SOLUTION FOR HOMEWORK 4, STAT 6331

Welcome to your fourth homework. Reminder: if you find a mistake/misprint, do not e-mail or call me. Write it down on the first page of your solutions and you may give yourself a partial credit — but keep in mind that the total for your homeworks cannot exceed 20 points.

Now let us look at your problems.

1. a) Here $f(x|\theta) = (2\pi)^{1/2} e^{-(x-\theta)^2/2} I(-\infty < x < \infty)$ and $-\infty < \theta < \infty$. Write

$$L(\mathbf{x}, \mathbf{y}|\theta) := \frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = e^{-(1/2)\sum_{l=1}^n (x_l^2 - y_l^2) + \theta n(\bar{x} - \bar{y})}.$$

We see that $L(\mathbf{x}, \mathbf{y}|\theta) \equiv K(\mathbf{x}, \mathbf{y})$ for all θ and all pairs $(\mathbf{x}, \mathbf{y})$ iff $\bar{x} = \bar{y}$. Thus $T = \bar{X}$ is Minimal SS (MSS).

b) Here $f(x|\theta) = e^{-(x-\theta)} I(x > \theta)$ and $\theta \in (-\infty, \infty)$. Write,

$$L(\mathbf{x}, \mathbf{y}|\theta) := \frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{e^{-\sum_{l=1}^n x_l} I(x_{(1)} > \theta)}{e^{-\sum_{l=1}^n y_l} I(y_{(1)} > \theta)}. \quad (1)$$

(i) Suppose that $L(\mathbf{x}, \mathbf{y}|\theta \equiv K(\mathbf{x}, \mathbf{y})$ for all $\theta \in (-\infty, \infty)$. This yields that $x_{(1)} = y_{(1)}$.

(ii) Suppose that $x_{(1)} = y_{(1)}$. Then (1) implies that $L(\mathbf{x}, \mathbf{y}|\theta)$ does not depend on θ.

The properties (i) and (ii) establish, according to Theorem 6.2.13, that $T = X_{(1)}$ is the minimal sufficient statistic.

e) Here $f(x|\theta) = (1/2)e^{-|x-\theta|} I(x \in (-\infty, \infty))$ for $\theta \in (-\infty, \infty)$. Write for ordered observations:

$$L(\mathbf{x}, \mathbf{y}|\theta) := \frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$$

$$= \frac{e^{\sum_{l:\, x_{(l)} \le \theta} (\theta - x_{(l)}) - \sum_{l:\, x_{(l)} > \theta} (x_{(l)} - \theta)}}{e^{\sum_{l:\, y_{(l)} \le \theta} (\theta - y_{(l)}) - \sum_{l:\, y_{(l)} > \theta} (y_{(l)} - \theta)}}.$$

Now note that if at least one pair of ordered observations is such that $x_{(l)} \neq y_{(l)}$ then $L(\mathbf{x}, \mathbf{y}|\theta$ depends on θ. Observing this, Theorem 6.2.13 and some straightfirward steps (similar to the previous case), show that $T = (X_{(1)}, \ldots, X_{(n)})$ is MSS.

2. Exerc. 6.11. (a) All the considered distributions (random variables) are from a so-called location family where

$$X \stackrel{D}{=} Z + \theta, \quad Z \sim f_X(x|\theta = 0). \quad (2)$$

You can aslo think about (2) in the following way: there is a random variable Z with distribution corresponding to $\theta = 0$, and then any other RV from the same distribution with parameter θ is generated as shown in (2).

Let us prove (2) using moment generating function approach (another approach is to calculate $f_X(x|\theta)$ using our techniques in Chapter 2). Recall that $M_X(t) = E(e^{tX})$ and X has the same distribution as Y iff $M_X(t) \equiv M_Y(t)$ for all t in some vicinity of $t = 0$, say $t \in (-\delta, \delta)$, $\delta > 0$.

Thus, to check (2), we need to show that

$$E(e^{tX}) = E(e^{t(Z+\theta)}), \quad t \in (-\delta, \delta) \tag{3}$$

for the case of a location family $f(x|\theta) =: \psi(x - \theta)$. Here $f_Z(z|\theta) = \psi(z)$ is the pdf of Z.

To prove (3) we write

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} \psi(x - \theta)e^{tx}dx$$

[make change of variable $z = x - \theta$ and continue]

$$= \int_{-\infty}^{\infty} \psi(z)e^{t(z+\theta)}dz$$

[now we continue for all t that the integral exists]

$$= E\{e^{t(Z+\theta)}\} = M_{Z+\theta}(t).$$

What was wished to show.

Now, if $Y_i = X_{(n)} - X_{(i)}$ then according to (2) we have $Y_i = Z_{(n)} - Z_{(i)}$, and this shows that the distribution of Y_i does not depend on θ because the distribition of Z does not depend on θ.

We conclude that in all examples of a location family of distributions, statistics Y_i are ancillary for the location parameter θ. By the way, can you propose several other ancillary statistics?

(b)

(i) For the case (a) - normal distribution - the MSS is $\bar{X}$, and it is *complete* because normal distribution belongs to an exponential family. Indeed, we can write

$$f(\mathbf{x}|\theta) = C(\theta)h(\mathbf{x})e^{\theta n\bar{x}}$$

and then use Theorem 6.2.25. Then Basu's Theorem yields that $\bar{X}$ and Y_i are independent.

(ii) For the case (b) - location exponential family - the MSS is $X_{(1)}$. Is it complete? Let us check. Suppose that for a function g

$$E_\theta(g(X_{(1)}) \equiv 0, \quad \theta \in (-\infty, \infty). \tag{4}$$

This is equivalent to

$$\int_{\theta}^{\infty} f_{X_{(1)}}(t|\theta)g(t)dt \equiv 0, \quad \theta \in (-\infty, \infty). \tag{5}$$

Well, now is the time to remember the density for $X_{(1)}$. While I know that you remember the formula, let us one more time deduce it. We begin with the corresponding cdf (cumulative distribution function)

$$F_{X_{(1)}}(t|\theta) = \mathrm{P}(X_{(1)} \le t) = 1 - \mathrm{P}(X_{(1)} > t)$$

[remember that the last probability is called the survivor function]

$$= 1 - \mathrm{P}(X_1 > t, X_2 > t, \dots, X_n > t) = 1 - [\mathrm{P}(X_1 > t)]^n$$

$$= \{1 - [\int_t^\infty e^{-(x-\theta)} dx]^n\} I(t > \theta) = [1 - e^{-n(t-\theta)}] I(t > \theta).$$

Taking the derivative of the cdf we get the pdf

$$f_{X_{(1)}}(t|\theta) = n e^{-n(t-\theta)} I(t > \theta).$$

Now we plug-in this density in (5),

$$n \int_\theta^\infty e^{-n(t-\theta)} g(t) dt \equiv 0, \quad \theta \in (-infty, \infty).$$

The last relation is equivalent to

$$q(\theta) := \int_\theta^\infty e^{-nt} g(t) dty \equiv 0, \ \ \theta \in (-\infty, \infty).$$

From here we get $dq(\theta)/d\theta \equiv 0$ as well. In its turn, this yields (take the derivative)

$$e^{-n\theta} g(\theta) \equiv 0, \quad \theta \in (-\infty, \infty),$$

which finally yields $g(\theta) \equiv 0$ for all $\theta \in (-\infty, \infty)$. We established that $X_{(1)}$ is complete!

Conclusion: By Basu's Theorem the CMSS $X_{(1)}$ is independent of the ancillary statistic $Y := \{X_{(n)} - X_{(i)}, i = 1, 2, \dots, n-1\}$.

REMARK: Do you see that this exponential distribution is not from the exponential family? This is the reason for the direct proof!

(iii) For the double-exponential distribution the statistic $T = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$ is MSS. Clearly it is not independent of $Y := \{X_{(n)} - X_{(i)}, i = 1, 2, \dots, n-1\}$ because if you know T then you know Y.

3. Exers. 6.19. Let us begin with distribution 1. Let $\psi(X)$ be such that $E_p(\psi(X)) \equiv 0$ for all $p \in (0, 1/4)$. This means that

$$p\psi(0) + 3p\psi(1) + (1 - 4p)\psi(2) \equiv 0,$$

or equivalently

$$p[\psi(0) + 3\psi(1) - 4\psi(2)] \equiv -\psi(2), \quad p \in (0, 1/4).$$

This is possible for $\psi(2) = 0$ and $\psi(0) = -3\psi(1) \ne 0$. We conclude that $\psi(k)$ is not necessarily equal to 0, so the family of distributions of X is not complete.

For the second distribution, similarly assume that $E_p\{\psi(X)\} \equiv 0$ for $p \in (0, 1/2)$. This yields

$$p\psi(0) + p^2\psi(1) + (1 - p - p^2)\psi(2) \equiv 0, \quad p \in (0, 1/2).$$

The last relations yields

$$p^2[\psi(1) - \psi(2)] + p[\psi(0) - \psi(2)] + \psi(2) \equiv 0, \;\; p \in (0, 1/2).$$

A polynom is identically equal to zero on an interval iff all its coefficients are zero. This yields that $\psi(k) \equiv 0$.

Conclusion: The distribution is complete.

4. Exerc. 6.20.

(b) We have

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

Note that here $\Omega = (0, \infty)$. Write,

$$f(\mathbf{x}|\theta) = \theta^n \prod_{l=1}^{n} (1 + x_l)^{1+\theta} = \theta^n e^{(1+\theta)\sum_{l=1}^{n} \ln(1+x_l)}.$$

We see that the distribution is from an exponential family and $\Omega := (0, \infty)$ contains an open set, say $(1, 2)$. Thus, by Theorem 6.2.25 the statistic $T := \sum_{l=1}^{n} \ln(1 + X_l)$ is CSS.

(c) Here

$$f(x|\theta) = \log(\theta)(\theta - 1)\theta^x, \quad x \in \Omega := (1, \infty).$$

Write,

$$f(\mathbf{x}|\theta) = [\log(\theta)/(\theta - 1)]^n \prod_{l=1}^{n} \theta^{x_l} = [\log(\theta)/(\theta - 1)]^n e^{\theta \sum_{l=1}^{n} x_l}.$$

This is the distribution from an exponential family and Ω contains an open set, say $(2, 5)$. Thus, by Theorem 6.2.25 the statistic $T = \sum_{l=1}^{n} X_l$ (or $\bar{X}$) is CSS.

(e). Here we have

$$f(x|\theta) = \left(\begin{array}{c} 2 \\ x \end{array} \right) \theta^x (1 - \theta)^{2-x}, \quad x = 0, 1, \ldots, \; \theta \in [0, 1].$$

Write,

$$f(\mathbf{x}|\theta) = \prod_{l=1}^{n} \left(\begin{array}{c} 2 \\ x_l \end{array} \right) \theta^{x_l} (1 - \theta)^{2-x_l}$$

$$= \prod_{l=1}^{n} \left(\begin{array}{c} 2 \\ x_l \end{array} \right) e^{\ln(\theta/(1-\theta)) \sum_{l=1}^{n} x_l} (1 - \theta)^{2n}.$$

This is again the distribution from an exponential class with Ω containing an open interval. This $\bar{X}$ is CSS by Theorem 6.2.25.

5. Exerc. 6.21. X is distributed according to

$$f(x|\theta) = (\theta/2)^{|x|}(1-\theta^{1-|x|}I(x \in \{-1,0,1\}), \quad \theta \in [0,1].$$

(a) X is sufficient (of course - this is the observation) but *not complete*. Let us show this. Suppose that $E_\theta\psi(X)) \equiv 0$, $\theta \in [0,1]$. This is equivalent to

$$(\theta/2)\psi(-1_+(1-\theta)\psi(0) + (\theta/2)\psi(1) \equiv 0, \ \ \theta \in [0,1],$$

or

$$\theta[(1/2)\psi(-1) - \psi(0) + (1/2)\psi(1)] + \psi(0) \equiv 0, \quad \theta \in [0,1].$$

A polynom is zero over an interval iff all its coefficients are zero. This implies that

$$\begin{cases} \psi(0) = 0 \\ (1/2)[\psi(-1) + \psi(1)] - \psi(0) = 0 \end{cases}$$

Solving this system yields $\psi(0) = 0$ and $\psi(-a) = -\psi(1)$. As a results, $\psi(k)$ is not necessarily identical zero. For instance, one may choose $\psi(-1) = 5$, $\psi(0) = 0$ and $\psi(1) = -5$.

b),c) Let us begin with c). Write

$$f(x|\theta) = e^{|x|\ln(\theta/2(1-\theta))}(1-\theta)I(x \in \{-1,0,1\}).$$

This is the distribution from an exponential class. Then by Theorem 6.2.25, the statistic $|X|$ is CSS. Clearly it is minimal as well (what can you reduce here?), or if unclear use Theorem 6.2.28.

6. Exerc. 6.30. We have $f(x|\mu) = e^{-(x-\mu)}I(x > \mu)$, $\mu \in R$.

(a) Write

$$f(\mathbf{x}|\mu) = e^{-\sum_{l=1}^n x_l + n\mu}I(x_{(1)} > \mu).$$

Here $X_{(1)}$ is clearly SS and minimal as well. But completeness we should verify via its definition because *this is not the distribution from an exponential class*, that is, we cannot use Theorem 6.2.25.

Let us find the distribution of $X_{(1)}$; we did it several times already so I just briefly note that

$$\mathrm{P}(X_{(1)} \le t) = [1 - e^{-n(t-\mu)}]I(t \ge \mu),$$

and this yields

$$f_{X_{(1)}}(t) = ne^{-(t-\mu)}I(t \ge \mu).$$

Then is $E_\mu\{\psi(X_{(1)})\} \equiv 0$ then

$$\int_\mu^\infty e^{-n(x-\mu)}\psi(x)dx \equiv 0, \quad \mu \in (-\infty, \infty).$$

Now, the last identity implies that $\psi(\mu) \equiv 0$. To prove this, note that we have $\int_\mu^\infty e^{-nx}\psi(x)dx \equiv 0$ and then take the derivative of the integral with respect to μ. Thus $X_{(1)}$ is complete.

(b) Using notation in the text, $S^2 := [1/(n-1)] \sum_{l=1}^n (X_l - \bar{X}))^2$.

We are dealing with the location parameter here, so we can again use the same trick as before. Set $X = Z + \mu$ where Z is distributed with pdf $f_Z(z) = e^{-z} I(z > 0)$. Then

$$S^2 = (n-1)^{-1} \sum_{l=1}^{n} (Z_l + \mu - (\bar{Z} + \mu))^2 = (n-1)^{-1} \sum_{l=1}^{n} (Z_l - \bar{Z})^2.$$

This shows us that S^2 is ancillary with respect to μ. Then Basu's Theorem implies that $X_{(1)}$ and S^2 are independent.