SOLUTION FOR HOMEWORK 3, STAT 6331

Welcome to your third 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. Let us show that $P((X,Y)=(i,j)|X-Y=0,p)$ depends on p. This will yield that $X-Y$ is not SS. Write (remember that we may write $A\cap B := (A,B)$ to shorten formulas)

$$\mathrm{P}((X,Y)=(i,j)|X-Y=0,p)=\frac{\mathrm{P}((X,Y)=(i,j),X-Y=0|p)}{P(X-Y=0|p)}$$

$$=\frac{\mathrm{P}((X,Y)=(i,j),X=Y|p)}{P(X-Y=0|p)}=\frac{\mathrm{P}(X=i,Y=i|p)I(i=j)}{P(X-Y=0|p)}.$$

We have a specific distribution for the random variables, so let us use it to calculate the probabilities. We have $\mathrm{P}(X=i,Y=i|p)=p^2(1-p)^{2i-2}$ because X and Y are iid Geometric(p). Our next step is to calculate $\mathrm{P}(X-Y=0|p)$. Write

$$\mathrm{P}(X-Y=0|p)=\mathrm{P}(X=Y|p)=\sum_{(i,j):\, i=j,i\geq 1}\mathrm{P}((X,Y)=(i,j)|p)$$

$$=\sum_{i=1}^{\infty}\mathrm{P}(X=i|p)\mathrm{P}(Y=i|p)=\sum_{i=1}^{\infty}p(1-p)^{i-1}p(1-p)^{i-1}=p^1\sum_{i=1}^{\infty}[(1-p)^2]^{i-1}.$$

Now remember that for a partial geometric sum we have a nice formula $\sum_{k=0}^{n-1}q^k=(1-q^n)/(1-q)$ for $q\in(0,1)$ (to prove it, multiply both sides by $(1-q)$ and then cancel the same terms). We continue,

$$\mathrm{P}(X-Y=0|p)=p^2\sum_{i=0}^{\infty}[(1-p)^2]^i=\frac{p^2}{1-(1-p)^2}=\frac{p^2}{p(1-p)}=\frac{p}{2-p}.$$

Thus, the probablity does depend on p and thus $X-Y$ is not SS.

2. First of all. let us stress that the likelihood function

$$T(\underline{X})=\frac{f_{\theta_1}(\underline{X})}{f_{\theta_0}(\underline{X})}$$

is indeed a statistic because it is a function of observations. The fact that it also depends on θ_0 and θ_1 is irrelevant because these are two known values of an underlying parameter θ which we do not know. In other words, a statistic may depend on Ω and any other given information but not on the underlying parameter θ which we infer about.

Now I present several possible solutions. You can think about more...

A. Set $h(\underline{X}) := [f_{\theta_0}(\underline{X}) f_{\theta_1}(\underline{X})]^{1/2}$ and $g(T(\underline{X}), \theta) := [T(\underline{X})]^{1/2}$ if $\theta = \theta_1$ and $g(T(\underline{X}), \theta) := [T(\underline{X})]^{-1/2}$ if $\theta = \theta_0$. Then we can write

$$f_\theta(\underline{X}) = f_\theta(g(T(\underline{X}), \theta))h(\underline{X}).$$

By the Factorization Theorem T is SS¿

B. Set $h(\underline{X}) := f_{\theta_0}(\underline{X})$, and also $g(T(\underline{X}), \theta) := 1$ if $\theta = \theta_0$ and $g(T(\underline{X}), \theta) := T(\underline{X})$ if $\theta = \theta_1$. Then

$$f_\theta(\underline{X}) = g(T(\underline{X}), \theta)h(\underline{X}).$$

By the Factorization Theorem T is SS.

3. It is given that $\underline{X} := \{X_1, \ldots, X_n\}$ is a sample from $U(\theta - 1/2, \theta + 1/2)$. Let us write the joint pdf so, may be, the Factorization Theorem will point on a SS. Write

$$f(\underline{x}|\theta) \prod_{l=1}^{n} I(\theta - 1/2 \leq x_l \leq \theta + 1/2)$$

$$= I(\theta - 1/2 \leq \min_{l \in \{1,\ldots,n\}} x_l) I(\max_{l \in \{1,\ldots,n\}} x_l \leq \theta + 1/2) = I(x_{(1)} \geq \theta - 1/2) I(x_{(n)} \leq \theta + 1/2)$$

$$=: g((x_{(1)}, x_{(n)}), \theta).$$

Thus, by the Factorization Theorem, the pair $T(\underline{X}) := (X_{(1)}, X_{(n)})$ of the smallest and largest observations is SS.

4. a) Let T be SS and $T := \psi(U)$. By the Factorization Theorem (and remember that it is a criteria with iff), there exist some functions g and h such that

$$f(\underline{X}|\theta) = g(T, \theta)h(\underline{X}) = g(\psi(U), \theta)h(\underline{X}) =: g_1(U, \theta)h(\underline{X}).$$

Here g_1 is a new function such that $g_1(z, \theta) := g(\psi(z), \theta)$. Again using the Factorization Theorem, we establish that U is SS.

Remark: Please note that I used both the sufficient and necessary parts of the Factorization Theorem. Do you remember which part is the necessary one and which is sufficient one?

b) Now, let U be not SS and $T = \psi(U)$. Assume that T is sufficient. Then part a) yields that U must be SS, and the contradictory proves that T is not sufficient.

5. Problem 6.1. Let $X \sim N(0, \sigma^2)$. Is $|X|$ SS? Below I present 3 possible solutions.

a) Yes, and it does not matter that X is normal, the important part is that X is symmetricically distributed about 0 because then

$$f(X|\ |X|, \sigma) = \begin{cases} 1/2 & \text{if } X = |X| \\ 1/2 & \text{if } X = -|X|. \end{cases}$$

Indeed, if $|X| =: a$ is given then X can be either a or $-a$, and this occurs with the same probability if the distribution is symmetric about 0.

b) Write

$$f(x|\,|x|,\sigma)=\frac{f(x|\,\sigma)}{q(|x|\,|\sigma)}=\frac{f(x|\,\sigma)}{f(-|x|\,|\sigma)+f(|x|:|\sigma)}.$$

Using the assumed symmetry we conclude that

$$f(x|\,|x|,\sigma)=1/2I(x\in\{-|x|,|x|\}).$$

c) Write using the specific of the normal distribution,

$$f(x|\sigma)=(2\pi\sigma^2)^{1/2}e^{-x^2/(2\sigma^2)}=(2\pi\sigma^2)^{1/2}e^{-|x|^2/(2\sigma^2)}.$$

As a result, by the Factorization Theorem, X is SS.

6. Problem 6.2. Write using the assumed independence,

$$f_{\underline{X}}(\underline{x}|\theta)=\prod_{i=1}^{n}f_{X_i}(x_i|\theta)$$

$$=\prod_{i=1}^{n}e^{i\theta-x_i}I(x_i\geq i\theta)=e^{\theta\sum_{i=1}^{n}i}e^{-\sum_{i=1}^{n}x_i}\prod_{i=1}^{n}I(x_i/i\geq\theta)$$

$$=e^{\theta\sum_{i=1}^{n}i}e^{-\sum_{i=1}^{n}x_i}\prod_{i=1}^{n}I(\min_i(x_i/i)\geq\theta)=[e^{-\sum_{i=1}^{n}x_i}][e^{\theta\sum_{i=1}^{n}x_i}I(\min_i(x_i/i)\geq\theta)].$$

By the Factorization Theorem the statistic $T:=\min_i(x_i/i)$ is SS.

7. Problem 6.3. By the iid property (note that below the indicator function is an extremely important part of the density — you must not skip it!),

$$f(\underline{x}|\mu,\sigma)=\prod_{l=1}^{n}[\sigma^{-1}e^{-(x+l-\mu)/\sigma}I(x_l>\mu)]=\sigma^{-n}e^{-[\sum_{l=1}^{n}x_l-n\mu]/\sigma}I(x_{(1)}>\mu).$$

We conclude that by the FT:

a) if σ is given then $T_1:=X_{(1)}$ is SS for μ.
b) if μ is given then $T_2:=\bar{X}$ is SS for σ.
c) (T_1,T_2) is the two-dimensional SS for (μ,σ).

8. Problem 6.5. By the independence

$$f(\underline{x}|\theta)=\prod_{l=1}^{n}[2l\theta]^{-1}I(-l(\theta-1)<x_l<l(\theta+1))=(2\theta)^{-n}\prod_{l=1}^{n}l^{-1}I(-(\theta-1)<x_l/l<\theta+1)$$

$$=(2\theta)^{-n}\prod_{l=1}^{n}l^{-1}I(-(\theta-1)<\min_l x_l/l)I(\max_l x_l/l<\theta+1)=:g(\min_l(x_l/l),\max_l x_l/l,\theta).$$

By the FT, the two-dimensional statistic $T:=(\min_l(x_l/l),\max_l x_l/l)$ is SS.

9. Problem 6.6. It is given that $X_1, \dots, X_n$ is a sample from $Gamma(\alpha, \beta)$. Then

$$f(\underline{x}|\alpha,\beta) = \prod_{l=1}^{n} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x_l/\beta} I(x_l \geq 0)$$

$$= [\Gamma(\alpha)\beta^{\alpha}]^{-n} I(\min_l x_l \geq 0) \prod_{l=1}^{n} x_l^{\alpha-1} e^{-\sum_{l=1}^{n} x_l/\beta}$$

$$= \{[\Gamma(\alpha)\beta^{\alpha}]^{-n} I(\min_l x_l \geq 0)\}\{\prod_{l=1}^{n} x_l^{\alpha-1} e^{-\sum_{l=1}^{n} x_l/\beta}\}.$$

The first factor in the curly brackets is a function in observations and the second is a function $g(\prod_{l=1}^{n} x_l, \sum_{l=1}^{n} x_l, \alpha, \beta)$.

Now we can conclude using the FT that:
a) if α is given/known then $T_1 := \sum_{l=1}^{n} X_l$ is SS for β.
b) if β is given/known then $T_2 := \prod_{l=1}^{n} X_l$ is SS for α.
c) $T := (T_1, T_2)$ is SS for the pair (α, β) of parameters.

10. Problem 6.7. Begin with drawing a diagram showing the support of the bivariate density in R^2 (this is just a square with $[\theta_1, \theta_3]$ being the horizontal side in x-axis, and $[\theta_2, \theta_4]$ being the vertical side in y-axis).

Then we can write for a sample from (X, Y),

$$f(\underline{(x,y)}|\theta_1, \theta_2, \theta_3\theta_4) = [(\theta_3 - \theta_1)(\theta_4 - \theta_2)]^n \prod_{l=1}^{n} I(\theta_1 \leq x_l \leq \theta_3) I(\theta_2 \leq y_l \leq \theta_4)$$

$$= [(\theta_3 - \theta_1)(\theta_4 - \theta_2)]^n \prod_{l=1}^{n} I(\theta_1 \leq \min_l x_l) I(\max_l x_l \leq \theta_3) I(\theta_2 \leq \min_l y_l) I(\max_l y_l \leq \theta_4).$$

By the FT, the statistic $T := (x_{(1)}, x_{(n)}, y_{(1)}, y_{(n)})$ is SS for $(\theta_1, \theta_2, \theta_3, \theta_4)$.