HOMEWORK 4
GRA6039 ECONOMETRICS WITH PROGRAMMING
AUTUMN 2020

EMIL A. STOLTENBERG

In all the exercises that follow, when you are asked to estimate something, or compute something involving actual data, do it in Matlab.

Exercise 1. Let $X_1, \ldots, X_n$ be random variables, numbers, observations.

(a) Convince yourself of the following,

$$\sum_{i=1}^{n-1} (X_{i+1} - X_i) = X_n - X_1.$$

A convenient 'trick' is to write the sum fully out for some small n, for example $n = 4$. This is called a telescoping sum.

(b) Let $a_i = \sum_{j=i}^{3} X_j$ for $i = 1, 2, 3$. Show that

$$\sum_{i=1}^{3} a_i = X_1 + 2X_2 + 3X_3.$$

(c) Convince yourself of the following,

$$\sum_{i=1}^{n} iX_i = \sum_{j=1}^{n} \sum_{i=j}^{n} X_i.$$

Exercise 2. Let $X_1, \ldots, X_n$ an $Y_1, \ldots, Y_m$ be random variables. As usual, we write

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \text{and} \quad \bar{Y}_m = \frac{1}{m} \sum_{i=1}^{m} Y_i,$$

for the empirical averages. Define the random variables $Z_1, \ldots, Z_{n+m}$ as follows,

$$Z_1 = X_1, \ldots, Z_n = X_n,\ Z_{n+1} = Y_1, \ldots, Z_{n+m} = Y_m.$$

(a) Derive an expression for

$$\bar{Z}_{n+m} = \frac{1}{n+m} \sum_{i=1}^{n+m} Z_i,$$

in terms of $\bar{X}_n$ and $\bar{Y}_m$.

(b) When is it true that the average of the averages is equal to $\bar{Z}_{n+m}$? That is, when is

$$\frac{1}{2}(\bar{X}_n + \bar{Y}_m) = \bar{Z}_{n+m}?$$

Date: September 16, 2020.

(c) As usual, the empirical variances are

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2, \quad \text{and} \quad s_Y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2.$$

and s_Z^2 is similarly defined. Let a be some constant, show that

$$\sum_{i=1}^{n}(X_i - a)^2 = (n-1)s_X^2 + n(\bar{X}_n - a)^2.$$

Hint: A 'trick' that makes this easier is to add $0 = -\bar{X}_n + \bar{X}_n$ inside the parenthesis:

$$\sum_{i=1}^{n}(X_i - a)^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n + \bar{X}_n - a)^2.$$

(d) Show that

$$s_Z^2 = \frac{(n-1)s_X^2}{n+m-1} + \frac{(m-1)s_Y^2}{n+m-1} + \frac{nm(\bar{X}_n - \bar{Y}_m)^2}{(n+m-1)(n+m)}.$$

(e) Check the right hand side expression for s_Z^2 given in (d) by running (and understanding!) this Matlab code a few times. Feel free to change n and m.

```
n = 83;
m = 12;
x = normrnd(0,1,1,n);
y = normrnd(0,1,1,m);

xbar = mean(x);ybar = mean(y);
sx2 = var(x);sy2 = var(y);

z = [x y];
var(z)

denom = (n-1)*sx2 + (m-1)*sy2 + n*m*(xbar - ybar)^2/(n+m);
denom/(n+m-1)
```

Exercise 3. Suppose you have a coin whose probability of showing heads equals θ (some unknown parameter). We represent one toss of this coin by the random variable

$$X = \begin{cases} 0, & \text{if tails,} \\ 1, & \text{if heads,} \end{cases}$$

which means that

$$\Pr(X = 1) = \theta.$$

We decide to toss this coin until we get a heads up, then stop. By so deciding, we can define a new random variable,

$$Y = \text{the numbers of tosses until we get heads up,}$$

so that Y takes its values in $\{1, 2, 3, \ldots\}$. For example, if we toss tails, tails, heads, then $Y = 3$.

(a) Explain why

$$\Pr_\theta(Y = 3) = (1 - \theta)^2\theta.$$

What is the probability $\Pr_\theta(Y = 4)$?

(b) Derive an expression for the probability mass function $f_\theta(y) = \Pr_\theta(Y = y)$ for $y = 1, 2, 3, \ldots$.

(c) Verify that the function $f(y)$ you found in (b) is indeed a pmf., i.e. that $f(y) \geq 0$ for all y, and that $\sum_{y=1}^{\infty} f(y) = 1$. To do this you need to know that

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x},$$

when $|x| < 1$. This sum is an infinite geometric series. A finite geometric series is

$$\sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1-x},$$

provided $x \neq 1$.

(d) This is a tough exercise. Skip it if you spend too much time on it. Given that you have the pmf. $f_\theta(y)$ from (c), show that

$$\mathrm{E}\, Y = \frac{1}{\theta}. \tag{1}$$

This can be shown in different ways, one of which uses the fact from (c) and the result from Ex. 1(c) as it applies to infinite sums. Below you will also need that

$$\mathrm{Var}(Y) = \frac{1-\theta}{\theta^2}.$$

(e) Let $Y_1, \ldots, Y_n$ be independent random variables with the pmf. $f_\theta(y)$ you found above. Write down an expression for the log-likelihood function,

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f_\theta(Y_i).$$

(f) Find the maximum likelihood estimator $\widehat{\theta}_n$.

(e) Show that $\widehat{\theta}_n$ converges in probability to θ, i.e. that for any $\varepsilon > 0$,

$$\Pr(|\widehat{\theta}_n - \theta| \geq \varepsilon) \to 0,$$

as $n \to \infty$. One often says that the estimator $\widehat{\theta}_n$ is *consistent* for θ.

Exercise 4. Let $Y_1, \ldots, Y_n$ be independent random variables; and let $x_1, \ldots, x_n$ be some numbers, at least one of which does not equal zero. Assume that $Y_i \sim \mathrm{N}(\theta x_i, \sigma^2)$ for $i = 1, \ldots, n$. That is, the density of the ith random variable Y_i is

$$f_i(y; \theta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big\{ -\frac{1}{2\sigma^2}(y - \theta x_i)^2 \Big\},$$

where $\sigma > 0$ and $\theta \in \mathbb{R}$. In this exercise we will study the maximum likelihood estimators of θ and σ^2.

x	Y
0.01	-1.72
0.02	0.66
0.03	0.02
0.04	-0.19
0.05	0.76
0.06	2.66
0.07	0.12
0.08	-0.58
0.09	-0.50
0.10	2.86
0.11	0.37
⋮	⋮

TABLE 1. The first eleven rows of the dataset in the file `hw4_data.txt`.

(a) Write down the log-likelihood function $\ell_n(\theta, \sigma^2) = \sum_{i=1}^n \log f_i(Y_i; \theta, \sigma^2)$ and show that

$$\frac{\partial}{\partial \theta} \ell_n(\theta, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \theta x_i) x_i.$$

What is the expectation of $\partial \ell(\theta, \sigma^2)/\partial \theta$?

(b) Find the maximum likelihood estimator for θ, say $\widehat{\theta}_n$, and show that it is unbiased, that is $\mathrm{E}\,[\widehat{\theta}_n] = \theta$.

(c) Show that

$$\mathrm{Var}\,(\widehat{\theta}_n) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

(d) Suppose that $\sum_{i=1}^n x_i^2 \to \infty$ as $n \to \infty$. Show that $\widehat{\theta}_n \to_p \theta$ as $n \to \infty$.

(e) Find the maximum likelihood estimator $\widehat{\sigma}_n^2$ for σ^2.

(f) Show that $\widehat{\sigma}_n^2$ can be expressed as

$$\widehat{\sigma}_n^2 = \frac{\sigma^2}{n} \Big\{ \sum_{i=1}^n \frac{(Y_i - \theta x_i)^2}{\sigma^2} - \frac{(\widehat{\theta}_n - \theta)^2}{\mathrm{Var}\,(\widehat{\theta}_n)} \Big\},$$

and use this to show that

$$\mathrm{E}\,[\widehat{\sigma}_n^2] = \frac{n-1}{n} \sigma^2,$$

which means that $\widehat{\sigma}_n^2$ is biased for σ^2.

(g) It can be shown that the variance of $\widehat{\sigma}^2$ is

$$\mathrm{Var}\,(\widehat{\sigma}_n^2) = \frac{2(n-1)\sigma^4}{n^2}.$$

Show that

$$\widehat{\sigma}_n^2 \xrightarrow{p} \sigma^2, \quad \text{as } n \to \infty.$$

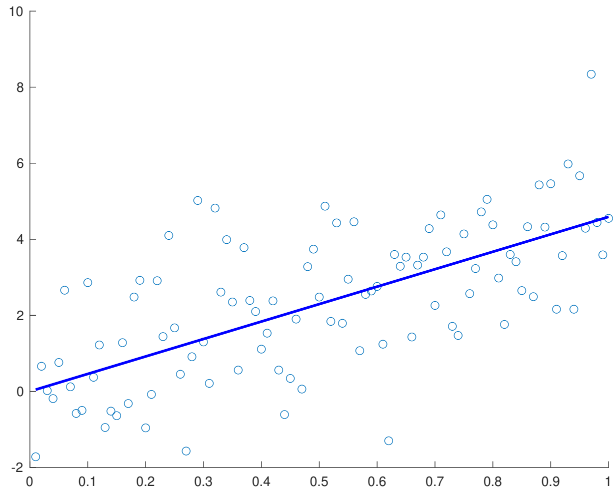

FIGURE 1. A scatter plot of the pairs of (x_i, Y_i) for $i = 1, \dots, n$ found in `hw4_data.txt`, with the fitted line $\widehat{g}_n(x) = \widehat{\theta}_n x$ overlaid.

(h) In the file `hw4_data.txt` (available on Itslearnig) you can find a dataset with $n = 100$ pairs (x_i, Y_i), whose first eleven rows are shown in Table 1. Go to Itslearning, download the full dataset, and read it into Matlab. You can use the following Matlab script

```
cd("~/Desktop/GRA6039");
data = readmatrix("hw4_data.txt");
x = data(1:100);
y = data(101:200);
```

The command `cd(" /Desktop/GRA6039");` makes sure that the Matlab script is set to the directory (folder) where `hw4_data.txt` can be found, in this example a folder named GRA6039 on your Desktop.

(i) Estimate θ and σ^2. Provide also an estimate of the standard error of $\widehat{\theta}_n$, that is an estimate of

$$\mathrm{se}(\widehat{\theta}_n) = \sqrt{\mathrm{Var}(\widehat{\theta}_n)}.$$

(j) If Z is a standard normal random variable, i.e. $Z \sim \mathrm{N}(0, 1)$, then $\Pr(-1.96 \leq Z \leq 1.96) = 0.95$. Here is a fact,

$$\frac{\widehat{\theta}_n - \theta}{\mathrm{se}(\widehat{\theta}_n)} \sim \mathrm{N}(0, 1).$$

Construct an interval that will contain θ with 95% probability. Use your estimators and the data to estimate this interval.

(k) Make a scatter plot of (x_i, Y_i) for $i = 1, \dots, n$, and add the line $\widehat{g}_n(x) = \widehat{\theta}_n x$ to your plot. Your plot should look something like the plot in Figure 1.

DEPARTMENT OF ECONOMICS, BI NORWEGIAN BUSINESS SCHOOL
Email address: `emil.a.stoltenberg@bi.no`