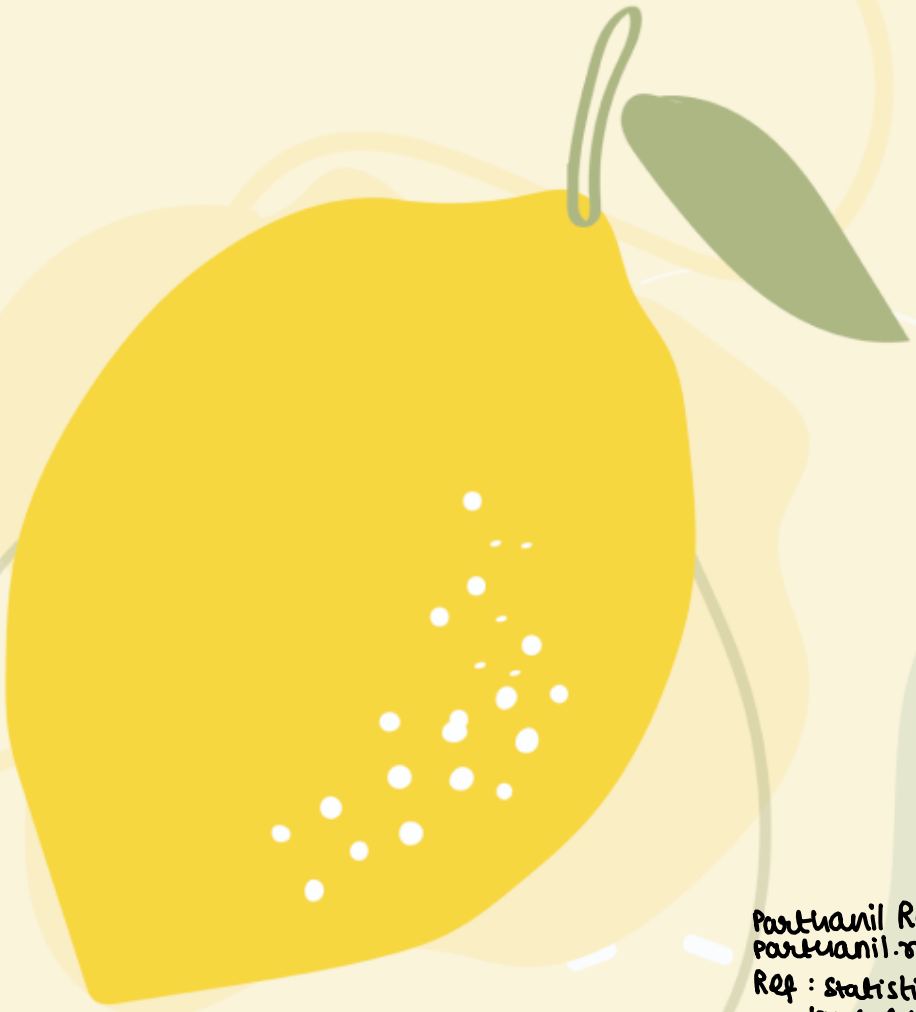
# SI 424
## Statistical Inference I



Parthanil Roy 211-A
parthanil.roy@gmail.com

Ref : Statistical Inference by G. Casella and R.L. Berger

Class notes

Quizes - 30%. 4 or more
Midsem - 30%. worst dropped
Endsem - 40%.

Exercises and quiz questions in exams

Close notes exam

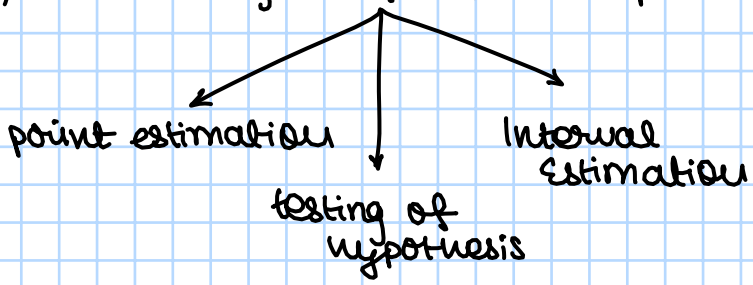
6th Jan:

Basic distributions should be known

Defn: (Statistical inference) Theory and methods using data analysis to infer about the properties of underlying probability distribution. We shall assume that underlying probability distribution has one or more parameters which are unknown but fixed. This leads to parametric inference.

Eg: for normal $\mu$, $\sigma$ one or both can be unknown but fixed, we want to estimate them.

we shall have data/observations from this probability distribution, based on this data, we shall try to infer about parameter(s)




Note: suppose $f(x) = f_\theta(x) = f(x|\theta)$ is either a probability mass function (pmf) or a probability density function (pdf) and $\theta$ (possibly a vector) is an unknown but fixed parameter

Eg: ① $f(x|\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$ if $x \in \{0,1,\dots\}$ pmf of poisson $\lambda$ distribution, here (exponential is inverse)

$\lambda \in (0,\infty)$ ← parameter space

↑ parameter

② $f(x|\alpha,\lambda) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1}; & x>0 \text{ (sum of exponential} \\ 0 ; & \text{otherwise} \end{cases}$

pdf of gamma$(\alpha,\lambda)$ ← scale parameter as $X \sim \text{gamma}(\alpha,\lambda)$, $\lambda X \sim \text{gamma}(\alpha,1)$

↑ shape parameter

here $(\alpha,\lambda) \in (0,\infty)\times(0,\infty)$
parameters parameter space

Defn: $X_1, X_2, \dots X_n$ is called a random sample of size n from $f(x|\theta)$ if

$X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ ($f$ is pmf/pdf)

Therefore, the joint pmf / joined pdf of $X_1, X_2, \dots X_n$ is

$$L(x_1,x_2,\dots x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = L_\theta(x_1,x_2,\dots x_n)$$

Defn: (Likelihood function) $L(x_1,x_2,\dots x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

Distributions of functions of random variables:

Eg: Suppose $X_1 \sim \text{Poi}(\lambda_1)$, $X_2 \sim \text{Poi}(\lambda_2) \dots X_n \sim \text{Poi}(\lambda_n)$

we want to find the distribution of $\sum_{i=1}^{n} X_i$

if $X_1, X_2, \dots X_n$ are independent, then

for $n=2$:

$X_1 \sim Poi(\lambda_1)$

$X_2 \sim Poi(\lambda_2)$ and $X_1, X_2$ independent

$X_1 \sim Poi(\lambda_1) \Rightarrow Range(X_1) = \mathbb{N} \cup \{0\}$

$X_2 \sim Poi(\lambda_2) \Rightarrow Range(X_2) = \mathbb{N} \cup \{0\}$

$Range(X_1 + X_2) \subseteq \mathbb{N} \cup \{0\}$

take $k \in \mathbb{N} \cup \{0\}$, then $P(X_1 + X_2 = k) = P\left[\bigcup_{l=0}^{k} (X_1 = l, X_2 = k-l)\right]$

$= \sum_{l=0}^{k} P(X_1 = l, X_2 = k-l)$ ($\because$ pairwise disjoint)

$= \sum_{l=0}^{k} P(X_1 = l) P(X_2 = k-l)$

$= \sum_{l=0}^{k} e^{-\lambda_1} \frac{(\lambda_1)^l}{l!} e^{-\lambda_2} \frac{(\lambda_2)^{k-l}}{(k-l)!}$

$= e^{-(\lambda_1+\lambda_2)} \sum_{l=0}^{k} \frac{1}{l!(k-l)!} (\lambda_1)^l (\lambda_2)^{k-l}$

$= e^{-(\lambda_1+\lambda_2)} \sum_{l=0}^{k} \frac{k!}{l!(k-l)!} \frac{(\lambda_1)^l (\lambda_2)^{k-l}}{k!}$

$= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^k}{k!}$ for $k = 0, 1, \dots$

$\Rightarrow X_1 + X_2 \sim Poi(\lambda_1 + \lambda_2)$

Exe: using induction on $n$, show that $\sum_{i=1}^{n} X_i \sim Poi\left(\sum_{i=1}^{n} \lambda_i\right)$ → done down

Consequence: If $X_1, X_2, \dots X_n$ is a random sample (of size $n$) from poisson distribution with parameter $\lambda$ (i.e $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\lambda)$) then

$$\sum_{i=1}^{n} X_i \sim Poi(n\lambda)$$

Theorm: If $X$ and $Y$ are two independent discrete random variables (r.v.s) with probability mass functions with $f$ and $g$ respectively then, $Z := X + Y$ is also a discrete random variable and has pmf

$$h(z) = P[Z = z] = \sum_{x \in Range(x)} f(x) g(z-x) \quad z \in \mathbb{R}$$

Note: ① $h(z) = P[Z=z] = \sum_{x \in Range(x)} f(x) g(z-x)$ is called discrete convolution of $f$ and $g$

② we can also write it like

$$h(z) = \sum_{y \in Range(Y)} f(z-y) g(y)$$

**Note:** $h(z) = \sum_{\substack{x \in \text{Range}(X) \\ z-x \in \text{Range}(Y)}} f(x) g(z-x)$

**Exe:** Suppose $X_1, X_2 \ldots X_n$ are independent r.v.s and $X_i \sim \text{Bin}(m_i, p)$ $\forall i = 1, 2 \ldots n$ show that $\sum_{i=1}^{n} X_i \sim \text{Bin}(\sum_{i=1}^{n} m_i, p)$. In particular if $X_1, X_2, \ldots X_n$ is a random sample from $\text{Bin}(m, p)$ distribution, then $\sum_{i=1}^{n} X_i \sim \text{Bin}(mn, p)$

→ done down

**Exe:** Suppose $X_1, X_2, \ldots X_n$ is a random sample from $\text{poi}(\lambda)$ dist$^n$. Fix $s_0 \in \mathbb{N} \cup \{0\}$ define $S = \sum_{i=1}^{n} X_i$

(a) Find conditional pmf of $X_1$ given $S = s_0$

(b) compute conditional expectation $\mathbb{E}[X_1 | S = s_0]$ → done down

**Theorm:** (convolution formula) if $X$ and $Y$ are independent continious r.v.s with pdfs $f$ and $g$ respectively then $Z := X + Y$ is also a continious r.v with pdf

$$h(z) = \int_{\mathbb{R}} f(x) g(z-x)\, dx = \int_{\mathbb{R}} f(z-y) g(y)\, dy \quad z \in \mathbb{R}$$

$h(z)$ is called convolution of $f$ and $g$

**proof:** It will be skipped

**Note:** If above $X, Y$ are not ind then $Z$ can/cannot be continious, it can be forced to become zero, $X \sim N(0,1)$, $Y = -X$ then $X + Y = 0$

**Exe:** Suppose

ind — $X \sim \text{gamma}(\alpha, \lambda)$, $Y \sim \text{gamma}(\beta, \lambda)$

we want to find $Z := X + Y$

$\text{Range}(X) = \text{Range}(Y) = (0, \infty) \Rightarrow \text{Range}(Z) \subseteq (0, \infty)$

clearly $Z$ is cont r.v, let $h$ be pdf of $Z$ then $h(z) = 0$ if $z \notin (0, \infty)$

Let $z \in (0, \infty)$, then by convolution formula

$$h(z) = \int_{\mathbb{R}} f(x) g(z-x)\, dx \quad \text{where} \quad f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} \text{ if } x > 0$$

$$g(y) = \frac{\lambda^\beta}{\Gamma(\beta)} e^{-\lambda y} y^{\beta-1} \text{ if } y > 0$$

integral from $0$ to $z$ as $f(x) > 0$ if $x > 0$ and $g(z-x) > 0$ if $z - x > 0$

$\Rightarrow 0 < x < z$

so, $h(z) = \int_0^z f(x) g(z-x)\, dx$

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

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

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda\xi} \int_0^{1} (\xi u)^{\alpha-1}(\xi-\xi u)^{\beta-1}\xi\,du \quad \left(u=\frac{x}{\xi} \Rightarrow du=\frac{dx}{\xi};\ x=0,\ u=0;\ x=\xi,\ u=1\right)$$

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

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

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

$$= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} e^{-\lambda\xi}(\xi)^{\alpha+\beta-1} \quad \left(\because B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)$$

$\forall\, \xi \in (0,\infty)$

$\Rightarrow Z := X+Y \sim \text{gamma}(\alpha+\beta, \lambda)$

Exe: using induction on $n$, show that $\sum_{i=1}^{n} X_i \sim Poi\left(\sum_{i=1}^{n} \lambda_i\right)$

Ans: $n=2$: we have already seen $X_1 + X_2 \sim Poi(\lambda_1+\lambda_2)$

let's assume true for $n=k-1$, true for

$n=k$: $\sum_{i=1}^{k} X_i = \left(\sum_{i=1}^{k-1} X_i\right) + X_k$

as $\sum_{i=1}^{k-1} X_i \sim Poi\left(\sum_{i=1}^{k-1} \lambda_i\right)$ — ①

$X_k \sim Poi(\lambda_k)$ — ②

from $n=2$ case $\left(\sum_{i=1}^{k-1} X_i\right) + X_k \sim Poi\left(\sum_{i=1}^{k-1} \lambda_i + \lambda_k\right)$

so, by induction true for all $n$

Exe: If $X$ and $Y$ are two independent discrete random variables (r.v.s) with probability mass functions with $f$ and $g$ respectively then, $Z := X+Y$ is also a discrete random variable and has pmf

$$h(z) = P[Z=z] = \sum_{x \in Range(x)} f(x)\, g(z-x) \quad z \in \mathbb{R}$$

Ans:

$h(z) = P[Z=z]$ $\quad Z := X+Y$

true as $X, Y$ independent

$$P[Z=\xi] = P[X+Y=\xi]$$

$$= P\left[\bigcup_{l \in Range(X)} (X=l \cap Y=\xi-l)\right]$$

$$= \sum_{l \in Range(X)} P[X=l \cap Y=\xi-l]$$

$$= \sum_{l \in \text{Range}(X)} f(l)\, g(\xi - l) \quad (\because X, Y \text{ independent})$$

**Exe:** Suppose $X_1, X_2 \ldots X_n$ are independent r.v.s and $X_i \sim \text{Bin}(m_i, p)$ $\forall i = 1, 2 \ldots n$ show that $\sum_{i=1}^{n} X_i \sim \text{Bin}(\sum_{i=1}^{n} m_i, p)$. In particular if $X_1, X_2, \ldots X_n$ is a random sample from $\text{Bin}(m, p)$ distribution, then $\sum_{i=1}^{n} X_i \sim \text{Bin}(mn, p)$

**Ans:**

$$P(X = k) = \binom{n}{k} (p)^k (1-p)^{n-k} \quad \text{when } X \sim \text{Bin}(n, p)$$

now, let $X_1 \sim \text{Bin}(m_1, p)$ $X_2 \sim \text{Bin}(m_2, p)$

then $P(X_1 + X_2 = k) = P(\bigcup_{l \in \text{Range}(X_1)} X_1 = l, X_2 = k - l)$ $(\because \text{theorm})$

$$= \sum_{l \in \text{Range}(X_1)} P(X_1 = l, X_2 = k - l)$$

$$= \sum_{l \in \text{Range}(X_1)} \binom{n_1}{l} (p)^l (1-p)^{n_1 - l} \times \binom{n_2}{k-l} (p)^{k-l} (1-p)^{n_2 - k + l}$$

$(\because X_1, X_2 \text{ are independent})$

for $0 \le l \le n_1$
$0 \le k - l \le n_2$
$-n_2 \le l - k \le 0$
$\Rightarrow 0 \le l \le n_1$
$k - n_2 \le l \le k$

$$P(X_1 + X_2 = k) = \sum_{\substack{0 \le l \le n_1 \\ k - n_2 \le l \le k}} \binom{n_1}{l} \binom{n_2}{k-l} p^k (1-p)^{n_1 + n_2 - k}$$

$\binom{n_1}{l}\binom{n_2}{k-l}$: group $\binom{n_1+n_2}{k}$ any $l$ from $n_1$ group and any $k-l$ from $n_2$ group

so, combinatorially

$$\sum_{\substack{0 \le l \le n_1 \\ k - n_2 \le l \le k}} \binom{n_1}{l} \binom{n_2}{k-l} = \binom{n_1 + n_2}{k}$$

splitting a group and selecting $l$, $k-l$ or selecting $k$ together

$$\Rightarrow P(X_1 + X_2 = k) = \binom{n_1 + n_2}{k} (p)^k (1-p)^{n_1 + n_2 - k}$$

$$\Rightarrow X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$$

So for $n=2$ case: $X_1 + X_2 \sim Bin(n_1+n_2, p)$

then let's assume true for $n=k-1$ case i.e

$$\sum_{i=1}^{k-1} X_i \sim Bin\left(\sum_{i=1}^{k-1} n_i, p\right) \quad \text{for } X_i \sim Bin(n_i, p)$$

for $n=k$ case:

$$\sum_{i=1}^{k} X_i = \left(\sum_{i=1}^{k-1} X_i\right) + (X_n)$$

then from $n=k-1$ case:

$$\sum_{i=1}^{k-1} X_i \sim Bin\left(\sum_{i=1}^{k-1} n_i, p\right)$$

and from $n=2$ case:

$$\left(\sum_{i=1}^{n-1} X_i\right) + (X_n) \sim Bin\left(\sum_{i=1}^{k-1} n_i + n_k, p\right) = Bin\left(\sum_{i=1}^{k} n_i, p\right)$$

So by induction, true for all $n$

Ex: suppose $X_1, X_2, \ldots X_n$ is a random sample from $Poi(\lambda)$ dist$^n$. Fix $s_0 \in \mathbb{N} \cup \{0\}$
define $S = \sum_{i=1}^{n} X_i$

(a) Find conditional pmf of $X_1$ given $S = s_0$

(b) compute conditional expectation $\mathbb{E}[X_1 \mid S = s_0]$

Ans:

$$S = \sum_{i=1}^{n} X_i \qquad X_1, \ldots X_n \overset{iid}{\sim} Poi(\lambda)$$

$$P(X_1 = k \mid S = s_0) = \frac{P(X_1 = k \text{ and } S = s_0)}{P(S = s_0)}$$

as $S \sim Poi(n\lambda)$ (∵ from above)

and $P(X_1 = k \text{ and } \sum_{i=1}^{n} X_i = s_0)$

$= P(X_1 = k \text{ and } \sum_{i=2}^{n} X_i = s_0 - k)$

$= P(X_1 = k)\, P(\sum_{i=2}^{n} X_i = s_0 - k)$ (∵ independent)

and similarly $\sum_{i=2}^{n} X_i \sim Poi((n-1)\lambda)$

So, $P(X_1 = k \mid S = s_0) = \dfrac{P(X_1 = k)\, P(\sum_{i=2}^{n} X_i = s_0 - k)}{P(\sum_{i=1}^{n} X_i = s_0)}$

$$= \left(\frac{\lambda^{K} e^{-\lambda}}{K!_0}\right)\left(\frac{(n\lambda-\lambda)^{S_0-K} e^{-n\lambda+\lambda}}{(S_0-K)!_0}\right)$$

$$\frac{(n\lambda)^{S_0} e^{-n\lambda}}{(S_0)!_0}$$

$$P(X_1=K \mid S=S_0) = \frac{(S_0)!_0}{K!_0\,(S_0-K)!_0}\,\frac{(n-1)^{S_0-K}}{(n)^{S_0}}$$

$$= \binom{S_0}{K}\frac{(n-1)^{S_0-K}}{(n)^{S_0}}$$

$$\mathbb{E}[X_1 \mid S=S_0] = \sum_{K=0}^{\infty}(K)\,P(X_1=K \mid S=S_0)$$

$$= \sum_{K=0}^{\infty} K\binom{S_0}{K}\frac{(n-1)^{S_0-K}}{(n)^{S_0}}$$

9th Jan:

we have seen that

$$\text{ind}\begin{cases} X \sim \text{gamma}(\alpha,\lambda) \\ Y \sim \text{gamma}(\beta,\lambda) \end{cases}$$

$$\Rightarrow X+Y \sim \text{gamma}(\alpha+\beta,\lambda)$$

Exe: If $X \sim N(0,1)$, then show that $X^2 \sim \text{Gamma}\left(\frac{1}{2},\frac{1}{2}\right)$ → done down

Note: $\text{Gamma}\left(\frac{1}{2},\frac{1}{2}\right)$ is also called $\chi^2$ distribution with degree of freedom 1

Theorem: Suppose $X_1, X_2, \ldots X_n$ are ind r.v.s such that

$$X_i \sim \text{gamma}(\alpha_i,\lambda) \text{ for } \forall i \in \{1,2\ldots n\}$$

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

proof: using induction on $n=2$ case, trivially we get the result

Cor: If $X_1, X_2, \ldots X_n$ is a random sample from the exponential distribution with rate parameter $\lambda$ (Expection is $\frac{1}{\lambda}$)

i.e $X_1, \ldots X_n \overset{iid}{\sim} \text{Exp}(\lambda)$, then

$$\sum_{i=1}^{n} X_i \sim \text{gamma}(n,\lambda)$$

proof: $\text{Exp}(\lambda) = \text{gamma}(1,\lambda)$ so result follows from prev theorm

Cor: If $Z_1, Z_2, \ldots Z_n \overset{iid}{\sim} N(0,1)$ then $\sum_{i=1}^{n} Z_i^2 \sim \text{gamma}\left(\frac{n}{2},\frac{1}{2}\right)$

proof: $Z_1, \ldots Z_n \overset{iid}{\sim} N(0,1)$

$$\Rightarrow Z_1^2, \ldots Z_n^2 \overset{iid}{\sim} \text{gamma}\left(\frac{1}{2},\frac{1}{2}\right)$$

then from theorm above with $\lambda=\frac{1}{2}$, $\alpha_i=\frac{1}{2}$ $\forall i \in \{1\ldots n\}$

$$\Rightarrow \sum_{i=1}^{n}\alpha_i = \frac{n}{2}$$

so, $\sum_{i=1}^{n} Z_i^2 \sim \text{gamma}\left(\frac{n}{2},\frac{1}{2}\right)$

Note: $\text{gamma}\left(\frac{n}{2},\frac{1}{2}\right)$ is also called the $\chi^2$ distribution with $n$ degrees of freedom

so, $\sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$ if $Z_i \sim N(0,1)$ $\forall i \in \{1,2\ldots n\}$

Exe: Suppose $U,V \overset{iid}{\sim} \text{Unif}(0,1)$, then find the pdf of $Z=U+V$ using convolution formula → done down

Theorm: Suppose $X$ and $Y$ are ind r.v.s with pdfs $f$ and $g$ respectively then $Z := \frac{X}{Y}$ is also a cont r.v with pdf

$$h(\xi) = \int_{-\infty}^{\infty} |y| f(y\xi)\, g(y)\, dy, \quad \xi \in \mathbb{R}$$

further if $\text{Range}(X) \subseteq (0,\infty)$ and $\text{Range}(Y) \subseteq (0,\infty)$, then

$$h(\xi) = \int_{0}^{\infty} y f(y\xi) g(y) dy, \quad \xi \in (0,\infty)$$

proof: It will be skipped

Eg: Suppose $X, Y \overset{iid}{\sim} \text{Exp}(\lambda)$, we want to find pdf of $Z = \frac{X}{Y}$

clearly in this case

$$f(x) = \lambda e^{-\lambda x}, \; x > 0$$

$$g(y) = \lambda e^{-\lambda y}, \; y > 0$$

clearly, $Z$ has

$$h(\xi) = \begin{cases} \int_0^{\infty} y \lambda e^{-\lambda y\xi}\, \lambda e^{-\lambda y}\, dy & ; \xi > 0 \\ 0 & ; \xi \leq 0 \end{cases}$$

therefore for any $\xi \in (0,\infty)$ we have

$$h(\xi) = \int_0^{\infty} y \lambda^2 e^{-\lambda(\xi+1)y} dy$$

$$= \frac{\lambda}{\xi+1} \int_0^{\infty} y \underbrace{\lambda(\xi+1) e^{-\lambda(\xi+1)y}}_{\text{pdf of } Y' \sim \text{Exp}(\lambda(\xi+1))} dy$$

$$= \frac{\lambda}{\xi+1} \underbrace{\mathbb{E}[Y']}_{\mathbb{E}[Y']} = \frac{\lambda}{\xi+1} \times \frac{1}{\lambda(\xi+1)} = \frac{1}{(\xi+1)^2}$$

therefore, pdf of $Z = \frac{X}{Y}$ is $h(\xi) = \frac{1}{(\xi+1)^2}$ if $\xi > 0$

Note: in above example $\lambda$ vanishes out

Exe: $X, Y \overset{iid}{\sim} N(0,1)$ find pdf of $Z := \frac{X}{Y}$ → done down

Theorem: If $X$ and $Y$ are ind cont r.v.s with pdfs $f$ and $g$ respectively then $W := XY$ is also a r.v with pdf

$$\varphi(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x) g\left(\frac{w}{x}\right) dx = \int_{-\infty}^{\infty} \frac{1}{|y|} f\left(\frac{w}{y}\right) g(y) dy$$

$w \in \mathbb{R}$

proof: It will be skipped here

Towards multivariate change of density formula:

$(X_1, X_2, \ldots X_n)$ with joint pdf $\varphi(x_1, x_2, \ldots x_n)$

$\downarrow g$

$(Y_1, Y_2, \ldots Y_n) = g(X_1, X_2, \ldots X_n)$

we want to find out the joint distribution of $(Y_1, Y_2, \ldots Y_n)$

## $n=1$ Univariate:

**Theorem:** (Univariate change of density formula) Suppose $I, J \subseteq \mathbb{R}$ are two open (possibly unbounded) intervals and

$g: I \to J$ is a differentiable bijective function s.t $g' > 0$ on $I$ or $g' < 0$ on $I$

If $X$ is a cont r.v with pdf $\varphi$ that vanishes on $I^c$ ($\text{Range}(X) \subseteq I$)

then

$Y := g(X)$ is also a continious r.v with

pdf

$$\psi(y) = \begin{cases} \varphi(g^{-1}(y)) \left| \frac{d g^{-1}(y)}{dy} \right| & ; \text{if } y \in J \\ 0 & ; \text{if } y \notin J \end{cases}$$

**Note:** ① conditions on $g$ ensure that $g^{-1}$ is differtiable on $J$

② If range$(X) = I$, then Range$(Y) = J$

**Theorem:** (Multivariate change of density Formula) Suppose $I, J \subseteq \mathbb{R}^n$ are both open and connected sets and $g: I \to J$ is a bijective map with

$g^{-1}: J \to I$ of the form

$$g^{-1}(y_1, \dots y_n) = (h_1(y_1 \dots y_n), h_2(y_1 \dots y_n), \dots h_n(y_1 \dots y_n))$$
$$\forall (y_1, y_2 \dots y_n) \in J$$

Assume that all the partial derivatives $\frac{\partial h_i}{\partial y_j}$ $\left(\begin{matrix} i=1,2\dots n \\ j=1,2\dots n \end{matrix}\right)$

Exist and are continious on $J$, and

$\det\left(J_{g^{-1}}(\underset{\sim}{y})\right) \neq 0 \ \forall \ \underset{\sim}{y} = (y_1 \dots y_n) \in J$

where $J_{g^{-1}}(\underset{\sim}{y}) = \left( \left( \frac{\partial h_i}{\partial y_j} \right) \right)_{1 \le i,j \le n}$ is the Jacobian matrix of $g^{-1}$

If $\underset{\sim}{X} = (X_1, X_2 \dots X_n)$ is a continious random vector with joint pdf $\varphi$ that vanishes on $I^c$ (i.e Range$(\underset{\sim}{X}) \subseteq I$)

then $\underset{\sim}{Y} = (Y_1, \dots Y_n) := g(X) = g(X_1, \dots X_n)$ is also a continious random vector with joint pdf density function

$$\psi(\underset{\sim}{y}) = \begin{cases} \varphi(g^{-1}(\underset{\sim}{y})) \left| \det(J_{g^{-1}}(y)) \right| & ; \ \underset{\sim}{y} \in J \\ 0 & ; \ \underset{\sim}{y} \notin J \end{cases}$$

**Note:** ① If range$(\underset{\sim}{X}) = I$, then Range$(\underset{\sim}{Y}) = J$

② $\det(J_{g^{-1}}(\underset{\sim}{y}))$ → Jacobian of $g^{-1}$ (in many references)

③ Figureing out $I$ and $J$ can be very complex

④ It is enough to check that $I$ and $J$ are open and path connected as path-connected

⇓

connected

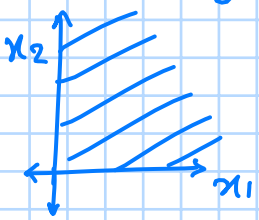
Eg: Suppose $X_1 \perp\!\!\!\perp X_2$ (i.r.v), $X_1 \sim \text{gamma}(\alpha_1, \lambda)$ and $X_2 \sim \text{gamma}(\alpha_2, \lambda)$

let $S := X_1 + X_2$ $\quad Y := \frac{X_1}{S} = \frac{X_1}{X_1 + X_2}$

we want to find joint dist of $Y$ and $S$

$X_1 \perp\!\!\!\perp X_2 \Rightarrow \text{Range}(X_1, X_2) = \text{Range}(X_1) \times \text{Range}(X_2)$

$= (0,\infty) \times (0,\infty)$




clearly, $I$ is open and path connected

define $g : I \to \mathbb{R}^2$ by $g(x_1, x_2) = \left( \frac{x_1}{x_1 + x_2}, x_1 + x_2 \right)$

$= \left( \frac{x_1}{S}, S \right)$ where $S = x_1 + x_2$

$g$ is one-one as if $g(x_1, x_2) = g(y_1, y_2)$

then $\left( \frac{x_1}{S_1}, S_1 \right) = \left( \frac{y_1}{S_2}, S_2 \right) \Rightarrow S_1 = S_2,\ y_1 = x_1,\ x_2 = y_2$

$\Rightarrow (x_1, x_2) = (y_1, y_2)$

now, $g(I) = \left\{ g(x_1, x_2) \;\middle|\; (x_1, x_2) \in I = (0,\infty)^2 \right\}$

$= \left\{ \left( \frac{x_1}{x_1 + x_2}, x_1 + x_2 \right) \;\middle|\; (x_1, x_2) \in I = (0,\infty)^2 \right\} \subseteq (0,1) \times (0,\infty)$

if $(0,1) \times (0,\infty) \subseteq g(I)$ then we will get $(0,1) \times (0,\infty) = J$




Exe: If $X \sim N(0,1)$, then show that $X^2 \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$

Ans: as $X \sim N(0,1)$

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

now, $P(X \le t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$

$X^2 \subset (0,\infty)$ range

then $P(X^2 \le t)$

$= P(X \le \sqrt{t} \text{ and } X \ge -\sqrt{t})$

$= \int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$

$= 2 \int_{0}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ ($\because$ even function)

let $u = x^2$ $\quad du = 2x\,dx$

$$= 2\int_0^t \frac{1}{\sqrt{2\pi}} e^{-u/2} \frac{du}{2(u)^{1/2}}$$

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

now, $\frac{d}{dt} P(X^2 \le t) = f_{X^2}(t) = \frac{1}{\sqrt{2\pi}} (t)^{-1/2} e^{-t/2}$ ($\because$ FTC)

for $t \in (0,\infty)$

so, $f_{X^2}(t) = \begin{cases} \frac{1}{\sqrt{2\pi}} (t)^{-1/2} e^{-t/2} & ; t \in (0,\infty) \\ 0 & ; \text{otwise} \end{cases}$

also gamma$(\alpha, \lambda)$ has

$$f(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & ; x > 0 \\ 0 & ; \text{otherwise} \end{cases}$$

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

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

so, gamma$\left(\frac{1}{2}, \frac{1}{2}\right)$ $f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-x/2} x^{-1/2} & ; x > 0 \\ 0 & ; \text{otwise} \end{cases}$

so, $X^2 \sim$ gamma$\left(\frac{1}{2}, \frac{1}{2}\right)$

Exe: suppose $U, V$ iid unif$(0,1)$, then find the pdf of $Z = U+V$ using convolution formula

Ans:

now $U, V \sim$ unif$(0,1)$

$$f_U(x) = \mathbb{I}_{x \in [0,1]}$$

$$f_V(x) = \mathbb{I}_{x \in [0,1]}$$

then $f_{U+V}(x) = \int_{\mathbb{R}} f_U(s) f_V(x-s) ds$ ($\because$ convolution)

$0 < s < 1$

$0 < x - s < 1$

as range of $U, V$ is $(0,1)$

range of $U+V$ is $(0,2)$

if $x \in (0,1)$ then $0 < s < x$ $\quad s < x < 1+s$

if $x \in [1,2)$ then $x-1 < s < 1$

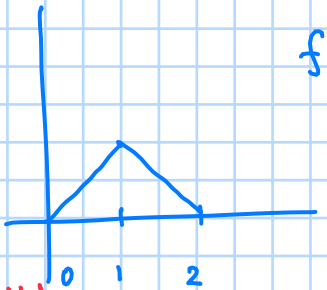
$\otimes$, in convolution formula, from case I: $x \in (0,1)$

$$f_{U+V}(x) = \int_0^x (1)(1)\,dx = x$$

Case II: $x \in [1,2)$

$$f_{U+V}(x) = \int_{x-1}^{1} (1)(1)\,dx = 1-(x-1) = 2-x$$

$$f_{U+V}(x) = \begin{cases} x & ; x \in (0,1) \\ 2-x & ; x \in [1,2) \\ 0 & ; \text{otherwise} \end{cases}$$




Exe: $X, Y \overset{iid}{\sim} N(0,1)$ find pdf of $Z := \frac{X}{Y}$

Ans:

$X, Y \sim N(0,1)$ $\quad Z := \frac{X}{Y}$

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

$$= \int_{-\infty}^{\infty} |y| \frac{1}{\sqrt{2\pi}} e^{-(yz)^2/2} \frac{1}{\sqrt{2\pi}} e^{-(y)^2/2}\, dy$$

$$h(z) = \int_{-\infty}^{\infty} \frac{1}{2\pi} |y|\, e^{-y^2z^2/2} e^{-y^2/2}\, dy$$

$$= 2\int_0^{\infty} \frac{1}{2\pi} y\, e^{-y^2z^2/2} e^{-y^2/2}\, dy \quad (\because \text{even function})$$

$$y^2/2 = u$$
$$y\,dy = du$$

$$= \frac{1}{\pi} \int_0^{\infty} e^{-u(1+z^2)}\, du$$

$$= \frac{1}{\pi} \left[ \frac{e^{-u(1+z^2)}}{-(1+z^2)} \right]_0^{\infty}$$

$$h(z) = \frac{1}{\pi} \left[ \frac{1}{1+z^2} \right]$$

13th Jan:

left g further

Eg: $X_1 \sim \text{gamma}(\alpha_1, \lambda)$, $X_2 \sim \text{gamma}(\alpha_2, \lambda)$ ind

$S = X_1 + X_2$

$Y = \frac{X_1}{S} = \frac{X_1}{X_1 + X_2}$

we wanted to find joint pdf of Y and S

Range $(X_1, X_2) = (0,\infty)^2 = I$ (open and connected)

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

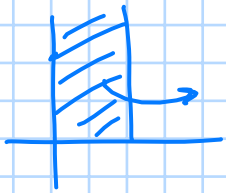
$$g(x_1, x_2) = \left(\frac{x_1}{x_1+x_2},\ x_1+x_2\right) \quad (x_1, x_2) \in I$$

clearly $(Y, S) = g(X_1, X_2)$, we saw that g is one-one

clearly, $g(I) = \{ g(x_1, x_2) \mid (x_1, x_2) \in I = (0,\infty)^2 \}$

$$= \left\{ \left(\frac{x_1}{x_1+x_2},\ x_1+x_2\right) \,\middle|\, x_1, x_2 \in (0,\infty) \right\}$$

$$\subseteq (0,1) \times (0,\infty) \quad (\because \text{trivial})$$

now  $(0,1) \times (0,\infty) =$ open and connected $= J$

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

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

$\exists (x_1, x_2) \in I$ s.t $(y, s) = \left(\frac{x_1}{x_1+x_2},\ x_1+x_2\right) = g(x_1, x_2)$

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

$\frac{x_1}{x_1+x_2} = y$ $\quad x_1 = ys$

$x_1 + x_2 = s$ $\quad x_2 = s - ys = s(1-y)$

$ys \in (0,\infty)$

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

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

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

so this shows $g(I) = J$

now $g: I \to J$ is one-one onto map

in the process of proving the claim, we have also computed the inverse map $g^{-1}$, namely the map

$g^{-1}: J \to I$ is given by

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

therefore $\forall (y, s) \in J = (0,1) \times (0,\infty)$

$h_1(y, s) = ys \quad h_2(y, s) = (1-y)s$

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

clearly partial derivatives $\frac{\partial h_1}{\partial y}$, $\frac{\partial h_2}{\partial s}$, $\frac{\partial h_2}{\partial y}$ and $\frac{\partial h_2}{\partial s}$ exist and are all continious on $J$

infact $\left.\begin{matrix} \frac{\partial h_1}{\partial y} = s & \frac{\partial h_1}{\partial s} = y \\ \frac{\partial h_2}{\partial y} = -s & \frac{\partial h_2}{\partial s} = 1-y \end{matrix}\right\} \forall (y,s) \in J$

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

$$\det(J_{g^{-1}}(y,s)) = (s)(1-y) + (y)(s) = s \in (0,\infty) > 0 \ \forall (y,s) \in J$$

this shows that all the analytic conditions of the multivariate change of density formula are satisfied

also $(X_1, X_2)$ is a continious random vector with range being $I$ having joint pdf

$$\varphi(x_1,x_2) = \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda(\alpha_1+\alpha_2)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} & ; \text{if } (x_1,x_2) \in I \\ 0 & ; \text{if } (x_1,x_2) \in I^c \end{cases}$$

so, now by the multivariate change of density formula

$(Y,S) = g(X_1, X_2)$ is also a cont random variable with range$(Y,S) = J$ having joint pdf

$$\psi(y,s) = \begin{cases} \varphi(g^{-1}(y,s)) |\det(J_{g^{-1}}(y,s))| & ; (y,s) \in J \\ 0 & ; \text{otherwise} \end{cases}$$

$$= \begin{cases} \varphi(ys, (1-y)s)\, s & ; (y,s) \in J \\ 0 & ; (y,s) \in J^c \end{cases}$$

$$= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\lambda(ys+(1-y)s)} (ys)^{\alpha_1-1} ((1-y)s)^{\alpha_2-1} s & ; (y,s) \in J \\ 0 & ; (y,s) \in J^c \end{cases}$$

$$= \begin{cases} \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)} e^{-\lambda s} (s)^{\alpha_1+\alpha_2-1} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y^{\alpha_1-1}(1-y)^{\alpha_2-1} & ; (y,s) \in J \\ 0 & ; (y,s) \in J^c \end{cases}$$

($\frac{1}{B(\alpha_1,\alpha_2)}$: pdf of beta$(\alpha_1,\alpha_2)$; pdf of gamma$(\alpha_1+\alpha_2, \lambda)$)

by integrating out $s$, $Y \sim \text{beta}(\alpha_1, \alpha_2)$
by integrating out $y$, $S \sim \text{gamma}(\alpha_1+\alpha_2, \lambda)$

Note: in above part ind $\leftarrow \begin{matrix} Y \sim \text{Beta}(\alpha_1,\alpha_2) \\ S \sim \text{gamma}(\alpha_1+\alpha_2, \lambda) \end{matrix}$

Note: ① the distribution of $Y$ does not depend on the scale parameter $\lambda$

② If $X_1, X_2 \overset{iid}{\sim} \text{Exp}(\lambda) = \text{gamma}(1, \lambda)$ then

$$\frac{X_1}{X_1+X_2} \sim \text{beta}(1,1) = \text{uniform}(0,1)$$

$$X_1 + X_2 \sim \text{gamma}(2, \lambda)$$

Exe: Suppose $n \geq 2$ and $X_1, X_2, \ldots X_n$ are i.i.d with $X_i \sim \text{gamma}(\alpha_i, \lambda)$ $\forall i = 1, 2 \ldots n$, let $S = \sum_{i=1}^{n} X_i$

$$Y_1 = \frac{X_1}{S} \cdots Y_{n-1} = \frac{X_{n-1}}{S}$$

then show that $S \perp\!\!\!\perp (Y_1, Y_2, \ldots Y_{n-1})$

$S \sim \text{gamma}(\sum_{i=1}^{n} \alpha_i, \lambda)$ and

$(Y_1, \ldots Y_{n-1})$ has joint pdf

$$f(y_1 \ldots y_{n-1}) = \frac{\Gamma(\sum_{i=1}^{n} \alpha_i)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} (y_1)^{\alpha_1 - 1} (y_2)^{\alpha_2 - 1} \ldots (y_{n-1})^{\alpha_{n-1}} \times (1 - \sum_{i=1}^{n-1} y_i)^{\alpha_n - 1}$$

→ love doun

if $(y_1, \ldots y_{n-1}) \in (0,1)^{n-1}$ and $\sum_{i=1}^{n-1} y_i < 1$

$S \sim \text{gamma}(\sum_{i=1}^{n} \alpha_i, \lambda)$, $\underset{\sim}{Y} \sim \text{Dirichlet}(\alpha_1 \ldots \alpha_{n-1}; \alpha_n)$ (ind)

Defn: fix $n \geq 2$, $\underset{\sim}{Y} = (Y_1, \ldots Y_{n-1})$ is said to follow $(n-1)$-dimensional Dirichlet distn with parameters $\alpha_1, \alpha_2, \ldots \alpha_{n-1}$ and $\alpha_n$ if $\underset{\sim}{Y}$ has joint pdf of form

$$f(y_1 \ldots y_{n-1}) = \frac{\Gamma(\sum_{i=1}^{n} \alpha_i)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} (y_1)^{\alpha_1 - 1} (y_2)^{\alpha_2 - 1} \ldots (y_{n-1})^{\alpha_{n-1}} \times (1 - \sum_{i=1}^{n-1} y_i)^{\alpha_n - 1}$$

if $(y_1, \ldots y_{n-1}) \in (0,1)^{n-1}$ and $\sum_{i=1}^{n-1} y_i < 1$

we write $\underset{\sim}{Y} \sim \text{Dirichlet}(\alpha_1, \alpha_2, \ldots \alpha_{n-1}; \alpha_n)$

Note: Dirichlet$(\alpha_1; \alpha_2)$ distn is same as Beta$(\alpha_1, \alpha_2)$ distn

Note: The joint distribution of $Y_1, Y_2, \ldots Y_{n-1}$ does not depend on $\lambda$

Theorem: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf) such that

$$\mathbb{E}[X_1] = \mu \text{ and } \text{Var}(X_1) = \sigma^2 \text{ (finite)}$$

let $\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i :=$ sample mean

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

then we have (a) $\mathbb{E}(\bar{X}_n) = \mu$

(b) $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$

(c) $\mathbb{E}[S^2] = \sigma^2$

proof: (a) as $\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{\sum X_i}{n}\right] = \sum \frac{\mathbb{E}[X_i]}{n} = \frac{n \cdot \mu}{n} = \mu$

(b) $\text{Var}(\bar{X}_n) = \mathbb{E}\left[\left(\frac{\sum X_i}{n}\right)^2\right] - \left(\mathbb{E}\left[\frac{\sum X_i}{n}\right]\right)^2 = \frac{1}{n^2}\left(\sum(\mathbb{E}[X_i^2]) + 2\sum_{i\neq j}\mathbb{E}[X_i X_j]\right)$

$= \frac{1}{n^2}\left(n\,\mathbb{E}[X_1^2] + 2\binom{n}{2}(\mathbb{E}[X_1])^2\right) - \mu^2$

$= \frac{1}{n}\mathbb{E}[X_1^2] + \frac{(n-1)}{n}\mathbb{E}[X_1]^2 - \mu^2 = \frac{1}{n}\sigma^2$

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

$\Rightarrow \mathbb{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] = \mathbb{E}\left[\sum_{i=1}^n X_i^2\right] - n\,\mathbb{E}[\bar{X}^2]$

$= n\,\mathbb{E}[X_1^2] - n\,\mathbb{E}[\bar{X}^2]$

$= n\left(\text{var}(X_1) + (\mathbb{E}[X_1])^2\right) - n\left(\text{var}(\bar{X}) + (\mathbb{E}(\bar{X}))^2\right)$

$= n[\sigma^2 + \mu^2] - n\left[\frac{\sigma^2}{n} + \mu^2\right]$

$= (n-1)\sigma^2$

so, $\mathbb{E}[S^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right] = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$

Theorem: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

(a) $\bar{X} \perp S^2$

(b) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

(c) $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2_{n-1}$

proof:

step 1: wlog we can assume that

$\mu = 0, \ \sigma = 1$

in general case work with

$X_i^* = \frac{X_i - \mu}{\sigma}, \ i = 1, 2, \ldots n \quad X_i^* \sim N(0,1)$

as $P(X^* \leq x) = P\left(\frac{X - \mu}{\sigma} \leq x\right) = P(X \leq \sigma x + \mu)$

$= \int_{-\infty}^{\sigma x + \mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$ let $u = \frac{t-\mu}{\sigma}$

$du = \frac{dt}{\sigma}$

$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u)^2} dt \Rightarrow X^* \sim N(0,1)$

Exe: suppose $n \geq 2$ and $X_1, X_2 \ldots X_n$ are i.r.v with $X_i \sim \text{gamma}(\alpha_i, \lambda)$ $\forall i = 1,2 \ldots n$, let $S = \sum_{i=1}^{n} X_i$

$$Y_1 = \frac{X_1}{S} \quad \cdots \quad Y_{n-1} = \frac{X_{n-1}}{S}$$

then show that $S \perp\!\!\!\perp (Y_1, Y_2, \ldots Y_{n-1})$

$S \sim \text{gamma}(\sum_{i=1}^{n} \alpha_i, \lambda)$ and

$(Y_1 \ldots Y_{n-1})$ has joint pdf

$$f(y_1 \ldots y_{n-1}) = \frac{\Gamma(\sum_{i=1}^{n} \alpha_i)}{\prod_{i=1}^{n} \Gamma(\alpha_i)} (y_1)^{\alpha_1 - 1} (y_2)^{\alpha_2 - 1} \ldots (y_{n-1})^{\alpha_{n-1}} \times (1 - \sum_{i=1}^{n-1} y_i)^{\alpha_n - 1}$$

if $(y_1 \ldots y_{n-1}) \in (0,1)^{n-1}$ and $\sum_{i=1}^{n-1} y_i < 1$

$S \sim \text{gamma}(\sum_{i=1}^{n} \alpha_i, \lambda)$, $Y \sim \text{Dirichlet}(\alpha_1 \ldots \alpha_{n-1}; \alpha_n)$ (ind)

Ans:

let $(X_1 \ldots X_n)$ $I = (0, \infty)^n$

$$J = (0, \infty) \times \{(y_1, y_2 \ldots, y_{n-1}) \mid \sum_{i=1}^{n-1} y_i < 1, \ y_i \in (0,1) \ \forall i\}$$

then $g : I \to J$ s.t

$$g(x_1, x_2 \ldots x_n) = (S, \frac{x_1}{S}, \frac{x_2}{S}, \ldots \frac{x_{n-1}}{S})$$

as $S = \sum_{i=1}^{n} x_i$ $\sum_{i=1}^{n-1} \frac{x_i}{S} = \frac{S - x_n}{S} = 1 - \frac{x_n}{S} < 1$

as $\frac{x_n}{S} > 0$

then ① g is one-one:

if $g(x_1, \ldots x_n) = g(y_1 \ldots y_n)$

$\Rightarrow (S_x, \frac{x_1}{S_x}, \ldots \frac{x_{n-1}}{S_x}) = (S_y, \frac{y_1}{S_y}, \ldots \frac{y_{n-1}}{S_y})$

$\Rightarrow S_x = S_y$ and $\frac{x_i}{S_x} = \frac{y_i}{S_y} \Rightarrow x_i = y_i \ \forall i \in \{1,2 \ldots n-1\}$

$\Rightarrow x_i = y_i \ \forall i \in \{1 \ldots n\}$

② g is onto:

as $\forall (S, y_1, y_2 \ldots y_{n-1}) \in J$

as $\sum_{i=1}^{n-1} y_i < 1$, and $y_i \in (0,1)$

$\exists$ $x_i = S y_i \ \forall i \in \{1,2 \ldots n-1\}$

$x_n = (1 - \sum_{i=1}^{n-1} y_i) S$ s.t $x_i \in (0, \infty) \ \forall i \in \{1 \ldots n\}$

and $g(sy_1, sy_2, \ldots sy_{n-1}, (1-\sum_{i=1}^{n-1} y_i)s) = (s, y_1 \ldots y_{n-1})$

so, $\forall\ \underset{\sim}{y} \in J$, $\exists\ \underset{\sim}{x} \in I$ s.t $g(\underset{\sim}{x}) = \underset{\sim}{y}$

$\Rightarrow g$ is onto

so, $g$ is bijective map, now

$g^{-1}(s, y_1, y_2, \ldots y_{n-1}) = (sy_1, sy_2, \ldots sy_{n-1}, (1-\sum_{i=1}^{n-1} y_i)s)$

$\Rightarrow h_i(s, y_1 \ldots y_{n-1}) = sy_i \quad \forall i \in \{1,2,\ldots n-1\}$

$h_n(s, y_1, y_2, \ldots y_{n-1}) = s(1-\sum_{i=1}^{n-1} y_i)$

$\frac{dh_i}{ds} = y_i \quad \frac{dh_i}{dy_i} = s\delta_{ij} \quad \forall i,j \in \{1,2\ldots n-1\}$

$\frac{dh_n}{ds} = 1-\sum_{i=1}^{n-1} y_i \quad \frac{dh_n}{dy_i} = -s \quad \forall i \in \{1,2\ldots n-1\}$

now trivial to see all $\frac{dh_i}{ds}, \frac{dh_i}{dy_j}$ are continuous

moreover,

$$J(g^{-1}(s, y_1, y_2, \ldots y_{n-1})) = \begin{pmatrix} y_1 & s & 0 & \cdots & 0 \\ y_2 & 0 & s & 0 \cdots & 0 \\ \vdots & 0 & & \ddots & \vdots \\ y_{n-1} & 0 & \cdots & 0 & s \\ 1-\sum_{i=1}^{n-1} y_i & -s & \cdots & -s & -s \end{pmatrix}$$

$$\det(J) = s^{n-1}\begin{pmatrix} y_1 \\ \vdots \\ y_{n-1} \\ 1-\sum y_i \end{pmatrix}\begin{pmatrix} I \\ -1 \cdots -1 \end{pmatrix}$$

$$= s^{n-1}\left[-\sum_{i=1}^{n-1} y_i - 1 + \sum_{i=1}^{n-1} y_i\right]$$

$$= -s^{n-1}$$

but $|\det(J)| = s^{n-1} \in (0,\infty)$ as $s \in (0,\infty)$

so, now $X_i \sim \text{gamma}(\alpha_i, \lambda)\ \forall i \in \{1,2\ldots n\}$

the joint PDF:

$$\varphi(x_1, x_2, \ldots x_n) = \frac{\lambda^{\sum_{i=1}^{n}\alpha_i}}{\prod_{i=1}^{n}\Gamma(\alpha_i)} e^{-\lambda(\sum_{i=1}^{n}\alpha_i)} \prod_{i=1}^{n} x_i^{\alpha_i - 1} \quad \begin{cases} \text{if } (x_i) \text{ in domain} \\ \text{else } 0 \end{cases}$$

joint PDF of $(S, Y_1, Y_2, \ldots Y_{n-1})$

s.t $S \sim \sum_{i=1}^{n} X_i \quad Y_i \sim \frac{X_i}{S} \quad \forall i \in \{1 \ldots n-1\}$

$\psi(s, y_1, y_2 \ldots y_{n-1}) = \varphi(g^{-1}(s, y_1, \ldots y_{n-1}))\,|\det(J)|$ $\begin{cases} \text{if in domain} \\ \text{else } 0 \end{cases}$

$= \varphi(y_1 s, y_2 s, \ldots y_{n-1} s, s(1-\sum_{i=1}^{n-1} y_i))\, s^{n-1}$

$$= \frac{\lambda^{\Sigma\alpha_i}}{\pi\Gamma(\alpha_i)} e^{-\lambda(\Sigma\alpha_i)} \pi(y_i s)^{\alpha_i - 1} s^{\alpha_{n+1}} [1 - \Sigma y_i]^{\alpha_{n+1}}$$

$$= \frac{\lambda^{\Sigma\alpha_i}}{\pi\Gamma(\alpha_i)} e^{-\lambda(\Sigma\alpha_i)} s^{\Sigma\alpha_i - n + n - 1} \pi(y_i)^{\alpha_i - 1} (1 - \Sigma y_i)^{\alpha_{n+1}}$$

$$= \underbrace{\left(\frac{\lambda^{\Sigma\alpha_i}}{\Gamma(\Sigma\alpha_i)} e^{-\lambda(\Sigma\alpha_i)} s^{\Sigma\alpha_i - 1}\right)}_{\text{distribution of } s \text{ is } \perp \text{ from of } (y_1, y_2 \ldots y_{n+1}) \text{ by taking integration}} \left(\frac{\Gamma(\Sigma\alpha_i)}{\pi\Gamma(\alpha_i)} \pi(y_i)^{\alpha_i - 1} (1 - \Sigma y_i)^{\alpha_{n+1}}\right)$$

$\Rightarrow s \sim \text{gamma}(\Sigma\alpha_i, \lambda)$

$(y_1, y_2, \ldots y_{n+1}) \sim \text{dirichlet}(\alpha_1, \alpha_2 \ldots \alpha_{n+1}; \alpha_n)$

and $s \perp\!\!\!\perp (y_1, y_2 \ldots y_{n+1})$

16th Jan:

**Theorem:** Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

(a) $\bar{X} \perp\!\!\!\perp S^2$

(b) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

(c) $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2_{n-1}$

**Proof:**

**Step 1:** WLOG, let's assume that $\mu = 0$ and $\sigma = 1$ as the general case works with

$$X_i^* = \frac{X_i - \mu}{\sigma} \quad i = 1, 2, \ldots n$$

then $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(0,1)$ we want to show

(a) $\bar{X} \perp\!\!\!\perp S^2$

(b) $\bar{X} \sim N(0, \frac{1}{n})$

(c) $(n-1)S^2 \sim \chi^2_{n-1}$

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

$$= (X_1 - \bar{X})^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2$$

now $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\Rightarrow n\bar{X} = \sum_{i=1}^{n} X_i$

$\Rightarrow 0 = \sum_{i=1}^{n}(X_i - \bar{X})$

$\Rightarrow -\sum_{i=2}^{n}(X_i - \bar{X}) = X_1 - \bar{X}$

So, $(n-1)S^2 = \left(\sum_{i=2}^{n}(X_i - \bar{X})\right)^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2$

So, $(n-1)S^2$ (and so $S^2$ also) can be written as a function of the random vector

$(X_2 - \bar{X}, X_3 - \bar{X}, \ldots X_n - \bar{X})$ So, we will show that

$\bar{X} \perp\!\!\!\perp (X_2 - \bar{X}, X_3 - \bar{X}, \ldots X_n - \bar{X})$ and hence $\bar{X} \perp\!\!\!\perp S^2$

to show this, we will compute joint PDF of $(\bar{X}, X_2 - \bar{X}, \ldots X_n - \bar{X})$

**Step 3:** Joint pdf of $X_1, X_2, \ldots X_n$

$$\varphi(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\}$$

if $(x_1, x_2, \ldots x_n) \in \mathbb{R}^n$

Let $g: I \xrightarrow{I = \mathbb{R}^n} J$ $I = \mathbb{R}^n$, $J = \mathbb{R}^n$ by

$$g(x_1, x_2, \ldots x_n) = (\bar{x},\ x_2 - \bar{x},\ x_3 - \bar{x}, \ldots\ x_n - \bar{x})$$

g is one-one: as

if $g(x_1, x_2 \ldots x_n) = g(y_1, y_2, \ldots y_n)$

then

$$(\bar{x}, x_2 - \bar{x}, \ldots x_n - \bar{x}) = (\bar{y}, y_2 - \bar{y}, \ldots y_n - \bar{y})$$

$$\Rightarrow \bar{x} = \bar{y} \text{ and } x_i = y_i \ \forall\ i = 2, 3, \ldots n$$

$$\Rightarrow x_i = y_i \ \forall\ i \in \{1, \ldots n\}$$

g is onto: as for any $(y_1, y_2, \ldots y_n) \in \mathbb{R}^n = J$

if $\bar{x} = y_1,\ x_2 - \bar{x} = y_2, \ldots\ x_n - \bar{x} = y_n$

$$\Rightarrow x_2 = y_2 + y_1,\ x_3 = y_3 + y_1, \ldots\ x_n = y_n + y_1$$

and so

$$\sum_{i=1}^{n} x_i = n y_1 = x_1 + \sum_{i=2}^{n} (y_i + y_1) = n y_1$$

$$\Rightarrow x_1 + (n-1) y_1 + \sum_{i=2}^{n} y_i = n y_1$$

$$\Rightarrow x_1 = y_1 - \sum_{i=2}^{n} y_i$$

and so, $\forall\ \underset{\sim}{y} = (y_1, y_2 \ldots y_n) \in \mathbb{R}^n = J$ s.t. $\exists\ (x_1, x_2 \ldots x_n) \in \mathbb{R}^n = I$

$$x_1 = y_1 - \sum_{i=2}^{n} y_i$$

$$x_2 = y_1 + y_2$$

$$\vdots$$

$$x_n = y_1 + y_n$$

$$g(x_1, x_2 \ldots x_n) = (y_1, y_2, \ldots y_n)$$

so, g is onto function

we have also computed $g^{-1} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

$$g^{-1}(y_1, y_2, \ldots y_n) = \left(y_1 - \sum_{i=2}^{n} y_i,\ y_1 + y_2, \ldots\ y_1 + y_n\right)$$

$$\forall\ (y_1 \ldots y_n) \in \mathbb{R}^n$$

$$h_1(y_1, y_2, \ldots y_n) = y_1 - \sum_{i=2}^{n} y_i$$

$$h_i(y_1, y_2, \ldots y_n) = y_1 + y_i \quad \forall\ i \in \{2, 3, \ldots n\}$$

so that, $g^{-1}(\underset{\sim}{y}) = (h_1(\underset{\sim}{y}), h_2(\underset{\sim}{y}), \ldots h_n(\underset{\sim}{y}))\ \forall\ \underset{\sim}{y} \in \mathbb{R}^n$

here $\forall\ \underset{\sim}{y} \in \mathbb{R}^n$, we have $h_1(y_1, y_2, \ldots y_n) = y_1 - \sum_{i=2}^{n} y_i$

$$\frac{\partial h_1}{\partial y_1} = 1,\ \frac{\partial h_1}{\partial y_2} = -1, \ldots\ \frac{\partial h_1}{\partial y_n} = -1$$

$$h_2(y_1, y_2 \ldots y_n) = y_1 + y_2$$

$$\frac{\partial h_2}{\partial y_1} = 1,\ \frac{\partial h_2}{\partial y_2} = 1,\ \frac{\partial h_2}{\partial y_3} = 0, \ldots \frac{\partial h_2}{\partial y_n} = 0$$

similarly $\frac{\partial h_i}{\partial y_1} = 1 \quad \forall\ i \in \{2, 3, \ldots n\}$

$$\frac{dh_i}{dy_1} = 1 \quad \forall i \in \{2,3,\dots n\}$$

$$\frac{dh_i}{dy_p} = 0 \text{ for } p \neq 1 \quad p, i \in \{2,3,\dots n\}$$

so, $\forall \underset{\sim}{y} \in \mathbb{R}^n$

$$J_{g^{-1}}(\underset{\sim}{y}) = \begin{pmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 1 & 0 & \cdots & 0 & 1 \end{pmatrix} \Rightarrow \det(J_{g^{-1}}(\underset{\sim}{y}))$$

$$= \det \begin{pmatrix} n & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ 1 & 0 & \ddots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 1 \end{pmatrix} \quad (\because \text{adding all 2 to n rows to 1})$$

$$= n > 0$$

so, all the conditions of multivariate change of density formula are satisfied, and so joined PDF of $(Y_1, Y_2, \dots Y_n)$ is

(← from prev result)

$$\psi(y_1, y_2, \dots y_n) = \varphi(g^{-1}(y_1, y_2, \dots y_n)) \left|\det(J_{g^{-1}}(y_1, y_2, \dots y_n))\right|$$

$$= \varphi\left(y_1 - \sum_{i=2}^{n} y_i, y_1 + y_2, \dots y_1 + y_n\right)(n)$$

$$= \frac{n}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\left\{\left(y_1 - \sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n}(y_1 + y_i)^2\right\}\right]$$

$$\left(y_1 - \sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n}(y_1 + y_i)^2 = n y_1^2 + \left(\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2$$

then $\psi(y_1, y_2, \dots y_n) = \left(\frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{n}}} e^{-\frac{y_1^2}{2\left(\frac{1}{\sqrt{n}}\right)^2}}\right)\left(\frac{\sqrt{n}}{(2\pi)^{(n-1)/2}} \exp\left\{-\frac{1}{2}\left(\left(\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{2n} y_i^2\right)\right\}\right)$

density is normal $\sim N(0, \frac{1}{n})$ — independent of $y_1$

so, $\bar{X} \sim N(0, \frac{1}{n})$ $\quad \bar{x} = y_1$ $\quad$ joint pdf of $(Y_2, \dots Y_n)$ $f(y_2, y_3, \dots y_n)$

and $Y_1 \perp\!\!\!\perp (Y_2, Y_3, \dots Y_n)$

i.e $\bar{X} \perp\!\!\!\perp (X_2 - \bar{X}, X_3 - \bar{X}, \dots X_n - \bar{X})$

$\Rightarrow \bar{X} \perp\!\!\!\perp S^2$ as $S^2$ is fution of $(X_2 - \bar{X}, \dots X_n - \bar{X})$

Step 4: now we have to show that when $\mu = 0$, $\sigma = 1$

$$X_1, X_2, \dots X_n \overset{iid}{\sim} N(0,1) \Rightarrow (n-1) S_n^2 \sim \chi^2_{n-1} \quad \forall n > 1/2$$

we shall show this using induction on n

now, $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(0,1) \Rightarrow \sum_{i=1}^{n} \alpha_i X_i \sim N(0, \sum_{i=1}^{n} \alpha_i^2)$ $\forall \alpha_1, \alpha_2 \ldots \alpha_n \in \mathbb{R}$ ——①

as if $X \sim N(0,1)$ then $\alpha X \sim N(0, \alpha^2)$ for $\alpha \neq 0$ w.l.o.g $\alpha > 0$

$$P(\alpha X \leq x) = P(X \leq \frac{x}{\alpha}) = \int_{-\infty}^{x/\alpha} \frac{1}{\sqrt{2\pi}} e^{t^2/2} dt \quad \text{let } \alpha t = u, \ \alpha\, dt = du$$

$$= \int_{-\infty}^{x} \frac{1}{\alpha\sqrt{2\pi}} e^{-u^2/2\alpha^2} du$$

$$\Rightarrow \alpha X \sim N(0, \alpha^2)$$

now, $M_{\alpha_1 X_1}(t) = \mathbb{E}[e^{t(\alpha_1 X_1)}]$

$M_{\alpha_2 X_2}(t) = \mathbb{E}[e^{t(\alpha_2 X_2)}]$ $\Rightarrow M_{\alpha_1 X_1 + \alpha_2 X_2}(t) = M_{\alpha_1 X_1}(t)\, M_{\alpha_2 X_2}(t)$

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

$$M_X(t) = \int_{-\infty}^{\infty} e^{xt} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

$$xt - \frac{(x-\mu)^2}{2\sigma^2} = \frac{1}{2\sigma^2}\left[2\sigma^2 xt - x^2 - \mu^2 + 2x\mu\right]$$

$$= -\frac{1}{2\sigma^2}\left[x^2 + \mu^2 - 2\sigma^2 xt - 2x\mu\right]$$

$$= -\frac{1}{2\sigma^2}\left[x^2 - 2x(\sigma^2 t + \mu) + (\sigma^2 t + \mu)^2 + \mu^2 - (\sigma^2 t + \mu)^2\right]$$

$$= -\frac{1}{2\sigma^2}\left[(x - \sigma^2 t - \mu)^2 + \mu^2 - \sigma^4 t^2 - \mu^2 - 2\sigma^2 t\mu\right]$$

$$= -\frac{1}{2\sigma^2}\left[(x - \sigma^2 t - \mu)^2\right] + \frac{1}{2}\sigma^2 t^2 + t\mu$$

$$\Rightarrow M_X(t) = e^{\frac{1}{2}\sigma^2 t^2 + \mu t} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x - \sigma^2 t - \mu)^2} dx$$

$$= e^{\frac{1}{2}\sigma^2 t^2 + \mu t}$$

then $M_{\alpha_1 X_1}(t) \cdot M_{\alpha_2 X_2}(t) = e^{\frac{1}{2}\alpha_1^2 t^2} e^{\frac{1}{2}\alpha_2^2 t^2} = e^{\frac{1}{2}(\alpha_1^2 + \alpha_2^2)t^2}$

$$\Rightarrow M_{\alpha_1 X_1 + \alpha_2 X_2}(t) = e^{\frac{1}{2}(\alpha_1^2 + \alpha_2^2)t^2} \Rightarrow \alpha_1 X_1 + \alpha_2 X_2 \sim N(0, \alpha_1^2 + \alpha_2^2)$$

then by induction (trivial as $(\sum_{i=1}^{n-1} \alpha_i X_i) + \alpha_n X_n \sim N(0, \sum_{i=1}^{n} \alpha_i^2)$ )
true for general case
now with our original equation

case I: $n = 2$: $(2-1)S_2^2 \sim \chi_1^2$

as $S_2^2 = \left(X_1 - \left(\frac{X_1 + X_2}{2}\right)\right)^2 + \left(X_2 - \left(\frac{X_1 + X_2}{2}\right)\right)^2$

$$= \left(\frac{X_1 - X_2}{2}\right)^2 + \left(\frac{X_1 - X_2}{2}\right)^2$$

$$S_2^2 = \left(\frac{X_1 - X_2}{\sqrt{2}}\right)^2$$

as $X_1 \sim N(0,1)$, $X_2 \sim N(0,1) \Rightarrow \left(\frac{X_1 - X_2}{\sqrt{2}}\right) \sim N\left(0, \frac{1}{2} + \frac{1}{2}\right)$

$\Rightarrow \left(\frac{X_1 - X_2}{\sqrt{2}}\right)^2 \sim \chi_1^2$ ($\because$ ①)

$\Rightarrow (2-1) S_2^2 \sim \chi_1^2$

this shows the result for $n = 2$

case II: $n = k+1$:
lets assume statement for $n = k$, then we have to show it for $n = k+1$

in other words, lets assume that $(k-1) S_k^2 \sim \chi_{k-1}^2$

we have to show $k S_{k+1}^2 \sim \chi_k^2$

as $k S_{k+1}^2 = \underbrace{(k-1) S_k^2}_{\sim \chi_{k-1}^2} + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2$ ($\because$ result below)

↑ we want to show $\frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2 \sim \chi_1^2$

and $(k-1) S_k^2$ ind $\frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2$

ex: show that $k S_{k+1}^2 = (k-1) S_k^2 + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2$

Ans: $k S_{k+1}^2 = \cancel{k}\left(\frac{1}{\cancel{k+1}-1}\right)\left[\sum_{i=1}^{k+1}(X_p - \bar{X})^2\right]$

$= \sum_{i=1}^{k+1}(X_i - \bar{X})^2 \qquad \bar{X} = \bar{X}_{k+1}$

$= \sum_{i=1}^{k+1}(X_i)^2 + (k+1)(\bar{X})^2 - 2\left(\sum_{i=1}^{k+1} X_i\right)(\bar{X})$

$= \sum_{i=1}^{k+1}(X_i)^2 + (k+1)(\bar{X})^2 - 2(k+1)(\bar{X})^2$

$= \sum_{i=1}^{k+1}(X_i)^2 - (k+1)(\bar{X})^2$

$= (k-1) S_k^2 - \sum_{i=1}^{k}(X_i)^2 + (k)(\bar{X}_k)^2 + \sum_{i=1}^{k+1}(X_i)^2 - (k+1)(\bar{X})^2$ ($\because$ prov step k instead of k+1)

$= (k-1) S_k^2 + k(\bar{X}_k)^2 + (X_{k+1})^2 - (k+1)(\bar{X}_{k+1})^2$

$= (k-1) S_k^2 + k(\bar{X}_k)^2 + (X_{k+1})^2 - (k+1)\left[\frac{k\bar{X}_k + X_{k+1}}{k+1}\right]^2$

$= (k-1) S_k^2 + k(\bar{X}_k)^2 + (X_{k+1})^2 - \underbrace{\left[k^2(\bar{X}_k)^2 + X_{k+1}^2 + 2k\bar{X}_k X_{k+1}\right]}_{k+1}$

$$= (k-1)S_k^2 + \frac{k^2(\bar{X}_k)^2 + k(X_k)^2}{k+1} + \frac{k(X_{k+1})^2 + X_{k+1}^2}{k+1} - \left[\frac{k^2(\bar{X}_k)^2 + X_{k+1}^2 + 2k\bar{X}_k X_{k+1}}{k+1}\right]$$

$$= (k-1)S_k^2 + \frac{k}{k+1}\left[(\bar{X}_k)^2 + (X_{k+1})^2 - 2\bar{X}_k X_{k+1}\right]$$

$$= (k-1)S_k^2 + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2$$

20th Jan:

**Theorem:** Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

(a) $\bar{X} \perp S^2$

(b) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

(c) $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2_{n-1}$

**proof:** we showed $\bar{X} \perp S^2$
and $\bar{X} \sim N(0, \frac{1}{n})$

then let's again assume $X_i \sim N(0,1)$

we also showed
if $X_i \sim N(0,1)\ \forall i \in \{1, \dots n\}$
$\Rightarrow \sum_{i=1}^{n} a_i X_i \sim N(0, \sum_{i=1}^{n} a_i^2)$ ——— (1)

$n=2$:

$$(2-1)S_2^2 = \left(X_1 - \left(\frac{X_1+X_2}{2}\right)\right)^2 + \left(X_2 - \left(\frac{X_1+X_2}{2}\right)\right)^2$$
$$\sim N(0,1)$$

let's assume statement true for $n=k$

$n=k+1$:

$$kS_{k+1}^2 = (k-1)S_k^2 + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2$$

($\because$ we have shown this before)

$$X_{k+1} - \bar{X}_k = X_{k+1} - \frac{X_1}{k} - \frac{X_2}{k} \cdots - \frac{X_k}{k}$$

$$\sim N\left(0, \underbrace{1 + \frac{1}{k^2} + \frac{1}{k^2} + \cdots \frac{1}{k^2}}_{1+\frac{1}{k}}\right) \quad (\because \text{proposition } (1))$$

so, $X_{k+1} - \bar{X}_k \sim N(0, \frac{k+1}{k})$

then $\sqrt{\frac{k}{k+1}}\left(X_{k+1} - \bar{X}_k\right) \sim N(0,1)$ ($\because$ proposition (1))

$$\Rightarrow \left(\frac{k}{k+1}\right)\left(X_{k+1} - \bar{X}_k\right)^2 \sim \chi_1^2$$

and induction Hypothesis $\Rightarrow (k-1)S_k^2 \sim \chi^2_{k-1}$

so, if $(k-1)S_k^2$ and $\left(\frac{k}{k+1}\right)(X_{k+1} - \bar{X}_k)^2$ are independent, then their addition will be $\chi_k^2$ and we are done

so, we have to show $S_k^2 \perp (X_{k+1} - \bar{X}_k)^2$

as $s_k^2$ is fuction of $(X_2-\bar{X}_k, \ldots X_k-\bar{X}_k)$ and $(X_{k+1}-\bar{X}_k)^2$ is a fuction of

$(X_{k+1}, \bar{X}_k)$

$\Rightarrow S_k^2 \perp\!\!\!\perp (X_{k+1}-\bar{X}_k)^2$ ($\because$ exercise above below)

so, $(n-1)S^2 \sim \chi^2_{n-1}$

exe: show the following

$X_1, X_2, \ldots X_{k+1} \overset{iid}{\sim} N(0,1)$

$\Rightarrow (X_2-\bar{X}_k, X_3-\bar{X}_k, \ldots X_k-\bar{X}_k) \perp\!\!\!\perp (\bar{X}_k, X_{k+1})$

Ans:

if $U \perp\!\!\!\perp V$ and $(U,V) \perp\!\!\!\perp W$ and $V \perp\!\!\!\perp W$

$\Rightarrow U \perp\!\!\!\perp (V,W)$

here $U=(X_2-\bar{X}_k, \ldots X_k-\bar{X}_k)$

then we have already shown that

$U \perp\!\!\!\perp \bar{X}_k = V$

and

$U \perp\!\!\!\perp X_{k+1} = W$ is trivial from given data

and $(U,V) \perp\!\!\!\perp W$ is also trivial as $U,V$ is fuction of $\{X_1, \ldots X_k\}$ $W$ is $X_{k+1}$

$\Rightarrow U \perp\!\!\!\perp (V,W)$

$\Rightarrow (X_2-\bar{X}_k, \ldots X_k-\bar{X}_k) \perp\!\!\!\perp (\bar{X}_k, X_{k+1})$

two important distributions:

1) (Student's) t distribution

2) Fisher-snedecor distribution or F distribution

t-distribution:

suppose, $Z \sim N(0,1)$, $V \sim \chi^2_m$ } ind

then

$T := \dfrac{Z}{\sqrt{V/m}}$ is said to follow t distribution with m degree of freedom

notation: $T \sim t_m$

eg: $X_1 \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\quad V := \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

we know $\bar{X} \perp\!\!\!\perp V$ ($\because$ previous theorem)

**Theorem:** If $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ then

$$\frac{\sqrt{n}\,(\bar{X}-\mu)}{S} \sim t_{n-1}$$

**Proof:** $X_1 \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\quad V := \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

we know $\bar{X} \perp\!\!\!\perp V$ ($\because$ previous theorem)

$$Z = \sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1) \qquad Z \perp\!\!\!\perp V$$

$$T := Z / \sqrt{V/(n-1)} \sim t_{n-1} \quad (\because \text{definition})$$

$$T := \frac{\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)}{+\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} \sim t_{n-1}$$

$$= \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{S^2}} \sim t_{n-1}$$

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

**F - distribution:**

let

$$\text{ind} \begin{cases} U_1 \sim \chi^2_{d_1} \\ U_2 \sim \chi^2_{d_2} \end{cases}$$

then $W := \dfrac{U_1/d_1}{U_2/d_2}$ is said to follow F distribution with $d_1$ and $d_2$ degrees of freedom

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

**Note:** $W \sim F_{d_1,d_2} \Rightarrow \dfrac{1}{W} \sim F_{d_2,d_1}$

**Eg:** Suppose we have two independent random samples

$$\text{ind} \begin{cases} X_1, X_2, \ldots X_m \overset{iid}{\sim} N(\mu_1, \sigma_1^2) \\ Y_1, Y_2, \ldots Y_n \overset{iid}{\sim} N(\mu_2, \sigma_2^2) \end{cases}$$

Let's define

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

$$S_y^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$U_1 := \frac{(m-1)S_X^2}{\sigma_1^2} \sim \chi^2_{m-1}$$

$$U_2 := \frac{(n-1)S_Y^2}{\sigma_2^2} \sim \chi^2_{n-1}$$

Hence, $W := \dfrac{U_1/(m-1)}{U_2/(n-1)} \sim F_{m-1,n-1}$

NOTE: In above, we can write W as

$$W = \frac{U_1/(m-1)}{U_2/(n-1)} = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}$$

$$= \left(\frac{S_X^2}{S_Y^2}\right) \times \left(\frac{\sigma_2^2}{\sigma_1^2}\right) \sim F_{m-1,n-1}$$

THEOREM: if we have two independent random samples

and $X_1, X_2 \ldots X_m$ iid $N(\mu_1, \sigma_1^2)$

$Y_1, Y_2, \ldots Y_n$ iid $N(\mu_2, \sigma_2^2)$

then $\left(\dfrac{S_X^2}{S_Y^2}\right)\cdot\left(\dfrac{\sigma_2^2}{\sigma_1^2}\right) \sim F_{m-1,n-1}$

if $\sigma_1^2 = \sigma_2^2$ then

$$\frac{S_X^2}{S_Y^2} \sim F_{m-1,n-1}$$

proof:

as

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

$$S_Y^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

$$U_1 := \frac{(m-1)S_X^2}{\sigma_1^2} \sim \chi^2_{m-1}$$

$$U_2 := \frac{(n-1)S_Y^2}{\sigma_2^2} \sim \chi^2_{n-1}$$

$$W = \frac{U_1/(m-1)}{U_2/(n-1)} = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}$$

$$= \left(\frac{S_X^2}{S_Y^2}\right) \times \left(\frac{\sigma_2^2}{\sigma_1^2}\right) \sim F_{m-1,n-1}$$

and trivial if $\sigma_1^2 = \sigma_2^2$

Quiz 1 - Thursday / Friday next week

No class this friday

**Order statistics:**

Suppose $X_1, X_2, \ldots X_n$ is a random sample from some distribution

let $X_{(1)} = \min_{1 \le i \le n} \{X_i\}$

$X_{(2)}$ = second smallest among $\{X_1, X_2, \ldots X_n\}$

$\vdots$

$X_{(k)}$ = $k^{th}$ smallest among $\{X_1, X_2, \ldots X_n\}$

$1 \le k \le n$ $\vdots$

$X_{(n)} = \max_{1 \le i \le n} \{X_i\}$

**Note:** If there is a tie choose any

So, $X_{(1)} \le X_{(2)} \le X_{(3)} \cdots \le X_{(n)}$

**Defn:** $X_{(1)}, X_{(2)}, \ldots X_{(n)}$ are called order statistics

in general, $X_{(k)}$ = $k^{th}$ order statistic

**Defn:** sample range $R := X_{(n)} - X_{(1)}$

**Defn:** Sample median $M := \begin{cases} X_{(\frac{n+1}{2})} & ; n \text{ is odd} \\ \dfrac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2} & ; n \text{ is even} \end{cases}$

**Note:** median is not unique for even $n$

as if $\{15, \underbrace{21, 36}, 47\}$

any number between 21 and 36 can be median

**Discrete case:**

**Theorm:** Suppose $X_1, X_2, \ldots X_n$ is a random sample from a discrete distribution with pmf

$$f(x) = \begin{cases} p_1 & ; \text{if } x = x_1 \\ p_2 & ; \text{if } x = x_2 \\ p_3 & ; \text{if } x = x_3 \\ \vdots \end{cases}$$

where each $p_i > 0$ and $x_1 < x_2 < x_3 < \cdots$ are the possible values of $X_1$ arranged in increasing order

$p_i$ = prob that $X_1$ takes $x_i$ value, $\sum_{i=1}^{\infty} p_i = 1$

let $P_0 = 0$, $P_1 = p_1$, $P_2 = p_1 + p_2$, ... $P_p = p_1 + p_2 + \cdots + p_p$

$P_i^p = P[X_1 \leq x_i^p]$, true for each $j \in \{1, 2, \ldots n\}$ and for each $p$

$$P[X_{(j)} = x_p] = \sum_{k=j}^{n} \binom{n}{k} \left[ P_i^{p\,k} (1 - P_i^p)^{n-k} - P_{i-1}^{p\,k} (1 - P_{i-1}^p)^{n-k} \right]$$

this gives us pmf of the $j^{th}$ order statistic $X_{(j)}$

proof: fix $p$:

$$Y := \sum_{i=1}^{n} I_{(X_i \leq x_i)}$$

where $I_{(X_i \leq x_p)} = \begin{cases} 1 \; ; X_i \leq x_i \\ 0 \; ; X_i > x_i \end{cases} \quad \forall i \in \{1, 2, \ldots n\}$

Y counts the no of sample points that are less than or equal to $x_i^p$

27th Jan :

Quiz 1 : Friday (30th Jan) LA 002 , 15:45–16:15 , syllabus : uptil order statistics
No tutorial this week

## Order statistics :

Suppose $X_1, X_2, \dots X_n$ is a random sample

$$X_{(i)} = i^{th} \text{ min out of } \{x_1, x_2, \dots x_n\}$$

## Discrete case :

**Theorem:** suppose $X_1, X_2, \dots X_n$ is a random sample from a discrete distribution with pmf

$$f(x) = \begin{cases} p_1 & ; \text{ if } x = x_1 \\ p_2 & ; \text{ if } x = x_2 \\ p_3 & ; \text{ if } x = x_3 \\ \vdots \end{cases}$$

where each $p_i > 0$ and $x_1 < x_2 < x_3 < \dots$ are the possible values of $X_1$ arranged in increasing order

$p_i$ = prob that $X_1$ takes $x_i$ value, $\sum_{i=1}^{\infty} p_i = 1$

Let $P_0 = 0$, $P_1 = p_1$, $P_2 = p_1 + p_2$, $\dots$ $P_i = p_1 + p_2 + \dots + p_i$

$P_i = P[X_1 \le x_i]$, true for each $j \in \{1, 2, \dots n\}$ and for each $i$

$$P[X_{(j)} = x_i] = \sum_{k=j}^{n} \binom{n}{k} \left[ P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k} \right]$$

this gives us pmf of the $j^{th}$ order statistic $X_{(j)}$

**proof:**

fix $i$ $\quad Y := \sum_{l=1}^{n} I_{(X_l \le x_i)}$

$$I_{(X_l \le x_i)} = \begin{cases} 1 & ; X_l \le x_i \\ 0 & ; X_l > x_i \end{cases}$$

$Y$ counts no. of $\{X_1, X_2, \dots X_l\}$ that are $\le x_i$

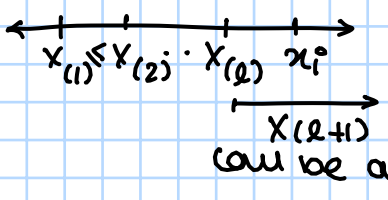
as $X_1, X_2, X_3, \dots X_n$ are iid rv's

$$I_{(X_1 \le x_i)}, I_{(X_2 \le x_i)}, \dots I_{(X_n \le x_i)} \overset{iid}{\sim} \text{Ber}(P(X_1 \le x_i]) = \text{Ber}(P_i)$$

$$\Rightarrow Y = \sum_{l=1}^{n} I_{(X_l \le x_i)} \sim \text{Bin}(n, P_i) \quad (\because \text{sum of Ber is Binomial})$$

also, event $(X_{(j)} \le x_i)$ is same as the event $Y \ge j$

($Y$ = no of variables s.t. $X_l \le x_i$ could be more than $j$)



So, $P[Y \geq j] = P[X_{(j)} \leq x_i] \quad \forall j = \{1,2,\dots n\}$

$$P[X_{(j)} \leq x_i] = \sum_{k=j}^{n} P[Y=k]$$

$$= \sum_{k=j}^{n} \binom{n}{k} (p_i)^k (1-p_i)^{n-k} \quad (\because Y \sim Bin(n, p_i))$$

Therefore $\forall\ i \geq 2 \ \forall\ j \in \{1,2..n\}$

$$P[X_{(j)} = x_i] = P[X_{(j)} \leq x_i] - P[X_{(j)} < x_i]$$

$$= P[X_{(j)} \leq x_i] - P[X_{(j)} \leq x_{i-1}]$$

$(\because x_1, x_2, \dots$ are ordered$)$

$$P[X_{(j)} = x_i] = \sum_{k=j}^{n} \binom{n}{k} \left[ (p_i)^k (1-p_i)^{n-k} - (p_{i-1})^k (1-p_{i-1})^{n-k} \right]$$

if $i=1$, then $\forall\ j \in \{1,2..n\}$

$$P[X_{(j)} = x_1] = P[X_{(j)} \leq x_1]$$

$$= \sum_{k=j}^{n} \binom{n}{k} (p_1)^k (1-p_1)^{n-k}$$

$$= \sum_{k=j}^{n} \binom{n}{k} \left[ (p_1)^k (1-p_1)^{n-k} - (p_0)^k (1-p_0)^{n-k} \right]$$

as $p_0 = 0$

Continious case:

Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x)$ (pdf) and $F(x) = \int_{-\infty}^{x} f(u)du$ is cdf of $X_1$

in this case $P[X_{(1)} < X_{(2)} < \cdots < X_{(n)}] = 1$, i.e we dont have to worry about ties

distribution of $X_{(n)}$:

fix $x \in \mathbb{R}$, then cdf of $X_{(n)}$ is $F_{X_{(n)}}(x) = P[X_{(n)} \leq x]$

$$= P\left[\max_{1 \leq j \leq n} X_j \leq x\right]$$

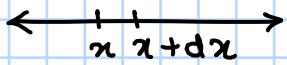
$$= \prod_{j=1}^{n} P[X_j \leq x] = (F(x))^n$$

then pdf of $X_{(n)}$

$$\frac{d}{dx} (F(x))^n = f_{X_{(n)}}(x)$$

$$\Rightarrow f_{X_{(n)}}(x) = n(F(x))^{n-1} f(x) = n f(x) \left[\int_{-\infty}^{x} f(u)du\right]^{n-1}$$

$(\because$ by chain rule$)$

Intuitively: 


$$f_{X_{(n)}}(x)\,dx \approx P[x \le X_{(n)} < x+dx]$$

$$\approx (F(x))^{n-1} f(x)\,dx + (\ ) + (\ ) \quad \text{(for } X_2\text{, for } X_3\text{)}$$

($(F(x))^{n-1}$: prob that $X_2, X_3 \ldots X_n \le x$; $f(x)\,dx$: prob that $x \le X_1 < x+dx$; $\ldots$)

$$\approx n\,(F(x))^{n-1} f(x)\,dx$$

($\binom{n}{1}$; rest $\le x$; this one b/w $x$ and $x+dx$)

**distribution of $X_{(1)}$:**

Fix $x \in \mathbb{R}$, $P[X_{(1)} \le x] = 1 - P[X_{(1)} > x]$

$$= 1 - P[\min_{1 \le j \le n} X_j > x]$$

$$= 1 - P[X_1 > x,\ X_2 > x, \ldots X_n > x]$$

$$= 1 - \prod_{j=1}^{n} P[X_j > x]$$

$$= 1 - \left(P[X_1 > x]\right)^n$$

$$F_{X_{(1)}}(x) = 1 - (1-F(x))^n$$

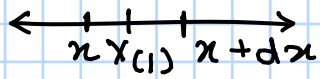
then pdf of $X_{(1)}$ is $f_{X_{(1)}}(x) = \frac{d}{dx} F_{X_{(1)}}(x)$

$$= \frac{d}{dx}\left[1 - (1-F(x))^n\right]$$

$$= +n(1-F(x))^{n-1}(+1)\frac{d}{dx}F(x)$$

$$f_{X_{(1)}}(x) = n(1-F(x))^{n-1} f(x)$$

so, pdf of $X_{(1)}$ is $f_{X_{(1)}}(x) = n[1-F(x)]^{n-1} f(x) \quad \forall x \in \mathbb{R}$

here also, 


$\binom{n}{1}(1-F(x))^{n-1} f(x)\,dx$ (1 − prob of ↑)

$$f_{X_{(1)}}(x)\,dx \approx P[x \le X_{(1)} \le x+dx]$$

$$\approx \underbrace{(1-F(x))^{n-1} f(x)\,dx}_{\text{any one } X_j \text{ is } x \text{ and rest all } > n} \times \underbrace{n}_{\binom{n}{1}\text{ choosing any one}}$$

**pdf of $X_{(j)}$:**

now $j \in \{1,2 \ldots n\}$, fix $j \in \{1,2 \ldots n\}$ and fix $x \in \mathbb{R}$

let $Y := \sum_{l=1}^{n} I_{(X_l \le x)}$ $\quad I_{(X_l \le x)} = \begin{cases} 1 ; & X_l \le x \\ 0 ; & X_l > x \end{cases}$

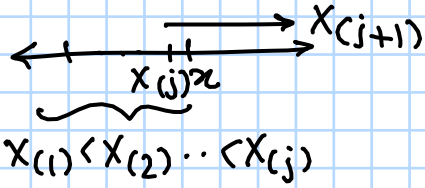
clearly $Y$ counts no. of $X_l$'s that are $\le x$

as $X_1, X_2, \ldots X_n$ are iid $\quad P(X_i \le x) = F(x)$

$I_{(X_1 \le x)}, \ldots I_{(X_n \le x)} \overset{iid}{\sim} Ber(F(x))$

$\Rightarrow Y = \sum_{i=1}^{n} I_{(X_i \le x)} \sim Bin(n, F(x))$

now, event $(X_{(j)} \le x)$ is same as $(Y \ge j)$




$X_{(1)} < X_{(2)} \cdots < X_{(j)}$

so, cdf of $X_{(j)}$ is $F_{X_{(j)}}(x) = P[X_{(j)} \le x]$

$= P[Y \ge j]$

$= \sum_{k=j}^{n} \binom{n}{k} (F(x))^k (1-F(x))^{n-k}$

now, pdf of $X_{(j)}$

$f_{X_{(j)}}(x) = \frac{d}{dx} F_{X_{(j)}}(x) = \frac{d}{dx} \sum_{k=j}^{n} \binom{n}{k} (F(x))^k (1-F(x))^{n-k}$

$= \sum_{k=j}^{n} \binom{n}{k} \frac{d}{dx} \left( (F(x))^k (1-F(x))^{n-k} \right)$

$= \sum_{k=j}^{n} \binom{n}{k} \Big[ k(F(x))^{k-1} (1-F(x))^{n-k} - (n-k)(F(x))^k (1-F(x))^{n-k-1} \Big] f(x)$

$= \binom{n}{j} j (F(x))^{j-1} (1-F(x))^{n-j} f(x)$

$+ \sum_{k=j+1}^{n} \binom{n}{k} k (F(x))^{k-1} (1-F(x))^{n-k} f(x)$

$- \sum_{k=j}^{n-1} \binom{n}{k} (n-k) (F(x))^k (1-F(x))^{n-k-1} f(x)$

Ex: show that two later two terms above (sigma terms) cancel out by using → done down

$\binom{n}{k}(n-k) = \binom{n}{k+1}(k+1)$

so, $f_{X_{(j)}}(x) = \binom{n}{j} j (F(x))^{j-1} (1-F(x))^{n-j} f(x)$

Theorem: If $X_1, X_2, \ldots X_n$ is a random sample from dist$^n$ with pdf $f(x)$ and cdf $F(x)$, then $j^{th}$ order statistic

$X_{(j)}$ has pdf $f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} (F(x))^{j-1} (1-F(x))^{n-j} f(x)$ $x \in \mathbb{R}$

**proof**: It follows from above calculations

**Exe**: show $X_{(j)}$ has pdf as above intuitively → **do in class**

**Exe**: show that the later two terms above (sigma terms) cancel out by using

$$\binom{n}{k}(n-k) = \binom{n}{k+1}(k+1)$$

**Ans**:

$$\sum_{k=j+1}^{n} \binom{n}{k} k (F(x))^{k-1} (1-F(x))^{n-k} f(x)$$

$$= \sum_{k=j}^{n} \binom{n}{k+1} (k+1) (F(x))^{k} (1-F(x))^{n-k-1} f(x) \quad (k \to k+1)$$

$$= \sum_{k=j}^{n} \binom{n}{k} (n-k) (F(x))^{k} (1-F(x))^{n-k-1} f(x)$$

$$\left( \because \binom{n}{k+1}(k+1) = \frac{n!}{(n-k-1)!(k+1)!}(k+1) = \binom{n}{k}(n-k) \right)$$

$$= \sum_{k=j}^{n} \binom{n}{k} (n-k) (F(x))^{k} (1-F(x))^{n-k-1} f(x)$$

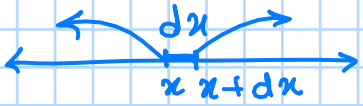
$(\because k = n$ term is $0)$

$\Rightarrow$ both sigma terms are same

**Exe**: show $X_{(j)}$ has pdf as above intuitively

**Ans**:

$$f_{X_{(j)}}(x) = \binom{n}{j} j (F(x))^{j-1} (1-F(x))^{n-j} f(x)$$




$$f_{X_{(j)}}(x)\,dx \approx P[x \le X_{(j)} \le x+dx]$$

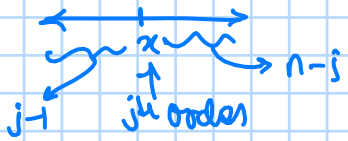
$$\approx \binom{n}{j} (j) f(x)dx \, (F(x))^{j-1} (1-F(x))^{n-j}$$

- $\binom{n}{j}$: choosing $j$ many
- $(j)$: one out of $j$ many is s.t. $x \le X \le x+dx$
- $f(x)dx$: one out of $j$ many is s.t. $x \le X \le x+dx$
- $(F(x))^{j-1}$: Rest $j-1$ many $\le x$
- $(1-F(x))^{n-j}$: other $n-j$ many $> x$

$$\approx \binom{n}{j} (j) (F(x))^{j-1} (1-F(x))^{n-j} f(x)dx$$

3rd Feb:

we have seen, by intuition




$$\binom{n}{j-1}(n-j+1)(F(x))^{j-1}(1-F(x))^{n-j}f(x)dx = \binom{n}{j}(j)(F(x))^{j-1}(1-F(x))^{n-j}f(x)dx$$

↑ choose $j-1$ all left ↑ choose any one

so from intuition $f_{X_{(j)}}(x) = \binom{n}{j} j (F(x))^{j-1}(1-F(x))^{n-j} f(x)$

Q9: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0,1)$

$$f(x) = \begin{cases} 1 ; & 0<x<1 \\ 0 ; & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} x ; & x \in (0,1) \\ 0 ; & x \le 0 \\ 1 ; & x \ge 1 \end{cases}$$

$$f_{X_{(j)}}(x) = \binom{n}{j} j (F(x))^{j-1}(1-F(x))^{n-j} f(x)$$

If $x \le 0$ then $f(x) = 0 \Rightarrow f_{X_{(j)}}(x) = 0$

similarly $x \ge 1$ then $f(x) = 0 \Rightarrow f_{X_{(j)}}(x) = 0$

So for $x \in (0,1)$

$$f_{X_{(j)}}(x) = \binom{n}{j}(j)(x)^{j-1}(1-x)^{n-j}$$

0 otherwise

$$= \frac{n!}{(j-1)!(n-j)!} x^{j-1}(1-x)^{n-j}$$

$$= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} (x)^{j-1}(1-x)^{(n-j+1)-1}$$

$$B(j, n-j+1) = \frac{\Gamma(j)\Gamma(n-j+1)}{\Gamma(n+1)}$$

So, $f_{X_{(j)}}(x) = \frac{1}{B(j, n-j+1)} (x)^{j-1}(1-x)^{(n-j+1)-1}$

$\Rightarrow X_{(j)} \sim \text{Beta}(j, n-j+1)$

Note: Recall that $Y \sim \text{Beta}(\alpha, \beta)$

$\Rightarrow \mathbb{E}[Y] = \frac{\alpha}{\alpha+\beta}$

$$\text{var}(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

so, using above formula

$$\forall j \in \{1,2,\dots,n\}$$

$$\mathbb{E}[X_{(j)}] = \frac{j}{j+n-j+1} = \frac{j}{n+1}$$

$$\text{var}(X_{(j)}) = \frac{(j)(n-j+1)}{(n+1)^2(n+1+1)}$$

$$= \frac{(j)(n-j+1)}{(n+1)^2(n+2)}$$

in particular, for $j = 1, n$:

$$\mathbb{E}[X_{(1)}] = \frac{1}{n+1}$$

$$\mathbb{E}[X_{(n)}] = \frac{n}{n+1}$$

$$\text{var}(X_{(1)}) = \frac{n}{(n+1)^2(n+2)}$$

$$\text{var}(X_{(n)}) = \frac{n}{(n+1)^2(n+2)}$$

**Theorem:** If $X_1, X_2, \dots, X_n$ is a random sample from a continious distribution with CDF $F(x)$ and pdf $f(x)$, then

$U = X_{(1)}$, $V = X_{(n)}$ have joint pdf

$$f_{U,V}(u,v) = \begin{cases} n(n-1)(F(v)-F(u))^{n-2} f(u) f(v) ; & u < v \\ 0 ; & u \geq v \end{cases}$$

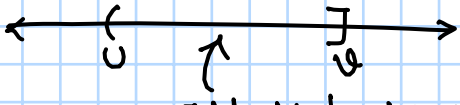
**proof:** we want to find joint pdf

$$P[U \leq u, V \leq v] \quad \forall u < v$$

let $u < v$ fixed
then its better to compute

$$P[u < \underbrace{U}_{X_{(1)}}, \underbrace{V}_{X_{(n)}} \leq v]$$

← equality sign does not matter here, but matters in discrete case




all $X_i$'s in this

$$P[u < U, V \leq v] = P[u < X_1 \leq v, u < X_2 \leq v, \dots u < X_n \leq v]$$

$$= P[u < X_1 \leq v] P[u < X_2 \leq v] \dots P[u < X_n \leq v]$$

$$= P[u < X_1 \leq v]^n$$

$$= \left(P[X_1 \leq v] - P[X_1 \leq u]\right)^n$$

$$= (F(v) - F(u))^n \quad u < v \quad \text{——} (1)$$

this gives joint pdf: $u<v$

$$F_{U,V}(u,v) = P[U \le u, V \le v]$$

now, the event $(V \le v) = (U \le u, V \le v) \cup (u < U, V \le v)$ (disjoint events)

$$\Rightarrow P(V \le v) = P(U \le u, V \le v) + P(u < U, V \le v)$$

so, $P[U \le u, V \le v] = P[V \le v] - P[u < U, V \le v]$

$$= P[X_{(n)} \le v] - (F(v)-F(u))^n \quad (\because ①)$$

$$= F_{X_{(n)}}(v) - (F(v)-F(u))^n$$

$$P[U \le u, V \le v] = (F(v))^n - (F(v)-F(u))^n \quad u < v$$

$(\because F_{X_{(n)}}(v) = \text{Prob}(\max \le v) = \text{prob}(\text{all} \le v) = F(v) \cdot F(v) \cdots F(v))$

so, if $u<v$ the joint pdf of U and V is

$$f_{U,V}(u,v) = \frac{\partial}{\partial v}\frac{\partial}{\partial u}\left(F(v)^n - (F(v)-F(u))^n\right)$$

$$= \frac{\partial}{\partial v}\left[+n(F(v)-F(u))^{n-1} f(u)\right]$$

$$f_{U,V}(u,v) = (n)(n-1)(F(v)-F(u))^{n-2} f(u) f(v)$$

this means,

$$f_{U,V}(u,v) = \begin{cases} n(n-1)(F(v)-F(u))^{n-2} f(u) f(v) & ; u<v \\ 0 & ; u \ge v \end{cases}$$

Note: minimal and maximum not independent, also from this we can compute distribution of range

Exe: using change of joint density formula, show that $U = X_{(1)}$, $R = X_{(n)} - X_{(1)}$ have joint probability density function

$$f_{U,R}(u,r) = f_{U,V}(u, u+r) \quad \text{if } -\infty < u < \infty,\ r > 0 \rightarrow \text{done down}$$

Exe: using above, prove that if

$X_1, X_2, \ldots X_n$ is a random from continuous distribution with cdf $F(x)$ and pdf $f(x)$, then the sample range $R := X_{(n)} - X_{(1)} = V - U$ has

$$\text{pdf } f_R(r) = \int_{-\infty}^{\infty} n(n-1)(F(u+r)-F(u))^{n-2} f(u+r) f(u)\, du$$

if $r>0$, 0 otherwise

→ done down

Note: If $X_1, X_2, \ldots X_n$ is a random from continuous distribution with cdf $F(x)$ and pdf $f(x)$, then the sample range

$R := X_{(n)} - X_{(1)} = V - U$ has

pdf $f_R(r) = \int_{-\infty}^{\infty} n(n-1)(F(u+r) - F(u))^{n-2} f(u+r) f(u)\, du$

if $r > 0$, 0 otherwise

proof: It follows from the exercises

exe: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Unif}(0,1)$

(a) find the joint pdf of $U := X_{(1)}$, $V := X_{(n)}$ → done down

(b) compute the pdf of the sample range $R := V - U = X_{(n)} - X_{(1)}$

Sufficiency:

setup: $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ ← possibly a multidimensional

random sample — pmf / pdf

$\theta$ = parameter unknown but non-random

$\Theta$ = parameter space (set of all possible values of $\theta$)

$\Theta \subseteq \mathbb{R}^n$

def: (Sample space) the joint range of $(X_1, X_2, \ldots X_n)$ = Range$(X_1, X_2, \ldots X_n)$

$$\text{Range}(X_1 \ldots X_n) = \text{Range}(X_1) \times \text{Range}(X_2) \underset{(\because \text{iid})}{\times} \ldots \times \text{Range}(X_n)$$
$$= (\text{Range}(X_1))^n$$
$$= \mathcal{X}$$

$\mathcal{X}$ = set of all values that the random vector $\underset{\sim}{X} = (X_1, X_2, \ldots X_n)$ can take

As usual, values in $\underset{\sim}{X}$ can be denoted by $\underset{\sim}{x} = (x_1, x_2, \ldots x_n)$

Def$^n$: Any function $T(X)$ of the random sample $\underset{\sim}{X} = (X_1, X_2, \ldots X_n)$ is called a Statistic

Eg: $T_1(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ sum

$T_2(\underset{\sim}{X}) = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}_n$

$T_3(\underset{\sim}{X}) = S^2(\underset{\sim}{X}) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$

$T_4(\underset{\sim}{X}) = X_{(1)} = \max_{1 \le i \le n} X_i$

$T_5(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$

$T_6(\underset{\sim}{X}) = R(\underset{\sim}{X}) = X_{(n)} - X_{(1)}$

**Defn:** A statistic $T = T(\underset{\sim}{X})$ is called a sufficient statistic of $\theta$ if the conditional distribution of the random sample $\underset{\sim}{X}$ given $T$ does not depend on $\theta$

**Note:** such a $T$ is sufficient for $\theta$ because of following reason if we know value of $T$, then the data does not contain any extra information about $\theta$

**Exe:** using usage of joint density formula, show that $U = X_{(1)}$, $R = X_{(n)} - X_{(1)}$ have joint probability density function

$$f_{U,R}(u,r) = f_{U,V}(u, u+r) \quad \text{if } -\infty < u < \infty, \; r > 0$$

**Ans:** $V = X_{(n)}$
$U = X_{(1)}$

now, $(U, V) \xrightarrow{g} (U, V-U) = (U, R)$

$g(u,v) = (u, v-u)$ then if range of $U, V = \mathbb{R}$ then trivial to see $v > u$
$(u, v-u)$ is $\mathbb{R} \times (0, \infty)$

let
$I = \mathbb{R}^2$
$J = (0, \infty)$

then $g(u,v) = (u, v-u)$

$g$ is one-one is trivial as
$g(u_1, v_1) = g(u_2, v_2)$
$\Rightarrow u_1 = u_2$ and $v_1 - u_1 = v_2 - u_2$
$\Rightarrow v_1 = v_2$

so, $(u_1, v_1) = (u_2, v_2)$

now, we $\forall (x,y) \in \mathbb{R} \times (0, \infty)$
$g(x, x+y) = (x, y+x-x)$
$= (x, y)$
so, $\exists (x, x+y) \in I$ s.t. $g(x, x+y) = (x, y)$

so, $g$ is a bijective function from $I$ to $J$

also $g^{-1}(x,y) = (x, x+y)$ from above steps

then
$h_1(x,y) = x$ → $\frac{\partial h_1}{\partial x} = 1$, $\frac{\partial h_1}{\partial y} = 0$
$h_2(x,y) = x+y$ → $\frac{\partial h_2}{\partial x} = 1$, $\frac{\partial h_2}{\partial y} = 1$
→ all are cont

Jacobian $J(x,y) = \begin{bmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$

$|J(x,y)| = 1 > 0 \; \forall \; (x,y)$

so, now by usage of joint density formula

$$f_{U,R}(u,r) = f_{u,v}(g^{-1}(u,r))\,|J(u,r)| \qquad \begin{matrix} u\in\mathbb{R} \\ r>0 \end{matrix}$$

$$= f_{u,v}(u, r+u) \quad \text{for } u\in\mathbb{R}, r>0$$

**Exe:** using above, prove that if

$X_1, X_2, \dots X_n$ is a random from continious distribution with cdf $F(x)$ and pdf $f(x)$, then the sample range $R := X_{(n)} - X_{(1)} = V - U$ has

pdf $f_R(r) = \int_{-\infty}^{\infty} n(n-1)(F(u+r)-F(u))^{n-2} f(u+r) f(u)\,du$

if $r>0$, 0 otherwise

**Ans:** $R = X_{(n)} - X_{(1)} = V - U$

as, $f_{U,R}(u,r) = f_{U,V}(u, u+r)$ (∵ previous exercise)

$$= n(n-1)[F(u+r)-F(u)]^{n-2} f(u+r) f(u)$$

now, $f_R(r) = \int_{-\infty}^{\infty} n(n-1)[F(u+r)-F(u)]^{n-2} f(u+r) f(u)\,du$ $\quad r>0$

otherwise 0

**Exe:** Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{Unif}(0,1)$

(a) find the joint pdf of $U := X_{(1)}$, $V := X_{(n)}$

(b) compute the pdf of the sample range $R := V - U = X_{(n)} - X_{(1)}$

**Ans:**

(a) $f_{U,V}(u,v) = (n)(n-1)[F(v)-F(u)]^{n-2} f(v) f(u)$

now for $u, v \in (0,1)$

$u < v$

$$= (n)(n-1)[v-u]^{n-2}(1)(1)$$

as otherwise 0

$$f_{U,V}(u,v) = \begin{cases} (n)(n-1)[v-u]^{n-2} & ; u<v,\ u,v\in(0,1) \\ 0 & ; \text{otherwise} \end{cases}$$

(b) $f_{U,R}(u,r) = f_{U,V}(u, u+r)$

$r>0$, $u\in(0,1)$, $u+r<1$, $r<1-u$

$0<r<1-u$, $0<u<1-r$, $r\in(0,1)$

then $f_R(r) = \int_0^{1-r} (n)(n-1)(r)^{n-2}\,du = (1-r)(r)^{n-2}(n)(n-1)$, $r\in(0,1)$

6th Feb:

Note: In discrete case, the conditional distribution is basic to understand, we just have look at the conditional pmf. In continious case a more sophisticated notion of conditional probability is required to fully understand and appreciate the defination of a sufficient statistic.

we will prove the results in discrete case and state the results without proof in continious case

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} Ber(p)$ $p \in (0,1)$, here $\theta = p$, $\Theta = (0,1)$, sample space $\mathcal{X} = \{0,1\} \times \{0,1\} \times \cdots \{0,1\} = \{0,1\}^n$, then

$T(X) = \sum_{i=1}^{n} X_i$ is a sufficient statistic for $p$

we shall compute conditional pmf of $(X_1, X_2, \dots X_n)$ given $T$

$T = \sum_{i=1}^{n} X_i \sim Bin(n, p)$ so, range $(T) = \{0,1,2,\dots n\}$

fix $t \in \{0,1,\dots n\}$

then conditional pmf

$f_{(X_1, X_2, \dots X_n)|T=t}(x_1, x_2, \dots x_n)$ of $(X_1, X_2, \dots X_n)$ given $T = t$

before we find conditional pmf, we first find conditional range

$\underset{\sim}{X} = (X_1, X_2, \dots X_n)$ given $T = t$ is

$\{(x_1, x_2, \dots x_n) \in \{0,1\}^n \mid \sum_{i=1}^{n} x_i = t\}$

for all other $f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = 0$ if $\underset{\sim}{x}$ s.t. $\sum_{i=1}^{n} x_i \neq t$ or $\underset{\sim}{x} \notin \{0,1\}^n$

now, $f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = P[X_1 = x_1, X_2 = x_2 \dots X_n = x_n | T = t]$

$= \dfrac{P[X_1 = x_1, X_2 = x_2, \dots X_n = x_n, T = t]}{P[T = t]}$ ← remove this part as all the $x_i$'s $= t$

$= \dfrac{P[X_1 = x_1]\, P[X_2 = x_2] \cdots P[X_n = x_n]}{P[T = t]}$

$= \dfrac{(p)^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}}{\binom{n}{t}(p)^t (1-p)^{n-t}}$

$= \dfrac{(p)^t (1-p)^{n-t}}{\binom{n}{t}(p)^t (1-p)^{n-t}}$

fixing sum makes all possible arrangements become equally likely

$f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = \dfrac{1}{\binom{n}{t}}$ so, $f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = \begin{cases} \dfrac{1}{\binom{n}{t}} ; & \underset{\sim}{x} \in \{0,1\}^n, \sum_{i=1}^{n} x_i = t \\ 0 ; & \text{otherwise} \end{cases}$

so, for each $t \in$ Range $(T) = \{0,1,2,\dots n\}$

conditional dist of $(X_1, X_2, \dots X_n)$ given $T = t$ does not depend on $p$

$\Rightarrow T = \sum_{i=1}^{n} X_i$ is sufficient statistic of $\theta$

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\lambda)$, show from definition that same statistic

$T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i \sim Poi(n\lambda)$ is sufficient statistic for $\lambda$ → done above

now, we already have seen that the conditional distribution of $X_1$ given

$T = \sum_{i=1}^{n} X_i = t$ does not depend on $\lambda$ (∵ tutorial Exercise)

The joint pmf / pdf of $(X_1, X_2, \dots X_n) = \underset{\sim}{X}$ is called the (joint) likelihood function

$$L(\underset{\sim}{x}|\theta) = L_\theta(\underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i|\theta) \quad \forall \ \underset{\sim}{x} \in \mathcal{X}$$

we want to know if we can use the likelihood fution to prove that a statistic T is sufficient for $\theta$ or not

**Theorem**: let $L(\underset{\sim}{x}|\theta)$ be a likelihood fution for a random sample $\underset{\sim}{x}$ and $T(\underset{\sim}{X}) = T$ is a statistic with pmf/pdf $q(t|\theta)$ then

$T(\underset{\sim}{X})$ is sufficient for $\theta$ if $\forall \ \underset{\sim}{x} \in \mathcal{X}$, the ratio

$$\frac{L(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}$$ is constant as a fution of $\theta$

Eg: let $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, 1)$ $\mu$ is unknown parameter, we want to show using above result for $\mu$ that $T(\underset{\sim}{X}) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{1}{n})$ is sufficient

$\theta = \mu$, $\Theta = \mathbb{R}$, $\mathcal{X} = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R} = \mathbb{R}^n$, then likelihood fuction

$$L(\underset{\sim}{x}|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x_i - \mu)^2}{2}\right\}$$

$$= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

$$= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}((x_i - \bar{x}) + (\bar{x} - \mu))^2\right\}$$

$$(x_i - \bar{x})^2 + (\bar{x} - \mu)^2 + 2(x_i - \bar{x})(\bar{x} - \mu)$$

$$= (x_i - \bar{x})^2 + (\bar{x} - \mu)^2 + 2[x_i\bar{x} - x_i\mu - \bar{x}^2 + \bar{x}\mu]$$

sigma cancels all this term

$$= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)\right\} \quad \underset{\sim}{x} \in \mathbb{R}^n$$

$$q(t|\mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n}} \exp\left\{-\frac{(t-\mu)^2}{2\frac{1}{n}}\right\}$$

$$= \left(\frac{2\pi}{n}\right)^{-1/2} \exp\left\{-\frac{n}{2}(t-\mu)^2\right\} \quad t \in \mathbb{R}$$

$$q(T(\underset{\sim}{x})|\theta) = q(\bar{x}|\theta) = \left(\frac{2\pi}{n}\right)^{-1/2} \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\} \quad \underset{\sim}{x} \in \mathbb{R}^n$$

now, fixing $\underset{\sim}{x} \in \mathbb{R}^n$, the ratio

$$\frac{L(\underset{\sim}{x}|\mu)}{q(T(\underset{\sim}{x})|\mu)} = \frac{(2\pi)^{-n/2} \exp\left\{-\frac{n}{2}\sum_{i=1}^{n}(x_i-\bar{x}) - \frac{n}{2}(\bar{x}-\mu)^2\right\}}{\left(\frac{2\pi}{n}\right)^{-1/2} \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\}}$$

$$= \frac{(2\pi)^{-n/2}}{\left(\frac{2\pi}{n}\right)^{-1/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}$$

which is a constant function of $\mu$, so by the theorem, $T(\underset{\sim}{x}) = \bar{x}$ is sufficient condition of $\mu$

**Theorem:** let $L(\underset{\sim}{x}|\theta)$ be a likelihood function for a random sample $\underset{\sim}{x}$ and $T(\underset{\sim}{x}) = T$ is a statistic with pmf/pdf $q(t|\theta)$ then $T(\underset{\sim}{x})$ is sufficient for $\theta$ if $\forall\, \underset{\sim}{x} \in \mathcal{X}$, the ratio $\frac{L(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}$ is constant as a function of $\theta$

**proof:**

for each $\underset{\sim}{x} \in \mathcal{X}$, the ratio likelihood $\frac{L(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}$ does not depend on $\theta$

to show the conditional pmf of $\underset{\sim}{X}$ given $T=t$ does not depend on $\theta$ $\forall t \in \text{Range}(T)$

let $t \in \text{Range}(T)$, then conditional pmf $f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x})$ of $\underset{\sim}{X}$ given $T=t$ is

$$f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = P_\theta[\underset{\sim}{X} = \underset{\sim}{x} | T=t]$$

$$= \frac{P_\theta[\underset{\sim}{X} = \underset{\sim}{x}, T(\underset{\sim}{X}) = t]}{P_\theta[T=t]}$$

then $\underset{\sim}{X} = \underset{\sim}{x}$ and $T(\underset{\sim}{X}) = t$ cannot happen together

$$= \begin{cases} 0 & ; \text{ if } T(\underset{\sim}{x}) \neq t \\ \dfrac{P_\theta[\underset{\sim}{X} = \underset{\sim}{x}]}{P_\theta[T=t]} & ; \text{ if } T(\underset{\sim}{x}) = t \end{cases}$$

as $T(\underset{\sim}{x}) = t$ becomes redundant

$$= \begin{cases} 0 & ; T(\underset{\sim}{x}) \neq t \\ \dfrac{L(\underset{\sim}{x}|\theta)}{q(t|\theta)} & ; T(\underset{\sim}{x}) = t \end{cases}$$

$$= \begin{cases} 0 & ; T(\underset{\sim}{x}) \neq t \\ \dfrac{L(\underset{\sim}{x}|\theta)}{q(t|\theta)} & ; T(\underset{\sim}{x}) = t \end{cases}$$

which does not depend on $\theta$, so $T$ is sufficient for $\theta$

**Note:** above is a sufficient condition, not a necessary one, we still need to find $T$

**Exe:** $X_1, X_2, \ldots X_n \overset{iid}{\sim} Poi(\lambda)$, show from definition that same statistic $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i \sim Poi(n\lambda)$ is sufficient statistic for $\lambda$

$x_i \in \{0, 1, 2, \ldots\} = \mathbb{Z}_{\geq 0}$

then $f_{\underset{\sim}{X}|T=t}(\underset{\sim}{x}) = \dfrac{P[X_1 = x_1, X_2 = x_2 \ldots X_n = x_n, T = \sum_{i=1}^{n} x_i]}{P[T=t]}$

well $t = \sum_{i=1}^{n} x_i$ otherwise 0

$$= \frac{P[X_1 = x_1]\ P[X_2 = x_2] \cdots P[X_n = x_n]}{P[T=t]}$$

$$= \frac{e^{-\lambda}\frac{(\lambda)^{x_1}}{x_1!}\ e^{-\lambda}\frac{(\lambda)^{x_2}}{x_2!} \ldots e^{-\lambda}\frac{(\lambda)^{x_n}}{x_n!}}{e^{-n\lambda}\frac{(n\lambda)^t}{t!}}$$

$$= \frac{(x_1 + x_2 + \cdots + x_n)!}{x_1!\, x_2! \ldots x_n!} \frac{1}{(n)^t}$$

so, $T(\underset{\sim}{X})$ is sufficient statistic

10th Feb:

Quiz 2 - friday 13th Feb 3:45pm to 4:15pm LA002 (ordeal statistic)

now we saw given T, we can find if T is sufficient or not, now we want to guess sufficient statistic

**theorm**: (Factorisation Criterion) A statistic $T=T(\underset{\sim}{X})$ is a sufficient statistic for $\theta$ iff $\exists\ g(t|\theta)$ depends on $\theta$ and $h(\underset{\sim}{x})$ does not depend on $\theta$ s.t

$$\forall\ \theta\in\Theta,\ \forall\ \underset{\sim}{x}\in\mathcal{X},\ L(\underset{\sim}{x}|\theta)=g(T(\underset{\sim}{x})|\theta)\,h(\underset{\sim}{x})$$

**proof**: statement true in all cases, but proof in discrete case

($\Rightarrow$) let $T=T(\underset{\sim}{X})$ be sufficient for $\theta$ then $\forall\theta\in\Theta$ and $\forall\underset{\sim}{x}\in\mathcal{X}$ we have

(this does not change probability)

$$L(\underset{\sim}{x}|\theta)=P_\theta(\underset{\sim}{X}=\underset{\sim}{x})=P_\theta(\underset{\sim}{X}=\underset{\sim}{x},T(\underset{\sim}{X})=T(\underset{\sim}{x}))$$
$$=\underbrace{P_\theta(T(\underset{\sim}{X})=T(\underset{\sim}{x}))}_{\text{depends on }\theta}\ \underbrace{P\left[\underset{\sim}{X}=\underset{\sim}{x}\,\middle|\,T(\underset{\sim}{X})=T(\underset{\sim}{x})\right]}_{\text{free of }\theta}$$

(pmf of T)

choose $g(t|\theta)=P_\theta(T(\underset{\sim}{X})=t)\quad t\in\text{Range}(T)$

and $h(\underset{\sim}{x})=P\left[\underset{\sim}{X}=\underset{\sim}{x}\,\middle|\,T(\underset{\sim}{X})=T(\underset{\sim}{x})\right]$

so, $L(\underset{\sim}{x}|\theta)=g(T(\underset{\sim}{x})|\theta)\,h(\underset{\sim}{x})=P_\theta(T(\underset{\sim}{X})=T(\underset{\sim}{x}))\,P\left[\underset{\sim}{X}=\underset{\sim}{x}\,\middle|\,T(\underset{\sim}{X})=T(\underset{\sim}{x})\right]$

$$=P_\theta(\underset{\sim}{X}=\underset{\sim}{x},\ T(\underset{\sim}{X})=T(\underset{\sim}{x}))=P_\theta(\underset{\sim}{X}=\underset{\sim}{x})$$

($\Leftarrow$) Suppose factorisation of $L(\underset{\sim}{x}|\theta)=g(T(\underset{\sim}{x})|\theta)\,h(\underset{\sim}{x})$ exist

then let $q(t|\theta)=$ pmf of $T$

to show T is sufficient for $\theta$, we have to show, $\forall\ \underset{\sim}{x}\in\mathcal{X}$ the ratio:

$$\frac{L(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)}\quad\text{does not depend on }\theta$$

fix $\underset{\sim}{x}\in\mathcal{X}$, then $q(t|\theta)=P_\theta[T(\underset{\sim}{X})=t]$

$$=\sum_{\underset{\sim}{y}:T(\underset{\sim}{y})=t}P_\theta[\underset{\sim}{X}=\underset{\sim}{y}]$$

$$\Rightarrow q(T(\underset{\sim}{x})|\theta)=\sum_{\underset{\sim}{y}:T(\underset{\sim}{y})=T(\underset{\sim}{x})}P_\theta[\underset{\sim}{X}=\underset{\sim}{y}]$$

$$=\sum_{\underset{\sim}{y}:T(\underset{\sim}{y})=T(\underset{\sim}{x})}L(\underset{\sim}{y}|\theta)$$

$$=\sum_{\underset{\sim}{y}:T(\underset{\sim}{y})=T(\underset{\sim}{x})}g(T(\underset{\sim}{y})|\theta)\,h(\underset{\sim}{y})\quad(\because\text{factorisation given above})$$

$$=\sum_{\underset{\sim}{y}:T(\underset{\sim}{y})=T(\underset{\sim}{x})}g(T(\underset{\sim}{x})|\theta)\,h(\underset{\sim}{y})$$

$$=g(T(\underset{\sim}{x})|\theta)\sum_{\underset{\sim}{y}:T(\underset{\sim}{y})=T(\underset{\sim}{x})}h(\underset{\sim}{y})$$

so, $q(T(\underset{\sim}{x})|\theta) = g(T(\underset{\sim}{x})|\theta) \sum_{\underset{\sim}{y}: T(\underset{\sim}{y})=T(\underset{\sim}{x})} h(\underset{\sim}{y})$

now, $\frac{L(\underset{\sim}{x}|\theta)}{q(T(\underset{\sim}{x})|\theta)} = \frac{g(T(\underset{\sim}{x})|\theta)\, h(\underset{\sim}{x})}{q(T(\underset{\sim}{x})|\theta)}$

$= \frac{g(T(\underset{\sim}{x})|\theta)\, h(\underset{\sim}{x})}{g(T(\underset{\sim}{x})|\theta) \sum_{\underset{\sim}{y}: T(\underset{\sim}{y})=T(\underset{\sim}{x})} h(\underset{\sim}{y})}$ ($\because$ Factorization)

$g(T(\underset{\sim}{x})|\theta)$ is non-zero as $T(\underset{\sim}{x}) \in \text{Range}(T)$

so, ratio is free of $\theta$ and so, $T$ is sufficient of $\theta$ ($\because$ previous theorem)

Note: Now with the above result, we can guess the statistic

Exe: Let $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$ (any pmf/pdf), show that

① $T_1(\underset{\sim}{X}) = (X_1, X_2, \ldots X_n)$ is sufficient for $\theta$ → done down

② $T_2(\underset{\sim}{X}) = (X_{(1)}, X_{(2)}, \ldots X_{(n)})$ is also sufficient for $\theta$

Eg: suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Poi}(\lambda)$, now let's try a one dimensional summary which is sufficient

In this case, $\theta = \lambda$ $\Theta = (0, \infty)$ $\mathcal{X} = \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0} \ldots \times \mathbb{Z}_{\geq 0}$

$= (\mathbb{Z}_{\geq 0})^n$

also the likelihood function $L(\underset{\sim}{x}|\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$ if $\underset{\sim}{x} = (x_1, x_2, \ldots x_n) \in \mathcal{X}$

$= e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \frac{1}{\prod_{i=1}^{n} (x_i)!}$ if $\underset{\sim}{x} \in \mathcal{X}$

$= g(T(\underset{\sim}{x})|\lambda)\, h(\underset{\sim}{x})$ $\underset{\sim}{x} \in \mathcal{X}$

where, $\forall\, \underset{\sim}{x} \in \mathcal{X}$, $T(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$

$h(\underset{\sim}{x}) = \frac{1}{\prod_{i=1}^{n} (x_i)!}$ ← free of $\lambda$

$g(t|\lambda) = e^{-n\lambda} \lambda^t$ $\forall\, t \in \text{Range}(T) = \mathbb{N} \cup \{0\}$

So, by factorization criterion, $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is sufficient for $\lambda$

Exe: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{Exp}(\lambda)$ with mean $\frac{1}{\lambda}$, then find a one-dim sufficient statistic for $\lambda$ → done down

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{unif}(0,\theta)$ $\quad \Theta = (0,\infty)$ $\quad \mathcal{X} = (0,\theta)^n$ ← depends on parameter $\theta$

$$\text{Now, } L(\underset{\sim}{x}|\theta) = \begin{cases} \frac{1}{\theta^n} ; \text{ if } x_1, x_2, \dots x_n \in (0,\theta) \\ 0 ; \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n} ; x_{(1)} := \min_{1\le i\le n} x_i > 0 \text{ and } x_{(n)} = \max_{1\le i\le n} x_i < \theta \\ 0 ; \text{otherwise} \end{cases}$$

$$\Rightarrow L(\underset{\sim}{x}|\theta) = \underbrace{\frac{1}{\theta^n} \amalg_{(x_{(n)}<\theta)}}_{\text{depends on } \theta} \underbrace{\amalg_{(x_{(1)}>0)}}_{\text{does not depend on } \theta}$$

$$= g(T(\underset{\sim}{x})|\theta)\, h(\underset{\sim}{x}) \text{ where}$$

$T(\underset{\sim}{x}) = x_{(n)} = \max_{1\le i\le n} x_i$ ← maximum $\quad h(\underset{\sim}{x}) = \amalg_{(x_{(1)}>0)}$ ← minimum

$$g(t|\theta) = \frac{1}{\theta^n} \amalg_{(t<\theta)}$$

So by factorisation criterion, $T(\underset{\sim}{X}) = X_{(n)} = \max_{1\le i\le n} X_i$ is sufficient statistic for $\theta$

Note: Above in unif $(0,\theta)$, the sample space $\mathcal{X}$ depends on $\theta$

Exe: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{unif}(\tau, 0)$ where $\tau < 0$ is a parameter, write down parameter space and sample space, also show that $T(\underset{\sim}{X}) = X_{(1)}$ is sufficient for $\tau$ → done down
$X_{(1)} = \min_{1\le i\le n} X_i$

Exe: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{unif}(\theta_1, \theta_2)$ where $\theta_1 < \theta_2$ are true unknown parameters, write down the parameter space and sample space and show that
$T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is sufficient for $(\theta_1, \theta_2)$ → done down

Exe: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\lambda,\beta)$ where $\lambda, \beta$ are unknown parameters
$\lambda > 0$, $\beta \in \mathbb{R}$ and
$$f(x|\lambda,\beta) = \begin{cases} \lambda \exp\{-\lambda(x-\beta)\} ; x > \beta \\ 0 ; x \le \beta \end{cases}$$ → done down

Shifted Exponential density

(a) write down the parameter space and sample space

(b) show that $T(\underset{\sim}{X}) = (\min_{1\le i\le n} X_i, \sum_{i=1}^{n} X_i)$ is sufficient for $(\beta, \lambda)$

Exe: Let $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (any pmf/pdf), show that

① $T_1(\underset{\sim}{X}) = (X_1, X_2, \dots X_n)$ is sufficient for

② $T_2(\underset{\sim}{X}) = (X_{(1)}, X_{(2)}, \ldots X_{(n)})$ is also sufficient for $\theta$

Ans: $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$

$T_1(\underset{\sim}{X}) = (X_1, X_2, \ldots X_n)$

then $L(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

$T_1(\underset{\sim}{x}) = (x_1, x_2, \ldots x_n)$

let $g((x_1, x_2, \ldots x_n)|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

$h((x_1, x_2, \ldots x_n)) = 1$

then, $g(T_1(\underset{\sim}{x})|\theta) = \prod_{i=1}^{n} f(x_i|\theta)\ h(\underset{\sim}{x}) = 1$

$\exists\ g(T_1(\underset{\sim}{x})|\theta),\ h(\underset{\sim}{x})$ s.t $L(\underset{\sim}{x}|\theta) = g(T(\underset{\sim}{x})|\theta)h(\underset{\sim}{x})$

$\forall\ \underset{\sim}{x} \in \mathcal{X}, \forall \theta \in \Theta$ (∵ arbitrary)

now, as $g((x_1, x_2, \ldots x_n)|\theta)$

$= \prod_{i=1}^{n} f(x_i|\theta)$

$= \prod_{i=1}^{n} f(x_{(i)}|\theta)$

$= g((x_{(1)}, x_{(2)}, \ldots x_{(n)})|\theta)$

$= g(L_2(\underset{\sim}{x})|\theta)$

So, $g(L_1(\underset{\sim}{x})|\theta) = g(L_2(\underset{\sim}{x})|\theta)$, i.e both $L_1, L_2$ sufficient

(∵ above theorem)

Exa: Suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} Exp(\lambda)$ with mean $\frac{1}{\lambda}$, then find a one-dim sufficient statistic for $\lambda$

Ans:

$\theta = \lambda,\ \Theta = (0, \infty)$ Range$(Exp(\lambda)) = (0, \infty)$

$\mathcal{X} = (0, \infty)^n$

now $L(\underset{\sim}{x}|\lambda) = \prod_{i=1}^{n} f(x_i|\theta)$ $\quad \forall \underset{\sim}{x} \in \mathcal{X},\ \lambda \in \Theta$

$= \prod_{i=1}^{n} (\lambda e^{-\lambda x_i})$

$= \lambda^n e^{-\lambda(\sum_{i=1}^{n} x_i)}$

let $T(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$ then

$\forall \underset{\sim}{x} \in \mathcal{X}, \lambda \in \Theta$

$L(\underset{\sim}{x}|\lambda) = \lambda^n e^{-\lambda T(\underset{\sim}{x})}$

$= g(T(\underset{\sim}{x})|\lambda)\ h(\underset{\sim}{x})$

$g(t|\lambda) = \lambda^n e^{-\lambda t}$ $\quad h(\underset{\sim}{x}) = 1$

$\forall \underset{\sim}{x} \in \mathcal{X}, \lambda \in \Theta, \exists T(\underset{\sim}{x}) = \sum_{i=1}^{n} x_i$ s.t

$L(\underset{\sim}{x}|\lambda) = g(T(\underset{\sim}{x})|\lambda)h(\underset{\sim}{x})$ so, $T(\underset{\sim}{x})$ is sufficient

**Exe:** suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{unif}(\tau, 0)$ where $\tau < 0$ is a parameter, write down parameter space and sample space, also show that $T(\underset{\sim}{X}) = X_{(1)}$

$X_{(1)} = \min_{1 \le i \le n} X_i$ is sufficient for $\tau$

**Ans:** $\mathcal{X} = (\tau, 0)^n$ $\Theta = \tau \in (-\infty, 0)$ then

$$L(\underset{\sim}{x}|\theta) = \begin{cases} \frac{1}{(-\tau)^n} & ; x_{(n)} < 0, x_{(1)} > \tau \\ 0 & ; \text{otherwise} \end{cases}$$

so, $L(\underset{\sim}{x}|\theta) = \frac{1}{(-\tau)^n} \mathbb{I}_{(x_{(n)} < 0)} \mathbb{I}_{(x_{(1)} > \tau)}$

i.e $\forall \underset{\sim}{x} \in (\tau, 0)^n, \tau \in (-\infty, 0)$

$T(\underset{\sim}{x}) = x_{(1)}$ then

$\exists g(T(\underset{\sim}{x})|\tau) = \frac{1}{(-\tau)^n} \mathbb{I}_{x_{(1)} > \tau}$

$h(\underset{\sim}{x}) = \mathbb{I}_{x_{(n)} < 0}$

s.t $L(\underset{\sim}{x}|\tau) = g(T(\underset{\sim}{x})|\tau) h(\underset{\sim}{x})$

so, $T$ is sufficient

**Exe:** suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{unif}(\theta_1, \theta_2)$ where $\theta_1 < \theta_2$ are the unknown parameters, write down the parameter space and sample space and show that

$T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is sufficient for $(\theta_1, \theta_2)$

**Ans:**

$X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{unif}(\theta_1, \theta_2)$

$\theta = (\theta_1, \theta_2)$ $\Theta = \{(x, y) \mid x < y, x, y \in \mathbb{R}\}$

$\underset{\sim}{x} \in \mathcal{X} = (\theta_1, \theta_2)^n$

now, $T(\underset{\sim}{x}) = (X_{(1)}, X_{(n)})$ then

$$L(\underset{\sim}{x}|\theta) = \begin{cases} \frac{1}{(\theta_2 - \theta_1)^n} & ; x_{(1)} > \theta_1, x_{(n)} < \theta_2 \\ 0 & ; \end{cases}$$

then $g(T(\underset{\sim}{x})|\theta) = \frac{1}{(\theta_2 - \theta_1)^n} \mathbb{I}_{x_{(1)} > \theta_1} \mathbb{I}_{x_{(n)} < \theta_2}$

$h(\underset{\sim}{x}) = 1$

s.t $\forall \theta \in \Theta, \underset{\sim}{x} \in \mathcal{X}$

$L(\underset{\sim}{x}|\theta) = g(T(\underset{\sim}{x})|\theta) h(\underset{\sim}{x})$

so, $T$ is sufficient

**Exe:** suppose $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\lambda, \beta)$ where $\lambda, \beta$ are unknown parameters $\lambda > 0$, $\beta \in \mathbb{R}$ and

$$f(x|\lambda,\beta) = \begin{cases} \lambda \exp\{-\lambda(x-\beta)\} & ; x > \beta \\ 0 & ; x \le \beta \end{cases}$$

Shifted exponential density

(a) write down the parameter space and sample space

(b) show that $T(\underset{\sim}{X}) = (\min_{1\le i\le n} X_i, \sum_{i=1}^{n} X_i)$ is sufficient for $(\beta, \lambda)$

Ans:

$X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\lambda,\beta)$

$(\lambda,\beta) \in \Theta = (0,\infty) \times \mathbb{R}$

Range$(X_i) = (\beta,\infty) \Rightarrow \mathcal{X} = (\beta,\infty)^n$

then $L(\underset{\sim}{x}|\theta) = \begin{cases} \prod_{i=1}^{n} \lambda e^{-\lambda(x_i-\beta)} & ; x_i > \beta \ \forall i \in \{1,2,\ldots n\} \\ 0 & ; \text{otherwise} \end{cases}$

$= \lambda^n e^{-\lambda(\sum x_i - n\beta)} \mathbb{I}_{x_{(1)} > \beta}$

$T(\underset{\sim}{x}) = (x_{(1)}, \sum x_i)$

then $\forall \underset{\sim}{x} \in \mathcal{X}, (\lambda,\beta) = \theta \in \Theta$

$\exists\, g(t_1,t_2|\lambda,\beta) = \lambda^n e^{-\lambda(t_2-n\beta)} \mathbb{I}_{t_1>\beta}$

$h(\underset{\sim}{x}) = 1$

s.t $L(\underset{\sim}{x}|\theta) = g(T(\underset{\sim}{x})|\lambda\beta)\, h(\underset{\sim}{x})$

So, T is sufficient for $(\lambda,\beta)$ (same for $\beta,\lambda$ trivial)

but $(\beta,\lambda) = \mathbb{R} \times (0,\infty) = \Theta$

17th Feb:

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ $\sigma \in (0,\infty)$ are both unknown parameters, in this case

$\theta = (\mu, \sigma)$ $\Theta = \mathbb{R} \times (0,\infty)$ also $\mathcal{X} = \mathbb{R}^n$

$$L(\underset{\sim}{x} | \mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\} \quad \underset{\sim}{x} \in \mathcal{X}$$

$$= (2\pi)^{-n/2} \sigma^{-n} \exp\left\{ -\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{2\sigma^2} \right\} \times \exp\left\{ -\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right\}$$

$$= (2\pi)^{-n/2} \sigma^{-n} \exp\left\{ -\frac{(n-1)s^2}{2\sigma^2} \right\} \exp\left\{ -\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right\}$$

$$= h(\underset{\sim}{x})\, g(T(\underset{\sim}{x}) | \mu, \sigma)$$

$$T(\underset{\sim}{x}) = (\bar{x}, s^2) = \left( \frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \right)$$

$$h(\underset{\sim}{x}) = (2\pi)^{-n/2}$$

$$g((t_1, t_2) | \mu, \sigma) = \sigma^{-n} \exp\left\{ -\frac{n(t_1 - \mu)}{2\sigma^2} \right\} \exp\left\{ -\frac{(n-1)t_2}{2\sigma^2} \right\}$$

by factorisation criterion,

$T(\underset{\sim}{X}) = (\bar{X}, S^2)$ is sufficient for $(\mu, \sigma^2)$

Exe: show that $\left( \sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2 \right)$ pair is sufficient for $(\mu, \sigma^2)$

for $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ → done down

Eg: $X_1, X_2, \dots X_n \overset{iid}{\sim} \text{unif}(\theta, \theta+1)$ where $\theta \in \mathbb{R}$ is unknown parameter

$\Theta = \mathbb{R}$, $\mathcal{X} = (\theta, \theta+1)^n$ (depends on $\theta$) and likelihood

$$L(\underset{\sim}{x} | \theta) = \begin{cases} 1 & ; \theta < x_1, x_2, \dots x_n < \theta + 1 \\ 0 & ; \text{otherwise} \end{cases}$$

$$= \begin{cases} 1 & ; x_{(1)} > \theta, \; x_{(n)} < \theta + 1 \\ 0 & ; \text{otherwise} \end{cases}$$

$$= \begin{cases} 1 & ; x_{(n)} - 1 < \theta < x_{(1)} \\ 0 & ; \text{otherwise} \end{cases}$$

$$= \mathbb{I}_{(x_{(n)} - 1 < \theta < x_{(1)})}$$

$$= g(T(\underset{\sim}{x}) | \theta)\, h(\underset{\sim}{x})$$

where $T(\underset{\sim}{x}) = (x_{(1)}, x_{(n)})$ $h(\underset{\sim}{x}) = 1$

$$g((t_1, t_2) | \theta) = \mathbb{I}_{t_1 - 1 < \theta < t_2}$$

so, by factorization criterion

$T(\underset{\sim}{X}) = (X_{(1)}, X_{(n)})$ is sufficient for $\theta$

Note: above we see a sufficient statistic with more degree than no of variables

## Exponential family of univariate distributions:

suppose we have $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots \theta_n)$ a vector of parameters taking values in $\Theta \subseteq \mathbb{R}^d$, $\Theta$ is called the parameter space

a family of univariate pmf/pdf $f(x|\theta)$ parametrised by $\theta \in \Theta$ is said to belong to Exponential family if $f$ is of the form

$$E = f(x|\underset{\sim}{\theta}) = \begin{cases} h(x) \exp\left\{ \sum_{i=1}^{k} \eta_i(\underset{\sim}{\theta}) T_i(x) - \psi(\underset{\sim}{\theta}) \right\} ; & x \in \mathcal{R} \subseteq \mathbb{R} \\ 0 \quad ; \text{otherwise} \end{cases}$$

(Range)

where $\mathcal{R} \subseteq \mathbb{R}$ does not depend on $\underset{\sim}{\theta}$

$\mathcal{R} (\subseteq \mathbb{R}) = \{x \in \mathbb{R} \mid f(x|\underset{\sim}{\theta}) > 0\}$ does not depend on $\theta$

$k \in \mathbb{Z}$, $k \geq d$, $h : \mathcal{R} \to (0, \infty)$

$\psi : \Theta \to \mathbb{R}$, $\eta_i : \Theta \to \mathbb{R}$ $\quad i = 1, 2, \ldots k$

$T_i : \mathcal{R} \to \mathbb{R}$ $\quad i = 1, 2 \ldots k$

} functions

functions are NOT unique

Note: ① $\mathcal{R} = \text{Range}(X)$, $X \sim f(x|\underset{\sim}{\theta})$, it is very crucial that the range $\mathcal{R}$ does not depend on $\underset{\sim}{\theta} = (\theta_1, \ldots \theta_n)$

② in many cases, it is possible to reparametrise family using $\eta_1, \eta_2, \ldots \eta_k$ as parameters, this is called the natural parametrisation of family and $\underset{\sim}{\eta} = (\eta_1, \eta_2, \ldots \eta_k)$ is called natural parameter

③ In many cases, $k = d$

Eg: suppose $X \sim \text{Bin}(n, p)$ where $n$ is known and $p$ is unknown parameter then $X$ has pmf

$$f(x|p) = \begin{cases} \binom{n}{x} p^x (1-p)^x ; & 0 \leq x \leq n \quad x \in \mathbb{Z} \\ 0 \quad ; \text{otherwise} \end{cases}$$

$$= \begin{cases} \binom{n}{x} \exp\left\{ x \log \frac{p}{1-p} + n \log(1-p) \right\} ; & x \in \{0, 1 \ldots n\} \\ 0 \quad ; \text{otherwise} \end{cases}$$

taking $d = k = 1$, $\mathcal{R} = \{0, 1, 2, \ldots n\}$ $\theta_1 = p$, $\Theta = (0, 1)$ $\eta_1(p) = \log \frac{p}{1-p}$

$T_1(x) = x$

$\psi(p) = -n \log(1-p)$

$h(x) = \binom{n}{x}$ we get that pmf of Bin(n,p) distribution belongs to one parameter exponential family

$n=1 \Rightarrow$ Bern(p) also one parameter Exponential family

Eg: Suppose $X \overset{iid}{\sim} Exp(\lambda)$ then X has pdf $f(x|\lambda) = \lambda e^{-\lambda x}, x>0$

$= \exp\{\lambda(-x) + \log\lambda\}$ $x>0$

taking $d=k=1$, $\mathcal{X}=(0,\infty)$, $\theta_1=\lambda$, $\Theta=(0,\infty)$

$\eta_1(\lambda)=\lambda \quad T_1(x)=-x \quad \psi(\lambda)=-\log\lambda \quad h(x)=1$

we see that pdf of exponential distribution is of form E
so, it belongs to the exponential family

Exe: Show that the Poi(λ) dist$^n$ belongs to the exponential family → done down

Eg: $X \sim N(\mu,\sigma^2)$ → Both unknown $\quad \underset{\sim}{\theta} = (\theta_1,\theta_2)=(\mu,\sigma)$

$\Theta = \mathbb{R}\times(0,\infty)$

also, pdf is $f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad x\in\mathbb{R}$

$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\left[\frac{x^2+\mu^2-2\mu x}{2\sigma^2}+\log\sigma\right]\right\}$

$= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2\sigma^2}+\frac{\mu x}{\sigma^2}-\frac{\mu^2}{2\sigma^2}-\log\sigma\right\}$ $x\in\mathbb{R}$

$d=2=k$, $\mathcal{X}=\mathbb{R}$, $\eta_1(\mu,\sigma)=\frac{1}{2\sigma^2}$ $\eta_2(\mu,\sigma)=\frac{\mu}{\sigma^2}$

$T_1(x)=-x^2$ $T_2(x)=x$ $\psi(\mu,\sigma)=\frac{\mu^2}{2\sigma^2}+\log\sigma$ $h(x)=\frac{1}{\sqrt{2\pi}}$

We observe that $N(\mu,\sigma^2)$ dist$^n$ belongs to Exponential family

Exe: Show that gamma(α,λ) dist$^n$ (with $\theta=(\alpha,\lambda)$) belongs to exponential family → done down

Exe: Show that Beta(a,b) dist$^n$ (with $\theta=(a,b)$) belongs to the exponential family → done down

Exe: Show that negative binomial(r,p) dist$^n$ with r known, p unknown belongs to the exponential family → done down

Exe: Show that $N(\mu,\mu^2)$ μ is unknown belongs to the exponential family (in this case $k=2>1=d$) → done down

Note: $(\mu,\mu^2)$ are called curve parameters

Theorem: Let $X_1,X_2,\dots X_n \overset{iid}{\sim} f(x|\underset{\sim}{\theta})$ (pmf/pdf) where $\underset{\sim}{\theta}=(\theta_1,\dots\theta_k)$ is of form (E), where notations are as before (in particular $k\geq d$) then $T(\underset{\sim}{x}) = \left(\sum_{j=1}^n T_1(X_j), \sum_{j=1}^n T_2(X_j), \dots \sum_{j=1}^n T_k(X_j)\right)$

then $T(\underset{\sim}{X})$ is sufficient statistic for $\underset{\sim}{\theta}$, dim is $k\geq d$

Note: dimension of the sufficient statistic (=k) can be strictly bigger than d=dim of parameter like $(\mu,\mu^2)$ from exercise

Ex: show that $(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$ pair is sufficient for $(\mu, \sigma^2)$

for $x_1, x_2, \ldots x_n \overset{iid}{\sim} N(\mu, \sigma^2)$

Ans:

$$L(x_1, x_2, \ldots x_n \mid \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)}$$

$$= (2\pi)^{-n/2}\sigma^{-n} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n}\mu^2 - 2\mu\sum_{i=1}^{n} x_i\right)}$$

$$= (2\pi)^{-n/2}\sigma^{-n} e^{-n\mu^2/2\sigma^2} e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i\right)}$$

$$h(\underset{\sim}{x}) = (2\pi)^{-n/2}$$

$$g(t_1, t_2 \mid \mu, \sigma^2) = \sigma^{-n} e^{-n\mu/2\sigma^2} e^{\frac{-1}{2\sigma^2}(t_2 - 2\mu t_1)}$$

then $L(\underset{\sim}{x} \mid \mu, \sigma^2) = h(\underset{\sim}{x})\, g(T(\underset{\sim}{x}) \mid \mu, \sigma^2)$

$$T(\underset{\sim}{x}) = \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$$

by factorisation criterion it is sufficient

Ex: show that the poi($\lambda$) distⁿ belongs to the exponential family

Ans:

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

then $f(x \mid \lambda) = e^{-\lambda}\frac{\lambda^x}{(x)!}$ $x \in \{0, 1, 2, \ldots\} = S$

$\Theta = (0, \infty)$ $\lambda \in (0, \infty)$

$$f(x \mid \lambda) = \frac{1}{x!} e^{-\lambda} e^{\ln(\lambda^x)}$$

$$= \frac{1}{x!} e^{x\ln\lambda - \lambda}$$

$\psi(\lambda) = \lambda$

$k = 1$, $d = 1$ $\eta_1(\lambda) = \ln\lambda$ $T_1(x) = x$

$h(x) = \frac{1}{x!}$

Ex: show that gamma($\alpha, \lambda$) distⁿ (with $\theta = (\alpha, \lambda)$) belongs to exponential family

Ans:

$\theta = (\alpha, \lambda)$ $f(x \mid \alpha, \lambda) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} = e^{\alpha\ln\lambda + (\alpha-1)\ln x - \lambda x - \ln\Gamma(\alpha)}$

$(\alpha, \lambda) \in (0, \infty) \times (0, \infty) = \Theta$

$x \in (0, \infty) = S$

$\forall x \in S$

$h(x) = 1$ $\psi(\alpha, \lambda) = \ln\Gamma(\alpha) - \alpha\ln\lambda$

$k = 2 = d$

$\eta_1(\alpha, \lambda) = \alpha - 1$ $\eta_2(\alpha, \lambda) = -\lambda$

$T_1(x) = \ln(x)$ $T_2(x) = x$

Exe: Show that Beta(a,b) distⁿ (with $\theta=(a,b)$) belongs to the exponential family

Ans: $a,b \in (0,\infty)$

$\Theta = (0,\infty)\times(0,\infty)$ $\quad \theta=(a,b)\in\Theta$

$x\in(0,1)=\mathcal{X}$ $\quad f(x|a,b)=\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$

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

$= e^{(a-1)\ln x+(b-1)\ln(1-x)-\ln(B(a,b))}$

$h(x)=1$ $\quad \psi(\theta)=-\ln(B(a,b))$

$\eta_1(a,b)=a-1$ $\quad T_1(x)=\ln x$

$\eta_2(a,b)=b-1$ $\quad T_2(x)=\ln(1-x)$

Exe: Show that negative binomial $(r,p)$ distⁿ with $r$ known, $p$ unknown belongs to the exponential family

Ans:

$f(x|p)=\binom{x+r-1}{x}(1-p)^x p^r = \binom{x+r-1}{r} e^{x\ln(1-p)+r\ln(p)}$

$p\in(0,1)$

$r$: no of success till experiment is stopped $\quad h(x)=\binom{x+r-1}{r}$

$x=0,1,2,\ldots$ no of failures

last success

$\binom{x+r-1}{x}$ = failures $\quad \psi(p)=-r\ln(p)$

$(1-p)^x$ = probability of failure $\quad \eta_1(p)=\ln(1-p)$

$\Theta=(0,1)$ $\quad T_1(x)=x$

Exe: Show that $N(\mu,\mu^2)$ $\mu$ is unknown belongs to the exponential family (in this case $k=2>1=d$)

Ans:

$N(\mu,\mu^2)$ $\quad \mu\in\mathbb{R}$, $\Theta=\mathbb{R}$ $\quad \mathcal{X}=\mathbb{R}$

$f(x|\mu)=\frac{1}{\sqrt{2\pi}\mu^2}e^{\frac{-(x-\mu)^2}{2\mu^2}}$

$=\frac{1}{\sqrt{2\pi}}\frac{1}{|\mu|}e^{-\frac{(x-\mu)^2}{2\mu^2}}$

$=\frac{1}{\sqrt{2\pi}}e^{\ln\left(\frac{1}{|\mu|}\right)-\frac{(x-\mu)^2}{2\mu^2}}$

$-\ln(|\mu|)-\frac{(x^2+\mu^2-2\mu x)}{2\mu^2}$

$=-\frac{x^2}{2\mu^2}-\frac{1}{2}+\frac{x}{\mu}-\ln(|\mu|)$

$h(x)=\frac{1}{\sqrt{2\pi}}$ $\quad \psi(\mu)=\frac{1}{2}+\ln(|\mu|)$ $\quad \eta_1(\mu)=\frac{-1}{2\mu^2}$ $\quad T_1(x)=x^2$

$\eta_2(\mu)=\frac{1}{\mu}$ $\quad T_2(x)=x$

20th Feb:

midsem is 27th Feb (8:00 - 10:00 am) LA001, till todays class syllabus
3 questions, 10 marks each

theorem: let $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf) where $\theta = (\theta_1 .. \theta_k)$ is of form (E), where notations are as before (in particular $K \geq d$) then $T(\underset{\sim}{x}) = \left( \sum_{j=1}^{n} T_1(X_j), \sum_{j=1}^{n} T_2(X_j), \dots \sum_{j=1}^{n} T_K(X_j) \right)$

then $T(\underset{\sim}{X})$ is sufficient statistic for $\underset{\sim}{\theta}$, also $K \geq d$

proof:

$$L(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{j=1}^{n} f(x_j|\underset{\sim}{\theta}) \text{ as form (E)} \quad x \in \mathcal{X}^n$$

$$= \prod_{j=1}^{n} \left[ h(x_j) \exp\left\{ \sum_{i=1}^{K} \eta_i(\underset{\sim}{\theta}) T_i(x_j) - \psi(\underset{\sim}{\theta}) \right\} \right]$$

$$= \left( \prod_{j=1}^{n} h(x_j) \right) \exp\left\{ \sum_{j=1}^{n} \left( \sum_{i=1}^{K} \eta_i(\underset{\sim}{\theta}) T_i(x_j) - \psi(\underset{\sim}{\theta}) \right) \right\}$$

$$= \left( \prod_{j=1}^{n} h(x_j) \right) \exp\left\{ \sum_{i=1}^{K} \left( \sum_{j=1}^{n} \eta_i(\underset{\sim}{\theta}) T_i(x_j) \right) - n\psi(\underset{\sim}{\theta}) \right\}$$

$$= \left( \prod_{j=1}^{n} h(x_j) \right) \exp\left\{ \sum_{i=1}^{K} \eta_i(\underset{\sim}{\theta}) \sum_{j=1}^{n} T_i(x_j) \right\} \exp\{-n\psi(\underset{\sim}{\theta})\}$$

$$= H(\underset{\sim}{x}) g(T(\underset{\sim}{x})|\underset{\sim}{\theta})$$

where $H(\underset{\sim}{x}) = \prod_{j=1}^{n} h(x_j)$

$$T(\underset{\sim}{x}) = \left( \sum_{j=1}^{n} T_1(x_j), \sum_{j=1}^{n} T_2(x_j), \dots \sum_{j=1}^{n} T_K(x_j) \right)$$

$$g((t_1, t_2, \dots t_K)|\underset{\sim}{\theta}) = \exp\left\{ \sum_{i=1}^{K} t_i \eta_i(\underset{\sim}{\theta}) \right\} \exp\{-n\psi(\underset{\sim}{\theta})\}$$

so, by factorisation criterion,

$T(\underset{\sim}{x}) = \left( \sum_{j=1}^{n} T_1(x_j), \sum_{j=1}^{n} T_2(x_j), \dots \sum_{j=1}^{n} T_K(x_j) \right)$ is sufficient for $\underset{\sim}{\theta}$

Note: above condition gives a sufficient statistic for a large family of random variables (N, Ber, Bin, E)

Ex: suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \mu^2)$ then using above result, find a two dimensional ($K=2$) sufficient statistic for $\mu$ → done above

Note: in almost all cases, we will get above T to be minimal sufficient statistic (theorem later)

Minimal sufficiency:

suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$ (pmf/pdf)
($\theta$ can be a vector also)

and $T = T(X)$ is a sufficient statistic for $\theta$
suppose $T^* = T^*(X)$ is another statistic s.t. T is a function of $T^*$

i.e. $\exists$ a function $\varphi$: Range($T^*$) → Range($T$)

s.t
$$T(\underset{\sim}{x}) = \varphi(T^*(\underset{\sim}{x})) \quad \forall\ \underset{\sim}{x} \in \mathcal{X}$$

**Note:** Above is equivalent to saying $T^*(\underset{\sim}{x}) = T^*(\underset{\sim}{y})$ for $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$
$$\Rightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{y})$$

**Lemma:** In above setup, $T^*$ is also sufficient for $\theta$

**proof:**
$T$ is given to be sufficient for $\theta$

$\Rightarrow$ by factorisation criterion the likelihood $L(\underset{\sim}{x}|\theta) = g(T(\underset{\sim}{x})|\theta)\, h(\underset{\sim}{x})\ \forall\ \underset{\sim}{x} \in \mathcal{X}$

$$\Rightarrow L(\underset{\sim}{x}|\theta) = g(\varphi(T^*(\underset{\sim}{x}))|\theta)\, h(\underset{\sim}{x}) \quad \forall \underset{\sim}{x} \in \mathcal{X}$$
$$= g^*(T^*(\underset{\sim}{x})|\theta) h(\underset{\sim}{x})$$
$$g^*(T^*(\underset{\sim}{x})|\theta) = g(\varphi(T^*(\underset{\sim}{x}))|\theta)$$

$\Rightarrow$ $T^*$ is sufficient statistic by factorisation criterion

**Note:** The above proof uses both if and only if part of factorisation criterion

**Eg:** $X_1, X_2, \ldots X_{100} \overset{iid}{\sim} Poi(\lambda)$ then we already know that
$$T(\underset{\sim}{X}) = \sum_{i=1}^{100} X_i \text{ is sufficient for } \lambda$$
In this case, $T^*(\underset{\sim}{X}) = \left(\sum_{i=1}^{34} X_i, \sum_{i=35}^{100} X_i\right)$ as
$$\varphi(t_1, t_2) = t_1 + t_2 \Rightarrow \varphi(T^*(\underset{\sim}{X})) = T(\underset{\sim}{X})$$
so, $T^*(\underset{\sim}{X})$ is also sufficient for $T^*$
so, $T$ can be recovered / calculated from $T^*$ and so, $T^*$ is also sufficient

**defn:** A sufficient statistic, $T = T(\underset{\sim}{X})$ is called a minimal sufficient statistic for $\theta$ if for any other sufficient statistic $T' = T'(\underset{\sim}{X})$, $T$ is a function of $T'$

i.e $\exists$ a function $\varphi$: Range$(T') \longrightarrow$ Range$(T)$ s.t $T(\underset{\sim}{x}) = \varphi(T'(\underset{\sim}{x}))$ $\forall\ \underset{\sim}{x} \in \mathcal{X}$ (i.e $T'(\underset{\sim}{x}) = T'(\underset{\sim}{y}) \Rightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{y})$)

**Note:** In above example, as $\exists\ \varphi$ s.t $\varphi(T^*(\underset{\sim}{X})) = T(\underset{\sim}{X})$
from definition $\Rightarrow T^*(\underset{\sim}{X})$ is not minimal

**Note:** A minimal sufficient statistic is expected to have attained "the best possible reduction of the data"

**Note:** Minimal sufficient statistic may not exist

Now, we need a way to verify that a given statistic is minimal or not and we also want to know a way to directly guess a minimal statistic

**theorm**: Suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} f(x|\theta)$, let $L(\underset{\sim}{x}|\theta)$ be the likelyhood function, suppose

$T = T(\underset{\sim}{X})$ is a statistic s.t the following condition holds:

① for two sample points $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$, (any two points)

$$\{\theta \in \Theta : L(\underset{\sim}{x}|\theta) > 0\} = \{\theta \in \Theta : L(\underset{\sim}{y}|\theta) > 0\}$$

② for any two sample points $\underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}$ the likelyhood ratio

$$\frac{L(\underset{\sim}{x}|\theta)}{L(\underset{\sim}{y}|\theta)} \text{ is constant as a function of } \theta \text{ iff}$$

$$T(\underset{\sim}{x}) = T(\underset{\sim}{y})$$

then $T = T(\underset{\sim}{X})$ is a minimal sufficient statistic

**Note**: when $\mathcal{X}$ does not depend on value of $\theta$, then (1) is true if $\mathcal{X}$ depends on value of $\theta$, then (1) needs to be checked

**Eg**: $X_1, X_2, \dots X_n \overset{iid}{\sim} Poi(\lambda)$, in this case $\theta = \lambda$ $\Theta = (0, \infty)$

$\mathcal{X} = (\mathbb{N} \cup \{0\})^n$, also the likelyhood function

$$L(\underset{\sim}{x}|\lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{(x_i)!} \quad ; x_1, x_2 \dots x_n \in \mathbb{N} \cup \{0\}$$

$$\Rightarrow L(\underset{\sim}{x}|\lambda) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} (x_i)!} \quad ; \underset{\sim}{x} \in \mathcal{X} \text{ for all } \lambda \in (0, \infty)$$

$> 0$

$$\Rightarrow \forall \underset{\sim}{x} \in \mathcal{X}, \ \{\lambda \in (0,\infty) \mid L(\underset{\sim}{x}|\lambda) > 0\} = (0, \infty)$$

$$\Rightarrow \forall \underset{\sim}{x}, \underset{\sim}{y} \in \mathcal{X}, \ \{\lambda \in (0,\infty) \mid L(\underset{\sim}{x}|\lambda) > 0\} = \{\lambda \in (0,\infty) \mid L(\underset{\sim}{y}|\lambda) > 0\}$$

$$= (0, \infty)$$

$\Rightarrow$ ① is trivially satisfied

now take $\underset{\sim}{x}, \underset{\sim}{y} \in (\mathbb{N} \cup \{0\})^n$ then likelyhood ratio:

$$\frac{L(\underset{\sim}{x}|\lambda)}{L(\underset{\sim}{y}|\lambda)} = \left(\frac{e^{-n\lambda} \lambda^{\sum x_i}}{\prod (x_i)!}\right)\left(\frac{\prod (y_i)!}{e^{-n\lambda} \lambda^{\sum y_i}}\right)$$

$$= \frac{\prod (y_i)!}{\prod (x_i)!} \frac{\lambda^{\sum x_i}}{\lambda^{\sum y_i}}$$

will be constant function of $\lambda$

$$\Leftrightarrow \lambda^{\sum x_i} = \lambda^{\sum y_i}$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

so, by theorm, it follows that $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is minimal sufficient

above theorem proof in Casella and Berger

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} Bor(p)$ where $p \in (0,1)$ is the unknown parameter

show that $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for $p$ → deadline

Exe: suppose $X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \mu^2)$ then using above result, find a two dimensional ($k=2$) sufficient statistic for $\mu$

Ans:

$X_1, X_2, \dots X_n \overset{iid}{\sim} N(\mu, \mu^2)$

$\mu \in \mathbb{R}, x \in \mathbb{R}$ $\Theta = \mathbb{R} \setminus \{0\} = \mathbb{R}$ $\mu \neq 0$

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}|\mu|} e^{\frac{-(x-\mu)^2}{2\mu^2}} \quad \forall x \in \mathbb{R}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\ln|\mu| - \frac{(x^2+\mu^2-2x\mu)}{2\mu^2}}$$

$$= -\ln|\mu| - \frac{x^2}{2\mu^2} - \frac{1}{2} + \frac{2x}{\mu}$$

$\eta_1(\mu) = \frac{2}{\mu}$ $T_1(x) = x$

$\eta_2(\mu) = \frac{-1}{2\mu^2}$ $T_2(x) = x^2$ $\psi(\mu) = \ln|\mu| + \frac{1}{2}$

$$\Rightarrow T(\underset{\sim}{X}) = \left( \sum_{i=1}^{n} T_1(x_i), \sum_{i=1}^{n} T_2(x_i) \right)$$

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

by theorem this is sufficient

Exe: $X_1, X_2, \dots X_n \overset{iid}{\sim} Bor(p)$ where $p \in (0,1)$ is the unknown parameter

show that $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for $p$

Ans: $X \sim Bern(p)$

then

$$f(x|p) = (1-p)^{1-x} (p)^x$$

$x=1 \to p$

$x=0 \to 1-p$

$x \in \{0,1\} = \mathcal{X}$

$\Theta = (0,1) \ni p$

then,

$R = \{0,1\}^n$

$\forall \underset{\sim}{x}, \underset{\sim}{y} \in \{0,1\}^n$ $L(\underset{\sim}{x}|p) = \prod_{i=1}^{n} (1-p)^{1-x_i} (p)^{x_i}$

$= (1-p)^{n - \sum x_i} (p)^{\sum x_i}$

$$\left\{ p \in \Theta \,\middle|\, L(\underset{\sim}{x}|p) > 0 \right\} = (0,1)$$

$\forall \underset{\sim}{x} \in \{0,1\}^n$

so, ① satisfied

and, $\frac{L(\underset{\sim}{y}|p)}{L(\underset{\sim}{x}|p)} = (1-p)^{\sum y_i - \sum x_i} (p)^{\sum x_i - \sum y_i}$ const w.r.t $p$

$\Leftrightarrow \sum x_i = \sum y_i \Leftrightarrow T(\underset{\sim}{x}) = T(\underset{\sim}{y})$

so, ① is satisfied and then

$T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i$ is minimal-sufficient statistic for Bern(P)