Statistics 134 - Instructor: Hank Ibser

MIDTERM

FRIDAY, OCTOBER 6, 2017

PRINT YOUR NAME ________________________

SIGN YOUR NAME ________________________

CIRCLE YOUR SECTION TIME:

9-10 1-2 2-3 3-4 4-5 5-6 6-7

CIRCLE YOUR GSI'S NAME:

Jason Zhang Andy Palaniappan Maxwell Weinstein

Kazu Kogachi Brian Thorsten Biyonka Liang Dibya Ghosh

Jessica Gao Dhruvil Badani Dhrushil Badani

TURN OFF YOUR CELL PHONE!
You may not use a calculator or any notes or books.
For full credit, give reasons and/or show work.
Each part of every problem is worth 10 points, for a total of 80.
Normal approximation answers should be left in terms of Φ, the normal cdf.
Answers need not be simplified, but any infinite sums should be.
You may use the backs of sheets as scratch, but write anything you want graded on the front.
Distribution summaries are on the last page, you need not hand this page in.
The exam will be collected at 1pm. When I call time, close your exam and stand up.
If you continue to work after I call time, I will give you a 0 on that problem.
GOOD LUCK!

Scores:

1:__________

2:__________

3:__________

Total:__________

1. You have a bag that contains 6 yellow skittles, 4 red skittles, and 5 purple skittles. You eat them randomly one-by-one. You eat a purple one and don't like it very much, but you keep eating. Once you eat a second purple one, you stop.

   (a) What is the chance that you eat exactly 7 skittles? Let X = number of skittles until seconds purple skittle.

   If you ate 7 skittles, you must have eaten one purple skittle in the first six draws, and then another purple on the seventh. So,

$$P(X=7) = \frac{\binom{5}{1}\binom{10}{5}}{\binom{15}{6}} \cdot \frac{4}{9}$$

   (b) What is the expected number of skittles that you eat? We will use the method of indicators here. Label each of the non-purple skittles from $\{1, 2, \ldots, 10\}$. Let $\mathbb{I}_j$ be the indicator that the non-purple skittle $\#j$ appears before the second skittle. We hence have that

$$X = \mathbb{I}_1 + \ldots + \mathbb{I}_{10} + 2$$

   Now, we find $\mathbb{E}\mathbb{I}_j$. There are six ways to place the non-purple skittle among the five purple skittles and of those six ways, only two of them are before the second purple. Hence, $\mathbb{E}\mathbb{I}_j = \frac{2}{6}$.

$$\begin{aligned}
\mathbb{E}X &= \mathbb{E}[\mathbb{I}_1 + \ldots + \mathbb{I}_{10} + 2] \\
&= \mathbb{E}\mathbb{I}_1 + \ldots + \mathbb{E}\mathbb{I}_{10} + 2 \\
&= 10 \cdot \frac{2}{6} + 2
\end{aligned}$$

   (c) What is the variance of the number of skittles that you eat? Note that $Var(X) = Var(\mathbb{I}_1 + \ldots + \mathbb{I}_{10} + 2) = Var(\mathbb{I}_1 + \ldots + \mathbb{I}_{10})$. Unfortunately, we cannot split this terms into a sum of variances since the indicators are not independent.

   Now, let $Y = \mathbb{I}_1 + \ldots + \mathbb{I}_{10} = X - 2$. Note that $Var(X) = Var(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2$. We already know $\mathbb{E}Y = \mathbb{E}X - 2$, so it simply suffices to find $\mathbb{E}Y^2$.

   Note that

$$\begin{aligned}
\mathbb{E}Y^2 &= \mathbb{E}[(\mathbb{I}_1 + \ldots + \mathbb{I}_{10})^2] \\
&= \mathbb{E}[\sum_j \mathbb{I}_j^2 + \sum_{j \neq k} \mathbb{I}_j \mathbb{I}_k] \\
&= \sum_j \mathbb{E}\mathbb{I}_j^2 + \sum_{j \neq k} \mathbb{E}[\mathbb{I}_j \mathbb{I}_k] \\
&= 10\mathbb{E}I_j^2 + (10)(9)\mathbb{E}[\mathbb{I}_j \mathbb{I}_k]
\end{aligned}$$

Now, we know that $\mathbb{E}I_j^2 = \mathbb{E}I_j$, which we have from part b. All that remains is to compute $\mathbb{E}[\mathbb{I}_j\mathbb{I}_k]$. There are $(7)(6) = 42$ ways to place the two non-purple skittles among the five purple skittles and of those 42 ways, only $(3)(2) = 6$ of them are such that both the non-purple skittles are before the second purple skittle. Hence, $\mathbb{E}[\mathbb{I}_j\mathbb{I}_k] = \frac{3\cdot 2}{7\cdot 6}$. The answer follows by plugging all the values in.

2. In the game of "BEARS" two basketball players take turns attempting free throws. In one round, each player attempts a free throw. If one player makes the free throw and the other doesn't, the player that makes it scores a point. If both make it or neither makes it, no one scores a point. First player to 5 points wins. Suppose Stephen has chance p_1 of making a free throw and Kevin has chance p_2, and all free throws are independent.

   (a) What is the chance that Stephen gets to 5 points first? (Answer should be in terms of p_1 and p_2. The answer may be given as a sum.) If Stephen gets to 5 points first, then Kevin must have between 0 and 4 points; hence, N must be in between 5 and 9. Further note that the game terminates as soon as someone gets to 5 points; so Stephen must have won the last game.

   The probability that Stephen gets a point in any round, by the Craps principle, is $\frac{p_1q_2}{p_1q_2+p_2q_1}$. Similarly, the probability that Kevin gets a point is $\frac{p_2q_1}{p_1q_2+p_2q_1}$. Hence,

$$
\begin{aligned}
&P(\text{Stephen gets 5 before Kevin}) \\
&= \sum_{k=0}^{4} P(\text{Stephen gets 4 points in the first (5 + k - 1) games, then wins the last game}) \\
&= \sum_{k=0}^{4} \binom{5+k-1}{4} \left(\frac{p_1q_2}{p_1q_2+p_2q_1}\right)^4 \left(\frac{p_2q_1}{p_1q_2+p_2q_1}\right)^k \cdot \frac{p_1q_2}{p_1q_2+p_2q_1} \\
&= \sum_{k=0}^{4} \binom{5+k-1}{4} \left(\frac{p_1q_2}{p_1q_2+p_2q_1}\right)^5 \left(\frac{p_2q_1}{p_1q_2+p_2q_1}\right)^k
\end{aligned}
$$

   (b) Describe the distribution of the number of total points scored. (Name it and give parameter(s), give distribution table, or write a formula.)

   Let X denote the number of total points scored. The minimum number of points is 5 (e.g., Stephen wins the first five games) and the maximum number of points is 9 (e.g., Kevin wins the first four games and Stephen wins the next five). It is not hard to see that X can take any value in this range; hence, $X \in \{5, 6, 7, 8, 9\}$.

   Now, if $X = k$, then we must have that either Stephen wins on the k^{th} game or Kevin does. Hence,

$$
\begin{aligned}
P(X = k) &= P(\text{Stephen wins in } k \text{ games}) + P(\text{Kevin wins in } k \text{ games}) \\
&= \binom{k-1}{4} \left(\frac{p_1q_2}{p_1q_2+p_2q_1}\right)^5 \left(\frac{p_2q_1}{p_1q_2+p_2q_1}\right)^{k-5} \\
&\quad + \binom{k-1}{4} \left(\frac{p_2q_1}{p_1q_2+p_2q_1}\right)^5 \left(\frac{p_1q_2}{p_1q_2+p_2q_1}\right)^{k-5}
\end{aligned}
$$

   (c) Suppose Stephen and Kevin each shoot free throws until they each miss one. Let X be the number of free throws that Stephen shoots, and Y is the number that Kevin shoots. In terms of p_1 and p_2, what is the chance that $X > Y$?

Since Stephen throws *until he misses*, we have that $X \sim Geom(q_1)$. Similarly, $Y \sim Geom(q_2)$.

$$\begin{aligned} P(X > Y) &= \sum_{i=1}^{\infty} P(X > k) \cdot P(Y = k) \\ &= \sum_{i=1}^{\infty} p_1^k \cdot p_2^{k-1} q_2 \\ &= p_1 q_2 \sum_{i=1}^{\infty} (p_1 p_2)^{k-1} \\ &= \frac{p_1 q_2}{1 - p_1 p_2} \end{aligned}$$

3. Suppose that in a group of dogs, the average age is 2 years.

   (a) i. What is the greatest possible proportion of the dogs that is older than 7 years? Let X denote the age of a sampled dog. By Markov's inequality,

   $$P(X > 7) = P(X \geq 8) \leq \frac{\mathbb{E}X}{8} = \frac{2}{8} = \frac{1}{4}$$

   (We also accepted answers that interpreted the question as $P(X \geq 7)$.

   ii. Suppose that the SD of the ages of the dogs is 2 years. Now what is the greatest possible proportion of the dogs that is older than 7 years?

   $$P(X > 7) = P(X \geq 8) = P(X-2 \geq 6) \leq P(|X-2| \geq 6) = P(|X-\mathbb{E}X| \geq 3 \cdot SD(X)) \leq \frac{1}{3^2}$$

   (b) Suppose we take the average of the ages of 64 random dogs. Approximately what is the chance that the average is more than 2.5 years? Let X_i denote the age of dog i, and let $\bar{X} = \frac{1}{64}\sum_{i=1}^{6} 4X_i$. We have that

   $$\mathbb{E}\bar{X} = \mathbb{E}\left[\frac{1}{64}\sum_{i=1}^{64} X_i\right] = \frac{1}{64}\sum_{i=1}^{64}\mathbb{E}X_i = \frac{1}{64}\cdot 64 \cdot \mathbb{E}X_i = 2 = \mu$$

   $$Var(\bar{X}) = Var\left(\frac{1}{64}\sum_{i=1}^{64} X_i\right) = \frac{1}{64^2}\sum_{i=1}^{64} Var(X_i) = \frac{1}{64^2}\cdot 64 \cdot Var(X_i) = \frac{4}{64} = \frac{1}{16} = \sigma^2$$

   Hence, $SD(X) = \sigma = \frac{1}{4}$. By the Central Limit Theorem, $\bar{X}$ is approximately distributed as $N(\mu, \sigma^2)$. Hence,

   $$P(\bar{X} > 2.5) = 1 - P(\bar{X} \leq 2.5) \approx 1 - \Phi\left(\frac{2.5-\mu}{\sigma}\right) = 1 - \Phi(2)$$