

Saddlepoint Method to Cumulative Distribution Function for Poisson-Binomial Model

Al Mutairi Alya O.^{1,2} & Heng Chin Low¹

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

Received: April 9, 2013

Accepted: April 30, 2013

Online Published: May 24, 2013

doi:10.5539/mas.v7n6p101

URL: <http://dx.doi.org/10.5539/mas.v7n6p101>

The study is supported financially by Taibah University in Al-Madinah, Kingdom of Saudi Arabia

Abstract

The random sum distribution plays an important key in statistical science as well as with insurance program, biotechnology and applied medical science. Saddlepoint methods are considered to be random sum variables with dependent elements supposing presence of the Moment Generation Function (MGF). Saddlepoint methods are influential instruments for getting precise terms for distribution functions in closed form. However, the paper also, discusses the Saddlepoint methods to the Cumulative Distribution Function (CDF) for Poisson-Binomial model in discrete form.

Keywords: random-sum distribution, poisson-binomial model, saddlepoint approximation, cumulative distribution function

1. Introduction

Saddlepoint method is influential way in having precise terms for distribution function which is not recognized in closed sitting. Saddlepoint approximation almost surpass other techniques regarding calculating expenses; while it does not inevitably surpass them concerning accuracy. The most basic Saddlepoint method was launched by Daniels (1954) and is fundamentally expression for approximating CDF and discrete distribution function through its MGF. Saddlepoint methods are constructed by supposing existence of the MGF or, equally, the CGF, of random variable. However, for improvement to the Saddlepoint methodology and associated techniques, as references, Skovagaard in 1987 proposed a conditional version of this approximation, Daniels (1954, 1987) reported details concerning density and mass approximation, in 1999 Borowaik discussed a tail-area approximation with a uniform relative error, Reid in 1998 indicated applications to inference, and Terrell (2003) proposed a stabilized Lugannani-Rice formula. However, this paper will use Saddlepoint method to estimate the random-sum based on MGF for Poisson-Binomial model. For all real values of x , random variable X was presumed to have mass function $p(x)$ identified. Subsequently, the MGF is identified as

$$M(s) = E(e^{sx}) = \sum_{x=-\infty}^{\infty} e^{sx} p(x), \text{ (Hogg \& Craig, 1978)} \quad (1)$$

Over values of s for which the integral converges and the convergence is constantly certain at $s=0$ and it should be supposed that $M(s)$ coverage over largest open neighborhood zero as (a,b) . However, the CGF is known as

$$K(s) = \ln M(s), s \in (a,b), \text{ (Johnson et al., 2005)} \quad (2)$$

For discrete integral-valued random variable X , the saddlepoint approximation for its mass function $p(x)$, based on the CGF K is given by

$$\hat{p}(x) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp(k(\hat{s}) - \hat{s}x) \quad (3)$$

where $\hat{s} = \hat{s}(x)$ denote the unique outcome to the Saddlepoint equation

$$K'(\hat{s}) = x, x \in \tau_x$$

$x \in \tau_x$ and τ_x is the inner part of the span that powered by X (Butler, 2007). Saddlepoint expression (3) is calculable for any value in τ_x , however, the plot of $\hat{p}(x)$ is significant as an approximation to $p(x)$ merely for integer-value.

Daniels (1987) initiated, two stability approved adjustments for discrete integral-valued random variable X for univariate cumulative distribution functions (CDF) that are presented below.

1.1 First Continuity-Modification

Assume $x \in \tau_x$, in order that the saddlepoint equation could be solved at value x , the first approximation is

$$\hat{P}_{r_1}(X \geq x) = \begin{cases} 1 - \Phi(\hat{w}) - \phi(\hat{w})\left(\frac{1}{\hat{w}} - \frac{1}{\hat{u}}\right) & \text{if } x \neq \mu \\ 0.5 - \frac{1}{\sqrt{2\pi}} \left\{ \frac{K'''(0)}{6K''(0)^{\frac{3}{2}}} - \frac{1}{2\sqrt{K'(0)}} \right\} & \text{if } x = \mu \end{cases} \quad (4)$$

\hat{w} and \hat{u}_1 are known by

$$\hat{w} = \text{sgn}(\hat{s})[\hat{s}x - K(\hat{s})]^{0.5}$$

$$\hat{u}_1 = (1 - \exp(-\hat{s})) [K''(\hat{s})]^{0.5}$$

and the saddlepoint \hat{s} solves $K(\hat{s}) = x$. The symbol ϕ and Φ indicate the normal probability distribution function (PDF) and the CDF correspondingly as well as $\text{sgn}(\hat{s})$ takes $\text{sign}(\pm)$ for \hat{s} (Butler, 2007).

1.2 Second Continuity-Modification

Describe $x^- = x - .05 \in \tau_x$ as the continuity-modification or offset value of x . The second approximation solves the offset value of saddlepoint equation $K'(\tilde{s}) = x^-$. The saddlepoint \tilde{s} and x^- are employed to change the inputs into the cumulative distribution function approximation consistent with

$$\tilde{w}_2 = \text{sgn}(\tilde{s})\sqrt{2[\tilde{s}x^- - K(\tilde{s})]} \quad (5)$$

$$\tilde{u}_2 = 2 \sinh\left(\frac{\tilde{s}}{2}\right)\sqrt{K''(\tilde{s})}$$

This leads to the second continuity-modified approximation

$$\hat{P}_{r_2}(X \geq x) = \begin{cases} 1 - \Phi(\tilde{w}_2) - \phi(\tilde{w}_2)\left(\frac{1}{\tilde{w}_2} - \frac{1}{\tilde{u}_2}\right) & \text{if } x^- \neq \mu \\ \frac{1}{2} - \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{\frac{3}{2}}} & \text{if } x^- = \mu \end{cases}, \text{ (Butler, 2007)} \quad (6)$$

1.3 Third Approximation

This approximation is indicated as $\hat{P}_{r_3}(X \geq x)$ and employs expression (6) with \tilde{w}_2 as in (5) and \tilde{u}_2 substituted with

$$\tilde{u}_3 = \tilde{s}\sqrt{K''(\tilde{s})}, \text{ (Butler, 2007)} \quad (7)$$

2. Claim Frequency Models (Daniel, 2008)

Numerous procedures in daily life that calculate events up to a special stage in time could be precisely explained by the supposed Poisson process; now a Poisson process will be defined. Firstly, a Poisson process known as a compilation of random variables $N(t)$ per t about particular group. Explicitly, Poisson process is a reliable form: in every $t > 0$ they calculate the incidents that occur through time 0 and time t . The type of incidents relies in the function. One may desire to calculate how many incidents cases claimed via a private made, or how many calls for assistance and how many persons quit from a special company etc. No matter what you may consider by an "incidents", $N(t)$ indicates the incident numbers that happened following time 0 to that moment and comprising time $t > 0$. Consider that $N(0)$ is regarded to exact 0 signifies no incidents could be happened sooner than you begin calculating. As when the incident take a place and supposed to be a random, $N(t)$ is a random variable to every value of t . Consider that N itself is presented *random process*, however, differentiating it from random variable $N(t)$ at every value of $t > 0$. Furthermore, to recognize the counting process, it required to comprehend the connotation. In addition, the probability response through the amount of

$N(t+h) - N(t)$ in time t to time $t+h$, where $h > 0$ and certainly $t \geq 0$. Because $N(t+h)$ equals to the number of incidents up through $t+h$ and $N(t)$ equals the random number up through t , the growth is merely the random number of incidents happening firmly after time t and up through and comprising time $t+h$. Consider that $N(t)$ itself could be regarded as an increase, explicitly from time (0) to time t , given that $N(t) = N(t) - N(0)$ (Daniel, 2008).

In Poisson process the balance that incidents are generally happening is very significant.

The proportion of function λ reports the average as $\lambda(t)$ at time t . Consider that the average could vary with time, however, the rapidity of a vehicle the average at which it is passing distance could differ with time. You obtain the total distance passed by a vehicle through a time period by multiplying the average by the length of the time interval, provided the average is a constant; if the average varies, to merge the average function over the interval, which gives the same outcome as the simpler formula when the average is constant. As well as, to merge Poisson process's average function over an interval we gain the average of incidents in that period (Daniel, 2008).

2.1 Definition and Properties of Poisson Random Variable

A Poisson random variable Y with mean λ is a random variable with the subsequent characteristics:

1) It is only potential values are the positive integers 0, 1, 2, 3, ... and

$$2) \Pr[Y = x] = \frac{e^{-\lambda} \lambda^x}{x!}$$

Within these characteristics we can find that

$$3) E[Y] = \text{Var}[Y] = \lambda.$$

The meaning of the previous explanation of Poisson processes.

2.2 Definition and Characteristics of Poisson Process

A Poisson process N with rate function λ has the next properties:

1) N is a counting process $N(0) = 0$ and for $t > 0$ $N(t)$ is non decreasing and accept only positive integer values 0, 1, 2, 3, ... and, thus, it could be regarded as the random number of incidents of significance happening after time 0 and by time t .

2) N has separate increases of any group of increases $N(t_i + h_i) - N(t_i)$ for $i = 1, 2, \dots, n$ is separate for any positive integer n , given that the time intervals $(t_i, t_i + h_i)$ are interference (interference at an endpoint is acceptable).

3) Every $t \geq 0$ and $h > 0$, the increases $N(t+h) - N(t)$ is a Poisson random variable with mean

$$\lambda = \int_t^{t+h} \lambda(z) dz.$$

4) If the rate function λ is indeed stable, afterward, N is described a homogeneous Poisson process. In usual practice, Poisson process has typically implied "homogeneous Poisson process", whereas "non-homogeneous Poisson process" has been employed to signify a rate function that isn't stable. However, the terms must not puzzle you, because you could constantly observe to perceive whether λ is stable. Moreover, you observe declarations such as "events happen at the Poisson rate 3 per hour"; this is stenography for "events are happening consistent with a Poisson process with stable rate function $\lambda = 3$ per hour".

2.3 Random-Sum Poisson Process

Companies offer medical insurance to their staff, they are certainly worried from frequent prompts, the random number of prompts occurred. Furthermore, they are worried from prompts strictness, the random size of every claim. However, they are particularly worried from the total sum of every one of the prompts. The total of the random numbers of random variables, and as it should really difficult to be analyzed. However, the probability distribution is identified as the random sum distribution.

2.3.1 Definition of Random-Sum Poisson Process

The random-sum Poisson process Y has the subsequent properties:

$$1) \text{ For } t > 0, Y(t) = \sum_{j=1}^{N(t)} X_j.$$

- 2) N is the Poisson process with rate function λ ;
- 3) Every random variable X_j has similar distribution as a single random variable X ;
- 4) For all t , the random variables $N(t)$ and all the X_j from independent group;
- 5) If $N(t)$ equal 0 for a special value of t , after that the empty sum $Y(t)$ is regarded to be equal 0, thus, specifically, $Y(0) = 0$.

The number of N is linked with the frequent prompts in the specified group of the policies and regulation. The general distribution of X_1, X_2, \dots is indicated by X . Consider X forms the sum of a random prompts made in this file of insurance contract. Once the frequent prompts N with a constant parameter λ follow a Poisson distribution, the total prompts Y is supposed to have a random-sum Poisson distribution with mean $\lambda = E[N]$ and the variance $Var[Y] = \lambda E[X^2]$. The MGF is given by

$$M_Y(s) = M_N[\ln M_X(s)] = \exp[\lambda(M_X(s) - 1)] \quad (8)$$

3. Numerical Example

Assume that vehicle accident prompts are submitted with a vehicle insurer at the Poisson rate $\lambda = 5$ per hour, and that independent X of persons badly hurt in every accident are Binomial random variables with parameters $n = 2$ and $p = 0.2$. Subsequently, the total number Y of those badly hurt is a random-sum Poisson process. As previously showed, random sum Poisson process are extremely difficult and hard to investigate, and, therefore, approximation techniques are frequently employed. Saddlepoint methods defeat this difficulty. Saddlepoint methods are influential tool for getting precise expressions for distribution function that isn't recognized in closed form. Saddlepoint methods roughly surpass other techniques regarding calculating expenses, while it does not inevitably surpass them concerning correctness.

The total $Y(t)$ are supposed to have a Poisson-Binomial random-sum distribution when i.i.d. random variables X_i 's follow Binomial (n, p) distribution. The MGF of a total of prompts Y is identified as

$$M_{Y(t)}(s) = M_N[\ln M_X(s)], \text{ (Hogg \& Tanis, 1983)} \quad (9)$$

where "ln" is the natural log function.

Furthermore, the CGF of a total of prompts Y identified as the CGF for N is known by

$$K_N(s) = \ln[M_N(s)] = \lambda(e^s - 1) \quad (10)$$

And for X_i 's follow Binomial (n, p) distribution, the CGF is described as

$$K_X(s) = \ln[M_X(s)] = \ln(pe^s + q)^n \quad (11)$$

Where $p + q = 1$. Subsequently, it possible to drive the CGF for the Poisson-Binomial random sum distribution as

$$K_{Y(t)}(s) = K_N(K_X(s)) = \lambda[(pe^s + q)^n - 1] \quad (12)$$

In this case the saddlepoint equation is

$$K'_{Y(t)}(\hat{s}) = n\lambda p e^{\hat{s}} (pe^{\hat{s}} + q)^{n-1} \quad (13)$$

Next, the saddlepoint can be found as $\hat{s} = \hat{s}(x)$ which indicates as the only conclusion to the saddlepoint formula $K'(\hat{s}) = x$.

However, second and third derivative of the CGF $K''(\hat{s})$ as well as $K'''(\hat{s})$ also

$$K''_{Y(t)}(\hat{s}) = n(n-1)\lambda p^2 e^{2\hat{s}} (pe^{\hat{s}} + q)^{n-2}$$

$$K'''_{Y(t)}(\hat{s}) = n(n-1)(n-2)\lambda p^3 e^{3\hat{s}} (pe^{\hat{s}} + q)^{n-3}$$

This leads to the saddlepoint mass function for Poisson-Binomial random sum which is known as

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi n(n-1)\lambda p^2 e^{2\hat{s}} (pe^{\hat{s}} + q)^{n-2}}} \exp[\lambda[(pe^{\hat{s}} + q)^n - 1] - \hat{s}x] \quad (14)$$

Also the saddlepoint method to CDF for Poisson- Binomial random sum presented below.

3.1 First Continuity-Modification

The first continuity-modification will be described as:

$$\hat{w} = \text{sgn}(\hat{s})\sqrt{2(x\hat{s} - \lambda[(pe^{\hat{s}} + q)^n - 1])}$$

$$\tilde{u}_1 = (1 - \exp(-\hat{s}))\sqrt{n(n-1)\lambda p^2 e^{2\hat{s}}(pe^{\hat{s}} + q)^{n-2}}$$
(15)

This leads to the first continuity-modified method known in expression (4).

3.2 Second Continuity-Modification

The second continuity modification will describe $x^- = x - .05 \in \tau_x$ as the continuity-modified or offset value of x . The second method solves the offset saddlepoint equation $K'(\tilde{s}) = x - 0.5$; and the saddlepoint \tilde{s} and x^- are employed to change the inputs into the CDF method consistent with

$$\hat{w}_2 = \text{sgn}(\hat{s})\sqrt{2\{\hat{s}(x - 0.5) - \lambda[(pe^{\hat{s}} + q)^n - 1]\}}$$

$$\tilde{u}_1 = 2 \sinh\left(\frac{\tilde{s}}{2}\right)\sqrt{n(n-1)\lambda p^2 e^{2\tilde{s}}(pe^{\tilde{s}} + q)^{n-2}}$$
(16)

Subsequently, the second continuity-modified approximation specified in expression (6) will be found.

3.3 Third Approximation

This approximation is indicated as $\hat{P}_{\tilde{s}}(X \geq x)$ and employs expression (6) with \tilde{w}_2 as in (16) and \tilde{u}_2 substituted with

$$\tilde{u}_3 = \tilde{s}\sqrt{n(n-1)\lambda p^2 e^{2\tilde{s}}(pe^{\tilde{s}} + q)^{n-2}}$$
(17)

4. Conclusion

This study indicated that, the saddlepoint methods to the CDF for random sum Poisson-Binomial distribution in discrete form. Furthermore, the methods estimate the random-sum variable components assuming presence of the MGF.

Acknowledgements

The study is supported financially by Taibah University in Al-Madinah, Kingdom of Saudi Arabia. The authors Thanks the referees for constructive cimments.

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>
- Daniel, J. W. (2008). *Poisson processes (and mixture distributions)*. Retrieved from <http://www.actuarialseminars.com/Misc/PPjwd.pdf>
- Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (3rd ed.). John Wiley & Sons, Inc., pp. 386-388. <http://dx.doi.org/10.1002/0471715816>
- Hogg, R. V., & Craig, A. T. (1978). *Introduction to mathematical statistics* (4th ed.). USA: Collier Macmillan Publisher.
- Hogg, R. V., & Tanis, E. A. (1983). *Probability and Statistical Inference* (2nd ed.). New York: Macmillan Publishing Co.
- 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 ayposium on the Interface. *Computing Science and Statistics, Annual report department of statistics Virginia tech March 14*.