•) For $u, v \in \mathbb{R}^n$, $A \in \mathbb{R}_n$

$$\tilde{A}(u, v) := u^T A v$$

<u>Special case:</u> when $A = Id \Rightarrow \tilde{B}(u, v) = u^T v = u \cdot v$

• In bilinear : $\tilde{B}(\lambda u, \lambda v) = \lambda^2 B(u, v)$

$\tilde{B}$ is linear in u and in v but not (u, v)

e.g : $\tilde{B}(u, \lambda v_1 + v_2) = \lambda \tilde{B}(u, v_1) + \tilde{B}(u, v_2)$

$(\forall u, v_1, v_2 \in E)$

$$\tilde{A}(u+v, w+r) = \tilde{A}(u+v, w) + \tilde{A}(u+v, r)$$
$$= \tilde{A}(u, w) + \tilde{A}(v, w) + \tilde{A}(u, r) + \tilde{A}(v, r)$$

•) $(X, Y) \mapsto Cov(X, Y)$ is bilinear

<u>Proof</u> :

•) We have $Cov(X, Y) = Cov(Y, X) \Rightarrow Cov$ is symmetric

•) We only need to prove the linearity wrt X :

① $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$

② $Cov(\lambda X, Y) = \lambda Cov(X, Y)$

① : $Cov(X_1 + X_2, Y) = \mathbb{E}((X_1 + X_2) \cdot Y) - \mathbb{E}(X_1 + X_2) \cdot \mathbb{E}(Y)$

$= \mathbb{E}(X_1 Y) + \mathbb{E}(X_2 Y) - \mathbb{E}(X_1) \cdot \mathbb{E}(Y) - \mathbb{E}(X_2)\mathbb{E}(Y)$

$= Cov(X_1, Y) + Cov(X_2, Y)$

② : $Cov(\lambda X, Y) = \mathbb{E}(\lambda X \cdot Y) - \mathbb{E}(\lambda X) \cdot \mathbb{E}(Y)$

$= \lambda \mathbb{E}(XY) - \lambda \mathbb{E}(X) \cdot \mathbb{E}(Y) = \lambda Cov(X, Y)$

• <u>Note</u> : $\mathbb{E}$ is linear $\Rightarrow Cov$ is bilinear

•) $Var(X) = \mathbb{E}(|\vec{X}|^2) = \mathbb{E}((X - \mathbb{E}(X))^2) = Cov(X, X)$

Var is not linear since:

• $Var(2X) = Cov(2X, 2X) = 4\,Cov(X, X) = 4\,Var(X)$

• $Var(X+Y) \neq Var(X) + Var(Y)$

$\quad\;\; \parallel$

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

$= Var\,X + Var(Y) + 2Cov(X, Y)$

•) By thogoras Theorem: $Var(X+Y) = Var\,X + Var\,Y$

$\Leftrightarrow$ X, Y uncorrelated $(Cov(X, Y) = 0)$

•) Hilbert Space (Geometry): $u, v \in \mathbb{R}^n$, $u \cdot v = 0 \Leftrightarrow u \cdot u + v \cdot v$
$= (u+v) \cdot (u+v)$

•) Cov is similar to dot product except:

• $u \cdot u = 0 \Rightarrow u = 0$

• $Cov(X, X) = 0 \Rightarrow X$ is self independence

Since $Cov(X, X) = \mathbb{E}(X \cdot X) - \mathbb{E}^2(X)$ (1)

Since X is self-inde $\Rightarrow$ the image of X is $\{a_0\}$, $a_0 \in \mathbb{R}$

$\Rightarrow Cov(X, X) = a_0^2 - a_0^2 = 0$

•) $X = max(X_1, X_2)$, $Image(X) = \{1, 2, \dots, 6\}$

$$P_X(t) = \frac{2 \cdot t - 1}{36}$$

•) For general X:

dis	cont
pmf	pdf

CDF

CDF of X is $F_X(t) = P(X \le t)$

• X cont $\Rightarrow F_X(t) = \int_{-\infty}^{t} f_X \overset{FTC}{\Rightarrow} F_X' = f_X$ cdf $\underset{int}{\overset{diff}{\rightleftarrows}}$ pdf

• X disc $\Rightarrow F_X(t) = \sum_{a \le t} p_X(a)$

:) If X is $C^0 \iff F_X$ is C^0 and diff

:) X disc, $Y = 2X + 3$

$\forall\, t \in Im(Y),\ t = 2t_X + 3 \quad (t_X \in Im(X))$

$\Rightarrow p_Y(Y = t) = p_X\left(\frac{t_X - 3}{2}\right)$

$\mathbb{E}(Y) = \mathbb{E}(2X + 3) = 2\mathbb{E}(X) + 3$

$$\begin{aligned} Var(Y) &= Var(2X+3) \\ &= Cov(2X+3,\ 2X+3) \\ &= Var(2X) + Var(3) + 2\,Cov(2X, 3) \\ &= Var(2X) = 4\,Var(X) \end{aligned}$$

Problem: X cont, known $(f_X, \mathbb{E}X, \mathrm{var}\, X)$

$Y = 2X + 3$, find f_Y?

Solution:

- $F_Y(t) = P(Y \le t) = P(2X+3 \le t) = P\left(X \le \frac{t-3}{2}\right)$

$$= F_X\left(\frac{t-3}{2}\right)$$

$$\Rightarrow f_Y(t) = F'_X\left(\frac{t-3}{2}\right) \cdot \frac{1}{2} = f_X\left(\frac{t-3}{2}\right) \cdot \frac{1}{2}$$

- IID (inde, identically distributed): X and Y are iid when they are inde and $P(X \in [a,b]) = P(Y \in [a,b])\ \forall a, b \in \mathbb{R}$

- $\mathrm{Var}\left(\frac{X+Y}{2}\right) = \mathrm{Cov}\left(\frac{X+Y}{2}, \frac{X+Y}{2}\right) = \frac{1}{4}\mathrm{Var}(X+Y)$

$$\underset{=}{\text{Pythagoras}}\ \frac{1}{4}(\mathrm{Var}\, X + \mathrm{Var}\, Y) = \frac{2\,\mathrm{Var}\, X}{4} = \frac{\mathrm{Var}\, X}{2}$$

- $\mathbb{E}_{Freq}(X) = \lim_{\#\text{tries} \to +\infty}$ (Average over all tries for X)

Modern notation: $\mathbb{E}_{Freq}(X) = \lim_{n \to +\infty} \underbrace{\frac{X_1 + \dots + X_n}{n}}_{\text{Sample mean}}$ where (X_n) are iid

Clones of X

- Law of large numbers: $\mathbb{E}_{Freq} = \mathbb{E}$

$$P\left(\frac{X_1 + \dots + X_n}{n} \xrightarrow{n \to +\infty} \mathbb{E}(X_1)\right) = 1$$

•) Indicator function : Let A be an event (a subset of Ω), define

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

$$\mathbb{E}(1_A) = 0 \cdot P(1_A = 0) + 1 \cdot P(1_A = 1) = 1 \cdot P(A) = P(A)$$

•) Monte Carlo method : Approximate $\mathbb{E}_{\text{Freq}} X$ by computer simulations
let # tries be large

•) Markov's inequality : $\forall a \geq 0$

$$\mathbb{E}|X| \geq a P(|X| \geq a)$$

Proof : $\mathbb{E}|X| =$

$$\sum_{w : |X(w)| \geq a} |X(w)| \cdot P(X = w) + \sum_{w : |X(w)| < a} |X(w)| P(X = w)$$

$$\geq a \sum_{w : |X(w)| \geq a} P(X = w)$$

•) Outliers : the cases that are rare
e.g: $\mathbb{E}(X) = 2$

$$P(|X| \geq 1000) \leq \frac{\mathbb{E}|X|}{1000} = \frac{2}{1000}$$

•) Chebyshev's inequality : $\text{var}(X) \geq a^2 \cdot P(|X - \mathbb{E}(X)| \geq a)$

$\forall a \geq 0$

Proof : $\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2 \right) \geq a^2 P(|X - E(X)| \geq a)$

•) Sample mean :

$(X_n)_{n=1}^{\infty}$ iid clones of X

$$X_n = \frac{X_1 + \dots + X_n}{n}$$

↓

Sample mean

• $\mathbb{E}\,\overline{X}_n = \dfrac{\mathbb{E}(X_1) + \dots + \mathbb{E}(X_n)}{n} = \dfrac{\mathbb{E}(X)\cdot n}{n} = \mathbb{E}(X)$

• $\operatorname{Var} \overline{X}_n = \dfrac{\operatorname{Var}(X_1 + \dots + X_n)}{n^2} = \dfrac{\operatorname{Var}(X_1) + \dots + \operatorname{Var}(X_n)}{n^2}$

$$= \frac{n \cdot \operatorname{Var} X}{n^2} = \frac{\operatorname{Var} X}{n} \xrightarrow[n \to +\infty]{} 0$$

• $\mathbb{P}\left(\left| \overline{X} - \mathbb{E}\overline{X}_n \right| \geq a \right) \leq \dfrac{\operatorname{Var}(\overline{X})}{a^2} \quad (\forall a > 0)$

$$\underline{\mathbb{P}\left(\left| \overline{X} - \mathbb{E}\overline{X}_n \right| \geq a \right)} \xrightarrow[n \to \infty]{} 0 \qquad \frac{\operatorname{Var}(\overline{X})}{a^2} \xrightarrow[n \to \infty]{} 0$$

(Squeeze Theorem)

•) Weak law of large numbers :

Small varience → concentration near center ($\overline{X}$ goes closer to $\mathbb{E}(X)$)

- iid : best case

worst case : $X_1 = X_2 = \dots = X$

$$\bar{X}_n = X \;,\; \operatorname{Var} \bar{X}_n = \operatorname{Var} X \not\to 0$$

- <u>Cauchy-Schwarz inequality</u>: $|u.v|^2 \leq (u.u)(v.v) \ \forall u, v$

Trick : $\forall t \in \mathbb{R}$

$$(u+tv).(u+tv) \geq 0$$

$$\Rightarrow \vec{u}.\vec{u} + 2t(\vec{u}.\vec{v}) + t^2 \vec{v}.\vec{v} \geq 0$$

$$\Rightarrow \quad (\vec{u}\vec{v})^2 - \vec{u}.\vec{u}\ \vec{v}.\vec{v} \leq 0 \quad \Rightarrow (\vec{u}.\vec{v})^2 \leq (\vec{u}.\vec{u})(\vec{v}.\vec{v})$$

- In Prob & Stats : $|\operatorname{Cov}(X,Y)|^2 \leq \operatorname{Var}(X) \operatorname{Var}(Y)$
$\forall$ RV X, Y

$$\Rightarrow -1 \leq \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X) \operatorname{Var} Y}} \leq 1$$

$$\Downarrow$$

$$\operatorname{Cor}(X,Y) \ \left(\sim \cos(\widehat{\vec{u}, \vec{v}}) \right)$$

- <u>Bernoulli and binomial</u>

- Bernoulli RV : X where $\operatorname{Im} X = \{0, 1\}$

- $\mathbb{E}(X) = 0.\,P(X=0) + 1.\,P(X=1)$
$= p$

•) $\mathbb{E}(X^2) = 0^2 \cdot \mathbb{P}(X=0) + 1^2 \cdot \mathbb{P}(X=1)$

$= p$

• Binomial : Z with para p and n is the sum of n inde Bernoulli variables with para p

$Z = X_1 + X_2 + \dots + X_n$

•) $X \sim Bin(n, p)$

$\mathbb{P}(X=k) = \mathbb{P}(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k}$

•) Geometric RV : Let $X_1, X_2, \dots$ be a sequence of Bernoulli vars with parameter p. Let T be the 1st one we get a 1, then T is called a geometric RV with parameter p

$\mathbb{P}(T=k) = (1-p)^{k-1} p$

$\mathbb{E}(T) = \sum_{a=1}^{\infty} a \, p_T(a) = \sum_{a=1}^{\infty} a(1-p)^{a-1} p = p \sum_{a=1}^{\infty} a(1-p)^{a-1}$

Let $r \in [0,1]$

$\sum_{a=1}^{\infty} a r^{a-1} = \sum_{a=1}^{\infty} (r^a)' = \left(\sum_{a=1}^{\infty} r^a\right)' = \left(\frac{r}{1-r}\right)' = \frac{(1-r)+r}{(1-r)^2} = \frac{1}{(1-r)^2}$

$= \frac{1}{p^2}$

When $r = 1 \Rightarrow p = 0 \Rightarrow T = \infty$

•) Uniform RV : A RV is called exponential with parameter $\lambda > 0$ when the pdf

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

$\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

•) $\int_{\mathbb{R}} e^{-t^2+4t+5} \, dt = \int e^{-(t-2)^2+9} \, dt$

•) Normal distribution :

• Gaussian func : $f(t) = e^{-t^2}$

• $X \sim \mathcal{N}(\mu, \sigma^2)$, $f_X(t) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$

① $\int_{\mathbb{R}} f_X = 1$

$$\int \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \, dt = 1$$

standard normal variable

② let $Z \sim \mathcal{N}(0,1)$, $f_Z(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

$Z \sim \frac{X-\mu}{\sigma}$

Check : $X = \sigma Z + \mu$

know f_Z , find $f_{\sigma Z + \mu}$

$F_X(t) = \mathbb{P}(\sigma Z + \mu \le t) = \mathbb{P}\left(Z \le \frac{t-\mu}{\sigma}\right) = F_Z\left(\frac{t-\mu}{\sigma}\right)$

Diff in t $f_X(t) = f_Z\left(\frac{t-\mu}{\sigma}\right) \cdot \frac{1}{\sigma}$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \Rightarrow \text{Correct}$$

- $\mathbb{E}(Z) = 0$, $\operatorname{Var} Z = 1$

$$X = \sigma Z + \mu \Rightarrow \mathbb{E}(X) = \sigma$$

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

- Gaussian = heat kernel = eigenfunction of Fourier transform

- Random vectors: For 2 RV X, Y, we can define a function

$$(X, Y) : \Omega \to \mathbb{R}^2 \qquad , (X, Y) \text{ is a random vector}$$
$$w \mapsto (X(w), Y(w))$$

- For discrete: $p_{X,Y}(a, b) = P(X = a, Y = b) = P((X, Y) = (a, b))$

If X, Y is inde $\Rightarrow P(X = a, Y = b) = P(X = a) \cdot P(Y = b)$

- Inde $\Leftrightarrow p_{X,Y}(a, b) = p_X(a)\, p_Y(b) \quad \forall a, b$

<u>Note</u>: $p_{X,Y} = p_X p_Y$ is wrong notation since it is ~ pointwise product

$$(f \otimes g)(a, b) = f(a) f(b) \rightarrow \text{tensor product}$$

$\Rightarrow$ Correct notation: $p_{X,Y} = p_X \otimes p_Y$

- For C^0: X, Y C^0; $f_{X,Y}$: pdf

$$\int_c^d \int_a^b f_{X,Y}(t, s)\, dt\, ds = P(a \leq X \leq b, c \leq Y \leq d)$$

$F_{X,Y}$: cdf ; $F_{X,Y}(b, d) = P(X \leq b, Y \leq d)$

$$= \int_{-\infty}^{d}\int_{-\infty}^{b} f_{X,Y}(t,s)\, dt\, ds$$

$$\Rightarrow f_{X,Y} = \partial_{12} F_{xy} = \partial_{21} F_{xy}$$

- Inde for $\mathcal{C}^0$: $P(X \leq b, Y \leq d) = P(X \leq b) \cdot P(Y \leq d)$

$$\Rightarrow F_{X,Y}(b,d) = F_X(b) \cdot F_Y(d)$$

$$\Rightarrow \partial_b \partial_d F_{X,Y}(b,d) = F_X'(b)\, F_Y'(d)$$

$$\Rightarrow f_{X,Y}(b,d) = f_X(b)\, f_Y(d) \quad \forall b, d \in \mathbb{R}$$

$$\Rightarrow f_{X,Y} = f_X \otimes f_Y$$

•) Addition : Let X, Y : $\begin{cases} \text{discrete, integer-valued} \\ \text{inde} \end{cases}$

$$P(X+Y=k) = \sum_{a \in \mathbb{Z}} P(X=a, Y=k-a)$$

$$= \sum_{a \in \mathbb{Z}} P(X=a) \cdot P(Y=k-a)$$

$$= \sum_{a \in \mathbb{Z}} p_X(a)\, p_Y(k-a)$$

$$= \left(p_X \overset{\mathbb{Z}}{*} p_Y\right)(k) \rightarrow \text{convolution product}$$

•) Convolution product : $f, g : \mathbb{Z} \rightarrow \mathbb{R}$, then $\sum_{a \in \mathbb{Z}} f(a)\, g(k-a)$

is a convolution product, denoted as $(f \overset{\mathbb{Z}}{*} g)(k)$

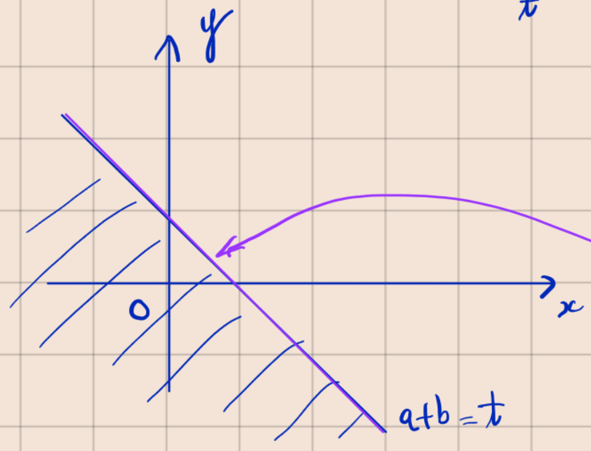
Ex : $g(x) = \begin{cases} 1/3 & \text{for } x \in \{-1,0,1\} \\ 0 & \text{otherwise} \end{cases}$

$$f * g(x) = \frac{1}{3}\left(f(x-1) + f(x) + f(x+1)\right)$$

- **Note** : Convolution product form needs independence

- For $\mathcal{C}^0$ RV: X, Y $\begin{cases} \mathcal{C}^0, \text{ real-valued} \\ \text{independt} \end{cases}$

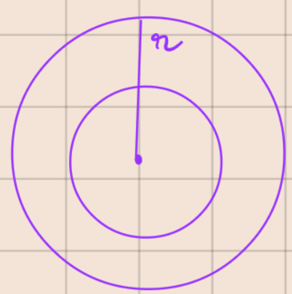
$$F_{X+Y}(t) = \mathbb{P}(X + Y \leq t) = \iint_{D_t} f_{X,Y} = \iint_{D_t} f_X \otimes f_Y$$



$$\Rightarrow f_{X+Y}(t) = \partial_t\left(\iint_{D_t} f_X \otimes f_Y\right) = \lim_{h \to 0} \frac{1}{h}\left(\iint_{D_{t+h}} - \iint_{D_t}\right)(f_X \otimes f_Y)$$

$$= \int_{a+b=t} f_X \otimes f_Y(a,b)$$

$\boxed{\text{Coarea formula}}$



$A(r)$: area of a circle

$$A'(r) = \lim_{h \to 0} \frac{1}{h}\left(A(r+h) + A(r)\right) = \text{perimeter}$$

•) Stokes theorem : $\displaystyle\int_{\partial A} w = \int_A dw$

$$\int_{[a,b]} f' = f(b) - f(a) = \int_{\partial [a,b]} f$$

•) X is normal, prove that $3X+4$ is normal

$$F_{3X+4}(t) = P(3X+4 \leq t) = P\left(X \leq \frac{t-4}{3}\right) = F_X\left(\frac{t-4}{3}\right)$$

$$\rightarrow f_{3X+4}(t) = f_X\left(\frac{t-4}{3}\right)\frac{1}{3}$$

$$f_X(t) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2\sigma^2}(t-\mu)^2\right)$$

$$\Rightarrow f_{3X+4}(t) = \frac{1}{\sqrt{2\pi}\,3\sigma}\exp\left(-\frac{1}{2\cdot 9\sigma^2}\left(t-4-3\mu\right)^2\right)$$

$$\Rightarrow 3X+4 \sim \mathcal{N}(3\mu+4,\ 9\sigma^2)$$

•) $\displaystyle\int_{\mathbb{R}} \exp\left(-ax^2+bx+c\right)dx = \int_{\mathbb{R}} \exp\left(-a\left(x-\frac{b}{2a}\right)^2 + c + \frac{b^2}{4a}\right)dx$

$$= \exp\left(c+\frac{b^2}{4a}\right)\cdot\sqrt{\frac{\pi}{a}}$$

$$-ax^2+bx+c$$

$$-a\left(x^2-\frac{b}{a}x-\frac{c}{a}\right)$$

$$-a\left(\left(x-\frac{b}{2a}\right)^2 - \frac{b^2}{4a^2} - \frac{c}{a}\right)$$

$$= -a\left(x-\frac{b}{2a}\right)^2 + \frac{b^2}{4a} + c$$

$$\Downarrow$$

$$\int \exp(\ldots) = \exp\left(\frac{b^2}{4a}+c\right)\cdot\frac{1}{\sqrt{a}}\cdot\sqrt{\pi}$$

•) $\displaystyle f_{X+Y}(t) = (f_X * f_Y) = \int_{\mathbb{R}} f_X(t-s)\,f_Y(s)\,ds$

$\downarrow$

{ Independence
Normal

→ Merry

However, $f_X \overset{\mathbb{R}}{*} f_Y$ is messy

$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ (different center)

$$\tilde{X} = \frac{X - \mu_X}{\sigma_X}, \quad \tilde{Y} = \frac{Y - \mu_Y}{\sigma_Y}$$

$\tilde{X}\sigma_X + \tilde{Y}\sigma_Y$ is normal

$\tilde{X}\sigma_X \sim \mathcal{N}(0, \sigma_X^2)$, $\tilde{Y}\sigma_Y \sim \mathcal{N}(0, \sigma_Y^2)$

$$f_{\mathcal{N}(0,\sigma_X^2)} * f_{\mathcal{N}(0,\sigma_Y^2)}(t) = \int_{\mathbb{R}} f_{\dots}(s) f_{\dots}(t-s)\, ds$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{\mathbb{R}} \exp\left(-\frac{s^2}{2\sigma_X^2}\right) \exp\left(-\frac{(t-s)^2}{2\sigma_Y^2}\right) ds$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{t^2}{2\sigma_Y^2}\right) \int_{\mathbb{R}} \exp\left(-s^2 \cdot \frac{\sigma_X^2 + \sigma_Y^2}{2\sigma_X^2\sigma_Y^2} + \frac{st}{\sigma_Y^2}\right) ds$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{t^2}{2\sigma_Y^2}\right) \exp\left(\frac{t^2}{\sigma_Y^4} \cdot \frac{1}{4} \cdot \frac{2\sigma_X^2\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}\right) \sqrt{\frac{2\pi}{\sigma_X^2 + \sigma_Y^2}}$$

$\sigma_X\sigma_Y$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2}} \exp\left(\frac{-t^2}{2\sigma_Y^2}\left(1 - \frac{\sigma_X^2}{\sigma_X^2 + \sigma_Y^2}\right)\right)$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_X^2 + \sigma_Y^2}} \exp\left(\frac{-t^2}{2(\sigma_X^2 + \sigma_Y^2)}\right)$$

- Moment generating function : $M_X(z) = \mathbb{E}\left(\exp(zX)\right)$

$$\exp(t) = 1 + t + \frac{t^2}{2!} + \dots = \sum_{n=0}^{\infty} \frac{t^n}{n!}$$

$$M_X(t) = \mathbb{E}\left(1 + tX + \frac{t^2 X^2}{2!} + \dots\right)$$

$\mathbb{E}(X)$, $\mathbb{E}(X^2)$, $\mathbb{E}(X^k)$: moments of X

$$M_X(0) = 1$$
$$M'_X(t) = \mathbb{E}(X) + t\,\mathbb{E}(X^2) + \dots$$

$$M'_X(0) = \mathbb{E}(X)$$

$$M''_X(0) = \mathbb{E}(X^2)$$
$$M^{(k)}_X(0) = \mathbb{E}(X^k)$$

Sometimes moments are hard but mgf are easy

- $X \sim \text{Bin}(n, p)$, $Z_i \overset{iid}{\sim} \text{Ber}(p)$

$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$
If X and Y are inde
(uncorrelation)

$X = Z_1 + \dots + Z_n$, $\mathbb{E}(X^5)$: hard

$$M_X(z) = \mathbb{E}\left(\exp(zX)\right) = \mathbb{E} \exp\left(zZ_1 + zZ_2 + \dots + zZ_n\right)$$

$$= \mathbb{E}\exp(zZ_1) \cdot \mathbb{E}\exp(zZ_2) \cdot \dots \cdot \mathbb{E}\exp(zZ_n)$$

$$= \left(e^z p + 1 - p\right)^n = \left(M_{\text{Ber}(p)}(z)\right)^n$$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

$$M_X(z) = \mathbb{E}\exp(zX) = \int_{\mathbb{R}} \exp(zt) f_X(t)\,dt$$

$$= \int_{\mathbb{R}} \exp(zt) \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\mathbb{R}} \exp\left(zt - \frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \int_{\mathbb{R}} \exp\left(zt - \frac{t^2}{2\sigma^2} + \frac{t\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right) dt \quad (*)$$

$$a = \frac{1}{2\sigma^2}, \quad b = z + \frac{\mu}{\sigma^2}, \quad c = \frac{-\mu^2}{2\sigma^2}$$

$$\Rightarrow (*) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(c + \frac{b^2}{4a}\right) \sqrt{\frac{\pi}{a}}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-\mu^2}{2\sigma^2} + \frac{\left(z + \frac{\mu}{\sigma^2}\right)^2}{2} \sigma^2\right) \cdot \sqrt{2\pi\sigma^2}$$

$$= \exp\left(\frac{-\mu^2}{2\sigma^2} + \frac{z^2\sigma^2}{2} + z\mu + \frac{\mu^2}{2\sigma^2}\right)$$

$$= \exp\left(\frac{z^2\sigma^2}{2} + z\mu\right) = e^{z\mu} \cdot e^{\frac{z^2\mu^2}{2}}$$

- <u>Caveat</u> : $\mathbb{E}(\exp(zX)) \overset{?}{=} \infty$

  M_X may not exist

  ↓ Fourier transform, spectral theorem of Laplacian

- <u>Characteristic func</u> : $\varphi_X(\tau) = \mathbb{E}(\exp(i\tau X))$ " $= M_X(i\tau)$ "

  ↓ $(\exp(i\tau X) = \cos(\tau X) + i\sin(\tau X))$

  $|\varphi_X(\tau)| \leq 1 \quad \forall \tau \in \mathbb{R}$ (exist always)

- CDF $\underset{1:1}{\overset{\text{Fourier}}{\longleftrightarrow}}$ char. func (black box)

Same CDF = identically distributed (The id in iid)

- $M_X^{(k)}(0) = \mathbb{E}(X^k)$

$\varphi_X^{(k)}(0) = i^k \, \mathbb{E}(X^k)$

- $X \sim \mathcal{N}(\mu, \sigma^2)$

$M_X(\tau) = \exp(\tau\mu + \frac{1}{2}\tau^2\sigma^2) \quad \forall x \in \mathbb{R}$

$\Rightarrow a_0 + a_1\tau + a_2\tau^2 + \dots = b_0 + b_1\tau + b_2\tau^2 + \dots \quad \forall \tau \in \mathbb{R}$

$\downarrow$ Guess

$a_j = b_j \quad \forall j$

$\downarrow$ Trick

$a_0 + a_1\tau + a_2\tau^2 + \dots = b_0 + b_1\tau + b_2\tau^2 + \dots \quad \forall \tau \in \mathbb{C}$

$\downarrow$

$M_X(\tau) = \exp(\tau\mu + \frac{1}{2}\tau^2\sigma^2)$ for all complex τ

$$\Rightarrow \boxed{\varphi(\tau) = M_X(i\tau) = \exp(i\tau\mu - \frac{1}{2}\tau^2\sigma^2) \quad \forall \tau \in \mathbb{R}}$$

- Application of char func

$$\begin{cases} X \sim \mathcal{N}(\mu_X, \sigma_X^2) \\ Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \\ X, Y \text{ inde} \end{cases}$$

Previously, we showed $X+Y \sim \mathcal{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$

<u>2nd method</u> : $\varphi_{X+Y}(\tau) = \mathbb{E}\exp(i\tau(X+Y))$
$= \mathbb{E}(\exp(i\tau X)\exp(i\tau Y))$
$= \mathbb{E}(\exp(i\tau X))\,\mathbb{E}(\exp(i\tau Y))$

(X, Y inde $\Rightarrow$ $\exp i\tau X$, $\exp i\tau Y$ inde)

$= \varphi_X(\tau).\varphi_Y(\tau) = \exp\left(i\tau\mu_X - \frac{1}{2}\tau^2\sigma_X^2\right)\exp\left(i\tau\mu_Y - \frac{1}{2}\tau^2\sigma_Y^2\right)$

$= \exp\left(i\tau(\mu_X+\mu_Y) - \frac{1}{2}\tau^2(\sigma_X^2+\sigma_Y^2)\right) =$

$\downarrow$ Fourier blackbox

$X+Y \sim \mathcal{N}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$

•) X: RV

Need to normalize to compare shapes $\rightarrow Z = \dfrac{X-\mu}{\sigma}$, $\mathbb{E}(Z)=0$,

$Var(Z) = 1$

<u>Central Limit Theorem</u> : let (X_i) be iid, $\mathbb{E}(X_i)=0$, $Var(X_i)=1$

$Var(X_1+\dots+X_n) = n.1 = n$ (by Pythogoras)

$\Rightarrow$ Normalize to $\dfrac{X_1+\dots+X_n}{\sqrt{n}}$ (square-averaged sample mean)

($\overline{X_n} = \dfrac{X_1+\dots+X_n}{n}$ is sample mean)

What shape does the distribution of $\sqrt{n}\cdot\overline{X}_n$ converge to?

$\rightarrow$ Use char func

$$\varphi_{\sqrt{n}\,\bar{x}}(\tau) = \varphi_{\frac{x_1}{\sqrt{n}}}(\tau)\,\varphi_{\frac{x_2}{\sqrt{n}}}(\tau)\dots\varphi_{\frac{x_n}{\sqrt{n}}}(\tau) = \left(\varphi_{\frac{x_1}{\sqrt{n}}}(\tau)\right)^n$$

Since $\varphi_{\frac{x_1}{\sqrt{n}}}(\tau) = \mathbb{E}\left(\exp\left(\frac{i\tau x_1}{\sqrt{n}}\right)\right) = \varphi_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)$

$$\Rightarrow \varphi_{\sqrt{n}\,\bar{x}_n}(\tau) = \left(\varphi_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)\right)^n$$

$$\lim_{n\to+\infty}\left(\varphi_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)\right)^n = \lim_{n\to+\infty} e^{\left(\ln \varphi_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)\right)\cdot n}$$

$$\left(\ln \varphi_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)\right)\cdot n \xrightarrow[+\infty]{\text{L'hopital}} \frac{1}{\varphi_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)}\cdot \varphi'_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)\cdot \tau\cdot\frac{1}{2}\cdot n^{-3/2}$$

$$\cdot n^2 = \frac{\tau}{2}\lim_{n\to\infty}\frac{\varphi'_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)}{n^{-1/2}} \overset{\text{L'Hopital}}{=}$$

$$\frac{\tau}{2}\cdot\lim\frac{\varphi''_{x_1}\left(\frac{\tau}{\sqrt{n}}\right)\cdot\left(-\frac{1}{2}\right)n^{-3/2}\,\tau}{\left(\frac{-1}{2}\right)n^{-3/2}} = \frac{\tau^2}{2}\cdot\varphi''_{x_1}(0) = \frac{-1}{2}\tau^2$$

$$\log A_n \xrightarrow{n\to\infty} -\frac{1}{2}\tau^2$$

Looks like pdf of $\mathcal{N}(0,1)$ (diff by $\frac{1}{\sqrt{2\pi}}$)

$$A_n \xrightarrow{n\to\infty} \exp\left(-\frac{1}{2}\tau^2\right) = \varphi_{\mathcal{N}(0,1)}(\tau)$$

$$\boxed{\varphi_{\sqrt{n}\,\bar{x}_n} \xrightarrow{n\to+\infty} \varphi_{\mathcal{N}(0,1)}}$$

CLT

Gaussian = eigenfunc of $\mathcal{N}(0,1)$

Fourier $\downarrow$

- $\forall t \in \mathbb{R}$

$$F_{\sqrt{n}\cdot \bar{X}_n}(t) \xrightarrow{n\to\infty} F_{\mathcal{N}(0,1)}(t) \quad (\not\equiv \text{CV of pdf})$$

→ Convergence in distribution (≠ Almost-sure CV - law of large numbers : $P(X_n \xrightarrow{n\to+\infty} X) = 1$)

- Modeling rare events :
  - X is very large, p is very small , but $np = \lambda$ (λ is constant)

→ Binomial

- $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$

$$\xrightarrow{n\to\infty} \ ?$$

$$\left(1-\frac{\lambda}{n}\right)^{n-k} = e^{\ln\left(1-\frac{\lambda}{n}\right)(n-k)}$$

$$\lim_{n\to+\infty} \frac{\ln\left(1-\frac{\lambda}{n}\right)}{(n-k)^{-1}} \overset{\text{l'hopital}}{=} \lim_{n\to\infty} \frac{\frac{1}{1-\frac{\lambda}{n}}\cdot\frac{\lambda}{n^2}}{-(n-k)^{-2}}$$

$$= \lim_{n\to\infty} \frac{n}{n-\lambda}\cdot\frac{\lambda}{n^2}\cdot(-1)\cdot(n-k)^2 = -\lambda$$

$$\to \left(1-\frac{\lambda}{n}\right)^{n-k} \xrightarrow{n\to\infty} e^{-\lambda}$$

$$\Rightarrow \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \xrightarrow{n\to\infty} \underbrace{\binom{n}{k}\left(\frac{\lambda}{n}\right)^k}\cdot e^{-\lambda}$$

$$\frac{n!}{(n-k)!\,k!} \cdot \frac{\lambda^k}{n^k} = \frac{\lambda^k}{k!} \cdot \frac{n!}{(n-k)!\,n^k}$$

$$= \frac{\lambda^k}{k!} \cdot \frac{n(n-1)\cdot \ldots \cdot (n-k+1)}{n^k} \xrightarrow{n\to\infty} \frac{\lambda^k}{k!}$$

$\Rightarrow P(X=k) \xrightarrow{n\to\infty} \frac{\lambda^k}{k!} e^{-\lambda}$ (Poisson approx / Stirling approx)

-) <u>Poisson distribution</u> : $Y \sim \text{Poisson}(\lambda)$ when $P(Y=k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ $\forall k \in \mathbb{N}_0$ (k could be 0)

-) <u>Poisson limit theorem</u> : $P_{Bin(n,\frac{\lambda}{n})}(k) \xrightarrow{n\to\infty} P_{Poisson(\lambda)}(k)$

CLT says $F_{\frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}}(t) \xrightarrow{n\to\infty} F_{\mathcal{N}(0,1)}(t)$

-) <u>Negative binomial distribution</u> : $NB(m,p)$ = # of fails to get to m-th success

$$P_{NB(m,p)}(k) = \binom{k+m-1}{m-1} \cdot (1-p)^k p^m$$

- From HW4, we have $T_3 - 3 \sim NB(3,p)$
  $T_n - n \sim NB(n,p)$

$X \sim NB(m,p)$

$$\mathbb{E}(X) = n\left(\frac{1}{p} - 1\right)$$

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

:) Alternative form of PLT (lecture)

:) <u>Gamma func</u> : $\Gamma(k) = \int_0^\infty t^{k-1} e^{-t} dt$ (where $k>0$) $=$

$$-\left(e^{-t} t^{k-1}\Big|_{t=0}^{\infty} - \int_0^\infty e^{-t}(k-1) t^{k-2} dt\right)$$

$$= (k-1)\int_0^\infty t^{k-2} e^{-t} dt = (k-1)\,\Gamma(k-1)$$

$$\Rightarrow \Gamma(n) = (n-1)!$$

- $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{1}{\sqrt{t}} e^{-t} dt$, let $u = \sqrt{t} \Rightarrow du = \frac{1}{2u} dt$

$$\Rightarrow \Gamma\left(\frac{1}{2}\right) = 2\int_0^\infty e^{-u^2} du = \int_{-\infty}^{+\infty} e^{-u^2} du = \sqrt{\pi}$$

- $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right)$

:) $B(k_1, k_2) = \int_0^1 t^{k_1-1}(1-t)^{k_2-1} dt \longrightarrow$ Beta function

- $$B(k_1, k_2) = \frac{\Gamma(k_1)\cdot\Gamma(k_2)}{\Gamma(k_1+k_2)} = \frac{\int_0^\infty t^{k_1-1}\cdot e^{-t} dt \cdot \int_0^\infty t^{k_2-1} e^{-t} dt}{\int_0^\infty t^{k_1+k_2-1} e^{-t} dt}$$

In convolution : $f * g(t) = \int_{\mathbb{R}} f(t-s)\, g(s)\, ds$

$\Big\downarrow$ Int in t

$$\int_{\mathbb{R}} f*g = \int_{\mathbb{R}}\int_{\mathbb{R}} f(t-s)\, g(s)\, ds\, dt$$

$$\overset{\text{Fubini}}{\underset{\mathbb{R}}{=}} \int_{\mathbb{R}} g(s)\Big(\int_{\mathbb{R}} f(t-s)\, dt\Big) ds = \Big(\int_{\mathbb{R}} g\Big)\Big(\int_{\mathbb{R}} f\Big)$$

•) $\int_{\mathbb{R}} f(x)\, 1_{[0,\infty)}(x)\, dx = \int_0^\infty f(x)\, dx$

$$\Rightarrow \Gamma(k_1) \cdot \Gamma(k_2) = \int_0^\infty t^{k_1-1} \cdot e^{-t}\, dt \cdot \int_0^\infty t^{k_2-1} e^{-t}\, dt$$

$$= \Big(\int_{\mathbb{R}} t^{k_1-1}\, 1_{[0,\infty)}\, e^{-t}\, dt\Big) \cdot \Big(\int_{\mathbb{R}} t^{k_2-1}\, 1_{[0,\infty)}\, e^{-t}\, dt\Big)$$

$$= f * g(t) = \int_{\mathbb{R}} (t-s)^{k_1-1}\, 1_{[0,\infty)}(s)\, e^{-t+s}\, s^{k_2-1}\, 1_{[0,\infty)}(s)\, e^{-s}\, ds$$

$1_{[0,\infty)}(s) \cdot 1_{[0,\infty)}(s) \neq 0$ when $0 \leq s \leq t$

$$\Rightarrow f*g(t) = e^{-t} \int_0^t (t-s)^{k_1-1}\, s^{k_2-1}\, ds \quad (*)$$

Let $u = \frac{s}{t} \Rightarrow du = \frac{ds}{t}$

$$(*) = e^{-t} \cdot \int_0^1 (t-tu)^{k_1-1}\, (ut)^{k_2-1}\, t\, du$$

$$= e^{-t} \cdot t^{k_1+k_2-1} \int_0^1 (1-u)^{k_1-1} \cdot u^{k_2-1} \, du$$

$$= e^{-t} \cdot t^{k_1+k_2-1} \cdot B(k_1, k_2)$$

$$\Rightarrow \int_{\mathbb{R}} f * g = B(k_1, k_2) \int_0^\infty e^{-t} \cdot t^{k_1+k_2-1} \, dt$$

$$\Rightarrow \Gamma(k_1)\, \Gamma(k_2) = B(k_1, k_2) \cdot \Gamma(k_1+k_2)$$

• <u>Gamma distribution</u> : $X \sim \text{Gamma}(k, \lambda)$ $(k>0, \lambda>0)$

when $f_X(t) = \dfrac{\lambda^k}{\Gamma(k)} t^{k-1} \exp(-\lambda t) \, 1_{[0,\infty)}(t)$

$$\int_0^\infty \frac{\lambda^k}{\Gamma(k)} t^{k-1} \exp(-\lambda t) \, dt \quad (*)$$

Let $u = \lambda t \Rightarrow (*) = \displaystyle\int_0^\infty \frac{\lambda^k}{\Gamma(k)} \left(\frac{u}{\lambda}\right)^{k-1} \exp(-u) \frac{du}{\lambda}$

$$= \int \frac{u^{k-1}}{\Gamma(k)} \exp(-u) \, du = \frac{\Gamma(k)}{\Gamma(k)} = 1$$

• MGF $\rightarrow$ CF :

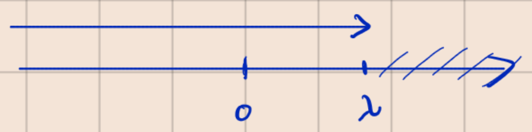
$$M_X(\tau) \overset{\text{Lotus}}{=} \int_{\mathbb{R}} \exp(\tau \cdot t) \, f_X(t) \, dt$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^\infty e^{\tau t} \cdot t^{k-1} e^{-\lambda t} \, dt$$

$$= \frac{\lambda^k}{\Gamma(k)} \int_0^\infty e^{(\tau-\lambda)t} \cdot t^{k-1} \, dt$$

To make sure the integral is finite, we need $\tau < \lambda$

So $M_X(\tau)$ is only defined for $\tau \in (-\infty, \lambda)$



Let $u = (\lambda - \tau) t$

$$\Rightarrow M_X(\tau) = \frac{\lambda^k}{\Gamma(k)} \int_0^\infty \frac{u^{k-1}}{(\lambda-\tau)^{k-1}} \exp(-u) \cdot \frac{1}{\lambda-\tau} \, du$$

$$= \frac{\lambda^k}{(\lambda-\tau)^k} \cdot \frac{\Gamma(k)}{\Gamma(k)} = \left(\frac{\lambda}{\lambda-\tau}\right)^k$$ for $\tau \in (-\lambda, \lambda)$

Taylor series

$$\frac{1}{1-\frac{\tau}{\lambda}}$$

$\left|\frac{\tau}{\lambda}\right| \le 1$

Taylor series

$$a_0 + a_1\tau + a_2\tau^2 + \dots = b_0 + b_1\tau + b_1\tau^2 + \dots$$

$\Rightarrow a_i = b_i \quad \forall i$

$\Rightarrow M_X(\tau) = \left(\frac{\lambda}{\lambda-\tau}\right)^k$ $\forall$ complex τ s.t. $\mathrm{Re}(\tau) \in (-\lambda, \lambda)$

<u>**Worry**</u>: is $M_X(\tau)$ finite (complex number)

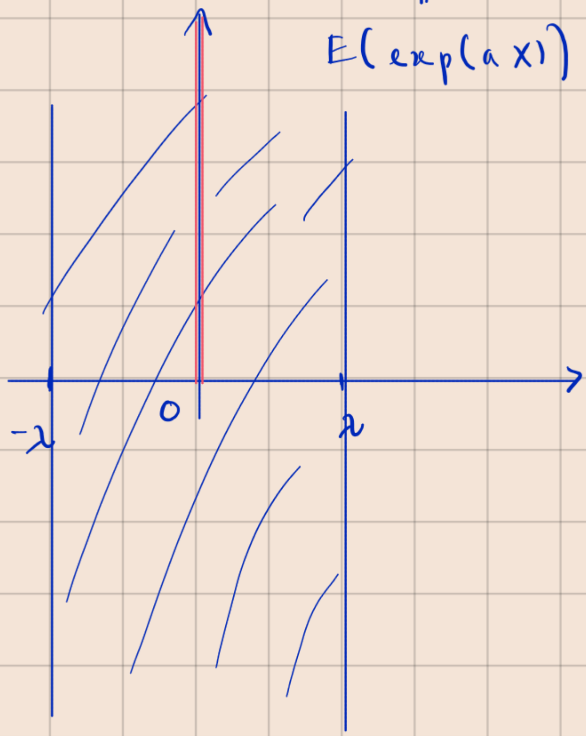
$\tau = a + bi$ where $a, b \in \mathbb{R}$, $a \in (-\lambda, \lambda)$

$$M_X(\tau) = \mathbb{E}(\exp(aX)\exp(ibX))$$

$\downarrow$ Triangle inequality

$$|M_X(\tau)| \leq \mathbb{E}(|\exp(aX)|) = M_X(a) < \infty \quad \text{as } a \in (-\lambda, \lambda)$$

$$\mathbb{E}(|\exp(aX)|) = E(\exp(aX))$$



So $M_X(ib) = \varphi_X(b)$

$$= \left(\frac{\lambda}{\lambda - ib}\right)^k$$

- Data set :
  - Sample mean : $\bar{x} = \frac{\sum_i x_i}{n}$
  - Unbiased sample varience :

(Maximum likelihood estimator)

- MLE for Discrete :

$\vec{\theta}$: parameter vectors (unknown, to be estimated)

$\vec{X}$: discrete RV

Previously : given $\vec{\theta} = \vec{b}$, we calculated

p of $X = \vec{a}$ given $\vec{b}$

$$\vec{a} \mapsto p(\vec{X} = \vec{a} ; \vec{\theta} = \vec{b}) \text{ (denoted as } p_{\vec{X}, \vec{b}}(\vec{a}) \text{)}$$

<u>Now</u> : given observed data set $\vec{X} = \vec{a}$ and we define the <u>likelihood function</u>

$$\vec{b} \longmapsto \mathcal{L}(\vec{X} = \vec{a} \,;\, \vec{\theta} = \vec{b})$$

<u>Example</u> : $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$, $X_i \sim Ber(\theta)$

$$\mathcal{L}\left(\vec{X} = (1,0,1) \,;\, \theta = p\right) = p^2(1-p)$$

→ the likelihood func from the observed data $\vec{X} = \vec{a}$

<u>Gauss MLE</u> : Find p to maximize $\mathcal{L}$ $(p \in [0,1])$

<u>Optimization</u> :

- Solve $f'(x) = 0 \Rightarrow$ critical points
- 2nd derivative test at crit points $f'(a) = 0$, $f''(a) < 0$

$\Rightarrow$ a is a local max point

check all critical points and "boundary"

↓

"global" max point

Review 2^{nd} derivative test on $\mathbb{R}^n$

Take log first : $2\ln p + \ln(1-p) = G(p)$

$$G'(p) = \frac{2}{p} - \frac{1}{1-p} = 0 \Rightarrow p = \frac{2}{3}$$

$$G''(p) = \frac{-2}{p^2} - \frac{1}{(1-p)^2} < 0 \ \forall p \Rightarrow p = \frac{2}{3} \text{ is local max point}$$

- $G(0) = -\infty$, $G(1) = -\infty$

$\Rightarrow$ $p = \frac{2}{3}$ is the local max point

- MLE of $\mathcal{C}^{\circ}$ RV : Given observed data set $\vec{X} = \vec{a}$, find $\vec{b}$ to maximize $f_{\vec{X}}(\vec{t}, \vec{b})$

Ex1 : $\vec{X} = (X_1, X_2)$, $X_i \sim$ Uniform over $(0, \theta)$

Observed data $\vec{X} = (t_1, t_2)$, $\vec{\theta} = \vec{b}$

$$f_X((t_1, t_2), \vec{b}) = f_{X_1, b}(t_1) \cdot f_{X_2, b}(t_2)$$

$$= \frac{1}{b} \cdot 1_{[0,b]}(t_1) \cdot \frac{1}{b} \cdot 1_{[0,b]}(t_2) = \underbrace{\frac{1}{b^2}}_{\text{decreasing}} \cdot \underbrace{1_{[0,b]}(t_1)\, 1_{[0,b]}(t_2)}_{b \geq t_1 \ \& \ b \geq t_2}$$

$\Rightarrow$ $b = \max(t_1, t_2)$

What if $\max(t_1, t_2) \leq 0$

- $\max(t_1, t_2) < 0$ cannot happen as observed data
- $\max(t_1, t_2) = 0$ is an event with 0 probability

Ex2 : $\vec{X} = (X_1, X_2, X_3)$, $X_i \sim \mathcal{N}(\mu, \sigma^2)$

unknown

$$g(\mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^3 \exp\left(-\sum_i \frac{(t_i - \mu)^2}{2\sigma^2}\right)$$

$3 \ln \frac{1}{\sqrt{2\pi}} - 3\ln\sigma$

Let $G(\mu, \sigma) = \ln g(\mu, \sigma) = \left(-\sum_i \frac{(t_i - \mu)^2}{2\sigma^2}\right) + 3 \ln \frac{1}{\sqrt{2\pi}\,\sigma}$

$$\vec{\nabla} G(\mu, \sigma) = \vec{0} \Rightarrow \begin{cases} \partial_\mu G = \sum_{i=1}^{3} \frac{t_i - \mu}{\sigma^2} = 0 \\ \partial_\sigma G = \frac{-3}{\sigma} + \sum_i \frac{(t_i - \mu)^2}{\sigma^3} = 0 \end{cases}$$

$$\begin{cases} \mu = \frac{t_1 + t_2 + t_3}{3} \leftarrow \text{Sample mean} \\ \sigma^2 = \frac{1}{3} \sum_i (t_i - \mu)^2 \leftarrow \text{biased sample var} \end{cases}$$

∘ Hessian Matrix : $\begin{pmatrix} \partial_{\mu\mu} & \partial_{\sigma\mu} \\ \partial_{\mu\sigma} & \partial_{\sigma\sigma} \end{pmatrix}$

$$\partial_{\mu\mu} G = \frac{-3}{\sigma^2}$$

$$\partial_{\mu\sigma} = \partial_{\sigma\mu} = \frac{-2}{\sigma^3} \sum_i (t_i - \mu)$$

$$\partial_{\sigma\sigma} = \frac{3}{\sigma^2} - \frac{3}{\sigma^4} \sum_i (t_i - \mu)^2$$

$$\begin{cases} \mu_0 = \frac{t_1 + t_2 + t_3}{3} \leftarrow \text{Sample mean} \\ \sigma_0^2 = \frac{1}{3} \sum_i (t_i - \mu)^2 \leftarrow \text{biased sample var} \end{cases}$$

$$H\Big|_{(\mu_0, \sigma_0)} = \begin{pmatrix} \frac{-3}{\sigma_0^2} & 0 \\ 0 & \frac{-6}{\sigma_0^2} \end{pmatrix}$$ is negative definite

$\begin{cases} \text{symmetric} \\ \text{negative eigval} \end{cases}$

$\Updownarrow$ Spectral Theo

$$\left(A\vec{x} \cdot \vec{x} = \langle A\vec{x}, \vec{x} \rangle < 0 \right)$$

$\Rightarrow$ max at (μ_0, σ_0)

Domain : $\begin{cases} \mu \in \mathbb{R} \\ \sigma > 0 \end{cases}$

$$-3 \ln \sigma - \frac{3}{2} \ln (2\pi) - \sum_i \frac{(t_i - \mu)^2}{2\sigma^2}$$

• $\lim \left(- \ln \sigma - \frac{1}{\sigma^2} \right) = -\infty$

Prove : $\lim\limits_{\sigma \to 0^+} \left(- \ln \sigma - \frac{1}{\sigma^2} \right) \xrightarrow{\exp} \frac{1}{\sigma \exp\left(\frac{1}{\sigma^2}\right)} \xrightarrow{x = \frac{1}{\sigma}} \frac{x}{e^{x^2}}$

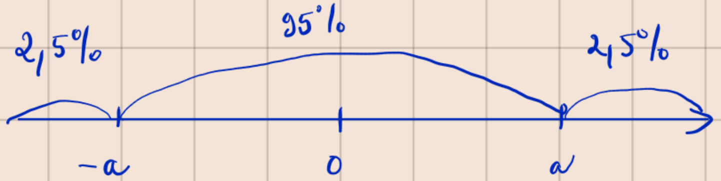
$$\overset{\text{L'Hopital}}{=} \lim_{x \to +\infty} \frac{1}{e^{x^2} 2x} = 0$$

$$\overset{\log}{\Rightarrow} -\ln \sigma - \frac{1}{\sigma^2} \xrightarrow{\sigma \to 0} -\infty$$

$G = -\infty$ on boundary $\rightarrow$ (μ_0, σ_0) is global max

•) <u>z-table</u> : gives the value for $P(Z < z)$

Find a s.t $P(-a < z < a) = 95\%$



$\Rightarrow P(Z < a) = 97,5\%$

$\Downarrow$

$a = 1,96$

When $a = 1,96$; $P(z < a) = 97,5\%$

• <u>z-score</u> : let $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$. $\forall a \in \mathbb{R}$, the z-score (or standard score) of a wrt X : $\frac{a-\mu}{\sigma}$

•) <u>95% confidence interval for normal mean with known var :</u>

• $X \sim \mathcal{N}(\mu, \sigma^2)$, know σ^2, don't know μ (just that it is some constant). Need to estimate μ by an interval

$\vec{X} = (X_1, \dots, X_n)$ (random sample vector before observing)
(theoretical model)

where $X_i \overset{iid}{\sim} X \sim \mathcal{N}(\mu, \sigma^2)$

• After experiment, we got $\vec{X}_{obs} = ((X_1)_{obs}, \dots, (X_n)_{obs})$

= constant (reality / data)

We want to find $c_1(\vec{x}), c_2(\vec{x})$ s.t

$$\boxed{\underbrace{\mathbb{P}}_{\text{for } X}\left(\underbrace{c_1(\vec{X})}_{\text{Random}} < \underbrace{\mu}_{\text{unknown const}} < \underbrace{c_2(\vec{X})}_{\text{Random}}\right) = 95\%}$$
theoretical

• Look at the (theoretical) sample mean $\bar{X} = \frac{X_1 + \dots + X_n}{n}$

$X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

$\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ $\left(\begin{array}{l} \mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\mu}{n} \\ Var(\bar{X}) = Var\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{\sigma^2}{n} \end{array} \right)$

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

$$95\% = P\left(-1,96 < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < 1,96\right) = P\left(\bar{X} - 1,96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1,96\frac{\sigma}{\sqrt{n}}\right)$$

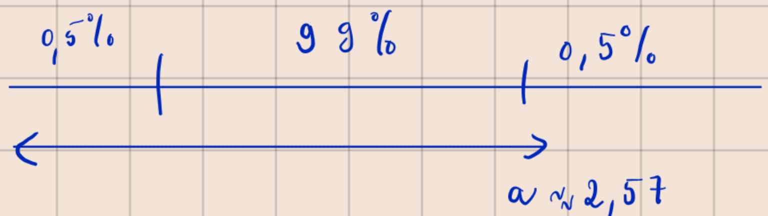
95% CI for μ is $\left(\bar{X} - 1,96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96\frac{\sigma}{\sqrt{n}}\right)$

(theoretical)

Every thing is constant except $\bar{X}$ ($\bar{X}$ is random)

•) The choice for $(c_1(\vec{X}), c_2(\vec{X}))$ is not unique

• Only unique when restrict $\frac{c_1 + c_2}{2} = \bar{x}$

•) 

Find 95% interval for this RV

•) Previously : know σ^2, estimate μ by the expression $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$

Now : Do not know σ^2, estimate μ

Replace σ^2 by s^2

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$ (Theoretical) sample variance

Use this expression $\dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

Find pdf s^2 then $\dfrac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

<u>pdf of s^2</u> : $s^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$x_i \sim \mathcal{N}(\mu, \sigma^2)$, $\bar{x} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

z- normalize : $z_i = \dfrac{x_i - \mu}{\sigma} \sim \mathcal{N}(0,1) \Rightarrow \boxed{x_i = \sigma z_i + \mu}$

$$\bar{z} = \frac{1}{n}(z_1 + \dots + z_n) \sim \mathcal{N}(0, \tfrac{1}{n}) \quad \Rightarrow \boxed{\bar{x} = \sigma \bar{z} + \mu}$$

$$\Rightarrow s^2 = \frac{\sigma^2}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})^2$$

$$= \frac{\sigma^2}{n-1} \sum_{i=1}^{n} \left(z_i^2 - 2 z_i \bar{z} + \bar{z}^2 \right)$$

$$= \frac{\sigma^2}{n-1} \left(\left(\sum_{i=1}^{n} z_i^2 \right) - 2\bar{z}\, n\bar{z} + n\,\bar{z}^2 \right)$$

$$= \frac{\sigma^2}{n-1} \left(\left(\sum_i z_i^2 - n\,\bar{z}^2 \right) \right) \qquad (\sqrt{n}\,\bar{z})^2 \sim \chi^2(1)$$

$$\chi^2(1) + \dots + \chi^2(1) = \chi^2(n)$$

Inde sum formula for Gamma

$$\sum_i z_i^2 = n\bar{z}^2 + \frac{n-1}{\sigma^2} s^2$$

Gamma

Inde by HW7

$$\varphi_{X+Y} = \varphi_X \varphi_Y \quad \text{when } X, Y \text{ inde}$$

$$\Rightarrow \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right) = \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) + A$$

$\downarrow$ CF

$$\varphi_{\text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)}(t) = \varphi_{\text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)}(t) \cdot \varphi_A(t)$$

$$\left(\frac{\frac{1}{2}}{\frac{1}{2} - it}\right)^{\frac{n}{2}} = \left(\frac{1/2}{1/2 - it}\right)^{1/2} \varphi_A(t)$$

$$\Rightarrow \varphi_A(t) = \left(\frac{1/2}{1/2 - it}\right)^{\frac{n-1}{2}} \Rightarrow A \sim \chi^2(n-1)$$

$$s^2 = \frac{\sigma^2}{n-1} A \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right)$$

<u>HW8</u> : $\dfrac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1} \rightarrow$ Student's t-distribution

deg of freedom

→ t - table

$$P(-a < t_9 < a) = 95\%$$

$$P(-a < t_{10} < a) = 99\%$$

·) <u>Student t-dist</u>: $\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim \dfrac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n}\chi^2(n)}}$ (random expression) $\left(\begin{array}{l} X \sim \mathcal{N}(\mu, \sigma^2) \\ \bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}) \end{array} \right)$

→ CI for μ when σ^2 is unknown

· Standard normal dist: $\dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ → CI for μ when σ^2 is known

Main idea: to get CI for unknown quantity a, find some random expression that involves a (and no other unknown)

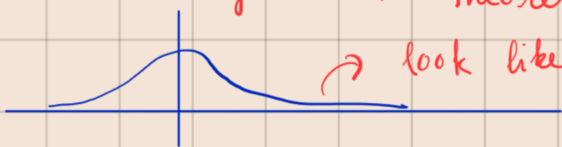
<u>Example</u>: $n = 10$, σ^2 is unknown $\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{10}}} \sim t_9$

$$\to 95\% = \mathbb{P}\left(-2{,}262 < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{10}}} < 2{,}262\right)$$

$$= \mathbb{P}\left(\frac{S}{\sqrt{10}}(-2{,}262) < \bar{X}-\mu < \frac{S}{\sqrt{10}} \cdot 2{,}262\right)$$

$$= \mathbb{P}\left(\bar{X} - \frac{S}{\sqrt{10}}\, 2{,}262 < \mu < \bar{X} + \frac{S}{\sqrt{10}}\, 2{,}262\right)$$

Theoretical — will be constant after exp — Theoretical

· Graph of t-dist:  look like, but not normal

③ $X \sim \mathcal{N}(\mu, \sigma^2)$, give CI for σ^2

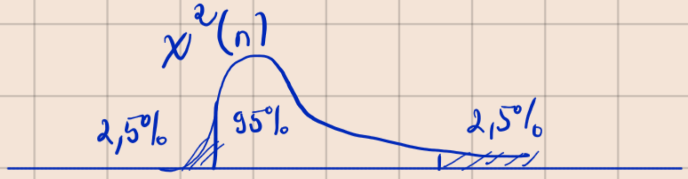
- Need to give random expression that involve σ^2 (and no other unk)

<u>Intuition</u> : S^2 is related to σ^2

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \sim \frac{\sigma^2}{n-1} \chi^2(n-1)$$

$$\Rightarrow \frac{s^2(n-1)}{\sigma^2} \sim \chi^2(n-1) \qquad \left(\mathcal{N}(0,1)^2 \sim \chi^2(1) \right)$$

$$95\% = \mathbb{P}(a_0 < \chi^2(n-1) < a_n)$$



Not symmetric over 0

But we can still define "symmetric" 2 tailed

<u>Example</u> : When $n = 10$, $\chi^2(9)$. Find a_0, a_1

$$95\% = \mathbb{P}\left(2{,}7 < \frac{s^2 \cdot 9}{\sigma^2} < 19{,}023\right)$$

$$= \mathbb{P}\left(\frac{s^2 \cdot 9}{19{,}023} < \sigma^2 < \frac{s^2 \cdot 9}{2{,}7}\right)$$

$\rightarrow$ CI for σ^2 ✓

$\cdot)$ $X_1, \ldots, X_{n_1} \overset{iid}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$

$Y_1, \ldots, Y_{n_2} \overset{iid}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$

Give CI for $\mu_1 - \mu_2$

① σ_1, σ_2 are known

② σ_1, σ_2 are unknown $\begin{cases} \sigma_1 = \sigma_2 \\ \sigma_1 \neq \sigma_2 \end{cases}$

①

Need random expression that has $\mu_1 - \mu_2$

Intuition : $\bar{X} - \bar{Y}$ is related to $\mu_1 - \mu_2$

$$\bar{X} \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\bar{Y} \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

only unknown

$$\Rightarrow \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1)$$

$95\% = P(-1{,}96 < \ldots < 1{,}96)$

② σ_1, σ_2 are unknown (but $\sigma_1 = \sigma_2$) :

Intuition : need to make χ^2 appear , and use the formula for χ^2

$$\frac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n}\chi^2(n)}} \sim t_n$$

Replace σ_1^2 by s_1^2
σ_2^2 by s_2^2

$$s_1^2 \sim \frac{\sigma_1^2}{n_1 - 1} \chi^2(n_1 - 1)$$

$$s_2^2 \sim \frac{\sigma_2^2}{n_2 - 1} \chi^2(n_2 - 1)$$

$$\chi^2(n_1 - 1) + \chi^2(n_2 - 1) \sim \chi^2(n_1 + n_2 - 2)$$

$$\frac{s_1^2(n_1 - 1)}{\sigma^2} + \frac{s_2^2(n_2 - 1)}{\sigma^2} \sim \chi^2(n_1 + n_2 - 2)$$

Want to make $\sqrt{\frac{1}{n_1 + n_2 - 2} \cdot \chi^2(n_1 + n_2 - 2)}$

We have $\frac{1}{n_1 + n_2 - 2} \chi^2(n_1 + n_2 - 2) \sim \frac{1}{n_1 + n_2 - 2} \cdot \frac{1}{\sigma^2} \left(s_1^2(n_1 - 1) + s_2^2(n_2 - 1)\right)$

Define $s_p^2 = \frac{1}{n_1 + n_2 - 2}\left(s_1^2(n_1 - 1) + s_2^2(n_2 - 1)\right)$

$$\Rightarrow \frac{1}{n_1 + n_2 - 2} \chi^2(n_1 + n_2 - 2) \sim \frac{1}{\sigma^2} s_p^2$$

Look at

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{s_p^2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \frac{\mathcal{N}\left(0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}{\sqrt{s_p^2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$\sigma^2 = \sigma_1^2 = \sigma_2^2$

$$\sim \frac{\mathcal{N}\left(0,\ \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)}{\sqrt{s_p^2}\ \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim \frac{\mathcal{N}(0,\ \sigma^2)}{\sqrt{\frac{\sigma^2}{n_1+n_2-2}\ \chi^2(n_1+n_2-2)}}$$

$$\sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n_1+n_2-2}\ \chi^2(n_1+n_2-2)}} \sim t_{n_1+n_2-2}$$

③ σ_1, σ_2 unknown, $\sigma_1 \overset{???}{=} \sigma_2$

$\Rightarrow$ Give up

·) <u>Prediction interval</u> : " From n data points, predict the next data point "

Let $X_1, \dots, X_n, X_{n+1} \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

where $\begin{cases} \mu \text{ is known} \\ \sigma^2 \text{ is } \begin{cases} \text{known} & ① \\ \text{unknown} & ② \end{cases} \end{cases}$

Let $\vec{X} = (X_1, \dots, X_n)$ be the current sample (which collapse into the n data points after observation)

X_{n+1} is the future we need to predict (unlike CI where we estimated constansts)

We want to find " symmetrical " theoretical interval $(c_1(\vec{X}), c_2(\vec{X}))$ s.t

$$P(\underbrace{c_1(\vec{X})}_{X_1, \dots, X_n \text{ are random objects}} < X_{n+1} < c_2(\vec{X})) = 95\%$$

Find a random expression that has X_{n+1} and no unknown (and other RVs that will collapse into const)

① μ unknown, σ is unknown

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$X_{n+1} \sim \mathcal{N}(\mu, \sigma^2)$$

Then $\bar{X} - X_{n+1} \sim \mathcal{N}\left(0, \sigma^2 + \frac{\sigma^2}{n}\right)$

$$\frac{\bar{X} - X_{n+1}}{\sqrt{\sigma^2 + \frac{\sigma^2}{n}}} \sim \mathcal{N}(0, 1)$$

$$\Rightarrow \frac{\bar{X} - X_{n+1}}{\sigma\sqrt{1 + \frac{1}{n}}} \sim \mathcal{N}(0,1)$$

$$P\left(-1{,}96 < \frac{\bar{X} - X_{n+1}}{\sigma\sqrt{1+\frac{1}{n}}} < 1{,}96\right) = 95\,\%$$

$$\Rightarrow P\left(\bar{X} - \sigma\sqrt{1+\tfrac{1}{n}} \cdot 1{,}96 < X_{n+1} < \bar{X} + \sigma\sqrt{1+\tfrac{1}{n}} \cdot 1{,}96\right)$$

② Both μ, σ are unknown :

Formula $\dfrac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n}\chi^2(n)}} \overset{\text{equal in dist}}{\sim} t_n$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$ current theoretical sample variance

$$\sim \frac{\sigma^2}{n-1}\chi^2(n-1) \rightarrow \frac{s^2}{\sigma^2} \sim \frac{1}{n-1}\chi^2(n-1)$$

$$\frac{x_{n+1} - \overline{X}}{\sqrt{\sigma^2(1+\frac{1}{n})} \cdot \sqrt{\frac{S^2}{\sigma^2}}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n-1} \chi^2(n-1)}} \sim t_{n-1}$$

$$= \frac{x_{n+1} - \overline{X}}{\sqrt{(1+\frac{1}{n}) s^2}}$$

$\Rightarrow \quad 95\% = P(\dots)$

bounds : $\bar{X} \pm a\sqrt{S^2(1+\frac{1}{n})}$

- rstudio and R - programming ✓

- <u>Fisher's null hypothesis testing</u> ✓ :

• <u>Theorem</u> : There is no biggest natural number

<u>Proof</u> : Assume the opposite (there is a biggest n)

Note a contradiction ($n+1 > n$)
→ Proof by contradiction

• <u>In stats</u> : Assume the opposite ← H_0 (null hypothesis)

→ show a "statistical contradiction" (5% chance of happening)
→ reject H_0

- "Not a contradiction" (95% chance of happening)
  → Conclude nothing (fail to reject H_0)

- <u>Ex</u> : Scientist believes $\mu \neq 10$

$H_0 : \mu = 10$ → Try to reject H_0

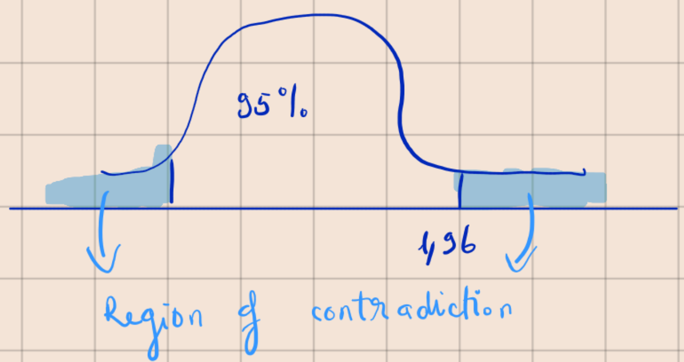
Assume $\sigma^2 = 4$

Take a random sample $\vec{X} = (X_1, \dots, X_n)$, $n = 100$

Then $\boxed{\dfrac{\bar{X} - \mu}{\sqrt{\dfrac{\sigma^2}{n}}} \sim \mathcal{N}(0,1)}$ No unknown

Random expression $\xrightarrow{\text{collapse}}$ test number

Contradiction <=> Test num is in $\mathbb{R} \setminus [-1{,}96, 1{,}96]$



<u>Ex 1</u>: $\bar{X}_{obs} = 13$

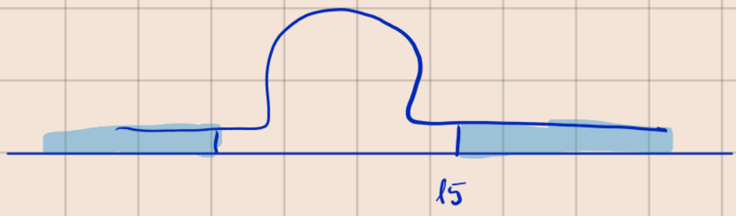
$$\frac{\bar{X}_{obs} - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{13 - 10}{\sqrt{\frac{4}{100}}} = 15 \in C \rightarrow \text{Reject } H_0$$

Too large

•) <u>p-value</u> : $\mathbb{P}$ (given H_0, $\vec{X}$ is more extreme then $\vec{X}_{obs}$)

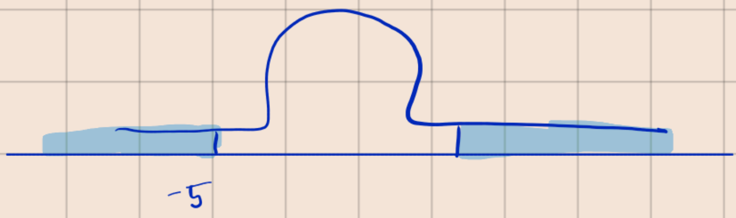
<u>Ex 1</u> : Test num $= 15$

Random dist $= \mathcal{N}(0, 1)$

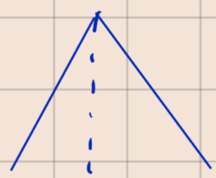


$$p = 2\,(1 - \mathrm{CDF}(15))$$

<u>Ex 2</u> : Test num $= -5$



$$p = 2\ \mathrm{CDF}(-5)$$

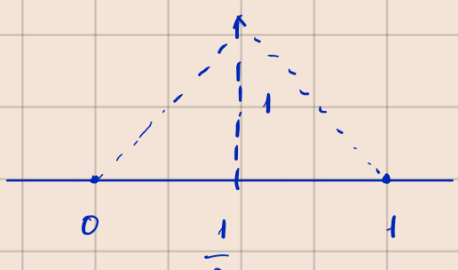


<u>Ex 3</u> : Test num $= 100$

random dist $= \chi^2(3)$

$$|x| + \left|\tfrac{1}{2} - x\right|$$

$\cdot)$ 2 sided p-value $= 2 \min\left(\mathrm{cdf}(\text{test}),\ 1 - \mathrm{cdf}(\text{test})\right)$



•) When σ^2 is unknown :

Use $\dfrac{\bar{x} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$

<u>Ex</u> : change the scientist example to

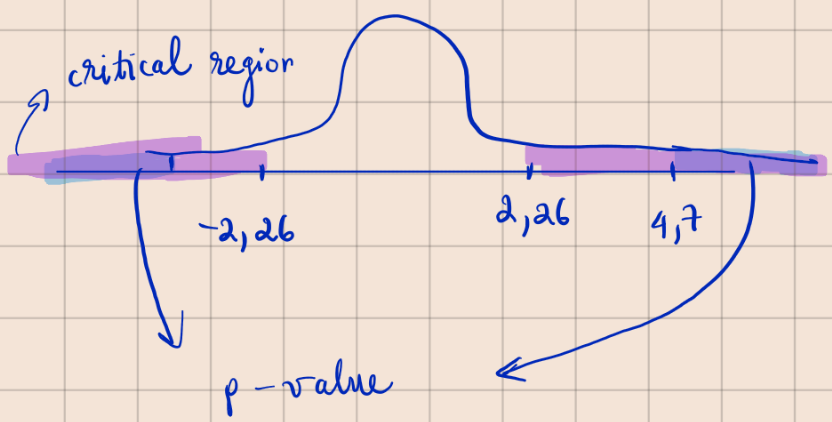
$$\begin{cases} \sigma^2 \text{ is unknown} \\ \bar{x}_{obs} = 13 \ , \ s^2_{obs} = 4 \\ n = 10 \end{cases}$$

$$\frac{13 - 10}{\sqrt{\frac{4}{10}}} = \frac{3\sqrt{10}}{2} \approx 4,7 \text{ (test num)}$$

Reject region : $(-\infty, 2,26) \cup (2,26, \infty)$

→ reject

p-value $= 2(1 - \text{cdf}(4,7)) \approx 0,001$



• What if $H_0 : \mu \neq 10$ (composite), then it is complicated (don't care)

•) Recall : p-value < 5% ← Test num is in crit region
⟺ reject H_0

<u>Problem</u>: We don't know how large the effect size

•) CI : estimate the unknown
HT : assume unknown = ... ← simple H_0 , try to show contradiction

•) "Two normal samples"

CI : estimate $\mu_1 - \mu_2$
HT : $H_0 : \mu_1 - \mu_2 = 0$ (goal is proving $\mu_1 \neq \mu_2$)

•) <u>Hypothesis testing for variance</u> :

$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

μ, σ^2 unknown

• Random expression : $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ → will collapse to Test num

Ex : $S^2_{obs} = 5$

$\dfrac{5 \cdot 9}{3} = 15$ = Test num → $\chi^2(9)$

95% interval (2,7 ; 19,02) → fail to reject H_0

• p-value ✓ $= 2 \cdot \min(\text{pchisq}(15, 9), 1 - \text{pchisq}(15, 9))$
$\approx 0{,}182 > 5\%$

• $CDF^{-1} \rightarrow$ magic nums for the interval
$CDF(\text{test}) \rightarrow$ p-value

upper and lower are inde RVs

•) <u>F-distribution</u> : $F(d_1, d_2) := \dfrac{\text{“}\chi^2(d_1)\text{”}}{\chi^2(d_2)} \cdot \dfrac{d_2}{d_1}$

• $F^{-1}_{F(10, 20)}(97{,}5\%) = \text{qf}(0{,}975, 10, 20)$ for R-prog

•) <u>Type I Error</u> : Given H_0

In HT, P(wrongly rejecting null hypothesis) $= 5\%$
$\parallel$
P(false positive)
$\parallel$
P(Type I Error) $= \alpha =$ significance level

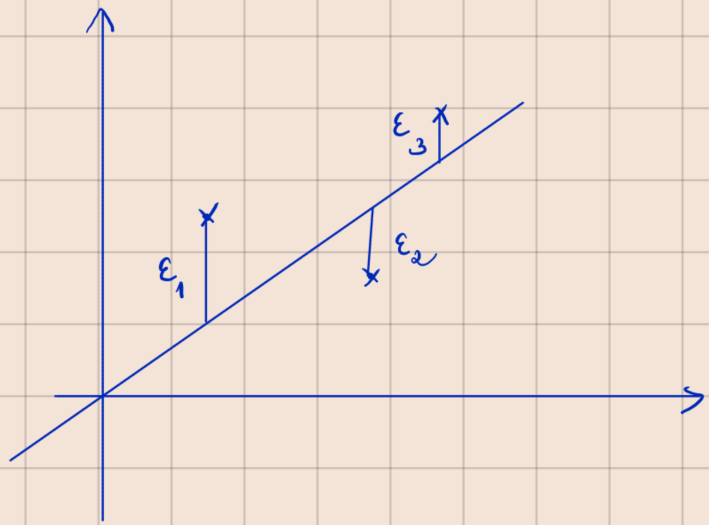
* P(false negative) $= P$(wrongly retain null hypo)
$= P$(Type II Error) $= \beta \rightarrow P$(H_0 can still be rejected)

$\rightarrow$ power of test $= 1 - \beta$

<u>Note</u> : “ 5% false positive rate does not mean positives are 5% false” (because of Bayes's Theorem)

- Linear Regression : Have $(x_1, Y_1), \dots, (x_n, Y_n)$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad , \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$



General relation (linear ver)

$$Y_i = \underbrace{\alpha + \beta x_i}_{\text{line}} + \underbrace{\varepsilon_i}_{\text{error}}$$

α, β : unknown constant

ε_i : error terms

Previously in LA : ε_i are not random

Now in Stats : ε_i are random (before observation)

- LA problem : $\hat{\alpha}, \hat{\beta}$ are the values of α, β to minimize

$$SSR = \sum_i \varepsilon_i^2 = \sum_i (Y_i - \alpha - \beta x_i)^2$$

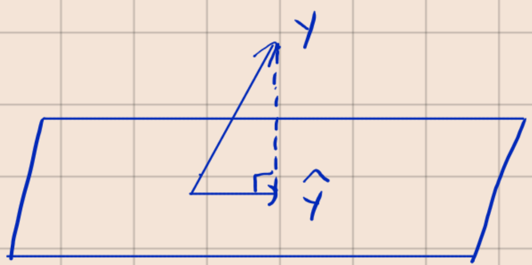
Define $\boxed{\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i}$ → sample LA model

How to find $\hat{\alpha}$ and $\hat{\beta}$

$$Y = \underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}}_{\text{fixed}} \Bigg\| \begin{pmatrix} \alpha + \beta x_1 \\ \vdots \\ \alpha + \beta x_n \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{\text{fixed}} \overbrace{\begin{pmatrix} \alpha \\ \beta \end{pmatrix}}^{\text{unknown}}$$

Find α, β to minimize

$$\left\| A \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - Y \right\|^2$$



•) <u>Projection problem</u> :

Given $\begin{cases} A & n \times k \text{ matrix} \\ Y & \in \mathbb{R}^n \end{cases}$ (fixed)

Find $v \in \mathbb{R}^k$ to minimize $\| Av - Y \|$

<u>Solution</u> : let v be the desired solution

Then $Av - Y \perp \text{Range}(A)$

$\Rightarrow Av - Y \in \text{Range}(A)^{\perp} = \text{Null}(A^T)$

Important id comes from $u \cdot v = v^T u$

$(A\vec{v})\vec{u} = \vec{v} A^T \vec{u}$

so if $\vec{u} \perp A\vec{v}$

then $A^T \vec{u} = 0$

$\Rightarrow u \perp A\vec{v} \ \forall \vec{v} \in V$

$$\Updownarrow$$

$$A^T(Av - Y) = 0$$

$$\Rightarrow A^T A v = A^T Y \Rightarrow v = (A^T A)^{-1} A^T Y$$

$A^T A$ (square) invertible $\Leftrightarrow$ $\mathcal{M}(A^T A)$ is bijection

$\Leftrightarrow$ $\mathcal{M}(A^T A)$ is injective $\Leftrightarrow$ $\mathcal{M}(A^T A)$ is surjective

$\Leftrightarrow$ $\text{Null}(A^T A) = 0$ $\Leftrightarrow$ $\text{Null}(A) = 0$ $\Leftrightarrow$ cols of A are inde

$$\left(Av = 0 \Leftrightarrow A^T A v = 0 \right)$$

<u>Summary</u> : if cols of A are inde then $v = (A^T A)^{-1} A^T Y$

<u>Cramer</u> $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix} \rightarrow \begin{cases} x = \dfrac{\begin{vmatrix} e & b \\ f & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \\ y = \dfrac{\begin{vmatrix} a & e \\ c & f \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \end{cases}$

LR : $A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$, $Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$

If X_i are diff then cols of A are inde

Find $\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$ to minimize $\left\| A \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} - Y \right\|$

$$(A^T A) \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = A^T Y$$

$$A^T A = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{pmatrix} \qquad A^T Y = \begin{pmatrix} n\bar{Y} \\ \sum_i x_i Y_i \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} n\bar{Y} \\ \sum x_i Y_i \end{pmatrix}$$

$$\hat{\alpha} = \frac{\begin{vmatrix} n\bar{Y} & n\bar{x} \\ \sum x_i Y_i & \sum x_i^2 \end{vmatrix}}{\begin{vmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{vmatrix}} = \frac{n\bar{Y} \sum_i x_i^2 - n\bar{x} \sum x_i Y_i}{n \sum_i x_i^2 - n^2 \bar{x}^2}$$

$$= \frac{\bar{Y} \sum_i x_i^2 - \bar{x} \sum_i x_i Y_i}{\sum_i x_i^2 - n\bar{x}^2}$$

$$\hat{\beta} = \ldots = \frac{\sum x_i Y_i - n\bar{x}\bar{Y}}{\sum_i x_i^2 - n\bar{x}^2}$$

- <u>Covariance Formula</u> : $S_{xy} = \sum_i (x_i - \bar{x})(Y_i - \bar{Y})$

$$= \sum_i x_i Y_i - \sum_i x_i \bar{Y} - \sum_i \bar{x} Y_i + n\bar{x}\bar{Y}$$

$$= \sum_i x_i Y_i - \bar{Y} \sum_i x_i - \bar{x} \sum Y_i + n \bar{x} \bar{Y}$$

$$= \sum_i x_i Y_i - n \bar{x} \bar{Y} = \sum_i \left((x_i - \bar{x}) Y_i \right)$$

$$\hat{\beta} \overset{\text{Simplified}}{=} \frac{S_{xY}}{S_{xx}}$$

·) Want to show : $\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x}$

$$\begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_i x_i^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} n\bar{Y} \\ \sum x_i Y_i \end{pmatrix}$$

$$\Rightarrow n\hat{\alpha} + n\bar{x}\hat{\beta} = n\bar{Y}$$

$$\Rightarrow \hat{\alpha} + \bar{x}\hat{\beta} = \bar{Y}$$

·) <u>linear regression in R:</u>

- <u>M1</u>: $f = lm(Y \sim x)$

<u>M2</u> : use formulas

·) <u>Randomness in regression</u> : Assume there is a (linear) relation:

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

α, β : unknown constants

ε_i : random vars (before observation)

$\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2) \to \begin{cases} Y \text{ become RV} \\ \hat{\alpha}, \hat{\beta} \text{ (calculated from } x_i, Y_i) \\ \text{also become RV} \end{cases}$

(After observing the sample $(x_1, Y_1), \dots, (x_n, Y_n)$

they collarpse into constr)

$Y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma_\varepsilon^2)$ (inde, but not iid)

$$\overline{Y} \sim \mathcal{N}\left(\frac{\sum_i (\alpha + \beta x_i)}{n}, \frac{\sigma_\varepsilon^2}{n}\right) \sim \mathcal{N}\left(\alpha + \beta \bar{x}, \frac{\sigma_\varepsilon^2}{n}\right)$$

$$S_{xY} = \sum_i (x_i - \bar{x}) Y_i \sim \mathcal{N}(\beta S_{xx}, S_{xx} \sigma_\varepsilon^2)$$

$$\left(\begin{aligned} &Var(S_{xY}) = \sum_i Var(x_i - \bar{x}) Y_i = \sum (x_i - \bar{x})^2 \sigma_\varepsilon^2 = S_{xx} \sigma_\varepsilon^2 \\ &\mathbb{E}(S_{xY}) = \sum_i (x_i - \bar{x})(\alpha + \beta x_i) = \alpha \underbrace{\sum_i (x_i - \bar{x})}_{\to 0} \\ &+ \beta \sum_i (x_i - \bar{x}) x_i = \beta S_{xx} \end{aligned} \right)$$

- $$\hat{\beta} = \frac{S_{xY}}{S_{xx}} \sim \mathcal{N}\left(\beta, \frac{\sigma_\varepsilon^2}{S_{xx}}\right)$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta} \bar{x} = \sum_i \frac{1}{n} Y_i - \sum_i \frac{\bar{x}(x_i - \bar{x})}{S_{xx}} Y_i$$

$$= \sum_i \left(\frac{1}{n} - \frac{\bar{x}}{S_{xx}} (x_i - \bar{x}) \right) Y_i \sim \mathcal{N}(\dots$$

$$\mathbb{E}(\hat{\alpha}) = \mathbb{E}\bar{Y} - \bar{x}\,\mathbb{E}\hat{\beta} = (\alpha+\beta\bar{x}) - \bar{x}\beta$$

$\mathbb{E}(\hat{\alpha}) = \alpha$
$\mathbb{E}(\hat{\beta}) = \beta$
$\downarrow$ unbiased estimator

$$Var(\hat{\alpha}) = \sum_i \left(\frac{1}{n} - \frac{\bar{x}}{S_{xx}}(x_i-\bar{x})\right)^2 Var\,Y_i$$

$$= \sigma_\varepsilon^2 \sum_i \left(\frac{1}{n} - \frac{\bar{x}}{S_{xx}}(x_i-\bar{x})\right)^2$$

$$= \sigma_\varepsilon^2 \frac{1}{n^2 S_{xx}^2} \sum_i \left(S_{xx} - n\bar{x}(x_i-\bar{x})\right)^2$$

$$= \sigma_\varepsilon^2 \cdot \frac{1}{n^2 S_{xx}^2} \sum_i \left(S_{xx}^2 - 2n\bar{x}(x_i-\bar{x}) + n^2\bar{x}^2(x_i-\bar{x})^2\right)$$

$$= \sigma_\varepsilon^2 \cdot \frac{1}{n^2 S_{xx}^2}\left(\sum_i S_{xx}^2 - 0 + n^2\bar{x}^2\sum_i (x_i-\bar{x})^2\right)$$

$$= \sigma_\varepsilon^2 \cdot \frac{1}{n^2 S_{xx}^2}\cdot\left(n S_{xx}^2 + n^2\bar{x}^2 S_{xx}\right)$$

$$= \frac{\sigma_\varepsilon^2}{n S_{xx}}\left(\sum_i x_i^2\right)$$

- <u>z - normal</u> :

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{\sigma_\varepsilon^2}{S_{xx}}\right) \Rightarrow \frac{\hat{\beta}-\beta}{\frac{\sigma_\varepsilon}{\sqrt{S_{xx}}}} \sim \mathcal{N}(0,1)$$

- <u>what if the σ_ε^2 is unknown</u> :

$$\frac{\hat{\beta}-\beta}{\sqrt{\frac{SSR}{(n-2)S_{xx}}}} \sim \frac{\mathcal{N}\left(0, \frac{\sigma_\varepsilon^2}{S_{xx}}\right)}{\sqrt{\frac{\sigma_\varepsilon^2}{S_{xx}(n-2)}\chi^2(n-2)}} \sim \frac{\mathcal{N}(0,1)}{\sqrt{\frac{\chi^2(n-2)}{n-2}}}$$

$\sim t_{n-2}$

•) Let ξ be a constant $\in \mathbb{R}$

$$\hat{\alpha} + \hat{\beta}\,\xi = \bar{Y} - \hat{\beta}(\bar{x} - \xi) = \text{lin com of } Y_i$$

•) <u>Prediction</u> : Have $(x_1, Y_1), \dots (x_n, Y_n)$

Predict (x_{n+1}, Y_{n+1})

$$Y_{n+1} \sim \mathcal{N}(\alpha + \beta x_{n+1}, \sigma_\varepsilon^2)$$

$$\underbrace{Y_{n+1} - \underbrace{\hat{\alpha} - \hat{\beta}\, x_{n+1}}_{\text{lin com of } Y_1 \dots Y_n}}_{\text{lin com of } Y_1, \dots, Y_{n+1}}$$

•) <u>Sample corellation coeff</u> :

$$r = \frac{\sum_i (x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \cdot \sqrt{\sum_i (Y_i - \bar{Y})^2}} \quad , i \in \{1, \dots, n\}$$

• $$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var} X}\sqrt{\text{Var} Y}}$$

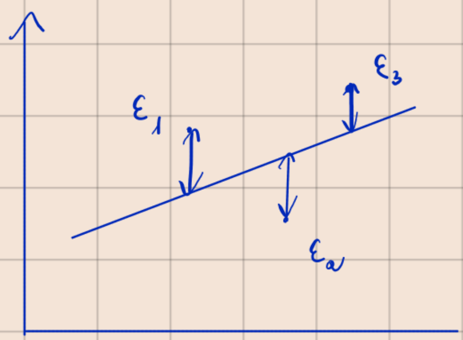
- $r \in [-1, 1]$

Proof : let $\vec{u} = (x_i - \bar{x})_{i=1}^{n}$

$$\vec{v} = (y_i - \bar{y})_{i=1}^{n}$$

$$1 \leq r = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|} \leq 1 \quad (\text{Cauchy schwarz})$$

- If $(\varepsilon_i)_{obs} = 0$ $\quad Y_i = \alpha + \beta x_i + \varepsilon_i$



$$(Y_i)_{obs} = \alpha + \beta x_i$$

$$(\bar{Y})_{obs} = \alpha + \beta \bar{x}$$

$$(Y_i - \bar{Y})_{obs} = \beta (x_i - \bar{x})$$

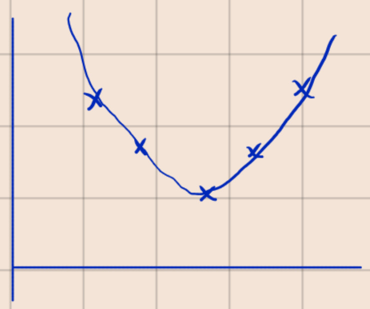
$$\rightarrow (\vec{v})_{obs} = \beta (\vec{u})$$

$$\Rightarrow r_{obs} = \frac{\vec{u} \, (\beta \vec{u})}{\|\vec{u}\| \, \|\beta \vec{u}\|} = \frac{\beta \|\vec{u}\|^2}{|\beta| \, \|\vec{u}\|^2} = \frac{\beta}{|\beta|}$$

$$\rightarrow r^2_{obs} = 1$$

• **General regression**: Sometimes linear regression is not suitable

Ex :



$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 \cos(x_i) + \varepsilon_i$$

where β_i are const and ε_i are random var

• **Get least-squared estimators $\hat{\beta_0}, \hat{\beta_1}, \hat{\beta_2}, \hat{\beta_3}$:**

Find $\begin{pmatrix} \hat{\beta_0} \\ \vdots \\ \hat{\beta_3} \end{pmatrix}$ for $\underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 & \cos x_1 \\ & & \vdots & \\ 1 & x_n & x_n^2 & \cos x_n \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \hat{\beta_0} \\ \hat{\beta_1} \\ \hat{\beta_2} \\ \hat{\beta_3} \end{pmatrix}}_{\hat{\beta}} \underset{\approx}{\text{closest}} \underbrace{\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}}_{\vec{Y}}$

$$\hat{\beta} = (A^T A)^{-1} A^T \vec{Y} \qquad \text{Assuming } \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$$

$\hat{\beta_i}$?

To show pdf of $\hat{\beta}$, we need <u>random linear algebra</u>

• $\mathbb{E}$: $\mathbb{E}\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix}$

$$\mathbb{E}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \mathbb{E}(a_{11}) & \mathbb{E}(a_{12}) \\ \mathbb{E}(a_{21}) & \mathbb{E}(a_{22}) \end{pmatrix}$$

• Cov : $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$, $\vec{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$

$$\text{Cov}(\vec{X}, \vec{Y}) = \begin{pmatrix} \text{Cov}(X_1, Y_1) & \cdots & \text{Cov}(X_1, Y_n) \\ \vdots & \ddots & \\ \text{Cov}(X_n, Y_1) & \cdots & \text{Cov}(X_n, Y_n) \end{pmatrix}$$

① $\text{Cov}(\vec{X}, \vec{Y})_{ij} = \text{Cov}(X_i, Y_j) \quad \forall i, j$

② $\text{Cov}(\vec{X}, \vec{Y}) = \mathbb{E}\left((\vec{X} - \mathbb{E}\vec{X})(\vec{Y} - \mathbb{E}(\vec{Y}))^T \right)$

• ② $\Rightarrow \text{Cov}(\vec{X}, \vec{Y})_{ij} = \mathbb{E}\left(\left(X_i - \mathbb{E}(X_i)\right)\left(Y_j - \mathbb{E}(Y_j)\right) \right)$

$= \mathbb{E}(X_i Y_j) - \mathbb{E}(Y_j)\mathbb{E}(X_i) - \mathbb{E}(X_i)\mathbb{E}(Y_j)$

$+ \mathbb{E}(X_i)\mathbb{E}(Y_j) = \mathbb{E}(X_i Y_j) - \mathbb{E}(Y_j)\mathbb{E}(X_i) = \text{Cov}(X_i, Y_j)$

$\text{Var}(\vec{X}) = \text{Cov}(\vec{X}, \vec{X})$

(want to show)

•) WTS $\text{Var}(\vec{X}) \succcurlyeq 0$

$M \succcurlyeq 0 \iff \begin{cases} M \text{ sym} \\ \text{eigvals} \geq 0 \end{cases}$

$\iff \begin{cases} M \text{ sym} \\ (M\vec{v})\,\vec{v} \geq 0 \quad \forall \vec{v} \in \mathbb{R}^n \setminus \{0\} \end{cases}$

$M > 0 \iff \begin{cases} \ldots \\ \ldots > 0 \end{cases}$

$\iff \begin{cases} \ldots \\ \ldots > 0 \end{cases}$

If $M \geq 0$: M is invertible $\iff M > 0$

$$\mathrm{Cov}(X, X) = \mathbb{E}\big(\underbrace{(X - \mathbb{E}(X))}_{Y}(X - \mathbb{E}(X))^T\big) \geq 0$$

WTS : $\mathbb{E}(Y Y^T) \geq 0$

$\forall \vec{v} \in \mathbb{R}^n \setminus \{0\}$, we want to show that $(\mathbb{E}(YY^T)\vec{v})\vec{v} \geq 0$

$\cdot\ (\mathbb{E}(YY^T)\vec{v})\vec{v} \overset{\text{Linearity}}{=} \mathbb{E}((YY^T\vec{v})\vec{v})$ $\quad (A\vec{u})\vec{v} = \vec{u}(A^T\vec{v})$

$= \mathbb{E}((Y^T\vec{v})(Y^T\vec{v})) = \mathbb{E}(\|Y^T\vec{v}\|^2) \geq 0$

$\cdot)\ \mathrm{Var}(\underset{\substack{\downarrow \\ \text{fixed} \\ \text{matrix}}}{A}\vec{x} + \underset{\substack{\downarrow \\ \text{fixed} \\ \text{vector}}}{\vec{v}}) = A\,\mathrm{Var}(\vec{x})\,A^T$

<u>Sol</u> : $\mathrm{Var}(A\vec{x} + \vec{v}) = \mathrm{Var}(A\vec{x}) = \mathrm{Cov}(A\vec{x}, A\vec{x})$

$$= \mathbb{E}\left(\left(A\ X - \mathbb{E} A X\right)\left(A X - \mathbb{E} A X\right)^T\right)$$

$$= \mathbb{E}\left(\left(A X - A \mathbb{E} X\right)\left(A X - A \mathbb{E} X\right)^T\right)$$

$$= \mathbb{E}\left(A(X - \mathbb{E} X)(X - \mathbb{E} X)^T A^T \right) = A\, \mathbb{E}\left(\left(X - \mathbb{E} X\right)\left(X - \mathbb{E}(X)\right)^T\right)$$

$$A^T = A \operatorname{Var}(X)\, A^T$$

•) $\vec{Z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$. Assume $f_{\vec{Z}}(\vec{x}) = \dfrac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$

$$= f_{z_1}(x_1) \dots f_{z_n}(x_n) \rightarrow \text{tensor product} \rightarrow \text{inde}$$

•) Assume $\vec{X}$ has CF $\varphi_{\vec{X}}(\vec{t}) = \mathbb{E}\left(\exp(i \vec{t} \cdot \vec{x})\right)$

$$= \exp\left(i t \vec{\mu} - \frac{1}{2} t^T \underbrace{\Sigma}_{n \times n \text{ matrix}} t\right)$$

where $\mu \in \mathbb{R}^n$, Σ is a $n \times n$ matrix

Then we say X is normal RV $X \sim \mathcal{N}(\vec{\mu}, \Sigma)$

<u>Note</u> : X might not have any pdf, $\operatorname{var} X = \Sigma$ might be non-invertible $\Leftrightarrow$ no pdf

When Σ is invertible the pdf is

$$f_{\vec{x}}(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})\right)$$

1D : $\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-1}{2\sigma^2} (x-\mu)^2 \right)$

Link between pdf and cf :

$$\text{pdf} \xrightarrow[\text{integral}]{\text{Gaussian}} \text{MGF} \xrightarrow{\text{complexify}} \text{CF}$$

n-dim Gaussian int ?

·) Assume $\vec{x} \sim \mathcal{N}(\mu, \Sigma)$

Let $\vec{Y} = \underbrace{A}_{\text{fixed matrix}} \vec{x} + \underbrace{\vec{b}}_{\text{fixed vector}}$

What RV is $\vec{Y}$?

Solution : $\varphi_{\vec{Y}}(\vec{t}) = \mathbb{E}(\exp(i\vec{t}\cdot\vec{Y}))$

$$= \mathbb{E}\left(\exp\left(i\vec{t}\cdot(A\vec{x}+\vec{b})\right)\right)$$

$$= \mathbb{E}\left(\exp(i\vec{t}\cdot A\vec{x})\exp(i\vec{t}\cdot\vec{b})\right)$$

$$= \exp(i\vec{t}\cdot\vec{b}) \cdot \mathbb{E}\left(\exp(iA^T\vec{t}\cdot\vec{x})\right)$$

$$= \exp(i\vec{t}\cdot\vec{b}) \cdot \varphi_{\vec{x}}(A^T\vec{t})$$

$$= \exp\left(i\vec{t}\cdot\vec{b} + iA^T\vec{t}\cdot\vec{\mu} - \frac{1}{2}(A^Tt)^T\Sigma A^Tt\right)$$

$$= \exp\left(i\vec{t}\cdot\vec{b} + i\vec{t}\cdot A\vec{\mu} - \frac{1}{2}t^TA\Sigma A^Tt\right)$$

$$= \exp\left(i\vec{t}(\vec{b}+A\vec{\mu}) - \frac{1}{2}t^TA\Sigma A^T\vec{t}\right)$$

$\Rightarrow Y \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$

True without independence !

Ex let X, Y be jointly normal RV

Find pdf of $X+Y$

$$(1 \quad 1)\begin{pmatrix} X \\ Y \end{pmatrix} = X + Y$$

Note : X, Y normal doesn't imply $\begin{pmatrix} X \\ Y \end{pmatrix}$ is normal

$\begin{pmatrix} X \\ Y \end{pmatrix}$ is normal

$\Updownarrow$

X, Y jointly normal

•) Assume $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ is normal

$\operatorname{Var} \vec{X} = \Sigma$ is invertible

Thm : the X_i are inde

$\Leftrightarrow$ the X_i are uncorrellated

Proof : • indep $\Rightarrow$ uncorr

• Assume X_i are uncorr

•) Lin alg :

Symmetric :

Properties :

① $(Ax)\, y = x\, Ay$

② Spectral Theorem :

preserves dot product

Symmetric = orthogonally diagonalizable

There is an orthogonal basis (change of basis)

Where matrix becomes diagonal

In matrix form

A sym $\Rightarrow A = P D P^{-1}$

where $\begin{cases} D \text{ is diagonal} \\ D \text{ is orthogonal } (P^{-1} = P^{T}) \end{cases}$

③ Square root :

Let $A \geq 0$ (sym + eig vals > 0)

By spectral : $A = P D P^{-1}$

Then define $\sqrt{A} = P \sqrt{D} P^{-1}$

where $\sqrt{D}$ is entry wise sqrt of D

$$D = \begin{pmatrix} 1 & & \\ & 4 & \\ & & 9 \end{pmatrix}, \quad \sqrt{D} = \begin{pmatrix} 1 & & \\ & 2 & \\ & & 3 \end{pmatrix}$$

$$\sqrt{A}\sqrt{A} = P\sqrt{D}P^{-1} P\sqrt{D}P^{-1} = P\sqrt{D}\sqrt{D}P^{-1} = PDP^{-1} = A$$

•) Let $\vec{X}$ be n-dim normal

Assume $Var\,\vec{X}$ is invertible

Then we have $$\big(\underbrace{\sqrt{Var X}}_{\text{const}}\big)^{-1}\big(X - \underbrace{\mathbb{E}(X)}_{\text{const}}\big) \sim \mathcal{N}(0, I_n)$$

<u>Proof</u> :

linear transform of normal is normal

$$\mathbb{E}(\dots) = 0$$

$$\begin{aligned} Var(\dots) &= \big(\sqrt{Var X}\big)^{-1} Var(X - \mathbb{E}X)\big(\sqrt{Var X}\big)^{-T} \\ &= \big(\sqrt{Var X}\big)^{-1} Var X \big(\sqrt{Var X}\big)^{-1} \\ &= \big(\sqrt{Var X}\big)^{-1}\sqrt{Var X}\sqrt{Var X}\big(\sqrt{Var X}\big)^{-1} \\ &= I_n \end{aligned}$$

<u>Rewrite</u> : $X - \mathbb{E}X = \big(\sqrt{Var X}\big) Z$

where Z is n-dim standard normal

⊛ When $Var\,X$ is not invertible (X normal)

Then there is $n \times k$ matrix A s.t

$$\begin{cases} Var\,X = AA^T \\ k = rank(Var\,X) \quad (k < n) \end{cases}$$

$$\begin{pmatrix} 1 & & \\ & 4 & \\ & & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}$$

$$D = B B^T$$

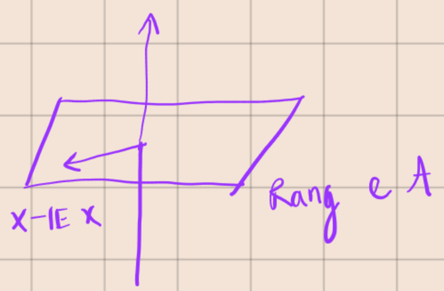
$$A = P D P^T = \underbrace{P B}_{A} \underbrace{B^T P^T}_{A^T}$$

<u>Z - normalization</u> : $X - \mathbb{E}X = A \cdot Z$ (black box)

where Z is k - dim standard normal

• <u>Picture</u> : Var X is not invertible $\Rightarrow$ Value of $X - \mathbb{E}X$ is contained in a $\underbrace{\text{subspace}}_{\downarrow}$

range of A



$$\overset{\mathcal{N}(0,1)^2}{\downarrow}$$

• <u>chi-squared</u> : 10 : $\dfrac{(X - \mathbb{E}X)^2}{\operatorname{Var} X} \sim \chi^2(1)$ if X is normal

Let X be n - dim normal, and assume Var X is invertible

Then $(X - \mathbb{E}X)^T (\operatorname{Var} X)^{-1} (X - \mathbb{E}X) \sim \chi^2(n)$

<u>Proof</u> By Z - normal, $X - \mathbb{E}(X) = \left(\sqrt{\operatorname{Var} X}\right) Z \quad \left(Z \sim \mathcal{N}(0,1)\right)$

$$\Rightarrow (X - \mathbb{E}X)^T (\mathrm{Var}\, X)^{-1} (X - \mathbb{E}X)$$

$$= \left(\sqrt{\mathrm{Var}X}\; Z\right)^T (\mathrm{Var}\, X)^{-1} \left(\sqrt{\mathrm{Var}X}\; Z\right)$$

$$= Z^T \sqrt{\mathrm{Var}\, X}\, (\mathrm{Var}\, X)^{-1} \sqrt{\mathrm{Var}\, X}\; Z$$

$$= Z^T Z = Z_1^2 + \dots + Z_n^2$$

$$\begin{cases} Z_i \sim \mathcal{N}(0,1) \Rightarrow Z_i^2 \sim \chi^2(1) \\ Z_i\text{'s are inde} \end{cases}$$

① jointly normal + uncorrellated

② pdf = tensor product

$$\chi^2(1) + \dots + \chi^2(1) = \chi^2(n)$$

<u>General case</u> :

let $\begin{cases} Z \sim \mathcal{N}(0, I_n) \\ B_1, B_2 \text{ sym} \end{cases}$ $\begin{cases} B_1 + B_2 = I_n \\ \mathrm{rank}(B_1) + \mathrm{rank}(B_2) = n \end{cases}$

Then $\begin{cases} Z^T B_1 Z \\ Z^T B_2 Z \end{cases}$ are inde

and $Z^T B_i Z \sim \chi^2(\mathrm{rank}(B_i))$

<u>Proof</u> : WLOG, B_1 is diagonal

$$B_1 = \begin{pmatrix} a & & & \\ & b & & \\ & & 0 & \\ & & & 0 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 0 & & & \\ & 0 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

$\Rightarrow \quad a = b = 1$

$$B_1 \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}$$

$$z^T B_1 z = z_1^2 + z_2^2$$

$$z^T B_2 z = z_3^2 + z_4^2$$