Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 6, 7431-7439
https://doi.org/10.28919/jmcs/6626
ISSN: 1927-5307

BOUNDS ON POISSON APPROXIMATION FOR THE MULTICOLOUR MATCHING

CHANOKGAN SAHATSATHATSANA[1,*], SATTRA SAHATSATHATSANA[2], NIRUN NITISUK[1]

[1]Department of Science and Mathematics, Faculty of Science and Health Technology,
Kalasin University, Kalasin, Thailand
[2]Department of Foreign Language, Faculty of Liberal Arts,
Kalasin University, Kalasin, Thailand

Abstract. In this paper, we were, consequently, interested in the problem of the number of the multicolour matching and provided the bounds in Poisson approximation for the number of the multicolour matching by using Stein-chen method.

Keywords: Poisson approximation; multicolour matching; Stein-Chen method.
2010 AMS Subject Classification: 60G07.

1. Introduction

Regarding the theory of probability, the birthday problem or birthday paradox concerns with the combinatorial probability proposed by Chen (1975) and it's variant was presented by Jason [4], Holst [3], and Diaconis and Mosteller [2]. It refers to a set of n randomly chosen people which some pairs of them have the similarity of their births. For the conventional birthday problem, k balls are thrown independently into n boxes which any of them are equally and probably to fall into any of the n boxes.

*Corresponding author
E-mail address: Chanokgan.na@ksu.ac.th
Received August 11, 2021

We, consequently, are interested in the following problem: supposing that there are k balls, c colours and n boxes. The experiment will be done by throwing k balls with c colors in all n boxes. After that, randomly pick up r boxes. The Multicolour Matching will happen only if the r boxes that were picked up have the same number of balls and the same colours. Let us set up this problem in framework of Stein-chen method, let Ω be the space of possible arrangements of k, $k \geq 2$ balls with c colour in n boxes and $r \geq 2$ with probability P given by independent and uniform placement of the balls.The Multicolour Matching will happen only if $|\{(i,j\ldots,r)\} : 1 \leq i < j < \ldots r \leq k, b_i = b_j = \ldots = b_r$ with the same colour $|$, where b_i is the box in which put ball i. One can construct the random variable to obtain the problem by the following, let

$$W_n = \sum_{i}^{n} X_i, \tag{1.1}$$

the random variable W_n will denote the total number of multicolour matches, where $b_i \in \{b_i, b_j, \ldots, b_r\}$ and

$$X_i = \begin{cases} 1 & \text{if } b_i \text{ was similar in both numbers and colours,} \\ 0 & \text{otherwise.} \end{cases}$$

For the numbers of n which are getting larger, the poission distribution can be used to logically approximate the distribution of W_n with mean $\lambda = \mathrm{E}W_n = \dfrac{c^{2-k}(n-r)(c^{k-1} - n^{k-1})}{n^k(c-n)}$.

In this study, we derived bounds for the error on $|P(W_n \in A) - Poi_\lambda(A)|$, where W_n be the total number of multicolour matching. The tool for giving our main results consisted of the so-called poisson approximation and the Stein-chen method, which we mentioned them in Section 2. In Section 3, we use the Stein-chen method to derive the main results.

2. Method

Stein-Chen method [6],[1] was created to present the probabilities of rare event which can be approximated by Poission probabilities. The Normal distribution can often be used to approximate the common events. It introduces a new great technique to obtain the rate of convergence of the standard normal distribution. which gives,

Theorem 2.1. *If $W_n = \sum_{i=1}^{n} X_i$, $p_i = \mathrm{E}(X_i) = P(X_i = 1)$, $\lambda = \mathrm{E}(W_n)$ and $W_{n,i}$ be the random variable on same probability space as W_n, then*

$$|\mathrm{P}(W_n \in A) - \mathscr{P}oi_{\lambda}(A)| \leq \parallel g_{\lambda,A} \parallel \sum_{i=1}^{n} p_i \mathrm{E}|W_n - W_{n,i}| \tag{2.1}$$

where $\parallel g_{\lambda,A} \parallel := \sup_{w} [g_{\lambda,A}(w+1) - g_{\lambda,A}(w)]$.

Stein's equation for the poisson distribution is of the form

$$I_A(j) - Poi_{\lambda}(A) = \lambda g_{\lambda,A}(j+1) - j g_{\lambda,A}(j) \tag{2.2}$$

where $\mathscr{P}_{\lambda}(I_A) = e^{-\lambda} \sum_{w=0}^{\infty} I_A(w) \frac{\lambda^w}{w!}$, $\lambda > 0$, $j \in \mathbb{N} \cup \{0\}$, $A \subseteq \mathbb{N} \cup \{0\}$ and $I_A : \mathbb{N} \cup \{0\} \to \mathbb{R}$ be defined by

$$I_A(w) = \begin{cases} 1 & ; w \in A, \\ 0 & ; w \notin A. \end{cases}$$

The solution $g_{\lambda,A}$ of (2.2) is the form

$$g_{\lambda,A}(w) = \begin{cases} (w-1)!\lambda^{-w} e^{\lambda} [\mathscr{P}_{\lambda}(I_{A \cap C_{w-1}}) - \mathscr{P}_{\lambda}(I_A)\mathscr{P}_{\lambda}(I_{C_{w-1}})] & ; w \geq 1, \\ 0 & ; w = 0 \end{cases}$$

and

$$C_{w-1} = \{0, 1, \ldots, w-1\}.$$

For the substitution of j and λ in (2.2) by any integer-valued random variables W_n and $\lambda = \mathrm{E}W_n$, we have

$$\mathrm{P}(W_n \in A) - \mathscr{P}oi_{\lambda} A = \mathrm{E}(\lambda g_{\lambda,A}(W_n + 1)) - \mathrm{E}(W_n g_{\lambda,A}(W_n)). \tag{2.3}$$

There were many authors defined the bound of $\parallel g_{\lambda,A} \parallel$. For $A \subseteq \mathbb{N} \cup \{0\}$, Chen [1] prove that

$$\parallel g_{\lambda,A} \parallel \leq \min\{1, \lambda^{-1}\}$$

and Janson [5] presented that

$$\parallel g_{\lambda,A} \parallel \leq \lambda^{-1}(1 - e^{-\lambda}). \tag{2.4}$$

For the non-uniform bound, Neammanee [7] proposed that

$$\parallel g_{\lambda,A} \parallel \leq \min\left\{\frac{1}{w_0}, \lambda^{-1}\right\}$$

Teerapabolarn and Neammanee [9] revealed the bound of $\| g_{\lambda,A} \|$ when $A = \{0,1,\ldots,w_0\}$ through the following:

$$\| g_{\lambda,A} \| \leq \lambda^{-1}(1-e^{-\lambda})\min\left\{1, \frac{e^{\lambda}}{w_0+1}\right\}.$$

For any subset A of $\{0,1,\ldots,n\}$, Santiwipanont and Teerapabolarn [8] proposed the bound in the form of

$$\| g_{\lambda,A} \| \leq \lambda^{-1}\min\left\{1, \lambda, \frac{\triangle(\lambda)}{M_A+1}\right\} \tag{2.5}$$

where

$$\triangle(\lambda) = \begin{cases} e^{\lambda}+\lambda-1 & \text{if } \lambda^{-1}(e^{\lambda}-1) \leq M_A, \\ 2(e^{\lambda}-1) & \text{if } \lambda^{-1}(e^{\lambda}-1) > M_A, \end{cases}$$

and

$$M_A = \begin{cases} \max\{w \mid C_w \subseteq A\} & \text{if } 0 \in A, \\ \min\{w \mid w \in A\} & \text{if } 0 \notin A. \end{cases}$$

The problem of applying Theorem 2.1 is to find out $W_{n,i}$ which makes
$\mathrm{E}|W_n - W_{n,i}|$ small enough. So, there, generally, has not been solution. In case of $X_1,\ldots,X_n$ are considered to be independent,we let $W_{n,i} = W_n - X_i$. Then $\mathrm{E}|W_n - W_{n,i}| = p_i$ and from (2.1), we have

$$|\mathrm{P}(W_n \in A) - \mathscr{P}oi_{\lambda}(A)| \leq \| g_{\lambda,A} \| \sum_{i=1}^{n} p_i^2.$$

The problematic of constructing the $W_{n,i}$ is finding out the dependent indicator. Theorem 2.1 will be presented in the next part to prove our main result through the construction of the random variable $W_{n,i}$.

3. RESULT

The following theorem present bounds in poisson approximation for multicolour matching by using the Stein-Chen method.

Theorem 3.1. *Let W_n be the total number of the multicolour matching and $A \subseteq \mathbb{N} \cup \{0\}$. Then we have*

1. $|P(W_n \in A) - Poi_{\lambda}(A)| \leq C_{\lambda,n,r,A} \frac{(n-r)(n+c)^r}{n^r c^k} \left\{ \frac{c^2(n-1-r)(n^{k-1}-c^{k-1})}{n^{k-1}c^k(n-1)(n-c)} \right\}$

2. $|P(W_n \in A) - Poi_{\lambda}(A)| \leq (1-e^{-\lambda})\frac{(n-r)(n+c)^r}{n^r c^k}\left\{\frac{c^2(n-1-r)(n^{k-1}-c^{k-1})}{n^{k-1}c^k(n-1)(n-c)}\right\}$,

where $\lambda = \dfrac{c^{2-k}(n-r)(c^{k-1}-n^{k-1})}{n^k(c-n)}$, $\mathscr{P}_{\lambda}(I_A) = e^{-\lambda}\sum_{w=0}^{\infty} I_A(w)\frac{\lambda^w}{w!}$,

$$I_A(w) = \begin{cases} 1 & ; w \in A, \\ 0 & ; w \notin A. \end{cases}$$

$C_{\lambda,n,r,A} = \min\left\{1, \lambda, \frac{\triangle(\lambda)}{M_A+1}\right\}$,

$$\triangle(\lambda) = \begin{cases} e^{\lambda}+\lambda-1 & \text{if } \lambda^{-1}(e^{\lambda}-1) \leq M_A, \\ 2(e^{\lambda}-1) & \text{if } \lambda^{-1}(e^{\lambda}-1) > M_A, \end{cases}$$

and

$$M_A = \begin{cases} \max\{w \mid C_w \subseteq A\} & \text{if } 0 \in A, \\ \min\{w \mid w \in A\} & \text{if } 0 \notin A. \end{cases}$$

when $C_w = \{0,1,2,\ldots,w\}$

Proof. After the k balls with c colors are thrown into the n boxes; they are randomly picked up r, $r \geq 2$ boxes. Let $\{(i,j\ldots,r)\} : 1 \leq i < j < \ldots r \leq k, b_i = b_j = \ldots = b_r$ have similar colors, whereas b_i is the box for putting the ball i. For each $i,j \in \{1,2,\ldots,n\}$ such that $i \neq j$, we take those balls which have been thrown into the b_i box. We, then, remove the b_i box and throw the balls independently into other boxes. We define the indicator random variable X_{ij} as follow.

$$X_{ij} = \begin{cases} 1 & \text{if } b_j \in \{b_j,\ldots,b_r\} \text{ was similar in both numbers and colours,} \\ & \text{after removing } b_i \text{ box} \\ 0 & \text{otherwise.} \end{cases}$$

Let $W_{n,i} = \sum_{i=1,j\neq i}^{n} X_{ij}$ be the total number of multicolour matching after we removed the b_i box which similar in both number and colour .

Suppose that $\{j_s | s = 1,2,3,\ldots,w_0\}$ be the set of w_0 multicolour matching, $w_0 < k$, so for each $w_0 \in \{0,1,2,\ldots,k\}$, we got

$$\mathrm{P}(W_{n,i} = w_0) = \frac{\binom{n-1-w_0}{r}}{\binom{n}{r}}\{\sum_{r=2}^{k}\frac{c^r}{c^k n^r}\}^{w_0}$$

and

$$\begin{aligned}
\mathrm{P}(W_n - X_i = w_0 \mid X_i = 1) &= \frac{\mathrm{P}(W_n - X_i = w_0, X_i = 1)}{\mathrm{P}(X_i = 1)} \\
&= \frac{\mathrm{P}(W_n = w_0 + 1, X_i = 1)}{\mathrm{P}(X_i = 1)} \\
&= \frac{\frac{\binom{n-(w_0+1)}{r}}{\binom{n-1}{r}} \{\sum_{r=2}^{k} \frac{c^r}{c^k n^r}\}^{w_0+1}}{\frac{\binom{n-1}{r}}{\binom{n}{r}} \{\sum_{r=2}^{k} \frac{c^r}{c^k n^r}\}^{w_0}}
\end{aligned}$$

Then $W_{n,i}$ had the same distribution as $W_n - X_i$ conditional on $X_i = 1$. In order to bound $\mathrm{E}|W_n - W_{n,i}|$, we observed that

- In case $X_i = 1$, b_i is the set of multicolour matching boxes, so the number of boxes that are similar in both numbers and colours after removing the b_i box, equals to the number of multicolour matching minus 1, that was

$$W_{n,i} = W_n - 1. \tag{3.1}$$

- In case $X_i = 0$, we had the total number of multicolour matching boxes after we removing the b_i box and we throw the balls again as defined, equals to the total number of multicolour matching boxes minus the sum of number of the b_j box, where $j \neq i$, was similar in both numbers and colours at the first-throw and there are not similar in both numbers and colours after we throw them again, that was

$$W_{n,i} = W_n - \sum_{i,j=1, i\neq j}^{n} X_j Y_{ij}. \tag{3.2}$$

For each $j \in \{0, 1, 2, \ldots, n\}$, such that $j \neq i$, we defined the indicator random variable Y_{ij} as follow:

$$Y_{ij} = \begin{cases} 1 & \text{if } b_j \in \{b_j, \ldots, b_r\} \text{ was not similar in both numbers and colours,} \\ & \text{after we throw the balls again, in which these balls exactly} \\ & \text{used to land the } b_i \text{ box before} \\ 0 & \text{otherwise.} \end{cases}$$

We knew that

$$\mathrm{E} \mid W_n - W_{n,i} \mid = \mathrm{E}(W_n - W_{n,i})^+ + \mathrm{E}(W_n - W_{n,i})^-.$$

Where

$$(W_n - W_{n,i})^+ = \max\{W_n - W_{n,i}, 0\},$$

and

$$(W_n - W_{n,i})^- = -\min\{W_n - W_{n,i}, 0\}.$$

Form (3.1) and (3.2).

- In case $X_i = 1$, we had $(W_n - W_{n,i})^+ = 1$ and $(W_n - W_{n,i})^- = 0$

- In case $X_i = 0$, we had $(W_n - W_{n,i})^+ = \sum_{i,j=1,i\neq j}^{n} X_i Y_{ij}$ and $(W_n - W_{n,i})^- = 0$.

Therefore,
$(W_n - W_{n,i})^+ = \sum_{i,j=1,i\neq j}^{n} X_i Y_{ij}$ and $(W_n - W_{n,i})^- = 0$.

$$\begin{aligned}
\mathrm{E}(\mathrm{W_n} - \mathrm{W_{n,i}})^+ &\leq \mathrm{E}\{\sum_{\mathrm{i,j=1,i\neq j}}^{\mathrm{n}} \mathrm{X_j Y_{ij}}\} \\
&= \sum_{i,j=1,i\neq j}^{n} \mathrm{E}\{\mathrm{X_{ij} Y_{ij}}\} \\
&= \sum_{i,j=1,i\neq j}^{n} \mathrm{P}(\mathrm{X_{ij}} = 1, \mathrm{Y_{ij}} = 1) \\
&= \sum_{i,j=1,i\neq j}^{n} \mathrm{P}(\mathrm{X_{ij}} = 1)\mathrm{P}(\mathrm{Y_{ij}} = 1) \\
&= \sum_{i,j=1,i\neq j}^{n} \left\{ \frac{\binom{n-2}{r}}{\binom{n-1}{r}} \sum_{r=2}^{k} \frac{c^r}{c^k n^r} \right\} \left\{ 1 - \frac{\binom{n-1}{r}}{\binom{n}{r}} \sum_{r=2}^{k} \frac{c^r}{c^k n^r} \right\} \\
&= \sum_{i,j=1,i\neq j}^{n} \frac{c^2(n-1-r)(n^{k-1} - c^{k-1})}{(nc)^k (n-1)(n-c)} \left\{ 1 - \frac{c^2(n-r)(n^{k-1} - c^{k-1})}{c^k (n-c) n^{k+1}} \right\} \\
&\leq \frac{c^2(n-1-r)(n^{k-1} - c^{k-1})}{n^{k-1} c^k (n-1)(n-c)} \qquad (3.3)
\end{aligned}$$

Hence, by (2.1), (3.1), (2.5) and (3.3), we had

$$|P(W_n \in A) - Poi_\lambda(A)| \leq C_{\lambda,n,r,A} \left\{ \frac{c^2(n-1-r)(n^{k-1} - c^{k-1})}{n^{k-1} c^k (n-1)(n-c)} \right\}$$

and

$$|P(W_n \in A) - Poi_{\lambda}(A)| \leq (1-e^{-\lambda})\left\{\frac{c^2(n-1-r)(n^{k-1}-c^{k-1})}{n^{k-1}c^k(n-1)(n-c)}\right\}$$

where $\lambda = \dfrac{c^{2-k}(n-r)(c^{k-1}-n^{k-1})}{n^k(c-n)}$, and $C_{\lambda,n,r,A} = \min\left\{1, \lambda, \frac{\triangle(\lambda)}{M_A+1}\right\}$. □

Remark 3.2. We see that the result in Theorem 3.1, for n beginning sufficiently large and $A \subseteq \mathbb{N} \cup \{0\}$, we can approximate the cumulative probability of the number of the multicolour matching, $P(W_n \in A)$, by the cumulative Poisson probability,
$Poi_{\lambda}(A)$ with $\lambda = \dfrac{c^{2-k}(n-r)(c^{k-1}-n^{k-1})}{n^k(c-n)}$., i.e.,

$$P(W_n \in A) \approx Poi_{\lambda}(A), \text{ when } n \to \infty \tag{3.4}$$

ACKNOWLEDGEMENT

We would like to sincerely thanks all reviewers for the valuable and creative comments for improving the quality of the paper. Our appreciation also goes to Mr. Jonathan Wary for his assistance in editing the language usage in the paper.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] L.H.Y. Chen, Poisson approximation for dependent trials, Ann. Probab. 3 (1975), 534–545.

[2] P. Diaconis, and F. Mosteller, Methods for studying coincidences, J. Amer. Stat. Assoc. 84 (1989), 853–861.

[3] L. Holst, On birthday, collectors', occupancy and other classical urn problems, Int. Stat. Rev. 54 (1986), 15–27.

[4] S. Janson, Birthday problems, randomly colored graphs, and Poisson limits of dissociated variables, Univ., Dept. of Math 16, (1986).

[5] S. Janson, Coupling and Poisson approximation, Acta Appl. Math. 34 (1994), 7–15.

[6] C.M. Stein, A bound for the error in normal approximation to distribution of a sum of dependent random variables, Proc. Sixth Berkeley Symapos, Math. Statist. Probab. 3 (1972), 583–602.

[7] K. Neammanee, Pointwise approximation of Poisson binomial by Poisson distribution, Stoch. Model. Appl. 6 (2003), 20–26.

[8] T. Santiwipanont and K. Teerabolarn, Two formulas of nonuniform bounds on dependent indicators, Thai J. Math. 5 (2006), 5–39.

[9] K. Teerapabolarn and K. Neammanee, Poisson Approximation for Sums of dependent Bernulli Random, Acta Math. 22 (2006), 87–89.