BDA - Assignment 1

Loaded packages

```
# To install aaltobda, see the General information in the assignment.
library(aaltobda)
```

Basic probability theory notation and terms

probability: the chance that a given event will occur (or is true), quantified between 0 and 1 where 0 is impossibility and 1 is certainty.

probability mass: probability distributions of discrete random variables

probability density: probability distributions of continuous random variables

probability mass function (pmf): function that provides the probability that a discrete random variable is exactly equal to a given value.

probability density function (pdf): function that provides the probability that a continuous random variable is exactly equal to a given value.

probability distribution

discrete probability distribution

continuous probability distribution

cumulative distribution function (cdf)

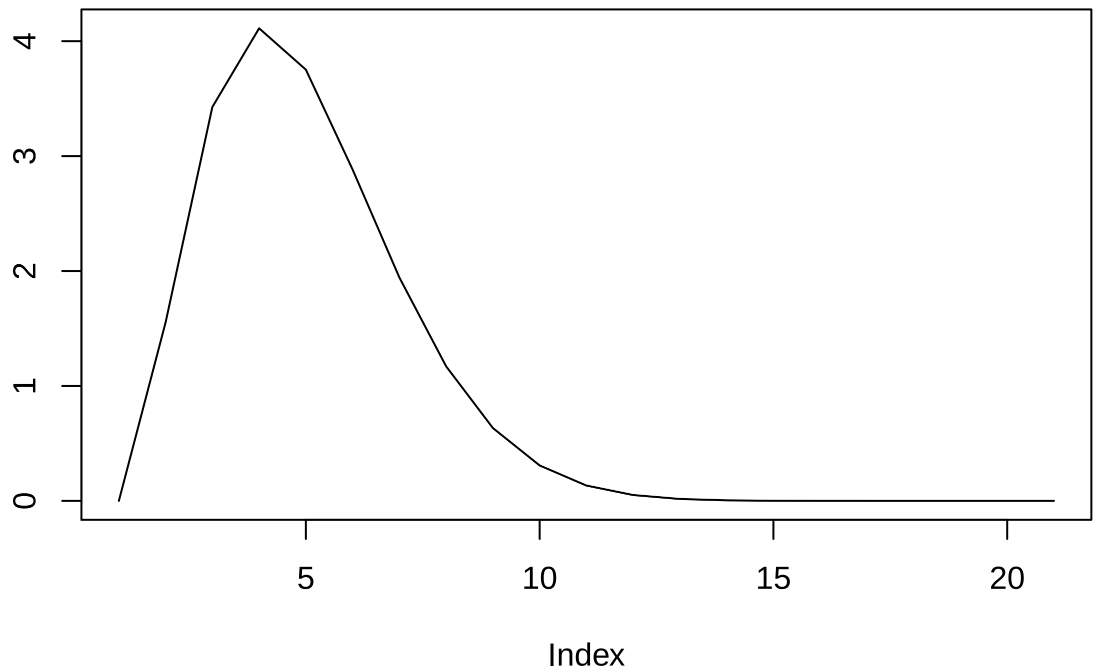
likelihood

Basic computer skills

a) Plot the density function of Beta-distribution, with mean $\mu = 0.2$ and variance $\sigma^2 = 0.01$.

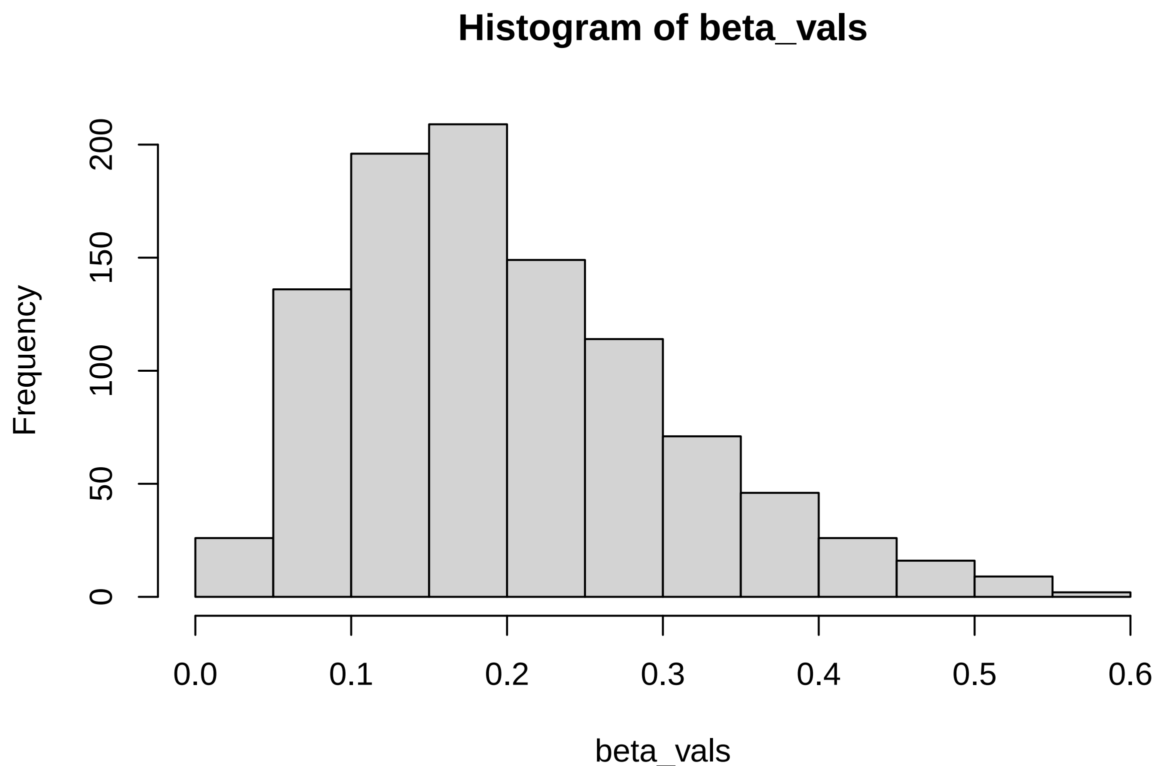
```
x <- seq(0,1, length = 21)
mu <- 0.2
sig <- 0.01

alpha <- mu*((mu*(1-mu)/sig)-1)
beta <- (alpha*(1-mu))/mu
beta_dist <- dbeta(x, shape1 = alpha, shape2 = beta)
plot(beta_dist, type = "l", ylab = "")
```



b) Take a sample of 1000 random numbers from the above distribution and plot a histogram of the results.

```
beta_vals <- rbeta(1000, alpha, beta)
hist(beta_vals)
```



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.

```
# sample mean and variance from the drawn sample
mean(beta_vals)
```

```
## [1] 0.2020462
```

```
mu
```

```
## [1] 0.2
```

```
var(beta_vals)
```

```
## [1] 0.01089246
```

```
sig
```

```
## [1] 0.01
```

```
# they match!
```

d) Estimate the central 95% probability interval of the distribution from the drawn samples.

```
quantile(beta_vals, probs = c(0.025, 0.975))
```

```
##       2.5%      97.5%
## 0.04974042 0.45457867
```

3. Bayes' theorem

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

where

$ P(A|B) $ is the conditional probability that A will occur given that B is true $ P(B|A) $ is the conditional probability that B will occur given that A is true $ P(A) $ and $ P(B) $ is the marginal probability of observing A and B respectively

Here, $A = Cancer =$ Subject has lung cancer $B = Positive =$ Test gives positive

So, we are looking for:

$$P(Cancer|Positive) = \frac{P(Positive|Cancer)P(Cancer)}{P(Positive)}$$

$$P(Cancer|Positive) = \frac{P(Positive|Cancer)P(Cancer)}{P(Positive|Cancer)P(Cancer) + P(Positive|NoCancer)P(NoCancer)}$$

We know that:

$P(Positive|Cancer) = 0.98$

$P(Cancer) = 1/1000 = 0.001$

$P(Positive|NoCancer) = 1 - 0.96 = 0.04$

$P(NoCancer) = 999/1000 = 0.999 $

so $P(Positive) = 0.98 * 0.001 + 0.04 * 0.999$

Meaning,

$$P(Cancer|Positive) = \frac{0.98 * 0.001}{0.04}$$

```
((0.98*0.001)/(0.98*0.001 + 0.04*0.999))*100
```

```
## [1] 2.393747
```

Which is very low. This means the joint probability of someone getting a positive test and having lung cancer is ~4%, which should be improved before getting the test to market.

4.

5.

```
p_identical_twin <- function(fraternal_prob, identical_prob){
  identical_prob/(fraternal_prob/2 + identical_prob)
}
p_identical_twin(identical_prob = 1/400, fraternal_prob = 1/150)
```

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