Assignment 1

anonymous

1 General information

2 Basic probability theory notation and terms

1. Probability: Probability is a measure of how likely an event or chance is to occur, typically expressed as a number between 0 (impossible) and 1 (certain).

2. Probability Mass Function: A probability mass function (PMF) assigns probabilities to discrete random variables, specifying the probability associated with each possible outcome.

3. Probability Density Function: A probability density function (PDF) characterizes the probability distribution of continuous random variables, indicating the likelihood of a variable falling within a specific range.

4. Probability Distribution: A probability distribution describes the set of all possible outcomes of a random event along with their associated probabilities.

5. Discrete Probability Distribution: A discrete probability distribution models random variables with distinct, separate outcomes and assigns probabilities to each individual outcome.

6. Continuous Probability Distribution: A continuous probability distribution models random variables with an infinite number of potential outcomes within a given range, and it is characterized by a probability density function.

7. Cumulative Distribution Function (CDF): The cumulative distribution function (CDF) provides the probability that a random variable takes on a value less than or equal to a given value, encompassing both discrete and continuous distributions.

8. Likelihood: Likelihood represents the probability of observing a set of data or evidence given a specific hypothesis or model

3 Basic computer skills

In 3a. I compute Alpha and Beta by formulas given in the assignment. Then compute the PDF of a beta distribution with dbeta command. Lastly the distribution is plotted.

```
# Do some setup:
distribution_mean = .2
distribution_variance = .01
```

```r
# You have to compute the parameters below from the given mean and variance
p = seq(0, 1, length=1000)

alpha <- (distribution_mean * (1 - distribution_mean) / distribution_variance - 1)*distribution_mean
beta <- alpha * (1 - distribution_mean) / distribution_mean
alpha
```

```
[1] 3
```

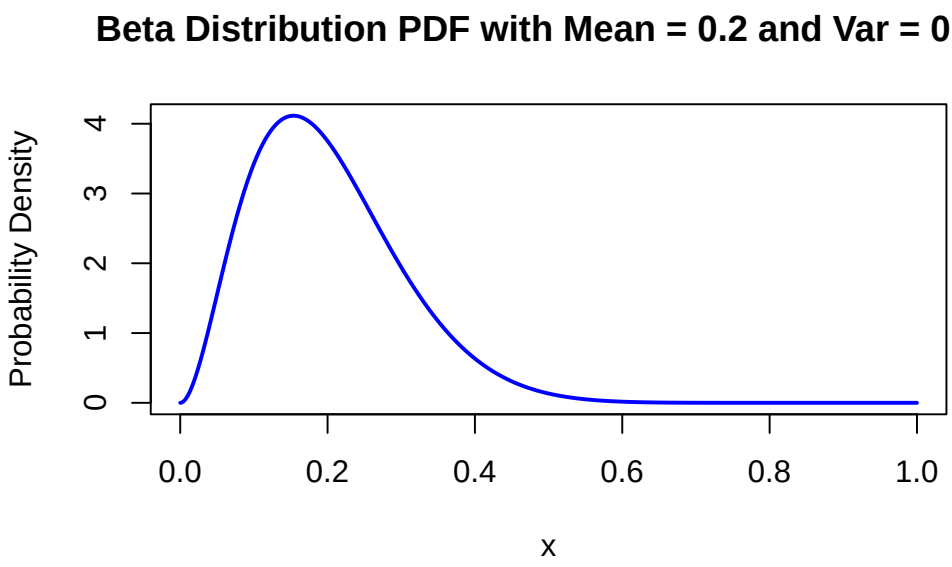
```r
beta
```

```
[1] 12
```

3.1 (a)

Plot the PDF here.

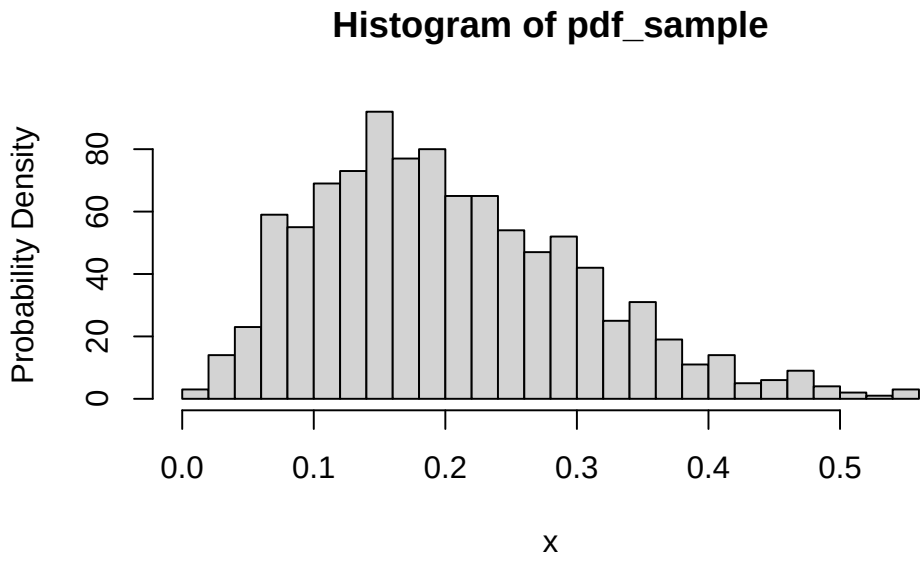
```r
p = seq(0, 1, length=1000)

pdf <- dbeta(p, alpha, beta)

# Plot the PDF
plot(p, pdf, type="l", col="blue", lwd=2,
     xlab="x", ylab="Probability Density",
     main=paste("Beta Distribution PDF with Mean =", round(distribution_mean, 2), "and Var =", round
```



3.2 (b)

Plot a histogram of 1000 random numbers from the Beta distribution. Looks clearly lookalike to the above graph but there is some variation because of random numbers.

```r
pdf_sample <- rbeta(1000, alpha, beta)
hist(pdf_sample , xlab="x",breaks= 30, ylab="Probability Density")
```




```r
# Useful functions: rbeta() and hist()
```

3.3 (c)

Printing the sample mean and variance which are nearly the same as the underlying variance and mean for the distribution.

```r
sample_mean = mean( pdf_sample)
sample_var = var(pdf_sample)

cat("sample mean:", sample_mean," sample variance: ", sample_var)
```

```
sample mean: 0.2031805  sample variance:  0.0102421
```

```r
# Useful functions: mean() and var()
```

3.4 (d)

Estimate the central 95% probability interval of the distribution from the drawn sample. The x postion for the 95% probability interval is 0.045 and 0.41

```r
confidence_interval = quantile(pdf_sample,c(0.025, 0.975))
cat("Central 95% Probability Interval: [", confidence_interval[1], ", ", confidence_interval[2], "]\
```

```
Central 95% Probability Interval: [ 0.04984255 ,  0.433158 ]
```

```
# Useful functions: quantile()
```

4 Bayes' theorem 1

4.1 (a) True positive = 98%, True negative = 96%, False

True positive + False negative = 1 => false negative = 2%

false positive + true negative = 1 = > false positive = 4%

Avg lung cancer 1/1000

P(has cancer | test result postive) = P(test result positive | has cancer) * P(has cancer) / P(test result is positive) =

```
# P(has cancer | test result postive) =
 0.98 * 1/1000 / (0.98*0.001+0.04*0.999)
```

```
[1] 0.02393747
```

In this assignment, the goal is to exploring the effectivness of the test. To test the effectivnesswe compute the probablity with Bayes theorem the probablity that a paitent has cancer and the test result is negative. The result is 2%, which means 2% of tests that show has cancer actually has cancer. This is low and would result in giving treatment to alot of people who dont need it. I would recommend to conclude a better test.

5 Bayes' theorem 2

Completing assignment 2 with Bayes theorem

```
boxes_test <- matrix(c(2,4,1,5,1,3), ncol = 2,
    dimnames = list(c("A", "B", "C"), c("red", "white")))
```

5.1 (a)

Keep the below name and format for the function to work with `markmyassignment`:

```
p_red <- function(boxes) {

    v <- c(0.4, 0.1, 0.5) # prbability of each box
    row_total = c(sum(boxes[1,]), sum(boxes[2,]), sum(boxes[3,]))
    boxes <- boxes/row_total # probablility for occurence of each  P(Red occurance) / P(all occuranc
    boxes <- v*boxes

    return(sum(boxes[,1])) # return sum of red occurance from each box

}
p_red(boxes_test)
```

```
[1] 0.3192857
```

5.2 (b)

Keep the below name and format for the function to work with `markmyassignment`:

```
boxes <- matrix(c(2,2,1,5,5,1), ncol = 2,
    dimnames = list(c("A", "B", "C"), c("red", "white")))

p_box <- function(boxes) {

    # P(B|red) = P(red|box) * P(box) / P(red)
    v <- c(0.4, 0.1, 0.5)
    row_sum = c(sum(boxes[1,]), sum(boxes[2,]), sum(boxes[3,]))
    prob <- boxes/row_sum
    return(prob[,1]*v/p_red(boxes))
    #c(0.29090909,0.07272727,0.63636364)
}

p_box(boxes_test)
```

```
        A         B         C
0.3579418 0.2505593 0.3914989
```

The most probable is that the red ball is from box C as it has highest probablity with 39%

6 Bayes' theorem 3

6.1 (a)

```
fraternal_prob = 1/125
identical_prob = 1/400 # change to 1/400
```

Keep the below name and format for the function to work with `markmyassignment`:

```
p_identical_twin <- function(fraternal_prob, identical_prob) {
    # Using Bayes rule
    # P(identical twins | twin brother) = P(Identical twin and twin brother)/ P(twin brother) = 1/2
    return( 1/2* identical_prob / (1/2*identical_prob + 1/4 * fraternal_prob) )
    0.4545455
}
p_identical_twin(fraternal_prob = fraternal_prob, identical_prob = identical_prob)
```

```
[1] 0.3846154
```

7 The three steps of Bayesian data analysis

7.1 (a)

1. Establishing a comprehensive probability model that encompasses the joint probability distribution of all known and unknown variables, aligning it with our understanding of the underlying scientific issue and the data collection process.
2. Incorporating observed data to compute and interpret the relevant posterior distribution, which represents the conditional probability distribution of the variables of primary interest, given the available data.
3. Assessing the model's fit and the ramifications of the resulting posterior distribution: gauging the model's fit with the data,are the substantive conclusions reasonable, and how sensitive are the results to the modeling assumptions in step 1? As a response, one can modify or expand by redoing the 3 step process.