BDA - Assignment 1

Anonymous

Contents

Exercise 1 1

Exercise 2 2

Exercise 3 4

Exercise 4 4

Exercise 5 5

```r
library(markmyassignment)
exercise_path = "https://github.com/avehtari/BDA_course_Aalto/blob/master/exercises/tests/ex1.yml"
set_assignment(exercise_path)
```

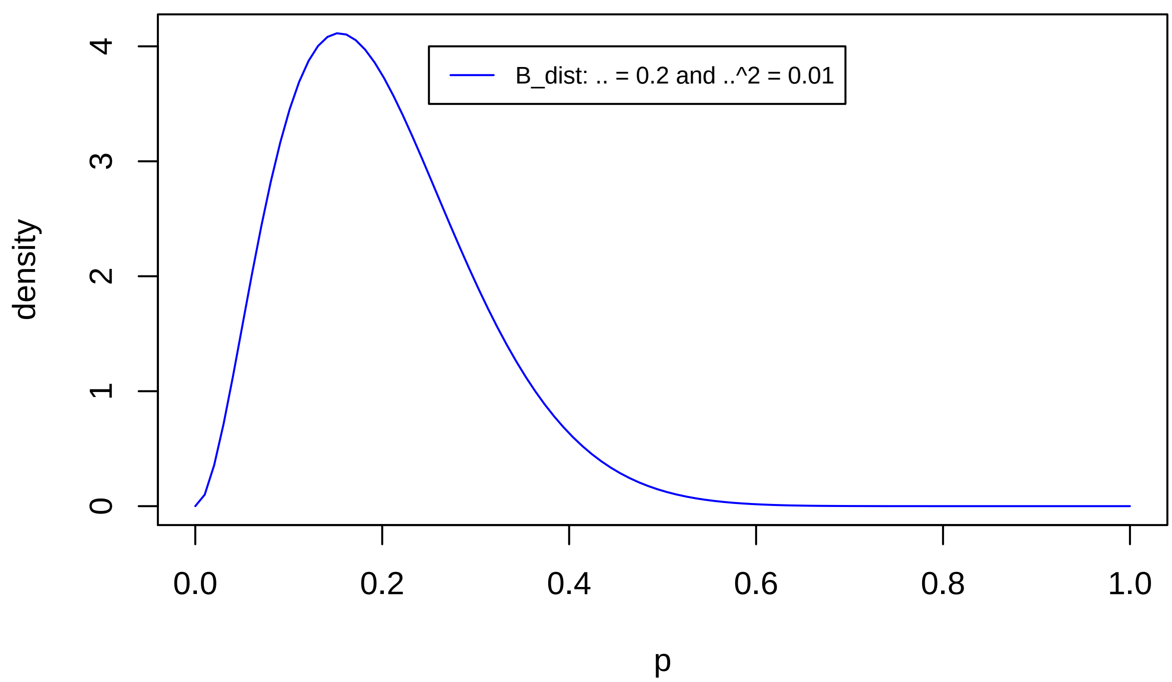
Exercise 1

- Probability: The likelihood of an event happening expressed between 0 and 1.
- Probability mass: A set of discrete random variables
- Probability density: A set of continuous random variables
- Probability mass function (pmf): Discrete probability distribution that gives the probabilities for discrete random variables.
- Probability density function (pdf): Continuous probability distribution that is used to spesify the probability of a real valued random variable belonging to a spesific range of values.
- Probability distribution: A function that describes all possible values and probabilitites that a random varaible can obtain with a given range.
- Discrete probability distribution: Describes the probability of each value of a discrete random variable
- Continuous probability distribution: Describes the probabilities of the possible values of a continuous random variable
- Cumulative distribution function (cdf) Describes a distribution of a real valued random variable Z, where Z returns the probability of it being less or equal than an event z.
- Likelihood Describes the probability of observations for different set of parameters.

Exercise 2

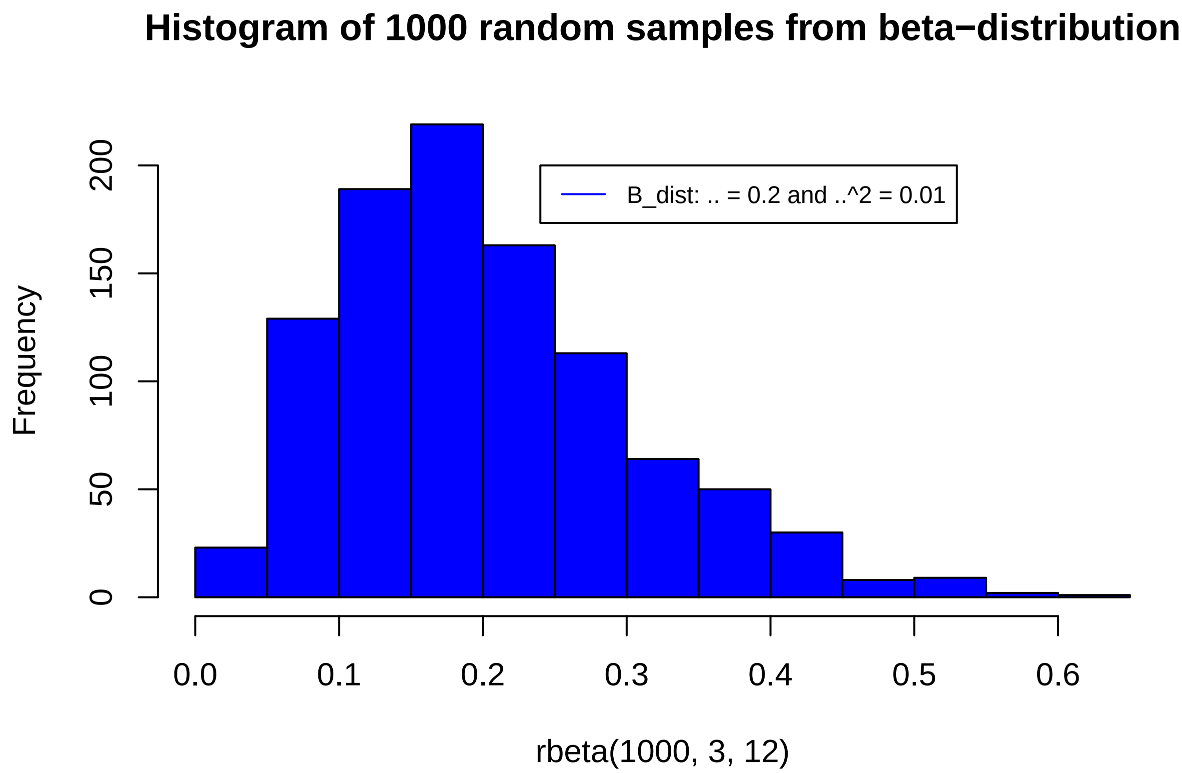
2a)

```
p = seq(from = 0, to = 1, length = 100)
plot(p, dbeta(p, 3, 12), ylab = "density", type = "l", col = 4)
legend(0.25, 4, c("B_dist:   = 0.2 and  ^2 = 0.01"), lty = c(1,
    1, 1, 1), col = c(4), cex = 0.75)
```



2b)

```
hist(rbeta(1000, 3, 12), main = "Histogram of 1000 random samples from beta-distribution",
    col = 4)
legend(0.24, 200, c("B_dist:   = 0.2 and  ^2 = 0.01"), lty = c(1,
    1, 1, 1), col = c(4), cex = 0.75)
```



2c)

```r
paste("Sample mean: ", mean(rbeta(1000, 3, 12)))
```

```
## [1] "Sample mean:  0.197624490329996"
```

```r
paste("Sample variance: ", var(rbeta(1000, 3, 12)))
```

```
## [1] "Sample variance:  0.00869805208113186"
```

```r
print("The obtained mean and variance are roughly the same as in the above distribution")
```

```
## [1] "The obtained mean and variance are roughly the same as in the above distribution"
```

2d)

```r
l = quantile(rbeta(1000, 3, 12), probs = 0.025)
u = quantile(rbeta(1000, 3, 12), probs = 0.975)
cat("The central 95%-interval of the distribution from the drawn\n samples is [",
    l, u, "]")
```

```
## The central 95%-interval of the distribution from the drawn
##  samples is [ 0.05145375 0.4396211 ]
```

Exercise 3

```r
result = function(pos_result, neg_result, cancer_odds) {
    result = cancer_odds * pos_result/((cancer_odds * pos_result) +
        (1 - cancer_odds) * (1 - neg_result))
    cat("By using Bayes formula P(A|B) = P(B|A)*P(A) / P(B|A)*P(A) + P(~A)*P(B |~A),",
        "we conclude\n that the probability of having cancer given a positive test",
        "result, formalized as\n p(has cancer|test result   is positive) is",
        result * 100, "%. This indicates\n that it is not reasonable to perform the test",
        "because there is   approximately only\n a 2.4% chance that the patient",
        "has a cancer if the test results is positive.")
}
result(0.98, 0.96, 1/1000)
```

```
## By using Bayes formula P(A|B) = P(B|A)*P(A) / P(B|A)*P(A) + P(~A)*P(B |~A), we conclude
##  that the probability of having cancer given a positive test result, formalized as
##  p(has cancer|test result   is positive) is 2.393747 %. This indicates
##  that it is not reasonable to perform the test because there is   approximately only
##  a 2.4% chance that the patient has a cancer if the test results is positive.
```

Exercise 4

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

A_prob = 0.4
B_prob = 0.1
C_prob = 0.5

p_red = function(boxes) {
    A_red = A_prob * boxes[1, 1]/(sum(boxes[1, 1], boxes[1, 2]))
    B_red = B_prob * boxes[2, 1]/(sum(boxes[2, 1], boxes[2, 2]))
    C_red = C_prob * boxes[3, 1]/(sum(boxes[3, 1], boxes[3, 2]))

    return(A_red + B_red + C_red/(A_red + B_red + C_red + (1 -
        sum(A_red + B_red + C_red))))
}

paste("Probability of drawing a red ball is", p_red(boxes))
```

```
## [1] "Probability of drawing a red ball is 0.319285714285714"
```

```r
p_box = function(boxes) {
```

```
    A_red = A_prob * boxes[1, 1]/(sum(boxes[1, 1], boxes[1, 2]))
    B_red = B_prob * boxes[2, 1]/(sum(boxes[2, 1], boxes[2, 2]))
    C_red = C_prob * boxes[3, 1]/(sum(boxes[3, 1], boxes[3, 2]))

    red_drawn = p_red(boxes)

    A_box = A_red/red_drawn
    B_box = B_red/red_drawn
    C_box = C_red/red_drawn
    cat("If a red ball was drawn, the probability that is came from box A is:",
        A_box, ", box B: \n", B_box, ", and box C: ", C_box,
        ". Hence, it is most", "likely that the red ball ws drawn\n from box C.")
}
p_box(boxes)
```

```
## If a red ball was drawn, the probability that is came from box A is: 0.3579418 , box B:
##  0.2505593 , and box C:  0.3914989 . Hence, it is most likely that the red ball ws drawn
##  from box C.
```

Exercise 5

```
p_identical_twin = function(fraternal_prob, identical_prob) {
    identical_twin = 1/2
    fraternal_twin = 1/4

    bb_it = identical_twin * identical_prob
    bb_ft = fraternal_twin * fraternal_prob

    it_tb = bb_it/sum(bb_it + bb_ft)
    paste("The probability that Elvis was an identical twin is: ",
        it_tb)
}

p_identical_twin(1/150, 1/400)
```

```
## [1] "The probability that Elvis was an identical twin is:  0.428571428571429"
```