1 Preliminary Discussion

1.1 Main Idea

The main idea of Skorohod embedding is that, given a random walk \( S_n = X_1 + \cdots + X_n \), with \( X_i \) independent and \( X_i \overset{d}{=} X \) with \( \mathbb{E}[X] = 0 \) and \( \mathbb{E}[X^2] < \infty \), we may embed the random walk in a Brownian motion \((B_t, t \geq 0)\). More specifically, we may find an increasing sequence of stopping times \( T_n \) so that

\[
(S_1, S_2, \ldots) \overset{d}{=} (B(T_1), B(T_2), \ldots)
\]

with \( \mathbb{E}[T_n] < \infty \). It is important to emphasize that the condition \( \mathbb{E}[T_n] < \infty \) is essential for all applications. Recall Wald’s identity and note then that we have

\[
\mathbb{E}[T_n] = \mathbb{E}[B^2_{T_n}] = \mathbb{E}[S^2_n] = n\mathbb{E}[X^2].
\]

It may be easily seen that to prove this result we need only embed a single variable \( X \), i.e., we need only find a stopping time \( T \) with \( \mathbb{E}[T] < \infty \) and \( B_T \overset{d}{=} X \). Why is this? Because once we have achieved this, then we may use the strong Markov property to start things anew and then the same technique may be applied repeatedly. Thus,

\[
T_1 = T = T(B)
\]

then

\[
B^{(2)}(t) = B(T_1 + t) - B(T_1)
\]

is a again a BM independent of \( B(T_1) \).

so we can

\[
T_2 = T_1 + T(B^{(2)})
\]

and so on. Thus for each \( n \) we have

\[
T_{n+1} = T_n + T(B^{(n+1)})
\]
and
\[ B^{(n+1)}(t) = B(T_n + t) - B(T_n) \]

**Remark:** There is nothing like uniqueness in embedding. There are many, genuinely different, ways to do this (see [4]).

**Method in Text:** Durrett uses the original method due to Skorohod. This method does not give a \( T = T(B) \), but instead gives a \( T = T(B, U) \) where \( U \sim U[0, 1] \) and independent of \( B \). We will instead use a method due to Dubins [5].

## 2 Embedding

### 2.1 Foundation

The building block for both Skorohod's technique and Dubins' technique is to first consider 2-point distributions. Suppose \( X \in \{-a, b\} \) and \( \mathbb{E}[X] = 0 \). Thus,
\[ \mathbb{P}(X = b) = \frac{a}{a + b}, \quad \mathbb{P}(X = -a) = \frac{b}{a + b}. \]

This is the most general 2-point distribution with mean zero. How do we embed this? First let,
\[ T = \inf \{ t : B_t \in \{-a, b\} \}. \]

Then \( B_T \in \{-a, b\} \) and \( T \) is the earliest time with this property. However, do we have that \( \mathbb{E}[T] < \infty \)? Yes! Take \( C > a + b \) and consider \( \{ n : |B(n + 1) - B(n)| > c \} \). This is a geometrically distributed stopping time which occurs after \( T \) which has finite expectation. Thus it must be that \( T \) has finite expectation. Finally, By Wald’s identity we have
\[ \mathbb{E}[B_T] = 0 \Rightarrow \mathbb{P}(B_T = b) = \frac{a}{a + b}, \quad \mathbb{P}(B_T = -a) = \frac{b}{a + b}. \]

We have found the (unique) stopping time \( T \) so \( B_T \) has a 2-point stopping time with mean zero.

**Remark:** This is where Skorohod and Dubins’ techniques diverge.

### 2.2 A Simple Example

In this example, let us embed a random walk where \( X \stackrel{d}{=} U[-1, 1] \). To do so we start the Brownian motion at 0 and we draw gridlines at \( -\frac{1}{2} \) and \( \frac{1}{2} \) and we see which of the two we hit first. Specifically,
\[ B(T_1) \stackrel{d}{=} U \left[ -\frac{1}{2}, \frac{1}{2} \right]. \]
Next we draw gridlines at \(-\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \text{ and } \frac{3}{4}\). Again, we see which of the four we hit first. Specifically,

\[ B(T_2) \stackrel{d}{=} U \left[ \left[ -\frac{3}{4}, -\frac{1}{4} \right], \left[ \frac{1}{4}, \frac{3}{4} \right] \right]. \]

Clearly, we may continue on in this fashion and at each \(n\) we will be considering a uniform distribution over \(2^n\) values. Note also that these \(2^n\) values are evenly spaced on a lattice \(\subset [-1, 1]\).

\(B(T_n)\) converges in distribution to a \(U[-1, 1]\). To see this, note that \(T_\infty\) hits before the Brownian motion leaves \([-1, 1]\) so \(\mathbb{E}\left[ T_n \right] \uparrow \mathbb{E}\left[ T_\infty \right] < \infty\) by a similar argument as above. Also, \(\mathbb{P}(T_\infty < \infty) = 1\) so \(B(T_n) \xrightarrow{a.s.} B(T_\infty)\). Hence, \(B(T_\infty) = d U[-1, 1]\) and we have achieved the embedding.

### 2.3 Embedding Random Variables in Brownian Motion

**Theorem 1** Let \(X\) be a r.v. with \(\mathbb{E}X = 0\) and \(\mathbb{E}X^2 < \infty\). Then there exists a stopping time \(T\) of Brownian Motion such that \(\mathbb{E}T < \infty\) and \(B_T = d X\). Hence (by previous theorem) \(\mathbb{E}B_T = 0\) and \(\mathbb{E}B_T^2 = \mathbb{E}T = \mathbb{E}X^2\)

First a little work with no Brownian Motion in view. We make a nice discrete approximations of any r.v. \(X\) on \((\Omega, \mathcal{F}, \mathbb{P})\) with \(\mathbb{E}X = 0\) and \(\mathbb{E}X^2 < \infty\).

First, let \(X_0 = 0\). Next let

\[ G_1 = \sigma(X > X_0), \text{ i.e., } G_1 = \{\emptyset, (X > X_0), (X \leq X_0), \Omega\} \]

and

\[ X_1 = \mathbb{E}(X|G_1) \]

Then \(X_1\) has two values \(\mathbb{E}(X|X > 0)\) and \(\mathbb{E}(X|X \leq 0)\) and can be rewritten as

\[ X_1 = \mathbb{E}(X|X > 0)1(X > 0) + \mathbb{E}(X|X \leq 0)1(X \leq 0). \]

Moreover, by previous arguments we find \(T_1\) so that \(B(T_1) = d X_1\) with \(\mathbb{E}[T_1] < \infty\). Now, let

\[ G_2 = \sigma(G_1, X > X_1) \]

\[ X_2 = \mathbb{E}(X|G_2) \]

So \(G_2\) has 4 atoms \((X > 0, X > X_1), (X > 0, X \leq X_1), (X \leq 0, X > X_1)\) and \((X \leq 0, X \leq X_1)\) and thus \(X_2\) typically has four possible values and \(B(T_2) = d X_2\) with \(\mathbb{E}[T_2] < \infty\). Inductively

\[ G_{n+1} = \sigma(G_n, X > X_n) \]
\[ X_{n+1} = \mathbb{E}(X|\mathcal{G}_{n+1}) \]

Then \( \mathcal{G}_n \) is generated by a partition of the probability space into at most \( 2^n \) sets, and \( B(T_n) =^d X_n \) with \( \mathbb{E}[T_n] < \infty \).

Observe that \( \mathcal{G}_n \uparrow \mathcal{G}_\infty \). In addition, 
\[
\mathbb{E}(X|\mathcal{G}_n) \uparrow \mathbb{E}(X|\mathcal{G}_\infty) \leq \mathbb{E}(X^2) < \infty
\]
by assumption. Thus, \( \mathbb{E}(X|\mathcal{G}_n) = \mathbb{E}(T_n) \) by Wald’s identity and note that \( \mathbb{E}(T_n) \uparrow \mathbb{E}(T_\infty) \leq \mathbb{E}(X^2) < \infty \). Finally, \( T_n \uparrow T_\infty \) with probability 1 and so we have,
\[
B(T_\infty) =^d \mathbb{E}(X|\mathcal{G}_\infty)
\]
by the fact that \( B(T_n) \to^{as} B(T_\infty) \) and the claim below. However, also by the claim below we have that \( X_\infty =^{as} X \) so that the result follows.

**Claim 1** \( X_n \to X \) both almost surely and in \( L^2 \) as \( n \to \infty \).

**Proof:** Using \( X_n = \mathbb{E}(X|\mathcal{G}_n) \) and Jensen’s inequality for conditional expectation, we obtain
\[
\mathbb{E}(X_n^2) = \mathbb{E}([\mathbb{E}(X|\mathcal{G}_n)]^2) \leq \mathbb{E}(\mathbb{E}(X^2|\mathcal{G}_n)) = \mathbb{E}(X^2)
\]
Hence,
\[
\sup_n \mathbb{E}(X_n^2) < \infty
\]
So as for the martingale \( (X_n, \mathcal{G}_n) \), by \( L^2 \) convergence theorem, we know
\[
X_n \to X_\infty
\]
both almost surely and in \( L^2 \) for some square-integrable limit \( X_\infty \). But \( X_n = \mathbb{E}(X|\mathcal{G}_n) \) implies that
\[
X_n \to^{as} \mathbb{E}(X|\mathcal{G}_\infty)
\]
where \( \mathcal{G}_\infty \) is the \( \sigma \)-field generated by \( \cup_n \mathcal{G}_n \). Thus \( X_\infty = \mathbb{E}(X|\mathcal{G}_\infty) \) and our goal is to prove \( X_\infty = X \) a.s.

One proof is given by Billingsley [1]. A nicer argument is suggested by J. Neveu [2], (p. 34, Exercise II-7).

Notice the following facts
\[
(X > X_\infty) \subseteq \cup_n \cap_{m \geq n} (X > X_n) \subseteq (X \geq X_\infty)
\]
and
\[
\mathbb{E}((X - X_\infty)1(X > X_\infty)) = \mathbb{E}((X - X_\infty)1(X \geq X_\infty))
\]
Then if let
\[
G = \cup_n \cap_{m \geq n} (X > X_n)
\]
Well have
\[ \mathbb{E}((X - X_\infty)1(X > X_\infty)) = \mathbb{E}((X - X_\infty)1_G) = (\mathbb{E}((X - X_\infty)1(X \geq X_\infty)) \]

But the fact that \( X_\infty = \mathbb{E}(X|G_\infty) \) makes
\[ \mathbb{E}(X1_G) = \mathbb{E}(X_\infty1_G) \]
since \( G \in G_\infty \). Hence
\[ \mathbb{E}((X - X_\infty)1(X > X_\infty)) = 0 \]
Same argument on \( (X < X_\infty) \) leads to
\[ \mathbb{E}((X - X_\infty)1(X < X_\infty)) = 0 \]
The last two observations imply
\[ \mathbb{E}|X - X_\infty| = 0 \]
We immediately obtain the desired result
\[ X \overset{a.s.}{=} X_\infty \]
The proof of the claim is complete.

### 3 References