ขอบเขตในการประมาณปกติของตัวแบบถ้วยอนันต์

นายสุนทร บุญตา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-53-2027-7

BOUNDS IN A NORMAL APPROXIMATION

OF AN INFINITE URN MODEL

Mr. Soontorn Boonta

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Mathematics

Department of Mathematics

Faculty of Science

Chulalongkorn University

Academic Year 2005

ISBN 974-53-2027-7

| | |
|---|---|
| Thesis title | Bounds in a Normal Approximation of an Infinite Urn Model |
| By | Mr. Soontorn Boonta |
| Field of study | Mathematics |
| Thesis Advisor | Associate Professor Kritsana Neammanee, Ph.D. |

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree.

.................................. Dean of the Faculty of Science

(Professor Piamsak Menasveta, Ph.D.)

THESIS COMMITTEE

Imchit Termwuttipong ............................................. Chairman

(Assistant Professor Imchit Termwuttipong, Ph.D.)

Kritsana Neammanee. ............................................. Thesis Advisor

(Associate Professor Kritsana Neammanee, Ph.D.)

.............................................. Member

(Songkiat Sumetkijakan, Ph.D.)

.............................................. Member

(Kittipat Wong, Ph.D.)

สุนทร บุญตา : ขอบเขตในการประมาณปกติของตัวแบบถ้วยอนันต์. (BOUNDS IN A NORMAL APPROXIMATION OF AN INFINITE URN MODEL) อ. ที่ปรึกษา : รศ. ดร. กฤษณะ เนียมมณี, 42 หน้า. ISBN 974-53 -2027 -7.

ให้ $N(n)$ เป็นตัวแปรสุ่มปัวซงที่มีพารามิเตอร์ $n$ เรานิยามตัวแบบถ้วยอนันต์เป็นตัวแบบของการโยนลูกบอลอย่างอิสระจำนวน $N(n)$ ลูก ลงไปในถ้วยที่มีจำนวนอนันต์โดยความน่าจะเป็นที่ลูกบอลแต่ละลูกจะถูกโยนลงในถ้วยที่ $k$ มีค่าเท่ากับ $p_k$ โดยที่ $p_k > 0, p_k \ge p_{k+1}$ และ $\sum_{k=1}^{\infty} p_k = 1$ กำหนดให้ $Z_{N(n)}$ เป็นตัวแปรสุ่มที่มีค่าเป็นจำนวนถ้วยที่มีลูกบอลภายหลังการโยนลูกบอลจำนวน $N(n)$ ลูก Dutko ได้แสดงในปี 1989 ว่า ภายใต้เงื่อนไข

$$\lim_{n\to\infty} Var(Z_{N(n)}) = \infty$$

เราจะได้ว่า

$$\lim_{n\to\infty} F_n(x) = \Phi(x)$$

เมื่อ $F_n$ เป็นฟังก์ชันการแจกแจงของ $\dfrac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}$ และ $\Phi$ คือฟังก์ชันการแจกแจงปกติมาตรฐาน อย่างไรก็ตาม Dutko ไม่ได้ให้ขอบเขตการประมาณค่า ในวิทยานิพนธ์นี้ เราใช้เทคนิคของ Chen และ Shao ในปี 2001 เพื่อหาขอบเขตการประมาณค่าแบบสม่ำเสมอและแบบไม่สม่ำเสมอของการประมาณดังกล่าว



ภาควิชา **คณิตศาสตร์**..................................ลายมือชื่อนิสิต......สุนทร บุญตา......

สาขาวิชา **คณิตศาสตร์**..................................ลายมือชื่ออาจารย์ที่ปรึกษา..........................

ปีการศึกษา 2548..............................

## # 4672453123 : MAJOR MATHEMATICS

KEY WORDS : BOUND / NORMAL APPROXIMATION/ INFINITE URN MODEL

SOONTORN BOONTA : BOUNDS IN A NORMAL APPROXIMATION OF AN INFINITE URN MODEL. THESIS ADVISOR : ASSOC.PROF. KRITSANA NEAMMANEE, Ph.D., 42 pp. ISBN 974-53-2027-7.

Let $N(n)$ be a Poisson random variable with parameter $n$. An infinite urn model is defined as follows: $N(n)$ balls are independently placed in an infinite set of urns and each ball has probability $p_k > 0$ of being assigned to the $k$-th urn. We assume that $p_k \geq p_{k+1}$ for all $k$ and $\sum_{k=1}^{\infty} p_k = 1$. Let $Z_{N(n)}$ be the number of occupied urns after $N(n)$ balls have been thrown. Dutko showed in 1989 that under the condition

$$\lim_{n\to\infty} Var(Z_{N(n)}) = \infty,$$

we have

$$\lim_{n\to\infty} F_n(x) = \Phi(x)$$

where $F_n$ is the distribution function of $\dfrac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}$ and $\Phi$ is the standard normal distribution function. However, Dutko did not give a bound of his approximation. In our work, we use the technique in Chen and Shao (2001) to give uniform and non-uniform bounds of the approximation.



Department **Mathematics** ........................ Student's signature Soontorn Boonta

Field of study **Mathematics** ...................... Advisor's signature Kritsana Neammanee

Academic year **2005**......................................

# ACKNOWLEDGEMENTS

I would like to express my profound gratitude and deep appreciation to Associate Professor Dr. Kritsana Neammanee, my thesis advisor, for his suggestions and helpful advice in preparing the writing of my thesis. Sincere thanks and deep appreciation are also extended to Assistant Professor Dr. Imchit Termwuttipong, the chairman, Dr. Songkiat Sumetkijakan, and Dr. Kittipat Wong, the commitee members, for their comments and suggestions. Finally, I thank all teachers who have taught me all along.

In particular, I would like to express my deep gratitude to my family and friends for their encouragement throughout my graduate study.

# CONTENTS

# CHAPTER I

# INTRODUCTION

Let $N(n)$ be a Poisson random variable with parameter $n$, then we obtain that $P(N(n) = k) = \dfrac{e^{-n}n^k}{k!}$ for $k = 0, 1, 2, \ldots$. An infinite urn model is defined as follows: $N(n)$ ball are independently placed in an infinite set of urns and each ball has probability $p_k > 0$ of being assigned to the $k$ $th$ urn. We assume that $p_k \geq p_{k+1}$ for all $k$ and $\sum_{k=1}^{\infty} p_k = 1$. Let $Z_n$ be the number of occupied urns after $n$ balls have been thrown. So $Z_{N(n)}$ is the number of occupied urns after $N(n)$ ball have been thrown. Since the number of urns is infinite and the number of thrown is random, we cannot apply the usual central limit theorem to $Z_n$ and $Z_{N(n)}$. In 1967, Karlin gave the condition on $(p_k)$ for the convergence of $\dfrac{Z_n - E(Z_n)}{b_n}$ to $\mathcal{N}(0,1)$ where $\mathcal{N}(0,1)$ is the standard normal random variable and $b_n^2 \sim Var(Z_n)$. In 1989, Dutko considered in case of random thrown. Under the condition

$$\lim_{n\to\infty} Var(Z_{N(n)}) = \infty, \tag{1.1}$$

he showed that

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \tag{1.2}$$

where $F_n$ is the distribution function of $\dfrac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}$ and $\Phi$ is the standard normal distribution function. Examples of $(p_k)$ which satisfy (1.1) are $p_k = \dfrac{C}{k^{\log k}}$ and

$p_k = \dfrac{C}{k^r}$ where $C$ is a normalizing constant and $r > 1$ (see [1], page 1257-1258). In 1972, Stein gave a new technique to find a bound in normal approximation. His technique relied instead on the elementary differential equation. In 2001, Chen and Shao combined truncation with Stein's method and by taking the concentration inequality approach to find uniform and non-uniform bounds on Berry-Esseen theorem. In our work, we use the technique in Chen and Shao to obtain bounds on the convergence of (1.2).

Let $F_n$ and $\Phi$ be the distribution function of $\dfrac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}$ and $\mathcal{N}(0,1)$ respectively. The followings are our main results.

**Theorem 1**

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}}.$$

**Theorem 2** *There exists an absolute constant $C$ such that for every real number $x$,*

$$|F_n(x) - \Phi(x)| \leq \frac{C}{(1+|x|)^3\sqrt{Var(Z_{N(n)})}}.$$

Furthermore, under the condition (1.1) we have the bounds in Theorem 1 and Theorem 2 tend to zero as $n \to \infty$.

# CHAPTER II

# PRELIMINARIES

In this chapter, we present some basic concepts and facts of probability theory that are needed in this thesis. The proofs of the statements are omitted as they can be found in [2] and [3].

## 2.1 Random Variables and Distribution functions

A **probability space** is a measure space $(\Omega, \mathcal{F}, P)$ for which $P(\Omega) = 1$. The set $\Omega$ will be refered as a **sample space**. The elements of $\mathcal{F}$ are called **events**. For any event $A$, the value $P(A)$ is called the **probability of** $A$.

A function $X$ from the probability space $(\Omega, \mathcal{F}, P)$ to the set of complex numbers $\mathbb{C}$ is said to be a **complex-valued random variable** if for every borel set $B$ in $\mathbb{C}$, $X^{-1}[B]$ belongs to $\mathcal{F}$. In case that $X$ is real-valued, we say that it is a **real-valued random variable**, or simply a **random variable**. We note that the composition beetween a Borel function and a complex-valued random variable is also a complex-valued random variable.

We will use the notation $P(X \le x), P(X \ge x)$ and $P(|X| \ge x)$ to denote $P(\{\omega | X(\omega) \le x\}), P(\{\omega | X(\omega) \ge x\})$ and $P(\{\omega | \ |X|(\omega) \ge x\})$, respectively.

We define the **expectation** of a complex-valued random variable $X$ to be

$$\int_{\Omega} X dP,$$

provided that the integral $\int_{\Omega} X dP$ exists. It will be denoted by $E(X)$ or $EX$. The expectation of a random variable $X$ is known as the **mean.** The expectation of $(X - E(X))^2$ is known as the **variance** of $X$ and is denoted by $Var(X)$.

Let $\{\mathcal{F}_\alpha | \alpha \in \Lambda\}$ be a family of $\sigma$-algebras. We say that $\{\mathcal{F}_\alpha | \alpha \in \Lambda\}$ is a family of **independent $\sigma$-algebras** if for every finite subset $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of $\Lambda$

$$P(\bigcap_{i=1}^{n} A_{\alpha_i}) = \prod_{i=1}^{n} P(A_{\alpha_i})$$

for all $A_{\alpha_i} \in \mathcal{F}_{\alpha_i}$ and all $i = 1, 2, \ldots, n$.

We say that $\{E_\alpha \in \mathcal{F} | \alpha \in \Lambda\}$ is a family of **independent events** if $\{\sigma(\{E_\alpha\}) | \alpha \in \Lambda\}$ is a family of independent $\sigma$-algebras where

$$\sigma(\{E_\alpha\}) = \{\varnothing, E_\alpha, E_\alpha^C, \Omega\}$$

and $\{X_\alpha | \alpha \in \Lambda\}$ is a family of **independent random variables** if $\{\sigma(X_\alpha) | \alpha \in \Lambda\}$ is a family of independent $\sigma$-algebras where

$$\sigma(X_\alpha) = \{X_\alpha^{-1}(B) | B \text{ is a Borel set in } \mathbb{R}\}.$$

**Proposition 2.1.** ([3], p.55) *Let $\{X_\alpha,\ \alpha \in \Lambda\}$ be a family of independent random variables. For every $\alpha \in \Lambda$, let $g_\alpha$ be a Borel measurable function defined on $\mathbb{R}$. Then $\{g_\alpha(X_\alpha),\ \alpha \in \Lambda\}$ is also a family of independent random variables.*

Let $X$ be a random variable. A function $F : \mathbb{R} \to [0, 1]$ is defined by

$$F(x) = P(X \leq x),$$

for each real number $x$. $F$ is called the **distribution function of the random variable** X.

Let $X$ be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. $X$ is said to be a **discrete random variable** if the image of $X$ is countable and $X$ is called a **continuous random variable** if $F$ can be written in the form

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

for some nonnegative integrable function $f$ on $\mathbb{R}$ and in this case, we say that $f$ is the **probability function** of $X$.

Now we will give some examples of random variable.

**Example 2.2.** *$I_E$ is said to be an* ***indicator random variable*** *with respect to event $E$ if*

$$I_E(\omega) = \begin{cases} 1, & \text{if } \omega \in E; \\ 0, & \text{if } \omega \notin E. \end{cases}$$

*We note that $E(I_E) = P(E)$.*

**Example 2.3.** *$X$ is said to be* ***Poisson random variable*** *with parameter $\lambda$, written as $X \sim Poi(\lambda)$, if its image is $\{0, 1, 2, \ldots\}$ and*

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

**Example 2.4.** *We say that $X$ is a* ***normal random variable*** *with parameter $\mu$ and $\sigma^2$, written as $X \sim \mathcal{N}(\mu, \sigma^2)$, if its probability function is*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{1}{2\sigma^2}(x-\mu)^2).$$

*Specifically, $X$ is a* ***standard normal random variable*** *if $X \sim \mathcal{N}(0, 1)$.*

Proposition 2.5 is the property of $\Phi$ where $\Phi$ is the distribution function of $\mathcal{N}(0,1)$.

**Proposition 2.5.** *([4], p.295-297 and [5], p.246) Let $W$ be a random variable such that $E(W)=0$ and $Var(W)=1$. Then*

$$|P(W \le x) - \Phi(x)| \le 0.55$$

*for all $x \ge 0$.*

The following are properties of expectation and variance which we need in our work.

**Proposition 2.6.** *([2], p.59) Let $(X_j)$ be a sequence of random variables from $(\Omega, \mathcal{F}, P)$ to $[0,\infty)$. Then $E(\sum_{j=1}^{\infty} X_j) = \sum_{j=1}^{\infty} E(X_j)$.*

**Proposition 2.7.** *([2], p.64) Let $(X_j)$ be a sequence of random variables in $L^1(\Omega,\ P)$ such that $\sum_{j=1}^{\infty} E|X_j| < \infty$. Then*

$$E(\sum_{j=1}^{\infty} X_j) = \sum_{j=1}^{\infty} E(X_j).$$

**Proposition 2.8.** *([3], p.55) Let $X_1, X_2$ be two independent random variables. If $E|X_1| < \infty$ and $E|X_2| < \infty$, then $E|X_1X_2| < \infty$, and $E(X_1X_2) = EX_1EX_2$.*

Now we will give the definition condition expectation of a random variable.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{D} \subseteq \mathcal{F}$ be a $\sigma$-algebra. Let $P_{\mathcal{D}}$ be a probability measure induced by $P$, that is, $P_{\mathcal{D}}(E) = P(E)$ for all $E \in \mathcal{D}$. Let $X$ be a random variable defined on $(\Omega, \mathcal{F}, P)$ such that $E(X)$ exists. Then for every $E \in \mathcal{D}$ we can define the indefinite integral

$$Q_X(E) = \int_E X dP = \int_\Omega X I_E dP.$$

Clearly $Q_X$ is a finite signed measure on $\mathcal{D}$ such that $Q_X(E) = 0$ for every $E \in \mathcal{D}$ for which $P_D(E) = 0$. Hence $Q_X << P_{\mathcal{D}}$, so that in view of the Radon-Nikodym theorem there exists a $\mathcal{D}$- measurable function defined on $\Omega$, which we denote by $E(X|\mathcal{D})$, such that the relation

$$\int_E E(X|\mathcal{D})dP_{\mathcal{D}} = Q_X(E) = \int_E XdP$$

holds for every $E \in \mathcal{D}$. Here the function $E(X|\mathcal{D})$ is determined uniquely with respect to $P_{\mathcal{D}}$ in the sense that, if there exists another $\mathcal{D}$-measurable function $g$ on $\Omega$ satisfying

$$\int_E gdP_{\mathcal{D}} = Q_X(E)$$

for every $E \in \mathcal{D}$, then $g = E(X|\mathcal{D})$ a.s. $[P_{\mathcal{D}}]$. The measurable function $E(X|\mathcal{D})$ is called the **conditional expection of $X$ with respect to $\mathcal{D}$.**

Let $X$ and $Y$ be random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that $E(|X|) < \infty$. $E(X|\sigma(Y))$ is called the **conditional expectation of $X$ with respect to $Y$**, where $\sigma(Y) = \{Y^{-1}(B)|B \text{ is a Borel set in } \mathbb{R}\}$. We denote $E(X|\sigma(Y))$ by $E(X|Y)$ or $E^Y(X)$.

The **conditional probability** $P(A|\mathcal{D})$ of an event $A \in \mathcal{F}$, given $\mathcal{D}$, is defined by

$$P(A|\mathcal{D}) = E(I_A|\mathcal{D}).$$

**Proposition 2.9.** ([3], p.365) *Let $X$ be a random variable on $(\Omega, \mathcal{F}, P)$ with $E(X)$ exists and let $\mathcal{D} \subseteq \mathcal{F}$ be a $\sigma$-algebra and $\sigma(X)$ and $\mathcal{D}$ are independent. Then $E(X|\mathcal{D}) = E(X)$ a.s. $[P_{\mathcal{D}}]$*

## 2.2 An Infinite Urn Model

An infinite urn model is defined that $n$ ball are independently placed in an infinite set of urns and each ball has probability $p_k > 0$ of being assigned to the $k$ $th$ urn.We assume that $p_k \geq p_{k+1}$ and $\sum_{k=1}^{\infty} p_k = 1$, we define the random variable $S_{n,k}$ by

$$S_{n,k} = \text{the number of balls in the } k \text{ } th \text{ urn after } n \text{ throws.}$$

We need to consider the case where the number of throws is not fixed in advance but depends on the outcome of a random experiment. Specifically, suppose that the number of balls thrown is a Poisson random variable with means $n$, denoted by $N(n)$, then we have

$$P(N(n) = r) = \frac{e^{-n} n^r}{r!} \text{ for all } r = 0, 1, 2, \dots \text{ .}$$

We define the random variable $S_{N(n),k}$ by

$$S_{N(n),k} = \text{the number of balls in the } k \text{ } th \text{ urn after } N(n) \text{ throws.}$$

The random variables $(S_{N(n),k})$, $k = 1, 2, \dots$ are mutually independent Poisson random variables with respective mean $(np_k)$ (see [1], p.1259), so that

$$P(S_{N(n),k} = r) = \frac{e^{-np_k}(np_k)^r}{r!} \text{ for all } r = 0, 1, 2, \dots \text{ .}$$

We next define the random variable $Z_n$ by

$$Z_n = \sum_{k=1}^{\infty} I(S_{n,k}), \text{ where } I(u) = \begin{cases} 1, & \text{if } u > 0, \\ 0, & \text{if } u = 0. \end{cases}$$

Also, we define the random variable $Z_{N(n)}$ by

$$Z_{N(n)} = \sum_{k=1}^{\infty} I(S_{N(n),k}).$$

The random variable $Z_n$ is the number of occupied urn after $n$ balls have been thrown, and the random variable $Z_{N(n)}$ is the number of occupied urn after $N(n)$ balls have been thrown. From Dutko[1] we know

$$(I(S_{N(n),k}))_{k\in\mathbb{N}} \text{ is the sequence of independent random variables,} \tag{2.1}$$

$$E(I(S_{N(n),k})) = 1 - e^{-np_k},$$

$$Var(I(S_{N(n),k})) = e^{-np_k} - e^{-2np_k},$$

$$E(Z_{N(n)}) = \sum_{k=1}^{\infty}(1 - e^{-np_k}),$$

$$Var(Z_{N(n)}) = \sum_{k=1}^{\infty}(e^{-np_k} - e^{-2np_k}), \tag{2.2}$$

$$E(Z_{N(n)}) \text{ is finite}$$

$$\text{and } Var(Z_{N(n)}) \text{ is finite.} \tag{2.3}$$

Let

$$X_{n,k} = \frac{I(S_{N(n),k}) - E(I(S_{N(n),k}))}{\sqrt{Var(Z_{N(n)})}}. \tag{2.4}$$

Then

$$|X_{n,k}| \leq \frac{1}{\sqrt{Var(Z_{N(n)})}}, \tag{2.5}$$

$$\frac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}} = \sum_{k=1}^{\infty} X_{n,k}, \tag{2.6}$$

$$E(\sum_{k=1}^{\infty} X_{n,k}) = 0 \text{ and } Var(\sum_{k=1}^{\infty} X_{n,k}) = 1.$$

**Proposition 2.10.**

1. $\displaystyle\sum_{k=1}^{\infty} E|X_{n,k}|^3 \leq \frac{1}{\sqrt{Var(Z_{N(n)})}}$ *which is finite.*
2. *If* $\displaystyle\lim_{n\to\infty} Var(Z_{N(n)}) = \infty$, *then* $\displaystyle\lim_{n\to\infty} \sum_{k=1}^{\infty} E|X_{n,k}|^3 = 0.$

*Proof.*

1. We note that

$$I(S_{N(n),k}) = \begin{cases} 0 & \text{if } S_{N(n),k} = 0, \\ 1 & \text{if } S_{N(n),k} > 0. \end{cases}$$

So $P(I(S_{N(n),k}) = 0) = P(S_{N(n),k} = 0) = e^{-np_k}$

and $P(I(S_{N(n),k}) = 1) = P(S_{N(n),k} > 0) = 1 - P(S_{N(n),k} = 0) = 1 - e^{-np_k}$, i.e.,

$$I(S_{N(n),k}) = \begin{cases} 0 & \text{with the probability } e^{-np_k}, \\ 1 & \text{with the probability } 1 - e^{-np_k}. \end{cases}$$

Thus

$$P\Big(X_{n,k} = \frac{e^{-np_k} - 1}{\sqrt{Var(Z_{N(n)})}}\Big) = P(I(S_{N(n),k}) = 0) = e^{-np_k} \quad \text{and}$$
$$P\Big(X_{n,k} = \frac{e^{-np_k}}{\sqrt{Var(Z_{N(n)})}}\Big) = P(I(S_{N(n),k}) = 1) = 1 - e^{-np_k}.$$

Hence,

$$\begin{aligned} E\ |X_{n,k}|^3 &= \sum_{x \in Im\ X_{n,k}} |x|^3 P(X_{n,k} = x) \\ &= \Big|\frac{e^{-np_k} - 1}{\sqrt{Var(Z_{N(n)})}}\Big|^3 e^{-np_k} + \Big|\frac{e^{-np_k}}{\sqrt{Var(Z_{N(n)})}}\Big|^3 (1 - e^{-np_k}) \end{aligned}$$

$$= \frac{(1-e^{-np_k})^3}{(Var(Z_{N(n)}))^{\frac{3}{2}}}e^{-np_k} + \frac{e^{-3np_k}}{(Var(Z_{N(n)}))^{\frac{3}{2}}}(1-e^{-np_k})$$
$$= \frac{(1-3e^{-np_k}+3e^{-2np_k}-e^{-3np_k})e^{-np_k}+e^{-3np_k}(1-e^{-np_k})}{(Var(Z_{N(n)}))^{\frac{3}{2}}}$$
$$= \frac{e^{-np_k}-3e^{-2np_k}+3e^{-3np_k}-e^{-4np_k}+e^{-3np_k}-e^{-4np_k}}{(Var(Z_{N(n)}))^{\frac{3}{2}}}$$
$$= \frac{-2e^{-4np_k}+4e^{-3np_k}-3e^{-2np_k}+e^{-np_k}}{(Var(Z_{N(n)}))^{\frac{3}{2}}}.$$

We observe that

$$\begin{aligned}&-2e^{-4np_k}+4e^{-3np_k}-3e^{-2np_k}+e^{-np_k}\\ &\quad= -2e^{-4np_k}+2e^{-3np_k}+2e^{-3np_k}-2e^{-2np_k}-e^{-2np_k}+e^{-np_k}\\ &\quad= 2e^{-2np_k}(e^{-np_k}-e^{-2np_k})-2e^{-np_k}(e^{-np_k}-e^{-2np_k})+(e^{-np_k}-e^{-2np_k}),\end{aligned}$$

then

$$\sum_{k=1}^{\infty} E|X_{n,k}|^3 = A_n + B_n + C_n$$

where

$$A_n = \frac{2\sum_{k=1}^{\infty} e^{-2np_k}(e^{-np_k}-e^{-2np_k})}{(Var(Z_{N(n)}))^{\frac{3}{2}}},$$

$$B_n = \frac{-2\sum_{k=1}^{\infty} e^{-np_k}(e^{-np_k}-e^{-2np_k})}{(Var(Z_{N(n)}))^{\frac{3}{2}}} \quad \text{and}$$

$$C_n = \frac{\sum_{k=1}^{\infty}(e^{-np_k}-e^{-2np_k})}{(Var(Z_{N(n)}))^{\frac{3}{2}}}.$$

By (2.2), (2.3) and $A_n + B_n < 0$, $\sum_{k=1}^{\infty} E|X_{n,k}|^3 \leq \frac{1}{\sqrt{Var(Z_{N(n)})}}$.

2. This follows from 1. $\square$

## 2.3 Stein's equation for standard normal distribution function

In this section, we will introduce Stein's method which is based on the differential equation

$$f'(\omega) - \omega f(\omega) = h(\omega) - E(h(Z)) \tag{2.7}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a continuous function, $h : \mathbb{R} \to \mathbb{R}$ is a test function and $Z$ is the standard normal random variable. The equation (2.7) is called **Stein's equation for normal approximation**. For any real number $x$, let $h$ be an indicator function defined by

$$h_x(\omega) = \begin{cases} 1, & \text{if} \quad \omega \leq x, \\ 0, & \text{if} \quad \omega > x, \end{cases} \tag{2.8}$$

then Stein's equation (2.7) has a unique solution $f_x : \mathbb{R} \to \mathbb{R}$ defined by

$$f_x(\omega) = \begin{cases} \sqrt{2\pi}\ e^{\frac{1}{2}\ \omega^2}\Phi(\omega)[1-\Phi(x)], & \text{if} \quad \omega \leq x; \\ \sqrt{2\pi}\ e^{\frac{1}{2}\ \omega^2}\Phi(x)[1-\Phi(\omega)], & \text{if} \quad \omega > x \end{cases} \tag{2.9}$$

(see [6], p.22).

By (2.7) – (2.9) we get

$$f_x'(W) - Wf_x(W) = h_x(W) - \Phi(x)$$

which implies that

$$P(W \leq x) - \Phi(x) = Ef_x'(W) - EWf_x(W). \tag{2.10}$$

Hence we can find a bound of $Ef'_x(W) - EWf_x(W)$ instead of $P(W \leq x) - \Phi(x)$. To find a bound of $Ef'_x(W) - EWf_x(W)$, we need the following properties of the solution $f_x$ of Stein's equation (2.7).

For $\omega, s, t \in \mathbb{R}$,

$$f'_x(\omega+s) - f'_x(\omega+t) \leq \begin{cases} 1, & \text{if } \omega+s \leq x,\ \omega+t > x; \\ (|\omega| + \frac{\sqrt{2\pi}}{4})(|s|+|t|), & \text{if } s \geq t; \\ 0, & \text{elsewhere,} \end{cases} \tag{2.11}$$

$$f'_x(\omega+s) - f'_x(\omega+t) \geq \begin{cases} -1, & \text{if } \omega+s > x,\ \omega+t \leq x; \\ -(|\omega| + \frac{\sqrt{2\pi}}{4})(|s|+|t|), & \text{if } s < t; \\ 0, & \text{elsewhere;} \end{cases} \tag{2.12}$$

and

$$|f'_x(s) - f'_x(t)| \leq 1 \tag{2.13}$$

(see [5], p.246-247).

# CHAPTER III

# A UNIFORM BOUND ON AN INFINITE URN MODEL

Let $Z_{N(n)}$ be defined as in section 2.2 of Chapter II. The purpose of this chapter is to give a uniform bound in the approximation of the distribution function of $\dfrac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}$ by $\Phi$ is stated in Theorem 3.5. To prove Theorem 3.5, we need the following results.

**Proposition 3.1.** *Let $\delta \in \mathbb{R}^+$ and $M: \mathbb{R}^2 \to \mathbb{R}$ be defined by*

$$M(\omega, t) = \omega I(-\omega \leq t \leq 0) - \omega I(0 < t \leq -\omega),$$

*then* $\displaystyle\int_{-\delta}^{\delta} M(\omega, t)dt = |\omega| \min(\delta, |\omega|)$.

*Proof.*

**Case 1** $\min(\delta, |\omega|) = \delta$.

If $\omega \geq 0$, then

$$\begin{aligned}
\int_{-\delta}^{\delta} M(\omega, t)dt &= \int_{-\delta}^{\delta} \omega I(-\omega \leq t \leq 0) - \omega I(0 < t \leq -\omega)dt \\
&= \int_{-\delta}^{\delta} \omega I(-\omega \leq t \leq 0)dt \\
&= \int_{-\delta}^{0} \omega dt \\
&= |\omega|\delta
\end{aligned}$$

$$= |\omega| \min(\delta, |\omega|).$$

If $\omega < 0$, then

$$\begin{aligned}
\int_{-\delta}^{\delta} M(\omega,t)dt &= \int_{-\delta}^{\delta} \omega I(-\omega \le t \le 0) - \omega I(0 < t \le -\omega)dt \\
&= \int_{-\delta}^{\delta} -\omega I(0 < t \le -\omega)dt \\
&= \int_{0}^{\delta} -\omega dt \\
&= |\omega|\delta \\
&= |\omega| \min(\delta, |\omega|).
\end{aligned}$$

**Case 2** $\min(\delta, |\omega|) = |\omega|$.

If $\omega \ge 0$, then

$$\begin{aligned}
\int_{-\delta}^{\delta} M(\omega,t)dt &= \int_{-\delta}^{\delta} \omega I(-\omega \le t \le 0) - \omega I(0 < t \le -\omega)dt \\
&= \int_{-\delta}^{\delta} \omega I(-\omega \le t \le 0)dt \\
&= \int_{-\omega}^{0} \omega dt \\
&= |\omega||\omega| \\
&= |\omega| \min(\delta, |\omega|).
\end{aligned}$$

If $\omega < 0$, then

$$\begin{aligned}
\int_{-\delta}^{\delta} M(\omega,t)dt &= \int_{-\delta}^{\delta} \omega I(-\omega \le t \le 0) - \omega I(0 < t \le -\omega)dt \\
&= \int_{-\delta}^{\delta} -\omega I(0 < t \le -\omega)dt \\
&= \int_{0}^{-\omega} -\omega dt
\end{aligned}$$

$$= (-\omega)(-\omega)$$

$$= |\omega||\omega|$$

$$= |\omega| \min(\delta, |\omega|).$$

Then the proof is completed. □

**Proposition 3.2.** *Let $(a_i)$ be a sequence of real numbers such that the both series $\sum_{i=1}^{\infty} a_i$ and $\sum_{i=1}^{\infty} a_i^2$ are finite. Then*

$$(\sum_{i=1}^{\infty} a_i)^2 = \sum_{i=1}^{\infty} a_i^2 + \sum_{i=1}^{\infty} a_i \sum_{\substack{j=1 \\ j\neq i}}^{\infty} a_j.$$

*Proof.* We note that for each $i \in \mathbb{N}$

$$a_i \sum_{\substack{j=1 \\ j\neq i}}^{\infty} a_j = a_i(\sum_{j=1}^{\infty} a_j - a_i) = a_i \sum_{j=1}^{\infty} a_j - a_i^2.$$

Since $\sum_{i=1}^{\infty} a_i$ and $\sum_{i=1}^{\infty} a_i^2$ are finite, we have

$$\begin{aligned}
\sum_{i=1}^{\infty} a_i \sum_{\substack{j=1 \\ j\neq i}}^{\infty} a_j &= \sum_{i=1}^{\infty}(a_i \sum_{j=1}^{\infty} a_j - a_i^2) \\
&= \sum_{i=1}^{\infty} a_i \sum_{j=1}^{\infty} a_j - \sum_{i=1}^{\infty} a_i^2 \\
&= (\sum_{i=1}^{\infty} a_i)^2 - \sum_{i=1}^{\infty} a_i^2.
\end{aligned}$$

Hence

$$(\sum_{j=1}^{\infty} a_i)^2 = \sum_{i=1}^{\infty} a_i^2 + \sum_{i=1}^{\infty} a_i \sum_{\substack{j=1 \\ j\neq i}}^{\infty} a_j.$$

□

**Proposition 3.3.** *Let $X_{n,k}$ be defined in* (2.4) *and* $\delta > 0$. *Then we have*

$$Var\Big(\sum_{j=1}^{\infty}|X_{n,j}|\min(\delta,|X_{n,j}|)\Big) = \sum_{j=1}^{\infty} Var\Big(|X_{n,j}|\min(\delta,|X_{n,j}|)\Big).$$

*Proof.* By (2.1), (2.4) and Proposition 2.1, $(|X_{n,j}|\min(\delta,|X_{n,j}|))_{j\in\mathbb{N}}$ is the sequence of independent random variables. Now we will show that

$\sum\limits_{j=1}^{\infty}|X_{n,j}|\min(\delta,|X_{n,j}|)$, $\sum\limits_{j=1}^{\infty}(|X_{n,j}|\min(\delta,|X_{n,j}|))^2$, $\sum\limits_{j=1}^{\infty}E|X_{n,j}|\min(\delta,|X_{n,j}|)$, and $\sum\limits_{j=1}^{\infty}(E|X_{n,j}|\min(\delta,|X_{n,j}|))^2$ are finite. It is clear that for every $k\in\mathbb{N}$,

$$\sum_{j=1}^{\infty}|I(S_{N(n),j})|^k \text{ is finite}$$

and, by the fact that $1-e^{-np_j}\le np_j$,

$$\sum_{j=1}^{\infty}(1-e^{-np_j})^k \le \sum_{j=1}^{\infty}(1-e^{-np_j}) \le \sum_{j=1}^{\infty}np_j = n.$$

From this facts and (2.3), it follows that

$$\begin{aligned}\sum_{j=1}^{\infty}|X_{n,j}| &= \sum_{j=1}^{\infty}\Big|\frac{I(S_{N(n),j})-(1-e^{-np_j})}{\sqrt{Var(Z_{N(n)})}}\Big| \\ &\le \frac{1}{\sqrt{Var(Z_{N(n)})}}\{\sum_{j=1}^{\infty}|I(S_{N(n),j})| + \sum_{j=1}^{\infty}(1-e^{-np_j})\} \\ &< \infty\end{aligned}$$

and

$$\begin{aligned}\sum_{j=1}^{\infty}|X_{n,j}|^2 &= \sum_{j=1}^{\infty}\Big|\frac{I(S_{N(n),j})-(1-e^{-np_j})}{\sqrt{Var(Z_{N(n)})}}\Big|^2 \\ &\le \frac{1}{Var(Z_{N(n)})}\{\sum_{j=1}^{\infty}|I(S_{N(n),j})|^2 + \sum_{j=1}^{\infty}(1-e^{-np_j})^2\} \\ &< \infty.\end{aligned}$$

Hence

$$\sum_{j=1}^{\infty} |X_{n,j}| \min(\delta, |X_{n,j}|) \leq \delta \sum_{j=1}^{\infty} |X_{n,j}| \text{ and}$$
$$\sum_{j=1}^{\infty} (|X_{n,j}| \min(\delta, |X_{n,j}|))^2 \leq \delta^2 \sum_{j=1}^{\infty} |X_{n,j}|^2 \text{ are finite.}$$

Observe that

$$\begin{aligned}
E|X_{n,j}| &= \sum_{x \in Im\ X_{n,j}} |x| P(X_{n,j} = x) \\
&= \Big| \frac{e^{-np_j} - 1}{\sqrt{Var(Z_{N(n)})}} \Big| e^{-np_j} + \Big| \frac{e^{-np_j}}{\sqrt{Var(Z_{N(n)})}} \Big| (1 - e^{-np_j}) \\
&= \frac{(1 - e^{-np_j})e^{-np_j} + (1 - e^{-np_j})e^{-np_j}}{\sqrt{Var(Z_{N(n)})}} \\
&= \frac{2(e^{-np_j} - e^{-2np_j})}{\sqrt{Var(Z_{N(n)})}},
\end{aligned}$$

hence, by (2.2),

$$\sum_{j=1}^{\infty} E|X_{n,j}| = 2\sqrt{Var(Z_{N(n)})} \tag{3.1}$$

which implies

$$\sum_{j=1}^{\infty} E|X_{n,j}| \min(\delta, |X_{n,j}|) \leq \delta \sum_{j=1}^{\infty} E|X_{n,j}| = 2\delta \sqrt{Var(Z_{N(n)})} < \infty$$

and

$$\begin{aligned}
\sum_{j=1}^{\infty} (E|X_{n,j}| \min(\delta, |X_{n,j}|))^2 &\leq \delta^2 \sum_{j=1}^{\infty} (E|X_{n,j}|)^2 \\
&= \frac{4\delta^2}{Var(Z_{N(n)})} \sum_{j=1}^{\infty} (e^{-np_j} - e^{-2np_j})^2 \\
&\leq \frac{4\delta^2}{Var(Z_{N(n)})} \sum_{j=1}^{\infty} (e^{-np_j} - e^{-2np_j})
\end{aligned}$$

$$= 4\delta^2$$

$$< \infty.$$

By Proposition 3.2, we obtain

$$\begin{aligned}
&Var(\sum_{j=1}^{\infty} |X_{n,j}| \min(\delta, |X_{n,j}|)) \\
&= E(\sum_{j=1}^{\infty} |X_{n,j}| \min(\delta, |X_{n,j}|))^2 - (\sum_{j=1}^{\infty} E|X_{n,j}| \min(\delta, |X_{n,j}|))^2 \\
&= E\sum_{i=1}^{\infty} (|X_{n,i}| \min(\delta, |X_{n,i}|))^2 \\
&\quad + E\sum_{i=1}^{\infty} |X_{n,i}| \min(\delta, |X_{n,i}|) \sum_{\substack{j=1 \\ j\neq i}}^{\infty} |X_{n,j}| \min(\delta, |X_{n,j}|) \\
&\quad - \sum_{i=1}^{\infty} (E|X_{n,i}| \min(\delta, |X_{n,i}|))^2 \\
&\quad - \sum_{i=1}^{\infty} E|X_{n,i}| \min(\delta, |X_{n,i}|) \sum_{\substack{j=1 \\ j\neq i}}^{\infty} E|X_{n,j}| \min(\delta, |X_{n,j}|) \\
&= \sum_{j=1}^{\infty} E(|X_{n,j}| \min(\delta, |X_{n,j}|))^2 - \sum_{j=1}^{\infty} (E|X_{n,j}| \min(\delta, |X_{n,j}|))^2 \\
&= \sum_{j=1}^{\infty} (E(|X_{n,j}| \min(\delta, |X_{n,j}|))^2 - (E|X_{n,j}| \min(\delta, |X_{n,j}|))^2 \\
&= \sum_{j=1}^{\infty} Var(|X_{n,j}| \min(\delta, |X_{n,j}|)).
\end{aligned}$$

□

In what following, we let

$$W := \sum_{j=1}^{\infty} X_{n,j} \text{ and } \beta := \sum_{j=1}^{\infty} E|X_{n,j}|^3.$$

By (2.6), we see that

$$W = \frac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}.$$

**Proposition 3.4.** (*concentration inequality for a uniform bound*) *Let $a < b$ and for each $k \in \mathbb{N}$, let $W^{(k)} = W - X_{n,k}$. If $\beta < 0.14$, then*

$$P(a \leq W^{(k)} \leq b) \leq 1.5(b-a) + 4.21\beta,$$

*Proof.* Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(t) = \begin{cases} -\frac{1}{2}(b-a) - \frac{\beta}{2}, & \text{if} \quad t < a - \frac{\beta}{2}; \\ t - \frac{1}{2}(b+a), & \text{if} \quad a - \frac{\beta}{2} \leq t \leq b + \frac{\beta}{2}; \\ \frac{1}{2}(b-a) + \frac{\beta}{2}, & \text{if} \quad t > b + \frac{\beta}{2} \end{cases}$$

and let $M : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$M(\omega, t) = \omega I(-\omega \leq t \leq 0) - \omega I(0 < t \leq -\omega).$$

By (3.1),

$$\begin{aligned} \sum_{\substack{j=1 \\ j \neq k}}^{\infty} E|X_{n,j} f(W^{(k)})| &\leq \frac{1}{2}(b - a + \beta) \sum_{j=1}^{\infty} E|X_{n,j}| \\ &= (b - a + \beta)\sqrt{Var(Z_{N(n)})} \\ &< \infty. \end{aligned}$$

Hence it follows from Proposition 2.7 that

$$E \sum_{\substack{j=1 \\ j \neq k}}^{\infty} X_{n,j} f(W^{(k)}) = \sum_{\substack{j=1 \\ j \neq k}}^{\infty} E X_{n,j} f(W^{(k)}). \tag{3.2}$$

Since $X_{n,j}$ and $W^{(k)} - X_{n,j}$ are independent for all $j \neq k$, $EX_{n,k} = 0$ for all $k$ and

$M(\omega, t) \geq 0$, we have

$$EW^{(k)}f(W^{(k)})$$

$$= \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{X_{n,j}f(W^{(k)})\} \quad \text{(from (3.2))}$$

$$= \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{X_{n,j}f(W^{(k)})\} - E\{X_{n,j}f(W^{(k)} - X_{n,j})\},$$

$$\text{(because } X_{n,j},\ W^{(k)} - X_{n,j} \text{ are independent and } EX_{n,j} = 0)$$

$$= \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{X_{n,j}[f(W^{(k)}) - f(W^{(k)} - X_{n,j})]\}$$

$$= \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{X_{n,j} \int_{-X_{n,j}}^{0} f'(W^{(k)} + t)dt\}$$

$$= \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{\int_{-\infty}^{\infty} f'(W^{(k)} + t)X_{n,j}[I(-X_{n,j} \leq t \leq 0) - I(0 < t \leq -X_{n,j})]dt\}$$

$$= \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{\int_{-\infty}^{\infty} f'(W^{(k)} + t)M(X_{n,j}, t)dt\} \tag{3.3}$$

$$\geq \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E \int_{|t|\leq \frac{\beta}{2}} M(X_{n,j}, t)dt$$

$$\geq \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{I(a \leq W^{(k)} \leq b) \int_{|t|\leq \frac{\beta}{2}} M(X_{n,j}, t)dt\}$$

$$= E\{I(a \leq W^{(k)} \leq b) \sum_{\substack{j=1 \\ j\neq k}}^{\infty} |X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|)\} \quad \text{(by Proposition 3.1)}$$

$$= E\{I(a \leq W^{(k)} \leq b)S\} - P(a \leq W^{(k)} \leq b)E|X_{n,k}| \min(\frac{\beta}{2}, |X_{n,k}|)$$

$$\geq E\{I(a \leq W^{(k)} \leq b)SI(S \geq 0.38)\} - P(a \leq W^{(k)} \leq b)E|X_{n,k}| \min(\frac{\beta}{2}, |X_{n,k}|)$$

$$\geq 0.38E\{I(a \leq W^{(k)} \leq b)(1 - I(S < 0.38)\}$$

$$- P(a \leq W^{(k)} \leq b)E|X_{n,k}| \min(\frac{\beta}{2}, |X_{n,k}|)$$

$$= 0.38E\{I(a \le W^{(k)} \le b)\} - 0.38E\{I(a \le W^{(k)} \le b)I(S < 0.38)\}$$

$$- P(a \le W^{(k)} \le b)E|X_{n,k}| \min(\frac{\beta}{2}, |X_{n,k}|)$$

$$\ge 0.38P(a \le W^{(k)} \le b) - 0.38P(S < 0.38)$$

$$- P(a \le W^{(k)} \le b)E|X_{n,k}| \min(\frac{\beta}{2}, |X_{n,k}|)$$

$$\ge 0.34P(a \le W^{(k)} \le b) - 0.38P(S < 0.38), \tag{3.4}$$

where $S = \sum_{j=1}^{\infty} |X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|)$ and the last inequality follows from the fact that

$$E|X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|) \le \frac{\beta}{2} E|X_{n,j}| \le \frac{\beta}{2}[E|X_{n,j}|^3]^{\frac{1}{3}} \le \frac{\beta^{\frac{4}{3}}}{2} \le 0.04.$$

By the fact that

$$\min(a, b) \ge b - \frac{b^2}{4a} \tag{3.5}$$

for positive number $a$ and $b$ (see [5], p.238), we obtain that

$$E(S) = \sum_{j=1}^{\infty} E|X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|) \ge \sum_{j=1}^{\infty} EX_{n,j}^2 - \frac{1}{2\beta} \sum_{j=1}^{\infty} E|X_{n,j}|^3 = 0.5.$$

From this fact and the fact that

$$\begin{aligned}
Var(S) &= Var \sum_{j=1}^{\infty} |X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|) \\
&= \sum_{j=1}^{\infty} Var(|X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|) \quad \text{(by Proposition 3.3)} \\
&= \sum_{j=1}^{\infty} \{E(|X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|))^2 - (E|X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|))^2\} \\
&\le \sum_{j=1}^{\infty} E(|X_{n,j}| \min(\frac{\beta}{2}, |X_{n,j}|))^2 \\
&\le \frac{\beta^2}{4} \sum_{j=1}^{\infty} EX_{n,j}^2 \\
&= \frac{\beta^2}{4},
\end{aligned}$$

we have

$$
\begin{aligned}
P(S < 0.38) &= P(ES - S > ES - 0.38) \\
&\leq P(ES - S \geq 0.12) \\
&\leq \frac{Var(S)}{(0.12)^2} \quad \text{(by Chebyshev's inequality)} \\
&\leq \frac{\beta^2}{4(0.12)^2} \\
&\leq 2.45\beta, \qquad (3.6)
\end{aligned}
$$

where we have used the fact that $\frac{\beta}{2} < 0.07$ in the last inequality.

Observe that

$$|EW^{(k)}f(\ W^{(k)})| \leq \frac{1}{2}(b - a + \beta)E|W^{(k)}| \leq \frac{1}{2}(b - a + \beta).$$

Hence, by (3.4) and (3.6),

$$
\begin{aligned}
P(a \leq W^{(k)} \leq b) &\leq \frac{1}{0.34}\{\frac{1}{2}(b - a + \beta) + 0.931\beta\} \\
&= \frac{1}{0.68}(b - a) + \frac{1.431}{0.34}\beta \\
&\leq 1.5(b - a) + 4.21\beta.
\end{aligned}
$$

$\square$

Next, we will prove the main result of this chapter.

**Theorem 3.5.** *Let $F_n$ and $\Phi$ be the distribution function of $\frac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}$ and $\mathcal{N}(0,1)$, respectively. Then*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}}.$$

*Moreover, under the additional condition that $\lim_{n\to\infty} Var(Z_{N(n)}) = \infty$, the bound tends to zero.*

*Proof.* We devide the proof into 2 cases.

**Case 1** $x \geq 0$.

By Proposition 2.10(1) we know that $\beta \leq \dfrac{1}{\sqrt{Var(Z_{N(n)})}}$. If $\beta > 0.14$, then

$$0.9317 < 6.655\beta \leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}}. \tag{3.7}$$

From (3.7) and Proposition 2.5,

$$|F_n(x) - \Phi(x)| \leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}}.$$

Next, we assume that $\beta \leq 0.14$. For each $k \in \mathbb{N}$, let

$$K_k(t) = E\{X_{n,k}[I(0 \leq t \leq X_{n,k}) - I(X_{n,k} \leq t < 0)]\}.$$

Observe that

$$\sum_{k=1}^{\infty} \int_{-\infty}^{\infty} K_k(t)dt = \sum_{k=1}^{\infty} EX_{n,k}^2 = 1. \tag{3.8}$$

Let $f$ be a real-valued, bounded, continuous and piecewise differentiable function defined on the real line. Then

$$\begin{aligned}
&EWf(W)\\
&= \sum_{k=1}^{\infty} EX_{n,k}f(W)\\
&= \sum_{k=1}^{\infty} E\{X_{n,k}f(W^{(k)} + X_{n,k}) - X_{n,k}f(W^{(k)})\} \ (EX_{n,k}f(W^{(k)}) = 0)\\
&= \sum_{k=1}^{\infty} EX_{n,k} \int_0^{X_{n,k}} f'(W^{(k)} + t)dt\\
&= \sum_{k=1}^{\infty} E \int_{-\infty}^{\infty} f'(W^{(k)} + t)X_{n,k}\{I(0 \leq t \leq X_{n,k}) - I(X_{n,k} \leq t < 0)\}dt
\end{aligned}$$

$$= \sum_{k=1}^{\infty} E \int_{-\infty}^{\infty} f'(W^{(k)} + t) E\{X_{n,k}[I(0 \le t \le X_{n,k}) - I(X_{n,k} \le t < 0)]\} dt$$

$$= \sum_{k=1}^{\infty} E \int_{-\infty}^{\infty} f'(W^{(k)} + t) K_k(t) dt. \tag{3.9}$$

Let $f$ in (3.9) be the unique bound solution $f_x$ of the Stein equation (2.7) for $h_x$ which is defined by (2.8). Then, by (2.10), (3.8) and (3.9),

$$\begin{aligned}
&F_n(x) - \Phi(x) \\
&= Ef_x'(W) - EWf_x(W) \\
&= \sum_{k=1}^{\infty} E \int_{-\infty}^{\infty} f_x'(W) K_k(t) dt - \sum_{k=1}^{\infty} E \int_{-\infty}^{\infty} f_x'( W^{(k)} + t) K_k(t) dt \\
&= \sum_{k=1}^{\infty} E\{ \int_{-\infty}^{\infty} [ f_x'( W^{(k)} + X_{n,k}) - f_x'( W^{(k)} + t)] K_k(t) dt\}. \qquad (3.10)
\end{aligned}$$

By (3.10) and (2.11), we have

$$F_n(x) - \Phi(x) \le R_1 + R_2,$$

where

$$R_1 = \sum_{k=1}^{\infty} E \int_{\substack{W^{(k)}+t>x \\ W^{(k)}+X_{n,k} \le x}} K_k(t) dt \quad \text{and}$$

$$R_2 = \sum_{k=1}^{\infty} E \int_{X_{n,k} \ge t} (| W^{(k)}| + 0.63)(|X_{n,k}| + |t|) K_k(t) dt.$$

We observe that if $W^{(k)} + t > x$ and $W^{(k)} + X_{n,k} \le x$, then

$x - t < W^{(k)} \le x - X_{n,k}$ and $t > X_{n,k}$. Then

$$\begin{aligned}
R_1 &\le \sum_{k=1}^{\infty} E \int_{t > X_{n,k}} I(x - t < W^{(k)} \le x - X_{n,k}) K_k(t) dt \\
&= \sum_{k=1}^{\infty} E \int_{\mathbb{R}} I(t > X_{n,k}) I(x - t < W^{(k)} \le x - X_{n,k}) K_k(t) dt \\
&= \sum_{k=1}^{\infty} \int_{\mathbb{R}} EE^{X_{n,k}} I(t > X_{n,k}) I(x - t < W^{(k)} \le x - X_{n,k}) K_k(t) dt
\end{aligned}$$

$$= \sum_{k=1}^{\infty} \int_{\mathbb{R}} EI(t > X_{n,k}) K_k(t) E^{X_{n,k}} I(x - t < W^{(k)} \le x - X_{n,k}) dt$$
$$= \sum_{k=1}^{\infty} E \int_{t > X_{n,k}} K_k(t) E^{X_{n,k}} I(x - t < W^{(k)} \le x - X_{n,k}) dt$$
$$= \sum_{k=1}^{\infty} E \int_{t > X_{n,k}} P(x - t < \ W^{(k)} < x - X_{n,k} | X_{n,k}) K_k(t) dt. \tag{3.11}$$

By (3.11) and Proposition 3.4, we have

$$R_1 \le \sum_{k=1}^{\infty} E \int_{t > X_{n,k}} \{1.5(|t| + |X_{n,k}|) + 4.21\beta\} K_k(t) dt$$
$$\le 1.5 \sum_{k=1}^{\infty} E \int_{t > X_{n,k}} (|t| + |X_{n,k}|) K_k(t) dt + 4.21\beta. \tag{3.12}$$

Since $W^{(k)}$ and $X_{n,k}$ are independent, we have

$$R_2 = \sum_{k=1}^{\infty} E(|\ W^{(k)}| + 0.63) E \int_{X_{n,k} \ge t} (|X_{n,k}| + |t|) K_k(t) dt$$
$$\le 1.63 \sum_{k=1}^{\infty} E \int_{X_{n,k} \ge t} (|X_{n,k}| + |t|) K_k(t) dt. \tag{3.13}$$

By (3.12) and (3.13), we get

$$R_1 + R_2 \le 1.63 \sum_{k=1}^{\infty} E \int_{-\infty}^{\infty} (|X_{n,k}| + |t|) K_k(t) dt + 4.21\beta.$$

Since $\int_{-\infty}^{\infty} K_k(t) dt = EX_{n,k}^2$ and $\int_{-\infty}^{\infty} |t| K_k(t) dt = \frac{1}{2} E|X_{n,k}|^3$, we have

$$\begin{aligned} F_n(x) - \Phi(x) &\le 1.63 \sum_{k=1}^{\infty} (E|X_{n,k}| EX_{n,k}^2 + \frac{1}{2} E|X_{n,k}|^3) + 4.21\beta \\ &\le 1.63 \sum_{k=1}^{\infty} E|X_{n,k}|^3 + 0.815 \sum_{k=1}^{\infty} E|X_{n,k}|^3 + 4.21\beta \\ &= 2.445 \sum_{k=1}^{\infty} E|X_{n,k}|^3 + 4.21\beta \\ &\le 6.655\beta. \end{aligned}$$

Similarly, we use (2.12) to get,

$$F_n(x) - \Phi(x) \geq -6.655\beta.$$

Then

$$|F_n(x) - \Phi(x)| \leq 43.7\beta \leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}}.$$

**Case 2** $x < 0$.

Since $P(W \leq x) = P(\bigcap_{n=1}^{\infty}\{W < x + \frac{1}{n}\}) = \lim_{n\to\infty} P(W < x + \frac{1}{n})$,

$P(-W < -x) = P(\bigcup_{n=1}^{\infty}\{-W \leq -x - \frac{1}{n}\}) = \lim_{n\to\infty} P(-W \leq -x - \frac{1}{n})$ and

$\Phi(x) = \lim_{n\to\infty} \Phi(x + \frac{1}{n})$, we have

$$\begin{aligned}
|P(W \leq x) - \Phi(x)| &= |\lim_{n\to\infty}[P(W < x + \frac{1}{n}) - \Phi(x + \frac{1}{n})]| \\
&= |\lim_{n\to\infty}[\{1 - \Phi(x + \frac{1}{n})\} - \{1 - P(W < x + \frac{1}{n})\}]| \\
&= |\lim_{n\to\infty}[\Phi(-x - \frac{1}{n}) - P(W \geq x + \frac{1}{n})]| \\
&= |\lim_{n\to\infty}[\Phi(-x - \frac{1}{n}) - P(-W \leq -x - \frac{1}{n})]| \\
&= |\Phi(-x) - P(-W < -x)| \\
&\leq |P(-W \leq -x) - \Phi(-x)| \\
&\leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}} \quad \text{(by case 1).}
\end{aligned}$$

Therefore,

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}}.$$

□

# CHAPTER IV

# A NON-UNIFORM BOUND ON AN INFINITE URN MODEL

In this chapter, we give a non-uniform bound in normal approximation of $\dfrac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}$ which is stated in Theorem 4.4. To prove Theorem 4.4, we need the following results.

Throughout this chapter, $C$ stands for an absolute constant with possibly different values in different places.

**Proposition 4.1.** *(Rosenthal inequality, [7], p.59) Let $p \geq 2$ and let $X_1, X_2, \ldots, X_n$ be independent random variables such that $EX_i = 0,\ E|X_i|^p < \infty$. Then there exists $C(p)$ such that*

$$E|\sum_{i=1}^{n} X_i|^p \leq C(p)\{\sum_{i=1}^{n} E|X_i|^p + (\sum_{i=1}^{n} EX_i^2)^{\frac{p}{2}}\}$$

*where $C(p)$ is a positive constant depending only on $p$.*

In the following, we let

$$W := \sum_{k=1}^{\infty} X_{n,k},\ \ \beta := \sum_{k=1}^{\infty} E|X_{n,k}|^3 \text{ and } \delta_a := \frac{\beta}{(1+a)^3} \text{ for } a \geq 0.$$

By (2.6), we have

$$W = \frac{Z_{N(n)} - E(Z_{N(n)})}{\sqrt{Var(Z_{N(n)})}}.$$

**Proposition 4.2.** (*concentration inequality for non-uniform bound*) *For any* $0 \leq a < b < \infty$*, we have*

$$P(a \leq W^{(k)} \leq b) \leq C\{\frac{b-a}{(1+a)^3} + \delta_a\}$$

*where* $C$ *is a positive constant.*

*Proof.* We devide the proof into 2 cases

$$\textbf{Case 1} \qquad (1+a)\beta \geq \frac{1}{64}. \tag{4.1}$$

By the Rosenthal inequality, we have

$$\begin{aligned}
E(W^{(k)})^4 &= E(\sum_{\substack{j=1\\ j\neq k}}^{\infty} X_{n,j})^4 \\
&= \int_{\Omega} \lim_{m\to\infty} (\sum_{\substack{j=1\\ j\neq k}}^{m} X_{n,j})^4 dP \\
&= \lim_{m\to\infty} \int_{\Omega} (\sum_{\substack{j=1\\ j\neq k}}^{m} X_{n,j})^4 dP \\
&= \lim_{m\to\infty} E(\sum_{\substack{j=1\\ j\neq k}}^{m} X_{n,j})^4 \\
&\leq \lim_{m\to\infty} C\{\sum_{\substack{j=1\\ j\neq k}}^{m} E|X_{n,j}|^4 + (\sum_{\substack{j=1\\ j\neq k}}^{m} EX_{n,j}^2)^2\} \\
&= C\{\sum_{\substack{j=1\\ j\neq k}}^{\infty} E|X_{n,j}|^4 + (\sum_{\substack{j=1\\ j\neq k}}^{\infty} EX_{n,j}^2)^2\} \qquad (4.2) \\
&\leq C\{\sum_{\substack{j=1\\ j\neq k}}^{\infty} E|X_{n,j}|^4 + 1\}. \qquad (4.3)
\end{aligned}$$

and hence

$$P(a \leq W^{(k)} \leq b) \leq P(W^{(k)} \geq a)$$

$$
\begin{aligned}
&= P(a+1 \leq W^{(k)}+1) \\
&\leq E\frac{|1+W^{(k)}|^4}{(1+a)^4} \\
&\leq \frac{8+8E|W^{(k)}|^4}{(1+a)^4} \\
&\leq \frac{8}{(1+a)^4} + \frac{C}{(1+a)^4}\{\sum_{\substack{j=1\\ j\neq k}}^{\infty} E|X_{n,j}|^4 + 1\} \quad \text{(by (4.3))} \\
&\leq \frac{C}{(1+a)^4} + \frac{C}{(1+a)^4}\frac{1}{\sqrt{Var(Z_{N(n)})}}\sum_{j=1}^{\infty} E|X_{n,j}|^3 \quad \text{(by (2.5))} \\
&\leq \frac{C}{(1+a)^4} + \frac{C\beta}{(1+a)^3\sqrt{Var(Z_{N(n)})}} \\
&\leq \frac{C\beta}{(1+a)^3} + \frac{C\beta}{(1+a)^3\sqrt{Var(Z_{N(n)})}} \quad \text{(by (4.1))} \\
&= \frac{C\beta}{(1+a)^3} \qquad (4.4)
\end{aligned}
$$

**Case 2** $\quad (1+a)\beta < \frac{1}{64}. \qquad (4.5)$

Let $\kappa = 16\beta$, $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$
f(x) = \begin{cases} 0 & \text{for } \; x < a-\kappa, \\ (1+x+\kappa)^3(x-a+\kappa) & \text{for } \; a-\kappa \leq x \leq b+\kappa, \\ (1+x+\kappa)^3(b-a+2\kappa) & \text{for } \; x > b+\kappa, \end{cases}
$$

and $M :\ \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$
M(\omega, t) = \omega I(-\omega \leq t \leq 0) - \omega I(0 < t \leq -\omega).
$$

Note that $f$ is a non-decreasing function satisfying

$$
f'(x) \geq \begin{cases} (1+a)^3 & \text{for } \; a-\kappa < x < b+\kappa, \\ 0 & \text{otherwise.} \end{cases} \qquad (4.6)
$$

and

$$\kappa < \frac{1}{4(1+a)}. \tag{4.7}$$

$$\begin{aligned}
&E\{W^{(k)}f(W^{(k)})\} \\
&= \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{\int_{-\infty}^{\infty} f'(W^{(k)}+t)M(X_{n,j},t)dt\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(using the same argument of (3.3))} \\
&\geq \sum_{\substack{j=1 \\ j\neq k}}^{\infty}(1+a)^3 E\{I(a \leq W^{(k)} \leq b)\int_{|t|\leq\kappa} M(X_{n,j},t)dt\} \quad \text{(by (4.6))} \\
&= (1+a)^3 E\{I(a \leq W^{(k)} \leq b)\sum_{\substack{j=1 \\ j\neq k}}^{\infty} |X_{n,j}|\min(\kappa,|X_{n,j}|)\} \\
&= (1+a)^3 E\{I(a \leq W^{(k)} \leq b)\sum_{\substack{j=1 \\ j\neq k}}^{\infty} \eta_j\} \\
&\geq 0.5(1+a)^3\{P(a \leq W^{(k)} \leq b) - P(U \leq 0.5)\} \qquad (4.8)
\end{aligned}$$

where $\eta_j = |X_{n,j}|\min(\kappa,|X_{n,j}|),\ U = \sum_{\substack{j=1 \\ j\neq k}}^{\infty} \eta_j$ and we have use the fact that for $a < b,\ y \geq 0$ and $c > 0$,

$$I(a \leq \omega \leq b)y \geq c[I(a \leq \omega \leq b) - (1-\frac{y}{c})I(y \leq c))]$$

in the last inequality (see [5], p.238). By (3.5) and the fact that

$$EX_{n,k}^2 \leq (E|X_{n,k}|^3)^{\frac{2}{3}} \leq (\sum_{k=1}^{\infty} E|X_{n,k}|^3)^{\frac{2}{3}} = \beta^{\frac{3}{2}} \leq \frac{1}{16},$$

we have

$$EU = E\sum_{\substack{j=1 \\ j\neq k}}^{\infty} |X_{n,j}|\min(\kappa,|X_{n,j}|)$$

$$\geq E \sum_{\substack{j=1 \\ j\neq k}}^{\infty} (X_{n,j}^2 - \frac{|X_{n,j}|^3}{4\kappa})$$

$$= E \sum_{j=1}^{\infty} X_{n,j}^2 - EX_{n,k}^2 - E \sum_{\substack{j=1 \\ j\neq k}}^{\infty} \frac{|X_{n,j}|^3}{4\kappa}$$

$$\geq 1 - \frac{1}{16} - \frac{\beta}{4\kappa}$$

$$> 0.75.$$

Using the same argument in (4.2),

$$P(U \leq 0.5)$$

$$\leq P(EU - U \geq 0.75 - 0.5)$$

$$\leq \frac{E|U - EU|^4}{(0.25)^4}$$

$$= \frac{E| \sum_{\substack{j=1 \\ j\neq k}}^{\infty} \{|X_{n,j}| \min(\kappa, |X_{n,j}|) - E|X_{n,j}| \min(\kappa, |X_{n,j}|)\}|^4}{(0.25)^4}$$

$$\leq C\Big\{ \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{|X_{n,j}| \min(\kappa, |X_{n,j}|) - E|X_{n,j}| \min(\kappa, |X_{n,j}|)\}^4$$

$$+ \big( \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{|X_{n,j}| \min(\kappa, |X_{n,j}|) - E|X_{n,j}| \min(\kappa, |X_{n,j}|)\}^2 \big)^2 \Big\}$$

$$\leq C\Big\{ \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{|X_{n,j}| \min(\kappa, |X_{n,j}|)\}^4 + \big( \sum_{\substack{j=1 \\ j\neq k}}^{\infty} E\{|X_{n,j}| \min(\kappa, |X_{n,j}|)\}^2 \big)^2 \Big\}$$

$$\leq C\{\kappa^4 \sum_{j=1}^{\infty} EX_{n,j}^4 + \kappa^4\}$$

$$\leq C\{\kappa^4 \frac{1}{\sqrt{Var(Z_{N(n)})}} \sum_{j=1}^{\infty} E|X_{n,j}|^3 + \kappa^4\} \quad \text{(by (2.5))}$$

$$\leq \frac{C\beta}{4^4(1+a)^4\sqrt{Var(Z_{N(n)})}} + C\kappa^3\kappa \quad \text{(by (4.7))}$$

$$\le \frac{C\beta}{(1+a)^3\sqrt{Var(Z_{N(n)})}} + \frac{C\beta}{(1+a)^3} \quad \text{(by (4.7) and } \kappa = 16\beta)$$
$$= \frac{C\beta}{(1+a)^3}.$$

Combining this with (4.8), we have

$$
\begin{aligned}
&P(a \le W^{(k)} \le b)\\
&\le P(U \le 0.5) + \frac{2E\{W^{(k)}f(W^{(k)})\}}{(1+a)^3}\\
&\le \frac{C\beta}{(1+a)^3} + \frac{2E\{W^{(k)}f(W^{(k)})\}}{(1+a)^3}\\
&\le C(1+a)^{-3}\{\beta + (b-a+2\kappa)E|W^{(k)}(1+W^{(k)}+\kappa)^3|\}\\
&\le C(1+a)^{-3}\{\beta + (b-a+\kappa)E|W^{(k)}((1+\kappa)^3+|W^{(k)}|^3)|\}\\
&\le C(1+a)^{-3}\{\beta + (b-a+\kappa)E|W^{(k)}(1+\kappa)^3+(W^{(k)})^4|\}\\
&\le C(1+a)^{-3}\{\beta + (b-a+\kappa)(E|W^{(k)}|+E(W^{(k)})^4)\}\\
&\le C(1+a)^{-3}\{\beta + (b-a+\kappa)(E|W^{(k)}|+C(\sum_{\substack{j=1\\ j\ne k}}^{\infty} E|X_{n,j}|^4+1))\} \quad \text{(by (4.3))}\\
&\le C(1+a)^{-3}\Big\{\beta + (b-a+\kappa)(E|W^{(k)}|\\
&\qquad\qquad + C(\frac{1}{\sqrt{Var(Z_{N(n)})}}\sum_{j=1}^{\infty} E|X_{n,j}|^3+1))\Big\} \quad \text{(by (2.5))}\\
&\le C(1+a)^{-3}\{\beta + b - a + \kappa\} \quad \text{(by (4.5) and } E|W^{(k)}| \le 1)\\
&\le C(1+a)^{-3}\{\beta + b - a\} \quad (\kappa = 16\beta). \qquad (4.9)
\end{aligned}
$$

Hence, by (4.4) and (4.9), the concentration inequality is proved. □

**Proposition 4.3.** ([5], p.248, 250) *Let $g(\omega) = (wf_x(\omega))'$. Then the followings hold.*

1. $f_x'(\omega+s) - f_x'(\omega+t) \le \int_t^s g(\omega+u)du + I(x - \max(s,t) < \omega \le x - \min(s,t)).$

2. *For* $x \geq 4$ *and* $|u| \leq 1 + \dfrac{x}{4}$, *we have*

$$E(g(W^{(k)}) + u) \leq C((1+x)^{-3} + x\delta_{\frac{x}{4}}),$$

*where* $f_x$ *is given by* (2.9).

We are now ready to prove the main result of this chapter.

**Theorem 4.4.** *There exists an absolute constant* $C$ *such that for every real number* $x$,

$$|F_n(x) - \Phi(x)| \leq \frac{C}{(1+|x|)^3\sqrt{Var(Z_{N(n)})}}.$$

*Moreover, under the additional condition that* $\lim\limits_{n\to\infty} Var(Z_{N(n)}) = \infty$, *the bound tends to zero.*

*Proof.* By the same arguments on Theorem 3.5, it suffice to prove the theorem in case of $x \geq 0$.

**Case 1.** $0 \leq x \leq 4$.

Note that $(1+|x|)^3 \leq 125$ so $1 \leq \dfrac{125}{(1+|x|)^3}$. By Theorem 3.5,

$$|F_n(x) - \Phi(x)| \leq \frac{6.655}{\sqrt{Var(Z_{N(n)})}} \cdot \frac{125}{(1+x)^3} \leq \frac{C}{(1+x)^3\sqrt{Var(Z_{N(n)})}}.$$

**Case 2.** $x \geq 4$.

For $\omega > 0$, we note from [6] on page 23 that

$$1 - \Phi(\omega) \leq \frac{e^{-\frac{1}{2}\omega^2}}{\sqrt{2\pi}\omega}.$$

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(\omega) = \frac{e^{-\frac{1}{2}\omega^2}}{\sqrt{2\pi}\omega}(1+\omega)^4$$

then

$$f'(\omega) = -\frac{1}{\sqrt{2\pi}\omega^2} e^{-\frac{1}{2}\omega^2}(1+\omega)^3(\omega-1)(\omega^2+2\omega-1) < 0$$

for all $\omega \geq 4$. It follows that for each $\omega \geq 4$,

$$\frac{e^{-\frac{1}{2}\omega^2}}{\sqrt{2\pi}\omega}(1+\omega)^4 = f(\omega) \leq f(4) = \frac{156.25e^{-8}}{\sqrt{2\pi}}$$

and hence

$$1 - \Phi(x) \leq \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}x} \leq \frac{C}{(1+x)^4}.$$

If $(1+x)\beta \geq \dfrac{1}{64}$, then, by the same arguments of (4.4), we have

$$P(W \geq x) \leq \frac{C\beta}{(1+x)^3}$$

which implies

$$\begin{aligned}
|F_n(x) - \Phi(x)| &\leq P(W \geq x) + 1 - \Phi(x) \\
&\leq \frac{C\beta}{(1+x)^3} + \frac{C}{(1+x)^4} \\
&\leq \frac{C\beta}{(1+x)^3} + \frac{C\beta}{(1+x)^3} \quad (\text{because } \ \frac{1}{(1+x)^4} \leq \frac{64\beta}{(1+x)^3}) \\
&\leq \frac{C}{(1+x)^3\sqrt{Var(Z_{N(n)})}} \quad (\text{by Proposition 2.10(1)}).
\end{aligned}$$

Assume that

$$(1+x)\beta < \frac{1}{64}. \tag{4.10}$$

By noting that

$$\delta_{\frac{x}{4}} = \frac{(1+x)^3}{(1+\frac{x}{4})^3}\delta_x \leq C\delta_x \leq \frac{C}{(1+|x|)^3\sqrt{Var(Z_{N(n)})}} \quad \text{for all } x \geq 4,$$

it suffice to show that

$$|F_n(x) - \Phi(x)| \le C\delta_{\frac{x}{4}}. \tag{4.11}$$

Let

$$\overline{X}_{n,k,\frac{x}{4}} = X_{n,k} I(|X_{n,k}| \le 1 + \frac{x}{4})$$

and

$$K_{k,\frac{x}{4}}(t) = E\{\overline{X}_{n,k,\frac{x}{4}}[I(0 \le t \le \overline{X}_{n,k,\frac{x}{4}}) - I(\overline{X}_{n,k,\frac{x}{4}} \le t < 0)]\}$$

Note that $\quad |\overline{X}_{n,k,\frac{x}{4}}| \le |X_{n,k}|$ and $\displaystyle\int_{|t|>1+\frac{x}{4}} K_{k,\frac{x}{4}}(t)dt = 0.$

Let $f_x$ be the unique solution of the Stein's equation for $h_x$ which given by (2.8) and (2.9). Using the same arguments as for (3.10), we have

$$\begin{aligned}
&F_n(x) - \Phi(x) \\
&= \sum_{k=1}^{\infty} E\Big\{ I(|X_{n,k}| \le 1 + \frac{x}{4}) \\
&\qquad\qquad \int_{|t|\le 1+\frac{x}{4}} [f_x'(W^{(k)} + X_{n,k}) - f_x'(W^{(k)} + t)] K_{k,\frac{x}{4}}(t)dt \Big\} \\
&\quad + \sum_{k=1}^{\infty} E\Big\{ I(|X_{n,k}| > 1 + \frac{x}{4}) \\
&\qquad\qquad \int_{|t|\le 1+\frac{x}{4}} [f_x'(W^{(k)} + X_{n,k}) - f_x'(W^{(k)} + t)] K_{k,\frac{x}{4}}(t)dt \Big\} \\
&:= R_1 + R_2.
\end{aligned}$$

Then

$$\begin{aligned}
&|R_1| \\
&\le \sum_{k=1}^{\infty} \Big| E\Big\{ I(|X_{n,k}| \le 1 + \frac{x}{4}) \\
&\quad \times \int_{|t|\le 1+\frac{x}{4}} \big[f_x'(W^{(k)} + X_{n,k}) - f_x'(W^{(k)} + t)\big] K_{k,\frac{x}{4}}(t)dt \Big\} \Big|
\end{aligned}$$

$$\leq \sum_{k=1}^{\infty}\Big|E\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})\int_{|t|\leq 1+\frac{x}{4}}\big[\int_{t}^{X_{n,k}}g(W^{(k)}+u)du$$

$$+I\big(x-\max(t,X_{n,k})\leq W^{(k)}\leq x-\min(t,X_{n,k})\big]K_{k,\frac{x}{4}}(t)dt\Big\}\Big|$$

(by Proposition 4.3(1))

$$\leq \sum_{k=1}^{\infty}\Big|E\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})\int_{|t|\leq 1+\frac{x}{4}}K_{k,\frac{x}{4}}(t)\int_{t}^{X_{n,k}}g(W^{(k)}+u)dudt\Big\}\Big|$$

$$+\sum_{k=1}^{\infty}\Big|E\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})$$

$$\times\int_{|t|\leq 1+\frac{x}{4}}K_{k,\frac{x}{4}}(t)I\big(x-\max(t,X_{n,k})\leq W^{(k)}\leq x-\min(t,X_{n,k})\big)dt\Big\}\Big|$$

$$\leq \sum_{k=1}^{\infty}\Big|E\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})\int_{|t|\leq 1+\frac{x}{4}}K_{k,\frac{x}{4}}(t)\int_{t}^{X_{n,k}}g(W^{(k)}+u)dudt\Big\}\Big|$$

$$+\sum_{k=1}^{\infty}E\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})$$

$$\times\int_{|t|\leq 1+\frac{x}{4}}K_{k,\frac{x}{4}}(t)I\big(x-\max(t,X_{n,k})\leq W^{(k)}\leq x-\min(t,X_{n,k})\big)dt\Big\}$$

$$=\sum_{k=1}^{\infty}\Big|\int_{|t|\leq 1+\frac{x}{4}}\int_{\mathbb{R}}E\Big\{K_{k,\frac{x}{4}}(t)I(t\leq u\leq X_{n,k})I(|X_{n,k}|\leq 1+\frac{x}{4})g(W^{(k)}+u)\Big\}dudt\Big|$$

$$+\sum_{k=1}^{\infty}\int_{|t|\leq 1+\frac{x}{4}}E\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})$$

$$\times K_{k,\frac{x}{4}}(t)I\big(x-\max(t,X_{n,k})\leq W^{(k)}\leq x-\min(t,X_{n,k})\big)\Big\}dt$$

$$=\sum_{k=1}^{\infty}\Big|\int_{|t|\leq 1+\frac{x}{4}}\int_{\mathbb{R}}EE^{X_{n,k}}\Big\{K_{k,\frac{x}{4}}(t)I(t\leq u\leq X_{n,k})I(|X_{n,k}|\leq 1+\frac{x}{4})g(W^{(k)}+u)\Big\}dudt\Big|$$

$$+\sum_{k=1}^{\infty}\int_{|t|\leq 1+\frac{x}{4}}EE^{X_{n,k}}\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})$$

$$\times K_{k,\frac{x}{4}}(t)I\big(x-\max(t,X_{n,k})\leq W^{(k)}\leq x-\min(t,X_{n,k})\big)\Big\}dt$$

$$=\sum_{k=1}^{\infty}\Big|\int_{|t|\leq 1+\frac{x}{4}}\int_{\mathbb{R}}E\Big\{K_{k,\frac{x}{4}}(t)I(t\leq u\leq X_{n,k})I(|X_{n,k}|\leq 1+\frac{x}{4})E^{X_{n,k}}g(W^{(k)}+u)\Big\}dudt\Big|$$

$$+\sum_{k=1}^{\infty}\int_{|t|\leq 1+\frac{x}{4}}E\Big\{I(|X_{n,k}|\leq 1+\frac{x}{4})$$

$$\times K_{k,\frac{x}{4}}(t)E^{X_{n,k}}I\big(x-\max(t,X_{n,k})\leq W^{(k)}\leq x-\min(t,X_{n,k})\big)\Big\}dt$$

$$
\begin{aligned}
&= \sum_{k=1}^{\infty} \Big| E\Big\{ I(|X_{n,k}| \le 1 + \frac{x}{4} \int_{|t| \le 1+\frac{x}{4}} \int_{\mathbb{R}} K_{k,\frac{x}{4}}(t) I(t \le u \le X_{n,k}) Eg(W^{(k)} + u) dudt \Big\} \Big| \\
&\quad + \sum_{k=1}^{\infty} E\Big\{ I(|X_{n,k}| \le 1 + \frac{x}{4}) \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t) \\
&\quad \times E^{X_{n,k}} I\big(x - \max(t, X_{n,k}) \le W^{(k)} \le x - \min(t, X_{n,k})\big) dt \Big\} \\
&= \sum_{k=1}^{\infty} \Big| E\Big\{ I(|X_{n,k}| \le 1 + \frac{x}{4}) \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t) \int_{t}^{X_{n,k}} Eg(W^{(k)} + u) dudt \Big\} \Big| \\
&\quad + \sum_{k=1}^{\infty} E\Big\{ I(|X_{n,k}| \le 1 + \frac{x}{4}) \\
&\quad \times \int_{|t| \le 1+\frac{x}{4}} P\big(x - \max(t, X_{n,k}) \le W^{(k)} \le x - \min(t, X_{n,k}) | X_{n,k}\big) K_{k,\frac{x}{4}}(t) dt \Big\} \\
&:= R_{11} + R_{12}.
\end{aligned}
\tag{4.12}
$$

By Proposition 4.3(2), we have

$$
\begin{aligned}
& R_{11} \\
&\le C \sum_{k=1}^{\infty} \Big| E\{ I(|X_{n,k}| \le 1 + \frac{x}{4}) \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t) \int_{t}^{X_{n,k}} ((1+x)^{-3} + x\delta_{\frac{x}{4}}) dudt \} \Big| \\
&\le C((1+x)^{-3} + x\delta_{\frac{x}{4}}) \sum_{k=1}^{\infty} \Big| E\{ \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t) \int_{t}^{X_{n,k}} dudt \} \Big| \\
&\le C((1+x)^{-3} + x\delta_{\frac{x}{4}}) \sum_{k=1}^{\infty} E \Big| \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t)(X_{n,k} - t) dt \Big| \\
&\le C((1+x)^{-3} + x\delta_{\frac{x}{4}}) \sum_{k=1}^{\infty} E \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t)|X_{n,k}| + K_{k,\frac{x}{4}}(t)|t| dt \\
&= C((1+x)^{-3} + x\delta_{\frac{x}{4}}) \Big\{ \sum_{k=1}^{\infty} E \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t)|X_{n,k}| dt \\
&\qquad\qquad + \sum_{k=1}^{\infty} E \int_{|t| \le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t)|t| dt \Big\} \\
&= C((1+x)^{-3} + x\delta_{\frac{x}{4}}) \Big\{ \sum_{k=1}^{\infty} E(|X_{n,k}| E\overline{X}^{2}_{n,k,\frac{x}{4}}) + \frac{1}{2} \sum_{k=1}^{\infty} E|\overline{X}_{n,k,\frac{x}{4}}|^3 \Big\} \\
&\le C((1+x)^{-3} + x\delta_{\frac{x}{4}}) \Big\{ \sum_{k=1}^{\infty} E(|X_{n,k}| EX^{2}_{n,k}) + \frac{1}{2} \sum_{k=1}^{\infty} E|X_{n,k}|^3 \Big\}
\end{aligned}
$$

$$\le C((1+x)^{-3} + x\delta_{\frac{x}{4}})\Big\{\sum_{k=1}^{\infty} E|X_{n,k}|^3 + \frac{1}{2}\sum_{k=1}^{\infty} E|X_{n,k}|^3\Big\}$$

$$= C((1+x)^{-3} + x\delta_{\frac{x}{4}})\sum_{k=1}^{\infty} E|X_{n,k}|^3 \tag{4.13}$$

$$= C(\frac{\beta}{(1+x)^3} + \frac{x\beta^2}{(1+\frac{x}{4})^3})$$

$$\le C(\frac{\beta}{(1+\frac{x}{4})^3} + \frac{x\beta^2}{(1+\frac{x}{4})^3})$$

$$\le C(\frac{\beta}{(1+\frac{x}{4})^3}) \ \ (\text{by (4.10), we get } 1+x\beta < \frac{65}{64})$$

$$\le C\delta_{\frac{x}{4}}. \tag{4.14}$$

By Proposition 4.2, $\delta_x$ is decreasing in $x$ and the fact that

$$x - \max(t, X_{n,k}) \ge x - (1+\frac{x}{4}) \ \text{ for } |t| \le 1+\frac{x}{4} \text{ and } |X_{n,k}| \le 1+\frac{x}{4},$$

we have

$$R_{12}$$

$$\le \sum_{k=1}^{\infty} E\{\int_{|t|\le 1+\frac{x}{4}} P\big(x - \max(t, X_{n,k}) \le$$

$$W^{(k)} \le x - \min(t, X_{n,k})\big)K_{k,\frac{x}{4}}(t)dt\}$$

$$\le C\sum_{k=1}^{\infty} E\{\int_{|t|\le 1+\frac{x}{4}} [\delta_{x-\max(t,X_{n,k})} +$$

$$(1+x-\max(t,X_{n,k}))^{-3}(|t|+|X_{n,k}|)]K_{k,\frac{x}{4}}(t)dt\}$$

$$\le C\sum_{k=1}^{\infty} E\{\int_{|t|\le 1+\frac{x}{4}} [\delta_{\frac{x}{4}} + (1+x)^{-3}(|t|+|X_{n,k}|)]K_{k,\frac{x}{4}}(t)dt\}$$

$$= C\Big\{\sum_{k=1}^{\infty} E\{\int_{|t|\le 1+\frac{x}{4}} \delta_{\frac{x}{4}} K_{k,\frac{x}{4}}(t)dt\}$$

$$+ (1+x)^{-3}\sum_{k=1}^{\infty} E\{\int_{|t|\le 1+\frac{x}{4}} (|t|+|X_{n,k}|))K_{k,\frac{x}{4}}(t)dt\}\Big\}$$

$$\le C\Big\{\delta_{\frac{x}{4}} + (1+x)^{-3}\sum_{k=1}^{\infty} E\{\int_{|t|\le 1+\frac{x}{4}} (|t| + |X_{n,k}|))K_{k,\frac{x}{4}}(t)dt\}\Big\}$$

$$\le C\Big\{\delta_{\frac{x}{4}} + (1+x)^{-3}\sum_{k=1}^{\infty} E|X_{n,k}|^3\Big\} \text{ (using the same argument of (4.13))}$$

$$\le C\delta_{\frac{x}{4}} \text{ (because } \delta_x \le \delta_{\frac{x}{4}}). \tag{4.15}$$

Next, by (2.13),

$$\begin{aligned}
|R_2| &\le \sum_{k=1}^{\infty} E\Big\{I(|X_{n,k}| > 1 + \frac{x}{4}) \int_{|t|\le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t)dt\Big\} \\
&\le \sum_{k=1}^{\infty} E\Big\{I(|X_{n,k}| > 1 + \frac{x}{4})\Big\} \quad (\int_{|t|\le 1+\frac{x}{4}} K_{k,\frac{x}{4}}(t)dt \le 1) \\
&= \sum_{k=1}^{\infty} P(I(|X_{n,k}| > 1 + \frac{x}{4}) \\
&\le \frac{\sum_{k=1}^{\infty} E|X_{n,k}|^3}{(1+\frac{x}{4})^3} \\
&= \delta_{\frac{x}{4}}.
\end{aligned} \tag{4.16}$$

By (4.11), (4.12) and (4.14)-(4.16), we have

$$|F_n(x) - \Phi(x)| \le \frac{C}{(1+|x|)^3\sqrt{Var(Z_{N(n)})}}.$$

□

# REFERENCES


1. Dutko, M. Central limit theorems for infinite urn models. *Annals of Prob.* 17,3(1989) : 1255-1263.

2. De Barra, G. *Measure Theory and Integration.* Chichester: John Wiley & son, 1981.

3. Laha, R.G. and Rohatgi, V.K. *Probability Theory.* New York: John Wiley & son, 1979.

4. กฤษณะ เนียมมณี. *ทฤษฎีความน่าจะเป็นขั้นสูงและขอบเขตการประมาณค่า.* พิทักษ์การพิมพ์, 2548.

5. Chen, L.H.Y and Shao, Q.M. A non-uniform Berry-Esseen bound, via Stein's method. *Probab. Theory Relat. Fields.* 120(2001) : 236-254.

6. Stein, C. *Approximation Computation of Expectations.* Lecture Notes 7. California : Inst. Math. Statist, 1986.

7. Petrov, V.V. *Limit Theorem of Probability Theory:Sequences of Indepen dent Random Variables.* Clarendon Press, 1995.

8. Karlin, S. Central limit theorems for certain infinite urn schemes. *S. Math. Mech.* 17(1967) : 373-401.

9. Stein, C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, Proc, Sixth Berkeley Symp. *Math. Stat. Prob.* 2(1972) : 583-602.

# VITA

Name : Mr. Soontorn Boonta

Degree : B.Sc. (Mathematics) (2nd Class Degree Honours), 2002,

Khon Kaen University, Thailand.