Advanced Stochastic Calculus I
Fall 2007
Prof. K. Ramanan*

Chris Almost†

Course website available under Dr. Ramanan's website. These notes were originally compiled during the Fall semester of 2007, with updates made during the Fall semester of 2009.

*kramanan@math.cmu.edu
†cdalmost@cmu.edu

Contents

0 Review of some probability theory

For this course we let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $(S, \mathscr{S})$ be a (sufficiently nice) topological space. The *Borel σ-algebra* $\mathscr{B}(S)$ is the σ-algebra generated by the open sets, $\mathscr{B}(S) = \sigma(\mathscr{S})$.

0.0.1 Definition. $X : (\Omega, \mathscr{F}) \to (S, \mathscr{S})$ is called a *random element* (or an *$\mathscr{F}$-measurable function*) if $X^{-1}(A) \in \mathscr{F}$ for all $A \in \mathscr{S}$.

We will also use the terms *random variable*, *random vector*, *random process*, or *random measure*, as appropriate for the codomain.

0.0.2 Definition. (Comparison of random elements)
(i) If $\mathbb{P}[X = X'] = 1$ then we say that X and X' are *equal a.s.* or *indistinguishable*.
(ii) If $\mathbb{P}[X \in A] = \mathbb{P}[X' \in A]$ for all $A \in \mathscr{S}$ then we say that X and X' are *equal in distribution*.

Equality in distribution can be defined for random elements defined on unequal probability spaces (but they must have the same codomain).

0.0.3 Example. Let $\Omega = \{H, T\}$ and $\mathbb{P}$ be the uniform probability (i.e. Bernoulli with parameter $\frac{1}{2}$). Let $X(H) = 0 = X'(T)$ and $X(T) = 1 = X'(H)$. Then $\mathbb{P}[X = 1] = \mathbb{P}[X' = 1]$ so they are equal in distribution, but they are not equal a.s., in fact they are unequal a.s.

0.0.4 Definition. (Convergence of random variables)
(i) X_n *converges $\mathbb{P}$-a.s.* to X if $\mathbb{P}[\limsup_n |X_n - X| > 0] = 0$.
(ii) X_n *converges in probability* to X (or $X_n \xrightarrow{(p)} X$) if

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| \geq \varepsilon] = 0$$

for all $\varepsilon > 0$.
(iii) X_n *converges in distribution* to X (or $X_n \xrightarrow{(d)} X$) if

$$\mathbb{P}[X_n \in A] \to \mathbb{P}[X \in A]$$

for all A such that $\mathbb{P}[X \in \partial A] = 0$. Equivalently, $X_n \xrightarrow{(d)} X$ if the distribution functions of X_n converge pointwise to the distribution function of X at ever point of continuity of that function. Equivalently, $X_n \xrightarrow{(d)} X$ if

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$$

for all bounded continuous functions f.

Remark. When trying to prove a sequence of random variables converges in probability, *Markov's inequality* is often very useful.

Now let us recall some important theorems. Let X_i, $i \in \mathbb{N}$ be i.i.d. r.v.'s such that $\mu = \mathbb{E}|X_i| < \infty$ and let $S_n = \sum_{i=1}^n X_i$. Khintchine's *weak law of large numbers* says that

$$\frac{S_n}{n} \xrightarrow{(p)} \mu,$$

and Kolmogorov's *strong law of large numbers* says that

$$\frac{S_n}{n} \to \mu \ \mathbb{P}\text{-a.s.}$$

If $\sigma^2 = \mathbb{E}[|X_1 - \mu|^2] < \infty$ (i.e. the r.v.'s have finite variance) then the *central limit theorem* states

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{(d)} \mathcal{N}(0,1),$$

where of course the right hand side may be replaced by any standard Gaussian random element. The appearance in this theorem of the normal distribution is a big part of why this distribution comes up everywhere.

1 Brownian Motion

1.1 Introduction to stochastic processes

1.1.1 Definition. A *stochastic process* on $(\Omega, \mathscr{F}, \mathbb{P})$ with *state space* $(S, \mathscr{S})$ is (equivalently) a

(i) (one-parameter) family of random variables, $\{X_t : \Omega \to S \mid t \in [0,\infty)\}$;

(ii) random element of $\mathbb{R}^{[0,\infty)}$, $X = \{X_{(\cdot)}(\omega) : [0,\infty) \to S \mid \omega \in \Omega\}$;

(iii) random element of two variables $X : [0,\infty) \times \Omega \to S$.

Notice that the concept of "measurability" would seem *a priori* to be different for each of the three definitions.

1.1.2 Definition. A stochastic process is said to be *measurable* if X is measurable as a random element of two variables, i.e. if for all $A \in \mathscr{S}$,

$$\{(t,\omega) \mid X(t,\omega) \in A\} \in \mathscr{B}[0,\infty) \times \mathscr{F}.$$

Warning: In K&S the authors write $\mathscr{B}(\mathbb{R}^{[0,\infty)}) = \bigotimes_{[0,\infty)} \mathscr{B}(\mathbb{R})$, which may or may not be true according to our definitions. (Think about this?)

1.1.3 Definition. (Comparison of random processes) Let X and X' be random processes.

(i) X and X' are *indistinguishable* if $\mathbb{P}[X_t = X'_t \text{ for all } t] = 1$.

(ii) X is said to be a *modification* of X' if $\mathbb{P}[X_t = X'_t] = 1$ for all t.

(iii) If $\mathbb{P}[X \in A] = \mathbb{P}[X' \in A]$ for all $A \in \mathscr{B}(\mathbb{R}^{[0,\infty)})$ then X and X' are *equal in distribution*.

(iv) If for every $t_1, \dots, t_n$ then we have

$$\mathbb{P}[(X_{t_1}, \dots, X_{t_n}) \in A] = \mathbb{P}[(X'_{t_1}, \dots, X'_{t_n}) \in A]$$

for all $A \in \mathscr{B}(\mathbb{R}^n)$ they X and X' are said to have the same *finite dimensional marginal distributions* (or *fi.di. distributions*).

As before, the definitions of equality in distribution and having the same fi.di. distributions can be extended to processes defined on different probability spaces.

1.1.4 Definition. A sequence of random elements X_n *converges weakly* to X (or $X_n \xrightarrow{(w)} X$) if for every bounded continuous function $f : S \to \mathbb{R}$ then $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$.

1.2 Construction of Brownian motion

For reasons that will become clear in the section on Donsker's invariance principle, we look for a process $B = \{B_t, t \geq 0\}$ with the following properties.

(i) $B_0 = 0$ and $B_t \sim \mathcal{N}(0, t)$ for all t;

(ii) B has *stationary increments* (or *homogeneous increments*), i.e. $B_t - B_s$ has the same distribution as B_{t-s} for all $s < t$;

(iii) B has *independent increments*, i.e. for $s < t$, $B_t - B_s$ is independent of $\{B_u \mid 0 \leq u \leq s\}$; and

(iv) B has continuous paths.

The *canonical version* of a random element $X : (\Omega, \mathscr{F}, \mathbb{P}) \to (S, \mathscr{S})$ is constructed as follows. The *image measure* $\mathbb{P}X^{-1}$ of X is defined by $\mathbb{P}X^{-1}(A) = \mathbb{P}[X \in A]$ for $A \in \mathscr{B}(S)$. There is always a random element defined on $(S, \mathscr{S}, \mathbb{P}X^{-1})$ with the same distribution as X, namely the identity function. Thus our goal is to construct a measure on $(C[0,1], \mathscr{B}(C[0,1]))$ that satisfies properties 1–3.

1.2.1 Lemma. *Given $0 \leq t_1 < t_2 < t_3$, any process B that satisfies properties 1–3 must have the following joint probability density function.*

$$f_{B_{t_1}, B_{t_2}, B_{t_3}}(x, y, z) = p(t_1; 0, x) p(t_2 - t_1; x, y) p(t_3 - t_2; y, z),$$

where $p(t; x, y) := \frac{1}{\sqrt{2\pi t}} \exp(-\frac{(y-x)^2}{2t})$.

This lemma extends easily to any finite number of times and it is seen that properties 1–3 determine the fi.di. distributions of B.

To use Carathéodory's extension theorem to construct a measure on $(G, \mathscr{G})$ one must:

(i) Define a finitely additive set function μ_0 on some algebra or π-system $\mathscr{C} \subseteq \mathscr{G}$.

(ii) Show that μ_0 is countably additive on $\mathscr{C}$.

(iii) Carathéodory's theorem allows us to conclude that μ_0 may be extended as a measure to the completion of $\sigma(\mathscr{C}) \subseteq \mathscr{G}$.

One could also start with a reference measure and define another measure via a density function. Or one could define a measure as a "limit" of "simple" measures.

Construction of Brownian motion: method 1

Let I be the space of finite increasing sequences of times

$$I := \{(t_1, \dots, t_n) \mid n \in \mathbb{N}, 0 \le t_1 < \cdots < t_n < \infty\}.$$

From the lemma, it is clear that $\mathbb{P}$ has finite dimensional distributions given by

$$Q_{\underline{t}}(A) := \mathbb{P}\{\omega \in C[0,\infty) \mid (\omega_{t_1}, \dots, \omega_{t_n}) \in A\}$$

for all $A \in \mathscr{B}(\mathbb{R}^n)$, where

$$Q_{\underline{t}}(A) = \int_A p(t_1; 0, x_1) \cdots p(t_n - t_{n-1}; x_{n-1}, x_n) dx_1 \cdots dx_n.$$

This method will yield a unique measure on $(C[0,\infty), \mathscr{B}(C[0,\infty)))$, *Wiener measure*, with fi.di. distributions given by $\{Q_{\underline{t}}, \underline{t} \in I\}$. (But I don't see how, yet.)

Let $\mathscr{C}'$ be the set containing each so-called *cylinder set*, sets of the form

$$\{\omega \in C[0,\infty) \mid (\omega_{t_1}, \dots, \omega_{t_n}) \in A\}$$

for $A \in \mathscr{B}(\mathbb{R}^n)$, for $n \in \mathbb{N}$. We have a finitely additive set function $Q_{\underline{t}}$ defined on $\mathscr{C}'$. As an exercise, prove that the collection $\{Q_{\underline{t}} \mid \underline{t} \in I\}$ is consistent (see below). The next step is to show that $Q_{\underline{t}}$ is countably additive on $\mathscr{C}'$ (see Itô-McKean), and the final step is to show that $\sigma(\mathscr{C}') = \mathscr{B}(C[0,\infty))$. (I must be missing something here.)

The canonical version of Brownian motion is defined to be the canonical process on $(C[0,\infty), \mathscr{B}(C[0,\infty)), \mathbb{P})$, where $\mathbb{P}$ is Wiener measure.

Construction of Brownian motion: method 2

1.2.2 Theorem (Daniell-Kolmogorov). *Let $\{Q_{\underline{t}}(\cdot), \underline{t} \in I\}$ be a family of finite dimensional distributions that satisfies*

(i) $Q_{\underline{t}}(A)$ is invariant under permutation of the elements of $\underline{t}$; and

(ii) For any $\underline{t} = (t_1, \dots, t_n)$, if $\underline{s} = (t_1, \dots, t_{n-1})$ and $A \in \mathscr{B}(\mathbb{R}^{n-1})$ then $Q_{\underline{t}}(A \times \mathbb{R}) = Q_{\underline{s}}(A)$. This is the requirement that the family is consistent.

Then there is a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^{[0,\infty)}, \mathscr{B}(\mathbb{R}^{[0,\infty)}))$ such that

$$\mathbb{P}(\omega \in \mathbb{R}^{[0,\infty)} \mid (\omega_{t_1}, \dots, \omega_{t_n}) \in A\} = Q_{\underline{t}}(A)$$

for all $A \in \mathscr{B}(\mathbb{R}^n)$ and $\underline{t} \in I$.

Let $\tilde{\mathscr{C}}$ be the collection of cylinder sets

$$\{\omega \in \mathbb{R}^{[0,\infty)} \mid (\omega_{t_1}, \dots, \omega_{t_n}) \in A\}$$

for $A \in \mathscr{B}(\mathbb{R}^n)$, for $n \in \mathbb{N}$, from $\mathbb{R}^{[0,\infty)}$. As in method 1, $Q_{\underline{t}}$ is defined on $\tilde{\mathscr{C}}$ and it is finitely additive and consistent. It is countably additive by the Daniell-Kolmogorov extension theorem. Carathéodory's theorem gives a measure satisfying properties 1–3 on $\sigma(\tilde{\mathscr{C}})$. The only thing remaining is to deal with the continuity of the paths.

1.2.3 Theorem (Kolmogorov-Čentsov). *If $\{X_t \mid t \in [0,T]\}$ is a real valued stochastic process defined on $(\Omega, \mathscr{F}, \mathbb{P})$ that satisfies*

$$\mathbb{E}[|X_t - X_s|^\alpha] \le C|t-s|^{1+\beta}$$

for all $0 \le s \le t \le T$ and some positive constants α, β, and C, then there is a continuous modification of X which is locally Hölder continuous with exponent γ, for all $\gamma \in (0, \frac{\beta}{\alpha})$.

A stochastic process $\{X_t \mid t \in [0,T]\}$ is said to be *locally Hölder continuous* with exponent γ if an a.s. positive r.v. h such that

$$\mathbb{P}\left(\sup_{\substack{0<t-s<h \\ s,t\in[0,T]}} \frac{|X_t - X_s|}{|t-s|^\gamma} \le \delta \right) = 1$$

for some appropriate constant $\delta > 0$.

Proof: The first step is to choose a countable dense subset of $[0,T]$. We use the *dyadic rationals*

$$D = \{k2^{-n} \mid k = 0, \dots, 2^{n-1}, n \in \mathbb{N}\}.$$

Define $\tilde{\Omega}^* = \{\omega \mid t \mapsto X_t(\omega) \text{ is uniformly continuous}\}$, where X is the canonical process on $\mathbb{R}^{[0,\infty)}$. We would like to show that $\mathbb{P}(\Omega^*) = 1$. We will show instead that $\mathbb{P}(\Omega^*) = 1$, where

$$\Omega^* = \{\omega \mid t \mapsto X_t(\omega) \text{ is locally Hölder continuous on } D \text{ with coefficient } \gamma\}.$$

By definition, $\omega \in \Omega^*$ holds if there is $n^*(\omega)$ such that

$$\max_{1\le k\le 2^n} |X_{k2^{-n}}(\omega) - X_{(k-1)2^{-n}}(\omega)| < 2^{-n\gamma}$$

for all $n \ge n^*(\omega)$. Call this property (P). Let

$$E_n = \{\omega \mid \max_{1\le k\le 2^n} |X_{k2^{-n}}(\omega) - X_{(k-1)2^{-n}}(\omega)| \ge 2^{-n\gamma}\}.$$

The set of ω for which property (P) does not hold is the set of ω for which E_n occurs infinitely often. Whence

$$\Omega \setminus \Omega^* = \bigcap_{m\in\mathbb{N}} \bigcup_{n\ge m} E_n.$$

Now

$$\begin{aligned}
\mathbb{P}(E_n) &= \mathbb{P}\left(\bigcup_{k=1}^{2^n} |X_{k2^{-n}} - X_{(k-1)2^{-n}}| \ge 2^{-n\gamma} \right) \\
&\le \sum_{k=1}^{2^n} \mathbb{P}(|X_{k2^{-n}} - X_{(k-1)2^{-n}}|^\alpha \ge 2^{-n\gamma\alpha})
\end{aligned}$$

$$\leq 2^{n\gamma\alpha} \sum_{k=1}^{2^n} \mathbb{E}(|X_{k2^{-n}} - X_{(k-1)2^{-n}}|^\alpha)$$
$$\leq C 2^{n\gamma\alpha} \sum_{k=1}^{2^n} \left(\frac{1}{2^n}\right)^{1+\beta}$$
$$\leq 2^{(\gamma\alpha-\beta)n}$$

so $\sum_{n=1}^\infty \mathbb{P}(E_n) < \infty$ if $\gamma \in (0, \frac{\beta}{\alpha})$. Therefore by the Borel-Cantelli Lemma we have $\mathbb{P}(\Omega^*) = 1$.

The next step is to define the modification. We define

$$\tilde{X}(\omega) = \begin{cases} X_t(\omega) & t \in D, \omega \in \Omega^* \\ \lim_{\substack{s_n \to t \\ \{s_n\} \subseteq D}} X_{s_n}(\omega) & t \notin D, \omega \in \Omega^* \\ 0 & \omega \notin \Omega^* \end{cases}$$

and we must show that it is truly a modification of X. For $t \in D$, $\mathbb{P}(X_t = \tilde{X}_t) = \mathbb{P}(\Omega^*) = 1$. For $t \notin D$, we know by construction that $\tilde{X}_{s_n} \to \tilde{X}_t$ a.s. for any $\{s_n\} \subseteq D$ converging to t. We know by K-C inequality that $X_{s_n} \to X_t$ in probability. Therefore $\tilde{X}_{s_n} \to X_t$ in probability, and so $\mathbb{P}(X_t = \tilde{X}_t) = 1$. □

To complete the construction we must check the K-C inequality holds for Brownian motion for some α and β.

Weak convergence

There are a few books on convergence of processes: Billingsley (but not *Probability*), Parthasarathy, Jacod-Shiryaev.

For this section we let $(S, \mathscr{S})$ be a metrizable space.

1.2.4 Definition. A sequence of measures μ_n *converges weakly* to μ, all defined on the same space $(S, \mathscr{B}(S))$, denoted $\mu_n \xrightarrow{(w)} \mu$, if

$$\mathbb{E}_{\mu_n}[f] = \int_S f(x)\mu_n(dx) \to \int_S f(x)\mu(dx) = \mathbb{E}_\mu[f],$$

for all $f \in C_b(S)$.

It can be seen that when $S = \mathbb{R}$ this reduces to convergence in distribution.

A norm that may be defined on a collection of measures absolutely continuous with respect to some measure λ is the *total variation norm* defined by

$$||\mu||_{\mathrm{TV}} = \int_S \left|\frac{d\mu}{d\lambda}\right| d\lambda.$$

1.2.5 Exercise. Suppose that μ_n is the distribution of

$$\frac{S_n}{\sqrt{n}} = \frac{\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])}{\sqrt{n}},$$

where the X_i are i.i.d. with finite means and variances. Does μ_n converge in the total variation metric?

1.2.6 Definition. Given a sequence of probability measures on $(S, \mathscr{B}(S))$ and a probability measure μ on $(S, \mathscr{B}(S))$, μ_n converges weakly to μ if and only if

$$\lim_{n\to\infty} \int_S f\, d\mu_n = \int_S f\, d\mu$$

for every bounded continuous function f on S. We write $\mu_n \xrightarrow{(w)} \mu$.

This is the weak* topology induced from considering the space of finitely additive measures as the dual of the space of bounded measurable functions. Note that

$$\|\mu\|_{\text{TV}} = \sup_{\substack{f\in L^\infty(S)\\ \|f\|_\infty \le 1}} \left| \int_S f\, d\mu \right|,$$

so the total variation norm is the operator norm induced by this duality.

1.2.7 Theorem (Portmanteau). *The following are equivalent.*

(i) $\mu_n \xrightarrow{(w)} \mu$*;*
(ii) $\liminf_{n\to\infty} \mu_n(U) \ge \mu(U)$ *for every open set* U*;*
(iii) $\limsup_{n\to\infty} \mu_n(C) \le \mu(C)$ *for every closed set* C*;*
(iv) $\lim_{n\to\infty} \mu_n(A) = \mu(A)$ *for every measurable set* A *such that* $\mu(A) = 0$.

1.2.8 Continuous mapping theorem. *If* $\Phi : S \to S'$ *is continuous and* $\mu_n \xrightarrow{(w)} \mu$ *in* $(S, \mathscr{B}(S))$ *then* $\Phi(\mu_n) \xrightarrow{(w)} \Phi(\mu)$ *in* $(S', \mathscr{B}(S'))$.

PROOF: Let $g \in C_b(S')$. Then $g \circ \Phi \in C_b(S)$, so by definition

$$\int_{S'} g\, d\Phi(\mu_n) = \int_S g \circ \Phi\, d\mu_n \to \int_S g \circ \Phi\, d\mu = \int_{S'} g\, d\Phi(\mu),$$

so $\Phi(\mu_n) \xrightarrow{(w)} \Phi(\mu)$. □

1.2.9 Definition. Let Π be a family of probability measures on $(S, \mathscr{B}(S))$.

(i) Π is said to be *relatively compact* if every sequence in Π contains a weakly convergent subsequence. (This is just relative compactness with respect to the topology induced by weak convergence).

(ii) Π is *tight* if for every $\varepsilon > 0$ there is a compact set $K \subseteq S$ such that $P(K) \geq 1 - \varepsilon$ for all $P \in \Pi$.

1.2.10 Theorem (Prohorov). *Suppose that $(S, \mathscr{S})$ is a Polish space (i.e. a complete, separable, metrizable space). Then a family Π of probability measures on $(S, \mathscr{B}(S))$ is tight if and only if Π is relatively compact.*

We will be interested in the case when S is the space of continuous functions on $[0, \infty)$ with the metric ρ associated with the uniform norm

$$\rho(f, g) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\sup_{t \in [0,n]} |f(t) - g(t)|}{1 + \sup_{t \in [0,n]} |f(t) - g(t)|}.$$

$(C[0, \infty), \rho)$ is a Polish space.

1.2.11 Theorem (Arzela-Ascoli). *A set $A \subseteq C[0, \infty)$ is relatively compact if and only if the following two conditions hold.*

(i) $\sup_{\omega \in A} |\omega(0)| < \infty$*; and*

(ii) $\lim_{\delta \searrow 0} \sup_{\omega \in A} m^T(\omega, \delta) = 0$ *for all $T < \infty$, where*

$$m^T(\omega, \delta) = \sup_{\substack{0 \leq s \leq t \leq T \\ |t-s| \leq \delta}} |\omega(t) - \omega(s)|$$

is the modulus of continuity. (A is said to be equicontinuous).

1.2.12 Theorem. *A sequence $\{\mathbb{P}_n\}$ of probability measures on $C[0, \infty)$ is tight if and only if*

(i) $\lim_{\lambda \nearrow \infty} \sup_{n \geq 1} \mathbb{P}_n\{\omega \mid |\omega(0)| \geq \lambda\} = 0$*; and*

(ii) $\lim_{\delta \searrow 0} \sup_{n \geq 1} \mathbb{P}_n\{\omega \mid m^T(\omega, \delta) > \varepsilon\} = 0$ *for all $\varepsilon > 0$ and all $T < \infty$.*

PROOF: Suppose that $\{\mathbb{P}_n\}$ is tight. Then given any $\eta > 0$ there is some compact set $K_\eta \subseteq C[0, \infty)$ such that $\mathbb{P}_n(K_\eta) > 1 - \eta$. By the Arzela-Ascoli theorem, given $T > 0$, $\varepsilon > 0$ there is $\lambda < \infty$ and $\delta_0 > 0$ such that

$$K = \{\omega \mid |\omega(0)| \leq \lambda, m^T(\omega, \delta) \leq \varepsilon \text{ for } \delta \in (0, \delta_0)\}.$$

Now suppose that $\{\mathbb{P}_n\}$ satisfies the conditions. Then given $T > 0$ and $\eta > 0$, choose $\lambda > 0$ such that

$$\sup_{n \geq 1} \mathbb{P}_n\{\omega \mid |\omega(0)| > \lambda\} \leq \frac{\eta}{2^{T+1}}$$

Choose $\delta_k > 0$ such that for each $k = 1, 2, \ldots$

$$\sup_{n \geq 1} \mathbb{P}_n\{\omega \mid m^T(\omega, \delta_k) > \tfrac{1}{k}\} \leq \frac{\eta}{2^{T+k+1}}.$$

Define

$$A^T := \{\omega \mid |\omega(0)| \leq \lambda, m^T(\omega, \delta_k) \leq \tfrac{1}{k} \text{ for all } k = 1, 2, \ldots\}$$

and let $A := \bigcap_{T=1}^{\infty} A_T$. Then

$$\mathbb{P}_n(A_T) \geq 1 - \sum_{k=0}^{\infty} \frac{\eta}{2^{T+k+1}} = 1 - \frac{\eta}{2^T}$$

so $\mathbb{P}_n(A) \geq 1-\eta$ for all $n \geq 1$. By the Arzela-Ascoli theorem A is relatively compact, so $\{\mathbb{P}_n\}$ is tight. □

It turns out that the topology on $M_1(S)$ induced by weak convergence is metrizable, and $M_1(S)$ with this so-called *Prohorov metric* is complete and separable.

Another important fact is that we will use often is that probability measures on $(C[0,\infty), \mathscr{B}(C[0,\infty)))$ are uniquely characterized by their finite dimensional distributions, i.e. their values on cylinders

$$\{\omega \in [0,\infty) \mid (\omega_{t_1}, \dots, \omega_{t_n}) \in A\},$$

for $A \in \mathscr{B}(\mathbb{R}^n)$, for $n \in \mathbb{N}$.

Donsker's invariance principle

Let $\xi_1, \xi_2, \dots$ be i.i.d. r.v.'s with mean zero and variance one. Let $S_n = \sum_{i=1}^n \xi_i$. Recall the *central limit theorem*, that $S_n/\sqrt{n} \xrightarrow{(w)} N(0,1)$. Following KS, let

$$Y_t := S_{\lfloor t \rfloor} + (t - \lfloor t \rfloor)\xi_{\lfloor t \rfloor + 1}$$

be the continuous time process that is the linear interpolation between the partial sums. For each $n \geq 1$, scale Y in space by a factor of $\sqrt{n}$ and in time by a factor of n (the choice of these scaling factors will become clear in a moment) to get a process $\{X_t^{(n)} \mid t \in [0,1]\}$,

$$X_t^{(n)} = \frac{1}{\sqrt{n}} Y_{nt} = \frac{1}{\sqrt{n}}(S_{\lfloor nt \rfloor} + \{nt\}\xi_{\lfloor nt \rfloor + 1}).$$

Notice that for $s = k/n$ and $t = (k+1)/n$ we have $X_t^{(n)} - X_s^{(n)} = \xi_{k+1}/\sqrt{n}$ which is independent of $\sigma(X_u^{(n)} \mid u \leq s) = \sigma(\xi_1, \dots, \xi_k)$ and it has mean zero and variance 1. At $t = 1$ we have $X_1^{(n)} = S_n/\sqrt{n}$, which converges weakly to $\mathcal{N}(0,1)$. For $t = \frac{1}{2}$ we have (approximately)

$$X_{\frac{1}{2}}^{(n)} = \frac{1}{\sqrt{n}} S_{\lfloor \frac{n}{2} \rfloor} = \frac{1}{\sqrt{2}} \frac{S_{\lfloor \frac{n}{2} \rfloor}}{\sqrt{\frac{n}{2}}}$$

which is $\mathcal{N}(0, \frac{1}{2})$ in the limit, by the CLT. These computations lead us to believe that the $X_t^{(n)}$ "converge" to Brownian motion. This is made precise below.

1.2.13 Theorem. $\{X^{(n)}\}$ *converges weakly to Brownian motion.*

First we show that $\{X^{(n)}\}$ is tight. From Exercise 4.11 of KS, if $\{X^{(n)}\}$ is a sequence of continuous stochastic processes with $X_0^{(n)} = 0$ and if

$$\sup_{n\geq 1} \mathbb{E}[|X_t^{(n)} - X_s^{(n)}|^\alpha] \leq C_T |t-s|^{1+\beta}$$

for all $T > 0$ and $0 \leq s, t \leq T$ for some constants α, β, and C_T, then $\mathbb{P}_n = \mathbb{P}(X^{(n)})^{-1}$ form a tight sequence. To show that $\{X^{(n)}\}$ is tight we will show that it satisfies the conditions of this problem.

The next step is to apply Prohorov's theorem to see that $\{X^{(n)}\}$ is relatively compact. Thus there is a weakly convergent subsequence $\{X^{(n_k)}\}$. Let X be such that $X^{(n_k)} \xrightarrow{(w)} X$ and apply the continuous mapping theorem and central limit theorem to conclude that X has the required fi.di. distributions.

1.3 Sample path properties

1.3.1 Proposition. *If B is a Brownian motion then the following processes X are also Brownian motions with respect to their natural filtrations.*

(i) $X_t := \frac{1}{\sqrt{c}} B_{ct}$ *for $c > 0$ (scaling property)*
(ii) $X_t := B_{t+c} - B_c$ *for $c \geq 0$ (simple Markov property)*
(iii) $X_t := B_T - B_{T-t}$ *for $t \in [0, T]$ (time reversal property)*
(iv) $X_t := tB_{\frac{1}{t}}$ *for $t > 0$ and $X_0 := 0$ (time inversion property)*
(v) $X_t := UB_t$ *for an orthogonal matrix U (where B is d-dimensional Brownian motion). In particular we have the reflection property, that $-B$ is a Brownian motion.*

1.3.2 Lemma. $\mathbb{P}[\sup_{t\geq 0} B_t = \infty$ *and* $\inf_{t\geq 0} B_t = -\infty] = 1$.

Proof: Let $Z = \sup_{t\geq 0} B_t$. For any $c > 0$ we have

$$cZ = \sup_{t\geq 0} cB_t \stackrel{(d)}{=} \sup_{t\geq 0} B_{\frac{t}{c^2}} = Z.$$

Therefore the law of Z is concentrated on $\{0, \infty\}$. Let $p = \mathbb{P}(Z = 0)$. Then

$$\begin{aligned}
p &\leq \mathbb{P}[B_1 \leq 0 \text{ and } \sup_{t\geq 0} B_{1+t} - B_1 < \infty] \\
&= \mathbb{P}[B_1 \leq 0]\, \mathbb{P}[\sup_{t\geq 0} B_{1+t} - B_1 < \infty] && B_1 \perp (B_{1+t} - B_1, t \geq 0) \\
&= \mathbb{P}(B_1 \leq 0)\, \mathbb{P}(Z = 0) && (B_{1+t} - B_1, t \geq 0) \text{ is a BM} \\
&= \tfrac{1}{2} p
\end{aligned}$$

so $p = 0$ and $\mathbb{P}(Z = \infty) = 1$. □

Remark. A direct consequence of this is that a.s. for all $a \in \mathbb{R}$, $\{t \mid B_t = a\}$ is not bounded above.

1.3.3 Lemma. *Brownian motion is a.s. not differentiable at zero.*

Proof: The last lemma and time inversion together imply that

$$\mathbb{P}[\forall\, \varepsilon > 0, \exists\, s, t \leq \varepsilon \text{ s.t. } B_s < 0 < B_t] = 1.$$

Indeed, if this were not the case then there would be a set A with $\mathbb{P}[A] > 0$ and the property that for all $\omega \in A$ there is $\varepsilon = \varepsilon(\omega)$ such that either $B(u) > 0$ or $B(u) < 0$ for all $u \in (0, \varepsilon]$. By the time inversion property this implies that for all $\omega \in A$, $\tilde{B}_u = uB_{\frac{1}{u}}$ satisfies $\tilde{B}_u > 0$ or $\tilde{B}_u < 0$ for all $s \in [\frac{1}{\varepsilon}, \infty)$, which contradicts the previous lemma.

Therefore the only possible (right) derivative of Brownian motion at zero is 0. If this were the case on a set A of positive probability, then then for $\omega \in A$, $|B_t(\omega)| \leq t$ for all $0 \leq t \leq T(\omega)$. Once again, using time inversion, $\tilde{B}_t := tB_{\frac{1}{t}}$ is a Brownian motion. On A, for all $0 < t \leq T(\omega)$, $\tilde{B}_{\frac{1}{t}} = \frac{B_t}{t} \leq 1$, which is impossible. □

1.3.4 Lemma. *Brownian paths are monotone on no interval, a.s.*

Proof: We must show that the set $\bigcup_{s,t\in\mathbb{Q}}\{\omega \mid B(\omega) \text{ is monotone on } [s,t]\}$ has probability zero. By the symmetry properties of Brownian motion it suffices to show that $A := \{\omega \mid B(\omega) \text{ is non-decreasing on } [0,1]\}$ has probability zero. Let $A_n := \bigcap_{i=0}^{n-1}\{B_{\frac{i+1}{n}} - B_{\frac{i}{n}} \geq 0\}$ and notice that $A = \bigcap_{n=1}^{\infty} A_n$ since B has continuous paths. (It follows in particular that A is measurable.) Since B has independent, normally distributed increments $\mathbb{P}[A_n] = (\frac{1}{2})^n$, so by continuity of measure, $\mathbb{P}[A] = \lim_{n\to\infty} \mathbb{P}[A_n] = 0$. □

Given a partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_{k_\Pi} = 1\}$ of $[0,1]$, and a real-valued function $f : [0,1] \to \mathbb{R}$, let

$$V^{(p)}(\Pi)(f) = \sum_{i=1}^{k_\Pi} |f(t_i) - f(t_{i-1})|^p.$$

The (classical) *p-variation* of f is defined to be

$$\widetilde{V}^{(p)}(f) = \sup_{\Pi} V^{(p)}(\Pi)(f).$$

If f is continuous and $p = 1$ then $V^{(p)}(f) = \lim_{\|\Pi_n\|\to 0} V^{(p)}(\Pi_n)(f)$ where Π_n is any sequence of partitions such that $\|\Pi_n\| \to 0$ (i.e. the mesh size goes to zero). When $p \neq 1$, these quantities need not be the same. Regardless, the p-variation of a stochastic process X is defined to be

$$V^{(p)}(X) = \lim_{\|\Pi_n\|\to 0} V^{(p)}(\Pi_n)(X)$$

where the limit is taken in probability.

1.3.5 Lemma. *The quadratic ($p = 2$) variation of Brownian motion on the interval $[0, t]$ is the deterministic value t, and Brownian motion does not have finite variation ($p = 1$) on any interval.*

Proof: Fix t and fix a partition $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ of $[0, t]$. Then

$$\mathbb{E}(V^{(2)}(\Pi)(B) - t)^2 = \mathbb{E}\left(\sum_{j=1}^{n} \Delta_j\right)^2$$

where $\Delta_j = (B_{t_j} - B_{t_{j-1}})^2 - (t_j - t_{j-1})^2$. It can be shown that $\mathbb{E}[\Delta_j \Delta_k] = 0$ for $j \neq k$ and $\mathbb{E}[\Delta_j^2] = 2(t_j - t_{j-1})^2$. Whence

$$\mathbb{E}(V^{(2)}(\Pi)(B) - t)^2 = 2\sum_{j=1}^{n}(t_j - t_{j-1})^2 \le \|\Pi\| t,$$

so $V^{(2)}(\Pi)(B)$ converges to t in $L^2(0, t)$ (and hence in probability). It follows that Brownian motion cannot have finite first variation because it has continuous paths. □

1.3.6 Sample path properties. *The following properties are true of a.e. sample path of Brownian motion.*

(i) Unboundedness.
(ii) Of unbounded first variation. (This is a consequence of the fact that $V^{(2)}(B(\omega)) \neq 0$ a.s.)
(iii) Non-differentiable at zero (or anywhere).
(iv) Monotone on no interval. (This is a consequence of (ii))
(v) Nowhere differentiable, i.e.

$$\{\omega \mid \forall\ t \in [0, \infty) \text{ either } D^+ W_t(\omega) = \infty \text{ or } D_- W_t(\omega) = -\infty\}$$

contains an event F with $\mathbb{P}(F) = 1$, where

$$D^+ f_t = \limsup_{h \to 0} \frac{1}{h}(f(t+h) - f(h))$$

and

$$D^- f_t = \liminf_{h \to 0} \frac{1}{h}(f(t+h) - f(h)).$$

(vi) Law of the Iterated Logarithm:

$$\limsup_{t \to \infty} \frac{W_t(\omega)}{\sqrt{2t \log\log t}} = 1 \quad \text{and} \quad \liminf_{t \to \infty} \frac{W_t(\omega)}{\sqrt{2t \log\log t}} = -1$$

(vii) Exact modulus of continuity (Lévy), see 2.9.F.

1.4 Distributional properties

1.4.1 Definition. A stochastic process X is a *Gaussian process* if for every $0 < t_1 < t_2 \cdots < \infty$ the $\mathbb{R}^n$ valued random vector $(X_{t_1}, \dots, X_{t_n})$ has a (multi-variate) Gaussian distribution. The *covariance function* of a Gaussian process is defined to be $\rho(s,t) := \mathbb{E}[(X_s - \mathbb{E}[X_s])(X_t - \mathbb{E}[X_t])]$.

Brownian motion is a Gaussian process mean zero and covariance function $\rho(s,t) = s \wedge t$.

1.4.2 Lemma. *The CDF* $F_{t_1,\dots,t_n}(x_1,\dots,x_n)$ *of* $(B_{t_1},\dots,B_{t_n})$ *is*

$$\int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} p(t_1;0,y)p(t_2-t_1;y_1,y_2)\cdots p(t_n-t_{n-1};y_{n-1},y_n)dy_n\cdots dy_1.$$

1.5 Markov property

Informally, X is a *Markov process* if there is a family of Borel measurable functions $\{f_{s,t}\}$ such that

$$\mathbb{P}(X_t \in A \mid \mathscr{F}_s^X) = f_{s,t}(X_s, A).$$

1.5.1 Definition. Let $(\Omega, \mathscr{F})$ be a measurable space. A *kernel* on Ω is a map $N : \Omega \times \mathscr{F} \to [0,1]$ such that
(i) $A \mapsto N(x,A)$ is a probability measure on $(\Omega,\mathscr{F})$ for all $x \in \mathscr{F}$; and
(ii) $x \mapsto N(x,A)$ is $\mathscr{F}$-measurable for every $A \in \mathscr{F}$.
A kernel N is called a *transition probability* or *stochastic kernel* if $N(x,\Omega) = 1$ for all $x \in \Omega$.

Notation. If f is a non-negative $\mathscr{F}$-measurable function and N is a kernel then the function

$$Nf(x) := \int_\Omega N(x,dy)f(y) = \mathbb{E}_{N(x,\cdot)}[f].$$

Likewise, if M and N are two kernels then

$$MN(x,A) = \int_\Omega M(x,dy)N(y,A).$$

Suppose there is a process X for which, for any $s < t$ there is a transition probability $P_{s,t}$ such that a.s.

$$\mathbb{P}(X_t \in A \mid \mathscr{F}_s^X) = P_{s,t}(X_s, A).$$

Then for any positive $\mathscr{F}$-measurable function f, using standard approximation arguments,

$$\mathbb{E}[f(X_t) \mid \mathscr{F}_s^X] = P_{s,t}f(X_s).$$

So if $s < t < v$ then

$$\mathbb{P}(X_v \in A \mid \mathscr{F}_s^X) = \mathbb{E}[\mathbb{P}(X_v \in A \mid \mathscr{F}_t^X) \mid \mathscr{F}_s^X] = \int P_{s,t}(X_s, dy) P_{t,v}(y, A).$$

This should equal $P_{s,v}(X_s, A)$.

1.5.2 Definition. A *transition function* on $(\Omega, \mathscr{F})$ is a family $\{P_{s,t} \mid 0 \le s < t\}$ of transition probabilities on $(\Omega, \mathscr{F})$ such that for all $s < t < v$,

$$\int_\Omega P_{s,t}(X_s, dy) P_{t,v}(y, A) = P_{s,v}(x, A).$$

These are the *Chapman-Kolmogorov equations*.

1.5.3 Definition. Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t, t \ge 0\}, \mathbb{P})$ be a filtered probability space. An adapted process X is a *Markov process* with respect to another filtration $\{\mathscr{G}_t\}$ (containing $\{\mathscr{F}_t^X\}$) with transition functions $P_{s,t}$ if for all non-negative $\mathscr{F}$-measurable functions f and $0 \le s \le t$, $\mathbb{E}[f(X_t) \mid \mathscr{G}_s] = P_{s,t} f(X_s)$.

Remark.

(i) Given a transition function, one can always construct a Markov process with that transition function using Kolmogorov's extension theorem.
(ii) The transition function is said to be *homogeneous* if for all $s < t$, $P_{s,t}$ depends on s and t only through $t - s$. In this case the C-K equation takes the form $P_{t+s}(x, A) = \int P_s(x, dy) P_t(y, A)$.

1.5.4 Theorem. *Brownian motion is a Markov process with respect to its natural filtration. (With what transition function?)*

Intuition: $B_t = B_t - B_s + B_s$ and $B_t - B_s$ is $\mathscr{N}(0, t - s)$, distributed as $\mathbb{P}^y$ if $B_s = y$.

1.5.5 Lemma. *Suppose that X and Y are d-dimensional random vectors on $(\Omega, \mathscr{F}, \mathbb{P})$, $\mathscr{G}$ is a sub-σ-algebra of $\mathscr{F}$, X is independent of $\mathscr{G}$ and Y is $\mathscr{G}$-measurable. Then for every $\Gamma \in \mathscr{B}(\mathbb{R}^d)$,*

$$\mathbb{P}[X + Y \in \Gamma \mid \mathscr{G}] = \mathbb{P}[X + Y \in \Gamma \mid Y] \quad \mathbb{P}\text{-a.s.}$$

and

$$\mathbb{P}[X + Y \in \Gamma \mid Y = y] = \mathbb{P}[X + Y \in \Gamma] \quad \text{a.e. for } \mathbb{P} Y^{-1}.$$

PROOF: We will show that for $D \in \mathscr{B}(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\mathbb{P}[(X, Y) \in D \mid \mathscr{G}] = \mathbb{P}[(X, Y) \in D \mid Y].$$

First look at $D = D_1 \times D_2$ for $D_1, D_2 \in \mathscr{B}(\mathbb{R}^d)$. The left hand side is

$$\mathbb{P}[X \in D_1, Y \in D_2 \mid \mathscr{G}] = \mathbf{1}_{\{Y \in D_2\}} \mathbb{P}[X \in D_1 \mid \mathscr{G}] = \mathbf{1}_{\{Y \in D_2\}} \mathbb{P}[X \in D_1]$$

and the right hand side is equal to the same thing by the same logic. Since the measurable rectangles form a Dynkin system and generate $\mathscr{B}(\mathbb{R}^d \times \mathbb{R}^d)$, we are done. □

Let $\Omega^0 = (C[0,\infty) : \mathbb{R})^d$, $\mathscr{F}^0 = \mathscr{B}(\Omega^0)$, and $P^0 = P^{(1)} \times \cdots \times P^{(d)}$, where each $P^{(i)}$ is Wiener measure. Let X be the canonical process on $(\Omega^0, \mathscr{F}^0, P^0)$, so X is a d-dimensional Brownian motion started at zero. Let μ be an arbitrary initial distribution on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$. Consider the random variable on $(\mathbb{R}^d \times \Omega^0, \mathscr{B}(\mathbb{R}^d) \otimes \mathscr{F}^0, \mu \otimes P^0)$ defined by $X(x, \omega_1, \dots, \omega_d) = x + (\omega_1, \dots, \omega_d)$, and so is Brownian motions with initial distribution μ.

Another way to think about this is to think of P^μ as the image measure. How can we explicitly write P^μ in terms of P^0 and μ? Naturally, take $P^x(F) = P^0(F - x)$, and write $P^\mu(F) = \int P^x(F)\mu(dx)$. This integral is well-defined if for every $F \in \mathscr{F}^0$ the map $x \mapsto P^x(F)$ is "universally measurable." The following fact is true: for every $F \in \mathscr{F}^B_\infty$ the map $x \mapsto P^x(F)$ is $\mathscr{B}(\mathbb{R}^d)$-measurable. (Universal measurability is introduced so that we get this nice property for slightly larger filtrations, such as the augmented natural filtration.)

Define

$$\mathscr{U}(\mathbb{R}^d) := \bigcap_{\mu \text{ prob}} \overline{\mathscr{B}(\mathbb{R}^d)}^{\mu} \supseteq \mathscr{B}(\mathbb{R}^d).$$

A mapping $\mathbb{R}^d \to \mathbb{R}$ is said to be *universally measurable* if it is $\mathscr{U}(\mathbb{R}^d)/\mathscr{B}(\mathbb{R}^d)$-measurable.

Note that if F is a set of the form $\{\omega \in \Omega^0 \mid \omega(0) \in \Gamma_0, \omega(t_1) \in \Gamma_1\}$ then $P^x(F) = \mathbf{1}_{\Gamma_0}(x) \int_{\Gamma_1} p_d(t_1; x_1, y_1) dy_1$, where as always

$$p_d(t; x, y) = \frac{1}{(2\pi t)^{\frac{d}{2}}} \exp\left(-\frac{\|x - y\|^2}{2t} \right).$$

As a consequence, Brownian motion is a homogeneous Markov process with transition function $P_t(x, A) = \int_A p_d(t; x, y) dy$. Here we call $p_d(t; x, y)$ the *transition density* of Brownian motion. We have seen $\mathbb{E}[f(B_{t+s}) \mid B_s] = P_t f(B_s)$. The *infinitesimal generator* of a homogeneous Markov process is $\mathscr{G} := \lim_{s \searrow} \frac{1}{s}(P_s - I)$.

1.5.6 Definition. A *Markov family* is an adapted process $\{S_t, \mathscr{F}_t \mid t \geq 0\}$ on some $(\Omega, \mathscr{F})$ together with a family of probability measures $\{P^x \mid x \in \mathbb{R}^d\}$ on $(\Omega, \mathscr{F})$ such that

(i) $x \mapsto P^x(F)$ is universally measurable for all $F \in \mathscr{F}$;
(ii) $P^x[X_0 = x] = 1$ for all $x \in \mathbb{R}^d$;
(iii) $P^x[X_{s+\cdot} \in F \mid \mathscr{F}_s] = P^x[X_{s+\cdot} \in F \mid X_s]$ for all $x \in \mathbb{R}^d$, for all $F \in \mathscr{B}(\mathbb{R}^d)^{[0,\infty)}$;
(iv) $P^x[X_{s+t} \in \Gamma \mid X_s = y] = P^y[X_t \in \Gamma]$ for all $x \in \mathbb{R}^d$, a.s.-$P^x X_s^{-1}$

1.5.7 Definition. A process X is $\mathscr{F}$*-progressively measurable* if the restricted map $X : [0,t] \times \Omega \to \mathbb{R}^d$ is $(\mathscr{B}[0,t] \otimes \mathscr{F}_t)/\mathscr{B}(\mathbb{R}^d)$-measurable for all $t < \infty$.

1.5.8 Definition. The σ-algebra generated by a random time T is the σ-algebra generated by $\{\{X_T \in A\} \mid A \in \mathscr{B}(\mathbb{R}^d)\} \cup \{T = \infty\}$.

1.5.9 Definition. A random time T is an *$\mathscr{F}$-stopping time* (resp. *$\mathscr{F}$-optional time*) if $\{T \le t\} \in \mathscr{F}_t$ (resp. $\{T < t\} \in \mathscr{F}_t$) for all t. The σ-algebra generated by a stopping time T is

$$\mathscr{F}_T := \{A \in \mathscr{F} \mid A \cap \{T \le t\} \in \mathscr{F}_t \text{ for all } t \ge 0\}.$$

1.5.10 Exercise. Suppose that X is an adapted process with right continuous paths and $A \in \mathscr{B}(\mathbb{R}^d)$. The *hitting time* is $H_A = \inf\{t \ge 0 \mid X_t \in A\}$.
(i) If A is open show that H_A is an optional time.
(ii) If A is closed and X is continuous show that H_A is a stopping time.

1.5.11 Definition. Let $\{\overline{\mathscr{F}}_t\}_{t\ge 0}$ be the completion of $\{\mathscr{F}_t^B\}_{t\ge 0}$, and for each $t \ge 0$ let $\mathscr{F}_t = \bigcap_{s>t} \overline{\mathscr{F}}_s$. Then $\{\mathscr{F}_t\}_{t\ge 0}$ is a right continuous filtration, the *Brownian filtration*.

1.5.12 Theorem. *Let $\{B_t, \mathscr{F}_t\}_{t\ge 0}$ be a Brownian motion and let T be a finite valued stopping time. Then the process defined by $B_t^{(T)} = B_{T+t} - B_T$ for $t \ge 0$ is a Brownian motion independent of $\mathscr{F}_T$.*

PROOF: A *simple stopping time* is a stopping time whose image is countable. We claim that given any finite time there is a non-increasing sequence of simple stopping times $T_1 \ge T_2 \ge \cdots$ such that $\lim_{n\to\infty} T_n = T$ point-wise. In addition, $\mathscr{F}_T = \bigcap_n \mathscr{F}_{T_n}$. (The reason we take the approximation from above is for this latter property.) Indeed, define

$$T_n = \sum_{k=0}^{\infty} \frac{k}{2^n} \mathbf{1}_{[k2^{-n},(k+1)2^{-n})} \circ T.$$

Then clearly $T_n \ge T_{n+1}$ for all n, and T_n converges to T since $0 \le T_n - T \le 2^{-n}$. Further, $\mathscr{F}_T \subseteq \mathscr{F}_{T_n}$ for all n as a consequence of the general fact that $S \le T$ implies $\mathscr{F}_S \subseteq \mathscr{F}_T$. (Indeed, if $A \in \mathscr{F}_S$ then

$$A \cap \{T \le t\} = A \cap \{S \le t\} \cap \{T \le t\} \in \mathscr{F}_t$$

so $A \in \mathscr{F}_T$.) If $A \in \bigcap_n \mathscr{F}_{T_n}$ then $A \cap \{T_n \le t\} \in \mathscr{F}_t$ for all $n \ge 1$ and all $t \ge 0$. Therefore

$$A \cap \{T \le t\} = \bigcap_{\varepsilon>0} \bigcup_{m\ge 1} \bigcap_{n\ge m} (A \cap \{T_n \le t + \varepsilon\}) \in \bigcap_{\varepsilon>0} \mathscr{F}_{t+\varepsilon} = \mathscr{F}_t$$

by the right continuity of $\{\mathscr{F}_t\}$.

Back to the proof of the theorem. If T is a simple stopping time then let $\{\tau_1, \tau_2, \dots\}$ be the range of T. For any $A \in \mathscr{F}_T$ and for all $C_1, \dots, C_n \in \mathscr{B}(\mathbb{R})$,

$$\begin{aligned}
&\mathbb{P}(A \cap (\bigcap_{i \le m} \{B_{T+t_i} - B_T \in C_i\})) \\
&\qquad = \sum_{k=0}^{\infty} \mathbb{P}(A \cap (\bigcap_{i \le m} \{B_{\tau_k + t_i} - B_{\tau_k} \in C_i, T = \tau_k\})) \\
&\qquad = \sum_{k=0}^{\infty} \mathbb{P}(\bigcap_{i \le m} \{B_{\tau_k + t_i} - B_{\tau_k} \in C_i\}) \mathbb{P}(\{T = \tau_k\} \cap A) \\
&\qquad = \mathbb{P}(\bigcap_{i \le m} \{B_{t_i} \in C_i\}) \mathbb{P}(A)
\end{aligned}$$

Now set $A = \mathbb{R}$ to deduce that $t \mapsto B_{T+t} - B_T$ is a BM. Since $A \in \mathscr{F}_T$ was arbitrary, we also get independence.

For a general topping time T, consider the approximating sequence of simple stopping times $T_n \searrow T$ defined above. For any $A \in \mathscr{F}_T$ and for all open $C_1, \dots, C_n \in \mathscr{B}(\mathbb{R}^n)$ we have

$$\begin{aligned}
\mathbb{P}(A \cap (\bigcap_{i \le m} \{B_{T+t_i} - B_T \in C_i\})) &= \lim_{n \to \infty} \mathbb{P}(A \cap (\bigcap_{i \le m} \{B_{T_n + t_i} - B_{T_n} \in C_i\})) \\
&= \lim_{n \to \infty} \mathbb{P}(\bigcap_{i \le m} \{B_{t_i} \in C_i\}) \mathbb{P}(A) \\
&= \mathbb{P}(\bigcap_{i \le m} \{B_{t_i} \in C_i\}) \mathbb{P}(A)
\end{aligned}$$

since $A \in \mathscr{F}_{T_n}$ for all n. □

1.5.13 Example. This theorem is not true for general random times. Take T to be the last time before 1 that B_t is zero.

1.5.14 Lemma. *Under P^0, $|B| = \{|B_t|, \mathscr{F}_t\}$ is a Markov process with transition density $P^0[|W_{t-s}| \in dy \mid |W_t| = x] = p_+(s; x, y)$, where $p_+(s; x, y) = p(s; x, y) + p(s; x, -y)$.*

1.5.15 Lemma. *Define $Y_t = M_t - B_t$. Under P^0, the process $\{Y_t, \mathscr{F}_t\}$ is a Markov process and has transition density*

$$P^0[Y_{t+s} \in dy \mid Y_t = z] = (p(s; z, y) + p(s; z, -y))dy = p_+(s; z, y)dy.$$

Proof: For $s > 0$, $t \ge 0$, $b \ge a$, $b \ge 0$,

$$\begin{aligned}
P^0[B_{t+s} \ge a, M_{t+s} \le b \mid \mathscr{F}_t] &= P^0[B_{t+s} \ge a, M_t \le b, (\sup_{u \in [0,s]} B_{t+u}) \le b \mid \mathscr{F}_t] \\
&= \mathbf{1}_{\{M_t \le b\}} P^0[B_{t+s} \ge a, (\sup_{u \in [0,s]} B_{t+u}) \le b \mid \mathscr{F}_t]
\end{aligned}$$

$$= \mathbf{1}_{\{M_t \le b\}} P^0[B_{t+s} \ge a, (\sup_{u\in[0,s]} B_{t+u}) \le b \mid B_t]$$

since $\{B_t\}$ is a Markov process under P^0. This calculation shows that (M_t, B_t) is a Markov process under P^0. Since Y_t is a function of M_t and B_t, it follows that for every $\Gamma \in \mathscr{B}(\mathbb{R})$,

$$P^0[Y_{t+s} \in \Gamma \mid \mathscr{F}_t] = P^0[Y_{t+s} \in \Gamma \mid B_t, M_t].$$

It suffices to show that

$$P^0[Y_{t+s} \in dy \mid B_t = x, M_t = m] = p_+(s; m-x, y)dy.$$

For $b > m > x$, $b \ge a$, $m \ge 0$,

$$\begin{aligned}
P^0&[B_{t+s} \in da, M_{t+s} \in db \mid B_t = x, M_t = m] \\
&= P^0[B_{t+s} \in da, \max_{0\le u\le s} B_{t+u} \in db \mid B_t = x, M_t = m] \\
&= P^x[B_s \in da, M_s \in db] \\
&= P^0[B_s \in da - x, M_s \in db - x] \\
&= \frac{2}{\sqrt{2\pi s^3}}(2b-a-x)e^{-\frac{(2b-a-x)^2}{2s}}\, da\, db
\end{aligned}$$

For $m > x$, $m \ge a$, $m \ge 0$,

$$\begin{aligned}
P^0&[B_{t+s} \in da, M_{t+s} = m \mid B_t = x, M_t = m] \\
&= P^0[B_{t+s} \in da, \sup_{u\in[0,s]} B_{t+u} \le m \mid B_t = x, M_t = m] \\
&= P^x[B_s \in da, M_s \le m] \\
&= P^x[B_s \in da] - P^x[B_s \in da, M_s \ge m] \\
&= \frac{1}{\sqrt{2\pi s}}\left(e^{-\frac{(a-x)^2}{2s}} - e^{-\frac{(2m-a-x)^2}{2s}}\right) da
\end{aligned}$$

Therefore, since either the maximum increases to a new level $b > m$ over the interval $[t, t+s]$ or it stays the same, we have

$$\begin{aligned}
P^0&[Y_{t+s} \in dy \mid B_t = x, M_t = m] \\
&= \int_m^\infty P^0[B_{t+s} \in b - dy, M_{t+s} \in db \mid B_t = x, M_t = m] \\
&\qquad + P^0[B_{t+s} \in m - dy, M_{t+s} = m \mid B_t = x, M_t = m] \\
&= \int_m^\infty \frac{2}{\sqrt{2\pi s^3}}(b+y-x)e^{-\frac{(b+y-x)^2}{2s}}\, dy\, db \\
&\qquad + \frac{1}{\sqrt{2\pi s}}\left(e^{-\frac{(m-y-x)^2}{2s}} - e^{-\frac{(m+y-x)^2}{2s}}\right) dy \\
&= p_+(s; m-x, y)dy
\end{aligned}$$

□

We have the following Markov processes:

(i) Brownian motion B_t
(ii) Poisson process N_t
(iii) Reflected Brownian motion $|B_t|$
(iv) (M_t, B_t)
(v) $Y_t = M_t - B_t$

What about

$$T_b = \inf\{t \geq 0 \mid B_t = b\} = \inf\{t \geq 0 \mid B_t \geq b\} = \inf\{t \geq 0 \mid M_t \geq b\}?$$

1.5.16 Theorem. $\{T_b, 0 < b < \infty\}$ *is a non-decreasing left-continuous (strong Markov) process that has stationary independent increments and is purely discontinuous (i.e. there is no interval on which* $b \mapsto T_b$ *is right-continuous).*

PROOF: Notice that

$$\{T_b \leq t\} = \{M_t \geq b\} = \bigcap_{n \in \mathbb{N}} \{M_t \geq b - \tfrac{1}{n}\} = \bigcap_{n \in \mathbb{N}} \{T_{b-\frac{1}{n}} \leq t\}.$$

which implies left-continuity. That it is non-decreasing is obvious.

The *time shift operator*, defined by $\theta_s(\omega)(t) := \omega(s+t)$ for $s, t \geq 0$. The operator can also be defined for random times. We have for $0 < a < b$, $T_b = T_a + T_b \circ \theta_{T_a}$ a.s. For all $\mathscr{F}_{T_b}$-measurable functions f,

$$\mathbb{E}[f(T_b - T_a) \mid \mathscr{F}_{T_a}] = \mathbb{E}[f(T_b \circ \theta_{T_a}) \mid \mathscr{F}_{T_a}] = \mathbb{E}^a[f(T_b)] = \mathbb{E}[f(T_{b-a})]$$

using the continuity of Brownian motion and the fact that Brownian motion has stationary, independent increments. This implies that $T_b - T_a$ is independent of $\mathscr{F}_a$ and has the same distribution as T_{b-a}.

For the last part it is enough to show that for $p, q \in \mathbb{Q}$,

$$P[\omega \mid b \mapsto T_b(\omega) \text{ is cts on } [p,q]] = 0.$$

However, $b \mapsto T_b$ is continuous on this interval if and only if M_t is strictly increasing on $[T_p, T_q]$, and for this to happen we would require $B_{T_p+t} - B_{T_p}$ to be strictly increasing on that interval. But this last process is a Brownian motion by the strong Markov property, and so is not strictly increasing anywhere. □

1.5.17 Lemma. $\mathbb{E}^0[\exp(-uT_b)] = \exp(-b\sqrt{2u})$.

1.5.18 Proposition. *Almost surely, the set* $Z = \{t \in [0,\infty) \mid B_t = 0\}$ *has no isolated points.*

PROOF: Earlier we showed that zero is not an isolated point, indeed

$$\mathbb{E}^0[\exp(-u(t + T_0 \circ \theta_t))] = e^{-ut}\,\mathbb{E}^0[\mathbb{E}^{B_t}[\exp(-uT_0)]] = e^{-ut}\,\mathbb{E}^0[\exp(-|B_t|\sqrt{2u})]]$$

and as $t \to 0$, the right hand side goes to one. Therefore by Fatou's Lemma,

$$\mathbb{P}[\liminf_{t\searrow 0}(t + T_0 \circ \theta_t) = 0] \geq \liminf_{t\searrow 0}\mathbb{P}[(t + T_0 \circ \theta_t) = 0] = 1,$$

so zero is a limit point of Z a.s. For any rational q, the define the time d_q to be $q + T_q \circ \theta_q$, the first point in Z after q. However, $B_{d_q} = 0$, so $\{B_{d_q+t} \mid t \geq 0\}$ is a standard BM by the strong Markov property. Therefore the set

$$\bigcup_{q\in\mathbb{Q}} \{d_q \text{ is not a limit point of } Z\}$$

has $\mathbb{P}$-measure zero. If $h \in Z$ and $h = d_q$ then h is a limit point of Z. If not, choose a sequence $\{q_n\} \subseteq \mathbb{Q}$ such that $q_n \nearrow h$. Then $d_{q_n} \in [q_n, h]$, so $d_{q_n} \to h$ and h is a limit point of Z. □

2 Martingales

2.1 Martingale convergence theorem

The definitions of sub- and super-martingales are analogous in discrete- and continuous-time.

2.1.1 Definition. A stochastic process X is *integrable* if $\mathbb{E}|X_a| < \infty$ for all $a \in I$. An adapted stochastic process $\{X_a, \mathscr{F}_a \mid a \in I\}$ is a

(i) *sub-martingale* if $\mathbb{E}[X_b \mid \mathscr{F}_a] \geq X_a$ for all $a < b$;
(ii) *super-martingale* if $\mathbb{E}[X_b \mid \mathscr{F}_a] \leq X_a$ for all $a < b$;
(iii) *martingale* if it is both a sub- and super-martingale.

2.1.2 Doob's Upcrossing Lemma. *Let X be a super-martingale and let $U_N[a,b]$ be the number of up-crossings of $[a,b]$ by time N. Then*

$$(b - a)\,\mathbb{E}\,U_N[a,b] \leq \mathbb{E}(X_N - a)^-$$

2.1.3 Definition. A *predictable process* is a process $\{C_n, \mathscr{F}_n\}$ such that C_n is $\mathscr{F}_{n-1}$-measurable for all n.

PROOF: Let $\{X_n, \mathscr{F}_n\}$ be a super-martingale, $C_1 = \mathbf{1}_{\{X_0<a\}}$ and

$$C_n = \mathbf{1}_{\{C_{n-1}=1\}}\mathbf{1}_{\{X_{n-1}\leq b\}} + \mathbf{1}_{\{C_{n-1}=0\}}\mathbf{1}_{\{X_{n-1}<a\}}$$

for $n > 1$, and $Y_n := \sum_{i=1}^n C_i(X_i - X_{i-1})$. Then $\{C_n, \mathscr{F}_n\}$ is predicable and $\{Y_n, \mathscr{F}_n\}$ is a super-martingale. We have fundamental inequality

$$Y_N \geq (b - a)U_N[a,b] - (X_N - a)^-.$$

Therefore we may conclude that

$$0 \geq \mathbb{E}[Y_N] \geq (b - a)\,\mathbb{E}\,U_N[a,b] - \mathbb{E}(X_N - a)^-.$$ □

Now for the continuous-time analog. For $a < b$ and $F \subseteq [0, \infty)$ finite, let

(i) $\tau_1 = \min\{t \in F \mid X_t \le a\}$

(ii) $\sigma_j = \min\{t \in F \mid t \ge \tau_j, X_t > b\}$

(iii) $\tau_{j+1} = \min\{t \in F \mid t \ge \sigma_j, X_t < a\}$

Given an interval $I \subseteq [0, \infty)$ let

$$U_I(a, b; X) = \sup\{U_F[a, b] \mid F \subseteq I \text{ finite}\}.$$

2.1.4 Theorem. *Let $\{X_t, \mathscr{F}_t\}$ be a right-continuous sub-martingale, $a < b$, and $\lambda > 0$.*

(i) $\lambda \mathbb{P}[\sup_{t\in[\sigma,\tau]} X_t \ge \lambda] \le \mathbb{E} X_\tau^+$.

(ii) $\lambda \mathbb{P}[\inf_{t\in[\sigma,\tau]} X_t \le -\lambda] \le \mathbb{E} X_\tau^+ - \mathbb{E} X_\sigma$.

(iii) Up-crossings: $(b - a)\mathbb{E} U_{[\sigma,\tau]}(a, b; X) \le |a| + \mathbb{E} X_\tau^+$.

(iv) If X is non-negative then $\mathbb{E}[(\sup_{t\in[\sigma,\tau]} X_t)^p] \le (\frac{p}{p-1})^p \mathbb{E} X_\tau^p$.

PROOF: Exercise. (Approximation arguments.) □

2.1.5 Doob's Forward Convergence Theorem. *Let $\{X_n, \mathscr{F}_n\}$ be a super-martingale bounded in L^1 (i.e. $\sup_n \mathbb{E}|X_n| < \infty$). Then $X_\infty = \lim_{n\to\infty} X_n$ exists and is finite a.s., and it is $\mathscr{F}_\infty$-measurable.*

PROOF: Let

$$\begin{aligned}
\Lambda &= \{X_n \text{ does not converge to a limit in } [-\infty, \infty]\} \\
&= \{\liminf_n X_n \ne \limsup_n X_n\} \\
&= \bigcup_{a<b\in\mathbb{Q}} \{\liminf_n X_n < a < b < \limsup_n X_n\} \\
&=: \bigcup_{a<b\in\mathbb{Q}} \Lambda_{a,b}.
\end{aligned}$$

But $\Lambda_{a,b} \subseteq \{\lim_{N\to\infty} U_N[a, b] = \infty\}$. The probability of this set is zero, so $\mathbb{P}(\Lambda) = 0$. By Fatou's Lemma

$$\mathbb{E}|X_\infty| = \mathbb{E}[\liminf_n |X_n|] \le \liminf_n \mathbb{E}|X_n| < \infty.$$

□

2.1.6 Doob's Forward Convergence Theorem. *Let $\{X_t, \mathscr{F}_t\}$ be a cadlag super-martingale bounded in L^1 (i.e. $\sup_t \mathbb{E}|X_t| < \infty$). Then $X_\infty = \lim_{t\to\infty} X_t$ exists and is finite a.s., and it is $\mathscr{F}_\infty$-measurable.*

2.1.7 Discrete Optional Sampling Theorem. *Let M be a uniformly integrable martingale and T be a stopping time. Then $\mathbb{E}[M_\infty \mid \mathscr{F}_T] = M_T$ a.s.*

2.1.8 Corollary. $\mathbb{E}|M_T| < \infty$ *and* $\mathbb{E} M_T = \mathbb{E} M_0$.

2.1.9 Corollary. *If S is another stopping time and $S \leq T$ then $\mathbb{E}[M_T \mid \mathscr{F}_S] = M_S$.*

2.1.10 Continuous Optional Sampling Theorem. *Let $\{X_t, \mathscr{F}_t\}$ be a right-continuous sub-martingale with a last element and let S and T be $\mathscr{F}_t$-optional times. Then $\mathbb{E}[X_T \mid \mathscr{F}_{S+}] \geq M_S$ a.s. If S is a stopping time then we may replace $\mathscr{F}_{S+}$ by $\mathscr{F}_S$.*

PROOF: Define

$$S_n = \begin{cases} \infty & S = \infty \\ k2^{-n} & (k-1)2^{-n} \leq S < k2^{-n} \end{cases}$$

and T_n similarly. Then S_n and T_n are stopping times and $S_n \searrow W$ and $T_n \searrow T$, and by the discrete optional sampling theorem, $\mathbb{E}[X_{T_n} \mid \mathscr{F}_{S_n}] \geq X_{S_n}$. For all $A \in \mathscr{F}_{S_n}$,

$$\int_A X_{T_n} \, d\,\mathbb{P} \geq \int_A X_{S_n} \, d\,\mathbb{P}.$$

Therefore it also holds for all $A \in \bigcap_{n=1}^{\infty} \mathscr{F}_{S_n} = \mathscr{F}_{S+}$. Also, since $S \leq S_n$, $\mathscr{F}_S \subseteq \mathscr{F}_{S_n}$.

Observe that $\{X_{S_n}, \mathscr{F}_{S_n}\}$ are backward sub-martingales and $\mathbb{E}X_{S_n}$ is decreasing and bounded below by $\mathbb{E}X_0$. Therefore $\{X_{S_n}\}$ are u.i., and likewise for $\{X_{T_n}\}$.

Since the process is right continuous

$$X_S = \lim_{n\to\infty} X_{S_n} \quad \text{and} \quad X_T = \lim_{n\to\infty} X_{T_n},$$

so we can take limits in the equation above and interchange the limits to get $\mathbb{E}[X_T \mid \mathscr{F}_{S+}] \geq \mathbb{E}X_S$. □

2.2 Continuous Martingales

For this section let $\{X_t, \mathscr{F}_t\}$ be a process with right continuous paths.

Remark. (i) We automatically know that it has limits from the left a.s. since

$$\{\exists\ t \in [0,n]\ \liminf_{s \nearrow t} X_s < \limsup_{s\nearrow t} X_s\} \subseteq \bigcup_{a<b\in\mathbb{Q}} \{\omega \mid U_{[0,n]}(a,b,X(\omega)) = \infty\}.$$

(ii) Recall that $\{\mathscr{F}_t\}$ is said to satisfy the *usual conditions* if $\mathscr{F}_0$ contains all the $\mathbb{P}$-negligible sets and $\{\mathscr{F}_t\}$ is right continuous. If $\{X_t, \mathscr{F}_t\}$ is a sub-martingale and $\{\mathscr{F}_t\}$ satisfies the usual conditions then $t \mapsto \mathbb{E}X_t$ is right continuous then X_t has a right continuous modification such that $\{X_t, \mathscr{F}_t\}$ is a sub-martingale.

(iii) Continuous martingale results that are derived from discrete martingale results (e.g. OST) rely on approximation arguments for which a key ingredient is the backward sub-martingale convergence theorem. This theorem (given below) allows one to justify limits of the following kind.

Let $T_n \searrow T$, $S_n \searrow S$, and $\{X_n, \mathscr{F}_n\}$ is a right-continuous sub-martingale. If $X_{S_n} \leq \mathbb{E}[X_{T_n} \mid \mathscr{F}_{S_n}]$ for all n then $X_S \leq \mathbb{E}[X_T \mid \mathscr{F}_{S+}]$. Namely, for all $A \in \mathscr{F}_{S+} \subseteq \mathscr{F}_{S_n}$ we have

$$\lim_{n\to\infty} \int_A X_{T_n} d\,\mathbb{P} =^{u.i.} \int_A \lim_{n\to\infty} X_{T_n} d\,\mathbb{P} =^{rt\ cts} \int_A X_T d\,\mathbb{P}$$

using the fact that $\mathbb{E} X_{T_n} \geq \mathbb{E} X_0$.

2.2.1 Theorem (BS-MCT). *Let $\{\mathscr{F}_n\}_{n=1}^{\infty}$ be a decreasing sequence of σ-algebra and suppose that $\{X_n, \mathscr{F}_n\}$ is a backward sub-martingale (i.e. $\mathbb{E}|X_n| < \infty$ and $\mathbb{E}[X_n \mid \mathscr{F}_{n+1}] \geq X_{n+1}$ a.s. for all n). If $\lim_{n\to\infty} \mathbb{E} X_n > -\infty$ then $\{X_n\}$ is u.i.*

PROOF: Step 1: Show that $\{X_n^+, \mathscr{F}_n\}$ is a backward sub-martingale. This is the case since $X_{n+1} \leq \mathbb{E}[X_n \mid \mathscr{F}_{n+1}]$ implies, since $x \mapsto x^+$ is non-decreasing,

$$X_{n+1}^+ \leq \mathbb{E}[X_n \mid \mathscr{F}_{n+1}]^+ \leq \mathbb{E}[X_n^+ \mid \mathscr{F}_{n+1}]$$

by the condition Jensen inequality.

Step 2: $\lim_{\lambda\to\infty} \sup_{n\geq 1} \mathbb{P}[|X_n| > \lambda] = 0$ since

$$\lambda\,\mathbb{P}[|X_n| > \lambda] \leq \mathbb{E}|X_n| = -\,\mathbb{E} X_n + 2\,\mathbb{E} X_n^+ < \infty$$

by the assumed bound and the fact that $\mathbb{E} X_n^+ \leq \mathbb{E} X_1^+$.

Step 3: X_n^+ is u.i. Indeed, since $\{X_n^+, \mathscr{F}_n\}$ is a sub-martingale we have $\mathbb{E}[X_{n-1}^+ \mid \mathscr{F}_n] \geq X_n^+$ so

$$\int_{\{|X_n|>\lambda\}} X_n^+ d\,\mathbb{P} = \int_{\{|X_n|>\lambda\}} \mathbb{E}[X_1^+ \mid \mathscr{F}_n] d\,\mathbb{P} \leq \int_{\{|X_n|>\lambda\}} X_1^+ d\,\mathbb{P}$$

Step 4: X_n^- is u.i. Indeed, for $\lambda > 0$ and $n > m$ we have

$$\begin{aligned}
0 &\geq \int_{X_n<-\lambda} X_n d\,\mathbb{P} \\
&= \mathbb{E} X_n - \int_{X_n\geq-\lambda} X_n d\,\mathbb{P} \\
&\geq \mathbb{E} X_n - \int_{X_n\geq-\lambda} X_m d\,\mathbb{P} \\
&= \mathbb{E} X_n - \mathbb{E} X_m + \int_{X_n<-\lambda} X_m d\,\mathbb{P}
\end{aligned}$$

Given $\varepsilon > 0$, there is m large enough so that

$$0 \leq \mathbb{E} X_m - \mathbb{E} X_n \leq \frac{\varepsilon}{2}$$

for all $n > m$ (since X is L^1-bounded and $\mathbb{E}X_n$ is monotonic). For that m choose $\lambda > 0$ such that

$$\sup_{n>m} \int_{X_n<-\lambda} |X_m| \, d\,\mathbb{P} < \frac{\varepsilon}{2}$$

so $\sup_{n>m} \int_{X_n^- > \lambda} X_n^- \, d\,\mathbb{P} < \varepsilon$ and X_n^1 is u.i. □

2.2.2 Theorem (Convergence). *If $\{X_t, \mathscr{F}_t\}$ is a sub-martingale and*

$$\sup_{t\geq 0} \mathbb{E} X_t^+ < \infty$$

then X_t has a limit a.s. and in L^1.

2.2.3 Corollary. *If $\{X_t, \mathscr{F}_t\}_{t\geq 0}$ is a right continuous non-negative super-martingale then $X_\infty = \lim_{t\to\infty} X_t$ exists and is in L^1.*

2.2.4 Optional Sampling Theorem. *If $\{X_t, \mathscr{F}_t\}$ is a right-continuous sub-martingale with a last element (i.e. $X_\infty = \lim_{t\to\infty} X_t$ exists a.s. and is in L^1) and S and T are $\{\mathscr{F}_t\}$-optional times then $\mathbb{E}[X_T \mid \mathscr{F}_{S+}] \geq M_S$ a.s. If S is a stopping time then we may replace $\mathscr{F}_{S+}$ by $\mathscr{F}_S$.*

2.3 Applications

Note first that if $\{B_t, \mathscr{F}_t\}$ is a Brownian motion then it is a martingale. Indeed,

$$\mathbb{E}[B_t - B_s + B_s \mid \mathscr{F}_s] = \mathbb{E}[B_t - B_s] + B_s = B_s.$$

2.3.1 Lemma. *Let $\tau = \inf\{t \geq 0 \mid B_t \notin (a, b)\}$, where $a < 0 < b$. Then*
(i) $\mathbb{P}(B_\tau = b) = \frac{-a}{b-a}$
(ii) $\mathbb{E}[\tau] = -ab$

Proof: τ is a stopping time, but we cannot naively apply the OST since Brownian motion does not have a last element. Instead we look at the *stopped process* $\{B_{t\wedge n}\}$, which is a right continuous martingale with last element. Applying the OST we get

$$0 = \mathbb{E} B_{\tau\wedge n} = b\,\mathbb{P}[B_\tau = b, \tau \leq n] + a\,\mathbb{P}[B_\tau = a, \tau \leq n] + \mathbb{E}[B_n; \tau > n].$$

Taking limits as $n \to \infty$ we get

$$0 = b\,\mathbb{P}[B_\tau = b] + a\,\mathbb{P}[B_\tau = a]$$

For the next part, we show that $M_t := (B_t - a)(b - B_t) + t$ is a martingale. (This is not too hard.) Then again applying the OST to the stopped process

$$-ab = \mathbb{E} M_{\tau\wedge n} = \mathbb{E}[\tau \wedge n] + \mathbb{E}[(B_{\tau\wedge n} - a)(b - B_{\tau\wedge n})]$$

and taking limits we get the result. □

2.3.2 Example. Let $X_t = B_t + ct$ (Brownian motion with drift). We are interested in $H_x = \inf\{t > 0 \mid X_t = x\}$. We need the fact that $\exp(\theta B_t - \frac{1}{2}\theta^2 t)$ is a martingale. Fix $\lambda > 0$. Then from the exponential martingale it follows that

$$\exp(\theta X_t - \lambda t) = \exp(\theta B_t - (\lambda - \theta c)t)$$

is a martingale provided that $\lambda - \theta c = \frac{1}{2}\theta^2$. Let $\beta, \alpha = -c \pm \sqrt{c^2 + 2\lambda}$. Note that $\alpha < 0 < \beta$. Thus for any $\lambda > 0$ and β as given the martingale $\exp(\beta X_t - \lambda t)$ is bounded on $[0, H_x]$. We can use the OST to conclude that

$$\mathbb{E}[\exp(\beta X_{H_x} - \lambda H_x)] = e^{\beta x}\,\mathbb{E}[e^{-\lambda H_x}]$$

from which it follows that $\mathbb{E}[e^{-\lambda H_x}] = \exp(-x(\sqrt{c^2 + 2\lambda} - c))$. The Laplace transform can be inverted explicitly to give

$$\mathbb{P}(H_x \in dt) = \frac{x}{\sqrt{2\pi t^3}} \exp\left(-\frac{(x - ct)^2}{2t}\right).$$

Take limits as $\lambda \searrow 0$ in the Laplace transform to conclude that

$$\mathbb{P}[H_x < \infty] = \begin{cases} 1 & \text{if } c \geq 0 \\ e^{-2|c|x} & \text{if } c < 0 \end{cases}$$

Now we calculate $\mathbb{E}[e^{-\lambda T}]$ where $T = H_a \wedge H_b$ for $a < 0 < b$. Recall that the θ that we used previously was found as a root of $\lambda - \theta c = \frac{1}{2}\theta^2$. We know that any process of the form

$$M_t = C_1 e^{\alpha X_t - \lambda t} + C_2 e^{\beta X_t - \lambda t}$$

is a martingale for any constants C_1, C_2. Choose M_t of the form $M_t = f(X_t)e^{-\lambda t}$ such that $f(a) = f(b)$, say

$$M_t = (e^{\beta b} - e^{\beta a})e^{\alpha X_t - \lambda t} + (e^{\alpha a} - e^{\alpha b})e^{\beta X_t - \lambda t}.$$

With this choice, M_t is bounded on $[0, T]$, and so the OST implies

$$f(0) = \mathbb{E}[M_0] = \mathbb{E}[M_T] = \mathbb{E}[f(a)e^{-\lambda T}]$$

so

$$\mathbb{E}[e^{-\lambda T}] = \frac{e^{\beta b} - e^{\beta a} + e^{\alpha a} - e^{\alpha b}}{e^{\beta b + \alpha a} - e^{\beta a + \alpha b}}$$

In the special case $c = 0$ and $a = -b$ this reduces to $\mathbb{E}[e^{-\lambda T}] = \operatorname{sech}(b\sqrt{2\lambda})$.

2.3.3 Law of the Iterated Logarithm.

$$\mathbb{P}\left(\limsup_{t \searrow 0} \frac{B_t}{\sqrt{2t \log\log(\frac{1}{t})}}\right) = 1$$

PROOF: Write $h(t) = \sqrt{2t \log\log(\frac{1}{t})}$. The first step is to show that

$$\limsup_{t \searrow 0} \frac{B_t}{h(t)} \le 1$$

$\mathbb{P}$-a.s. Apply Doob's maximal inequality to the exponential martingale $Z_t = \exp(\alpha B_t - \frac{\alpha^2}{2} t)$ yielding for $\alpha > 0$

$$\mathbb{P}\left(\sup_{s \in [0,t]} (B_t - \tfrac{1}{2}\alpha s) > \beta\right) = \mathbb{P}\left(\sup_{s \in [0,t]} Z_t > e^{\alpha\beta}\right) \le e^{-\alpha\beta}\, \mathbb{E}[Z_t] = e^{-\alpha\beta}.$$

Now fix $\theta, \delta \in (0,1)$ and apply the inequality with $t = \theta^n$, $\alpha = \theta^{-n}(1+\delta)h(\theta^n)$, and $\beta = \frac{1}{2}h(\theta^n)$. Then

$$\alpha\beta = \frac{1}{2}(1+\delta)\theta^n h^2(\theta^n) = (1+\delta)\log\log\left(\frac{1}{\theta}\right)^n$$

and $e^{\alpha\beta} = \log(n\log(\frac{1}{\theta}))^{1+\delta} = O(n^{1+\delta})$. So

$$\sup_{s \in [0,\theta^n]} \mathbb{P}(B_s - \frac{1}{2}s(1+\delta)\theta^{-n}h(\theta^n)) \ge \frac{1}{2}h(\theta^n)) \le Cn^{-(1+\delta)}.$$

By the Borel-Cantelli Lemma there is $\Theta'_{\theta,\delta} \in \mathscr{F}$ with $\mathbb{P}\Omega' = 1$ such that for all $\omega \in \Omega'$ there is $N_{\theta,\delta}(\omega)$ such that for all $n \ge N_{\theta,\delta}(\omega)$

$$\max_{x \in [0,\theta^n]} (B_s - \frac{1}{2}s(1+\delta)\theta^{-n}h(\theta^n)) < \frac{1}{2}h(\theta^n)$$

Thus for $\theta^{n+1} < t \le \theta^n$

$$B_t \le \sup_{s \in [0,\theta^n]} B_s \le \frac{1}{2}(2+\delta)\theta^{-n}h(\theta^n) \le \frac{1}{2}(2+\delta)\theta^{-\frac{1}{2}}h(t)$$

where the last inequality uses the fact that $h(\theta^n) \le \theta^{-\frac{1}{2}}h(\theta^{n+1}) \le \theta^{-\frac{1}{2}}h(t)$, so

$$\limsup_{t \searrow 0} \frac{B_t}{t} \le (1 + \tfrac{\delta}{2})\theta^{-\frac{1}{2}}.$$

Letting $\delta \searrow 0$ and $\theta \nearrow 1$ along countable sequences we complete the proof of the first step.

For the second step, see KS p.112–113 □

Recall the generator of a Markov process is $\mathscr{G} = \lim_{t \searrow 0} \frac{P_t - I}{t}$ where $P_t f(x) = \mathbb{E}[f(X_t) \,|\, X_0 = x]$. The Brownian transition function is

$$p(t; x, y) = \frac{1}{(2\pi t)^{\frac{d}{2}}} \exp(-\frac{\|x-y\|^2}{2t})$$

and satisfies $\frac{\partial p_t}{\partial t} = \frac{1}{2}\Delta p_t(x)$. For Brownian motion, check that $\mathscr{G}f(t,x) = \frac{\partial f}{\partial t} + \frac{1}{2}\Delta f$. Define

$$C_t^f = f(t, B_t) - f(0, B_t) - \int_0^t \mathscr{G}f(s, B_s)ds.$$

If $\{B_t, \mathscr{F}_t\}$ is a Brownian motion and $f \in C^{1,2}$ then C_t^f is an $\{\mathscr{F}_t\}$-martingale.

2.4 Lévy processes

2.4.1 Definition. A *Poisson process* $\{N_t, \mathscr{F}_t\}$ with intensity λ is a right continuous $\{\mathscr{F}_t\}$-adapted process with $N_0 = 0$ and $N_t - N_s \sim \text{Poisson}(\lambda(t-s))$, i.e. for $i = 0, 1, \dots,$

$$\mathbb{P}[N_t - N_s = i] = e^{-\lambda(t-s)} \frac{\lambda^i (t-s)^i}{i!}$$

The Poisson process has stationary independent increments and $\{N_t - \lambda t\}$ and $\{e^{\alpha N_t - \lambda t(e^{\alpha}-1)}\}$, for any $\alpha \in \mathbb{R}$, are martingales. This is because $\mathbb{E}[e^{\alpha N_t}] = e^{\lambda t(e^{\alpha}-1)}$ is the moment generating function for the Poisson distribution.

2.4.2 Definition. A *Lévy process* is a right-continuous process with stationary independent increments.

2.4.3 Examples.
(i) Brownian motion (with drift)
(ii) Poisson process
(iii) T_a, the hitting time of Brownian motion to a level a.

2.4.4 Definition. A probability measure μ on $\mathbb{R}$ is said to be *infinitely divisible* if for all n there is a probability measure ν on $\mathbb{R}$ such that $\mu = \nu^{*n}$. Equivalently, if $Y \sim \mu$ then for every n there are i.i.d. r.v.'s $Y_i \sim \nu$ such that $Y = \sum_{i=1}^n Y_i$.

If $\{X_t, \mathscr{F}_t\}$ is a Lévy process with $X_0 = 0$ then for all t, X_t is infinitely divisible since it may be written as a sum of n i.i.d. increments,

$$X_t = \sum_{i=1}^n (X_{t\frac{i}{n}} - X_{t\frac{i-1}{n}}).$$

Conversely, given any infinitely divisible r.v. Y there is a Lévy process $\{X_t, \mathscr{F}_t\}$ such that $Y \stackrel{(d)}{=} X_1$.

Analytical methods can be used to show that if μ is infinitely divisible then its Fourier transform is equal to $e^{\Psi(\theta)}$, where

$$\Psi(\theta) = \overbrace{i\beta\theta}^{\text{drift}} - \overbrace{\frac{1}{2}\sigma^2\theta^2}^{BM} + \underbrace{\int \left(\overbrace{e^{i\theta x} - 1}^{\text{Poisson}} - \frac{i\theta x}{1+x^2} \right) \nu(dx)}_{\text{pure jump process}}$$

and ν is a Radon measure on $\mathbb{R}\backslash\{0\}$ such that $\int \frac{x}{1+x^2}\nu(dx) < \infty$. This immediately gives you a complex exponential martingale associated with a Lévy process. This is the *Lévy-Khintchine formula*

2.4.5 Definition. A r.v. Y is *stable* if for all n there are independent r.v.'s with the same law as Y and constants $a_n > 0$ and b_n such that $Y_1 + \cdots + Y_n \stackrel{(d)}{=} a_n Y + b_n$.

2.4.6 Lemma. *Stable r.v.'s are infinitely divisible.*

2.4.7 Exercise. It must be the case that $a_n = n^{\frac{1}{\alpha}}$ for some $\alpha \in (0,2]$. When $\alpha = 2$ we get the Gaussian distribution. When $\alpha \in (0,2]$ then $\sigma = 0$ in the L-K formula and the Lévy measure has density

$$(m_1 \mathbf{1}_{\{x<0\}} + m_2 \mathbf{1}_{\{x>0\}})|x|^{-(1+\alpha)}$$

for some $m_1, m_2 \geq 0$.

2.5 Doob-Meyer decomposition

2.5.1 Lemma. *Any non-constant continuous martingale* $\{M_t\}$ *a.s. has infinite variation.*

PROOF: Let V_t be the variation of M on $[0,t]$ and define

$$S_n := \inf\{s \geq 0 \mid V_s \geq n\} \wedge \inf\{s \geq 0 \mid |M_s| \geq n\}.$$

Then the stopped process M^{S_n} is of bounded variation and is a martingale by the OST. Therefore it is enough to prove that M is constant whenever it and its variation are bounded. Assume further that $M_0 = 0$ a.s. Fix $t < \infty$ and let $\Pi = \{0 = t_0 < t_1 < \cdots < t_k = t\}$ be a subdivision of $[0,t]$. Then

$$\mathbb{E}[M_t^2] = \mathbb{E}[\sum_{i=0}^{k-1}(M_{t_{i+1}}^2 - M_{t_i}^2)] = \mathbb{E}[\sum_{i=0}^{k-1}(M_{t_{i+1}} - M_{t_i})^2]$$

As a result

$$\mathbb{E}[M_t^2] \leq \mathbb{E}[V_t \sup_i |M_{t_{i+1}} - M_{t_i}|]$$

and since M is of bounded variation and it is continuous, this quantity goes to zero as the mesh of the partition goes to zero, so $M \equiv 0$. □

Remark. Continuity is required for this proof. The compensated Poisson process is a right continuous process of bounded variation (it is seen to be of bounded variation since it is the difference of two increasing processes).

2.5.2 Definition. A_n is said to be an *increasing sequence* if $0 = A_n \leq A_1 \leq \cdots$ $\mathbb{P}$-a.e. and $\mathbb{E}[A_n] < \infty$ for all $n \geq 0$. A_t is said to be an *increasing sequence* if $0 = A_0$, $t \mapsto A_t$ is non-decreasing $\mathbb{P}$-a.e., right continuous and $\mathbb{E}A_t < \infty$ for all $t \geq 0$. Such a thing is said to be *integrable* if $\mathbb{E}[A_\infty] < \infty$.

2.5.3 Definition. In discrete time, a sequence is said to be *natural* if for every bounded martingale $\{M_n\}$,

$$\mathbb{E}[M_n A_n] = \mathbb{E}[\sum_{k=1}^{n} M_{k-1}(A_n - A_{n-1})]$$

Let $Y_n = \sum_{k=1}^{n} M_{k-1}(A_n - A_{n-1})$. Then a sequence A is natural if and only if $\mathbb{E}[Y_n] = 0$ for all n, if and only if $\{A_n\}$ is predictable.

2.5.4 Definition. In continuous time, A is natural if for all bounded martingales

$$\mathbb{E}[M_t A_t] = \mathbb{E}[\int_{(0,t]} M_{s-} dA_s].$$

2.5.5 Lemma.

$$\mathbb{E}[\int_{(0,t]} M_s dA_s = \mathbb{E}[\int_{(0,t]} M_{s-} dA_s]$$

It will be a consequence of the definition of the stochastic integral that $\int_{(0,t]} A_s dM_s = M_t A_t - \int_{(0,t]} M_{s-} dA_s$ and so $\mathbb{E}[(A \cdot M)_t] = 0$.

2.5.6 Proposition. *(In discrete time) an increasing random sequence is predictable if and only if it is natural.*

PROOF: Suppose that A is natural and M is a bounded martingale. Let Y_n be the martingale transform, as defined above. Then

$$\mathbb{E}[A_n(M_n - M_{n-1})] = \mathbb{E}[Y_n] - \mathbb{E}[Y_{n-1}] = 0 \qquad \square$$

2.5.7 Definition. An increasing process $\{A_t\}$ is said to be *natural* if

$$\mathbb{E}[\int_{(0,t]} M_s dA_s] = \mathbb{E}[\int_{(0,t]} M_{s-} dA_s]$$

for all bounded martingale $\{M_t \mathscr{F}_t\}$.

This is analogous to the discrete time version because

$$\mathbb{E}[M_t A_t] = \mathbb{E}[\int_{(0,t]} M_{s-} dA_s]$$

(see text for the approximation argument). An increasing process is increasing if and only if it is natural. See notes for the proof of existence.

2.5.8 Definition. A right continuous process X is of *class D* (resp. *class DL*) if $\{X_\tau\}_{\tau\in S}$ (resp. $\{X_\tau\}_{\tau\in S_a}$ for all $a \in \mathbb{R}_+$) is u.i., where S is the set of all finite stopping times (resp. S_a is the set of all stopping times bounded by a).

2.5.9 Theorem. *If X is a right continuous sub-martingale of class DL, then $X = M + A$, where M is a martingale and A is a natural increasing (predictable) process. If X is class D then M is u.i. and A is integrable.*

2.5.10 Definition. A *local martingale* is a right continuous process M such that there exists a *localizing sequence* of stopping times $\{T_n\}$ with $T_n \nearrow \infty$ a.s. and such that $(M - M_0)^{T_n}$ is a martingale for all n.

2.5.11 Theorem. *A process X is a local sub-martingale if and only if it has a decomposition $X = M + A$, where M is a local martingale and A is a locally integrable increasing process. The decomposition is unique when A is required to be predictable/natural.*

2.5.12 Definition. A process X is a *semi-martingale* if $X = M + A$, where M is a local martingale and A is locally of finite variation.

2.5.13 Definition. A process X is said to be *regular* if for all $a > 0$ and every non-decreasing sequence of stopping times $\{T_n\}$ bounded by a, if $T = \lim_{n\to\infty} T_n$ then $\lim_{n\to\infty} \mathbb{E}[X_{T_n}] = \mathbb{E}[X_T]$.

A continuous sub-martingale is regular.

2.5.14 Theorem. *For a right continuous sub-martingale X of class DL, the compensator is continuous if and only if X is regular.*

2.5.15 Lemma. *Non-negative sub-martingales are of class DL.*

PROOF: Fix $a > 0$ and suppose that T is such that $\mathbb{P}[T \le a] = 1$. Then apply OST to $\{X_{T\wedge a}\}$ to get $\mathbb{E}[X_a \mid \mathscr{F}_T] \ge X_T$. Multiply both sides by $\mathbf{1}_{\{X_T>\lambda\}}$ and take expectations to get

$$\mathbb{E}[X_T \mathbf{1}_{\{X_T>\lambda\}}] \le \mathbb{E}[X_a \mathbf{1}_{\{X_T>\lambda\}}].$$

Since $X_a \in L^1$ and

$$\mathbb{P}[X_T > \lambda] \le \frac{1}{\lambda}\mathbb{E}[X_T] \le \frac{1}{\lambda}\mathbb{E}[X_a] \to 0 \text{ as } \lambda \to \infty,$$

it follows that $\{X_T\}_{T\in S_a}$ is u.i. $\square$

3 Stochastic Integration

Naive stochastic integration (i.e. via Riemann sums) is impossible. Let X be a right continuous function on $[0,1]$ and Π_n be a refining sequence of dyadic rationals such that $\|\Pi_n\| \to 0$. What conditions are needed on X so that the sums $S_n = \sum_{\Pi_n} h(t_k)(x(t_{k+1}) - x(t_k))$ converge to a finite limit for all continuous h?

3.0.16 Theorem. *Finite variation is necessary.*

PROOF: Let $X = C[0,\infty)$ and $Y = \mathbb{R}$, and for $h \in X$ let

$$T_n(h) = \sum_{\Pi_n} h(t_k)(x(t_{k+1}) - x(t_k)).$$

Construct h_n in X such that $h_n(t_k) = \operatorname{sgn}(x(t_{k+1}) - x(t_k))$ over Π_n and $\|h_n\| = 1$. For such an h_n we have $T_n(h_n) = \sum_{\Pi_n} |x(t_{k+1}) - x(t_k)|$, so $\|T_n\| \geq \operatorname{Var}_{[0,1]}(x)$. On the other hand, if for all $h \in X$ $\lim_{n\to\infty} T_n(h)$ exists then by the Banach-Steinhaus theorem the total variation over $[0,1]$ of x is finite. □

3.1 Riemann-Stieltjes Integration

Let $\mathscr{FV}$ be the class of finite variation processes (differences of increasing processes) started at 0.

3.1.1 Theorem. *Let $A \in \mathscr{FV}$ and H be a (jointly) measurable process such that a.s. $s \mapsto H(s,\omega)$ is continuous. Let Π_n be a sequence of random finite partitions of $[0,t]$ such that $\lim_{n\to\infty} \|\Pi_n\| \to 0$. Then for and $\{S_k\}$ with $T_k \leq S_k \leq T_{k+1}$, a.s.*

$$\lim_{n\to\infty} \sum_{\Pi_n} H_{S_k}(A_{T_{k+1}} - A_{T_k}) = \int_0^t H_s \, dA_s.$$

3.1.2 Theorem. *Let $A \in \mathscr{FV}$ be right continuous . For $f \in C^1$, the process $(f(A_t))_{t\geq 0}$ is in $\mathscr{FV}$ and is equal to*

$$\int_0^t f'(A_s)\, dA_s + \sum_{0<s\leq t} \Delta f(A_s) + f'(A_{s-})\Delta A_s.$$

3.1.3 Example. Let N be a Poisson process of parameter λ and $M_t = N_t - \lambda t$ be the compensated Poisson process. Let H be jointly measurable and (say) bounded. The natural way to define the integral of H with respect to M is

$$I_t^M(H) = \int_0^t H_s \, dM_s = \int_0^t H_s \, d(N_s - \lambda t) = \int_0^t H_s \, dN_s - \lambda \int_0^t H_s \, ds.$$

Let $\{T_i\}$ be the jump times of the Poisson process, so that $N_t = \sum_{i=0}^\infty \mathbf{1}_{T_i \leq t}$. Then

$$I_t^M(H) = \sum_{n=1}^\infty H_{T_n} \mathbf{1}_{T_n \leq t} - \lambda \int_0^t H_s \, ds.$$

If H is continuous and adapted, then

$$\mathbb{E}[I_t^M(H) - I_s^M(H) \mid \mathscr{F}_s] = \mathbb{E}[\int_s^t H_u dM_u \mid \mathscr{F}_s] = 0$$

applying the first theorem in the section, so the integral is seen to be a martingale. What if H is not continuous but right continuous? Let $\tilde{H}_t = \mathbf{1}_{[0,T_1)}(t)$, so

$$\int_0^t H_s dM_s = \sum_{i=1}^{\infty} \tilde{H}_{T_i} \mathbf{1}_{T_i \le t} - \lambda \int_0^t \tilde{H}_s ds = -\lambda(t \wedge T_1),$$

which is not a martingale.

Now for the harder case of continuous martingales (necessarily of unbounded variation). Let B be standard Brownian motion and consider a sequence $\{\Pi_n\}$ of dyadic partitions of $[0, \infty)$ with $\|\Pi_n\| \to 0$ as $n \to \infty$. Define

$$B_t^{(n)} = \sum_{\Pi_k} B_{t_k} \mathbf{1}_{(t_k, t_{k+1}]}.$$

We know $B^{(n)}$ is càglàd, and $B^{(n)} \to B$ u.c.p. i.e. for all T,

$$\sup_{t \in [0,T]} |B_t^{(n)} - B_t| \xrightarrow{\text{(p)}} 0$$

as $n \to \infty$. The martingale transform is

$$\begin{aligned}
I_t^B(B^{(n)}) &= \sum_{\Pi_n} B_{t_k}(B_{t_{k+1}} - B_{t_k}) \\
&= \sum_{\Pi_n} \frac{1}{2}(B_{t_{k+1}} + B_{t_k})(B_{t_{k+1}} - B_{t_k}) - \frac{1}{2}(B_{t_{k+1}} - B_{t_k})(B_{t_{k+1}} - B_{t_k}) \\
&= \frac{1}{2}B_t^2 - \frac{1}{2}\sum_{\Pi_n}(B_{t_{k+1}} - B_{t_k})^2
\end{aligned}$$

This converges u.c.p. to $\frac{1}{2}B_t^2 - \frac{1}{2}t$. Note that this is *not* what we would expect from the usual change of variable formula. There is an extra term of $-\frac{1}{2}t$,

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t.$$

3.2 Construction of the Itô integral

See text, §3.1 and §3.2.

3.2.1 Definition. A process X is called *simple* if there is a strictly increasing sequence of real numbers $\{t_n\}_{n\geq 0}$ with $t_0 = 0$ and $\lim_{n\to\infty} t_n = \infty$ and a sequence of random variables $\{\xi_n\}_{n\geq 0}$ with $\sup_{n\geq 0} |\xi_n| \leq C < \infty$ and ξ_n is $\mathscr{F}_{t_n}$-measurable for all n and

$$X_t = \xi_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{\infty} \xi_i \mathbf{1}_{(t_i, t_{i+1}]}(t)$$

This class of processes will be denoted $\mathscr{L}_0$. We define the stochastic integral of $X \in \mathscr{L}_0$ with respect to M by the martingale transform

$$I_t^M(X) = \sum_{i=0}^{\infty} \xi_i (M_{t\wedge t_{i+1}} - M_{t\wedge t_i}).$$

Some obvious properties of I_t^M for simple processes are

(i) $I_0^M(X) = 0$;

(ii) $I_t^M(\alpha X + \beta Y) = \alpha I_t^M(X) + I_t^M(Y)$ (linearity);

(iii) $\{I_t^M(X), \mathscr{F}_t\}_{t\geq 0}$ is a martingale.

(iv) $\mathbb{E}[(I_t^M(X))^2] = \mathbb{E}[\int_0^t X_s^2 d[M]_s]$ (see below)

Proof (of 4.):

$$\begin{aligned} \mathbb{E}[(I_t^M(X))^2] &= \text{fill in all details as an exercise} \\ &= \mathbb{E}[\int_0^t X_s^2 d[M]_s] \end{aligned}$$ □

It follows from the last property that

$$\mathbb{E}[(I_t^M(X))^2 - (I_s^M(X))^2 \mid \mathscr{F}_s] = \mathbb{E}[\int_s^t X_s^2 d[M]_s],$$

so if $X \in \mathscr{L}_0$ and $M \in \mathscr{M}_2$ then $I^M(X) \in \mathscr{M}_2$ and

$$\|I^M(X)\| = [X] = \sum_{n=0}^{\infty} \frac{1}{2^n} 1 \wedge [X]_n,$$

where $[X]_n^2 = \mathbb{E}[\int_0^n X_s^2 d[M]_s]$.

It is a fact that $(\mathscr{M}_2, \|\cdot\|)$ and $(\mathscr{L}, [\cdot])$ are complete metric spaces.

3.2.2 Lemma. *Let X be a bounded, measurable, adapted process. Then there is a sequence $\{X^{(m)}\}_{m\geq 1}$ of simple processes such that*

$$\sup_{T>0} \lim_{m\to\infty} \mathbb{E}\left[\int_0^T |X_t^{(m)} - X_t|^2 dt\right] = 0.$$

Proof: Fix $T > 0$.

(i) Suppose that X has continuous paths. Then X can be approximated by

$$X_t^{(m)} = X_0 \mathbf{1}_{\{0\}}(t) + \sum_{k=1}^{\infty} X_{k2^{-m}} \mathbf{1}_{(k2^{-m},(k+1)2^{-m}]}(t).$$

Indeed, $X^{(m)} \to X$ a.s. as $m \to \infty$ and by the BCT $[X - X^{(m)}]_T \to 0$.

(ii) Suppose that X is progressively measurable. Define, for $t \in [0, T]$,

$$F_t = \int_0^t X_s ds \quad \text{and} \quad \tilde{X}_t^{(m)} = m(F_t - F_{t-\frac{1}{m}}).$$

Note that for F to be well-defined we require only that X is measurable. F is continuous, so $\tilde{X}^{(m)}$ is continuous for all m. Progressive measurability of F follows from that of X. Indeed, if

$$g : ([0, t] \times \Omega, \mathscr{B}[0, t] \otimes \mathscr{F}_t) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

is measurable then $\int_0^t g(s, \omega) ds$ is also $(\mathscr{B}[0, t] \otimes \mathscr{F}_t)$-measurable by Fubini(?). Therefore F is continuous and progressively measurable. For all $\omega \in \Omega$, $X_t^{(m)}(\omega) \to X_t(\omega)$ by the fundamental theorem of calculus. The BCT gives you the rest. Standard diagonalization gives you that X can be approximated by a sequence in $\mathscr{L}_0$.

(iii) Let X be measurable and adapted. As before, F is continuous and measurable but we can not be sure that it is progressively measurable. We will show that F is indeed adapted. By KS Proposition 1.1.2, X has a progressively measurable modification Y. Let $G_t = \int_0^t Y_s ds$ for $t \in [0, T]$. We know from the second part that G is $\mathscr{F}_t$-adapted. It suffices to show that F is a modification of G since the filtration is complete. Fix $t \in [0, T]$.

$$\{F_t \neq G_t\} \subseteq \{\int_0^T \mathbf{1}_{X_t \neq Y_t} dt > 0\}$$

so

$$\mathbb{P}[F_t \neq G_t] \le \int_0^T \mathbb{P}[X_t \neq Y_t] dt = 0.$$

Since $\{\mathscr{F}_t\}$ is complete, F_t is adapted. Proceed as in the previous step. □

3.2.3 Proposition. *If $t \mapsto [M]_t$ is absolutely continuous then $\mathscr{L}_0$ is dense in $(\mathscr{L}(M), [\cdot]_M)$.*

PROOF: If $X \in \mathscr{L}$ is bounded then the assertion essentially follows from the previous lemma. Choose a subsequence of $\{X^{(m)}\}$ along which

$$\left\{ \lim_{k\to\infty} X^{(m_k)} = X \right\}^C$$

has zero μ_B measure, and therefore zero μ_M-measure. By BCT we have convergence in $[\cdot]_M$. If not bounded then use DCT instead (truncation, take limits). □

Integrand	Integrator
$\mathscr{L}(M)$ (meas, adapted)	$M \in \mathscr{M}_2^c$, $t \mapsto [M]_t$ a.c. (Itô)
$\mathscr{L}^*(M)$ (prog meas)	$M \in \mathscr{M}_2^c$ (Itô)
$\mathscr{P}(M)$ (predictable)	$M \in \mathscr{M}_2$ (Kunita-Watanabe)

Using localization we can replace $\mathscr{M}_2^c$ by $\mathscr{M}^{c,loc}$ and replace

$$\mathbb{E}[\int_0^T X_s^2 d[M]_s] < \infty \quad \text{by} \quad \mathbb{P}[\int_0^T X_s^2 d[M]_s] < \infty.$$

We must deal with questions like whether $I(X^T) = (I(X))^T$ for stopping times T.

3.3 Characterization of the Stochastic Integral

For $M \in \mathscr{M}_2^c$, $X \in \mathscr{L}^*(M)$ we have shown that $I_t^M(X) = \int_0^t X_s dM_s$ is well-defined. We know what $I^M(X) \in \mathscr{M}_2^c$ with quadratic variation $\int_0^t X_s^2 d[M]_s$.

What is the cross variation of $I^M(X)$ and $I^Y(N)$? Recall that the cross variation may be characterized as the unique predictable process (of finite variation) such that

$$I^M(X)I^N(Y) - [I^M(X), I^N(Y)]$$

is a martingale. First, when X and Y are simple, suppose without loss of generality that

$$X = \xi_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{\infty} \xi_i \mathbf{1}_{(t_i, t_{i+1}]}(t)$$

and

$$Y = \eta_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=1}^{\infty} \eta_i \mathbf{1}_{(t_i, t_{i+1}]}(t).$$

Remember that

$$I_t^M(X) = \sum_{i=1}^{\infty} \xi_i (M_{t_{i+1} \wedge t} - M_{t_i \wedge t}) \quad \text{and} \quad I_t^N(Y) = \sum_{i=1}^{\infty} \eta_i (N_{t_{i+1} \wedge t} - N_{t_i \wedge t}).$$

Fix $0 \le s \le t < t$ and suppose that n and m are such that $t_m \le s < t_{m+1}$ and $t_n \le t < t_{n+1}$ (in fact, suppose for now that $s = t_m$ and $t = t_{m+1}$).

$$\begin{aligned}
&\mathbb{E}[(I_t^M(X) - I_s^M(X))(I_t^N(Y) - I_s^N(Y)) \mid \mathscr{F}_s] \\
&\qquad = \mathbb{E}[\sum_{i=m}^{n} \sum_{j=m}^{n} \xi_i \eta_j (M_{t_{i+1}} - M_{t_i})(N_{t_{j+1}} - N_{t_j}) \mid \mathscr{F}_s] \\
&\qquad = \mathbb{E}[\sum_{i=m}^{n} \xi_i \eta_i (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}) \mid \mathscr{F}_s]
\end{aligned}$$

$$= \sum_{i=m}^{n} \mathbb{E}[\xi_i \eta_i (M_{t_{i+1}} - M_{t_i})(N_{t_{i+1}} - N_{t_i}) \mid \mathscr{F}_s]$$

$$= \sum_{i=m}^{n} \mathbb{E}[\xi_i \eta_i \, \mathbb{E}[M_{t_{i+1}} N_{t_{i+1}} - M_{t_i} N_{t_i} \mid \mathscr{F}_{t_i}] \mid \mathscr{F}_s]$$

$$= \sum_{i=m}^{n} \mathbb{E}[\xi_i \eta_i \, \mathbb{E}[[M,N]_{t_{i+1}} - [M,N]_{t_i} \mid \mathscr{F}_{t_i}] \mid \mathscr{F}_s]$$

$$= \mathbb{E}[\sum_{i=m}^{n} \xi_i \eta_i ([M,N]_{t_{i+1}} - [M,N]_{t_i}) \mid \mathscr{F}_s]$$

$$= \mathbb{E}[\int_s^t X_u Y_u d[M,N]_u \mid \mathscr{F}_s]$$

3.3.1 Proposition. *Let α, β, γ be right continuous functions $[0,\infty) \to \mathbb{R}$ with $\alpha(0) = \beta(0) = \gamma(0) = 0$. Let α be of finite variation and β and γ be increasing. Suppose further that for all $s \le t$ we have*

$$\left| \int_s^t d\alpha_u \right| \le \left(\int_s^t d\beta_u \right)^{\frac{1}{2}} \left(\int_s^t d\gamma_u \right)^{\frac{1}{2}}.$$

Then for any measurable functions f, g we have

$$\int_s^t |fg| d|\alpha| \le \left(\int_s^t f^2 d\beta \right)^{\frac{1}{2}} \left(\int_s^t g^2 d\gamma \right)^{\frac{1}{2}}.$$

PROOF: Monotone class theorem. □

3.3.2 Theorem (Kunita-Watanabe inequality, 1967).
If $M, N \in \mathscr{M}_2^c$, $X \in \mathscr{L}^*(M)$, $Y \in \mathscr{L}^*(N)$ then a.s.

$$\int_0^t |X_s Y_s| d|[M,N]_s| \le \left(\int_0^t X_s^2 d[M]_s \right)^{\frac{1}{2}} \left(\int_0^t Y_s^2 d[N]_s \right)^{\frac{1}{2}}.$$

PROOF: By the previous result we only need to show that there is a negligible set Z such that

$$\int_s^t d|[M,N]_u| \le \left(\int_s^t d[M]_u \right)^{\frac{1}{2}} \left(\int_s^t d[N]_u \right)^{\frac{1}{2}}$$

holds path-wise for all s, t. Let Z be the null set such that if $\omega \notin Z$ then

$$0 \le \int_s^t d[M + rN, M + rN]_u$$

for all r, s, t with $s \le t$ and $r, s, t \in \mathbb{Q}$. Then

$$0 \le \int_s^t d[M + rN, M + rN]_t - \int_s^t d[M + rN, M + rN]_s$$

$$= r^s([N]_t - [N]_s) + 2r([N,M]_t - [N,M]_s) + ([M]_t - [M]_s)$$

The right hand side is non-negative for all rational r, so it holds for all real r by continuity. The discriminant of the quadratic equation must be non-negative, which gives us the desired inequality. Since we have it for rational s, t, by right continuity of the paths we have it for all s, t. □

3.3.3 Lemma. *If $M, N \in \mathscr{M}_2^c$, $X \in \mathscr{L}^*(M)$, $\{X^{(n)}\}_{n=1}^{\infty} \subseteq \mathscr{L}^*(M)$, with*

$$\lim_{n\to\infty} \int_0^T |X_s^{(n)} - X_s|^2 d[M]_s = 0$$

a.s.-$\mathbb{P}$, then, for all $0 \le t \le T$,

$$\lim_{n\to\infty} [I^M(X^{(n)}), N]_t = [I^M(X), N]_t.$$

PROOF:

$$|[]_t| \le []_t []_t \le \int_0^T []_T$$ □

3.3.4 Lemma. *If $M, N \in \mathscr{M}_2^c$ and $X \in \mathscr{L}^*(M)$ then*

$$[I^M(X), N]_t = \int_0^t X_s d[M,N]_s.$$

PROOF: We showed there is a sequence of simple processes such that the condition in the above lemma holds. But we showed that the condition in this lemma holds for simple processes. □

3.3.5 Theorem. *Consider a martingale $M \in \mathscr{M}_2^c$ and $X \in \mathscr{L}^*(M)$. Then $I^M(X)$ is the unique martingale $\Phi \in \mathscr{M}_2^c$ such that*

$$[\Phi, N]_t = \int_0^t X_s d[M,N]_s$$

for all $N \in \mathscr{M}_2^c$.

3.3.6 Corollary. *If $M \in \mathscr{M}_2^c$, $X \in \mathscr{L}^*(M)$, $N = I^M(X)$, $Y \in \mathscr{L}^*(N)$, then $XY \in \mathscr{L}^*(M)$ and $I^N(Y) = I^M(XY)$.*

PROOF: $[N]_t = \int_0^t X_s^2 d[M]_s$, so

$$\mathbb{E}[\int_0^T X_s^2 Y_s^2 d[M]_s] = \mathbb{E}[\int_0^T Y_s^2 d[N]_s] < \infty.$$

So $X, Y \in \mathscr{L}^*(M)$. For any $\tilde{N} \in \mathscr{M}_2^c$, the previous theorem showed that

$$d[N,\tilde{N}]_s = X_s d[M,\tilde{N}]_s$$

and so

$$[I^M(XY),\tilde{N}]_t = \int_0^t X_s Y_s d[M,\tilde{N}]_s = \int_0^t Y_s d[N,\tilde{N}]_s = [I^N(Y),\tilde{N}]_t.$$

By the characterization of the integral, $I^M(XY) = I^N(Y)$. □

Today we have shown that $[I^M(X), I^N(Y)] = \int_0^t X_s Y_s d[M,N]_s$.
Read the proof of Itô's formula.

3.4 Stochastic Integration

Today we extend the definition of the stochastic integral to all of $\mathscr{M}^{c,loc}$. For $M \in \mathscr{M}_2^c$ with $t \mapsto [M]_t$ absolutely continuous and $X \in \mathscr{L}^*(M)$. We used the fact that for all $T < \infty$ there is a sequence $\{X^{(m)}\} \subseteq \mathscr{L}_0$ such that

$$\mathbb{E}[\int_0^T |X_t^{(m)} - X_t|^2 dt] \to 0$$

as $m \to \infty$. We use "time changes" to do the general case.

3.4.1 Theorem. *For $M \in \mathscr{M}_2^c$, $\mathscr{L}_0(M)$ is dense in $\mathscr{L}^*(M)$ with respect to $[\cdot]$.*

PROOF: The proof in the general case follows from the following more general lemma. □

3.4.2 Lemma. *Let $\{A_t\}$ is a continuous (resp. right-continuous) increasing, $\mathscr{F}$-adapted process. If X is progressively measurable and satisfies*

$$\mathbb{E}[\int_0^T X_t^2 dA_t] < \infty$$

for all $T > 0$, then there exists a.s. a sequence $\{X^{(n)}\}_{n=1}^\infty$ of simple processes such that

$$\sup_{T>0} \lim_{n\to\infty} \mathbb{E}[\int_0^t |X_t^{(n)} - X_t|^2 dA_t] = 0.$$

PROOF: Assume without loss of generality that X is bounded, say by C. It suffices to fix $T > 0$ and show there exists $\{X^{(n)}\}_{n=1}^\infty \subseteq \mathscr{L}_0$ such that

$$\lim_{n\to\infty} \mathbb{E}[\int_0^t |X_t^{(n)} - X_t|^2 dA_t] = 0.$$

The process $A_t + t$ is strictly increasing and continuous, so it has a continuous strictly inverse function T_s defined by $A_{T_s} + T_s = s$ for all ω. In particular, $T_s \le s$ and

$$\{T_s \le t\} = \{A_t + t \ge s\} \in \mathscr{F}_t.$$

Therefore for all $s \ge 0$, T_s is an $\mathscr{F}$-stopping time. Define $\mathscr{G}_s = \mathscr{F}_{T_s}$ and $Y_s = X_{T_s}$. Since X is progressively measurable, Y_s is $\mathscr{G}$-adapted. Without loss of generality, assume $X_t \equiv 0$ for $t \ge T$. Also,

$$\begin{aligned}
\mathbb{E}[\int_0^\infty Y_s^2 ds &= \mathbb{E}[\int_0^\infty \mathbf{1}_{\{T_s \le T\}} X_{T_s}^2 ds] \\
&= \mathbb{E}[\int_0^{A_T+T} \mathbf{1}_{\{T_s \le T\}} X_{T_s}^2 ds] \\
&\le C(\mathbb{E}[A_T] + T) < \infty.
\end{aligned}$$

For any $N \in \mathbb{N}$, choose $R < \infty$ such that $\mathbb{E}[\int_R^\infty Y_s^2 ds] < \frac{1}{2n}$. By the old result, there is a simple process $\tilde{Y}^{(n)}$ such that

$$\mathbb{E}[\int_0^R |\tilde{Y}_s^{(n)} - Y_s|^2 ds] < \frac{1}{2n}.$$

Define $Y_s^{(n)} = \mathbf{1}_{[0,R]}(s)\tilde{Y}_s^{(n)}$. Then

$$\mathbb{E}[\int_0^R |Y_s^{(n)} - Y_s|^2 ds] < \frac{1}{n}.$$

But

$$Y_s^{(n)} = \xi_0 \mathbf{1}_{\{0\}}(s) + \sum_{i \ge 0} \xi_i \mathbf{1}_{(s_i, s_{i+1}]}(s)$$

where each ξ_i is $\mathscr{G}_{s_i}$-measurable. Define

$$X_t^{(n)} = Y_{t+A_t}^{(n)} = \xi_0 \mathbf{1}_{\{0\}}(s) + \sum_{i \ge 0} \xi_i \mathbf{1}_{(T_{s_i}, T_{s_{i+1}}]}(s).$$

To see that $X_t^{(n)}$ is $\mathscr{F}_t$-adapted, simply observe that ξ_i restricted to $(T_{s_i}, T_{s_{i+1}}]$ is $\mathscr{F}_t$-measurable. □

Now we extend to $\mathscr{M}^{c,loc}$. For simplicity we assume that $M_0 = 0$. Define $T_n := \inf\{t > 0 \mid |M_t| > n\}$ for $n \in \mathbb{N}$. Then M^{T_n} is a bounded martingale. Recall the (generalized) Doob decomposition. If $M \in \mathscr{M}^{c,loc}$ then there is a unique, continuous, increasing process $[M]$ such that $M^2 - [M] \in \mathscr{M}^{c,loc}$. For $M, N \in \mathscr{M}_0^{c,loc}$ define

$$[M,N] = \frac{1}{4}([M+N] - [M-N]).$$

3.4.3 Definition. Let $M \in \mathscr{M}_0^{c,loc}$ and X progressively measurable with

$$\mathbb{P}\left[\int_0^T X_t^2 d[M]_t\right] = 1$$

for all $T < \infty$. Then $I_t^M(X)$ is defined to be

$$I_t^M(X) = I_t^{M^{T_n}}(X^{T_n}) = \int_0^t X^{T_n} dM^{T_n}$$

for all $t \in [0, T_n]$, where $T_n = S_n \wedge R_n$ and

$$S_n = \inf\{t > 0 \mid \int_0^t X_t^2 d[M]_t \geq n\} \quad \text{and} \quad R_n = \inf\{t > 0 \mid |M_t| \geq n\}.$$

Note that $M^{T_n} \in \mathscr{M}_2^c$ and $X^{T_n} \in \mathscr{L}^*(M^{T_n})$ by the definition of S_n. So $I^{M^{T_n}}(X^{T_n})$ is well-defined.

3.4.4 Definition. A (continuous) *semi-martingale* X is a process that admits a decomposition $X = M + A$, where $M \in \mathscr{M}^{c,loc}$ and $A \in \mathscr{F}\mathscr{V}^{c,loc}$, the collection of continuous adapted processes that are of finite variation on every bounded interval.

This decomposition is unique. Now we can define the stochastic integral with respect to a continuous semi-martingale in the obvious way.

3.4.5 Proposition. *We have the following properties of $I^M(H)$.*
(i) Linearity;
(ii) $[I^M(H)]_t = \int_0^t H_s^2 d[M]_s$;
(iii) $[I^M(H), I^N(K)]_t = \int_0^t H_s K_s d[M,N]_s$;
Furthermore, the stochastic integral $I^M(H)$ is characterized as the unique $\Phi \in \mathscr{M}^{c,loc}$ such that $[\Phi, N]_t = \int_0^t H_s d[M,N]_s$ for all $N \in \mathscr{M}_2^c$.

In particular we *cannot* say things regarding the conditional expectations.

3.5 Integration by parts formula for stochastic integrals

Recall that

$$\begin{aligned} M_t^2 &= \sum_\Pi (M_{t_k}^2 - M_{t_{k-1}}^2) \\ &= \sum_\Pi (M_{t_k} - M_{t_{k-1}})(M_{t_k} + M_{t_{k-1}}) \\ &= 2\sum_\Pi M_{t_{k-1}}(M_{t_k} - M_{t_{k-1}}) + \sum_\Pi (M_{t_k} - M_{t_{k-1}})^2 \end{aligned}$$

3.5.1 Lemma. *Let M be a bounded continuous martingale and A be a continuous adapted process of finite variation. Then*

(i) $M_t^2 = 2\int_0^t M_s dM_s + [M]_t$*; and*

(ii) $M_t A_t = \int_0^t A_s dM_s + \int_0^t M_s dA_s$.

PROOF: Define $T_0^n := 0$ and

$$T_{k+1}^n := \inf\{t > T_k^n \mid |M_t - M_{T_k^n}| > \frac{1}{2^n}\},$$

and define $t_k^n = t \wedge T_k^n$. Then we have

$$M_t^2 = 2\sum_{k\geq 1} M_{t_{k-1}^n}(M_{t_k^n} - M_{t_{k-1}^n}) + \sum_{k\geq 1}(M_{t_k^n} - M_{t_{k-1}^n})^2$$

Define

$$X^n := \sum_{k\geq 1} M^{T_{k-1}^n}\mathbf{1}_{(T_{k-1}^n, T_k^n]} \quad \text{and} \quad A^n := \sum_{k\geq 1}(M_{t_k^n} - M_{t_{k-1}^n})^2,$$

so we can rewrite this as $M_t^2 = 2I^M(X^n) + A^n$. We showed earlier that $A^n \to [M]$ a.s. (at least along a subsequence). Since $\sup_t |X_t^n - X_t^{n+1}| \leq 2^{-n-1}$ and $\sup_t |X_t^n - M_t| \leq 2^{-n}$. Taking limits, we have $M_t^2 = 2I_t^M(M) + [M]_t$.

The second part is a similar argument, taking $t_k^n = (k2^{-n}) \wedge t$ and $X^n = \sum_{k\geq 0} A_{t_{k-1}^n}\mathbf{1}_{(t_{k-1}^n, t_k^n]}$. □

3.5.2 Theorem. *Let X and Y be continuous semi-martingales. Then*

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X, Y]_t.$$

PROOF: We assume without loss of generality that $X_0 = Y_0 = 0$. Suppose that $X = M + A$ and $Y = N + V$. Using the first part of the last lemma applied to $M + N$ and $M - N$ to get

$$M_t N_t = \int_0^t M_s dN_s + \int_0^t N_s dM_s + [M, N]_t.$$

Combine this with ordinary Lebesgue-Stieltjes integration to get the result. □

Recall the following results.

3.5.3 Theorem (Chain Rule).
If X is a continuous semi-martingale and U and V are progressively measurable processes with $V \in \mathscr{L}^*(X)$ then $U \in \mathscr{L}^*(V \cdot X)$ if and only if $UV \in \mathscr{L}^*(X)$, in which case $U \cdot (V \cdot X) = (UV) \cdot X$.

3.5.4 Theorem (Integration-by-parts).
For continuous semi-martingales X and Y,

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X,Y]_s.$$

We prove today the following.

3.5.5 Theorem (Itô's Formula).
For a continuous semi-martingale X and a smooth function $f \in C^2(\mathbb{R}^d)$,

$$f(X_t) - f(X_0) = \sum_i \int_0^t \frac{\partial f}{\partial x_i}(X_s) dX_s^i + \frac{1}{2}\sum_{i,j}\int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d[X^i, X^j]_s.$$

Proof: We prove the case $d = 1$. Fix X and let $\mathscr{C}$ be the collection of smooth functions f for which the formula holds. Clearly $\mathscr{C}$ is a linear subspace of $C^2(\mathbb{R})$ what contains all linear functions. We show that $\mathscr{C}$ is closed under multiplication. Let $f, g \in \mathscr{C}$ and define $F = f(X)$ and $G = g(X)$. Then F and G are continuous semi-martingales. By the integration-by-parts formula,

$$\begin{aligned}
(fg)(X_t) - (fg)(X_0) &= F_t G_t - F_0 G_0 \\
&= (F \cdot G)_t + (G \cdot F)_t + [F,G]_t \\
&= f(X)\cdot(g'(X)\cdot X + \tfrac{1}{2}g''(X)\cdot[X]) + g(X)\cdot(f'(X)\cdot X + \tfrac{1}{2}f''(X)\cdot[X]) + [f(X), g(X)] \\
&= f(X)\cdot(g'(X)\cdot X + \tfrac{1}{2}g''(X)\cdot[X]) + g(X)\cdot(f'(X)\cdot X + \tfrac{1}{2}f''(X)\cdot[X]) + f'(X)g'(X)\cdot[\\
&= (fg' + gf')(X)\cdot X + \tfrac{1}{2}(2f'' + 2g'f' + 2g'')(X)\cdot[X] \\
&= (fg)'(X)\cdot X + \tfrac{1}{2}(fg)''(X)\cdot[X]
\end{aligned}$$

since

$$[f(X), g(X)] = [f'(X)\cdot X, g'(X)\cdot X] = f'(X)g'(X)\cdot[X]$$

by Kunita-Watanabe. Therefore $\mathscr{C}$ contains all polynomials. Let $f \in \mathscr{C}^2$ be arbitrary. By the Weierstrass approximation theorem there are polynomials p_n such that

$$\sup_{|x|\le c} |p_n(x) - f''(x)| \to 0$$

as $n \to \infty$ for all $c > 0$. Integrate p_n twice to get polynomials F_n such that

$$\sup_{|x|\le c} |F_n(x) - f(x)| \vee |F_n'(x) - f'(x)| \vee |F_n''(x) - f''(x)| \to 0$$

as $n \to \infty$ for all $c > 0$. In particular, $F_n(X_t) \to f(X_t)$ for all $t \ge 0$. Letting $X = \tilde{M} + \tilde{A}$, by the dominated convergence theorem for Stieltjes integral,

$$(F_n'(X)\cdot\tilde{A} + \tfrac{1}{2}F_n''(X)\cdot[\tilde{M}]) \to (f'(X)\cdot\tilde{A} + \tfrac{1}{2}f''(X)\cdot[\tilde{M}]).$$

All that remains to show is that

$$\int_0^t F_n'(X) d\tilde{M}_s \to \int_0^t f'(X) d\tilde{M}_s.$$

This sequence does converge to this limit in L^2 because

$$\mathbb{E}\left(\int_0^t F_n'(X) - f'(X)d\tilde{M}_s\right)^2 = \mathbb{E}\int_0^t (F_n'(X) - f'(X))^2 d[\tilde{M}_s] \to 0$$

as $n \to \infty$ (this is the Itô Isometry). Therefore there is a subsequence along which the convergence is a.s. □

3.6 Fisk-Stratonovich integral

3.6.1 Definition. The *Fisk-Stratonovich integral* of X with respect to Y is

$$\int_0^t X_s \circ dY_s := \int_0^t X_s dY_s + \frac{1}{2}[X,Y]_t.$$

We have an Itô rule and integration-by-parts rule for this type of integral. We also have the following fact.

$$S_\varepsilon(\Pi) = \sum_{i=0}^{m-1} \frac{1}{2}(B_{t_i} + B_{t_{i+1}})(B_{t_i} - B_{t_{i+1}}) \xrightarrow{\text{(p)}} \int_0^t B_s \circ dB_s.$$

3.6.2 Example. Consider the ODE $\frac{dN}{dt} = a(t)N(t)$, where a is the rate of growth and N is the number of people. For whatever reason, we may think of $a(t)$ as $r(t)+\xi(t)$, a deterministic part plus a process of random fluctuations. Empirically, "$\xi(t) = \frac{dB_t}{dt}$", so we write

$$dN_t = r(t)N_t dt + \sigma N_t dB_t.$$

It is not clear which integral we should use to write N_t. The choice of integral depends on the model. In finance the Itô integral is used because we cannot look into the future. For stochastic processes on a manifold the Stratonovich integral is used.

3.7 Applications of Itô's formula

Regular conditional probabilities

Given a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$, its characteristic function is

$$f(\theta) = \int_\Omega e^{i\theta x}\, \mathbb{P}(dx) = \mathbb{E}[e^{i\theta Z}]$$

for any r.v. Z with law $\mathbb{P}$. Now suppose you were able to show that for $\mathbb{P}$-a.e. $\omega \in \Omega$,

$$f(\theta) = \mathbb{E}[e^{i\theta Z} \mid \mathscr{F}_T](\omega).$$

We would expect that $f(\theta)$ should be the characteristic function of the conditional law of Z given $\mathscr{F}_T$. In order to do that we would like to be able to write

$$\mathbb{E}[e^{i\theta Z} \mid \mathscr{F}_T](\omega) = \int_\Omega e^{i\theta x}(\mathbb{P} \mid \mathscr{F}_T)(\omega \mid dx)$$

for some measure $(\mathbb{P} \mid \mathscr{F}_T)$. This motivates the following.

3.7.1 Definition. Given a r.v. Z on $(\Omega, \mathscr{F}, \mathbb{P})$ taking values in $(S, \mathscr{S})$, and a sub-σ-algebra $\mathscr{F}_T \subseteq \mathscr{F}$, we will say $(\mathbb{P} \mid \mathscr{F}_T)$ is a *regular conditional probability* if

(i) for all $\omega \in \Omega$, $(\mathbb{P} \mid \mathscr{F}_T)(\omega, \cdot)$ defines a probability measure on $(S, \mathscr{S})$;
(ii) for all $A \in \mathscr{S}$, $(\mathbb{P} \mid \mathscr{F}_T)(\cdot, A)$ is $\mathscr{F}_T$-measurable; and
(iii) for all $A \in \mathscr{S}$ and $\omega \in \Omega$, $\mathbb{P}(Z \in A \mid \mathscr{F}_T)(\omega) = (\mathbb{P} \mid \mathscr{F}_T)(\omega, A)$.

For fixed A, we notice $\mathbb{P}(Z \in A \mid \mathscr{F}_T)(\omega) = \mathbb{E}[\mathbf{1}_A \mid \mathscr{F}_T](\omega)$ $\mathbb{P}$-a.s. The existence of a regular conditional probability is asking whether there a "modification" of $\{\mathbb{P}(A \mid \mathscr{F}_T) \mid A \in \mathscr{S}\}$ (defined as above) that satisfies the first condition (as it certainly satisfies the second).

3.7.2 Theorem. *If $(S, \mathscr{S})$ is a complete, separable, metric space and $\mathscr{S} = \mathscr{B}(\mathscr{S})$ then regular conditional probabilities exist.*

3.7.3 Lemma. *Let X be a d-dimensional random vector on $(\Omega, \mathscr{F}, \mathbb{P})$. Suppose that $\mathscr{G}$ is a sub-σ-field of $\mathscr{F}$ and suppose that for each $\omega \in \Omega$ there is a function $\varphi(\omega, \cdot) : \mathbb{R}^d \to \mathbb{C}$ such that for all $u \in \mathbb{R}^d$,*

$$\varphi(\omega, u) = \mathbb{E}[e^{i\langle u, X\rangle} \mid \mathscr{G}](\omega)$$

$\mathbb{P}$-a.e. If for each ω, $\varphi(\omega, u)$ is the characteristic function of some probability measure $\mathbb{P}^\omega$ on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, i.e. if

$$\varphi(\omega, u) = \int_{\mathbb{R}^d} e^{i\langle u, x\rangle}\, \mathbb{P}^\omega(dx)$$

then for all $A \in \mathscr{B}(\mathbb{R}^d)$,

$$\mathbb{P}[X \in A \mid \mathscr{G}](\omega) = \mathbb{P}^\omega(A) =: (\mathbb{P} \mid \mathscr{G})(\omega, A)$$

$\mathbb{P}$-a.e. ω.

PROOF: Let $(Q \mid \mathscr{G})$ be a regular conditional probability for X given $\mathscr{G}$, so that for each fixed $u \in \mathbb{R}^d$,

$$\varphi(\omega, u) = \mathbb{E}[e^{i\langle u, X\rangle} \mid \mathscr{G}](\omega) = \int_{\mathbb{R}^d} e^{i\langle u, x\rangle}(Q \mid \mathscr{G})(\omega, dx)$$

$\mathbb{P}$-a.e. ω. The set of ω for which this holds could depend on u. Take a countable dense subset D and $\tilde{\Omega} \in \mathscr{F}$ with $\mathbb{P}(\tilde{\Omega}) = 1$ so that the equation above holds for all $u \in D$ and all $\omega \in \tilde{\Omega}$. Use continuity with respect to u of both sides to conclude for all $u \in \mathbb{R}^d$ and all $\omega \in \tilde{\Omega}$. (See page 85 in KS.) □

This can be used to prove the strong Markov property in a different way.

Martingale characterization of Brownian motion

Recall that if B is a standard d-dimensional Brownian motion then the covariation among the components is $[B^{(j)}, B^{(j)}]_t = \delta_{ij} t$.

3.7.4 Theorem (Lévy, 1948, Kunita-Watanabe, 1967).
Suppose that X is a d-dimensional continuous adapted process such that for every component $1 \le k \le d$ the process $M_t^{(k)} := X_t^{(k)} - X_0^{(k)}$ is a continuous local martingale and $[M^{(j)}, M^{(j)}]_t = \delta_{ij}$. Then X is a Brownian motion.

PROOF: We will show that for all $0 \le s \le t$, $X_t - X_s$ is independent of $\mathscr{F}_s$ and has the d-variate normal distribution with mean zero and covariance matrix $(t-s)I_d$. To do this we will show that

$$\mathbb{E}[e^{i\langle u, X_t - X_s\rangle} \mid \mathscr{F}_s] = e^{-\frac{1}{2}\|u\|^2(t-s)}.$$

For fixed u, $f(x) = e^{i\langle u,x\rangle}$ satisfies $\frac{\partial f}{\partial x_j}(x) = iu_j f(x)$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = -u_i u_j f(x)$. Applying Itô's formula to the real and imaginary parts we have

$$e^{i\langle u,X_t\rangle} = e^{i\langle u,X_s\rangle} + i\sum_{j=1}^d u_j \int_s^t e^{i\langle u,X_v\rangle} dM_v^{(j)} - \frac{1}{2}\sum_{j=1}^d u_j^2 \int_s^t e^{i\langle u,X_v\rangle} dv.$$

Now $|f(x)| \le 1$ for all $x \in \mathbb{R}^d$, and because $[M^{(j)}]_t = t$ is bounded on any interval, $M^{(j)} \in \mathscr{M}_2^c$. Thus the real and imaginary parts of $\{\int_0^t e^{i\langle u,X_v\rangle} dM_v^{(j)}\}$ lie in $\mathscr{M}_2^c$ (not just $\mathscr{M}^{c,loc}$). Taking expectations,

$$\mathbb{E}[\int_s^t e^{i\langle u,X_v\rangle} dM_v^{(j)} \mathscr{F}_s] = 0$$

$\mathbb{P}$-a.s. For $A \in \mathscr{F}_s$, multiplying by $e^{-i\langle u,X_s\rangle}\mathbf{1}_A$ gives the result. (Fill this in.) □

Itô's formula for general $f \in C^{1,2}$ and $X = M + A$ a semi-martingale is

$$f(t,X_t) = f(0,X_0) + \int_0^{\frac{\partial f}{\partial t}} (x,X_s)ds + \frac{1}{2}\sum_{i,j}\int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(s,X_s)d[X_i,X_j]_s + \sum_i \int_0^t \frac{\partial f}{\partial x_i}(x,X_s)dX_s^i$$

Bessel process

Let B be a d-dimensional Brownian motion and let

$$R_t = \|B_t\| = \sqrt{(B_t^{(1)})^2 + \cdots + (B_t^{(d)})^2}.$$

By the rotation (orthogonal transformation property) of Brownian motion, if $\|y\| = \|x\|$ then R has the same distribution under $\mathbb{P}^x$ and $\mathbb{P}^y$. Use this to show that R is a Markov process (see hand-written notes).

Index