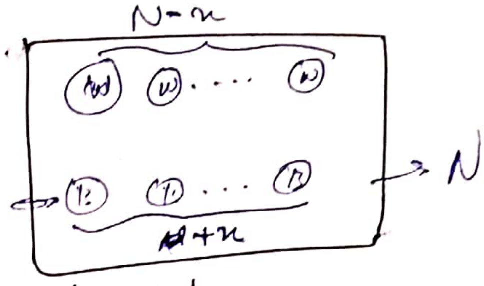
RANDOM SAMPLING : (#R1)



Population

(No of black & white balls)

$\frac{\#W}{\#B} = p$ (parameter)

$p = \frac{N-n}{n}$

$= N/n - 1$

→ Parameter p estimated by drawing smaller sets.

→ This is called sampling.

→ Suppose n-objects collected, it is called sample of size n.

Samples are mostly drawn randomly.

⊛ Random samples: Each individual has identical probability of being included in the sample.

Such Samples are called simple random samples. Process is Simple Random Sampling.

① Sim. Rand. Sampling with replacement:

(SRSWR)

② Simple Random Sampling without replacement (SRSWOR)

③ Polya Sampling.

After every draw, another same ball is added along with picked ball.

If ball drawn is W:

Like: $N-K+1 \to W$

$K \to B$

In k^{th} draw:

Total balls: $N+K$.

RANDOM VARIABLES / VECTORS

Random experiment with sample space Ω is given.

Any measurable function $X : \Omega \to \mathbb{R}^k$

is called a random $\begin{cases} \text{variable if } k = 1 \; (k \in \mathbb{Z}^+) \\ \text{vector if } k > 1 \end{cases}$

Scanned by CamScanner

Ex: $\Omega_n \rightarrow$ sample space of n independent tosses of a coin.

H $\rightarrow$ success,

$X : \Omega_n \longrightarrow \mathbb{R}$ by $X(\omega) = \#\, H\text{'s in } \omega$

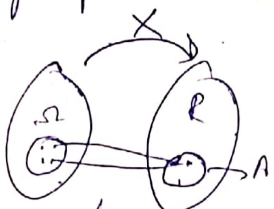
$P(H) = p \in [0, 1]$

$X(\omega) = \{0, 1, \ldots n\} \rightarrow$ Range or Spectrum of random Variable.

$A \subset \mathbb{R},\ P(X \in A) = \{\underset{\downarrow}{\omega \in \Omega} \mid X(\omega) \in A\}$

Subspace of sample space (event)

$= P_0(X^{-1}(A))$



Inverse of range taken.

$X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$

$$P_0 \underbrace{\left[X^{-1}(\mathbb{R})\right]}_{= \Omega} = 1 \quad \text{(Identity)}$$

$\Omega_n \rightarrow \mathbb{R}$, $P(H) = P$

$X \in \{0, 1, \ldots n\}$ [pre image is not NULL case]

$$P(\underbrace{H, H, \cdots H}_{k}, \overbrace{T, \cdots T}^{n-k}) = P(H \cap H \cap \cdots H \cap T \cap T \cdots T)$$

or any permutation of $= \dfrac{\lfloor n}{\lfloor k \ \lfloor n-k}$ $= (P(H))^k \cdot (P(T))^{n-k}$

$= p^k \cdot q^{n-k}$, $q \rightarrow (1-p)$

$P(X = k) = \binom{n}{k}$

In case of dice: (Multinomial Distribution)

$X : \Omega_n$

$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix} : \underset{\omega}{\Omega_n} \rightarrow \mathbb{R}$

$\Omega_n = (\omega_1 \ldots \omega_n)$, $\omega_n \in \{1, \ldots 6\}$

$\omega_i \in \{1, 2 \ldots 6\}$

For dice, $p_1 = P(\text{Face} = 1)$

$p_2 = P(\text{Face} = 2)$

$$\begin{pmatrix} X_1 \\ \vdots \\ X_6 \end{pmatrix} (\omega_1, \dots, \omega_6) = \begin{pmatrix} \#1\text{s in } n \text{ rolls} \\ \#2\text{s} \;\; " \;\; " \;\; " \\ \vdots \\ \#6\text{s} \;\; " \;\; " \;\; " \end{pmatrix}$$

$$X_1(\omega_k) = \begin{cases} 1 & \text{if } \omega_k = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$X_2(\omega_k) = \begin{cases} 1 & \text{if } \omega_k = 2 \\ 0 & " \end{cases}$$

$$X_5(\omega_k) = 1 \text{ if } \omega_k = 5$$
$$0 \text{ otherwise.}$$

$$\underset{\sim}{\omega} = (\omega_1, \omega_2 \dots \omega_n)$$

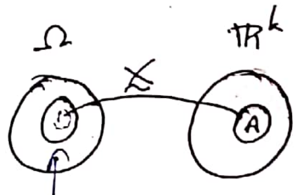
$$\underset{\sim}{X}(\underset{\sim}{\omega}) = \begin{pmatrix} X_1(\underset{\sim}{\omega}) \\ \vdots \\ X_6(\underset{\sim}{\omega}) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} X_1(\omega_k) \\ \vdots \\ \sum_{k=1}^{n} X_6(\omega_k) \end{pmatrix}$$ [Total respective numbers obtained]

(Mass function) | Calculate Multinomial Distribution.

$$\binom{n}{k_1 \; k_2 \dots k_6} = \frac{\lfloor n}{\lfloor k_1 \lfloor k_2 \dots \lfloor k_6}$$

BINOMIAL AND MULTINOMIAL

Binary vector — But other coord. obtained from first. | ↳ Random Vector



$$\underset{\sim}{X} : \Omega \to \mathbb{R}^k$$

inv of A $\underset{\sim}{X}^{-1}(A)$

Prob. Mass Function (pmf) ← $P(\underset{\sim}{X} \in A)$, $A \subseteq \mathbb{R}^k$

$$= P\{\omega \in \Omega \mid \underset{\sim}{X}(\omega) \in A\}$$

For binomial, Ex: $n = 100$, $X \leq 10$ $\quad p = P(H)$

$\cdot\; X \in (-\infty, 10]$

X is r.v, it takes $\mathbb{Z}^+, 0$.

$$\therefore P(X) = P((X=0) \cup (X=1) \cup (X=2) \dots \cup (X=10))$$

Scanned by CamScanner

At times, a r.v is defined solely by defining its p.m.f.

A r.v X is defined by its pmf

$$p_X : \mathbb{R} \to \mathbb{R}$$

where $p_X(i) = P_r(X=i), \forall i \in \mathbb{R}$

Poisson Random Variable

$$p_X(i) = \frac{e^{-\lambda} \lambda^i}{\lfloor i}, \quad i = 0, 1, 2 \ldots$$

A r.v X is said to follow Poisson Distribution with parameter λ $(\lambda > 0)$, if its p.m.f is given by above p_X.

f_i	E_i

Expected Probability.

# of typos in a page

0	f_0	$f_0/1000$
1	f_1	$f_1/1000$
...	...	...
9 or more	f_9	$f_9/100$

↳ Relative freq. table.

Total pages $= \sum_{i=0} f_i = 1000$

Karl Pearson's Test

$$\chi^2 = \sum_{i=0}^{9} \frac{(O_i - E_i)^2}{E_i}$$

Gives measure of variation.
If variations less χ^2 is less,
else it is more varied data.

Find mean: $\sum x_i \frac{f_i}{1000} \doteq \hat{\lambda}$

$P_\lambda(i) = \frac{e^{-\lambda} \lambda^i}{\lfloor i}$, $i = 0, 1, \ldots 8$ ~~cal~~ separately

For $i = 9$, add its ~~to the~~ prob.

Actuals $= 1000 \times P_x$

Ω = A coin is tossed repeatedly & independently until a H comes up.
As soon as 1 H comes up the exp. is stopped and no. of tosses req. are counted.

X = # tosses to get one head for the first time

$X = \{1, 2, \ldots\} \cup \{+\infty\}$

$\Omega = \{H, TH \ldots, T \ldots TH\} \cup \{T \ldots T \ldots\}$

$P(X=i) = q^{i-1}p$

$P(H) = p$, ~~P(H)~~ $P(T) = 1-p = q$

$$P(X=i) = P(\underbrace{T \dots T}_{(i-1)} H)$$
$$= (P(T))^{i-1} \cdot P(H)$$
$$= q^{i-1} \cdot p$$

If $p = 1$

$P(X=1) = 1$ All other X $P(X) = 0$

$p \to 0 \Rightarrow P(X = +\infty) = 1, = ?$

It is called geometric distribution (due to G.P.) and variable is called geometric random ~~for~~ variable.

~~for getting k heads~~

~~For K tosses, for the~~

K-heads obtained:

$X = \{k, k+1, \dots\} \cup \{+\infty\}$

$\binom{k+l-1}{l}$ $k+l$

$[X = n]$ $n \geq k$ n

$X = k + l$ (heads, tails)

$P(X) = p^k q^l \binom{n-1}{k-1}$

Negative Binomial Random Variable (p, k) $\Rightarrow kGEO(p)$

$$\boxed{(1+x)^{-n} = \sum_{r=0}^{\infty} \binom{-n}{r} x^r \quad -1 < x \leq 1}$$ → Negative binomial theorem

Hyper-Geometric Random Variable

$\boxed{M \text{ (W) balls} + N \text{ (B) balls}}$

I draw balls at random simultaneously from the box. What is the probability that $\exists$ k white balls in my sample?

r = no. of white balls in sample

$$p(X=k) = \frac{\binom{M}{k}\binom{N}{r-k}}{\binom{M+N}{r}} \quad \_ \leq k \leq \_$$

Hypergeometric (M, N, r)

$p_n(i)$ is p.m.f of any discrete R V.

$$p_n(i) \geq 0 \ \forall i$$

$$\sum_i p_x(i) = 1$$

$$P(\Omega) = 1$$

$M-k \geq 0$

~~Hal.~~ $k \leq M$

$N-r-k \geq 0$

$k \geq r-N$

$M+N-r \geq 0$

$r \leq M+N$

Ex: Find Mean, Variance of each of above random variables.
For Multinomial r.v find mean vector, Dispersion Matrix (Variance Co-variance mat.)

$$\underset{\sim}{\mu} = \begin{pmatrix} E(X_i) \\ E(X_k) \end{pmatrix}, \quad \sigma_{ij} = Cov(X_i, X_j)$$

$$\Sigma = [\ ]_{k \times k}$$

Multivariate Normal Random Vector (MVN) : (N_F)

$N_F(\mu, \Sigma) \rightarrow$ Dispersion Matrix, Vari. Cov. Matrix

μ: $p \times 1$, Σ: $p \times p$

If $\underset{\sim}{X}_{p \times 1}$ is a random vector with Joint ~~property~~ prob. density fn given by: $f(\underset{\sim}{x}) = f(x_1, \ldots x_p) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(\frac{(x-\mu)^t \Sigma^{-1} (x-\mu)}{-2}\right)$

$$= \left\{ 2\pi^{-p/2} |\Sigma|^{-1/2} e^{-\frac{(x-\mu)^t \Sigma^{-1} (x-\mu)}{2}} \right\}$$

$$N(\mu, a^2) \Rightarrow f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad x \in \mathbb{R}$$

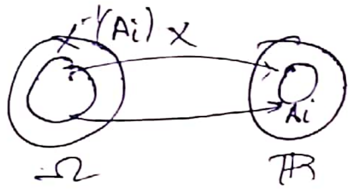
Characterisation: $\underset{\sim}{X}_{p \times 1}$ is MVN $\Leftrightarrow$ every linear combination of $\underset{\sim}{x}$ say $l^t \cdot \underset{\sim}{X} = \sum_{i=1}^{p} l_i x_i$ is univariate Normal.

$$P\left(\bigcap_i A_i\right) = \prod_{i=1} (P(A_i));$$

Scanned by CamScanner

A, B are independent,
if $P(A \cap B) = P(A) \cdot P(B)$
$\Rightarrow P(A/B) = P(A)$
$\Rightarrow P(B/A) = P(B)$

$X : \Omega \to \mathbb{R}$



$X : \Omega \to \mathbb{R}$

Independence of Random Variables:

$X_1, \ldots, X_p$ are independent if $\cdot\ P\left[\bigcap_{i=1}^{p} (X_i \in A_i)\right]$, $A_i \subseteq \mathbb{R}$
$= \prod_{i=1}^{p} P(X_i \in A_i)\ \forall A_i \subseteq \mathbb{R}$

$(X_i \in A_i) = \{\omega \in \Omega \mid X_i(\omega) \in A_i\}$
$= X^{-1}(A_i)$

Eg: X_i: $1 \to H$, $0 \to T$

$\forall a_i = 0,1 \mid P\left(\bigcap_{i=1}^{p} (x_i = a_i)\right) = \prod_{i=1}^{p} P(x_i = a_i)$ [Null event not considered $\because P(\phi) = 0$]

$$\boxed{P(X=a) = \int_a^a f(x)\,dx}$$

Defn Two r.v.s X & Y are uncorrelated (iff $Cov(X,Y) = 0$) (if $\mu_{X,Y} = 0$ iff)
(if $V(X) = Cov(X,X) = 0$)

$var(X) = 0$ iff $X = k$, ~~w.p.1~~ with probability 1.

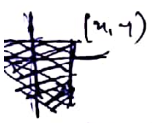
Lemma: 1 If X and Y are independent then $Cov(X,Y) = 0$, i.e X & Y must be uncorrelated. $(E(X)^{2} + V(E(x)))^2$

X, Y are continous

$$Cov(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f_{X,Y}(x,y)\,dx\,dy \to (1)$$

X, Y are independent
$P((X \in A) \cap (Y \in B)) = P(X \in A)\, P(Y \in B),\ \forall A, B \subseteq \mathbb{R}$

$A = (-\infty, x],\ B = (-\infty, y]$



$P(X \in (-\infty, x], Y \in (-\infty, y]) = P(X \in (-\infty, x]) \cdot P(Y \in (-\infty, y])$

$P((X,Y) \in (-\infty, x] \times (-\infty, y]) = \frac{d}{dx} \quad \frac{d}{dy}$

$\Rightarrow \frac{\partial^2}{\partial x \partial y} \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)\,ds\,dt = \frac{\partial^2}{\partial x \partial y} \int_{-\infty}^{x} f_X(s)\,ds \int_{-\infty}^{y} f_Y(t)\,dt$

$\Rightarrow f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$

For descrete case: $P_{X,Y}(s,t) = P_X(s) P_Y(t)$

From (1)

$$Cov(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y) f_X(x) f_Y(y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty}(x-\mu_x) f_x(x)\, dx \int_{-\infty}^{\infty}(y-\mu_y) f_Y(y)\, dy$$

$$\underbrace{\int_{-\infty}^{\infty} x f_x(x)\, dx}_{\mu_x} - \mu_x \underbrace{\int_{-\infty}^{\infty} f_x(x)\, dx}_{1}$$

$$= 0$$

$\therefore Cov(X,Y) = 0$

(*) $X \sim N(0,1)$, $E(X) = 0$

$Y = X^2$, $E(X^2) = (H/W) = 1$

$\int_{-\infty}^{\infty} x f(x)\, dx$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$$Cov(X,Y) = E(X,Y) - E(X)E(Y)$$
$$= \underset{0}{E(X^2)} - \underset{0}{E(X)}\underset{1}{E(Y)}$$
$$= 0$$

Uncorrelated but not independent.

Lemma 2: If X, Y are uncorrelated and both X & Y are normal $\Rightarrow$ X and Y are independent.

$\Sigma = \begin{pmatrix} \ddots \end{pmatrix}$ diagonal

then x, y are uncorrelated

Proof: Let $\underset{\sim}{X}_{p\times 1}$ $(X_1, \ldots X_p)$ be MVN with Σ, a diagonal matrix (with +ve entries $\because$ they are variances)

Then $X_1 \ldots X_p$ must be independent.

$$P\left(\bigcap_{i=1}^{p} (X_i \in A_i)\right)$$
$$= P(X_1 \in A_1 \cap \ldots \cap X_p \in A_p) = \prod_{i=1}^{p} P(X_i \in A_i)$$

$$\int_{A_1}\int_{A_2}\cdots\int_{A_p} f_{\underset{\sim}{X}}(\underset{\sim}{\mu}, \Sigma)(\underset{\sim}{x})\, d\underset{\sim}{x} = \int_{A_1}\cdots\int_{A_p} (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp\left(\frac{(\underset{\sim}{x}-\mu)^t \Sigma^{-1}(\underset{\sim}{x}-\mu)}{-2}\right) d\underset{\sim}{x} \quad (1)$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & & \\ & \ddots & \\ & & \sigma_{pp} \end{bmatrix} \quad \Sigma^{-1} = \begin{bmatrix} \sigma^{11} & & \\ & \ddots & \\ & & \sigma^{pp} \end{bmatrix}$$

$$-\frac{1}{2}(x-\mu)^t(\Sigma)^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) = -\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}(x_i-\mu_i)\sigma^{ij}(x_j-\mu_j)$$

$$= -\frac{1}{2}\sum_{i=1}^{p}(x_i-\mu_i)^2\sigma^{ii}$$

$$= -\frac{1}{2}\sum_{i=1}^{p}\frac{(x_i-\mu_i)^2}{\sigma_{ii}}$$

From ②:

$$= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{11}\cdots\sigma_{pp}}}e^{-\frac{1}{2}\sum\frac{(x_i-\mu_i)^2}{\sigma_{ii}}}$$

$$= \prod_{i=1}^{p}\int_{A_i}\frac{1}{\sqrt{2\pi}\sqrt{\sigma_{ii}}}e^{-\frac{(x_i-\mu_i)^2}{2\sigma_{ii}}}dx_i = \prod_{i=1}^{p}P(X_i\in A_i)$$

MVN

① Conditional Distribution

② Max Likelihood Estimation.

$$(x_1,\dots,x_n)^t = \underset{\sim 1\times p}{X} \sim MVN(\underset{\sim}{\mu}, \Sigma_{p\times p})$$

if the jt p.d.f of $\underset{\sim}{X}$ is given,

by ⓚ $\exp -\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^t\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})$

① Conditional Distribution

Let $\underset{\sim}{X}\begin{pmatrix}\underset{\sim}{X}_{1\,q\times 1}\\ \underset{\sim}{X}_{2\,(p-q)\times 1}\end{pmatrix}$ be a Random Vector

1. If $\underset{\sim}{X}$ is DISCRETE, then the conditional Jt. pmf (prob. mass fn) of $\underset{\sim}{X}_1$ given $\underset{\sim}{X}_2 = \underset{\sim}{x}_2$ is defined as:

$$f(\underset{\sim}{X}_1=\underset{\sim}{x}_1 \mid \underset{\sim}{X}_2=\underset{\sim}{x}_2) = P(\underset{\sim}{X}_1=\underset{\sim}{x}_1 \mid \underset{\sim}{X}_2=\underset{\sim}{x}_2)$$

$$= \frac{P(\underset{\sim}{X}_1=\underset{\sim}{x}_1, \underset{\sim}{X}_2=\underset{\sim}{x}_2)}{P(\underset{\sim}{X}_2=\underset{\sim}{x}_2)} = \frac{f(\underset{\sim}{x}_1,\underset{\sim}{x}_2)}{f_{X_2}(\underset{\sim}{x}_2)}$$

$f(\underset{\sim}{x}_1,\underset{\sim}{x}_2)$ → value of the jt. pmf of $\underset{\sim}{X}$ at $\begin{pmatrix}\underset{\sim}{x}_1\\ \underset{\sim}{x}_2\end{pmatrix}$

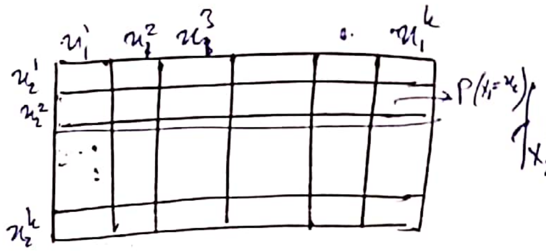
$f_{X_2}(\underset{\sim}{x}_2)$ → value of the marginal pmf of $\underset{\sim}{X}_2$ at $\underset{\sim}{x}_2$

Marginal pmf of X_2, $f_{X_2}(\cdot)$ is defined as

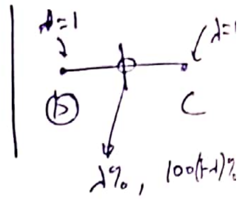
$$f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2) = \sum_{\underset{\sim}{x}_1} f(\underset{\sim}{x}_1, \underset{\sim}{x}_2)$$

$$\boxed{\underset{\sim}{X} = \lambda \underset{\sim}{X}^{P} + (1-\lambda)\underset{\sim}{X}^{C}}_{0<\lambda<1}$$

incl. both discrete and cont. state together.



$$f_{X_2}(x_2^1) = \sum_i p_{ij}$$



$$\underset{\sim}{X} \sim MVN \qquad \underset{\sim}{X}_p = \begin{pmatrix} \underset{\sim}{X}_1 & q \\ \underset{\sim}{X}_2 & p-q \end{pmatrix} \sim MVN\left[\begin{pmatrix} \underset{\sim}{\mu}_1 \to q \\ \underset{\sim}{\mu}_2 \end{pmatrix}_{p-q}, \ \Sigma = \begin{pmatrix} \Sigma_{11}^{q\times q} & \Sigma_{12}^{q\times(p-q)} \\ \Sigma_{12}^{t} & \Sigma_{22} \end{pmatrix}_{(p\times p)}\right]$$

$(p-q)\times q \qquad (q-p)\times(q-p)$

Exercise: Let $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix}$ be MVN as above. P.T the conditional distribution of $\underset{\sim}{X}_2$ given $\underset{\sim}{X}_1 = \underset{\sim}{x}_1$ is again MVN with

$$E[\underset{\sim}{X}_2 \mid \underset{\sim}{X}_1 = \underset{\sim}{x}_1] = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\underset{\sim}{x}_1 - \underset{\sim}{\mu}_1)$$

$$\& \ Cov(\underset{\sim}{X}_2 \mid \underset{\sim}{X}_1 = \underset{\sim}{x}_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

$\Sigma_{ij} = Cov(X_i, X_j)$

$\Sigma_{11} = Var(\underset{\sim}{X}_1)$

$\Sigma_{22} = Var(\underset{\sim}{X}_2)$

$\Sigma_{12} = Var(\underset{\sim}{X}_1, \underset{\sim}{X}_2)_{q\times(p-q)}$

$\Sigma_{21} = Var(\underset{\sim}{X}_2, \underset{\sim}{X}_1)_{(p-q)\times q}$

$\boxed{\text{P.f}}$ $\quad T: (\underset{\sim}{X}_1, \underset{\sim}{X}_2) \longrightarrow (X_1', X_2')$

$$T\begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix} = \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}\begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix} = \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{X}_1 \end{pmatrix} \sim MVN \ (*)$$

(*) Since linear vector obtained

$$E(X_1') = E(\underset{\sim}{X}_1) = \mu_1$$

$$E(X_2') = E(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1) = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1$$

$$Cov(X_1', X_1') = Cov(X_1, X_1) = \Sigma_{11}$$

$$Cov(X_1', X_2') = Cov(X_1, X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1)$$

[Dimension is erroneous (no correction)]

$$= Cov(X_1, X_2) - \Sigma_{21}\Sigma_{11}^{-1}Cov(X_1, X_1)$$

$$= \underset{q\times(p-q)}{\Sigma_{12}} - \left[\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{11}\right]^{t} \to \Sigma_{12} = 0_{q\times(p-q)}$$

$q\times(p-q)$

X_1^t & X_2' are uncorrelated, M.VN R.V. Hence X_1' & X_2' are independent.

Since, X_1 & X_2' are independent, $E(X_2' \mid \underset{\sim}{X}_1 = \underset{\sim}{x}_1) \overset{\text{Ind.}}{=} E(X_2')$

$$= \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1$$

$$Cov\left((X_2' \mid \underset{\sim}{X}_1 = \underset{\sim}{x}_1) \right) = Cov(X_2')$$

$$= Cov(X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1, X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1).$$

$$= Cov(X_2, X_2) - Cov(\Sigma_{21}\Sigma_{11}^{-1}X_1, X_2) - Cov(X_2, \Sigma_{21}\Sigma_{11}^{-1}X_1)$$

$$+ Cov(\Sigma_{21}\Sigma_{11}^{-1}X_1, \Sigma_{21}\Sigma_{11}^{-1}X_1)$$

$$= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} - \cancel{\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}} + \Sigma_{21}\Sigma_{11}^{-1}\cancel{\Sigma_{11}\Sigma_{11}^{-1}\Sigma_{12}}$$

$$= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

~~$Cov(x_1, x_2) = Cov(Ax_1, Bx_2)$~~

Maximum Likelihood Estimation:

Given $\underset{\sim}{x}_1, \underset{\sim}{x}_2 \dots \underset{\sim}{x}_n \in \mathbb{R}^p \to$ ①, $n \in \mathbb{N}$

To find out which μ, Σ (from $M.VN(\underset{\sim}{\mu}, \Sigma)_p$) matches the ~~of~~ given ①.

$$f(\underset{\sim}{x}_i) = k\exp\left[-\tfrac{1}{2}(x_i - \mu)^t \Sigma^{-1} (\underset{\sim}{x}_i - \mu)\right]$$

$$f(\underbrace{\underset{\sim}{x}_1, \dots \underset{\sim}{x}_n}_{\left(\substack{p\times n \\ \text{matrix}}\right)} \mid \underset{\sim}{\mu}, \underset{\sim}{\Sigma}) = \prod_{i=1} f(\underset{\sim}{x}_i \mid \mu, \Sigma), \quad \boxed{\mu, \Sigma \text{ are unknown}}$$

$$L(\mu, \Sigma \mid \underset{\sim}{x}_1, \underset{\sim}{x}_2 \dots \underset{\sim}{x}_n)$$

$$= \prod_{i=1}^{n} f(\underset{\sim}{x}_i \mid \mu, \Sigma)$$

choose $\hat{\mu} \in \mathbb{R}^p$ & $\hat{\Sigma} \in M^t_{p\times p}$

$\ni L(\hat{\mu}, \hat{\Sigma} \mid \underset{\sim}{x}_1, \dots \underset{\sim}{x}_n)$ is Max $L(\quad)$

$\underset{\sim p\times 1}{\mu} \in \mathbb{R}^p$ +ve definite

$\Sigma_{p\times p} \in M_{p\times p}$

↳ invertible, each eigen value is +ve.

A Matrix is Positive definite: $A_{p\times p}$ if $x^t A x > 0 \ \forall \ \underset{p\times 1}{x} \neq 0$.

i.e $\sum_i \sum_j x_i a_{ij} x_j > 0 \ \forall \ x_i \in \mathbb{R}$ not all 0

$L(\max) = \log_e L(\max)$

$$\log L(\mu, \Sigma \mid \underset{\sim}{x}_1, \underset{\sim}{x}_2 \cdots \underset{\sim}{x}_n) = \sum_{i=1}^{n} \log f(\underset{\sim}{x}_i \mid \mu, \Sigma)$$

$$L = \prod_{r=1}^{n} f(x_r \mid \mu, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2}\sum_{r=1}^{n} (x_r-\mu)^t \Sigma^{-1} (x_r-\mu)\right\}$$

Maximising L is equivalent to Minimizing the -ve of log likelihood.

$$-2\ln L = k + n\log|\Sigma| + \sum_{r=1}^{n} (x_r-\mu)^t \Sigma^{-1}(x_r-\mu), \quad k = \text{const independent of } \mu, \Sigma.$$

<u>Result</u>: $l(\mu, \Sigma) = n\{\log|\Sigma| + tr[\Sigma^{-1}(S + dd^t)]\} + c$ const.

$(d = \bar{x} - \mu)$

In ⊛ the dependence of l on μ is totally through d.

Thus $\forall\, d \neq 0$, $d^t \Sigma^{-1} d > 0$ (as Σ & Σ^{-1} are assumed to be +ve definite)

$\boxed{\text{MLE of } \mu \text{ is } \hat{\mu} = \bar{x}}$

min 0

x is a r.v $= 0$ when $d = 0$

then $var\, X = 0$ i.e when $\bar{x} = \mu$,

$\hat{\mu} = \bar{x}$

iff x is a constant.

i.e iff $x = k$, $\omega \neq 1$

<u>Lemma 1</u>

$\Sigma^{-1} S$ is +ve semi-definite [A is +ve semi-definite if, $x^t A x \geq 0 \;\forall x$]

sample dispersion matrix

The characteristic values (eigen) of $\Sigma^{-1} S$ are +ve.

$$\left[\begin{array}{l} E[(X-\mu)(X-\mu)^t] = \Sigma \\ \frac{1}{(n-1)} (X-\bar{X})(X-\bar{X})^t = S \end{array}\right.$$

<u>Lemma 2</u>: For any set of +ve nos.

Let $y_1, y_2 \ldots$ be all +ve $\{y_1 \ldots y_n\}$

$e^x \geq 1 + x \;\forall x$ to prove.

$y_i \geq 1 + \log y_i$

$\Rightarrow \sum y_i \geq n + \sum_1^n \log y_i$

$A \geq \log G + 1$ — ①

$A = AM$, $G = GM$

$\Rightarrow \frac{\sum y_i}{n} \geq 1 + \frac{1}{n} \sum \ln y_i$ ..

$\Rightarrow A \geq 1 + \ln(\prod y_i)^{1/n}$ So

Recall that $trA = \sum_{1}^{n} a_{ii} = \sum_{1}^{n} \lambda_i$ λ_i is an eigen value of A.

& $\det A = \prod_{i=1}^{n} \lambda_i = |A|$

from ①: $\frac{1}{n} \sum_{1}^{n} a_{ii} \geq 1 + \frac{1}{n} \log \cancel{|A|} \prod_{i=1}^{n} \lambda_i$

$\Rightarrow \frac{1}{n} trA \geq 1 + \frac{1}{n} \log |A|$

Let $\lambda_1, \ldots \lambda_n$ be +ve ch. values of $\Sigma^{-1} S$,

Then the above $\Rightarrow$

We have taken Σ & Σ^{-1} are both +ve definite:

$\Sigma = VDV^t$, where D is a Diagonal Matrix

& $D = \text{Diag}[\lambda_1', \ldots \lambda_n']$ $\boxed{B = P^t D P \text{ similar}}$

$\lambda_i' \to$ ch. value of Σ

$\Sigma^{-1} = VD^{-1}V^t$

$(\because \hat{\mu} = \bar{x}, 1^{st}$ vanishes)

$l(\hat{\mu}, \Sigma) = n\{\log|\Sigma| + tr(\Sigma^{-1}S)\} = \Phi(\Sigma)$

Claim, $\Phi(\Sigma) - \Phi(S) = n\{\log|\Sigma| + tr(\Sigma^{-1}S) - \log|S| - \cancel{tr(I)}_p\}$

$\cancel{= n\{\log|\Sigma|}$ $= n\{tr(\Sigma^{-1}S) - \log|\Sigma^{-1}S| - p\} \geq 0$ (Claim)

Let $e_1, e_2 \ldots e_p$ be the ch. values of $(\Sigma^{-1}S)_{p\times p}$

$\log|\Sigma^{-1}S| = \log|\prod_{1}^{p} e_i| = p \log G$.

$tr|\Sigma^{-1}S| = \sum_{1}^{p} e_i = pA$

$\frac{1}{p} tr(\Sigma^{-1}S) \geq 1 + \frac{1}{p} \log|\Sigma^{-1}S|$

$tr(\Sigma^{-1}S) \geq p + \log|\Sigma^{-1}S|$

$\left.\begin{array}{l}\hat{\mu} = \bar{x} \\ \hat{\Sigma} = S\end{array}\right\}$ MLE of μ, Σ

if $\underset{\sim}{X} \sim N_p(\mu, \Sigma)$