

TFALDS Lab 1

Klaudia Weigel

Contents

1	Definitions	1
2	Exercise 1	2
3	Exercise 2	11

1 Definitions

Definition 1 For a probability density function $f(x; \theta), \theta \in \Omega$ we define the *regularity conditions* as follows

(R0) The pdfs are distinct: $\theta \neq \theta' \implies f(x_i; \theta) \neq f(x_i; \theta')$.

(R1) The pdfs have common support for all θ .

(R2) The point θ_0 is an interior point in Ω .

(R3) The pdf $f(x; \theta)$ is twice differentiable as a function of θ .

(R4) The integral $\int f(x; \theta) dx$ can be differentiated twice under the integral sign as a function of θ .

(R5) The pdf $f(x; \theta)$ is three times differentiable as a function of θ . Further, for all $\theta \in \Omega$, there exists a constant c and a function $M(x)$ such that

$$\left| \frac{\sigma^3}{\sigma \theta^3} \log f(x; \theta) \right| \leq M(x)$$

with $E_{\theta_0}[M(X)] < \infty$, for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X .

Definition 2 The *Fisher Information* of the parameter α for a sample of size one is

$$\mathcal{I}(\alpha) = \text{Var} \left(\frac{\partial \log f(x; \alpha)}{\partial \alpha} \right)$$

For a random sample of size n we get the Fisher information is $n\mathcal{I}(\alpha)$, since

$$\text{Var} \left(\frac{\partial \log L(\mathbf{x}; \alpha)}{\partial \alpha} \right) = \text{Var} \left(\sum_{i=1}^n \frac{\partial \log f(x_i; \alpha)}{\partial \alpha} \right) = n\mathcal{I}(\alpha)$$

The Fisher Information can also be obtained from

$$\mathcal{I}(\alpha) = -E \left[\frac{\partial^2 \log f(x; \alpha)}{\partial \alpha^2} \right].$$

The above equality can be obtained from the fact that

$$1 = \int_{-\infty}^{\infty} f(x; \alpha) dx.$$

We are interested in a random variable X_1 which has pdf or pmf $f(x; \theta)$ where $\theta \in \Omega$. We assume that $\theta \in \omega_0$ or $\theta \in \omega_1$ where ω_0 and ω_1 are subsets of Ω and $\omega_0 \cup \omega_1 = \Omega$. We label the hypotheses as

$$H_0 : \theta \in \omega_0 \text{ versus } H_1 : \theta \in \omega_1$$

Let \mathcal{S} denote the support of the random sample $X = (X_1, \dots, X_n)$.

Definition 3 A test of H_0 versus H_1 is based on a subset C of \mathcal{S} . This set C is called the **critical region** and its corresponding decision rule is:

$$\text{Reject } H_0, \text{ (Accept } H_1), \quad \text{if } X \in C$$

$$\text{Retain } H_0, \text{ (Reject } H_1), \quad \text{if } X \in C^c$$

Definition 4 The **size** or **significance level** α is a number between 0 and 1, that indicates the probability of a Type I error, namely

$$\alpha = \max_{\theta \in \omega_0} P(X \in C)$$

Definition 5 The **power** of the test is defined as probability of rejecting H_0 , when it is not true

$$\gamma(\theta) = P_{H_1}(X \in C).$$

Theorem 1 (Neyman-Pearson Theorem.) Let X_1, X_2, \dots, X_n , where n is a fixed positive integer, denote a random sample from a distribution that has pdf or pmf $f(x; \theta)$. Then the likelihood of X_1, X_2, \dots, X_n is

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i; \theta).$$

Let θ_0 and θ_1 be distinct fixed values of θ so that $\Omega = \{\theta : \theta = \theta_0, \theta_1\}$, and let k be a positive number. Let C be a subset of the sample space such that:

- (a) $\frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})} \leq k$, for each point $\mathbf{x} \in C$.
- (b) $\frac{L(\theta_0; \mathbf{x})}{L(\theta_1; \mathbf{x})} \geq k$, for each point $\mathbf{x} \in C^c$.
- (c) $\alpha = P_{H_0}(\mathbf{X} \in C)$.

Then C is a best critical region of size α for testing the simple hypothesis $H_0 : \theta = \theta_0$ against the alternative simple hypothesis $H_1 : \theta = \theta_1$.

2 Exercise 1

Density of the beta distribution $Beta(\alpha, \beta)$:

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}; \quad x \in (0, 1), \quad \alpha, \beta > 0$$

In the exercise we are given a random sample X_1, \dots, X_n from the beta distribution $Beta(\alpha + 1, 1)$, namely the density function is $f(x; \alpha) = (\alpha + 1)x^\alpha$, for $x \in (0, 1)$ and $\alpha > 0$.

- a) Find the maximum likelihood estimator $\hat{\alpha}_{MLE}$ of the parameter α .

The likelihood function is

$$L(\mathbf{x}; \alpha) = \prod_{i=1}^n (\alpha + 1)x_i^\alpha = (\alpha + 1)^n \left[\prod_{i=1}^n x_i \right]^\alpha.$$

Taking the natural logarithm we get

$$l(\mathbf{x}; \alpha) = n \log(\alpha + 1) + \alpha \sum_{i=1}^n \log(x_i)$$

Now taking the derivative with respect to α and setting it to zero

$$\begin{aligned} \frac{\partial l(\mathbf{x}; \alpha)}{\partial \alpha} &= \frac{n}{\alpha + 1} + \sum_{i=1}^n \log(x_i) = 0 \\ \frac{n}{\alpha + 1} &= - \sum_{i=1}^n \log(x_i) \\ \frac{n}{\sum_{i=1}^n \log(x_i)} &= -1 - \alpha \\ \alpha &= - \frac{n}{\sum_{i=1}^n \log(x_i)} - 1. \end{aligned}$$

Finally let's check if we have indeed found the maximum, using the second derivative test

$$\frac{\partial^2 l(\alpha; \mathbf{x})}{\partial \alpha^2} = - \frac{n}{(\alpha + 1)^2}.$$

Since $n > 0$ and $\alpha > 0$, then the second derivative is negative, hence we have found the local maximum. So the MLE estimator is

$$\hat{\alpha}_{MLE} = - \frac{n}{\sum_{i=1}^n \log(x_i)} - 1.$$

- b) Find the Fisher Information and the asymptotic distribution of this estimator. How would you estimate the mean squared error of $\hat{\alpha}_{MLE}$?

$$\begin{aligned} \frac{1}{n} \mathcal{I}(\alpha) &= -E \left[\frac{\partial^2 \log f(x; \alpha)}{\partial \alpha^2} \right] \\ &= -E \left[\frac{\partial^2 \log(\alpha + 1) + \alpha \log x}{\partial \alpha^2} \right] \\ &= -E \left[\frac{\partial \left(\frac{1}{\alpha + 1} + \log x \right)}{\partial \alpha} \right] \\ &= -E \left[-\frac{1}{(\alpha + 1)^2} \right] = \frac{1}{(\alpha + 1)^2} \end{aligned}$$

Under regularity conditions (R0) - (R5) we have

$$\sqrt{n}(\hat{\alpha}_{MLE} - \alpha_0) \xrightarrow{d} N(0, \frac{1}{I(\alpha_0)}).$$

Where α_0 is the true value of α . Hence

$$\hat{\alpha}_{MLE} \xrightarrow{d} N(\alpha_0, \frac{(1 + \alpha_0)^2}{n^2}).$$

The mean squared error of an estimator α is $E_\alpha[(\hat{\alpha} - \alpha)^2] = \text{Var}(\hat{\alpha}) + [E(\hat{\alpha}) - \alpha]^2$. We can estimate the error from the asymptotic distribution of the MLE estimator. If the true value of α is α_0 , and n is the sample size, then the error is $\frac{(1 + \alpha_0)^2}{n^2}$.

c) Calculate $E(X_1)$ and use it to derive the moment estimator $\hat{\alpha}_M$ of α .

$$E[X_1] = \int_0^1 x(\alpha + 1)x^\alpha dx = (\alpha + 1) \frac{1}{\alpha + 2} x^{(\alpha+2)} \Big|_0^1 = \frac{\alpha + 1}{\alpha + 2}.$$

The first sample moment is

$$E[X] = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Equating the moments we get

$$\begin{aligned}\bar{X} &= \frac{\hat{\alpha}_M + 1}{\hat{\alpha}_M + 2} \\ (\hat{\alpha}_M + 2)\bar{X} &= \hat{\alpha}_M + 1 \\ \hat{\alpha}_M(\bar{X} - 1) &= 1 - 2\bar{X} \\ \hat{\alpha}_M &= \frac{1 - 2\bar{X}}{\bar{X} - 1}.\end{aligned}$$

d) Fix $\alpha = 5$ and generate one random sample with $n = 20$.

```
set.seed(2020)
```

```
sample <- rbeta(n = 20, shape1 = 6, shape2 = 1)
```

e) Calculate both estimators and the respective values of $\hat{\alpha} - \alpha$ and $(\hat{\alpha} - \alpha)^2$. Which estimator is more accurate ?

```
mle <- function(sample) {
  n = length(sample)
  return(-n/sum(log(sample)) - 1)
}
```

```
moment <- function(sample) {
  return((1-2*mean(sample))/(mean(sample)-1))
}
```

```
alpha_mle <- mle(sample)
alpha_moment <- moment(sample)
```

```
alpha_mle
```

```
## [1] 5.841329
```

```
alpha_moment
```

```
## [1] 6.07012
```

```
alpha_mle - 5
```

```
## [1] 0.8413285
```

```
alpha_moment - 5
```

```
## [1] 1.07012
```

```
(alpha_mle - 5)^2
```

```
## [1] 0.7078337
```

```
(alpha_moment - 5)^2
```

```
## [1] 1.145156
```

For this sample the mle estimator is more accurate, as both the difference and the difference squared is smaller for that estimator. Therefore it is closer to the true value of α .

f) *Generate 1000 samples.*

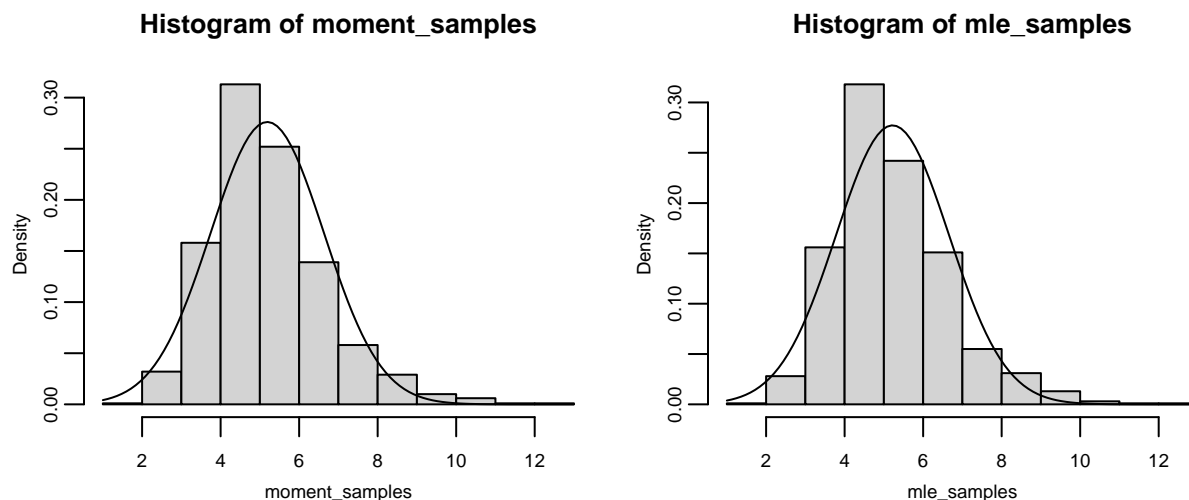
(i) *Draw histograms, box-plots and q-q plots for both estimators.*

Histograms for the mle estimators and the moment estimators, together with a curve from the normal distribution.

```
n = 20
samples = matrix(rbeta(1000 * n, shape1 = 6, shape2 = 1), n)
# MLE estimators for each the 1000 samples
mle_samples = apply(samples, 2, mle)
# Moment estimators for each of the 1000 samples
moment_samples = apply(samples, 2, moment)

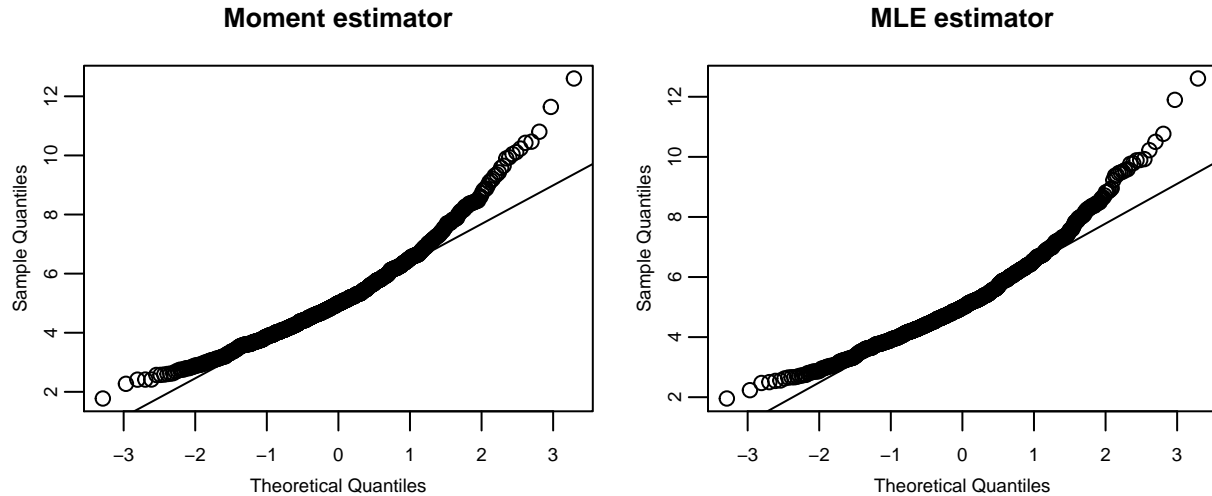
par(mfrow = c(1,2), mgp=c(1.3,0.5,0), mar=c(5,4,4,2)-1.5)
hist(moment_samples, probability = TRUE, cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
x<-seq(min(moment_samples),max(moment_samples),by=0.02)
curve(dnorm(x, mean = mean(moment_samples), sd = sd(moment_samples)), add=TRUE)

hist(mle_samples, probability = TRUE, cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
x<-seq(min(mle_samples),max(mle_samples),by=0.02)
curve(dnorm(x, mean = mean(mle_samples), sd = sd(mle_samples)), add=TRUE)
```



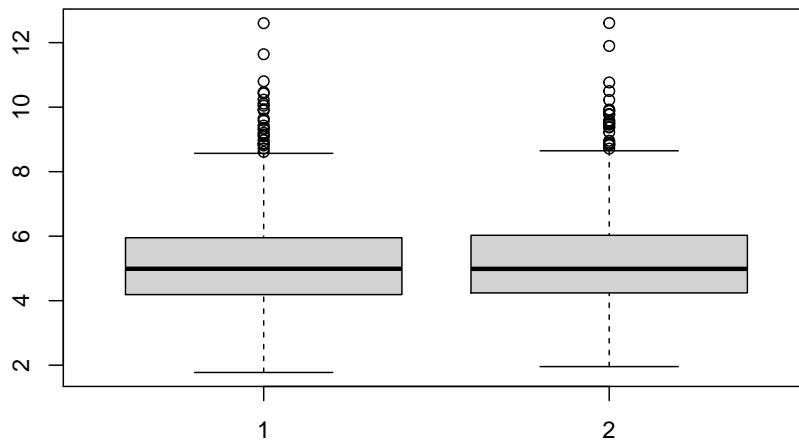
In both histograms we see that the data is mostly symmetric around the true value of $\alpha = 5$. It is slightly skewed to the right as there are more values to the right of the central point. The data for both estimators resembles the normal distribution and there is no significant difference between the distributions of the two estimators.

```
par(mfrow = c(1,2), mgp=c(1.3,0.5,0), mar=c(5,4,4,2)-1.5)
qqnorm(moment_samples, main = "Moment estimator", cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
qqline(moment_samples)
qqnorm(mle_samples, main = "MLE estimator", cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
qqline(mle_samples)
```



Both the moment estimators and the mle estimators lie close to the normal line, however the data is skewed especially towards the right edge. This corresponds to the fact that there is more data concentrated to the right, therefore the quantiles have higher values than the quantiles of the normal distribution. There is no significant difference between the two estimators.

```
boxplot(moment_samples, mle_samples)
```



The boxplots show no significant distortion of symmetry, therefore there is no reason reject normality of the data. The outliers confirm what we've observed before, namely that the data for both estimators has a right tail.

We see that for both estimators we cannot reject the assumption of normality.

(ii)

An estimator $\hat{\theta}$ is unbiased if $E[\hat{\theta}] = \theta$. Therefore the bias of an estimator is $E[\hat{\theta}] - \theta$.

```
get_bias = function(estimate, truth) {
  mean(estimate) - truth
}
```

```
get_mse = function(estimate, truth) {
  mean((estimate - truth) ^ 2)
}
```

```
var(moment_samples)
```

```
## [1] 2.086224
```

```
get_bias(moment_samples, 5)
```

```
## [1] 0.1889805
```

```
get_mse(moment_samples, 5)
```

```
## [1] 2.119852
```

```
var(mle_samples)
```

```
## [1] 2.07049
```

```
get_bias(mle_samples, 5)
```

```
## [1] 0.2208381
```

```
get_mse(mle_samples, 5)
```

```
## [1] 2.117189
```

The variance of the moment estimators is higher than that of the mle estimators. The bias is lower for the moment estimators, which is to be expected from a bias-variance tradeoff. The mean squared error is comparable for both the estimators.

From the Student's theorem we have that $Q = \frac{(n-1)S^2}{\sigma^2}$, where S^2 is the sample variance, has a $\chi^2(n-1)$ distribution. From that we can estimate the confidence interval for the variance:

$$0.95 = P(\chi_{0.025}^2 \leq Q \leq \chi_{0.975}^2) = P\left(\frac{(n-1)S^2}{\chi_{0.975}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{0.025}^2}\right)$$

```
n_samples = 1000
```

```
v_mle = var(mle_samples)
```

```
(n_samples - 1)*v_mle/qchisq(c(0.975, 0.025), n_samples-1)
```

```
## [1] 1.900270 2.264773
```

```
v_moment = var(moment_samples)
```

```
(n_samples - 1)*v_moment/qchisq(c(0.975, 0.025), n_samples-1)
```

```
## [1] 1.914711 2.281984
```

The confidence intervals for the variance, are close to each other, so we might expect to get similar results in terms variability.

For the bias and mse confidence intervals we will use the fact that if X is a random sample of size n from a distribution with a mean μ and a sample variance S^2 , then $\frac{\sqrt{n}(\bar{X}-\mu)}{S}$ has a Student t-distribution with $n-1$ degrees of freedom. So the 95% confidence interval for the mean will be

$$(\bar{X} - \frac{t_{0.975, df=n-1}S}{\sqrt{n}}, \bar{X} - \frac{t_{0.025, df=n-1}S}{\sqrt{n}}).$$

```
get_confint = function(sample_data) {
  s = sd(sample_data)
  n_data = length(sample_data)
  return(mean(sample_data) - qt(c(0.975, 0.025), df = n_data-1)*s/sqrt(n_data))
}
```

```
biases_mle = sapply(mle_samples, get_bias, truth=5)
get_confint(biases_mle)
```

```
## [1] 0.1315464 0.3101298
```

```
mse_mle = sapply(mle_samples, get_mse, truth=5)
get_confint(mse_mle)
```

```
## [1] 1.848345 2.386032
```

```
biases_moment = sapply(moment_samples, get_bias, truth=5)
get_confint(biases_moment)
```

```
## [1] 0.0993502 0.2786109
```

```
mse_moment = sapply(moment_samples, get_mse, truth=5)
get_confint(mse_moment)
```

```
## [1] 1.852282 2.387422
```

There is a better chance of the bias being smaller for the moment estimator. Since the mean squared error and the variance are comparable, the moment estimator might perform better.

The asymptotic distribution of $\hat{\alpha}_{MLE}$ is $N(5, 1.8)$. The bias is equal to $E[\hat{\alpha}] - \alpha_0 = 0$, mse and variance are both 1.8.

g) Repeat for $n=200$.

```
sample <- rbeta(n = 200, shape1 = 6, shape2 = 1)
```

```
alpha_mle <- mle(sample)
alpha_moment <- moment(sample)
```

```
alpha_mle
```

```
## [1] 5.25772
```

```
alpha_moment
```

```
## [1] 5.244323
```

```
alpha_mle - 5
```

```
## [1] 0.2577201
```

```
alpha_moment - 5
```

```
## [1] 0.2443234
```

```
(alpha_mle - 5)^2
```

```
## [1] 0.06641967
```

```
(alpha_moment - 5)^2
```

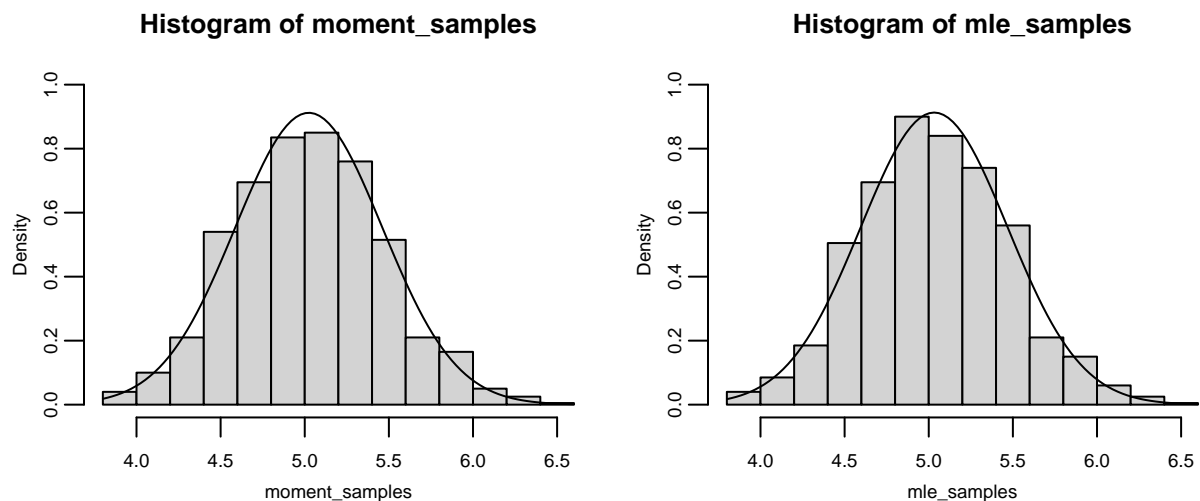
```
## [1] 0.05969393
```


Both estimators are closer to the true value. This time the moment estimator achieved better results and is closer to the true value than the mle estimator.

```
n = 200
# One sample in each column
samples = matrix(rbeta(1000 * n, shape1 = 6, shape2 = 1), n)
# MLE estimators for each the 1000 samples
mle_samples = apply(samples, 2, mle)
# Moment estimators for each of the 1000 samples
moment_samples = apply(samples, 2, moment)

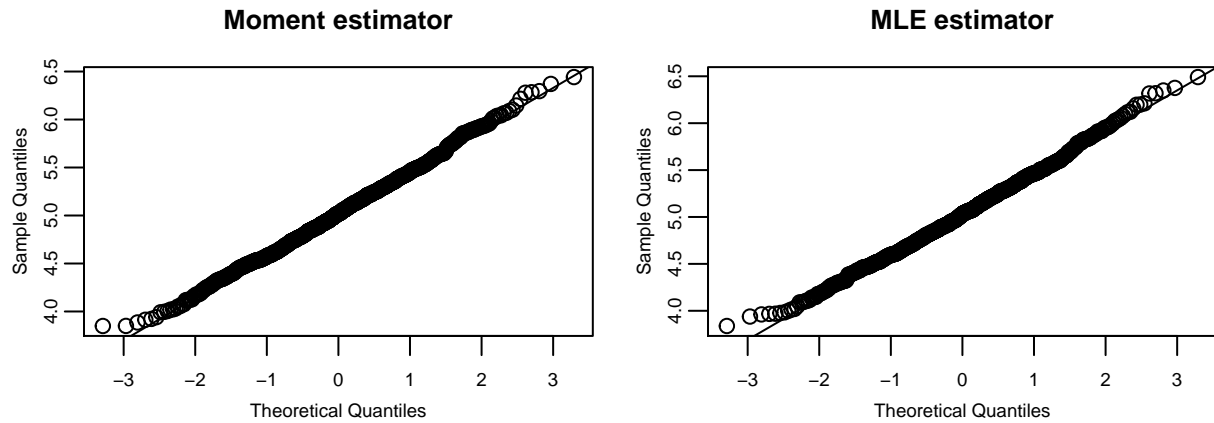
par(mfrow = c(1,2), mgp=c(1.3,0.5,0), mar=c(5,4,4,2)-1.5)
hist(moment_samples, probability = TRUE, cex.main = 0.8,
      cex.lab = 0.6, cex.axis = 0.6, ylim = c(0, 1))
x<-seq(min(moment_samples),max(moment_samples),by=0.02)
curve(dnorm(x, mean = mean(moment_samples), sd = sd(moment_samples)), add=TRUE)

hist(mle_samples, probability = TRUE, cex.main = 0.8,
      cex.lab = 0.6, cex.axis = 0.6, ylim = c(0, 1))
x<-seq(min(mle_samples),max(mle_samples),by=0.02)
curve(dnorm(x, mean = mean(mle_samples), sd = sd(mle_samples)), add=TRUE)
```



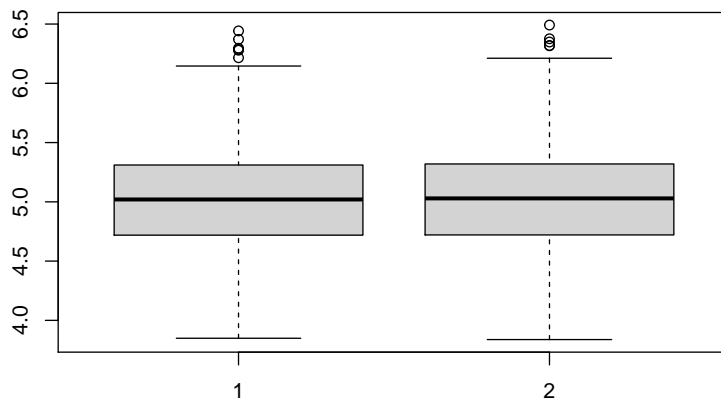
The histograms for both estimators visibly stabilized. There is no significant skewness and the data fits better to the normal distribution.

```
par(mfrow = c(1,2), mgp=c(1.3,0.5,0), mar=c(5,4,4,2)-1.5)
qqnorm(moment_samples, main = "Moment estimator", cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
qqline(moment_samples)
qqnorm(mle_samples, main = "MLE estimator", cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
qqline(mle_samples)
```



The data for both estimators fits better to the normal line. There is no significant difference between the estimators.

```
boxplot(moment_samples, mle_samples)
```



There are less outliers than for $n = 20$, both plots are symmetric.

From all the plots we see that the estimators are more closely centered around the true value than they were for $n = 20$. They fit better to the normal distribution.

```
var(moment_samples)
```

```
## [1] 0.1913273
```

```
get_bias(moment_samples, 5)
```

```
## [1] 0.02392704
```

```
get_mse(moment_samples, 5)
```

```
## [1] 0.1917085
```

```
var(mle_samples)
```

```
## [1] 0.1909974
```

```
get_bias(mle_samples, 5)
```

```
## [1] 0.03222589
```

```
get_mse(mle_samples, 5)
```

```
## [1] 0.1918449
```

The variance in the data is smaller when compared to $n = 20$. The bias and mse are also smaller when compared to $n = 20$.

```
n_samples = 1000
v_mle = var(mle_samples)
(n_samples - 1)*v_mle/qchisq(c(0.975, 0.025), n_samples-1)

## [1] 0.1752950 0.2089195

v_moment = var(moment_samples)
(n_samples - 1)*v_moment/qchisq(c(0.975, 0.025), n_samples-1)

## [1] 0.1755978 0.2092803

biases_mle = sapply(mle_samples, get_bias, truth=5)
get_confint(biases_mle)

## [1] 0.005105977 0.059345803

mses_mle = sapply(mle_samples, get_mse, truth=5)
get_confint(mses_mle)

## [1] 0.1748602 0.2088296

biases_moment = sapply(moment_samples, get_bias, truth=5)
get_confint(biases_moment)

## [1] -0.00321628 0.05107037

mses_moment = sapply(moment_samples, get_mse, truth=5)
get_confint(mses_moment)

## [1] 0.1750006 0.2084163
```

Similarly as it was for $n = 20$, the confidence intervals for the variance and the mse are comparable, however in case of the moment estimator it is more likely to get a smaller bias.

3 Exercise 2

Let X_1, \dots, X_n be a simple random sample from the exponential distribution with the density $f(x; \lambda) = \lambda e^{-\lambda x}$, for $x > 0$, $\lambda > 0$. Find the uniformly most powerful test at the level $\alpha = 0.05$ for testing the hypothesis $H_0 : \lambda = 5$ against $H_1 : \lambda = 3$.

We reject the null hypothesis if

$$\begin{aligned} \frac{L(\mathbf{X}; \lambda_0)}{L(\mathbf{X}; \lambda_1)} &\leq k \\ \frac{5^n \exp(-5 \sum_{i=1}^n x_i)}{3^n \exp(-3 \sum_{i=1}^n x_i)} &\leq k \\ \left(\frac{5}{3}\right)^n \exp(-2 \sum_{i=1}^n x_i) &\leq k \\ \sum_{i=1}^n x_i &\geq -\frac{n}{2} \log \left(\frac{3k^{1/n}}{5} = c \right) \end{aligned}$$

The uniformly most powerful critical region is

$$C = \{T(X) = \sum_{i=1}^n X_i : T(X) \geq c\}.$$

Where the test statistic $T(X)$ has $\Gamma(n, \frac{1}{5})$ under H_0 and $\Gamma(n, \frac{1}{3})$ under H_1 . a)

$$\begin{aligned} P_{H_0}(T \in C) &= \alpha \\ P_{H_0}\left(\sum_{i=1}^n X_i \geq -\frac{n}{2} \log\left(\frac{3k^{1/n}}{5}\right)\right) &= 0.05 \\ c &= -\frac{n}{2} \log\left(\frac{3k^{1/n}}{5}\right) \\ \left(\frac{5}{3}\right)^n e^{-2c} &= k \end{aligned}$$

We can get c from the distribution tables of the $\text{Gamma}(n, \frac{1}{5})$ distribution, as it is the 0.95th quantile ($c = \Gamma_{n, 1/5, 0.95}$). Therefore the critical region is

$$C = \{T(X) = \sum_{i=1}^n X_i : T(X) \geq \Gamma_{n, 1/5, 0.95}\}.$$

b)

The power of the test is the probability that we reject the null hypothesis when the alternative hypothesis is true

$$\begin{aligned} P_{H_1}(T \in C) &= \gamma \\ P_{H_1}\left(\sum_{i=1}^n X_i \geq \Gamma_{n, 1/5, 0.95}\right) &= \gamma \end{aligned}$$

c)

The p value is the probability that a test statistic, as a random variable, has values more extreme than the currently observed value of the test statistic under the null hypothesis. If the critical region is of the form $T(\mathbf{X}) \geq c$, then the p value is the probability that $T(\mathbf{X}) \geq T(\mathbf{x})$.

```
n=20
c = qgamma(0.95, shape = n, scale = 1/5)
exp_sample = rexp(n=n, rate = 5)
pgamma(sum(exp_sample), shape=n, scale = 1/5, lower.tail = FALSE)
```

```
## [1] 0.9962134
```

The p-value is large, which is to be expected since we have drawn a sample from the null hypothesis.

d)

The observed value of p is $p = P_{H_0}(T(\mathbf{X}) \geq T(\mathbf{x}))$. Let F be the CDF of $T(\mathbf{X})$ under H_0 (in our case $F = \Gamma(n, 1/5)$). We can write p as a random variable $P = 1 - F(T(\mathbf{X}))$.

$$\begin{aligned}
P_{H_0}(P \leq x) &= P_{H_0}(1 - F(T(\mathbf{X})) \leq x) \\
&= P_{H_0}(F(T(\mathbf{X})) \geq 1 - x) \\
&= P_{H_0}(T(\mathbf{X}) \geq F^{-1}(1 - x)) \\
&= 1 - F(F^{-1}(1 - x)) = x
\end{aligned}$$

The p-value has a uniform distribution on $[0, 1]$.

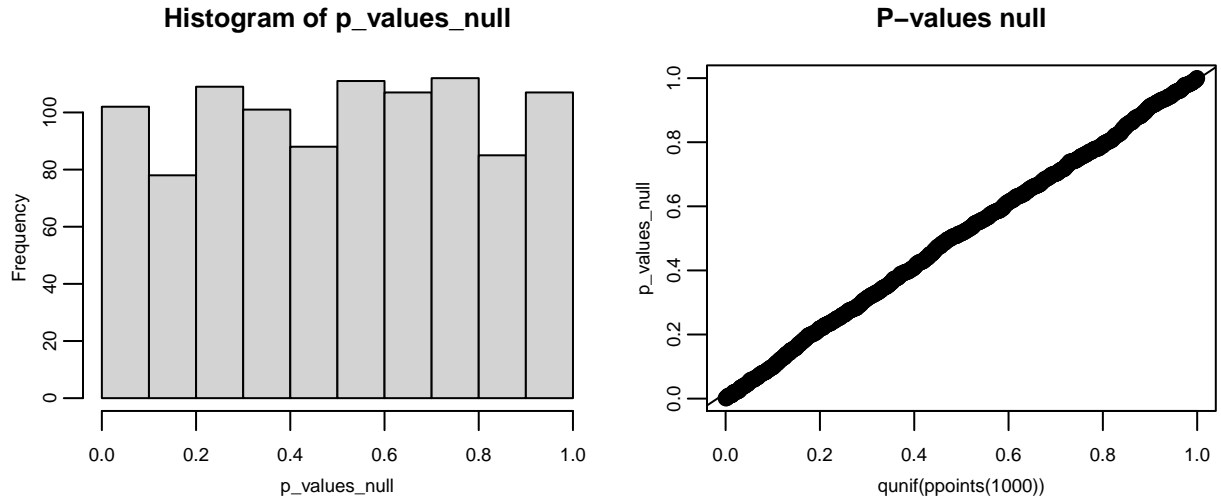
e) (i)

```

n = 20
samples_exp_null = matrix(rexp(1000 * n, rate = 5), n) # samples from H_0 distribution
get_prob <- function(sample, shape_param, scale_param) {
  return(pgamma(sum(sample), shape=shape_param, scale = scale_param, lower.tail = FALSE))
}
p_values_null = apply(samples_exp_null, 2, get_prob, shape_param=n, scale_param = 1/5)

par(mfrow = c(1,2), mgp=c(1.3,0.5,0), mar=c(5,4,4,2)-1.5)
hist(p_values_null, cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
## Q-Q plot for the p-value data against true theoretical distribution:
qqplot(qunif(ppoints(1000)), p_values_null,
       main = "P-values null", cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
qqline(p_values_null, distribution = function(p) qunif(p))

```



We can see that both the histogram and the qqplot confirm the uniform distribution of the p-values.

- (ii) To construct a confidence interval for the type I error of our test we will first set up a Bernoulli trial where success signifies the fact that we've committed a type I error. This occurs when a p-value under the null hypothesis is less than the significance level $\alpha = 0.05$. Let p be the probability of success (committing a type I error) and p_i a p-value in an i -th sample, $i = 1, \dots, k$ (in our case $k=1000$). Then our Bernoulli variables are $X_1 = \mathbb{I}_{\{p_1 < \alpha\}}, \dots, X_k = \mathbb{I}_{\{p_k < \alpha\}}$ and it is satisfied that $P(X_i = 1) = p$. The unbiased estimator for p is $\hat{p} = \sum_{i=1}^k X_i / k = \bar{X}_k$, and we have $\mu_{\hat{p}} = E[\hat{p}] = p$, $\sigma_{\hat{p}}^2 = Var[\hat{p}] = \frac{1}{k}p(1-p)$. From the central limit theorem $(\hat{p} - \mu_{\hat{p}})/\sigma_{\hat{p}}$ has the $N(0, 1)$ distribution, so to get the 95% confidence interval for p :

$$\begin{aligned}
0.95 &= P(-1.96 < \frac{(\hat{p} - p)}{\sqrt{\frac{1}{k}p(1-p)}} < 1.96) \\
&= P(-1.96\sqrt{p(1-p)}/\sqrt{k} < \hat{p} - p < 1.96\sqrt{p(1-p)}/\sqrt{k}) \\
&= P(\hat{p} - 1.96\sqrt{p(1-p)}/\sqrt{k} < p < \hat{p} + 1.96\sqrt{p(1-p)}/\sqrt{k}) \\
&\approx P(\hat{p} - 1.96\sqrt{\hat{p}(1-\hat{p})}/\sqrt{k} < p < \hat{p} + 1.96\sqrt{\hat{p}(1-\hat{p})}/\sqrt{k}).
\end{aligned}$$

```

k=1000
y = sum(p_values_null < 0.05)
p_hat = y/k
p_hat + c(-1.96, 1.96)*(sqrt((p_hat*(1-p_hat))/k))

```

```
## [1] 0.03562036 0.06237964
```

f) Generate 1000 samples of size $n = 20$ from the alternative distribution and calculate respective p -values.

(i) Compare the distribution of p -values under H_0 and under the alternative.

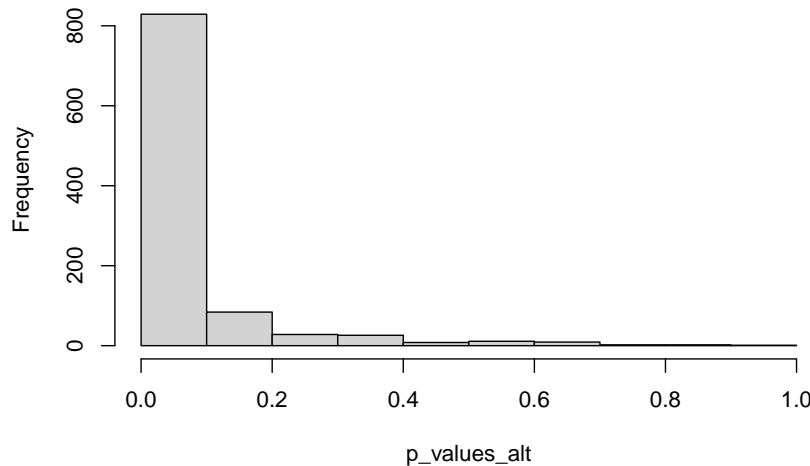
```

samples_exp_alt = matrix(rexp(1000 * n, rate = 3), n) # samples from H_1
p_values_alt = apply(samples_exp_alt, 2, get_prob, shape_param=n, scale_param = 1/5)

```

```
hist(p_values_alt)
```

Histogram of p_values_alt



We see that under the alternative hypothesis most of p -values are close to zero, which is to be expected since we want to reject the null hypothesis.

(ii) We can get the power of a test by first computing the probability of type II error (accepting H_0 when H_1 is true) β . The power is then defined as $1 - \beta$. In order to accept the null hypothesis when it is not true the p -value under the alternative hypothesis has to be larger than the significance level. Knowing that we can use the same method as in the previous point.

```

k=1000
y = sum(p_values_alt > 0.05)
p_hat = y/k
int = 1 - (p_hat + c(1.96, -1.96)*(sqrt((p_hat*(1-p_hat))/k)))
int

```

```
## [1] 0.7231616 0.7768384
```

Theoretically calculated power

```
1 - pgamma(qgamma(0.95, shape = 20, scale = 1/5), shape = 20, scale = 1/3)
```

```
## [1] 0.7581204
```

g) Repeat points c)-f) for $n = 200$.

```
n=200
```

```
exp_sample_200 = rexp(n=n, rate = 5)
```

```
pgamma(sum(exp_sample_200), shape=n, scale = 1/5, lower.tail = FALSE)
```

```
## [1] 0.5832104
```

The p-value is still larger than the significance level, however is it smaller than for $n = 20$.

```
samples_exp_null_200 = matrix(rexp(1000 * n, rate = 5), n)
```

```
p_values_null_200 = apply(samples_exp_null_200, 2, get_prob, shape_param=n, scale_param = 1/5)
```

```
par(mfrow = c(1,2), mgp=c(1.3,0.5,0), mar=c(5,4,4,2)-1.5)
```

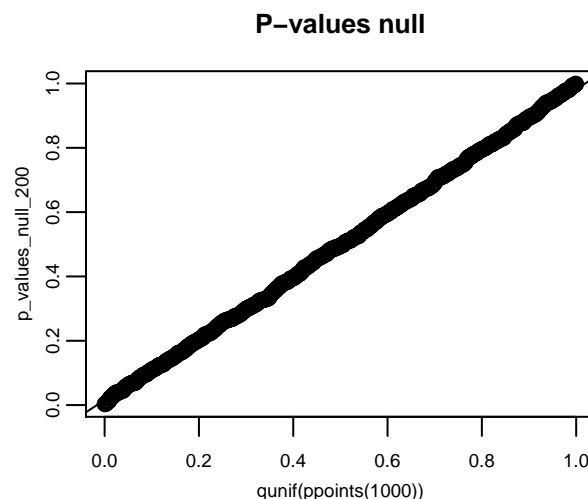
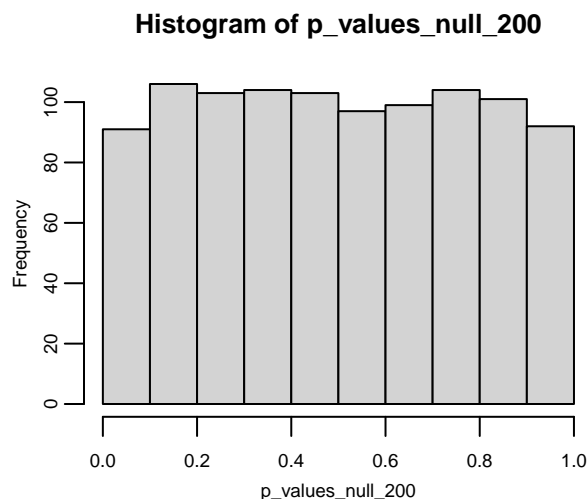
```
hist(p_values_null_200, cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
```

```
## Q-Q plot for the p-value data against true theoretical distribution:
```

```
qqplot(qunif(ppoints(1000)), p_values_null_200,
```

```
main = "P-values null", cex.main = 0.8, cex.lab = 0.6, cex.axis = 0.6)
```

```
qqline(p_values_null_200, distribution = function(p) qunif(p))
```



The histogram plot is closer to the uniform distribution.

```
k=1000
```

```
y = sum(p_values_null_200 < 0.05)
```

```
p_hat = y/k
```

```
int_200 = p_hat + c(-1.96, 1.96)*(sqrt((p_hat*(1-p_hat))/k))
```

```
int_200
```

```
## [1] 0.02870985 0.05329015
```

```
int - int_200
```

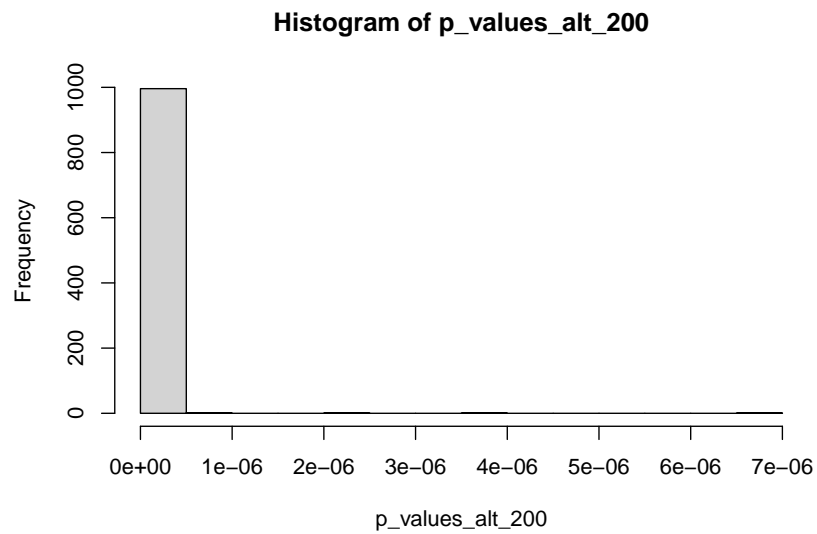
```
## [1] 0.6944517 0.7235483
```

It is more probable that the type I error is smaller for $n = 200$, since the boundary values are smaller than they were for $n = 20$.

```
samples_exp_alt_200 = matrix(rexp(1000 * n, rate = 3), n)
```

```
p_values_alt_200 = apply(samples_exp_alt_200, 2, get_prob, shape_param=n, scale_param = 1/5)
```

```
hist(p_values_alt_200)
```



```
k=1000
y = sum(p_values_alt_200 > 0.05)
p_hat = y/k
1 - (p_hat + c(1.96, -1.96)*(sqrt((p_hat*(1-p_hat))/k)))
```

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

```
# Theoretical power of the test
```

```
1 - pgamma(qgamma(0.95, shape = 200, scale = 1/5), shape = 200, scale = 1/3)
```

```
## [1] 0.9999999
```