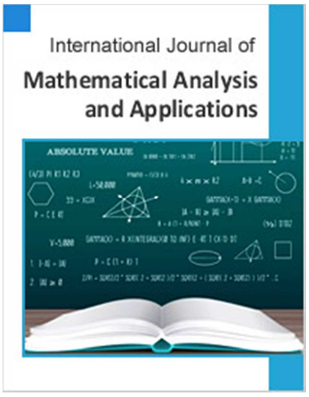
International Journal of Mathematical Analysis and Applications
2017; 4(6): 42-46
http://www.aascit.org/journal/ijmaa
ISSN: 2375-3927

American Association for Science and Technology



Improved Approximation Methods to the Stopped Sum Distribution

Amani Al Rashidi*, Bashair Al Juhani, Tahani Al Saeidi, Hanan Al Ahmadi, Mashael Al Harbi, Mawada Brri, Nada Al Johani, Alya Almutairi

Department of Mathematics, Faculty of Science, Taibah University, Medina, Saudi Arabia

Email address

mosra99@yahoo.com (A. Al Rashidi)

*Corresponding author

Citation

Amani Al Rashidi, Bashair Al Juhani, Tahani Al Saeidi, Hanan Al Ahmadi, Mashael Al Harbi, Mawada Brri, Nada Al Johani, Alya Almutairi. Improved Approximation Methods to the Stopped Sum Distribution. *International Journal of Mathematical Analysis and Applications.* Vol. 4, No. 6, 2017, pp. 42-46.

Keywords

Approximations,
Stopped Sum Distribution,
Geometric,
Negative Binomial-Bernoulli,
Statistics,
Probabilities

Received: July 24, 2017
Accepted: October 16, 2017
Published: November 9, 2017

Abstract

Most of the modern statistical models used the method that require the computation of probabilities from the complicated distribution such as (stopped sum) which can lead to intractable computation. However, in this study we will use a modern method to solve this problem, this method is called "Saddlepoint Approximation". However, In this study, we will derive the Saddlepoint Approximation (CDF) for some very complicated statistics such as stopped sum Geometric- Geometric distribution and stopped sum Negative binomial-bernoulli distribution.

1. Introduction

The method of approximations are very important in statistics because sometimes we cannot derive the exact CDF or the PDF for some complicated distributions see Ref. [1]. So, using the approximations methods can save time, effort and cost see Ref. [2]. Some basics and fundamentals about Saddlepoint Approximation and the stopped sum are presented below to simplify understanding them. In most cases, the distribution of the stopped 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 see Ref. [3].

2. Stopped Sum Distribution

The stopped sum distribution one of the very important statistics and has many application in our life such as in branching process and in insurance system like the work of [4]. Unfortunately, most of these distributions are still unknown until now see Ref. [5]. This technique is based on the moment generating function as given in [6].

Let $X_1 + X_2 + \cdots + X_N$ are iid random variable and X's and N are independent random variables. This stopped sum can be expressed mathematically as the following formula Ref. [7]

$$S_N(X) = X_1 + X_2 + \cdots + X_N$$

The mean and variance of stopped sum of the random variable $S_N = \sum_{i=1}^{N} X_i$ are presented respectively by:

$$E(S_N) = E(N)E(X)$$

$$Var(S_N) = E(N)\,var(X) + (E(X))^2\,var(N)$$

The moment generating function of the stopped sum S_N is given by Ref. [8]:

$$M_{S_N}(s) = M_N[\ln M_X(s)]$$

3. Stopped Sum of the Negative Binomial-Bernoulli Distribution

The stopped sum of the Negative Binomial-Bernoulli distribution where N ~Negative Binomial (r, p) and X′s~Bernoulli (p).

The MGF for Negative Binomial distribution is given by [9] as:

$$M_N(s) = \left(\frac{q}{1 - pe^s}\right)^r$$

where $p + q = 1$ as well, the MGF for Bernoulli distribution is given by:

$$M_X(s) = (pe^s + q).$$

The cumulant generating function for N is given by:

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

$$K_N(s) = \left(\frac{q}{1 - pe^s}\right)^r$$

The cumulant generating function for X is given by:

$$K_X(s) = \ln(M_X(s))$$

$$K_X(s) = \ln(pe^s + q)$$

The cumulant generating for the stopped sum Negative Binomial–Bernoulli distribution as follows:

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

$$K_{S_N} = r\ln\left(\frac{q}{1 - p^{2e^s} - pq}\right)$$

The Saddlepoint equation $K'_{S_N}(\hat{s})$ =x becomes

$$K'_{SN}(\hat{s}) = -rp^2e^{\hat{s}}p^2e^{\hat{s}} + pq - 1 = x$$

then $\hat{s} = \ln\left(\frac{x(pq-1)}{p^2(-r-x)}\right)$

The first continuity-correction uses

$$\widehat{w}_1 = sgn(\hat{s})\sqrt{2\,\{x\ln(x(pq - 1)p^2(-r - x)) - r\ln(q1 - pq - p^2e^{\hat{s}})\}},$$

$$\hat{u}_1 = \left\{1 - e^{-(\ln\left(\frac{x(pq-1)}{p^2(-r-x)}\right)}\right\}\sqrt{\left(-\frac{p^2(pq-1)re^{\hat{s}}}{(p^2e^{\hat{s}}+pq-1)^2}\right)}$$

see Ref. [10].

The second continuity-correction leads to

$$\widetilde{w}_2 = sgn(\tilde{s})\sqrt{2\,\{\ln(\bar{x}(pq - 1)p2(-r - \bar{x}))\bar{x} - r\ln(q1 - pq - p2e\tilde{s})\}},$$

where $\bar{x} = x - 0.5$,
then

$$\tilde{u}_2 = 2\sinh\left(\frac{\ln(\frac{\bar{x}(pq-1)}{p^2(-r-\bar{x})})}{2}\right)\sqrt{(-\frac{p^2(pq - 1)re^{\tilde{s}}}{(p^2e^{\tilde{s}} + pq - 1)^2})}.$$

And the third continuity-correction is used

$$\tilde{u}_3 = \tilde{s}\sqrt{(-p^2(pq - 1)re^{\tilde{s}}/(p^2e^{\tilde{s}} + pq - 1)^2)},$$

The mean and variance are given respectively as

$$E(N) = \frac{pr}{q}, E(X) = p$$

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

$$E_{S_N}(s) = \frac{pr}{q}p$$

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

$$V(N) = \frac{pr}{q^2}, V(X) = pq$$

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

then $V_{S_N}(s) = \left(\frac{pr}{q}\right)(pq) + (p)^2\left(\frac{pr}{q^2}\right)$

Table 1 shows the Saddlepoint Approximation $\hat{p}_r$ with its corresponding normal approximation $\bar{p}_r$ for the left tail, Table 2 show, the Saddlepoint approximation $\hat{p}_r$ with its corresponding normal approximation $\bar{p}_r$ for the center and Table 3 show, the Saddlepoint Approximation $\hat{p}_r$ with its corresponding normal approximation $\bar{p}_r$ for right tail for different values of X.

Table 1. *Saddlepoint Approximation of left tail for stopped sum of Negative Binomial (1, 0.8)-Bernoulli (0.8) Distribution.*

X	First-correction		Second-correction		Third-correction	
	$\hat{p}_{r_1}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_2}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_3}(X \geq x)$	$\bar{p}_r(X \geq x)$
1.4	0.678916	0.687933	0.6848876	0.732371	0.68072320	0.732371
1.6	0.6446826	0.666402	0.643150483	0.715661	0.643233575	0.715661
1.8	0.6103359	0.648027	0.6130298	0.694974	0.61071855	0.694974
2	0.579629	0.625516	0.579312124	0.677242	0.577568404	0.677242
2.5	0.505601686	0.575345	0.506191663	0.625516	0.505346702	0.625516

Table 2. *Saddlepoint Approximation of center for stopped sum of Negative Binomial (1, 0.8)-Bernoulli (0.8) Distribution.*

X	First-correction		Second-correction		Third-correction	
	$\hat{p}_{r_1}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_2}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_3}(X \geq x)$	$\bar{p}_r(X \geq x)$
3.1	0.428685625	0.507978	0.429188026	0.563559	0.428901781	0.563559
3.2*	0.4201983525*	0.5*	0.4178001145*	0.551717*	0.4175755231*	0.551717*
3.3	0.41181527	0.492022	0.406470738	0.539828	0.406301386	0.539828
3.4	0.399547177	0.480061	0.395202748	0.531881	0.395082909	0.531881
3.5	0.391368183	0.468119	0.383999512	0.519939	0.38392406	0.519939

(*) It is the value of X equal the value of μ.

Table 3. *Saddlepoint Approximation of right tail for stopped sum of Negative Binomial (1, 0.8)-Bernoulli (0.8) Distribution.*

X	First-correction		Second-correction		Third-correction	
	$\hat{p}_{r_1}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_2}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_3}(X \geq x)$	$\bar{p}_r(X \geq x)$
4	0.340131398	0.416834	0.341007223	0.468119	0.341093312	0.468119
4.5	0.299365433	0.363169	0.295913956	0.416834	0.296091559	0.416834
5	0.19870192	0.312067	0.263780486	0.363169	0.263999343	0.363169
5.5	0.236048875	0.267629	0.19870192	0.312067	0.23353654	0.312067
6	0.209477656	0.223627	0.236048875	0.267629	0.209430994	0.267629

4. Stopped Sum of Geometric-Geometric Distribution

The stopped sum of Geometric- Geometric distribution where N~Geometric (p), X′s~Geometric (p)

The MGF for N is given by:

$$M_N(s) = \frac{pe^s}{1-qe^s}$$

The MGF for X is given by:

$$M_X(s) = \frac{pe^s}{1-qe^s}$$

The cumulant generating function for N is given by:

$$K_N(s) = \ln\left(M_N(s)\right)$$

$$K_N(s) = \ln\left(\frac{pe^s}{1-qe^s}\right)$$

The cumulant generating function for X is given by:

$$K_X(s) = \ln(M_X(s))$$

$$K_X(s) = \ln\left(\frac{pe^s}{1-qe^s}\right)$$

The cumulant generating function for the stopped sum Geometric–Geometric distribution as follows:

$$K_{S_N}(s) = K_N\left(K_X(s)\right)$$

$$K_{S_N}(s) = \ln\left(\frac{p^2e^s}{1-qe^s} \div \left(1-\frac{qpe^s}{1-qe^s}\right)\right)$$

The Saddlepoint equation $K'_{S_N}(\hat{s}) = x$ is becomes

$$K'_{S_N}(\hat{s}) = -\frac{1}{(p+1)qe^{\hat{s}}-1} = x$$

$$\hat{s} = \ln\left(\frac{-1+x}{x(p+1)q}\right)$$

For the first continuity-correction, we find

$$\widehat{W}_1 = \operatorname{sgn}(\hat{s})\sqrt{2\left\{x\ln\left(\frac{-1+x}{x(p+1)q}\right) - \ln\left(\frac{p^2e^{\hat{s}}}{1-qe^{\hat{s}}} \div 1 - \frac{qpe^{\hat{s}}}{1-qe^{\hat{s}}}\right)\right\}},$$

$$\hat{u}_1 = \left\{1 - e^{-\left(\ln\left(\frac{-1+x}{x(p+1)q}\right)\right)}\right\}\sqrt{\left(\frac{(p+1)qe^{\hat{s}}}{\left((p+1)qe^{\hat{s}}\right)^2}\right)}.$$

The second continuity-correction implies

$$\widetilde{W}_2 = \operatorname{sgn}(\tilde{s})\sqrt{2\left\{\ln\left(\frac{-1+x^-}{x^-(p+1)q}\right)x^- - \ln\left(\frac{p^2e^{\tilde{s}}}{1-qe^{\tilde{s}}} \div \left(1-\frac{qpe^{\tilde{s}}}{1-qe^{\tilde{s}}}\right)\right)\right\}},$$

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

$$\tilde{u}_2 = 2\sinh\left(\frac{\ln\left(\frac{-1+x^-}{x^-(p+1)q}\right)}{2}\right)\sqrt{\left(\frac{(p\pm1)qe^{\tilde{s}}}{\left((p+1)qe^{\tilde{s}}\right)^2}\right)}.$$

And the third continuity-correction is used

$$\tilde{u}_3 = \tilde{s}\sqrt{\left(\frac{(p+1)qe^{\tilde{s}}}{\left((p+1)qe^{\tilde{s}}\right)^2}\right)}.$$

The mean and the variance are given as

$$E(N) = \frac{1}{p}, E(X) = \frac{1}{p}$$

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

$$E_{S_N}(s) = \frac{1}{p^2}$$

and $V(N) = \frac{q}{p^2}, V(X) = \frac{q}{p^2}$

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

$$V_{S_N}(s) = \left(\frac{1}{p}\right)\left(\frac{q}{p^2}\right) + \left(\frac{1}{p}\right)^2\left(\frac{q}{p^2}\right).$$

However, Table 4, 5 and 6 show the three continuity correction Saddlepoint $\hat{p}_r$ with normal approximation $\bar{p}_r$.

Table 4. *Saddlepoint Approximation of left tail for stopped sum of Geometric (0.2)-Geometric (0.2) Distribution.*

X	First-correction		Second-correction		Third-correction	
	$\hat{p}_{r_1}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_2}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_3}(X \geq x)$	$\bar{p}_r(X \geq x)$
5	0.84629541	0.79103	0.8457456	0.796731	0.845414516	0.796731
5.5	0.830123456	0.785236	0.828650153	0.79103	0.8283693	0.79103
10	0.688267006	0.79069	0.68746681	0.735653	0.687376791	0.735653
10.5	0.6768012	0.722405	0.664750462	0.729069	0.664662614	0.729069
15	0.562603004	0.655422	0.561957482	0.662757	0.561925557	0.662757

Table 5. *Saddlepoint Approximation of center for stopped sum of Geometric (0.2)-Geometric (0.2) Distribution.*

X	First-correction		Second-correction		Third-correction	
	$\hat{p}_{r_1}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_2}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_3}(X \geq x)$	$\bar{p}_r(X \geq x)$
15.5	0.550818232	0.648027	0.550437372	0.655422	0.55040861	0.655422
20	0.455870252	0.57926	0.457498426	0.587064	0.457488384	0.587064
20.5	0.448003771	0.571424	0.445908292	0.57926	0.445899577	0.57926
25*	0.375134914	0.5	0.374319562	0.507978	0.374318978	0.507978
25.6	0.367805786	0.492022	0.366539512	0.5	0.366539632	0.5

(*) It is the value of X equal the value of μ.

Table 6. *Saddlepoint Approximation of right tail for stopped sum of Geometric (0.2)-Geometric (0.2) Distribution.*

X	First-correction		Second-correction		Third-correction	
	$\hat{p}_{r_1}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_2}(X \geq x)$	$\bar{p}_r(X \geq x)$	$\hat{p}_{r_3}(X \geq x)$	$\bar{p}_r(X \geq x)$
30	0.306770173	0.42074	0.306998415	0.428576	0.307001993	0.428576
30.5	0.303713126	0.412936	0.300119086	0.42074	0.300122915	0.42074
35	0.250928221	0.344578	0.251887361	0.351973	0.251892614	0.351973
35.5	0.250928221	337243	0.245498595	0.344573	0.245503934	0.344573
40	0.207255669	0.270931	0.204893076	0.722405	0.204898775	0.722405

5. Conclusion

From the mathematical computations in the left, centre and right, it is clear that both of the Saddlepoint Approximation and the normal are very close to each other which leads to the performance of the Saddlepoint method.

Acknowledgements

The authors would like to thank the Editor and the referee for carefully reading the manuscript and for their valuable comments and suggestions which greatly improved this paper.

References

[1] Kuonen, Diego. "Computer-intensive statistical methods: saddlepoint approximations with applications in bootstrap and robust inference." *PhD Thesis, Swiss Federal Institute of Technology* (2001).

[2] Robin, Stéphane. "A compound Poisson model for word occurrences in DNA sequences." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51.4 (2002): 437-451.

[3] Johnson, NL, Kotz, S.; Kemp, AW: Univariate discrete distributions. 2nd Ed. New York: John Wiley & Sons 1992.

[4] Johnson, Norman L., Adrienne W. Kemp, and Samuel Kotz. *Univariate discrete distributions*. Vol. 444. John Wiley & Sons, 2005.

[5] Minkova, Leda D. "The Pólya-Aeppli process and ruin problems." *International Journal of Stochastic Analysis* 2004.3 (1900): 221-234.

[6] Hogg, R. V., and A. T. Craig. "Introduction to Mathematical Statistics. Macmillan Publishing Company." *NY, NY* (1978).

[7] Al Mutairi Alya, O., and Heng Chin Low. "Saddlepoint Approximation to Cumulative Distribution Function for Poisson–Exponential Distribution." *Modern Applied Science* 7.3 (2013): 26.

[8] Al Mutairi, O., and Heng Chin Low. "Estimations of the Central Tendency Measures of the Random-sum Poisson-Weibull Distribution using Saddlepoint Approximation." *Journal of Applied Sciences* 14 (2014): 1889-1893.

[9] Shell, Richard. *Handbook of industrial automation*. CRC Press, 2000.

[10] Butler, Ronald W. *Saddlepoint approximations with applications*. Vol. 22. Cambridge University Press, 2007.