Probability-3 Lecture-1

22 July 2024 10:12

Instructor: Prof. Alok Goswami

Office hours: Tuesday 4pm – 5:30 pm
Venue: Associate Dean's Office. (5th floor, S.N. Bose Bhavan)

Recall: (Qn from midsem of Sem–2)

$X \in \{0, 1, 2, \dots\}$.

$p_n = P(X = n)$.

$P(s) = \sum_{n=0}^{\infty} s^n \cdot p_n$

$P'(s) = \sum_{n=1}^{\infty} n s^{n-1} \cdot p_n$.

$s \in (-1, 1)$.

Q. Show that,

$\lim_{s \uparrow 1} P'(s)$ exists & is finite — ①

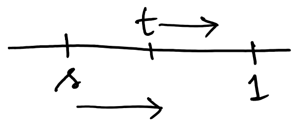
iff

$\lim_{s \uparrow 1} \dfrac{P(s) - 1}{s - 1}$ exists. — ②.

and is finite & equal.

then, show that ① $\Leftrightarrow$ ② $\Leftrightarrow$ $E(X)$ is finite & equals to the limit in ① or ②.

Soln. for $0 < s < 1$..



$$\frac{P(s) - 1}{s - 1} = P'(t) \quad \text{for } s < t < 1.$$

If ① is assumed, then $\lim_{t \uparrow 1} P'(t)$ exists & finite.

But, converse is NOT true!!

correction –

$P'(t)$ is increasing in t.

So, $\lim_{t \uparrow 1} P'(t)$ must exist.

$\therefore$ We have one such seq. $\{t_n\}$ given by

$\dfrac{P(s) - 1}{s - 1} = P'(t_n)$. $\therefore \lim_{s \uparrow 1} \dfrac{P(s) - 1}{s - 1}$ exists.

For each n, define $X_k = X \cdot (1 - 1/k)^{X-1}$

Now, what happens when $k \uparrow$?

X_k increases to X.

$\therefore X_k \geqslant 0$, $X_k \uparrow X$.

[... MCT]

$\therefore X_k \geq 0$, $X_k \uparrow X$.

$\therefore E(X_k) \uparrow E(X)$. [By MCT]

$\therefore \lim_{s \uparrow 1} P'(s) = E(X)$.

Revision of Sem-1 & Sem-2:

Random Variable:

$(\Omega, \mathcal{A}, P)$ ← Probability assignment over $\mathcal{A}$.

Ω: Sample space

$\mathcal{A}$: σ-field. "Class of events"

$X : \Omega \to \mathbb{R}$ (or $[-\infty, \infty]$) is a r.v.
if $\forall\, a \in \mathbb{R}$, $\{\omega : X(\omega) \leq a\} \in \mathcal{A}$.

This is equivalent to $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ for all borel sets $B \subset \mathbb{R}$.

Given any real r.v. X,
probability distribution of X, $P_X(B) = P(X^{-1}(B))$ $\forall$ borel sets $B \subset \mathbb{R}$

i.e., X acts like the "carrier" of probability mass.

$F_X : \mathbb{R} \to \mathbb{R}$.

$F_X(a) = P_X((-\infty, a])$

$= P(X \leq a)$, $a \in \mathbb{R}$.

$\Delta F_X(a) = P_X(\{a\}) = P(X = a)$.

Properties:

- Right continuous
- Non-decreasing
- $F_X(a) \to \begin{cases} 0, & a \to -\infty \\ 1, & a \to \infty \end{cases}$

The distribution function completely defines the probability assignment.

$$F_X \xleftrightarrow[\text{correspondence}]{1-1} P_X$$

Special types:

1. X - discrete.

$\exists$ countable $D \subset \mathbb{R}$ such that $P_X(D) = 1$.

① X - discrete.

$\exists$ countable $D \subset \mathbb{R}$, such that $P_X(D) = 1$.

$$p_X(x) = P(X = x), \; x \in \mathbb{R}. \quad \text{pmf}.$$

$$p_X(x) = 0, \; x \notin D. \; \& \; \forall x \in D, \; p_X(x) > 0.$$

&, $$P_X(B) = \sum_{x \in B \cap D} p_X(x).$$

$$F_X(a) = \sum_{x \leq a, x \in D} p_X(a)$$

② X (absolutely*) continuous.

if $\exists f \geq 0$ on $\mathbb{R}$.

s.t. $$P_X(B) = \int_B f(x)\,dx$$

$$F_X(a) = \int_{(-\infty, a]} f(x)\,dx = \int_{-\infty}^{a} f(x)\,dx.$$

* absolutely cts = has probability density.

[Note that, the "$\int$" need not always be a Riemann Integration.]

Say, X has density $f(x)$, $x \in I$ (open interval.)

take $Y := h(X)$.

Special case: $h: I \longrightarrow J$ is 1-1, onto.

$$g = h^{-1}: J \longrightarrow I.$$

Assume: g is continuously differentiable.

Then, $Y = h(X)$ has density:

$$f_Y(y) = f(g(y)) \cdot |g'(y)|.$$

In general, for $Y = h(X)$, try to find

$$F_Y(b) = P(h(X) \leq b) = \int_{x: h(x) \leq b} f(x)\,dx.$$

$x : h(x) \leq b$

try & see if
$F_Y(y)$ is onto.

Expected Value:

Step-1: Non-ve, simple, real r.v X.

(can be written in the form:

$$X = \sum_{i=1}^{n} c_i \cdot 1_{A_i}, \quad A_1, \dots, A_n \text{ is a partition of } \Omega.)$$

$$E(X) = \sum_{i=1}^{n} c_i \cdot P(A_i)$$

Step-2:

$X \geq 0$ extended real values.

$$E(X) = \sup \{ E(Y) : Y \text{ is a simple real non-ve r.v}, \ Y \leq X \}$$

Result:

For any sequence $\{X_n\}$ of non-ve real simple r.vs with $X_n \uparrow X$,

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

Step-3:

For general X, $E(X) = E(X^+) - E(X^-)$ if RHS is defined.

Properties of $E(X)$:

Expectation is a linear, order preserving operator.

Inequalities: Chebyshev, Holder's, Minkowski's, Jensen.

M.G.F:

X - r.v.

X - r.v.

$m_X(t) = E(e^{tX})$ if RHS is finite.

Holder's inequality gives -

The set $I = \{t : E(e^{tX}) < \infty\}$ is an interval with $0 \in I$.

Eg: $I = \{0\}$ for Cauchy distn.

$$f_X(x) = \frac{c}{1+x^2}, \quad x \in \mathbb{R}, \quad c \text{ - normalizing constant.}$$

Incase 0 is an interior point of I, then MGF characterizes (or determines) the distribution of X. ie, $m_X(t)$ determines F_X.

$$m_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)}{\underline{\lfloor n}} \cdot t^n \quad \text{for } t \in (-\varepsilon, \varepsilon).$$

radius of convergence

Probability-3 Lecture-2

26 July 2024 02:08

$(\Omega, \mathcal{A}, P)$ - a probability space.

$\underset{\sim}{X} = (X_1, \ldots, X_n)$ is a random vector.

$\Leftrightarrow X_1, \ldots, X_n$ are real random variables.

$$\underset{\sim}{X} : \Omega \longrightarrow \mathbb{R}^d$$

Joint probability distribution $\underset{\sim}{X}$:

$$P_{\underset{\sim}{X}}(B) = P(X \in B), \quad B \subset \mathbb{R}^k$$

$$F_{\underset{\sim}{X}}(\underset{\sim}{a}) = P(X_1 \leq a_1, \ldots, X_k \leq a_k),$$

$$\underset{\sim}{a} = (a_1, \ldots, a_k)$$

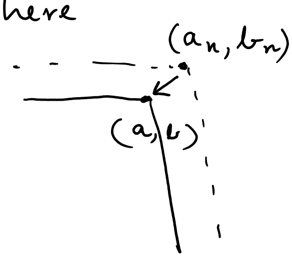
$$\therefore F_{\underset{\sim}{X}} : \mathbb{R}^k \longrightarrow \mathbb{R}.$$

Eg: $k=2$.

$*$ F is "right-continuous" everywhere

[ie, $a_n \searrow a$, $b_n \searrow b$, then

$$F(a_n, b_n) \longrightarrow F(a, b).$$



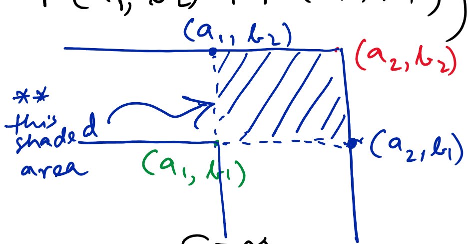
$**$ $\Delta F_{(a_1, b_1), (a_2, b_2)} = \Big(F(a_2, b_2) - F(a_2, b_1) - F(a_1, b_2) + F(a_1, b_1) \Big)$

(Stronger than "non-decreasing" property)

$a_1 \leq a_2, \quad b_1 \leq b_2$

$$\geq 0$$

$$\forall (a_1, b_1), (a_2, b_2).$$



$*$ $F(a, b) \longrightarrow \begin{cases} 0 \\ 1 \end{cases}$ as $\underbrace{a \wedge b}_{\min\{a, b\}} \longrightarrow \begin{cases} -\infty \\ \infty \end{cases}$

$a \wedge b \longrightarrow -\infty$

$\Rightarrow$ <u>atleast one</u> of $a, b \rightarrow -\infty$

& , $a \wedge b \longrightarrow \infty$

$\Rightarrow$ atleast one of a, b

$\&$, $a \wedge b \longrightarrow \infty$

$\Rightarrow$ both $a, b \longrightarrow \infty$:

Important: In 2 or higher dimensions,

F is continuous everywhere

$\Updownarrow$

$P(X=a) = P(Y=b) = 0 \ \forall\, a, b \in \mathbb{R}$.

or, F has uncountably many discontinuities.

Marginals:

$$F_X(a) = \lim_{b \to \infty} F(a, b).$$

$$F_Y(b) = \lim_{a \to \infty} F(a, b).$$

Special cases:

① (X, Y) - discrete if $\exists$ countable $D \subset \mathbb{R}^2$

s.t. $P((X,Y) \in D) = 1$

$p(x, y) = P(X = x, Y = y)$, $(x, y) \in \mathbb{R}^2$; joint pmf

$p(x,y) = 0$ if $(x,y) \notin D$ $\rightarrow$ [ensures that $p(x,y) > 0 \ \forall (x,y) \in D$.]

$$P_{X,Y}(B) = \sum_{(x,y) \in D \cap B} p(x,y)$$

$$p_X(x) = \sum_y p(x,y), \quad p_Y(y) = \sum_x p(x,y) \quad \leftarrow \text{Marginals}$$

② (X, Y) - jointly (absolutely) continuous if $\exists f \geq 0$ on $\mathbb{R}^2$

s.t. $P_{X,Y}(B) = \iint_B f(x,y)\, dx\, dy$, $B \subset \mathbb{R}^2$ (borel)

$$F_{X,Y}(a, b) := \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\, dx\, dy = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y)\, dy\, dx$$

[Fubini's theorem — gives the conditions

[Fubini's theorem
— gives the conditions
for interchanging
the "$\int$"s or "$\sum$"s.]

$$F_X(a) = \int_{-\infty}^{\infty}\int_{-\infty}^{a} f(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{a}\Big(\underbrace{\int_{-\infty}^{\infty} f(x,y)\,dy}_{=\, f_X(x)}\Big)dx \quad = \int_{-\infty}^{a}\Big(\qquad\Big)\cdot dx$$

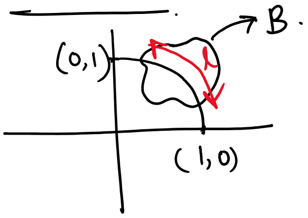
$\dashrightarrow f^n$ of X.

$f_X(x) \leftarrow$ density of X. (why? ⤴)

ie, "integrating out" y gives density of X.

* X, Y absolutely continuous $\not\Rightarrow$ (X,Y) jointly (abs) continuous.

Picture: 1

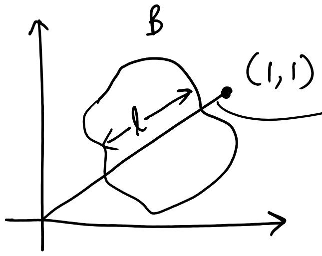


Arc from unit circle. $0 \leq x \leq 1$, $0 \leq y \leq 1$.

ie, entire probability mass lies on the arc.

$$P_{X,Y}(B) = \frac{\text{Arc length}(l)}{\pi/2}$$

Picture 2:



$\rightarrow$ entire probability mass distributed on this line.

Here, $P_{X,Y}(B) = \left(\frac{l}{\sqrt{2}}\right)$

(show that, both the marginals are Unif(0,1), but they do not have a joint density.)

Special result:

(X,Y) has a density f on $I \hookrightarrow$ open region in $\mathbb{R}^2$.

(X,Y) has a density f ... $\hookrightarrow$ open region in $\mathbb{R}^2$.

$h: I \to \tilde{I}$ (an open region in $\mathbb{R}^2$.)
(bijection)

$g = h^{-1}: \tilde{I} \to I$.

$h: (x,y) \longmapsto (u,v)$

$g: (u,v) \longmapsto (x,y)$

$\frac{\partial x}{\partial u}, \frac{\partial x}{\partial v}, \frac{\partial y}{\partial u}, \frac{\partial y}{\partial v}$ exist and are continuous in I.

The Jacobian, $J(u,v) = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$

Suppose, $\det(J(u,v)) \neq 0$ on $\tilde{I}$.

Under all these conditions:

$$f_{U,V}(u,v) = f_{X,Y}(y(u,v)) \cdot \left|\det(J(u,v))\right| \quad \forall (u,v) \in \tilde{I}$$

. . . .

$\underset{\sim}{X} = (X_1, X_2, \ldots, X_k)$ has density $f_{\underset{\sim}{X}}$ on $I \leftarrow$ open region in $\mathbb{R}^k$.

$h: I \longrightarrow \tilde{I} \subset \mathbb{R}^k$

$g := h^{-1}: \tilde{I} \to I$.

$(y_1, \ldots, y_k) \longmapsto (x_1, \ldots, x_k)$.

$J(\underset{\sim}{y}) = \left(\left(\frac{\partial x_i}{\partial y_j}\right)\right)_{k \times k}$ exists and continuous on I,

& $\det(J(\underset{\sim}{y})) \neq 0$ on I.

then $\underset{\sim}{Y} = h(\underset{\sim}{X})$ has density $f_{\underset{\sim}{Y}}(y_1, \ldots, y_k) = f_{\underset{\sim}{X}}(g(y))$.

<u>Special Case:</u>

Let A be a non-singular matrix
$\underset{\sim}{\beta}$ be a k-dimensional column vector.

$h(\underset{\sim}{x}) = A\underset{\sim}{x} + \underset{\sim}{\beta} \longleftarrow$ "translation"

$g(\underset{\sim}{y}) = A^{-1}(\underset{\sim}{y} - \underset{\sim}{\beta})$

$$g(\underset{\sim}{y}) = A^{-1}(\underset{\sim}{y} - \underset{\sim}{\beta})$$

$$\det(J(\underset{\sim}{y})) = \det(A^{-1}).$$

$\Rightarrow$ $\underset{\sim}{Y} = A\underset{\sim}{X} + \underset{\sim}{\beta}$ has density

$$f_Y(\underset{\sim}{y}) = \frac{1}{|\det A|} \cdot f_{\underset{\sim}{X}}(A^{-1}(\underset{\sim}{y} - \underset{\sim}{\beta}))$$

A general approach if this method fails:

$\underset{\sim}{X}$ has density $f_{\underset{\sim}{X}}$ on I.

$\underset{\sim}{Y} = h(\underset{\sim}{X})$. "Special case" doesn't hold.

$h : I \longrightarrow \tilde{I}$.

Fix a point $\underset{\sim}{y} \in \tilde{I}$, take $\underset{\sim}{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n) > 0$ (ie, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\vdots$, $\varepsilon_n > 0$)

small enough so that —

$$[\underset{\sim}{y}, \underset{\sim}{y} + \varepsilon) \subset \tilde{I}$$

$$\parallel$$

$$[y_1, y_1 + \varepsilon) \times \cdots \times [y_k, y_k + \varepsilon).$$

Compute:

$$P(\underset{\sim}{Y} \in [\underset{\sim}{y}, \underset{\sim}{y} + \varepsilon)) = P(h(\underset{\sim}{X}) \in [\underset{\sim}{y}, \underset{\sim}{y} + \varepsilon))$$

look at $\frac{1}{(\prod_{i=1}^{n} \varepsilon_i)} \cdot P(\underset{\sim}{Y} \in [\underset{\sim}{y}, \underset{\sim}{y} + \varepsilon))$

$\hookrightarrow$ volume of that box.

"Lebesgue's Differentiation Theorem"

now, let $\varepsilon_i \longrightarrow 0$ & see if there exists a limit.

Conditional Distribution:

X, Y – random variables.

(X, Y) has a joint distribution.

How to define: Conditional Distribution of Y given $X = x$.

Aim: to define $Q(x, B) \overset{??}{=\!=} P(Y \in B \mid X = x)$

a function

set

point.

ie, both a set & point f^n.

Such that, Q must satisfy:

1. for each $x \in \mathbb{R}$, $Q(x, B)$ is a prob. on $\mathbb{R}$.
2. for each $B \subset \mathbb{R}$, $Q(x, B)$ is measurable f^n of x.
3. For every pair of borel sets $A, B \subset \mathbb{R}$.

$$P(X \in A, Y \in B) = E(\underbrace{Q(X,B)}_{\text{A random variable: } P(Y \in B \mid X)} \cdot 1_A(X))$$

does the job of keeping $X \in A$.

Cases where we can identify a legitimate "Q"?

Case-1: X- discrete.

Case-2: When (X,Y) has a joint density then,

$$h(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{f_X(x)} & \forall x, \text{ s.t } f_X(x) > 0 \\ \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}y^2} & \text{otherwise} \end{cases}$$

can be anything else as well !!

for these 2 cases above —

$$\therefore Q(x, B) = \int_B h(x,y)\, dy.$$

[Exercise: prove that this specific $Q(x,B)$ follows the three conditions.]

Case-3: X, Z - independent r.v.s.

independent → necessary.

$$Y := h(X,Z)$$

$$Q(x,B) = P(h(x,Z) \in B)$$

[if x is fixed, $h(x,Z)$ is a random variable]

Claim:

Q is a candidate for $P(Y \in B \mid X = x)$.

Exercise: prove the claim. (hint: starting from very def^n of independence)

Why does this makes sense intuitively?

$$P(h(X,Z) \in B \mid X = x) = P(h(x,Z) \in B)$$

[$\because X, Z$ are independent, the condition that $X = x$

[$\because X, Z$ are independent, the condition that $X = x$ only affects X, <u>not</u> Z.]

Probability-3 Lecture-3

30 July 2024 12:56

(X, Y) - random vector.

Conditional distribution of Y, given X.

Fix $x \in \mathbb{R}$. $B \subset \mathbb{R}$.

Suppose $Q(x, B)$ satisfies:

$$(*)\begin{cases} 1) \text{ for each } x \in \mathbb{R},\ Q(x,B) \text{ is a probability in } B. \\ 2) \ \forall \text{ borel } B \subset \mathbb{R},\ Q(x,B) \text{ is a measurable function.} \\ 3) \ \forall \text{ borel } B, C \subset \mathbb{R}, \\ \qquad P(Y \in B, X \in C) = E\left(Q(X,B) \cdot 1_C(X)\right) \end{cases}$$

$$Q(x,B) = P(Y \in B \mid X = x)$$

(interpretation)

Two "trivial" cases:

1) X is discrete.

Say, with $D \subset \mathbb{R}$ countable, satisfying $P(X=x) > 0 \ \forall x \in D \subset \mathbb{R}$.

Define
$$Q(x,B) = P(Y \in B \mid X = x) = \begin{cases} \dfrac{P(Y \in B, X = x)}{P(X=x)}, & x \in D. \\ \delta_{\{x\}}(B), & x \notin D. \end{cases}$$

Dirac mass.

(doesn't matter!!!)

2. (X,Y) jointly (absolutely) continuous with joint density f

has joint density.

Define $g(x,y) = \begin{cases} \dfrac{f(x,y)}{}, & \text{if } \int f(x,y)\,dy > 0 \end{cases}$

Define $g(x,y) = \begin{cases} \dfrac{f(x,y)}{\int f(x,y)\,dy}, & \text{if } \int f(x,y)\,dy > 0 \\ \dfrac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}y^2}, & \text{if } \int f(x,y)\,dy = 0. \end{cases}$

Marginal density of x.

Can be anything else as well!!

Define $Q(x,B) = \int_B g(x,y)\,dy$

A Special Case: X, Y – independent r.v.s.

$h : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a measurable function.

$\therefore h(X,Y)$ is a random variable.

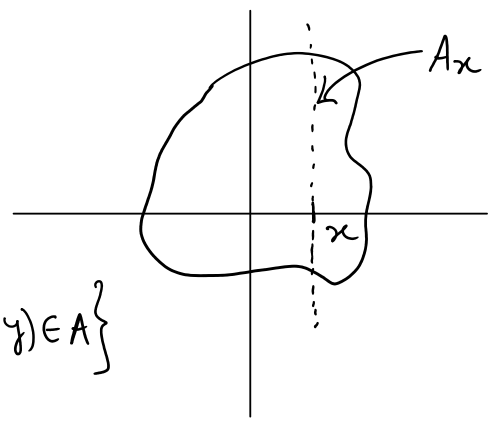
Q. What is the conditional distribution of $h(X,Y)$ given X.

- Proposed conditional distribution:

define $Q(x,B) = P(h(x,Y) \in B)$.

to show: this $Q(x,B)$ follows all the 3 results (*)

Consider the following setting:



Let $A \subset \mathbb{R}^2$ → borel subset of $\mathbb{R}^2$.

For $x \in \mathbb{R}$, denote $A_x = \{ y \in \mathbb{R} \mid (x,y) \in A \}$ ← "x-section"

* $\forall A \subset \mathbb{R}^2$ borel,

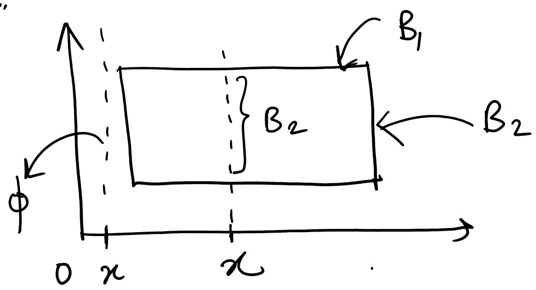
$A_x \subset \mathbb{R}$ is borel $\forall x \in \mathbb{R}$.

[ie, pick a borel in $\mathbb{R}^2$. take an x-section. that $A_x \subset \mathbb{R}$ is also borel.]

Take $A = B_1 \times B_2$ ← "rectangle"

Take $A = B_1 \times B_2 \leftarrow$ "rectangle"

Clearly, $A_x = \begin{cases} B_2 & , \text{ if } x \in B_1 \\ \phi & , \text{ if } x \notin B_1 \end{cases}$



So, we can talk about
$P(Y \in A_x)$, $\forall$ borel $A \subset \mathbb{R}^2$
& $x \in \mathbb{R}$.

$= \varphi(x, A)$

Now, fix a borel $A \subset \mathbb{R}^2$, & keep on varying x.
ie, look at x-section.

look at
$x \longmapsto \varphi(x, A)$

$\therefore$ This is measurable $\forall$ $A \subset \mathbb{R}^2$ borel.

Clearly, this map $x \longmapsto \varphi(x, A)$ is a f^n from borel $\mathbb{R}$ to $[0,1]$.

$\therefore$ for $A = B_1 \times B_2$.
$P(Y \in A_x) = P(Y \in B_2) \cdot 1_{B_1}(x)$

Now, to show,

$\mathcal{A} = \{ A \subset \mathbb{R}^2 : \varphi_x(A)$ is measurable in x $\}$
$\forall$ $B_1 \times B_2 \in \mathcal{A}$

"Monotone class Theorem"

Theorem (Monotone Class Theorem):

Ω is a set. $\mathcal{F}$ is a field on Ω
$\to$ NOT σ-field !!
$\to$ closed under
- complementation
- finite intersection.

If M is a class of sets such that,
$\mathcal{F} \subset M$, and M is closed under monotone limits, then $M \supset \sigma(\mathcal{F})$

$\mathcal{F} \subset M$, and M is closed under monotone limits, then $M \supset \sigma(\mathcal{F})$

σ-field generated by field $\mathcal{F}$.

In our case, we have to first show - the set of rectangles in $\mathbb{R}^2$ form a field.

Where do we stand now? let $A \subset \mathbb{R}^2$ any Borel set.
$x \in \mathbb{R}$. $A_x = \{ y \in \mathbb{R} \mid (x,y) \in A \}$

Step-1: We showed:
$A_x \subset \mathbb{R}$ is Borel $\forall x$.
(first for rectangles, then ....)

then, $P(Y \in A_x)$ — an event

Borel set

$\therefore$ we can talk about this probability

Step-2: the f^n: $x \longmapsto \varphi(x, A)$ is measurable.

Step-3: $P((X,Y) \in A)$

Borel in $\mathbb{R}^2$
Hence, (X,Y) is an event.
So, we can talk about this probability.

Claim:

$$P((X,Y) \in A) = E(\varphi(X, A)).$$

a r.v. [$\because$ from step-2: $x \longmapsto \varphi(x,A)$ is measurable]
hence, we can talk about this expectation.

expectation.

Main thing: we can't do serious Probability without Measure Theory !!

Here, $A \subset \mathbb{R}^2$.

if $A = B_1 \times B_2$.

LHS $= P((X,Y)\in A) = P((X,Y) \in B_1 \times B_2)$

$= P(X\in B_1, Y \in B_2)$

$\left[\because X, Y \text{ independent}\right] = P(X\in B_1)\cdot P(Y\in B_2)$.

&, RHS $= E(\varphi(X,A))$ &, $\varphi(x,A) = P(Y\in A_x)$

$\therefore \varphi(X,A) = P(Y \in B_2)\cdot 1_{B_1}(X)$

the X-section.

$= E(P(Y\in B_2)\cdot 1_{B_1}(X))$

$= P(Y\in B_2)\cdot E(1_{B_1}(X))$

$= P(Y\in B_2)\cdot P(X\in B_1)$.

$\therefore$ LHS = RHS is immediate if $A = B_1\times B_2 \subset \mathbb{R}^2$ borel.

Now, back to our last qn from last lecture (Lec-2)

proposed $Q(x,B) := P(h(x,Y)\in B)$

Is $Q(x,B)$ measurable in x?

$Q(x,B) = \varphi(x,A)$, where $A = h^{-1}(B)$.

$= \{(x,y) \mid h(x,y)\in B\}$

$\therefore$ the X-section,

$A_x = \{y \mid h(x,y) \in B\}$.

$\therefore \varphi(x,A) = P(Y\in A_x)$

$= P(h(x,Y)\in B)$, which is measurable.

$\therefore Q(x,B)$ is measurable.

Now, Is $Q(x,B)$ a probability in B?

$Q(x,B) = P(h(x,Y)\in B)$. • if $B = \mathbb{R}$, Is $\varphi(x,B)$ ... :

$$\varphi(x,B) = P(h(x,Y)\in B).$$

- if $B=\mathbb{R}$, $P(h(x,Y)\in\mathbb{R})=1$ ✓
- take $B, B^c \subset \mathbb{R}$. $P(h(x,Y)\in B) = 1-P(h(x,Y)\in B^c)$ ✓
- B_1, B_2 - disjoint borels in $\mathbb{R}$. $P(h(x,Y)\in B_1 \sqcup B_2) = P(h(x,Y)\in B_1) + P(h(x,Y)\in B_2)$ ✓

& now, finally, is

$$P(h(X,Y)\in B, X\in C) = E(\varphi(X,B)\cdot 1_C(X)) \quad ??$$

take $A = \{(x,y) \mid h(x,y)\in B, x\in C\}$.

$$\therefore A_x = \begin{cases} \{y \mid h(x,y)\in B\}, & x\in C \\ \phi, & x\notin C \end{cases}$$

(empty set)

Now, the 3 steps above can be put to use.

Little relief: 🙂

All these proofs are <u>not</u> a part of the syllabus. Only the final result (i.e, the proposition) is!!

<u>A consequence:</u>

X, Y - independent r.vs.

$h(x,y) = x+y$.

finding distⁿ of $Z = X+Y$

$$F_Z(a) = P(Z\le a) = P(X+Y\le a)$$

$$= E(\varphi(X) \mid \text{where } \varphi(x) = P(x+Y\le a))$$

$F_Z(a) = P(Z \leq a) = \ldots$

$$= E\left(\varphi(X)\right) \quad \text{where } \varphi(x) = P(X+Y \leqslant a) = P(Y \leqslant a-x) = F_Y(a-x)$$

$\therefore$

$$= E\left(F_Y(a-X)\right)$$

$$= E\left(F_X(a-Y)\right)$$

usually called

$F_X * F_Y$. This is called the <u>convolution</u>.

So, in general, X has distⁿ fⁿ F
Y " " " G.

Then, $X+Y$ has distⁿ $F * G$

<u>Fact</u>: take distⁿ of X degenerate at 0.
It acts as identity !!

Also, in general, F does <u>not</u> have an inverse.
But, if F is degenerate, then inverse exists.

<u>Prove that</u>: No other distⁿ will have an inverse.

Suppose F has density f $X \sim F$
G " " g. $Y \sim G$, X, Y - ind.

$$F * G(a) = P(X+Y \leqslant a)$$
$$= E\left(G(a-X)\right)$$
$$= \int G(a-x) f(x)\, dx$$
$$= \int^{\infty} \int^{a-x} g(u)\, du \, f(x)\, dx.$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a-x} g(y)\,dy\; f(x)\,dy\,dx$$

$u = y + x$
$\Rightarrow y = u - x$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{a} g(u-x)\cdot f(x)\,dy\cdot dx$$

$$= \int_{-\infty}^{a} \underbrace{\int_{-\infty}^{\infty} g(u-x)\cdot f(x)\,dx}_{\text{Acts as density !!}} \cdot dy$$

$\therefore$ $F * G$ has density given by—

$$f*g(u) = \int_{-\infty}^{\infty} g(u-x)\cdot f(x)\,dx\,.$$

More general cases:

Given general X, Y.

Try to find $P(Y \leq a \mid X = x)$

Take x s.t.

$$P(X \in (x-\varepsilon, x+\varepsilon)) > 0 \quad \forall\, \varepsilon > 0$$

Find $P(Y \leq a \mid X \in (x-\varepsilon, x+\varepsilon))$ $\forall\, \varepsilon > 0$ using "classical definition"

In general, the classical defn does not work as the conditioning event $X = x$ might not have positive probability.

Now, we'll do $\varepsilon \xrightarrow{\text{lt}} 0$ & check if a limit exists. → [Lebesgue Differentiation Theorem ⚠ (out of syllabus)]

(If it exists, then it works!! i.e., it gives a candidate for the conditional distn.)

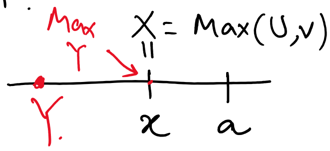
$P(Y \leq a \mid X = x)$.

Eg: $U \sim \mathrm{Exp}(1)$ } Independent.

Eg: $U \sim \text{Exp}(1)$, $V \sim \text{Exp}(2)$ } Independent.

$X = \text{Max.}\{U, V\}$ $\quad X > 0$, $\; Y > 0$, $\; X \geqslant Y$.

$Y = V$



So, Y always $\leq$ a.

to find: $P(Y \leq a \mid X = x)$

Fix $x > 0$. $\quad P(Y \leq a \mid X = x) = 1$ if $a \geqslant x$

for $a < x$, take $\varepsilon > 0$ st.

$P(Y \leq a \mid X \in (x - \varepsilon, x + \varepsilon))$

$= P(Y \leq a, X \in (x - \varepsilon, x + \varepsilon))$

$\equiv V \leq a, \; U \in (x - \varepsilon, x + \varepsilon)$

$$\therefore f(a|x) = \begin{cases} \text{density}, & a < x \\ 1, & a \geqslant x. \end{cases}$$

Probability-3 Lecture-4

02 August 2024 11:20

$X \sim F_X$
$Y \sim F_Y$ } Independent.

$Z = X+Y \sim F_X * F_Y$.

(Convolution).

Special Case:

X - density f_X.

Y - " f_Y.

$Z = X+Y$ has density:

$$f_Z(z) = \int f_Y(z-x) \cdot f_X(x)\, dx$$

$$= \int f_X(z-y) \cdot f_Y(y)\, dy$$

$$f_Z = f_X * f_Y.$$

Suppose $X \sim \text{Gamma}(\lambda, \alpha_1)$
$Y \sim \text{Gamma}(\lambda, \alpha_2)$ } independent.

$$Z = X+Y \sim \text{Gamma}(\lambda, \alpha_1+\alpha_2).$$

$$f_Z(z) = \int_0^z f_Y(z-x) \cdot f_X(x)\, dx \quad , z>0.$$

$x>0 \to$ lower limit,

$z-x>0$
$\Rightarrow x<z \to$ upper limit

Eg.: $X, Y \sim \text{Unif}(0,1)$ are independent.

$Z = X+Y$. $\quad 0<Z<2$

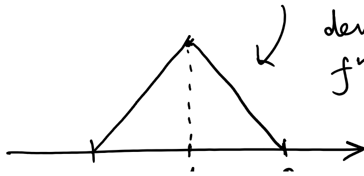
$$f_Z(z) = \int f_Y(z-x) \cdot f_X(x)\, dx.$$

$$= \begin{cases} \int_0^z 1 \cdot 1 \cdot dx, & 0<z<1 \\ \int_{z-1}^1 1 \cdot 1 \cdot dx, & 1<z<2 \end{cases}$$

$0 \leq x < 1$
$0 < z-x < 1$
$x < z < 1$

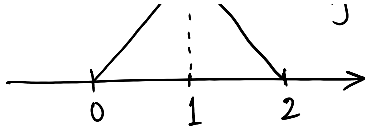
$0 < x \leq 1$
$0 < z-x < 1$
$x > z-1 > 0$

$$= \begin{cases} z, & 0<z<1 \\ 2-z, & 0 \leq z < 2 \end{cases}$$



that's how density f^n looks like.

$\lfloor 2-z, \quad 0 \leq z < 2$



looks like.

<u>Review of Multivariate:</u>

$\underset{\sim}{X} = (X_1, \ldots, X_k)^T$.

$E(\underset{\sim}{X}) = \left(E(X_1), \ldots, E(X_k)\right)^T$ - mean vector.

$D(\underset{\sim}{X}) = ((Cov(X_i, X_j)))_{k \times k}$ - "dispersion matrix", or "Variance - Covariance Matrix".

$E(A\underset{\sim}{X} + \underset{\sim}{\beta}) = A \cdot E(\underset{\sim}{X}) + \underset{\sim}{\beta}$ [Linearity of Expectation].

$D(A\underset{\sim}{X} + \beta) = A\, D(\underset{\sim}{X}) \cdot A^T$.

Clearly, $D(\underset{\sim}{X}) = E\left((X - E(X))(X - E(X))^T\right)$

* $D(\underset{\sim}{X})$ is a real, symmetric, non-negative definite (nnd)

$\underset{\sim}{a}^T D(X)\, \underset{\sim}{a} = Var(\underset{\sim}{a}^T X) \geq 0$

* $D(\underset{\sim}{X})$ is singular.

$\Rightarrow P(\underset{\sim}{X} - E(\underset{\sim}{X}) \in \text{Column space}(D(\underset{\sim}{X}))) = 1$

X has density.

$\Rightarrow D(X)$ is non-singular, positive definite.

Conversely, every p.d is matrix is the dispersion matrix of some random vector, namely the normal random vector.

[take the $\underbrace{\text{quadratic form}}$ of the pd matrix.

$e^{-\frac{1}{2}(\quad)}$ is integrable.

(normalise by some constant to get the density of the normal random vector)]

Some special Multivariate Distributions:

① Discrete Multinomial $(n; p_1, \ldots, p_k)$: $\quad 0 < p_i < 1$

k

① Discrete Multinomial $(n;\ p_1, \ldots, p_k)$: $\quad 0<p_i<1$, $\sum_{i=1}^{k} p_i = 1$

$$P(X_1=x_1, \ldots, X_k=x_n) = \frac{\lfloor n}{\lfloor n_1 \cdot \lfloor n_2 \cdots \cdots \lfloor n_k} \cdot p_1^{n_1} \cdot p_2^{n_2} \cdots \cdots p_k^{n_k}.$$

(Classical example: rolling a die.)

Eg.:

$X \sim \text{Bin}(n,p)$.

$\therefore X, n-X \sim \text{Multi}(n, p, 1-p)$.

<u>Marginals</u>:

$j<k$.

$$(X_1, \ldots, X_j,\ X_{j+1} + \cdots + X_k) \sim \text{Multi}(n;\ p_1, \ldots, p_j,\ p_{j+1} + \cdots + p_k).$$

(Exercise: conditional distributions.)

② <u>Dirichlet</u>:

$D_k(\alpha_1, \ldots, \alpha_k;\ \alpha_{k+1})$, $\quad \alpha_i > 0 \quad \forall i$

Density is given by:

$$f(x_1, x_2, \ldots, x_k) = \frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_{k+1})}{\Gamma(\alpha_1)\cdot\Gamma(\alpha_2)\cdots\Gamma(\alpha_{k+1})} \cdot x_1^{\alpha_1-1} \cdots x_k^{\alpha_k-1} (1-x_1-\cdots-x_k)^{\alpha_{k+1}-1}$$

for $\underset{\sim}{x} \in \underbrace{S}_{\text{"Simplex"}} = \{(x_1, \ldots, x_k):\ x_i>0,\ \forall i,\ x_1+\cdots+x_k<1\}$.

$X \sim \text{Beta}(\alpha_1, \alpha_2) \Longleftrightarrow X \sim D_1(\alpha_1; \alpha_2)$.

$$\therefore \iint_S \cdots \int x_1^{\alpha_1-1} \cdots \cdots x_k^{\alpha_k-1} \cdot (1-x_1-\cdots-x_k)^{\alpha_{k+1}-1}\, dx_k \cdots dx_1$$

the inner-most integral.

$$\int^{1-x_1-\cdots-x_{k-1}}$$

$$\int_0^{1-x_1-\cdots-x_{k-1}} x_k^{(\alpha_k-1)} \cdot (1-x_1-\cdots-x_{k-1}-x_k)^{\alpha_{k+1}-1} dx_k.$$

change of variable:

$$u = \frac{x_k}{(1-x_1-\cdots-x_{k-1})}$$

$$\Rightarrow x_k = u(1-x_1-\cdots-x_{k-1})$$

$$= (1-x_1-\cdots-x_{k-1})^{\alpha_k+\alpha_{k+1}-1} \cdot \underbrace{\int_0^1 u^{\alpha_k-1} \cdot (1-u)^{\alpha_{k+1}-1} \cdot du}_{\beta(\alpha_k,\alpha_{k+1})}$$

$$= (1-x_1-\cdots-x_{k-1})^{\alpha_k+\alpha_{k+1}-1} \cdot \frac{\Gamma(\alpha_k)\cdot\Gamma(\alpha_{k+1})}{\Gamma(\alpha_k+\alpha_{k+1})}$$

(proceed further... Exc.)

$(X_1,\ldots,X_j)^T \sim D_j(\alpha_1,\ldots,\alpha_j, \alpha_{j+1}+\cdots+\alpha_{k+1})$.
$1 \leq j < k$.

* $\pi : \{1,\ldots,k\} \longrightarrow \{1,\ldots,k\}$ is a permutation.

$$(X_{\pi(1)},\ldots,X_{\pi(k)})^T \sim D_k(\alpha_{\pi(1)},\ldots,\alpha_{\pi(k)}; \alpha_{k+1})$$

* $Y_1,\ldots,Y_{k+1}$ — independent.

$$Y_i \sim \text{Gamma}(\alpha_i), \quad 1 \leq i \leq k+1.$$

Show that (exercise): $(X_1,\ldots,X_k)^T \sim D_k(\alpha_1,\ldots,\alpha_k; \alpha_{k+1})$,

where $X_j = \dfrac{Y_j}{Y_1+\cdots+Y_{k+1}}$.

[Best book for Dirichlet (or just multivariate dist^ns in general)
• Samuel Wilks : Mathematical

L • Samuel Wilks : Mathematical Statistics.

dist^ns in general)

"Ordered" Dirichlet :

$\underset{\sim}{Y} \sim D_k^*(\alpha_1, \ldots, \alpha_k, \alpha_{k+1})$, where $Y_1 = X_1$,

$Y_2 = X_1 + X_2$

$\vdots$

$Y_k = X_1 + \cdots + X_k$.

$$f(\underset{\sim}{Y}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})} \cdot y_1^{\alpha_1 - 1} \cdot (y_2 - y_1)^{\alpha_2 - 1} \cdots (y_k - y_{k-1})^{\alpha_k - 1} (1 - y_k)^{\alpha_{k+1} - 1}$$

for $0 < y_1 < y_2 < \cdots < y_k < 1$

Order Statistics:

$X_1, X_2, \ldots, X_k$ i.i.d.

The order statistic obtained from $(X_1, \ldots, X_k)$

$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(k)}$.

$X_{(j)}$ is called the j^{th} order statistic.

Assume, each X_j has density f.

Q. How to find joint density of $(X_{(1)}, \ldots, X_{(n)})^T$?

Strategy – Take a $y_1 < y_2 < \cdots < y_k$

take a "rectangle"

$R = (y_1, y_1 + \varepsilon] \times \cdots \times (y_k, y_k + \varepsilon]$.

$$P\left((X_{(1)}, \ldots, X_{(k)})^T \in R\right) = \underline{|k} \cdot \prod_{j=1}^{k} \int_{y_j}^{y_j + \varepsilon} f(x) \cdot dx$$

$\because X_{(1)}, \ldots, X_{(k)}$ is some permutation of $X_1, \ldots, X_k$, & # total permutations $= \underline{|n}$

Also, they can be factored because the events

$X_{(i)} \in (y_i, y_i + \varepsilon]$

because the events
$X_{(j)} \in (y_i ; y_i+\varepsilon]$
are <u>disjoint</u>. $\therefore$

$$\therefore \lim_{\varepsilon \downarrow 0} \frac{P((X_{(1)}, \ldots\ldots, X_{(k)}) \in \mathbb{R})}{\varepsilon^k} = \underline{\lfloor k} \cdot \prod_{j=1}^{k} f(y_j) \quad \text{for } y_1 < y_2 < \cdots < y_k.$$

(By FTC)

<u>Note:</u>

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\iint_{y_1<y_2} f(y_1)\cdot f(y_2) = \frac{1}{\underline{\lfloor 2}}$$

$$\iiint_{y_1<y_2<y_3} f(y_1)\cdot f(y_2)\cdot f(y_3) = \frac{1}{\underline{\lfloor 3}}$$

$\vdots$

and so on.

[in case of iid rv, order statistic $\equiv$ sufficient statistic.]

$X_1, \ldots, X_k$ – independent with a continuous distn. fn F.
then, the order statistic $(F(X_{(1)}), F(X_{(2)}) \ldots, F(X_{(k)}))$
$\sim D_k^*(1, 1, \ldots, 1 ; 1)$.

(<u>Exercise:</u> Verify)

[Recall X – r.v. has cont. distn fn F. then, define: $Y := F(X)$ then, $Y \sim \text{Unif}(0,1)$]

<u>Multivariate Normal:</u>

$\sum_{k\times k}$ symmetric, p.d. matrix.

$\Sigma_{k\times k}$ symmetric, p.d. matrix.

$\underset{\sim}{\mu}$ is a vector.

X has density:

$$f_X(\underset{\sim}{x}) = (2\pi)^{-k/2} \cdot \Sigma^{-1/2} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}, \quad \underset{\sim}{x} \in \mathbb{R}^k.$$

* $\underset{\sim}{\mu} = E(\underset{\sim}{X})$; $\Sigma = D(\underset{\sim}{X})$.

* $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma) \Rightarrow$ for non-singular A & vector $\underset{\sim}{\beta}$,

$$A\underset{\sim}{X} + \underset{\sim}{\beta} \sim N(A\underset{\sim}{\mu} + \underset{\sim}{\beta}, A\Sigma A^T)$$

$$(X_{\pi(1)}, \ldots, X_{\pi(k)}) \sim N_k(\mu_\pi, \Sigma_\pi).$$

&, for $m < k$,

$$(X_1, \ldots, X_m) \sim N_m(\cdot\,,\,\cdot)$$

* for non-null $\underset{\sim}{a}$,

$$a^T\underset{\sim}{X} = N(a^T\mu, \underset{\sim}{a}^T \Sigma \underset{\sim}{a})$$

* $\underset{\sim}{X} \sim N_k(\underset{\sim}{\mu}, \Sigma) \Rightarrow \exists$ non-singular A s.t. $A(\underset{\sim}{X} - \underset{\sim}{\mu}) \sim N_k(\underset{\sim}{0}, I_k)$ $\rightarrow$ [Here, $A = \Sigma^{-1}$]

* $\sigma_{ij} = 0 \Rightarrow X_i$ & X_j are independent.

* $(\underset{\sim}{X} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu}) \sim \chi^2_{(k)} \equiv \text{Gamma}\left(\frac{1}{2}, \frac{k}{2}\right)$.

$X_1, \ldots, X_k \overset{iid}{\sim} N(0,1)$.

$Q_1(\underset{\sim}{x}), \ldots, Q_n(\underset{\sim}{x})$ are quadratic forms. | say, $Q_j(\underset{\sim}{x}) = \underset{\sim}{x}^T A_j \underset{\sim}{x}$.

$$Q_1(\underset{\sim}{x}) + \cdots + Q_n(\underset{\sim}{x}) = \underset{\sim}{x}^T x.$$

Let $r = \text{rank}(A_j)$, $1 \leq j \leq n$

$Q_1(\underset{\sim}{X}), \ldots\ldots, Q_n(\underset{\sim}{X})$ are independent with

$Q_j(\underset{\sim}{X}) \sim \chi^2_{(r_j)} \iff r_1 + r_2 + \cdots + r_n = k.$

(Version of Fisher-Cochran theorem)

Exercise: prove it!!! (think)
$\Rightarrow$ is trivial
$\Leftarrow$ Linear algebra.

One "easy" consequence:

$\underset{\sim}{X}^T A \underset{\sim}{X}$ is χ^2

$\iff A^2 = A$. (ie, A-idempotent).

In that case,
degrees of freedom,
$d.f = r(A) = tr(A)$.

Q. $\underset{\sim}{X}^T A \underset{\sim}{X}$ - quadratic form.

$\underset{\sim}{X}^T \sim N(\underset{\sim}{0}, \Sigma)$ not necessarily I.

What is a necessary & sufficient condition for the above to hold??

$\Sigma^{-1/2} \cdot \underset{\sim}{X}^T \sim N(0, I)$. (Now think!!!).

Probability-3 Lecture-5

06 August 2024 14:21

Fisher-Cochran Theorem:

$Q_1(x) = \underset{\sim}{x}^T A_1 \underset{\sim}{x}, \ldots, Q_n(x) = \underset{\sim}{x}^T A_n \underset{\sim}{x}$.

are quadratic forms satisfying:

$Q_1(x) + \cdots + Q_n(x) = \underset{\sim}{x}^T \underset{\sim}{x}$.

$\left[\text{ie, } A_1 + \cdots + A_n = I\right]$.

Let $r_j := r(A_j)$, $1 \le j \le n$.

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(0,1)$

Then a necessary & sufficient condition for $Q_1(\underset{\sim}{X}), \ldots, Q_n(\underset{\sim}{X})$ to be independent with

$$Q_j(\underset{\sim}{X}) \sim \chi^2_{r_j}, \quad 1 \le j \le n.$$

is: $\boxed{r_1 + \cdots + r_n = k}$.

<u>Proof</u>: "Necessity" is obvious.

$$Q_j(\underset{\sim}{X}) \sim \chi^2_{r_j}$$

$$\therefore \sum Q_j(\underset{\sim}{X}) \sim \chi^2_{\Sigma r_j} = \chi^2_k$$

"Sufficiency": $\forall\, j = 1, \ldots, n$,

$\exists$ linearly independent vectors

$\underset{\sim}{l}_{j,1}, \underset{\sim}{l}_{j,2}, \ldots, \underset{\sim}{l}_{j,r_j}$ s.t.

$$Q_j(\underset{\sim}{x}) = \pm\left(\underset{\sim}{l}_{j}^T \cdot \underset{\sim}{x}\right)^2 \pm \cdots \pm \left(l_{j r_j}^T \cdot \underset{\sim}{x}\right)^2$$

[Diagonalizing. Refer V.M.-2]

Let B be the matrix.

first r_1 rows: $\underset{\sim}{l}_{1,1}, \ldots, \underset{\sim}{l}_{1,r_1}$

next r_2 rows: $\underset{\sim}{l}_{2,1}, \ldots, l_{2,r_2}$,

$\vdots$

& so on.

So, B is a $k \times k$ matrix.

$$\sum_j Q_j(\underset{\sim}{x}) = \underset{\sim}{x}^T B^T \Delta B \underset{\sim}{x},$$

where $\Delta = \text{diag}(\pm 1, \pm 1, \ldots, \pm 1)$.

$\sum_j Q_j(\underset{\sim}{x}) = x^T x$

$\Rightarrow x^T x = x^T B^T \Delta B x.$

$\Rightarrow B^T \Delta B = I.$

$\Rightarrow \boxed{r(B) = k}$.

$\Rightarrow$ B is non-singular.

$\Delta = (B^T)^{-1} \cdot B^{-1}$

$\Rightarrow \Delta$ is positive definite

$\therefore$ &, $\Delta = \text{diag}(\pm 1, \ldots, \pm 1)$

this forces that $\Delta = \text{diag}(1, 1, \ldots, 1)$

$\Rightarrow \boxed{\Delta = I}$.

Also, B is orthogonal

$\Rightarrow \underset{\sim}{Y} = B \underset{\sim}{X} \Rightarrow Y_1, \ldots, Y_k \overset{iid}{\sim} N(0,1)$.

$$\therefore \quad Q_j(\underset{\sim}{x}) = Y^2_{\sum_{i=1}^{j-1} r_i + 1} + \cdots + Y^2_{\sum_{i=1}^{j-1} r_i + r_j}$$

$$\sim \chi^2_{r_j}$$ ▨

<u>Corollary</u>: $X_1, \ldots, X_k \overset{iid}{\sim} N(0,1)$. A - real, symmetric matrix.

$Q(\underset{\sim}{X}) = \underset{\sim}{X}^T A \underset{\sim}{X}$ has a χ^2 distribution

$\Longleftrightarrow$ A is idempotent,

& in that case, degrees of freedom of

$\chi^2 = r(A) = tr(A)$

& in that case, 'degrees of freedom of $\chi^2 = r(A) = tr(A)$

Proof: ($\Leftarrow$)

Assume $A^2 = A$.

$\Rightarrow A(I-A) = 0$.

$\Rightarrow k \le r(A) + r(I-A)$.

$\le r(A + (I-A)) = k$.

Sylvester's inequality

$\Rightarrow$ By Fisher-Cochran $\underset{\sim}{X}^T A \underset{\sim}{X} \sim \chi^2_{r(A)}$.

"$\Rightarrow$" Suppose, $\underset{\sim}{X}^T A \underset{\sim}{X} \sim \chi^2_d$

To prove: A is idempotent, & $d = r(A)$.

Let $r = r(A)$.

$\exists$ orthogonal B st.

$$B^T A B = diag(\lambda_1, \dots, \lambda_r, 0 \dots 0),$$

where $\lambda_1, \dots, \lambda_r$ are the non-zero eigenvalues of A.

We make the orthogonal transformation:

$$\underset{\sim}{Y} = B\underset{\sim}{X} \Rightarrow Y_1, \dots, Y_k \overset{iid}{\sim} N(0,1)$$

Further,

$$\underset{\sim}{X}^T A \underset{\sim}{X} = \lambda_1 y_1^2 + \dots + \lambda_r y_r^2$$

Also, $\because$ the transformation is orthogonal,

$$\underset{\sim}{Y} = B\underset{\sim}{X}$$

$$\therefore \underset{\sim}{Y}^T \cdot \underset{\sim}{Y} = (B\underset{\sim}{X})^T (BX)$$

$$= X^T \underbrace{B^T B}_{I} X$$

$$= X^T B^T B X$$ with $\underbrace{B^T B}_{I}$

$$\underset{\sim}{Y}^T \underset{\sim}{Y} = \underset{\sim}{X}^T \underset{\sim}{X} \,..$$

Now, mgf: $E(e^{t\underset{\sim}{X}^T A \underset{\sim}{X}}) = E(e^{t(\lambda_1 y_1^2 + \cdots + \lambda_r y_r^2)})$

$$= \prod_{j=1}^{r} E(e^{t\lambda_j Y_j})$$

$$= \prod_{j=1}^{r} (1-2\lambda_j t)^{-1/2} \longleftarrow \text{Mgf of } \chi^2$$

$\therefore$ Now, suppose $\underset{\sim}{X}^T A X \sim \chi^2_d$,

$$\prod_{j=1}^{r} (1-2\lambda_j t)^{-1/2} = (1-2t)^{-d/2}$$

$\Rightarrow \boxed{d = r}$. & $\lambda_j = 1 \;\; \forall j$

equatting highest powers of t on both sides.

equatting the coefficients on both sides

Also, $B^T A B = \text{diag}(1,1,\ldots,1,0,\ldots,0)$.

$\therefore$ Clearly, $B^T A B$ is idempotent.

$$\Rightarrow B^T A \underbrace{B B^T}_{I} A B = B^T A B$$

$$\Rightarrow B^T A^2 B = B^T A B$$

$$\Rightarrow A^2 = A. \;\; \therefore A \text{ is idempotent.}$$

(*) Suppose, $\{a_n\}$ - non-ve sequence.

Refer

Suppose, $\sum^{\infty} a_n t^n$ converges for $0 \le t < 1$

Refer pgf question from Sem-2 midsem.

Suppose, $\sum_{n=1}^{\infty} a_n t^n$ converges for $0 \le t < 1$

then, $\lim_{t \uparrow 1} \sum_{n=1}^{\infty} a_n t^n$ exists & equals to $\sum_{n=1}^{\infty} a_n$.

firstly, as $t \uparrow 1$, $\sum_{n=1}^{\infty} a_n t^n \uparrow$. hence, limit exists & is the supremum.

$$\therefore \lim_{t \uparrow 1} \sum_{n=1}^{\infty} a_n t^n = \sup \left\{ \sum_{n=1}^{\infty} a_n t^n, t < 1 \right\}.$$

$$\stackrel{?}{=} \sum_{n=1}^{\infty} a_n$$

One side is clear:

$$\sup \left\{ \sum_{n=1}^{\infty} a_n t^n, t < 1 \right\} \le \sum_{n=1}^{\infty} a_n$$

i.e, this is an upper bound.

Take $\alpha < \sum_{n=1}^{\infty} a_n$. [if $\sum_n a_n$ diverges, take any $\alpha \in \mathbb{R}_{\geq 0}$]

we want to show, $\sum_{i=1}^{\infty} a_n$ is the lowest upper bound.

$\exists n_0$ s.t.

$$\sum_{n=1}^{N_0} a_n > \alpha \quad \text{[Archimedean Property]}$$

$$\Rightarrow \sum_{n=1}^{N_0} a_n t^n \longrightarrow \sum_{n=1}^{N_0} a_n. \text{ as } t \nearrow 1.$$

$\therefore$ for some $t \in [0,1)$,

$$\sum_{n=1}^{N_0} a_n t^n > \alpha.$$

$$\Rightarrow \sum_{n=1}^{\infty} a_n t^n > \alpha. \checkmark$$

$$\Rightarrow \sum_{n=1}^{\infty} a_n = \sup_{t \in [0,1)} \left\{ \sum_{n=1}^{\infty} a_n t^n \right\}.$$

PROBABILITY THEORY - 3 (officially begins)

(officially begins)

$(\Omega, \mathcal{A}, P)$ - Probability space.

$\{X_n\}$ - a sequence of real random variables, <u>all</u> on $(\Omega, \mathcal{A}, P)$.

<u>Definition</u>: Say that:

$\{X_n\}$ converges to X almost surely (a.s.) or, with probability 1 (w.p. 1)

if $P\left(\{w : X_n(w) \longrightarrow X(w)\}\right) = 1$

$x_n \longrightarrow x$
<u>iff</u> $\forall \varepsilon > 0$,
$\exists n \in \mathbb{N}$ s.t.,
$\forall m > N$,
$|x_m - x| < \varepsilon \equiv \frac{1}{j}$
(can be replaced)

Now, $\{w : X_n(w) \longrightarrow X(w)\} =$

$$\bigcap_{j \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \left\{w : |X_m(w) - X(w)| < \frac{1}{j}\right\}$$

$= \{w :$ for every $j \in \mathbb{N}$, $\exists n \in \mathbb{N}$ s.t., for every $m \geq n$, $|X_m(w) - X(w)| < \frac{1}{j}\}$

Now, for every $\{w : |X_m(w) - X(w)| < \frac{1}{j}\}$,

$X_m \to$ r.v. $\quad X \to$ r.v

$|X_m - X|$: also an r.v. in $(\Omega, \mathcal{A}, P)$.

$\therefore (X_m - X)^{-1} = (-1/j, 1/j)$, which is a Borel set

$\therefore$ each such set $\in \mathcal{A}$.

$\therefore$ Countable unions & intersections of such sets $\in \mathcal{A}$.

<u>Fact</u>: $\left.\begin{array}{l} X_n \longrightarrow X \text{ a.s.} \\ X_n \longrightarrow Y \text{ a.s.} \end{array}\right\} X = Y$ a.s.

<u>Proof</u>. $X_n \longrightarrow X$ a.s. $\Rightarrow P(\underbrace{\{w : X_n(w) \longrightarrow X(w)\}}_{\text{"A"}}) = 1$.

$X \longrightarrow Y$ a.s $\Rightarrow P(\{w : X_n(w) \longrightarrow Y(w)\}) = 1$.

Proof. $X_n \to X$ a.s. $\Rightarrow P(\{w : X_n(w) \to X(w)\}) = 1.$

$X_n \to Y$ a.s $\Rightarrow P(\{w : X_n(w) \to Y(w)\}) = 1.$

"B"

$\therefore P(A) = 1 \Rightarrow P(A^c) = 0$

$P(B) = 1 \Rightarrow P(B^c) = 0$

$\Rightarrow P(A^c \cup B^c) = 0$

$\Rightarrow P((A^c \cup B^c)^c) = 1$

$\Rightarrow P(A \cap B) = 1.$

$\Rightarrow P(\{w : X(w) \geq Y(w)\}) = 1$

&, $P(\{w : X(w) \leq Y(w)\}) = 1$

$\therefore P(\{w : X(w) = Y(w)\}) = 1$

Results:

(1) $X_n \xrightarrow{a.s.} X \Rightarrow c \cdot X_n \xrightarrow{a.s} cX$

Proof: $\{w : X_n(w) \to X(w)\} \subseteq \{w : c \cdot X_n(w) \to c \cdot X(w)\}$

$\therefore P(\{w : X_n(w) \to X(w)\}) = 1 = P(\{w : cX_n(w) \to cX(w)\})$

[ie, if $c = 0$, then this becomes entire Ω. $\because 0 \to 0 \ \forall w \in \Omega$]

(2) $\left.\begin{array}{l} X_n \xrightarrow{a.s} X \\ Y_n \xrightarrow{a.s} Y \end{array}\right\} \Rightarrow X_n + Y_n \xrightarrow{a.s.} X + Y$

Proof: $P(\underbrace{\{w : X_n(w) \to X(w)\}}_{A}) = 1 \quad \Rightarrow P(A) = 1.$

$P(\underbrace{\{w : Y_n(w) \to Y(w)\}}_{B}) = 1 \quad \Rightarrow P(B) = 1.$

$\Rightarrow P(A \cap B) = 1.$

$\Rightarrow P\{w : X_n(w) + Y_n(w) \to X(w) + Y(w)\} = 1$

$$\Rightarrow P\{w: X_n(w)+Y_n(w) \longrightarrow X(w)+Y(w)\}=1$$

$$\Rightarrow X_n+Y_n \longrightarrow X+Y \text{ a.s.}$$

(3) $X_n \xrightarrow{a.s.} X \Rightarrow f(X_n) \xrightarrow{a.s} f(X)$

if f – continuous.

Proof: Exercise.

(4) $X_n \xrightarrow{a.s.} X$, $P(X=0)=0$. ie, $P(X\neq 0)=1$.

ie, X is non-zero a.s.

ie, $\frac{1}{X}$ is defined a.s.

then, $\frac{1}{X_n} \xrightarrow{a.s} \frac{1}{X}$.

Proof: Exercise.

[Here, note: $\frac{1}{X_n}$ are not "just" real r.v, but extended real r.v.]

Probability-3 Lecture-6

09 August 2024 10:16

Definition: $X_n \xrightarrow{a.s} X$ if
$P(\{w : X_n(w) - X(w)\}) = 1.$
ie, $\exists$ a P- null set N s.t.
for $w \notin N$, $X_n(w) \longrightarrow X(w)$.

* $X_n \xrightarrow{a.s} X$, $Y_n \xrightarrow{a.s} Y \Rightarrow$
  - $X_n + Y_n \xrightarrow{a.s} X + Y$.
  - $X_n \cdot Y_n \xrightarrow{a.s} X \cdot Y$.

* $\left.\begin{array}{l} X_n \xrightarrow{a.s} X \\ X_n \xrightarrow{a.s} Y \end{array}\right\} \Rightarrow X = Y$ a.s .

* $X_n \xrightarrow{a.s} X \Rightarrow f(X_n) \xrightarrow{a.s} f(X)$ for any continuous function f.

* $X_n \xrightarrow{a.s} X$, $X \neq 0$ a.s.
$\Rightarrow \frac{1}{X_n} \longrightarrow \frac{1}{X}$.

Proof: $P(X \neq 0) = 1$. &, $X_n \xrightarrow{a.s} X$
$\Rightarrow P\left(\bigcup_k \{|X| > 1/k\} \cap \{X_n \to X\}\right) = 1$

intersection of 2 sets having probability 1.

$$\Rightarrow P\Big(\underbrace{\bigcup_k \{|X| > 1/k\} \cap \{X_n \to X\}}_{A}\Big) = 1$$

&,
$\left\{\frac{1}{X_n} \longrightarrow \frac{1}{X}\right\} = B$ (say).

$A \subseteq B$. [think why?]

$\Rightarrow P\left(\left\{\frac{1}{X_n} \longrightarrow \frac{1}{X}\right\}\right) = 1$ ▨

Definition: (Converges in Probability)
$\{X_n\}$ - real r.vs.
X real r.v.

$\{X_n\}$ - real r.v.s.

X - real r.v.

Say that, $\{X_n\}$ converges in probability to X,
if $\forall \varepsilon > 0$, $\forall \delta > 0$, $\exists\, n_0 \in \mathbb{N}$ s.t $\forall n > n_0$,

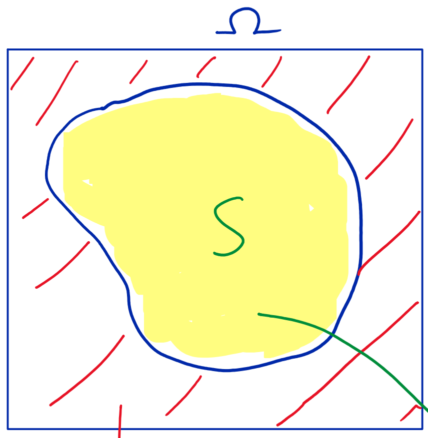
$$P(|X_n - X| > \varepsilon) < \delta .$$

ie, if $\forall \varepsilon > 0$, $P(|X_n - X| > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$

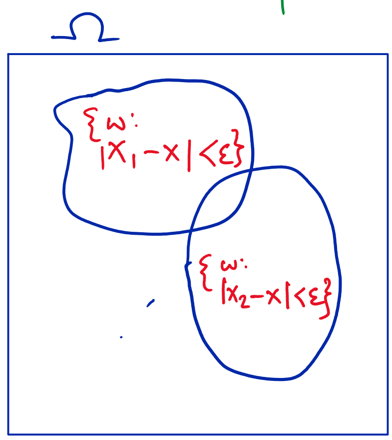
Notice that:
for $\varepsilon' = \min\{\varepsilon, \delta\}$, this holds.

Notation: $X_n \xrightarrow{P} X$

Bit of explanation (a.s convergence v/s convergence in probability)



this shaded area is that one P-Null set.
If we "throw this away"
in this
$\forall w \in$ remaining set $= S$ (say)
$X_n(w) \longrightarrow X(w)$. $P(S^c) = 0$.



fix $\varepsilon > 0$.
we look at the sets
$\{w: |X_i - X| < \varepsilon\}$.
& hence, the sequence
$P(\{w: |X_i - X| < \varepsilon\})$
converges.

$*$ $\left.\begin{array}{l} X_n \xrightarrow{P} X \\ X_n \xrightarrow{P} Y \end{array}\right\} \Rightarrow X = Y$

replacement for ε.

ie, to show, $P(X \neq Y) = 0 \iff P\left(\bigcup_k \left\{|X - Y| > 1/k\right\}\right) = 0$

countable

Proof: fix $\varepsilon > 0$.

ie, to show, $P(X \neq Y) = 0 \Leftrightarrow P(\bigcup_k \ldots)$

countable union.

So, we just need to show,

$P(\{|X-Y| > 1/k\}) = 0 \ \forall k$

[then, by countable additivity, the result would follow.]

Proof: fix $\varepsilon > 0$.

$$P(|X-Y| > \varepsilon) = 0.$$

$$P(\underbrace{|X-Y| > \varepsilon}_{A}) \leq P(\underbrace{|X_n - X| > \varepsilon/2}_{B}) + P(\underbrace{|X_n - Y| > \varepsilon/2}_{C})$$

to show this, its enough to show:

$A \subset B \cup C$,

or $A^c \supset B^c \cap C^c \rightarrow$ we use Δ-inequality to show this.

Here, $B^c =$ event that $|X_n - X| \leq \varepsilon/2 = \{\omega : |X_n(\omega) - X(\omega)| \leq \varepsilon/2\}$.

$C^c =$ event that $|X_n - Y| \leq \varepsilon/2 = \{\omega : |X_n(\omega) - Y(\omega)| \leq \varepsilon/2\}$.

$\therefore$ By Δ-inequality,

$$|X-Y| \leq \underbrace{|X_n - X|}_{<\varepsilon/2} + \underbrace{|X_n - Y|}_{<\varepsilon/2} < \varepsilon.$$

$\therefore A \subset B \cup C$.

$$\therefore P(A) \leq P(B \cup C) \leq \underbrace{P(B)}_{\substack{<\delta/2 \\ \forall n > N_1}} + \underbrace{P(C)}_{\substack{<\delta/2 \\ \forall n > N_2}} < \delta.$$

$[\forall n > N = \text{Max.}\{N_1, N_2\}]$

(*) $X_n \xrightarrow{P} X \Rightarrow c \cdot X_n \xrightarrow{P} cX$.

Case-I:

$c = 0$.

(trivial: nothing to prove)

Case-II.

$c \neq 0$.

then, for $\varepsilon > 0$.

$$P(|cX_n - cX| > \varepsilon)$$
$$= P\left(|X_n - X| > \varepsilon/|c|\right) \rightarrow 0$$

as $X_n \xrightarrow{P} X$

(*) $\left.\begin{array}{l} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{array}\right\} X_n + Y_n \xrightarrow{P} X + Y$.

$$P(\underbrace{|X_n + Y_n - X - Y|}_{A} > \varepsilon)$$

$\ldots P(|X_n - X| > \ldots) + P(|Y_n - Y| > \varepsilon/\ldots)$

$$P(|X_n+Y_n-X-Y|>\varepsilon)$$

$$\leq P(\underbrace{|X_n-X|>\varepsilon/2}_{B}) + P(\underbrace{|Y_n-Y|>\varepsilon/2}_{C})$$

Again, enough to show, $A \subset B \cup C$

$\Rightarrow A^c \supset B^c \cap C^c$

B^c = event that $|X_n-X| \leq \varepsilon/2$

C^c = event that $|Y_n-Y| \leq \varepsilon/2$.

Clearly, $|X_n-X| \leq \varepsilon/2$ & $|Y_n-Y| \leq \varepsilon/2$

Again, by Δ-inequality, $|X_n+Y_n-X-Y| < \underbrace{|X_n-X|}_{\leq \varepsilon/2} + \underbrace{|Y_n-Y|}_{\leq \varepsilon/2} < \varepsilon$.

$\therefore A \subset B \cup C$.

$\therefore P(A) \leq P(B \cup C) \leq \underbrace{P(B)}_{<\delta/2 \ \forall n>N_1} + \underbrace{P(C)}_{<\delta/2 \ \forall n>N_2} < \delta$. $[\forall n > N = \text{Max}\{N_1, N_2\}]$ ▨

(*) $X_n \xrightarrow{P} X \Rightarrow f(X_n) \xrightarrow{P} f(X)$ for any continuous function f.

Proof: We're going to show, that

$X_n \xrightarrow{P} X$ implies: why compact set? $\because$ on a compact set, every cont. f^n is bounded

for every $\varepsilon > 0$, we can find atleast one compact set $K_\varepsilon \subset \mathbb{R}$ s.t $\forall n \geq 1$,

Called "Tightness Property". (we'll prove this later)

$$P(X_n \in K_\varepsilon, X \in K_\varepsilon) > 1 - \varepsilon/2$$

$$\Leftrightarrow P(X_n \notin K_\varepsilon \text{ or } X \notin K) < \varepsilon/2$$

for the time being, we'll assume: proving this (above) $\Rightarrow$ we're done.

So, for $\varepsilon > 0$.

$$P(|f(X_n) - f(X)| > \varepsilon).$$

Get K_ε as stated:

$$P(|f(X_n) - f(X)|)$$

$$P(|f(X_n) - f(X)|)$$

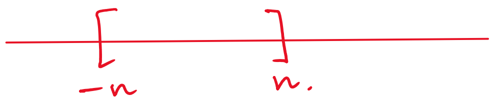
$$= P(X_n \notin K_\varepsilon \text{ or } X \notin K_\varepsilon,\ |f(X_n)-f(X)| > \varepsilon)$$

$$+ P(X_n \in K_\varepsilon, X \in K_\varepsilon,\ |f(X_n)-f(X)|>\varepsilon)$$

$$< \varepsilon/2 + P(|X_n - X| > \delta),$$ where $\delta > 0$ is s.t, $\forall\ x, y \in K_\varepsilon$, $|x-y| < \delta$ $\Rightarrow |f(x) - f(y)| < \varepsilon$

[$\because$ On a compact set. f continuous $\Rightarrow$ f-uniformly continuous]

Notion:



* Given a real r.v X, we can find a compact set K_ε s.t $P(X \in K_\varepsilon) < 1 - \varepsilon$.

ie, $P(X \in \mathbb{R}) = 1$,

& $\bigcup_n [-n, n] = \mathbb{R}$. ie, $[-n,n] \nearrow$ to $\mathbb{R}$

$\therefore\ P(X \in [-n,n]) \nearrow 1$

s.t, $\forall\ 0 < \varepsilon < 1$, $\exists\ N$ s.t., $\forall\ n > N$,
$$P(X \in [-n, n]) > 1 - \varepsilon.$$

Exercise:

Seq. $\{X_n\}$.

$\exists$ compact set K_n . s.t. $P(X \in K_n) > 1 - \varepsilon$.

prove that:

You cannot find a single $K \subset \mathbb{R}$ (a compact set) s.t. $\forall\ n$, $P(X_n \in K) > 1 - \varepsilon$.

Now, we prove our assumption: $\hookrightarrow$ that tightness $\Rightarrow$ "we're done."

Get $M \in \mathbb{N}$ s.t.
$$\alpha = P(X \in [-M, M]) > 1 - \varepsilon/2$$

$$X_n \xrightarrow{P} X \Rightarrow P(|X_n - X| > 1) \longrightarrow 0.$$

get n_0 s.t,
$$P(|X_n - X| > 1) < \alpha - (1 - \varepsilon/2) \quad \forall\ n \geq n_0$$

get n_0 s.t,

$$P(|X_n - X| > 1) < \alpha - (1 - \varepsilon/2) \quad \forall n \geq n_0$$

$\therefore$ for $n \geq n_0$,

$$* \quad \underbrace{P(|X_n| > M+1)}_{A} \leq \underbrace{P(|X_n - X| > 1, |X| \leq M)}_{B} + \underbrace{P(|X| > M)}_{C}$$

$B \to \alpha - (1 - \varepsilon/2)$, $C \to 1 - \alpha$. $< \varepsilon/2$

Again, we'll try using the same argument.

Explanation.

$$* \quad P(|X_n| > M+1) = P(|X_n| > M+1, |X| \leq M) + P(|X_n| > M+1, |X| > M)$$

$\Longleftarrow$ $P(|X_n - X| > 1)$.

$|X_n| > M+1 \Rightarrow |X| > M$.

$\therefore P(|X_n| > M+1) < \varepsilon/2$

$\Rightarrow P(|X_n| \leq M+1) > 1 - \varepsilon/2$

$\therefore$ We found an $M_0 \in \mathbb{R}$ s.t. for $\varepsilon > 0$,

$$P(X \in [-M_0, M_0]) > 1 - \varepsilon/2,$$

$M_0 = \text{Max.}\{M, M+1\}$.

(*). $X_n \xrightarrow{P} X \Rightarrow X_n^2 \xrightarrow{P} X$

$$\left.\begin{array}{l} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{array}\right\} X_n \cdot Y_n \xrightarrow{P} X \cdot Y$$

Proof: $X_n \cdot Y_n = \frac{1}{4} \cdot \left[(X_n + Y_n)^2 - (X_n - Y_n)^2 \right]$.

$$\downarrow P \qquad \downarrow P$$

$$(X+Y)^2 - (X-Y)^2 = XY. \quad \blacksquare$$

Exercise: $X_n \xrightarrow{P} X$ $\quad X \neq 0$ a.s.

Prove that: $\frac{1}{X_n} \xrightarrow{P} \frac{1}{X}$.

(*) $X_n \xrightarrow{a.s} X$

$$\iff P\left(\bigcap_{j=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k\geq n}\left\{\omega : |X_k(\omega)-X(\omega)| < \frac{1}{j}\right\}\right) = 1$$

take complement

$$\iff P\left(\bigcup_{j=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k\geq n}\left\{\omega : |X_k(\omega)-X(\omega)| > \frac{1}{j}\right\}\right) = 0$$

$$\iff P\left(\bigcap_{n=1}^{\infty}\bigcup_{k\geq n}\left\{\omega : |X_k(\omega)-X(\omega)| > \frac{1}{j}\right\}\right) = 0 \quad \forall j\in\mathbb{N}$$

$$\iff P\left(\bigcap_{n=1}^{\infty}\left\{\sup_{k\geq n}|X_k(\omega)-X(\omega)| > \frac{1}{j}\right\}\right) = 0 \quad \forall j\in\mathbb{N}$$

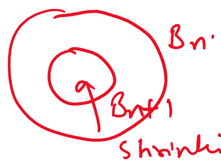
(*)

$$\iff P\left(\sup_{k\geq n}|X_k - X| > 1/j\right) \longrightarrow 0 \quad \text{as } n\longrightarrow\infty \quad \forall j\in\mathbb{N}$$

(* if $B_n = \left\{\sup_{k\geq n}|X_k(\omega)-X(\omega)| > 1/j\right\}$,

& $B_{n+1} = \left\{\sup_{k\geq n+1}|X_k(\omega)-X(\omega)| > 1/j\right\}$,

then, $B_{n+1}\subseteq B_n$. ∴



shrinking, so, $n\to\infty$ makes sense.)

$$\iff P\left(\sup_{k\geq n}|X_k - X| > \varepsilon\right) \longrightarrow 0 \quad \forall\, \varepsilon > 0.$$

$$\iff \forall\, \varepsilon>0,\ P\left(\sup_{k\geq n}|X_n - X| > \varepsilon\right) \longrightarrow 0$$

$$\equiv \sup_{k\geq n}|X_n - X| \xrightarrow{P} 0.$$

∴ <u>Theorem:</u>

$$X_n \xrightarrow{a.s.} X \quad \text{iff} \quad \sup_{k\geq n}|X_k - X| \xrightarrow{P} 0$$

$\Downarrow$

P [∵ sup → 0]

$$\Downarrow$$

$$|X_k - X| \xrightarrow{P} 0 \quad [\because \sup \to 0]$$

ie, a.s. convergence $\Rightarrow$ convergence in p.

However, converse is <u>NOT</u> true !!

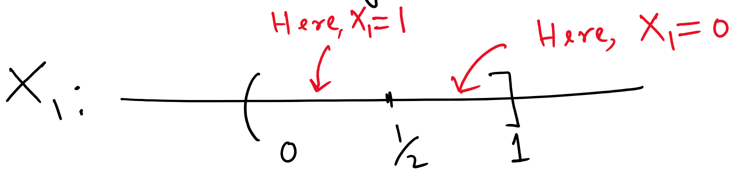
Counter eg:

to show: $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{a.s} X$

Take $\Omega = (0,1]$. $\mathcal{A}$ = Borel sets on $\mathbb{R}$.

P = Lebesgue measure.

fix we define X_1 & X_2



ie, $X_1 := \mathbb{1}_{(0,1/2]}$, &

$X_2 := \mathbb{1}_{(1/2,1]}$.

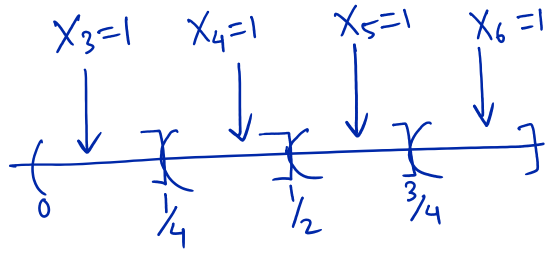
Now, we define X_3, X_4, X_5, X_6.

$X_3 := \mathbb{1}_{(0,1/4]}$

$X_4 := \mathbb{1}_{(1/4,1/2]}$

$X_5 := \mathbb{1}_{(1/2,3/4]}$

$X_6 := \mathbb{1}_{(3/4,1]}$.



Now, it's clear how next 8 X_i's would be defined.

$\therefore$ from the defn, the seqn $X_n \xrightarrow{P} 0$.

Q. Does $X_n \xrightarrow{a.s} 0$?

— NO !!! a.s .

Q. Does $X_n \xrightarrow{\cdot\cdot} 0$?

– NO!!! $X_n \xrightarrow{a.s} 0$ ~~X~~.

for any $\omega \in (0,1]$,

Exercise:

to show: in a discrete probability space (ie, Ω – countable).

$$X_n \xrightarrow{a.s} X \iff X_n \xrightarrow{P} X$$

(ie, the reverse implication becomes true in a discrete probability space.)

"Strict non-atomicity" property:

A – a set.

if $P(A) = \alpha > 0$,

$\forall \beta \in (0, \alpha)$ $\exists$ $B \subseteq A$ s.t. $P(B) = \beta$.

("kind of" I.V.T in probability).

Probability-3 Lecture-7

13 August 2024 14:22

$(\Omega, \mathcal{A}, P)$ - probability space.

$X_n, n \geq 1$. - seq. of real r.vs on $(\Omega, \mathcal{A}, P)$.

Defn: $X_n \xrightarrow{P} X$ if $\forall \varepsilon > 0$, $P(|X_n - X| > \varepsilon) \longrightarrow 0$ as $n \longrightarrow \infty$.

ie, $\forall \varepsilon > 0$, $\forall \delta > 0$,

$\exists n_0$ st $\forall n > n_0$,

$P(|X_n - X| > \varepsilon) < \delta$

(Enough to this for $\varepsilon = \delta$)

(*) $\left.\begin{array}{l} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{array}\right\} X = Y$.

(*) $\left.\begin{array}{l} X_n \longrightarrow X \\ Y_n \longrightarrow Y \end{array}\right\} \Rightarrow$ • $cX_n \xrightarrow{P} cX \quad \forall c \in \mathbb{R}$

• $X_n + Y_n \longrightarrow X + Y$.

• $X_n \cdot Y_n \longrightarrow X \cdot Y$.

$\underset{D}{\underbrace{|a+b| > \varepsilon}} \Rightarrow$ either $\overset{A}{\overbrace{|a| > \varepsilon/2}}$, or $\underset{= B}{\underbrace{|b| > \varepsilon/2}}$.

$D \subset A \cup B \Rightarrow$

$P(D) \leq P(A) + P(B)$

(*) $X_n \xrightarrow{P} X$

$\Rightarrow f(X_n) \xrightarrow{P} f(X)$ for any continuous function f.

if this is proved,

then, $X_n Y_n = \frac{1}{4}\left((X_n+Y_n)^2 - (X_n - Y_n)^2\right)$

then, $\left.\begin{array}{l} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{array}\right\} X_n Y_n \xrightarrow{P} XY$

becomes obvious by simple algebra.

quick review of the proof:

"Tightness":

Suppose, Y is a real r.v.

$P(|Y| > M) \longrightarrow 0$ as $M \to \infty$.

$\therefore \forall \varepsilon > 0$, $\exists M_\varepsilon$ s.t. $P(|Y| > M_\varepsilon) < \varepsilon$.

$Y_1, Y_2, \ldots, Y_k$ are real r.v's

$\therefore \forall i = 1, \ldots, k$,

$\therefore \forall i = 1, \dots, k,$

$\exists M_i = M_i(\varepsilon)$ s.t. $P(|Y_i| > M_i) < \varepsilon$.

$\therefore M = \text{Max}\{M_1, \dots, M_k\}$.

$\Rightarrow P(|Y_i| > M) < \varepsilon.\ \forall\ i = 1, \dots, k$

We did this for a finite no. of r.vs.

Q: Given a sequence of $\{Y_n\}$ of real random r.vs,

(exercise given in Lec-6) Can we get, $\forall\ \varepsilon > 0$, an M_ε s.t. $P(|Y_i| > M_\varepsilon) < \varepsilon$?

In general, NO!!!

→ Eg: $Y_n = \begin{cases} n & \text{with prob.} = 1/2 \\ -n & \text{with prob} = 1/2 \end{cases}$

Here, for every M,

$$P(|Y_n| > M) = 1 \quad \forall\ n \geq M.$$

So, here's a counter example.

But, what if, $\{Y_n\}$ is such a sequence s.t. $Y_n \xrightarrow{P} Y$, then? (this is what we did in last class)

Definition: (Tightness)

A sequence of $\{X_n\}$ - real r.v is said to be tight if $\forall\ \varepsilon > 0$, $\exists\ M = M_\varepsilon$, such that:

Independent of n

$$P(|X_n| > M) < \varepsilon \quad \forall\ n.$$

ie, the tail probability goes to 0 uniformly.

What we proved last time (lec-6) is:

$$X_n \xrightarrow{P} X \Rightarrow \{X_n\} \text{ is tightness.}$$

Proof:

Let $\varepsilon > 0$ be given.

X is a real r.v., get M_0 st $P(|X| > M_0) < \varepsilon$.

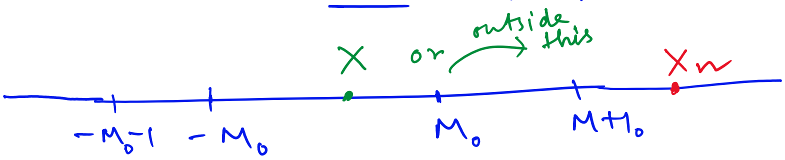
look at: $P(|X_n| > M_0 + 1) \leq P(|X| > M_0) + P(|X_n - X| > 1)$

look at: $P(|X_n| > M_0 + 1) \leq P(|X| > M_0) + P(|X_n - X| > 1)$

(just a convenient choice. i.e. could have been something else as well.)

Claim: $|X_n| > M_0 + 1 \Rightarrow$ Either $|X| > M_0$ or $|X_n - X| > 1$

$\hookrightarrow$ when $|X| \leq M_0$



So, $P(|X_n| > M_0+1) \leq \underbrace{P(|X| > M_0)}_{< \varepsilon} + \underline{P(|X_n - X| > 1)}$

we need this small

So, get n_0 large enough, st, RHS $< \varepsilon$.

$\therefore P(|X| > M_0 + 1) < \varepsilon$

$\& \; P(|X_n| > M_0+1) < \varepsilon \;\; \forall \, n \geq n_0$.

$\&$ $X_1, \ldots, X_{n_0 - 1}$ are finitely many. So, we can take care of that.

Result: $X_n \xrightarrow{a.s} X \Leftrightarrow \sup_{k \geq n} |X_k - X| \xrightarrow{p} 0$

$\Rightarrow |X_k - X| \xrightarrow{p} 0 \Leftrightarrow X_n \xrightarrow{p} X$

So, almost sure converges

$\Downarrow$

converges in probability.

The converse is NOT true!!! (Refer L-6 for counter example).

(*) One special case, where $X_n \xrightarrow{a.s} X \Leftarrow\Rightarrow X_n \xrightarrow{a.s} X$

(*) One special case, where $X_n \xrightarrow{a.s} X \iff X_n \xrightarrow{a.s} X$ is when we have a Discrete Probability Space.

(Proof: Exercise)

Theorem:

$X_n \xrightarrow{P} X \implies \exists$ a subsequence $X_{n_k} \xrightarrow{a.s} X$

ie, convergence in $P \not\Rightarrow$ a.s convergence, but we can get a subsequence that exhibits a.s convergence.

Proof: $X_n \xrightarrow{P} X$

$\therefore$ By definition, $\forall k \geq 1, \exists n_k \geq 1$ s.t.

$$P(|X_m - X| > 2^{-k}) < 2^{-k} \quad \forall m \geq n_k$$

We may, & do assume $1 \leq n_1 < n_2 < \cdots < n_k < n_{k+1} < \cdots$

Denote $A_k = \{|X_{n_k} - X| > 2^{-k}\}$,
event

then, $P(A_k) < 2^{-k}$ $[\because n_k > k]$

$$\therefore \sum_k P(A_k) < \infty$$

(Recall)

* Borel-Cantelli Lemma:

If $A_k, k \geq 1$ - events on the same probability space.

s.t, $\sum_k P(A_k) < \infty$.

Then, $P(\limsup_k A_k) = 0$.

ie, $P\left(\bigcap_{k\geq 1} \bigcup_{j \geq k} A_j\right) = 0$.

(just a consequence of continuity of Probability)

$\Rightarrow \lim_{k \to \infty} P\left(\bigcup_{j\geq k} A_j\right)$

$\therefore$ By Borel-Cantelli lemma—

$\therefore$ By Borel-Cantelli lemma —

$$P\left(\bigcap_{k} \bigcup_{k' \geq k} \left\{ |X_{n_{k'}} - X| > 2^{-k'} \right\}\right) = 0.$$

$$\Rightarrow P\underbrace{\left(\bigcup_{k\geq 1} \bigcap_{k' \geq k} \left\{ |X_{n_{k'}} - X| \leq 2^{-k'} \right\}\right)}_{A \text{ (say)}.} = 1.$$

then, $\forall\ \omega \in A$, $X_{n_k}(\omega) \longrightarrow X(\omega)$

$\Downarrow$

$\exists\ k_0 = k_0(\omega)$

comes from $\bigcup_{k\geq 1}$

s.t. $\left| X_{n_{k'}}(\omega) - X(\omega) \right| \leq 2^{-k'}$ $\forall\ k' \geq k_0(\omega)$

comes from $\bigcap_{k' \geq k_0}$

$\therefore$ This has reduced to the definition of a.s. convergence.

<u>Theorem</u>:

$X_n \xrightarrow{P} X \iff$ For every subsequence $\{X_{n_k}\} \subseteq \{X_n\}$ $\exists$ a further subsequence $\{X_{n_{k''}}\}$ s.t. $X_{n_{k''}} \xrightarrow{a.s} X$.

$(\Rightarrow)$

* clearly, $X_n \xrightarrow{P} X \Rightarrow X_{n_k} \xrightarrow{P} X$. $\therefore$ By prev. theorem, $\exists\ \{X_{n_{k''}}\} \subseteq \{X_{n_k}\}$ s.t. $X_{n_{k''}} \xrightarrow{a.s.} X$

(proof: exercise)

$(\Leftarrow)$ Fix $\varepsilon > 0$.

Have to show:

$$a_n = P(|X_n - X| > \varepsilon) \longrightarrow 0$$

Now, let $\{a_{n_k}\}$ a subsequence of $\{a_n\}$.

$$\therefore a_{n_k} = P(|X_{n_k} - X| > \varepsilon)$$

$\therefore \exists\ n_{k''}$ s.t. $X_{n_{k''}} \xrightarrow{a.s.} X$. $\Rightarrow X_{n_{k''}} \xrightarrow{P} X$

$\therefore \exists n_k''$ s.t. $X_{n_k''} \xrightarrow{a.s.} X. \Rightarrow X_{n_k''} \xrightarrow{P} X$

$\Rightarrow a_{n_k''} \to 0$.

Corollary: $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$ $\forall$ continuous f^n g.
(Exercise)

Definition: (Convergence in the p^{th} moment).

$\{X_n\}$ is said to converge to X in L_p (or, in p^{th} moment)

if $\|X_n - X\|_p \longrightarrow 0$.

$$\left(E|X_n - X|^p \longrightarrow 0\right).$$

We write, $X_n \xrightarrow{L_p} X$

$\left.\begin{array}{l} X_n \xrightarrow{L_p} X \\ X_n \xrightarrow{L_p} Y \end{array}\right\}$

$\Rightarrow X = Y$ a.s.

why? By Minkowski's inequality,

$\|X-Y\|_p \le \underbrace{\|X_n - X\|_p}_{\downarrow\, 0} + \underbrace{\|X_n - Y\|_p}_{\downarrow\, 0}$.

$\Rightarrow \|X-Y\|_p \to 0$

$\Rightarrow X = Y$ a.s.

(*) $X_n \xrightarrow{L_p} X \Rightarrow cX_n \longrightarrow cX$

(*) $X_n \xrightarrow{L_p} X, \; Y_n \xrightarrow{L_p} Y$

$\Rightarrow X_n + Y_n \xrightarrow{L_p} X + Y$.

(*) $X_n \xrightarrow{L_p} X$.

$\therefore P(|X_n - X| > \varepsilon) \le \dfrac{\|X_n - X\|_p^p}{\varepsilon^p} \longrightarrow 0$.

$\| \; P(|X_n - X|^p > \varepsilon^p)$ [By Chebyshev's inequality]

$\Rightarrow X_n \xrightarrow{P} X$. ie, convergence in L_p $\Rightarrow$ convergence in p.

However, converge is not true!!

Counter example.

α_n

0 — $1/n$ — 1]

$X_n = \alpha_n \cdot \mathbb{1}_{(0, 1/n]}$, $\alpha_n > 0$

$X_n = \alpha_n \cdot \mathbb{1}_{(0, 1/n]}$, $\alpha_n > 0$

$\therefore X_n \xrightarrow{p} 0$

$$E(|X_n|^p) = \alpha_n^p \cdot \frac{1}{n}.$$

Clearly, we can choose α_n such that this does not converge.

infact, we can choose α_n s.t. $E(|X_n|^p)$ diverges to ∞.

So, convergence in $p \not\Rightarrow$ convergence in L_p

$\therefore$ Convergence in probability is the "weakest" in a sense.

Probability-3 Lecture-8

20 August 2024 14:19

"Learn it, give exams, but when it comes to real life, Probability doesn't matter." — Prof. AG, 20th Aug. '24

$(\Omega, \mathcal{A}, P)$

$p > 0$.

Defn: $L_p := \{X \text{ - r.v} : E|X|^p < \infty\}$ → set of all r.vs with finite p^{th} moment.

Firstly, L_p is a Vector Space.

- Scalar multiplication & pt-wise addition are to be checked

$X \in L_p \Rightarrow cX \in L_p$ ✓ (trivial)

Now if $X, Y \in L_p$, does $X+Y \in L_p$

$$\|X+Y\|_p = \left(E|X+Y|^p\right)^{1/p} \leq \|X\|_p + \|Y\|_p < \infty$$

[By Minkowski's inequality]

$\therefore X+Y \in L_p$ ✓.

$\therefore L_p$ is a vector space. ▨

Inequality: for any two real nos. a, b.

$$|a+b|^p \leq |a|^p + |b|^p \quad (*)$$

Then: for $\omega \in \Omega$, $X(\omega), Y(\omega) \in \mathbb{R}$. $a = X(\omega)$, $b = Y(\omega)$

$$\therefore |X(\omega) + Y(\omega)|^p \leq |X(\omega)|^p + |Y(\omega)|^p.$$

$\therefore$ Taking Expectation on both sides:

$$E|X(\omega) + Y(\omega)|^p \leq \underbrace{E|X(\omega)|^p}_{<\infty} + \underbrace{E|Y(\omega)|^p}_{<\infty} \quad \forall \omega \in \Omega.$$

Proof: Case-I if any of $a = 0$ or $b = 0$, then nothing to prove!! (trivial)

Case-II $a \neq 0,\ b \neq 0$.

for (*), enough to prove $(|a|+|b|)^p \leq \underbrace{|a|^p+|b|^p}_{(**)}$.

then (*) follows by Δ-inequality.

WLOG, assume

$$0 < |a| \leq |b|$$

$$\therefore r = \frac{|b|}{|a|} \geq 1.$$

$f(r) = (1+r)^p - (1+r^p)$

Exercise: Show that $f'(r) < 0$, i.e, $f(r) \downarrow$ $r \geq 1$.

$\Rightarrow f(r) \leq f(1)$
$= 2^p - 2 \leq 0$
$[\because 0 < p \leq 1]$

$\therefore$ LHS of (**)

$$= |a|^p (1+r)^p$$

RHS of (**)

$$= |a|^p (1+r^p)$$

$\therefore$ L_p is a vector space $\forall p > 0$.

But specifically: L_p is a Normed Linear Space $\forall p \geq 1$.

Definition: (Convergence in L_p)

For $p > 0$, we say that $\{X_n\}$ converges to X in L_p norm

OR,

$\{X_n\}$ converges in p^{th} moment

[for $p=1$, $\{X_n\}$ converges in mean]

if $X_n \in L_p,\ n \geq 1$.

& $E|X_n - X|^p \longrightarrow 0$.

Observe:

* $X_n, n \geq 1$ in L_p and $X_n \xrightarrow{L_p} X \Rightarrow X \in L_p$. (Proof: Exercise)

* $X_n \xrightarrow{L_p} X \Rightarrow cX_n \xrightarrow{L_p} cX$.

* $X_n \xrightarrow{L_p} X,\ Y_n \xrightarrow{L_p} Y \Rightarrow X_n + Y_n \xrightarrow{L_p} X + Y$

$*\quad X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{p} X$

[Proof by chebyshev's inequality : $\forall Z$ - r.v. $P(|Z|>\lambda) \leq \frac{E(|Z|)}{\lambda}$

$P(|X_n - X| > \varepsilon) \leq \frac{E|X_n - X|^p}{\varepsilon^p} \to 0$]

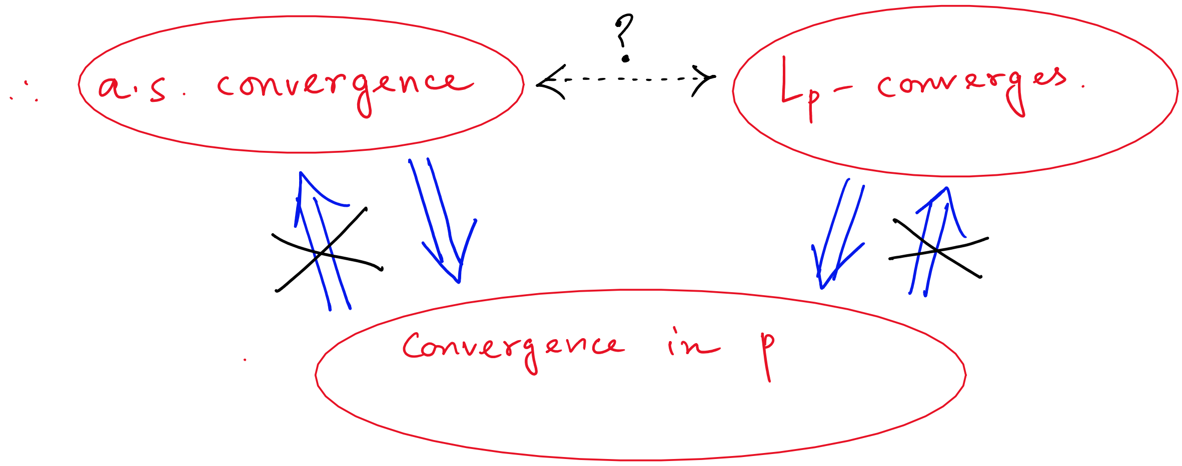
$\therefore X_n \xrightarrow{p} X$.

$*\quad X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{L_r} X \quad \forall\ r \leq p.$

[Consequence of Holder's Inequality.

We had proved: rth moment $\leq$ pth moment $\forall\ r \leq p$.]

What we did so far.....



We already know,

L_p convergence $\not\Rightarrow$ a.s. convergence

Recall that example of convergence in p $\not\Rightarrow$ a.s convergence.

Also, Eg...

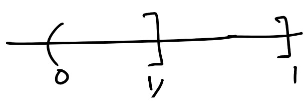
$\left((0,1], \mathcal{B}(0,1], P = leb\right)$

$$X_n = n^{\alpha} \cdot 1_{(0, 1/n]} \xrightarrow{a.s.} X \equiv 0 .$$

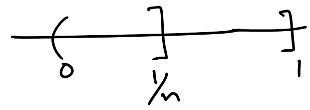
take $\omega \in (0,1]$.

$\Rightarrow \exists\ n_0$ s.t $1/n < \omega$, $\forall\ n \geq n_0$

$X_n = 0$.



$$X_n = 0.$$



$\therefore \forall \omega, \exists$ such n_0 exists.

$$\therefore X_n \xrightarrow{a.s} 0$$

BUT: $E\,|X_n|^p = n^{\alpha p} \cdot P(X \in (0, 1/n])$

$= n^{\alpha p} \cdot \frac{1}{n}. \not\longrightarrow 0$ for certain values of α.

Precisely, $\alpha p > 1$. $\Rightarrow \boxed{\alpha > 1/p}$

$\therefore$ a.s convergence $\not\Longrightarrow$ L.p convergence

L.p convergence $\not\Longrightarrow$ a.s convergence.

Now, we only consider $\boxed{p \geq 1}$.

We define $\|X\|_p = (E|X|^p)^{1/p}$

Minkowski $\Rightarrow \|\cdot\|_p$ is a norm

(Modulo identification)

(ie, 2 r.vs are "same" if they are equal a.s., ie, equal except over a measure zero set.)

Result: $\|\cdot\|_p$ is complete.

Proof: let $X_n, n \geq 1$ be a Cauchy sequence in L_p.

We can get a subsequence: $1 \leq n_1 < n_2 < \cdots < n_k < n_{k+1} < \cdots$ such that, for each $k \geq 1$,

$$\|X_m - X_n\|_p < 3^{-k} \quad \forall m, n \geq n_k.$$

Now, we're going to show, this subsequence converges a.s.

$\therefore \forall k \geq 1$,

$$P(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) \underset{\text{(Chebyshev)}}{\leq} \frac{E\,|X_{n_{k+1}} - X_{n_k}|^p}{2^{-k\cdot p}}$$

$$= \frac{\left(\|X_{n_{k+1}} - X_{n_k}\|_p\right)^p}{2^{-kp}}$$

$$\leq \frac{3^{-kp}}{\ } \quad (2^p)^k$$

$$\leq \frac{3^{-kp}}{2^{-kp}} = \left(\frac{2^p}{3^p}\right)^k < \infty$$

2^{-kp}

$\therefore$ This is a convergent series.

$$\Rightarrow \sum_k P(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}) < \infty.$$

$\therefore$ By Borel-Cantelli Lemma,

$$P\left(|X_{n_{k+1}} - X_{n_k}| > 2^{-k} \text{ for infinitely many } k\right) = 0$$

converse

$$\Rightarrow P\left(|X_{n_{k+1}} - X_{n_k}| \leq 2^{-k} \text{ for all large } k\right) = 1$$

(will hold for finitely many k's. So, won't hold for large k's.)

$\Rightarrow X = \lim X_{n_k}$ exists P- a.s.

Now, take $X = \limsup X_{n_k}$

$$X_{n_k} \xrightarrow{a.s.} X.$$

[Every Cauchy seq. in L_p has a subsequence which convergences a.s.]

Now, we're left to show, this "X" is the limit of the entire sequence.

ie, we'll show, $X_{n_k} \xrightarrow{L_p} X$. Then,

{ Sequence - Cauchy
Subsequence - Converges to X

$\Rightarrow$ Entire Sequence converges to X.

$\because$ a.s convergence holds
$\therefore \limsup X_{n_k} \equiv \liminf X_{n_k} \equiv$ limit X.

$$E(|X|^p) = E\left(\liminf |X_{n_k}|^p\right)$$

[Fatou's lemma] $\leq \liminf \left(E|X_{n_k}|^p\right) < \infty$

$\left[\begin{array}{l}\text{Fatou's} \\ \text{lemma}\end{array}\right] \leq \liminf \left(E\, |X_{n_k}|\right) < \infty$

$\Rightarrow \sup_{n} \|X_n\|^p < \infty$ $\quad \left[\because X_n \text{ is Cauchy in } L_p, \therefore X_n \text{ is bounded in } L_p.\right]$

$\therefore X \in L_p.$ ✓

Now, fix k.

$\forall\, j > k, \quad E\,|X_{n_j} - X_{n_k}|^p \leq 3^{-kp}$

$n_j > n_k$

$\because X_{n_j} \xrightarrow{a.s} X.$

Let $j \nearrow \infty$.

(remember, k is fixed)

$\therefore$ Again, by Fatou's lemma,

$$E\,|X - X_{n_k}|^p \leq 3^{-kp}.$$

Now, let $k \nearrow \infty$.

$$\therefore E\,|X - X_{n_k}|^p \longrightarrow 0.$$

Ques: trying to build something "extra" to build the bridge bet^n a.s convergence & L_p convergence.

Also, conv. in p + "Extra" $\equiv$ a.s convergence.

Some rephrasing: X - r.v

"X is integrable" $\equiv E|X| < \infty$. (finite expectation)

↳ just a name!!!

Exercise:

$\Leftrightarrow E\left(|X| \cdot 1_{|X| > \lambda}\right) \longrightarrow 0$ as $\lambda \longrightarrow \infty$.

$\Rightarrow$ DCT

$\Leftarrow$ (trivial).

(Sort of "tail integral" $\to 0$)

Definition:

A sequence of r.vs $\{X_n\}$ is said to be **uniformly integrable (u.i.)** if

$$E\left(|X_n| \cdot 1_{|X_n| > \lambda}\right) \longrightarrow 0 \quad \text{as } \lambda \longrightarrow \infty$$

$E\left(|X_n|\cdot \mathbb{1}_{|X_n|>\lambda}\right) \longrightarrow 0$ as $\lambda \longrightarrow \infty$

UNIFORMLY in n.

(ie, for a given ε, one single λ works $\forall n$) $\to \sup_n$.

Equivalently,

$$\forall \varepsilon>0, \exists \lambda>0 \text{ s.t } \sup_n E\left(|X_n|\cdot \mathbb{1}_{|X_n|>\lambda}\right)<\varepsilon$$

Simple observations:

① if X_i - "integrable" $\forall i=1,\dots,k$

finite collection of r.vs

then $\{X_i\}$ is uniformly integrable.

[for X_i, given ε, get $\lambda_i>0$ then choose $\lambda = \max_i \{\lambda_i\}$]

② If $\{X_n\}$ is u.i,

then $\{X_n\}$ is bounded in L_1.

ie, $\exists c>0$ s.t $E(|X_n|)<c \ \forall n \in \mathbb{N}$.

Exercise: show that, the converse is NOT true !!!

(try to get a counter example.

Hint: a simple r.v. might as well work !!!)

③ If $\exists Y$ with $E(|Y|)<\infty$,

s.t. $|X_n| \leq |Y| \ \forall n$.

then $\{X_i\}$ is uniformly integrable.

Converse is again NOT true !!!

ie, for $\{X_i\}$ u.i, there might not exist such a Y.

④ If $\sup_n E|X_n|^p<\infty$ for any $p>1$.

then $\{X_n\}$ is uniformly integrable.

AGAIN !! Converse is NOT true.

proved last sem

Note: $E|X|<\infty \iff E\left(|X|\cdot 1_{|X|>\lambda}\right) \longrightarrow 0$ as $\lambda \longrightarrow \infty$.

$\Updownarrow$

$E\left(|X|\cdot 1_A\right) \longrightarrow 0$ as $P(A) \longrightarrow 0$

(Proof needed)

$\hookrightarrow$ $\forall\, \varepsilon>0, \exists\, \delta>0$ s.t. $\forall A$,
$P(A)<\delta \Rightarrow E\left(|X|\cdot 1_A\right)<\varepsilon$.

⑤ **Theorem:**

$\{X_n\}$ - uniformly integrable.

$\iff \sup_n E\left(|X_n|\cdot 1_A\right) \longrightarrow 0$ as $P(A) \longrightarrow 0$.

ie, given $\varepsilon>0$, $\exists\, \delta>0$ s.t.
$E\left(|X_n|\cdot 1_A\right)<\varepsilon \quad \forall n$ whenever $P(A)<\delta$.

Recall: Uniform Integrability.

Definition:

A seq. $\{X_n\}$ of r.v.s is called **Uniformly Integrable** if $E\left(|X_n| \cdot 1_{|X_n|>\lambda}\right) \longrightarrow 0$ as $\lambda \to \infty$ (uniformly integrable)

$\Leftrightarrow \sup_n \left(E\, |X_n| \cdot 1_{|X_n|>\lambda}\right) \longrightarrow 0$ as $\lambda \longrightarrow \infty$

$\Leftrightarrow \forall \varepsilon>0, \exists \lambda = \lambda_\varepsilon$ s.t. $E\left(|X_n| \cdot 1_{|X_n|>\lambda}\right) < \varepsilon$ $\underline{\underline{\forall n}}$ → uniformity.

Recall: X is integrable $\Longleftrightarrow$ $E\left(|X| \cdot 1_{|X|>\lambda}\right) \to 0$ as $\lambda \to \infty$ (ie, tail expectation ↓ 0)

"X is integrate" $\equiv E|X| < \infty$

Observations:

① $\{X_n\}$ - uniformly integrable $\Rightarrow \{X_n\}$ is L-1 bounded. (ie, $\sup_n \|X_n\| < \infty$).

for proof: take $\varepsilon = 1$.

Converse not true !!

② If $\exists$ integrable r.v. Y such that, $|X_n| \le Y \ \forall n$, then $\{X_n\}$ is uniformly integrable.

Converse not true !! ie, DCT can be made stronger !!

③ $\{X_n\}$ is L_p bounded for some $p>1$, (say, even $p = 1+\delta$, $\delta > 0$.)
$\Rightarrow \{X_n\}$ is uniformly integrable

Converse not true !!

Ques: What is it then, that's equivalent ($\Leftrightarrow$) to uniform integrability?

Suppose, X - real-valued r.v. (*)

Proposition: X is "integrable" ~~iff~~. $\lim_{P(A) \to 0} E(X \cdot 1_A) = 0$

ie, $\forall \varepsilon > 0, \exists \delta > 0$ st, $\forall A \in \mathcal{A}$, $P(A) < \delta$ $\Rightarrow E(|X| \cdot 1_A) < \varepsilon$.

Proof: Exercise. (All necessary tools taught in Sem-2.)

(*) Why "real-valued" is needed? Counter eg: $\Omega = \mathbb{N}$. $\mathcal{A} = \mathbb{P}(\mathbb{N})$ (Power set of naturals)

$P(\{n\}) = \frac{1}{2^n}$.

$X(n) = \begin{cases} +\infty & \text{, if } n = 1 \\ 0 & \text{, otherwise.} \end{cases}$

take $\delta = \frac{1}{2}$. $\rightarrow$ this works $\forall \varepsilon$.

BUT, this X is certainly <u>NOT</u> integrable.

Result 1:

A sequence X_n is uniformly integrable

$\Leftrightarrow$ ($\{X_n\}$ is L_1- bounded and

$\lim_{P(A) \to 0} \sup_n E(|X_n| \cdot 1_A) = 0$.

ie, $\forall \varepsilon > 0$, $\exists \delta = \delta_\varepsilon > 0$ s.t. $\left.\begin{matrix} A \in \mathcal{A}, \\ P(A) < \delta \end{matrix}\right] \Rightarrow E(|X_n| \cdot 1_A) < \varepsilon$. $\forall n$)

this is referred to as "uniform absolute continuity" absolute continuity" $\quad P(A) < \delta \Big] \Rightarrow E(|X_n| \cdot 1_A) < \varepsilon. \quad \forall n.\Big)$

Proof: ($\Rightarrow$) Assume, $\{X_n\}$ - uniformly integrable.

to prove, X_n is L_1 bounded $\longrightarrow$ exercise !!

So, let $\varepsilon > 0$ be given.

$$E(|X_n| \cdot 1_A) = E(|X_n| \cdot 1_{|X_n| \le \lambda} \cdot 1_A) + E(|X_n| \cdot 1_{|X_n| > \lambda} \cdot 1_A)$$

$$\le \lambda \cdot P(A) + E(|X_n| \cdot 1_{|X_n| > \lambda} \cdot 1_A)$$

Now, choose $\lambda_0 = \lambda_0(\varepsilon)$ s.t.

(by uniform integrability) $\quad E(|X_n| \cdot 1_{|X_n| > \lambda}) < \varepsilon/2$

Now, choose $\delta > 0$ s.t.

$$\lambda_0 \delta < \varepsilon/2 \qquad \left(\delta < \frac{\varepsilon}{2\lambda_0}\right)$$

$P(A) \le \delta \quad \Rightarrow \therefore \lambda_0 P(A) \le \lambda_0 \delta < \lambda_0 \cdot \frac{\varepsilon}{2\lambda_0} < \varepsilon/2$

($\Leftarrow$) Let $\varepsilon > 0$ be given.

To show: We can λ_0 st $E(|X_n| \cdot 1_{|X_n| > \lambda_0}) < \varepsilon. \ \forall n$

ie, all that we have to do, is to find λ_0 s.t $P(|X_n| > \lambda_0) < \delta$.

By Chebyshev's Inequality:

$$P(|X_n| > \lambda_0) \le \frac{E|X_n|}{\lambda_0} = \frac{\|X_n\|_1}{\lambda_0}$$

$$\le \sup_n \frac{\|X_n\|_1}{}$$

this "removes"

this "removes" the dependence on n. $\leq \sup_n \frac{\|X_n\|_1}{\lambda_0}$

So now, we can choose λ_0 such that

$$\sup_n \frac{\|X_n\|_1}{\lambda_0} < \delta.$$

(from given condition)

Result 2:

$X_n, n \geq 1$ and X - real r.vs on a Probability Space.

(a) $X_n \xrightarrow{P} X$ and $\{X_n\}$ uniformly integrable

$\Rightarrow X \in L_1$, and $X_n \xrightarrow{L_1} X$

kind of an additional criteria over p-conv to imply L_1 conv.

(b) $X_n \xrightarrow{L_1} X \Rightarrow X_n \xrightarrow{P} X$ and $\{X_n\}$ is uniformly integrable.

Recall: DCT.

$X_n \xrightarrow{a.s} X$ & $|X_n| \leq Y$ for some Y-integrable is $E|Y| < \infty$

$\Rightarrow E(X_n) \longrightarrow E(X)$ $E|X_n| < \infty$.

$X_n \xrightarrow{P} X$ ie, $X_n \xrightarrow{L_1} X$.

So, note that, Result-2, part-a is a huge improvement over DCT. ie, we've weakened the conditions of DCT, yet getting the same result.

Proof: (a) $X_n \xrightarrow{P} X$

"same" result

Proof: (a) $X_n \xrightarrow{P} X$

$\Rightarrow X_{n_k} \xrightarrow{a.s.} X$ for a subsequence $\{n_k\}$.

Fatou's lemma $\Rightarrow E|X| \le \underbrace{\liminf E|X_{n_k}|}_{\liminf \|X_{n_k}\|_1} \le \sup_n \|X_n\|_1 < \infty$ (from u.i)

$E|X| = E|\liminf X_{n_k}|$ as $X_{n_k} \xrightarrow{a.s} X$.

(Eg. real seq. $\{x_n\}$, $|x_n| \le 10$. $x_n \to x$. then, $|x| \le 10$)

$\Rightarrow X \in L_1$.

now, to show, $X_n \xrightarrow{L_1} X$

$$E|X_n - X| = E\left(|X_n - X| \cdot 1_{|X_n - X| \le \varepsilon/3}\right) + E\left(|X_n - X| \cdot 1_{|X_n - X| > \varepsilon/3}\right)$$

$$\le (\varepsilon/3) + \left(E\left(|X_n| \cdot 1_{|X_n - X| > \varepsilon/3}\right) + E\left(|X| \cdot 1_{|X_n - X| > \varepsilon/3}\right)\right)$$

Δ - inequality.

Now, $E|X| < \infty \Rightarrow$

$\exists \delta_1 > 0$ st, $A \in \mathcal{a}$, $P(A) < \delta_1$ $\Rightarrow E(|X| \cdot 1_A) < \varepsilon/3$

$\delta_1 = \delta_{1, \varepsilon/3}$

Also $\{X_n\}$ - u.i,

$\Rightarrow \exists \delta_2 > 0$, s.t $A \in \mathcal{a}$, $P(A) < \delta_2$ $\Rightarrow E(|X_n| \cdot 1_A) < \varepsilon/3$ $\forall n$

(using result-1).

$X_n \xrightarrow{P} X$

$\exists n_0$ s.t., $\forall n \ge n_0$, $P(|X_n - X| > \varepsilon/3)$

$X_n \xrightarrow{\;\cdot\;} X$

$\exists\, n_0$ s.t., $\forall\, n \geq n_0$, $P(|X_n - X| > \varepsilon/3)$

$< \min\{\delta_1, \delta_2\} \quad \forall\, n \geq n_0.$

$\therefore E(|X| \cdot 1_{|X_n - X| > \varepsilon/3}) < \varepsilon/3 \quad \forall\, n \geq n_0,$

as, $\forall\, n \geq n_0$,

$P(|X_n - X| > \varepsilon/3) < \delta_2.$

&, $E(|X_n| \cdot 1_{|X_n - X| > \varepsilon/3}) < \varepsilon/3 \quad \forall\, n \geq n_0,$

as, $\forall\, n \geq n_0$,

$P(|X_n - X| > \varepsilon/3) < \delta_1$

$\therefore E|X_n - X| \leq \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.$

$X_n \xrightarrow{L_1} X.$

(b) To prove: $X_n \xrightarrow{L_1} X \Rightarrow X_n \xrightarrow{P} X$ & $\{X_n\}$ is u.i

we know, $\forall\, p \geq 1$. $X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{P} X$.

So, this part is trivial.

we're just left to show, $\{X_n\}$ is u.i (uniformly integrable)

Let $\lambda > 1$ [$\because$ we're supposed to choose "large" λ.]

$$E\left(|X_n| \cdot 1_{|X_n| > \lambda}\right) \leq E|X_n - X| \cdot 1_{|X_n| > \lambda} + E|X| \cdot 1_{|X_n| > \lambda}$$

(Δ- inequality)

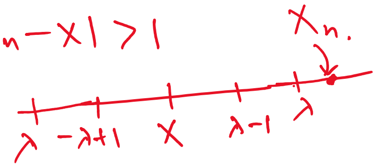
$$\leq \|X_n - X\|_1 + E\left(|X| \cdot 1_{|X_n| > \lambda,\ |X| \leq \lambda - 1}\right) + E\left(|X| \cdot 1_{|X_n| > \lambda,\ |X| > \lambda - 1}\right)$$

$$\leq \|X_n - X\|_1 + E \cdot \left(|X| \cdot 1_{|X_n - X| > 1,\ |X| \leq \lambda - 1}\right)$$

this just stays

(*) $\rightarrow \underbrace{|X_n| > \lambda}_{A}, \underbrace{|X| \leq \lambda - 1}_{B}.$

(C) $\Rightarrow |X_n - X| > 1$



$A \cap B \subseteq C$

$\therefore 1_{A \cap B} \leq 1_C$

i.e, 1.

i.e, $\mathbb{1}_{(X_n > \lambda, |X| \leq \lambda - 1)} \leq \mathbb{1}_{(|X_n - X| > 1)}$

$$\leq \|X_n - X\|_1 + \underset{(*)}{E}\left(|X| \cdot \mathbb{1}_{|X_n - X| > 1, |X| \leq \lambda - 1}\right)$$

$$\underset{(*)}{E}\left(|X| \mathbb{1}_{|X| > \lambda - 1}\right)$$

$$\leq \underbrace{\|X_n - X\|_1 + (\lambda - 1) \cdot P(|X_n - X| > 1)} + E\left(|X| \cdot \mathbb{1}_{|X| > \lambda - 1}\right)$$

$\|X_n - X\|_1 + (\lambda - 1) \cdot \|X_n - X\|_1 = \lambda \cdot \|X_n - X\|_1$.

$$\leq (\lambda - 1) \cdot \frac{E|X_n - X|}{1}$$

(Chebyshev)

$$= (\lambda - 1) \cdot \|X_n - X\|_1$$

$\therefore$

$$\leq \lambda \cdot \|X_n - X\|_1 + E\left(|X| \cdot \mathbb{1}_{|X| > \lambda - 1}\right).$$

Now, $X \in L_1$

tail expectation

$\therefore$ Choose $\lambda_0 > 1$ st

$$E\left(|X| \cdot \mathbb{1}_{|X| > \lambda_0 - 1}\right) < \varepsilon/2$$

$\therefore$ Choose n_0 s.t $\forall n \geq n_0$

$$\lambda_1 \|X_n - X\|_1 < \varepsilon/2$$

$\left[\because X_n \xrightarrow{L_1} X, \therefore \|X_n - X\| \text{ can be made arbitrarily small.}\right]$

&, choose $\lambda_1, \ldots, \lambda_{n_0 - 1}$ s.t.

$$E\left(|X_n| \cdot \mathbb{1}_{|X_n| > \lambda_n}\right) < \varepsilon \quad \forall n = 1, \ldots, n_0 - 1.$$

$\|X_n - X\|_1 < \varepsilon$.

Now, choose $\lambda = \text{Max.}\{\lambda_0, \lambda_1, \ldots, \lambda_{n_0 - 1}\}$.

$\therefore$ For this λ, $\forall n$,

$$E\left(|X_n| \cdot \mathbb{1}_{|X_n| > \lambda}\right) < \varepsilon.$$

— Midsem syllabus till here —

Probability-3 Lecture-10

03 September 2024 14:18

Set-2,
Q-9. $\{X_n\}$ $\{Y_n\}$ $\Rightarrow \{X_n \cdot Y_n\}$ - u.i

This is false!!!

$\{X_n\} \to$ u.i, $\{Y_n\} \to$ tight.

i.e, $\forall \varepsilon > 0, \exists K > 0$
s.t. $P(|Y_n| > K) < \varepsilon \ \forall n$

take $Y_n \sim U(0,1)$ with prob. $= \frac{n-1}{n}$

& , $Y_n = e^n$ with prob $= 1/n$.

this is NOT u.i.

$\because \|Y_n\|_{L_1} \geq \frac{E(e^n)}{n} \to \infty$

Correct question:
$\{X_n\}, \{Y_n\}. \Rightarrow \{X_n \cdot Y_n\}$ u.i
$\{X_n\} \to$ u.i, $\{Y_n\} \to$ bounded
(prove)

$\exists$ $\{X_n\}$ - u.i, but X_n is not bounded in any L_p for $p > 1$

I construct a r.v $X \geq 1$.
s.t. $E(X \log X) < \infty$.
But, $E(X^p) = \infty \ \forall p > 1$.

Now, seq. of r.vs,
$$X_n = X \wedge n$$
$\hookrightarrow$ Min.$\{X, n\}$

<u>claim</u>: $\{X_n\}$ is u.i,
but $\sup_n E(X_n^p) = \infty$.

clearly, X_n's are increasing $\uparrow$

$\therefore \sup_n E(X_n^p) = \lim_{n \to \infty} E(X_n^p)$
$= E(X^p) = \infty$.

$$E(X_n \cdot 1_{X_n > \lambda}) = E\left(\frac{X_n}{X_n \log X_n}(X_n \log X_n) 1_{X_n > \lambda_0}\right)$$

$\frac{x}{x \log x} \searrow 0$ as $x \nearrow \infty$

$\therefore$ Given $\varepsilon > 0$,
<u>get λ_0</u> s.t. $x > \lambda_0 \Rightarrow \frac{x}{x \log x} < \varepsilon / E(X \log X)$.

$$< \frac{\varepsilon}{E(X\log X)} \cdot E\left(X_n \log X_n \cdot 1_{X_n > \lambda_0}\right).$$

$$< \frac{\varepsilon}{E(X\log X)} \cdot E\left(X \log X\right).$$

$$< \varepsilon.$$

* Z has density.

$$F(z) = e^{1-z}, \quad z > 1.$$

$$X = \frac{e^Z}{Z^{2+\varepsilon}} \leftarrow \text{anything larger than 2 would do.}$$

Borel-Cantelli Lemma.

If $\{A_n\}_{n\geq 1}$ is a sequence of events s.t.

$$\sum_n P(A_n) < \infty \quad \left(\Rightarrow \sum_{k\geq n} P(A_k) \longrightarrow 0 \text{ as } n \to \infty \quad (\text{tail sum})\right)$$

then, $P(\limsup_n A_n) = P\left(\bigcap_{n=1} \bigcup_{k\geq n} A_k\right) = 0.$

Proof: Clearly, $\bigcup_{k\geq n} A_k$ - decreasing unions.

$$\therefore P\left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right) \underset{\substack{\downarrow \\ \text{Continuity of probability.}}}{=} \lim_{n\to\infty} P\left(\bigcup_{k\geq n} A_k\right).$$

$$\leq \lim_{n\to\infty} \sum_{k\geq n} P(A_k) = 0$$

* What are the ω's that belong to $\underbrace{\limsup A_n}_{\text{set}}$?

$\omega \in \limsup_n A_n \iff$ for every n,
$\omega \in \bigcup A_k$.

$\omega \in \limsup_n A_n \iff$ for every n,
$\omega \in \bigcup_{k \geq n} A_k$.

$\iff$ for every n, there exists $k \geq n$
s.t $\omega \in A_k$.

ie, ω should belong to infinitely many of these A_k's.

Borel-Cantelli Second Lemma

If $\{A_n\}$ - sequence of independent events s.t,
$\sum_n P_n = \infty$ (ie, sequence of partial sum diverges)
then $P(\limsup_n A_n) = 1$

(kind of, partial converse of Borel-Cantelli lemma)

* Kolmogorov's 0-1 law
(ie, if A_n - seq. of ind. events then $P(\limsup_n A_n) = 0$ or 1
ie, no intermediate value

Proof: $P((\limsup A_n)^c) = P\left(\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right)^c\right)$

$$= P\left(\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k^c\right)$$

countable union.

So, enough to show that

$P\left(\bigcap_{k \geq n} A_k^c\right) = 0$ for every n.

then $P((\limsup A_n)^c) = 0$.

$k \geq n$ $\qquad$ $\cup$

then, $P\left(\left(\limsup A_n\right)^c\right) = 0$.

$\because$ $\bigcap_{k \geq n} A_k^c$ is decreasing limit of partial intersections.

$\therefore$ fix n.

$$\therefore P\left(\bigcap_{k \geq n} A_k^c\right) = \lim_{m \to \infty} P\left(\bigcap_{k=n}^{n+m} A_k^c\right)$$

$$= \lim_{m \to \infty} \prod_{k=n}^{n+m} P(A_k^c)$$

$$= \lim_{m \to \infty} \prod_{k=n}^{n+m} \left(1 - P(A_k)\right)$$

$$* \; 1 - x < e^{-x} \longrightarrow \quad = \lim_{m \to \infty} \prod_{k=n}^{n+m} e^{-P(A_k)}$$

$$= \lim_{m \to \infty} e^{-\sum_{k=n}^{n+m} P(A_k)} \quad \left(\sum_{k=n}^{n+m} P(A_k) \to \infty\right)$$

tail diverges

$$= 0. \; \blacksquare$$

Set-1, Q.12 :

$$d(X,Y) := \inf\{\varepsilon > 0 : P(|X-Y| > \varepsilon) \leq \varepsilon\}.$$

$(\Omega, \mathcal{A}, P)$

$L_0 \left(= L_0(\Omega, \mathcal{A}, P)\right)$

$= \{$all real r.vs on $(\Omega, \mathcal{A}, P)\}$.

Another metric, $\rho(X,Y) := E\left(\frac{|X-Y|}{1+|X-Y|}\right)$

So, we have 2 diff. metrics which are complete in L_0.

Recall:

$L_p(\Omega, \mathcal{a}, P) = \{X - \text{real r.v. s.t. } E(|X|^p) < \infty\}$

For $X, Y \in L_p$. $d_p(X,Y) = E(|X-Y|^p)^{1/p} =: \|X-Y\|_p$

(we showed that this is a valid metric)

for L_p, $p \geq 1$,

to show: L_p is complete.

Let $\{X_n\}$ - Cauchy in metric d_p

ie, $d_p(X_m, X_n) = \|X_m - X_n\|_p \longrightarrow 0$ as $m,n \to \infty$

x_n - Cauchy $\Rightarrow$ x_n bounded

Proof:

$\forall \varepsilon > 0, \exists N$ st $\forall m,n \geq N$ $|x_n - x_m| < \varepsilon$.

$||x_n| - |x_N|| < \varepsilon$

$\Rightarrow |x_n| < \varepsilon + |x_N|$

$\forall$ $n > N$.

Take max

$x_0 = \max\{|x_1|, |x_2|, \dots, |x_N|\}$

$|x_n| \leq \varepsilon + x_0 < \infty$

Claim,

$\{X_n\}$ is bounded in L_p.

$\forall \varepsilon > 0, \exists N$ st $\forall m,n > N$.

$\|X_m - X_n\|_p < \varepsilon$.

$\therefore \|X_n - X_N\|_p < \varepsilon$.

$\Rightarrow \|X_n\|_p \leq \varepsilon + \|X_N\|_p$. $\forall n > N$.

(Minkowski's ineq.)

Let $X_0 = \max\{\|X_1\|_p, \dots, \|X_N\|_p\}$.

$\therefore \|X_n\|_p \leq \varepsilon + X_0$. $\boxed{\therefore X_n - \text{bounded in } L_p.}$

Using Cauchy property, we get a subsequence

Using Cauchy property, we get a subsequence

$1 \le n_1 < n_2 < \dots < n_k < \dots$

s.t. $\|X_m - X_n\|_p < 3^{-k}$ $\forall$ $m, n \geq n_k$.

ie, put $k=1$. get $\underline{n_1}$

put $k=2$, get n_2.

if $n_2 > n_1$ ✓

if $n_2 \le n_1$

take $n_2 = n_1 + n_2$

& so on.

$\therefore$ get these indices in this manner.

$\therefore$ By Chebyshev's inequality.

$$P\left(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\right) \le \frac{E\,|X_{n_{k+1}} - X_{n_k}|^p}{2^{-kp}}$$

$$< \left(\left(\frac{2}{3}\right)^p\right)^k \longrightarrow 0$$

infinite g.p, series - finite.

$$\therefore \sum_k P\left(|X_{n_{k+1}} - X_{n_k}| > 2^{-k}\right) < \infty$$

$\therefore$ By Borel-Cantelli Lemma,

$$P\left(|X_{n_{k+1}} - X_{n_k}| > 2^{-k} \text{ for infinitely many } k\right) = 0.$$

$$\therefore P\left(\left(|X_{n_{k+1}} - X_{n_k}| > 2^{-k} \text{ for infinitely many } k\right)^c\right) = 1$$

$$\Rightarrow P\left(|X_{n_{k+1}} - X_{n_k}| \le 2^{-k} \text{ for all } k \text{ after } \underbrace{\text{some stage}}_{\text{depending on } \omega}\right) = 1$$

$\lceil$ $\{a_n\}$ - real seq.

on ω.)

[$\{a_n\}$ - real seq.
$|a_{k+1} - a_k| < 2^{-k} \quad \forall k \geq k_0$
then, a_n must converge to some real limit: $a_n \to a$.]

$$\Rightarrow P\left(X_{n_k}(\omega) \longrightarrow X(\omega) \text{ real}\right) = 1$$

ie, $X_{n_k} \xrightarrow{a.s.} X$ define $X(\omega) = \lim_{n\to\infty} X_n(\omega)$, where lim exists

(verify that this works) $= 0$, otherwise.

We have shown,

if $\{X_n\}$ is Cauchy in L_p, then $\exists$ a subsequence

Step 1: $\{X_{n_k}\}$ s.t. $X_{n_k} \xrightarrow{a.s.} X$ - real r.v.

Step 2: next, to show: $X_n \xrightarrow{a.s.} X$

Step 3: then, to show: $X \in L_p$.

$$\therefore E(|X|^p) = E\left(\liminf_k |X_k|\right) \leq \liminf_k E(|X_k|)$$

$$\leq \sup_n \|X_n\|_p^p < \infty.$$

($\because X_n$ is Cauchy)

Fix k.

$$E\left(|X_m - X_{n_k}|^p\right) \leq 3^{-kp} \quad \forall m \geq n_k.$$

$$E\left(|X - X_{n_k}|^p\right) \leq \lim_j E\left(|X_{n_j} - X_{n_k}|^p\right) \leq 3^{-kp} \quad \forall j > k.$$

$\longrightarrow 0$ as $k \to \infty$.

$$\therefore X_{n_k} \xrightarrow{L_p} X, \ \& \ \{X_n\}\text{- Cauchy}$$

$$\Downarrow$$

$$X_n \xrightarrow{L_p} X.$$

Set - 0: X, Y - independent.

Y - cont $\longrightarrow P(Y=y) = 0 \ \forall y \in \mathbb{R}$.

?

Y - cont $\longrightarrow P(Y=y)=0 \ \forall y \in \mathbb{R}$.

(a) $P(X=Y) \overset{?}{=} 0$

$= P((X,Y) \in D)$

$= E(Q(X,D))$,

where

$Q(x,D) = P(Y \in D_x)$

$= P(Y=x)$

$= 0$.

where,

$D = \{(x,y) : x=y\}$.

$\therefore D_x = \{y : (x,y) \in D\}$

"x-section".

(b) $h : \mathbb{R}^2 \longrightarrow \mathbb{R}$

$\forall a, x \in \mathbb{R}$.

$h(x,y)=a$ has countably many sol^n.

$A_a = \{(x,y) : h(x,y)=a\}$.

$P(h(X,Y)=a) = P$

⋮

$\sup_{k \geq n} |X_k - X_n|$

$\leq \sup_{k \geq n} |X_k - X| + |X_n - X|$

$\sup_{k \geq n} |X_k - X| \to 0$, $|X_n - X| \to 0$

$X_n \xrightarrow{a.s.} X$

$X_n \xrightarrow{P} X$

$$P\left(\bigcup_{(a,b) \in \mathbb{Q} \times \mathbb{Q}} N_{(a,b)}\right) = 0.$$

$\downarrow$

- X_n does $) = 0$.

$P \quad X$.

Probability-3 Lecture-11

06 September 2024 11:25

- Respect to A·N·Kolmogorov !!!

(classical) Laws of Large Numbers

$(\Omega, \mathcal{A}, P)$ - probability

$X_1, \ldots, X_n$ iid r.v.s with common finite mean μ.

$\therefore$ Law of Large Numbers (LLN):

$$\frac{X_1 + \cdots + X_n}{n} \longrightarrow \mu$$

Weak Law of Large Numbers (WLLN):

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu$$

Strong Law of Large Numbers (SLLN):

[easier proof: done by Etemadi]

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s} \mu$$

"Real-life problem is just an abstract concept."
— Prof. AG, 6^{th} Sept. '24.

Weak Law of Numbers (WLLN):

$X_1, X_2, \ldots, X_n$ — sequence of i.i.d. r.vs with finite common mean

$(\Leftrightarrow E|X_i| < \infty \quad \forall i)$

n^{th} partial sum, $S_n := X_1 + X_2 + \cdots + X_n$

Aim: To show: $\underline{S_n} \xrightarrow{P} \mu$

Aim: To show: $\frac{S_n}{n} \xrightarrow{P} \mu$

ie, $\forall \varepsilon > 0$, $P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \longrightarrow 0$ as $n \longrightarrow \infty$

we don't have a distribution given. So, computing this isn't possible.

So, idea: to find an upper bound, & show that, that upper bound $\longrightarrow 0$.

Using Chebyshev's inequality,

$$\therefore P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{E\,|S_n - n\mu|}{n\varepsilon}$$

(Δ-inequality) $$\leq \frac{E\,|X_1-\mu| + E\,|X_2-\mu| + \cdots + E\,|X_n-\mu|}{n\varepsilon}$$

$$= \frac{n \cdot E\left(|X_1-\mu|\right)}{n\varepsilon} \quad \left[\because X_i\text{'s are i.i.d.}\right]$$

$$= \frac{E\left(|X_1-\mu|\right)}{\varepsilon}$$

?? We are stuck !!

this approach leads us nowhere.

Just to get a "feeling" of happiness, we'll assume a stronger assumption, ie, X_i's have finite 2^{nd} moment.

then, applying Chebyshev's inequality using 2^{nd} moments,

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{E\left(|S_n - n\mu|^2\right)}{n^2\varepsilon^2}$$

$$= \frac{V(S_n)}{n^2\varepsilon^2}$$

$\mu = E(X_i)$

$\therefore n\mu = E(S_n)$

ie, this is variance of S_n.

$$= \frac{V(S_n)}{n^2 \varepsilon^2}$$ variance of S_n.

$$= \frac{n \cdot V(X_1)}{n^2 \varepsilon^2}$$ [$\because V(S_n) = n \cdot V(X_i)$

for this to be true, only the covariances $= 0$

$$= \frac{V(X_1)}{n \varepsilon^2} \to 0 \text{ as } n \to \infty$$

ie, Remark:

At this stage, only pairwise independence of the r.v.s is needed. Total independence isn't needed.]

So, we proved a slightly weaker condition than WLLN.

... back to the hypothesis:

$X_1, \ldots, X_n$ iid with finite common mean

technique: replace original seq. $X_1, \ldots, X_n$ by a new seq: $Y_1, \ldots, Y_n$; X_n's truncated appropriate.

Truncation technique:

For each $n \geqslant 1$, $Y_n := \begin{cases} |X_n| , & |X_n| \leqslant n \\ 0 , & \text{else.} \end{cases}$

Y_n - function of X_n.

So, if X_n's are inpendent, Y_n's are independent too!!

What did we lose? identicality of the distributions.

[$\because$ truncation levels are different]

ie, Y_n's are not identically distributed.

$$Y_n \overset{d \leftarrow \text{in distributed}}{=\!=} |X_1| \cdot \mathbb{1}_{|X_1| \leqslant n}$$

Recall our aim: $\frac{S_n}{n} \xrightarrow{P} \mu$

$$\Leftrightarrow \frac{S_n}{n} - \mu \xrightarrow{P} 0$$

$$\Leftrightarrow \frac{S_n - n\mu}{n} \xrightarrow{P} 0$$

$$\Leftrightarrow \frac{S_n - E(S_n)}{n} \xrightarrow{P} 0. \quad \text{——— } (1)$$

Here, let $T_n := Y_1 + Y_2 + \cdots + Y_n$

<u>we will prove:</u>

$$\frac{T_n - E(T_n)}{n} \xrightarrow{P} 0 \quad \text{——— } (2)$$

firstly, why doing this suffices? ie, to show: $(2) \Rightarrow (1)$

<u>Observation - 1:</u>

$$P(Y_n \neq X_n) = P(|X_n| > n)$$
$$= P(|X_1| > n) \quad \left[\because X_1, \ldots, X_n \text{ are iids.}\right]$$

$\because E|X_n| < \infty \quad \forall n,$

$$\therefore \sum_n P(Y_n \neq X_n) = \sum_n P(|X_n| > n)$$
$$= \sum_n P(|X_1| > n) < \infty$$

$\Rightarrow$ By Borel-Cantelli lemma,

$P(Y_n \neq X_n \text{ for infinitely many } n) = 0$

$\Leftrightarrow P(Y_n = X_n \text{ for all but finitely many } n) = 1$

ie, $\exists\, n_0$ sufficiently large, s.t $\forall\, n > n_0$, this holds.

$$\Rightarrow P\left(\frac{T_n}{n} - \frac{S_n}{n} \to 0\right) = 1$$

$$\Rightarrow \frac{T_n}{n} - \frac{S_n}{n} \xrightarrow{a.s} 0$$

[Consider 2 real seq^n: a_n & b_n s.t, $\exists\, n_0$ large s.t $\forall\, n > n_0$, $a_n = b_n$. Then $\frac{a_n}{n}$ … $\frac{b_n}{n}$.

$$\Rightarrow \frac{T_n}{n} - \frac{S_n}{n} \xrightarrow{a.s} 0$$

$n > n_0$, $a_n = b_n$. Then $\frac{\sum_1^n a_k}{n} = \frac{\sum_1^n b_k}{n}$

$$\Rightarrow \frac{T_n}{n} - \frac{S_n}{n} \xrightarrow{p} 0$$

this + ② $\Rightarrow \frac{S_n}{n} - \frac{E(T_n)}{n} \xrightarrow{p} 0$

Observation 2 :

$$\frac{E(T_n)}{n} = \frac{1}{n} \cdot \sum_{k=1}^{n} E(Y_k)$$

$$= \frac{1}{n} \cdot \sum_{k=1}^{n} E\left(X_k \cdot 1_{|X_k| \le k}\right)$$

$$= \frac{1}{n} \cdot \sum_{k=1}^{n} E\left(X_1 \cdot 1_{|X_1| \le k}\right)$$

Now, $E\left(X_1 \cdot 1_{|X_1| \le n}\right) \longrightarrow \mu = \frac{E(S_n)}{n}$

$\because X_1 \cdot 1_{|X_1| \le n} \le X_1$, & $E|X_1| = \mu < \infty$

$\therefore$ By DCT, this follows.

Cesaro mean of that

$$\therefore \frac{1}{n} \sum_{k=1}^{n} E\left(X_1 \cdot 1_{|X_1| \le k}\right) \longrightarrow \frac{E(S_n)}{n}$$

$$\therefore \frac{E(T_n)}{n} \longrightarrow \frac{E(S_n)}{n}.$$

$$\Leftrightarrow \frac{E(T_n)}{n} - \frac{E(S_n)}{n} \longrightarrow 0.$$

Note that, this has all moments finite!!

Now, finally,

$$P\left(\left|\frac{T_n - E(T_n)}{n}\right| > \varepsilon\right) = P\left(|T_n - E(T_n)| > n\varepsilon\right)$$

$$\le \frac{E|T_n - E(T_n)|^2}{n^2 \varepsilon^2}$$

(Chebyshev's inequality, and

(Chebyshev's inequality, wrt 2^{nd} moment) $\leq \dfrac{E|T_n - E(T_n)|}{n^2\varepsilon^2}$

$$= \frac{V(T_n)}{n^2\varepsilon^2}$$

$$= \frac{\sum_{k=1}^{n} V(Y_k)}{n^2\varepsilon^2}$$

$$\leq \frac{1}{n^2\varepsilon^2} \cdot \sum_{k=1}^{n} E(Y_k^2) \quad \text{--- } (*)$$

$$\leq \frac{1}{n^2\varepsilon^2} \cdot \sum_{k=1}^{n} k \cdot E|Y_k|$$

$$\leq \frac{1}{n^2\varepsilon^2} \cdot \sum_{k=1}^{n} k \cdot \underbrace{E|X_1|}_{\text{common bound for all } Y_k}$$

Again, we're stuck!! both numerator & denominator have orders of n^2.

$\therefore$ ② $\Rightarrow$ ①

Choose a sequence $\{a_n\}$ of +ve real nos, s.t $a_n \nearrow \infty$, but $\frac{a_n}{n} \longrightarrow 0$

(ie, $a_n \nearrow \infty$ "slower" than $n \to \infty$. Eg, $a_n = \log(n)$, $a_n = \left(\frac{2}{3}\right)^n$, etc.)

back to $(*)$:

$$P\left(\left|\frac{T_n - ET_n}{n}\right| > \varepsilon\right) \leq \frac{1}{n^2\varepsilon^2} \sum_{k=1}^{n} E(Y_k^2)$$

$$= \frac{1}{n^2\varepsilon^2} \cdot \sum_{k=1}^{n} \left(E X_1^2 \cdot 1_{|X_1| \leq k}\right)$$

$$= \frac{1}{n^2\varepsilon^2} \cdot \sum^{n} E\left(X_1^2 \cdot 1_{|X_1| < a}\right) \quad +$$

$$= \frac{1}{n^2\varepsilon^2}\cdot\sum_{k=1}^{n} E\left(X_1^2\cdot 1_{|X_1|\le a_k}\right) \quad +$$

1st term

$$\frac{1}{n^2\cdot\varepsilon^2}\cdot\sum_{k=1}^{n} E\left(X_1^2\cdot 1_{a_k<|X_1|\le k}\right)$$

2nd term.

1st term:

$$\frac{1}{n^2\varepsilon^2}\cdot\sum_{k=1}^{n} E\left(X_1^2\cdot 1_{|X_1|\le a_k}\right) = \frac{1}{n^2\varepsilon^2}\sum_{k=1}^{n} E\Big(X_1\cdot \underbrace{X_1\cdot 1_{|X_1|\le a_k}}_{\le a_k}\Big)$$

$$\le \frac{1}{n^2\varepsilon^2}\cdot\sum_{k=1}^{n} a_k\cdot E(|X_1|)$$

$$\longrightarrow 0. \checkmark$$

2nd term:

$$\frac{1}{n^2\varepsilon^2}\cdot\sum_{k=1}^{n} E\left(|X_1|^2\cdot 1_{a_k<|X_1|\le k}\right)$$

$$\to |X_1|\cdot\underbrace{|X_1|\cdot 1_{a_k<|X_1|\le k}}_{\le k\cdot 1_{a_k<|X_1|}\le n\cdot 1_{a_k<|X_1|}}$$

$$\le \frac{1}{n^{\cancel{2}}\varepsilon^2}\cdot\sum_{k=1}^{n}\cancel{n}\cdot E\left(|X_1|\cdot 1_{a_k<|X_1|}\right)$$

$$= \frac{1}{n\varepsilon^2}\cdot\sum_{k=1}^{n} E\left(|X_1|\cdot 1_{a_k<|X_1|}\right)$$

$$= \frac{1}{\varepsilon^2}\cdot\frac{\sum_{k=1}^{n} E\left(|X_1|\cdot 1_{|X_1|>a_k}\right)}{n} \longrightarrow 0.$$

Cesaro mean

$[\because a_k\to 0]$

(analysis-1 result. Suppose $x_n\to x$ $\Leftrightarrow$ $\frac{x_1+\cdots+x_n}{n}\to x$)

$\blacksquare$

Probability-3 Lecture-12

17 September 2024 14:25

$X_1, X_2, \ldots$ i.i.d. sequence of r.vs, all on same probability space $(\Omega, \mathcal{A}, P)$ with finite common mean μ.

$E|X_1| < \infty$.

$$S_n = \frac{X_1 + \cdots + X_n}{n}$$

Weak Law of Large numbers (WLLN):

$$\frac{S_n}{n} \xrightarrow{P} \mu$$

Strong law of large numbers (SLLN):

$$\frac{S_n}{n} \xrightarrow{a.s} \mu.$$

for WLLN,

we need to prove: $P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \longrightarrow 0 \quad \forall \varepsilon > 0.$

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = P\left(|S_n - n\mu| > n\varepsilon\right)$$

tail probability of the r.v: $(S_n - n\mu)$.

Assume: 2^{nd} moment finite.

ie, $E|X_1|^2 < \infty$

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = P\left(|S_n - n\mu| > n\varepsilon\right) \leq \frac{E|S_n - n\mu|^2}{n^2\varepsilon^2}$$

(Chebyshev)

$$= \frac{Var(S_n)}{n^2\varepsilon^2}$$

pairwise independence

$$= \frac{\cancel{n} \cdot Var(X_1)}{n^{\cancel{2}}\varepsilon^2}$$

$\longrightarrow 0$ as $n \to \infty$

Q: How to prove if only $E|X_1| < \infty$ is given?

i.e, only first moment finite is given.

New idea: "Truncation".

↳ Replace X_n's by Y_n's, where Y_n's are X_n's "truncated" at an "appropriate threshold".

Here, define $Y_n := X_n \cdot \mathbb{1}_{|X_n| \le n}$.

$Y_1, Y_2, \ldots\ldots$ while are still independent

they are no longer identically distributed, as each Y_i has different truncation levels.

Define $T_n := Y_1 + Y_2 + \cdots + Y_n$.

We will show: $\dfrac{T_n - E T_n}{n} \xrightarrow{P} 0$ ——— $(*)$

& this will then imply, $\dfrac{S_n - n\mu}{n} \xrightarrow{P} 0$? Why? *

* firstly, $\sum P(X_n \neq Y_n) = \sum P(|X_n| > n)$

$= \sum P(|X_1| > n)$ ($\because X_i$'s - iid)

$< \infty$.

because: $E|X_1| < \infty$

$\therefore$ By Borel-Cantelli lemma:

$P(X_n \neq Y_n$ for infinitely$) = 0$

$\therefore$ By Borel-Cantelli lemma:

$$P\left(X_n \neq Y_n \text{ for infinitely many } n\right) = 0$$

take complement

$$\Rightarrow P\left(\{\omega : X_n(\omega) = Y_n(\omega) \text{ for all } n \geq n_0(\omega)\}\right) = 1$$

$$\Rightarrow P\left(\left\{\omega : \left(\frac{1}{n} S_n - \frac{1}{n} T_n\right) \longrightarrow 0\right\}\right) = 1.$$

$$\therefore \frac{S_n}{n} - \frac{T_n}{n} \xrightarrow{a.s} 0 \quad \text{——— } (1)$$

$$\Downarrow$$

Now, $\underbrace{\frac{T_n - ET_n}{n} \xrightarrow{P} 0 \quad \& \quad \frac{S_n}{n} - \frac{T_n}{n} \xrightarrow{P} 0}_{\text{sum up these two}}$

$$\frac{T_n}{n} - \frac{ET_n}{n} + \frac{S_n}{n} - \frac{T_n}{n} \xrightarrow{P} 0 + 0$$

$$\Rightarrow \frac{S_n}{n} - \frac{ET_n}{n} \xrightarrow{P} 0 \quad \text{——— } (2)$$

$$\therefore E(Y_n) = E(X_1 \cdot 1_{|X_1| \leq n}) \xrightarrow[DCT]{} \mu$$

$$\frac{E(T_n)}{n} = \frac{E(Y_1) + \cdots + E(Y_n)}{n} \longrightarrow \mu.$$

(real seq. $\to \mu$ $\Rightarrow$ their cesaro mean $\to \mu$) ——— (3)

$*$ & (3) $\Rightarrow$

$$\frac{S_n}{n} - \mu \xrightarrow{P} 0$$

$\therefore$ Now, we have to show: $\dfrac{T_n - ET_n}{n} \xrightarrow{P} 0$.

Fix $\varepsilon > 0$.

$$P\left(\left|\frac{T_n - ET_n}{n}\right| > \varepsilon\right) = P\left(|T_n - ET_n| > n\varepsilon\right)$$

$$\leq \frac{Var(T_n)}{n^2\varepsilon^2}$$

pairwise independent

$$= \frac{1}{n^2\varepsilon^2} \cdot \sum_{k=1}^{n} V(Y_k)$$

$$\leq \frac{1}{n^2\varepsilon^2} \cdot \sum_{k=1}^{n} E|X_1|^2 \cdot \mathbb{1}_{|X_1| \leq k}$$

we're stuck —.

(New trick:
fix any non-ve seq. $\{a_n\} \nearrow \infty$, but $\frac{a_n}{n} \to 0$,
eg. $a_n = \sqrt{n}$.)

$$\leq \frac{1}{n^2\varepsilon^2} \cdot \left(\sum_{k=1}^{n} E\left(|X_1|^2 \cdot \mathbb{1}_{|X_1| \leq a_k}\right) + \sum_{k=1}^{n} E\left(|X_1|^2 \cdot \mathbb{1}_{a_k \leq |X_1| \leq n}\right) \right)$$

2nd term

1st term:

$$\leq \sum_{k=1}^{n} a_k \cdot E\left(|X_1| \cdot \mathbb{1}_{|X_1| \leq a_k}\right)$$

$$\leq \sum_{k=1}^{n} a_k \cdot E(|X_1|) =$$

$$\therefore \frac{1}{n^2\varepsilon^2} \cdot (1^{st} \text{ term}) = \frac{1}{n^2\varepsilon^2} \cdot \sum_{k=1}^{n} a_n \cdot E(|X_1|)$$

$$\therefore \frac{1}{n^2\varepsilon^2}\cdot(1^{st}\text{ term}) = \frac{1}{n^2\varepsilon^2}\cdot\sum_{k=1} a_n \cdot E(|X_1|)$$

$$= \frac{\cancel{n} a_n \cdot E|X_1|}{n^{\cancel{2}}\varepsilon^2} \longrightarrow 0$$

&, 2^{nd} term,

$$\sum_{k=1}^{n} E\left(X_1^2 \cdot \mathbb{1}_{a_n < |X_1| \leq n}\right)$$

$$\leq \sum_{k=1}^{n} n\, E\left(|X_1| \cdot \mathbb{1}_{a_n < |X_1| \leq n}\right)$$

$$\leq \sum_{k=1}^{n} n \cdot E\left(|X_1| \cdot \mathbb{1}_{|X_1| > a_n}\right)$$

$$\therefore \frac{1}{n^2\varepsilon^2}(2^{nd}\text{ term}) = \frac{1}{n^{\cancel{2}}\varepsilon^2}\cdot \cancel{n}\cdot \sum E\left(|X_1| \cdot \mathbb{1}_{|X_1| > a_n}\right)$$

$$= \frac{1}{\cancel{n}\varepsilon^2} \times \cancel{n} \cdot E\left(|X_1| \cdot \mathbb{1}_{|X_1| > a_n}\right)$$

$$= \frac{E\left(|X_1| \cdot \mathbb{1}_{|X_1| > a_n}\right)}{\varepsilon^2} \xrightarrow{(D.C.T.)^*} 0$$

$\because X_1$ – finite 1^{st} moment $\therefore E|X_1| \leq M < \infty$

Hence, we are done. $\blacksquare$

Claim: for WLLN, we only need to show,

$$P\left(\left|\frac{S_n - ES_n}{n}\right| > \varepsilon\right) \longrightarrow 0$$

for SLLN, we need to show

$$\sum P\left(\left|\frac{S_n - ES_n}{n}\right| > \varepsilon\right) < \infty .$$

$$\sum_n P\left(\left|\frac{S_n - ES_n}{n}\right| > \varepsilon\right) < \infty .$$

Borel's Strong Law of Large Numbers:

Proved (supposedly) SLLN under stronger assumption that,

$$E|X_1|^4 < \infty \qquad \left(4^{th} \text{ moments finite}\right)$$

essentially centering the X_is.

Denote $Y_n := X_n - EX_n = X_n - \mu$

$T_n = Y_1 + \cdots + Y_n$.

We have to now show: $\frac{T_n}{n} \xrightarrow{a.s} 0$.

ie, $$\sum_n P\left(\left|\frac{T_n}{n}\right| > \varepsilon\right) < \infty .$$

$$P\left(\left|\frac{T_n}{n}\right| > \varepsilon\right) \leq \frac{E(T_n^4)}{n^4 \varepsilon^4} \qquad [\text{Chebyshev.}]$$

Y_i's . independent

$$E(T_n^4) = E\left(\left(\sum_{i=1}^n Y_i\right)^4\right)$$

$$4 . \sum_{i \neq j} E(Y_i^3 . Y_j) = 4 . \sum E(Y_i^3) . E(\cancelto{0}{Y_j})$$

Now, $$E\left(\sum_{i=1}^4 Y_i\right)^4 = E\left(\sum_{i=1}^n Y_i^4\right) + \cancelto{0}{4 . E\left(\sum_{i \neq j}^n Y_i^3 . Y_j\right)} +$$

$$\cancelto{0}{4\left(\quad\right)} + \cancelto{0}{\left(\quad\right)} +$$

(similarly)

$$6 . E\left(\sum_{i \neq j} Y_i^2 . Y_j^2\right)$$

$$= n\, E(Y_1^4) + 6 . n(n-1) . E(Y_1^2 . Y_2^2)$$

$$= n\,E(Y_1^4) + 6\cdot n(n-1)\cdot E(Y_1^2\cdot Y_2^2)$$

$$\therefore P\left(\left|\frac{T_n}{n}\right| > \varepsilon\right) \le \frac{E(T_n^4)}{n^4\varepsilon^4} = \frac{n\,E(Y_1^4) + 6\,n(n-1)\,E(Y_1^2 Y_2^2)}{n^4\varepsilon^2}$$

of the order $\frac{1}{n^2}$

$$\therefore \sum(\cdot) < \infty \;\checkmark$$

$$\therefore \frac{S_n}{n} \xrightarrow{a.s} E|X_1|.$$

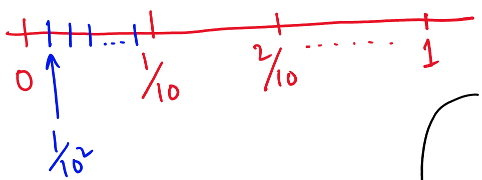
$\Omega = (0,1]$, $\mathcal{A}$ = Borel σ-field

P = leb. measure

$$X(\omega) = \omega$$

$$X_1(\omega) = \begin{cases} 0, & \text{if } \omega \in (0, 1/10] \\ 1, & \text{if } \omega \in (1/10, 2/10] \\ \vdots & \\ 9, & \text{if } \omega \in (9/10, 1] \end{cases}$$

$$|\omega - X_1(\omega)| \le \frac{1}{10}$$



$$X_2(\omega) = \begin{cases} 0, & \text{if } \omega - \frac{X_1(\omega)}{10} \in \left(0, \frac{1}{10^2}\right] \\ 1, & \text{if } \omega - \frac{X(\omega)}{10} \in \left(\frac{1}{10^2}, \frac{2}{10^2}\right] \\ \vdots & \\ 9, & \text{if } \omega - \frac{X(\omega)}{10} \in \left(\frac{9}{10^2}, \frac{1}{10}\right] \end{cases}$$

This is basically the non-terminating

This is basically the non-terminating decimal expansion of ω

ie, $\omega = 0 . X_1(\omega) X_2(\omega) \cdots$

decimal

1st place after decimal

2nd place after decimal.

Note that, X_i's are identically distributed.

Q. are they independent ?

$P(X_1=2, X_2=5, X_3=1)$

$= P\left(\omega : \omega \in \left(\frac{2}{10}+\frac{5}{10^2}+\frac{1}{10^3}, \frac{2}{10}+\frac{5}{10^2}+\frac{2}{10^3}\right)\right)$

$= \frac{1}{10^3} = P(X_1=2) \cdot P(X_2=5) \cdot P(X_3=1)$

Yes ✓.

$\therefore$ We have i.i.d seq of r.vs, bounded

ie, all moments finite.

Common mean $= \frac{0+1+2+\cdots+9}{10}$

$= 9/2$

$\therefore$ By Borel SLLN,

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow{a.s} 9/2$$

Now, say, $Z_n := 1_{X_n=3}$.

Q. are Z_n's iid = ? Yes!!

Yes!!

$\because P(Z_n=1)$

$= P(X_n=3)$,

& X_n's are

i.i.d.

$$E(Z_n) = 1 \cdot \frac{1}{10} + 0 + 0 + \cdots + 0$$

$\frac{1}{10} \downarrow P(X_n=3)$

this is the actually proportion of 3's in the first n decimal places

$$\left(\frac{Z_1+Z_2+\cdots+Z_n}{n}\right) \xrightarrow{a.s} \frac{1}{10}.$$

$$P\left(\omega : \text{For any } i \in \{0,1,\ldots,10\}. \text{ proportion of } \underline{i} \text{ in decimal expansion of } \omega \longrightarrow \frac{1}{10}\right) = 1$$

* Borel's Normal Number Theorem:

↳ "almost every no. is normal".

↓ a no. for which, for every k, the k-adic expansion of the that no., proportion of every no. in that expansion $\to \frac{1}{k}$

Probability-3 Lecture-13

20 September 2024 11:23

Again, full respect to A.N. Kolmogorov

Classical SLLN:

$\{X_n\}$ - sequence of i.i.d r.vs with finite common mean μ.

Then, $\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} \mu$

$$\Leftrightarrow \frac{(X_1-\mu) + (X_2-\mu) + \cdots + (X_n-\mu)}{n} \xrightarrow{a.s.} 0$$

So, we'll prove, if $\{X_n\}$ - sequence of i.i.d r.vs with common mean, $EX_1 = 0$

then, $\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} 0$

Here, $\forall \varepsilon > 0$,

$$\sum_n P\left(\left| \frac{X_1 + \cdots + X_n}{n} \right| > \varepsilon \right) < \infty$$

ie, to show that, this series converges [which is stronger than WLLN!!]

Kronecker's Lemma:

$\to$ a result on real sequences.

Let $\{x_n\}$ be any real sequence.

Then, if the series $\sum_n \frac{x_n}{n}$ converges, then

$$\frac{1}{n}(x_1 + \cdots + x_n) \longrightarrow 0.$$

Proof:

Let $b_n = \sum_{k=1}^{n} x_k/k$, $n = 1, 2, 3, \ldots$

partial sums

$b_0 = 0.$

Hypothesis: b_n converges. say, $b_n \longrightarrow b$.

$\therefore \forall k \geq 1$, $b_k - b_{k-1} = k^{th}$ term $= \frac{x_k}{k}$

$$\Rightarrow x_k = k(b_k - b_{k-1})$$

$$x_1 + \cdots + x_n = \sum_{k=1}^{n} k \cdot b_k - \sum_{k=1}^{n} k \cdot b_{k-1}$$

$$= \sum_{k=1}^{n} k b_k - \sum_{k=1}^{n} (k-1) \cdot b_{k-1} - \sum_{k=1}^{n} b_{k-1}$$

change of variable $k-1 \to j$

$$= \sum_{k=1}^{n} k \cdot b_k - \sum_{j=1}^{n-1} j b_j - \sum_{k=1}^{n-1} b_k$$

change of variable $k-1 \to k$.

$$= n b_n - \sum_{k=1}^{n-1} b_k$$

$$\therefore \frac{x_1 + \cdots + x_n}{n} = \underbrace{\frac{\cancel{n} b_n}{\cancel{n}}}_{\downarrow\, b} - \underbrace{\left(\frac{1}{n} \sum_{k=1}^{n-1} b_k\right)}_{\downarrow\, b} \longrightarrow 0. \checkmark \quad \blacksquare$$

back to SLLN...

By Kronekar's lemma, its enough to show, (ie, sufficient) this "random series"

$$\sum_n \frac{X_n}{n} \text{ converges a.s.}$$

Q: So, what should I prove to get that

$$\sum_n \frac{X_n}{n} \text{ converges a.s. ??}$$

Let $\{y_n\}$ be a real sequence.

the series $\sum_n y_n$ converges <u>iff</u> the sequence of partial sums are cauchy.

ie, $\forall j \geq 1, \exists n \geq 1$ s.t $\forall$ $m' > m \geq n$,

$$\left| \sum_{k=1}^{m'} y_k - \sum_{k=1}^{m} y_k \right| \leq 1/j$$

$$\Leftrightarrow \left| \sum_{k=m+1}^{m'} y_k \right| \leq 1/j$$

So, back again...

$$Y_n := \frac{X_n}{n}.$$

$\{Y_n\}$ - sequence of real r.vs, then the series

$$\sum_n Y_n \text{ converges a.s.}$$

$$\Leftrightarrow P\left(\bigcap_{j\geq 1} \bigcup_{n \geq 1} \bigcap_{m' > m \geq n} \left\{ \omega : \left| \sum_{k=m+1}^{m'} Y_k(\omega) \right| \leq \frac{1}{j} \right\} \right) = 1$$

take complement

$$\Leftrightarrow P\left(\bigcup_{j\geq 1} \bigcap_{n \geq 1} \bigcup_{m' > m \geq n} \left\{ \omega : \left| \sum_{k=m+1}^{m'} Y_k(\omega) \right| > \frac{1}{j} \right\} \right) = 0$$

$$\Leftrightarrow P\left(\bigcup_{j\geq 1}\bigcap_{n\geq 1}\bigcup_{m'>m\geq n}\left\{\omega : \left|\sum_{k=m+1}^{m'} Y_k(\omega)\right| > \frac{1}{j}\right\}\right)=0$$

$$\Leftrightarrow P\left(\bigcup_{j\geq 1}\bigcap_{n\geq 1}\left\{\omega : \sup_{m'>m\geq n}\left|\sum_{k=m+1}^{m'} Y_k(\omega)\right| > \frac{1}{j}\right\}\right)=0.$$

this set is decreasing in n

ie, as $n\uparrow$, this set shrinks.

$\Leftrightarrow \forall j\geq 1,$

$$P\left(\bigcap_{n\geq 1}\left\{\omega : \sup_{m'>m\geq n}\left|\sum_{k=m+1}^{m'} Y_k(\omega)\right| > \frac{1}{j}\right\}\right)=0$$

decreasing events. $\therefore$ By continuity of probability.

$$\Leftrightarrow \forall j\geq 1,\ P\left(\left\{\omega : \sup_{m'>m\geq n}\left|\sum_{k=m+1}^{m'} Y_k(\omega)\right| > \frac{1}{j}\right\}\right)\to 0 \text{ as } n\to\infty$$

take $m=n$, $m'=m$ $\Downarrow$ $\Uparrow$ Δ-inequality (* Exc: verify this)

$$\Leftrightarrow \forall j\geq 1,\ P\left(\left\{\omega : \sup_{m>n}\left|\sum_{k=n+1}^{m} Y_k(\omega)\right| > \frac{1}{j}\right\}\right)\to 0 \text{ as } n\to\infty.$$

So now, fix $\varepsilon>0$.

then,

$$\sup_{m>n}\left|\sum_{k=n+1}^{m} Y_k\right| = \lim_{m\to\infty}\uparrow \max_{n<l\leq m}\left|\sum_{k=n+1}^{l} Y_k\right|$$

Supremum $\equiv$ (increasing limit of partial maximums)

$$\therefore \sup_{m>n}\left|\sum_{k=n+1}^{m} Y_k\right| > \varepsilon$$

$m > n \quad | \quad \overline{k=n+1} \quad |$

$$\iff \max_{n<l\le m} \left| \sum_{k=n+1}^{l} Y_k \right| > \varepsilon \quad \text{for some } m > n.$$

$\left[\text{Since } \max_{n<l\le m} \left\{ \left| \sum_{k=n+1}^{l} Y_k \right| \right\} \right.$ is an increasing sequence in m, $\therefore \forall$ for all $m' > m$, it holds. $\left.\right]$

Kolmogorov's Maximal Inequality

$\xi_1, \xi_2, \ldots, \xi_n$ - independent r.v.s with 0 means & finite variances.

$$P\left(\max_{1\le k\le n} \left| \sum_{i=1}^{k} \xi_i \right| > \varepsilon \right) \le \frac{\operatorname{Var}\left(\sum_{i=1}^{n} \xi_i\right)}{\varepsilon^2}$$

this set $\equiv A$ (say)

(stronger than Chebyshev !!)

[Compare:

$$P\left(\left| \sum_{i=1}^{n} \xi_i \right| > \varepsilon \right) \le \frac{\operatorname{Var}\left(\sum_{i=1}^{n} \xi_i\right)}{\varepsilon^2}$$

But yes, Chebyshev doesn't need independence, but Kolmogorov's Maximal ineq. assumes independence.]

Proof: "Think" of this as a random walk.

ξ_i - increment in the ith step.

$$A_k := \left\{ \left|\sum_{i=1}^{k-1} \xi_i\right| \le \varepsilon \quad \forall\, i=1,2,\dots,k-1, \quad \underline{\left|\sum_{i=1}^{k} \xi_i\right| > \varepsilon} \right\}$$

k is the first time the threshold $(-\varepsilon, \varepsilon)$ has been crossed

Clearly, $\bigcup_{i=1}^{n} A_k = A = P\left(\max_{1\le k\le n} \left|\sum_{i=1}^{k} \xi_i\right| > \varepsilon\right)$.

& , $A_1, A_2, \dots, A_n$ – disjoint (trivial)

$$\therefore P(A) = \sum_{k=1}^{n} \underbrace{P(A_k)}_{= E(A_k)} \le \frac{1}{\varepsilon^2} \cdot \sum_{k=1}^{n} E\left(\left|\sum_{i=1}^{k} \xi_i\right|^2 \cdot 1_{A_k}\right)$$

How ?

at k, $\left|\sum_{i=1}^{k} \xi_i\right| > \varepsilon$

$\therefore \left|\sum \xi_i\right|^2 > \varepsilon^2$

Hence, this inequality holds.

$$= \frac{1}{\varepsilon^2} \cdot \left[\sum_{k=1}^{n} E\left(\left|\sum_{i=1}^{n} \xi_i\right|^2 \cdot 1_{A_k}\right) - \sum_{k=1}^{n} E\left(\left|\sum_{i=k+1}^{n} \xi_i\right|^2 \cdot 1_{A_k}\right) - 2\sum_{k=1}^{n} E\left(\left(\sum_{i=1}^{n} \xi_i\right) \cdot \left(\sum_{i=1}^{k} \xi_i\right) \cdot 1_{A_k}\right)\right.$$

$\left[\because\ a^2 = b^2 - (b-a)^2 - 2a(b-a)\right.$

[$\because a^2 = b^2 - (b-a)^2 - 2a(b-a)$

Here,

$$a = \sum_{i=1}^{k} \xi_i$$

$$b = \sum_{i=1}^{n} \xi_i$$
]

Now, last term: (i.e, 3rd term)

$$-2\sum_{k=1}^{n} E\left(\underbrace{\left(\sum_{i=k+1}^{n} \xi_i\right)}_{\text{depends on rest of the } \xi_i\text{'s}} \cdot \underbrace{\left(\sum_{i=1}^{k} \xi_i\right) \cdot 1_{A_k}}_{\text{depends on } \xi_1, \ldots, \xi_k}\right)$$

Independent

$$= -2 \sum \cdot \cancel{E\left(\sum_{i=k+1}^{n} \xi_i\right)}^{\,0} \cdot E\left(\quad \cdot \quad \right)$$

[as per hypothesis.]

& 2nd term ≥ 0. So, removing ($-$ 2nd term) preserves the inequality.

$$\therefore P(A) \leq \frac{1}{\varepsilon^2} \cdot \sum_{k=1}^{n} E\left(\left|\sum_{i=1}^{n} \xi_i\right|^2 \cdot 1_{A_k}\right)$$

$$\leq \frac{1}{\varepsilon^2} \cdot E\left(\left|\sum_{i=1}^{n} \xi_i\right|^2 \cdot 1_A\right) \qquad \left[\because A = \bigcup_{k=1}^{n} A_k\right]$$

$$\leq \frac{1}{\varepsilon^2} \cdot Var\left(\sum_{i=1}^{n} \xi_i\right).$$

[$\because$ all cross terms $= 0$

"Fundamental law of Statistics"]

law of Statistics"

... Again, back to SLLN :

Have to show : $\forall \varepsilon > 0, \quad P\left(\sup_{m > n} \left| \sum_{k=n+1}^{m} \frac{X_k}{k} \right| > \varepsilon \right) \longrightarrow 0.$

from Kolmogorov's maximal inequality,

$$P\left(\max_{n \leq l \leq m} \left| \sum_{k=n+1}^{l} \frac{X_k}{k} \right| > \varepsilon \right) \leq \frac{1}{\varepsilon^2} \cdot \operatorname{Var}\left(\sum_{k=n+1}^{m} \frac{X_k}{k} \right)$$

$$= \frac{1}{\varepsilon^2} \cdot \sum_{k=n+1}^{m} \frac{1}{k^2} \cdot \operatorname{Var}(X_k)$$

bound by the entire series further

$$\leq \frac{1}{\varepsilon^2} \cdot \sum_{k=n+1}^{\infty} \frac{1}{k^2} \cdot \operatorname{Var}(X_k)$$

to say, this $\downarrow 0$,

Goal: (next week) only thing that remains to show is that,

if $X_1, X_2, \ldots$ iid with 0 common mean,

then $\displaystyle\sum_{k=1}^{\infty} \frac{\operatorname{Var}(X_k)}{k^2} < \infty.$

(ie, only then, the tail

$$\sum_{k=n+1}^{\infty} \frac{\operatorname{Var}(X_k)}{k^2} \to 0,$$

and that's precisely what we need.)

Probability-3 Lecture-14

24 September 2024 14:17

Recall:

<u>Result:</u>

$\{\xi_i\}$ – an independent, mean zero seq. of r.vs.

then, $\sum_n V(\xi_n) < \infty$ is a sufficient condition

for random series $\sum_n \xi_n$ to converge a.s.

<u>Corollary:</u>

Say, $\eta_n = \pm 1$, with prob $= \frac{1}{2}$ each, independent

$\therefore$ By the result above: $\sum_n \eta_n \cdot \frac{1}{n}$ converges a.s.

$\left[\because \{\eta_n\} \text{ – mean } 0 \text{ seq.} \quad E(\eta_n) = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0 \right]$

[check that: $\sum_n V(\eta_n) < \infty$]

"The Random Signs Problem"

$\sum_n V(\eta_n) = \sum \left(\pm \frac{1}{n}\right)^2 = \sum \frac{1}{n^2} < \infty$. ✓

<u>Corollary:</u>

$\{\xi_n\}$ – independent seq.

Then, $\underbrace{\sum E(\xi_n) \text{ converges } \& \sum V(\xi_n) < \infty}$

(Exercise) $\Rightarrow \sum_n \xi_n$ converges a.s.

* $\sum_n (\xi - E(\xi))$ converges a.s.

complete this.

[Converse <u>not</u> true]

But, partial converse:

if $|\xi_n| \le K$,

then the converse holds.

Back to SLLN:

$\{X_n\}$ - iid sequence with finite mean μ.

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} \mu .$$

Step 1: Truncation.

for each $n \geq 1$, let $Y_n = X_n \cdot \mathbb{1}_{|X_n| \leq n}$

clearly, Y_n's - independent, but not identically distributed.

$$Y_n \stackrel{d}{=} X_1 \cdot \mathbb{1}_{|X_1| \leq n} .$$

Let $T_n = \sum_{k=1}^{n} \left(Y_k - E(Y_k) \right)$ ← i.e., centering the Y_n's & taking their partial sums.

clearly, $\left\{ \underbrace{\frac{Y_n - E(Y_n)}{n}} \right\}$ is an independent, 0-mean sequence.

↓

think of this as ξ_n.

We want to show,

$$\sum_n \frac{Y_n - E(Y_n)}{n} \text{ converges a.s.}$$

If we can do that, then by Kronekar's lemma,

$$\Rightarrow \frac{1}{n} \cdot \sum_{k=1}^{n} \left(Y_k - E(Y_k) \right) \xrightarrow{a.s.} 0, \quad \text{—— } (1)$$

& that suffices. ??

& that suffices.
??

& $\frac{1}{n}\sum_{k=1}^{n}\left(X_k - E(Y_k)\right) \xrightarrow{a.s} 0 \overset{??}{\Longrightarrow}$ classical SLLN

Observation 1:

$$\sum_n P(X_n \neq Y_n) = \sum_n P(|X_n| > n)$$

$$= \sum_n P(|X_1| > n) < \infty$$

$[\because E|X_1| < \infty]$

$\Rightarrow$ By Borel-Cantelli Lemma.

$P(X_n \neq Y_n$ for infinitely many n's$) = 0$

$\Rightarrow P(X_n = Y_n$ for all "large" n's$) = 1.$

$\Rightarrow \frac{1}{n}\sum_{k=1}^{n}(Y_k - X_k) \xrightarrow{a.s} 0$ —— 2

[think why? only finitely many terms in numerator are significant, while denominator $\nearrow \infty$]

$\therefore$ 1 & 2 $\Rightarrow$

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - EY_k) \xrightarrow{a.s.} 0 \text{ —— } 3$$

Observation 2:

$$E(Y_n) = E\left(X_1 \cdot 1_{|X_1| \le n}\right)$$

$\downarrow$

$E(X_1)$ as $n \to \infty$ [By DCT]

$= \mu.$

Also, $\frac{1}{n}\sum_{k=1}^{n} E(Y_k) \longrightarrow \mu = E(X_k).$ —— 4

Also, $\frac{1}{n}\sum_{k=1}^{} E(Y_k) \longrightarrow \mu = E(X_k)$. —— (4)

(3) + (4) $\Rightarrow$

$$\frac{1}{n}\cdot\sum_{k=1}^{n}(X_k - E(X_k)) \xrightarrow{a.s} 0$$

$$\Longleftrightarrow \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s} \mu.$$

SLLN ✓.

So, left to show:

$$\frac{1}{n}\sum_{k=1}^{n}(Y_k - E(Y_k)) \xrightarrow{a.s} 0$$

$$\sum_n V\left(\frac{Y_n - EY_n}{n}\right) \leq \sum_n \frac{1}{n^2}\cdot E(Y_n^2)$$

$$= \sum_n \frac{1}{n^2}\cdot E\left(|X_1|^2\cdot 1_{|X_1|\leq n}\right)$$

$$= \left(\sum_n \frac{1}{n^2}\right)\sum_{j=1}^{n} E\left(|X_1|^2\cdot 1_{j-1<|X_1|\leq j}\right)$$

take inside.

$$= \sum_{j=1}^{\infty}\left(E\left(|X_1|^2\cdot 1_{j-1<|X_1|\leq j}\right)\cdot \underbrace{\sum_{n=j}^{n}\frac{1}{n^2}}_{\text{(check!!)} \leq \frac{2}{j}}\right)$$

[$|X_1| = 0$ does not contribute to the expection.]

$$\leq \sum_{j=1}^{\infty}\frac{2}{j}\cdot E\left(|X_1|^2\cdot 1_{j-1<|X_1|\leq j}\right)$$

pull $|X_1|$ out

$$\leq \sum_{j=1}^{\infty}\frac{2}{\cancel{j}}\cdot\cancel{j}\cdot E\left(|X_1|\cdot 1_{j-1<|X_1|\leq j}\right)$$

$$j=1$$

$$= 2 \cdot \sum_{j=1}^{\infty} \underbrace{E\left(|X_1| \cdot 1_{j-1 < |X_1| \le j}\right)}_{E|X_1|}$$

$$= 2 \cdot E|X_1| < \infty . \quad \blacksquare$$

SLLN proved !!!

A Trivial Application: (of SLLN).

F - unknown distribution. (ie, kind of "non-parametric": ie, not just the parameter, but also the distribution is not known.)

$X_1, X_2, \dots \overset{iid}{\sim} F$.

Fix $x \in \mathbb{R}$.

Goal: to estimate $F(x)$.

Approach: take a sample of size n.

$$X_1, \dots, X_n .$$

$$\therefore F_n(x) := \frac{1_{X_1 \le x} + \dots + 1_{X_n \le x}}{n}$$

$\leftarrow$ (proportion of X_i's $\le x$)

each of these r.vs,

$1_{X_i \le x}$ - iid, with common mean

$$E(1_{X_i \le x}) = P(X_i \le x) = F(x).$$

ie, by SLLN,

$= F(x)$.

ie, by SLLN,

$$F_n(x) \xrightarrow{a.s.} F(x)$$ & hence, as well as: $$F_n(x) \xrightarrow{P} F(x)$$

F_n : Empirical distribution f^n based on n observations.

→ discrete distⁿ.

Glivenko - Cantelli Lemma:

$$\sup_x |F_n(x) - F(x)| \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.$$

"the Kolmogorov Metric"

ie, ∃ a single P-null set, s.t, ∀ w ∉ that P-null set, the function $F_n(x) \longrightarrow F(x)$ uniformly.

Recall: Ana-2

Proof:

Trivial consequence of SLLN:

for each $x \in \mathbb{R}$, $F_n(x) \xrightarrow{a.s} F(x)$

Let N_1 be a P-null set s.t. ∀ w outside N_1.

$$F_n(r)(\omega) \longrightarrow F(r)(\omega) \quad \forall\ r \in \mathbb{Q}.$$

[for each r, we get a P-null set N_r that countable union & define

$$N_1 = \bigcup_{r \in \mathbb{Q}} N_r \quad \text{——} \quad (1)$$]

Another consequence of SLLN:

If x is a discontinuity point of F,

then, $$F_n(x) - F_n(x^-) \xrightarrow{a.s} F(x) - F(x^-).$$

$$F_n(x) - F_n(x^-) \longrightarrow F(x) - F(x^-)$$

our required iid seq: $\mathbb{1}_{X_1 = x}, \mathbb{1}_{X_2 = x}, \ldots$

$$E\left(\mathbb{1}_{X_1 = x}\right) = F(x) - F(x^-).$$
$$= P(X \le x) - P(X < x)$$

& , $$F_n(x) - F_n(x^-) = \frac{\#\{k : X_k \le x\} - \#\{k : X_k < x\}}{n}.$$

$\therefore$ Let N_2 be a null set s.t. outside N_2,

$$F_n(x) - F_n(x^-) \longrightarrow F(x) - F(x^-). \qquad \text{——} (2)$$

$\bigcup_x N_x$, where x - discont. pts [countably many at most]

$\therefore$ By taking the required p-null set to be $N = N_1 \cup N_2$,

we get, outside of N, both (1) & (2) hold.

Claim: Outside of N,

$$\sup_x \left| F_n(x) - F(x) \right| \xrightarrow{a.s} 0 \quad \text{as} \quad n \to \infty.$$

Proof by contradiction:

Suppose not.

ie, $\exists\, \varepsilon > 0$, & a subsequence

$$1 < n_1 < n_2 < \cdots < n_k < \cdots \uparrow \infty$$

& a real sequence $\{x_k\}$

s.t. $$\left| F_{n_k}(x_k) - F(x_k) \right| \geq \varepsilon.$$

ie, n_1 is the first index where,

first index where, the sup$|\cdot| > \varepsilon$.

s.t. $|F_{n_k}(x_k) - F(x_k)| \geq \varepsilon$.

n_2 is the next index where this happens.

Now, for $\sup_x |F_{n_1}(x) - F(x)| > \varepsilon$

$\exists\, x = x_1$ for which this holds.

Similarly, $\sup_x |F_{n_2}(x) - F(x)| > \varepsilon$

for some $x = x_2$.

that's how we obtain $\{x_k\}$.

Claim: the sequence $\{x_k\}$ - cannot be unbounded.

i.e, It has to be both bounded above & bounded below.

$\because F(x) \to 0$ as $x \to -\infty$.

$\therefore \exists\, r \in \mathbb{Q}$ (largely -ve)

s.t

$F(r) < \varepsilon/2$

$\Rightarrow F_{n_k}(r) < \varepsilon/2$

just a subsequence of $F_n(r)$

$\left[\because F_{n_k}(r) \to F(r)\right]$

If $\{x_k\}$- not bounded below, then $x_k < r$ for infinitely many k

$\therefore$ (for this r, there must exist x_{n_k} st. $x_{n_k} < r$, as we've assumed x_{n_k} to be unbounded below.)

$\Rightarrow F(x_k) \leq F(r) < \varepsilon/2$

$F_{n_k}(x_k) \leq F(r) < \varepsilon/2$

for infinitely

$F_{n_k}(x_k) \le F(r) < \varepsilon/2$) for infinitely many k. below.

$\therefore \left| F_{n_k}(x_k) - F(x_k) \right| < \varepsilon/2$, which is a contradiction!!!

$\therefore \{x_k\}$ must be bounded below.

Similarly, $\because F(x) \to 1$ as $x \to \infty$

$\therefore \exists r \in \mathbb{Q}$ st,

$F(r) > 1 - \varepsilon/2$

$\Rightarrow 1 - F(r) < \varepsilon/2$

By similar argument $\Rightarrow 1 - F_{n_k}(r) < \varepsilon/2$ $\forall$ large k.

then, we can conclude that $\{x_k\}$ is bounded above.

Now, by passing to a subsequence, if necessary, (& still calling it $\{x_k\}$), we may assume $x_k \longrightarrow a \in \mathbb{R}$.

Exercise: (Ana-1) By again passing through subsequences, if necessary, we assume atleast one of the following holds:

(1) $x_k < a$, $x_k \uparrow a$, $F_{n_k}(x_k) - F(x_k) > \varepsilon$.

(2) $x_k < a$, $x_k \uparrow a$, $F(x_k) - F_{n_k}(x_k) \geq \varepsilon$

(3) $x_k \geq a$, $x_k \downarrow a$, $F_{n_k}(x_k) - F(x_k) \geq \varepsilon$

(4) $x_k \geqslant a$, $x_k \downarrow a$, $F(x_k) - F_{n_k}(x_k) \geqslant \varepsilon$

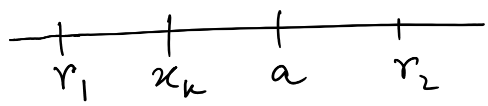
$\therefore$ No. of "k"s for which all 4 above holds is all $k \in \mathbb{N}$. ← infinite

$\therefore$ Atleast one of the 4 cases above must happen for infinitely many k's.

W.L.O.G, say, (1) holds for infinitely many k's.

So, Suppose, we are in case-(1) this $a \in \mathbb{R}$.

first, let's assume, 'a' is a continuity point of F



$$\varepsilon \leqslant \underbrace{F_{n_k}(x_k)}_{\leqslant} - \underbrace{F(x_k)}_{\geqslant}$$

$$\leqslant F_{n_k}(r_2) - F(r_1) + F(r_2) - F(r_2)$$

$$= \underbrace{F_{n_k}(r_2) - F(r_2)}_{\to 0} - \big(F(r_1) - F(r_2)\big)$$

$$= \underbrace{F(r_2) - F(r_1)}_{\to 0 \text{ as } r_1 \downarrow a}$$

[$\because$ 'a' is a continuity point.]

as $r_1 \downarrow a$

$r_2 \nearrow a$

"You'd see it if you want to see it,

you'd not see if you don't want to see it."

— Prof. AG,
24th Sept, '24

Now, assume a to be a point of discontinuity of F

$$\varepsilon \leq F_{n_k}(x_k) - F(x_k)$$

$$\leq F_{n_k}(\bar{a}) - F_{n_k}(a) + F_{n_k}(\bar{a}) - F(x_k)$$

Probability-3 Lecture-15

27 September 2024 11:23

Lemma:

Let $F_n, n \geq 1$ and F be cdf on $\mathbb{R}$

Assume, on a dense set $D \subset \mathbb{R}$

$$F_n(x) \longrightarrow F(x) \quad \forall\, x \in D, \text{ \&,}$$

$$\forall\, x \in J(F) = \text{set of all discontinuities of } F.$$

$$F_n(x) - F_n(x^-) \longrightarrow F(x) - F(x^-)$$

Then, $F_n \longrightarrow F$ on $\mathbb{R}$ uniformly, i.e,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \longrightarrow 0.$$

Definitions: $(\Omega, \mathcal{A}, P)$ — probability space (fixed).

$\mathcal{G}_1$ & $\mathcal{G}_2$ be sub-σ fields of $\mathcal{A}$ are said to be independent if

* In practice, proving independence through this might be difficult.

$$P(G_1 \cap G_2) = P(G_1) \cdot P(G_2)$$

$$\forall\, G_1 \in \mathcal{G}_1,\ G_2 \in \mathcal{G}_2$$

Result:

$\mathcal{G}_2$ - smallest σ-field gen. by $\mathcal{S}_2$

If $\mathcal{S}_1$ & $\mathcal{S}_2$ are semifields s.t.

$$\sigma(\mathcal{S}_1) = \mathcal{G}_1, \quad \sigma(\mathcal{S}_2) = \mathcal{G}_2,$$

and if $P(S_1 \cap S_2) = P(S_1) \cdot P(S_2) \quad \forall\, S_1 \in \mathcal{S}_1, S_2 \in \mathcal{S}_2$

then, $\mathcal{G}_1$ & $\mathcal{G}_2$ are independent.

Definition:

Let Λ- indexing set.

Let $\{\mathcal{G}_\alpha, \alpha \in \Lambda\}$ be a family of sub-σ fields of $\mathcal{A}$.

$\{ \mathcal{G}_\alpha, \alpha \in \Lambda \}$ ... sub-σ fields of $\mathcal{a}$.

Then, $\{ \mathcal{G}_\alpha, \alpha \in \Lambda \}$ are said to be **mutually independent** if for any choice of $\alpha_1, \ldots, \alpha_n \in \Lambda$,

$$P(G_1 \cap G_2 \cap \ldots \cap G_n) = \prod_{i=1}^{n} P(G_i) \quad \forall$$

$$G_1 \in \mathcal{G}_{\alpha_1}, \ldots, G_n \in \mathcal{G}_{\alpha_n}.$$

Result:

If for each $\alpha \in A$, $\mathcal{S}_\alpha$ is a semi-field s.t. $\sigma(\mathcal{S}_\alpha) = \mathcal{G}_\alpha$.

then
$$P\left(\bigcap_{i=1}^{n} S_i\right) = \prod_{i=1}^{n} P(S_i) \quad \text{for all choices of } \alpha_1, \ldots, \alpha_n \in \Lambda.$$

is sufficient for $\{ \mathcal{G}_\alpha, \alpha \in \Lambda \}$, and all choices of $S_i \in \mathcal{S}_{\alpha_i}$ to be independent. $\forall i = 1, \ldots, n$.

$(\Omega, \mathcal{a}, P)$.

Given a family $\{ X_\alpha, \alpha \in \Lambda \}$ of r.vs, the smallest σ-field on Ω w.r.t. which all $X_\alpha, \alpha \in \Lambda$ are measurable, is called the σ-field generated by $\{X_\alpha, \alpha \in A\}$, denoted by

$$\sigma(\{X_\alpha, \alpha \in A\}).$$

$\mathcal{C} = \{ S \subset \Omega : S = \bigcap^{n} X^{-1}(B) \ldots \alpha \ldots \alpha \in \Lambda \}$

$$\mathcal{S} = \left\{ S \subset \Omega : S = \bigcap_{i=1}^{n} X_{\alpha_i}^{-1}(B_i), \ \begin{array}{l} \alpha_1, \ldots, \alpha_n \in \Lambda \\ B_1, \ldots, B_n \in \mathcal{B} \end{array} \right\}$$

$$\sigma(\mathcal{S}) = \sigma(\{ X_\alpha, \alpha \in A \}).$$

KOLMOGOROV'S 0-1 LAW

Setup: Let $\{X_n, n \geq 1\}$ - sequence of independent r.vs

ie, $\{\sigma(X_n), n \geq 1\}$ - is an independent sequence of σ-fields.

For each $n \geq 1$, define

$$\mathcal{A}_n = \sigma(X_1, \ldots, X_n)$$

these are $\uparrow$ in n.

Any event determined by the first n random variables.

check: $\bigcup_n \mathcal{A}_n$ is a field.

($\because$ increasing union of σ-fields is a field.)

$\therefore$ the σ-field generated by this,

$$\sigma\left(\bigcup_n \mathcal{A}_n\right) = \mathcal{A}_\infty.$$

Check: this is the smallest σ-field wrt which all the X_n's are measurable.

Now, take $\mathcal{T}_n := \sigma(X_{n+1}, X_{n+2}, \ldots)$

$\hookrightarrow$ any event that depends only on the tail.

Note: $\mathcal{T}_n$ decreases $\downarrow$ with n.

Note: $\mathcal{T}_n$ decreases $\downarrow$ with n.

[Eg. of an event that belong to $\mathcal{T}_n$:
$\{\omega : X_n(\omega) = 0$ for infinitely many $n\}$]
i.e, $X_n = 0$ a.s.

Also, <u>note</u>: $\sum X_n \geq 0$ is <u>not</u> a tail event.

$\mathcal{T}$ = the "tail" σ-field.

Any set $A \in \mathcal{T}$ is called a <u>"tail event"</u>.

Any r.v X measurable w.r.t $\mathcal{T}$ is a tail r.v.

<u>K's 0-1 law:</u>

If $\{X_n\}$ - independent seq. of r.vs, then for every tail event A, $P(A)$ is either 0 or 1.

<u>Proof</u>: <u>Step 1</u>: for every $n \geq 1$, $\mathcal{A}_n$ - independent of $\mathcal{T}_n$. [Exercise]

$\Downarrow$

<u>Step 2</u>: $\mathcal{A}_n$ is independent of $\mathcal{T}$ $\forall n$.

$\Downarrow$

<u>Step 3</u>: $\sigma\left(\bigcup_n \mathcal{A}_n\right)$ independent of $\mathcal{T}$.

$\Downarrow$

<u>Step 4</u>: $\mathcal{T}$ is independent of $\mathcal{T}$.

$\therefore$ any $A \in \mathcal{T}$.

$\therefore P(A \cap A) = P(A) \cdot P(A)$

$\mathcal{T} = \bigcap \mathcal{T}_n$

$\mathcal{T} := \bigcap_n \mathcal{T}_n$

& , any $\mathcal{T}_n$ is
a sub-σ field
of $\sigma(\bigcup_n a_n)$

$\therefore \bigcap_n \mathcal{T}_n = \mathcal{T} \subseteq \sigma(\bigcup_n a_n)$.

$\therefore P(A \cap A) = P(A) \cdot P(A)$
$\Rightarrow P(A) = (P(A))^2$
$\Rightarrow P(A) = 0$ or 1 $\boxtimes$

* X - tail r.v.
then $P(X \leq c)$ = either 0
or 1.
ie, X - degenerate r.v.

Jessen - Wintner.

Suppose $\{X_n\}$ is an independent seq. of r.vs,
each of which is discrete. s.t.
$\sum_n X_n$ converges a.s.

Then, the limit r.v. X is of "pure" type.
ie, either • X is discrete,
or • X is continuous (ie, distⁿ fⁿ continuous)
ie, no point mass,
but supported by a set of measure 0.

or • X is absolutely continuous
(ie, has a density fⁿ.)

Proof:

Let $D_n, n \geq 1$ be the countable set of possible values (ie, support) of X_n,
& let $D = \bigcup D_n$

& let $D = \bigcup_n D_n$

Let G be a <u>Subgroup</u> (!!!) of $\mathbb{R}$, generated by D.

$$G = \left\{ g : g = \sum_{i=1}^{n} k_i x_i \ : \ \begin{matrix} x_1, \ldots, x_n \in D. \\ k_1, \ldots, k_n \in \mathbb{Z}. \end{matrix} \right\}$$

G - countable

Observe:

for any Borel Set B, the set $\{X \in B + G\}$ is a tail set. (How !?)

$$\sum X_n \in B + G.$$

$\Leftrightarrow \sum X_n - b \in G$ for some $b \in B$.

$\Leftrightarrow \sum_{i=1}^{n} X_i + \sum_{i=n+1}^{\infty} X_i - b \in G.$

$\left[\begin{matrix} \text{take } k_i = 1 \\ g = \sum x_i \in G \end{matrix} \right]$

$\Leftrightarrow \sum_{i=n+1}^{\infty} X_i - b \in G.$ ✓

$\Rightarrow \sum_{i=n+1}^{\infty} X_i \in B + G.$

$\therefore \sum_{i=n+1}^{\infty} X_i$ - tail r.v. ✓.

$\therefore \forall$ borel set B,

$$P(X \in B + G) = 0 \text{ or } 1$$

[By K's 0-1 law]

Case 1: $\exists$ a countable set B,

Case 1: $\exists$ a countable set B,

s.t $P(X \in B + G) = 1$.

$\therefore X$ - discrete. ✓

Case 2: Otherwise,

$P(X \in B + G) = 0$ for every countable B.

$\therefore$ take $B = \{x\}$.

$\therefore P(X = x) \leq P(X \in \{x\} + G)$

$= 0$

$= P(X \in \{x\})$

$= P(B)$

$0 \in G$ (assumed)

$\Rightarrow X$ - continuous.

Case 2a: $\exists$ a Borel set B of $leb(B) = 0$

$\hookrightarrow$ lebesgue measure.

s.t, $P(X \in B + G) = 1$.

$$leb(B+G) = leb\left(\bigcup_{g \in G}(B+g)\right)$$

$$\leq \sum_{g \in G} leb(B+g)$$

each $= 0$.

$= 0$. $\left[\because G \text{ - ctbl}, \therefore \sum_{g \in G} \text{ is a countable sum.}\right]$

Case 2b: for every borel set B with $leb(B) = 0$.

$P(X \in B + G) = 0$.

$P(X \in B + G) = 0.$

? $\Rightarrow P(X \in B) = 0,$

? $\Rightarrow X$ - absolutely continuous.

Convergence in Distributions.

Def^n: Let $\{X_n\}_{n\geq 1}$ and X be real r.vs.

Let $F_n, n\geq 1$ and F be cdfs of $X_n, n\geq 1$ & X respectively

We say, $\{X_n\}$ converges in distribution to X, or

$$X_n \xrightarrow{d} X \quad \text{if} \quad F_n(x) \longrightarrow F(x) \ \forall \ x \in C_F$$

$C_F \equiv$ set of all continuity pts. of F.

If F is a cdf, then C_F denotes the set of continuity points of F.

Since $\mathbb{R}\setminus C_F$ is countable, $\therefore$ C_F is dense in $\mathbb{R}$.

Example to show why pointwise convergence is relaxed to only convergence at pts. in C_F.

$X_n \sim \text{Unif}\left(-\frac{1}{n}, \frac{1}{n}\right)$.

X_n has density $f_n(x) = \frac{n}{2}$, $-\frac{1}{n} < x < \frac{1}{n}$.

$$X_n \xrightarrow{d} X \equiv 0$$

$$F_n = \begin{cases} 0, & x \leq -\frac{1}{n} \\ \dfrac{x+\frac{1}{n}}{2n}, & -\frac{1}{n} < x < \frac{1}{n} \\ 1, & x \geq \frac{1}{n} \end{cases}$$

$\quad \llcorner \; 1 \;, \quad x \geq 1/n$

Let $n \to \infty$,

$x < 0$. $\quad F_n(x) = 0$ for all n large enough

$F_n(x) \to 0$

& , $x > 0$. $\quad F_n(x) = 1$ for all n large enough.

$F_n(x) \longrightarrow 1$

at $x = 0$, $\quad F_n(0) = \frac{1}{2} \; \forall n$.

So, $F = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0. \end{cases}$

This is the limit required for pointwise convergence.

$\therefore$ If we had "wanted" ptwise conv.,

then $X_n \not\longrightarrow X \equiv 0$

$\downarrow$

r.v degenerate at 0.

$\therefore$ Hence, we relax pt-wise convergence.

Remark:

(1) Conv. in distr does not require X_n's, X all to lie in the same probability space.

② $\left.\begin{array}{l} X_n \xrightarrow{d} X \\ X_n \xrightarrow{d} Y \end{array}\right\} \iff F_X = F_Y$, or $X \overset{d}{=} Y$ (ie, X, Y must have the same cdf.)

③ $\left.\begin{array}{l} X_n \xrightarrow{d} X \\ Y_n \xrightarrow{d} Y \end{array}\right\} \not\Longrightarrow X_n \pm Y_n \overset{d}{\not\longrightarrow} X \pm Y$

[counter eg: let $Z \sim N(0,1)$. take $X_n = Z$, $Y_n = -Z$
$\forall n \geq 1$ $\forall n \geq 1$.

$X_n \to Z$ $\quad$ $X_n + Y_n = 0$.

$Y_n \to -Z$ $\quad$ But, $(X_n + Y_n) \longrightarrow 2Z \neq 0$.

$\because Z \overset{d}{=} -Z$.]

Result:

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$

Proof: take any $x \in \mathbb{R}$.

$$\liminf_n F_n(x) \geq F(x^-)$$

$$\limsup_n F_n(x) \leq F(x^+).$$

Proving these two are enough, as, for $x \in C_F$, $F(x^+) = F(x^-)$.

Remark: this is <u>necessary & sufficient</u> (ie, equivalent) condition for convergence in condition for convergence in distribution.

$$F_n(x) = P(X_n \le x)$$

$$= P(X_n \le x, X \le x+\varepsilon) + P(X_n \le x, X \geq x+\varepsilon)$$

$$\le F(x+\varepsilon) + P(|X_n - X| > \varepsilon)$$

$\downarrow$ 0 (ie, limit exists, $\because X_n \xrightarrow{P} X$) $\therefore \limsup P(|X_n - X| > \varepsilon) = 0$

taking limsup,

$$\limsup_n F_n(x) \le F(x+\varepsilon)$$

letting $\varepsilon \searrow 0$,

$$\limsup_n F_n(x) \le F(x^+)$$

the other side:

$$P(X \le x-\varepsilon)$$

$$= P(X \le x-\varepsilon, X_n \le x) + P(X \le x-\varepsilon, X_n > x).$$

$$\le F_n(x) + P(|X_n - X| > \varepsilon)$$

$$F_n(x) \geq F(x-\varepsilon) - P(|X_n - X| > \varepsilon).$$

$\rightarrow \liminf P(|X_n - X| > \varepsilon) = 0$

$\therefore$ taking liminf.

$$\liminf_n F_n(x) \geq F(x-\varepsilon)$$

letting $\varepsilon \downarrow 0$

$$\therefore \liminf_n F_n(x) \geq F(x^-).$$

Conv. in dist^n $\equiv$ "Weak convergence".

$X_n \xrightarrow{a.s} X \underset{\not\Leftarrow}{\Rightarrow} X_n \xrightarrow{d} X$

$X_n \xrightarrow{P} X \underset{\not\Leftarrow}{\Rightarrow} X_n \xrightarrow{d} X$

Exercise: $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{P} X$

$Z \overset{d}{=} N(0,1)$.

Take $X_{2n} = Z$, $X_{2n-1} = -Z$. $\quad n = 1, 2, \dots$

$\therefore \quad X_{2n} \overset{d}{=} X_{2n-1} \overset{d}{=} Z \qquad Z \overset{d}{=} -Z$ (symm. around 0)

$\therefore X_n \xrightarrow{d} Z$.

But, $X_{2n} - X_{2n-1} = Z - (-Z)$
$= 2Z$.

$\therefore X_{2n} - X_{2n-1}$ is <u>not</u> even Cauchy !!

Result: $X_n \xrightarrow{d} X \equiv c$ (X - degenerate at c. ie, $X = c$ with prob $= 1$)

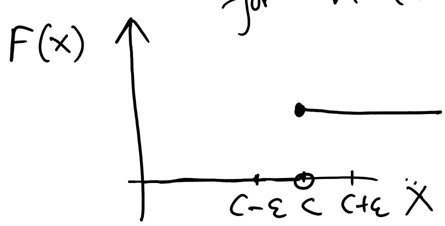
$\Rightarrow X_n \xrightarrow{P} X$

Proof: Fix $\varepsilon > 0$.

$$P(|X_n - X| > \varepsilon) = P(|X_n - c| > \varepsilon)$$
$$= P(X_n < c - \varepsilon) + 1 - P(X_n \leq c + \varepsilon)$$
$$\leq F_n(c-\varepsilon) + 1 - F_n(c+\varepsilon)$$
$$\to 0 + 1 - 1 = 0$$

$\left[\because F_n(c-\varepsilon) \longrightarrow F(c-\varepsilon) = 0 \right.$
$\left. F_n(c+\varepsilon) \longrightarrow F(c+\varepsilon) = 1 \text{ for } n \text{ large.} \right]$

$\therefore \boxed{X_n \xrightarrow{P} X}$.



$\blacksquare$

Slutsky "type" result:

(we saw just now, if limit r.v X is degenerate at some c, then, $X_n \xrightarrow{d} X \Leftrightarrow X_n \xrightarrow{p} X$)

$X_n \xrightarrow{d} X$, $Y_n \xrightarrow{p} 0$.

$\Rightarrow X_n + Y_n \xrightarrow{d} X$.

Proof: Let $X_n \sim F_n$. $n \geq 1$.

$X_n + Y_n \sim G_n$.

$X \sim F$.

Fix pts. $x, y \in C_F$

WLOG, $x < y$.

We'll prove:

$$(*) \begin{cases} \liminf G_n(y) \geq F(x) \\ \limsup G_n(x) \leq F(y). \end{cases}$$

We'll first accept this (*) & complete the proof.

We'll prove this later then. (*)

Take any $z \in C_F$.

We have to show,

$$\lim_n G_n(z) = F(z).$$

We'll consider 2 sequences,

one strictly increasing to z,

another strictly decreasing to z.

another strictly decreasing to z.

So, let $\{x_k\}_{k\geq 1}$ be a sequence in C_F.

$x_k < z \quad \forall k$

$x_k \nearrow z$

&, let $\{y_k\}_{k\geq 1}$ be a sequence in C_F,

$y_k < z \quad \forall k$

$y_k \searrow z$

By (*): $\liminf G_n(z) \geq F(x_k) \quad \forall k$

[for pair (x_k, y_k), $x_k \in C_F$, $y_k \in C_F$]

& $\limsup G_n(z) \leq F(y_k) \quad \forall k$

$\therefore$ By letting $k \nearrow \infty$, & $z \in C_F$,

we get,

$$F(z) \leq \liminf G_n(z) \leq \limsup G_n(z) \leq F(z)$$

$$\therefore \liminf_n G_n(z) = \limsup_n G_n(z) = \lim_n G_n(z) = F(z)$$

So, we're done !!!

(modulo this (*)).

Now, proof of (*):

$$G_n(x) = P(X_n + Y_n \leq x)$$

$$= P(X_n + Y_n \leq x,\ X_n \leq y) + P(X_n + Y_n \leq x,\ X_n > y)$$

$$\leq F_n(y) + \underbrace{P(|Y_n| > y - x)}_{\to\, 0}.$$

$y > x$,

$Y_n \xrightarrow{P} 0$.

taking limsup,

$$\limsup G_n(x) \le F(y)$$

Similarly, $\liminf G_n(y) \geq F(x)$. [Prove this: exercise].

$F_n, n \geq 1$ - cdfs of $X_n, n \geq 1$

F - cdf of X.

Let $S \subset \mathbb{R}$ be a dense set.

Suppose, (1) $F_n(x) \to F(x)$.

(2) $\forall x \in J \ (= C_F)$.

$$F_n(x) - F_n(x^-) \longrightarrow F(x) - F(x^-)$$

Then $\sup_{x} |F_n(x) - F(x)| \longrightarrow 0$

Corollary: Polya's Theorem.

If $X_n \xrightarrow{d} X$, and if X has a continuous cdf F,

then, $F_n \longrightarrow F$ uniformly on $\mathbb{R}$.

(From previous result, set $S = \mathbb{R}$. & check that, this thm. is proved.) ✓

*. Some "characterization" of convergence in distribution:

Step 1:

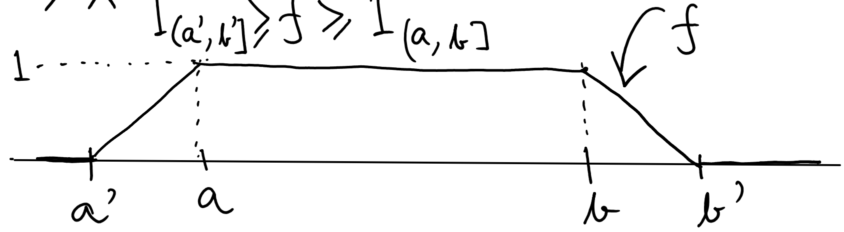
Suppose, $X_n, n \geq 1$ is a seq. of real r.vs and X - a real r.v.

Given,

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

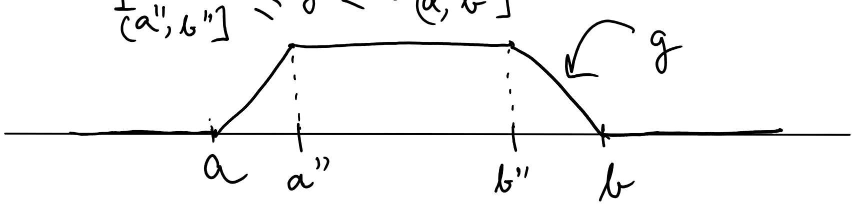
set of all bounded, continuous functions over $\mathbb{R}$.

We'll show: $X_n \xrightarrow{d} X$ $\quad \mathbb{1}_{(a',b']} \geq f \geq \mathbb{1}_{(a,b]}$



$f, g \in C_b(\mathbb{R})$

$\mathbb{1}_{(a'',b'']} \leq g \leq \mathbb{1}_{(a,b]}$



$$\therefore E(f(X)) = \lim_n E(f(X_n))$$

Using f, $\Rightarrow F(b) - F(a) \leq \lim_n E(f(X_n))$

$$\leq \liminf_n \left(F_n(b') - F_n(a')\right)$$

Using g,

$$\limsup_n \left(F_n(b'') - F_n(a'')\right) \leq F(b) - F(a).$$

Fix $a, b \in C_F$.

Using this, show that,

$$\boxed{\lim_n F_n(b) - F_n(a) = F(b) - F(a)}$$

[Hint / sketch: C_F is dense. So, we can take $a', a'' \in C_F$ as close to 'a' we wish to, & same for $b', b'' \in C_F$.]

Now, take $a \to -\infty$.

Now, take $a \searrow -\infty$.

$$\therefore \lim_n F_n(b) = F(b).$$

$$\therefore X_n \xrightarrow{d} X.$$

Step-2: (the other side)

Given $X_n \xrightarrow{d} X$

We'll show, $E f(X_n) \longrightarrow E f(X)$. for all $f \in C_b(\mathbb{R})$

Proof: WLOG, we assume, $|f(x)| \le 1 \;\; \forall x \in \mathbb{R}$.

Let $\varepsilon > 0$ be given.

Pick $M > 0$ with $M, -M \in C_F$ s.t.

$$F(-M) + (1 - F(M)) < \varepsilon.$$

ie, we want $-M$ very large $-$ve, & M very large $+$ve s.t. the probability mass of X beyond M & $-M$ adds up to $< \varepsilon$.

Realize this: it is enough to show,

$$E\left(f(X_n) \cdot \mathbb{1}_{X_n \in (-M, M]}\right) - E\left(f(X) \cdot \mathbb{1}_{X \in (-M, M]}\right) \longrightarrow 0$$

call $J = (-M, M]$.

$$\therefore \left| E(f(X)) \cdot \mathbb{1}_{X \notin J} \right| \le E\Big(\underbrace{|f(x)|}_{\substack{<1 \\ \text{(assumed)}}} \cdot \mathbb{1}_{X \notin J}\Big)$$

$$\le E\left(\mathbb{1}_{X \notin J}\right)$$

$$= P(X \notin J)$$

&,

$$\left| E\left(f(X_n) \cdot 1_{X \notin J}\right)\right| \leq E\left(\underbrace{|f(X_n)|}_{\substack{<1 \\ \text{(assumed)}}} \cdot 1_{X \notin J}\right)$$

$$\leq E\left(1_{X_n \notin J}\right)$$

$$= P(X_n \notin J)$$

$$= P(X_n \leq -M) + 1 - P(X_n < M)$$

$$= F_n(-M) + 1 - F_n(M)$$

$$\downarrow \qquad\qquad \downarrow$$

$$F(-M) \qquad F(M) \quad \left[\because M \in C_F \text{ \& } -M \in C_F\right]$$

$$\to F(-M) + 1 - F(M)$$

$$= P(X \notin J) < \varepsilon' \quad \forall\, n \geq N_1$$

Choose $\delta > 0$ s.t. $x, y \in [-M, M]$
& $|x-y| < \delta$
$\Rightarrow |f(x) - f(y)| < \varepsilon'$.

[using uniform continuity of f on compact set $[-M, M]$.]

Divide $[-M, M]$ into finitely many intervals, each of length $\delta/2$.

Let the partition points be called:

$$-M = y_0 < y_1 < y_2 < \cdots < y_k < y_{k+1} = M.$$

Now, between y_0 & y_1,

Now, between y_0 & y_1,
i.e., in $[y_0, y_1]$,
pick $x_1 \in [y_0, y_1] = [-M, y_1]$
s.t. $x_1 \in C_F$. [We can do that, $\because C_F$ is dense.]

Similarly, pick
$x_2 \in (y_1, y_2)$,
$x_3 \in (y_2, y_3)$
$\vdots$ each $x_i \in C_F$.
$x_{k+1} \in (y_k, y_{k+1}) = (y, M)$,

$\therefore$ Due to this choices of x_i's,
$|x_{i+1} - x_i| < \delta/2$

Now, define $g(x) := \sum_{i=1}^{k+1} f(x_i) \cdot \mathbb{1}_{(x_i, x_{i+1}]}(x)$

function on $[-M, M]$, $\underline{\text{not}}$ cont, agrees with f on every "left" point.

Convince yourself that — $|f(x) - g(x)| < \varepsilon \quad \forall x \in J = [-M, M]$

$$\left| E\left(f(X_n) \cdot \mathbb{1}_{X_n \in J}\right) - E\left(g(X_n) \cdot \mathbb{1}_{X_n \in J}\right) \right| < \varepsilon',$$

&,

$$\left| E\left(f(X)\cdot \mathbb{1}_{X\in J}\right) - E\left(g(X)\cdot \mathbb{1}_{X\in J}\right)\right| < \varepsilon',$$

$$\leq \left| E\left(g(X_n)\cdot \mathbb{1}_{X_n\in J}\right) - E\left(\left(g(X)\cdot \mathbb{1}_{X\in J}\right)\right)\right| + 2\varepsilon'$$

[adding & subtracting terms. Δ-ineq.]

$$= \left| \sum_{i=1}^{k+1} f(x_i)\cdot\left(F_n(x_{i+1}) - F_n(x_i)\right) - \sum_{i=1}^{k+1} f(x_i)\cdot\left(F(x_{i+1}) - F(x_i)\right)\right| + 2\varepsilon'$$

$$= \left| \sum_{i=1}^{k+1} f(x_i)\left[\left(F_n(x_{i+1}) - F_n(x_i)\right) - \left(F(x_{i+1}) - F(x_i)\right)\right]\right| + 2\varepsilon'$$

(How? think!!) $< \varepsilon' + 2\varepsilon'$ $\forall$ $n \geq N_2$ large enough.

$= 3\varepsilon'$

$\blacksquare$

Probability-3 Lecture-17

04 October 2024 11:19

Definition:

$X_n \xrightarrow{d} X$ (also denoted by $X_n \xrightarrow{w} X$ ← weak convergence)

if $F_n(x) \longrightarrow F(x)$ $\forall\, x \in C_F$ ← (set of all continuity points of F.)

recall, **Theorem:**

$X_n \xrightarrow{d} X$ iff $E(f(X_n)) \longrightarrow E f(X)$, $\forall\, f \in C_B(\mathbb{R})$ → [set of all bounded, continuous fns.]

for complete proof, refer prev. Lec. (L-16)

Corollary: (Continuous Mapping Theorem)

$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$ $\forall\, g \in C(\mathbb{R})$ → set of all continuous fns over $\mathbb{R}$.

to show: $E(f(g(X_n))) \longrightarrow E(f(g(X)))$

If $f \in C_b(\mathbb{R})$,
& $g \in C(\mathbb{R})$,
then $f \circ g \in C_b(\mathbb{R})$ &, $X_n \xrightarrow{d} X$
then, just simply apply the above characterization of conv. of distribution.

$\therefore E(f(g(X_n))) \longrightarrow E(f(g(X)))$.

$\Leftrightarrow f(g(X_n)) \xrightarrow{d} f(g(X))$.

Corollary:

Suppose K is a compact interval.
If $X_n, n \geq 1$ and X are r.vs, all taking values in K, then

$$X_n \xrightarrow{d} X \iff E(X_n^j) \longrightarrow E(X^j) \quad \forall j = 1, 2, \ldots$$

($\Rightarrow$) Here, $f(x) = x^j$

On a compact interval, this is bounded.

($\Leftarrow$) for any polynomial p,
$$E(p(X_n)) \longrightarrow E(p(X))$$

Exercise: This thm. can be made "stronger".

Show that: 1 $X_n \xrightarrow{d} X$ iff. $E(f(X_n)) \longrightarrow E(f(X))$ $\forall f \in C_K(\mathbb{R})$.

Set of all continuous f^ns with compact support.

2 &, $X_n \xrightarrow{d} X$ iff. $E(f(X_n)) \longrightarrow E(f(X))$ $\forall f \in C_\infty(\mathbb{R})$,

infinitely diff^ble f^ns.

Probability-3 Lecture-18

15 October 2024 14:19

Recall - result:

$$X_n \xrightarrow{d} X \quad \text{iff} \quad E(f(X_n)) \longrightarrow E(f(X)) \ \forall f \in C_b(\mathbb{R})$$

$C_b(\mathbb{R})$ → bounded real f^ns.

This is a characterization of convergence of distributions.

Some standard notations:

- $X_n \xrightarrow{d} F$ → a cdf

  this means, X_n converges to some r.v whose CDF is F.

- $F_n \xrightarrow{d} F$

  → F_n - cdf of X_n. $\xrightarrow{d}$ cdf F (of some r.v).

Complex Random Variables.

Z is a complex valued r.v on $(\Omega, \mathcal{A}, P)$ if $Z = X + iY$, where X, Y are real valued r.v.s

Define expectation,

$$E(Z) := E(X) + i \cdot E(Y), \text{ if}$$

$E(X)$ & $E(Y)$ are both finite.

Properties:

(a) Z_1, Z_2 - complex r.vs.

α_1, α_2 - complex nos.

$E(Z_1)$ & $E(Z_2)$ both exist.

then, $\Rightarrow E(\alpha_1 Z_1 + \alpha_2 Z_2) = \alpha_1 \cdot E(Z_1) + \alpha_2 \cdot E(Z_2)$

[Easy to prove.
take $\alpha_1 = u_1 + i \cdot v_1$, $\alpha_2 = u_2 + i v_2$
$Z_1 = X_1 + i \cdot Y_1$; $Z_2 = X_2 + i \cdot Y_2$
& do some algebra.]

(b) For any complex r.v Z,

$$\underbrace{|E(Z)|}_{\text{a complex no.}} \leq E(|Z|).$$

⚠ looks analogous to $|EX| \leq E(|X|)$, but it isn't !!

Proof of (b):

For every complex w, $\exists$ a complex no. α with $|\alpha| = 1$

s.t. $|w| = \alpha . w$.

if $w = 0$, take any α.

[Easy to check]

$\therefore$ Applying with $w = E(Z)$,

$\exists$ $\alpha \in \mathbb{C}$ with $|\alpha| = 1$.

s.t, $\underbrace{|E(Z)|}_{\text{Real}} = \underbrace{\alpha . E(Z)}_{\text{complex}}$.

$$\Rightarrow |E(Z)| = |Re(\alpha E(Z))|$$

Easy to check !!

$$= |Re \cdot E(\alpha Z)|$$

$$= |E(Re(\alpha Z))|$$

$$\leq E\,\underbrace{|Re(\alpha Z)|}_{\leq}.$$

$$\leq E|\alpha Z|$$

$$= E|Z| \quad [\because |\alpha| = 1]$$

Suppose, $Z = X + iY$.

then, $E(Z) := E(X) + iE(Y)$.

$$\therefore |E(Z)| = \sqrt{(E(X))^2 + (E(Y))^2},$$

whereas,

$$E|Z| = E\left(\sqrt{X^2 + Y^2}\right)$$

$\therefore$ for real r.vs X, Y,

we're saying:

$$\sqrt{(E(X))^2 + (E(Y))^2} \leq E\left(\sqrt{X^2+Y^2}\right)$$

This is <u>not</u> a L_1 ineq.

This is stronger.

Consequence of (b):

DCT holds for complex r.vs as well !!

Z_n, $n \geq 1$ is a sequence of complex r.vs &

Z is a complex r.v,

Z is a complex r.v,

and if $Z_n \xrightarrow{a.s} Z$, & if

$\exists Y$ with $E(Y) < \infty$ s.t,

$|Z_n| \le |Y| \quad \forall n$,

then $E|Z_n - Z| \longrightarrow 0$. [the "real-DCT"]

in particular,

$EZ_n \longrightarrow EZ$.

[using part (b)].

Definition:

Complex r.vs

$$Z_1 = X_1 + i \cdot Y_1$$
$$\vdots$$
$$Z_n = X_n + i \cdot Y_n.$$

are said to be independent if

$(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent random vectors.

Consequence:

Z_1, Z_2 - complex independent r.vs,

then $E(Z_1 \cdot Z_2) = E(Z_1) \cdot E(Z_2)$.

Proof: Write $Z_1 = X_1 + i \cdot Y_1$

$Z_2 = X_2 + i \cdot Y_2$

, just do it !!

just do it !!

Definition: Characteristic Function. (C.F.)

For a real r.v X, its characteristic f^n φ_X is defined by

$$\varphi_X(t) = E\left(e^{itX}\right), \quad \varphi_X : \mathbb{R} \longrightarrow \mathbb{C}.$$

$$\varphi_X(t) = E(\cos tX) + i\,E(\sin tX).$$

* C.F is a function $\varphi_X : \mathbb{R} \longrightarrow \mathbb{C}$,

$\varphi_X(t)$ exists $\forall\ t \in \mathbb{R}$. [$\because$ $E(\cos tX)$ & $E(\sin tX)$ are always bounded within $[-1,1]$]

[Unlike MGF.]

Properties:

(a) $\varphi_X(0) = 1$

(b) $|\varphi_X(t)| \leq 1$.

Proof:

$$|\varphi_X(t)| = \left|E\left(e^{itX}\right)\right| \leq E\left|e^{itX}\right| = 1.$$

(c) φ_X is a uniformly continuous f^n on $\mathbb{R} \longrightarrow \mathbb{C}$.

Proof:

$$|\varphi_X(t+h) - \varphi_X(t)| = |E(e^{i(t+h)X}) - E(e^{itX})|$$

$$= |E(e^{itX}(e^{ihX}-1))|$$

$$\leq E|e^{itX}\cdot(e^{ihX}-1)|$$

$[\because E|e^{itX}|=1]$

$$= E|e^{ihX}-1| \xrightarrow[(DCT)]{} 0$$

$\because |e^{ihX}-1| \to 0,$

& $|e^{ihX}-1| \leq \underset{\substack{\uparrow \\ \text{constant r.v.}}}{2} < \infty.$

$\boxtimes$

(d) $\varphi_{a+bX}(t) = e^{iat}\cdot\varphi_X(bt)$

Proof: easy to check.

$$\varphi_{a+bX}(t) = E\left(e^{it(a+bX)}\right)$$

$$= E\left(e^{iat}\cdot e^{i(bt)\cdot X}\right) = e^{iat}\cdot E\left(e^{i(bt)X}\right)$$

$$= e^{iat}\cdot\varphi_X(bt).$$

(e) $\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$

* if 2 r.vs have same CDF, they have same C.F.

(f) If $X \stackrel{d}{=} -X$ [ie, X has symmetric dist^n eg. $X = N(0,1)$]

then $\varphi_X(t) = \varphi_{-X}(t)$

$$\varphi_X(t) = \overline{\varphi_X(t)}$$

$\therefore \varphi_X(t)$ is real-valued.

Is the converse true?

ie if $\varphi_X(t)$ is real-valued, can we say: X has symmetric dist^n?

Not until we prove that C.F. uniquely determines the distribution.

(g) X, Y - independent.

$$\varphi_{X+Y}(t) = \varphi_X(t) + \varphi_Y(t).$$

---

* For a discrete distribution with values $x_1, x_2, \ldots$ & pmf $p(x_1), p(x_2), \ldots$

$$\varphi_X(t) = \sum_i e^{itx_i} \cdot p(x_i)$$

For a random variables with density f,

$$\varphi_X(t) = \int e^{itx} \cdot f(x)\, dx$$

Some examples.

(a) $X \sim \text{Bin}(n, p)$. Write $q = 1 - p$.

$$\varphi_X(t) = \sum_{k=0}^{n} \binom{n}{k} e^{itk} \cdot p^k \cdot q^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \cdot (pe^{it})^k \cdot q^{n-k}$$

$$= (pe^{it} + q)^n$$

(b) $X \sim \text{Poi}(\lambda)$.

$$\varphi_X(t) = e^{-\lambda} \cdot \sum_{k=0}^{\infty} e^{itk} \cdot \lambda^k / \lfloor k$$

$$= e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{\lfloor k} = e^{-\lambda} \cdot e^{\lambda e^{it}}$$

$$= e^{-\lambda(1 - e^{it})}$$

(c) $X \sim \exp(\lambda)$.

$$\varphi_X(t) = \lambda \cdot \int_0^{\infty} e^{itX} \cdot e^{-\lambda x} \cdot dx$$

$$= \lambda \cdot \int_0^{\infty} e^{-(\lambda - it)} dx \overset{?}{\underset{\text{check}}{=}} \left(\frac{\lambda}{\lambda - it} \right)$$

quick thinking: assume $\lambda - it = \theta$.

$\therefore \int_0^{\infty} e^{-\theta} = \frac{1}{\theta}$.

that shortcut works!!

0 $\quad$ $)$

exc: work out: $= \left(\lambda \int \cos tx \cdot e^{-\lambda x} dx + i\lambda \int \sin tx \cdot e^{-\lambda x} \cdot dx \right)$

(d) $X \sim N(0,1)$.
$\searrow$ symmetric.

$\therefore \varphi_X(t)$ should be real-valued.

$$\varphi_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot \cos tx \cdot e^{-x^2/2} dx$$

$$= \frac{2}{\sqrt{2\pi}} \cdot \int_0^{\infty} \cos tx \cdot e^{-x^2/2} dx.$$

$$\varphi_X'(t) = \frac{-2}{\sqrt{2\pi}} \cdot \int_0^{\infty} \sin tx \cdot x e^{-x^2/2} \cdot dx \qquad \left[* \int x e^{-x^2/2} dx = = e^{-x^2/2} \right]$$

Int. by parts.

$$= \frac{-2}{\sqrt{2\pi}} \cdot \left[\left(\sin tx \left(-e^{-x^2/2}\right) \right) \Big|_{x=0}^{x=\infty} + t \int_0^{\infty} \cos tx \cdot e^{-x^2/2} dx \right]$$

$\therefore \varphi_X'(t) = -t \cdot \varphi_X(t)$, $\varphi_X(0) = 1$.

$\searrow$ a simple differential eq^n.

Sol^n: $\boxed{\varphi_X(t) = e^{-t^2/2}.}$

f density of $N(0,1)$

$$\hat{f}(t) = \sqrt{2\pi}\,.\, f(t).$$

$\therefore$ for $X \sim N(\mu, \sigma^2)$

$$\varphi_X(t) = e^{i\mu t}\,.\, e^{-\sigma^2 t^2/2}.$$

- - - - - - - - - - - - - - - - -

Characteristic f^n $\xleftrightarrow[\text{correspondance}]{1-1}$ distribution.

Inversion formula:

Let φ_X be the characteristic function of X & F_X be the CDF.

then, for any $a < b$ in $\mathbb{R}$,

$$P(a < X < b) + \frac{1}{2}\Big(P(X=a) + P(X=b)\Big) = \underbrace{\frac{1}{2\pi}\cdot \lim_{T\to\infty} \int_{-\infty}^{\infty} \frac{e^{ita} - e^{itb}}{it}\cdot \varphi_X(t)\, dt.}$$

In particular, for $a, b \in C(F_X)$, $F(b) - F(a) =$

$C(F_X)$: set of all continuity pts. of cdf of X.

* fact to recall from Sem - 2:

Suppose, let X, Y - independent r.vs.

Y has a density f.

Y has a density f.

then, $P(X+Y \leq a) = E(g(X))$, where $g(x) = P(Y \leq a-X)$

$$= \int_{-\infty}^{a-X} f(y)\,dy. \qquad = \int_{-\infty}^{a-x} f(y)\,dy$$

change of variable:

call $v = y + X$.

v	y
$-\infty$	$-\infty$
a	$a-X$

$$= E\left(\int_{-\infty}^{a} f(v-X)\cdot dv \right)$$

$$= \int_{-\infty}^{a} E\big(f(v-X)\big)\,dv$$

$\therefore$ X, Y - ind. rvs, Y has density f. $\Rightarrow$

$\boxed{X+Y \text{ has a density } g \text{ given by } g(v) = E\big(f(v-X)\big).}$

Now, Feller's way of proving:

C.f $\xleftrightarrow{}$ CDF
1-1 correspondance.

Let X have a c.f φ_X.

$$\frac{1}{\sqrt{2\pi n}} \int_{\mathbb{R}} e^{-itv} \varphi_X(v)\, e^{-v^2/2n}.\,dv \qquad , \ t \in \mathbb{R}. \text{ some real.}$$

$$= \frac{1}{\sqrt{2\pi n}} \cdot \int_{\mathbb{R}} e^{-itv} \cdot E\left(e^{itv}\right) \cdot e^{-v^2/2n} \, dv$$

$$= \frac{1}{\sqrt{2\pi n}} \cdot \int_{\mathbb{R}} E\left(e^{-iv(t-X)}\right) \cdot e^{-v^2/2n} \cdot dv$$

$$= E\left(\frac{1}{\sqrt{2\pi n}} \cdot \int_{\mathbb{R}} e^{-iv(t-X)} \cdot e^{-v^2/2n} \, dv\right)$$

$\to$ C.F of $N(0,1)$ evaluated at $(t-X)$

$\sigma^2 = n$.

$$= E\left(e^{-\frac{n}{2}(t-X)^2}\right)$$

$$= E\left(\frac{\sqrt{n}}{\sqrt{2} \cdot \sqrt{n}} \cdot e^{-n/2 (t-X)^2}\right)$$

$|||$

density of $X + \frac{1}{\sqrt{n}} \cdot Z$,

$Z \sim N(0,1)$

$\therefore$ LHS $\to$ a crazy integral using C.F. of X

RHS $\to$ density of $X + \frac{1}{\sqrt{n}} \cdot Z$.

So now, if

$$\varphi_X = \varphi_Y ,$$

then,

$$\Rightarrow X + \frac{1}{\sqrt{n}} Z \overset{d}{=} Y + \frac{1}{\sqrt{n}} Z .$$

Now, as $n \longrightarrow \infty$, $\frac{1}{\sqrt{n}} \longrightarrow 0$.

$\therefore$ $\therefore$ By slutsky, $\frac{1}{\sqrt{n}} \cdot Z \xrightarrow{d} 0$

$$\therefore \boxed{X \overset{d}{=} Y .} .$$

$\blacksquare$

Probability-3 Lecture-19

18 October 2024 11:14

(Recall):

Definition:

for a real r.v, its characteristic f^n φ_X is defined as

$$\varphi_X(t) = E\left(e^{itX}\right), \ t \in \mathbb{R}$$

$$= E(\cos tX + i \sin tX)$$

$$= E(\cos tX) + i\, E(\sin tX)$$

$$\varphi_X : \mathbb{R} \to \mathbb{C}$$

We showed, $X \stackrel{d}{=} Y \iff \varphi_X \equiv \varphi_Y$.

Properties:

1. $\varphi_X(0) = 1$
2. $|\varphi_X(t)| \leq 1$ $\left[\because |EZ| \leq E|Z|\right]$
3. $\varphi_{a+bX}(t) = e^{iat} \cdot \varphi_X(bt)$.
4. $\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$
5. If X has a symmetric distribution, $X \stackrel{d}{=} -X$.

   then, $\varphi_{-X}(t) = \varphi_X(t)$

   $\Rightarrow \overline{\varphi_X(t)} = \varphi_X(t)$

   $\Rightarrow \varphi_X(t)$ is real-valued.

6. $\varphi_X(t)$ is uniformly continuous.

Proof: $\quad |\varphi_X(t+y) - \varphi_X(t)| \leq E\left|e^{itX} - 1\right| \xrightarrow[DCT]{} 0$

(b) $\varphi_X(t)$ is uniformly continuous.

Proof: (detailed proof in last lecture) $\quad |\varphi_X(t+y) - \varphi_X(t)| \leq E\,|e^{itX} - 1| \xrightarrow[\text{DCT}]{} 0$

Some examples:

Eg 1. $X \sim Bin(n, p)$

$$\varphi_X(t) = (pe^{it} + q)^n \quad , \; q = 1-p$$

Eg. 2. $X \sim Poi(\lambda)$

$$\varphi_X(t) = e^{-\lambda(1-e^{it})}$$

Eg. 3. $X \sim Exp(\lambda)$.

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}$$

Eg. 4. $Z \sim N(0,1)$

$$\varphi_Z(t) = e^{-t^2/2}$$

if $X \sim N(\mu, \sigma^2)$, $\quad \varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$

In particular, if $Y \sim N(0, \sigma^2)$

$$\varphi_Y(t) = e^{-\frac{1}{2}\sigma^2 t^2}$$

Recall (a result from last lecture)

for independent random variables X, Y, we use the notation $X * Y$ for the random variable $(X+Y)$.

If Y has density function f, then $X * Y$ has density $E(f(v - X))$.

Suppose,

X has characteristic f^n $\varphi_X(t)$.

... with the integral

X has characteristic f^n $\varphi_X(t)$.

We start with the integral

$$\frac{1}{\sqrt{2\pi n}} \cdot \int_{\mathbb{R}} \varphi_X(v) \cdot e^{-itv} \cdot e^{-v^2/2n} \cdot dv$$

$$= \frac{1}{\sqrt{2\pi n}} \int_{\mathbb{R}} E\left(e^{ivx}\right) \cdot e^{-itv} \cdot e^{-v^2/2n}\, dv$$

$$= E\left(\frac{1}{\sqrt{2\pi n}} \cdot \int_{\mathbb{R}} e^{-iv(t-x)} \cdot e^{-v^2/2n}\, dv\right)$$

Note that: $\frac{1}{\sqrt{2\pi n}} \cdot e^{-v^2/2n} \rightarrow$ Density of $N(0, n)$

$\therefore$ this is nothing but, the characteristic f^n of $N(0,n)$ evaluated at $-(t-x)$.

$$= E\left(e^{-\frac{1}{2}(t-x)^2 \cdot n}\right)$$

$\therefore$ Multiply both sides by $\frac{\sqrt{n}}{\sqrt{2\pi}}$,

$$\frac{1}{2\pi} \cdot \int_{\mathbb{R}} \varphi_X(v) \cdot e^{-itv} \cdot e^{-v^2/2n}\, du = \frac{\sqrt{n}}{\sqrt{2\pi}} \cdot E\left(e^{-\frac{1}{2} \cdot (t-x)^2 \cdot n}\right)$$

$$= E\Big(\underbrace{\frac{\sqrt{n}}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \cdot (t-x)^2 / (1/\sqrt{n})^2}}_{f(t-x)}\Big)$$

$f(t-x)$, where

$f(t-x)$, where f is the density of $N(0, 1/n)$

$= E(f(t-X))$

$\hookrightarrow$ density of $X_n = X * N(0, \frac{1}{n})$

All of these imply,

if $\varphi_X = \varphi_Y$, then

$$\begin{array}{ccc} X * N(0, \frac{1}{n}) & = & Y * N(0, \frac{1}{n}) \\ \Big\downarrow d & \text{(by Slutsky)} & \Big\downarrow d \\ X & & Y \end{array}$$

$\therefore X \overset{d}{=} Y$. ✓

ie, that's Feller's way of showing, that characteristic function uniquely determines a distribution.

Eg. C.F of Binomial (n, p)

$X \sim Bin(n, p)$. $\qquad q = 1-p$

$\varphi_X(t) = (pe^{it} + q)^n$

$\varphi_X(0) = 1$.

$\varphi_X(2\pi) = (pe^{2i\pi} + q)^n$

$= (p+q)^n = 1$.

infact, $\varphi_X(t) = 1$ for any $t = 2n\pi$, $n \in \mathbb{Z}$.

Eg: $X \sim Poi(\lambda)$

$$\varphi_X(t) = e^{-\lambda(1-e^{it})}$$

Again, $t = 2n\pi$, $n \in \mathbb{Z}$,

$$\varphi_X(t) = \varphi_X(2n\pi) = e^{-\lambda(1-e^{i2n\pi})}$$
$$= e^{-\lambda(1-1)} = e^0 = 1.$$

(Again!!)

Now,

take $X \sim Exp(\lambda)$.

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}$$

$\varphi_X(t) = 1$ only at $t = 0$,

& $\forall\, t \neq 0$, $|\varphi_X(t)| < 1$

} - same conclusion in the case that $X \sim N(\mu, \sigma^2)$

Eg. take this discrete case:

$X = 0, \pm\sqrt{2}, \pm 1$; each with $p = 1/5$.

$$\therefore \varphi_X(t) = \frac{1}{5}\left(1 + 2\cos\sqrt{2}t + 2\cos t\right).$$

check that, Here too,

$\varphi_X(t) = 1$ only if $t = 0$,

& $\forall\, t \neq 0$, $|\varphi_X(t)| < 1$.

* Such distributions, like Bin(n,p), or Exp(λ) are called Lattice distributions, or Arithmetic distributions (as used by Feller)

Def^n: A real r.v X is said to be **Lattice distribution**, or **Arithmetic distribution** if:

X is discrete, and

its countable support is <u>contained in</u>.

$$S = \{ a + jd : j \in \mathbb{Z} \},$$

for some real a, d.

* **Result:**

Let X be a real r.v. with characteristic f^n - φ_X, then $\exists\, t_0 \neq 0$, s.t. $|\varphi_X(t_0)| = 1$ **iff**.

X has a Lattice distribution with its support contained in

$$S = \left\{ a_0 + j \cdot \frac{2\pi}{t_0} \;\middle|\; j \in \mathbb{Z} \right\}$$

further, In this case -

$|\varphi_X(t)|$ is periodic with period $= t_0$.

Proof: <u>"if" part:</u>

Let support of X be contained in S.

denote $p_j = P\left(X = a_0 + j \cdot \frac{2\pi}{t_0}\right)$

$$\varphi_X(t_0) = E\left(e^{it_0 X}\right)$$

$$= \sum_j e^{it_0(a_0 + j \cdot 2\pi/t_0)} \cdot p_j$$

corresponding probability

[we can have $p_j = 0$ for many j's. Eg. for Bin (n, p), $p_j = 0 \;\forall j > n$]

$$= e^{ia_0 t_0} \sum_j \underbrace{e^{ij \cdot 2\pi}}_{= 1} \cdot p_j$$

$$= e^{ia_0 t_0} \qquad \left[\because \sum p_j = 1\right]$$

$$= e^{ia_0 t_0} \cdot 1. \qquad \left[\because \sum p_j = 1\right]$$

$$\therefore |\varphi_X(t_0)| = 1 \qquad \blacksquare$$

"only if" part:

WLOG, take $t_0 > 0$. $\left[\because \text{ if } t_0 < 0, \ -t_0 > 0, \ \& \ |\varphi_X(t_0)| = 1 \iff |\overline{\varphi_X(t_0)}| = 1\right]$

why can we take ?

$$|\varphi_X(t_0)| = 1$$

$$\Rightarrow \varphi_X(t_0) = e^{i\theta_0}$$

$$\Rightarrow E\left(e^{it_0 X}\right) = e^{i\theta_0}$$

$$\Rightarrow E\left(e^{it_0 X - i\theta_0}\right) = 1$$

$$\Rightarrow E\big(\underbrace{e^{it_0(X - \theta_0/t_0)}}_{\text{complex r.v}}\big) = 1.$$

Real expected value

$$\therefore E(\text{Real part}) = 1, \qquad E(\text{imaginary part}) = 0$$

$$\Rightarrow E\left(\cos t_0\left(X - \frac{\theta_0}{t_0}\right)\right) = 1. \qquad \text{ie, X must take values - multiples of } 2\pi \text{ with prob} = 1.$$

2nd sem endsem paper

$$\Rightarrow P\left(X \in \left\{\frac{\theta}{t_0} + \frac{2\pi j}{t_0} : j \in \mathbb{Z}\right\}\right) = 1. \qquad \blacksquare$$

Now, take $\varphi_X(t+t_0)$

$$= E\left(e^{i(t+t_0)X}\right)$$

$$= \sum_j e^{i(t+t_0)\cdot\left(\frac{\theta_0}{t_0}+\frac{2\pi j}{t_0}\right)} \cdot p_j$$

$$= \sum_j e^{it\left(\frac{\theta_0}{t_0}+\frac{2\pi j}{t_0}\right)} \cdot \underbrace{e^{it_0\left(\frac{\theta_0}{t_0}+\frac{2\pi j}{t_0}\right)}}_{e^{i\theta_0}\cdot \underbrace{e^{ij2\pi}}_{\overset{\shortparallel}{1}}} \cdot p_j$$

$$= e^{i\theta_0}\cdot\sum_j e^{it\left(\theta_0/t_0+2\pi j/t_0\right)} \cdot p_j$$

$$= \underbrace{e^{i\theta_0}}_{1\cdot 1=1}\cdot\varphi_X(t). \qquad \therefore |\varphi_X(t+t_0)| = |\varphi_X(t)|.$$

ie, for every t_0,

$|\varphi_X(t)|$ is periodic with period $= t_0$.

$\boxtimes$

Theorem:

Let φ be a characteristic f^n.

Then,

either: (i) $|\varphi(t)| < 1 \quad \forall\, t \neq 0.$

<u>or</u> (ii) $|\varphi(t)| \equiv 1$

<u>or</u> (iii) $|\varphi(t)| = 1$ for a countable set of isolated points.

Proof: We'll simply prove, if there is a

Proof: We'll simply prove, if there is a sequence $\{t_n\}$ of distinct reals, with $t_n \longrightarrow s$ real..
s.t, $|\varphi(t_n)| = 1 \ \forall n$,
then $|\varphi(t)| \equiv 1$.

[Proving this works. Think why !!]

if such t's are uncountable,
then $\exists$ a compact set in which there are infinitely many such t's, which imply $|\varphi(t)| \equiv 1$

[$\because$ By Bolzano Weierstrass, $\exists$ a seq that converges.]

else, if t's are countable. there are no limit pts. Hence, the t's are isolated.

Observe: for any integer k, and any $m \neq n$,

$$|\varphi(k(t_n - t_m))| = |\varphi(kt_m)| = 1.$$

Take any integer in (a,b).

$\exists\, n \neq m$ s.t., $0 < (t_n - t_m) < b - a$

$\Rightarrow k$ - integer s.t. $k(t_n - t_m) \in (a,b)$

$\therefore$ Every interval contains a t s.t $|\varphi(t)| = 1$

So, the set of such t's is dense.

But, φ is continuous.

But, φ is continuous,
So, that set is closed.
$\therefore$ that set of t's, where $|\varphi(t)|=1$ has to be the entire real line $\mathbb{R}$.

$\blacksquare$

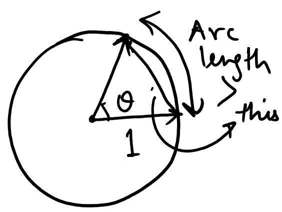
[Exercise: if $t_1, t_2 \in \mathbb{R}$ s.t $|\varphi(t_1)| = 1 = |\varphi(t_2)|$, (WLOG, $t_1 > t_2$) &, $t_1/t_2 \in \mathbb{Q}^c$ (irrational), then show that: $|\varphi(t)| \equiv 1$.]

Inversion formula.

If φ_X is a characteristic function of X, then for any two pts. $a<b$ in the "continuity of X",

$$F_X(b) - F_X(a) = \frac{1}{2\pi} \cdot \lim_{T\to\infty} \underbrace{\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \cdot \varphi_X(t)\, dt}_{\text{is it integrable? } \underline{\text{YES}}.}$$

$\because |e^{i\theta}-1| \le |\theta|$



$\therefore |e^{-ita} - e^{-itb}|$

$$\therefore \left| \frac{e^{-ita} - e^{-itb}}{it} \right| \leq |b-a|.$$

$$\left| \frac{e^{-it(a-b)} - 1}{it} \right|$$

$\therefore$ it is integrable over any compact set.

Suppose, the characteristic function φ_X of a r.v. X is integrable over $\mathbb{R}$, ie,

$$\int_{\mathbb{R}} |\varphi_X(t)|\, dt < \infty.$$

In this case, we have, for continuity pts $a < b$.

$$F_X(b) - F_X(a) = \frac{1}{2\pi} \cdot \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \cdot \varphi_X(t)\, dt$$

* verify:

$$\int_a^b e^{-itu}\, du = \frac{-\left(e^{-itb} - e^{-ita}\right)}{it}.$$

[$\because$ "pretend" that, this is real.]

$$F_X(b) - F_X(a) = \frac{1}{2\pi} \cdot \underbrace{\int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it}}_{\parallel} \cdot \varphi_X(t)\, dt$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_a^b e^{-itu}\, du \right) \cdot \varphi_X(t)\, dt$$

$$= \int_a^b \left(\frac{1}{2\pi} \int e^{-itu} \cdot \varphi_X(t)\, dt \right) du$$

$\rightarrow$ "Inverse Fourier

$$F_X(b) - F_X(a) = \int_a^b \underbrace{\left(\frac{1}{2\pi} \cdot \int_{\mathbb{R}} e^{-itv} \cdot \varphi_X(t)\, dt \right)}_{} dv$$

"Inverse Fourier Transform".

it "looks like", this is some density !!

Denote $f(v) = \frac{1}{2\pi} \cdot \int_{\mathbb{R}} e^{-itv} \cdot \varphi_X(t)\, dt \quad , v \in \mathbb{R}$.

this is integrable because,

$$\left| e^{-itv} \cdot \varphi_X(t) \right| \leq \left| \varphi_X(t) \right| < \infty$$

[$\because \varphi_X$ is integrable.]

* firstly, f is continuous. (Proof: take seq. $v_n \to v$.
$\Rightarrow e^{-itv_n} \cdot \varphi_X(t) \to e^{-itv} \cdot \varphi_X(t)$
$\therefore$ By DCT, integral converges)

* for every pair $a < b$ of continuity points, (which is dense)
& $F_X(b) - F_X(a) = \int_a^b f(v)\, dv$.

Right cont. at every (a, b)

cont. for every (a, b)

& LHS, RHS agree on a dense set,

then, by Lebesgue criterion, this holds.

In particular, F is continuous.

In particular, F is continuous.

Also, $F_X(b) - F_X(a) = \int_a^b f(v)\,dv$ $\forall (a,b)$.

$\therefore$ By Fundamental Theorem of Calculus,

F is differentiable,

& $F'(v) = f(v)$,

& Hence, we're done. $\boxtimes$

Exercise: $f_X(x) = \frac{1}{2} e^{-|x|}$. $\leftarrow$ symmetric 2-sided Exponential.

- ✓ compute C.F,
- ✓ observe that C.F is integrable
- Write the inverse fourier transform formula,

[Original density can be written as inverse transform formula of the C.F.]

Something should click !!.

$$\varphi_X(t) = E\left(e^{itX}\right) = \int_{\mathbb{R}} e^{itx} \cdot \frac{1}{2} e^{-|x|}\,dx$$

$$= \int_{-\infty}^{0} e^{itx} \cdot \frac{1}{2} \cdot e^{x}\,dx + \int_0^\infty e^{itx} \cdot \frac{1}{2} e^{-x}\,dx$$

$$= \int_{-\infty}^{0} \frac{1}{2} e^{(it+1)x}\,dx + \int_0^\infty \frac{1}{2} e^{(it-1)x}\,dx$$

$$= \frac{1}{2} \cdot \frac{e^{(it+1)x}}{it+1} \Bigg|_{-\infty}^{0} + \frac{1}{2} \cdot \frac{e^{(it-1)x}}{it-1} \Bigg|_{0}^{\infty}$$

$$= \frac{1}{2} \left[\frac{1}{it-1} + \frac{1}{it+1} \right] = \frac{1}{1+t^2}$$

$$\boxed{\therefore \varphi_X(t) = \frac{1}{1+t^2}}$$

$$\int_{\mathbb{R}} |\varphi_X(t)| \, dt = \int_{\mathbb{R}} \frac{1}{1+t^2} \cdot dt = \tan^{-1} t \Big|_{-\infty}^{\infty}$$

$$= \frac{\pi}{2} - \left(-\pi/2\right) = \pi.$$

$\therefore \varphi_X(t)$ is integrable.

Now, By Inverse Fourier Transform formula -

$$f(v) = \int_{\mathbb{R}} e^{-itv} \cdot \varphi_X(t) \, dt$$

Result: φ is a characteristic function which is integrable on $\mathbb{R}$.

Then, the underlying distribution function is continuously differentiable on $\mathbb{R}$, & has a continuous density given by —

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

"Inverse-Fourier Transform".

⚠ Converse NOT true !!!

ie, a distribution f^n with continuously differentiable density need not have a integrable characteristic f^n.

Result: Suppose X is a r.v. taking only integer values,

then, the characteristic f^n φ_X would be periodic with period $= 2\pi$, & for $k \in \mathbb{Z}$,

$$P(X=k) = \frac{1}{2\pi} \cdot \int_{-\pi}^{\pi} e^{-itk} \varphi_X(t)\, dt .$$

[Proof: Exercise.]

Here, $\varphi_X(t) = E(e^{itX})$

$$= \sum_{m \in \mathbb{Z}} e^{itm} \cdot P(X=m) .$$

Moments & Characteristic Functions:

Let X - a real r.v and φ - its characteristic f^n.

Result 1: If $E|X| < \infty$, then φ is continuously differentiable, and $\varphi'(t) = E(iXe^{itX})$, $\forall t \in \mathbb{R}$.

Proof:

Now, $\left|\frac{e^{i\theta}-1}{i\theta}\right| \leq 1$, and $\frac{e^{i\theta}-1}{i\theta} \longrightarrow 1$ as $\theta \longrightarrow 0$.

$$\frac{\cos\theta - 1}{i\theta} + \frac{\cancel{i}\sin\theta}{\cancel{i}\theta}$$

$$\frac{\cos\theta-1}{i\theta} \to 0, \qquad \frac{\sin\theta}{\theta} \to 1$$

as $\theta \longrightarrow 0$.

$$\therefore \frac{\varphi(t+h) - \varphi(t)}{h}$$

$$= E\left(\frac{e^{itX}(e^{ihX}-1)}{h}\right)$$

$$= E\left(\underbrace{e^{itX}}_{\leq 1} \underbrace{\frac{(e^{ihX}-1)}{ihX}}_{\leq 1} \cdot \underbrace{iX}_{\leq |X|}\right), \quad \& \; E|X| < \infty$$

$\therefore$ By DCT $\xrightarrow{\text{this}} E\left(e^{itX} \cdot iX\right)$ $\left[\because \frac{e^{ihx}-1}{ihx} \to 1 \text{ as } h \to 0\right]$

$\therefore \varphi'(t) = E\left(e^{itX} \cdot iX\right)$. ▨

Result-2:

Suppose $E(X^2) < \infty$.

Suppose $E(X^2) < \infty$.

Then, $\varphi'(t)$ is continuously differentiable

& , $\varphi''(t) = -E(X^2 \cdot e^{itX})$.

Proof: $$\frac{\varphi'(t+h) - \varphi'(t)}{h} = E\left(iX \cdot e^{itX} \frac{(e^{ihX} - 1)}{ihX} \cdot iX \right)$$

$$= E\left(-X^2 \cdot e^{itX} \cdot \underbrace{\left(\frac{e^{ihX} - 1}{ihX} \right)}_{\downarrow \, 1} \right)$$

Again, DCT : $\downarrow$

$$E(-X^2 e^{itX})$$

$$\therefore \varphi'(t) = -E(X^2 e^{itX})$$

$\therefore$ Result 'k'th :

If $E(|X|^k) < \infty$,

then φ is k-times continuously differentiable, and

$$\varphi^{(j)}(t) = E(i^j \cdot X^j \cdot e^{itX})$$

$$\forall \; j = 1, 2, \ldots, k$$

In particular, if $E(X^k) < \infty$ for some $k \geq 1$.

then,

$$\varphi(t) = 1 + \sum^{k} \frac{i^j}{j!} \cdot E(X^j) + o(t^k)$$

$$\varphi(t) = 1 + \sum_{j=1}^{n} \frac{i^j}{\lfloor j} \cdot E(X^j) + o(t^k)$$

$$\frac{o(t^k)}{t^k} \to 0 \text{ as } t \downarrow 0$$

(finite Taylor Expansion)

"The ultimate aim is to do well in exams. Don't deny that!" – Prof. AG, 22nd Oct.

$X_n, n \geq 1$, X – r.vs.

$\varphi_n, n \geq 1$ φ – c.Fs.

Levy's Continuity Theorem:

$X_n \xrightarrow{d} X$ iff

$\varphi_n \longrightarrow \varphi$ pointwise.

Proof:

"only if": $X_n \xrightarrow{d} X$.

$\Rightarrow E(\cos tX_n) \longrightarrow E(\cos tX)$

& $E(\sin tX_n) \longrightarrow E(\sin tX)$. $\forall t \in \mathbb{R}$.

$\therefore E(\cos tX_n + i \sin tX_n) \longrightarrow E(\cos tX + i \sin tX)$.

$\therefore$ Clearly, $\varphi_n(t) \longrightarrow \varphi(t)$ $\forall t \in \mathbb{R}$.

Levy's Theorem.

Suppose, $\varphi_n, n \geq 1$ – sequence of c.Fs.

Suppose, $\varphi_n \longrightarrow$ some function g pointwise suppose, $\varphi_n \longrightarrow$ some function g pointwise

If g is continuous at 0, then g is a characteristic f^n and

$$F_n \xrightarrow{d} F.$$

["if" part will follow from here.]

Helly's Selection Theorem:

Given any sequence $\{F_n\}$ of probability distribution f^ns, there is a subsequence $\{F_{n_k}\}$ and a non-decreasing right continuous f^n $G : \mathbb{R} \to [0,1]$ s.t.

$$F_{n_k}(a) \longrightarrow G(a)$$
$$\forall\, a \in C(G)$$

$C(G)$: set of all cont. pts of G.

* A subset $A \subseteq \mathbb{R}$ is called "Relatively Compact" if A has a compact closure.
[characterized by: every seq. has a subseq. which converges, (not necessarily within A.)]

Eg. F_n = pdf of $X_n = \pm n$, each with prob $= \frac{1}{2}$.

verify that: We can never get a "G" s.t. G is a cdf.

Eg. F_n = pdf of Unif$(-n, n)$.

Again, verify this

[Intuition: as $n \nearrow \infty$, the probability mass is

[Intuition: as $n \nearrow \infty$, the probability mass is "dissipated" over ∞.]

* In addition, if $\{F_n\}$ is "tight", i.e,
$\forall \varepsilon > 0, \exists K_\varepsilon > 0$ s.t.
$$F_n(K_\varepsilon) - F_n(K_\varepsilon^-) > 1 - \varepsilon \quad \forall n.$$

then, the limit f^n G is indeed a probability dist^n f^n.

[i.e, the compact set $[-K_\varepsilon, K_\varepsilon]$ covers more than $1-\varepsilon$ mass. i.e, this prevents the "escaping" of probability mass.]

- -

Let $r_1, r_2, \ldots$ be an enumeration of $\mathbb{Q}$.

$\{F_n(r_1)\}$ – seq. of cdfs.

$\therefore$ By Bolzano-Weierstrass thm

$\exists$ subsequence $n(1,j)$ s.t.

$$F_{n(1,j)}(r_1) \longrightarrow l_1.$$

Now, take $\{F_{n(1,j)}(r_2)\}$ – new seq.

Again, by Bolzano-Weierstrass thm,

$\Rightarrow \exists$ subseq. $n(2,j)$

s.t.

$$F_{n(2,j)}(r_i) \longrightarrow l_i, \quad i = 1, 2$$

$\Big[\longrightarrow l_1$ as, any further subsequence of a convergent subsequence is also convergent,

& $\longrightarrow l_2$, by Bolzano - Weierstrass (applied freshly) $\Big]$

Proceeding this way, and then looking at the diagonal subsequence will give us $\{F_{n_k}\}$

$$F_{n_k}(r_i) \longrightarrow l_i \quad \forall i$$

as $k \longrightarrow \infty$

Define $H : \mathbb{Q} \longrightarrow [0,1]$.

$$H(r_i) = l_i$$

H - non decreasing.

why? say, $r, s \in \mathbb{Q}$, $r < s$.

$$\begin{array}{ccc} F_{n_k}(r) & < & F_{n_k}(s) \\ \downarrow & & \downarrow \\ H(r) & & H(s). \end{array}$$

Define $G : \mathbb{R} \longrightarrow [0,1]$ by $G(x) = \inf\left\{ H(r) : \begin{array}{l} r > x \\ r \in \mathbb{Q} \end{array} \right\}$

Now, <u>show that</u>. if 'a' is a continuity point of G, then $F_{n_k}(a) \longrightarrow G(a)$

- -

∴ Levy's real thm:

φ_n's - characteristic f^{n}s.

$\varphi_n \longrightarrow \varphi$ pointwise, &

$$\boxed{\varphi \text{ is continuous at } 0}$$

$\xrightarrow{d.} F$

g is continuous at 0

Then, g is a C.F. & $F_n \xrightarrow{d.} F$

Sketch of proof:

$\{F_n\}$ - seq. of cdfs corresponding to $\{\varphi_n\}_{n\geq 1}$.

$\therefore$ By Helly, $\exists$ a subsequence n_k & a sub p.d.f $G : \mathbb{R} \to [0,1]$

s.t. $F_{n_k}(a) \longrightarrow G(a) \ \forall a \in C(G)$.

Now, $\varphi_n \longrightarrow g$ pointwise, & g is cont. at 0 $\Big\} \Rightarrow \{F_n\}$ is tight

[Show: by "delicate" integral estimates]

Now, as $\{F_n\}$ - tight

$\Rightarrow G$ is a cdf.

(Helly, additional part.)

$\therefore$ subseq. $F_{n_k}(a) \longrightarrow G(a)$

$\therefore \varphi_{n_k} \longrightarrow \varphi_G$ ptwise.

But, $\varphi_n \longrightarrow g$ ptwise $\Bigg\} \Rightarrow \boxed{\varphi_G = g}$.

We can apply Helly's further on sub-subsequences.

CENTRAL LIMIT THEOREM (C.L.T.)

$\{X_n\}$ - iid with mean μ, variance σ^2.

$\{X_n\}$ - iid with mean μ, variance σ^2.

$$\Rightarrow \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1), \quad S_n = X_1 + \cdots + X_n$$

Proof: Because of Levy's continuity theorem, its enough to show:

$$\varphi_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) \longrightarrow e^{-t^2/2}$$

Can't compute this!!

So, Denote $Y_k := \frac{X_k - \mu}{\sigma}$ ($\equiv$ "Y")

$\therefore Y_k$ - iid with mean 0, variance 1.

$$\therefore \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{T_n}{\sqrt{n}}, \quad \text{where } T_n = Y_1 + \cdots + Y_n.$$

$$\therefore \text{to show: } \varphi_{T_n/\sqrt{n}}(t) \longrightarrow e^{-t^2/2}$$

Now, $\varphi_{T_n/\sqrt{n}}(t) = \varphi_{T_n}(t/\sqrt{n})$

$$= \left(\varphi_Y(t/\sqrt{n})\right)^n$$

Now, $\varphi_Y = 1 - \frac{h^2}{2} + o(h^2)$ $\left[\because \begin{array}{l} E(Y) = 0 \\ E(Y^2) = 1 \end{array}\right]$

[Finite Taylor Expansion]

$$\therefore \left(\varphi_Y(t/\sqrt{n})\right)^n = \left(1 - t^2/2 + R_n\right)^n,$$

$$\therefore \underbrace{\left(\varphi_Y(t/\sqrt{n})\right)^n}_{\substack{\shortparallel \\ a_n \text{ (say)}}} = \left(1 - t^2/_{2n} + R_n\right)^n,$$

where $nR_n(t) \longrightarrow 0$ for $t \neq 0$ [fixed]

$$\therefore \lim_{n\to\infty} a_n \overset{?}{=} \lim_{n\to\infty} \left(1 - t^2/_{2n}\right)^n$$

$$= e^{-t^2/2}. \quad \text{(done!!)}$$

Left to show: this

Lemma: for complex numbers $z_1, z_2, \ldots, z_n$ & $w_1, w_2, \ldots, w_n$, all with modulus ≤ 1,

$$|z_1 \cdot z_2 \cdot \ldots \cdot z_n - w_1 \cdot w_2 \cdot \ldots \cdot w_n| \leq \sum_{k=1}^{n} |z_k - w_k|$$

$\therefore$ By Induction,

$$|z_1 \cdots z_n - w_1 \cdots w_n|$$

$$\leq |z_1 \cdots z_{n-1} \cdot z_n - z_1 \cdots z_{n-1} \cdot w_n| + |z_1 \cdots z_{n-1} \cdot w_n - w_1 \cdots w_{n-1} \cdot w_n|$$

$$\leq |z_n - w_n| + |z_1 \cdots z_{n-1} - w_1 \cdots w_{n-1}|$$

$[\because |w_n| < 1]$

$$\therefore \left| \left(1 - t^2/_{2n} + R_n\right)^n - \left(1 - t^2/_{2n}\right)^n \right|$$

$$\leq n \cdot |R_n| \longrightarrow 0. \quad \left[\because R_n = o\left(\tfrac{1}{n}\right)\right]$$

Hence, done!!

Hence, done!!

Course ends here!!

Respect to Prof. AG!

Probability-3 Lecture-21

29 October 2024 11:09

Recall: qs. from midsem.

$(\Omega, \mathcal{A}, P)$ – a discrete probability space.

ie, Ω – countable.

$\Omega = \{\omega_1, \omega_2, \ldots\}$

&, $P(A) = \sum_{i : \omega_i \in A} P(\{\omega_i\})$

$X_n \xrightarrow{P} X$.

to show: $X_n \xrightarrow{a.s} X$

Proof: Suppose,

$X_n \overset{a.s}{\not\longrightarrow} X$

$\Rightarrow P\Big(\underbrace{\{\omega : X_n(\omega) \not\longrightarrow X(\omega)\}}_{A \text{ (say)}}\Big) > 0 \quad (\text{ie}, \neq 0)$

$\because \Omega$ is discrete.

$\Rightarrow \exists\, \omega_0 \in A$ s.t,

$P(\{\omega_0\}) = \underset{(\text{say})}{\delta} > 0$.

$X_n(\omega_0) \not\longrightarrow X(\omega_0)$

$\Rightarrow \exists\, \varepsilon > 0$ and a subsequence $\{n_k\}$ such that:

$$|X_{n_k}(\omega) - X(\omega)| > \varepsilon$$

$$P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \xrightarrow{?} 0$$

But, $\exists\, n_k \geq N$,

s.t. $P(\{\omega : |X_{n_k}(\omega) - X(\omega)| > \varepsilon\}) \geq \delta$

$\forall\, n \geq N$.

$\therefore X_n \overset{P}{\not\longrightarrow} X$ [which is a contradiction !!].

$\therefore X_n \xrightarrow{a.s} X$.

$\therefore X_n \xrightarrow{a.s} X.$

[contradiction !!]

■

* whenever you write density functions, write their supports (always !!)

$\underset{\sim}{X} = (X_1, X_2, \dots, X_m)$ is a random vector which has a joint density.

$f_{\underset{\sim}{X}}(x_1, \dots, x_m)$ for $(x_1, \dots, x_m) \in I$ (a connected open subset $\subseteq \mathbb{R}^m$)

Suppose, we have

$$\varphi : \underset{\mathbb{R}^m}{\underset{\cap}{I}} \longrightarrow \underset{\mathbb{R}^m}{\underset{\cap}{J}}$$

[Jacobian rule applies only on same dim. spaces $\mathbb{R}^m \longrightarrow \mathbb{R}^m$.]

$\therefore \underset{\sim}{Y} = \varphi(\underset{\sim}{X}) \in J$

$= (Y_1, Y_2, \dots\dots, Y_m)$
(say)

Sufficient conditions for Y to have joint density:

1. $\varphi : I \longrightarrow J$ is 1-1, onto.

: let $\varphi^{-1} : J \longrightarrow I$.

then, ψ^{-1} is 1-1, onto.

(2) Let us write:

$$\psi(y_1, y_2, \ldots, y_m) = (x_1, x_2, \ldots, x_m)$$

[ie, each x_i is a real-valued function of $y_1, y_2, \ldots, y_m$.]

So, we can talk about $\left(\frac{\partial x_i}{\partial y_j}\right)$

Condition: $\left(\frac{\partial x_i}{\partial y_j}\right)$ exist & is 1-1, onto in J for all (i, j)

Define $J(\underset{\sim}{y}) = \det\left(\left(\frac{\partial x_i}{\partial y_j}\right)\right)_{m \times m}$

"Jacobian"

(Has to be a square matrix..)

J is continuous.

(3) $J(\underset{\sim}{y}) \neq 0 \quad \forall \underset{\sim}{y} \in J.$

ie, $J(\underset{\sim}{y}) \neq 0$, & J continuous.

$\Rightarrow$ J non-zero

Under all these conditions, $\underset{\sim}{Y}$ has a density

$$f_{\underset{\sim}{Y}}(y_1, \ldots, y_m) = f_{\underset{\sim}{X}}(\underbrace{\psi}_{\varphi^{-1}}(y_1, \ldots, y_n)) \cdot |J(y)|, \quad \underset{\sim}{y} \in J.$$

Eg. $X, Y \overset{iid}{\sim} Exp(\lambda)$.

$U = X+Y$
$V = X-Y$. find the joint density of (U,V).

Here, $\underset{\sim}{X} = (X,Y)$ has a density—

$$f(x,y) = \lambda^2 \cdot e^{-\lambda(x+y)}, \quad x>0, \; y>0.$$

or, $I = (0,\infty)\times(0,\infty)$

$\varphi : I \longrightarrow J$

$(x,y) \longmapsto (U,V)$.

$= (x+y, \; x-y)$

ie, $\begin{bmatrix} U \\ V \end{bmatrix} = A \cdot \begin{bmatrix} x \\ y \end{bmatrix}$,

$A - 2\times 2$.

$A = ?$

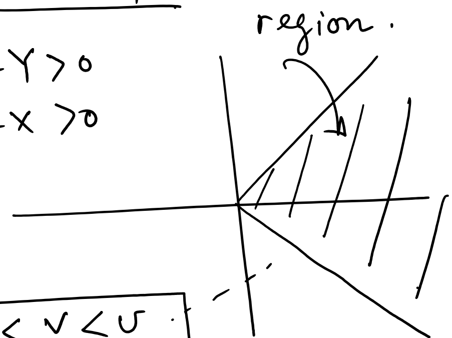
$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Now, $U = X+Y$
$V = X-Y$

$U - V = 2Y > 0$
$U + V = 2X > 0$
$\Rightarrow V < U$

$\Rightarrow V > -U$

$\boxed{\therefore -U < V < U \text{ \& }, U > 0}$



$\therefore \psi : (U,V) \longmapsto (x,y)$

$$= \left(\tfrac{1}{2}(U+V), \; \tfrac{1}{2}(U-V)\right)$$

$$\therefore D\psi(U,V) = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$$

$$\therefore |J| = \left|\det\left(D\psi(U,V)\right)\right|$$
$$= 1/2.$$

$\therefore$ The density of (U,V) is $g(u,v) = \lambda^2 \cdot e^{-\lambda u} \cdot \frac{1}{2}$; $\forall (u,v) \in J$.

$$J = \left\{ (u,v) : \begin{array}{l} u > 0, \\ -u < v < u \end{array} \right\}.$$

Eg. Y_1, Y_2, Y_3, Y_4 i.i.d Gamma with parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$

ie, $Y_i \sim \text{Gamma}(1, \alpha_i) \equiv \text{Gamma}(\alpha_i)$

scale parameter

$$X_1 = \frac{Y_1}{Y_1 + Y_2 + Y_3 + Y_4}$$

$$X_2 = \frac{Y_2}{Y_1 + Y_2 + Y_3 + Y_4}$$

$$X_3 = \frac{Y_3}{Y_1 + Y_2 + Y_3 + Y_4}$$

find joint density of X_1, X_2, X_3.

take: $Z = Y_1 + Y_2 + Y_3 + Y_4$
(to make a $\mathbb{R}^4 \longrightarrow \mathbb{R}^4$ transformation, hence to get a square matrix for det $|\cdot|$.

$$\therefore f(y_1, y_2, y_3, y_4) = \frac{e^{-(y_1+y_2+y_3+y_4)}}{\Gamma(\alpha_1) \cdot \Gamma(\alpha_2) \cdot \Gamma(\alpha_3)\, \Gamma(\alpha_4)} \cdot y_1^{\alpha_1 - 1}\, y_2^{\alpha_2 - 1}\, y_3^{\alpha_3 - 1}\, y_4^{\alpha_4 - 1}$$

∴ consider the transformation

$\varphi : I \longrightarrow J$

$\varphi : (y_1, y_2, y_3, y_4) \longmapsto (x_1, x_2, x_3, z)$;

$\qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \quad \parallel$

$I = (0,\infty)^4$

$$\left(\frac{y_1}{\sum y_i}, \frac{y_2}{\sum y_i}, \frac{y_3}{\sum y_i}, \sum y_i \right)$$

$J = ?$

∴ $0 < x_1 < 1$

$0 < x_2 < 1$

$0 < x_3 < 1$

$x_1 + x_2 + x_3 \leq 1$

$$J = \left\{ (x_1, x_2, x_3, z) \;\middle|\; \begin{array}{l} x_1 > 0 \\ x_2 > 0 \\ x_3 > 0 \\ x_1 + x_2 + x_3 \leq 1 \end{array} \right\}$$

∴ the inverse transformation:

∴ $\psi : J \longrightarrow I$.

$(x_1, x_2, x_3, z) \longmapsto (y_1, y_2, y_3, y_4)$

$\qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \parallel$

$$\left(x_1 z, x_2 z, x_3 z, (1 - x_1 - x_2 - x_3) \cdot z \right)$$

Notice ↗
all coordinates > 0 ✓

[∴ J, as we evaluated is correct.]

$$J = \det \begin{pmatrix} z & 0 & 0 & x_1 \\ 0 & z & 0 & x_2 \\ 0 & 0 & z & x_3 \\ -z & -z & -z & 1 - x_1 - x_2 - x_3 \end{pmatrix}$$

$= \qquad 3$

$$J = z^3$$

$$\left[\because z = y_1 + y_2 + y_3 + y_4\right]$$

$$\therefore g(x_1, x_2, x_3, z) = \frac{1}{\Gamma(\alpha_1)\cdot\Gamma(\alpha)\cdot\Gamma(\alpha_3)\cdot\Gamma(\alpha_4)} \cdot e^{-z} \cdot$$

$$(x_1 z)^{\alpha_1 - 1} (x_2 z)^{\alpha_2 - 1} \cdot (x_3 z)^{\alpha_3 - 1} \cdot z^{\alpha_4 - 1} \cdot (1 - x_1 - x_2 - x_3)^{\alpha_4 - 1} \cdot \underset{|J|}{\underset{\uparrow}{z^3}}$$

$$(x_1, x_2, x_3, x_4) \in J.$$

$$= \frac{1}{\Gamma(\alpha_1 + \cdots + \alpha_4)} \cdot \frac{\Gamma(\alpha_1 + \cdots + \alpha_4)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_4)} \cdot e^{-z} \cdot z^{\alpha_1 + \cdots + \alpha_4 - 1} \cdot x_1^{\alpha_1 - 1} \cdot x_2^{\alpha_2 - 1} \cdot x_3^{\alpha_3 - 1} \cdot (1 - x_1 - x_2 - x_3)^{\alpha_4 - 1}$$

itself a density. ($\int (\cdot)\, dz = 1$).
integrating out this whole expression w.r.t z, this becomes the density of Dirichlet $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$.

Concept : Tail event.

$\{X_1, X_2, \ldots\}$.
An event A is a <u>tail event</u> of $A \in \sigma(X_n, X_{n+1}, \ldots)$ $\forall\, n \geq 1$.

Eg. Event $A := \{ \omega : X_1(\omega) + X_2(\omega) + X_3(\omega) \leq 3 \}$.

then, $A \in \sigma(X_1, X_2, X_3)$.

ie, if A is a tail event, with just any given tail of the seq. $X_1, X_2, \ldots$ we can say whether A occurs or not.

Eg. $A = \{ \omega : \lim_n X_n(\omega) \text{ exists} \}$.

$A = \{ \omega : \sum_n X_n(\omega) \text{ converges} \}$.

these are tail events.

$A = \{ \omega : \sum_n X_n > 0 \quad \text{converges} \}$ → NOT a tail event.

for $\sum X_n > 0$ we require to know each of the X_i's.

Eg. $A = \{ \omega : \sup_n X_n > 23 \}$.

→ Not tail event.

why?

we can have

$$X_i = \begin{cases} 25, & i = 1 \\ 0, & i \geq 2 \end{cases}$$

So, information on any tail isn't enough.

What's a tail r.v ?

Say, Y is a r.v.

if the event $A_B = Y \in B$ $\forall B$ is

Borel set.

a tail event, then

Y is a tail r.v.

In practice, we don't need to see for all such Borel set 'B'

We have to check for things whether,

- the event $Y \geq a$ for some a, or $\forall a$

is a tail event or not

- the event $Y \leq a$ for some a, or $\forall a$

is a tail event or not.

Eg: $Y = \sum_n X_n$ is NOT a tail r.v,

$\because A := \{ \omega : \sum_n X_n > 0 \}$ is NOT a tail event.

Probability-3 Lecture-22

01 November 2024 11:25

<u>Set - 3</u>

2. (i) $A = \{\omega : \{X_n(\omega)\}$ remains bounded$\}$.

take <u>any $k \in \mathbb{N}$.</u>

$$A^{(k)} = \{\omega : \{X_{n+k}(\omega)\}_{n \geqslant 1} \text{ is bounded.}\}.$$

Clearly, $A = A^{(k)}$ for every k.

$$\therefore A^{(k)} \in \underbrace{\sigma(X_{k+1}, X_{k+2}, \ldots)}_{\mathcal{T}_k}$$

a subset $A \subseteq \mathbb{R}$ is bounded iff $\forall$ finite subsets $F \subseteq A$, $A \setminus F$ is bounded.

$\therefore A \in \mathcal{T}_k \quad \forall k$

$\therefore A \in \bigcap_k \mathcal{T}_k = \mathcal{T}$. ▨

(ii) $\{S_n > 0 \text{ i.o}\} \equiv \{X_1 + \cdots + X_n > 0 \text{ i.o}\}$

$\equiv \{X_1 > -(X_2 + \cdots + X_n)$ for infinitely many $n.\}$

$\notin \sigma(X_{k+1}, \ldots) \quad \forall k \geqslant 1$

$\therefore$ this is <u>NOT</u> a tail event.

3.(b) $\{X_n\}$-independent seq.

$\Leftarrow$: Fix <u>$k \geqslant 1$</u>.

Fix $k \geq 1$.

$$\limsup_n \left(\frac{S_n - S_k}{a_n} \right)$$ → this r.v is measurable w.r.t $\sigma(X_{k+1}, \ldots, X_{k+n}) = \mathcal{T}_k$.

$$\overset{?}{=} \limsup_n \frac{S_n}{a_n}$$

Yes.

2 seq: $(a_n - b_n)$ s.t $b_n \to 0$

$\therefore \limsup_n (a_n - b_n) = \limsup_n a_n$.

$\because \lim_{n\to 0} \frac{S_k}{a_n} = 0$, as S_k - fixed, & $a_n \nearrow \infty$.

4.(b) $\{X_n\}$ - i.i.d.

(by contradiction)

Suppose, $\{X_n\}$ - not degenerate at 0.

$\Rightarrow \exists\ a > 0$ s.t. either

$$P(X_n > a) = P(X_1 > a) = \delta > 0$$

or $P(X_n < -a) = P(X_1 < -a) = \delta > 0$

$$\therefore \sum_n P(X_n > a) = \sum_n P(X_1 > a) = \infty$$

$\therefore$ By B.C-II,

$$P(X_n > a \ \text{ i.o}) = 1.$$

$$\Rightarrow P\left(\sum_n X_n \text{ conv.}\right) = 0. \quad \blacksquare$$

(d) Again $\{X_n\}$ - i.i.d.

(d) Again $\{X_n\}$ - i.i.d.

Here, X_n's are "symmetric".

i.e, X_n & $-X_n$ have same r.v.

So, such an r.v, if it is degenerate, it has to be " at only & only $X_n = 0$

So, by contradiction: if X_n - <u>not</u> degenerate,

$\Rightarrow$ X_n - <u>not</u> degenerate at 0.

$\exists\ a > 0$, s.t $P(X_n > a) = P(X_n < -a) = \delta > 0$.

$\therefore P(\limsup S_n = \infty) = 1$

$P(\liminf S_n = -\infty) = 1$.

Qs. 7. think, but don't waste much time on that.

Qs. 8 $\rightarrow$ out of syllabus for the time being.
(requires knowledge of "infinite products.")

* Unless specified, r.v $\equiv$ real r.v
(ie, <u>NOT</u> extended real)

9. (a) $\{X_n\}$ - i.i.d.

$M_n = \text{Max.}\{|X_1|, |X_2|, \ldots, |X_n|\}$.

(a) $\dfrac{M_n}{n} \xrightarrow{P} 0 \iff n \cdot P(|X_1| > n) \to 0$.

" $\Leftarrow$ " $\dfrac{M_n}{n} \xrightarrow{P} 0$

$\therefore P\left(\frac{M_n}{n} > \varepsilon\right) = P(M_n > n\varepsilon)$

$= 1 - P(M_n \le n\varepsilon)$

$= 1 - P(|X_1| \le n\varepsilon,\ |X_2| \le n\varepsilon,\ \cdots |X_n| \le n\varepsilon)$

$\{X_i\}$ - iid

$= 1 - [P(|X_1| \le n\varepsilon)]^n$

$= 1 - (1 - P(|X_1| > n\varepsilon))^n$

$\overset{?}{\le} n P(|X_1| > n\varepsilon)$

[Expand using Taylor (upto 2nd term ?!)]

$\le n P(|X_1| > [n\varepsilon])$ Greatest integer f^n.

$= \frac{1}{\varepsilon} \cdot \frac{n\varepsilon}{[n\varepsilon]} \cdot [n\varepsilon] \cdot P(|X_1| > [n\varepsilon]) \to 0$

from hypothesis.

" $\Rightarrow$ " $\frac{M_n}{n} \xrightarrow{P} 0$ ..

$\therefore \frac{X_n}{n} \xrightarrow{P} 0$.

to prove:

[$(1-a_n)^n \to 1 \Leftrightarrow n a_n \to 0$]

" $\Leftarrow$ "

$(1-a_n)^n \ge 1 - n a_n$.

$\Rightarrow n a_n \ge 1 - (1-a_n)^n$

$\Rightarrow na_n \geq 1-(1-a_n)^n$

we used this previously.

9. (b) (discussion)

$\{x_n\}$ - real seq.

$$m_n := \operatorname{Max}\{|x_1|, \dots, |x_n|\}.$$

$$\frac{x_n}{n} \to 0 \iff \frac{m_n}{n} \to 0$$

"$\Leftarrow$" trivial.

"$\Rightarrow$" $\frac{x_n}{n} \to 0$.

take $\varepsilon > 0$. $\therefore \exists N$ s.t, $\frac{|x_n|}{n} < \varepsilon \quad \forall n \geq N$.

$$\therefore \frac{m_n}{n} < \frac{(|x_1| \vee |x_2| \vee \dots \vee |x_{n-1}|)}{n} \vee \varepsilon.$$

$$\therefore \frac{X_n}{n} \xrightarrow{a.s} 0$$

$$\iff \forall j \geq 1, \; P\left(\bigcap_n \bigcup_{k \geq n} \left\{ \left|\frac{X_k}{k}\right| > \frac{1}{j} \right\}\right) = 0.$$

$$E|X_1| < \infty \iff \forall j \geq 1, \; E|j \cdot X_1| < \infty$$

[ie, any r.v has finite mean $\Rightarrow$ any multiple of it has finite mean.]

$$\Rightarrow \sum_n P(|jX_1| > n) < \infty$$

B.C-I: $\Leftrightarrow P(|jX_1| > n \text{ i.o.}) = 0$

for independent case

(as X_i's iid), this becomes $\Leftrightarrow$.

10. Hint: Truncate the seq. at λ. Apply SLLN.

$\lambda \to \infty$ (MCT or something)

$$\frac{S_n}{n} \xrightarrow{a.s.} E(X^\lambda) .$$

$$\liminf \frac{S_n}{n} \geqslant \lim \frac{S_n^*}{n} \overset{a.s.}{=} E(X_1^\lambda) \quad \forall \lambda.$$

$\lambda \uparrow \infty$. MCT.

$E(X) = +\infty$.

[Instead of λ,
use M - some large integer.
Why? Some "Null-sets" or smth.
(dealing with uncountable null sets)]

11. (b) $\{X_n\}$ - i.i.d.

$\frac{X_n - c_n}{n} \xrightarrow{a.s.} 0$ for some real $\{c_n\}$.

$\Leftrightarrow E(|X_1|) < \infty$.

& in that case, $\frac{c_n}{n} \to 0$.

($\Leftarrow$) $E(|X_1|) < \infty$.

$\therefore$ we exhibit one real seq,

$c_n = 0$. $\quad \frac{X_n - c_n}{n} = \frac{X_n}{n} \xrightarrow{a.s} 0$

So, done. ✓

✓.

($\Longrightarrow$) we now have, for some real seq. $\{c_n\}$,

$$\frac{X_n - c_n}{n} \xrightarrow{a.s} 0 \quad \text{for some } \{c_n\}.$$

$$\therefore \frac{X_n - c_n}{n} \xrightarrow{P} 0$$

11. part (a)

$$\frac{X_n}{n} \xrightarrow{P} 0 \text{ (always)}$$

$$\therefore \frac{c_n}{n} \longrightarrow 0.$$

Probability-3 Lecture-23

05 November 2024 14:18

Set-3 (contd...)

4.(d) $P(\sup S_n = \infty, \inf S_n = -\infty) = 1.$

$\exists \varepsilon > 0,\ P(X_n > \varepsilon \text{ i.o},\ X_n < -\varepsilon \text{ i.o}) = 1.$

Special Case:

$X_n = \pm$ with prob $= 1/2$ (S.S.R.W)

fix an integer $j \geq 1$ $S_0 = 0$

$\theta_j = P(\sup_n S_n \geq j)$ $\theta_0 = 1$

$$\theta_j = \frac{1}{2}\theta_{j-1} + \frac{1}{2}\theta_{j+1}.$$

$$\theta_j - \theta_{j-1} = d$$

$$\therefore \theta_j = \theta_0 + j \cdot d$$

$\to$ Here, $\theta_j \to$ a prob.

& $0 \leq \theta_j \leq 1 \quad \forall j$

$\therefore \Rightarrow d = 0$ (forced)

$\Rightarrow \theta_j = 1 \quad \forall j$

$$\therefore \Rightarrow P\left(\left\{\bigcap_j \sup_n S_n \geq j\right\}\right) = 1$$

$$\Rightarrow P(\sup_n S_n = \infty) = 1.$$

Similarly, $P(\inf_n S_n = -\infty) = 1$

Similarly, $P\left(\inf_n S_n = -\infty\right) = 1$ $\blacksquare$

General case:

$\exists\, \varepsilon > 0$ s.t. $P(X_n > \varepsilon) = \delta > 0$. $\longrightarrow$ ie, X_n <u>not</u> degenerate at 0.

take any $k \geq 1$.

$$P(X_1 > \varepsilon, X_2 > \varepsilon, \dots, X_n > \varepsilon) = \delta^k > 0$$

$[\because X_i$ - iid$]$

Define:
$$\left.\begin{aligned} A_1 &= (X_1 > \varepsilon, \dots, X_k > \varepsilon) \\ A_2 &= (X_{k+1} > \varepsilon, \dots, X_{2k} > \varepsilon) \\ A_3 &= (X_{2k+1} > \varepsilon, \dots, X_{3k} > \varepsilon) \\ &\ \ \vdots \end{aligned}\right\} \to \text{Breaking into blocks, each of } k$$

$\therefore A_n$'s - independent.

$\therefore P(A_n) = \delta^k > 0 \quad \therefore \sum P(A_n) = \infty$

$\therefore$ By B.C-II:

$$P(\underbrace{A_n \text{ happens i.o}}_{C_k}) = 1.$$

ie, for every k, we can get a k-long run i.o s.t every term within that is $> \varepsilon$.

fix $M > 0$.

<u>Claim</u>: $\underbrace{\{|S_n| \leq M \ \forall\, n\} \cap C_k}_{} = \phi$ if $k > \frac{2M}{\varepsilon}$

Why? take $\omega \in C$

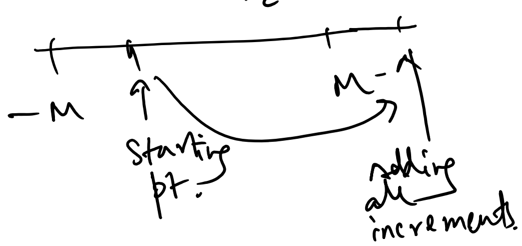
Why? take $\omega \in$

i.e., $\exists$ a k-long run s.t. every $X_i > \varepsilon$ in that run. So, if $S_n < -M$, done.

if $S_n > -M$, & $k > 2M/\varepsilon$, $S_n > M$!!

$\therefore$ done.

$\varepsilon \times 2M/\varepsilon = 2M$



$$\therefore P(|S_n| \leq M) = 0.$$

$$\therefore P(|S_n| = \infty) = 1$$

$$\Rightarrow P(\sup S_n = \infty \text{ or } \inf S_n = -\infty) = 1$$

12. $X_n = 2$ or n^α each with $p = \theta_n$.

$X_n = \theta_n$ with prob. $= 1 - 2\theta_n$.

$P(\sum X_n \text{ conv}) = 1$ iff $\sum \theta_n < \infty$.

$$\Rightarrow \sum P(X_n \neq \theta_n) < \infty.$$

Set - 4

1.(b) Note: (ii) & (iii) are just complements.

$\therefore$ (ii) $\Leftrightarrow$ (iii).

to show: (i) $\Leftrightarrow$ (ii).

"$\Leftarrow$" Assume, $\liminf P(X_n \in V) \geq P(X \in V)$ $\forall V$ open.

take $V = (-\infty, a)$

take $V = (-\infty, a)$

$$F_n(a^-) = \lim P(X_n \in (-\infty, a))$$
$$\geq P(X \in (-\infty, a))$$
$$= F(a^-).$$

&, take $V = (a, \infty)$.

Using $\underrightarrow{\liminf} (1 - F_n(a)) \geq 1 - F(a)$

$\Rightarrow \cancel{1} - \limsup F_n(a) \geq \cancel{1} - F(a)$

$\Rightarrow \limsup F_n(a) \leq F(a)$.

Now, refer to part 1.(a)

$X_n \xrightarrow{d} X$.

($\Rightarrow$) Assume, $X_n \xrightarrow{d} X$.

Use part 1. (a) to show that,

$\lim P(X_n \in I) \geq P(X \in I)$ $\forall$ open interval I.

Let V- open set.

Write $V = \bigcup_n I_n$

disjoint open intervals.

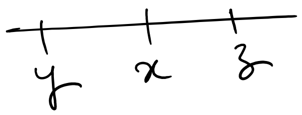
Given $\varepsilon > 0$, $\exists M > 0$ s.t.

$$P(X \in V) \leq \sum_{j=1}^{M} P(X \in I_j)$$

$$\liminf P(X_n \in V) \geq \liminf \sum_{j=1}^{M} P(X_n \in I_j)$$

$$\geq \sum_{j=1}^{M} P(X \in I_j) \geq P(X \in V) - \varepsilon$$

2. Idea: $x \in C(F)$. i.e. x is a fixed continuity pt.



Find $y < x < z$, $y, z \in C(F)$

s.t. $F(x) - F(y) < \varepsilon$.

$F(z) - F(x) < \varepsilon$.

(we can always do so, $\because C(F)$ is dense)

Now, $X_n \longrightarrow X$.

$\therefore$ after a large n, all x_n's are "trapped" within (y, z).

$\therefore \limsup_n F_n(x_n) \leq F_n(z)$

$\liminf_n F_n(x_n) \geq F_n(y)$

3. direct definition, & computation.

4. Let D be the countable set.

$P(X \in D) = 1$.

for any $\varepsilon > 1$.

$\exists$ a finite set $F \subseteq D$.

s.t. $P(X \notin F) < \varepsilon$.

$P(X \in V) > P(X_n \in V \cap F)$.

$$P(X \in V) > P(X_n \in V \cap F).$$

$$\downarrow$$

$$P(X \in V \cap F) > P(X \in V) - \varepsilon$$

$$\uparrow$$

By our choice of F.

$$\therefore \liminf P(X_n \in V \cap F) = P(X \in V \cap F).$$

5. (a) "local limit theorem"

6. (a) showing Δ - inequality :

F, G, H - 3 cdfs.

Suppose $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ are such that,

$$G(x - \varepsilon_1) - \varepsilon_1 < F(x) < G(x + \varepsilon_1) + \varepsilon_1 \quad \forall x \in \mathbb{R}.$$

and, $H(y - \varepsilon_2) - \varepsilon_2 < G(y) < H(y + \varepsilon_2) + \varepsilon_2 \quad \forall y \in \mathbb{R}.$

(should imply)

$$\Rightarrow \rho(F, H) \leq \varepsilon_1 + \varepsilon_2.$$

7. for any cdf F,

define $a_y = \sup\{x : F(x) < y\}$, $0 < y < 1$.

$\rightarrow$ left half

$$b_y = \inf\{x : \underset{\substack{\shortparallel \\ G(y)}}{F(x) > y}\}.$$

Here, $\forall y$, $a_y \geq b_y$ [think & check]

& $y_1 < y_2 \Rightarrow b_{y_1} \leq a_{y_2}$

if $a_y = b_y$, $(a_y, b_y) = \phi$.

ie, collect those $0 < y < 1$ s.t,

$(a_y, b_n) \neq \phi$,

ie, $a_y < b_y$

ie, $\{y : (a_y, b_y) \neq \phi\} =: D$ (say)

this is countable (think)

$\therefore P(Y \in D) = 0$

Now, show: for $y \notin D$,

$G_n(y) \longrightarrow G(y)$

Step-1. $\lim G_n(y) \geq G(y) - \varepsilon$.

Step. -2. $\lim G_n(y) \leq G(y) + \varepsilon$.

for step 1: pick $x \in C(F)$

$G(y) > x$.

$G(y) > x.$

& then, show, $\lim G_n(y) \geq x$.

$\rightarrow$ Not of use for exam.

9. (b) (converse)

$Y_n := e^{-X_n}, \quad Y = e^{-X}.$ $\rightarrow$ takes values in

All moments bounded.

Set-5

7. $1 - \frac{\lambda}{n} + \frac{\lambda}{n} \varphi(t)$ $\rightarrow$ convex combination of C.F $\therefore$ C.F ✓

a C.F (or r.v degenerate at 0)

11. (b). $X_1, X_2, \ldots \ldots = 1$ w.p $\left(\frac{1}{2}\right)$

$\varphi_n(t) =$ Char. f^n $\sum_{k=1}^{n} X_k/2^n$

$\xi_k = \frac{X_k + 1}{2}$. $\sum_{k=1}^{n} \frac{\xi_k}{2^k} = \frac{1}{2} \sum_{k=1}^{n} \frac{(X_k+1)}{2^k}$.

(first n terms of binary expansion)

$\sum_{k=1}^{n} \frac{\xi_k}{2^k} \xrightarrow{a.s.}$ Unif(0,1).

$\therefore \frac{1}{2}\sum_{k=1}^{n} \frac{(X_k+1)}{2^k} \xrightarrow{a.s}$ Unif(-1,1)

13-(c) $\varphi(t) = \int_{-\infty}^{\infty} e^{itx} \cdot h(x)\,dx$

Suppose, $\varphi(t)$ is a non-ve seq, real-valued integrable s.t, $\exists c > 0$