Stat S-110 Homework 3, Summer 2015

Due: Wednesday 7/15 at noon (Eastern Time), submitted as a PDF file via the course website or as a hard copy in class. Show your work and explain your steps.

Collaboration Policy: You are welcome to discuss the problems with others, but *you must write up your solutions yourself and in your own words.* See the syllabus for further information about academic integrity.

The following problems are from Chapter 4 of the book.

1. (BH 4.7) A certain small town, whose population consists of 100 families, has 30 families with 1 child, 50 families with 2 children, and 20 families with 3 children. The *birth rank* of one of these children is 1 if the child is the firstborn, 2 if the child is the secondborn, and 3 if the child is the thirdborn.

(a) A random family is chosen (with equal probabilities), and then a random child within that family is chosen (with equal probabilities). Find the PMF, mean, and variance of the child's birth rank.

(b) A random child is chosen in the town (with equal probabilities). Find the PMF, mean, and variance of the child's birth rank.

2. (BH 4.8) A certain country has four regions: North, East, South, and West. The populations of these regions are 3 million, 4 million, 5 million, and 8 million, respectively. There are 4 cities in the North, 3 in the East, 2 in the South, and there is only 1 city in the West. Each person in the country lives in exactly one of these cities.

(a) What is the average size of a city in the country? (This is the arithmetic mean of the populations of the cities, and is also the expected value of the population of a city chosen uniformly at random.)

Hint: Give the cities *names* (labels).

(b) Show that without further information it is impossible to find the variance of the population of a city chosen uniformly at random. That is, the variance depends on how the people within each region are allocated between the cities in that region.

(c) A region of the country is chosen uniformly at random, and then a city within that region is chosen uniformly at random. What is the expected population size of this randomly chosen city?

Hint: First find the selection probability for each city.

(d) Explain intuitively why the answer to (c) is larger than the answer to (a).

3. (BH 4.19) Let $X \sim \text{Bin}(100, 0.9)$. For each of the following parts, construct an example showing that it is possible, or explain clearly why it is impossible. In this problem, Y is a random variable on the same probability space as X; note that X and Y are not necessarily independent.

(a) Is it possible to have $Y \sim \text{Pois}(0.01)$ with $P(X \geq Y) = 1$?

(b) Is it possible to have $Y \sim \text{Bin}(100, 0.5)$ with $P(X \geq Y) = 1$?

(c) Is it possible to have $Y \sim \text{Bin}(100, 0.5)$ with $P(X \leq Y) = 1$?

4. (BH 4.25) 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. (BH 4.34) Each of $n \geq 2$ people puts his or her name on a slip of paper (no two have the same name). The slips of paper are shuffled in a hat, and then each person draws one (uniformly at random at each stage, without replacement). Find the average number of people who draw their own names.

6. (BH 4.36) In a sequence of n independent fair coin tosses, what is the expected number of occurrences of the pattern HTH (consecutively)? Note that overlap is allowed, e.g., $HTHTH$ contains two overlapping occurrences of the pattern.

7. (BH 4.43) 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. (See Diaconis (1978) for much more about the statistics of testing for psychic powers.)

(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.

8. (BH 4.66) Use Poisson approximations to investigate the following types of coincidences. The usual assumptions of the birthday problem apply, such as that there are 365 days in a year, with all days equally likely.

(a) How many people are needed to have a 50% chance that at least one of them has the same birthday as *you*?

(b) How many people are needed to have a 50% chance that there are two people who not only were born on the same day, but also were born at the same *hour* (e.g., two people born between 2 pm and 3 pm are considered to have been born at the same hour).

(c) Considering that only 1/24 of pairs of people born on the same day were born at the same hour, why isn't the answer to (b) approximately $24 \cdot 23$? Explain this intuitively, and give a simple approximation for the factor by which the number of people needed to obtain probability p of a birthday match needs to be scaled up to obtain probability p of a birthday-birthhour match.

(d) With 100 people, there is a 64% chance that there are 3 with the same birthday (according to R, using `pbirthday(100,classes=365,coincident=3)` to compute it). Provide two different Poisson approximations for this value, one based on creating an indicator r.v. for each triplet of people, and the other based on creating an indicator r.v. for each day of the year. Which is more accurate?