THE RATE OF CONVERGENCE IN A CENTRAL LIMIT THEOREM FOR DEPENDENT RANDOM VARIABLES WITH ARBITRARY INDEX SET

(Abbreviated title: RATE OF CONVERGENCE IN A CLT)

BY LOUIS H. Y. CHEN

National University of Singapore

Institute for Mathematics and Its Applications
University of Minnesota

and

Mathematics Research Center
University of Wisconsin-Madison

May 1986

SUMMARY

The classical rate of convergence in the central limit theorem for independent random variables with finite third moments has been found to be too crude for dependent random variables in general. This paper has two objectives. The first is to find a generalization of the classical rate of convergence. The second is to apply Stein's method and in doing so develop general techniques in the application.

We obtain results for a class of dependent random variables with arbitrary index set which includes m-dependent random fields as special cases and is relevant to spatial statistics. ~~Our results~~ can easily be extended to mixing random fields and in particular certain Gibbs random fields.

Research partially sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

AMS 1980 Subject Classifications: Primary 60F05, 60G60.

Key Words and Phrases: Central limit theorem, rate of convergence, normal approximation, Stein's method, exchangeability argument, conditional concentration inequality, dependent random variables with arbitrary index set, mixing random field, m-dependent random field, spatial statistic.

THE RATE OF CONVERGENCE IN A CENTRAL LIMIT THEOREM FOR DEPENDENT RANDOM VARIABLES WITH ARBITRARY INDEX SET

BY LOUIS H. Y. CHEN

National University of Singapore

Institute for Mathematics and Its Applications
University of Minnesota

and

Mathematics Research Center
University of Wisconsin-Madison

1. Introduction. Let $X_1,\ldots,X_n$ be random variables with $EX_i = 0$, $\mathrm{Var}(\sum_{i=1}^{n} X_i) = B_n^2$ and $E|X_i|^3 = \gamma_i < \infty$. Let F_n be the distribution function of $B_n^{-1} \sum_{i=1}^{n} X_i$ and Φ the standard normal distribution function. The Berry-Esseen theorem (Berry (1941) and Esseen (1945)) states that if $X_1,\ldots,X_n$ are independent then there exists an absolute constant C such that

$$\|F_n - \Phi\|_\infty < CB_n^{-3} \sum_{i=1}^{n} \gamma_i \quad .$$

Here and throughout the rest of this paper $\|\cdot\|_p$, $1 < p < \infty$, denotes the usual L_p norm on the real line and C denotes an absolute constant. If in addition $X_1,\ldots,X_n$ are identically distributed, then the bound on $\|F_n - \Phi\|_\infty$ if of the order $n^{-1/2}$. It is known that the rate of convergence $n^{-1/2}$ is best possible.

There has been considerable interest in the literature in obtaining a bound for $\|F_n - \Phi\|_\infty$ in the case when $X_1,\ldots,X_n$ are dependent. Many attempts have been made to obtain a bound having an order as close to $n^{-1/2}$ as possible and in many cases existence of moments higher than the third was assumed. Some of the early works in this direction were by Philipp (1969), Petrov (1970), Egorov (1970), Stein (1972), Erickson (1973, 1974) and Leonenko (1975). More recent works include those of Maejima (1978), Neaderhouser (1978b), Shergin (1979), Tikhomirov (1980), Riauba (1980), Prakasa Rao (1981), Schneider (1981), Bolthausen (1982a), Takahata (1981, 1983) and Guyon and Richardson (1984). Of these the better results for sequences of random variables are by Shergin, Tikhomirov, Schneider and Bolthausen, and the better results for random fields are by Takahata (1983) and Guyon and Richardson. Except for Bolthausen and Takahata these authors assumed existence of the third and sometimes lower moments. Of the latter we will state only those results assuming the existence of the third moment. Under this assumption, Tikhomirov obtained a bound of order $n^{-1/2}(\log n)^2$ for strongly mixing strictly stationary sequences of random variables and a bound of order $n^{-1/2}(\log n)^{3/2}$ for completely regular (φ-mixing) strictly stationary sequences. In each case the mixing rate is assumed to be exponentially decreasing. The latter result was extended to nonstationary sequences by Schneider who also improved the order of the bound to $n^{-1/2}\log n$. Tikhomirov's result for strictly stationary m-dependent sequences was generalized by Shergin who proved that if $X_1,\ldots,X_n$ is an m-dependent sequence with $EX_i = 0$, $\mathrm{Var}(\sum_{i=1}^{n} X_i) = B_n^2$ and $E|X_i|^3 = \gamma_i < \infty$, then

$$\|F_n - \Phi\|_\infty < C(m+1)^2 B_n^{-3} \sum_{i=1}^{n} \gamma_i \quad . \tag{1.1}$$

By a result of Berk (1973), the order for m in (1.1) is best possible. Bolthausen was able to obtain a bound of order $n^{-1/2}$ for strongly mixing stationary Harris recurrent Markov chain by assuming certain mixing rate and the existence of a moment higher than the third.

To state the results of Takahata (1983) and Guyon and Richardson (1984), we regard the indices $1,\ldots,n$ of $X_1,\ldots,X_n$ as being from $\mathbb{Z}^d$ for notational convenience. Takahata proved that for strongly mixing random fields with exponentially decreasing mixing rate and bounded absolute $(8+\delta)$th moments for some $\delta > 0$ such that

(1.2) $$\liminf_{n\to\infty} n^{-1}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) > 0,$$

we have

(1.3) $$\|F_n - \Phi\|_\infty = O(n^{-1/2}(\log n)^d).$$

Guyon and Richardson obtained a rate of $n^{-1/2}(\log n)^{2d}$ for strongly mixing random fields with exponentially decreasing mixing rate assuming (1.2) and only the boundedness of the absolute third moments. But the mixing condition assumed by Guyon and Richardson is different from that assumed by Takahata who took into account the sizes of two disjoint subsets of $\mathbb{Z}^d$ in addition to the distance between them. The mixing condition of Guyon and Richardson cannot be applied to Gibbs random fields (see Dobrushin (1968)). For m-dependent random fields, Takahata and Guyon and Richardson proved under condition (1.2) that

(1.4) $$\|F_n - \Phi\|_\infty = O(n^{-1/2}).$$

While Takahata assumed the boundedness of the eighth moments for this result, Guyon and Richardson assumed only the boundedness of the absolute third moments. The authors did not indicate how the constants in (1.3) and (1.4) depend on d, the moments and the mixing rate.

If an m-dependent sequence is stationary with $B_n \to \infty$, then (1.1) yields a rate of $n^{-1/2}$. Similarly if an m-dependent random field is stationary and $\mathrm{Cov}(X_i, Y_i) > 0$ where Y_i is the sum of those X_j dependent of X_i, then condition (1.2) is satisfied and the rate of $n^{-1/2}$ in (1.4) holds. However, without stationarity, even if the random variables are bounded and identically distributed, the bound is (1.1) may tend to infinity and condition (1.2) may not be satisfied. Here is an example due to Erickson (1974). Define a sequence of bounded and identically distributed random variables $X_1, \ldots, X_n$ with $EX_i = 0$ and $EX_i^2 = 1$ as follows. Let $B_k^2 = \mathrm{Var}(\sum_{i=1}^{k} X_i)$ and let X_1, X_2 be independent. For $k \geq 2$, define $X_{k+1} = -X_k$ if $B_k^2 > k^{1/2}$ and define X_{k+1} to be independent of $X_1, \ldots, X_k$ if $B_k^2 \leq k^{1/2}$. It is not difficult to see that $X_1, \ldots, X_n$ is a 1-dependent sequence, $|B_n^2 - n^{1/2}| < 2$ and $\sum_{i=1}^{n} X_i$ is a sum of $B_n^2 \approx n^{1/2}$ terms of independent and identically distributed random variables. By the Berry-Esseen theorem

$$\|F_n - \Phi\|_\infty < C\gamma n^{-1/4} \tag{1.5}$$

where $\gamma = E|X_i|^3$. But the bound in (1.1) is asymptotically $C\gamma n^{1/4}$ which $+\infty$, and condition (1.2) for (1.4) is not satisfied. In view of the example of Erickson it is natural to ask whether a better bound exists and if it does, what is it? One of the objectives of this paper is to answer this question.

We consider random variables $X_1,\ldots,X_n$ with $EX_i = 0$, $\mathrm{Var}(\sum_{i=1}^{n} X_i) = B^2$ and $E|X_i|^3 = \gamma_i < \infty$ and satisfying the following dependence assumption. For each $i \in \{1,\ldots,n\}$, there exist nonempty $A_i \subset B_i \subset C_i \subset \{1,\ldots,n\}$ such that $\{X_i\}$ is independent of $X_{A_i^c}$, X_{A_i} is independent of $X_{B_i^c}$ and X_{B_i} is independent of $X_{C_i^c}$. Here X_A denotes $\{X_i : i \in A\}$. We also assume that each C_i is relatively small compared to $\{1,\ldots,n\}$. One can see that the index set $\{1,\ldots n\}$ is actually arbitrary although the natural numbers are used. This class of dependent random variables has an interesting structure. If $X_1,\ldots,X_n$ is the n^{th} row $X_{n1},\ldots,X_{nn}$ of a triangular array with A_{ni} replacing A_i and so on, it has been proved in Chen (1978, theorem 4.1) that under very natural conditions, $\mathcal{L}(\sum_{i=1}^{n} X_{ni})$ converges weakly to an infinitely divisible distribution with finite canonical measure ν if and only if the signed measure ν_n converges vaguely to ν where $Y_{ni} = \sum_{j \in A_{ni}} X_{nj}$ and $\nu_n(A) = \sum_{i=1}^{n} \mathrm{Cov}(X_{ni}, Y_{ni} I(Y_{ni} \in A))$ for $A \in \mathcal{B}(\mathbb{R})$. It is clear that this class of dependent random variables includes m-dependent random fields as special cases and many spatial statistics can be represented by a sum of such dependent random variables. (See for example Mantel (1967), David (1972), Sen (1976), Shapiro and Hubert (1979), and Cliff and Ord

(1981)). Furthermore the results proved in this paper can be extended in a fairly routine manner to mixing random fields and in particular certain Gibbs random fields. (See Dobrushin (1968), Holley and Stroock (1976) and Neaderhouser (1978a)). We have chosen not to include results for mixing random fields in this paper because we want to minimize the amount of tedious calculations in order to focus on the ideas and techniques of our proofs.

The main result in this paper is

$$(1.6) \qquad \|F-\Phi\|_p \le C\lambda^{\frac{p-1}{p}} B^{-3}\beta \quad \text{for} \quad 1 \le p \le \infty \ ,$$

where F is the distribution function of $B^{-1}\sum_{i=1}^{n} X_i$, $X_1,\ldots,X_n$ satisfy our dependence assumption, and λ and β are given by (2.6) and (2.7) respectively. We note that λ is a dimensionless quantity and β is a generalization of $\sum_{i=1}^{n} \gamma_i$. It is reasonable to assume that λ remains bounded as $n \to \infty$. (See Remarks in Section 5.) It is shown in Section 5 that the bound in (1.6) coincides with that in (1.5) when it is applied to the example of Erickson. It is also shown that (1.6) yields

$$(1.7) \qquad \|F-\Phi\|_p \le C\theta^2 B^{-3} \sum_{i=1}^{n} \gamma_i \quad \text{for} \quad 1 \le p \le \infty \ ,$$

where θ is given by (2.5). When applied to m-dependent random fields on $\mathbb{Z}^d$, $\theta \le (10m+1)^d$ (see (5.24)). For $d = 1$ we obtain a bound of the same order as that in (1.1) with the best possible order for m.

Since $\|\cdot\|_p^p \le \|\cdot\|_\infty^{p-1}\|\cdot\|_1$, it suffices to prove (1.6) for $p = \infty$ and 1. An L_1 bound for independent random variables was first proved by Esseen (1958) and that for m-dependent sequences by Erickson (1973, 1974) who obtained a bound of the order $(m+1)^3 \sum_{i=1}^{n} \gamma_i$ in the second paper. The result of Erickson was generalized by Loh (1975) to completely regular sequences. Takahata (1983) considered the L_1 problem under the same dependence assumptions as his L_∞ problem, and under the condition (1.2), obtained L_1

bounds of the same orders as in (1.3) and (1.4) assuming the boundedness of the absolute $(4+\delta)$th moments $(\delta>0)$ and the fourth moments respectively.

Most of the authors mentioned above used the characteristic function method and Bernstein's technique of dividing the sum into blocks and approximating the blocks by independent random variables. Stein (1972) invented his own method. Erickson (1974) and Loh (1975) adapted the method for L_1 bounds. Takahata's works (1981, 1983) are based on Stein's results. Tikhomirov (1980) and Guyon and Richardson (1984) combined techniques associated with Stein's method with the use of characteristic functions. Such an approach was also used by Chen (1978) and Bolthausen (1982b) but these authors were only concerned with limit theorems and not the rate of convergence.

Apart from obtaining new results, another objective of this paper is to apply Stein's method and in doing so develop general techniques which we hope are applicable to other problems. Stein's method may be briefly described as follows. Let $W = B^{-1}\sum_{i=1}^{n} X_i$ where $X_1,\dots,X_n$ are dependent random variables with $EX_i = 0$ and $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = B^2$. We derive an identity of the form

$$E\{f'(W) - Wf(W)\} = ER_f(X_1,\dots,X_n) \tag{1.8}$$

for absolutely continuous f such that $|f(w)| < C(1+|w|)$ for some constant C, where the right hand side of (1.8) is an error term. By choosing f in (1.8) to be f_h, the unique bounded solution of the differential equation

$$f'(w) - wf(w) = h(w) - Nh \tag{1.9}$$

where h is a bounded piecewise continuous function and $Nh = (2\pi)^{-1/2}\int_{-\infty}^{\infty} h(x)e^{-x^2/2}dx$, we obtain

$$Eh(W) - Nh = ER_{f_h}(X_1,\dots,X_n) \tag{1.10}$$

If $h = I_{(-\infty,z]}$, then the left hand side of (1.10) is $F(z) - \Phi(z)$. We will be concerned only with $h = I_{(-\infty,z]}$ and $h = h_{z,\alpha}$ which is defined by (5.3). Bounds for $\|F-\Phi\|_\infty$ and $\|F-\Phi\|_1$ can then be obtained by appropriately bounding the right hand side of (1.10). There are two major steps in the application of Stein's method: the derivation of an identity of the form (1.8) and the bounding of the error term. The techniques involved in these two steps depend on the nature of dependence among $X_1,\ldots,X_n$ and may vary from problem to problem. Our aim is to develop these techniques in such a way that they may also be applicable to other problems.

For $X_1,\ldots,X_n$ satisfying our dependence assumption, we use an exchangeability argument to derive an identity of the form (1.8). Our exchangeability argument is related to Stein's idea of an exchangeability pair and an antisymmetric function but appears to be more intuitive and flexible in applications. (See Diaconis (1977) and Stein (1986).) By using (1.9) and the boundedness properties of f_h, a bound is immediately obtained for $\|F-\Phi\|_1$ whereas the problem of bounding $\|F-\Phi\|_\infty$ is reduced to that of bounding $P^\zeta(a_\zeta < W < b_\zeta)$ where ζ is a random vector depending on a relatively small subset of $\{X_1,\ldots,X_n\}$.

To bound $P^\zeta(a_\zeta < W < b_\zeta)$, we use a conditional version of our identity and choose f in the identity to be dependent on ζ and such that its derivative f' is an appropriate trapezoidal function. We call the resulting inequality for $P^\zeta(a_\zeta < W < b_\zeta)$ a conditional concentration inequality. The use of a concentration inequality in bounding $\|F-\Phi\|_\infty$ was originally due to Stein in proving the Berry-Esseen theorem for independent and identically distributed random variables. A simplified version of Stein's proof is given in Ho and Chen (1978). However Stein's proof uses the symmetry inherited from the fact that the random variables are independent and identically distributed and extension of the concentration inequality by using an identity of the form (1.8) even to independent but non-identically distributed random variables is difficult and remained unsolved for some time. In a discarded chapter of his monograph (1986), Stein has a proof of a concentration inequality involving an exhchangeable pair. The inequality is then applied to independent but non-identically distributed random variables. The proof of our conditional concentration inequality which also covers the independent but non-identically distributed case is different.

The foregoing discussion shows that bounding $\|F-\Phi\|_\infty$ is more difficult than bounding $\|F-\Phi\|_1$ using Stein's method. This fact was first observed by Erickson (1974).

We have only discussed the application of Stein's method to normal approximation. For an abstract formulation of Stein's method and other applications, see Stein (1986) or Chen (1979).

We have mentioned earlier that some of the above authors obtained their results assuming the existence of moments lower than the third. In particular, for m-dependent sequences Shergin (1979) obtained

$$\|F_n - \Phi\|_\infty \le \inf_{\varepsilon>0} C(m+1)^2[R_{2,n}(\varepsilon) + R_{3,n}(\varepsilon)] \tag{1.11}$$

and Erickson (1974) obtained

$$\|F_n - \Phi\|_1 \le \inf_{\varepsilon>0} C(m+1)^3[R_{2,n}(\varepsilon) + R_{3,n}(\varepsilon)] \tag{1.12}$$

where $R_{2,n}(\varepsilon) = B_n^{-2} \sum_{i=1}^n EX_i^2 I(|X_i| > \varepsilon B_n)$ and $R_{3,n}(\varepsilon) = B_n^{-3} \sum_{i=1}^n E|X_i|^3 I(|X_i| \le \varepsilon B_n)$. It is possible to extend our techniques to obtain a bound for $\|F-\Phi\|_p$ involving our β but in a way analogous to the bound in (1.11). However we sacrifice refinement for simplicity in this paper. Perhaps we will consider refinements and extension to mixing random fields in another paper.

2. A class of dependent random variables.

In this section we state our dependence assumption and define notations which will be used throughout the paper. Let $X_1,\ldots,X_n$ be random variables with $EX_i = 0$, $\mathrm{Var}(\sum_{i=1}^{n} X_i) = B^2$ and $E|X_i|^3 = \gamma_i < \infty$. We assume that for each $i \in \{1,\ldots,n\}$, there exist nonempty $A_i \subset B_i \subset C_i \subset \{1,\ldots,n\}$ such that $\{X_i\}$ is independent of $X_{A_i^c}$, X_{A_i} is independent of $X_{B_i^c}$ and X_{B_i} is independent of $X_{C_i^c}$. Here X_A denotes $\{X_i : i \in A\}$. It is clear that the index set of $X_1,\ldots,X_n$ is arbitrary although we use the natural numbers for notational convenience.

For $A \subset \{1,\ldots,n\}$, we define

(2.1) $N(A) \subset \{1,\ldots,n\}$ to be such that for each $i \in N(A)^c$, $B_i \cap A = \phi$.

For each $j \in \{1,\ldots,n\}$, we define

$$N^*(C^j) = \{i : j \in N(C_i)\} \quad . \tag{2.2}$$

We assume that the C_i are relatively small compared to $\{1,\ldots,n\}$ so that each $N(C_i)$ is a proper subset of $\{1,\ldots,n\}$.

For each $i \in \{1,\ldots,n\}$, let

$$\begin{cases} V_i = \sum_{j\in B_i^c} X_j \ , \quad Y_i = \sum_{j\in A_i} X_j \ , \\ Z_i = \sum_{j\in B_i} X_j \ , \quad T_i = \sum_{j\in C_i} X_j \ . \end{cases} \tag{2.3}$$

As usual $|A|$ denotes the cardinality of the set A. We define

$$\tilde{B}_i^2 = \mathrm{Var}\Big(\sum_{j\in N(C_i)^c} X_j\Big) \tag{2.4}$$

$$\theta = \max_{1\le i,j\le n} \{|N(C_i)|, |N^*(C^j)|\} \tag{2.5}$$

$$\lambda = \max_{1\le i\le n} \left(\frac{B}{\tilde{B}_i}\right)^3 \vee 1 \tag{2.6}$$

$$(2.7)\qquad \beta = \sum_{i=1}^{n} \sum_{j\in N(C_i)\cup N^*(C^i)} \{E|X_jX_iY_i| + E|X_jX_iZ_i|\}$$

$$+ \theta \sum_{i=1}^{n} \{E|X_i^2Y_i| + EX_i^2E|Z_i|\}$$

$$+ \sum_{i=1}^{n} \{E|X_iY_iZ_i| + E|X_iY_i^2| + E|X_iZ_i^2|$$

$$+ E|X_iY_i|E|Z_i| + E|X_iY_i|E|T_i| + E|X_iZ_i|E|T_i|$$

$$+ E|X_i|EY_i^2 + E|X_i|EZ_i^2$$

$$+ E|X_i|E|Y_iZ_i| + E|X_i|E|Z_iT_i| + E|X_i|E|T_iY_i|\} .$$

Let $\Lambda_0 \subset \Lambda \subset \{1,\ldots,n\}$ be such that X_{Λ_0} is independent of X_{Λ^c} and let ζ be a random vector depending on X_{Λ_0}. Let $X_1',\ldots,X_n'$ be an independent copy of $X_1,\ldots,X_n$ and for $i \in \{1,\ldots,n\}$, let

$$(2.8)\qquad Y_i' = \sum_{j\in A_i} X_j' .$$

We define

$$(2.9)\qquad \tilde{W} = \sum_{i\in N(\Lambda)^c} X_i$$

$$(2.10)\qquad \tilde{B}^2 = \mathrm{Var}(\tilde{W})$$

$$(2.11)\qquad \Gamma(y,t) = I(y > t > 0) - I(y < t < 0)$$

$$(2.12)\qquad \hat{M}(t) = \sum_{i\in N(\Lambda)^c} X_i\Gamma(Y_i - Z_i,t)$$

$$(2.13)\qquad \hat{K}(t) = \sum_{i\in N(\Lambda)^c} X_i'\Gamma(Y_i' - Z_i,t)$$

$$(2.14)\qquad K(t) = E\hat{K}(t) .$$

Note that by the definitions of Λ_0, Λ and $N(\Lambda)$, $\hat{M}(t)$ and ζ are independent and by the dependence assumption X_i and $Y_i - Z_i$ are independent for each i. Consequently

$$(2.15)\qquad E^{\zeta}\hat{M}(t) = E\hat{M}(t) = 0 .$$

Likewise $\hat{K}(t)$ and ζ are independent and so

(2.16) $$E^{\zeta}\,\hat{K}(t) = E\hat{K}(t) = K(t) \quad .$$

By direct computations

(2.17) $$\int K(t)dt = \sum_{i\in N(\Lambda)^c} EX_i'(Y_i' - Z_i) = \sum_{i\in N(\Lambda)^c} EX_i'Y_i'$$

$$= \sum_{i\in N(\Lambda)^c} EX_iY_i = \mathrm{Var}(\tilde{W}) = \tilde{B}^2 \quad .$$

If $\Lambda_0 = \phi$, then we take $N(\Lambda) = \phi$.

3. An exchangeability argument. Let $X_1,\ldots,X_n$ be dependent random variables with $EX_i = 0$ and $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = 1$ and let $W = \sum_{i=1}^{n} X_i$ (note $B = 1$). In this section we describe a method for deriving identities of the form (1.8). For the time being we do not assume that $X_1,\ldots,X_n$ satisfy the dependence assumption of Section 2.

Suppose there exists a probability space $(\Omega,\mathcal{B},P)$ on which the random variables X_i, ξ_i, X_i', ξ_i' and U_i, $i = 1,\ldots,n$, are defined such that for each i,

(i) $\mathcal{L}(X_i,\xi_i,X_i',\xi_i',U_i) = \mathcal{L}(X_i',\xi_i',X_i,\xi_i,U_i)$.

Then we have

$$\sum_{i=1}^{n} EX_i f(U_i + \xi_i) = \sum_{i=1}^{n} EX_i' f(U_i + \xi_i') \quad , \tag{3.1}$$

for absolutely continuous f such that $|f(w)| \leq C(1 + |w|)$ for some constant C. Suppose, in addition, for each i,

(ii) U_i is close to W in some sense,

(iii) X_i, ξ_i, X_i', ξ_i' are small in some sense,

(iv) (X_i',ξ_i') are weakly correlated to functions of (X_i,ξ_i,U_i,W),

(v) $\sum_{i=1}^{n} EX_i\xi_i \approx 1$.

Then (3.1) can be rewritten as an identity of the form (1.8) with the error term small in some sense. We show heuristically how this can be done. We will derive explicit identities of the form (1.8) later when we apply this exchangeability argument to dependence random variables satisfying the dependence assumption of Section 2. By (ii) and (iii), the left hand side of (3.1)

$$\approx \sum_{i=1}^{n} EX_i f(W) = EWf(W) \quad .$$

By (iv), the right hand side of (3.1)

$$= \sum_{i=1}^{n} EX_i'[f(U_i + \xi_i') - f(U_i)]$$

$$= \sum_{i=1}^{n} EX_i' \int \Gamma(\xi_i',t)f'(U_i+t)dt$$

where Γ is defined by (2.11). By (ii) and (iv), the last sum

$$\approx \int\Big(\sum_{i=1}^{n} EX_i'\Gamma(\xi_i',t)\Big)Ef'(W+t)dt$$

$$= \int\Big(\sum_{i=1}^{n} EX_i\Gamma(\xi_i,t)\Big)Ef'(W+t)dt$$

$$= \int \tilde{K}(t)Ef'(W+t)dt$$

where $\tilde{K}(t) = \sum_{i=1}^{n} EX_i\Gamma(\xi_i,t)$. Now $\tilde{K}$ is the density of a signed measure and $\int \tilde{K}(t)dt = \sum_{i=1}^{n} EX_i\xi_i$ which by (v) ≈ 1. If the X_i and ξ_i are sufficiently small, then the mass of $\tilde{K}$ is concentrated near 0 and $\int \tilde{K}(t)Ef'(W+t)dt \approx Ef'(W)$. Consequently (3.1) leads to

$$E\{f'(W) - Wf(W)\} \approx 0$$

which is of the form (1.8) with small error term.

Sometimes it is more convenient to use

$$\sum_{i=1}^{n} EX_i f(U_i + \xi_i + \xi_i') = \sum_{i=1}^{n} EX_i' f(U_i + \xi_i + \xi_i') \tag{3.2}$$

instead of (3.1) although in this paper we will use (3.1). It is easy to see that (i) also implies (3.2).

We now see how our exchangeability argument is related to Stein's use of an exchangeable pair and an antisymmetric function. Suppose W takes a finite number of values and there exists a probability space with W and W' defined on it such that $L(W,W') = L(W',W)$. Stein (1986) used homology theory to prove that under certain connectedness condition if $Eg(W) = 0$ then there exists an antisymmetric function $F : \mathbb{R}^2 \to \mathbb{R}$ with $E|F(W,W')| < \infty$ such that $g(W) = E^W F(W,W')$. Guided by this result he used the antisymmetric function $(w,w') \mapsto (w'-w)[f(w) + f(w')]$ and the easily verified fact $E\{(W'-W)[f(W)+f(W')]\} = 0$ to derive an identity of the form (1.8). See Stein (1986, Lecture I). We show that (3.1) can be rewritten as

$$EF_f(S,S') = 0 \tag{3.3}$$

where $L(S,S') = L(S',S)$, S and S' are approximately equal to W and F_f is an antisymmetric function similar to $(w,w) \to (w'-w)[f(w)+f(w')]$.

Let I be uniformly distributed on $\{1,\ldots,n\}$ and independent of X_i, ξ_i, X_i', ξ_i', U_i, $i = 1,\ldots,n$. Define $S = U_I + X_I$ and $S' = U_I + X_I'$. By (ii) and (iii), $S \approx S' \approx W$ and by (i), $L(S,S') = L(S',S)$. Also by (i)

$$EU_I f(U_I + \xi_I) = EU_I f(U_I + \xi_I') \tag{3.4}$$

where f is as in (3.1). Now (3.1) can be rewritten as

$$EX_I f(U_I + \xi_I) = EX_I' f(U_I + \xi_I') \quad . \tag{3.5}$$

Combining (3.4) and (3.5), we obtain

$$ESf(U_I + \xi_I) = ES'f(U_I + \xi_I')$$

which is

$$ESE^S f(U_I + \xi_I) = ES'E^{S'} f(U_I + \xi_I') \quad . \tag{3.6}$$

By (i) again, the regular conditional distribution of $U_I + \xi_I$ given $S = s$ is the same as that of $U_I + \xi_I'$ given $S' = s$. By letting $\varphi_f(s)$ to be the regular conditional expectation $E^{S=s} f(U_I + \xi_I)$, (3.6) becomes

$$ES\varphi_f(S) = ES'\varphi_f(S')$$

which is equivalent to

$$E\{(S'-S)[\varphi_f(S) + \varphi_f(S')]\} = 0 \quad . \tag{3.7}$$

The function $(s,s') \mapsto (s'-s)[\varphi_f(s) + \varphi_f(s')]$ is clearly antisymmetric.

Now we apply the exchangeability argument to $X_1,\ldots,X_n$ satisfying the dependence assumption of Section 2. We use the same definition and notations as in Section 2. We assume $B = 1$ so that we have $EX_i = 0$, $\mathrm{Var}(\sum_{i=1}^{n} X_i) = 1$ and $E|X_i|^3 = \gamma_i < \infty$. Let $W = \sum_{i=1}^{n} X_i$. Since $(X_1',\ldots,X_n')$ is an independent copy of $(X_1,\ldots,X_n)$ and for each $i \in N(\Lambda)^c$, $B_i \cap \Lambda = \phi$, it follows that for each $i \in N(\Lambda)^c$,

$$(X_i, Y_i),\ (X_i', Y_i') \text{ and } (V_i, \zeta) \text{ are independent}$$

and

$$L(X_i, Y_i, X_i', Y_i', V_i, \zeta) = L(X_i', Y_i', X_i, Y_i, V_i, \zeta) \quad .$$

Let $(t,w) \to f_t(w)$ be Borel measurable such that $|f_\zeta(w)| \le C(1 + |w|)$ for some constant C. By the exchangeability argument

$$\sum_{i \in N(\Lambda)^c} E^\zeta X_i f_\zeta(V_i + Y_i) = \sum_{i \in N(\Lambda)^c} E^\zeta X_i' f_\zeta(V_i + Y_i') \quad . \tag{3.8}$$

Here and throughout we drop the term "a.s." for simplicity. Now based on (3.8) we derive two slightly different identities of the form (1.8) which will be used separately for the next two sections. Write the left hand side of (3.8) as

$$\begin{aligned} &\sum_{i \in N(\Lambda)^c} E^\zeta X_i [f_\zeta(V_i + Y_i) - f_\zeta(W)] + E^\zeta \tilde{W} f_\zeta(W) \\ &= \sum_{i \in N(\Lambda)^c} E^\zeta X_i \int_0^{Y_i - Z_i} f_\zeta'(W+t)\,dt + E^\zeta \tilde{W} f_\zeta(W) \\ &= E^\zeta \int f_\zeta'(W+t)\hat{M}(t)\,dt + E^\zeta \tilde{W} f_\zeta(W) \quad . \end{aligned} \tag{3.9}$$

On the other hand, the right hand side of (3.8) equals

$$\begin{aligned} &\sum_{i \in N(\Lambda)^c} E^\zeta X_i' [f_\zeta(V_i + Y_i') - f_\zeta(W)] \\ &= \sum_{i \in N(\Lambda)^c} E^\zeta X_i' \int_0^{Y_i' - Z_i} f_\zeta'(W+t)\,dt \\ &= E^\zeta \int f_\zeta'(W+t)\hat{K}(t)\,dt \\ &= E^\zeta \int f_\zeta'(W+t)[\hat{K}(t) - K(t)]\,dt \\ &\quad + E^\zeta \int [f_\zeta'(W+t) - f_\zeta'(W)]K(t)\,dt + \tilde{B}^2 E^\zeta f_\zeta'(W) \end{aligned} \tag{3.10}$$

where we have used (2.17). Combining (3.8), (3.9) and (3.10), we obtain the first identity

$$\tilde{B}^2 E^\zeta f_\zeta'(W) = E^\zeta \tilde{W} f_\zeta(W) + E^\zeta \int f_\zeta'(W+t)\hat{M}(t)dt \tag{3.11}$$

$$- E^\zeta \int f_\zeta'(W+t)[\hat{K}(t) - K(t)]dt$$

$$- E^\zeta \int [f_\zeta'(W+t) - f_\zeta'(W)]K(t)dt \quad .$$

For the second identity we go back to (3.8) and take $\Lambda_0 = \phi$. Then we have

$$\sum_{i=1}^{n} EX_i f(V_i + Y_i) = \sum_{i=1}^{n} EX_i' f(V_i + Y_i') \quad . \tag{3.12}$$

By using the fact that X_i and $(Z_i - Y_i, V_i)$ are independent and $EX_i = 0$ we write the left hand side of (3.12) as

$$\sum_{i=1}^{n} EX_i [f(V_i + Y_i) - f(W)] + EWf(W) \tag{3.13}$$

$$= - \sum_{i=1}^{n} E \int f'(V_i + Y_i + t) X_i \Gamma(Z_i - Y_i, t)dt + EWf(W)$$

$$= - \sum_{i=1}^{n} E \int [f'(V_i + Y_i + t) - f'(V_i + t)] X_i \Gamma(Z_i - Y_i, t)dt + EWf(W) \quad .$$

Since (X_i', Y_i') and (V_i, W) are independent and $EX_i' = EX_i = 0$, the right hand side of (3.12) equals

$$\sum_{i=1}^{n} EX_i' [f(V_i + Y_i') - f(V_i)] \tag{3.14}$$

$$= \sum_{i=1}^{n} E \int f'(V_i + t) X_i' \Gamma(Y_i', t)dt$$

$$= \sum_{i=1}^{n} E \int [f'(V_i + t) - f'(W)] EX_i' \Gamma(Y_i', t)dt$$

$$+ Ef'(W) \sum_{i=1}^{n} \int EX_i' \Gamma(Y_i', t)dt$$

$$= - \sum_{i=1}^{n} E \int [f'(W) - f'(V_i+t)]EX_i\Gamma(Y_i,t)dt + Ef'(W)$$

where we have used the fact that $\sum_{i=1}^{n} \int EX_i'\Gamma(Y_i',t)dt = \sum_{i=1}^{n} EX_i'Y_i'$ $= \sum_{i=1}^{n} EX_iY_i = Var(W) = 1$. Combining (3.12), (3.13) and (3.14), we obtain the second identity

(3.15) $\quad E\{f'(W) - Wf(W)\}$

$$= - \sum_{i=1}^{n} E \int [f'(V_i+Y_i+t) - f'(V_i+t)]X_i\Gamma(Z_i-Y_i,t)dt$$

$$+ \sum_{i=1}^{n} E \int [f'(W) - f'(V_i+t)]EX_i\Gamma(Y_i,t)dt \quad .$$

4. A conditional concentration inequality.

As was mentioned in Section 1, a conditional concentration inequality is required to bound $\|F - \Phi\|_\infty$. Since such a concentration inequality is of interest in itself we prove it in a separate section here.

Let $X_1, \ldots, X_n$ be random variables with $EX_i = 0$, $\mathrm{Var}\,(\sum_{i=1}^n X_i) = B^2$ and $E|X_i|^3 = \gamma_i < \infty$ and satisfying the dependence assumption of Section 2. Let $W = B^{-1} \sum_{i=1}^n X_i$ and let Λ_0, Λ and ζ be as defined in Section 2.

PROPOSITION 4.1. For Borel measurable functions a_ζ and b_ζ of ζ such that $a_\zeta \leq b_\zeta$, we have

$$(4.1) \qquad P^\zeta(a_\zeta \leq W \leq b_\zeta) \leq (3/2)\,\tilde{B}^{-1}(b_\zeta - a_\zeta) + 15\tilde{B}^{-3}\beta \quad \text{a.s.}$$

where β is defined by (2.7) and $\tilde{B}$ by (2.10).

If $\Lambda_0 = \phi$, we take $P^\zeta = P$, $a_\zeta = a$, $b_\zeta = b$ and $\tilde{B} = B$.

To prove the proposition we assume without loss of generality that $B = 1$ and $b_\zeta - a_\zeta$ bounded and use the identity (3.11). Denote the terms on the right hand side of (3.11) by $R_i^\zeta(f_\zeta)$, $i = 1,2,3,4$, respectively. Now choose f_ζ in (3.11) to be such that $f_\zeta(2^{-1}(a_\zeta + b_\zeta)) = 0$, f_ζ' is continuous and given by

$$f_\zeta'(w) = \begin{cases} 1 & , a_\zeta \leq w \leq b_\zeta , \\ 0 & , w \leq a_\zeta - \alpha \text{ or } \geq b_\zeta + \alpha , \\ \text{linear} & , a_\zeta - \alpha \leq w \leq a_\zeta \text{ or } b_\zeta \leq w \leq b_\zeta + \alpha \end{cases}$$

where α is a positive number to be determined later. Clearly $|f_\zeta(w)| \leq 2^{-1}(b_\zeta - a_\zeta) + \alpha$. With this f_ζ,

$$(4.2)\qquad R_1^\zeta(f_\zeta) < [2^{-1}(b_\zeta - a_\zeta) + \alpha]E^\zeta|\tilde{W}|$$

$$= [2^{-1}(b_\zeta - a_\zeta) + \alpha]E|\tilde{W}|$$

$$< \tilde{B}[2^{-1}(b_\zeta - a_\zeta) + \alpha] \quad .$$

By the geometric mean-arithmetic mean (GM-AM) inequality,

$$(4.3)\qquad R_2^\zeta(f_\zeta) < 2^{-1}\tilde{B}\; E^\zeta \int [f_\zeta'(W+t)]^2 dt + 2^{-1}\,\tilde{B}^{-1}\; E^\zeta \int [\hat{M}(t)]^2 dt$$

$$< 2^{-1}\,\tilde{B}\; E^\zeta(b_\zeta - a_\zeta + 2\alpha) + 2^{-1}\,\tilde{B}^{-1}\; E\int [\hat{M}(t)]^2 dt$$

$$= \tilde{B}[2^{-1}(b_\zeta - a_\zeta) + \alpha] + 2^{-1}\,\tilde{B}^{-1}\int \mathrm{Var}[\hat{M}(t)]dt \quad .$$

Likewise

$$(4.4)\qquad R_3^\zeta(f_\zeta) < \tilde{B}[2^{-1}(b_\zeta - a_\zeta) + \alpha] + 2^{-1}\,\tilde{B}^{-1}\int \mathrm{Var}[\hat{K}(t)]dt \quad .$$

Now we abuse the notation a little and regard E^ζ and P^ζ as regular conditional expectation and probability given a particular value of ζ until we state otherwise. Define

$$L_\zeta(\alpha) = \sup_{x\in\mathbb{R}} P^\zeta(x < W < x+\alpha) \quad .$$

Then by writing

$$R_4^\zeta(f_\zeta) = -\,E^\zeta\int_0^\infty\int_0^t f_\zeta''(W+s)ds K(t)dt + E^\zeta\int_{-\infty}^0\int_t^0 f_\zeta''(W+s)dsK(t)dt$$

$$= -\,\alpha^{-1}\int_0^\infty\int_0^t [P^\zeta(a_\zeta - \alpha < W < a_\zeta) - P^\zeta(b_\zeta < W < b_\zeta + \alpha)]dsK(t)dt$$

$$+\,\alpha^{-1}\int_{-\infty}^0\int_t^0 [P^\zeta(a_\zeta - \alpha < W < a_\zeta) - P^\zeta(b_\zeta < W < b_\zeta + \alpha)dsK(t)dt \quad ,$$

we have

$$(4.5)\qquad R_4^\zeta(f_\zeta) < \alpha^{-1}\int_0^\infty\int_0^t L_\zeta(\alpha)ds\;|K(t)|dt + \alpha^{-1}\int_{-\infty}^0\int_t^0 L_\zeta(\alpha)ds\;|K(t)|dt$$

$$= \alpha^{-1}\,L_\zeta(\alpha)\int_{-\infty}^\infty |tK(t)|dt \quad .$$

Since $P^{\zeta}(a_{\zeta} < W < b_{\zeta}) < E^{\zeta} f_{\zeta}'(W)$, it follows from (3.11), (4.2), (4.3), (4.4) and (4.5) that

$$(4.6) \qquad P^{\zeta}(a_{\zeta}<W<b_{\zeta}) < (3/2)\,\tilde{B}^{-1}(b_{\zeta}-a_{\zeta}+2\alpha) + 2^{-1}\,\tilde{B}^{-3}\int \mathrm{Var}[\hat{M}(t)]dt + 2^{-1}\,\tilde{B}^{-3}\int \mathrm{Var}[\hat{K}(t)]dt + \alpha^{-1}L_{\zeta}(\alpha)\tilde{B}^{-2}\int |tK(t)|dt \ .$$

Let $\rho = \tilde{B}^{-3}\int \mathrm{Var}[\hat{M}(t)]dt + \tilde{B}^{-3}\int \mathrm{Var}[\hat{K}(t)]dt$. Substituting $a_{\zeta} = x$ and $b_{\zeta} = x+\alpha$ in (4.6), we obtain

$$(4.7) \qquad L_{\zeta}(\alpha) < 9\tilde{B}^{-1}\alpha/2 + \rho/2 + \alpha^{-1}L_{\zeta}(\alpha)\tilde{B}^{-2}\int |tK(t)|dt \ .$$

Now let $\alpha = 2\tilde{B}^{-2}\int |tK(t)|dt = \alpha^*$ say. Then (4.7) yields

$$(4.8) \qquad L_{\zeta}(\alpha^*) < 9\tilde{B}^{-1}\alpha^*/2 + \rho/2 + L_{\zeta}(\alpha^*)/2 \ .$$

From this we get

$$(4.9) \qquad L_{\zeta}(\alpha^*) < 9\tilde{B}^{-1}\alpha^* + \rho = 18\tilde{B}^{-3}\int |tK(t)|dt + \rho \ .$$

Finally combining (4.6) and (4.9) and letting $\alpha = \alpha^*$ we get

$$(4.10) \qquad P^{\zeta}(a_{\zeta}<W<b_{\zeta}) < (3/2)\,\tilde{B}^{-1}(b_{\zeta}-a_{\zeta}) + 15\tilde{B}^{-3}\int |tK(t)|dt + \tilde{B}^{-3}\int \mathrm{Var}[\hat{M}(t)]dt + \tilde{B}^{-3}\int \mathrm{Var}[\hat{K}(t)]dt \ .$$

Now we can go back and regard the left hand side of (4.10) as ordinary conditional probability and the right hand side as a random variable. What remains to do is to bound the last three terms on the right hand side of (4.10) which we denote by R_1, R_2 and R_3 respectively. For R_1 we have

$$(4.11) \qquad \int |tK(t)|dt < \int |t| \sum_{i\in N(\Lambda)^c} E|X_i'\Gamma(Y_i' - Z_i,\ t)|dt < \sum_{i=1}^{n} E|X_i'| \int |t|[I(Y_i'-Z_i > t>0) + I(Y_i'-Z_i < t<0)]dt$$

$$= 2^{-1} \sum_{i=1}^{n} E|X_i'(Y_i'-Z_i)^2|$$

$$< \sum_{i=1}^{n} \{E|X_i'Y_i'^2| + E|X_i'Z_i^2|\}$$

$$= \sum_{i=1}^{n} \{E|X_iY_i^2| + E|X_i|EZ_i^2\} \quad .$$

For R_2, direct computations yield

(4.12) $$\int Var[\hat{M}(t)]dt$$

$$= \sum_{i\epsilon N(\Lambda)^c} \sum_{j\epsilon N(C_i)} E\{X_iX_j[(Y_i-Z_i)\wedge(Y_j-Z_j)I((Y_i-Z_i)\wedge(Y_j-Z_j) > 0)$$
$$+ (Y_i-Z_i)\vee(Y_j-Z_j)I((Y_i-Z_i)\vee(Y_j-Z_j) < 0)]\}$$

$$< \sum_{i=1} \sum_{j\epsilon N(C_i)} E\{|X_iX_j|[|Y_i-Z_i|\wedge|Y_j-Z_j|]\}$$

$$< 2^{-1} \sum_{i=1}^{n} \sum_{j\epsilon N(C_i)} E\{|X_iX_j|[|Y_i| + |Z_i| + |Y_j| + Z_j|]\} \quad .$$

For R_3, we have

(4.13) $$\int Var[\hat{K}(t)]dt$$

$$= \int \sum_{i\epsilon N(\Lambda)^c} \sum_{j\epsilon N(C_i)} Cov(X_i'\Gamma(Y_i'-Z_i,t),X_j'\Gamma(Y_j'-Z_j,t))dt$$

$$< \int \sum_{i=1}^{n} \sum_{j\epsilon N(C_i)} \{Var[X_i'\Gamma(Y_i'-Z_i,t)]Var[X_j'\Gamma(Y_j'-Z_j,t)]\}^{1/2}dt$$

$$< 2^{-1} \int \sum_{i=1}^{n} \sum_{j\epsilon N(C_i)} \{Var[X_i'\Gamma(Y_i'-Z_i,t)] + Var[X_j'\Gamma(Y_j'-Z_j,t)]\}dt$$

$$< 2^{-1} \sum_{i=1}^{n} \sum_{j\epsilon N(C_i)} \int \{E|X_i'^2\Gamma(Y_i'-Z_i,t)| + E|X_j'^2\Gamma(Y_j'-Z_j,t)|\}dt$$

$$< 2^{-1} \sum_{i=1}^{n} \sum_{j\epsilon N(C_i)} \{E|X_i'^2(Y_i'-Z_i)| + E|X_j'^2(Y_j'-Z_j)|\}$$

$$< 2^{-1} \sum_{i=1}^{n} \sum_{j \in N(C_i)} \{E|X_i^2 Y_i| + EX_i^2 E|Z_i| + E|X_j^2 Y_j| + EX_j^2 E|Z_j|\} .$$

By (4.10), (4.11), (4.12) and (4.13), the proposition is proved.

By using the Hölder and the GM-AM inequalities, $\beta < 210\theta^2 \sum_{i=1}^{n} \gamma_i$ and Proposition 4.1 yields

$$P^{\zeta}(a_{\zeta} < W < b_{\zeta}) < (3/2)\,\tilde{B}^{-1}(b_{\zeta} - a_{\zeta}) + 210\theta^2\,\tilde{B}^{-3} \sum_{i=1}^{n} \gamma_i \text{ a.s.} \tag{4.14}$$

Since β contains more terms than are contributed by (4.11), (4.12) and (4.13), the second absolute constant in (4.14) can be reduced by using (4.11), (4.12) and (4.13) directly. The result is the following corollary.

COROLLARY 4.1. Let ζ, a_{ζ} and b_{ζ} be as in Proposition 4.1. Then

$$P^{\zeta}(a_{\zeta} < W < b_{\zeta}) < (3/2)\,\tilde{B}^{-1}(b_{\zeta} - a_{\zeta}) + 34\theta^2\,\tilde{B}^{-3} \sum_{i=1}^{n} \gamma_i \quad \text{a.s.} \tag{4.15}$$

If $X_1, \ldots, X_n$ are independent, then $\theta = 1$. By taking $\Lambda_0 = \Lambda = N(\Lambda) = \phi$, we have the following classical result.

COROLLARY 4.2. For $a < b$,

$$P(a < W < b) < (3/2)\,B^{-1}(b-a) + 34B^{-3} \sum_{i=1}^{n} \gamma_i . \tag{4.16}$$

Of course the absolute constants in (4.16) can be substantially reduced if we prove the inequality for independent random variables directly.

5. The main theorem. Let $X_1,\ldots,X_n$ be random variables with $EX_i = 0$, $\operatorname{Var}(\sum_{i=1}^{n} X_i) = B^2$ and $E|X_i|^3 = \gamma_i < \infty$ and satisfying the dependence assumption of Section 2. Let F be the distribution function of $W = B^{-1}\sum_{i=1}^{n} X_i$ and Φ the standard normal distribution function.

THEOREM 5.1. There exists an absolute constant C such that for $1 < p < \infty$,

$$\|F-\Phi\|_p < C\lambda^{\frac{p-1}{p}} B^{-3}\beta \tag{5.1}$$

where λ is given by (2.6) and β by (2.7).

To prove Theorem 5.1 we assume without loss of generality that $B = 1$ and use the identity (3.15). Denote the error terms on the right hand side of (3.15) by $R_1(f)$ and $R_2(f)$ respectively. Now choose f in (3.15) to be $f_{z,\alpha}$, the unique bounded solution of the differential equation

$$f'(w) - wf(w) = h_{z,\alpha}(w) - Nh_{z,\alpha} \tag{5.2}$$

where $\alpha > 0$ is to be determined later and for $\alpha > 0$,

$$h_{z,\alpha}(w) = \begin{cases} 1 & , \quad w < z \ , \\ 1 + \frac{z-w}{\alpha} & , \quad z < w < z+\alpha \ , \\ 0 & , \quad w > z+\alpha \ , \end{cases} \tag{5.3}$$

and $Nh_{z,\alpha} = (2\pi)^{-1/2}\int_{-\infty}^{\infty} h_{z,\alpha}(x)e^{-x^2/2}dx$. For $\alpha = 0$, $h_{z,0} = I_{(-\infty,z]}$ and $Nh_{z,0} = \Phi(z)$. Then (3.15) yields

$$Eh_{z,\alpha}(W) - Nh_{z,\alpha} = R_1(f_{z,\alpha}) + R_2(f_{z,\alpha}) \ . \tag{5.4}$$

The solution of (5.2) is given by

$$f_{z,\alpha}(w) = e^{w^2/2} \int_{-\infty}^{w} [h_{z,\alpha}(x) - Nh_{z,\alpha}] e^{-x^2/2} dx$$

$$= -e^{w^2/2} \int_{w}^{\infty} [h_{z,\alpha}(x) - Nh_{z,\alpha}] e^{-x^2/2} dx \quad .$$

We need the boundedness properties of $f_{z,\alpha}$ which we state in the following lemmas.

LEMMA 5.1. For all w, z and α,

(5.5) $$0 \leq f_{z,\alpha}(w) \leq 1 \quad ,$$

(5.6) $$|wf_{z,\alpha}(w)| \leq 1 \quad ,$$

(5.7) $$|f'_{z,\alpha}(w)| \leq 1 \quad .$$

PROOF. For $\alpha = 0$, see Stein (1972) or Stein (1986). For $\alpha > 0$, use the fact that $h_{z,\alpha} = \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} h_{z+\alpha n^{-1},0}$.

LEMMA 5.2. For all w, z and $\alpha > 0$,

(5.8) $$|f'_{z,0}(w+s) - f'_{z,0}(w+t)| \leq [2\wedge(|s|+|t|)][|w|+1] + I(z-s\leq w\leq z-t)I(s\geq t) + I(z-t\leq w\leq z-s)I(s\leq t) \quad .$$

(5.9) $$|f'_{z,\alpha}(w+s) - f'_{z,\alpha}(w+t)| \leq [2\wedge(|s|+|t|)][|w|+1] + \alpha^{-1}\left|\int_t^s I(z\leq w+u\leq z+\alpha)du\right| \quad .$$

PROOF. Use (5.2), (5.5) and (5.7), keeping in mind that the left hand side of (5.8) or (5.9) is at most 2.

LEMMA 5.3. For all w,

(5.10) $$\int f_{z,0}(w)dz = 1 \quad ,$$

$$(5.11) \qquad \int |f'_{z,0}(w)|dz \le 1 \quad .$$

PROOF. See Ho and Chen (1978, Lemma 2.1).

LEMMA 5.4. For all w,

$$(5.12) \qquad \int |f'_{z,0}(w+s) - f'_{z,0}(w+t)|dz \le [1\wedge(|s|+|t|)][|w|+2] \quad .$$

PROOF. Use (5.2), (5.10) and (5.11), keeping in mind that the left hand side of (5.12) is at most 2.

We also need a lemma which relates $\|F-\Phi\|_\infty$ to $\sup_{z\in\mathbb{R}} |Eh_{z,\alpha}(W) - Nh_{z,\alpha}|$.

LEMMA 5.5. For $\alpha > 0$,

$$(5.13) \qquad \|F-\Phi\|_\infty \le \sup_{z\in\mathbb{R}} |Eh_{z,\alpha}(W) - Nh_{z,\alpha}| + (2\pi)^{-1/2}\alpha \quad .$$

PROOF. See Bolthausen (1984, Section 2) or use the fact that $Eh_{z-\alpha,\alpha}(W) \le F(z) \le Eh_{z,\alpha}(W)$ and that $\Phi(z+\alpha) - \Phi(z) \le (2\pi)^{-1/2}\alpha$.

Now we bound the error terms in (5.4). Since $\|\cdot\|_p^p \le \|\cdot\|_p^{p-1}\|\cdot\|_1$, it suffices to prove (5.1) for $p = \infty$ and 1. To facilitate calculations, we let C denote an absolute constant with possibly different magnitudes at different places. By Lemma 5.2,

$$(5.14) \qquad |R_1(f_{z,\alpha})|$$

$$\le \sum_{i=1}^{n} E\int [2\wedge(|Y_i+t| + |t|)](E^{\zeta_i}|V_i| + 1]|X_i\Gamma(Z_i-Y_i,t)|dt$$

$$+ \alpha^{-1} \sum_{i=1}^{n} E\int \{I(Y_i>0) \int_t^{Y_i+t} P^{\zeta_i}(Z_i+z\le W+s\le Z_i+z+\alpha)ds$$

$$+ I(Y_i<0) \int_{Y_i+t}^{t} P^{\zeta_i}(Z_i+z\le W+s\le Z_i+z+\alpha)ds\} |X_i\Gamma(Z_i-Y_i,t)|dt$$

$$= R_{11} + R_{12} \quad \text{say,}$$

where $\zeta_i = (X_i,Y_i,Z_i)$. Now

$$E^{\zeta_i}|V_i| < E^{\zeta_i}|W-T_i| + E^{\zeta_i}|T_i-Z_i|$$

$$= E|W-T_i| + E^{\zeta_i}|T_i-Z_i|$$

$$< E|W| + E|T_i| + E^{\zeta_i}|T_i-Z_i|$$

$$< 1 + E|T_i| + E^{\zeta_i}|T_i-Z_i| \quad .$$

Using this, we have

$$(5.15) \qquad |R_{11}| < C\sum_{i=1}^{n} E\int [|Y_i| + 2|t|]|X_i\Gamma(Z_i-Y_i,t)|dt$$

$$+ C\sum_{i=1}^{n} E|T_i|E\int |X_i\Gamma(Z_i-Y_i,t)|dt$$

$$+ C\sum_{i=1}^{n} E\int |T_i-Z_i|\ |X_i\Gamma(Z_i-Y_i,t)|dt$$

$$< C\sum_{i=1}^{n} \{E|X_iY_i(Z_i-Y_i)| + E|X_i(Z_i-Y_i)^2|$$

$$+ C\sum_{i=1}^{n} E|X_i(Z_i-Y_i)|E|T_i|$$

$$+ C\sum_{i=1}^{n} E|X_i(Z_i-Y_i)(T_i-Z_i)|$$

$$< C\beta \quad .$$

By Proposition 4.1,

$$(5.16) \qquad |R_{12}| < \lambda\alpha^{-1}\sum_{i=1}^{n} E\int |Y_i|(\alpha+\beta)|X_i\Gamma(Z_i-Y_i,t)|dt$$

$$= \lambda\alpha^{-1}(\alpha+\beta)\sum_{i=1}^{n} E|X_iY_i(Z_i-Y_i)|$$

$$< C\lambda\alpha^{-1}(\alpha+\beta)\beta \quad .$$

Combining (5.14), (5.15) and (5.16), we obtain

(5.17) $$|R_1(f_{z,\alpha})| \le C\lambda[1+\alpha^{-1}(\alpha+\beta)]\beta \quad .$$

By similar arguments,

(5.18) $$|R_2(f_{z,\alpha})| \le C\lambda[1+\alpha^{-1}(\alpha+\beta)]\beta \quad .$$

Now choose $\alpha = \beta$. Then by (5.4), (5.17), (5.18) and Lemma 5.5, we obtain

(5.19) $$\|F-\Phi\|_\infty \le C\lambda\beta \quad .$$

This proves (5.1) for $p = \infty$.

By Lemma 5.4,

(5.20) $$\int |R_1(f_{z,0})|dz \le \sum_{i=1}^{n} E\int [1\wedge(|Y_i+t|+|t|)][E^{\zeta_i}|V_i|+2]|X_i\Gamma(Z_i-Y_i,t)|dt$$

which is the same as R_{11} except for the constants and therefore $\le C\beta$.

Likewise

(5.21) $$\int |R_2(f_{z,0})|dz \le C\beta \quad .$$

Combining (5.4), (5.20) and (5.21), we obtain

(5.22) $$\|F-\Phi\|_1 \le C\beta \quad .$$

This proves (5.1) for $p = 1$ and therefore theorem 5.1.

REMARKS. (1) It is reasonable to assume that λ remains bounded as $n \to \infty$. In fact $\lambda \to 1$ as $n \to \infty$ if $\max_{1\le i\le n} B^{-2}\mathrm{Var}\left(\sum_{j\in N(C_i)} X_j\right) \to 0$ as $n \to \infty$ (here λ should be λ_n, $N(C_i)$ should be $N(C_{ni})$ and B^2 should be B_n^2), which is an infinitesimality condition and is natural to assume in order for the central limit theorem to hold.

(2) In bounding $\|F-\Phi\|_\infty$, we could use $h_{z,0}$ instead of $h_{z,\alpha}$ in (5.4) and use (5.8) instead of (5.9) of Lemma 5.2. But in doing so we would end up having the quantity $\max\{1, B^{-2}\sum_{i=1}^{n} E|X_iY_i|, B^{-2}\sum_{i=1}^{n} E|X_iZ_i|\}$ as a factor in the bound.

COROLLARY 5.1. There exists an absolute constant C such that for

$1 < p < \infty$,

$$\|F-\Phi\|_p < C\theta^2 B^{-3} \sum_{i=1}^{n} \gamma_i \tag{5.23}$$

where θ is given by (2.5).

PROOF. Without loss of generality we assume $B = 1$. Let $\rho = \theta^2 \sum_{i=1}^{n} \gamma_i$. First, by the Hölder and the GM-AM inequalities we obtain $\|F-\Phi\|_p < C\lambda^{\frac{p-1}{p}} \rho$ from (5.1). It suffices to prove (5.23) for $p = \infty$. Now choose $C > 8$ so that for $\rho > \frac{1}{8}$, $\|F-\Phi\|_\infty < C\rho$ is trivially true. For $\rho < \frac{1}{8}$, we use $\tilde{B}_i > 1 - [\mathrm{Var}(\sum_{j\in N(C_i)} X_j)]^{1/2}$ which by the Hölder and the GM-AM inequalities $> 1 - [\theta^2 \sum_{j\in N(C_i)} \gamma_i]^{1/3} > 1 - \rho^{1/3} > \frac{1}{2}$, so that $\lambda < 8$. This proves the corollary.

If $X_1,\ldots,X_n$ are independent, then $\theta = 1$ and Corollary 5.1 yields the classical result.

Recall that the distance between two points $i = (i_1,\ldots,i_d)$ and $j = (j_1,\ldots,j_d)$ in $\mathbb{Z}^d$ is defined by $|i-j| = \max_{1<\ell<d} |i_\ell - j_\ell|$ and the distance between two subsets A and B of $\mathbb{Z}^d$ is defined by $\rho(A,B) = \inf\{|i-j| : i\in A,\ j\in B\}$. A set of random variables $\{X_i : i\in\mathbb{Z}^d\}$ is said to be an m-dependent random field if X_A and X_B are independent whenever $\rho(A,B) > m$ where A and B are finite sets.

Suppose $\{X_i : i\in\mathbb{Z}^d\}$ is an m-dependent random field with $EX_i = 0$ and $E|X_i|^3 = \gamma_i < \infty$. If $\{X_1,\ldots,X_n\} = \{X_i : i\in\Lambda\}$ where $\Lambda \subset \mathbb{Z}^d$, then in applying Theorem 5.1 or Corollary 5.1, we have $A_i = \{j : |j-i| < m\} \cap \Lambda$, $B_i = \{j : |j-i| < 2m\} \cap \Lambda$, $C_i = \{j : |j-i| < 3m\} \cap \Lambda$ and $N(C_i) = \{j : |j-i| < 5m\} \cap \Lambda = N^*(C^i)$. Therefore

$$\theta < (10m+1)^d \quad . \tag{5.24}$$

Now suppose further that $\{X_i : i\in\mathbb{Z}^d\}$ is stationary and $\Lambda = \{i\in\mathbb{Z}^d : |i| < N\}$.

Assume $EX_i(\sum_{|j-i|\leq m} X_j) = \tau^2 > 0$. Then $B^2 \sim \tau^2|\Lambda|$. Let $\gamma_i = \gamma$ for all i. The following corollary follows from Corollary 5.1.

COROLLARY 5.2. There exists an absolute constant C such that for sufficiently large N,

$$||F - \Phi||_p \leq C(10m+1)^{2d} \gamma\tau^{-1}|\Lambda|^{-1/2}, \quad 1 \leq p \leq \infty .$$

Finally we show that Theorem 5.1 gives a rate of $n^{-1/4}$ if $X_1, \ldots, X_n$ is the 1-dependence sequence example of Erickson (1974). Recall that the rate of $n^{-1/4}$ is the correct one. It is not difficult to see that in this example if X_i is independent of all the other random variables then $A_i = B_i = C_i = N(C_i) = \{i\}$ and $X_i = Y_i = Z_i = T_i$. On the other hand if $X_i = -X_{i-1}$ or $-X_{i+1}$, then $A_i = B_i = C_i = N(C_i) = \{i, i-1\}$ or $\{i, i+1\}$ accordingly. In this case $Y_i = Z_i = T_i = 0$ identically. Consequently

$$\begin{aligned}\beta &\leq C\gamma \text{ (number of } X_i \text{ which is independent of all the other random variables)} \\ &= C\gamma B_n^2\end{aligned}$$

where $B_n^2 = \text{Var}(\sum_{i=1}^{n} X_i)$ and $\gamma = E|X_i|^3$ for all i. Similarly, $\lambda \leq C$ for sufficiently large. Hence for $1 \leq p \leq \infty$,

$$\begin{aligned}||F - \Phi||_p &\leq CB_n^{-3}\gamma B_n^2 = C\gamma B_n^{-1} \\ &\leq C\gamma n^{-1/4}\end{aligned}$$

for sufficiently large n.

Acknowledgement. The author would like to thank Andrew Barbour for simplifying some of the steps in the proof of the concentration inequality in Section 4.

REFERENCES

BERK, K. N. (1973). A central limit theorem for m-dependent random variables with unbounded m. Ann. Probab. 1, 352-354.

BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of the independent variates. Trans. Amer. Math. Soc. 49, 122-136.

BOLTHAUSEN, E. (1982). The Berry-Esseen theorem for strongly mixing Harris recurrent Markov chains. Z. Wahrscheinlichkeitstheorie verw. Gebiete 60, 283-289.

BOLTHAUSEN, E. (1982). On the central limit theorem for stationary mixing random fields. Ann. Probab. 4, 1047-1050.

BOLTHAUSEN, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. Z. Washrscheinlichkeitstheorie verw. Gebiete 66, 379-386.

CHEN, L. H. Y. (1978). Two central limit problems for dependent random variables. Z. Washrscheinlichkeitstheorie verw. Gebiete 43, 223-243.

CHEN, L. H. Y. (1979). Stein's method in limit theorems for dependent random variables. Proc. First Franco-SEA Math. Conf., SEA Bull. Math. (Special Issue), 36-50.

CLIFF, A. D. and ORD, J. K. (1981). Spatial Processes: Models and Applications. Pion, London.

DAVID, F. N. (1972). Measurement of diversity. Proc. Sixth Berkeley Symp. Math. Statist. Probab. 1, 631-648.

DIACONIS, P. (1977). The distribution of leading digits and uniform distribution mod 1. Ann. Probab. 5, 72-81.

DOBRUSHIN, R. L. (1968). The description of the random field by its conditional distribution. Theory Probab. Appl. 13, 201-229.

EGOROV, V. A. (1970). Some limit theorems for m-dependent random variables. Litovsk. Matem. Sb. 10, 51-59. (In Russian).

ERICKSON, R. V. (1973). On an L_p version of the Berry-Esseen theorem for independent and m-dependent variables. Ann. Probab. 1, 497-503.

ERICKSON, R. V. (1974). L_1 bounds for asymptotic normality of m-dependent sums using Stein's technique. Ann. Probab. 2, 522-529.

ESSEEN, C. G. (1945). Fourier analysis of distribution functions, a mathematical study of the Laplace-Gaussian law. Acta Math. 77, 1-125.

ESSEEN, C. G. (1958). On mean central limit theorems. Kungl. Tekn. Högsk. Handl. Stockholm 121, 31pp.

GUYON, X. and RICHARDSON, S. (1984). Vitesse de convergence du thérème de la limite centrale pour des champs faiblement dépendants. Z. Washrscheinlichkeitstheorie verw. Gebiete 66, 297-314.

HO, S. T. and CHEN, L. H. Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. Ann. Probab. 6, 231-249.

HOLLEY, R. A. and STROOCK, D. W. (1976). Applications of the stochastic Ising model to the Gibbs states. Comm. Math. Phys. 48, 249-265.

LEONENKO, N. N. (1975). Estimate of the rate of convergence in the central limit theorem for m-dependent random fields. Mat. Zametki 17, 129-132. (In Russian).

LOH, W. Y. (1975). On the normal approximation for sums of mixing random variables. M. Sc. Thesis. University of Singapore.

MAEJIMA, M. (1978). A non-uniform estimate in the central limit theorem for m-dependent random variables. Keio Engineering Reports 31, 15-20.

MANTEL, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Research 27, 209-220.

NEADERHOUSER, C. C. (1978). Limit theorems for multiply indexed mixing random variables, with application to Gibbs random fields. Ann. Probab. 6, 207-215.

NEADERHOUSER, C. C. (1978). Some limit theorems for random fields. Comm. Math. Phys. 61, 293-305.

PETROV. V. V. (1970). On the central limit theorem for m-dependent variables. Select. Transl. Math. Statist. Probab. 9, 83-88.

PHILIPP, W. (1969). The remainder in the central limit theorem for mixing stochastic processes. Ann. Math. Statist. 40, 601-609.

PRAKASA RAO, B. L. S. (1981). A non-uniform estimate of the rate of convergence in the central limit theorem for m-dependent random fields. Z. Wahrscheinlichkeitstheorie verw. Gebiete 58, 247-256.

RIAUBA, B. (1980). On the remainder term in the central limit theorem for m-dependent random fields. Litovsk. Mat. Sb. 20, 157-163. (In Russian).

SCHNEIDER, E. (1981). On the speed of convergence in the random central limit theorem for φ-mixing processes. Z. Washrscheinlichkeitstheorie verw. Gebiete 58, 125-138.

SEN, A. (1976). Large sample-size distribution of statistics used in testing for spatial correlation. Geographical Analysis 9, 175-184.

SHAPIRO, C. P. and HUBERT, L. (1979). Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. Ann. Statist. 7, 788-794.

SHERGIN, V. V. (1979). On the convergence rate in the central limit theorem for m-dependent random variables. Theory Probab. Appl. 24, 782-796.

STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. Sixth Berkeley Symp. Math. Statist. Probab. 2, 583-602.

STEIN, C. (1986). Lectures on the Approximate Computation of Expectations. IMS monograph series. To appear.

TAKAHATA, H. (1981). L_∞-bound for asymptotic normality of weakly dependent summands using Stein's result. Ann. Probab. 9, 676-683.

TAKAHATA, H. (1983). On the rates in the central limit theorem for weakly dependent random fields. Z. Wahrscheinlichkeitstheorie verw. Gebiete 64, 445-456.

TIKHOMIROV, A. N. (1980). On the convergence rate in the central limit theorem for weakly dependent random variables. Theory Probab. Appl. 25, 790-809.

Department of Mathematics
National University of Singapore
Lower Kent Ridge Road
Singapore 0511
Republic of Singapore

LHYC/jvs