Probability Theory II

MAT 5171

Hanan Ather

Winter 2024

Contents

⋆ These notes were created during my review process to aid my own understanding and not written for the purpose of instruction. I originally wrote them only for myself, and they may contain typos and errors [a]. *No professor has verified or confirmed the accuracy of these notes.* With that said, I've decided to share these notes on the off chance they are helpful to anyone else.

[a] Any corrections are greatly appreciated.

§1 January 8, 2024

§1.1 Sums of independent random variables

Strong Law of Large Numbers: Let $(X_i)_{i\geq 1}$ be independent and identically distributed (i.i.d.) random variables with finite expected value $\mathbb{E}[X_1]$. Define $S_n = \sum_{i=1}^n X_i$. Then, the Strong Law of Large Numbers states:

$$\frac{S_n}{n} \to \mathbb{E}[X_1] \quad \text{almost surely as } n \to \infty.$$

Kolmogorov 0-1 Law: If $(X_n)_{n\geq 1}$ are independent random variables, then for any event A in the tail σ-field $\mathcal{T}$, defined as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \dots),$$

we have $\mathbb{P}(A) \in \{0, 1\}$.

Corollary 1.1

If $(X_n)_{n\geq 1}$ are independent random variables, and $A = \left\{\lim_{n\to\infty} \frac{S_n}{n} = 0\right\}$ and $B = \{S_n \text{ converges}\}$, then $\mathbb{P}(A) \in \{0, 1\}$ and $\mathbb{P}(B) \in \{0, 1\}$.

Theorem 1.2 (Kolmogorov Maximal Inequality) — Let $(X_n)_{n\geq 1}$ be independent random variables with $\mathbb{E}(X_n) = 0$ and $\mathbb{E}(X_n^2) < \infty$ for all n. Then, for any $\alpha > 0$,

$$\mathbb{P}\left(\max_{k\leq n} |S_k| \geq \alpha\right) \leq \frac{1}{\alpha^2}\mathbb{E}(S_n^2).$$

Proof. Let $\tilde{A}_k = \{|S_k| \geq \alpha\}$ and note that $\{\max_{k\leq n} |S_k| \geq \alpha\} = \bigcup_{k=1}^n \tilde{A}_k$. We disjointize the events $\tilde{A}_k$ by taking:

$$\tilde{A}_k = \tilde{A}_k \setminus \left(\bigcup_{i=1}^{k-1} \tilde{A}_i\right) \quad \text{for } k = 2, \dots, n,$$

and

$$\tilde{A}_k = \bigcup_{i=1}^{k} \tilde{A}_i \quad \text{for } k = 1, \dots, n.$$

It can be proven that

$$\max_{k\leq n} |S_k| \geq \alpha \text{ is equivalent to } \bigcup_{k=1}^{n} \tilde{A}_k.$$

Note that:

$$\mathbb{E}(S_n^2) = \int_\Omega S_n^2 \, dP \geq \sum_{k=1}^n \int_{\tilde{A}_k} S_n^2 \, dP = \sum_{k=1}^n \int_{\tilde{A}_k} (S_k^2 + (S_n - S_k)^2) \, dP,$$

where $(\tilde{A}_k)_{k=1,\dots,n}$ are disjoint.

$$\mathbb{E}(S_n^2) \geq \sum_{k=1}^{n} \int_{A_k} \left(S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right) \, dP.$$

Since $(S_n - S_k)^2 \geq 0$, this simplifies to:

$$\mathbb{E}(S_n^2) \geq \sum_{k=1}^{n} \int_{A_k} \left(S_k^2 + 2S_k(S_n - S_k)\right) \, dP.$$

Noting that

$$\int_{A_k} S_k(S_n - S_k) \, dP = \int_{A_k} \left(\sum_{i=1}^{k} X_i\right) \left(\sum_{j=k+1}^{n} X_j\right) \, dP,$$

and since $\{X_i\}_{i=1}^n$ are independent, we have

$$\mathbb{E}\left[\left(\sum_{i=1}^{k} X_i\right) \left(\sum_{j=k+1}^{n} X_j\right)\right] = 0.$$

Thus,

$$\int_{A_k} S_k(S_n - S_k) \, dP = 0,$$

and

$$\mathbb{E}(S_n^2) = \sum_{k=1}^{n} \mathbb{E}(X_k^2) = 0.$$

It follows that:

$$\mathbb{E}(S_n^2) \geq \sum_{k=1}^{n} \alpha^2 \mathbb{P}(A_k) = \alpha^2 \sum_{k=1}^{n} \mathbb{P}(A_k),$$

where $A_k = \{|S_k| \geq \alpha\}$, and the events A_k are disjoint.

In summary, we obtained:

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \leq \frac{1}{\alpha^2} \mathbb{E}(S_n^2).$$

The conclusion follows from (1) and (2). □

Theorem 1.3 (Etemadi's Inequality) — Let $(X_n)_{n\geq 1}$ be independent random variables and let $S_n = \sum_{i=1}^n X_i$. Then, for any $\alpha > 0$, we have

$$P\left(\max_{1\leq r\leq n} |S_r| \geq 3\alpha\right) \leq 3 \max_{1\leq r\leq n} P(|S_r| \geq \alpha).$$

Proof. Omitted. □

Theorem 1.4 (Kolmogorov's Criterion) — Let $(X_n)_{n\geq 1}$ be independent random variables with $E(X_n) = 0$ for all n and $\sum_{n=1}^\infty E(X_n^2) < \infty$. Then, the series $\sum_{n=1}^\infty X_n$ converges almost surely.

Proof: Step 1. Note that by Kolmogorov's maximal inequality, for each integer $n \geq 1$ and $\epsilon > 0$, we have

$$P\left(\max_{1\leq r\leq n} |S_{n+r} - S_n| > \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{i=n+1}^{n+r} E(X_i^2),$$

where $S_{n+r} - S_n = \sum_{i=n+1}^{n+r} X_i$ and (X_i) are independent random variables with $E(X_i) = 0$.

Letting $r \to \infty$, we get

$$P\left(\sup_{r\geq 1} |S_{n+r} - S_n| > \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{i=n+1}^{\infty} E(X_i^2).$$

Finally, letting $n \to \infty$, we obtain

$$\lim_{n\to\infty} P\left(\sup_{r\geq 1} |S_{n+r} - S_n| > \epsilon\right) = 0 \quad \forall \epsilon > 0.$$

This completes the proof of the assertion. □

§2 January 10, 2024

§2.1 Convergence of Random Series continued

Proof of Theorem 22.6 (Continued from last time). Step 1 concluded with:

$$\lim_{n\to\infty} P\left(\sup_{r\geq 1} |S_{n+r} - S_n| > \epsilon\right) = 0 \quad \forall \epsilon > 0. \tag{1}$$

Step 2: Define $E_n(\epsilon) = \left\{\sup_{s,r\geq n} |S_s - S_r| > 2\epsilon\right\}$ and let $E(\epsilon) = \bigcap_{n=1}^{\infty} E_n(\epsilon)$.

Note that $P(E_n(\epsilon)) \downarrow P(E(\epsilon))$ as $n \to \infty$.

Furthermore, observe that if $|S_j - S_n| > 2\epsilon$ then $|S_i - S_n| > \epsilon$ or $|S_R - S_n| > \epsilon$ for some $i, R \geq n$. To see this, assume by contradiction that both $|S_i - S_n| \leq \epsilon$ and $|S_R - S_n| \leq \epsilon$. Then

$$|S_j - S_R| = |(S_j - S_n) + (S_n - S_R)| \leq |S_j - S_n| + |S_n - S_R| \leq 2\epsilon,$$

which contradicts our assumption that $|S_j - S_R| > 2\epsilon$.

Hence,

$$\sup_{j,R\geq n} |S_j - S_R| > 2\epsilon \implies \bigcup_{j,R\geq n} (|S_j - S_n| > \epsilon) \text{ or } (|S_R - S_n| > \epsilon),$$

and so, $E_n(\epsilon) = \bigcup_{j\geq n} \{|S_j - S_n| > \epsilon\}$, which we denote by $A_n(\epsilon)$.

Therefore, we can summarize that

$$P\left(\bigcup_{R\geq n} A_R\right) \leq \frac{1}{\epsilon^2} E(S_n^2),$$

Recall that $A_n(\epsilon) = \left\{\sup_{j\geq n} |S_j - S_n| > \epsilon\right\}$ and by equation (1), $P(A_n(\epsilon)) \to 0$ as $n \to \infty$.

Since $P(E_n(\epsilon)) \leq P(A_n(\epsilon))$ by the squeeze principle, we have $P(E_n(\epsilon)) \to 0$ as $n \to \infty$. Thus,

$$P(E(\epsilon)) = 0 \quad \forall \epsilon > 0. \tag{3}$$

Finally, define $E = \bigcup_{\epsilon>0} E(\epsilon)$. Then, by countable additivity,

$$P(E) \leq \sum_{\epsilon>0} P(E(\epsilon)) = 0.$$

To summarize, we have shown that $P(E) = 0$ (equation 3).

Note that

$$E = \left\{\exists \epsilon > 0 \text{ such that } \forall n, \sup_{j\geq n} |S_j - S_n| > 2\epsilon\right\} = \{(S_n)_n \text{ is not a Cauchy sequence}\}.$$

Hence, $P(E^c) = 1$. This proves that $(S_n)_n$ is a convergent sequence almost surely. □

Theorem 2.1 (22.7) — Let $(X_n)_{n\geq 1}$ be a sequence of independent random variables and $S_n = \sum_{i=1}^{n} X_i$. If $S_n \to S$ almost surely, then $S_n \xrightarrow{\text{a.s.}} S$.

Proof. The main effort will be to prove again that (1) holds. Then, exactly as in the proof of Theorem 22.6, we conclude that $(S_n)_{n\geq 1}$ converges almost surely to a limit that we may call T. Since $S_n \xrightarrow{\text{a.s.}} T$ implies that $S_n \to P$, and by uniqueness of the limit, $T = S$ almost surely. Hence $S_n \to S$ almost surely.

Let us prove (1). The probability that the partial sums deviate from S by at least ϵ can be bounded by

$$P(|S_{n+j} - S_n| \geq \epsilon) \leq P(|S_{n+j} - S| \geq \frac{\epsilon}{2}) + P(|S_n - S| \geq \frac{\epsilon}{2}).$$

Taking the supremum over $j \geq 1$, we obtain

$$\sup_{j\geq 1} P(|S_{n+j} - S_n| \geq \epsilon) \leq \sup_{j\geq 1} P(|S_{n+j} - S| \geq \frac{\epsilon}{2}) + P(|S_n - S| \geq \frac{\epsilon}{2}).$$

As $n \to \infty$, both terms on the right-hand side tend to zero since $S_n \to S$ almost surely. Recall that $S_n \to S$ almost surely means that for every $\epsilon > 0$, $P(|S_n - S| > \epsilon/2) \to 0$ as $n \to \infty$. Hence, for $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that $P(|S_j - S| > \epsilon/2) < \delta$ for all $j \geq N_\epsilon$. Therefore, if $h > N_\epsilon$, then $\sup_{j\geq h} P(|S_j - S| > \epsilon/2) < \delta$. Thus, $\limsup_{h\to\infty} \sup_{j\geq h} P(|S_j - S| > \epsilon/2) = 0$, which proves (1).

We return to (5). Taking the limit as $n \to \infty$ in (5), we obtain:

$$\limsup_{n\to\infty} \sup_{j\geq 1} P(|S_{n+j} - S_n| > \epsilon) = 0 \quad (6)$$

By Etemadi's Maximal Inequality, we have

$$P(\max_{1\leq j\leq n} |S_{n+j} - S_n| > \epsilon) \leq 3 \max_{1\leq j\leq n} P(|S_{n+j} - S_n| > \epsilon/3).$$

Let $n \to \infty$; we get

$$P(\sup_{j\geq 1} |S_{n+j} - S_n| > \epsilon) \leq 3 \sup_{j\geq 1} P(|S_{n+j} - S_n| > \epsilon/3) \to 0 \text{ as } n \to \infty \text{ by (6).}$$

By the Squeeze Principle, (1) follows. □

Theorem 22.8 (Three Series Theorem). Let (X_n) be independent random variables, and define $X_n^{(c)}$ as the truncated random variable at level c:

$$X_n^{(c)} = \begin{cases} X_n & \text{if } |X_n| \leq c, \\ 0 & \text{if } |X_n| > c. \end{cases}$$

Here, $c > 0$.

a) If $\sum X_n$ converges almost surely, then $\sum P(|X_n| > c)$, $\sum E[X_n^{(c)}]$, and $\sum \mathrm{Var}[X_n^{(c)}]$ converge for all $c > 0$.

b) If there exists $c > 0$ such that all three series $\sum P(|X_n| > c)$, $\sum E[X_n^{(c)}]$, and $\sum \mathrm{Var}[X_n^{(c)}]$ converge, then $\sum X_n$ converges almost surely.

Proof. In order that $\sum X_n$ converge with probability 1 it is necessary that the three series converge for all positive c and sufficient that they converge for some positive c.

Proof of Sufficiency. Suppose that the series (22.13) converge, and put $m_n^{(c)} = E[X_n^{(c)}]$. By Theorem 22.6, $\sum(X_n - m_n^{(c)})$ converges with probability 1, and since $\sum m_n^{(c)}$ converges, so does $\sum X_n$. Since $P(X_n \neq X_n^{(c)} \text{ i.o.}) = 0$ by the first Borel–Cantelli lemma, it follows finally that $\sum X_n$ converges with probability 1. □

§2.2 Weak Convergence

Recall (from MAT5170) let $(\Omega, \mathcal{F}, P)$ be a prob. space, and $X : \Omega \to \mathbb{R}$ r.v. i.e.

$$\{X \in A\} = \{\omega \in \Omega; X(\omega) \in A\} \in \mathcal{F} \text{ for any } A \in \mathcal{R}$$

Here $\mathcal{R}$ is the class of Borel sets of $\mathbb{R}$.

- The law of X is a prob. measure on $(\mathbb{R}, \mathcal{R})$ given by:

$$\mu(A) := \mu_X(A) \overset{\text{def}}{=} P(X \in A) \quad \forall A \in \mathcal{R}$$

- The distribution function (c.d.f) of X is a function $F = F_X : \mathbb{R} \to [0,1]$ given by:

$$F(x) = P(X \leq x) \text{ for all } x \in \mathbb{R}$$
$$= \mu((-\infty, x])$$

  where μ is the law of X

Note that:

$$\mu((-\infty, x)) = F(x^-) = \lim_{y \nearrow x} F(y)$$
$$\mu(\{x\}) = F(x) - F(x^-) \text{ the jump of } F \text{ at } x$$

Properties of F:

1. F is non-decreasing
2. F is right-continuous
3. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$

Definition 2.2 (Convergence in Distribution) Let $(X_n)_n$ be a sequence of random variables defined on probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$ and X be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. We say that (X_n) converges in distribution to X, denoted as $X_n \overset{d}{\Rightarrow} X$ or $X_n \overset{d}{\to} X$, if for all points $x \in \mathbb{R}$ at which $F_X(x) = P(X \leq x)$ is continuous, we have

$$F_{X_n}(x) = P_n(X_n \leq x) \to F_X(x) \quad \text{as} \quad n \to \infty.^a$$

[a] This implies that the cumulative distribution functions (c.d.f.'s) satisfy $F_{X_n}(x) \to F_X(x)$, and for the associated probability measures μ_n, μ, we have $\mu_n((-\infty, x]) \to \mu((-\infty, x])$ for all x such that $\mu(\{x\}) = 0$.

Remark: If $\mu_n(-\infty, x] = P_n(X_n \leq x)$ and $\mu(-\infty, x] = P(X \leq x)$ then $\mu_n \Rightarrow \mu$.

Example 2.3 (Example 25.1)**.** Let X_n be a sequence of random variables in $\mathcal{F}$ with $P(X_n = 1)$. Define

$$X_n = \begin{cases} n & \text{on } -n, \\ 0 & \text{otherwise.} \end{cases}$$

The c.d.f. of X_n is:

$$F_n(x) = P(X_n \leq x) = \begin{cases} 0 & \text{if } x < n, \\ 1 & \text{if } x \geq n. \end{cases}$$

For any $x \in \mathbb{R}$ fixed,

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 1 & \text{if } n > x, \\ 0 & \text{otherwise.} \end{cases} = 0.$$

So we will be tempted to say that $F_n \Rightarrow F$ where $F(x) = 0$ for all x. But F is **not** a distribution function! (since $\lim_{x\to\infty} F(x) \neq 1$)

Therefore, we cannot say $F_n \Rightarrow F$.

§3 January 15, 2024

§3.1 Convergence of Distributions, Probability, & Almost Sure

Definition 3.1 (Convergence in Distribution) Let $X_n : \Omega_n \to \mathbb{R}$ be a random variable defined on probability space $(\Omega_n, \mathcal{F}_n, P_n)$, and $X : \Omega \to \mathbb{R}$ be defined on probability space $(\Omega, \mathcal{F}, P)$. We say that $(X_n)_n$ converges in distribution to X if

$$F_{X_n}(x) = P_n(X_n \le x) \to P(X \le x) = F_X(x) \quad \text{for all points} \quad x \in \mathbb{R} \quad \text{s.t.} \quad P(X = x) = 0$$

We write $X_n \Rightarrow X$ or $X_n \xrightarrow{d} X$.

Remark: If $\mu_n(-\infty, x] = P_n(X_n \le x)$ and $\mu(-\infty, x] = P(X \le x)$, then $\mu_n \Rightarrow \mu$.

Definition 3.2 Let (X_n) be random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$.

a) We say that (X_n) converges in probability to X if

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0 \quad \forall \varepsilon > 0.$$

We write $X_n \xrightarrow{P} X$.

b) We say that (X_n) converges to X almost surely (a.s.) or with probability 1 if

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

We write $X_n \xrightarrow{\text{a.s.}} X$.

Theorem 3.3 (25.2) — We will prove the following two claims:

a) If $X_n \to X$ a.s., then $X_n \xrightarrow{P} X$.

b) If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.

Proof. a) Fix $\varepsilon > 0$. Let $A_n = \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \ge \varepsilon\}$.

Recall Theorem 1.1:

$$P(\limsup A_n) \le \limsup P(A_n) \le \liminf P(A_n) \le P(\liminf A_n)$$

It is enough to prove that

$$P(\limsup A_n) = 0 \quad (4)$$

Recall that:

$$\limsup A_n = \bigcap_{N=1}^{\infty} \bigcup_{n\ge N} A_n = \{\omega \mid \exists N, \forall n \ge N, \omega \in A_n\}$$
$$= \{\omega \mid \exists N, \forall n \ge N, |X_n(\omega) - X(\omega)| \ge \varepsilon\}$$

Note that:

$$(\limsup A_n)^c = \bigcup_{N=1}^{\infty} \bigcap_{n\ge N} A_n^c = \{\omega \mid \forall \varepsilon > 0, \exists N, \forall n \ge N, |X_n(\omega) - X(\omega)| < \varepsilon\}$$

by De Morgan's Law, which implies $\{X_n\}$ converges to X hence $P((\limsup A_n)^c) = 1$. So (4) holds. Let $X \in \mathbb{R}$ be such that $P(X = x) = 0$. Let ε_0 be arbitrary.

b)

1. Note that:
$$\{X_n \leq x\} \subseteq \{|X_n - X| \geq \varepsilon\} \cup \{X \leq x - \varepsilon\}$$
To see this, assume by contradiction that $|X_n - X| < \varepsilon$ and $X > x + \varepsilon$. Then $X_n - X > -\varepsilon$ and $X > x + \varepsilon$. Hence
$$X_n = (X_n - X) + X > -\varepsilon + (x + \varepsilon) = x.$$
This is a contradiction.

2. From 1, we deduce that:
$$P(X_n \leq x) \leq P(|X_n - X| \geq \varepsilon) + P(X \leq x - \varepsilon)$$
which can be written as:
$$P(X \leq x - \varepsilon) \leq \lim_{n\to\infty} P(X_n \leq x) \leq \lim_{n\to\infty} P(X_n \leq x + \varepsilon) \quad \text{for all } \varepsilon > 0.$$

3. Finally, let $\varepsilon \to 0$. We get
$$P(X \leq x) \leq \lim_{n\to\infty} P(X_n \leq x) \leq P(X \leq x)$$
Hence,
$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x).$$

This completes the proof since $P(X = x) = 0$. □

Theorem 3.4 (Convergence in Distribution Implies Convergence in Probability) — Let (X_n) be a sequence of random variables defined on the same probability space. If $X_n \xrightarrow{d} X$ for all $\omega \in \Omega$, where $a \in \mathbb{R}$, then $X_n \xrightarrow{P} X$.

Proof. Let $\varepsilon > 0$ be arbitrary. We want to prove that $P(|X_n - a| > \varepsilon) \to 0$ as $n \to \infty$.

Note that
$$\{X_n - a > \varepsilon\} = \{X_n > a + \varepsilon\} \cup \{X_n < a - \varepsilon\} = \{X_n > a + \varepsilon\} \cup \{X_n < a - \varepsilon\}$$
and
$$P(|X_n - a| > \varepsilon) = P(X_n > a + \varepsilon) + P(X_n < a - \varepsilon) \quad (7)$$

We know that $X_n \xrightarrow{d} X$ i.e., $F_{X_n}(x) \to F_X(x)$ for all $x \in \mathbb{R}$ where $P(X = x) = 0$ (i.e., F_X is continuous at x).

Recall that
$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < a, \\ 1 & \text{if } x \geq a. \end{cases}$$

Hence
$$P(X_n \leq x) \to 0 \text{ for all } x < a.$$
and
$$P(X_n \geq x) \to 1 \text{ for all } x > a.$$

We let $n \to \infty$ in (7):
$$\begin{aligned} &P(X_n > a + \varepsilon) = 1 - P(X_n \leq a + \varepsilon) = 1 - F_{X_n}(a + \varepsilon) \to 1 - 0 = 0, \\ &P(X_n < a - \varepsilon) \leq P(|X_n - a| > \varepsilon) \to 0. \end{aligned}$$

In summary, both terms converge to 0. This proves that $P(|X_n - a| > \varepsilon) \to 0$ as $n \to \infty$. □

Theorem 3.5 (Slutsky's Theorem) — If $X_n \xrightarrow{d} X$ and $Y_n - X_n \xrightarrow{P} 0$, then $Y_n \xrightarrow{d} X$.

Proof. Let F be the distribution function of X, i.e., $F(x) = P(X \leq x)$, and let x be a continuity point of F, i.e., $P(X = x) = 0$. Let $\varepsilon > 0$ be arbitrary. Choose y' and y'' continuity points of F such that $y' < x < y''$ and

$$F(x) - F(y') < \varepsilon \quad \text{and} \quad F(y'') - F(x) < \varepsilon$$

where

$$\lim_{y \uparrow x} F(y) = F(x-) = F(x) \quad \text{and} \quad \lim_{y \downarrow x} F(y) = F(x+).$$

Let $\varepsilon > 0$ be such that y' is $x - \varepsilon$ and y'' is $x + \varepsilon$. Similarly to (5) and (6), it can be proved that:

$$P(X_n \leq y') - P(|X_n - X| \geq \varepsilon) \leq P(Y_n \leq x) \leq P(X_n \leq y'') + P(|X_n - X| \geq \varepsilon) \quad \text{(exercise)}$$

Taking $n \to \infty$, we get:

$$P(X \leq y') = \lim_{n\to\infty} P(X_n \leq x) = \lim_{n\to\infty} P(X_n \leq y'') \leq F(y'') = F(x) + \varepsilon$$

Finally, letting $\varepsilon \to 0$, we get:

$$F(x) = \lim_{n\to\infty} P(Y_n \leq x) \leq \lim_{n\to\infty} P(X_n \leq x) \leq F(x)$$

This proves that:

$$\lim_{n\to\infty} P(X_n \leq x) = F(x).$$

□

§4 January 17, 2024

§4.1 Fundamental Theorems

Theorem 4.1 (Skorohod Representation Theorem) — Let $\{\mu_n\}$ and μ be probability measures on $(\mathbb{R}, \mathcal{R})$ such that $\mu_n \Rightarrow \mu$. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables $(Y_n)_n$ on this space such that [a]:

- The distribution of Y_n is μ_n for all n, i.e., $P \circ Y_n^{-1} = \mu_n$ for all n.
- Distribution of Y is μ.
- $Y_n(\omega) \to Y(\omega)$ for all $\omega \in \Omega$.

[a]Recall:
$$(P \circ X^{-1})(A) \stackrel{\text{def}}{=} P(X^{-1}(A)) \text{ where } X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\}$$

Proof: Omitted.

Theorem 4.2 (Continuous Mapping Theorem) — Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable function and D_h be the discontinuity points of h. Let $\{\mu_n\}, \mu$ be probability measures on $(\mathbb{R}, \mathcal{R})$ such that $\mu_n \Rightarrow \mu$. Assume that $\mu(D_h) = 0$. Then

$$\mu_n \circ h^{-1} \Rightarrow \mu \circ h^{-1}.$$

Recall:

$$h : \mathbb{R} \to \mathbb{R} \quad \mu \circ h^{-1}(A) \stackrel{\text{def}}{=} \mu(h^{-1}(A))$$

where

$$h^{-1}(A) = \{x \in \mathbb{R}; h(x) \in A\}.$$

[a]

[a]**Remark:** Note that $D_h \in \mathcal{R}$. See the proof in the textbook.

Proof. By Theorem 25.6 (Skorohod Representation Theorem), there exists a probability space $(\Omega', \mathcal{F}', P')$ and random variables $\{Y_n, Y\}$ on this space such that $P \circ Y_n^{-1} = \mu_n$ and $P \circ Y^{-1} = \mu$, and $Y_n(\omega) \to Y(\omega)$ for all $\omega \in \Omega'$.

Let $\omega \in \Omega'$ but $Y(\omega) \notin D_h$. Then h is continuous at $Y(\omega)$ and hence $h(Y_n(\omega)) \to h(Y(\omega))$.

Denote by $\Omega'_\sim$ the set $\{\omega \in \Omega'; Y(\omega) \notin D_h\}$. Then

$$P(\Omega'_\sim) = P(\{\omega \in \Omega', Y(\omega) \notin D_h\}) = P(Y^{-1}(D_h^c)) = 1 - P(Y^{-1}(D_h)) = 1 - \mu(D_h) = 1.$$

and so $P(\Omega'_\sim) = 1$. This proves that $h(Y_n) \to h(Y)$ almost surely.

Hence $h(Y_n) \xrightarrow{d} h(Y)$ by Theorem 25.2 (a.s. convergence implies convergence in probability), which in turn implies convergence in distribution. This means that $P \circ (h(Y_n))^{-1} \to P \circ (h(Y))^{-1}$.

This proves that $\mu_n \circ h^{-1} \to \mu \circ h^{-1}$. □

Corollary 4.3

If $X_n \xrightarrow{d} X$ and $h : \mathbb{R} \to \mathbb{R}$ is a measurable function such that $P(X \in D_h) = 0$, then $h(X_n) \xrightarrow{d} h(X)$.

Proof. Note that $X_n \xrightarrow{d} X$ means that $\mu_n \to \mu$ where $P \circ X_n^{-1} = \mu_n$ for n and $P \circ X^{-1} = \mu$, and $P(X \in D_h) = (P \circ X^{-1})(D_h) = \mu(D_h)$. Then by Theorem 25.7, $\mu_n \circ h^{-1} \to \mu \circ h^{-1}$. So $h(X_n) \xrightarrow{d} h(X)$.

Law of h_n: Law of $h(X)$ (see below). □

Recall:

$$
\begin{aligned}
P \circ (h(X))^{-1}(A) &= P(\{\omega \in \Omega; h(X(\omega)) \in A\}) \\
&= P(\{\omega \in \Omega; X(\omega) \in h^{-1}(A)\}) \\
&= (P \circ X^{-1})(h^{-1}(A)) \\
&= \mu(h^{-1}(A)) \\
&= (\mu \circ h^{-1})(A).
\end{aligned}
$$

Corollary 4.4

Suppose that $X_n \xrightarrow{P} a$, where $a \in \mathbb{R}$ is a constant. Let $h : \mathbb{R} \to \mathbb{R}$ be measurable and continuous at a. Then $h(X_n) \xrightarrow{P} h(a)$.

Proof. By Theorem 25.2, $X_n \xrightarrow{P} a$, hence, we let $X(\omega) = a$ for all $\omega \in \Omega$. Note that $\{X \in D_h\} = \{a \in D_h\} = \varnothing$, so $P(X \in D_h) = 0$. So by Corollary 1, $h(X_n) \xrightarrow{d} h(a)$. By Theorem 25.3, $h(X_n) \xrightarrow{P} h(a)$. □

Example 4.5 (25.8)**.** Suppose that $X_n \xrightarrow{d} X$ and $\{a_n\}, \{b_n\}$ are real numbers such that $a_n \to a \in \mathbb{R}$ and $b_n \to b \in \mathbb{R}$. Then

$$a_n X_n + b_n \xrightarrow{d} aX + b.$$

(See also problem 25.2 for a generalization.)

Proof. Recall Slutsky's Theorem: If $X_n \xrightarrow{d} X$, and $Y_n - X_n \xrightarrow{P} 0$, then $Y_n \xrightarrow{d} X$.

Example 25.7: If $X_n \xrightarrow{d} X$ and $s_n \to 0$, then $s_n X_n \xrightarrow{d} 0$.

Note that

$$(a_n X_n + b_n) - (aX + b) = (a_n - a)X_n + (b_n - b) \xrightarrow{d} 0 \text{ (by ex. 25.7)}$$

by TRS 25.5.

In addition, because $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = ax + b$ is continuous since $X_n \xrightarrow{d} X$, we also have $h(X_n) \xrightarrow{d} h(X)$, i.e.,

$$a_n X_n + b_n \xrightarrow{d} aX + b.$$

In summary, we proved:

$$
\begin{cases}
(a_n X_n + b_n) - (aX + b) \xrightarrow{d} 0 & \text{(which is equivalent to } P \to 0) \\
a_n X_n + b_n \xrightarrow{d} aX + b.
\end{cases}
$$

By Slutsky's Theorem, we can take the sum and conclude that $a_n X_n + b_n \xrightarrow{d} aX + b$. □

Theorem 4.6 (Portmanteau Theorem) — Let μ_n, μ be probability measures on $\mathbb{R}$. The following statements are equivalent:

(i) $\mu_n \to \mu$

(ii) $\int f d\mu_n \to \int f d\mu$ for any $f : \mathbb{R} \to \mathbb{R}$ which is continuous and bounded

(iii) $\mu_n(A) \to \mu(A)$ for any set $A \in \mathbb{R}$ which is a continuity set, i.e., $\mu(\partial A) = 0$ where $\partial A = \bar{A} \setminus A^\circ$ is the boundary of A

Proof. $(i) \Rightarrow (ii)$: By Skorohod Representation Theorem, there exists a probability space $(\Omega', \mathcal{F}', P')$ and random variables $\{Y_n, Y\}$ on this space such that:

$$P \circ Y_n^{-1} = \mu_n \text{ and } P \circ Y^{-1} = \mu,$$

and $Y_n(\omega) \to Y(\omega)$ for all $\omega \in \Omega'$.

Let $f : \mathbb{R} \to \mathbb{R}$ which is continuous and bounded. Then the discontinuity set of f is $D_f = \varnothing$, hence $\mu(D_f) = 0$.

Moreover, if $Y_n(\omega) \to Y(\omega)$ for all $\omega \in \Omega'$, then:

$$\int_{\mathbb{R}} f d\mu_n = \int_{\Omega'} f(Y_n) dP' \to \int_{\Omega'} f(Y) dP' = \int_{\mathbb{R}} f d\mu$$

by Bounded Convergence Theorem (Thm 16.5) and Change of Variables for $P \circ Y_n^{-1}$ and $P \circ Y^{-1}$. □

Recall: Change of Variable (21.1)

$$\Omega \xrightarrow{P} \mathbb{R} \xrightarrow{f} \mathbb{R}, \quad f(X) = f \circ X$$

$$\int_{\Omega} f(X) dP = \int_{\mathbb{R}} f d(P \circ X^{-1})$$

We can also write this as:

$$\int_{\Omega} f(X(\omega)) dP(\omega) = \int_{\mathbb{R}} f(x) d(P \circ X^{-1})(x)$$

§5 January 22, 2024

§5.1 Intergration to the limit

Theorem 5.1 (25.11) — If $X_n \xrightarrow{d} X$, then $E(|X_n|)$ is bounded above by $\liminf E(|X_n|)$. If $X_n \xrightarrow{d} X$, then $E(|X_n|) \leq \liminf_{n\to\infty} E(|X_n|)$.

Proof. Let μ_n be the law of X_n. Then $\mu_n \to \mu$ where μ is the law of X.
By Skorohod Representation Theorem, there exists a probability space $(\Omega', \mathcal{F}', P')$ and random variables $\{Y_n, Y\}$ on this space such that:

$$P \circ Y_n^{-1} = \mu_n \text{ and}$$
$$P \circ Y^{-1} = \mu,$$

and $Y_n(\omega) \to Y(\omega)$ for all $\omega \in \Omega'$. By Fatou's Lemma, $E'(|Y|) \leq \liminf_{n\to\infty} E'(|Y_n|)$. (Here E' is expectation w.r.t. P') But $E(|X|) = E'(|Y|)$ and $E(|X_n|) = E'(|Y_n|)$ for all n. Let μ_n be the law of X_n and μ the law of X. By the Skorohod Representation Theorem, there exists a probability space $(\Omega', \mathcal{F}', P')$ and random variables $\{Y_n\}$ and Y on this space such that Y_n converges to Y almost surely and the law of Y_n under P' is μ_n and the law of Y under P' is μ. By Fatou's Lemma, $E'(|Y|) \leq \liminf E'(|Y_n|)$. Here E' denotes expectation with respect to P'. But $E(|X_n|) = E'(|Y_n|)$ and $E(|X|) = E'(|Y|)$. □

The Fatou Lemma (Thm 16.3) states that if $\{f_n\}$ are non-negative measurable functions, then $\int \liminf f_n d\mu \leq \liminf \int f_n d\mu$.

Recall (MAT5170) Fatou's Lemma (Thm.16.3). Let $(\Omega, \mathcal{F}, \mu)$ be a measure space such that $\mu(\Omega) < \infty$. Assume (f_n) are measurable $\mathbb{R}$-valued functions such that $f_n \to f$ almost everywhere (w.r.t. μ).
If (f_n) is uniformly integrable and f is integrable, then

$$\int_\Omega f_n d\mu \to \int_\Omega f d\mu.$$

Theorem 5.2 (15.12) — If $X_n \xrightarrow{d} X$ and (X_n) is uniformly integrable, then X is integrable and $E(X_n) \to E(X)$.

Proof. Let μ_n be the law of X_n and μ the law of X. Then $\mu_n \to \mu$. By Skorohod Representation Theorem, there exists a probability space $(\Omega', \mathcal{F}', P')$ and random variables Y_n, Y on this space such that the law of Y_n under P' is μ_n and the law of Y under P' is μ, and $Y_n(\omega) \to Y(\omega)$ for all $\omega \in \Omega'$.
By Fatou's Lemma, since $E(|X_n|)$ is uniformly integrable, it is bounded, hence $E(X_n) \to E(X)$. □

Recall (MAT5170) Fatou's Lemma (Thm.16.3). Let $(\Omega, \mathcal{F}, \mu)$ be a measure space such that $\mu(\Omega) < \infty$. Assume (f_n) are measurable $\mathbb{R}$-valued functions such that $f_n \to f$ almost everywhere (w.r.t. μ).
If (f_n) is uniformly integrable and f is integrable, then

$$\int_\Omega f_n d\mu \to \int_\Omega f d\mu.$$

Theorem 5.3 (15.12) — If $X_n \xrightarrow{d} X$ and (X_n) is uniformly integrable, then X is integrable and $E(X_n) \to E(X)$.

Proof. By Skorohod Representation Theorem (as in the proof of Th.25.11), there exists a probability space $(\Omega', \mathcal{F}', P')$ and random variables Y_n, Y on $(\Omega', \mathcal{F}', P')$ such that

- the law of Y_n is μ_n (where μ_n is the law of X_n),
- the law of Y is μ (where μ is the law of X),
- $Y_n(\omega) \to Y(\omega)$ for all $\omega \in \Omega'$.

Note that Y_n are uniformly integrable since

$$\int_{\Omega'} |Y_n| dP' = \int_{\{|Y|>\alpha\}} |Y_n| dP' = \int_{\{|X|>\alpha\}} |X_n| dP = \int_{\Omega} |X_n| dP$$

when $|Y_n| > \alpha$.

Change of variables (Th.16.13)

$$\int_{\Omega} f(X) dP = \int_{\mathbb{R}} f(z) d(P \circ X^{-1})(z) = \int_{\mathbb{R}} f d\mu$$

By Theorem 16.14, $E'(Y_n) \to E'(Y)$. This gives us the desired conclusion since:

$$E'(Y_n) = E(X_n) \text{ for all } n \text{ and } E'(Y) = E(X).$$

Here E' is expectation with respect to P'. □

§5.2 Characteristic Functions

Definition 5.4 a) Let μ be a probability measure on $(\mathbb{R}, \mathcal{R})$. The characteristic function of μ is:

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \mu(dx) = \int_{-\infty}^{\infty} \cos(tx) \mu(dx) + i \int_{-\infty}^{\infty} \sin(tx) \mu(dx)$$

for all $t \in \mathbb{R}$.
(Recall: e^{it} is defined as $\cos t + i \sin t$ for all $t \in \mathbb{R}$.)

b) Let $X : \Omega \to \mathbb{R}$ be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. Let μ be the law of X. Then the characteristic function of X is:

$$\varphi(t) = E(e^{itX}) = \int_{\mathbb{R}} e^{itx} dP = \int_{\mathbb{R}} e^{itx} \mu(dx)$$

Observation: Since $|e^{itx}|^2 = \cos^2(tx) + \sin^2(tx) = 1$,

$$|\varphi(t)| = \left| \int_{\mathbb{R}} e^{itx} \mu(dx) \right| \leq \int_{\mathbb{R}} |e^{itx}| \mu(dx) = \mu(\mathbb{R}) = 1.$$

1. $\varphi(0) = E(e^{i \cdot 0}) = E(1) = 1$
2. φ is uniformly continuous on $\mathbb{R}$:

$$\begin{aligned} |\varphi(t+\varepsilon) - \varphi(t)| &= \left| \int_{\mathbb{R}} (e^{i(t+\varepsilon)x} - e^{itx}) \mu(dx) \right| \\ &\leq \int_{\mathbb{R}} |e^{i(t+\varepsilon)x} - e^{itx}| \mu(dx) \\ &= \int_{\mathbb{R}} |e^{itx}| \cdot |e^{i\varepsilon x} - 1| \mu(dx) \\ &= \int_{\mathbb{R}} |e^{i\varepsilon x} - 1| \mu(dx) \to 0 \quad \text{by Bounded Convergence Theorem since} \\ |e^{i\varepsilon x} - 1| &\leq |e^{i\varepsilon x}| + 1 = 2 \quad \text{for all } x \text{ as } \varepsilon \to 0. \end{aligned}$$

§6 January 24, 2024

§6.1 Computing Characteristic Function

Example 6.1. Example: Let $X \sim N(0,1)$. We aim to compute $\varphi(t) = \mathbb{E}[e^{itX}]$ for $t \in \mathbb{R}$.

The characteristic function $\varphi(t)$ is given by:

$$\varphi(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \mathbb{E}[X^k] \quad (1)$$

We use the property: for differentiable functions $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g'(X)] = \mathbb{E}[Xg(X)] \quad (2)$$

Since

$$\mathbb{E}[g'(X)] = \int_{-\infty}^{\infty} g'(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \int_{-\infty}^{\infty} g(x) x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \mathbb{E}[Xg(X)],$$

by integration by parts.

Applying (2) for $g(x) = x^k$, then $g'(x) = kx^{k-1}$. So (2) becomes:

$$\mathbb{E}[kX^{k-1}] = \mathbb{E}[X \cdot X^{k-1}] \quad (3)$$

Hence,

$$\mathbb{E}[X^k] = k\mathbb{E}[X^{k-1}] \quad \text{for } k \geq 1 \quad (4)$$

By symmetry of the standard normal distribution, all odd powers of X have an expected value of zero, i.e., $\mathbb{E}[X^k] = 0$ for k odd.

For even powers, using the property from before:

$$\begin{aligned} k = 2: &\quad \mathbb{E}[X^2] = 1, \\ k = 4: &\quad \mathbb{E}[X^4] = 3 \cdot \mathbb{E}[X^2] = 3, \\ k = 6: &\quad \mathbb{E}[X^6] = 5 \cdot \mathbb{E}[X^4] = 5 \cdot 3 = 15, \end{aligned}$$

and so on.

In general, for $k = 2n$:

$$\mathbb{E}[X^{2n}] = 1 \cdot 3 \cdot 5 \cdots (2n-1) = (2n-1)!! \quad \text{(double factorial)}$$

Characteristic Function: Returning to the characteristic function:

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!} \mathbb{E}[X^{2n}] = \sum_{n=0}^{\infty} \frac{(it)^{2n}}{(2n)!} (2n-1)!! = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{2^n n!}$$

where we used the relation $(2n)!/(2n-1)!! = 2^n n!$.

Recalling the Taylor series expansion for $e^{-t^2/2}$, we have:

$$e^{-t^2/2} = \sum_{n=0}^{\infty} \frac{(-1)^n (t^2/2)^n}{n!} = \sum_{n=0}^{\infty} \frac{(-1)^n t^{2n}}{2^n n!}$$

Thus, $\varphi(t) = e^{-t^2/2}$.

Remark: The characteristic function of a random variable $aX + b$ (where $a, b \in \mathbb{R}$) is given by:

$$\varphi_{aX+b}(t) = \mathbb{E}\left[e^{it(aX+b)}\right] = e^{itb}\mathbb{E}\left[e^{itaX}\right] = e^{itb}\varphi_X(at).$$

This expression uses the fact that the characteristic function of X evaluated at at can be modified by a shift in the variable corresponding to the addition of b.

In particular, if $a=-1$ and $b=0$, the characteristic function of $-X$ is:

$$\varphi_{-X}(t)=\varphi_X(-t) \quad \text{for all } t\in\mathbb{R}.$$

Next goal: Our next goal is to show that the characteristic function determines uniquely the law (or the distribution) of a random variable.

Theorem 6.2 (Theorem 26.2.) — Two parts of the theorem:

(a) Let μ be a probability measure on $\mathbb{R}$. Let $\varphi(t)$ be the characteristic function of μ. If $a,b\in\mathbb{R}$ are such that $\mu(\{a\})=0$ and $\mu(\{b\})=0$, then

$$\mu((a,b])=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,dt.$$

Convention: In this formula, the function $\frac{e^{-ita}-e^{-itb}}{it}$ is defined for $t=0$ to be equal to $b-a$ (by l'Hopital's Rule).

(b) Let μ and ν be probability measures on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. If μ and ν have the same characteristic function, then $\mu=\nu$.

Proof. **(a)** Let $I_T=\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}dt$. Then, by Fubini's theorem,

$$\begin{aligned} I_T &= \frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\left(\int_{-\infty}^{\infty}e^{itx}\mu(dx)\right)dt \\ &= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itx}dt\right)\mu(dx) \\ &= \int_{-\infty}^{\infty}\phi_T(x)\mu(dx) \end{aligned}$$

We can apply Fubini's Theorem since:

$$\left|\frac{e^{-ita}-e^{-itb}}{it}\cdot e^{itx}\right|=\left|\frac{e^{-ita}-e^{-itb}}{it}\right|\cdot\left|e^{itx}\right|\le b-a$$

$$\left|e^{-ita}-e^{-itb}\right|=\left|e^{-ita}\left(1-e^{it(b-a)}\right)\right|=\left|e^{-ita}\right|\cdot\left|1-e^{it(b-a)}\right|\le t(b-a)$$

And

$$\int_{-T}^{T}(b-a)\mu(dx)e^{itx}\le(b-a)(2T)\epsilon$$

(Note: It was crucial for this argument to work with $[-T,T]$.)

We compute $\phi_T(x)$ explicitly, as follows:

$$\begin{aligned} \phi_T(x) &= \frac{1}{2\pi}\left[\int_{-T}^{T}\frac{e^{it(x-a)}}{it}dt-\int_{-T}^{T}\frac{e^{it(x-b)}}{it}dt\right] \\ &= \frac{1}{2\pi}\left[-i\int_{-T}^{T}\frac{\cos(t(x-a))}{t}dt+i\int_{-T}^{T}\frac{\sin(t(x-a))}{t}dt\right] \\ &\quad +i\int_{-T}^{T}\frac{\cos(t(x-b))}{t}dt-i\int_{-T}^{T}\frac{\sin(t(x-b))}{t}dt \\ &= \frac{1}{2\pi}\left[(-i)\cdot 2\int_{0}^{T}\frac{\sin(t(x-a))}{t}dt+i\cdot i\cdot 2\int_{0}^{T}\frac{\sin(t(x-b))}{t}dt\right] \end{aligned}$$

$$= \frac{1}{\pi}\left[\int_0^T \frac{\sin(t(x-a))}{t}dt - \int_0^T \frac{\sin(t(x-b))}{t}dt\right]$$

Recall:

$$I_T = \int_{-\infty}^{\infty} \phi_T(x)\mu(dx)$$

We want to let $T \to \infty$, and apply the Dominated Convergence Theorem (D.C.T.)

It can be proved that

$$\lim_{T\to\infty}\int_0^T \frac{\sin(\theta t)}{t}dt = \begin{cases} \frac{\pi}{2} & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -\frac{\pi}{2} & \text{if } \theta < 0 \end{cases}$$

In our case,

$$\lim_{T\to\infty}\int_0^T \frac{\sin(t(x-a))}{t}dt = \begin{cases} -\frac{\pi}{2} & \text{if } x < a \\ 0 & \text{if } x = a \\ \frac{\pi}{2} & \text{if } x > a \end{cases}$$

$$\lim_{T\to\infty}\int_0^T \frac{\sin(t(x-b))}{t}dt = \begin{cases} -\frac{\pi}{2} & \text{if } x < b \\ 0 & \text{if } x = b \\ \frac{\pi}{2} & \text{if } x > b \end{cases}$$

Hence

$$\lim_{T\to\infty}\phi_T(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{2} & \text{if } x = a \\ 1 & \text{if } a < x < b \\ \frac{1}{2} & \text{if } x = b \\ 0 & \text{if } x > b \end{cases}$$

Recall:

$$I_T = \int_{-\infty}^{\infty} \phi_T(x)\mu(dx)$$

We want to let $T \to \infty$, and apply Dominated Convergence Theorem (D.C.T.)

It can be proved that

$$\lim_{T\to\infty}\int_0^T \frac{\sin(\theta t)}{t}dt = \begin{cases} \frac{\pi}{2} & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -\frac{\pi}{2} & \text{if } \theta < 0 \end{cases}$$

Next time!

□

§7 January 29, 2024

§7.1 Characteristic Functions Continued

Corollary 7.1

Let μ be a probability measure with characteristic function φ. If

$$\int_{-\infty}^{\infty} \frac{|\varphi(t)|}{|t|} dt < \infty$$

then μ has a continuous density f given by:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt \quad \text{(Inversion Formula)}$$

Proof. Let $F(x) = \mu((-\infty, x])$ be the cumulative distribution function corresponding to μ. We have to prove that F is differentiable. Then, for $\varepsilon > 0$,

$$\frac{F(x+\varepsilon) - F(x)}{\varepsilon} = \frac{\mu((-\infty, x+\varepsilon]) - \mu((-\infty, x])}{\varepsilon} = \frac{\mu((x, x+\varepsilon])}{\varepsilon}$$

$$= \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-it(x+\varepsilon)} - e^{-itx}}{it\varepsilon} \varphi(t) dt$$

By Theorem 26.2, as $T \to \infty$, this limit exists and hence, the function F is differentiable. **By D.C.T.,**

$$\frac{F(x+\varepsilon) - F(x)}{\varepsilon} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - e^{-it(x+\varepsilon)}}{it\varepsilon} \varphi(t) dt \quad (2) \tag{1}$$

To justify the application of D.C.T, we note:

$$\left| \frac{e^{-itx} - e^{-it(x+\varepsilon)}}{it\varepsilon} \right| = \left| \frac{e^{-itx}(1 - e^{-it\varepsilon})}{it\varepsilon} \right| \le |t| \text{ (since } |1 - e^{-it\varepsilon}| \le |t\varepsilon|)$$

Recall:

$$\left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| \le \min \left\{ \frac{|t|^{n+1}}{(n+1)!}, \frac{2|t|^n}{n!} \right\}$$

$$\left| \frac{e^{-itx} - e^{-it(x+\varepsilon)}}{it\varepsilon} \varphi(t) \right| \le \frac{|t\varepsilon|}{|\varepsilon|} |\varphi(t)| = |\varphi(t)| \text{ and } |\varphi(t)| \text{ is an integrable function.}$$

Note that (2) also holds for $\varepsilon < 0$. By another application of D.C.T.,

$$F'(x) = \lim_{\varepsilon\to 0} \frac{F(x+\varepsilon) - F(x)}{\varepsilon} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{\varepsilon\to 0} \frac{e^{-itx} - e^{-it(x+\varepsilon)}}{it\varepsilon} \varphi(t) dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt$$

Note that f is continuous on $\mathbb{R}$:

$$\begin{aligned}
|f(x+\varepsilon)-f(x)| &= \left|\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it(x+\varepsilon)}\varphi(t)\,dt - \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt\right| \\
&= \left|\frac{1}{2\pi}\int_{-\infty}^{\infty} (e^{-it(x+\varepsilon)} - e^{-itx})\varphi(t)\,dt\right| \\
&\le \frac{1}{2\pi}\int_{-\infty}^{\infty} |e^{-itx}(e^{-it\varepsilon}-1)||\varphi(t)|\,dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} |e^{-it\varepsilon}-1|\cdot|\varphi(t)|\,dt \quad \text{by D.C.T. as } \varepsilon\to 0.
\end{aligned}$$

□

1. If $X \sim N(0,1)$, then X has density $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, x\in\mathbb{R}$ and characteristic function:

$$\varphi(t) = e^{-\frac{t^2}{2}} \quad \text{(used the power series expansion).}$$

2. If $X \sim \text{Uniform}(0,1)$ then X has density $f(x) = \begin{cases} 1 & \text{if } x\in[0,1], \\ 0 & \text{if } x\notin[0,1]. \end{cases}$ and characteristic function:

$$\varphi(t) = \int_0^1 e^{itx}dx = \frac{e^{it}-1}{it} \quad \left(\text{or} \quad \frac{1}{it}(e^{it}-1)'\right).$$

3. If $X \sim \text{Exponential}(\lambda)$, then X has density $f(x) = \lambda e^{-\lambda x}\mathbb{1}_{(0,\infty)}(x)$ and characteristic function:

$$\varphi(t) = \int_0^{\infty} e^{itx}e^{-\lambda x}dx = \left.\frac{e^{(it-\lambda)x}}{it-\lambda}\right|_0^{\infty} = \frac{1}{1-it}. \quad \text{(since the limit as } x\to\infty \text{ is } 0).$$

4. If $X \sim$ Double-Exponential, then X has density $f(x) = \frac{1}{2}e^{-|x|}, x\in\mathbb{R}$ and characteristic function:

$$\begin{aligned}
\varphi(t) &= \int_{-\infty}^{\infty} e^{itx}\cdot\frac{1}{2}e^{-|x|}dx \\
&= \frac{1}{2}\left(\int_0^{\infty} e^{-(1-it)x}dx + \int_{-\infty}^{0} e^{-(1+it)x}dx\right) \\
&= \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) \\
&= \frac{1}{2}\left(\frac{1+it+1-it}{1+t^2}\right) \\
&= \frac{1}{1+t^2}.
\end{aligned}$$

5. If $X \sim$ Cauchy, then X has density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, x\in\mathbb{R}$ and characteristic function:

$$\begin{aligned}
\varphi(t) &= \int_{-\infty}^{\infty} e^{itx}\frac{1}{\pi}\frac{1}{1+x^2}dx \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{1}{1+x^2}dx \\
&= \frac{1}{\pi}\left[e^{-itx}\frac{1}{1+(-it)^2}\right] \\
&= \frac{1}{\pi}\frac{e^{-itx}}{1+t^2}.
\end{aligned}$$

(Note that the characteristic function of a Cauchy distribution is an exercise in some texts and can be derived using complex analysis techniques.)

Theorem 7.2 (Continuity Theorem) — Let $\{\mu_n\}$ and μ be probability measures on $\mathbb{R}$, with characteristic functions $\{\varphi_n\}$ and φ respectively. Then

$$\mu_n \to \mu \text{ if and only if } \varphi_n(t) \to \varphi(t) \text{ for all } t \in \mathbb{R}.$$

Proof. **Part 1 "Only If":** Suppose that $\mu_n \to \mu$. Then, by Portmanteau theorem, we know that

$$\int f d\mu_n \to \int f d\mu \text{ for all } f : \mathbb{R} \to \mathbb{R} \text{ continuous and bounded.}$$

In our case,

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{-itx}\mu_n(dx) = \int_{-\infty}^{\infty} \cos(tx)\mu_n(dx) + i\int_{-\infty}^{\infty} \sin(tx)\mu_n(dx)$$

implies that as $n \to \infty$,

$$\int_{-\infty}^{\infty} \cos(tx)\mu_n(dx) + i\int_{-\infty}^{\infty} \sin(tx)\mu_n(dx) \to \int_{-\infty}^{\infty} e^{-itx}\mu(dx) = \varphi(t).$$

Part 2 "If": We do not discuss this. It uses "tightness". Details are in the book. □

§7.2 Central Limit Theorem

Theorem 7.3 (Lindeberg–Lévy Theorem) — Let $\{X_i\}_{i\geq 1}$ be a sequence of independent and identically distributed (i.i.d.) random variables, with $\mathbb{E}[X_i^2] < \infty$. We denote $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{Var}(X_i)$. Let $S_n = \sum_{i=1}^n X_i$. Then

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

Proof. Let $I = \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\varphi(t)\,dt$. Then, by Fubini's Theorem,

$$\begin{aligned} I_T &= \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\int_{-\infty}^{\infty} e^{itx}\,\mu(dx)\,dt \\ &= \int_{-\infty}^{\infty}\left(\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-it(a-x)}-e^{-it(b-x)}}{it}\,dt\right)\mu(dx) \\ &= \int_{-\infty}^{\infty}\Phi_T(x)\mu(dx), \end{aligned}$$

where $\Phi_T(x)$ is defined as $\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-it(a-x)}-e^{-it(b-x)}}{it}\,dt$.

We can apply Fubini's Theorem since:

$$\left|\frac{e^{-ita}-e^{-itb}}{it}\cdot e^{itx}\right| = \left|\frac{e^{-it(a-x)}-e^{-it(b-x)}}{it}\right| \leq b-a,$$

$$\left|e^{ita}-e^{itb}\right| = \left|e^{itb}(e^{it(a-b)}-1)\right| \leq |t(b-a)|,$$

which is integrable over t in the interval $[-T,T]$ and measurable with respect to μ.

□

Theorem 7.4 (Central Limit Theorem for Triangular Arrays with Lyapunov condition) — For each $n \geq 1$, let $X_{n1}, X_{n2}, \ldots, X_{nn}$ be independent random variables with $\mathbb{E}(X_{ni}) = 0$ for all $i = 1, \ldots, n$ and

$$\sigma_{ni}^2 = \mathbb{E}(X_{ni}^2) < \infty \quad \forall i = 1, \ldots, n.$$

Let $S_n = \sum_{i=1}^n X_{ni}$ and $\lambda_n^2 = \mathbb{E}(S_n^2) = \sum_{i=1}^n \sigma_{ni}^2$. Assume that $\lambda_n^2 \geq 0$ for all n. Suppose that there exists $\delta > 0$ such that

$$\mathbb{E}(|X_{ni}|^{2+\delta}) < \infty \quad \text{for all } i = 1, \ldots, n,$$

and

$$\lim_{n\to\infty} \frac{1}{\lambda_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}(|X_{ni}|^{2+\delta}) = 0 \quad \text{(Lyapunov condition)}.$$

Then

$$\frac{S_n}{\lambda_n} \xrightarrow{d} Z \sim N(0,1).$$

Proof. It suffices to show that the Lyapunov condition holds, and then we apply Theorem 27.2. We have:

$$\begin{aligned}
\frac{1}{\lambda_n^2} \sum_{i=1}^n \int_{\{|X_{ni}| \geq \epsilon\lambda_n\}} X_{ni}^2 dP &= \frac{1}{\lambda_n^2} \sum_{i=1}^n \mathbb{E}\left[X_{ni}^2 \mathbf{1}_{\{|X_{ni}| \geq \epsilon\lambda_n\}}\right] \\
&\leq \frac{1}{\epsilon^\delta \lambda_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\left[|X_{ni}|^{2+\delta}\right] \\
&= \frac{1}{\epsilon^\delta \lambda_n^{2+\delta}} \mathbb{E}\left[\sum_{i=1}^n |X_{ni}|^{2+\delta}\right] \to 0 \quad \text{by the Lyapunov condition.}
\end{aligned}$$

Hence the Lyapunov condition holds. □

§8 February 7, 2024

§8.1 Section 33: Conditional Probability (continued)

Example 8.1. If $P(B) > 0$, $\mathcal{G} = \sigma(\{B\}) \to \{\emptyset, \Omega, B, B^c\}$

$$f(\omega) = \begin{cases} P(A \mid B) & \text{if } \omega \in B \\ P(A \mid B^c) & \text{if } \omega \in B^c \end{cases}$$

We prove that f satisfies conditions (i) and (ii) from the definition of $P(A \mid \mathcal{G})$, i.e.,

(i) f is $\mathcal{G}$-measurable (we checked this last time)

(ii) $\int_G f\, dP = P(A \cap G) \quad \forall G \in \mathcal{G}$

$$\int_G f\, dP = P(A \cap G) \quad \forall G \in \mathcal{G} \tag{1}$$

Last time, we checked that (1) holds for $G = \emptyset$ and $G = \Omega$.
Assume that $G = B$. Then

$$\int_B f\, dP = \int_B (P(A \mid B)\mathbf{1}_B + P(A \mid B^c)\mathbf{1}_{B^c})\, dP$$

$$= \int_B P(A \mid B) dP = P(A \mid B)P(B) = \frac{P(A \cap B)}{P(B)} P(B)$$

$$= P(A \cap B)$$

This proves (1) for $G = B$.
The fact that (1) also holds for $G = B^c$ is similar (exercise).

Example 8.2. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $A \in \mathcal{F}$, and $\mathcal{G} = \sigma(\{B_i\}_{i\geq 1})$, where $\{B_i\}_{i\geq 1}$ is a partition of Ω, $B_i \in \mathcal{F}$, $P(B_i) > 0$ for all $i \geq 1$. We claim that

$$P(A \mid \mathcal{G}) = \sum_{i\geq 1} P(A \mid B_i)\mathbf{1}_{B_i} \quad \text{a.s.} \tag{2}$$

We prove (2): Let $f = \sum_{i\geq 1} P(A \mid B_i)\mathbf{1}_{B_i}$. We check that f satisfies conditions (i) and (ii) from the definition of $P(A \mid \mathcal{G})$.

Condition (i): f is $\mathcal{G}$-measurable since $\mathbf{1}_{B_i}$ is $\mathcal{G}$-measurable for all $i \geq 1$.

Condition (ii): We have to check that

$$\int_G f\, dP = P(A \cap G) \quad \forall G \in \mathcal{G} \tag{1}$$

Note that $\mathcal{G} = \left\{\bigcup_{j\in I} B_j \mid I \subseteq \{1, 2, \ldots\}\right\}$. Taking $G = \bigcup_{j\in I} B_j$, we have

$$\int_G f\, dP = \sum_{j\in I} \int_{B_j} f\, dP = \sum_{j\in I} \int_{B_j} P(A \mid B_j) dP = \sum_{j\in I} P(A \mid B_j)P(B_j)$$

$$= \sum_{j\in I} P(A \cap B_j) = P\left(A \cap \left(\bigcup_{j\in I} B_j\right)\right) = P(A \cap G)$$

This proves (1).

Example 8.3. If $A \in \mathcal{G}$, then $P(A \mid \mathcal{G}) = \mathbf{1}_A$ a.s.
Recall:

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

Proof: We show that $\mathbf{1}_A$ satisfies conditions (i) and (ii) from the definition of $P(A \mid \mathcal{G})$.

(i) $\mathbf{1}_A$ is $\mathcal{G}$-measurable since $A \in \mathcal{G}$.
(ii) Let $G \in \mathcal{G}$ be arbitrary. Then

$$\int_G \mathbf{1}_A \, dP = \int_\Omega \mathbf{1}_{G \cap A} \, dP = P(G \cap A)$$

Example 8.4. If $\mathcal{G} = \{\emptyset, \Omega\}$, then $P(A \mid \mathcal{G}) = P(A)$ a.s.
Proof: Let $f = P(A)$. We prove that f satisfies conditions (i) and (ii).
(i) f is $\mathcal{G}$-measurable since f is a constant random variable and so $\forall B \in \mathbb{R}$,

$$f^{-1}(B) = \{\omega \in \Omega; f(\omega) \in B\} = \begin{cases} \Omega & \text{if } P(A) \in B \\ \emptyset & \text{if } P(A) \notin B \end{cases} \in \mathcal{G}$$

(ii) We have to show that

$$\int_G f \, dP = P(A \cap G) \quad \forall G \in \mathcal{G} \tag{1}$$

We have two cases:

- $G = \emptyset$. Then
$$\int_G f \, dP = \int_\emptyset P(A) \, dP = 0 = P(A \cap \emptyset) = P(A \cap G)$$
- $G = \Omega$. Then
$$\int_G f \, dP = \int_\Omega P(A) \, dP = P(A) = P(A \cap \Omega) = P(A \cap G)$$

Definition 8.5 We say that event A is *independent* of the σ-field $\mathcal{G}$ if A is independent of G, $\forall G \in \mathcal{G}$, i.e.,

$$P(A \cap G) = P(A) \cdot P(G) \quad \forall G \in \mathcal{G}$$

Observation: Any event A is independent of the trivial σ-field $\mathcal{G} = \{\emptyset, \Omega\}$. (Exercise)

Example 8.6. The event A is independent of $\mathcal{G} \iff P(A \mid \mathcal{G}) = P(A)$ a.s.
Proof: $\Rightarrow$ Assume that A is independent of $\mathcal{G}$. Let $f = P(A)$. We prove that f satisfies conditions (i) and (ii) from the definition of $P(A \mid \mathcal{G})$.
(i) $f = P(A)$ is a constant random variable. Hence, f is $\mathcal{G}$-measurable.
(ii) We have to check that

$$\int_G f \, dP = P(A \cap G) \quad \forall G \in \mathcal{G} \tag{1}$$

Let $G \in \mathcal{G}$ be arbitrary. Then

$$\int_G f \, dP = \int_G P(A) \, dP = P(A) \int_G dP = P(A) \cdot P(G) = P(A \cap G)$$

So (1) holds.
$\Leftarrow$ Suppose that $P(A \mid \mathcal{G}) = P(A)$ a.s. Let $G \in \mathcal{G}$ be arbitrary. Then, by property (ii) of conditional probability, we know that

$$\int_G f \, dP = P(A \cap G), \quad \text{where } f = P(A)$$

Note that

$$\int_G f \, dP = \int_G P(A) \, dP = P(A) \cdot P(G)$$

So, $P(A) \cdot P(G) = P(A \cap G)$.

Definition 8.7 Let $(\Omega, \mathcal{F}, P)$ be a probability space, $A \in \mathcal{F}$. Let $X : \Omega \to \mathbb{R}$ be a random variable (i.e., X is $\mathcal{F}$-measurable).

Let $\mathcal{G} = \sigma(X) = \{X^{-1}(B); B \in \mathcal{B}(\mathbb{R})\}$ where

$$X^{-1}(B) = \{\omega \in \Omega; X(\omega) \in B\} = \{X \in B\}$$

We say that $P(A \mid \mathcal{G})$ is a version of the conditional probability of A given X, and we denote this by $P(A \mid X)$, i.e.,

$$P(A \mid X) := P(A \mid \sigma(\{X\}))$$

This means that:

$$\begin{cases} (i) & P(A \mid X) \text{ is } \sigma(X)\text{-measurable} \\ (ii) & \int_B P(A \mid X)\, dP = P(A \cap \{X \in B\}) \quad \forall B \in \mathcal{B}(\mathbb{R}) \end{cases}$$

Theorem 8.8 — Let $(X, \mathcal{X}, \mu)$ and $(Y, \mathcal{Y}, \nu)$ be measure spaces. μ and ν are σ-finite. $X \times Y = \{(x, y); x \in X, y \in Y\}$.

$$\mathcal{X} \otimes \mathcal{Y} = \sigma(\{A \times B; A \in \mathcal{X}, B \in \mathcal{Y}\}) \quad \text{product } \sigma\text{-field}$$

If $E \in \mathcal{X} \otimes \mathcal{Y}$, then

$$\begin{cases} E_x = \{y \in Y; (x, y) \in E\} & \forall x \in X \\ E^y = \{x \in X; (x, y) \in E\} & \forall y \in Y \end{cases}$$

Proposition 8.9. (i) If $E \in \mathcal{X} \otimes \mathcal{Y}$ then

$$\begin{cases} E_x \in \mathcal{Y} & \forall x \in X \\ E^y \in \mathcal{X} & \forall y \in Y \end{cases}$$

(ii) If $f : X \times Y \to \mathbb{R}$ is $\mathcal{X} \otimes \mathcal{Y}$-measurable then

$$\begin{cases} y \mapsto f(x, y) \text{ is } \mathcal{Y}\text{-measurable} & \forall x \in X \\ x \mapsto f(x, y) \text{ is } \mathcal{X}\text{-measurable} & \forall y \in Y \end{cases}$$

Proposition 8.10. For any set $E \in \mathcal{X} \otimes \mathcal{Y}$

$$\begin{cases} x \mapsto \nu(E_x) \text{ is } \mathcal{X}\text{-measurable} \\ y \mapsto \mu(E^y) \text{ is } \mathcal{Y}\text{-measurable} \end{cases}$$

Define

$$\pi'(E) = \int_X \nu(E_x)\mu(dx) \quad \text{and} \quad \pi''(E) = \int_Y \mu(E^y)\nu(dy)$$

Then π' and π'' are measures on $(X \times Y, \mathcal{X} \otimes \mathcal{Y})$ and

$$\pi'(E) = \pi''(E) =: \pi(E) \quad \forall E \in \mathcal{X} \otimes \mathcal{Y}$$

Moreover, π is the only measure on $X \times Y$ s.t.

$$\pi(A \times B) = \mu(A) \cdot \nu(B) \quad \forall A \in \mathcal{X}, \forall B \in \mathcal{Y}$$

We denote $\pi = \mu \times \nu$ and we say that π is the product measure.

Theorem 8.11 — (i) If $f : X \times Y \to [0, \infty)$ is $\mathcal{X} \otimes \mathcal{Y}$-measurable, then

$$g : X \to \mathbb{R}, \quad g(x) = \int_Y f(x,y)\nu(dy) \text{ is } \mathcal{X}\text{-measurable}$$

$$h : Y \to \mathbb{R}, \quad h(y) = \int_X f(x,y)\mu(dx) \text{ is } \mathcal{Y}\text{-measurable}$$

and

$$\begin{aligned} \int_X \left(\int_Y f(x,y)\nu(dy) \right) \mu(dx) &= \int_Y \left(\int_X f(x,y)\mu(dx) \right) \nu(dy) \\ &= \int_{X \times Y} f(x,y)(\mu \times \nu)(dx, dy) \end{aligned} \tag{4}$$

(ii) If $f : X \times Y \to \mathbb{R}$ is $\mathcal{X} \otimes \mathcal{Y}$-measurable and integrable w.r.t. $\mu \times \nu$, then

$$\begin{cases} g(x) \text{ is finite for } \mu\text{-almost all } x \in X, \quad g \text{ is } \mathcal{X}\text{-measurable} \\ h(y) \text{ is finite for } \nu\text{-almost all } y \in Y, \quad h \text{ is } \mathcal{Y}\text{-measurable} \end{cases}$$

and (4) holds.

§9 February 12, 2024

§9.1 Conditional probability continued

Theorem 9.1 — Let X and Y be independent random variables and $\mu = P \circ X^{-1}$, $\nu = P \circ Y^{-1}$. Then

a)
$$P((X,Y) \in B) = \int_{\mathbb{R}} P((x,Y) \in B)\mu(dx) \quad \forall B \in \mathbb{R}^2 \tag{2}$$

b)
$$P((X \in A, (X,Y) \in B) = \int_{\mathbb{R}} P((x,Y) \in B)\mu(dx) \quad \forall A \in \mathbb{R} \quad \forall B \in \mathbb{R}^2 \tag{4}$$

Proof. a) Since X, Y are independent, the law of (X,Y) is $\mu \times \nu$, i.e.,

$$P \circ (X,Y)^{-1} = (P \circ X^{-1}) \times (P \circ Y^{-1}) = \mu \times \nu$$

Recall:
$$B_x = \{y \in \mathbb{R}; (x,y) \in B\} \text{ is the section of } B \text{ at } x$$

By Fubini's Theorem,
$$(\mu \times \nu)(B) = \int_{\mathbb{R}} \nu(B_x)\mu(dx) \tag{1}$$

Note that
$$(\mu \times \nu)(B) = P((X,Y) \in B)$$
$$\nu(B_x) = (P \circ Y^{-1})(B_x) = P(Y \in B_x) = P(\{\omega \in \Omega; Y(\omega) \in B_x\})$$

So
$$\nu(B_x) = P(\{\omega \in \Omega; (x, Y(\omega)) \in B\}) = P((x,Y) \in B)$$

Hence (1) gives our desired conclusion for a). □

Proof. b) We write (1) for set B replaced by $B' = (A \times \mathbb{R}) \cap B$, relation (1) becomes:

$$(\mu \times \nu)(B') = \int_{\mathbb{R}} \nu(B'_x)\mu(dx) \tag{3}$$

Note that

$$(\mu \times \nu)(B') = (P \circ (X,Y)^{-1})(B') = P((X,Y) \in B') = P((X,Y) \in (A \times \mathbb{R}) \cap B) = P(X \in A, (X,Y) \in B) = \text{LHS of (4)}$$

$$B'_x = \{y \in \mathbb{R}; (x,y) \in B'\} = \{y \in \mathbb{R}; x \in A \text{ and } (x,y) \in B\} = \begin{cases} \emptyset & \text{if } x \notin A \\ B_x & \text{if } x \in A \end{cases}$$

$$\nu(B'_x) = \begin{cases} 0 & \text{if } x \notin A \\ \nu(B_x) & \text{if } x \in A \end{cases}$$

So
$$\nu(B'_x) = \begin{cases} 0 & \text{if } x \notin A \\ P((x,Y) \in B) & \text{if } x \in A \end{cases}$$

Relation (3) gives exactly (4).

□

Theorem 9.2 — Let X and Y be independent random variables, and $J \subseteq \mathbb{R}$. Consider the function

$$f(x) = P((x,Y) \in J) \quad \text{for all } x \in \mathbb{R}.$$

a) Then

$$P((X,Y) \in J \mid X) = f(X) \quad \text{a.s.}$$

b) Let $M = \max(X,Y)$. Then for all $m \in \mathbb{R}$,

$$P(M \le m \mid X) = \mathbf{1}\{X \le m\} P(Y \le m) \quad \text{a.s.}$$

Proof. a) We check that $f(X)$ satisfies conditions (i) and (ii) from the definition of conditional probability. Here $\mathcal{G} = \sigma(X)$.

(i) $f(X)$ is $\sigma(X)$-measurable. This is clear.

(ii) Let $G \in \sigma(X)$ be arbitrary. Then $G = \{X \in H\}$ for some $H \in \mathcal{B}(\mathbb{R})$. Let $P \circ X^{-1} = \mu$.

$$\int_G f(X)\, dP = \int_{\{X \in H\}} f(X)\, dP = \int_H f(x)\, \mu(dx) \quad \text{(change of variable, Th 16.13)}$$

$$\int_G f(X)\, dP = \int_\Omega f(X(\omega)) \mathbf{1}_G(\omega)\, dP(\omega) = \int_H f(x)\, \mu(dx) = \int_H P((x,Y) \in J)\, \mu(dx) \quad \text{(definition of } f)$$
$$= P(X \in H, (X,Y) \in J) \quad \text{(by (4))}$$

In summary, we proved that:

$$\int_G f(X)\, dP = P(A \cap G) \quad \forall G \in \sigma(X)$$

□

Proof. b) We use the result in part a). Note that

$$\{M \le m\} = \{\max(X,Y) \le m\} = \{X \le m, Y \le m\} = \{(X,Y) \in J\}$$

where $J = \{(x,y) \in \mathbb{R}^2; x \le m \text{ and } y \le m\}$.

By a),

$$P(M \le m \mid X) = P((X,Y) \in J \mid X) = f(X) \quad \text{a.s.} \tag{5}$$

where $f(x) = P((x,Y) \in J)$.

Let us calculate $f(x)$:

$$f(x) = P((x,Y) \in J) = P(\{\omega \in \Omega; x \le m \text{ and } Y(\omega) \le m\})$$
$$= \begin{cases} 0 & \text{if } x > m \\ P(Y \le m) & \text{if } x \le m \end{cases} = \mathbf{1}_{\{x \le m\}} P(Y \le m)$$

Then

$$f(x) = \mathbf{1}_{\{x \le m\}} P(Y \le m)$$

Relation (5) becomes:

$$P(M \le m \mid X) = \mathbf{1}_{\{X \le m\}} P(Y \le m).$$

□

Recall: (MAT 5170): A family $\mathcal{P}$ of subsets of a set Ω is called a π-system if it is closed under finite intersections, i.e., if $A, B \in \mathcal{P}$ then $A \cap B \in \mathcal{P}$.

If μ and ν are measures on $(\Omega, \mathcal{F})$ and $\mu(A) = \nu(A)$ for all $A \in \mathcal{P}$, then $\mu = \nu$.

Theorem 9.3 — Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ is a sub σ-field of $\mathcal{F}$, $A \in \mathcal{F}$. Assume that $\mathcal{G} = \sigma(\mathcal{P})$ where $\mathcal{P}$ is a π-system and $\Omega = \bigcup_{i \geq 1} A_i$ with $A_i \in \mathcal{P}$.
Let $f : \Omega \to [0, \infty)$ be a function which satisfies:

(i) f is $\mathcal{G}$-measurable and integrable

(ii) $\int_G f \, dP = P(A \cap G) \quad \forall G \in \mathcal{P}$

Then $f = P(A \mid \mathcal{G})$ a.s.

Proof. Define

$$\mu(G) = \int_G f \, dP, \quad G \in \mathcal{G}$$

$$\nu(G) = P(A \cap G), \quad G \in \mathcal{G}$$

Both μ and ν are measures on $(\Omega, \mathcal{G})$.

By (ii), $\mu(G) = \nu(G) \quad \forall G \in \mathcal{P}$.

Hence, by Theorem 10.4, $\mu(G) = \nu(G) \quad \forall G \in \mathcal{G}$. The conclusion follows since f satisfies the two conditions (i) and (ii) from the definition of $P(A \mid \mathcal{G})$.

The next result shows that $P(\cdot \mid \mathcal{G})$ satisfies the same properties as the classical probability measure P. □

Theorem 9.4 — Theorem 33.2 (Properties of Conditional Probability) Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-field.

1) $P(\emptyset \mid \mathcal{G}) = 0$ a.s. and $P(\Omega \mid \mathcal{G}) = 1$ a.s.

2) $P(A \mid \mathcal{G}) \geq 0$ a.s. and $P(A \mid \mathcal{G}) \leq 1$ a.s. $\forall A \in \mathcal{F}$

3) If $\{A_n\}_{n \geq 1}$ are disjoint sets in $\mathcal{F}$, then

$$P\left(\bigcup_{n \geq 1} A_n \mid \mathcal{G}\right) = \sum_{n \geq 1} P(A_n \mid \mathcal{G}) \quad \text{a.s.}$$

4) If $A, B \in \mathcal{F}$ and $A \subseteq B$, then

$$P(B \setminus A \mid \mathcal{G}) = P(B \mid \mathcal{G}) - P(A \mid \mathcal{G}) \quad \text{a.s.}$$

$$P(A \mid \mathcal{G}) \leq P(B \mid \mathcal{G}) \quad \text{a.s.}$$

5) *Inclusion-exclusion principle:* For any $A_1, \ldots, A_n \in \mathcal{F}$,

$$P\left(\bigcup_{i=1}^n A_i \mid \mathcal{G}\right) = \sum_{i=1}^n P(A_i \mid \mathcal{G}) - \sum_{i<j} P(A_i \cap A_j \mid \mathcal{G}) + \ldots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i \mid \mathcal{G}\right) \quad \text{a.s.}$$

6) If $\{A_n\}_{n \geq 1}$ are subsets of $\mathcal{F}$ such that $A_n \uparrow A \in \mathcal{F}$ (i.e., $A_n \subseteq A_{n+1}$ and $A = \bigcup_{n \geq 1} A_n$), then

$$P(A_n \mid \mathcal{G}) \uparrow P(A \mid \mathcal{G}) \quad \text{a.s.}$$

Similarly, if $A_n \downarrow A$ (i.e., $A_n \supseteq A_{n+1}$ and $A = \bigcap_{n \geq 1} A_n$), then

$$P(A_n \mid \mathcal{G}) \downarrow P(A \mid \mathcal{G}) \quad \text{a.s.}$$

7) If $A \in \mathcal{F}$ is such that $P(A) = 1$, then $P(A \mid \mathcal{G}) = 1$ a.s.
If $A \in \mathcal{F}$ is such that $P(A) > 0$, then $P(A \mid \mathcal{G}) > 0$ a.s.

Proof. 1) 1 is trivial: $f = 0$ satisfies conditions (i) and (ii) from the definition of $P(\emptyset \mid \mathcal{G})$.

$$f = 1 \quad \text{satisfies } P(\Omega \mid \mathcal{G})$$

2) Use the following result: If $f : \Omega \to \mathbb{R}$ is a $\mathcal{G}$-measurable function and

$$\int_G f\,dP \geq 0 \quad \forall G \in \mathcal{G} \text{ then } f \geq 0 \text{ a.s.} \quad \text{(Section 15)}$$

In our case, $f = P(A \mid \mathcal{G})$ satisfies:

$$\int_G f\,dP = P(A \cap G) \geq 0 \quad \forall G \in \mathcal{G}. \text{ Hence, } f \geq 0 \text{ a.s.}$$

Similarly, the function $f' = 1 - P(A \mid \mathcal{G})$ satisfies:

$$\int_G f'\,dP = \int_G (1 - P(A \mid \mathcal{G}))\,dP = P(G) - \int_G P(A \mid \mathcal{G})\,dP = P(G) - P(A \cap G) = P(G \setminus A) \geq 0$$

Hence $f' \geq 0$ a.s., that is $P(A \mid \mathcal{G}) \leq 1$ a.s.

3) Let $f = \sum_{n\geq 1} P(A_n \mid \mathcal{G})$. We check that f satisfies conditions (i) and (ii) from the definition of $P(\bigcup_{n\geq 1} A_n \mid \mathcal{G})$.

(i) f is $\mathcal{G}$-measurable (limit of a seq. of $\mathcal{G}$-measurable functions is $\mathcal{G}$-measurable).

(ii) Let $G \in \mathcal{G}$ be arbitrary, and denote $A = \bigcup_{n\geq 1} A_n$. We want to prove that:

$$\int_G f\,dP = P(A \cap G) \tag{7}$$

$$\int_G f\,dP = \int_G \sum_{n\geq 1} P(A_n \mid \mathcal{G})\,dP \geq 0 \quad \text{(Corollary to Theorem 16.7)}$$

$$\int_G \sum_{n\geq 1} P(A_n \mid \mathcal{G})\,dP = \sum_{n\geq 1} \int_G P(A_n \mid \mathcal{G})\,dP = \sum_{n\geq 1} P(A_n \cap G) \quad \text{(by condition (ii) in the def. of } P(A_n \mid \mathcal{G}))$$

$$= P\left(\bigcup_{n\geq 1} (A_n \cap G)\right) = P\left(\left(\bigcup_{n\geq 1} A_n\right) \cap G\right) = P(A \cap G)$$

This proves (7).

4) - 7) Exercise.

□

§10 February 14, 2024

§10.1 Conditional Distributions continued

Theorem 10.1 — Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : \Omega \to \mathbb{R}$ is a random variable, and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-field. Then there exists a function $\mu(H, \omega)$ defined for any $H \in \mathcal{B}(\mathbb{R}), \omega \in \Omega$ such that the following conditions hold:

(a) $\mu(\cdot, \omega)$ is a probability measure on $\mathbb{R}$, $\forall \omega \in \Omega$

(b) $\mu(H, \cdot)$ is a version of $P(X \in H \mid \mathcal{G})$, $\forall H \in \mathcal{B}(\mathbb{R})$

We say that μ is the conditional distribution of X given $\mathcal{G}$. In particular, if $\mathcal{G} = \sigma(Y)$, we say that μ is the conditional distribution of X given Y.

For each $r \in \mathbb{Q}$, let $F(r, \cdot)$ be a version of $P(X \leq r \mid \mathcal{G})$, i.e.,

$$F(r, \omega) = P(X \leq r \mid \mathcal{G})(\omega) \quad \text{for } P\text{-almost all } \omega \in \Omega.$$

Properties of F:
1) If $r, s \in \mathbb{Q}$ with $r \leq s$, then $F(r, \omega) \leq F(s, \omega)$ with probability 1.

$$P(X \leq r \mid \mathcal{G})(\omega) \leq P(X \leq s \mid \mathcal{G})(\omega) \quad \text{since } \{X \leq r\} \subseteq \{X \leq s\}.$$

Let $E_{r,s} = \{\omega \in \Omega; F(r, \omega) \leq F(s, \omega)\}$.
Then $E_{r,s} \in \mathcal{G}$ and $P(E_{r,s}) = 1$.
2) For every $r \in \mathbb{Q}$ fixed,

$$\lim_{n\to\infty} F\left(r + \frac{1}{n}, \omega\right) = \lim_{n\to\infty} P\left(X \leq r + \frac{1}{n} \mid \mathcal{G}\right)(\omega) = P(X \leq r \mid \mathcal{G})(\omega) = F(r, \omega)$$

by property 6) in Theorem 33.2.
Let $E_r = \left\{\omega \in \Omega; \lim_{n\to\infty} F\left(r + \frac{1}{n}, \omega\right) = F(r, \omega)\right\}$. Then $E_r \in \mathcal{G}$ with $P(E_r) = 1$.
3)

$$\lim_{r\to\infty} F(r, \omega) = \lim_{r\to\infty} P(X \leq r \mid \mathcal{G})(\omega) = P(\Omega \mid \mathcal{G})(\omega) = 1 \quad \text{with probability 1}$$

$$\{\{X \leq r\}\}_{r\in\mathbb{Q}} \uparrow \Omega$$

Let $D_1 = \{\omega \in \Omega; \lim_{r\to\infty} F(r, \omega) = 1\}$. Then $D_1 \in \mathcal{G}$ and $P(D_1) = 1$.
4)

$$\lim_{r\to-\infty} F(r, \omega) = \lim_{r\to-\infty} P(X \leq r \mid \mathcal{G})(\omega) = P(\emptyset \mid \mathcal{G})(\omega) = 0 \quad \text{with probability 1}$$

$$\{\{X \leq r\}\}_{r\in\mathbb{Q}} \downarrow \emptyset$$

Let $D_2 = \{\omega \in \Omega; \lim_{r\to-\infty} F(r, \omega) = 0\}$. Then $D_2 \in \mathcal{G}$ and $P(D_2) = 1$.
Let $S = \left(\bigcap_{r\in\mathbb{Q}} E_r\right) \cap \left(\bigcap_{r,s\in\mathbb{Q}} E_{r,s}\right) \cap D_1 \cap D_2$. Then $S \in \mathcal{G}$ and $P(S) = 1$.

- For $\omega \in S$, extend $F(r, \omega)$ to $\mathbb{R}$ by setting

$$\bar{F}(x, \omega) := \inf_{r > x, r\in\mathbb{Q}} F(r, \omega)$$

  Clearly, if $x \in \mathbb{Q}$ then $\bar{F}(x, \omega) = F(x, \omega)$.

- For $\omega \notin S$, let $\bar{F}(\cdot, \omega) := F^*$ where F^* is a fixed cumulative distribution function on $\mathbb{R}$.

- For $\omega \in S$, we check that $\bar{F}(\cdot, \omega) : \mathbb{R} \to [0, 1]$ is a probability distribution function:
  (a) right-continuity: $\lim_{n\to\infty} \bar{F}(x_n, \omega) = \bar{F}(x, \omega)$ if $x_n \uparrow x$
  (b) non-decreasing: if $x \leq y$, then $\bar{F}(x, \omega) \leq \bar{F}(y, \omega)$

(c) $\lim_{x\to\infty} \bar{F}(x,\omega) = 1$

(d) $\lim_{x\to-\infty} \bar{F}(x,\omega) = 0$

Hence, by Theorem 1.2, there exists a unique probability measure $\bar{\mu}(\cdot,\omega)$ on $\mathbb{R}$ such that

$$\bar{\mu}((-\infty,x],\omega) = \bar{F}(x,\omega) \quad \forall x \in \mathbb{R}$$

- For $\omega \notin S$, let $\bar{\mu}^*$ be the probability measure corresponding to F^*, i.e.

$$\bar{\mu}^*((-\infty,x]) = F^*(x) = F^*(x) \quad \forall x \in \mathbb{R}$$

Define

$$\mu(H,\omega) = \begin{cases} \bar{\mu}(H,\omega) & \text{if } \omega \in S \\ \bar{\mu}^*(H) & \text{if } \omega \notin S \end{cases}$$

Then $\mu(H,\omega)$ is a probability measure on $\mathbb{R}$ $\forall \omega \in \Omega$, i.e. condition (a) holds.

We now prove that μ satisfies condition (b):

We will prove that $\mu(H,\cdot) = P(X \in H \mid \mathcal{G})$ a.s. by checking that $\mu(H,\cdot)$ satisfies conditions (i) and (ii) from the definition of $P(X \in H \mid \mathcal{G})$.

(i) *We have to prove that $\mu(H,\cdot)$ is $\mathcal{G}$-measurable,* $\forall H \in \mathcal{B}(\mathbb{R})$.

Let $\mathcal{L} = \{H \in \mathcal{B}(\mathbb{R}); \mu(H,\cdot) \text{ is } \mathcal{G}\text{-measurable}\}$ is a λ-system, i.e.

1) $\mathbb{R} \in \mathcal{L}$

2) If $H \in \mathcal{L}$ then $H^c \in \mathcal{L}$

3) If $(H_n)_{n\geq 1}$ are disjoint then $\bigcup_{n\geq 1} H_n \in \mathcal{L}$

$\mathcal{P} = \{(-\infty,r]; r \in \mathbb{Q}\}$ is a π-system, i.e.

- if $A_1, A_2, \ldots, A_n \in \mathcal{P}$ then $A_1 \cap A_2 \cap \ldots \cap A_n \in \mathcal{P}$

$\mathcal{P} \subseteq \mathcal{L}$ since $\mu((-\infty,r],\cdot) = F(r,\cdot) = P(X \leq r \mid \mathcal{G})(\cdot)$ if $\omega \in S$, and hence $\mu((-\infty,r],\cdot) = P(X \leq r \mid \mathcal{G})$ with probability 1.

Because $P(X \leq r \mid \mathcal{G})$ is $\mathcal{G}$-measurable, it follows that $\mu((-\infty,r],\cdot)$ is $\mathcal{G}$-measurable.

To summarize, we have:

$$\mathcal{L} = \lambda\text{-system}$$

$$\mathcal{P} = \pi\text{-system}$$

$\mathcal{P} \subseteq \mathcal{L}$

Then, by Dynkin's π-λ theorem (Theorem 3), it follows that:

$$\sigma(\mathcal{P}) = \mathcal{L}$$

Hence,

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{P}) \subseteq \mathcal{L} \subseteq \mathcal{B}(\mathbb{R}) \text{ i.e. } \mathcal{L} = \mathcal{B}(\mathbb{R})$$

This means that $\mu(H,\cdot)$ is $\mathcal{G}$-measurable $\forall H \in \mathcal{B}(\mathbb{R})$.

(ii) *We want to prove that*

$$P(\{X \in H\} \cap G) = \int_G \mu(H,\omega) P(d\omega) \quad \forall G \in \mathcal{G}, \forall H \in \mathcal{B}(\mathbb{R})$$

$$P(\{X \in H\} \cap G) = \int_G \mu(H,\omega) P(d\omega) \quad \forall G \in \mathcal{G}, \forall H \in \mathcal{B}(\mathbb{R})$$

Fix $G \in \mathcal{G}$. Define

$$\varphi_1(H) = P(\{X \in H\} \cap G)$$

$$\varphi_2(H) = \int_G \mu(H,\omega) P(d\omega)$$

Note that $\varphi_1(H) = \varphi_2(H) \forall H \in \mathcal{P}$, since if $H = (-\infty, r]$ with $r \in \mathbb{Q}$

$$\varphi_1((-\infty, r]) = P(\{X \le r\} \cap G)$$
$$\varphi_2((-\infty, r]) = \int_G \mu((-\infty, r], \omega) P(d\omega)$$

$$\begin{aligned}\varphi_2((-\infty, r]) &= \int_G \mu((-\infty, r], \omega) P(d\omega) = \int_G F(r, \omega) P(d\omega) = \int_G P(X \le r \mid \mathcal{G})(\omega) P(d\omega) \\ &= P(X \le r \mid \mathcal{G}) P(d\omega) = P(\{X \le r\} \cap G)\end{aligned}$$

By the definition of conditional probability.

Since $\mathcal{P}$ is a π-system, $\varphi_1(H) = \varphi_2(H) \forall H \in \mathcal{B}(\mathbb{R})$.

$$\begin{aligned}\varphi_1((-\infty, r]) &= P(\{X \le r\} \cap G) \\ \varphi_2((-\infty, r]) &= \int_G \mu((-\infty, r], \omega) P(d\omega) = \int_G F(r, \omega) P(d\omega) = \int_G P(X \le r \mid \mathcal{G})(\omega) P(d\omega) \\ &= P(X \le r \mid \mathcal{G}) P(d\omega) = P(\{X \le r\} \cap G)\end{aligned}$$

By the definition of conditional probability.

Since $\mathcal{P}$ is a π-system, $\varphi_1(H) = \varphi_2(H) \forall H \in \mathcal{B}(\mathbb{R})$.

$$P(\{X \in H\} \cap G) = \int_G \mu(H, \omega) P(d\omega) \quad \forall G \in \mathcal{G}, \forall H \in \mathcal{B}(\mathbb{R}) \quad \square$$

Example 10.2. Let X, Y be r.v.'s on $(\Omega, \mathcal{F}, P)$ s.t. the law of (X, Y) has density $f(x, y)$, i.e.

$$P((X, Y) \in A) = \int_A f(x, y)\, dx\, dy \quad \forall A \subseteq \mathbb{R}^2$$

Let $f_X(x) = \int_{\mathbb{R}} f(x, y)\, dy$ be the marginal density of X:

$$P(X \in B) = \int_B f_X(x)\, dx \quad \forall B \subseteq \mathbb{R}$$

Define

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \quad \text{if } f_X(x) \neq 0$$

Observation:

$$\int_{\mathbb{R}} f_{Y|X}(y \mid x)\, dy = 1 \quad \text{(exercise)}$$

Define

$$Q(x, H) = \begin{cases} \int_H f_{Y|X}(y \mid x)\, dy & \text{if } f_X(x) \neq 0 \\ Q^*(H) & \text{if } f_X(x) = 0 \end{cases}$$

Set

$$\mu(H, \omega) = Q(X(\omega), H)$$

Claim: $\mu(H, \omega)$ is the conditional distribution of Y given X.

Proof of this claim: We check properties a) and b) of Theorem 33.3
a) $\mu(\cdot, \omega) = Q(X(\omega), \cdot)$ is indeed a probability measure $\forall \omega \in \Omega$
b) We have to check that $\mu(H, \cdot)$ is a version of $P(Y \in H \mid X)$, i.e.

$$\mu(H, \cdot) = P(Y \in H \mid X) \quad \text{a.s.}$$

For this, we have to check that conditions (i) and (ii) are verified:

(i) $\mu(H, \cdot) = Q(X(\cdot), H)$ is $\sigma(X)$-measurable. This is clear since Q is a function of X.

(ii) We have to prove that

$$P(\{Y \in H\} \cap G) = \int_G \mu(H,\omega) P(d\omega) \quad \forall G \in \sigma(X) = \mathcal{G} \quad (2)$$

Let us prove (2). Let $G = \{X \in E\} \in \sigma(X)$ be arbitrary, with $E \in \mathcal{B}(\mathbb{R})$. Then

Let $G = \{X \in E\} \in \sigma(X)$ be arbitrary, with $E \in \mathcal{R}$.

$$\begin{aligned}
&\int_G \mu(H,\omega)\, P(d\omega) = \int_{\{X \in E\}} Q(X(\omega), H)\, P(d\omega) \\
&= \int_{\{X \in E\}} 1_E(X(\omega))\, Q(X(\omega), H)\, P(d\omega) \\
&= \int_\Omega 1_E(X(\omega))\, Q(X(\omega), H)\, P(d\omega) \\
&= \int_E Q(x, H)\, (P \circ X^{-1})(dx) \quad \text{(change of variables theorem 16.13)} \\
&= \int_E Q(x, H)\, f_X(x)\, dx \\
&= \int_{E \cap \{f_X(x) \neq 0\}} Q(x, H)\, f_X(x)\, dx \\
&= \int_{E \cap \{f_X(x) \neq 0\}} \left(\int_H f_{Y|X}(y|x)\, dy \right) f_X(x)\, dx \\
&= \int_{E \cap \{f_X(x) \neq 0\}} \int_H f(x,y)\, dy\, dx \\
&= \int_E \int_H f(x,y)\, dy\, dx \\
&= P((X,Y) \in E \times H) \\
&= P(\{X \in E\} \cap \{Y \in H\}) \quad \text{(by definition of } E \text{ and } H\text{)}
\end{aligned}$$

§11 February 28, 2024

§11.1 Conditional Expectation

⋆ **Recall**: We say that a r.v. $P(A|\mathcal{G})$ is the conditional probability of A given $\mathcal{G}$ if:

1. $P(A|\mathcal{G})$ is $\mathcal{G}$-measurable and integrable
2. $\int_G P(A|\mathcal{G})\,dP = P(A \cap G) \quad \forall G \in \mathcal{G}$

Note that $P(A \cap G) = \int_G \mathbf{1}_A\,dP$, (ii) can be stated as:

$$\int_G P(A|\mathcal{G})\,dP = \int_G \mathbf{1}_A\,dP \quad \forall G \in \mathcal{G}$$

Theorem 11.1 — Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-field, and $X : \Omega \to \mathbb{R}$ an integrable r.v. Then, there exists a r.v. $g : \Omega \to \mathbb{R}$ such that:

1. g is $\mathcal{G}$-measurable and integrable
2. $\int_G g\,dP = \int_G X\,dP \quad \forall G \in \mathcal{G}$

If $g' : \Omega \to \mathbb{R}$ is another r.v. satisfying (i) and (ii), then $g = g'$ a.s., i.e.

$$P(\{\omega \in \Omega;\, g(\omega) = g'(\omega)\}) = 1$$

We say that g is a (version of) the conditional expectation of X given $\mathcal{G}$, and we denote

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

Proof. **Proof: Existence Case 1, $X \geq 0$**

Define

$$\mathcal{D}(G) = \int_G X\,dP \quad \text{for all } G \in \mathcal{G}.$$

Clearly, $\mathcal{D}$ is a measure on $(\Omega, \mathcal{G})$.

Note that $\mathcal{D}$ is a finite measure:

$$\mathcal{D}(\Omega) = \int_\Omega X\,dP = \mathbb{E}(X) < \infty.$$

Moreover, $\mathcal{D}$ is absolutely continuous with respect to P:

$$\text{if } P(G) = 0 \text{ then } \mathcal{D}(G) = 0.$$

By the Radon-Nikodym Theorem (Theorem 32.3), there exists a $\mathcal{G}$-measurable function $g : \Omega \to \mathbb{R}$ such that:

$$\mathcal{D}(G) = \int_G g\,dP \quad \forall G \in \mathcal{G}.$$

From (1) and (2),

$$\int_G X\,dP = \int_G g\,dP \quad \forall G \in \mathcal{G}.$$

Thus, g is clearly integrable. So, g satisfies (i) and (ii). □

Proof. **Case 2: X is arbitrary**

Recall that any $a \in \mathbb{R}$ can be written as:

$$a = a^+ - a^- \quad \text{where} \quad a^+ = \begin{cases} a & \text{if } a \geq 0 \\ 0 & \text{if } a < 0 \end{cases}, \quad a^- = \begin{cases} 0 & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$$

(Note: $a^+ \geq 0, a^- \geq 0$)

Hence, for $X(\omega) \in \mathbb{R}$, we have:

$$X(\omega) = X^+(\omega) - X^-(\omega) \quad \forall \omega \in \Omega.$$

Both X^+ and X^- are non-negative r.v.'s. By Case 1,

- there exists a function $g_1 : \Omega \to \mathbb{R}$ $\mathcal{G}$-measurable and integrable s.t.

$$\int_G g_1 \, dP = \int_G X^+ \, dP \quad \forall G \in \mathcal{G} \tag{3}$$

- there exists a function $g_2 : \Omega \to \mathbb{R}$ $\mathcal{G}$-measurable and integrable s.t.

$$\int_G g_2 \, dP = \int_G X^- \, dP \quad \forall G \in \mathcal{G} \tag{4}$$

Take the difference between (3) and (4), we get:

$$\int_G (g_1 - g_2) \, dP = \int_G (X^+ - X^-) \, dP = \int_G X \, dP \quad \forall G \in \mathcal{G}.$$

Taking $g = g_1 - g_2$, we see that g satisfies (i) and (ii). □

Lemma 11.2 — Lemma 1 If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$ a.s. (and integrable)

Proof. It is clear that $g = X$ satisfies (ii) and (iii) of Theorem 1. □ □

Lemma 11.3 — Lemma 2 If X is independent of $\mathcal{G}$ (i.e. $\{X \in B\}$ and G are independent for any $B \in \mathcal{R}, G \in \mathcal{G}$), then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$ a.s.

Proof. We check that $g = \mathbb{E}(X)$ satisfies (i) and (ii) from Theorem 1:

(i) $g = \mathbb{E}(X)$ is a constant r.v., so it is measurable w.r.t. any σ-field, and in particular it is $\mathcal{G}$-measurable. Clearly, g is integrable.

(ii)

$$\int_G g \, dP = \int_G \mathbb{E}(X) \, dP = \mathbb{E}(X) \int_G dP = \mathbb{E}(X) \cdot P(G) \quad \forall G \in \mathcal{G}.$$

$$\int_G X \, dP = \int_\Omega 1_G X \, dP = \mathbb{E}(1_G X) = \mathbb{E}(1_G) \cdot \mathbb{E}(X) = P(G) \cdot \mathbb{E}(X) \quad \text{for any } G \in \mathcal{G}.$$

(independent since X is indep. of $\mathcal{G}$)

□

Example 11.4. Let X be an integrable r.v. on $(\Omega, \mathcal{F}, P)$ and $\mathcal{G} = \sigma(\{B_i\}_{i\geq 1})$ where $\{B_i\}_{i\geq 1}$ is a partition of Ω, with $P(B_i) > 0$. Recall that an arbitrary set in $\mathcal{G}$ is of the form $G = \bigcup_{i\in I} B_i$ for some $I \subset \{1, 2, \ldots\}$. **Find** $\mathbb{E}(X|\mathcal{G})$.

Solution It can be proved that since $\mathbb{E}(X|\mathcal{G})$ is $\mathcal{G}$-measurable and $\mathcal{G} = \sigma(\{B_i\}_{i\geq 1})$, then

$$\mathbb{E}(X|\mathcal{G}) = \sum_{i\geq 1} \alpha_i 1_{B_i}$$

for some $\alpha_i \in \mathbb{R}$.

Let us find the constants $\alpha_i \in \mathbb{R}$. We write property (ii) for $G = B_i$:

$$\int_{B_i} \alpha_i \, dP = \int_{B_i} X \, dP,$$

i.e. $\alpha_i \int_{B_i} dP = \int_{B_i} X\,dP$, or equivalently $\alpha_i P(B_i) = \int_{B_i} X\,dP$. So $\alpha_i = \frac{1}{P(B_i)} \int_{B_i} X\,dP$.
Hence,

$$\mathbb{E}(X|\mathcal{G}) = \sum_{i\geq 1} \left(\frac{1}{P(B_i)} \int_{B_i} X\,dP \right) 1_{B_i}.$$

Remark: If there exist some $i \geq 1$ such that $P(B_i) = 0$, for those values i we can choose $d_i \in \mathbb{R}$ arbitrarily. In that case,

$$\mathbb{E}(X|\mathcal{G}) = \sum_{\{i\geq 1; P(B_i)>0\}} \left(\frac{1}{P(B_i)} \int_{B_i} X\,dP \right) 1_{B_i} + \sum_{\{i\geq 1; P(B_i)=0\}} d_i 1_{B_i}$$

Example 11.5. For any event $A \in \mathcal{F}$ and for any σ-field $\mathcal{G} \subset \mathcal{F}$,

$$\mathbb{E}(1_A|\mathcal{G}) = P(A|\mathcal{G}) \text{ a.s.}$$

Proof: We show that $g = P(A|\mathcal{G})$ satisfies (i) and (ii) in Theorem 1:

(i) g is $\mathcal{G}$-measurable (clear).

(ii) $\int g\,dP = \int P(A|\mathcal{G})\,dP = P(A \cap G) = \int 1_A\,dP \quad \forall G \in \mathcal{G}$.

Theorem 11.6 — Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : \Omega \to \mathbb{R}$ an integrable random variable. Suppose that $\mathcal{G} = \sigma(\mathcal{P})$ where

$$\mathcal{P} \text{ is a } \pi\text{-system, i.e., if } A, B \in \mathcal{P} \text{ then } A \cap B \in \mathcal{P}$$

and

$$\Omega = \bigcup_{i\geq 1} P_i \text{ for some } P_i \in \mathcal{P}.$$

Let $g : \Omega \to \mathbb{R}$ be a function which satisfies:

$$\begin{cases} (i) & g \text{ is } \mathcal{G}\text{-measurable and integrable} \\ (ii)' & \int_G g\,dP = \int_G X\,dP \quad \forall G \in \mathcal{P} \end{cases}$$

Then $g = \mathbb{E}(X|\mathcal{G})$ a.s.

Proof.

$$\int_G g\,dP = \int_G X\,dP = \int_G \mathbb{E}(X|\mathcal{G})\,dP \quad \forall G \in \mathcal{P}.$$

By **Theorem 16.10(iii)**, $g = \mathbb{E}(X|\mathcal{G})$ a.s. □ □

Theorem 11.7 — Properties of Conditional Expectation Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-field; let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ be integrable random variables.
If $X = a$ a.s. where $a \in \mathbb{R}$, then $\mathbb{E}(X|\mathcal{G}) = a$ a.s.

(ii) (Linearity) $\mathbb{E}(aX + bY|\mathcal{G}) = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$ a.s. $\quad \forall a, b \in \mathbb{R}$

(iii) (Monotonicity) If $X \leq Y$ a.s., then $\mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(Y|\mathcal{G})$ a.s.

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

Proof. (i) Clearly $g = a$ satisfies (i) and (ii) from Theorem 1.

(ii) We let $g = a\mathbb{E}(X|\mathcal{G}) + b\mathbb{E}(Y|\mathcal{G})$. We show that g satisfies properties (i) and (ii) from the definition of $\mathbb{E}(aX + bY|\mathcal{G})$ (Theorem 1):

(a) g is $\mathcal{G}$-measurable. This is clear since g is a linear combination of $\mathcal{G}$-measurable functions. Similarly, g is integrable.

(b) $\int_G g\,dP = \int_G (a\mathbb{E}(X|\mathcal{G})+b\mathbb{E}(Y|\mathcal{G}))\,dP = a\int_G \mathbb{E}(X|\mathcal{G})\,dP + b\int_G \mathbb{E}(Y|\mathcal{G})\,dP = a\int_G X\,dP + b\int_G Y\,dP = \int_G (aX+bY)\,dP$

$\forall G \in \mathcal{G}$.

(iii) $(\mathbb{E}(Y|\mathcal{G}) - \mathbb{E}(X|\mathcal{G}))\,dP = \int_G \mathbb{E}(Y|\mathcal{G})dP - \int_G \mathbb{E}(X|\mathcal{G})dP = \int_G Y\,dP - \int_G X\,dP = \int_G (Y - X)\,dP \geq 0$

for all $G \in \mathcal{G}$. Hence $\mathbb{E}(Y|\mathcal{G}) - \mathbb{E}(X|\mathcal{G}) \geq 0$ a.s.

(iv)

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

This is true because

$$-|X| \leq X \leq |X|$$

and then we apply monotonicity:

$$\mathbb{E}(-|X||\mathcal{G}) \leq \mathbb{E}(X|\mathcal{G}) \leq \mathbb{E}(|X||\mathcal{G}).$$

□

§12 March 4, 2024

§12.1 Conditonal Expectation Continued

Theorem 12.1 — Suppose that X, Y, X_n are integrable.

(i) If $X = a$ with probability 1, then $E[X \mid \mathcal{G}] = a$.

(ii) For constants a and b, $E[aX + bY \mid \mathcal{G}] = aE[X \mid \mathcal{G}] + bE[Y \mid \mathcal{G}]$.

(iii) If $X \leq Y$ with probability 1, then $E[X \mid \mathcal{G}] \leq E[Y \mid \mathcal{G}]$.

(iv) $|E[X \mid \mathcal{G}]| \leq E[|X| \mid \mathcal{G}]$.

(v) If $\lim_n X_n = X$ with probability 1, $|X_n| \leq Y$, and Y is integrable, then $\lim_n E[X_n \mid \mathcal{G}] = E[X \mid \mathcal{G}]$ with probability 1.

Proof. If $X = a$ with probability 1, the function identically equal to a satisfies conditions (i) and (ii) in the definition of $E[X \mid \mathcal{G}]$, and so (i) above follows by uniqueness.

As for (ii), $aE[X \mid \mathcal{G}] + bE[Y \mid \mathcal{G}]$ is integrable and measurable $\mathcal{G}$, and

$$\int_G (aE[X \mid \mathcal{G}] + bE[Y \mid \mathcal{G}])\, dP = a\int_G E[X \mid \mathcal{G}]dP + b\int_G E[Y \mid \mathcal{G}]dP = a\int_G XdP + b\int_G YdP = \int_G (aX + bY)dP$$

for all G in $\mathcal{G}$, so that this function satisfies the functional equation.

If $X \leq Y$ with probability 1, then

$$\int_G (E[Y \mid \mathcal{G}] - E[X \mid \mathcal{G}])\, dP = \int_G (Y - X)dP \geq 0$$

for all G in $\mathcal{G}$. Since $E[Y \mid \mathcal{G}] - E[X \mid \mathcal{G}]$ is measurable $\mathcal{G}$, it must be nonnegative with probability 1 (consider the set G where it is negative). This proves (iii), which clearly implies (iv) as well as the fact that $E[X \mid \mathcal{G}] = E[Y \mid \mathcal{G}]$ if $X = Y$ with probability 1.

To prove (v), consider $Z_n = \sup_{k \geq n} |X_k - X|$. Now $Z_n \downarrow 0$ with probability 1, and by (ii), (iii), and (iv),

$$|E[X_n \mid \mathcal{G}] - E[X \mid \mathcal{G}]| \leq E[Z_n \mid \mathcal{G}].$$

It suffices, therefore, to show that $E[Z_n \mid \mathcal{G}] \downarrow 0$ with probability 1. By (iii) the sequence $E[Z_n \mid \mathcal{G}]$ is nonincreasing and hence has a limit Z; the problem is to prove that $Z = 0$ with probability 1, or, Z being nonnegative, that $E[Z] = 0$. But $0 \leq Z_n \leq 2Y$, and so (34.1) and the dominated convergence theorem give

$$E[Z] = \int E[Z \mid \mathcal{G}]dP \leq \int E[Z_n \mid \mathcal{G}]dP = E[Z_n] \to 0.$$

□

Theorem 12.2 (Theorem 34.2 (v) Dominated Convergence Theorem for Conditional Expectation) — Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-field. Let $(X_n), X, Y$ be integrable random variables. If $X_n \to X$ a.s. and $|X_n| \leq Y$ a.s.$\forall n$, then

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

Proof. We proved it above. □

Theorem 12.3 — If X is integrable and the σ-fields $\mathcal{G}_1$ and $\mathcal{G}_2$ satisfy $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then

$$E[E[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = E[X \mid \mathcal{G}_1]$$

with probability 1.

Proof. It will be shown first that the right side of (34.4) is a version of the left side if $X = I_{G_0}$ and $G_0 \in \mathcal{G}$. Since $I_{G_0}E[Y \mid \mathcal{G}]$ is certainly measurable $\mathcal{G}$, it suffices to show that it satisfies the functional equation

$$\int_G I_{G_0}E[Y \mid \mathcal{G}]dP = \int_G I_{G_0}Y dP, \quad G \in \mathcal{G}.$$

But this reduces to

$$\int_{G\cap G_0} E[Y \mid \mathcal{G}]dP = \int_{G\cap G_0} Y dP,$$

which holds by the definition of $E[Y \mid \mathcal{G}]$. Thus (34.4) holds if X is the indicator of an element of $\mathcal{G}$.

It follows by Theorem 34.2(ii) that (34.4) holds if X is a simple function measurable $\mathcal{G}$. For the general X that is measurable $\mathcal{G}$, there exist simple functions X_n, measurable $\mathcal{G}$, such that $|X_n| \leq |X|$ and $\lim_n X_n = X$ (Theorem 13.5). Since $|X_nY| \leq |XY|$ and $|XY|$ is integrable, Theorem 34.2(v) implies that

$$\lim_n E[X_nY \mid \mathcal{G}] = E[XY \mid \mathcal{G}]$$

with probability 1. But $E[X_nY \mid \mathcal{G}] = X_nE[Y \mid \mathcal{G}]$ by the case already treated, and of course $\lim_n X_nE[Y \mid \mathcal{G}] = XE[Y \mid \mathcal{G}]$. (Note that $X_nE[Y \mid \mathcal{G}] = E[X_nY \mid \mathcal{G}] \leq E[|XY| \mid \mathcal{G}]$, so that the limit $XE[Y \mid \mathcal{G}]$ is integrable.) Thus (34.4) holds in general. Notice that X has not been assumed integrable.

□

Theorem 12.4 (Tower Property) — If X is measurable $\mathcal{G}$, and if Y and XY are integrable, then

$$E[XY \mid \mathcal{G}] = XE[Y \mid \mathcal{G}]$$

with probability 1.

Proof. Let $X' = E(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1)$. We check that X' satisfies properties (i) and (ii) in the definition of $E(X \mid \mathcal{G}_1)$.

(i) X' is $\mathcal{G}_1$-measurable and integrable. This is clear.

(ii) We have to prove that:

$$\int_G X'\, dP = \int_G X\, dP \quad \forall G \in \mathcal{G}_1$$

Let $G \in \mathcal{G}_1$ be arbitrary. Then

$$\int_G X'\, dP = \int_G E(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1)\, dP = \int_G E(Y \mid \mathcal{G}_1)\, dP = \int_G Y\, dP$$

where $Y = E(X \mid \mathcal{G}_2)$. By property (ii) in the definition of $E(Y \mid \mathcal{G}_1)$, since $G \in \mathcal{G}_1$,

$$\int_G E(X \mid \mathcal{G}_2)\, dP = \int_G X\, dP \quad \text{(using property (ii) in the def. of } E(X \mid \mathcal{G}_2))$$

$$\Rightarrow \int_G X\, dP.$$

Therefore,

$$\int_G X' \, dP = \int_G X \, dP \quad \forall G \in \mathcal{G}_1.$$

$\square$

If $\mathcal{G}_1 \subseteq \mathcal{G}_2$ then trivially $E(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1) = E(X \mid \mathcal{G}_1)$.

$$Y = E(X \mid \mathcal{G}_2)$$

Y is $\mathcal{G}_1$-measurable, hence $\mathcal{G}_2$-measurable.

Lemma 12.5 — If X is $\mathcal{G}$-measurable then $E(X \mid \mathcal{G}) = X$ a.s.

Recall Jensen's inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, then

$$\varphi(E(X)) \leq E(\varphi(X)) \tag{5}$$

for any r.v. X for which $X, \varphi(X)$ are integrable.

Example: $\varphi(X) = |X|^p$, $p \geq 1$

Then (5) says:

$$|E(X)|^p \leq E(|X|^p) \quad \forall p \geq 1$$

In particular, $|E(X)|^2 \leq E(X^2)$.

Recall the following basic properties of convex functions:

1. Definition: φ is convex if

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y) \quad \forall t \in (0, 1)$$

Remark: If $\mathcal{G}_1 \subseteq \mathcal{G}_2$ then trivially $E(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1) = E(X \mid \mathcal{G}_1)$.

$$Y = E(X \mid \mathcal{G}_2)$$

Y is $\mathcal{G}_1$-measurable, hence $\mathcal{G}_2$-measurable.

Lemma (Feb 28): If X is $\mathcal{G}$-measurable then $E(X \mid \mathcal{G}) = X$ a.s.

Recall: Jensen's inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function, then

$$\varphi(E(X)) \leq E(\varphi(X)) \tag{5}$$

for any r.v. X for which $X, \varphi(X)$ are integrable.

Example: $\varphi(X) = |X|^p$, $p \geq 1$

Then (5) says:

$$|E(X)|^p \leq E(|X|^p) \quad \forall p \geq 1$$

In particular, $|E(X)|^2 \leq E(X^2)$.

Recall the following basic properties of convex functions:

1. Definition: φ is convex if

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y) \quad \forall t \in (0, 1)$$

2. If φ is convex, then φ is continuous.

3. If φ is convex,

$$\varphi'(x_0^+) = \lim_{\epsilon \to 0^+} \frac{\varphi(x_0 + \epsilon) - \varphi(x_0)}{\epsilon} \text{ exists and is finite}$$

$$\varphi'(x_0^-) = \lim_{\epsilon \to 0^-} \frac{\varphi(x_0 - \epsilon) - \varphi(x_0)}{\epsilon} \text{ exists and is finite}$$

4. If φ is convex and $\varphi'(x_0^-) \leq A(x_0) \leq \varphi'(x_0^+)$, then

$$\varphi(x) \geq \varphi(x_0) + A(x_0)(x - x_0) \quad \forall x \in \mathbb{R} \tag{6}$$

(6) says that the graph of φ stays above any support line through $(x_0, \varphi(x_0))$. This happens for any $x_0 \in \mathbb{R}$.

Lemma 3 (Jensen's Inequality):

$$\varphi(E(X)) \leq E(\varphi(X)) \tag{2}$$

for any convex function φ and any random variable X such that the expectations exist.

§13 March 6, 2024

§13.1 Proof of Conditional Jensen Inequality

Recall: Jensen Inequality says for any convex function φ,

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$$

Goal: Extend this inequality to $\mathbb{E}(\cdot \mid \mathcal{G})$

Lemma 13.1 (Jensen Inequality for Conditional Expectations) — For any convex function $\varphi : \mathbb{R} \to \mathbb{R}$ and for any random variable X such that X and $\varphi(X)$ are integrable,

$$\varphi(\mathbb{E}(X \mid \mathcal{G})) \leq \mathbb{E}(\varphi(X) \mid \mathcal{G}) \quad \text{a.s.}$$

Proof. Recall from last time that $\forall x_0 \in \mathbb{R}$, $\forall x \in \mathbb{R}$, $\varphi'(x_0^-) \leq A(x_0) \leq \varphi'(x_0^+)$,

$$\varphi(x) \geq \varphi(x_0) + A(x_0)(x - x_0) \tag{2}$$

Fix $\omega \in \Omega$. We apply (2) to $\begin{cases} x_0 = \mathbb{E}(X \mid \mathcal{G})(\omega) \\ x = X(\omega) \end{cases}$. We obtain:

$$\varphi(X(\omega)) \geq \varphi(\mathbb{E}(X \mid \mathcal{G})(\omega)) + A(\mathbb{E}(X \mid \mathcal{G})(\omega))(X(\omega) - \mathbb{E}(X \mid \mathcal{G})(\omega))$$

We drop ω from the writing. We write:

$$\varphi(X) \geq \varphi(\mathbb{E}(X \mid \mathcal{G})) + A(\mathbb{E}(X \mid \mathcal{G}))(X - \mathbb{E}(X \mid \mathcal{G})) \tag{2}$$

□

Case 1

Assume that $\mathbb{E}(X \mid \mathcal{G})$ is bounded, i.e. $|\mathbb{E}(X \mid \mathcal{G})| \leq M$ for some $M \geq 0$.

Note that if φ is convex, then φ and A are bounded on bounded sets. Hence $\varphi(\mathbb{E}(X \mid \mathcal{G}))$ and $A(\mathbb{E}(X \mid \mathcal{G}))$ are bounded (hence integrable).

Take $\mathbb{E}(\cdot \mid \mathcal{G})$ in (2). We use monotonicity of cond. expect. (Th.34.2.(iii)). We get:

$$\mathbb{E}(\varphi(X) \mid \mathcal{G}) \geq \mathbb{E}[\varphi(\mathbb{E}(X \mid \mathcal{G})) \mid \mathcal{G}] + \mathbb{E}[A(\mathbb{E}(X \mid \mathcal{G}))(X - \mathbb{E}(X \mid \mathcal{G})) \mid \mathcal{G}]$$

Case 2: General Case

Let $G_n = \{\omega \in \Omega; |\mathbb{E}(X \mid \mathcal{G})(\omega)| \leq n\}$. Note that $G_n \in \mathcal{G}$ and

$$\mathbb{E}(\mathbb{I}_{G_n} X \mid \mathcal{G}) = \mathbb{I}_{G_n} \mathbb{E}(X \mid \mathcal{G})$$

$$\mathbb{E}(X \mid \mathcal{G}) = \begin{cases} \mathbb{E}(X \mid \mathcal{G}) & \text{on } G_n \\ 0 & \text{on } G_n^c \end{cases}$$

Hence $\mathbb{E}(\mathbb{I}_{G_n} X \mid \mathcal{G})$ is bounded. By applying Case 1 (to $\mathbb{I}_{G_n} X$ instead of X), we obtain:

$$\varphi\left(\mathbb{E}(\mathbb{I}_{G_n} X \mid \mathcal{G})\right) \leq \mathbb{E}\left(\varphi(\mathbb{I}_{G_n} X) \mid \mathcal{G}\right) \quad \text{a.s. } \forall n \geq 1 \tag{3}$$

We evaluate separately the two sides of (3):

LHS (left hand side) is equal to:

$$\text{LHS of (3)} = \varphi\left(\mathbb{E}(\mathbb{I}_{G_n} X \mid \mathcal{G})\right) = \varphi\left(\mathbb{I}_{G_n} \mathbb{E}(X \mid \mathcal{G})\right) \tag{4}$$

because $\mathbb{I}_{G_n}$ is $\mathcal{G}$-measurable.

RHS: Note that

$$(\mathbb{I}_{G_n}X)(\omega) = \mathbb{I}_{G_n}(\omega)X(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in G_n \\ 0 & \text{if } \omega \in G_n^c \end{cases}$$

$$\varphi(\mathbb{I}_{G_n}X)(\omega) = \begin{cases} \varphi(X(\omega)) & \text{if } \omega \in G_n \\ \varphi(0) & \text{if } \omega \in G_n^c \end{cases}$$

This means that $\varphi(\mathbb{I}_{G_n}X) = \varphi(X)\mathbb{I}_{G_n} + \varphi(0)\mathbb{I}_{G_n^c}$. Hence,

$$\text{RHS of (3)} = \mathbb{E}[\varphi(X)\mathbb{I}_{G_n}+\varphi(0)\mathbb{I}_{G_n^c} \mid \mathcal{G}] = \mathbb{E}[\varphi(X)\mathbb{I}_{G_n} \mid \mathcal{G}]+\mathbb{E}[\varphi(0)\mathbb{I}_{G_n^c} \mid \mathcal{G}] = \mathbb{I}_{G_n}\mathbb{E}[\varphi(X) \mid \mathcal{G}]+\mathbb{I}_{G_n^c}\mathbb{E}[\varphi(0) \mid \mathcal{G}] = \mathbb{I}_{G_n}\mathbb{E}[\varphi(X \tag{5}$$

We will use (4) and (5) in inequality (3). We obtain:

$$\varphi(\mathbb{I}_{G_n}\mathbb{E}(X \mid \mathcal{G})) \leq \mathbb{I}_{G_n}\mathbb{E}[\varphi(X) \mid \mathcal{G}] + \varphi(0)\mathbb{I}_{G_n^c} \quad \forall n \geq 1 \quad \text{a.s.}$$

We take the limit as $n \to \infty$. We use the fact that $\{G_n \subseteq G_{n+1} \forall n\}$, $\bigcup_{n=1}^{\infty} G_n = \Omega$.

Hence, $\mathbb{I}_{G_n} \to \mathbb{I}_\Omega = 1$ and $\mathbb{I}_{G_n^c} \to 0$.

Since φ is convex, φ is continuous. Hence $\varphi(\mathbb{I}_{G_n}\mathbb{E}(X \mid \mathcal{G})) \to \varphi(\mathbb{E}(X \mid \mathcal{G}))$ as $n \to \infty$.

Therefore,

$$\varphi(\mathbb{E}(X \mid \mathcal{G})) \leq \mathbb{E}(\varphi(X) \mid \mathcal{G}) \quad \text{a.s.}$$

Recall (Th.33.3) X = r.v., $\mathcal{G} \subseteq \mathcal{F}$ sub σ-field. The *conditional distribution* of X given $\mathcal{G}$ is $\mu(H, \omega)$ for $H \in \mathcal{R}, \omega \in \Omega$ such that:

$$(i) \quad \mu(\cdot, \omega) \text{ is a probability measure on } \mathcal{R} \text{ for } \omega \in \Omega.$$

$$(ii) \quad \mu(H, \cdot) = P(X \in H \mid \mathcal{G}) \quad \text{a.s. } \forall H \in \mathcal{R}$$

§13.2 Conditional Distribution and Conditional Expectation

Theorem 13.2 (Th.34.5: Conditional Distribution and Conditional Expectation) — Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{G} \subseteq \mathcal{F}$ is a sub σ-field, X is an *integrable* r.v. Let $\mu(H, \omega)$ be the cond. distrib. of X given $\mathcal{G}$.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a *measurable* function s.t. $\varphi(X)$ is *integrable*. Then

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}](\omega) = \int_{\mathbb{R}} \varphi(\xi)\mu(d\xi, \omega) \quad \text{for almost all } \omega \in \Omega.$$

In particular, if $\varphi(\xi) = \xi$, then

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \int_{\mathbb{R}} \xi\mu(d\xi, \omega) \quad \text{for almost all } \omega \in \Omega.$$

Proof. **Case 1** $\varphi = \mathbb{1}_H$

For some Borel set $H \in \mathcal{R}$.

$$\text{RHS of (6)} = \int_{\mathbb{R}} \mathbb{1}_H(x)\mu(dx \times \omega) = \mu(H, \omega) = \mathbb{P}(X \in H|\mathcal{G}) = \mathbb{E}\left[\mathbb{1}_{\{X \in H\}}|\mathcal{G}\right]$$

$$\mathbb{1}_{\{X \in H\}}(\omega) = \begin{cases} 1 & \text{if } \omega \in \{X \in H\} \\ 0 & \text{if } \omega \notin \{X \in H\} \end{cases} = \begin{cases} 1 & \text{if } X(\omega) \in H \\ 0 & \text{if } X(\omega) \notin H \end{cases}$$

$$\mathbb{1}_H(X)(\omega) = \mathbb{1}_H(X(\omega)) = \begin{cases} 1 & \text{if } X(\omega) \in H \\ 0 & \text{if } X(\omega) \notin H \end{cases}$$

So $\mathbb{1}_{\{X \in H\}} = \mathbb{1}_H(X)$ and $\mathbb{E}\left[\mathbb{1}_{\{X \in H\}}|\mathcal{G}\right] = \mathbb{E}\left[\mathbb{1}_H(X)|\mathcal{G}\right] = \mathbb{E}\left[\varphi(X)|\mathcal{G}\right]$

Case 2 φ is a simple function i.e., $\varphi = \sum_{i=1}^{k} \alpha_i \mathbb{1}_{H_i}$ with $\alpha_i \in \mathbb{R}, H_i \in \mathbb{R}$.

Follows by Case 1, using linearity. **Case 3** $\varphi \geq 0$. By **Theorem 13.5**, there exists a sequence $\{\varphi_n\}$ of simple functions s.t. $\varphi_n(x) \uparrow \varphi(x)$ as $n \to \infty$, for any $x \in \mathbb{R}$. By Case 2,

$$\mathbb{E}[\varphi_n(X)|\mathcal{G}](\omega) = \int_{\mathbb{R}} \varphi_n(x)\mu(dx \times \omega) \quad \forall n \text{ for a.a. } \omega$$

Let $n \to \infty$ in (7). We have:

$\mathbb{E}[\varphi_n(X)|\mathcal{G}] \xrightarrow{a.s.} \mathbb{E}[\varphi(X)|\mathcal{G}]$ by D.C.T.
To justify the application of this theorem, note that
$\varphi_n(X) \leq \varphi(X) \forall n$ and $\varphi(X)$ is integrable (by hypothesis)
$\int_{\mathbb{R}} \varphi_n(X)\mu(dx \times \omega) \to \int_{\mathbb{R}} \varphi(X)\mu(dx \times \omega)$ by MCT.

We obtain:

$$\mathbb{E}[\varphi(X)|\mathcal{G}] = \int_{\mathbb{R}} \varphi(X)\mu(dx \times \omega) \quad \text{for a.a. } \omega$$

Case 4 φ is arbitrary. We write $\varphi = \varphi^+ - \varphi^-$ where

$$\varphi^+(x) = \begin{cases} \varphi(x) & \text{if } \varphi(x) \geq 0 \\ 0 & \text{if } \varphi(x) < 0 \end{cases}$$

$$\varphi^-(x) = \begin{cases} 0 & \text{if } \varphi(x) \geq 0 \\ -\varphi(x) & \text{if } \varphi(x) < 0 \end{cases}$$

The conclusion follows by applying Case 3 to φ^+, φ^- and use linearity. □

Using **Theorem 3.15**, we can give another proof of **Jensen's Inequality** for **Conditional Expectation**: for any convex function φ,

$$\varphi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}[\varphi(X)|\mathcal{G}] \quad \text{a.s.}$$

To see this, let $\mu(dx, \omega)$ be the cond. distr. of X given $\mathcal{G}$. Then

$$\mathbb{E}(X|\mathcal{G})(\omega) = \int_{\mathbb{R}} x\mu(dx \times \omega) \quad \text{by (6)'}$$

$$\varphi(\mathbb{E}(X|\mathcal{G})(\omega)) = \varphi\left(\int_{\mathbb{R}} x\mu(dx \times \omega)\right) \quad \text{for a.a. } \omega \in \mathbb{R}$$

On the other hand, by (6)

$$\mathbb{E}[\varphi(X)|\mathcal{G}](\omega) = \int_{\mathbb{R}} \varphi(X)\mu(dx \times \omega) \quad \text{for a.a. } \omega \in \mathbb{R}$$

So it suffices to prove that:

$$\varphi\left(\int_{\mathbb{R}} x\mu(dx \times \omega)\right) \leq \int_{\mathbb{R}} \varphi(x)\mu(dx \times \omega) \quad \text{for a.a. } \omega$$

This is in fact the **(Classical) Jensen's Inequality** which says that

$$\varphi(\mathbb{E}(X')) \leq \mathbb{E}[\varphi(X')] \quad \text{for r.v. } X'$$

So here we choose X' to be a r.v. with law $\mu(dx, \omega)$ for fixed ω. Then

$$\begin{cases} \mathbb{E}[X'] = \int_{\mathbb{R}} x\mu(dx, \omega) \\ \mathbb{E}[\varphi(X')] = \int_{\mathbb{R}} \varphi(x)\mu(dx, \omega) \end{cases}$$

§14 March 11, 2024

Recall from last time:

Theorem 14.1 (Theorem 34.5) —

$$\mathbb{E}[\varphi(X)|\mathcal{G}] = \int_{\mathbb{R}} \varphi(z)\mu(dz;\omega)$$

for almost all ω. For all $\varphi : \mathbb{R} \to \mathbb{R}$ measurable s.t. $\varphi(X)$ is integrable.

Here $\mu(H, \omega)$ is the cond. distr. of X given $\mathcal{G}$:

$$\begin{cases} \text{(i)} & \mu(\cdot, \omega) \text{ is a probab. measure } \forall \omega \in \Omega \\ \text{(ii)} & \mu(H, \cdot) = P(X \in H|\mathcal{G}) \text{ a.s.} \end{cases}$$

We will use the following result (see the proof of Th 25.6):

Lemma 14.2 — Let μ be an arb. probab. measure on $(\mathbb{R}, \mathcal{R})$. Then there exists a probab. space $(\Omega, \mathcal{F}, P)$ and a r.v. $X : \Omega \to \mathbb{R}$ s.t. μ is the law of X, i.e.

$$P(X \in B) = \mu(B) \quad \forall B \in \mathcal{R},$$

or equivalently

$$P(X \le x) = F(x) \quad \forall x \in \mathbb{R} \text{ where } F(x) = \mu((-\infty, x]).$$

Proof. Let $(\Omega, \mathcal{F}, P) = ((0,1), \mathcal{B}(0,1), \lambda)$ where λ is the Lebesgue measure.
Define the generalized inverse of F by:

$$F^{-1}(u) = \inf\{x \in \mathbb{R}; F(x) \ge u\} \quad \forall u \in (0,1)$$

It can be proved that: (exercise)

$$u \le F(x) \iff F^{-1}(u) \le x \quad \forall x \in \mathbb{R} \forall u \in (0,1)$$

Take $X(\omega) := F^{-1}(\omega) \quad \forall \omega \in (0,1)$. Then (1) holds:

$$P(X \le x) = P(\{\omega \in (0,1); X(\omega) \le x\}) = P(\{\omega \in (0,1); F^{-1}(\omega) \le x\})$$
$$= \lambda((0, F(x)]) = F(x)$$

□

§14.1 Markov Inequality for Cond. Expectation

Lemma 14.3 (Markov Inequality for Cond. Expectation) — For any integr. r.v. X and any sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$, we have:

$$P(|X| \ge \alpha|\mathcal{G}) \le \frac{1}{\alpha^p}\mathbb{E}(|X|^p|\mathcal{G}) \quad a.s.$$

Proof. Let $\varphi(x) = 1_{\{|X| \ge \alpha\}}, x \in \mathbb{R}$. Clearly $\varphi : \mathbb{R} \to \mathbb{R}$ is measurable. Let $\mu(H, \omega)$ be the conditional distr. of X given $\mathcal{G}$.
For every $\omega \in \Omega$ fixed, let Z_ω be a r.v. defined on probab. space $(\Omega', \mathcal{F}', P') = ((0,1), \mathcal{B}(0,1), \lambda)$ such that the law of Z_ω (under P') is $\mu(\cdot; \omega)$, i.e.

$$P' \circ Z_\omega^{-1} = \mu(\cdot, \omega) \quad \text{(see Lemma 1)}$$

Then

$$P(|X| \geq \alpha|\mathcal{G})(\omega) = \mathbb{E}\left[1_{\{|X|\geq\alpha\}}|\mathcal{G}\right](\omega) = \mathbb{E}[\varphi(X);\mathcal{G}](\omega)$$

Applying Theorem 34.5,

$$\int_{\mathbb{R}} \varphi(x)\mu(dx;\omega) = \int_{\Omega'} \varphi(Z_\omega)dP' = P'(|Z_\omega| \geq \alpha)$$

By the classical Markov inequality,

$$P'(|Z_\omega| \geq \alpha) \leq \frac{1}{\alpha^p}\mathbb{E}'(|Z_\omega|^p) = \frac{1}{\alpha^p}\int_{\mathbb{R}} |x|^p \mu(dx;\omega)$$

Thus,

$$P(|X| \geq \alpha|\mathcal{G})(\omega) \leq \frac{1}{\alpha^p}\mathbb{E}(|X|^p|\mathcal{G})(\omega)$$

□

§14.2 Inequalites for Cond. Expectation

Corollary 14.4 (Chebyshev's Inequality for Cond. Expectation)
For any integrable r.v. X and for any sub-σ-field $\mathcal{G}$,

$$P(|X - \mathbb{E}(X|\mathcal{G})| \geq \alpha|\mathcal{G}) \leq \frac{1}{\alpha^2}\text{Var}(X|\mathcal{G}) \quad \forall\alpha > 0, \text{ if } X^2 \text{ is integrable}$$

where

$$\text{Var}(X|\mathcal{G}) = \mathbb{E}((X - \mathbb{E}(X|\mathcal{G}))^2|\mathcal{G}) = \mathbb{E}(X^2|\mathcal{G}) - (\mathbb{E}(X|\mathcal{G}))^2$$

Proof. Let $Y = X - \mathbb{E}(X|\mathcal{G})$. Then Y is integrable since it is a linear combination of integrable r.v.'s. We apply Lemma 2 to Y with $p = 2$. We obtain:

$$P(|Y| \geq \alpha|\mathcal{G}) \leq \frac{1}{\alpha^2}\mathbb{E}(Y^2|\mathcal{G}) = \frac{1}{\alpha^2}\text{Var}(X|\mathcal{G})$$

$$P(|X - \mathbb{E}(X|\mathcal{G})| \geq \alpha|\mathcal{G})$$

□

Note that Theorem 34.5 has a multivariate extension:

$$\mathbb{E}[\varphi(X,Y)|\mathcal{G}](\omega) = \int_{\mathbb{R}^2} \varphi(x,y)\mu(dx,dy;\omega) \quad \text{for a.a. } \omega$$

where $\mu(H,\omega)$ is the cond. distribution of (X,Y) given $\mathcal{G}$, i.e.

$$\begin{cases} \text{(i)} & \mu(\cdot,\omega) \text{ is a prob. measure on } \mathbb{R}^2 \forall\omega \in \Omega \\ \text{(ii)} & \mu(H,\cdot) = P((X,Y) \in H|\mathcal{G}) \text{ a.s. } \forall H \in \mathcal{R}^2 \end{cases}$$

Lemma 14.5 (Hölder Inequality for Cond. Expectations) — Let X, Y be two r.v.'s s.t. XY is integrable ($\mathbb{E}(|X|^p|\mathcal{G})$ is integrable and $\mathbb{E}(|Y|^q|\mathcal{G})$ is integrable).
For some $p, q > 1$ s.t.

$$\frac{1}{p} + \frac{1}{q} = 1$$

let $\mathcal{G}$ be an arb. sub-σ-field of $\mathcal{F}$. Then

$$\mathbb{E}[|XY||\mathcal{G}] \leq (\mathbb{E}(|X|^p|\mathcal{G}))^{\frac{1}{p}} (\mathbb{E}(|Y|^q|\mathcal{G}))^{\frac{1}{q}}$$

Proof. Let $\mu(H, \omega)$ be the cond. distr. of (X, Y) given $\mathcal{G}$. Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$\varphi(x, y) = |xy|$$

Clearly φ is measurable. For any $\omega \in \Omega$ fixed, let $Z_\omega = (Z^1_\omega, Z^2_\omega)$ be a random vector defined on a probab. space $(\Omega', \mathcal{F}', P')$ s.t. the law of Z_ω under P' is $\mu(\cdot, \omega)$, i.e.

$$P' \circ Z_\omega^{-1} = \mu(\cdot, \omega)$$

Then

$$\mathbb{E}[|XY||\mathcal{G}](\omega) = \mathbb{E}[\varphi(X, Y)|\mathcal{G}](\omega) = \mathbb{E}[\varphi(Z^1_\omega, Z^2_\omega)]$$

By change of variable,

$$\int_{\mathbb{R}^2} \varphi(z^1_\omega, z^2_\omega) dP' = \int_{\mathbb{R}} |z^1_\omega z^2_\omega| d\mu(z_\omega; \omega)$$

By Hölder's inequality,

$$\mathbb{E}[|XY||\mathcal{G}] \leq (\mathbb{E}(|X|^p|\mathcal{G}))^{\frac{1}{p}} (\mathbb{E}(|Y|^q|\mathcal{G}))^{\frac{1}{q}}$$

□

Finally, we define the Markov process:

Definition 14.6 Let $(\Omega, \mathcal{F}, P)$ be a prob. space and $X_t : \Omega \to \mathbb{R}$ a r.v.
For all $t \geq 0$, the collection $(X_t)_{t \geq 0}$ is a Markov process if

$$P(X_u \in H | X_s, s \leq t) = P(X_u \in H | X_t) \quad \forall t < u$$

Here the cond. probab. is w.r.t. $\sigma\{X_s; s \leq t\}$ on the RHS and $\sigma\{X_t\}$ on the LHS.

§15 March 13, 2024

§15.1 Markov Decision Process

Recall the following definition from last time:
A process $(X_t)_{t\geq 0}$ (i.e. a collection of r.v.'s defined on $(\Omega, \mathcal{F}, P)$) is called a Markov process if

$$P(X_u \in H | X_s, s \in [0,t]) = P(X_u \in H | X_t) \quad \forall 0 \leq t < u \tag{3}$$

Denote $\mathcal{G}_1 = \sigma(\{X_s; s \in [0,t]\})$ "the history" (or the past) of the process up to time t
$\mathcal{G}_2 = \sigma(\{X_t\})$ "the present"
$\mathcal{G}_3 = \sigma(\{X_u\})$ where $u > t$ "the future"
Relation (1) says that for every $A \in \mathcal{G}_3$

$$P(A|\sigma(\mathcal{G}_1 \cup \mathcal{G}_2)) = P(A|\mathcal{G}_2) \tag{4}$$

which is denoted by $\mathcal{G}_1 \vee \mathcal{G}_2$ (notation).

Lemma 15.1 (Problem 3.11) — Let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ be sub-σ-fields of $\mathcal{F}$. The following conditions are equivalent:

(i) $P(A|\mathcal{G}_1 \vee \mathcal{G}_2) = P(A|\mathcal{G}_2)$ for all $A \in \mathcal{G}_3$.

(ii) $P(A \cap B|\mathcal{G}_2) = P(A|\mathcal{G}_2) \cdot P(B|\mathcal{G}_2)$ for all $A \in \mathcal{G}_1$, $B \in \mathcal{G}_3$, i.e., A and B are "conditionally independent" given $\mathcal{G}_2$.

(iii) $P(A|\mathcal{G}_2 \vee \mathcal{G}_3) = P(A|\mathcal{G}_2)$ for all $A \in \mathcal{G}_1$.

Proof. It is enough to prove (i) $\implies$ (ii). The argument for (ii) $\implies$ (i) is the same. We have

$$\begin{aligned}
P(A \cap A_3|\mathcal{G}_2) &= E\left[\mathbf{1}_{A\cap A_3}|\mathcal{G}_2\right] \\
&= E\left[E\left[\mathbf{1}_A \mathbf{1}_{A_3}|\mathcal{G}_1 \vee \mathcal{G}_2\right]|\mathcal{G}_2\right] \quad \text{(Tower Property)} \\
&= E\left[\mathbf{1}_A E\left[\mathbf{1}_{A_3}|\mathcal{G}_1 \vee \mathcal{G}_2\right]|\mathcal{G}_2\right] \\
&= E\left[\mathbf{1}_A P(A_3|\mathcal{G}_1 \vee \mathcal{G}_2)|\mathcal{G}_2\right] \\
&\qquad (\mathbf{1}_{A_3} \text{ is } \mathcal{G}_1\text{-measurable, hence } \mathcal{G}_1 \vee \mathcal{G}_2\text{-measurable}) \\
&= E\left[\mathbf{1}_A P(A_3|\mathcal{G}_2)|\mathcal{G}_2\right] \quad \text{(from (i))} \\
&= E\left[\mathbf{1}_A|\mathcal{G}_2\right] P(A_3|\mathcal{G}_2) \\
&= P(A|\mathcal{G}_2)P(A_3|\mathcal{G}_2).
\end{aligned}$$

This shows that (i) implies (ii). (ii) $\implies$ (i) We show that $P(A|\mathcal{G}_2)$ satisfies the two conditions from the def of $P(A_3|\mathcal{G}_1 \vee \mathcal{G}_2)$:

1) $P(A|\mathcal{G}_2)$ is $\mathcal{G}_2$-measurable, hence $\mathcal{G}_1 \vee \mathcal{G}_2$-measurable

2) We have to show that

$$\int_G P(A|\mathcal{G}_2)\, dP = P(A \cap G) \quad \forall G \in \mathcal{G}_1 \vee \mathcal{G}_2$$

By Theorem 33.1, it is enough to prove that (i) holds $\forall G \in \mathcal{F}$ where $\{F = A \cap A' : A \in \mathcal{G}_1, A' \in \mathcal{G}_2\}$ is a π-system (exer) and $\sigma(F) = \mathcal{G}_1 \vee \mathcal{G}_2$ (exer). Ω is a countable union of sets in F $(\Omega \in \mathcal{G}_1, \mathcal{G}_2)$.

Let $G = A \cap A'$ with $A \in \mathcal{G}_1, A' \in \mathcal{G}_2$. Then on the left-hand side of (1) we have: **LHS of (1):**

$$\begin{aligned}
\text{LHS of (1)} &= \int_{A_1 \cap A_2} P(A_3|\mathcal{G}_2)\, dP \\
&= E\left[\mathbf{1}_{A_1 \cap A_2} P(A_3|\mathcal{G}_2)\right] \\
&= E\left[E\left[\mathbf{1}_{A_1} \frac{P(A_3|\mathcal{G}_2)}{\mathbf{1}_{A_2}} \middle| \mathcal{G}_2\right]\right] \\
&\quad \text{by } \mathcal{G}_2\text{-measurability (product of } \mathcal{G}_2\text{-meas. rv's)} \\
&= E\left[\mathbf{1}_{A_2} P(A_3|\mathcal{G}_2) \cdot E\left[\mathbf{1}_{A_1}|\mathcal{G}_2\right]\right] \\
&= E\left[\mathbf{1}_{A_2} P(A_3|\mathcal{G}_2)\right] \cdot P(A_1|\mathcal{G}_2) \\
&\quad \text{using (ii)} \\
&= E\left[\mathbf{1}_{A_1 \cap A_2 \cap A_3}|\mathcal{G}_2\right] \\
&= P(A_1 \cap A_2 \cap A_3).
\end{aligned}$$

RHS of (1):

$$\begin{aligned}
\text{RHS of (1)} &= P(A_1 \cap (A_2 \cap A_3)) \\
&= E\left[\mathbf{1}_{A_1 \cap A_2 \cap A_3}\right] \\
&= E\left[E\left[\mathbf{1}_{A_1 \cap A_3} \mathbf{1}_{A_2} \middle| \mathcal{G}_2\right]\right] \\
&= E\left[\mathbf{1}_{A_2}\right] \cdot E\left[\mathbf{1}_{A_1 \cap A_3}|\mathcal{G}_2\right] \\
&= P(A_2) \cdot P(A_1 \cap A_3|\mathcal{G}_2).
\end{aligned}$$

□

§15.2 Discrete Time Martingales

Definition 15.2 Let $X_1, X_2, \ldots$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$, and let $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be a sequence of σ-fields in $\mathcal{F}$. The sequence $\{(X_n, \mathcal{F}_n) : n = 1, 2, \ldots\}$ is a martingale if the following four conditions hold:

1. $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$,
2. X_n is measurable with respect to $\mathcal{F}_n$,
3. $E[|X_n|] < \infty$ for all n,
4. with probability 1, $E[X_{n+1}|\mathcal{F}_n] = X_n$.

We simply say that $\{X_n\}_{n \geq 1}$ is a *martingale* if (X_n) is a martingale with respect to the natural filtration

$$\mathcal{F}_n^X = \sigma(X_1, X_2, \ldots, X_n)$$

which is the “smallest” σ-filtration which satisfies (i) and (ii).

Remark: If (ii) holds, then (iv) is equivalent to:

$$\int_A X_n\, dP - \int_A X_{n+1}\, dP = 0 \quad \forall A \in \mathcal{F}_n$$

(by the def. of $E[X_n|\mathcal{F}_n]$).

Motivation: Bets placed at horse races

- X_n = fortune of the gambler after the n-th race
- $\mathcal{F}_n$ = information accumulated by the gambler up to the n-th race.
- $E[X_{n+1}|\mathcal{F}_n]$ = expected fortune after the $(n+1)$-th race.

The game is fair if $E[X_{n+1}|\mathcal{F}_n] = X_n$.

§16 March 18, 2024

§16.1 Section 35 Martingales Continued

Definition 16.1 Let $(X_n)_{n\geq 1}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. The sequence is a martingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 1}$ if:

(i) $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \geq 1$.

(ii) X_n is $\mathcal{F}_n$-measurable for all $n \geq 1$.

(iii) $E[|X_n|] < \infty$ for all $n \geq 1$.

(iv) $E[X_{n+1}|\mathcal{F}_n] = X_n$ almost surely for all $n \geq 1$.

Basic Example: Let $(S_n)_{n\geq 1}$ be independent random variables with $E[\Delta_n] = 0$ where $X_n = \frac{1}{2}\Delta_n$ and $\mathcal{F}_n = \sigma(\Delta_1, \ldots, \Delta_n)$. Then $(X_n)_{n\geq 1}$ is a martingale with respect to $(\mathcal{F}_n)_{n\geq 1}$.

Example 16.2 (Martingale Representation with Respect to Filtration)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, let ν be a finite measure on $\mathcal{F}$, and let $\mathcal{F}_1, \mathcal{F}_2, \ldots$ be a nondecreasing sequence of σ-fields in $\mathcal{F}$. Suppose that P dominates ν when both are restricted to $\mathcal{F}_n$—that is, suppose that $A \in \mathcal{F}_n$ and $P(A) = 0$ together imply that $\nu(A) = 0$. There is then a density or Radon-Nikodym derivative X_n of ν with respect to P when both are restricted to $\mathcal{F}_n$. X_n is a function that is measurable $\mathcal{F}_n$ and integrable with respect to P, and it satisfies

$$\int_A X_n \, dP = \nu(A), \quad A \in \mathcal{F}_n. \tag{5}$$

If $A \in \mathcal{F}_n$ then $A \in \mathcal{F}_{n+1}$ as well, so that

$$\int_A X_{n+1} \, dP = \nu(A); \tag{6}$$

this and (35.9) give (35.3). Thus the X_n are a martingale with respect to the $\mathcal{F}_n$.

Definition 16.3 We say that a sequence $(X_n)_{n\geq 1}$ is a submartingale with respect to the filtration $(\mathcal{F}_n)_{n\geq 1}$ if it satisfies conditions (i)–(iii) in Definition 1, and the following property:

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n \text{ a.s.} \quad \text{for all } n \geq 1.$$

Condition (iv) is equivalent to:

$$\int_A X_n \, dP \leq \int_A X_{n+1} \, dP \quad \forall A \in \mathcal{F}_n.$$

Example 16.4 (Basic Example)**.** Let $(\Delta_n)_{n\geq 1}$ be i.i.d. random variables with $\mathbb{E}[\Delta_n] \geq 0$ for all $n \geq 1$. Let $X_n = \sum_{i=1}^n \frac{\Delta_i}{2}$ and $\mathcal{F}_n = \sigma(\Delta_1, \ldots, \Delta_n)$, then $(X_n)_{n\geq 1}$ is a submartingale with respect to $(\mathcal{F}_n)_{n\geq 1}$.

To see this, we note that for all $n \geq 1$,

$$\begin{aligned}
\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_n + \frac{\Delta_{n+1}}{2}|\mathcal{F}_n] \\
&= X_n + \mathbb{E}[\frac{\Delta_{n+1}}{2}|\mathcal{F}_n] \\
&= X_n + \frac{\mathbb{E}[\Delta_{n+1}]}{2} \geq X_n \text{ a.s.},
\end{aligned}$$

since Δ_{n+1} is independent of $\mathcal{F}_n$ and hence $\mathbb{E}[\Delta_{n+1}|\mathcal{F}_n] = \mathbb{E}[\Delta_{n+1}]$.

If $(X_n)_{n\geq 1}$ is a submartingale with respect to $(\mathcal{F}_n)_{n\geq 1}$, then $(X_n)_{n\geq 1}$ is also a submartingale with respect to $(\mathcal{G}_n)_{n\geq 1}$ where $\mathcal{G}_n = \sigma(X_1, \dots, X_n)$ is the minimal σ-field generated by $(X_1, \dots, X_n)$.

Properties of Submartingales (exercise):

1. $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$ almost surely for all $n \geq 1$.
2. $\mathbb{E}[X_1] \leq \mathbb{E}[X_2] \leq \mathbb{E}[X_3] \leq \dots$
3. If $X_n - X_{n-1} = \Delta_n$ for all $n \geq 1$, then Δ_n is integrable and $\mathbb{E}[\Delta_n|\mathcal{F}_{n-1}] \geq 0$ almost surely for all $n \geq 1$.

Theorem 16.5 — (i) If $(X_n)_{n\geq 1}$ is a martingale with respect to $(\mathcal{F}_n)_{n\geq 1}$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a convex function such that $\phi(X_n)$ is integrable for all $n \geq 1$, then $(\phi(X_n))_{n\geq 1}$ is a submartingale with respect to $(\mathcal{F}_n)$.

(ii) If $(X_n)_{n\geq 1}$ is a submartingale with respect to $(\mathcal{F}_n)$ and $\phi : \mathbb{R} \to \mathbb{R}$ is a convex non-decreasing function such that $\phi(X_n)$ is integrable for all $n \geq 1$, then $(\phi(X_n))_{n\geq 1}$ is a submartingale with respect to $(\mathcal{F}_n)$.

Proof. Properties (i)-(ii) from the definition of submartingale are clearly satisfied. To prove (iv') we have the following:

(i) $\mathbb{E}[\phi(X_{n+1})|\mathcal{F}_n] \geq \phi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) = \phi(X_n)$ by Jensen's Inequality for Conditional Expectation.

(ii) $\mathbb{E}[\phi(X_{n+1})|\mathcal{F}_n] \geq \phi(\mathbb{E}[X_{n+1}|\mathcal{F}_n]) \geq \phi(X_n)$ as ϕ is convex and ϕ is non-decreasing.

□

Observation: If $(X_n)_{n\geq 1}$ is a martingale then $(X_n^2)_{n\geq 1}$ and $(|X_n|)_{n\geq 1}$ are sub-martingales.

Definition 16.6 Let $(\mathcal{F}_n)_{n\geq 1}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\tau : \Omega \to \{1, 2, \dots\}$ be a random variable such that $\{\tau \leq n\} \in \mathcal{F}_n$ for all $n \geq 1$. We say that τ is a stopping time with respect to $(\mathcal{F}_n)_{n\geq 1}$ and define

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \text{ for all } n \geq 1\}.$$

If $(X_n)_{n\geq 1}$ is a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, we define a new random variable $X_\tau : \Omega \to \mathbb{R}$ by

$$X_\tau(\omega) := X_{\tau(\omega)}(\omega) \quad \text{for all } \omega \in \Omega.$$

Lemma 16.7 — Let $\mathcal{F} = (\mathcal{F}_n)_{n\geq 1}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the following statements:

(a) τ is a stopping time with respect to $(\mathcal{F}_n)$ if $\{\tau = n\} \in \mathcal{F}_n$ for all $n \geq 1$.

(b) $\mathcal{F}_\tau$ is a σ-field if τ is a stopping time with respect to $(\mathcal{F}_n)$.

(c) τ is $\mathcal{F}_\tau$-measurable and X_τ is $\mathcal{F}_\tau$-measurable if X_n is $\mathcal{F}_n$-measurable.

(d) If $\tau(\omega) = k$ for some fixed $k \in \mathbb{N}$, then $\mathcal{F}_\tau = \mathcal{F}_k$.

(e) If $\tau_1 \leq \tau_2$ are stopping times with respect to $(\mathcal{F}_n)$, then $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$.

Proof. **a)** We have that $\{\tau = n\} = \bigcap_{m\geq n}\{\tau \leq m\} \subseteq \mathcal{F}_m \subseteq \mathcal{F}_n$ for all $m \geq n$, hence $\{\tau = n\} \in \mathcal{F}_n$.
Conversely, $\{\tau \leq n\} = \bigcup_{k=1}^{n}\{\tau = k\} \in \mathcal{F}_k \subseteq \mathcal{F}_n$ for all $k \leq n$, therefore $\{\tau \leq n\} \in \mathcal{F}_n$.
b) $\mathcal{F}_\tau$ satisfies the following axioms:

1. $\emptyset \in \mathcal{F}_\tau$: $\emptyset \cap \{\tau \leq n\} = \emptyset \in \mathcal{F}_n$ for all $n \geq 1$.
2. If $A \in \mathcal{F}_\tau$ then $A^c \in \mathcal{F}_\tau$: $A^c \cap \{\tau \leq n\} = \{\tau \leq n\} \setminus A \in \mathcal{F}_n$ because $\{\tau \leq n\}$ and A are in $\mathcal{F}_n$.
3. If $\{A_k\} \subseteq \mathcal{F}_\tau$ then $\bigcup_k A_k \in \mathcal{F}_\tau$: $(\bigcup_k A_k) \cap \{\tau \leq n\} = \bigcup_k (A_k \cap \{\tau \leq n\}) \in \mathcal{F}_n$ by the closure of $\mathcal{F}_n$ under countable unions.

We continue with parts c) and e) next time. $\square$

§17 March 20, 2024

Recall: Let $(\mathcal{F}_n)_{n\geq 1}$ be a filtration on a probability space $(\Omega, \mathcal{F}, P)$. A random variable $\tau : \Omega \to \{1, 2, \ldots\}$ is called a *stopping time* with respect to $(\mathcal{F}_n)_{n\geq 1}$ if

$$\{\tau = n\} \in \mathcal{F}_n \text{ for all } n \geq 1.$$

In this case, we define $\mathcal{F}_\tau \equiv \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n \text{ for all } n \geq 1\}$.

We proved the following properties:

1. τ is a stopping time if $\{\tau = n\} \in \mathcal{F}_n$ for all $n \geq 1$.
2. $\mathcal{F}_\tau$ is a σ-field.
3. τ is $\mathcal{F}_\tau$-measurable.
4. If $\tau = k$ (constant) then $\mathcal{F}_\tau = \mathcal{F}_k$.

Exercise: Show that $\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \text{ for all } n \geq 1\}$.

Property: If $\tau_1 \leq \tau_2$ are stopping times with respect to $(\mathcal{F}_n)_{n\geq 1}$, then $\mathcal{F}_{\tau_1} \subseteq \mathcal{F}_{\tau_2}$.

Proof. Let $A \in \mathcal{F}_{\tau_1}$. We want to prove that $A \in \mathcal{F}_{\tau_2}$, i.e., $A \cap \{\tau_2 = n\} \in \mathcal{F}_n$ for all n.

$$A \cap \{\tau_2 = n\} = (A \cap \{\tau_1 = n\}) \cap \{\tau_2 = n\} \in \mathcal{F}_n \text{ since } \{\tau_2 = n\} \in \mathcal{F}_n.$$

□

Property: If $(X_n)_{n\geq 1}$ are r.v.'s such that X_n is $\mathcal{F}_n$-measurable for all $n \geq 1$, then $\mathbf{1}_{\{X_\tau \in B\}}$ is $\mathcal{F}_\tau$-measurable.

Proof. Let $B \in \mathbb{R}$ be an arbitrary Borel set. We have to prove that $\mathbf{1}^{-1}_{\{X_\tau \in B\}}(1) = \{X_\tau \in B\} \in \mathcal{F}_\tau$. Using property 5, this is equivalent to showing that $\{X_\tau \in B\} \cap \{\tau = n\} \in \mathcal{F}_n$ for all $n \geq 1$.

Note that:

$$\begin{aligned}\{X_\tau \in B\} \cap \{\tau = n\} &= \{\omega \in \Omega : X_{\tau(\omega)}(\omega) \in B, \tau(\omega) = n\} \\ &= \{\omega \in \Omega : X_n(\omega) \in B\} \cap \{\tau = n\} \in \mathcal{F}_n, \quad \text{for any } n \geq 1.\end{aligned}$$

□

Theorem 17.1 (Optional Sampling Theorem) — Let $(X_i)_{i=1,\ldots,n}$ be a submartingale with respect to the filtration $(\mathcal{F}_i)_{i=1,\ldots,n}$. Let τ_1 and τ_2 be stopping times with respect to $(\mathcal{F}_i)_{i=1,\ldots,n}$ with $\tau_1, \tau_2 \colon \Omega \to \{1, 2, \ldots, n\}$. Then

$$\mathbb{E}[X_{\tau_2}|\mathcal{F}_{\tau_1}] \geq X_{\tau_1} \quad \text{a.s.} \tag{7}$$

that is, (X_{τ_1}, X_{τ_2}) is a submartingale with respect to $(\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2})$.

Proof. Let $X_{\tau_i} = \sum_{k=1}^n X_k \mathbf{1}_{\{\tau_i = k\}}$ then $|X_{\tau_i}| \leq \sum_{k=1}^n |X_k| \mathbf{1}_{\{\tau_i = k\}} \leq \sum_{k=1}^n |X_k|$. So $\mathbb{E}[|X_{\tau_i}|] \leq \sum_{k=1}^n \mathbb{E}[|X_k|] < \infty$, i.e., X_{τ_i} is integrable. (for $i = 1, 2$)

To show (2), we must prove that:

$$\left| \int_A X_{\tau_2} dP \right| \geq \int_A X_{\tau_1} dP \quad \forall A \in \mathcal{F}_{\tau_1} \quad (3) \tag{8}$$

Let $\Delta_k = X_k - X_{k-1}$ for $k = 2, \ldots, n$, and $\Delta_1 = X_1$. Then $X_{\tau_2} - X_{\tau_1} = \sum_{k=\tau_1+1}^{\tau_2} \Delta_k = \sum_{k=\tau_1+1}^{n} \Delta_k \mathbf{1}_{\{\tau_1 < k \leq \tau_2\}}$.

(Use: $\sum_{k=\tau_1+1}^m (X_k - X_{k-1}) = (X_{m-1} - X_{\tau_1}) + (X_{m-2} - X_{m-1}) + \ldots + (X_m - X_{m-1}) = X_m - X_{\tau_1}$ for any $m, \tau_1 \in \{1, \ldots, n\}, \tau_1 \leq m$)

In our case, $L = \tau_1(\omega), M = \tau_2(\omega)$. Hence, for $A \in \mathcal{F}_{\tau_1}$,

$$\int_A (X_{\tau_2} - X_{\tau_1})\, dP = \int_A \sum_{k=\tau_1+1}^{\tau_2} \Delta_k\, dP = \int_A \sum_{k=\tau_1+1}^{n} \Delta_k \mathbf{1}_{\{\tau_1 < k \leq \tau_2\}}\, dP.$$

Note that

$$\mathbf{1}_{B_{\tau_2}} := A \cap \{\tau_1 < k \leq \tau_2\} = A \cap \{\tau_1 < k\} \cap \{k \leq \tau_2\} \in \mathcal{F}_{\tau_2},$$

where $B_{\tau_2} \in \mathcal{F}_{\tau_2}$ by the definition of $\mathcal{F}_{\tau_2}$. Recall that $(\Delta_k)_{k=1,\ldots,n}$ is a submartingale difference:

$$\mathbb{E}[X_k|\mathcal{F}_{k+1}] \geq X_k \quad \text{so} \quad \mathbb{E}[X_{\tau_2} - X_{\tau_1}|\mathcal{F}_{\tau_2}] \geq 0 \quad \text{i.e.} \quad \mathbb{E}[\Delta_k|\mathcal{F}_{k+1}] \geq 0 \quad \text{a.s.}$$

This means that for any set $B \in \mathcal{F}_{\tau_1}$,

$$\int_B \Delta_k\, dP \geq 0.$$

In particular, this is true for $B = B_{\tau_2}$ above. Hence

$$\int_A \Delta_k\, dP \geq 0, \quad \text{for all} \quad A \in \mathcal{F}_{\tau_1}, \{\tau_1 < k \leq \tau_2\}.$$

Hence

$$\int_A (X_{\tau_2} - X_{\tau_1})\, dP \geq 0.$$

□

If $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_m$ are stopping times with respect to $(\mathcal{F}_k)_{k=1,\ldots,n}$, and $(X_k)_{k=1,\ldots,n}$ is a submartingale with respect to $(\mathcal{F}_k)_{k=1,\ldots,n}$, then $(X_{\tau_1}, X_{\tau_2}, \ldots, X_{\tau_m})$ is a submartingale with respect to $(\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}, \ldots, \mathcal{F}_{\tau_m})$.

Theorem 17.2 (Kolmogorov's Maximal Inequality) — Let $(X_k)_{k\geq 1}$ be i.i.d. random variables with $\mathbb{E}(X_k^2) < \infty$ for all k. Let

$$S_n = \sum_{k=1}^{n} X_k,$$

and we know that (S_n) is a martingale. Then Kolmogorov's inequality states that

$$\mathbb{P}\left(\max_{k\leq n} |S_k| > \alpha\right) \leq \frac{1}{\alpha^2}\mathbb{E}(S_n^2) \quad \text{for all } \alpha > 0.$$

Note that $\max_{k\leq n} |S_k| > \alpha$ is equivalent to $\max_{k\leq n} S_k^2 > \alpha^2$. Hence, we can write the inequality as:

$$\mathbb{P}\left(\max_{k\leq n} S_k^2 > \alpha^2\right) \leq \frac{\mathbb{E}(S_n^2)}{\alpha^2}.$$

Recall that (S_n^2) is a submartingale. The next result extends this inequality to an arbitrary submartingale.

Theorem 17.3 (Maximal Inequality) — Let $(X_k)_{k=1,\ldots,n}$ be a submartingale with respect to $(\mathcal{F}_k)_{k=1,\ldots,n}$. Then for any $\alpha > 0$,

$$\mathbb{P}\left(\max_{k\leq n} |X_k| \geq \alpha\right) \leq \frac{1}{\alpha}\mathbb{E}(|X_n|).$$

Proof. Define: $\tau\colon \Omega \to \{1, 2, \ldots, n\}$ as

$$\tau(\omega) = \begin{cases} \min\{j \leqslant n : X_j(\omega) \geq \alpha\} & \text{if there exists } j \leqslant n \text{ s.t. } X_j(\omega) \geq \alpha, \\ n & \text{otherwise } (i.e., X_i(\omega) < \alpha \ \forall i \leqslant n). \end{cases}$$

Clearly, τ is a stopping time w.r.t. $(\mathcal{F}_k)_{k=1,\dots,n}$.

Proof of Claim: We have to prove that $\{\tau = k\} \in \mathcal{F}_k$ for all $k = 1, \dots, n$ (see property 1 on page 1).

Let $\{r_j\}_{j=1,\dots,n}$ be arbitrary. We have two cases:

Case 1: $\{r_j \leqslant m\}$

For $\{\tau = k\} = \bigcap_{j=1}^{k} \{X_j < \alpha\} \cap \{X_k \geq \alpha\} \in \mathcal{F}_k$

Case 2: $\{r_j = n\}$

For $\{\tau = n\} = \bigcap_{j=1}^{n} \{X_j < \alpha\} \in \mathcal{F}_n$.

Define $\tau \geqslant n$ (also a stopping time). Clearly, $\tau_1 \leq \tau_2$. By Optional Sampling Theorem (Theorem 35.2)

$$\mathbb{E}[X_{\tau_2} | \mathcal{F}_{\tau_1}] \geq X_{\tau_1} \text{ a.s.}$$

Let $M_\tau = \max\{X_i, i \leq \tau\}$, for $\tau = 1, \dots, n$. Clearly, $M_{\tau_1} \leq M_{\tau_2} \leq \cdots \leq M_{\tau_n}$.

Let us examine the event $\{M_n \geq \alpha\}$.

Claim: $\{M_n \geq \alpha\} \in \mathcal{F}_{\tau_1}$, i.e., $\{M_n \geq \alpha\} \cap \{\tau_1 \leq \tau_2\} \in \mathcal{F}_{\tau_2}$ for all $\tau_2 = 1, \dots, n$.

Proof of Claim: We will show that: $\forall \tau_2 = 1, \dots, n$.

$$\{M_n \geq \alpha\} \cap \{\tau_1 \leq \tau_2\} = \{M_{\tau_2} \geq \alpha\}$$

To prove (7), we use double-inclusion:

($\subseteq$) Let $\omega \in \{M_{\tau_2} \geq \alpha\}$. Then $M_{\tau_2}(\omega) \geq \alpha$. But since $M_{\tau_2}(\omega) = \max\{X_i(\omega), i \leq \tau_2\}$ and $\tau_1(\omega)$ is the smallest index i for which $X_i(\omega) \geq \alpha$, we have $\{\tau_1(\omega) \leq \tau_2\}$.

($\supseteq$) If $\tau_2 = n$, the inclusion is clear. If $\tau_2 = n - 1$, by the definition of τ_1, $X_{\tau_1} \geq \alpha$. But $M_{\tau_2} \geq X_{\tau_1}$, so $M_{\tau_2} \geq \alpha$. On the event $\{\tau_1 \leq \tau_2\}$, we have $M_{\tau_1} \leq M_{\tau_2}$. Hence, $\{M_{\tau_2} - X_{\tau_1} \geq 0\}$.

Remark: If $\tau_1, \tau_2, \dots, \tau_n$ are stopping times w.r.t. $(\mathcal{F}_\tau)_{\tau=1,\dots,n}$, then $(X_{\tau_1}, X_{\tau_2}, \dots, X_{\tau_n})$ is a submartingale w.r.t. $(\mathcal{F}_{\tau_1}, \mathcal{F}_{\tau_2}, \dots, \mathcal{F}_{\tau_n})$.

Coming back to (8), we recall that (8) means that

$$\int_A X_{\tau_2} \, dP \geq \int_A X_{\tau_1} \, dP \quad \forall A \in \mathcal{F}_{\tau_1},$$

we will this inequality with $A = \{M_n \geq \alpha\} \in \mathcal{F}_{\tau_1}$, hence

$$\int \mathbf{1}_{\{M_n \geq \alpha\}} X_{\tau_2} \, dP \geq \int \mathbf{1}_{\{M_n \geq \alpha\}} X_{\tau_1} \, dP.$$

To summarize, we obtain that:

$$\int_{\{M_n \geq \alpha\}} X_{\tau_2} \, dP \leq \int_{\{M_n \geq \alpha\}} X_n \, dP \tag{9}$$

On the other hand, $\{M_n \geq \alpha\} = \bigcup_{k=1}^{n} \{X_k \geq \alpha\}$. So if $\omega \in \{M_n \geq \alpha\}$, then $\tau_2 = n$ such that $X_{\tau_2}(\omega) \geq \alpha$ and $\tau_1(\omega) \leq \tau_2$.

Hence

$$\int_{\{M_n \geq \alpha\}} X_{\tau_2} \, dP = \alpha P(M_n \geq \alpha) \tag{10}$$

Putting (9) and (10) together, we get:

$$\alpha P(M_n \geq \alpha) \leq \int_{\{M_n \geq \alpha\}} X_n^+ \, dP - \int_{\{M_n \geq \alpha\}} X_n^- \, dP \leq \int_\Omega (X_n^+ + X_n^-) \, dP = \mathbb{E}(|X_n|)$$

$\square$

§18 March 27, 2024

§18.1 Martingales Continued

Let $[a,b]$ be an interval, and $X_1, X_2, \ldots, X_n$ are random variables. Inductively, we define variables $\sigma_1, \sigma_2, \ldots, \sigma_n$ as follows:

$$\sigma_1 = \begin{cases} \min\{j \leq n : X_j \leq \alpha\} & \text{if there exists } j \leq n \text{ s.t. } X_j \leq \alpha \\ n & \text{otherwise} \end{cases}$$

For any $k \leq n$:

- if k is even,

$$\sigma_k = \begin{cases} \min\{j \leq n; j > \sigma_{k-1} \text{ and } X_j \geq \beta\} & \text{if there exists } j \leq n \text{ s.t. } j > \sigma_{k-1} \text{ and } X_j \geq \beta \\ n & \text{otherwise} \end{cases}$$

- if k is odd,

$$\sigma_k = \begin{cases} \min\{j \leq n; j > \sigma_{k-1} \text{ and } X_j \leq \alpha\} & \text{if there exists } j \leq n \text{ s.t. } j > \sigma_{k-1} \text{ and } X_j \leq \alpha \\ n & \text{otherwise} \end{cases}$$

We define the number U of upcrossings of $[a,b]$ by $X_1, \ldots, X_n$ as the largest index i s.t.

$$X_{\sigma_{2i-1}} \leq \alpha < \beta \leq X_{\sigma_{2i}}$$

Example: $n = 17$. Fix $\omega \in \Omega$.
In this picture,

$$U(\omega) = 2,$$

$$\sigma_1(\omega) = 4, \quad \sigma_2(\omega) = 6, \quad \sigma_3(\omega) = 10, \quad \sigma_4(\omega) = 12, \quad \sigma_5(\omega) = 16, \quad \sigma_6 = \ldots = \sigma_{17} = 17$$

Theorem 18.1 (Doob's Upcrossing Theorem) — Let $(X_k)_{k=1,\ldots,n}$ be a submartingale w.r.t. $(\mathcal{F}_k)_{k=1,\ldots,n}$ and U be the number of upcrossings of $[a,b]$ by $X_1, \ldots, X_n$. Then

$$E(U) \leq \frac{E(|X_n|) + |a|}{\beta - \alpha}$$

Proof. Let

$$Y_k = \max\{X_k - \alpha, 0\}$$

Note that $\psi(x) = \max\{x - \alpha, 0\}$ is a convex and non-decreasing function $\psi : \mathbb{R} \to \mathbb{R}$.
By Theorem 35.1 (iii), $(Y_k)_{k=1,\ldots,n}$ is a submartingale w.r.t. $(\mathcal{F}_k)_{k=1,\ldots,n}$.
Note that $\sigma_1, \ldots, \sigma_n$ are stopping times w.r.t. $(\mathcal{F}_k)_{k=1,\ldots,n}$ (exercise).
Moreover,

- for $k = 1$,

$$\sigma_k = \begin{cases} \min\{j \leq n; X_j = 0\} & \text{if there exists } j \leq n \text{ s.t. } X_j = 0 \\ n & \text{otherwise} \end{cases}$$

- for k even,

$$\sigma_k = \begin{cases} \min\{j \leq n; j > \sigma_{k-1} \text{ and } X_j \geq \beta\} & \text{if there exists } j \leq n \text{ s.t. } j > \sigma_{k-1} \text{ and } X_j \geq \beta \\ n & \text{otherwise} \end{cases}$$

- for k odd,

$$\sigma_k = \begin{cases} \min\{j \le n; j > \sigma_{k-1} \text{ and } X_j = 0\} & \text{if there exists } j \le n \text{ s.t. } j > \sigma_{k-1} \text{ and } X_j = 0 \\ n & \text{otherwise} \end{cases}$$

Then U is the number of upcrossings of $[0, \theta]$ by $Y_1, \ldots, Y_n$.
Note that $1 \le \sigma_1 \le \sigma_2 \le \ldots \le \sigma_n = n$. By the Optional Stopping Theorem (Th. 35.2),

$$(Y_{\sigma_k})_{k=1,\ldots,n} \text{ is a submartingale w.r.t. } (\mathcal{F}_{\sigma_k})_{k=1,\ldots,n}.$$

Hence,

$$E(Y_{\sigma_k} \mid \mathcal{F}_{\sigma_{k-1}}) \ge Y_{\sigma_{k-1}} \quad \forall k = 2, \ldots, n.$$

In particular,

$$E(Y_{\sigma_k}) \ge E(Y_{\sigma_{k-1}}) \quad \forall k = 2, \ldots, n.$$

It follows that

$$Y_n \ge Y_{\sigma_n} \ge Y_{\sigma_n} - Y_{\sigma_1} = \sum_{k=2}^{n} (Y_{\sigma_k} - Y_{\sigma_{k-1}})$$

$$\sum_{k=2}^{n} (Y_{\sigma_k} - Y_{\sigma_{k-1}}) = \sum_{\substack{k=2 \\ k \text{ even}}}^{n} (Y_{\sigma_k} - Y_{\sigma_{k-1}}) + \sum_{\substack{k=2 \\ k \text{ odd}}}^{n} (Y_{\sigma_k} - Y_{\sigma_{k-1}})$$

Hence,

$$E(Y_n) \ge E\left(\sum_{\substack{k=2 \\ k \text{ even}}}^{n} (Y_{\sigma_k} - Y_{\sigma_{k-1}})\right) + E\left(\sum_{\substack{k=2 \\ k \text{ odd}}}^{n} (Y_{\sigma_k} - Y_{\sigma_{k-1}})\right) \ge 0$$

If $Y_{\sigma_{2i}} \ge \theta$, then

$$Y_{\sigma_{2i}} - Y_{\sigma_{2i-1}} \ge \theta$$

Since there are U such differences, we get

$$\sum_{e} \ge \theta U$$

and so

$$E(\sum_{e}) \ge \theta E(U) \tag{3}$$

From (2) and (3), we get

$$E(U) \le \frac{E(|X_n|) + |a|}{\theta}$$

Finally,

$$E(Y_n) = \int_{\Omega} \max\{X_n - \alpha, 0\} dP \le \int_{\Omega} |X_n - \alpha| dP \le E(|X_n|) + |\alpha| \tag{5}$$

So

$$E(U) \le \frac{E(|X_n|) + |\alpha|}{\beta - \alpha}$$

□

§18.2 Martingale Convergence Theorem

If $(X_n)_{n \ge 1}$ is a submartingale w.r.t. $(\mathcal{F}_n)$ and

$$K := \sup_{n \ge 1} E(|X_n|) < \infty,$$

then there exists an integrable random variable X such that $X_n \to X$ a.s. Moreover, $E(|X|) \le 1$.

Proof

Fix $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$. Let $U_n^{\alpha,\beta}$ be the number of upcrossings of $[\alpha, \beta]$ by $X_1, \ldots, X_n$. By Theorem 35.4,

$$E(U_n^{\alpha,\beta}) \leq \frac{E(|X_n|) + \alpha}{\beta - \alpha} \leq \frac{K + \alpha}{\beta - \alpha} \quad \forall n \geq 1.$$

Note that $(U_n^{\alpha,\beta})$ is a non-decreasing sequence. Hence

$$\lim_{n\to\infty} U_n^{\alpha,\beta} \text{ exists (but may be } \infty).$$

By Monotone Convergence Theorem,

$$E(U_n^{\alpha,\beta}) \uparrow E(\lim_{n\to\infty} U_n^{\alpha,\beta}).$$

By (7),

$$E(\lim_{n\to\infty} U_n^{\alpha,\beta}) \leq \frac{K + \alpha}{\beta - \alpha} < \infty.$$

Hence

$$\lim_{n\to\infty} U_n^{\alpha,\beta} < \infty \text{ a.s.} \quad (8).$$

For $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$, let

$$X^* = \limsup_{n\to\infty} X_n \quad \text{and} \quad X_* = \liminf_{n\to\infty} X_n.$$

Then,

$$X^* = \inf_{n\geq 1} \sup_{k\geq n} X_k \quad \text{and} \quad X_* = \sup_{n\geq 1} \inf_{k\geq n} X_k.$$

Claim

$$\{\omega \in \Omega : X_*(\omega) < \alpha < \beta < X^*(\omega)\} \subset \{\omega \in \Omega : \lim_{n\to\infty} U_n^{\alpha,\beta}(\omega) = \infty\}$$

with probability 0.

Proof of Claim

$$X_*(\omega) = \sup \inf_{k\geq n} X_k(\omega) < \alpha$$

implies

$$\forall n, \inf_{k\geq n} X_k(\omega) < \alpha.$$

Similarly,

$$X^*(\omega) > \beta$$

implies

$$\forall n, \sup_{k\geq n} X_k(\omega) > \beta.$$

By (8),

$$P(X_* < \alpha < \beta < X^*) = 0 \quad \forall \alpha, \beta \in \mathbb{R}, \alpha < \beta.$$

From here,

$$0 \leq P(X_* < X^*) = P\left(\bigcup_{\alpha,\beta\in\mathbb{Q},\alpha<\beta} \{X_* < \alpha < \beta < X^*\}\right) \leq \sum_{\alpha,\beta\in\mathbb{Q},\alpha<\beta} P(X_* < \alpha < \beta < X^*) = 0.$$

So,

$$P(X_* < X^*) = 0 \quad \text{and} \quad P(X_* = X^*) = 1.$$

Hence, $\lim_{n\to\infty} X_n = X$ exists.