STA 3000, Fall 2009 — Assignment #2

Due November 26, 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.

Question 1: Consider the following procedure for randomly generating a sequence $Y_1, Y_2, Y_3, \ldots$, with each $Y_i \in R$, by successively generating from the conditional distributions $Y_i | Y_1, \ldots, Y_{i-1}$:

Draw Y_1 from the $N(0, 1)$ distribution.

For $i = 2, 3, 4, \ldots$

With probability $\dfrac{c + d\,\mathrm{card}(\{Y_1, \ldots, Y_{i-1}\})}{c + i - 1}$,

draw Y_i from the $N(0, 1)$ distribution (independently of previous draws).

Otherwise, for $j \in \{1, \ldots, i-1\}$, with probability $\dfrac{\mathrm{card}(\{k : Y_k = Y_j,\, k \in \{1, \ldots, i-1\}\}) - d}{c + i - 1}$,

let Y_i be equal to Y_j.

Here, $c \in (0, \infty)$ and $d \in [0, 1)$ are fixed constants. The card function gives the number of distinct elements in a set.

a) Show that the distribution for $Y_1, Y_2, \ldots$ produced by this procedure is exchangeable.

b) For some fixed a, and for $i = 1, 2, \ldots$, let $X_i = I(Y_i < a)$. Show that the distribution for $X_1, X_2, \ldots$ is exchangeable.

c) By DeFinetti's representation theorem, there must therefore be a distribution, $D(a)$, over $\theta \in (0, 1)$ such that the distribution of $X_1, X_2, \ldots$ can be expressed as

$$\begin{array}{rcl} \theta & \sim & D(a) \\ X_1, X_2, \ldots | \theta & \stackrel{\text{IID}}{\sim} & \text{Bernoulli}(\theta) \end{array}$$

Find the mean and variance of $D(a)$, and check whether or not it is possible for a beta distribution to have such a mean and variance.

Question 2: Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be pairs in R^2 that are drawn independently from the distribution with the following density function:

$$f_{X_i, Y_i}(x, y) \;=\; c(\theta) \exp(-\theta(|x| + |y|))$$

where $\theta \in (0, \infty)$ is an unknown model parameter, and $c(\theta)$ is the function that makes the density integrate to one. This is an exponential family model.

Find the function $c(\theta)$, find the distribution of the natural sufficient statistic, and find the expected value of the natural sufficient statistic (by differentiating $c(\theta)$).

Question 3: Let $A_1, \ldots, A_k$ be measureable subsets of R^2. Let $\theta_1, \ldots, \theta_k \in R$ be unknown model parameters. Let the density of $Y \in R^2$ be

$$f_Y(y) \;=\; c(\theta)\phi(y_1)\phi(y_2) \exp\Big(\sum_{j=1}^{k} \theta_j I(y \in A_j)\Big)$$

where ϕ is the density function of the standard normal distribution, and $c(\theta)$ is the function that makes the density integrate to one.

Under what conditions is this exponential family model minimal?