

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. $X_1 \sim \text{Bin}(1, 0.5)$.
Let X_2 denote the value by flipping the coin 2 times, i.e., drawing ball 2. $X_2 \sim \text{Bin}(2, 0.5)$.
Let X_3 denote the value by flipping the coin 3 times, i.e., drawing ball 3. $X_3 \sim \text{Bin}(3, 0.5)$.
The probability of drawing ball 1, 2, 3 is equal to $1/3$.

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

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

(1)

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

$$\begin{aligned} Pr(Y = 3) &= \frac{1}{3} \times Pr(X_3 = 3) \\ &= \frac{1}{3} \times \left(\binom{3}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^0 \right) \\ &= \frac{1}{24} \end{aligned}$$

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:

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

Standard Deviation:

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

(c) The 3 experiments are independent.

Let Y_1 denote the result we get from the first experiment. From (b) we know that:

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

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

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

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

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

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 $X_1 \sim \text{NegBin}(3, 0.5)$.

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 $X_2 \sim \text{NegBin}(3, 0.7)$.

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,

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

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:

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

(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

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

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

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

(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

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

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

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

(3) Finally, we compute the variance:

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

(b) Let Y denote the number of tosses needed to obtain 3 heads. $Y \sim \text{NegBin}(3, p)$.

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. $Y \sim \text{Geom}(p)$.

Let W denote the event that it 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 identifies 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:** (a) A is a Hypergeometric random variable with parameters $N = 20$, $n = 10$, $m = 12$.
So $A \sim \text{Hypergeom}(20, 10, 12)$.

$$\begin{aligned} \Pr(5 \leq A \leq 7) &= \Pr(A = 5) + \Pr(A = 6) + \Pr(A = 7) \\ &= \frac{\binom{12}{5}\binom{8}{5}}{\binom{20}{10}} + \frac{\binom{12}{6}\binom{8}{4}}{\binom{20}{10}} + \frac{\binom{12}{7}\binom{8}{3}}{\binom{20}{10}} = \frac{153384}{184756} = 0.8301976661 \end{aligned}$$

B is a Binomial random variable with parameters $n = 10$, $p = 12/20 = 0.6$.

So $B \sim \text{Bin}(10, 0.6)$.

$$\begin{aligned} \Pr(5 \leq B \leq 7) &= \Pr(B = 5) + \Pr(B = 6) + \Pr(B = 7) \\ &= \binom{10}{5} 0.6^5 0.4^5 + \binom{10}{6} 0.6^6 0.4^4 + \binom{10}{7} 0.6^7 0.4^3 = 0.6664716288 \end{aligned}$$

A 's sampling scheme seems better because there is no "strengthen" of repeated opinion in the final result, so A 's scheme can possibly represent a larger number of students than B does.

- (b) A is a Hypergeometric random variable with parameters $N = 20000$, $n = 1000$, $m = 20000 \times 60\% = 12000$.

So $A \sim \text{Hypergeom}(20000, 1000, 12000)$.

$$\mathbb{E}(A) = \frac{nm}{N} = \frac{1000 \times 12000}{20000} = 600$$

B is a Binomial random variable with parameters $n = 1000$, $p = 60\% = 0.6$.

So $B \sim \text{Bin}(1000, 0.6)$.

$$\mathbb{E}(B) = np = 1000 \times 0.6 = 600$$

- (c)

$$\Pr(\mathbb{E}(A) - 15 \leq A \leq \mathbb{E}(A) + 15) = \sum_{k=585}^{615} \frac{\binom{12000}{k}\binom{8000}{n-k}}{\binom{20000}{1000}} = 1$$

$$= \text{phyper}(615, 12000, 8000, 1000) - \text{phyper}(584, 12000, 8000, 1000) = 0.695355$$

$$\Pr(\mathbb{E}(B) - 15 \leq B \leq \mathbb{E}(B) + 15) = \sum_{k=585}^{615} \binom{1000}{k} 0.6^k 0.4^{1000-k}$$

$$= \text{pbinom}(615, \text{size}=1000, \text{prob}=0.6) - \text{pbinom}(584, \text{size}=1000, \text{prob}=0.6) = 0.6829448$$

$$\text{Diff}_a = 0.8301976661 - 0.6664716288 = 0.1637260373$$

$$\text{Diff}_c = 0.695355 - 0.6829448 = 0.0124102$$

There difference in probabilities found in (c) is significantly smaller than the difference in probabilities found in (a). Therefore, I do not prefer one sampling scheme to the other. When the sample space gets bigger and bigger, the resulting probabilities chosen between binomial and hypergeometric gets closer and closer. If the sample space is large enough, the difference in resulting probabilities becomes insignificant.

(d) [Optional bonus question] **Proof:**

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}} &= \lim_{N \rightarrow \infty} \frac{\frac{m!}{x!(m-x)!} \cdot \frac{(N-m)!}{(n-x)!(N-m-(n-x))!}}{\frac{N!}{n!(N-n)!}} \\ &= \lim_{N \rightarrow \infty} \frac{n!}{x!(n-x)!} \cdot \frac{m!(N-m)!(N-n)!}{(m-x)!(N-m-(n-x))!N!} \\ &= \frac{n!}{x!(n-x)!} \lim_{N \rightarrow \infty} \frac{(m-x+1)\dots m \cdot (N-m-(n-x)+1)\dots(N-m)}{(N-n+1)\dots N} \\ &= \frac{n!}{x!(n-x)!} \left(\frac{m}{N}\right)^x \left(\frac{N-m}{N}\right)^{n-x} \\ &= \binom{n}{x} p^x (1-p)^{n-x} \end{aligned} \tag{2}$$

line 1 just expand the combinations.

line 2 rearranges the division and pulls out $\frac{n!}{x!(n-x)!}$

line 3 cancels same items in $m!$ with $(m-x)!$, $(N-m)!$ with $(N-m-(n-x))!$, $N!$ with $(N-n)!$.

line 4, notice that there are x items in $(m-x+1)\dots m$, and there are $(n-x)$ items in $(N-m-(n-x)+1)\dots(N-m)$, and there are n items in $(N-n+1)\dots N$, also $n = x + n - x$. So when we take the limit, we can replace all items in $(m-x+1)\dots m$ with m^x , $(N-m-(n-x)+1)\dots(N-m)$ with $(N-m)^{n-x}$, $(N-n+1)\dots N$ with $N^x N^{n-x}$ because they are asymptotically equivalent when taking the limit. Therefore we get the result in line 4.

line 5, notice that expression in line 4 is just the expanded form of line 5, after replacing $\frac{m}{N}$ with p .

Q.E.D.

Question 4: (a) Let X be the number of cormorant pairs appearing at the colony. $X \sim \text{Poisson}(20)$

$$p(x) = \frac{20^x e^{-20}}{x!}$$

We would like to know the value of k , such that:

$$\begin{aligned} p(0) + p(1) + p(2) + \dots + p(k) &\geq 90\% \\ \frac{20^0 e^{-20}}{0!} + \frac{20^1 e^{-20}}{1!} + \frac{20^2 e^{-20}}{2!} + \dots + \frac{20^k e^{-20}}{k!} &\geq 90\% \end{aligned}$$

Using R, we notice that `ppois(25, lambda=20) = 0.887815`, and `ppois(26, lambda=20) = 0.9221132`. Therefore $k = 26$.

So the expected number of the team size is: $26/5 = 5.2$ member.

- (b) 3 members can observe
- $3 \times 5 = 15$
- cormorant pairs.

Let X be the number of cormorant pairs appearing at the colony. $X \sim \text{Poisson}(20 \times 2)$

$$p(x) = \frac{40^x e^{-40}}{x!}$$

The probability that 3 members can observe all the cormorant pairs at any one time is:

$$p(0) + p(1) + p(2) + \dots + p(15) = \sum_{k=0}^{15} \frac{40^k e^{-40}}{k!} = \text{ppois}(15, \text{lambda}=40) = 5.463981 \times 10^{-6}$$

Question 5: (a)

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\pi} c x \sin(x) dx = 1$$

$$\int_0^{\pi} c x \sin(x) dx = c((x \cdot -\cos(x))|_0^{\pi} - \int_0^{\pi} -\cos(x) dx) = c(\pi - 0 + \sin(x)|_0^{\pi}) = c\pi$$

$$c\pi = 1 \Rightarrow c = \frac{1}{\pi}$$

(b)

$$\int \frac{1}{\pi} x \sin(x) dx = \frac{1}{\pi}(-x \cos(x) + \int \cos(x) dx) = \frac{1}{\pi}(-x \cos(x) + \sin(x) + C)$$

In this case $C = 0$. Therefore, the CDF of E is:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{\pi}(-x \cos(x) + \sin(x)), & 0 < x < \pi \\ 1, & x \geq \pi \end{cases}$$

(c)

$$p = \Pr(X > 3) = 1 - \Pr(X \leq 3) = 1 - \frac{1}{\pi}(-3 \cos(3) + \sin(3)) = 0.00970690954 \approx 0.01$$

Let Y denote the number of samples we take until we find the first with an energy intake exceeding 3 kilocalories.

$$Y \sim \text{Geom}(0.01)$$

The probability that we have to sample more than 10 trees is:

$$\Pr(Y > 10) = 1 - \left(\sum_{k=1}^{10} \Pr(Y = k)\right) = 1 - \left(\sum_{k=1}^{10} 0.99^{k-1} 0.01\right) = 1 - \text{pgeom}(10, 0.01) = 0.8953383$$