Week 2 Note (字跡潦草先預防性道歉)

if there's any question or mistake, mail me:
b09303052@ntu.edu.tw
(蔡尚恩)

Outline: (first 2 hours)

1. Review: matrix
2. Review: univariate distribution
3. Multivariate distribution
4. Review: Covariance
5. PCA: definition

for the 3rd hour:

- Peer review
- HW1
- Data mini workshop

Data v.s. matrix

Age	Gender	Salary

observation

$$\begin{array}{ccc} & n & \\ X = & \begin{pmatrix} \quad & \quad \end{pmatrix} & p \end{array}$$

n variables

p observations

Quiz: $\underset{n\times p}{\underline{\mathbb{X}}}$, $\underset{p\times 1}{\underline{X_i}}$

$$\mathbb{X} = \textcircled{1}\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}_{np\times 1} \text{ or } \textcircled{2}\left(X_1 . X_2 \cdots X_n\right)_{p\times n} \text{ or } \textcircled{3}\begin{pmatrix} X_1^T \\ X_2^T \\ \vdots \\ X_n^T \end{pmatrix}_{n\times p} ?$$

The answer is ③

Other examples:

① Picture

② Text: word count

Univariate distribution ($p=1$)

$$\mathbb{X} = \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{n1} \end{pmatrix} \qquad \underset{1\times 1}{\underline{X_i}} = (X_{i1}) \in \mathbb{R}$$

$$\mathbb{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

$$X_1, X_2, X_n \overset{iid.}{\sim} F_X(\cdot)$$

CDF: Cumulative Distribution Function

CDF fully characterizes the stochastic behavior of X

$F_X(x) = P(X \leq x)$

$P(X > a) = 1 - F_X(a)$

$P(a < X \le b) = P(X \le b) - P(X \le b) = F(b) - F(a)$

CDF fully determines probabilities of events regarding X

Rmk: if the distribution is known, it's easier to work with P.D.F or P.M.F

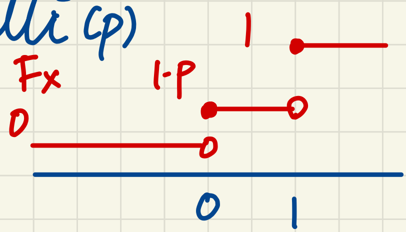
PDF exists in discrete
PMF exists in cont.
CDF always exists

$X \sim \text{Bernoulli}(p)$

$P(X=1) = p$

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

$P(X<0) = 0$

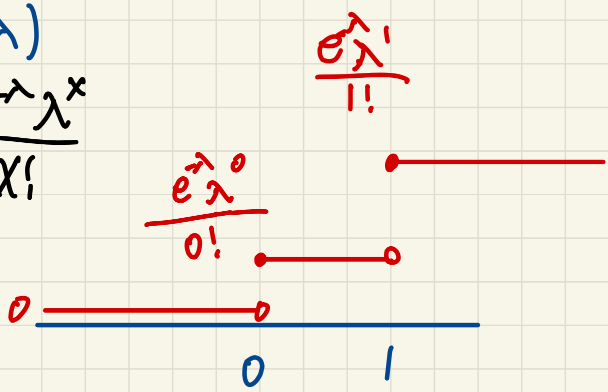


$X \sim \text{Poisson}(\lambda)$

$$P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$x = 0, 1, 2, \ldots$

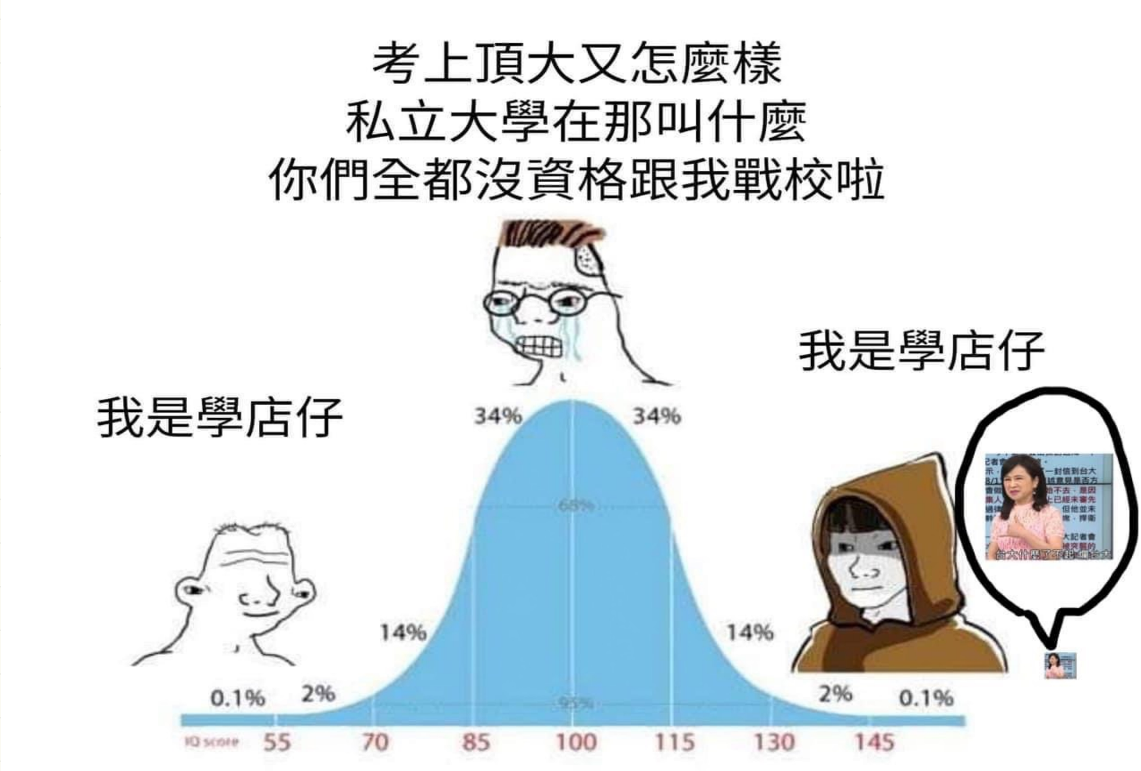
$\lambda > 0$



$X \sim N(\mu, \sigma^2)$

$$F_X = \int_{-\infty}^{x} \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}_{\text{Pdf}}$$

homemade meme describing Normal distribution
↓



Ex: (non discrete, non continuous.)

Distribution of Payment (課金)

Def: a function $f_X(x)$ is said to be a p.d.f. of a r.v. X if

$$P(X\le x) = \int_{-\infty}^{x} f(x)\,dx$$

fact: p.d.f is not unique

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$$

$\tilde{f}_X(x)$ is still pdf of $N($

ex: $$\tilde{f}_X(x) = \begin{cases} f(x) & x\neq 0 \\ 10^{100}, & x=0 \end{cases}$$

$X \sim \exp(\lambda)$ if

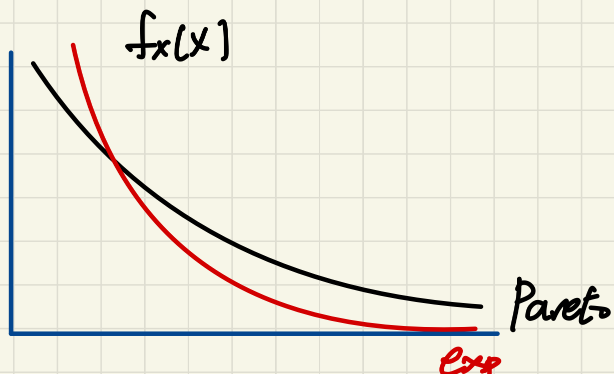
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{o.w.} \end{cases}$$

"o.w." means "otherwise", not "oh wow"

Pareto distribution

$X \sim \text{Pareto}(\alpha)$ if

$$f_X(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}} & x \geq 0 \\ 0, & \text{o.w.} \end{cases}$$



$\alpha = 1$, $f_X(x) = \dfrac{1}{x^2}$

$E(X) = \infty$

	$f(1)$	$f(10)$	$f(100)$
exp(1)	e^{-1}	e^{-10}	e^{-100}
Pareto(1)	1	0.01	0.0001

Remark

Distributions are not purely human constructs.

They emerge naturally depending physical mechanism.

$\exp(\lambda)$: model lifetime of a light bulb

Pareto(α): household wealth. city size

Central limit Theorem (CLT)

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

even X not not normal

CLT

$X_1, X_2 .. X_n \overset{iid}{\sim} F_X(\cdot)$ if $E(X^2) < \infty$

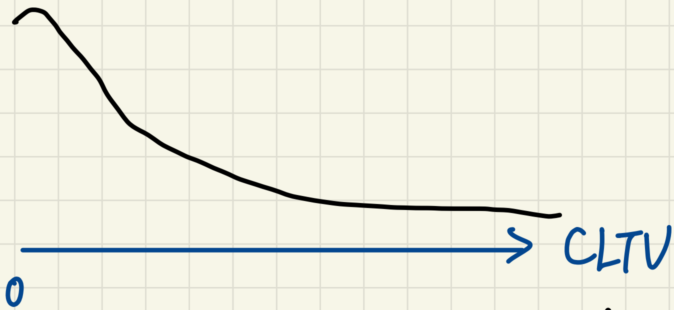
Then $\sqrt{n}(\bar{X}_n - \mu) \to N(0, \sigma^2)$

Customer Lifetime Value (CLTV)

= Contribution of a customer from entry to exit

Application

acquisition cost < CLTV
(獲客成本)



pdf of CLTV, usually it's heavy tail

Multivariate distribution

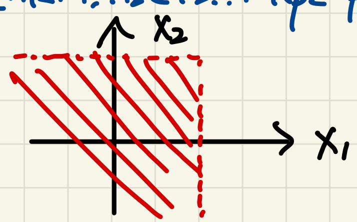
<u>Def</u> A random vector is a vector of random variable $\underset{p\times 1}{X_i} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$

$x_1, x_2, \ldots, x_p$ are random variables

Multivariate CDF

$F_X(x_1, x_2 \ldots x_p) = P(X_1 \le x_1, X_2 \le x_2 \ldots X_p \le x_p)$

e.g. $p=2$



Ex: (multivariate Normal)

X is a multivariate normal if

pdf $f(x_1, x_2, \ldots x_p) = \frac{1}{(\sqrt{2\pi})^p}\left(\frac{1}{\det(\Sigma)}\right)^{\frac{1}{2}}$

$\cdot \exp\left[\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$

$X \sim N_p(\mu, \Sigma)$

$\Sigma \rightarrow$ cov. matrix

$X_{p\times 1} \sim N_p(\mu, \Sigma)$

$\mu = \begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{pmatrix}$ $\Sigma_{p\times p} = (\sigma_{ij})$

$\sigma_{ij} = \begin{cases} Cov(X_i, X_j) & i \neq j \\ Var(X_i) & i = j \end{cases}$

$\Sigma_{p\times p} = \begin{pmatrix} Var(X_1) & & \\ Cov(2,1) & Var(X_2) & \\ & & \ddots \\ & & & Var(X_n) \end{pmatrix}$

$Cov(X_i, X_j) = E[(X_i - E_i)(X_j - EX_j)]$

$= E[(X_j - EX_j)(X_i - EX_i)]$

$= Cov(X_j, X_i) \Rightarrow$ implies Σ is a symmetric matrix

Special Case:

Σ is a diagonal matrix

$$\Sigma = \begin{pmatrix} Var(X_1) & & & & 0 \\ & Var(X_2) & & & \\ & & Var(X_3) & & \\ & 0 & & \ddots & \\ & & & & Var(X_p) \end{pmatrix}$$

$$\det(\Sigma) = Var(X_1)\,Var(X_2) \cdots Var(X_p)$$
$$= \sigma_{11}\,\sigma_{22} \cdots \sigma_{pp}$$

$$(x-\mu)^T \Sigma^{-1} (x-\mu)$$

$$\Sigma^{-1} = \begin{pmatrix} \sigma_{11}^{-1} & & 0 \\ & \sigma_{22}^{-1} & \\ 0 & & \ddots \ \sigma_{pp}^{-1} \end{pmatrix}$$

$$(x-\mu)^T \Sigma^{-1} = (x_1-\mu_1, x_2-\mu_2 \ldots x_p-\mu_p) \cdot \begin{pmatrix} \sigma_{11}^{-1} & & 0 \\ & \sigma_{22}^{-1} & \\ 0 & & \ddots \ \sigma_{pp}^{-1} \end{pmatrix}$$

$$= \left(\sigma_{11}^{-1}(x_1-\mu_1) \cdot \sigma_{22}^{-1}(x_2-\mu_2) \cdot \sigma_{pp}^{-1}(x_p-\mu_p) \right)_{1\times p}$$

$$\Downarrow$$

$$\cdots \cdot \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_p-\mu_p \end{pmatrix}_{p\times 1}$$

$$= \sigma_{11}^{-1}(x_1-\mu_1)^2 + \sigma_{22}^{-1}(x_2-\mu_2)^2 \cdots + \sigma_{pp}^{-1}(x_p-\mu_p)^2$$

When Σ is diagonal.

When Σ diagonal

$$f_X(x_1, \dots x_p) = \left(\frac{1}{\sqrt{2\pi}}\right)^p \left(\frac{1}{\sigma_{11} \cdot \sigma_{22} \cdots \sigma_{pp}}\right)$$

$$\exp\left[\frac{-1}{2}\left(\sigma_{11}^{-1}(x_1-\mu_1)^2 + \dots + \sigma_{pp}^{-1}(x_p-\mu_p)^2\right)\right]$$

$$= \exp\left(\frac{-1}{2}\frac{(x-\mu)^2}{\sigma_{11}}\right) \cdot \exp\left(\frac{-1}{2}\left(\frac{(x_2-\mu_2)^2}{\sigma_{22}}\right)\right) \cdots \exp\left(\frac{-1}{2}\frac{(x_p-\mu_p)^2}{\sigma_{pp}}\right)$$

$$f_X(x_1, x_2, \dots x_p)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right) \cdot \frac{1}{\sqrt{\sigma_{11}}} \cdot \exp\left(\frac{-1}{2}\frac{(x_1-\mu)^2}{\sigma_{11}}\right) \sim N(\mu_1, \sigma_{11})$$

$$\cdot \left(\frac{1}{\sqrt{2\pi}}\right) \cdot \frac{1}{\sqrt{\sigma_{22}}} \exp\left(\frac{-1}{2}\frac{(x_2-\mu_2)^2}{\sigma_{22}}\right) \sim N(\mu_2, \sigma_{22})$$

$$\vdots$$

$$\left(\frac{1}{\sqrt{2\pi}}\right) \frac{1}{\sqrt{\sigma_{pp}}} \exp\left(\frac{-1}{2}\frac{(x_p-\mu_p)^2}{\sigma_{pp}}\right) \sim N(\mu_p, \sigma_{pp})$$

Conclusion

When Σ diagonal $\Sigma = \begin{pmatrix} \sigma_{11} & & 0 \\ & \sigma_{22} & \\ 0 & & \ddots \ \sigma_{pp} \end{pmatrix}$

pdf $f(x_1 x_2 \dots x_n) = f_X(x_1) \cdot f_X(x_2) \cdots f_p(x_p)$

pdf of $N(\mu_p, \sigma_{pp})$

$(p=2)$ $\Sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$

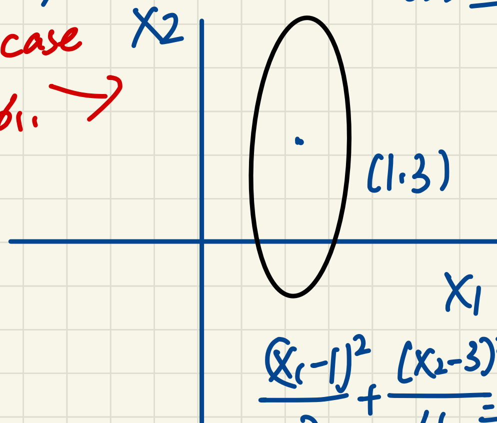
$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \cdot \left(\frac{1}{\sqrt{\sigma_{11}\cdot\sigma_{22}}}\right) \cdot \exp\left[\frac{-(x_1-\mu_1)^2}{\sigma_{11}} + \frac{-(x_2-\mu_2)^2}{\sigma_{22}}\right]$$

$\exists R \in \mathbb{R}$

in this case $\sigma_{22} > \sigma_{11}$ →

$f_X(x_1, x_2) = k$

$$\Longleftrightarrow \frac{(x_1-\mu_1)^2}{\sigma_{11}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}} = R'$$



$$\frac{(x_1-1)^2}{2} + \frac{(x_2-3)^2}{4} = k$$

HW1

How to estimate CDF?

$F(x) = P(X \le x)$ $\quad X_1, X_2 \dots X_n$

$\hat{F}_n = \frac{1}{n} \sum \mathbb{1}\{X_i \le x\}$ "sample analogue

$$\mathbb{1}\{X_i \le x\} = \begin{cases} 1 & X_i \le x \\ 0 & \text{o.w.} \end{cases}$$

$\mathbb{1}\{X_i \le x\} \sim Ber(p)$

$p = P(X_i \le x)$

$\Rightarrow E\hat{F}_n(x) = p$