

Advanced Probability

=====

(5th October 2018, Friday)

0. Review

0.1. Measure Spaces

Let E be a set.

Let \mathcal{E} be a set of subsets of E . We say that \mathcal{E} is a **σ -algebra** if $\emptyset \in \mathcal{E}$ and for $A \in \mathcal{E}$ and for any sequence $(A_n : n \in \mathbb{N})$ in \mathcal{E} , we have $E \setminus A \in \mathcal{E}$ and $\cup_n A_n \in \mathcal{E}$. Then (E, \mathcal{E}) is called a **measurable space** and we call the elements of \mathcal{E} **measurable**.

Let μ be a function $\mathcal{E} \rightarrow [0, \infty]$. We say that μ is a **measure** if $\mu(\emptyset) = 0$ and for all sequences $(A_n : n \in \mathbb{N})$ of disjoint elements, we have $\mu(\cup_n A_n) = \sum_n \mu(A_n)$. (**countable additivity**). The triple (E, \mathcal{E}, μ) is called the **measure space**.

Given a topological space E there is a smallest σ -algebra containing all the open sets. This is the **Borel σ -algebra**, denoted $\mathcal{B}(E)$.

E.g. For \mathbb{R} , has $\mathcal{B}(\mathbb{R}) = \mathcal{B}$, and $[0, \infty]$ (the extended real positive line) has analogous Borel σ -algebra.

0.2. Integration of Measurable Functions

Let (E, \mathcal{E}) and (E', \mathcal{E}') be measurable spaces. A function $f : E \rightarrow E'$ is **measurable** if

$$f^{-1}(A) = \{x \in E : f(x) \in A\} \in \mathcal{E} \quad \forall A \in \mathcal{E}'$$

If we refer to a **measurable function** f without specifying range, the default is $(\mathbb{R}, \mathcal{B})$.

If we refer to f as a **non-negative measurable function**, then we mean $E' = [0, \infty]$, $\mathcal{E}' = \mathcal{B}([0, \infty])$. Write $m\mathcal{E}^+$ for the set of non-negative measurable functions.

Theorem 0.2.1.) Let (E, \mathcal{E}, μ) be a measure space. Then there exists a *unique* map $\tilde{m}u : m\mathcal{E}^+ \rightarrow [0, \infty]$ such that

- (a) $\tilde{m}u(1_A) = \mu(A)$ for all $A \in \mathcal{E}$.
- (b) $\tilde{m}u(\alpha f + \beta g) = \alpha \tilde{m}u(f) + \beta \tilde{m}u(g)$ for all $\alpha, \beta \in [0, \infty)$ and all $f, g \in m\mathcal{E}^+$.
- (c) (**monotone convergence**) $\tilde{m}u(f) = \lim_{n \rightarrow \infty} \tilde{m}u(f_n)$ for a non-decreasing sequence of functions $(f_n)_{n \in \mathbb{N}}$ in $m\mathcal{E}^+$ such that $f_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ for all $x \in E$.

(End of statement) \square

For now on, write μ for $\tilde{m}u$. Call $\mu(f)$ the **integral of f** w.r.t. μ . Also write $\int_E f d\mu = \int_E f(x) \mu(dx)$.

A **simple function** is one which is a finite linear combination of indicator functions of measurable sets with positive coefficients. So f is simple if $f = \sum_{k=1}^n \alpha_k 1_{A_k}$ for some $n \geq 0$, $\alpha_k \in (0, \infty)$ and $A_k \in \mathcal{E}$ for $k \in \{1, \dots, n\}$.

Form (a) and (b) for f simple, we have

$$\mu(f) = \sum_{k=1}^n \alpha_k \mu(A_k)$$

Also, if $f, g \in m\mathcal{E}^+$ with $f \leq g$ then $f + h = g$ where $h = g - f1_{f < \infty} \in m\mathcal{E}^+$. Since $\mu(h) \geq 0$, (b) implies $\mu(f) \leq \mu(g)$.

Take $f \in m\mathcal{E}^+$. Define for $x \in E$, $n \in \mathbb{N}$,

$$f_n(x) = 2^{-n} [2^n f(x)] \wedge n$$

Then $(f_n : n \in \mathbb{N})$ is a non-decreasing sequence of simple functions such that $f_n(x) \rightarrow f(x)$ for all $x \in E$. By monotone convergence(property (c)), we have

$$\mu(f) = \lim_{n \rightarrow \infty} \mu(f_n)$$

These procedure characterizes $\mu(f)$ uniquely in terms of the measure $\mu : \mathcal{E} \rightarrow [0, +\infty]$, (a), (b) and (c), so we have shown the uniqueness in **Theorem 0.2.1**.

When is $\mu(f) = 0$? (for $f \in m\mathcal{E}^+$)

For measurable functions f, g we say $f = g$ **almost everywhere**(or a.e.) if

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0$$

Can show, for $f \in m\mathcal{E}^+$, we have $\mu(f) = 0$ if and only if $f = 0$ a.e.

Let f be a measurable function. We say that f is **integrable** if $\mu(|f|) < \infty$. Write $L^1 = L^1(E, \mathcal{E}, \mu)$ for the set of all integrable functions. We extend the integral to L^1 by setting $\mu(f) = \mu(f^+) - \mu(f^-)$ where $f^\pm(x) = 0 \vee (\pm f(x))$ so that $f = f^+ - f^-$. Then can show L^1 is a vector space $\mu : L^1 \rightarrow \mathbb{R}$ is linear.

Lemma 0.2.2.) (Fatou's lemma) Let $(f_n : n \in \mathbb{N})$ be a sequence in $m\mathcal{E}^+$. Then

$$\mu(\liminf_{n \rightarrow \infty} f_n) \leq \liminf_{n \rightarrow \infty} \mu(f_n)$$

(End of statement) \square

The proof is done by applying monotone convergence to the sequence $(\inf_{m \geq n} f_m : n \in \mathbb{N})$.

=====

8th October, Monday

Theorem 0.2.3) (Dominated Convergence) Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions on (E, \mathcal{E}) . Suppose $f_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ for some function f and for every $x \in E$. Suppose furthermore that $|f_n| \leq g$ for all n , for some measurable function g . Then f_n is integrable for all n and so is f . Moreover, we have $\mu(f_n) \rightarrow \mu(f)$ as $n \rightarrow \infty$.

We call a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$ a **probability space**, with the following terms replaced:

measurable functions(f) \leftrightarrow random variable(X)

measurable sets \leftrightarrow events

almost everywhere \leftrightarrow almost surely

integral \leftrightarrow expectation($\mathbb{E}X = \int_{\Omega} X d\mathbb{P}$ in place of $\mathbb{P}(X)$)

0.3. Product measure and Fubini's theorem

Take finite (or σ -finite) measure spaces $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$. Write $\mathcal{E}_1 \otimes \mathcal{E}_2$ for the σ -algebra on $E_1 \times E_2$ generated by sets of the form $A_1 \times A_2$ for $A_1 \in \mathcal{E}_1$ and $A_2 \in \mathcal{E}_2$ and call this **product σ -algebra**.

Theorem 0.3.1.) There exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \quad \forall A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2$$

Theorem 0.3.2.) (Fubini's theorem) Let f be a non-negative measurable function on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$.

- $x_1 \in E_1$, set $f_{x_1}(x_2) = f(x_1, x_2)$, then f_{x_1} is \mathcal{E} -measurable for all $x_1 \in E_1$.
- Set $f_1(x_1) = \mu_2(f_{x_1})$. Then f_1 is \mathcal{E}_1 -measurable, and $\mu_1(f_1) = \mu(f)$.

(End of statement) \square

Now define \hat{f} on $E_2 \times E_1$ by $\hat{f}(x_2, x_1) = f(x_1, x_2)$. Then we can show that \hat{f} is $\mathcal{E}_2 \otimes \mathcal{E}_1$ -measurable and $\mu_2 \otimes \mu_1(\hat{f}) = (\mu_1 \otimes \mu_2)(f)$. So by Fubini, we have $\mu_2(f_2) = \hat{\mu}(\hat{f}) = \mu(f)\mu_1(f_1)$ where f_2 is defined as for f_1 .

$$\text{i.e. } \int_{E_2} \left(\int_{E_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2) = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1).$$

Note, f being integrable is also sufficient to for Fubini's theorem.

Chapter 1. Conditional Expectation

1.1. Discrete Case

Suppose $(G_n : n \in \mathbb{N})$ is a sequence of disjoint sets in \mathcal{F} such that $\cup_n G_n = \Omega$. Let X be an integrable random variable.

- Set $\mathcal{G} = \sigma(G_n : n \in \mathbb{N}) = \{\cup_{n \in I} G_n : I \subset \mathbb{N}\}$.
- Define a new random variable $Y = \sum_{n \in \mathbb{N}} \mathbb{E}(X|G_n) 1_{G_n}$ where $\mathbb{E}(X|G_n) = \mathbb{E}(X 1_{G_n}) / \mathbb{P}(G_n)$ (used convention $0/0=0$). Then,
 - (a) Y is \mathcal{G} -measurable.
 - (b) Y is integrable and $\mathbb{E}(Y 1_A) = \mathbb{E}(X 1_A)$ for all $A \in \mathcal{G}$.

We denote $Y = \mathbb{E}(X|\mathcal{G})$ and call (a version of) the **conditional expectation** of X given \mathcal{G} .

1.2. Gaussian Case

Let (W, X) be a Gaussian random variable in \mathbb{R}^2 that has distribution $N(\mu, \Sigma)$. Take $\mathcal{G} = \sigma(W) = \{W^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\}$.

Consider, for $a, b \in \mathbb{R}$, the random variable $Y = aW + b$.

We can choose a, b so that $\mathbb{E}(Y) = a\mathbb{E}(X) + b = \mathbb{E}(X)$ and $\text{Cov}(Y - X, W) = a\text{Var}(W) - \text{Cov}(X, W) = 0$. Then

- (a) Y is \mathcal{G} -measurable.
- (b) Y is integrable and $\mathbb{E}(Y 1_A) = \mathbb{E}(X 1_A)$ for all $A \in \mathcal{G}$.

To see this, note $Y - X$ and W are independent as $\text{Cov}(Y - X, W) = 0$ so $\mathbb{E}((Y - X) 1_A) = \mathbb{E}(Y - X) \mathbb{P}(A) = 0$ for all $A \in \mathcal{G}$.

1.3. Conditional density functions

Let (U, V) be a random variable in \mathbb{R}^2 with density function $f(u, v)$. That is, $\mathbb{P}((U, V) \in A) = \int_A f(u, v) du dv$. Take $\mathcal{G} = \sigma(U) = \{U^{-1}(B) : B \in \mathcal{B}\}$. Take a Borel measurable function h on \mathbb{R} and set $X = h(V)$. Assume $X \in L^1(\mathbb{P})$. Note U has density function $f(u) = \int f(u, v) dv$. Define the **conditional density function** $f(u|v) = f(u, v)/f(v)$ (again with convention $0/0=0$). Now set $Y = g(U)$ where $g(u) = \int_{\mathbb{R}} h(v) f(v|u) dv$. Then g is a Borel function on \mathbb{R} so

- (a) Y is a \mathcal{G} -measurable random variable.
- (b) Y is integrable and for all $A = U^{-1}(B) \in \mathcal{G}$, we have $\mathbb{E}(Y 1_A) = \mathbb{E}(X 1_A)$.

To see this,

$$\begin{aligned} \mathbb{E}(Y 1_A) &= \int_{\mathbb{R}} g(u) 1_B(u) f(u) du \\ &= \int \int h(v) f(v|u) dv 1_B(u) f(u) du \\ &= \mathbb{E}(X 1_A) \end{aligned}$$

using Fubini's theorem to go from second to the third line.

=====

(10th October, Wednesday)

1.4. Existence and uniqueness of conditional expectation

$(\Omega, \mathcal{F}, \mathbb{P})$ is always the probability space behind.

Theorem 1.4.1.) Let X be integrable random variable and let \mathcal{G} be a sub- σ -algebra of \mathcal{F} . There exist a random variable Y such that

- (a) Y is \mathcal{G} -measurable,
- (b) Y is integrable and for all $A \in \mathcal{G}$, $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$.

Moreover, if Y' is another random variable satisfying (a) and (b) then $Y' = Y$ a.s.

We will write $Y = \mathbb{E}(X|\mathcal{G})$ a.s. (because $\mathbb{E}(X|\mathcal{G})$ is not a fully specified random variable) and we say Y is a **(version of) the conditional expectation of X given \mathcal{G}** . In case $X = 1_A$ write $Y = \mathbb{P}(A|\mathcal{G})$ a.s.

An analogous statement holds with 'integrable' replaced by 'non-negative' throughout the statement.

proof) (Uniqueness) Suppose Y satisfies (a) and (b), and Y' satisfies (a) and (b) for an integrable random variable X' , with $X \leq X'$. Consider the non-negative random variable $Z = (Y - Y')1_A$ where $A = \{Y \geq Y'\} \in \mathcal{G}$. Then

$$\mathbb{E}(Y1_A) = \mathbb{E}(X1_A) \leq \mathbb{E}(X'1_A) = \mathbb{E}(Y'1_A)$$

so $\mathbb{E}(Z) \leq 0$ so $Z = 0$ a.s. and so $Y \leq Y'$ a.s.

By symmetry, if $X = X'$, we see that $Y = Y'$ a.s.

(Existence) Consider for now the case $X \in L^2(\mathcal{F})$ (that is, $\mathbb{E}(|X|^2) < \infty$, which is stronger than being integrable). Since $L^2(\mathcal{F})$ is a Hilbert space, and in fact $L^2(\mathcal{G})$ is a closed subspace of $L^2(\mathcal{F})$, there exists an orthogonal projection Y of X on $L^2(\mathcal{G})$. That is, $Y \in L^2(\mathcal{G})$ and for all $Z \in L^2(\mathcal{G})$, we have $\mathbb{E}((Y - X)Z) = 0$. (This is because $(X - Y)$ should be orthogonal to $L^2(\mathcal{G})$, and the inner product is given by $\langle X - Y, Z \rangle = \mathbb{E}((X - Y)Z) = 0$.) So for any $A \in \mathcal{G}$, take $Z = 1_A$ to see $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$. So Y satisfies (a) and (b).

Suppose now that $X \geq 0$. Set $X_n = X \wedge n$ then $X_n \in L^2(\mathcal{F})$. We have shown there exist $Y_n \in L^2(\mathcal{G})$ such that for all $A \in \mathcal{G}$, we have $\mathbb{E}(Y_n1_A) = \mathbb{E}(X_n1_A)$, $0 \leq X_n \leq X_{n+1}$ and $0 \leq Y_n \leq Y$ for all n . Set $\Omega_0 = \{\omega \in \Omega : 0 \leq Y_n(\omega) \leq Y_{n+1}(\omega) \text{ for all } n\}$. Then $\mathbb{P}(\Omega_0) = 1$ (because it is a countable intersection of sets of measure 1). Define a non-negative \mathcal{G} -measurable random variable

$$Y_\infty = \lim_{n \rightarrow \infty} 1_{\Omega_0} Y_n$$

Then $0 \leq 1_{\Omega_0} Y_n \nearrow Y_\infty$ as $n \rightarrow \infty$ so by monotone convergence, we have for all $A \in \mathcal{G}$

$$\mathbb{E}(Y_\infty 1_A) = \lim_{n \rightarrow \infty} \mathbb{E}(Y_n 1_A) = \lim_{n \rightarrow \infty} \mathbb{E}(X_n 1_A) = \mathbb{E}(X 1_A)$$

In particular, taking $A = \Omega$, we see Y_∞ is integrable, so $\mathbb{P}(Y_\infty) = 0$. So if we set $Y = Y_\infty 1_{Y_\infty < \infty}$, then Y is a random variable and satisfies (a) and (b).

In the general case we can apply the proceeding to $X^\pm = 0 \vee (\pm X)$ (so $X = X^+ - X^-$) to obtain Y^\pm and then check $Y = Y^+ - Y^-$ satisfies (a) and (b).

(End of proof) \square

1.5. Properties of conditional expectation

Conditional expectations plays a central role in many core materials of the course. e.g. martingales and Brownian motion. So it would be useful to investigate some properties of conditional expectation.

Fix an integrable random variable X and sub- σ -algebra \mathcal{G} . **Theorem 1.4.1** has useful consequences:

- (i) $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$

(ii) If X is \mathcal{G} -measurable then $\mathbb{E}(X|\mathcal{G}) = X$ a.s.

(iii) If X is independent of \mathcal{G} then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ a.s.

: check for $A \in \mathcal{G}$, $\mathbb{E}(X1_A) = \mathbb{E}(X)\mathbb{P}(A) = \mathbb{E}(\mathbb{E}(X)1_A)$ so $\mathbb{E}(X)$ satisfies condition (b).

(iv) If $X \geq 0$ a.s. then $\mathbb{E}(X|\mathcal{G}) \geq 0$ a.s. (follows from the proof of **Theorem 1.4.1**)

(v) For all $\alpha, \beta \in \mathbb{R}$, all integrable random variables X, Y , has

$$\mathbb{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha \mathbb{E}(X|\mathcal{G}) + \beta \mathbb{E}(Y|\mathcal{G}) \quad \text{a.s.}$$

: check that the expression on the RHS satisfies condition (b) to show that it is a version of conditional expectation of $\alpha X + \beta Y$ given \mathcal{G} .

Now consider a sequence of random variables $(X_n)_n$.

(vi) (*conditional monotone convergence*) If $0 \leq X_n \nearrow X$ pointwise then $\mathbb{E}(X_n|\mathcal{G}) \rightarrow \mathbb{E}(X|\mathcal{G})$ a.s.

: To see this, we know $\mathbb{E}(X_n|\mathcal{G}) \nearrow Y$ a.s. for some non-negative \mathcal{G} -measurable random variable Y . Take $A \in \mathcal{G}$ then by monotone convergence, we have

$$\begin{aligned} \mathbb{E}(Y1_A) &= \lim_{n \rightarrow \infty} \mathbb{E}(\mathbb{E}(X_n|\mathcal{G})1_A) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(X_n1_A) = \mathbb{E}(X1_A) \end{aligned}$$

so Y satisfies (b) for X .

(vii) (*Conditional Fatou's lemma*) For any non-negative random variables X_n ,

$$\mathbb{E}(\liminf_{n \rightarrow \infty} X_n|\mathcal{G}) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{G})$$

: By conditional monotone convergence, we have

$$\begin{aligned} \mathbb{E}(\liminf_{n \rightarrow \infty} X_n|\mathcal{G}) &= \lim_{n \rightarrow \infty} \mathbb{E}(\inf_{m \geq n} X_m|\mathcal{G}) \leq \lim_{n \rightarrow \infty} \inf_{m \geq n} \mathbb{E}(X_m|\mathcal{G}) \\ &= \liminf_{n \rightarrow \infty} \mathbb{E}(X_n|\mathcal{G}) \end{aligned}$$

so we have the result.

(viii) (*Conditional dominated convergence*) If $X_n(\omega) \rightarrow X(\omega)$ for all ω as $n \rightarrow \infty$ and there exist Y integrable so $|X_n| \leq Y$ for all n , then

$$\mathbb{E}(X_n|\mathcal{G}) \rightarrow \mathbb{E}(X|\mathcal{G}) \quad \text{a.s.}$$

: First, note that X_n and X are integrable, since they are bounded by Y . Consider the sequence of non-negative random variables $Z_n = Y - X_n$. By Fatou's lemma,

$$\mathbb{E}(\liminf_n Z_n|\mathcal{G}) = \mathbb{E}(Y - X|\mathcal{G}) \leq \liminf_n \mathbb{E}(Y - X_n|\mathcal{G}) = \mathbb{E}(Y|\mathcal{G}) - \limsup_n \mathbb{E}(X_n|\mathcal{G})$$

and hence

$$\limsup_n \mathbb{E}(X_n|\mathcal{G}) \leq \mathbb{E}(X|\mathcal{G})$$

On the other hand, if we let $W_n = Y + X_n$, applying Fatou's lemma gives

$$\mathbb{E}(\liminf_n W_n|\mathcal{G}) = \mathbb{E}(Y|\mathcal{G}) + \mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G}) + \liminf_n \mathbb{E}(X_n|\mathcal{G})$$

and so

$$\liminf_n \mathbb{E}(X_n|\mathcal{G}) \geq \mathbb{E}(X|\mathcal{G})$$

The two inequalities give the result.

=====
(12th October, Friday)

(ix) (*Conditional Jensen*) Let $c : \mathbb{R} \rightarrow (-\infty, \infty]$ be convex. Then $c(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(c(X)|\mathcal{G})$ a.s.

: To see this, since c is convex there exist real sequences $(a_n), (b_n)$ such that $c(x) = \sup_n (a_n x + b_n)$. Then $c(X) \geq a_n X + b_n$ for all n . So

$$\mathbb{E}(c(X)|\mathcal{G}) \geq \mathbb{E}(a_n X + b_n|\mathcal{G}) = a_n \mathbb{E}(X|\mathcal{G}) + b_n \quad \text{a.s.}$$

so $\mathbb{E}(c(X)|\mathcal{G}) \geq c(\mathbb{E}(X|\mathcal{G}))$ a.s.

(x) $\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p$ for all $p \in [0, \infty)$ (where $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$)

: To see this, take $c(x) = |x|^p$ in conditional Jensen's inequality.

(xi) (*Tower property*) This is also important for martingales.

Suppose $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ be sub- σ -algebras. Then

$$\mathbb{E}(X|\mathcal{H}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) \quad \text{a.s.}$$

: Proof is an exercise.

(xii) (*Taking out what is known*) This is related to the 'filtration' that is to be introduced in the theory of martingales.

Let Y be a bounded \mathcal{G} -measurable random variable. Then

$$\mathbb{E}(YX|\mathcal{G}) = Y\mathbb{E}(X|\mathcal{G}) \quad \text{a.s.} \quad \dots\dots (\star)$$

: To see this, consider first the case when $Y = 1_B$, $B \in \mathcal{G}$. Take $A \in \mathcal{G}$, then

$$\mathbb{E}(Y\mathbb{E}(X|\mathcal{G})1_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})1_A) = \mathbb{E}(X1_{AB}) = \mathbb{E}(YX1_A)$$

so above identity (\star) holds. The general case follows by a monotone class argument. i.e. (\star) extends by linearity to case where Y is simple. Next consider $X, Y \geq 0$ and set $Y_n = (2^{-n} \lfloor 2^n Y \rfloor) \wedge n$. Then Y_n is simple, so $\mathbb{E}(Y_n X 1_A) = \mathbb{E}(Y_n \mathbb{E}(X|\mathcal{G}))$ for all $A \in \mathcal{G}$. Let $n \rightarrow \infty$ using monotone convergence to obtain (\star) for Y . Finally extend to X, Y general using $X = X^+ - X^-$ and $Y = Y^+ - Y^-$. (Or instead, use *monotone class theorem*, that abstracts this process)

(xiii) Let \mathcal{H} be a σ -algebra and suppose \mathcal{H} is independent of $\sigma(X, \mathcal{G})$. Then

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

: To see this, put $Z = \mathbb{E}(X|\sigma(\mathcal{G}, \mathcal{H}))$ and $Y = \mathbb{E}(X|\mathcal{H})$. Take $A \in \mathcal{G}$, $B \in \mathcal{H}$.

$$\begin{aligned} \mathbb{E}(Y1_{A \cap B}) &= \mathbb{E}(X1_A 1_B) = \mathbb{E}(X1_A)\mathbb{P}(B) \\ &= \mathbb{E}(Y1_A)\mathbb{P}(B) = \mathbb{E}(Y1_{A \cap B}) \end{aligned}$$

Hence $\mathbb{E}((Z - Y)1_{A \cap B}) = 0$ for all $A \in \mathcal{G}$, $B \in \mathcal{H}$. Also, $(Z - Y)$ is $\sigma(\mathcal{G}, \mathcal{H})$ -measurable, and therefore $Y = Z$ by standard argument, e.g. using the following lemma.

Lemma) Suppose X is integrable and $\mathbb{E}(X1_A) = 0$ for all $A \in \mathcal{A}$ where \mathcal{A} is a π -system generating \mathcal{F} . Then $X = 0$ a.s.

proof) Consider $\mathcal{D} = \{A \in \mathcal{F} : \mathbb{E}(X1_A) = 0\}$ then \mathcal{D} is a d -system, contains \mathcal{A} so $\mathcal{D} = \mathcal{F}$ by Dynkin's lemma. Finally, $\mathbb{E}(X1_{\{X > 0\}}) = 0$ so $X = 0$ a.s.

We end the chapter with a lemma that is going to be useful.

Lemma 1.5.1.) Let $X \in L^1(\mathbb{P})$. Set $\mathcal{Y} = \{\mathbb{E}(X|\mathcal{G}) : \mathcal{G} \subset \mathcal{F} \text{ a sub-}\sigma\text{-algebra}\}$. Then \mathcal{Y} is UI(*uniformly integrable*). That is,

$$\sup_{Y \in \mathcal{Y}} \mathbb{E}(|Y|1_{|Y| \geq \lambda}) \rightarrow 0 \quad \text{as } \lambda \rightarrow \infty$$

proof) Since X is integrable, given $\epsilon > 0$, there exists $\delta > 0$ s.t. $\mathbb{E}(|X|1_A) \leq \epsilon$ whenever $\mathbb{P}(A) \leq \delta$. (Suppose not, then we should find $\epsilon > 0$ and $A_n \in \mathcal{F}$ s.t. $\mathbb{P}(A_n) \leq 2^{-n}$ and $\mathbb{E}(|X|1_{A_n}) > \epsilon$. Set $A = \bigcap_n \bigcup_{m \geq n} A_m$. Then $\mathbb{P}(A) = 0$ by Borel-Cantelli lemma. But $\mathbb{E}(|X|1_{\bigcup_{m \geq n} A_m}) \geq \epsilon$, while this converges to $\mathbb{E}(|X|1_A) = 0$ as $n \rightarrow \infty$, which is a contradiction.)

Choose $\lambda \in (0, \infty)$ so that $\mathbb{E}(|X|) \leq \lambda\delta$. Take $Y = \mathbb{E}(X|\mathcal{G}) \in \mathcal{Y}$. Since $x \mapsto |x|$ is a convex function, by conditional Jensen's inequality,

$$Y \leq \mathbb{E}(|X||\mathcal{G}) \quad \text{so} \quad \mathbb{E}(|Y|) \leq \mathbb{E}(|X|)$$

(In fact, by decomposition $X = X^+ + X^-$, we see that $|Y| = \mathbb{E}(X^+|\mathcal{G}) + \mathbb{E}(X^-|\mathcal{G}) = \mathbb{E}(|X||\mathcal{G})$ a.s.... but does not need this at this point)

By Chebyshev's inequality, $\mathbb{P}(|Y| \geq \lambda) \leq \lambda^{-1}\mathbb{E}(|Y|) \leq \delta$ hence

$$\begin{aligned} \mathbb{E}\left(|Y|1_{|Y| \geq \lambda}\right) &\leq \mathbb{E}\left(\mathbb{E}(|X||\mathcal{G})1_{|Y| \geq \lambda}\right) \quad \text{as } 1_{Y \geq \lambda} \in \mathcal{G} \\ &= \mathbb{E}(|X|1_{|Y| \geq \lambda}) \leq \epsilon \end{aligned}$$

So $\sup_{x \in \mathcal{Y}} \mathbb{E}(|Y|1_{|Y| \geq \lambda}) \leq \epsilon$. This is true for any $\epsilon > 0$, therefore we have the uniform integrability.

(End of proof) \square