This document contain's solution's to Exercises 6.10 to 6.24 from Chapter 6, of Statistical inference by Casella and Berger, focusing on sufficient statistic, minimal sufficient statistic, Complete statistic

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

6.10 Show that the minimal sufficient statistic for the uniform $(\theta, \theta+1)$ found in Example 6.2.5, is not complete.

Soln

$X_1, X_2, \ldots, X_n \sim \text{Unif}(\theta, \theta+1)$

$-\infty < \theta < \infty$

$$\Rightarrow f(x|\theta) = \begin{cases} 1 & \theta < x_i < \theta+1 \\ & i = 1, 2, \ldots n \\ 0 & \text{o.w} \end{cases}$$

$$\Rightarrow f(x|\theta) = \begin{cases} 1 & \min\limits_i x_i > \theta \\ & \max\limits_i x_i < \theta+1 \\ 0 & \text{o.w} \end{cases}$$

$$f(x|\theta) = I_{(\theta, \infty)}\left(\min_i x_i\right) I_{(-\infty, \theta+1)}\left(\max_i x_i\right)$$

$T_1(x) = \min_i x_i$

$T_2(x) = \max_i x_i$

$$\Rightarrow \quad f(x|\theta) = \underbrace{I_{(\theta,\infty)}(T_1(x))\, I_{(-\infty,\theta+1)}(T_2(x))}_{g(T_1(x),T_2(x)|\theta)\,\cdot\, h(x)}$$

$h(x) \Rightarrow 1$

$\Rightarrow (X_{(1)}, X_{(2)})$ is

sufficient statistic

$\Downarrow$

This is minimal sufficient statistic as well showed in EX 6.2.15

To show this is not complete we need to find the Joint distribution of minimal sufficient Statistic

$$f_{X_{(1)},X_{(n)}}(x_{(1)},x_{(n)}) = \begin{cases} n(n-1)(x_{(n)}-x_{(1)})^{n-2} & \theta < x_{(1)} < x_{(n)} < \theta+1 \\ 0 & o.w \end{cases}$$

let's replace $x_{(1)} = u$

$x_{(n)} = v$

$$f_{U,V}(u,v) = \begin{cases} n(n-1)(v-u)^{n-2} & \theta < u < v < \theta+1 \\ 0 & o.w \end{cases}$$

$$E_\theta\left[g(U,V)\right] = \int_\theta^{\theta+1}\int_\theta^{v} g(u,v)\, n(n-1)(v-u)^{n-2}\, du\, dv = 0$$

$$\Rightarrow n(n-1)\iint g(u,v)(v-u)^{n-2}\, du\, dv$$

difficult to solve in this approach.

we know Range R is independent of θ

$$\Rightarrow \quad \mathbb{R} = X_{(n)} - X_{(1)}$$

$$f_R(r|\theta) = n(n-1)r^{n-2}(1-r)$$

$$0 < r < 1$$

does not depend on θ

$$R \sim \text{beta}(n-1, 2)$$

$$\mathbb{E}[R] = \frac{n-1}{n+1} = \frac{\alpha}{\alpha+\beta}$$

therefore lets choose

$$g(X_{(1)}, X_{(n)}) = X_{(n)} - X_{(1)} - \frac{n-1}{n+1}$$

$$E\left[g(X_{(1)}, X_{(n)})\right] = 0 \quad \forall \theta$$

$\Rightarrow (X_{(1)}, X_{(n)})$ is not complete statistic

6.11 Refer to the Pdf's given in Exercise 6.9. For each, let $X_{(1)} < \dots < X_{(n)}$ be the ordered sample, and define

$$Y_i = X_{(n)} - X_{(i)}, \quad i = 1, 2, \dots, n-1$$

(a) For each of the Pdf's in exercise 6.9, verify that the set $(Y_1, \dots, Y_{n-1})$ is ancillary for θ. Try to prove a general theorem, Like Ex 6.2.18, that handles all these families at once.

Sol^n

$$Y_1 = X_{(n)} - X_{(1)}$$

$$Y_2 = X_{(n)} - X_{(2)}$$

$$\vdots$$

$$Y_{n-1} = X_{(n)} - X_{(n-1)}$$

$$Y_n = X_{(n)}$$

$$x_{(1)} = y_n - y_1$$

$$x_{(2)} = y_n - y_2$$

$$\vdots$$

$$x_{(n-1)} = y_n - y_{n-1}$$

$$x_{(n)} = y_n$$

$$|J| = \begin{vmatrix} -1 & 0 & 0 & \cdots & 1 \\ 0 & -1 & 0 & \cdots & 1 \\ 0 & 0 & -1 & \cdots & 1 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots -1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{vmatrix} = 1$$

The Joint distribution of order statistic

$$f_{X_{(1)}, X_{(2)}, \ldots X_{(n)}}(x_{(1)} x_{(2)}, \ldots x_{(n)} | \theta)$$

$$= n! \prod_{i=1}^{n} f(x_{(i)} | \theta)$$

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

$-\infty < \theta < \infty$

(normal)

$$f_{X_{(1)}, X_{(2)}, \ldots X_{(n)}}(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$$

$$= \frac{n!}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n}(x_{(i)}-\theta)^2}$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\left(\sum_{i=1}^{n} x_{(i)}^2 + n\theta^2 - 2\theta\sum_{i=1}^{n} x_{(i)}\right)}$$

$$f_{y_1, y_2, \ldots, y_n}(y_1, y_2, \ldots y_n|\theta) = \frac{n!}{(2\pi)^{n/2}} e^{-\frac{n}{2}\theta^2} \cdot e^{-\frac{1}{2}\left(\sum_{i=1}^{n}(x_{(i)})^2 - 2\theta\sum_{i=1}^{n} x_{(i)}\right)}$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{n}{2}\theta^2} \cdot e^{-\frac{1}{2}\left(\sum_{i=1}^{n-1}(y_n - y_i)^2 + y_n^2 - 2n\theta y_n + 2\theta\sum_{i=1}^{n-1} y_i\right)}$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{n}{2}\theta^2} \cdot e^{-\frac{1}{2}\left(n y_n^2 + \sum_{i=1}^{n} y_i^2 - 2y_n\sum_{i=1}^{n-1} y_i - 2n\theta y_n + 2\theta\sum_{i=1}^{n-1} y_i\right)}$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\left(n(y_n-\theta)^2 + \sum_{i=1}^{n} y_i^2 - 2(y_n-\theta)\sum_{i=1}^{n-1} y_i \right)}$$

$$f_{y_1, y_2, .. y_{n-1}}(y_1, y_2, .. y_{n-1})$$

$$= \frac{n!}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n-1} y_i^2} \cdot \int_{y_n=-\infty}^{\infty} e^{-\frac{1}{2}\left[n(y_n-\theta)^2 - 2(y_n-\theta)\sum_{i=1}^{n-1} y_i \right]} dy_n$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n-1} y_i^2} \int_{y_n=-\infty}^{\infty} e^{-\frac{n}{2}\left[(y_n-\theta)^2 - 2(y_n-\theta)\frac{\sum_{i=1}^{n-1} y_i}{n} + \left(\frac{1}{n}\sum_{i=1}^{n-1} y_i\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n-1} y_i\right)^2 \right]}$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n-1} y_i^2} \cdot e^{+\frac{1}{2n}\left(\sum_{i=1}^{n-1} y_i\right)^2}$$

$$\cdot \int_{y_n\infty}^{\infty} e^{-\frac{n}{2}\left[(y_n-\theta)^2 - 2(y_n-\theta)\frac{1}{n}\sum_{i=1}^{n-1} y_i + \left(\frac{1}{n}\sum_{i=1}^{n-1} y_i\right)^2 \right]}$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n-1} y_i^2} \cdot e^{+\frac{1}{2n}\left(\sum_{i=1}^{n-1} y_i\right)^2}$$

$$\cdot \int_{y_n=-\infty}^{\infty} e^{-\frac{1}{2\left(\frac{1}{\sqrt{n}}\right)^2}\left[y_n - \theta - \frac{1}{n}\sum_{i=1}^{n-1} y_i\right]^2}$$

$$= \frac{n!}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n-1} y_i^2} \cdot e^{\frac{1}{2n}\left(\sum_{i=1}^{n-1} y_i\right)^2} \cdot \frac{\sqrt{2\pi}}{\sqrt{n}}$$

$$= \frac{n!}{(2\pi)^{\frac{n-1}{2}}} \cdot \frac{1}{\sqrt{n}} \cdot e^{-\frac{1}{2}\left[\sum_{i=1}^{n-1} y_i^2 - \left(\sum_{i=1}^{n-1} y_i\right)^2\right]}$$

The Joint distribution of $Y_1, Y_2, \ldots, Y_n$ is independent of θ

$\Rightarrow (Y_1, Y_2, \ldots, Y_n)$ is ancillary statistic.

for location famili distribution's

$$f(x|\theta) = f(x-\theta)$$

let $Z_1, Z_2, \dots, Z_n \sim f(x)$

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

let's take an order statistic

$$X_{(1)}, X_{(2)}, X_{(3)}, \dots X_{(n)}$$

the the CDF of the Joint distribution of $Y_1, Y_2, \dots Y_{n-1}$

$$F(y_1, y_2, \dots y_{n-1} | \theta)$$

$$= \mathbb{P}_\theta (Y_1 \le y_1, Y_2 \le y_2, \dots Y_{n-1} \le y_{n-1})$$

$$= \mathbb{P}_\theta\big(X_{(n)} - X_{(1)} \le y_1,\ X_{(n)} - X_{(2)} \le y_2,\ \dots,\ X_{(n)} - X_{(n-1)} \le y_{n-1}\big)$$

$$= \mathbb{P}_\theta\big(Z_{(n)} - Z_{(1)} \le y_1,\ Z_{(n)} - Z_{(2)} \le y_2,\ \dots,\ Z_{(n)} - Z_{(n-1)} \le y_{n-1}\big)$$

$\Rightarrow$ This Joint CDF is independent of Parameter θ

$\Rightarrow$ $\big(X_{(n)} - X_{(1)},\ X_{(n)} - X_{(2)},\ \dots\ X_{(n)} - X_{(1)}\big)$ is an ancillary statistic

6.11 (a) $f(x|\theta) = e^{-(x-\theta)}$

$\theta < x < \infty$

$-\infty < \theta < \infty$

$$f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_{(1)}, x_{(2)}, x_{(3)} \ldots x_{(n)} | \theta)$$

$$= \begin{cases} n!\, f_X(x_{(1)})\, f_X(x_{(2)}) \cdots f_X(x_{(n)}) & \theta < x_{(1)} < x_{(2)} < \ldots < x_{(n)} < \infty \\ & -\infty < \theta < \infty \\ 0 & \text{o.w} \end{cases}$$

$$= n!\, e^{n\theta - \sum_{i=1}^{n} x_{(i)}}$$

$\theta < x_{(1)} < x_{(2)} < \ldots < x_{(n)} < \infty$

$-\infty < \theta < \infty$

$$f_{X_{(1)}, X_{(2)}, \ldots X_{(n)}}(x_{(1)}, x_{(2)}, \ldots x_{(n)}) = n!\, e^{n\theta - \sum_{i=1}^{n} x_{(i)}}\, I_{(\theta, \infty)}(x_{(1)})$$

$$\left.\begin{aligned}
Y_1 &= X_{(n)} - X_{(1)} \\
Y_2 &= X_{(n)} - X_{(2)} \\
&\vdots \\
Y_{n-1} &= X_{(n)} - X_{(n-1)} \\
Y_n &= X_{(1)}
\end{aligned}\right\} \Longrightarrow
\begin{aligned}
X_{(1)} &= y_n \\
X_{(2)} &= y_1 + y_n - y_2 \\
&\vdots \\
x_{(n-1)} &= y_1 + y_n - y_{n-1} \\
x_{(n)} &= y_1 + y_n
\end{aligned}$$

$$|J| = \begin{vmatrix}
0 & 0 & \cdots & & 0 & 1 \\
1 & -1 & 0 & \cdots & 0 & 1 \\
1 & 0 & -1 & \cdots & 0 & 1 \\
1 & 0 & 0 & -1 \cdots & 0 & 1 \\
\vdots & & & & & \\
1 & 0 & 0 & 0 \cdots & -1 & 1
\end{vmatrix} = 1$$

$$f_{y_1, y_2, \ldots y_n}(y_1, y_2, \ldots y_n) = n!\, e^{n\theta} \cdot e^{-(n-1)y_1 - n y_n + \sum_{i=2}^{n-1} y_i}$$

$$0 < y_{n-1} < y_{n-2} < \cdots < y_1 < \infty$$

$$\theta < y_n < \infty$$

$$= n!\, e^{n\theta} \cdot e^{-(n-1)y_1 + \sum_{i=2}^{n-1} y_i}\, e^{-n y_n}\, I_{(\theta,\infty)}(y_n)$$

$$= n! \cdot e^{-(n-1)y_1 + \sum_{i=2}^{n-1} y_i}\, e^{n(\theta - y_n)}\, I_{(\theta,\infty)}(y_n)$$

from Joint distribution

$(Y_1, Y_2, \dots Y_{n-1})$ & Y_n

are separable function $\Rightarrow$ Independent

& $(Y_1, Y_2, \dots Y_{n-1})$ independent

of θ $\Rightarrow$ ancillary statistic.

6.11 b

in each case determine whether the set $(Y_1, \dots Y_{n-1})$ is independent of the minimal sufficient statistic.

Sol^n

from BASU's theorem if a statistic is complete then it is independent of ancillary statistic & it is minimal sufficient statistic

(a) (normal)

$T(X) = \bar{X}$ (minimal sufficient statistic)

$\Downarrow$

independent of ancillary statistic $(Y_1, Y_2, \ldots Y_{n-1})$

(b) (location exponential)

$T(X) = X_{(1)}$ we proved

$T(X)$ is independent of $(Y_1, Y_2, \ldots Y_{n-1})$ in (6.11 a) using joint distribution

(c) (logistic) (d) Cauchy (e) double Exponential

$T(X) = (X_{(1)}, X_{(2)} \ldots, X_{(n)})$

$\Rightarrow T(X)$ is complete $\Rightarrow$ independent of ancillary statistic

6.12 A natural ancillary statistic in most problem's is the sample size. for example, let N be a random variable taking values $1, 2, \ldots$ with known probabilities $p_1, p_2, \ldots$, where $\sum p_i = 1$. Having observed $N = n$, perform

(a) Prove that the pair (X, N) is minimal sufficient statistic and N is ancillary for θ. (Note the similarity to some of the hierarchical models discussed in sec 4.4)

Soln

$$N \in \{1, 2, 3, \ldots\}$$

each has PROB $\{p_1, p_2, \ldots\}$

such that $\sum_{i=1}^{\infty} p_i = 1$

$$X \sim \text{Bernoulli}(N, \theta)$$

$$f(X \mid N, \theta) = \binom{N}{x} \theta^{x} (1-\theta)^{N-x}$$

$$x = \{1, 2, \dots N\}$$

$$f(X, N \mid \theta) = \mathbb{P}(X \mid N=n, \theta)\, \mathbb{P}(N=n \mid \theta)$$

$$= \mathbb{P}(X \mid N=n, \theta)\, \mathbb{P}(N=n)$$

$$= \binom{n}{x} \theta^{x} (1-\theta)^{n-x} \cdot P_n$$

$$\frac{f(x, N_1 \mid \theta)}{f(y, N_2 \mid \theta)} = \frac{\binom{n_1}{x} \theta^{x} (1-\theta)^{n_1-x} P_{n_1}}{\binom{n_2}{x} \theta^{y} (1-\theta)^{n_2-y} P_{n_2}}$$

$$= \frac{\binom{n_1}{x}}{\binom{n_2}{x}} \cdot \theta^{x-y} (1-\theta)^{(n_1-n_2)-(x-y)} \cdot \frac{P_{n_1}}{P_{n_2}}$$

for this ratio to be independent

of θ $\quad X = Y,\quad N_1 = N_2$

$\Rightarrow$ (X, N) minimal sufficient statistic.

$$\mathbb{P}(N=n) = P_n \quad \text{s.t} \quad \sum_{n=1}^{\infty} P_n = 1$$

$\Rightarrow$ $\mathbb{P}(N)$ does not depend on θ

$\Rightarrow$ N is ancillary statistic.

6.12
(b) Prove that the estimator $\frac{X}{N}$ is unbaised for θ and has variance $\theta(1-\theta)\,\mathbb{E}\left[\frac{1}{N}\right]$

Sol^n

$$f(x, N \mid \theta) = \binom{N}{x} \theta^{x} (1-\theta)^{N-x} \cdot P_N$$

using theorem 4.4.3

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$$

$$\Rightarrow \mathbb{E}\left[\frac{X}{N}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{X}{N} \mid N\right]\right]$$

$$= \mathbb{E}\left[\frac{1}{N}\mathbb{E}[X \mid N]\right]$$

$$= \mathbb{E}\left[\frac{1}{N} N\theta\right]$$

$$= \mathbb{E}[\theta] = \theta$$

$\Rightarrow \mathbb{E}\left[\frac{X}{N}\right] = \theta \Rightarrow$ unbaised estimator of θ

Using theorem 4.4.7

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{var}(X \mid Y)] + \mathrm{var}(\mathbb{E}[X \mid Y])$$

$$\Rightarrow \mathrm{var}\left(\frac{X}{N}\right) = \mathbb{E}\left[\mathrm{var}\left(\frac{X}{N} \mid N\right)\right] + \mathrm{var}\left[\mathbb{E}\left[\frac{X}{N} \mid N\right]\right]$$

$$= \mathbb{E}\left[\frac{1}{N^2} \cancel{N}\theta(1-\theta)\right]$$

$$+ var\left[\frac{1}{\cancel{N}}\cancel{N}\theta\right]$$

$$= \theta(1-\theta)\, E\left[\frac{1}{N}\right] + \underbrace{var(\theta)}_{=0}$$

$$= \theta(1-\theta)\, E\left[\frac{1}{N}\right]$$

6.13 Suppose X_1 and X_2 are iid observations from the pdf $f(x|\alpha) = \alpha x^{\alpha-1} e^{-x^\alpha}$

$x>0, \alpha>0$

Show that $\dfrac{\log X_1}{\log X_2}$ is an ancillary statistic

Sol^n

$$f(x|\alpha) = \alpha x^{\alpha-1} \cdot e^{-x^\alpha} \quad x>0, \alpha>0$$

$y = \log x$

$\Rightarrow \quad x = e^y \qquad \dfrac{dx}{dy} = e^y$

$$f_Y(y|\alpha) = f_X(e^y|\alpha) \cdot \left|\frac{dx}{dy}\right|$$

$$= \alpha \cdot e^{y(\alpha-1)} \cdot e^{-e^{y\alpha}} \cdot e^{y}$$

$$f_Y(y|\alpha) = \alpha \cdot e^{y\alpha} \cdot e^{-e^{y\alpha}}$$

$$= \alpha \cdot \exp\{y\alpha - e^{y\alpha}\}$$

we know that $f\left(\frac{x}{\alpha}\right) = f(x) \cdot \alpha$

let $x = \frac{y}{\alpha}$

the $f_X(x|\alpha) = \exp\{x - e^{x}\}$

$\Rightarrow$ $f_Y(y|\alpha)$ is a scale family

for a scale family distribution the

ratio is independent of α

i.e $\frac{Y_1}{Y_2}$ is independent of α

$\Rightarrow$ $\frac{\log X_1}{\log X_2}$ is ancillary statistic.

(6.14) Let $X_1, X_2, \dots X_n$ be a random sample from a location famility. Show that $M - \bar{X}$ is an ancillary statistic, where M is the sample median

Sol^n $X_1, X_2, \dots, X_n \overset{iid}{\sim} f_X(x)$

Given that X is a location family distribution's

$\Rightarrow X_i = Z_1 + \theta$ form

$Z_1, Z_2, \dots, Z_n \overset{iid}{\sim} Z$

Where $f_Z(z)$ is standard distribution

$$M = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{if } n \text{ is odd} \\ \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{if } n \text{ is even} \end{cases}$$

$$\Rightarrow M = \begin{cases} Z_{\left(\frac{n+1}{2}\right)} + \theta & \text{if } n \text{ is odd} \\ \dfrac{Z_{\left(\frac{n}{2}\right)} + Z_{\left(\frac{n}{2}+1\right)}}{2} + \theta & \text{if } n \text{ is even} \end{cases}$$

$$\Rightarrow \quad M(X) = M(Z) + \theta$$

$$\bar{X} = \bar{Z} + \theta$$

therefore $\quad M(X) - \bar{X} = \underbrace{M(Z) - \bar{Z}}_{\text{independent of } \theta}$

$\Rightarrow \quad M(X) - \bar{X} = M(Z) - \bar{Z}$ is ancillary statistic.

6.15 Let $X_1, X_2, \ldots, X_n$ be iid $n(\theta, a\theta^2)$, where $a > 0$ is a known constant and $\theta > 0$.

(a) Show that the parameter space does not contain a two-dimentional open set.

$$X_1, X_2, \ldots X_n \sim N(\theta, a\theta^2)$$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi a}\,\theta} \exp\left\{-\frac{1}{2a\theta^2}(x-\theta)^2\right\}$$

$$\Rightarrow f(x|\theta) = \frac{1}{\sqrt{2\pi a}\,\theta} \exp\left\{-\frac{1}{2a\theta^2}(x^2+\theta^2-2\theta x)\right\}$$

$$= \frac{1}{\sqrt{2\pi a}\,\theta} \exp\left\{-\frac{x^2}{2a\theta^2} - \frac{1}{2a} + \frac{x}{a\theta}\right\}$$

$$= \underbrace{\frac{1}{\sqrt{2\pi a}} \cdot \exp\left(\frac{-1}{2a}\right)}_{h(x)} \cdot \underbrace{\frac{1}{\theta}}_{c(\theta)} \exp\left(\frac{-x^2}{2a\theta^2} + \frac{x}{a\theta}\right)$$

natural parameters: $w_1(\theta) = \frac{1}{\theta^2}$ $\quad w_2(\theta) = \frac{1}{\theta}$

$t_1(x) = \frac{-x^2}{2a}$ $\quad t_2(x) = \frac{x}{a}$

here $k = 2$, But $d = 1$

$\Rightarrow$ This is not fully exponential family, This is curved exponential family.

The Parameter space looks like a Parabola

(b) Show that the statistic $(\bar{x}, s^2)$ is a sufficient statistic for θ, but the family of Distributions is not complete.

Sol^n

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

$$= \frac{1}{(2\pi a)^{n/2}} \cdot \frac{1}{\theta^n} \cdot \exp\left\{ \frac{-1}{2a\theta^2} \sum_{i=1}^{n} (x_i - \theta)^2 \right\}$$

$$= \frac{1}{(2\pi a)^{n/2}} \cdot \frac{1}{\theta^n} \cdot \exp\left\{ \frac{-1}{2a\theta^2} \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \theta)^2 \right\}$$

$$\sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \theta)^2 = \sum_{i=1}^{n} \left\{ (x_i - \bar{x})^2 + (\bar{x} - \theta)^2 + 2(x_i - \bar{x})(\bar{x} - \theta) \right\}$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2 + 2(\bar{x}-\theta)\underbrace{\cancel{\left(\sum x_i-\bar{x}\right)}}_{0}$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\theta)^2$$

$$= (n-1)\,S^2 + n(\bar{x}-\theta)^2$$

$\Longrightarrow$

$$f(X|\theta) = \underbrace{\frac{1}{(2\pi a)^{n/2}} \cdot \frac{1}{\theta^n} \cdot \exp\left\{\frac{-1}{2a\theta^2}\left[(n-1)s^2 + n(\bar{x}-\theta)^2\right]\right\}}_{g(\bar{x},\, s^2 \,|\, \theta)}$$

$$h(x) = 1$$

Hence $(\bar{x}, s^2)$ is sufficient statistic.

To show that this $T(X) = (\bar{x}, s^2)$ is not complete statistic we need to find $E[g(T(x))] = 0$ $\forall\theta$ but $g(T(x)) \neq 0$ for some t

$$\Rightarrow \quad \mathbb{E}[\bar{X}^2] = \operatorname{var}(\bar{X}) + E[\bar{X}]^2$$

$$= \frac{a\theta^2}{n} + \theta^2$$

$$= \frac{a+n}{n}\,\theta^2$$

$$\mathbb{E}[S^2] = a\theta^2$$

$$\Rightarrow \quad \mathbb{E}\left[\bar{X}^2 \cdot \frac{n}{a+n} - \frac{S^2}{a}\right] = 0 \quad \forall \theta$$

$$\Rightarrow \quad g(\bar{X}, S^2) = \bar{X}^2 \cdot \frac{n}{a+n} - \frac{S^2}{a} \neq 0$$ for some x

But $E\left[g(\bar{x}, S^2)\right] = 0 \quad \forall \theta$

$\Rightarrow \quad (\bar{X}, S^2)$ is not complete statistic.

6.16 A famous example in genetic modeling is a genetic linkage multinomial model, where we observe the multinomial vector (X_1, X_2, X_3, X_4)

with cell Prob given by

$$\left(\frac{1}{2}+\frac{\theta}{4},\ \frac{1}{4}(1-\theta),\ \frac{1}{4}(1-\theta),\ \frac{\theta}{4}\right)$$

(a) Show that this is a curved exponential family.

Sol^n

$$f(x_1, x_2, x_3, x_4) =$$

$$= \frac{m!}{x_1!\,x_2!\,x_3!\,x_4!}\left(\frac{1}{2}+\frac{\theta}{4}\right)^{x_1}\left(\frac{1}{4}(1-\theta)\right)^{x_2}\left(\frac{1}{4}(1-\theta)\right)^{x_3}\left(\frac{\theta}{4}\right)^{x_4}$$

$e^{\log x}$ where $x_1+x_2+x_3+x_4=m.$

$$= \frac{m!}{x_1!\,x_2!\,x_3 x_4!}\exp\left\{x_1\log\left(\frac{1}{2}+\frac{\theta}{4}\right)+x_2\log\left(\frac{1}{4}(1-\theta)\right)+x_3\log\left(\frac{1}{4}(1-\theta)\right)+x_4\log\frac{\theta}{4}\right\}$$

$$= \frac{m!}{x_1!\, x_2!\, x_3!\, x_4!} \exp\left\{ x_1 \log\left(\frac{1}{2}+\frac{\theta}{4}\right) + (x_2+x_3)\log\left(\frac{1}{4}(1-\theta)\right) + x_4 \log \frac{\theta}{4} \right\}$$

natural Parameter's

$$\left.\begin{aligned} w_1(\theta) &= \log\left(\frac{1}{2}+\frac{\theta}{4}\right) \\ w_2(\theta) &= \log\left(\frac{1}{4}(1-\theta)\right) \\ w_3(\theta) &= \log\frac{\theta}{4} \end{aligned}\right\} \Rightarrow \begin{aligned} t_1(x) &= x_1 \\ t_2(x) &= x_2+x_3 \\ t_3(x) &= \log\frac{\theta}{4} \end{aligned}$$

$\Rightarrow \quad k=3, \quad d=1$

$\Rightarrow$ this is Curved-exponential family.

(b) find a sufficient statistic for θ

let $X_1, X_2, X_3, \ldots X_n$ iid multinomial

$$f(x|\theta) = \frac{(m!)^n}{\prod x_1^i!\, x_2^i!\, x_3^i!\, x_4^i!}\cdot$$

$$\cdot \left(\frac{1}{2}+\frac{\theta}{4}\right)^{\sum x_1^i} \left(\frac{1}{4}(1-\theta)\right)^{\sum x_2^i}$$

$$\left(\frac{1}{4}(1-\theta)\right)^{\sum x_3^i} \left(\frac{\theta}{4}\right)^{\sum x_4^i}$$

$$\Rightarrow \frac{(m!)^n}{\prod x_1^i!\, x_2^i!\, x_3^i!\, x_4^i!} \cdot \left(\frac{1}{2}+\frac{\theta}{4}\right)^{\sum x_1^i} \left(\frac{1}{4}(1-\theta)\right)^{\sum x_2^i + \sum x_3^i} \left(\frac{\theta}{4}\right)^{\sum x_4^i}$$

$T(X) = \sum X_i$ is sufficient statistic.

$$\frac{f(X|\theta)}{f(Y|\theta)} \propto \left(\frac{1}{2}+\frac{\theta}{4}\right)^{\sum X_1^i - \sum Y_1^i} \cdot \left(\frac{1}{4}(1-\theta)\right)^{\sum X_2^i + \sum X_3^i - \sum Y_2^i + \sum Y_3^i} \cdot \left(\frac{\theta}{4}\right)^{\sum X_4^i - \sum Y_4^i}$$

for this ratio need to be independe

we need to have

$$\left.\begin{array}{l} \sum x_1^i = \sum y_1^i \\ \sum x_4^i = \sum y_4^i \end{array}\right\}$$ this is not equal to $\sum x = \sum y$

$\Rightarrow$ $T(x) = \sum x$ is not minimal sufficient statistic

6.17 Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ geometric(θ)

$$\mathbb{P}_\theta(X = x) = \theta(1-\theta)^{x-1}$$

$$x = 1, 2, \ldots, \quad 0 < \theta < 1$$

Show that $\sum X_i$ is sufficient for θ, and find the family of distribution of $\sum X_i$. is the family complete?

Sol^n

$$f(X = x \mid \theta) = \theta(1-\theta)^{x-1}$$

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

$X = (X_1, X_2, \ldots, X_n)$ random sample

$$\Rightarrow f(X \mid \theta) = \theta^n (1-\theta)^{\sum x_i - n}$$

$$= \underbrace{\theta^n (1-\theta)^{\sum x_i - n}}_{g(\sum x_i \mid \theta)} \cdot \underbrace{1}_{h(x)}$$

hence $T(X) = \sum X_i$ is

Sufficient Statistic

$$\mathbb{P}_\theta(X=x) = \theta(1-\theta)^{x-1}$$

$$= \exp\left(\log\theta + (x-1)\log(1-\theta)\right)$$

$$= \exp\left(\log\theta + x\log(1-\theta) - \log(1-\theta)\right)$$

$$= \exp\left(\log\frac{\theta}{1-\theta} + x\log(1-\theta)\right)$$

$$= \frac{\theta}{1-\theta}\cdot\exp\left(x\log(1-\theta)\right)$$

$$t_1(x) = x$$

then from theorem 6.2.10

$$T(x) = \sum_{i=1}^{n} t_1(x_i) = \sum x_i$$

is a sufficient statistic

from theorem 6.2.25 $T(x) = \sum x_i$ is complete statistic.

6.18 Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim}$ Poisson(λ)

Show that family of distribution's of $\sum x_i$ is Complete. Prove Completeness without using theorem 6.2.25

Sol^n $X_1, X_2, \ldots X_n \sim$ Poisson(λ)

$$P(X=x|\lambda) = \frac{e^{-\lambda} \lambda^{x}}{x!} \quad x = 0,1,2,3,\ldots$$

Joint

$$\mathbb{P}(X|\lambda) = \frac{e^{-n\lambda} \cdot \lambda^{\sum x_i}}{x_1! x_2! \ldots x_n!}$$

$$= \underbrace{\frac{1}{x_1! x_2! \ldots x_n!}}_{h(x)} \cdot \underbrace{e^{-n\lambda} \cdot \lambda^{\sum x_i}}_{g(\sum x_i | \lambda)}$$

$\Rightarrow \sum x_i = T(x)$ sufficient statistic.

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

$\sum x_i \sim \text{Poisson}(n\lambda)$

$$E_\lambda\{g(t)\} = \sum_{t=0}^{\infty} g(t) \frac{e^{-n\lambda} (n\lambda)^t}{(t)!}$$

$$= e^{-n\lambda} \sum_{t=0}^{\infty} \frac{g(t) \cdot (n\lambda)^t}{t!}$$

for this value to be $0 \ \forall t$

$g(t) = 0 \quad \forall t$

$\Rightarrow \sum X$ is a complete statistic.

6.19

The random variable X takes the value $0, 1, 2$ according to one of the following distribution's

	$P(X=0)$	$P(X=1)$	$P(X=2)$	
Dist 1	p	$3p$	$1-4p$	$0<p<1/4$
Dist 2	p	p^2	$1-p-p^2$	$0<p<1/2$

in each case determine whether the family of distribution's of X is complete.

Sol^n Distribution 1

$$\mathbb{E}[g(X)] = \sum_{x=0}^{2} g(x)\, P(X=x)$$

$$= g(0)\cdot p + g(1)\,3p + g(2)\,(1-4p)$$

$$= p\,g(0) + 3p\,g(1) + g(2) - 4p\,g(2)$$

take $g(0) = g(1) = g(2) = 1$

then

$$\mathbb{E}[g(X)] = 1$$

Now take $\mathbb{E}[g(x)-1] = E[g(x)]-1$

$= 1-1 = 0 \quad \forall p$

$\Rightarrow$ Dist 1 is not complete

Dist 2

$$E_p[g(x)] = p\,g(0) + p^2 g(1) + (1-p-p^2)\,g(2)$$

$$= p(g(0)-g(2)) + p^2(g(1)-g(2)) + g(2)$$

for this to be zero for all p

$g(0)-g(2)=0$

$g(1)-g(2)=0 \quad \Rightarrow \quad g(x)=0 \quad \forall x=1,2,3$

$g(2)=0$

$\Rightarrow$ Dist 2 is complete statistic.

6.20 For each of the following pdf's let $X_1, X_2, \dots X_n$ be iid observation's.

Find a complete sufficient statistic, or show that one does not exist's.

(a) $f(x|\theta) = \frac{2x}{\theta^2}$, $0 < x < \theta$

$\theta > 0$

Sol^n

$f(x|\theta) = \frac{2x}{\theta^2}$, $0 < x < \theta$

$\theta > 0$

$$f(x|\theta) = \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^{n} x_i \quad , 0 < x_i < \theta \;\; \forall i$$

$$\Rightarrow f(x|\theta) = \left(\frac{2}{\theta^2}\right)^n \prod_{i=1}^{n} x_i \; I_{(0,\theta)}(\max_i x_i)$$

$$\Rightarrow f(x|\theta) = \left(\frac{2}{\theta^2}\right)^n I_{(0,\theta)}(\max x_i) \cdot \prod_{i=1}^{n} x_i$$

$$\Rightarrow \quad T(X) = \max_i x_i$$

is a sufficient statistic.

To prove that $X_{(n)}$ is complete statistic, we need to find its distribution

Pdf = $f(x|\theta) \quad \frac{2x}{\theta^2} \qquad 0 < x < \theta, \; \theta > 0$

CDF $\quad F_X(x|\theta) = \int_0^x \frac{2x}{\theta^2}\, dx \qquad 0 < x < \theta$

$$= \frac{2}{\theta^2} \cdot \frac{x^2}{2} \Big]_0^x \, dx \qquad 0 < x < \theta$$

$$= \left(\frac{x}{\theta}\right)^2$$

Therefore the distribution of $X_{(n)}$ from Theorem 5.4.4

$$f_{X_{(n)}}(x) = \frac{n!}{(n-1)!} \frac{2x}{\theta^2} \cdot \frac{\left(\left(\frac{x}{\theta}\right)^{2 \cdot n-1} \right)}{\left(1 - \frac{x^2}{\theta^2}\right)^{n-n}}$$

$$= n \cdot \frac{2x}{\theta^2} \left(\frac{x}{\theta}\right)^{2n-2}$$

$$= \frac{2n}{\theta^{2n}} \cdot x^{2n-1}$$

$$\Rightarrow \quad f_{X_{(n)}}(x) = \frac{2n}{\theta^{2n}} \cdot x^{2n-1} \qquad 0 < x < \theta$$

$$\mathbb{E}[g(X)] = \int_0^\theta g(x) \cdot \frac{2n}{\theta^{2n}} \cdot x^{2n-1} \, dx$$

$$= \frac{2n}{\theta^{2n}} \int_0^\theta g(x) \, x^{2n-1} \, dx$$

for $\mathbb{E}[g(X)] = 0 \quad \forall \theta$ the

$$\Rightarrow \quad \frac{d}{d\theta} \mathbb{E}[g(x)] = 0$$

$$\Rightarrow \quad 0 = \frac{d}{d\theta} \left[\frac{2n}{\theta^{2n}} \int_0^\theta g(x)\, x^{2n-1}\, dx \right]$$

$$\Rightarrow \quad \frac{2n}{\theta^{2n}} \frac{d}{d\theta} \int_0^\theta g(x)\, x^{2n-1}\, dx + \int_0^\theta g(x)\, x^{2n-1}\, dx \cdot \frac{-2n}{\theta^{2n+1}} = 0$$

$$\Rightarrow \quad \frac{2n}{\theta^{2n}} \cdot g(\theta) \cdot \theta^{2n-1} - \frac{2n}{\theta^{2n+1}} \cdot 0 = 0$$

$$\Rightarrow \quad \frac{2n}{\theta} \cdot g(\theta) \quad - 0 = 0$$

$$\Rightarrow \quad g(\theta) = 0 \quad \forall \theta$$

Hence $T(x) = \max_i x_i$

is complete statistic.

(b) $$f(x|\theta) = \frac{\theta}{(1+x)^{1+\theta}}, \quad 0 < x < \infty, \quad \theta > 0$$

Sol^n

$$f(x|\theta) = \theta \cdot (1+x)^{-(1+\theta)} \qquad 0 < x < \infty, \quad \theta > 0$$

$$= \theta \cdot \exp\left\{-(1+\theta)\log(1+x)\right\}$$

$$\Rightarrow \quad c(\theta) = \theta, \quad h(x) = 1$$

$$w_1(\theta) = -(1+\theta)$$

$$t_1(x) = \log(1+x)$$

this is a exponential family

$$\Rightarrow \quad T(x) = \sum_{i=1}^{n} \log(1+x)$$

minimal sufficient statistic

(c) $$f(x|\theta) = \frac{(\log\theta)}{\theta - 1}\,\theta^{x}, \quad 0 < x < 1, \quad \theta > 1$$

Sol^n

$$f(x|\theta) = \frac{\log\theta}{\theta - 1}\exp(x \log\theta)$$

$0 < x < 1$

$\theta > 1$

$\Rightarrow$ $T(x) = \sum x$ is complete sufficient statistic

(d) $$f(x|\theta) = e^{-(x-\theta)} \exp\left(-e^{-(x-\theta)}\right)$$

$-\infty < x < \infty$

$-\infty < \theta < \infty$

Sol^n

$$f(x|\theta) = e^{-x} \cdot e^{\theta} \cdot \exp\left(-e^{-(x-\theta)}\right)$$

$$= e^{-x} \cdot e^{\theta} \cdot \exp\left(-e^{-x} \cdot e^{+\theta}\right)$$

Exponential family

$$\Rightarrow \underbrace{T(X) = \sum_{i=1}^{n} e^{-x}}$$

Complete sufficient statistic of θ.

(e)

$$f(x|\theta) = \binom{2}{x} \theta^{x} (1-\theta)^{2-x}$$

$x = 0, 1, 2$

$0 \leq \theta \leq 1$

Sol^n

$$f(x|\theta) = \binom{2}{x} \exp\left(x \log\theta + (2-x) \log(1-\theta) \right)$$

$$= \binom{2}{x} \exp\left(x \log\theta + 2 \log(1-\theta) - x \log(1-\theta) \right)$$

$$= \binom{2}{x} \exp\left(x \log \frac{\theta}{1-\theta} \right) \cdot (1-\theta)^{2}$$

$\Rightarrow$ Exponential family

$T(x) = \sum x_i$ is complete sufficient statistic.

6.21 Let X be one observation from the pdf

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

$x = -1, 0, 1$

$0 \leq \theta \leq 1$

c) does $f(x|\theta)$ belong to the Exponential class?

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

$x = -1, 0, 1$

$$= \left(\frac{1}{2}\right)^{|x|} \cdot (\theta)^{|x|} \frac{(1-\theta)}{(1-\theta)^{|x|}}$$

$$= \left(\frac{1}{2}\right)^{|x|} (1-\theta) \exp\left\{ |x| \cdot \log \frac{\theta}{1-\theta} \right\}$$

$$= h(x)\, c(\theta) \exp\left\{ t(x) \cdot w(\theta) \right\}$$

$f(x|\theta)$ is exponential family.

(b) is $|X|$ a complete sufficient statistic?

Soln from exponential distribution in (c) $|X|$ is complete sufficient statistic.

(b) is X a complete sufficient statistic.

Soln $\mathbb{E}[g(X)] =$

$$g(-1)\cdot\frac{\theta}{2} + g(0)(1-\theta)$$

$$g(1)\cdot\frac{\theta}{2} = 0$$

$$\Rightarrow \theta\left(\frac{g(-1)}{2} - g(0) + \frac{g(1)}{2}\right) + g(0) = 0$$

for this to be zero for all θ

$$g(0) = 0, \quad \frac{g(-1)}{2} + \frac{g(1)}{2} = 0$$

$\Rightarrow$ X is not complete

6.22 Let $X_1, X_2, \ldots X_n$ be a random sample from a population with pdf

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0<x<1, \quad \theta>0$$

(a) is $\sum x_i$ sufficient for θ?

$$f(x|\theta) = \theta^n \cdot \left(\prod x_i\right)^{\theta-1} \quad 0<x<1, \quad \theta>0$$

$T(x) = \prod x_i$ is sufficient statistic.

$\Rightarrow$ $\sum x_i$ is not sufficient statistic

(b) Find a complete sufficient statistic for θ

Sol^n $f(x|\theta) = \theta \exp\{(\theta-1)\log x\}$ $\quad 0<x<1$

from theorem 6.2.25

$$T(x) = \sum_{i=1}^{n} \log x_i$$

$$= \log\left(\prod_{i=1}^{n} x_i\right)$$

Therefore $\log\left(\prod_{i=1}^{n} x_i\right)$ is complete sufficient statistic.

& x & $\log x$ have one-to-one function

$\Rightarrow \prod_{i=1}^{n} x_i$ is complete sufficient statistic.

6.23 Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform distribution on the interval $(\theta, 2\theta)$, $\theta > 0$.

Find a minimal sufficient statistic for θ. is the statistic complete?

Sol^n

$X_1, X_2, \ldots, X_n \sim \text{Unif}(\theta, 2\theta)$

$\Rightarrow \theta < x_1, x_2, \ldots, x_n < 2\theta$

$\theta < \min x_i < \max x_i < 2\theta$

$\Rightarrow \theta < \min x_i, \frac{\max x_i}{2} < \theta$

$$\Rightarrow f(x|\theta) = \left(\frac{1}{\theta}\right)^n I_{(\theta,\infty)}(\min x_i)\, I_{(0,\theta)}\left(\frac{\max x_i}{2}\right)$$

$\Rightarrow T(x) = \left(\min x_i, \frac{\max x_i}{2}\right)$ sufficient statistic let X, Y be random samples

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{I_{(\theta,\infty)}(\min x_i)\, I_{(0,\infty)}\left(\frac{\max x_i}{2}\right)}{I_{(\theta,\infty)}(\min y_i)\, I_{(0,\infty)}\left(\frac{\max y_i}{2}\right)}$$

for this function to be independent of θ

then $\min x_i = \min y_i$

$\max x_i = \max y_i$

$\Rightarrow T(x) = (X_{(1)}, X_{(n)})$ minimal sufficient statistic.

is the statistic complete?

Soln

$X \sim \text{unif}(\theta, 2\theta)$

$\Rightarrow$ Unif $(\theta, 2\theta)$ Scale family distribution.

$Z \sim$ Unif$(1,2)$

$\Rightarrow$ $X_{(1)} = \theta Z_{(1)}$

$X_{(2)} = \theta Z_{(2)}$

$\vdots$

$X_{(n)} = \theta Z_{(n)}$

the $T(x) = \left(\theta Z_{(1)}, \theta Z_{(n)}\right)$

is minimal sufficient statistic

$$\Rightarrow g\left(T_1(x), T_2(x)\right) = \frac{T_1(x)}{T_2(x)}$$

$$= \frac{\theta Z_{(1)}}{\theta Z_{(n)}}$$

$$= \frac{Z_{(1)}}{Z_{(n)}}$$

Independent of θ

$\Rightarrow \dfrac{X_{(1)}}{X_{(n)}}$ is a ancillary statistic

$\Rightarrow T(X) = (X_{(1)}, X_{(n)})$

is not complete.

6.24 Consider the following family of distribution's.

$$P = \left\{ P_\lambda(X=x) : P_\lambda(X=x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0,1,2,\dots; \quad \lambda = 0 \text{ or } 1 \right\}$$

This is a Poisson family with λ restricted to be 0 or 1. Show that the family P is not complete, demonstrating that completeness can be dependent on the range of the Parameter

Sol^n

let's take a function $g(x)$

$$\mathbb{E}[g(x) \mid \lambda] = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} g(x)$$

if $\lambda = 0$

then $\mathbb{E}[g(x) \mid \lambda = 0] = 0$

if $\lambda = 1$

$$\mathbb{E}[g(x) \mid \lambda = 1] = \frac{1}{e} \sum_{x=0}^{\infty} \frac{g(x)}{x!}$$