LSTAT2180: Resampling methods with applications

Bootstrap method evaluation

Lambda^2 estimator from a Weibull distribution

Lise Léonard 79802000, Mehrdaad Doraghi 55141800
STAT2M
12-14-2020

1 Introduction

In this project, we will study the performance of the different bootstrap methods seen in class on an estimator of a parameter of interest.

The iid. population studied $X_1...X_n$ follows $X_i \sim$ Weibull(λ, c) where λ is the scale of the distribution and c is its shape. The parameter c will be assumed to be known and to be equal to 2. Our parameter of interest is λ^2, also known as $\frac{1}{m_2}$ where m_2 is the second moment of the Weibull distribution for $c = 2$.

We will first try to find an appropriate estimator of λ^2 and study some of its properties.

Secondly, different bootstrap methods will be used to obtain a confidence interval in the case where $n = 30$. The performance of these methods will be then evaluated using a Monte-Carlo simulation.

Finally, we will compare the power of the different bootstrap methods with a sample of $n = 30$.

2 Estimator of λ^2

In this section, we will propose an estimator and show some of its properties as its real distribution. A bootstrap method will be used to estimate its bias and its variance.

The following estimator for λ^2 is found by using the Maximum Likelihood method (proof in the appendix) :

$$\hat{\lambda}^2 = \frac{n}{\Sigma_i x_i^2} \tag{1}$$

If X_i is an Weilbull, we have this exact pivotal root and confidence interval :

$$2\lambda^2 \Sigma_i x_i^2 \sim \chi^2(2n) \tag{2}$$

and

$$\lambda^2 \in \left[\frac{q_{\frac{\alpha}{2}}}{2\Sigma_i x_i^2}; \frac{q_{1-\frac{\alpha}{2}}}{2\Sigma_i x_i^2}\right] \tag{3}$$

where q_a is the a-quantile of a $\chi^2(2n)$.

As $\hat{\lambda}^2$ is a maximum likelihood estimator we can apply the following theorem to find its asymptotic distribution.

Theorem : Under some regularity assumptions the maximum likelihood estimator distribution follows an asymptotic normal distribution when n increases such that : $\sqrt{n}(\hat{\lambda}^2 - \lambda^2) \longrightarrow \mathcal{N}(0, I_1(\lambda^2)^{-1})$ where $I_1(\lambda^2)$ is the Fisher information.

$I_1(\lambda^2)$ is defined in the case where $c = 2$ as $m_4 - m_2^2$ where m_4 is the fourth moment of the Weibull distribution and m_2 its second moment (proof in the appendix). We can then write that the asymptotical Normal distribution of the estimator is :

$$\sqrt{n}(\hat{\lambda}^2 - \lambda^2) \longrightarrow \mathcal{N}\left(0, \frac{1}{m_4 - m_2^2}\right) \tag{4}$$

The following confidence interval is obtained :

$$\lambda^2 \in \left[\hat{\lambda}^2 - z_{1-\alpha/2}\frac{n^{-1/2}}{(m_4 - m_2^2)^{1/2}}; \hat{\lambda}^2 + z_{1-\alpha/2}\frac{n^{-1/2}}{(m_4 - m_2^2)^{1/2}}\right] \tag{5}$$

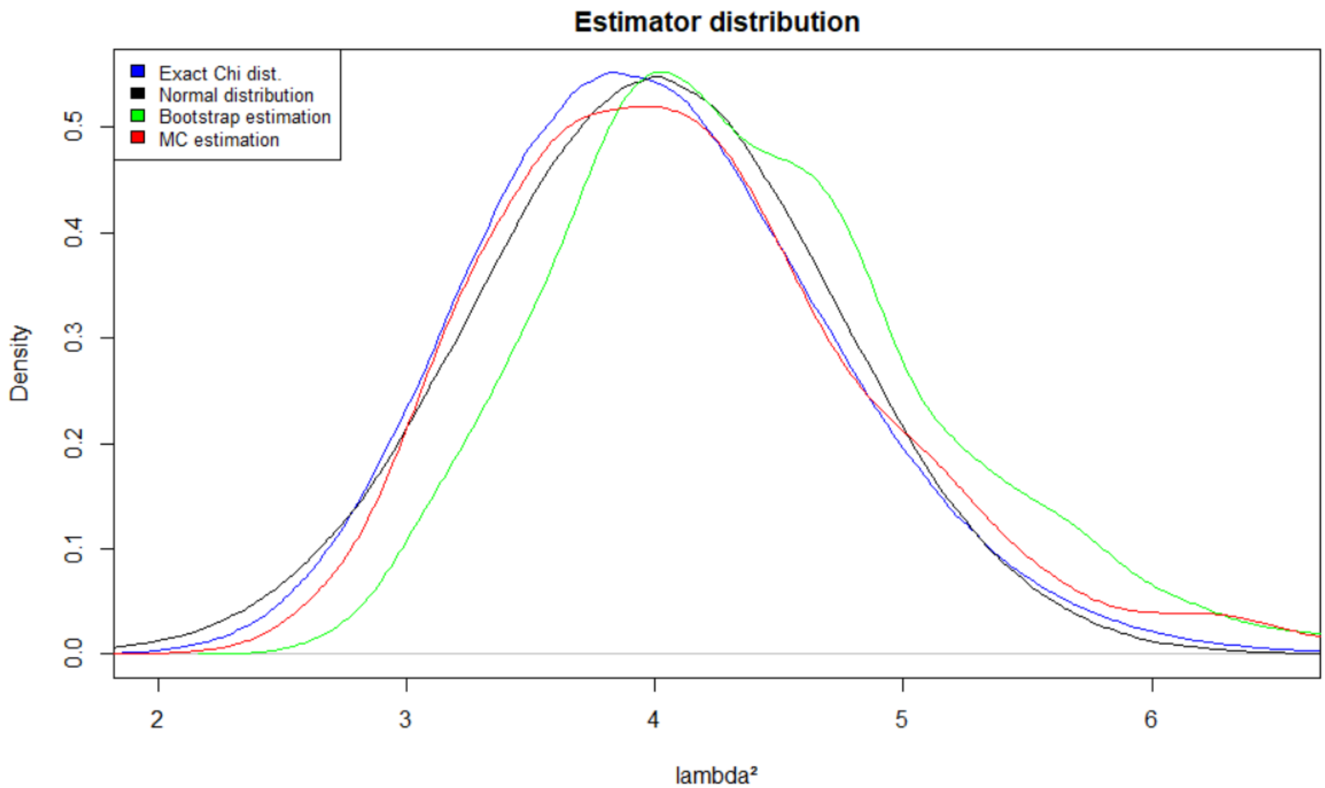

Figure 1: Estimator distribution with n=30

The Figure 1 illustrates the distribution of $\hat{\lambda}^2$ with respect to the equations 2 and 4 for the the exact Chi^2 distribution and the Normal distribution. The Monte-Carlo (MC) simulation uses the Weilbull distribution defined in the beginning with $n = 30$ while the bootstrap uses the sample with 30 observations generated in the beginning of our R code. The Normal and MC are closed to the exact distribution but the Bootstrap distribution is a bit shifted to the right.

Finally we will check the bias and variance of the estimator using a nonparametric bootstrap method with different values of B.

The bootstrap estimation of the bias is given by :

$$Bias^*[T] \approx \frac{1}{B}\sum_{b=1}^{B} T(\mathcal{X}^{*(b)}) - \hat{\lambda}^2 \tag{6}$$

The bootstrap estimation of the variance given by :

$$Var^*[T] \approx \frac{1}{B}\sum_{b=1}^{B} T^2(\mathcal{X}^{*(b)}) - \left(\frac{1}{B}\sum_{b=1}^{B} T(\mathcal{X}^{*(b)})\right)^2 \tag{7}$$

where $T(\mathcal{X})$ is our estimator computed with the sample $\mathcal{X}$.

Table 1 shows the result of the bias and the variance of $\hat{\lambda}^2$ obtained with a bootstrap method. The bias is stable around [0.135;0.153] after at least 100 resamplings. Its standard error is also quite stable around [0.74;0.80]. Increasing B, the number of resampling, decreases the error bound meaning we are close to the true bias. According to Effron's rule of thumb, we do not correct our estimator if the ratio bias over standard error is low ($\frac{Bias}{std} \leq 0.25$). In our case, with B=100, $\frac{Bias}{std} = 0.20$ meaning there is no need for any correction.

Table 1: Bias and variance estimation of $\hat{\lambda^2}$ using a bootstrap method

Value of B	Bias	Standard Error	Variance	error bound
10	0,0430	1,0555	1,1140	0,6675
100	0,1523	0,7443	0,5540	0,1489
1000	0,1430	0,7846	0,6155	0,0496
10000	0,1354	0,8021	0,6434	0,0160

3 Confidence intervals using bootstrap methods and performance evaluation of these methods

In this section, first we will present the different methods to obtain a confidence interval for our estimator presented in the previous section. Then, we will compare the intervals obtained. The third step of the analysis will be to compare the performance of these methods and see how behaves the coverage and the length of our interval when n increases.

Four different bootstrap methods will be used to construct our confidence interval namely the Basic bootstrap, the percentile bootstrap, the t-bootstrap and the iterated bootstrap.

The confidence interval for the basic bootstrap is given by $[2T(\mathcal{X}) - v^*_{1-\frac{\alpha}{2}}, 2T(\mathcal{X}) - v^*_{\frac{\alpha}{2}}]$ where v_a is the a-quantile of $T(\mathcal{X}^*)$. Concretely to compute the quantiles we order the B bootstrap estimations of λ^2 where B is the number of resampling and the $B(\frac{\alpha}{2})$ and the $B(1-\frac{\alpha}{2})$ ordered estimations are our quantiles. On the other side, the percentile bootstrap is defined as $[v^*_{1-\frac{\alpha}{2}}, v^*_{\frac{\alpha}{2}}]$ using the same quantiles as the basic bootstrap. To use the percentile bootstrap, we suppose that there is a monotonic transformation of λ^2 that changes its distribution to a symmetric distribution. In the previous section, on Figure 1, we observe that the distributions are almost symmetric and that we can assume that this condition is fulfilled.

The t-bootstrap confidence interval is given by $\left[T(\mathcal{X}) - \hat{\sigma}(\mathcal{X})n^{-\frac{1}{2}}y^*_{1-\frac{\alpha}{2}}, T(\mathcal{X}) - \hat{\sigma}(\mathcal{X})n^{-\frac{1}{2}}y^*_{\frac{\alpha}{2}}\right]$ where y^*_a is the a-quantile of $\sqrt{n}\frac{T(\mathcal{X}^*)-\hat{\lambda}^2}{\hat{\sigma}(\mathcal{X}^*)}$. $\hat{\sigma}(\mathcal{X}^*)$ is the variance compute from the resample. As $X_i \sim Weibull$, we know that, for $c = 2$, $\sigma(X)^2 = \frac{1}{\lambda^2} - \left(\frac{1}{\lambda}\frac{\sqrt{\pi}}{2}\right)^2$. Thus we defined $\hat{\sigma}(X)^2$ as $\frac{1}{\hat{\lambda}^2} - \left(\frac{1}{\hat{\lambda}}\frac{\sqrt{\pi}}{2}\right)^2$. The iterated bootstrap confidence interval is $\left[T(\mathcal{X}) - \hat{\sigma}(\mathcal{X}^*)n^{-\frac{1}{2}}y^*_{1-\frac{\alpha}{2}}, T(\mathcal{X}) - \hat{\sigma(\mathcal{X}^*)}n^{-\frac{1}{2}}y^*_{\frac{\alpha}{2}}\right]$ where y^*_a is the a-quantile of $\sqrt{n}\frac{T(\mathcal{X}^*)-\hat{\lambda}^2}{\hat{\sigma}(\mathcal{X}^{**})}$. That means we use the bootstrap estimation for the variance estimation instead of the statistic estimator $\hat{\sigma}(X)^2$ defined previously. As a consequence to compute the quantile of y^*_a, we need to conduct a double bootstrap method to obtain the variance estimation $\hat{\sigma}(\mathcal{X}^{**})$.

Table 2 summarizes the confidence interval at $\alpha = 0.05$ obtained with a sample of $n = 30$ with the exact confidence interval of equation 3, the asymptotic normal confidence interval of equation 5 and the four bootstrap methods. First, we note that the true value of λ^2, $\lambda^2 = 4$, is contained in all confidence intervals. We observe that the difference between all upper and lower bound is almost the same across all methods. Compare to the exact chi, the asymptotic normal, the basic bootstrap and T-bootstrap confidence interval are shifted to the left. On the other side, the percentile bootstrap is shifted to the right. The iterated bootstrap method gives an interval that is more centered and encompasses the exact chi interval.

By the Monte-Carlo simulation, we will study the performances of the different methods used to obtain the confidence intervals. We fix the number of Monte-Carlo simulation and resampling to respectively $M = 1000$ and $B = 5000$. n, the sample size, will vary across the simulations. We compare the methods using the average length and the

Table 2: 5% Confidence intervals for the estimator of λ^2

method	Lower bound	Upper bound
Exact Chi	2,846	5,857
ASN	2,647	5,791
Basic B	2,316	5,337
Perc B	3,101	6,122
T-Bootstap	1,926	5,177
Iterated B	2,792	5,905

coverage. The average length is the mean length obtained by MC of the confidence interval for a given method. The coverage is the number of times that the true value of the estimator λ^2 falls into the confidence interval divided by the number of MC simulation. We expect the coverage to be close to 0.95 as the confidence intervals is construct with $\alpha = 5\%$. We want to see with a low n how each methods perform compare to the true distribution. With a high n, we will check if the coverage converges to 0.95 and if we have a low length.

As t-bootstrap method is second order accurate, we expect it to give better results than other methods when n increases. The asymptotic normal, the basic bootstrap and the percentile bootstrap are first order accurate which means that we expect them to converge to the true coverage and to a small length slower than the t-bootstrap method.

Table 3: Performance evaluation using Monte-Carlo

sample size	n=5		n=10		n=100		n=500	
method	Length	Coverage	Length	Coverage	Length	Coverage	Length	Coverage
Exact Chi	8,61	0,943	5,45	0,939	1,58	0,949	0,70	0,948
ASN	7,01	0,861	4,96	0,888	1,57	0,949	0,70	0,948
Basic B	12,21	0.810	6,22	0,859	1,58	0,933	0,70	0,94
Perc B	12,21	0,778	6,22	0,843	1,58	0,945	0,70	0,946
T-Bootstap	22,61	0,763	8,14	0,82	1,63	0,929	0,70	0,938
Interated B	30,30	0,952	7,55	0,931	1,60	0,947	0,70	0,944

In Table 3[1], we can see that the exact Chi have a coverage that is always almost equal to 0.95. This is coherent with the fact that its pivotal root from equation 2 does not depend on n. Its length decreases with n meaning that to obtain this interval of 95%, smaller interval is needed. For a small sample, $n = 5, 10$, the asymptotic normal has the lowest length (7.01 for $n = 5$ and 4.96 for $n = 10$). The three first bootstrap methods perform more poorly than the exact and the asymptotic distribution. They have higher length and lower coverage in all cases. The iterated bootstrap has a coverage close to 0.95 but has a high length. Compare to the t-bootstrap, for $n = 10$, the iterated has a higher coverage (0.93 vs 0.82 for t-bootstrap) and a lower length (7.55 vs 8.14 for the t-boostrap). We can say from this result that the iterated bootstrap performs better than the t-bootstrap. The t-bootstrap have the worst length and coverage with respectively 22.61 and 0.763 for $n = 5$ and 8.14 and 0.82 for $n = 10$. As n growths (for $n = 100, 500$), the asymptotic normal and the bootstrap methods catch up the exact chi coverage and length. We still observe that the t-bootstrap performs more poorly than the other methods. For $n = 100$, its lenght is a bit higher than the others (1.63) and its coverage lower (0.929).

[1]The iterated bootstrap method is calculated with $M = 1000$ and $B = 1000$ due to the high time necessary for the computation. N, the number of resampling in the second level bootstrap is $N = 500$.

With respect to the expected performance of the methods, the t-bootstrap does not meet our expectations. Moreover, the iterated bootstrap seems to be better than the t-bootstrap. One possible explanation would be that the $\hat{\sigma^2}$ used in the t-bootstrap is not appropriate in this case. Another explanation would be that the high shift to the left on table 2 could explain the poor performance of the t-bootstrap.

4 Hypothesis test and power

In this section, we will present a hypothesis test for which we will reject one hypothesis and do not reject another one. Then we will compare with the rejected hypothesis the power of the different bootstrap methods.

We have chosen to conduct a hypothesis parametric test on the parameter λ^2 with a sample of n=30 from the Weibull distribution. We performed a Monte-Carlo simulation of the Weibull to compute the p-value $\hat{p} = \frac{1}{B+1}\left(\sum_{b=1}^{B} I(\hat{\lambda}_b^{2*} \geq \hat{\lambda}^2) + 1\right)$ where $\hat{\lambda}_b^{2*}$ is the estimator of λ^2 in each Monte-Carlo simulation and $\hat{\lambda^2}$ the estimator of the original sample.

We define the test where we want H_0 not to be rejected as

[H0] $\lambda^2 = 2^2$
[Ha] $\lambda^2 \neq 2^2$

We perform a bilateral test where the null hypothesis is rejected if the $\hat{p}$ is less than $\frac{\alpha}{2}$ where $\alpha = 0.05$. The p-value obtained from this test is 0.4112, and as a consequence, we do not reject the null hypothesis.

And we define the test where we want H_0 to be rejected as

[H0] $\lambda^2 = 1.6^2$
[Ha] $\lambda^2 \neq 1.6^2$

We reject the null hypothesis if the $\hat{p}$ is less than $\frac{\alpha}{2}$ where $\alpha = 0.05$. The p-value obtained from this test is 0.0069, and as a consequence, we reject the null hypothesis. With this test result, we can now discuss the power of the bootstrap methods using the rejected hypothesis.

We use our sample of $n = 30$ to construct the confidence intervals with $\alpha = 0.05$ for the different bootstrap methods. These confidence intervals are the same as in the table 2. With these intervals we can count the number of time (over $M = 1000$ replications) that the $\hat{\lambda}_m^{2*}$ from a sample generated under the alternative hypothesis (for example $\lambda^2 = 1.6^2$ instead $\lambda^2 = 2^2$) falls into theses intervals. This number gives us the probability of non-reject of the null hypotheses under the alternative hypothesis $P(NRH_0|H_a)$. The power of the test under the alternative hypothesis is given by $P(RH_0|H_a)$. We can know have this quantity in subtracting : $1 - P(NRH_0|H_a) = P(RH_0|H_a)$. The power of the test under the null hypothesis is given by $P(RH_a|H_0)$. By definition this quantity is 0.05 for all bootstrap methods.

We want a high power of the test because the method would reject the null hypothesis when it is wrong. On table 4, at $\lambda^2 = 1.6^2$, the percentile bootstrap has the highest power when on the other side, the t-bootstrap have a low power. This result is coherent with the previous section when we discussed the performance of different methods. We illustrated the power of the bootstrap methods on the figure 2 for all values of λ^2. On the line with $lambda = 1.6$, we see the percentile bootstrap performs better. In general on the left of $\lambda^2 = 4$ the percentile bootstrap has a higher power whereas the t-bootstrap has a higher power on its right. These results seem coherent with the table 2 where the percentile bootstrap has the more shifted to the right intervals whereas the T-bootstrap has the more shifted to the left intervals.

Table 4: Power of the bootstap methods with $\lambda^2 = 1.6^2$ with n=30

method	Power
Basic B	0.262
Perc B	0.827
T-Bootstap	0.036
Iterated B	0.542

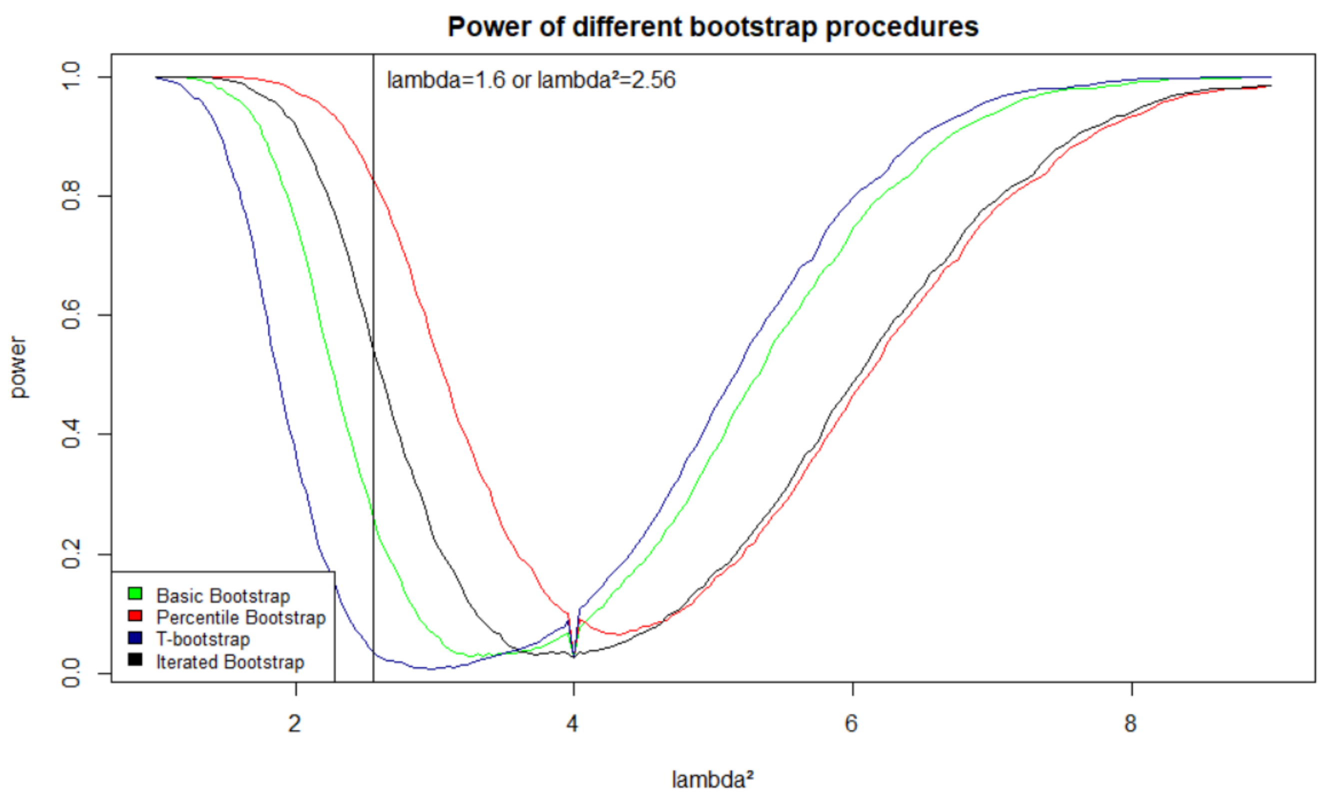

Figure 2: Power of the bootstrap procedures with n=30

5 Conclusion

In this project, we have studied the performance of the different bootstrap methods seen in class with the estimator of the parameter of interest, λ^2 from a Weibull(λ, 2).

We used its MLE estimator and we evaluated the bias and the variance with the bootstrap method and we observed a small bias. By constructing and comparing the confidence intervals, one can see that, in the case where $n = 30$, they are shifted either on the left or on the right of the exact Chi interval except for the iterated bootstrap that surround closely this exact interval.

The performance of the methods tells that the t-bootstrap is not the best bootstrap method in our case compare to the other way to construct the intervals. This result is maybe explained by the strong shift to the left of its confidence interval compare to the exact Chi. The asymptotic normal performs better than the simple bootstrap methods. The iterated bootstrap seems to have an interval that contains the Exact Chi's interval according to the table 2. The interval seems however to be longer than other methods.

The power of the test is also influenced by the way the intervals are constructed. The percentile bootstrap performs better for lower λ^2 than 4 and the t-bootstrap has a higher power with higher values of λ^2.

6 Appendix

Maximum likelihood estimator :

Let $X_1, ..., X_n$ be $W(\lambda, c)$. The density function is : $f(x, \lambda, c) = c\lambda(x\lambda)^{c-1}exp(-(x\lambda)^c)$.

Then the likelihood is $L(X, \lambda, c) = \prod_{i=1}^n f(x_i, \lambda, c) = \prod_{i=1}^n c\lambda(x_i\lambda)^{c-1}exp(-(x_i\lambda)^c)$

We can now compute explicitly the log-likelihood :

$$\Longrightarrow \ln L(X, \lambda, c) = \sum_{i=1}^n (\ln(c\lambda) + (c-1)\ln(x_i\lambda) - (x_i\lambda)^c) = n\ln(c\lambda) + (c-1)\sum_{i=1}^n \ln(x_i\lambda) - \lambda^c \sum_{i=1}^n x_i^c$$

Assuming that c=2, we look for the MLE estimator for λ^2 :

$$\frac{\partial \ln L(X, \lambda, 2)}{\partial \lambda^2} = 0 \iff \frac{\partial \left(n\ln(2\lambda) + \sum_{i=1}^n \ln(x_i\lambda) - \lambda^2 \sum_{i=1}^n x_i^2\right)}{\partial \lambda^2} = 0$$

$$\iff \frac{2n}{2\hat{\lambda}}\frac{1}{2\hat{\lambda}} + \sum_{i=1}^n \frac{x_i}{2\hat{\lambda}}\frac{1}{x_i\hat{\lambda}} - \sum_{i=1}^n x_i^2 = 0$$

$$\iff \frac{n}{2\hat{\lambda}^2} + \sum_{i=1}^n \frac{1}{2\hat{\lambda}^2} - \sum_{i=1}^n x_i^2 = 0$$

$$\iff \frac{n}{2\hat{\lambda}^2} + \frac{n}{2\hat{\lambda}^2} = \sum_{i=1}^n x_i^2$$

$$\iff \frac{n}{\hat{\lambda}^2} = \sum_{i=1}^n x_i^2$$

$$\iff \hat{\lambda}^2 = \frac{n}{\sum_{i=1}^n x_i^2}$$

The maximum likelihood estimator for λ^2 is $\hat{\lambda}^2 = \frac{n}{\sum_{i=1}^n x_i^2}$

Asymptotic distribution of $\hat{\lambda}^2$ and Fisher information :

$$I_1(\lambda^2) = Var(\frac{\partial \ln L_1(X, \lambda, 2)}{\partial \lambda^2})$$

$$\iff I_1(\lambda^2) = VAR\left[\frac{1}{\lambda^2} - x_1^2\right] = \mathbb{E}\left[(\frac{1}{\lambda^2} - x_1^2)^2\right] - \mathbb{E}\left[\frac{1}{\lambda^2} - x_1^2\right]^2 = \left(\frac{1}{\lambda^4} + \mathbb{E}[x_1^4] - \frac{2}{\lambda^2}\mathbb{E}[x_1^2]\right) - \left(\frac{1}{\lambda^2} - \mathbb{E}[x_1^2]\right)^2$$

$$\iff I_1(\lambda^2) = m_4 - m_2^2$$

where m_2 is the moment of order 2 and m_4 the moment of order 4