Songklanakarin J. Sci. Technol.
39 (3), 355-358, May - Jun. 2017

SJST
http://www.sjst.psu.ac.th

Original Article

A non-uniform bound on Poisson approximation for a sum of negative binomial random variables

Kanint Teerapabolarn[1,2*]

[1] *Department of Mathematics, Faculty of Science, Burapha University, Chonburi, 20131 Thailand*

[2] *Centre of Excellence in Mathematics, Commission on Higher Education, Ratchathewi, Bangkok, 10400 Thailand*

Received: 7 March 2016; Revised: 1 May 2016; Accepted: 8 June 2016

Abstract

This paper uses the Stein–Chen method to determine a non-uniform bound on the point metric between the distribution of a sum of independent negative binomial random variables and a Poisson distribution with mean $\lambda = \sum_{i=1}^{n} r_i q_i$, where r_i and $p_i = 1 - q_i$ are parameters of each negative binomial distribution. The result gives a good Poisson approximation when all q_i are small or λ is small.

Keywords: negative binomial distribution, non-uniform bound, point metric, Poisson approximation, Stein–Chen method.

1. Introduction

Let $X_1, ..., X_n$ be n independently distributed negative binomial random variables, each with probability function $p_{X_i}(x) = \frac{\Gamma(r_i + x)}{\Gamma(r_i)x!} q_i^x p_i^{r_i}$, $x \in \mathbb{N} \cup \{0\}$, mean $E(X_i) = \frac{r_i q_i}{p_i}$ and variance $Var(X_i) = \frac{r_i q_i}{p_i^2}$, where $q_i = 1 - p_i$. Let $S_n = \sum_{i=1}^{n} X_i$ and $\wp_\lambda$ denote the Poisson random variable with mean $\lambda > 0$. It can be seen that if all $r_i q_i$ are small, then the distribution of S_n can be approximated by a Poisson distribution with mean $\sum_{i=1}^{n} \frac{r_i q_i}{p_i}$ or $\sum_{i=1}^{n} r_i q_i$. For $\lambda = \sum_{i=1}^{n} r_i q_i$, Vellaisamy and Upadhye (2009) used the method of exponents to give a uniform bound in the form

$$d_A(S_n, \wp_\lambda) \le \sum_{i=1}^{n} \frac{r_i q_i^2}{p_i} \min\left\{1, \frac{1}{\sqrt{2\lambda e}}\right\}, \tag{1.1}$$

where $d_A(S_n, \wp_\lambda) = \sup_{A \subseteq \mathbb{N} \cup \{0\}} |P(S_n \in A) - P(\wp_\lambda \in A)|$ is the total variation distance between the distribution of S_n and the Poisson distribution with mean λ. For $\lambda = \sum_{i=1}^{n} \frac{r_i q_i}{p_i}$, Teerapabolarn (2014) used the Stein-Chen method and the *w*-function associated with each negative binomial random variable to give a uniform bound in the form

$$d_A(S_n, \wp_\lambda) \le \frac{1 - e^{-\lambda}}{\lambda} \sum_{i=1}^{n} \frac{r_i q_i^2}{p_i^2}. \tag{1.2}$$

It should be noted that, if we ignore coefficient factors in (1.1), $\min\left\{1, \frac{1}{\sqrt{2\lambda e}}\right\}$, and (1.2), $\frac{1-e^{-\lambda}}{\lambda}$, the bound corresponds to $\lambda = \sum_{i=1}^{n} r_i q_i$ in (1.1) is sharper than that correspond-

* Corresponding author.
Email address: kanint@buu.ac.th

ing to $\lambda = \sum_{i=1}^{n} \frac{r_i q_i}{p_i}$ in (1.2). In addition, the result in (1.1) is more applicable than that presented in (1.2), because it is easy to calculate the accuracy of the Poisson approximation. A non-uniform counterpart of the bound in (1.2) was determined by Teerapabolarn (2015). He used the Stein-Chen method and the *w*-function associated with each negative binomial random variable to give a non-uniform bound as follows:

$$d_{x_0}\left(S_n, \wp_\lambda\right) \le \min\left\{\frac{1-e^{-\lambda}}{\lambda}, \frac{1}{x_0}\right\} \sum_{i=1}^{n} \frac{r_i q_i^2}{p_i^2}, \quad (1.3)$$

where $d_{x_0}\left(S_n, \wp_\lambda\right) = \left|P(S_n = x_0) - P(\wp_\lambda = x_0)\right| = \left|P(S_n = x_0) - \frac{\lambda^{x_0} e^{-\lambda}}{x_0!}\right|$, $x_0 \in \mathbb{N} \cup \{0\}$, is the point metric between the distribution of S_n and the Poisson distribution with mean $\lambda = \sum_{i=1}^{n} \frac{r_i q_i}{p_i}$. For this case, another interesting point is determining a non-uniform counterpart of the bound in (1.1). Note that, for $x_0 = 0$, we can compute the exact probability of $S_n = x_0$, that is, $P(S_n = x_0) = \prod_{i=1}^{n} p_i^{r_i}$. So, in this paper, we are interested to determine a non-uniform bound with respect to the Poisson mean $\lambda = \sum_{i=1}^{n} r_i q_i$ for $d_{x_0}\left(S_n, \wp_\lambda\right)$ when $x_0 \in \mathbb{N}$.

The Stein-Chen method is the tool for giving the main result, which is utilized to provide the desired result as mentioned in Section 2. In Section 3, we use the Stein-Chen method to determine a non-uniform bound for $d_{x_0}\left(S_n, \wp_\lambda\right)$ and the conclusion of this study is presented in the last section.

2. Methods

Stein (1972) introduced a powerful and general method for bounding the error in the normal approximation. This method was developed and applied in the setting of the Poisson approximation by Chen (1975), and is referred to as the Stein-Chen method. Stein's equation for the Poisson distribution with mean $\lambda > 0$, for given *h*, is of the form

$$h(x) - P_\lambda(h) = \lambda f(x+1) - x f(x), \quad (2.1)$$

where $P_\lambda(h) = e^{-\lambda} \sum_{k=0}^{\infty} h(k) \frac{\lambda^k}{k!}$ and f and h are real valued bounded functions defined on $\mathbb{N} \cup \{0\}$. For $A \subseteq \mathbb{N} \cup \{0\}$, let $h_A : \mathbb{N} \cup \{0\} \to \mathbb{R}$ be defined by

$$h_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases} \quad (2.2)$$

For convenience, we will write $h_{\{x\}}$ by h_x, and for $x_0 \in \mathbb{N} \cup \{0\}$, the solution f_{x_0} of (2.1) can be expressed in the form

$$f_{x_0}(x) = \begin{cases} \frac{(x-1)!}{x_0!} \lambda^{x_0-x} P_\lambda(1 - h_{C_{x-1}}), & \text{if } x_0 < x, \\ -\frac{(x-1)!}{x_0!} \lambda^{x_0-x} P_\lambda(h_{C_{x-1}}), & \text{if } x_0 \ge x > 0, \\ 0, & \text{if } x = 0, \end{cases} \quad (2.3)$$

where $x \in \mathbb{N}$ and $C_{x-1} = \{0, 1, ..., x-1\}$.

The following lemma gives a non-uniform bound for $\left|f_{x_0}\right|$, which is also need for giving the desired result.

Lemma 2.1. *Let* $x_0, x \in \mathbb{N}$, *then the following inequality holds*:

$$\sup_{x \ge 2}\left|f_{x_0}(x)\right| \le \begin{cases} \frac{1-e^{-\lambda}(1+\lambda)}{\lambda}, & \text{if } x_0 = 1, \\ \max\left\{\frac{1-P(\wp_\lambda \le x_0-1)}{x_0+1}, \frac{P(\wp_\lambda \le x_0-1)}{x_0}\right\}, & \text{if } x_0 \ge 2, \end{cases} \quad (2.4)$$

where $P(\wp_\lambda \le x_0 - 1) = \sum_{j=0}^{x_0-1} \frac{e^{-\lambda}\lambda^j}{j!}$.

Proof. For $x_0 = 1$, when $x \ge 2$, it follows from (2.3) that $f_{x_0}(x) > 0$, thus

$$\begin{aligned} \left|f_{x_0}(x)\right| &= f_1(x) \\ &= (x-1)! \sum_{j=x}^{\infty} \frac{e^{-\lambda}\lambda^{j-x+1}}{j!} \\ &= (x-1)! e^{-\lambda}\left\{\frac{\lambda}{x!} + \frac{\lambda^2}{(x+1)!} + \frac{\lambda^3}{(x+2)!} + \cdots\right\} \\ &= e^{-\lambda}\left\{\frac{\lambda}{x} + \frac{\lambda^2}{x(x+1)} + \frac{\lambda^3}{x(x+1)(x+2)} + \cdots\right\} \\ &\le \frac{e^{-\lambda}}{\lambda}\left\{\frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \cdots\right\} \\ &= \frac{1-e^{-\lambda}(1+\lambda)}{\lambda}. \end{aligned} \quad (2.5)$$

For $x_0 \ge 2$, we divide the proof into two cases as follows:

Case 1. $x_0 < x$, Teerapabolarn and Neammanee (2005) showed that

$$\begin{aligned} \left|f_{x_0}(x)\right| &\le \frac{1}{x} e^{-\lambda}\left\{\frac{\lambda^{x_0}}{x_0!} + \frac{\lambda^{x_0+1}}{(x_0+1)!} + \frac{\lambda^{x_0+2}}{(x_0+2)!} + \cdots\right\} \\ &\le \frac{1}{x_0+1} \sum_{j=x_0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \\ &= \frac{1-P(\wp_\lambda \le x_0-1)}{x_0+1}. \end{aligned} \quad (2.6)$$

Case 2. $x_0 \ge x$, following Barbour, *et al*. (1992), f_{x_0} is a negative and decreasing function for $x \in \{1, ..., x_0\}$, thus

$$\begin{aligned} \left|f_{x_0}(x)\right| &= -f_{x_0}(x) \\ &\le -f_{x_0}(x_0) \end{aligned}$$

$$= \frac{1}{x_0} P_\lambda(h_{C_{x_0-1}}) \quad \text{(by (2.3))}$$

$$= \frac{1}{x_0} P(\wp_\lambda \le x_0 - 1). \qquad (2.7)$$

Therefore, from case 1 and case 2, it follows that

$$\left| f_{x_0}(x) \right| \le \max\left\{ \frac{1-P(\wp_\lambda \le x_0-1)}{x_0+1}, \frac{P(\wp_\lambda \le x_0-1)}{x_0} \right\}. \qquad (2.8)$$

Hence, by (2.5) and (2.8), the inequality (2.4) holds.

3. Main Results

The following theorem presents a non-uniform bound for the point metric between the distribution of S_n and the Poisson distribution with mean $\lambda = \sum_{i=1}^{n} r_i q_i$.

Theorem 3.1. *Let* $x_0 \in \mathrm{N}$ *and* $\lambda = \sum_{i=1}^{n} r_i q_i$, *then we have the following.*

$$d_{x_0}(S_n, \wp_\lambda) \le \begin{cases} \frac{1-e^{-\lambda}(1+\lambda)}{\lambda} \sum_{i=1}^{n} \frac{r_i q_i^2}{p_i} & , \textit{if } x_0 = 1, \\ \max\left\{ \frac{1-P(\wp_\lambda \le x_0-1)}{x_0+1}, \frac{P(\wp_\lambda \le x_0-1)}{x_0} \right\} \sum_{i=1}^{n} \frac{r_i q_i^2}{p_i}, & \textit{if } x_0 \ge 2. \end{cases} \qquad (3.1)$$

Proof. Substituting h by h_{x_0} and x by S_n and also taking expectation in (2.1), we have

$$P(S_n = x_0) - P(\wp_\lambda = x_0)$$

$$= E\left[\lambda f(S_n+1) - S_n f(S_n)\right]$$

$$= E\left[\sum_{i=1}^{n} r_i q_i f(S_n+1) - \sum_{i=1}^{n} X_i f(S_n) \right]$$

$$= \sum_{i=1}^{n} E\left[r_i q_i f(S_n+1) - X_i f(S_n) \right], \qquad (3.2)$$

where $f = f_{x_0}$ is defined in (2.3). For $i = 1, \ldots, n$, let $S_n^i = S_n - X_i$, then we have

$$E\left[r_i q_i f(S_n+1) - X_i f(S_n) \right]$$

$$= E\left[r_i q_i f(S_n^i + X_i + 1) - X_i f(S_n^i + X_i) \right]$$

$$= E\left\{ E\left[(r_i q_i f(S_n^i + X_i + 1) - X_i f(S_n^i + X_i)) \middle| X_i \right] \right\}$$

$$= \sum_{x=0}^{\infty} E\left[(r_i q_i f(S_n^i + X_i + 1) - X_i f(S_n^i + X_i)) \middle| X_i = x \right] p_{X_i}(x)$$

$$= E\left[r_i q_i f(S_n^i + 1) \right] p_{X_i}(0) + E\left[r_i q_i f(S_n^i + 2) - f(S_n^i + 1) \right] p_{X_i}(1)$$

$$+ E\left[r_i q_i f(S_n^i + 3) - 2 f(S_n^i + 2) \right] p_{X_i}(2)$$

$$+ E\left[r_i q_i f(S_n^i + 4) - 3 f(S_n^i + 3) \right] p_{X_i}(3)$$

$$+ E\left[r_i q_i f(S_n^i + 5) - 4 f(S_n^i + 4) \right] p_{X_i}(4) + \cdots$$

$$= r_i q_i p_i^{r_i} E\left[f(S_n^i + 1) \right] + r_i^2 q_i^2 p_i^{r_i} E\left[f(S_n^i + 2) \right]$$

$$- r_i q_i p_i^{r_i} E\left[f(S_n^i + 1) \right]$$

$$+ \frac{r_i^2 (r_i+1) q_i^3 p_i^{r_i} E\left[f(S_n^i + 3) \right]}{2} - \frac{2 r_i (r_i+1) q_i^2 p_i^{r_i} E\left[f(S_n^i + 2) \right]}{2}$$

$$+ \frac{r_i^2 (r_i+1)(r_i+2) q_i^4 p_i^{r_i} E\left[f(S_n^i + 4) \right]}{3!} - \frac{r_i (r_i+1)(r_i+2) q_i^3 p_i^{r_i} E\left[f(S_n^i + 3) \right]}{2}$$

$$+ \frac{r_i^2 (r_i+1)(r_i+2)(r_i+3) q_i^5 p_i^{r_i} E\left[f(S_n^i + 5) \right]}{4!}$$

$$- \frac{r_i (r_i+1)(r_i+2)(r_i+3) q_i^4 p_i^{r_i} E\left[f(S_n^i + 4) \right]}{3!} + \cdots$$

$$= - r_i q_i^2 p_i^{r_i} E\left[f(S_n^i + 2) \right] - r_i (r_i+1) q_i^3 p_i^{r_i} E\left[f(S_n^i + 3) \right]$$

$$- \frac{r_i (r_i+1)(r_i+2) q_i^4 p_i^{r_i} E\left[f(S_n^i + 4) \right]}{2}$$

$$- \frac{r_i (r_i+1)(r_i+2)(r_i+3) q_i^5 p_i^{r_i} E\left[f(S_n^i + 5) \right]}{3} - \cdots$$

$$- \frac{r_i (r_i+1)(r_i+2)(r_i+3) \cdots (r_i+x-1) q_i^{x+1} p_i^{r_i} E\left[f(S_n^i + x + 1) \right]}{x-1} - \cdots$$

$$= - \sum_{x=1}^{\infty} x q_i p_{X_i}(x) E\left[f(S_n^i + x + 1) \right]. \qquad (3.3)$$

Putting the result (3.3) into (3.2), gives

$$P(S_n = x_0) - P(\wp_\lambda = x_0)$$

$$= - \sum_{i=1}^{n} \sum_{x=1}^{\infty} x q_i p_{X_i}(x) E\left[f(S_n^i + x + 1) \right].$$

From which, it follows that

$$d_{x_0}(S_n, \wp_\lambda) \le \sum_{i=1}^{n} \sum_{x=1}^{\infty} x q_i p_{X_i}(x) E\left| f(S_n^i + x + 1) \right|$$

$$\le \sum_{i=1}^{n} \sum_{x=1}^{\infty} x q_i p_{X_i}(x) \sup_x \left| f(x+1) \right|$$

$$\le \sup_{k \ge 2} \left| f(k) \right| \sum_{i=1}^{n} \sum_{x=1}^{\infty} x q_i p_{X_i}(x)$$

$$= \sup_{k \ge 2} \left| f(k) \right| \sum_{i=1}^{n} \frac{r_i q_i^2}{p_i}$$

By using Lemma 2.1, the inequality (3.1) is obtained.

If $r_1 = r_2 = \cdots = r_n = 1$, then $\lambda = \sum_{i=1}^{n} q_i$, and the result in Theorems 3.1 becomes to be a Poisson approximation for a sum of independent geometric random variables.

Corollary 3.1. *For $r_1 = r_2 = \cdots = r_n = 1$ and $\lambda = \sum_{i=1}^{n} q_i$, then we have the following.*

$$d_{x_0}\left(S_n, \wp_\lambda\right) \le \begin{cases} \frac{1-e^{-\lambda}(1+\lambda)}{\lambda} \sum_{i=1}^{n} \frac{q_i^2}{p_i} & , \text{ if } x_0 = 1, \\ \max\left\{ \frac{1-P(\wp_\lambda \le x_0 - 1)}{x_0+1}, \frac{P(\wp_\lambda \le x_0 - 1)}{x_0} \right\} \sum_{i=1}^{n} \frac{q_i^2}{p_i}, & \text{if } x_0 \ge 2. \end{cases} \tag{3.2}$$

4. Conclusions

In this study, a non-uniform bound for the point metric between the distribution of a sum of independent negative binomial random variables and a Poisson distribution with mean $\lambda = \sum_{i=1}^{n} r_i q_i$ was obtained. In view of this bound, it is found that the result give a good Poisson approximation when all q_i are small or λ is small. In other words, if all q_i are small or λ is small, then the Poisson distribution with mean $\lambda = \sum_{i=1}^{n} r_i q_i$ can be used as a good estimate of the distribution of these independent summands.

Acknowledgements

The author would like to thank the referees for many helpful comments which have led to the improvement.

References

Barbour, A. D., Holst, L., & Janson, S. (1992). *Poisson approximation* (Oxford studies in probability 2). Oxford, England: Clarendon Press.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Annals of Probability*, *3*(3), 534-545

Stein, C. M. (1972). A bound for the error in normal approximation to the distribution of a sum of dependent random variables. In L. M. Le Cam, J. Neyman, & E. L. Scott (Eds), *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 583-602), Berkeley, CA: University of California Press.

Teerapabolarn, K. (2014). Poisson approximation for independent negative binomial random variables. *International Journal of Pure and Applied Mathematics*, *93*(6), 779-781.

Teerapabolarn, K. (2015). Pointwise Poisson approximation for independent negative binomial random variables. *Global Journal of Pure and Applied Mathematics*, *11*(4), 1979-1982.

Teerapabolarn K., & Neammanee K. (2005). A non-uniform bound on Poisson approximation for dependent trials. *Stochastic Modelling and Applications*, *8*(1), 17-31.

Vellaisamy, P., & Upadhye, N.S. (2009). Compound negative binomial approximations for sums of random variables. *Probability and Mathematical Statistics*, *29*(2), 205-226.