

625.603 - Statistical Methods and Data Analysis
Homework #1

Due on September 7, 2018 at 11:55pm

Professor Barry Bodt Fall 2018

Cameron McIntyre

Problem 1

2.2.4: Suppose that two cards are dealt from a standard 52-card poker deck. Let A be the event that the sum of the two cards is 8 (assume that aces have a numerical value of 1). How many outcomes are in A?

Answer: We use the counting Principle:

There are $\binom{6}{1}$ ways of choosing the number of the first card. This corresponds to the cards A,2,3,5,6,7.
There are $\binom{4}{1}$ Ways to choose the suit of the second card.

For the special case of the number 4, There are $\binom{4}{1}$ ways to choose the first 4, only $\binom{3}{1}$ to choose the suit of the second 4.

This makes the total number of ways to choose 2 cards that add up to 8:

$$\binom{6}{1} \cdot \binom{4}{1} + \binom{4}{1} \cdot \binom{3}{1} = 108$$

Problem 2

2.2.11: A woman has her purse snatched by two teenagers. She is subsequently shown a police lineup consisting of five suspects, including the two perpetrators. What is the sample space associated with the experiment? Woman picks two suspects out of lineup?? Which outcomes are in the event A: She makes at least one incorrect identification?

Answer:

Let B_i $1 \leq i \leq 3$ Represent the bystanders and P_j $1 \leq j \leq 2$ represent the perpetrators.

We enumerate the sample space **S**:

$$\mathbf{S} = \{(B_1, B_2), (B_1, B_3), (B_2, B_3), (B_1, P_1), (B_1, P_2), (B_2, P_1), (B_2, P_2), (B_3, P_1), (B_3, P_2), (P_1, P_2)\}$$

The outcomes associate with the event **A** = At least 1 incorrect perpetrator:

$$\mathbf{A} = \{(B_1, B_2), (B_1, B_3), (B_2, B_3), (B_1, P_1), (B_1, P_2), (B_2, P_1), (B_2, P_2), (B_3, P_1), (B_3, P_2)\}$$

Problem 3

2.2.28: Let events A and B and sample space S be defined as the following intervals:

$$\mathbf{S} = \{x : 0 \leq x \leq 10\}$$

$$\mathbf{A} = \{x : 0 < x < 5\}$$

$$\mathbf{B} = \{x : 3 \leq x \leq 7\}$$

Characterize the following events:

(a) $A^C = \{x : 0, 5 \leq x \leq 10\}$

(b) $A \cap B = \{x : 3 \leq x < 5\}$

(c) $A \cup B = \{x : 0 < x \leq 7\}$

(d) $A \cap B^C = \{x : 0 < x < 3\}$

(e) $A^C \cup B = \{x : 3 \leq x \leq 10\}$

(f) $A^C \cap B^C = \{x : 0, 7 < x \leq 10\}$

Problem 4

2.2.40: For two events A and B defined on a sample space **S**:

$$N(A \cap B^C) = 15$$

$$N(A^C \cap B) = 50$$

and

$$N(A \cap B) = 2$$

Given that $N(S) = 120$, how many outcomes belong to neither A nor B?

Answer:

We approach this problem by partitioning the sample space into mutually exclusive sets.

$$S = (A \cup B) \cup (A \cup B)^C \quad \textbf{(1)}$$

$$\text{By Definition } (A \cup B) \cap (A \cup B)^C = \emptyset$$

Taking the left hand side of the above union, we can break it up as follows:

$$(A \cup B) = (A \cap B^C) \cup (A^C \cap B) \cup (A \cap B)$$

This means we can represent **(1)** as:

$$S = (A \cup B) \cup (A \cup B)^C = [(A \cap B^C) \cup (A^C \cap B) \cup (A \cap B)] \cup (A \cup B)^C$$

Applying **N** as an operator, and knowing that the sets are disjoint we get:

$$N(S) = N([(A \cap B^C) \cup (A^C \cap B) \cup (A \cap B)] \cup (A \cup B)^C)$$

$$N(S) = N((A \cap B^C)) + N((A^C \cap B)) + N((A \cap B)) + N((A \cup B)^C)$$

$$120 = 15 + 50 + 2 + N((A \cup B)^C)$$

$$N((A \cup B)^C) = 120 - 15 - 50 - 2$$

$$N((A \cup B)^C) = 53$$

Problem 5

2.3.2: 2.3.2. Let A and B be any two events defined on S. Suppose that $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.1$. What is the probability that A or B but not both occur?

Answer:

The goal is to find $P((A \cup B) \setminus (A \cap B)) = P(A \cap B^C) + P(A^C \cap B)$ since they are disjoint sets.

Consider the following equations

$$P(A) = P(A \cap B) + P(A \cap B^C) \quad \textbf{(1)}$$

$$P(B) = P(A \cap B) + P(A^C \cap B) \quad \textbf{(2)}$$

From **(1)**

$$.4 = .1 + P(A \cap B^C)$$

$$.3 = P(A \cap B^C) \quad \textbf{(3)}$$

From **(2)**

$$.5 = .1 + P(A^C \cap B)$$

$$.4 = P(A^C \cap B) \quad \textbf{(4)}$$

$$\text{Back to this statement } P((A \cup B) \setminus (A \cap B)) = P(A \cap B^C) + P(A^C \cap B)$$

We substitute using **(4)** **(3)** to get:

$$P((A \cup B) \setminus (A \cap B)) = .3 + .4 = .7$$

Problem 6

2.3.12: Events A_1 and A_2 are such that $A_1 \cup A_2 = S$ and $A_1 \cap A_2 = \emptyset$.

Find p_2 if $P(A_1) = p_1$, $P(A_2) = p_2$, and $3p_1 - p_2 = \frac{1}{2}$.

Answer:

Using the information given $\mathbf{S} = A \cup B$

Then

$$P(\mathbf{S}) = 1 \Leftrightarrow 1 = P(A \cup B) \Leftrightarrow 1 = P(A) + P(B) - P(A \cap B)$$

We know $A \cap B = \emptyset$ from the definitions in the question. So, $P(A \cap B) = 0$.

Therefore,

$$1 = P(A \cup B) \Leftrightarrow 1 = P(A) + P(B) - P(A \cap B) \Leftrightarrow 1 = P(A) + P(B) \Leftrightarrow 1 = p_1 + p_2$$

Now we have 2 equations and 2 unknowns. It is an exercise of substitution to arrive at a solution.

$$(1) \quad 1 = p_1 + p_2$$

$$(2) \quad 3p_1 - p_2 = \frac{1}{2}$$

From (1), Let $p_1 = 1 - p_2$

Substitute into (2)

$$3(1 - p_2) - p_2 = \frac{1}{2}$$

$$3 - 4 \cdot p_2 = \frac{1}{2}$$

$$p_2^* = \frac{5}{8}$$

Substituting p_2^* in into (1)

$$1 - \frac{5}{8} = p_1^*$$

We reach our answers: $P(A_1) = p_1 = \frac{3}{8}$ and $P(A_2) = p_2 = \frac{5}{8}$

Problem 7

2.3.16: Two dice are tossed. Assume that each possible outcome has a $\frac{1}{36}$ probability. Let A be the event that the sum of the faces showing is 6, and let B be the event that the face showing on one die is twice the face showing on the other. Calculate $P(A \cap B^C)$.

Answer:

There are $\binom{5}{1}$ ways to roll the first die so that they sum to 6 and 1 way to roll the second die to make that sum to 6. The problem is that this includes the event where $d_1 = 2 \cdot d_2$ or $d_2 = 2 \cdot d_1$.

To address this we must remove the events where the number on one die is twice that of the number displayed on the other die. This occurs in the sample space twice with the pairs (2, 4) and (4, 2).

This leaves the total number of events in $A \cap B^C$ as $\binom{5}{1} \cdot 1 - 2 = 3$.

Therefore the probability $P(A \cap B^C) = \frac{3}{36} = \frac{1}{12}$

Problem 8

2.4.7: An urn contains one red chip and one white chip. One chip is drawn at random. If the chip selected is red, that chip together with two additional red chips are put back into the urn. If a white chip is drawn, the chip is re- turned to the urn. Then a second chip is drawn. What is the probability that both selections are red?

Answer: Let D_1 be the first draw and D_2 be the second draw.

$$P(D_2 = R | D_1 = R) = P(D_2 = R \text{ and } D_1 = R) \cdot P(D_1 = R) = \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8}$$

Problem 9

2.4.10: 2.4.10. Suppose events A and B are such that $P(A \cap B) = 0.1$ and $P((A \cup B)^C) = 0.3$. If $P(A) = 0.2$, what does $P[(A \cap B) | (A \cup B)^C]$ equal? (Hint: Draw the Venn diagram.)

Answer: By inspection this set is Null. Therefore $P[(A \cap B)|(A \cup B)^C] = 0$

We can see this using DeMorgans law on the event we are conditioning on (to the right of the |).

$(A \cap B)^C = (A^C \cap B^C)$ By Demorgans Law.

And $(A \cap B) \cap (A^C \cap B^C) = \emptyset$.

So for some x it is impossible for $x \in (A \cap B)$ and $x \in (A^C \cap B^C)$. So $(A \cap B) \cap (A^C \cap B^C) = \emptyset$

And $P(\emptyset) = 0$ By definition.

I hope an argument using demorgans law is ok instead of a venn diagram.

Problem 10

2.4.16: Given that $P(A) + P(B) = 0.9$, $P(A|B) = 0.5$, and $P(B|A) = 0.4$, find $P(A)$.

Answer:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = .4$$

$$P(A|B) = \frac{P(B \cap A)}{P(B)} = .5$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \leftrightarrow P(A \cap B) = .4P(A)$$

$$P(A|B) = \frac{P(B \cap A)}{P(B)} \leftrightarrow P(A \cap B) = .5P(B)$$

Using the above results we can equate $P(A \cap B)$:

$$.4P(A) = .5P(B) \leftrightarrow .4P(A) - .5P(B) = 0$$

And we use $P(B) = .9 - P(A)$ to substitute into the equation above.

$$.4P(A) - .5 \cdot (.9 - P(A)) = 0 \leftrightarrow .4P(A) - .45 + .5P(A) = 0$$

Simplifying, we arrive at $P(A) = \frac{.45}{.9} = \frac{1}{2}$

Problem 11

2.4.23: Given that $P(A) + P(B) = 0.9$, $P(A|B) = 0.5$, and $P(B|A) = 0.4$, find $P(A)$.

Answer: Your favourite college football team has had a good season so far but they need to win at least two of their last four games to qualify for a New Year's Day bowl bid. Oddsmakers estimate the team's probabilities of winning each of their last four games to be 0.60, 0.50, 0.40, and 0.70, respectively.

(a) What are the chances that you will get to watch your team play on Jan. 1?

(b) Is the probability that your team wins all four games given that they have won at least three games equal to the probability that they win the fourth game? Explain.

(c) Is the probability that your team wins all four games given that they won the first three games equal to the probability that they win the fourth game?

Answer:

A) Let G_1, G_2, G_3, G_4 correspond to winning the 1st, 2nd 3rd and 4th game, respectively. Then

$P(\text{Wins at least 2 Games}) = 1 - P(\text{wins at most one game})$

$P(\text{Wins at least 2 Games}) = 1 - (1-.6)(1-.5)(1-.4)(1-.7) + .6(1-.5)(1-.4)(1-.7) + (1-.6)(.5)(1-.4)(1-.7) + (1-.6)(1-.5)4(1-.7) + (1-.6)(1-.5)(1-.4).7$

$P(\text{Wins at least 2 Games}) = .766$

B and C) Is the probability that your team wins all four games given that they have won at least three games equal to the probability that they win the fourth game? Explain.

No. The probabilities are not equal. See Below:

$P(\text{Win at least 3}) = (.6)(.4)(.5)(.7) + (.6)(.4)(.5)(1-.7) + (.6)(.4)(1-.5)(.7) + (.6)(1-.4)(.5)(.7) + (1-.6)(.4)(.5)(.7)$

$P(\text{Win 4th} | \text{Won 3 already}) = .386$

$P(\text{Win all 4}) = (.6)(.4)(.5)(.7) = .08$

$P(\text{Win 4th} | \text{Won 3 already}) = .08 / .386 = .2176$

We observe that the probability of winning the fourth game conditional on winning the first three is higher. This is because we have limited the sample space to the set of events where the team has won the first 3 games. Because we have made the "denominator" smaller the probability increases.

Problem 12

2.4.46: Brett and Margo have each thought about murdering their rich Uncle Basil in hopes of claiming their inheritance a bit early. Hoping to take advantage of Basil's predilection for immoderate desserts, Brett has put rat poison into the cherries flambe; Margo, unaware of Brett's activities, has laced the chocolate mousse with cyanide. Given the amounts likely to be eaten, the probability of the rat poison being fatal is 0.60; the cyanide, 0.90. Based on other dinners where Basil was presented with the same dessert options, we can assume that he has a 50 chance of asking for the cherries flambe, a 40 chance of ordering the chocolate mousse, and a 10 chance of skipping dessert altogether. No sooner are the dishes cleared away than Basil drops dead. In the absence of any other evidence, who should be considered the prime suspect?

Answer:

We label our events Br as Brad being the killer and M margo being the killer. We need to find who has a greater chance of being the killer. We define the event as $P(B|Basil \text{ is Dead})$.

$$P(Br|BasilDead) = \frac{P(BasilDead \cap Br)}{P(BasilDead)} = \frac{P(BasilDead \cap Br)}{P(BasilDead|Br) + P(BasilDead|M)}$$

\leftrightarrow

$$P(Br|BasilDead) = \frac{.6 \cdot .5}{.6 \cdot .5 + .4 \cdot .9}$$

\leftrightarrow

$$P(Br|BasilDead) = \frac{.6 \cdot .5}{.6 \cdot .5 + .4 \cdot .9} = \frac{.3}{.66} = .45$$

$$P(M) = .55$$

The more likely suspect is Margot.

Problem 13

2.5.5: Dana and Cathy are playing tennis. The probability that Dana wins at least one out of two games is 0.3. What is the probability that Dana wins at least one out of the 4?

Answer: The event "Dana wins at least one out of two games" will be denoted as D , and we know $P(D) = .3$. We are trying to find $P(D_1 \cup D_2)$ Where 1 and 2 represent a set of 2 games. This event can be expressed as $P(D_1 \cup D_2) = 1 - P(D_1^C \cap D_2^C)$.

This gives us probability:

$$P(D_1 \cup D_2) = 1 - P(D_1^C \cap D_2^C) = 1 - P(D_1^C) \cdot P(D_2^C) = 1 - .7 \cdot .7 = .51$$

Problem 14

2.5.16: On her way to work, a commuter encounters four traffic signals. Assume that the distance between each of the four is sufficiently great that her probability of getting a green light at any intersection is independent of what happened at any previous intersection. The first two lights are green for forty seconds of each minute; the last two, for thirty seconds of each minute. What is the probability that the commuter has to stop at least three times?

Answer:

Let L_i^C Represent a green light at the i th light and L_i represent stopping at the i th light. Then:

$$P(\text{Stop at least 3 times}) = P(L_1)L_2P(L_3)P(L_4) + P(L_1^C)L_2P(L_3)P(L_4) + P(L_1)L_2P(L_3^C)P(L_4) + P(L_1)L_2P(L_3)P(L_4^C)$$

$$P(\text{Stop at least 3 times}) = \left(\frac{1}{3} \frac{1}{3} \frac{1}{2} \frac{1}{2}\right) + \left(\frac{2}{3} \frac{1}{3} \frac{1}{2} \frac{1}{2}\right) + \left(\frac{1}{3} \frac{2}{3} \frac{1}{2} \frac{1}{2}\right) + \left(\frac{1}{3} \frac{1}{3} \frac{1}{2} \frac{1}{2}\right)$$

$$P(\text{Stop at least 3 times}) = \left(\frac{1}{36} + \frac{2}{36} + \frac{2}{36} + \frac{1}{36}\right)$$

$$P(\text{Stop at least 3 times}) = \frac{7}{36}$$

Problem 15

2.5.32:

What is the smallest number of switches wired in parallel that will give a probability of at least 0.98 that a circuit will be completed? Assume that each switch operates independently and will function properly 60 of the time.

Answer: A parallel circuit fails if all of the switches fail. We are trying to find the number of switches needed to make the circuit 98 percent fail-proof.

Define a the circuit working as C and failing as C^C . Define the switch working as W and failing as W^C . Then the probability of the circuit working can be expressed as:

$$P(C) = 1 - P(C^C) = 1 - P(\cap_i^n W^C)$$

Where n is the number of switches.

To find n such that $P(C) \geq .98$, we use some algebraic manipulation and independence to isolate for n .

$$1 - .98 = \prod_i^n P(W^C) \leftrightarrow .02 = .4^n$$

$$\frac{\log .02}{\log .4} = n \leftrightarrow n = 4.26$$

Therefore we use at least 5 switches to make the circuit 98 percent fail proof.

Problem 16

2.5.32:

Monte Carlo exercise: John has integers 1:10. He randomly draws 5 without replacement and reasons that he could estimate the 80th percentile of his 10 integers, the value 8, by taking the 2nd largest sampled value; that is the 4th value in order from smallest to largest. (a) Applying this approach repetitively, what proportion of the time will he accurately estimate the value 8? (b) Underestimate? (c) Overestimate? The answer is easily accessible using combinations in the next module, but until then, simulation is the preferred approach.

Answer: Running the simulation (code Below In Python). John underestimates about 50.128 percent of the time, Correctly estimates 27.57 percent of the time and overestimates about 22.302.

Using Python Code:

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
```

Created on Sat Sep 1 13:32:43 2018

@author: cpmcintyre

"""

```
import numpy as np
from numpy.random import choice
from numpy import sort
samplechoice=list(range(1,11))
Results = []
for i in range(100000):
    k=sort(choice(samplechoice,5, replace=False))
    Results.append(k[3])
counts=[0,0,0]
for i in Results:
    if i < 8:
        counts[0]+=1.0
    elif i > 8:
        counts[2]+=1.0
    else:
        counts[1]+=1.0

print((counts[0]/10000, counts[1]/10000, counts[2]/100000))
```