III Advanced Probability

Ishan Nath, Michaelmas 2024

Based on Lectures by Prof. Jason Miller

October 21, 2024

Contents

0 Introduction

We will be following the lecture notes by Perla Sousi.

This course is divided into seven main topics:

Chapter 1. Conditional expectation. We have seen how to define $\mathbb{E}[Y|X=x]$ for X a discrete-valued random variable, or one with continuous density wrt. the Lebesgue measure. We will define more generally $\mathbb{E}[Y|\mathcal{G}]$, for $\mathcal{G}$ a σ-algebra.

Chapter 2. Discrete-time martingales. A martingale is a sequence (X_n) of integrable random variables such that $\mathbb{E}[X_{n+1} \mid X_1, \ldots, X_n] = X_n$, for all n. A basic example is a simple symmetric random walk. We will be looking at properties, bounds and computations.

Chapter 3. Continuous time processes. Examples are

- Martingales in continuous time.
- Some continuous continuous-time processes.

Chapter 4. Weak convergence. This is a notion of convergence which extends convergence in distribution, and is related to other notions of convergence.

Chapter 5. Large deviations. Here we estimate how unlikely rare events are. An example if the sample mean for (ξ_j) iid, with mean μ, which converges to μ by SLLN. We can ask what the probability is that the sample mean deviates from μ by ε, which can be answered by Cramer's theorem.

Chapter 6. Brownian motion. This is a fundamental object, which is a continuous time stochastic process. We will look at the definition, Markov processes, and its relationship with PDEs.

Chapter 7. Poisson random measures. This is the generalization of the standard Poisson process on $\mathbb{R}_+$.

1 Conditional Expectation

Recall that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of a set Ω, a σ-algebra on Ω, meaning:

- $\Omega \in \mathcal{F}$,
- $A \in \mathcal{F} \implies A^c \in \mathcal{F}$,
- if (A_n) is in $\mathcal{F}$, then $\bigcup A_n \in \mathcal{F}$,

and $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$, meaning:

- $\mathbb{P} : \mathcal{F} \to [0,1]$ such that
- $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$,
- if (A_n) are pairwise disjoint in $\mathcal{F}$, then $\mathbb{P}(\bigcup A_n) = \sum \mathbb{P}(A_n)$.

An important σ-algebra is the Borel σ-algebra $\mathcal{B}$, which for $\mathbb{R}$ is the intersection of all σ-algebra on $\mathbb{R}$ which contain the open sets in $\mathbb{R}$.

Recall that $X : \Omega \to \mathbb{R}$ is a random variable if $X^{-1}(U) \in \mathcal{F}$ for all $U \subseteq \mathbb{R}^n$ open. This is equivalent to $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}$.

Definition 1.1. Suppose that $\mathcal{A}$ is a collection of sets in Ω. Then

$$\sigma(\mathcal{A}) = \bigcap \{\mathcal{E} \mid \mathcal{E} \text{ is a } \sigma\text{-algebra containing } \mathcal{A}\}.$$

If (X_i) is a collection of random variables, then

$$\sigma(X_i \mid i \in I) = \sigma(\{\omega \in \Omega \mid X_i(\omega) \in B\}, i \in I, B \in \mathcal{B}).$$

This is the smallest σ-algebra which makes X_i measurable for all $i \in I$.

Let $A \in \mathcal{F}$. Set

$$\mathbb{1}_A(x) = \mathbb{1}(x \in A) = \begin{cases} 1 & x \in A, \\ 0 & \text{else.} \end{cases}$$

We can define expectation as follows:

- For $C_i \geq 0$, $A_i \in \mathcal{F}$, set
$$\mathbb{E}\left[\sum_{i=1}^n C_i \mathbb{1}_{A_i}\right] = \sum_{i=1}^n C_i \mathbb{P}(A_i).$$
- For $X \geq 0$, set $X_n = (2^{-n} \lfloor 2^n X \rfloor) \wedge n$, so that $X_n \uparrow X$ as $n \to \infty$. Set
$$\mathbb{E}[X] = \lim_{n \to \infty} \mathbb{E}[X_n].$$

- If X is a general random variable, write $X = X^+ - X^-$, where $X^+ = \max(X, 0)$, $X^- = \max(-X, 0)$. If $\mathbb{E}[X^+]$ or $\mathbb{E}[X^-]$ is finite, set
$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

Call X integrable if $\mathbb{E}[|X|] < \infty$.

Recall that if $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, then
$$\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B).$$
If X is integrable and $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, then
$$\mathbb{E}[X|A] = \mathbb{E}[X\mathbb{1}_A]/\mathbb{P}(A).$$
Our goal is to extend this definition to
$$\mathbb{E}[X|\mathcal{G}],$$
where $\mathcal{G}$ is a σ-algebra in $\mathcal{F}$. The main idea is that $\mathbb{E}[X|\mathcal{G}]$ is the best prediction of X given $\mathcal{G}$.

Hence it should be a $\mathcal{G}$-measurable random variable Y, which minimizes
$$\mathbb{E}[(X - Y)^2].$$

1.1 Discrete Case

As a warm-up, we consider the discrete case.

Suppose that X is integrable, (B_i) is countable and disjoint with $\Omega = \bigcup B_i$, and $\mathcal{G} = (B_i, i \in I)$. Set $X' = \mathbb{E}[X|\mathcal{G}]$ by
$$X' = \sum_{i \in I} \mathbb{E}[X|B_i]\mathbb{1}_{B_i},$$
with the convention that $\mathbb{E}[X|B_i] = 0$ if $\mathbb{P}(B_i) = 0$. If $\omega \in \Omega$, then
$$X'(\omega) = \sum_i \mathbb{E}[X|B_i]\mathbb{1}(\omega \in B_i).$$
We look at some important properties of X'.

1. X' is $\mathcal{G}$-measurable, as it is a linear combination of the $\mathcal{G}$-measurable random variables $\mathbb{1}_{B_i}$.
2. X' is integrable, as
$$\mathbb{E}[|X'|] \leq \sum_{i \in I} \mathbb{E}[|X|\mathbb{1}_{B_i}] = \mathbb{E}[|X|] < \infty.$$
3. If $G \in \mathcal{G}$, then
$$\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[X'\mathbb{1}_G].$$

1.2 Existence and Uniqueness

We need a couple of important facts from measure theory:

Theorem 1.1 (Monotone Convergence Theorem). *If (X_n) is a sequence of random variables with $X_n \geq 0$ for all n and $X_n \uparrow X$ almost-surely, then $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$.*

Theorem 1.2 (Dominated Convergence Theorem). *Suppose that (X_n) is a sequence of random variables with $X_n \to X$ almost-surely and $|X_n| \leq Y$ for all n where Y is an integrable random variable, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$ as $n \to \infty$.*

The construction of $\mathbb{E}[X|\mathcal{G}]$ requires a few steps; in the case that $X \in L^2$, then this can be defined by orthogonal projection.

We need to recall a few things about L^p spaces. Suppose that $p \in [1, \infty)$, and X is a random variables in $(\Omega, \mathcal{F})$. Then,

$$\|X\|_p = \mathbb{E}[|X|^p]^{1/p}.$$

Now

$$L^p = L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X \text{ with } \|X\|_p < \infty\}.$$

For $p = \infty$,

$$\|X\|_\infty = \inf\{\lambda \mid |X| \leq \lambda \text{ almost surely}\},$$

and

$$L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P}) = \{X \text{ with } \|X\|_\infty < \infty\}.$$

Two random variables in L^p are equivalent if they agree almost-surely. Then $\mathcal{L}^p$ is the set of equivalence classes under this equivalence relation.

Theorem 1.3. *The space $(\mathcal{L}^2, \|\cdot\|_2)$ is a Hilbert space with inner product*

$$\langle X, Y\rangle = \mathbb{E}[XY].$$

If $\mathcal{H}$ is a closed subspace of $\mathcal{L}^2$, then for all $X \in \mathcal{L}^2$, there exists $Y \in \mathcal{H}$ such that

$$\|X - Y\|_2 = \inf_{Z \in \mathcal{H}} \|X - Z\|_2,$$

and $\langle X - Y, Z\rangle = 0$ for all $Z \in \mathcal{H}$. Y is the orthogonal projection *of X onto $\mathcal{H}$.*

Theorem 1.4. *Let X be an integrable random variable, and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. Then there exists a random variable Y such that:*

(i) *Y is $\mathcal{G}$-measurable.*

(ii) *Y is integrable, with*

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$$

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

If Y' satisfies these two properties, then $Y = Y'$ almost surely. Y is called (a version of) the conditional expectation *of X given $\mathcal{G}$, and we write*

$$Y = \mathbb{E}[X|\mathcal{G}].$$

If $\mathcal{G} = \sigma(G)$ for a random variable G, we write $Y = \mathbb{E}[X|G]$.

Proof: We begin by proving uniqueness. Suppose Y, Y' are two conditional expectation. Then

$$A = \{Y > Y'\} \in \mathcal{G},$$

so by (ii),

$$\mathbb{E}[(Y - Y')\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A] - \mathbb{E}[X\mathbb{1}_A] = 0,$$

but the LHS is non-negative, so the only way this can hold is if $Y \le Y'$ almost surely. Similarly, $Y \ge Y'$ almost surely, so $Y = Y'$ almost surely.

Now we are ready to tackle existence. The first step is by restricting to $\mathcal{L}^2$ function.

Take $X \in \mathcal{L}^2$. Note that $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$.

Set $Y = \mathbb{E}[X|\mathcal{G}]$ to be the orthogonal projections of X onto $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$.

It is automatically true that (i) holds. Suppose $A \in \mathcal{G}$. Then

$$\mathbb{E}[(X - Y)\mathbb{1}_A] = 0,$$

by orthogonality, since $\mathbb{1}_A$ is $\mathcal{G}$-measurable. So

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A],$$

so (ii) holds. Finally Y is integrable as it is specified to be in $\mathcal{L}^2$.

As an aside, notice if $X \ge 0$, then $Y = \mathbb{E}[X|\mathcal{G}] \ge 0$ almost-surely. Since $\{Y < 0\} \in \mathcal{G}$ and

$$\mathbb{E}[Y\mathbb{1}_{\{Y<0\}}] = \mathbb{E}[X\mathbb{1}_{\{Y<0\}}] \ge 0,$$

which can only happen if $\mathbb{P}(Y < 0) = 0$.

Our second step is to prove this for $X \ge 0$. In this case, let $X_n = X \wedge n$. Then $X_n \in \mathcal{L}^2$ for all n, so there exists $\mathcal{G}$-measurable random variables Y_n so that

$$\mathbb{E}[Y_n\mathbb{1}_A] = \mathbb{E}[(X \wedge n)\mathbb{1}_A],$$

for all $A \in \mathcal{G}$. Now (X_n) is increasing in n, so (Y_n) are also increasing by the above aside. Set

$$Y = \limsup_n Y_n.$$

We now have to check the definitions. Note Y is $\mathcal{G}$-measurable as it is a limsup of $\mathcal{G}$-measurable functions. So we check the second definition.

For $A \in \mathcal{G}$,

$$\mathbb{E}[Y\mathbb{1}_A] \overset{MCT}{=} \lim_n \mathbb{E}[Y_n \mathbb{1}_A] = \lim_n \mathbb{E}[(X \wedge n)\mathbb{1}_A] \overset{MCT}{=} \mathbb{E}[X\mathbb{1}_A].$$

We did not check integrability, but notice that setting $A = G$, $\mathbb{E}[Y] = \mathbb{E}[X]$, and X is non-negative, hence so is Y. Thus Y is integrable.

Finally we prove the result for $X \in \mathcal{L}^1$. We can apply step 2 to X^+ and X^-, and so can set

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X^+|\mathcal{G}] - \mathbb{E}[X^-|\mathcal{G}].$$

This is $\mathcal{G}$-measurable and integrable as it is the difference of two $\mathcal{G}$-measurable and integrable random variables, and satisfies (ii) since both of the random variables satisfy (ii).

Remark.

- The proof also works for $X \geq 0$, and not necessarily integrable. Then $\mathbb{E}[X|\mathcal{G}]$ is not necessarily integrable.
- The property
$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$$
for all $A \in \mathcal{G}$, is equivalent to
$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]Y] = \mathbb{E}[XY],$$
for all Y bounded and $\mathcal{G}$-measurable.

1.3 Properties of Expectation

Definition 1.2. A collection of σ-algebras $(\mathcal{G}_i)$ in $\mathcal{G}$ is *independent* if whenever $G_i \in \mathcal{G}_i$ and $i_1, \dots, i_n$ distinct, then

$$\mathbb{P}(G_{i_1} \cap \cdots \cap G_{i_n}) = \prod_{k=1}^n \mathbb{P}(G_{ik}).$$

We say that a random variable X is *independent* of a σ-algebra $\mathcal{G}$ if $\sigma(X)$ is independent of $\mathcal{G}$.

Proposition 1.1. *Let $X, Y \in \mathcal{L}^2$, and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. Then:*

(i) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.

(ii) *If* X *is* $\mathcal{G}$*-measurable, then* $\mathbb{E}[X|\mathcal{G}] = X$ *almost surely.*

(iii) *If* X *is independent of* $\mathcal{G}$*, then* $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

(iv) *If* $X \geq 0$ *almost surely, then* $\mathbb{E}[X|\mathcal{G}] \geq 0$ *almost surely.*

(v) *For any* $\alpha, \beta \in \mathbb{R}$, $\mathbb{E}[\alpha X + \beta Y \mid \mathcal{G}] = \alpha\mathbb{E}[X|\mathcal{G}] + \beta\mathbb{E}[Y|\mathcal{G}]$.

(vi) $|\mathbb{E}[X|\mathcal{G}]| \leq \mathbb{E}[|X||\mathcal{G}]$ *almost-surely.*

Theorem 1.5 (Fatou's Lemma)**.** *If* (X_n) *is a sequence of random variables with* $X_n \geq 0$*, then*

$$\mathbb{E}[\liminf_n X_n] \leq \liminf_n \mathbb{E}[X_n].$$

Theorem 1.6 (Jensen's Inequality)**.** *Let* $X \in \mathcal{L}^2$ *and let* $\phi : \mathbb{R} \to (-\infty, \infty]$ *be a convex function. Then,*

$$\mathbb{E}[\phi(X)] \geq \phi(\mathbb{E}[X]).$$

Proposition 1.2. *Let* $\mathcal{G} \subseteq \mathcal{F}$ *be a* σ*-algebra.*

1. *If* (X_n) *is an increasing sequence of random variables with* $X_n \geq 0$ *for all* n *and* $X_n \uparrow X$*, then*
$$\mathbb{E}[X_n|\mathcal{G}] \uparrow \mathbb{E}[X|\mathcal{G}],$$
almost-surely (conditional MCT).

2. *If* $X_n \geq 0$*, then*
$$\mathbb{E}[\liminf_n X_n|\mathcal{G}] \leq \liminf_n \mathbb{E}[X_n|\mathcal{G}]$$
(conditional Fatou's lemma).

3. *If* $X_n \to X$ *and* $|X_n| \leq Y$ *almost-surely for all* n *and* $Y \in \mathcal{L}^1$*, then*
$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}],$$
almost-surely (conditional DCT).

4. *If* $X \in \mathcal{L}^\infty$ *and* $\phi : \mathbb{R} \to (-\infty, \infty]$ *is convex such that either* $\phi(X) \in \mathcal{L}^1$ *or* $\phi(X) \geq 0$*, then:*
$$\mathbb{E}[\phi(X)|\mathcal{G}] \geq \phi(\mathbb{E}[X|\mathcal{G}])$$
almost surely (convex Jensen's). In particular, for all $1 \leq p < \infty$*,*
$$\|\mathbb{E}[X|\mathcal{G}]\|_p \leq \|X\|_p.$$

Proof:

1. Let Y_n be a version of of $\mathbb{E}[X_n|\mathcal{G}]$. Since $0 \le X_n \uparrow X$ as $n \to \infty$, we have that $Y_n \ge 0$ and are increasing.

Define $Y = \limsup_{n\to\infty} Y_n$. We will show that $= \mathbb{E}[X|\mathcal{G}]$.

- Y is $\mathcal{G}$-measurable as it is a lim sup of $\mathcal{G}$-measurable random variables.
- For $A \in \mathcal{G}$,

$$\mathbb{E}[X\mathbb{1}_A] \overset{MCT}{=} \lim_{n\to\infty} \mathbb{E}[X_n\mathbb{1}_A] = \lim_{n\to\infty} \mathbb{E}[Y_n\mathbb{1}_A] \overset{MCT}{=} \mathbb{E}[Y\mathbb{1}_A].$$

2. The sequence $\inf_{k\ge n} X_k$ is increasing in n. Moreover,

$$\lim_{n\to\infty} \inf_{k\ge n} X_k = \liminf_{n\to\infty} X_n.$$

By 1,

$$\lim_{n\to\infty} \mathbb{E}[\inf_{k\ge n} X_k|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}].$$

But also,

$$\mathbb{E}[\inf_{k\ge n} X_k|\mathcal{G}] \le \inf_{k\ge n} \mathbb{E}[X_k|\mathcal{G}]$$

almost-surely, by monotonicity. Hence taking limits, we get Fatou's lemma.

3. Since $X_n + Y$, $Y - X_n$ give a sequence of random variables which are non-negative,

$$\mathbb{E}[X+Y|\mathcal{G}] = \mathbb{E}[\liminf_n (X_n+Y)|\mathcal{G}] \le \liminf_{n\to\infty} \mathbb{E}[X_n+Y|\mathcal{G}]$$

almost-surely, and similarly

$$\mathbb{E}[Y-X|\mathcal{G}] = \mathbb{E}[\liminf_{n\to\infty} (Y-X_n)|\mathcal{G}] \le \liminf_{n\to\infty} \mathbb{E}[Y-X_n|\mathcal{G}]$$

almost-surely. Combining these inequalities,

$$\limsup_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \le \mathbb{E}[X|\mathcal{G}] \le \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}].$$

This can only hold if $\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]$, almost-surely.

4. We use the fact that a convex function can be written as a supremum of countably many affine functions:

$$\phi(x) = \sup_i (a_i x + b_i).$$

Hence we get

$$\mathbb{E}[\phi(X)|\mathcal{G}] \geq a_i \mathbb{E}[X|\mathcal{G}] + b_i$$

for all i almost-surely, hence

$$\mathbb{E}[\phi(X)|\mathcal{G}] \geq \sup_i (a_i \mathbb{E}[X|\mathcal{G}] + b_i) = \phi(\mathbb{E}[X|\mathcal{G}]),$$

almost-surely.

In particular, for $1 \leq p < \infty$,

$$\begin{aligned} \|\mathbb{E}[X|\mathcal{G}]\|_p^p &= \mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|^p] \leq \mathbb{E}[\mathbb{E}[|X|^p|\mathcal{G}]] \\ &= \mathbb{E}[|X|^p] = \|X\|_p^p. \end{aligned}$$

Proposition 1.3 (Tower Property)**.** *Let $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ be σ-algebras, and $X \in \mathcal{L}^1$. Then,*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$$

almost-surely.

Proof: Note that $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ is $\mathcal{H}$-measurable. For $A \in \mathcal{H}$, we have

$$\begin{aligned} \mathbb{E}[\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]\mathbb{1}_A] &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_A] \\ &= \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbb{1}_A], \end{aligned}$$

since $A \in \mathcal{G}$.

Proposition 1.4 (Taking out what is known)**.** *Let $X \in \mathcal{L}^1$, $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. If Y is bounded and $\mathcal{G}$-measurable, then*

$$\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y \qquad \text{a.s.}$$

Proof: Y is $\mathcal{G}$-measurable, so $\mathbb{E}[X|\mathcal{G}]Y$ is $\mathcal{G}$-measurable.

For $A \in \mathcal{G}$,

$$\mathbb{E}[(\mathbb{E}[X|\mathcal{G}]Y)\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}](Y\mathbb{1}_A)] = \mathbb{E}[XY\mathbb{1}_A].$$

Definition 1.3. Let $\mathcal{A}$ be a collection of subsets of Ω. Then $\mathcal{A}$ is a *π-system* if for all $A, B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$, and $\emptyset \in \mathcal{A}$.

Theorem 1.7. *Let μ_1, μ_2 be measures on $(E, \mathcal{E})$. Suppose that $\mathcal{A}$ is a π-system which generates $\mathcal{E}$, and*

$$\mu_1(A) = \mu_2(A)$$

for all $A \in \mathcal{A}$, and $\mu_1(E) = \mu_2(E)$ is finite.

Then $\mu_1 = \mu_2$.

Proposition 1.5. *Let $X \in \mathcal{L}^1$, and $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ be σ-algebras. If $\sigma(X, \mathcal{G})$ are independent of $\mathcal{H}$, then*

$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}] \qquad \text{a.s.}$$

Proof: Without loss of generality, $X \geq 0$, since the general case comes from decomposing $X = X^+ - X^-$. Let $A \in \mathcal{G}$, $B \in \mathcal{H}$. Then,

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})]\mathbb{1}_{A\cap B}] &= \mathbb{E}[X\mathbb{1}_{A\cap B}] = \mathbb{E}[X\mathbb{1}_A]\mathbb{P}(B) \text{ (independence)} \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_A]\mathbb{P}(B) = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_{A\cap B}].\end{aligned}$$

Define two measures

$$\begin{aligned}\mu_1(F) &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_F], \\ \mu_2(F) &= \mathbb{E}[\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})]\mathbb{1}_F].\end{aligned}$$

These are measures on $\mathcal{F}$ which agree on the π-system $\{A \cap B \mid A \in \mathcal{G}, B \in \mathcal{H}\}$, generating $\sigma(\mathcal{G}, \mathcal{H})$. Moreover since $X \in \mathcal{L}^1$,

$$\mu_1(F) = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X] = \mu_2(F) < \infty.$$

We can now use uniqueness of measures to see these two measures agree on $\sigma(\mathcal{G}, \mathcal{H})$, which can only occur if

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] \qquad \text{a.s.}$$

1.4 Examples of Conditional Expectation

Example 1.1. (Gaussians)

Let (X, Y) be a Gaussian random vector in $\mathbb{R}^2$. Our goal is to compute

$$X' = \mathbb{E}[X|\mathcal{G}],$$

where $\mathcal{G} = Y$. Since X' is a $\mathcal{G}$-measurable function, there exists a Borel measurable function f so that $X' = f(Y)$. We want to find f.

We guess that $X' = aY + b$, for $a, b \in \mathbb{R}$. Then,

$$a\mathbb{E}[Y] + b = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X],$$

and moreover

$$\begin{aligned} \mathbb{E}[(X - X')Y] &= 0 \\ \implies \operatorname{Cov}(X - X', Y) &= 0 \\ \implies \operatorname{Cov}(X, Y) = \operatorname{Cov}(X', Y) &= a \operatorname{Var}(Y). \end{aligned}$$

We can now take a to satisfy this equation, so $\operatorname{Cov}(X - X', Y) = 0$.

But since $(X - X', Y)$ is a Gaussian random variable with covariance 0, $X - X'$ and Y are independent.

Suppose that Z is a $\sigma(Y)$-measurable random variable. Then Z is independent of $X - X'$, so

$$\mathbb{E}[(X - X')Z] = \mathbb{E}[X - X']\mathbb{E}[Z] = 0.$$

This shows the projection property. Hence

$$\mathbb{E}[X|\mathcal{G}] = aY + b,$$

where a, b are determined as before.

Example 1.2. (Conditional Density Functions)

Suppose that X, Y are random variables with a joint density function $f_{X,Y}$ on $\mathbb{R}^2$. Let $h : \mathbb{R} \to \mathbb{R}$ be Borel measurable so that $h(X)$ is integrable.

Our goal is to compute

$$\mathbb{E}[h(X)|Y] = \mathbb{E}[h(X)|\sigma(Y)].$$

The density for Y is given by

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, \mathrm{d}x.$$

Let g be a bounded, measurable function. Then,

$$\begin{aligned}\mathbb{E}[h(X)g(Y)] &= \int h(x)g(y)f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int \left(\int h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathrm{d}x\right) g(y)f_Y(y)\,\mathrm{d}y,\end{aligned}$$

where if $f_Y(y) = 0$, the inner integral is 0. We set

$$\phi(y) = \begin{cases} \int h(x)f_{X,Y}(x,y)/f_Y(y)\,\mathrm{d}x & \text{if } f_Y(y) > 0, \\ 0 & \text{else.} \end{cases}$$

Then,

$$\mathbb{E}[h(X)|Y] = \phi(Y) \qquad \text{a.s.}$$

since $\phi(Y)$ is $\sigma(Y)$-measurable and satisfies the property defining the conditional expectation.

We interpret this computation as giving that

$$\mathbb{E}[h(X)|Y] = \int_{\mathbb{R}} h(x)\nu(Y, \mathrm{d}x),$$

where

$$\begin{aligned}\nu(y, \mathrm{d}x) &= \frac{f_{X,Y}(x,y)}{f_Y(y)}\mathbb{1}_{f_Y(y)\geq 0}\,\mathrm{d}x \\ &= f_{X|Y}(x|y)\,\mathrm{d}x.\end{aligned}$$

$\nu(y, \mathrm{d}x)$ gives the *conditional distribution* of X given $Y = y$, and $f_{X|Y}(x|y)$ is the *conditional density function* of X given $Y = y$.

In this case, the conditional expectation corresponds to an actual expectation. This corresponds to a regular conditional probability distribution.

Also note $f_{X|Y}(x|y)$ is only defined up to a set of measure 0.

2 Discrete Time Martingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(E, \mathcal{E})$ a measurable space.

A sequence of random variables $X = (X_n)$ with values in E is a (discrete time) *stochastic process*. A *filtration* $(\mathcal{F}_n)$ is a sequence of σ-algebras with $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}$.

The *natural filtration* $(\mathcal{F}_n^X)$ associated with X is

$$\mathcal{F}_n^X = \sigma(X_1, \dots, X_n).$$

We say that X is *adapted* to $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable for all n. X is always adapted to its natural filtration. Say that X is *integrable* if $X_n \in \mathcal{L}^1$ for all n.

Definition 2.1. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ be a filtered probability space and let X be a stochastic process which is integrable, adapted and real valued.

- X is a *martingale* (MG) if

  $$\mathbb{E}[X_n | \mathcal{F}_m] = X_m \qquad \text{a.s.}$$

  whenever $n \geq m$.
- X is a *supermartingale* (sup MG) if

  $$\mathbb{E}[X_n | \mathcal{F}_m] \leq X_m \qquad \text{a.s.}$$

  whenever $n \geq m$.
- X is a *submartingale* (sub MG) if

  $$\mathbb{E}[X_n | \mathcal{F}_m] \geq X_m \qquad \text{a.s.}$$

  whenever $n \geq m$.

If X is a MG/sup MG/sub MG, then it is also a MG/sup MG/sub MG with respect to its natural filtration.

Example 2.1.

1. Say

   $$X_n = \sum_{i=1}^n \xi_i,$$

   where (ξ_i) are IID, integrable and $\mathbb{E}[\xi_i] = 0$.

2. We could also have
$$X_n = \prod_{i=1}^{n} \xi_i,$$
where (ξ_i) are IID, integrable and $\mathbb{E}[\xi_i] = 1$.
3. Choosing
$$X_n = \mathbb{E}[Z|\mathcal{F}_n],$$
where $\mathcal{F} \in \mathcal{L}^1$ and $(\mathcal{F}_n)$ is a filtration, this gives a martingale.

Martingales are very useful for:

- computations (optimal stopping theorem),
- bounds (Doob's inequalities),
- proving theorems (martingale convergence theorem).

2.1 Stopping Times

Definition 2.2. Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ be a filtered probability space. A *stopping time* is a random variable $T : \Omega \to \mathbb{Z}_+ \cup \{\infty\}$ with
$$\{T \leq n\} \in \mathcal{F}_n,$$
for all n.

This is equivalent to $\{T = n\} \in \mathcal{F}_n$, for discrete time.

Example 2.2.

1. Constant (deterministic) times.
2. First hitting times: if (X_n) is an adapted stochastic process with values in $\mathbb{R}$, and $A \in \mathcal{B}(\mathbb{R})$, then
$$T_A = \inf\{n \geq 0 \mid X_n \in \mathcal{A}\}.$$
3. Last exit times are not always stopping times.

Proposition 2.1. *Suppose that $S, T, (T_n)$ are stopping times on $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$. Then:*

$$\begin{array}{lll} S \wedge T, & S \vee T & \inf T_n, \\ \sup T_n, & \liminf T_n, & \limsup T_n, \end{array}$$

are stopping times

Definition 2.3. Let T be a stopping time on $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$. Then

$$\mathcal{F}_T = \{A \in \mathcal{F} \mid A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for all } n\}$$

is called the *stopped σ-algebra*.

For $T = n$ deterministic, $\mathcal{F}_T = \mathcal{F}_n$. For X a stochastic process, $X_T = X_{T(\omega)}(\omega)$, whenever $T(\omega) < \infty$.

The *stopped process* X^T is defined by

$$X_n^T = X_{n \wedge T}.$$

Proposition 2.2. *Let S, T be stopping times and X an adapted process. Then,*

(i) *If $S \leq T$, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.*

(ii) *$X_T \mathbb{1}_{\{T<\infty\}}$ is $\mathcal{F}_T$-measurable.*

(iii) *X^T is adapted.*

(iv) *If X is integrable, so is X^T.*

Proof:

(i) This is immediate from definition.

(ii) Take $A \in \mathcal{E}$, then

$$\{X_T \mathbb{1}_{\{T<\infty\}} \in A\} \cap \{T \leq n\} = \bigcup_{k=0}^{n} (\{X_k \in A\} \cap \{T = k\}) \in \mathcal{F}_n.$$

Since X is adapted, $\{T = k\} \in \mathcal{F}_k$.

(iii) For all n, $X_{T\wedge n}$ is $\mathcal{F}_{T\wedge n}$-measurable by (ii), so they are $\mathcal{F}_n$-measurable by (i).

(iv) Note

$$\begin{aligned}\mathbb{E}[|X_{T\wedge n}|] &= \mathbb{E}\left[\sum_{k=0}^{n-1} |X_k| \mathbb{1}(T = k)\right] + \mathbb{E}\left[\sum_{k=n}^{\infty} |X_n| \mathbb{1}(T = k)\right] \\ &\leq \sum_{k=0}^{n} \mathbb{E}[|X_k|] < \infty.\end{aligned}$$

Theorem 2.1 (Optional Stopping Theorem)**.** *Let (X_n) be a MG. Then:*

(i) *If T is a stopping time, then X^T is a MG, hence*

$$\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_0]$$

for all n.

(ii) *If $S \le T$ are bounded stopping times, then*

$$\mathbb{E}[X_T | \mathcal{F}_S] = X_S \qquad \text{a.s.}$$

(iii) *If $S \le T$ are bounded stopping times, then*

$$\mathbb{E}[X_T] = \mathbb{E}[X_S].$$

(iv) *If there exists an integrable random variable Y such that $|X_n| \le Y$ for all n, then for all almost-surely finite stopping times T,*

$$\mathbb{E}[X_T] = \mathbb{E}[X_0].$$

(v) *If X has bounded increments, so $|X_{n+1} - X_n| \le M$, and T is a stopping time with finite expectation, then*

$$\mathbb{E}[X_T] = \mathbb{E}[X_0].$$

Proof:

(i) By the tower property, we only need to show that

$$\mathbb{E}[X_{T\wedge n} | \mathcal{F}_{n-1}] = X_{T\wedge(n-1)} \qquad \text{a.s.}$$

We have

$$\begin{aligned}\mathbb{E}[X_{T\wedge n} | \mathcal{F}_{n-1}] &= \mathbb{E}\left[\sum_{k=0}^{n-1} X_k \mathbb{1}(T = k) \;\middle|\; \mathcal{F}_{n-1}\right] + \mathbb{E}[X_n \mathbb{1}(T > n-1) | \mathcal{F}_{n-1}] \\ &= X_T \mathbb{1}_{(T<n-1)} + X_{n-1}\mathbb{1}_{(T>n-1)} = X_{T\wedge(n-1)}.\end{aligned}$$

Since $(T > n-1) \in \mathcal{F}_{n-1}$, $\mathbb{E}[X_n | \mathcal{F}_{n-1}] = X_{n-1}$.

(ii) Suppose that $T \leq n$, and $S \leq T$. Then,

$$\begin{aligned} X_T &= (X_T - X_{T-1}) + \cdots + (X_{S+1} - X_S) + X_S \\ &= X_S + \sum_{k=0}^{n} (X_{k+1} - X_k) \mathbb{1}_{S \leq k < T}. \end{aligned}$$

Let $A \in \mathcal{F}_S$. Then,

$$\begin{aligned} \mathbb{E}[X_T \mathbb{1}_A] &= \mathbb{E}[X_S \mathbb{1}_A] + \mathbb{E}\left[\sum_{k=0}^{n} (X_{k+1} - X_k) \mathbb{1}_A \mathbb{1}(S \leq k < T)\right] \\ &= \mathbb{E}[X_S \mathbb{1}_A], \end{aligned}$$

since $\{S \leq k < T\} \cap A \in \mathcal{F}_k$ for all k, and X is a MG.

(iii) This follows from taking expectations in (ii).

(iv) and (v) are on the example sheet.

Index