BDA - Assignment 1

Anonymous

Table of Contents

# Loaded packages

library(aaltobda)  
library(ggplot2)  
library(rstan)

## Warning: package 'rstan' was built under R version 4.1.3

## Loading required package: StanHeaders

## Warning: package 'StanHeaders' was built under R version 4.1.3

##   
## rstan version 2.26.13 (Stan version 2.26.1)

## For execution on a local, multicore CPU with excess RAM we recommend calling  
## options(mc.cores = parallel::detectCores()).  
## To avoid recompilation of unchanged Stan programs, we recommend calling  
## rstan\_options(auto\_write = TRUE)  
## For within-chain threading using `reduce\_sum()` or `map\_rect()` Stan functions,  
## change `threads\_per\_chain` option:  
## rstan\_options(threads\_per\_chain = 1)

## Do not specify '-march=native' in 'LOCAL\_CPPFLAGS' or a Makevars file

library(markmyassignment)

## Warning: package 'markmyassignment' was built under R version 4.1.3

assignment\_path <-  
paste("https://github.com/avehtari/BDA\_course\_Aalto/",  
"blob/master/assignments/tests/assignment1.yml", sep="")  
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 (Basic probability theory notation and terms)

This can be trivial or you may need to refresh your memory on these concepts (see, e.g. Aalto course First Course in Probability and Statistics). Note that some terms may be different names for the same concept. Explain each of the following terms with one sentence: - probability: Probability is a way of quantifying the belief of something being true or false. - probability mass - probability density - probability mass function (pmf) - probability density function (pdf) - probability distribution - discrete probability distribution - continuous probability distribution - cumulative distribution function (cdf) - likelihood

# Exercise 2 (Basic computer skills)

This task deals with elementary plotting and computing skills needed during the rest of the course. You can use either R or Python, although R is the recommended language and we will only guarantee support in R. For documentation in R, just type ?{function name here}.

## a)

Plot the density function of Beta-distribution, with mean and variance . The parameters α and β of the Beta-distribution are related to the mean and variance according to the following equations

## b)

Take a sample of 1000 random numbers from the above distribution and plot a histogram of the results. Compare visually to the density function. Hint! Useful R functions: rbeta() and hist() ## c) Compute the sample mean and variance from the drawn sample. Verify that they match (roughly) to the true mean and variance of the distribution. Hint! Useful R functions: mean() and var() ## d) Estimate the central 95% probability interval of the distribution from the drawn samples. Hint! Useful R functions: quantile() # Plots

# Exercise 3 (Bayes’ theorem)

A group of researchers has designed a new inexpensive and painless test for detecting lung cancer. The test is intended to be an initial screening test for the population in general. A positive result (presence of lung cancer) from the test would be followed up immediately with medication, surgery or more extensive and expensive test. The researchers know from their studies the following facts: Test gives a positive result in 98% of the time when the test subject has lung cancer. - Test gives a negative result in 96 % of the time when the test subject does not have lung cancer. - In general population approximately one person in 1000 has lung cancer. The researchers are happy with these preliminary results (about 97% success rate), and wish to get the test to market as soon as possible. How would you advise them? Base your answer on Bayes’ rule computations. Hint : Relatively high false negative (cancer doesn’t get detected) or high false positive (unnecessarily administer medication) rates are typically bad and undesirable in tests. Hint : Here are some probability values that can help you fgure out if you copied the right conditional probabilities from the question. - P(Test gives positive | Subject does not have lung cancer) = 4% - P(Test gives positive and Subject has lung cancer) = 0.098% this is also referred to as the joint probability of test being positive and the subject having lung cancer

# Exercise 4 (Bayes’ theorem)

We have three boxes, A, B, and C. There are - 2 red balls and 5 white balls in the box A, - 4 red balls and 1 white ball in the box B, and - 1 red ball and 3 white balls in the box C. Consider a random experiment in which one of the boxes is randomly selected and from that box, one ball is randomly picked up. After observing the color of the ball it is replaced in the box it came from. Suppose also that on average box A is selected 40% of the time and box B 10% of the time (i.e. P(A) = 0.4). a) What is the probability of picking a red ball? b) If a red ball was picked, from which box it most probably came from? Implement two functions in R that computes the probabilities.

p\_red <- function(boxes){  
 13  
}  
p\_box <- function(boxes){  
 12  
}  
# To check your code/functions, just run  
mark\_my\_assignment()

## Warning: package 'testthat' was built under R version 4.1.3

## v | F W S OK | Context  
##   
## / | 0 | task-1-subtask-1-tests   
## / | 0 | p\_red()   
## x | 2 2 | p\_red()  
## --------------------------------------------------------------------------------  
## Failure (test-task-1-subtask-1-tests.R:16:3): p\_red()  
## p\_red(boxes = boxes) not equivalent to 0.3928571.  
## 1/1 mismatches  
## [1] 13 - 0.393 == 12.6  
## Error: Incorrect result for matrix(c(2,2,1,5,5,1), ncol = 2)  
##   
## Failure (test-task-1-subtask-1-tests.R:21:3): p\_red()  
## p\_red(boxes = boxes) not equivalent to 0.5.  
## 1/1 mismatches  
## [1] 13 - 0.5 == 12.5  
## Error: Incorrect result for matrix(c(1,1,1,1,1,1), ncol = 2)  
## --------------------------------------------------------------------------------  
##   
## / | 0 | task-2-subtask-1-tests   
## / | 0 | p\_box()   
## x | 2 2 | p\_box()  
## --------------------------------------------------------------------------------  
## Failure (test-task-2-subtask-1-tests.R:15:3): p\_box()  
## p\_box(boxes = boxes) not equivalent to c(0.29090909, 0.07272727, 0.63636364).  
## Lengths differ: 1 is not 3  
## Error: Incorrect result for matrix(c(2,2,1,5,5,1), ncol = 2)  
##   
## Failure (test-task-2-subtask-1-tests.R:19:3): p\_box()  
## p\_box(boxes = boxes) not equivalent to c(0.4, 0.1, 0.5).  
## Lengths differ: 1 is not 3  
## Error: Incorrect result for matrix(c(1,1,1,1,1,1), ncol = 2)  
## --------------------------------------------------------------------------------  
##   
## / | 0 | task-3-subtask-1-tests   
## / | 0 | p\_identical\_twin()   
## - | 1 0 | p\_identical\_twin()   
## x | 2 0 | p\_identical\_twin() [0.1s]  
## --------------------------------------------------------------------------------  
## Failure (test-task-3-subtask-1-tests.R:8:3): p\_identical\_twin()  
## exists("p\_identical\_twin") is not TRUE  
##   
## `actual`: FALSE  
## `expected`: TRUE   
## Error: p\_identical\_twin() is missing  
##   
## Error (test-task-3-subtask-1-tests.R:10:3): p\_identical\_twin()  
## Error in `checkFunction(x, args, ordered, nargs, null.ok)`: object 'p\_identical\_twin' not found  
## Backtrace:  
## 1. checkmate::expect\_function(...)  
## at test-task-3-subtask-1-tests.R:10:2  
## 2. checkmate::checkFunction(x, args, ordered, nargs, null.ok)  
## --------------------------------------------------------------------------------  
##   
## == Results =====================================================================  
## Duration: 0.2 s  
##   
## [ FAIL 6 | WARN 0 | SKIP 0 | PASS 4 ]

# Exercise 5 (Bayes’ theorem)

Assume that on average fraternal twins (two fertilized eggs and then could be of different sex) occur once in 150 births and identical twins (single egg divides into two separate embryos, so both have the same sex) once in 400 births (Note! This is not the true value, see Exercise 1.6, page 28, in BDA3). American male singer-actor Elvis Presley (1935 - 1977) had a twin brother who died in birth. Assume that an equal number of boys and girls are born on average. What is the probability that Elvis was an identical twin? Show the steps how you derived the equations to compute that probability. Implement this as a function in R that computes the probability.

p\_identical\_twin <- function(fraternal\_prob, identical\_prob){  
 return(14)  
}  
  
mark\_my\_assignment()

## v | F W S OK | Context  
##   
## / | 0 | task-1-subtask-1-tests   
## / | 0 | p\_red()   
## x | 2 2 | p\_red()  
## --------------------------------------------------------------------------------  
## Failure (test-task-1-subtask-1-tests.R:16:3): p\_red()  
## p\_red(boxes = boxes) not equivalent to 0.3928571.  
## 1/1 mismatches  
## [1] 13 - 0.393 == 12.6  
## Error: Incorrect result for matrix(c(2,2,1,5,5,1), ncol = 2)  
##   
## Failure (test-task-1-subtask-1-tests.R:21:3): p\_red()  
## p\_red(boxes = boxes) not equivalent to 0.5.  
## 1/1 mismatches  
## [1] 13 - 0.5 == 12.5  
## Error: Incorrect result for matrix(c(1,1,1,1,1,1), ncol = 2)  
## --------------------------------------------------------------------------------  
##   
## / | 0 | task-2-subtask-1-tests   
## / | 0 | p\_box()   
## x | 2 2 | p\_box()  
## --------------------------------------------------------------------------------  
## Failure (test-task-2-subtask-1-tests.R:15:3): p\_box()  
## p\_box(boxes = boxes) not equivalent to c(0.29090909, 0.07272727, 0.63636364).  
## Lengths differ: 1 is not 3  
## Error: Incorrect result for matrix(c(2,2,1,5,5,1), ncol = 2)  
##   
## Failure (test-task-2-subtask-1-tests.R:19:3): p\_box()  
## p\_box(boxes = boxes) not equivalent to c(0.4, 0.1, 0.5).  
## Lengths differ: 1 is not 3  
## Error: Incorrect result for matrix(c(1,1,1,1,1,1), ncol = 2)  
## --------------------------------------------------------------------------------  
##   
## / | 0 | task-3-subtask-1-tests   
## / | 0 | p\_identical\_twin()   
## x | 3 2 | p\_identical\_twin()  
## --------------------------------------------------------------------------------  
## Failure (test-task-3-subtask-1-tests.R:13:3): p\_identical\_twin()  
## p\_identical\_twin(fraternal\_prob = (1/125), identical\_prob = (1/300)) not equivalent to 0.4545455.  
## 1/1 mismatches  
## [1] 14 - 0.455 == 13.5  
## Error: Incorrect result for fraternal\_prob = 1/125 and identical\_prob = 1/300  
##   
## Failure (test-task-3-subtask-1-tests.R:16:3): p\_identical\_twin()  
## p\_identical\_twin(fraternal\_prob = 1/100, identical\_prob = 1/500) not equivalent to 0.2857143.  
## 1/1 mismatches  
## [1] 14 - 0.286 == 13.7  
## Error: Incorrect result for fraternal\_prob = 1/100 and identical\_prob = 1/500  
##   
## Failure (test-task-3-subtask-1-tests.R:19:3): p\_identical\_twin()  
## p\_identical\_twin(fraternal\_prob = 1/10, identical\_prob = 1/20) not equivalent to 0.5.  
## 1/1 mismatches  
## [1] 14 - 0.5 == 13.5  
## Error: Incorrect result for fraternal\_prob = 1/10 and identical\_prob = 1/20  
## --------------------------------------------------------------------------------  
##   
## == Results =====================================================================  
## [ FAIL 7 | WARN 0 | SKIP 0 | PASS 6 ]