# A REFINEMENT OF NORMAL APPROXIMATION TO POISSON BINOMIAL

K. NEAMMANEE

*Received 8 August 2004 and in revised form 6 December 2004*

Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with $P(X_j = 1) = 1 - P(X_j = 0) = p_j$ and let $S_n := X_1 + X_2 + \cdots + X_n$. $S_n$ is called a Poisson binomial random variable and it is well known that the distribution of a Poisson binomial random variable can be approximated by the standard normal distribution. In this paper, we use Taylor's formula to improve the approximation by adding some correction terms. Our result is better than before and is of order $1/n$ in the case $p_1 = p_2 = \cdots = p_n$.

## 1. Introduction and main result

Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with $P(X_j = 1) = p_j$ and $P(X_j = 0) = q_j$, where $0 < p_j < 1$ and $p_j + q_j = 1$ for $j = 1, 2, \ldots, n$. Let $S_n := X_1 + X_2 + \cdots + X_n$, $\mu := ES_n = p_1 + p_2 + \cdots + p_n$, and $\sigma^2 := \operatorname{Var} S_n = p_1 q_1 + p_2 q_2 + \cdots + p_n q_n$. In connection with Bernoulli's theorem, the following important question arises: when the number of trials is large, how can one find, at least approximately, the probability

$$P(a \le S_n \le b), \tag{1.1}$$

where $a, b = 1, 2, \ldots, n$?

De Moivre [3] was the first one who successfully attacked this difficult problem in case of $p_1 = p_2 = \cdots = p_n$ by using the standard normal distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(1/2)t^2} dt. \tag{1.2}$$

After him, in essentially the same way, but using more powerful analytical tools, Laplace [7] succeeded in establishing a simple approximation formula which is given in all books on probability. A general bound was given by Feller [5] for independent and nonidentically distributed random variables with finite third moments and by Chen and Shao [2] without assuming the existence of third moments. In the case when the random variables are identically distributed, it has long been known that the best bound is of order $1/\sqrt{n}$.


International Journal of Mathematics and Mathematical Sciences 2005:5 (2005) 717–728
DOI: 10.1155/IJMMS.2005.717

In this paper, we investigate the approximation of $S_n$ by its asymptotic expansions. If two terms,

$$G(x) := \Phi(x) + \frac{1}{6\sqrt{2\pi}\sigma^3} \sum_{j=1}^{n} p_j q_j (p_j - q_j)(1 - x^2) e^{-x^2/2}, \tag{1.3}$$

are used, then the accuracy of the approximation is better. The use of asymptotic expansions is one of the most natural ways of refinement (see, e.g., Uspensky [12], Kolassa [6], Petrov [11], and Bhattacharya and Rao [1]). The refinement of the central limit theorem for sums of independent Bernoulli random variables has a long history.

In what follows, let

$$\Delta_n := \left| P(a \le S_n \le b) - (G(x_2) - G(x_1)) \right|, \tag{1.4}$$

where

$$x_1 = \frac{1}{\sigma}\left(a - \mu - \frac{1}{2}\right), \qquad x_2 = \frac{1}{\sigma}\left(b - \mu + \frac{1}{2}\right). \tag{1.5}$$

In the case when $S_n$ is a binomial random variable, that is, $p_1 = p_2 = \cdots = p_n$, Uspensky [12] shows that

$$\Delta_n \le \frac{0.26 + 0.36\,|q - p|}{\sigma^2} + 2e^{-(3/2)\sigma} \tag{1.6}$$

under the condition that

$$\sigma^2 \ge 25 \tag{1.7}$$

and in 1955, Makabe [8] improved the result of Uspensky in the form of

$$\Delta_n \le \frac{0.106 + 0.054(q - p) + 0.108(q - p)^2}{\sigma^2} + 2e^{-(3/2)\sigma} \tag{1.8}$$

under the conditions that

$$p < \frac{1}{2}, \quad \sigma^2 \ge 25, \quad n \ge 100. \tag{1.9}$$

In this paper, we consider the correction terms in the case when $p_i$'s are not necessarily equal, that is, $S_n$ is a Poisson binomial random variable. In this case, Makabe [9] shows that

$$\Delta_n \le \frac{C}{\sigma^2} \tag{1.10}$$

for some constant $C > 0$ under the conditions that

$$\sigma^2 \ge 25, \qquad p_i < \frac{1}{2} \quad \text{for } i = 1, 2, \ldots, n. \tag{1.11}$$

Mikhaĭlov [10] calculated the constant $C$ of Makabe and found out that

$$\Delta_n \leq \frac{2(\sigma+3)}{\sigma^3} \tag{1.12}$$

under the condition that

$$\sigma^2 \geq 100. \tag{1.13}$$

In 1995, Volkova showed that

$$\begin{aligned}\Delta_n \leq{}& \frac{2}{\sigma^2}\left(0.05\beta_4+0.1\beta_3^2+0.08\right)+\frac{2}{\sigma^3}\left(0.05\beta_3+0.17\beta_3\beta_4+0.056\right)\\ &+\frac{2}{\sigma^4}\left(0.06\beta_4^2+0.27\beta_4+0.002\right),\end{aligned} \tag{1.14}$$

where $\beta_3$ and $\beta_4$ are the third and fourth semi-invariants of $S_n$, respectively. The bound of Volkova is valid under the condition (1.13).

In the present work, we improve the bounds in (1.6), (1.8), (1.10), (1.12), and (1.14) under the condition (1.13). Here is our main results.

Theorem 1.1. *For $\sigma^2 \geq 100$,*

$$\Delta_n \leq \frac{0.1618}{\sigma^2}. \tag{1.15}$$

*In the proof of the main theorem, we use the idea of Uspensky which uses only Taylor's formula without using any high-power analytical tools.*

*Remarks 1.2.* (1) In Theorem 1.1, we estimate $\Delta_n$ in case of $\sigma^2 \geq 100$. In fact, the bound is valid in the range $0 < \sigma^2 < 100$ as well. For example, if $\sigma^2 \in [25,100)$, by using the argument of Theorem 1.1, one can get the bound of the form

$$\Delta_n \leq \frac{0.3056}{\sigma^2}. \tag{1.16}$$

In this case, Volkova [14] showed that

$$\Delta_n \leq \frac{2(2\sigma+5)}{5\sigma^3}, \tag{1.17}$$

which is larger than our result.

(2) The bound in Theorem 1.1 is correct in order (see Deheuvels et al. [4]) and in the case $p_1 = p_2 = \cdots = p_n$, the order of the bound is $1/n$.

## 2. Proof of main result

Let $\varphi_1, \varphi_2, \ldots, \varphi_n$ and $\varphi$ be the characteristic functions of $X_1, X_2, \ldots, X_n$ and $S_n$, respectively. Hence

$$\varphi_j(t) = q_j + p_j e^{it} \quad \text{for } j = 1,2,\ldots,n, \qquad \varphi(t) = \prod_{j=1}^{n}\left(q_j + p_j e^{it}\right), \tag{2.1}$$

where $i = \sqrt{-1}$. We note that the complex number $\varphi(t)$ can be represented in the form

$$\varphi_j(t) = \rho_j(t)e^{i\Theta_j(t)}, \tag{2.2}$$

where

$$\rho_j(t) := |\varphi_j(t)| = \left(p_j^2 + q_j^2 + 2p_jq_j\cos t\right)^{1/2} = \left(1 - 4p_jq_j\sin^2\frac{t}{2}\right)^{1/2}, \tag{2.3}$$

$$\Theta_j(t) := \text{argument of } \varphi_j(t) = \arctan\left(\frac{p_j\sin t}{q_j + p_j\cos t}\right). \tag{2.4}$$

Hence

$$\varphi(t) = \rho(t)e^{i\Theta(t)}, \tag{2.5}$$

where $\rho(t) = \prod_{j=1}^{n}\rho_j(t)$ and $\Theta(t) = \sum_{j=1}^{n}\Theta_j(t)(\operatorname{mod} 2\pi)$.

Let

$$\begin{gathered}\alpha(t) := \Theta(t) - \mu t,\\ R(x) := \frac{1}{2\pi}\int_0^{\pi}\frac{\rho(t)\sin\left(\sigma xt - \alpha(t)\right)}{\sin(t/2)}dt.\end{gathered} \tag{2.6}$$

From Uspensky [12], we know that

$$P(a \le S_n \le b) = R(x_2) - R(x_1), \tag{2.7}$$

where $x_1$ and $x_2$ are defined in (1.5).

Lemma 2.1. *For $j = 1,2,\dots,n$,*

(1) $\rho_j(t) \le e^{-(2/\pi^2)p_jq_jt^2}$ *for* $t \in [0,\pi)$,

(2) $\rho_j(t) \le e^{-(1/2)p_jq_jt^2+(1/24)p_jq_jt^4}$ *for* $t \in [0,\pi]$,

(3) $\rho_j(t) \ge e^{-(1/2)p_jq_jt^2-(1/4)p_j^2q_j^2t^4}$ *for* $t \in [0,\pi/2]$.

*Proof.* (1) By (2.3) and the fact that $|4p_jq_j\sin^2(t/2)| < 1$, we have

$$\ln\rho_j(t) = \frac{1}{2}\ln\left(1 - 4p_jq_j\sin^2\frac{t}{2}\right) \tag{2.8}$$

$$= -\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k}\left(4p_jq_j\sin^2\frac{t}{2}\right)^k \tag{2.9}$$

$$\le -2p_jq_j\sin^2\frac{t}{2} \tag{2.10}$$

$$\le -\frac{2}{\pi^2}p_jq_jt^2, \tag{2.11}$$

where we have used the fact that

$$\sin\frac{t}{2} \ge \frac{t}{\pi} \quad \text{on } [0,\pi) \tag{2.12}$$

in the last inequality. Hence we have (1).

(2) By Taylor's formula, we know that $\sin^2(t/2) \geq t^2/4 - t^4/48$. So (2) follows by this fact and (2.10).

(3) Let $t \in [0, \pi/2]$. Using Taylor's formula for $f(t) = (1/3)\sin^4(t/2) + \sin^2(t/2)$, we can show that

$$\frac{t^2}{4} - \sin^2\frac{t}{2} \geq \frac{1}{3}\sin^4\frac{t}{2}. \tag{2.13}$$

Hence, by (2.9), (2.13) and the fact that $0 \leq \sin(t/2) \leq t/2$, we have

$$\begin{aligned}
\ln\rho_j(t) &= \left\{-\frac{1}{2}p_jq_jt^2 - \frac{1}{4}(4p_jq_j)^2\sin^4\frac{t}{2}\right\} + 2p_jq_j\left\{\frac{t^2}{4} - \sin^2\frac{t}{2}\right\} \\
&\quad - \frac{1}{2}\sum_{k=3}^{\infty}\frac{1}{k}\left(4p_jq_j\sin^2\frac{t}{2}\right)^k \\
&\geq \left\{-\frac{1}{2}p_jq_jt^2 - \frac{1}{4}p_j^2q_j^2t^4\right\} + \frac{2}{3}p_jq_j\sin^4\frac{t}{2} - \frac{1}{6}\sum_{k=3}^{\infty}\left(4p_jq_j\sin^2\frac{t}{2}\right)^k \\
&= \left\{-\frac{1}{2}p_jq_jt^2 - \frac{1}{4}p_j^2q_j^2t^4\right\} + \frac{2}{3}p_jq_j\sin^4\frac{t}{2} - \frac{(1/6)(4p_jq_j)^3\sin^6(t/2)}{1 - 4p_jq_j\sin^2(t/2)} \\
&\geq \left\{-\frac{1}{2}p_jq_jt^2 - \frac{1}{4}p_j^2q_j^2t^4\right\} + \frac{2}{3}p_jq_j\sin^4\frac{t}{2} - \frac{1}{3}(4p_jq_j)^3\sin^6\frac{t}{2} \\
&= \left\{-\frac{1}{2}p_jq_jt^2 - \frac{1}{4}p_j^2q_j^2t^4\right\} + \frac{2}{3}p_jq_j\sin^4\frac{t}{2}\left(1 - 32p_j^2q_j^2\sin^2\frac{t}{2}\right) \\
&\geq -\frac{1}{2}p_jq_jt^2 - \frac{1}{4}p_j^2q_j^2t^4
\end{aligned} \tag{2.14}$$

which implies that $\rho_j(t) \geq e^{-(1/2)p_jq_jt^2 - (1/4)p_j^2q_j^2t^4}$. □

We are now ready to prove the main result of this section. For convenience, we assume $\sigma^2 \geq 100$ and divide the proof into 5 steps as follows.

*Step 1.* We will show that $|\rho(t) - e^{-(1/2)\sigma^2t^2}| \leq (1/16)\sigma^2t^4e^{-(1/2)\sigma^2t^2}$ for $t \in [0, \sqrt{3/\sigma}]$.

From Lemma 2.1(2), we have

$$\rho(t) = \prod_{j=1}^{n}\rho_j(t) \leq e^{-(1/2)\sigma^2t^2 + (1/24)\sigma^2t^4}, \tag{2.15}$$

which implies that

$$\rho(t) - e^{-(1/2)\sigma^2t^2} \leq e^{-(1/2)\sigma^2t^2}\left(e^{(1/24)\sigma^2t^4} - 1\right) \leq \frac{1}{24}\sigma^2t^4e^{-(1/2)\sigma^2t^2 + (1/24)\sigma^2t^4}, \tag{2.16}$$

where we have used the fact that $e^x - 1 \leq xe^x$ for $x > 0$ in the last inequality.

By Lemma 2.1(3) and the fact that $0 \le p_j q_j \le 1/4$, we have

$$\begin{aligned}\rho(t) - e^{-(1/2)\sigma^2 t^2} &\ge e^{-(1/2)\sigma^2 t^2}\left(e^{-(1/4)\sum_{j=1}^{n} p_j^2 q_j^2 t^4} - 1\right)\\ &\ge e^{-(1/2)\sigma^2 t^2}\left(e^{-(1/16)\sigma^2 t^4} - 1\right)\\ &\ge -\frac{1}{16}\sigma^2 t^4 e^{-(1/2)\sigma^2 t^2},\end{aligned} \tag{2.17}$$

where we have used the fact that $e^{-x} - 1 > -x$ for $x > 0$ in the last inequality.

From (2.16), (2.17) and the fact that $e^{(1/24)\sigma^2 t^4} \le 3/2$ for $t \in [0, \sqrt{3/\sigma}]$, we have

$$\left|\rho(t) - e^{-(1/2)\sigma^2 t^2}\right| \le \frac{1}{16}\sigma^2 t^4 e^{-(1/2)\sigma^2 t^2} \quad \text{on } \left[0, \sqrt{\frac{3}{\sigma}}\right]. \tag{2.18}$$

*Step 2.* We will show that

$$\sin\left(\sigma x t - \alpha(t)\right) = \sin(\sigma x t) - \frac{1}{6}\sum_{j=1}^{n} p_j q_j (p_j - q_j) t^3 \cos(\sigma x t) + \triangle_1, \tag{2.19}$$

where $|\triangle_1| \le 0.0285 t^5 + 0.0035 t^6$ and $t \in [0, \sqrt{3/\sigma}]$.

From Uspensky [12, page 124], we see that

$$\Theta_j^{(1)}(0) = p_j, \qquad \Theta_j^{(2)}(0) = 0, \qquad \Theta_j^{(3)}(0) = p_j q_j (p_j - q_j), \tag{2.20}$$

and for $t \in [0, \pi/2]$,

$$\begin{aligned}\left|\Theta_j^{(3)}(t)\right| &\le \frac{9}{8} p_j q_j \left|p_j - q_j\right| \left(1 - 4 p_j q_j \sin^2 \frac{t}{2}\right)^{-3},\\ \left|\Theta_j^{(4)}(t)\right| &\le 2 p_j q_j \left|p_j - q_j\right| \left(1 - 4 p_j q_j \sin^2 \frac{t}{2}\right)^{-4} t.\end{aligned} \tag{2.21}$$

Hence, for $t \in [0, \sqrt{3/\sigma}]$ and $\sigma^2 \ge 100$, we have

$$\begin{aligned}\left|\Theta_j^{(3)}(t)\right| &\le \frac{9 p_j q_j}{8(1 - 3/4\sigma)^3} \le 1.4215 p_j q_j,\\ \left|\Theta_j^{(4)}(t)\right| &\le \frac{2 p_j q_j t}{(1 - 3/4\sigma)^4} \le 2.7319 p_j q_j t.\end{aligned} \tag{2.22}$$

Hence

$$\begin{aligned}\alpha(t) &= \frac{1}{6}\sum_{j=1}^{n} p_j q_j (p_j - q_j) t^3 + M_1(t) t^5,\\ \alpha(t) &= M_2(t) t^3,\end{aligned} \tag{2.23}$$

where $|M_1(t)| \le 0.0285$ and $|M_2(t)| \le 0.0593$. So

$$\begin{aligned}
&\sin(\sigma x t - \alpha(t)) \\
&\quad = \sin(\sigma x t)\cos(\alpha(t)) - \sin(\alpha(t))\cos(\sigma x t) \\
&\quad = \sin(\sigma x t)\left(1 - \frac{1}{2}\cos(t_0)\alpha^2(t)\right) \\
&\qquad - \left(\alpha(t) - \frac{1}{2}\sin(t_1)\alpha^2(t)\right)\cos(\sigma x t), \quad \text{for some } t_0 \text{ and } t_1, \\
&\quad = \sin(\sigma x t) - \frac{1}{6}\sum_{j=1}^{n} p_j q_j (p_j - q_j) t^3 \cos(\sigma x t) + \triangle_1,
\end{aligned} \tag{2.24}$$

where $|\triangle_1| \le |M_1(t)|t^5 + |\alpha^2(t)| \le 0.0285t^5 + 0.0035t^6$.

*Step 3.* We will show that

$$R(x) = \frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} e^{-(1/2)\sigma^2 t^2} \frac{\sin(\sigma x t - \alpha(t))}{t} dt + \triangle_2, \tag{2.25}$$

where $|\triangle_2| \le 0.0713/\sigma^2$.

By Lemma 2.1(1) and (2.12), we have

$$\begin{aligned}
\frac{1}{2\pi}\int_{\sqrt{(3/4\sigma)}\pi}^{\pi} \frac{\rho(t)\sin(\sigma x t - \alpha(t))}{\sin(t/2)} dt &\le \frac{1}{2}\int_{\sqrt{(3/4\sigma)}\pi}^{\infty} \frac{e^{-(2/\pi^2)\sigma^2 t^2}}{t} dt \\
&= \frac{1}{2}\int_{\sqrt{3\sigma/2}}^{\infty} \frac{e^{-t^2}}{t} dt \\
&\le \frac{1}{3\sigma}\int_{\sqrt{3\sigma/2}}^{\infty} t e^{-t^2} dt \\
&= \frac{1}{6\sigma} e^{-3\sigma/2} \\
&\le 0.0167 e^{-3\sigma/2}.
\end{aligned} \tag{2.26}$$

By (2.12) and the fact that $\rho(t)$ is decreasing on $[0, \pi/2]$, we have

$$\begin{aligned}
&\frac{1}{2\pi}\int_{\sqrt{3/\sigma}}^{\sqrt{(3/4\sigma)}\pi} \frac{\rho(t)\sin(\sigma x t - \alpha(t))}{\sin(t/2)} dt \\
&\quad \le \frac{1}{2}\int_{\sqrt{3/\sigma}}^{\sqrt{(3/4\sigma)}\pi} \frac{\rho(t)}{t} dt \\
&\quad \le \frac{1}{2}\rho\left(\sqrt{\frac{3}{\sigma}}\right)\int_{\sqrt{3/\sigma}}^{\sqrt{(3/4\sigma)}\pi} \frac{1}{t} dt \quad \text{which by Lemma 2.1(2)} \\
&\quad \le \frac{3}{4} e^{-(3/2)\sigma} \ln\frac{\pi}{2} \\
&\quad = 0.3383 e^{-(3/2)\sigma}.
\end{aligned} \tag{2.27}$$

From (2.26), (2.27) and the fact that

$$e^{(3/2)\sigma} \geq \frac{1}{10!}\left(\frac{3}{2}\sigma\right)^{10} \geq 1589\sigma^2 \quad \text{for } \sigma^2 \geq 100, \tag{2.28}$$

we have

$$R(x) = \frac{1}{2\pi}\int_0^{\sqrt{3/\sigma}} \frac{\rho(t)\sin\left(\sigma x t - \alpha(t)\right)}{\sin(t/2)}dt + \triangle_{21}, \tag{2.29}$$

where $|\triangle_{21}| \leq 0.00024/\sigma^2$.

Since $\sin x = x - \cos(x_0)(x^3/6)$ for some $x_0$, we have

$$\left|\frac{1}{\sin x} - \frac{1}{x}\right| = \left|\frac{\sin x - x}{x\sin x}\right| \leq \frac{x^2}{6\sin x}, \tag{2.30}$$

which implies that

$$\frac{1}{2\pi}\int_0^{\sqrt{3/\sigma}} \frac{\rho(t)\sin\left(\sigma x t - \alpha(t)\right)}{\sin(t/2)}dt = \frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} \frac{\rho(t)\sin\left(\sigma x t - \alpha(t)\right)}{t}dt + \triangle_{22}, \tag{2.31}$$

where

$$\begin{aligned} |\triangle_{22}| &\leq \frac{1}{48\pi}\left|\int_0^{\sqrt{3/\sigma}} \frac{\rho(t)\sin\left(\sigma x t - \alpha(t)\right)t^2}{\sin(t/2)}dt\right| \quad \text{which by (2.12)} \\ &\leq \frac{1}{48}\int_0^{\sqrt{3/\sigma}} t\rho(t)dt \quad \text{which by Lemma 2.1(2)} \\ &\leq \frac{1}{32}\int_0^{\infty} te^{-(1/2)\sigma^2 t^2}dt \\ &= \frac{1}{32\sigma^2}. \end{aligned} \tag{2.32}$$

By Step 1, we see that

$$\frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} \frac{\rho(t)\sin\left(\sigma x t - \alpha(t)\right)}{t}dt = \frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} \frac{e^{-(1/2)\sigma^2 t^2}\sin\left(\sigma x t - \alpha(t)\right)}{t}dt + \triangle_{23}, \tag{2.33}$$

where

$$\begin{aligned} |\triangle_{23}| &\leq \frac{\sigma^2}{16\pi}\int_0^{\sqrt{3/\sigma}} t^3 e^{-(1/2)\sigma^2 t^2}dt \\ &\leq \frac{\sigma^2}{16\pi}\int_0^{\infty} t^3 e^{-(1/2)\sigma^2 t^2}dt \\ &= \frac{0.0398}{\sigma^2}. \end{aligned} \tag{2.34}$$

Hence, by (2.29)–(2.34), we have the conclusion of Step 3.

*Step 4.* We will show that

$$\frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} e^{-(1/2)\sigma^2t^2}\frac{\sin(\sigma xt)}{t}dt = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-(1/2)t^2}dt + \triangle_3, \tag{2.35}$$

where $|\triangle_3| \le (6.68\times 10^{-6})/\sigma^2$.

Note that

$$\frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} e^{-(1/2)\sigma^2t^2}\frac{\sin(\sigma xt)}{t}dt = \frac{1}{\pi}\int_0^{\infty} e^{-(1/2)\sigma^2t^2}\frac{\sin(\sigma xt)}{t}dt + \triangle_3, \tag{2.36}$$

where

$$\begin{aligned} |\triangle_3| &\le \frac{1}{\pi}\int_{\sqrt{3/\sigma}}^{\infty}\frac{e^{-(1/2)\sigma^2t^2}}{t}dt \\ &\le \frac{\sigma}{3\pi}\int_{\sqrt{3/\sigma}}^{\infty} te^{-(1/2)\sigma^2t^2}dt \\ &= \frac{0.1061}{\sigma}e^{-(3/2)\sigma} \quad \text{which by (2.28)} \\ &\le \frac{6.68\times 10^{-6}}{\sigma^2}. \end{aligned} \tag{2.37}$$

Let $L(x) = \int_0^\infty e^{-(1/2)t^2}(\sin(xt)/t)dt$. From the well-known integral

$$\int_0^\infty e^{-at^2}\cos(bt)dt = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-b^2/4a} \quad \text{for } a>0, \tag{2.38}$$

we have

$$L'(x) = \int_0^\infty e^{-(1/2)t^2}\cos(xt)dt = \frac{\sqrt{\pi}}{\sqrt{2}}e^{-x^2/2} \tag{2.39}$$

which implies that

$$L(x) = \frac{\sqrt{\pi}}{\sqrt{2}}\int_0^x e^{-(1/2)t^2}dt. \tag{2.40}$$

Hence

$$\begin{aligned} \frac{1}{\pi}\int_0^\infty e^{-(1/2)\sigma^2t^2}\frac{\sin(\sigma xt)}{t}dt &= \frac{1}{\pi}\int_0^\infty e^{-(1/2)t^2}\frac{\sin(xt)}{t}dt \\ &= \frac{1}{\sqrt{2\pi}}\int_0^x e^{-(1/2)t^2}dt. \end{aligned} \tag{2.41}$$

From (2.36) and (2.41), Step 4 is proved.

*Step 5.* We will show that

$$\frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} e^{-(1/2)\sigma^2 t^2}\frac{\sin(\sigma x t-\alpha(t))}{t}dt=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-(1/2)t^2}dt-G(x)+\triangle_4, \tag{2.42}$$

where $|\triangle_4|\le 0.0218/\sigma^2$.

Differentiating (2.38) twice with respect to $b$, we have

$$\int_0^\infty t^2 e^{-at^2}\cos(bt)dt=\frac{1}{4a}\sqrt{\frac{\pi}{a}}\left(1-\frac{b^2}{2a}\right)e^{-b^2/4a}. \tag{2.43}$$

Putting $a=1/2$ and $b=x$, we have

$$\int_0^\infty t^2 e^{-(1/2)t^2}\cos(xt)dt=\sqrt{\frac{\pi}{2}}(1-x^2)e^{-x^2/2}. \tag{2.44}$$

Hence

$$\begin{aligned}
&\frac{1}{6\pi}\sum_{j=1}^n p_jq_j(p_j-q_j)\int_0^{\sqrt{3/\sigma}} t^2e^{-(1/2)\sigma^2t^2}\cos(\sigma xt)dt\\
&\quad=\frac{1}{6\pi\sigma^3}\sum_{j=1}^n p_jq_j(p_j-q_j)\int_0^\infty t^2e^{-(1/2)t^2}\cos(xt)dt\\
&\qquad-\frac{1}{6\pi\sigma^3}\sum_{j=1}^n p_jq_j(p_j-q_j)\int_{\sqrt{3\sigma}}^\infty t^2e^{-(1/2)t^2}\cos(xt)dt\\
&\quad=\frac{1}{6\sqrt{2\pi}\sigma^3}(1-x^2)\sum_{j=1}^n p_jq_j(p_j-q_j)e^{-x^2/2}+\triangle_{41},
\end{aligned} \tag{2.45}$$

where

$$\begin{aligned}
|\triangle_{41}|&\le\frac{1}{6\pi\sigma}\int_{\sqrt{3\sigma}}^\infty t^2e^{-(1/2)t^2}dt\\
&\le\frac{1}{6\sqrt{3}\pi\sigma^{3/2}}\int_{\sqrt{3\sigma}}^\infty t^3e^{-(1/2)t^2}dt\\
&=\frac{2}{3\sqrt{3}\pi\sqrt{\sigma}}e^{-(3/2)\sigma}\quad\text{which by (2.28)}\\
&\le\frac{2.4\times10^{-5}}{\sigma^2}.
\end{aligned} \tag{2.46}$$

From (2.45), Steps 2 and 4,

$$\begin{aligned}
&\frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} e^{-(1/2)\sigma^2t^2}\frac{\sin(\sigma xt-\alpha(t))}{t}dt\\
&\quad=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-(1/2)t^2}dt-\frac{1}{6\sqrt{2\pi}\sigma^3}(1-x^2)\sum_{j=1}^n p_jq_j(p_j-q_j)e^{-x^2/2}+\triangle_4,
\end{aligned} \tag{2.47}$$

where

$$\begin{aligned} |\triangle_4| &\le \frac{1}{\pi}\int_0^{\sqrt{3/\sigma}} \left(0.0285t^4 + 0.0035t^5\right) e^{-(1/2)\sigma^2 t^2}\,dt + \frac{6.68\times 10^{-6}}{\sigma^2} + \frac{2.4\times 10^{-5}}{\sigma^2} \\ &\le \frac{0.0095}{\sigma^2} + \frac{3.068\times 10^{-5}}{\sigma^2} \le \frac{0.0096}{\sigma^2}. \end{aligned} \tag{2.48}$$

From Steps 3 and 5,

$$R(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-(1/2)t^2}\,dt - G(x) + \triangle_5, \tag{2.49}$$

where $|\triangle_5| < |\triangle_2| + |\triangle_4| \le 0.0809/\sigma^2$. Hence, by (2.7), we have Theorem 1.1.

## 3. Example

We will demonstrate a possible application of approximation in Theorem 1.1 with the problem of estimating the distribution function of the number of empty cells in an equiprobable scheme for group allocation of particles introduced by Vatutin and Mikhaĭlov [13] as follows.

Suppose that $n$ groups of $s$ particles are allocated independently in $N$ cells labelled by the numbers $1, 2, \ldots, N$. It is assumed that these particles are allocated one to a cell. Let

$$S_n := \text{number of cells remaining empty after } n \text{ groups are allocated.} \tag{3.1}$$

Vatutin and Mikhaĭlov [13] showed that the distribution function of $S_n$ coincides with that for a sum of independent Bernoulli random variables with

$$\begin{aligned} \mu &= N\left(1 - \frac{s}{N}\right)^n, \\ \sigma^2 &= N\left(1 - \frac{s}{N}\right)^n \left[1 - N\left(1 - \frac{s}{N}\right)^n + (N-1)\left(1 - \frac{s}{N-1}\right)^n\right]. \end{aligned} \tag{3.2}$$

From Theorem 1.1, we see that

$$\Delta_n \le \frac{0.1618}{\sigma^2}, \tag{3.3}$$

where $\sigma^2$ is defined in (3.2). We note that our bound is simpler than that in Volkova [14] and easy to evaluate.

## Acknowledgments

The author would like to thank the referees for useful comments and acknowledge Thailand Research Fund for a financial support.

## References

[1] R. N. Bhattacharya and R. Ranga Rao, *Normal Approximation and Asymptotic Expansions*, John Wiley & Sons, New York, 1976.

[2] L. H. Y. Chen and Q.-M. Shao, *A non-uniform Berry-Esseen bound via Stein's method*, Probab. Theory Related Fields **120** (2001), no. 2, 236–254.

[3] A. De Moivre, *The Doctrine of Chances: or, a Method of Calculating the Probability of Events in Play*, W. Pearson, London, 1718.

[4] P. Deheuvels, M. L. Puri, and S. S. Ralescu, *Asymptotic expansions for sums of nonidentically distributed Bernoulli random variables*, J. Multivariate Anal. **28** (1989), no. 2, 282–303.

[5] W. Feller, *On the Berry-Esseen theorem*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **10** (1968), 261–268.

[6] J. E. Kolassa, *Series Approximation Methods in Statistics*, Lecture Notes in Statistics, vol. 88, Springer-Verlag, New York, 1994.

[7] P.-S. Laplace, *Theórie Analytique des Probabilités*, Mme Ve Courcier, Paris, 1812.

[8] H. Makabe, *A normal approximation to binomial distribution*, Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs. **4** (1955), 47–53.

[9] _____, *The Approximations to the Poisson Binomial Distribution with Their Applications to the Sampling Inspection Theory. II*, Memo No. 19610602, Kawada Branch of Union of Japanese Scientists and Engineers, Kawada, Japan, 1961.

[10] V. G. Mikhaĭlov, *On a refinement of the central limit theorem for sums of independent random indicators*, Theory Probab. Appl. **38** (1993), no. 3, 479–489.

[11] V. V. Petrov, *Limit Theorems of Probability Theory. Sequences of Independent Random Variables*, Oxford Studies in Probability, vol. 4, The Clarendon Press, Oxford University Press, New York, 1995.

[12] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.

[13] V. A. Vatutin and V. G. Mikhaĭlov, *Limit theorems for the number of empty cells in an equiprobable scheme for group allocation of particles*, Theory Probab. Appl. **27** (1982), 734–743.

[14] A. Yu. Volkova, *A refinement of the central limit theorem for sums of independent random indicators*, Theory Probab. Appl. **40** (1995), no. 4, 791–794 .

K. Neammanee: Department of Mathematics, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand

*E-mail address*: kritsana.n@chula.ac.th