STA 3000, Fall 2009 — Assignment #1

Due October 29, at start of lecture. Worth 8% of the course grade.

This assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own. In particular, you should not leave any discussion of this assignment with any written notes or other recordings, nor receive any written or other material from anyone else by other means such as email.

For all questions, show both the final answer and how you obtained it.

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.

Question 2: Consider the composite null hypothesis, H_0, according to which $Y_1, \ldots, Y_n$ are IID from a density function parameterized by θ having the following form:

$$f(y;\theta) \quad \propto \quad \frac{1 \;-\; I(\theta - 1 < y < \theta + 1)}{1 + (y-\theta)^2}$$

Ie, the data points are from a Cauchy distribution with a "hole" in the middle of width two.

Devise a pure significance test of H_0 that is sensitive to the alternative that there is no hole in the middle of the distribution, or that the hole is smaller than it is under H_0. It is not necessary for your test to allow p-values to be computed by any simple formula, as long as it would be possible in principle to compute the p-values on a computer.

There is no single correct answer to this question. You should both present your proposed test, and discuss what properties it has (such details about exactly which alternatives it is or is not sensitive to).

Question 3: Consider a model with $X_1, \ldots, X_n \overset{\text{IID}}{\sim}$ Poisson(θ) and $Y_1, \ldots, Y_m \overset{\text{IID}}{\sim}$ Poisson$(1-\theta)$, with the X's and Y's being independent. Here n and m are positive integers and θ is an unknown parameter in the interval $(0, 1)$.

Answer the following questions:

a) Find a simple form for the minimal sufficient statistic of this model.

b) Determine whether or not the minimal sufficient statistic is complete.

c) Suppose we use a $U(0, 1)$ prior for θ. Find the posterior distribution for θ given values for $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$.

Answer the following questions assuming that $n = m$:

d) Find the posterior mean and standard deviation of θ from the posterior distribution you found in part (c).

e) Find an ancillary statistic that is a functon of the minimal sufficient statistic.

f) Find an estimator for θ that is unbiased conditional on the ancillary statistic that you found in part (e), and find its standard deviation conditional on that ancillary statistic.

g) Discuss how the Bayesian posterior mean and standard deviation of (d) and the unbiased estimator of (f) and its standard deviation compare as ways of inferring the value of θ.

Finally,

h) Discuss how Bayesian and frequentist inference should be done when $n \neq m$.