STAT 302: Assignment 2

Ziyang Jin
34893140

February 2018

Question 1: (a) Let Y denote the sum of the values of flipping the coin B times.
Let X_1 denote the value by flipping the coin 1 time, i.e., drawing ball 1. $X_1 \sim Bin(1, 0.5)$.
Let X_2 denote the value by flipping the coin 2 times, i.e., drawing ball 2. $X_2 \sim Bin(2, 0.5)$.
Let X_3 denote the value by flipping the coin 3 times, i.e., drawing ball 3. $X_3 \sim Bin(3, 0.5)$.
The probability of drawing ball 1, 2, 3 is equal to 1/3.

$$\begin{aligned}
Pr(Y=0) &= \frac{1}{3} \times (Pr(X_1 = 0) + Pr(X_2 = 0) + Pr(X_3 = 0)) \\
&= \frac{1}{3} \times (\binom{1}{0}(\frac{1}{2})^0(\frac{1}{2})^1 + \binom{2}{0}(\frac{1}{2})^0(\frac{1}{2})^2 + \binom{3}{0}(\frac{1}{2})^0(\frac{1}{2})^3) \\
&= \frac{7}{24} \\
\\
Pr(Y=1) &= \frac{1}{3} \times (Pr(X_1 = 1) + Pr(X_2 = 1) + Pr(X_3 = 1)) \\
&= \frac{1}{3} \times (\binom{1}{1}(\frac{1}{2})^1(\frac{1}{2})^0 + \binom{2}{1}(\frac{1}{2})^1(\frac{1}{2})^1 + \binom{3}{1}(\frac{1}{2})^1(\frac{1}{2})^2) \\
&= \frac{11}{24} \\
\\
Pr(Y=2) &= \frac{1}{3} \times (Pr(X_2 = 2) + Pr(X_3 = 2)) \\
&= \frac{1}{3} \times (\binom{2}{2}(\frac{1}{2})^2(\frac{1}{2})^0 + \binom{3}{2}(\frac{1}{2})^2(\frac{1}{2})^1) \\
&= \frac{5}{24} \\
\\
Pr(Y=3) &= \frac{1}{3} \times Pr(X_3 = 3) \\
&= \frac{1}{3} \times (\binom{3}{3}(\frac{1}{2})^3(\frac{1}{2})^0) \\
&= \frac{1}{24}
\end{aligned} \tag{1}$$

Therefore, the probability mass function is:

$$p(y) = \begin{cases} 7/24, & y = 0 \\ 11/24, & y = 1 \\ 5/24, & y = 2 \\ 1/24, & y = 3 \end{cases}$$

(b) Expectation:

$$\mathbb{E}(Y) = 0 \times \frac{7}{24} + 1 \times \frac{11}{24} + 2 \times \frac{5}{24} + 3 \times \frac{1}{24} = 1$$

Variance:

$$Var(Y) = (0-1)^2 \times \frac{7}{24} + (1-1)^2 \times \frac{11}{24} + (2-1)^2 \times \frac{5}{24} + (3-1)^2 \times \frac{1}{24} = \frac{2}{3}$$

Standard Deviation:

$$SD(Y) = \sqrt{Var(Y)} = \sqrt{\frac{2}{3}}$$

(c) The 3 experiments are independent.
Let Y_1 denote the result we get from the first experiment. From (b) we know that:

$$\mathbb{E}(Y_1) = 1,\ SD(Y_1) = \sqrt{\frac{2}{3}}$$

Let Y_2 denote the result we get from the second experiment.$Y_2 = 2Y_1$, so

$$\mathbb{E}(Y_2) = 2,\ SD(Y_2) = 2\sqrt{\frac{2}{3}}$$

Let Y_3 denote the result we get from the third experiment.$Y_3 = 3Y_1$, so

$$\mathbb{E}(Y_3) = 3,\ SD(Y_3) = 3\sqrt{\frac{2}{3}}$$

So the expectation and standard deviation for the total sum are:

$$\mathbb{E}(Y_1 + Y_2 + Y_3) = 1 + 2 + 3 = 6$$

$$SD(Y_1 + Y_2 + Y_3) = \sqrt{\frac{2}{3}} + 2\sqrt{\frac{2}{3}} + 3\sqrt{\frac{2}{3}} = 6\sqrt{\frac{2}{3}}$$

Question 2: (a) i. The probability of picking the first type of coin is $7/10$. Let X_1 be the number of trials Peter needs to have 3 heads when he picks the first type of coin, then $X_1 \sim NegBin(3, 0.5)$.

The probability of picking the first type of coin is $3/10$. Let X_2 be the number of trials Peter needs to have 3 heads when he picks the first type of coin, then $X_2 \sim NegBin(3, 0.7)$.

Let X denote the number of trials Peter needs to get 3 heads after he randomly picks a coin. Then the probability mass function is:

$$p(x) = 0.7 \cdot p(x_1) + 0.3 \cdot p(x_2) = \frac{7}{10}\binom{x-1}{2}0.5^3 0.5^{x-3} + \frac{3}{10}\binom{x-1}{2}0.7^3 0.3^{x-3}$$

Therefore,

$$p(x) = 0.7\binom{x-1}{2}(0.5^x + 0.7^2 \cdot 0.3^{x-2}),\ x = 3, 4, 5, ...$$

ii. The expectation is:

$$\mathbb{E}(X) = \sum_x x \cdot p(x) = 0.7\ \mathbb{E}(X_1) + 0.3\ \mathbb{E}(X_2) = 0.7 \times \frac{3}{0.5} + 0.3 \times \frac{3}{0.7} = \frac{192}{35} = 5.486$$

The variance is:

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 0.7\ \mathbb{E}(X_1^2) + 0.3\ \mathbb{E}(X_2^2) - [\mathbb{E}(X)]^2$$

(1) Let's compute $\mathbb{E}(X_1^2)$

$$\mathbb{E}(X_1^2) = \sum_x x^2 \binom{x-1}{2} 0.5^x = 3 \sum_x x \binom{x}{3} 0.5^x$$

let $y = x + 1$, we have

$$\mathbb{E}(X_1^2) = 3 \sum_y (y-1) \binom{y-1}{3} 0.5^{y-1} = 6[\sum_y y \binom{y-1}{3} 0.5^y - \sum \binom{y-1}{3} 0.5^y]$$

suppose we have $Y_1 \sim NegBin(4, 0.5)$

$$\mathbb{E}(X_1^2) = 6[\mathbb{E}(Y_1) - 1] = 6(\frac{4}{0.5} - 1) = 42$$

(2) Let's compute $\mathbb{E}(X_2^2)$

$$\mathbb{E}(X_2^2) = \sum_x x^2 \binom{x-1}{2} 0.7^3\ 0.3^{x-3} = \frac{30}{7} \sum_x x \binom{x}{3} 0.7^4\ 0.3^{x-3}$$

let $y = x + 1$, we have

$$\mathbb{E}(X_2^2) = \frac{30}{7} \sum_y (y-1) \binom{y-1}{3} 0.7^4\ 0.3^{y-4} = \frac{30}{7} (\sum_y y \binom{y-1}{3} 0.7^4\ 0.3^{y-4} - \sum_y \binom{y-1}{3} 0.7^4\ 0.3^{y-4})$$

suppose we have $Y_2 \sim NegBin(4, 0.7)$

$$\mathbb{E}(X_2^2) = \frac{30}{7}[\mathbb{E}(Y_2) - 1] = \frac{30}{7}(\frac{4}{0.7} - 1) = \frac{990}{49}$$

(3) Finally, we compute the variance:

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 0.7 \times 42 + 0.3 \times \frac{990}{49} - (\frac{192}{35})^2 = \frac{6576}{1225} = 5.368$$

(b) Let Y denote the number of tosses needed to obtain 3 heads. $Y \sim NegBin(3, p)$.
Let X_1 denote the event that Peter picks the first type of coin. $Pr(X_1) = 0.7,\ p = 0.5$.
Let X_2 denote the event that Peter picks the second type of coin. $Pr(X_2) = 0.3,\ p = 0.7$.

$$Pr(Y = 6\ |X_1) = \binom{5}{2} 0.5^3\ (1 - 0.5)^{6-3} = \frac{5}{32} = 0.15625$$

$$Pr(Y = 6\ |X_2) = \binom{5}{2} 0.7^3\ (1 - 0.7)^{6-3} = \frac{9261}{100000} = 0.09261$$

According to the PMF obtained in (a),

$$Pr(Y = 6) = 0.7 \binom{5}{2} (0.5^6 + 0.7^2 \cdot 0.3^4) = \frac{68579}{100000} = 0.137158$$

The question is asking about $Pr(X_1\ |Y = 6)$. According to *Bayes' Theorem*:

$$Pr(X_1\ |Y = 6) = \frac{Pr(Y = 6\ |X_1) Pr(X_1)}{Pr(Y = 6)} = \frac{0.15625 \times 0.7}{0.137158} = \frac{109375}{137158} = 0.797438$$

(c) Let Y denote the number of trials until the first head is obtained. $Y \sim Geom(p)$.
Let W denote the event that is takes at most 4 tosses to obtain the first head.
Then we have $Pr(W) = Pr(Y = 1) + Pr(Y = 2) + Pr(Y = 3) + Pr(Y = 4)$.
Let X_1 denote the event that Peter picks the first type of coin. $Pr(X_1) = 0.7,\ p = 0.5$.

$$Pr(W|X_1) = (1-0.5)^{1-1}\,0.5+(1-0.5)^{2-1}\,0.5+(1-0.5)^{3-1}\,0.5+(1-0.5)^{4-1}\,0.5 = \frac{15}{16} = 0.9375$$

So the probability that Peter correctly identifiers a randomly chosen coin is:

$$Pr(W \cap X_1) = Pr(W|X_1)Pr(X_1) = 0.9375 \times 0.7 = 0.65625$$

Question 3: Suppose there are two separate polling agencies interested in gauging the opinions of UBC students regarding how the current provincial goverrnment has been dealing with the proposed Kinder Morgan pipeline expansion from Edmonton to Burnaby. Both polling agencies will ask random UBC students to answer "Yes" or "No" to the following question: *I generally approve of how the current BC government is handling the proposed Kinder Morgan pipeline expansion.*

(a) The two agencies first run a test poll on a random sample of students from a UBC Statistics class containing 20 students. Polling agency A samples 10 students *without replacement* from the class to answer the question, whereas polling agency B samples 10 students *with replacement* from the same class. Identify the random variables, A and B, that could model these two polling procedures and calculate $\Pr(5 \leq A \leq 7)$ and $\Pr(5 \leq B \leq 7)$. Which sampling scheme seems better and why? Suppose that 12 out of the 20 students in the class would answer "Yes" to the polling question.

(b) Now suppose the polling agencies are ready to sample a large proportion of UBC students on Main Mall. Assume the student population of UBC is 20,000 and that 60% of the student population would answer "Yes" to the polling question. If the two polling agencies sample 1000 student opinions using their respective sampling schemes, and we use A and B to denote the number of sampled students answering "Yes" to the polling question from each agency respectively, what is $\mathbb{E}(A)$ and $\mathbb{E}(B)$?

(c) Now find $\Pr(\mathbb{E}(A) - 15 \leq A \leq \mathbb{E}(A) + 15)$ and $\Pr(\mathbb{E}(B) - 15 \leq B \leq \mathbb{E}(B) + 15)$. How do the differences in these probabilities compare to the differences in the probabilities you found in part (a)? Do you still prefer one sampling scheme to the other? Why or why not?

(d) *[Optional bonus question]* Show that as the population size N grows, the probability mass function of a hypergeometric random variable approaches the probability mass function of a binomial random variable, assuming a constant chance of success. That is, show

$$\lim_{N\to\infty} \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} = \binom{n}{x} p^x (1-p)^{n-x},$$

where $p = \frac{m}{N}$ is constant. *Hint:* Rewrite the choose functions in terms of factorials and remember that, as we have seen in class, the highest order terms will dominate in the limit.

Question 4: You are part of a field team in the Strait of Georgia tasked with surveying the nesting habits of a colony of Pelagic Cormorants (*Phalacrocorax pelagicus*), an iconic and precipitously declining species native to coastal British Columbia. Your supervisor wants to ensure the team is big enough to monitor all cormorant pairs in the colony at least 90% of the time. Each member of the field team can reliably observe 5 cormorant pairs simultaneously.

(a) The expected number of cormorant pairs at the colony at any one time is known to be 20. How big should the field team be to ensure that at least 90% of the time all pairs can be observed? (*Hint:* use R or an online applet to calculate the appropriate probabilities.)

(b) Due to budget considerations, your supervisor was only able to recruit 3 people for the field team (yourself and two others). Unfortunately, when you arrive at the colony of study, you find that the colony is twice the size as anticipated. with twice the average number of cormorant pairs present at any one time. What is the probability that your team of 3 can observe all the cormorant pairs at any one time? (Your supervisor is busy with other tasks so can't help with the observations....)

Question 5: Suppose the amount of energy (in kilocalories), E, provided by the sun to a random arbutus tree on a winter day in Victoria follows a random variable characterized by the following probability density function:

$$f(x) = \begin{cases} cx\sin(x) & \text{if } 0 \leq x \leq \pi \\ 0 & \text{otherwise} \end{cases}$$

where c is a fixed constant.

(a) Find the value of c that makes $f(x)$ an honest PDF.

(b) Find the cumulative distribution function of E.

(c) Suppose we would like to study how much energy arbutus trees around James Bay in Victoria are absorbing. We start sampling trees at random in James Bay and recording their energy intake from the sun. How likely is it that we have to sample more than 10 trees to find the first with an energy intake exceeding 3 kilocalories?