CHAPTER I

Introduction

Let (X_{ij}) be an $n \times n (n \geq 20)$ matrix of independent random variables with finite third moment and $\pi = (\pi(1), \pi(2), ..., \pi(n))$ be a random permutation of $\{1, 2, ..., n\}$ such that π and $X_{ij}^{'}s$ are independent. This work is concerned with the normal approximation to the distribution function of $W_n = \sum_{i=1}^{n} X_{i\pi(i)}$. The special cases of W_n are the statistics $\eta_n = \sum_{i=1}^{n} a_i b_{i\pi(i)}$ and $\xi_n = \sum_{i=1}^{n} c_{i\pi(i)}$ where a_i, b_{ij} and c_{ij} $(i, j = 1, 2, ..., n)$ are real numbers. Both statistics η_n and ξ_n arise in permutation tests in nonparametric inference (see,for examples, Fraser(1957), Puri and Sen(1971), Does(1982)). The literature concerning the limit behavior of W_n dates back to 1944 when Wald and Wolfowitz(1944) first established the asymptotic normality of η_n with some strong sufficient conditions. After that, a theorem which has been proved under various conditions by Hoeffding(1951), Matoo(1957), Hájek(1961), Robinson(1972), Kolchim and Chistyakov (1973), Ho and Chen(1978), Does(1982), Bolthausen(1984), Schneller(1988) and Loh(1996). Most of the literatures gave a bound when X_{ij} are constants and the best bound of order $\frac{1}{n}$ is given by Bolthausen(1984). In the case when $X_{ij}^{'}s$ are any random variables, the estimations have been obtained by Von Bahr(1976) and Ho and Chen (1978) but they yield the convergence rate $\frac{1}{\sqrt{n}}$ only under some boundedness condition like $\sup_{i,j}|X_{ij}| = O(n^{-\frac{1}{2}})$. But in this work we give the rate $\frac{1}{n}$ by using Stein's method.

Stein(1972) originally introduced his method for obtainning rates of convergence in a central limit theorem for sums of nearly independent random variables. There are at least 3 approaches to use Stein's method when the limit distribution is normal, i.e. concentration inequality approach(see for examples, Ho and Chen(1978) and Chen and Shao(2001)), inductive approach (see for example,Bolthausen(1984)) and coupling approach (see for example, Stein(1986)). In this work we use concentration inequality and coupling approaches.

For each $i, j \in \{1, 2, ..., n\}$, let μ_{ij} and σ_{ij}^2 be the expected value and variance of X_{ij}, respectively and

$$\begin{aligned} \mu_{i.} &= \frac{1}{n}\sum_{j}\mu_{ij}\ , \quad \mu_{.j} = \frac{1}{n}\sum_{i}\mu_{ij}, \quad \mu_{..} = \frac{1}{n^2}\sum_{i,j}\mu_{ij} \\ d^2 &= \frac{1}{(n-1)}\sum_{i,j}(\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..})^2 \quad \text{and } \sigma^2 = \frac{1}{n}\sum_{i,j}\sigma_{ij}^2. \end{aligned}$$

From Ho and Chen(1978) we know that $VarW_n = d^2 + \sigma^2$.

Define

$$W = \frac{1}{\sqrt{d^2+\sigma^2}}\sum_{i}(X_{i\pi(i)} - \mu_{..}).$$

So $EW = 0, VarW = 1$ and

$$W = \frac{W_n - EW_n}{\sqrt{VarW_n}} = \sum_{i=1}^{n}\hat{X}_{i\pi(i)}$$

where $\hat{X}_{ij} = \frac{1}{\sqrt{d^2+\sigma^2}}(X_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..}).$

Our main result is the following.

Theorem

$$\sup_{\mathbb{R}} |F_n(x) - \Phi(x)| \leq 210\beta + \frac{18}{n}$$

where $\beta = \frac{1}{n} \sum_{i,j} E|\hat{X}_{ij}|^3$, F_n is the distribution function of $\frac{W_n - EW_n}{\sqrt{VarW_n}}$ and Φ is the standard normal distribution function.