BDA - Assignment 1

Anonymous

Exercise 1)

Probability: Likelihood of an outcome.

Probability mass: Probability of some discrete outcome.

Probability density: Probability of a continuous random variable being within a certain interval.

Probability mass function: Function that outputs the probability of a discrete outcome

Probability density function: Function that outputs the probability of a continuous outcome.

Probability distribution: Describes the probabilities of many/all outcomes.

Discrete probability distribution: Describes the probabilities of discrete outcomes.

Continuous probability distribution: Describes the probabilities of continuous outcomes.

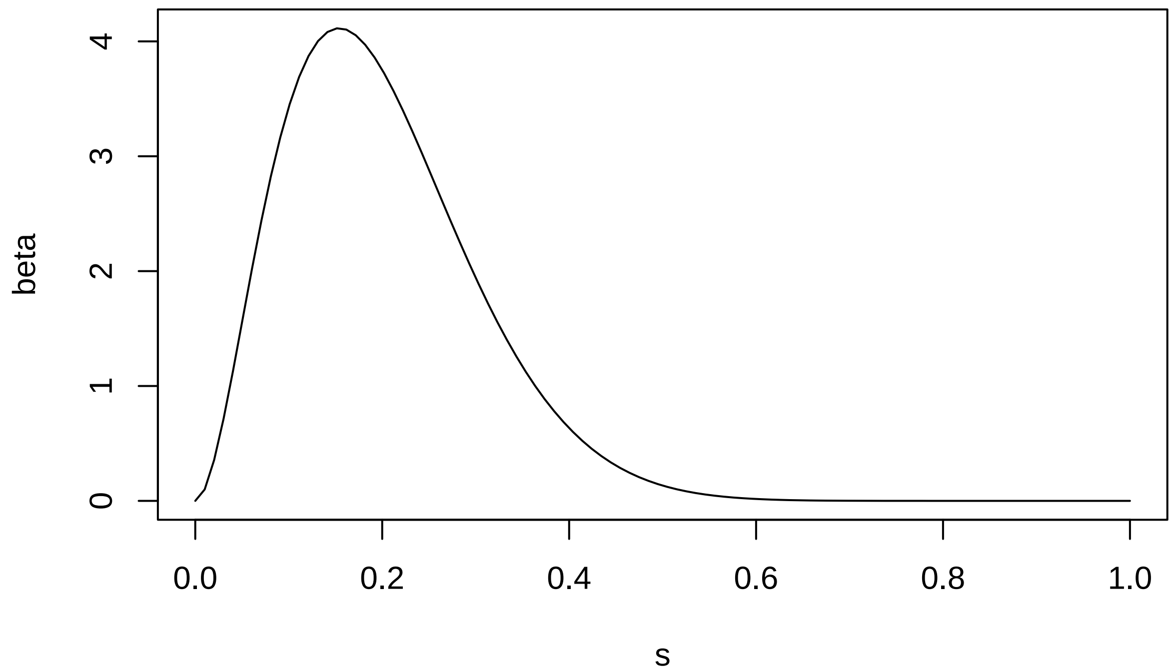
Cumulative distribution function: Describes the probability that an outcome is less than or equal to a certain value.

Likelihood: Likelihood of an outcome based on the past observation data.

Exercise 2)

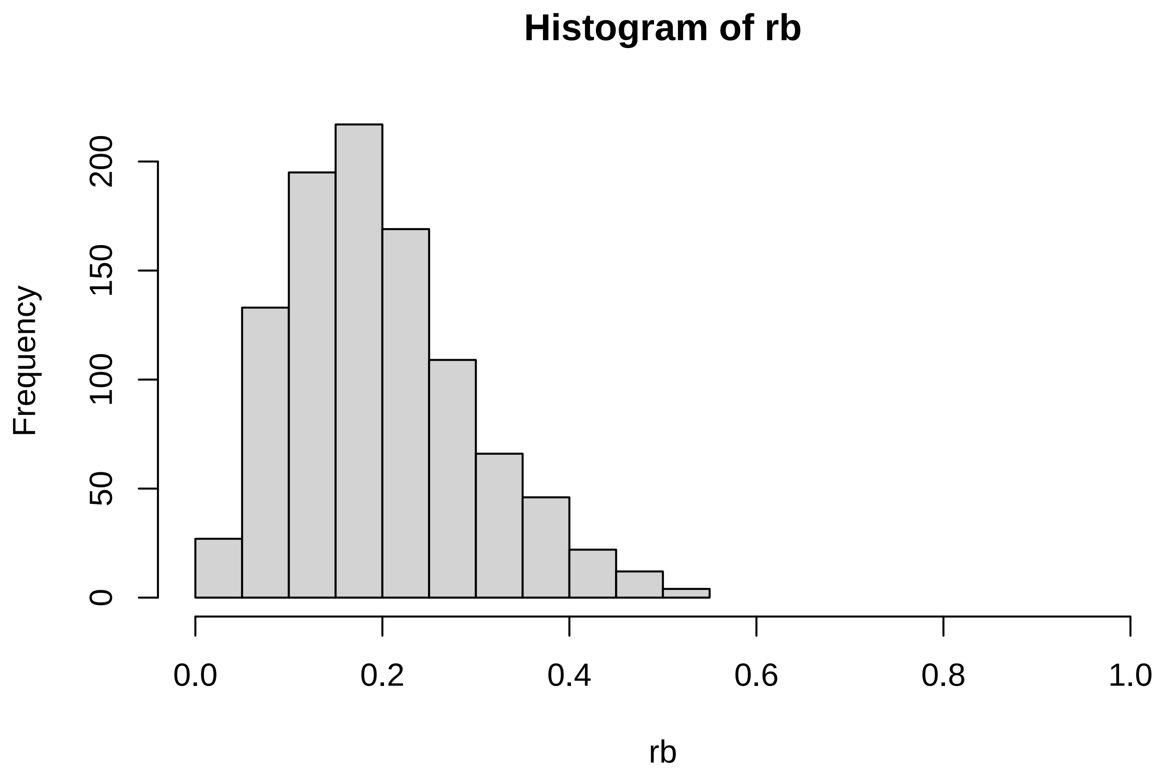
a)

```
mean <- 0.2
var <- 0.01

a <- mean * (mean * ( 1 - mean ) / var - 1 )
b <- a * (1 - mean) / mean

s <- seq(0,1, length=100)
beta <- dbeta(s, a, b)
plot(s, beta, type='l')
```



b)

```r
rb <- rbeta(1000, a, b)
hist(rb, xlim=c(0,1))
```



Visually the two plots look similar.

c)

```
mean(rb)
```

```
## [1] 0.1963122
```

```
var(rb)
```

```
## [1] 0.00951368
```

d)

```
quantile(rb, c(0.05, 0.95))
```

```
##         5%       95%
## 0.0619315 0.3835336
```

Exercise 3)

Define

P(T) = positive test result

P(!T) = negative test result

P(C) = has cancer

P(!C) = doesn't have cancer

and using Bayes rule we get

$P(C|T) = \frac{P(T|C)P(C)}{P(T)} = \frac{P(T|C)P(C)}{P(T|C)P(C)+P(T|!C)P(!C)}$

where

P(T|C) = 0.98

P(!T|!C) = 0.96

P(C) = 0.001

P(!C) = 0.999

P(T|!C) = 1 - 0.96 = 0.04

```
true_positive_rate <- 0.98
true_negative_rate <- 0.96
false_negative_rate <- 1 - true_negative_rate
cancer_rate <- 1 / 1000

true_positive_rate*cancer_rate/(true_positive_rate*cancer_rate + false_negative_rate*(1-cancer_rate))
```

```
## [1] 0.02393747
```

So even if the test indicates the person has cancer, they actually will have it only ~2.4% of the time. Therefore, I would advise against using the test.

Exercise 4)

P(A) = 0.4

P(B) = 0.1

P(C) = 0.5

a)

$P(R) = P(R|A)P(A) + P(R|B)P(B) + P(R|C)P(C)$

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

p_red <-function(boxes) {
  box_probabilities <- c(0.4, 0.1, 0.5)
  p_r <- colSums(boxes / rowSums(boxes)*box_probabilities)
```

```
  return(p_r['red'])
}

p_red(boxes)
```

```
##       red
## 0.3192857
```

b)

$P(Box|R) = \frac{P(R|Box)P(Box)}{P(R)}$

```
p_box <-function(boxes) {
  box_probabilities <- c(0.4, 0.1, 0.5)
  p_boxes <- (boxes / rowSums(boxes)*box_probabilities)/p_red(boxes)
  return(p_boxes[, 'red'])
}

p_box(boxes)
```

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

Box C is the most probable one.

Exercise 5)

P(Boy) = 0.5

P(Girl) = 0.5

P(Boy, Boy | Identical) = 0.5

P(Girl, Girl | Identical) = 0.5

P(Boy, Girl | Fraternal) = 0.5

P(Boy, Boy | Fraternal) = 0.25

P(Girl, Girl | Fraternal) = 0.25

Apply Bayes' rule

$P(Identical|Boy, Boy) = \frac{P(Boy,Boy|Identical)P(Identical)}{P(Boy,Boy)}$

where

$P(Boy, Boy) = P(Boy, Boy|Identical)P(Identical) + P(Boy, Boy|Fraternal)P(Fraternal)$

we get

$P(Identical|Boy, Boy) = \frac{P(Boy,Boy|Identical)P(Identical)}{P(Boy,Boy|Identical)P(Identical)+P(Boy,Boy|Fraternal)P(Fraternal)}$

$P(Identical|Boy, Boy) = \frac{0.5*1/400}{0.5*1/400+0.25*1/150} = 0.4285714$

```r
p_identical_twin <-function(fraternal_prob, identical_prob) {
  p_identical <- 0.5*identical_prob/(0.5*identical_prob + 0.25*fraternal_prob)
  return(p_identical)
}
p_identical_twin(fraternal_prob = 1/150, identical_prob = 1/400)
```

```
## [1] 0.4285714
```

The probability that Elvis was an identical twin is ~43%.