Lecture 12

Lecture date: May 23, 2011 — *Scribe: Naomi Feldheim*

In this lecture we define continuous time martingales, and prove Wald's identities concerning Brownian motion. Later on we discuss Skorohod embedding, which is the problem of sampling a certain random walk out of a sample of the Brownian motion. We define the so-called Azéme-Yor embedding, and show it solves the problem. At last we discuss the Donsker's invariance principle, which roughly states that Brownian motion is the universal scaling limit of any mean-zero finite-variance random walk.

Tags for today's lecture: continuous time martingales, Wald identities on BROWNIAN MOTION, Skorohod embedding problem, Azéme-Yor embedding, Donsker's invariance principle.

1 Continuous Time Martingales

Definition 1 *A real valued stochastic process* $\{X(t)\}_{t\geq 0}$ *adapted to a filteration* $\{\mathcal{F}(t)\}_{t\geq 0}$ *is a* martingale *if:*

- $\forall t,\ \mathbb{E}|X(t)| < \infty$
- $\forall 0 < s \leq t,\ \mathbb{E}(X(t)|\mathcal{F}(s)) = X(s)$ *a.s.*

A *submartingale* is defined similarly, but with $\mathbb{E}(X(t)|\mathcal{F}(s)) \geq X(s)$ a.s., and a *supermartingale* has $\mathbb{E}(X(t)|\mathcal{F}(s)) \leq X(s)$ a.s..

Definition 2 *A martingale* X *is called* continuous *if almost surely, the function* $t \mapsto X(t)$ *is continuous.*

Examples:

1. 1D Brownian motion (1-dim. Brownian motion) is a martingale w.r.t. $\mathcal{F}^+$. Indeed, $\forall t \geq 0\ \mathbb{E}|B(t)| < \infty$, B is adapted to $\mathcal{F}^+$, and $\forall 0 \leq s \leq t$, we have

   $$\mathbb{E}(B(t)|\mathcal{F}^+(s)) = \mathbb{E}(B(t) - B(s)|\mathcal{F}^+(s)) + B(s) = \mathbb{E}(B(t) - B(s)) + B(s) = 0 + B(s)$$

   The second equality holds by the Markov property w.r.t. $\mathcal{F}^+$. Furthermore, B is continuous.

2. If $\{p(t)|t \geq 0\}$ is the counting function of a Poisson process, then $p(t)-t$ is a martingale w.r.t. $\mathcal{F}(t) = \sigma(p(s)|s \leq t)$.

   Indeed, note that $p(t) \sim \text{Poisson}(t)$, so $\mathbb{E}p(t) - t = 0 = \mathbb{E}p(0)$. The other properties are easily verified. This is not a continuous process.

3. Let B be a $1D$ Brownian motion, then $\{B(t) - t\}_{t \geq 0}$ is a martingale w.r.t. $\mathcal{F}^+$. Adapteness and integrability are clear. For $0 \leq s \leq t$:

$$
\begin{aligned}
&B(t)^2 = (B(t) - B(s))^2 + B(s)^2 + 2B(t)(B(t) - B(s)) \\
&\mathbb{E}[(B(t) - B(s))^2 | \mathcal{F}^+(s)] = t - s \ \ \text{(by Markov prop.)} \\
&\mathbb{E}[B(t) - B(s) | \mathcal{F}^+(s)] = 0 \ \ \text{(again by Markov prop.)} \\
&\Rightarrow \mathbb{E}[B(t)^2 \,|\, \mathcal{F}^+(s)] = B(s)^2 + t - s, \ \ \text{as needed.}
\end{aligned}
$$

A fundamental tool in this subject is

Theorem 3 (continuous time optional stopping lemma - COSL). *If $\{X(t)\}_{t \geq 0}$ is a cont. martingale w.r.t. $\mathcal{F}$, T is a stopping time w.r.t. $\mathcal{F}$, and there exists an integrable random variable Z s.t. $|X(T \wedge t)| \leq Z$ (a.s.), then:*

$$\mathbb{E}X(T) = \mathbb{E}X(0).$$

Remark 4 *Proof uses discrete time result and approximation.*

There are many other results of similar sort, e.g., if S, T are stopping times, $S \leq T$, and $|X(t \wedge T)| \leq Z$ for some integrable random variable Z, then $\mathbb{E}(X(T) \,|\, \mathcal{F}^+(S)) = X(S)$ almost surely.

Proposition 5 (Wald's lemma for Brownian motion). *If B is a $1D$ Brownian motion and T is a stopping time w.r.t. $\mathcal{F}^+$ s.t. $\mathbb{E}T \leq \infty$, then:*

1. $\mathbb{E}B(T) = 0$.

2. $\mathbb{E}B(T)^2 = \mathbb{E}T$.

Proof Plan: We shall use the COSL, first for the martingale $B(t)$ and then for $B(t)^2 - t$. The problem is to find an integrable majorant Z (first for $B(T \wedge t)$).

Let $M_k = \sup_{t \in [0,1]} |B(k+t) - B(k)|$, and $M = \sum_{k=1}^{\lceil T \rceil} M_k$.
Note $|B(T \wedge t)| \leq M$ a.s. for all t. Hence if $\mathbb{E}M < \infty$, we are done for the first identity.

$$\mathbb{E}M = \mathbb{E}\sum_{k=1}^{\infty} \mathbb{1}_{\{T>k-1\}} M_k = \sum_{k=1}^{\infty} \mathbb{E}(M_k \mathbb{1}_{\{T>k-1\}}) = \star$$

Notice that M_k depends only on $B(t+k) - B(k)$ for $t \geq 0$, and the indicator $\mathbb{1}_{\{T>k-1\}} \in \mathcal{F}^+(k-1)$. So, by the Markov property we get:

$$\mathbb{E}M = \star = \sum_{k=1}^{\infty} \mathbb{E}M_k \, \mathbf{P}(T > k-1) = \mathbb{E}M_0 \sum_{k=1}^{\infty} \mathbf{P}(T > k-1) = \mathbb{E}M_0 \, \mathbb{E}(T+1).$$

Hence it is enough to show $\mathbb{E}M_0 < \infty$. This f ollows since we showed $\mathbf{P}(M_0 > a) = \mathbf{P}(|B(1)| > a)$. This proves identity no. 1.

For no. 2, we first define the stopping times:

$$H_n = \min(t \geq 0 : \, |b(t)| = n), \qquad\qquad T_n = t \wedge H_n.$$

Notice that $|B(t \wedge T_n)^2 - (t \wedge T_n)| \leq n^2 + T$, and $\mathbb{E}(n^2 + T) < \infty$, hence by the COST $\mathbb{E}B(T_n)^2 = \mathbb{E}T_n$. By Fatou's lemma (which says that if X_n converges a.s. to X, then $\mathbb{E}X \leq \liminf_{n\to\infty} \mathbb{E}X_n$), we get:

$$\mathbb{E}B(T)^2 \leq \liminf_{n\to\infty} \mathbb{E}B(T_n)^2 = \liminf_{n\to\infty} \mathbb{E}T_n = \mathbb{E}T,$$

where the very last step is achieved by the monotone converging theorem.

For the other direction, we note:

$$\mathbb{E}B(T)^2 = \mathbb{E}(B(T) - B(T_n))^2 + \mathbb{E}B(T_n) + 2\mathbb{E}B(T_n)(B(T) - B(T_n))$$

the first term is clearly non-negative, while the last term we claim to be 0. This is due to the strong Markov property and the first Wald identity for $T - T_n$, which yield:

$$\mathbb{E}(B(T) - B(T_n) \,|\, \mathcal{F}^+(T_n)) = 0.$$

All together we have $\mathbb{E}B(T)^2 \geq \mathbb{E}B(T_n)^2$ for every n, thus

$$\mathbb{E}B(T)^2 \geq \lim_{n\to\infty} \mathbb{E}B(T_n)^2 = \lim_{n\to\infty} \mathbb{E}T_n = \mathbb{E}T.$$

□

Corollary 6 (exit time from an interval.) *For a 1D Brownian motion and $a, b > 0$, let*

$$T_{a,b} = \min(t \geq 0 \,|\, B(t) \in \{-a, b\}).$$

Then:

1. $\mathbf{P}(B(T_{a,b}) = -a) = \frac{b}{a+b}$, $\mathbf{P}(B(T_{a,b}) = b) = \frac{a}{a+b}$.
2. $\mathbb{E}T_{a,b} = a \cdot b$.

Proof Check that $T_{a,b}$ is integrable, and apply Wald's lemma. □

Remark 7 *The condition for the first Wald identity can be weakened to $\mathbb{E}\sqrt{T} < \infty$ (we will not show this). This result is in some sense sharp: $T_1 = \min(t \geq 0 \,|\, B(t) = 1)$ satisfies $\mathbb{E}T_1^{\frac{1}{2}-\epsilon} < \infty$ for $\epsilon > 0$ but $\mathbb{E}B(T_1) = 1 \neq 0$.*

2 Skorohod Embedding Problem

Given a random variable X, is there an integrable stopping time T s.t. $B(T) \sim X$? If so, we may embed a random walk with steps distributed like X into a Brownian motion, step by step.

By Wald's lemma, it is necessary that $\mathbb{E}X = 0$, and $\mathbb{E}X^2 = \mathbb{E}T < \infty$. It turns out to be also a sufficient condition:

Theorem 8 *If X satisfies $\mathbb{E}X = 0$ and $\mathbb{E}X^2 < \infty$, then there exists an integrable stopping time T (w.r.t. $\mathcal{F}^+$) s.t. $B(T) \sim X$.*

Example 9 *If X takes only two values $-a$ and b, we can use $T_{a,b}$.*

There are two approaches to the proof: Dubins embedding (may be found in Durret [1] or Mörters and Peres [2]), and Azéma-Yor embedding, which we will present here.

Theorem 10 *(*Azéma-Yor (AY) embedding.*) Let X satisfy the premises of Theorem 8, and let*

$$\Psi(x) = \left\{ \begin{array}{cc} \mathbb{E}(X \,|\, X \geq x), & \mathbf{P}(X \geq x) > 0 \\ 0, & \textit{otherwise} \end{array} \right\}$$

for a 1D-Brownian motion B, let M be its maximum process: $M(t) = \max_{s \in [0,t]} B(s)$. Define the stopping time

$$T = \inf(t \geq 0 \,|\, M(t) \geq \Psi(B(t))\,)$$

Then T is integrable and $B(T) \sim X$.

intuition: $\Psi^{-1}(x) = \sup(b \,|\, \Psi(b) \leq x)$. We are waiting for the process $\Psi^{-1}(M(t))$ to collide with $B(t)$.

We will prove the theorem for random variables supported on finitely many points. The general case follows by a limiting process. This specific case is covered by the following lemma:

Lemma 11 *Suppose X with $\mathbb{E}X = 0$ takes values $x_1 < x_2 < \cdots < x_n$. Define $y_1 < y_2 < \cdots < y_{n-1}$ by $y_i = \Psi(x_{i+1})$. Define stopping times recursively:*

$$T_0 = 0, \;\; T_i = \inf(t \geq T_{i-1} \,|\, B(t) \notin (x_i, y_i)\,)$$

Then T_{n-1} satisfies $\mathbb{E}T_{n-1} = \mathbb{E}X^2$ and $B(T_{n-1} \sim X$.

Example 12 *Suppose $X \sim Unif\{-2,-1,0,1,2\}$. We begin by waiting for the Brownian motion to exit $(x_1, y_1) = (-2, 1/2)$. Suppose it exists from above. Now we wait till the Brownian motion exists $(x_2, y_2) = (-1, 1)$. Suppose now it existed from below (i.e., $B(T_2) = -1$). The next step (T_3) would be the exist time from $(x_3, y_3) = (0, 1.5)$, but since we are already out of this interval, we set $T_3 = T_2$. For a similar reason $T_4 = T_2$, and our sampled step is $B(T_4) = -1$.*

Proof First note that $y_i \geq x_{i+1}$ with equality if and only if $i = n-1$. Also note that $\mathbb{E}T_{n-1} < \infty$. Hence we only need to show that $B(T_{n-1}) \sim X$, and use Wald's lemma.

Write $Y_i = \left\{ \begin{array}{cc} \mathbb{E}(X \,|\, X \geq x_{i+1}), & X \geq x_{i+1} \\ X & \text{otherwise.} \end{array} \right\}$ Note that Y_1 satisfies $\mathbb{E}Y_1 = 0$ and $Y_1 \in \{x_1, y_1\}$. For $i > 2$ there are two cases:
- If $Y_{i-1} = x_j$ for some $j \leq i-1$, then $Y_i = x_j$.
- If $Y_{i-1} = y_{i-1}$, then $\mathbb{E}(Y_i \,|\, Y_{i-1} = y_i) = y_{i-1}$ and $Y_i \in \{x_i, y_i\}$.

Finally, note that $Y_{n-1} = X$. It follows that

$$(B(T_1), B(T_2), \ldots B(T_{n-1)} \sim (Y_1, Y_2, \ldots Y_{n-1})$$

(i.e., those tuples have the same joint-distribution), since a random variable supported on two values is determined by its expectation. (we have checked that $B(T_1) \sim Y_1$, $B(T_2)|B(T_1) \sim Y_2|Y_1$, etc.) $\square$

Lemma 13 *The stopping time T_{n-1} we constructed above equals the stopping time T of the AY embedding (in Theorem 10).*

Proof Let j be such that

$$B(T_{n-1}) = x_j$$

(the sample in the previous lemma). Note

$$\Psi(B(T_{n-1})) = y_{j-1}$$

by definition of y_{j-1}. If $j \le n-1$, then $T_{j-1} < T_j = T_{j+1} = \cdots = T_{n-1}$, and $B(T_{j-1}) = y_{j-1}$. If $j = n$ then $B(T_{n-1}) = x_n = y_{n-1}$. In both cases $B(T_{j-1}) = y_{j-1}$, hence

$$M(T_{n-1}) \ge y_{j-1} = \Psi(B(T_{n-1})).$$

This means $T \le T_{n-1}$ since T is the first time to pass $\Psi(B(\cdot))$.

Conversely, if $T_{i-1} \le t < T_i$ for some $i \le j$ then $B(t) \in (x_i, y_i)$, and so

$$M(t) < y_i \le \Psi(B(t)) \Rightarrow T \ge T_{n-1}.$$

□

3 The Donsker's Invariance Principle

Let $\{X_n\}_{n\ge 0}$ be a sequence of i.i.d. random variables with finite variance. WLOG, $\mathbb{E}X_1 = 0$ and $\mathrm{var}X_1 = 1$. Let $S_n = \sum_{i=1}^n X_i$. We also let $S(t)$ be its linear interpolation:

$$S(t) = S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)(S_{\lfloor t \rfloor + 1} - S_{\lfloor t \rfloor}).$$

Note $S \in C[0,\infty)$. Define $S_n^*(t) = \frac{S(nt)}{\sqrt{n}}$ for $t \in [0,1]$.

Theorem 14 (Donsker) *S_n^* converges in distribution to $B|_{[0,1]}$ where B is a 1D Brownian motion, in the space $C[0,1]$ (with the sup.-norm).*

Remark 15 *See Portmanteau Theorem for convergence in distribution in Polish spaces. Here we mean that for any bounded continuous function $g : C[0,1] \to \mathbb{R}$, it holds that* $\lim_{n\to\infty} \mathbb{E}g(S_n^*) = \mathbb{E}g(B|_{[0,1]})$.

This is equivalent to: for every bounded measure μ s.t. $\mathbf{P}(\mu$ is discontinuous at $B|_{[0,1]}) = 0$, it holds that $\lim_{n\to\infty} \mathbb{E}\mu(S_n^*) = \mathbb{E}\mu(B|_{[0,1]})$.

This is also equiv. to: for any closed set $K \subset C[0,1]$, $\limsup_{n\to\infty} \mathbf{P}(S_n^* \in K) \le \mathbf{P}(B|_{[0,1]} \in K)$.

Donsker's Theorem follows from

Proposition 16 *Let B be a 1D Brownian motion and X a random variable with $\mathbb{E}X = 0$, $\mathbb{E}X^2 < \infty$. Then there exists a sequence $0 = T_0 \le T_1 \le T_2 \le \dots$ of stopping times for B w.r.t. $\mathcal{F}^+$ s.t. :*

1. *$\{B(T_n)\}_{n\ge 0}$ is a random walk with increments distributed like X.*

2. *Letting S_n^* be constructed from this walk as before, then:*

$$\lim_{n\to\infty} \mathbf{P}\left(\sup_{0\le t\le 1}\left|\frac{B(nt)}{\sqrt{n}} - S_n^*(t)\right| > \epsilon\right) = 0, \;\; \forall \epsilon > 0.$$

The last item roughly means convergence in probability, which is stronger then convergence in distribution.

Proof Let T_1 be the sopping time given by the Skorohod embedding theorem. Define the Brownian motion $B_1(t) := B(t+T_1)-B(T_1)$. Let T_2' be the stopping time of the embedding theorem for B_1; $T_2 := T_1 + T_2'$. Similarly define B_2 (the Brownian motion from time T_2), and T_3' (the stopping time of it according to emb. thm), and $T_3 := T_2 + T_3'$. This gives the T_i of the theorem, where property (1) is clear.

Let $W_n(t) = \frac{B(nt)}{\sqrt{n}}$, which is BM by scale invariance. Define the event

$$A_n = \{\exists t \in [0,1] \text{ s.t. } |S_n^*(t) - W_n(t)| > \epsilon\}.$$

Need to show: $\forall\epsilon$, $\lim_{n\to\infty}\mathbf{P}(A_n) = 0$.

Let $k = k(t)$ be such that $\frac{k-1}{n} \le t < \frac{k}{n}$. Since S_n^* is piecewise linear:

$$A_n \subset \big\{\exists t \in [0,1] \text{ s.t. } |S_{k(t)}(t) - W_n(t)| > \epsilon\big\} \cup \big\{\exists t \in [0,1] \text{ s.t. } |S_{k(t)-1}(t) - W_n(t)| > \epsilon\big\} =$$
$$\left\{\exists t \in [0,1] \text{ s.t. } \left|W_n\left(\frac{T_k}{n}\right) - W_n(t)\right| > \epsilon\right\} \cup \left\{\exists t \in [0,1] \text{ s.t. } \left|W_n\left(\frac{T_{k-1}}{n}\right) - W_n(t)\right| > \epsilon\right\}$$

The equality is true since $S_k = B(T_k) = W_n\left(\frac{T_k}{n}\right)\sqrt{n}$.

Claim For a given $\delta \in (0,1)$, the last event is contained in the following one:

$$\begin{aligned}&\{\exists t,s \in [0,2] \text{ s.t. } |s-t| < \delta,\;\; |W_n(s) - W_n(t)| > \epsilon\} \cup \\ &\left\{\exists t \in [0,1] \text{ s.t. } \left|\frac{T_{k(t)}}{n} - t\right| \vee \left|\frac{T_{k(t)-1}}{n} - t\right| \ge \delta\right\} \\ &:= I_1 \cup I_2\end{aligned}$$

We do not have time to prove the claim (but it should not be difficult).

Roughly, the claim says that the "bad" event A_n is contained in one of two cases: either values of the BM W_n fluctuated more than ϵ in a small interval (I_1), or the stopping times T_i are far apart (I_2).

- For every n, $\mathbf{P}(I_1) \to 0$ as $\delta \downarrow 0$, by equicontinuity of W_n.
- Need: for fixed δ, $\mathbf{P}(I_2) \to 0$ as $n \to \infty$. For this it is enough to show that $\frac{T_n}{n} \to 1$ a.s., which is the law of large numbers (LLN) for the variables $\{T_i\}$. $\square$

Applications:
- Analyzing the maximum of a random walk is similar to analyzing the maximum of BM.
- Deriving of arcsine laws and iterated logarithm law for BM.

References

[1] R. Durret. Probability: Theory and Examples.

[2] Mörters and Peres. Brownian Motion.