Stein’s Method for Diffusion Approximations

A.D. Barbour
Institut für Angewandte Mathematik, Universität Zürich, Rämistrasse 74, CH-8001 Zürich, Switzerland

Summary. Stein’s method of obtaining distributional approximations is developed in the context of functional approximation by the Wiener process and other Gaussian processes. An appropriate analogue of the one-dimensional Stein equation is derived, and the necessary properties of its solutions are established. The method is applied to the partial sums of stationary sequences and of dissociated arrays, to a process version of the Wald-Wolfowitz theorem and to the empirical distribution function.

1. Introduction

In 1970, Stein introduced a new technique for obtaining rates of convergence in the neighbourhood of a normal limit. The heart of the technique is an ingenious method of estimating the discrepancy between the expectation of a smooth function \( g \) of a random variable \( W \) and the expectation of the same function \( g \) of a standard normal random variable \( N \). The idea is to find a function \( f \) such that

\[
g(w) - \mathbb{E}g(N) = f'(w) - wf(w) \tag{1.1}
\]

and then to use that structure of \( W \) which makes a normal limit plausible to estimate

\[
\mathbb{E}\{f'(W) - Wf(W)\},
\]

and hence

\[
\mathbb{E}g(W) - \mathbb{E}g(N). \tag{1.2}
\]

The technique, described in detail in Stein (1986), has subsequently been widely exploited, and is most effective in situations where a certain symmetry, rather than an evolution with time, characterizes the structure of \( W \).

One of the attractive features of the method is that it works equally well near limits other than the normal. Chen (1975) demonstrated that the method can be used near a Poisson limit. In the context of sums of dependent zero-one random variables, it has emerged as one of the most powerful and flexible methods available for estimating the accuracy of Poisson approximation in the total variation metric. Other limiting distributions have also been considered.

These results all concern the approximation of distributions on \( \mathbb{R} \). However, it has recently been shown in Barbour (1988) and in Arratia et al. (1989) that it is
possible to use Stein’s method to obtain rates of convergence in the total variation metric when making Poisson process approximation. The approach taken in Barbour (1988) is developed in Barbour, Holst and Janson (in preparation) into a direct method for Poisson process approximation, using an appropriate representation analogous to (1.1), which is in many ways as easy to use as the method for one-dimensional Poisson approximation. The aim of this paper is to explore what can be achieved using the same ideas in the context of approximation by Gaussian processes.

In Sect. 2, an analogue of Eq. (1.1) for a Wiener process in introduced, and the necessary properties of the solutions $f$ in terms of those of $g$ are established. The set of test functions $g$ for which the method is then directly applicable depends on the choice of norm on the function space under consideration—there is an obvious choice on $C[0, 1]$, but not necessarily on $D[0, \infty)$, and when considering empirical distribution functions, even on $[0, 1]$, the usual norm may not be the most useful—since they are required to be smooth with respect to the chosen norm, and it is also necessary to impose growth conditions upon them, reflecting the conditions under which approximations are to be established. The sets of test functions considered are rich enough to ensure that the estimates obtained for the process version of (1.2) imply a corresponding estimate of the distance between the process and the limiting Wiener process, with respect to a certain metric. This metric is not the Skorohod metric. However, it is also shown that, in most cases, the estimates obtained are at least enough to imply weak convergence in the usual sense.

Section 3 is concerned with applications in a variety of contexts, in which, from time to time, Gaussian processes other than the Wiener process arise. The problems addressed concern partial sums derived from (stationary) mixing random processes and random fields and from dissociated arrays, a process version of the Wald-Wolfowitz theorem, and the empirical distribution function of a sequence of independent identically distributed random variables, together with generalizations relaxing (to some extent) each of these restrictions.

In Götzte (1989), Stein’s method is used to obtain rates of convergence in the multivariate central limit theorem in $\mathbb{R}^k$, $1 \leq k < \infty$. The Stein equation used by Götzte is of the same form as Eq. (2.1) below, with $\mathcal{A}$ the generator of a $k$-dimensional Ornstein-Uhlenbeck process, and his solution $\psi_o$ of the equation corresponds, by change of variable, to the general expression (2.20). By an inductive argument reminiscent of that in Bolthausen (1984), he is for instance able to show that

$$|\mathbb{P}[S_n \in C] - \mathbb{P}[\mathcal{A}(0, I_k) \in C]| \leq C_k \Gamma_n$$

for any measurable convex set $C \subset \mathbb{R}^k$, where $S_n$ is a sum of independent random vectors, $\Gamma_n$ is the corresponding Lyapounov ratio and $C_k$ can be explicitly estimated by $C_k < 10 + 147.83k^{1/2}$.

2. Methods

The first step in using Stein’s method to derive rates of convergence to a Wiener process limit is to construct an analogue of (1.1). As observed in Barbour (1988),
one way of doing so is to find a Markov process whose equilibrium distribution is Wiener measure. If \( \mathcal{A} \) is the generator of such a process, the equation
\[
\mathcal{A}f = g - \mathbb{E}_g(Z),
\]
where \( Z \) denotes a standard Brownian motion, can be used as a Stein equation. Indeed, (1.1) can be thought of in this way, if \( f \) is replaced by \( f' \), since the LHS is then of the form \( \mathcal{A}(f)(w) \), where \( \mathcal{A} \) is the generator of an Ornstein-Uhlenbeck process with equilibrium distribution \( \mathcal{N}(0, 1) \).

In order to make use of this idea, the construction of Brownian motion using the Schauder functions (Lamperti 1966, p. 104, Theorem 1) can be extended to define an appropriate process. Let \( \{X_k\}_{k \geq 0} \) be a collection of independent, identically distributed Ornstein-Uhlenbeck processes on \([0, \infty)\), as in the one-dimensional case, with equilibrium distribution \( \mathcal{N}(0, 1) \), and define
\[
W(t, u; \omega) = \sum_{k \geq 0} X_k(u; \omega)S_k(t); \quad 0 \leq t \leq 1, \quad u \geq 0,
\]
where the Schauder functions \( S_k \) are defined by
\[
S_0(t) = t; \quad S_k(t) = \int_0^t H_k(u)du, \quad k \geq 1,
\]
and where, for \( 2^n \leq k < 2^{n+1} \),
\[
H_k(u) = 2^{n/2} \left[ I[2^{-n}k - 1 \leq u \leq 2^{-n}(k + 1/2) - 1] - I[2^{-n}(k + 1/2) - 1 \leq u \leq 2^{-n}(k + 1) - 1] \right].
\]
Note that the sum in (2.2) converges a.s. uniformly in \([0, 1] \times [0, U]\) for each \( U > 0 \), by an extension of the argument in Lamperti (1966), so that \( W \) is a.s. continuous as a function of \((t, u)\), is Gaussian, and, for each \( u, W(\cdot, u) \) is distributed according to Wiener measure. Furthermore, letting \( \mathcal{F}_u = \sigma(W(\cdot, v), v \leq u) \), \( W \) inherits from the Ornstein-Uhlenbeck processes the property that
\[
W(\cdot, u + v) - e^{-v}W(\cdot, u) \overset{D}{=} \sigma(v)Z(\cdot)
\]
where \( \sigma^2(v) = 1 - e^{-2v} \) and \( Z \) is a standard Brownian motion, independent of \( \mathcal{F}_u \). This makes computing the generator of \( W \), considered as a Markov process on \( C \subset D = D[0, 1] \) with filtration \( (\mathcal{F}_u, u \geq 0) \), simpler.

Let \( L \) be the Banach space of those continuous functions \( f: D \to \mathbb{R} \) for which the norm
\[
\| f \|_L = \sup_{w \in D} |f(w)|/(1 + \|w\|^3) < \infty,
\]
where \( \| . \| \) denotes the sup norm on \( D \). For \( f \in L \), define
\[
(T_u f)(w) = \mathbb{E}\{ f(W(\cdot, u))W(\cdot, 0) = w \} = \mathbb{E}\{ f(we^{-u} + \sigma(u)Z) \},
\]
and observe that \( T_u f \in L \), and that \( \lim(L)_{t \to 0} T_u f = f \), so that \( (T_u)_{u \geq 0} \) is a strongly continuous semi-group on \( L \). Let \( M \subset L \) consist of those twice Fréchet differenti-
able functions \( f \) which satisfy
\[
\| D^2 f(w + h) - D^2 f(w) \| \leq k_f \| h \| ,
\]
for some constant \( k_f \), uniformly in \( w, h \in D \); here, \( D^k f \) denotes the \( k \)-th derivative of \( f \), the norm of a \( k \)-linear form \( B \) on \( L \) is defined to be \( \| B \| = \sup_{\| h \| \leq 1} |B[h^{(k)}]| \), and \( h^{(k)} \) denotes the \( k \)-tuple \((h, \ldots, h)\). As a consequence of (2.5), it follows that, for each \( f \in M \), there exists a constant \( K_f \) such that
\[
\| Df(w) \| < K_f (1 + \| w \|^2); \quad \| D^2 f(w) \| < K_f (1 + \| w \|);
\]
(2.6)
\[
| f(w + h) - f(w) - Df(w)[h] - \frac{1}{2} D^2 f(w)[h, h] | \leq K_f \| h \|^3 ,
\]
uniformly in \( w, h \in D \), and that
\[
\| f \|_M = \sup_{w \in D} |f(w)|/(1 + \| w \|^3) + \sup_{w \in D} \| Df(w) \|/(1 + \| w \|^2)
\]
\[
+ \sup_{w \in D} \| D^2 f(w) \|/(1 + \| w \|) + \sup_{w, h \in D} \{ \| D^2 f(w + h) - D^2 f(w) \|/\| h \| \} \]
(2.7)
can be taken as a norm on \( M \). Then, for \( f \in M \), it follows from (2.6) that
\[
| (T_u f)(w) - f(w) - Edf(w)[\sigma(u)Z - w(1 - e^{-u})] \\
- \frac{1}{2} Ed^2 f(w)[\{\sigma(u)Z - w(1 - e^{-u})\}^{(2)}] |
\]
\[
\leq \frac{1}{2} k_f \| \sigma(u)Z - w(1 - e^{-u}) \|^3 \leq K_1 (1 + \| w \|^3) u^{3/2} ,
\]
so that
\[
u^{-1} (T_u f - f)(w) + uDf(w)[w] - uEd^2 f(w)[Z^{(2)}] \leq K_2 (1 + \| w \|^3) u^{1/2} ,
\]
for constants \( K_f \) depending only on \( f \). Thus each \( f \in M \) is in the domain of the generator \( \mathcal{A} \) of \( W_t \), and
\[
(\mathcal{A} f)(w) = - Df(w)[w] + Ed^2 f(w)[Z^{(2)}] .
\]
(2.9)
The last expression in the generator is somewhat cumbersome. However, if \( Y \) is another random element defined on the same probability space as \( Z \), by the bilinearity of \( D^2 f(w) \),
\[
Ed^2 f(w)[Z^{(2)}] - Ed^2 f(w)[Y^{(2)}] \leq Ed^2 f(w)[(Z - Y)^{(2)}] + 2D^2 f(w)[Y, Z - Y] \]
\[
\leq K_f (1 + \| w \|) E \{ \| Z - Y \|^2 + \| Z - Y \| \| Y \| \} .
\]
(2.10)
Hence, using the representation
\[
Z(t) = \sum_{k \geq 0} Z_k S_k(t) ,
\]
where the \((Z_k)_{k \geq 0}\) are independent \(\mathcal{N}(0, 1)\) random variables, we also have

\[
(\mathcal{A} f)(w) = -Df(w)[w] + \sum_{k \geq 0} D^2f(w)[S_k^{(2)}].
\] (2.11)

Furthermore, when exploiting (2.9) or (2.11) using Stein’s method, it often turns out to be natural to approximate the process under consideration by a pre-limiting Gaussian process. Thus, for example, for the normalized partial sums of \(n\) independent and identically distributed random variables, a natural approximating generator is given by

\[
(\mathcal{A}_n f)(w) = -Df(w)[w] + \mathbb{E} D^2f(w)[Z_n^{(2)}],
\] (2.12)

where

\[
Z_n(t) = n^{-1/2} \sum_{k=1}^{[nt]} Z_k,
\] (2.13)

and the \((Z_k)_{k \geq 0}\) are once again independent \(\mathcal{N}(0, 1)\) random variables. The process corresponding to \(\mathcal{A}_n\) can for instance be realized by

\[
W_n(t, u; \omega) = n^{-1/2} \sum_{k=1}^{n} X_k(u; \omega) J_{k/n}(t),
\] (2.14)

with the \(\{X_k\}_{k=1}^{n}\) distributed as for (2.2), where

\[
J_\alpha(t) = \begin{cases}
1 & \text{if } t \geq \alpha, \\
0 & \text{if } t < \alpha.
\end{cases}
\] (2.15)

Thus the expectation in (2.12) can be computed, yielding

\[
(\mathcal{A}_n f)(w) = -Df(w)[w] + n^{-1} \sum_{k=1}^{n} D^2f(w)[J_{k/n}^{(2)}].
\] (2.16)

This last result suggests a further representation for \(\mathcal{A}\), in addition to (2.9) and (2.11). Let \(M' \subset M\) consist of those \(f \in M\) for which

\[
\lim_{n \to \infty} \sup_{w \in D} \left\{ \int_0^1 \left[ D^2f(w)[J_{t/n}^{(2)}] - D^2f(w)[J_{[nt]/n}^{(2)}] \right] dt \|(1 + \|w\|^2) \right\} = 0.
\] (2.17)

Then, for \(f \in M'\),

\[
(\mathcal{A} f)(w) = -Df(w)[w] + \int_0^1 D^2f(w)[J_t^{(2)}] dt.
\] (2.18)

Thus Eq. (2.1) can be formulated using any one of the forms (2.9), (2.11) and (2.18) for \(\mathcal{A}\).

The next step is to solve (2.1), with the generator \(\mathcal{A}\) defined as above, for a suitable class of smooth functions \(g\). This is easy to do, since, for all \(t > 0\) and for all \(g \in L\),

\[
T_t g - g = \mathcal{A} \left( \int_0^t T_u g \, du \right),
\] (2.19)
by Ethier and Kurtz (1986), p. 9, Proposition 1.5. Take any $g \in L$ such that $\mathbb{E}g(Z) = 0$. Then, using (2.4) and the fact that $\mathcal{A}$ is closed (Ethier and Kurtz 1986, p. 10, Corollary 1.6), it follows immediately that

$$ f = \phi(g) = -\int_{0}^{\infty} T_{u} g \, du \tag{2.20} $$

exists and is in the domain of $\mathcal{A}$ if also, for instance,

$$ |g(w) - g(x)| \leq C_{g} \{ 1 + \|w\|^{2} + \|x\|^{2} \} \|w - x\| \tag{2.21} $$

uniformly in $w, x \in D$; and

$$ \mathcal{A} f = g - \lim_{t \to \infty} T_{t} g = g \tag{2.22} $$

This is true, in particular, for all $g \in M$ such that $\mathbb{E}g(Z) = 0$. For such $g$,

$$ \phi(g)(w + h) - \phi(g)(w) = -\int_{0}^{\infty} \{ g((w + h)e^{-u} + \sigma(u)Z) - g(we^{-u} + \sigma(u)Z) \} \, du, $$

and hence, using dominated convergence,

$$ D^{k} \phi(g)(w) = -\int_{0}^{\infty} e^{-ku} D^{k} g(we^{-u} + \sigma(u)Z) du, \quad k = 1, 2 \tag{2.23} $$

so that

$$ \phi(g) \in M \quad \text{and} \quad K_{\phi(g)} \leq CK_{g} \tag{2.24} $$

for some constant $C$, uniformly in $g \in M$.

The third step is to establish a technique for applying the method. We consider the simplest example. Suppose that random elements $(Y_{n})_{n \geq 1}$ of $D$ are defined by

$$ Y_{n}(t) = n^{-1/2} \sum_{j=1}^{[nt]} X_{j} = n^{-1/2} \sum_{j=1}^{n} X_{j} J_{jn}(t), \quad 0 \leq t \leq 1, $$

where the random variables $(X_{j})_{j \geq 1}$ are independent and identically distributed, with zero mean, unit variance and finite third moment. Then we can prove the following result.

**Theorem 1.** There exists a universal constant $C$ such that, for all $g \in M$,

$$ |\mathbb{E}g(Y_{n}) - \mathbb{E}g(Z)| \leq C n^{-1/2} \|g\|_{M} \{ \sqrt{\log n} + \mathbb{E}|X_{1}|^{3} \}. $$

**Remark.** Here and subsequently, $C$ will be used to denote a generic universal constant, not necessarily the same at each appearance.

**Proof.** Without loss of generality, let $g \in M$ satisfy $\mathbb{E}g(Z) = 0$. Then

$$ \mathbb{E}g(Y_{n}) = \mathbb{E}(\mathcal{A}f)(Y_{n}) = \mathbb{E} \{ -Df(Y_{n})[Y_{n}] + \sum_{k \geq 0} D^{2} f(Y_{n})[S_{k}^{(2)}] \}, $$

where $f = \phi(g) \in M$ and $\|f\|_{M} \leq C \|g\|_{M}$. Clearly,

$$ \mathbb{E}Df(Y_{n})[Y_{n}] = n^{-1/2} \sum_{j=1}^{n} \mathbb{E} \{ X_{j} Df(Y_{n})[J_{jn}] \}. $$
Letting \( Y_n^j = n^{-1/2} \sum_{k+j} X_k J_{kn} \), it follows by Taylor's expansion and (2.5) that
\[
|n^{-1/2} \mathbb{E} J f(Y_n)[J_{jn}] - \mathbb{E} [n^{-1/2} J f(Y_n)[J_{jn}] + n^{-2} J^2 f(Y_n)[J_{jn}^2]]| \\
\leq \frac{1}{2} n^{-3/2} \| f \|_M \mathbb{E} |X_j|^3,
\]
since \( \| J_{jn} \| = 1 \). Hence, since \( X_j \) is independent of \( Y_n^j \) and has zero mean and unit variance,
\[
|\mathbb{E} D f(Y_n)[Y_n] - n^{-1} \sum_{j=1}^n \mathbb{E} D^2 f(Y_n)[J_{jn}^2]| \leq \frac{1}{2} n^{-1/2} \| f \|_M \mathbb{E} |X_1|^3,
\]
and thus, again by Taylor's expansion,
\[
|\mathbb{E} D f(Y_n)[Y_n] - n^{-1} \sum_{j=1}^n \mathbb{E} D^2 f(Y_n)[J_{jn}^2]| \leq \frac{3}{2} n^{-1/2} \| f \|_M \mathbb{E} |X_1|^3. \tag{2.25}
\]
This actually proves that, for all \( g \in M \),
\[
|\mathbb{E} g(Y_n) - \mathbb{E} g(Z_n)| \leq C n^{-1/2} \| g \|_M \mathbb{E} |X_1|^3,
\]
for the Gaussian process \( Z_n \) defined by (2.13). Since also \( Z_n \) and \( Z \) can be realized together, by first realizing \( Z \) and then defining \( Z_n(j/n) = Z(j/n) \) for all \( 0 \leq j \leq n \), in such a way that
\[
\mathbb{E} \left\{ \sup_{0 \leq t \leq 1} |Z_n(t) - Z(t)| \right\} \leq C n^{-1/2} \sqrt{\log n}, \tag{2.26}
\]
the above argument yields
\[
|\mathbb{E} g(Y_n) - \mathbb{E} g(Z)| \leq C n^{-1/2} \| g \|_M (\sqrt{\log n + \mathbb{E} |X_1|^3}) \tag{2.27}
\]
for all \( g \in M \), for some universal constant \( C \). \( \Box \)

Remarks. 1. The computations required to establish (2.27) are as easy as those required to prove the one-dimensional result, that
\[
|\mathbb{E} g(Y_n(1)) - \mathbb{E} g(N)| \leq C k_g n^{-1/2} \mathbb{E} |X_1|^3,
\]
for smooth functions \( g: \mathbb{R} \rightarrow \mathbb{R} \) and suitable constants \( k_g \). The advantage of the more general setting is that, with no more effort, rates of approximation are immediately established for quite a broad range of functionals of the whole path \( Y_n \).

2. The choice of norm \( \| . \| \) on \( D \) actually plays little formal rôle in the preceding discussion: all that is important is that \( \mathbb{E} |Z|^3 < \infty \). However, different choices of norm lead to classes of smooth functions different from \( M \). The sharper the norm, the more functions \( f \) satisfy the condition (2.5), and hence the wider the scope of the approximations, provided always that the paths both of the approximating processes and of the limit process—here, Brownian motion—have finite norm a.s. However, (2.26) suggests that the \( \sup \) norm on \( D \) is also a little too strong, when aiming for the natural \( n^{-1/2} \) convergence rate. In particular, functions satisfying (2.5) for certain integral norms—for example, the \( L_1 \) norm—yield an \( n^{-1/2} \).
convergence rate. More simply, the sup norm on $D$ can be retained, by further restricting the class of functions $g$ to those elements of $M$ for which
\begin{equation}
\left| \int_0^1 \{ D^2 g(w)[J^{(2)}_r] - D^2 g(w)[J^{(2)}_s] \} \, dt \right| \leq k_g n^{-1/2}(1 + \|w\|^3),
\end{equation}
for some constant $k_g$, so that an order $n^{-1/2}$ approximation follows directly from (2.18), (2.24) and (2.25). A convenient condition, sufficient to ensure (2.28) and compatible with estimation in terms of $\|g\|_M$, is that
\begin{equation}
\sup_{w \in D} |D^2 g(w)[J_r , J_s - J_r]| \leq C \|g\|_M |t - s|^{1/2},
\end{equation}
for all $r, s, t$.

3. In place of $Y_n$, one could define $C$-valued approximands $\tilde{Y}_n$ by linearly interpolating between the points $(j/n, Y_n(j/n))_{j=0}^n$, giving, by a similar argument,
\begin{equation}
\left| \mathcal{E} Df(\tilde{Y}_n)[\tilde{Y}_n] \right| + \mathcal{E} D^2 f(\tilde{Y}_n)[\tilde{Z}_n^{(2)}] \leq C n^{-1/2} \|g\|_M \mathbb{E}|X_1|^3,
\end{equation}
where $\tilde{Z}_n$ is the linearly interpolated modification of $Z_n$, and is independent of $\tilde{Y}_n$. This implies that
\begin{equation}
|\mathcal{E} g(\tilde{Y}_n) - \mathcal{E} g(\tilde{Z}_n)| \leq C n^{-1/2} \|g\|_M \mathbb{E}|X_1|^3
\end{equation}
for all $g \in M$. Now, for each fixed $w \in D$, by the bilinearity of $D^2 f$, and by realizing $Z$ and $Z_n$ together as for (2.26), we have
\begin{equation}
\mathcal{E} D^2 f(w)[Z^{(2)}] - \mathcal{E} D^2 f(w)[\tilde{Z}_n^{(2)}] = 2 \mathcal{E} D^2 f(w)[Z - \tilde{Z}_n, Z] - \mathcal{E} D^2 f(w)[(Z - \tilde{Z}_n)^{(2)}],
\end{equation}
in which the first term is zero, since $Z - \tilde{Z}_n$ and $Z$ are independent and $Z - \tilde{Z}_n$ has mean 0 in $D$, and the second is $o(n^{-1/2})$. Hence, for the linear approximands $\tilde{Y}_n$,
\begin{equation}
|\mathcal{E} g(\tilde{Y}_n) - \mathcal{E} g(Z)| \leq C n^{-1/2} \|g\|_M \mathbb{E}|X_1|^3
\end{equation}
for all $g \in M$.

4. The preceding arguments could all be carried through on the space of functions $g \in M(\delta) \subset L$, for which
\begin{equation}
\sup_{w, h \in D} \left\{ \|D^2 g(w + h) - D^2 g(w)\|, \|h\|^{-\delta} \right\} < \infty,
\end{equation}
for some $0 < \delta < 1$. The conclusion corresponding to (2.27) would then be
\begin{equation}
|\mathcal{E} g(Y_n) - \mathcal{E} g(Z)| \leq C n^{-\delta/2} \|g\|_{M(\delta)} \mathbb{E}|X_1|^{2+\delta}
\end{equation}
for all $g \in M(\delta)$, for some universal constant $C = C(\delta)$, requiring less of the summands $X_j$ and less smoothness of $g$, at the expense of a more restrictive growth condition on $g$ and a less precise error estimate.

5. Instead of looking at the space $D[0, 1]$, one could equally well look at $D[0, \infty)$. Here a candidate norm would be
\begin{equation}
\|w\| = \sup_{t \geq 0} \{|w(t)|/(1 + t)\}
\end{equation}
on that part of $D[0, \infty)$ where it is finite.
6. Define a function \( d \) on the space of Borel probability measures over \( D[0, 1] \) with the Skorohod topology by
\[
d(P, Q) = \sup_{g \in M^0} \| \int g \, dP - \int g \, dQ \|_{M^e},
\]
where \( M^0 \subset M \) consists of those functions \( g \) for which
\[
\| g \|_{M^o} = \| g \|_M + \sup_{w \in D} |g(w)| + \sup_{w \in D} \| Dg(w) \| + \sup_{w \in D} \| D^2 g(w) \| < \infty.
\]
d is clearly symmetric, positive and finite, and satisfies the triangle inequality. Furthermore, considering all complex functions \( g \) with real and imaginary parts in \( M^0 \) of the form
\[
g(w) = \exp \left\{ i \sum_{j=1}^{k} \theta_j w(t_j) \right\},
\]
for \( k \in \mathbb{N}, \theta \in \mathbb{R}^k \) and \( t \in [0, 1]^k \), it follows that, if \( d(P, Q) = 0 \), \( P \) and \( Q \) have the same finite dimensional distributions, and are hence identical. Thus \( d \) is a metric over the Skorohod-Borel measures on \( D \), and estimates such as (2.27) imply statements about the closeness of the distributions \( P \) and \( Q \) induced by \( Y_n \) and \( Z \), measured with respect to the \( d \)-metric.

7. As is demonstrated in the examples in Sect. 3, the basic method is not restricted to approximation by a Wiener process. To see that other processes can be used, suppose that \( (\phi_j)_{j \geq 0} \) are a family of functions in \( D \) which are orthonormal with respect to Lebesgue measure, and let \( (\lambda_j)_{j \geq 0} \) be a sequence of constants satisfying \( \sum \lambda_j^2 < \infty \). Then the Gaussian random function \( Z \) given by
\[
Z(t) = \sum_{j \geq 0} Z_j \phi_j(t) \lambda_j
\]
exists in the \( L_2 \) sense, and has covariance function
\[
\rho(s, t) = \sum_{j \geq 0} \phi_j(s) \phi_j(t) \lambda_j^2.
\]  
(2.30)
Now set
\[
W(t, u) = \sum_{j \geq 0} X_j(u) \phi_j(t) \lambda_j,
\]
with the \( X_j \)'s defined as for (2.2). Then, as before,
\[
W(., u + v) - e^{-v} W(., u) \equiv \sigma(v) Z(.)
\]
and, for suitably smooth functions \( f \), the generator \( \mathcal{A} \) of \( W \), considered as a Markov process with \( u \) as time, satisfies
\[
(\mathcal{A} f)(w) = -Df(w)[w] + \sum_{j \geq 0} \lambda_j^2 D^2 f[\phi_j^{(2)}].
\]  
(2.31)
Thus, approximation by any Gaussian process \( W \) with paths in \( D \) and with
covariance function representable as in (2.30) can be analysed in much the same way.

8. There is no need to restrict attention to processes with a one-dimensional time parameter. For instance, $D\{[0, 1]^d\}$ with the sup norm can be used as an underlying function space in the same way as $D$ was used above.

9. Functions $g \in M$ include functions of the form

$$g(w) = \int_0^1 k(t, w(t))d\mu(t),$$

where $\mu$ is a probability measure on $[0, 1]$, and $k$ satisfies

$$\sup_t |k(t, 0)| + \sup_t |\partial k(t, 0)| + \sup_t |\partial^2 k(t, 0)| < K$$

$$\sup_{x, y \in \mathbb{R}} \sup_t |\partial^2 k(t, x) - \partial^2 k(t, y)| < K|x - y|$$

for some $K < \infty$, where $\partial$ denotes differentiation with respect to the second argument. Suitably smooth functions of such functions are also included in $M$. In particular, the function $f(w) = w(s)w(t)$ is of this form, and can be used with expressions such as (2.31) to identify the covariance function of the limiting process from its generator.

10. The set of functions $M$ is considerably smaller than the set of all bounded, Skorohod-continuous functions, so that a statement such as (2.27) does not of itself imply weak convergence of $\tilde{Y}_n$ to $Z$ in the usual sense in $D$, although convergence of the finite dimensional distributions follows by using the characteristic functions of Remark 6. However, the following result shows that weak convergence is indeed a consequence, under rather mild conditions.

**Theorem 2.** Let $(Y_n)_{n \geq 0}$ be piecewise constant random elements of $D$ such that each interval of constancy of $Y_n$ is of length at least $r_n$, and let $Z$ be a random element of $C[0, 1]$. Suppose that

$$|E g(Y_n) - E g(Z)| \leq C_{\tau_n} \|g\|_{M^0}$$

for all $g \in M^0$ satisfying the smoothness condition

$$\sup_{w \in D} |D^2 g(w)\{J_r, J_s - J_r\}| \leq C_1 \|g\|_{M^0} |t - s|^{1/2},$$

(2.32)

for some fixed $C_1 > 0$. Then, if $\tau_n \{ - \log r_n\}^2 \to 0$, it is possible to construct random elements $(\tilde{Y}_n)_{n \geq 0}$ and $\tilde{Z}_n$ on the same probability space, in such a way that $\tilde{Y}_n \overset{d}{=} Y_n,$ $\tilde{Z}_n \overset{d}{=} Z$ and $\|\tilde{Y}_n - \tilde{Z}_n\| \to 0$ a.s.

**Proof.** Let $\phi: \mathbb{R} \to [0, 1]$ be a three times continuously differentiable decreasing function satisfying

$$\phi(y) = \begin{cases} 1, & \text{if } y \leq 0; \\
0, & \text{if } y \geq 1, \end{cases}$$

for some $K < \infty$, where $\partial$ denotes differentiation with respect to the second argument.
and define $\phi_{p, \eta}: \mathbb{R}^+ \to [0, 1]$ by $\phi_{p, \eta}(y) = \phi(\eta(y - \rho))$: then clearly, for some constant $K > 0$,

$$
\sup_{y > 0} |\phi^{(k)}_{p, \eta}(y)| \leq K\eta^{-k}, \quad k = 1, 2, 3.
$$

Then, for any $\varepsilon, p > 0$ and $s \in C[0, 1]$, define $g = g(\varepsilon, p, \rho, \eta, s)$ by

$$
g(y) = \phi_{p, \eta}\left\{ \left[ \frac{\varepsilon^2 + (\eta(t) - s(t))^2}{t^{p/2}} \right]^{1/p} \right\}.
$$

Routine calculation shows that $g$ is bounded and belongs to $M^0$, with $\|g\|_{M^0} \leq Cp^2\varepsilon^{-2} \eta^{-3}$, and that $g$ satisfies (2.32), with the constants $C$ and $C_1$ not depending on $\varepsilon, p, \rho, \eta, s$. Furthermore, the same is true for any finite product of such functions $g$, with constants $C$ and $C_1$ depending only on $C, C_1$ and the number of factors, and with $p^2\varepsilon^{-2} \eta^{-3}$ construed taking the largest of the $p$'s and the smallest of the $\varepsilon$'s and $\eta$'s.

Now, given $\gamma > 0$, pick a compact $K \subset C[0, 1]$ such that $\mathbb{P}[Z \in K] > 1 - \gamma$, and let $B_1, \ldots, B_L$ be a finite covering of $K$, where

$$
B_l = \{ x \in D : \| x - s_l \| < \gamma_l \}
$$

for some $s_l \in K$, the $\gamma_l$ being chosen so that $\mathbb{P}[Z \in \partial B_l] = 0$ and so that $\gamma/2 \leq \gamma_l \leq \gamma$.

We wish first to show that $\mathbb{P}[Y_n \in B_l] \to \mathbb{P}[Z \in B]$ for all $B$ in the algebra generated by $B_1, \ldots, B_L$. To do this, it is enough to show that $\mathbb{P}[Y_n \in B] \to \mathbb{P}[Z \in B]$ for any $B = \bigcap_{l \in L'} B_l$ with $L' \subset \{1, 2, \ldots, L\}$, because of inclusion/exclusion. So we approximate each random variable $I[Y_n \in B_l]$ above and below by functions from the family $g(\varepsilon, p, \rho, \eta, s_l) (Y_n)$, and then products of these functions will suffice to approximate the indicators $I[Y_n \in \bigcap_{l \in L'} B_l]$.

It is immediate that $x \in B_l$ implies that $g_l(x) = 1$, where

$$
g_l = g(\varepsilon\gamma_l, p, \gamma_l(1 + \varepsilon^2)^{1/2}, \eta, s_l),
$$

for all $\varepsilon, p, \eta$: hence

$$
\mathbb{P}\left[ Y_n \in \bigcap_{l \in L'} B_l \right] \leq \mathbb{E}\left\{ \prod_{l \in L'} g_l(Y_n) \right\} \leq \mathbb{E}\left\{ \prod_{l \in L'} g_l(Z) \right\} + 4C_{L'} \tau_n p^2(\varepsilon\gamma_l)^{-2} \eta^{-3}.
$$

Fixing $\varepsilon$ and choosing $p = p_n$ and $\eta = \eta_n$ in such a way that $p_n \uparrow \infty$, $\eta_n \downarrow 0$ and $\tau_n p_n^2 \eta_n^{-3} \to 0$, the second of these terms converges to zero as $n \to \infty$, while the first decreases to

$$
\mathbb{P}\left[ \bigcap_{l \in L'} \{ \| Z - s_l \| \leq \gamma_l \} \right] = \mathbb{P}[Z \in B],
$$

because $\mathbb{P}[Z \in \partial B_l] = 0$ for each $l$. Hence

$$
\lim_{n \to \infty} \sup_l \mathbb{P}\left[ Y_n \in \bigcap_{l \in L'} B_l \right] \leq \mathbb{P}[Z \in B].
$$

To approximate in the other direction, note that the finite collection of functions $s_l$ is uniformly equicontinuous, and that the functions $Y_n$ are piecewise
constant with intervals of constancy of length at least \( r_n \). Hence, for any \( 0 < \theta < 1 \), there exists \( \delta = \delta(\theta/\gamma/2) \) such that almost surely

\[
\{ \sup_t |Y_n(t) - s_t(t)| \geq \gamma \} \subset \{ \text{leb}[t: |Y_n(t) - s_t(t)| \geq \gamma(1 - \theta)] \geq \delta \wedge \frac{1}{2} r_n \}
\]

\[
\subset \left\{ \left( \int_0^1 [\gamma^2 + (Y_n(t) - s_t(t))^2]^{p/2} dt \right)^{1/p} \geq \gamma \left[ \int_0^1 [\gamma^2 + (1 - \theta)^2]^{1/2} (\delta \wedge \frac{1}{2} r_n)^{1/p} \right] \right\}.
\]

where

\[
g(Y_n) = g(\epsilon Y_n, p, \gamma Y_n).\]

Thus \( I[Y_n \in B_r] \geq g(Y_n) \), and hence

\[
P \left[ Y_n \in \bigcap_{t \in \mathbb{L}'} B_t \right] \geq \mathbb{E} \left[ \prod_{t \in \mathbb{L}'} g(Y_n) \right] = \mathbb{E} \left[ \prod_{t \in \mathbb{L}'} g(Z) \right] - 4C_{L'} \tau_n p^2 (\epsilon \gamma)^2 \eta^{-3} .
\]

Fix \( \epsilon \) and \( \theta \), and let \( p = p_n \uparrow \infty \) and \( \eta = \eta_n \downarrow 0 \) in such a way that \( \tau_n p_n^2 \eta_n^{-3} \to 0 \) and \( p_n \tau_n \to 1 \), which is possible because \( \tau_n (-\log r_n)^2 \to 0 \). Then the second term in the estimate converges to zero as \( n \to \infty \), whereas the first increases to \( \mathbb{P}[\bigcap_{t \in \mathbb{L}'} \| Z - s_t \| < \gamma(1 - \theta)] \). Hence, since \( \theta \) was arbitrary,

\[
\liminf_{n \to \infty} \mathbb{P} \left[ Y_n \in \bigcap_{t \in \mathbb{L}'} B_t \right] \geq \mathbb{P}[Z \in B],
\]

and thus \( \mathbb{P}[Y_n \in B] \to \mathbb{P}[Z \in B] \) as required.

The rest of the proof follows by the standard arguments needed to prove the Skorohod a.s. representation theorem: see, for example, Pollard (1984), pp. 72–73. \( \square \)

**Corollary.** Under the conditions of Theorem 2, \( Y_n \Rightarrow Z \) in the Skorohod topology. \( \square \)

**Remarks.** 1. In typical applications, \( \tau_n \leq n^{-\alpha} \) for some \( 0 < \alpha \leq 1/2 \) and \( r_n \asymp n^{-1} \), so that the condition \( \tau_n (-\log r_n)^2 \to 0 \) is easily satisfied. However, in limit theorems for non-identically distributed summands, this may not be automatic. For instance, in the independent case and with the usual standardization

\[
Y_n = s_n^{-1} \sum_{i=1}^n X_i \sigma_i s_i / s_n^2 ,
\]

the condition \( \tau_n (-\log r_n)^2 \to 0 \) could be expressed as

\[
\gamma_n \left( -\log \left( \min_{1 \leq i \leq n} \sigma_i / s_n \right) \right)^2 \to 0 ,
\]

for some Lyapounov estimate \( \gamma_n \). If a few of the variances of the \( X_i \) are very small, the condition may not hold. Of course, in this example, the problem can be avoided by amalgamating summands with small variance into their neighbours: nonetheless, a little extra over and above the Stein argument is needed.
2. The condition that $Y_n$ should remain constant over intervals of length $r_n$ is convenient, but not essential. The argument above can be adapted to allow the following weaker condition: that, given any $\delta > 0$, there exists a sequence $(r_n(\delta))_{n \geq 1}$ such that $\lim_{n \to \infty} P[w_{r_n}(r_n(\delta)) > \delta] = 0$ and such that $\tau_n(- \log r_n(\delta))^2 \to 0$, where $w_\cdot$ is the modulus of continuity on $D$ defined in Billingsley (1968), p. 110, (14.6) and (14.7). This is useful, for instance, when considering weak convergence for empirical processes.

Applications

3.1. Short Range Dependence

Let $X_1, \ldots, X_n$ be random variables with zero mean and finite third moment. Suppose that $G \subset \{ (i,j): 1 \leq i < j \leq n \}$ is a graph on the vertices $\{1,2,\ldots,n\}$, and define

$$N_i = \{ i \} \cup \{ j: (i,j) \in G \} \cup \{ j: (j,i) \in G \}$$

(3.1)

to be the set of neighbours of $i$. Define

$$\sigma_{ij} = \mathbb{E}(X_i X_j); \quad \sigma_i^2 = \sum_{j=1}^n \sigma_{ij}; \quad s_n^2 = \sum_{i=1}^n \sigma_i^2 ,$$

(3.2)

and let

$$Y_n = s_n^{-1} \sum_{i=1}^n X_i J_i^* ,$$

(3.3)

where the functions $J_i^* \in D$ are as yet free to be chosen, but will for convenience be assumed to satisfy $\| J_i^* \| \leq 1$. Then, for any $f \in M$, define the quantities

$$\varepsilon_1 = s_n^{-1} \sum_{i=1}^n | \mathbb{E}(X_i Df(Y_n^i)[J_i^*]) | ,$$

$$\varepsilon_2 = s_n^{-2} \sum_{i=1}^n \mathbb{E} \{ (X_i X_j - \sigma_{ij}) D^2 f(Y_n) [J_i^*, J_j^*] \} ,$$

$$\varepsilon_3 = \frac{3}{2} s_n^{-3} \sum_{i=1}^n \sum_{j,k \in N_i} | \sigma_{ij} | \| X_i X_j X_k \| f_M ,$$

(3.4)

$$\varepsilon_4 = 2 s_n^{-3} \sum_{i=1}^n \sum_{j,k \in N_i} | \sigma_{ij} | \| X_i \| f_M ,$$

$$\varepsilon_5 = s_n^{-2} \sum_{i=1}^n \sum_{j \not\in N_i} | \sigma_{ij} | \| D^2 f(Y_n) [J_i^*, J_j^*] | ,$$

where

$$Y_n^i = s_n^{-1} \sum_{k \not\in N_i} X_k J_k^* , \quad Y_n^{ij} = s_n^{-1} \sum_{k \not\in N_i \cup N_j} X_k J_k^* .$$

Then we can prove the following result.
Lemma 3.1. With the above definitions, for any \( f \in M \),
\[
\left| \mathbb{E}\{Df(Y_n)[Y_n] - s_n^{-2} \sum_{i,j} \sigma_{ij} D^2 f(Y_n)[J_i^*, J_j^*]\} \right| \leq \sum_{i=1}^{s} \varepsilon_i . \tag{3.6}
\]

Remark. This lemma describes how well the process \( Y_n \) is approximated by a process \( Z_n \) constructed as in (3.3), but with the \( X_i \)'s replaced by Gaussian random variables with the same covariance structure, in the sense that, for \( g \in M \),
\[
\mathbb{E}g(Y_n) - \mathbb{E}g(Z_n) = -Df(Y_n)[Y_n] + s_n^{-2} \sum_{i,j} \sigma_{ij} D^2 f(Y_n)[J_i^*, J_j^*]
\]
for an appropriately chosen \( f \in M \).

Proof. Proceeding as in the proof of Theorem 1, we have
\[
s_n^{-1} \mathbb{E}X_i Df(Y_n)[J_i^*] = s_n^{-1} \mathbb{E}\{X_i Df(Y_n + U)^t [J_i^*]\},
\]
where \( U^t = Y_n - Y_n^t \). Hence, by Taylor's expansion,
\[
\left| s_n^{-1} \mathbb{E}\{X_i Df(Y_n)[J_i^*]\} - s_n^{-1} \mathbb{E}\{X_i Df(Y_n^t)[J_i^*]\} \right| + s_n^{-2} \mathbb{E}\left\{X_i \sum_{j \in N_i} X_j D^2 f(Y_n^t)[J_i^*, J_j^*]\right\} \leq \frac{1}{2} s_n^{-3} \mathbb{E}\left\|X_i \left( \sum_{j \in N_i} X_j \right)^2 \right\| f_M . \tag{3.7}
\]

In a similar way, from Taylor's expansion,
\[
|s_n^{-2} \mathbb{E}\{X_i X_j D^2 f(Y_n)[J_i^*, J_j^*]\} - s_n^{-2} \mathbb{E}\{X_i X_j D^2 f(Y_n^t)[J_i^*, J_j^*]\}| \leq s_n^{-3} \mathbb{E}\left\|X_i X_j \sum_{k \in N_j \setminus N_i} X_k \right\| f_M \tag{3.8}
\]
and
\[
s_n^{-2} \mathbb{E}\{X_i X_j D^2 f(Y_n^t)[J_i^*, J_j^*]\} - \sigma_{ij} \mathbb{E}\{D^2 f(Y_n^t)[J_i^*, J_j^*]\}| \leq s_n^{-2} \mathbb{E}\{(X_i X_j - \sigma_{ij}) D^2 f(Y_n^t)[J_i^*, J_j^*]\} + s_n^{-3} |\sigma_{ij}| \mathbb{E}\left\| \sum_{k \in N_i \cup N_j} X_k \right\| f_M . \tag{3.9}
\]

Combining (3.7)–(3.9), we obtain
\[
\left| \mathbb{E}\{Df(Y_n)[Y_n] - s_n^{-2} \sum_{i,j} \sigma_{ij} D^2 f(Y_n)[J_i^*, J_j^*]\} \right| \leq \sum_{i=1}^{s} \varepsilon_i , \tag{3.10}
\]
from which the lemma follows.

In order to make use of Lemma 3.1, it is necessary to make some assumptions about the structure of \( Y_n \). Suppose, for example, that the graph \( G \) is linear, in the sense that
\[(i,j) \in G \text{ implies } (i,k) \in G \text{ and } (k,j) \in G \text{ for all } i < k < j :\]
then it is natural to take \( J_i^* = J_{s_i^{|i|}} \), where \( J_a \) is as defined for (2.14) and
\( s_t^2 = \sum_{j=1}^i \sigma_j^2 \), and to look at a one-dimensional process of partial sums. We can then prove the following theorem.

**Theorem 3.** Suppose that \( g \in M \) satisfies

\[
\sup_{w \in B} |D^2 g(w)[J_r, J_s - J_r]| \leq C_1 \| g \|_M |t - s|^{1/2} \tag{3.11}
\]

for all \( r, s, t \in [0, 1] \). Define \( Y_n = s_n^{-1} \sum_{i=1}^n X_i J_{s_i^2/s_n^2} \), and let \( Z \) denote a Wiener process. Then there exists a universal constant \( C \) such that

\[
|\mathbb{E} g(Y_n) - \mathbb{E} g(Z)| \leq C \left( \sum_{i=1}^5 \epsilon_i + C_1 (\epsilon_6 + \epsilon_7) \right),
\]

where \( \epsilon_1 - \epsilon_5 \) are as in (3.4) with \( \phi(g) \) defined in (2.20) for \( f \),

\[
\epsilon_6 = s_n^{-3} \| g \|_M \left\{ \sum_{i,j} \left| \sigma_{ij} \right| s_j^2 - s_i^2 |1/2 \right\}
\]

and

\[
\epsilon_7 = 2s_n^{-1} \max_i \sigma_i \| g \|_M.
\]

**Remark.** For many functions \( g \), the factor \( |t - s|^{1/2} \) in Condition (3.11) could be replaced by \( |t - s| \), which would lead to smaller values of \( \epsilon_6 \) and \( \epsilon_7 \). The current condition makes \( \epsilon_6 \) and \( \epsilon_7 \) comparable to \( \sum_i \epsilon_i \) in the simplest cases.

**Proof.** Given \( g \in M \), construct \( f = \phi(g) \in M \) according to (2.20). Then, from (2.18), (2.22), and Lemma 3.1, it remains to prove that

\[
\left| \mathbb{E} \left\{ s_n^{-2} \sum_{i,j} \sigma_{ij} D^2 f(Y_n)[J_{s_i^2/s_n^2}, J_{s_j^2/s_n^2}] - \int_0^1 D^2 f(Y_n)[J^{(2)}_t] \, dt \right\} \right| \leq CC_1 (\epsilon_6 + \epsilon_7),
\]

for some universal \( C \). This, however, follows from (3.11) applied first to the differences \( D^2 f(y)[J_{s_i^2/s_n^2}, J_{s_j^2/s_n^2}] - D^2 f(y)[J^{(2)}_t] \), and then twice to estimate differences of the form

\[
D^2 f(y)[J^{(2)}_t] - D^2 f(y)[J^{(2)}_t], \quad s_i^2/s_n^2 \leq t \leq s_{i+1}^2/s_n^2.
\]

This sort of structure occurs in particular when the \( X_i \)'s arise as a mixing sequence indexed by the integers, and the elements of \( N_i \) denote those \( X_j \) which are 'significantly' dependent on \( X_i \). The quantities \( \epsilon_1, \epsilon_2, \epsilon_5 \) and \( \epsilon_6 \) are then usually controlled by the mixing assumptions and by the choice of \( G \). For example, defining the strong mixing coefficients

\[
\alpha(i,j) = \sup_{A \in \sigma(X_i, X_j), B \in \sigma(X_i; k \in N_i \cup N_j)} \left| \mathbb{P}[A \cap B] - \mathbb{P}[A] \mathbb{P}[B] \right|,
\]

and

\[
\alpha_{ij} = \sup_{A \in \sigma(X_i), B \in \sigma(X_j)} \left| \mathbb{P}[A \cap B] - \mathbb{P}[A] \mathbb{P}[B] \right|,
\]
it follows from Hall and Heyde (1980), p. 277, Theorem A5, that for $f \in M^0$,

$$
\varepsilon_1 \leq 8s_n^{-1} \sum_{i=1}^{\infty} \left( \mathbb{E} |X_i|^3 \right)^{1/3} \alpha(i, i)^{2/3} \| f \|_{M^0};
$$

$$
\varepsilon_2 \leq 8s_n^{-2} \sum_{i=1}^{\infty} \sum_{j \in N_i} \left( \mathbb{E} |X_i^j| \right)^{1/3} \alpha(i, j)^{1/3} \| f \|_{M^0};
$$

$$
\varepsilon_3 \leq 8s_n^{-2} \sum_{i=1}^{\infty} \sum_{j \in N_i} \left( \mathbb{E} |X_i^j| \right)^{1/3} \left( \mathbb{E} |X_j| \right)^{1/3} \alpha(i, j)^{1/3} \| f \|_{M^0},
$$

and, furthermore,

$$
\varepsilon_6 \leq \varepsilon_5 + \varepsilon_4 \quad \text{and} \quad \varepsilon_4 \leq \varepsilon_4^* \quad \text{where}
$$

$$
\varepsilon_4 = 2s_n^{-3} \sum_{i=1}^{\infty} \sum_{j, k \in N_i} |\alpha(i, j)| \left( \mathbb{E} X_k^j \right)^{1/2} \| f \|_{M^0}
$$

$$
\leq 16s_n^{-3} \sum_{i=1}^{\infty} \sum_{j, k \in N_i} \left( \mathbb{E} |X_i^j| \right)^{1/3} \left( \mathbb{E} |X_j^k| \right)^{1/3} \alpha(i, j)^{1/3} \| f \|_{M^0}.
$$

Thus, in the case of stationary, mixing summands, letting

$$
G = \{ (i, j) : |j - i| \leq m \},
$$

and setting

$$
\alpha_r(m) = \sup_{A \in \sigma(X_0, X_r, B \in \sigma(X_s : s \neq [m, r + m])} \left| \mathbb{P}[A \cap B] - \mathbb{P}[A] \mathbb{P}[B] \right|,
$$

we have

$$
\varepsilon_1 \leq 8\gamma^{1/3} s_n^{-1} \alpha_0^{2/3}(m) \| f \|_{M^0};
$$

$$
\varepsilon_2 \leq 64\gamma^{2/3} s_n^{-2} \sum_{r \leq m} \{ \alpha_r(m) \wedge \alpha_0(r) \}^{1/3} \| f \|_{M^0};
$$

$$
\varepsilon_3 \leq \frac{3}{2} (2m + 1)^3 \gamma s_n^{-3} \| f \|_{M^0};
$$

$$
\varepsilon_4 \leq 32 \gamma^{2/3} s_n^{-3} \sum_{r \geq m} \alpha_0(r)^{1/3} \| f \|_{M^0};
$$

$$
\varepsilon_5 \leq 16\gamma^{2/3} s_n^{-2} \sum_{r \geq m} \alpha_0(r)^{1/3} \| f \|_{M^0},
$$

where $\gamma = \mathbb{E}|X_0|^3$.

This implies that, if $\sum_{r \geq 0} \alpha_0(r)^{1/3} < \infty$ and $\alpha_r(m) \to 0$ for each $r$ as $m \to \infty$; and if

$$
\sigma^2 = \sum_{j = -\infty}^{\infty} \sigma_{0j} > 0,
$$

so that, as $n \to \infty$, $s_n^2 \sim \sigma^2 n^2$; then $m = m(n)$ can be chosen in such a way that $m \to \infty$, $n^{1/2} \alpha_0(m)^{2/3} \to 0$ and $n^{-1/2} m^2 \to 0$, and hence so that $\varepsilon_1 - \varepsilon_7$ all converge
to zero as \( n \to \infty \). For example, if, for all \( r \geq 0 \), \( \alpha_r(m) \leq Km^{-\beta} \) for some \( \beta > 3 \) and \( K < \infty \), then

\[
\varepsilon_1 \asymp n^{1/2} m^{-2\beta/3}; \quad \varepsilon_2, \varepsilon_5 \asymp m^{1-\beta/3}; \quad \varepsilon_3, \varepsilon_4 \asymp n^{-1/2} m^2; \quad \varepsilon_7 \asymp n^{-1/2},
\]

so that, picking \( m = n^\eta(\beta) \), where \( \eta(\beta) = [2(1 + \beta/3)]^{-1} \), it follows that

\[
\varepsilon_l = O(n^{-\theta_l(\beta)}), \quad 1 \leq l \leq 7, \quad \text{where} \quad \theta_l(\beta) = (\beta/3 - 1)\eta(\beta).
\]

These conditions are reminiscent of the sufficient condition \( \sum m^{1/3} < \infty \) for the one-dimensional central limit theorem when \( \mathbb{E}|X_0|^3 < \infty \) (Hall and Heyde 1980, p. 132, Corollary 5.1), where

\[
\alpha_m = \sup_{A \in \sigma(X_k; k \leq 0), B \in \sigma(X_k; k > m)} |\mathbb{P}(A \cap B) - \mathbb{P}[A] \mathbb{P}[B]|
\]

denote the usual strong mixing coefficients. The coefficients \( \alpha_r(m) \) are different from the \( \alpha_m \)'s, though both embody the idea that two events depending on well separated index sets should not influence each other very much. The \( \alpha_r(m) \) coefficients have the advantage that they are easily generalized to higher dimensional index sets. Furthermore, in the present formulation, the way in which the coefficients \( \alpha_r(m) \) influence the accuracy of the approximation by a Wiener process, measured in the \( d \)-metric, is made explicit.

In the case of \( m \)-dependent summands, taking \( G \) as in (3.12), the quantities \( \varepsilon_1, \varepsilon_2 \) and \( \varepsilon_5 \) are zero for all \( f \in M \), and the error estimates can be correspondingly simplified.

**Theorem 4.** Let \( (X_i) \) be a stationary, \( m \)-dependent sequence such that \( \mathbb{E}X_0 = 0 \) and \( \gamma = \mathbb{E}|X_0|^3 < \infty \), and set \( Y_n = s_n^{-1} \sum_{i=1}^{n} X_i J^n \). Then there exists a universal constant \( C \) such that

\[
|\mathbb{E}g(Y_n) - \mathbb{E}g(Z)| \leq C \| g \|_M \left[ (m^2 + 1)\gamma n s_n^{-3} \right. \\
\left. + [(m + 1)^{1/3} s_n^{-1} + C_1 (m/n)^{1/2}] \left( \max_i \left\{ \alpha_i^{-2} \sum_j |\sigma_{ij}| \right\} \right) \right. \\
\left. + C_1 \left\{ \max_{0 \leq i \leq n} \left\{ \max_{0 \leq i \leq n} \{ s_i^2/s_n^2 - i/n \}, \max_{0 \leq i \leq n} \{(i + 1)/n - s_i^2/s_n^2 \} \right\} \right\}^{1/2} \right],
\]

(3.14)

for all \( g \in M \) satisfying (3.11).

**Remark.** In the usual asymptotic situation, with \( m \) fixed and letting \( n \to \infty \), if

\[
\sigma^2 = \sum_{j=-m}^{m} \sigma_{0j} > 0,
\]

this gives an approximation error of order \( n^{-1/2} \).

**Proof.** Similar to that of Theorem 3. \( \square \)
Remarks. 1. In the case of independent summands \((m = 0)\), the last term is just \(C_1 n^{-1/2}\).

2. If the summands are not identically distributed, a similar result can still be proved. Letting \(Y_n = s_n^{-1} \sum_{i=1}^n X_i J_{i/s_n^2}\), an estimate corresponding to (3.14) is

\[
C \| g \|_M \left[ (m^2 + 1) s_n^{-3} \sum_{i=1}^n \mathbb{E}|X_i|^3 + s_n^{-3} \sum_{i=1}^n \mathbb{E}|X_i| \sum_{j \in N_i, k \in N_j} |\sigma_{jk}| 
+ C_1 \left\{ m^{1/2} \max_i \left[ \sigma_i^{-2} \sum_j |\sigma_{ij}| \right] + 1 \right\} \max_i (\sigma_i/s_n) \right].
\]

The method used to prove Theorems 3 and 4 relies little on the nature of the underlying index set, and is equally appropriate for estimating the error when approximating the distribution of the partial sum process defined on a mixing random field by a Gaussian random process. The definition of the partial sum process is of course different: for a stationary field in \(d\) dimensions, one might consider

\[
Y_n = s_n^{-1} \sum_{0 \leq j \leq n} X_j J_{jn^{-1}},
\]

where \(n\) and \(j\) are elements of \(\mathbb{N}^d, \leq\) is the usual partial order, \(s_n^2 = \text{Var}\left( \sum_{0 \leq j \leq n} X_j \right)\), \(jn^{-1}\) denotes componentwise division, and \(J_t \in D^{(d)}\), for any \(t \in \mathbb{R}^d\), is the function defined by

\[
J_t(u) = \begin{cases} 1 & \text{if } u \geq t, \\ 0 & \text{otherwise}. \end{cases}
\]

\(G\) could then be taken to be \(\{(i, j) \in \mathbb{N}^d \times \mathbb{N}^d, |i - j| \leq m\}\), the smoothness condition (3.11) can be reinterpreted in the natural way, and the limiting Gaussian process has a generator which, for such functions \(g\), can again be expressed as

\[
(\mathcal{A} g)(w) = -D g(w)[w] + \int_0^1 D^2 g(w)[J_t^{(2)}] dt,
\]

though now \(w\) and \(J_t\) belong to \(D^{(d)}\) and \(t \in [0, 1]^d\); that is, the limiting process is the \(d\)-dimensional Brownian sheet.

3.2. Dissociated Random Variables

Let \(A = A(n)\) denote the set of all \(d\)-subsets of \(\{1, 2, \ldots, n\}\), and let \((X_t)_{t \in A}\) be a dissociated family of random variables with zero means and finite third moments: that is, the joint distributions of the families of random variables \(\{X_{I_a}\}_{a \in A}\) and \(\{X_{I_\beta}\}_{\beta \in B}\) are independent of one another whenever

\[
\left( \bigcup_{a \in A} I_a \right) \cap \left( \bigcup_{\beta \in B} I_\beta \right) = \emptyset
\]
(McGinley and Sibson 1975). Then the methods of the previous section can be applied to processes of the form \( Y_n = s_n^{-1} \sum_{i \in \Delta} X_i J^\tau_i \), for suitable functions \( J^\tau_i \), taking

\[
G = \{(I, J) \in \Delta \times \Delta; I \cap J + \emptyset\},
\]

which implies that \( \varepsilon_1 = \varepsilon_2 = \varepsilon_5 = 0 \). In the case of exchangeably dissociated families such that \( \mathbb{E}(X_i X_j) = \tau > 0 \) when \( |I \cap J| = 1 \), as, for example, centred non-degenerate U-statistics,

\[
\sigma^2 = \tau d \left\{ \frac{n^d}{(d-1)!} \right\} (1 + O(n^{-1})); \quad s_n^2 = \tau n^{d-2} \left\{ \frac{(d-1)!}{(d-2)!} \right\} (1 + O(n^{-1})),
\]

so that \( \varepsilon_3 \asymp n^{-1/2} \) and \( \varepsilon_4 \asymp n^{-1/2} \). Thus Lemma 3.1 may be applied to give an estimate of the form

\[
\left| \mathbb{E}\left\{ Df(Y_n)[Y_n] - \tilde{s}_n^{-2} \sum_{I, J \in \Delta} \tilde{\sigma}_{IJ} D^2 f(Y_n)[J^\tau_I, J^\tau_J] \right\} \right| \leq K n^{-1/2} \| f \|_M
\]

for any \( f \in M \), where, in this section, \( K \) denotes a constant depending on the product moments of the \( X_i \) of order at most three.

In this form, however, the approximating Gaussian process is not so easy to analyse as if one takes \( \tilde{\sigma}_{IJ} = |I \cap J| \), which also yields

\[
\tilde{s}_n^2 = n^{d-1} \left\{ \frac{(d-1)!}{(d-2)!} \right\} (1 + O(n^{-1})),
\]

and hence, for \( f \in M \),

\[
\left| \mathbb{E}\left\{ Df(Y_n)[Y_n] - \tilde{s}_n^{-2} \sum_{I, J \in \Delta} \tilde{\sigma}_{IJ} D^2 f(Y_n)[J^\tau_I, J^\tau_J] \right\} \right| \leq K n^{-1/2} \| f \|_{M^0}
\]

\[
+ n^{-1} \mathbb{E}\| D^2 f(Y_n) \|.
\]

Now the approximating Gaussian process can be realized as

\[
\tilde{Z}_n = \tilde{s}_n^{-1} \sum_{I \in \Delta} \left( \sum_{k \in I} Z_k \right) J^\tau_I,
\]

where \( (Z_k)_{k=1}^\infty \) are independent \( \mathcal{N}(0, 1) \) random variables. This leads to the following result.

**Theorem 5.** Let \( (X_i)_{i \in \Delta} \) be exchangeably dissociated, with zero mean and finite third moment, and set \( Y_n = s_n^{-1} \sum_{I \in \Delta} X_i \). Let \( \tilde{Z}(t) = t^{d-1} Z(t) \), where \( Z \) denotes standard Brownian motion. Then, for all \( g \in M^0 \) satisfying the smoothness condition (3.11),

\[
|\mathbb{E}g(Y_n) - \mathbb{E}g(Z)| \leq K n^{-1/2} \| g \|_{M^0},
\]

where the constant \( K \) depends only on the product moments of the \( X_i \) of order at most three.
Proof. Take $J^*_i = J_i/n \in D$, where $i = i(l) = \max\{i' : i' \in I\}$. Then

\[
\hat{Z}_n = \hat{s}_n^{-1} \sum_{k=1}^n Z_k \sum_{l \leq k} J^*_l
\]

\[
= \hat{s}_n^{-1} \sum_{k=1}^n Z_k \left[ \binom{k-1}{d-1} J_{k/n} \sum_{i=k+1}^n \binom{i-2}{d-2} J_{i/n} \right],
\]

so that

\[
\hat{Z}_n(v) = \hat{s}_n^{-1} \left( \sum_{k=1}^n Z_k \left( \binom{[nv]}{d-1} - 1 \right) \right) = n^{-1/2} \sum_{k=1}^n Z_k u_{k/n}(v)(1 + O(n^{-1})) ,
\]

where $u_i(v) = v^{d-1} J_i(v)$. Since the generator $\mathscr{A}$ of $\hat{Z}$ can be expressed as

\[
(\mathscr{A}f)(w) = -Df(w)[w] + \frac{1}{0} D^2 f(w)[u^{(2)}] dt ,
\]

the rest of the proof can be accomplished by comparing the values of $D^2 f(w)[y^{(2)}]$ for the following choices of $y$:

\[
y(v) = n^{1/2} \hat{s}_n^{-1} \left( \binom{[nv]}{d-1} - 1 \right) f[k \leq nv]; \quad y = u_{k/n}; \quad \text{and} \quad y = u_t, (k-1)/n \leq t \leq k/n.
\]

Alternatively, a process with $d$-dimensional argument could be considered.

3.3. The Wald-Wolfowitz Theorem

Let $(a_i)_{i=1}^n$ and $(b_i)_{i=1}^n$ be real numbers, chosen so that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i = 0$, and define $s_n^2 = (n-1)^{-1} \sum_{i=1}^n a_i^2 \sum_{i=1}^n b_i^2$. Under mild conditions on the $a$'s and $b$'s, the distribution of $Y = s_n^{-1} \sum_{i=1}^n a_i b_{n(i)}$ is approximately standard normal, when $\pi$ is drawn at random from the uniform distribution on $\Sigma_n$, the permutations of $\{1, 2, \ldots, n\}$. Stein's method has previously been successfully applied in this context by Bolthausen (1984), to obtain a Berry-Esséen theorem. Here, we demonstrate a convergence rate for smooth functionals in a process analogue of the theorem, expressed in terms of the Lyapounov estimate

\[
g_n = n^{-1} s_n^{-3} \sum_{i=1}^n |a_i|^3 \sum_{i=1}^n |b_i|^3 .
\]

It would be possible to consider approximating the process

\[
Y_n^d = s_n^{-1} \sum_{i=1}^n a_i b_{n(i)} J_{i/n} ,
\]

where $J_x$ is defined as in (2.15), but this formulation treats the sequences $a$ and
$b$ asymmetrically. We therefore define

$$Y_n = Y_n(\sigma, \pi) = s_n^{-1} \sum_{i=1}^{n} X_i J_{i/n},$$

where $X_i = a_{\sigma(i)} b_{\pi(i)}$, and $\sigma$ and $\pi$ are independently drawn from the uniform distribution on $\Sigma_n$. Note that $Y_n(1)$ has the same distribution as $Y$ above.

**Theorem 6.** Let $Y_n$ be defined as above, and let $Z$ denote a standard Wiener process. Then, for any $g \in M$ satisfying (3.11),

$$|\mathbb{E}g(Y_n) - \mathbb{E}g(Z)| \leq C\gamma_n \|g\|_M.$$

**Proof.** We take $f \in M$ and evaluate $\mathbb{E}(X_k Df(Y_n)[J_{k/n}])$. First, note that

$$\mathbb{E}(X_k Df(Y_n)[J_{k/n}]) = n^{-2} \sum_{i,j} a_i b_j \mathbb{E}(Df(Y_n)[J_{k/n}]) \sigma(k) = i, \pi(k) = j.$$

In order to evaluate the conditional expectation, note that, if $\sigma$ is uniformly distributed on $\Sigma_n$, the permutation $\tilde{\sigma}$ defined by

$$\tilde{\sigma}(k) = i; \tilde{\sigma}(\sigma^{-1}(i)) = \sigma(k); \tilde{\sigma}(l) = \sigma(l)$$

has the distribution obtained from the uniform distribution by conditioning on the event $\tilde{\sigma}(k) = i$. Thus, with a similar observation for $\pi$, we can write

$$\mathbb{E}(Df(Y_n)[J_{k/n}]) \sigma(k) = i, \pi(k) = j = \mathbb{E}(Df(Y_n(\sigma, \pi) + A_{ijk})[J_{k/n}]),$$

where

$$s_n A_{ijk} = (a_i b_j - a_{\sigma(k)} b_{\pi(k)})(J_{k/n} - J_{\sigma^{-1}(i)/n} I[\sigma^{-1}(i) = \pi^{-1}(j)]) + [(a_{\sigma(k)} - a_i)$$

$$\times b_{\pi(\sigma^{-1}(i))} J_{\sigma^{-1}(i)/n} + a_{\sigma(\pi^{-1}(j))}(b_{\pi(k)} - b_j) J_{\pi^{-1}(j)/n}] I[\sigma^{-1}(i) \neq \pi^{-1}(j)].$$

Then

$$\left| \mathbb{E}Df(Y_n)[Y_n] - s_n^{-1} \sum_{k=1}^{n} n^{-2} \sum_{i,j} a_i b_j \{\mathbb{E}Df(Y_n)[J_{k/n}] + \mathbb{E}D^2f(Y_n)[J_{k/n}, A_{ijk}] \} \right|$$

$$= \left| \mathbb{E}Df(Y_n)[Y_n] - s_n^{-1} \sum_{k=1}^{n} n^{-2} \sum_{i,j} a_i b_j \mathbb{E}D^2f(Y_n)[J_{k/n}, A_{ijk}] \right|$$

$$\leq n^{-2} s_n^{-1} \sum_{i,j,k} |a_i b_j| \mathbb{E}\|A_{ijk}\|^2 \|f\|_M \leq C\gamma_n \|f\|_M.$$

Now observe that $A_{ijk}$ depends only on the values $l, m, r, s, t, u$ taken by $\sigma(k), \pi(k), \sigma^{-1}(i), \pi^{-1}(j), \sigma(\pi^{-1}(j))$ and $\pi(\sigma^{-1}(i))$ respectively, and that a random element with the distribution of $Y_n$, given these values of $\sigma$ and $\pi$, can be realized as

$$Y_n(\sigma, \pi) + A(i,j,k; l, m, r, s, t, u),$$

using a construction similar to that of $A_{ijk}$, where $\sigma$ and $\pi$ are uniform on $\Sigma_n$. It is
then not difficult to see that
\[ n^{-2} s_n^{-1} \sum_{i,j,k} |a_i b_j| \mathbb{E} \| \Delta_{i,j,k} \Delta(i,j,k; \sigma(k), \pi(k), \sigma^{-1}(i), \pi^{-1}(j)) \| \leq C \gamma_n , \]
and it remains to compute
\[ \mathbb{E} D^2 f(Y_n)[J_{k/n}, \mathbb{E} A_{i,j,k}] . \]
This is achieved by routine calculations, based on the probabilities of the different combinations of \( l, m, r, s, t, u \): for instance,
\[ \mathbb{E}(a_{\sigma(i)} b_{\pi(i)} J_{\sigma^{-1}(i)/n} I[\pi^{-1}(j) = \sigma^{-1}(i)]) \]
\[ = \frac{1}{n^2(n - 1)} \sum_{s + k m + j} a_i b_m J_{k/n} + \frac{1}{n^2(n - 1)^2} \sum_{r + k l + i u + j} a_i b_u J_{r/n} \]
\[ + \frac{1}{n^2(n - 1)^2(n - 2)} \sum_{l + i m + j} \sum_{r + k s + r, u + m, j} a_i b_u J_{r/n} \]
\[ = \{a_i b_j/(n(n - 1))\} (L - J_{k/n}) , \]
using the fact that \( \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} b_i = 0 \), where \( L = n^{-1} \sum_{k=1}^{n} J_{k/n} \). The complete calculation yields
\[ n^{-2} s_n^{-1} \sum_{i,j} \sum_{k=1}^{n} a_i b_j \mathbb{E} D^2 f(Y_n)[J_{k/n}, \mathbb{E} A_{i,j,k}] \]
\[ = \frac{n - 2}{n(n - 1)} \sum_{i=1}^{n} \mathbb{E} D^2 f(Y_n)[J_{k/n}^{(2)}] + \frac{1}{n - 1} \mathbb{E} D^2 f(Y_n)[L^{(2)}] , \]
\[ = n^{-1} \sum_{k=1}^{n} \mathbb{E} D^2 f(Y_n)[J_{k/n}^{(2)}] + O(n^{-1/2} f \| M) , \]
since
\[ \mathbb{E} \| Y_n \| \leq n^{-1} s_n^{-1} \sum_{i=1}^{n} |a_i| \sum_{i=1}^{n} |b_i| \leq n^{1/2} . \]
Now, arguing as for Theorem 1, one obtains
\[ |\mathbb{E} g(Y_n) - \mathbb{E} g(Z)| \leq C \| g \|_M (n^{-1/2} \sqrt{\log n} + \gamma_n) \]
for all \( g \in M \), and
\[ |\mathbb{E} g(Y_n) - \mathbb{E} g(Z)| \leq C \| g \|_M \gamma_n \]
for all \( g \in M \) satisfying the smoothness condition (3.11). \( \square \)

Remark. For the process \( Y_n^1 \), the argument is much simpler, because there is only the permutation \( \pi \) to consider. However, the resulting penultimate approximation, with error at most \( C \gamma_n \| g \|_M \) for \( g \in M \), is to the Gaussian process
\[ Z_n^1 = \sum_{k=1}^{n} \left( \sum_{i=1}^{n} a_i^2 \right)^{-1/2} Z_k \{a_k J_{k/n} - \bar{J} \} , \]
where
\[ \bar{J}(t) = N^{-1} \sum_{k=1}^{n} a_k J_{k/n}(t), \]
and a neat approximation by a limiting process is only possible if, for instance, \( \|\bar{J}\| \) is small. In this case, it would be natural to replace \( J_{k/n} \) in the definition of \( Y_n^1 \) by \( J_{A(k)/A(0)} \), where \( A(k) = \sum_{i=1}^{k} a_i^2 \), and to look at a Wiener process limit.

### 3.4. The Empirical Distribution Function

Let \( X_1, \ldots, X_n \) be independent \( U[0, 1] \) random variables, and set
\[ Y_n(t) = n^{1/2} \left\{ n^{-1} \sum_{i=1}^{n} I[X_i \leq t] - t \right\}. \tag{3.15} \]

Then, writing \( G_j(t) = I[X_j \leq t] - t \) and \( Y_n^j = n^{-1/2} \sum_{k+j} G_k \), it follows from Taylor’s expansion that, for any \( f \in M \),
\[ |\mathbb{E}\{n^{-1/2} Df(Y_n)[G_j] - n^{-1} D^2f(Y_n^1)[G_j^2]\}| \leq \frac{1}{2} n^{-3/2} \mathbb{E}\|G_j\|\|f\|_M, \]
since \( G_j \) and \( Y_n^j \) are independent and \( G_j \) has mean zero in \( D \): similarly,
\[ |\mathbb{E}D^2f(Y_n)[h, k] - \mathbb{E}D^2f(Y_n^1)[h, k]| \leq n^{-1/2} \mathbb{E}\|G_j\|\|f\|_M \mathbb{E}\|h\| \mathbb{E}\|k\|. \]

This implies in turn that
\[ \left| \mathbb{E}\left\{ -Df(Y_n)[Y_n] + \int_{0}^{1} D^2f(Y_n)[(J_u - l)^2] du \right\} \right| \leq \frac{3}{2} n^{-1/2} \|f\|_M, \]
where \( l \) denotes the identity function on \([0, 1]\) and \( J_u \) is as in (2.15), since \( \|G_j\| \leq 1 \). Hence we have the following result.

**Theorem 7.** For all \( g \in M \),
\[ |\mathbb{E}g(Y_n) - \mathbb{E}g(Z^0)| \leq Cn^{-1/2} \|g\|_M, \]
where \( Z^0 \) denotes the Brownian bridge.

**Remarks.** 1. Note that the generator associated with \( Z^0 \) as equilibrium measure is given by
\[ \mathcal{A}^0 g(w) = -Df(w)[w] + \int_{0}^{1} D^2f(w)[(J_u - l)^2] du, \]
and that one construction of the corresponding Markov process is to take Formula (2.2) without the \( k = 0 \) term in the sum.

2. In some applications, when proving rates of convergence for integral functionals of \( Y_n \), designed to be sensitive to the values taken by \( Y_n \) near 0 and 1,
the uniform norm on $D$ is not quite sharp enough. In such cases, the norm
\[\|w\|' = \sup_t |w(t)| + \left\{ \frac{1}{0} \int_0^t \frac{w^2(t)}{t(1-t)} \, dt \right\}^{1/2}\]
on the set $D' = \{w \in D: \|w\|' < \infty\}$ is more convenient.

3. The case of non-identically distributed $X_j$'s can be handled in a very similar way.

4. The same argument formally applies to the standardized empirical measure
\[Y_n = n^{-1/2} \sum_{i=1}^n (\delta_{X_i} - \lambda),\]
where $\lambda$ denotes Lebesgue measure on $[0, 1]$. Here, it is necessary to choose a suitable class of functions to index $Y_n$, with respect to which there is a tractable norm.

5. The generalization to multivariate empirical distribution functions is straightforward.

6. For the process defined by
\[Y_n(s, t) = n^{-1/2} \sum_{i=1}^{[nt]} (I[X_i \leq t] - t), \quad 0 \leq s, t \leq 1,
\]
a similar argument establishes an order $n^{-1/2}$ approximation by the Kiefer process, which has generator
\[(\mathcal{A}f)(w) = -Df(w)[w] + \int_0^1 \int_0^1 D^2f(w)[\{J_u^1(J_v^2 - l)\}] \, du \, dv,
\]
where
\[J_u^1(s, t) = I[s \geq u], \quad J_u^2(s, t) = I[t \geq u], \quad \text{and} \quad l(s, t) = t.
\]
For this result, the functions $g$ should also satisfy the smoothness condition
\[\sup_{w \in D} |D^2g(w)[\{J_u^1(J_v^2 - l), J_u^1(J_v^2 - l)\}]| \leq C_1 \|g\|_M |u - v|^{1/2},\]
for all $r, u, v, t \in [0, 1]$.

The assumption of independence of the $X_j$'s can also be relaxed, using methods similar to those of Sect. 3.1. A typical example is given by the next result.

**Theorem 8.** Let $(X_i)$ be a stationary $m$-dependent sequence with $U[0, 1]$ marginals, and define $Y_n$ as in (3.15). Let
\[\rho(u, v) = \sum_{r = -m}^m \{P[X_0 < u, X_r < v] - uv\},\]
and suppose that $\rho(u, u) > 0$. Then, for all $g \in M^0$,
\[|Eg(Y_n) - Eg(Z)| \leq C(m + 1)^2 n^{-1/2} \|g\|_{M^0},\]
where $Z$ denotes a Gaussian process with covariance function $\rho$.

**Proof.** We suppress the subscript $n$ throughout. With $G_j$ defined as before, set
\[W^j = n^{-1/2} \sum_{|k - j| \leq m} G_k, \quad Y^j = n^{-1/2} \sum_{|k - j| > m} G_k,
\]
so that, in particular, \( Y^j \) and \( G_j \) are independent. Then, arguing in the usual way, we have

\[
|\mathbb{E}Df(Y)[G_j] - \mathbb{E}D^2f(Y^j)[G_j, W^j]| \leq \frac{1}{2} n^{-1/2} \mathbb{E} \| W^j \|_2 \| f \|_M
\]

\[
\leq \frac{1}{2} (2m + 1)^2 n^{-3/2} \| f \|_M ,
\]

and, for \(|j - k| \leq m\),

\[
|\mathbb{E}D^2f(Y^j)[G_j, G_k] - \mathbb{E}D^2f(Y^k)[G_j, G_k]| \leq mn^{-1/2} \| f \|_M ,
\]

where \( Y^k = n^{-1/2} \sum_{|r - k| > m, |r - j| > m} G_r \) is independent of \( G_j \) and \( G_k \). Now the observation

\[
|\mathbb{E}D^2f(Y)[h_1, h_2] - \mathbb{E}D^2f(Y^k)[h_1, h_2]| \leq (3m + 1)n^{-1/2} \| f \|_M \| h_1 \| \| h_2 \|
\]

yields

\[
|\mathbb{E}Df(Y)[Y] - n^{-1} \sum_{\{j, k : |j - k| \leq m\}} \mathbb{E}D^2f(Y)[(J_u - l), (J_v - l)]dF_{j-k}(u, v)|
\]

\[
\leq n^{-1/2} \left\{ \frac{1}{2} (2m + 1)^2 + m(2m + 1) + (2m + 1)(3m + 1) \right\} \| f \|_M ,
\]

where \( dF_r \) denotes the joint distribution of \( X_0 \) and \( X_r \), and \( l \) denotes the identity function on \([0, 1]\). Finally, for \( f \in M^0 \),

\[
\left| n^{-1} \sum_{\{j, k : |j - k| \leq m\}} \mathbb{E}D^2f(Y)[(J_u - l), (J_v - l)]dF_{j-k}(u, v) \right|
\]

\[
- \sum_{r = -m}^{m} \mathbb{E}D^2f(Y)[J_u - l, J_v - l]dF_r(u, v) | \leq m(m + 1)n^{-1} \| f \|_{M^0} ,
\]

and the result follows upon identifying the covariance function.

Acknowledgements. It is a great pleasure to thank G.K. Eagleson for many helpful discussions, and the University of New South Wales for making possible the visit to Australia, during which most of this work was accomplished.

References


Received October 11, 1988