International Journal of Mathematics and
Computer Science, **17**(2022), no. 2, 601–609

The Exact Formula for the Supremum of Poisson-Poisson Binomial Relative Error

Kanint Teerapabolarn

Department of Mathematics
Faculty of Science
Burapha University
Chonburi, Thailand

Centre of Excellence in Mathematics
Bangkok, Thailand

email: kanint@buu.ac.th

(Received September 22, 2021, Accepted October 25, 2021)

Abstract

A mathematical inequality is used to determine the exact formula for the supremum of relative error between the Poisson binomial cumulative distribution function with parameter $\mathbf{p} = (p_1, ..., p_n)$ and a Poisson cumulative distribution function with mean $\sum_{i=1}^{n} \frac{p_i}{1-p_i}$. With this formula, the Poisson cumulative distribution function with this mean can be used as a good estimate of the Poisson binomial cumulative distribution function when all p_i are small.

1 Introduction

Let $X = \sum_{i=1}^{n} Y_i$, where $Y_i,\ i = 1, 2, ..., n$, are independent Bernoulli random variables with probability of success $p_i = P(Y_i = 1) = 1 - P(Y_i = 0)$. Then the distribution of X is called the *Poisson binomial* distribution with parameter $\mathbf{p} = (p_1, ..., p_n)$. The distribution may be referred to as the distribution of the number of successes in a sequence of n independent Bernoulli trials,

---

Key words and phrases: Cumulative distribution function, Poisson distribution, Poisson binomial distribution, relative error.
AMS (MOS) Subject Classifications: 60F05, 60G50, 62E17.
ISSN 1814-0432, 2022, http://ijmcs.future-in-tech.net

where success occurs on the i^{th} trial with a probability of p_i and failure occurs on the i^{th} trial with a probability of $q_i = 1 - p_i$. Following Chen and Liu [4], the probability mass function of X can be expressed as

$$P(X = x) = \sum_{d_1+\cdots+d_n=x} \prod_{i=1}^{n} \left(\frac{p_i}{q_i}\right)^{d_i} \left(\prod_{i=1}^{n} q_i\right), \ x = 0, 1, ..., n, \tag{1.1}$$

where $d_1, ..., d_n \in \{0, 1\}$, and the mean and variance of X are $E(X) = \sum_{i=1}^{n} p_i$ and $Var(X) = \sum_{i=1}^{n} p_i q_i$, respectively. For limiting distribution of X, it is well known that if $n \to \infty$ and all $p_i \to 0$ while $\lambda = \sum_{i=1}^{n} p_i$ remains fixed, then the Poisson binomial distribution with parameter $\mathbf{p}$ converges to a Poisson distribution with mean λ due to the well known *Law of Small Numbers.* In practice, the Poisson distribution with mean λ can be used as an approximation of the Poisson binomial distribution with parameter $\mathbf{p}$ when all p_i are sufficiently small. In the past few years, many authors have tried to approximate this distribution by the Poisson distribution with this mean, and they also investigated each upper bound as a criteria for measuring the accuracy of the approximation, such as Le Cam [8], Hodges and Le cam [6], Kerstan [7], Chen [3], Barbour and Hall [1], Barbour et al. [2], Neammanee [9], Daley and Vere-Jones [5], Teerapabolarn [11].

Let $\wp_\lambda$ denote the Poisson random variable with mean $\lambda = \sum_{i=1}^{n} \frac{p_i}{q_i}$, $q_i \neq 0$, and its probability mass function is of the form

$$P(\wp_\lambda = j) = \frac{e^{-\lambda}\lambda^j}{j!}, \tag{1.2}$$

where $j \in \mathbb{N} \cup \{0\}$. Following the multinomial theorem, it follows that

$$\begin{aligned} P(\wp_\lambda = j) &= e^{-\lambda} \frac{\left(\sum_{i=1}^n \frac{p_i}{q_i}\right)^j}{j!} \\ &= e^{-\lambda} \frac{1}{j!} \left(\sum_{i=1}^n \frac{p_i}{q_i}\right)^j \\ &= e^{-\lambda} \frac{1}{j!} \sum_{k_1+\cdots+k_n=j} \frac{j!}{k_1! \cdots k_n!} \prod_{i=1}^n \left(\frac{p_i}{q_i}\right)^{k_i} \\ &= e^{-\lambda} \sum_{k_1+\cdots+k_n=j} \frac{1}{k_1! \cdots k_n!} \prod_{i=1}^n \left(\frac{p_i}{q_i}\right)^{k_i} \\ &= e^{-\lambda} \sum_{d_1+\cdots+d_n=j} \prod_{i=1}^n \left(\frac{p_i}{q_i}\right)^{d_i} \\ &\quad + \sum_{\substack{k_1+\cdots+k_n=j \\ (k_1,\ldots,k_n) \neq (d_1,\ldots,d_n)}} \frac{e^{-\lambda}}{k_1! \cdots k_n!} \prod_{i=1}^n \left(\frac{p_i}{q_i}\right)^{k_i}, \end{aligned} \tag{1.3}$$

where $k_1, ..., k_n \in \{0, 1, .., j\}$. From (1.1) and (1.3), we observe that if $n \to \infty$ and all $p_i \to 0$ while $\lambda = \sum_{i=1}^n \frac{p_i}{q_i}$ remains fixed and $\prod_{i=1}^n q_i \to e^{-\lambda}$, then

$$\sum_{\substack{k_1+\cdots+k_n=j \\ (k_1,\ldots,k_n) \neq (d_1,\ldots,d_n)}} \frac{e^{-\lambda}}{k_1! \cdots k_n!} \prod_{i=1}^n \left(\frac{p_i}{q_i}\right)^{k_i} \to 0,$$

which implies that $P(X = x) \to P(\wp_\lambda = x)$ for every $x \in \{0, 1, .., n\}$. However, in practice, it is possible to approximate the Poisson binomial distribution with parameter $\mathbf{p}$ by a Poisson distribution with mean $\lambda = \sum_{i=1}^n \frac{p_i}{q_i}$ as well when all p_i are sufficiently small. Similarly, the Poisson binomial cumulative distribution function can also be approximated by a Poisson cumulative distribution function with mean $\lambda = \sum_{i=1}^n \frac{p_i}{q_i}$. In this case, Teerapabolarn [10] gave the result in the Poisson approximation as follows:

$$0 \leq 1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\mathbf{p}}(x_0)} \leq \frac{\lambda^{-2}(e^\lambda - \lambda - 1)}{x_0 + 1} \sum_{i=1}^n \frac{p_i^2}{q_i} \tag{1.4}$$

for $x_0 \in \{1, ..., n\}$, where $1 - \frac{\mathbb{P}_\lambda(0)}{\mathbb{PB}_{\mathbf{p}}(0)} = \frac{\mathbb{PB}_{\mathbf{p}}(0) - \mathbb{P}_\lambda(0)}{\mathbb{PB}_{\mathbf{p}}(0)} = 1 - \frac{e^{-\lambda}}{\Pi_{i=1}^n q_i}$ and $\mathbb{PB}_{\mathbf{p}}(x_0)$ and $\mathbb{P}_\lambda(x_0)$ are the Poisson binomial and Poisson cumulative distribution functions at x_0. Before 2013, there were no articles related to the Poisson mean, $\lambda = \sum_{i=1}^{n} \frac{p_i}{q_i}$, and the approximation form, $1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)} = \frac{\mathbb{PB}_\mathbf{p}(x_0) - \mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)}$. With this bound, the Poisson distribution with mean λ can be used as a good approximation of the Poisson binomial distribution with parameter $\mathbf{p}$ when $\frac{\lambda^{-2}(e^\lambda - \lambda - 1)}{x_0+1} \sum_{i=1}^{n} \frac{p_i^2}{q_i}$ is small; that is, all p_i are small. Considering the relative error between the Poisson binomial cumulative distribution function with parameter $\mathbf{p}$ and the Poisson cumulative distribution function with mean $\lambda = \sum_{i=1}^{n} \frac{p_i}{q_i}$, $\left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)}\right| = \left|\frac{\mathbb{PB}_\mathbf{p}(x_0) - \mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)}\right|$, we see that

$$|\mathbb{PB}_\mathbf{p}(x_0) - \mathbb{P}_\lambda(x_0)| \leq \left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)}\right| \tag{1.5}$$

and if $\left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)}\right| \to 0$, then $|\mathbb{PB}_\mathbf{p}(x_0) - \mathbb{P}_\lambda(x_0)| \to 0$. Thus $\mathbb{P}_\lambda(x_0)$ is a good approximation of $\mathbb{PB}_\mathbf{p}(x_0)$ when $\left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)}\right|$ tends to 0 for $x_0 \in \mathbb{N} \cup \{0\}$. Similar to that of (1.4), we must always consider the value of x_0, where, in some cases, consideration of the value of x_0 is not a convenience. In this study, we aim to determine the exact formula for $\sup_{x_0 \geq 0} \left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_\mathbf{p}(x_0)}\right|$ by using a mathematical inequality in Section 2.

2 Main result

Our aim is an exact formula for the supremum of relative error between the Poisson binomial and Poisson cumulative distribution functions using a mathematical inequality. Before giving the result, the following lemmas are needed.

Lemma 2.1. *Let m be any positive integer. For $i \in \{1, ..., n\}$, let $q_i \neq 0$. Then*

$$\frac{1}{m!}\left(\sum_{i=1}^{n} \frac{p_i}{q_i}\right)^m \geq \sum_{d_1+\cdots+d_n=m} \prod_{i=1}^{n} \left(\frac{p_i}{q_i}\right)^{d_i}, \tag{2.6}$$

where $d_1, ..., d_n \in \{0, 1\}$.

Proof. For $k_1, ..., k_n \in \{0, 1, .., m\}$, from (1.3) it follows that

$$\begin{aligned}\frac{1}{m!}\left(\sum_{i=1}^{n}\frac{p_i}{q_i}\right)^m &= \sum_{d_1+\cdots+d_n=m}\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{d_i} \\ &\quad+ \sum_{\substack{k_1+\cdots+k_n=m \\ (k_1,\ldots,k_n)\neq(d_1,\ldots,d_n)}}\frac{1}{k_1!\cdots k_n!}\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{k_i} \\ &\geq \sum_{d_1+\cdots+d_n=m}\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{d_i},\end{aligned}$$

□

Lemma 2.2. *Let* $x_0 \in \mathbb{N} \cup \{0\}$ *and* $\lambda = \sum_{i=1}^{n} \frac{p_i}{q_i}$. *Then*

$$0 \leq \mathbb{PB}_{\boldsymbol{p}}(x_0) - \mathbb{P}_\lambda(x_0). \tag{2.7}$$

Proof. For $x_0 \in \{0, 1, ..., n\}$, Teerapabolarn [10] showed that

$$0 \leq \mathbb{PB}_{\mathbf{p}}(x_0) - \mathbb{P}_\lambda(x_0), \tag{2.8}$$

and for $x_0 > n$, we see that

$$0 \leq 1 - \mathbb{P}_\lambda(x_0) = \mathbb{PB}_{\mathbf{p}}(x_0) - \mathbb{P}_\lambda(x_0). \tag{2.9}$$

Hence, from (2.8) and (2.9), the inequality (2.7) holds. □

The following theorem presents a result to approximate the Poisson binomial cumulative distribution function with parameter $\mathbf{p} = (p_1, ..., p_n)$ by a Poisson cumulative distribution function with mean $\lambda = \sum_{i=1}^{n} \frac{p_i}{q_i}$ in terms of the supremum relative error together with its exact formula.

Theorem 2.3. *Let* $\lambda = \sum_{i=1}^{n} \frac{p_i}{q_i}$. *Then*

$$\sup_{x_0\geq 0}\left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\boldsymbol{p}}(x_0)}\right| = 1 - \frac{e^{-\lambda}}{\prod_{i=1}^{n} q_i}. \tag{2.10}$$

Proof. For $x_0 = 0$ and 1, it is easy to see that $\frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\mathbf{p}}(x_0)} = \frac{e^{-\lambda}}{\prod_{i=1}^{n} q_i}$. For $x_0 \geq 2$, we have

$$\begin{aligned}
\frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\mathbf{p}}(x_0)} &= \left(\frac{e^{-\lambda}}{\prod_{i=1}^n q_i}\right) \frac{\sum_{j=0}^{x_0} \frac{\lambda^j}{j!}}{\displaystyle\sum_{j=0}^{x_0} \sum_{d_1+\cdots+d_n=j} \prod_{i=1}^{n} \left(\frac{p_i}{q_i}\right)^{d_i}} \\
&= \left(\frac{e^{-\lambda}}{\prod_{i=1}^n q_i}\right) \frac{\displaystyle\sum_{j=0}^{x_0} \frac{1}{j!}\left(\sum_{i=1}^{n}\frac{p_i}{q_i}\right)^j}{\displaystyle\sum_{j=0}^{x_0} \sum_{d_1+\cdots+d_n=j} \prod_{i=1}^{n} \left(\frac{p_i}{q_i}\right)^{d_i}} \\
&\geq \left(\frac{e^{-\lambda}}{\prod_{i=1}^n q_i}\right) \frac{\displaystyle\sum_{j=0}^{x_0} \sum_{d_1+\cdots+d_n=j} \prod_{i=1}^{n} \left(\frac{p_i}{q_i}\right)^{d_i}}{\displaystyle\sum_{j=0}^{x_0} \sum_{d_1+\cdots+d_n=j} \prod_{i=1}^{n} \left(\frac{p_i}{q_i}\right)^{d_i}} \quad \text{(by Lemma 2.1)} \\
&= \frac{e^{-\lambda}}{\prod_{i=1}^n q_i}. \qquad (2.11)
\end{aligned}$$

and so $\frac{\mathbb{P}_\lambda(0)}{\mathbb{PB}_{\mathbf{p}}(0)} = \frac{\mathbb{P}_\lambda(1)}{\mathbb{PB}_{\mathbf{p}}(1)} = \frac{e^{-\lambda}}{\prod_{i=1}^n q_i} \leq \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\mathbf{p}}(x_0)}$. Therefore, using Lemma 2.2, we have

$$0 \leq 1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\mathbf{p}}(x_0)} \leq 1 - \frac{\mathbb{P}_\lambda(1)}{\mathbb{PB}_{\mathbf{p}}(1)} = 1 - \frac{\mathbb{P}_\lambda(0)}{\mathbb{PB}_{\mathbf{p}}(0)} = 1 - \frac{e^{-\lambda}}{\prod_{i=1}^n q_i}$$

and

$$\left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\mathbf{p}}(x_0)}\right| \leq \left|1 - \frac{\mathbb{P}_\lambda(1)}{\mathbb{PB}_{\mathbf{p}}(1)}\right| = \left|1 - \frac{\mathbb{P}_\lambda(0)}{\mathbb{PB}_{\mathbf{p}}(0)}\right| = 1 - \frac{e^{-\lambda}}{\prod_{i=1}^n q_i}$$

for every $x_0 \geq 2$. Consequently,

$$\sup_{x_0 \geq 0} \left|1 - \frac{\mathbb{P}_\lambda(x_0)}{\mathbb{PB}_{\mathbf{p}}(x_0)}\right| = 1 - \frac{e^{-\lambda}}{\prod_{i=1}^n q_i}.$$

and the proof is now complete. □

As a consequence of Theorem 2.1, consider an additional form of the result in Theorem 2.3: $\left|1 - \frac{\mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)}\right| = \left|\frac{\mathbb{P}_\lambda(x_0) - \mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)}\right|$. Then

$$|\mathbb{PB}_{\mathbf{p}}(x_0) - \mathbb{P}_\lambda(x_0)| \leq \left|1 - \frac{\mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)}\right|. \qquad (2.12)$$

If $\left|1 - \frac{\mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)}\right| \to 0$, then $|\mathbb{PB}_{\mathbf{p}}(x_0) - \mathbb{P}_\lambda(x_0)| \to 0$. So $\mathbb{P}_\lambda(x_0)$ can also be used as a good estimate of $\mathbb{PB}_{\mathbf{p}}(x_0)$ when $\left|1 - \frac{\mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)}\right|$ is close to 0. Therefore, applying Theorem 2.3, we obtain the following corollary.

Corollary 2.4. *With the above definitions, we have*

$$\sup_{x_0 \geq 0} \left|1 - \frac{\mathbb{PB}_{\boldsymbol{p}}(x_0)}{\mathbb{P}_\lambda(x_0)}\right| = e^\lambda \prod_{i=1}^n q_i - 1. \tag{2.13}$$

Proof. From (2.11), it follows that

$$\frac{\mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)} \leq \frac{\mathbb{PB}_{\mathbf{p}}(1)}{\mathbb{P}_\lambda(1)} = \frac{\mathbb{PB}_{\mathbf{p}}(0)}{\mathbb{P}_\lambda(0)} = e^\lambda \prod_{i=1}^n q_i$$

and by Lemma 2.2

$$0 \leq \frac{\mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)} - 1 \leq \frac{\mathbb{PB}_{\mathbf{p}}(1)}{\mathbb{P}_\lambda(1)} - 1 = \frac{\mathbb{PB}_{\mathbf{p}}(0)}{\mathbb{P}_\lambda(0)} - 1 = e^\lambda \prod_{i=1}^n q_i - 1$$

for every $x_0 \geq 2$. This yields

$$\sup_{x_0 \geq 0} \left|1 - \frac{\mathbb{PB}_{\mathbf{p}}(x_0)}{\mathbb{P}_\lambda(x_0)}\right| = e^\lambda \prod_{i=1}^n q_i - 1.$$

Hence the result in (2.13) follows. □

If $p_1 = \cdots = p_n = p$, then the distribution of X is the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, and each result in Theorem 2.3 and Corollary 2.4 is a Poisson approximation to the binomial cumulative distribution function. The following corollary gives these results.

Corollary 2.5. *Let* $\mathbb{B}_{n,p}(x_0)$ *be the binomial cumulative distribution function with parameters* $n \in \mathbb{N}$ *and* $p = 1 - q \in (0, 1)$ *at* $x_0 \in \mathbb{N} \cup \{0\}$*. Then*

$$\sup_{x_0 \geq 0} \left|1 - \frac{\mathbb{P}_{\frac{np}{q}}(x_0)}{\mathbb{B}_{n,p}(x_0)}\right| = 1 - e^{-\frac{np}{q}} q^{-n} \tag{2.14}$$

and

$$\sup_{x_0 \geq 0} \left|1 - \frac{\mathbb{B}_{n,p}(x_0)}{\mathbb{P}_{\frac{np}{q}}(x_0)}\right| = e^{\frac{np}{q}} q^n - 1. \tag{2.15}$$

Considering the bounds in (2.14) and (2.15), observe that $e^{\frac{np}{q}}q^n - 1 > 1 - e^{-\frac{np}{q}}q^{-n}$. Therefore, the bounds approach 0 when $e^{\frac{np}{q}}q^n$ approaches 1 or p approaches 0; that is, the bounds are small when p is small.

3 Conclusion

In the present study, the exact formula for the supremum of relative error between the Poisson binomial cumulative distribution function with parameter $\mathbf{p} = (p_1, ..., p_n)$ and a Poisson cumulative distribution with mean $\lambda = \sum_{i=1}^{n} \frac{p_i}{q_i}$ was derived using a mathematical inequality. This formula is easy to compute for measuring the accuracy of the approximation and is appropriate to check how well the Poisson cumulative distribution function with this mean approximates the Poisson binomial cumulative distribution function with parameter $\mathbf{p}$. Additionally, when λ is relatively large and all p_i are sufficiently small, the Poisson cumulative distribution function with mean λ can also be used as a good estimate of the Poisson binomial cumulative distribution function.

References

[1] A. D. Barbour, P. Hall, On the rate of Poisson convergence, Mathematical Proceedings of the Cambridge Philosophical Society, **95**, no. 3, (1984), 473–480.

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

[3] L. H. Y. Chen, Poisson approximation for dependent trials, Annals of Probability, **3**, no. 3, (1975), 534-545.

[4] S. X. Chen, J. S. Liu, Statistical applications of the Poisson-binomial and condition Bernoulli distributions, Statistica Sinica, **7**, no. 4, (1997), 875–892.

[5] D. J. Daley, D. Vere-Jones, An introduction to the theory of point processes, **II.** General Theory and Structure, 2nd edition, Springer, New York, 2008.

[6] J. L. Hodges, L. Le Cam, The Poisson approximation to the Poisson binomial distribution, Annals of Mathematical Statistics, **31**, no. 3, (1960), 737–740.

[7] J. Kerstan, Verallgemeinerung eines Satzes von Prochorow und Le Cam, Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete, **2**, no. 2, (1964), 173–179.

[8] L. Le Cam, An approximation theorem for the Poisson binomial distribution, Pacific Journal of Mathematics, **10**, no. 4, (1960), 1181–1197.

[9] K. Neammanee, Pointwise approximation of Poisson binomial by Poisson distribution, Stochastic Modeling and Applications, **6**, no. 1, (2003), 20–26.

[10] K. Teerapabolarn, A bound on the Poisson binomial-Poisson relative error, International Journal of Pure and Applied Mathematics, **87**, no. 4, (2013), 535–540.

[11] K. Teerapabolarn, New non-uniform bounds on Poisson approximation for dependent Bernoulli trials, Bulletin of the Malaysian Mathematical Sciences Society, **38**, no. 1, (2005), 231–248.