Report 1

Klaudia Balcer

7 10 2021

Task 1

In this excersize, we are working with the distribution $Beta(\alpha+1,1)$. We will focus on estimating the parameter α. First we will do some theoretical calculations, than some simulations will be executed.

theoretical calculations

MLE

Deriving the MLE

pdf: $f(x,\alpha)=(\alpha+1)x^{\alpha}$.

random sample: $X=X_1,\dots,X_n$

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

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

We are looking for the maximum of the likelihood function, equivalently the maximum of the loglikelihood function.

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

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

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:

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

MLE distribution

Fisher Information

We can calculate the Fisher Information from:

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

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

The second derivative is contans, so the constant value is the mean:

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

Moment estimator

Deriving the moment estimator

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

Let's use $u_1 = \mathbb{E}X_1$, than $u_1 = \frac{\alpha+1}{\alpha+2}$, what leads us to $\hat{\alpha} = \frac{1-2\hat{u_1}}{\hat{u_1}-1}$

Diagnosing estimators

TODO: describe the Fisher Info connection with Var and MSE (Cromer Rao band)

Simulations

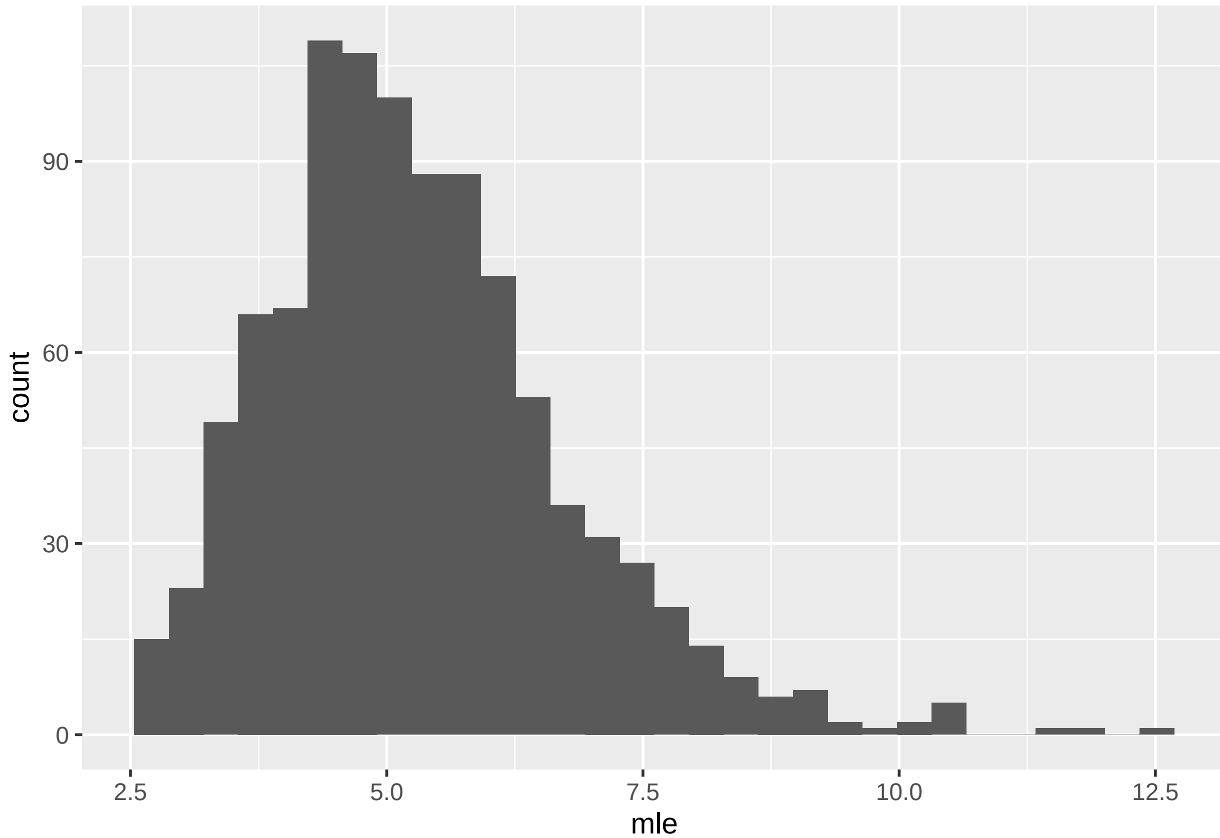
```
sim_1_20 <- simulation_1(5, 20,  1000)
sim_1_200 <- simulation_1(5, 200,  1000)
```

for both estimators:

Histogram

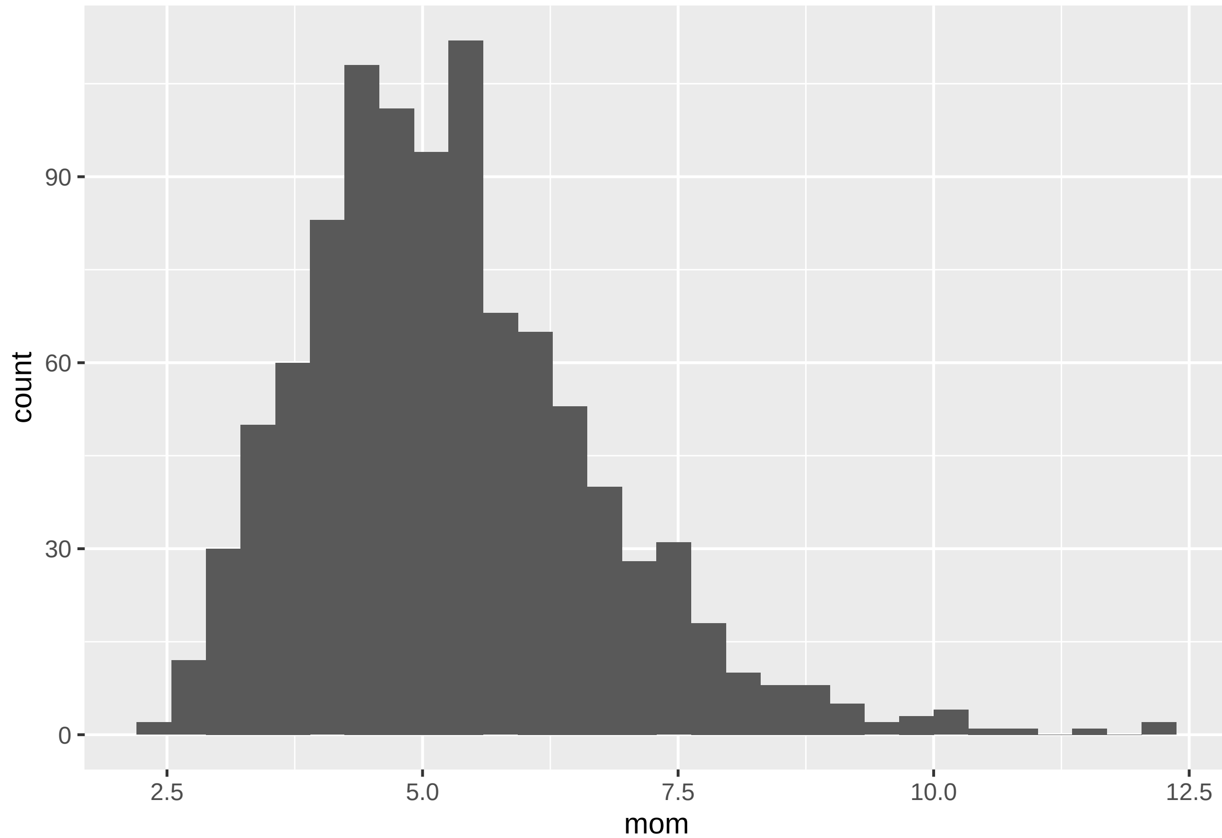
```
par(mfrow=c(2,2))
estimators_20 <- as.data.frame.array(t(sim_1_20[,  1, ]))
colnames(estimators_20) <- c("mle",  "mom")
estimators_200 <- as.data.frame.array(t(sim_1_200[,  1, ]))
colnames(estimators_200) <- c("mle",  "mom")

ggplot(estimators_20,  aes(x=mle)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



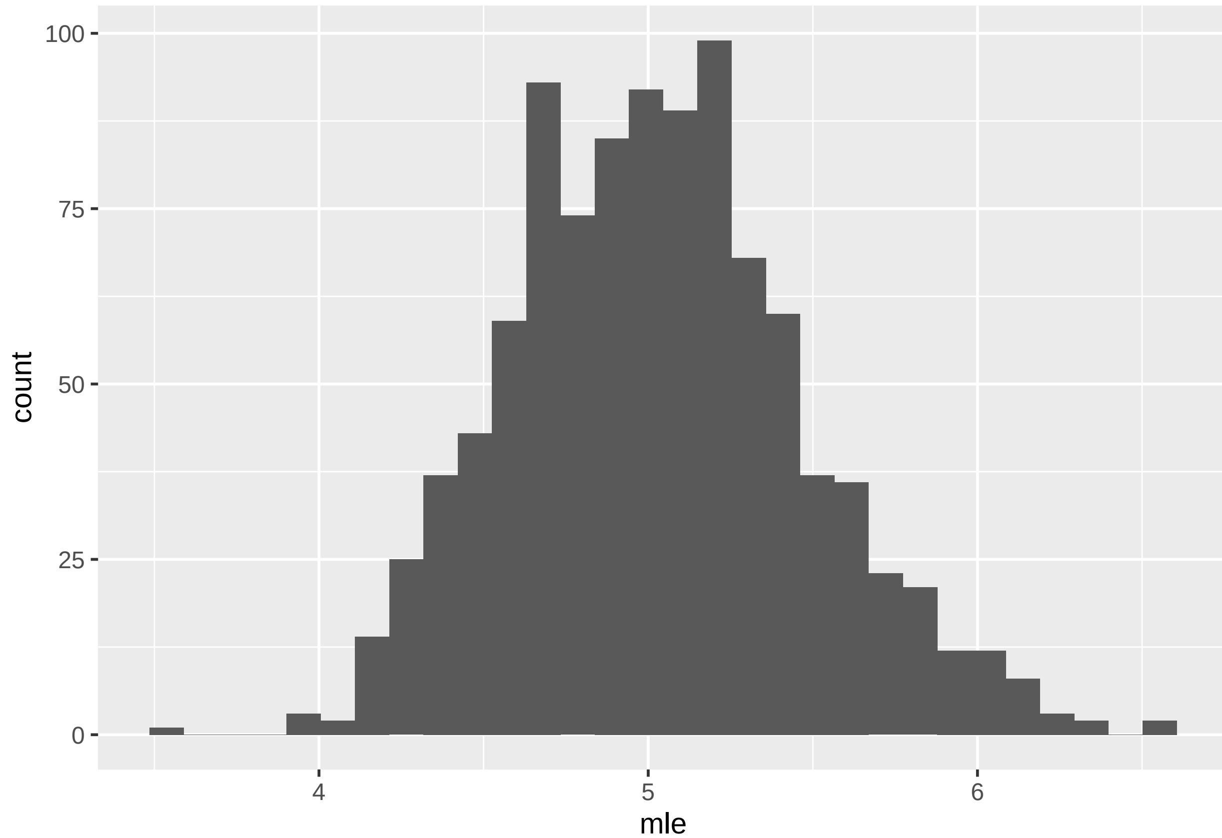

```r
ggplot(estimators_20,  aes(x=mom)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



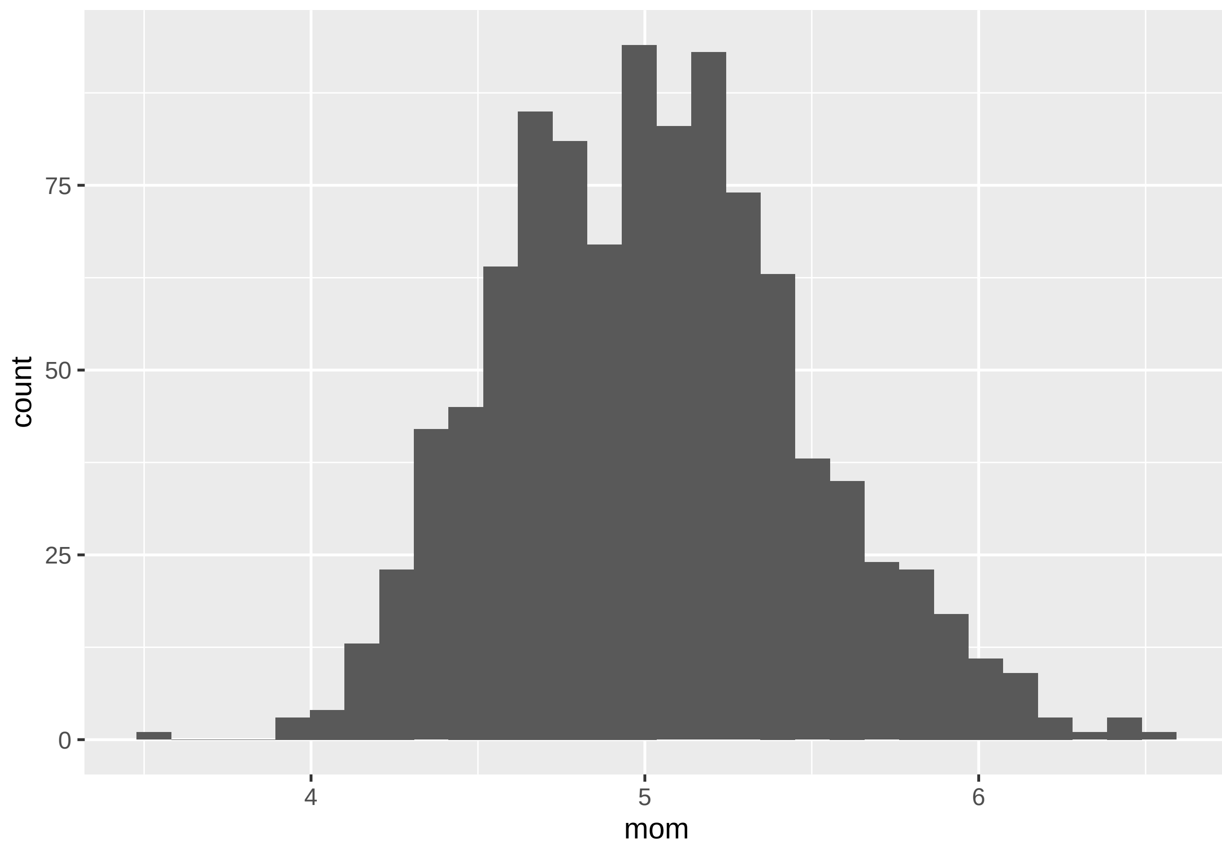

```r
ggplot(estimators_200,  aes(x=mle)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
ggplot(estimators_200,  aes(x=mom)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Box plot

Q-Q plot

Bias + confidence intervals **Variance** + confidence intervals

MSE + confidence intervals

n = 200

Conclusions

Task 2