

## STA 3000, Fall 2014 — Assignment #4

To be handed in by 2 April 2015 (or earlier). 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:** [ 34 marks ] Let  $X_1, X_2, \dots$  be an infinite sequence of exchangeable random variables with the same range (which for simplicity we'll assume is discrete), and let  $Y_1, Y_2, \dots$  be another such infinite exchangeable sequence, independent of  $X_1, X_2, \dots$ , with each  $Y_i$  having the same range as the  $X_i$ , but not necessarily with the same distribution as  $X_1, X_2, \dots$ . By De Finetti's representation theorem, there exist probability measures  $\mu_X$  and  $\mu_Y$ , over probability measures,  $\nu$ , on the range of the  $X_i$  and  $Y_i$ , such that for any  $n$ ,

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \int \prod_{i=1}^n \nu(x_i) d\mu_X(\nu)$$
$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_n = y_n) = \int \prod_{i=1}^n \nu(y_i) d\mu_Y(\nu)$$

- a) Let  $B$  be a Bernoulli(1/2) random variable, independent of  $X_1, X_2, \dots$  and  $Y_1, Y_2, \dots$ . Define  $Z_1, Z_2, \dots$  to be an infinite sequence of random variables, with

$$Z_i = \begin{cases} X_i & \text{if } B = 0 \\ Y_i & \text{if } B = 1 \end{cases}$$

Prove (in detail) that  $Z_1, Z_2, \dots$  is exchangeable, and describe the probability distribution over distributions that must therefore underly this exchangeable sequence by De Finetti's representation theorem.

- b) Let  $x_1^*$  be some value in the range of  $X_i$  for which  $P(X_1 = x_1^*) > 0$ . Define the sequence of random variables  $\hat{X}_1, \hat{X}_2, \dots$  as being independent of  $X_1, X_2, \dots$  and with distribution such that, for any  $n$ ,

$$P(\hat{X}_1 = \hat{x}_1, \hat{X}_2 = \hat{x}_2, \dots, \hat{X}_n = \hat{x}_n) = P(X_2 = \hat{x}_1, X_3 = \hat{x}_2, \dots, X_{n+1} = \hat{x}_n \mid X_1 = x_1^*)$$

Prove (in detail) that  $\hat{X}_1, \hat{X}_2, \dots$  is exchangeable, and describe the probability distribution over distributions that must therefore underly this exchangeable sequence by De Finetti's representation theorem.

**Question 2:** [ 32 marks ] Suppose that we model positive real data  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_m$  (with  $n > 0$  and  $m > 0$ ) in terms of a parameter  $\mu$ , with all the  $X_i$  and  $Y_j$  being independent given  $\mu$ , with the  $X_i$  having the exponential distribution with mean  $\mu$  and the  $Y_j$  having the exponential distribution with mean  $\mu + 1$ .

- a) Write down the likelihood function for this data and model.
- b) Find a simple form of the minimal sufficient statistic for this model, and prove (in detail) that it is minimal sufficient.
- c) Determine whether or not the minimal sufficient statistic is complete.
- d) Find simple expressions for the observed information and Fisher information. Are they the same (for all data sets) when evaluated at the MLE for  $\mu$ ? Discuss what implications (if any) this has for drawing inferences about  $\mu$  from this data.

**Question 3:** [ 34 marks ] In order to study the population of rabbits in an certain area, a researcher has hired a trapper to set out 6 rabbit traps, and check them each morning, counting how many rabbits have been caught, and resetting the traps for the next day. The data collected this way is a sequence of counts,  $r_1, r_2, \dots, r_n$ , each of which is an integer between 0 and 6 giving the number of rabbits found in the traps on successive days.

The researcher plans to model these counts as being independent from one day to another, with distributions over the integers 0 to 6 on each day that may vary from one day to another, but only slowly. (That is, the distribution for  $r_i$  and  $r_{i+1}$  will be similar, though the distributions for  $r_i$  and  $r_j$  when  $|i - j|$  is large may be quite different.) Note that these distributions are *not* assumed to be binomial, because the 6 traps may not be equally effective at trapping rabbits.

However, the researcher suspects that the trapper may not have checked the traps every day, but instead, may have recorded a count of zero on some days without looking at or resetting the traps. The number of rabbits counted on the next day would then be for two consecutive days, rather than one. If the trapper does behave this way, the researcher does not know what determines which days the trapper doesn't check the traps, nor does the researcher know how the distribution of number of rabbits caught in two days would relate to the distribution of the number caught in one day, except that of course one would expect to catch more rabbits in two days than one (a rabbit never escapes a trap).

The researcher would like to see if there is good evidence that the trapper hasn't been checking the traps every day, by computing the  $p$ -value for a pure significance test of the null hypothesis that the trapper checks the traps every day, so that the counts are independent from day to day, with a distribution that varies only slowly.

Design such a pure significance test for this problem, that is sensitive to the trapper not checking the traps every day. Your test should ideally produce a distribution for the  $p$ -value that is uniform over  $(0, 1)$ , if the null hypothesis is true, but a test that produces an approximately uniform distribution is OK. It must be possible to compute the  $p$ -value for your test from the data, but you don't need to find a fast way to compute it.

Hint: Since, under the null hypothesis, the distribution for  $r_i$  changes only slowly, it stays approximately constant over a short subsequence for  $r_1, \dots, r_n$  (eg, the distributions for  $r_1, r_2$ , and  $r_3$  are almost the same).