Stat88 midterm

February 2020

Problem 1 (Conditional expectation and Geometric)

A box has three coins with probability of landing head 0.3, 0.5, 0.9. You randomly reach in the box and grab one of the coins. What is the expected numbers of flips till you get a head?

Solution

We condition on which coin we have. The expected number of p-coin tosses till we get a heads is $\frac{1}{p}$. Hence the expectation is $1/3(10/3 + 10/5 + 10/9)$.

Mistakes
- Solutions where all the probabilities were summed up received no credit.
- Some students mistakenly thought the probabilities given were for drawing a specific coin and proceeded to incorrectly assume all the coins were fair coins.
- Some students mistakenly modeled the event as binomial distribution with three trials
- Lots of students correctly used the inverse of the probability for the expected number of heads for each coin but did not weigh it by the conditional probability of actually drawing that coin.
- Lots of students correctly knew to weight each coin by (the conditional probability of drawing that coin) but then used the probability as the expectation of the geometric distribution.
- Some students tried to find a new "p" value by weighting the probabilities by 1/3 and took the inverse of that but that was considered as the same error as the above common mistake.
- Some students used indicators to sum up the total heads instead of using expectation by conditioning.

Problem 2 (Prosecutor's fallacy,base rate and likelihood) A woman's DNA matches that of a sample found at a crime scene. If the woman is guilty the chances of a DNA match is virtually 100% (i.e. the chance of a false negative is essentially zero). Can we say that given a DNA match, the woman must be guilty? Explain your reasoning fully using the concepts discussed in Stat 88.

Solution No, this is the prosecutor's fallacy. You must take the base rate and the chance of a false positive into account using Bayes' rule.

Mistakes - Incorrectly assuming that P(Guilty—Match) = P(Match—Guilty) and therefore saying yes
- Just stating that P(Guilty—Match) does not equal P(Match—Guilty) without using some sort of Bayes Rule/False Positive explanation

Problem 3 (Binomial, Poisson approximation, complement rule, $P(X = Y)$) You're curious to find out the chance that UC Berkeley students are ambidextrous (the rare ability to be both left and right handed). Using historical data, you find out that men students have a $\frac{1}{500}$ chance of being ambidextrous, independent from other men. On the other hand, women students have a $\frac{1}{250}$ chance of being ambidextrous, independent from other women. Assume independence across men and women.

a) Suppose there's a sample of 100 men and another sample of 100 women. What is the chance that both samples have an equal number of students that are ambidextrous?

b) Suppose now that there's a sample of 1000 men and 1000 women. Let T be the total number of students that are ambidextrous from both samples. What is the **approximate** distribution of T? Use this distribution to find $P(T > 10)$.

Solution

a) $\sum_{k=0}^{100} \binom{100}{k}(\frac{1}{500})^k * (\frac{499}{500})^{100-k} \cdot \binom{100}{k}(\frac{1}{250})^k * (\frac{249}{250})^{100-k}$

b) Poisson(6) (sum of independent Poisson is Poisson and since when n is large and p is small Poisson(np) approximates Binomial(n,p)). $P(T > 10) = 1 - P(T <= 10) = 1 - \sum_{i=0}^{10} \frac{e^{-6} * 6^i}{i!}$

Mistakes For part a
- Using poisson instead of binomial
- Forgot the summation, or using two summations one before each binomial

For part b
- Not using poisson
- Wrong summation indices, typically summing up to 2000 for poisson(6)]
- Not summing poisson parameters and instead multiplied poisson(2) * poisson(4)

Problem 4 (conditional probability)

Two jars each contains r red marbles and b blue marbles. A marble is chosen at random from the first jar and placed in the second jar. A marble is then randomly chosen from the second jar. Find the probability this marble is red.

Solution

Let X=the first marble color and Y=the second marble color. Let R be red and B be blue. By the multiplication rule and addition rules,

$$P(Y = R) = P(Y = R|X = R)P(X = R) + P(Y = R|X = B)P(Y = B).$$

This gives

$$\left(\frac{r+1}{r+1+b}\right)\left(\frac{r}{r+b}\right) + \left(\frac{r}{r+1+b}\right)\left(\frac{b}{r+b}\right)$$

Mistakes

- r / (r+b) with either no justification or "by symmetry"
- r / (r + b + 1) + (r + 1) / (r + b + 1), where there is no weighting of these probabilities
- misunderstanding the question as one jar full of red balls, the other all blue balls
- some say P(red second) = P(red second given blue first) + P(red second given red first), which is wrong because P(red second) = P(red second, blue first) + P(red second, red first)

Problem 5 (indicator method, expectation, hypergeometric)

A class of 60 students includes 20 seniors. For a group project, the class is split at random without replacement into 10 groups of 6 students each. Find the expected number of groups that contain no seniors.

Solution

We follow the 5 steps for method of indicators discussed in class.

Step 1: Let X=the number of groups, out of 10, with no seniors.
Step 2: I_2 is 1 if the second group has no seniors.
Step 3:

$$p = E(I_2) = \frac{\binom{20}{0} * \binom{40}{6}}{\binom{60}{6}}$$

Step 4: $X = I_1 + \ldots + I_{10}$
Step 5: $E(X) = 10p$ where p is defined above.

Mistakes - Found the expectation of seniors as if drawing 6 people, instead of 10 groups of 6 people Used 6 indicators instead of 10 (Very common, be sure to know how to figure out how many indicators to use)

- Used method of indicators but found incorrect $P(I_j = 1)$, which equals to $E(I_j)$

Problem 6 (Baye's rule, binomial, hypergeometric)

A die has 2 red faces and 4 green faces. The die is rolled 13 times. Given that green faces appeared exactly 7 times in 13 rolls, what is the chance that the green faces appeared exactly 3 times in the first 5 rolls?

Please look to see if you can simplify your answer algebraically.

Solution

Let X= the number of green faces in the first 5 rolls.
Let Y= the number of green faces in all 13 rolls.
You can think of your population as 13 rolled die, 7 of which have a green face (Good) lined up randomly (SRS) into a group of 5 followed by a group of 8. We are asked to find the chance that 3 out of the group of 5 are green. Since we have population of size 13, 7 of which are Good, and a sample size of 5 we may apply the HG formula. More precisely, $X|Y = 7 \sim HG(13, 7, 5)$ so by HG formula

$$P(X = 3|Y = 7) = \frac{\binom{7}{3}\binom{6}{2}}{\binom{13}{5}}.$$

Alternatively, the answer can be worked out using Baye's rule and the binomial formula.

$$P(X = 3|Y = 7) = \frac{P(Y = 7|X = 3)P(X = 3)}{P(Y = 7)}.$$

Notice that $P(Y = 7|X = 3)$ is the probability of 4 green in the last 8 rolls. Each of the terms can be found with the binomial formula.

$$P(X = 3|Y = 7) = \frac{\binom{8}{4}(2/3)^4(1/3)^4\binom{5}{3}(2/3)^3(1/3)^2}{\binom{13}{7}(2/3)^7(1/3)^6} = \frac{\binom{8}{4}\binom{5}{3}}{\binom{13}{7}} = \frac{\binom{7}{3}\binom{6}{2}}{\binom{13}{5}}.$$

The last equality can be seen by writing out the definition of the choose terms and doing some algebra, but either answer will receive full credit.

Mistakes - Calculated joint probability instead of conditional probability
- Did not take into account the later 8 rolls in numerator
- Add up instead of multiplying the two independent binomials