Autumn Semester 2021 **December 31, 2021**

IE 605: Engineering Statistics

Solutions of tutorial 8

Solution 1

The sample density is given by

$$\prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \frac{1}{2i\theta} I(-i(\theta-1) \leq x_i \leq i(\theta+1))$$
$$= \left(\frac{1}{2\theta}\right)^n \left(\prod_{i=1}^{n} \frac{1}{i}\right) I\left(\min \frac{x_i}{i} \geq -(\theta-1)\right) I\left(\max \frac{x_i}{i} \leq \theta+1\right).$$

Thus $(\min \frac{X_i}{i}, \max \frac{X_i}{i})$ is sufficient for θ.

Solution 2

Let X_1, X_2 be iid $Poisson(\lambda)$ RVs.

1. To check: $T_1 = X_1 + X_2$ is sufficient for λ or not.

   Solution:

$$\mathbb{P}\{X_1 = x_1, X_2 = x_2 | T = t\} = \mathbb{P}\{X_1 = x_1, X_2 = x_2 | X_1 + X_2 = t\}$$
$$= \begin{cases} \frac{\mathbb{P}\{X_1 = x_1, X_2 = t - x_1\}}{\mathbb{P}\{X_1 + X_2 = t\}}, & \text{if } t = x_1 + x_2, x_i = 0, 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

   Thus, for $x_i = 0, 1, 2, \ldots, i = 1, 2, x_1 + x_2 = t$, we have

$$\mathbb{P}\{X_1 = x_1, X_2 = x_2 | X_1 + X_2 = t\} = \binom{t}{x_1} \left(\frac{1}{2}\right)^t,$$

   which is independent of λ. Hence, T is sufficient statistics.

2. To check: $T_2 = X_1 + 2X_2$ is sufficient for λ or not.

   Solution:

$$\mathbb{P}\{X_1 = 0, X_2 = 1 | X_1 + 2X_2 = 2\} = \frac{\mathbb{P}\{X_1 = 0, X_2 = 1\}}{\mathbb{P}\{X_1 + 2X_2 = 2\}}$$
$$= \frac{e^{-\lambda}(\lambda e^{-\lambda})}{\mathbb{P}\{X_1 = 0, X_2 = 1\} + \mathbb{P}\{X_1 = 2, X_2 = 0\}}$$
$$= \frac{\lambda e^{-2\lambda}}{\lambda e^{-2\lambda} + (\lambda^2/2) e^{-2\lambda}}$$

$$= \frac{1}{1 + (\lambda/2),}$$

and we see that $T = X_1 + 2X_2$ is not sufficient for λ.

Solution 3

To show: $T = \sum_{i=1}^{n} X_i$ is sufficient for θ

We have

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n | T = t) = \frac{P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)}{P(T = t)}$$

Bearing in mind that the X_i can take on only the values 0s or 1s, the probability in the numerator is the probability that some particular set of t X_i are equal to 1s and the other $n - t$ are 0s. Since the X_i are independent, the probability of this is $\theta^t(1-\theta)^{n-t}$. To find the denominator, note that the distribution of T, the total number of ones, is binomial with n trials and probability of success θ. Therefore the ratio in the above equation is

$$\frac{\theta^t(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$

The conditional distribution thus does not involve θ at all. Given the total number of ones, the probability that they occur on any particular set of t trials is the same for any value of θ so that set of trials contains no additional information about θ.

To show: $T = \sum_{i=1}^{n} X_i$ is complete

$$\mathbb{E}\left[g(T)\right] = \sum_{t=0}^{n} g(t)\binom{n}{t}\theta^t(1-\theta)^{n-t} = 0 \qquad \text{for all } \theta \in (0,1)$$

may be rewritten as

$$(1-\theta)^n \sum_{t=0}^{n} g(t)\binom{n}{t}\left(\frac{\theta}{1-\theta}\right)^t = 0 \qquad \text{for all } \theta \in (0,1).$$

This is a polynomial in $\frac{\theta}{1-\theta}$. Hence the coefficients must vanish, and it follows that $g(t) = 0$ for $t = 0, 1, 2, \ldots, n$, as required. Hence $T = \sum_{i=1}^{n} X_i$ is complete.

Solution 4

For $x = (x_1, ..., x_n)$, the joint pdf of $X_1, ..., X_n$ is

$$f_n(\mathbf{x}|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

This could be rewritten as

$$f_n(\mathbf{x}|\mu) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right) \exp\left(\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2}\right)$$

It can be seen that $f_n(\mathbf{x}|\mu)$ has now been expressed as the product of a function that does not depend on μ and a function the depends on **x** only through the value of $\sum_{i=1}^n x_i$. It follows from the factorization theorem that $T = \sum_{i=1}^n X_i$ is a sufficient statistic for μ.

Since $\sum_{i=1}^n x_i = n\bar{x}$, we can state equivalently that the final expression depends on **x** only through the value of $\bar{x}$, therefore $\bar{X}$ is also a sufficient statistic for μ. More generally, every one to one function of $\bar{X}$ will be a sufficient statistic for μ.

Solution 5

The p.d.f. $f(x|\beta)$ of each individual observation X_i is

$$f(x|\beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the joint p.d.f. $f_n(x|\beta)$ of $X_1, X_2, ..., X_n$ is

$$f(\mathbf{x}|\beta) = \prod_{i=1}^n \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1}$$

$$\begin{aligned} f(\mathbf{x}|\beta) &= \prod_{i=1}^n \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1} \\ &= \Gamma(\alpha)^{-n}\left(\prod_{i=1}^n x_i\right)^{\alpha-1}\left[\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right)^n\left(\prod_{i=1}^n (1-x_i)\right)^{\beta-1}\right] \end{aligned}$$

We define $T' = (X_1, X_2, ...X_n) = \prod_{i=1}^n (1 - X_i)$, and because α is known, so we can define

$$u(\mathbf{x}) = \Gamma(\alpha)^{-n}\left(\prod_{i=1}^n x_i\right)^{\alpha-1}, \quad v(T', \beta) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right)^n T'(x_1, x_2, ..., x_n)^{\beta-1}$$

We can see that the function v depends on **x** only through T', therefore T' is a sufficient statistic. It is easy to see that

$$T = g(T') = \frac{\log(-T')^3}{n}$$

and the function g is a one-to-one mapping. Therefore T is a sufficient statistic.

Solution 6

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

$$= e^{-\sum_{i=1}^{n}(y_i-x_i)} \frac{\prod_{i=1}^{n} 1+e^{-(y_i-\theta)}}{\prod_{i=1}^{n} 1+e^{-(x_i-\theta)}}$$

This is constant as a function of θ if and only if x and y have the same order statistics. Therefore, the order statistics are minimal sufficient for θ.

Solution 7

Suppose $X_1, X_2, \ldots, X_n$ are iid uniform observation on the interval $(\theta, \theta+1)$, $-\infty < \theta < \infty$. Thus the joint pdf of X is,

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

which can be written as,

$$f(x|\theta) = \begin{cases} 1, & \max_i x_i - 1 < \theta < \min_i x_i \\ 0, & \text{otherwise} \end{cases}$$

Thus for two sample points x and y, the numerator and denominator of the ratio $\frac{f(x|\theta)}{f(y|\theta)}$ will be positive for the same values of θ if and only if $\min_i x_i = \min_i y_i$ and $\max_i x_i = \max_i y_i$. And, if the minima and maxima are equal, then the ratio is constant and, in fact, equals 1. Thus, letting $X_{(1)} = \min_i X_i$ and $X_{(n)} = \max_i X_i$, we have $T(X) = (X_{(1)}, X_{(n)})$ is a minimal sufficient statistic. This is a case in which the dimension of a minimal sufficient statistic does not match the dimension of the parameter.

To prove $T(X) = (X_{(1)}, X_{(n)})$ is not complete, we want to find $g[T(X)]$ such that $\mathbb{E}\left[g[T(X)]\right] = 0$ for all θ, but $g[T(X)] \neq 0$. A natural candidate is $R = X_{(n)} - X_{(1)}$, the range of R does not depend on θ [Verify]. It can be shown that $R \sim beta(n-1, 2)$ [Verify]. Thus $\mathbb{E}\left[R\right] = (n-1)/(n+1)$ does not depend on θ, and $\mathbb{E}\left[R - \mathbb{E}\left[R\right]\right] = 0$ for all θ.

Thus

$$g[X_{(n)}, X_{(1)}] = X_{(n)} - X_{(1)} - (n-1)/(n+1) = R - \mathbb{E}\left[R\right]$$

is a non-zero function whose expected value is always 0. So, $(X_{(1)}, X_{(n)})$ is not complete.

NOTE: This problem can be generalized to show that if a function of a sufficient statistic is ancillary, then the sufficient statistic is not complete, because the expectation of that function does not depend on θ. That provides the opportunity to construct an unbiased, nonzero estimator of zero.

Solution 8

Let $X_1, X_2, \ldots, X_n$ be a sample from $N(\theta, \theta^2)$ where $\theta > 0$.

To show: $T = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ is sufficient for θ but T is not complete.

Solution: The joint distribution function of $X_1, \ldots, X_n$ is

$$f_\theta(x_1, \ldots, x_n) = \frac{1}{(2\pi\theta^2)^n} exp\left\{\frac{-1}{2\theta^2}\sum_{i=1}^{n} x_i^2 + \frac{1}{\theta}\sum_{i=1}^{n} x_i - \frac{1}{2}\right\}.$$

By factorisation theorem, $T = (\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i)$ is sufficient statistic.

Note that

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=}^{n} X_i^2\right] &= n\mathbb{E}\left[X_1^2\right] \\
&= 2n\theta^2 \\
\text{and } \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]^2 &= n\theta^2 + (n\theta)^2 \\
&= (n + n^2)\theta^2. \\
\text{Let } h(t_1, t_2) &= \frac{1}{2n}t_1 - \frac{1}{n(n+1)}t_2^2. \\
\text{Then } h(t_1, t_2) \neq 0 \text{ but } \mathbb{E}\left[h(T)\right] &= 0 \text{ for any } \theta.
\end{aligned}$$

Hence T is not complete.

Solution 9

Let $X_1, X_2, \ldots, X_n$ be a sample from $N(\theta, \alpha\theta^2)$, where α is known constant and $\theta > 0$.

To show: $T = (\bar{X}, S^2)$ is sufficient statistics for θ.

Solution: the joint pdf of $x = (X_1, X_2, \ldots, X_n)$ is given by

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

By Factorization Theorem, the joint pdf can be written as

$$
\begin{aligned}
& f_\theta(x) = \left(\frac{1}{2\pi\alpha\theta^2}\right)^{n/2} \exp\left\{\frac{-1}{2\alpha\theta^2}\left((n-1)t_2 + n(t_1-\theta)^2\right)\right\} \\
\implies & f_\theta(x) = g(T_1(x), T_2(x)|\theta)h(x)
\end{aligned}
$$

where $T_1 = \bar{X}, T_2 = S^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$ and $h(X) = 1$.

Therefore, by Factorisation Theorem, $T(X) = (T_1(X), T_2(X)) = (\bar{X}, S^2)$ is a sufficient statistic for θ.

To show: $T = (\bar{X}, S^2)$ is not complete.

Solution:

$$\mathbb{E}\left[S^2\right] = \alpha\theta^2 \text{ and } \mathbb{E}\left[\bar{X}^2\right] = Var\bar{X} + (\mathbb{E}\left[\bar{X}\right])^2 = \frac{\alpha\theta^2}{n} + \theta^2 = \frac{(\alpha+n)\theta^2}{n}.$$

Therefore, $\mathbb{E}\left[\frac{n}{\alpha+n}\bar{X}^2 - \frac{S^2}{\alpha}\right] = \left(\frac{n}{\alpha+n}\right)\left(\frac{\alpha+n}{n}\theta^2\right) - \frac{1}{\alpha}\alpha\theta^2 = 0, \qquad$ for all θ.

Thus $g(\bar{X}, S^2) = \dfrac{n}{\alpha+n}\bar{X}^2 - \dfrac{S^2}{\alpha}$ has zero expectation

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

Solution 10

Let $Y_1 = \log(X_1)$ and $Y_2 = \log(X_2)$. Then Y_1 and Y_2 are iid and, the pdf of each is

$$
\begin{aligned}
f(y|\alpha) &= \alpha\exp\{\alpha y - e^{\alpha y}\} \\
&= \frac{1}{1/\alpha}\exp\left\{\frac{y}{1/\alpha} - e^{y/(1/\alpha)}\right\}, -\infty < y < \infty.
\end{aligned}
$$

We see that the family of distributions of Y_i is a scale family with scale parameter $1/\alpha$. Thus, by the following Theorem which states that,

Theorem 1. *Let $f(.)$ be the pdf. Let μ be any real number, and let σ be any positive real number. Let X is a random variable with pdf $(1/\sigma)f\left(\frac{x-\mu}{\sigma}\right)$ iff there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.*

We can write $Y_i = \frac{1}{\alpha}Z_i$, where Z_1 and Z_2 are a random sample from $f(z|1)$. Then

$$
\begin{aligned}
\frac{\log X_1}{\log X_2} &= \frac{Y_1}{Y_2} \\
&= \frac{(1/\alpha)Z_1}{(1/\alpha)Z_2} \\
&= \frac{Z_1}{Z_2}
\end{aligned}
$$

Because the distribution of $\frac{Z_1}{Z_2}$ does not depend on α, $\frac{\log X_1}{\log X_2}$ is an ancillary statistic.

Solution 11

Suppose $X_1, X_2, \ldots, X_n$ are iid uniform observation on the interval $(\theta, \theta+1)$, $-\infty < \theta < \infty$. Thus the joint pdf of X is,

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

which can be written as,

$$f(x|\theta) = \begin{cases} 1, & \max_i x_i - 1 < \theta < \min_i x_i \\ 0, & \text{otherwise} \end{cases}$$

Thus for two sample points x and y, the numerator and denominator of the ratio $\frac{f(x|\theta)}{f(y|\theta)}$ will be positive for the same values of θ if and only if $\min_i x_i = \min_i y_i$ and $\max_i x_i = \max_i y_i$. And, if the minima and maxima are equal, then the ratio is constant and, in fact, equals 1. Thus, letting $X_{(1)} = \min_i X_i$ and $X_{(n)} = \max_i X_i$, we have $T(X) = (X_{(1)}, X_{(n)})$ is a minimal sufficient statistic. This is a case in which the dimension of a minimal sufficient statistic does not match the dimension of the parameter.

A minimal sufficient statistics is not unique. A one-to-one function of minimal sufficient statistic is also a minimal sufficient statistic.
So, $T(X) = ((X_{(n)} - X_{(1)}), (X_{(n)} + X_{(1)})/2)$ is also a minimum sufficient statistics.

Solution 12

Let $X_1, X_2, \ldots, X_n$ be i.i.d. Uniform observations on the interval $(\theta, \theta+1)$, $-\infty < \theta < \infty$. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics from the sample. We show below that the range statistics $R = X_{(n)} - X_{(1)}$ is an ancillary statistic by showing that the pdf of R does not depend on θ.

The cdf of each X_i is

$$F(x|\theta) = \begin{cases} 0, & x \leq \theta \\ x - \theta, & \theta < x < \theta + 1 \\ 1, & \theta + 1 \leq x \end{cases}$$

Thus, the joint pdf of $X_{(1)}$ and $X_{(n)}$

$$g_{(x_{(1)}, x_{(n)})}(x_1, x_n|\theta) = \begin{cases} n(n-1)(x_n - x_1)^{n-2}, & \theta < x_1 < x_n < \theta + 1 \\ 0, & \text{otherwise} \end{cases}$$

Making the transformation $R = X_{(n)} - X_{(1)}$ and $(M = X_{(n)} + X_{(1)})/2$ which has the inverse transformation $X_{(1)} = (2M - R)/2$ and $X_{(n)} = (2M + R)/2$ with

Jacobin 1, joint pdf of R and M :

$$h(r,m|\theta) = \begin{cases} n(n-1)r^{n-2}, & 0 < r < 1, \theta + (r/2) < m < \theta + 1 - (r/2) \\ 0, & \text{otherwise} \end{cases}$$

Thus pdf for R is:

$$\begin{aligned} h(r|\theta) &= \int_{\theta+(r/2)}^{\theta+1-(r/2)} n(n-1)r^{n-2} dm \\ &= n(n-1)r^{n-2}(1-r) \end{aligned}$$

This is pdf with $\alpha = n-1$ and $\beta = 2$. More important, the pdf is the same for all θ. Thus, the distribution of R does not depend on θ, and R is ancillary.

Solution 13

- (a) Using the definition of minimal sufficient statistics,

$$\begin{aligned} \frac{f(x,n|\theta)}{f(y,n'|\theta)} &= \frac{f(x|\theta, N=n)P(N=n)}{f(y|\theta, N=n')P(N=n')} \\ &= \frac{\binom{n}{x}\theta^x(1-\theta)^{n-x}p_n}{\binom{n'}{y}\theta^y(1-\theta)^{n'-y}p_{n'}} \\ &= \theta^{x-y}(1-\theta)^{n-n'-x+y}\frac{\binom{n}{x}p_n}{\binom{n'}{x}p_{n'}} \end{aligned}$$

  The last ratio does not depend on θ. The other terms are constant as a function of θ if and only if $n = n'$ and $x = y$. So (X, N) is minimal sufficient for θ. Because $\mathbb{P}\{N = n\} = p_n$ does not depend on θ, N is ancillary for θ. The point is that although N is independent of θ, the minimal sufficient statistic contains N in this case. A minimal sufficient statistic may contain an ancillary statistic.

- (b)

$$\begin{aligned} \mathbb{E}\left[\frac{X}{N}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{X}{N}\middle| N\right]\right] \\ &= \mathbb{E}\left[\frac{1}{N}\mathbb{E}[X|N]\right] \\ &= \mathbb{E}\left[\frac{1}{N}N\theta\right] \\ &= \theta \end{aligned}$$

$$\begin{aligned}
Var\left(\frac{X}{N}\right) &= Var\left(\mathbb{E}\left[\frac{X}{N}\middle|N\right]\right) + \mathbb{E}\left[Var\left(\frac{X}{N}\middle|N\right)\right] \\
&= Var\left(\theta\right) + \mathbb{E}\left[\frac{1}{N^2}Var\left(X|N\right)\right] \\
&= 0 + \mathbb{E}\left[\frac{N\theta(1-\theta)}{N^2}\right] \\
&= \theta(1-\theta)\mathbb{E}\left[\frac{1}{N}\right]
\end{aligned}$$

We used the fact that $X|N \sim binomial(N, \theta)$.

Solution 14

The likelihood function is given by,

$$L(\theta|x,n) = f(x,n|\theta) = f(x|\theta, N=n)\mathbb{P}\left\{N=n\right\} = \binom{n}{x}\theta^x(1-\theta)^{n-x}p_n \tag{1}$$

. It can be said that $X|N = n \sim Bin(n, \theta)$. Therefore, the likelihood function of $X|N = n$ is given by,

$$L(\theta, N=n|x) = f(x|\theta, N=n) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \tag{2}$$

. For a fixed sample point (x, n) by eq. (1) and (2) we get,

$$L(\theta|x,n) = p_n L(\theta, N=n|x) \Rightarrow L(\theta|x,n) \infty L(\theta, N=n|x)$$

Therefore, it implies that by the Formal Likelihood Principle we can conclude that θ should not depend on the fact that the sample size n was chosen randomly.

Solution 15

Let $1 =$ success and $0 =$ failure. The four sample points are $\{0, 10, 110, 111\}$. From the likelihood principle, inference about p is only through $L(p|x)$. The values of the likelihood are $1, p, p^2,$ and p^3, and the sample size does not directly influence the inference.

Solution 16

Let X have a negative binomial distribution with $r = 3$ and success probability p. If $x = 2$ is observed, then the likelihood function is the fifth-degree polynomial on $0 \leq p \leq 1$ defined by

$$L(p|2) = P_p(X=2) = \binom{4}{2}p^3(1-p)^2.$$

In general, if $X = x$ is observed, then the likelihood function is the polynomial of degree $3 + x$,

$$L(p|x) = \binom{3+x-1}{x} p^3(1-p)^x.$$

Solution 17

- (a) This pdf can be written as

  $$f(x|,\lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2} \left(\frac{1}{x^3}\right)^{1/2} \exp\left(\frac{\lambda}{\mu}\right) \exp\left\{-\frac{\lambda}{2\mu^2}x - \frac{\lambda}{2}\frac{1}{x}\right\}$$

  This is an exponential family with $t_1(x) = x$ and $t_2(x) = 1/x$.

  We use the Theorem of the property of complete statistics of the exponential family as stated below,

  Theorem 2. *Let $X_1, X_2, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form*

  $$f(x\theta) = h(x)c(\theta)\exp\left(\sum_{j=1}^{k} w(\theta_j)t_j(x)\right)$$

  where $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. Then the statistics

  $$T(X) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots, \sum_{i=1}^{k} t_k(X_i)\right)$$

  is complete as long as the parameter space Θ contains an open set $\mathcal{R}^k$.

  By using the above theorem, the statistic $(\sum_i X_i, \sum_i(1/X_i))$ is a complete sufficient statistic. $(\bar{X}, T)$ given in the problem is a one-to-one function of $(\sum_i X_i, \sum_i(1/X_i))$. Thus, $(\bar{X}, T)$ is also a complete sufficient statistic.

- This can be accomplished using the methods from (Refer Section 4.3 of the book Statistical Inference by George Casella) by a straightforward but messy two-variable transformation $U = (X_1 + X_2)/2$ and $V = 2\lambda/T = \lambda[(1/X_1) + (1/X_2)(2/[X_1 + X_2])]$. This is a two-to-one transformation.

Solution 18

Suppose that $(X_1, Y_1), ..., (X_n, Y_n)$ are independent and identically distributed random 2-vectors having the normal distribution with $\mathbb{E}[X_1] = \mathbb{E}[Y_1] = 0, Var(X_1) = Var(Y_1) = 1$, and $Cov(X_1, Y_1) = \theta \in (-1, 1)$.

1. <u>To find:</u> Find a minimal sufficient statistic for θ.

Solution: The joint distribution function of $(X_1, Y_1), \ldots, (X_n, Y_n)$ is

$$\left(\frac{1}{2\pi\sqrt{1-\theta^2}}\right)^n \exp\left\{-\frac{1}{1-\theta^2}\sum_{i=1}^{n}(x_i^2+y_i^2)+\frac{2\theta}{1-\theta^2}\sum_{i=1}^{n}x_iy_i\right\}.$$

Let $\eta = \left(-\frac{1}{1-\theta^2}, \frac{2\theta}{1-\theta^2}\right)$.

The parameter space $\{\eta : -1 < \theta < 1\}$ is a curve in R^2. We can find that $(T_1, T_2) = \left(\sum\limits_{i=1}^{n}(X_i^2+Y_i^2), \sum\limits_{i=1}^{n}X_iY_i\right)$ is minimal sufficient (Check!).

2. To show: The minimal sufficient statistic obtained in (i) is complete or not.

   Solution:

   Here (X, Y) follows Bivariate Normal distribution with parameters as $(0, 0; 1, 1; \theta)$. By the properties of Bivariate Normal distribution, $X + Y \sim Normal(0, 2(1+\theta)) \implies \sum\limits_{i=1}^{n}\frac{(X_i+Y_i)^2}{2(1+\theta)} \sim \chi^2_n$.

$$\begin{aligned}\text{Note that } \mathbb{E}\left[\frac{1}{1+\theta}\left(\sum_{i=1}^{n}(X_i^2+Y_i^2)+2\sum_{i=1}^{n}X_iY_i\right)\right]-2n&=0,\\ \mathbb{E}\left[\frac{1}{1+\theta}(T_1+2T_2)\right]-2n&=0,\\ \text{but } \frac{1}{1+\theta}(T_1+2T_2)-2n&\neq 0\end{aligned}$$

   Therefore, the minimal sufficient statistic is not complete.

3. To prove: $T_1 = \sum\limits_{i=1}^{n}X_i^2$ and $T_2 = \sum\limits_{i=1}^{n}Y_i^2$ are both ancillary but (T_1, T_2) is not ancillary.

   Solution: Both T_1 and T_2 have the chi-square distribution χ^2_n, which does not depend on θ. Hence both T_1 and T_2 are ancillary. Note that

$$\begin{aligned}\mathbb{E}[T_1T_2] &= \mathbb{E}\left[\left(\sum_{i=1}^{n}X_i^2\right)\left(\sum_{j=1}^{n}Y_j^2\right)\right]\\ &= \mathbb{E}\left[\sum_{i=1}^{n}X_i^2Y_i^2\right]+\mathbb{E}\left[\sum_{i\neq j}X_i^2Y_j^2\right]\\ &= n\mathbb{E}\left[X_1^2Y_1^2\right]+n(n-1)\mathbb{E}\left[X_1^2\right]\mathbb{E}\left[Y_1^2\right]\\ &= n(1+2\theta^2)+2n(n-1),\end{aligned}$$

   which depends on θ. Therefore the distribution of (T_1, T_2) depends on θ and (T_1, T_2) is not ancillary.

Solution 19

Given that T_1 is sufficient and T_2 is minimal sufficient, U is an unbiased estimator of θ, and define $U_1 = \mathbb{E}[U|T_1]$ and $U_2 = \mathbb{E}[U|T_2]$.

- (a) This is seen by first noting that because $T_2 = \phi(T_1)$ for some function ϕ, then

$$U_2 = \mathbb{E}[U|T_2] = \mathbb{E}[\mathbb{E}[U|T_2]|T_1] = \mathbb{E}[\mathbb{E}[U|T_1]|T_2] = \mathbb{E}[U_1|T_2].$$

- (a) Hence, by applying the we obtain

$$Var_\theta(U_1) = \mathbb{E}[Var(U_1|T_2)] + Var(\mathbb{E}[U_1|T_2]) \geq Var(\mathbb{E}[U_1|T_2]) = Var(U_2).$$

  If T_1 and T_2 are both minimally sufficient, then there is a one-to-one function such that $T_2 = \phi(T_1)$, so it follows that $U_1 = U_2$.

Solution 20

Let $(X_1, ..., X_n)$ be a random sample of random variables having the Cauchy distribution with location parameter μ and scale parameter σ, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are unknown parameters.

To show: The vector of order statistics is minimal sufficient for (μ, σ).

Solution: The joint distribution function of $(X_1, \ldots, X_n)$ is

$$f_{\mu,\sigma}(x) = \frac{\sigma^n}{\pi^n} \prod_{i=1}^{n} \frac{1}{\sigma^2 + (x_i - \mu)^2}, \qquad x = (x_1, \ldots, x_n).$$

For any $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, suppose that

$$\frac{f_{\mu,\sigma}(x)}{f_{\mu,\sigma}(y)} = \psi(x, y)$$

holds for any μ and σ, where ψ does not depend on (μ, σ). Let $\sigma = 1$. Then we must have

$$\prod_{i=1}^{n} [1 + (y_i - \mu)^2] = \psi(x, y) \prod_{i=1}^{n} [1 + (x_i - \mu)^2]$$

for all μ. Both sides of the above identity can be viewed as polynomials of degree $2n$ in μ. Comparison of the coefficients to the highest terms gives $\psi(x, y) = 1$. Thus,

$$\prod_{i=1}^{n} [1 + (y_i - \mu)^2] = \prod_{i=1}^{n} [1 + (x_i - \mu)^2]$$

for all μ. As a polynomial of μ, the left-hand side of the above identity has $2n$ complex roots $x_i \pm \sqrt{-1}, i = 1, \ldots, n$, while the right-hand side of the above identity has $2n$ complex roots $y_i \pm \sqrt{-1}, i = 1, \ldots, n$. By the unique factorization theorem for the entire functions in complex analysis, we conclude that the two sets of roots must agree. This means that the ordered values of x_i's are the same as the ordered values of y_i's. Therefore, the order statistics of $X_1, \ldots, X_n$ is minimal sufficient for (μ, σ).