STA 3000, Fall 2009, Assignment #1, Answer to Question 1

Question 1: For the model $Y \sim P_\theta$, suppose that $s(Y)$ is a sufficient statistic. Prove that for the model $Y_1, \ldots, Y_n \overset{\text{IID}}{\sim} P_\theta$, the order statistics, $S_{(1)}, \ldots, S_{(n)}$, of $S_1, \ldots, S_n$, where $S_i = s(Y_i)$, are sufficient. Prove this directly from the (classical) definition of a sufficient statistic, without using any theorems about sufficient statistics.

Answer: We need to prove that the conditional distribution of $Y_* = (Y_1, \ldots, Y_n)$ given $S_{(*)} = (S_{(1)}, \ldots, S_{(n)})$ is the same for all θ. I will assume that we can naively express conditional distributions by conditional densities that are ratios of joint densities. This will certainly be fine when the Y_i are discrete. The density function for the original model for a single Y will be written as f_Y, and the density function for $s(Y)$ will be written as as f_S.

To start,

$$
\begin{aligned}
f_{Y_*|S_{(*)}}(y_*|s_{(*)}) &= \frac{f_{Y_*,S_{(*)}}(y_*, s_{(*)})}{f_{S_{(*)}}(s_{(*)})} \\
&= \frac{f_{Y_*}(y_*)\ I(\text{there is a permutation } \pi \text{ such that for all } i,\ s_{(i)} = s(y_{\pi(i)}))}{f_{S_{(*)}}(s_{(*)})}
\end{aligned}
$$

We can now use independence of the Y_i to write $f_{Y_*}(y_*) = \prod_i f_Y(y_i)$. Using independence of the S_i, we can write

$$
f_{S_{(*)}}(s_{(*)}) = \sum_{\substack{s_1,\ldots,s_n \\ \text{permuting} \\ s_{(1)},\ldots,s_{(n)}}} \prod_i f_S(s_i) = N(s_{(*)}) \prod_i f_S(s_{(i)})
$$

where the sum is over distinct permuted sequences of $s_{(1)}, \ldots, s_{(n)}$, and $N(s_{(*)})$ is the number of such distinct permutations (which will be $n!$ if the $s_{(i)}$ are all distinct). With these substitutions, we get

$$
\begin{aligned}
f_{Y|S_{(*)}}(y|s_{(*)}) &= \frac{\prod_i f_Y(y_i)}{N(s_{(*)}) \prod_i f_S(s_{(i)})}\ I(\text{there is a permutation } \pi \text{ such that for all } i,\ s_{(i)} = s(y_{\pi(i)})) \\
&= \frac{\prod_i f_Y(y_{\pi(i)})}{N(s_{(*)}) \prod_i f_S(s_{(i)})}\ I(\text{there is a permutation } \pi \text{ such that for all } i,\ s_{(i)} = s(y_{\pi(i)})) \\
&= \left[\prod_i \frac{f_Y(y_{\pi(i)})}{f_S(s_{(i)})}\right]\ \frac{I(\text{there is a permutation } \pi \text{ such that for all } i,\ s_{(i)} = s(y_{\pi(i)}))}{N(s_{(*)})}
\end{aligned}
$$

where if the indicator function $I(\cdots)$ above is one, the π appearing in $y_{\pi(i)}$ is one of the permutations for which $s_{(i)} = s(y_{\pi(i)})$ for all i (and otherwise is any arbitrary permutation).

When $s_{(i)} = s(y_{\pi(i)})$, we can write $f_Y(y_{\pi(i)}) = f_{Y,S}(y_{\pi(i)}, s_{(i)})$, and we can therefore also write $f_Y(y_{\pi(i)})/f_S(s_{(i)}) = f_{Y|S}(y_{\pi(i)}|s_{(i)})$, which does not depend on θ since S is sufficient for the model of a single observation, Y.

The second factor above — the ratio of $I(\cdots)$ and $N(s_{(*)})$ — does not depend on θ. When this factor is not zero, the first factor — the product over i of $f_Y(y_{\pi(i)})/f_S(s_{(i)}) = f_{Y|S}(y_{\pi(i)}|s_{(i)})$ — also does not depend on θ. It follows that $f_{Y|S_{(*)}}$ does not depend on θ, and therefore $S_{(*)}$ is sufficient.