

IEOR-4709

A. Capponi
Spring 2025

Problem Set #2

Issued: February 5, 2025
Due: **BEFORE CLASS** February 19, 2025

Note: Please put the number of hours that you spent on this homework set on top of the first page of your homework. The CA in charge of grading this homework is Yu Gao. The TA in charge of this homework is Jose Antonio Sidaoui.

Ex. 1.

Suppose $\{X_n\}_{n \geq 1}$ is a sequence of independent and identically distributed random variables from a distribution with density $f_X(x) = \frac{1}{\beta} e^{-\frac{x-\theta}{\beta}}$, for $\theta \leq x < \infty$. Here, θ, β are unknown parameters. Do the following, making all your arguments complete and precise.

- Find the maximum likelihood (MLE) estimator for θ and β .
- Using the method of moments (MM), calculate an estimator for θ and β .
- Using Python or R (or your favourite programming language), randomly draw 5 values from the distribution with parameter $\theta = 3$ and $\beta = 2$. Calculate the MLE for both θ and β . Repeat this 1000 times. Calculate the mean and variance of these estimates, as well as the approximate mean squared error (i.e. the mean squared difference between the estimate and the true value).

Solution

- The likelihood function and the log-likelihood function are

$$L(\theta, \beta | X) = \prod_{i=1}^n \frac{1}{\beta} e^{-\frac{x_i - \theta}{\beta}},$$

$$\log(\theta, \beta | X) = -n \log(\beta) - \frac{1}{\beta} \sum_{i=1}^n (x_i - \theta).$$

We see it is an increasing function of θ , so the optimal $\hat{\theta} = \min\{x_1, \dots, x_n\}$. As for β , differentiating the log-likelihood with respect to β and setting it to zero gives:

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^n \left(-\frac{1}{\beta} + \frac{X_i - \theta}{\beta^2} \right) = 0$$

Solving for β yields:

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\theta})$$

- We calculate

$$\mathbb{E}(f(\theta, \beta)) = -e^{\frac{\theta-x}{\beta}} - \beta e^{\frac{\theta-x}{\beta}} \Big|_{\theta}^{\infty} = \theta + \beta$$

$$\mathbb{E}(f(\theta, \beta)^2) = (-e^{\frac{\theta-x}{\beta}})(2\beta^2 + x^2 + 2\beta x) \Big|_{\theta}^{\infty} = \theta^2 + 2\beta\theta + 2\beta^2$$

Then

$$\frac{1}{n} \sum_{i=1}^n X_i = \bar{\theta} + \bar{\beta}$$

$$\frac{1}{n} \sum_{i=1}^n X_i^2 = \bar{\theta}^2 + 2\bar{\beta}\bar{\theta} + 2\bar{\beta}^2$$

We have a quadratic equation

$$\bar{\theta}^2 + 2\left(\frac{1}{n} \sum_{i=1}^n X_i - \bar{\theta}\right)^2 \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n X_i^2$$

From here we can solve

$$\bar{\theta} = \frac{\frac{4}{n^2} (\sum_{i=1}^n X_i)^2 \pm \sqrt{\frac{16}{n^4} (\sum_{i=1}^n X_i)^4 - 8(1 - \frac{2}{n} \sum_{i=1}^n X_i)(2\frac{1}{n^3} (\sum_{i=1}^n X_i)^3 - \frac{1}{n} \sum_{i=1}^n X_i)}}{2(1 - \frac{2}{n} \sum_{i=1}^n X_i)}$$

$$\bar{\beta} = \frac{1}{n} \sum_{i=1}^n X_i - \bar{\theta}$$

- The result varies. The sample code can be found on Canvas. A sample result looks like

```
Mean of Theta Estimates: 3.3805051627893694
Variance of Theta Estimates: 0.14852391644678542
Standard Error of Theta Estimates: 0.012193136964531368
Mean of Beta Estimates: 1.59901111106551281
Variance of Beta Estimates: 0.632808886630117
Standard Error of Beta Estimates: 0.025168280214569214
```

Remark. The asymptotic normality does not always hold for MLE estimators, as it requires some extra conditions to be satisfied (see STATS 200, Lecture 14, Theorem 14.1 at <https://web.stanford.edu/class/stats200/>).

Theorem. Let $\{f(x|\theta_0) : \theta_0 \in \Omega\}$ be a parametric model, where $\theta_0 \in R$ is a single parameter. Let X_1, X_2, \dots, X_n be i.i.d samples drawn from $f(x|\theta)$ for some $\theta \in \Omega$. Let $\hat{\theta}$ be the MLE computed from X_1, X_2, \dots, X_n . Suppose certain regularity conditions hold, including:

1. All pdfs/pmfs $f(x|\theta_0)$ in the model have the same support,
2. θ is an interior point (i.e., not on the boundary) of Ω ,

3. The log-likelihood function $L(\theta_0|X)$ is differentiable in θ_0 , and

4. $\hat{\theta}$ is the unique value of $\theta_0 \in \Omega$ that solves the equation $L'(\theta_0|X) = 0$.

Then $\hat{\theta}$ is consistent and asymptotically normal, with

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow \mathcal{N}(0, \frac{1}{I(\theta)}). \quad (1)$$

In the question, we see the support of $f_X(x)$ is $[\theta, \infty)$, which depends on the choice of θ . The $L(\theta|X)$ is an increasing function with derivative equal to n when $\theta = \min\{x_1, \dots, x_n\}$. Thus the first and last conditions are not satisfied.

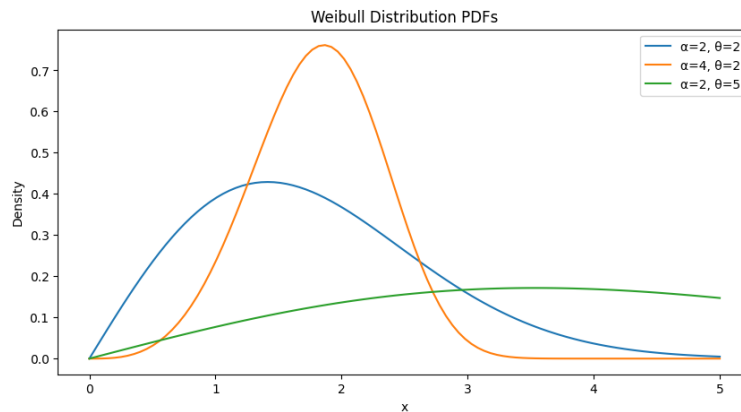
Ex. 2.

Suppose $\{X_n\}_{n=1}^N$ is a sequence of independent and identically distributed random variables drawn from the Weibull distribution with probability density function $f_X(x) = \frac{\alpha x^{\alpha-1}}{\theta^\alpha} e^{-(x/\theta)^\alpha}$, for $x \geq 0$. Do the following, making all your arguments complete and precise.

- (i) Plot the probability density function for three choices of α, θ , $\alpha = \theta = 2$, $\alpha = 4$ and $\theta = 2$, $\alpha = 2$ and $\theta = 5$.
- (ii) Find the maximum likelihood estimator for θ assuming α is known.
- (iii) Assume $N = 50$, $\alpha = 1$ and $\theta = 2$. Simulate N independent samples from the Weibull distribution with these parameters α and θ .
- (iv) Find the maximum likelihood estimator for θ, α , i.e., when both are assumed to be unknown. **(a closed-form expression is unlikely to exist and you will need to solve for it numerically)**. Report the numerical estimates obtained using the N independent samples from part (iii).
- (v) Using the method of moments (MM), calculate an estimator for θ, α , and evaluate it using the samples from part (iii).

Solution

- The plot is as follows.



- The likelihood function is given by

$$L(\theta|X) = \frac{\alpha^n}{\theta^{n\alpha}} \exp\left(-\sum_{i=1}^n (x_i/\theta)^\alpha\right) \prod_{i=1}^n x_i^{\alpha-1},$$

and the log-likelihood function is

$$\log(L(\theta|X)) = n \log \alpha - n\alpha \log \theta - \sum_{i=1}^n \left(\frac{x_i}{\theta}\right)^\alpha + (\alpha - 1) \sum_{i=1}^n \log(x_i).$$

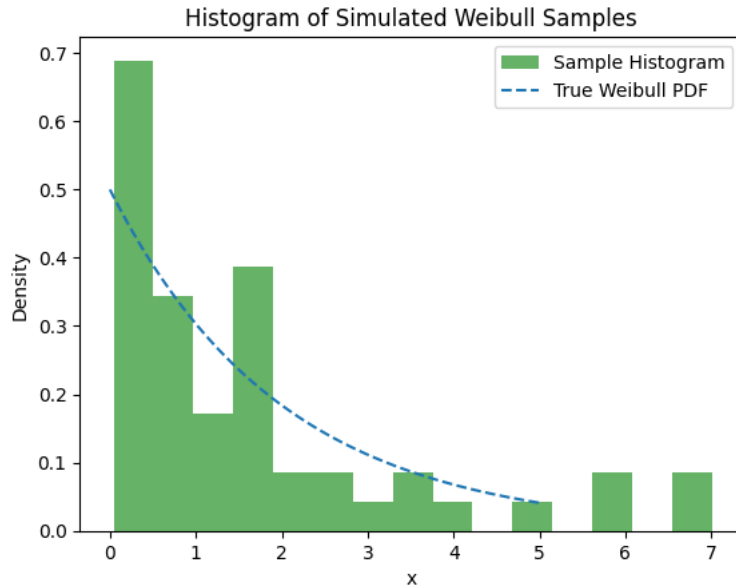
Taking the derivative of log-likelihood function with respect to θ , and setting it to zero,

$$\frac{\partial l}{\partial \theta} = -\frac{n\alpha}{\theta} + \alpha \sum_{i=1}^n \frac{x_i^\alpha}{\theta^{\alpha+1}} = 0.$$

It follows

$$-1 + \frac{1}{n} \sum_{i=1}^n \frac{x_i^\alpha}{\theta^\alpha} = 0, \theta^* = \left(\frac{1}{n} \sum_{i=1}^n x_i^\alpha\right)^{1/\alpha}.$$

- We sample 50 independent samples from the Weibull distribution and draw the the histogram as follows.



- Taking the derivative of the log-likelihood function with respect to α , and setting it to zero,

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} - n \log(\theta) - \sum_{i=1}^n \log\left(\frac{x_i}{\theta}\right) \exp(\alpha \log(\frac{x_i}{\theta})) + \sum_{i=1}^n \log(x_i) = 0.$$

Solving the above equation, we have

$$\begin{aligned}
\alpha^* &= \left(\log(\theta^*) + \frac{1}{n} \sum_{i=1}^n \log\left(\frac{x_i}{\theta^*}\right) \exp(\alpha^* \log(\frac{x_i}{\theta^*})) - \frac{1}{n} \sum_{i=1}^n \log(x_i) \right)^{-1} \\
&= \left(\log(\theta^*) \left(1 - \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\theta^*}\right)^{\alpha^*}\right) + \frac{1}{n} \sum_{i=1}^n \frac{\log(x_i) x_i^{\alpha^*}}{(\theta^*)^{\alpha^*}} - \frac{1}{n} \sum_{i=1}^n \log(x_i) \right)^{-1} \\
&= \left(\frac{\sum_{i=1}^n \log(x_i) x_i^{\alpha^*}}{\sum_{i=1}^n x_i^{\alpha^*}} - \frac{1}{n} \sum_{i=1}^n \log(x_i) \right)^{-1}
\end{aligned}$$

Given the samples we sample in the previous question, we use Newton steps to find the root α^* such that $f(\alpha^*) = 0$ where

$$f(\alpha) \equiv \left(\frac{\sum_{i=1}^n \log(x_i) x_i^\alpha}{\sum_{i=1}^n x_i^\alpha} - \frac{1}{n} \sum_{i=1}^n \log(x_i) \right) - 1/\alpha.$$

Starting with $\alpha_0 = 1$, the iteration process can be finished by doing

$$\alpha_{k+1} = \alpha_k - \frac{f(\alpha_k)}{f'(\alpha_k)}.$$

We find the final $\alpha^* = 0.96808$. Thus $\theta^* = 1.6668$.

- We match the population first moment with sample first two moments to estimate

$$E(X) = \theta \Gamma\left(1 + \frac{1}{\alpha}\right) = \frac{1}{n} \sum_{i=1}^n x_i,$$

and

$$E(X^2) = \theta^2 \Gamma\left(1 + \frac{2}{\alpha}\right) = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

Hence, α is the root of the function

$$\frac{\Gamma(1 + \frac{2}{\alpha})}{(\Gamma(1 + \frac{1}{\alpha}))^2} - \frac{\overline{X^2}}{(\bar{X})^2}.$$

We can use binary search to finally find $\alpha^* = 0.95638, \theta^* = 1.6584$.

Ex. 3.

Consider i.i.d samples X_1, X_2, \dots, X_n from a Gaussian distribution with *unknown mean* μ and unknown variance σ^2 . Our objective is to construct an interval estimator with a high confidence coefficient. Let us define the sample variance estimator

$$\sigma_s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Do the following:

- Consider the statistics $T_{n-1} := \frac{\bar{X} - \mu}{\sigma_s / \sqrt{n}}$, which we can rewrite as $T_{n-1} := \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma_s}$. What is the distribution of the numerator of T_{n-1} ? What is the distribution of the denominator of T_{n-1} ? Assume that the numerator and denominator are independent (this is true and was an important discovery by Gossett). What is the distribution of T_{n-1} ?
- Use the above statistics to construct an interval estimator for μ . Specify the left and right extreme point of the interval.
- Suppose we want to have confidence coefficient $1 - \alpha$. How do we set the constant c in the interval to achieve that?

Solution.

- The distribution of the random variable of the numerator is a Gaussian random variable with mean 0 and variable σ^2 . First of all, it's a linear combination of Gaussian random variables, so it's still Gaussian. Its mean is

$$\mathbb{E}[\sqrt{n}(\bar{X} - \mu)] = \sqrt{n}(\mathbb{E}[\bar{X}] - \mu) = 0$$

and its variance is

$$\text{Var}[\sqrt{n}(\bar{X} - \mu)] = n\text{Var}[\bar{X}] = \text{Var}[X] = \sigma^2.$$

The denominator of T_{n-1} is the square root of a constant times a Chi-square random variable with degrees of freedom $n - 1$. Namely,

$$\frac{(n-1)\sigma_s^2}{\sigma^2}$$

is a Chi-square random variable with degrees of freedom $n - 1$.

Assume that the numerator and denominator are independent (which is true),

$$\begin{aligned} T_{n-1} &= \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma_s} \\ &= \frac{\sigma Z}{\sigma \sqrt{\frac{\chi_{n-1}^2}{n-1}}} \\ &= \frac{Z}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}}, \end{aligned}$$

where Z is a standard normal random variable, and χ_{n-1}^2 is a Chi-square random variable with degrees of freedom $n - 1$. Thus T_{n-1} is a Student-t random variable with degrees of freedom $n - 1$.

- We have that

$$P\left(\left|\frac{\bar{X} - \mu}{\sigma_s / \sqrt{n}}\right| \leq c\right) = P(\bar{X} - c\sigma_s / \sqrt{n} \leq \mu \leq \bar{X} + c\sigma_s / \sqrt{n}) = 1 - 2P(T_{n-1} \leq -c),$$

so the left and right bounds of the interval are $\bar{X} - c\sigma_s / \sqrt{n}$ and $\bar{X} + c\sigma_s / \sqrt{n}$ respectively.

- We set

$$c = t_{1-\alpha/2, n-1},$$

which is the t-statistic with degree of freedom $n - 1$ and tail probability $\frac{\alpha}{2}$.

Ex. 4.

Consider a model for the stock price process of the form:

$$S_{t_n} = S_{t_{n-1}} e^{(\mu - \frac{1}{2}\sigma^2)\Delta t_n + \sigma Z_n},$$

where $\Delta t_n = t_n - t_{n-1}$ is the time interval between two consecutive observations, and Z_n is a Gaussian random variable with mean zero and variance Δt_n . **Hint: We can express time in years, so that a one day time interval can be chosen to be 1/365.**

Go on Yahoo Finance. Use the stock price data of IBM from January 1, 2021 to January 1, 2022. Do the following:

- Implement the maximum-likelihood estimator method, and report the maximum likelihood estimates for μ and σ .
- Using the Fisher Information matrix, compute an approximate 95% confidence interval for the parameters μ and σ . Please, specify how you have approximated the Fisher information matrix of the true (unknown) parameters μ and σ .

Solution

- We know that the difference between log-prices $X_i \triangleq \log(S_{t_i}) - \log(S_{t_{i-1}})$ is Gaussian with mean $(\mu - \frac{\sigma^2}{2})\Delta t$ and variance $\sigma^2\Delta t$ where $\Delta t = 1/365$. The MLE computation for Gaussian suggests that:

$$\bar{X} = (\hat{\mu} - \frac{\hat{\sigma}^2}{2})\Delta t \tag{2}$$

and

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \hat{\sigma}^2 \Delta t. \tag{3}$$

which gives

$$\hat{\sigma}^2 = \frac{1}{n\Delta t} \sum_{i=1}^n (X_i - \bar{X})^2 \tag{4}$$

and

$$\hat{\mu} = \frac{\hat{\sigma}^2}{2} + \frac{\bar{X}}{\Delta t}. \tag{5}$$

- First we calculate the Fisher Information $I(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \log L(\theta|X)}{\partial \theta \partial \theta^\top} \right]$ in this case where $\theta = (\mu, \sigma^2)$ and

$$L(\theta|X) = \frac{1}{(\sqrt{2\pi\sigma^2\Delta t})^n} e^{-\sum_{i=1}^n \frac{(X_i - (\mu - \frac{\sigma^2}{2})\Delta t)^2}{2\sigma^2\Delta t}}$$

. Thus, we have

$$\begin{aligned}\log L(\theta|X) &= -\frac{n}{2}\log(2\pi\Delta t) - \frac{n}{2}\log(\sigma^2) - \sum_{i=1}^n \frac{(X_i - (\mu - \frac{\sigma^2}{2})\Delta t)^2}{2\sigma^2\Delta t} \\ &= -\frac{n}{2}\log(2\pi\Delta t) - \frac{n}{2}\log(\sigma^2) - \sum_{i=1}^n \left(\frac{(X_i - \mu\Delta t)^2}{2\sigma^2\Delta t} + \frac{(X_i - \mu\Delta t)}{2} \right) - \frac{n\sigma^2\Delta t}{8}\end{aligned}\quad (6)$$

which gives

$$\frac{\partial \log L(\theta|X)}{\partial \mu} = \sum_{i=1}^n \frac{(X_i - (\mu - \frac{\sigma^2}{2})\Delta t)}{\sigma^2}$$

and

$$\frac{\partial \log L(\theta|X)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(X_i - \mu\Delta t)^2}{2(\sigma^2)^2\Delta t} - \frac{n\Delta t}{8}.\quad (7)$$

Thus

$$\frac{\partial^2 \log L(\theta|X)}{\partial \mu^2} = -\frac{n\Delta t}{\sigma^2}$$

,

$$\frac{\partial^2 \log L(\theta|X)}{\partial \mu \partial \sigma^2} = \frac{\partial^2 \log L(\theta|X)}{\partial \sigma^2 \partial \mu} = -\sum_{i=1}^n \frac{X_i - \mu\Delta t}{\sigma^4}$$

and

$$\frac{\partial^2 \log L(\theta|X)}{(\partial \sigma^2)^2} = \frac{n}{2\sigma^4} - \sum_{i=1}^n \frac{(X_i - \mu\Delta t)^2}{(\sigma^2)^3\Delta t}$$

Thus we have

$$I(\theta)_{1,1} = -\mathbb{E}_\theta \left[\frac{\partial^2 \log L(\theta|X)}{\partial \mu^2} \right] = \mathbb{E}_\theta \left[\frac{n\Delta t}{\sigma^2} \right] = \frac{n\Delta t}{\sigma^2}$$

,

$$I(\theta)_{1,2} = I(\theta)_{2,1} = -\mathbb{E}_\theta \left[\frac{\partial^2 \log L(\theta|X)}{\partial \mu \partial \sigma^2} \right] = \mathbb{E}_\theta \left[\frac{n\Delta t}{\sigma^2} \right] = -\frac{n\Delta t}{2\sigma^2}$$

and

$$I(\theta)_{2,2} = -\mathbb{E}_\theta \left[\frac{\partial \log L(\theta|X)}{\partial (\sigma^2)^2} \right] = -\left(\frac{n}{2\sigma^4} - n \frac{\sigma^2\Delta + \frac{\sigma^4}{4}\Delta t^2}{\sigma^6\Delta t} \right) = \frac{n}{2\sigma^4} + \frac{n\Delta t}{4\sigma^2}$$

- We use the following method to estimate $I(\theta)$ as $V_E(\hat{\theta})$, where $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$ where

$$V_E(\hat{\theta})_{1,1} = \frac{n\Delta t}{\hat{\sigma}^2},$$

$$V_E(\hat{\theta})_{1,2} = -\frac{n\Delta t}{2\hat{\sigma}^2},$$

$$V_E(\hat{\theta})_{2,1} = -\frac{n\Delta t}{2\hat{\sigma}^2},$$

$$V_E(\hat{\theta})_{2,2} = \frac{n}{2\hat{\sigma}^4} + \frac{n\Delta t}{4\hat{\sigma}^2}$$

According to Section 3 of Lecture notes 5, we know that one possible confidence region for θ is the hyper-rectangle $[\hat{\theta} + z_{\frac{\alpha}{2}} \text{diag}(I(\theta)^{-1/2}), \hat{\theta} + z_{1-\frac{\alpha}{2}} \text{diag}(I(\theta)^{-1/2})]$. In our cases, $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)$, $\alpha = 5\%$ thus $z_{\frac{\alpha}{2}} = -1.96$ and $z_{1-\frac{\alpha}{2}} = 1.96$ and we substitute $V_E(\hat{\theta})$ for $I(\theta)$. We have $\hat{\mu} = -0.20167$ and $\sqrt{\hat{\sigma}^2} = 0.2879$, thus the hyper-rectangle is $[-0.8823, 0.4790] \times [0.2734, 0.3024]$.

Ex. 5.

In class, we have seen how to estimate the drift μ and the volatility σ of the stock when we fix the time δ elapsing between two consecutive observations. An important quantity that investors and portfolio managers are interested in estimating is the Sharpe ratio. This is defined as $\frac{\mu-r}{\sigma}$, i.e. the excess expected return over the risk-free rate r , measured in units of the volatility σ . Do the following

- Using the delta method, provide an estimator for the Sharpe ratio. Determine the asymptotic distribution of the estimator.
- Go on Google or Yahoo finance, and download the time series of Citigroup stock prices from January 1, 2024 through January 1, 2025. Compute the estimate of the drift μ , volatility σ , and Sharpe ratio for the log-returns of the Citigroup stock. Additionally, provide an estimate for the variance of the estimator using the data (you will need to estimate the Fisher information matrix using one of the three methods outlined in class). Assume $r = 0.01$. Use the close prices data for your analysis.

Solution

- Let parameter $\theta = (\mu, \sigma^2)^\top$ and we want to estimate sharpe ratio $g(\theta) = \frac{\mu-r}{\sqrt{\sigma^2}}$. We use the maximum likelihood estimator of θ ,

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2)^\top,$$

where

$$\hat{\mu} = \frac{1}{2}\hat{\sigma}^2 + \frac{1}{\delta}\bar{X} = \frac{1}{2n\delta} \sum_{i=1}^n (X_i - \bar{X})^2 + \frac{1}{\delta}\bar{X}$$

and

$$\hat{\sigma}^2 = \frac{1}{n\delta} \sum_{i=1}^n (X_i - \bar{X})^2,$$

Here $\{X_i\}$ are log-returns time series data.

Since the asymptotic distribution of $\hat{\theta}$ is $N((\mu, \sigma^2)^\top, I(\theta)^{-1})$, the asymptotic distribution of $g(\hat{\theta})$ is $N(\frac{\mu-r}{\sqrt{\sigma^2}}, \xi^2)$, where

$$\xi^2 = \left(\frac{\partial g(\theta)}{\partial \theta}\right)^\top I(\theta)^{-1} \frac{\partial g(\theta)}{\partial \theta} = \left(\frac{1}{\sqrt{\sigma^2}}, -\frac{\mu-r}{2\sqrt{\sigma^6}}\right) I(\theta)^{-1} \left(\frac{1}{\sqrt{\sigma^2}}, -\frac{\mu-r}{2\sqrt{\sigma^6}}\right)^\top.$$

- We are using the log-returns over the close prices and $\delta = 1/365$ here. The estimate for μ is -0.3612, the estimate for σ is 0.3170 and the estimate for sharpe ratio is -1.1712.

We use the similar method as in Problem 4 to estimate $I(\theta)$ as $V_E(\hat{\theta})$, where

$$\begin{aligned} V_E(\hat{\theta})_{1,1} &= \frac{n\delta}{\hat{\sigma}^2}, \\ V_E(\hat{\theta})_{1,2} &= -\frac{n\delta}{2\hat{\sigma}^2}, \\ V_E(\hat{\theta})_{2,1} &= -\frac{n\delta}{2\hat{\sigma}^2}, \\ V_E(\hat{\theta})_{2,2} &= \frac{n}{2\hat{\sigma}^4} + \frac{n\delta}{4\hat{\sigma}^2} \end{aligned}$$

Therefore

$$\xi^2 = \left(\frac{1}{\sqrt{\hat{\sigma}^2}}, -\frac{\hat{\mu} - r}{2\sqrt{\hat{\sigma}^6}} \right) I(\boldsymbol{\theta})^{-1} \left(\frac{1}{\hat{\sigma}^2}, -\frac{\hat{\mu} - r}{2\sqrt{\hat{\sigma}^6}} \right)^\top = 1.2068.$$