Estimation theory – Report 4

Marta Frankowska, 208581
Agnieszka Szkutek, 208619

December 15, 2017

Contents

1 Exercise 1

We generate a time series of 20 observations according to the model $y_t = \alpha + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2)$ and *iid*.

1.1 Part 1

The density function of a single observation is

$$f(y_t, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_t-\alpha)^2}{2\sigma^2}},$$

the likelihood function is

$$L(\theta; y_1, \ldots, y_N) = f(y_1, \ldots, y_N; \theta) = \prod_{i=1}^{N} f(y_i; \theta) =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_1-\alpha)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_N-\alpha)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i-\alpha)^2},$$

and the log-likelihood is as follows

$$l(\theta; y_1, \ldots, y_N) = \ln L(\theta; y_1, \ldots, y_N) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \alpha)^2.$$

1.2 Part 2

The contour plot of log-likelihood function is displayed below.

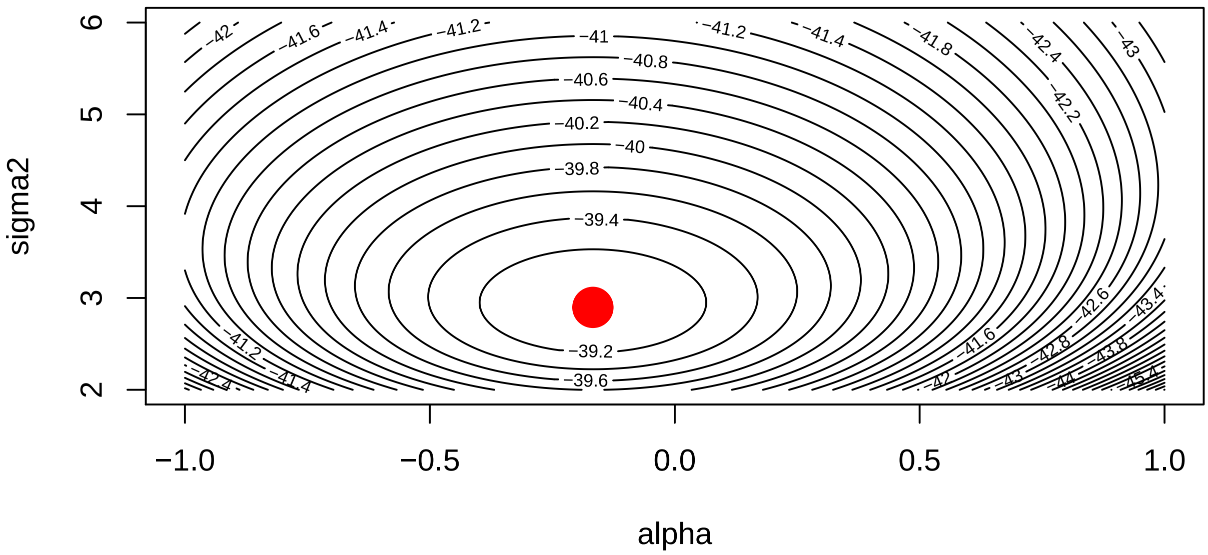

Figure 1: Contour plot of log-likelihood function. The red point marks the maksimum of the function.

1.3 Part 3

The First Order Condition is as follows $\frac{\partial \ln L}{\partial \theta}$, where vector θ is equal to $\theta = \begin{bmatrix} \alpha \\ \sigma^2 \end{bmatrix}$. It gives the following set of equations

$$\begin{cases} \frac{\partial \ln L}{\partial \alpha} = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = 0 \end{cases} \Rightarrow \begin{cases} \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i - \alpha) = 0 \\ -\frac{N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{N}(y_i-\alpha)^2 = 0 \end{cases} \Rightarrow \begin{cases} \frac{N}{N}\sum_{i=1}^{N} y_i - N\alpha = 0 \\ \frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i-\alpha)^2 = N \end{cases}$$

$$\Rightarrow \begin{cases} N(\bar{y} - \alpha) = 0 \\ \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i-\alpha)^2 \end{cases} \Rightarrow \begin{cases} \alpha = \bar{y} \\ \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2 \end{cases}$$

So the ML estimators of the model parameters are

$$\begin{cases} \hat{\alpha} = \bar{y} \\ \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \end{cases}$$

1.4 Part 4

1.4.1 Variance-covariance matrix

The variance-covariance matrix of parameters $\hat{\alpha}$ and $\hat{\sigma}^2$:

$$\Sigma_{(\hat{\alpha},\hat{\sigma}^2)} = \begin{bmatrix} Var\hat{\alpha} & Cov(\hat{\alpha},\hat{\sigma}^2) \\ Cov(\hat{\alpha},\hat{\sigma}^2) & Var\hat{\sigma}^2 \end{bmatrix},$$

where $Var\hat{\alpha} = Var\bar{y} = Var\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) = \frac{\sigma^2}{N}$.

We know that $\frac{N\hat{\sigma}^2}{\sigma^2} \sim \chi^2(N-1)$ and $Var[\chi^2(N-1)] = 2(N-1)$ so

$$Var\left(\frac{N\hat{\sigma}^2}{\sigma^2}\right) = \frac{N^2}{\sigma^4}Var(\hat{\sigma}^2) = 2(N-1) \quad \Rightarrow \quad Var(\hat{\sigma}^2) = \frac{2(N-1)\sigma^4}{N^2}$$

To calculate $Cov(\hat{\alpha},\hat{\sigma}^2)$ we will use the following fact.

Lemma 1 *Let $X_1,...,X_N \sim N(\mu,\sigma^2)$. Then $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$ are independent. That means $Cov(\bar{X},\hat{\sigma}^2) = 0$.*

Thus the variance-covariance matrix of parameters $\hat{\alpha}$ and $\hat{\sigma}^2$ is

$$\Sigma_{(\hat{\alpha},\hat{\sigma}^2)} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \frac{2(N-1)\sigma^4}{N^4} \end{bmatrix}.$$

Substituting $\sigma^2 = Var(Y)$ and N equal to length of the data we obtain the sample variance-covariance matix

```
##          [,1]        [,2]
## [1,] 2.991618 0.000000000
## [2,] 0.000000 0.002125572
```

1.4.2 The asymptotic distribution of ML estimator

The asymptotic distribution of ML estimator is

$$\hat{\theta} \xrightarrow{d} N(\theta_0, I(\theta_0)^{-1}),$$

where $I(\theta_0) = -E_0(H(\theta_0)) = -E_0\left(\frac{\partial^2 \ln L}{\partial\theta_0\partial\theta_0'}\right)$.

We calculate partial derivatives:

$$\frac{\partial \ln L}{\partial\alpha} = \frac{1}{\sigma^2}N(\bar{y}-\alpha), \quad \frac{\partial^2 \ln L}{\partial\alpha^2} = -\frac{N}{\sigma^2}$$

$$\frac{\partial \ln L}{\partial\sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(y_i-\alpha)^2, \quad \frac{\partial^2 \ln L}{\partial(\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{N}(y_i-\alpha)^2$$

$$\frac{\partial^2 \ln L}{\partial\alpha\partial\sigma^2} = \frac{\partial^2 \ln L}{\partial\sigma^2\partial\alpha} = -\frac{1}{\sigma^4}\sum_{i=1}^{N}(y_i-\alpha)$$

Now we calculate expected value of the above derivatives:

$$
\begin{aligned}
E\frac{\partial^2 \ln L}{\partial\alpha^2} &= -\frac{N}{\sigma^2}\\
E\frac{\partial^2 \ln L}{\partial(\sigma^2)^2} &= E\left(\frac{N}{2\sigma^4}\right) - E\left(\frac{1}{\sigma^6}\sum_{i=1}^{N}(y_i-\alpha)^2\right) = \frac{N}{2\sigma^4} - \frac{N}{\sigma^4} = -\frac{N}{2\sigma^4}\\
E\frac{\partial^2 \ln L}{\partial\alpha\partial\sigma^2} &= -\frac{1}{\sigma^4}E\sum_{i=1}^{N}(y_i-\alpha) = -\frac{1}{\sigma^4}\sum_{i=1}^{N}E(y_i-\alpha) = 0
\end{aligned}
$$

So

$$
I^{-1}(\theta_0) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}
$$

We see that asymptotic covariance of $\hat{\alpha}$ and $\hat{\sigma}^2$ is equal to 0.

1.5 Part 5

To calculate ML estimator of $1+\alpha+\alpha^2$ we use below property.

Lemma 2 (Invariance property) $\hat{g}(\theta) = g(\hat{\theta})$, *where* $\hat{\theta}$ *is MLE of* θ *and* g *is continous and continously differentaible funcion.*

ML estimator of $\beta = 1+\alpha+\alpha^2$ is

$$
\hat{\beta} = \hat{g}(\alpha) = g(\hat{\alpha}) = 1+\hat{\alpha}+\hat{\alpha}^2 = 1+\bar{y}+\bar{y}^2.
$$

Now we calculate $Var\hat{\beta}$ using fact that $\hat{\alpha} \sim N(\alpha, \frac{\sigma^2}{N})$.

$$
\begin{aligned}
Var(\hat{\beta}) = Var(1+\hat{\alpha}+\hat{\alpha}^2) = Var(\hat{\alpha}+\hat{\alpha}^2) = E(\hat{\alpha}+\hat{\alpha}^2)^2 - (E(\hat{\alpha}+\hat{\alpha}^2))^2 =\\
= E\hat{\alpha}^2 + 2E\hat{\alpha}^3 + E\hat{\alpha}^4 - (E\hat{\alpha})^2 - 2E\hat{\alpha}E\hat{\alpha}^2 - (E\hat{\alpha}^2)^2
\end{aligned}
$$

Using formula for normal distribution moments we get

$$
E\hat{\alpha} = \alpha, \quad E\hat{\alpha}^2 = \alpha^2 + \frac{\sigma^2}{N}, \quad E\hat{\alpha}^3 = \alpha^3 + 3\alpha\frac{\sigma^2}{N}, \quad \text{and} \quad E\hat{\alpha}^4 = \alpha^4 + 6\alpha^2\frac{\sigma^2}{N} + 3\frac{\sigma^4}{N^2},
$$

and

$$
\begin{aligned}
Var(\hat{\beta}) = \alpha^2 + \frac{\sigma^2}{N} + 2\alpha^3 + 6\alpha\frac{\sigma^2}{N} + \alpha^4 + 6\alpha^2\frac{\sigma^2}{N} + 3\frac{\sigma^4}{N^2} - \alpha^2 - 2\alpha\left(\alpha^2 - \frac{\sigma^2}{N}\right) - \left(\alpha^2 + \frac{\sigma^2}{N}\right)^2 =\\
= \frac{\sigma^2}{N} + 4\alpha\frac{\sigma^2}{N} + 4\alpha^2\frac{\sigma^2}{N} + 2\frac{\sigma^4}{N^2}.
\end{aligned}
$$

2 Exercise 2

In this exercise we will be using data set from file *datalab4-1.xlsx*. We will assume that y has a mixed normal distribution,

$$
y_n \sim \begin{cases} N(0,1), & \text{for } p \\ N(\mu,\sigma^2), & \text{for } 1-p, \end{cases} \qquad \text{depending on } \theta = \begin{bmatrix} \mu \\ \sigma^2 \\ p \end{bmatrix}.
$$

2.1 Part 1

The density function of y is equal to

$$f(y;\theta) = f(y;\mu,\sigma^2,p) = p\cdot f(y;0,1) + (1-p)\cdot f(y;\mu,\sigma^2),$$

where

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

is the density function for normal distribution with parameters μ and σ^2.

The likelihood function of y is obtained in the following way

$$L(\theta;y_1,\dots,y_N) = \prod_{i=1}^{N} f(y_i;\theta) = \prod_{i=1}^{N}\left(p\cdot f(y_i;0,1) + (1-p)\cdot f(y_i;\mu,\sigma^2)\right) =$$

$$= \prod_{i=1}^{N}\left(p\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{y_i^2}{2}} + (1-p)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}\right).$$

The log-likelihood function of y is as follows

$$l(\theta;y_1,\dots,y_N) = \log L(\theta;y_1,\dots,y_N) = \log\prod_{i=1}^{N} f(y_i;\theta) = \sum_{i=1}^{N}\log f(y_i;\theta) =$$

$$\sum_{i=1}^{N}\log\left(p\cdot f(y_i;0,1) + (1-p)\cdot f(y_i;\mu,\sigma^2)\right).$$

2.2 Part 2

In the left plot we took $\sigma = [0.01, 0.02, 0.03, \dots, 4]$ and in the right $\sigma = [e^{-1}, e^{-2}, e^{-3}, \dots, e^{-N}]$, where N is the length of the data.

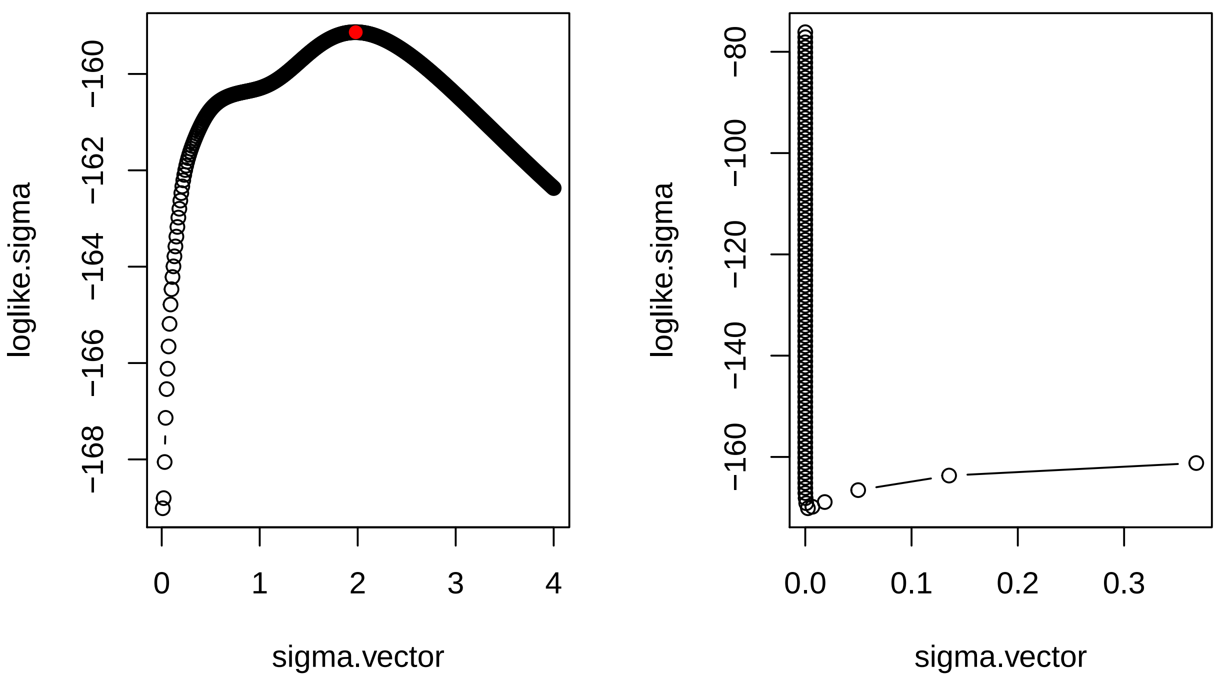

Figure 2: Log-likelihood function for different values of sigma

$$L(\theta; y_1, \ldots, y_N) = \prod_{i=1}^{N} \left(p\frac{1}{\sqrt{2\pi}} e^{-\frac{y_i^2}{2}} + (1-p)\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}} \right) =$$

$$= \left(p\frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} + (1-p)\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_1-\mu)^2}{2\sigma^2}} \right) \cdots \left(p\frac{1}{\sqrt{2\pi}} e^{-\frac{y_N^2}{2}} + (1-p)\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_N-\mu)^2}{2\sigma^2}} \right)$$

We took $\mu = y_1$, so

$$\mu - y_1 = 0 \quad \Rightarrow \quad e^{-\frac{(y_1-\mu)^2}{2\sigma^2}} = e^0 = 1.$$

Then the second component of the sum is $(1-p)\frac{1}{\sqrt{2\pi\sigma^2}}$. But if we take σ close to 0, then the whole fraction goes to infinity.

$$(1-p)\frac{1}{\sqrt{2\pi\sigma^2}} \to^{\sigma\to 0} \infty \quad \Rightarrow \quad \left(p\frac{1}{\sqrt{2\pi}} e^{-\frac{y_1^2}{2}} + (1-p)\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_1-\mu)^2}{2\sigma^2}} \right) \to^{\sigma\to 0} \infty \quad \Rightarrow$$
$$\Rightarrow \quad L(\theta; y_1, \ldots, y_N) \to^{\sigma\to 0} \infty \quad \Rightarrow \quad \ln L(\theta; y_1, \ldots, y_N) \to^{\sigma\to 0} \infty$$

Then $l(\theta; y_1, \ldots, y_N)$ goes to infinity, so maximum of $l(\theta; y_1, \ldots, y_N)$ doesn't exist.

3 Exercise 3

We suppose that y has a Bernoulli distribution

$$y = \begin{cases} 1, & F(x_n; \theta) \\ 0, & 1 - F(x_n; \theta), \end{cases}$$

where $F(x_n; \theta) = \frac{e^{\theta_1 + x_n\theta_2}}{1+e^{\theta_1 + x_n\theta_2}}$ is the probability of success.

3.1 Part 1

The marginal effect of variable x on the probability of success is given by the partial derivative

$$\frac{\partial F(x;\theta)}{\partial x} = \frac{\partial}{\partial x}\left(\frac{e^{\theta_1+x_n\theta_2}}{1+e^{\theta_1+x_n\theta_2}}\right) = \frac{\theta_2 e^{\theta_1+x_n\theta_2}(1+e^{\theta_1+x_n\theta_2}) - e^{\theta_1+x_n\theta_2}\theta_2 e^{\theta_1+x_n\theta_2}}{(1+e^{\theta_1+x_n\theta_2})^2} =$$
$$= \theta_2 e^{\theta_1+x_n\theta_2}\frac{1+e^{\theta_1+x_n\theta_2} - e^{\theta_1+x_n\theta_2}}{(1+e^{\theta_1+x_n\theta_2})^2} = \frac{\theta_2 e^{\theta_1+x_n\theta_2}}{(1+e^{\theta_1+x_n\theta_2})^2}.$$

3.2 Part 2

The likelihood function is as follows

$$L(\theta; (x_1, y_1), ..., (x_N, y_N)) = \prod_{k=1}^{N} (F(x_k;\theta))^{y_k}(1 - F(x_k;\theta))^{1-y_k}$$

so the log-likelihood function is

$$l(\theta;(x_1,y_1),...,(x_N,y_N)) = \sum_{k=1}^{N}(y_k \ln F(x_k;\theta) + (1-y_k)\ln(1-F(x_k;\theta))) =$$

$$= \sum_{k=1}^{N}(y_k \ln \frac{e^{\theta_1+x_k\theta_2}}{1+e^{\theta_1+x_k\theta_2}} + (1-y_k)\ln(1-\frac{e^{\theta_1+x_k\theta_2}}{1+e^{\theta_1+x_k\theta_2}})) =$$

$$= \sum_{k=1}^{N}(y_k(\theta_1+x_k\theta_2) - y_k\ln(1+e^{\theta_1+x_k\theta_2}) + \ln 1 - \ln(1+e^{\theta_1+x_k\theta_2}) -$$

$$-y_k \ln 1 + y_k \ln(1+e^{\theta_1+x_k\theta_2})) = \sum_{k=1}^{N}(y_k(\theta_1+x_k\theta_2) - \ln(1+e^{\theta_1+x_k\theta_2}))$$

3.3 Part 3

We can see that we are not able to solve this set of equations analitically.

$$\frac{\partial l}{\partial \theta_1} = \sum_{k=1}^{N}(y_k - \frac{e^{\theta_1+x_k\theta_2}}{1+e^{\theta_1+x_k\theta_2}}) = 0$$

$$\frac{\partial l}{\partial \theta_2} = \sum_{k=1}^{N}(y_k x_k - \frac{e^{\theta_1+x_k\theta_2}x_k}{1+e^{\theta_1+x_k\theta_2}}) = 0$$

3.4 Part 4

Because we can't calculate θ_1 and θ_2 analitically, we will calculate the maximum of the log-likelihood function numerically using Newton's method.

The $\hat{\theta}_1$ and $\hat{\theta}_2$ are equal to

```
## [1] -0.2913384  1.0768976
```

We tested various starting points for the Newton's method to find $\hat{\theta}_1$ and $\hat{\theta}_2$ and the results were almost always the same, the error was $O(10^{-4})$.

3.5 Part 5

Using Newton's method we obtained also the hesian matix of the log-likelihood function and to estimate the variance-covariance matrix we use the following formula $\Sigma_{\hat{\theta}} = (-H(\hat{\theta}))^{-1}$.

```
##              [,1]         [,2]
## [1,]  0.071714628 -0.007825794
## [2,] -0.007825794  0.046786323
```

3.6 Part 6

The estimated marginal effect of x on y using $\hat{\theta}_1$, $\hat{\theta}_2$ and $\bar{x}$ is equal to

```
## [1] 0.4643691
```