BDA - Assignment 1

Anonymous

Contents


```r
# To install aaltobda, see the General information in the assignment.
remotes::install_github("avehtari/BDA_course_Aalto", subdir = "rpackage", upgrade = "never")
```

```
## Skipping install of 'aaltobda' from a github remote, the SHA1 (38f34d35) has not changed since last i
##   Use `force = TRUE` to force installation
```

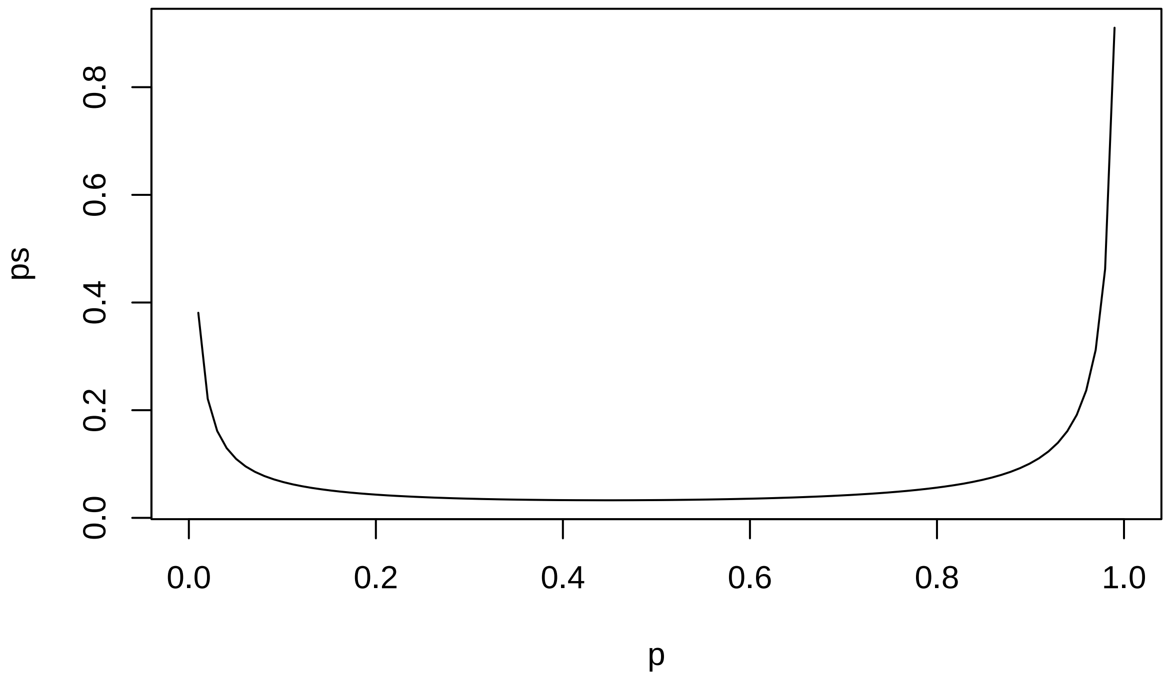
Exercise 1 (Basic probability theory notation and terms):

- **Probability:** Based on *known* parameters, how likely an specific event occurs.
- **Probability mass:** For discrete variables, how likely a series of events (could be a single event) occurs.
- **Probability density** For continuous variables, how likely an interval of events (probability at a single point is zero) occurs.
- **Probability mass function (pmf):** Related to *Probability mass*, it is the function that maps the event to the probability that it occurs.
- **Probability density function (pdf):** Related to *Probability density*, it is the function that maps the interval of events to the probability that it occurs.
- **Probability distribution:** The mass/density assigned by the pmf/pdf to each outcome occurrence.
- **Discrete probability distribution:** The mass assigned by the pmf to each outcome occurrence.
- **Continuous probability distribution:** The density assigned by the pdf to each outcome occurrence.
- **Cumulative distribution function:** The summed of the mass/density from all outcomes occurrence as they are arranged from left to right.
- **Likelihood:** The probability of observing the data given the parameters.

Exercise 2 (Basic computer skills):

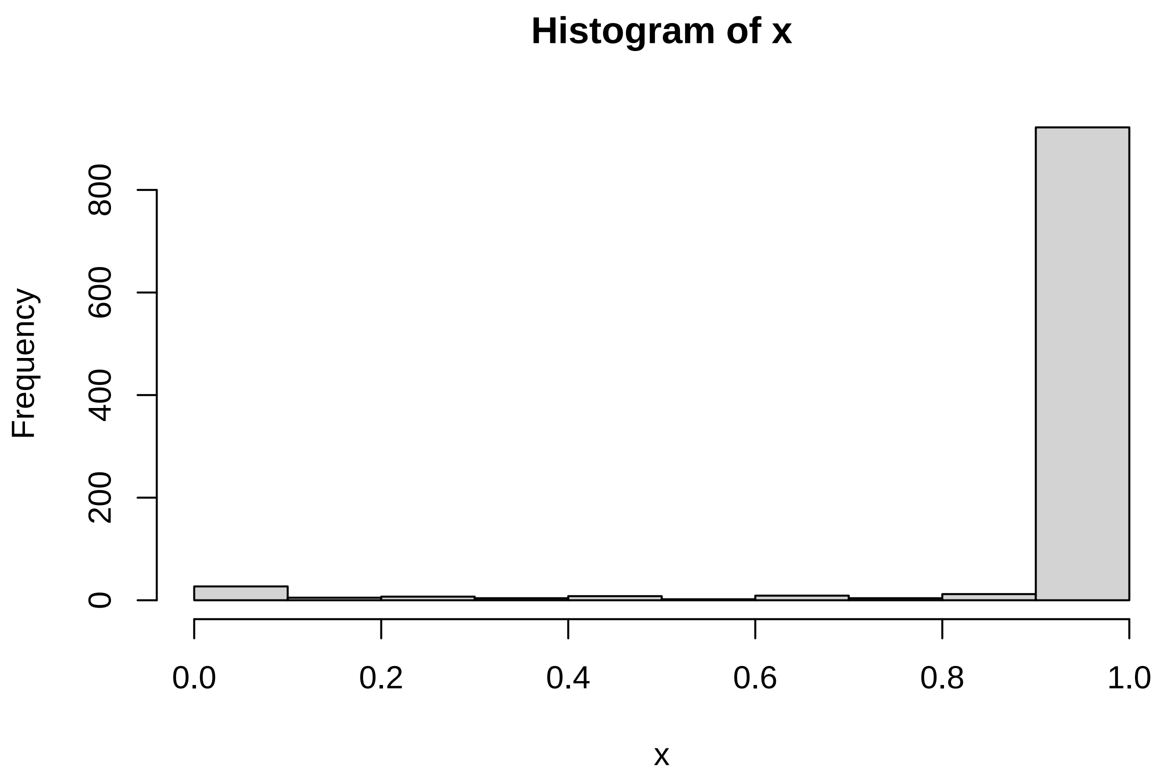
- a)

```r
p  <- seq(0, 1, length = 100)
mu <- 0.2
sd <- .01
ps <-  dbeta(p, mu, sd)
plot(p, ps, type = "l")
```



- b)

```r
n <- 1000
x <- rbeta(n, mu, sd)
hist(x)
```



- c)

```
cat("This is the mean:", mean(x))
```

```
## This is the mean: 0.9492692
```

```
cat("\n")
```

```
cat("This is the variance:", var(x))
```

```
## This is the variance: 0.03726147
```

- d)

```
quantile(x, prob = seq(0.1, 1, by = .05), CI.type = "two.sided")
```

```
##       10%       15%       20%       25%       30%       35%       40%       45% 
## 0.9937794 0.9999824 0.9999999 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
##       50%       55%       60%       65%       70%       75%       80%       85% 
## 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 
##       90%       95%      100% 
## 1.0000000 1.0000000 1.0000000
```

Exercise 3 (Bayes' theorem):

According from my notes for the exercise required in the pre-requisites, I found that $P(cancerdoesnotgetdetected) = P(cancer|negativeresult) = 0.000002$, and $P(positivetestbutnocancer) = P(\ cancer|positiveresult) = 0.97$. My conclusion was that while $P(C|N)$ is low, I am concerned about $P(\ C|P)$ being close to 1. I would pass on investing on this company.

Exercise 4 (Bayes' theorem):

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

```
## The following is the boxes matrix
```

```
boxes
```

```
##   red white
## A   2     5
## B   4     1
## C   1     3
```

```
p_red <- function(boxes) {
  Pr_boxes_balls <- boxes/rowSums(boxes)
  P_R_box <- matrix(Pr_boxes_balls[,1], ncol = 1)
  Pr_boxes <- matrix(c(.4,.1,.5), ncol = 3)
  p_red <- Pr_boxes %*% P_R_box
  return(p_red)
}

pred <- p_red(boxes)

cat("The probability of a red ball is P(R):", pred)
```

```
## The probability of a red ball is P(R): 0.3192857
```

```
p_box <- function(boxes) {
  Pr_boxes_balls <- boxes/rowSums(boxes)
  P_R_box <- matrix(Pr_boxes_balls[,1], ncol = 1)
  Pr_boxes <- matrix(c(.4,.1,.5), ncol = 3)
  P_box_R_list <- list()
  for(i in 1:3) {
    P_box_R <- (P_R_box[i]*Pr_boxes[i])/pred
    P_box_R_list[i] <- P_box_R
  }
  return(P_box_R_list)
}

cat("These are the probabilities of each box being picked given grabbing a red ball")
```

```
## These are the probabilities of each box being picked given grabbing a red ball
```

```
p_box(boxes)
```

```
## [[1]]
## [1] 0.3579418
##
## [[2]]
## [1] 0.2505593
##
## [[3]]
## [1] 0.3914989
```

```r
cat("\n")
```

```r
cat("The most likely box to be picked if we observed a red ball is c")
```

```
## The most likely box to be picked if we observed a red ball is c
```

Exercise (Bayes' theorem):

First thing to notice is that probabilities here dependent on Elvis and the other facts that we know. That is $P(B|I, Elvis) = 1$ because we know that Elvis was male. For sake of exposition, I dropped the *Elvis* from the notation.

Our bayes theorem formula is:

$P(I|B) = \frac{P(B|I)P(I)}{P(B)} = \frac{P(B|I)P(I)}{P(B|I)P(I)+P(B|F)P(F)}$

where I stands for identical twin, B is for fraternal. Let us assume the following (no strict) assumption holds:

$P(B|F) = \frac{P(B,F)}{P(F)} = \frac{P(B)P(F)}{P(F)} = P(B)$

The latter simply assumes that the probability of Elvis having a fraternal brother is idependent of the probability of the brother being a boy.

Therefore, $P(I|B)$ is equal to:

```r
fraternal_prob <- 1/150
identical_prob <- 1/400
p_identifical_twin <- function() {
  identical_prob/(identical_prob + .5*fraternal_prob)
}
p_identifical_twin()
```

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