First report

Klaudia Puchała

23 October 2020

Part One

Let $X_1, X_2, \ldots, X_n$ be the simple random sample from the distribution with the density $f(x;\alpha) = (\alpha+1)x^{\alpha}$, for $x \in (0;1),\ \alpha > 0$, then:

a) The maximum likelihood estimator $\hat{\alpha}_{MLE}$ of the parameter α is given by the formula:

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

It comes from the calculations:

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

$$l(x,\alpha) = \log\left((\alpha+1)^n \prod_{i=1}^{n} x_i^{\alpha}\right) = n\log(\alpha+1) + \sum_{i=1}^{n}\log(x_i^{\alpha}) = n\log(\alpha+1) + \alpha\sum_{i=1}^{n}\log(x_i)$$

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha+1} + \sum_{i=1}^{n}\log(x_i) = 0 \quad \Rightarrow \quad \alpha = -\frac{n}{\sum_{i=1}^{n}\log(x_i)} - 1$$

$$\frac{\partial^2 l}{\partial \alpha^2} = -\frac{n}{(\alpha+1)^2} < 0 \text{ for all } n > 0$$

From this it follows:

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

b) The Fisher Information of maximum likelihood estimator is:

$$\mathrm{I}(\alpha) = \frac{1}{(\alpha+1)^2}$$

This can be calculated from the formula:

$$\mathrm{I}(\theta) = \mathrm{Var}\left(\frac{\partial \log f(x,\theta)}{\partial \theta}\right) = \mathrm{E}\left(\frac{\partial \log f(x,\theta)}{\partial \theta}\right)^2 = -\,\mathrm{E}\left(\frac{\partial^2 \log f(x,\theta)}{\partial \theta^2}\right)$$

We use the last equality, then:

$$\log f(x,\alpha) = \log\left((\alpha+1)x^{\alpha}\right) = \log(\alpha+1) + \alpha\log x$$

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

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

$$\mathrm{I}(\theta) = -\,\mathrm{E}\left(-(\alpha+1)^{-2}\right) = (\alpha+1)^{-2}$$

Asymptotic distribution of maximum likelihood estimator is:

$$\sqrt{n}\,(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, (\alpha+1)^2\right)$$

It follows from the formula

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathrm{D}} \mathrm{N}\left(0, \frac{1}{\mathrm{I}(\theta)}\right)$$

Mean squared error of maximum likelihood estimator it's:

$$MSE(\hat{\alpha}) \approx \frac{(\alpha+1)^2}{n}$$

The formula is taken from

$$MSE(\hat{\alpha}) = \mathrm{E}\left((\hat{\alpha}-\alpha)^2\right) = \mathrm{E}(\hat{\alpha}^2 - 2\hat{\alpha}\alpha + \alpha^2) = \mathrm{E}(\hat{\alpha}^2) - 2\alpha\,\mathrm{E}(\hat{\alpha}) + \alpha^2$$

it's known that

$$\mathrm{Var}(\hat{\alpha}) = \mathrm{E}(\hat{\alpha}^2) - \mathrm{E}^2(\hat{\alpha}) \Rightarrow \mathrm{E}(\hat{\alpha}^2) = \mathrm{Var}(\hat{\alpha}) + \mathrm{E}^2(\hat{\alpha})$$

so:

$$MSE(\hat{\alpha}) \approx \mathrm{Var}(\hat{\alpha}) + \mathrm{E}^2(\hat{\alpha}) - 2\alpha\,\mathrm{E}(\hat{\alpha}) + \alpha^2$$

Form asymptotic distribution it's known that:

$$\hat{\alpha}_n \sim \mathrm{N}\left(\alpha, \frac{(\alpha+1)^2}{n}\right)$$

Therefore:

$$MSE(\hat{\alpha}) = \frac{(\alpha+1)^2}{n} + \alpha^2 - 2\alpha^2 + \alpha^2 = \frac{(\alpha+1)^2}{n}$$

c) The moment estimator $\hat{\alpha}_M$ is:

$$\hat{\alpha}_M = -2 - \frac{1}{\overline{X}-1} \text{ where } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

It's known that:

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

and equality must be satisfied

$$\mu_k = \mathrm{E}(X_1^k).$$

So:

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}\,\mathrm{E}(X_1) = \int_0^1 x(\alpha+1)x^{=} \int_0^1 (\alpha+1)x^{\alpha+1}dx = \frac{\alpha+1}{\alpha+2} = 1 - \frac{1}{\alpha+2}$$

Therefore

$$\overline{X} = 1 - \frac{1}{\alpha+2} \;\Rightarrow\; \hat{\alpha} = -2 - \frac{1}{\overline{X}-1}$$

e) After generating 20 and 200 random samples, we get the following estymators:

- for n=20

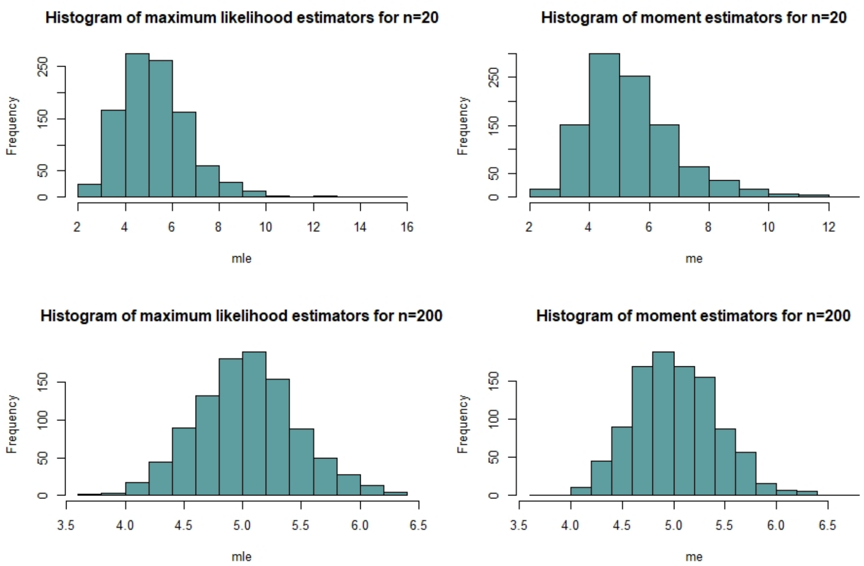
	mle	m
estymators	5.817264	5.805158
$\hat{\alpha} - \alpha$	0.8172641	0.8051576
$(\hat{\alpha} - \alpha)^2$	0.6679206	0.6482788

- for n=200

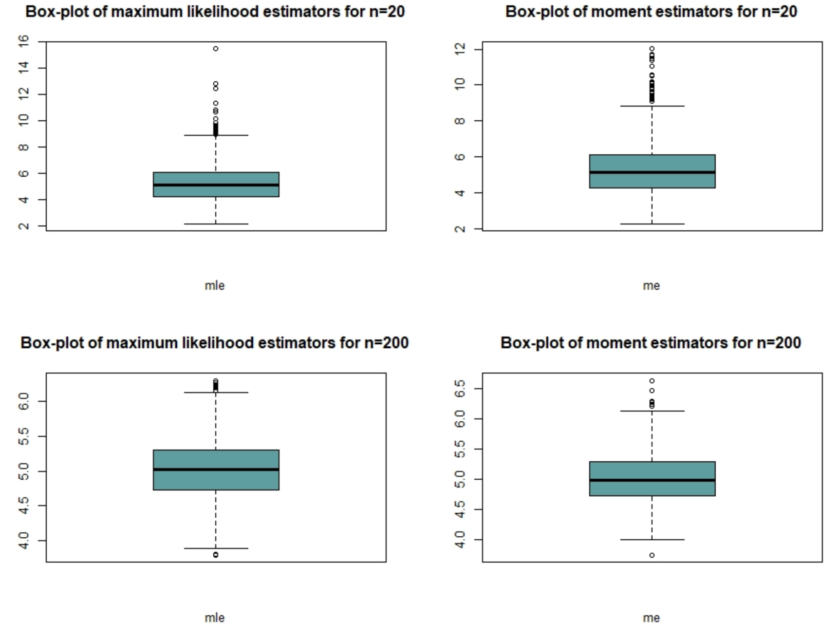
	mle	m
estymators	5.61764	5.720145
$\hat{\alpha} - \alpha$	0.6176398	0.720145
$(\hat{\alpha} - \alpha)^2$	0.3814789	0.5186088

For $n = 20$, the differences between the estimators are so small that it is difficult to draw any conclusions. However, for $n = 200$, it can be seen that the $\hat{\alpha}_{MLE}$ is better because it has a smaller bias.

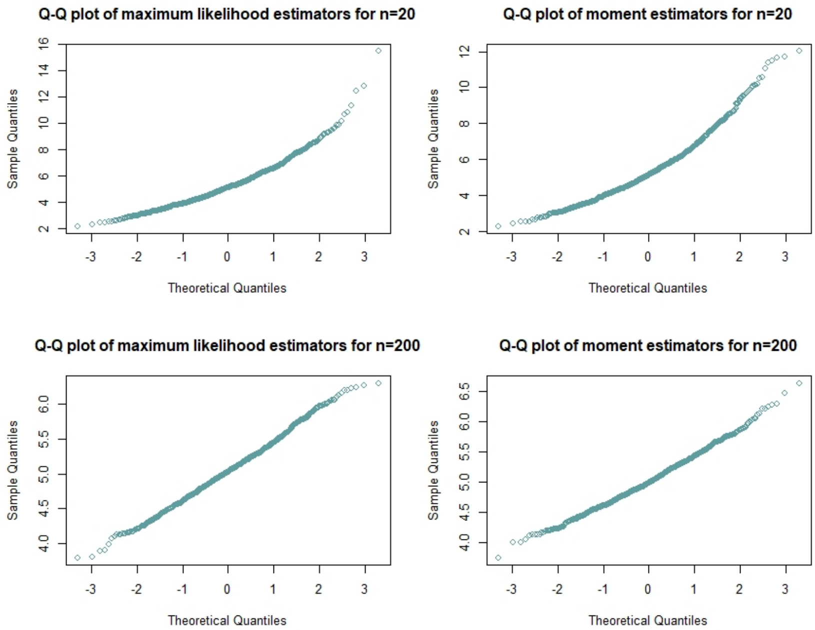
f) After generating 1000 $\hat{\alpha}_{MLE}$ and $\hat{\alpha}_M$ estimators 20 random samples, we get the following histograms, box-plots and q-q plots:



The histograms of both estimators are similar and the right tail is longer. Even so, these plots closely resemble a normal distribution.When n grows, the distribution of both converges to the normal distribution.



As n increases, the distribution of both estimators becomes more central. This is because of the convergence of the normal distribution which is central.



Q-Q plots also shows, that for bigger MLE and Me becom almost normal distributed.

for n=20

	mle	m
bias	0.3605339	0.2303039
CI of bias	0.09791296	0.09185486
variance	2.493157	2.194187
CI of variance	0.3021082	0.2857841
mean-squared error	2.623142	2.247227
CI of mean-squared error	0.3450464	0.3100766

for n=200

	mle	m
bias	0.03543171	0.0481983
CI of bias	0.02635594	0.02666668
variance	0.1806452	0.1849299
CI of variance	0.0154616	0.01821846
mean-squared error	0.1819006	0.187253
CI of mean-squared error	0.01586568	0.01888354

The above parameters are similar to each other, so the parameters can be from similar distributions.

Part Two

Let $X_1, X_2, \ldots, X_n$ be the simple random sample from the exponential distribution with the density $f(x,\theta) = \lambda e^{-\lambda x}$ 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 $H_1 : \lambda = 3$.

$$\frac{H_0}{H_1} = \frac{\prod_{i=1}^n 5e^{-5x_i}}{\prod_{i=1}^n 3e^{-3x_i}} = \frac{5}{3}e^{-2\sum_{i=1}^n x_i} \leqslant k$$

$$e^{-2\sum_{i=1}^n x_i} \leqslant \frac{3}{5}k$$

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

$$\sum_{i=1}^n x_i \geqslant -\frac{1}{2}\log\left(\frac{3}{5}k\right) = k^*$$

Thus, the test statistic of $\mathrm{T} = \sum_{i=1}^n x_i$ and critical area $\mathrm{C} = [k^*, \infty)$.

$$\mathrm{P}_{H_0}\left(T \geqslant k^*\right) = 0.05$$

To find this out, let's consider the distribution of T statistics. It's known that:

$$X_i \sim \exp(\lambda)$$

because the density function of the Gamma distribution $\Gamma(k,\theta)$ is given by the formula:

$$f(x) = \frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}}$$

it is easy to see that:

$$X_i \sim \Gamma\left(1, \frac{1}{\lambda}\right).$$

If Xi has a $\Gamma(k_i,\theta)$ distribution for $4i = 1, 2, ..., n$, then

$$\sum_{i=1}^n X_i \sim \Gamma\left(\sum_{i=1}^n k_i, \theta\right)$$

So our test statistic:

$$T \sim \Gamma\left(n, \frac{1}{\lambda}\right).$$

a) The formula for the critical value for this test is:

$$\mathrm{P}_{H_0}\left(T > k^*\right) = 1 - \mathrm{P}_{H_0}\left(T \leqslant k^*\right) = 0.05 \ \Rightarrow\ k^* = F^{-1}_{\Gamma\left(n,\frac{1}{5}\right)}(0.95).$$

b) The formula for the power of this test:

$$\mathrm{P}_{H_1}\left(T > k^*\right) = 1 - \mathrm{P}_{H_1}\left(T \leqslant k^*\right) = 1 - F_{\Gamma\left(n,\frac{1}{3}\right)}\left(F^{-1}_{\Gamma\left(n,\frac{1}{5}\right)}(0.95)\right)$$

c) The table of p-value for one random sample from H_0 of size n :

$n = 20$	$n = 200$
0.7916712	0.9930976

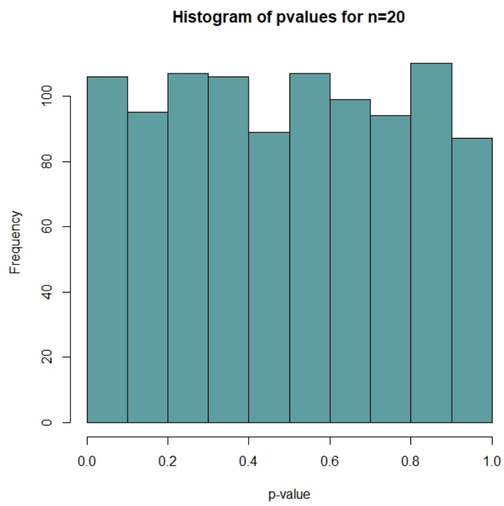
Based on these p-values, we have no grounds to reject H_0, which means that any deviation from the norm is most likely a coincidence.

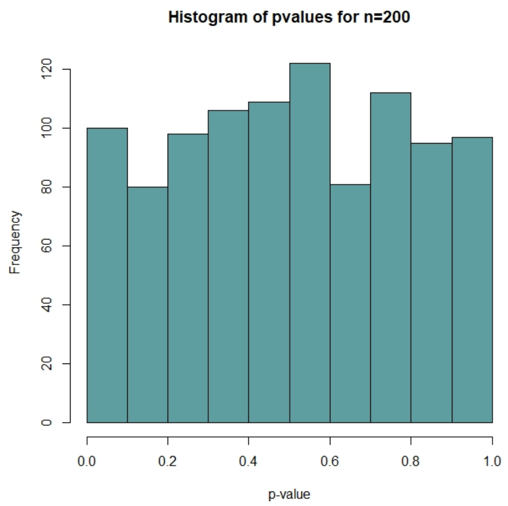
d) Under the null hypothesis, the test statistic T has the distribution $F(t)\,\mathrm{P}(T < t)$ (here Gamma distribution) for all t. Since $F(t)$ is invertible, we can derive the distribution of the random value of p as follows:

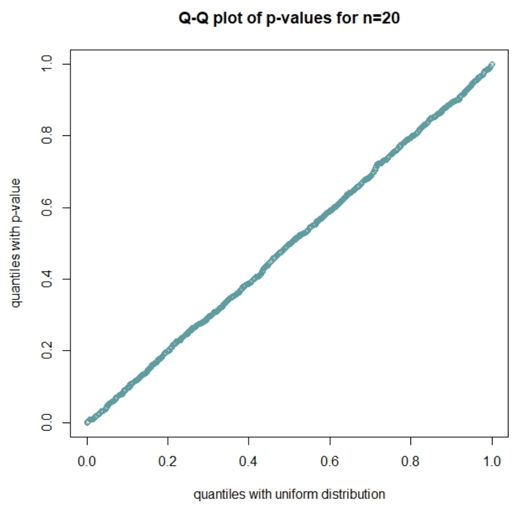
$$\mathrm{P}\left(F(T) < p\right) = \mathrm{P}\left(T < F^{-1}(p)\right) = F\left(F^{-1}(p)\right) = p$$

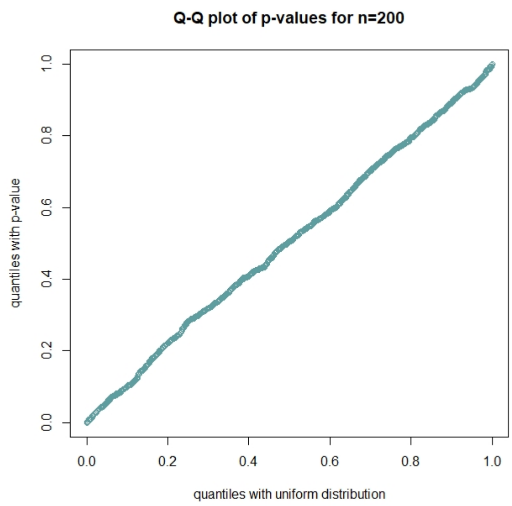
which shows that the distribution $F(t)$ is uniform on the interval [0,1].

e) After generating 1000 times random samples of size n, we get:









From the above graphs, we can conclude that the p-value is uniform distribution, regardless of n. This is in line with the theory in the lecture.

Let's use these simulations to construct the 95% confidence interval for the type I error of the test. We will use the formula for the confidence interval for fractions:

$$\mathrm{P}\left(\frac{m}{n} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}} < p < \frac{m}{n} + u_{-\frac{\alpha}{2}}\sqrt{\frac{\frac{m}{n}(1-\frac{m}{n})}{n}}\right) = 1-\alpha$$

where:

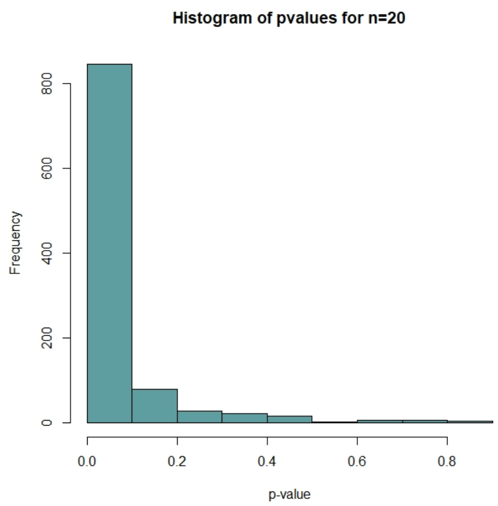
- m - number of favorable events (here, we reject the null hypothesis when it is true, i.e. when p-value < 0.05)
- n - number of all events
- $u_{1-\frac{\alpha}{2}}$ - is the value of the quantile $1-\frac{\alpha}{2}$ of the standardized normal distribution for the significance level α.

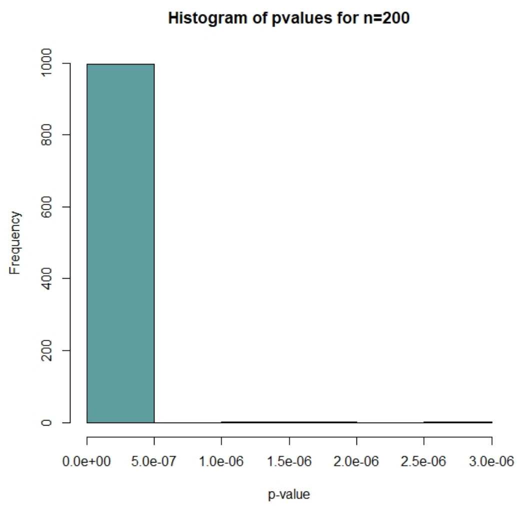
For our data will give the following results:

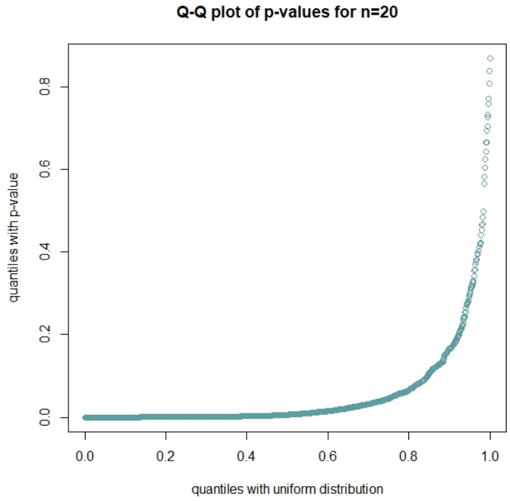
	percentage of favorable events ($\hat{\alpha}$)	confidence interval (+/-)
$n = 20$	0.054	0.01400846
$n = 200$	0.045	0.01284861

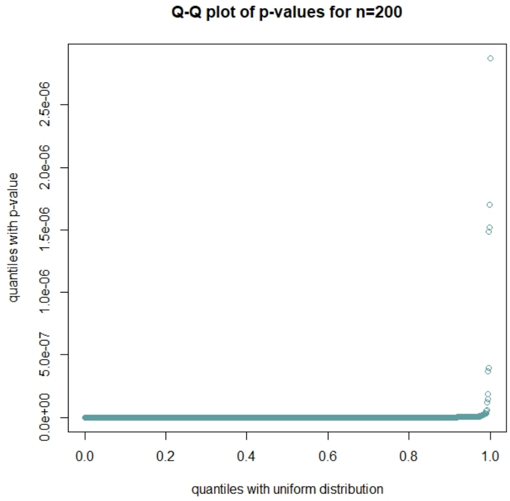
There is $\alpha = 0.05$ in both confidence intervals. The width of the confidence intervals does not depend on n.

f) We generate 1000 samples of size n from the alternative distribution and calculate p-values.









As n increases, p-values approach zero.

Let's use these simulations to construct a confidence interval of 95% for for the power of this test. For our data will give the following results:

	power	percentage of favorable events	confidence interval (+/-)
$n = 20$	0.7581204	0.759	0.02650803
$n = 200$	0.9999999	1.000	0.00000000

As n grows, the power also grows.