DR.RUPNATHJI(DR.RUPAK NATH)

STOCHASTIC ANALYSIS

NOTES

DR.RUPNATHJI(DR.RUPAK NATH)

Chapter 1

Introduction

We know from elementary probability theory that probabilities of disjoint events "add up", that is, if A and B are events with $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$. For infinitely-many events the situation is a little more subtle. For example, suppose that X is a random variable with uniform distribution on the interval $(0, 1)$. Then $\{ X \in (0,1) \} = \bigcup_{s \in (0,1)} \{ X = s \}$ and $P(X \in (0,1)) = 1$ but $P(X = s) = 0$ for every $s \in (0,1)$. So the probabilities on the right hand side do not "add up" to that on the left hand side.

A satisfactory theory must be able to cope with this.

Continuing with this uniform distribution example, given an arbitrary subset $S \subset (0,1)$, we might wish to know the value of $P(X \in S)$. This seems a reasonable request but can we be sure that there is an answer, even in principle? We will consider the following similar question.

Does it make sense to talk about the length of every subset of $\mathbb{R}$? More precisely, does there exist a "length" or "measure" m defined on *all* subsets of $\mathbb{R}$ such that the following hold?

1. $m(A) \geq 0$ for all $A \subset \mathbb{R}$ and $m(\varnothing) = 0$.

2. If $A_1, A_2, \dots$ is any sequence of pairwise disjoint subsets of $\mathbb{R}$, then $m\big(\bigcup_n A_n\big) = \sum_n m(A_n)$.

3. If I is an interval $[a, b]$, then $m(I) = \ell(I) = b - a$, the "length" of I.

4. $m(A + a) = m(A)$ for any $A \subset \mathbb{R}$ and $a \in \mathbb{R}$ (translation invariance).

The answer to this question is "no", it is not possible. This is a famous "no-go" result.

Proof. To see this, we first construct a special collection of subsets of $[0, 1)$ defined via the following equivalence relation. For $x, y \in [0, 1)$, we say that $x \sim y$ if and only if $x - y \in \mathbb{Q}$. Evidently this is an equivalence relation on $[0, 1)$ and so $[0, 1)$ is partitioned into a family of equivalence classes.

DR.RUPNATHJI(DR.RUPAK NATH)

One such is $\mathbb{Q} \cap [0, 1)$. No other equivalence class can contain a rational. Indeed, if a class contains one rational, then it must contain all the rationals which are in $[0, 1)$. Choose one point from each equivalence class to form the set A_0, say. Then no two elements of A_0 are equivalent and every equivalence class has a representative which belongs to A_0.

For each $q \in \mathbb{Q} \cap [0, 1)$, we construct a subset A_q of $[0, 1)$ as follows. Let $A_q = \{ x + q - [x + q] : x \in A_0 \}$ where $[t]$ denotes the integer part of the real number t. In other words, A_q is got from $A_0 + q$ by translating that portion of $A_0 + q$ which lies in the interval $[1, 2)$ back to $[0, 1)$ simply by subtracting 1 from each such element.

If we write $A_0 = B'_q \cup B''_q$ where $B'_q = \{ x \in A_0 : x + q < 1 \}$ and $B''_q = A_0 \setminus B'_q$, then we see that $A_q = (B'_q + q) \cup (B''_q + q - 1)$. The translation invariance of m implies that $m(A_q) = m(A_0)$.

Claim. If $r, s \in \mathbb{Q} \cap [0, 1)$ with $r \neq s$, then $A_r \cap A_s = \varnothing$.

Proof. Suppose that $x \in A_r \cap A_s$. Then there is $\alpha \in A_0$ such that $x = \alpha + r - [\alpha + r]$ and $\beta \in A_0$ such that $x = \beta + s - [\beta + s]$. It follows that $\alpha \sim \beta$ which is not possible unless $\alpha = \beta$ which would mean that $r = s$. This proves the claim.

Claim. $\bigcup_{q \in \mathbb{Q} \cap [0,1)} A_q = [0, 1)$.

Proof. Since $A_q \subset [0, 1)$, we need only show that the right hand side is a subset of the left hand side. To establish this, let $x \in [0, 1)$. Then there is $a \in A_0$ such that $x \sim a$, that is $x - a \in \mathbb{Q}$.

Case 1. $x \geq a$. Put $q = x - a \in \mathbb{Q} \cap [0, 1)$. Then $x = a + q$ so that $x \in A_q$.

Case 2. $x < a$. Since both $x \in [0, 1)$ and $a \in [0, 1)$, it follows that $2 > 1 + x > a$. Put $r = 1 + x - a$. Then $0 < r < 1$ (because $x < a$) and $r \in \mathbb{Q}$. Hence $x = a + r - 1$ so that $x \in A_r$. The claim is proved.

Finally, we get our contradiction. We have $m([0, 1)) = \ell([0, 1)) = 1$. But

$$m([0, 1)) = m\Big(\bigcup_{q \in \mathbb{Q} \cap [0,1)} A_q \Big) = \sum_{q \in \mathbb{Q} \cap [0,1)} m(A_q)$$

which is not possible since $m(A_q) = m(A_0)$ for all $q \in \mathbb{Q} \cap [0, 1)$. ∎

We have seen that the notion of length simply cannot be assigned to every subset of $\mathbb{R}$. We might therefore expect that it is not possible to assign a probability to every subset of the sample space, in general.

The following result is a further very dramatic indication that we may not be able to do everything that we might like.

Theorem 1.1 (Banach-Tarski). *It is possible to cut-up a sphere of radius one (in $\mathbb{R}^3$) into a finite number of pieces and then reassemble these pieces (via standard Euclidean motions in $\mathbb{R}^3$) into a sphere of radius 2.*

Moral - all the fuss with σ-algebras and so on really is necessary if we want to develop a robust (and rigorous) theory.

DR.RUPNATHJI(DR.RUPAK NATH)

Some useful results

Frequent use is made of the following.

Proposition 1.2. *Let* (A_n) *be a sequence of events such that* $P(A_n) = 1$ *for all* n*. Then* $P(\bigcap_n A_n) = 1$.

Proof. We note that $P(A_n^c) = 0$ and so

$$P(\bigcup_n A_n^c) = \lim_{m\to\infty} P(\bigcup_{n=1}^m A_n^c) = 0$$

because $P(\bigcup_{n=1}^m A_n^c) \le \sum_{n=1}^m P(A_n^c) = 0$ for every m. But then

$$P(\bigcap_n A_n) = 1 - P((\bigcap_n A_n)^c) = 1 - P(\bigcup_n A_n^c) = 1$$

as required. ∎

Suppose that (A_n) is a sequence of events in a probability space (Ω, Σ, P). We define the event $\{ A_n$ infinitely-often $\}$ to be the event

$$\{ A_n \text{ infinitely-often} \} = \{ \omega \in \Omega : \forall N \;\; \exists n > N \text{ such that } \omega \in A_n \}.$$

Evidently, $\{ A_n$ infinitely-often $\} = \bigcap_n \bigcup_{k\ge n} A_k$ and so $\{ A_n$ infinitely-often $\}$ really is an event, that is, it belongs to Σ.

Lemma 1.3 (Borel-Cantelli).

(i) (First Lemma) *Suppose that* (A_n) *is a sequence of events such that* $\sum_n P(A_n)$ *is convergent. Then* $P(A_n$ *infinitely-often*$) = 0$.

(ii) (Second Lemma) *Let* (A_n) *be a sequence of independent events such that* $\sum_n P(A_n)$ *is divergent. Then* $P(A_n$ *infinitely-often*$) = 1$.

Proof. (i) By hypothesis, it follows that for any $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $\sum_{n=N}^\infty P(A_n) < \varepsilon$. Now

$$P(A_n \text{ infinitely-often}) = P(\bigcap_n \bigcup_{k\ge n} A_k) \le P(\bigcup_{k\ge N} A_k).$$

But for any m

$$P(\bigcup_{k=N}^m A_k) \le \sum_{k=N}^m P(A_k) < \varepsilon$$

and so, taking the limit $m \to \infty$, it follows that

$$P(A_n \text{ infinitely-often}) \le P(\bigcup_{k\ge N} A_k) = \lim_m P(\bigcup_{k=N}^m A_k) \le \varepsilon$$

and therefore $P(A_n$ infinitely-often$) = 0$, as required.

DR.RUPNATHJI(DR.RUPAK NATH)

(ii) We have

$$\begin{aligned} P(A_n \text{ infinitely-often}) &= P\big(\textstyle\bigcap_n \bigcup_{k \ge n} A_k \big) \\ &= \lim_n \, P\big(\textstyle\bigcup_{k \ge n} A_k \big) \\ &= \lim_n \, \lim_m \, P\big(\textstyle\bigcup_{k=n}^{m+n} A_k \big) \, . \end{aligned}$$

The proof now proceeds by first taking complements, then invoking the independence hypothesis and finally by the inspirational use of the inequality $1 - x \le e^{-x}$ for any $x \ge 0$. Indeed, we have

$$\begin{aligned} P\big((\textstyle\bigcup_{k=n}^{m+n} A_k)^c \big) &= P\big(\textstyle\bigcap_{k=n}^{m+n} A_k^c \big) \\ &= \prod_{k=n}^{m+n} P(A_k^c) \, , \quad \text{by independence,} \\ &\le \prod_{k=n}^{m+n} e^{-P(A_k)} \, , \quad \text{since } P(A_k^c) = 1 - P(A_k) \le e^{-P(A_k)}, \\ &= e^{-\sum_{k=n}^{m+n} P(A_k)} \\ &\to 0 \end{aligned}$$

as $m \to \infty$, since $\sum_k P(A_k)$ is divergent. It follows that $P\big(\bigcup_{k \ge n} A_k \big) = 1$ and so $P(A_n \text{ infinitely-often}) = 1$. ■

Example 1.4. A fair coin is tossed repeatedly. Let A_n be the event that the outcome at the n^{th} play is "heads". Then $P(A_n) = \frac{1}{2}$ and evidently $\sum_n P(A_n)$ is divergent (and $A_1, A_2, \dots$ are independent). It follows that $P(A_n \text{ infinitely-often}) = 1$. In other words, in a sequence of coin tosses, there will be an infinite number of "heads" with probability one.

Now let B_n be the event that the outcomes of the five consecutive plays at the times $5n$, $5n+1$, $5n+2$, $5n+3$ and $5n+4$ are all "heads". Then $P(B_n) = (\frac{1}{2})^5$ and so $\sum_n P(B_n)$ is divergent. Moreover, the events $B_1, B_2, \dots$ are independent and so $P(B_n \text{ infinitely-often}) = 1$. In particular, it follows that, with probability one, there is an infinite number of "5 heads in a row".

DR.RUPNATH(DR.RUPAKNATH)

Functions of a random variable

If X is a random variable and $g : \mathbb{R} \to \mathbb{R}$ is a Borel function, then $Y = g(X)$ is $\sigma(X)$-measurable (where $\sigma(X)$ denotes the σ-algebra generated by X). The converse is true. If X is discrete, then one can proceed fairly directly. Suppose, by way of illustration, that X assumes the finite number of distinct values $x_1, \dots, x_m$ and that $\Omega = \bigcup_{k=1}^m A_k$ where $X = x_k$ on A_k. Then $\sigma(X)$ is generated by the finite collection $\{ A_1, \dots, A_m \}$ and so any $\sigma(X)$-measurable random variable Y must have the form $Y = \sum_k y_k \, 1_{A_k}$ for some $y_1, \dots, y_m$ in $\mathbb{R}$. Define $g : \mathbb{R} \to \mathbb{R}$ by setting $g(x) = \sum_{k=1}^m y_k \, 1_{\{ x_k \}}(x)$. Then g is a

Borel function and $Y = g(X)$. To prove this for the general case, one takes a far less direct approach.

Theorem 1.5. *Let X be a random variable on $(\Omega, \mathcal{S}, P)$ and suppose that Y is $\sigma(X)$-measurable. Then there is a Borel function $g : \mathbb{R} \to \mathbb{R}$ such that $Y = g(X)$.*

Proof. Let $\mathcal{C}$ denote the class of Borel functions of X,

$$\mathcal{C} = \{ \varphi(X) : \text{ for some Borel function } \varphi : \mathbb{R} \to \mathbb{R} \}$$

Then $\mathcal{C}$ has the following properties.

(i) $1_A \in \mathcal{C}$ for any $A \in \sigma(X)$.

To see this, we note that $\sigma(X) = X^{-1}(\mathcal{B})$ and so there is $B \in \mathcal{B}$ such that $A = X^{-1}(B)$. Hence $1_A(\omega) = 1_B(X(\omega))$ and so $1_A \in \mathcal{C}$ because $1_B : \mathbb{R} \to \mathbb{R}$ is Borel.

(ii) Clearly, any linear combination of members of $\mathcal{C}$ also belongs to $\mathcal{C}$. This fact, together with (i) means that $\mathcal{C}$ contains all simple functions on $(\Omega, \sigma(X))$.

(iii) $\mathcal{C}$ is closed under pointwise convergence.

Indeed, suppose that (φ_n) is a sequence of Borel functions on $\mathbb{R}$ such that $\varphi_n(X(\omega)) \to Z(\omega)$ for each $\omega \in \Omega$. We wish to show that $Z \in \mathcal{C}$, that is, that $Z = \varphi(X)$ for some Borel function $\varphi : \mathbb{R} \to \mathbb{R}$.

Let $B = \{ s \in \mathbb{R} : \lim_n \varphi_n(s) \text{ exists} \}$. Then

$$\begin{aligned} B &= \{ s \in \mathbb{R} : (\varphi_n(s)) \text{ is a Cauchy sequence in } \mathbb{R} \} \\ &= \bigcap_{k \in \mathbb{N}} \bigcup_{N \in \mathbb{N}} \bigcap_{m,n > N} \underbrace{\{ \varphi_m(s) - \tfrac{1}{k} < \varphi_n(s) < \varphi_m(s) + \tfrac{1}{k} \}}_{\{ \varphi_n(s) < \varphi_m(s) + \frac{1}{k} \} \cap \{ \varphi_m(s) - \frac{1}{k} < \varphi_n(s) \}} \end{aligned}$$

and it follows that $B \in \mathcal{B}$. Furthermore, by hypothesis, $\varphi_n(X(\omega))$ converges (to $Z(\omega)$) and so $X(\omega) \in B$ for all $\omega \in \Omega$.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be given by

$$\varphi(s) = \begin{cases} \lim_n \varphi_n(s), & s \in B \\ 0, & s \notin B\,. \end{cases}$$

We claim that φ is a Borel function on $\mathbb{R}$. To see this, let $\psi_n(s) = \varphi_n(s)\, 1_B(s)$. Each ψ_n is Borel and $\psi_n(s)$ converges to $\varphi(s)$ for each $s \in \mathbb{R}$. It follows that φ is also a Borel function on $\mathbb{R}$. Indeed,

$$\{ s : \varphi(s) < \alpha \} = \bigcup_{k \in \mathbb{N}} \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} \{ s : \psi_n(s) < \alpha - \tfrac{1}{k} \}$$

is a Borel subset of $\mathbb{R}$ for any $\alpha \in \mathbb{R}$.

To complete the proof of (iii), we note that for any $\omega \in \Omega$, $X(\omega) \in B$ and so $\varphi_n(X(\omega)) \to \varphi(X(\omega))$. But $\varphi_n(X(\omega)) \to Z(\omega)$ and so we conclude that $Z = \varphi(X)$ and is of the required form.

DR RUPNATHJI(DR RUPAK NATH)

Finally, to complete the proof of the theorem, we note that any Borel measurable function $Y : \Omega \to \mathbb{R}$ is a pointwise limit of simple functions and so must belong to $\mathcal{C}$, by (ii) and (iii) above. In other words, any such Y has the form $Y = g(X)$ for some Borel function $g : \mathbb{R} \to \mathbb{R}$. ■

$\mathcal{L}$ versus L

Let $(\Omega, \mathcal{S}, P)$ be a probability space. For any $1 \le p < \infty$, the space $\mathcal{L}^p(\Omega, \mathcal{S}, P)$ is the collection of measurable functions (random variables) $f : \Omega \to \mathbb{R}$ (or $\mathbb{C}$) such that $E(|f|^p) < \infty$, i.e., $\int_\Omega |f(\omega)|^p \, dP$ is finite. One can show that $\mathcal{L}^p$ is a linear space. Let $\|f\|_p = \left(\int_\Omega |f(\omega)|^p \, dP\right)^{1/p}$. Then (using Minkowski's inequality), one can show that

$$\|f + g\|_p \le \|f\|_p + \|g\|_p$$

for any $f, g \in \mathcal{L}^p$. As a consequence, $\|\cdot\|_p$ is a semi-norm on $\mathcal{L}^p$ — but not a norm. Indeed, if $g = 0$ almost surely, then $\|g\|_p = 0$ even though g need not be the zero function. It is interesting to note that if q obeys $1/p + 1/q = 1$ (q is called the exponent conjugate to p), then

$$\|f\|_p = \sup\{ \int_\Omega |fg| \, dP : \|g\|_q \le 1 \}.$$

Of further interest is Hölder's inequality

$$\int_\Omega |fg| \, dP \le \|f\|_p \, \|g\|_q$$

for any $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^q$ and with $1/p + 1/q = 1$. This reduces (essentially) to the Cauchy-Schwarz inequality when $p = 2$ (so that $q = 2$ also).

To "correct" the norm-seminorm issue, consider equivalence classes of functions, as follows. Declare functions f and g in $\mathcal{L}^p$ to be equivalent, written $f \sim g$, if $f = g$ almost surely. In other words, if $\mathcal{N}$ denotes the collection of random variables equal to 0 almost surely, then $f \sim g$ if and only if $f - g \in \mathcal{N}$. (Note that every element h of $\mathcal{N}$ belongs to every $\mathcal{L}^p$ and obeys $\|h\|_p = 0$.) One sees that $\sim$ is an equivalence relation. (Clearly $f \sim f$, and if $f \sim g$ then $g \sim f$. To see that $f \sim g$ and $g \sim h$ implies that $f \sim h$, note that $\{ f = g \} \cap \{ g = h \} \subset \{ f = h \}$. But $P(f = g) = P(g = h) = 1$ so that $P(f = h) = 1$, i.e., $f \sim h$.)

For any $f \in \mathcal{L}^p$, let $[f]$ be the equivalence class containing f, so that $[f] = f + \mathcal{N}$. Let $L^p(\Omega, \mathcal{S}, P)$ denote the collection of equivalence classes $\{ [f] : f \in \mathcal{L}^p \}$. $L^p(\Omega, \mathcal{S}, P)$ is a linear space equipped with the rule

$$\alpha[f] + \beta[g] = [\alpha f + \beta g].$$

One notes that if $f_1 \sim f$ and $g_1 \sim g$, then there are elements h', h'' of $\mathcal{N}$ such that $f_1 = f + h'$ and $g_1 = g + h''$. Then $\alpha f_1 + \beta g_1 = \alpha f + \beta g + \alpha h' + \beta h''$

DR.RUPNATHJI(DR.RUPAK NATH)

which shows that $\alpha f_1 + \beta g_1 \sim \alpha f + \beta g$. In other words, the definition of $\alpha[f] + \beta[g]$ above does not depend on the particular choice of $f \in [f]$ or $g \in [g]$ used, so this really does determine a linear structure on L^p.

Next, we define $|||[f]|||_p = \|f\|_p$. If $f \sim f_1$, then $\|f\|_p = \|f_1\|_p$, so $||| \cdot |||_p$ is well-defined on L^p. In fact, $||| \cdot |||_p$ is a norm on L^p. (If $|||[f]|||_p = 0$, then $\|f\|_p = 0$ so that $f \in \mathcal{N}$, i.e., $f \sim 0$ so that $[f]$ is the zero element of L^p.) It is usual to write $||| \cdot |||_p$ as just $\| \cdot \|_p$. One can think of L^p and $\mathcal{L}^p$ as "more or less the same thing" except that in L^p one simply identifies functions which are almost surely equal.

Note: this whole discussion applies to any measure space — not just probability spaces. The fact that the measure has total mass one is irrelevant here.

Riesz-Fischer Theorem

Theorem 1.6 (Riesz-Fischer). *Let $1 \le p < \infty$ and suppose that (f_n) is a Cauchy sequence in $\mathcal{L}^p$, then there is some $f \in \mathcal{L}^p$ such that $f_n \to f$ in $\mathcal{L}^p$ and there is some subsequence (f_{n_k}) such that $f_{n_k} \to f$ almost surely as $k \to \infty$.*

Proof. Put $d(N) = \sup_{n,m \ge N} \|f_n - f_m\|_p$. Since (f_n) is a Cauchy sequence in $\mathcal{L}^p$, it follows that $d(N) \to 0$ as $N \to \infty$. Hence there is some sequence $n_1 < n_2 < n_3 < \dots$ such that $d(n_k) < 1/2^k$. In particular, it is true that $\|f_{n_{k+1}} - f_{n_k}\|_p < 1/2^k$. Let us write g_k for f_{n_k}, simply for notational convenience.

Let $A_k = \{\, \omega : |g_{k+1}(\omega) - g_k(\omega)| \ge 1/k^2 \,\}$. By Chebyshev's inequality,

$$P(A_k) \le k^{2p}\, E(|g_{k+1} - g_k|^p) = k^{2p}\, \|g_{k+1} - g_k\|_p^p \le k^{2p}/2^k\,.$$

This estimate implies that $\sum_k P(A_k) < \infty$ and so, by the Borel-Cantelli Lemma,

$$P(A_k \text{ infinitely-often}) = 0.$$

In other words, if $B = \{\, \omega : \omega \in A_k \text{ for at most finitely many } k \,\}$, then clearly $B = \{\, A_k \text{ infinitely often} \,\}^c$ so that $P(B) = 1$. But for $\omega \in B$, there is some N (which may depend on ω) such that $|g_{k+1}(\omega) - g_k(\omega)| < 1/k^2$ for all $k > N$. It follows that

$$g_{j+1}(\omega) = g_1(\omega) + \sum_{k=1}^{j} (g_{k+1}(\omega) - g_k(\omega))$$

converges (absolutely) for each $\omega \in B$ as $j \to \infty$. Define the function f on Ω by $f(\omega) = \lim_j g_j(\omega)$ for $\omega \in B$ and $f(\omega) = 0$, otherwise. Then $g_j \to f$ almost surely. (Note that f is measurable because $f = \lim_j g_j \, 1_B$ on Ω.)

DR.RUPNATH JI(DR.RUPAK NATH)

We claim that $f_n \to f$ in $\mathcal{L}^1$. To see this, first observe that

$$\|g_{m+1} - g_j\|_p = \Big\|\sum_{k=j}^{m} (g_{k+1} - g_k)\Big\|_p \le \sum_{k=j}^{m} \|g_{k+1} - g_k\|_p < \sum_{k=j}^{\infty} 2^{-k}.$$

Hence, by Fatou's Lemma, letting $m \to \infty$, we find

$$\|f - g_j\|_p \le \sum_{k=j}^{\infty} 2^{-k}.$$

In particular, $f - g_j \in \mathcal{L}^p$ and so $f = (f - g_j) + g_j \in \mathcal{L}^p$.

Finally, let $\varepsilon > 0$ be given. Let N be such that $d(N) < \frac{1}{2}\varepsilon$ and choose any j such that $n_j > N$ and $\sum_{k=j}^{\infty} 2^{-k} < \frac{1}{2}\varepsilon$. Then for any $n > N$, we have

$$\|f - f_n\|_p \le \|f - f_{n_j}\|_p + \|f_{n_j} - f_n\|_p \le \sum_{k=j}^{\infty} 2^{-k} + d(N) < \varepsilon\,,$$

that is, $f_n \to f$ in $\mathcal{L}^p$. ∎

Proposition 1.7 (Chebyshev's inequality). *For any $c > 0$ and $1 \le p < \infty$*

$$P(|X| \ge c) \le c^{-p}\, \|X\|_p^p$$

for any $X \in \mathcal{L}^p$.

Proof. Let $A = \{\, \omega : |X| \ge c \,\}$. Then

$$\begin{aligned}
\|X\|_p^p &= \int_\Omega |X|^p \; dP \\
&= \int_A |X|^p \; dP + \int_{\Omega \setminus A} |X|^p \; dP \\
&\ge \int_A |X|^p \; dP \\
&\ge c^p P(A)
\end{aligned}$$

as required. ∎

Remark 1.8. For any random variable g which is bounded almost surely, let

$$\|g\|_\infty = \inf\{\, M : |g| \le M \text{ almost surely} \,\}.$$

Then $\|g\|_p \le \|g\|_\infty$ and $\|g\|_\infty = \lim_{p \to \infty} \|g\|_p$. To see this, suppose that g is bounded with $|g| \le M$ almost surely. Then

$$\|g\|_p^p = \int_\Omega |g|^p \; dP \le M^p$$

DR.RUPNATHJI(DR.RUPAK NATH)

and so $\|g\|_p$ is a lower bound for the set $\{ M : |g| \le M$ almost surely $\}$. It follows that $\|g\|_p \le \|g\|_\infty$.

For the last part, note that if $\|g\|_\infty = 0$, then $g = 0$ almost surely. ($g \neq 0$ on the set $\bigcup_n \{ |g| > 1/n \} = A$, say. But if for each $n \in \mathbb{N}$, $|g| \le 1/n$ almost surely, then $P(A) = 0$ because $P(|g| > 1/n) = 0$ for all n.) It follows that $\|g\|_p = 0$ and there is nothing more to prove. So suppose that $\|g\|_\infty > 0$. Then by replacing g by $g/\|g\|_\infty$, we see that we may suppose that $\|g\|_\infty = 1$. Let $0 < r < 1$ be given and choose δ such that $r < \delta < 1$. By definition of the $\|\cdot\|_\infty$-norm, there is a set B with $P(B) > 0$ such that $|g| > \delta$ on B. But then

$$1 = \|g\|_\infty \ge \|g\|_p^p = \int_\Omega |g|^p \, dP \ge \delta^p \, P(B)$$

so that $1 \ge \|g\|_p \ge \delta \, P(B)^{1/p}$. But $P(B)^{1/p} \to 1$ as $p \to \infty$, and so $1 \ge \|g\|_p \ge r$ for all sufficiently large p. The result follows.

DR.RUPNATHJI(DR.RUPAK NATH)

Monotone Class Theorem

It is notoriously difficult, if not impossible, to extend properties of collections of sets directly to the σ-algebra they generate, that is, "from the inside". One usually has to resort to a somewhat indirect approach. The so-called Monotone Class Theorem plays the rôle of the cavalry in this respect and can usually be depended on to come to the rescue.

Definition 1.9. A collection $\mathcal{A}$ of subsets of a set X is an algebra if

(i) $X \in \mathcal{A}$,
(ii) if $A \in \mathcal{A}$ and $B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$,
(iii) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.

Note that it follows that if $A, B \in \mathcal{A}$, then $A \cap B = (A^c \cup B^c)^c \in \mathcal{A}$. Also, for any finite family $A_1, \dots, A_n \in \mathcal{A}$, it follows by induction that $\bigcup_{i=1}^n A_i \in \mathcal{A}$ and $\bigcap_{i=1}^n A_i \in \mathcal{A}$.

Definition 1.10. A collection $\mathcal{M}$ of subsets of X is a monotone class if

(i) whenever $A_1 \subseteq A_2 \subseteq \dots$ is an increasing sequence in $\mathcal{M}$, then $\bigcup_{i=1}^\infty A_i \in \mathcal{M}$,

(ii) whenever $B_1 \supseteq B_2 \supseteq \dots$ is a decreasing sequence in $\mathcal{M}$, then $\bigcap_{i=1}^\infty B_i \in \mathcal{M}$.

One can show that the intersection of an arbitrary family of monotone classes of subsets of X is itself a monotone class. Thus, given a collection $\mathcal{C}$ of subsets of X, we may consider $\mathcal{M}(\mathcal{C})$, the monotone class generated by the collection $\mathcal{C}$ — it is the "smallest" monotone class containing $\mathcal{C}$, i.e., it is the intersection of all those monotone classes which contain $\mathcal{C}$.

Theorem 1.11 (Monotone Class Theorem). *Let $\mathcal{A}$ be an algebra of subsets of X. Then $\mathcal{M}(\mathcal{A}) = \sigma(\mathcal{A})$, the σ-algebra generated by $\mathcal{A}$.*

Proof. It is clear that any σ-algebra is a monotone class and so $\sigma(\mathcal{A})$ is a monotone class containing $\mathcal{A}$. Hence $\mathcal{M}(\mathcal{A}) \subseteq \sigma(\mathcal{A})$. The proof is complete if we can show that $\mathcal{M}(\mathcal{A})$ is a σ-algebra, for then we would deduce that $\sigma(\mathcal{A}) \subseteq \mathcal{M}(\mathcal{A})$.

If a monotone class $\mathcal{M}$ is an algebra, then it is a σ-algebra. To see this, let $A_1, A_2, \dots \in \mathcal{M}$. For each $n \in \mathbb{N}$, set $B_n = A_1 \cup \dots \cup A_n$. Then $B_n \in \mathcal{M}$, if $\mathcal{M}$ is an algebra. But then $\bigcup_{i=1}^\infty A_i = \bigcup_{n=1}^\infty B_n \in \mathcal{M}$ if $\mathcal{M}$ is a monotone class. Thus the algebra $\mathcal{M}$ is also a σ-algebra. It remains to prove that $\mathcal{M}$ is, in fact, an algebra. We shall verify the three requirements.
(i) We have $X \in \mathcal{A} \subseteq \mathcal{M}(\mathcal{A})$.

(iii) Let $A \in \mathcal{M}(\mathcal{A})$. We wish to show that $A^c \in \mathcal{M}(\mathcal{A})$. To show this, let

$$\widetilde{\mathcal{M}} = \{\, B : B \in \mathcal{M}(\mathcal{A}) \text{ and } B^c \in \mathcal{M}(\mathcal{A}) \,\}.$$

Since $\mathcal{A}$ is an algebra, if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$ and so

$$\mathcal{A} \subseteq \widetilde{\mathcal{M}} \subseteq \mathcal{M}(\mathcal{A}).$$

We shall show that $\widetilde{\mathcal{M}}$ is a monotone class. Let (B_n) be a sequence in $\widetilde{\mathcal{M}}$ with $B_1 \subseteq B_2 \subseteq \dots$. Then $B_n \in \mathcal{M}(\mathcal{A})$ and $B_n^c \in \mathcal{M}(\mathcal{A})$. Hence $\bigcup_n B_n \in \mathcal{M}(\mathcal{A})$ and also $\bigcap_n B_n^c \in \mathcal{M}(\mathcal{A})$, since $\mathcal{M}(\mathcal{A})$ is a monotone class (and (B_n^c) is a decreasing sequence).

But $\bigcap_n B_n^c = \big(\bigcup_n B_n\big)^c$ and so both $\bigcup_n B_n$ and $\big(\bigcup_n B_n\big)^c$ belong to $\mathcal{M}(\mathcal{A})$, i.e., $\bigcup_n B_n \in \widetilde{\mathcal{M}}$.

Similarly, if $B_1 \supseteq B_2 \supseteq \dots$ belong to $\widetilde{\mathcal{M}}$, then $\bigcap_n B_n \in \mathcal{M}(\mathcal{A})$ and $\big(\bigcap_n B_n\big)^c = \bigcup_n B_n^c \in \mathcal{M}(\mathcal{A})$ so that $\bigcap_n B_n \in \widetilde{\mathcal{M}}$. It follows that $\widetilde{\mathcal{M}}$ is a monotone class. Since $\mathcal{A} \subseteq \widetilde{\mathcal{M}} \subseteq \mathcal{M}(\mathcal{A})$ and $\mathcal{M}(\mathcal{A})$ is the monotone class generated by $\mathcal{A}$, we conclude that $\widetilde{\mathcal{M}} = \mathcal{M}(\mathcal{A})$. But then this means that for any $B \in \mathcal{M}(\mathcal{A})$, we also have $B^c \in \mathcal{M}(\mathcal{A})$.

DR.RUPNATHJI(DR.RUPAK NATH)

(ii) We wish to show that if A and B belong to $\mathcal{M}(\mathcal{A})$ then so does $A \cup B$. Now, by (iii), it is enough to show that $A \cap B \in \mathcal{M}(\mathcal{A})$ (using $A \cup B = (A^c \cap B^c)^c$). To this end, let $A \in \mathcal{M}(\mathcal{A})$ and let

$$\mathcal{M}_A = \{\, B : B \in \mathcal{M}(\mathcal{A}) \text{ and } A \cap B \in \mathcal{M}(\mathcal{A}) \,\}.$$

Then for $B_1 \subseteq B_2 \subseteq \dots$ in $\mathcal{M}_A$, we have

$$A \cap \bigcup_{i=1}^\infty B_i = \bigcup_{i=1}^\infty A \cap B_i \in \mathcal{M}(\mathcal{A})$$

since each $A \cap B_i \in \mathcal{M}(\mathcal{A})$ by the definition of $\mathcal{M}_A$.

Similarly, if $B_1 \supseteq B_2 \supseteq \dots$ belong to $\mathcal{M}_A$, then

$$A \cap \bigcap_{i=1}^{\infty} B_i = \bigcap_{i=1}^{\infty} A \cap B_i \in \mathcal{M}(\mathcal{A}).$$

Therefore $\mathcal{M}_A$ is a monotone class.

Suppose $A \in \mathcal{A}$. Then for any $B \in \mathcal{A}$, we have $A \cap B \in \mathcal{A}$, since $\mathcal{A}$ is an algebra. Hence $\mathcal{A} \subseteq \mathcal{M}_A \subseteq \mathcal{M}(\mathcal{A})$ and therefore $\mathcal{M}_A = \mathcal{M}(\mathcal{A})$ for each $A \in \mathcal{A}$.

Now, for any $B \in \mathcal{M}(\mathcal{A})$ and $A \in \mathcal{A}$, we have

$$A \in \mathcal{M}_B \iff A \cap B \in \mathcal{M}(\mathcal{A}) \iff B \in \mathcal{M}_A = \mathcal{M}(\mathcal{A}).$$

Hence, for every $B \in \mathcal{M}(\mathcal{A})$,

$$\mathcal{A} \subseteq \mathcal{M}_B \subseteq \mathcal{M}(\mathcal{A})$$

and so (since $\mathcal{M}_B$ is a monotone class) we have $\mathcal{M}_B = \mathcal{M}(\mathcal{A})$ for every $B \in \mathcal{M}(\mathcal{A})$.

Now let $A, B \in \mathcal{M}(\mathcal{A})$. We have seen that $\mathcal{M}_B = \mathcal{M}(\mathcal{A})$ and therefore $A \in \mathcal{M}(\mathcal{A})$ means that $A \in \mathcal{M}_B$ so that $A \cap B \in \mathcal{M}(\mathcal{A})$ and the proof is complete. ∎

Example 1.12. Suppose that P and Q are two probability measures on $\mathcal{B}(\mathbb{R})$ which agree on sets of the form $(-\infty, a]$ with $a \in \mathbb{R}$. Then $P = Q$ on $\mathcal{B}(\mathbb{R})$.

Proof. Let $\mathcal{S} = \{ A \in \mathcal{B}(\mathbb{R}) : P(A) = Q(A) \}$. Then $\mathcal{S}$ includes sets of the form $(a, b]$, for $a < b$, $(-\infty, a]$ and (a, ∞) and so contains $\mathcal{A}$ the algebra of subsets generated by those of the form $(-\infty, a]$. However, one sees that $\mathcal{S}$ is a monotone class (because P and Q are σ-additive) and so $\mathcal{S}$ contains $\sigma(\mathcal{A})$. The proof is now complete since $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

DR.RUPNATHJI(DR.RUPAK NATH)

Chapter 2

Conditional expectation

Consider a probability space $(\Omega, \mathcal{S}, P)$. The conditional expectation of an integrable random variable X with respect to a sub-σ-algebra $\mathcal{G}$ of $\mathcal{S}$ is a $\mathcal{G}$-measurable random variable, denoted by $E(X \mid \mathcal{G})$, obeying

$$\int_A E(X \mid \mathcal{G})\, dP = \int_A X\, dP$$

for every set $A \in \mathcal{G}$.
Where does this come from?
Suppose that $X \geq 0$ and define $\nu(A) = \int_A X\, dP$ for any $A \in \mathcal{G}$. Then

$$\nu(A) = \int_\Omega X\, 1_A\, dP = E(X\, 1_A)\,.$$

Proposition 2.1. *ν is a finite measure on* $(\Omega, \mathcal{G})$.

Proof. Since $X\, 1_A \geq 0$ almost surely for any $A \in \mathcal{G}$, we see that $\nu(A) \geq 0$ for all $A \in \mathcal{G}$. Also, $\nu(\Omega) = E(X)$ which is finite by hypothesis (X is integrable).

Now suppose that $A_1, A_2, \ldots$ is a sequence of pairwise disjoint events in $\mathcal{G}$. Set $B_n = A_1 \cup \cdots \cup A_n$. Then

$$\begin{aligned}\nu(B_n) &= \int_\Omega X\, 1_{B_n}\, dP = \int_\Omega X(1_{A_1} + \ldots + 1_{A_n})\, dP \\ &= \nu(A_1) + \cdots + \nu(A_n)\,.\end{aligned}$$

Letting $n \to \infty$, $1_{B_n} \uparrow 1_{\bigcup_n A_n}$ on Ω and so by Lebesgue's Monotone Convergence Theorem,

$$\int_\Omega X\, 1_{B_n}\, dP \uparrow \int_\Omega X\, 1_{\bigcup_n A_n}\, dP = \nu(\bigcup_n A_n)\,.$$

It follows that $\sum_k \nu(A_k)$ is convergent and $\nu(\bigcup_n A_n) = \sum_n \nu(A_n)$. Hence ν is a finite measure on $(\Omega, \mathcal{G})$. ■

DR.RUPNATHJI(DR. RUPAK NATH)

If $P(A) = 0$, then $X 1_A = 0$ almost surely and (since $X1_A \geq 0$) we see that $\nu(A) = 0$. Thus $P(A) = 0 \implies \nu(A) = 0$ for $A \in \mathcal{G}$. We say that ν is absolutely continuous with respect to P on $\mathcal{G}$ (written $\nu << P$).

The following theorem is most relevant in this connection.

Theorem 2.2 (Radon-Nikodym). *Suppose that μ_1 and μ_2 are finite measures (on some $(\Omega, \mathcal{G})$) with $\mu_1 << \mu_2$. Then there is a $\mathcal{G}$-measurable μ_2-integrable function g ($g \in \mathcal{L}^1(\mathcal{G}, d\mu_2)$) on Ω such that $g \geq 0$ (μ_1-almost everywhere) and $\mu_1(A) = \int_A g \, d\mu_2$ for any $A \in \mathcal{G}$.*

With $\mu_1 = \nu$ and $\mu_2 = P$, we obtain the conditional expectation $E(X \,|\, \mathcal{G})$ as above. Note that if X is not necessarily positive, then we can write $X = X_+ - X_-$ with $X_\pm \geq 0$ to give $E(X \,|\, \mathcal{G}) = E(X_+ \,|\, \mathcal{G}) - E(X_- \,|\, \mathcal{G})$.

DR.RUPNATHJI(DR.RUPAK NATH)

Properties of the Conditional Expectation

Various basic properties of the conditional expectation are contained in the following proposition.

Proposition 2.3. *The conditional expectation enjoys the following properties.*

(i) *$E(X \,|\, \mathcal{G})$ is unique, almost surely.*

(ii) *If $X \geq 0$, then $E(X \,|\, \mathcal{G}) \geq 0$ almost surely, i.e., the conditional expectation is (almost surely) positivity preserving.*

(iii) *$E(X + Y \,|\, \mathcal{G}) = E(X \,|\, \mathcal{G}) + E(Y \,|\, \mathcal{G})$ almost surely.*

(iv) *For any $a \in \mathbb{R}$, $E(aX \,|\, \mathcal{G}) = aE(X \,|\, \mathcal{G})$ almost surely. Also $E(a \,|\, \mathcal{G}) = a$ almost surely.*

(v) *If $\mathcal{G} = \{\Omega, \varnothing\}$, the trivial σ-algebra, then $E(X \,|\, \mathcal{G}) = E(X)$ everywhere, i.e., $E(X \,|\, \mathcal{G})(\omega) = E(X)$ for every $\omega \in \Omega$.*

(vi) *(Tower Property) If $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $E(E(X \,|\, \mathcal{G}_2) \,|\, \mathcal{G}_1) = E(X \,|\, \mathcal{G}_1)$.*

(vii) *If $\sigma(X)$ and $\mathcal{G}$ are independent σ-algebras, then $E(X \,|\, \mathcal{G}) = E(X)$ almost surely.*

(viii) *For any $\mathcal{G}$, $E(E(X \,|\, \mathcal{G})) = E(X)$.*

Proof. (i) Suppose that f and g are positive $\mathcal{G}$-measurable and satisfy

$$\int_A f \, dP = \int_A g \, dP$$

for all $A \in \mathcal{G}$. Then $\int_A (f - g) \, dP = 0$ for all $A \in \mathcal{G}$ and so $f - g = 0$ almost surely, that is, if $A = \{\omega : f(\omega) \neq g(\omega)\}$, then $A \in \mathcal{G}$ and $P(A) = 0$. This last assertion follows from the following lemma.

Lemma 2.4. *If h is integrable on some finite measure space $(\Omega, \mathcal{S}, \mu)$ and satisfies $\int_A h \, d\mu = 0$ for all $A \in \mathcal{S}$, then $h = 0$ μ-almost everywhere.*

Proof of Lemma. Let $A_n = \{ \omega : h(\omega) \ge \frac{1}{n} \}$. Then

$$0 = \int_{A_n} h \, d\mu \ge \tfrac{1}{n} \int_{A_n} d\mu = \tfrac{1}{n} \, \mu(A_n)$$

and so $\mu(A_n) = 0$ for all $n \in \mathbb{N}$. But then it follows that

$$\mu(\underbrace{\{ \omega : h(\omega) > 0 \}}_{\bigcup_n A_n}) = \lim_n \mu(A_n) = 0 \, .$$

Let $B_n = \{ \omega : h(\omega) \le -\frac{1}{n} \}$. Then

$$0 = \int_{B_n} h \, d\mu \le -\tfrac{1}{n} \int_{B_n} d\mu = -\tfrac{1}{n} \, \mu(B_n)$$

so that $\mu(B_n) = 0$ for all $n \in \mathbb{N}$ and therefore

$$\mu(\underbrace{\{ \omega : h(\omega) < 0 \}}_{\bigcup_n B_n}) = \lim_n \mu(B_n) = 0 \, .$$

Hence

$$\mu(\{ \omega : h(\omega) \ne 0 \}) = \mu(h < 0) + \mu(h > 0) = 0 \, ,$$

that is, $\mu(h = 0) = 1$. ■

Remark 2.5. Note that we have used the standard shorthand notation such as $\mu(h > 0)$ for $\mu(\{ \omega : h(\omega) > 0 \})$. There is unlikely to be any confusion.

(ii) For notational convenience, let $\widehat{X}$ denote $E(X \,|\, \mathcal{G})$. Then for any $A \in \mathcal{G}$,

$$\int_A \widehat{X} \, dP = \int_A X \, dP \ge 0.$$

Let $B_n = \{ \omega : \widehat{X}(\omega) \le -\frac{1}{n} \}$. Then

$$\int_{B_n} \widehat{X} \, dP \le -\tfrac{1}{n} \int_{B_n} dP = -\tfrac{1}{n} \, P(B_n) \, .$$

However, the left hand side is equal to $\int_{B_n} X \, dP \ge 0$ which forces $P(B_n) = 0$. But then $P(\widehat{X} < 0) = \lim_n P(B_n) = 0$ and so $P(\widehat{X} \ge 0) = 1$.

(iii) For any choices of conditional expectation, we have

$$\int_A E(X + Y \,|\, \mathcal{G}) \, dP = \int_A (X + Y) \, dP$$

DR.RUPNATHJI(DR.RUPAK NATH)

$$= \int_A X\,dP + \int_A Y\,dP$$
$$= \int_A E(X\,|\,\mathcal{G})\,dP + \int_A E(Y\,|\,\mathcal{G})\,dP$$
$$= \int_A E(X+Y\,|\,\mathcal{G})\,dP$$

for all $A \in \mathcal{G}$. We conclude that $E(X+Y\,|\,\mathcal{G}) = E(X\,|\,\mathcal{G}) + E(Y\,|\,\mathcal{G})$ almost surely, as required.

(iv) This is just as the proof of (iii).

(v) With $A = \Omega$, we have

$$\int_A X\,dP = \int_\Omega X\,dP = E(X)$$
$$= \int_\Omega E(X)\,dP$$
$$= \int_A E(X)\,dP\,.$$

DR.RUPNATHJI(DR.RUPAK NATH)

Now with $A = \varnothing$,

$$\int_A X\,dP = \int_\varnothing X\,dP = 0 = \int_A E(X)\,dP\,.$$

So the everywhere constant function $\widehat{X} : \omega \mapsto E(X)$ is $\{\,\Omega, \varnothing\,\}$-measurable and obeys

$$\int_A \widehat{X}\,dP = \int_A X\,dP$$

for every $A \in \{\,\Omega, \varnothing\,\}$. Hence $\omega \mapsto E(X)$ is a conditional expectation of X. If X' is another, then $X' = \widehat{X}$ almost surely, so that $P(X' = \widehat{X}) = 1$. But the set $\{\,X' = \widehat{X}\,\}$ is $\{\,\Omega, \varnothing\,\}$-measurable and so is equal to either $\varnothing$ or to Ω. Since $P(X' = \widehat{X}) = 1$, it must be the case that $\{\,X' = \widehat{X}\,\} = \Omega$ so that $\widehat{X}(\omega) = X'(\omega)$ for all $\omega \in \Omega$.

(vi) Suppose that $\mathcal{G}_1$ and $\mathcal{G}_2$ are σ-algebras satisfying $\mathcal{G}_1 \subseteq \mathcal{G}_2$. Then, for any $A \in \mathcal{G}_1$,

$$\int_A E(X\,|\,\mathcal{G}_1)\,dP = \int_A X\,dP$$
$$= \int_A E(X\,|\,\mathcal{G}_2)\,dP \quad \text{since } A \in \mathcal{G}_2,$$
$$= \int_A E(E(X\,|\,\mathcal{G}_2)\,|\,\mathcal{G}_1)\,dP \quad \text{since } A \in \mathcal{G}_1$$

and we conclude that $E(X\,|\,\mathcal{G}_1) = E(E(X\,|\,\mathcal{G}_2)\,|\,\mathcal{G}_1)$ almost surely, as claimed.

(vii) For any $A \in \mathcal{G}$,

$$\begin{aligned}\int_A E(X\,|\,\mathcal{G})\,dP &= \int_A X\,dP = \int_\Omega X\,1_A\,dP \\ &= E(X\,1_A) = E(X)\,E(1_A) \\ &= \int_A E(X)\,dP.\end{aligned}$$

The result follows.

(viii) Denote $E(X\,|\,\mathcal{G})$ by $\widehat{X}$. Then for any $A \in \mathcal{G}$,

$$\int_A \widehat{X}\,dP = \int_A X\,dP\,.$$

In particular, with $A = \Omega$, we get

$$\int_\Omega \widehat{X}\,dP = \int_\Omega X\,dP\,,$$

that is, $E(\widehat{X}) = E(X)$. ∎

The next result is a further characterization of the conditional expectation.

Theorem 2.6. *Let $f \in \mathcal{L}^1(\Omega, \mathcal{S}, P)$. The conditional expectation $\widehat{f} = E(f\,|\,\mathcal{G})$ is characterized (almost surely) by*

$$\int_\Omega fg\,dP = \int_\Omega \widehat{f}\,g\,dP \qquad (*)$$

for all bounded $\mathcal{G}$-measurable functions g.

Proof. With $g = 1_A$, we see that $(*)$ implies that

$$\int_A f\,dP = \int_A \widehat{f}\,dP$$

for all $A \in \mathcal{G}$. It follows that $\widehat{f}$ is the required conditional expectation.

For the converse, suppose that we know that $(*)$ holds for all $f \geq 0$. Then, writing a general f as $f = f^+ - f^-$ (where $f^\pm \geq 0$ and $f^+ f^- = 0$ almost surely), we get

$$\begin{aligned}\int_\Omega fg\,dP &= \int_\Omega f^+ g\,dP - \int_\Omega f^- g\,dP \\ &= \int_\Omega \widehat{f^+}\,g\,dP - \int_\Omega \widehat{f^-}\,g\,dP\end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

$$= \int_\Omega (\,\widehat{f^+} - \widehat{f^+}\,)\, g\, dP$$
$$= \int_\Omega \widehat{f}\, g\, dP$$

and we see that $(*)$ holds for general $f \in \mathcal{L}^1$. Similarly, we note that by decomposing g as $g = g^+ - g^-$, it is enough to prove that $(*)$ holds for $g \geq 0$. So we need only show that $(*)$ holds for $f \geq 0$ and $g \geq 0$. In this case, we know that there is a sequence (s_n) of simple $\mathcal{G}$-measurable functions such that $0 \leq s_n \leq g$ and $s_n \to g$ everywhere. For fixed n, let $s_n = \sum a_j 1_{A_j}$ (finite sum). Then

$$\int_\Omega f\, s_n\, dP = \sum \int_{A_j} f\, dP = \sum \int_{A_j} \widehat{f}\, dP = \int_\Omega \widehat{f}\, s_n\, dP$$

giving the equality

$$\int_\Omega f\, s_n\, dP = \int_\Omega \widehat{f}\, s_n\, dP\,.$$

By Lebesgue's Dominated Convergence Theorem, the left hand side converges to $\int_\Omega f\, g\, dP$ and the right hand side converges to $\int_\Omega \widehat{f}\, g\, dP$ as $n \to \infty$, which completes the proof. ∎

Jensen's Inequality

We begin with a definition.

Definition 2.7. The function $\varphi : \mathbb{R} \to \mathbb{R}$ is said to be convex if

$$\varphi(a + s(b-a)) \leq \varphi(a) + s(\varphi(b) - \varphi(a)),$$

that is,

$$\varphi((1-s)a + sb)) \leq (1-s)\varphi(a) + s\varphi(b) \tag{2.1}$$

for any $a, b \in R$ and all $0 \leq s \leq 1$. The point $a + s(b-a) = (1-s)a + sb$ lies between a and b and the inequality (2.1) is the statement that the chord between the points $(a, \varphi(a))$ and $(b, \varphi(b))$ on the graph $y = \varphi(x)$ lies above the graph itself.

Let $u < v < w$. Then $v = u + s(w-u) = (1-s)u + sw$ for some $0 < s < 1$ and from (2.1) we have

$$(1-s)\varphi(v) + s\varphi(v) = \varphi(v) \leq (1-s)\varphi(u) + s\varphi(w) \tag{2.2}$$

which can be rearranged to give

$$(1-s)(\varphi(v) - \varphi(u)) \leq s(\varphi(w) - \varphi(v)). \tag{2.3}$$

DR.RUPNATHJI(DR.RUPAK NATH)

But $v = (1-s)u + sw = u + s(w-u) = (1-s)(u-w) + w$ so that $s = (v-u)/(w-u)$ and $1-s = (v-w)/(u-w) = (w-v)/(w-u)$. Inequality (2.3) can therefore be rewritten as

$$\frac{\varphi(v)-\varphi(u)}{v-u} \le \frac{\varphi(w)-\varphi(v)}{w-v}\,. \tag{2.4}$$

Again, from (2.2), we get

$$\varphi(v)-\varphi(u) \le (1-s)\varphi(u) + s\varphi(w) - \varphi(u) = s(\varphi(w)-\varphi(u))$$

and so, substituting for s,

$$\frac{\varphi(v)-\varphi(u)}{v-u} \le \frac{\varphi(w)-\varphi(u)}{w-u}\,. \tag{2.5}$$

Once more, from (2.2),

$$\varphi(v)-\varphi(w) \le (1-s)\varphi(u) + s\varphi(w) - \varphi(w) = (s-1)(\varphi(w)-\varphi(u))$$

which gives

$$\frac{\varphi(w)-\varphi(u)}{w-u} \le \frac{\varphi(w)-\varphi(v)}{w-v}\,. \tag{2.6}$$

These inequalities are readily suggested from a diagram.

Now fix $v = v_0$. Then by inequality (2.5), we see that the ratio (Newton quotient) $(\varphi(w)-\varphi(v_0))/(w-v_0)$ decreases as $w \downarrow v_0$ and, by (2.4), is bounded below by $(\varphi(v_0)-\varphi(u))/(v_0-u)$ for any $u < v_0$. Hence φ has a right derivative at v_0, i.e.,

$$\exists \lim_{w\downarrow v_0} \frac{\varphi(w)-\varphi(v_0)}{w-v_0} \equiv D^+\varphi(v_0).$$

Next, we consider $u \uparrow v_0$. By (2.6), the ratio $(\varphi(v_0)-\varphi(u))/(v_0-u)$ is increasing as $u \uparrow v_0$ and, by the inequality (2.4), is bounded above (by $(\varphi(w)-\varphi(v_0))/(w-v_0)$ for any $w > v_0$). It follows that φ has a left derivative at v_0, i.e.,

$$\exists \lim_{u\uparrow v_0} \frac{\varphi(v_0)-\varphi(u)}{v_0-u} \equiv D^-\varphi(v_0).$$

It follows that φ is continuous at v_0 because

$$\varphi(w)-\varphi(v_0) = (w-v_0)\left(\frac{\varphi(w)-\varphi(v_0)}{w-v_0}\right) \to 0 \quad \text{as } w \downarrow v_0$$

and

$$\varphi(v_0)-\varphi(u) = (v_0-u)\left(\frac{\varphi(v_0)-\varphi(u)}{v_0-u}\right) \to 0 \quad \text{as } u \uparrow v_0.$$

DR.RUPNATHJI(DR.RUPAK NATH)

By (2.5), letting $u \uparrow v_0$, we get

$$\begin{aligned} D^-\varphi(v_0) &\le \frac{\varphi(w)-\varphi(v_0)}{w-v_0} \\ &\le \frac{\varphi(w)-\varphi(\lambda)}{w-\lambda} \quad \text{for any } v_0 \le \lambda \le w, \text{ by (2.6)} \\ &\uparrow D^-\varphi(w) \quad \text{as } \lambda \uparrow w. \end{aligned}$$

Hence

$$D^-\varphi(v_0) \le D^-\varphi(w)$$

whenever $v \le w$. Similarly, letting $w \downarrow v$ in (2.6), we find

$$\frac{\varphi(v)-\varphi(u)}{v-u} \le D^+\varphi(v)$$

and so

$$\frac{\varphi(\lambda)-\varphi(u)}{\lambda-u} \le \frac{\varphi(v)-\varphi(u)}{v-u} \le D^+\varphi(v).$$

Letting $\lambda \downarrow u$, we get

$$D^+\varphi(u) \le D^+\varphi(v)$$

whenever $u \le v$. That is, both $D^+\varphi$ and $D^-\varphi$ are non-decreasing functions. Furthermore, letting $u \uparrow v_0$ and $w \downarrow v_0$ in (2.4), we see that

$$D^-\varphi(v_0) \le D^+(v_0)$$

at each v_0.

Claim. Fix v and let m satisfy $D^-\varphi(v) \le m \le D^+\varphi(v)$. Then

$$m(x-v)+\varphi(v) \le \varphi(x) \tag{2.7}$$

for any $x \in \mathbb{R}$.

Proof. For $x > v$, $(\varphi(x)-\varphi(v))/(x-v) \downarrow D^+\varphi(v)$ and so

$$\frac{\varphi(x)-\varphi(v)}{x-v} \ge D^+\varphi(v) \ge m$$

which means that $\varphi(x)-\varphi(v) \ge m(x-v)$ for $x > v$.

Now let $x < v$. Then $(\varphi(v)-\varphi(x))/(v-x) \uparrow D^-\varphi(v) \le m$ and so we see that $\varphi(v)-\varphi(x) \le m(v-x)$, i.e., $\varphi(x)-\varphi(v) \ge m(x-v)$ for $x < v$ and the claim is proved. ∎

Note that the inequality in the claim becomes equality for $x = v$.

DR.RUPNATHJI(DR.RUPAK NATH)

Let $\mathcal{A} = \{ (\alpha, \beta) : \varphi(x) \geq \alpha x + \beta \text{ for all } x \}$. From the remark above (and using the same notation), we see that for any $x \in \mathbb{R}$,

$$\varphi(x) = \sup\{ \alpha x + \beta : (\alpha, \beta) \in \mathcal{A} \}.$$

We wish to show that it is possible to replace the uncountable set $\mathcal{A}$ with a countable one.

For each $q \in \mathbb{Q}$, let $m(q)$ be a fixed choice such that

$$D^- \varphi(q) \leq m(q) \leq D^+ \varphi(q)$$

(for example, we could systematically choose $m(q) = D^- \varphi(q)$). Set $\alpha(q) = m(q)$ and $\beta(q) = \varphi(q) - m(q)q$. Let $\mathcal{A}_0 = \{ (\alpha(q), \beta(q)) : q \in \mathbb{Q} \}$. Then the discussion above shows that

$$\alpha x + \beta \leq \varphi(x)$$

for all $x \in \mathbb{R}$ and that $\varphi(q) = \sup\{ \alpha q + \beta : (\alpha, \beta) \in \mathcal{A}_0 \}$.

Claim. For any $x \in \mathbb{R}$, $\varphi(x) = \sup\{ \alpha x + \beta : (\alpha, \beta) \in \mathcal{A}_0 \}$.

Proof. Given $x \in \mathbb{R}$, fix $u < x < w$. Then we know that for any $q \in \mathbb{Q}$ with $u < q < w$,

$$D^- \varphi(u) \leq D^- \varphi(q) \leq D^+ \varphi(q) \leq D^+ \varphi(w). \tag{2.8}$$

Hence $D^- \varphi(u) \leq m(q) \leq D^+ \varphi(w)$. Let (q_n) be a sequence in $\mathbb{Q}$ with $u < q_n < w$ and such that $q_n \to x$ as $n \to \infty$. Then by (2.8), $(m(q_n))$ is a bounded sequence. We have

$$\begin{aligned}
0 &\leq \varphi(x) - (\alpha(q_n)x + \beta(q_n)) \\
&= \varphi(x) - \varphi(q_n) + \varphi(q_n) - (\alpha(q_n)x + \beta(q_n)) \\
&= (\varphi(x) - \varphi(q_n)) + \alpha(q_n)q_n + \beta(q_n) - (\alpha(q_n)x + \beta(q_n)) \\
&= (\varphi(x) - \varphi(q_n)) + m(q_n)(q_n - x)\,.
\end{aligned}$$

Since φ is continuous, the first term on the right hand side converges to 0 as $n \to \infty$ and so does the second term, because the sequence $(m(q_n))$ is bounded. It follows that $(\alpha(q_n)x + \beta(q_n)) \to \varphi(x)$ and therefore we see that $\varphi(x) = \sup\{ \alpha x + \beta : (\alpha, \beta) \in \mathcal{A}_0 \}$, as claimed. ∎

We are now in a position to discuss Jensen's Inequality.

DR.RUPNATHJI(DR.RUPAK NATH)

Theorem 2.8 (Jensen's Inequality). *Suppose that φ is convex on $\mathbb{R}$ and that both X and $\varphi(X)$ are integrable. Then*

$$\varphi(E(X \,|\, \mathcal{G})) \le E(\varphi(X) \,|\, \mathcal{G}) \quad \textit{almost surely.}$$

Proof. As above, let $\mathcal{A} = \{\, (\alpha, \beta) : \varphi(x) \ge \alpha x + \beta \text{ for all } x \in \mathbb{R} \,\}$. We have seen that there is a countable subset $\mathcal{A}_0 \subset \mathcal{A}$ such that

$$\varphi(x) = \sup_{(\alpha,\beta) \in \mathcal{A}_0} (\alpha x + \beta).$$

For any $(\alpha, \beta) \in \mathcal{A}$,

$$\alpha X(\omega) + \beta \le \varphi(X(\omega))$$

for all $\omega \in \Omega$. In other words, $\varphi(X) - (\alpha X + \beta) \ge 0$ on Ω. Hence, taking conditional expectations,

$$E(\varphi(X) - (\alpha X + \beta) \,|\, \mathcal{G}) \ge 0 \quad \text{almost surely.}$$

It follows that

$$E((\alpha X + \beta) \,|\, \mathcal{G}) \le E(\varphi(X) \,|\, \mathcal{G}) \quad \text{almost surely}$$

and so

$$\alpha \widehat{X} + \beta \le E(\varphi(X) \,|\, \mathcal{G}) \quad \text{almost surely,}$$

where $\widehat{X}$ is any choice of $E(X \,|\, \mathcal{G})$.

Now, for each $(\alpha, \beta) \in \mathcal{A}_0$, let $A(\alpha, \beta)$ be a set in $\mathcal{G}$ such that $P(A(\alpha, \beta)) = 1$ and

$$\alpha \widehat{X}(\omega) + \beta \le E(\varphi(X) \,|\, \mathcal{G})(\omega)$$

for every $\omega \in A(\alpha, \beta)$. Let $A = \bigcap_{(\alpha,\beta) \in \mathcal{A}_0}$. Since $\mathcal{A}_0$ is countable, $P(A) = 1$ and

$$\alpha \widehat{X}(\omega) + \beta \le E(\varphi(X) \,|\, \mathcal{G})(\omega)$$

for all $\omega \in A$. Taking the supremum over $(\alpha, \beta) \in \mathcal{A}_0$, we get

$$\underbrace{\sup_{(\alpha,\beta) \in \mathcal{A}_0} \alpha \widehat{X}(\omega) + \beta}_{\varphi(\widehat{X}(\omega)) = \varphi(\widehat{X})(\omega)} \le E(\varphi(X) \,|\, \mathcal{G})(\omega)$$

on A, that is,

$$\varphi(\widehat{X}) \le E(\varphi(X) \,|\, \mathcal{G})$$

almost surely and the proof is complete. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

Proposition 2.9. *Suppose that* $X \in \mathcal{L}^r$ *where* $r \geq 1$. *Then*

$$\| E(X \,|\, \mathcal{G}) \|_r \leq \| X \|_r .$$

In other words, the conditional expectation is a contraction on every $\mathcal{L}^r$ *with* $r \geq 1$.

Proof. Let $\varphi(s) = |s|^r$ and let $\widehat{X}$ be any choice for the conditional expectation $E(X \,|\, \mathcal{G})$. Then φ is convex and so by Jensen's Inequality

$$\varphi(\widehat{X}) \leq E(\varphi(X) \,|\, \mathcal{G}) \,, \quad \text{almost surely,}$$

that is

$$|\, \widehat{X} \,|^r \leq E(\, |X|^r \,|\, \mathcal{G}) \quad \text{almost surely.}$$

Taking expectations, we get

$$E(\, |\widehat{X}|^r \,) \leq E(\, E(\, |X|^r \,|\, \mathcal{G}) \,) = E(\, |X|^r \,) \,.$$

Taking r^{th} roots, gives

$$\| \, \widehat{X} \, \|_r \leq \| X \|_r$$

as claimed. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

Functional analytic approach to the conditional expectation

Consider a vector $\boldsymbol{u} = a\boldsymbol{i} + b\boldsymbol{j} + c\boldsymbol{k} \in \mathbb{R}^3$, where $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ denote the unit vectors in the Ox, Oy and Oz directions, respectively. Then the distance between $\boldsymbol{u}$ and the x-y plane is just $|c|$. The vector $\boldsymbol{u}$ can be written as $\boldsymbol{v} + \boldsymbol{w}$, where $\boldsymbol{v} = a\boldsymbol{i} + b\boldsymbol{k}$ and $w = c\boldsymbol{k}$. The vector $\boldsymbol{v}$ lies in the x-y plane and is orthogonal to the vector $\boldsymbol{w}$.

The generalization to general Hilbert spaces is as follows. Recall that a Hilbert space is a complete inner product space. (We consider only real Hilbert spaces, but it is more usual to discuss complex Hilbert spaces.)

Let X be a real linear space equipped with an "inner product", that is, a map $x, y \mapsto (x, y) \in \mathbb{R}$ such that

(i) $(x, y) \in \mathbb{R}$ and $(x, x) \geq 0$, for all $x \in X$,

(ii) $(x, y) = (y, x)$, for all $x, y \in X$,

(iii) $(ax + by, z) = a(x, z) + b(y, z)$, for all $x, y, z \in X$ and $a, b \in \mathbb{R}$.

Note: it is usual to also require that if $x \neq 0$, then $(x, x) > 0$ (which is why we have put the term "inner product" in quotation marks). This, of course, leads us back to the discussion of the distinction between $\mathcal{L}^2$ and L^2.

Example 2.10. Let $X = \mathcal{L}^2(\Omega, \mathcal{S}, P)$ with $(f, g) = \int_\Omega f(\omega) g(\omega)\, dP$ for any $f, g \in \mathcal{L}^2$.

In general, set $\|x\| = (x, x)^{1/2}$ for $x \in X$. Then in the example above, we observe that $\|f\| = \left(\int_\Omega f^2\, dP\right)^{1/2} = \|f\|_2$. It is important to note that $\|\cdot\|$ is not quite a norm. It can happen that $\|x\| = 0$ even though $x \neq 0$. Indeed, an example in $\mathcal{L}^2$ is provided by any function which is zero almost surely.

Proposition 2.11 (Parallelogram law). *For any $x, y \in X$,*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\,\|x\|^2 + \|y\|^2\,).$$

Proof. This follows by direct calculation using $\|w\|^2 = (w, w)$. ■

Definition 2.12. A subspace $V \subset X$ is said to be complete if every Cauchy sequence in V converges to an element of V, i.e., if (v_n) is a Cauchy sequence in V (so that $\|v_n - v_m\| \to 0$ as $m, n \to \infty$) then there is some $v \in V$ such that $v_n \to v$, as $n \to \infty$.

Theorem 2.13. *Let $x \in X$ and suppose that V is a complete subspace of X. Then there is some $v \in V$ such that $\|x - v\| = \inf_{y \in V} \|x - y\|$, that is, there is $v \in V$ so that $\|x - v\| = dist(x, V)$, the distance between x and V.*

Proof. Let (v_n) be any sequence in V such that

$$\|x - v_n\| \to d = \inf_{y \in V} \|x - y\|\,.$$

We claim that (v_n) is a Cauchy sequence. Indeed, the parallelogram law gives

$$\begin{aligned}
\|v_n - v_m\|^2 &= \|(x - v_m) + (v_n - x)\|^2 \\
&= 2\|x - v_m\|^2 + \|v_n - x\|^2 - \|\underbrace{(x - v_m) - (v_n - x)}_{2(x - \frac{1}{2}(v_m + v_n))}\|^2 \\
&= 2\|x - v_m\|^2 + \|v_n - x\|^2 - 4\|x - \tfrac{1}{2}(v_m + v_n)\|^2 \qquad (*)
\end{aligned}$$

Now

$$\begin{aligned}
d &\le \|x - \tfrac{1}{2}(\underbrace{v_m + v_n}_{\in V})\| = \|\tfrac{1}{2}(x - v_m) + \tfrac{1}{2}(x - v_m)\| \\
&\le \tfrac{1}{2}\underbrace{\|x - v_m\|}_{\to d} + \tfrac{1}{2}\underbrace{\|x - v_n\|}_{\to d}
\end{aligned}$$

so that $\|x - \frac{1}{2}(v_m + v_n)\| \to d$ as $m, n \to \infty$. Hence $(*) \to 2d^2 + 2d^2 - 4d^2 = 0$ as $m, n \to \infty$. It follows that (v_n) is a Cauchy sequence. By hypothesis, V is complete and so there is $v \in V$ such that $v_n \to v$. But then

$$d \le \|x - v\| \le \underbrace{\|x - v_n\|}_{\to d} + \underbrace{\|v_n - v\|}_{\to 0}$$

and so $d = \|x - v\|$. ■

DR.RUPNATHJI(DR.RUPAK NATH)

Proposition 2.14. *If $v' \in V$ satisfies $d = \|x - v'\|$, then $x - v' \perp V$ and $\|v - v'\| = 0$ (where v is as in the previous theorem).*

Conversely, if $v' \in V$ and $x - v' \perp V$, then $\|x - v'\| = d$ and $\|v - v'\| = 0$.

Proof. Suppose that there is $w \in V$ such that $(x - v', w) = \lambda \neq 0$. Let $u = v' + \alpha w$ where $\alpha \in R$. Then $u \in V$ and

$$\begin{aligned} \|x - u\|^2 = \|x - v' - \alpha w\|^2 &= \|x - v'\|^2 - 2\alpha(x - v', w) + \alpha^2 \|w\|^2 \\ &= d^2 - 2\alpha\lambda + \alpha^2 \|w\|^2. \end{aligned}$$

But the left hand side is $\geq d^2$, so by the definition of d we obtain

$$-2\alpha\lambda + \alpha^2\|w\|^2 = \alpha(-2\lambda + \alpha\|w\|^2) \geq 0$$

for any $\alpha \in R$. This is impossible. (The graph of $y = x(xb^2 - a)$ lies below the x-axis for all x between 0 and a/b^2.)

We conclude that there is no such w and therefore $x - v' \perp V$. But then we have

$$\begin{aligned} d^2 &= \|x - v\|^2 = \|(x - v') + (v' - v)\|^2 \\ &= \|x - v'\|^2 + 2\underbrace{(x - v', v' - v)}_{=0 \ \text{since } v' - v \in V} + \|v' - v\|^2 = d^2 + \|v' - v\|^2. \end{aligned}$$

It follows that $\|v' - v\| = 0$.

For the converse, suppose that $x - v' \perp V$. Then we calculate

$$\begin{aligned} d^2 &= \|x - v\|^2 = \|(x - v') + (v' - v)\|^2 \\ &= \|x - v'\|^2 + 2\underbrace{(x - v', v' - v)}_{=0 \ \text{since } v' - v \in V} + \|v - v'\|^2 \\ &= \underbrace{\|x - v'\|^2}_{\geq d^2} + \|v - v'\|^2 \qquad (*) \\ &\geq d^2 + \|v - v'\|^2. \end{aligned}$$

It follows that $\|v - v'\| = 0$ and then $(*)$ implies that $\|x - v'\| = d$. ∎

Suppose now that $\|\cdot\|$ satisfies the condition that $\|x\| = 0$ if and only if $x = 0$. Thus we are supposing that $\|\cdot\|$ really is a norm not just a seminorm on X. Then the equality $\|v - v'\| = 0$ is equivalent to $v = v'$. In this case, we can summarize the preceding discussion as follows.

Given any $x \in X$, there is a unique $v \in V$ such that $x - v \perp V$. With $x = v + x - v$ we see that we can write x as $x = v + w$ where $v \in V$ and $w \perp V$. This decomposition of x as the sum of an element $v \in V$ and an element $w \perp V$ is unique and means that we can define a map $P : X \to V$ by the formula $P : x \mapsto Px = v$. One checks that this is a linear

DR.RUPNATHJI(DR.RUPAK NATH)

map from X onto V. (Indeed, $Pv = v$ for all $v \in V$.) We also see that $P^2 x = PPx = Pv = v = Px$, so that $P^2 = P$. Moreover, for any $x, x' \in X$, write $x = v + w$ and $x' = v' + w'$ with $v, v' \in V$ and $w, w' \perp V$. Then

$$(Px, x') = (v, x') = (v, v' + w') = (v, v') = (v + w, v') = (x, Px').$$

A linear map P with the properties that $P^2 = P$ and $(Px, x') = (x, Px')$ for all $x, x' \in X$ is called an orthogonal projection. We say that P projects onto the subspace $V = \{ Px : x \in X \}$.

Suppose that Q is a linear map obeying these two conditions. Then we note that $Q(1-Q) = Q - Q^2 = 0$ and writing any $x \in X$ as $x = Qx + (1-Q)x$, we find that the terms Qx and $(1-Q)x$ are orthogonal

$$(Qx, (1-Q)x) = (Qx, x) - (Qx, Qx) = (Qx, x) - (Q^2 x, Qx) = 0.$$

So Q is the orthogonal projection onto the linear subspace $\{ Qx : x \in X \}$.

We now wish to apply this to the linear space $\mathcal{L}^2(\Omega, \mathcal{S}, P)$. Suppose that $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{S}$. Just as we construct $\mathcal{L}^2(\Omega, \mathcal{S}, P)$, so we may construct $\mathcal{L}^2(\Omega, \mathcal{G}, P)$. Since every $\mathcal{G}$-measurable function is also $\mathcal{S}$-measurable, it follows that $\mathcal{L}^2(\Omega, \mathcal{G}, P)$ is a linear subspace of $\mathcal{L}^2(\Omega, \mathcal{S}, P)$. The Riesz-Fisher Theorem tells us that $\mathcal{L}^2(\Omega, \mathcal{G}, P)$ is complete and so the above analysis is applicable, where now $V = \mathcal{L}^2(\Omega, \mathcal{G}, P)$.

Thus, any element $f \in \mathcal{L}^2(\Omega, \mathcal{S}, P)$ can be written as

$$f = \widehat{f} + g$$

where $\widehat{f} \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ and $g \perp \mathcal{L}^2(\Omega, \mathcal{G}, P)$. Because $\| \cdot \|_2$ is not a norm but only a seminorm on $\mathcal{L}^2(\Omega, \mathcal{S}, P)$, the functions $\widehat{f}$ and g are unique only in the sense that if also $f = f' + g'$ with $f' \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ and $g' \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ then $\| \widehat{f} - f' \|_2 = 0$, that is, $\widehat{f} \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ is unique almost surely.

If we were to apply this discussion to $L^2(\Omega, \mathcal{S}, P)$ and $L^2(\Omega, \mathcal{G}, P)$, then $\| \cdot \|_2$ is a norm and the object corresponding to $\widehat{f}$ should now be unique and be the image of $[f] \in L^2(\Omega, \mathcal{S}, P)$ under an orthogonal projection. However, there is a subtle point here. For this idea to go through, we must be able to identify $L^2(\Omega, \mathcal{G}, P)$ as a subspace of $L^2(\Omega, \mathcal{S}, P)$. It is certainly true that any element of $\mathcal{L}^2(\Omega, \mathcal{G}, P)$ is also an element of $\mathcal{L}^2(\Omega, \mathcal{S}, P)$, but is every element of $L^2(\Omega, \mathcal{G}, P)$ also an element of $L^2(\Omega, \mathcal{S}, P)$?

The answer lies in the sets of zero probability. Any element of $L^2(\Omega, \mathcal{G}, P)$ is a set (equivalence class) of the form $[f] = \{ f + \mathcal{N}(\mathcal{G}) \}$, where $\mathcal{N}(\mathcal{G})$ denotes the set of null functions that are $\mathcal{G}$-measurable. On the other hand, the corresponding element $[f] \in L^2(\Omega, \mathcal{S}, P)$ is the set $\{ f + \mathcal{N}(\mathcal{S}) \}$, where now $\mathcal{N}(\mathcal{S})$ is the set of $\mathcal{S}$-measurable null functions. It is certainly true that $\{ f + \mathcal{N}(\mathcal{G}) \} \subset \{ f + \mathcal{N}(\mathcal{S}) \}$, but in general there need not be equality. The notion of almost sure equivalence depends on the underlying σ-algebra. If we demand that $\mathcal{G}$ contain all the null sets of $\mathcal{S}$, then we do have equality

DR.RUPNATHJI(DR.RUPAK NATH)

$\{ f + \mathcal{N}(\mathcal{G}) \} = \{ f + \mathcal{N}(\mathcal{S}) \}$ and in this case it is true that $L^2(\Omega, \mathcal{G}, P)$ really is a subspace of $L^2(\Omega, \mathcal{S}, P)$. For any $x \in L^2(\Omega, \mathcal{S}, P)$, there is a unique element $v \in L^2(\Omega, \mathcal{G}, P)$ such that $x - v \perp L^2(\Omega, \mathcal{G}, P)$. Indeed, if $x = [f]$, with $f \in \mathcal{L}^2(\Omega, \mathcal{S}, P)$, then v is given by $v = [\,\widehat{f}\,]$.

Definition 2.15. For given $f \in \mathcal{L}^2(\Omega, \mathcal{S}, P)$, any element $\widehat{f} \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$ such that $f - \widehat{f} \perp \mathcal{L}^2(\Omega, \mathcal{G}, P)$ is called a version of the conditional expectation of f and is denoted $E(f \,|\, \mathcal{G})$.

On square-integrable random variables, the conditional expectation map $f \mapsto \widehat{f}$ is an orthogonal projection (subject to the ambiguities of sets of probability zero). We now wish to show that it is possible to recover the usual properties of the conditional expectation from this $\mathcal{L}^2$ (inner product business) approach.

Proposition 2.16. *For $f \in \mathcal{L}^2(\Omega, \mathcal{S}, P)$, (any version of) the conditional expectation $\widehat{f} = E(f \,|\, \mathcal{G})$ satisfies $\int_\Omega f(\omega) g(\omega)\, dP = \int_\Omega \widehat{f}(\omega) g(\omega)\, dP$ for any $g \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$. In particular,*

$$\int_A f\, dP = \int_A \widehat{f}\, dP$$

for any $A \in \mathcal{G}$.

Proof. By construction, $f - \widehat{f} \perp \mathcal{L}^2(\Omega, \mathcal{G}, P)$ so that

$$\int_\Omega (f - \widehat{f})\, g\, dP = 0$$

for any $g \in \mathcal{L}^2(\Omega, \mathcal{G}, P)$. In particular, for any $A \in \mathcal{G}$, set $g = 1_A$ to get the equality

$$\int_A f\, dP = \int_\Omega fg\, dP = \int_\Omega \widehat{f} g\, dP = \int_A \widehat{f}\, dP.$$

■

This is the defining property of the conditional expectation except that f belongs to $\mathcal{L}^2(\Omega, \mathcal{S}, P)$ rather than $\mathcal{L}^1(\Omega, \mathcal{S}, P)$. However, we can extend the result to cover the $\mathcal{L}^1$ case as we now show. First note that if $f \geq g$ almost surely, then $\widehat{f} \geq \widehat{g}$ almost surely. Indeed, by considering $h = f - g$, it is enough to show that $\widehat{h} \geq 0$ almost surely whenever $h \geq 0$ almost surely. But this follows directly from

$$\int_A \widehat{h}\, dP = \int_A h\, dP \geq 0$$

for all $A \in \mathcal{G}$. (If $B_n = \{ \widehat{h} < -1/n \}$ for $n \in \mathbb{N}$, then the inequalities $0 \leq \int_{B_n} \widehat{h}\, dP \leq -\frac{1}{n} P(B_n)$ imply that $P(B_n) = 0$. But then $P(\widehat{h} < 0) = \lim_n P(B_n) = 0$.)

DR.RUPNATHJI(DR.RUPAK NATH)

Proposition 2.17. *For $f \in \mathcal{L}^1(\Omega, \mathcal{S}, P)$, there exists $\widehat{f} \in \mathcal{L}^1(\Omega, \mathcal{G}, P)$ such that*

$$\int_A f \, dP = \int_A \widehat{f} \, dP$$

for any $A \in \mathcal{G}$. The function $\widehat{f}$ is unique almost surely.

Proof. By writing f as $f^+ - f^-$, it is enough to consider the case with $f \geq 0$ almost surely. For $n \in N$, set $f_n(\omega) = f(\omega) \wedge n$. Then $0 \leq f_n \leq n$ almost surely and so $f_n \in \mathcal{L}^2(\Omega, \mathcal{S}, P)$. Let $\widehat{f_n}$ be any version of the conditional expectation of f_n with respect to $\mathcal{G}$. Now, if $n > m$, then $f_n \geq f_m$ and so $\widehat{f_n} \geq \widehat{f_m}$ almost surely. That is, there is some event $B_{mn} \in \mathcal{G}$ with $P(B_{mn}) = 0$ and such that $\widehat{f_n} \geq \widehat{f_m}$ on B_{mn}^c (and $P(B_{mn}^c) = 1$). Let $B = \bigcup_{m,n} B_{mn}$. Then $P(B) = 0$, $P(B^c) = 1$ and $\widehat{f_n} \geq \widehat{f_m}$ on B^c for all m, n with $m \leq n$. Set

$$\widehat{f}(\omega) = \begin{cases} \sup_n \widehat{f_n}(\omega), & \omega \in B^c \\ 0, & \omega \in B. \end{cases}$$

Then for any $A \in \mathcal{G}$,

$$\int_A f_n \, dP = \int_\Omega f_n \, 1_A \, dP = \int_\Omega \widehat{f_n} \, 1_A \, dP = \int_\Omega \widehat{f_n} \, 1_{B^c} \, 1_A \, dP$$

because $P(B^c) = 1$. The left hand side $\to \int_\Omega f \, 1_A \, dP$ by Lebesgue's Dominated Convergence Theorem. Applying Lebesgue's Monotone Convergence Theorem to the right hand side, we see that

$$\int_\Omega \widehat{f_n} \, 1_{B^c} 1_A \, dP \to \int_\Omega \widehat{f} \, 1_{B^c} \, 1_A \, dP = \int_A \widehat{f} \, dP$$

which gives the equality $\int_A f \, dP = \int_A \widehat{f} \, dP$.

Taking $A = \Omega$, we see that $\widehat{f} \in \mathcal{L}^1(\Omega, \mathcal{G}, P)$. ∎

Of course, the function $\widehat{f}$ is called (a version of) the conditional expectation $E(f \,|\, \mathcal{G})$ and we have recovered our original construction as given via the Radon-Nikodym Theorem.

DR.RUPNATHJI(DR.RUPAK NATH)

Chapter 3

Martingales

Let $(\Omega, \mathcal{S}, P)$ be a probability space. We consider a sequence $(\mathcal{F}_n)_{n\in\mathbb{Z}^+}$ of sub-σ-algebras of a fixed σ-algebra $\mathcal{F} \subset \mathcal{S}$ obeying

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}.$$

Such a sequence is called a filtration (upward filtering). The intuitive idea is to think of $\mathcal{F}_n$ as those events associated with outcomes of interest occurring up to "time n". Think of n as (discrete) time.

Definition 3.1. A martingale (with respect to $(\mathcal{F}_n)$) is a sequence (ξ_n) of random variables such that

1. Each ξ_n is integrable: $E(|\xi_n|) < \infty$ for all $n \in \mathbb{Z}^+$.
2. ξ_n is measurable with respect to $\mathcal{F}_n$ (we say (ξ_n) is *adapted*).
3. $E(\xi_{n+1} \mid \mathcal{F}_n) = \xi_n$, almost surely, for all $n \in \mathbb{Z}^+$.

Note that (1) is required in order for (3) to make sense. (The conditional expectation $E(\xi \mid \mathcal{G})$ is not defined unless ξ is integrable.)

Remark 3.2. Suppose that (ξ_n) is a martingale. For any $n > m$, we have

$$\begin{aligned} E(\xi_n \mid \mathcal{F}_m) &= E(E(\xi_n \mid \mathcal{F}_{n-1}) \mid \mathcal{F}_m) \text{ almost surely, by the tower property} \\ &= E(\xi_{n-1} \mid \mathcal{F}_m) \text{ almost surely} \\ &= \cdots \\ &= E(\xi_{m+1} \mid \mathcal{F}_m) \text{ almost surely} \\ &= \xi_m \text{ almost surely.} \end{aligned}$$

That is,

$$E(\xi_n \mid \mathcal{F}_m) = \xi_m \text{ almost surely}$$

for all $n \geq m$. (This could have been taken as part of the definition.)

DR.RUPNATHJI(DR.RUPAK NATH)

Example 3.3. Let X_0, X_1, X_2, ... be independent integrable random variables with mean zero. For each $n \in \mathbb{Z}^+$ let $\mathcal{F}_n = \sigma(X_0, X_1, \ldots, X_n)$ and let $\xi_n = X_0 + X_1 + \cdots + X_n$. Evidently (ξ_n) is adapted and

$$\begin{aligned} E(|\xi_n|) &= E(|X_0 + X_1 + \cdots + X_n|) \\ &\le E(|X_0|) + E(|X_1|) + \cdots + E(|X_n|) \\ &< \infty \end{aligned}$$

so that ξ_n is integrable. Finally, we note that (almost surely)

$$\begin{aligned} E(\xi_{n+1} \,|\, \mathcal{F}_n) &= E(X_{n+1} + \xi_n \,|\, \mathcal{F}_n) \\ &= E(X_{n+1} \,|\, \mathcal{F}_n) + E(\xi_n \,|\, \mathcal{F}_n) \\ &= E(X_{n+1}) + \xi_n \\ &\qquad \text{since } \xi_n \text{ is adapted and } X_{n+1} \text{ and } \mathcal{F}_n \text{ are independent} \\ &= \xi_n \end{aligned}$$

and so (ξ_n) is a martingale.

Example 3.4. Let $X \in \mathcal{L}^1(\mathcal{F})$ and let $\xi_n = E(X \,|\, \mathcal{F}_n)$. Then $\xi_n \in \mathcal{L}^1(\mathcal{F}_n)$ and

$$\begin{aligned} E(\xi_{n+1} \,|\, \mathcal{F}_n) &= E(E(X \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n) \\ &= E(X \,|\, \mathcal{F}_n) \text{ almost surely} \\ &= \xi_n \text{ almost surely} \end{aligned}$$

so that (ξ_n) is a martingale.

Proposition 3.5. *For any martingale* (ξ_n), $E(\xi_n) = E(\xi_0)$.

Proof. We note that $E(X) = E(X \,|\, \mathcal{G})$, where $\mathcal{G}$ is the trivial σ-algebra, $\mathcal{G} = \{\, \Omega, \varnothing \,\}$. Since $\mathcal{G} \subset \mathcal{F}_n$ for all n, we can apply the tower property to deduce that

$$E(\xi_n) = E(\xi_n \,|\, \mathcal{G}) = E(E(\xi_n \,|\, \mathcal{F}_0) \,|\, \mathcal{G}) = E(\xi_0 \,|\, \mathcal{G}) = E(\xi_0)$$

as required. ■

Definition 3.6. The sequence $X_0, X_1, X_2, \ldots$ of random variables is said to be a supermartingale with respect to the filtration $(\mathcal{F}_n)$ if the following three conditions hold.

1. Each X_n is integrable.
2. (X_n) is adapted to $(\mathcal{F}_n)$.
3. For each $n \in \mathbb{Z}^+$, $X_n \ge E(X_{n+1} \,|\, \mathcal{F}_n)$ almost surely.

DR.RUPNATHJI(DR.RUPAK NATH)

The sequence $X_0, X_1, \ldots$ is said to be a submartingale if both (1) and (2) above hold and also the following inequalities hold.

3′. For each $n \in \mathbb{Z}^+$, $X_n \le E(X_{n+1} \,|\, \mathcal{F}_n)$ almost surely.

Evidently, (X_n) is a submartingale if and only if $(-X_n)$ is a supermartingale.

Example 3.7. Let $X = (X_n)_{n\in\mathbb{Z}^+}$ be a supermartingale (with respect to a given filtration $(\mathcal{F}_n)$) such that $E(X_n) = E(X_0)$ for all $n \in \mathbb{Z}^+$. Then X is a martingale. Indeed, let $Y = X_n - E(X_{n+1} \,|\, \mathcal{F}_n)$. Since (X_n) is a supermartingale, $Y \ge 0$ almost surely. Taking expectations, we find that $E(Y) = E(X_n) - E(X_{n+1}) = 0$, since $E(X_{n+1}) = E(X_n) = E(X_0)$. It follows that $Y = 0$, almost surely, that is, (X_n) is a martingale.

Example 3.8. Let (Y_n) be a sequence of independent random variables such that Y_n and e^{Y_n} are integrable for all $n \in \mathbb{Z}^+$ and such that $E(Y_n) \ge 0$. For $n \in \mathbb{Z}^+$, let

$$X_n = \exp(Y_0 + Y_1 + \cdots + Y_n)\,.$$

Then (X_n) is a submartingale with respect to the filtration $(\mathcal{F}_n)$ where $\mathcal{F}_n = \sigma(Y_0, \ldots, Y_n)$, the σ-algebra generated by $Y_0, Y_1, \ldots, Y_n$.

To see this, we first show that each X_n is integrable. We have

$$X_n = e^{(Y_0 + \ldots + Y_n)} = e^{Y_0}\, e^{Y_1} \ldots e^{Y_n}\,.$$

Each term e^{Y_j} is integrable, by hypothesis, but why should the product be? It is the independence of the Y_js which does the trick — indeed, we have

$$\begin{aligned} E(X_n) &= E(e^{(Y_0 + \cdots + Y_n)}) \\ &= E(e^{Y_0} e^{Y_1} \ldots e^{Y_n}) \\ &= E(e^{Y_0})\, E(e^{Y_1}) \ldots E(e^{Y_n})\,, \end{aligned}$$

by independence.

[Let $f_j = e^{Y_j}$ and for $m \in \mathbb{N}$, set $f_j^m = f_j \wedge m$. Each f_j^m is bounded and the f_j^ms are independent. Note also that each f_j is non-negative. Then $E(f_0^m \ldots f_n^m) = E(f_0^m)\,E(f_1^m) \ldots E(f_n^m)$, by independence. Letting $m \to \infty$ and applying Lebesgue's Monotone Convergence Theorem, we deduce that the product $f_0 \ldots f_n$ is integrable (and its integral is given by the product $E(f_0) \ldots E(f_n)$).]

By construction, (X_n) is adapted. To verify the submartingale inequality, we consider

$$\begin{aligned} E(X_{n+1} \,|\, \mathcal{F}_n) &= E(e^{Y_0 + Y_1 + \cdots + Y_n + Y_{n+1}} \,|\, \mathcal{F}_n) = E(X_n\, e^{Y_{n+1}} \,|\, \mathcal{F}_n) \\ &= X_n\, E(e^{Y_{n+1}} \,|\, \mathcal{F}_n), \quad \text{since } X_n \text{ is } \mathcal{F}_n\text{-measurable,} \\ &= X_n\, E(e^{Y_{n+1}}), \quad \text{by independence,} \\ &\ge X_n\, e^{E(Y_{n+1})}, \quad \text{by Jensen's inequality, } (s \mapsto e^s \text{ is convex}) \\ &\ge X_n, \quad \text{since } E(Y_{n+1}) \ge 0. \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

Proposition 3.9.

(i) *Suppose that* (X_n) *and* (Y_n) *are submartingales. Then* $(X_n \vee Y_n)$ *is also a submartingale.*

(ii) *Suppose that* (X_n) *and* (Y_n) *are supermartingales. Then* $(X_n \wedge Y_n)$ *is also a supermartingale.*

Proof. (i) Set $Z_n = X_n \vee Y_n$. Then $Z_k \geq X_k$ and $Z_k \geq Y_k$ for all k and so

$$E(Z_{n+1} \,|\, \mathcal{F}_n) \geq E(X_{n+1} \,|\, \mathcal{F}_n) \geq X_n \quad \text{almost surely}$$

and similarly $E(Z_{n+1} \,|\, \mathcal{F}_n) \geq E(Y_{n+1} \,|\, \mathcal{F}_n) \geq Y_n$ almost surely. It follows that $E(Z_{n+1} \,|\, \mathcal{F}_n) \geq Z_n$ almost surely, as required.

(If $A = \{ E(Z_{n+1} \,|\, \mathcal{F}_n) \geq X_n \}$ and $B = \{ E(Z_{n+1} \,|\, \mathcal{F}_n) \geq Y_n \}$, then $A \cap B = \{ E(Z_{n+1} \,|\, \mathcal{F}_n) \geq Z_n \}$. However, $P(A) = P(B) = 1$ and so $P(A \cap B) = 1$.)

(ii) Set $U_n = X_n \wedge Y_n$, so that $U_k \leq X_k$ and $U_k \leq Y_k$ for all k. It follows that

$$E(U_{n+1} \,|\, \mathcal{F}_n) \leq E(X_{n+1} \,|\, \mathcal{F}_n) \leq X_n \quad \text{almost surely}$$

and similarly $E(U_{n+1} \,|\, \mathcal{F}_n) \leq E(Y_{n+1} \,|\, \mathcal{F}_n) \leq Y_n$ almost surely. We conclude that $E(X_{n+1} \wedge Y_{n+1} \,|\, \mathcal{F}_n) \leq X_n \wedge Y_n$ almost surely, as required. ∎

Proposition 3.10. *Suppose that* $(\xi_n)_{n \in \mathbb{Z}^+}$ *is a martingale and* $\xi_n \in \mathcal{L}^2$ *for each* n*. Then* (ξ_n^2) *is a submartingale.*

Proof. The function $\varphi : t \mapsto t^2$ is convex and so by Jensen's inequality

$$\varphi(\underbrace{E(\xi_{n+1} \,|\, \mathcal{F}_n)}_{\xi_n}) \leq E(\varphi(\xi_{n+1}) \,|\, \mathcal{F}_n) \text{ almost surely.}$$

That is,

$$\xi_n^2 \leq E(\xi_{n+1}^2 \,|\, \mathcal{F}_n) \text{ almost surely}$$

as required. ∎

Theorem 3.11 (Orthogonality of martingale increments). *Suppose that* (X_n) *is an* $\mathcal{L}^2$*-martingale with respect to the filtration* $(\mathcal{F}_n)$*. Then*

$$E(\,(X_n - X_m)\, Y\,) = 0$$

whenever $n \geq m$ *and* $Y \in \mathcal{L}^2(\mathcal{F}_m)$*. In particular,*

$$E(\,(X_k - X_j)\,(X_n - X_m)\,) = 0$$

for any $0 \leq j \leq k \leq m \leq n$*. In other words, the increments* $(X_k - X_j)$ *and* $(X_n - X_m)$ *are orthogonal in* $\mathcal{L}^2$*.*

DR.RUPNATHJI(DR.RUPAK NATH)

Proof. Note first that $(X_n - X_m)Y$ is integrable. Next, using the "tower property" and the $\mathcal{F}_m$-measurability of Y, we see that

$$\begin{aligned} E((X_n - X_m)\,Y) &= E(E((X_n - X_m)\,Y \,|\, \mathcal{F}_m)) \\ &= E(E((X_n - X_m \,|\, \mathcal{F}_m)\,Y) \\ &= 0 \end{aligned}$$

since $E(X_n - X_m \,|\, \mathcal{F}_m) = X_m - X_m = 0$.

The orthogonality of the martingale increments follows immediately by taking $Y = X_k - X_j$. ∎

Gambling

It is customary to mention gambling. Consider a sequence $\eta_1, \eta_2, \dots$ of random variables where η_n is thought of as the "winnings per unit stake" at game play n. If a gambler places a unit stake at each game, then the total winnings after n games is $\xi_n = \eta_1 + \eta_2 + \cdots + \eta_n$.

For $n \in \mathbb{N}$, let $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$ and set $\xi_0 = 0$ and $\mathcal{F}_0 = \{\,\Omega, \varnothing\,\}$. To say that (ξ_n) is a martingale is to say that

$$E(\xi_{n+1} \,|\, \mathcal{F}_n) = \xi_n \text{ almost surely,}$$

or

$$E(\xi_{n+1} - \xi_n \,|\, \mathcal{F}_n) = 0 \text{ almost surely.}$$

We can interpret this as saying that knowing everything up to play n, we expect the gain at the next play to be zero. We gain no advantage nor suffer any disadvantage at play $n+1$ simply because we have already played n games. In other words, the game is "fair".

On the other hand, if (ξ_n) is a super martingale, then

$$\xi_n \ge E(\xi_{n+1} \,|\, \mathcal{F}_n) \text{ almost surely.}$$

This can be interpreted as telling us that the knowledge of everything up to (and including) game n suggests that our winnings after game $n+1$ will be less than the winnings after n games. It is an unfavourable game for us.

If (ξ_n) is a submartingale, then arguing as above, we conclude that the game is in our favour.

Note that if (ξ_n) is a martingale, then

$$E(\xi_n) = E(\xi_1)\ (\,= E(\xi_0)) = 0 \text{ in this case)}$$

so the expected total winnings never changes.

DR.RUPNATHJI(DR.RUPAK NATH)

Definition 3.12. A stochastic process $(\alpha_n)_{n\in\mathbb{N}}$ is said to be previsible (or predictable) with respect to a filtration $(\mathcal{F}_n)$ if α_n is $\mathcal{F}_{n-1}$-measurable for each $n \in \mathbb{N}$.
Note that one often extends the labelling to $\mathbb{Z}^+$ and sets $\alpha_0 = 0$, mainly for notational convenience.

Using the gambling scenario, we might think of α_n as the number of stakes bought at game n. Our choice for α_n can be based on all events up to and including game $n-1$. The requirement that α_n be $\mathcal{F}_{n-1}$-measurable is quite reasonable and natural in this setting.

The total winnings after n games will be

$$\zeta_n = \alpha_1\eta_1 + \cdots + \alpha_n\eta_n .$$

However, $\xi_n = \eta_1 + \cdots + \eta_n = \xi_{n-1} + \eta_n$ and so

$$\zeta_n = \alpha_1(\xi_1 - \xi_0) + \alpha_2(\xi_2 - \xi_1) + \cdots + \alpha_n(\xi_n - \xi_{n-1}) .$$

This can also be expressed as

$$\zeta_{n+1} = \zeta_n + \alpha_{n+1}\eta_{n+1} = \zeta_n + \alpha_{n+1}(\xi_{n+1} - \xi_n) .$$

Example 3.13. For each $k \in \mathbb{Z}^+$, let B_k be a Borel subset of $\mathbb{R}^{k+1}$ and set

$$\alpha_{k+1} = \begin{cases} 1, & \text{if } (\eta_0, \eta_1, \dots, \eta_k) \in B_k \\ 0, & \text{otherwise.} \end{cases}$$

Then α_{k+1} is $\sigma(\eta_0, \eta_1, \dots, \eta_k)$-measurable. We have a "gaming strategy". One decides whether to play at game $k+1$ depending on the outcomes up to and including game k, namely, whether the outcome belongs to B_k or not.

DR.RUPNATHJI(DR.RUPAK NATH)

In the following discussion, let $\mathcal{F}_0 = (\Omega, \varnothing)$ and set $\zeta_0 = 0$.

Theorem 3.14. *Let (α_n) be a predictable process and as above, let (ζ_n) be the process $\zeta_n = \alpha_1(\xi_1 - \xi_0) + \cdots + \alpha_n(\xi_n - \xi_{n-1})$.*

(i) *Suppose that (α_n) is bounded and that (ξ_n) is a martingale. Then (ζ_n) is a martingale.*

(ii) *Suppose that (α_n) is bounded, $\alpha_n \geq 0$ and that the process (ξ_n) is a supermartingale. Then (ζ_n) is a supermartingale.*

(iii) *Suppose that (α_n) is bounded, $\alpha_n \geq 0$ and that the process (ξ_n) is a submartingale. Then (ζ_n) is a submartingale.*

Proof. First note that

$$|\zeta_n| \le |\alpha_1|\,(|\xi_1| + |\xi_0|) + \dots + |\alpha_n|\,(|\xi_n| + |\xi_{n-1}|)$$

and so ζ_n is integrable because each ξ_k is and (α_n) is bounded.

Next, we have

$$\begin{aligned} E(\zeta_n \,|\, \mathcal{F}_{n-1}) &= E(\zeta_{n-1} + \alpha_n(\xi_n - \xi_{n-1}) \,|\, \mathcal{F}_{n-1}) \\ &= \zeta_{n-1} + E(\alpha_n(\xi_n - \xi_{n-1}) \,|\, \mathcal{F}_{n-1}) \\ &= \zeta_{n-1} + \alpha_n \, \underbrace{\underbrace{E((\xi_n - \xi_{n-1}) \,|\, \mathcal{F}_{n-1})}_{E(\xi_n \,|\, \mathcal{F}_{n-1}) - \xi_{n-1}}}_{\Phi_n} \end{aligned}$$

That is,

$$E(\zeta_n \,|\, \mathcal{F}_{n-1}) - \zeta_{n-1} = \alpha_n \, \Phi_n \,.$$

Now, in case (i), $\Phi_n = 0$ almost surely, so $E(\zeta_n \,|\, \mathcal{F}_{n-1}) = \zeta_{n-1}$ almost surely.

In case (ii), $\Phi_n \le 0$ almost surelyand therefore $\alpha_n \Phi_n \le 0$ almost surely. It follows that $\zeta_{n-1} \ge E(\zeta_n \,|\, \mathcal{F}_{n-1})$ almost surely.

In case (iii), $\Phi_n \ge 0$ almost surely and so $\alpha_n \Phi_n \ge 0$ almost surely and therefore $\zeta_{n-1} \le E(\zeta_n \,|\, \mathcal{F}_{n-1})$ almost surely. ■

Remarks 3.15.

1. This last result can be interpreted as saying that no matter what strategy one adopts, it is not possible to make a fair game "unfair", a favourable game unfavourable or an unfavourable game favourable.

2. The formula

$$\zeta_n = \alpha_1(\xi_1 - \xi_0) + \alpha_2(\xi_2 - \xi_1) + \dots + \alpha_n(\xi_n - \xi_{n-1})$$

is the martingale transform of ξ by α. The general definition follows.

Definition 3.16. Given an adapted process X and a predictable process C, the (martingale) transform of X by C is the process $C \cdot X$ with

$$\begin{aligned} (C \cdot X)_0 &= 0 \\ (C \cdot X)_n &= \sum_{1 \le k \le n} C_k \, (X_k - X_{k_1}) \end{aligned}$$

(which means that $(C \cdot X)_n - (C \cdot X)_{n-1} = C_n(X_n - X_{n-1})$).

We have seen that if C is bounded and X is a martingale, then $C \cdot X$ is also a martingale. Moreover, if $C \ge 0$ almost surely, then $C \cdot X$ is a submartingale if X is and $C \cdot X$ is a supermartingale if X is.

DR.RUPNATHJI(DR.RUPAK NATH)

Stopping Times

A map $\tau : \Omega \to \mathbb{Z}^+ \cup \{\, \infty \,\}$ is said to be a stopping time (or a Markov time) with respect to the filtration $(\mathcal{F}_n)$ if

$$\{\, \tau \le n \,\} \in \mathcal{F}_n \quad \text{for each } n \in \mathbb{Z}^+.$$

One can think of this as saying that the information available by time n should be sufficient to tell us whether something has "stopped by time n" or not. For example, we should not need to be watching out for a company's profits in September if we only want to know whether it went bust in May.

Proposition 3.17. *Let $\tau : \Omega \to \mathbb{Z}^+ \cup \{\, \infty \,\}$. Then τ is a stopping time (with respect to the filtration $(\mathcal{F}_n)$) if and only if*

$$\{\, \tau = n \,\} \in \mathcal{F}_n \quad \textit{for every } n \in \mathbb{Z}^+.$$

Proof. If τ is a stopping time, then $\{\, \tau \le n \,\} \in \mathcal{F}_n$ for all $n \in \mathbb{Z}^+$. But $\{\, \tau = 0 \,\} = \{\, \tau \le 0 \,\}$ and

$$\{\, \tau = n \,\} = \underbrace{\{\, \tau \le n \,\}}_{\in \mathcal{F}_n} \cap \underbrace{\{\, \tau \le n-1 \,\}^c}_{\in \mathcal{F}_{n-1} \subset \mathcal{F}_n}$$

for any $n \in \mathbb{N}$. Hence $\{\, \tau = n \,\} \in \mathcal{F}_n$ for all $n \in \mathbb{Z}^+$.

For the converse, suppose that $\{\, \tau = n \,\} \in \mathcal{F}_n$ for all $n \in \mathbb{Z}^+$. Then

$$\{\, \tau \le n \,\} = \bigcup_{0 \le k \le n} \{\, \tau = k \,\}.$$

Each event $\{\, \tau = k \,\}$ belongs to $\mathcal{F}_k \subset \mathcal{F}_n$ if $k \le n$ and therefore it follows that $\{\, \tau \le n \,\} \in \mathcal{F}_n$. ∎

Proposition 3.18. *For any fixed $k \in \mathbb{Z}^+$, the constant time $\tau(\omega) = k$ is a stopping time.*

Proof. If $k > n$, then $\{\, \tau \le n \,\} = \varnothing \in \mathcal{F}_n$. On the other hand, if $k \le n$, then $\{\, \tau \le n \,\} = \Omega \in \mathcal{F}_n$. ∎

Proposition 3.19. *Let σ and τ be stopping times (with respect to $(\mathcal{F}_n)$). Then $\sigma + \tau$, $\sigma \vee \tau$ and $\sigma \wedge \tau$ are also stopping times.*

Proof. Since $\sigma(\omega) \ge 0$ and $\tau(\omega) \ge 0$, we see that

$$\{\, \sigma + \tau = n \,\} = \bigcup_{0 \le k \le n} \{\, \sigma = k \,\} \cap \{\, \tau = n - k \,\}.$$

Hence $\{\, \sigma + \tau \le n \,\} \in \mathcal{F}_n$.

DR.RUPNATHJI(DR.RUPAK NATH)

Next, we note that

$$\{\, \sigma \vee \tau \le n \,\} = \{\, \sigma \le n \,\} \cap \{\, \tau \le n \,\} \in \mathcal{F}_n$$

and so $\sigma \vee \tau$ is a stopping time.

Finally, we have

$$\{\, \sigma \wedge \tau \le n \,\} = \{\, \sigma \wedge \tau > n \,\}^c = (\{\, \sigma > n \,\} \cap \{\, \tau > n \,\})^c \in \mathcal{F}_n$$

and therefore $\sigma \wedge \tau$ is a stopping time. ■

Definition 3.20. Let X be a process and τ a stopping time. The process X stopped by τ is the process (X_n^τ) with

$$X_n^\tau(\omega) = X_{\tau(\omega)\wedge n}(\omega) \quad \text{for } \omega \in \Omega.$$

So if the outcome is ω and say, $\tau(\omega) = 23$, then $X_n^\tau(\omega)$ is given by

$$\begin{aligned} &X_{\tau(\omega)\wedge 1}(\omega), X_{\tau(\omega)\wedge 2}(\omega), \dots \\ &\qquad = X_1(\omega), X_2(\omega), \dots, X_{22}(\omega), X_{23}(\omega), X_{23}(\omega), X_{23}(\omega), \dots \end{aligned}$$

$X_{\tau\wedge n}$ is constant for $n \ge 23$. Of course, for another outcome ω', with $\tau(\omega') = 99$ say, then $X_{\tau\wedge n}$ takes values

$$X_1(\omega'), X_2(\omega'), \dots, X_{98}(\omega'), X_{99}(\omega'), X_{99}(\omega'), X_{99}(\omega'), \dots$$

Proposition 3.21. *If (X_n) is a martingale (resp., submartingale, supermartingale), so is the stopped process (X_n^τ).*

Proof. Firstly, we note that

$$X_n^\tau(\omega) = X_{\tau(\omega)\wedge n}(\omega) = \begin{cases} X_k(\omega) & \text{if } \tau(\omega) = k \text{ with } k \le n \\ X_n(\omega) & \text{if } \tau(\omega) > n, \end{cases}$$

that is,

$$X_n^\tau = \sum_{k=0}^{n} X_k \, 1_{\{\,\tau = k\,\}} + X_n \, 1_{\{\,\tau \le n\,\}^c} \,.$$

Now, $\{\, \tau = k \,\} \in \mathcal{F}_k$ and $\{\, \tau \le n \,\}^c \in \mathcal{F}_n$ and so it follows that X_n^τ is $\mathcal{F}_n$-measurable. Furthermore, each term on the right hand side is integrable and so we deduce that (X_n^τ) is adapted and integrable.

DR.RUPNATHJI(DR.RUPAK NATH)

Next, we see that

$$\begin{aligned} X^\tau_{n+1} - X^\tau_n &= \sum_{k=0}^{n+1} X_k \, 1_{\{\,\tau=k\,\}} + X_{n+1} \, 1_{\{\,\tau>n+1\,\}} \\ &\qquad - \sum_{k=0}^{n} X_k \, 1_{\{\,\tau=k\,\}} - X_n \, 1_{\{\,\tau>n\,\}} \\ &= X_{n+1} \, 1_{\{\,\tau=n+1\,\}} + X_{n+1} \, 1_{\{\,\tau>n+1\,\}} - X_n \, 1_{\{\,\tau>n\,\}} \\ &= X_{n+1} \, 1_{\{\,\tau\ge n+1\,\}} - X_n \, 1_{\{\,\tau>n\,\}} \\ &= (X_{n+1} - X_n) \, 1_{\{\,\tau\ge n+1\,\}} \,. \end{aligned}$$

Taking the conditional expectation, the left hand side gives

$$E(X^\tau_{n+1} - X^\tau_n \,|\, \mathcal{F}_n) = E(X^\tau_{n+1} \,|\, \mathcal{F}_n) - X^\tau_n$$

since X^τ_n is adapted. The right hand side gives

$$\begin{aligned} E((X_{n+1} - X_n) \, 1_{\{\,\tau\ge n+1\,\}} \,|\, \mathcal{F}_n) &= 1_{\{\,\tau\ge n+1\,\}} \, E((X_{n+1} - X_n) \,|\, \mathcal{F}_n) \\ &\qquad \text{since } 1_{\{\,\tau\ge n+1\,\}} \in \mathcal{F}_n \\ &= 1_{\{\,\tau\ge n+1\,\}} \, (E(X_{n+1} \,|\, \mathcal{F}_n) - X_n) \\ &\begin{cases} = 0, & \text{if } (X_n) \text{ is a martingale} \\ \ge 0, & \text{if } (X_n) \text{ is a submartingale} \\ \le 0, & \text{if } (X_n) \text{ is a supermartingale} \end{cases} \end{aligned}$$

and so it follows that (X^τ_n) is a martingale (respectively, submartingale, supermartingale) if (X_n) is. ■

Definition 3.22. Let $(X_n)_{n\in\mathbb{Z}^+}$ be an adapted process with respect to a given filtration $(\mathcal{F}_n)$ built on a probability space $(\Omega, \mathcal{S}, P)$ and let τ be a stopping time such that $\tau < \infty$ almost surely. The random variable stopped by τ is defined to be

$$X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega), & \text{for } \omega \in \{\,\tau \in \mathbb{Z}^+\,\} \\ X_\infty, & \text{if } \tau \notin \mathbb{Z}^+, \end{cases}$$

where X_∞ is any arbitrary but fixed constant. Then X_τ really is a random variable, that is, it is measurable with respect to $\mathcal{S}$ (in fact, it is measurable with respect to $\sigma\big(\bigcup_n \mathcal{F}_n\big)$). To see this, let B be any Borel set in $\mathbb{R}$. Then (on $\{\,\tau < \infty\,\}$)

$$\begin{aligned} \{\,X_\tau \in B\,\} = \{\,X_\tau \in B\,\} \cap \bigcup_{k\in\mathbb{Z}^+} \{\,\tau = k\,\} &= \bigcup_{k\in\mathbb{Z}^+} (\{\,X_\tau \in B\,\} \cap \{\,\tau = k\,\}) \\ = \bigcup_{k\in\mathbb{Z}^+} \{\,X_k \in B\,\} \in \bigcup_{k\in\mathbb{Z}^+} \mathcal{F}_k & \end{aligned}$$

which shows that X_τ is a bone fide random variable.

DR.RUPNATHJI(DR.RUPAK NATH)

Example 3.23 (Wald's equation). Let $(X_j)_{j\in\mathbb{N}}$ be a sequence of independent identically distributed random variables with finite expectation. For each $n \in \mathbb{N}$, let $S_n = X_1 + \cdots + X_n$. Then clearly $E(S_n) = n\,E(X_1)$.

For $n \in \mathbb{N}$, let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ be the filtration generated by the X_js and suppose that N is a bounded stopping time (with respect to $(\mathcal{F}_n)$).

We calculate

$$\begin{aligned} E(S_N) &= E(S_1 1_{\{N=1\}} + S_2 1_{\{N=2\}} + \ldots) \\ &= E(X_1 1_{\{N=1\}} + (X_1+X_2) 1_{\{N=2\}} + \\ &\qquad\qquad + (X_1+X_2+X_3) 1_{\{N=3\}} + \ldots) \\ &= E(X_1 1_{\{N\ge 1\}} + X_2 1_{\{N\ge 2\}} + X_3 1_{\{N\ge 3\}} + \ldots) \\ &= E(X_1 g_1) + E(X_2 g_2) + \ldots \end{aligned}$$

where $g_j = 1_{\{N\ge j\}}$. But $1_{\{N\ge j\}} = 1 - 1_{\{N<j\}}$ and so (g_j) is predictable and, by independence,

$$E(X_j\, g_j) = E(X_j)\,E(g_j) = E(X_j)\,P(N \ge j) = E(X_1)\,P(N \ge j)$$

since $E(X_j) = E(X_1)$ for all j. Therefore

$$\begin{aligned} E(S_N) &= E(X_1)\,(\,P(N\ge 1) + P(N \ge 2) + P(N\ge 3) + \ldots) \\ &= E(X_1)\,(P(N=1) + 2\,P(N=2) + 3\,P(N=3) + \ldots) \\ &= E(X_1)\,E(N) \end{aligned}$$

— known as Wald's equation.

DR.RUPNATHJI(DR.RUPAK NATH)

Theorem 3.24 (Optional Stopping Theorem). *Let $(X_n)_{n\in\mathbb{Z}^+}$ be a martingale and τ a stopping time such that:*

(1) $\tau < \infty$ *almost surely (i.e., $P(\tau \in \mathbb{Z}^+) = 1$);*

(2) X_τ *is integrable;*

(3) $E(X_n\, 1_{\{\tau > n\}}) \to 0$ *as $n \to \infty$.*

Then $E(X_\tau) = E(X_0)$.

Proof. We have $X_{\tau\wedge n} = \begin{cases} X_\tau & \text{on } \{\tau \le n\} \\ X_n & \text{on } \{\tau > n\} \end{cases}$

and so

$$\begin{aligned} X_t &= X_{\tau\wedge n} + (X_\tau - X_{\tau\wedge n}) \\ &= X_{\tau\wedge n} + (X_\tau - X_{\tau\wedge n})(1_{\tau\le n} + 1_{\tau>n}) \\ &= X_{\tau\wedge n} + (X_\tau - X_n) 1_{\tau>n}\,. \end{aligned}$$

Taking expectations, we get

$$E(X_\tau) = E(X_{\tau\wedge n}) + E(X_\tau 1_{\{\tau>n\}}) - E(X_n 1_{\{\tau>n\}}) \qquad (*)$$

But we know $(X_{\tau\wedge n})$ is a martingale and so $E(X_{\tau\wedge n}) = E(X_{\tau\wedge 0}) = E(X_0)$ since $\tau(\omega) \geq 0$ always.

The middle term in $(*)$ gives (using $1_{\{\tau=\infty\}} = 0$ almost surely)

$$\begin{aligned} E(X_\tau 1_{\{\tau>n\}}) &= \sum_{k=n+1}^{\infty} E(X_k 1_{\{\tau=k\}}) \\ &\to 0 \quad \text{as } n \to \infty \end{aligned}$$

because, by (2),

$$E(X_\tau) = \sum_{k=0}^{\infty} E(X_k 1_{\{\tau=k\}}) < \infty\,.$$

Finally, by hypothesis, we know that $E(X_n 1_{\{\tau>n\}}) \to 0$ and so letting $n \to \infty$ in $(*)$ gives the desired result. ■

DR.RUPNATHJI(DR.RUPAK NATH)

Remark 3.25. If X is a submartingale and the conditions (1), (2) and (3) of the theorem hold, then we have

$$E(X_\tau) \geq E(X_0)\,.$$

This follows just as for the case when X is a martingale except that one now uses the inequality $E(X_{\tau\wedge n}) \geq E(X_{\tau\wedge 0}) = E(X_0)$ in equation $(*)$. In particular, we note these conditions hold if τ is a bounded stopping time.

Remark 3.26. We can recover Wald's equation as a corollary. Indeed, let $X_1, X_2, \dots$ be a sequence of independent identically distributed random variables with means $E(X_j) = \mu$ and let $Y_j = X_j - \mu$. Let $\xi_n = Y_1 + \dots + Y_n$. Then we know that (ξ_n) is a martingale with respect to the filtration $(\mathcal{F}_n)$ where $\mathcal{F}_n = \sigma(X_1, \dots, X_n) = \sigma(Y_1, \dots, Y_n)$. By the Theorem (but with index set now $\mathbb{N}$ rather than $\mathbb{Z}^+$), we see that $E(\xi_N) = E(\xi_1) = E(Y_1) = 0$, where N is a stopping time obeying the conditions of the theorem. If $S_n = X_1 + \dots + X_n$, then we see that $\xi_N = S_N - N\,\mu$ and we conclude that

$$E(S_N) = E(N)\,\mu = E(N)\,E(X_1)$$

which is Wald's equation.

Lemma 3.27. *Let (X_n) be a submartingale with respect to the filtration $(\mathcal{F}_n)$ and suppose that τ is a bounded stopping time with $\tau \le m$ where $m \in \mathbb{Z}$. Then*

$$E(X_m) \ge E(X_\tau)\,.$$

Proof. We have

$$\begin{aligned}
E(X_m) &= \sum_{j=0}^{m} \int_\Omega X_m \, 1_{\{\,\tau=j\,\}} \, dP \\
&= \sum_{j=0}^{m} \int_{\{\,\tau=j\,\}} X_m \, dP \\
&= \sum_{j=0}^{m} \int_{\{\,\tau=j\,\}} E(X_m \,|\, \mathcal{F}_j) \, dP, \quad \text{since } \{\,\tau=j\,\} \in \mathcal{F}_j\,, \\
&\ge \sum_{j=0}^{m} \int_{\{\,\tau=j\,\}} X_j \, dP, \quad \text{since } E(X_m \,|\, \mathcal{F}_j) \ge X_j \text{ almost surely,} \\
&= E(X_\tau)
\end{aligned}$$

as required. ■

Theorem 3.28 (Doob's Maximal Inequality). *Suppose that (X_n) is a non-negative submartingale with respect to the filtration $(\mathcal{F}_n)$. Then, for any $m \in \mathbb{Z}^+$ and $\lambda > 0$,*

$$\lambda \, P(\max_{k \le m} X_k \ge \lambda) \le E(X_m \, 1_{\{\,\max_{k\le m} X_k \ge \lambda\,\}})\,.$$

Proof. Fix $m \in \mathbb{Z}^+$ and let $X_m^* = \max_{k \le m} X_k$.
For $\lambda > 0$, let

$$\tau(\omega) = \begin{cases} \min\{\,k \le m : X_k(\omega) \ge \lambda\,\}, & \text{if } \{\,k : k \le m \text{ and } X_k(\omega) \ge \lambda\,\} \ne \varnothing \\ m, & \text{otherwise.} \end{cases}$$

Then evidently $\tau \le n$ almost surely (and takes values in $\mathbb{Z}^+$). We shall show that τ is a stopping time. Indeed, for $j < m$

$$\{\,\tau = j\,\} = \{\,X_0 < \lambda\,\} \cap \{\,X_1 < \lambda\,\} \cap \cdots \cap \{\,X_{j-1} < \lambda\,\} \cap \{\,X_j \ge \lambda\,\} \in \mathcal{F}_j$$

and

$$\{\,\tau = m\,\} = \{\,X_0 < \lambda\,\} \cap \{\,X_1 < \lambda\,\} \cap \cdots \cap \{\,X_{m-1} < \lambda\,\} \in \mathcal{F}_m$$

and so we see that τ is a stopping time as claimed.

DR.RUPNATH(DR.RUPAK NATH)

Next, we note that by the Lemma (since $\tau \le m$)

$$E(X_m) \ge E(X_\tau)\,.$$

Now, set $A = \{\, X_m^* \ge \lambda \,\}$. Then

$$E(X_\tau) = \int_A X_\tau \, dP + \int_{A^c} X_\tau \, dP\,.$$

If $X_m^* \ge \lambda$, then there is some $0 \le k \le m$ with $X_k \ge \lambda$ and $\tau = k_0 \le k$ in this case. (τ is the minimum of such ks so $X_{k_0} \ge \lambda$.) Therefore

$$X_\tau = X_{k_0} \ge \lambda\,.$$

On the other hand, if $X_m^* < \lambda$, then there is no j with $0 \le j \le m$ and $X_j \ge \lambda$. Thus $\tau = m$, by construction. Hence

$$\begin{aligned}\int_\Omega X_m \, dP = E(X_m) \ge E(X_\tau) &= \int_A X_\tau \, dP + \int_{A^c} X_\tau \, dP \\ &\ge \lambda\, P(A) + \int_{A^c} X_m \, dP\,.\end{aligned}$$

Rearranging, we find that

$$\lambda\, P(A) \le \int_A X_m \, dP$$

and the proof is complete. ■

Corollary 3.29. *Let (X_n) be an $\mathcal{L}^2$-martingale. Then*

$$\lambda^2\, P(\max_{k\le m} |X_k| \ge \lambda) \le E(X_m^2)$$

for any $\lambda \ge 0$.

Proof. Since (X_n) is an $\mathcal{L}^2$-martingale, it follows that the process (X_n^2) is a submartingale (Proposition 3.10). Applying Doob's Maximal Inequality to the submartingale (X_n^2) (and with λ^2 rather than λ), we get

$$\begin{aligned}\lambda^2\, P(\max_{k\le m} X_k^2 \ge \lambda^2) &\le \int_{\{\max_{k\le m} X_k^2 \ge \lambda^2\,\}} X_m^2 \, dP \\ &\le \int_\Omega X_m^2 \, dP\end{aligned}$$

that is,

$$\lambda^2\, P(\max_{k\le m} |X_k| \ge \lambda) \le E(X_m^2)$$

as required. ■

DR.RUPNATHJI(DR.RUPAK NATH)

Proposition 3.30. *Suppose that $X \geq 0$ and X^2 is integrable. Then*

$$E(X^2) = 2\int_0^\infty t\,P(X \geq t)\,dt\,.$$

Proof. For $x \geq 0$, we can write

$$\begin{aligned} x^2 &= 2\int_0^x t\,dt \\ &= 2\int_0^\infty t\,1_{\{x \geq t\}}\,dt\,. \end{aligned}$$

Setting $x = X(\omega)$, we get

$$X(\omega)^2 = 2\int_0^\infty t\,1_{X(\omega) \geq t}\,dt$$

so that

$$\begin{aligned} E(X^2) &= 2\int_0^\infty t\,E(1_{X \geq t})\,dt \\ &= 2\int_0^\infty t\,P(X \geq t)\,dt\,. \end{aligned}$$

■

Theorem 3.31 (Doob's Maximal $\mathcal{L}^2$-inequality). *Let (X_n) be a non-negative $\mathcal{L}^2$-submartingale. Then*

$$E(\,(X_n^*)^2\,) \leq 4\,E(\,X_n^2\,)$$

where $X_n^* = \max_{k \leq n} X_k$. In alternative notation,

$$\|\,X_n^*\,\|_2 \leq 2\,\|X_n\|_2.$$

Proof. Using the proposition, we find that

$$\begin{aligned} \|\,X_n^*\,\|_2^2 &= E(\,(X_n^*)^2\,) \\ &= 2\int_0^\infty t\,P(X_n\,1_{\{\,X_n^* \geq t\,\}})\,dt \\ &\leq 2\int_0^\infty E(X_n\,1_{\{\,X_n^* \geq t\,\}})\,dt\,, \quad \text{by the Maximal Inequality,} \\ &= 2\int_0^\infty \Big(\int_{\{\,X_n^* \geq t\,\}} X_n\,dP\Big)\,dt \\ &= 2\int_\Omega X_n \Big(\int_0^{X_n^*} dt\Big)\,dP\,, \quad \text{by Tonelli's Theorem,} \\ &= 2\int_\Omega X_n\,X_n^*\,dP \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

$$= 2\,E(X_n\,X_n^*)$$
$$\le 2\,\|X_n\|_2\,\|X_n^*\|_2\,, \quad \text{by Schwarz' inequality.}$$

It follows that

$$\|\,X_n^*\,\|_2 \le 2\,\|\,X_n\,\|_2$$

or

$$E(\,(X_n^*)^2\,) \le 4\,E(\,X_n^2\,)$$

and the proof is complete. ∎

Doob's Up-crossing inequality

We wish to discuss convergence properties of martingales and an important first result in this connection is Doob's so-called up-crossing inequality. By way of motivation, consider a sequence (x_n) of real numbers. If $x_n \to \alpha$ as $n \to \infty$, then for any $\varepsilon > 0$, x_n is eventually inside the interval $(\alpha-\varepsilon, \alpha+\varepsilon)$. In particular, for any $a < b$, there can be only a finite number of pairs $(n, n+k)$ for which $x_n < a$ and $x_{n+k} > b$. In other words, the graph of points (n, x_n) can only cross the semi-infinite strip $\{\,(x,y) : x > 0,\ a \le y \le b\,\}$ finitely-many times. We look at such crossings for processes (X_n).

Consider a process (X_n) with respect to the filtration $(\mathcal{F}_n)$. Fix $a < b$. Now fix $\omega \in \Omega$ and consider the sequence $X_0(\omega), X_1(\omega), X_2(\omega), \dots$ (although the value of $X_0(\omega)$ will play no rôle here). We wish to count the number of times the path $(X_n(\omega))$ crosses the band $\{\,(x,y) : a \le y \le b\,\}$ from below a to above b. Such a crossing is called an up-crossing of $[a,b]$ by $(X_n(\omega))$.

DR.RUPNATHJI(DR.RUPAK NATH)

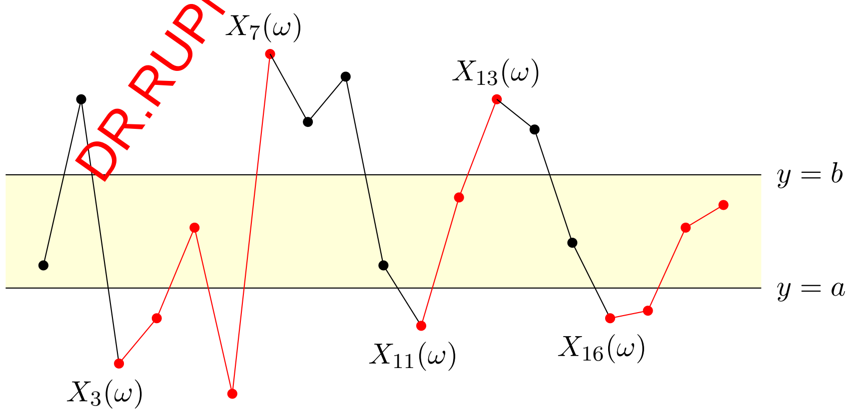

Figure 3.1: Up-crossings of $[a,b]$ by $(X_n(\omega))$.

In the example in the figure 3.1, the path sequences $X_3(\omega), \dots, X_7(\omega)$ and $X_{11}(\omega), X_{12}(\omega), X_{13}(\omega)$ each constitute an up-crossing. The path sequence

$X_{16}(\omega), X_{17}(\omega), X_{18}(\omega), X_{19}(\omega)$ forms a partial up-crossing (which, in fact, will never be completed if $X_n(\omega) \le b$ remains valid for all $n > 16$).

As an aid to counting such up-crossings, we introduce the process $(g_n)_{n\in\mathbb{N}}$ defined as follows:

$$g_1(\omega) = 0$$

$$g_2(\omega) = \begin{cases} 1, & \text{if } X_1(\omega) < a \\ 0, & \text{otherwise} \end{cases}$$

$$g_3(\omega) = \begin{cases} 1, & \text{if } g_2(\omega) = 0 \text{ and } X_2(\omega) < a \\ 1, & \text{if } g_2(\omega) = 1 \text{ and } X_2(\omega) \le b \\ 0, & \text{otherwise} \end{cases}$$

$$\vdots$$

$$g_{n+1}(\omega) = \begin{cases} 1, & \text{if } g_n(\omega) = 0 \text{ and } X_n(\omega) < a \\ 1, & \text{if } g_n(\omega) = 1 \text{ and } X_n(\omega) \le b \\ 0, & \text{otherwise.} \end{cases}$$

We see that $g_n = 1$ when an up-crossing is in progress. Fix m and consider the sum (up to time m) (— the transform of X by g at time m)

$$\sum_{j=1}^{m} g_j(\omega)\,(X_j(\omega) - X_{j-1}(\omega)\,)$$
$$= g_1(\omega)\,(X_1(\omega) - X_0(\omega)\,) + \cdots + g_m(\omega)\,(X_m(\omega) - X_{m-1}(\omega)\,)\,. \quad (*)$$

The g_j s take the values 1 or 0 depending on whether an up-crossing is in progress or not. Indeed, if

$$g_r(\omega) = 0, \quad g_{r+1}(\omega) = \cdots = g_{r+s}(\omega) = 1, \quad g_{r+s+1}(\omega) = 0,$$

then the path $X_r(\omega), \ldots, X_{r+s}(\omega)$ forms an up-crossing of $[a,b]$ and we see that

$$\begin{aligned}\sum_{j=r+1}^{r+s} g_j(\omega)(X_j(\omega) - X_{j-1}(\omega)) &= \sum_{j=r+1}^{r+s} (X_j(\omega) - X_{j-1}(\omega)\,) \\ &= X_{r+s}(\omega) - X_r(\omega) \\ &> b - a\end{aligned}$$

because $X_r < a$ and $X_{r+s} > b$.

Let $U_m[a,b](\omega)$ denote the number of completed up-crossings of $(X_n(\omega))$ up to time m. If $g_j(\omega) = 0$ for each $1 \le j \le m$, then the sum $(*)$ is zero. If not all g_j s are zero, we can say that the path has made $U_m[a,b](\omega)$ up-crossings

DR.RUPNATHJI(DR.RUPAK NATH)

(which may be zero) followed possibly by a partial up-crossing (which will be the case if and only if $g_m(\omega) = g_{m+1}(\omega) = 1$). Hence we may estimate $(*)$ by

$$\sum_{j=1}^{m} g_j(\omega)\,(X_j(\omega) - X_{j-1}(\omega)\,) \geq U_m[a,b](\omega)(b-a) + R$$

where $R = 0$ if there is no residual partial up-crossing but otherwise is given by

$$R = \sum_{j=k}^{m} g_j(\omega)\,(X_j(\omega) - X_{j-1}(\omega)\,)$$

where k is the largest integer for which $g_j(\omega) = 1$ for all $k \leq j \leq m+1$ and $g_{k-1} = 0$. This means that $X_{k-1}(\omega) < a$ and $X_j(\omega) \leq b$ for $k \leq j \leq m$. The sequence $X_{k-1}(\omega) < a, X_k(\omega) \leq b, \dots, X_m(\omega) \leq b$ forms the partial up-crossing at the end of the path $X_0(\omega), X_1(\omega), \dots, X_m(\omega)$. Since we have $g_k(\omega) = \cdots = g_m(\omega) = 1$, we see that

$$R = X_m(\omega) - X_{k-1}(\omega) > X_m(\omega) - a$$

where the inequality follows because $g_k(\omega) = 1$ and so $X_{k-1}(\omega) < a$, by construction. Now, any real-valued function f can be written as $f = f^+ - f^-$ where $f^\pm$ are the positive and negative parts of f, defined by $f^\pm = \frac{1}{2}(|f| \pm f)$. Evidently $f^\pm \geq 0$. The inequality $f \geq -f^-$ allows us to estimate R by

$$R \geq -(X_m - a)^-(\omega)$$

which is valid also when there is no partial up-crossing (so $R = 0$).

Putting all this together, we obtain the estimate

$$\sum_{j=1}^{m} g_j\,(X_j - X_{j-1}\,) \geq U_m[a,b](b-a) - (X_m - a)^- \qquad (**)$$

on Ω.

Proposition 3.32. *The process (g_n) is predictable.*

Proof. It is clear that g_1 is $\mathcal{F}_0$-measurable. Furthermore, $g_2 = 1_{\{\,X_1 < a\,\}}$ and so g_2 is $\mathcal{F}_1$-measurable. For the general case, let us suppose that g_m is $\mathcal{F}_{m-1}$-measurable. Then

$$g_{m+1} = 1_{\{\,g_m = 0\,\} \cap \{\,X_m < a\,\}} + 1_{\{\,g_m = 1\,\} \cap \{\,X_m \leq b\,\}}$$

which is $\mathcal{F}_m$-measurable. By induction, it follows that $(g_n)_{n \in \mathbb{N}}$ is predictable, as required. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

Theorem 3.33. *For any supermartingale (X_n), we have*

$$(b-a)\, E(\, U_m[a,b]\,) \le E((X_m - a)^-)\,.$$

Proof. The non-negative, bounded process (g_n) is predictable and so the martingale transform $((g\cdot X)_n)$ is a supermartingale. The estimate $(**)$ becomes

$$(g\cdot X)_m \ge (b-a)\, U_m[a,b] - (X_m - a)^-\,.$$

Taking expectations, we obtain the inequality

$$E(\,(g\cdot X)_m\,) \ge (b-a)\, E(\, U_m[a,b]\,) - E((X_m - a)^-\,)\,.$$

But $E(\,(g\cdot X)_m\,) \le E(\, g\cdot X)_1\,)$ since $((g\cdot X)_n)$ is a supermartingale, and $(g\cdot X)_1 = 0$, by construction, so we may say that

$$0 \ge (b-a)\, E(\, U_m[a,b]\,) - E((X_m - a)^-)$$

and the result follows. ∎

Lemma 3.34. *Suppose that (f_n) is a sequence of random variables such that $f_n \ge 0$, $f_n \uparrow$ and such that $E(f_n) \le K$ for all n. Then*

$$P(\lim_n f_n < \infty) = 1\,.$$

Proof. Set $g_n = f_n \wedge m$. Then $g_n \uparrow$ and $0 \le g_n \le m$ for all n. It follows that $g = \lim_n g_n$ exists and obeys $0 \le g \le m$. Let $B = \{\, \omega : f_n(\omega) \to \infty\,\}$. Then $g = m$ on B (because $f_n(\omega)$ is eventually $> m$ if $\omega \in B$). Now

$$0 \le g_n \le f_n \implies E(g_n) \le E(f_n) \le K\,.$$

By Lebesgue's Monotone Convergence Theorem, $E(g_n) \uparrow E(g)$ and therefore $E(g) \le K$. Hence

$$m\, P(B) \le E(g) \le K\,.$$

This holds for any m and so it follows that $P(B) = 0$. ∎

Theorem 3.35 (Doob's Martingale Convergence Theorem). *Let (X_n) be a supermartingale such that $E(X_n) < M$ for all $n \in \mathbb{Z}^+$. Then there is an integrable random variable X such that $X_n \to X$ almost surely as $n \to \infty$.*

Proof. We have seen that

$$E(\, U_n[a,b]\,) \le \frac{E(\,(X_n - a)^-\,)}{b-a}$$

for any $a < b$. However,

$$\begin{aligned}(X_n - a)^- &= \tfrac{1}{2}\,(|X_n - a| - (X_n - a))\\ &\le |X_n| + |a|\end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

and so, by hypothesis,

$$\begin{aligned} E(\,(X_n - a)^-\,) &\le E|X_n| + |a| \\ &\le M + |a| \end{aligned}$$

giving

$$E(\,U_n[a,b]\,) \le \frac{M + |a|}{b - a}$$

for any $n \in \mathbb{Z}^+$. However, by its very construction, $U_n[a,b] \le U_{n+1}[a,b]$ and so $\lim_n E(\,U_n[a,b]\,)$ exists and obeys $\lim_n E(\,U_n[a,b]\,) \le \dfrac{M+|a|}{b-a}$. By the lemma, it follows that $U_n[a,b]$ converges almost surely (to a finite value).

Let $A = \bigcap_{\substack{a<b \\ a,b \in \mathbb{Q}}} \{\, \lim_n U_n[a,b] < \infty \,\}$. Then A is a countable intersection of sets of probability 1 and so $P(A) = 1$.

Claim: (X_n) converges almost surely.
For, if not, then

$$B = \{\, \liminf X_n < \limsup X_n \,\} \subset \Omega$$

would have positive probability. For any $\omega \in B$, there exist $a, b \in \mathbb{Q}$ with $a < b$ such that

$$\liminf X_n(\omega) < a < b < \limsup X_n(\omega)$$

which means that $\lim_n U_n[a,b](\omega) = \infty$ (because $(X_n(\omega))$ would cross $[a,b]$ infinitely-many times). Hence $B \cap A = \varnothing$ which means that $B \subset A^c$ and so $P(B) \le P(B^c) = 0$. It follows that X_n converges almost surely as claimed.

Denote this limit by X, with $X(\omega) = 0$ for $\omega \notin A$. Then

$$\begin{aligned} E(\,|X|\,) &= E(\,\liminf_n |X_n|\,) \\ &\le \liminf_n E(|X_n|\,, \quad \text{by Fatou's Lemma,} \\ &\le M \end{aligned}$$

and the proof is complete. ∎

Corollary 3.36. *Let (X_n) be a positive supermartingale. Then there is some $X \in \mathcal{L}_1$ such that $X_n \to X$ almost surely.*

Proof. Since $X_n \ge 0$ almost surely (and is a supermartingale), it follows that

$$0 \le E(X_n) \le E(X_0)$$

and so $(E(X_n))$ is bounded. Now apply the theorem. ∎

Remark 3.37. These results also hold for martingales and submartingales (because $(-X_n)$ is a supermartingale whenever (X_n) is a submartingale).

DR.RUPNATHJI(DR.RUPAK NATH)

Remark 3.38. The supermartingale (X_n) is $\mathcal{L}^1$-bounded if and only if the process (X_n^-) is, that is

$$\sup_n E(|X_n|) < \infty \iff \sup_n E(X_n^-) < \infty \,.$$

We can see this as follows. Since $E(X_n^-) \le E(X_n^+ + X_n^-) = E(|X_n|)$, we see immediately that if $E(|X_n|) \le M$ for all n, then also $E(X_n^-) \le M$ for all n. Conversely, suppose there is some positive constant M such that $E(X_n^-) \le M$ for all n. Then

$$E(|X_n|) = E(X_n^+ + X_n^-) = E(X_n) + 2\,E(X_n^-) \le E(X_0) + 2M$$

and so (X_n) is $\mathcal{L}^1$-bounded. ($E(X_n) \le E(X_0)$ because (X_n) is a supermartingale.)

Remark 3.39. Note that the theorem gives convergence almost surely. In general, almost sure convergence does not necessarily imply $\mathcal{L}^1$ convergence. For example, suppose that Y has a uniform distribution on $(0,1)$ and for each $n \in \mathbb{N}$ let $X_n = n 1_{\{ Y < 1/n \}}$. Then $X_n \to 0$ almost surely but $\|X_n\|_1 = 1$ for all n and so it is false that $X_n \to 0$ in $\mathcal{L}^1$. However, under an extra condition, $\mathcal{L}^1$ convergence is assured.

Definition 3.40. The collection $\{ Y_\alpha : \alpha \in I \}$ of integrable random variables labeled by some index set I is said to be uniformly integrable if for any given $\varepsilon > 0$ there is some $M > 0$ such that

$$E(\,|Y_\alpha| 1_{\{ |Y_\alpha| > M \}}\,) = \int_{\{ |Y_\alpha| > M \}} |Y_\alpha| \, dP < \varepsilon$$

for all $\alpha \in I$.

Remark 3.41. Note that any uniformly integrable family $\{ Y_\alpha : \alpha \in I \}$ is $\mathcal{L}^1$-bounded. Indeed, by definition, for any $\varepsilon > 0$ there is M such that $\int_{\{ |Y_\alpha| > M \}} |Y_\alpha| \, dP < \varepsilon$ for all α. But then

$$\int_\Omega |Y_\alpha| \, dP = \int_{\{ |Y_\alpha| \le M \}} |Y_\alpha| \, dP + \int_{\{ |Y_\alpha| > M \}} |Y_\alpha| \, dP \le M + \varepsilon$$

for all $\alpha \in I$.

Before proceeding, we shall establish the following result.

Lemma 3.42. *Let $X \in \mathcal{L}^1$. Then for any $\varepsilon > 0$ there is $\delta > 0$ such that if A is any event with $P(A) < \delta$ then $\int_A |X| \, dP < \varepsilon$.*

Proof. Set $\xi = |X|$ and for $n \in \mathbb{N}$ let $\xi_n = \xi \, 1_{\{ \xi \le n \}}$. Then $\xi_n \to \xi$ almost surely and by Lebesgue's Monotone Convergence Theorem

$$\int_{\{ \xi > n \}} \xi \, dP = \int_\Omega \xi \, (1 - 1_{\{ \xi \le n \}}) \, dP = \int_\Omega \xi \, dP - \int_\Omega \xi_n \, dP \to 0$$

DR. RUPNATHJI(DR. RUPAK NATH)

as $n \to \infty$. Let $\varepsilon > 0$ be given and let n be so large that $\int_{\{\xi > n\}} \xi \, dP < \frac{1}{2}\varepsilon$. For any event A, we have

$$\begin{aligned}\int_A \xi \, dP &= \int_{A \cap \{\xi \le n\}} \xi \, dP + \int_{A \cap \{\xi > n\}} \xi \, dP \\ &\le \int_A n \, dP + \int_{\{\xi > n\}} \xi \, dP \\ &< n\, P(A) + \tfrac{1}{2}\varepsilon \\ &< \varepsilon\end{aligned}$$

whenever $P(A) < \delta$ where $\delta = \varepsilon / 2n$. ∎

Theorem 3.43 ($\mathcal{L}^1$-convergence Theorem). *Suppose that (X_n) is a uniformly integrable supermartingale. Then there is an integrable random variable X such that $X_n \to X$ in $\mathcal{L}^1$.*

Proof. By hypothesis, the family $\{ X_n : n \in \mathbb{Z}^+ \}$ is uniformly integrable and we know that X_n converges almost surely to some $X \in \mathcal{L}^1$. We shall show that these two facts imply that $X_n \to X$ in $\mathcal{L}^1$. Let $\varepsilon > 0$ be given. We estimate

$$\begin{aligned}\|X_n - X\|_1 &= \int_\Omega |X_n - X| \, dP \\ &= \int_{\{|X_n| \le M\}} |X_n - X| \, dP + \int_{\{|X_n| > M\}} |X_n - X| \, dP \\ &\le \int_{\{|X_n| \le M\}} |X_n - X| \, dP \\ &\qquad + \int_{\{|X_n| > M\}} |X_n| \, dP + \int_{\{|X_n| > M\}} |X| \, dP \,.\end{aligned}$$

We shall consider separately the three terms on the right hand side.

By the hypothesis of uniform integrability, we may say that the second term $\int_{\{|X_n| > M\}} |X_n| \, dP < \varepsilon$ for all sufficiently large M.

As for the third term, the integrability of X means that there is $\delta > 0$ such that $\int_A |X| \, dP < \varepsilon$ whenever $P(A) < \delta$. Now we note that (provided $M > 1$)

$$P(\, |X_n| > M \,) = \int_\Omega 1_{\{|X_n| > M\}} \, dP \le \int_\Omega 1_{\{|X_n| > M\}} |X_n| \, dP < \delta$$

for large enough M, by uniform integrability. Hence, for all sufficiently large M,

$$\int_{\{|X_n| > M\}} |X| \, dP < \varepsilon \,.$$

DR.RUPNATHJI(DR.RUPAK NATH)

Finally, we consider the first term. Fix $M > 0$ so that the above bounds on the second and third terms hold. The random variable $1_{\{\,|X_n| \le M\,\}} |X_n - X|$ is bounded by the integrable random variable $M + |X|$ and so it follows from Lebesgue's Dominated Convergence Theorem that

$$\int_{\{\,|X_n| \le M\,\}} |X_n - X| \, dP = \int_\Omega 1_{\{\,|X_n| \le M\,\}} |X_n - X| \, dP \to 0$$

as $n \to \infty$ and the result follows. ■

Proposition 3.44. *Let (X_n) be a martingale with respect to the filtration $(\mathcal{F}_n)$ such that $X_n \to X$ in $\mathcal{L}^1$. Then, for each n, $X_n = E(X \,|\, \mathcal{F}_n)$ almost surely.*

Proof. The conditional expectation is a contraction on $\mathcal{L}^1$, that is,

$$\| E(X \,|\, \mathcal{F}_m) \|_1 \le \| X \|_1$$

for any $X \in \mathcal{F}^1$. Therefore, for fixed m,

$$\| E(X - X_n \,|\, \mathcal{F}_m) \|_1 \le \| X - X_n \|_1 \to 0$$

as $n \to \infty$. But the martingale property implies that for any $n \ge m$

$$E(X - X_n \,|\, \mathcal{F}_m) = E(X \,|\, \mathcal{F}_m) - E(X_n \,|\, \mathcal{F}_m) = E(X \,|\, \mathcal{F}_m) - X_m$$

which does not depend on n. It follows that $\| E(X \,|\, \mathcal{F}_m) - X_m \|_1 = 0$ and so $E(X \,|\, \mathcal{F}_m) = X_m$ almost surely. ■

Theorem 3.45 ($\mathcal{L}^2$-martingale Convergence Theorem). *Suppose that (X_n) is an $\mathcal{L}^1$-martingale such that $E(X_n^2) < K$ for all n. Then there is $X \in \mathcal{L}^2$ such that*

(1) $X_n \to X$ *almost surely,*

(2) $E(\,(X - X_n)^2\,) \to 0$.

In other words, $X_n \to X$ almost surely and also in $\mathcal{L}^2$.

Proof. Since $\|X_n\|_1 \le \|X_n\|_2$, it follows that (X_n) is a bounded $\mathcal{L}^1$-martingale and so there is some $X \in \mathcal{L}^1$ such that $X_n \to X$ almost surely. We must show that $X \in \mathcal{L}^2$ and that $X_n \to X$ in $\mathcal{L}^2$. We have

$$\begin{aligned} \|X_n - X_0\|_2^2 &= E(\,(X_n - X_0)^2\,) \\ &= E\big(\sum_{k=1}^n (X_k - X_{k-1}) \sum_{j=1}^n (X_j - X_{j-1}) \big) \\ &= \sum_{k=1}^n E\big(\,(X_k - X_{k-1})^2\,\big)\,, \text{ by orthogonality of increments.} \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

However, by the martingale property,

$$E(\,(X_n - X_0)^2\,) = E(X_n^2) - E(X_0^2) < K$$

and so $\sum_{k=1}^n E\big((X_k - X_{k-1})^2\big)$ increases with n but is bounded, and so must converge.

Let $\varepsilon > 0$ be given. Then (as above)

$$E(\,(X_n - X_m)^2\,) = \sum_{k=m+1}^{n} E\big((X_k - X_{k-1})^2\big) < \varepsilon$$

for all sufficiently large m, n. Letting $n \to \infty$ and applying Fatou's Lemma, we get

$$\begin{aligned} E(\,(X - X_m)^2\,) &= E(\,\liminf_n (X_n - X_m)^2\,) \\ &\le \liminf_n E(\,(X_n - X_m)^2\,) \\ &\le \varepsilon \end{aligned}$$

for all sufficiently large m. Hence $X - X_n \in \mathcal{L}^2$ and so $X \in \mathcal{L}^2$ and $X_n \to X$ in $\mathcal{L}^2$, as required. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

Doob-Meyer Decomposition

The (continuous time formulation of the) following decomposition is a crucial idea in the abstract development of stochastic integration.

Theorem 3.46. *Suppose that (X_n) is an adapted $\mathcal{L}^1$ process. Then (X_n) has the decomposition*

$$X = X_0 + M + A$$

where (M_n) is a martingale, null at 0, and (A_n) is a predictable process, null at 0. Such a decomposition is unique, that is, if also $X = X_0 + M' + A'$, then $M = M'$ almost surely and $A = A'$ almost surely.

Furthermore, if X is a supermartingale, then A is increasing.

Proof. We define the process (A_n) by

$$\begin{aligned} &A_0 = 0 \\ &A_n = A_{n-1} + E(X_n - X_{n-1} \,|\, \mathcal{F}_{n-1})\,. \end{aligned}$$

Evidently, A is null at 0 and A_n is $\mathcal{F}_{n-1}$-measurable, so (A_n) is predictable. Furthermore, we see (by induction) that A_n is integrable.

Now, from the construction of A, we find that

$$\begin{aligned} E((X_n - A_n) \mid \mathcal{F}_{n-1}) &= E(X_n \mid \mathcal{F}_{n-1}) \\ &\qquad - E(A_{n-1} \mid \mathcal{F}_{n-1}) - E((X_n - X_{n-1}) \mid \mathcal{F}_{n-1}) \\ &= X_{n-1} - A_{n-1} \quad \text{a.s.} \end{aligned}$$

which means that $(X_n - A_n)$ is an $\mathcal{L}^1$-martingale. Let $M_n = (X_n - A_n) - X_0$. Then M is a martingale, M is null at 0 and we have the Doob decomposition

$$X = X_0 + M + A\,.$$

To establish uniqueness, suppose that

$$X = X_0 + M' + A' = X_0 + M + A\,.$$

Then

$$\begin{aligned} E(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) &= E((M_n + A_n) - (M_{n-1} + A_{n-1}) \mid \mathcal{F}_{n-1}) \\ &= A_n - A_{n-1} \quad \text{a.s.} \end{aligned}$$

since M is a martingale and A is predictable. Similarly,

$$E(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) = A'_n - A'_{n-1} \quad \text{a.s.}$$

and so

$$A_n - A_{n-1} = A'_n - A'_{n-1} \quad \text{a.s.}$$

Now, both A and A' are null at 0 and so $A_0 = A'_0$ $(= 0)$ almost surely and therefore

$$A_1 = A'_1 + \underbrace{(A_0 - A'_0)}_{=0\text{a.s.}} \implies A_1 = A'_1 \quad \text{a.s.}$$

Continuing in this way, we see that $A_n = A'_n$ a.s. for each n. It follows that there is some $\Lambda \subset \Omega$ with $P(\Lambda) = 1$ and such that $A_n(\omega) = A'_n(\omega)$ for all n for all $\omega \in \Lambda$. However, $M = X - X_0 - A$ and $M' = X - X_0 - A'$ and so $M_n(\omega) = M'_n(\omega)$ for all n for all $\omega \in \Lambda$, that is $M = M'$ almost surely.

Now suppose that $X = X_0 + M + A$ is a submartingale. Then $A = X - M - X_0$ is also a submartingale so, since A is predictable,

$$A_n = E(A_n \mid \mathcal{F}_{n-1}) \geq A_{n-1} \quad \text{a.s.}$$

Once again, it follows that there is some $\Lambda \subset \Omega$ with $P(\Lambda) = 1$ and such that

$$A_n(\omega) \geq A_{n-1}(\omega)$$

for all n and all $\omega \in \Lambda$, that is, A is increasing almost surely. ■

DR.RUPNATHJI(DR.RUPAK NATH)

Remark 3.47. Note that if A is any predictable, increasing, integrable process, then $E(A_n \,|\, \mathcal{F}_{n-1}) = A_n \ge A_{n-1}$ so A is a submartingale. It follows that if $X = X_0 + M + A$ is the Doob decomposition of X and if the predictable part A is increasing, then X must be a submartingale (because M is a martingale).

Corollary 3.48. *Let X be an $\mathcal{L}^2$-martingale. Then X^2 has decomposition*

$$X^2 = X_0^2 + M + A$$

where M is an $\mathcal{L}^1$-martingale, null at 0 and A is an increasing, predictable process, null at 0.

Proof. We simply note that X^2 is a submartingale and apply the theorem. ∎

Definition 3.49. The almost surely increasing (and null at 0) process A is called the quadratic variation of the $\mathcal{L}^2$-process X.

The following result is an application of a Monotone Class argument.

Proposition 3.50 (Lévy's Upward Theorem). *Let $(\mathcal{F}_n)_{n\in\mathbb{Z}^+}$ be a filtration on a probability space $(\Omega, \mathcal{S}, P)$. Set $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$ and let $X \in \mathcal{L}^2$ be $\mathcal{F}_\infty$-measurable. For $n \in \mathbb{Z}^+$, let $X_n = E(X \,|\, \mathcal{F}_n)$. Then $X_n \to X$ almost surely.*

Proof. The family (X_n) is an $\mathcal{L}^2$-martingale with respect to the filtration $(\mathcal{F}_n)$ and so the Martingale Convergence Theorem (applied to the filtered probability space $(\Omega, \mathcal{F}_\infty, P, (\mathcal{F}_n))$) tells us that there is some $Y \in \mathcal{L}^2(\mathcal{F}_\infty)$ such that $X_n \to Y$ almost surely, and also $X_n \to Y$ in $\mathcal{L}^2(\mathcal{F}_\infty)$. We shall show that $Y = X$ almost surely.

For any $n \in \mathbb{Z}^+$ and any $B \in \mathcal{F}_n$, we have

$$\int_B X_n \, dP = \int_B E(X \,|\, \mathcal{F}_n) \, dP = \int_B X \, dP.$$

On the other hand,

$$\Big| \int_B X_n \, dP - \int_B Y \, dP \Big| = \Big| \int_\Omega 1_B (X_n - Y) \, dP \Big| \le \| X_n - Y \|_2 .$$

by the Cauchy-Schwarz inequality. Since $\|X_n - Y\|_2 \to 0$, we must have the equality $\int_B (X - Y) \, dP = 0$ for any $B \in \bigcup_n \mathcal{F}_n$.

DR.RUPNATHJI(DR.RUPAK NATH)

Let $\mathcal{M}$ denote the subset of $\mathcal{F}_\infty$ given

$$\mathcal{M} = \{\, B \in \mathcal{F}_\infty : \int_B (X - Y)\, dP = 0 \,\}.$$

We claim that $\mathcal{M}$ is a monotone class. Indeed, suppose that $B_1 \subset B_2 \subset \dots$ is an increasing sequence in $\mathcal{M}$. Then $1_{B_n} \uparrow 1_B$, where $B = \bigcup_n B_n$. The function $X - Y$ is integrable and so we may appeal to Lebesgue's Dominated Convergence Theorem to deduce that

$$\begin{aligned} 0 = \int_{B_n} (X - Y)\, dP &= \int_\Omega 1_{B_n} (X - Y)\, dP \\ \to \int_\Omega 1_B (X - Y)\, dP &= \int_B (X - Y)\, dP \end{aligned}$$

which proves the claim. Now, $\bigcup_n \mathcal{F}_n$ is an algebra and belongs to $\mathcal{M}$. By the Monotone Class Theorem, $\mathcal{M}$ contains the σ-algebra generated by this algebra, namely $\mathcal{F}_\infty$. But then we are done, $\int_B (X - Y)\, dP$ for every $B \in \mathcal{F}_\infty$ and so we must have $X = Y$ almost surely. ∎

Remark 3.51. This result is also valid in $\mathcal{L}^1$. In fact, if $X \in \mathcal{L}^1(\mathcal{F}_\infty)$ then $(E(X \,|\, \mathcal{F}_n))$ is uniformly integrable and the Martingale Convergence Theorem applies. X_n converges almost surely and also in $\mathcal{L}^1$ to some Y. An analogous proof shows that $X = Y$ almost surely.

In the $\mathcal{L}^2$ case, one would expect the result to be true on the following grounds. The subspaces $\mathcal{L}^2(\mathcal{F}_n)$ increase and would seem to "exhaust" the space $\mathcal{L}^2(\mathcal{F}_\infty)$. The projections P_n from $\mathcal{L}^2(\mathcal{F}_\infty)$ onto $\mathcal{L}^2(\mathcal{F}_n)$ should therefore increase to the identity operator, $P_n g \to g$ for any $g \in \mathcal{L}^2(\mathcal{F}_\infty)$. But $X_n = P_n X$, so we would expect $X_n \to X$. Of course, to work this argument through would bring us back to the start.

DR.RUPNATH(DR.RUPAK NATH)

DR.RUPNATHJI(DR.RUPAK NATH)

Chapter 4

Stochastic integration - informally

We suppose that we are given a probability space $(\Omega, \mathcal{S}, P)$ together with a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ of sub-σ-algebras indexed by $\mathbb{R}^+ = [0, \infty)$ (so $\mathcal{F}_s \subset \mathcal{F}_t$ if $s < t$). (Recall that our notation $A \subset B$ means that $A \cap B^c = \varnothing$ so that $\mathcal{F}_s = \mathcal{F}_t$ is permissible here.) Just as for a discrete index, one can define martingales etc.

Definition 4.1. The process $(X_t)_{t\in\mathbb{R}^+}$ is adapted with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ if X_t is $\mathcal{F}_t$-measurable for each $t \in \mathbb{R}^+$.

The adapted process $(X_t)_{t\in\mathbb{R}^+}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ if X_t is integrable for each $t \in \mathbb{R}^+$ and if

$$E(X_t \,|\, \mathcal{F}_s) = X_s \text{ almost surely for any } s \le t.$$

Supermartingales indexed by $\mathbb{R}^+$ are defined as above but with the inequality $E(X_t \,|\, \mathcal{F}_s) \le X_s$ almost surely for any $s \le t$ instead of equality above. A process is a submartingale if its negative is a supermartingale.

Remark 4.2. If (X_t) is an L^2-martingale, then (X_t^2) is a submartingale. The proof of this is just as for a discrete index, as in Proposition 3.10.

Remark 4.3. It must be stressed that although it may at first appear quite innocuous, the change from a discrete index to a continuous one is anything but. There are enormous technical complications involved in the theory with a continuous index. Indeed, one might immediately anticipate measure-theoretic difficulties simply because $\mathbb{R}^+$ is not countable.

Remark 4.4. Stochastic processes (X_n) and (Y_n), indexed by $\mathbb{Z}^+$, are said to be indistinguishable if

$$P(X_n = Y_n \text{ for all } n) = 1.$$

The process $(Y_n)_{n\in\mathbb{Z}^+}$ is said to be a version (or a modification) of $(X_n)_{n\in\mathbb{Z}^+}$ if $X_n = Y_n$ almost surely for every $n \in \mathbb{Z}^+$.

DR.RUPNATHJI(DR.RUPAK NATH)

If (Y_n) is a version of (X_n) then (X_n) and (Y_n) are indistinguishable. Indeed, for each $k \in \mathbb{Z}^+$, let $A_k = \{ X_k = Y_k \}$. We know that $P(A_k) = 1$, since (Y_n) is a version of (X_n), by hypothesis. But then

$$P(X_n = Y_n \text{ for all } n \in \mathbb{Z}^+) = P(\textstyle\bigcap_n A_n) = 1.$$

Warning: the extension of these definitions to processes indexed by $\mathbb{R}^+$ is clear, but the above implication can be false in the situation of continuous time, as the following example shows.

Example 4.5. Let $\Omega = [0,1]$ equipped with the Borel σ-algebra and the probability measure P determined by (Lebesgue measure) $P([a,b]) = b - a$ for any $0 \le a \le b \le 1$. For $t \in \mathbb{R}^+$, let $X_t(\omega) = 0$ for all $\omega \in \Omega = [0,1]$ and let

$$Y_t(\omega) = \begin{cases} 0, & \omega \ne t \\ 1, & \omega = t. \end{cases}$$

Fix $t \in \mathbb{R}^+$. Then $X_t(\omega) = Y_t(\omega)$ unless $\omega = t$. Hence $\{ X_t = Y_t \} = \Omega \setminus \{ t \}$ or Ω depending on whether $t \in [0,1]$ or not. So $P(X_t = Y_t) = 1$. On the other hand, $\{ X_t = Y_t \text{ for all } t \in \mathbb{R}^+ \} = \varnothing$ and so we have

$$\begin{aligned} &P(X_t = Y_t) = 1, \text{ for each } t \in \mathbb{R}^+, \\ &P(X_t = Y_t \text{ for all } t \in \mathbb{R}^+) = 0\,, \end{aligned}$$

which is to say that (Y_t) is a version of (X_t) but these processes are far from indistinguishable.

Note also that the path $t \mapsto X_t(\omega)$ is constant for every ω, whereas for every ω, the path $t \mapsto Y_t(\omega)$ has a jump at $t = \omega$. So the paths $t \mapsto X_t(\omega)$ are continuous almost surely, whereas with probability one, no path $t \mapsto Y_t(\omega)$ is continuous.

Example 4.6. A filtration $(\mathcal{G}_t)_{t \in \mathbb{R}^+}$ of sub-σ-algebras of a σ-algebra $\mathcal{S}$ is said to be right-continuous if $\mathcal{G}_t = \bigcap_{s>t} \mathcal{G}_s$ for any $t \in \mathbb{R}^+$.

For any given filtration $(\mathcal{F}_t)$, set $\mathcal{G}_t = \bigcap_{s>t} \mathcal{F}_s$. Fix $t \ge 0$ and suppose that $A \in \bigcap_{s>t} \mathcal{G}_s$. For any $r > t$, choose s with $t < s < r$. Then we see that $A \in \mathcal{G}_s = \bigcap_{v>s} \mathcal{F}_v$ and, in particular, $A \in \mathcal{F}_r$. Hence $A \in \bigcap_{r>t} \mathcal{F}_r = \mathcal{G}_t$ and so $(\mathcal{G}_t)_{t \in \mathbb{R}^+}$ is right-continuous.

Remark 4.7. Let $(X_t)_{t \in \mathbb{R}^+}$ be a process indexed by $\mathbb{R}^+$. Then one might be interested in the process given by $Y_t = \sup_{s \le t} X_t$ for $t \in \mathbb{R}^+$. Even though each X_t is a random variable, it is not at all clear that Y_t is measurable. However, suppose that $t \mapsto X_t$ is almost surely continuous and let $\Omega_0 \subset \Omega$ be such that $P(\Omega_0) = 1$ and $t \mapsto X_t(\omega)$ is continuous for each $\omega \in \Omega_0$. Then we can define

$$Y_t(\omega) = \begin{cases} \sup_{s \le t} X_t(\omega), & \text{for } \omega \in \Omega_0, \\ 0, & \text{otherwise.} \end{cases}$$

DR.RUPNATHJI(DR.RUPAK NATH)

Claim: Y_t is measurable.

Proof of claim. Suppose that $\varphi : [a,b] \to \mathbb{R}$ is a continuous function. If $D = \{ t_0, t_1, \dots, t_k \}$ with $a = t_0 < t_1 < \dots < t_k = b$ is a partition of $[a,b]$, let $\max_D \varphi$ denote $\max\{ \varphi(t) : t \in D \}$. Let

$$D_n = \{ a = t_0^{(n)} < t_1^{(n)} < \dots < t_{m_n}^{(n)} = b \}$$

be a sequence of partitions of $[a,b]$ such that $\mathrm{mesh}(D_n) \to 0$, where $\mathrm{mesh}(D)$ is given by $\mathrm{mesh}(D) = \max\{ t_{j+1} - t_j : t_j \in D \}$.

Then it follows that $\sup\{ \varphi(s) : s \in [a,b] \} = \lim_n \max_{D_n} \varphi$. This is because a continuous function φ on a closed interval $[a,b]$ is uniformly continuous there and so for any $\varepsilon > 0$, there is some $\delta > 0$ such that $|\varphi(t) - \varphi(s)| < \varepsilon$ whenever $|t - s| < \delta$. Suppose then that n is sufficiently large that $\mathrm{mesh}(D_n) < \delta$. Then for any $s \in [a,b]$, there is some $t_j \in D_n$ such that $|s - t_j| < \delta$. This means that $|\varphi(s) - \varphi(t_j)| < \varepsilon$. It follows that

$$\max_{D_n} \varphi \le \sup\{ \varphi(s) : s \in [a,b] \} < \max_{D_n} \varphi + \varepsilon$$

and so $\sup\{ \varphi(s) : s \in [a,b] \} = \lim_n \max_{D_n} \varphi$.

Now fix $t \ge 0$, set $a = 0$, $b = t$, fix $\omega \in \Omega$ and let $\varphi(s) = 1_{\Omega_0}(\omega)\, X_s(\omega)$. Evidently $s \mapsto \varphi(s)$ is continuous on $[0,t]$ and so

$$Y_t(\omega) = \sup\{ \varphi(s) : s \in [0,t] \} = \lim_n \max_{D_n} \varphi$$

as above. But for each n, $\omega \mapsto \max_{D_n} \varphi = \max\{ X_t(\omega) : t \in D_n \}$ is measurable and so the pointwise limit $\omega \mapsto Y_t(\omega)$ is measurable, which completes the proof of the claim. ∎

Example 4.8 (Doob's Maximal Inequality). Suppose now that (X_t) is a non-negative submartingale with respect to a filtration $(\mathcal{F}_t)$ such that $\mathcal{F}_0$ contains all sets of probability zero. Then $\Omega_0 \in \mathcal{F}_0$ and so the process $(\xi_t) = (1_{\Omega_0} X_t)$ is also a non-negative submartingale. Fix $t \ge 0$ and $n \in \mathbb{N}$. For $0 \le j \le 2^n$, let $t_j = tj/2^n$. Set $\eta_j = \xi_{t_j}$ and $\mathcal{G}_j = \mathcal{F}_{t_j}$ for $j \le 2^n$ and $\eta_j = \eta_t$ and $\mathcal{G}_j = \mathcal{F}_t$ for $j > 2^n$. Then (η_j) is a discrete parameter non-negative submartingale with respect to the filtration $(\mathcal{G}_j)$. According to the discussion above,

$$Y_t = \sup_{s \le t} \xi_s = \lim_n \max_{j \le 2^n} \eta_j = \sup_n \max_{j \le 2^n} \eta_j$$

on Ω. Now, by adjusting the inequalities in the proof of Doob's Maximal inequality, Theorem 3.28, we see that the result is also valid if the inequality $\max_k X_k \ge \lambda$ is replaced by $\max_k X_k > \lambda$ on both sides and so

$$\lambda\, P(\max_{j \le 2^n} \eta_j > \lambda) \le E(\, \eta_{2^n} 1_{\{ \max_{j \le 2^n} \eta_j > \lambda \}} \,)\,.$$

DR. RUPNATHJI(DR. RUPAK NATH)

Since $\eta_{2^n} = X_t$, we may say that

$$\lambda\, E(1_{\{\max_{j\le 2^n}\eta_j > \lambda\}}) \le E(\, X_t\, 1_{\{\max_{j\le 2^n}\eta_j > \lambda\}}\,)\,.$$

But $\max_{j\le 2^n}\eta_j \le Y_t$ and $\max_{j\le 2^n}\eta_j \to Y_t$ on Ω as $n\to\infty$ and therefore $1_{\{\max_{j\le 2^n}\eta_j>\lambda\}} \to 1_{\{Y_t>\lambda\}}$. Hence, letting $n\to\infty$, we find that

$$\lambda\, E(1_{\{Y_t>\lambda\}}) \le E(\, X_t\, 1_{\{Y_t>\lambda\}}\,) \qquad (*)$$

Replacing λ by λ_n in $(*)$ where now $\lambda_n \downarrow \lambda > 0$ and letting $n\to\infty$, we obtain

$$\lambda\, E(1_{\{Y_t\ge\lambda\}}) \le E(\, X_t\, 1_{\{Y_t\ge\lambda\}}\,),$$

a continuous version of Doob's Maximal inequality.

Similarly, we can obtain a version of Corollary 3.29 for continuous time. Suppose that $(X_t)_{t\in\mathbb{R}^+}$ is an $\mathcal{L}^2$ martingale such that the map $t\mapsto X_t(\omega)$ is almost surely continuous. Then for any $\lambda>0$

$$P(\,\sup_{s\le t}|X_s| \ge \lambda\,) \le \frac{1}{\lambda^2}\,\|X_t\|_2^2.$$

Proof. Let (D_n) be the sequence of partitions of $[0,t]$ as above and let $f_n(\omega) = \max_{j\le 2^n}|\xi_j(\omega)|$. Then $f_n \to f = \sup_{s\le t}|X_s|$ almost surely and since $f_n \le f$ almost surely, it follows that $1_{\{f_n>\mu\}} \to 1_{\{f>\mu\}}$ almost surely. By Lebesgue's Dominated Convergence Theorem, it follows that $E(1_{\{f_n>\mu\}}) \to E(1_{\{f>\mu\}})$, that is, $P(f_n>\mu)\to P(f>\mu)$.

Let $\mu<\lambda$. Applying Doob's maximal inequality, corollary 3.29, for the discrete-time filtration $(\mathcal{F}_s)$ with $s\in\{\,0=t_0^n, t_1^n,\dots,t_{m_n}^n = t\,\}$, we get

$$P(\, f_n > \mu\,) \le P(\, f_n \ge \mu\,) \le \frac{1}{\mu^2}\,\|X_t\|_2^2.$$

Letting $n\to\infty$, gives

$$\mu^2\, P(\, f>\mu\,) \le \|X_t\|_2^2\,. \qquad (*)$$

For $j\in\mathbb{N}$, let $\mu_j \uparrow \lambda$ and set $A_j = \{\, f>\mu_j\,\}$. Then $A_j \downarrow \{\, f\ge\lambda\,\}$ so that $P(A_j)\downarrow P(f\ge\lambda)$. Replacing μ in $(*)$ by μ_j and letting $j\to\infty$, we get the inequality

$$\lambda^2\, P(\, f\ge\lambda\,) \le \|X_t\|_2^2$$

as required. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

Stochastic integration – first steps

We wish to try to perform some kind of abstract integration with respect to martingales. It is usually straightforward to integrate step functions, so we shall try to do that here. First, we recall that if Z is a random variable, then its (cumulative) distribution function F is defined to be $F(t) = P(Z \le t)$ and we have

$$P(a < Z \le b) = P(Z \le b) - P(Z \le a) = F(b) - F(a)\,.$$

This can be written as $E(1_{\{\,a<Z\le b\,\}}) = F(b) - F(a)$ or as

$$\int_{-\infty}^{\infty} 1_{(a,b]}(s)\; dF(s) = F(b) - F(a)\,.$$

Notice that step-functions of the form $1_{(a,b]}$ (left-continuous) appear quite naturally here.

Now, suppose that $(X_t)_{t\in\mathbb{R}^+}$ is a square-integrable martingale (that is, $X_t \in \mathcal{L}^2$ for each t). We want to set-up something like $\int_0^T f\; dX_t$ so we begin with integrands which are step-functions.

Definition 4.9. A process $g : \mathbb{R}^+ \times \Omega \to \mathbb{R}$ is said to be elementary if it has the form

$$g(s,\omega) = h(\omega)\, 1_{(a,b]}(s)$$

for some $0 \le a < b$ and some bounded random variable $h(\omega)$ which is $\mathcal{F}_a$-measurable. Note that g is piecewise-constant in s (and continuous from the left).

The stochastic integral $\int_0^T g\; dX$ of the elementary process g with respect the $\mathcal{L}^2$-martingale (X_t) is defined to be the random variable

$$\int_0^T g\; dX = \int_0^T h\, 1_{(a,b]}(s)\; dX_s$$
$$= \begin{cases} h\,(X_b - X_a), & \text{if } T \ge b \\ h\,(X_T - X_a), & \text{if } a < T < b \\ 0, & \text{if } T \le a. \end{cases}$$

Proposition 4.10. *For a fixed elementary process g, the family $(\int_0^t g\, dX)_{t\in\mathbb{R}^+}$ is an $\mathcal{L}^2$-martingale.*

Proof. Suppose that $g = h\, 1_{(a,b]}$ and let

$$Y_t = \int_0^t g\, dX = h\,(X_{b\wedge t} - X_{a\wedge t})\,.$$

Now, if $t \ge a$, then h is $\mathcal{F}_t$-measurable and so is $X_{b\wedge t} - X_{a\wedge t}$. Since $Y_t = 0$ for $t < a$, we see that (Y_t) is adapted. Furthermore, since h is bounded, it

DR.RUPNATHJI(DR.RUPAK NATH)

follows that $Y_t \in \mathcal{L}^2$ for each $t \in \mathbb{R}^+$. To show that Y_t is a martingale, let $0 \le s \le t$ and consider three cases.

Case 1. $s \le a$.
Using the tower property $E(\cdot \,|\, \mathcal{F}_s) = E(E(\cdot \,|\, \mathcal{F}_a) \,|\, \mathcal{F}_s)$, we find that

$$\begin{aligned} E(Y_t \,|\, \mathcal{F}_s) &= E(h\,(X_{t\wedge b} - X_{t\wedge a}) \,|\, \mathcal{F}_s) \\ &= E(E(h\,(X_{t\wedge b} - X_{t\wedge a}) \,|\, \mathcal{F}_a) \,|\, \mathcal{F}_s) \\ &= E(h\ \underbrace{E((X_{t\wedge b} - X_{t\wedge a}) \,|\, \mathcal{F}_a)}_{=0 \text{ since } (X_t) \text{ is a martingale}}\ \mathcal{F}_s \,|\,) \quad \text{a.s.} \end{aligned}$$

But $s \le a$ implies that $Y_s = h\,(X_{s\wedge b} - X_{s\wedge a}) = 0$ and so

$$E(Y_t \,|\, \mathcal{F}_s) = 0 = Y_s \text{ almost surely.}$$

Case 2. $a \le s \le b$.
We have

$$\begin{aligned} E(Y_t \,|\, \mathcal{F}_s) &= E(h\,(X_{t\wedge b} - X_a) \,|\, \mathcal{F}_s) \\ &= h\,E((X_{t\wedge b} - X_a) \,|\, \mathcal{F}_s) \quad \text{a.s.} \\ &= h\,(X_{t\wedge b\wedge s} - X_a) \quad \text{a.s.} \\ &= h\,(X_{s\wedge b} - X_{s\wedge a}) \\ &= Y_s . \end{aligned}$$

Case 3. $b < s$.
We find

$$\begin{aligned} E(Y_t \,|\, \mathcal{F}_s) &= E(h\,(X_b - X_a) \,|\, \mathcal{F}_s) \\ &= h\,E((X_b - X_a) \,|\, \mathcal{F}_s) \quad \text{a.s.} \\ &= h\,(X_b - X_a) \quad \text{a.s.} \\ &= h\,(X_{s\wedge b} - X_{s\wedge a}) \\ &= Y_s \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

and the proof is complete. ■

Notation Let $\mathcal{E}$ denote the real linear span of the set of elementary processes. So any element $h \in \mathcal{E}$ has the form

$$h(\omega, s) = \sum_{i=1}^{n} g_i(\omega) 1_{(a_i, b_i]}(s)$$

for some n, pairs $a_i < b_i$ and bounded random variables g_i where each g_i is $\mathcal{F}_{a_i}$-measurable. Notice that $h(\omega, 0) = 0$. In fact, we are not interested in the value of h at $s = 0$ as far as integration is concerned. We could have included random variables of the form $g_0(\omega) 1_{\{0\}}(s)$ in the construction of $\mathcal{E}$, where g_0 is $\mathcal{F}_0$-measurable, but such elements play no rôle.

Definition 4.11. For $h = \sum_{i=1}^n g_i 1_{(a_i,b_i]} \in \mathcal{E}$ and $T \ge 0$, the stochastic integral $\int_0^T h\,dX$ is defined to be the random variable

$$\int_0^T h\,dX = \sum_{i=1}^n \int_0^T h_i\,dX = \sum_{i=1}^n g_i\,(X_{T\wedge b_i} - X_{T\wedge a_i})$$

where $h_i = g_i 1_{(a_i,b_i]}$.

Remark 4.12. It is crucial to note that the stochastic integral is a sum of terms $g_i(X_{T\wedge b_i} - X_{T\wedge a_i})$ where g_i is $\mathcal{F}_{a_i}$-measurable and $b_i > a_i$. The "increment" $(X_{T\wedge b_i} - X_{T\wedge a_i})$ "points to the future". This will play a central rôle in the development of the theory.

Proposition 4.13. *For $h \in \mathcal{E}$, the process $(\int_0^t g\,dX)_{t\in\mathbb{R}^+})$ is an $\mathcal{L}^2$-martingale.*

Proof. If $h = \sum_{i=1}^n h_i$, where each h_i is elementary as above, then

$$\int_0^t h\,dX = \sum_{i=1}^n \int_0^t h_i\,dX$$

which is a linear combination of $\mathcal{L}^2$-martingales, by Proposition 4.10. The result follows. ■

So far so good. Now if $Y_t = \int_0^t h\,dX$, can anything be said about $\|Y_t\|_2$ (or $E(Y_t^2)$)? We pursue this now.

Let $h \in \mathcal{E}$. Then h can be written as

$$h = \sum_{i=1}^n g_i\,1_{(t_i,t_{i+1}]}$$

for some m, $0 = t_1 < \cdots < t_m$ and $\mathcal{F}_{t_i}$-measurable random variables g_i (which may be 0). Let $Y_t = \int_0^t h\,dX$ and consider $E(Y_t^2)$. First we note that Y_t is unchanged if h is replaced by $h\,1_{(0,t]}$ which means that we may assume, without loss of generality, that $t_m = t$. This done, we consider

$$\begin{aligned} E(Y_t^2) &= E(\,(\sum_{i=1}^{m-1} g_i\,(X_{t_i+1} - X_{t_i})\,)^2\,) \\ &= E(\sum_{j=1}^{m-1}\sum_{i=1}^{m-1} g_i g_j\,(X_{t_i+1} - X_{t_i})(X_{t_j+1} - X_{t_j})\,)\,. \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

Now, suppose $i \neq j$, say $i < j$. Writing ΔX_i for $(X_{t_i+1} - X_{t_i})$, we find that

$$\begin{aligned} E(\, g_i g_j \, \Delta X_i \, \Delta X_j \,) &= E(\, E(g_i g_j \, \Delta X_i \, \Delta X_j \,|\, \mathcal{F}_{t_j})\,) \\ &= E(\, g_i g_j \, \Delta X_i \, E(\Delta X_j \,|\, \mathcal{F}_{t_j})\,)\,, \quad \text{a.s.,} \\ &\qquad \text{since } g_i g_j \, \Delta_i \text{ is } \mathcal{F}_{t_i}\text{-measurable,} \\ &= 0\,, \quad \text{since } (X_t) \text{ is a martingale.} \end{aligned}$$

Of course, this also holds for $i > j$ (simply interchange i and j) and so we may say that

$$E(\,(\int_0^t g \, dX\,)^2\,) = E(\,\sum_{j=1}^{m-1} g_j^2 \, \Delta X_j^2\,) = \sum_{j=1}^{m-1} E(\, g_j^2 \, \Delta X_j^2\,)\,.$$

Next, consider

$$\begin{aligned} E(\, g_j^2 \, \Delta X_j^2\,) &= E(\, g_j^2 \, (X_{t_{j+1}} - X_{t_j})^2\,) \\ &= E(\, g_j^2 \, (X_{t_{j+1}}^2 + X_{t_j}^2 - 2X_{t_{j+1}} X_{t_j})\,) \\ &= E(\, g_j^2 \, (X_{t_{j+1}}^2 + X_{t_j}^2)\,) - 2E(\, g_j^2 \, X_{t_{j+1}} X_{t_j}\,) \\ &= E(\, g_j^2 \, (X_{t_{j+1}}^2 + X_{t_j}^2)\,) - 2E(\, E(g_j^2 \, X_{t_{j+1}} X_{t_j} \,|\, \mathcal{F}_{t_j})\,) \\ &= E(\, g_j^2 \, (X_{t_{j+1}}^2 + X_{t_j}^2)\,) - 2E(\, g_j^2 \, X_{t_j} E(\, X_{t_{j+1}} \,|\, \mathcal{F}_{t_j})\,) \\ &= E(\, g_j^2 \, (X_{t_{j+1}}^2 + X_{t_j}^2)\,) - 2E(\, g_j^2 \, X_{t_j} X_{t_j}\,), \\ &\qquad \text{since } (X_t) \text{ is a martingale,} \\ &= E(\, g_j^2 \, (X_{t_{j+1}}^2 - X_{t_j}^2)\,)\,. \end{aligned}$$

What now? We have already seen that the square of a discrete-time $\mathcal{L}^2$-martingale has a decomposition as the sum of an $\mathcal{L}^1$-martingale and a predictable increasing process. So let us assume that (X_t^2) has the Doob-Meyer decomposition

$$X_t^2 = M_t + A_t$$

where (M_t) is an $\mathcal{L}^1$-martingale and (A_t) is an $\mathcal{L}^1$-process such that $A_0 = 0$ and $A_s(\omega) \le A_t(\omega)$ almost surely whenever $s \le t$.

How does this help? Setting $\Delta M_j = M_{t_{j+1}} - M_{t_j}$ and similarly $\Delta A_j = A_{t_{j+1}} - A_{t_j}$, we see that

$$\begin{aligned} E(\, g_j^2 \, (X_{t_{j+1}}^2 - X_{t_j}^2)\,) &= E(\, g_j^2 \, \Delta M_j\,) + E(\, g_j^2 \, \Delta A_j\,) \\ &= E(\, E(g_j^2 \, \Delta M_j \,|\, \mathcal{F}_{t_j})\,) + E(\, g_j^2 \, \Delta A_j\,) \\ &= E(\, g_j^2 \underbrace{E((M_{t_{j+1}} - M_{t_j}) \,|\, \mathcal{F}_{t_j})}_{= 0 \text{ since } (M_t) \text{ is a martingale}}\,) + E(\, g_j^2 \, \Delta A_j\,) \\ &= E(\, g_j^2 \, \Delta A_j)\,. \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

Hence

$$E\Big(\Big(\int_0^t g\,dX\Big)^2\Big) = \sum_{j=1}^{m-1} E\big(\,g_j^2\,(A_{t_{j+1}} - A_{t_j})\,\big)\,.$$

Now, on a set of probability one, $A_t(\omega)$ is an increasing function of t and we can consider Stieltjes integration using this. Indeed,

$$\begin{aligned}\int_0^t g(s,\omega)^2\,dA_s(\omega) &= \sum_{j=1}^{m-1} \int_0^t g_j^2(\omega)\,1_{(t_j,t_{j+1}]}(s)\,dA_s(\omega)\\ &= \sum_{j=1}^{m-1} g_j^2(\omega)\,(\,A_{t_{j+1}}(\omega) - A_{t_j}(\omega)\,)\,.\end{aligned}$$

Taking expectations,

$$E\Big(\int_0^t g^2\,dA\,\Big) = E\Big(\sum g_j^2\,\Delta A\,\Big)$$

and this leads us finally to the formula

$$\boxed{E\Big(\Big(\int_0^t g\,dX_s\Big)^2\Big) = E\Big(\int_0^t g^2\,dA_s\,\Big)}$$

known as the isometry property. This isometry relation allows one to extend the class of integrands in the stochastic integral. Indeed, suppose that (g_n) is a sequence from $\mathcal{E}$ which converges to a map $h : \mathbb{R}^+ \times \Omega \to \mathbb{R}$ in the sense that $E(\int_0^t (g_n - h)^2\,dA_s\,) \to 0$. The isometry property then tells us that the sequence $(\int_0^t g_n\,dX)$ of random variables is a Cauchy sequence in $\mathcal{L}^2$ and so converges to some Y_t in $\mathcal{L}^2$. This allows us to define $\int_0^t h\,dX$ as this Y_t. We will consider this again for the case when (X_t) is a Wiener process.

Remark 4.14. The Doob-Meyer decomposition of X_t^2 as $M_t + A_t$ is far from straightforward and involves further technical assumptions. The reader should consult the text of Meyer for the full picture.

DR.RUPNATHJI(DR.RUPAK NATH)

DR.RUPNATHJI(DR.RUPAK NATH)

Chapter 5

Wiener process

DR.RUPNATHJI(DR.RUPAK NATH)

We begin, without further ado, with the definition.

Definition 5.1. An adapted process $(W_t)_{t\in\mathbb{R}^+}$ on a filtered probability space $(\Omega, \mathcal{S}, P, (\mathcal{F}_t))$ is said to be a Wiener process starting at 0 if it obeys the following:

(a) $W_0 = 0$ almost surely and the map $t \mapsto W_t$ is continuous almost surely.

(b) For any $0 \le s < t$, the random variable $W_t - W_s$ has a normal distribution with mean 0 and variance $t - s$.

(c) For all $0 \le s < t$, $W_t - W_s$ is independent of $\mathcal{F}_s$.

Remarks 5.2.

1. From (b), we see that the distribution of $W_{s+t} - W_s$ is normal with mean 0 and variance t; so

$$P(W_{s+t} - W_s \in A) = \frac{1}{\sqrt{2\pi t}} \int_A e^{-x^2/2t}\, dx$$

for any Borel set A in $\mathbb{R}$. Furthermore, by (a), $W_t = W_{t+0} - W_0$ almost surely, so each W_t has a normal distribution with mean 0 and variance t.

2. For each fixed $\omega \in \Omega$, think of the map $t \mapsto W_t(\omega)$ as the path of a particle (with t interpreted as time). Then (a) says that all paths start at 0 (almost surely) and are continuous (almost surely).

3. The independence property (c) implies that any collection of increments

$$W_{t_1} - W_{s_1},\ W_{t_2} - W_{s_2},\ \dots,\ W_{t_n} - W_{s_n}$$

with $0 \le s_1 < t_1 \le s_2 < t_2 \le s_3 < \cdots \le s_n < t_n$ are independent. Indeed, for any Borel sets $A_1, \dots, A_n$ in $\mathbb{R}$, we have

$$\begin{aligned} P(&\underbrace{W_{t_1} - W_{s_1}}_{\Delta W_1} \in A_1, \dots, \underbrace{W_{t_n} - W_{s_n}}_{\Delta W_n} \in A_n) \\ &= P(\{\, \Delta W_1 \in A_1, \dots, \Delta W_{n-1} \in A_{n-1} \,\} \cap \{\, \Delta W_n \in A_n \,\}) \end{aligned}$$

$$
\begin{aligned}
&= P(\Delta W_1 \in A_1, \dots, \Delta W_{n-1} \in A_{n-1})\, P(\Delta W_n \in A_n), \\
&\qquad \text{since } \{\, \Delta W_j \in A_j \,\} \in \mathcal{F}_{s_n} \text{ for all } 1 \le j \le n-1 \\
&\qquad \text{and } \Delta W_n \text{ is independent of } \mathcal{F}_{s_n}, \\
&= \quad \cdots \\
&= P(\Delta W_1 \in A_1)\, P(\Delta W_2 \in A_2) \dots P(\Delta W_n \in A_n)
\end{aligned}
$$

as required.

4. A d-dimensional Wiener process (starting from $0 \in \mathbb{R}^d$) is a d-tuple $(W_t^1, \dots W_t^d)$ where the $(W_t^1), \dots (W_t^d)$ are independent Wiener processes in $\mathbb{R}$ (starting at 0).

5. A Wiener process is also referred to as Brownian motion.

6. Such a Wiener process exists. In fact, let $p(x,t) = \frac{1}{\sqrt{2\pi t}}\, e^{-x^2/2t}$ denote the density of $W_t = W_t - W_0$ and let $0 < t_1 < \cdots < t_n$. Then to say that $W_{t_1} = x_1$, $W_{t_2} = x_2$, $\dots$, $W_{t_n} = x_n$ is to say that $W_{t_1} = x_1$, $W_{t_2} - W_{t_2} = x_2 - x_1$, $\dots$, $W_{t_n} - W_{t_{n-1}} = x_n - x_{n-1}$. Now, the random variables $W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_n} - W_{t_{n-1}}$ are independent so their joint density is a product of individual densities. This suggests that the joint probability density of $W_{t_1}, \dots, W_{t_n}$ is

$$
\begin{aligned}
&p(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) \\
&\qquad = p(x_1, t_1)\, p(x_2 - x_1, t_2 - t_1) \;\dots\; p(x_n - x_{n-1}, t_n - t_{n-1})\,.
\end{aligned}
$$

Let $\Omega_t = \dot{\mathbb{R}}$, the one-point compactification of $\mathbb{R}$, and let $\Omega = \prod_{t \in \mathbb{R}^+} \Omega_t$ be the (compact) product space. Suppose that $f \in C(\Omega)$ depends only on a finite number of coordinates in Ω, $f(\omega) = f(x_{t_1}, \dots, x_{t_n})$, say. Then we define

$$
\rho(f) = \int_{\mathbb{R}^n} p(x_1, \dots, x_n; t_1, \dots, t_n) f(x_1, \dots, x_n)\, dx_1 \dots dx_n\,.
$$

It can be shown that ρ can be extended into a (normalized) positive linear functional on the algebra $C(\Omega)$ and so one can apply the Riesz-Markov Theorem to deduce the existence of a probability measure μ on Ω such that

$$
\rho(f) = \int_\Omega f(\omega)\, d\mu\,.
$$

Then $W_t(\omega) = \omega_t$, the t-th component of $\omega \in \Omega$.
(This elegant construction is due to Edward Nelson.)

DR.RUPNATHJI(DR.RUPAK NATH)

We collect together some basic facts in the following theorem.

Theorem 5.3. *The Wiener process enjoys the following properties.*

(i) $E(W_s W_t) = \min\{s, t\} = s \wedge t$.

(ii) $E((W_t - W_s)^2) = |t - s|$.

(iii) $E(W_t^4) = 3t^2$.

(iv) (W_t) *is an* $\mathcal{L}^2$*-martingale.*

(v) *If* $M_t = W_t^2 - t$*, then* (M_t) *is a martingale, null at* 0.

Proof. (i) Suppose $s \le t$. Using independence, we calculate

$$\begin{aligned} E(W_s W_t) &= E(W_s (W_t - W_s)) + E(W_s^2) \\ &= \underbrace{E(W_s)}_{=0} \underbrace{E(W_t - W_s)}_{=0} + \underbrace{\operatorname{var} W_s}_{\text{since } E(W_s)=0} \\ &= s\,. \end{aligned}$$

(ii) Again, suppose $s \le t$. Then

$$\begin{aligned} E((W_t - W_s)^2) &= \operatorname{var}(W_t - W_s), \text{ since } E(W_t - W_s) = 0, \\ &= t - s\,, \text{ by definition.} \end{aligned}$$

(iii) We have

$$\int_{-\infty}^{\infty} e^{-\alpha x^2/2}\, dx = \alpha^{-1/2} \sqrt{2\pi}\,.$$

Differentiating both sides twice with respect to α gives

$$\int_{-\infty}^{\infty} x^4\, e^{-\alpha x^2/2}\, dx = 3\alpha^{-5/2} \sqrt{2\pi}\,.$$

Replacing α by $1/t$ and rearranging, we get

$$E(W_t^4) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x^4\, e^{-x^2/2t}\, dx = 3t^2\,.$$

and the proof is complete.

(iv) For $0 \le s < t$, we have (using independence)

$$\begin{aligned} E(W_t \mid \mathcal{F}_s) &= E(W_t - W_s \mid \mathcal{F}_s) + E(W_s \mid \mathcal{F}_s) \\ &= E(W_t - W_s) + W_s \quad \text{a.s.} \\ &= W_s \quad \text{a.s.} \end{aligned}$$

$W_t = W_t - W_0$ has a normal distribution, so $W_t \in \mathcal{L}^2$.

DR.RUPNATHJI(DR.RUPAK NATH)

(v) Let $0 \le s < t$. Then, with probability one,

$$\begin{aligned} E(W_t^2 \mid \mathcal{F}_s) &= E((W_t - W_s)^2 \mid \mathcal{F}_s) + 2E(W_t W_s \mid \mathcal{F}_s) - E(W_s^2 \mid \mathcal{F}_s) \\ &= E((W_t - W_s)^2) + 2W_s E(W_t \mid \mathcal{F}_s) - W_s^2 \\ &= t - s + 2W_s^2 - W_s^2\,, \quad \text{by (ii) and (iv),} \\ &= t - s + W_s^2 \end{aligned}$$

so that $E(\,(W_t^2 - t) \mid \mathcal{F}_s) = (W_s^2 - s)$ almost surely. ■

Example 5.4. For $a \in \mathbb{R}$, $(\,e^{aW_t - \frac{1}{2}a^2 t}\,)$ (and so $(\,e^{W_t - \frac{1}{2}t}\,)$, in particular) is a martingale. Indeed, for $s \le t$,

$$\begin{aligned} E(e^{aW_t - \frac{1}{2}a^2 t} \mid \mathcal{F}_s) &= E(e^{a(W_t - W_s) + aW_s - \frac{1}{2}a^2 t} \mid \mathcal{F}_s) \\ &= e^{-\frac{1}{2}a^2 t} E(e^{a(W_t - W_s)} e^{aW_s} \mid \mathcal{F}_s) \\ &= e^{-\frac{1}{2}a^2 t} e^{aW_s} E(e^{a(W_t - W_s)} \mid \mathcal{F}_s) \\ &= e^{-\frac{1}{2}a^2 t} e^{aW_s} E(e^{a(W_t - W_s)})\,, \text{ by independence,} \\ &= e^{-\frac{1}{2}a^2 t} e^{aW_s}\, e^{\frac{1}{2}a^2 (t-s)} \\ &= e^{aW_s - \frac{1}{2}a^2 s} \end{aligned}$$

since we know that $W_t - W_s$ has a normal distribution with mean zero and variance $t - s$.

DR.RUPNATHJI(DR.RUPAK NATH)

Example 5.5. For $k \in \mathbb{N}$, $E(W_t^{2k}) = \dfrac{(2k)!\, t^k}{2^k\, k!}$.

To see this, let $I_k = E(W_t^{2k})$ and for $n \in \mathbb{N}$, let $P(n)$ be the statement that $E(W_t^{2n}) = \dfrac{(2n)!\, t^n}{2^n\, n!}$. Since $I_1 = t$, we see that $P(1)$ is true. Integration by parts, gives

$$E(W_t^{2k}) = E(W_t^{2k+2})/t(2k+1)$$

and so the truth of $P(k)$ implies that

$$E(W_t^{2k+2}) = t(2k+1)\, E(W_t^{2k}) = t(2k+1)\, \frac{(2k)!\, t^k}{2^k\, k!} = \frac{(2(k+1))!\, t^{(k+1)}}{2^{k+1}\, (k+1)!}$$

which says that $P(k+1)$ is true. By induction, $P(n)$ is true for all $n \in \mathbb{N}$.

Example 5.6. Let $D_n = \{ 0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{m_n}^{(n)} = T \}$ be a sequence of partitions of the interval $[0,T]$ such that $\operatorname{mesh}(D_n) \to 0$, where $\operatorname{mesh}(D) = \max\{ t_{j+1} - t_j : t_j \in D \}$. For each n and $0 \le j \le m_n - 1$, let $\Delta^n W_j$ denote the increment $\Delta^n W_j = W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}}$ and set $\Delta^n t_j = t_{j+1}^{(n)} - t_j^{(n)}$. Then

$$\sum_{j=0}^{m_n} (\Delta^n W_j)^2 \to T$$

in $\mathcal{L}^2$ as $n \to \infty$.

To see this, we calculate

$$\begin{aligned}
\| \sum_{j=0}^{m_n} (\Delta^n W_j)^2 - T \|^2 &= E\Big(\Big(\sum_{j=0}^{m_n} (\Delta^n W_j)^2 - \sum_{j=0}^{m_n} \Delta^n t_j \Big)^2 \Big) \\
&= \sum_{i,j} E\big(((\Delta^n W_i)^2 - \Delta^n t_i)((\Delta^n W_j)^2 - \Delta^n t_j) \big) \\
&= \sum_{j} E\big(((\Delta^n W_j)^2 - \Delta^n t_j)^2 \big), \\
&\qquad \text{(off-diagonal terms vanish by independence)}, \\
&= \sum_{j} E\big((\Delta^n W_j)^4 - 2 (\Delta^n W_j)^2 \Delta^n t_j + (\Delta^n t_j)^2 \big) \\
&= \sum_{j} 2 (\Delta^n t_j)^2 \\
&\le 2 \operatorname{mesh}(D_n) \sum_{j} \Delta^n t_j \\
&= 2 \operatorname{mesh}(D_n)\, T \to 0
\end{aligned}$$

as required. This is the quadratic variation of the Wiener process.

DR.RUPNATHJI(DR. RUPAK NATH)

Example 5.7. For any $c > 0$, $Y_t = \frac{1}{c} W_{c^2 t}$ is a Wiener process with respect to the filtration generated by the Y_ts.

We can see this as follows. Clearly, $Y_0 = 0$ almost surely and the map $t \mapsto Y_t(\omega) = W_{c^2 t}(\omega)/c$ is almost surely continuous because $t \mapsto W_t(\omega)$ is. Also, for any $0 \le s < t$, the distribution of the increment $Y_t - Y_s$ is that of $(W_{c^2 t} - W_{c^2 s})/c$, namely, normal with mean zero and variance $(c^2 t - c^2 s)/c^2 = t - s$. Let $\mathcal{G}_t = \mathcal{F}_{c^2 t}$ which is equal to the σ-algebra generated by the random variables $\{ Y_s : s \le t \}$. Then $c(Y_t - Y_s) = (W_{c^2 t} - W_{c^2 s})$ is independent of $\mathcal{F}_{c^2 s} = \mathcal{G}_s$, and so therefore is $Y_t - Y_s$. Hence (Y_t) is a Wiener process with respect to $(\mathcal{G}_t)$.

Remark 5.8. For $t > 0$, let $Y_t = t\,X_{1/t}$ and set $Y_0 = 0$. Then for any $0 < s < t$

$$Y_t - Y_s = t\,X_{1/t} - s\,X_{1/s} = (t-s)\,X_{1/t} - s\,(X_{1/s} - X_{1/t})\,.$$

Now, $0 < 1/t < 1/s$ and so $X_{1/t}$ and $(X_{1/s} - X_{1/t})$ are independent normal random variables with zero means and variances given by $1/t$ and $1/s - 1/t$, respectively. It follows that $Y_t - Y_s$ is a normal random variable with mean zero and variance $(t-s)^2/t + s^2(1/s - 1/t) = (t-s)$. When $s = 0$, we see that $Y_t - Y_0 = Y_t = t\,X_{1/t}$ which is a normal random variable with mean zero and variance $t^2/t = t$.

Let $(\mathcal{G}_t)$ be the filtration where $\mathcal{G}_t$ is the σ-algebra generated by the family $\{\,Y_s : s \le t\,\}$. Again, suppose that $0 < s < t$. Then for any $r < s$

$$\begin{aligned}
E((Y_t - Y_s)\,Y_r) &= E(\,(t\,X_{1/t} - s\,X_{1/s})\,r\,X_{1/r}) \\
&= rt\,E(X_{1/t}\,X_{1/r}) - rs\,E(X_{1/s}\,X_{1/r}) \\
&= rt\,E(X_{1/t}\,(X_{1/r} - X_{1/t})) + rt\,E(X_{1/t}\,X_{1/t}) \\
&\qquad - rsE(X_{1/s}\,(X_{1/r} - X_{1/s})) - rs\,E(X_{1/s}\,X_{1/s}) \\
&= rt\,E(X_{1/t})\,E(X_{1/r} - X_{1/t}) + rt\,\mathrm{var}(X_{1/t}) \\
&\qquad - rsE(X_{1/s})\,E(X_{1/r} - X_{1/s}) - rs\,\mathrm{var}(X_{1/s}) \\
&= 0 + rt\,(1/t) - 0 - rs\,(1/s) \\
&= 0\,.
\end{aligned}$$

This shows that $(Y_t - Y_s)$ is orthogonal in $\mathcal{L}^2$ to each Y_r. But orthogonality for jointly normal random variables implies independence and so it follows that the increment $Y_t - Y_s$ is independent of $\mathcal{G}_s$.

Finally, it can be shown that the map $t \mapsto Y_t$ on $\mathbb{R}^+$ is continuous almost surely. In fact, it is evidently continuous almost surely on $(0,\infty)$ because this is true of $t \mapsto X_t$. The continuity at $t = 0$ requires additional argument. Notice that it is easy to show that $t \mapsto Y_t$ is $\mathcal{L}^2$-continuous. For $s > 0$ and $t > 0$, we find that

$$\begin{aligned}
\|Y_t - Y_s\|_2 &= \|\,t\,X_{1/t} - s\,X_{1/s}\|_2 \\
&\le \|\,t\,X_{1/t} - t\,X_{1/s}\|_2 + \|\,t\,X_{1/s} - s\,X_{1/s}\|_2 \\
&= t\,|1/s - 1/t|^{1/2} + |t-s|\;(1/s)^{1/2} \\
&\to 0 \text{ as } s \to t.
\end{aligned}$$

At $t = 0$, we have

$$\|Y_t - Y_0\|_2^2 = \|Y_t\|_2^2 = t^2\,\|X_{1/t}\|_2^2 = t^2\;\mathrm{var}\,X_{1/t} = t \to 0$$

as $t \downarrow 0$. This is example is useful in that it relates large time behaviour of the Wiener process to small time behaviour. The behaviour of X_t for large t is related to that of Y_s for small s and both X_t and Y_s are Wiener processes.

DR.RUPNATHJI(DR.RUPAK NATH)

Example 5.9. Let (W_t) be a Wiener process and let $X_t = \mu t + \sigma W_t$ for $t \ge 0$ (where μ and σ are constants). Then (X_t) is a martingale if $\mu = 0$ but is a submartingale if $\mu \ge 0$.

We see that for $0 \le s < t$,

$$E(X_t \,|\, \mathcal{F}_s) = E(\mu t + \sigma W_t \,|\, \mathcal{F}_s) = \mu t + \sigma W_s > \mu s + \sigma W_s = X_s \,.$$

Non-differentiability of Wiener process paths

Whilst the sample paths $W_t(\omega)$ are almost surely continuous in t, they are nevertheless extremely jagged, as the next result indicates.

Theorem 5.10. *With probability one, the sample path $t \mapsto W_t(\omega)$ is nowhere differentiable.*

Proof. First we just consider $0 \le t \le 1$. Fix $\beta > 0$. Now if a given function f is differentiable at some point s with $f'(s) \le \beta$, then certainly we can say that

$$|f(t) - f(s)| \le 2\beta\,|t - s|$$

whenever $|t-s|$ is sufficiently small. (If t is sufficiently close to s, then $(f(t)-f(s))/(t-s)$ is within β of $f'(s)$ and so is smaller than $f'(s)+\beta \le 2\beta$). So let

$$A_n = \{\, \omega : \exists s \in [0,1) \text{ such that } |W_t(\omega) - W_s(\omega)| \le 2\beta\,|t-s|\,, \quad \text{whenever } |t-s| \le \tfrac{2}{n} \,\}\,.$$

Evidently $A_n \subset A_{n+1}$ and $\bigcup_n A_n$ includes all $\omega \in \Omega$ for which the function $t \mapsto W_t(\omega)$ has a derivative at some point in $[0,1)$ with value $\le \beta$.

Once again, suppose f is some function such that for given s there is some $\delta > 0$ such that

$$|f(t) - f(s)| \le 2\beta\,|t-s| \qquad (*)$$

if $|t-s| \le 2\delta$. Let $n \ge 1/\delta$ and let k be the largest integer such that $k/n \le s$. Then

$$\max\{\,|\,f(\tfrac{k+2}{n}) - f(\tfrac{k+1}{n})\,|\,,\; |\,f(\tfrac{k+1}{n}) - f(\tfrac{k}{n})\,|\,,\; |\,f(\tfrac{k}{n}) - f(\tfrac{k-1}{n})\,|\,\} \le \tfrac{6\beta}{n} \quad (**)$$

To verify this, note first that $\frac{k-1}{n} < \frac{k}{n} \le s < \frac{k+1}{n} < \frac{k+2}{n}$ and so we may estimate each of the three terms involved in $(**)$ with the help of $(*)$. We find

$$\begin{aligned} |\,f(\tfrac{k+2}{n}) - f(\tfrac{k+1}{n})\,| &\le |\,f(\tfrac{k+2}{n}) - f(s)\,| + |\,f(s) - f(\tfrac{k+1}{n})\,| \\ &\le 2\beta|\tfrac{k+2}{n} - s| + 2\beta|s - \tfrac{k+1}{n}| \\ &\le \tfrac{6\beta}{n} \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

and

$$\begin{aligned}
|\, f(\tfrac{k+1}{n}) - f(\tfrac{k}{n})\,| &\le |\, f(\tfrac{k+1}{n}) - f(s)\,| + |\, f(s) - f(\tfrac{k}{n})\,| \\
&\le 2\beta|\tfrac{k+1}{n} - s| + 2\beta|s - \tfrac{k}{n}| \\
&\le \tfrac{4\,\beta}{n}
\end{aligned}$$

and

$$\begin{aligned}
|\, f(\tfrac{k}{n}) - f(\tfrac{k-1}{n})\,| &\le |\, f(\tfrac{k}{n}) - f(s)\,| + |\, f(s) - f(\tfrac{k-1}{n})\,| \\
&\le 2\beta|\tfrac{k}{n} - s| + 2\beta|s - \tfrac{k-1}{n}| \\
&\le \tfrac{6\,\beta}{n}
\end{aligned}$$

which establishes the inequality $(**)$.

For given ω, let

$$\begin{aligned}
g_k(\omega) = \max\{\, |W_{\frac{k+2}{n}}(\omega) - W_{\frac{k+1}{n}}(\omega)|\ ,\ |W_{\frac{k+1}{n}}(\omega) - W_{\frac{k}{n}}(\omega)|\ , \\
|W_{\frac{k}{n}}(\omega) - W_{\frac{k-1}{n}}(\omega)|\,\}
\end{aligned}$$

and let

$$B_n = \{\, \omega :\ g_k(\omega) \le \tfrac{6\beta}{n} \text{ for some } k \le n-2\ \}\,.$$

Now, if $\omega \in A_n$, then $W_t(\omega)$ is differentiable at some s and furthermore $|W_t(\omega) - W_s(\omega)| \le 2\beta\,|t-s|$ if $|t-s| \le 2/n$. However, according to our discussion above, this means that $g_k(\omega) \le 6\beta/n$ where k is the largest integer with $k/n \le s$. Hence $\omega \in B_n$ and so $A_n \subset B_n$. Now,

$$B_n = \bigcup_{k=1}^{n-2} \{\, \omega : g_k(\omega) \le \tfrac{6\,\beta}{n}\,\}$$

and so

$$P(B_n) \le \sum_{k=1}^{n-2} P(\, g_k \le \tfrac{6\,\beta}{n}\,)\,.$$

We estimate

$$\begin{aligned}
P(\, g_k \le \tfrac{6\,\beta}{n}\,) &= P(\,\max\{\, |W_{\frac{k+2}{n}} - W_{\frac{k+1}{n}}|\,, \\
&\qquad |W_{\frac{k+1}{n}} - W_{\frac{k}{n}}|\,,\, |W_{\frac{k}{n}} - W_{\frac{k-1}{n}}|\,\} \le \tfrac{6\,\beta}{n}\,) \\
&= P(\,\{\, |W_{\frac{k+2}{n}} - W_{\frac{k+1}{n}}| \le \tfrac{6\,\beta}{n}\,\} \cap \\
&\qquad \{\, |W_{\frac{k+1}{n}} - W_{\frac{k}{n}}| \le \tfrac{6\,\beta}{n}\,\} \cap \{\, |W_{\frac{k}{n}} - W_{\frac{k-1}{n}}|\,\} \le \tfrac{6\,\beta}{n}\,\}\,) \\
&= P(\,\{\, |W_{\frac{k+2}{n}} - W_{\frac{k+1}{n}}| \le \tfrac{6\,\beta}{n}\,\}\,) \times \\
&\qquad \times P(\,\{\, |W_{\frac{k+1}{n}} - W_{\frac{k}{n}}| \le \tfrac{6\,\beta}{n}\,\}\,)\, P(\,\{\, |W_{\frac{k}{n}} - W_{\frac{k-1}{n}}| \le \tfrac{6\,\beta}{n}\,\}\,)\,, \\
&\qquad\qquad \text{by independence of increments,}
\end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

$$= \left(\sqrt{\tfrac{n}{2\pi}} \int_{-6\beta/n}^{6\beta/n} e^{-n\,x^2/2}\,dx \right)^3$$
$$\le \left(\sqrt{\tfrac{n}{2\pi}}\ \tfrac{12\,\beta}{n} \right)^3$$
$$= C\,n^{-3/2}$$

where C is independent of n. Hence

$$P(B_n) \le n\,C\,n^{-3/2} = C\,n^{-1/2} \to 0$$

as $n \to \infty$. Now for fixed j, $A_j \subset A_n$ for all $n \ge j$, so

$$P(A_j) \le P(A_n) \le P(B_n) \to 0$$

as $n \to \infty$, and this forces $P(A_j) = 0$. But then $P(\bigcup_j A_j) = 0$.

What have we achieved so far? We have shown that $P(\bigcup_n A_n) = 0$ and so, with probability one, W_t has no derivative at any $s \in [0,1)$ with value smaller than β. Now consider any unit interval $[j, j+1)$ with $j \in \mathbb{Z}^+$. Repeating the previous argument but with W_t replaced by $X_t = W_{t+j}$, we deduce that almost surely X_t has no derivative at any $s \in [0,1)$ whose value is smaller than β. But to say that X_t has a derivative at s is to say that W_t has a derivative at $s+j$ and so it follows that with probability one, W_t has no derivative in $[j, j+1)$ whose value is smaller than β. We have shown that if

$$C_j = \{\,\omega : W_t(\omega) \text{ has a derivative at some } s \in [j, j+1) \text{ with value} \le \beta\,\}$$

then $P(C_j) = 0$ and so $P(\bigcup_{j \in \mathbb{Z}^+} C_j) = 0$, which means that, with probability one, W_t has no derivative at any point in $\mathbb{R}^+$ whose value is $\le \beta$. This holds for any given $\beta > 0$.

To complete the proof, for $m \in \mathbb{N}$, let

$$S_m = \{\,\omega : W_t(\omega) \text{ has a derivative at some } s \in \mathbb{R}^+ \text{ with value} \le m\,\}\,.$$

Then we have seen above that $P(S_m) = 0$ and so $P(\bigcup_m S_m) = 0$. It follows that, with probability one, W_t has no derivative anywhere. ∎

DR RUPNATHJI(DR RUPAK NATH)

Itô integration

We wish to indicate how one can construct stochastic integrals with respect to the Wiener process. The resulting integral is called the Itô integral. We follow the strategy discussed earlier, namely, we first set-up the integral for integrands which are step-functions. Next, we establish an appropriate isometry property and it then follows that the definition can be extended abstractly by continuity considerations.

We shall consider integration over the time interval $[0,T]$, where $T>0$ is fixed throughout. For an elementary process, $h \in \mathcal{E}$, we define the stochastic integral $\int_0^T h\, dW$ by

$$\int_0^T h\, dW \equiv I(h)(\omega) = \sum_{i=1}^n g_{i+1}(\omega)\,(\,W_{t_{i+1}}(\omega) - W_{t_i}(\omega)\,)\,.$$

where $h = \sum_{i=1}^n g_i\, 1_{(t_i,t_{i+1}]}$ with $0 = t_1 < \dots < t_{n+1} = T$ and where g_i is bounded and $\mathcal{F}_{t_i}$-measurable. Now, we know that W_t^2 has Doob-Meyer decomposition $W_t^2 = M_t + t$, where (M_t) is an $\mathcal{L}^1$-martingale. Using this, we can calculate $E(I(h)^2)$ as in the abstract set-up and we find that

$$E(I(h)^2) = E\Big(\int_0^T h^2(t,\omega)\, dt\,\Big)$$

which is the isometry property for the Itô-integral.

For any $0 \le t \le T$, we define the stochastic integral $\int_0^t h\, dW$ by

$$\int_0^t h\, dW = I(h\, 1_{[0,t]}) = \sum_{i=1}^n g_{i+1}(\omega)\,(\,W_{t_{i+1}\wedge t}(\omega) - W_{t_i\wedge t}(\omega)\,)$$

and it is convenient to denote this by $I_t(h)$. We see (as in the abstract theory discussed earlier) that for any $0 \le s < t \le T$

$$E(I_t(h)\,|\,\mathcal{F}_s) = I_s(h)\,,$$

that is, $(I_t(h))_{0\le t\le T}$ is an $\mathcal{L}^2$-martingale.

Next, we wish to extend the family of allowed integrands. The right hand side of the isometry property suggests the way forward. Let $\mathcal{K}_T$ denote the linear subspace of $\mathcal{L}^2((0,T]\times\Omega)$ of adapted processes $f(t,\omega)$ such that there is some sequence (h_n) of elementary processes such that

$$E\Big(\int_0^T (\,f(s,\omega) - h_n(s,\omega)\,)^2\, ds\Big) \to 0\,. \qquad (*)$$

We construct the stochastic integral $I(f)$ (and $I_t(f)$) for any $f \in \mathcal{K}_T$ via the isometry property. Indeed, we have

$$\begin{aligned} E(\,(I(h_n) - I(h_m))^2\,) &= E(\,I(h_n - h_m)^2\,) \\ &= E\big(\int_0^T (h_n - h_m)^2\, ds\,\big)\,. \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

But by $(*)$, (h_n) is a Cauchy sequence in $\mathcal{K}_T$ (with respect the norm $\|h\|_{\mathcal{K}_T} = E(\int_0^T h^2\,ds)^{1/2}$) and so $(I(h_n))$ is a Cauchy sequence in $\mathcal{L}^2$. It follows that there is some $F \in \mathcal{L}^2(\mathcal{F}_T)$ such that

$$E(\,(F - I(h_n))^2\,) \to 0\,.$$

We denote F by $I(f)$ or by $\int_0^T f\,dW_s$. One checks that this construction does not depend on the particular choice of the sequence (h_n) converging to f in $\mathcal{K}_T$. The Itô-stochastic integral obeys the isometry property

$$\boxed{E(\,(\int_0^T f\,dW_s\,)^2\,) = \int_0^T E(f^2)\,ds}$$

for any $f \in \mathcal{K}_T$.

For any $f, g \in \mathcal{K}_T$, we can apply the isometry property to $f \pm g$ to get

$$E(\,(I(f) \pm I(g))^2\,) = E(\,(I(f \pm g))^2\,) = E\Big(\int_T (f \pm g)^2\,ds\Big).$$

On subtraction and division by 4, we find that

$$E(\,I(f)I(g)\,) = E\Big(\int_0^T f(s)g(s)\,ds\Big)\,.$$

Replacing f by $f1_{(0,t]}$ in the discussion above and using $h_n 1_{(0,t]} \in \mathcal{E}$ rather than h_n, we construct $I_t(f)$ for any $0 \le t \le T$. Taking the limit in $\mathcal{L}^2$, the martingale property $E(I_t(h_n)\,|\,\mathcal{F}_s) = I_s(h_n)$ gives the martingale property of $I_t(f)$, namely,

$$E(I_t(f)\,|\,\mathcal{F}_s) = I_s(f)\,, \quad 0 \le s \le t \le T\,.$$

For the following important result, we make a further assumption about the filtration $(\mathcal{F}_t)$.

Assumption: each $\mathcal{F}_t$ contains all events of probability zero.

Note that by taking complements, it follows that each $\mathcal{F}_t$ contains all events with probability one.

Theorem 5.11. *Let $f \in \mathcal{K}_T$. Then there is a continuous modification of $I_t(f)$, $0 \le t \le T$, that is, there is an adapted process $J_t(\omega)$, $0 \le t \le T$, for which $t \mapsto J_t(\omega)$ is continuous for all $\omega \in \Omega$ and such that for each t, $I_t(f) = J_t$, almost surely.*

Proof. Let (h_n) in $\mathcal{E}$ be a sequence of approximations to f in $\mathcal{K}_T$, that is

$$E\Big(\int_0^T (h_n - f)^2\,ds\Big) \to 0$$

DR.RUPNATHJI(DR.RUPAK NATH)

so that $I(h_n) \to I(f)$ in $\mathcal{L}^2$ (and also $I_t(h_n) \to I_t(f)$ in $\mathcal{L}^2$). It follows that $(I(h_n))$ is a Cauchy sequence in $\mathcal{L}^2$,

$$\|I(h_n) - I(h_m)\|_2^2 = \int_0^T E(h_n - h_m)^2 \, ds \to 0$$

as $n, m \to \infty$. Hence, for any $k \in \mathbb{N}$, there is N_k such that if $n, m \geq N_k$ then

$$\|I(h_n) - I(h_m)\|_2^2 < \tfrac{1}{2^k} \, \tfrac{1}{4^k} \, .$$

If we set $n_k = N_1 + \cdots + N_k$, then clearly $n_{k+1} > n_k \geq N_k$ and

$$\|I(h_{n_{k+1}}) - I(h_{n_k})\|_2^2 < \tfrac{1}{2^k} \, \tfrac{1}{4^k}$$

for $k \in \mathbb{N}$. Now, the map $t \mapsto I_t(h_n)$ is almost surely continuous (because this is true of the map $t \mapsto W_t$) and $I_t(h_n) - I_t(h_m) = I_t(h_n - h_m)$ is an almost surely continuous $\mathcal{L}^2$-martingale. So by Doob's Martingale Inequality, we have

$$P(\sup_{0 \le t \le T} |\, I_t(h_n - h_m)\,| > \varepsilon) \leq \tfrac{1}{\varepsilon^2} \, \|I(h_n - h_m)\|_2^2$$

for any $\varepsilon > 0$. In particular, if we set $\varepsilon = 1/2^k$ and denote by A_k the event

$$A_k = \{ \sup_{0 \le t \le T} |\, I_t(h_{n_{k+1}} - h_{n_k})\,| > \tfrac{1}{2^k} \}$$

then we get

$$P(A_k) \leq (2^k)^2 \, \|I(h_n - h_m)\|_2^2 < (2^k)^2 \, \frac{1}{2^k} \, \frac{1}{4^k} = \frac{1}{2^k} \, .$$

But then $\sum_k P(A_k) < \infty$ and so by the Borel-Cantelli Lemma (Lemma 1.3), it follows that

$$P(\underbrace{A_k \text{ infinitely-often}}_{B}) = 0 \, .$$

For $\omega \in B^c$, we must have

$$\sup_{0 \le t \le T} |\, I_t(h_{n_{k+1}})(\omega) - I_t(h_{n_k})(\omega)\,| \leq \tfrac{1}{2^k}$$

for all $k > k_0$, where k_0 may depend on ω. Hence, for $j > k$

$$\begin{aligned}
\sup_{0 \le t \le T} &|\, I_t(h_{n_j}(\omega) - I_t(h_{n_k})(\omega)\,| \\
&\leq \sup_{0 \le t \le T} |\, I_t(h_{n_j})(\omega) - I_t(h_{n_{j-1}})(\omega)\,| \\
&\qquad + \sup_{0 \le t \le T} |\, I_t(h_{n_{j-1}})(\omega) - I_t(h_{n_{j-2}})(\omega)\,| \\
&\qquad\quad + \ldots + \sup_{0 \le t \le T} |\, I_t(h_{n_{k+1}})(\omega) - I_t(h_{n_k})(\omega)\,|
\end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

$$\leq \frac{1}{2^{j-1}} + \frac{1}{2^{j-2}} + \cdots + \frac{1}{2^k}$$
$$< \frac{2}{2^k}\,.$$

This means that for each $\omega \in B^c$, the sequence of functions $(I_t(h_{n_k})(\omega))$ of t is a Cauchy sequence with respect to the norm $\|\varphi\| = \sup_{0\leq t\leq T} |\varphi(t)|$. In other words, it is uniformly Cauchy and so must converge uniformly on $[0,T]$ to some function of t, say $J_t(\omega)$. Now, for each k, there is a set E_k with $P(E_k) = 1$ such that if $\omega \in E_k$ then $t \mapsto I_t(h_{n_k})(\omega)$ is continuous on $[0,T]$. Set $E = \bigcap_k E_k$, so $P(E) = 1$. Then $P(B^c \cap E) = 1$ and if $\omega \in B^c \cap E$ then $t \mapsto J_t(\omega)$ is continuous on $[0,T]$. We set $J_t(\omega) = 0$ for all t if $\omega \notin B^c \cap E$ which means that $t \mapsto J_t(\omega)$ is continuous for all $\omega \in \Omega$.

However, $I_t(h_{n_k}) \to I_t(f)$ in $\mathcal{L}^2$ and so there is some subsequence $(I_t(h_{n_{k_j}}))$ such that $I_t(h_{n_{k_j}})(\omega) \to I_t(f)(\omega)$ almost surely, say, on S_t with $P(S_t) = 1$. But $I_t(h_{n_{k_j}})(\omega) \to J_t(\omega)$ on $B^c \cap E$ and so $I_t(h_{n_{k_j}})(\omega) \to J_t(\omega)$ on $B^c \cap E \cap S_t$ and therefore $J_t(\omega) = I_t(f)(\omega)$ for $\omega \in B^c \cap E \cap S_t$. Since $P(B^c \cap E \cap S_t) = 1$, we may say that $J_t = I_t(f)$ almost surely.

We still have to show that the process (J_t) is adapted. This is where we use the hypothesis that $\mathcal{F}_t$ contains all events of zero probability. Indeed, by construction, we know that

$$J_t = \lim_{k\to\infty} \underbrace{I_t(h_{n_k})\,1_{B^c \cap E}}_{\mathcal{F}_t\text{-measurable}}$$

and so it follows that J_t is $\mathcal{F}_t$-measurable and the proof is complete. ■

DR.RUPNATHJI(DR.RUPAK NATH)

DR.RUPNATHJI(DR.RUPAK NATH)

Chapter 6

Itô's Formula

We have seen that $E((W_t - W_s)^2) = t - s$ for any $0 \le s \le t$. In particular, if $t-s$ is small, then we see that $(W_t - W_s)^2$ is of first order (on average) rather than second order. This might suggest that we should not expect stochastic calculus to be simply calculus with $\omega \in \Omega$ playing a kind of parametric rôle.

Indeed, the Itô stochastic integral is not an integral in the usual sense. It is constructed via limits of sums in which the integrator points to the future. There is no reason to suppose that there is a stochastic fundamental theorem of calculus. This, of course, makes it difficult to evaluate stochastic integrals. After all, we know, for example, that the derivative of x^3 is $3x^2$ and so the usual fundamental theorem of calculus tell us that the integral of $3x^2$ is x^3. By differentiating many functions, one can consequently build up a list of integrals. The following theorem allows a similar construction for stochastic integrals.

Theorem 6.1 (Itô's Formula). *Let $F(t,x)$ be a function such that the partial derivatives $\partial_t F$ and $\partial_{xx} F$ are continuous. Suppose that $\partial_x F(t, W_t) \in \mathcal{K}_T$. Then*

$$F(T, W_T) = F(0, W_0) + \int_0^T \Big(\partial_t F(t, W_t) + \tfrac{1}{2}\, \partial_{xx} F(t, W_t) \Big)\, dt + \int_0^T \partial_x F(t, W_t)\, dW_t$$

almost surely.

Proof. Suppose first that $F(t,x)$ is such that the partial derivatives $\partial_x F$ and $\partial_{xx} F$ are bounded on $[0,T] \times \mathbb{R}$, say, $|\partial_x F| < C$ and $|\partial_{xx} F| < C$. Let $\Omega_0 \subset \Omega$ be such that $P(\Omega_0) = 1$ and $t \mapsto W_t(\omega)$ is continuous for each $\omega \in \Omega_0$. Fix $\omega \in \Omega_0$. Let $t_j^{(n)} = jT/n$, so that $0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = T$ partitions the interval $[0,T]$ into n equal subintervals. Suppressing the n dependence,

DR.RUPNATHJI(DR.RUPAK NATH)

let $\Delta_j t = t_{j+1}^{(n)} - t_j^{(n)}$ and $\Delta_j W = W_{t_{j+1}}(\omega) - W_{t_j}(\omega)$. Then we have

$$
\begin{aligned}
F(T, W_T(\omega)) &- F(0, W_0(\omega)) \\
&= \sum_{j=0}^{n-1} \big(F(t_{j+1}, W_{t_{j+1}}(\omega)) - F(t_j, W_{t_j}(\omega)) \big) \\
&= \sum_{j=0}^{n-1} \big(F(t_{j+1}, W_{t_{j+1}}(\omega)) - F(t_j, W_{t_{j+1}}(\omega)) \big) \\
&\qquad + \sum_{j=0}^{n-1} \big(F(t_j, W_{t_{j+1}}(\omega)) - F(t_j, W_{t_j}(\omega)) \big) \\
&= \sum_{j=0}^{n-1} \quad \partial_t F(\tau_j, W_{t_{j+1}}(\omega))\, \Delta_j t \\
&\qquad + \sum_{j=0}^{n-1} \big(F(t_j, W_{t_{j+1}}(\omega)) - F(t_j, W_{t_j}(\omega)) \big)\,,
\end{aligned}
$$

for some $\tau_j \in [t_j, t_{j+1}]$, by Taylor's Theorem (to 1^{st} order),

$$
\begin{aligned}
&= \sum_{j=0}^{n-1} \quad \partial_t F(\tau_j, W_{t_{j+1}}(\omega))\, \Delta_j t \\
&\qquad + \sum_{j=0}^{n-1} \big\{ \partial_x F(t_j, W_{t_j}(\omega))\, \Delta_j W + \tfrac{1}{2}\, \partial_{xx} F(t_j, z_j)\, (\Delta_j W)^2 \big\}\,,
\end{aligned}
$$

for some z_j between $W_{t_j}(\omega)$ and $W_{t_{j+1}}(\omega)$, by Taylor's Theorem (to 2^{nd} order),

$$
\equiv \quad \Gamma_1(n) \quad + \quad \Gamma_2(n) \quad + \quad \Gamma_3(n)\,.
$$

We shall consider separately the behaviour of these three terms as $n \to \infty$. Consider first $\Gamma_1(n)$. We write

$$
\begin{aligned}
\partial_t F(\tau_j, W_{t_{j+1}}(\omega))\, \Delta_j t = \big(&\partial_t F(\tau_j, W_{t_{j+1}}(\omega)) - \partial_t F(t_{j+1}, W_{t_{j+1}}(\omega)) \big) \\
&+ \partial_t F(t_{j+1}, W_{t_{j+1}}(\omega))\,.
\end{aligned}
$$

By hypothesis, $\partial_t F(t,x)$ is continuous and so is uniformly continuous on any rectangle in $\mathbb{R}^+ \times \mathbb{R}$. Also $t \mapsto W_t(\omega)$ is continuous and so is bounded on the interval $[0,T]$; say, $|W_t(\omega)| \le M$ on $[0,T]$. In particular, then, $\partial_t F(t,x)$ is uniformly continuous on $[0,T] \times [-M,M]$ and so, for any given $\varepsilon > 0$,

$$
|\, \partial_t F(\tau_j, W_{t_{j+1}}(\omega)) - \partial_t F(t_{j+1}, W_{t_{j+1}}(\omega))\, | < \varepsilon
$$

for sufficiently large n (so that $|\tau_j - t_{j+1}| \le 1/n$ is sufficiently small).

DR.RUPNATHJI(DR.RUPAK NATH)

Hence, for sufficiently large n,

$$\begin{aligned}\Gamma_1(n) &= \sum_{j=0}^{n-1} \big(\partial_t F(\tau_j, W_{t_{j+1}}(\omega)) - \partial_t F(t_{j+1}, W_{t_{j+1}}(\omega)) \big)\, \Delta_j t \\ &\qquad + \sum_{j=0}^{n-1} \partial_t F(t_{j+1}, W_{t_{j+1}}(\omega))\, \Delta_j t\,.\end{aligned}$$

For large n, the first summation on the right hand side is bounded by $\sum_j \varepsilon\, \Delta_j t = \varepsilon T$ and, as $n \to \infty$, the second summation converges to the integral $\int_0^T \partial_s F(s, W_s(\omega))\, ds$. It follows that

$$\Gamma_1(n) \to \int_0^T \partial_s F(s, W_s(\omega))\, ds$$

almost surely as $n \to \infty$.

Next, consider term $\Gamma_2(n)$. This is

$$\sum_{j=0}^{n-1} \partial_x F(t_j, W_{t_j}(\omega))\, \underbrace{\Delta_j W}_{W_{t_{j+1}}(\omega) - W_{t_j}(\omega)} = I(g_n)(\omega)$$

where $g_n \in \mathcal{E}$ is given by

$$g_n(s,\omega) = \sum_{j=0}^{n-1} \partial_x F(t_j, W_{t_j}(\omega))\, 1_{(t_j, t_{j+1}]}(s)\,.$$

Now, the function $t \mapsto \partial_x F(t, W_{t_j}(\omega))$ is continuous (and so uniformly continuous) on $[0,T]$ and $g_n(\cdot,\omega) \to g(\cdot,\omega)$ uniformly on $(0,T]$, where $g(s,\omega) = \partial_x F(s, W_s(\omega))$. It follows that

$$\int_0^T |g_n(s,\omega) - g(s,\omega)|^2\, ds \to 0$$

as $n \to \infty$ for each $\omega \in \Omega_0$. But $\int_0^T |g_n(s,\omega) - g(s,\omega)|^2\, ds \le 4TC^2$ and therefore (since $P(\Omega_0) = 1$)

$$E\Big(\int_0^T |g_n - g|^2\, ds \Big) \to 0$$

as $n \to \infty$, by Lebesgue's Dominated Convergence Theorem.

Applying the Isometry Property, we see that

$$\| I(g_n) - I(g) \|_2 \to 0\,,$$

that is, $\Gamma_2(n) = I(g_n) \to I(g)$ in $\mathcal{L}^2$ and so there is a subsequence (g_{n_k}) such that $I(g_{n_k}) \to I(g)$ almost surely. That is, $\Gamma_2(n_k) \to I(g)$ almost surely.

DR.RUPNATHJI(DR.RUPAK NATH)

We now turn to term $\Gamma_3(n) = \frac{1}{2} \sum_{j=0}^{n-1} \partial_{xx} F(t_j, z_j)(\Delta_j W)^2$. For any $\omega \in \Omega$, write

$$\begin{aligned}\tfrac{1}{2}\,\partial_{xx}F(t_i,z_i)(\Delta_i W)^2 &= \tfrac{1}{2}\,\partial_{xx}F(t_i,W_i(\omega))\big((\Delta_i W)^2-\Delta_i t\big)\\ &\quad+\tfrac{1}{2}\,\partial_{xx}F(t_i,W_i(\omega))\,\Delta_i t\\ &\quad\quad+\tfrac{1}{2}\,\big(\partial_{xx}F(t_i,z_i)-\partial_{xx}F(t_i,W_i(\omega))\big)\,(\Delta_i W)^2\,.\end{aligned}$$

Summing over j, we write $\Gamma_3(n)$ as

$$\Gamma_3(n) \equiv \Phi_1(n) + \Phi_2(n) + \Phi_3(n)$$

where $\Phi_1(n) = \sum_{j=0}^{n-1} \frac{1}{2}\,\partial_{xx}F(t_i, W_i(\omega))\big((\Delta_i W)^2 - \Delta_i t\big)$ etc. As before, we see that for $\omega \in \Omega_0$

$$\Phi_2(n) = \textstyle\sum_{i=0}^{n-1} \tfrac{1}{2}\,\partial_{xx}F(t_i, W_i(\omega))\,\Delta_i t \to \tfrac{1}{2}\displaystyle\int_0^T \partial_{xx}F(s, W_s(\omega))\,ds$$

as $n \to \infty$. Hence $\Phi_2(n_k) \to \frac{1}{2}\int_0^T \partial_{xx}F(s, W_s(\omega))\,ds$ almost surely as $k \to \infty$.

To discuss $\Phi_1(n_k)$, consider

$$\begin{aligned}E(\Phi_1(n_k)^2) &= E((\textstyle\sum_i \alpha_i\,)^2)\\ &= E(\textstyle\sum_{i,j}\alpha_i\,\alpha_j\,)\end{aligned}$$

where $\alpha_i = \partial_{xx}F(t_i, W_i(\omega))\big((\Delta_i W)^2 - \Delta_i t\big)$. By independence, if $i < j$,

$$E(\alpha_i\,\alpha_j) = E(\,\alpha_i\,\partial_{xx}F(t_j, W_j)\,)\;\underbrace{E(\,(\Delta_j W)^2 - \Delta_j t\,)}_{=0}$$

and so

$$\begin{aligned}E(\Phi_1(n_k)^2) &= \sum_i E(\,\alpha_i^2\,)\\ &= \sum_i E\big(\,(\partial_{xx}F(t_i,W_i))^2\,(\,(\Delta_i W)^2-\Delta_i t)^2\,\big)\\ &= \sum_i \underbrace{E\big(\,(\partial_{xx}F(t_i,W_i))^2\,\big)}_{\le C^2}\;E\big(\,(\Delta_i W)^2-\Delta_i t)^2\,\big),\\ &\qquad\qquad\text{by independence,}\\ &\le C^2\;\sum_i E\big(\,(\Delta_i W)^2-\Delta_i t)^2\,\big)\\ &= C^2\;\sum_i E\big(\,(\Delta_i W)^4-2\Delta_i t(\Delta_i W)^2+(\Delta_i t)^2\,\big)\\ &= C^2\;\sum_i\big(\,E\big(\,(\Delta_i W)^4\,\big)-(\Delta_i t)^2\,\big)\end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

$$
\begin{aligned}
&= C^2 \sum_i \left(3(\Delta_i t)^2 - (\Delta_i t)^2 \right) \\
&= C^2\, 2 \sum_i (\Delta_i t)^2 \\
&= C^2\, 2 \sum_i \left(\tfrac{T}{n_k} \right)^2 \\
&= 2\, C^2\, \tfrac{T^2}{n_k} \to 0\,,
\end{aligned}
$$

as $n \to \infty$. Hence $\Phi_1(n_k) \to 0$ in $\mathcal{L}^2$.

Now we consider $\Phi_3(n) = \sum_i (\, \partial_{xx} F(t_i, z_i) - \partial_{xx} F(t_i, W_i(\omega))\,)(\Delta_i W)^2$.

Fix $\omega \in \Omega_0$. Then (just as we have argued for $\partial_x F$), we may say that the function $t \mapsto \partial_{xx} F(t, W_t(\omega))$ is uniformly continuous on $[0, T]$. It follows that for any given $\varepsilon > 0$,

$$
|\partial_{xx} F(t_i, z_i) - \partial_{xx} F(t_i, W_i(\omega))| < \varepsilon
$$

for all $0 \le i \le n-1$, for sufficiently large n. Hence, for all sufficiently large n,

$$
\Big| \sum_{i=0}^{n-1} (\, \partial_{xx} F(t_i, z_i) - \partial_{xx} F(t_i, W_i(\omega))\,)(\Delta_i W)^2 \Big| \le \varepsilon\, S_n \qquad (*)
$$

where $S_n = \sum_{i=0}^{n-1} (\Delta_i W)^2$.

Claim: $S_n \to T$ in $\mathcal{L}^2$ as $n \to \infty$.

To see this, we consider

$$
\begin{aligned}
\|S_n - T\|_2^2 &= E(\, (S_n - T)^2 \,) \\
&= E(\, (S_n - \textstyle\sum_i \Delta_i t)^2 \,) \\
&= E(\, (\textstyle\sum_i \{(\Delta W_i)^2 - \Delta_i t\})^2 \,) \\
&= \textstyle\sum_{i,j} E(\beta_i\, \beta_j)\,, \quad \text{where } \beta_i = (\Delta W_i)^2 - \Delta_i t, \\
&= \textstyle\sum_i E(\beta_i^2) + \sum_{i \ne j} E(\beta_i\, \beta_j) \\
&= \textstyle\sum_i E(\beta_i^2) + \sum_{i \ne j} E(\beta_i)\, E(\beta_j)\,, \quad \text{by independence}, \\
&= \textstyle\sum_i E(\beta_i^2)\,, \quad \text{since } E(\beta_i) = 0, \\
&= \textstyle\sum_i E(\, (\Delta_i W)^4 - 2\Delta_i t\, (\Delta_i W)^2 + (\Delta_i t)^2 \,) \\
&= \textstyle\sum_i 2\, (\Delta_i t)^2 \\
&= 2\, \tfrac{T^2}{n} \\
&\to 0
\end{aligned}
$$

as $n \to \infty$, which establishes the claim.

DR.RUPNATHJI(DR. RUPAK NATH)

It follows that $S_{n_k} \to T$ in $\mathcal{L}^2$ and so there is a subsequence such that both $\Phi_1(n_m) \to 0$ almost surely and $S_{n_m} \to T$ almost surely. It follows from $(*)$ that

$$\Phi_3(n_m) \to 0$$

almost surely as $m \to \infty$.

Combining these results, we see that

$$\begin{aligned}\Gamma_1(n_m) + \Gamma_2(n_m) + \Gamma_3(n_m) \\ \to \int_0^T \partial_s F(s, W_s(\omega))\, ds + \int_0^T \partial_x F(s, W_s)\, dW_s \\ + \tfrac{1}{2} \int_0^T \partial_{xx} F(s, W_s(\omega))\, ds\end{aligned}$$

almost surely, which concludes the proof for the case when $\partial_x F$ and $\partial_{xx} F$ are both bounded. To remove this restriction, let $\varphi_n : \mathbb{R} \to \mathbb{R}$ be a smooth function such that $\varphi_n(x) = 1$ for $|x| \le n + 1$ and $\varphi_n(x) = 0$ for $|x| \ge n + 2$. Set $F_n(t, x) = \varphi_n(x)\, F(t, x)$. Then $\partial_x F_n$ and $\partial_{xx} F_n$ are continuous and bounded on $[0, T] \times \mathbb{R}$. Moreover, $F_n = F$, $\partial_t F_n = \partial_t F$, $\partial_x F_n = \partial_x F$ and $\partial_{xx} F_n = \partial_{xx} F$ whenever $|x| \le n$.

We can apply the previous argument to deduce that for each n there is some set $B_n \subset \Omega$ with $P(B_n) = 1$ such that

$$\begin{aligned}F_n(T, W_T) = F_n(0, W_0) + \int_0^T \Big(\partial_t F_n(t, W_t) + \tfrac{1}{2}\, \partial_{xx} F_n(t, W_t) \Big)\, dt \\ + \int_0^T \partial_x F_n(t, W_t)\, dW_t \qquad (**)\end{aligned}$$

for $\omega \in B_n$.

Let $A = \Omega_0 \cap \big(\bigcap_n B_n \big)$ so that $P(A) = 1$. Fix $\omega \in A$. Then $\omega \in \Omega_0$ and so $t \mapsto W_t(\omega)$ is continuous. It follows that there is some $N \in \mathbb{N}$ such that $|W_t(\omega)| < N$ for all $t \in [0, T]$ and so $F_N(t, W_t(\omega)) = F(t, W_t(\omega))$ for $t \in [0, T]$ and the same remark applies to the partial derivatives of F_N and F. But $\omega \in B_N$ and so $(**)$ holds with F_n replaced by F_N which in turn means that $(**)$ holds with now F_N replaced by F – simply because F_N and F (and the derivatives) agree for such ω and all t in the relevant range.

We conclude that

$$\begin{aligned}F(T, W_T) = F(0, W_0) + \int_0^T \Big(\partial_t F(t, W_t) + \tfrac{1}{2}\, \partial_{xx} F(t, W_t) \Big)\, dt \\ + \int_0^T \partial_x F(t, W_t)\, dW_t\end{aligned}$$

for all $\omega \in A$ and the proof is complete. ■

DR.RUPNATHJI(DR. RUPAK NATH)

Example 6.2. Let $F(s,x) = s\,x$. Then we find that

$$F(t, W_t) - F(0, W_0) = t\,W_t - 0 = \int_0^t W_s\, ds + \int_0^t s\, dW_s\,,$$

that is,

$$\int_0^t s\, dW_s = t\,W_t - \int_0^t W_s\, ds\,.$$

Example 6.3. Let $F(t,x) = x^2$. Then Itô's Formula tells us that

$$W_t^2 = W_0^2 + \int_0^t \tfrac{1}{2}\, 2\, ds + \int_0^t 2W_s\, dW_s\,,$$

that is,

$$W_t^2 = t + 2\int_0^t W_s\, dW_s\,.$$

This can also be expressed as

$$\int_0^t W_s\, dW_s = \tfrac{1}{2}\, W_t^2 - \tfrac{1}{2}\, t\,.$$

Example 6.4. Let $F(t,x) = \sin x$, so we find that $\partial_t F = 0$, $\partial_x F = \cos x$ and $\partial_{xx} F = -\sin x$. Applying Itô's Formula, we get

$$F(t, W_t) - F(0, W_0) = \sin W_t = \int_0^t \cos(W_s)\, dW_s - \tfrac{1}{2}\int_0^t \sin(W_s)\, ds$$

or

$$\int_0^t \cos(W_s)\, dW_s = \sin W_t + \tfrac{1}{2}\int_0^t \sin(W_s)\, ds\,.$$

Example 6.5. Suppose that the function $F(t,x)$ obeys $\partial_t F = -\frac{1}{2}\,\partial_{xx} F$. Then if $\partial_x F(t, W_t) \in \mathcal{K}_T$, Itô's formula gives

$$F(T, W_T) = F(0, W_0) + \int_0^T \Big(\underbrace{\partial_t F(t, W_t) + \tfrac{1}{2}\,\partial_{xx} F(t, W_t)}_{=0} \Big)\, dt + \underbrace{\int_0^T \partial_x F(t, W_t)\, dW_t}_{\text{martingale}}$$

Taking $F(t,x) = e^{\alpha x - \frac{1}{2} t \alpha^2}$, we may say that $e^{\alpha W_t - \frac{1}{2} t\alpha^2}$ is a martingale.

DR.RUPNATHJI(DR.RUPAK NATH)

Itô's formula also holds when W_t is replaced by the somewhat more general processes of the form

$$X_t = X_0 + \int_0^t u(s,\omega)\,ds + \int_0^t v(s,\omega)\,dW_s$$

where u and v obey $P(\int_0^t |u|^2\,ds < \infty) = 1$ and similarly for v. One has

$$\begin{aligned} F(t,X_t) - F(0,X_0) = \int_0^t \Big\{\, \partial_s F(s,X_s) + u\,\partial_x F(s,X_s) \\ + \tfrac{1}{2}\,v^2\,\partial_{xx}F(s,X_s)\Big\}\,ds \\ + \int_0^t v(s,\omega)\,\partial_x F(s,X_s)\,dW_s\,. \end{aligned}$$

In symbolic differential form, this can be written as

$$dF(t,X_t) = \partial_t F\,dt + u\,\partial_x F\,dt + \tfrac{1}{2}v^2\partial_{xx}F\,dt + v\,\partial_x F\,dW\,. \qquad (*)$$

On the other hand, on the grounds that $dt\,dX$ and $dt\,dt$ are negligible, we might also write

$$dF(t,X) = \partial_t F\,dt + \partial_x F\,dX + \tfrac{1}{2}\,\partial_{xx}F\,dX\,dX\,. \qquad (**)$$

However, in the same spirit, we may write

$$dX = u\,dt + v\,dW$$

which leads to

$$dX\,dX = u^2\underbrace{dt^2}_{=0} + 2uv\underbrace{dt\,dW}_{=0} + v^2\,dW\,dW\,.$$

But we know that $E((\Delta W)^2) = \Delta t$ and so if we replace $dW\,dW$ by dt and substitute these expressions for dX and $dX\,dX$ into $(**)$, then we recover the expression $(*)$.

It would appear that standard practice is to manipulate such symbolic differential expressions (rather than the proper integral equation form) in conjunction with the following Itô table:

Itô table

$$dt^2 = 0$$
$$dt\,dW = 0$$
$$dW^2 = dt$$

DR.RUPNATHJI(DR.RUPAK NATH)

Example 6.6. Consider $Z_t = e^{\int_0^t g\,dW - \frac{1}{2}\int_0^t g^2\,ds}$. Let

$$X_t = X_0 + \int_0^t g\,dW - \int_0^t \tfrac{1}{2}\,g^2 ds$$

so that $dX = -\frac{1}{2}\,g^2\,ds + g\,dW$. Then, with $F(t,x) = e^x$, we have $\partial_t F = 0$, $\partial_x F = e^x$ and $\partial_{xx} F = e^x$ so that

$$\begin{aligned} dF &= \partial_t F(X)\,dt + \partial_x F(X)\,dX + \tfrac{1}{2}\,\partial_{xx}F(X)\,dX\,dX \\ &= 0 + e^X\left(-\tfrac{1}{2}\,g^2\,dt + g\,dW\right) + \tfrac{1}{2}\,e^X \underbrace{dX\,dX}_{g^2\,dt} \\ &= e^X g\,dW\,. \end{aligned}$$

We deduce that $F(X_t) = e^{X_t} = Z_t$ obeys $dZ = e^X\,g\,dW = Z\,g\,dW$, so that

$$Z_t = Z_0 + \int_0^t g\,Z_s\,dW_s\,.$$

It can be shown that this process is a martingale for suitable g – such a condition is given by the requirement that $E(e^{\frac{1}{2}\int_0^T g^2\,ds}) < \infty$, the so-called Novikov condition.

Example 6.7. Suppose that $dX = u\,dt + dW$, $M_t = e^{-\int_0^t u\,dW - \frac{1}{2}\int_0^t u^2\,ds}$ and consider the process $X_t\,M_t$. Then

$$dY_t = X_t\,dM_t + M_t\,dY_t + dX_t\,dM_t\,.$$

The question now is what is dM_t? Let $dV_t = -\frac{1}{2}u^2\,dt - u\,dW_t$ so that $M_t = e^{V_t}$. Applying Itô's formula (equation $(*)$ above) with $F(t,x) = e^x$, (so $\partial_t F = 0$ and $\partial_x F = \partial_{xx} F = e^x$ and $F(V_t) = M_t$), we obtain

$$\begin{aligned} dM_t = dF(V_t) &= \underbrace{\left(0 - \tfrac{1}{2}\,u^2 e^V + \tfrac{1}{2}\,u^2 e^V\right)}_{=0} ds - u\,e^V\,dW \\ &= -u\,e^V\,dW_t\,. \end{aligned}$$

So

$$\begin{aligned} dY_t &= -X_t\,u\,e^{V_t}\,dW_t + M_t(u\,dt + dW_t) - u\,e^{V_t}\,dW_t(u\,dt + dW_t) \\ &= -u\,X_t\,M_t\,dW_t + u\,M_t\,dt + M_t\,dW_t - u\,e^{V_t}\,dt \\ &= (-u\,X_t\,M_t + M_t)\,dW_t\,, \quad \text{since } M_t = e^{V_t}, \end{aligned}$$

or, in integral form,

$$Y_t = Y_0 + \int_0^t (M_s - u\,X_s\,M_s)\,dW_s$$

which is a martingale.

DR.RUPNATHJI(DR.RUPAK NATH)

Remark 6.8. There is also an m-dimensional version of Itô's formula. Let $(W^{(1)}, W^{(2)}, \dots, W^{(m)})$ be an m-dimensional Wiener process (that is, the $W^{(j)}$ are independent Wiener processes) and suppose that

$$\begin{aligned} dX_1 &= u_1\,dt + v_{11}\,dW^{(1)} + \cdots + v_{1m}\,dW^{(m)} \\ &\;\vdots \\ dX_n &= u_n\,dt + v_{n1}\,dW^{(1)} + \cdots + v_{nm}\,dW^{(m)} \end{aligned}$$

and let $Y_t(\omega) = F(t, X_1(\omega), \dots, X_n(\omega))$. Then

$$dY = \partial_t F(t,X)\,dt + \sum_{i=1}^{n} \partial_{x_i} F(t,X)\,dX_i + \tfrac{1}{2} \sum_{i,j} \partial_{x_i x_j}\,dX_i\,dX_j$$

where the Itô table is enhanced by the extra rule that $dW^{(i)}\,dW^{(j)} = \delta_{ij}\,dt$.

DR.RUPNATHJI(DR.RUPAK NATH)

Stochastic Differential Equations

The purpose of this section is to illustrate how Itô's formula can be used to obtain solutions to so-called stochastic differential equations. These are usually obtained from physical considerations in essentially the same way as are "usual" differential equations but when a random influence is to be taken into account. Typically, one might wish to consider the behaviour of a certain quantity as time progresses. Then one theorizes that the change of such a quantity over a small period of time is approximately given by such and such terms (depending on the physical problem under consideration) plus some extra factor which is supposed to take into account some kind of random influence.

Example 6.9. Consider the "stochastic differential equation"

$$dX_t = \mu\,X_t\,dt + \sigma\,X_t\,dW_t \qquad (*)$$

with $X_0 = x_0 > 0$. Here, μ and σ are constants with $\sigma > 0$. The quantity X_t is the object under investigation and the second term on the right is supposed to represent some random input to the change dX_t. Of course, mathematically, this is just a convenient shorthand for the corresponding integral equation. Such an equation has been used in financial mathematics (to model a risky asset).

To solve this stochastic differential equation, let us seek a solution of the form $X_t = f(t, W_t)$ for some suitable function $f(t,x)$. According to Itô's formula, such a process would satisfy

$$dX_t = df = \big\{ f_t(t,W_t) + \tfrac{1}{2} f_{xx}(t,W_t) \big\}\,dt + f_x(t,W_t)\,dW_t \qquad (**)$$

Comparing coefficients in $(*)$ and $(**)$, we find

$$\mu\, f(t, W_t) = f_t(t, W_t) + \tfrac{1}{2}\, f_{xx}(t, W_t)$$
$$\sigma\, f(t, W_t) = f_x(t, W_t)\,.$$

So we try to find a function $f(t, x)$ satisfying

$$\mu\, f(t, x) = f_t(t, x) + \tfrac{1}{2}\, f_{xx}(t, x) \qquad \text{(i)}$$
$$\sigma\, f(t, x) = f_x(t, x) \qquad \text{(ii)}$$

Equation (ii) leads to $f(t, x) = e^{\sigma x} C(t)$. Substitution into equation (i) requires $C(t) = e^{t(\mu - \frac{1}{2}\sigma(2)}\, C(0)$ which leads to the solution

$$f(t, x) = C(0)\, e^{\sigma x + t(\mu - \frac{1}{2}\sigma^2)}\,.$$

But then $X_0 = f(0, W_0) = f(0, 0) = C(0)$ so that $C(0) = x_0$ and we have found the required solution

$$X_t = x_0\, e^{\sigma\, W_t + t(\mu - \frac{1}{2}\sigma^2)}\,.$$

From this, we can calculate the expectation $E(X_t)$. We find

$$E(X_t) = x_0\, e^{t\mu}\, \underbrace{E\big(e^{\sigma\, W_t - \frac{1}{2} t\sigma^2)}\big)}_{=1}$$
$$= x_0\, e^{t\mu}$$

and so we see that if $\mu > 0$ then the expectation $E(X_t)$ grows exponentially as $t \to \infty$.

However, we can write

$$\frac{W_n}{n} = \frac{(W_n - W_{n-1}) + (W_{n-1} - W_{n-2}) + \cdots + (W_1 - W_0)}{n}$$
$$= \frac{Z_1 + Z_2 + \cdots + Z_n}{n}$$

where $Z_1, \ldots, Z_n$ are independent, identically distributed random variables with mean zero (in fact, standard normal). So we can apply the Law of Large Numbers to deduce that

$$P\big(\tfrac{W_n}{n} \to 0 \text{ as } n \to \infty\big) = 1\,.$$

Now, we have

$$X_n = x_0\, e^{\sigma\, W_n + n(\mu - \frac{1}{2}\sigma^2)}$$
$$= x_0\, e^{\sigma n\,(\frac{W_n}{n} - (\frac{1}{2}\sigma^2 - \mu))}\,.$$

If $\sigma^2 > 2\mu$, the right hand side above $\to 0$ almost surely as $n \to \infty$.

We see that X_n has the somewhat surprising behaviour that $X_n \to 0$ with probability one, whereas $E(X_n) \to \infty$ (exponentially quickly).

DR.RUPNATHJI(DR.RUPAK NATH)

Example 6.10. Ornstein-Uhlenbeck Process

Consider

$$dX_t = -\alpha\, X_t\, dt + \sigma\, dW_t$$

with $X_0 = x_0$, where α and σ are positive constants. We look for a solution of the form $X_t = g(t)\, Y_t$ where $dY_t = h(t)\, dW_t$. In differential form, this becomes (using the Itô table)

$$\begin{aligned} dX_t &= g\, dY_t + dg\; Y_t + dg\, dY_t \\ &= gh\, dW_t + g' Y_t\, dt + 0\,. \end{aligned}$$

Comparing this with the original equation, we require

$$\begin{aligned} g'Y &= -\alpha g\, Y \\ gh &= \sigma\,. \end{aligned}$$

The first of these equations leads to the solution $g(t) = C\, e^{-\alpha t}$. This gives $h = \sigma/g = \sigma e^{\alpha t}/C$ and so $dY = \frac{\sigma}{C}\, e^{\alpha t}\, dW$, or

$$Y_t = Y_0 + \tfrac{\sigma}{C} \int_0^t e^{\alpha s}\, dW_s\,.$$

Hence

$$\begin{aligned} X_t &= C\, e^{-\alpha t}\Big(Y_0 + \tfrac{\sigma}{C} \int_0^t e^{\alpha s}\, dW_s\Big) \\ &= e^{-\alpha t}\Big(C\, Y_0 + \sigma \int_0^t e^{\alpha s}\, dW_s\Big)\,. \end{aligned}$$

When $t = 0$, $X_0 = x_0$ and so $x_0 = CY_0$ and we finally obtain the solution

$$X_t = x_0\, e^{-\alpha t} + \sigma \int_0^t e^{\alpha(s-t)}\, dW_s\,.$$

Example 6.11. Let $(W^{(1)}, W^{(2)})$ be a two-dimensional Wiener process and let $M_t = \varphi(W_t^{(1)}, W_t^{(2)})$. Let $f(t,x,y) = \varphi(x,y)$ so that $\partial_t f = 0$. Itô's formula then says that

$$\begin{aligned} dM_t = df &= \partial_x \varphi(W_t^{(1)}, W_t^{(2)})\, dW_t^{(1)} + \partial_y \varphi(W_t^{(1)}, W_t^{(2)})\, dW_t^{(2)} \\ &\qquad + \tfrac{1}{2}\big\{\, \partial_x^2 \varphi(W_t^{(1)}, W_t^{(2)}) + \partial_y^2 \varphi(W_t^{(1)}, W_t^{(2)})\,\big\}\, dt\,. \end{aligned}$$

In particular, if φ is harmonic (so $\partial_x^2 \varphi + \partial_y^2 \varphi = 0$) then

$$dM_t = \partial_x \varphi(W_t^{(1)}, W_t^{(2)})\, dW_t^{(1)} + \partial_y \varphi(W_t^{(1)}, W_t^{(2)})\, dW_t^{(2)}$$

and so $M_t = \varphi(W_t^{(1)}, W_t^{(2)})$ is a martingale.

DR.RUPNATHJI(DR.RUPAK NATH)

Feynman-Kac Formula

Consider the diffusion equation

$$u_t(t,x) = \tfrac{1}{2}\, u_{xx}(t,x)$$

with $u(0,x) = f(x)$ (where f is well-behaved). The solution can be written as

$$u(t,x) = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} f(v)\, e^{(v-x)^2/2t}\, dv\,.$$

However, the right hand side here is equal to the expectation $E(\, f(x+W_t)\,)$ and $x + W_t$ is a Wiener process started at x, rather than at 0.

Theorem 6.12 (Feynman-Kac). *Let $q : \mathbb{R} \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be bounded. Then (the unique) solution to the initial-value problem*

$$\partial_t u(t,x) = \tfrac{1}{2}\, \partial_{xx} u(t,x) + q(x)\, u(t,x)\,,$$

with $u(0,x) = f(x)$, has the representation

$$u(t,x) = E\Big(\, f(x+W_t)\, e^{\int_0^t q(x+W_s)\, ds} \,\Big)\,.$$

Proof. Consider Itô's formula applied to the function $f(s,y) = u(t-s, x+y)$. In this case $\partial_s f = -\partial_1 u = -u_1$, $\partial_y f = \partial_2 u = u_2$, $\partial_{yy} f = \partial_{22} u = u_{22}$ and so we get

$$\begin{aligned} df(s,W_s) &= -u_1(t-s, x+W_s)\, ds \\ &\qquad + \tfrac{1}{2}\, u_{22}(t-s, x+W_s)\, \underbrace{dW_s\, dW_s}_{ds} + u_2(t-s, x+W_s)\, dW_s \\ &= -q(x+W_s)\, u(t-s, x+W_s)\, ds + u_2(t-s, x+W_s)\, dW_s \end{aligned}$$

where we have made use of the partial differential equation satisfied by u.

For $0 \le s \le t$, set

$$M_s = u(t-s, x+W_s)\, e^{\int_0^s q(x+W_v)\, dv}$$

Applying the rule $d(XY) = (dX)\,Y + X\,dY + dX\,dY$ together with the Itô table, we get

$$\begin{aligned} dM_s &= df\, e^{\int_0^s q(x+W_v)\, dv} + f\, d\big(\, e^{\int_0^s q(x+W_v)\, dv} \,\big) + df\, d\big(\, e^{\int_0^s q(x+W_v)\, dv} \,\big) \\ &= -q(x+W_s)\, u(t-s, x+W_s)\, e^{\int_0^s q(x+W_v)\, dv}\, ds \\ &\qquad + u_2(t-s, x+W_s)\, e^{\int_0^s q(x+W_v)\, dv}\, dW_s \\ &\qquad\quad + f\, q(x+W_s)\, e^{\int_0^s q(x+W_v)\, dv}\, ds + 0 \\ &= u_2(t-s, x+W_s)\, e^{\int_0^s q(x+W_v)\, dv}\, dW_s\,. \end{aligned}$$

DR.RUPNATHJI(DR.RUPAK NATH)

It follows that

$$M_\tau = M_0 + \int_0^\tau u_2(t-s, x+W_s)\, e^{\int_0^s q(x+W_v)\,dv}\, dW_s$$

and so M_τ is a martingale and, in particular, $E(M_t) = E(M_0)$. However, by construction, $M_0 = u(t,x)$ almost surely and so $E(M_0) = u(t,x)$ and

$$E(M_t) = E\Big(\underbrace{u(0, x+W_t)}_{=f(x+W_t)}\, e^{\int_0^t q(x+W_v)\,dv}\Big)$$

by the initial condition. The result follows. ■

Martingale Representation Theorems

Let $(\mathcal{F}_s)$ be the standard Wiener process filtration (the minimal filtration generated by the W_t's).

Theorem 6.13 ($\mathcal{L}^2$-martingale representation theorem). *Let $X \in \mathcal{L}^2(\Omega, \mathcal{F}_T, P)$. Then there is $f \in \mathcal{K}_T$ (unique almost surely) such that*

$$X = \alpha + \int_0^T f(s,\omega)\, dW_s$$

where $\alpha = E(X)$.

If $(X_t)_{t \le T}$ is an $\mathcal{L}^2$-martingale, then there is $f \in \mathcal{K}_T$ (unique almost surely) such that

$$X_t = X_0 + \int_0^t f\, dW_s$$

for all $0 \le t \le T$.

We will not prove this here but will just make some remarks. To say that X in the Hilbert space $\mathcal{L}^2(\mathcal{F}_T)$ obeys $E(X) = 0$ is to say that X is orthogonal to 1, so the theorem says that every element of $\mathcal{L}^2(\mathcal{F}_T)$ orthogonal to 1 is a stochastic integral. The uniqueness of $f \in \mathcal{K}_T$ in the theorem is a consequence of the isometry property. Moreover, if we denote $\int_0^T f\, dW$ by $I(f)$, then the isometry property tells us that $f \mapsto I(f)$ is an isometric isomorphism between $\mathcal{K}_T$ and the subspace of $\mathcal{L}^2(\mathcal{F}_T)$ orthogonal to 1.

Note, further, that the second part of the theorem follows from the first part by setting $X = X_T$. Indeed, by the first part, $X_T = \alpha + \int_0^T f\, dW$ and so

$$X_t = E(X_T \,|\, \mathcal{F}_t) = \alpha + \int_0^t f\, dW_s$$

for $0 \le t \le T$. Evidently, $\alpha = X_0 = E(X_T)$.

DR.RUPNATHJI(DR.RUPAK NATH)

Recall that if $dX_t = u\,dt + v\,dW_t$ then Itô's formula is

$$df(t,X_t) = \partial_t f(t,X_t)\,dt + \tfrac{1}{2}\,\partial_{xx} f(t,X_t)\,dX_t\,dX_t + \partial_x f(t,X_t)\,dX_t$$

where, according to the Itô table, $dX_t\,dX_t = v^2\,dt$.

Set $u = 0$ and let $f(t,x) = x^2$. Then $\partial_t f = 0$, $\partial_x f = 2x$ and $\partial_{xx} f = 2$ so that

$$d(X_t^2) = 0 + \underbrace{dX_t\,dX_t}_{=v^2\,dt} + 2X\,dX_t\,.$$

That is,

$$X_t^2 = X_0^2 + \int_0^t 2X_s\,dX_s + \int_0^t v^2\,ds\,.$$

But now $dX_s = u\,ds + v\,dW_s = 0 + v\,dW_s$ and X_s is a martingale and

$$X_t^2 = \underbrace{X_0^2 + \int_0^t 2X_s v\,dW_s}_{\text{martingale part, } Z_t} \quad + \quad \underbrace{\int_0^t v^2\,ds}_{\text{increasing process part, } A_t}$$

This is the Doob-Meyer decomposition of the submartingale (X_t^2).

Note that $X_t^2 - \int_0^t v^2\,ds$ is the martingale part. If $v = 1$, then $\int_0^t v^2\,ds = t$ and so this says that $X_t^2 - t$ is the martingale part. But if $v = 1$, we have $dX = v\,dW = dW$ so that $X_t = X_0 + W_t$ and in this case (X_t) is a Wiener process.

The converse is true (Lévy's characterization) as we show next using the following result.

Proposition 6.14. *Let $\mathcal{G} \subset \mathcal{F}$ be σ-algebras. Suppose that X is $\mathcal{F}$-measurable and for each $\theta \in \mathbb{R}$*

$$E(e^{i\theta X} \mid \mathcal{G}) = e^{-\theta^2\sigma^2/2}$$

almost surely. Then X is independent of $\mathcal{G}$ and X has a normal distribution with mean 0 and variance σ^2.

Proof. For any $A \in \mathcal{G}$,

$$\begin{aligned} E(e^{i\theta X}\,1_A) &= E(e^{-\theta^2\sigma^2/2}\,1_A) \\ &= e^{-\theta^2\sigma^2/2}\,E(1_A) \\ &= e^{-\theta^2\sigma^2/2}\,P(A)\,. \end{aligned}$$

In particular, with $A = \Omega$, we see that

$$E(e^{i\theta X}) = e^{-\theta^2\sigma^2/2}$$

and so it follows that X is normal with mean zero and variance σ^2.

To show that X is independent of $\mathcal{G}$, suppose that $A \in \mathcal{G}$ with $P(A) > 0$. Define Q on $\mathcal{F}$ by

$$Q(B) = \frac{P(B \cap A)}{P(A)} = \frac{E(1_A\,1_B)}{P(A)}\,.$$

DR.RUPNATHJI(DR.RUPAK NATH)

Then Q is a probability measure on $\mathcal{F}$ and

$$\begin{aligned} E_Q(e^{i\theta X}) &= \int_\Omega e^{i\theta X}\, dQ \\ &= \int_\Omega e^{i\theta X}\, \frac{dP}{P(A)} \\ &= \frac{E(e^{i\theta X}\, 1_A)}{P(A)} \\ &= e^{-\theta^2\sigma^2/2} \end{aligned}$$

from the above. It follows that X also has a normal distribution, with mean zero and variance σ^2 with respect to Q. Hence, if Φ denotes the standard normal distribution function, then for any $x \in \mathbb{R}$ we have

$$\begin{aligned} Q(X \le x) = \Phi(x/\sigma) &\implies \frac{P(\{\, X \le x \,\} \cap A)}{P(A)} = \Phi(x/\sigma) = P(X \le x) \\ &\implies P(\{\, X \le x \,\} \cap A) = P(X \le x)\, P(A) \end{aligned}$$

for any $A \in \mathcal{G}$ with $P(A) \ne 0$. This trivially also holds for $A \in \mathcal{G}$ with $P(A) = 0$ and so we may conclude that this holds for all $A \in \mathcal{G}$ which means that X is independent of $\mathcal{G}$. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

Using this, we can now discuss Lévy's theorem where as before, we work with the minimal Wiener filtration.

Theorem 6.15 (Lévy's characterization of the Wiener process). *Let $(X_t)_{t \le T}$ be a (continuous) $\mathcal{L}^2$-martingale with $X_0 = 0$ and such that $X_t^2 - t$ is an $\mathcal{L}^1$-martingale. Then (X_t) is a Wiener process.*

Proof. By the $\mathcal{L}^2$-martingale representation theorem, there is $\beta \in \mathcal{K}_T$ such that

$$X_t = \int_0^t \beta(s, \omega)\, dW_s\,.$$

We have seen that $d(X_t^2) = dZ_t + \beta^2\, dt$ where (Z_t) is a martingale and so by hypothesis and the uniqueness of the Doob-Meyer decomposition $d(X^2) = dZ + dA$, it follows that $\beta^2\, dt = dt$ (the increasing part is $A_t = t$).

Let $f(t, x) = e^{i\theta x}$ so that $\partial_t f = 0$, $\partial_x f = i\theta\, e^{i\theta x}$ and $\partial_{xx} f = -\theta^2\, e^{i\theta x}$ and apply Itô's formula to $f(t, X_t)$, where now $dX = \beta\, dW$, to get

$$d(e^{i\theta X_t}) = -\tfrac{1}{2}\,\theta^2\, e^{i\theta X_t} \underbrace{dX_t\, dX_t}_{\beta^2\, dt = dA_t = dt} + i\theta\, e^{i\theta X_t} \underbrace{dX_t}_{\beta\, dW_t}\,.$$

So

$$e^{i\theta X_t} - e^{i\theta X_s} = -\tfrac{1}{2}\,\theta^2 \int_s^t e^{i\theta X_u}\,du + i\theta \int_s^t e^{i\theta X_u}\,\beta\,dW_u$$

and therefore

$$e^{i\theta(X_t - X_s)} - 1 = -\tfrac{1}{2}\,\theta^2 \int_s^t e^{i\theta(X_u - X_s)}\,du + i\theta \int_s^t e^{i\theta(X_u - X_s)}\,\beta\,dW_u \qquad (*)$$

Let $Y_t = \int_0^t e^{i\theta X_u}\beta\,dW_u$. Then (Y_t) is a martingale and the second term on the right hand side of equation $(*)$ is equal to $i\theta\,e^{-i\theta X_s}\,(Y_t - Y_s)$. Taking the conditional expectation with respect to $\mathcal{F}_s$, this term then drops out and we find that

$$E(e^{i\theta(X_t - X_s)}\,|\,\mathcal{F}_s) - 1 = -\tfrac{1}{2}\,\theta^2 \int_s^t E(e^{i\theta(X_u - X_s)}\,|\,\mathcal{F}_s)\,du\,.$$

Letting $\varphi(t) = E(e^{i\theta(X_t - X_s)}\,|\,\mathcal{F}_s)$, we have

$$\begin{aligned} \varphi(t) - 1 &= -\tfrac{1}{2}\,\theta^2 \int_s^t \varphi(u)\,du \\ \implies\ \varphi'(t) &= -\tfrac{1}{2}\,\theta^2\,\varphi(t) \text{ and } \varphi(s) = 1 \\ \implies\ \varphi(t) &= C\,e^{-(t-s)\theta^2/2} \end{aligned}$$

for $t \ge s$. The result now follows from the proposition. ■

DR.RUPNATHJI(DR.RUPAK NATH)

Cameron-Martin, Girsanov change of measure

The process $W_t + \mu t$ is a Wiener process "with drift" and is not a martingale. However, it becomes a martingale if the underlying probability measure is changed. To discuss this, we consider the process

$$X_t = W_t + \int_0^t \mu(s,\omega)\,ds$$

where μ is bounded and adapted on $[0,T]$. Let $M_t = e^{-\int_0^t \mu\,dW - \frac{1}{2}\int_0^t \mu^2\,ds}$.

Claim: $(M_t)_{t \le T}$ is a martingale.

Proof. Apply Itô's formula to $f(t,Y)$ where $f(t,x) = e^x$ and where Y_t is the process $Y_t = -\int_0^t \mu\,dW - \frac{1}{2}\int_0^t \mu^2\,ds$ so $M_t = e^{Y_t}$ and $dY = -\mu\,dW - \frac{1}{2}\,\mu^2\,ds$. We see that $\partial_t f = 0$ and $\partial_x f = \partial_{xx} f = e^x$ and therefore

$$\begin{aligned} dM = d(e^Y) &= 0 + \tfrac{1}{2}\,e^Y\,dY\,dY + e^Y\,dY \\ &= \tfrac{1}{2}\,e^Y\,\mu^2\,ds + e^Y(-\mu\,dW - \tfrac{1}{2}\,\mu^2\,ds) \\ &= -\mu e^Y\,dW \end{aligned}$$

and so $M_t = e^{Y_t}$ is a martingale, as claimed. ■

Claim: $(X_t\, M_t)_{t \le T}$ is a martingale.

Proof. Itô's product formula gives

$$\begin{aligned} d(X\,M) &= (dX)\,M + X\,dM + dX\,dM \\ &= M(dW + \mu\,ds) + X(-\mu M\,dW) + (-\mu M)\,ds \\ &= (M - \mu X M)\,dW \end{aligned}$$

and therefore

$$X_t\,M_t = X_0\,M_0 + \int_0^t (M - \mu X M)\,dW$$

is a martingale, as required. ∎

Theorem 6.16. *Let Q be the measure on $\mathcal{F}$ given by $Q(A) = E(1_A\,M_T)$ for $A \in \mathcal{F}$. Then Q is a probability measure and $(X_t)_{t \le T}$ is a martingale with respect to Q.*

Proof. Since Q is the map $A \mapsto Q(A) = \int_A M_T\,dP$ on $\mathcal{F}$, it is clearly a measure. Furthermore, $Q(\Omega) = E(M_T) = E(M_0)$ since M_t is a martingale. But $E(M_0) = 1$ and so we see that Q is a probability measure on $\mathcal{F}$.

We can write $Q(A) = \int_\Omega 1_A\,M_T\,dP$ and therefore $\int_\Omega f\,dQ = \int_\Omega f\,M_T\,dP$ for Q-integrable f (symbolically, $dQ = M_T\,dP$.) (Note, incidentally, that $Q(A) = 0$ if and only if $P(A) = 0$.)

To show that X_t is a martingale with respect to Q, let $0 \le s \le t \le T$ and let $A \in \mathcal{F}_s$. Then, using the facts shown above that M_t and $X_t\,M_t$ are martingales, we find that

$$\begin{aligned} \int_A X_t\,dQ &= \int_A X_t\,M_T\,dP \\ &= \int_A E(X_t\,M_T \,|\, \mathcal{F}_t)\,dP\,, \text{ since } A \in \mathcal{F}_s \subset \mathcal{F}_t, \\ &= \int_A X_t\,M_t\,dP \\ &= \int_A X_s\,M_s\,dP \\ &= \int_A E(X_s\,M_T \,|\, \mathcal{F}_s)\,dP \\ &= \int_A X_s\,M_T\,dP \\ &= \int_A X_s\,dQ \end{aligned}$$

and so

$$E^Q(X_t|\mathcal{F}_s) = X_s$$

and the proof is complete. ∎

DR.RUPNATHJI(DR.RUPAK NATH)

Textbooks on Probability and Stochastic Analysis

There are now many books available covering the highly technical mathematical subject of probability and stochastic analysis. Some of these are very instructional.

L. Arnold, *Stochastic Differential Equations, Theory and Applications*, John Wiley, 1974. Very nice — user friendly.

R. Ash, *Real Analysis and Probability*, Academic Press, 1972. Excellent for background probability theory and functional analysis.

L. Breiman, *Probability*, Addison-Wesley, 1968. Very good background reference for probability theory.

P. Billingsley, *Probability and Measure*, 3rd edition, John Wiley, 1995. Excellent, well-written and highly recommended.

N. H. Bingham and R. Kiesel, *Risk-Neutral Valuation, Pricing and Hedging of Financial Derivatives*, Springer, 1998. Useful reference for those interested in applications in mathematical finance, but, as the authors point out, many proofs are omitted.

Z. Brzeźniak and Tomasz Zastawniak, *Basic Stochastic Processes*, SUMS, Springer, 1999. Absolutely first class textbook — written to be understood by the reader. This book is a must.

K. L. Chung and R. J. Williams, *Introduction to Stochastic Integration*, 2nd edition, Birkhäuser, 1990. Advanced monograph with many references to the first author's previous works.

R. Durrett, *Probability: Theory and Examples*, 2nd edition, Duxbury Press, 1996. The author clearly enjoys his subject — but be prepared to fill in many gaps (as exercises) for yourself.

DR.RUPNATHJI(DR.RUPAK NATH)

R. Durrett, *Stochastic Calculus, A Practical Introduction*, CRC Press, 1996. Good reference book — with constant reference to the author's other book (and with many exercises).

R. J. Elliott, *Stochastic Calculus and Applications*, Springer, 1982. Advanced no-nonsense monograph.

R. J. Elliott and P. E. Kopp, *Mathematics of Financial Markets*, Springer, 1999. For the specialist in financial mathematics.

T. Hida, *Brownian Motion*, Springer, 1980. Especially concerned with generalized stochastic processes.

N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, 2nd edition, 1989. Advanced monograph.

I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd edition, Springer, 1991. Many references to other texts and with proofs left to the reader.

F. Klebaner, *Introduction to Stochastic Calculus with Applications*, Imperial College Press, 1998. User-friendly — but lots of gaps.

P. E. Kopp, *Martingales and Stochastic Integrals*, Cambridge University Press, 1984. For those who enjoy functional analysis.

R. Korn and E. Korn, *Option Pricing and Portfolio Optimization, Modern Methods of Financial Mathematics*, American Mathematical Society, 2000. Very good — recommended.

D. Lamberton and B. Lapeyre, *Introduction to Stochastic Calculus Applied to Finance*, Chapman and Hall/CRC 1996. For the practitioners.

R. S. Lipster and A. N. Shiryaev, *Stochastics of Random Processes, I General Theory*, 2nd edition, Springer, 2001. Advanced treatise.

X. Mao, *Stochastic Differential Equations and their Applications*, Horwood Publishing, 1997. Good — but sometimes very brisk.

A. V. Mel´nikov, *Financial Markets, Stochastic Analysis and the Pricing of Derivative Securities*, American Mathematical Society, 1999. Well worth a look.

DR.RUPNATHJI(DR.RUPAK NATH)

P. A. Meyer, *Probability and Potentials*, Blaisdell Publishing Company, 1966. Top-notch work from one of the masters (but not easy).

T. Mikosch, *Elementary Stochastic Calculus with Finance in View*, World Scientific, 1998. Excellent — a very enjoyable and accessible account.

B. Øksendal, *Stochastic Differential Equations, An Introduction with Applications*, 5th edition, Springer, 2000. Good — but there are many results of interest which only appear as exercises.

D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, Springer, 1991. Advanced treatise.

J. M. Steele, *Stochastic calculus and Financial Applications*, Springer, 2001. Very good — recommended reading.

H. von Weizsäcker and G. Winkler, *Stochastic Integrals*, Vieweg and Son, 1990. Advanced mathematical monograph.

D. Williams, *Diffusions, Markov Processes and Martingales*, John Wiley, 1979. Very amusing — but expect to look elsewhere if you would like detailed explanations.

D. Williams, *Probability with Martingales*, Cambridge University Press, 1991. Very good, with proofs of lots of background material — recommended.

J. Yeh, *Martingales and Stochastic Analysis*, World Scientific, 1995. Excellent — very well-written careful logical account of the theory. This book is like a breath of fresh air for those interested in the mathematics.

DR.RUPNATHJI(DR.RUPAK NATH)