KYUNGPOOK Math. J. 46(2006), 489-496

A Non-uniform Bound on Matching Problem

Kanint Teerapabolarn
Department of Mathematics, Faculty of Science, Burapha University, Chonburi 20131, Thailand
e-mail : kanint@buu.ac.th

Kritsana Neammanee
Department of Mathematics, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand
e-mail : Kritsana.N@chula.ac.th

Abstract. The aim of this paper is to use the Stein-Chen method to obtain a non-uniform bound on Poisson approximation in matching problem.

1. Introduction and main result

Suppose that n cards, numbered $1, 2, \cdots, n$, are placed at random onto n places on a table where the places are numbered $1, 2, \cdots, n$. Each place is to hold one and only one card. We say that a match (or coincidence or rencontre) occurs at the ith place if the card numbered i is placed there. For each $i \in \{1, \cdots, n\}$, let

$$X_i = \begin{cases} 1 & \text{if the card numbered } i \text{ is at the } i\text{th place,} \\ 0 & \text{otherwise.} \end{cases}$$

The probability that $X_i = 1$ is given by

$$P(X_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}. \tag{1.1}$$

Let $W_n = \sum_{i=1}^{n} X_i$ be the total number of matches. When n is sufficiently large, it is logical to approximate the distribution of W_n by Poisson distribution with mean $\lambda = E[W_n] = 1$.

In 1992, Barbour, Holst and Janson [1] gave a uniform bound for approximating the distribution of W_n by Poisson distribution with parameter 1 in the form of

$$\left| P(W_n \leq w_0) - \frac{1}{e} \sum_{k=0}^{w_0} \frac{1}{k!} \right| \leq \frac{2(1-e^{-1})}{n}, \tag{1.2}$$

Received April 21, 2005, and, in revised form, August 28, 2006.
2000 Mathematics Subject Classification: 60F05, 60G05.
Key words and phrases: Poisson distribution, non-uniform bound, matching problem, Stein-Chen method.

where $w_0 \in \{0, 1, \cdots, n\}$.

In this paper, we give a non-uniform bound for this approximation and the following theorem is our main result.

Theorem 1.1. *Let W_n be defined as above. Then we have*

$$\left| P(W_n \leq w_0) - \frac{1}{e}\sum_{k=0}^{w_0}\frac{1}{k!} \right| \leq \triangle(n, w_0), \tag{1.3}$$

where

$$\triangle(n, w_0) = \begin{cases} \frac{2}{en} & \text{if } \; w_0 = 0, \\ \frac{2(1-2e^{-1})}{n} & \text{if } \; w_0 = 1, \\ \frac{2.08}{(w_0+1)n} & \text{if } \; w_0 = 2, 3, \cdots, n. \end{cases} \tag{1.4}$$

We see that the result in (1.3) improves that in 1.2) and the bound tends to 0 when n is large. Hence, for all $w_0 \in \{0, 1, \cdots, n\}$, we can approximate the cumulative probability of the number of matches, $P(W_n \leq w_0)$, by the cumulative Poisson probability, $\frac{1}{e}\sum_{k=0}^{w_0}\frac{1}{k!}$, i.e.,

$$P(W_n \leq w_0) \approx \frac{1}{e}\sum_{k=0}^{w_0}\frac{1}{k!}, \quad \text{as} \quad n \to \infty.$$

Example 1.1. In Table 1.1, we give an example of Poisson approximation in matching problem according to formulas (1.2) and (1.3) in the case when $n = 100$.

Table 1.1 Poisson Estimate of $P(W_n \leq w_0)$ for $n = 100$

w_0	Estimate $\frac{1}{e}\sum_{k=0}^{w_0}\frac{1}{k!}$	Uniform Bound (1.2)	Non-Uniform Bound (1.3)
0	0.36787944	0.01264241	0.00735759
1	0.73575944	0.01264241	0.00528482
2	0.91969860	0.01264241	0.00693333
3	0.98101184	0.01264241	0.00520000
4	0.99634015	0.01264241	0.00416000
5	0.99940582	0.01264241	0.00346667
6	0.99991676	0.01264241	0.00297143
7	0.99998975	0.01264241	0.00260000
8	0.99999887	0.01264241	0.00231111
9	0.99999989	0.01264241	0.00208000
10	0.99999999	0.01264241	0.00189091
11	1.00000000	0.01264241	0.00173333

Example 1.2. Suppose that a secretary drops 500 matching pairs of letters and envelopes down the stairs, and then randomly each places letter into one of the empty the envelopes. What is the probability of having at least one correctly matched pair?

We have to find $1 - P(W_n = 0)$, where W_n is the number of correctly matched pairs. Since it is difficult to find the exact result, the probability approximation should be used in this case. For $n = 500$, by Theorem 1.1, a bound for this approximation is

$$\left|P(W_n = 0) - \frac{1}{e}\right| \leq 0.001471518.$$

Hence the probability that there is at least one correctly matched pair is

$$0.630649041 \leq 1 - P(W_n = 0) \leq 0.633592077.$$

2. Proof of main result

In 1972, Stein [3] introduced a powerful and general method for obtaining an explicit bound for the error in the normal approximation for dependent random variables. This method was adapted and applied to the Poisson approximation by Chen [2]. It is usually referred to as the Stein-Chen or Chen-Stein method. This method started by Stein's equation for Poisson distribution with parameter $\lambda = 1$ which is defined by

$$f(w+1) + wf(w) = h(w) - \mathcal{P}(h), \tag{2.1}$$

where $\mathcal{P}(h) = \frac{1}{e}\sum_{l=0}^{\infty} h(l)\frac{1}{l!}$ and f and h are real valued 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

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

From [1] p.7, we know that the solution Uh_A of (2.1) is of the form

$$Uh_A(w) = \begin{cases} e(w-1)![\mathcal{P}(h_{A\cap C_{w-1}}) - \mathcal{P}(h_A)\mathcal{P}(h_{C_{w-1}})] & \text{if } \ w \geq 1, \\ 0 & \text{if } \ w = 0, \end{cases} \tag{2.3}$$

where $C_w = \{0, \cdots, w\}$.

Hence, by (2.3), we have

$$Uh_{C_{w_0}}(w) = \begin{cases} e(w-1)![\mathcal{P}(h_{C_{w_0}})\mathcal{P}(1 - h_{C_{w-1}})] & \text{if } \ w > w_0, \\ e(w-1)![\mathcal{P}(h_{C_{w-1}})\mathcal{P}(1 - h_{C_{w_0}})] & \text{if } \ w \leq w_0, \\ 0 & \text{if } \ w = 0. \end{cases} \tag{2.4}$$

In the proof of main result, we need the following lemma.

Lemma 2.1. *Let* $w_0 \in \{0, 1, 2, \cdots, n\}$.

(1) *For any* $s, t \in \mathbb{N}$,

$$|Vh_{C_{w_0}}(t,s)| \le \sup_{w\ge 1} |Vh_{C_{w_0}}(w+1,w)||t-s|,$$

where $Vh_{C_{w_0}}(t,s) = Uh_{C_{w_0}}(t) - Uh_{C_{w_0}}(s)$.

(2) *For* $w \ge 1$,

$$|Vh_{C_{w_0}}(w+1,w)| \le \frac{n\triangle(n,w_0)}{2}, \tag{2.5}$$

where $\triangle(n, w_0)$ *is defined in* (1.4).

Proof. (1). See [4], p.90.

(2) From (2.4) we note that

$$\begin{aligned} & Vh_{C_{w_0}}(w+1,w) \\ = & \begin{cases} e(w-1)!\mathcal{P}(h_{C_{w_0}})[w\mathcal{P}(1-h_{C_w}) - \mathcal{P}(1-h_{C_{w-1}})] & \text{if } w \ge w_0+1, \\ e(w-1)!\mathcal{P}(1-h_{C_{w_0}})[w\mathcal{P}(h_{C_w}) - \mathcal{P}(h_{C_{w-1}})] & \text{if } w \le w_0. \end{cases} \end{aligned} \tag{2.6}$$

Case 1. $w_0 = 0$.

By the fact that

$$\begin{aligned} Vh_{C_0}(w+1,w) &= (w-1)![w\mathcal{P}(1-h_{C_w}) - \mathcal{P}(1-h_{C_{w-1}})] \\ &= \frac{(w-1)!}{e}\left\{ w\sum_{k=w+1}^{\infty}\frac{1}{k!} - \sum_{k=w}^{\infty}\frac{1}{k!}\right\} \\ &= \frac{(w-1)!}{e}\sum_{k=w+1}^{\infty}\frac{w-k}{k!} \\ &< 0 \end{aligned}$$

and

$$\begin{aligned} 0 < -Vh_{C_0}(w+1,w) &= \frac{(w-1)!}{e}\sum_{k=w+1}^{\infty}\frac{k-w}{k!} \\ &= \frac{(w-1)!}{e}\left\{\frac{1}{(w+1)!} + \frac{2}{(w+2)!} + \cdots\right\} \\ &= \frac{(w-1)!}{ew!}\left\{\frac{1}{w+1} + \frac{2}{(w+1)(w+2)} + \cdots\right\} \\ &\le \frac{1}{e}\left\{\frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \cdots\right\} \\ &= \frac{1}{e}\left\{\left[1 + \frac{1}{2!} + \frac{1}{3!}\cdots\right] - \left[\frac{1}{2!} + \frac{1}{3!} + \cdots\right]\right\} = \frac{1}{e}, \end{aligned}$$

we have

$$|Vh_{C_0}(w+1,w)| \le \frac{1}{e}.$$

Case 2. $w_0 = 1$.
From (2.6), we have $Vh_{C_1}(2,1) = 1 - \frac{2}{e}$ and, for $w \ge 2$,

$$\begin{aligned} Vh_{C_1}(w+1,w) &= 2(w-1)!e^{-1}\sum_{k=w+1}^{\infty}\frac{w-k}{k!} \\ &< 0 \end{aligned}$$

and

$$\begin{aligned} 0 < -Vh_{C_1}(w+1,w) &= \frac{2}{e}(w-1)!\left\{\frac{1}{(w+1)!}+\frac{2}{(w+2)!}+\frac{3}{(w+3)!}+\cdots\right\} \\ &= \frac{2}{e}\frac{(w-1)!}{(w+1)!}\left\{1+\frac{2}{w+2}+\frac{3}{(w+2)(w+3)}+\cdots\right\} \\ &\le \frac{1}{3e}\left\{1+\frac{2}{4}+\frac{3}{4^2}+\frac{4}{4^3}+\cdots\right\} \\ &= \frac{16}{27e}. \end{aligned}$$

Hence,

$$|Vh_{C_1}(w+1,w)| \le \max\{1-\frac{2}{e},\frac{16}{27e}\} = 1-\frac{2}{e}.$$

Case 3. $w_0 \ge 2$.
For $w \ge w_0+1$, we see that

$$\begin{aligned} 0 < -Vh_{C_{w_0}}(w+1,w) &\le (w-1)!\sum_{k=w+1}^{\infty}\frac{k-w}{k!} \\ &= (w-1)!\left\{\frac{1}{(w+1)!}+\frac{2}{(w+2)!}+\frac{3}{(w+3)!}+\cdots\right\} \\ &= \frac{(w-1)!}{(w+1)!}\left\{1+\frac{2}{w+2}+\frac{3}{(w+2)(w+3)}+\cdots\right\} \\ &\le \frac{1}{(w_0+1)(w_0+2)}\left\{1+\frac{2}{5}+\frac{3}{5^2}+\frac{4}{5^3}+\cdots\right\} \\ &= \frac{25}{16(w_0+1)(w_0+2)} \end{aligned}$$

and, for $w \le w_0$, we have

$$
\begin{aligned}
0 &< Vh_{C_{w_0}}(w+1,w)\\
&= (w-1)!\mathcal{P}(1-h_{C_{w_0}})\left\{w\sum_{k=0}^{w}\frac{1}{k!}-\sum_{k=0}^{w-1}\frac{1}{k!}\right\}\\
&\le (w_0-1)!\mathcal{P}(1-h_{C_{w_0}})\left\{(w_0-1)\sum_{k=0}^{w_0}\frac{1}{k!}+\frac{1}{w_0!}\right\}\\
&\le \left\{\frac{1}{w_0}+e(w_0-1)(w_0-1)!\right\}\mathcal{P}(1-h_{C_{w_0}})\\
&= \frac{e^{-1}+(w_0-1)w_0!}{w_0}\left\{\frac{1}{(w_0+1)!}+\frac{1}{(w_0+2)!}+\frac{1}{(w_0+3)!}+\cdots\right\}\\
&= \frac{e^{-1}+(w_0-1)w_0!}{w_0(w_0+1)!}\left\{1+\frac{1}{w_0+2}+\frac{1}{(w_0+2)(w_0+3)}+\cdots\right\}\\
&\le \frac{e^{-1}+(w_0-1)w_0!}{w_0(w_0+1)!}\left\{1+\frac{1}{w_0+2}+\frac{1}{(w_0+2)^2}+\cdots\right\}\\
&= \frac{\{e^{-1}+(w_0-1)w_0!\}(w_0+2)}{w_0(w_0+1)(w_0+1)!}.
\end{aligned}
$$

So,

$$
\begin{aligned}
|Vh_{C_{w_0}}(w+1,w)| &\le \max\{\frac{25}{16(w_0+1)(w_0+2)},\frac{\{e^{-1}+(w_0-1)w_0!\}(w_0+2)}{w_0(w_0+1)(w_0+1)!}\}\\
&= \frac{\{e^{-1}+(w_0-1)w_0!\}(w_0+2)}{w_0(w_0+1)(w_0+1)!}\\
&\le \frac{1.04}{w_0+1}.
\end{aligned}
$$

Hence, from case 1 to case 3, we have (2.5). □

Proof of Theorem 1.1. From (2.1), when $h=h_{C_{w_0}}$, we have

$$
P(W_n\le w_0)-\frac{1}{e}\sum_{k=0}^{w_0}\frac{1}{k!}=E[f(W_n+1)-W_nf(W_n)], \tag{2.7}
$$

where $f=Uh_{C_{w_0}}$ is defined by (2.4).
By the same argument in the proof of Theorem 1.2 of [4], we have

$$
E[W_nf(W_n)]=\sum_{i=1}^{n}E[X_if(W_n)]
$$

and for each i,

$$\begin{aligned} E[X_i f(W_n)] &= E[X_i f(W_n)|X_i = 0]P(X_i = 0) + E[X_i f(W_n)|X_i = 1]P(X_i = 1) \\ &= E[f(W_n)|X_i = 1]P(X_i = 1) \text{ ;which by (1.1)} \\ &= \frac{1}{n} E[f(W^*_{n,i} + 1)], \end{aligned}$$

where $W^*_{n,i} \sim (W_n - X_i)|X_i = 1$, i.e., $W^*_{n,i}$ has the same distribution as $W_n - X_i$ conditional on $X_i = 1$. Thus

$$\begin{aligned} E[f(W_n + 1) - W_n f(W_n)] &= \sum_{i=1}^{n} \frac{1}{n} E[f(W_n + 1)] - \sum_{i=1}^{n} \frac{1}{n} E[f(W^*_{n,i} + 1)] \\ &= \frac{1}{n} \sum_{i=1}^{n} E[f(W_n + 1) - f(W^*_{n,i} + 1)]. \end{aligned}$$

By Lemma 2.1(1 and 2), we have

$$\begin{aligned} |E[f(W_n + 1) - W_n f(W_n)]| &\leq \frac{1}{n} \sum_{i=1}^{n} E|f(W_n + 1) - f(W^*_{n,i} + 1)| \\ &\leq \frac{\triangle(n, w_0)}{2} \sum_{i=1}^{n} E|W_n - W^*_{n,i}|. \end{aligned} \tag{2.8}$$

In order to bound $E|W_n - W^*_{n,i}|$, we observe that $W^*_{n,i}$ has the same distribution as W_{n-1}. From this fact we can see that $E[W^*_{n,i}] = E[W_{n-1}] = 1$ and $W^*_{n,i} = W_n - 1$ in case that $X_i = 1$ and $W_n \leq W^*_{n,i}$ if $X_i = 0$. These imply that

$$|W^*_{n,i} - W_n + X_i| = W^*_{n,i} - W_n + X_i$$

and

$$\begin{aligned} E|W_n - W^*_{n,i}| &\leq E|W^*_{n,i} - W_n + X_i| + E[X_i] \\ &= E|W^*_{n,i} - W_n + X_i| + E[X_i] \\ &= 2E[X_i] - E[W_n] + E[W^*_{n,i}] \\ &= \frac{2}{n} - 1 + 1 \\ &= \frac{2}{n}. \end{aligned} \tag{2.9}$$

From this fact and (2.8), we have

$$|E[f(W_n + 1) - W_n f(W_n)]| \leq \triangle(n, w_0).$$

Then the theorem follows from this fact and (2.7). □

Remark. If we apply (2.9) to Theorem 1.2 of [4], then we have

$$\left|P(W_n \le w_0) - \frac{1}{e}\sum_{k=0}^{w_0} \frac{1}{k!}\right| \le \frac{2(e-1)}{(w_0+1)n} \quad \text{for } w_0 = 2, 3, \cdots, n. \tag{2.10}$$

We can observe that, our constant in (1.4) is sharper than the constant $2(1-e)$ in (2.10).

Acknowledgements. The authors would like to thank the referees for their insightful comments.

References

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

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

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

[4] K. Teerapabolarn and K. Neammanee, *Poisson Approximation for Sums of Dependent Bernoulli Random Variables*, Acta Math., **22**(2006), 87-99.