This document contain's solution's to Exercises 6.1 to 6.9 from Chapter 6, of Statistical inference by Casella and Berger, focusing on sufficient statistic, minimal sufficient statistic.

Some problem's are still in progress, and the document will be updated with their solution's soon.

6.1 Let X be one observation from a $N(0, \sigma^2)$ population. Is $|X|$ a sufficient statistic?

Sol^n

$$X \sim N(0, \sigma^2)$$

$$f(x|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2} x^2\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2} |x|^2\right)$$

$$= \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2} (T(x))^2\right)$$

$$= g(T(x)|\sigma^2) \cdot h(x)$$

where $h(x) = 1$ & the above entire function is $g(T(x)|\sigma^2)$

$\Rightarrow$ $|x|$ is sufficient statistic

(6.2) Let $X_1, X_2, \ldots X_n$ be independent r.v with densities

$$f_{X_i}(x|\theta) = \begin{cases} e^{i\theta - x} & x \geq i\theta \\ 0 & x < i\theta \end{cases}$$

Prove that $T = \min_i \left(\frac{X_i}{i}\right)$

is a sufficient statistic for θ.

Sol^n

$$f_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots x_n | \theta)$$

$$= \begin{cases} e^{\theta(\Sigma i) - \Sigma x_i} & \text{if } \forall i \quad x_i \geq i\theta \\ 0 & \end{cases}$$

$$= \begin{cases} e^{\theta \cdot \frac{n(n+1)}{2} - \Sigma x_i} & \text{if } \forall i \quad x_i \geq i\theta \\ 0 & \end{cases}$$

$\theta, 2\theta, 3\theta, \cdots, n\theta$

$\Rightarrow \quad x_1 \geq \theta, \; x_2 \geq 2\theta, \; x_3 \geq 3\theta$

$\ldots \quad x_n \geq n\theta$

$$\Rightarrow \quad \underbrace{x_1 \geq \theta, \; \frac{x_2}{2} \geq \theta, \; x_3 \geq \theta, \; \ldots \; \frac{x_n}{n} \geq \theta}_{\text{this can be re-expressed as}}$$

$$\min_i \left(\frac{x_i}{i} \right) \geq \theta$$

$$\Rightarrow \quad f(x|\theta) = \begin{cases} e^{\theta \cdot \frac{n(n+1)}{2} - \sum x_i} & \text{if } \min_i \left(\frac{x_i}{i}\right) \geq \theta \\ 0 & \text{o.w} \end{cases}$$

Take the statistic as

$$T(x) = \min_i \left(\frac{x_i}{i} \right)$$

$$g(t|\theta) = \begin{cases} e^{\theta \frac{n.(n+1)}{2}} & \text{if } t \geq \theta \\ 0 & \text{o.w} \end{cases}$$

$$h(x) = e^{-\sum x_i}$$

$$\Rightarrow f(x|\theta) = g(T(x)|\theta)\, h(x)$$

$\Rightarrow T(x) = \min_i \left(\frac{x_i}{i}\right)$ is a sufficient statistic.

(6.3) Let $X_1, X_2, \ldots X_n$ be a random sample from the pdf,

$$f(x|\mu,\sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma}$$

$\mu < x < \infty$

$0 < \sigma < \infty.$

Find a two-dimentional sufficient statistic for (μ, σ)

Sol^n

$$X_1, X_2, \ldots X_n \sim f(x|\mu,\sigma)$$

$$f(x|\mu,\sigma) = \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}}$$

Joint distribution

$$f(x \mid \mu, \sigma) = \begin{cases} \frac{1}{\sigma^n} \cdot e^{-\frac{1}{\sigma}(\sum x_i - n\mu)} & \forall i,\ x_i > \mu \\ 0 & \text{o.w} \end{cases}$$

This can be re-expressed as

$$f(x \mid \mu, \sigma) = \begin{cases} \frac{1}{\sigma^n} \cdot \exp\left(-\frac{1}{\sigma}(\sum x_i - n\mu)\right) & \min_i x_i > \mu \\ 0 & \text{o.w} \end{cases}$$

$$T_1(x) = \min_i x_i$$

$$T_2(x) = \sum x_i$$

$$\Rightarrow g(T_1(x), T_2(x) \mid \mu, \sigma) = \frac{1}{\sigma^n} \exp\left(-\frac{1}{\sigma}(T_2(x) - n\mu)\right)$$

$$h(x) = 1$$

Therefore $(\min_i x_i, \sum x_i)$ are the sufficient statistic's for μ, σ

6.4 Prove theorem 6.2.10

Sol^n Let $x_1, \dots, x_n$ iid observation's form a pdf or pmf $f(x|\theta)$ that belong's to an exponential family given by

$$f(x|\theta) = h(x)\, c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x)\right)$$

where $\theta = (\theta_1, \theta_2, \dots \theta_d)$, $d \le k$

Sol^n

$$f_{x_1 \dots x_n}(x_1, x_2, \dots x_n | \theta)$$

$$= \prod_{j=1}^{n} h(x_j)\, c(\theta) \exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x)\right)$$

$$= C(\theta)^n \left[\prod_{j=1}^{n} h(x_j)\right] \left[\exp\left(\sum_{j=1}^{n} \sum_{i=1}^{k} w_i(\theta)\, t_i(x_j)\right)\right]$$

Example: $n=3$, $k=3$

$$w_1(\theta).t_1(x_1) + w_2(\theta).t_2(x_1) + w_3(\theta)\, t_3(x_1)$$

$$+ \; w_1(\theta) \cdot t_1(x_2) + w_2(\theta).\, t_2(x_2) + w_3(\theta)\, t_3(x_2)$$

$$+ \; w_1(\theta) \cdot t_1(x_3) + w_2(\theta).t_2(x_3) + w_3(\theta)\, t_3(x_3)$$

$$\Rightarrow \; w_1(\theta) \cdot \left(\sum t_1(x_i)\right)$$

$$+ \; w_2(\theta) \left(\sum t_2(x_i)\right)$$

$$+ \; w_3(\theta) \left(\sum t_3(x_i)\right)$$

$$\Rightarrow \; \sum_{i=1}^{k} w_i(\theta) \cdot \left[\sum_{j=1}^{n} t_i(x_j)\right]$$

$f(x|\theta)$

$$= \underbrace{\left[\prod_{j=1}^{n} h(x_j)\right]}_{h(x)} \underbrace{c(\theta)^n \left[\exp\left(\sum_{i=1}^{k} w_i(\theta) \cdot \left[\sum_{j=1}^{n} t_i(x_j)\right]\right)\right]}_{g(T(x)|\theta)}$$

where $T(x) = \left(\sum_{j=1}^{n} t_1(x_j), \sum_{j=1}^{n} t_2(x_j), \dots, \sum_{j=1}^{n} t_k(x_j)\right)$

6.5 Let $X_1, X_2, \dots, X_n$ be independent random variables with pdf's.

$$f(x_i|\theta) = \begin{cases} \frac{1}{2i\theta} & -i(\theta-1) < x_i < i(\theta+1) \\ 0 & \text{o.w} \end{cases}$$

where $\theta > 0$, Find a two-dimentional sufficient statistic for θ.

<u>Soln</u>

$$f(x_i|\theta) = \begin{cases} \frac{1}{2i\theta} & -i(\theta-1) < x_i < i(\theta+1) \\ 0 & o.w \end{cases}$$

$$\Rightarrow f(X|\theta) = \begin{cases} \frac{1}{(2\theta)^n} \cdot \prod_{i=1}^{n} \frac{1}{i} & \forall i = \{1, 2, \cdots n\} \\ & -i(\theta-1) < x_i < i(\theta+1) \\ 0 & o.w \end{cases}$$

$$-i(\theta-1) < x_i < i(\theta+1) \quad \text{for all } i$$

$$\Rightarrow -(\theta-1) < x_1 < \theta+1, \quad -2(\theta-1) < x_2 < 2(\theta+1)$$

$$\vdots$$

$$\Rightarrow -(\theta-1) < \frac{x_i}{i} < \theta+1 \quad \forall i$$

$$\Rightarrow -\theta+1 < \frac{x_i}{i} < \theta+1 \quad \forall i$$

$$\Rightarrow \quad -\theta < \frac{x_i}{i} - 1 < \theta \qquad \forall i$$

$$\Rightarrow \quad \left| \frac{x_i}{i} - 1 \right| < \theta \qquad \forall i \;\; \& \;\; \theta > 0$$

$$\Rightarrow \quad \max_i \left| \frac{x_i}{i} - 1 \right| < \theta$$

$$\Rightarrow f(X|\theta) = \begin{cases} \frac{1}{(2\theta)^n} \prod_{i=1}^{n} \frac{1}{i} & \text{if } \max_i \left|\frac{x_i}{i} - 1\right| < \theta \\ 0 & \text{o.w} \end{cases}$$

therefore $h(x) = \begin{cases} \prod_{i=1}^{n} \frac{1}{i} & \text{for } i = 1, 2, \ldots n \\ 0 & \text{o.w} \end{cases}$

$$\underbrace{T_1(x) = \max_i \left| \frac{x_i}{i} - 1 \right|}_{\text{sufficient statistic}}$$

as we want 2D sufficient statistic

$$-\theta < \frac{x_i}{i} - 1 < \theta$$

$$\Rightarrow \max_i \left(\frac{x_i}{i}\right) < \theta + 1$$

$$\& \min_i \left(\frac{x_i}{i}\right) > -\theta - 1$$

we can take

$$T(x) = \left(\min_i \left(\frac{x_i}{i}\right), \max_i \left(\frac{x_i}{i}\right) \right)$$

6.6 Let $X_1, X_2, \ldots, X_n$ be random sample from a gamma (α, β) Population. Find a two-Dimensional sufficient statistic of α, β

Sol

$$f(x_i|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x_i^{\alpha-1} \cdot e^{-\frac{x_i}{\beta}}$$

$x_i \in (0,\infty)$

$$f(x|\alpha,\beta) = \left[\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right]^n \cdot e^{-\frac{\sum x_i}{\beta}} \cdot \prod_{i=1}^{n} x_i^{\alpha-1}$$

$$= \underbrace{\left[\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right]^n \cdot e^{-\frac{\sum x_i}{\beta}} \cdot \left(\prod_{i=1}^{n} x_i\right)^{\alpha-1}}_{g\left(\sum x_i,\ \prod_{i=1}^{n} x_i \,\middle|\, \alpha,\beta\right)}$$

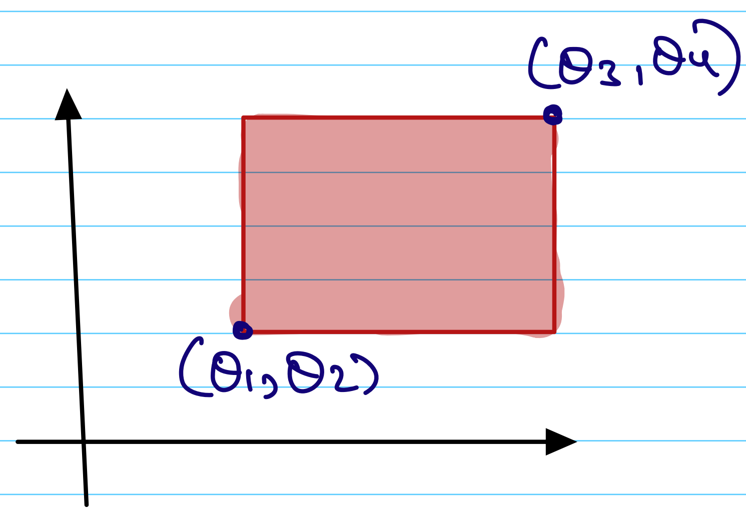
&

$h(x) = 1$

$\Rightarrow$ the 2 Dim soffient statistic for (α,β) are $\left(\sum_{i=1}^{n} x_i,\ \prod_{i=1}^{n} x_i\right)$

(6.7) Let $f(x,y \mid \theta_1, \theta_2, \theta_3, \theta_4)$ be the bivariate pdf for the uniform distribution on the rectangle with lower left corner (θ_1, θ_2) and upper right corner (θ_3, θ_4) in $\mathbb{R}^2$. The parameter's satisfy $\theta_1 < \theta_3$ and $\theta_2 < \theta_4$. Let (X_1, Y_1), ..., (X_n, Y_n) be a random sample from this pdf. find a four-dimensional sufficient statistic for $(\theta_1, \theta_2, \theta_3, \theta_4)$

Sol/n



$$f_X(x \mid \theta_1, \theta_3) = \begin{cases} \dfrac{1}{\theta_3 - \theta_1} & \text{if } \theta_1 < x_i < \theta_3 \\ 0 & \text{o.w} \end{cases}$$

$$f_Y(y \mid \theta_2, \theta_4) = \begin{cases} \dfrac{1}{\theta_4 - \theta_2} & \text{if } \theta_2 < y_i < \theta_4 \\ 0 & \text{o.w} \end{cases}$$

$$f_{X,Y}(x, y \mid \theta_1, \theta_2, \theta_3, \theta_4) = \begin{cases} \left(\dfrac{1}{\theta_3 - \theta_1}\right)\left(\dfrac{1}{\theta_4 - \theta_2}\right) & \text{if } \begin{matrix} \theta_1 < x < \theta_3 \\ \theta_2 < y < \theta_4 \end{matrix} \\ 0 & \text{o.w} \end{cases}$$

the Joint Pdf of $(X, Y) = \left((x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\right)$

$$f_{X,Y}(x, y \mid \theta_1, \theta_2, \theta_3, \theta_4) = \begin{cases} \left[\dfrac{1}{(\theta_3 - \theta_1)(\theta_4 - \theta_1)}\right]^n & \text{if } \begin{matrix} \forall i \in \{1, 2, \ldots n\} \\ \theta_1 < x_i < \theta_3 \\ \theta_2 < y_i < \theta_4 \end{matrix} \\ 0 & \text{o.w} \end{cases}$$

$\forall i,$ $\quad \theta_1 < x_i < \theta_3$

$\theta_2 < y_i < \theta_4$

This can be re expressed as

$$\min_i x_i > \theta_1$$

$$\max_i x_i < \theta_3$$

$$\min_i y_i > \theta_2$$

$$\max_i y_i < \theta_4$$

$$h(x) = \begin{cases} 1 & x_i \in \mathbb{R}, \text{ for } i = 1, 2, \ldots n \\ 0 & \text{o.w} \end{cases}$$

$$T_1(x) = \min_i x_i$$

$$T_2(x) = \min_i y_i$$

$$T_3(x) = \max_i x_i$$

$$T_4(x) = \max_i y_i$$

$$\Rightarrow g(T_1, T_2, T_3, T_4 \mid \theta_1, \theta_2, \theta_3, \theta_4)$$

$$= \begin{cases} \left[\dfrac{1}{(\theta_3-\theta_1)(\theta_4-\theta_2)}\right]^n & \text{for } \begin{matrix} t_1 > \theta_1 \\ t_2 > \theta_2 \\ t_3 < \theta_3 \\ t_4 < \theta_4 \end{matrix} \\ 0 & \text{o.w} \end{cases}$$

Therefore the 4-Dim sufficient statistic

are $\left(\min\limits_i x_i, \min\limits_i y_i, \max\limits_i x_i, \max\limits_i y_i\right)$

for $(\theta_1, \theta_2, \theta_3, \theta_4)$

6.8 Let $X_1, X_2, \dots X_n$ be a random sample from a population with location pdf $f(x-\theta)$. Show that the order statistic's

$$T(X_1, \dots X_n) = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$$

are a sufficient statistic for θ and no further reduction is possible we use Theorem 3.5.6 and work with $Z_1, Z_2, \dots Z_n$

$Z_1, Z_2, \dots, Z_n \overset{iid}{\sim} f(x) \quad (\theta = 0)$

$X_1 = Z_1 + \theta, \quad X_2 = Z_2 + \theta, \quad \dots, \quad X_n = Z_n + \theta$

where $-\infty < \theta < \infty$.

(6.9) for each of the following distribution's let $X_1, X_2, \ldots X_n$ be a random sample. Find a minimal sufficient statistic for θ

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

$-\infty < x < \infty$
$-\infty < \theta < \infty$
(normal)

$X_1, X_2, \ldots, X_n \sim N(\theta, 1)$

$$f(x|\theta) = \frac{1}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\left[\sum x_i^2 + n\theta^2 - 2\theta \sum x_i\right]}$$

$$\Rightarrow f(x|\theta) = \frac{1}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\left[\sum x_i^2 + n\theta^2 - 2\theta \sum x_i\right]}$$

if X, Y are two random sample points then

$$\frac{f(x|\theta)}{f(y,\theta)} = e^{-\frac{1}{2}\left[\sum x_i^2 - \sum y_i^2 - 2\theta\left(\sum x_i - \sum y_i\right)\right]}$$

The ratio will be constant as a function of $\theta \iff \sum x_i = \sum y_i$

Thus $\sum x_i$ is minimal sufficient statistic

(b) $f(x|\theta) = e^{-(x-\theta)}$. $\theta < x < \infty$

$-\infty < \theta < \infty$

(location Exponent.)

Sol^n

$x_1, x_2, \ldots, x_n$ iid $f(x|\theta)$

$x = (x_1, x_2, \ldots, x_n)$ The Joint Pdf

$f(x|\theta) = e^{n\theta - \sum x_i}$ $x_i = 1,2,\ldots n$

$x_i > \theta$

$$\Rightarrow \quad f(x|\theta) = e^{n\theta - \sum x_i} \qquad \min_i x_i > \theta$$

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{e^{n\theta} \cdot e^{-\sum x_i}}{e^{n\theta} \cdot e^{-\sum x_i}} \quad \frac{I_{(\theta,\infty)}(\min x_i)}{I_{(\theta,\infty)}(\min_i x_i)}$$

this will be in dependend

$$\Longleftrightarrow \quad \min_i x_i = \min_i y_i$$

$$\Longleftrightarrow \quad T(x) = T(y)$$

there fore $\min_i x_i$ is the minimum sufficient statistic.

(c) $$f(x|\theta) = \frac{e^{-(x-\theta)}}{\left(1+e^{-(x-\theta)}\right)^2}$$

$$-\infty < x < \infty$$

$$-\infty < \theta < \infty$$

(Logistic)

Sol^n

let $X = (X_1, X_2, \dots, X_n)$ iid

then the Joint distribution

$$f(X|\theta) = \prod_{i=1}^{n} \frac{e^{-(x_i-\theta)}}{\left(1+e^{-(x_i-\theta)}\right)^2}$$

$$= e^{n\theta} \cdot e^{-\sum x_i} \cdot \prod_{i=1}^{n} \frac{1}{\left(1+e^{-(x_i-\theta)}\right)^2}$$

$$\frac{f(x|\theta)}{f(y|\theta)} = e^{-(\sum x_i - \sum y_i)} \cdot \prod_{i=1}^{n} \left[\frac{1+e^{-(y_i-\theta)}}{1+e^{-(x_i-\theta)}}\right]^2$$

this need to be

1

$\Rightarrow$ entire order statistis is minimal sufficient statistic.

(d) $$f(x|\theta) = \frac{1}{\pi[1+(x-\theta)^2]} \qquad -\infty < x < \infty,\; -\infty < \theta < \infty$$

(Cauchy)

Sol^n

$$f(x|\theta) = \frac{1}{\pi^n} \cdot \frac{1}{\prod_{i=1}^{n} (1+(x_i-\theta)^2)}$$

$$\frac{f(x|\theta)}{f(y|\theta)} = \prod_{i=1}^{n} \frac{1+(y_i-\theta)^2}{1+(x_i-\theta)^2}$$

$\Downarrow$

To make this independent of θ, then order statistic is minimal sufficient statistic

(e) $$f(x|\theta) = \frac{1}{2} e^{-|x-\theta|} \qquad -\infty < x < \infty,\; -\infty < \theta < \infty$$

(double exponential)

Sol^n $f(x|\theta) = \frac{1}{2^n} \cdot e^{-\sum_{i=1}^{n} |x-\theta|}$

$$\frac{f(x|\theta)}{f(y|\theta)} = e^{-\left(\sum_{i=1}^{n} |x-\theta| - \sum_{i=1}^{n} |y-\theta|\right)}$$

$\shortparallel$

To make this independent of θ

Both x, y should have same number of sample's $< \theta$, $> \theta$

i.e # of sample's of $x < \theta$

= # of samplee's of $y < \theta$