)Var(X) + (E(X))^2 Var(N)$$

$$= (2)(0.3)(0.7) + ((0.3))^2(2)$$

$$= 0.6. \tag{25}$$

Then

$$\hat{p}_r(X \geq x) = 1 - cdf\left(\frac{0.5-0.6}{\sqrt{0.6}}\right) = 1 - 0.5422$$

$$= 0.5478.$$

The MATLAB program is used to obtain various CDF approximations for any value in support of x. The exact probabilities are obtained using the empirical distribution with generating one million observations.

Table 1 shows the continuity-corrected CDF approximations for the random sum Poisson (2)-Bernoulli (0.3) model with its corresponding exact and normal approximation for 10 different values of x.

Table 1. Continuity-corrected CDF approximations for the random sum Poisson (2)-Bernoulli (0.3) model

	Exact	**First-corrected**		**Second-corrected**		**Third-corrected**	
x	$P(S_N \geq x)$	$\hat{P}_1(S_N \geq x)$	$P'(S_N \geq x)$	$\hat{P}_2(S_N \geq x)$	$P'(S_N \geq x)$	$\hat{P}_3(S_N \geq x)$	$P'(S_N \geq x)$
1	0.4533	0.4538	0.5517	0.4775	0.5478	0.4751	0.5478
2	0.1222	0.1239	0.0354	0.1250	0.1226	0.13200	0.1226
3	0.221	0.2330	0.0010	0.0238	0.0071	0.02655	0.0071
4	0.0501	0.0616	0.0000	0.00350	0.0001	0.0040	0.0001
5	0.0511	0.05004	0.0000	0.05009	0.0000	0.050900	0.0000
6	$1.755*10^{-5}$	$1.855*10^{-5}$	0.0000	$6.715*10^{-4}$	0.0000	$6.898*10^{-4}$	0.0000
7	$8.663*10^{-6}$	$8.798*10^{-6}$	0.0000	$2.657*10^{-7}$	0.0000	$1.745*10^{-6}$	0.0000
8	$7.000*10^{-7}$	$7.027*10^{-7}$	0.0000	$4.657*10^{-13}$	0.0000	$3.194*10^{-14}$	0.0000
9	$4.129*10^{-8}$	$4.952*10^{-8}$	0.0000	$1.532*10^{-8}$	0.0000	$9.771*10^{-9}$	0.0000
10	$3.021*10^{-9}$	$3.122*10^{-9}$	0.0000	$9.727*10^{-10}$	0.0000	$6.162*10^{-10}$	0.0000

To investigate the performance of this method we used different parameters for this model given in Table 2.

Table 2. Continuity-corrected CDF approximations for the random sum Poisson (3)-Bernoulli (0.4) model

	Exact	**First-corrected**		**Second-corrected**		**Third-corrected**	
x	$P(S_N \geq x)$	$\hat{P}_1(S_N \geq x)$	$P'(S_N \geq x)$	$\hat{P}_2(S_N \geq x)$	$P'(S_N \geq x)$	$\hat{P}_3(S_N \geq x)$	$P'(S_N \geq x)$
1	0.6534	0.6515	0.5714	0.7076	0.7357	0.6352	0.7357
2	0.5112	0.5000	0.2327	0.4209	0.1190	0.3436	0.1190
3	0.1219	0.1218	0.0054	0.12251	0.0183	0.1269	0.0183
4	0.02998	0.0343	0.0003	0.0194	0.0013	0.02468	0.0013

5	0.03110	0.03005	0.0000	$7.802*10^{-3}$	0.0000	$8.499*10^{-3}$	0.0000
6	$5.888*10^{-3}$	$5.986*10^{-3}$	0.0000	$2.350*10^{-3}$	0.0000	$1.722*10^{-3}$	0.0000
7	$8.999*10^{-4}$	$8.908*10^{-4}$	0.0000	$2.776*10^{-4}$	0.0000	$3.123*10^{-4}$	0.0000
8	$1.266*10^{-5}$	$1.370*10^{-5}$	0.0000	$1.994*10^{-5}$	0.0000	$1.400*10^{-5}$	0.0000
9	$2.023*10^{-6}$	$2.135*10^{-6}$	0.0000	$1.136*10^{-4}$	0.0000	$3.443*10^{-6}$	0.0000
10	$2.322*10^{-6}$	$2.312*10^{-6}$	0.0000	$1.3941*10^{-5}$	0.0000	$1.009*10^{-6}$	0.0000

5. The Performance in Discrete Distributions

This section presents the performance of continuity-corrected CDF saddle point approximations in a discrete distribution for the Poisson (λ)-Bernoulli (p) model. Tables 1 and 2 show the three saddle point approximations for the random sum Poisson-Bernoulli model, corresponding to the exact and the normal approximation. The numerical results indicate in general that the saddle point approximation is accurate and seems to suggest that both $\hat{P}_1(S_N \geq x)$ and $\hat{P}_2(S_N \geq x)$ are consistently accurate with $\hat{P}_3(S_N \geq x)$. Throughout the study of this issue for example [4], no final decision could arrive as to which approximation $\hat{P}_1$, $\hat{P}_2$ and $\hat{P}_3$ are better. Ultimately, this choice depends upon the application. In certain cases, it is found that $\hat{P}_1(S_N \geq x)$ is better than $\hat{P}_2(S_N \geq x)$; and, in others, it is found to be otherwise. Still, in most cases, $\hat{P}_3(S_N \geq x)$ is better than either of the former options. In [6], it is found, in general, that the second and third corrections are better than the first one. For this reason, this study suggests to calculate $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_3$. If all are functioning well and share the same accuracy (very close together), then their accuracy is supported. However, if they differ, a choice must be made depending on the particular application at hand [7]. Furthermore, among all the three continuity-corrected approximations, the saddle point approach is close to exact than the normal approximation.

6. Conclusion

We have introduced a simple technique to find the CDF for the discrete random sum distribution using the saddle point distribution. In conclusion, this study confirmed the accuracy of the saddle point approximation for the random sum models.

Acknowledgement

Authors declare that we have no competing interest, and they thank the referees for constructive comments.

References

[1] L. J. Norman, S. Kotz and A. W. Kemp, Univariate Discrete Distributions, 2nd ed., John Wiley and Sons Inc., New York, 1992.

[2] E. D. Henry, Saddle point approximations in statistics, The Annals of Mathematical Statistics (1954), 631-650.

[3] E. D. Henry, Tail probability approximations, International Statistical Review/Revue Internationale de Statistique (1987), 37-48.

[4] O. A. Alya and H. C. Low, Improved measures of the spread of data for some unknown complex distributions using saddle point approximations, Communications in Statistics-Simulation and Computation 45(1) (2016), 33-47.

[5] R. W. Butler, Saddle Point Approximations with Applications, Cambridge University Press, 2007.

[6] F. A. Ehab, Bivariate stopped-sum distributions using saddle point methods, Statistics and Probability Letters 78(13) (2008), 1857-1862.

[7] O. A. Alya and H. C. Low, Saddle point approximation to cumulative distribution function for damage process, Journal of Asian Scientific Research 3(5) (2013), 485-492.