ON THE SKOROKHOD EMBEDDING THEOREM

BY

W. J. HALL

TECHNICAL REPORT NO. 33
OCTOBER 4, 1968

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-5705
AND
U.S. PUBLIC HEALTH SERVICE GRANT GM-10397

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE SKOROKHOD EMBEDDING THEOREM

by

W. J. Hall

TECHNICAL REPORT NO. 33
OCTOBER 4, 1968

PREPARED UNDER THE AUSPICES OF
THE NATIONAL SCIENCE FOUNDATION GRANT GP-5705
AND U.S. PUBLIC HEALTH SERVICE GRANT GM-10397

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE SKOROKHOD EMBEDDING THEOREM

by W. J. Hall

Stanford University and University of North Carolina

1. Introduction. The Skorokhod embedding theorem (Skorokhod, 1961) states that, if \( \{S_n\} \) is a sequence of cumulative sums of independent r.v.'s with zero means, and if \( W(\cdot) \) is a standard Wiener process, then there is a sequence \( \{T_n\} \) of successive stopping times defined on \( W(\cdot) \) and on a sequence of r.v.'s \( \{R_n\} \) independent of \( W(\cdot) \) such that the two sequences \( \{W(T_n)\} \) and \( \{S_n\} \) have the same law.

In this note we present a short proof of this theorem, without recourse to the randomization sequence \( \{R_n\} \). Other proofs have been given by Dubins (1968) and Root (1968). Proofs using randomization appear in Skorokhod (1961), Freedman (1967) and Breiman (1968). All proofs coincide for simple random walks \( \{S_n\} \).

In Dubins' construction of \( \{T_n\} \), each \( T_n \) is the least upper bound of an infinite sequence of successive stopping times \( T_{n1}, T_{n2}, \ldots \). We use Dubins' first stage \( T_{n1} \) and then a randomized stage \( T_{n2} = T_n \), basing the randomization on the path of \( W(\cdot) \) prior to \( T_{n1} \) rather than on an independent \( R_n \).

Some properties of the \( \{T_n\} \) sequence are given in Section 4. In Section 5, the embedding theorem is applied to the study of martingale

---

1 Research supported by National Science Foundation Grant GP-5705 and U.S. Public Health Service Grant GM-10397.
sequences with two absorbing barriers and bounded increments. Other applications of the Skorokhod Theorem appear, for example, in Skorokhod (1961), Strassen (1966), Freedman (1967), Breiman (1968), Root (1968), and Hall (1968b).

An Appendix provides a proof of the following fact, appearing in Lévy (1937) and used herein: there exists a measurable transformation from a single random variable with a continuous distribution to a random vector with an arbitrary distribution.

2. **Lemma.** There exists a finite-mean stopping time $T$ on the path of $W(\cdot) = \{W(t); t \geq 0, W(0) = m\}$ for which

\[(1) \quad W(T) \overset{d}{=} X \quad ("\text{equal in law})"

iff $E(X) = m$ and $\text{var } X < \infty$; moreover, $E(T) = \text{var } X$.

First note that, without the requirement that $T$ have finite mean, the problem of finding a $T$ to satisfy (1) is simply solved for any r.v. $X$. For, let $T'$ be the hitting time by $W(\cdot)$ of a level $L(\neq m)$, and let $X' = f(T')$ where $f(\cdot)$ (measurable) is so chosen that $X' \overset{d}{=} X$ (see Appendix). Let $T$ be the subsequent ($T' \leq T$) hitting time of $X'$. Then, clearly, (1) holds. But $ET = \infty$ (although $T < \infty$ a.s.). In applications, it is frequently desirable to control the variability of $T$.

The necessity proof of the Lemma is immediate from Wald's first and second equations (Shepp (1967) and Hall (1968a)): For any finite-mean stopping time $T$ on $W(\cdot)$ (with $W(0) = m$), $E(W(T)) = m$ and $\text{var } W(T) = ET < \infty$. 

2.
Now turning to the sufficiency proof, we first give a short constructive proof of (1) using randomization, defining $T$ on the path of $W(\cdot)$ and on r.v.'s $U$, $V$, independent of $W(\cdot)$. Let $(U,V)$ have distribution function (d.f.) $H(\cdot,\cdot)$ defined by $dH(u,v) = k^{-1}(u-v)\,dF(u)dF(v)$ for $v \leq m < u$, where $F$ is the d.f. of $X$ and $k = E[I(X > m)(X-m)]$, $I(\cdot)$ being the indicator function. Let $T$ be the hitting time of the set $[U,V]$. Then, for $x > m$,

$$
\Pr[W(T) > x] = E\Pr[W(T) > x | U,V] = E[I(U > x)\Pr[W(T) = U|U,V]}
$$

$$
= \int_{-\infty}^{m} \int_{x}^{\infty} \frac{m-v}{u-v} dH(u,v) \quad \text{since}
$$

(2) $\Pr[W(\cdot) \text{ hits } a \text{ before } b, \text{ starting from } m(b \leq m < a)] = \frac{m-b}{a-b}$

(e.g., Skorokhod, pg.166; also see (iv) in Section 4 below). The double integral readily reduces to $1 - F(x)$. Similarly, for $x \leq m$, $\Pr[W(T) \leq x] = F(x)$. Hence (1) holds.

We now give a two-stage randomized construction of $T$. Let $A^+$ denote the event $[X > m]$ and $A^- = [X \leq m]$, and let $m^i = E(X|A^i)$ where $i = \pm$. Let $T'$ be the hitting time of $(m^+,m^-)$. Note that $\Pr[W(T') = m^+] = (m-m^-)/(m^+-m^-) = p^+$, defined as $\Pr[A^+]$, since

$m = p^+ m^+ + p^- m^-$ where $p^- = \Pr[A^-]$. We now argue conditionally on $W(T') = m^+$ or $m^-$ and on $A^+$ or $A^-$, respectively. Let $W'(t) = W(t+T')$ with initial value $W(T')$, and, using a suitable $(U,V)$, define $T''$ as in the previous paragraph so that $\Pr[W'(T'') \leq x | W'(0) = m^i] = \Pr[X \leq x | A^i](i = \pm)$. Then, with $T = T' + T''$, 3.
\[ \Pr[W(T) \leq x] = \sum_{i=1}^{m} \Pr[W(T) \leq x | W'(0) = m^i]p^i = \sum_{i=1}^{m} \Pr[X \leq x | A^i]p^i = F(x), \]
so that (1) holds.

Finally, we make the randomization internal. Let \( T^* \) be the hitting time of \( \{m+a, m-a\} \) where \( a = \min(m^+, m^-) \), so that \( T^* \leq T' \), and so that \( T^* \) and \( W'(\cdot) \) are independent; also \( T^* \) is well-known to have a continuous distribution. Let \( (U,V) = f^i(T^*) \)
\((i=+)\) where \( f^i \) is so chosen that \((U,V)\) has the appropriate joint d.f., say \( H^i \), required for the construction of \( T'' \). The existence of such an \( f^i \) is verified in the Appendix. (It may be conceptually simpler to replace the role of \( T^* \) by \( (T^*, T^{**}) \) where \( T^{**} \) is the hitting time of \( \{m+\frac{1}{2}a, m-\frac{1}{2}a\} \), say, so that \( f^i: (T^*, T^{**}) \to (U,V) \) rather than \( f^i: T^* \to (U,V) \).

3. THEOREM. There exists a sequence of successive finite-mean stopping times \( \{T_n\} \) on \( W(\cdot) = (W(t); t \geq 0) \) for which

\[ (W(T_n)) \sim (Y_n) \]

iff \((Y_n)\) is an \( L_2 \)-martingale sequence and \( W(0) = EY_1 \); moreover,
\( E\) \( T_n = \text{var} Y_n \).

As in the proof of the Lemma, \((W(T_n))\) — for any finite-mean \( T_n \)'s — is readily seen to be an \( L_2 \)-martingale sequence. For,
\( W(T_n) = Z_n \), and \( \mathcal{G}_n \) the Borel field generated by \( Z_1, \ldots, Z_n \), \( E(Z_{n+1} | \mathcal{G}_n) = E(E[Z_{n+1} | \mathcal{G}_n, T_n] | \mathcal{G}_n) = E(Z_n | \mathcal{G}_n) \) (by Wald's equation) = \( Z_n \) a.s. And \( \text{var} W(T_n) < \infty \) as before.

The construction of \( T_n \) to satisfy (3) is done recursively, applying the Lemma successively to the conditional distribution of
Y_n given \( \mathcal{B}_n \) and to \( W_n(\cdot) = (W_n(t) = W(t+T_{n-1}); t \geq 0) \); see Dubins (1968) for details.

Note that an \( L_2 \)-martingale sequence \( \{ Y_n, \mathcal{F}_n \} \) can be embedded in \( W(\cdot) \) since \( \mathcal{F}_n \supseteq \mathcal{B}_n \) so that \( \{ Y_n \} = \{ Y_n, \mathcal{B}_n \} \) is also a martingale sequence.

4. Remarks. (i) T of the Lemma may be so chosen that \( ET^k < \infty \) iff \( EX^{2k} < \infty \) (k a positive integer); in fact, \( ET^k < c_k EX^{2k} \). For, if \( ET^k < \infty \), then \( EX^{2k} < \infty \) by a continuous-time version of Brown's (1968) results on moment identities (see Hall, 1968a). The converse may be proved for the \( T \) constructed in Section 2 very much as in Skorokhod (1961).

Root (1968) proved this for his construction of \( T \) also.

(ii) If the increments in \( \{ Y_n \} \) are bounded, so is the path of \( W(\cdot) \) between stopping times. Specifically, if \( b_n \leq Y_n - Y_{n-1} \leq a_n \) a.s., then \( b_n \leq W(t) - W(T_{n-1}) \leq a_n \) a.s. for \( t \in [T_{n-1}, T_n] \). This is obvious from the construction of \( \{ T_n \} \) and indeed must hold if the \( T_n \)'s are to have finite means.

(iii) Bounds on the variance of \( T \) (upper and lower) for any \( T \) satisfying (i) are given in Corollary 4 of Hall (1968a), together with conditions for equality.

These bounds are \( 2\sigma_4^4 [\sqrt{(1+\frac{1}{2} \gamma)} + \sqrt{(1+\frac{1}{2} \gamma)}]^2 \) where \( \gamma \) is the kurtosis coefficient of \( X \) (fourth cumulant/\( \sigma^4 \), \( \geq -2 \)). For example, if \( X \) is normal \( (m, \sigma^2) \), the bounds on \( \text{var } T \) are \( 8\sigma_4^4 \) and 0, the lower bound being attained by Root's construction \( (T_0, \sigma^2 \text{ a.s.}) \), whereas a one-stage randomized \( T \) (see proof of Lemma) may be shown to have \( \text{var } T = 8\sigma_4^4 / 3j \), our two-stage \( T = T' + T'' \), and Dubins' \( T \).
would have smaller variances. How typical this example is is not known. Our $T''$ can be so chosen to be negatively correlated with $T'$, with a consequent reduction of $\text{var } T$.

The bounds also apply to $\text{var } T_n$. Each $T_n$ can presumably be made to be negatively correlated with $T_{n-1}$, by letting the path prior to $T_{n-1}$ generate the randomization for $T_n$ — reducing $\text{var } T_n$. But explicit evaluations do not seem feasible.

(iv) Other embedding processes may be used. This was pointed out by Dubins (1968). In our construction of Section 2, for example, the properties of $W(\cdot)$ that were needed for (2) were that it have continuous sample paths, hit $\{U,V\}$ a.s., and be a martingale. (That these are sufficient for (2) to hold is readily verified. If $T$ is the hitting time of $(a,b)$, then the r.v. $W(T)$ can only equal $a$ or $b$, and therefore its distribution is completely determined by its mean; and $E W(T) = m$ from martingale theory.) In addition, the strong Markov property of $W(\cdot)$ was needed to combine stages in the construction; this was implicitly used again in Section 3. Hence, the same construction could be used to embed $X$, or a martingale $(Y_n)$, in most any diffusion process $Z(\cdot)$ with the martingale property. Characterizations of such processes have been given by Arbib (1965). Note, however, that all claims about moments of $T$ utilized other properties specific to the Wiener process.

The martingale $\{e^{-8W(t)}; t \geq 0\}$, where $W(\cdot)$ is a Wiener process with drift $8$, is used as the embedding process in Hall (1968b).
5. **Martingales with two absorbing barriers.** We shall study a martingale sequence \( \{Y_n\} \) with two absorbing "barriers" \((-\infty, bc]\) and \([ac, +\infty)\) for large \(c\). Since \(\{Y_n\}\) may be embedded in \(W(\cdot)\), we can equivalently consider the classical problem of a Wiener process with two absorbing barriers, except that the barriers are applied only at the embedding times \(\{T_n\}\), rather than for all \(t\). However, if \(W(\cdot)\) cannot vary much between the \(T_n\)'s, the classical problem should serve as a good approximation to the martingale problem. This indeed is verified in the theorem below, under the assumptions that \(\{Y_n\}\) has (i) uniformly bounded increments and (ii) non-random conditional variances converging in Cesàro-mean, and \(c \to \infty\). Some of the results, presented in the lemmas, are valid under milder assumptions. The results on expected absorption time are reminiscent of those obtained by Chow and Robbins (1963) for a related problem. Relevant results on the classical Wiener process problem may be found in Cox and Miller (1965) and in Feller (1966).

Throughout, \(\{Y_n\}\) is a zero-mean martingale sequence with increments \(\{X_n\}\), and \(W(\cdot)\) is a standard Wiener process. Superscripts denote conditioning. We denote, for given \(\beta < 0 < \alpha\),

\[
N(\alpha, \beta) = \min(n \mid Y_n \leq \beta \text{ or } \geq \alpha) \leq \infty
\]

\[
T(\alpha, \beta) = \min(t \mid W(t) \leq \beta \text{ or } \geq \alpha).
\]

**Lemma 5.1:** If

\[
(4) \quad \lim_{n \to \infty} \sup_n \Pr^n[|X_n| \geq \eta] \geq \delta \text{ a.s. for some } \eta > 0, \delta > 0,
\]
then

\[ N(\alpha, \beta) < \infty \text{ a.s. for } -\infty < \beta < 0 < \alpha < +\infty. \]

**Proof:** (4) implies \( \sum_{n=1}^{\infty} \Pr^{n-1}[|X_n| > \eta] = \infty \) so that, by the Lévy form of the Borel-Cantelli Lemma, \(|X_n| > \eta\) infinitely often a.s.

Hence \( \sup_n |Y_n| = \infty \) and (5) follows.

**Lemma 5.2:** If \(|X_n| \leq K < \infty \) a.s. for all \( n \), then (6) or (7) below implies (5):

\[ \limsup_{n \to \infty} E^{n-1}|X_n|^r \geq \epsilon (> 0) \text{ a.s. for some } r > 0; \]

\[ \limsup_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} E^{m-1}X_m^2 \geq \epsilon (> 0) \text{ a.s.} \]

**Proof:** From (6), we have \( \epsilon \leq E^{n-1}|X_n|^r = f(|X_n| \geq \epsilon/2)|X_n|^r \]

\[ + f(|X_n| < \epsilon/2)|X_n|^r \leq K^{r} P^{n-1}[|X_n| \geq \epsilon/2] + (\epsilon/2)^r \text{ so that (4)} \]

holds with \( \eta = \epsilon/2 \) and \( \delta = [\epsilon-(\epsilon/2)^r]K^{-r} \). That (7) implies (6) with \( r = 2 \) is immediate.

**Lemma 5.3:** If \(|X_n| \leq K \) a.s. and \( N(ac, bc) < \infty \) a.s. \((b < 0 < a)\), then

\[ \lim_{c \to \infty} \Pr[Y_N \leq bc] = a/(a-b) = \Pr[W_T \leq b] \]

8.
where \( N = N(ac, bc) \) and \( T = T(a, b) \). (Bounds on \( \Pr[Y_N \leq bc] \) appear in the proof.)

**Proof:** We embed \( \{Y_n\} \) in \( W(\cdot) \). Then \( \Pr[Y_N \geq ac] = \Pr[W(T_N) \geq ac] \leq \Pr[W(T(ac, bc-K)) = ac] \) for, whenever \( W(T_m) > bc \) for all \( m < n \), \( W(t) \) will not have reached \( bc-K \) between stopping times \( T_1, \ldots, T_n \) (Remark (ii) of Section 4). But \( \Pr[W(T(ac, bc-K)) = ac] = -(bc-K)/(ac-bc+K) \) (by (2)) = \(-b/(a-b) + O(1/c)\). Similarly, \( \Pr[Y_N \leq bc] = \Pr[W(T(ac+K, bc)) = bc] = \frac{(ac+K)}{(ac-bc+K)} = ab/(a-b) + O(1/c) \). Equation (8) thus follows. (The second equality in (8) is equivalent to (2).)

**Lemma 5.4:** Suppose \( |X_n| \leq K \) a.s. If \( n^{-1} \sum_{m=1}^{n} E^{m-1} X_m^2 \leq \sigma_U^2 \leq K^2 \) a.s., then \( -ab \leq \frac{2}{c} - 2E(ac, bc) \leq \sigma_L^2 \) a.s., then \( \frac{2}{c} - 2E(ac, bc) \leq \frac{1}{K} \), so that \( \sigma_L^2 - \frac{1}{K} \leq ab \leq 0(1/c) \) as \( c \to \infty \).

**Proof:** Let \( \{Y_n\} \) be embedded in \( W(\cdot) \). Denote \( \Sigma_n = \sum_{m=1}^{n} E^{m-1} X_m^2 \) and \( \Sigma'_n = \sum_{m=1}^{n} E^{m-1} [W(T_n) - W(T_m-1)]^2 \) so that \( \Sigma_n \leq \Sigma'_n \). We use the fact that \( \{V_n\} \), defined by \( V_n = T_n - \Sigma_n' \), is a zero-mean martingale sequence since, according to Lemma 2, the conditional mean of the increment in \( T_n \) equals the conditional variance of the increment of \( Y_n \) or, equivalently, the increment in \( W(T_n) \).

Let \( T = T(ac, bc), T^* = T(ac+K, bc-K), N = N(ac, bc) \), and \( N' = \min(n \mid W(T_n) \geq ac \text{ or } \leq bc) = N \). Using Remark (ii) of Section 4, \( T \leq T_N \leq T^* \). Also, \( ET = -abc^2 \) and \( ET^* = -(ac+K)(bc-K) \) (by Wald’s equation), and \( ET_{N'} = ET_N \) lies between.
Now \( V'_N \leq V'_N^+ \leq T'_N \leq T^* \) so that \( EV'_N \leq ET^* < \infty \) and thus \( EV'_N \) exists and \( EV'_N < \infty \).

Next, \( \int [N' > n]V'_n \leq \int [N' > n]T_n \leq \int [N' > n]T^* = ET^* - \int [N' \leq n]T^* \to 0 \) as \( n \to \infty \). The martingale times theorem (Loeve, 1963, pg. 532) therefore implies that \( EV'_N \geq EV_1 = 0 \). Hence, \( ET'_N \geq EV'_N \geq EV'_N \geq ENc_L^2 \) from which the second part of the lemma follows. (It is convenient to proceed with part of the proof of (iii) of Theorem 5 below. Under the assumptions of Theorem 5, the last inequality may be replaced by: \( EV'_N > \int [N > n]N \geq (\gamma - \epsilon) \int [N > n]^N N \) for \( n > n_\epsilon \), so chosen that \(|n^{-1}N - \gamma^2| \leq \epsilon \) for \( n > n_\epsilon \), and the \( RHS = (\gamma - \epsilon)[EN - \int [N \leq n]N] \geq (\gamma - \epsilon)(EN - n) \); hence \( EN < \infty \) and, in fact, \( c^{-2}EN \leq c^{-2}(\gamma - \epsilon)^{-1}[-abc^2 + o(c)] + c^{-2}n \to -ab(\gamma - \epsilon)^{-1} \) as \( c \to \infty \). Since \( \epsilon \) is arbitrary, \( \lim sup \sigma_c^{-2}EN \leq -ab \).

We now consider the first part of Lemma 5,4. There is nothing to prove unless \( EN < \infty \), which we now assume. Then \( \int [N' > n]V'_n \leq \int [N' > n]N' \leq n^2 u Pr[N' > n] \to 0 \) since \( EN' = EN < \infty \). The martingale times theorem therefore implies \( EV'_N \leq EV_1 = 0 \), i.e., \( ET'_N \leq EV'_N = EV'_N \). Therefore \( -abc^2 = ET \leq ET'_N \leq EV'_N \leq ENc_U^2 \). (Under the assumptions of Theorem 5, the last inequality may be replaced by \( EV'_N \leq EV'_{N+n} \leq (\gamma + \epsilon)E(N+n) \) for \( n > n_\epsilon \), so that \( c^{-2}EN \geq -ab(\gamma + \epsilon)^{-1}c^{-2}n \to -ab(\gamma + \epsilon)^{-1} \), completing the proof of (iii) of Theorem 5.)

The following is a generalization of Skorokhod's Lemma 1, pg. 170,
needed in the proof of Theorem 5 to follow. Skorokhod's proof may be
easily generalized:

**Lemma 5.5 (Skorokhod):** Suppose for each positive integer \( m \)
\( \{Y_{mn}, n=1, \ldots, m\} \) is a zero-mean martingale sequence with increments
\( \{X_{mn}\} \) where \( E^{n-1}X_{mn}^2 \leq B/m \) and \( E^{n-1}X_{mn}^{1/4} \leq C/m^{1/2} \). Then
\[
\Pr[\sup_n \frac{|Y_{mn}|}{m} > 2 \forall m] \leq A/m.
\]

**Theorem 5:** If \( |X_n| \leq k \) a.s., if \( E^{n-1}X_n^2 = \sigma_n^2 \) (constant) a.s.,
and if \( n^{-1} \sum_{m=1}^n \sigma_n^2 \to \sigma^2 (> 0) \) as \( n \to \infty \), then

(i) \( N(ac,bc) < \infty \) a.s.,

(ii) (8) holds,

(iii) \( \sigma^2 c^{-2} E N(ac,bc) \to -ab = ET(a,b) \) as \( c \to \infty \), and

(iv) \( \sigma^2 c^{-2} N(ac,bc) \overset{L}{\to} T(a,b) \) as \( c \to \infty \).

The second-order conditions in the theorem can be relaxed to
\( n^{-1} \sum_{m=1}^n E^{m-1}X_m^2 \to \sigma^2 \) unif. a.s., but we have not succeeded in proving
all of these results without the uniformity.

**Proof:** (i) follows from Lemma 5.2(7), and (ii) holds by Lemma 5.3;
(iii) was proved parenthetically along with Lemma 5.4. Only (iv)
remains to be proved. Our proof is patterned after Skorokhod's proof
of his theorem on pg. 170.

For fixed positive \( x \), let \( m = m(c) = [xc^2/\sigma^2] \) and \( Y_{mn} = m^{-1/2}Y_{mn} \).
Embed \( \{Y_{mn}, n \leq m\} \) in \( W(\cdot) \) so that \( \{Y_{mn}\} \overset{D}{=} \{W(T_{mn})\} \) for each \( m \).
We also consider the zero-mean martingale \( \{Y'_{nm}; n \leq m\} \) with increments \( X'_{nm} = \sqrt{m} \left( \frac{T_{nm} - T_{m,n-1}}{-\hat{\sigma}^2/m} \right) \). It may be shown to satisfy the conditions of Lemma 5.5, using Remark (i) of Section 4. We note for later use that, with \( t_m = m^{-1/2} \Sigma_m - 2^{-1/2} \Sigma_m \) and \( \Sigma_m = \sum_{n=1}^{m} \hat{\sigma}_n^2 \),

\[ \Pr[T_{nm} < t_m] = \Pr[-Y'_{nm} > 2 \Delta n m] \leq \Pr[\sup_{n \leq m} |Y'_{nm}| > 2 \Delta n m] \leq A/m \text{ by Lemma 5.5.} \]

The following notation will be used:

\[ \alpha = a + \sqrt{x} \quad \quad \alpha_m = \alpha + K/\sqrt{m} \]

\[ \alpha' = \alpha_m \sqrt{(x/t_m)} = a + o(1) \text{ as } c \to \infty, \]

and likewise \( \beta, \beta'_m, \beta'_m \) with \((a,K)\) replaced by \((b,-K)\). We assume \( c \) sufficiently large that \( t_m > 0 \).

Now \( P_c(x) = \Pr[\sigma^2 c^{-2}N(ac,bc) > x] = \Pr[bc < Y_n < ac \text{ for all } n \leq m] = \Pr[\beta < Y_{mn} < \alpha \text{ for } n \leq m] \leq \Pr[\beta_m < W(t) < \alpha_m \text{ for } t \leq T_{mm} \} \text{ (using Remark (ii))} \leq \Pr[\beta_m < W(t) < \alpha_m \text{ for } t \leq t_m] + \Pr[T_{mm} < t_m] \text{ for any } t_m, \text{ and in particular for } t_m \text{ defined above. The second term on the RHS is } O(m^{-1}). \]

Considering the first term, which we denote \( A_m \), we replace the index \( t \) by \( st_m / x \) and note that \( W'(\cdot) = (\sqrt{(x/t_m)} W(st_m / x); s \geq 0) \) is a standard Wiener process. Hence, \( A_m = \Pr[\beta'_m < W'(s) < \alpha'_m \text{ for } s \leq x] = \Pr[T(a+o(1), b+o(1)) > x] = \Pr[T(a,b) > x] + o(1) \text{ as } c \to \infty. \) (Formal proof of the last step is straightforward and omitted.) We thus have proved that \( \limsup P_c(x) \leq \Pr[T(a,b) > x] \).

That \( \liminf P_c(x) \geq \Pr[T(a,b) > x] \) may be likewise proved.
Acknowledgement. The author is grateful to L. Dubins for introducing him to the Skorokhod Theorem, to D. Siegmund for discussions leading to Lemmas 5.1 and 5.2, and to S. G. Ghurye and R. Olshen for discussion of the Appendix problem.
REFERENCES


BROWN, B. M. (1968). Moments of a stopping rule related to the central limit theorem. Mimeo Series. No. 159, Department of Statistics, Purdue University.


SKOROKHOD, A. V. (1961). Translated from the Russian in 1965 as:  
Studies in the Theory of Random Processes. Addison-Wesley,  
Reading, Mass. (Chapter 7.)

APPENDIX: ON GENERATING RANDOM VECTORS

The following fact appears in Lévy (1937):

**THEOREM:** If $U$ is a random variable, uniformly distributed on the unit interval, and $H(\cdot, \cdot)$ is an arbitrary distribution function on the plane, then there exists a measurable transformation $T: U \to (X,Y)$ so that $(X,Y)$ has joint distribution function $H$.

Alternatively, $U$ could have an arbitrary continuous distribution. The generalization to higher dimensions is obvious. The purpose here is to record a proof of this theorem.

First note, as did Lévy, that $T$ can be constructed as follows when $(X,Y)$ are to be independent and uniformly distributed: Let the $i^{th}$ decimal place in $X[Y$, resp.] be given by the $(2i-1)^{th}$ $[(2i)^{th}$, resp.] decimal place in $U$. Hence, it is sufficient to prove the existence of a measurable $T': (U,V) \to (X,Y)$ where $(U,V)$ are independent and uniformly distributed and $H$ is arbitrary.

If the distribution function (d.f.) $H$ of $(X,Y)$ is absolutely continuous, with $X$ having marginal d.f. $F(\cdot)$ and $Y$ having conditional d.f. $G(\cdot | \cdot)$ given $X$, then $U = F(X)$ and $V = G(Y | X)$ are readily seen to be independent and uniformly distributed. The inverse transformation, $X = F^{-1}(U)$ and $Y = G^{-1}(V | F^{-1}(U))$, may be used to define $T'$ — without the absolute continuity assumption, so long as the inverses are carefully defined. This too is discussed by Lévy. However, a measurability question arises, which we first settle in

16.
LEMMA A: Suppose $G(\cdot \mid \cdot)$ is a conditional d.f. for which $G(\cdot \mid x)$ is a d.f. for every real $x$ and $G(y \mid \cdot)$ is a Borel-measurable function for every real $y$. Then $G(\cdot \mid \cdot)$ is a Borel-measurable function on the plane.

Proof: Without loss of generality, assume the range of $y$ is in $[0,1]$. Let $G_n(y \mid x) = G(r_n(y) \mid x)$ where $r_n(y) = k2^{-n}$ and $k$ is the integer for which $y \in I_{kn} = [(k-1)2^{-n}, k2^{-n})$. Then $r_n(y) \downarrow y$ as $n$ increases, and $\lim_{n \to \infty} G_n(y \mid x) = G(y \mid x)$ by the right continuity of $G(\cdot \mid x)$ for each $x$. Hence, it is sufficient to show that $G_n(\cdot \mid \cdot)$ is measurable for every $n$.

For fixed $c$ and $n$, let $B_k = \{x \mid G(k2^{-n} \mid x) \leq c\}$, a measurable set. Then $\{x,y \mid G_n(y \mid x) \leq c\} = \bigcup_{k=1}^{2^n+1} \{x,y \mid y \in I_{kn} \text{ and } G_n(y \mid x) \leq c\} = \bigcup \{x,y \mid y \in I_{kn} \text{ and } G(k2^{-n} \mid x) \leq c\} = \bigcup (I_{kn} \times B_k)$, a union of measurable sets, and hence measurable.

For completeness, we record the rest of the proof of the theorem. It is convenient to handle the one-dimensional version separately as

LEMMA B: Let $U$ be uniform on $(0,1)$ and $F$ an arbitrary d.f. on the real line. There exists a transformation $T: U \to X$ so that $X$ has d.f. $F$.

Proof: Let $F^{-1}(u) = \inf(x \mid F(x) \geq u)$, a measurable function. Then, for each $x$ in $\mathbf{R}$, the support of $F$, $u \leq F(x)$ iff $F^{-1}(u) \leq x$. (A diagram makes it obvious.) With $X = F^{-1}(U)$, we thus have
Pr(X ≤ x) = Pr(U ≤ F(x)) = F(x) for x in \mathcal{X}_F, and hence for all x.

**Proof of Theorem:** (X,Y) is to have d.f. H. Let G(y|·) represent the corresponding conditional probability of \{Y ≤ y\} given X, chosen to satisfy the hypothesis of Lemma A; thus H(x,y) = E[I(X ≤ x) G(y|X)] where I(·) is the indicator function. Let F be the corresponding marginal d.f. of X. Let G⁻¹(v|x) = inf(y|G(y|x) ≥ v). Then, for each x in \mathcal{X}_F, the support of F, and each y in \mathcal{Y}_G(x), the support of G(·|x), v ≤ G(y|x) iff G⁻¹(v|x) ≤ y. Let X = F⁻¹(U) (defined in the proof of Lemma B) and Y = G⁻¹(V|F⁻¹(U)), a measurable function by Lemma A. Then, for x in \mathcal{X}_F and y in \mathcal{Y}_G(x), Pr(X ≤ x, Y ≤ y) = E[I[F⁻¹(U) ≤ x] · Pr[G⁻¹(V|F⁻¹(U)) ≤ y|U]) = E[I(F⁻¹(U) ≤ x] Pr[V ≤ G(y|F⁻¹(U))|U]) = E[I(X ≤ x)G(y|X)] = H(x,y), and hence for all (x,y).