Journal of Asian Scientific Research, 2013, 3(5):485-492

Journal of Asian Scientific Research

journal homepage: http://aessweb.com/journal-detail.php?id=5003

SADDLE POINT APPROXIMATION TO CUMULATIVE DISTRIBUTION FUNCTION FOR DAMAGE PROCESS

AL MUTAIRI ALYA O.

Applied Statistics Department, Faculty of Applied Science, Taibah University, AlMadinah-M., Kingdom of Saudi Arabia

HENG CHIN LOW

School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia

ABSTRACT

The random sum distribution plays a key role in the application of statistics, in addition, it can be used with insurance program, biotechnology and applied medical science. The statistical significance of this distribution arises from its applicability in real life situations. Saddle point approximations are powered tool in obtaining accurate expressions for distribution functions in closed form. Approximations almost outperform other methods with respect to computational costs, though not necessarily with respect to accuracy. However, the paper also, discusses the Saddle point methods to the cumulative distribution function (CDF) for damage process in discrete form. Furthermore, it shows approximations to random sum variable with dependent components assuming presence of the moment generating function (MGF).

Keywords: Random sum, Poisson-Bernoulli distributions, Saddle point approximation, Cumulative distribution function.

INTRODUCTION

Saddle point methods are influential tools in getting precise terms for distribution function which are not recognized in closed sitting. Saddle point approximations almost surpass other techniques regarding calculating expenses; while it does not inevitably surpass them concerning accuracy. The most basic Saddle point approximation was launched by Daniels (1954) and is fundamentally expression for approximating CDF and discrete distribution function through its MGF. Saddle point approximations are constructed by supposing existence of the MGF or, equally, the CGF, of random variable. However, in purpose of improvement of saddle point and associated methods, one can refer to (Daniels, 1954; Daniels, 1987) for facts of the continuous and discrete method distribution function, study conducted by Skovgaard (1987) for a conditional methods.

Furthermore, Reid (1988) reported for applications to inference. As well as, Borowiak (1999) study, for discussion of a tail area approximation that has uniform relative error and Terrell (2003) for a stabilized Lugannani-Rice formula. In this paper, we will use the saddle point approximation to estimate the random sum based on MGF for damage process. Presume a random variable X has mass function $f(x)$ identified for all real values of x. Subsequently, the MGF is identified as

$$M(s) = E(e^{sx}) = \sum_{x=-\infty}^{\infty} e^{sx} p(x) \qquad (1)$$

(Hogg and Craig, 1978) Over values of s for which the integral converges and the convergence is constantly certain at $s = 0$ and it should be supposed that $M(s)$ converges over largest open neighborhood at zero as (a,b). The CGF is known by

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

(Johnson *et al.*, 2005).

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

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

where $K'(\hat{s}) = x \qquad , x \in \tau_\chi$

and τ_χ is the inner part of the span that powered by χ, see Butler (2007). Saddle point expression (3) is computable for any value in τ_χ, but the plot of $\hat{p}(x)$ is meaningful as an approximation to $p(x)$ only for integer-valued arguments.

Saddle point Approximation for Univariate Cumulative Distribution Functions

Daniels (1987) initiated two continuity correct modifications for discrete integral-valued random variable X for univariate cumulative distribution functions (CDF) that are presented below.

First continuity-correction

Suppose $x \in \tau_{\chi}$ so, that the saddle point equation can be solved at value x, the first approximation is

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

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

$$\hat{w} = \operatorname{sgn}(\hat{s})\left[2(\hat{s}x - K(\hat{s}))\right]^{0.5}$$
$$\hat{u}_1 = (1 - \exp(-\hat{s}))\left[K''(\hat{s})\right]^{0.5}$$

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

Second Continuity-correction

Define $x^{-} = x - .05 \in \tau_{\chi}$ as the continuity-corrected or offset value of x. The second approximation solves the offset saddle point equation $K'(\tilde{s}) = x^{-}$. The saddle point $\tilde{s}$ and x^{-} are used to alter the inputs into the CDF approximation according to

$$\tilde{w}_2 = \operatorname{sgn}(\tilde{s})\sqrt{2[\tilde{s}x^{-} - K(\tilde{s})]}$$
$$\tilde{u}_2 = 2\sinh(\frac{\tilde{s}}{2})\sqrt{K''(\tilde{s})} \tag{5}$$

This leads to the second continuity-modified approximation

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

(Butler, 2007).

Third Approximation

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

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

(Butler, 2007).

THE RANDOM-SUM DISTRIBUTIONS

The random sum distributions have many natural applications. We motivate the notion of random distributions with damage process. The random variable Y is said to have a random sum distribution if Y is of the following form

$$Y = X_1 + X_2 + X_3 + \ldots + X_N \quad (8)$$

However, the number of terms N is uncertain, the random variables X_i's are independent and identically distributed (with common distribution X) and every X_i is independent of N. If N = 0 is realized, then we have $Y = 0$. Even though, this implicit in definition, we want to call this out for clariety. The concepts of damage process are inherent in the work of catcheside (1048). For a prescribed dosage and length of exposure to radio-therapy, the number of N of chromosome breakages in individual cells can be assumed to have Poisson random variables.

Each breakage has fixed independent probability. This leads to the random sum Poisson- Bernoulli distribution. The distribution function of Y is given by

$$f_Y(y) = \sum_{n=0}^{\infty} G_n(y) P[N = n] \quad (9)$$

(Johnson *et al.*, 2005). Where for $n \geq 1, G_n(y)$, 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:

$$f_Y(y) = \sum_{n=0}^{\infty} f^{*n}(y) P[N = n] \quad (10)$$

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 is discrete. On the other hand, if X is continuous and if $P[N = n] > 0$, then the random variable Y will have a mixed distribution, as is often the case in insurance applications. The mean of the random sum Y is $E[Y]$ given by

$$E[Y] = E[N]E[X] \quad (11)$$

The variance of the random sum Y is $Var[Y]$ and is defined as

$$Var[Y] = E[N]Var[X] + Var[N]E[X]^2 \quad (12)$$

The moment generating function of random sum Y is given by

$$M_Y(s) = M_N[\ln M_X(s)] \quad (13)$$

(Hogg and Tanis, 1983) where the function $\ln$ is the natural log function. As well as, the cumulant generating function of random sum Y is defined as

$$K_Y(s) = \ln M_Y(s) = \ln M_N[K_X(s)] = K_N[K_X(s)] \quad (14)$$

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

$$M_Y(s) = M_N[\ln M_X(s)] = \exp[\lambda(M_X(s) - 1)] \quad (15)$$

If the distribution X_i's are i.i.d. random variables follows Bernoulli (p) distribution, the sum Y is said to have a Poisson- Bernoulli random sum distribution. The cumulant generating function for N is given by

$$K_N(s) = \ln[M_N(s)] = \lambda(e^s - 1) \tag{16}$$

And for X_i's which are i.i.d. random variables follows Bernoulli (p) distribution, The CGF for X_i's is defined as

$$K_X(s) = \ln[M_X(s)] = \ln(pe^s + q) \tag{17}$$

Where $p + q = 1$. Then, we can drive the cumulant generating function for the Poisson- Bernoulli random sum distribution as follows

$$K_Y(s) = K_N(K_X(s)) = \lambda[(pe^s + q) - 1)] \tag{18}$$

Then the saddle point equation is

$$K'_Y(\hat{s}) = K''_Y(\hat{s}) = \lambda pe^{\hat{s}} = x \tag{19}$$

Then, we can find the saddle point as $$\hat{s} = \ln\frac{x}{\lambda p} \tag{20}$$

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

$$\hat{p}(x) = \frac{1}{\sqrt{2\pi x}}\exp(x - \lambda p - x\ln\frac{x}{\lambda p}) \tag{21}$$

And the saddle point approximation to cumulative distribution functions for Poisson- Bernoulli random sum presented below.

First Continuity-correction

For first continuity-correction we will define

$$\hat{w} = \sqrt{2(x\ln\frac{x}{\lambda p} - x + \lambda p)}$$
$$\tilde{u}_1 = (1 - \frac{\lambda p}{x})\sqrt{x} \tag{22}$$

This leads to the First continuity-corrected approximation given in expression (4).

Second continuity-correction

For Second continuity correction we will define $x^- = x - .05 \in \tau_\chi$ as the continuity-corrected or offset value of x. The second approximation solves the offset saddle point equation

$K'(\tilde{s}) = x - 0.5$ and the saddle point $\tilde{s}$ and x^- is used to alter the inputs into the CDF approximation according to

$$\hat{w}_2 = \sqrt{2\{(x-0.5)\ln\frac{x-.5}{\lambda p} - \lambda(x-0.5-\lambda p)\}}$$
$$\tilde{u}_1 = 2\sinh(0.5\ln\frac{x-0.5}{\lambda p})\sqrt{x-0.5} \tag{23}$$

Then, we'll find the second continuity-corrected approximation given in expression (6).

Third Approximation

This approximation is denoted as $\hat{P}_{r_3}(X \geq x)$ and uses expression (6) with $\tilde{w}_2$ as in (23) and $\tilde{u}_2$ replaced with $\tilde{u}_3 = \ln\frac{x-0.5}{\lambda p}\sqrt{x-0.5}$.

CONCLUSION

This study indicated that, the saddle point approximation to the CDF for random sum Poisson-Binomial Distribution in discrete settings. Furthermore, we discussed approximations to random sum random variable with dependent components assuming presence of the MGF.

ACKNOWLEDGEMENT

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

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.

Daniels, H.E., 1954. Saddlepoint approximations in statistics. Annals of Mathematical Statistics, 25: 573–650.

Daniels, H.E., 1987. Tail probability approximations. International Statistical Review 55: 37–48.

Hogg, R.V. and A.T. Craig, 1978. Introduction to mathematical statistics. 4th Edn., USA: Collier Macmillan Publisher.

Hogg, R.V. and E.A. Tanis, 1983. Probability and statistical inference. 2nd Edn., New York: Macmillan Publishing Co.

Johnson, N.L., A.W. Kemp and S. Kotz 2005. Univariate discrete distributions. 3rd Edn.: John Wiley & Sons, Inc.

Reid, N., 1988. Saddlepoint methods and statistical inference (with discussion). Statist. Sci, 3: 213--238.

Skovgaard, I.M., 1987. Saddlepoint expansions for conditional distributions Journal of Applied Probability 24: 875–887.

Terrell, G.R., 2003. A stabilized lugannani-rice formula aymposium on the interface. In: Computing Science and Statistics, Annual report department of statistics Virginia tech.