ECS 132 Spring 2024: Midterm 1

Instructions –	There are 6 questions on the exam in total

1. Don't panic. There are no trick questions. Explain your intent if you cannot directly answer the question to enable partial credit.

1 Problem – Counting and sampling

Provide your answers to this problem as mathematical expressions which can be in terms of factorials or powers of numbers. Suppose that a lottery consists of picking 7 winning numbers from the set of integers 1-100. How many different lottery outcomes are possible if:

1. The order the numbers are drawn in matters and each number 1-100 can appear only once at most.

   $$P(100, 7) = \frac{100!}{93!}$$

2. The order the numbers are drawn in does not matter and each number 1-100 can appear only once at most.

   $$C(100, 7) = \frac{100!}{93!7!}$$

3. The order the numbers are drawn in matters and each number 1-100 can appear repeatedly, an unlimited number of times.

   $$(100)^7$$

4. The order the numbers are drawn in does not matter and each number 1-100 can appear repeatedly, an unlimited number of times.

   $$\binom{n+k-1}{k} = \binom{106}{7} = \frac{106!}{7!99!}$$

2 Problem – Counting applications

Provide your answers to this problem as mathematical expressions.

1. *Birthday paradox:* You encounter 5 people standing in a line and go through in order and ask each one their birthday. What is the probability that the first four people have distinct birthdays but the 5th person's birthday is the same as the first person's?

   $$P = \frac{365 \cdot 364 \cdot 363 \cdot 362 \cdot 1}{(365)^5} = \frac{364 \cdot 363 \cdot 362}{(365)^4}$$

2. *Poker hands:* You are playing a standard game of poker with a standard deck of 52 cards. You are dealt 5 cards at random, and the order does not matter. What is the probability that you receive 4 Aces **or** that you receive 4 Jacks?

$$P(4A \cup 4J) = P(4A) + P(4J) \text{ events are mutually exclusive}$$

$$\text{Consider the individual events}: P(4A) = P(4J) = \frac{\binom{4}{4}\binom{52-4}{1}}{\binom{52}{5}} = \frac{\binom{48}{1}}{\binom{52}{5}} = \frac{48 \cdot 5! \cdot 47!}{52!}$$

$$P(4A \cup 4J) = P(4A) + P(4J) = \frac{2\binom{48}{1}}{\binom{52}{5}} = \frac{96 \cdot 5! \cdot 47!}{52!}$$

3 Problem – Simple conditional probability

You are playing a game where there are two slots (a left slot and a right slot). Each slot contains a ball that is colored red or green with equal probability.

1. Both slots are closed, hiding their contents. What is the probability that both balls are red?

$$S = \{\{R, R\}, \{R, G\}, \{G, R\}, \{G, G\}\} \text{ so } P(\{R, R\}) = \frac{1}{4}$$

2. The left slot is now opened revealing a green ball. What is the probability that the right slot contains a red ball?

$$\text{Reduced sample space } S' = \{\{G, R\}, \{G, G\}\} \text{ so } P(R) = \frac{1}{2}$$

4 Problem – Bayes' Rule

A hacker has compromised your computer and is using it to mine bitcoin. Consider that the hacker wants to remain undetected so they do not send packets to your CPU too frequently especially because the bitcoin mining packets tend to be longer than average. Define the two events:

- B – the packet is a bitcoin mining attempt.
- T – the packet is twice as long as the average packet.

The hacker makes sure that only one-in-ten packets sent is a bitcoin request meaning that with probability 0.1 a random packet is a bitcoin mining attempt (otherwise it is a legitimate packet). It is also the case that the probability of a packet being twice as long as average is 0.5. Finally, the probability of a packet being twice as long as average is 0.8 **given that** the packet is known to be a bitcoin mining attempt.

1. What is the probability of the packet being a bitcoin mining attempt **given that** it is twice as long as average? Give a mathematical expression using the events defined above (and their complements if necessary) and then the final numerical answer.

   Givens and implications:

$$
\begin{aligned}
P(B) &= 0.1 &\rightarrow\quad & P(B^{\complement}) = 0.9 \\
P(T) &= 0.5 &\rightarrow\quad & P(T^{\complement}) = 0.5 \\
P(T|B) &= 0.8 &\rightarrow\quad & P(T^{\complement}|B) = 0.2
\end{aligned}
$$

   Solve for $P(B|T)$:

$$P(B|T) = \frac{P(B)P(T|B)}{P(T)} = \frac{(0.1)(0.8)}{0.5} = 0.16$$

2. What is the probability of the packet being legitimate **given that** it is twice as long as average? Give a mathematical expression using the events defined above (and their complements if necessary) and then the final numerical answer.

   Solve for $P(B^{\complement}|T)$:

$$
\begin{aligned}
P(B^{\complement}|T) &= 1 - P(B|T) \quad \text{(solved for in 1)} \\
&= 1 - 0.16 = 0.84
\end{aligned}
$$

3. What is the probability of the packet being a bitcoin mining attempt **and** being twice as long as average? Give a mathematical expression using the events defined above (and their complements if necessary) and then the final numerical answer.

   Solve for $P(B \cap T)$:

$$
\begin{aligned}
P(B \cap T) &= P(B)P(T|B) = P(T)P(B|T) \\
&= (0.5)(0.16) = 0.08
\end{aligned}
$$

5 Problem – Conditional probability with false positives/negatives

You are training a machine learning algorithm to classify whether a given image is a cat or not a cat. You input a series of images to the algorithm such that with probability 0.55 the image is a cat. Define the two events:

- C – the input image is a cat.
- A – the algorithm classifies the image as a cat.

The algorithm is not perfectly accurate: (1) With probability 0.95 the algorithm classifies the image as a cat given that the input image was a cat; (2) With probability 0.1 the algorithm classifies the image as a cat given that the input image was not a cat.

1. What is $P(A)$, the probability that the algorithm classifies an image as a cat? Give a mathematical expression using the events defined above (and their complements if necessary) and the final numerical answer.

   Givens and implications:

$$\begin{aligned} P(C) &= 0.55 \quad \to \quad P(C^{\complement}) = 0.45 \\ P(A|C) &= 0.95 \quad \to \quad P(A^{\complement}|C) = 0.05 \\ P(A|C^{\complement}) &= 0.1 \quad \to \quad P(A^{\complement}|C^{\complement}) = 0.9 \end{aligned}$$

   Solve for

$$P(A) = P(A|C)P(C) + P(A|C^{\complement})P(C^{\complement})$$
$$= (0.95)(0.55) + (0.1)(0.45) = 0.5675$$

2. What is $P(A^{\complement})$, the probability that the algorithm classifies the image as not a cat? You can give this as a simple number.

$$P(A^{\complement}) = 1 - P(A) = 0.4325$$

3. What is the probability the image is a cat, given that the algorithm has classified it as not a cat? Give a mathematical expression using the events defined above (and their complements if necessary) and the final numerical answer.

   Solve for $P(C|A^{\complement})$:

$$P(C|A^{\complement}) = \frac{P(C)P(A^{\complement}|C)}{P(A^{\complement})}$$

   Note $P(A^{\complement}|C) = 1 = P(A|C) = 1 - 0.95 = 0.05$

$$P(C|A^{\complement}) = \frac{(0.55)(0.05)}{0.4325} = 0.0636$$

6 Problem – Distributions of discrete random variables

Consider a scenario we have encountered several times by now. There are two sensors who can be active or inactive in a given time slot. Each sensor has the probability of being active in a time slot of q=0.4. The sensors are independent of one another.

1. Let the random variable $X = 1$ if both sensors are active in one time slot. What is $P(X = 1)$?

$$p_1 = P(X = 1) = q^2 = 0.16$$

2. To what parametric family of distributions does the distribution of X above belong (provide this in terms of $X \sim$ distribution_name(parameters)).

$$X \sim Bern(0.16)$$

Now consider that there is a sequence of 100 independent time slots, each one defined as above. Use the notation p_1 to denote $p_1 = P(X = 1)$ for each individual slot.

3. Let the random variable X denote the number of time slots in which both sensors are active. What is $P(X = k)$? Express this as a mathematical equation in terms of p_1 and other necessary variables.

$$P(X = k) = \binom{n}{k} p_1^k (1 - p_1)^{n-k} = \binom{100}{k} p_1^k (1 - p_1)^{100-k}$$

4. To what parametric family of distributions does the distribution of X above belong (provide this in terms of $X \sim$ distribution_name(parameters)).

$$X \sim Binom(100, p_1)$$

5. What R function would return the probability that $X = k$ for the distribution you specified in part(4)? Write out the function with the proper arguments.

   dbinom(k,100,p_1)

6. What is the R function that would return the probability that $X \leq k$ for the distribution you specified in part(4)? Write out the function with the proper arguments.

   pbinom(k,100,p_1)

7. How would you use the R function above to calculate $P(2 \leq X \leq 10)$ for the distribution you specified in part(4)? Write this out using the R functions.

   pbinom(10,100,p_1) - pbinom(1,100,p_1)