CS544, Fundamentals of Analysis Homework 2

Jefferson Parker

[japarker@bu.edu](mailto:japarker@bu.edu)

Part 1, Probability

1. From the data in Section 3.10, we have the following data:

7% of the population has lung disease; of these 90 % are smokers.

Of those not having lung disease (in the population), 25% are smokers.

P(A1) = 0.07 (has lung disease)

P(A2) = P(!A1) = 0.93 (does not have lung disease)

P(B|A1) = 0.9 (has lung disease AND is a smoker)

P(B|A2) = 0.25 (does not have lung disease AND is a smoker)

* 1. The above data present two pairs of mutually exclusive and collectively exhaustive possibilities: (Smoker OR Non-Smoker) AND (Lung disease OR No Lung Disease). The probability of a person having lung disease AND NOT being a smoker is P(B2|A1) = 10%. Similarly, the probability of a person not having lung disease AND NOT being a smoker is P(B2|A2) = 75%

P(A1|B2) =

P(A1) \* P(B2|A1) / P(A1) \* P(B2|A1) + P(A2) \* P(B2|A2) =

(0.07 \* 0.1) / (0.07 \* 0.1) + (0.93 \* 0.75) =

0.0099 =

**1%**

* 1. Similarly, the probability of a Non-smoker not having lung disease is

P(A2|B2) =

P(A2) \* P(B2|A2) / P(A1) \* P(B2|A1) + P(A2) \* P(B2|A2) =

(0.93 \* 0.75) / (0.07 \* 0.1) + (0.93 \* 0.75) =

0. 99 =

**99%**

See corresponding section (#Part 1.a) in the R code for programmatic support.

1. For the given sales tax by political party question, the distribution of probabilities is:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Democrat (A1) | Republican (A2) | Independent (A3) |
| Party | 0.4 | 0.5 | 0.1 |
| Support (B|An) | 0.7 | 0.4 | 0.2 |

The probability that a random sales tax supporter is a:

Democrat = P(A1|B) = (0.4 \* 0.7) / (0.4 \* 0.7) + (0.5 \* 0.4) + (0.1 \* 0.2) = 0.56 = 56%

Republican = P(A2|B) = (0.5 \* 0.4) / (0.4 \* 0.7) + (0.5 \* 0.4) + (0.1 \* 0.2) = 0.4 = 40%

Independent = P(A3|B) = (0.1 \* 0.2) / (0.4 \* 0.7) + (0.5 \* 0.4) + (0.1 \* 0.2) = 0.04 = 4%

See corresponding section (#Part 1.b) in the R code for programmatic support.

Part 2, Random Variables.

1. See corresponding section (#Part 2.a) in the R code for example of custom function using absolute value (abs()).
2. The probability of the absolute difference being at most (maximum) of 2 is 0.67 (2/3). The probability of the absolute difference being at least 3 is 0.33 (1/3).
3. The marginal probability of the absolute difference variable (A) is:

|  |  |
| --- | --- |
| A | probs |
| 0 | 0.16666667 |
| 1 | 0.27777778 |
| 2 | 0.22222222 |
| 3 | 0.16666667 |
| 4 | 0.11111111 |
| 5 | 0.05555556 |

1. The probability of the sum of two die rolls being even is 0.5 (1/2). The marginal distribution of the new random variable is:

|  |  |
| --- | --- |
| E | probs |
| FALSE | 0.5 |
| TRUE | 0.5 |

Part 3, Functions

1. See corresponding section (#Part 3) in the R code for solutions.

Part 4, R

1. See corresponding section (Part 4.a.) in the R code for solutions.
2. For the data, the Dow closed higher for 30 days and closed lower for 21 days.
3. The subset of days for which the Dow closed above 400 points from the previous day is:

DATE VALUE DIFF

25 2/6/18 24913 567

29 2/12/18 24601 410

47 3/9/18 25336 441

1. See corresponding section (Part 4.d) in the R code for programmatic portion of the solution. The longest gaining streak is:

DATE VALUE DIFF RUN

36 2/22/18 24962 164 Y

37 2/23/18 25310 348 Y

38 2/26/18 25709 399 Y