Math 318 – homework 2 – due 2023-01-27

Problem 1. A government wants to find out how many of its citizens are filling out fraudulent tax returns. It creates a survey with the single question "Do you fill out your tax return honestly?" Along with the question come the following *Instructions:* Toss a fair coin. If the result of the coin toss is heads, answer the question truthfully. If the result of the coin toss is tails, answer the question with NO regardless of whether you cheat or not.

Randomising the response makes it impossible to deduce for any single individual that they cheat. The result of the survey is that 45% of the respondents say YES. We interpret this as meaning that a randomly chosen member of the population will answer the survey YES with probability 0.45. Find the proportion of the population that fills out fraudulent returns.

Solution. Suppose the proportion of people who cheat is p. The only people who answer YES are those who are honest and get heads. The probability that a random person is honest is $1-p$, and the probability of heads is $\frac{1}{2}$. These are independent, so $\frac{1-p}{2} = 0.45$, and so $p = 0.1$.

A bit more precisely, the sample space used is $S = P \times \{H, T\}$ where P is the set of people, and all people are equally likely, and the coin is independent. We have $0.45 = P(\mathit{honest} \cap H) = P(\mathit{honest})P(\mathit{head})$, and so $P(\mathit{honest}) = 0.9$.

Problem 2. Let A, B be independent events. Show that A, B^c are independent; that A^c, B are independent; and that A^c, B^c are independent.

Solution. We have $A = (A \cap B) \cup (A \cap B^c)$ as a disjoint union. Therefore $P(A) = P(A \cap B) + P(A \cap B^c)$. If A, B are independent then $P(A \cap B) = P(A)P(B)$, and this becomes

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

so that A, B^c are independent.

An alternative: $P(B^c|A) = 1 - P(B|A)$. If A, B are independent then this is $1 - P(B)$ so $P(B^c|A) = P(B^c)$ and A, B^c are independent.

We have seen that for independent events, replacing one event by its complement maintains independence. Repeating this we get that also A^c, B^c are independent.

Note. This works also for more than two events: you can replace some or all by their complements.

Problem 3. Let A, B, C be independent events. Show that

$$P(A \cup B \cup C) = 1 - (1 - P(A))(1 - P(B))(1 - P(C))$$

.

Solution. Using inclusion-exclusion we have

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= P(A) + P(B) + P(C) - P(A)P(B) - P(A)P(C) - P(B)P(C) + P(A)P(B)P(C) \\ &= 1 - (1 - P(A))(1 - P(B))(1 - P(C)). \end{aligned}$$

The second step uses independence, and the third is verified by expanding the parentheses.

Alternatively, we have $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$. From the previous problem, we know these are independent, so

$$P((A \cup B \cup C)^c) = P(A^c)P(B^c)P(C^c).$$

Writing $P(E^c) = 1 - P(E)$ for every event gives the desired identity.

Problem 4. Two teams play a series of games, each of which is won by Team A with probability p and by Team B with probability $1-p$. The winner of the series is the first team to win 5 games, so at most 9 games are played. Find the probability that a total of 9 games are played, and show that this probability is maximal when $p = \frac{1}{2}$.

Solution. The 9th game is played if and only if the first 8 games are 4-4. The probability of this is $\binom{8}{4}p^4(1-p)^4 = 70(p-p^2)^4$.

We have that $p - p^2$ is maximized when $p = 1/2$, and so the same holds for $70(p-p^2)^4$.

Problem 5. A True/False question is posed to a team with two members. Each team member independently gives the correct answer with probability p.

(a) Which of the following is the better strategy for the team?
 (i) Choose one team member at random and let that member answer the question.
 (ii) The two members decide on their answers and consult. If they agree, that is the team's answer. If they disagree, they flip a fair coin to pick the team answer.

(b) Suppose $p = 0.6$ and the team adopts strategy (ii). What is the conditional probability that the team gives the correct answer given that the team members agree? What is the conditional probability that the team gives the correct answer given that the team members disagree?

Solution.

(a) In the first strategy, they get the answer correctly with probability p. In the second, they get it correctly if both members are right (prob. p^2), wrong if both are wrong (prob. $(1-p)^2$). If the members disagree (prob. $2p(1-p)$) they choose based on a coin toss, and get the point with conditional probability $1/2$. Let $X \in \{0,1,2\}$ be the number of members who are correct, and A the event that they answer correctly. Using the law of total probability we get

$$P(A) = \sum_i P(X=i)P(A|X=i) = (1-p)^2 \cdot 0 + 2p(1-p) \cdot \tfrac{1}{2} + p^2 \cdot 1 = p.$$

Therefore the probability is exactly the same.

(b) The team agrees is X is 0 or 2. We have

$$P(A, X \in \{0,2\}) = (1-p)^2 \cdot 0 + p^2 \cdot 1 = p^2,$$

and $P(X \in \{0,2\}) = (1-p)^2 + p^2$, so

$$P(A|X \in \{0,2\}) = \frac{p^2}{(1-p)^2 + p^2}.$$

If $p = 0.6$ this is $\frac{0.36}{0.52} = 0.692\ldots$.
Similarly, if they disagree,

$$P(A|X=1) = \frac{p(1-p)}{2p(1-p)} = \frac{1}{2}.$$

(Logical, since if they disagree, their answer is determined by a coin toss.)

Note. As we see, this does not increase the probability of being right, but if the team agrees increases their confidence in the answer. If there are more than 2 team members, using a majority will increase the probability of being right.

Problem 6. Credit card transactions can be legitimate or fraudulent, and the proportion of fraudulent transactions is assumed to be one per thousand. Prior to approval, credit card transactions are tested and classified as legitimate or fraudulent. The test used classifies 99.5% of legitimate transactions as legitimate, and classifies 99% of fraudulent transactions as fraudulent.

(a) Determine the probability that a transaction classified as fraudulent is in fact fraudulent.

(b) The transactions all originate in regions A and B. The proportions of fraudulent transactions in these regions are assumed to be respectively 1/2000 (in A) and 1/500 (in B). What fraction of transactions are in region A?

Solution.

(a) Let F denote the event that a transaction is fraudulent, T the event that the test says it is fraudulent. Then by Bayes':

$$P(F|T) = \frac{P(T|F)P(F)}{P(T|F)P(F) + P(T|F^c)P(F^c)}.$$

By assumption, $P(F) = .001$, $P(T|F) = .99$, and $P(T|F^c) = .005$, so

$$P(F|T) = \frac{(.99)(.001)}{(.99)(.001) + (.005)(.999)}.$$

(b) Let A denote the event that a transaction originates in region A, and let B denote that it originates in region B. Then $P(F|A) = .0005$ and $P(F|B) = .002$, so

$$P(F) = P(F|A)P(A) + P(F|B)P(B) = (.0005)P(A) + (.002)(1 - P(A)).$$

However, we are given $P(F) = 0.001$. Solving gives $P(F) = \frac{2}{3}$.

Problem 7. In Python do the following:

(a) Sample 100,000 geometric random variables with parameter $p = 0.01$ and create a histogram of the resulting values, with buckets for each of the values 1 to 1000.

(b) Create a separate plot of the probability mass function of the geometric random variable, over the integers 1 to 1000. Briefly describe how this plot compares to the histogram from part (a).

(c) Finally, plot the function $f(t) = e^{-t}$ for t between 0 and 10. Compare this plot to the two plots above[1]

Submit your code, plots, and your written answers to the questions in (b) and (c).

[1] the relation between the plots in (b) and (c) will be explored next week.