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 Question (a)

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

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

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

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

1.2 Question (b)

Our goal is to prove the following:

$$\mathbb{P}\left(\bigcap_{i=1}^{N}\{y_i = \nu_i\}\right) = \prod_{i=1}^{N}\mathbb{P}\left(\{y_i = \nu_i\}\right) \tag{1}$$

and we have that for a given $1 \leq i \leq N$:

$$\begin{aligned}
\mathbb{P}(\{y_i = \nu_i\}) &= \mathbb{P}\left(\{a + bx_i + \epsilon_i = \nu_i\}\right) \\
&= \mathbb{P}(\{\epsilon_i = \nu_i - \underbrace{(a + bx_i)}_{\text{constant}}\})
\end{aligned}$$

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

$$\mathbb{P}\left(\bigcap_{i=1}^{N} \{y_i = \nu_i\}\right) = \mathbb{P}\left(\bigcap_{i=1}^{N} \{\epsilon_i = \nu_i - (a + bx_i)\}\right)$$
$$= \prod_{i=1}^{N} \mathbb{P}\left(\{\epsilon_i = \nu_i - (a + bx_i)\}\right) \qquad (2)$$

Now, the $\nu_i - (a + bx_i)$ in equation (2) are just deterministic values. Hence, by the fact that the ϵ_i's, $1 \leq i \leq N$ are Independent and Identically Distributed (i.i.d.), we have that (2) holds.
Thus the y_i's, $1 \leq i \leq N$ are i.i.d..

1.3 Question (c)

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

$$\mathcal{L}(\theta) := \prod_{i=1}^{N} g(y_1, \ldots, y_N; \theta)$$

We also define the log-likelihood as:

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

and since the y_i's are i.i.d.:

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

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

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

We also know that for a given i, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. Therefore its PDF is given by:

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

which we can rename to $f(t;\theta)$, to stress the fact that we have an influence on our parameters $\theta = (a,b) \in \mathbb{R}^2$, not over our observations. Thus by (2), we obtain that the log-likelihood becomes:

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

and we want find $\hat{a}_{ML}$ by maximising ℓ with respect to a. Therefore we get:

$$\begin{aligned}
\hat{a}_{ML} &= \arg\max_a \ell(\theta) \\
&= \arg\max_a -\frac{N}{2\sigma^2} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \right] \sum_{i=1}^{N} (\epsilon_i - (a + bx_i))^2 \\
&= \arg\min_a \sum_{i=1}^{N} (\epsilon_i - (a + bx_i))^2 \\
&= \arg\min_a \sum_{i=1}^{N} (\epsilon_i - bx_i - a)^2
\end{aligned}$$

1.4 Question (d)

1.5 Question (e)

2 Exercise 2

2.1 Question (a)

2.2 Question (b)