Lecture 1

Sep 4th, 2024

Introduction

Office Hours: W 1–2, F 2–3

Ex: Game of Chance → flip a coin
- heads ⇒ get \$1
- tails ⇒ get \$0

X = payout ⇝ $\mathbb{E}X = \$0.50$.

Ex: Keep flipping until get heads. Payout $X = 2^n$, where n is the # of flips it took to get H.

H \$1
TH \$2
TTTH \$3
⋮

$$\mathbb{E}X = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 4 + \cdots = \frac{1}{2} + \frac{1}{2} + \cdots = \infty$$

kinda unrealistic and not really matching reality...

Law of Large Numbers: if $X_1, X_2, X_3, \ldots$ are payouts from independent plays of this game, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mathbb{E}X_i = \infty \implies \frac{1}{n}\sum_{i=1}^{n} X_i \sim \frac{1}{2}\underbrace{\log_2 n}_{\text{growing but grows very slowly...}}$$

If you pay \$10 to play, you should be able to play it $n = 2^{20} \approx 1\,000\,000$ times to break even.

Measure Theory

Why measure theory? Take an ∞ seq. of coin flips (e.g., HTTHTHHHT...). Let ω = HTTHTHHHT... ← specified to ∞

Then, $P(\omega) = 0$.

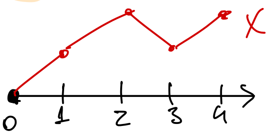
Know: $P(A \sqcup B) = P(A) + P(B)$

So,

ILLEGAL BY MEASURE THEORY

$$P(\text{first flip} = H) = \sum_{\substack{\omega \text{ first} \\ \text{flip is } H}} P(\omega) = \sum 0 = 0 \neq \frac{1}{2}$$

Ex: (SIMPLE RANDOM WALK) Prob $\frac{1}{2}$ go up and $\frac{1}{2}$ go down



$X(t)$ = this function of the random walk

$$X_n(t) = \frac{1}{\sqrt{n}} X(nt), \quad t \in [0,1]$$

Donsker's Invariance Principle: $\lim_{n \to \infty} X_n(t) =$ Brownian motion

Def: (PROBABILITY SPACE) A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where $(\Omega, \mathcal{F})$ is a measurable space and $\mathbb{P}$ is a measure on $(\Omega, \mathcal{F})$ s.t. $\mathbb{P}(\Omega) = 1$.

$\Omega \to$ set (possible outcomes)

$\mathcal{F} \to$ σ-algebra of subsets of Ω

(events)

- $\emptyset \in \mathcal{F}$
- if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
- if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$

Algebra is the same except it's closed only under finite unions

$\mu : \mathcal{F} \to \mathbb{R} \cup \{\infty\}$ is a measure

- $\mu(\emptyset) = 0$
- $\mu(A) \geq 0 \quad \forall A \in \mathcal{F}$
- if $\{A_i\}_{i \in \mathbb{N}} \subset \mathcal{F}$ are all disjoint, $\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mu(A_i)$

$\downarrow$

Probability measure: if $\mu(\Omega) = 1$

Ex: 3 coin flips $\longrightarrow$ $\Omega = \{T, H\}$

$$\mathcal{F} = \{TTT, HHH, THH, \dots\}$$

Ex: (DISCRETE PROB. SPACE) $\Omega = \mathbb{N}$, $\mathcal{F} =$ all subsets of Ω.

Construct a prob. measure by letting $(p_i)_{i\in\mathbb{N}}$ be a sequence s.t. $p_i \geq 0$ and $\sum_{i\in\mathbb{N}} p_i = 1$. Then,

$$\boxed{\mathbb{P}(A) := \sum_{i\in A} p_i \text{ is a probability measure.}}$$

Thm: Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$. Then

(i) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ (monotonicity)

(ii) $A \subset \bigcup_{i\in\mathbb{N}} A_i \Rightarrow \mathbb{P}(A) \leq \sum_{i\in\mathbb{N}} \mathbb{P}(A_i)$ (union bound)

(iii) If $A_i \nearrow A$ (i.e., $A_1 \subset A_2 \subset A_3 \subset \cdots$ and $A := \bigcup_{i\in\mathbb{N}} A_i$) then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ (continuity from below)

(iv) If $B_i \searrow B$ (i.e., $B_1 \supset B_2 \supset \cdots$ and $B := \bigcap_{i\in\mathbb{N}} B_i$), then $\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(B)$. (continuity from above)

Pf: (i) $\mathbb{P}(B) = \mathbb{P}(B\cap A \sqcup B \cap A^c) = \mathbb{P}(B\cap A) + \mathbb{P}(B\cap A^c)$

$$\geq \mathbb{P}(B\cap A) = \mathbb{P}(A).$$

(ii) Let $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i$ then $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$. So,

$$\mathbb{P}(A) \le \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n\right) \overset{B_n\text{'s are all disjoint}}{=} \sum_{n\in\mathbb{N}} \mathbb{P}(B_n) \le \sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$$

(iii) Let $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i$. So,

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(iv) Similar. ■

LECTURE 2

Sep 6^{th}, 2024

Last time: $(\Omega, \mathcal{F}, \mathbb{P})$ probability triple

- Ω: set (possible outcomes)
- $\mathcal{F}$: σ-algebra (events)
- $\mathbb{P}$: probability measure $\mathbb{P}(\Omega) = 1$

Def: If $\mathcal{A}$ is a collection of subsets of Ω, denote by $\sigma(\mathcal{A})$ the σ-algebra generated by $\mathcal{A}$.

(i.e., $\sigma(\mathcal{A})$ is the smallest σ-algebra containing $\mathcal{A}$)

Def: If Ω is a topological space, then denote by $\mathcal{B}(\Omega)$ the σ-algebra generated by open sets (called BOREL σ-algebra)

e.g.: $\Omega = \mathbb{R} \longrightarrow \mathcal{B}(\mathbb{R})$.

Ex: If Q is a probability measure on $(\Omega, \mathcal{F})$ and if $f: \Omega \to \mathbb{R}$ is a measurable function s.t.

f is measurable iff $\forall A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) := \{\omega : f(\omega) \in A\} \in \mathcal{F}$.

(i) $f(\omega) \geq 0 \quad \forall \omega \in \Omega$

(ii) $\int_\Omega f(\omega)\, dQ(\omega) = 1$

then $P(A) := \int_A \underbrace{f(\omega)}\, dQ(\omega)$ defines a probability measure.

"f = DENSITY OF P W.R.T. Q"

* **LEBESGUE-STIELTJES MEASURE:** Given a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, define

$$\boxed{F(t) := \mu\big((-\infty, t]\big)}.$$

Then,

(i) $F(t) \geq 0$

(ii) F is increasing

(iii) F is right-continuous: if $t_n \downarrow t$ then $F(t_n) = \mu((-\infty, t_n])$

so,

$$\lim_{n\to\infty} \mu((-\infty, t_n]) = \mu\left(\bigcap_{n=1}^{\infty} (-\infty, t_n]\right) = \mu((-\infty, t_n]) = F(t).$$

(iv) $\lim_{t \to -\infty} F(t) = 0$ and $\lim_{t \to \infty} F(t) = 1$.

Thm: If $F: \mathbb{R} \to \mathbb{R}$ obeys (i)–(iv) above, then $\exists! \, \mu_F$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ s.t. $\mu_F((-\infty, t]) = F(t)$.

- Terminology: F is the cummulative distribution function (cdf) of μ_F.

Recall: (Carathéodory Extension Thm) If $\mathcal{A}$ is an algebra of sets in Ω and if $\mu: \mathcal{A} \to \mathbb{R}$ is a nonnegative countably additive function, then $\exists$ an extension μ to $\sigma(\mathcal{A})$ s.t. μ is a measure. The extension is unique if $(\Omega, \sigma(\mathcal{A}), \mu)$ is σ-finite.

(i) Countably additive: if $A_i \in \mathcal{A}$ are disjoint sets in $\mathcal{A}$ and if $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$ then $\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mu(A_i)$

(ii) σ-finite: $\Omega = \bigcup_{i \in \mathbb{N}} A_i$ s.t. $A_i \in \sigma(\mathcal{A})$ and $\mu(A_i) < \infty \; \forall i$.

Pf (of Thm): • Let $\mathcal{A}$ be the algebra of sets that are finite disjoint unions of intervals $(a, b]$ ($a = -\infty$ & $b = \infty$ are allowed).

- If $A = \bigsqcup_{i=1}^{n} (a_i, b_i]$, then define $\mu(A) = \sum_{i=1}^{n} F(b_i) - F(a_i)$.

$$F(t) = \mu((-\infty, t])$$
$$F(b) - F(a) = \mu((-\infty, b]) - \mu((-\infty, a]) = \mu((a, b]).$$

- Also, clear that μ is finitely additive on $\mathcal{A}$ and if $A \subset B$ then $\mu(A) \leq \mu(B)$.

- Just need to check that $\mu(A) = \sum_{i=1}^{\infty} \mu(A_i)$ if $A = \bigsqcup_{i=1}^{\infty} A_i$

First, by monotonicity,

$$\mu(A) \geq \mu\left(\bigcup_{i=1}^{n} A_i\right) \overset{\text{finite additivity}}{=} \sum_{i=1}^{n} \mu(A_i) \Longrightarrow \mu(A) \geq \sum_{i=1}^{\infty} \mu(A_i).$$

Check upper-bound:

WLOG, $A = (a, b]$, $A_i = (a_i, b_i]$. Let $\varepsilon > 0$ and choose δ st. $F(a+\delta) \leq F(a) + \varepsilon$ (can do this b/c we assume F is right-continuous)

Choose $\eta_i > 0$ st. $F(b_i + \eta_i) \leq F(b_i) + \varepsilon 2^{i}$.

Thus,

$$\mu((a, b]) = F(b) - F(a) \leq F(b) - F(a+\delta) + \varepsilon$$
$$= \mu((a+\delta, b]) + \varepsilon.$$

Now,

$$\underbrace{[a+\delta, b]}_{\text{compact}} \subset \bigcup_{i=1}^{\infty} (a_i, b_i + \eta_i)$$

$\Rightarrow \exists J < \infty$ st. $[a+\delta, b] \subset \bigcup_{i=1}^{J} (a_i, b_i + \eta_i)$

$$\Rightarrow (a+\delta, b] \subset \bigcup_{i=1}^{J} (a_i, b_i + \eta_i] .$$

Thus,

$$\mu((a+\delta, b]) + \varepsilon \leq \mu\left(\bigcup_{i=1}^{J} (a_i, b_i+\eta_i]\right) + \varepsilon$$

$$\leq \sum_{i=1}^{J} \mu((a_i, b_i+\eta_i]) + \varepsilon$$

$$= \sum_{i=1}^{J} F(b_i-\eta_i) - F(a_i) + \varepsilon$$

$$\leq \sum_{i=1}^{J} F(b_i) - F(a_i) + 2^{i}\varepsilon$$

$$\leq \sum_{i=1}^{\infty} F(b_i) - F(a_i) + 2\varepsilon$$

$$= \sum_{i=1}^{\infty} \mu((a_i, b_i]) + 2\varepsilon .$$

$\Rightarrow \mu$ countably additive $\Rightarrow$ Use Carathéodory to get the measure. ■

* **DYNKIN'S $\pi-\lambda$ THEOREM**

! VERY IMPORTANT IN PROBABILITY !

Def: A collection P of sets is called a π-system if it's closed under finite intersections

Def: A collection of sets L is called a λ-system if

(i) $\Omega \in L$

(ii) If $A \in L$, then $A^c \in L$

(iii) If $A_i \in L$ are disjoint, then $\bigcup_{i=1}^{\infty} A_i \in L$

Obs: λ-system is almost a σ-algebra.

Thm: (Dinkin's π-λ) If P is a π-system and L is a λ-system such that $P \subset L$, then $\sigma(P) \subset L$.

Ex: Take $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and two prob. measures μ, ν. Then $\mu \equiv \nu$ iff $\mu(\theta) = \nu(\theta)$ $\forall$ open sets $\theta \in \mathcal{B}(\mathbb{R})$.

π-system = open sets = P

λ-system = $\{A \in \mathcal{B}(\mathbb{R}) : \nu(A) = \mu(A)\} = L$

$\Rightarrow \mathcal{B}(\mathbb{R}) = \sigma(P) \subset L$.

Dinkin Thm

Ex: $\Omega = \Omega_i^{\mathbb{Z}}$, Ω_i is some set, $\mathcal{F}_i$ σ-algebra on Ω_i.

$\mathcal{E}$:= cylinder sets Ω i.e: $A \in \mathcal{E}$ if $A = A_1 \times A_2 \times \cdots \times A_n \times \Omega_i \times \cdots \times \Omega_i \times \cdots$ where $A_i \in \mathcal{F}_i$

$\mathcal{E}$ are a π-system $\Rightarrow$ if two measures agree on cylinder sets then they agree on $\sigma(\mathcal{E})$.

Pf: (Dynkin's Thm) • Let $\lambda(P)$ be the λ-system generated by P
• We just need ~~need~~ to check that $\lambda(P)$ is also a π-system
$\Rightarrow \sigma(P) \subset \lambda(P) \subset L$.

• For any $A \subset \Omega$, define $M_A := \{B \subset \Omega \text{ s.t. } A \cap B \in \lambda(P)\}$

NTS: if $A \in \lambda(P)$, then $\lambda(P) \subset M_A$.

Check the definition of λ-system. ■

* RANDOM VARIABLE: If $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ are both measurable spaces, a measurable map $f: \Omega \to S$ is a fct. s.t.

$$f^{-1}(A) \in \mathcal{F} \quad \forall A \in \mathcal{S}.$$

Def: If $(\Omega, \mathcal{F}, P)$ is a probability space then a random variable (taking values in S) is just a measurable map $f: \Omega \to S$.

→ Usually, $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and a random variable w/out further quantifiers means real-valued.

Ex: 3 COIN FLIPS $\Omega = \{HHH, THH, TTH, \ldots\}$
and a random variable $X(\omega)$ could be the # of heads.
ω = outcome
$X(\omega)$ = "realization" of $X(\omega)$

Ex: RANDOM WALK $\Omega = \{0,1\}^{\mathbb{N}}$

$\omega \in \Omega \longrightarrow$ simple random walk paths

$(S, \mathcal{S}) = (\mathbb{Z}^{\mathbb{N}}, \sigma(\text{cylinder sets}))$.

NOTATION: Usually use X, Y, Z to denote random variables and typically one forgets about Ω. $X = X(\omega)$

- If X takes values on $(S, \mathcal{S})$, then we write

$$\{X \in B\} = \{\omega : X(\omega) \in B\}, \quad B \in \mathcal{S}.$$

Def: If X takes values on $(S, \mathcal{S})$, then the measure μ_X on $(S, \mathcal{S})$ defined by

$$\mu_X(A) := \mathbb{P}(X \in A)$$

(Pushforward of measure via X)

is called a distribution or law of X.

- In the real variable case, we get a probability measure μ_X on $\mathbb{R}$,

$$F(t) = \mu_X((-\infty, t]) = \mathbb{P}(X \leq t)$$

(Here, we say that F is the cdf of X)

Ex: Whenever we have a Lebesgue–Stieltjes measure μ_F on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the function $X : \mathbb{R} \to \mathbb{R}$ given by $X(x) = x$ is a random variable w/ distribution μ_F.

Ex: (NORMAL DISTRIBUTION) Take

$$d\mu_F(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$$

If Z has this measure $d\mu_F(x)$ as its distribution, then we say Z is a standard normal random variable.

- Let $(\Omega, \mathcal{F}) = ([0,1], \mathcal{B}([0,1]))$. Then $U: \Omega \to \mathbb{R}$ given by $U(x) = x$ is a uniform $[0,1]$-random variable

$$P(U \le t) = t \quad \text{for } t \in [0,1].$$

Let F be a cdf. For $\omega \in \Omega$, define

$$X(\omega) := \inf \{ s \in \mathbb{R} : F(s) \ge \omega \}$$

Claim: The distribution of X is μ_F.

Ex: for $t \in (0,1)$ and $X_\theta = \mathbb{1}_{\{U \le t\}}$ then $X_\theta \in \{0,1\}$

$P(X_\theta = 1) = P(U \le t) = \theta$.

LECTURE 3

INEQUALITIES & INDEPENDENCE

Sep 11th, 2024

Def: (EXPECTATION) Given a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $X: \Omega \to \mathbb{R}$, we define the expectation of X as:

$$\mathbb{E}X := \int_\Omega X(\omega)\, d\mathbb{P}(\omega).$$

Assuming of course X is integrable or $X \geq 0$.

Rmk: $\mathbb{E}$ inherits properties from the integral.

Ex: If A is an event (i.e., $A \in \mathcal{F}$), then

$$\mathbb{P}(A) = \mathbb{E}\mathbb{1}_A, \quad \mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}.$$

Def: (VARIANCE) If $\mathbb{E}X < \infty$, then we can define the <u>variance</u> of X by

$$\mathrm{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

[$\mathrm{Var}(X)$ finite $\iff \mathbb{E}X^2 < \infty$]

<u>TERMINOLOGY</u>: $\mathbb{E}X^k$ is usually called the k-th moment of X

Recall: μ_X is the measure on $\mathbb{R}$ s.t. $\mu_X((-\infty, t]) = \mathbb{P}(X \leq t)$

Thm: If X is a random variable with law μ_X and if $g: \mathbb{R} \to \mathbb{R}$ is measurable and $g(X)$ is integrable, then

$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(s)\, d\mu_X(s)$$

⚠ Compute the exp. of any fct. of r.v. X

Pf: True for simple fcts $\Rightarrow$ true for bdd fcts $\Rightarrow$ true for general fcts. $\square$

USEFUL INEQUALITIES:

- **Hölder's Inequality**: X, Y r.v's and $\frac{1}{p} + \frac{1}{q} = 1$ with $p, q \in [1, \infty]$, then

$$\mathbb{E}|XY| \leq \left[\mathbb{E}|X|^p\right]^{\frac{1}{p}} \left[\mathbb{E}|Y|^q\right]^{\frac{1}{q}}.$$

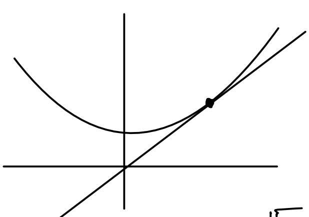
Special case: if $Y \equiv 1$, then $\mathbb{E}|X| \leq \left(\mathbb{E}|X|^p\right)^{\frac{1}{p}}$ (i.e., harder for a r.v. to have lots of moments that are not finite)

- **Jensen's Inequality**: if $\varphi: \mathbb{R} \to \mathbb{R}$ is convex and if X and $\varphi(X)$ are both integrable, then

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

Special case: $\varphi(x) = e^x$, then $\mathbb{E}X \leq \log(\mathbb{E}e^X)$.

Pf: (Jensen's Inequality) Let $f(x) = ax + b$ s.t.



$$f(\mathbb{E}X) = \varphi(\mathbb{E}X) \text{ and } f(x) \leq \varphi(x) \ \forall x \in \mathbb{R}$$

(This f exists by analysis......) Thus, this ineq. gives

$$\mathbb{E}\varphi(X) \geq \mathbb{E}f(X) = \mathbb{E}(aX+b) = a\,\mathbb{E}(X) + b = f(\mathbb{E}X) = \varphi(\mathbb{E}X).$$

$\blacksquare$

• **Markov's Inequality**: If X is a r.v. and $t \geq 0$, then

$$\mathbb{P}(|X| \geq t) \leq \frac{1}{t} \mathbb{E}|X| .$$

(i.e., probability of X being very large is very small)

Pf: $\mathbb{P}(|X| \geq t) = \mathbb{E}\left[\mathbb{1}_{\{|X| \geq t\}}\right]$.

If $|X| \geq t$ then $\frac{|X|}{t} \geq 1$, so

$$\mathbb{E}\left(\mathbb{1}_{\{|X| \geq t\}}\right) \leq \mathbb{E}\left(\frac{|X|}{t}\mathbb{1}_{\{|X| \geq t\}}\right)$$

$$\leq \mathbb{E}\left(\frac{|X|}{t}\right).$$

■

Ex: (Chebyshev's Inequality)

$$\mathbb{P}(|X| \geq t) = \mathbb{P}(X^2 \geq t^2) \overset{\text{Markov}}{\leq} \frac{\mathbb{E}X^2}{t^2}.$$

(can change $2 \to p$ to get the decay of the higher moments of X)

Ex: $\mathbb{P}(X \geq t) = \mathbb{P}(e^X \geq e^t) \leq e^{-t}\,\mathbb{E}(e^X)$.

⚠ Much harder to check if X is not zero... only tool:

Thm: (Paley–Zygmund Inequality) If $Z \geq 0$ is a r.v. and $\theta \in (0,1)$, then

$$\mathbb{P}(Z \geq \theta\, \mathbb{E}Z) \geq (1-\theta)^2 \frac{(\mathbb{E}Z)^2}{\mathbb{E}(Z^2)}.$$

Pf: (Basically Cauchy–Schwartz)

$$\mathbb{E}Z = \mathbb{E}Z\mathbb{1}_{\{Z<\theta\mathbb{E}Z\}} + \underline{\mathbb{E}Z\mathbb{1}_{\{Z\geq\theta\mathbb{E}Z\}}}$$

$$\leq \theta\,\mathbb{E}Z + \underline{(\mathbb{E}Z^2)^{1/2}\left(\mathbb{E}\mathbb{1}^2_{\{Z\geq\theta\mathbb{E}Z\}}\right)^{1/2}}$$ Hölder

$$= \theta\,\mathbb{E}Z + (\mathbb{E}Z^2)^{1/2}\left(\mathbb{P}(Z\geq\theta\mathbb{E}Z)\right)^{1/2}.$$ $\mathbb{1}_A^2 = \mathbb{1}_A$

∎

---

* **INDEPENDENT VARIABLES**

Def: Two events A, B are <u>independent</u> if $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Def: Two r.v.'s X, Y are independent if $\forall\, C, D$ Borel, we have

$$\mathbb{P}\left[(X\in C)\cap(Y\in D)\right] = \mathbb{P}(X\in C)\,\mathbb{P}(Y\in D).$$

A and B are independent $\iff$ $\mathbb{1}_A, \mathbb{1}_B$ are indep. (in the sense of r.v.'s.

Def: Two σ-algebras $\mathcal{F}, \mathcal{G}$ are independent iff $\forall\, C\in\mathcal{F}$ and $\forall\, D\in\mathcal{G}$ we have $\mathbb{P}(C\cap D) = \mathbb{P}(C)\cdot\mathbb{P}(D)$

X, Y are independent r.v.'s $\iff$ $\underline{\sigma(X)}$ and $\underline{\sigma(Y)}$ are independent as σ-algebras.

$\sigma(X) := \{A \in \Omega : A = X^{-1}(B), B \text{ Borel}\}$

(smallest σ-algebra that makes X measurable)

GENERALIZE THIS TO MORE R.V.'s:

Def: n σ-algebras $\mathcal{F}_1, \dots, \mathcal{F}_n$ are independent if $\forall A_i \in \mathcal{F}_i$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i).$$

- n r.v.'s $X_1, \dots, X_n$ are independent if $\sigma(X_i)$ are indep.

$\iff$ $\forall$ Borel B_i we have

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in B_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i).$$

- n events $A_1, \dots, A_n$ are independent $\iff$ $\mathbb{1}_{A_i}$ are independent

$$\iff \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i), \quad \forall I \subset \{1, \dots, n\}.$$

LECTURE 4

Sep 13th, 2024

* **PRODUCT MEASURE**: If we have two σ-finite measure spaces $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$, then $\exists!$ μ on $\Omega_1 \times \Omega_2$

w/ $\sigma(A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2)$ s.t.

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

Thm: (Fubini) If f on $\Omega_1 \times \Omega_2$ is measurable then if either $f \geq 0$ or

$$\int_{\Omega_1 \times \Omega_2} |f(x,y)| \, d\mu_1 \otimes \mu_2 (x,y) < \infty$$

then

$$\int_{\Omega_1} \left(\int_{\Omega_2} f(x,y) \, d\mu_2(y) \right) d\mu_1(x) = \int_{\Omega_1 \times \Omega_2} f(x,y) \, d\mu_1 \otimes \mu_2 (x,y)$$

- If we have $X_1, \ldots, X_n$ then $(X_1, \ldots, X_n) \in \mathbb{R}^n$ is a random variable in $\mathbb{R}^n$.
- $\exists!$ μ on $\mathbb{R}^n$ s.t.

$$\mu(A) = \mathbb{P}\big((X_1, \ldots, X_n) \in A\big).$$

Joint law (or joint distribution) of the X_i's

- In particular, for $g: \mathbb{R}^n \to \mathbb{R}$ measurable, then

$$\mathbb{E}[g(X_1, \dots, X_n)] = \int_{\mathbb{R}^n} g(x_1, \dots, x_n)\, d\mu(x_1, \dots, x_n)$$

Thm: A collection of r.v.'s $X_1, \dots, X_n$ is independent iff their joint law μ can be written as:

$$\mu = \mu_1 \times \mu_2 \times \dots \times \mu_n$$

where μ_i is the law of X_i.

Application of $\pi - \lambda$

Pf: ($\Leftarrow$) Assume μ is the product, then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \mathbb{E}\left(\prod_{i=1}^{n} \mathbb{1}_{\{X_i \in A_i\}}\right)$$

$$= \int_{\mathbb{R}^n} \prod_{i=1}^{n} \mathbb{1}_{\{x_i \in A_i\}}\, d\mu_1(x_1) \cdots d\mu_n(x_n)$$

$$= \prod_{i=1}^{n} \left(\int_{\mathbb{R}} \mathbb{1}_{\{x_i \in A_i\}}\, d\mu_i(x_i) \right)$$

$$= \prod_{i=1}^{n} \mathbb{P}(\{X_i \in A_i\}).$$

($\Rightarrow$) Assume the r.v.'s are independent. Then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(\{X_i \in A_i\})$$

$$\mu(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mu_i(A_i) = \mu_1 \times \cdots \times \mu_n (A_1 \times \cdots \times A_n)$$

Thus, $\mu = \mu_1 \times \cdots \times \mu_n$ by $\pi - \lambda$ thm (cylinder sets are a π-system) ∎

Ex: If X, Y are independent r.v.'s and integrable then

$$\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$$

$$\mathbb{E}[XY] = \int_{\mathbb{R}^2} xy \, d\mu_1(x) \, d\mu_2(y)$$

∎

Rmk: " $\mathbb{E}[XY] = \mathbb{E}X \, \mathbb{E}Y$ " is <u>NOT</u> equiv to X, Y being independent.

This is called "uncorrelated"

Def: Given r.v.'s X, Y s.t. $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$, define the covariance of X and Y by

$$\boxed{\begin{aligned}\mathrm{Cov}(X,Y) &:= \mathbb{E}\big[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\big] \\ &= \mathbb{E}(XY) - \mathbb{E}X\,\mathbb{E}Y\end{aligned}}$$

FACTS:

- If X, Y are indep. and $f, g : \mathbb{R} \to \mathbb{R}$ are measurable then

$$f(X) \text{ and } g(Y) \text{ are independent r.v.'s.}$$

Pf: $\sigma(X)$ and $\sigma(Y)$ are independent and $f(X)$ is measurable w.r.t. $\sigma(X)$. ■

- If X, Y, Z, W are independent r.v.'s and $f, g : \mathbb{R}^2 \to \mathbb{R}$ are measurable, then

$$f(X,Y) \text{ and } g(Z,W) \text{ are indep. r.v.'s}$$

Pf: Suffices to show that $\sigma(X,Y)$ is indep. from $\sigma(Z,W)$.

$$\sigma(X,Y) = \sigma\big(\sigma(X) \cup \sigma(Y)\big)$$

$$\sigma(Z,W) = \sigma\big(\sigma(Z) \cup \sigma(W)\big)$$

$$\mathbb{P}\big((X,Y)\in A, (Z,W)\in B\big) = \mathbb{P}\big((X,Y)\in A\big)\,\mathbb{P}\big((Z,W)\in B\big).$$

Every set in $\sigma(X,Y)$ is of the form $(X,Y)\in A$ for Borel $A \subset \mathbb{R}^2$.

Check:

(1) $\sigma(X,Y)$ is generated by π-system $\{X\in A\}\cap\{Y\in B\}$

(2) If A_1 and A_2 are indep. π-systems then $\sigma(A_1)$, $\sigma(A_2)$ are also indep. (π-λ thm) ■

Ex: We say X has the <u>Bernoulli</u> <u>p-distribution</u> if

$$\mathbb{P}(X=1)=p=1-\mathbb{P}(X=0)$$

If we have n indep. Bernoulli p-r.v.'s $X_1,\dots,X_n$ then $Y=X_1+\cdots+X_n$ is said to have the <u>binomial</u> <u>Binom(n,p)</u> <u>distribution</u> iff

$$\mathbb{P}(Y=k)=p^k(1-p)^{n-k}\binom{n}{k}$$

Ex: (MULTINOMIAL DISTRIBUTION) Let $\{Y_i\}_{i=1}^n$ be independent r.v.'s s.t.

$$\mathbb{P}(Y_i=j)=P_j\ ,\quad 1\le j\le m$$

and $\sum_{j=1}^m P_j=1$. Let $X_k=\#Y_i$ s.t. $Y_i=k$.

Then the X_k's are <u>NOT</u> indep r.v.'s $\rightsquigarrow \sum_{k=1}^m X_k=n \neq 1$

cannot happen for indep. r.v.'s

$$\mathbb{P}(X_1 = t_1, \dots, X_m = t_m)$$

$$= \left(p_1^{t_1} p_2^{t_2} \cdots p_m^{t_m} \frac{n!}{t_1!\, t_2! \cdots t_m!} \right) \mathbb{1}_{\{t_1 + \dots + t_m = n\}}$$

Thm: If $X_1, \dots, X_n$ are r.v.'s, then they are independent

iff $\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \geq t_i\} \right) = \prod_{i=1}^{n} \mathbb{P}(\{X_i \geq t_i\}) \quad \forall t_i \in [-\infty, \infty)$

iff $\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \leq t_i\} \right) = \prod_{i=1}^{n} \mathbb{P}(\{X_i \leq t_i\}) \quad \forall t_i \in (-\infty, \infty]$

Pf: Sets of the form $\prod_{i=1}^{n} [t_i, \infty)$ w/ $t_i \in [-\infty, \infty)$ are a π-system covering $\mathbb{R}^n$ that generates $\mathcal{B}(\mathbb{R}^n)$

$\Longrightarrow$ apply π-λ thm. ∎

Ex: We say X has the EXPONENTIAL DISTRIBUTION $\text{Exp}(\lambda)$, $\lambda > 0$, if $\mathbb{P}(X \geq t) = e^{-\lambda t} \quad \forall t \geq 0$.

$\iff$ X having density $= \mathbb{1}_{\{t \geq 0\}} e^{-\lambda t}$.

- Let $X_1, \dots, X_n$ be n indep $\mathrm{Exp}(\lambda)$ r.v.'s and set $Y := \min_i X_i$. Then

(i) Y is an $\mathrm{Exp}(n\lambda)$

Pf: $$\mathbb{P}(Y \geq t) = \mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \geq t\}\right) \overset{\text{indep.}}{=} \prod_{i=1}^{n} \mathbb{P}(\{X_i \geq t\}) = e^{-\lambda n t}$$ $\square$

(ii) Let $I = \arg\min_i X_i$ and let $Z_1, \dots, Z_{n-1}$ be the $n-1$ r.v.'s $\{X_i - X_I\}_{i \notin I}$. Then $Y, Z_1, \dots, Z_{n-1}$ are independent and each Z_i is $\mathrm{Exp}(\lambda)$. ← n clocks ringing w/ prob. $\mathrm{Exp}(\lambda)$

Compute: $t \geq 0$, $s_i \geq 0 \ \forall i$

$$\mathbb{P}(Y \geq t, Z_1 \geq s_1, \dots, Z_n \geq s_{n-1})$$

$$= n\,\mathbb{P}(I = n, Y \geq t, Z_1 \geq s_1, \dots, Z_{n-1} \geq s_{n-1})$$

$$= n\,\mathbb{P}(I = n, X_n \geq t, X_1 - X_n \geq s_1, \dots, X_{n-1} - X_n \geq s_{n-1})$$

$$= n\,\mathbb{P}(X_n \geq t, X_1 \geq s_1 + X_n, \dots, X_{n-1} \geq s_{n-1} + X_n)$$

$$= n \int_{\mathbb{R}^n} \mathbb{1}_{\{x_n \geq t\}} \prod_{i=1}^{n} \mathbb{1}_{\{x_i \geq s_i + x_n\}} \underbrace{\lambda^n e^{-\lambda(x_1 + \dots x_n)}\, dx_1 \cdots dx_n}_{\text{joint probability of } X_i\text{'s}}$$

Decide which integrals to compute first (Fubini)

$$= n\int_t^\infty \left(\underbrace{\int_{\mathbb{R}^{n-1}} \prod_{i=1}^{n-1} \mathbb{1}_{\{x_i \geq s_i + x_n\}} \lambda^{n-1} e^{-\lambda x_1} \cdots e^{-\lambda x_{n-1}} dx_1 \cdots dx_{n-1}}_{= \prod_{i=1}^{n-1} \int_{s_i+x_n}^{\infty} \lambda e^{-\lambda x_i} dx_i = \prod_{i=1}^{n-1} e^{-\lambda(s_i+x_n)}} \right) \lambda e^{-\lambda x_n} dx_n$$

$$= n \int_t^\infty \lambda e^{-\lambda n x_n} dx_n \cdot \prod_{i=1}^{n-1} e^{-\lambda(s_i + x_n)}$$

$$= e^{-\lambda n t} \prod_{i=1}^{n-1} e^{-\lambda s_i}.$$

$\blacksquare$

INFINITE SEQUENCES of R.V.'s: Do they exist?

- Let $X_1, X_2, \ldots$ be an infinite seq. of r.v.'s. Then, their joint distribution should be a measure on $\mathbb{R}^{\mathbb{N}}$ w/

$$\mathcal{F} = \sigma(A_1 \times A_2 \times \cdots \times A_n \times \mathbb{R} \times \mathbb{R} \times \cdots).$$

Suppose this μ exists. Then we get a measure μ_n on $\mathbb{R}^n$ by

$$\mu_n(A_1 \times \cdots \times A_n) := \mu(A_1 \times \cdots \times A_n \times \mathbb{R} \times \mathbb{R} \times \cdots).$$

So, $\forall n$, $\mu_{n+1}(A_1 \times \cdots \times A_n \times \mathbb{R}) = \mu(A_1 \times \cdots \times A_n \times \mathbb{R} \times \cdots)$

$$\overset{(*)}{=} \mu_n(A_1 \times \cdots \times A_n)$$

Thm: (Kolmogorov Extension) Let μ_n, $n \geq 1$, be a family of prob. measures on $\mathbb{R}^n$ s.t. (*) holds $\forall n$. Then, there exists a unique $\mathbb{P}$ on $\mathbb{R}^{\mathbb{N}}$ w/ $\mathcal{F} = \sigma(\text{cylinder sets})$ s.t.
$$\mathbb{P}(A_1 \times \cdots \times A_n) = \mu_n(A_1 \times \cdots \times A_n).$$

Cor: $\exists$ prob. spaces supporting inf. seq. of indep. r.v.'s.

Thm: Let X, Y be indep. r.v.'s w/ cdf's F and G. Then
$$\mathbb{P}(X+Y \leq t) = \int_{\mathbb{R}} F(t-y)\, d\mu_G(y), \qquad (*)$$
$\mu_G((-\infty, t)) = G(t)$. In particular, if X has density $f(x)$ and Y has density $g(x)$, then $X+Y$ has density given by
$$x \longmapsto (f * g)(x) \overset{\text{def}}{=} \int f(x-y) g(y)\, dy.$$

Pf: (of (*))
$$\mathbb{P}(X+Y \leq t) = \int \mathbb{1}_{\{x+y \leq t\}}\, d\mu_F(x)\, d\mu_G(y)$$
$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_{\{x+y \leq t\}}\, d\mu_F(x) \right) d\mu_G(y)$$

$$= \int_{\mathbb{R}} \mathbb{P}(X \le t - y)\, d\mu_G(y)$$

$$\overset{\text{def}}{=} \int_{\mathbb{R}} F(t-y)\, d\mu_G(y).$$

$\blacksquare$

Ex: Z has $N(\mu, \sigma^2)$ if its density $= \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\dfrac{-(z-\mu)^2}{2\sigma^2}\right)$.

If Z_1, Z_2 are indep. $N(\mu_i, \sigma_i^2)$, then $Z_1 + Z_2$ is

$$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Pf: Enough to do $\mu_1 = \mu_2 = 0$ and $\sigma_1^2 = \sigma_2^2 = 1$:

$$\int e^{-z^2/2\sigma^2} e^{-(x-z)^2/2\sigma^2}\, dz = \int \exp\left(-\frac{z^2}{2}\left(\frac{1}{\sigma^2} + 1\right) + xz - \frac{x^2}{2}\right) dz$$

$$= \int \exp\left(-\frac{1}{2}\left(z - \frac{x}{1+\frac{1}{\sigma^2}}\right)^2 - \frac{x^2}{2} + \frac{x^2}{2(1+1/\sigma^2)}\right) dz$$

$$\propto \exp\left(-\frac{x^2}{2} + \frac{x^2}{2(1+\frac{1}{\sigma^2})}\right)$$

$$= \exp\left(\frac{-x^2}{2(1+\sigma^2)}\right).$$

$\square$

<u>ASIDE</u>: Motivated by HW01

Panchenko 1.3: $X \geq 0$ r.v., $\mathbb{E}X = \int_0^\infty \mathbb{P}(X \geq t)\, dt$.

$$x \mathbb{1}_{\{x>0\}} \xrightarrow{\text{derivative}} \mathbb{1}_{\{x>0\}} \text{ everywhere except for } x=0.$$

$$\Rightarrow \quad x \mathbb{1}_{\{x>0\}} = \int_0^x \mathbb{1}_{\{t>0\}}\, dt = \int_0^\infty \mathbb{1}_{\{x>t\}}\, dt$$

So,

$$X = \int_0^\infty \mathbb{1}_{\{X>t\}}\, dt \quad \text{a.s.}$$

$= x\mathbb{1}_{\{x>0\}}$ and $\{x=0\}$ has Lebesgue measure 0.

$X \geq 0 \Rightarrow x\mathbb{1}_{\{x>0\}} = X$ a.s.

$$\mathbb{E}X = \mathbb{E}\left(\int_0^\infty \mathbb{1}_{\{X>t\}}\, dt\right)$$

$$= \int_{-\infty}^\infty \int_0^\infty \mathbb{1}_{\{X>t\}}\, dt\, d\mu(x)$$

$$\overset{\text{Fubini}}{=} \int_0^\infty \underbrace{\int_{-\infty}^\infty \mathbb{1}_{\{X>t\}}\, d\mu(x)}_{=\mathbb{P}(X>t)}\, dt$$

$$= \int_0^\infty \mathbb{P}(X>t)\, dt.$$

Now, look at the square:

$$\mathbb{1}_{\{x^2>0\}} \xrightarrow{\text{derivative}} 2x\, \mathbb{1}_{\{x^2>0\}} \overset{x \geq 0}{=} 2x\, \mathbb{1}_{\{x>0\}}$$

$$\Longrightarrow \mathbb{1}\{x^2>0\} = \int_0^x 2t\,\mathbb{1}\{t>0\}\,dt = \int_0^\infty 2t\,\mathbb{1}\{x>t\}\,dt$$

So, for r.v. $X \geq 0$

$$X^2 = \int_0^\infty 2t\,\mathbb{1}\{X>t\}\,dt$$

$$\mathbb{E}X^2 = \mathbb{E}\left(\int_0^\infty 2t\,\mathbb{1}\{X>t\}\,dt\right)$$

$$= \int_{-\infty}^\infty \int_0^\infty 2t\,\mathbb{1}\{x>t\}\,dt\,d\mu(x)$$

$$\overset{\text{Fubini}}{=} \int_0^\infty 2t\left(\int_{-\infty}^\infty \mathbb{1}\{x>t\}\,d\mu(x)\right)dt$$

$$= \int_0^\infty 2t\,\mathbb{P}(X>t)\,dt$$

<u>Generalize</u>: Take a r.v. $X \geq 0$.

$$\mathbb{1}\{x^n>0\} \xrightarrow{\text{derivative}} n x^{n-1}\,\mathbb{1}\{x^n>0\} \overset{x\geq 0}{=} n x^{n-1}\,\mathbb{1}\{x>0\}$$

$$X^n = \mathbb{1}\{X^n>0\} = \int_0^\infty n t^{n-1}\,\mathbb{1}\{X>t\}\,dt$$

$\downarrow$ Expectation on both sides

$$\mathbb{E}X^n = \mathbb{E}\left(\int_0^\infty n t^{n-1} \mathbb{1}_{\{X>t\}}\, dt\right)$$

$$= \int_{-\infty}^{\infty} \int_0^\infty n t^{n-1} \mathbb{1}_{\{x>t\}}\, dt\, d\mu(x)$$

$$\overset{\text{Fubini}}{=} \int_0^\infty n t^{n-1} \left(\int_{-\infty}^{\infty} \mathbb{1}_{\{x>t\}}\, d\mu(x)\right) dt$$

$$= \int_0^\infty n t^{n-1}\, \mathbb{P}(X>t)\, dt .$$

LECTURE 5

Sep 18th, 2024

TERMINOLOGY: • iid → "independent identically distributed"

- $X \sim \text{Exp}(1) \longrightarrow \mathbb{P}(X \geq t) = e^{-t}$, $t \geq 0$.
- Almost surely (a.s.) → Event A holds a.s. if $\mathbb{P}(A) = 1$.
  If $X \leq Z$ a.s., then $\mathbb{E}(X) \leq \mathbb{E}(Z)$

Ex: (RANDOM GRAPH) Let $\{X_i\}_{i=1}^n$ be iid w/ $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1-p$. Let $Y = \#$ pairs of X_i that are both equal to 1

Claim: $\mathbb{E}Y$ is increasing in p.

Pf: Let $\{U_i\}_{i=1}^n$ be iid uniform $(0,1)$ r.v.'s and let $X_i^{(p)} := \mathbb{1}_{\{U_i \le p\}}$.

So, $X_i^{(p)}$ are iid w/ distribution $\mathbb{P}(X_i^{(p)}=1) = p = 1-\mathbb{P}(X_i^{(p)}=0)$

Let $Y^{(p)} :=$ # of pairs of $X_i^{(p)}$ that are both $=1$. $\Longrightarrow \mathbb{E}Y^{(p)} = \mathbb{E}Y$.

Point: If $p_1, p_2 > 0$, then $X_i^{(p_1)} = 1 \Longrightarrow X_i^{(p_2)} = 1$

In particular, $Y^{(p_1)} \le Y^{(p_2)}$ a.s.

$\Rightarrow \mathbb{E}Y^{(p_1)} \le \mathbb{E}Y^{(p_2)}$. ← USE IN HW01

Ex: Suppose $\{X_i^{(p_1)}\}_{i=1}^n$ and $\{X_i^{(p_2)}\}_{i=1}^n$ are $2n$ independent r.v.'s with Bernoulli(p) distribution.

Construct $\hat{Y}^{(p_1)}$ and $\hat{Y}^{(p_2)}$ as above.

It is not true that $\hat{Y}^{(p_2)} \le \hat{Y}^{(p_2)}$ a.s.

independent

---

* **WEAK LAW OF LARGE NUMBERS**:

Mode of convergence in the theorem. Weak ⟷ convergence in probability.

Def: We say $\{X_n\}_{n=1}^{\infty}$ converges in probability to X if $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n \nearrow \infty} 0$.

FACTS: (1) This convergence is <u>metrizable</u>. Namely, it induces a topology on the space of r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$ that is metrizable.

(2) If $X_n \xrightarrow{\text{prob}} X$ and $Y_n \xrightarrow{\text{prob}} Y$ then $X_n + Y_n \xrightarrow{\text{prob}} X + Y$.

(3) If $X_n \xrightarrow{\text{prob}} X$ and $X_n \xrightarrow{\text{prob}} Z$, then $X = Z$ a.s.

IMPORTANT

FACT: If $\{X_i\}_{i=1}^n$ are independent r.v.'s w/ $\mathrm{Var}(X_i) < \infty$ $\forall i$, then

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Pf: For any X, $\mathrm{Var}(X) \overset{\text{def}}{=} \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \mathbb{E}(X - \mathbb{E}X)^2$

Moreover, $\forall$ scalars α, $\mathrm{Var}(X+a) = \mathrm{Var}(X)$.

Assume $\mathbb{E}X_i = 0$ ("centered"). Thus

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \mathbb{E}\left(\sum_{i=1}^n X_i\right)^2 = \sum_{i,j} \mathbb{E}X_iX_j$$

$$= \sum_{i=1}^n \mathbb{E}X_i^2 + \sum_{i \neq j} \underbrace{\mathbb{E}(X_iX_j)}_{= \underbrace{(\mathbb{E}X_i)}_{= 0 \text{ by assumption}} \mathbb{E}X_j}$$

$$= \sum_{i=1}^n \mathbb{E}X_i^2$$

$$= \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

$\blacksquare$

Upshot: $\operatorname{Var}(\alpha X) = \alpha^2 \operatorname{Var}(X)$

$\operatorname{Var}(nX) = n^2 \operatorname{Var}(X)$

WEAK LLN

Thm: If $\{X_i\}_{i=1}^{n}$ are independent r.v.'s are s.t.

[Also holds if we assume X_i are uncorrelated; i.e., $\operatorname{Cov}(X_i, X_j) = 0$]

$$\frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var}(X_i) \xrightarrow{n \nearrow \infty} 0$$

then

$$\frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} X_i \xrightarrow[n \nearrow \infty]{\text{in probability}} 0$$

(and also in L^P: $Y_n \xrightarrow{L^P} Y$ if $\mathbb{E}|Y - Y_n|^P \to 0$)

Cor: If X_i are iid and $\operatorname{Var}(X_i) < \infty \ \forall i$, then

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[\& L^2]{\text{in prob}} \mathbb{E} X_i \qquad \left(\frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var}(X_i) = \frac{1}{n} \operatorname{Var}(X_i) \right)$$

Pf: (of Thm) Can assume $\mathbb{E} X_i = 0$. Thus

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i > \varepsilon \right| \right) \overset{\text{Chebyshev}}{\leq} \frac{1}{\varepsilon^2} \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2$$

$$= \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n} \mathbb{E}X_i^2$$

$$= \frac{1}{\varepsilon^2}\underbrace{\left(\frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i)\right)}_{\xrightarrow[n\nearrow\infty]{} 0} \qquad \blacksquare$$

⚠ **Important trick**: assume $\mathbb{E}X_i = 0$ and then we have that

$$\mathbb{E}\left(\sum_i X_i\right)^2 = \sum_i \mathbb{E}X_i^2 \quad \left[\text{IF THE } X_i\text{'S ARE INDEPENDENT}\right]$$

Note: in the centered iid case of $\{X_i\}_{i=1}^{n}$, we get

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \varepsilon\right) \le \frac{1}{\varepsilon^2}\frac{1}{n}\mathrm{Var}(X_i).$$

⚠

Thm: Let $\{X_i\}_{i=1}^{n}$ be iid and assume $\mathbb{E}X_i = 0$ and $\mathbb{E}|X_i|^4 < \infty$. Then, $\exists C > 0$ st.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \varepsilon\right) \le \frac{1}{\varepsilon^4}\frac{C}{n^2}$$

LECTURE 6

Sep 20th, 2024

Today: Improve bound from last time; i.e., $X_1, \dots, X_n$ indep. centered r.v.'s and $\mathrm{Var}(X_i) \le C$, then

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i \right| > \varepsilon \right) \le \frac{C}{n\varepsilon^2}.$$

Thm: Let $X_1, \dots, X_n$ be centered indep. r.v.'s s.t.

$$\mathbb{E} X_i^4 \le C \quad \forall i.$$

Then

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i \right| > \varepsilon \right) \le \frac{C}{n^2\varepsilon^4}.$$

Pf:
$$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^4 = \frac{1}{n^4} \sum_{i,j,k,l} \mathbb{E}(X_i X_j X_k X_l)$$

$\binom{4}{2}$ but doesn't matter here...

Centered $\Leftrightarrow$ $\mathbb{E} X_i = 0$ $\forall i$

$$= \frac{1}{n^4} \sum_i \mathbb{E} X_i^4 + \frac{1}{n^4} \sum_{i \ne j} \mathbb{E}(X_i^2 X_j^2)$$

$$+ \frac{1}{n^4} \sum_{i \ne j} \cancel{\mathbb{E}(X_i X_j^3)}_{=0} + \frac{1}{n^4} \sum_{i \ne j \ne k} \cancel{\mathbb{E}(X_i X_j X_k^2)}_{=0}$$

$=0$ b/c indep & centered

$$+ \frac{1}{n^4} \sum_{i \ne j \ne k \ne l} \cancel{\mathbb{E}(X_i X_j X_k X_l)} = 0$$

b/c the X_i's are indep and centered

$$= \frac{1}{n^4} \sum_i \mathbb{E} X_i^4 + \frac{1}{n^4} \sum_{i \neq j} (\mathbb{E} X_i^2)(\mathbb{E} X_j^2)$$

$$\leq \frac{nC}{n^4} + \frac{n^2 \tilde{C}}{n^4} \leq \frac{C'}{n^4}.$$

$\blacksquare$

FACT: Let $p \in \mathbb{N}$ and assume $\mathbb{E} X_i^{2p} \leq C \ \forall i$. Let $a_i \in \mathbb{R}$
X_i centered & independent
then

$$\mathbb{E} \left(\sum_{i=1}^n a_i X_i \right)^{2p} \leq C_p \left(\sum_{i=1}^n a_i^2 \right)^p .$$

Pf: Similar to thm above by breaking up the sum.

Cor: $$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| > \varepsilon \right) \leq \frac{C_p}{n^p \varepsilon^{2p}} = \frac{C_p}{(\varepsilon^2 n)^p}$$ (Markov...)

Suggests that $\varepsilon := u/\sqrt{n}$, then

$$\mathbb{P}\left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right| > u \right) \leq \frac{C_p}{u^p}.$$

Def: Given a r.v. X, the <u>moment generating function</u> (MGF) of X is defined as

$$t \longmapsto \mathbb{E}\left(e^{tX}\right) =: \varphi_X(t)$$

FACT: If MGF is finite for $t \in (-a,a)$, then X has moments of all orders and they can be found by differentiating the MGF:

requires proof

$$\frac{d}{dt}\varphi_X(t) = \frac{d}{dt}\mathbb{E}(e^{tX}) = \mathbb{E}\left(\frac{d}{dt}(e^{tX})\right) = \mathbb{E}(Xe^{tX})$$

$\Rightarrow \varphi_X'(0) = \mathbb{E}X$ and so on for the higher moments.

Lemma: If X is centered and $|X| \leq 1$ a.s., then

$$\mathbb{E}(e^{sX}) \leq e^{s^2/2} \quad \forall s \in \mathbb{R}.$$

Pf: $$sX = \underbrace{\left(\frac{1+X}{2}\right)}_{\in [0,1]} s + \underbrace{\left(\frac{1-X}{2}\right)}_{\in [0,1]}(-s).$$

By convexity of the exp.,

$$e^{sX} \leq \frac{1+X}{2}e^{s} + \frac{1-X}{2}e^{-s}$$

$$\Rightarrow \mathbb{E}(e^{sX}) \leq \frac{e^{s}+e^{-s}}{2} \leq e^{s^2/2}$$

← By comparing power series.

∎

Use this to prove

Thm: (Hoeffding) Let $X_1, \ldots, X_n$ be centered independent r.v.'s such that $|X_i| \leq 1$ $\forall i$. Let $a_i \in \mathbb{R}$, then

$$\mathbb{P}\left(\left|\sum_{i=1}^n a_i X_i\right| > t\right) \leq 2 e^{-t^2/2\sum_{i=1}^n a_i^2}.$$

↳ Rmk: $\mathrm{Var}\left(\sum_i a_i X_i\right) = \sum_i a_i^2 \mathrm{Var}(X_i) \approx \sum_i a_i^2$

CLT ⇒ we expect that $\sum_i a_i X_i \sim N(0, \sigma^2)$

If $Z \sim N(0,\sigma^2)$, then $\mathbb{P}(|Z| > t) \lesssim e^{-t^2/2\sigma^2}$.

Pf: $$\mathbb{P}\left(\sum a_i X_i > t\right) \underset{\substack{\uparrow \\ \text{Chebyshev} \\ \text{w/ } \lambda > 0}}{=} \mathbb{P}\left(e^{\lambda \sum a_i X_i} > e^{\lambda t}\right)$$

Markov
$$\leq e^{-\lambda t} \mathbb{E}\left(e^{\lambda \sum_{i=1}^n a_i X_i}\right)$$

$$= e^{-\lambda t} \prod_{i=1}^n \mathbb{E}\left(e^{\lambda a_i X_i}\right)$$

Lemma
$$\leq e^{-\lambda t} \prod_{i=1}^n e^{\lambda^2 a_i^2/2}$$

$$= \exp\left(-\lambda t + \frac{\lambda^2}{2}\sum_{i=1}^n a_i^2\right)$$

Minimized at $\lambda = t/\sum_i a_i^2$.

Plug in

$$= \exp\left(-\frac{t^2}{2\sum_i a_i^2}\right).$$

concentration result → $\blacksquare$

(b/c it shows that the distrib. of $\left|\sum a_i X_i\right|$ is concentrated around 0)

* **APPLICATIONS**:

1) APPROXIMATION BY BERNSTEIN POLYNOMIALS

Let f be a continuous function on $[0,1]$. Define

$$f_n(p) := \sum_{j=0}^{n} f\left(\frac{j}{n}\right)\binom{n}{j} p^j (1-p)^{n-j}.$$

Claim: $\lim_{n\to\infty} \| f_n - f \|_{L^\infty[0,1]}$.

$P(B_i = 1) = p = 1 - P(B_i = 0)$

Pf: Let $B_1, \dots, B_n$ be n indep. Bernoulli(p) r.v.'s. Let

$$S_n = \frac{1}{n}\sum_{i=1}^{n} B_i$$

$\left(\implies nS_n \sim \text{Binom}(n,p) \text{ i.e., } P(nS_n = j) = \binom{n}{j} p^j (1-p)^{n-j} \right)$

Then

$$f_n(p) = \mathbb{E} f(S_n).$$

Now, $\mathbb{E}B_i = p$ and $\mathrm{Var}(B_i) = \mathbb{E}B_i^2 - (\mathbb{E}B_i)^2$

$$= p - p^2 = p(p-1).$$

Moreover,

$$
\begin{aligned}
\mathbb{P}(|S_n - p| > \delta) &\overset{\text{Chebyshev}}{\leq} \frac{1}{\delta^2}\mathrm{Var}(S_n) \\
&= \frac{1}{\delta^2}\frac{1}{n}\mathrm{Var}(B_i) \\
&= \frac{1}{\delta^2}\frac{1}{n}\underbrace{p(1-p)}_{\leq 1} \\
&\leq \frac{1}{n\delta^2}.
\end{aligned}
$$

f is abs. cont.

Let $\varepsilon > 0$ and choose $\delta > 0$ s.t. $|f(x) - f(y)| < \varepsilon \ \forall |x-y| < \delta$.

Then

$$
\begin{aligned}
|f_n(p) - f(p)| &= |\mathbb{E}f(S_n) - \mathbb{E}f(p)| \\
&= |\mathbb{E}(f(S_n) - f(p))| \\
&\leq \mathbb{E}(|f(S_n) - f(p)|) \\
&= \mathbb{E}\left(|f(S_n) - f(p)| \mathbb{1}_{\{|S_n - p| < \delta\}}\right) \\
&\quad + \mathbb{E}\left(|f(S_n) - f(p)| \mathbb{1}_{\{|S_n - p| > \delta\}}\right)
\end{aligned}
$$

$$\leq \varepsilon + 2\|f\|_\infty \mathbb{P}(|S_n - p| > \delta)$$

$$\leq \varepsilon + \frac{2\|f\|_\infty}{n\delta^2} \leq 2\varepsilon \text{ if } n \text{ is large enough.}$$

■

general sum of r.v.'s

OBS: If we have S_n and we define $\mu_n := \mathbb{E} S_n$, $\sigma_n^2 := \mathrm{Var}(S_n)$. Then, for any sequence b_n s.t. $\frac{b_n^2}{\sigma_n^2} \to \infty$ as $n \to \infty$. We have $\frac{S_n - \mu_n}{b_n} \xrightarrow{\text{in prob.}} 0$

$\frac{S_n}{\mu_n} \to 1$ in prob. (if you can take $b_n = \mu_n$).

Ex: (COUPON COLLECTOR PROBLEM) Let X_i be iid unif. distrib. on $\{1, 2, \ldots, n\}$ (i.e., $\mathbb{P}(X_i = j) = 1/n$).

Define $\tau_k := \inf\{m : |\{X_1, \ldots, X_m\}| \geq k\}$. e.g.: $\tau_1 = 1$
set $\tau_0 := 0$.

1^{st} time something happens

Define $Y_k := \tau_k - \tau_{k-1}$. ← If I have $k-1$ cards, how long it takes for me to get a new one

Note: $\{Y_k\}_{k=1}^n$ are independent.

$Y_k \sim \mathrm{Geom}\left(1 - \frac{k-1}{n}\right)$. $\left[\text{We say } Z \sim \mathrm{Geom}(p),\ p \in (0,1) \text{ if } \mathbb{P}(Z = k) = p(1-p)^{k-1}.\right]$

$$\mathbb{P}(X_i = \text{new}) = 1 - \frac{k-1}{n}$$

If $B_1, B_2, \ldots \sim \text{Bern}(p)$, then Z is the first time you see 1

Now, we are interested in how long it takes to get the whole set, namely:

$$T_n = \sum_{j=1}^{n} Y_j$$

How long it takes to get the full set of cards

<u>Lemma</u>: if $Z \sim \text{Geom}(p)$, then $\mathbb{E}Z = \frac{1}{p}$, $\text{Var}(Z) = \frac{1-p}{p^2} \leq \frac{1}{p^2}$.

$$\text{So, } \mathbb{E}T_n = \sum_{j=1}^{n} \mathbb{E}Y_j = \sum_{j=1}^{n} \frac{1}{1 - \frac{j-1}{n}}$$

$$= n \sum_{j=1}^{n} \frac{1}{n+1-j} = n \sum_{j=1}^{n} \frac{1}{j} = n \log n (1 + o(1))$$

Y_j indep

$$\text{Var}(T_n) = \sum_{j=1}^{n} \text{Var}(Y_j) \leq \sum_{j=1}^{n} \frac{1}{\left(1 - \frac{j-1}{n}\right)^2} = n^2 \sum_{j=1}^{n} \frac{1}{(n+1-j)^2}$$

$$= n^2 \sum_{j=1}^{n} \frac{1}{j^2} \leq Cn^2.$$

Here $\mu_n = n \log n (1 + o(1))$

$\sigma_n^2 \leq Cn^2$

$$\xRightarrow{\text{observation above}} \frac{T_n - n}{n \log n} \xrightarrow{\text{in prob.}} 0$$

$$\Longrightarrow \frac{T_n}{n \log n} \xrightarrow{\text{prob}} 1 \Longrightarrow T_n \simeq n \log n.$$

e.g.: $n = 365$, $365 \log 365 \simeq 2163$.

LECTURE 7

Sep 25th, 2024

SEE PROBABILITY THEORY - STROOCK

Def: We say $X_n \to X$ almost surely if

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

$$\omega \in \left\{\lim_{n\to\infty} X_n = X\right\} \iff \forall \varepsilon > 0 \;\exists N > 0 \text{ s.t. } |X_n(\omega) - X(\omega)| < \varepsilon \quad \forall n \geq N.$$

FACTS: • supremum of countably many meas. fcts. is a measurable fct.

- limit of meas. fcts. is meas.
- $\limsup_{x\to\infty} X_n(\omega)$ and $\liminf_{x\to\infty} X_n(\omega)$ are meas.
- $\left\{\lim_{n\to\infty} X_n(\omega) \text{ exists}\right\}$ is measurable.

Prop: If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{\text{prob.}} X$.

Pf: If $X_n \xrightarrow{a.s.} X$, then $\mathbb{P}\Big(\bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m\geq n}^{\infty} \{|X_m - X| > \varepsilon\}}_{=: B_n}\Big) = 0 \quad \forall \varepsilon.$

So, B_n is a decreasing seq. of events. So,

elementary property of prob. measures

$$0 = \lim_{n\to\infty} \mathbb{P}(B_n).$$

But
$$\mathbb{P}\left(|X_n - X| > \varepsilon\right) \leq \mathbb{P}(B_n) \xrightarrow{n\to\infty} 0.$$

$\blacksquare$

Def: Let A_n be a sequence of events. We define

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m\geq n}^{\infty} A_m = \left\{ \omega \in \Omega : \begin{array}{c} \omega \in A_n \text{ for} \\ \infty\text{-many } n \end{array} \right\}$$

(This is the event that A_n occurs ∞-often; i.e., $\{A_n \text{ i.o.}\}$)

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m\geq n}^{\infty} A_m = \left\{ \omega \in \Omega : \begin{array}{c} \omega \in A_n \text{ except for} \\ \text{at most finitely} \\ \text{many } n \end{array} \right\}$$

(This is the event that all but finitely many A_n occur)

FACT:
- $\limsup_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup_{n\to\infty} A_n}$
- $\liminf_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf_{n\to\infty} A_n}$

Ex: Let $\{X_i\}_{i=1}^{\infty}$ be iid Bernoulli-p r.v.'s, $p \in (0,1)$.
$\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Consider the event
$$A_n := \{X_n = 1\}.$$

Thus, $\underbrace{\mathbb{P}\left(\limsup_{n\to\infty} A_n\right)}$ $= 1$

Let $\omega = (X_1, X_2, X_3, \dots)$, then $\limsup_{n\to\infty} A_n = \left\{ \begin{array}{l} \text{the sequence } \omega \\ \text{having } \infty\text{-many } 1\text{s} \end{array} \right\}$

$\Omega = \{0,1\}^{\mathbb{N}}$.

$\underbrace{\mathbb{P}\left(\liminf_{n\to\infty} A_n\right)}$ $= 0$

$\liminf_{n\to\infty} A_n = \left\{ \begin{array}{l} \text{the sequence } \omega \text{ eventually} \\ \text{becomes only } 1\text{s}. \end{array} \right\}$ e.g.: $\omega = (1,0,0,1,1,1,\dots)$

!

$\boxed{\begin{array}{l} \textbf{Borel-Cantelli I: If } A_n \text{ is a sequence of events s.t.} \\ \qquad \displaystyle\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \\ \text{then} \\ \qquad \mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}\left(\limsup_{n\to\infty} A_n\right) = 0. \end{array}}$

Converse is false!

Pf: $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m\geq n}^{\infty} A_m$. So, $\forall n$,

$$\mathbb{P}\left(\limsup_{n\to\infty} A_n\right) \leq \mathbb{P}\left(\bigcup_{m\geq n}^{\infty} A_m\right) \leq \sum_{m=n}^{\infty} \mathbb{P}(A_m) \xrightarrow{n\nearrow\infty} 0.$$

intersection of decreasing sequence of events

uniform bound by assumption

■

REMARK: Let $N := \sum_{k=1}^{\infty} \mathbb{1}_{\{A_k\}}$. By MCT,

$$\mathbb{E}N = \sum_{k=1}^{\infty} \mathbb{E}\,\mathbb{1}_{A_k} = \sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$$

and $N = \#$ of A_k that occur.

CONVERSE TO BOREL-CANTELLI I IS FALSE: If U is uniform on $(0,1)$ r.v. and

$$A_n = \{U \leq \tfrac{1}{n}\}$$

$$\{A_n \text{ i.o.}\} = \{U = 0\} \Longrightarrow \mathbb{P}(A_n \text{ i.o.}) = 0$$

However,
$$\sum_{m=1}^{\infty} \mathbb{P}(A_m) = \sum_{m=1}^{\infty} \frac{1}{m} = \infty .$$

$\blacktriangle$

Thm: If $\{X_i\}_{i=1}^{\infty}$ are centered, independent r.v.'s s.t.

$$\sup_i \mathbb{E}X_i^4 < \infty .$$

Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s} 0 .$$

Pf: Already used Markov's ineq. to bound:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \varepsilon\right) \leq \frac{C}{n^2\varepsilon^4}.$$

summable

But, by Borel-Cantelli I,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \varepsilon \text{ for } \infty\text{-many } n\right) = 0$$

$$\|$$

$$\mathbb{P}\left(\bigcup_{m=1}^{\infty}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| > \frac{1}{m} \text{ for } \infty\text{-many } n\right\}\right) = 0$$

$\blacksquare$

Thm: Let X_n be a sequence of r.v.'s. Then

$X_n \xrightarrow{\text{prob.}} X \iff$ every subsequence has a further subsequence that converges a.s. to X

Recall: (useful lemma) $y_n \in \mathbb{R}$, $y_n \to y \iff$ every subsequence has a further subseq. converging to y

Pf: ($\Rightarrow$) Say $X_n \xrightarrow{\text{prob.}} X$. Let X_{n_k} be a subseq.. Choose some $\varepsilon_j \to 0$ and a subseq. n_{k_j} s.t.

$$\mathbb{P}\left(|X_{n_{k_j}} - X| > \varepsilon_j\right) \leq \frac{1}{2^j}.$$

summable.

By Borel-Cantelli I, we have

$$\mathbb{P}\left(|X_{n_{k_j}} - X| > \varepsilon_j \text{ i.o.}\right) = 0 \Rightarrow X_{n_{k_j}} \xrightarrow{\text{a.s.}} X.$$

($\Leftarrow$) Let $\varepsilon > 0$ and let $y_n := \mathbb{P}(|X_n - X| > \varepsilon)$.

WTS: $y_n \to 0$.

But $y_n \in \mathbb{R}$ and, for any subseq. n_k, $\exists\, n_{k_j}$ s.t. $X_{n_{k_j}} \xrightarrow{a.s.} X$

As shown before, a.s. $\Rightarrow$ prob., so $X_{n_{k_j}} \xrightarrow{prob.} X$

i.e., $y_{n_{k_j}} \xrightarrow{j \nearrow \infty} 0$. So, by the real anal. lemma in blue,

we are done. $\blacksquare$

LECTURE 8

BOREL CANTELLI II: If events A_n are independent and

$$\sum_n \mathbb{P}(A_n) = +\infty,$$

then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Pf: $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m \geq n} A_m}_{=: B_n}$.

Note that B_n's are decreasing. So, by continuity of the measure, suffices to show that $\mathbb{P}(B_n) = 1 \Leftrightarrow \mathbb{P}(B_n^c) = 0$.

So, $$\mathbb{P}(B_n^c) = \mathbb{P}\Big(\bigcap_{m\geq n} A_m^c\Big) \leq \mathbb{P}\Big(\bigcap_{m=n}^{k} A_m^c\Big)$$

$$\overset{\text{independence}}{=} \prod_{m=n}^{k} \big(1 - \mathbb{P}(A_m)\big)$$

We have: $1 - x \leq e^{-x} \quad \forall x \geq 0$

$$\overset{\text{concavity of log}}{\iff} \log(1-x) \leq -x.$$

$$e^{\log(1-x)} \;(\text{"} = \text{" } 1-x)$$

Using that inequality, we find that

$$\mathbb{P}(B_n^c) \leq \prod_{m=n}^{k} \big(1 - \mathbb{P}(A_m)\big)$$

$$\leq \prod_{m=n}^{k} e^{-\mathbb{P}(A_m)}$$

$$= \exp\Big(-\sum_{m=n}^{k} \mathbb{P}(A_m)\Big) \xrightarrow{k \nearrow \infty} 0.$$

$\blacksquare$

APPLICATION OF BOREL CANTELLI II:

Thm: If $X_1, X_2, X_3, \ldots$ are independent random variables and $\mathbb{E}|X_i| = \infty$, then

$$\mathbb{P}\Big(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \text{ exists in } (-\infty,\infty)\Big) = 0$$

Pf: Look at the increments (if lim exists, the increments $\to 0$)

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n+1}\sum_{i=1}^{n+1} X_i = -\frac{X_{n+1}}{n+1} + \left(\frac{1}{n} - \frac{1}{n+1}\right)\sum_{i=1}^{n} X_i$$

$$= -\frac{X_{n+1}}{n+1} + \frac{1}{n(n+1)}\sum_{i=1}^{n} X_i$$

On the event that $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i$ exists, the equality tells us that $\lim_{n\to\infty} \frac{X_n}{n} = 0$.

Suffices to show that $\mathbb{P}\left(\lim_{n\to\infty} \frac{X_n}{n} = 0\right) = 0$.

X_n indep $\Rightarrow$ A_n indep

Let $A_n := \{|X_n| \geq n\}$. By Borel Cantelli II, it suffices to show that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$.

Note that

$$+\infty = \mathbb{E}|X_i| = \int_0^{\infty} \mathbb{P}(|X_i| > s)\, ds$$

$$= \sum_{j=0}^{\infty} \int_j^{j+1} \mathbb{P}(|X_i| > s)\, ds$$

$$\leq \sum_{j=0}^{\infty} \mathbb{P}(|X_i| \geq j)\, ds \overset{X_i \text{ are iid}}{=} \sum_{j=0}^{\infty} \mathbb{P}(A_j).$$

∎

* Strengthen Borel - Cantelli to get the rate of how things go to infinity:

Thm: Let $A_1, A_2, A_3, \ldots$ be a sequence of pairwise independent events s.t.

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty .$$

Then

$$\frac{\sum_{j=1}^{n} \mathbb{1}_{A_j}}{\sum_{j=1}^{n} \mathbb{P}(A_j)} \xrightarrow{a.s.} 1$$

Pf: Let $S_n := \sum_{j=1}^{n} \mathbb{1}_{A_j}$. Then

$$\mathrm{Var}(S_n) \overset{\substack{\text{pairwise}\\ \text{independence}}}{=} \sum_{j=1}^{n} \mathrm{Var}(\mathbb{1}_{A_j}) = \sum_{j=1}^{n} \mathbb{P}(A_j) - \mathbb{P}(A_j)^2$$

$$\leq \sum_{j=1}^{n} \mathbb{P}(A_j) = \mathbb{E} S_n .$$

Use Chebyshev: let $\delta > 0$

$$\mathbb{P}\left(\frac{|S_n - \mathbb{E}S_n|}{\mathbb{E}S_n} > \delta\right) \overset{\text{Chebyshev}}{\leq} \frac{1}{\delta^2 (\mathbb{E}S_n)^2} \mathrm{Var}(S_n)$$

$$\leq \frac{1}{\delta^2 \mathbb{E}S_n} \longrightarrow 0 \text{ as } n \to \infty.$$

This shows convergence in probability. Need to strengthen this to a.s. Standard technique to do that

$\forall k \geq 1$, let $n_k := \inf\{n \text{ s.t. } \mathbb{E}S_n \geq k^2\}$ and

let $T_k := S_{n_k}$

<u>Note</u>: $k^2 \leq \mathbb{E}T_k \leq (k+1)^2$

b/c $|\mathbb{E}S_n - \mathbb{E}S_{n+1}| \leq 1$ since the S_n's are just a bunch of indicator fcts. Can't jump more than 1.

By Borel-Cantelli I, $\frac{T_k}{\mathbb{E}T_k} \xrightarrow{a.s.} 1$

$\left(Z_n := \frac{T_k}{\mathbb{E}T_k} - 1 \quad \{Z_n\}_{n=1}^\infty, \forall \delta > 0 \right.$

$\mathbb{P}(|Z_n| > \delta \ \infty\text{-many } n) = 0$. Then

$\left. \delta = \frac{1}{m}, \ \mathbb{P}\left(\bigcup_m \{|Z_n| > 1/m\} \ \infty\text{-many } n\right) = 0 \right)$

$$\mathbb{P}\left(\left|\frac{T_k}{\mathbb{E}T_k} - 1\right| > \delta\right) \leq \frac{1}{k^2 \delta^2}$$ summable

For $n \in [n_k, n_{k+1}]$, we have $T_k \le S_n \le T_{k+1}$. So,

$$\frac{T_k}{\mathbb{E}T_{k+1}} \le \frac{S_n}{\mathbb{E}S_n} \le \frac{T_{k+1}}{\mathbb{E}T_k}. \quad (*)$$

Since $\dfrac{\mathbb{E}T_k}{\mathbb{E}T_{k+1}} \longrightarrow 1$. Both sides of $(*)$ tend to 1. Thus, by the Squeeze Thm, $\dfrac{S_n}{\mathbb{E}S_n} \longrightarrow 1$ too. $\blacksquare$

FACT: If X and Y are independent and X has a continuous distribution fct, then $P(X=Y)=0$.

X has continuous cdf $\Longrightarrow \mathbb{P}(X=s)=0 \ \forall s \in \mathbb{R}$.

Note: $P(X=Y) = \iint \mathbb{1}_{\{x=y\}} \, d\mu_X(x) \, d\mu_Y(y)$ Fubini...

Lecture 9

Oct 4th, 2024

Ex: (Record Values) Suppose $X_1, X_2, \dots$ are iid w/ continuous distribution $\Longrightarrow X_i \neq X_j \ \forall i \neq j$ a.s. "long-jump attempts"

(i.e., no ties in the jumps)

Let $A_k := \{X_k > X_j \ \forall 1 \le j \le k\}$

Cl: The events A_k are independent and $P(A_k) = \frac{1}{k}$.

these only depend on the relative ordering of the X's.

Pf: Let $\sigma \in S_n$ be a permutation s.t.

$$X_{\sigma(1)} > X_{\sigma(2)} > X_{\sigma(3)} > \dots$$

i.e., $\sigma(j)$ is the index of the j-th largest value.

Cl: σ is uniformly distributed in S_n.

Pf: This is the case b/c the X's are iid $\Rightarrow$ no permutation can be more likely than the other ones

For fixed $\pi \in S_n$

$$P(\sigma = \pi) = \int_{\mathbb{R}^n} \mathbb{1}_{\{X_{\pi(1)} > X_{\pi(2)} > \dots > X_{\pi(n)}\}} d\mu_X(x_1) \cdots d\mu_X(x_n)$$

$$= \int_{\mathbb{R}^n} \mathbb{1}_{\{x_1 > x_2 > \dots > x_n\}} d\mu_X(x_1) \cdots d\mu_X(x_n)$$

$\Rightarrow$ independent of π $\Rightarrow \frac{1}{n!}$ i.e., unif. distrib.

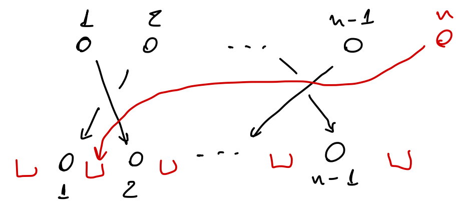
Generate σ by first picking $\sigma(1) \in \{1, \_, n\}$ uniformly.
Then $\sigma(2) \in \{1, \_, n\} \setminus \{\sigma(1)\}$ uniformly and so on.

Thus, $P(A_n) = P\big(\sigma(1) = n\big) = \frac{1}{n}$.

<u>Check independence of A_k's</u>: since σ is uniformly distrib. in S_n, σ^{-1} is also uniformly distributed (b/c $\sigma \mapsto \sigma^{-1}$ is a bijection).

$$\sigma^{-1}(j) = \text{ranking of } X_j \text{ amongst } X_1, X_2, \dots, X_n.$$

Set $\sigma^{-1} =: \pi$. Generate π by first picking a uniformly distrib. element of S_{n-1} and then insert $\pi(n)$ uniformly between the first $n-1$ elements. Then since



$\pi(1) = 3$, $\pi(2) = 1$, $\pi(n) = 2$

$\mathbb{P}(A_n) = \frac{1}{n}$, A_n independent of the values of the permutation on $(n-1)$ letters we used.

$\Longrightarrow$ The events $A_1, \dots, A_{n-1}$ are functions of the permutation on $(n-1)$ letters and are indep of A_n.

$\Longrightarrow$ $(A_1, A_2, \dots, A_n)$ are indep $\forall n$.

<u>Q</u>: How many records have been set?

<u>A</u>: $\displaystyle \lim_{n\to\infty} \frac{\sum_{j=1}^{n} \mathbb{1}_{A_j}}{\log(n)} = 1$ a.s. b/c $\displaystyle \sum_{j=1}^{n} \frac{1}{j} = \log n\,(1+o(1))$

Theorem from last class

$\blacktriangle$

* COMPUTING THE TAILS :

Def: (Tail σ-algebra) Let $X_1, X_2, \dots$ be a sequence of independent r.v.'s. Set

$$\mathcal{F}_n := \sigma(X_n, X_{n+1}, \dots).$$

Then, the tail σ-algebra is defined as

$$\mathcal{T} := \bigcap_{n=1}^{\infty} \mathcal{F}_n$$

Ex: $S_n := \sum_{i=1}^{n} X_i$. Then $\{\lim_{n\to\infty} S_n \text{ exists}\} \in \mathcal{T}$

value of finitely many terms in the series does not affect the existence of the lim

$$\{\lim_{n\to\infty} S_n \text{ exists}\} = \{\lim_{n\to\infty} (S_n - S_N) \text{ exists}\} \in \mathcal{F}_N \quad \forall N.$$

Ex: $\exists$ choices of $\{X_i\}$ such that $\{\limsup_{n\to\infty} S_n > 0\} \notin \mathcal{T}$.

Ex: If $c_n \in \mathbb{R}$, $c_n \to \infty$, then $\{\limsup_{n\to\infty} \frac{S_n}{c_n} > x\} \in \mathcal{T}$

Ex: (Borel–Cantelli) $A_n := \{X_n \in B_n\}$, B_n Borel set, then

$$\limsup_{n\to\infty} A_n \in \mathcal{T}.$$

Thm: (KOLMOGOROV 0-1 LAW) If $A \in \mathcal{T}$, then $P(A) = \{0,1\}$.

Pf: We claim that A is independent of itself b/c, in that case,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2 \Rightarrow \mathbb{P}(A) = \{0,1\}.$$

Cl: A indep. of itself.

- If $B \in \sigma(X_1, \ldots, X_n)$ and $C \in \sigma(X_{n+1}, X_{n+2}, \ldots)$. then B and C are independent.

Pf: If $C \in \sigma(X_{n+1}, X_{n+2}, \ldots, X_{n+m})$, then B and C are indep by definition.

$\Rightarrow$ B indep of any event in $\underbrace{\bigcup_{m=1}^{\infty} \sigma(X_{n+1}, \ldots, X_{n+m})}_{\pi\text{-system that contains } \Omega \text{ and generates } \sigma(X_{n+1}, X_{n+2}, \ldots)}$

λ-system := events in $\sigma(X_{n+1}, X_{n+2}, \ldots)$ that are indep. of B.

$\Rightarrow$ B and C indep. by π-λ thm. $\square$

- If $B \in \sigma(X_1, X_2, \ldots)$ and $C \in \mathcal{T}$, then B and C are independent.

Pf: $\sigma(X_1, X_2, \ldots)$ is generated by the π-system $\bigcup_{n \geq 1} \sigma(X_1, \ldots, X_n)$

So, by the previous claim, this π-system is independent of $\mathcal{T}$.

$\Rightarrow$ B and C indep. by π-λ thm. $\square$

Thus, A is independent of itself. $\blacksquare$

Cor: Series of indep. r.v.'s either converge or diverge.

---

* STRONG LAW OF LARGE NUMBERS :

STRONG LAW OF LARGE NUMBERS: Let $X_1, X_2, \ldots$ be iid r.v.'s such that $\mathbb{E}|X_i| < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}X_1 \quad a.s.$$

↓ Need some inequalities to prove this

Thm: (KOLMOGOROV'S MAXIMAL INEQUALITY) Let $X_1, X_2, \ldots$ be independent and centered ($\mathbb{E}X_i = 0$). Set

$$S_k = \sum_{j=1}^{k} X_j .$$

Then

$$\mathbb{P}\left(\max_{1\le k\le n} |S_k| > \varepsilon\right) \le \frac{1}{\varepsilon^2} \mathrm{Var}(S_n)$$

Pf: Let

$$A_k := \{ |S_k| > \varepsilon_k \text{ but } |S_j| \le \varepsilon \text{ for } 1 \le j \le k \}.$$

Note that $\{\max_{1\le j\le n} |S_j| > \varepsilon\} = \bigsqcup_{j=1}^{n} A_k$.

Moreover,

$$\mathrm{Var}(S_n) = \mathbb{E}S_n^2 \geq \mathbb{E}\left(S_n^2 \mathbb{1}\{\bigcup_{j=n}^{\infty} A_j\}\right) \overset{A_k\text{'s are disjoint}}{=} \sum_{j=1}^{n} \mathbb{E}(S_n^2 \mathbb{1}_{A_j})$$

$$= \sum_{j=1}^{n} \mathbb{E}\left[(S_n + S_j - S_j)^2 \mathbb{1}_{A_j}\right]$$

$$= \sum_{j=1}^{n} \mathbb{E}S_j^2 \mathbb{1}_{A_j} + \underbrace{\mathbb{E}(S_n - S_j)^2 \mathbb{1}_{A_j}}_{\geq 0}$$

$$+ 2\cancel{\mathbb{E}S_j(S_n - S_j)\mathbb{1}_{A_j}}_{=0}$$

b/c $\mathbb{E}S_j(S_n - S_j) = 0$ since $(S_n - S_j)$ indep. of $S_j \mathbb{1}_{A_j}$

$\Rightarrow \; = (\mathbb{E}(S_n - S_j))\, \mathbb{E}(S_j \mathbb{1}_{A_j})$

$$\geq \sum_{j=1}^{n} \mathbb{E}S_j^2 \mathbb{1}_{A_j} \geq \varepsilon^2 \sum_{j=1}^{n} \mathbb{P}(A_j)$$

$$= \varepsilon^2 \mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > \varepsilon\right).$$

■

Lemma: Let $X_1, X_2, \ldots$ be indep r.v.'s s.t. $\sum_{i=1}^{\infty} \mathrm{Var}(X_i) < \infty$

(in particular, X_i's can't be iid b/c then $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \ldots \Rightarrow \sum \mathrm{Var} = \infty$)

then $\sum_{j=1}^{\infty} (X_j - \mathbb{E}X_j)$ converges a.s.

Pf: Assume $\mathbb{E}X_i = 0 \ \forall i$. Suffices to show that the partial sums

$$S_n := \sum_{j=1}^{n} X_j$$

form a Cauchy sequence. For fixed n, estimate

$$\mathbb{P}\Big(\sup_{m \geq n} |S_m - S_n| \geq \varepsilon\Big) = \lim_{N\to\infty} \mathbb{P}\Big(\sup_{N \geq m \geq n} |S_m - S_n| \geq \varepsilon\Big)$$

Kolmogorov max ineq →

$$\leq \lim_{N\to\infty} \frac{1}{\varepsilon^2} \sum_{j=n+1}^{N} \mathrm{Var}(X_j)$$

$$= \frac{1}{\varepsilon^2} \sum_{j=n+1}^{\infty} \mathrm{Var}(X_j) \xrightarrow[n\nearrow\infty]{a.s.} 0$$

Let $F_n := \{\sup_{m \geq n} |S_m - S_n| \geq \varepsilon\}$. Set n_k s.t. $\mathbb{P}(F_{n_k}) \leq \frac{1}{k^2}$

Can do this b/c we checked above that these probs $\to 0$

By Borel-Cantelli I,

$$\mathbb{P}(F_{n_k} \text{ i.o.}) = 0.$$

$\Longrightarrow S_n$ is Cauchy a.s.

■

Lemma: (Kronecker's Lemma)

Pf: (SLLN)

LECTURE 10

Oct 9th, 2024

STRONG LAW: X_i iid and $\mathbb{E}|X_i| < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}X_1 .$$

Pf: $Y_k := X_k \mathbb{1}_{\{|X_k| \leq k\}}$ (truncate)

(i) $\mathbb{P}(Y_k \neq X_k \text{ i.o.}) = 0$ by Borel-Cantelli

(ii) $\frac{1}{n}\sum_{k=1}^{n} Y_k - \mathbb{E}X_i$ by showing $\sum_j \frac{Var(Y_j)}{j} < \infty$

Maximal Ineq. $\Longrightarrow \sum_j \frac{Y_j - \mathbb{E}Y_j}{j} \overset{\text{Kronecker}}{\Longrightarrow} \frac{1}{n}\sum_j Y_j - \mathbb{E}Y_j$

$\square$

Thm: (KOLMOGOROV'S 3-SERIES) Let $X_1, X_2, \ldots$ be independent and fix $A > 0$ then $\sum_{j=1}^{\infty} X_j$ converges a.s. iff

(i) $\sum_{j=1}^{\infty} \mathbb{P}(|X_j| > A) < \infty$

(ii) Let $Y_j := X_j \mathbb{1}_{\{|X_j| \leq A\}}$, $\sum_{j=1}^{\infty} \mathbb{E}Y_j$ converges

(iii) $\sum_{j=1}^{\infty} \mathrm{Var}(Y_j) < \infty$.

If (i), (ii), (iii) hold for any $A>0$, then it holds $\forall A>0$.

Pf: $(\Leftarrow)$ (i) implies that $\sum_{k=1}^{\infty} X_k$ converges $\Leftrightarrow \sum_{k=1}^{\infty} Y_k < \infty$

(ii), (iii) imply $\sum Y_j = \left(\sum_j Y_j - \mathbb{E} Y_j\right) + \sum_j \mathbb{E} Y_j$ converges

$\square$

FACT: Already checked that

$$\mathbb{E}|X_i| = \infty \Rightarrow \mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \text{ exists in } (-\infty,\infty)\right) = 0$$

Thm: If X_i's iid s.t. $\mathbb{E}(X_i)_+ = 0$ and $\mathbb{E}(X_i)_- < \infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \infty$$

If use $\mathbb{1}\{X_n \le n\}$ the Y_k's are not iid

Pf: Truncate: $Y_k = X_k \mathbb{1}\{X_k \le M\}$ for some $M>0$. Then we have $\{Y_k\}$ is iid and we can use the strong law:

$$\liminf_{n\to\infty} \frac{1}{n}\sum_k X_k \geq \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Y_k = \mathbb{E} Y_1 = \mathbb{E} X_1 \mathbb{1}\{X_1 \le M\}$$

Take $M \to \infty$ then by MCT $\mathbb{E} X_1 \mathbb{1}_{\{X_1 \le M\}} \to \mathbb{E} X_1 = \infty$.
assumption

Thus, $\liminf\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k} X_k \ge \infty$. ∎

IDEA: TRUNCATE SOMEWHERE & USE STRONG LAW

Application of Strong Law

Ex: (RENEWAL THEORY) Take iid $X_i \in [0, \infty]$. Define

$$T_n := \sum_{i=1}^{n} X_i$$

(= if X_i are lifetimes of lightbulbs, T_n is how long it takes to go through n lightbulbs)

Define $N_t := \sup\{n : T_n \le t\}$ (N_t = # of bulbs that burned out before time t)

Let $\mu = \mathbb{E} X_1 \in (0, \infty]$ (mean lifetime)

By the Strong Law, $\frac{T_n}{n} \longrightarrow \mu$ a.s.

Cl: $\frac{N_t}{t} \xrightarrow{a.s.} \frac{1}{\mu}$ as $t \uparrow \infty$.

Pf: Note $T_{N_t} \le t < T_{N_t + 1}$ by definitions. So,

$$\frac{T_{N_t}}{N_t} \le \frac{t}{N_t} < \frac{T_{N_t+1}}{N_t+1} \frac{N_t+1}{N_t}.$$

Since $T_n < \infty \ \forall n$, we have that $N_t \to \infty$ as $t \to \infty$

Thus, $\lim\limits_{t\to\infty} \frac{T_{N_t}}{N_t} = \lim\limits_{n\to\infty} \frac{T_n}{n} = \mu$ and

$\lim_{t\to\infty} \frac{N_t + 1}{N_t} = 1$. So, by the Squeeze theorem we're done. ◢

FACT: If $X_n \to X$ a.s. and $N_t \to \infty$ as $t \to \infty$, then $X_{N_t} \to X$ a.s. even if the N_t are random.

Exercise: This is false if we replace "a.s." to "in probability".

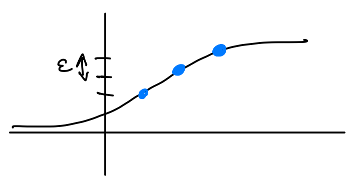
⚠ IMPORTANT (application of Strong Law)

Ex: Let X_i be iid w/ cdf $F(t) := \mathbb{P}(X_i \le t)$.
Define the **EMPIRICAL CDF** by

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}}$$

Cl: $\lim_{n\to\infty} \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| = 0$ a.s.. (uniform convergence)

Pf: If $F(t)$ was continuous, then we would be able to go back & forth.



Without the supremum, it's just the Strong Law b/c the indicators are a bunch of Bernoulli r.v.'s.
Now, with the supremum,

WTS: $\limsup\limits_{n\to\infty} \sup\limits_{t\in\mathbb{R}} |F_n(t) - F(t)| = 0$ a.s.

Since F(t) is monotone, $\exists$ only finitely many $x_1, \ldots, x_n$ s.t. $F(x_i^+) - F(x_i^-) > \frac{\varepsilon}{2}$.

Let $\varepsilon > 0$ and let $X_1, \ldots, X_n$ be finitely many pts. at which $F(x_i^+) - F(x_i^-) > \varepsilon/2$

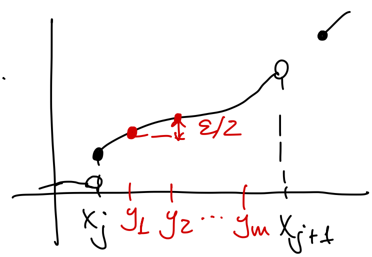
$\Rightarrow$ F is continuous between (x_j, x_{j+1})

Have to show $\limsup\limits_{n\to\infty} \sup\limits_{t\in[x_j, x_{j+1}]} |F_n(t) - F(t)| = 0$ a.s.

Also need to check $(-\infty, x_1]$ and $[x_n, \infty)$ but these are similar to how we prove the $[x_j, x_{j+1}]$'s.

$y_1 < y_2 < \cdots < y_p$

For any interval $[x_j, x_{j+1}]$, let $y_1, \ldots, y_p$ be finitely many pts. s.t. $F(y_{j+1}) \leq F(y_j) + \varepsilon/2$ and $F(y_1) < F(x_j) - \frac{\varepsilon}{2}$ and $F(x_{j+1}) < F(y_p) + \frac{\varepsilon}{2}$.



Define $G_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_j < t\}}$,

$G(t) := \mathbb{P}(X_j < t)$

strictly less $\neq F_n$

By the Strong Law, $\exists N(\omega)$ s.t. $n \geq N$,

$|F_n(t) - F(t)| + |G_n(t) + G(t)| < \varepsilon$

for all $t \in \{x_j, x_{j+1}, y_1, \ldots, y_p\}$.

For $t \in (y_k, y_{k+1})$ and $n \geq N$,

$$F_n(t) \leq G_n(y_{k+1}) \leq G(y_{k+1}) + \varepsilon \leq F(t) + 2\varepsilon \quad (*)$$

and

$$F_n(t) \geq F_n(y_k) \geq F(y_k) - \varepsilon \geq F(t) - 2\varepsilon .$$

Namely, we need G_n b/c F might jump on the rightmost point.

(*) Used G_n b/c (*) also works for $k = p$ and $y_{p+1} = x_{j+1}$

LECTURE 11

Oct 11, 2024

- EXAMPLE OF 0-1 LAW: (HEWITT - SAVAGE 0-1 LAW)

If X_i are iid r.v.'s, say $P(X_i = 1) = p = 1 - P(X_i = -1)$.

Let

$$S_n = \sum_{j=1}^{n} X_j . \quad \text{Then } \{\limsup_{n\to\infty} S_n = +\infty\} \in \mathcal{T}$$

$$\{S_n = 0 \text{ i.o.}\} \notin \mathcal{T} .$$

This event is invariant under shuffling the j's

Def: A set $B \in \mathbb{R}^{\mathbb{N}}$ (i.e., ∞ seqs. of real #'s) is SYMMETRIC if for all n and all permutations $\pi \in S_n$ we have

$$(X_1, X_2, X_3, \dots) \in B \implies (X_{\pi(1)}, X_{\pi(2)}, X_{\pi(3)}, \dots) \in B .$$

An event $A \in \sigma(X_1, X_2, \dots)$ is symmetric if $A = \{X \in B\}$ with $B \in \mathbb{R}^{\mathbb{N}}$ symmetric and $X := (X_1, X_2, \dots)$.

> **Thm**: (Hewitt–Savage 0-1 Law) If $X_1, X_2, \dots$ are iid and A is symmetric, then $P(A) = \{0, 1\}$.

FACT: For any event $A \in \sigma(X_1, X_2, \dots)$ and $\forall \varepsilon > 0$, $\exists A_n \in \sigma(X_1, \dots, X_n)$ s.t. $P(A \triangle A_n) \le \varepsilon$.

$\leftarrow B \triangle C = B \cup C \setminus B \cap C$.

Pf: Define $A_n := \{(X_1, \dots, X_n) \in B_n\}$ for some $B_n \subset \mathbb{R}^n$.

Set $\hat{A}_n := \{(X_{n+1}, \dots, X_{2n}) \in B_n\}$.

Cl: $P(A_n \triangle \hat{A}_n) \le \varepsilon$.

Pf: Let $\Gamma : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ by

Bijection $\Gamma(X_1, X_2, \dots) := (X_{n+1}, X_{n+2}, \dots, X_{2n}, X_1, \dots, X_n, X_{2n+1}, \dots)$

Then

$$P(A_n \triangle \hat{A}_n) = P(\{(X_{n+1}, \dots, X_{2n}) \in B_n\} \triangle \{X \in B\})$$

Γ is bijective
$$= P(\{(X_{n+1}, \dots, X_{2n}) \in B_n\} \triangle \{\Gamma X \in \Gamma B\})$$

B symmetric $\stackrel{\downarrow}{=} \mathbb{P}(\{(X_{n+1},\ldots,X_{2n})\in B_n\} \Delta \{\Gamma X\in B\})$

$\left[\mathbb{P}(\Gamma X\in C) \overset{\text{cid}}{\Downarrow}{=} \mathbb{P}(X\in C) \right]$ $\stackrel{\downarrow}{=} \mathbb{P}(\{(X_1,\ldots,X_n)\in B_n\} \Delta \{X\in B\})$

$= \mathbb{P}(A_n \Delta A) \le \varepsilon$

Fact above

$\square$

Thm:

In general:

$|\mathbb{P}(B) - \mathbb{P}(C)| \le \mathbb{P}(B\Delta C)$.

$$
\begin{aligned}
|\mathbb{P}(A) - \mathbb{P}(A_n \cap \hat{A}_n)| &\le \mathbb{P}(A \Delta (A_n \cap \hat{A}_n)) \\
&\le \mathbb{P}(A\Delta A_n) + \mathbb{P}(A \Delta \hat{A}_n) \\
&\le 2\varepsilon
\end{aligned}
$$

On the other hand, indep b/c they depend on diff. X's.

$$
\mathbb{P}(A_n \cap \hat{A}_n) = \mathbb{P}(A_n)\,\mathbb{P}(\hat{A}_n) = (\mathbb{P}(A) + O(\varepsilon))(\mathbb{P}(A) + O(\varepsilon))
$$

$\Rightarrow \mathbb{P}(A) = \mathbb{P}(A)^2 + O(\varepsilon)$ now take $\varepsilon \to 0$.

$\blacksquare$

Fact used in the proof is a conseq. of this general fact:

Proposition: If $\mathcal{A}$ is an algebra and $\mathcal{B} = \sigma(\mathcal{A})$ then $\forall B \in \mathcal{B}$ there exists $A_n \in \mathcal{A}$ s.t. $\lim_{n\to\infty} \mathbb{P}(A_n \Delta B) = 0$.

Pf: Let

$$D = \{ B \in \mathcal{B} \text{ s.t. } \exists A_n \in \mathcal{A} \text{ s.t. } \lim_{n\to\infty} \mathbb{P}(A_n \triangle B) = 0 \}$$

Cl: D is a σ-algebra ← since $\mathcal{B}$ is the smallest σ-alg. containing $\mathcal{A}$, this means $D = \mathcal{B}$.

(i) $\Omega \in D$

(ii) $B \in D \Rightarrow B^c \in D$. ✓ not hard

(iii) First check it's closed under finite unions. Let $E = \bigcup_{i=1}^{n} E_i \in D$, then $\forall E_i \; \exists C_j^i \in \mathcal{A}$ s.t.

$$\lim_{j\to\infty} (C_j^i \triangle E_i) = 0.$$

So, $P(A \cup B \triangle C \cup E) \leq \mathbb{P}(A \triangle C) + \mathbb{P}(B \triangle E)$

$\Longrightarrow C_j = \bigcup_{i=1}^{n} C_j^i$ satisfies $\lim_{j\to\infty} \mathbb{P}(E \triangle C_j) = 0$

$\Longrightarrow E \in D$.

Check it's closed under countable unions. Let $E = \bigcup_{i\in\mathbb{N}} E_i$, $E_i \in D$. Then $\forall \varepsilon > 0 \; \exists n > 0$ s.t. $\mathbb{P}(E \setminus \bigcup_{j=1}^{n} E_j) \leq \varepsilon$ (by measure continuity)

$\Longrightarrow E \in D$.

∎

Ex: X_i iid and $P(X_i = 1) = p = 1 - P(X_i = -1)$.

Set $S_n := \sum_{i=1}^n X_i$. By the 0-1 law

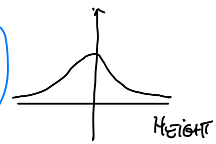
$$P(\underbrace{S_n = 0 \text{ i.o.}}_{\text{this event is symmetric}}) \in \{0,1\}$$

If $p = \frac{1}{2}$, expect particle to return to zero (i.e., $P=1$)

If $p \neq \frac{1}{2}$, by the strong law $\frac{S_n}{n} \longrightarrow (2p-1)$ a.s. ◢

---

* **CENTRAL LIMIT THEOREM** (Add a bunch of r.v.'s and we should get smth that looks like a Gaussian)



"CLT": If $X_1, \dots, X_n$ are centered and independent and not too different from each other in size (i.e., $EX_i^2 = 1$) + some other assumptions on X_i's

then $$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \approx N(0,1)$$ weak convergence (i.e., in distribution)

sqrt has to do with the variance of X_i's so that $\sum X_i$ has fluctuations of order 1

* **WEAK CONVERGENCE**: Let (S,d) be a metric space with the Borel σ-algebra $\mathcal{B}$.

Def: Given $\{P_n\}$ and P probability measures on S, we say $P_n \longrightarrow P_n$ weakly if $\forall f \in C_b(S)$, we have

continuous & bounded

$$\int f(\omega)\, d\mathbb{P}_n(\omega) \xrightarrow{n\nearrow\infty} \int f(\omega)\, d\mathbb{P}$$

Def: For a sequence X_n of r.v.'s we say $X_n \to X$ weakly iff $\mu_{X_n} \longrightarrow \mu_X$ weakly

$$\Longleftrightarrow \mathbb{E} f(X_n) \longrightarrow \mathbb{E} f(X) \quad \forall f \in C_b(S).$$

Rmk: Weak convergence does not require the r.v.'s to be in the same probability space (unlike $\xrightarrow{\mathbb{P}}$)

NOTATION: Weak convergence is called CONVERGENCE IN DISTRIBUTION

$$\boxed{\textbf{Thm}: \text{Let } S=\mathbb{R}. \text{ Then } \mathbb{P}_n \xrightarrow{d} \mathbb{P} \text{ weakly iff } \forall t \in \mathbb{R} \text{ s.t. the cdf } F(t) \text{ of } \mathbb{P} \text{ is continuous we have } F_n(t) \to F(t). \text{ Here } F_n(t) = \mathbb{P}_n((-\infty,t]),\ F(t) = \mathbb{P}((-\infty,t]).}$$

Pf: $(\Rightarrow)$

LECTURE 12

Oct 18^{th}, 2024

WEAK CONVERGENCE: If $\mathbb{P}_n$ are prob. measures on (S,d), then $\mathbb{P}_n \longrightarrow \mathbb{P}$ weakly if $\forall f \in C_b(S)$,

$$\int f(x)\, d\mathbb{P}_n(x) \longrightarrow \int f(x)\, d\mathbb{P}(x) \quad \text{as } n \to \infty.$$

Remark: If $S = \mathbb{R}$, then weak convergence is equivalent to $F_n(t) \to F(t)$ $\forall t$ s.t. F is continuous at t, where

$$F_n(t) = \mathbb{P}_n((-\infty, t]), \quad F(t) = \mathbb{P}((-\infty,t]).$$

Rmk: If X_n are r.v.'s, then $X_n \longrightarrow X$ weakly if $\mu_{X_n} \to \mu_X$ weakly; which is equivalent to

law of X_n

$$\mathbb{P}(X_n \le t) \longrightarrow \mathbb{P}(X \le t)$$

whenever $t \mapsto \mathbb{P}(X \le t)$ is continuous.

Ex: If X is a r.v., then $X_n := X + \frac{1}{n}$ converges weakly to X since

$$\mathbb{P}(X_n \le t) = \mathbb{P}(X_n \le t - \tfrac{1}{n}) = \underbrace{F(t - \tfrac{1}{n})}_{\longrightarrow F(t) \text{ at continuity pts. of } F}$$

But, $X = 0$ and $X_n = \frac{1}{n}$. Then $\mathbb{P}(X_n \le 0) = 0$ and $\mathbb{P}(X \le 0) = 1$

Def: We say $\mathbb{P}_n$ on (S, d) is <u>tight</u> if $\forall \varepsilon > 0$ $\exists K$ compact s.t. $\mathbb{P}_n(K) \ge 1 - \varepsilon$ $\forall n$.

Ex: $S = \mathbb{R}$, the laws of $X_n = n$ are <u>not</u> tight.

Lemma: If $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, then the $\mathbb{P}_n$'s are tight. ⚠

Thm: If $\mathbb{P}_n$ are probability measures on (S, d) that are tight, then $\exists$ subsequence $\mathbb{P}_{n_k}$ that converges to some $\mathbb{P}$.

Pf: ($S = \mathbb{R}$) Let $F_n(t)$ be the cdf of $\mathbb{P}_n$. By Cantor's diagonalization argument, $\exists n_k$ s.t. $F_{n_k}(q) \to G(q)$ $\forall q \in \mathbb{Q}$ ...... HELLY SELECTION CRITERION $\square$

Thm: If $\mathbb{P}_n$ and $\mathbb{P}$ have the property that any subseq. of $\mathbb{P}_n$ has a further subseq. that converges weakly to $\mathbb{P}$, then $\mathbb{P}_n \to \mathbb{P}$ weakly.

IMPORTANT STRATEGY: to show $\mathbb{P}_n \longrightarrow \mathbb{P}$ weakly,

- 1^{st}: show $\mathbb{P}_n$ is tight
- 2^{nd}: apply the tight $\Rightarrow \exists$ subseq $\xrightarrow{\text{weakly}} \mathbb{P}$
- 3^{rd}: use the above to conclude that $\mathbb{P}_n \xrightarrow{\text{weakly}} \mathbb{P}$

Pf: If $\mathbb{P}_n$ does not converge to $\mathbb{P}$, then $\exists f \in C_b(S)$ s.t.

$$\int f(x)\, \mathbb{P}_n(x) \not\longrightarrow \int f(x)\, d\mathbb{P}(x).$$

$\Rightarrow \exists$ subseq. n_k and $\varepsilon > 0$ s.t.

$$\left| \int f(x)\, d\mathbb{P}_{n_k}(x) - \int f(x)\, d\mathbb{P} \right| \geq \varepsilon$$

$\Rightarrow$ By hypothesis, $\exists n_{k_j}$ s.t. $\mathbb{P}_{n_{k_j}} \to \mathbb{P}$ weakly $\rightarrow\!\!/\!/\!\!\leftarrow$

■

Ex: Let $\{U_i\}_{i=1}^{n}$ be iid uniform $\text{Unif}(0,1)$ r.v.'s.

Set $X_n := \max_{1 \leq i \leq n} U_i$.

Cl: $n(1 - X_n) \xrightarrow{\text{weakly}} \text{Exp}(1)$ as $n \to \infty$.

Pf: Compute $\mathbb{P}(X_n \leq t)$. Note that

$$\mathbb{P}(X_n \leq t) = \mathbb{P}(U_i < t \quad \forall i = 1, \dots, n)$$

intersection of a bunch of indep events

$$= \prod_{i=1}^{n} \mathbb{P}(U_i < t)$$

$U_i \sim \mathrm{Unif}(0,1)$

$$= t^n \quad \text{if } t \in (0,1).$$

So,

$$\mathbb{P}\left(X_n \le 1 - \frac{s}{n}\right) = \left(1 - \frac{s}{n}\right)^n \xrightarrow{n \nearrow \infty} e^{-s}.$$

LECTURE 13

Oct 18th, 2024

$$\mathbb{P}_n \to \mathbb{P} \text{ weakly} \iff \int f(x)\, d\mathbb{P}_n(x) \to \int f(x)\, d\mathbb{P}(x) \quad \forall f \in C_b(S)$$

$$\iff F_n(t) \to F(t) \ \forall t \text{ at which } F \text{ is continuous}$$

Thm: (PORTMONTEAU) TFAE

(1) $\mathbb{P}_n \to \mathbb{P}$ weakly

(2) $\int f\, d\mathbb{P}_n \to \int f\, d\mathbb{P}$ $\forall$ Lipschitz and bounded f

(f is Lipschitz iff $\exists C > 0$ s.t. $|f(x) - f(y)| \le C\, d(x,y)$)

(3) $\forall$ lower semicontinuous f bounded from below,

$$\liminf_{n\to\infty} \int f\, d\mathbb{P}_n \ge \int f\, d\mathbb{P}.$$

(f is lower semicont. iff $\forall x_n,\ x_n \to x_0$ for $x_0 \in S$ we have $\liminf_{n\to\infty} f(x_n) \ge f(x_0)$)

(4) $\forall$ upper semicontinuous f bounded from above,

$$\limsup_{n\to\infty} \int f \, d\mathbb{P}_n \le \int f \, d\mathbb{P}$$

(5) $\forall$ open sets $\mathcal{O} \subset S$, $\liminf_{n\to\infty} \mathbb{P}_n(\mathcal{O}) \ge \mathbb{P}(\mathcal{O})$

(6) $\forall$ closed sets $C \subset S$, $\limsup_{n\to\infty} \mathbb{P}_n(C) \le \mathbb{P}(C)$

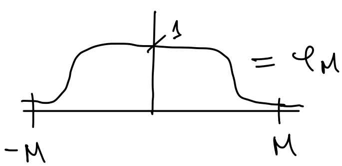
(7) $\forall$ continuity sets A of $\mathbb{P}$, $\lim_{n\to\infty} \mathbb{P}_n(A) = \mathbb{P}(A)$

(A is a continuity set of $\mathbb{P} \iff \mathbb{P}(\partial A) = 0$)

Rmk: If $S = \mathbb{R}$, we can replace (2) by $C_0^\infty(\mathbb{R})$ fcts.

Pf: $(1) \Rightarrow (2)$ ✓

$(2) \Rightarrow (1)$ Deduce that $\mathbb{P}_n$ is tight by using test functions



Then,

$$\int f(x) \, d\mathbb{P}_n = \int f(x) \varphi_M(x) \, d\mathbb{P}_n(x) + \int \underbrace{f(x)(1-\varphi_M(x))}_{\le C\varepsilon} \, d\mathbb{P}_n(x)$$

Since $\varphi_M f$ is compactly supported, we can approx. it in L^∞ by smooth functions

$(3) \iff (4)$ by sending $f \longmapsto -f \rightsquigarrow \liminf \longmapsto \limsup$.

$(3) \iff (4) \Rightarrow (1)$

$(1) \Rightarrow (3)$ Thm (Baire) If f is lsc $\exists$ bdd & cont. seq. f_n s.t. $f_n \xrightarrow{\text{ptwise}} f$.

Thus, $\liminf_{n\to\infty} \int f \, d\mathbb{P}_n \geq \liminf_{n\to\infty} \int f_m \, d\mathbb{P}_n = \int f_m \, d\mathbb{P}$

$$\int f_m \, d\mathbb{P} \xrightarrow[\text{by MCT}]{\text{as } m\to\infty} \int f \, d\mathbb{P}$$

$(5) \Leftrightarrow (6)$ by taking complements & noting that

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$$

$(4) \Rightarrow (6)$ $\mathbb{1}_C$ is upper semicontinuous b/c C is closed.

$(5)+(6) \Rightarrow (7)$ $A^\circ \subset A \subset \overline{A}$. Since A is a continuity set of $\mathbb{P}$, then $\mathbb{P}(A^\circ) = \mathbb{P}(A) = \mathbb{P}(\overline{A})$. So,

$$\limsup_{n\to\infty} \mathbb{P}_n(A) \leq \limsup_{n\to\infty} \mathbb{P}_n(\overline{A}) \leq \mathbb{P}(\overline{A}) = \mathbb{P}(A).$$

$(7) \Rightarrow (1)$ Assume $S = \mathbb{R}$

If F is continuous at a point $t \in \mathbb{R}$, then $(-\infty, t]$ is a continuity set for $\mathbb{P}$. Thus $F_n(t) \longrightarrow F(t)$.

$\blacksquare$

Upshot: only useful part is $\mathbb{P}_n \to \mathbb{P} \Leftrightarrow \int f \, d\mathbb{P}_n \to \int f \, d\mathbb{P}$ for f Lipschitz ?

Thm: (CENTRAL LIMIT THEOREM) If $\{X_i\}_{i=1}^n$ is a sequence of iid r.v.'s s.t. $\mathbb{E}X_i = 0$ and $\mathbb{E}X_i^2 = 1$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \longrightarrow N(0,1) \text{ weakly}$$

($Z \sim N(0,1)$ if it's distrib. is: $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$)

sqrt to ensure the variance remains 1

Note: • CLT is trivial if $X_i \sim N(0,1)$ b/c sums of $N(0,1)$ are again $N(0,1)$. So it's an equality rather than "$\longrightarrow$".

Pf: (Lindeberg) First, assume $\mathbb{E}|X_i|^3 \le C$. Let Z_i be iid $N(0,1)$ independent from X_i's. Suffices to show that

$$\lim_{n\to\infty} \left| \mathbb{E}\left(f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right)\right) - \mathbb{E}f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\right)\right| = 0$$

$\forall$ smooth f w/ bdd derivatives of all orders (by the first equiv. in Portmanteau).

Let $S := \frac{1}{\sqrt{n}} \sum_{i=2}^n X_i$.

Cl: $\left| \mathbb{E}f\left(\frac{X_1}{\sqrt{n}} + S\right) + \mathbb{E}f\left(\frac{Z_1}{\sqrt{n}} + S\right)\right| \le C n^{-3/2}$

Note: $f\left(\frac{X_1}{\sqrt{n}} + S\right) \overset{\text{Taylor series}}{=} f(S) + \frac{X_1}{\sqrt{n}} f'(S) + \frac{X_1^2}{2n} f''(S) + O\left(\frac{|X_1|^3}{n^{3/2}}\right)$

$$f\left(\frac{Z_1}{\sqrt{n}}+S\right) = f(S) + \frac{Z_1}{\sqrt{n}} f'(S) + \frac{Z_1^2}{2n} f''(S) + O\left(\frac{|Z_1|^3}{n^{3/2}}\right).$$

Thus,

$$\left| \mathbb{E} f\left(\frac{X_1}{\sqrt{n}}+S\right) + \mathbb{E} f\left(\frac{Z_1}{\sqrt{n}}+S\right)\right|$$

$$= \left| \mathbb{E}\left(\cancel{f(S) - f(S)}\right)^{=0} + \mathbb{E}\left[\left(\frac{X_1}{\sqrt{n}} - \frac{Z_1}{\sqrt{n}}\right) f'(S)\right] + \mathbb{E}\left[\left(\frac{X_1^2}{\sqrt{n}} - \frac{Z_1^2}{\sqrt{n}}\right)\frac{f''(S)}{2}\right]\right| + O\left(\frac{\mathbb{E}|X_1|^3}{n^{3/2}} + \frac{\mathbb{E}|Z_1|^3}{n^{3/2}}\right)$$

$\leq C n^{-3/2}$.

But

X_i iid and for a fct of the X's

$$\mathbb{E}\left(\frac{X_1}{\sqrt{n}} f'(S)\right) = \left(\cancel{\mathbb{E}\frac{X_1}{\sqrt{n}}}\right)^{=0 \text{ b/c } X_1 \text{ centered}} \left(\mathbb{E} f'(S)\right)$$

$$\mathbb{E}\left(X_1^2 f''(S)\right) = \underbrace{\mathbb{E}X_1^2}_{=1} \mathbb{E} f''(S) = \mathbb{E} f''(S).$$

Now, just replace the first two X's in S:

$$\hat{S} = \frac{Z_1}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{i=3}^{n} X_i.$$

Now, dropping the assumption on the 3rd moment, replace the Taylor expansion argument as follows:

- Pick $M>0$. If $|X_1| \leq M$, then we do the following expansion

$$f\left(\frac{X_1}{\sqrt{n}}+S\right)\mathbb{1}_{\{|X_1|\le M\}} = \left(f(S)+\frac{f'(S)}{\sqrt{n}}X_1+\frac{f''(S)X_1^2}{2n} + O\left(\frac{|X_1|^3}{n^{3/2}}\right)\right)\mathbb{1}_{\{|X_1|\le M\}}$$

- If $|X_1|>M$,

$$f\left(\frac{X_1}{\sqrt{n}}+S\right)\mathbb{1}_{\{|X_1|>M\}} = \left(f(S)+\frac{X_1 f'(S)}{\sqrt{n}}+O\left(\frac{|X_1|^2}{n}\right)\right)\mathbb{1}_{\{|X_1|>M\}}$$

Add the two expansions:

$$f\left(\frac{X_1}{\sqrt{n}}+S\right) = f(S)+\frac{X_1}{\sqrt{n}}f'(S)+\frac{X_1^2}{\sqrt{n}}\frac{f''(S)}{2}\left(1-\cancel{\mathbb{1}_{\{|X_1|>M\}}}\right) + O\left(\frac{|X_1|^2}{n}\mathbb{1}_{\{|X_1|>M\}}+\frac{M^3}{n^{3/2}}\right)$$

absorb this into the error

Thus

$$\left|\mathbb{E}f\left(\frac{X_1}{\sqrt{n}}+S\right)-\mathbb{E}f\left(\frac{Z_1}{\sqrt{n}}+S\right)\right| \le C\left(\frac{M^3}{n^{3/2}}+\frac{\mathbb{E}\left(|X_1|^2\mathbb{1}_{\{|X_1|>M\}}\right)}{n}\right)$$

Doing this n times:

$$\left|\mathbb{E}f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right)-\mathbb{E}f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\right)\right|$$

$$(**) \le \frac{CM^3}{n^{1/2}}+C\,\mathbb{E}\left[|X_1|^2\mathbb{1}_{\{|X_1|>M\}}\right]$$

$$\limsup_{n\to\infty} (**) \le C\,\mathbb{E}\left[X_1^2 \mathbb{1}_{\{|X_1|>M\}}\right] \longrightarrow 0 \text{ as } M\nearrow\infty \text{ by DCT.}$$

$\blacksquare$

APPLICATIONS OF CLT

Ex: (Normal approx. to Poisson distrib.)

$Z_\lambda \sim \text{Poiss}(\lambda)$, $\mathbb{P}(Z_\lambda = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$ $\forall k \in \mathbb{Z}$, $k \geq 1$.

$Z_{\lambda_1}, Z_{\lambda_2}$ indep Poisson $\Rightarrow Z_1 + Z_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Claim: $\dfrac{Z_\lambda - \lambda}{\sqrt{\lambda}} \longrightarrow N(0,1)$ weakly as $\lambda \to \infty$.

If $\lambda = n$, then $Z_n \overset{d}{=} X_1 + \dots + X_n$, $X_i \sim \text{Poisson}(1)$

If $\lambda = n$, suffices to show

$$\left.\frac{\sum_{i=1}^{n} X_i - n}{\sqrt{n}} \xrightarrow{\text{weakly}} N(0,1).\right] \text{CLT}$$

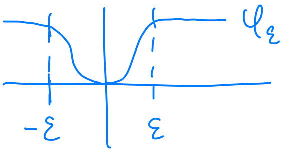
For general $\lambda \in \mathbb{R}$, $Z_\lambda = Z_{\lfloor\lambda\rfloor} + Z_{\lambda - \lfloor\lambda\rfloor}$. So,

$$\frac{Z_\lambda - \lambda}{\sqrt{\lambda}} \overset{d}{=} \underbrace{\frac{Z_{\lfloor\lambda\rfloor} - \lambda}{\sqrt{\lambda}}}_{\substack{\downarrow \\ N(0,1) \text{ weakly as} \\ \lambda\nearrow\infty \text{ by the CLT}}} + \underbrace{\frac{Z_{\lambda - \lfloor\lambda\rfloor}}{\sqrt{\lambda}}}_{\substack{\downarrow \\ 0 \text{ weakly as } \lambda\to\infty}}$$

FACT: If $X_n \to X$ weakly, $Y_n \to 0$ weakly then $X_n + Y_n \to X$ weakly. ← False if we substitute $0 \mapsto Y$ b/c then $X+Y$ does not make sense in general...

<u>Pf of fact:</u> Since $Y_n \xrightarrow{\text{weakly}} 0$, it actually $Y_n \xrightarrow{\text{Prob.}} 0$ i.e., $\forall \varepsilon > 0$, $\lim_{n\to\infty} P(|Y_n| \geq \varepsilon) = 0$

$$\mathbb{P}(|Y_n| \geq \varepsilon) \leq \mathbb{E}\varphi_\varepsilon(Y_n) \longrightarrow \mathbb{E}(\varphi_\varepsilon(0)) = 0.$$



Let f be Lipschitz, then

$$\left| \mathbb{E}[f(X_n + Y_n)] - \mathbb{E} f(X_n) \right| \leq C\varepsilon + C\,\mathbb{P}(|Y_n| \geq \varepsilon) \quad \forall \varepsilon > 0.$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{E} f(X_n + Y_n) = \lim_{n\to\infty} \mathbb{E} f(X_n) = \mathbb{E} f(X).$$

Lemma: If $X_n \to X$ weakly, then $\exists$ probability space and r.v.'s $\hat{X}_n$, $\hat{X}$ s.t. $\hat{X}_n \overset{d}{=} X_n$ and $\hat{X} \overset{d}{=} X$ and $\hat{X}_n \xrightarrow{a.s} \hat{X}$.

Lemma: Let $F_n(t)$ and $F(t)$ be the cdfs of X_n and X.

Define
$$G_n(x) := \sup\{y : F_n(y) < x\}$$
$$G(x) := \sup\{y : F(y) < x\}$$
← "Inverse" cdf

If $U \sim \text{Unif}(0,1)$ r.v., and
$\hat{X}_n = G_n(U)$, then $\hat{X}_n \xrightarrow{a.s.} \hat{X} = G(U)$
and $\hat{X}_n \overset{d}{=} X_n$ and $\hat{X} \overset{d}{=} X$.

LECTURE 14

Oct 23rd, 2024

Lemma from last time: If $X_n \xrightarrow{w} X$ and if $U \sim \text{Unif}(0,1)$ and if $G_n(x) = \sup\{y : F_n(y) < x\}$, then
$G_n(U) \sim X_n$ and $G_n \longrightarrow$ a.s. to a r.v. having the same distribution as X.

$\left(G(U),\ G(x) = \sup\{y : F(y) < x\}\right)$.

Useful for proving this

Lemma: If g is continuous and $g \geq 0$, then
$$\liminf_{n\to\infty} \mathbb{E}\, g(X_n) \geq \mathbb{E}\, g(X)$$
if $X_n \to X$ weakly.

Pf: Can assume $X_n \xrightarrow{a.s.} X$ by the lemma from last lecture and then just use the usual Fatou.

Lemma: If $X_n \xrightarrow{w} X$ and $Y_n \xrightarrow{w} Y$ and if X_n and Y_n are <u>independent</u> $\forall n$, then $X_n + Y_n \longrightarrow X + Y$ and X and Y are independent.

Pf: Let $\hat{X}_n := G_n^{X_n}(U)$ ← G's like the ones above

$\hat{Y}_n := G_n^{Y_n}(U')$, where $U' \sim \text{Unif}(0,1)$ indep. of U.

Note: $\begin{pmatrix} \hat{X}_n \\ \hat{Y}_n \end{pmatrix} \sim \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ since $\hat{X}_n$ and $\hat{Y}_n$ are indep $\forall n$.

By the lemma, $\hat{X}_n \sim X_n$ and $\hat{Y}_n \sim Y_n$. So, using the product measure, we get the right result.

Since $\hat{X}_n \xrightarrow{a.s.} \hat{X}$ and $\hat{Y}_n \xrightarrow{a.s.} \hat{Y}$, where $\hat{X}$ and $\hat{Y}$ are independent (from the lemma, $\hat{X} \sim X$ and $\hat{Y} \sim Y$).

Fact: $\mathbb{E} f(\hat{X}_n, \hat{Y}_n) \longrightarrow \mathbb{E} f(\hat{X}, \hat{Y})$ $\forall f: \mathbb{R}^2 \to \mathbb{R}$ continuous and bounded

(by DCT)

So, take $f(x,y) = x + y$ to get the claim. ∎

* CHARACTERISTIC FUNCTIONS:

Def: For a real-valued r.v. X, its characteristic fct is defined as

$$\varphi_X(t) := \mathbb{E}(e^{itX}) = \int e^{itx}\, d\mu_X(x) \quad \forall t \in \mathbb{R}$$

Note: if X and Y are independent, then

$$\varphi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX}\, e^{itY})$$

$$\overset{\text{independence}}{=} \mathbb{E}(e^{itX})\,\mathbb{E}(e^{itY})$$

$$= \varphi_X(t)\,\varphi_Y(t).$$

⚠ **PROPERTIES**: If $\varphi_X(t) = \mathbb{E}(e^{itX})$ for some r.v. X, then

(i) $\varphi_X(0) = 1$

(ii) $\overline{\varphi_X(t)} = \overline{\mathbb{E}(e^{itX})} \overset{\text{linearity of } \mathbb{E}}{=} \mathbb{E}(e^{-itX}) = \varphi_X(-t)$

(iii) $|\varphi_X(t)| = |\mathbb{E}(e^{itX})| \leq \mathbb{E}|e^{itX}| \leq \mathbb{E}\,1 = 1$

(iv) $\varphi_X(t)$ is uniformly continuous on $\mathbb{R}$ since

$$|\varphi_X(t+h) - \varphi_X(t)| = |\mathbb{E}(e^{i(t+h)X}) - \mathbb{E}(e^{itX})|$$

$$= \left| \mathbb{E}\left(e^{itX}\left(e^{ihX} - 1\right)\right) \right|$$

$$\leq \mathbb{E}\left| e^{ihX} - 1 \right|$$

(v) If $a, b \in \mathbb{R}$, then

$$\varphi_{aX+b}(t) = \mathbb{E}\left(e^{it(aX+b)}\right) = e^{itb}\, \mathbb{E}\, e^{itaX}$$

$$= e^{itb} \varphi_X(at).$$

$\boxed{\text{Thm: If two random variables } X \text{ and } Y \text{ have the same characteristic function, then they have the same distribution.}}$

FACT 1: If $Z \sim N(0, \sigma^2)$, then $\varphi_Z(t) = \mathbb{E}\, e^{itZ} = e^{-t^2\sigma^2/2}$

Pf 1: Suffices to show when $\sigma^2 = 1$ b/c we can scale char. fcts. using the property above. Compute:

$$\mathbb{E}\left(e^{itZ}\right) \overset{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itz} e^{-z^2/2}\, dz$$

complete the square

$$= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(z-it)^2/2} e^{-t^2/2}\, dz$$

$$= \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_{\mathbb{R}} e^{-(z-it)^2/2}\, dz$$

$$= e^{-t^2/2}$$

$\square$

$$\int_{\mathbb{R}} e^{-(z-it)^2/2}\,dz = \int_{\Gamma_t} F(z)\,dz$$

$F(z) = e^{-z^2}$ analytic and $\Gamma_t = \{\mathbb{R} - it\}$. By Cauchy's

Thm,

$$\int_{\Gamma_t} F(z)\,dz = \int_{\Gamma_0} F(z)\,dz$$

$$= \int_{\mathbb{R}} e^{-z^2/2}\,dz$$

$$= \sqrt{2\pi}.$$

FACT 2: If $\varphi_X(t) = \mathbb{E}e^{itX}$ and $f \in C_0^\infty(\mathbb{R})$ (smooth and compactly supported), then

$$\mathbb{E}f(X) = \int \overline{\hat{f}(t)}\,\varphi_X(t)\,\frac{dt}{2\pi}, \quad \hat{f}(t) := \int_{\mathbb{R}} e^{ixt} f(x)\,dx.$$

Pf: Let $\varepsilon > 0$ and $Z \sim N(0,1)$ independent of X.

Thus

$$\mathbb{E}f(X) = \lim_{\varepsilon \to 0} \mathbb{E}f(X + \varepsilon Z).$$

On the other hand, the r.v. $X + \varepsilon Z$ has a bounded density given by

$$P_\varepsilon(x) = \int \frac{1}{\varepsilon\sqrt{2\pi}} e^{-(x-y)^2/2\varepsilon^2}\,d\mu_X(y)$$

laws of sums is the convolution of the laws

So,

$$\mathbb{E} f(X+\varepsilon Z) \overset{\text{def}}{=} \int f(x)\, p_\varepsilon(x)\, dx \overset{\text{Plancherel}}{=} \int \hat{f}(t)\, \hat{p}_\varepsilon(t) \frac{dt}{2\pi}$$

$$= \int \hat{f}(t)\, \varphi_X(t)\, e^{-\varepsilon^2 t^2/2} \frac{dt}{2\pi}$$

$$\xrightarrow{\varepsilon \to 0} \int \hat{f}(t)\, \varphi_X(t) \frac{dt}{2\pi}.$$

Alternatively, can compute explicitly:

$$p_\varepsilon(x) = \frac{1}{2\pi} \int \varphi_X(t)\, e^{-itX} e^{-\varepsilon^2 t^2/2}\, dt$$

compute this next time

$\square$

LECTURE 15

Oct 25th, 2024

Lemma: If $Z \sim N(0,1)$ independent of X, then the density of $X+\varepsilon Z$ is

$$p_\varepsilon(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(t)\, e^{-itx} e^{-\varepsilon^2 t^2/2}\, dt.$$

Corollary: If X, Y are r.v.'s s.t. $\mathbb{E} e^{itX} = \mathbb{E} e^{itY}$ (i.e., $\varphi_X = \varphi_Y$) then X, Y have the same distribution.

Pf: If same characteristic function, then $\forall f \in C_b(\mathbb{R})$,

$\mathbb{E} f(X+\varepsilon Z) = \mathbb{E} f(Y+\varepsilon Z)$ and as $\varepsilon \to 0$ $\mathbb{E} f(X+\varepsilon Z) = \mathbb{E} f(X)$.

$\square$

Pf of Lemma: density of sum = convolution

$$P_\varepsilon(x) = \frac{1}{\varepsilon\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-y)^2/2\varepsilon^2} \, d\mu_X(y)$$

Fourier transf. of Gaussian is itself

$$= \frac{1}{\varepsilon\sqrt{2\pi}} \int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-iz(x-y)/\varepsilon} e^{-z^2/2} \, dz \right) d\mu_X(y)$$

$$= \frac{1}{2\pi\varepsilon} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-iz(x-y)/\varepsilon} \, d\mu_X(y) \, e^{-z^2/2} \, dz$$

$$= \frac{1}{2\pi\varepsilon} \int e^{-izx/\varepsilon} \varphi_X\left(\frac{z}{\varepsilon}\right) e^{-z^2/2} \, dz$$

Set $t := z/\varepsilon$

$$= \frac{1}{2\pi} \int e^{-itx} \varphi_X(t) \, e^{-\varepsilon^2 t^2/2} \, dt.$$

$\square$

Corollary: If $\varphi_X(t) \in L^1(\mathbb{R}, dt)$, then X has a density given by

$$P_X(t) = \frac{1}{2\pi} \int e^{-itX} \varphi_X(t) \, dt.$$

Pf: Take $\varepsilon \to 0$ on P_ε using DCT.

$\square$

Lemma: (INVERSION FORMULA) Given a probability measure μ and characteristic fct $\varphi(t) = \int e^{itx}\, d\mu(x)$, then

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

$$= \mu\big((a,b)\big) + \frac{1}{2}\mu(\{a,b\}) \quad \text{for } a<b.$$

i.e., recover measure of intervals.

Pf: Compute

$$\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\, \varphi(t)\, dt$$

$$= \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \int_{\mathbb{R}} e^{itx}\, d\mu(x)\, dt$$

$$= \int_{\mathbb{R}} \left(\int_{-T}^{T} \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\, dt \right) d\mu(x)$$

symmetry of interval

$$= \int_{\mathbb{R}} \underbrace{\left(\int_{-T}^{T} \frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t}\, dt \right)}_{=:\, F_T(x)} d\mu(x)$$

Claim 1: $F_T(x) \longrightarrow \begin{cases} 2\pi, & x\in(a,b) \\ \pi, & x = a \text{ or } b \\ 0, & \text{otherwise} \end{cases}$

Claim 2: $|F_T(x)|$ is uniformly bounded in T and x.

Now, for $\theta \in \mathbb{R}$, consider

sgn(0) = 0

$$\int_{-T}^{T} \frac{\sin \theta t}{t}\, dt = \operatorname{sgn}\theta \int_{-T}^{T} \frac{\sin(|\theta| t)}{t}\, dt$$

let $u := |\theta| t$

$$= \operatorname{sgn}\theta \int_{-|\theta|T}^{|\theta|T} \frac{\sin u}{u}\, du.$$

Recall: $\lim_{T\to\infty} \int_{-T}^{T} \frac{\sin u}{u}\, du = \pi$

This shows Claim 2

Thus, $F_T(x) \to \pi\left(\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b)\right)$ which proves Claim 1.

■

Ex: If $Z_1 \sim N(0, \sigma_1^2)$ and $Z_2 \sim N(0, \sigma_2^2)$ and Z_1, Z_2 are independent. Then $Z_1 + Z_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$

via convolution before but now can do it w/ char. fcts.

Compute:
$$\varphi_{Z_1+Z_2}(t) = E\left(e^{it(Z_1+Z_2)}\right)$$
$$= E\left(e^{itZ_1}\right) E\left(e^{itZ_2}\right)$$
$$= e^{-t^2\sigma_1^2/2}\, e^{-t^2\sigma_2^2/2}$$
$$= e^{-t^2(\sigma_1^2+\sigma_2^2)/2} = \varphi_{N(0,\sigma_1^2+\sigma_2^2)}(t)$$

$\Rightarrow Z_1 + Z_2 \sim N(0, \sigma_1^2 + \sigma_2^2).$

Ex: $X \sim Exp(\lambda)$ $\left(\text{i.e., } \mathbb{P}(X \geq u) = e^{-\lambda u}\right)$, then

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \int_0^\infty e^{itx} e^{-\lambda x} \lambda dx = \frac{\lambda}{\lambda - it}$$

$\Rightarrow$ sum of indep exp. r.v.'s is not exp.

Check: if $Z_1 \sim \text{Poisson}(\lambda_1)$, $Z_2 \sim \text{Poisson}(\lambda_2)$, Z_1 and Z_2 independent, then $Z_1 + Z_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

* CHARACTERISTIC FUNCTIONS & WEAK CONVERGENCE

FACT: Clearly, if $X_n \xrightarrow{w} X$, then $\varphi_{X_n}(t) \longrightarrow \varphi_X(t)$ $\forall t \in \mathbb{R}$.

Converse ?

i.e. $\mathbb{E}(e^{iX_n t}) \to \mathbb{E}(e^{itX})$ $\forall t \in \mathbb{R}$.

Lemma: If X_n is a sequence of r.v.'s that is tight, and if $\varphi_{X_n}(t) = \mathbb{E}\, e^{itX_n} \longrightarrow f(t)$ $\forall t \in \mathbb{R}$, then

CRUCIAL ASSUMPTION

1. $f(t) = \mathbb{E}\, e^{itX}$ for some r.v X

2. $X_n \to X$ weakly.

Ex: If $X_n \sim N(0,n)$, then $\mathbb{E}e^{itX_n} \longrightarrow \underbrace{\mathbb{1}\{t=0\}}_{}$

from before $\mathbb{E}e^{itX_n} = e^{-t^2 n/2}$

Not the char. fct. of the limit. These do not converge to r.v. $=0$, it converges to measure $=0$.

Pf: Since X_n is tight, all of its subseqs have a further subseq. that converges to some r.v. X. So, by assumption, $f(t) = \mathbb{E}e^{itX}$. Since distributions are determined by char. fcts., and all sub-subsequential limits of X_n coincide, then $X_n \xrightarrow{\text{weakly}} X$. ∎

since the space of r.v.'s is metrizable, it follows that the whole seq $X_n \longrightarrow X$.

Rmk: Tightness is crucial here. So, it would be nice to have a way to check tightness from the char. fct.

⚠ **Lemma**: Suppose we have a sequence of r.v.'s X_n s.t. $\mathbb{E}e^{itX_n} \longrightarrow f(t)$ $\forall t \in \mathbb{R}$. If $f(t)$ is continuous at $t=0$, then the X_n's are tight.

CONSEQUENCES:

LEVY'S CONTINUITY THEOREM: If X_n is a sequence of r.v.'s s.t. $\mathbb{E}e^{itX_n} \longrightarrow f(t)$ and $f(t)$ is continuous at $t=0$, then $X_n \xrightarrow{\text{weakly}} X$ s.t. $f(t) = \mathbb{E}e^{itX}$.

$\downarrow$ Direct consequence

Thm: If X_n and X are r.v.'s,

$$X_n \xrightarrow{\text{weakly}} X \iff \mathbb{E}e^{itX_n} \longrightarrow \mathbb{E}e^{itX} \quad \forall t \in \mathbb{R}$$

Intermediate lemmas to prove the lemma above.

Lemma 1: For any r.v. X, $\varphi(t) = \mathbb{E}e^{itX}$ and $u>0$, we have

$$\mathbb{P}\left(|X| > \frac{1}{u}\right) \leq 7\,\frac{1}{u}\int_0^u \big(1 - \operatorname{Re}(\varphi(t))\big)\,dt.$$

Pf:
$$\frac{1}{u}\int_0^u 1-\operatorname{Re}(\varphi(t))\,dt = \frac{1}{u}\int_0^u\int_{\mathbb{R}} 1-\cos(tx)\,d\mu_X(x)\,dt$$

$$= \int_{\mathbb{R}} \frac{1}{u}\int_0^u 1-\cos(tx)\,dt\,d\mu_X(x)$$

$$= \int_{\mathbb{R}} \frac{1}{u}\left(u - \frac{\sin ux}{x}\right)\,d\mu_X(x)$$

$$= \int_{\mathbb{R}} 1 - \frac{\sin xu}{xu} \, d\mu_X(x)$$

$1 - \frac{\sin y}{y} \geq 0 \quad \forall y \in \mathbb{R}$

$$\geq \int_{|x| \geq \frac{1}{u}} 1 - \frac{\sin xu}{xu} \, d\mu_X(x)$$

Check that for $y > 1$, $\frac{\sin y}{y} \leq \frac{\sin 1}{1} \leq \frac{6}{7}$

$$\geq \int_{|x| \geq \frac{1}{u}} \left(1 - \frac{\sin 1}{1}\right) d\mu_X(x)$$

$$\geq \frac{1}{7} \int_{|x| \geq \frac{1}{u}} d\mu_X(x) = \frac{1}{7} \mathbb{P}\left(|X| \geq \frac{1}{u}\right).$$

$\blacksquare$

Pf of lemma: Suffices to prove that $\forall \varepsilon > 0$, $\exists M > 0$ s.t. $\limsup_{n \to \infty} \mathbb{P}(|X_n| > M) \leq \varepsilon$ for tightness.

By continuity assumption, $\forall \varepsilon > 0 \ \exists \delta > 0$ s.t. $\forall |t| < \delta$,

$\underbrace{|f(t) - f(0)|}_{= |f(t) - \mathbb{E} e^{i0X_n}| = |f(t) - 1|} < \varepsilon$. b/c of ptwise limit assumption

So,

$$\limsup_{n \to \infty} \mathbb{P}(|X_n| > 1/\delta) \leq \limsup_{n \to \infty} \frac{7}{\delta} \int_0^\delta \overbrace{\left(1 - \mathrm{Re}(\mathbb{E} e^{itX_n})\right)}^{\text{bounded}} dt$$

Bdd Convergence Thm $\longrightarrow$ $$= \frac{7}{\delta}\int_0^\delta 1 - \mathrm{Re}(f(t))\, dt$$

$$\leq \frac{7}{\delta}\int_0^\delta \varepsilon\, dt = 7\varepsilon.$$

$\Longrightarrow$ X_n are tight. ∎

ANOTHER PROOF OF CLT

Thm (CLT) If X_i are iid w/ $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \xrightarrow{\text{weakly}} N(0,1)$$

Lemma: $\mathbb{E}\, e^{itX_i/\sqrt{n}} = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$

Pf: Assuming the lemma,

$$\mathbb{E}\exp\left(it\frac{1}{\sqrt{n}}\sum_{j=1}^n X_j\right) = \left(\mathbb{E}\exp\left(i\frac{t}{\sqrt{n}}X_j\right)\right)^n$$

$$= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \longrightarrow e^{-t^2/2}.$$

∎

LECTURE 16

Oct 30th, 2024

* **MOMENT METHOD**: $\{X_n\}$ real-valued r.v.'s. Suppose that $\forall k$,

$$\mathbb{E}X_n^k \xrightarrow{(*)} \mathbb{E}X^k \quad \text{as } n\to\infty$$

← implicitly assume that all of these exist...

Is it true that if $X_n \xrightarrow{w} X$ then

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X) \quad \forall f \in C_b(\mathbb{R}) \quad ?$$

↳ Converse not true: $\mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$

$\mathbb{P}(X_n = e^n) = \frac{1}{n}$

Ex: If $\exists M > 0$ s.t. $X_n \in (-M, M)$ $\forall n$, then

$$X_n \longrightarrow X \quad \text{weakly.}$$

$\Rightarrow$ Approximate f uniformly by poly. on compact sets.

COMPACTNESS & UNIQUENESS ARGUMENT: Since $\mathbb{E}X_n^2 \to \mathbb{E}X^2$ by assumption, we know that

$$\sup_n \mathbb{E}X_n^2 \le C < \infty$$

Chebyshev

$$\Rightarrow \mathbb{P}(|X_n| > M) \le \frac{C}{M^2} \Rightarrow X_n\text{'s are tight.}$$

Claim: If $X_{n_k} \to \hat{X}$ weakly, then $\mathbb{E}X_{n_k}^{\ell} \to \mathbb{E}\hat{X}^{\ell}$ (in the setup of (*)).

Q: Do the moments of a r.v. determine its distribution?

A: No! Counterexample:

$Z \sim N(0,1)$ and $X \sim e^{Z}$. Then the density of X is

$$f_0(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\log(x)^2/2} \mathbb{1}_{\{x>0\}}.$$

← all moments for this are finite.

For all $a \in (-1,1)$, let $f_a(x) := f_0(x)\left(1 + a\sin(2\pi \log x)\right)\mathbb{1}_{\{x>0\}}$

Cl: $f_a(x)$ is a density $\forall a \in (-1,1)$ and it has the same moments as X_0; i.e.,

$$\int f_a(x)\, x^r\, dx = \int f_0(x)\, x^r\, dx \quad \forall r \in \mathbb{N}.$$

Need to check: $\displaystyle\int_0^{\infty} x^r \frac{1}{x} e^{-\log(x)^2/2} \sin(2\pi \log x)\, dx = 0 \quad \forall r \geq 0$

let $x := \exp(s+r)$

$\frac{dx}{x} = ds$

$$\| \qquad \int_{\mathbb{R}} \exp(sr + r^2) \exp\left(-\frac{(s+r)^2}{2}\right) \sin(2\pi(s+r))\, ds$$

$$\| \qquad e^{r^2/2} \int_{\mathbb{R}} e^{-s^2/2} \underbrace{\sin(2\pi(s+r))}_{=\sin 2\pi s = 0 \text{ since } r \in \mathbb{N}\cup\{0\}}\, ds = 0$$

<u>Upshot</u>: the f_a have same moments but they give different distributions.

<u>Compute the moments</u>: $\mathbb{E}(X^r) = \mathbb{E}(e^{rZ}) = \underbrace{e^{r^2/2}}$

$r! \leq r^r = e^{r\log r}$. increases faster than $r!$

Thm: Let $\mu_k := \mathbb{E}X^k < \infty$ $\forall k$. If $\exists C>0$ s.t. $\mu_k \leq (Ck)^k$ $\forall$ even k, then X is the distribution with these moments k.

Ex: Normal distribution $\mathbb{E}|Z|^k \leq (Ck)^{k/2} \leq (Ck)^k$. So normal distribution is an example of r.v. that is determined by its moments. $\Rightarrow$ Prove CLT.

Pf: <u>Check that</u>: $\mathbb{E}e^{itX}$ is analytic.

Note that, by Cauchy-Schwarz, we have that, for odd moments:

$$\mathbb{E}X^{2k+1} \leq \left(\mathbb{E}X^{2k+2}\right)^{\frac{1}{2}}\left(\mathbb{E}X^{2k}\right)^{\frac{1}{2}}$$

$$\leq \left(C(k+2)\right)^{\frac{2k+2}{2}}\left(C\cdot 2k\right)^{2k/2} \leq \left(\tilde{C}(2k+1)\right)^{2k+1}$$

$\Longrightarrow$ Can assume we have growth $\mathbb{E}|X|^k \le (Ck)^k \ \forall k$ odd as well.

Thm: given $t \in \mathbb{R}$, $\theta \in \mathbb{C}$ small,

Taylor series

$$\left| e^{i(t+\theta)X} - e^{itX}\left(\sum_{j=0}^{n-1} \frac{(i\theta X)^j}{j!}\right) \right|$$

remainder...

$$\le \left| e^{i\theta X} - \sum_{j=0}^{n-1} \frac{(i\theta X)^j}{j!} \right| \le \frac{|\theta X|^n}{n!}$$

$$(n! \ge n^n e^{-n}) \longrightarrow \le \frac{|\theta X|^n}{n^n} e^n .$$

So,

$$\left| \mathbb{E} e^{i((t+\theta)X)} - \left(\sum_{j=0}^{n-1} \frac{(i\theta)^j}{j!}\right) \mathbb{E}(X^j e^{itX}) \right|$$

$$\le \frac{e^n}{n^n} |\theta|^n \mathbb{E}|X|^n$$

$$\le \frac{e^n}{n^n} |\theta|^n C^n n^n \le (Ce|\theta|)^n \xrightarrow[n \nearrow \infty]{} 0 \quad \text{if } |\theta| < \frac{1}{Ce}.$$

Upshot: The function $t \longmapsto \varphi(t)$ is analytic in a nbhd of $\mathbb{R}$.

$\Rightarrow$ For small t, $\mathbb{E}e^{itX}$ is just a fct of the moments themselves since

$$\mathbb{E}e^{itX} = \sum_{j=0}^{\infty} \frac{(it)^j}{j!} \mathbb{E}X^j$$

$\Rightarrow$ $\mathbb{E}e^{itX}$ is determined by the moments of X.
(as shown before, char. fct. determines distribution) $\blacksquare$

- From HW02, showed that

$$\mathbb{E}|X|^k \le (Ck)^k \iff \mathbb{P}(|X| > u) \le 2e^{-\tilde{C}u}$$

for some $\tilde{C} > 0$ (sub-exp.)

If r.v. is sub-exponential, then $\mathbb{E}e^{\alpha X} < \infty$ $\forall |\alpha| < \varepsilon$.

- <u>Example</u> <u>of</u> <u>moment</u> <u>method</u>:

Thm: If X_n is a seq. of r.v.'s having moments for all orders s.t.

$$\mathbb{E}X_n^k \longrightarrow \mathbb{E}X^k =: \mu_k$$

and $\mu_k \le (Ck)^k$ $\forall$ even k, then $X_n \xrightarrow{\text{weakly}} X$.

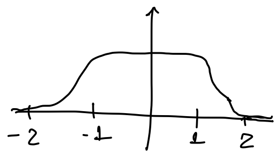
Pf: By compactness & uniqueness argument from before, suffices to prove that if $Y_n \to Y$ and

$$\mathbb{E} Y_n^\ell \to \mu_\ell, \quad \text{then } \mathbb{E} Y^\ell = \mu_\ell .$$

Fatou: $\mathbb{E} Y^{2\ell} \le \liminf_{n\to\infty} \mathbb{E} Y_n^{\ell} < \infty$.

Let ρ be a C^∞ cutoff fct of $(-1,1)$; i.e.,

$$\rho(x) = \begin{cases} 1, & x \in (-1,1) \\ 0, & x \notin (-2,2) \end{cases}$$



Let $M > 0$

$$\mathbb{E} Y_n^\ell = \underbrace{\mathbb{E} Y_n^\ell \rho\left(\frac{Y_n}{M}\right)} + \underbrace{\mathbb{E}\left(Y_n^\ell\left(1-\rho\left(\frac{Y_n}{M}\right)\right)\right)}$$

by assumption of $Y_n \xrightarrow{w} Y$ $\quad n \nearrow \infty$

$$\downarrow$$

$$\mathbb{E} Y^\ell \rho\left(\frac{Y}{M}\right)$$

Cauchy-Schwarz $\le \left(\mathbb{E} Y_n^{2\ell}\right)^{\frac{1}{2}} \left(\mathbb{P}\left(|Y_n| > \frac{M}{2}\right)\right)^{\frac{1}{2}}$

Chebyshev $\le \left(\mathbb{E} Y_n^{2\ell}\right)^{\frac{1}{2}} \left(\mathbb{E}|Y_n|^2 \frac{4}{M^2}\right)^{1/2}$

$$\Rightarrow \limsup_{M\to\infty} \sup_n \left| \mathbb{E} Y_n^\ell (1-\rho)\left(\frac{Y_n}{M}\right)\right| = 0$$

■

LECTURE 17

Nov 13th, 2024

LIMIT THEOREMS & CHARACTERISTIC FUNCTIONS IN $\mathbb{R}^k$

If we have a sequence $X^{(n)} = (X_1^{(n)}, \ldots, X_k^{(n)}) \in \mathbb{R}^k$ of random vectors in $\mathbb{R}^k$. We say $X^{(n)}$ converges weakly to $X \in \mathbb{R}^k$ if $\forall f \in C_b(\mathbb{R}^k)$ we have $\mathbb{E} f(X^{(n)}) \longrightarrow \mathbb{E} f(X)$ as $n \to \infty$

Much stronger than component-wise convergence

Counterexample to component-wise weak convergence: X, Y same distrib. but indip. Take sequence $(X,Y), (X,X), (X,Y), (X,X), \ldots$

Rmk: We have a Portmanteau for arbitrary m. spaces.

Def: For a random vector $X \in \mathbb{R}^k$, we define a characteristic function by

$$\varphi_X(t) := \mathbb{E}\, e^{it\cdot X} = \int_{\mathbb{R}^k} e^{it\cdot x}\, d\mu_X(x) \quad \text{for } t \in \mathbb{R}^k.$$

Thm: If X and Y are random vectors in $\mathbb{R}^k$, then

$$X \sim Y \iff \mathbb{E} e^{it\cdot X} = \mathbb{E} e^{it\cdot Y} \quad \forall t \in \mathbb{R}^k.$$

$Z = (Z_1, \dots, Z_k)$ w/ Z_j iid and $Z \sim N(0,1)$

Lemma: If $Z \sim N(0, \mathbb{1}_k)$, then the density of $X + \varepsilon Z$, where X indep. of Z, is given by

$$p_\varepsilon(x) = \frac{1}{(2\pi)^k} \int_{\mathbb{R}^k} \varphi_X(t)\, e^{-it\cdot x}\, e^{-\varepsilon^2 \|t\|^2/2}\, dt_1 \cdots dt_k$$

Pf of Thm: If X, Y have the same Fourier transform, then $\forall \varepsilon > 0$, $X + \varepsilon Z \sim Y + \varepsilon Z$.

As $\varepsilon \to 0$, $X + \varepsilon Z \to X$ weakly

$Y + \varepsilon Z \to Y$ weakly. ∎

Def: (Tightness) A seq. of r.v.'s $X^{(n)} \in \mathbb{R}^k$ is <u>tight</u> if $\forall \varepsilon > 0$ $\exists M > 0$ s.t.

$$\sup_n \mathbb{P}\left(\|X^{(n)}\| > M\right) \le \varepsilon$$

Since we're in finite dimensions

$\iff$ each component $(X_i^{(n)})_n$ is tight $\forall i$.

Thm: (LEVY'S CONTINUITY) Let $X^{(n)} \in \mathbb{R}^k$ and suppose $\mathbb{E}\, e^{it\cdot X^{(n)}} \longrightarrow f(t) \ \forall t \in \mathbb{R}^k$ s.t. $f(t)$ is continuous at $t=0$. Then $X^{(n)} \xrightarrow{\text{weakly}} X$, s.t. $X \in \mathbb{R}^k$ and $\mathbb{E}\, e^{it\cdot X} = f(t)$.

Pf: Cl 1: $X_i^{(n)}$ is tight $\forall i$.

The assumptions here imply that, for each component,

$$\mathbb{E}\, e^{iuX_i^{(n)}} \longrightarrow f(0,0,\_\_,0,\underset{\substack{\uparrow \\ i\text{th component}}}{u},0,\_\_,0) \quad \forall u \in \mathbb{R}.$$

Now, $u \longmapsto f(0,\_,0,u,0,\_,0)$ is continuous at 0. So, by 1-dim Levy Cont. Thm, $X_i^{(n)}$ converges weakly, hence are tight $\forall i$. $\underset{\substack{\uparrow \\ \text{finite dimensions}}}{\Rightarrow}$ the whole seq. $X^{(n)}$ is tight.

By functional analysis... in higher dimensions,

$$\text{tightness} \iff \text{sequential compactness}$$

So, every subseq. of $X^{(n)}$ has a further subseq. converging to some $X \in \mathbb{R}^k$.

But X must have char. fct. equal to $f(t)$. So, <u>all</u>

Subsequential limits of $X^{(n)}$ are the same

$\Rightarrow X^{(n)} \xrightarrow{\text{weakly}} X.$ ∎

Immediate Consequence

$$\textbf{Thm: } X^{(n)} \xrightarrow{\text{weakly}} X \iff \mathbb{E}e^{it\cdot X^{(n)}} \xrightarrow[\text{ptwise}]{} \mathbb{E}e^{it\cdot X} \quad \forall t\in\mathbb{R}^k$$

Thm: (Cramer-Wald Device)

random variable ($\theta\cdot X^{(n)}$ is just a #)

$$X^{(n)} \xrightarrow{\text{weakly}} X \iff \theta\cdot X^{(n)} \xrightarrow{\text{weakly}} \theta\cdot X \quad \forall\theta\in\mathbb{R}^k$$

random vector

$$\iff \theta\cdot X^{(n)} \xrightarrow{\text{weakly}} \theta\cdot X \quad \forall\theta\in\mathbb{R}^k,\ \|\theta\|=1.$$

Pf: ($\Rightarrow$) Trivial.

($\Leftarrow$) Suppose $\theta\cdot X^{(n)} \xrightarrow{\text{weakly}} \theta\cdot X$. Then by 1-d Levy,

$$\Updownarrow$$

$$\mathbb{E}e^{it\theta\cdot X^{(n)}} \longrightarrow \mathbb{E}e^{it\theta\cdot X} \quad \forall t\in\mathbb{R}.$$ □

Def: (MULTIVARIATE NORMAL DISTRIBUTION) Let $m\in\mathbb{R}^k$ and let C be a symmetric and positive definite $k\times k$ matrix (i.e., $C^T=C$ and $\alpha^T C\alpha>0$ $\forall\alpha\in\mathbb{R}^k$ nonzero)

Then we say $Z \sim N(m, C)$ if its density is

$$\frac{1}{(2\pi)^{k/2}\sqrt{\det C}} \exp\left(-(x-m)^T \frac{C^{-1}}{2}(x-m)\right) dx_1 \cdots dx_k .$$

Exercise: 1. The above is a density (i.e., its integral $= 1$)

2. $\mathbb{E} Z_i = m_i$

3. $\mathrm{Cov}(Z_i, Z_j) = C_{ij}$.

4. If $j < k$, then $(Z_1, \_, Z_j) \sim N(\hat{m}, \hat{C})$, where $\hat{m} = (m_1, \_, m_j)$ and $\hat{C} =$ top left $j \times j$ block of matrix C

Counterexample: $X_1 \sim N(0,1)$, $X_2 \sim |X_1| (\pm 1$ w/ prob. 50/50$)$
then (X_1, X_2) does not have a Gaussian distrib. in $\mathbb{R}^2$.

5. If $Z \sim N(0, \mathbb{1}_{k\times k})$, then $Z = (Z_1, \_, Z_k)$ and $Z_i \sim N(0,1)$ and Z_i's are all iid.

LECTURE 18

Nov 15th, 2024

MULTIVARIATE NORMAL: $Z \in \mathbb{R}^k$, $Z \sim N(m, C)$ if has density:

$$\frac{1}{(2\pi)^{k/2}} \frac{1}{\sqrt{\det C}} \exp\left(-(z-m)^T \frac{C^{-1}}{2} (z-m) \right) d\vec{z}.$$

$$\begin{cases} m_i = \mathbb{E} Z_i \\ C_{ij} = \mathrm{Cov}(Z_i, Z_j) \end{cases}$$

$$\mathrm{Var}(\alpha^T Z) = \mathrm{Cov}\left(\sum_i \alpha_i Z_i, \sum_j \alpha_j Z_j \right)$$

Cov(·,·) bilinear ↓

$$= \sum_{i,j} \alpha_i \alpha_j \mathrm{Cov}(Z_i, Z_j)$$

$$= \sum_{i,j} \alpha_i \alpha_j C_{ij} = \alpha^T C \alpha$$

Lemma: (Change of Variables) If $Z \sim N(m, C)$ and if A is an invertible $\overset{k \times k}{\vee}$ matrix, then $X := AZ$ is s.t.

$$X \sim N(Am, ACA^T)$$

Pf: WLOG (via shifts), $m = 0$. Then

$$\mathbb{E} f(X) = \mathbb{E} f(AZ) = \frac{1}{(2\pi)^{k/2}} \frac{1}{\sqrt{\det C}} \int f(Az) \exp\left(-z^T \frac{C^{-1}}{2} z \right) dz$$

Change variables: $x = Az$ $\longrightarrow z^T C^{-1} z = x^T (A^{-1})^T C^{-1} A^{-1} x$

$$dx = |\det A|\, dz \qquad = x^T (ACA^T)^{-1} x$$

$$= \frac{1}{(2\pi)^{k/2}\sqrt{\det C}\,|\det A|} \int f(x) \exp\left(-x^T \frac{(ACA^T)^{-1}}{2} x\right) dx .$$

$\square$

Corollary: If $X \sim N(0, C)$ and $Z \sim N(0, \mathbb{1}_k)$ then $X \sim \sqrt{C}\, Z$.

(Well-def b/c C pos. def.)

Corollary: If $X \sim N(m, C)$, then the char. fct of X is:

$$\varphi_X(t) = \mathbb{E} \exp(it \cdot X) = e^{it\cdot m}\, \mathbb{E} \exp(it \cdot (X - m))$$

$$= e^{it\cdot m}\, \mathbb{E} \exp(it \cdot \sqrt{C} Z)$$

$$= e^{it\cdot m}\, \mathbb{E} \exp(i (\sqrt{C} t) \cdot Z)$$

$$= e^{it\cdot m} \prod_{j=1}^{k} \mathbb{E}\left(i (\sqrt{C} t)_j \cdot Z_j\right)$$

$$= e^{it\cdot m} \prod_{j=1}^{k} e^{-(\sqrt{C} t)_j^2/2}$$

$$\boxed{= \exp\left(-t^T C t/2\right) e^{it\cdot m}}$$

$$\boxed{= \mathbb{E} \exp(i t \cdot X)}$$

Corollary: $Z \sim N(0, C) \iff \theta \cdot Z \sim N(0, \theta^T C \theta)$ $\forall \theta \in \mathbb{R}^k$ s.t. $\|\theta\| = 1$.

MULTIVARIATE CLT: Let $X_i^{(i)} \in \mathbb{R}^k$ be a seq. of iid random vectors w/ $\mathbb{E} X_i^{(i)} = 0$ and $C_{ij} = \mathbb{E} X_i^{(i)} X_j^{(i)}$ w/ $C = (C_{ij})_{i,j}$ positive-definite. Then

element of sequence

vec. compon.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X^{(i)} \xrightarrow{\text{weakly}} N(0, C)$$

Pf: $\forall \theta \in \mathbb{R}^k$ we know by the 1-dim CLT that

$$\theta \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X^{(i)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \theta \cdot X^{(i)} \xrightarrow{\text{weakly}} N(0, \theta^T C \theta).$$

In particular, $\forall \theta \in \mathbb{R}^k$, this converges to $\theta \cdot Z$ where $Z \sim N(0, C)$. Thus, by Cramer–Wald device, we get

Cor. above

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X^{(i)} \xrightarrow{\text{weakly}} Z \sim N(0, C).$$

■

CONDITIONAL EXPECTATION

Ex: (CONDITIONAL PROBABILITY) X_1, X_2 are 6-sided dice. If we know $X_1 + X_2 = 7$, then what is $\mathbb{P}$ that X or Y is 3?

6 possibilities: $(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)$

$\Rightarrow$ Answer $= \frac{1}{3}$.

Def: Probability of event A occurring conditional on B occurring is defined by

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

Note: $\mathbb{P}(A \mid B) + \mathbb{P}(A^c \mid B) = 1$

In ex. above: $A = \{X \text{ or } Y = 3\}$, $B = \{X + Y = 7\}$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/18}{1/6} = \frac{1}{3}.$$

FORMALIZE TO CONTINUOUS R.V.'S

⋮

Ex: Say X and Y have joint density $f(x,y): \mathbb{R}^2 \to \mathbb{R}$. What is the density of X conditional on $Y = y$?

⋮

Things can go wrong if we divide by zero...

⚠

Def: (Conditional Expectation) Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a prob. space and let X be a r.v. $\mathbb{E}|X| < \infty$. Let $\mathcal{B} \subset \mathcal{A}$ be a sub-σ-algebra of $\mathcal{A}$. We say that a r.v. Y is <u>the conditional expectation</u> of X w.r.t. $\mathcal{B}$ if

(i) $\sigma(Y) \in \mathcal{B}$ (i.e., Y is $\mathcal{B}$-measurable) (i.e., we can actually compute Y ...)

(ii) $\forall B \in \mathcal{B}$, we have $\mathbb{E}\, Y \mathbb{1}_B = \mathbb{E} X \mathbb{1}_B \iff \int_B Y \, d\mathbb{P} = \int_B X \, d\mathbb{P}$.

i.e., Y gives you the best guess for value of X in $\mathcal{B}$

⚠ Rmk: think about this as: if $\mathcal{B} = \mathcal{A}$ then $\mathcal{A}$ encodes all information

$\mathcal{B} = \sigma(X+Y) \rightsquigarrow$ cannot separate events $(2,5)$ and $(4,3)$ here for us.

NOTATION: $Y = \mathbb{E}[X \mid \mathcal{B}]$

Def: If Z is another r.v., $\mathbb{E}[X \mid Z] := \mathbb{E}[X \mid \sigma(Z)]$ $= f(Z)$

⚠ ⚠

Exercise: If Y is $\sigma(Z)$-measurable, then $Y = f(Z)$ for some measurable $f: \mathbb{R} \to \mathbb{R}$.

Cl: Existence & Uniqueness of condi exp.

Pf: (Existence) Define a measure μ on $(\Omega, \mathcal{B})$ by $\mu(B) := \mathbb{E}(X \mathbb{1}_B)$. Since $\mathbb{E}|X| < \infty$, this defines a σ-finite signed measure.

On the other hand, define ν on $(\Omega, \mathcal{B})$ by $\nu := \mathbb{P}|_{\mathcal{B}}$ i.e., $\nu(B) = \mathbb{P}(B)$. If $\nu(B) = 0$, then $\mathbb{P}(B) = 0$, i.e., $\mathbb{E} X \mathbb{1}_B = 0 = \mu(B) \Rightarrow \mu$ is absolute cont. w.r.t. ν

By Radon-Nikodym Thm, $\exists! \frac{d\mu}{d\nu} \in L^1(\Omega, \mathcal{B}, d\nu)$ s.t.

$$\mu(B) = \int_B \frac{d\mu}{d\nu} \, d\nu . \qquad \text{Let } Y := \frac{d\mu}{d\nu} .$$

Then $\mu(B) = \mathbb{E} X \mathbb{1}_B = \int_B Y \, d\mathbb{P} = \mathbb{E} Y \mathbb{1}_B$.

$\Rightarrow$ (i) and (ii) are satisfied

(Uniqueness) $Y_1, Y_2 = \mathbb{E}[X | \mathcal{B}]$. Let

$$B := \{ Y_1 - Y_2 > 0 \} \in \mathcal{B} .$$

Then
$$\mathbb{E}\big((Y_1 - Y_2) \mathbb{1}_B\big) = \mathbb{E} Y_1 \mathbb{1}_B - \mathbb{E} Y_2 \mathbb{1}_B$$
$$= \mathbb{E} X \mathbb{1}_B - \mathbb{E} X \mathbb{1}_B = 0 .$$

$\Rightarrow \mathbb{P}(B) = 0$.

Apply same argument to $\{Y_2 - Y_1 > 0\} \rightsquigarrow Y_1 = Y_2$ a.s. $\square$

TERMINOLOGY: • Conditional expectation is an equivalence class of measurable functions.

• Any specific choice of conditional expectation representative is called a VERSION.

PROPERTIES / EXAMPLES: (1) If X is $\mathcal{B}$-measurable, then the conditional exp. of X w.r.t. $\mathcal{B}$ is: $\mathbb{E}[X|\mathcal{B}] = X$ by uniqueness.

(2) If $\mathcal{B} = \{\emptyset, \Omega\}$, then $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}X$.

If $\mathcal{B} = \mathcal{A}$, then $\mathbb{E}[X|\mathcal{B}] = X$.

(3) Linearity: $\mathbb{E}(aX + bY \mid \mathcal{B}) = a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B})$

Cl: Clearly (i) $a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B})$ is $\mathcal{B}$-measurable

(ii) $\mathbb{E}\left(\left(a\mathbb{E}(X|\mathcal{B}) + b\mathbb{E}(Y|\mathcal{B})\right)\mathbb{1}_B\right)$

$$= a\,\mathbb{E}\left(\mathbb{E}(X|\mathcal{B})\,\mathbb{1}_B\right) + b\,\mathbb{E}\left(\mathbb{E}(Y|\mathcal{B})\,\mathbb{1}_B\right)$$

$$= a\,\mathbb{E}\left(X\mathbb{1}_B\right) + b\,\mathbb{E}\left(Y\mathbb{1}_B\right)$$

$$= \mathbb{E}\left((aX+bY)\,\mathbb{1}_B\right).$$

(4) If $\sigma(X)$ is independent of $\mathcal{B}$, then $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}X$

$\forall B \in \mathcal{B}$, $\mathbb{E}(X\mathbb{1}_B) = \mathbb{E}X\,\mathbb{E}\mathbb{1}_B = \mathbb{E}(\mathbb{E}(X)\,\mathbb{1}_B)$

(5) (WARNING) If $X = \mathbb{1}_A$, $A \in \mathcal{A}$, then

$$\mathbb{E}(\mathbb{1}_A \mid \mathcal{B}) =: \underbrace{\mathbb{P}(A \mid \mathcal{B})}_{\text{still a random variable ?!}} \text{ is called}$$ CONDITIONAL PROBABILITY !

Rmk: $\mathbb{P}(A|\mathcal{B}) + \mathbb{P}(A^c|\mathcal{B}) = \mathbb{E}(\mathbb{1}_A|\mathcal{B}) + \mathbb{E}(\mathbb{1}_{A^c}|\mathcal{B})$

$$= \mathbb{E}[1|\mathcal{B}] = 1.$$

Ex: Suppose $\Omega = \coprod_{i=1}^{\infty} \Omega_i$ such that $\mathbb{P}(\Omega_i) > 0$. Let

$\mathcal{B} := \sigma(\Omega_1, \Omega_2, \dots)$. Then given $A \subset \Omega$, $A \subset \mathcal{A}$,

$$\mathbb{P}(A|\mathcal{B}) = \sum_{i=1}^{\infty} \frac{\mathbb{P}(A \cap \Omega_i)}{\mathbb{P}(\Omega_i)} \mathbb{1}_{\Omega_i}$$ ← This is a random variable.

Check: (i) Clear

(ii) Any $B \in \mathcal{B}$ is the disjoint union of some subcollection of the Ω_i's. By linearity, take $B = \Omega_j$. Then

$$\mathbb{E}\left(\mathbb{1}_{\Omega_j} \sum_{i=1}^{\infty} \frac{\mathbb{P}(A\cap\Omega_i)}{\mathbb{P}(\Omega_i)} \mathbb{1}_{\Omega_i}\right) = \mathbb{E}\left(\frac{\mathbb{P}(A\cap\Omega_j)}{\mathbb{P}(\Omega_j)} \mathbb{1}_{\Omega_j}\right)$$

$$= \mathbb{P}(A \cap \Omega_j) = \mathbb{E}(\mathbb{1}_A \mathbb{1}_{\Omega_j}).$$

(6) (TOWER PROPERTY) Say $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A}$. Then

$$\mathbb{E}(X|\mathcal{C}) \underset{\text{Check?}}{=} \mathbb{E}(\mathbb{E}(X|\mathcal{B}) \mid \mathcal{C}) \underset{\text{Clear}}{=} \mathbb{E}(\mathbb{E}(X|\mathcal{C}) \mid \mathcal{B}).$$

NTC: $\mathbb{E}(X|\mathcal{C})$ satisfies (i), (ii) for $\mathbb{E}(X|\mathcal{B})$:

(i) ok

(ii) $\forall C \in \mathcal{C}$, $\mathbb{E}\left(\mathbb{E}(X|\mathcal{C}) \mathbb{1}_C\right) = \mathbb{E}(X \mathbb{1}_C) = \mathbb{E}\left(\mathbb{E}(X|\mathcal{B}) \mathbb{1}_C\right)$

conditioning on $\mathcal{C}$ — conditioning on $\mathcal{B}$ since $\mathcal{C} \subset \mathcal{B}$.

(7) If $X \geq 0$ then $\mathbb{E}(X|\mathcal{B}) \geq 0$. Take $B := \{\mathbb{E}(X|\mathcal{B}) < 0\}$.

then $\mathbb{E}\left(\mathbb{E}(X|\mathcal{B}) \mathbb{1}_B\right) = \mathbb{E} X \mathbb{1}_B \geq 0 \Rightarrow \mathbb{P}(B) = 0$.

Cor: $X \leq Y \Rightarrow \mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(Y|\mathcal{B})$.

LECTURE 19

Nov 20th, 2024

Recall: If X is a r.v. $\mathbb{E}|X| < \infty$ and $\mathcal{B}$ is a sub-σ-algebra then $Y = \mathbb{E}[X|\mathcal{B}]$ is a r.v. called conditional expectation of X w.r.t. $\mathcal{B}$ if (i) $\sigma(Y) \subset \mathcal{B}$ (i.e., Y is $\mathcal{B}$-measurable)

(ii) $\forall B \in \mathcal{B}$, $\mathbb{E}(Y \mathbb{1}_B) = \mathbb{E}(X \mathbb{1}_B)$.

⚠

Rmk: ⚠ • if $\mathcal{B} = \sigma(Z)$, $\mathbb{E}[X|\mathcal{B}] =: \mathbb{E}[X|Z] = h(Z)$ for some h.

⚠ • $\mathbb{E}\left(\mathbb{E}[X|\mathcal{B}]\right) = \mathbb{E}(X)$

⚠ • $\left|\mathbb{E}[X|\mathcal{B}]\right| \leq \mathbb{E}(|X| \,|\, \mathcal{B})$

Pf: $X \leq |X| \Rightarrow \mathbb{E}(X|\mathcal{B}) \leq \mathbb{E}(|X|\,|\,\mathcal{B})$

$-X \le |X| \Rightarrow -\mathbb{E}(X|\mathcal{B}) \le \mathbb{E}(|X||\mathcal{B})$ $\square$

$$\boxed{\Rightarrow \underbrace{\mathbb{E}(|\mathbb{E}(X|\mathcal{B})|)}_{L^1\text{-norm of } \mathbb{E}(X|\mathcal{B})} \le \underbrace{\mathbb{E}|X|}_{} \quad L^1\text{-norm of } X}$$

⚠

- Let $X_n \nearrow X$ w/ $\mathbb{E}|X| < \infty$ and $\mathbb{E}|X_n| < \infty$ $\forall n$.

  Then, $\mathbb{E}(X_n|\mathcal{B}) \nearrow \mathbb{E}(X|\mathcal{B})$ as $n \to \infty$.

Pf: Subtract X_1 from X_n and X. Assume $0 \le X_n \nearrow X$

$\Rightarrow \mathbb{E}(X_n|\mathcal{B})$ increasing seq. as $n \to \infty$.

$\Rightarrow Y = \lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{B})$ exists and is measurable w.r.t. $\mathcal{B}$ b/c each of the $\mathbb{E}(X_n|\mathcal{B})$ are $\mathcal{B}$-measurable.

By MCT,

$$\mathbb{E}(Y) = \mathbb{E}\left(\lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{B})\right) \overset{\text{MCT}}{=} \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(X_n|\mathcal{B}))$$

$$= \lim_{n\to\infty} \mathbb{E}X_n$$

$$\overset{\text{MCT}}{=} \mathbb{E}\lim_{n\to\infty} X_n = \mathbb{E}X < \infty$$

$\Rightarrow \mathbb{E}Y < \infty$.

To check that Y is indeed a conditional expectation, need to check $\mathbb{E}Y\mathbb{1}_B = \mathbb{E}X\mathbb{1}_B$ $\forall B \in \mathcal{B}$. But just repeat equation above switching $Y \to Y \mathbb{1}_B$. $\Rightarrow Y = \mathbb{E}(X \mid \mathcal{B})$.

$\square$

⚠

- DCT holds for cond. exp.: Assume $X_n \to X$ as $n \to \infty$ and $\exists Y$ s.t. $|X_n| \le Y$ $\forall n$, and $\mathbb{E} Y < \infty$. Then
$$\lim_{n \to \infty} \mathbb{E}(X_n \mid \mathcal{B}) = \mathbb{E}(X \mid \mathcal{B}).$$

Pf: Same as the usual proof for DCT using MCT.

Define $h_n := \inf_{m \ge n} X_m$. Then $h_n \nearrow X$. Moreover, $|h_n| \le Y$ $\forall n$

So, $\mathbb{E}|h_n| < \infty$ $\forall n$. By MCT for conditional expectations,
$$\lim_{n \to \infty} \mathbb{E}(h_n \mid \mathcal{B}) = \mathbb{E}(X \mid \mathcal{B}).$$

On the other hand, since $h_n \le X_n$ $\forall n$,
$$\mathbb{E}(h_n \mid \mathcal{B}) \le \mathbb{E}(X_n \mid \mathcal{B}) \quad \forall n$$
$$\Rightarrow \liminf_{n \to \infty} \mathbb{E}(X_n \mid \mathcal{B}) \ge \mathbb{E}(X \mid \mathcal{B})$$

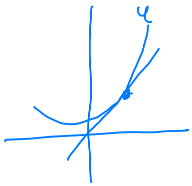
Same argument for $-X_n$ shows that
$$\limsup_{n \to \infty} \mathbb{E}(X_n \mid \mathcal{B}) \le \mathbb{E}(X \mid \mathcal{B}).$$

$\square$

- (JENSEN'S INEQUALITY) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex and s.t. $\mathbb{E}|\varphi(X)| < \infty$. Then

$$\varphi\big(\mathbb{E}(X|\mathcal{B})\big) \leq \mathbb{E}\big(\varphi(X)\,|\,\mathcal{B}\big) \quad \text{a.s.}$$

Pf: Can represent convex functions as:



$$\varphi(x) = \sup\{ax+b,\ a,b\in\mathbb{R} : ay+b \leq \varphi(y)\ \forall y\}$$
$$= \sup\{ax+b,\ a,b\in\mathbb{Q} : ay+b \leq \varphi(y)\ \forall y\}$$

Let $a,b\in\mathbb{Q}$ be s.t. $\varphi(x) \geq ax+b$ $\forall x\in\mathbb{R}$. Then $\varphi(X) \geq aX+b$,

i.e., $\varphi\big(\mathbb{E}(X|\mathcal{B})\big) \geq a\,\mathbb{E}(X|\mathcal{B}) + b$ a.s.

by linearity & monotonicity of conditional expectations.

Can assume this ineq. holds $\forall a,b\in\mathbb{Q}$ s.t. $\varphi(y) \geq ay+b$ $\forall y$.

Take sup on RHS over a.b.

$$\mathbb{E}\big(\varphi(X)\,|\,\mathcal{B}\big) \geq \sup_{a,b}\ a\,\mathbb{E}(X|\mathcal{B}) + b = \varphi\big(\mathbb{E}(X|\mathcal{B})\big).$$

$\square$

Jensen ineq. $\Rightarrow$ conditional expectations are contractions on L^p spaces

Ex: If $p\geq 1$, then $x \mapsto |x|^p$ is convex. So,

$$\big|\mathbb{E}(X|\mathcal{B})\big|^p \leq \mathbb{E}\big(|X|^p\,|\,\mathcal{B}\big)$$

Recall: $\|Y\|_p := \big(\mathbb{E}|Y|^p\big)^{1/p}$.

$$\boxed{\text{Thus: } \quad \| \mathbb{E}(X|\mathcal{B}) \|_p \leq \| X \|_p}$$

CONDITIONAL EXPECTATIONS & HILBERT SPACES: Take the Hilbert space $\mathcal{H} = L^2(\Omega, \mathcal{A}, \mathbb{P})$. This is real Hilbert space w.r.t. $\langle X, Y \rangle = \mathbb{E}(XY)$.

- Given $\mathcal{B} \subset \mathcal{A}$, define $V := L^2(\Omega, \mathcal{B}, \mathbb{P})$. This is a closed linear subspace of $\mathcal{H}$.
- Let P_V be the orthogonal projection onto V. Then

$$X = P_V X + (1 - P_V) X$$

b/c $\forall Y \in V$, $\langle Y, (1-P_V)X \rangle = 0$.

Claim: $P_V X = \mathbb{E}(X|\mathcal{B})$.

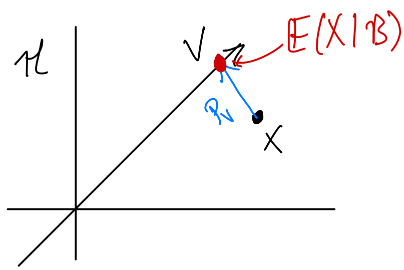
Pf: (i) Since $P_V X \in V$, $\sigma(P_V X) \subset \mathcal{B}$.

(ii) Take $Y = \mathbb{1}_B$ then

$$\begin{array}{ccc} \langle Y, X \rangle & = & \langle Y, P_V X \rangle \\ \| & & \| \\ \mathbb{E}(\mathbb{1}_B X) & & \mathbb{E}(\mathbb{1}_B P_V X) \end{array}$$

$\Rightarrow P_V X = \mathbb{E}(X|\mathcal{B})$. $\square$

Intuition:



LECTURE 20

Nov 22^{nd}, 2024

(i) $\mathbb{E}(X|\mathcal{B})$ is $\mathcal{B}$-measurable.

(ii) $\mathbb{E}(\mathbb{1}_B \mathbb{E}(X|\mathcal{B})) = \mathbb{E}(\mathbb{1}_B X) \ \forall B \in \mathcal{B}$.

!

Lemma: If X, Y are r.v.'s s.t. $\mathbb{E}|X| < \infty$, $\mathbb{E}|XY| < \infty$ and $\sigma(Y) \subset \mathcal{B}$, then

$$\mathbb{E}(XY|\mathcal{B}) = Y\,\mathbb{E}(X|\mathcal{B})$$

Pf: Do case $Y = \mathbb{1}_C$, $C \in \mathcal{B}$. Then

WTS: $\mathbb{1}_C \mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X\mathbb{1}_C|\mathcal{B})$

LHS clearly satisfies (i) above b/c prod. of $\mathcal{B}$-measurable

Check (ii): $\mathbb{E}(\mathbb{1}_{B\cap C}\, \mathbb{E}(X|\mathcal{B})) = \mathbb{E}(\mathbb{1}_{B\cap C} X)$

$$= \mathbb{E}(\mathbb{1}_B (\mathbb{1}_C X)).$$

So, LHS satisfies (ii).

$\Rightarrow$ By linearity, lemma holds if Y is a simple fct.

$\Rightarrow$ To conclude for general Y, assume $X \geq 0$, $Y \geq 0$.

Then, by linearity, take $Y_n \nearrow Y$ with Y_n simple.

Then, since both $X, Y \geq 0$, $XY_n \nearrow XY$.

$$\Rightarrow \mathbb{E}(XY \mid \mathcal{B}) \overset{\text{MCT}}{=} \lim_{n\to\infty} \mathbb{E}(XY_n \mid \mathcal{B})$$

$$\overset{\text{result for simple fcts?}}{=} \lim_{n\to\infty} Y_n \, \mathbb{E}(X \mid \mathcal{B})$$

$$= Y \, \mathbb{E}(X \mid \mathcal{B}).$$

$\square$

---

* **CONDITIONAL DISTRIBUTIONS**

Ex: Let X, Y be r.v.'s with density $f(x,y)\,dx\,dy$.

Then, if $g: \mathbb{R} \to \mathbb{R}$ s.t. $\mathbb{E}|g(X)| < \infty$ then

$$\mathbb{E}(g(X) \mid Y) = \frac{\int_{\mathbb{R}} g(x)\, f(x, Y(\omega))\, dx}{\int_{\mathbb{R}} f(x, Y(\omega))\, dx} \quad \text{a.s.}$$

denom $= 0$
w/ prob. $= 0$

In general, if X and Y are r.v.'s w/ density $f(x,y)\,dx\,dy$, then the density of Y is given by

$$p_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$$ "Marginal of Y"

$\Rightarrow$ Can write

$$\mathbb{E}(g(X)|Y) \overset{\text{claim}}{=} \frac{\int g(x) f(x,Y)\,dx}{p_Y(Y)} =: h(Y)$$

Pf: (Claim) $h(Y)$ is $\sigma(Y)$-measurable. Want to compute $\mathbb{E}(\mathbb{1}_B h(Y))$.

If $B \in \sigma(Y)$, then $B = \{Y \in R\}$, $R \in \mathcal{B}_{\mathbb{R}}$ (Borel)

So,

$$\mathbb{E}(\mathbb{1}_B h(Y)) = \mathbb{E}\left(\mathbb{1}_{\{Y\in R\}} h(Y)\right)$$

since $R \in \mathcal{B}_{\mathbb{R}}$

$$= \int \mathbb{1}_{\{y\in R\}} h(y)\, p_Y(y)\,dy$$

$$= \int \mathbb{1}_{\{y\in R\}} \cdot \frac{\int f(x,y) g(x)\,dx}{\cancel{p_Y(y)}} \cdot \cancel{p_Y(y)}\,dy$$

$$= \int \mathbb{1}_{\{y \in B\}} \, g(x) \, f(x,y) \, dx \, dy$$

$$= \mathbb{E}\left(g(X) \, \mathbb{1}_B\right).$$

$\square$

Rem: If we set $g(X) = \mathbb{1}_{\{X \in C\}}$ we get

$$\mathbb{E}\left(\mathbb{1}_{\{X \in C\}} \mid Y\right) = \frac{\int \mathbb{1}_{\{x \in C\}} f(x, Y) \, dx}{P_Y(Y)}$$

So, for a fixed value of Y, the map

$$C \longmapsto \frac{\int \mathbb{1}_{\{x \in C\}} f(x, Y) \, dx}{P_Y(Y)}$$

defines a measure on $\mathbb{R}$.

In this setting, this measure is called the conditional distribution of X w.r.t. Y. $\longleftarrow$ ⚠ Not always possible to define this measure

↙ Generalize

- In general, think of the map

$$C \mapsto \mathbb{E}(\mathbb{1}_{\{X \in C\}} \mid \mathcal{B}) =: \mathbb{P}(X \in C \mid \mathcal{B})$$

when $\mathcal{B}$ is some σ-algebra and X takes values in a general measure space.

If C_1 is <u>disjoint</u> from C_2, then, by additivity of measure,

$$\mathbb{P}(X \in C_1 \mid \mathcal{B}) + \mathbb{P}(X \in C_2 \mid \mathcal{B}) = \mathbb{P}(X \in C_1 \cup C_2 \mid \mathcal{B}).$$

This also holds for any countable disjoint union.

<u>PROBLEM</u>: Want to say: $\exists A \subset \Omega$ s.t. $\mathbb{P}(A) = 1$ and $\forall C$ in measure space where X takes values in,

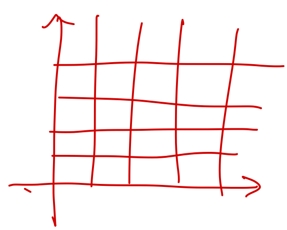
(i) $\mathbb{P}(X \in C \mid \mathcal{B})$ are defined $\forall \omega \in A$ in a way that

$$C \longmapsto \mathbb{P}(X \in C \mid \mathcal{B}) \text{ defines a measure}$$

ok if target(X) is finite or power set of $\mathbb{N}$

Can define horiz/vert. but can't always make sense of both coherently

$\iff$ Fubini failing to hold.



Def: (REGULAR CONDITIONAL DISTRIBUTION) "rdc" Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let X be a r.v. on $(\Omega, \mathcal{A})$ that takes values in $(S, \mathcal{M})$. Let $\mathcal{B} \subset \mathcal{A}$ be a sub-σ-alg..

Then a function

$$\mu(C, \omega) = \mathcal{M} \times \Omega \longrightarrow \mathbb{R}$$

is called a <u>regular</u> <u>conditional</u> <u>distribution</u> of X w.r.t. $\mathcal{B}$ if

(1) $\forall C \in \mathcal{M}$, $\omega \mapsto \mu(C, \omega)$ is a version of $\mathbb{P}(X \in C \mid \mathcal{B})$

(2) $\forall \omega \in \Omega$, $C \mapsto \mu(C, \omega)$ is a probability measure.

Thm: If $(S, \mathcal{M})$ is a separable and complete metric space with Borel σ-algebra, then regular conditional distribution exist and are unique (up to a.s.).

Ex: (Product Spaces) Let $\begin{cases} (\Omega, \mathcal{A}) = (\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}) \\ \mathbb{P} = \text{law of } (X, Y) \end{cases}$

The r.v. of definition will be X where $\mathbb{R}^2 \ni (x, y) \overset{X = \pi_1}{\longmapsto} x \in \mathbb{R}$ and $\mathcal{B}$ will be $\sigma(Y)$. Note $\sigma(Y) = \mathbb{R} \times \mathcal{B}_{\mathbb{R}}$.

Let $\mu(C,\omega)$ be the rcd of X w.r.t. $\mathcal{B}$. Then

$\omega = (x,y)$

$$\mu(C,\omega) = \mathbb{P}_y(C)$$

where $\mathbb{P}_y$ is a y-dependent prob. measure.

$\mathbb{P}_y(C)$ is the "probability of $X \in C$ conditioned on $Y = y$".

e.g.: $\mathbb{P}(X \in C,\ Y \in D) = \mathbb{E}(\mathbb{1}_{X\in C}\mathbb{1}_{Y \in D})$

$\quad \mathbb{P}(C \times D)$

$$= \mathbb{E}\left(\mathbb{E}(\mathbb{1}_{X\in C}\mathbb{1}_{Y\in D} \mid \mathcal{B})\right)$$

$\mathbb{1}_{Y\in D}$ is $\mathcal{B}$-measurable

$$= \mathbb{E}\left(\mathbb{1}_{Y\in D}\left(\mathbb{E}(\mathbb{1}_{X\in C}) / \mathcal{B}\right)\right)$$

$$= \mathbb{E}\left(\mathbb{1}_{Y\in D}\, \mathbb{P}_Y(C)\right)$$

$$= \int \mathbb{1}_{y\in D}\, \mathbb{P}_y(C)\, d\mu_Y(y)$$

$$= \int \mathbb{1}_{x\in C}\mathbb{1}_{y\in D}\, \underbrace{d\mu_{X\times Y}(x,y)}.$$

We're not assuming this is a product measure anymore, so Fubini might not hold!

In the product space case,

$$\boxed{\mathbb{E}(f(X)g(Y)) = \int\left(\int f(x)\, d\mathbb{P}_y(x)\right) g(y)\, d\mu_Y(y)}$$

Ex: If $\mu(C,\omega)$ is a rcd, then

$$\mathbb{E}(f(X) \mid \mathcal{B})(\omega) = \int f(x)\, d\mu(x,\omega) \quad \text{a.s.}$$

Ex: Y_i iid, N integer-valued r.v..

$$S := \sum_{i=1}^{N} Y_i .$$

Conditional distrib. of S w.r.t. N: if $N = n$, $S = \sum_{i=1}^{n} Y_i$ is the rdc.

Ex: Claim: Let X, Z be indep. $\mathrm{Exp}(1)$ r.v.'s

i.e., $\mathbb{P}(X \geq t) = \mathbb{P}(Z \geq t) = e^{-t} \quad \forall t \geq 0$.

Take $Y := X + Z$. The conditional distribution of X w.r.t. Y is $\mathrm{Unif}(0, Y)$.

Pf: First: compute joint distribution of X, Y:

density of X and Z is

$$e^{-x} e^{-z} \mathbb{1}_{\{x \geq 0\}} \mathbb{1}_{\{z \geq 0\}} \, dx\, dz$$

Change of variables: $x = x$, $y = x + z$ $\longrightarrow$ Jacobian $= \begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix}$

$\Rightarrow \det = 1$

$\Rightarrow$ Density of X and Y is

$$e^{-x} e^{-(y-x)} \mathbb{1}_{\{x \geq 0\}} \mathbb{1}_{\{y-x \geq 0\}} \, dx\,dy = e^{-y} \mathbb{1}_{\{0 \leq x \leq y\}} \, dx\,dy$$

From beginning of lecture, for two r.v.'s w/ density $f(x,y)$ the cond. distrib. of X w.r.t. Y is

$$C \longmapsto \frac{\int \mathbb{1}_{x \in C} f(x, Y)\, dx}{\int f(x, Y)\, dx}$$

Here, if Y is fixed, the density is $\propto e^{-y} \mathbb{1}_{\{0 \leq x \leq Y\}}$

$\propto \text{Unif}(0, Y)$.

CONDITIONAL DISTRIBUTIONS NOT in FINAL

(DISCRETE-TIME) MARTINGALES

Def: An increasing sequence of σ-algebras $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots$ is called a FILTRATION

index = time and $\mathcal{F}_3 \approx$ how much information we have at time $t=3$

Def: A seq. of r.v.'s $\{X_n\}_{n=1}^{\infty}$ is called ADAPTED (to filtration $\mathcal{F}_n$) if $\sigma(X_n) \subset \mathcal{F}_n$ $\forall n$. Call $\{X_n\}$ a PROCESS.

Def: A process $\{X_n\}_{n=1}^{\infty}$ is called a <u>MARTINGALE</u> w.r.t. some filtration $\mathcal{F}_n$ if

- X_n is adapted to $\mathcal{F}_n$
- $\mathbb{E}|X_n| < \infty \quad \forall n$
- $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n \quad \forall n$

Def: X_n is called a submartingale (resp. supermartingale) if the first two hold and the third holds w/ $\geq$ (resp. $\leq$).

LECTURE 21

Nov 27^{th}, 2024

Ex: (Martingale & Analysis) Let X_n be a simple random walk on $\mathbb{Z}$, i.e., $X_0 = 0$ and $X_n = \sum_{i=1}^{n} \xi_i$, where $\mathbb{P}(\xi_i = \pm 1) = \frac{1}{2}$.

If $h : \mathbb{Z} \to \mathbb{R}$, when is $h(X_n)$ a martingale? w.r.t. $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$

If martingale, then $\mathbb{E}(h(X_{n+1}) \mid \mathcal{F}_n) = h(X_n)$

$$\parallel$$

$$\frac{1}{2} h(1+X_n) + \frac{1}{2} h(-1+X_n) \stackrel{!}{=} h(X_n)$$

$$\Rightarrow \quad 0 = \frac{1}{2}\Big(h(x+1) + h(x-1) - 2h(x) \Big) \quad \forall x \in \mathbb{Z}$$

Discretization of $\left(\frac{d}{dx}\right)^2$

"discretized Laplacian"

Lemma: If X_n is a submartingale, then $\forall k \geq 1$,

$$\mathbb{E}(X_{n+k} \mid \mathcal{F}_n) \geq X_n .$$

Pf: Tower property

$$\mathbb{E}(X_{n+k} \mid \mathcal{F}_n) = \mathbb{E}\Big(\underbrace{\mathbb{E}(X_{n+k} \mid \mathcal{F}_{n+k-1})}_{\geq X_{n+k-1} \text{ b/c submartingale}} \Big| \mathcal{F}_n \Big)$$

Monotonicity of cond. exp.

$$\geq \mathbb{E}(X_{n+k-1} \mid \mathcal{F}_n)$$

iterate ...

$\square$

Ex: For
- submartingales, $\mathbb{E}X_n$ is increasing
- martingales, " " constant
- supermartingales, " " decreasing

Lemma: assuming $\mathbb{E}|\varphi(X_n)| < \infty \ \forall n$ ...

(1) If φ is a convex function, and X_n is a martingale, then $\varphi(X_n)$ is a submartingale.

(2) If φ is convex and increasing, and if X_n is a submartingale, then $\varphi(X_n)$ is a submartingale.

Pf: By Jensen,

$$\mathbb{E}\left(\varphi(X_{n+1}) \mid \mathcal{F}_n\right) \overset{\text{Jensen}}{\geq} \varphi\left(\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)\right)$$

- in (1), this is $= \varphi(X_n)$
- in (2), this is $\geq \varphi(X_n)$ b/c φ is increasing and $X_n \leq \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)$

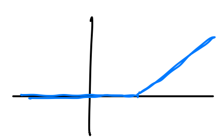
$\blacksquare$

USEFUL

Ex:

- If X_n is a martingale, then $|X_n|$ is a submartingale
- If X_n is a submartingale, then so is $(X_n - a)_+$, $a \in \mathbb{R}$ constant, since $(X_n - a)_+$ is convex.

$x \longmapsto (x-a)_+$



Def: • We say a sequence $\{H_n\}_{n \geq 1}$ is <u>predictable</u> if $\sigma(H_n) \subset \mathcal{F}_{n-1}$ (i.e., can figure out the entirely from information already available before)

• Given a predictable H, the <u>martingale transform</u> is defined by

$$(H \cdot X)_n := \sum_{i=1}^{n} H_i (X_i - X_{i-1}) .$$

X_n = winnings at time n

$X_n - X_{n-1}$ = winnings at n^{th} play of the game $\longrightarrow$ $X_n = \sum_{i=1}^{n} (X_i - X_{i-1}) + \underbrace{X_0}_{\text{amount you started with}}$

<u>Lemma</u>: (1) If H_n is predictable, $H_n \geq 0$ and bounded $\forall n$ (i.e., $\exists C_n$ s.t. $H_n \leq C_n$ $\forall n$), and if X_n is a submartingale, then $(H \cdot X)_n$ is a submartingale.

(2) Same thing holds if X_n is a supermartingale, here $(H \cdot X)_n$ will be a supermartingale

(3) If X_n is a martingale, so is $(H \cdot X)_n$ and H can take any sign in this case.

Pf: $\mathbb{E}((H \cdot X)_{n+1} | \mathcal{F}_n) = \mathbb{E}\Big(\underbrace{(H \cdot X)_n}_{\mathcal{F}_n\text{-measurable}} + H_{n+1}(X_{n+1} - X_n) \Big| \mathcal{F}_n\Big)$

$\mathbb{E}(XY|\mathcal{F}) = Y\mathbb{E}(X|\mathcal{F})$ if Y is $\mathcal{F}$-measurable

$= (H \cdot X)_n + \underbrace{H_{n+1}}_{\geq 0} \underbrace{\mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n)}_{\geq 0 \text{ b/c } X_n \text{ submartingale}}$

$$\geq (H \cdot X)_n .$$

Similar for other cases. $\blacksquare$

Bad b/c going bankrupt

Ex: (MARTINGALE BETTING STRATEGY)

$$X_n = \sum_{i=1}^{n} \xi_i \quad , \quad \xi_i \text{ ind s.t. } \mathbb{P}(\xi_i = \pm 1) = \frac{1}{2}$$

$$H_n = \begin{cases} 1, & \text{if } \xi_{n-1} = 1 \\ 2H_{n-1}, & \text{if } \xi_{n-1} = -1 \end{cases} ,$$

H_n fulfills the criteria of Lemma above

$$\Longrightarrow \mathbb{E}(H \cdot X)_n = 0 .$$

$\blacksquare$

Def: A random variable $N \in \mathbb{N} \cup \{0\}$ is a <u>stopping time</u> if $\forall n$, $\{N = n\} \in \mathcal{F}_n$.

(i.e., not allowed to use info. from time 6 to decide if you stop at time 5
i.e., can use only past information to decide whether to stop playing or not)

Properties: (1) If N is a stopping time, then

$$\bigcup_{j=0}^{n} \{N = j\} = \{N \leq n\} \in \mathcal{F}_n$$

(2) $\{N > n\} \in \mathcal{F}_n$ b/c $\{N > n\} = (\{N \leq n\})^c$.

Thm: If X_n is a submartingale and N is a stopping time, then

$$Y_n := X_{\min\{n, N\}}$$

is also a submartingale.

LECTURE 22

Def: A r.v. $N \in \mathbb{Z}_{\geq 0} \cup \{+\infty\}$ is a stopping time if $\{N = n\} \in \mathcal{F}_n$ for $0 \leq n < \infty$.

Ex: $\{N \leq n\} \in \mathcal{F}_n$

$$\{N \leq n\} = \bigcup_{j=0}^{n} \{N = j\}, \qquad \{N = j\} \in \mathcal{F}_j \subset \mathcal{F}_n$$

Ex: If $a \in \mathbb{R}$, then $N = \inf\{m : X_m \geq a\}$ is a stopping time if X_m is adapted

$$\{N = m\} = \{X_m \geq a\} \cap \bigcap_{j=0}^{m-1} \{X_j < a\} \in \mathcal{F}_j \subset \mathcal{F}_m$$

$$\{X_m \geq a\} \in \mathcal{F}_m$$

Lemma: If N is a stopping time, and X_n is a submartingale, then $Y_n := X_{\min\{n,N\}}$ is a submartingale.

$n \wedge N := \min\{n, N\}$

Pf:
$$X_{n\wedge N} = \sum_{j=1}^{n\wedge N} (X_j - X_{j-1}) + X_0$$
$$= \sum_{j=1}^{n} \mathbb{1}_{\{n \le N\}} (X_j - X_{j-1}) + X_0$$

In order to conclude, need that
$$H_n = \mathbb{1}_{\{n \le N\}} \text{ is predictable.}$$
i.e., $\{N \ge n\} \in \mathcal{F}_n$. But this is true b/c
$$\{N \ge n\} = \{N < n-1\}^c \quad \underline{\text{Cl}}: \{N < n-1\}^c \in \mathcal{F}_{n-1}$$
Since N is stopping time, $\{N < n-1\} \in \mathcal{F}_{n-1}$ by def.

$\square$

Thm: (MARTINGALE CONVERGENCE) Let X_n be a submartingale s.t. $\underline{\sup_n \mathbb{E}(X_n)_+ < \infty}$, then

sort of like assuming increasing seq. is bdd above. Submartingales tend to go up.

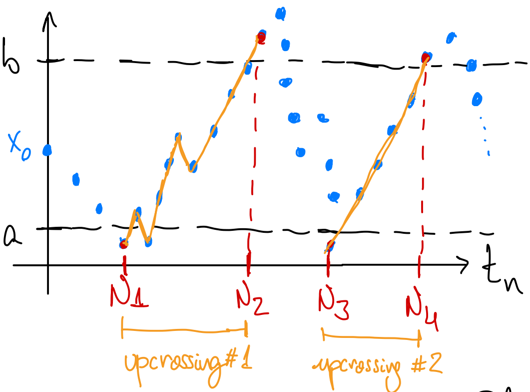
$$X_\infty := \lim_{n\to\infty} X_n \text{ a.s.,}$$

(i.e., X_n converge a.s. to some r.v. X_∞)

and $\mathbb{E}|X_\infty| < \infty$.

Need a few lemmas:

UPCROSSING INEQUALITY: Define the stopping times



Stopping times when X_n crosses out of (a,b) (only "out crossings")

$N_0 := -1$

"first time after N_{2k-2} s.t. $X_m \le a$"

$N_{2k-1} := \inf\{m > N_{2k-2} : X_m \le a\}$

$N_{2k} := \inf\{m > N_{2k-1} : X_m \ge b\}$

Cl: These are stopping times

Pf: $\{N_{2k-1} = m\} = \bigcup_{j=0}^{m-1} \{N_{2k-2} = j\}$ ($\mathcal{F}_j \subset \mathcal{F}_m$)

$\cap \{X_\ell > a \ \forall j < \ell < m\}$ ($\in \mathcal{F}_m$)

$\cap \{X_m \le a\} \in \mathcal{F}_m$

Let $U_n :=$ # of upcrossings before time n

$= \sup\{k \ge n \text{ s.t. } N_{2k} \le n\}$

Thm: (UPCROSSING INEQUALITY) Let X_n be a submartingale and let $a<b$ and U_n as above. Then

$$\mathbb{E}U_n \le \frac{1}{b-a}\left[\mathbb{E}(X_n-a)_+ - \mathbb{E}(X_0-a)_+\right]$$

Pf: Let $Y_n = (X_n-a)_+ + a$. Then Y_n is also a submartingale.

Let H_n be the process

$$H_n = \begin{cases} 1, & \text{if } N_{2k-1} < n \le N_{2k} \text{ for some } k \\ 0, & \text{else} \end{cases},$$

Then, the martingale transform is:

$$(H\cdot Y)_n = \sum_{j=1}^{n} H_j (Y_j - Y_{j-1}).$$

Cl: H_n is predictable.

Indeed, $\{H_n = 1\} = \bigsqcup_k \underbrace{\{N_{2k-1} = n-1\}}_{\in\, \mathcal{F}_{n-1}} \cap \underbrace{\{N_{2k} \le n-1\}^c}_{\in\, \mathcal{F}_{n-1}}$

$\Rightarrow$ H_n is predictable.

$\Rightarrow$ $(H\cdot Y)_n$ is also a submartingale.

Cl: $(b-a)\,U_n \le (H\cdot Y)_n$

But $(H\cdot Y)_n = \underbrace{(Y_{N_2} - Y_{N_1})}_{\ge b-a} + \underbrace{(Y_{N_4} - Y_{N_3})}_{\ge b-a}$

$+ \underbrace{(Y_{N_{2U_n}} - Y_{N_{2U_n-1}})}_{\ge b-a} + \underbrace{(Y_{2n} - Y_{N_{2U_{n+1}}})}_{\ge 0}$

$\ge (b-a)\,U_n.$

Upshot: $(b-a)\,\mathbb{E}U_n \overset{\text{taking } \mathbb{E}}{\le} \mathbb{E}(H\cdot Y)_n.$

But $\mathbb{E}(Y_n - Y_0) = \mathbb{E}(1\cdot Y)_n = \mathbb{E}(H\cdot Y)_n + \underbrace{\mathbb{E}((1-H)\cdot Y)_n}_{\text{submartingale}}$

$$\geq \mathbb{E}(H\cdot Y)_n + \mathbb{E}((1-H)\cdot Y)_0$$
$$= \mathbb{E}(H\cdot Y)_n \geq (b-a)\,\mathbb{E}(U_n).$$

Lt

$$(b-a)\,\mathbb{E}U_n \leq \mathbb{E}Y_n - Y_0 = \mathbb{E}(X_n-a)_+ - \mathbb{E}(X_0-a)_+ .$$

$\blacksquare$

Pf: (MARTINGALE CONVERGENCE) We know that $\forall a<b$,

$$\mathbb{E}\,U_n \leq \frac{1}{b-a}\,\mathbb{E}(X_n-a)_+ \leq \frac{1}{b-a}\left(\mathbb{E}(X_n)_+ + |a|\right)$$
$$\leq C \quad \forall n.$$

$\Rightarrow$ By MCT, # of upcrossings of (a,b) is finite a.s..

$$\Rightarrow \mathbb{P}\left(\liminf_n X_n < a < b < \limsup_n X_n\right) = 0$$

$$\Rightarrow \mathbb{P}\left(\bigcup_{\substack{a<b\\ a,b\in\mathbb{Q}}} \left\{\liminf_n X_n < a < b < \limsup_n X_n\right\}\right) = 0$$

$$\Rightarrow \limsup_n X_n = \liminf_n X_n \ \text{ a.s.} \Rightarrow X_\infty \text{ exists}.$$

<u>NTC</u>: $X_\infty < \infty$.

By Fatou, $$\infty > \liminf_{n\to\infty} \mathbb{E}(X_n)_+ \overset{\text{Fatou}}{\geq} \mathbb{E}\liminf_{n\to\infty}(X_n)_+ = \mathbb{E}(X_\infty)_+$$

Can do the same for negative part of X_n:

$$\mathbb{E}(X_n)_- = -\mathbb{E}X_n + \mathbb{E}(X_n)_+ \leq -\mathbb{E}X_0 + \sup_n (X_n)_+$$

USEFUL TRICK FOR SUBMARTINGALES

$$\leq C < \infty.$$

$$\Rightarrow \mathbb{E}(X_\infty)_- < \infty$$

Thus, $\mathbb{E}|X_\infty| < \infty$.

■

Cor: If X_n is a non-negative supermartingale, then $X_\infty = \lim_{n\to\infty} X_n$ a.s., and $\mathbb{E}|X_\infty| < \infty$.

Ex: $X_n = 1 + \sum_{j=1}^{n} \xi_j$, $\mathbb{P}(\xi_j = \pm 1) = \frac{1}{2}$, ξ_j iid.

$N := \inf\{m > 0 : X_m = 0\}$ $\rightsquigarrow$
- $N < \infty$ a.s.
- N is a stopping time

Let $Y_n := X_{n\wedge N}$. Then Y_n is a martingale.

$$\lim_{n\to\infty} Y_n = Y_\infty \text{ exists a.s.}$$

$$= X_N = 0$$

$\Rightarrow$ The convergence does not occur in L^1 in general, $\mathbb{E}Y_0 = 1$.

Thm: (Doob's decomposition for submartingales) If X_n is a submartingale, then we can write

$$X_n = M_n + A_n,$$

where M_n is a martingale and A_n is increasing and predictable.

Pf: (Construction)

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

So, set $\begin{cases} A_n - A_{n-1} := \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} \\ A_0 := 0 \end{cases}$

Note: this implies A_n is increasing and predictable. So, if we define $M_n := X_n - A_n$, we get that M_n is a martingale.

$\square$

LECTURE 23

* <u>HITTING PROBABILITIES OF RANDOM WALKS</u>

$X_n = \sum_{j=1}^{n} \xi_j$ simple random walk $\quad \mathbb{P}(\xi_j = \pm 1) = \frac{1}{2}$

ξ_j iid.

$\Rightarrow X_n$ is a martingale.

- For $x \in \mathbb{Z}$, $T_x := \inf \{n : X_n = x\}$ is a stopping time.
- From Midterm Q1, we know that $T_x < \infty$ a.s.
- For $T := T_a \wedge T_b$, $a < 0 < b$, T is a stopping time.
- Use T to compute $\mathbb{P}(T_a < T_b) = 1 - \mathbb{P}(T_b < T_a)$.

Let $Y_n := X_{n \wedge T}$. Then Y_n is a martingale (Optimal Stopping Theorem)

So, $\forall n$, $\mathbb{E} Y_n = \mathbb{E} Y_0 = \mathbb{E} X_0 = 0$.

$$\mathbb{E} Y_n \overset{\text{def}}{=} \mathbb{E}(X_{n \wedge T}) \qquad \Rightarrow \mathbb{E}(X_{n \wedge T}) = 0$$

Want to take $n \to \infty$ but it's not always possible to put this limit inside $\mathbb{E}$ (argue case-by-case)!

Here we can do this b/c since $T < \infty$ a.s.

$$\lim_{n \to \infty} T_{n \wedge T} = X_T. \quad \forall n \ |X_{n \wedge T}| \le |a| + b$$

By DCT, $\lim\limits_{n\to\infty} \mathbb{E}(X_{n\wedge T}) = \mathbb{E}\left(\lim\limits_{n\to\infty} X_{n\wedge T}\right)$

$$= \mathbb{E} X_T = 0.$$

$$\Rightarrow \quad 0 = \mathbb{E}X_T = a\,\mathbb{P}(T_a < T_b) + b\,\mathbb{P}(T_b < T_a)$$

$$= ap + b(1-p)$$

$$\Rightarrow \quad p = \frac{b}{b-a} = \mathbb{P}(T_a < T_b) \quad \text{(i.e., prob. of hitting } a \text{ first)}$$

$$\text{and} \quad \mathbb{P}(T_b < T_a) = -\frac{a}{b-a}.$$

Note: if $a, b > 0$, $\mathbb{E}T_n = +\infty$ (bad).

Note: $Z_n := X_n^2 - n$ is a martingale

$$\Rightarrow \quad Z_{T\wedge n} \text{ is a martingale}$$

$$\overset{*}{\Rightarrow} \quad \mathbb{E}_{T\wedge n} = \mathbb{E}Z_{T\wedge 0} = 0$$

$$\text{i.e.,} \quad \mathbb{E}X^2_{T\wedge n} = \mathbb{E}(T\wedge n)$$

$$\left.\begin{array}{c}\downarrow\\ \mathbb{E}X_T^2\end{array}\right. \qquad \left.\begin{array}{c}\downarrow \text{ MCT}\\ \mathbb{E}T\end{array}\right.$$

DCT b/c $X^2_{T\wedge n}$ is bdd as before

Thus: $\mathbb{E}T = \mathbb{E}X_T^2 = a^2 \frac{b}{b-a} + b^2 \frac{(-a)}{b-a} = -ab.$

$\blacksquare$

Lemma: Let N be a bounded stopping time, i.e., $N \le k$ a.s. $k \in \mathbb{Z}_{\ge 0}$. Let X_n be a submartingale. Then we have

$$\mathbb{E}X_0 \overset{1}{\le} \mathbb{E}X_N \overset{2}{\le} \mathbb{E}X_k$$

Pf: 1 Since $X_{N \wedge n}$ is a submartingale

$$\begin{array}{c} \mathbb{E}X_{0 \wedge N} \le \mathbb{E}X_{N \wedge k} = \mathbb{E}X_N \\ \| \\ \mathbb{E}X_0 \end{array}$$

2 $$\mathbb{E}X_k - \mathbb{E}X_N = \mathbb{E}\left(\sum_{j=N+1}^{k} X_j - X_{j-1}\right)$$

$$= \mathbb{E}\left(\sum_{j=1}^{k} \mathbb{1}_{\{j > N+1\}}(X_j - X_{j-1})\right)$$

$$= \mathbb{E}(H \cdot X)_k \ge \mathbb{E}(H \cdot X)_0 = 0$$

claim: $H_j = \mathbb{1}_{\{j > N+1\}}$ is predictable

NTC: $\{j \ge N+1\} \in \mathcal{F}_{j-1}$

$$\begin{array}{c} \{j \ge N+1\} \\ \| \\ \{j < N-1\} \end{array}$$

$\Rightarrow \mathbb{E}X_k \ge \mathbb{E}X_N.$ $\blacksquare$

Def: If N is a stopping time, the σ-algebra associated to N is denoted by $\mathcal{F}_N$ and is defined as

$$\mathcal{F}_N := \{ A \in \Omega : \underline{A \cap \{N = n\} \in \mathcal{F}_n} \ \forall n \geq 0 \} .$$

"if N were to be 5, we would have" all information for that in $\mathcal{F}_5$.

Exercise: 1. $\mathcal{F}_N$ is a σ-algebra

2. If X_n is an adapted process, then X_N is $\mathcal{F}_N$-measurable.

Thm: (OPTIONAL STOPPING) Let $S \leq T$ be bounded stopping times and X_n be a submartingale. Then

$$\mathbb{E}(X_T \mid \mathcal{F}_S) \geq X_S \quad \text{a.s.}$$

(Just like the definition of submartingale but now with stopping times, instead of deterministic time)

Pf: Let $n \geq T \geq S \geq 0$ (b/c stopping times are bdd), $n \in \mathbb{Z}_{\geq 0}$.

Cl: $\mathbb{E}(X_T) \geq \mathbb{E}(X_S)$.

Let $Y_n = X_{n \wedge T}$. Then Y_n is a submartingale. By lemma "stopped process" above w/ $n=S$, $\mathbb{E}Y_S \le \mathbb{E}Y_k$

$$\begin{array}{ccc} \mathbb{E}Y_S & \le & \mathbb{E}Y_k \\ \text{def}\,\| & & \|\,\text{def} \\ \mathbb{E}X_{S\wedge T} & & \mathbb{E}_{T\wedge k} \\ S\le T \longrightarrow \| & & \| \longleftarrow T\le k \\ \mathbb{E}X_S & & \mathbb{E}_T \end{array}$$

So, for the thm, it suffices to check that $\forall A\in\mathcal{F}_S$, we have $\mathbb{E}(\mathbb{1}_A X_T) \ge \mathbb{E}(\mathbb{1}_A X_S)$.

$$\mathbb{E}(\mathbb{1}_A X_T) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X_T \mid \mathcal{F}_S))$$

b/c then $0 \le \mathbb{E}(\mathbb{1}_A(\mathbb{E}(X_T\mid\mathcal{F}_S) - X_S))$ and, in general, if $0 \le \mathbb{E}(\mathbb{1}_B Z)$ then we have $Z\ge 0$ a.s.

Let $A\in\mathcal{F}_S$ and set

Common technique

$$N := \begin{cases} S & \text{on } A \\ T & \text{on } A^c \end{cases}$$

Check N is stopping time later

Since $\mathbb{E}X_T \ge \mathbb{E}X_S$, we have $\mathbb{E}X_N \le \mathbb{E}X_T$ b/c N is a stopping time that is always $\le T$. But note:

$$\mathbb{E}X_N \le \mathbb{E}X_T = \mathbb{E}(\mathbb{1}_A X_T) + \cancel{\mathbb{E}(\mathbb{1}_{A^c} X_T)}$$

$$\| $$

$$\mathbb{E}(\mathbb{1}_A X_S) + \cancel{\mathbb{E}(\mathbb{1}_{A^c} X_T)} \qquad \Longrightarrow \mathbb{E}\,\mathbb{1}_A X_S \le \mathbb{E}\,\mathbb{1}_A X_T$$

as desired.

Cl: N is stopping time

$\Updownarrow$

Cl: $\{N=n\} \in \mathcal{F}_n \ \forall n.$

But $\{N=n\} = \underbrace{\{N=n\}\cap A}_{\substack{\| \\ \{S=n\}\cap A \\ \in \\ \mathcal{F}_n}} \sqcup \underbrace{\{N=n\}\cap A^c}_{\| }$

by def. of $\mathcal{F}_S$

$$\begin{aligned}
\{N=n\}\cap A^c &= \{T=n\}\cap A^c \\
&= \{T=n\}\cap\left(\{S\le n\}\cap A^c\right) \\
&= \{T=n\}\cap\left(\bigcup_{j=0}^{n} \underbrace{\{S\le j\}\cap A^c}_{\in\ \mathcal{F}_j \subset \mathcal{F}_n}\right)
\end{aligned}$$

∎