BDA - Assignment 1

Anonymous

Contents

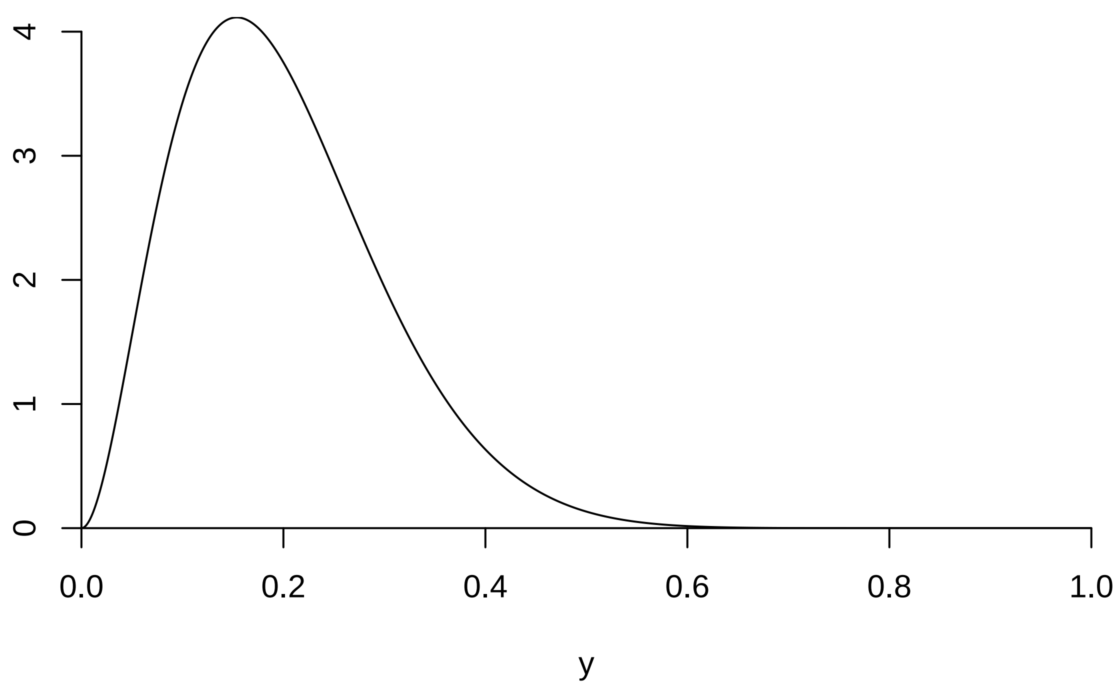
Exercise 1

- *Probabilities* - numerical quantities, defined on a set of 'outcomes,' that are nonnegative, additive over mutually exclusive outcomes,and sum to 1 over all possible mutually exclusive outcomes.[1]
- *Probability mass* - one of the ways of setting the distribution of a discrete random variable.
- *Probability density* - one of the ways of setting the distribution of a continuous random variable.
- *Probability mass function (pmf)* - a function that gives the probability that a discrete random variable is exactly equal to some value.
- *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 gives the probabilities of occurrence of different possible outcomes for an experiment.
- *Discrete probability distribution* - a probability distribution whose support is a countable set.
- *Continuous probability distribution* - a probability distribution whose support is an uncountable set, such as an interval in the real line.
- *Cumulative distribution function (cdf)* of a real-valued random variable X evaluated at x, is the probability that X will take a value less than or equal to x.
- *Likelihood* - a function $p(y|\theta)$, which, regarded as a function of θ, for fixed y, is used to affect posterior probability with data y.

Exercise 2

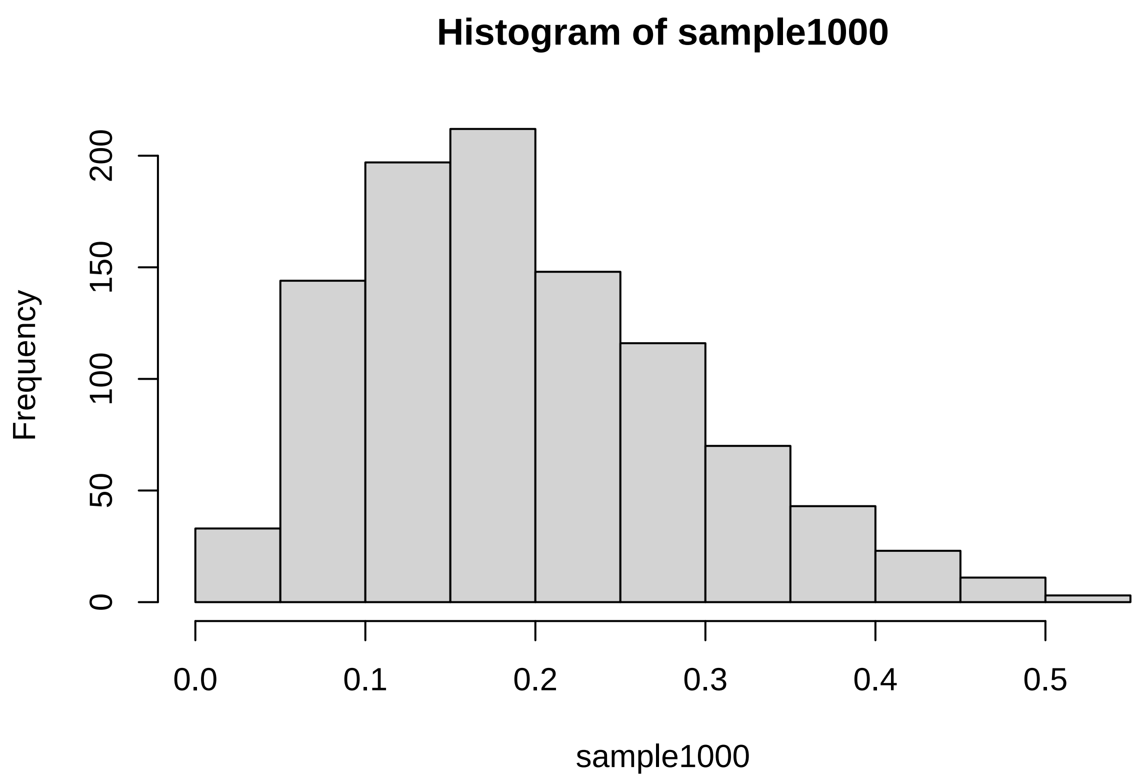
a)

```
x <- seq(0, 1, length.out = 10000)
m = .2
sigmasq = .01

alpha = m * (m * (1 - m) / sigmasq - 1)
beta = alpha * (1 - m) / m

plot(x, dbeta(x, alpha, beta), type="l", xlab="y", ylab="", xaxs="i",
     yaxs="i",  bty="n", cex=2)
```



b)

```
sample1000 = rbeta(1000, alpha, beta, ncp=0)
hist(sample1000)
```



c)

```
mean(sample1000)
```

```
## [1] 0.1946379
```

```
var(sample1000)
```

```
## [1] 0.009713001
```

d)

```
left_bound = mean(sample1000) - quantile(sample1000, probs = 0.95)
right_bound = mean(sample1000) + quantile(sample1000, probs = 0.95)

print(sprintf('The interval is [%f; %f]', left_bound, right_bound))
```

```
## [1] "The interval is [-0.184481; 0.573757]"
```

Exercise 3

```
options(width = 60)
# Event A = "Subject has cancer", event B = "Test is positive"
prob_A = 0.001
prob_BcondA = 0.98
prob_NotBcondNotA = 0.96

prob_AcondB = (prob_A * prob_BcondA) / (prob_A * prob_BcondA + (1 - prob_A) *
                                           (1 - prob_NotBcondNotA))

cat(sprintf('P(Subject has cancer | Test gives positive) = %f.
That means,that P(Subject does not have cancer | Test gives positive) = %f.
This is a high false positive rate, which means that there might be quite
a lot of unnecessarily administer medications, which is undesirable.',
prob_AcondB, 1-prob_AcondB))
```

```
## P(Subject has cancer | Test gives positive) = 0.023937.
## That means,that P(Subject does not have cancer | Test gives positive) = 0.976063.
## This is a high false positive rate, which means that there might be quite
## a lot of unnecessarily administer medications, which is undesirable.
```

Exercise 4

a) What is the probability of picking a red ball?

```
boxes <-
  matrix(
    c(2, 5, 4, 1, 1, 3),
    ncol = 2,
    byrow = TRUE,
    dimnames = list(c("A", "B", "C"), c("red", "white"))
  )
p_red <- function(boxes) {
  probA = 0.4
  probB = 0.1
  probC = 0.5
  prob = boxes[1, 1] / (boxes[1, 1] + boxes[1, 2]) * probA + boxes[2, 1] /
    (boxes[2, 1] + boxes[2, 2]) * probB + boxes[3, 1] / (boxes[3, 1] + boxes[3, 2]) *
    probC
  return(prob)
}
p_red(boxes = boxes)
```

```
## [1] 0.3192857
```

b) If a red ball was picked, from which box it most probably came from?

```
p_box <- function(boxes) {
  probA = 0.4
  probB = 0.1
  probC = 0.5

  boxA = (boxes[1, 1] / (boxes[1, 1] + boxes[1, 2])) * probA / p_red(boxes = boxes)
  boxB = (boxes[2, 1] / (boxes[2, 1] + boxes[2, 2])) * probB / p_red(boxes = boxes)
  boxC = (boxes[3, 1] / (boxes[3, 1] + boxes[3, 2])) * probC / p_red(boxes = boxes)

  return (c(boxA, boxB, boxC))
}
p_box(boxes = boxes)
```

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

Exercise 5

What is the probability that Elvis was an identical twin?

```
p_identical_twin <- function(fraternal_prob, identical_prob) {
  boy = .5
  girl = .5

  identical_twin = boy * identical_prob / (boy * identical_prob + boy * boy *
                                             fraternal_prob)
  return(identical_twin)
}

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

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

List of references

1. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). Bayesian Data Analysis (3rd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b16018