International Journal of Pure and Applied Mathematics
Volume 94 No. 5 2014, 697-700
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: http://dx.doi.org/10.12732/ijpam.v94i5.7
ijpam.eu

BINOMIAL APPROXIMATION FOR RANDOM SUMS OF BERNOULLI RANDOM VARIABLES

K. Teerapabolarn
Department of Mathematics
Faculty of Science
Burapha University
Chonburi, 20131, THAILAND

Abstract: In this paper, we give a bound for the total variation distance between the distribution of random sums of independent Bernoulli random variables and a binomial distribution. Two examples have been given to illustrate the result obtained.

AMS Subject Classification: 62E17, 60F05, 60G05
Key Words: Bernoulli random variable, binomial approximation, random sums, total variation distance

1. Introduction

Let $X_1, X_2, ...$ be a sequence of independent Bernoulli random variables, each with probability $p_i = 1 - q_i = P(X_i = 1) = 1 - P(X_i = 0)$. Let $S_N = \sum_{i=1}^{N} X_i$, where N is a non-negative integer-valued random variable and independent of the X_i's. The random summands is usually called *random sums*. Let $B_{n,p}$ be a binomial random variable with parameters n and p. For $N = n \in \mathbb{N}$ is fixed, Ehm [2] gave a bound for approximating the distribution of S_n by a binomial distribution in the form of

Received: March 26, 2014

$$d_{TV}(S_n, B_{n,p}) \leq \frac{1-p^{n+1}-q^{n+1}}{(n+1)pq}\sum_{i=1}^{n}(p_i-p)^2,$$

where $d_{TV}(S_n, B_{n,p}) = \sup_{A\subseteq\{0,1,\ldots,n\}} |P(S_n \in A) - P(B_{n,p} \in A)|$ is the total variation distance between the distribution of S_n and the binomial distribution with parameters n and $p = 1-q = \frac{1}{n}\sum_{i=1}^{n} p_i$. Let $\widehat{n} = E(N)$ and $\widehat{p} = 1-\widehat{q} = \frac{\lambda}{\widehat{n}}$, where $\widehat{n} \in \mathbb{N}$ and $\lambda = E(\lambda_N) = E\left(\sum_{i=1}^{N} p_i\right)$. In this study, we are interested to determine a bound for $d_{TV}(S_N, B_{\widehat{n},\widehat{p}})$, which is in Section 2. In Section 3, two examples have been given to illustrate the desired result, and the conclusion of this study is presented in the last section.

2. Result

We give a bound for the total variation distance between the distributions of S_N and $B_{\widehat{n},\widehat{p}}$ as follows.

Theorem 2.1. *For* $\lambda_N = \sum_{i=1}^{N} p_i$, $\lambda = E(\lambda_N)$ *and* $\widehat{n} \in \mathbb{N}$, *then*

$$d_{TV}(S_N, B_{\widehat{n},\widehat{p}}) \leq (1-e^{-\lambda})\left\{\frac{Var(\lambda_N)}{\lambda} + \frac{\lambda}{\widehat{n}}\right\} + E\left(\frac{1-e^{-\lambda_N}}{\lambda_N}\sum_{i=1}^{N} p_i^2\right). \tag{2.1}$$

Proof. Let U_λ be the Poisson random variable with mean λ. It follows the fact that

$$\begin{aligned} d_{TV}(S_N, B_{\widehat{n},\widehat{p}}) &\leq d_{TV}(S_N, U_\lambda) + d_{TV}(U_\lambda, B_{\widehat{n},\widehat{p}}) \\ &\leq d_{TV}(S_N, U_{\lambda_N}) + d_{TV}(U_{\lambda_N}, U_\lambda) + d_{TV}(U_\lambda, B_{\widehat{n},\widehat{p}}). \end{aligned} \tag{2.2}$$

Following [3], we have

$$d_{TV}(S_N, U_{\lambda_N}) \leq E\left(\frac{1-e^{-\lambda_N}}{\lambda_N}\sum_{i=1}^{N} p_i^2\right). \tag{2.3}$$

Applying Theorem 1.C and using inequality (1.23) in [1], we also obtain

$$d_{TV}(U_{\lambda_N}, U_\lambda) \leq (1-e^{-\lambda})\frac{Var(\lambda_N)}{\lambda} \tag{2.4}$$

and

$$d_{TV}(U_\lambda, B_{\widehat{n},\widehat{p}}) \leq (1-e^{-\lambda})\widehat{p} = (1-e^{-\lambda})\frac{\lambda}{\widehat{n}}, \tag{2.5}$$

respectively. Hence, the inequality (2.1) is obtained by taking the bounds in (2.3), (2.4) and (2.5) into (2.2). □

If X_i's are identically distributed, then the following corollary is an immediately consequence of the Theorem 2.1

Corollary 2.1. *For $\widehat{n} \in \mathbb{N}$, if $p_1 = p_2 = \cdots = p$, then we have the following:*

$$d_{TV}(S_N, B_{\widehat{n},p}) \leq (1 - e^{-\widehat{n}p}) \left\{ \frac{Var(N)}{\widehat{n}} + 1 \right\} p + E\left(1 - e^{-Np}\right) p. \qquad (2.6)$$

3. Examples

This section, two examples are given to illustrate the result in the case of X_i's are identically distributed.

Example 3.1. For n ($n \in \mathbb{N}$) is fixed, let N be a positive integer-valued random variable with probability function

$$P(N = k) = \begin{cases} \frac{1}{2} & , k = 2n, \\ \frac{1}{2} & , k = 4n, \\ 0 & , \text{otherwise}. \end{cases}$$

Therefore $\widehat{n} = 3n$ and $Var(N) = n^2$. Let $p_1 = p_2 = \cdots = p$, then we have

$$d_{TV}(S_N, B_{3n,p}) \leq (1 - e^{-3np}) \frac{(n+3)p}{3} + \left(1 - \frac{e^{-2np} + e^{-4np}}{2}\right) p.$$

Example 3.2. Let N be a positive integer-valued random variable with probability function

$$P(N = n) = \frac{1}{20}, \; n = 1, 2, ..., 21,$$

then we have $\widehat{n} = 11$ and $Var(N) = \frac{891}{20}$. If $p_1 = p_2 = \cdots = p$, then we obtain

$$d_{TV}(S_N, B_{11,p}) \leq \left\{ (1 - e^{-11p}) \frac{101}{20} + 1 \right\} p.$$

4. Conclusion

In this study, a bound for the total variation distance between the distribution of random sums of independent Bernoulli random variables and an appropriate binomial distribution was obtained. It is indicated that the binomial with parameters $\widehat{n}$ and $\widehat{p}$ can be used as an estimate of the distribution of random sums of independent Bernoulli random variables when $\widehat{p}$ is small.

References

[1] A.D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Oxford Studies in probability 2, Clarendon Press, Oxford, 1992.

[2] W. Ehm, Binomial approximation to the Poisson binomial distribution, *Statist. Probab. Lett.*, **11** (1991), 7–16.

[3] N. Yannaros, Poisson approximation for random sums of Bernoulli random variables, *Statist. Probab. Lett.*, **11** (1991), 161–165.