HW4

Sam Olson

Outline

- Q1: Draft
- Q2: Draft
- Q3: Draft
- Q4: Draft
- Q5: Draft

Problem 1

Problem 6.2, Casella and Berger (2nd Edition)

6.2 Let $X_1, \ldots, X_n$ be independent random variables with densities

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

Prove that $T = \min_i(X_i/i)$ is a sufficient statistic for θ.

By the Factorization Theorem, $T(X) = \min(X_i/i)$ is sufficient because the joint pdf is

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} e^{i\theta - x_i} I_{(\theta,+\infty)}(x_i) = e^{in\theta} I_{(\theta,+\infty)}(T(X)) \cdot \underbrace{e^{-\sum_i x_i}}_{h(x)}.$$

Notice, we use the fact that $i > 0$, and the fact that all x_is $> i\theta$ if and only if $\min(x_i/i) > \theta$.

Problem 2

Example of Rao-Blackwell theorem, which is largely a STAT 5420 problem in computation.

Let X_1 and X_2 be iid Bernoulli(p), $0 < p < 1$.

a)

Show $S = X_1 + X_2$ is sufficient for p.

By the Factorization Theorem, the joint pmf of X_1, X_2 is

$$f(x_1, x_2|p) = p^{x_1}(1-p)^{1-x_1} \cdot p^{x_2}(1-p)^{1-x_2}$$

$$= p^{x_1+x_2}(1-p)^{2-(x_1+x_2)} = g(S,p)h(x_1,x_2),$$

where $S = X_1 + X_2$. Thus, S is sufficient for p.

b)

Identify the conditional probability $P(X_1 = x|S = s)$; you should know which values of x, s to consider.

We compute $P(X_1 = x|S = s)$ for possible values of x and s:

$$P(X_1 = x|S = s) = \frac{P(X_1 = x, X_1 + X_2 = s)}{P(S = s)}.$$

For $s = 0$, we must have $X_1 = X_2 = 0$, so $P(X_1 = 0|S = 0) = 1$.
For $s = 2$, we must have $X_1 = X_2 = 1$, so $P(X_1 = 1|S = 2) = 1$.
For $s = 1$, possible values of X_1 are 0 and 1, with equal probability:

$$P(X_1 = 1|S = 1) = P(X_1 = 0|S = 1) = \frac{1}{2}.$$

c)

Find the conditional expectation $T \equiv E(X_1|S)$, i.e., as a function of the possibilities of S. Note that T is a statistic.

Using the values computed in part (b),

$$T = E(X_1|S) = \begin{cases} 0, & S = 0 \\ \frac{1}{2}, & S = 1 \\ 1, & S = 2. \end{cases}$$

d)

Show X_1 and T are both unbiased for p.

The expectation of $ X_1 $ is:

$$E_p(X_1) = p.$$

For $ T $,

$$E_p(T) = \sum_{s=0}^{2} E(X_1|S = s)P(S = s).$$

Substituting values,

$$E_p(T) = 0 \cdot (1-p)^2 + \frac{1}{2} \cdot 2p(1-p) + 1 \cdot p^2 = p.$$

Thus, both X_1 and T are unbiased for p.

e)

Show $\text{Var}_p(T) \leq \text{Var}_p(X_1)$, for any p.

Since T is the Rao-Blackwellized estimator of X_1, we apply the Rao-Blackwell theorem, which states that conditioning on a sufficient statistic cannot increase variance:

$$\text{Var}_p(T) \leq \text{Var}_p(X_1).$$

Explicitly computing,

$$\text{Var}_p(X_1) = p(1-p).$$

For T,

$$\text{Var}_p(T) = E_p(T^2) - (E_p(T))^2.$$

Using the values for T,

$$E_p(T^2) = 0^2(1-p)^2 + \left(\frac{1}{2}\right)^2 2p(1-p) + 1^2p^2 = \frac{p(1-p)}{2} + p^2.$$

So,

$$\text{Var}_p(T) = \left(\frac{p(1-p)}{2} + p^2\right) - p^2 = \frac{p(1-p)}{2}.$$

Since

$$\frac{p(1-p)}{2} \leq p(1-p),$$

it follows that $\text{Var}_p(T) \leq \text{Var}_p(X_1)$, as required.

Problem 3

Problem 6.21 a)-b), Casella and Berger (2nd Edition)

6.21 Let X be one observation from the pdf

$$f(x|\theta) = \left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|}, \quad x = -1, 0, 1, \quad 0 \leq \theta \leq 1.$$

a)

Is X a complete sufficient statistic?

X is sufficient because it is the data. To check completeness, calculate

$$Eg(X) = \frac{\theta}{2}g(-1) + (1-\theta)g(0) + \frac{\theta}{2}g(1).$$

If $g(-1) = g(1)$ and $g(0) = 0$, then $Eg(X) = 0$ for all θ, but $g(x)$ need not be identically 0.
So the family is not complete.

b)

Is $|X|$ a complete sufficient statistic?

$|X|$ is sufficient by Theorem 6.2.6, because $f(x|\theta)$ depends on x only through the value of $|x|$.
The distribution of $|X|$ is Bernoulli, because $P(|X| = 0) = 1 - \theta$ and $P(|X| = 1) = \theta$.
By Example 6.2.22, a binomial family (Bernoulli is a special case) is complete.

Problem 4

Problem 6.24, Casella and Berger (2nd Edition)

6.24 Consider the following family of distributions:

$$\mathcal{P} = \{P_\lambda(X = x) : P_\lambda(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}; x = 0, 1, 2, \ldots; \lambda = 0 \text{ or } 1\}.$$

This is a Poisson family with λ restricted to be 0 or 1. Show that the family $\mathcal{P}$ is *not complete*, demonstrating that completeness can be dependent on the range of the parameter. (See Exercises 6.15 and 6.18.)

If $\lambda = 0$, $Eh(X) = h(0)$. If $\lambda = 1$,

$$Eh(X) = e^{-1}h(0) + e^{-1}\sum_{x=1}^{\infty} \frac{h(x)}{x!}.$$

Let $h(0) = 0$ and $\sum_{x=1}^{\infty} \frac{h(x)}{x!} = 0$, so $Eh(X) = 0$ but $h(x) \neq 0$.
(For example, take $h(0) = 0$, $h(1) = 1$, $h(2) = -2$, $h(x) = 0$ for $x \geq 3$.)

Problem 5

Problem 7.57, Casella and Berger (2nd Edition) You may assume $n \geq 3$.

One has to Rao-Blackwellize on the complete/sufficient statistic here

$$\sum_{i=1}^{n+1} X_i.$$

7.57 Let $X_1, \ldots, X_{n+1}$ be iid Bernoulli(p), and define the function $h(p)$ by

$$h(p) = P\left(\sum_{i=1}^{n} X_i > X_{n+1} \middle| p\right),$$

the probability that the first n observations exceed the $(n+1)$st.

a)

Show that

$$T(X_1, \ldots, X_{n+1}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i > X_{n+1} \\ 0 & \text{otherwise} \end{cases}$$

is an unbiased estimator of $h(p)$.

T is a Bernoulli random variable. Hence,

$$E_p T = P_p(T = 1) = P_p\left(\sum_{i=1}^{n} X_i > X_{n+1}\right) = h(p).$$

b)

Find the best unbiased estimator of $h(p)$.

$\sum_{i=1}^{n+1} X_i$ is a complete sufficient statistic for θ, so $E\left(T \middle| \sum_{i=1}^{n+1} X_i\right)$ is the best unbiased estimator of $h(p)$. We have

$$E\left(T \middle| \sum_{i=1}^{n} X_i = y\right) = \frac{P\left(\sum_{i=1}^{n} X_i > X_{n+1} \middle| \sum_{i=1}^{n} X_i = y\right)}{P\left(\sum_{i=1}^{n} X_i = y\right)}.$$

The denominator equals $\binom{n}{y} p^y (1-p)^{n-y}$. If $y = 0$ the numerator is

$$P\left(\sum_{i=1}^{n} X_i > X_{n+1} \middle| \sum_{i=1}^{n+1} X_i = 0\right) = 0.$$

If $y > 0$ the numerator is

$$P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y, X_{n+1} = 0\right) + P\left(\sum_{i=1}^{n} X_i > X_{n+1}, \sum_{i=1}^{n+1} X_i = y, X_{n+1} = 1\right),$$

which equals

$$P\left(\sum_{i=1}^{n} X_i > 0, \sum_{i=1}^{n} X_i = y\right) P(X_{n+1} = 0) + P\left(\sum_{i=1}^{n} X_i > 1, \sum_{i=1}^{n} X_i = y - 1\right) P(X_{n+1} = 1).$$

For all $y > 0$,

$$P\left(\sum_{i=1}^{n} X_i > 0, \sum_{i=1}^{n} X_i = y\right) = P\left(\sum_{i=1}^{n} X_i = y\right) = \binom{n}{y} p^y (1-p)^{n-y}.$$

If $y = 1$ or 2, then

$$P\left(\sum_{i=1}^{n} X_i > 1, \sum_{i=1}^{n} X_i = y - 1\right) = 0.$$

And if $y > 2$, then

$$P\left(\sum_{i=1}^{n} X_i > 1, \sum_{i=1}^{n} X_i = y - 1\right) = P\left(\sum_{i=1}^{n} X_i = y - 1\right) = \binom{n}{y-1} p^{y-1} (1-p)^{n-y+1}.$$

Therefore, the UMVUE is

$$E\left(T \,\middle|\, \sum_{i=1}^{n} X_i = y\right) = \begin{cases} 0, & \text{if } y = 0 \\ \frac{(n-y+1)p}{(1-p)+(n-y+1)p}, & \text{if } y = 1 \text{ or } 2 \\ \frac{(n-y+1)p}{(1-p)+(n-y+1)p} - 1, & \text{if } y > 2. \end{cases}$$