STA 3000, Fall 2008 — Assignment #1

Due October 23, 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: Recall that we formalized statistical modelling using a probability space $(S, \mathcal{A}, \mu)$ in which S is a set of "outcomes", $\mathcal{A}$ is a sigma algebra of subsets of S, and μ is a probability measure defined for members of $\mathcal{A}$. Two mappings $X: S \to \mathcal{X}$ and $\Theta: S \to \Omega$ are also assumed to be defined, which extract the "data" and the "parameters" of the model.

Suppose that S is the sset of pairs $\{(a, b) : a, b \in (-1, +1)\}$, that $\mathcal{A}$ is the usual Borel sigma algebra, and that the probability measure μ is defined as follows, for any $A \in \mathcal{A}$:

$$\mu(A) = \begin{cases} (1/5)\,\lambda(A) & \text{if } (1/2, 0) \notin A \\ (1/5)\,(\lambda(A) + 1) & \text{if } (1/2, 0) \in A \end{cases}$$

where λ is Lesbegue measure. Let $\mathcal{X}$ be the interval $(-2, +2)$, let Ω be the interval $(-1, +1)$, and define the mappings X and Θ as follows, for any pair $(a, b) \in S$:

$$\begin{aligned} X((a, b)) &= a + b \\ \Theta((a, b)) &= a \end{aligned}$$

Let μ_X and μ_Θ be the probability measures over $\mathcal{X}$ and Ω produced by μ and the mappings defined above, and let $\mu_{\Theta|X}$ be the conditional probability measure for Θ given X.

(a) Find as simple an expression as you can for $\mu_\Theta((p, q))$, where (p, q) is an interval with p, $q \in (-1, +1)$ and $p < q$.

(b) Find a measure ω for which $\mu_\Theta \lll \omega$, and find the Radon-Nikodym derivative, f_Θ, of μ_Θ with respect to ω. Verify that your Radon-Nikodym derivative is correct by computing $\mu_\Theta((p, q))$ by integrating it over the interval (p, q).

(c) Find a measure ν for which $\mu_{X|\Theta=\theta} = P_\theta \lll \nu$ for all $\theta \in (-1, +1)$, and find the Radon-Nikodym derivative, $f_{X|\Theta}$, of P_θ with respect to ν. Verify that your answer is correct by showing that for any x, $y \in (-2, +2)$ with $x < y$,

$$\mu_X((x, y)) = \int \left[\int_{(x,y)} f_{X|\Theta}(z|\theta)\, d\nu(z) \right] f_\Theta(\theta)\, d\omega(\theta)$$

(d) Using Bayes' Theorem, find $f_{\Theta|X}$, the Radon-Nikodym derivative of $\mu_{\Theta|X}$ with respect to ω, and use it to compute $\mu_{\Theta|X}(\Theta = 1|X = 1/2)$ and $\mu_{\Theta|X}(\Theta < 0|X = -1/2)$.

Question 2: Consider modeling a sequence of n real-valued observations, $X = (X_1, \ldots, X_n)$, using a parameter $\theta = (\mu, \delta)$, in which given θ the observations are independent and each has the distribution with the following density (w.r.t. Lesbegue measure):

$$f_{X_i|\Theta}(x|(\mu,\delta)) \quad = \quad \begin{cases} (1/2)\,\exp(-(x-(\mu+\delta)) & \text{if } x > \mu+\delta \\ (1/2)\,\exp(-((\mu-\delta)-x) & \text{if } x < \mu-\delta \\ 0 & \text{otherwise} \end{cases}$$

Below, two variations on this model are defined, which differ in the set of values for θ that make up the parameter space.

- Suppose the parameter space is $\Omega = \{\,(\mu,\delta)\,:\,\mu \in (-\infty,\infty),\ \delta \in (0,\infty)\,\}$.
  - (a) Give a description of a simple minimal sufficient statistic for this model, and prove that it is minimal sufficient.
- Suppose the parameter space is $\Omega = \{\,(\mu,\delta)\,:\,\mu = 0,\ \delta \in (0,\infty)\,\}$ — ie, μ is fixed at 0, and δ can be any positive real.
  - (b) Give a description of a simple minimal sufficient statistic for this model, and prove that it is minimal sufficient.
  - (c) Suppose that the prior density for δ is $\exp(-\delta)$. Find a simple expression for the posterior density of δ.

Question 3: Consider a model for data $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ in which these data points are independent given a parameter $\theta \in (0,\infty)$, with $X_i \sim \text{Exp}(\theta)$ for $i = 1, \ldots, n$ and $Y_j \sim \text{Exp}(\theta^2)$ for $j = 1, \ldots, m$. The density for a random variable X with $\text{Exp}(\theta)$ distribution is $I(x > 0)\theta\exp(-\theta x)$.

(a) Find a simple form of the minimal sufficient statistic for this problem, and prove that it is minimal sufficient.

(b) Find an ancillary statistic that is a function of the minimal sufficient statistic.

Question 4: Consider a model in which both the data space, $\mathcal{X}$, and the parameter space, Ω, are finite sets, and in which $P_\theta(\{x\}) > 0$ for all $\theta \in \Theta$ and $x \in \mathcal{X}$. Let $T(X)$ be some statistic. Prove that if the cardinality of the range of $T(X)$ is greater than the cardinality of Ω, then $T(X)$ is not a complete statistic.