A formalisation of Dubins’ proof to Skorokhod’s embedding theorem

Shaikh Ammar • 14 Dec 2025

Introduction

The Skorokhod Embedding theorem says the following:

Theorem 1 (Skorokhod’s embedding theorem). *Let ξ be a mean-zero finite variance random variable. Let $\{B(t) \mid t \geq 0\}$ be a Brownian motion. Then, there is a stopping time T with respect to the natural filtration of the Brownian motion such that*

$$B(T) \overset{\mathrm{d}}{=} \xi, \qquad \mathbb{E}T = \mathbb{E}\xi^2.$$

With this theorem, one can embed any random walk whose increments are given by the distribution of ξ into a Brownian motion, allowing us to study random walks through studying Brownian motion.

Skorokhod, however, did not prove this statement exactly. He relied on “external randomisation” to achieve this result. Dubins however realised the distribution of ξ within the *natural filtration* of Brownian motion itself—essentially proving a manner of probability integral transform for Brownian motion.

Despite the clear significance of the theorem, I have not managed to find an exact rigorous treatment of the proof. Proofs can be found in Mörters and Peres (2010), Dubins (1968), and Meyer (1971) among other resources, but they all have faults in rigour somewhere or the other, and the steps to make it rigorous is not completely straightforward either. Some of the unjustified statements actually require the optionally stopping theorem which is nowhere mentioned in these proofs. I present a clear formalisation.

Preliminary results

First, we shall state Wald’s lemma (for a proof, see Mörters and Peres (2010, theorems 2.44 and 2.48)), which allows us to prove the embedding theorem for simple random walks.

Theorem 2 (Wald's lemma)**.** *Let T be a stopping time of finite mean with respect to the natural filtration of a Brownian motion $\{B(t) \mid t \geq 0\}$. Then,*

$$\mathbb{E}B(T) = 0, \qquad \mathbb{E}[B(T)^2] = \mathbb{E}T.$$

Lemma 3 (Embedding theorem for simple random walks)**.** *Let ξ be a mean-zero random variable supported on two values $\{a, b\}$. Let $T = \tau_{\{a,b\}}$ be the hitting time of $\{a, b\}$. Then, ξ is embedded in $B(T)$ as in theorem 1.*

Proof. Hitting times of closed sets are stopping times, so T is indeed a stopping time. $\mathbb{E}\xi = 0$ mandates $a \leq 0 \leq b$. Define $T_n = T \wedge n$. Then, by Wald's lemma

$$\mathbb{E}B(T_n) = 0, \qquad \mathbb{E}[B(T_n)^2] = \mathbb{E}T_n.$$

Almost surely, $B(T_n) \to B(T)$, and $B(T_n)^2 \to B(T)^2$, and

$$|B(T_n)| \leq \max(-a, b), \qquad B(T_n)^2 \leq \max(a^2, b^2).$$

So, by DCT and MCT,

$$\begin{aligned} \mathbb{E}B(T) &= \lim \mathbb{E}B(T_n) \\ &= 0 \end{aligned} \qquad \begin{aligned} \mathbb{E}[B(T)^2] &= \lim \mathbb{E}[B(T_n)^2] \\ &= \lim \mathbb{E}T_n \\ &= \mathbb{E}T. \end{aligned}$$

This proves Wald's identities for T.
Now, note that there is only one probability distribution supported on $\{a, b\}$ and with expectation 0. ξ follows this distribution, and we have just proved $B(T)$ also does. So, they are distributed the same. Using this on the second identity, we get that

$$\mathbb{E}\xi^2 = \mathbb{E}[B(T)^2] = \mathbb{E}T. \qquad \square$$

Now, we shall state some standard results on martingales. Their proofs can be found in Durret (2019, theorems 4.4.6, 4.6.3) and Mörters and Peres (2010, proposition 2.42).

Theorem 4 (Martingale L^2-convergence theorem)**.** *A discrete-time martingale bounded in L^2 converges almost surely and in L^2 to an identical limit.*

Lemma 5. *If $\{X_n \mid n \in \mathbb{N}\}$ is uniformly integrable, and $X_n \to X$ in probability, then $\mathbb{E}X_n \to \mathbb{E}X$.*

Theorem 6 (Optional stopping theorem)**.** *Let $0 \leq S \leq T$ be stopping times for a continuous martingale $\{M_t \mid t \geq 0\}$, such that $|M(T \wedge t)|$ is bounded by an integrable random variable, then*

$$\mathbb{E}[M(T) \mid \mathcal{F}_S] = M(S),$$

where $\mathcal{F}_S$ denotes the stopped σ-algebra.

Dubins' approximation

Dubins (1968) showed how a random variable having finite variance can be approximated by employing a sort of binary search algorithm. Consider a random variable X. We let X_0 to be the expectation of X i.e., our "best guess" when no other information is present. Now, add the information whether $X \geq X_0$ or not. Our next "best guess" X_1 is the expectation on X conditioned on this information. We repeat this for successive n. The resultant sequence is a uniformly integrable martingale that converges to X both almost surely and in L^2.

Definition (Binary-splitting martingale)**.** Let $\{(X_n, \mathcal{G}_n) \mid n \in \mathbb{N}\}$ be a martingale. Then, it is said to be *binary-splitting* if whenever the event

$$A(x_0, x_1, \ldots, x_n) = \{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\}$$

has non-zero probability, X_{n+1} conditioned on this is supported on at most two values.

We will show that our sequence of guesses form a binary-splitting martingale.

Theorem 7 (Dubins' approximation)**.** *Let $X : \Omega \to \mathbb{R}$ be a random variable having finite variance. Then there is a binary-splitting martingale converging to it almost surely and in L^2.*

Define $\mathcal{G}_0 = \{\emptyset, \Omega\}$, $X_0 = \mathbb{E}X$. Recursively, define $\epsilon_n = \mathbf{1}_{\{X \geq X_{n-1}\}}$, $\mathcal{G}_n = \sigma(\epsilon_1, \ldots, \epsilon_n)$, and $X_n = \mathbb{E}[X \mid \mathcal{G}_n]$. It is clear that we have defined a martingale.

Let $E_n : \Omega \to \{0,1\}^n$ denote $(\epsilon_1, \ldots, \epsilon_n)$. Since the Borel σ-algebra on $\{0,1\}^n$ is generated by the singletons, and $\mathcal{G}_n$ is the smallest σ-aglebra making E_n measurable, it follows that $\mathcal{G}_n$ is generated by the 2^n-sized partition

$$\mathcal{P}_n = \{E_n^{-1}(e) \mid e \in \{0,1\}^n\}.$$

The conditional expectation X_n is given by

$$X_n(\omega) = \frac{\mathbb{E}[X\mathbf{1}_{\{E_n = e\}}]}{P\{E_n = e\}} \quad \text{if } E_n(\omega) = e,$$

provided $P\{E_n = e\} \neq 0$. We write $X_n(e)$ for the (constant) value of X_n on ω in $\{E_n = e\}$ i.e., $X_n(e) = \mathbb{E}[X \mid E_n = e]$.

Clearly, conditioned on $E_n = e$, X_{n+1} is supported on at most two values provided the former is not of zero probability. We want to show that such a condition is equivalent to conditioning on the values of $X_1, \ldots, X_n$ so as to prove that the martingale defined is binary-splitting. Consider the lexicographic ordering on $\{0,1\}^n$. We claim that for $e, m \in \{0,1\}^n$, $e < m$ if and only if $X_n(e) < X_n(m)$ provided probability of neither $\{E_n = e\}$ nor $\{E_n = m\}$ is zero. Let $e < m$ such that neither $\{E_n = m\}$ nor $\{E_n = e\}$ have probability zero. We write e_i for the i-th element of e and similarly m_i. Let i be maximal such that $e_j = m_j$ for all $j \leq i$ (define it to be 0 if it does not exist). Then $0 = e_{i+1} < m_{i+1} = 1$. For $\omega \in \{E_n = m\}$ and $\omega' \in \{E_n = e\}$, we have $X_i(\omega) = X_i(\omega')$ since by martingale property

$$X_i(\omega) = \mathbb{E}[X_n \mid E_i = (m_1, \dots, m_i)] = \mathbb{E}[X_n \mid E_i = (e_1, \dots, e_i)] = X_i(\omega').$$

If the $(i+1)$-th element of E_n is 1, that means $X \geq X_i$, and if 0, then $X < X_i$. Hence, for ω, ω' as defined before,

$$\begin{aligned} X_n(\omega) = \frac{\mathbb{E}[X\mathbf{1}_{\{E_n=m\}}]}{P\{E_n = m\}} &\geq X_i(\omega) \\ &= X_i(\omega') > \frac{\mathbb{E}[X\mathbf{1}_{\{E_n=e\}}]}{P\{E_n = e\}} = X_n(\omega'), \end{aligned}$$

which proves the necessity. To prove sufficiency, assume $X_n(e) < X_n(m)$. Then, e cannot equal m provided the probabilities are non-zero. Define i, ω and ω' as before. We still have $X_i(\omega) = X_i(\omega')$. If $0 = m_{i+1} < e_{i+1} = 1$, then, similarly as before,

$$\begin{aligned} X_n(\omega) = \frac{\mathbb{E}[X\mathbf{1}_{\{E_n=m\}}]}{P\{E_n = m\}} &< X_i(\omega) \\ &= X_i(\omega') \leq \frac{\mathbb{E}[X\mathbf{1}_{\{E_n=e\}}]}{P\{E_n = e\}} = X_n(\omega'), \end{aligned}$$

which contradicts that $X_n(\omega) > X_n(\omega')$. The claim is proved.

Write $A(e)$ for the event

$$\{X_0 = 0, X_1 = X_1(e_1), X_2 = X_2(e_1, e_2), \dots, X_n = X_n(e)\}.$$

Now, $A(e)$ is an event about $\mathcal{G}_n$ measurable functions and so, is in $\mathcal{G}_n$, and hence, is a union of events of type $\{E_n = e\}$ (since these form a partition that generate $\mathcal{G}_n$), but due to the claim just proved, it must be that $A(e) = \{E_n = e\}$. So, conditioning on $A(e)$ is same as conditioning on $\{E_n = e\}$, and under this condition X_{n+1} is supported on two values $X_n(e, 1)$ and $X_n(e, 0)$. Therefore, the martingale defined is binary-splitting.

Using that conditional expectations are projections, $\mathbb{E}X^2 \geq \mathbb{E}X_n^2$, which shows that the martingale is L^2 bounded, and hence, converges to some X_∞ almost surely and in L^2 by Martingale L^2-convergence theorem.

We shall show that $X_\infty = X$. Let $\varepsilon_n = 2\epsilon_n - 1$. We claim that

$$\varepsilon_n(X - X_n) \to |X - X_\infty| \quad \text{a.s.}$$

If $X \geq X_\infty$, then eventually $X \geq X_n$ and $\epsilon_n = 1$, and almost surely,

$$\epsilon_n(X - X_n) \to X - X_\infty = |X - X_\infty|.$$

If instead, $X < X_\infty$, then similarly $X < X_n$ and $\epsilon_n = -1$, and almost surely,

$$\epsilon_n(X - X_n) \to -X + X_\infty = |X - X_n|.$$

By *Hölder's inequality*, $\epsilon_n(X - X_n)$ is uniformly integrable as

$$\begin{aligned}
\mathbb{E}|\epsilon_n(X - X_n)\mathbf{1}_A| &= \mathbb{E}|(X - X_n)\mathbf{1}_A| \\
&\leq \mathbb{E}|X\mathbf{1}_A| + \mathbb{E}|X_n\mathbf{1}_A| \\
&\leq [(\mathbb{E}X^2)^{1/2} + (\mathbb{E}X_n^2)^{1/2}]P(A)^{1/2} \\
&\leq 2(\mathbb{E}X^2)^{1/2}P(A)^{1/2},
\end{aligned}$$

which goes to 0 as $P(A)$ goes to 0. Therefore, By lemma 5,

$$\mathbb{E}[\epsilon_n(X - X_n)] \to \mathbb{E}|X - X_\infty|,$$

but

$$\mathbb{E}[\epsilon_n(X - X_n)] = \mathbb{E}[\epsilon_n \mathbb{E}[X - X_n \mid \mathcal{G}_n]] = 0.$$

Therefore,

$$\mathbb{E}|X - X_\infty| = 0 \implies X \overset{\text{a.s.}}{=} X_\infty.$$

Proof of the Skhorokhod embedding theorem

Theorem 8 (Dubins' embedding)**.** *Consider the setup as in theorem 1. Let* $\{\xi_n \mid n \in \mathbb{N}\}$ *be the Dubins approximation of* ξ*. Define* T_1 *to be the hitting time of* $\{\xi_1(0), \xi_1(1)\}$*. Recursively, define*

$$T_n = \inf\{t \geq T_{n-1} \mid B(t) \in \{\xi_n(e) \mid e \in \{0,1\}^n\}\}.$$

Then, ξ *is embedded in* $B(\sup T_n)$*.*

It follows from lemma 3 that ξ_1 is embedded in $B(T_1)$ and $B(T_1 \wedge t)$ is uniformly bounded in t. Assume the induction hypothesis that $B(T_n) \overset{\text{d}}{=} \xi_n$ and $B(T_n \wedge t)$ is uniformly bounded in t by some integrable M_n. Then, $|B(T_{n+1} \wedge t)|$ is uniformly bounded since for $t \leq T_n$,

$$|B(T_{n+1} \wedge t)| = |B(T_n \wedge t)| \leq M_n$$

and for $t > T_n$, if $E_n \in \{0,1\}^n$ denotes the unique binary word in $\{0,1\}^n$ such that $B(T_n) = \xi_n(E_n)$,

$$\begin{aligned}
|B(T_{n+1} \wedge t)| &\leq \max_{i \in \{0,1\}} |\xi_{n+1}(E_n, i)| \\
&\leq \max_{x \in \{0,1\}^{n+1}} |\xi_{n+1}(x)|.
\end{aligned}$$

So, by Optional stopping theorem,

$$\mathbb{E}[B(T_{n+1}) \mid B(T_n) = \xi_n(e)] = \xi_n(e),$$

and by martingale property,

$$\mathbb{E}[\xi_{n+1} \mid \xi_n = \xi_n(e)] = \xi_n(e)$$

Both the conditional distributions of $B(T_{n+1}) \mid B(T_n) = \xi_n(e)$ and $\xi_{n+1} \mid \xi_n = \xi_n(e)$ are centred on $\xi_n(e)$ and supported on $\{\xi_{n+1}(e,0), \xi_{n+1}(e,1)\}$, which shows they are equal since there is only one such distribution. In the following equation, for ease of notation, we are going to write e and m for $\xi_n(e)$ and $\xi_{n+1}(m)$.

$$\begin{aligned} P(B(T_{n+1}) = m) &= \sum_{e \in \{0,1\}^n} P(B(T_{n+1}) = m \mid B(T_n) = e) P(B(T_n) = e) \\ &= \sum_{e \in \{0,1\}^n} P(\xi_{n+1} = m \mid \xi_n = e) P(\xi_n = e) \\ &= P(\xi_{n+1} = m), \end{aligned}$$

which proves that $B(T_{n+1}) \overset{\text{d}}{=} \xi_{n+1}$. Now, to prove the condition of second moment, we use the martingale $\{B(t)^2 - t \mid t \geq 0\}$. Assume the induction hypothesis that $\mathbb{E}T_n = \mathbb{E}\xi_n^2$. Define for all $e \in \{0,1\}^n$,

$$\tau_e = \inf\{t \geq T_n \mid B(t) \in \{(e,0),(e,1)\}\}.$$

Then,

$$T_{n+1} = \sum_{e \in \{0,1\}^n} \tau_e \mathbf{1}_{\{T_n = e\}}.$$

Define τ_e^k to be $\tau_e \wedge k$. Then, using MCT and Wald's lemma,

$$\mathbb{E}\tau_e = \lim_{k \to \infty} \mathbb{E}\tau_e^k = \lim_{k \to \infty} \mathbb{E}[B(\tau_e^k)^2].$$

For $k \leq T_n$,

$$B(\tau_e^k)^2 = B(T_n \wedge k)^2 \leq M_n^2,$$

and for $k > T_n$, similarly as before,

$$\begin{aligned} B(\tau_e^k)^2 &\leq \max_{i \in \{0,1\}} \xi_{n+1}(E_n, i)^2 \\ &\leq \max_{x \in \{0,1\}^n} \xi_{n+1}(x)^2. \end{aligned}$$

So, by DCT,

$$\mathbb{E}\tau_e = \lim \mathbb{E}[B(\tau_e^k)^2] = \mathbb{E}[B(\tau_e)^2] \leq \max_{x \in \{0,1\}^n} \xi_{n+1}(x)^2.$$

Hence, T_{n+1}—being a sum of finite expectation variables—is also of finite expectation. So, we have that

$$|B(T_{n+1} \wedge t)^2 - T_{n+1} \wedge t| \leq M_{n+1}^2 + T_{n+1},$$

where the bound is integrable. So, we can use Optional stopping theorem to obtain

$$\begin{aligned} &\mathbb{E}[B(T_{n+1})^2 - T_{n+1} \mid \mathcal{F}_{T_n}] = B(T_n)^2 - T_n \\ \implies &\mathbb{E}[B(T_{n+1})^2 - T_{n+1}] = \mathbb{E}[B(T_n)^2 - T_n] = 0. \end{aligned}$$

Hence, $\mathbb{E}\xi_{n+1}^2 = \mathbb{E}[B(T_{n+1})^2] = \mathbb{E}T_{n+1}$.

So, we have a non-decreasing sequence $\{T_n \mid n \in \mathbb{N}\}$ of stopping times such that

$$B(T_n) \overset{\mathrm{d}}{=} \xi_n, \qquad \mathbb{E}T_n = \mathbb{E}\xi_n^2.$$

Supremum of countable stopping times is also a stopping time, which makes T one. By MCT, and L^2 convergence,

$$\mathbb{E}T = \lim \mathbb{E}T_n = \lim \mathbb{E}\xi_n^2 = \mathbb{E}\xi^2.$$

Since $B(T_n) \overset{\mathrm{u}}{=} \xi_n$, and ξ_n converges almost surely to ξ, we get that $B(T_n)$ converges in distribution to ξ. Since $B(T_n)$ converges almost surely to $B(T)$, it follows that $B(T) \overset{\mathrm{u}}{=} \xi$.

References

Dubins, Lester E. 1968. "On a Theorem of Skorohod." *The Annals of Mathematical Statistics* 39 (6): 2094–97. https://doi.org/10.1214/aoms/1177698036.

Durret, Rick. 2019. *Probability - Theory and Examples*. 5th ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Meyer, Paul-André. 1971." Sur Un Article de Dubins." *Séminaire de Probabilités de Strasbourg* 5: 170–76. https://www.numdam.org/item/SPS_1971__5__170_0/.

Mörters, P., and Y. Peres. 2010. *Brownian Motion*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.