Report 1

Klaudia Balcer

7 10 2021

Contents

Task 1

In the first task, we will study the properties of the parameter estimators of a particular form of the beta distribution $Beta(\alpha+1, 1)$. First, some theoretical calculations will be provided. Then, the estimation will be made on simulated data. The goal of the task is to conclude the differences between the MLE and moment estimators for different sample sizes.

Theoretical Calculations

MLE

To provide the Maximum Likelihood Estimator (MLE), we need to find the argument maximizing the likelihood function of the random sample vector X.

Deriving the MLE

The distribution $Beta(\alpha+1, 1)$ has the PDF: $f(x, \alpha) = (\alpha+1)x^{\alpha}$.

We have a random sample: $X = X_1, \dots, X_n$.

The likelihood function of the random sample vector X is given by the formula:

$$L(X, \alpha) = \prod_{i=1}^{n} f(X_i, \alpha) = \prod_{i=1}^{n} (\alpha+1)x_i^{\alpha}$$

Finding the argument maximizing the likelihood function is equivalent to finding the argument maximizing the logarithm of the likelihood function (loglikelihood function).

The loglikelihood function of the random sample vector X is given by the formula:

$$l(X, \alpha) = log(L(X, \alpha)) = log(\prod_{i=1}^{n} (\alpha+1)x_i^{\alpha}) = nlog(\alpha+1) + \alpha \sum_{i=1}^{n} log(x_i)$$

To find the max-arg, we need to look at the first and second derivatives of the loglikelihood function. We need to find the zeros from the first derivative and check if the second derivative is negative in those points.

$$\frac{\partial l(X, \alpha)}{\partial \alpha} = \frac{n}{\alpha+1} + \sum_{i=1}^{n} log(x_i)$$

Let's look for the zeros:

$$\frac{\partial l(X, \alpha)}{\partial \alpha} = 0 \iff \frac{n}{\alpha+1} + \sum_{i=1}^{n} log(x_i) = 0$$

So, the point considered to be the extremum is $\alpha = -\frac{n}{\sum_{i=1}^{n} log(x_i)} - 1$

Let's look at the second deriver:

$$\frac{\partial^2 l(X, \alpha)}{\partial \alpha^2} = -\frac{n}{(\alpha+1)^2}$$

It's always negative, so we have the MLE:

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

Fisher Information

We can calculate the Fisher Information using the formula:

$$I(\alpha) = -\mathbb{E}(\frac{\partial^2 f(x,\alpha)}{\partial\alpha^2})$$

For the $Beta(\alpha+1,1)$ distribution, the second derivative of the PDF does not depend on X:

$$\frac{\partial^2 f(x,\alpha)}{\partial\alpha^2} = -\frac{1}{(\alpha+1)^2}$$

So, the Fisher Information is equal:

$$I(\alpha) = -\mathbb{E}\frac{\partial^2 f(x,\alpha)}{\partial\alpha^2} = \frac{1}{(\alpha+1)^2}$$

MLE distribution

Using the Theorem 6.2.2. from Hogg, McKean, Craig *Introduction to Mathematical Statstics*, we can find the asymptotical distribution of $\hat{\alpha}_{MLE}$:

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{1}{I(\alpha)}\Big)$$

In our case:

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 36)$$

MSE

$$MSE = \mathbb{E}[\hat{\alpha} - \alpha]^2$$

as $\hat{\alpha} - \alpha$ is asymptotically normal with mean 0, we can state, that:

$$MSE \approx Var(\mathbb{Z})$$

where $\sqrt{n}Z \sim \mathcal{N}\Big(0, \frac{1}{I(\alpha)}\Big)$, so:

$$Z \sim \mathcal{N}\Big(0, \frac{1}{nI(\alpha)}\Big)$$

$$MSE \approx \frac{1}{nI(\alpha)}$$

In our case:

$$MSE_{20} = 36/20 = 1.8$$
$$MSE_{200} = 36/200 = 0.18$$

Moment estimator

The idea behind the moment estimator is to provide a formula for the parameter using the moments of the distribution. Then, we need to estimate the moments and substitute the estimators for real moments in that formula.

Deriving the moment estimator

First, let's calculate the first moment of the distribution $Beta(\alpha+1,1)$:

$$\mathbb{E}X_1 = \int_0^1 (\alpha+1)x^\alpha \cdot x dx = \frac{\alpha+1}{\alpha+2}x^{\alpha+2}|_{x=0}^{x=1} = \frac{\alpha+1}{\alpha+2}$$

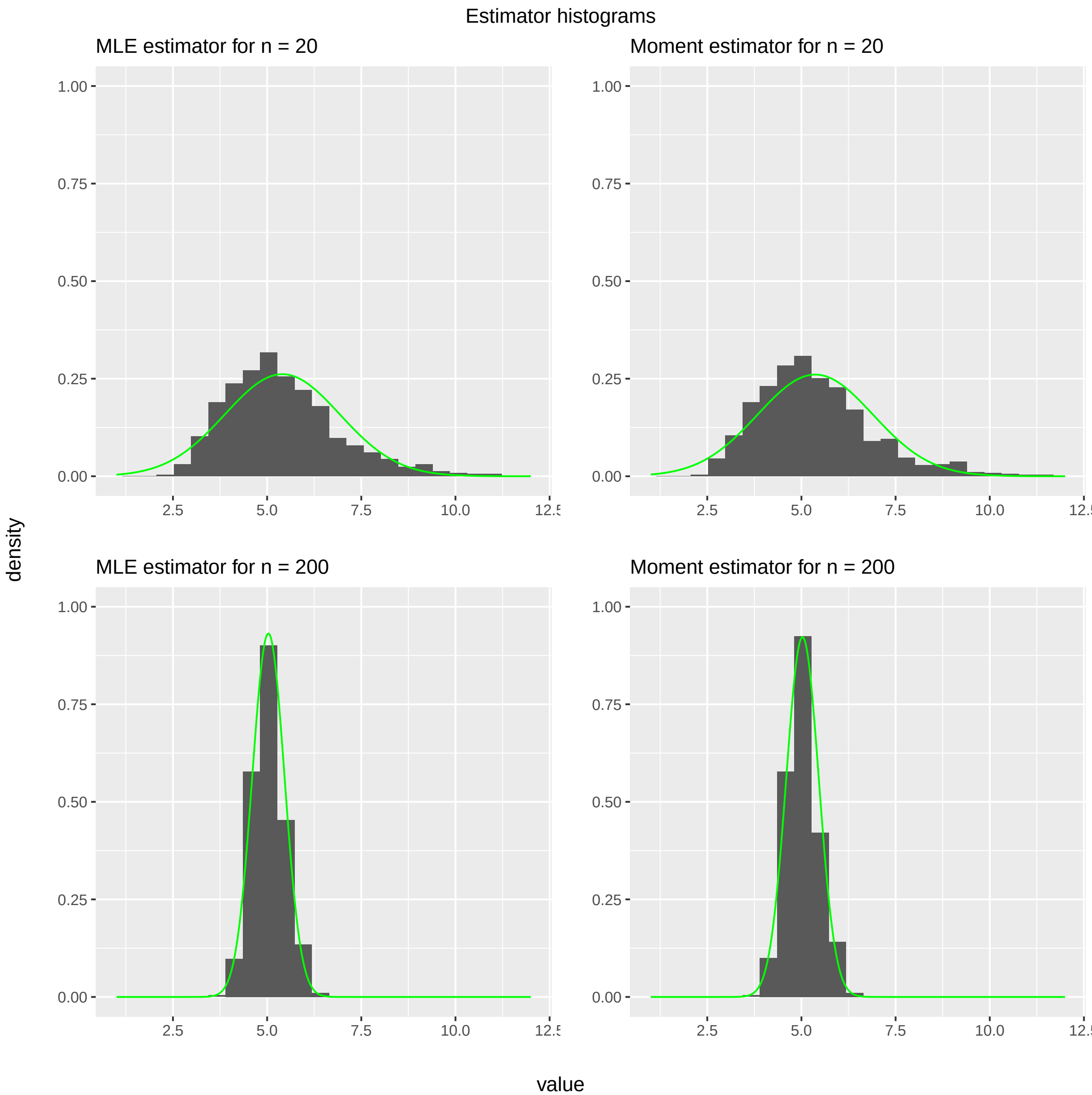
Let us denote $u_1 = \mathbb{E}X_1$. Than $u_1 = \frac{\alpha+1}{\alpha+2}$. This leads us to the formula:

$$\hat{\alpha} = \frac{1-2\hat{u_1}}{\hat{u_1}-1}$$

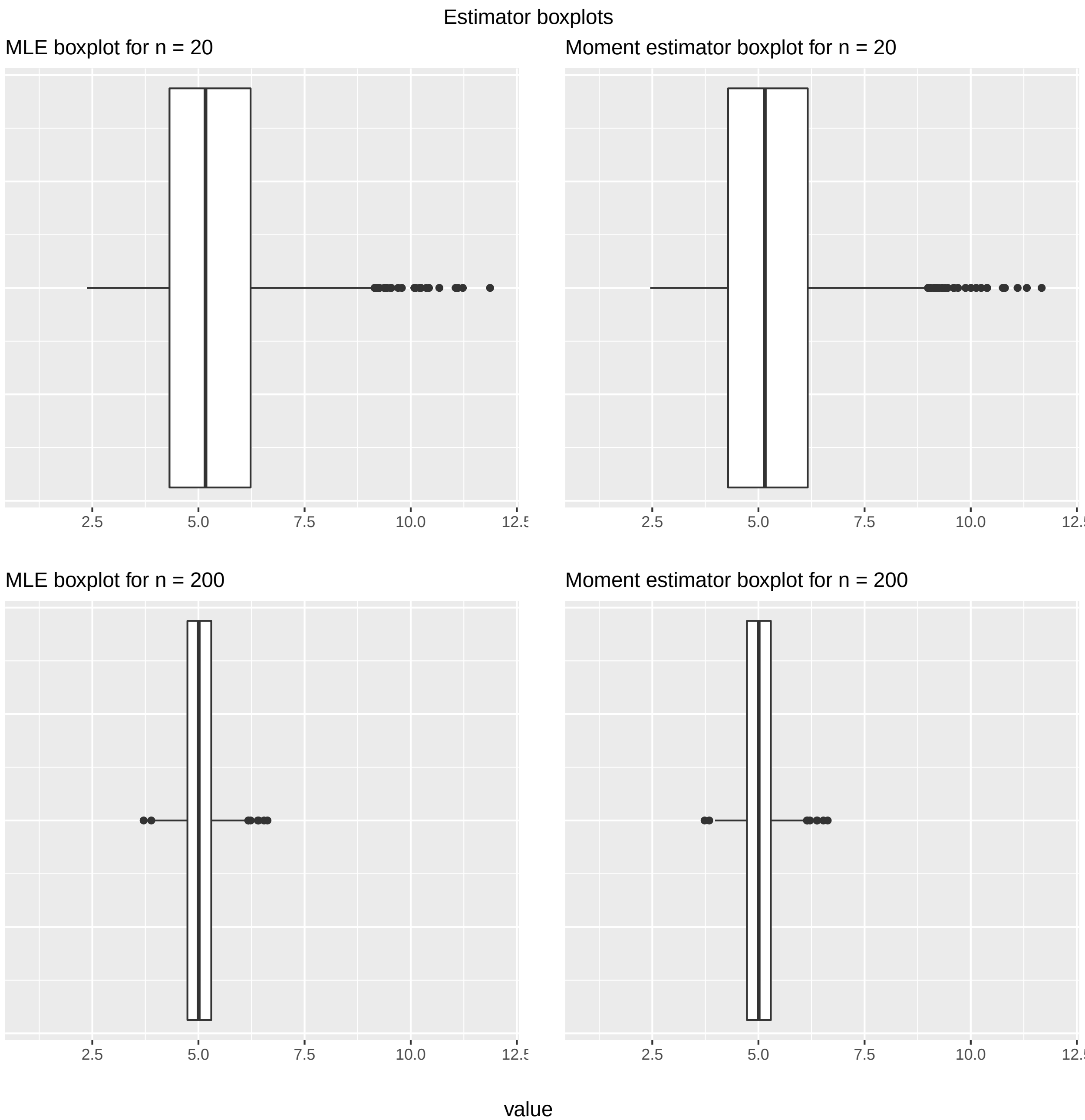
Simulations

We were estimating the parameter using two sample sizes: $n = 20, 200$. For each sample size, $m = 1000$ random sample vectors were provided. Each vector of observations leads us to estimations. Using m vectors, we can explore the distribution of the estimators.

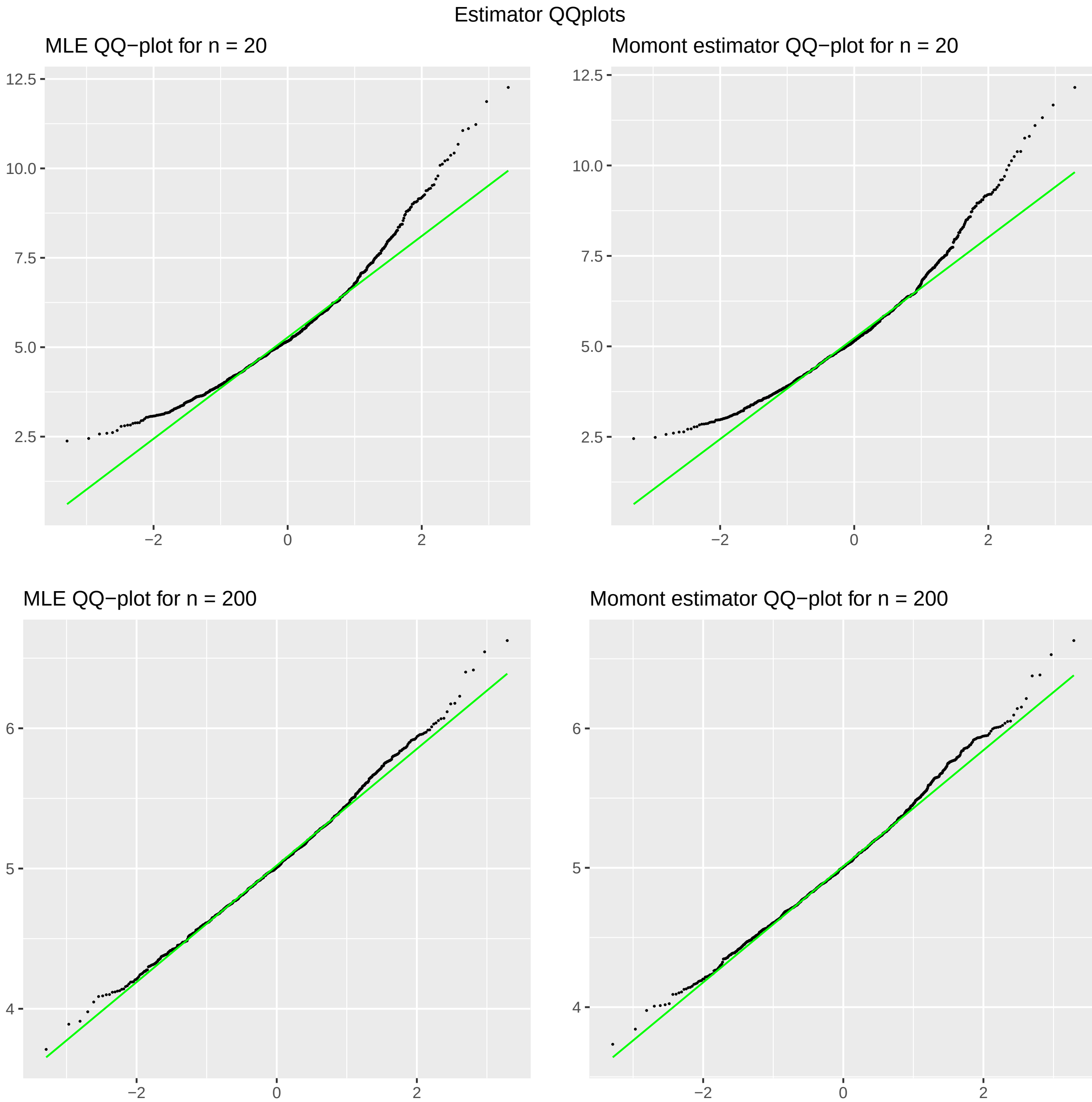
Histograms



Box plots


Estimator boxplots
MLE boxplot for n = 20
Moment estimator boxplot for n = 20
MLE boxplot for n = 200
Moment estimator boxplot for n = 200
2.5
5.0
7.5
10.0
12.5
value

Q-Q plots



Estimator QQplots
MLE QQ−plot for n = 20
Momont estimator QQ−plot for n = 20
MLE QQ−plot for n = 200
Momont estimator QQ−plot for n = 200

Bias

Both estimators are unbiased. We expect the bias to converge to zero.

For n = 20:

Estimated value of bias for MLE: 0.404

Confidence intervals for MLE: (0.31, 0.499)

Estimated value of bias for moment estimator: 0.368

Confidence intervals for moment estimator: (0.273, 0.463)

For n = 200:

Estimated value of bias for MLE: 0.033

Confidence intervals for MLE: (0.007, 0.06)

Estimated value of bias for moment estimator: 0.029

Confidence intervals for moment estimator: (0.002, 0.056)

Variance

From the Rao-Cramer Lower Bound, we expect the variance to be grater than the inverse of n times Fisher Information for real estimator value:

$$Var(\hat{\alpha}) \geqslant \frac{1}{n \cdot I(\alpha)}$$

For n = 20:

Estimated value of variance for MLE: 2.327

Confidence intervals for MLE: (2.136, 2.546)

Estimated value of variance for moment estimator: 2.346

Confidence intervals for moment estimator: (2.153, 2.566)

For n = 200:

Estimated value of variance for MLE: 0.183

Confidence intervals for MLE: (0.168, 0.201)

Estimated value of variance for moment estimator: 0.187

Confidence intervals for moment estimator: (0.172, 0.205)

MSE

We can estimate the MSE as $\sum_{i=1}^{n}(\hat{\alpha} - \alpha)^2$ (using the biased estimator of variance).

For n = 20:

Estimated value of MSE for MLE: 2.489

Confidence intervals for MLE: (2.179, 2.799)

Estimated value of MSE for moment estimator: 2.478

Confidence intervals for moment estimator: (2.174, 2.783)

For n = 20:

Estimated value of MSE for MLE: 0.184

Confidence intervals for MLE: (0.167, 0.201)

Estimated value of MSE for moment estimator: 0.188

Confidence intervals for moment estimator: (0.171, 0.205)

Conclusions

As we would expect, the bias converges to zero with n. For the bigger sample size, both estimators have much better results.

Both MLE and moment estimator have very similar results. For both sample sizes, the MLE has slightly greater bias and smaller variance. To combine those insights, we can look at the MSE. As we can see, the MLE performs a little better. Both estimators have MSE greater then expected for n=20 and MSE rougly equal to expected value for n=200.

Resuming: for most cases, MLE is the best choice. If we want to ensure the smallest bias, we can decide to use the moment estimator. Both estimator behave very similar, so the moment estimator is a good choice, when the MLE can by analitycaly derived.

Task 2

In the second task, we will focus on the problem of testing a simple hypothesis for the rate parameter of the exponential distribution. We will test the null hypothesis

$$H_0: \quad \lambda = 5$$

against the alternative

$$H_1: \quad \lambda = 3$$

on the significance level $\alpha = 0.05$.

We will consider rejecting the null hypothesis both when it is actually true and not. Rejecting the null hypothesis when it is true is the Type I Error. The probability of Type I Error should be lower than the significance level of the test α. The probability of rejecting the null hypothesis when it is false is called the power of the test. The test (on the given significance level α) is the better, the greater is the value of power.

First, some theoretical caculations will be presented. Then, we will run some simulations.

Theoretical Calculations

Test Definition

Let us start from finding the uniformly most powerfull test. We will use the formula from Neyman - Person Theorem.

PDF of the exponential distribution with rate parameter λ:

$$f(x, \lambda) = \lambda e^{-\lambda x}$$

To find the most powerfull test we will look at the likelihood ratio:

$$\frac{\prod_{i=1}^n f(x,3)}{\prod_{i=1}^n f(x,5)} = \frac{3^n e^{-3\sum_{i=1}^n x_i}}{5^n e^{-5\sum_{i=1}^n x_i}} = \left(\frac{3}{5}\right)^n e^{2\sum_{i=1}^n x_i}$$

We will recejt the null hypothesis, when the likelihood will be greater than some number k:

$$\left(\frac{3}{5}\right)^n e^{2\sum_{i=1}^n x_i} > k$$

Equivalenty:

$$\sum_{i=1}^n X_i > k'$$

The value k' (k) is the border of the critical region. We can find it using below equality:

$$P_{H_0}\left(\sum_{i=1}^n X_i > k'\right) = \alpha$$

So, the critical value of the test Satistics $T(X) := \sum_{i=1}^n X_i$ is:

$$k' = F^{-1}_{Gamma(n,5)}(1-\alpha)$$

We will consider $n = 20, 200$. For this sample sizes, the critical value is roughly equal:

$$k'_{20} \approx 5.576$$
$$k'_{200} \approx 44.763$$

Summarising, we have the test function

$$\varphi(X) = \mathbb{1}_{(k',\infty)}(T(X))$$

p-value

Definition of p-value:

$$p = P_{H_0}(T > T(X)) = 1 - F_{Gamma(n,5)}(T(X))$$

where T is a random variable from the same distribution as $T(X)$.

As the distribution is continuous, the test is of size α. This means, that the p-value (the probability of type 1 error) is equal to α.

$$p = P_{H_0}(T > T(X)) = 1 - F_{Gamma(n,5)}(T(X))$$

p-value distribution:

When the null hypothesis is true, the p-value is distributed uniformly under H_0.

Proof:

Let $T \sim Gamma(n, 5)$.

$$P\Bigg(P\Big(T > T(X)\Big) < a\Bigg) =$$

$$P\Big(1 - F_{Gamma(n,5)}(T(X)) < a\Big) =$$

$$P\Big(F_{Gamma(n,5)}(T(X)) \geq 1 - a\Big) =$$

$$P\Big(T(X) \geq F^{-1}_{Gamma(n,5)}(1 - a)\Big) =$$
$$1 - F_{Gamma(n,5)}(F^{-1}_{Gamma(n,5)}(1 - a)) =$$
$$1 - (1 - a) = a$$

We have shown that $P_{H_0}(p < a) = a$, so the p-value is uniformally distributed under H_0.

Power

The power is the probablity of receting the null hypothesis, when it is in fact false:

$$P_{H_1}\Big(\sum_{i=1}^{n} X_i > k'\Big) = P_{H_1}\Big(\sum_{i=1}^{n} X_i > F^{-1}_{Gamma(n,5)}(0.95)\Big) = 1 - F_{Gamma(n,3)}(F^{-1}_{Gamma(n,5)}(0.95))$$

For our ns, the powers are rougly equal:

$$\gamma_{20} \approx 0.764$$
$$\gamma_{200} \approx 1.0$$

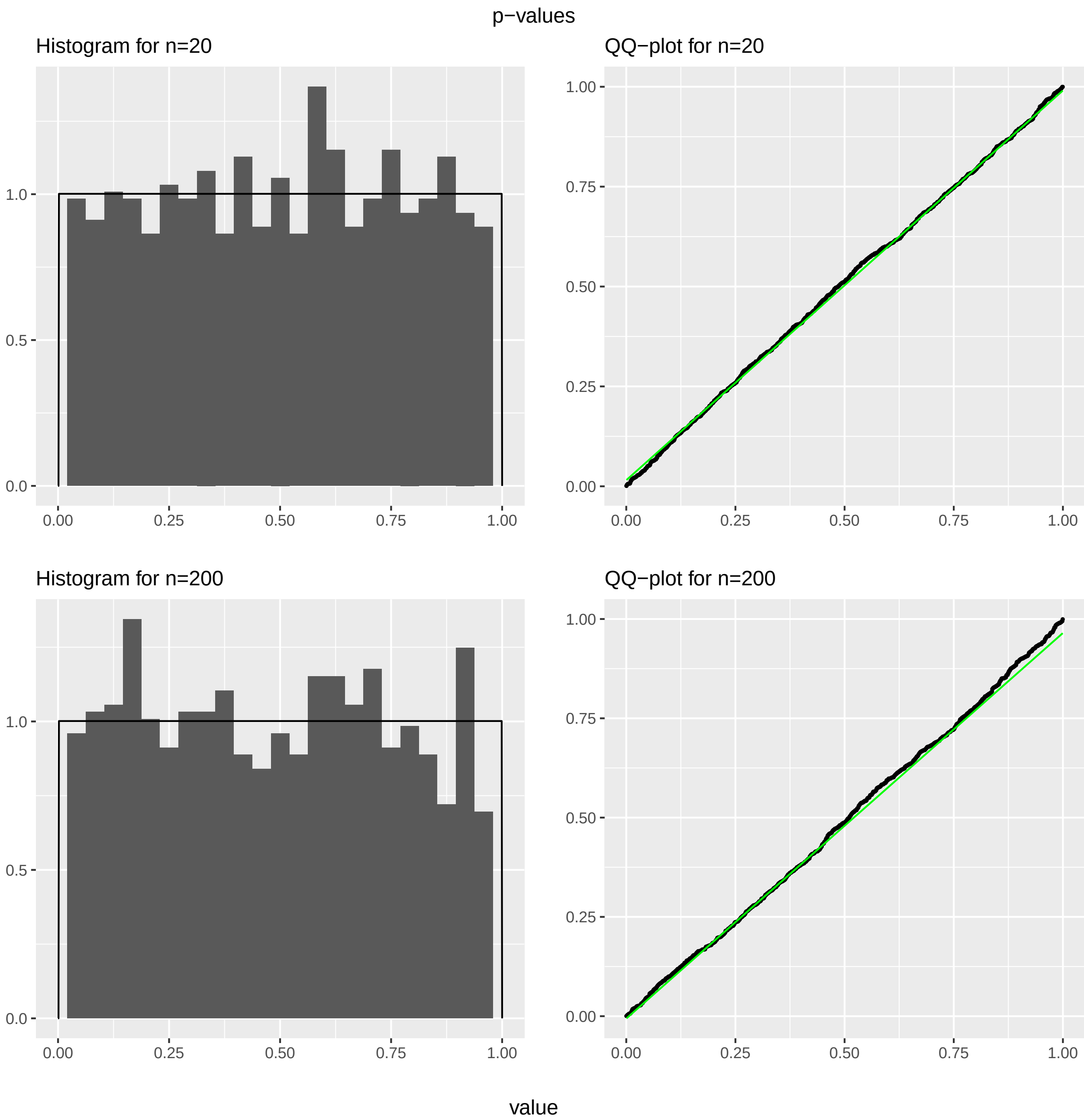
Simulations

We will look at the probability of rejecting the null hypothesis when the data is simulated from H_0 and H_1 distributions.

p-values

Data generated from H_0 distribution.

Distribution Below plots show the histogram and QQ plots of p-values.



Confidence Intervals **For n = 20:**

Estimated value of Type I Error: 0.048

Confidence intervals: (0.035, 0.061)

For n = 200:

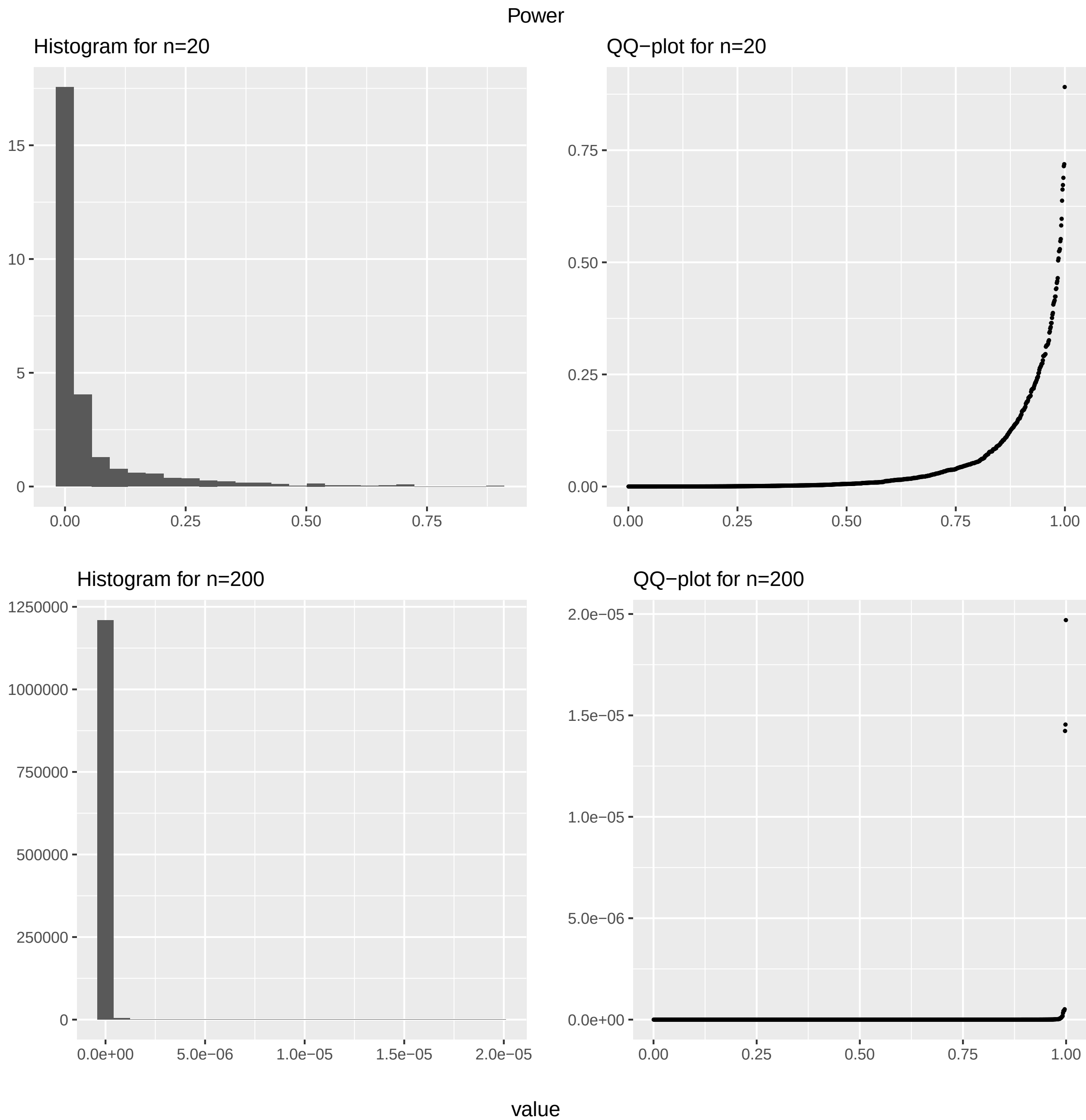
Estimated value of Type I Error: 0.052

Confidence intervals: (0.038, 0.066)

Power

Data generated from H_1 distribution.

Distribution Below plots show the histogram and QQ plots of power.



Confidence Intervals **For n = 20:**

Estimated value of power: 0.785

Confidence intervals: (0.76, 0.81)

For n = 200:

Estimated value of power: 1

Confidence intervals: (1, 1)

Conclusions

The results of the simulations match the theoretical expectations:

- the p-values are uniformally distributed under H_0;
- the p-values under H_0 are close to α;
- the power is not uniformly distributed under H_1, the mass is concentrated around 0;
- the value of power is close to the theoretically derived values.
- the bigger sample is considered, the better tests one can make.