Stat 110 Homework 4, Fall 2013

Due: Friday 10/4 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 3.4-3.9, 4.1-4.6.

1. A building has n floors, labeled $1, 2, \dots, n$. At the first floor, k people enter the elevator, which is going up and is empty before they enter. Independently, each decides which of floors $2, 3, \dots, n$ to go to and presses that button (unless someone has already pressed it).

(a) Assume for this part only that the probabilities for floors $2, 3, \dots, n$ are equal. Find the expected number of stops the elevator makes on floors $2, 3, \dots, n$.

(b) Generalize (a) to the case that floors $2, 3, \dots, n$ have probabilities $p_2, \dots, p_n$ (respectively); you can leave your answer as a finite sum.

2. The U.S. Senate consists of 100 senators, with 2 from each of the 50 states. There are d Democrats in the Senate. A committee of size c is formed, by picking a random set of senators such that all sets of size c are equally likely.

(a) Find the expected number of Democrats on the committee.

(b) Find the expected number of states represented on the committee (by at least one senator).

(c) Find the expected number of states such that both of the state's senators are on the committee.

3. The *Wilcoxon rank sum test* is a widely-used procedure for assessing whether two groups of observations come from the same distribution. Let group 1 consist of i.i.d. $X_1, \dots, X_m$ with CDF F and group 2 consist of i.i.d. $Y_1, \dots, Y_n$ with CDF G, with all of these r.v.s independent. Assume that the probability of 2 of the observations being equal is 0 (this will be true if the distributions are continuous).

After the $m+n$ observations are obtained, they are listed in increasing order, and each is assigned a *rank* between 1 and $m+n$: the smallest has rank 1, the second smallest has rank 2, etc. Let R_j be the rank of X_j among all the observations for $1 \leq j \leq m$, and let $R = \sum_{j=1}^{m} R_j$ be the sum of the ranks for group 1.

Intuitively, the Wilcoxon rank sum test is based on the idea that a very large value of R is evidence that observations from group 1 are usually larger than observations from group 2 (and vice versa if R is very small). But how large is "very large" and how small is "very small"? Answering this precisely requires studying the distribution of the *test statistic* R.

(a) The *null hypothesis* in this setting is that $F = G$. Show that if the null hypothesis is true, then $E(R) = m(m+n+1)/2$.

(b) The *power* of a test is an important measure of how good the test is about saying to reject the null hypothesis if the null hypothesis is false (see `http://www.youtube.com/watch?v=kMYxd6QeAss` for more about the power of a test). To study the power of the Wilcoxon rank sum stest, we need to study the distribution of R in general. So for this part, we do *not* assume $F = G$. Let $p = P(X_1 > Y_1)$. Find $E(R)$ in terms of m, n, p.

Hint: write R_j in terms of indicator r.v.s for X_j being greater than various other r.v.s.

4. (Exercise 4.10 in book) Nick and Penny are independently performing independent Bernoulli trials. For concreteness, assume that Nick is flipping a nickel with probability p_1 of Heads and Penny is flipping a penny with probability p_2 of Heads. Let $X_1, X_2, \dots$ be Nick's results and $Y_1, Y_2, \dots$ be Penny's results, with $X_i \sim \text{Bern}(p_1)$ and $Y_j \sim \text{Bern}(p_2)$.

(a) Find the distribution and expected value of the first time at which they are simultaneously successful, i.e., the smallest n such that $X_n = Y_n = 1$.

Hint: define a new sequence of Bernoulli trials and use the story of the Geometric.

(b) Find the expected time until at least one has a success (including the success).

Hint: define a new sequence of Bernoulli trials and use the story of the Geometric.

(c) For $p_1 = p_2$, find the probability that their first successes are simultaneous, and use this to find the probability that Nick's first success precedes Penny's.

5. (Exercise 4.18 in book) You have a well-shuffled 52-card deck. You turn the cards face up one by one, without replacement. What is the expected number of non-aces that appear before the first ace? Between the first ace and the second ace?

Hint: first consider for each non-ace the indicator of it preceding all the aces in the deck.

6. (Exercise 4.20 in book) You are being tested for psychic powers. Suppose that you do not have psychic powers. A standard deck of cards is shuffled, and the cards are dealt face down one by one. Just after each card is dealt, you name any card (as your prediction). Let X be the number of cards you predict correctly.

(a) Suppose that you get no feedback about your predictions. Show that no matter what strategy you follow, the expected value of X stays the same; find this value. (On the other hand, the *variance* may be very different for different strategies. For example, saying "Ace of Spades" every time gives variance 0.)

Hint: indicator r.v.s.

(b) Now suppose that you get partial feedback: after each prediction, you are told immediately whether or not it is right (but without the card being revealed). Suppose you use the following strategy: keep saying a specific card's name (e.g., "Ace of Spades") until you hear that you are correct. Then keep saying a different card's name (e.g., "Two of Spades") until you hear that you are correct (if ever). Continue in this way, naming the same card over and over again until you are correct and then switching to a new card, until the deck runs out. Find the expected value of X, and show that it is very close to $e - 1$.

Hint: indicator r.v.s.

(c) Now suppose that you get complete feedback: just after each prediction, the card is revealed. Call a strategy "stupid" if it allows, e.g., saying "Ace of Spades" as a guess after the Ace of Spades has already been revealed. Show that any non-stupid strategy gives the same expected value for X; find this value.

Hint: indicator r.v.s.

7. The dean of Blotchville University boasts that the average class size there is 20. But the reality experienced by the majority of students there is quite different: they find themselves in huge courses, held in huge lecture halls, with hardly enough seats or Haribo gummi bears for everyone. The purpose of this problem is to shed light on the situation. For simplicity, suppose that every student at Blotchville University takes only one course per semester.

(a) Suppose that there are 16 seminar courses, which have 10 students each, and 2 large lecture courses, which have 100 students each. Find the dean's-eye-view average class size (the simple average of the class sizes) and the student's-eye-view average class size (the average class size experienced by students, as it would be reflected by surveying students and asking them how big their classes are). Explain the discrepancy intuitively.

(b) Give a short proof that for *any* set of class sizes (not just those given above), the dean's-eye-view average class size will be strictly less than the student's-eye-view average class size, unless all classes have exactly the same size.

Hint: relate this to the fact that variances are nonnegative.