Stat 110 Homework 9, Fall 2013

Due: Friday 11/15 at 1:10 pm, in the Stat 110 dropbox outside of SC 105. No late homework will be accepted. Please write your name, and staple your homework. Show your work and make sure to give clear, careful, convincing explanations.

Reading: Sections 8.3-8.7

1. We wish to measure the lengths of two sticks as accurately and precisely as we can (there will always be uncertainty since we can't measure the lengths perfectly, to infinite precision). The first stick has length a and the second stick has length b, with a and b unknown. Using our lengthometer (also known as a "ruler"), we can measure any length c, obtaining a random variable with mean c and variance σ^2 not depending on c.

The most obvious method of measuring the lengths of the two sticks is to measure the length of the first stick and then measure the length of the second stick, and then report those two measurements. Call this Method 1. An alternative, Method 2, is to put the sticks together as one big stick and measure its length, and then align the sticks side-by-side and measure the difference in lengths, and then convert these measurements in a natural way back to estimates of a and b.

Write out the details of Method 2, and find the means and variances of the estimates obtained by both methods. Which method do you prefer, and why?

2. Show *without using calculus* that for every integer $j \geq 1$ and real number $t > 0$,

$$e^{-\lambda t} \sum_{k=j}^{\infty} \frac{(\lambda t)^k}{k!} = \frac{1}{(j-1)!} \int_0^t (\lambda x)^j e^{-\lambda x} \frac{dx}{x}.$$

3. (a) Let $p \sim \text{Beta}(a, b)$, where a and b are positive real numbers. Find $E(p^2(1-p)^2)$, fully simplified (Γ should not appear in your final answer).

Two teams, A and B, have an upcoming match. They will play five games and the winner will be declared to be the team that wins the majority of games. Given p, the outcomes of games are independent, with probability p of team A winning and $1-p$ of team B winning. But you don't know p, so you decide to model it as an r.v., with $p \sim \text{Unif}(0,1)$ a priori (before you have observed any data).

To learn more about p, you look through the historical records of previous games between these two teams, and find that the previous outcomes were, in chronological order, $AAABBAABAB$. (Assume that the true value of p has not been changing over time and will be the same for the match, though your *beliefs* about p may change over time.)

(b) Does your posterior distribution for p, given the historical record of games between A and B, depend on the specific order of outcomes or only on the fact that A won exactly 6 of the 10 games on record? Explain.

(c) Find the posterior distribution for p, given the historical data.

The posterior distribution for p from (c) becomes your new prior distribution, and the match is about to begin!

(d) Conditional on p, is the indicator of A winning the first game of the match positively correlated with, uncorrelated with, or negatively correlated of the indicator of A winning the second game of the match? What about if we only condition on the historical data?

(e) Given the historical data, what is the expected value for the probability that the match is not yet decided when going into the fifth game (viewing this probability as an r.v. rather than a number, to reflect our uncertainty about it)?

4. (a) Let $X_1, X_2, \dots$ be i.i.d. r.v.s with CDF F, and let $M_n = \max(X_1, X_2, \dots, X_n)$. Find the joint distribution of M_n and M_{n+1}, for each $n \geq 1$.

(b) Let $X_1, X_2, \dots, X_n$ be i.i.d. continuous r.v.s with CDF F and PDF f. Find the joint PDF of the order statistics $X_{(i)}$ and $X_{(j)}$, for $1 \leq i < j \leq n$.

5. We are about to observe random variables $Y_1, Y_2, \dots, Y_n$, i.i.d. from a continuous distribution. We will need to predict an independent future observation Y_{new}, which will also have the same distribution. The distribution is unknown, so we will construct our prediction using $Y_1, Y_2, \dots, Y_n$ rather than the distribution of Y_{new}. In forming a prediction, we do not want to report only a single number; rather, we want to give a *predictive interval* with "high confidence" of containing Y_{new}. One approach to this is via order statistics.

(a) For fixed j and k with $1 \leq j < k \leq n$, find $P(Y_{\text{new}} \in [Y_{(j)}, Y_{(k)}])$.

Hint: by symmetry, all orderings of $Y_1, \dots, Y_n, Y_{\text{new}}$ are equally likely.

(b) Let $n = 99$. Construct a predictive interval, as a function of $Y_1, \dots, Y_n$, such that the probability of the interval containing Y_{new} is 0.95.

6. Let $X_1, \dots, X_n$ be i.i.d. continuous r.v.s with n odd. Show that the median of the distribution of the sample median of the X_i's is the median of the distribution of the X_i's.

Hint: two approaches to evaluating a sum that might come up are (i) use the first story proof example and the first story proof exercise from Chapter 1, or (ii) use the fact that, by the story of the Binomial, $Y \sim \text{Bin}(n, 1/2)$ implies $n - Y \sim \text{Bin}(n, 1/2)$.