

# BDA - Assignment 1

Anonymous

## Contents

Exercise 1)	1
Exercise 2)	2
Exercise 3)	4
Exercise 4)	5
Exercise 5)	6

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

```
# To install aaltobda, see the General information in the assignment.
library(aaltobda)
library(markmyassignment)
assignment_path <- paste("https://github.com/avehtari/BDA_course_Aalto/", "blob/master/assignments/tests.",
set_assignment(assignment_path)
```

```
## Assignment set:
## assignment1: Bayesian Data Analysis: Assignment 1
## The assignment contain the following (3) tasks:
## - p_red
## - p_box
## - p_identical_twin
```

## Exercise 1)

Probability: Probability is the measure of the likelihood that an event will occur in a Random Experiment. Probability is quantified as a number between 0 and 1, where, loosely speaking, 0 indicates impossibility and 1 indicates certainty. The higher the probability of an event, the more likely it is that the event will occur.

Probability mass: Probability for a discrete random variable to have a single outcome in the sample space.

Probability density: Probability for a continuous random variable to have a single outcome or a range of outcomes in the sample space.

Probability mass function (pmf): Probability mass function (PMF) is a function that gives the probability that a discrete random variable is exactly equal to one possible outcome.

Probability density function (pdf): A function of a continuous random variable, whose integral across an interval gives the probability that the value of the variable lies within the same interval.

Probability distribution: A mathematical function that maps the all possible outcomes of an random experiment with it's associated probability.

Discrete probability distribution: Probability distribution of a discrete random variable is called discrete probability distribution.

Continuous probability distribution: Probability distribution of a continuous random variable is called continuous probability distribution.

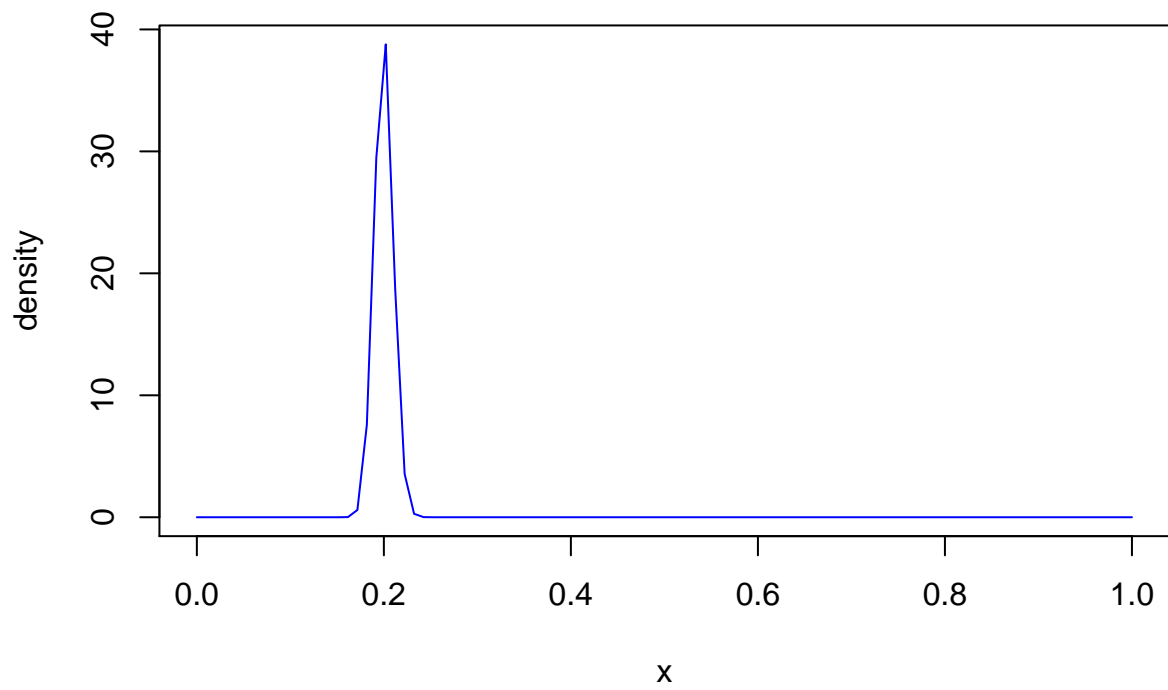
Cumulative distribution function (cdf): CDF gives the probability that a real-valued random variable is less than or equal to the independent variable of the function.

Likelihood: The state of being likely or probable.

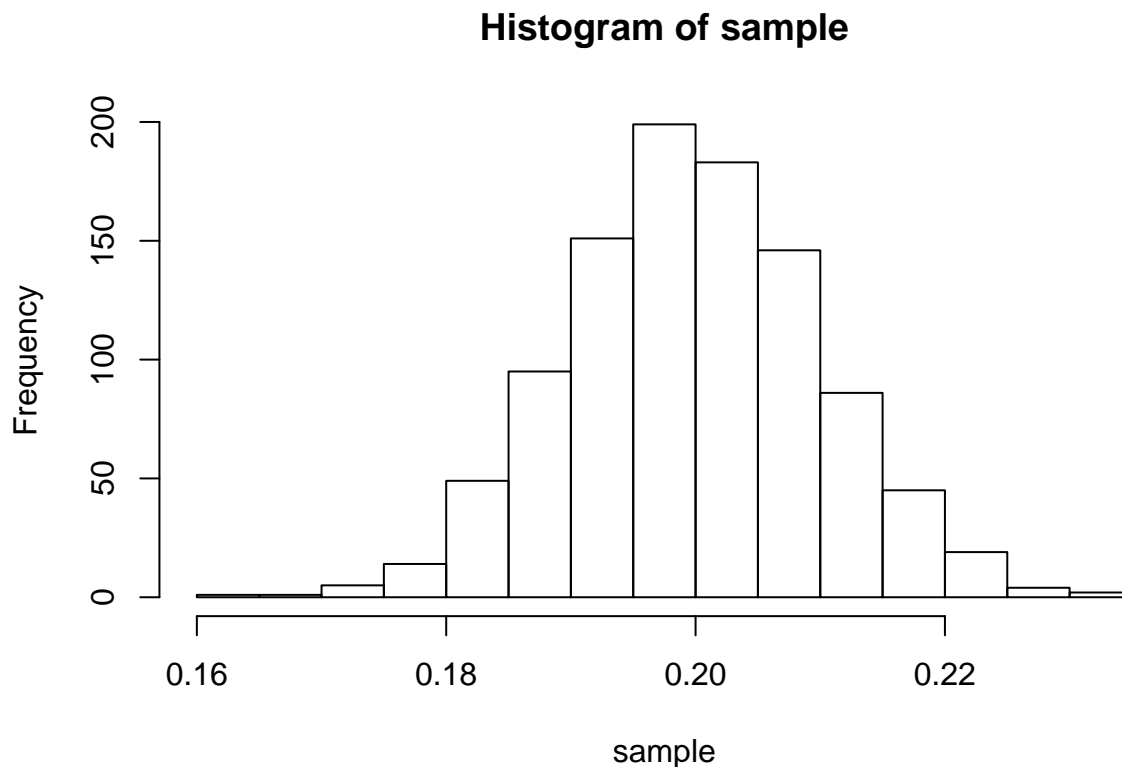
## Exercise 2)

a)

```
alpha <- 0.2*((0.2*(1-0.2)/0.01^2)-1)
beta <- alpha*(1-0.2)/0.2
x<-seq(0,1,length.out = 100)
dist<-dbeta(x, alpha, beta)
plot(x, dist, ylab="density", type ="l", col=4)
```



```
## b)
sample<-rbeta(1000,alpha,beta)
hist(sample)
```



The density function seemed to have most of the density around the mean 0.2. When taking random sample with  $n=1000$  from the same distribution, as one could assume, the histogram seems to follow a normal distribution around the mean 0.2. So, from this we can already assume that the sample mean and variance should be pretty close to the real value mean 0.2 at least. Let's see. ## c)

```
mean(sample)
```

```
## [1] 0.199886
```

```
# => 0.200482
```

```
var(sample)
```

```
## [1] 0.0001007889
```

```
# => 9.925429e-05
```

The mean roughly matches with the true mean but variance is a lot smaller compared to the true variance. ## d)

```
quantile(sample, c(0.025, 0.975))
```

```
##      2.5%      97.5%
```

```
## 0.1806697 0.2199822
```

## Exercise 3)

```
knitr::include_graphics("BDA_1_3.jpg")
```

Assignment 1  
# Exercise 3)

What is the probability of a patient having lung cancer, given he/she has gotten positive test results?

$P(+ | \text{cancer}) = 0,98$        $P(+ | \text{no-cancer}) = 0,04$   
 $P(- | \text{cancer}) = 0,02$        $P(- | \text{no-cancer}) = 0,96$   
 $P(\text{cancer}) = 0,001$        $P(\text{no-cancer}) = 0,999$   
 $P(\text{cancer} | +) = ?$

Bayes theorem

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

$$P(\text{cancer} | +) = \frac{P(+ | \text{cancer}) P(\text{cancer})}{P(+)}$$

$$= \frac{P(+ | \text{cancer}) P(\text{cancer})}{P(+ | \text{cancer}) P(\text{cancer}) + P(+ | \text{no-cancer}) P(\text{no-cancer})}$$

$$= \frac{0,98 \cdot 0,001}{0,98 \cdot 0,001 + 0,04 \cdot 0,999} \approx 0,02333 \dots \approx 2\%$$

So, there is only about 2% chance of a patient having lung cancer even if he or she is tested positive. I would advise the researchers not to start selling the product, as there will be many false positives (tested unnecessary medication, or further unwanted actions).

## Exercise 4)

a)

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

p_red<-function(boxes){
  a<-boxes[1,1]/sum(boxes[1, c(1,2)])*0.4
  b<-boxes[2,1]/sum(boxes[2, c(1,2)])*0.1
  c<-boxes[3,1]/sum(boxes[3, c(1,2)])*0.5

  a+b+c
}

p_red(boxes=boxes)

## [1] 0.3192857
```

b)

```
p_box<-function(boxes){
  a<-(boxes[1,1]/sum(boxes[1, c(1,2)])*0.4)/p_red(boxes = boxes)
```

```

b<-(boxes[2,1]/sum(boxes[2, c(1,2)])*0.1)/p_red(boxes = boxes)
c<-(boxes[3,1]/sum(boxes[3, c(1,2)])*0.5)/p_red(boxes = boxes)
c(a, b, c)
}
p_box(boxes=boxes)

```

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

If red ball was picked, it most probably came from box C.

```
mark_my_assignment()
```

```

## v | OK F W S | Context
##
/ | 0 | p_red()
v | 4 | p_red()
##
/ | 0 | p_box()
v | 4 | p_box()
##
/ | 0 | p_identical_twin()
x | 0 2 | p_identical_twin()
## -----
## test-3-1.R:8: failure: p_identical_twin()
## exists("p_identical_twin") isn't true.
## Error: p_identical_twin() is missing
##
## test-3-1.R:10: error: p_identical_twin()
## object 'p_identical_twin' not found
## Backtrace:
## 1. checkmate::expect_function(...) test-3-1.R:10:2
## 2. checkmate::checkFunction(x, args, ordered, nargs, null.ok)
## -----
##
## == Results =====
## Duration: 0.1 s
##
## OK:      8
## Failed:  2
## Warnings: 0
## Skipped: 0

```

## Exercise 5)

```

p_identical_twin<-function(fraternal_prob, identical_prob){
  pr_identical_and_twin_bros<-identical_prob*0.5
  pr_fraternal_and_twin_bros<-fraternal_prob*0.5*0.5

  pr_identical_and_twin_bros/(pr_identical_and_twin_bros+pr_fraternal_and_twin_bros)
}
p_identical_twin(fraternal_prob=1/150, identical_prob=1/400)

## [1] 0.4285714

```

```
mark_my_assignment()
```

```
## v | OK F W S | Context
```

```
##
```

```
/ | 0 | p_red()
```

```
v | 4 | p_red()
```

```
##
```

```
/ | 0 | p_box()
```

```
v | 4 | p_box()
```

```
##
```

```
/ | 0 | p_identical_twin()
```

```
v | 5 | p_identical_twin()
```

```
##
```

```
## == Results =====
```

```
## OK: 13
```

```
## Failed: 0
```

```
## Warnings: 0
```

```
## Skipped: 0
```

```
## Everything's correct!
```