K-dimentional Random Variable

Suppose $X_1, X_2 \ldots X_k$ are r.v.s defined on the sample space Ω. In this case, we say $(X_1, X_2 \ldots X_k)$ is k dimentional ~~vector~~ random space. We also say that $X_1, X_2, \ldots X_k$ are jointly distributed r.v.s.

Defn: $\underset{\sim}{X} = (X_1 \ldots X_k)$

$\underset{\sim}{U} = (U_1, U_2 \ldots U_k)$

Def^n: Joint CDF of $(X_1, X_2 \ldots X_k)$ is

$$F_{X_1, X_2 \ldots X_k}(U_1, U_2 \ldots U_k) = P(X_1 \leq U_1, X_2 \leq U_2 \ldots X_k \leq U_k).$$

$$(U_1, U_2 \ldots U_k) \in \mathbb{R}^k.$$

$$\| \quad F_{\underset{\sim}{X}}(\underset{\sim}{U})$$

Discrete case: This means range$(X_1, X_2 \ldots X_k)$ is ctble $\Leftrightarrow$ Each X_i is a discrete r.v. Then we say $(X_1, X_2 \ldots X_k)$ is a discrete random vector.

Clearly, $\text{Range}(X_1, X_2 \ldots X_k) \subseteq \text{Range}(X_1) \times \text{Range}(X_2) \ldots \text{Range}(X_k)$.

Def^n: In this case joint ~~pdf~~ pmf of $(X_1, X_2 \ldots X_k)$ is defined as $P_{X_1, X_2 \ldots X_k}(x_1, x_2 \ldots x_k)$

$$:= P(X_1 = x_1, X_2 = x_2 \ldots X_k = x_k)$$

For any $B \subseteq \mathbb{R}^k$

$$P[(x_1, x_2 \ldots x_k) \in B] = \sum_{(x_1, x_2 \ldots x_k) \in B} p_{x_1, x_2 \ldots x_k}(x_1, x_2 \ldots x_k)$$

$$p_{x_1}(x_1) = \sum_{x_2} \sum_{x_3} p_{x_1, x_2, x_3}(x_1, x_2, x_3)$$

and so on.

Continous Case

Def^n: A random vector $\underset{\sim}{X} = (x_1, x_2 \ldots x_k)$ is called (absolutely) continous if $\exists$ a function $\xrightarrow{}$ integrable

$f_{\underset{\sim}{X}} = f_{x_1 \ldots x_k} : \mathbb{R}^k \longrightarrow [0, \infty)$ such that for all $\underset{\sim}{u} = (u_1, u_2 \ldots u_k) \in \mathbb{R}^k$.

$$F_{\underset{\sim}{X}}(\underset{\sim}{u}) = P(X_1 \le u_1, X_2 \le u_2 \ldots) = \int_{-\infty}^{u_k} \int_{-\infty}^{u_{k-1}} \cdots \int_{-\infty}^{u_1} f_{\underset{\sim}{X}}(\underset{\sim}{x})\, dx_1\, dx_2 \ldots dx_k$$

can also write $d(\underset{\sim}{x})$

In this case we say that $f_{\underset{\sim}{X}} = f_{X_1, X_2 \ldots X_k}$ is a joint pdf of $(X_1 \ldots X_k)$.

In this case, Range $(X_1 \ldots X_k) = \{\underset{\sim}{x} \in \mathbb{R}^k : f_{\underset{\sim}{X}}(\underset{\sim}{x}) > 0\}$

It can be shown that for all "nice" $B \subseteq \mathbb{R}^k$;

$$(\star) \quad P[\underset{\sim}{X} \in B] = \int\int_B \cdots \int f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d(\underset{\sim}{x})$$

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(x_1,x_2,x_3)\,dx_2\,dx_3,$$

$$f_{X_1,X_3}(x_1,x_3) = \int_{-\infty}^{\infty} f_{X_1,X_2,X_3}(x_1,x_2,x_3)\,dx_2$$ and so on.

whenever $\underset{\sim}{x}$ is a continous point of $f_{\underset{\sim}{X}}$ we have

$$\lim_{\substack{\Delta x_1 \to 0 \\ \vdots \\ \Delta x_k \to 0}} \frac{P[x_1 \le X_1 < x_1 + \Delta x_1, \ldots x_k < X_k < x_k + \Delta x_k]}{\Delta x_1 \Delta x_2 \ldots \Delta x_k}$$

$$= f_{\underset{\sim}{X}}(\underset{\sim}{x}).$$

Any joint pdf $f_{\underset{\sim}{X}}$ satisfies

(I) $f_{\underset{\sim}{X}}$ is integrable,

(II) $f_{\underset{\sim}{X}}(\underset{\sim}{x}) \ge 0 \;\forall\; \underset{\sim}{x} \in \mathbb{R}^k$.

(III) $\int_{\mathbb{R}^k} f_{\underset{\sim}{X}}(\underset{\sim}{x})\,d\underset{\sim}{x} = \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}}\cdots\int_{\mathbb{R}} f_{X_1,X_2\ldots X_k}(x_1,x_2\ldots x_k)\,dx_1\,dx_2\ldots dx_k = 1.$

Independence of K random variables

Defn: $X_1, X_2 \ldots X_k$ are independent if $\forall$ $\underset{\sim}{U} = (U_1, U_2 \ldots U_k) \in \mathbb{R}^k$, we have

$$P(X_1 \le U_1, X_2 \le U_2 \ldots X_k \le U_k) = \prod_{i=1}^{k} P(X_i \le U_i)$$

$$= \cancel{P}\, F_{\underset{\sim}{X}}(\underset{\sim}{U}) = \prod_{i=1}^{k} F_{X_i}(U_i)$$

Thm D Suppose $\underline{X} = (X_1, X_2, \dots X_k)$ is a discrete random vector. Then $X_1 \dots X_k$ are independent iff

$$P_{X_1, X_2 \dots X_k}(x_1, x_2, \dots x_k) = P_{X_1}(x_1) P_{X_2}(x_2) \dots P_{X_k}(x_k)$$

$$\forall (x_1, x_2, \dots x_k) \in \mathbb{R}^k.$$

Thm C Suppose $X_1, X_2 \dots X_k$ are continous r.v.s with pdfs f_{X_1}, ~~&~~ $f_{X_2} \dots f_{X_k}$ respectively.

Then $X_1, X_2 \dots X_k$ are independent random vectors ~~with joint pdf~~ iff $(X_1, X_2 \dots X_k)$ is a cont random vector with a joint pdf.

$$h(x_1, x_2 \dots x_k) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_k}(x_k), \quad (x_1, x_2, \dots x_k) \in \mathbb{R}^k$$

Remark: Suppose $X_1, X_2 \dots X_4$ are independent r.v.s. Then we can show:

(i) $X_1 + X_2 \perp\!\!\!\perp X_3 + X_4$

(ii) e^{X_1}, $X_2 X_4$, X_3^2 are independent and so on.

Expectation of Cont r.V

Recall: If X is a discrete r.V, then $E(X)$ was depends as follows. X has finite mean provided.

$\sum_x |x| P_X(x) < \infty$ and, in this case,

$$E(x) = \mu_X = \sum_x x \, P_X(x)$$

__Def^n__: Suppose X is a continous random variable with pdf $f_X(x)$. We say that X has finite mean or expectation or expected value provided $\int_{-\infty}^{\infty} |x| f_X(x)\,dx$ is finite and in this case-

mean/expectation of X is defined as:

$$E(X) = \mu_X = \int_{\mathbb{R}} x f(x)\,dx$$

__Examples__

(1) $X \sim Unif(a,b)$ Here $E(X) = \int_a^b x \frac{1}{b-a}\,dx$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a,b) \\ 0 & \text{if } x \notin (a,b) \end{cases}$$

---

(2) $X \sim Gamma(\alpha, \lambda)$

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & x > 0 \\ 0, & o.w \end{cases}$$

$$E(X) = \frac{1}{\Gamma(\alpha)} \int_{\mathbb{R}} \lambda^\alpha x^\alpha e^{-\lambda x}\,dx$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^\infty \lambda^\alpha x^\alpha e^{-\lambda x}\,dx$$

$$= \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{(\alpha+1)-1} \, dx$$

$$= \frac{\alpha}{\lambda} \underbrace{\int_0^\infty \frac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)} e^{-\lambda x} x^{(\alpha+1)-1} \, dx}_{\text{equals } 1} = \frac{\alpha}{\lambda}$$

$$(\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha))$$

ie, $X \sim \text{Gamma}(\alpha, \lambda) \Rightarrow E(X) = \frac{\alpha}{\lambda}$.

$X \sim \text{Exp}(\lambda) \Rightarrow E(X) = \frac{1}{\lambda}$ (eae).

---

(3) $X \sim \text{Beta}(a, b) \Longrightarrow f_X(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}, \; 0 < x < 1$

$$\Rightarrow E(X) = \int_0^1 x \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} \, dx.$$

$$= \frac{B(a+1, b)}{B(a,b)} \int_0^1 \frac{1}{B(a+1,b)} x^{(a+1)-1} (1-b)^{b-1} \, dx$$

$$= \frac{B(a+1, b)}{B(a,b)}$$

Now,

$$\frac{B(a+1, b)}{B(a,b)} = \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

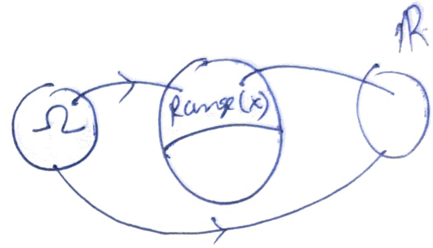
$$= \frac{a}{a+b}$$

- $X \sim Beta(a,b) \Rightarrow E(X) = \frac{a}{a+b}$
- $X \sim Unif(0,1) \Rightarrow E(X) = \frac{1}{2}$

Recall: Suppose X is a discrete r.v and $h: Range(X) \longrightarrow \mathbb{R}$. Then $h(X)$ is a r.v.

This r.v has finite mean provided

$$\sum_{x} |h(x)| \, P_X(x) < \infty$$



and in this case:

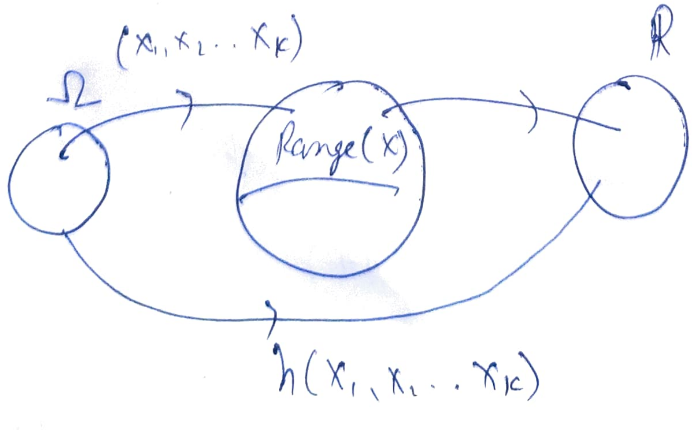
$$E(h(X)) = \sum_{x} h(x) \, P_X(x)$$

This result had a higher dimentional version

We have a discrete random vector $(X_1, X_2 \ldots X_k)$

We also have ~~a~~ a map $h: Range(X_1 \ldots X_k) \rightarrow \mathbb{R}$
$(\subseteq \mathbb{R}^k)$

Then $h(X_1, X_2 \ldots X_k)$ is a r.v.



This r.v has finite mean provided

$$\sum_{\underset{\sim}{x}} |h(\underset{\sim}{x})| \, p_{\underset{\sim}{X}}(x) < \infty \text{ and in this}$$

case, $$E(h(\underset{\sim}{x})) = \sum_{\underset{\sim}{x}} h(\underset{\sim}{x}) \left(p_{\underset{\sim}{X}}(\underset{\sim}{x}) \right) \quad (1)$$

Remark: We used (1) and Fubini's theorm to establish linerity of expectation in the discrete case. Now let us look into the cont case.

Thm: Suppose X is a cont r.v with a pdf $f_X(x)$ ($\Leftrightarrow$ Range $(X) := \{x \in \mathbb{R} : f_X(x) > 0\}$ and $h : \text{Range}(X) \to \mathbb{R}$ is a map such that
- $h(X)$ is also a r.v.
- In the above case, $E(h(X)) = \int_{-\infty}^{\infty} h(x) f_X(x) \, dx$.

Remark: For any r.v X and any map $h : \text{Range}(X) \to \mathbb{R}$ that has finitely many discontinuities, $h(X)$ can be shown to be a random variable.

Thm: Suppose $\underset{\sim}{X} = (X_1, X_2 \ldots X_k)$ is a continous random vector with a joint pdf $f_{\underset{\sim}{X}}(x)$

$(\Rightarrow$ (Range $(\underset{\sim}{X}) = \{f_{\underset{\sim}{X}}(x) > 0\}$) and

$h :$ Range $(\underset{\sim}{X})$ $(\subseteq \mathbb{R}^k)$ such that $h(\underset{\sim}{X})$ is also r.v.

• $h(\underset{\sim}{X})$ has finite mean provided $\int_{\mathbb{R}^k} |h_{\underset{\sim}{x}}(\underset{\sim}{x})| f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d(\underset{\sim}{x}) < \infty$

• In the above Case,

$$E\left[h(\underset{\sim}{X})\right] = \int_{\mathbb{R}^k} h(\underset{\sim}{x}) f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d\underset{\sim}{x}$$

<u>Fubinis Thm</u>: Suppose $g: \mathbb{R}^2 \to \mathbb{R}$ is a function then the followings holds:

① If $g(x,y) \geq 0 \ \forall \ (x,y) \in \mathbb{R}^2$, then

$$(*) \quad \cdots \quad \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)\, dx\, dy = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)\, dy\, dx$$

(both integrals are allowed to be $+\infty$).

② If either $\int_{\mathbb{R}^2} |g(x,y)|\, dx\, dy < \infty$ or $\int_{\mathbb{R}}\int_{\mathbb{R}} |g(x,y)|\, dy\, dx < \infty$

then also $(*)$ holds.

<u>Thm</u>: Suppose $X_1, X_2 \ldots X_k$ are r.v.s (defined on the same sample space) with finite mean. Then for any $a_1, a_2 \ldots a_k \in \mathbb{R}$, the linear combination

$\sum_{i=1}^{k} \alpha_i X_i$ has finite mean and

$$E\left(\sum_{i=1}^{k} \alpha_i X_i\right) = \sum_{i=1}^{k} \alpha_i E(X_i)$$

Remark: We have proved this result when $(X_1, X_2 \ldots X_k)$ is a discrete random vector.

<u>Exc</u>: Prove this theorm (linerity of expectation) when $(X_1, X_2 \ldots X_k)$ is a cont random vector.

<u>Cor 1</u>: If $(X_1, X_2 \ldots X_k)$ is a cont random vector and each X_i has finite mean, that $\sum_{i=1}^{k} X_i$ has finite mean and $E\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} E(X_i)$.

<u>Cor 2</u>:- If (X, Y) is a cont random vector. such that $X \leq Y$ and X, Y has finite mean, then $E(X) \leq E(Y)$.

Proof: Just use $E(Y-X) \geq 0$.

Cor: If X has finite mean, then so does $aX+b$ for any $a, b \in \mathbb{R}$. In this case $E(aX+b) = aE(X)+b$.

Ex: Suppose $X \sim N(\mu, \sigma^2)$. We want to compute $E(X)$.

Recall: $X \sim N(\mu, \sigma^2) \Rightarrow Z := \frac{X-\mu}{\sigma} \sim N(0,1)$

$$X = \mu + \sigma Z,$$

$$\xrightarrow{\text{Cor 3}} E(X) = \mu + \sigma E(Z).$$

Therefore it is enough to compute $E(Z)$, where $Z \sim N(0,1)$. A pdf of Z is –

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in \mathbb{R}.$$

Z has finite mean provided $\int_{-\infty}^{\infty} |z| \varphi(z)\, dz < \infty$

Now,

$$\int_{-\infty}^{\infty} |z| \varphi(z)\, dz = \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$$= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} z e^{-z^2/2}\, dz$$

Put $z^2/2 = u$

$z\,dz = du$

$$\frac{2}{\sqrt{2\pi}}\int_0^\infty e^{-u}\,du = \frac{2}{\sqrt{2\pi}} < \infty.$$

Hence ~~z~~ ~~to~~ Z has finite mean.

$$\cancel{E(Z) = \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = 0}$$

(Since $\int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is finite).

$$\Rightarrow E(X) = \mu + \sigma E(Z) = \mu.$$

$$X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu.$$

Variance: Suppose X is a r.v with finite mean μ.

Then we say X has finite variance provided $(X-\mu)^2$ has finite mean. And in this case, we define variance to be -

$$\mathrm{Var}(X) = V(X) = E[(X-\mu)^2]$$

$$\mathrm{Var}(X) = \begin{cases} \sum_x (x-\mu)^2 P_X(x) & \text{in the discrete case} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x) & \text{in the cont case.} \end{cases}$$

$(X-\mu)^2 = X^2 + \mu^2 - 2\mu X$

Assume that X has finite 2^{nd} moment,

i.e, X^2 has finite mean.

Then $\operatorname{Var}(X) = E[(X-\mu)^2]$

$= E(X^2) - 2\mu E(X) + \mu^2$

$= E(X^2) - \mu^2 = E(X^2) - (E(X))^2$

Also, $X^2 = (X-\mu)^2 + 2\mu X - \mu^2$

if X has finite variance then X^2 has finite mean.

<u>Thm</u>: Suppose X is a r.v with finite mean. Then X has finite variance if and only if X has finite 2^{nd} moment, and in this case,

$\operatorname{Var}(X) = E(X^2) - (E(X))^2$.

- Computation of Variance for various distributions.
- Some shortcuts for computing Variance

$\Longleftrightarrow$ Covariance.

(X, Y) has joint pdf $f_{X,Y}$

Shall compute pdfs of $X+Y$, X/Y, etc.

---

(1) ind $\left.\begin{array}{l} X_1 \sim \text{Gamma}(\alpha_1, \lambda) \\ X_2 \sim \text{Gamma}(\alpha_2, \lambda) \end{array}\right\} \Rightarrow X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \lambda)$

(2) If $X_1, X_2 \dots X_K$ are ind and each $X_i \sim \text{Gamma}(\alpha_i, \lambda)$

then $\sum_{i=1}^{K} X_i \sim \text{Gamma}\left(\sum_{i=1}^{K} \alpha_i, \lambda\right)$

(3) If $X_1, X_2 \dots X_K$ are ind and ~~and~~ identically distributed with each $X_i \sim \text{Exp}(\lambda)$, then

$\sum_{i=1}^{K} X_i \sim \text{Gamma}(K, \lambda)$.

$\boxed{\text{If } X \sim N(0,1) \text{ then } X^2 \sim \text{Gamma}\left(\tfrac{1}{2}, \tfrac{1}{2}\right).}$ already proved

Now suppose $X_1, X_2 \dots X_K \overset{iid}{\sim} N(0,1)$

then $\sum_{i=1}^{K} X_i^2 \sim \text{Gamma}\left(\frac{K}{2}, \frac{1}{2}\right)$

Chi-squared distribution with K degrees of freedom.

A comment on Linearity of Expectation 4/3/24

Thm 1:

$X_1, X_2 \ldots X_n$ has finite mean

$\Rightarrow \sum_{i=1}^{k} \alpha_i X_i$ also has finite mean $\forall \alpha_1, \alpha_2 \ldots \alpha_k \in \mathbb{R}$

and $E\left(\sum_{i=1}^{k} \alpha_i X_i\right) = \sum_{i=1}^{k} \alpha_i E(X_i)$

Remark: Thm 1 holds for all random vectors $(X_1, X_2 \ldots X_k)$ not necessarily discrete or continous.

However, we are able to prove only in two cases (i) when all X_i's are discrete.

(ii) when $(X_1, X_2 \ldots X_k)$ are cont r.v.

In particular, for any r.v X with finite mean, $aX+b$ also has finite mean for each $a, b \in \mathbb{R}$ and $E(aX+b) = aE(X) + b$.

Therefore monotonicity of expectation also holds for all kind of r.v.s.

Cor: If X, Y have finite mean then if $X \leq Y \Rightarrow E(X) \leq E(Y)$.

X has finite mean $\iff \begin{cases} \sum_x |x| P_X(x) < \infty & : \text{Discrete Case} \\ \int_{\mathbb{R}} |x| f_X(x) < \infty & : \text{Continous Case.} \end{cases}$

On the other hand, $|X|$ has finite mean iff

$$\sum_x |x| P_X(x) < \infty \quad \cancel{\text{and}} \quad - \text{Discrete Case}$$

$$\int_{\mathbb{R}} |x| f_X(x) dx < \infty \quad - \text{Continous Case.}$$

Therefore in both discrete and Continous cases,

X has finite mean $\iff$ $|X|$ has finite mean.

Suppose X is an non-negative ~~integer~~ integer valued r.v (ofcu discrete). This means $\text{Range}(X) \subseteq \mathbb{N} \cup \{0\}$

Such a random variable X has finite mean iff

$$\sum_{i=0}^{\infty} P(X > i) < \infty \text{ and in this Case}$$

$$E(X) = \sum_{i=0}^{\infty} P(X > i).$$

Thm: Suppose X is any non-negative r.v. Then X has finite mean if and only if, $\int_0^\infty P(X > u) dn$ is finite and in this Case $E(X) = \int_0^\infty P(X > u) du$.

Remark: This theorm is true for all r.v.s X, not necessarily discrete or continous However we will be able to show it only in the integer valued discrete and continous cases.

Suppose $\text{Range}(X) \subseteq \mathbb{N} \cup \{0\}$

Then,

$$P(X>u) = \begin{cases} P(X>0) & \text{if } 0 < u < 1 \\ P(X>1) & 1 \leq u < 2 \\ P(X>2) & 2 \leq u < 3 \\ \vdots & \vdots \end{cases}$$

Therefore, $\int_0^\infty P(X>u)\, dx$

$$= P(X>0) + P(X>1) + P(X>2) \ldots$$

$$= \sum_{i=0}^{\infty} P(X>i)$$

Proof of the theorm in the cont Case

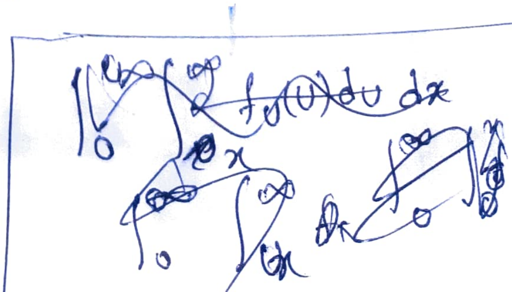
Since $X \geq 0$ with pdf f_X (say).

X has finite mean iff $\int_0^\infty x f_X(x)\, dx < \infty$, and in this case,

$$E(X) = \int_0^\infty x f_X(x)\, dx$$

Now,

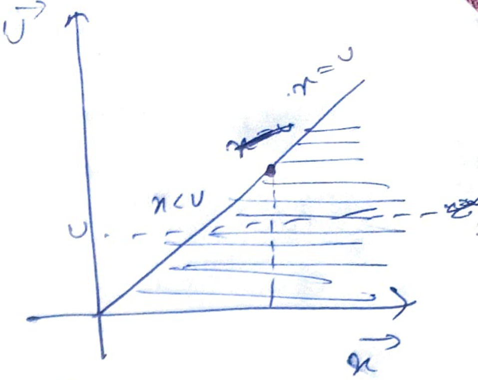
$$\int_0^\infty x f_X(x) = \int_0^\infty \int_0^x f_X(x)\, du\, dx$$



Fubini thm

$$\int_0^\infty \int_u^\infty f_X(x)\,dx\,du$$

$$= \int_0^\infty P(X>u)\,du$$



$\{(x,u)\in\mathbb{R}^2;\ 0<x<\infty,\ 0<u<x\}$
$=\{(x,u)\in\mathbb{R}^2;\ 0<u<\infty,\ u<x<\infty\}$

Therefore, X has finite mean

$$\Leftrightarrow \int_0^\infty x f_X(x)\,dx < \infty \Leftrightarrow \int_0^\infty P(X>u)\,du < \infty$$

and in this case, $E(X) = \int_0^\infty x f_X(x)\,dx = \int_0^\infty P(X>u)\,du$.

Exc: ① Using theorm above compute $E(X)$ when $X \sim Geo(p)$. $[\cancel{E(X)}\ E(X) = \sum_{i=0}^{\infty} P(X>i)]$.

② Also compute $E(Y)$ when $Y\sim Exp(\lambda)$ using the thm.

Cor: Suppose two r.v.s X and Y satisfy $0 \le X \le Y$. Then if Y has finite mean then X also has finite mean and $E(X) \le E(Y)$

Remark: In the previous result on the monotonicity of expectation we had to assume that both X

and Y have finite mean to conclude $E(X) \le E(Y)$.

In Cor1, however we just need to assume that the bigger one has finite mean as long as the r.v.s are non-negative.

<u>Cor2:</u> If X and Y are two random variables such that $|X| \le |Y|$ and Y has finite mean then X also has finite mean.

<u>Pf:</u> Cor1 $\Rightarrow$ Cor2 Y has finite mean $\Leftrightarrow$ $|Y|$ has finite mean. by cor1: $|X|$ has finite mean since $0 \le |X| \le |Y|$.

finite mean $|X| \Leftrightarrow$ X has finite mean.

<u>Proof of</u> Thm (Prev thm) $\Rightarrow$ Cor1

To show X has finite mean –

By thm, this is same as showing

$$\int_0^\infty P(X>u)\,du < \infty.$$

Now,

$$\int_0^\infty P(X>u)\,du.$$

$0 \le X \le Y \Rightarrow \forall\ u>0$

$(X>u) \subseteq (Y>u)$

$\Rightarrow \forall\, u > 0, \quad P(X > u) \leq P(Y < u)$

Hence,

$$\int_0^\infty P(X > u)\, du \leq \int_0^\infty P(Y > u)\, du < \infty.$$

[By the thm as Y has finite mean]

$$\Rightarrow \int_0^\infty P(X > u)\, dU < \infty \xrightarrow{\text{Thm}} X \text{ has finite mean.}$$

Hence $E(X) \stackrel{\text{Thm}}{=} \int_0^\infty P(X > u)\, du$

$$\leq \int_0^\infty P(Y > u)\, du.$$

$$\stackrel{\text{Thm}}{=} E(Y).$$

$$\Rightarrow E(X) \leq E(Y).$$

Application of Cor 2 :

Fact: If X has finite second moment then ~~X~~ it has finite mean.

Note that $|X| \leq \frac{1 + X^2}{2}$

Hence, X has finite second moment $\Rightarrow$

$\frac{1+X^2}{2}$ has finite mean $\Rightarrow$ X has finite mean by Cor 2.

Def^n: We say that X has finite m^{th} moment ($m \in \mathbb{N}$) if X^m has finite mean, and in this case $E(X^m)$ is called the m^{th} moment of X.

- Exc: If $m, n \in \mathbb{N}$, $m<n$ and X has n^{th} moment then show that X has finite m^{th} moment.

Variance:

Def^n: Suppose X is a r.v with finite mean μ. We say that X has finite variance if $(X-\mu)^2$ has finite mean and in this case, we define the variance of X as

$$Var(X) = V(X) = E[(X-\mu)^2].$$

Remark: (1) Variance is extremly outlier-sensitive.

(2) The unit of variance is the square of unit of X.

(3) Typically, with higher values of E(X), the values of Var(X) tends to be higher.

For example $Y = 2X$, then $Var(Y) = 4\,Var(X)$

(4) In many cases instead of variance, people use coefficient of variation.

$$\boxed{C.V(X) = \frac{+\sqrt{Var(X)}}{E(X)}}, \quad \left(\text{whenever } E(X) \neq 0\right)$$

as a measure of dispersion/spread of r.v X.

(5) Chebyshev's inequality

For all $a > 0$, $P(|X-\mu| \geq a\sigma)$

$\sigma_X := +\sqrt{Var(X)}$
= standard deviation of X

$$P[|X-\mu| \geq a\sigma] \leq \frac{1}{a^2}$$

$$\Rightarrow P[|X-\mu| < a\sigma] \geq 1 - \frac{1}{a^2}.$$

$$\Rightarrow P[\mu - a\sigma < X < \mu + a\sigma] \geq 1 - \frac{1}{a^2}.$$

$\Rightarrow$ X lies between $\mu \pm a\sigma$ with probability atleast $1 - \frac{1}{a^2}$.

($a = 3 \Rightarrow X$ lies between $\mu - 3\sigma$ and $\mu + 3\sigma$ with prob at least $\frac{8}{9} = 88.89\%$)

Thm: Suppose X is a random variable with finite mean. Then X has finite variance if and only if X has finite 2nd moment. And in this case,

$$Var(X) = E(X^2) - (E(X))^2.$$

Remark: In order to check if X has finite variance, it is enough to check if X has finite 2nd moment.

Proof: $\mu = E(X)$.

If part: $(X-\mu)^2 = X^2 - 2\mu X + \mu^2$.

therefore X has finite second moment.

$\Rightarrow (X-\mu)^2$ has finite mean $\Rightarrow X$ has finite variance.

And in this case

$$
\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - \cancel{2\mu E(X)}\ \mu^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

Only if part:

$$X^2 = (X-\mu)^2 + 2\mu X - \mu^2$$

X has finite ~~mean~~ variance $\Rightarrow (X-\mu)^2$ has finite mean.

$\Rightarrow X^2$ also has finite mean.

$\Rightarrow X$ has finite second moment.

Cor: If X has finite second moment, then X has finite variance, which is given by

$$Var(X) = E(X^2) - (E(X))^2$$

Proof: This follows from Thm + the fact that X has finite second moment $\Rightarrow$ X has finite mean

---

Cor: For a r.v X with finite 2nd moment,

$$E(X^2) \geq (E(X))^2$$

Proof: $E(X^2) - (E(X))^2 = Var(X) = E[(X-\mu)^2] \geq 0$

$$\Rightarrow E(X^2) \geq (E(X))^2$$

Remark: Equality holds in the last Corollary if and only if $Var(X) = E[(X-\mu)^2] = 0$, which can be shown to be equivalent to $P(X=\mu) = 1$.

Exc: (1) Suppose X is a discrete rv with finite 2^{nd} moment. Then show that $Var(X)=0$ iff X is a degenerate r.v.

Exc: (2) Suppose X is a cont r.v with a cont pdf and finite 2^{nd} moment then show that $Var(X) > 0$.

$$\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx$$

4/3/24

Computation of Variance: $Var(X) = E(x^2) - (E(x))^2$

(1) $X \sim Unif\{x_1, x_2 \dots x_n\}$

$$E(X) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \overline{X}.$$

$$E(X^2) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right)$$

$$\therefore Var(X) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - (\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - \bar{x}^2)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right) = E[(X-\bar{x})^2]$$

Just expand to check!

In particular, if $X \sim Unif\{1,2 \ldots n\}$, then

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n} i^2 - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \underline{\text{Check}} \; \frac{n^2-1}{12}$$

---

2. $X \sim Poi(\lambda) \Rightarrow P_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2 \cdots$

$E(X) = \lambda.$

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1)\, e^{-\lambda}\frac{\lambda^x}{x!}$$

$$= \sum_{x=2}^{\infty} e^{-\lambda}\frac{\lambda^x}{(x-2)!}$$

$$= \lambda^2 \sum_{x=2}^{\infty} e^{-\lambda}\frac{\lambda^{x-2}}{(x-2)!}$$

$$= \lambda^2 \sum_{y=0}^{\infty} e^{-\lambda}\frac{\lambda^y}{y!} = \lambda^2.$$

$$\Rightarrow E(X^2 - X) = \lambda^2 \Rightarrow E(X^2) = \lambda^2 + E(X)$$
$$= \lambda^2 + \lambda.$$

Exc: Show that if $X \sim Geo(p)$, then

$$Var(X) = \frac{1-p}{p^2}$$

3. $X \sim Unif(a,b)$

$$E(X) = \frac{a+b}{2}$$

$$E(X^2) = \int_a^b \frac{x^2}{b-a}\, dx = \frac{b^2+ab+a^2}{3}$$

$$Var(X) = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4}$$

check.

$$= \frac{(b-a)^2}{12}.$$

In part, if $X \sim Unif(0,1)$ then $Var(X) = \frac{1}{12}$.

---

4. $X \sim Gamma(\alpha, \lambda)$ $\qquad E(X) = \frac{\alpha}{\lambda}$.

$$E(X^2) = \int_0^\infty x^2 \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1}\, dx$$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\lambda^2} \int_0^\infty \frac{\lambda^{\alpha+2} e^{-\lambda x} x^{(\alpha+2)-1}}{\Gamma(\alpha+2)}$$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\lambda^2}$$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{1}{\lambda^2}$$ $\Gamma(\alpha+2) = (\alpha+1)\Gamma(\alpha+1)$

$= (\alpha+1)\alpha\,\Gamma(\alpha)$

$$= \frac{(\alpha+1)\alpha}{\lambda^2} = \frac{\alpha^2}{\lambda^2} + \frac{\alpha}{\lambda^2}.$$

$$\therefore Var(X) = E(X^2) - (E(X))^2$$
$$= \frac{\alpha}{\lambda^2}.$$

In particular ~~if~~ $X \sim Exp(\lambda) \Rightarrow Var(X) = \frac{1}{\lambda^2}$.

Exc: find ~~variance~~ ~~of~~ $Var(X)$ if $X \sim Beta(a, b)$.

Exc: $X \sim N(\mu, \sigma^2)$, $Z := \frac{X-\mu}{\sigma} \sim N(0,1)$

$X = \mu + \sigma Z$ when $Z \sim N(0,1)$.

Propn: If Z has finite variance, then for each $a, b \in \mathbb{R}$, $a + bZ$ also finite variance and

$$Var(a+bZ) = b^2 Var(Z). \text{ (Check)}$$

Using the propon' we get $Var(X) = \sigma^2 Var(Z)$.

Z has pdf $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

$$E(Z) = 0 \Rightarrow Var(Z) = E(Z^2),$$

$$E(Z^2) = \int_{-\infty}^{\infty} \cancel{\varnothing} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} Z^2 e^{-z^2/2} dZ.$$

(put $z^2/2 = u$)

($z dz = du$)

($u = \frac{z^2}{2}$)

$Z = \sqrt{2}\, u^{1/2}$

$$\cancel{\Rightarrow \frac{4}{\sqrt{2\pi}} \int_0^{\infty} \sqrt{\ } e^{-u} du}$$

$$= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \sqrt{2}\, u^{1/2} e^{-u} du.$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\infty} \cancel{u^{3/2}}\, u^{3/2-1} e^{-u} du$$

$$= \frac{2}{\sqrt{\pi}} \cancel{\beta} \Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}} \frac{1}{2} \Gamma\left(\frac{1}{2}\right)$$

$$= \cancel{2}\, 1.$$

$\left(\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right)$

$\therefore \mathrm{Var}(Z) = 1.$

Therefore $X \sim N(\mu, \sigma^2)$ $\Rightarrow$ $\begin{cases} E(X) = \mu \\ \mathrm{Var}(X) = \sigma^2 \end{cases}$

$\sigma > 0.$

$\Rightarrow \sigma = +\sqrt{Var(X)}$

~~Sutan~~ standard Deviation $= SD(X)$

Proof of Prop^n: Fix $a, b \in \mathbb{R}$.

Z has finite mean $\Rightarrow$ $a+bZ$ also has finite mean.

And $E(a+bZ) = a + b\cancel{E}(Z)$

$= \cancel{a+b\mu(Z)}\ a + b\mu_Z$

Z has finite variance $\Rightarrow (Z-\mu_Z)^2$ has finite mean.

To show: $((a+bZ)^{\cancel{2}} - E(a+bZ))^2$ has finite mean.

$(a+bZ) - E(a+bZ) = (a+bZ) - (a+b\mu_Z)$

$= b(Z-\mu_Z)$

$\Rightarrow ((a+bZ) - E(a+bZ))^2 = b^2(Z-\mu_Z)^2$.

Since Z has finite variance, it follows that $(a+bZ - E(a+bZ))^2 = b^2(Z-\mu_Z)^2$

$a+bZ$ has ~~finite mean~~ variance

$$Var(a+bZ) = E\left[(a+bZ) - E(a+bZ)^2\right]$$

$$= E\left[b^2(Z-\mu_Z)^2\right]$$

$$\Rightarrow Var(a+bZ) = b^2 E\left[(Z-\mu_Z)^2\right] = b^2 Var(Z)$$

This completes the proof of propⁿ.

Covariance

Suppose (X, Y) is a random vector, and we would like to measure the amount of association between X and Y. The most basic such measure is covariance.

Defⁿ: Suppose X and Y are jointly distributed r.v.s with finite means μ_X and μ_Y respectively. Then we say that X and Y have finite covariance if $(X-\mu_X)(Y-\mu_Y)$ has finite mean and in this case, we define the covariance of X and Y as follows -

$$Cov(X, Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

Remark: $Cov(X, X) = Var(X)$ whenever it is finite.

Remarks: ① If $X \equiv Y$, then

$$Cov(X,Y) = Cov(X,X)$$
$$= Var(X).$$

② Interpretation of Covariance

X and Y are positively associated (i.e, if a lower value of X tends to give rise to lower value of Y and a higher value of X tends to give rise to a higher value of Y), then $Cov(X,Y)$ is going to be positive. This is because whenever $X > \mu_X \Rightarrow Y > \mu_Y$ with high probability. Similarly whenever $X < \mu_X$ we will have $Y < \mu_Y$ with high probablity. Therefore the product $\cancel{(X-\mu_X)(Y-\mu_Y)}$ $(X-\mu_X)(Y-\mu_Y)$ will be strictly positive with high probablity making $Cov(X,Y) > 0$.

(b) If, on the other hand, X and Y are negatively associated then $(X-\mu_X)(Y-\mu_Y)$ will be negative with high prob making $Cov(X,Y) < 0$.

~~in this case we define~~

(c) Note that ⓐ and ⓑ give the interpretation of the sign of the covariance. The interpretation of the absolute value of Covariance is tricky. Covariance ~~of~~ gets affected by scaling. For example, $Cov(2X, 3Y) = 6\,Cov(X, Y)$.

Also Covarience is not Unit free.

③ In order to adress the problems of Covarience mentioned in Remark ② ⓒ, Covarience is divided by the product of the standard deviations of X and Y. This gives rise to the correlation coefficient of X and Y defined as follows.

$$Corr(X, Y) = \rho(X, Y) = \frac{Cov(X, Y)}{+\sqrt{Var(X)\,Var(Y)}}$$

(Provided $Var(X)\,Var(Y) \neq 0$)

This measure of association is unit free and always lies in $[-1, 1]$ making its value (not just the sign) easier to interpret.

Thm: Suppose X and Y are jointly distributed random variables with finite means μ_X and μ_Y respectively. Then X and Y have finite covariance iff XY has finite mean. And in this case

$$Cov(X,Y) = E(XY) - \mu_X \mu_Y$$
$$= E(XY) - E(X)E(Y)$$

Proof: Exc.

Remark: When $X \equiv Y$, this thm will boil down to a previous theorem on variance.

Exc: Suppose $X \sim Unif\{-1, 0, 1\}$, $Y := X^2$. Show $Cov(X,Y)$ is finite and zero even though X and Y are not independ. ($X \not\perp\!\!\!\perp Y$).

Exc: Suppose $(X,Y) \sim Unif(D)$ where $D = \{(x,y) \in \mathbb{R}^2 : x^2+y^2<1\}$ is the unit disk. Then show that $Cov(X,Y) = 0$ even though ($X \not\perp\!\!\!\perp Y$) ~~$Cov(X,Y)$~~$=0$

Remarks: ① Note that in the above version, X and Y are dependent (and hence they are "associated") even though $Cov(X,Y) = 0$.

② These examples clearly show a very

important drawback of Covariance (as well as Correlation) / dependence.

5/3/24

Thm: Suppose X and Y are jointly distributed r.v.s with finite means. Then X and Y have finite covariance iff XY has finite mean. In this case

$Cov(X,Y) = E(XY) - E(X)E(Y)$.

Cor: If X and Y are jointly distributed with finite second moments, then X and Y have finite covariance and:

$Cov(X,Y) = E(XY) - E(X)E(Y)$.

(All three quantities are finite and exist).

Proof: X has finite 2nd moment $\Rightarrow$ X has finite mean
Y " " " 2nd " $\Rightarrow$ Y " " ".

We have to show that $|XY|$ has finite mean.
This holds due to–

$$|XY| \le \frac{X^2+Y^2}{2}$$

$\Rightarrow |XY|$ has finite mean $\Rightarrow$ XY has finite mean.

by the above theorem X and Y have finite covariance given by–

$$E(XY) - E(X)E(Y).$$

Defn: Two jointly distributed random variables are called **uncorelated** if $Cov(X,Y) = 0$.

$(\iff Cor(X,Y) = 0)$

Thm: If $X \perp\!\!\!\perp Y$ and X, Y have finite means then XY also has finite mean and $E(XY) = E(X)E(Y)$, In particular X and Y are uncorelated.

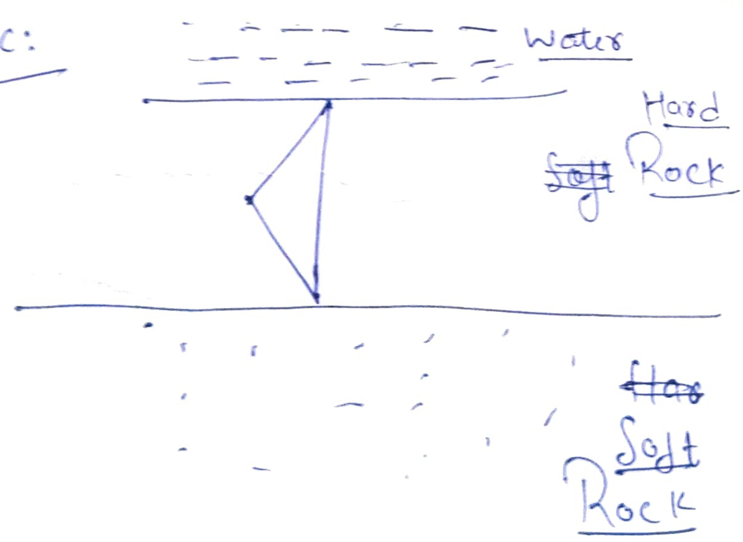
Exc: Prove this thm when (i) both X, Y are discrete.
(ii) " " " Continous.

Remark: ① The statement of this theorm holds for any random variable X and Y, not necessarily discrete or continous.

② X and Y are independent $(X \perp\!\!\!\perp Y) \underset{\not\Leftarrow}{\Longrightarrow}$ $\boxed{X, Y \text{ are uncorrelated}}$

($\Leftarrow$ (may not hold)

Exc:



Drinage Network Model

Each Path is Open with prob $1/2$.

X = No. of open paths

$Y = \begin{cases} 1 & \text{if water can pass through} \\ 0 & \text{o.w} \end{cases}$

Calculate $Cov(X,Y)$

(observe that $Cov(X,Y) > 0$)

Exc: Suppose you have three draws of Polya's Urn Scheme.

$X =$ No. of black balls in the first two draws

$Y =$ " red " last two ".

Compute $Cov(X, Y)$. (Observe $Cov(X,Y) < 0$).

Exc: Suppose (X, Y) is a cont random vector with a joint pdf

$$f_{X,Y}(x,y) = \begin{cases} cxy & \text{if } 0 < x < y < 1 \\ 0 & \text{o.w} \end{cases}$$

Properties of Covariance and Variance

(0) Suppose X is a r.v with finite 2^{nd} moment. Then $Var(X) = Cov(X, X)$.

(1) Suppose X and Y are jointly distributed r.v.s with finite covariance. Then

$$Cov(X,Y) = Cov(Y,X).$$

Symmetry of Covariance.

(2) Suppose X_1, X_2, Y are jointly distributed random variables (with finite means) such that for $i=1$ and 2 X_i and Y have finite Covariance.

Then for any $\alpha_1, \alpha_2 \in \mathbb{R}$ the random variables $\alpha_1 X_1 + \alpha_2 X_2$ and Y have finite covarience and ~~this~~

Case $Cov(\alpha_1 X_1 + \alpha_2 X_2, Y) = \alpha_1 Cov(X_1, Y) + \alpha_2 Cov(X_2, Y)$

(Bilinearity).

<u>Proof</u>: Take $\alpha_1, \alpha_2 \in \mathbb{R}$.

Then $\alpha_1 X_1 + \alpha_2 X_2$ and Y have finite means.

Also $|(\alpha_1 X_1 + \alpha_2 X_2)Y| \overset{①}{\leq} |\alpha_1||X_1 Y| + |\alpha_2||X_2 Y|$

X_1 and Y have finite mean covarience $\Rightarrow |X_1 Y|$ has finite mean.

X_2 and Y have finite covarience $\Rightarrow |X_2 Y|$ has finite mean.

Therefore by linearity of expectation,

$|\alpha_1||X_1 Y| + |\alpha_2||X_2 Y|$ has finite mean

(using the boddness theorm proved earlier)

$\overset{①}{\Longrightarrow} (\alpha_1 X_1 + \alpha_2 X_2)Y$ has finite mean.

$\Longrightarrow \alpha_1 X_1 + \alpha_2 X_2$ and Y have finite covarience.

Hence $Cov(\alpha_1 X_1 + \alpha_2 X_2, Y)$

$= E[(\alpha_1 X_1 + \alpha_2 X_2) Y] - E(\alpha X_1 + \alpha_2 X_2) E(Y)$

$= \alpha_1 E((X_1)Y) + \alpha_2 E(X_2 Y) - \alpha_1 E(X_1) E(Y)$

$- \alpha_2 E(X_2) E(Y)$.

$$= \alpha_1 \left[E(X_1 Y) - E(X_1)E(Y) \right] + \alpha_2 \left[E(X_2 Y) - E(X_2)E(Y) \right]$$

$$= \alpha_1 \operatorname{Cov}(X_1, Y) + \alpha_2 \operatorname{Cov}(X_2, Y).$$

(3) Suppose X, Y_1, Y_2 are jointly distributed random variables such that for $j = 1,2$ ~~and~~ X and Y_j have finite covariance. Then for any β_1 and β_2 $\in \mathbb{R}$, the r.v.s X and $\beta_1 Y_1 + \beta_2 Y_2$ have finite covariance and this covariance between X and $\operatorname{Cov}(X, \beta_1 Y_1 + \beta_2 Y_2) = \beta_1 \operatorname{Cov}(X, Y_1) + \beta_2 \operatorname{Cov}(X, Y_2)$.

Proof: (1) + (2) $\Rightarrow$ (3).

(4) <u>Bilinearity of Covariance</u>

Suppose $X_1, X_2 \dots X_m, Y_1, Y_2 \dots Y_n$ are jointly distributed random variables (with finite means) such that X_i and Y_j have finite covariance for each pair $(i,j) \in \{1 \dots m\} \times \{1, 2 \dots n\}$.

then for any $\alpha_1, \alpha_2 \dots \alpha_m, \beta_1, \beta_2 \dots \beta_n \in \mathbb{R}$ the random variables.

$\sum \alpha_i X_i$ and $\sum \beta_j Y_j$ have finite covariance and,

$$\operatorname{Cov}\left(\sum \alpha_i X_i, \sum \beta_j Y_j \right) = \sum_{i,j} \alpha_i \operatorname{Cov}(X_i, Y_j) \beta_j$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \beta_j \, Cov(X_i, Y_j)$$

Proof: (Sketch)

Use two fold induction on m and n and properties ② and ③.

(a) fix $m \in \mathbb{N}$ and show

$$Cov\left(\sum_{i=1}^{m} \alpha_i X_i, \sum_{i=1}^{n} \beta_j Y_j\right) = \sum_{j=1}^{n} \beta_j \, Cov\left(\sum_{i=1}^{m} \alpha_i X_i, Y_j\right)$$

Using (3) and induction on n.

(b) Then show that for each fixed $j \in \{1, 2, \ldots n\}$,

$$Cov\left(\sum_{i=1}^{m} \alpha_i X_i, Y_j\right) = \sum_{i=1}^{m} \alpha_i \, Cov(X_i, Y_j)$$

Using (2) + induction on m.

(c) Show that (a) + (b) $\Longrightarrow$ ~~Result~~ (4).

⑤ Under the assumptions of (4),

$$Cov\left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\right)$$
$$= \sum_{j=1}^{m} \sum_{j=1}^{n} Cov(X_i, Y_j)$$ ~~when~~

(Bi-additivity)

(6) Suppose $X_1, Y_2 \ldots, X_m$ are jointly distributed r.v.s with finite second moments. Then for any $\alpha_1, \alpha_2 \ldots \alpha_m \in \mathbb{R}$, then r.v $\sum_{i=1}^{m} \alpha_i X_i$ has finite variance and
$$Var\left(\sum_{i=1}^{m} \alpha_i X_i\right) = \sum_{i=1}^{m} \alpha_i^2 Var(X_i) + 2\sum_{i=1}^{m}\sum_{\substack{j=1 \\ i<j}}^{m} \alpha_i \alpha_j Cov(X_i, X_j)$$

Remark: Since each X_i has finite second moment, all pairs X_i and X_j have finite covariance.

Proof: Use (4) and $m=n$
$$X_i = Y_i \quad \forall \ 1 \le i \le m = n$$
$$\alpha_i = \beta_i \quad \forall \ 1 \le i \le n = m$$

Then, $Var\left(\sum_{i=1}^{m} \alpha_i X_i\right) = Cov\left(\sum_{i=1}^{m} \alpha_i X_i, \sum_{j=1}^{m} \alpha_j X_j\right)$
$$\overset{(4)}{=} \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j Cov(X_i, X_j).$$
$$= \sum_{i=1}^{m} \alpha_i^2 Var(X_i) + \sum_{j=1}^{m}\sum_{\substack{i=1 \\ i \ne j}}^{m} \alpha_i \alpha_j Cov(X_i, X_j)$$

$$= \sum_{i=1}^{m} \alpha_i^2 \, Var(X_i) + 2 \sum_{\substack{i=1 \\ i<j}}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j \, Cov(X_i, X_j)$$

(7) Under the assumptions of (6),

$$Var\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} Var(X_i) + \sum_{\substack{1 \le i,j \le m \\ i \ne j}} Cov(X_i, X_j)$$

Proof: follows from (6) with $\alpha_i = 1 \ \forall i$.

Defn: Jointly distributed r.v.s $X_1, X_2 \ldots X_m$ are called pairwise uncorelated ~~any~~ if $Cov(X_i, X_j)$ $\forall$ i, j s.t $i \ne j$.

Clearly $X_1, X_2 \ldots X_m$ are independent

$$\Rightarrow X_1, X_2 \ldots X_m \text{ are pairwise uncorrelated.}$$
$$\not\Leftarrow$$

---

(8) Under the assumption of (6)

$$Var\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i^2 \, Var(X_i) \quad \text{———} (*)$$

for any $\alpha_1, \alpha_2 \ldots \alpha_n \in \mathbb{R}$ and $Cov(X_i, X_j) = 0 \ \forall \ i \ne j$.

(ie provided $X_1, X_2 \dots X_n$ are pairwise uncorrelated).

In particular, if $X_1, X_2 \dots X_n$ are independent, then (*) holds.

(9) For pairwise uncorrelated random variables $X_1, X_2 \dots X_n$ satisfying assumptions of (6) we have,

$$\boxed{Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)} \quad \text{———} \quad (**)$$

In particular if $X_1, X_2 \dots X_n$ are independent then (**) holds.

Applications:

Exc: $X \sim Bin(n, p)$
Compute $E(X(X-1))$ and use this to compute $Var(X)$.

ex (1) Suppose $X \sim Bin(n, p)$

Then $X = \sum_{i=1}^{n} X_i$, where $X_1, X_2 \dots X_n \overset{iid}{\sim} Ber(p)$

(ie, $X_1, X_2 \dots X_n$ are independent and identically distributed and each $X_i \sim Ber(p)$)

Recall that for each i,

$$X_i = \begin{cases} 1 & \text{if } i\text{th trial yields } S \\ 0 & \text{if } i\text{th trial yields } F \end{cases}$$

Since the Bernoulli trials are assumed to be independent, it follows that $X_1, X_2 \dots X_n$ are ind.

Also $P(S) = p$ for each trial. Hence each $X_i \sim Ber(p)$.

$$\Rightarrow Var(X) = Var\left(\sum_{i=1}^{n} X_i\right) \overset{Ind}{=} \sum_{i=1}^{n} Var(X_i)$$

for each i,

$$\left.\begin{aligned} E(X_i) &= 1 \cdot p + 0(1-p) = p \\ E(X_i^2) &= 1^2 \cdot p + 0^2(1-p) = p \end{aligned}\right\} \Rightarrow \begin{aligned} &Var(X_i) \\ &= E(X_i^2) - (E(X_i))^2 \\ &= p - p^2 = p(1-p) \end{aligned}$$

$$\Rightarrow Var(X) = \sum_{i=1}^{n} Var(X_i) = np(1-p).$$

ex (2) Suppose $X \sim NB(r, p)$

Then $X = \sum_{i=1}^{n} X_i$ where $X_1, X_2 \dots X_r \overset{iid}{\sim} Geo(p)$

$$\Rightarrow Var(X) = Var\left(\sum_{i=1}^{r} X_i\right) \overset{ind}{=} \sum_{i=1}^{n} Var(X) = r\left(\frac{1-p}{p^2}\right)$$

ex (3) Maxwell-Boltzman Statistics.

Suppose ($r \in \mathbb{N}$) distinguishable particles are placed at random in $n (\in \mathbb{N})$ distinct cells whose capacities are unlimited. Let N_0 be the number of empty cells. Compute $E(N_0)$ and $Var(N_0)$.

Solution: Let us name the cells as Cell 1, Cell 2, ...
... Cell n. Let, for each $j \in \{1, 2 \dots n\}$,

$$I_j = \begin{cases} 1 & \text{if Cell } j \text{ is empty} \\ 0 & \text{if Cell } j \text{ is not empty.} \end{cases}$$

be the indicator that Cell j is empty.

Clearly, $N_0 = \sum_{j=1}^{n} I_j$

Also each $I_j \sim Ber(P(I_j = 1))$ or $\left(I_j \sim Ber\left(1 - \frac{1}{n}\right)^{r}\right)$

$$P(I_j = 1) = P(\text{Cell } j \text{ is empty})$$

$$= \frac{(n-1)^r}{n^r} = \left(1 - \frac{1}{n}\right)^r$$

for each $j \in \{1, 2 \dots n\}$, $E(I_j) = \frac{(n-1)^r}{n^r}$

Therefore $E(N_0) = E\left(\sum_{j=1}^{n} I_j\right)$

$$= \sum_{j=1}^{n} E(I_j)$$

$$= n\frac{(n-1)^r}{n^r} = \frac{(n-1)^r}{n^{r-1}}.$$

~~for each $j \in \{1,2,\dots,n\}$~~

Also $Var(N_0) = Var\left(\sum_{j=1}^{n} I_j\right)$

$$\overset{(7)}{=} \sum_{j=1}^{n} Var(I_j) + 2 \sum_{1 \le j < k \le n} Cov(I_j, I_k).$$

for each $j \in \{1,2,\dots,n\}$

$$Var(I_j) = \frac{(n-1)^r}{n^r}\left(1 - \frac{(n-1)^r}{n^r}\right).$$

for each (j,k) with $1 \le j < k \le n$,

$$Cov(I_j, I_k) = E(I_j . I_k) - E(I_j)E(I_k)$$

$= P($Cell j and Cell k are both empty$)$

$$- E(I_j)E(I_k)$$

$$\cancel{= \frac{(n-2)^r}{n^r}}$$

$$= \frac{(n-2)^r}{n^r} - \frac{(n-1)^r}{n^r}\frac{(n-1)^r}{n^r} \cancel{(n-1)^2} = \frac{(n-2)^r}{n^r} - \frac{(n-1)^{2r}}{n^{2r}}$$

(Show $Cov(I_j, I_k) < 0$)

Therefore, $Var(N_0)$

$$= \sum_{j=1}^{n} Var(I_j) + 2 \sum_{1 \le j < k \le n} Cov(I_j, I_k)$$

$$= n \frac{(n-1)^r}{n^r}\left(1 - \frac{(n-1)^r}{n^r}\right) + 2\binom{n}{2}\left[\frac{(n-2)^r}{n^r} - \frac{(n-1)^{2r}}{n^{2r}}\right]$$

$$= \frac{(n-1)^r}{n^{r-1}}\left(1 - \frac{(n-1)^r}{n^r}\right) + n(n-1)\left[\frac{(n-2)^r n^r - (n-1)^{2r}}{n^{2r}}\right]$$

Exc: Recall the sleepy secretary problem with n letters and n envelops.
Let X be the number of letters that are put inside the correct envelope.
Compute $E(X)$ and $Var(X)$.

Exc: Consider a maxwell Boltzman model with $r \geq 1$ particles and $n (\geq 3)$ cells. Let N_0 be the number of empty cells and N_1 be the number of cells with exactly one particle.

Show that the ~~covariance between~~ ~~N_0 and~~

$$Cov(N_0, N_1) = \frac{r(n-1)(n-2)^{r-1}}{n^{r-1}} - \frac{r(n-1)^{2r-1}}{n^{2r-2}}.$$

11/3/24

Scalar Transformation of Random Vectors

Given a random vector (X, Y) also a map

$T: \text{Range}(X, Y) \longrightarrow \mathbb{R}$.

We want to find $Z = T(x, y)$ distribution.

Special Cases:

$$T(x, y) = X \pm Y$$

$$T(x, y) = XY$$

$$T(x, y) = \frac{X}{Y} \quad \text{whenever possible.}$$

$T(x, y) = X + Y$ — Convolution formula.

(X, Y) discrete random vector with joint pmf $P_{X,Y}$ let $Z = X + Y$.

Then for any $z \in \mathbb{R}$ we have —

$$P_Z(z) = P[Z = z] = P[X + Y = z]$$

$$= \sum P[X = x, Y = z - x]$$

$$= \sum_{x \in \text{Range}(X)} P_{X,Y}(x, z - x).$$

Thm: (Discrete Convolution formula)

For a discrete random vector (X, Y), the pmf of $Z := X+Y$ is given by

$$p_Z(z) = \sum_{x \in Range(X,Y)} p_{X,Y}(x, z-x), \quad z \in \mathbb{R}.$$

In particular, if $X \perp\!\!\!\perp Y$, then

$$p_Z(z) = \underbrace{\sum_{x \in Range(X)} p_X(x)\, p_Y(z-x)}_{\text{Discrete Convolution of } p_X \text{ and } p_Y.}, \quad z \in \mathbb{R}$$

Remark: The actual range of the above sum can very well be proper subsets of Range(X). For example, if $X \perp\!\!\!\perp Y$, then instead of (1), we write $p_Z(z) = \sum_{x \in R} p_X(x)\, p_Y(z-x)$ where

$R = \{x \in Range(X) : z - x \in Range(Y)\}$

Exc: ① ind $\left.\begin{array}{l} X \sim Poi(\lambda) \\ Y \sim Poi(\mu) \end{array}\right\} \Rightarrow X+Y \sim Poi(\lambda+\mu)$

(2) ind $\begin{cases} X \sim Bin(m,p) \\ Y \sim Bin(n,p) \end{cases} \Rightarrow X+Y \sim Bin(m+n, p)$

Now, assume that (X,Y) is a continous random vector with a joint pdf $f_{X,Y}$. Then it is an obvious guess that $Z=X+Y$ will be cont r.v with a pdf $f_Z(z) = \int_{\mathbb{R}} f_{X,Y}(x; z-x)\,dx, x \in \mathbb{R}$.

This means, in particular, that $X \perp\!\!\!\perp Y \Rightarrow f_Z(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)\,dx, \ z \in \mathbb{R}$

$f_X * f_Y$ is convolution of f_X and f_Z.

Thm: ((Convolution Formula)

If X and Y are independent cont r.v.s with pdfs f_X and f_Y, respectively, then $Z := X+Y$ is also a cont r.v with a pdf $f_Z = \int_{\mathbb{R}} f_X(x) f(z-x)\,dx$, $z \in \mathbb{R}$

Remark: $f_Z(z) = \int_{\mathbb{R}} \cancel{f_X}(z-y) f_Y(y)\,dy$, $z \in \mathbb{R}$

which can be obtained for each $z \in \mathbb{R}$ by substituting $y = z - x$ in the previous formula.

Proof: We have to show that

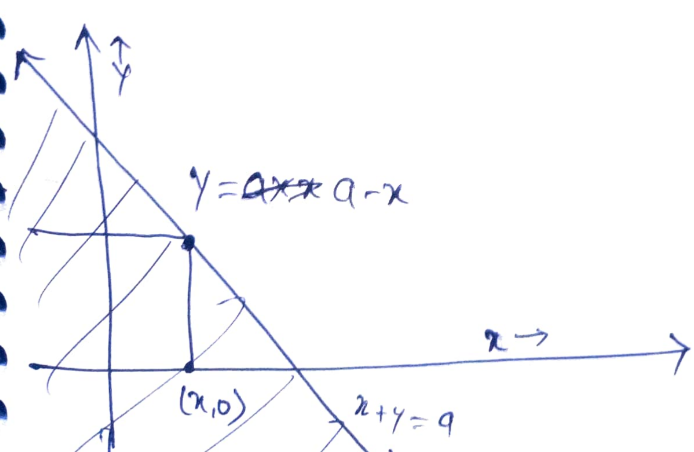
$$g(z) := \int_{\mathbb{R}} f_X(x) f_Y(z-x)\, dx, \quad z \in \mathbb{R}$$

is a pdf of Z. This means that we have to verify that $\forall\, a \in \mathbb{R}$ $P(Z \leq a) = \int_{-\infty}^{a} g(z)\, dz$

$$= \int_{-\infty}^{a} \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx\, dz$$

to this end observe that for each $a \in \mathbb{R}$,

$$P(Z \leq a) = P(X + Y \leq a) \underset{\text{end}}{\overset{(\star)}{=}} \iint_{x+y \leq a} f_X(x) f_Y(y)\, dx\, dy$$



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

In this integral inside put

$z = y + x$

$\Rightarrow y = z - x$

$\Rightarrow dy = dz$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{a} f_X(x)\, f_Y(z-x)\, dz\, dx.$$

$$\underset{\text{fubini's thm}}{=} \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx \right) dz$$

$$= \int_{-\infty}^{a} g(z)\, dz$$. This completes the proof.

Thm: (Generalization of Convolution theorem)

If (X, Y) is a continous random vector with a joint pdf $f_{X,Y}$, then $Z := X + Y$ is a Cont r.v with a pdf

$$f_Z(z) = \int_{\mathbb{R}} f_{X,Y}(x, z-x)\, dx, \quad z \in \mathbb{R}$$

$$= \int_{\mathbb{R}} f_{X,Y}(z-y, y)\,dy, \quad z \in \mathbb{R}.$$

Proof: Exc

Applications of Convolution formula

Proposition:

$$\text{ind}\begin{cases} X \sim Gamma(\alpha, \lambda) \\ Y \sim Gamma(\beta, \lambda) \end{cases} \Rightarrow X+Y = Gamma(\alpha+\beta, \lambda)$$

Proof:

$$Range(X) = Range(Y) = (0, \infty)$$

$$Range(Z) \subseteq (0, \infty)$$

By the convolution formula, Z is a continous r.v with pdf

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx, \quad z > 0.$$

Take $z \in (0,\infty)$, Since Range $(X) = (0,\infty)$ and Range $(Y) = (0,\infty)$, it follows that

$$f_Z(z) = \int_0^z f_X(x)\, f_Y(z-x)\, dx$$

$$= \int_0^z \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \frac{\lambda^\beta}{\Gamma(\beta)} e^{-\lambda(z-x)} (z-x)^{\beta-1} dx$$

$$= \int_0^z \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} x^{\alpha-1} (z-x)^{\beta-1} dx$$

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^z e^{-\lambda z} x^{\alpha-1} (z-x)^{\beta-1} dx$$

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^z x^{\alpha-1} (z-x)^{\beta-1} dx$$

Put $\frac{x}{y} = u$.

$\Rightarrow \frac{x}{z} = u$

$\Rightarrow x = zu \Rightarrow dx = z\, du$

$$\cancel{= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^1 (z-uz)^{\beta-1} z \, du}$$

$$\cancel{= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^1 (1-u)^{\beta-1} z^{\beta} \, du}$$

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} \int_0^1 (zu)^{\alpha-1} (z-zu)^{\beta-1} z \, du$$

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda z} z^{\alpha+\beta-1} \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} \, du$$

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} B(\alpha,\beta) \, e^{-\lambda z} z^{\alpha+\beta-1}$$

Using convolution formula, we get Z has a pdf $f_Z(z) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)/B(\alpha,\beta)} e^{-\lambda z} z^{\alpha+\beta-1}$, $z > 0$

On the other hand Gamma $(\alpha+\beta, \lambda)$ distn has a pdf $g(z) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} e^{-\lambda z} z^{\alpha+\beta-1}$

$z > 0$.

If 2 pdfs are constant multiples of each other then they are same pdfs. & Z is Gamma$(\alpha+\beta, 1)$

Therefore it follows that

① $\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\cancel{\Gamma(\alpha+\beta)}\; B(\alpha,\beta)} = \Gamma(\alpha+\beta) \Rightarrow B(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

② $Z \sim$ Gamma$(\alpha+\beta, 1)$.

This proves the proposition.

<u>Exc</u>: If $X, Y \overset{iid}{\sim}$ Unif$(0,1)$ then find pdf of $Z := X+Y$ using

(a) by computing the cdf of Z and then pdf of Z.

(b) with the help of convolution formula.

<u>Exc</u>: (i) If $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$, then show that $Z_1 + Z_2 \sim N(0,2)$

(ii) If $X_1 \perp\!\!\!\perp X_2$ and $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2$

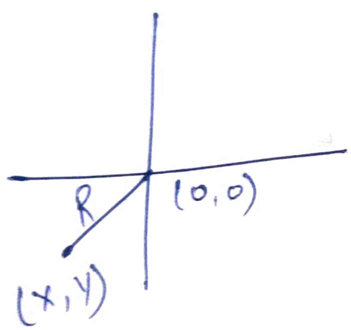
then show that $Z = X_1 + X_2$ is also normal.

and $Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

[Hint: Use (i) to prove (ii)].

<u>Exc:</u> If $X, Y \overset{iid}{\sim} N(0,1)$ then find a pdf of $R = +\sqrt{X^2 + Y^2}$

(Distribution of R is called the rally distribution.)



$X \sim N(0,1) \Rightarrow X^2 \sim Gamma(\frac{1}{2}, \frac{1}{2})$

ind $Y^2 \sim Gamma(\frac{1}{2}, \frac{1}{2})$

$\Rightarrow X^2 + Y^2 \sim Exp(\frac{1}{2})$

<u>Thm:</u> (Additivity of Gamma Distribution): fix $k \geq 2$

Suppose $X_1, X_2 \dots X_k$ are independent random variables such that $X_i \sim Gamma(\alpha_i, \lambda)$ for each $i \in \{1, 2, \dots, k\}$.

$X_1 + X_2 \dots + X_k \sim Gamma(\alpha_1 + \alpha_2 \dots \alpha_k, \lambda)$

Proof: : ~~k=2 Done!~~

We shall prove this using induction on $k \geq 2$.

For $k=2$, the thm is just a restatement of the prop^n. Now suppose that the theorem holds for $k=n$. Enough to show it holds for $k=m+1$.

Take $X_1, X_2 \dots X_{m+1}$ independent r.v.s such that $X_i \sim Gamma(\alpha_i, \lambda)$ for each $i \in \{1, 2 \dots m+1\}$

To show, $\sum_{i=1}^{m+1} X_i \sim Gamma\left(\sum_{i=1}^{m+1} \alpha_i, \lambda\right)$

~~Note this crucial step.~~

$$\boxed{X_1, X_2 \dots X_{m+1} \text{ are independent} \Rightarrow \sum_{i=1}^{m} X_i \perp\!\!\!\perp X_{m+1}}$$

Also, by induction hypothesis

$$\sum_{i=1}^{m} X_i \sim Gamma\left(\sum_{i=1}^{m} X_i, \lambda\right)$$

ind

$$X_{m+1} \sim Gamma(\alpha_{m+1}, \lambda)$$

$\xrightarrow{\text{Prop}^n} \sum_{i=1}^{m+1} X_i \sim \text{Gamma}\left(\sum_{i=1}^{m+1} \alpha_i, \lambda\right)$

Cor 1:

If $X_1, X_2 \dots X_K \overset{iid}{\sim} \text{Exp}(\lambda)$ then we have

$X_1 + X_2 \dots X_K \sim$ Gamma(K, λ).

Proof: Apply the theorm with $\alpha_i = 1 \ \forall i$

Cor 2:

If $Z_1, Z_2 \dots Z_K \overset{iid}{\sim} N(0,1)$, then $Z_1^2 + Z_2^2 \dots Z_K^2 \sim \text{Gamma}\left(\frac{K}{2}, \frac{1}{2}\right)$

Proof:

$Z_1, Z_2 \dots Z_K \overset{iid}{\sim} N(0,1)$

$\Rightarrow Z_1^2 + Z_2^2 \dots Z_K^2 \overset{iid}{\sim} \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$

[Proved in class]

$\xrightarrow{\text{Thm}} Z_1^2 + Z_2^2 + \dots Z_K^2 \sim \text{Gamma}\left(\frac{K}{2}, \frac{1}{2}\right)$

<u>Def^n</u>: Gamma$\left(\frac{k}{2},\frac{1}{2}\right)$ distribution is ~~also~~ also called Chi-squared distribution with k degrees of freedom.

<u>Notation</u>: We write $X \sim \chi^2_k$ whenever $X \sim \text{Gamma}\left(\frac{k}{2},\frac{1}{2}\right)$

~~therefore~~

<u>Cor 2</u> can be restated as follows:

$$Z_1 + Z_2 \dots Z_k \overset{iid}{\sim} N(0,1) \Rightarrow \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k$$

Chi-squared distributions are very important in statistics. Two other such important distributions are t-distribution and F-distribution.

<u>Student's distribution or t-distribution.</u>

<u>Def^n</u>: Suppose $Z \sim N(0,1)$, $X \sim \chi^2_n$ and $X \perp\!\!\!\perp Z$. Then the r.v $T := \frac{Z}{+\sqrt{X/n}}$

is said to follow t- distribution with n degrees of freedom.

Notation: $T \sim t_n$

Remark: Suppose $X_1, X_2 \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ then the following results can be proved.

(a) $$Z := \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

(b) $$V := \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Not imp now. first remark

(c) $\bar{X} \perp\!\!\!\perp S^2 \Rightarrow Z \perp\!\!\!\perp V$.

from (a), (b) and (c) we have

ind

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\Rightarrow T := \frac{Z}{+\sqrt{V/(n-1)}} = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{+\sqrt{S^2/\sigma^2}}$$

$$= \frac{\sqrt{n}(\bar{X}-\mu)}{S}$$

$$= \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

F-distribution or Snedecor's F-distribution: Or Fisher– Snedecor Distribution.

Defn: Suppose $S_1 \sim \chi^2_{d_1}$, $S_2 \sim \chi^2_{d_2}$ and $S_1 \perp\!\!\!\perp S_2$. Then the r.v ~~that~~

$$W := \frac{S_1/d_1}{S_2/d_2}$$ is said to follow to follow F distribution with d_1 and d_2 degrees of freedom.

Notation: $W \sim F_{d_1, d_2}$

18/4/24

The distribution of ratio of two jointly cont r.v

Suppose (X, Y) is a cont r.v with pdf $f_{X,Y}$. In particular, Y is a cont r.v and hence it satisfies $P(Y=0)=0$. Therefore $Z = \frac{X}{Y}$ is well defined.

Goal: To find distribution of Z.

Thm: If (X, Y) is a cont random vector with a joint pdf $f_{X,Y}(x,y)$, then $Z := \frac{X}{Y}$ is a cont random vector with pdf:

$$f_Z(z) = \int_{-\infty}^{\infty} |y|\, f_{X,Y}(zy, y)\, dy, \quad z \in \mathbb{R}.$$

Cor 1: If X, Y are ind cont random variables with pdfs f_X, f_Y res. then $Z = \frac{X}{Y}$ is also cont r.v with pdf $f_Z(z) = \int_{-\infty}^{\infty} |y|\, f_X(yz)\, f_Y(y)\, dy, \quad z \in \mathbb{R}$.

In particular if Range$(X) \subseteq (0,\infty)$ and Range$(Y) \subseteq (0,\infty)$ then $f_Z(z) = \int_0^{\infty} y\, f_X(yz)\, f_Y(y)\, dy, \quad z \in (0,\infty)$

Cor 2: In the setup above, if Range$(X) \subseteq (0,\infty)$ and Range$(Y) \subseteq (0,\infty)$ then ~~$f_Z(z) = \int_0^{\infty} y\, f_{X,Y}$~~

$$f_Z(z) = \int_0^\infty y\, f_{X,Y}(yz, y)\, dy, \quad z \in (0,\infty).$$

<u>Proof of thm</u>: We need to show that

$\forall\ a \in \mathbb{R}$

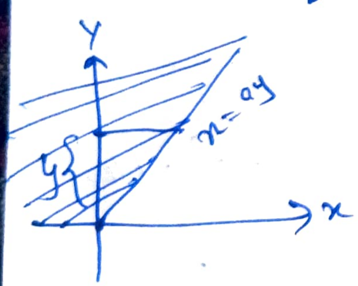
$$(*) \ldots\ P(Z \le a) = \int_{-\infty}^{a} h(z)\, dz.$$

where,

$$h(z) = \int_{-\infty}^{\infty} |y|\, f_{X,Y}(yz, y)\, dy, \quad z \in \mathbb{R}$$

Take $a \in \mathbb{R}$

LHS of $(*)$ $= P[Z \le a]$



$$= P\left[\frac{X}{Y} \le a\right]$$

$$= \underbrace{P\left[\frac{X}{Y} \le a, Y > 0\right]}_{\text{I}} + \underbrace{P\left[\frac{X}{Y} \le a, Y < 0\right]}_{\text{II}}$$

$$= P[X \le aY, Y > 0] + P[X \ge aY, Y < 0]$$

Now I $= P[X \le aY, Y > 0]$

$$= \iint_{\substack{y>0 \\ x \le ay}} f_{X,Y}(x,y)\, dx\, dy.$$

$$= \int_0^{\infty}\int_{-\infty}^{ay} f_{X,Y}(x,y)\,dx\,dy$$

In the inner integral Y is constant and x is the variable

Put $x = yz$ in the inner integral

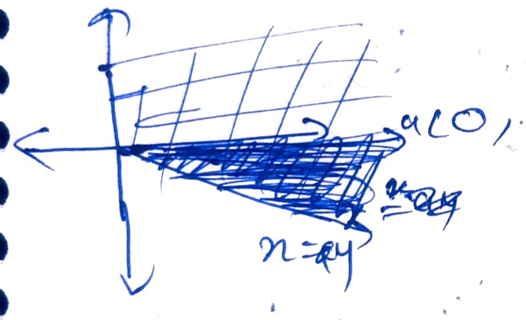
$$\Rightarrow \begin{cases} dx = y\,dz \\ z = \frac{x}{y} \end{cases}$$

$$= \int_0^{\infty}\int_{-\infty}^{a} y f_{X,Y}(yz,y)\,dy$$

$$\Rightarrow \text{(I)} = \int_0^{\infty}\int_{-\infty}^{a} |y| f_{X,Y}(yz,y)\,dz\,dy$$

Similarly

$$\text{(II)} = P[X \geq aY, Y < 0]$$



$$\stackrel{(A)}{=} \iint_{\substack{x \geq ay \\ y<0}} f_{X,Y}(x,y)\,dx\,dy$$

$$= \int_{-\infty}^{0}\int_{ay}^{\infty} f_{X,Y}(x,y)\,dx\,dy$$

Again put $x = yz$ in the inner integral

$$\Rightarrow \begin{cases} dx = y\,dz \\ z = x/y \end{cases}$$

$$\text{(II)} = \int_{-\infty}^{0}\int_{a}^{-\infty} y f_{X,Y}(yz,y)\,dz\,dy$$

$$= \int_{-\infty}^{0}\int_{-\infty}^{a} (-y) f_{X,Y}(yz,y)\,dz\,dy.$$

$$= \int_{-\infty}^{0} \int_{-\infty}^{a} |y| f_{X,Y}(yz, y)\, dz\, dy$$

$\text{I} + \text{II}$

$$\int_{0}^{\infty} \int_{-\infty}^{a} |y|\, f_{X,Y}(zy, y)\, dz\, dy$$

$$+ \int_{-\infty}^{0} \int_{-\infty}^{a} |y| f_{X,Y}(zy, y)\, dz\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{a} \underbrace{|y|}_{\geq 0}\, \underbrace{f_{X,Y}(zy, y)}_{\geq 0}\, dz\, dy$$

Fubini's Thm: $\displaystyle \int_{-\infty}^{a} \int_{-\infty}^{\infty} \underbrace{|y|\, f_{X,Y}(yz, y)\, dy}_{h(z)}\, dz$

$$= \int_{-\infty}^{a} h(z)\, dz$$

$= \text{RHS of } (*)$

This completes the proof.

EX: Suppose X and $Y \overset{iid}{\sim} Exp(\lambda)$

(a) Find the distribution of $Z := \frac{X}{Y}$.

(b) Using (a) or otherwise find the distribution of $U = \frac{X}{X+Y}$

Solution: (a) Note that Range (X) = Range $(Y) = (0, \infty)$ and $X, Y \overset{iid}{\sim} Exp(\lambda)$ ($\Rightarrow X \perp\!\!\!\perp Y$)

Hence (X, Y) has a joint pdf

$$f_{X,Y}(x,y) = \lambda^2 e^{-\lambda(x+y)} \quad \text{if } x>0, y>0.$$

Therefore Using Cor 2.0, we have, Z is cont r.v with pdf

$$f_Z(z) = \int_0^\infty y\, f_{X,Y}(yz, y)\, dy, \quad z>0$$

then we get,

$$f_Z(z) = \int_0^\infty y \lambda^2 e^{-\lambda(yz+y)}\, dy$$

$$= \lambda \int_0^\infty y \lambda e^{-\lambda y(1+z)}\, dy$$

$$= \frac{\lambda}{1+z} \int_0^\infty y \lambda (1+z) e^{-\lambda(1+z)y}\, dy$$

$$= \frac{\lambda}{1+z}\int_0^\infty y\, \underbrace{\lambda(1+z)e^{-\lambda(1+z)y}}_{\text{a pdf of } Exp(\lambda(1+z)) \text{ dist}^n.}\, dy$$

$$= \frac{\lambda}{1+z} E(Y') \quad \text{when } Y' \sim Exp(\lambda(1+z))$$

$$= \frac{\lambda}{1+z} \frac{1}{\lambda(1+z)} = \frac{1}{(1+z)^2}, \; z \in (0,\infty)$$

Therefore Z is cont r.v with a pdf

$$f_Z(z) = \frac{1}{(1+z)^2} \text{ if } z>0.$$

(b) Range $(X,Y) = (0,\infty) \times (0,\infty)$

$\Rightarrow$ Range$(U) \subseteq (0,1)$

Also $U = \frac{X}{X+Y} = \frac{X/Y}{1+X/Y} = \frac{Z}{1+Z}$

We shall use the change of density formula + part (a) to compute pdf of U.

To this end, take $I = \text{Range}(Z) = (0, \infty)$

and $g : I \to \mathbb{R}$ defined by $g(z) = \frac{z}{1+z}$, $z \in I = (0,\infty)$

Check that $\text{Range}(g) = (0,1) = J$

and $g : I \to J$ is a diff bijection such that $g'(z) = \frac{1}{(1+z)^2} > 0 \ \forall z \in I$.

<u>Exc:</u> Check that the inverse map of g is given by $g^{-1} : J \to I$

$$g^{-1}(u) = \frac{u}{1-u}, \ u \in J = (0,1)$$

In particular, $\frac{d}{du} g^{-1}(u) = \frac{1}{(1-u)^2}$, $u \in J$.

$\Rightarrow \text{Range}(U) = J = (0,1)$ $[\because U = g(Z)]$

By the change of density formula U is cont r.v with pdf

$$f_U(u) = \begin{cases} f_Z(g^{-1}(u)) \left| \frac{d}{du} g^{-1}(u) \right| & \text{if } u \in (0,1) \\ 0 & \text{if } u \notin (0,1) \end{cases}$$

Then we have

$$f_U(u) = f_Z\left(\frac{u}{1-u}\right)\frac{1}{(1-u)^2}$$

$$= \frac{1}{\left(1+\frac{u}{1-u}\right)^2} \times \frac{1}{(1-u)^2} = \frac{1}{\left(\frac{1}{1-u}\right)^2} \times \frac{1}{(1-u)^2} = 1$$

$\Rightarrow U$ has pdf $f_U(u) = \begin{cases} 1 & \text{if } 0<u<1 \\ 0 & \text{o.w} \end{cases}$ $U \sim Unif(0,1)$

Remarks: ① We have shown

$$X, Y \overset{iid}{\sim} Exp(\lambda) \Rightarrow \frac{X}{X+Y} \sim Unif(0,1).$$

We shall show later that $\frac{X}{X+Y} \perp\!\!\!\perp X+Y$

② Infact, we shall show the following:

$$\text{ind}\begin{cases} X \sim Gamma(\alpha, \lambda) \\ Y \sim Gamma(\beta, \lambda) \end{cases} \Rightarrow \text{ind}\begin{cases} X+Y \sim Gamma(\alpha+\beta, \lambda) \\ \frac{X}{X+Y} \sim Beta(\alpha, \beta) \end{cases}$$

Exc: Suppose $X, Y \overset{iid}{\sim} Exp(\lambda)$. Compute the cdf of $\frac{X}{X+Y}$ then show that $U \sim Unif(0,1)$.

Exc: If $X, Y \overset{iid}{\sim} N(0,1)$, then find the distribution $Z := \frac{X}{Y}$. Does ~~X~~ Z have finite mean? Justify.

Thm: If (X,Y) is a cont r.v with joint pdf $f_{X,Y}(x,y)$ then $W := XY$ is a cont random variable with pdf

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{w}{x}\right) dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X,Y}\left(\frac{w}{y}, y\right) dy$$

$w \in \mathbb{R}$.

Proof: Exc

Application of above thm:

Exc: Suppose $X, Y \overset{iid}{\sim} \text{Unif}(0,1)$ find dist of $W := XY$.

Exc: Suppose (X,Y) is a cont random vector. Show that $V := X - Y$ is a cont r.v. Write down a formula that gives a pdf of V in terms of a joint pdf of (X,Y). Prove this formula.

Towards Bivariate Change of Density formula

Recall the univariate change of density formula. It needs two open intervals $I, J \subseteq \mathbb{R}$ and a smooth map $g: I \to J$ which is bijective such that $g^{-1}: J \to I$ is also smooth and

$$\frac{dg^{-1}(y)}{dy} \neq 0 \text{ on } J. \text{ ~~That a pdf~~}$$

We also need that Range $(X) \subseteq I \Rightarrow$ Range $(Y) \subseteq J$.

Then pdf of $Y := g(X)$ can be written in terms of a pdf of X as follows:

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| & \text{if } y \in J. \\ 0 & \text{if } y \notin J. \end{cases}$$

Qn: How do we extend this result to the bivariate case?

We shall take I, J ~~to be two~~ $\subseteq R^2$ to be two open subsets of $\mathbb{R}^2$.

$\underline{\text{Def}^n}$: A path in $\mathbb{R}^2$ joining two points $\underset{\sim}{a}, \underset{\sim}{b}$ is a continuous map $f: [0,1] \to \mathbb{R}^2$ Satisfying $f(0) = \underset{\sim}{a}$, $\cancel{f(b)}$ $f(1) = \underset{\sim}{b}$.

We shall abuse the terminology and call the image $f([0,1]) \subseteq \mathbb{R}^2$ a path joining $\underset{\sim}{a}$ and $\underset{\sim}{b}$.

$\underline{\text{Def}^n}$: A set $A \subseteq \mathbb{R}^2$ is called ~~path connected~~ path connected if any two points in A can be joined by a path that lies entirely in A.

In other words, $A \subseteq \mathbb{R}^2$ is path connected if $\forall$ $\underset{\sim}{x}, \underset{\sim}{y} \in A$ $\exists$ a cont map $f: [0,1] \to \mathbb{R}^2$ $f(0) = \underset{\sim}{x}$ and $f(1) = \underset{\sim}{y}$.

In the change of bivariate joint density formula we shall take I and J $\subseteq \mathbb{R}^2$ to be <u>open</u> and path connected.

19/3/24

Setup: When $I, J \subseteq \mathbb{R}^2$ are open and path connected

[illegible] "smooth" map $g: \underset{(\subseteq \mathbb{R}^2)}{I} \longrightarrow \underset{(\subseteq \mathbb{R}^2)}{J}$ which is a bijection.

Aim: to find the distribution of random vector

$(Y_1, Y_2) = g(X_1, X_2)$, where (X_1, X_2) is a cont random vector with a joint pdf $f_{X_1,X_2}(x_1, x_2)$ that vanishes on I^c

(i.e. $\text{Range}(X_1, X_2) \subseteq I$)

Qns: (1) What will "smooth" mean in this setup?

(2) What will be the role of $\dfrac{d\, g^{-1}(y)}{dy}$ in bivariate case?

We shall first answer (2) and then answer (1).

Note that $g: I \longrightarrow J$ is a bijection.

$g: \underset{\in I}{(x_1, x_2)} \longmapsto \underset{\in J}{(y_1, y_2)} = \underset{\in J}{g(x_1, x_2)}$

Then exist an inverse map $g^{-1}: J \longrightarrow I$;

$g^{-1}: (y_1, y_2) \longmapsto (x_1, x_2)$

Suppose h_1 and h_2 are the first and second coordinate of the map g^{-1} i.e, $g^{-1} = (h_1, h_2)$

In other words, $h_1 : J \to \mathbb{R}$ and $h_2 : J \to \mathbb{R}$ are two maps such that

$$g^{-1}(y_1, y_2) = (h_1(y_1, y_2), h_2(y_1, y_2))$$

for all $(y_1, y_2) \in J$.

<u>Defn</u>: The jackobian matrix of g^{-1} is defined to be the 2×2 matrix.

$$J_{g^{-1}}(y_1, y_2) = \begin{pmatrix} \dfrac{\partial h_1(\underset{\sim}{y})}{\partial y_1} & \dfrac{\partial h_1(\underset{\sim}{y})}{\partial y_2} \\ \dfrac{\partial h_2(\underset{\sim}{y})}{\partial y_1} & \dfrac{\partial h_2(\underset{\sim}{y})}{\partial y_2} \end{pmatrix}, \quad \underset{\sim}{y} = (y_1, y_2) \in J.$$

This mtx is of partial derivatives whenever they all exist and are finite.

We shall assume that all of the 4 partial derivatives exist and are cont on J.

Just where,

$$\frac{\partial h_1(y_1, y_2)}{\partial y_1} = \lim_{k \to 0} \frac{h_1(y_1+k, y_2) - h(y_1, y_2)}{k}$$

whenever it exist and is finite.

Answer to Qn (2):

The role of $\frac{d g^{-1}(y)}{dy}$ played in the univariate change of density formula will be played in the bivariate case by the following one dimentional summary of the two dimentional function g^{-1}:

$$\frac{d g^{-1}(y)}{d \underset{\sim}{y}} := \det\left(J_{g^{-1}}(\underset{\sim}{y})\right), \underset{\sim}{y} = (y_1, y_2) \in J.$$

$$= \frac{\partial}{y_1}(\underset{\sim}{y}) \frac{\partial h_2}{\partial y_2}(\underset{\sim}{y}) - \frac{\partial h_1(\underset{\sim}{y})}{\partial y_2} \frac{\partial h_2(\underset{\sim}{y})}{\partial y_1}$$

$$, \underset{\sim}{y} \in J.$$

Ans to Qn (1):

We shall call g a "smooth" function if all the 4 partial derivatives $\frac{\partial h_i}{\partial y_j}$, $i, j \in \{1, 2\}$ exists and are cont on J, and the determinant

$$\frac{d g^{-1}(\underset{\sim}{y})}{d \underset{\sim}{y}} = \det\left(J_{g^{-1}}(\underset{\sim}{y})\right) \neq 0 \;\forall\; \underset{\sim}{y} \in J. \qquad (1)$$

Thm: (Change of bivariate joint density formula)

Suppose I and J are two open and path connected sets and g is a bijective and smooth (as described) map if $\underset{\sim}{X} = (X_1, X_2)$ is a continous random vector with a joint pdf $f_{\underset{\sim}{X}}(\underset{\sim}{x})$ that vanishes on I^c. (this implies Range $(\underset{\sim}{X}) \subseteq I$), then $\underset{\sim}{Y} = (Y_1, Y_2) = g(\underset{\sim}{X})$

$$= g(X_1, X_2)$$

is also continous random vector with joint pdf

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \begin{cases} f_{\underset{\sim}{X}}(g^{-1}(\underset{\sim}{y}))\left|\dfrac{d\,g^{-1}(y)}{d\,y}\right| & \text{if } \underset{\sim}{y} \in J, \\ 0 & \text{if } \underset{\sim}{y} \notin J. \end{cases}$$

Here $\dfrac{d g^{-1}(\underset{\sim}{y})}{d \underset{\sim}{y}}$ is the determinand defined (1).

Remarks: (1) If Range $(\underset{\sim}{X}) = I$, then Range $(\underset{\sim}{Y}) = J$.

(2) The computation of $\dfrac{d g^{-1}(\underset{\sim}{y})}{d(\underset{\sim}{y})}$ can be intensive.

Ex: Suppose $X_1, X_2 \overset{iid}{\sim} Unif(0,1)$

find the joint dist^n $Y_1 := X_1 + X_2$

and $Y_2 := X_1 - X_2$

Soln: X_1, X_2 iid $Unif(0,1) \Rightarrow$ a joint pdf of (X_1, X_2) is

$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} 1 & \text{if } (x_1,x_2) \in (0,1)^2 \\ 0 & \text{if } (x_1,x_2) \notin (0,1)^2 \end{cases}$$

It is easy to check that $I := Range(X_1, X_2) = (0,1)^2$ is open and path connected.

~~on exam just say that it is open and path connected~~

Define: $g : I \to \mathbb{R}^2$ by

$$g(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$$

$$(x_1, x_2) \in I = (0,1)^2.$$

Qn: What is $g(I) := J$.

Note that $g(I) \subseteq (0,2) \times (1,-1)$.

$$g(I) = \{ (y_1, y_2) \in \mathbb{R}^2 : (y_1, y_2) = g(x_1, x_2) \text{ for some } (x_1, x_2) \in I \}$$

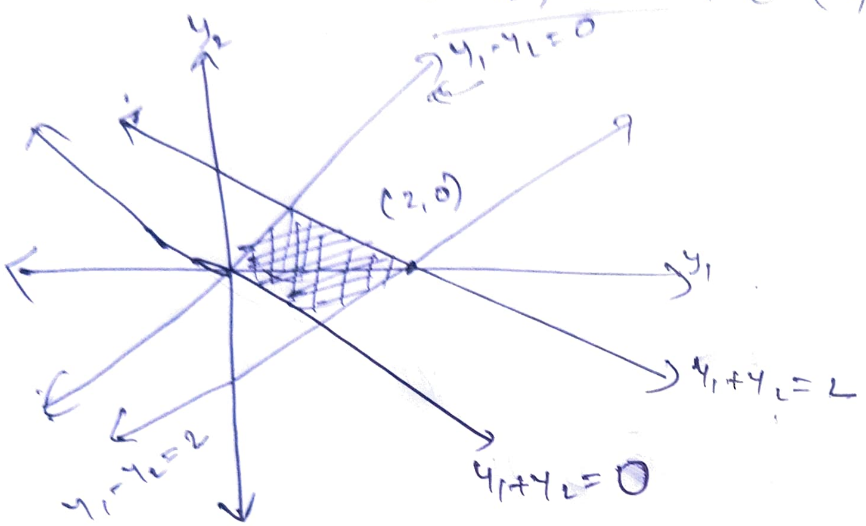
$$= \{ (y_1, y_2) \in \mathbb{R}^2 : (y_1, y_2) = (x_1 + x_2, x_1 - x_2) \text{ for some } (x_1, x_2) \in I = (0,1)^2 \}$$

$= \{(y_1, y_2) \in \mathbb{R}^2 : (x_1, x_2) = (\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}) \in (0,1)^2\}$

$= \{(y_1, y_2) \in \mathbb{R}^2, \; 0 < y_1+y_2 < 2, \; 0 < y_1 - y_2 < 2\}$

Therefore,

$$J = \{(y_1, y_2) \in \mathbb{R}^2 : 0 < y_1+y_2 < 2, \; 0 < y_1-y_2 < 2\}$$



Note that we have more or less solved the following exercise.

__Exc__: Show that $g : I \to J$ is ~~1-1 on~~ bijective. (One-one has to be checked onto already done and we found the range).

The inverse map is given by $g^{-1} : J \to I$ is given by $g^{-1}(y_1, y_2) = \left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right)$, $(y_1, y_2) \in J$.

Therefore in our notation the maps $h_1 : J \to \mathbb{R}$ and $h_2 : J \to \mathbb{R}$ are given by

$h_1(y_1, y_2) = \frac{y_1+y_2}{2}$ and $h_2(y_1, y_2) = \frac{y_1-y_2}{2}$

where ~~y_1 and y_2~~ $(y_1, y_2) \in J$ so that

$$g^{-1}(y_1, y_2) = (h_1(y_1, y_2), h_2(y_1, y_2)) \quad ((y_1, y_2) \in J).$$

Hence the Jacobian matrix of g^{-1} is.

$$J_{g^{-1}}(y_1, y_2) = \begin{pmatrix} \frac{\partial h_1}{y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{y_1} & \frac{\partial h_2}{\partial y_2} \end{pmatrix}, (y_1, y_2) \in J.$$

$$= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

$$\Longrightarrow \det\left(J_{g^{-1}}(\underset{\sim}{y})\right) = \frac{-1}{2}.$$

We have checked that g is "smooth" in the sense that all four partial derivatives exist and are continous, and $\det(J_{g^{-1}}(y)) \neq 0$ $\forall \underset{\sim}{y} \in J$.

We have checked all the assumptions of the bivariate change of joint density formula are satisfied. Therefore it follows that $\underset{\sim}{Y} = (Y_1, Y_2)$ is also a cont random vector with a joint pdf.

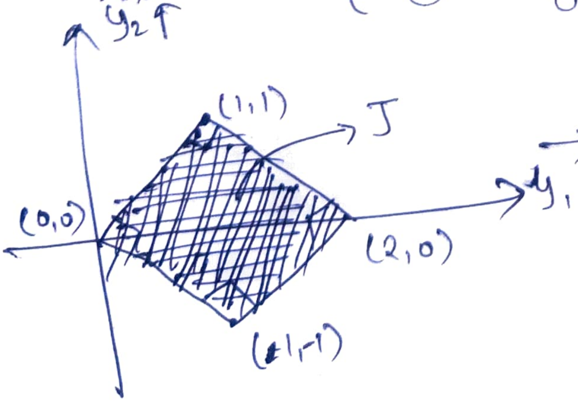
$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} f_{X_1,X_2}(g^{-1}(y_1,y_2)) \left| \dfrac{d\, g^{-1}(\underset{\sim}{y})}{d(\underset{\sim}{y})} \right|, & \underset{\sim}{y} = (y_1,y_2) \in J \\ 0 & \text{o.w.} \end{cases}$$

$$= \begin{cases} (1) \times \left|\left(-\frac{1}{2}\right)\right|, & \underset{\sim}{y} \in J \\ 0, & \text{o.w} \end{cases}$$

$$= \begin{cases} \frac{1}{2}, & \underset{\sim}{y} \in J \\ 0 & \text{o.w.} \end{cases}$$

Therefore, (Y_1, Y_2) have a joint pdf

$$f_{Y_1 Y_2} = \begin{cases} 1/2 & \text{if } 0 < y_1 + y_2 < 2,\ 0 < y_1 - y_2 < 2 \\ 0 & \text{o.w.} \end{cases}$$



$\rightarrow$ Area(J) = 2 sq units.

$(Y_1, Y_2) \sim \text{Unif}(J)$.

20/3/24

Ex: Suppose $X_1 \perp\!\!\!\perp X_2$, $X_1 \sim \text{Gamma}(\alpha_1, \lambda)$ and $X_2 \sim \text{Gamma}(\alpha_2, \lambda)$. Define $S = X_1 + X_2$ and $Y = \frac{X_1}{S} = \frac{X_1}{X_1 + X_2}$. find the joint distribution of Y and S.

Soll^n:

Clearly Range $(X_1, X_2) = (0, \infty)^2 := I$

It is easy to check that I is open and path connected.

Define $g : I \to \mathbb{R}^2$

$$(*) \text{ — } \quad g(x_1, x_2) = \left(\frac{x_1}{x_1 + x_2}, x_1 + x_2 \right); \quad (x_1, x_2) \in I = (0, \infty)^2$$

$$= \left(\frac{x_1}{s}, s \right), \text{ where } s = x_1 + x_2.$$

Qn: What is $g(I)$?

$(x_1, x_2) \in (0, \infty) \Rightarrow x_1 + x_2 \in (0, \infty)$ and

$\frac{x_1}{x_1 + x_2} \in (0, 1) \Rightarrow g(I) \subseteq (0, 1) \times (0, \infty)$

Claim:

Take $(y, s) \in (0, 1) \times (0, \infty)$

Rough Work

To find $(x_1, x_2) \in I$

s.t $g(x_1, x_2) = (y, s)$

ie, $\frac{x_1}{x_1+x_2} = y$

$x_1 + x_2 = s$

$\left\{ \begin{array}{l} x_1 = ys \\ x_2 (1-y) s \end{array} \right.$

Note that $(ys, (1-y)s) \in (0,\infty)^2 = I$

and ~~$g : I \to J$~~

$g(ys, (1-y)s) = (y, s)$

$\Rightarrow (y, s) \in g(I)$

$\Rightarrow (0, 1) \times (0, \infty) \subseteq g(I)$

which toger with the reverse inclusion proves the claim.

Therefore define $J := (0,1) \times (0,\infty)$

It is easy to check that $g : I \to J$ defined by $(*)$ is ~~1-1~~ bijection. [Just need to Check that $g(x_1, x_2) = g(x_1', x_2') \Rightarrow (x_1, x_2)\ (x_1', x_2')$]

In the process we have also computed the inverse function $g^{-1} : J \to I$, which is given by

$g^{-1}(y, s) = (ys, (1-y)s)$, $(y, s) \in J = (0,1) \times (0,\infty)$

$h_1 : J \to \mathbb{R}$ and $h_2 : J \to \mathbb{R}$. defined by

$h_1(y, s) = ys$, $(y, s) \in J = (0,1) \times (0,\infty)$

$h_2(y, s) = (1-y)s$, $(y, s) \in J = (0,1) \times (0,\infty)$

so that

$$g^{-1}(y,s) = (h_1(y,s), h_2(y,s)), \quad (y,s) \in J,$$

Therefore we get-

$$\frac{\partial h_1}{\partial y}(y,s) = s \qquad \frac{\partial h_1}{\partial s} = y$$

$$\frac{\partial h_2}{\partial y} = -s \qquad \frac{\partial h_2}{\partial s}(1-y)$$

$\Rightarrow$ ~~det g^{-1}~~

$$J_{g^{-1}}(y,s) = \begin{pmatrix} s & y \\ -s & 1-y \end{pmatrix} \qquad (y,s) \in J.$$

$$\det(J_{g^{-1}}(y,s)) = s.$$

$$\Rightarrow \frac{\partial g^{-1}(y,s)}{\partial(y,s)} = \det(J_{g^{-1}}(y,s)) = \det\left(\begin{pmatrix} s & y \\ -s & 1-y \end{pmatrix}\right)$$

$$= s(1-y) - (-sy)$$

$$= s. > 0$$

$\forall (y,s) \in J$

Note that g is "smooth" because all the partial derivatives exists and are continous functions of (y,s) on J, and $\det(J_{g^{-1}}(y,s)) \neq 0$ $\forall$ $(y,s) \in J$.

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$

$$= \frac{\lambda^{\alpha_1} e^{-\lambda x_1} x_1^{\alpha_1 - 1}}{\Gamma(\alpha_1)} \frac{\lambda^{\alpha_2} e^{-\lambda x_2} x_2^{\alpha_2 - 1}}{\Gamma(\alpha_2)}$$

$$(x_1, x_2) \in I = (0,\infty)^2$$

Since Range $(X_1, X_2) = I$, it follows Range $(Y,S) = J$.

Take $(y,s) \in J = (0,1) \times (0,\infty)$. Then by the change of bivarite joint density formula, a joint pdf of (Y,S) is $f_{Y,S}(Y,S)$

$$f_{Y,S}(y,s) = f_{X_1,X_2}(g^{-1}(y,s)) \left| \frac{\cancel{\det} d(g^{-1}(y,s))}{d(y,s)} \right|, (y,s) \in J.$$

$$= \left(f_{X_1,X_2}(ys, (1-y)s)\right) s$$

$$= \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \cancel{e^{-\lambda(x_1+x_2)}} e^{-\lambda(ys + (1-y)s)} (ys)^{\alpha_1 - 1} ((1-y)s)^{\alpha_2 - 1} s, \quad y \in (0,1), s \in (0,\infty)$$

$y \in (0,1), \cancel{s} s > 0$

$$= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda s} s^{\alpha_1+\alpha_2-1} y^{\alpha_1-1}(1-y)^{\alpha_2-1}$$

$1 \cancel{(y,s)}$
$(y \in (0,1), s > 0)$

$$= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1-1}(1-y)^{\alpha_2-1} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} e^{-\lambda s} s^{\alpha_1+\alpha_2-1}, \quad y \in (0,1), s \in (0,\infty)$$

$$\Rightarrow f_{Y,S}(y,s) = \underbrace{\left(\frac{1}{B(\alpha_1,\alpha_2)} y^{\alpha_1-1}(1-y)^{\alpha_2-1}\right)}\underbrace{\left(\frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} e^{-\lambda s} s^{\alpha_1+\alpha_2-1}\right)}$$

$y \in (0,1), s \in (0,\infty)$

$\Rightarrow$ $Y \sim \text{Beta}(\alpha_1, \alpha_2)$
$S \sim \text{Gamma}(\alpha_1+\alpha_2, \lambda)$.

we have proved the following result.

Propⁿ:

ind $X_1 \sim \text{Gamma}(\alpha_1, \lambda)$, $X_2 \sim \text{Gamma}(\alpha_2, \lambda)$

$\Rightarrow$ ind $X_1 + X_2 \sim \text{Gamma}(\alpha_1+\alpha_2, \lambda)$, $\dfrac{X_1}{X_1+X_2} \sim \text{Beta}(\alpha_1, \alpha_2)$

If $\alpha_1 = \alpha_2 = 1$ then

$$X_1, X_2 \overset{iid}{\sim} Exp(\lambda) \longrightarrow \begin{cases} X_1 + X_2 \sim Gamma(2, \lambda) \\ \dfrac{X_1}{X_1+X_2} \sim Beta(1,1) \end{cases}$$

i.e., $\dfrac{X_1}{X_1+X_2} \sim Unif(0,1)$.

Go back to Covariance and ~~Co~~ Correlation

Recall: If X and Y both have finite second moments, then ~~(X, Y) has~~ XY has finite mean. This follows from:

$$0 \le |XY| \le \cancel{\frac{|X|^2 + |Y|^2}{2}} \; \frac{X^2 + Y^2}{2}$$

Thm: (Cauchy - ~~Sch~~ Schwarz inequality for random Variables).

Suppose X, Y are jointly distributed random variables with finite second moments then, XY has finite mean and ~~??~~ $|E(XY)| \le {}^{+}\sqrt{E(X^2)}\,\sqrt{E(Y^2)}$

Equivalently

$$\left(E(XY)\right)^2 \leq E(X^2)\,E(Y^2).$$

<u>Remark</u>: CS inequality tell us

$$-\sqrt{E(X^2)E(Y^2)} \leq E(XY) \leq +\sqrt{E(X^2)E(Y^2)}.$$

After proving the CS ineq, we shall discuss when inequality holds on either side of (CS)

<u>Proof of Thm:</u>

$$q(t) = E\left[\{Y - tX\}^2\right] \geq 0 \quad \forall\, t \in \mathbb{R}$$

$$= E\left[Y^2 - 2tXY + t^2X^2\right]$$

$$\Rightarrow q(t) = E(X^2)\,t^2 - 2E(XY)\,t + E(Y^2) \geq 0 \quad \forall\, t \in \mathbb{R}$$

the disc have to be non-positive.

$$\Rightarrow (2E(XY))^2 - 4E(X^2)E(Y^2) \leq 0$$

$$\Rightarrow (E(XY))^2 \leq E(X^2)E(Y^2).$$

Qn: when do we get $E(XY) = \pm\sqrt{E(X^2)E(Y^2)}$?

Suppose $E(XY) = +\sqrt{E(X^2)E(Y^2)}$

$$\Rightarrow q(t) = E(X^2)t^2 - 2\sqrt{E(X^2)E(Y^2)}\,t + E(Y^2)$$

$$\Rightarrow q(t) = \left(\sqrt{E(X^2)}\,t - \sqrt{E(Y^2)}\right)^2$$

If $P(X=0)=1$ then $P(XY=0)=1$

$\Rightarrow$ equality holds in the second (very first one) inequality of CS.

Assume $P(X=0) < 1$

$$\Rightarrow E(X^2) > 0$$

Fact: $E(X^2) = 0 \Rightarrow P(X=0) = 1$

(obv is X is discrete/cont difficult to prove if X is ~~cont~~). any r.v.

Define $\gamma = \dfrac{\sqrt{E(Y^2)}}{\sqrt{E(X^2)}}$

$\Rightarrow \cancel{q(-\lambda\gamma^2)}\; q(\gamma) = 0$

$\Rightarrow E[(Y-\gamma X)^2] = 0$

$\cancel{\Rightarrow Y = \gamma X}$

$P[(Y-\gamma X) = 0] = 1$

Hence

$E(XY) = \pm\sqrt{E(X^2)E(Y^2)}$

$\Rightarrow P(X=0) = 1$ or

$\exists\ \gamma \geq 0$ such that

$P[Y = \gamma X] = 1.$

Therefore we get the following

$E(XY) = +\sqrt{E(X^2)E(Y^2)}$

$\Rightarrow$ (1) $P(X=0) = 1$ or $P(Y=0) = 1$

or $\exists\ \gamma > 0$ s.t $P[Y = \gamma X] = 1$

Conversely if $P[X=0]=1$ or $P[Y=0]=1$

or $P[\cancel{X}Y=\gamma X]=1$ for some $\gamma \in (0,\infty)$

then $E(XY)=+\sqrt{E(X^2)E(Y^2)}$

We have proved the following theorem-

Suppose X, Y are jointly distributed r.v with finite second moments then we have

(a) $E(XY)=+\sqrt{E(X^2)E(Y^2)}$ iff either ~~a~~

$P(X=0)=1$ or $P(Y=0)=1$ or

$\exists\ \gamma>0$ such that $P[X=\gamma Y]=1$.

(b) $E(XY)=-\sqrt{E(X^2)E(Y^2)}$ iff either

$P(X=0)=1$ or $P(Y=0)=1$ or $\exists\ \gamma<0$

such that $P[Y=\gamma X]=1$. [Just substitute $X'=-X$ to prove]

<u>Remark</u>: (b) follows from (a) by considering ~~φ~~ $X'=-X$.

Correlation Coefficient

Defn: Suppose X and Y are jointly distributed non-degenerate random variables having finite second moments. Then the correlation coefficient or the correlation coefficient of X and Y is defined as

$$\mathrm{Corr}(X,Y) \text{ or } \rho(X,Y) = \rho_{X,Y} := \frac{\mathrm{Cov}(X,Y)}{+\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Properties of Correlation Coefficient

⓪ If X is a non-degenerate random variable with finite second moment then $\rho_{X,X} = 1$.

① (Symmetry) Suppose X, Y are jointly distributed non-degenerate r.v.s having finite second moments. Then

$$\rho(X,Y) = \rho(Y,X)$$

Assume that X, Y are jointly distributed non, degenerate random variables with finite 2^{nd} moment.

② (Shift-invariance) For all $b, d \in \mathbb{R}$

$$\rho(X+b, Y+d) = \rho(X, Y).$$

Proof: This follows from:

$$Cov(X+b, Y+d) = Cov(X, Y)$$

$$Var(X+b) = Var(X)$$
$$Var(Y+d) = Var(Y).$$

③ (Scale invariance of the absolute value of the co-relation)

for all $a, c \in \mathbb{R}$, the following holds

$$\rho(aX, bY) = \begin{cases} \rho(X,Y) & \text{if } ac > 0 \\ \text{undefined} & \text{if } ac = 0 \\ -\rho(X,Y) & \text{if } ac < 0. \end{cases}$$

Pf: If $ac = 0$ the either $a = 0$ or $c = 0$

$\Rightarrow$ either $Var(aX) = 0$ or $Var(cY) = 0$

$\Rightarrow \rho(X,Y)$ is undefined

Suppose $ac \neq 0$. Then

$$\rho(aX, cY) = \frac{\operatorname{Cov}(aX, cY)}{\sqrt{\operatorname{Var}(aX)\operatorname{Var}(cY)}}$$

$$= \frac{ac \operatorname{Cov}(X,Y)}{|ac| \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

$$= \begin{cases} +\rho(X,Y) & \text{if } ac > 0 \\ -\rho(X,Y) & \text{if } ac < 0. \end{cases}$$

(4) (Correlation is invariant under change of units)

Take $a, b, c, d \in \mathbb{R}$ such that $a \neq 0, c \neq 0$ then

$$\rho(aX+b, cY+d) = \begin{cases} \rho(X,Y) & \text{if } ac > 0 \\ -\rho(X,Y) & \text{if } ac < 0 \end{cases}$$

<u>Proof</u>: 2 + 3 $\Longrightarrow$ 4

(5) ~~Coso~~ $-1 \leq \rho(X,Y) \leq 1$ (from C-S inequality)

Moreover (a) $\rho(X,Y) = +1$ holds iff $\exists$ $a > 0$ and $b \in \mathbb{R}$ such that $P[Y = aX + b] = 1$.

This is called (perfect positive linear association).

(b) $\rho(X,Y) = -1$ holds iff $\exists$ ~~b>0~~ $a < 0$ and $b \in \mathbb{R}$ s.t $P[Y = aX + b] = 1$

(Perfect negative linear association.)

Properties of Correlation Coefficient 25/3/24

X and Y are jointly distributed nondegenerate r.v.s having finite 2nd moment.

(5) $-1 \leq \rho(X,Y) \leq 1$ Moreover we have,

(a) $\rho(X,Y) = +1$ holds iff $\exists$ $a>0$ and $b \in \mathbb{R}$ such that $P[Y = aX+b] = 1$

(Perfect positive linear association).

(b) $\rho(X,Y) = -1$ holds if and only if $\exists$ $a<0$ and $b \in \mathbb{R}$ such that $P[Y = aX+b]$

[Perfect negative linear association].

<u>Remarks</u>: Correlation Coefficient is a measure of linear association between two r.v.s. Its sign gives the direction (positive or negative) of linear association and its absolute value ($\in [0,1]$) gives the amount of linear association.

$X \sim Unif\{-1,0,1\}$, $Y = X^2 \Rightarrow \rho(X,Y) = 0$

Here the association is "purely quadratic" and there is no linear association between X and Y.

<u>Proof of 5</u>:

Let $\mu_x = E(X)$ and $\mu_y = E(Y)$. Applying Cauchy Schwarz Inequality on $U = X - \mu_x$ and $V = Y - \mu_y$, we get:

$$|E(UV)| \le +\sqrt{E(U^2)E(V^2)}$$

$$\Rightarrow \left(E[(X-\mu_x)(Y-\mu_y)]\right) \le +\sqrt{E[(X-\mu_x)^2]\,E[(Y-\mu_y)^2]}$$

$$\Rightarrow |\mathrm{Cov}(X,Y)| \le +\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

$$\Rightarrow \rho(X,Y) \le 1$$

$$\Rightarrow -1 \le \rho(X,Y) \le 1.$$

(b) If part: Assume that $\exists\ a>0$ and $b \in \mathbb{R}$ s.t $P[Y = aX+b] = 1$. In this case,

$$\rho(X,Y) = \rho(X, aX+b) \overset{(4)}{=} \rho(X,X) \overset{(0)}{=} 1$$

Only if part: Suppose $\rho(X,Y) = +1$. This means

$$E[(X-\mu_x)(Y-\mu_y)] = +\sqrt{E[(X-\mu_x)^2]\,E[(Y-\mu_y)^2]}$$

where $U = X - \mu_x$ and

$V = Y - \mu_y$

X, Y non-degenerate $\Rightarrow P(U=0) \neq 1$ and $P(V=0) \neq 1$

$\Rightarrow \exists\ \gamma > 0$ such that ~~P~~ $P(V = \gamma U) = 1$

$\Rightarrow P[Y - \mu_Y = \gamma(X - \mu_X)] = 1$ for some $\gamma > 0 \Rightarrow P[Y = \underbrace{\gamma}_{a} X + \underbrace{(\mu_Y - \gamma\mu_X)}_{b}] = 1$

for some $\gamma > 0$.

$\Rightarrow \exists\ a > 0$ and $b \in \mathbb{R}$ s.t $P[Y = aX + b] = 1$

Proof of (b) is similar.

Linear Algebra of nnd and pd matrices with real entries.

Notation: ① For $m, n \in \mathbb{N}$

$$\mathbb{R}^{m\times n} = \{A : A \text{ is a } m\times n \text{ real matrix}\}$$

denotes the set of all $m\times n$ matrices with real entries.

② We shall identify the set $\mathbb{R}^{m\times 1}$ of all Column vectors of dimention n with $\mathbb{R}^{m}$

Def^n: A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is called non-negative definite (nnd) if $\forall \ \underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$

$\underset{\sim}{x}^t A \underset{\sim}{x} \geq 0$.

eg: ① I_n, λI_n with $\lambda \geq 0$, $\mathrm{diag}(\alpha_1, \alpha_2 \ldots \alpha_n)$ with $\alpha_1 \geq \alpha_2 \geq \ldots \alpha_n \geq 0$.

② Take any $U \in \mathbb{R}^{n\times p}$ (for any $p \in \mathbb{N}$) and define $\underset{n\times n}{A} = \underset{n\times p}{U}\,\underset{p\times n}{U^t}$. Then A is non negative definite.

Pf: Take any $x \in \mathbb{R}^n$. Then

$$x^T A x = x^T U U^T x = (U^T x)^T (U^T x)$$

$$= x^T (U U^T) x$$

$$= \cancel{(x^T U)}$$

$$(U^T x)^T (U^T x)$$

$$\geq 0.$$

Remark: It can be shown that any symmetric $n \times n$ non negative matrix is of the ~~form~~ arises in the fashion mentioned in (2).

Def^n: A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called positive definite (pd) if ~~for~~ $\forall x \in \mathbb{R}^n \setminus \{0\}$ $x^t A x > 0$.

Remark: (1) Any pd matrix is also nnd.
(2) A is pd iff A is non-negative

and $\underset{\sim}{x}^t A \underset{\sim}{x} = 0 \Longrightarrow \underset{\sim}{x} = 0$

$$\boxed{\text{Pd}} \; \overset{\Longrightarrow}{\not\Longleftarrow} \; \boxed{\text{nnd}}$$

eg:- ① I_d is Pd, λI_n ($\lambda > 0$), diag $(\alpha_1, \dots, \alpha_n)$ if $\alpha_i > 0$.

② Take any non-singular matrix $U \in \mathbb{R}^{n\times m}$ and define $A_{n\times n} = \underset{mn}{U}\,\underset{n\times n}{U^t}$

Then A is pd.

Proof: $A^T = (UU^T)^T = (U^T)^T U^T = UU^T = A$.

$\Rightarrow A$ is symmetric square mtx.

Take any $\underset{\sim}{x} \in \mathbb{R}^n / \{\underset{\sim}{0}\}$. Then $x^T A \underset{\sim}{x} =$

$$\underset{\sim}{x}^T U U^T \underset{\sim}{x}$$

$$= (U^T \underset{\sim}{x})^T \underbrace{(U^T \underset{\sim}{x})}_{n\times 1}$$

$\Rightarrow U^T \underset{\sim}{x} \neq \underset{\sim}{0}$.

& ~~$y_n = (y_1, y_2$~~

$\underset{\sim}{y} = (y_1, y_2 \dots y_n)^T \in \mathbb{R}^n \setminus \{0\}$.

$$\underset{\sim}{x}^T A \underset{\sim}{x} = \underset{\sim}{y}^T y = \sum_{i=1}^{n} y_i^2 \neq 0.$$

Thm: Any symmetric pd matrix $A \in \mathbb{R}^{n \times n}$ is of the form (*). $A = UU^T$ for some nonsingular matrix $U \in \mathbb{R}^{n \times n}$.

Defn: The mean vector of $\underset{\sim}{X}$ is defined as

$$\mu_{\underset{\sim}{X}} = E(\underset{\sim}{X}) = \left(E(X_1), E(X_2), \ldots E(X_m)\right)^T \in \mathbb{R}^{m \times 1}$$

provided each X_i has finite mean.

Remark: Note that if (*) holds, then

$$\det(A) = \det(UU^T) = \det(U)\det(U^T)$$
$$= (\det(U))^2$$
$$= |\det(U)| = +\sqrt{\det(A)}$$

Defn: Assume that each X_i and each Y_j have finite covariance. Then the covariance matrix between the random vectors $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ is defined as:

$$\underset{m\times n}{Cov}(\underset{\sim}{X}, \underset{\sim}{Y}) = ((Cov(X_i, Y_j)))_{\substack{1\le i\le m \\ 1\le j\le n}}.$$

$$A = ((a_{ij}))_{\substack{1\le i\le m \\ 1\le j\le n}} \qquad A\in\mathbb{R}^{m\times n},$$

with $(i,j)^{th}$ entry a_{ij}

Defn: ~~Assuming~~ Assume that each X_i and find 2nd moment. Then the variance-covariance matrix / Covariance matrix / dispersion matrix. Of the the random vector $\underset{\sim}{X}$ is defined as $\cancel{\underset{n\times n}{Cov}(\underset{\sim}{X}, \underset{\sim}{X})}$

$$Var(\underset{\sim}{X}) \text{ / or } Disp(\underset{\sim}{X}) = \underset{m\times m}{Cov}(\underset{\sim}{X_i}, \underset{\sim}{X_j})$$

$$= \mathbb{R}^{m\times m}.$$

Remark : Clearly, the i^{th} diagonal element of the dispersion matrix is $Var(X_i)$ for each i, also off diagonal elements are the Cross-Covariances. In other-words, $i \neq j$, the (i,j)th entry of $Var(\underset{\sim}{X})$ is $Cov(X_i, X_j)$. Clearly ~~the~~ $Var(\underset{\sim}{X})$ is Symmetric matrix.

Properties:

① Assume that each X_i and each Y_j have finite 2nd moments then the following properties holds.

① ~~The~~ $E(a\underset{\sim}{X} + \underset{\sim}{a}) = a\,E(\underset{\sim}{X}) + \underset{\sim}{a}$

$\forall A \in \mathbb{R}^{p \times m}$ and $\forall$ $\mathbb{R}^p$.

(Linearity of Mean)

2) $Cov(\underbrace{A\underset{\sim}{X} + \underset{\sim}{a}}_{p\times 1}, \underbrace{B\underset{\sim}{Y} + b}_{q\times 1}) = \overbrace{\underbrace{A}_{p\times m}\, \underbrace{Cov(\underset{\sim}{X}, \underset{\sim}{Y})}_{m\times n}\, \underbrace{B^T}_{n\times q}}^{p \times q}$

$\forall A \in \mathbb{R}^{p\times m}$, $\forall \underset{\sim}{a} \in \mathbb{R}^{q\times 1}$ $(p, q \in \mathbb{N})$

(Bilinearity of Covariance).

3) $\mathrm{Var}(\overbrace{A\underset{\sim}{X}}^{p\times 1}+\underset{\sim}{a}) = \overbrace{A}^{p\times m}\underbrace{\mathrm{Var}(\underset{\sim}{X})}_{m\times m}\underbrace{A^T}_{m\times p}$.

$\forall A \in \mathbb{R}^{p\times n}$ $(p\in\mathbb{N})$ and $\forall \underset{\sim}{a} \in \mathbb{R}^{p\times 1}$.

Exc: Show that for any random vector $\underset{\sim}{X}$, $\mathrm{disp}(\underset{\sim}{X})$ is nnd.

Exc: If $m=n$ then ~~Ts~~ $\mathrm{Var}(\underset{\sim}{X}+\underset{\sim}{Y})$

$= \mathrm{Var}(\underset{\sim}{X}) + \mathrm{Var}(\underset{\sim}{Y})$

$+ \mathrm{Cov}(\underset{\sim}{X},\underset{\sim}{Y}) + \mathrm{Cov}(\underset{\sim}{Y},\underset{\sim}{X})$

by bilinearity.

Bivariate Normal Distribution

Example: Suppose $X_1, X_2 \overset{iid}{\sim} N(0,1)$ and

$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is non-singular

~~Define~~ $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \underset{\sim}{Y} := A\underset{\sim}{X}$, where

$\underset{\sim}{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

~~Define~~ In other words,

$Y_1 := a_{11}X_1 + a_{12}X_2$ and

$Y_2 := a_{21}X_1 + a_{22}X_2$

Find the joint distribution of Y_1 and Y_2.

<u>Solution:</u>

$Y_1 := a_{11}X_1 + a_{12}X_2$

and $Y_2 := a_{21}X_1 + a_{22}X_2$.

Find the joint distribution of Y_1 and Y_2.

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2}$$

$x_1, x_2 \in \mathbb{R}$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)}$$

$$\Rightarrow f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{2\pi} e^{-\frac{1}{2}\underset{\sim}{x}^T\underset{\sim}{x}}, \quad \underset{\sim}{x} \in \mathbb{R}^2.$$

We shall use the bivariate change of joint density formula to compute a joint pdf of $\underset{\sim}{Y}$. In this case, $I = \text{Range}(\underset{\sim}{X}) = \mathbb{R}^2$ (open + path connected)

and $g : I \to \mathbb{R}^2$ is defined by

$$g(\underset{\sim}{x}) = A\underset{\sim}{x}, \quad \underset{\sim}{x} \in I = \mathbb{R}^2$$

Exc: Using the non-singularity of A, show that g is one-to-one and $g(I) = \mathbb{R}^2 = J$. Also show that inverse map $g^{-1} : J \to I$ is given by $g^{-1}(\underset{\sim}{y}) = A^{-1}\underset{\sim}{y}$, $\underset{\sim}{y} \in J \in \mathbb{R}^2$.

Exc: Show that $J_{g^{-1}}(\underset{\sim}{y}) = A^{-1} \; \forall \, \underset{\sim}{y} \in J = \mathbb{R}^2$

In particular, this implies that

$$\frac{d\, g^{-1}(\underset{\sim}{y})}{d\underset{\sim}{y}} = \det\left(J_{g^{-1}}(\underset{\sim}{y})\right)$$

$$= \det(A^{-1}) \;\forall\; y \in \mathcal{J} \subseteq \mathbb{R}^2$$

$$\neq 0 \;\forall\; \underset{\sim}{y} \in \mathcal{J}.$$

All the assumptions of bivariate change of density formula are ~~satisfy~~ satisfied.

Hence $\underset{\sim}{Y} = g(\underset{\sim}{X}) = A\underset{\sim}{X}$ is also a continuous random vector with Range$(\underset{\sim}{Y})$

$= \mathcal{J} = \mathbb{R}^2$ and a joint pdf

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = f_{\underset{\sim}{X}}(g^{-1}(\underset{\sim}{y}))\left|\frac{d\, g^{-1}(\underset{\sim}{y})}{d\underset{\sim}{y}}\right|, \quad \forall\; y \in \mathbb{R}^2$$

$$= f_{\underset{\sim}{X}}(A^{-1}\underset{\sim}{y})\,\left|\det(A^{-1})\right|, \; \underset{\sim}{y} \in \mathbb{R}^2.$$

$$= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(A^{-1}\underset{\sim}{y})^T (A^{-1}\underset{\sim}{y})\right\} \left|\det(A^{-1})\right|, \quad \left|\frac{1}{\det(A)}\right|$$

$\underset{\sim}{y} \in \mathbb{R}^2$

$$= \frac{1}{2\pi|\det(A)|} \exp\left\{-\frac{1}{2} y^T (A^T)^{-1} A^{-1} \underset{\sim}{y}\right\}, \; y \in \mathbb{R}.$$

$$= \frac{1}{2\pi\,|\det(A)|} \exp\left\{-\frac{1}{2} \underset{\sim}{y}^T (AA^T)^{-1} \underset{\sim}{y}\right\}, \; \underset{\sim}{y} \in \mathbb{R}.$$

<u>Define</u>: $\underset{2\times 2}{\Sigma} = \underset{2\times 2}{A}\underset{2\times 2}{A^T} \Rightarrow \Sigma \in \mathbb{R}^{2\times 2}$ symm pd.

(since A is non singular)

$$\Rightarrow |\det(A)| = +\sqrt{\det(\Sigma)}$$

$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ is a cont

r.v with a joint pdf

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{1}{2\pi \det(\Sigma)} \exp\left\{-\frac{1}{2} \underset{\sim}{y}^T \Sigma^{-1} \underset{\sim}{y}\right\},$$

$y \in \mathbb{R}^2$.

where $\Sigma = AA^T$.

<u>Fact</u>: For any pd matrix $\Sigma \in \mathbb{R}^{2\times 2}$

(**) gives a valid pdf on $\mathbb{R}^2$.

By mistake 1/4/24

Properties of Bivariate Normal Distribution

(6) If $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\underset{\sim}{\mu}, \Sigma)$, then for all $\underset{\sim}{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \in \mathbb{R}^2 \setminus \{\underset{\sim}{0}\}$, $\underset{\sim}{\alpha}^T \underset{\sim}{X} = \alpha_1 X_1 + a_2 X_2$

$$\sim N(\underset{\sim}{\alpha}^T \underset{\sim}{\mu}, \underset{\sim}{\alpha}^T \Sigma \underset{\sim}{\alpha})$$

<u>Remark</u>: Since Σ is pd and $\underset{\sim}{\alpha} \neq \underset{\sim}{0}$, it follows that $\underset{\sim}{\alpha}^T \Sigma \underset{\sim}{\alpha} > 0$.

A few Important Special cases of (6)

If $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\underset{\sim}{\mu}, \Sigma)$ where $\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ then we have:

(a) $X_1 \sim N(\mu_1, (\sqrt{\sigma_{11}})^2)$ (use $\underset{\sim}{c} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in (6))

(b) $X_2 \sim N(\mu_2, (\sqrt{\sigma_{22}})^2)$ (use $\underset{\sim}{c} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in (6))

(c) $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_{11} + \sigma_{22} + 2\sigma_{12})$ (Use $\underset{\sim}{c} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in (b))

from → continue

Conversely: if $\underset{\sim}{Z} = \binom{Z_1}{Z_2} \sim N_2(\underset{\sim}{0}, \underset{\sim}{I})$, then $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$. (check)

Therefore, we have proved the following result.

Propn: If $\underset{\sim}{Z} \sim N_2(\underset{\sim}{0}, I_2)$ then $\forall$ non-singular 2×2 matrix A with real entries and all $\underset{\sim}{\mu} \in \mathbb{R}^2$ we have ~~$A\underset{\sim}{Z} + \underset{\sim}{\mu} \sim N_2(\underset{\sim}{\mu}, AA^T)$~~

$$\underset{\sim}{X} = A\underset{\sim}{Z} + \underset{\sim}{\mu} \sim N_2(\underset{\sim}{\mu}, \Sigma) \quad \text{where } \Sigma = AA^T.$$

Remark: Any bivariate normal ~~dist~~ r.v $\underset{\sim}{X}$ ~~following~~ $\sim N_2(\underset{\sim}{\mu}, \Sigma)$ arises as follows $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$, $\underset{\sim}{Z} = \binom{Z_1}{Z_2}$ and $\underset{\sim}{X} = A\underset{\sim}{Z} + \underset{\sim}{\mu}$ where $A \in \mathbb{R}^{2\times 2}$ is nonsing such that $\Sigma = AA^T$.

eg. if we take $\underset{\sim}{\mu} = \underset{\sim}{0}$ and $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, then $X_1 = Z_1 + Z_2$ and $X_2 = Z_1 - Z_2$ are jointly normal. X_1 and X_2 are dependent but their association is linear.

This is true for any ~~X~~ bivariate

normal random vector $\underset{\sim}{X}$.

ie, the only possible association between X_1 and X_2 is linear association. Because of this coorelation coefficient between X_1 and X_2, $\rho(X_1, X_2)$ is good ~~mean~~ measure of association between X_1 and X_2.

Suppose $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$ and $\underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}. \text{ Here, } \mu_1 = E(X_1),\ \mu_2 = E(X_2)$$

$$\sigma_{11} = Var(X_1), \quad \sigma_{22} = Var(X_2)$$

$$\sigma_{12} = Cov(X_1, X_2), \quad \sigma_{21} = Cov(X_2, X_1)$$

Thm: Suppose $\underset{\sim}{X} \sim N_2(\underset{\sim}{\mu}, \Sigma)$. Then $X_1 \perp\!\!\!\perp X_2$

$\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

iff $\sigma_{12} = Cov(X_1, X_2) = 0$.

$\boxed{\text{Independent}} \underset{\text{when } X_1, X_2 \text{ are jointly normal}}{\overset{\longrightarrow}{\not\longleftarrow}} \boxed{\text{Uncoorelated}}$

Proof:

Only if part: $X_1 \perp\!\!\!\perp X_2 \Rightarrow Cov(X_1, X_2) = 0$

If part: Suppose $\sigma_{12} = Cov(X_1, X_2) = 0$.

then, $\sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$

Therefore, a joint pdf of $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ is

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{12}}} \exp\left\{ -\frac{1}{2} (x_1 - \mu_1, x_2 - \mu_2) \begin{pmatrix} \frac{1}{\sigma_{11}} & 0 \\ 0 & \frac{1}{\sigma_{12}} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right\}$$

$$= \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{12}}} \exp\left\{ -\frac{1}{2\sigma_1} (x_1 - \mu_1)^2 \right\} \cancel{\sqrt{2\pi}\sqrt{\sigma_{12}}}$$

$N(\mu_1, \sigma_1)$

$$\frac{1}{2\pi\sqrt{\sigma_{22}}} \exp\left\{ -\frac{1}{2\sigma_{22}} (x_2 - \mu_2)^2 \right\}$$

pdf of $N(\mu_2, \sigma_{22})$

Therefore $X_1 \sim N(\mu_1, \sigma_{11})$
$X_2 \sim N(\mu_2, \sigma_{22})$

Also, $f_{\underset{\sim}{X}}(\underset{\sim}{x}) = f_{X_1}(x_1) f_{X_2}(x_2) \Rightarrow X_1 \perp\!\!\!\perp X_2$.

Ex: $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

Define: $X_1 = Z_1 + Z_2$ and $X_2 = Z_1 - Z_2$

Then $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$

$= A \underset{\sim}{Z}$.

where, $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Clearly A is non-singular

Clearly X_1 and X_2 are jointly normal.

$Cov(X_1, X_2) = Cov(Z_1 + Z_2, Z_1 - Z_2)$

$= Var(Z_1) - Var(Z_2) + Cov(Z_2, Z_1)$
$- Cov(Z_1, Z_2) = 0$.

Hence, in this example $X_1 \perp\!\!\!\perp X_2$ (Use the thm).

Easy to check:

$$X_1, X_2 \sim N(0,2)$$

Hence, $X_1, X_2 \overset{iid}{\sim} N(0,2)$

Recall: $\underset{\sim}{Z} \sim N(\underset{\sim}{0}, I_2) \Rightarrow A\underset{\sim}{Z} + \underset{\sim}{\mu} \sim N_2(\underset{\sim}{\mu}, \overbrace{AA^T}^{\Sigma})$

$\forall$ non sing $A \in \mathbb{R}^{2\times 2}$

and $\forall\ \underset{\sim}{\mu} \in \mathbb{R}^2$

Exc: ① Suppose $\underset{\sim}{X} \sim N_2(\mu, \Sigma)$. Cut a non-sing $A \in \mathbb{R}^{2\times 2}$ such that $\Sigma = AA^T$. Define $\underset{\sim}{Z} = A^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$. Show that $\underset{\sim}{Z} \sim N_2(\underset{\sim}{0}, I_2)$

~~This means~~ $\left(\Leftrightarrow \underset{\sim}{X} = A\underset{\sim}{Z} + \underset{\sim}{\mu} \right)$.

(Use change of bivariate joint density formula).

This expression implies that ~~any~~ of the following

Any bivariate random vector

$\underset{\sim}{X} \sim N_2(\underset{\sim}{\mu}, \Sigma)$ is of the form

$\underset{\sim}{X} = A\underset{\sim}{Z} + \underset{\sim}{\mu}$, where $\underset{\sim}{Z} \sim N_2(\underset{\sim}{0}, I_2)$

and $\Sigma = AA^T$

ie, $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

(2) In the univariate case,

$$X \sim N(\mu, \sigma^2) \Rightarrow \underbrace{\left(\overset{+}{\sqrt{\sigma^2}}\right)(X-\mu) \sim N(0,1)}$$

$\frac{X-\mu}{\sigma}$ = the standardized value of X.

Therefore in the bivariate case the non-singular mtx A is playing the role of square-root of Variance - Covariance mtx Σ.

Also $\underset{\sim}{Z}$ is a ~~st~~ standardized value of $\underset{\sim}{X}$.

---

① $\underset{\sim}{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N_2(\underset{\sim}{0}, I_2) \Leftrightarrow Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

② $\underset{\sim}{Z} \sim N_2(\underset{\sim}{0}, I_2) \Rightarrow A\underset{\sim}{Z} + \underset{\sim}{\mu} \sim N_2(\mu, AA^T)$

$\forall$ non sing $A \in \mathbb{R}^{2\times 2}$ and

$\forall\ \underset{\sim}{\mu} \in \mathbb{R}^2$

<u>Cor 1</u>: $\underset{\sim}{X} \sim N_2(\mu_2, \Sigma)$

$\Rightarrow E(\underset{\sim}{X}) = \underset{\sim}{\mu}$

$Var(\underset{\sim}{X}) = \Sigma$

<u>Cor 2</u>: $\underset{\sim}{Z} \sim N_2(0, I_2)$.

$\forall$ orthogonal mtx $A \in \mathbb{R}^{2\times 2}$.

This mean the following

Take $Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

Define $Y_1 = \cos\theta Z_1 - \sin\theta Z_2$

$Y_2 = \sin\theta Z_1 - \cos\theta Z_2$

Then, $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = A\underset{\sim}{Z} = A\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$

where $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is an orthogonal matx.

$\xRightarrow{\text{Cor 2}} \underset{\sim}{Y} \sim N_2(\underset{\sim}{0}, I_2)$

$\Rightarrow Y_1, Y_2 \overset{iid}{\sim} N(0,1)$.

This means that the distribution of $\underset{\sim}{Z} \sim N_2(0, I_2)$ is rotation invariant.

3 If $\underset{\sim}{X} \sim N_2(\underset{\sim}{\mu}, \Sigma)$ then for all nonsingular $B \in \mathbb{R}^{2\times 2}$, we have

$B\underset{\sim}{X} \sim N_2(B\underset{\sim}{\mu}, B\Sigma B^T)$

(The family of bivariate normal distribution is closed under nonsingular linear transformations.)

<u>Proof of ③</u>: Let $A \in \mathbb{R}^{2\times 2}$ non-singular such that $\Sigma = AA^T$. then by Exe given today $\underset{\sim}{X} = A\underset{\sim}{Z} + \underset{\sim}{\mu}$ where $\underset{\sim}{Z} \sim N_2(\underset{\sim}{0}, I_2)$

$\Rightarrow B\underset{\sim}{X} = B(A\underset{\sim}{Z} + \underset{\sim}{\mu}) = \underbrace{BA\underset{\sim}{Z}}_{2\times 1} + \underbrace{B\underset{\sim}{\mu}}_{2\times 1}$.

A, B nonsingular $\Rightarrow BA$ is also nonsingular.

$\overset{②}{\Longrightarrow} B\underset{\sim}{X} \sim N_2(B\underset{\sim}{\mu}, (BA)(BA)^T)$

④ If $\underset{\sim}{X} \sim N_2(\underset{\sim}{\mu}, \Sigma)$, then for all non singular $B \in \mathbb{R}^{2\times 2}$ and for all $\underset{\sim}{k} \in \mathbb{R}^2$, we have

$$B\underset{\sim}{X} + \underset{\sim}{v} \sim N_2(B\underset{\sim}{\mu} + \underset{\sim}{v}, B\Sigma B^T)$$

(5) Suppose $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\underset{\sim}{\mu}, \Sigma)$

Then $X_1 \perp\!\!\!\perp X_2$ iff $Cov(X_1, X_2) = 0$.

Continue -

Proof of (6)

Step 1: Suppose $\underset{\sim}{\mu} = \underset{\sim}{0}$ and $\Sigma = I_2$.

Let $\underset{\sim}{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N_2(\underset{\sim}{0}, I_2) \iff Z_1, Z_2 \overset{iid}{\sim} N(0,1)$

Take $\underset{\sim}{\alpha} \in \mathbb{R}^2 \setminus \{\underset{\sim}{0}\}$

Shall show: $\underset{\sim}{\alpha}^T \underset{\sim}{Z} \sim N(0, \underset{\sim}{\alpha}^T \underset{\sim}{\alpha})$

$$\iff \alpha_1 Z_1 + \alpha_2 Z_2 \sim N(0, \alpha_1^2 + \alpha_2^2)$$

Exe: Show this directly using the convolution formula.

$Z_1, Z_2 \overset{iid}{\sim} N(0,1) \Rightarrow Z_1 + Z_2, Z_1 - Z_2 \overset{iid}{\sim} N(0,2)$

$(1,1)\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$ $(1,-1)\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$

take a vector $\underset{\sim}{\beta} \in \mathbb{R}^2 \setminus \{0\}$ such that $\underset{\sim}{\alpha} \perp \underset{\sim}{\beta}$,
i.e., $\underset{\sim}{\alpha}^T \underset{\sim}{\beta} = 0$.

Define a matrix $A = \begin{bmatrix} \underset{\sim}{\alpha}^T \\ \underset{\sim}{\beta}^T \end{bmatrix} = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$

$\underset{\sim}{\alpha} \perp \underset{\sim}{\beta}$, $\underset{\sim}{\alpha} \neq \underset{\sim}{0}$, $\underset{\sim}{\beta} \neq \underset{\sim}{0} \Rightarrow \underset{\sim}{\alpha}, \underset{\sim}{\beta}$ are lin ind.

Exc :

Show that $\boxed{\underset{\sim}{\alpha}^T \underset{\sim}{Z} \sim N(0, \underset{\sim}{\alpha}^T \underset{\sim}{\alpha})}$ ← This completes Step I

ind

$\underset{\sim}{\beta}^T \underset{\sim}{Z} \sim N(0, \underset{\sim}{\beta}^T \underset{\sim}{\beta})$

$\Rightarrow$ A is non-singular

Also $AA^T = \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{pmatrix} \begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\alpha}^T \underset{\sim}{\alpha} & 0 \\ 0 & \underset{\sim}{\beta}^T \underset{\sim}{\beta} \end{pmatrix}$

$\underset{\sim}{Z} \sim N_2(\underset{\sim}{0}, I_2)$, A non-sing $\Rightarrow AZ \sim$ $N_2(\underset{\sim}{0}, AA^T)$

$$\Rightarrow \begin{pmatrix} \alpha_1 & \alpha_2 \\ \beta_1 & \beta_2 \end{pmatrix}\begin{pmatrix} z_1 \\ z_1 \end{pmatrix} \sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} \underset{\sim}{\alpha}_1 \alpha_{\not 0} & 0 \\ 0 & \underset{\sim}{\beta}^T\beta \end{pmatrix}\right)$$

$$\Rightarrow \begin{pmatrix} \underset{\sim}{\alpha}^T Z \\ \underset{\sim}{\beta}^T Z \end{pmatrix} = \begin{pmatrix} \alpha_1 z_1 + \alpha_2 z_2 \\ \beta_1 z_1 + \beta_2 z_2 \end{pmatrix}$$

$$\sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} \underset{\sim}{\alpha}^T\underset{\sim}{\alpha} & 0 \\ 0 & \underset{\sim}{\beta}^T\underset{\sim}{\beta} \end{pmatrix}\right)$$

Note that $\underset{\sim}{\gamma} = B^T\underset{\sim}{\alpha} \in \mathbb{R}^2 \setminus \{\underset{\sim}{0}\}$

since $\underset{\sim}{\alpha} \neq \underset{\sim}{0}$ and B is non singular

Also $\underset{\sim}{Z} \sim N_2(\underset{\sim}{0}, I_2)$ Therefore by step 1,

$$\underset{\sim}{\alpha}^T B Z = (B^T\underset{\sim}{\alpha})^T \underset{\sim}{Z} \sim \cancel{N(0, (\underset{\sim}{\alpha}^T B)(B^T\underset{\sim}{\alpha}))}$$

$$N(0, \underset{\sim}{\gamma}^T\underset{\sim}{\gamma})$$

$$\Rightarrow \underset{\sim}{\alpha}^T B Z \sim N(0, \underset{\sim}{\alpha}^T \Sigma \underset{\sim}{\alpha})$$

$$\Rightarrow \underset{\sim}{\alpha}^T \underset{\sim}{X} = \underset{\sim}{\alpha}^T \underset{\sim}{\mu} + \underset{\sim}{\alpha}^T B \underset{\sim}{Z} \sim N(\underset{\sim}{\alpha}^T\underset{\sim}{\mu}, \underset{\sim}{\alpha}^T \Sigma \underset{\sim}{\alpha})$$

Conditional Distribution

Ex: Suppose $X_1 \sim Poi(\lambda_1)$ and $X_2 \sim Poi(\lambda_2)$ $\Big\}$ $X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2)$

Fix $n \in \mathbb{N} \cup \{0\}$. Then $P(X_1 + X_2 = n)$

$$= P\left(\bigcup_{k=0}^{n} (X_1 = k, X_2 = n-k)\right)$$

$$= \sum_{k=0}^{n} P(X_1 = k, X_2 = n-k)$$

$$= \sum_{k=0}^{n} P(X_1 = k)\, P(X_2 = n-k)$$

$$= \sum_{k=0}^{n} \left(e^{-\lambda_1} \frac{\lambda_1^k}{k!} \, e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!} \right)$$

$$= \sum_{k=0}^{n} \frac{e^{-(\lambda_1+\lambda_2)} \lambda^n (n!)}{k!(n-k)!(n!)}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{(n!)} \sum_{k=0}^{n} \frac{(\lambda_1^{k} \lambda_2^{n-k})(n!)}{k!(n-k)!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}$$

$n \in \mathbb{N} \cup \{0\}$.

$\Rightarrow X_1 + X_2 \sim \text{Poi}(\lambda_1+\lambda_2)$.

Fix $n_0 \in \mathbb{N} \cup \{0\}$. Suppose it is given $X_1+X_2 = n_0$

What is the "conditional distribution of X_1"?

Clearly, the "conditional range" of X_1 given $X_1+X_2=n_0$ is $\{1,2,\dots n_0\}$. Take $k \in \{0,\dots n_0\}$

We would like to find the conditional probability,

$P(X_1 = k \mid X_1+X_2 = n_0)$

If $n_0 = 0$, then given that $X_1+X_2 = n_0 = 0$, conditionally X_1 is degenerate at 0, ie,

$P[X_1 = 0 / X_1+X_2 = 0] = 1$.

Assume that $n_0 \in \mathbb{N}$ Then,

$$P(X_1 = k \mid X_1+X_2 = n_0)$$
$$= \frac{P(X_1 = k, X_1+X_2 = n_0)}{P(X_1+X_2 = n_0)}$$

$$= \frac{P(X_1 = k, X_2 = n_0 - k)}{P(X_1 + X_2 = n_0)}$$

$$= \frac{\dfrac{e^{-\lambda_1}\lambda_1^{k}}{k!} \; \dfrac{e^{-\lambda_2}\lambda_2^{n_0-k}}{(n_0-k)!}}{\dfrac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^{n_0}}{(n_0)!}}$$

$$= \frac{(n_0)!}{k!(n_0-k)!} \frac{\lambda_1^{k}\lambda_2^{n_0-k}}{(\lambda_1+\lambda_2)^{n_0}} = \binom{n_0}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{k}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n_0-k}$$

$$= \binom{n_0}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{k}\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n_0-k}$$

If $n_0 \in \mathbb{N}$, then given that $X_1 + X_2 = n_0$ conditionally

$$X_1 \sim Bin\left(n_0, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$

special Case ①:

Suppose $\lambda_1 = \lambda_2 = \lambda$.

then,

given that $X_1 + X_2 = n_0$

Conditionally $X_1 \sim Bin(n_0, 1/2)$

Conditional pmf of X_1 given $N = n_0$.

Define: $N := X_1 + X_2$

Note: We shall write $P_{X_1|N}(k|n_0) = P_{X_1|N}(k)$

$:= P[X_1 = k / N = n_0]$

$k \in \{0, 1, \ldots, n_0\}$

for $n_0 = 0$,

$$P_{X_1|N}(k|0) = P(X_1 = k / N = 0)$$

$$= \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

Conditional pmf of X_1 given $N = X_1 + X_2 = 0$

$\Rightarrow E(X_1 | N = 0) = 0$

for $n_0 \in \mathbb{N}$ $\quad N = X_1 + X_2$

$$P_{X_1|N}(k|n_0) = P(X_1 = k | N = n_0)$$

$$= \begin{cases} \binom{n}{k} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n_0-k} & \text{if } k\in\{1,\dots n\} \\ 0 & \text{o.w} \end{cases}$$

pmf of binomial

$B\left(n_0, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$ distribution

$$n_0 \frac{\lambda_1}{\lambda_1+\lambda_2}$$
$$\|$$

$\Rightarrow E(X_1 | N=n_0)$ = expectation of the conditional distribution of X_1 given

$\Rightarrow$ ~~E~~ $N=n_0$.

<u>General case</u>: <u>Conditional Distribution in the Discrete random variable</u>.

Suppose X and Y are two discrete random variables defined on the same Space.

Assume that ~~X has range~~ $\text{Range}(X) = \underset{(\text{Ctble})}{I} \subseteq \mathbb{R}$

$\text{Range}(Y) = \underset{(\text{Ctble})}{J} \subseteq (\mathbb{R})$

Define

$$P_{ij} = P(X=i, Y=j) \ \forall (i,j) \in I \times J.$$

~~(a few~~ (some of them P_{ij}'s may be zero one independence is not known),

Clearly $P_{ij} > 0$ iff $(i,j) \in \text{Range}(X,Y) \subseteq I \times J$.

for all $i \in I$,

$$P_{i\cdot} = \sum_{j \in J} P_{ij} = P(X=i) > 0.$$

for all $j \in J$, $P_{\cdot j} = P[Y=j]$

$$= \sum_{i \in I} P_{ij} > 0.$$

Fix $i \in I$ then, Then the conditional ~~probability mass function~~ pmf of Y given $X=i$ is defined as

$$P_{Y|X}(j|i) = \cancel{P_{Y|X=i}}\ P_{Y|X=i}(j) := P[Y=j \mid X=i]$$

$$= \frac{P_{ij}}{P_{i}}, \quad j \in J.$$

some of them can be zero.

Conditional range of Y given $X=i$ is the set $\{j \in J : P_{Y|X}(j|i) > 0\}$

$$= \{j \in J : P_{ij} > 0\}.$$

<u>Assume that</u> Y has finite mean $\Leftrightarrow \sum_{j \in J} |j| P[Y=j] < \infty$

$$\Leftrightarrow \sum_{j \in J} |j| P_j < \infty$$

<u>Recall</u>: We have fixed $i \in I$,

Given, $X=i$, the r.v Y conditionally follows the the ~~pdf~~ probability distribution $\{P_{Y|X}(j|i) : j \in J\}$

therefore the conditional expectation / conditional mean of Y given $X=i$ is nothing but expectation/mean of this conditional distribution.

... has finite mean ⟹ conditional expectation ...

... finite ... $X =$ ...

or

$$\sum_{j \in J} |j| \, P_{Y|X}(j|i) < \infty$$

Proof:

$$\sum_{j \in J} |j| \, P_{Y|X}(j|i) = \sum_{j \in J} |j| \frac{p_{ij}}{p_{i\cdot}}$$

$$= \frac{1}{p_{i\cdot}} \left(\sum_{j \in J} |j| \, p_{ij} \right)$$

$$< \frac{1}{p_{i\cdot}} \left(\sum_{j \in J} |j| \, p_{\cdot j} \right)$$

$$< \infty$$

by assumption

Therefore, the conditional distribution of Y given $X=i$ has finite mean. We define the conditional expectation mean/expectation of Y given $X=i$.

~~Fix~~

Def^n: Fix $i \in I$, ~~then~~ Assume that Y has finite mean. Then the conditional mean/expectation of Y given $X=i$ is denoted by $E(Y|X=i)$ and is defined by,

$$E(Y|X=i) = \sum_{j \in J} j \, P_{Y|X}(j|i).$$

$$= \sum_{j \in J} j \frac{P_{ij}}{P_{i\cdot}}.$$

Remark: As long as $E(Y) < \infty$

$E(Y|X=i)$ is well defined

thanks to claim ~~b~~ proved before.

Go back to the example

$$\text{and} \begin{cases} X_1 \sim Poi(\lambda_1) \\ X_2 \sim Poi(\lambda_2) \end{cases} \quad N := X_1 + X_2$$

for $n_0 \in \mathbb{N} \cup \{0\}$ we looked at the conditional distribution of $X_1 (= Y)$ given $N = n_0$ ($N = X$, $n_0 = i$)

We found that

$$E(X_1 | N = n_0) = n_0 \frac{d_1}{d_1 + d_2}$$

3/4/24

Recall: $\binom{X}{Y}$ is a discrete random vector. (i.e. Disc Random v. defined on same Probability space).

Range $(X) = \underset{(C \text{ or } bu)}{I} \subseteq \mathbb{R}$ and Range $(Y) = \underset{(C \text{ or } bu)}{J} \subseteq \mathbb{R}$

$$p_{ij} = P[X = i, Y = j], \quad (i, j) \in I \times J.$$

$$p_{i\cdot} = \sum_{j \in J} p_{ij} = P[X = i] > 0 \quad \forall \; i \in I$$

$$p_{\cdot j} = \sum_{i \in I} p_{ij} = P[Y = j] > 0 \quad \forall \; j \in \cancel{I} J.$$

for $i \in I$. The conditional pmf of Y given $X = i$ is $p_{Y|X}(j|i) = \dfrac{p_{ij}}{p_{i\cdot}}$, $j \in J$.

Claim: Y has finite mean $\Longrightarrow \sum_{j\in J} |j|\, P_{Y|X} < \infty$

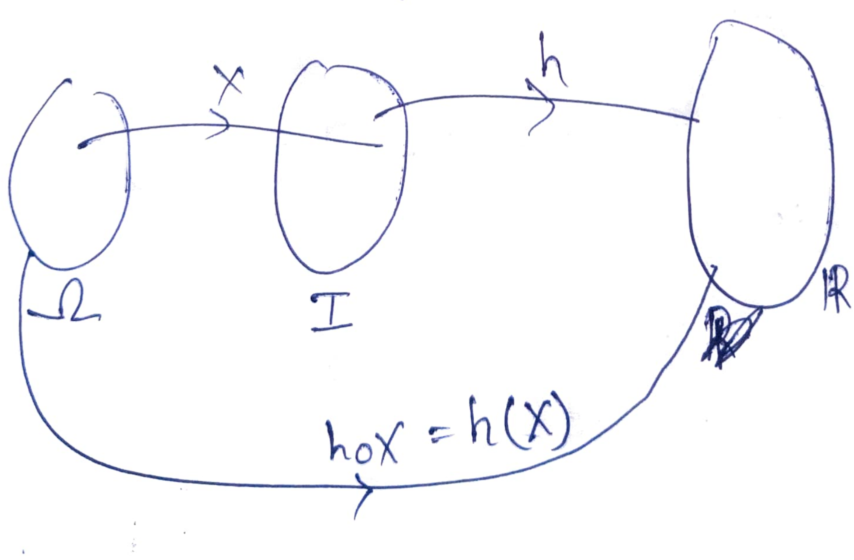
~~to define~~, this enables us to define,

$$E(Y|X=i) = \sum_{j\in J} j\, P_{Y|X}(j|i) = \sum_{j\in J} j \frac{P_{ij}}{P_{i\cdot}}$$

think of $E(Y|X=i)$ is the "best guess for Y" when it is known that X takes the value i".

Assumption: Y has finite mean. Then for each $i \in I = \text{Range}(X)$, $h(i) := E(Y|X=i)$ is well defined and takes real values.
This gives a map $h: I \longrightarrow \mathbb{R}$ defined by

$$h(i) := E(Y|X=i), \quad i \in I.$$



Clearly, the r.v $h(x)$ is well defined taking values in $\{h(i) : i \in I\}$.

$$\boxed{E(Y|X) := h(X)}$$

In other words, $E(Y|X)$ is a r.v that takes the value $h(i) = E(Y|X=i)$ whenever X takes the value i. Think of $E(Y|X)$ as the "best guess for Y when X is known".

In our example, $X_1 \sim Poi(\lambda_1)$
$X_2 \sim Poi(\lambda_2)$ $\quad N := X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2)$

then for all $n_0 \in Range(N) = \mathbb{N} \cup \{0\}$,

$$h(n_0) := \cancel{E(N)}\ E(X_1 | N = n_0) = \frac{n_0 \lambda_1}{\lambda_1 + \lambda_2}$$

$$(h : \mathbb{N} \cup \{0\} \to \mathbb{R})$$

$$h(n_0) = \frac{\lambda_1}{\lambda_1 + \lambda_2} n_0$$

$$\Rightarrow E(X_1 | N) = h(N) = \frac{\lambda_1}{\lambda_1 + \lambda_2} N$$

in this example we see that $E(X_1|N)$ has finite mean and $E[E(X_1|N)] = E\left[\frac{\lambda_1}{\lambda_1+\lambda_2} N\right]$

$$= \frac{\lambda_1}{\lambda_1+\lambda_2} E(N)$$

$$= \frac{\lambda_1}{\lambda_1+\lambda_2} (\lambda_1+\lambda_2)$$

$$= \lambda_1$$

$$= E(X_1).$$

<u>Thm</u>: If Y has finite mean, then the r.v $E(Y|X)$ also has finite mean, and $E[E(Y|X)] = E(Y)$

<u>Remake</u>: The above result say that the expectation Of the conditional expectation is the unconditional/ marginal expectation.

In other words, our "best guess" is Correct on ~~others~~ average.

<u>Pf</u>: <u>To prove</u>: $E(Y|X) := h(X)$ has finite mean.

That is, to prove $\sum_{i \in I} |h(i)| \, P[X=i] < \infty$.

Now, $\sum_{i \in I} |h(i)| \, P[X=i] = \sum_{i \in I} |h(i)| \, p_{i\cdot}$

$$= \sum_{i \in I} |E(Y|X=i)| \, p_{i\cdot}$$

$$= \sum_{i \in I} \left| \sum_{j \in J} j \frac{p_{ij}}{p_{i\cdot}} \right| p_{i\cdot} = \sum_{i \in I} \left| \sum_{j \in J} j \, p_{ij} \right|$$

$$\leq \sum_{i \in I} \sum_{j \in J} |j| \, p_{ij}$$

$$= \sum_{j \in J} \sum_{i \in I} |j| \, p_{ij}$$

[By fubini I part],

$$= \sum_{j \in J} |j| \sum_{i \in I} p_{ij}.$$

$$= \sum_{j \in J} |j| \, P[Y=j] < \infty$$

$\Rightarrow E(Y|X)$ has finite mean.

Also, $E(E(Y|X)) = E[h(X)]$

$$= \sum_{i \in I} h(i) \, p_{i\cdot}$$

$$\left[\because P[X=i] = p_{i\cdot} \right]$$

$$= \sum_{i \in I} \sum_{j \in J} j \frac{p_{ij}}{p_{i\cdot}} \, p_{i\cdot}$$

$$= \sum_{i \in I} \sum_{j \in J} j \, p_{ij}$$

By fubini thm (2nd part) as previous Calculations that ensures

$$\left\{ \sum_{j \in J} \sum_{i \in I} |j p_{ij}| < \infty \right\}$$

$$= \sum_{j \in J} j \sum_{i \in I} p_{ij}$$

$$= \sum_{j \in J} j \, P[Y=j] = E(Y).$$

Exc: Suppose you toss a fair coin n times independently and get X many heads. then you toss it X more times independently (and also independently of the previous tosses). Let T be the total number of heads that you get. Compute $E(T)$.

Remarks: If we assume that Y has finite p^{th} moment ($p \in \mathbb{N}$), then we define the conditional p^{th} moment of Y given $X=i$ $(=E(Y^p|X=i))$ for each $i \in I$.

Now assume that Y has finite second moment ($\Rightarrow Y$ has finite mean) for each $i \in I$,

define $\varphi(i) = Var(Y|X=i)$

$= $ Variance of the conditional distribution of Y given $X=i$.

$:= E(Y^2|X=i) - (E(Y|X=i))^2$

$Var(Y)$
$= E(Y^2) - (E(Y))^2$

This gives us a map.

$$\varphi : I \longrightarrow \mathbb{R} \text{ defined by}$$

$$\varphi(i) := \mathrm{Var}(Y \mid X = i), \ i \in I$$

$$\begin{tikzcd} \Omega \arrow[r] \arrow[rr, bend right, "\varphi(X)"'] & I \arrow[r] & \mathbb{R} \end{tikzcd}$$

$\mathrm{Var}(Y \mid X) := \varphi(X)$ is a random variable that takes a value $\varphi(i) = \mathrm{Var}(Y \mid X = i)$ whenever X takes the value i.

<u>Thm</u>: Suppose Y has finite second moment. Then the r.v $E(Y \mid X)$ has finite variance and the r.v $\mathrm{Var}(Y \mid X)$ has finite mean, and

$$\mathrm{Var}(Y) = E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(E(Y \mid X)).$$

Proof: Omitted.

Exc: Suppose $X \perp\!\!\!\perp Y$ are both discrete. Then show the following: for each $i \in I$,

(i) $p_{Y|X}(j|i) = p_Y(j)$, $j \in J$. (esy)

(ii) if Y has finite mean then

$$E(Y|X=i) = E(Y);$$

(iii) if Y has finite 2nd moment, then

$$Var(Y|X=i) = Var(Y).$$

Exc: ind $X_1 \sim Poi(\lambda_1)$, $X_2 \sim Poi(\lambda_2)$ $\quad N := X_1 + X_2$

(i) For each $n_0 \in \mathbb{N} \cup \{0\} = \text{Range}(N)$, Compute $Var(X_1|N=n_0)$.

(ii) Show directly (without using the thm)

$$Var(X_1) = E(Var(X_1|N)) + Var(E(X_1|N))$$

Exc: ind $X_1 \sim Bin(n_1, p)$, $X_2 \sim Bin(n_2, p)$ Define, $N := X_1 + X_2 \sim Bin(n_1+n_2, p)$

for $n \in \text{Range}(N) = \{0, 1, 2, \ldots, n_1 + n_2\}$

(a) Find the conditional distribution of X_1 given $N = n$.

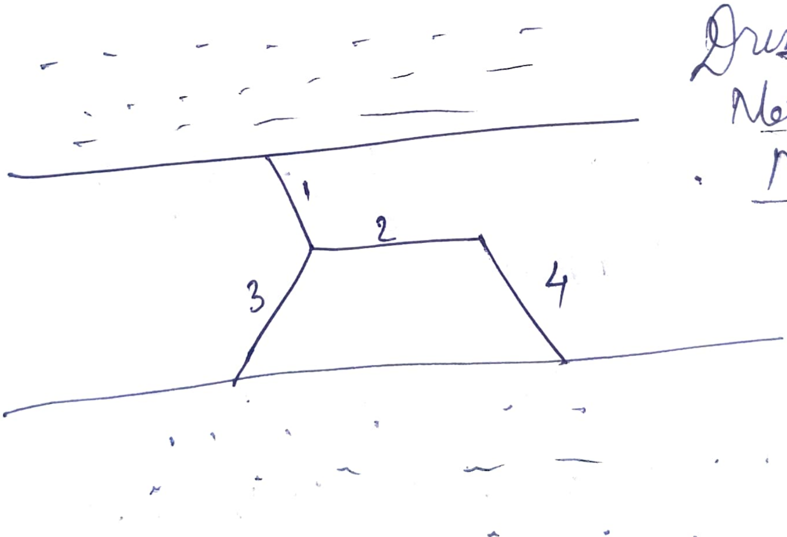
(b) Compute $E(X_1 | N = n)$ for each $n \in \text{Range}(N)$.

(c) Calculate $E(X_1 | N)$

(d) Show, by direct computation, show — (without using thm)

$$E(E(X_1|N)) = E(X_1).$$

Exc:



Drainage Network Model.

Assume:

(i) Each path is open with probability $P = 1/2$

(ii) Path are behaving independently

$X :=$ No of open paths.

$Y := \begin{cases} 1 & \text{water pass through} \\ 0 & \text{o.w} \end{cases}$

(i) find the conditional ~~distribution~~ pmf of X given $Y = 1$. Use this to compute $E(X \mid Y = 1)$.

(ii) find the conditional pmf of X given $Y = 0$. Use this to compute $E(X \mid Y = 0)$.

(iii) Verify directly that

$$E(X) = E[E(X \mid Y)],$$

i.e. Verify that $E(X) = E(X \mid Y = 0)P(Y=0) + E(X \mid Y = 1)P(Y = 1)$

(don't use them). Verify directly

Conditional Distribution in the Jointly Continuous Case

Suppose $\binom{X}{Y}$ is a cont random vector with a joint pdf $f_{X,Y}(x,y)$. From this we compute marginal pdfs

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \text{ and}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

Recall: Range $(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$ and

Range $(Y) = \{y \in \mathbb{R} : f_Y(y) > 0\}$

Suppose $5 \in$ Range (Y).

Que: How to define the conditional distribution of X given Y takes the value 5"?

Problem: $P(Y=5) = 0$.

Therefore ~~Unlike the discrete case~~, we cannot blindly follow the discrete case and define the conditional distribution with the help of conditional ~~pmf~~ probablities.

Take any $a \in \mathbb{R}$

$$\cancel{P(X \le a)} \; P(X \le a \mid Y=5) = \frac{P(X \le a)\,P(Y=5)}{P(Y=5)}$$

Nonsense

$$= \frac{0}{0}$$

Instead of trying to compute $\cancel{P(X \le a)}$ the conditional prob $X \le a$ given $Y=5$, let us compute the conditional probablity of $X \le a$ given $Y \approx 5$.

More ~~for~~ precisely, let us look at.

$$\lim_{\varepsilon \to 0^+} \frac{P(X \le a,\ 5-\varepsilon < Y < 5+\varepsilon)}{P(5-\varepsilon < Y < 5+\varepsilon)}$$

$$= \lim_{\varepsilon \to 0^+} \frac{\int_{5-\varepsilon}^{5+\varepsilon} \int_{-\infty}^{a} f_{X,Y}(x,y)\,dx\,dy}{\int_{5-\varepsilon}^{5+\varepsilon} f_Y(y)\,dy}$$

$$= \lim_{\varepsilon \to 0^+} \frac{\frac{1}{2\varepsilon} \int_{5-\varepsilon}^{5+\varepsilon} \int_{-\infty}^{a} f_{X,Y}(x,y)\,dx\,dy}{\frac{1}{2\varepsilon} \int f_Y(y)\,dy}$$

Assume $f_Y(y)$ is cont $y=5$.

Then $\frac{1}{2\varepsilon} \int_{5-\varepsilon}^{5+\varepsilon} f_Y(y)\,dy \longrightarrow f_Y(5)$ as $\varepsilon \to 0^+$.

(FTC)

Assume that $y \longmapsto \int_{-\infty}^{a} f_{X,Y}(x,y)\,dx$

is continuous $y=5$.

Then by FTC

$$\frac{1}{2\varepsilon} \int_{5-\varepsilon}^{5+\varepsilon} \boxed{\int_{-\infty}^{a} f_{X,Y}(x,y)\,dx}\,dy \longrightarrow \int_{-\infty}^{a} f_{X,Y}(x,5)\,dx$$

Under assumptions,

$$\text{"}P(X \le a \mid Y = s)\text{"}$$

$$:= \lim_{\varepsilon \to 0^+} \frac{P(X \le a,\ s-\varepsilon < Y < s+\varepsilon)}{P(s-\varepsilon < Y < s+\varepsilon)}$$

$$= \frac{\int_{-\infty}^{a} f_{X,Y}(x,s)\, \cancel{dy}\, dx}{f_Y(s)}$$

$$= \int_{-\infty}^{a} \frac{f_{X,Y}(x,s)\, \cancel{dy}}{f_Y(s)}\, dx$$

This should be thought of as Conditional Pdf of X given $Y = s$.

<u>Def^n</u>: Let $\binom{X}{Y}$ be a continuous random vector with joint Pdf with $f_{X,Y}(x,y)$ we compute marginal Pdf of Y, $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$, $y \in \mathcal{Y}$.

Fix $y_0 \in \mathbb{R}$ such that $f_Y(y_0) > 0$.

Then a conditional pdf of X given ~~Y=y~~ $Y = y_0$ (Not $Y = y_0$ but Y takes value close to y_0) is denoted by :-

(i) $f_{X|Y=y_0}(x)$

or ||

(ii) $f_{X|Y}(x|y_0)$

and is defined by

$$f_{X|Y}(x|y_0) = \frac{f_{X,Y}(x, y_0)}{f_Y(y_0)}, \quad x \in \mathbb{R}.$$

<u>Conditional Distn in Joint Cont Case</u>

Let $\binom{X}{Y}$ be a cont random variable with a joint pdf $f_{X,Y}(x,y)$. Compute a marginal pdf $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$, $y \in \mathbb{R}$ of Y. Fix y_0 such that $f_Y(y_0) > 0$.

Then the conditional pdf of X given $Y=y_0$ is defined as-

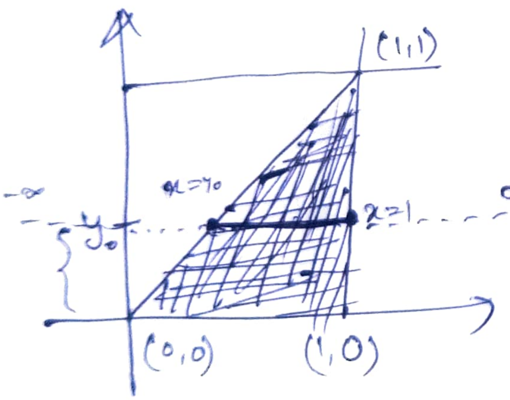
$$f_{X|Y=y_0}(x) = f_{X|Y}(x|y_0) := \frac{f_{X,Y}(x,y_0)}{f_Y(y_0)}, \quad x \in \mathbb{R}.$$

(Can be zero for many values of x).

It is going to be non-zero on the conditional range of X given $Y=y_0$, ie, on the set $\{x \in \mathbb{R} : f_{X,Y}(x,y_0) > 0\}$.

Ex: Let $\binom{X}{Y}$ be a cont random vector with a joint pdf $f_{X,Y}(x,y) = 15xy^2$, $0<y<x<1$.

Compute a conditional pdf of X given Y.



Clearly, Range $(X,Y) = \{(x,y) \in \mathbb{R}^2 : 0<y<x<1\}$

Hence Range $(Y) = (0,1)$.

We want to find the marginal pdf of Y, $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$

fix $y_0 \in (0,1)$. Given $Y=y_0$, the conditional range of X is $(y_0, 1)$.

Also $f_Y(y_0) = \int_{-\infty}^{\infty} f_{X,Y}(x,y_0)\,dx$

$$= \int_{y_0}^{1} f_{X,Y}(x, y_0)\, dx$$

$$= \int_{y_0}^{1} 15 x y_0^2\, dx = 15 y_0^2 \int_{y_0}^{1} x\, dx = 15 y_0^2 \left[\frac{x^2}{2}\right]_{y_0}^{1}$$

$$= \frac{15}{2} y_0^2 (1 - y_0^2)$$

$$> 0 \text{ as } y \in (0,1).$$

Therefore, for each $y_0 \in (0,1)$, the conditional distribution of X given $Y = y_0$ is well defined. Also, the conditional range of X given $Y = y_0$ is well defined. Hence fix $y_0 \in (0,1)$. The conditional range of X given $Y = y_0$ is,

$$f_{X|Y}(x|y_0) = \frac{f_{X,Y}(x, y_0)}{f_Y(y_0)} = \frac{15 x y_0^2}{\frac{15}{2} y_0^2 (1 - y_0^2)} = \frac{2x}{1 - y_0^2},$$

$$x \in (0,1).$$

for each fixed $y_0 \in (0,1)$, the conditional pdf $f_{X|Y}(x, y)$ of X given $Y = y_0$ is well defined and is of the form.

$$f_{X|Y}(x|y_0) = \begin{cases} \frac{2x}{1 - y_0^2} & \text{if } y_0 < x < 1 \\ 0 & \text{o.w} \end{cases}$$

$$\Rightarrow E(X|Y = y_0) = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y_0)\, dx$$

$$= \int_{y_0}^{1} \frac{2x^2}{1-y_0^2}\, dx$$

$$= \frac{2}{3}\, \frac{(x^3)\Big|_{y_0}^{1}}{1-y_0^2}$$

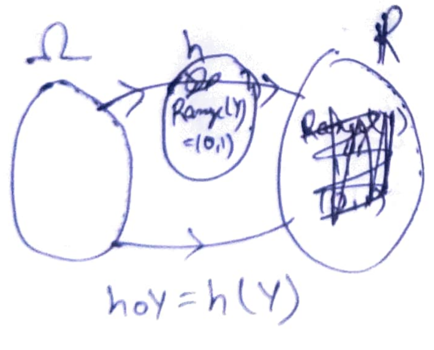
$$\frac{2\times(1-y_0^3)}{3\times(1-y_0^2)} = \frac{2}{3} \times \frac{(1+y_0+y_0^2)}{(1+y_0)}$$

for each $y_0 \in (0,1)$,

$\Longrightarrow$ for each $y_0 \in (0,1)$,

$$h(y_0) := E(X \mid Y = y_0) = \frac{2(1+y_0+y_0^2)}{3(1+y_0)}$$

this defines a map $h : \underset{\substack{\shortparallel \\ \text{Range}(Y)}}{(0,1)} \longrightarrow \mathbb{R}$



$$E(X \mid Y) := h(Y).$$

$$= \frac{2(1+Y+Y^2)}{3(1+Y)}$$

$h(Y)$ is well-defined r.v.

In this example, one can check that (don't its very nasty)

$$E(E(X \mid Y)) = E\left[\frac{2(1+Y+Y^2)}{3(1+Y)}\right] = E(X)$$

Thm: Suppose $\binom{X}{Y}$ is a continous random vector such that X has finite mean. Then -

$E(X|Y)$ is well defined and it also has finite mean $E(E(X|Y)) = E(X)$.

Thm: Suppose $\binom{X}{Y}$ is a continous random vector such that X has finite second moment. Then $E(X|Y)$ has finite ~~mean~~ variance and $Var(X|Y)$ has ~~fin~~ finite mean, and

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$$

Exc: Let Y be a continous r.v with a pdf $f_Y(y) = 2(1-y)$ if $0<y<1$. Also, given $Y=y$, conditionally $X \sim Unif(0, \sqrt{1-y^2})$.

(a) For what values of x_0, is the conditional distribution of Y given $X = x_0$ well defined?

(b) For each such value of x_0, find a conditional pdf of Y given $X = x_0$.

Exc: Suppose $\binom{X}{Y}$ is a cont random vector with a joint pdf $f_{X,Y}(x,y) = \begin{cases} \exp\{-(x-y)\} & \text{if } 0<y<1, x>y \\ 0 & \text{o.w.} \end{cases}$

(a) Verify that $f_{X,Y}$ is a valid joint pdf.

(b) Find marginal pdf's of X and Y.

(c) Find the conditional pdfs $f_{X|Y}$ and $f_{Y|X}$.

(d) For each $y \in (0,1)$ compute $E(X|Y=y)$ ~~Using this,~~ and $Var(X|Y=y)$.

(e) Using (d), compute $E(X|Y)$ and $Var(X|Y)$, and verify directly: (i) $E(X) = E[E(X|Y)]$

(ii) $Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$

Exc: Suppose X, Y are cont r.v.s and $X \perp\!\!\!\perp Y$ with pdf f_X, f_Y resp.

Then show that for each $x_0 \in Range(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$

$$f_{Y|X}(y|x_0) = f_Y(y) \quad \forall\ y \in \mathbb{R}$$

In other words the conditional distn of Y given $X=x_0$ is same as the unconditional distn of Y.

Conditional Distribution for bivariate Normal

~~Let~~ Suppose

Setup: $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim Nor\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right)$

fix $x_1 \in R = \text{Range}(X_1)$ $\Rightarrow X_1 \sim N(\mu_1, \sigma_{11})$

Goal: To compute the conditional distribution of X_2 given $X_1 = x_1$.

~~Sol~~ **Step 1:** Find α such that

$$Cov(X_2 - \alpha X_1, X_1) = 0.$$

$$\begin{aligned} 0 &= Cov(X_2 - \alpha X_1, X_1) \\ &= Cov(X_1, X_2) - \alpha\, Cov(X_1, X_1) \\ &= Cov(X_1, X_2) - \alpha\, Var(X_1) \end{aligned}$$

$$\Rightarrow \alpha\, Var(X_1) = Cov(X_1, X_2)$$

$$0 = \alpha \sigma_{11} - \sigma_{12} \Leftrightarrow \alpha = \frac{\sigma_{12}}{\sigma_{11}}$$

$$\boxed{\text{Conclusion } \cancel{\text{is that}}: Cov\left(X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1, X_1\right) = 0}$$

Step 2 :

$\left(X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1, X_1\right)^T$ is a bivariate normal random vector.

Note that

$$\begin{pmatrix} X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1 \\ X_1 \end{pmatrix} = \underbrace{\begin{pmatrix} -\frac{\sigma_{12}}{\sigma_{11}} & 1 \\ 1 & 0 \end{pmatrix}}_{= B} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1 \\ X_1 \end{pmatrix} = B \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad B \in \mathbb{R}^{2\times 2}$$

B is non singular. ~~Therefore~~

Also $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ follows a bivariate normal distribution. Therefore, $\left(X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1, X_1\right)^T$ also follows a bivariate normal distn.

Step 3: Set Step ① + Step ②

$$\Rightarrow \left(X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1\right) \perp\!\!\!\perp X_1 .$$

Step 4: Step 3 + Exe

$\Rightarrow$ Given $X_1 = x_1$, the conditional distribution of $X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1 = X_2 - \frac{\sigma_{12}}{\sigma_{11}} x_1$ is same as the unconditional (marginal) dist^n of $X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1$.

Step 5: Unconditionally / Marginally

$$X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1 \sim N\left(\mu_2 - \frac{\sigma_{12}}{\sigma_{11}} \mu_1 , \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)$$

$$E\left(X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1\right) = E(X_2) - \frac{\sigma_{12}}{\sigma_{11}} E(X_1)$$

$$= \mu_2 - \frac{\sigma_{12}}{\sigma_{11}} \mu_1$$

$$Var\left(X_2 - \frac{\sigma_{12}}{\sigma_{11}} X_1\right) = Var(X_2) + \frac{\sigma_{12}^2}{\sigma_{11}} Var(X_1)$$
$$- 2 \frac{\sigma_{12}}{\sigma_{11}} Cov(X_1, X_2)$$

Step 6: ~~Step 4 + Step 5~~ $\Rightarrow$

$$= \sigma_{22} + \frac{\sigma_{12}^2}{\sigma_{11}^2}\sigma_{11} - 2\frac{\sigma_{12}}{\sigma_{11}}\sigma_{12}$$

$$= \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}.$$

Step 7: Step 6 $\Rightarrow$ Given $X_1 = x_1$ conditionally

$$X_2 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1),\ \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)$$

Thm: Suppose $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}\right)$

then for each fixed $x_1 \in \mathbb{R}$, given that $X_1 = x_1$ conditionally

$$X_2 \sim N\left(\mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1 - \mu_1),\ \sigma_{12} - \frac{\sigma_{12}^2}{\sigma_{11}}\right)$$

Pf: Upon deth

Cor: $E(X_2|X_1=x_1) = \mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(x_1-\mu_1) \quad \forall x_1 \in \mathbb{R}$

↑

Your best guess for X_2 when it is given that $X_1=x_1$.

(Regression line of X_2 on X_1).

$$Var(X_2|X_1=x_1) = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}} \quad \forall x_1 \in \mathbb{R}$$

Remark: ① ~~Variance~~ $Var(X_2|X_1=x_1) \leq Var(X_2)$ with equality iff $X_1 \perp\!\!\!\perp X_2$. This is because, if we know the value of X_1, we predict the value of X_2 with more precision.

② $E(X_2|X_1) = \mu_2 + \frac{\sigma_{12}}{\sigma_{11}}(X_1-\mu_1)$

$$Var(X_2|X_1) = \sigma_{22} - \frac{\sigma_{12}^2}{\sigma_{11}}$$

Exc: Directly verify that

$$E(X_1) = E(X_2) \quad \text{and}$$

$$Var(X_2) = E[Var(X_2 | X_1)] + Var(E(X_2|X_1))$$

LAST CLASS

Convergence of Random Variables

Setting: Suppose $\{X_n\}_{n\geq 1}$ is a sequence of random variables defined on the same sample space and X is another such random variable. This means that we have a sample space Ω, a σ-field $\mathcal{A}$ on Ω. $\mathcal{A}$ = set of all events. and a probability $P : \mathcal{A} \to [0,1]$. In other words, we have a probability space $(\Omega, \mathcal{A}, P)$.

Moreover for ~~all~~ each $n \geq 1$,

$X_n : \Omega \to \mathbb{R}$ is a r.v and

$X : \Omega \to \mathbb{R}$ is also a r.v.

This means for each $n \geq 1$, $\{\omega : X_n(\omega) \in (a,b)\} \in \mathcal{A}$.

$$\underset{[X \in (a,b)]}{\{\omega : X(\omega) \in (a,b)\}} \in \mathcal{A}.$$

$\forall\ (a,b) \subseteq \mathbb{R}$.

Defⁿ: We say that X_n converges to X almost surely / Almost everywhere if

$$P\left(\{\omega \in \Omega : X_n(\omega) \not\to X(\omega)\}\right) = 0$$

---

* **Remark:** It can be shown that

$$\{\omega \in \Omega : X_n(\omega) \not\to X(\omega)\} \in \mathcal{A}.$$

(Too Adv, Dont Try).

Notation: $X_n \xrightarrow{a.s} X$ or $X_n \xrightarrow{a.e} X$

* **Remark:** ① $X_n \xrightarrow{a.s} X$ means that except for a subset of ~~with~~ Ω with probability zero $X_n \longrightarrow X$ pointwise.

② $X_n \xrightarrow{a.s} X \iff P[X_n \not\to X] = 0 \iff P[X_n = X] = 1.$

Almost Sure Convergence is also called Convergence with probability one.

Defⁿ: We say that X_n converges to X in probability if for all $\varepsilon > 0$

$$\cancel{P\left[|X_n - X| > \varepsilon\right]}$$

$$P\left[|X_n - X| > \varepsilon\right] \longrightarrow 0$$

as $n \to \infty$

or $P\left[\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\}\right] \longrightarrow 0$ as $n \to \infty$.

<u>Notation</u>: $X_n \xrightarrow{P} X$ or $X_n \xrightarrow{P} X$.

<u>Def^n</u>: We say that X_n converges to X in distribution/law if

$$(X_n \le a) = F_{X_n}(a) \to F_X(a) = P(X \le a)$$

as $n \to \infty$ for all continuity point of F_X.

<u>Notation</u>: $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{\mathcal{L}} X$.

<u>Ex</u>: Suppose $X_n \equiv \frac{1}{n}$ for all $n \geq 1$.

we expect $\boxed{X_n \xrightarrow{d} X \equiv 0.}$

this actually happens acc. to def^n

The cdf of X_n

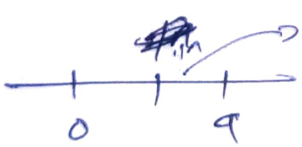
$$\cancel{F(X_n)}\ F_{X_n}(a) = P[X_n \le a]$$
$$= \begin{cases} 0, & a < \frac{1}{n} \\ 1, & a \geq 1/n \end{cases}$$

the cdf of X is

$$F_X(a) = \begin{cases} 0 & \text{if } a < 0 \\ 1 & \text{if } a \geq 0 \end{cases}$$

So, clearly for all $a<0$ $F_{X_n}(a) = 0 \longrightarrow 0$
as $n \to \infty$

Also for all $a>0$, $\cancel{F_X(X_n)}$ $F_{X_n}(a)$



after some stage,

$$\cancel{F(X_n)}\ F_{X_n}(a) = 1$$

So, $F_{X_n}(a) = 1$ as $n \to \infty$
$= F_X(a)$.

However, $F_{X_n}(0) = 0 \not\longrightarrow 1 = F_X(0)$.

for any r.v Y

$F_Y(b) = P[Y \leq b]$ is right Continous.

$\lim_{y \to b^-} F_Y(y) = P[Y < b] = P[Y \leq b]$

$= F_Y(b)$

if and only if $P[Y = b] = 0$

ie | F_Y is cont at b

iff $P[Y = b] = 0$.

We showed that $F_{X_n}(a) \to F_X(a)$

for all $a \in \mathbb{R} \setminus \{0\}$ = set of

all Continuity points of F_X.

$\Rightarrow F_{X_n}(a) \to F_X(a) \;\forall$ Continuity points

of F_X.

$\Rightarrow X_n \xrightarrow{d} X$.

* <u>Remark</u>: If X is a continuous random variable, then $X_n \xrightarrow{d} X \iff F_{X_n}(a) \to F_X(a)\ \forall\, a \in \mathbb{R}$.

i.e., $X_n \to d \iff F_{X_n} \to F_X$ pointwise.

$$\begin{tikzcd}
\boxed{X_n \overset{a.s}{\Longrightarrow} X} \arrow[r, Rightarrow] & \boxed{X_n \xrightarrow{P} X} \arrow[l, Rightarrow, bend left, "\text{subsequently}"] \arrow[d, Rightarrow] \\
 & \boxed{X_n \xrightarrow{d} X} \arrow[u, Rightarrow, bend right, "\text{subsequently}"']
\end{tikzcd}$$

Strongest

subsequently

subsequently

Weakest

<u>Exc</u>: $X_n \xrightarrow{d} 0$ (0 can be replaced by any real)

$\Rightarrow X_n \xrightarrow{P} 0$.

Weak Law of Large Numbers (WLLM)

Suppose $\{X_n\}_{n \geq 1}$ is a sequence of iid random variables

for each integer ($k \geq 2$, $X_1, X_2 \ldots X_k$ are iid).

each having finite mean μ. Then

$$\bar{X}_n := \frac{X_1 + X_2 \ldots X_n}{n} \xrightarrow{P} \mu.$$

This means $\forall \, \varepsilon > 0$.

$$P\left[\,|\bar{X}_n - \mu| > \varepsilon\right] \longrightarrow 0 \quad \text{as } n \to \infty$$

or, ~~for now~~

In other words for large n $\bar{X}_n \approx \mu$ with high probability.

Sample mean (of an iid sample) is a good estimator of population mean.

Thm : (Special case of weak law of large numbers)

Suppose $\{X_n\}_{n\geq 1}$ is a sequece of iid r.v.s each having finite mean μ and finite varience σ^2. Then $\overline{X}_n \xrightarrow{P} \mu$.

Proof: $\boxed{\text{Markovs Inequality}}$

$$\Downarrow$$

$$\boxed{\text{Chebychevs Inequality}} \Longrightarrow \boxed{\text{Thm}}$$

Markovs Inequality: Suppose X is a non-negative randomvariable with finite mean.. Then

Then,

$$P(X \geq a) \leq \frac{E(X)}{a} \quad \forall\ a > 0.$$

In particular if Y is any r.v with finite mean, then $P[|Y| > a] \leq \frac{E(|Y|)}{a}$

$\forall\ a > 0$.

Proof of Markovs Inequality: Fix $a > 0$

$$X = \underbrace{X I_{(x<a)}}_{\geq 0} + \underbrace{X I_{(x \geq a)}}_{\geq 0}$$

$$E(X) = \underbrace{E(X I_{(x<a)})}_{\geq 0} + E(X I_{(X \geq a)})$$

$$\geq E(X I_{(X \geq a)})$$

$$E(X) \geq E[X I_{(X \geq a)}]$$

$$X I_{(X \geq a)} \geq a I_{(X \geq a)}$$

$$= \begin{cases} X & \text{if } X \geq a \\ 0 & \text{if } X < a \end{cases} \qquad \begin{cases} a & \text{if } X \geq a \\ 0 & \text{if } x < a. \end{cases}$$

Therefore $E(X) \geq E(X I_{X \geq a})$

$$\geq \cancel{a}\, E(\cancel{a} I_{(X \geq a)})$$

$$= \cancel{a} \quad a E(I_{X \geq a})$$

$= a\,P(X \geq a)$

$$\boxed{\Rightarrow P(X \geq a) \leq \frac{E(X)}{a}}$$ Markov's Inequality

Suppose X is a random variable with finite mean μ and finite variance σ^2.

Fix $\varepsilon > 0$,

has finite mean

$$P\left[|X-\mu| > \varepsilon\right] = P\left[(X-\mu)^2 > \varepsilon^2\right]$$

$$\underset{\text{Markov's ineq}}{\leq} \boxed{\frac{E\left[(X-\mu)^2\right]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}}$$

Chebychev's inequality:

Suppose X is a r.v with finite mean μ and finite variance σ^2.

for all $\varepsilon > 0$ $P\left[|X-\mu| > \varepsilon\right]$

$$\leq \frac{\sigma^2}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}.$$

Proof:

"Chebyshev's inequality $\Rightarrow$ Thm"

fix $\varepsilon>0$. To show: $P[|\bar{X}_n-\mu|>\varepsilon] \longrightarrow 0$ as $n\to\infty$.

$$E(\bar{X}_n) = E\left(\frac{X_1+X_2\cdots X_n}{n}\right)$$

$$= \frac{1}{n}\left(E(X_1)+E(X_2)\cdots E(X_n)\right)$$

$$= \frac{n\mu}{n} = \mu$$

$$Var(\bar{X}_n) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \overset{ind}{=} \frac{1}{n^2} \cancel{Var} \sum_{i=1}^{n} Var(X_i)$$

$$= \frac{n\sigma^2}{n^2}$$

$$= \frac{\sigma^2}{n}$$

Summary: $E(\bar{X}_n) = \mu$

and $Var(\bar{X}_n) = \frac{\sigma^2}{n}$

Therefore

$$0 \le P[|\bar{X}_n - \mu| > \varepsilon] \le \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

$$\Rightarrow P[|\bar{X}_n - \mu| > \varepsilon] \to 0 \text{ as } n \to \infty.$$

$$\Rightarrow \bar{X}_n \xrightarrow{P} \mu.$$

Strong Law of Large Numbers (SLLN)

If $\{X_n\}_{n\geq 1}$ is a sequence of iid random variable each having finite mean μ then,

$$\bar{X}_n \xrightarrow{a.a} \mu.$$

* **Remark:** Strong law $\Rightarrow$ Weak law.

SLLN $\Rightarrow$ WLLN

Suppose $X_1, X_2 \ldots$ is a seq of iid random variables with finite mean μ and finite varience σ^2.

Then we know that $\overline{X}_n - \mu \xrightarrow{as} 0$

(By SLLN)

Qn: At what rate does $\overline{X}_n - \mu$ go to zero?

Ans: $\overline{X}_n - \mu \simeq \frac{1}{\sqrt{n}}$ " $\longleftarrow$? "

(CLT) Central Limit Theorm: Suppose $\{X_n\}_{n \geq 1}$ is a seq of iid r.v.s with finite mean μ and finite varience σ^2. Then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} X \sim N(0,1)$$

as $n \to \infty$.

$$\parallel$$

$$\frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}$$

$$\frac{n(\bar{X}_n - \mu)}{\sigma} = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)}{\sigma} \times \frac{\sqrt{n}}{\sqrt{n}}$$

$$= \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

~~Suppose~~ <u>Special Case</u>: Suppose $X_1, X_2 \ldots \overset{iid}{\sim} B(p)$ $(p \in (0,1))$

$$\sum_{i=1}^{n} X_i \sim Bin(n, p)$$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \longrightarrow N(0,1)$$

$$\Rightarrow \frac{\sum_{i=1}^{n} X_i - np}{\sqrt{p(1-p)n}} \xrightarrow{d} N(0,1).$$

Therefore when $Y_n \sim Bin(n, p)$

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1).$$

$$\Rightarrow \quad Y_n \overset{\text{approx}}{\sim} N\left(np, np(1-p)\right)$$

when n is large.