STAT 530 Homework 1

Hannah Butler

1/26/2022

1. (7 pts) Problem 6.2, Casella & Berger:

Let $X_1, \dots, X_n$ be independent random variables 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 $T = \min_i(X_i/i)$ is a sufficient statistic for θ.

To show that T is a sufficient statistic for θ, we can utilize the factorization theorem to show that $f(\boldsymbol{x}|\theta)$ can be factored into a product of a function of $\boldsymbol{x}$ and a function of only the statistic T and θ.

The joint density for $X_1, \dots, X_n$ is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} e^{i\theta - x_i} I(i\theta \leq x_i < \infty)$$

Consider first the case when $f \neq 0$. Then we have that $x_i \geq i\theta$ for $i = 1, 2, \dots, n$. Without loss of generality, let $x_k/k = \min_i(x_i/i)$. Then $x_k/k \geq \theta$ which implies that $x_i/i \geq \theta$ and $x_i \geq i\theta$ for $i = 1, 2, \dots, n$. Therefore the product of indicators can be written as

$$\prod_{i=1}^{n} I(i\theta \leq x_i < \infty) = I(\theta \leq \min_i(x_k/k) < \infty)$$

and the density as

$$f(\boldsymbol{x}|\theta) = \exp\left\{-\sum x_i\right\} \exp\left\{\frac{n(n+1)\theta}{2}\right\} I(\theta \leq \min_i(x_k/k) < \infty)$$

which can indeed be factored into $h(\boldsymbol{x})g(T, \theta)$ with

$$h(\boldsymbol{x}) = \exp\left\{-\sum x_i\right\}$$

and

$$g(T, \theta) = \exp\left\{\frac{n(n+1)\theta}{2}\right\} I(\theta \leq \min_i(x_k/k) < \infty).$$

If $f = 0$, then the factorization is trivial since f can be factored as $h(\boldsymbol{x}) = 0$ and $g(T, \theta)$.

Therefore, $T = \min_i(X_i)$ is a sufficient statistic.

2. (8 pts) Problem 6.7, Casella & Berger: Let $f(x, y|\theta_1, \theta_2, \theta_3, \theta_4)$ be the bivariate pdf for the uniform distribution on the rectangel with lower left corner (θ_1, θ_2) and upper right corner (θ_3, θ_4) in $\mathbb{R}^2$. The parameters satisfy $\theta_1 < \theta_3$ and $\theta_2 < \theta_4$. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from this pdf. Find a four-dimensional sufficient statistic for $(\theta_1, \theta_2, \theta_3, \theta_4)$.

We have

$$f(x, y|\theta_1, \theta_2, \theta_3, \theta_4) = \frac{1}{\theta_2 - \theta_1} I(\theta_1 < x < \theta_2) \frac{1}{\theta_4 - \theta_3} I(\theta_3 < y < \theta_4).$$

So the likelihood function can be written as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y}) &= \prod_{i=1}^{n} \frac{1}{\theta_2 - \theta_1} I(\theta_1 < x_i < \theta_2) \frac{1}{\theta_4 - \theta_3} I(\theta_3 < y_i < \theta_4) \\
&= \left(\frac{1}{\theta_2}\right)^n \left(\frac{1}{\theta_4 - \theta_3}\right)^n \prod_{i=1}^{n} I(\theta_1 < x_i) I(\theta_2 > x_i) I(\theta_3 < y_i) I(\theta_4 > y_i) \\
&= \left(\frac{1}{\theta_2}\right)^n \left(\frac{1}{\theta_4 - \theta_3}\right)^n I(\theta_1 < \min_i(x_i)) I(\theta_2 > \max_i(x_i)) I(\theta_3 < \min_i(y_i)) I(\max_i(y_i))
\end{aligned}$$

Then, we have that the likelihood can be factored $\mathcal{L}(\boldsymbol{\theta}|x, y) = h(\boldsymbol{x}, \boldsymbol{y}) g(\boldsymbol{T}, (\theta))$ where $g(\boldsymbol{T}, \boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{y})$ and $h(\boldsymbol{x}, \boldsymbol{y}) = 1$.

3. (10 pts) Problem 6.9 (b and d only), Casella & Berger: For each of the following distributions, let $X_1, \ldots, X_n$ be a random sample. Find a minimal sufficient statistic for θ.

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

The joint distribution for a sample $\boldsymbol{X}$ can be expressed as

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)} I(\theta < x_i) = e^{n\theta} e^{-\sum x_i} I(\theta < \min_i(x_i)).$$

Then the ratio of distributions from two samples would be written as

$$\frac{f(\boldsymbol{x}|\theta)}{f(\boldsymbol{y}|\theta)} = \frac{e^{n\theta} e^{-\sum x_i} I(\theta < \min_i(x_i))}{e^{n\theta} e^{-\sum y_i} I(\theta < \min_i(y_i))}.$$

Because the $e^{n\theta}$ in the numerator and denominator cancel out, we have that this ratio will be constant with respect to θ only if $\min_i(x_i) = \min_i(y_i)$. Therefore, $T(X) = \min_i(X_i)$ is a minimal sufficient statistic for θ.

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

The joint distribution of a sample $\boldsymbol{X}$ is

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

and the ratio of two samples

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

will be constant only if the order statistics for $\boldsymbol{x}$ are equal to the order statistics for $\boldsymbol{y}$. Therefore, a minimal sufficient statistic is the set of order statistics $(X_{(1)}, \ldots, X_{(n)})$.

4. (10 pts) Problem 6.10, Casella & Berger: Show that the minimal sufficient statistic for the uniform $(\theta, \theta+1)$, found in example 6.2.15 is not complete.

The minimal sufficient statistic found in example 6.2.15 is $T(\boldsymbol{X}) = (X_{(1)}, X_{(n)})$, ie, the minimum and maximum values of the sample. To show that T is not a complete statistic, we must find a function $g(T)$ such that for $E_\theta g(T) = 0$ for all θ, $g \neq 0$ for some θ.

First, using either Theorem 5.4.4 or deriving by hand using the cdf method, we have

$$f_{X_{(1)}}(x|\theta) = n(1 - x + \theta)^{n-1} I(\theta < x < \theta + 1).$$

and

$$f_{X_{(n)}}(x|\theta) = n(x - \theta)^{n-1} I(\theta < x < \theta + 1).$$

Consider the range transformation $R = X_{(n)} - X_{(1)}$. Then the expectation of R is

$$ER = 1 - \frac{2}{n+1}$$

and $E(R - ER) = 0$ for all θ. However, the function $g(\boldsymbol{T}) = X_{(n)} - X_{(1)} - \left(1 - \frac{2}{n+1}\right)$ is not zero for all θ since $\left(1 - \frac{2}{n+1}\right)$ is constant. Therefore, $\boldsymbol{T}$ is not a complete statistic.

5. (10 pts) Problem 6.11, Casella & Berger (only need to consider (b) and (d) in 6.9): Refer to the pdfs given in 6.9. For each, let $X_{(1)} < \cdots < X_{(n)}$ be the ordered sample, and define $Y_i = X_{(n)} - X_{(i)}$.

---

a. For each pdf, verify that the set $(Y_1, \dots, Y_{n-1})$ is ancillary for θ.

For part b of 6.9, the density of $X(i)$ is given by

$$f_{X_{(i)}}(x|\theta) = \frac{n!}{(i-1)!(n-i)!} e^{-(x-\theta)} \left[1 - e^{-(x-\theta)}\right]^{i-1} \left[e^{-(x-\theta)}\right]^{n-i}.$$

This is a member of a location family, so the random variable $Z_{(i)} = X_{(i)} + \theta$ has density

$$f_{Z_{(i)}}(z|\theta) = \frac{n!}{(i-1)!(n-i)!} e^{-x} \left[1 - e^{-x}\right]^{i-1} \left[e^{-x}\right]^{n-i},$$

which does not depend on the parameter θ. Therefore, the variable $Y_{(i)} = X_{(n)} - X_{(i)} = (X_{(n)} + \theta) - (X_{(i)} + \theta) = Z_{(n)} - Z_{(i)}$ does not depend on θ.

For part d of problem 6.9, a similar argument as above can be used in order to show that the distribution of $Y_{(i)}$ does not depend on the parameter θ.

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

By Basu's Theorem (Theorem 6.2.24), a complete and minimal sufficient statistic is independent of every ancillary statistic. So, if T is complete, then it is indeed independent of every $Y(i)$.

For part b, the minimal sufficient statistic is $T(X) = \min_i(X_i)$.

Suppose that we have a function $g(T)$ such that $Eg(T) = 0$ for all values of θ.

$$\begin{aligned} 0 = Eg(T) &= \int_{\theta}^{\infty} g(t) n e^{-(t-\theta)} \left[e^{-(t-\theta)}\right]^{n-1} dt \\ &= -\int_{\infty}^{\theta} g(t) n e^{-(t-\theta)} \left[e^{-(t-\theta)}\right]^{n-1} dt \end{aligned}$$

This expectation is a function of θ, so if we take the derivative of both sides with respect to θ, we get

$$\begin{aligned} 0 &= -\frac{d}{d\theta} \int_{\infty}^{\theta} g(t) n e^{-(t-\theta)} \left[e^{-(t-\theta)}\right]^{n-1} dt \\ &= -g(\theta) n e^{-(\theta-\theta)} \left[e^{-(\theta-\theta)}\right]^{n-1} \\ &= -g(\theta) n \end{aligned}$$

$n \geq 1$, so $g(\theta)$ must be 0 for all values of θ. Therefore, T is a complete statistic and is consequently independent of all of the ancillary statistics $Y_{(i)}$.

For part d, the minimal sufficient statistic is the set of order statistics $(X_{(1)}, \dots, X_{(n)})$. This is not independent of the $Y_{(i)}$.

6. (10 pts) Problem 6.15, Casella & Berger: Let $X_1, \dots, X_n$ be iid $N(\theta, a\theta^2)$, where a is a know constant and $\theta > 0$.

---

a. Show that the parameter space does not contain a two-dimensional open set.

We are estimating 2 parameters, however, the parameter space is defined by the parabola $a\theta^2$. This is a 1-dimensional subset of $\mathbb{R}^2$, and therefore does not have a 2-dimensional open set.

b. Show that the statistic $T = (\bar{X}, S^2)$ is a sufficient statistic for θ, but the family of distributions is not complete.

Using the factorization theorem, we can see that the joint density of the sample indicates that T is sufficient for θ:

$$\begin{aligned} f(\boldsymbol{x}|\theta, a\theta^2) &= \prod_{i=1}^{n} (2\pi a\theta^2)^{-1/2} e^{-\frac{1}{2a\theta^2}(x_i-\theta)^2} \\ &= (2\pi a\theta^2)^{-n/2} e^{-\frac{1}{2a\theta^2}\sum(x_i-\theta)^2} \\ &= (2\pi a\theta^2)^{-n/2} e^{-\frac{1}{2a\theta^2}\left((n-1)s^2+\sum(\bar{x}-\theta)^2\right)}. \end{aligned}$$

However, because the normal distribution is an exponential family, and we do not have a 2-d open set within the parameter space, T is not complete.

7. (15 pts) Problem 6.20, (a), (b) and (d), Casella & Berger: For each of the following pdfs, let $X_1, \ldots, X_n$ be iid observations. Find a complete sufficient statistic, or show that one does not exist.

---

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

The joint pdf is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \frac{2x_i}{\theta^2} I(0 < x_i < \theta) = 2^n (x_1 x_2 \ldots x_n) I(\min_i(x_i) > 0) I(\max_i(x_i) < \theta).$$

From this we can see that a sufficient statistic is $T(\boldsymbol{X}) = \max_i(X_i)$. Now suppose that there is a function $g(T)$ such that $Eg(T) = 0$ for all values of θ. Then we have

$$0 = \int_0^\theta g(t) n \frac{2t}{\theta^2} \left(\frac{t^2}{\theta^2} \right)^{n-1} dt \qquad \text{(Theorem 5.4.4)}$$

Pulling out the terms that do not depend on t in the integral, we have

$$0 = \int_0^\theta g(t) t^{2n-1} dt$$

Since this is a constant function of θ, we take the derivative of both sides with respect to θ to get

$$0 = \frac{d}{d\theta} \int_0^\theta g(t) t^{2n-1} dt = g(\theta) \theta^{2n-1}.$$

Since $\theta^{2n-1} > 0$ for all values of θ, then $g(\theta)$ must always be 0. Therefore, $T = \max_i(X_i)$ is a complete, sufficient statistic.

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

This is an exponential family, with a pdf which can be expressed as

$$f(x|\theta) = \frac{\theta}{1+x} e^{-\theta \log(1+x)}.$$

So the statistic $T = \log(1 + X_i)$ is a complete statistic.

Furthermore, we can use the factorization theorem to show

$$f(\boldsymbol{x}|\theta) = \frac{\theta^n}{\prod_{i=1}^n (1 + x_i)} e^{-\theta \sum \log(1+x_i)},$$

which can be factored into $h(\boldsymbol{x}) g(T, \theta)$. So T is also a sufficient statistic.

d. $f(x|\theta) = e^{-(x-\theta)} \exp\left(-e^{-(x-\theta)}\right), \quad -\infty < x < \infty, -\infty < \theta < \infty.$

Consider the statistic $T = (X_{(1)}, \ldots, X_{(n)})$. This is a minimal sufficient statistic. However, since f is a location family, we have that $R = X_{(n)} - X_{(1)}$ is ancillary. T is not independent of R, so by Basu's Theorem, we conclude that T can not be a complete statistic.

8. (10 pts) Problem 6.21 (a) and (b), Casella & Berger: Let X be one observation from the pdf

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

---

a. Is X a complete sufficient statistic?

X is certainly a sufficient statisic, since our sample includes only one observation.

Suppose now that there is a function $g(X)$ such that $Eg(X) = 0$ for all values of θ. We can then write out the expectation as

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

We can easily construct a non-zero function $g(-1) = -1, g(0) = 0, g(1) = 1$ that will satisfy the statement above. Therefore, X is not a complete statistic.

b. Is $|X|$ a complete sufficient statistic?

Again, it is easy to see using the factorization theorem that $|X|$ is a sufficient statistic. However, the expectation

$$\begin{aligned} Eg(X) &= g(|-1|)\left(\frac{\theta}{2}\right) + g(|0|)(1-\theta) + g(|1|)\left(\frac{\theta}{2}\right) \\ &= g(|-1|)\left(\frac{\theta}{2}\right) + g(|0|)(1-\theta) + g(|1|)\left(\frac{\theta}{2}\right) \\ &= g(1)\theta + g(0)(1-\theta) \end{aligned}$$

will only be zero for all values of θ if $g(0) = g(1) = 0$. So $|X|$ is a complete sufficient statistic.

c. Does $f(x|\theta)$ belong to the exponential class?

Yes. Taking the log and then exponentiatingn f, the pdf can be re-written as

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

9. (10 pts) Problem 6.24, Casella & Berger: Consider the following family of distributions:

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

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.

The joint pmf for a sample $\boldsymbol{X}$ is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{x_1! \ldots x_n!}.$$

A sufficient statistic would be $T = \sum_{i=1}^n x_i$ with $T \sim \text{Poisson}(n\lambda)$.

Suppose that $E_\lambda g(T) = 0$ for all values of λ, (ie $\lambda = 0, 1$). Then we have

$$0 = E_{\lambda=0} g(T) = 0$$

and

$$0 = E_{\lambda=1} g(T) = \sum_{i=0}^{\infty} g(i) \frac{(n\lambda)^i e^{-n\lambda}}{i!} = \sum_{i=0}^{\infty} g(i) \frac{(n)^i e^{-n}}{i!}.$$

If we define $g(t)$ such that $g(0) = 1$, $g(1) = -\frac{e}{n}$, and $g(i) = 0$ for $i = 2, 3, \ldots$, then we have that $E_\lambda g(T) = 0$ for $\lambda = 0, 1$, but $g(T)$ need not be zero. Therefore, $\mathcal{P}$ is not a complete family.

10. (10 pts) Problem 6.30, Casella & Berger: Let $X_1, \ldots, X_n$ be a random sample from the pdf $f(x|\mu) = e^{-(x-\mu)}$, where $-\infty < \mu < x < \infty$.

---

a. Show that $X_{(1)} = \min_i(X_i)$ is a complete sufficient statistic.

The joint distribution for the random sample $\boldsymbol{X}$ is

$$f(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} e^{\mu} e^{-x_i} I(\mu < x_i) = e^{n\mu - \sum x_i} I(\mu < \min_i(x_i)).$$

We can show that $X_{(1)}$ is minimal sufficient since

$$\frac{f(\boldsymbol{x}|\theta)}{f(\boldsymbol{y}|\theta)} = \frac{e^{n\mu - \sum x_i} I(\mu < \min_i(x_i))}{e^{n\mu - \sum y_i} I(\mu < \min_i(y_i))} = \frac{e^{\sum y_i} I(\mu < x_{(1)})}{e^{\sum x_i} I(\mu < y_{(1)})}$$

is constant with respect to θ only if $x_{(1)} = y_{(1)}$.

To show completeness, suppose that $E_\mu g(T) = 0$ for all values of μ. Then we have

$$0 = -\int_{\infty}^{\mu} g(t) n e^{-n(t-\mu)} dt \quad \forall \theta.$$

Pulling out the terms that do not depend on t, we have

$$0 = \int_{\infty}^{\mu} g(t) e^{-nt} dt \quad \forall \theta.$$

Taking the derivative of both sides with respect to μ, we have

$$0 = g(\mu) e^{-n\mu} \quad \forall \theta.$$

$e^{-n\mu} > 0$ for all μ, so $g(\mu)$ must be 0 for all μ. Therefore, $T = X_{(1)}$ is complete *and* minimal sufficient.

b. Use Basu's Theorem to show that $X_{(1)}$ and S^2 are independent.

If we can show that $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2$ is an ancillary statistic, then we can conclude, using Basu's Theorem that $X_{(1)}$ is independent of S^2. To show that S^2 is ancillary, we must show that the distribution of S^2 does not depend on μ. We notice that f is a location family, so we can write $X_i = Z_i + \mu$, where Z_i is a random variable with pdf $f(z|\mu) = e^{-x}$. That is, the distribution of Z_i does not depend on the parameter μ. Then we can write S^2 as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} Z_i^2$$

, ei, as a linear combination of random variables Z_i^2 whose distributions do not depend on μ. Therefore, the distribution of S^2 will also not depend on μ, and S^2 is ancillary. Therefore, it is independent of the complete and minimal sufficient statistic $X_{(1)}$.