



MSC. DATA SCIENCE & ARTIFICIAL INTELLIGENCE

STATISTICAL INFERENCE - PRACTICE

Dr. Marco CORNELLI

Final assignment

By: Joris LIMONIER

joris.limonier@gmail.com

Due: January 26, 2022

Contents

1	Exercise 1	1
1.1	Question (a)	1
1.2	Question (b)	1
1.3	Question (c)	3
1.4	Question (d)	5
1.5	Question (e)	6
2	Exercise 2	6
2.1	Question (a)	6
2.2	Question (b)	6

1 Exercise 1

Let $(x_1, y_1), \dots, (x_N, y_N)$ be observations assumed to be generated by a linear model:

$$y_i = a + bx_i + \epsilon_i \quad (\text{LM})$$

with $a, b \in \mathbb{R}$ and $1 \leq i \leq N$, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ independent and identically distributed (i.i.d.)

1.1 Question (a)

We compute $\mathbb{E}[y_i]$:

$$\begin{aligned} \mathbb{E}[y_i] &= \mathbb{E}[a + bx_i + \epsilon_i] && (\text{definition}) \\ &= \mathbb{E}[a] + \mathbb{E}[bx_i] + \mathbb{E}[\epsilon_i] && (\text{linearity of expectation}) \\ &= a + bx_i && (a, b, x_i \text{ deterministic, } \epsilon_i \text{ centered normal}) \end{aligned}$$

We compute $\text{Var}(y_i)$:

$$\begin{aligned} \text{Var}[y_i] &= \mathbb{E}[y_i^2] - \mathbb{E}[y_i]^2 \\ &= \mathbb{E}[(a + bx_i + \epsilon_i)^2] - (a + bx_i)^2 \\ &= a^2 + b^2x_i^2 + \mathbb{E}[\epsilon_i^2] && (\text{linearity of expectation}) \\ &\quad + 2abx_i + 2a\mathbb{E}[\epsilon_i] + 2bx_i\mathbb{E}[\epsilon_i] && (\text{and only } \epsilon_i \text{ random}) \\ &= (a^2 + b^2x_i^2 + 2abx_i) \\ &= \mathbb{E}[\epsilon_i^2] && (\epsilon_i \text{ centered normal}) \\ &= \mathbb{E}[\epsilon_i^2] - \underbrace{\mathbb{E}[\epsilon_i]^2}_{=0} \\ &= \text{Var}(\epsilon_i) \\ &= \sigma^2 \end{aligned}$$

1.2 Question (b)

First let us note that for a given $1 \leq i \leq N$, by (LM) we have that:

$$y_i = \underbrace{a + bx_i}_{\text{deterministic}} + \underbrace{\epsilon_i}_{\mathcal{N}(0, \sigma^2)} \quad (1)$$

Moreover, we know that the Probability Density Function (PDF) of a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is given by:

$$f_{\mu, \sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right)$$

and since for a given $1 \leq i \leq N$, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, its PDF is given by:

$$f_{0, \sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right)$$

Now applying the argument from (1) yields that for $1 \leq i \leq N$, the PDF of y_i is the following:

$$f(t; \theta) := f_{(a+bx_i), \sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t - (a + bx_i))^2}{2\sigma^2}\right)$$

with the parameter $\theta := (a, b) \in \mathbb{R}^2$. We write $f(t; \theta)$ (including θ), to stress the fact that we have an influence on our parameter, not over our observations.

Given this preamble, our goal is to prove the following:

$$f\left(\bigcap_{i=1}^N y_i; \theta\right) = \prod_{i=1}^N f(y_i; \theta) \quad (2)$$

We note that for a given $1 \leq i \leq N$, a, b and x_i are deterministic, so we can rewrite (2) as:

$$\begin{aligned} f\left(\bigcap_{i=1}^N y_i; \theta\right) &= f\left(\bigcap_{i=1}^N a + bx_i + \epsilon_i; \theta\right) \\ &= f_{0, \sigma^2}\left(\bigcap_{i=1}^N \epsilon_i\right) && \text{(shift of the distribution)} \\ &= \prod_{i=1}^N f_{0, \sigma^2}(\epsilon_i) && (\epsilon_i \text{'s are i.i.d.}) \\ &= \prod_{i=1}^N f(a + bx_i + \epsilon_i) \\ &= \prod_{i=1}^N f(y_i; \theta) \end{aligned} \quad (3)$$

Thus the y_i 's, $1 \leq i \leq N$ are i.i.d..

1.3 Question (c)

Let $g(y_i)$ denote the PDF of y_i . We define the likelihood function of y_i , $1 \leq i \leq N$ as:

$$\begin{aligned}\mathcal{L}(\theta) &:= f(y_1, \dots, y_N; \theta) \\ &= \prod_{i=1}^N f(y_i; \theta) \quad (y_i \text{'s are i.i.d.})\end{aligned}$$

We also define the log-likelihood as:

$$\ell(\theta) := \log \mathcal{L}(\theta)$$

which is equal to:

$$\ell(\theta) = \log \prod_{i=1}^N f(y_i; \theta) = \sum_{i=1}^N \log f(y_i; \theta)$$

Thus by (3), we obtain that the log-likelihood becomes:

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^N \log f(y_i; \theta) \\ &= \sum_{i=1}^N \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y_i - (a + bx_i))^2}{2\sigma^2} \right) \right] \\ &= -\frac{N}{2\sigma^2} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right] \sum_{i=1}^N (y_i - bx_i - a)^2\end{aligned}$$

We want find \hat{a}_{ML} which maximises ℓ with respect to a . To do so, we take the

partial derivative with respect to a and set it to 0. Therefore we get:

$$\begin{aligned}
& \frac{\partial \ell}{\partial a} = 0 \\
\Rightarrow & \frac{\partial}{\partial a} \left[-\frac{N}{2\sigma^2} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right] \sum_{i=1}^N (y_i - bx_i - \hat{a}_{ML})^2 \right] = 0 \\
\Rightarrow & -\frac{N}{2\sigma^2} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right] \sum_{i=1}^N \frac{\partial}{\partial a} [(y_i - bx_i - \hat{a}_{ML})^2] = 0 \\
\Rightarrow & \sum_{i=1}^N \frac{\partial}{\partial a} [(y_i - bx_i - \hat{a}_{ML})^2] = 0 \\
\Rightarrow & \sum_{i=1}^N -2(y_i - bx_i - \hat{a}_{ML}) = 0 \\
\Rightarrow & \sum_{i=1}^N (y_i - bx_i - \hat{a}_{ML}) = 0 \\
\Rightarrow & -N\hat{a}_{ML} + \sum_{i=1}^N y_i - bx_i = 0 \\
\Rightarrow & \hat{a}_{ML} = \frac{1}{N} \sum_{i=1}^N y_i - bx_i
\end{aligned}$$

where we used the linearity of differentiation.

Defining \bar{x}, \bar{y} as follows:

$$\begin{aligned}
\bar{x} &:= \frac{1}{N} \sum_{i=1}^N x_i \\
\bar{y} &:= \frac{1}{N} \sum_{i=1}^N y_i
\end{aligned}$$

we can rewrite \hat{a}_{ML} as:

$$\hat{a}_{ML} = \bar{y} - b\bar{x} \tag{4}$$

1.4 Question (d)

Now we want to find \hat{b}_{ML} . We proceed similarly, that is we differentiate ℓ with respect to b and equate it to 0:

$$\begin{aligned}
& \frac{\partial \ell}{\partial b} = 0 \\
\Rightarrow & \frac{\partial}{\partial b} \left[-\frac{N}{2\sigma^2} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right] \sum_{i=1}^N \left(y_i - a - \hat{b}_{ML} x_i \right)^2 \right] = 0 \\
\Rightarrow & -\frac{N}{2\sigma^2} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right] \sum_{i=1}^N \frac{\partial}{\partial b} \left[\left(y_i - a - \hat{b}_{ML} x_i \right)^2 \right] = 0 \\
\Rightarrow & \sum_{i=1}^N \frac{\partial}{\partial b} \left[\left(y_i - a - \hat{b}_{ML} x_i \right)^2 \right] = 0 \\
\Rightarrow & \sum_{i=1}^N -2x_i \left(y_i - a - \hat{b}_{ML} x_i \right) = 0 \\
\Rightarrow & \sum_{i=1}^N x_i (y_i - a) - \sum_{i=1}^N \hat{b}_{ML} x_i^2 = 0 \\
\Rightarrow & \sum_{i=1}^N x_i (y_i - a) = \hat{b}_{ML} \sum_{i=1}^N x_i^2 \\
\Rightarrow & \hat{b}_{ML} = \frac{\sum_{i=1}^N x_i (y_i - a)}{\sum_{j=1}^N x_j^2} \\
\Rightarrow & \hat{b}_{ML} = \frac{\sum_{i=1}^N x_i (y_i - a)}{\sum_{j=1}^N x_j^2}
\end{aligned}$$

where we used the linearity of differentiation.

We now plug in the value of \hat{a}_{ML} found in (4), which gives us the following system:

$$\begin{aligned} & \begin{cases} \hat{a}_{ML} = \bar{y} - \hat{b}_{ML}\bar{x} \\ \hat{b}_{ML} = \frac{1}{\sum_{j=1}^N x_j^2} \sum_{i=1}^N x_i (y_i - \hat{a}_{ML}) \end{cases} \\ \Rightarrow & \begin{cases} \hat{a}_{ML} = \bar{y} - \hat{b}_{ML}\bar{x} \\ \hat{b}_{ML} = \frac{1}{\sum_{j=1}^N x_j^2} \sum_{i=1}^N x_i (y_i - (\bar{y} - \hat{b}_{ML}\bar{x})) \end{cases} \\ \Rightarrow & \begin{cases} \hat{a}_{ML} = \bar{y} - \hat{b}_{ML}\bar{x} \\ \hat{b}_{ML} = \frac{1}{\sum_{j=1}^N x_j^2} \sum_{i=1}^N x_i y_i - x_i \bar{y} + \hat{b}_{ML} \bar{x} x_i \end{cases} \end{aligned}$$

1.5 Question (e)

2 Exercise 2

2.1 Question (a)

2.2 Question (b)