ON RATES OF CONVERGENCE FOR THE INVARIANCE PRINCIPLE

BY

WALTER A. ROSENKRANTZ

I. Introduction. Let \((X_i)_{i=1}^\infty\) denote a sequence of independent and identically distributed random variables satisfying the following condition:

\begin{align*}
(E(X_i) = 0, & V(X_i) = \sigma^2, \\
E(|X_i|^{2+a}) = b, & 0 < b < +\infty, \quad 0 < a \leq 2.
\end{align*}

For each integer \(n\), we define \(X_{nt}, i = 1, 2, \ldots, n\) as follows:

\[ X_{nt} = X_{(n-1)t}/n^{1/2}. \]

It follows at once from (1)-(3) that the random variables, \(X_{nt}\), have the following properties:

\begin{align*}
E(X_{nt}) = 0, & V(X_{nt}) = n^{-1}, \\
n^{r}E(|X_{nt}|^{2r}) = d < +\infty, & d = b\sigma^{-(2+a)} \quad \text{and} \quad 2r = 2+a.
\end{align*}

Let \(s_{nk} = \sum_{t=1}^{n} X_{nt}\) and denote by \(W(t)\), the one-dimensional Brownian motion process with covariance function \(r(s, t) = \min(s, t)\) and \(0 \leq t < +\infty\). We define probability distribution functions \(F_{n}(\lambda)\) and \(F(\lambda)\) as follows:

\begin{align*}
F_{n}(\lambda) = \Pr \left\{ \max_{1 \leq k \leq n} |s_{nk}| \leq \lambda \right\}, \\
F(\lambda) = \Pr \left\{ \max_{0 \leq t \leq 1} |W(t)| \leq \lambda \right\}.
\end{align*}

In their now classical paper [4], Kac and Erdös proved that \(\lim_{n \to \infty} F_{n}(\lambda) = F(\lambda)\) where

\[ F(\lambda) = \frac{4}{\pi} \sum_{i=0}^{\infty} (2i+1)^{-1} \exp \left[ -\frac{(2i+1)^2 \pi^2}{8\lambda^2} \right]. \]

These results were later generalized by Donsker [2] and Prochorov [8] and the method employed to obtain these limit theorems has become known as "the invariance principle." The purpose of this paper is to study the rate of convergence of \(F_{n}\) to \(F\). In particular we shall prove

**Theorem 1.**

\[ |F_{n}(\lambda) - F(\lambda)| \leq A(\log n)^{1/2} n^{-\mu}, \quad \text{where} \quad \mu = \frac{1}{2} a/(a+3) \]

and \(A\) is a constant depending on \(a\) and \(b\), but independent of \(\lambda\).
For the case $a = 1$, i.e., $E(|X|^3) < +\infty$, Theorem 1 predicts
\[ |F_n(\lambda) - F(\lambda)| \leq A(\log n)^{1/2}/n^{1/8}. \]

This case has also been studied by K. L. Chung [1] and Y. Prohorov [8, Theorem 4.1] who obtained respectively the following results:

**Theorem 2.1 (Chung).**
\[ |F_n(\lambda) - F(\lambda)| \leq A'(\log \log n^{1/2}/\log n^{1/2})^{1/2}. \]

**Theorem 2.2 (Prohorov).**
\[ |F_n(\lambda) - F(\lambda)| \leq A''(\log n)^{1/2}/n^{1/8}. \]

It is to be observed that the method used in this paper, which is based on a beautiful representation theorem of Skorokhod [9], is quite different from that of the above-mentioned authors—it is also much simpler. Next, a word about the organization of this paper: In Part II we state the Skorokhod representation theorem and apply it to the random variables defined at (3); the proof of Theorem 1 is given in Part III and an extension of Theorem 1 is given in Part IV. Finally it should be mentioned that it is still an open question what are the best possible rates of convergence. The author conjectures that $(\log n)^{1/2}/n^{1/2}$ is best possible.

**II. The Skorokhod representation.** In this section we present, without proofs, some important results due to Skorokhod. It is these results that we need to prove Theorem 1.

Let $\xi$ be a random variable defined over a probability space $(\Omega_1, B_1, P_1)$ and such that $E(\xi) = 0$. Let $\Omega_2 = \mathbb{C}[0, \infty)$ be endowed with the Wiener measure $P_w$ and we denote a Brownian motion path by $W(t)$, $0 \leq t < +\infty$. We now form the product space $\Omega = \Omega_1 \times \Omega_2$ and give it the product measure $P_1 \times P_w = P$. We can now assume that $\xi$ and $W(t)$ are random variables defined over $(\Omega, P)$ and we note that $\xi$ and $W(t)$ are independent relative to $P$. With these preliminaries out of the way we can now state the first part of the Skorokhod representation theorem (Chapter 7 of [9]):

**Theorem 3.** There exists a nonnegative random variable $\tau(\omega)$ which is a Markov time relative to the Brownian motion process $W(t, \omega)$ ($\omega \in \Omega$), with the following properties: The random variable $W(\tau(\omega), \omega)$, which we shall abbreviate to $W(\tau)$, has the same distribution as $\xi$; moreover the following moment inequalities are valid:
\[ E(\tau) = V(\xi) \]
and more generally
\[ E(\tau^k) \leq C_kE(|\xi|^{2k}), \quad k = 1, 2, \ldots, \]
where $C_k$ is an absolute constant independent of the distribution of $\xi$.

**Remarks.** The fact that $\tau$ is a Markov time for the Brownian motion process is important and is used in the following way: Let $Y_1$ and $Y_2$ be mutually independent random variables and such that $E(Y_1) = E(Y_2) = 0$. According to Skorokhod's
Theorem there exists a Markov time \( \tau_1 \) with the property that \( Y_1 \) and \( W(\tau_1) \) are identically distributed. Consider the process \( W_1(t) = W(\tau_1 + t) - W(\tau_1), 0 \leq t < \infty \); by the strong Markov property (see [6], [7]) \( W_1(t) \) is independent of both \( \tau_1 \) and \( W(\tau_1) \). Now apply the Skorokhod representation to \( Y_2 \) and \( W_1(t) \) to conclude there exists a Markov time \( \tau_2 \), relative to the process \( W_1(t) \), such that \( W_1(\tau_2) \) and \( Y_2 \) are identically distributed. It is clear that \( W_1(\tau_2) \) is independent of \( W(\tau_1) \) and that \( \tau_1 \) is also independent of \( \tau_2 \). Thus, the joint distribution of \( Y_1, Y_2 \) is the same as the joint distribution of \( W(\tau_1), W_1(\tau_2) = W(\tau_1 + \tau_2) - W(\tau_1) \). This naturally leads to the following general version of the Skorokhod representation theorem.

**Theorem 4.** Let \( Y_1, Y_2, \ldots, Y_n \) be mutually independent random variables with zero means and \( \text{Var}(Y_i) = \sigma_i^2 \). Then there exists a sequence of nonnegative, mutually independent random variables \( \tau_1, \tau_2, \ldots, \tau_n \) with the following properties.

The joint distributions of the random variables \( Y_1, Y_2, \ldots, Y_n \) are identical to the joint distributions of the random variables

\[
W(\tau_1), W(\tau_1 + \tau_2) - W(\tau_1), \ldots, W(\tau_1 + \cdots + \tau_n) - W(\tau_1 + \cdots + \tau_{n-1}).
\]

Moreover, we have

\[
E(\tau_k) = \sigma_k^2
\]

and

\[
E(\tau_k) \leq C_k E(|Y_i|^{2k}), \quad k = 1, 2, \ldots
\]

We now apply Theorem 4 to the random variables \( X_{ni}, i = 1, 2, \ldots, n \), defined at (3). Thus we can assert that there exist mutually independent, identically distributed and nonnegative random variables \( \tau_{n1}, \tau_{n2}, \ldots, \tau_{nn} \) with the following properties:

\[
E(\tau_{ni}) = n^{-1},
\]

and in addition the random variables

\[
W(\tau_{n1}), W(\tau_{n1} + \tau_{n2}) - W(\tau_{n1}), \ldots, W(\tau_{n1} + \cdots + \tau_{nn}) - W(\tau_{n1} + \cdots + \tau_{nn-1})
\]

have the same joint distribution as the random variables \( X_{ni}, i = 1, 2, \ldots, n \). In particular the joint distribution of \( s_{n1}, s_{n2}, \ldots, s_{nn} \) is the same as the joint distribution of the random variables:

\[
W(\tau_{n1}), W(\tau_{n1} + \tau_{n2}), \ldots, W(\tau_{n1} + \cdots + \tau_{nn}).
\]

We now define random variables \( z_{nk}, k = 1, 2, \ldots, n \) as follows:

\[
z_{nk} = \sum_{i=1}^{k} (\tau_{ni} - n^{-1}).
\]

Thus (18) can be rewritten in the following more useful form:

\[
W(1/n + z_{n1}), W(2/n + z_{n2}), \ldots, W(1 + z_{nn}).
\]
Since \(zn_k, k = 1, 2, \ldots, n\) is a sequence of partial sums of mutually independent, identically distributed random variables, each with zero means, it is reasonable to expect that if \(E(|\tau_{ni} - n^{-1}|) < +\infty\) for some \(\gamma > 1\) then we can obtain an estimate on the rate at which \(\Pr \{\text{Max}_{1 \leq k \leq n} |zn_k| > \delta\}\) tends to zero. Kolmogorov's inequality ([7, p. 235]) is not applicable because our hypotheses are not strong enough to insure the existence of a second moment for the \(\tau_{ni}, i = 1, 2, \ldots, n\). To get such estimates it is first of all necessary to extend inequality (15) as follows:

**Lemma 1.**

\[
E(|\tau|) \leq C \cdot E(|Y|^{2\gamma}) \quad \text{for any } \gamma \geq 1.
\]

The proof of this result is given in the appendix.

An immediate consequence of Lemma 1 (set \(\gamma = r\)) is the following result:

\[
v_n = E(|\tau_{ni} - n^{-1}|) \leq 2^{r-1}E(|\tau_{ni}| + n^{-r}) \leq 2^{r-1}C \cdot E(|X_{ni}|^{2r}) + 2^{r-1}n^{-r}.
\]

By (5) one easily concludes

\[
n'v_n \leq B < +\infty,
\]

where

\[
B = 2^{r-1}(C_d + 1).
\]

The Kolmogorov inequality for martingales (cf. [7]) yields:

\[
\Pr \left\{ \text{Max}_{1 \leq k \leq n} |zn_k| > \delta \right\} \leq \frac{E(|zn_n|^{r})}{\delta^{r}}.
\]

We now employ a result of von Bahr and Esseen [10] and obtain the following inequalities\(^{(2)}:\)

\[
E(|zn_n|^{r}) \leq 2nv_n,
\]

\[
n'^{-1}E(|zn_n|^{r}) \leq 2B < +\infty,
\]

and thus we obtain the estimate:

\[
\Pr \left\{ \text{Max}_{1 \leq k \leq n} |zn_k| > \delta \right\} \leq \frac{2B}{n'^{-1}\delta^{r}}.
\]

Summing up, then, we have the result that the joint distributions of the random variables \(s_{nk}, k = 1, 2, \ldots, n,\) are identical to the joint distributions of the random variables defined at (20), and so it is sufficient to prove Theorem 1 for these random variables instead of the original sequence \(s_{nk}, k = 1, 2, \ldots, n.\)

**III. Proof of Theorem 1.** The proof uses certain facts about the "tied down Brownian motion process"; we refer the reader to the paper by Doob [3] for a more complete discussion.

\(^{(2)}\) For this reference I am indebted to Professor Melvin Katz of the University of New Mexico.
Consider the following stochastic process $z_n(t)$: $z_n(t) = ntW(1/n) - W(t)$, $0 \leq t \leq 1/n$. We note that $z_n(0) = z_n(1/n) = 0$ and that $z_n(t)$ is a Gaussian process with covariance matrix $r_n(s, t) = s(1 - nt)$, $0 \leq s < t \leq 1/n$. We shall refer to the $z_n(t)$ process as the Brownian motion process "tied down at 0 and $n^{-1}$".

**Lemma 2.** $\Pr \{ \max_{0 \leq s \leq n^{-1}} |z_n(s)| \geq \epsilon \} \leq 2 \exp \left( -2n\epsilon^2 \right)$, $\epsilon > 0$.

**Proof.** Following Doob [3] one observes that $(1 + nt)z_n(t/(1 + nt)) = W(t)$, $0 \leq t < +\infty$. Thus

$$\Pr \left\{ \max_{0 \leq s \leq n^{-1}} |z_n(s)| \geq \epsilon \right\} = \Pr \left\{ \max_{0 \leq t < +\infty} |W(t)|/(1 + nt) \geq \epsilon \right\}.$$ 

Now using Doob's estimates [3, pp. 397–398] we conclude that

$$\Pr \left\{ \max_{0 \leq t < +\infty} |W(t)|/(1 + nt) \geq \epsilon \right\} \leq 2 \exp \left( -2n\epsilon^2 \right),$$

and this completes the proof of Lemma 2.

We now introduce a probability distribution function $\hat{F}_n(\lambda)$ as follows:

$$(28) \quad \hat{F}_n(\lambda) = \Pr \left\{ \max_{1 \leq k \leq n} \left| W(k/n) \right| \leq \lambda \right\},$$

where $W(k/n)$ denotes Brownian motion at time $t = k/n$. We now define a "random broken line" in the sense of Prochorov [8] as follows:

$$(29) \quad W_n(t) = W(k/n) \text{ for } t = k/n,$$

and

$$W_n(t) = W(k/n) + n \left[ W\left(\frac{k+1}{n}\right) - W\left(\frac{k}{n}\right) \right] \left( t - \frac{k}{n} \right) \text{ for } \frac{k}{n} \leq t \leq \frac{k+1}{n}. $$

In other words $W_n(t)$ is a polygonal approximation to the Brownian motion path $W(t)$. Moreover it is easily verified that

$$(30) \quad \hat{F}_n(\lambda) = \Pr \left\{ \max_{0 \leq t \leq 1} |W_n(t)| \leq \lambda \right\}. $$

**Lemma 3.**

$$\Pr \left\{ \max_{0 \leq t \leq 1} \left| W(t) - W_n(t) \right| \geq \epsilon_n \right\} \leq 2n \exp \left( -2n\epsilon_n^2 \right) = r_n,$$

where $\epsilon_n$ is a sequence tending to zero and whose precise dependence on $n$ will be specified later.

**Proof.** It is only necessary to observe that $W(t) - W_n(t)$ over each of the intervals $[k/n, (k+1)/n]$, $0 \leq k \leq n - 1$, is the $z_n(t)$ process defined in Lemma 2. Thus

$$\Pr \left\{ \max_{0 \leq t \leq 1} |W(t) - W_n(t)| \geq \epsilon_n \right\} \leq \sum_{k=0}^{n-1} \Pr \left\{ \max_{k/n \leq s \leq (k+1)/n} |W(t) - W_n(t)| \geq \epsilon_n \right\}$$

$$= \sum_{k=0}^{n-1} \Pr \left\{ \max_{0 \leq s \leq n^{-1}} |z_n(s)| \geq \epsilon_n \right\}$$

$$\leq 2n \exp \left( -2n\epsilon_n^2 \right).$$
An immediate consequence of Lemma 3 is

**Lemma 4.**

\[ F(\lambda - \varepsilon_n) - r_n \leq F_n(\lambda) \leq F(\lambda + \varepsilon_n) + r_n. \]

Now the distribution function \( F \) is quite smooth; it even has a bounded first derivative, and thus we know there exists a constant \( A_1 \) such that

\[ |F(\lambda + h) - F(\lambda)| \leq A_1|h|, \]

where \( A_1 \) is of course independent of \( \lambda \). We thus conclude

**Lemma 5.**

\[ F(\lambda) - A_1\varepsilon_n - r_n \leq F_n(\lambda) \leq F(\lambda) + A_1\varepsilon_n + r_n. \]

We now proceed to obtain a similar estimate for \( F_n \) and \( F_n \). We remind the reader that according to the Skorokhod representation theorem:

\[ F_n(\lambda) = \Pr \left\{ \max_{1 \leq k \leq n} \left| W\left( \frac{k}{n} + z_{nk} \right) - W\left( \frac{k}{n} \right) \right| \leq \lambda \right\}. \]

Now let \( B_n \) denote the following event:

\[ B_n = \left\{ \max_{1 \leq k \leq n} \left| W\left( \frac{k}{n} + z_{nk} \right) - W\left( \frac{k}{n} \right) \right| \geq \varepsilon_n \right\} \]

and note \( B_n = B_{n1} \cup B_{n2} \) where

\[ B_{n1} = B_n \cap \left\{ \max_{1 \leq k \leq n} |z_{nk}| \leq \delta_n \right\}, \]

and

\[ B_{n2} = B_n \cap \left\{ \max_{1 \leq k \leq n} |z_{nk}| > \delta_n \right\}, \]

where \( \delta_n > 0 \) is a sequence whose dependence on \( n \) will be specified later.

**Lemma 6.**

\[ \Pr \{ B_n \} \leq \frac{8}{(2\pi)^{1/2}} \frac{(\delta_n)^{1/n}}{\varepsilon_n} \exp \left( -\frac{\varepsilon_n^2}{2\delta_n^2} \right) + \frac{2B}{n^{r-1} \delta_n} = \rho_n. \]

**Proof.** We first note that

\[ \Pr \{ B_{n2} \} \leq \Pr \left\{ \max_{1 \leq k \leq n} |z_{nk}| > \delta_n \right\} \leq \frac{2B}{n^{r-1} \delta_n}; \]

this last inequality is an immediate consequence of (27).

As for \( B_{n1} \) we observe that \( B_{n1} \subset E_{n1} \cup E_{n2} \), where

\[ E_{n1} = \bigcup_{k=1}^n \left\{ \max_{0 \leq s \leq \delta_n} \left| W\left( \frac{k}{n} + z \right) - W\left( \frac{k}{n} \right) \right| \geq \varepsilon_n \right\} \]

and

\[ E_{n2} = \bigcup_{k=1}^n \left\{ \max_{0 \leq s \leq \delta_n} \left| W\left( \frac{k}{n} - z \right) - W\left( \frac{k}{n} \right) \right| \geq \varepsilon_n \right\}. \]
It is clear that \( \Pr \{ E_{n1} \} = \Pr \{ E_{n2} \} \) and thus we conclude
\[
\Pr \{ B_{n1} \} \leq 2n \Pr \left\{ \max_{0 \leq t \leq \delta_n} |W(t)| \geq \epsilon_n \right\}.
\]

It is easily shown that,
\[
\Pr \left\{ \max_{0 \leq t \leq \delta_n} |W(t)| \geq \epsilon_n \right\} = \Pr \left\{ \max_{0 \leq t \leq 1} |W(t)| \geq \frac{\epsilon_n}{(\delta_n)^{1/2}} \right\} \leq 4 \Pr \{ W(1) \geq \epsilon_n/(\delta_n)^{1/2} \}.
\]

As is well known, \( \Pr \{ W(1) \geq \epsilon_n/(\delta_n)^{1/2} \} \leq (1/(2\pi)^{1/2})((\delta_n)^{1/2}/\epsilon_n) \exp (-\epsilon_n^2 / 2\delta_n) \) (cf. [5, Chapter 7, Vol. 1]). Putting these inequalities together we conclude:
\[
\Pr \{ B_{n1} \} \leq (8n/(2\pi)^{1/2})((\delta_n)^{1/2}/\epsilon_n) \exp (-\epsilon_n^2 / 2\delta_n)
\]
and this completes the proof of Lemma 6. We have therefore the following inequality:
\[
(34) \quad F_n(\lambda - \epsilon_n) - \rho_n \leq F_n(\lambda) \leq F_n(\lambda + \epsilon_n) + \rho_n.
\]

Applying Lemma 5 with \( \lambda \pm \epsilon_n \) in the place of \( \lambda \) we conclude:
\[
(35) \quad F(\lambda - \epsilon_n) - A_1 \epsilon_n - r_n - \rho_n \leq F(\lambda) \leq F(\lambda + \epsilon_n) + A_1 \epsilon_n + r_n + \rho_n.
\]

Finally by using (31) we obtain the following estimate:
\[
(36) \quad |F_n(\lambda) - F(\lambda)| \leq 2A_1 \epsilon_n + r_n + \rho_n.
\]

It only remains to choose \( \epsilon_n \) and \( \delta_n \) tending to zero at just the correct rate of speed; our choices are:
\[
(37) \quad \delta_n = n^{-\alpha}, \quad \epsilon_n = 2(\log n)^{1/2} n^{-\alpha/2}, \quad 0 < \alpha < (r-1)/r.
\]

We assume \( n \geq 2 \) because for \( n = 1 \), \( \epsilon_1 = 0 \). With these choices for \( \delta_n \) and \( \epsilon_n \) it is easily seen that \( (\epsilon_n^2 / 2\delta_n) = 2 \log n \). Thus, for \( n \geq 2 \),
\[
(38) \quad \rho_n \leq \frac{4}{(2\pi)^{1/2}} \frac{n}{\log n}^{1/2} \frac{1}{n^{1/2}} + \frac{2B}{n^{r-1-\alpha r}} \leq \frac{2(B+1)}{n^{r-1-\alpha r}}.
\]

According to Lemma 3 we have the following estimate for \( r_n \):
\[
(39) \quad r_n = 2n \exp (-8n^{1-\alpha} \log n) \leq 2n \exp (-8 \log n) = 2/n^7.
\]

So \( r_n \) is negligible in comparison to \( \rho_n \) and \( \epsilon_n \). Thus we can assert that
\[
(40) \quad r_n + \rho_n \leq 2(B+2)/n^{r-1-\alpha r}, \quad n = 2, 3, \ldots .
\]

Hence (36) becomes:
\[
(41) \quad |F_n(\lambda) - F(\lambda)| \leq \frac{4A_1(\log n)^{1/2}}{\rho^{1/2}} + \frac{2(B+2)}{n^{r-1-\alpha r}}
\]
where \( A_1 \) and \( 2(B+2) \) are independent of our choice of \( \alpha \). The rate of the convergence is determined by \( \operatorname{Min}(\alpha/2, r-1-\alpha r) \) and so the best choice for \( \alpha \) is obviously
Max Min \((\alpha/2, r-1-\alpha r)\), \(0<\alpha<(r-1)/r\). The solution \(\mu\) satisfies the following simple linear equation: \(\alpha/2=r-1-\alpha r\), which means that \(\alpha=(r-1)/(r+1/2)\) and thus

\[
\mu = \frac{\alpha}{2} = \frac{r-1}{2r+1} = \frac{1}{2} \frac{\alpha}{a+3}.
\]

We have thus shown that we can choose \(\alpha\) so that:

\[
(42) \quad |F_n(\lambda)-F(\lambda)| \leq A(\log n)^{1/2}/n^\mu
\]

where

\[
A = 4A_1+2(B+2), \quad \text{and} \quad \mu = \frac{1}{2}a/(a+3).
\]

This completes the proof of Theorem 1. It is to be observed that \(A_1\) is an absolute constant, independent of the distribution of \(X_t\), whereas \(B\) depends on \(\sigma^2\) and the \((2+a)\)th moment of \(X_t\).

IV. An extension of Theorem 1. Let \(C(0; 1)\) denote the space of continuous functions on the interval \([0; 1]\) endowed with the sup norm topology i.e., if \(f(t)\) and \(g(t)\) are both elements of \(C(0; 1)\) then we denote the distance between them in the usual way:

\[
(43) \quad \|f-g\| = \max_{0 \leq t \leq 1} |f(t)-g(t)|.
\]

We denote by \(P_w\) the Wiener measure on \(C(0; 1)\). By a "functional" \(\mathcal{F}\) we mean a real-valued function \(\mathcal{F}\) with domain \(C(0; 1)\). The functional \(\mathcal{F}\) is said to be uniformly continuous if there exists a positive constant \(K\) with the property that

\[
(43') \quad |\mathcal{F}(f)-\mathcal{F}(g)| \leq K\|f-g\|.
\]

We define a "random broken line" \(S_n(t)\) as follows:

\[
S_n(t) = s_{nk} \quad \text{for} \quad t = \frac{k}{n},
\]

\[
S_n(t) = S_n\left(\frac{k}{n}\right) + n\left[ S_n\left(\frac{k+1}{n}\right) - S_n\left(\frac{k}{n}\right) \right] \left(t - \frac{k}{n}\right), \quad \text{for} \quad \frac{k}{n} \leq t \leq \frac{k+1}{n}.
\]

We define

\[
(45) \quad \Psi_n(\lambda) = \Pr\{\infty < \mathcal{F}(S_n(t)) \leq \lambda\}
\]

and

\[
\Psi(\lambda) = P_w(\infty < \mathcal{F}(f) \leq \lambda).
\]

**Theorem 5.** Let \(\mathcal{F}\) be a uniformly continuous functional and suppose there exists a constant \(L>0\) with the property that

\[
|\Psi(\lambda+h)-\Psi(\lambda)| \leq L|h|.
\]

Then there exists a constant \(A_1>0\) such that

\[
|\Psi_n(\lambda)-\Psi(\lambda)| \leq A_1(\log n)^{1/2}n^{-\mu}, \quad \mu = \frac{1}{2}a/(a+3).
\]
Proof. We first observe that

$$\Pr \{ |F(S_n(t)) - F(W_n(t))| \geq \epsilon_n \} \leq \Pr \{ |S_n - W_n| \geq \epsilon_n/K \} = \rho'_n$$

and similarly

$$\Pr \{ |F(W_n(t)) - F(W(t))| \geq \epsilon_n \} \leq \Pr \{ |W_n(t) - W(t)| \geq \epsilon_n/K \} = r'_n.$$ We thus obtain the following analogue of inequality (36):

$$|\Psi_n(\lambda) - \Psi(\lambda)| \leq 2L\epsilon_n + r'_n + \rho'_n. \tag{46}$$

By making the same choices for $\epsilon_n$ and $\delta_n$ as we did in the proof of Theorem 1 we see that $r'_n$ and $\rho'_n$ are of the same order of magnitude as $r_n$ and $\rho_n$ respectively. Thus, except for a different constant, we have shown that the same order of convergence is valid for the larger class of functionals $F$ satisfying the hypotheses of Theorem 5. It should also be remarked that rates of convergence can also be obtained for functionals $F$ which are assumed to be merely uniformly Hölder continuous of order $\nu$, $0 < \nu < 1$, i.e., replace (43') by the following condition (3):

$$|f-g| = K|f-g|^{\nu}, \quad 0 < \nu < 1.$$ We shall leave it to the reader to explore these and other refinements.

V. Appendix. The purpose of this part of the paper is to give a proof of inequality (21). A close examination of Skorokhod's proof of his inequality (15) shows that it suffices to establish the following inequality on the moments of the first passage times of Brownian motion. Let $a < 0 < b$ and let

$$\tau(w) = \inf_{t > 0} (w(t) - a)(w(t) - b) = 0,$$

i.e., $\tau$ is the first time that the Brownian motion starting at 0 hits $a$ or $b$. We shall prove the following inequality:

**Theorem 6.** There exists a constant $C_r, r \geq 1$ such that $E(\tau^r) \leq C_r|ab|^{r-2}$.

In order to avoid absolute value signs we make the following simple translation of the coordinate axis: we set $x = |a|$, $y = b + |a|$, we begin the Brownian motion at $x$ and denote by $\tau_x$, the first time that the Brownian motion starting at $x$ hits 0 or $y$. Then the inequality to be proved is the following:

$$E(\tau_x^r) \leq C_r(x(y-x))^{r-2}. \tag{47}$$

In the course of the proof of this result two elementary inequalities are needed, which we state now in order to avoid interrupting the argument at a later stage.

$$|\sin mx| \leq m|\sin x|, \quad m = 1, 2, \ldots. \tag{48}$$

(3) This observation is due to Dr. Stanley Sawyer of the Courant Institute of Mathematical Sciences.
Inequality (48) is trivially true for \( m = 1 \) and it is easily proved in general by mathematical induction. Another useful inequality is the following:

\[
\text{Max}_{0 \leq t \leq 1} \left| \frac{\sin \left( \pi t \right)}{t(1-t)} \right| = 4 \left| \frac{\pi}{y^2} \right| .
\]

**Proof.** Set \( t = x/y \), holding \( y \) fixed. Then

\[
\frac{\sin \left( \pi t \right)}{t(1-t)} = \frac{1}{y^2} \frac{\sin x}{x(y-x)} .
\]

Now as \( x \) varies between 0 and \( y \), \( t \) varies between 0 and 1. Thus,

\[
\text{Max}_{0 \leq t \leq 1} \left| \frac{\sin \left( \pi t \right)}{t(1-t)} \right| = \frac{1}{y^2} \text{Max}_{0 \leq t \leq 1} \left| \frac{\pi}{t(1-t)} \right| = 4 \left| \frac{\pi}{y^2} \right| ,
\]

and this completes the proof of (49).

Let \( H(t) = \Pr \{ \tau_x \leq t \} \). The following infinite series for \( 1 - H(t) \) is well known (cf. [5, Vol. 2, p. 330]):

\[
1 - H(t) = \frac{4}{\pi} \sum_{n=0}^{\infty} \left( 2n+1 \right)^{-1} \exp \left( -\frac{(2n+1)^2 \pi^2 t}{2y^2} \right) \sin \left( \frac{(2n+1)\pi x}{y} \right) .
\]

It follows at once from an integration by parts that

\[
\int_{0}^{\infty} t \, dH(t) = \int_{0}^{\infty} t^{-1}(1-H(t)) \, dt = E(\tau_x).
\]

We now apply inequalities (48) and (49) successively to conclude:

\[
\left| \frac{\sin \left( \frac{(2n+1)\pi x}{y} \right)}{y} \right| \leq (2n+1) \left| \frac{\pi x}{y} \right| \leq 4(2n+1) \frac{x(y-x)}{y^2} .
\]

Hence

\[
0 \leq 1 - H(t) \leq \frac{16x(y-x)}{\pi y^2} \sum_{n=0}^{\infty} \exp \left( -\frac{(2n+1)^2 \pi^2 t}{2y^2} \right) .
\]

Integrating term by term, as we may, we conclude

\[
\int_{0}^{\infty} t^{-1}(1-H(t)) \, dt \leq \frac{16x(y-x)}{\pi y^2} \sum_{n=0}^{\infty} \int_{0}^{\infty} t^{-1} \exp \left( -\frac{(2n+1)^2 \pi^2 t}{2y^2} \right) \, dt.
\]

Now a simple change of variables yields

\[
\int_{0}^{\infty} t^{-1} \exp \left( -\frac{(2n+1)^2 \pi^2 t}{2y^2} \right) \, dt = \frac{2y^2 \Gamma(r)}{(2n+1)^{2r} \pi^{2r}} ,
\]

where \( \Gamma \) denotes the Gamma function. Thus (54) becomes

\[
r \int_{0}^{\infty} t^{-1}(1-H(t)) \, dt \leq C_n x(y-x)y^{2r-2} ,
\]
where

\[ C_r = 2^{r+1} \Gamma(r+1) \pi^{-2r} \sum_{n=0}^{\infty} (2n+1)^{-2r}. \]

This completes the proof of (47) and, as we have already remarked, this result when applied to Skorokhod's proof of (15) yields inequality (21).

Acknowledgements. In the course of obtaining these results the author benefited from useful conversations with Professor S. R. S. Varadhan of the Courant Institute of Mathematical Sciences. I also wish to thank the Department of Statistics of the University of California, Berkeley, where parts of this paper were written, for the use of their facilities.

REFERENCES


NEW YORK UNIVERSITY,
NEW YORK, NEW YORK