Stat 88 Midterm Practice Solutions

Spring 2021 Course Staff

March 2, 2021

Cupcakes

- For the first part, let X be the number of chocolate cupcakes that Nancy takes. Then X follows $Hypergeometric(30, 10, 5)$. So the answer is $\sum_{k=3}^{5} \frac{\binom{10}{k}\binom{20}{5-k}}{\binom{30}{5}}$
- For the second part, just uses the expectation formula for Hypergeometric distribution: $n\frac{G}{N} = 5 * \frac{10}{30} = \frac{5}{3}$
- For the third part, notice that Adalie already takes 3 vanilla and 2 chocolate cupcakes, so there are 25 cupcakes left in total, among those 7 are vanilla, 8 are chocolate, and 10 are strawberry. Therefore, the conditional distribution of the number of vanilla cupcakes that Emily takes is $Hypergeometric(25, 7, 5)$ so the answer is $\frac{\binom{7}{2}\binom{18}{3}}{\binom{25}{5}}$

Geckos

- (1): Let X be the number of green or white geckos in Samarth's sample. We know the population size is N and the sample size is n. We also know 20% of the population is green and 10% is white, meaning the total number of green or white geckos is $(0.30)N$. Hence $X \sim Hypergeometric(N, (0.30)N, n)$.
- (2): Let Y be the number of red geckos in Samarth's sample. Then $Y \sim Hypergeometric(N, (0.40)N, n)$. Using the formula for the PMF of a hypergeometric, $P(Y < \frac{n}{2}) = \sum_{k=0}^{\frac{n}{2}-1} \frac{\binom{(0.40)N}{k}\binom{(0.60)N}{n-k}}{\binom{N}{n}}$.
- (3): Let C be the number of colors that don't appear in Samarth's sample. Then $C = I_G + I_R + I_B + I_W$, where I_i is the indicator that color i doesn't appear in the sample. By linearity of expectation, $E[C] = E[I_G + I_R + I_B + I_W] = E[I_G] + E[I_R] + E[I_B] + E[I_W]$. The expectation of any indicator is the probability said indicator equals one, which in this scenario occurs when the specific color does not appear in the sample. Hence $E[C] = (0.20)^n + (0.30)^n + (0.40)^n + (0.10)^n$.
- (4): The quantity in (c) is decreasing in n because as the sample size increases, it becomes less likely for a color to not be chosen. The quantity in (c) does not depend on N at all, since we define our problem in terms of population proportions that do not depend on the total population size.

GSIs and a Game of Chance

- For the first part, the answer is $\sum_{k=4}^{10} \binom{10}{k} \frac{1}{6}^k \frac{5}{6}^{10-k}$
- Lets call the above probability a.
- For the second part, the trials are now the 12 GSI's, each with probability a of winning. The number of GSI's winning has the Binomial(12,a) distribution. The expected value is $12 \times a$.
- For the third part, we notice that Nancy is the 10th GSI. For her to be the 4th GSI to win, there needs to be exactly 3 GSI's winning before her and then she wins. The answer is $\binom{9}{3}a^3(1-a)^6 a$

Statistical Stocks

- $0 \leq P(A^C \cap B^C \cap C^C) \leq 0.5$ or $1 - 1 \leq 1 - P(A \cup B \cup C) \leq 1 - 0.5$
- A has no overlap with either B or C but B and C can overlap. $0.1 \leq P(B, C) \leq 0.2$
- $P(B|C) = 0.4$ which tells us that event B is completely contained inside C. $0 \leq P(A, B, C) \leq 0.2$

Tough Enemy

- This question tests finding expectation by conditioning use the recursion technique. The most similar reference is Chapter 5 Exercise 15. Let x be the expected amount of hits for Shide to finish his enemy. By the setup, we have: $x = 7 + (0.05 \cdot 1 + 0.15 \cdot x + 0.8 \cdot 8)$. Solving this equation yields $x = 13.45/0.85 \approx 16$.

Monopoly

- Part a: We need to find the probability that he doesn't win a game in the first nine games. That is: $(0.8)^9$
- Part b: In this case, we need to find the probability that he wins his third game on exactly the 12th game. That is equal to : P(2 wins in 11 games) * P(win on the 12th game) (Binom (11, 0.2) evaluated at k = 2) * P(win on the 12th game) $\binom{11}{2} * (0.2^2) * (0.8^9) * (0.2)$

Witchcraft: The Meeting

All three parts of this question rely on the realization that for large samples we can approximate a Binomial distribution as a Poisson distribution with mean $\mu = np$. This will make the number of blue dragons $\sim Poisson(2)$ and the red dragons will be $\sim Poisson(1)$.

- For part a, if we call X the number of blue dragons we find:$P(X \geq 2) = 1 - P(X = 1) - P(X = 0) = 1 - \frac{e^{-2}2^0}{0!} - \frac{e^{-2}2^1}{1!} = 1 - 3e^{-2}$
- For part b, we use the additivity of Poisson random variables, to say that if X is the number of blue dragons, and Y is the number of red dragons, $X + Y \sim Poisson(2 + 1)$ Thus, $P(X + Y = 4) = \frac{e^{-3}3^4}{4!}$
- Finally, for part c, we need to first solve for how many dragons need to be found in packs in order to gain money (make at least \$0). $30(X+Y) - 100 \geq 0 \rightarrow X+Y \geq 100/30 = 10/3$. Thus, because we can only get integer numbers of cards, we will be finding $P(X + Y > 3)$. $P(X + Y > 3) = 1 - P(X + Y \leq 3) = 1 - \sum_{i=0}^{3} \frac{e^{-3}3^i}{i!}$

Zoning Zones

- Yes, Anton is correct. Recall that we assumed that the sample was i.i.d. and so each individual resident has $E(X_i) = \mu$ for each i
- The maximum is not usually an unbiased estimator of the mean because it will always return the largest value of the sample. It is unlikely that the expectation of the largest value of a sample is going to be equal to the average of the population.
- If for the sample every single resident has the exact same income than the the maximum of the sample is the average of the sample. This is obviously not a realistic nor practical assumption.
- In order to solve this problem we need the condition for unbiasedness to hold: parameter $= E(estimate)$. Our parameter value is 50000, so $50000 = E\frac{(a75000+b50000+c30000+d100000)}{100}$. Let a, b, c, d be any combination of values whose weighted average is 50,000. One example is to let b = 100 and the rest be 0.

Drawing Faces

- If Jamarcus has removed one face card from the deck, then the distribution of the number of face cards in Danny's hand is Hypergeometric(N=51, G=11, n=13). Since the expectation of a hypergeometric distribution is $n * \frac{G}{N}$ then the expected number of face cards in Danny's hand is $13 * \frac{11}{51}$.
- We condition on the number of face cards Jamarcus removes from the deck. Let X = number of face cards Jamarcus removes from the deck. The distribution of X is binomial(3, $\frac{1}{2}$). Let Y = number of face cards Danny has in his hand. $E(Y) = E(E(Y|X)) = P(X = 0)E(Y|X = 0) + P(X = 1)E(Y|X = 1) + P(X = 2)E(Y|X = 2) + P(X = 3)E(Y|X = 3)$. Using binomial probabilities, this expression becomes $\frac{1}{8}E(Y|X = 0) + \frac{3}{8}E(Y|X = 1) + \frac{3}{8}E(Y|X = 2) + \frac{1}{8}E(Y|X = 3)$. If Jamarcus has removed x cards from the deck, then the number of cards in Danny's hand has distribution Hypergeometric(N=52-x, G=12-x, n=13). Therefore, $E(Y) = \frac{1}{8} * 13 * \frac{12}{52} + \frac{3}{8} * 13 * \frac{11}{51} + \frac{3}{8} * 13 * \frac{10}{50} + \frac{1}{8} * 13 * \frac{9}{49} = 2.7$.

Mailman

- $P(\text{first mailbox is empty}) = (1 - \frac{1}{n})^{2n}$
- $\log(P(\text{first mailbox is empty})) = \log((1 - \frac{1}{n})^{2n}) = 2n \log(1 - \frac{1}{n})$.

  Since n is large, $\frac{1}{n}$ is small, so we have $\log(1 - \frac{1}{n}) \approx -\frac{1}{n}$.

  Taking the exponential, we have $(P(\text{first mailbox is empty})) \approx e^{2n(-\frac{1}{n})} = e^{-2}$
- Binomial($2n$, $\frac{n-1}{n}$), because for each of the $2n$ letters, the change of not being in box 1 is $\frac{n-1}{n}$, independent of other letters.
- The Law of Small Numbers tells us that we can approximate a Binomial(n, p) distribution to Poisson(np). However, the key thing is to recognize that $\frac{n-1}{n}$ is close to 1 as n gets larges, and is therefore a large probability. To use Law of Small Numbers, we need to use the complement event of "not being in the first box".

Let Y be the number of letters **in** the first mailbox. Then Y follows a Binomial($2n, \frac{1}{n}$) distribution.

$2n$ is large and $\frac{1}{n}$ is small, so we can now use the Poisson approximation. Using the approximation, Y follows a Poisson(2) distribution.

Going back to X, which is $n - Y$, we have $P(X = n) = P(Y = 0) \approx e^{-2}\frac{2^0}{0!} = e^{-2}$. This is the same as the probability in part (2).

Library Policies

- I is a Bernoulli random variable and $E[I] = P(I = 1)$. We need to condition a student's opinion on his or her current class, i.e. $P(I = 1) = P(I = 1|Freshman)P(Freshman) + P(I = 1|Sophomore)P(Sophomore) + P(I = 1|Junior)P(Junior) + P(I = 1|Senior)P(Senior) = \frac{1}{4}(0.5 - 2\theta + 0.5 - 1\theta + 0.5 + 0.5 + 3\theta) = 0.5$

- By linearity of expectation, $E[\bar{X}] = \frac{1}{300}(\sum_{k=1}^{100} I_{freshman,k} + \sum_{k=1}^{100} I_{sophomore,k} + \sum_{k=i}^{100} I_{junior,k}) = \frac{1}{300}(100 * E[I_{freshman,i}] + 100 * E[I_{sophomore,i}] + 100 * E[I_{junior,i}]) = \frac{1}{3}(1.5 - 3\theta) = 0.5 - \theta$. Therefore $E[0.5 - \bar{X}] = \theta$ and $0.5 - \bar{X}$ is an unbiased estimator of θ.

- Condition on $N = n$, $E[\frac{Y}{N}|N = n] = E[\frac{Y}{n}] = \frac{1}{n}E[Y] = \frac{n(0.5-2\theta)}{n} = 0.5 - 2\theta$. The second to last equation follows from the expectation of Binomial since $Y|N = n \sim Binomial(n, 0.5 - 2\theta)$. $E[\frac{Y}{N}] = \sum_{n=51}^{100} E[\frac{Y}{N}|N = n]P(N = n) = \sum_{n=51}^{100}(0.5-2\theta)\cdot\frac{1}{50} = 0.5-2\theta$. Therefore, $E[\frac{1}{2}(0.5-\frac{Y}{N})] = \theta$. $0.25 - 0.5 \cdot \frac{Y}{N}$ is an unbiased estimator of θ.

Biased Coins

- Using Bayes' theorem, we want to find: $P(headcoin \mid H)$, where:

$$P(head\ coin|H) = \frac{P(H|head\ coin) * P(head\ coin)}{P(H)}$$

where $P(head\ coin)$ is the probability of getting the two-headed coin, and by law of total probability, we can get $P(head)$:

$$\begin{aligned}
&= \frac{P(H|head\ coin) * P(head\ coin)}{P(H|head\ coin) * P(head\ coin) + P(H|fair\ coin) * P(fair\ coin) + P(H|tail\ coin) * P(tail\ coin)} \\
&= \frac{1 * \frac{1}{3}}{1 * \frac{1}{3} + \frac{1}{2} * \frac{1}{3} + 0 * \frac{1}{3}} \\
&= \frac{2}{3}
\end{aligned}$$

- We want to find $P(HHH|HH)$. Using the formula of conditional probability,

$$\begin{aligned}
P(HHH|HH) &= \frac{P(HHH, HH)}{P(HH)} \\
&= \frac{P(HHH)}{P(HH)}
\end{aligned}$$

since $P(HHH, HH) = P(HHH)$, as we are finding the intersection of HHH and HH. HH is a proper subset of HHH (HH is completely contained in HHH: for HHH to happen, HH must happen)

$$\begin{aligned}
&= \frac{P(HHH|head) * P(head) + P(HHH|fair) * P(fair) + P(HHH|tail) * P(tail)}{P(HH|head) * P(head) + P(HH|fair) * P(fair) + P(HH|tail) * P(tail)} \\
&= \frac{1^3 * \frac{1}{3} + (\frac{1}{2})^3 * \frac{1}{3} + 0^3 * \frac{1}{3}}{1^2 * \frac{1}{3} + (\frac{1}{2})^2 * \frac{1}{3} + 0^2 * \frac{1}{3}} \\
&= \frac{9}{10}
\end{aligned}$$