EMBEDDING SUBMARTINGALES IN WIENER PROCESSES WITH DRIFT,
WITH APPLICATIONS TO SEQUENTIAL ANALYSIS

by

W. J. Hall

TECHNICAL REPORT NO. 34
NOVEMBER 27, 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
EMBEDDING SUBMARTINGALES IN WIENER PROCESSES WITH DRIFT,
WITH APPLICATIONS TO SEQUENTIAL ANALYSIS

by W. J. Hall
Stanford University and University of North Carolina

0. Summary.

Skorokhod (1961) demonstrated how the study of martingale sequences (and zero-mean random walks) can be reduced to the study of the Wiener process (without drift) at a sequence of random stopping times. We show how the study of certain submartingale sequences, including certain random walks with drift and log likelihood ratio sequences, can be reduced to the study of the Wiener process with drift at a sequence of stopping times. Applications to absorption problems are given; specifically, new derivations of a number of the basic approximations and inequalities of classical sequential analysis - and some variations on them - appear.

1. Introduction.

The Skorokhod (martingale) embedding theorem (Skorokhod, 1961) states that, if \( \{Y_n\} \) is a zero-mean martingale sequence and if \( W(\cdot) \) is a Wiener process without drift, then there exists a sequence \( \{T_n\} \) of stopping times on \( W(\cdot) \) so that the two sequences \( \{W(T_n)\} \) and \( \{Y_n\} \) have the same law. This theorem is discussed and applied to absorption problems in Hall (1968b) (hereafter referred to as II).

In this paper we consider the analogous problem in which \( \{Y_n\} \) is a

---

1. Research supported by National Science Foundation Grant GP-5705 and U.S. Public Health Service Grant GM-10397.
submartingale (or supermartingale) sequence and \( W(\cdot) \) is replaced by a Wiener process with drift. The main result is presented in Theorem 4.1, the submartingale embedding theorem, being based on the basic construction theorem (Theorem 2.1).

Only certain submartingales can be embedded in a Wiener process with drift; we have termed them embedable submartingales. They are defined, characterized (Lemmas 3.2 and 3.3), and exemplified (Lemmas 3.7-3.9 and remarks following) in Section 3. They are related to families of r.v.'s introduced by de Finetti (1939) and described in Dubins and Savage (1965) under the name "exponential houses". Examples include (i) random walks with positive drift and moment generating functions, and (ii) log likelihood ratio sequences. Sections 5 and 7 discusses these examples, especially in relation to absorption problems and sequential analysis. These two sections are largely expository and may be read next as a continuation of this introduction.

The relevant absorption theory is derived in greater generality in Section 6. Section 5 also includes corresponding results for random walks without drift, the relevant absorption theory having been derived in \( \Pi \).

Thus, the theoretical results in this paper are contained in Sections 2, 3, 4 and 6. Some elementary results on stopped Wiener processes, used herein, appear in Hall (1968a) (hereafter referred to as \( I \)). The results of \( \Pi \) are also used, directly and by analogy.
2. Embedding $X$ in $W_\delta(\cdot)$.

The basic construction theorem is

**Theorem 2.1:** Suppose $\delta > 0$ and $W_\delta(\cdot) = \{W_\delta(t); t \geq 0, W_\delta(0) = 0\}$, a Wiener process with drift $\delta$ per unit time. There exists a stopping time $T$ on $W_\delta(\cdot)$ for which

(1) \[ W_\delta(T) \overset{d}{=} X \quad \text{(equal in law)} \]

iff $\mathbb{E}e^{-2\delta X} \leq 1$; moreover,

(2) \[ ET = \mathbb{E}X/\delta \quad (\leq \infty). \]

With $W_\delta(0) = w_0$, the theorem remains valid if $X$ is replaced by $X - w_0$; it is valid as stated if $\delta < 0$ rather than $\delta > 0$. However, we shall assume $\delta > 0$ and $W_\delta(0) = 0$ unless stated otherwise.

An alternative characterization of embedable $X$'s will be presented in Lemma 3.3 (with references to relevant history preceding it); specifically, $\mathbb{E}e^{-2\delta X} \leq 1$ for some $\delta > 0$ iff $X^-$ has an existing moment generating function (m.g.f.) and either $0 < \mathbb{E}X \leq \infty$ or $X = 0$ a.s.

**Proof of Theorem 2.1:** Necessity. For any $\delta > 0$ and any stopping time $T$, $\mathbb{E}e^{-\delta W_\delta(T)} \leq 1$ for $0 \leq \theta \leq 2\delta$ by Theorem 1 in I. Hence, (1) implies $\mathbb{E}e^{-\delta X} \leq 1$ for $0 \leq \theta \leq 2\delta$. Moreover, $\mathbb{E}W_\delta(T)$ exists (I) and $= \delta ET (\leq \infty)$ by Theorem 2 in I, so that (1) implies (2).

Sufficiency. This is proved in part as a corollary to Lemma 2 in II, modified according to Remark (iv) therein. Denote $Z = e^{-2\delta X}$, $\mathbb{E}Z \leq 1$, and $Z_\delta(t) = e^{-2\delta W_\delta(t)}$. Note that $Z_\delta(\cdot) = (Z_\delta(t); t \geq 0, Z_\delta(0) = 1)$ is a diffusion process (having continuous sample
paths and the strong Markov property) and a martingale (easily
checked). Moreover, \( \inf_t Z_\theta(t) = 0 \) a.s. so that, if \( a \) and \( b \) are
such that \( 0 < b < Z_\theta(0) < a \) and \( T \) is the hitting time by \( Z_\theta(\cdot) \)
of \( \{a, b\} \), then \( T \) is finite a.s. Therefore, as noted in Remark (iv)
of Section 4 of II, \( Z_\theta(\cdot) \) may be used as an embedding process and
we shall embed \( Z \) in it.

If \( m = 1 \), the construction of Lemma 2 of II is applicable since
\( Z_\theta(0) = EZ \); hence there exists a stopping time \( T \) on \( Z_\theta(\cdot) \) for
which \( Z_\theta(T) \downarrow Z \). Equivalently, \( W_\theta(T) \downarrow X \).

Now suppose \( m < 1 \), and first let \( Z_\theta(\cdot) \) drift from its initial
value 1 down to \( m \), requiring time \( T' \), say. (\( T' \) is finite a.s.
since \( \inf Z_\theta(t) = 0 \) a.s.) Then embed \( Z \) in \( Z'_\theta(\cdot) = \{Z'_\theta(t) = Z_\theta(t+T') ;
t \ge 0, Z'_\theta(0) = m\} \), according to Lemma 2 in II - i.e., let \( T'' \) be a
stopping time on \( Z'_\theta(\cdot) \) for which \( Z'_\theta(T'') \downarrow Z \). Let \( T = T' + T'' \);
then \( Z_\theta(T) \downarrow Z \), or \( W_\theta(T) \downarrow X \), completing the proof.

Let us describe explicitly such a construction, assuming
\( EZ = m < 1 \). We use \( T' \) defined above, followed by the 'one-stage'
randomized construction of II, but letting \( T' \) provide the
randomization. Let

\( T' = \text{hitting time (}>0\) by \( Z_\theta(\cdot) \) of \( \{m\} \),

and

\( T = \text{subsequent (}\ge T'\text{) hitting time of } \{U,V\} \)

where \( \{U,V\} \) is determined from \( T' \) and has joint d.f. \( H(\cdot,\cdot) \).
given by \(dH(u,v) \propto (u-v)dg(u)g(v)\) for \(v < m < u\), where \(g(\cdot)\) is the d.f. of \(Z = e^{-28X}\) (see II). This defines the construction in the \((t,Z_\delta)\)-plane; it is illustrated in Figure 1.

![Graphs showing \(Z_\delta(t) = e^{-28W_\delta(t)}\) and \(W_\delta(t) = -\frac{1}{28} \ln Z_\delta(t)\).]

**FIGURE 1:** Embedding \(Z\) in \(Z_\delta(\cdot)\) or \(X\) in \(W_\delta(\cdot)\) \((\delta > 0)\)

Equivalently, we could describe the construction in the \((t,W_\delta)\)-plane, by simply transforming the ordinate scale \(z\) into \(x = -(28)^{-1} \ln z\) (see Figure 1). Then \(W_\delta(\cdot)\) starts at 0, drifts up (assuming \(\delta > 0\)) to \(v = -(28)^{-1} \ln m\) in time \(T'\), and then drifts until \(U'\) or \(V'\) is reached, at time \(T\). Here, \(U' = -(28)^{-1} \ln U\) and similarly for \(V' (V' \geq w > U')\).

If \(m = 1\), then the 'two-stage' construction of II can be used in the \((t,Z_\delta)\)-plane, using the path of \(Z_\delta(\cdot)\) during the first stage to generate the necessary randomization for the second stage. (Alternatively, a 'one-stage' randomized construction can be used by generating \((U,V)\), defined above, externally.)
We now give two properties of the embedding; a third property is in Lemma 2.2 below.

**Property A.** If \( X \leq a \) a.s., then \( W_0(t) \leq a \) for all \( t \leq T \).

If \( E e^{-28X} = 1 \) and \( X \geq b \) a.s., then \( W_0(t) \geq b \) for all \( t \leq T \).

(See Remark (ii) in II.)

**Property B.** Denoting \( \mu = EX \) and \( \sigma^2 = \text{var } X \), then

\[
[s - \sqrt{\mu/5}]^2 \leq \sigma^2 \text{ var } T \leq [s + \sqrt{\mu/5}]^2
\]

so long as \( \text{var } T < \infty \) (see Lemma 2.2 below). This holds for any such \( T \) satisfying (1) by Corollary 3 of I.

**Lemma 2.2:** If (1) holds, there exists a \( T \) satisfying (1) and such that

\( E|X|^k < \infty \iff E|X|^k < \infty \) (k a positive integer); in fact, \( E|X|^k \leq c_k E|X|^k \).

To prove this, we shall need the following facts (see Cox and Miller (1965), e.g., for (3); the other assertions follow from it):

**Lemma 2.3:** If \( T \) is the hitting time by \( W_0(\cdot) \) of a level \( a (> 0) \), then

\[
(a \lambda e^{-(\lambda^2/2)})^k = a^b[1 - \sqrt{1 + 2 \lambda \delta^2 \lambda^2}] \text{ for } \lambda \geq e^{-\delta^2/2},
\]

the \( r \)th cumulant of \( T \) is \( c_r \delta^{2r+1} \) where \( c_r = 1!x1\ldots x|2r-3| \), and the \( r \)th moment of \( T \) is of the form \( ET^r = \sum_{j=1}^r c_{rj} a^j \).

**Proof of Lemma 2.2:** For any \( T \) on \( W_0(\cdot) \), \( ET^k < \infty \) implies \( E|W_0(T)|^k < \infty \) by Corollary 2 of I. We now prove the converse for \( T \) described in the proof of Theorem 2.1. We thus assume \( E|X|^k < \infty \).
First note that $T'$, defined above, has all moments finite by Lemma 2.3, so we can assume without loss of generality that $m = 1$ ($T' = 0$). Suppose first we consider a randomized $T$, defined as the hitting time by $Z_0(\cdot)$ of $\{(U, V)\}$ where $(U, V)$ has d.f. $H$ defined above.

Now let $(U', V')$ be the values corresponding to $(U, V)$ in the $(t, W(t))$-plane ($U' < 0$ ($= \nu$) $\leq V'$). Thus, $T$ is the hitting time by $W_0(\nu)$ of $\{(U', V')\}$. Let $T_0$ be the hitting time by $W_0(\cdot)$ of $V'$ so that $T \leq T_0$. Hence, conditional on $V' = a$, $E(T^k|V' = a) \leq E(T^k_0|V' = a) = E[a^k]_{c_j} a^j$ by Lemma 2.3. Since $E(T^k) = E(E(T^k|V'))$, it only remains to verify that $EV'^k < \infty$.

We now derive the distribution of $V'$. But first, the marginal d.f. of $V$ is found to be $H(\infty, v) = [\mu^+G(v) - p^+\int_{(0, v]}zdG(z)]/\mu^-p^+$ for $0 < v \leq 1$ (=m) where $p^+ = Pr[Z > 1] = Pr[X < 0]$, $\mu^+ = \int_{(1, \infty]}zdG(z) = \int_{(-\infty, 0)}e^{-2\delta x}dF(x)$, and $F$ is the d.f. of $X$.

Now, for $x \geq 0$, $Pr[V' \leq x] = Pr[V \geq e^{-2\delta x}]$

$$= 1 - H(\infty, e^{-2\delta x} - 0) = [\mu^+F(x) - p^+\int_{(-\infty, x]}e^{-2\delta y}dF(y)]/\mu^-p^+ = \lambda F^+(x)

+ (1-\lambda)K(x)$$

where $F^+(x) = Pr[X \leq x|X > 0], K(x) = \int_{[0, x]}e^{-2\delta y}dF(y)/\int_{(0, \infty)}e^{-2\delta y}dF(y)$, and $\lambda = \mu^+(1-p^+)/(\mu^-p^+)$ $\in (0, 1]$. Thus, the d.f. of $V'$ is the mixture of the two d.f.'s $F^+$ and $K$ on $[0, \infty)$. Hence, $V'$ has a finite $k\text{th}$ moment iff both the $F^+$ and $K$ distributions have finite $k\text{th}$ moments. That $F^+$ does follow from the assumption that $\int|x|^k dF < \infty$.

Now $\int_{[0, \infty)}x^k dK(x) = c \int_{[0, \infty)}x^k e^{-2\delta x}dF(x)$ (where $c^{-1}$

$= \int_{[0, \infty)}e^{-2\delta x}dF(x)) \leq c \int_{[0, \infty)}x^k dF(x) < \infty$, completing the proof.
that $E T^k < \infty$. In fact, it is readily seen that $E T^k \leq c_k E|X|^k$ for suitably chosen $c_k$ (not depending on the distribution of $X$).

If $T$ is not randomized, but is based on a two-stage construction as indicated above, analogous reasoning will yield the same conclusion.

3. Embeddable submartingales.

We now consider a sequence $\{Y_n; n = 1, 2, \ldots\}$ of r.v.'s on a probability space $(\Omega, \mathcal{F}, P)$ and an increasing sequence $\{\mathcal{F}_n\}$ of subfields of $\mathcal{F}$ such that $Y_n$ is $\mathcal{F}_n$-measurable ($\mathcal{F}_0 = \emptyset$). Denote $X_n = Y_n - Y_{n-1}$ ($n > 1$, $X_1 = Y_1$). Recall that $\{Y_n, \mathcal{F}_n\}$ is said to be a submartingale sequence [supermartingale sequence, resp.] if $E Y_n < \infty$ [$E Y_n^+ < \infty$, resp.] and $E^\theta Y_{n+1} \geq [\leq, \text{resp.}] Y_n$ a.s. for all $n$. ($E^\theta[\cdot]$ denotes conditional expectation given $\mathcal{F}_n$.) If $\{Y_n, \mathcal{F}_n\}$ is both a submartingale sequence and a supermartingale sequence, it is a martingale sequence.

Anticipating Lemma 3.1 and Theorem 4.1, we define:

**Definition**: $\{Y_n, \mathcal{F}_n; n = 1, 2, \ldots\}$ is an embeddable submartingale [supermartingale, resp.] sequence if there exists a $\theta > 0$ [$< 0$, resp.] for which, with $Z_n = e^{-\theta Y_n}$ ($n > 0; Z_0 = 1$), $\{Z_n, \mathcal{F}_n; n = 0, 1, 2, \ldots\}$ is a (positive) supermartingale sequence i.e., $E^\theta Z_{n+1} \leq Z_n$ a.s. for all $n = 0, 1, 2, \ldots$. Any such $\theta$ is an exponent of the sequence.

Note that any $\theta'$ smaller in magnitude than $\theta$ (but of the same sign) is also an exponent of the sequence since $\{Z_n^{\theta'/\theta}\}$ retains the supermartingale property of $\{Z_n\}$ by Jensen's inequality (see proof of Lemma 3.1 below).
LEMMA 3.1: An embeddable submartingale [supermartingale, resp.] sequence is a submartingale [supermartingale, resp.] sequence.

Proof: Since \( \{Z_n\} \) is a positive supermartingale, \( EZ_n < \infty \) by definition. Hence, \( Y_n^{-} \) has an m.g.f. (see Lemma 3.3 below) and therefore a finite mean. It remains to verify that \( E^n_{Y_n} Y_{n+1} \geq Y_n \) a.s.

Now \( Y_n = -\theta^{-1} \ln Z_n \), a convex decreasing function of \( Z_n \), which converts a supermartingale into a submartingale (Doob, 1953, pg. 213) — explicitly, \( E^n_{Y_{n+1}} = -\theta^{-1} E^n \ln Z_{n+1} \geq -\theta^{-1} E^n \ln Z_n \) by Jensen's inequality, and the RHS \( \geq -\theta^{-1} \ln Z_n = Y_n \) since \( \{Z_n\} \) is a supermartingale.

By changing the signs of \( \theta \) and \( Y_n \), the supermartingale assertion follows from the submartingale assertion.

Because of the strict nature of Jensen's inequality, an embeddable submartingale is not a martingale (except in degenerate cases).

Hereafter, we confine attention to the submartingale case.

Because the inequalities \( E^{n-1} e^{-\theta Y_n} \leq e^{-\theta Y_{n-1}} \) a.s. (n=1,2,...) and (4) below are equivalent, we have the following alternative characterization:

LEMMA 3.2: \( \{Y_n, \mathcal{F}_n\} \) is an embeddable submartingale sequence iff there is a positive \( \theta \) for which

(4) \[ E^{n-1} e^{-\theta X_n} \leq 1 \] a.s. for \( n = 1, 2, ... \).

In the terminology of Dubins and Savage (1965), (4) states that the conditional distributions of the \( X_n \)'s belong to a \((-\theta)\)-exponential house. The properties and relevance of 'exponential houses' are discussed in Dubins and Savage in gambling terminology. According to
them, such families of r.v.'s were first studied by de Finetti (1939), in an actuarial context.

Let $\theta_n$ be the non-zero root, if any, of $E e^{-\theta X_n} = 1$ and $\theta_n \to \infty$ whenever $\Pr_X(N+1 \geq D+1) = 1$. Hence, $\theta_n$ is $\mathcal{F}_n$-measurable.

If $\theta_n$ is bounded away from zero, say $\theta_n \geq \theta > 0$ a.s. for all $n$, then not only that is $(Y_n)$ an exchangeable submartingale.

Relevant to the interpretation of Lemma 3.2 is the following

**Lemma 3.3:** There exists $\theta > 0$ for which $E e^{-\theta X} \leq 1$ iff $X^+$ has an existing m.g.f. and either $\theta \in \Pr_X \to \infty$ or $X = 0$ a.s.

**Proof:** The identity $X = X^+ - X^-$ implies the identity $e^{sX} = e^{sX^+} + e^{sX^-}$ for all real $s$. Since $e^{sX^+}$ and $e^{sX^-}$ are bounded for $s < 0$, $E e^{sX^+} < \infty$ for $-\theta < s < 0$ iff $E e^{sX^-} < \infty$ for $-\infty < s \leq \theta$.

Now suppose $E e^{-\theta X} < 1$ for some $\theta > 0$, and $X$ not 0 a.s.

By the previous sentence, $X^+$ has an m.g.f. By the strict convexity of $E e^{sX^+}$ where finite, and the fact that $E e^{sX^+} = 1$ at $s = 0$, we have $E e^{sX^+} < 1$ for some interval $(s_0, 0)$ of $s$-values; moreover, upon differentiating, $E e^{sX^+} > E e^{sX^+} e^{s} (\text{say}) > 0$ for $-s_0 < s < 0$.

Let, after inserting the identity above and taking expectations, $E e^{sX^+} = e^{sX^-} e^{s}$ for $s_0 < s < 0$ (which may be seen to be finite - the first term on the RHS tends to $EX^+$ as $s \uparrow 0$ by monotone convergence, and the second term tends to $EX^-$. We thus have $E e^{-\theta X} = e^{sX^+} e^{s} > 0$, completing the proof of the forward implication of the Lemma.

Conversely, suppose $X^+$ has an existing m.g.f. (and $X$ not 0...
a.s.) so that, as proved above, \( E e^{-\theta X} < \infty \) for some positive \( \theta \).

Suppose \( E e^{sX} > 1 \) for all \( s < 0 \). Using the convexity again, we have
\[
0 < E e^{sX} = E^{+} e^{sX} - E^{-} e^{-sX} \quad \text{for} \quad -\theta < s < 0.
\]

Given \( \epsilon > 0 \) we can find \( s_{0} > 0 \) (and \( < \theta \)) so that \( E^{+} e^{-sX} < E^{-} + \epsilon \) for
\[
-s_{0} < s < 0.
\]

Hence, \( E^{+} e^{sX} < E^{-} + \epsilon \) for \( -s_{0} < s < 0 \) and, by
monotone convergence, we conclude that \( E^{+} \leq E^{-} + \epsilon \).

Since \( \epsilon \) is arbitrary, we conclude that \( E X = 0 \).

Hence, if \( E X > 0 \) and \( X^{-} \) has
an m.g.f., then \( E e^{sX} \leq 1 \) for some \( s < 0 \), completing the proof.

The next lemma asserts that an embedable submartingale sequence
must have a limit, and the subsequent two lemmas provide sufficient
conditions for that limit to be \( +\infty \) a.s. (Lemmas 3.5 and 3.6 are
essentially contained in Lemmas 5.1 and 5.2 of II; other sufficient
conditions (for (5)) are also given there.)

**Lemma 3.4.** If \( \{ Y_{n}, \tilde{F}_{n} \} \) is an embedable submartingale sequence with
exponent \( \theta \), then \( Y = \lim_{n \to \infty} Y_{n} \) exists a.s., \( -\infty < Y \leq +\infty \), and
\( E e^{-\theta Y} < \infty \).

**Proof:** This is an immediate consequence of Doob's submartingale
inequality [Doob, 1953, pg. 324, Theorem 4.1s(i)] applied to the
non-positive submartingale \( \{ -e^{-\theta Y_{n}}, \tilde{F}_{n} \} \).

**Lemma 3.5.** Suppose \( \{ Y_{n}, \tilde{F}_{n} \} \) is an embedable submartingale sequence
for which

\[
\limsup_{n \to \infty} \Pr^{n-1} \left[ X_{n} > \eta \right] \geq \epsilon \quad \text{a.s. for some} \quad \eta > 0, \epsilon > 0.
\]

Then \( Y_{n} \to \infty \) a.s.
Lemma 3.6: If \( X_n \leq K \) a.s. and \( \lim \sup E^{n-1}X_n \geq \epsilon > 0 \), then (5) holds.

The following three lemmas, which are corollaries to Lemmas 3.2 and 3.3, provide examples of embedable submartingale sequences.

Lemma 3.7: If \( X_n \geq 0 \) a.s. for all \( n \), then \( \{Y_n, \mathcal{F}_n\} \) is an embedable submartingale sequence (with arbitrary positive exponent).

Lemma 3.8: Suppose \( \{X_n\} \) is a sequence of i.i.d. r.v.'s for which \( 0 < EX_1(\leq \infty) \) and \( X_1 \) has an existing m.g.f. Then the random walk \( \{Y_n\} \) is an embedable submartingale sequence.

Lemma 3.9: Suppose, for each \( n = 1, 2, \ldots, \)

(i) \( X_n \geq b > -\infty \) a.s.,
(ii) \( \text{var}^{n-1}X_n \leq c^2 < \infty \) a.s., and
(iii) \( E^{n-1}X_n \geq \epsilon > 0 \) a.s.

Then \( \{Y_n, \mathcal{F}_n\} \) is an embedable submartingale sequence. If, for each \( n \), \( |X_n| \leq K \) a.s., then (i) and (ii) hold.

Proof: (i) implies \( E^{n-1}e^{-\theta X_n} < \infty \) a.s. for \( \theta > 0 \). Dropping all sub- and superscripts relating to \( n \), and denoting \( \mu = EX \) and \( \sigma^2 = \text{var} X \), we have, for \( \theta > 0 \),

\[
e^{-\theta \mu} = e^{-\theta \mu} \left(1 + \frac{1}{2} \theta^2 e^{-\theta (X-\mu)} \right) \text{ (for some r.v. } \theta \text{ in } (0, \theta)) \leq e^{-\theta \mu} + \frac{1}{2} \theta^2 e^{-b\theta} \text{ (by (i))} \leq e^{-\theta \mu} + \frac{1}{2} \theta^2 c^2 e^{-b\theta} \text{ (by (ii)}
\]

and (iii)) \( \leq 1 - \theta c + \theta f(\theta) \) where \( 2f(\theta) = \theta c^2 + \theta c^2 e^{-b\theta} \). Since \( f(\theta) \) increases continuously from \( 0 \) to \( \infty \), there is a \( \theta_0 > 0 \) (and not depending on \( n \)) for which \( f(\theta_0) = \epsilon \) and hence (4) holds for \( \epsilon > (0, \theta_0) \).

A simple example satisfying the conditions of Lemma 3.9 is the...
following: Each increment $X_n$ assumes one of two values, say $a$ and $b$ ($a > b$), and either $b \geq 0$ or, for some $\epsilon > 0$,

$$\Pr^n[X_n = a] \geq -b(a-b)^{-1} + c \text{ a.s. for each } n.$$

As a final example, to be made explicit in Section 7, it is well-known that a sequence of likelihood ratios form a martingale (the correct hypothesis being represented in the denominator and the distributions being mutually absolutely continuous), so that a sequence of log likelihood ratios is an embedable supermartingale sequence with exponent $\theta = -1$.

4. The submartingale embedding theorem.

The main theorem is

**Theorem 4.1:** Suppose $\delta > 0$ [\( \leq 0 \), resp.] and $W_\delta(\cdot) = \{W_\delta(t); t \geq 0,$ $W_\delta(0) = 0\}$. There exists a sequence $\{T_n\}$ of successive stopping times on $W_\delta(\cdot)$ for which

$$\{W_\delta[T_n]; n = 1, 2, \ldots\} \mathcal{F} \subseteq \{Y_n; n = 1, 2, \ldots\}$$

iff $\{Y_n, \mathcal{F}_n\}$ is an embedable submartingale [supermartingale, resp.] sequence for some $\{\mathcal{F}_n\}$, with exponent $2\delta$; moreover,

$$E T_n = E Y_n / \delta \quad (\leq \infty).$$

Proof: The necessity proof and the proof that (7) follows from (6) are almost identical to those of Theorem 2.1 and are therefore omitted. (See also the corresponding proof in II.)

The sufficiency proof is based on Theorem 2.1 and is analogous to the corresponding proof in II. The sequence $\{T_n\}$ is constructed
recursively by applying Theorem 2.1 to the conditional distribution of $X_n$ given $J_n$ with the origin in the $(t, W_0)$-plane shifted to $(T_{n-1}, W_0(T_{n-1}))$. (The 'randomization' required at the $n^{th}$ stage can be achieved as indicated in Section 2, or based on independent characteristics of the path of $W_0(\cdot)$ during earlier stages.)

We now give four properties of the embedding. They refer specifically to the construction of $\{T_n\}$ given above, but some of them would carry over to any other construction.

**Property 1.** Property A of Section 2 implies that

the variation in $W_0(\cdot)$ ($\delta > 0$) is bounded above between two embedding times if the corresponding increment in $\{Y_n\}$ is; if $\{e^{-2\delta} Y_n, T_n\}$ is a martingale sequence (rather than just a supermartingale sequence) and if an increment in $\{Y_n\}$ is bounded below, then so is the variation in $W_0(\cdot)$ between the corresponding embedding times.

**Property 2.** Property B of Section 2 likewise provides bounds on the variance of $T_n$; also, applying Lemma 2.2, the $T_n$'s may be so chosen that $E T_n^m < \infty$ iff $E |Y_n|^m < \infty$.

**Property 3.** If $\{Y_n\}$ has independent increments then $\{T_n\}$ may be so chosen to have independent increments.

**Property 4.** If $\lim \sup E^{-1} X_n < \epsilon > 0$ a.s., then $T_n \to \infty$ a.s.

For, the assumption is sufficient to prove that the distributions of the $T_n$'s are not all concentrated heavily at the origin.

We close this section with three remarks on variations in the embedding theorem. First, the sequence $\{Y_n\}$ with exponent $\delta$ can be embedded in $W_{\mu, \sigma}(\cdot)$, a Wiener process with drift $\mu$ and variance $\sigma^2$ per unit time, so long as $2|\mu| \sigma^2 \leq |\delta|$ and $\text{sgn} \mu = \text{sgn} \theta$; then (7) holds with $\delta$ replaced by $\mu$. This follows from the fact that $\{W_{\mu, \sigma}(t)\} \overset{d}{=} \{W_{\mu, \sigma_0}(\sigma^2 t)\}$.

The theorem also applies to extended-real-valued r.v.'s $\{Y_n\}$;
it is only necessary to assume \( Y_n > -\infty \) \([< +\infty, \text{ resp.}]\) and to assume that, if \( Y_n = \infty \) then \( Y_m = \infty \) for all \( m > n \). Then \( T_n \) need no longer be finite, but we interpret \( W_\delta(t) = \infty \) for \( t \geq \infty \).

The final remark concerns the embedding of an arbitrary positive supermartingale \( \{Z_n\} \) in \( Z_\delta(\cdot) = ce^{-2\delta W_\delta(\cdot)} \). The approach of the embedding theorem was to embed \( \{Y_n\} \) in \( W_\delta(\cdot) \) by embedding \( \{e^{-2\delta Y_n}\} \), assumed to be a positive supermartingale, in the positive martingale \( Z_\delta(\cdot) \) (with the scale factor \( c = 1 \)). The same approach works for any positive supermartingale, so long as \( c \geq EZ_1 \). We state this formally as

**Corollary 4.2:** Suppose \( \delta > 0 \), \( W_\delta(\cdot) \) as in Theorem 4.1, and \( Z_\delta(\cdot) = ce^{-2\delta W_\delta(\cdot)} \). Suppose \( \{Z_n, J_n; n = 0, 1, \ldots, Z_0 = c\} \) is a positive supermartingale sequence. Then there exists a sequence \( \{T_n\} \) of successive stopping times on \( Z_\delta(\cdot) \) for which \( \{Z_n\} \subseteq \{Z_\delta(T_n)\} \)

\((T_0 = 0); \) moreover, \( E_{\delta} = (\ln - E\ln Z_0)/(2\delta^2) \leq \infty \).

(The proof is immediate upon considering the embeddable submartingale \( \{Y_n\} = \{-(2\delta)^{-1}\ln(Z_n/c)\} \). Specifically, in gambling terminology, \( Z_n \) may represent the capital of a gambler who never risks his total capital after \( n \) unfair games, starting with initial capital \( Z_0 = c \). Choosing \( c = 1 \) and \( \delta = 1/2 \), the Corollary equivalently states that \( \{\ln Z_n\} \subseteq \{W_{-1/2}(T_n)\} \), so that the gambler's capital, in logarithmic units, behaves like a Wiener process with negative drift \((-1/2)\), at certain random time points \( \{T_n\} \). The usefulness of this representation has not been explored.
5. Application to random walks.

We first restate a slight variation of the two embedding theorems for the case of random walks; we have included both the Skorokhod (martingale) embedding theorem (II) and the submartingale embedding theorem (Theorem 4.1 above).

**THEOREM 5.1:** Suppose \( \{X_n\} \) is a sequence of i.i.d. r.v.'s, each distributed as \( X \) (not \( 0 \) a.s.), and \( Y_n = X_1 + \ldots + X_n \). Let \( W_{\mu,\sigma}(\cdot) \) be a Wiener process with drift \( \mu(\geq, =, \text{or } < 0) \) per unit time, variance \( \sigma^2 \) per unit time, and initial value \( 0 \). There exists a sequence of successive stopping times \( \{T_n\} \) on \( W_{\mu,\sigma}(\cdot) \) for which

\[
(8) \quad \{W_{\mu,\sigma}(T_n)\} \overset{d}{=} \{Y_n\} \quad \text{and} \quad \mathbb{E} T_n = n \quad \text{for all} \quad n
\]

iff \( \mathbb{E} X \) exists and \( \mu = \mu \), and

\[
(9) \quad \sigma^2 = \text{var} \ X < \infty \quad \text{if} \quad \mu = 0
\]

and

\[
(10) \quad \mathbb{E} e^{-2\mu X/\sigma^2} \leq 1 \quad \text{if} \quad \mu \neq 0.
\]

Also, the \( T_n \)'s may be so chosen to have i.i.d. increments (Property 3), and to have finite \( m \)th moments iff \( |X| \) has a finite \( (2m) \)th moment (if \( \mu = 0 \)) or \( m \)th moment (if \( \mu \neq 0 \) (Property 2)); if \( \mu \geq 0 \) and \( X \leq a \) a.s., then \( W_{\mu,\sigma}(t) - W_{\mu,\sigma}(T_n) \leq a \) for \( T_n \leq t \leq T_{n+1} \); and if \( \mu = 0 \), or \( \mu > 0 \) and equality holds in (10), and \( X \geq b \) a.s., then \( W_{\mu,\sigma}(t) - W_{\mu,\sigma}(T_n) \geq b \) for \( T_n \leq t \leq T_{n+1} \) (Property 1).
Also, if a Wiener process with unit variance is used, (8) would be replaced by \( W_{\theta}(T_n) \) \( \text{d} \) \( Y_n \) with \( \sigma = \mu / \sigma^2 \) and \( \mathbb{E} T_n = n \sigma^2 \).

It will facilitate exposition to define the **variability index** \( \sigma^2 \) of \( X \) by \( \sigma^2 = \text{var} X \) if \( \mathbb{E} X = \mu = 0 \), and \( \sigma^2 \) is such that \( \mathbb{E} e^{2 \mu X / \sigma^2} = 1 \) if \( \mathbb{E} X = \mu \neq 0 \), assuming such a \( \sigma^2 \) exists. Thus, \( 2 \mu / \sigma^2 \) is the non-zero root, if any, of \( \psi(\theta) = \ln \mathbb{E} e^{-\theta X} = 0 \). Because of the convexity of the m.g.f. in its interval of existence, \( \sigma^2 \) will be large or small according as the curvature of \( \psi(.) \) at the origin—the variance—is large or small. As noted by de Finetti (see Dubins and Savage, 1965, pg. 166), \( 2 \mu / \sigma^2 = (2 \mu / \sigma_o^2)[1 + O(\mu / \sigma_o)] \) where \( \sigma_o^2 = \text{var} X - \text{i.e.,} \), \( \sigma^2 \sim \sigma_o^2 \) as \( \mu / \sigma_o \to 0 \). (Dubins and Savage call \(-2 \mu / \sigma_o^2\) the 'inequity' and discuss its approximation to \(-2 \mu / \sigma^2\).) It is readily verified that, for normally distributed \( X \), \( \sigma^2 = \sigma_o^2 \) whether or not \( \mathbb{E} X = 0 \). Also, the variability index is seen to be additive for i.i.d. r.v.'s.

With this definition (and assuming the existence of \( \sigma^2 \)), the **random walk embedding theorem** (Theorem 5.1) has the following interpretation: The sequence \( \{Y_n\} \) may be approximated in continuous time by \( W_{\mu, \sigma}(. \) in such a way that, at random successive time points \( \{T_n\} \), with means \( \{n\} \), \( W_{\mu, \sigma}(. \) coincides exactly (in law) with \( \{Y_n\} \); while at the fixed time points \( \{n\} \), \( W_{\mu, \sigma}(. \) and \( \{Y_n\} \) have the same means \( \{n\mu\} \) and the same variability indices \( \{n\sigma^2\} \).

A more conventional approximation of a random walk \( \{Y_n\} \) by a Wiener process would be that obtained by matching means and variances at the fixed time points \( \{n\} \), that is, using \( W_{\mu, \sigma_o}(. \) with \( \mu = \mathbb{E} X \) and \( \sigma_o^2 = \text{var} X \). (This is the same approximation as above.
when \( \mu = 0 \), but with additional interpretation.) This approximation has been used, for example, by Armitage (1957, Sec. 3) in sequential analysis. Cox and Miller (1965, Sec. 5.2) use a limiting form of it in introducing the Wiener process as an approximation to the simple random walk. The relative merits of the two approximations \( (\mu \neq 0) \) may well depend on the context.

Let us look briefly at the simple asymmetric random walk, in which \( X = 1 \) with probability \( p \) \((1/2 < p < 1)\) and \( X = -1 \) with probability \( q = 1-p \). Then \( \mu = p-q, \sigma_0^2 = 4pq, \) and

\[
\sigma^2 = 2(\frac{p-q}{ln(p) - ln(q)} \quad (since \quad Ee^{-Xln(p/q)} = 1). \]

Note that \( \sigma_0^2 < \sigma^2 \) (in this example, but not generally); in fact, with \( p = (1+\epsilon)/2 \), we find \( \mu = \epsilon, \sigma_0^2 = 1-\epsilon^2, \) and \( \sigma^2 = 1-3^{-1}\epsilon^2-\frac{1}{45}\epsilon^4+O(\epsilon^6) \). Conventionally, one would approximate the simple random walk by a Wiener process with drift \( p-q \) and variance \( 4pq \) per unit time, whereas the embedding theorem suggests using a process with variance parameter equal to \( \sigma^2 \) given above. We plan some numerical comparisons of these two approximations in a sequential analysis setting.

The random walk embedding theorem may be applied to the study of random walks with absorbing barriers, and a number of the classical theorems thereby derived. Thus, if we are interested in evaluating the probability that \( \{Y_n\} \) reaches (or exceeds) a fixed positive level \( a \) before it reaches a fixed negative level \( b \), we might approximate it by the continuous-time absorption probability for the corresponding Wiener process with absorbing states \( \{a\} \) and \( \{b\} \). If the increments in the random walk are bounded, then the change in
the Wiener process is likewise bounded, at least below (assuming positive drift), between successive embedding times \( \{T_n\} \), at which the two processes coincide; so the approximation may be expected to be reasonable. It is important that the absorbing states are not time-dependent, however, for otherwise the distortion of the time axis in the embedding—replacing \( \{n\} \) by \( \{T_n\} \)—may destroy any validity of approximation. We shall pursue this absorption problem in a more general setting in the next two sections. The results there provide bounds on absorption probabilities and expected time of absorption, and asymptotic formulas for these and the distribution of absorption time when \(|X| \leq K \downarrow 0\).

Specifically, if \( \mu > 0 \) and the variability index \( \sigma^2 \) exists (it is sufficient that \( X \) is not of constant sign since \(|X| \leq K\)), (i)-(iv) of Theorem 6.7 hold with \( \delta = \mu/\sigma^2 \); according to Theorem 5 of II, they remain valid when \( \delta = \mu = 0 \) upon replacing \( \mu \) by \( \sigma^2 \).

Finally, it should be noted that throughout this section the i.i.d. assumption was not fully required. It is sufficient to assume (i) independent increments, (ii) common means \( EX_n = \mu \), and (iii) common variability indices, or \( Ee^{-2\mu X_n/\sigma^2} \leq 1 \) for some \( \sigma \) and all \( n \) (when \( \mu \neq 0 \)); in fact, (i) may be dropped if \( E \) is replaced by \( E^{n-1} \) (conditional expectation).


In this section we will be concerned with an embedable submartingale \( \{Y_n, \mathcal{F}_n\} \) with one or two absorbing "barriers" \( (-\infty, b] \) and \( [a, \infty) \) \((-\infty < b < 0 < a < \infty)\). We shall apply the submartingale embedding theorem, analogously to application of Skorokhod's (martingale)
embedding theorem in Section 5 of III, to this problem. We would expect behavior similar to that of \( W_6(\cdot) \) with one or two absorbing barriers; this behavior is well-known (see Cox and Miller (1965), for example).

We first state a basic property of the Wiener process with absorbing barriers:

**Lemma 6.1**: \( \Pr[W_6(\cdot) \text{ hits } a \text{ before } b, \text{ starting from } c] = (e^{-28b} \cdot e^{-28c}) / (e^{-28b} \cdot e^{-28a}) \) for \( b \neq 0, -\infty \leq b < c < a \leq \infty \).

This formula is well-known. A simple proof is as follows: The LHS = \( \Pr[e^{-28W_6(\cdot)} \text{ hits } e^{-28a} \text{ before } e^{-28b}, \text{ starting from } e^{-28c}] = \Pr[e^{-28Y_0(\cdot)} \text{ hits } A \text{ before } B, \text{ starting from } C] \) (by a change of notation) = \( (B-C)/(B-A) \) by an elementary martingale formula (see (2) in III).

Our first application of the submartingale embedding theorem gives a bound on the probability of absorption of \( \{Y_n\} \) when there is but one barrier \((-\infty, b]\); it is a known result (e.g., Dubins and Savage, 1965, p. 164). Of course, it also gives a bound on the probability of absorption at \((-\infty, b]\) in the two-barrier problem.

**Lemma 6.2**: Suppose \( \{Y_n; \mathcal{F}_n\} \) is embedable in \( W_6(\cdot) (\delta > 0) \). Then \( \Pr[\inf_n Y_n = b] \leq e^{-28b} E e^{-28Y_1} \leq e^{28b} (b < 0) \).

**Proof**: We embed \( \{Y_n\} \) in \( \{W_6(t); t \geq 0, W_6(0) = c\} \) where \( c \) satisfies \( e^{-28c} = E e^{-28Y_1} \). Then \( \Pr[\inf_n Y_n \leq b] = \Pr[\inf_n W_6(T_n) \leq b] \leq \Pr[\inf_t W_6(t) \leq b] = e^{28(b-c)} \) by Lemma 6.1 (with \( a = \infty \)). Since \( E e^{28Y_1} \leq 1 \) by assumption, the proof is complete.
An alternative proof, without use of embedding, is as follows:

\[ \Pr[\inf Y_n \leq b] = \Pr[\sup Z_n \geq \lambda] \quad \text{(where } Z_n = e^{-2\delta Y_n} \text{ and } \lambda = e^{-2\delta b}) \]

\[ \leq \lambda^{-1} E Z_1 \] by the basic inequality for positive supermartingales (Doob, 1953, pg. 314, (3.4')) with his \( x_n = -Z_n \) and assuming \( x_n \leq 0 \).

Many of the remaining results require that the sequence \( \{Y_n\} \) have uniformly bounded increments (bound \( K \)). To obtain asymptotic results, we actually consider a family of sequences with parameter \( K \), all being embedable in \( W_\delta(*) \) (\( \delta \) fixed), and let \( K \downarrow 0 \). The barrier boundaries \( a \) and \( b \) are also kept fixed. (Actually, we could add \( o(1) \) to \( \delta \), \( a \) and \( b \).)

The following notation will be used hereafter:

\[ N(a) = \min\{n | Y_n \geq a\} \leq \infty \quad \text{for } a > 0 \quad (< \infty) \]

\[ N(b) = \min\{n | Y_n \leq b\} \leq \infty \quad \text{for } b < 0 \quad (\geq -\infty) \]

\[ N = N(a,b) = \min[N(a), N(b)] = \min\{n | Y_n \geq a \text{ or } \leq b\} \leq \infty \]

\[ \rho = \rho_\delta(a,b) = \begin{cases} (1-e^{-2\delta a})/(e^{-2\delta b} - e^{-2\delta a}) & \text{if } \delta > 0 \\ a/(a-b) & \text{if } \delta = 0 \\ a-(a-b)p & \text{if } \delta > 0 \text{ and } b \text{ finite} \end{cases} \]

\[ \nu = \nu_\delta(a,b) = \begin{cases} a & \text{if } \delta > 0 \text{ and } b = -\infty \\ -ab & \text{if } \delta > 0 \end{cases} \]

\[ \pi = \Pr(N = \infty) \]

\[ \tau_\delta(\alpha,\beta) = \min\{t | W_\delta(t) \geq \alpha \text{ or } \leq \beta\} \quad (\beta < 0 < \alpha). \]
Thus, 1 - \pi is the probability of absorption. Lemmas 3.5 and 3.6 provide sufficient conditions for \pi to be zero (\delta > 0). We have included some definitions for the case \delta = 0 since all statements to follow have analogs for that case, already derived in II.

**Lemma 6.3**: Suppose \{Y_n, F_n\} is embedable in \(W_\delta(\cdot)(\delta > 0)\) and \(X_n \leq X\) a.s. for all \(n\). Then

\[
\Pr[N(b) < N(a)] < p + \Theta^+(K) \quad \text{as} \quad K \downarrow 0
\]

where \(\Theta^+(K) \geq 0\) and \(\Theta = O(K)\). If \(\{e^{-2S_n}, F_n\}\) is a martingale sequence and \(\left|X_n\right| < X\) a.s. for all \(n\), then

\[
\langle 2 \rangle \quad p - \pi - \Theta^+(K) < \Pr[N(b) < N(a)] < p + \Theta^+(K).
\]

If \(X < \infty\) a.s., we then have that the probability of absorption at \((-\infty, b)\) before \([a, \infty)\) is \(p + \Theta^+(K)\). Explicit bounds appear in the proof. If \(p = p_\delta(a, b)\) is small and \(p_\delta(-b, a)\) is also small, then \(\Theta^+(K)\) is approximately 28Xp.

**Proof**: We embed \(\{Y_n\}\) in \(W_\delta(\cdot)\) with embedding times \(\{T_n\}\). Then

\[
P = \Pr[X < a] \leq \Pr[N < \infty \text{ and } Y_n \leq b] = \Pr[N < \infty \text{ and } W_n \leq b] = \Pr[N < \infty \text{ and } W_{T_n} \leq b] = \Pr[W_{T^*_n} \leq b] = \Pr[W_\delta(T^*_n - a + K, b)] = \Pr[W_\delta(a + K, b)].
\]

The inequality is based on Property 1 of the embedding. By Lemma 6.1, the RHS is

\[
(1 - e^{-2S(a + K)}/(e^{-2S(b - K)} - e^{-2S(a - K)})\quad \text{which is readily seen to be} \quad < p + \Theta^+(K).
\]

Likewise, under the stronger assumptions whereby \(W_\delta(\cdot)\) is also bounded below between the embedding times (Property 1), we have

\[
Q = \Pr[N(b) > N(a)] \leq \Pr[W_\delta(T_\delta(a, b, K)) = a] = (e^{-2S(b - K)} - 1)/(e^{-2S(b - K)} - e^{-2S(a - K)}) = 1 - p + \Theta^+(K). \quad \text{Finally,} \quad Q = 1 - P - \pi.
\]
We now turn to the average absorption time for the one- and two-barrier problems.

**Lemma 6.4:** If \( \{ \zeta_n, \mathcal{F}_n \} \) is embeddable in \( W_\delta(\cdot) (\delta > 0) \) and

\[
\frac{1}{n} \sum_{m=1}^{n} E^{m-1} X_m \leq \mu_U \ a.s. \text{ for all } n,
\]

then

\[
\mu_{\overline{U}} \leq \nu_\delta(a,b).
\]

If \( \{ e^{-2\delta Y_n}, \mathcal{F}_n \} \) is a martingale sequence with \( |X_n| \leq K \) a.s. for all \( n \), and

\[
\frac{1}{n} \sum_{m=1}^{n} E^{m-1} X_m \geq \mu_L \ a.s. \text{ for all } n,
\]

then

\[
\mu_{\underline{U}} \leq \nu_\delta(a+K,b-K) = \nu_\delta(a,b) + O(K).
\]

The Lemma may be proved in the same way as Lemma 5.4 in II. One uses the martingale \( \{ \gamma_n = T_n - M_n/\delta \} \) where

\[
M_n = \sum_{m=1}^{n} E^{m-1} [W_\delta(T_m) - W_\delta(T_{m-1})]
\]

and

\[
\sum_{m=1}^{n} E^{m-1} X_m.
\]

Also see Corollary 7.3 below.

We now consider the asymptotic distribution of \( N \). The methods of II (essentially Skorokhod's) apparently do not generalize to the submartingale embedding case. However, Skorokhod's method can be applied directly, using martingale embedding, if we impose enough conditions. To this end, we first state a modification of the theorem in Section 3 of Skorokhod (1961):

**Theorem 6.5 (Skorokhod):** Suppose, for each \( n > 1 \), \( \{ Y_{kn}; k=1,2,\ldots,n \} \)

is a zero-mean martingale sequence with increments \( \{ X_{kn} \} \) for which

\[
\var k^{-1} X_{kn} = 1/n a.s. \text{ and } |X_{kn}| \leq Cn^{-1/2} a.s.
\]

Given \( \beta < 0 < \alpha \) and \( \lambda \), let \( Q_n = P(\beta < Y_{kn} + \lambda k/n < \alpha \text{ for all } k \leq n) \) and

\[
Q = P(\beta < Y_k(s) < \alpha \text{ for all } 0 \leq s \leq 1).
\]

Then there exists a
universal constant \( \lambda = \mathbb{L}(C, \alpha, \beta, \lambda) \) for which \(|Q_n - Q| \leq \ln^{-1/2} \ln n \) for each \( n > 1 \).

We obtain as a corollary

**THEOREM 6.6:** Suppose for each positive \( \sigma \), \( \{Y_n(\sigma)\} \) is a sequence of r.v.'s with increments \( \{X_n(\sigma)\} \) for which, for some positive \( \delta \) and \( \lambda \), and each \( n \),

(i) \( |X_n(\sigma)| \leq K \ a.s. \),

(ii) \( E^n \frac{1}{n} X_n(\sigma) = 5c^2 \ a.s., \) and

(iii) \( \text{var} \frac{1}{n} X_n(\sigma) = c^2 \ a.s. \).

Let \( N_\sigma = N(a, b) \) and \( T_\delta = T_\delta(a, b) \), defined above. Given

\(-\infty < b < 0, a < +\infty \) (and \( a, b, \delta \) such that \( T_\delta < \infty \) a.s.),

\( c^2 N_\sigma \to T_\delta \ a.s. \ as \ \sigma \to 0. \)

**Proof:** Given a fixed \( t > 0 \), let \( \sigma \) be so chosen that \( n, = t\sigma^{-2} \),

is integral. Let \( \{Y_{kn}; k \leq n\} \) be the sequence with increments defined

by \( X_{kn} = t^{-1/2}(k - \delta t^2) \), and note that this sequence satisfies the assumptions of Theorem 6.5.

Now \( Q_\sigma(t) = \mathbb{P}(\sqrt{2}N > t) = \mathbb{P}(b < Y_k < a \ for \ all \ k \leq n) = \)

\( \mathbb{P}(\beta < \chi_{kn} < \alpha \ for \ all \ k \leq n), \) where \( \beta = t^{-1/2}b, \alpha = t^{-1/2}a, \)

and \( \delta = t^{1/2} \). Also, \( Q(t) = \mathbb{P}(\beta < \chi(s) < \alpha \ for \ all \ 0 \leq s \leq 1) = \)

\( \mathbb{P}(b < \chi(s) < a \ for \ all \ s \leq t), \) since \( \{\sqrt{tW}(s/t); s \geq 0\} \sim \chi(\cdot). \)

Theorem 6.5 thus implies \( |Q_\sigma(t) - Q(t)| \leq 2c(-\ln t)n^{-1/2}L \) where \( L \)

depends only on \( t^{-1/2}C \) and \( \alpha, \beta, \lambda \) defined above. The theorem follows.

24
We now summarize some of the results of this section in

**THEOREM 6.7:** Suppose \( \{e^{-2bY_n}, \mathcal{F}_n\} \) is a martingale sequence \((b > 0)\)

with \(|X_n| \leq K \) a.s. for all \( n \), \( E^{n-1}X_n = \mu_n \) (constant) a.s. for all \( n \),
and \( n^{-1} \sum_{m=1}^{n} \mu_m \to \mu(> 0) \) as \( n \to \infty \). Then

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

(ii) \( \Pr[Y_n \leq b] \to p \) as \( k \to 0 \),

(iii) \( \mu EN \to v \) as \( k \to 0 \),

if, in addition, \( \mu_n = \mu \) for all \( n \), then \( \mu EN \geq v \) and

(iv) \( \mu N \xrightarrow{p} T_{\theta}(a,b) \).

This theorem is analogous to Theorem 5 in II. The one-barrier case of

(iii) \((b = -\infty)\) and then it is sufficient that \( X_n \leq K \) a.s.\) was

proved by Chow and Robbins (1963) under more general conditions. An

alternative proof (omitted) would parallel that in II. (Actually, \( \mu_n \)

need not be constant for (i)-(iii) so long as it converges in C\'esaro-

mean uniformly a.s.) Part (iv) follows from Theorem 6.6 since

assumption (iii) of that theorem is virtually equivalent to the

assumption that \( \{e^{-2bY_n}\} \) is a martingale sequence, with assumptions

(i) and (ii).

7. **Applications to sequential analysis.**

Classical sequential analysis, as developed by Wald (1947), is

concerned with a sequence of log likelihood ratios with two absorbing

barriers. Whenever such a sequence is an embedable submartingale
(or supermartingale, or martingale), the absorption theory of Section 6 (and II) will be applicable. We shall present a number of such results below, being immediate corollaries of results already presented. Included are some of the basic results of Wald, and some minor, even curious, variations of such results. This approach to sequential analysis through embedding theory provides new insight into basic properties and approximations. It has also made possible the development of some new results in nonparametric sequential analysis, to be presented in a subsequent paper.

The Wiener process has been used in several ways in sequential analysis. Dvoretzky, Kiefer and Wolfowitz (1953) presented sequential tests for hypotheses about the drift of a Wiener process. These provide a 'no overshoot' version of Wald's sequential probability ratio test (SPRT) of hypotheses about the mean of a normal distribution. This use of a continuous-time approximation to a discrete-time normal process has been used to advantage elsewhere in sequential analysis—in particular, by Armitage (1957) and by Anderson (1960). These have broader practical import, at least to the extent that many random variables of interest may be approximately normal. However, we suggest a much broader applicability of Wiener process approximations, by approximating an arbitrary log likelihood ratio sequence by a Wiener process. (As noted earlier, Armitage (1957) did use a Wiener process with drift to approximate a simple random walk, arising as a log likelihood ratio sequence, with intuitive and empirical justification. But in light of the remarks in Section 5 above, it is not clear that he used the best choice of parameters.)
We suppose \( \{Z_n\} \) is a sequence of random vectors (perhaps of varying dimensions) on a measure space \((\Omega, \mathcal{F})\), \(Z_n\) representing the data to be collected at stage \( n \) of the experiment (if performed). Let \( \mathcal{F}_n \) represent the subfield of \( \mathcal{F} \) generated by \( Z(n) = (Z_1, \ldots, Z_n) \) \( (\mathcal{F}_0 = \mathcal{F}) \). Let \( P_0 \) and \( P_1 \) represent two alternative probability measures on \((\Omega, \mathcal{F})\) and let \( E_1 \) represent expectation w.r.t. \( P_1 \). Let \( H_1 \) represent the hypothesis that \( P_1 \) is the correct measure. We consider testing \( H_0 \) vs. \( H_1 \) sequentially.

Let \( p_{ln} \) denote the density of \( Z(n) \) under \( H_1 \) (w.r.t. a mutually dominating measure), and let \( R_n \) denote the likelihood ratio \( R_n = \frac{P_{ln}}{P_{On}} \) (well-defined unless \( P_{ln} = P_{On} = 0 \)). We define the log likelihood ratio \( L_n = \ln R_n \), with increments \( X_n = L_n - L_{n-1} \) \( (L_0 = 0) \), which is well-defined, though possibly infinite, at least under \( H_0 \) and \( H_1 \).

Now \( \{R_n\} \) is readily seen to be a supermartingale when \( H_0 \) is true (Doob, 1953). For, in an obvious notation,

\[
E_{0}^{n-1} R_n = E_{0}^{n-1} \frac{P_{ln}(Z_n|Z_{n-1})}{P_{On}(Z_n|Z_{n-1})} \frac{P_{ln,n-1}}{P_{On,n-1}} \frac{P_{On}(Z_n|Z_{n-1})}{P_{On,n-1}} = \]

\[
\int_{\{P_{On} > 0\}} \frac{P_{On}(z_n|Z_{n-1})dP_{On}(z_n)}{P_{On,n-1}} \leq \frac{P_{ln,n-1}}{P_{On,n-1}} R_{n-1}.
\]

If \( P < P_0 \), the equality would hold above. Since, with \( \theta = -1 \), \( e^{-\theta L_n} = R_n \), we conclude that \( \{L_n, \mathcal{F}\} \) is an embedable supermartingale with exponent \(-1 \) \( (\theta = -1/2) \) under \( H_0 \). (Note that \(-\infty \leq L_n < \infty \) a.s. under \( H_0 \), with strict inequality if \( P_0 < P_1 \); see pentultimate remark in Section 4.) We thus conclude (see Theorem 4.1):

**Lemma 7.1:** There exist successive stopping times \( \{T_n\} \) on \( W_0(\cdot) \),
with \( \delta = -1/2 \) [or \( 1/2 \), resp.] for which \( [L_n] \overset{d}{=} (W_0(T_n)) \) under \( H_0 \) \([H_1, \text{ resp.}]. \) If \( P_1 < P_0 \) \([P_0 < P_1, \text{ resp.}], \) then \((e^{-2\delta L_n}, \mathcal{J}_n)\) is a martingale sequence.

We shall also use Properties 1 and 3 (Section 4) of the embedding.

Wald's SPRT terminates the experiment at stage \( n \) if \( L_n \leq b \) (in favor of \( H_0 \)) or \( L_n \geq a \) (in favor of \( H_1 \)) while \( b < L_m < a \) for all \( m < n. \) Wald proposed using \( a = \ln A \) and \( b = \ln B, \) with \( a = (1-\beta)/\alpha \) and \( B = \beta/(1-\alpha) \) where \((\alpha, \beta)\) are the desired (nominal) error probabilities. We denote \( P_1(H_j) = P_1(\text{SPRT decides in favor of } H_j), \)
\( \alpha' = P_0(H_1), \beta' = P_1(H_0); \alpha' \) and \( \beta' \) are the true error probabilities.

\[ \begin{align*}
W_{-1/2}(t) & \quad \text{accept } H_1 \\
& \quad \text{KEY: } W_0(t) \text{ vs. } t \\
& \quad W_0(T_n) \text{ vs. } n \quad (\mathcal{J}_n \text{ vs. } n) \\
& \quad b = \ln[\beta/(1-\alpha)] \\
& \quad \text{accept } H_0
\end{align*} \]

FIGURE 2: SPRT process embedded in a Wiener process (\( H_0 \) true)
The relevance of embedding theory is obvious from Figure 2 and Lemma 7.1. Instead of watching the discrete-time process \( \{L_n\} \), we can equivalently — for probability calculations under \( H_0 \) — watch a corresponding continuous-time Wiener process. As an approximation to the SPRT, we can terminate in favor of \( H_0 \) or \( H_1 \) according as \( W_5(\cdot) \) first crosses the levels \( b \) and \( a \), respectively. This process will terminate, at time \( T \), say, not later than \( T_N' \), where \( N' \) is the termination time for the discrete-time problem. It will be seen that Wald's approximation formulas can be derived from this analogy.

With the additional assumption that the increments in \( \{L_n\} \) are bounded (by \( K \)), another approximation can be obtained in continuous time by replacing one or both of the original boundaries by the wider boundaries \( a + K \) and \( b - K \), and the crossing time of one of these boundaries by \( W_5(\cdot) \) will be \( \geq T_N' \) (if mutually absolutely continuous). In this way additional bounds and asymptotic formulas (as \( K \to 0 \)) for the error probability and the average sample size (ASN) may be obtained.

Errors in the approximation may arise from one of two sources: one being the 'overshoot', \( W_5(T_N') - W_5(T) \), and the other being the possibility that \( W_5(\cdot) \) may cross \( a \) or \( b \) twice (or \( 2k \) times) between two earlier \( (n < N') \) embedding times \( T_{n-1} \) and \( T_n \). Both of these sources can be controlled in the case of bounded increments, however.

First, as a corollary to Lemma 6.2 (for extended-real-valued submartingales), we have
**COROLLARY 7.2:** \( \alpha' \leq \alpha/(1-\beta) \) and \( \beta' \leq \beta/(1-\alpha) \). Since the proof is so simple, we repeat it for expository value: \( \alpha' = P_0(L_n \text{ crosses a before } b) \leq P_0(\sup_n L_n \geq a) = Pr(\sup_n W_{-1/2}(T_n) \geq a) \) (upon embedding) \( \leq Pr(\sup_{t} W_{-1/2}(t) \geq a) = e^{-a} \) (by Lemma 6.1) = \( \alpha/(1-\beta) \).

These bounds appear in Wald (1947, (3:25) and (3:26)). He actually obtained the stronger inequalities \( \alpha'/(1-\beta') \leq \alpha/(1-\beta) \) and \( \beta'/(1-\alpha') \leq \beta/(1-\alpha) \). If the increments in \( L_n \) are bounded, asymptotic results (and other bounds) may be obtained; see Corollary 7.4.

Now considering the ASN, we have as a corollary to Lemma 6.4

**COROLLARY 7.3:** If \( E_n^{-1}X_n \geq \mu_0 \) (constant) \( \text{a.s. for all } n \), then \( E_n \geq [(1-\alpha)n\beta/(1-\alpha)] + \alpha n[(1-\beta)/\alpha]/\mu_0 \). If \( E_1^{-1}X_1 \leq \mu_1 \) (constant) \( \text{a.s. for all } n \), then \( E_1 \geq [\beta n\beta/(1-\beta)] + (1-\beta)n[(1-\beta)/\alpha]/\mu_1 \).

These lower bounds are not the same as Wald's (assuming independent and identically distributed (i.i.d.) stages of the experiment). Wald's bounds ((A:205) and (A:206)) have the true, but unknown, values \( (\alpha',\beta') \) where our bounds have the nominal values \( (\alpha,\beta) \). If \( \alpha' \leq \alpha \leq 1/2 \) and \( \beta' \leq \beta \leq 1/2 \), our bounds are \( \geq \) his.

It will be instructive to give a direct proof, confined to the i.i.d. case (and \( H_0 \)). We embed \( \{L_n\} \) in \( W_\delta(\cdot) \) with \( \delta = -1/2 \) (assuming \( H_0 \) true), and let \( T = \text{hitting time by } W_\delta(\cdot) \) of \( \{b,a\} \), \( N = \min(n|L_n \geq a \text{ or } \leq b) \), and \( N' = \min(n|W_\delta(T_n) \geq a \text{ or } \leq b) \). Then \( N \leq N' \), \( T \leq T_{N'} \), \( \delta ET = EW_\delta(T) \) (Wald's equation; see I) = \( (a-b)\alpha + b \) where \( \alpha = Pr[W_\delta(\cdot) \text{ hits a before } b] \) (the nominal error probability), and \( ET_{N'} = E_{j=1}^{N'}(T_{j} - T_{j-1}) = EN' \cdot ET_1 \leq EN' \cdot \mu_0/\delta \) (by Wald's equation and (7')). Putting these together, we obtain
$E_N \leq (s-b)a + b/n$, as claimed. (We have implicitly used the fact that $a_s < b < b_1$, by Jensen's inequality.)

We may also draw conclusions about the operating characteristic (OC) function $L(H) = \Theta(H)$ (SPRT accepts $H_0$), where $H$ is another hypothesis specifying a probability measure $P_H$ on $(\mathcal{F}, \mathcal{G})$. We assume there exists a $\delta > 0$ for which $(L_n, \mathcal{F}_n)$ is embedable in $W_0$ under $H$ (and implicitly that $L_n$ is well-defined).

A sufficient condition is that the stages are i.i.d. and either $E_X = 0$ (in which case $L(H) = 0$) or there exists $\delta \neq 0$ for which $E_X e^{-2\delta X} = 1$, $X$ representing an increment in $L_n$. This is an assumption considered by Wald.

Approximating the sequence $(L_n)$ under $H$ by $W_0(H)(\cdot)$, we have $L(H) = \text{Pr}[W_0(\cdot)]$ (say using Lemma 6.1 and assuming $\delta \neq 0$). This is precisely Wald's 'no overshoot' OC approximation (Wald, 1946); his $h$ is our $\delta^*$. As in Corollary 7.3, we also have $L(H) \leq A_2^\delta$ if $\delta < 0$ and $L(H) \geq 1 - B_{2\delta}$ if $\delta > 0$. An elaboration of Lemma 6.3 (and Lemma 5.3 in [3]) yields:

**COROLLARY 7.4**: Suppose $-\infty \leq c^1 \leq X_n \leq c \leq +\infty$ a.s. for all $n$, and let $C = e^c$, $C = e^{c^1}$. Suppose $(L_n, \mathcal{F}_n)$ is embedable in $W_0(\cdot)$ under $H$, and let $M$ denote the assumption that $(R_n = e^{-2\delta L_n}, \mathcal{F}_H)$ is a martingale sequence. If either $\delta \geq 0$, or $M$ holds and $\delta < 0$, then

$$L(H) \leq \frac{1}{B_{2\delta}} \frac{(AC)_{2\delta}^{2\delta}}{B_{2\delta}^{2\delta}} (= \frac{\delta + c}{\delta - c} \text{ if } \delta = 0).$$
If either $\delta \leq 0$, or $M$ holds and $\delta > 0$, then

$$L(H) \geq \frac{1-A^{-2\delta}}{(\beta c')^{-2\delta} - A^{-2\delta}} - \pi(H) \left(\frac{a}{a-b-c'} - \pi \text{ if } \delta = 0\right)$$

where $\pi(H) = P_H(N = \infty)$. Wald's methods in the i.i.d. case (then $\pi = 0$), with bounded increments, would yield these same bounds (see his (A:31) and (A:34)).

Likewise, the ASN under hypothesis $H$ may be bounded. From Lemma 6.4 (and Lemma 5.4 in II), we obtain

**COROLLARY 7.5:** Suppose the stages are i.i.d. under $H_0$, $H_1$ and $H$, and denote $\mu(H) = E_HX$, $\sigma^2(H) = \text{var}_HX$, and $\delta(H)$ as above. Then

$$E_HN \geq \begin{cases} 
\frac{[a - (a-b)L^*(H)]/\mu(H)}{\sqrt{\delta(H)}} & \text{if } \delta(H) \neq 0 \\
- ab/\sigma^2(H) & \text{if } \delta(H) = 0 
\end{cases}$$

These lower bounds are the same as Wald's approximations (3:57) and (A:99) to $E_HN$, except that we have the approximate OC $L^*(H)$ instead of the true OC $L(H)$ appearing. Since these lower bounds are explicit functions of $\alpha$, $\beta$, $\delta$ and $\mu$ (or $\sigma^2$), they are much simpler than Wald's lower bounds given by (A:77) and (A:78).

Moreover, our bound exceeds his if $\delta(H) < 0$ and $L^*(H) \geq L(H)$ or if $\delta(H) > 0$ and $L^*(H) \leq L(H)$.

If the increment $X$ is bounded, an upper bound on the ASN may be obtained as in Lemma 6.4 (and Lemma 5.4 in II). Specifically, under the assumptions of Corollary 7.5 and additionally $c' \leq X \leq c$.
a.s., the lower bounds on $\text{E}_H N$ given above become upper bounds upon replacing $a$ by $a + c$ and $b$ by $b + c'$.

Finally, in the i.i.d. case with bounded (by $K$) increments, we obtain from Theorem 6.7 (and Theorem 5 in II; see remarks at end of Section 5 above) all of the following: $\alpha' = \alpha + O(K)$, $\beta' = \beta + O(K)$, $L(H) = L^*(H) + O(K)$, and the lower bounds on the ASN given above are asymptotic equalities; also, the sequential sample size $N$ converges in law to an appropriately scaled hitting time by $W_b(\cdot)$ of $\{b,a\}$.

We have confined attention to SPRT's. Whether the approximations suggested by the embedding theorems are appropriate for other sequential tests is not apparent. For, while the distortion of the time axis in the embedding does not effect the horizontal line barriers of the SPRT, it would effect those of other sequential tests, for example, the closed sequential tests of Armitage (1957) and Anderson (1960).
REFERENCES


