Stats HW1 by Sejal Nimkar

2024-09-09

Homework 1 Questions and Answers

1. Suppose that $P(A) = 0.7$, $P(B) = 0.5$, and $P(A^c \cap B^c) = 0.15$

a. Is it possible for A and B to be disjoint events? Why or why not?

Ans: If two events, A and B are disjoint, then the sum of their probabilities should equal the probability of their union, that is $P(A \cup B)$.

To find out $P(A \cup B)$, let's use the Demorgan's law :

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

$\therefore P((A \cup B)^C) = 0.15$

According to the theorem of probability:

$1 - P(E) = P(E^c)$

Hence our equation becomes :

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

$\therefore P(A \cup B) = 1 - 0.15 = 0.85$

To check if the events are disjoint,let's calculate the sum of the probabilities of $P(A)$ and $P(B)$:

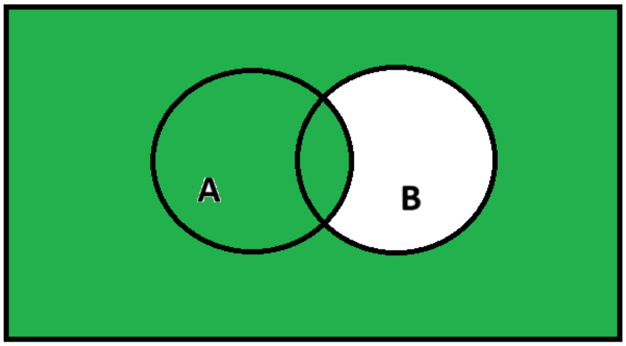
$P(A) + P(B) = 0.7 + 0.5 = 1.2$

And the probability of their union is 0.85 as calculated above.

Hence the events cannot be disjoint.

b. What is the probability of $A \cup B^c$?

Ans: We shall derive the answer to this question using venn diagram.



The above diagram shows the area (shaded region) of $A \cup B^c$. The probability of the non-shaded region can be given as:

$P(B) - P(A \cap B)$

To calculate probability of $A \cap B$, we will use the following equation:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\therefore P(A \cap B) = P(A) + P(B) - P(A \cup B)$

$\therefore P(A \cap B) = 0.7 + 0.5 - 0.85 = 0.35$

Thus the probability of the non-shaded region is

$P(B) - P(A \cap B)$ = 0.5 - 0.35 = 0.15

Now we can determine the probability of the shaded region by subtracting it from 1:

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

$\therefore P(A \cup B^c) = 1 - 0.15 = 0.85$

The probability of $A \cup B^c$ is 0.85.

c. Is it possible for A and B to be independent events? Why or why not?

Ans: If A and B are independent events, then probability of A does not change by the occurence and the following equation should hold true:

$P(A \cap B) = P(A).P(B)$

We know:

$P(A \cap B) = 0.35$

$P(A).P(B) = 0.5 * 0.7 = 0.35$

Thus LHS =RHS.

Therefore the events A and B are independent.

d. What is the conditional probability of A given B^c?

Ans: The conditional probability of A given B^c is represented as

$P(A|B^c)$

As we have established that the two events A and B are independent events. Thus the following equations will hold true:

$P(A|B) = P(A|B^c)$

and

$P(A|B) = P(A)$

$\therefore P(A|B^c) = P(A) = 0.7$

2. ISI Section 3.7 Exercise 8 (ISI is the acronym of our textbook), but use instead $P(+|D) = 0.93$, $P(-|D^c) = 0.98$ and $P(D) = 0.02$.

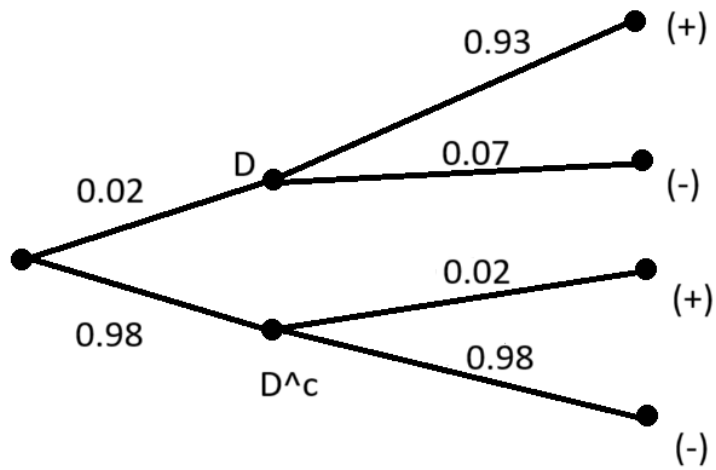
Pre-eclampsia is a disorder of pregnancy in which the interaction between mother and placenta is disrupted, causing a hypertensive response. Various studies have investigated the accuracy of using blood pressure to predict pre-eclampsia.

Consider a test that predicts pre-eclampsia when mean arterial pressure is at least 90 mm Hg. Assume that the sensitivity of this test is $P(+|D) = 0.93$ and that the specificity of this test is $P(-|D^c) = 0.98$.

a. What is the conditional probability of a false positive test result?

Ans: Let's draw a decision tree to represent the above problem.

```r
knitr::include_graphics("Decision2.png")
```



D denotes the women have pre-eclampsia while D^c denotes the pregnant women who do not have pre-eclampsia.

$+$ denotes that the test resulted as positive for pre-eclampsia and $-$ denotes that the test resulted negative.

Thus using the decision tree, we can determine that the probability of a false positive test i.e.

$P(+|D^c) = 0.02$

b. What is the conditional probability of a false negative test result?

Ans: Using the decision tree, we can see that probability of a false negative test i.e.

$P(-|D) = 0.07$

c. Suppose that a pregnant woman is selected at random from a population in which the prevalence of pre-eclampsia is $P(D) = 0.02$. Construct a tree diagram that describes this experiment.

Ans: Given above

d. What is the probability that the above test will diagnose the selected woman as having pre-eclampsia?

Ans: We are told to determine $P(+)$. We can calculate $P(+)$ using the below equation:

$P(+) = P(D \cap +) + P(D^c \cap +)$

Using the equation of conditional probability, we can find the values of $P(+ \cap D)$ and $P(+ \cap D^c)$

$P(D \cap +) = P(D) \times P(+|D)$

$\therefore P(D \cap +) = 0.02 \times 0.93 = 0.0186$

Similarly,

$P(D^c \cap +) = P(D) \times P(+|D^c)$

$\therefore P(D^c \cap +) = 0.02 \times 0.02 = 0.0004$

Thus, the value of $P(+)$ is:

$P(+) = P(D \cap +) + P(D^c \cap +)$

$\therefore P(+) = 0.0186 + 0.0004$

$\therefore P(+) = 0.0190$

Thus the probability that the test will diagnose the selected woman as having pre-eclampsia is 0.019.

e. Suppose that the above test does diagnose the selected woman as having pre-eclampsia. What then is the probability that this woman actually does have pre-eclampsia?

Ans: We have to find the value of $P(D|+)$.

$P(D|+) = \frac{P(D \cap +)}{P(+)}$

In the previous question, we found out the values of $P(+)$ and $P(D \cap +)$.

Therefore substituting those values, we get:

$P(D|+) = \frac{0.0186}{0.0190}$

$\therefore P(D|+) = 0.9789$

Thus the probability of a woman getting diagnosed have pre-eclampsia actually has pre-eclampsia is 0.97.

3. The experiment is to toss a fair coin eight times. Show your work to receive full credit.

a. What is $P(A)$ when $A = \{$Exactly five of the coins show heads$\}$

Ans: Calculating $P(A)$ using binomial distribution:

$P_x = \binom{n}{x} p^x (1-p)^{n-x}$

Here, we have:

Total number of trials i.e. coin tosses: $n = 8$

Number of coins that must show heads: $x = 5$

Probability of success that head will show on a fair coin : $p = \frac{1}{2}$ or 0.5

Probability of faiure that a head will show on the fair coin: (1-p = 1- 0.5 = 0.5)

Therefore:

$P(A) = \binom{8}{5}(0.5)^5(0.5)^{8-5}$

```r
dbinom(5,size=8,prob=0.5)
```

```
## [1] 0.21875
```

$\therefore P(A) = 0.21875$

b. What is $P(A|B)$ when $B = \{$There are more heads than tails$\}$?

Ans: Probability of event A given that B has already occurred i.e. $P(A|B)$ can be given by using the following equation:

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

Event A is a subset of event B since event B consists of outcomes where there are more heads than tails while A is the event where there are exactly 5 heads, i.e. the number of heads are greater than tails.

$\therefore P(A \cap B) = P(A)$

$\therefore P(A \cap B) = 0.21875$

while $P(B)$ can be calculated subtracting the cumulative probability distribution of outcomes where number of heads range from 0 to 4 from one.

Let X be the random variable where $X = \{assignthenumberofheadstoeveryoutcome\}$.

Th equation for cdf for probability for discrete random variable X is given by

$F(X) = P(X \leq x) = \sum_{t \leq x} P(X = t)$

Therefore $P(B)$ can be given as :

$P(B) = 1 - F(X \leq 4)$

We can use R to calculate the cdf of random variable X.

```r
pbinom(4,size=8,prob=0.5)
```

```
## [1] 0.6367187
```

$\therefore P(B) = 1 - 0.6367187$'

```r
1- 0.6367187
```

```
## [1] 0.3632813
```

$P(B) = 0.3632813$

Thus $P(A|B)$ can be calculated as:

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

$i.eP(A|B) = \frac{0.21875}{0.36329}$

$\therefore P(A|B) = 0.5996$

c. Identify an event E that produces a probability between 0.4 and 0.7. Do not repeat the event your classmates have identified.

Ans: Let event E be :

$E = \{\text{Number of Heads shown on the coins are at least 4}\}$

Probability of event E can be calculated by adding singular probabilities of outcomes that show zero heads, one head, two heads, three heads, and then subtracting those from one.

Thus we are calculating cumulative probabilities which can be given as :

$F(k; n, p) = P(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$

The function used in R to calculate cumulative probabilities for binomial distribution is pbinom()

```
pbinom(3,size=8,prob=0.5)
```

```
## [1] 0.3632813
```

Subtracting the probability from 1, we get:

```
1 - pbinom(3,size=8,prob=0.5)
```

```
## [1] 0.6367187
```

Thus event E has the probability of 0.6367 which lies between 0.4 and 0.7.

d. Find $A \cup E$

Ans: $A = \{HHHHHTTT, HTHTHTHH, TTTHHHHH, HHHTTTHH, HHTHTHHT, HTHHHTTH...\}$

For each coin, there are two possible outcomes, either Heads or Tails i.e. 2 outcomes.

The number of ways to get exactly 5 heads out of 8 coin tosses, we use the equation for combinations.

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Here n = 8, and k=5,

$\therefore \binom{8}{5} = \frac{8!}{5!(8-5)!}$

$\frac{8!}{5!3!} = \frac{8\times7\times6}{3\times2\times1} = 56$

Event A consists of 56 outcomes altogether.

Event E consists of outcomes that contain number of heads ranging from 4 to 8.

$E = \{HHHHHHHH, TTHTHHHH, THHHHHTT, HHHHHTTT, HHHHTHTH, THTHTHHH...\}$

To find the total number of outcomes in E, we sum the binomial coefficients for each possible number of heads.

$|E| = \binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \binom{8}{3} + \binom{8}{4}$

$\binom{8}{4} = 70,$

$\binom{8}{3} = 56,$

$\binom{8}{2} = 28,$

$\binom{8}{1} = 8,$

$\binom{8}{0} = 1$

Adding them together, we get 163 outcomes.

Thus $A \cup E$ will consist of $163 + 56 = 219$ outcomes.

$A \cup E = \{HHHHHTTT, HTHTHTHH, HHTHTHHT, HTHHHTTH, HHHHHHHH, TTHTHHHH...\}$

e. Can you identify an event F with a probability between 0.4 and 0.7 and disjoint from B? If Yes, find it. If not, explain why not.

Ans: Let F be an event such that:

$F = \{\text{Number of heads shown on the coins are at most 4}\}$

In this way, F is an event that is disjoint from B.

The probability of event F is:

```
pbinom(4,size=8,prob=0.5)
```

```
## [1] 0.6367187
```

$P(F) = 0.6367$

4. We'll continue working with the experiment in question 3: tossing a fair coin eight times. Show your work to receive full credit.

a.How many outcomes are there in the sample space? Show your work and write down one possible outcome.

Ans: For each coin toss, there are 2 possible outcomes: heads (H) or tails (T). To find the total number of possible outcomes for 8 coin tosses, we can use the principle of multiplication.

For the first coin, there can be two outcomes, either heads or tails.

Similarly, for the second coin there can be two outcomes.

Thus for each coin, there are two possible outcomes.

Hence total number of outcomes are $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2$

$\therefore$ total number of outcomes is 2^8

$2^8 = 256$

So, there are 256 possible outcomes for 8 coin tosses.

b. If the random variable X assigns the number of heads to each outcome, obtain $X(S)$, i.e., the range of X.

Ans: For 8 coin tosses, the maximum number of heads that can be obtained is 8 and the minimum number of heads that can be obtained is 0.

Thus range of X ranges from 0 to 8.

$X(S) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$

c. If the random variable Z is the largest between #Heads - #Tails OR #Tails - #Heads, what is $Z(S)$?

Ans: Z contains values of largest of either $\#H - \#T$ or $\#T - \#H$ i.e. it contains absolute values of #H-#T i.e. $|\#H - \#T|$.

1)For number of heads = 0, number of tails = 1:

$|\#H - \#T| = |0 - 8| = 8$

2)For number of heads = 1, number of tails = 7:

$|\#H - \#T| = |1 - 7| = 6$

3)For number of heads = 2, number of tails = 6:

$|\#H - \#T| = |2 - 6| = 4$

4)For number of heads = 3, number of tails = 5:

$|\#H - \#T| = |3 - 5| = 2$

5)For number of heads = 4, number of tails = 4:

$|\#H - \#T| = |4 - 4| = 0$

6)For number of heads = 5, number of tails = 3:

$|\#H - \#T| = |5 - 3| = 2$

7)For number of heads = 6, number of tails = 2:

$|\#H - \#T| = |6 - 2| = 4$

8)For number of heads = 7, number of tails = 1:

$|\#H - \#T| = |7 - 1| = 6$

9)For number of heads = 8, number of tails = 0:

$|\#H - \#T| = |8 - 0| = 8$

Thus $Z(S) = \{0, 2, 4, 6, 8\}$

d. Let's define the set $\{s \in S : X(s) \leq x\}$ .Identify the outcomes in this set when $x = 3$; you do not need to write them all down; simply describe what they are.

Ans: X is the random variable which assigns the number of heads to each outcome.

The set $\{s \in S : X(s) \leq x\}$ for $x = 3$ will contain outcomes for which $X(s) \leq 3$ i.e. the number of heads is less than 3.

Thus the set will contain all outcomes for which number of heads range from 0 to 3:-

$\{TTTTTTTT, HTTTTTTT, TTTTHTTT, TTHHTTTT, THTTTTTH, HTTHTHTT...\}$

e. Let's define the function $F(x) = P\{s \in S : X(s) \leq x\}$. Is $F(3)$ greater than, equal to or less than $P(B)$ from question 3b? Explain why.

Ans: $F(3)$ is the cumulative distributive function for random variable X which assigns the number of heads to every outcome.

The cdf of X for $x = 3$ can be given using R:

```
pbinom(3,size=8,prob=0.5)
```

```
## [1] 0.3632813
```

From above, we can collect that the value of $P(B)$ is :

$P(B) = 0.3632813$.

$\therefore P(B) = F(X \leq 3)$

The binomial probability distribution of a fair coin is symmetric. Thus the probability of getting 0, 1, 2, or 3 heads (less than or equal to 3 heads) is equal to the probability of getting 5, 6, 7, or 8 heads (more than 4 heads).

5. Use the urn {1, 3, 3, 3, 6, 6, 6, 6, 6, 6}. The experiment is to draw two tickets with replacement from this urn.

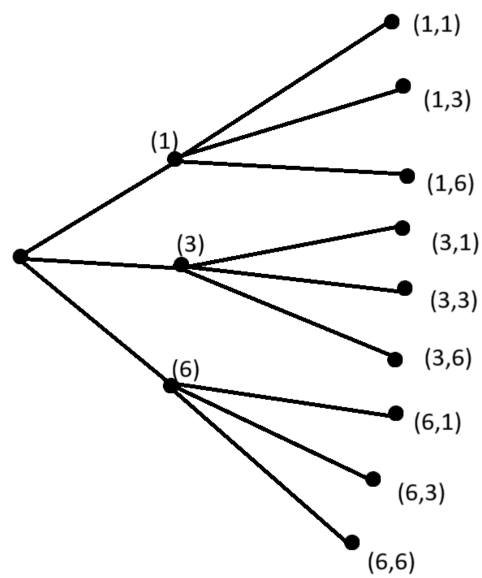
a. Write down the entire sample space for this experiment.

Ans: The sample space for the experiment is as follows:

$S = \{(1,1),(1,3),(1,6),(3,1),(3,3),(3,6),(6,1),(6,3),(6,6)\}$

b. Construct a decision tree where the first stage is to draw the first ticket and the second is to draw the second ticket (you'll have 9 branches)

Ans:



c. Let $A = \{\text{second ticket is a 6}\}$**.Use your decision tree to help you determine** $P(A)$**.**

Ans: The number of outcomes in which the second ticket is a 6 are 3. The total number of outcomes in the sample space of the experiment are 8.

Therefore the probability for event A is:

$P(A) = \frac{\text{number of outcomes for event A}}{\text{total number of outcomes for the experiment}}$

$\therefore P(A) = \frac{3}{9}$

$\therefore P(A) = 0.333$

d. Let the random variable X_1 **assign the sum of these two tickets, and the random variable** X_2 **assign the product of these two tickets. Write down**$X_1(S)$ **and** $X_2(S)$**.**

Ans: For X_1, the numbers assigned to the outcomes are:

$X_1(1,1) = 2$

$X_1(1,3) = 4$

$X_1(1,6) = 7$

$X_1(3,1) = 4$

$X_1(3,3) = 6$

$X_1(3,6) = 9$

$X_1(6,1) = 7$

$X_1(6,3) = 9$

$X_1(6,6) = 12$

Therefore Sample space for X_1 is :

(X_1(S)= {2,4,6,7,9,12}).

For X_2, the numbers assigned to the outcomes are:

$X_2(1,1) = 1$

$X_2(1,3) = 3$

$X_2(1,6) = 6$

$X_2(3,1) = 3$

$X_2(3,3) = 9$

$X_2(3,6) = 18$

$X_2(6,1) = 6$

$X_2(6,3) = 18$

$X_2(6,6) = 36$

Therefore sample space for X_2 is :

$X_2(S) = \{1, 3, 6, 9, 18, 36\}$

e. Let's define the function $F(x) = P(\{s \in S : X_2(s) \leq x\})$.What is $F(10)$? Use any information needed from the previous parts of this question. Show your work.

Ans: F(x) is defined as the probability that the product of two drawn tickets is less than or equal to 10.

From above, we can see that the values less than 10 from X_2 are : $1, 3, 6, 9$

Now to calculate the probabilities of the product of the drawn two tickets, we take account for their frequencies as well:

Total number of tickets is 10.

$\therefore$ total number of possible outcomes is $10 \times 10 = 100$.

Now, calculating probability for each product:

1) $1 \times 1 = 1$:

Probability of drawing 1 is $\frac{1}{10}$.

$\therefore$ probability of drawing tickets with product 1 is:

$P(X = 1) = \frac{1}{10} \times \frac{1}{100} = \frac{1}{10}$

2) $1 \times 3 = 3$ and $3 \times 1 = 3$:

Probability of drawing 1 is $\frac{1}{10}$.

Probability of drawing 3 is $\frac{3}{10}$.

$\therefore$ probability of drawing tickets with product 3 (either 1×3 or 3×1) is:

$P(X = 3) = 2 \times \left(\frac{1}{10} \times \frac{3}{10}\right) = \frac{6}{100}$

3) $1 \times 6 = 6$and $6 \times 1 = 6$:

Probability of drawing 1 is $\frac{1}{10}$.

Probability of drawing 6 is $\frac{6}{10}$.

$\therefore$ probability of drawing tickets with product 6 (either 1×6 or 6×1) is:

$P(X = 6) = 2 \times \left(\frac{1}{10} \times \frac{6}{10}\right) = \frac{12}{100}$

4) $3 \times 3 = 1$:

Probability of drawing 3 is $\frac{3}{10}$.

$\therefore$ probability of drawing tickets with product 3 is:

$P(X = 9) = \left(\frac{3}{10} \times \frac{3}{10}\right) = \frac{9}{100}$

Calculating the cumulative distributive function for X_2:

$F(X \leq 10) = P(X = 1) + P(X = 3) + P(X = 6) + P(X = 9)$

$F(X \leq 10) = \frac{1}{100} + \frac{6}{100} + \frac{12}{100} + \frac{9}{100}$

$\therefore F(X \leq 10) = \frac{28}{100} = 0.28$

Thus $F(10) = 0.28$ or 28