

Improved method for approximating the hazard rate for convolution Poisson random variables

Andrei Volodin¹, Alya Al Mutairi²

¹Department of Mathematics and Statistics, University of Regina, Regina, SK, S4S0A2, Canada; andrei@uregina.ca

²Mathematics Department, Faculty of Science, Taibah University, Medina, Saudi Arabia; amutairi@taibahu.edu.sa

Abstract In this study, we investigate the performance of the saddlepoint approximation of the probability mass function and the cumulative distribution function for the weighted sum of independent Poisson random variables. The goal is to approximate the hazard rate function for this complicated model. The better performance of this method is shown by numerical simulations and comparison with a performance of other approximation methods.

Key Words: Saddlepoint approximations; Probability mass function; Probability distribution function; Poisson random variable, Linear combination

AMS Classification: MSC Primary 62F15, Secondary 62F65

1 Introduction

The distribution of a linear combination of random variables appears in many applied problems and has attracted the attention of many scholars; specifically, the linear combination of two independent random variables has been investigated by many researchers (see e.g. Amari & Misra [1], Di Salvo [3], Almasi *et al.* [4], Al Mutairi and Low [5] among others). Currently, the demand for these models has increased due to their importance and wide range of applications.

This study improves on the methods of hazard rate function estimation for a linear combination of random variables by using the saddlepoint approximation.

It is well known that the Poisson distribution is stable; that is, the distribution of the sum of two independent Poisson random variables has the Poisson distribution. This means that if two random variables $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$ are independent, that the distribution of their sum $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

At the same time, rescaling of a Poisson random variable is not Poisson anymore. Obviously, random variable $Y = 0.5X$, where $X \sim \text{Poisson}(\lambda)$ is not integer valued and hence not Poisson. Thus, the exact distribution of a *linear combination* of independent Poisson random variables is not always known and, even if it is, in most cases it is very difficult to obtain.

Therefore, methods of approximation are very important and can provide a useful information about the distribution of these complicated models. The saddlepoint approximation method can replace the exact answer with high accuracy. This adds to its many advantages in terms of saving time and efforts.

2 Saddlepoint approximation for the linear combination of Poisson random variables

The linear combination of the Poisson random variables is given by

$$L_c = c_1X_1 + c_2X_2 + \cdots + c_nX_n,$$

where $c_i, i = 1, 2, \dots, n$ are real constants and $X_i \sim \text{Poisson}(\lambda_i), i = 1, 2, \dots, n$ are independent Poisson random variables. Note that we do not require that $X_i, i = 1, 2, \dots, n$ are identically distributed.

Remind that the *cumulant generating function* (CGF) of a random variable X is defined as a logarithm of its moment generating function

$$K_X(s) = \ln E e^{sX}.$$

The saddlepoint technique requires calculating the cumulant generating function for the linear combination L_c of Poisson random variables as

$$K_{L_c}(s) = \sum_{i=1}^n \lambda_i (e^{c_i s} - 1).$$

Note that in Butler [6] Sections 1.1.2 and 1.1.5 (formulae (1.13) and (1.14)), the saddlepoint approximation is defined only for continuous and integer-valued distributions

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

where \hat{s} the unique solution for the saddlepoint equation $K'(\hat{s}) = x$.

The distribution of the linear combination of Poisson random variables L_c is not continuous, nor integer valued, but hence we need a continuity correction for the saddlepoint approximation. For more detailed summary of this arguments, see Daniels [6] and Daniels [7], Section 2 for more details. The first continuity correction for the right tail is given as

$$\hat{P}_{r1}(X \geq x) = \begin{cases} 1 - \Phi(\tilde{w}_1) - \phi(\tilde{w}_1) \left(\frac{1}{\tilde{w}_1} - \frac{1}{\tilde{u}_1} \right), & \text{if } x \neq \mu \\ \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \left[\frac{K'''(0)}{6K''(0)^{3/2}} - \frac{1}{2\sqrt{K''(0)}} \right], & \text{if } x = \mu, \end{cases} \quad (2)$$

where μ is the mean of the random variable X , Φ and ϕ are the standard normal distribution and density functions, respectively, and

$$\tilde{w}_1 = \text{sgn}(\hat{s}) \sqrt{2(\hat{s}x - K(\hat{s}))}, \tilde{u}_1 = (1 - e^{-\hat{s}}) \sqrt{K''(\hat{s})}.$$

The second continuity correction is given as

$$\hat{P}_{r2}(X \geq x) = \begin{cases} 1 - \Phi(\tilde{w}_2) - \phi(\tilde{w}_2) \left(\frac{1}{\tilde{w}_2} - \frac{1}{\tilde{u}_2} \right), & \text{if } x^- \neq \mu \\ \frac{1}{2} - \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{2/3}}, & \text{if } x^- = \mu, \end{cases}$$

where x^- is the unique solution for the saddlepoint equation $K'(\tilde{s}) = x - 0.5 = x^-$, and

$$\tilde{w}_2 = \text{sgn}(\tilde{s}) \sqrt{2(\tilde{s}x^- - K(\tilde{s}))}, \tilde{u}_2 = 2 \sinh\left(\frac{\tilde{s}}{2}\right) \sqrt{K''(\tilde{s})}.$$

The third continuity correction suggested in Butler [5], Section 1, p. 17-18, is the same as formula \hat{P}_{r2} but with \tilde{u}_2 is replaced by $\tilde{u}_3 = \tilde{s} \sqrt{K''(\tilde{s})}$.

3 An example of the saddlepoint approximation for the distribution function

Suppose the linear combination of two independent Poisson random variables

$$L_{c1} = c_1X_1 + c_2X_2,$$

where $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$. The CGF is

$$K_{L_{c_1}}(s) = \lambda_1(e^{c_1 s} - 1) + \lambda_2(e^{c_2 s} - 1).$$

The saddlepoint equation is

$$K'_{L_{c_1}}(\hat{s}) = \lambda_1 c_1 e^{c_1 \hat{s}} + \lambda_2 c_2 e^{c_2 \hat{s}} = x$$

To show the performance of this method, we consider the sum of two independent Poisson random variables with means λ_1 and λ_2 , which is also a Poisson random variable with mean $\lambda_1 + \lambda_2$. Let

$$c_1 = c_2 = 1 \text{ and } \lambda_1 = 1, \lambda_2 = 2.$$

For $x = 1$ we have

$$K'_{L_{c_1}}(\hat{s}) = e^{\hat{s}} + 2e^{\hat{s}} = 1,$$

hence, the saddlepoint is $\hat{s} = -1.098612289$ and the PMS $\hat{f}(x)$ from equation (1) is

$$\hat{f}(1) = \frac{1}{\sqrt{2\pi}} \exp(-2 + 1.098612289) = 0.1619728996. \quad (3)$$

The exact $f(x)$ is given by calculating $\text{Poisson}(3)$ when $x = 1$

$$f(1) = 3e^{-3} = 0.1493612051. \quad (4)$$

The absolute error of our approximation is $|f(1) - \hat{f}(1)| = 0.01$.

Next we find the first continuity correction given in equation (2)

$$\hat{P}_{r1} = 1 - \Phi(-1.34267) - \phi(-1.34267) \left(\frac{1}{-1.34267} - \frac{1}{-2} \right) = 0.9428487449. \quad (5)$$

because

$$\tilde{w}_1 = -[2(\ln(1/3) + 2)]^{0.5} \text{ and } \tilde{u}_1 = 1 - e^{-1.098}.$$

The exact right tail is

$$P(X \geq 1) = 1 - f(0) = 0.9502129316 \quad (6)$$

with the absolute error of 0.007.

We note that saddlepoint approximations give accurate results in calculating the mass function and distribution function and can replace the exact answers. We can compare this with other approximation methods, for example, the normal approximation for the right tail when $x = 1$

$$1 - \Phi\left(\frac{1 - E(L_c)}{\sqrt{\text{Var}(L_c)}}\right) = 1 - \Phi\left(\frac{1-3}{\sqrt{3}}\right) = 0.8749 \quad (7)$$

with the absolute error 0.07.

There is also another methods of approximation, namely, the Haldane approximation. Pentikäinen [8] proved the accuracy of that method under certain circumstances. Suppose X is a random variable with mean μ_x , variance σ_x^2 , and the coefficient of skewness γ_x . The Haldane type A approximation for the right tail is given by

$$\hat{H} = 1 - \Phi[(1 + r\tilde{x}_o)^h - \mu(h, r)]/\sigma(h, r),$$

where

$$\tilde{x}_o = \frac{x_o - \mu_x}{\sigma_x}, r = \frac{\sigma_x}{\mu_x}, h = 1 - \frac{\gamma_x}{3r},$$

$$\mu(h, r) = 1 - \frac{1}{2}h(1-h) \left[1 - \frac{1}{4}(2-h)(1-3h)r^2 \right] r^2,$$

$$\sigma(h, r) = hr \sqrt{1 - \frac{1}{2}(1-h)(1-3h)r^2},$$

see Borowiak [9]. We get $\hat{H} = 0.8849$ with absolute error of 0.06.

From the calculations above, we conclude that the saddlepoint approximation for the distribution function is more accurate than the normal and Haldane approximations.

4 The hazard rate function based on the saddlepoint approximation

The *hazard rate function* for a discrete distribution taking values in $\mathbb{Z}^+ = \{0,1,2,3, \dots \}$ is defined as

$$h(x) = \frac{P(X=x)}{P(X\geq x)}, x \in \mathbb{Z}^+.$$

(see, e.g., Daly [10]). Consider the same example of the sum of two independent Poisson random variables with means 1 and 2 as in the previous section. Equations (3) and (5) lead to the saddlepoint approximation of hazard rate function with the first continuity correction

$$\hat{h}_{s1}(1) = \frac{0.16197288996}{0.9428487449} = 0.1717909589,$$

while the exact hazard rate based on equations (4) and (6) is

$$h(1) = \frac{0.1493612051}{0.9428487449} = 0.1571870895.$$

Here, the absolute error for the saddlepoint and exact values is 0.01. For the normal approximation, based on equation (7), the hazard rate approximation is

$$h_N(1) = \frac{0.0805}{0.8749} = 0.0920105$$

with absolute error of 0.06.

The discussion above pertains the special case of the linear combination of two independent not identically Poisson random variables with constant $c_1 = c_2 = 1$. We consider this case because we know already the exact solution for this sum, which gives us a possibility to compare the exact solution with the methods of approximation.

Now we consider other positive real values for c_1 and c_2 , which gives the linear combination of independent Poisson random variables. It has a complicated distribution and the exact mass and density functions in most cases are unknown or difficult to obtain. Hence, the methods of approximation are very important.

Table 0 shows the exact values, saddlepoint continuity corrections, and the normal approximations of the hazard rate function for the linear combination $L_c = 3X_1 + 5X_2$, where $X_1 \sim \text{Poisson}(2)$ and $X_2 \sim \text{Poisson}(3)$ for different values x . The exact value is found by simulations by generating 10^6 random samples from the linear combination using the R program.

Table 1: Saddlepoint approximations for three continuity corrections $\hat{h}_{s1}, \hat{h}_{s2}, \hat{h}_{s3}$ with corresponding exact value $h(x)$ and normal approximations $\hat{h}_{N1}, \hat{h}_{N2}, \hat{h}_{N3}$ for L_c .

X	Exact $h(x)$	First continuity correction		Second continuity correction		Third continuity correction	
		$\hat{h}_{s1}(x)$	$\hat{h}_{N1}(x)$	$\hat{h}_{s2}(x)$	$\hat{h}_{N2}(x)$	$\hat{h}_{s3}(x)$	$\hat{h}_{N3}(x)$
5	0.0516	0.0776	0.0116	0.0789	0.0088	0.0771	0.0088
10	0.0620	0.0820	0.0220	0.0883	0.0193	0.0751	0.0193
15	0.1031	0.0781	0.0431	0.0771	0.0404	0.0788	0.0404

20	0.1431	0.1371	0.0731	0.1306	0.0704	0.1361	0.0704
25	0.1574	0.1614	0.0874	0.1664	0.0847	0.1669	0.0847

Next, to show the effectiveness of this method of approximating, we consider four components linear combination of independent Poisson random variables

$$L_{c_2} = c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4,$$

Applying the same technique as for two independent Poisson random variable, we obtain that the CGF is given by

$$K_{L_{c_2}}(s) = \lambda_1(e^{c_1s} - 1) + \lambda_2(e^{c_2s} - 1) + \lambda_3(e^{c_3s} - 1) + \lambda_4(e^{c_4s} - 1),$$

and the saddlepoint equation is obtained as

$$K'_{L_{c_2}}(\hat{s}) = \lambda_1c_1e^{c_1\hat{s}} + \lambda_2c_2e^{c_2\hat{s}} + \lambda_3c_3e^{c_3\hat{s}} + \lambda_4c_4e^{c_4\hat{s}} = x.$$

For example, suppose that

$$X_1 \sim \text{Poisson}(1), X_2 \sim \text{Poisson}(2), X_3 \sim \text{Poisson}(5), \text{ and } X_4 \sim \text{Poisson}(3)$$

with constants $c_1 = c_2 = c_3 = c_4 = 1$. Consider the value $x = 1$. This also gives us a good chance to investigate the accuracy of this methods because the exact solution is already known as $\text{Poisson}(11)$.

From the saddlepoint equation $K'_{L_{c_2}}(\hat{s}) = x$, we obtain that $e^{\hat{s}} = \frac{1}{11}$ and the saddlepoint is $\hat{s} = -2.3978952$.

Based on equation (1)

$$\hat{f}(1) = \frac{1}{\sqrt{2\pi}} \exp(-10 + 2.3978952) = 0.0001992819,$$

with its corresponding exact answer $f(1) = 11e^{-11} = 0.0001837187$. This leads to a high level of accuracy for calculating the mass function with the absolute error of 0.00001556.

The exact right tail probability when $x = 1$ is obtained as

$$P(X \geq 1) = 1 - f(0) = 1 - e^{-11} = 0.99998.$$

Calculating the first continuity correction by equation (2), we obtain

$$\tilde{w}_1 = -[2(-2.3978952 + 10)]^{0.5} = -3.89925 \text{ and } \tilde{u}_1 = (1e^{2.3978952}) = -9.9999.$$

This provides the first continuity correction of saddlepoint approximation for the right tail with $x = 1$

$$\hat{P}_{r1}(X \geq 1) = 1 - \Phi(-3.89925) - \phi(-3.89925) \left(\frac{1}{-3.89925} + \frac{1}{9.9999} \right) = 1.000.$$

with the absolute error of 0.00002.

The corresponding normal approximation for $x = 1$ is

$$P_N(x \geq 1) = 1 - \Phi\left(\frac{1-11}{11}\right) = 0.8159$$

with the absolute error of 0.184. This means that the saddlepoint approximation is closer to the exact value than the normal approximation.

From the calculations above we can derive the saddlepoint approximations for the hazard rate for the first continuity correction when $x = 1$ as

$$\hat{h}_{s1} = \frac{\hat{f}(1)}{\hat{P}_{r1}(x \geq 1)} = 0.0001992819$$

with absolute error for the exact hazard rate $h_e(1) = 0.0001837$ of 0.00019.

Table 2: Presents the three continuity corrections for the saddlepoint $\hat{h}_s(x)$ with its corresponding exact $h_e(x)$ and normal approximation $\hat{h}_N(x)$ for L_{c_2}

X	Exact $h_e(x)$	First continuity correction		Second continuity correction		Third continuity correction	
		$\hat{h}_{s1}(x)$	$\hat{h}_{N1}(x)$	$\hat{h}_{s2}(x)$	$\hat{h}_{N2}(x)$	$\hat{h}_{s3}(x)$	$\hat{h}_{N3}(x)$
1	0.0001837	0.0001992819	0.0256314	0.0002893	0.0299	0.0001899	0.029910
2	0.0010102	0.0010203000	0.0360780	0.0011751	0.0410	0.0019980	0.040130
3	0.0037048	0.0007183000	0.0123490	0.0007710	0.0141	0.0008011	0.013991
4	0.0102380	0.0102496000	0.0853800	0.0192810	0.0779	0.0203300	0.090340
5	0.0229000	0.0229100000	0.0997130	0.0228100	0.0889	0.0389100	0.088950

From Table 2 we see that the saddlepoint method of approximation is doing well with high level of accuracy. To derive the exact mass and distribution function intractable calculations are needed. The saddlepoint provides a fairly quick and accurate method of calculating for tail probability.

5 Conclusion

The saddlepoint approximation technique is an efficient tool that can be used to derive an accurate approximation to the hazard rate function for complicated models, such as a weighted sum of independent not identically Poisson distributed random variables.

Acknowledgements

The authors declare no conflict of interest in preparing this article, authors thank the referee for constructive comments.

References

[1] Amari, S. V., & R. B. Misra. "Closed-form expressions for distribution of sum of exponential random variables," IEEE Transactions on reliability **46**(4), 519–522 (1997).

[2] F. Di Salvo. "A characterization of the distribution of a weighted sum of gamma variables through multiple hypergeometric functions," Integral Transforms and Special Functions **19**(8), 563–575 (2008).

[3] I. Almasi, R. Jalilian, & K. Sayehmiri. "The exact distribution of sums weights of Gamma variables," Journal of Iranian Statistical Society **11**(1), 23–37 (2012).

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

[5] R. W. Butler. Saddlepoint Approximations With Applications (Cambridge University Press 2007).

[6] H. E. Daniels. "Tail Probability Approximations," International Statistical Review **55**, 37–48 (1987).

[7] H. E. Daniels. "Saddlepoint approximations in statistics," The Annals of Mathematical Statistics **25**(4), 631–650 (1954).

- [8] T. Pentikäinen. "Approximative Evaluation of the Distribution Function of Aggregate Claims 1," *ASTIN Bulletin: The Journal of the IAA* **17**(1), 15–39 (1987).
- [9] D. S. Borowiak. "A saddlepoint approximation for tail probabilities in collective risk models," *Journal of Actuarial Practice* **7**(1), 239–249 (1999).
- [10] F. Daly. "On strong stationary times and approximation of Markov chain hitting times by geometric sums," *Statistics & Probability Letters*, **150**, 74–80. (2019).