HOMEWORK 2
GRA6039 ECONOMETRICS WITH PROGRAMMING
AUTUMN 2020

EMIL A. STOLTENBERG AND FREDRIK PIENE

Exercise 1. Let X have the Bernoulli distribution with success probability p. The pmf. $f(x) = \Pr(X = x)$ of X is then

$$f(x) = p^x(1-p)^{1-x}, \quad \text{for } x = 0, 1,$$

and zero otherwise.

(a) Show that $\mathrm{E}\, X = p$.
(b) Show that $\mathrm{Var}\, X = p(1-p)$.
(c) Let $g(x)$ be the function given by

$$g(x) = 2x - 1,$$

and suppose that $p = 1/2$ ('a fair coin'). Show that

$$\mathrm{E}\, g(X) = 0 \quad \text{and} \quad \mathrm{Var}\, g(X) = 1.$$

Exercise 2. Consider the function given by

$$f(x) = \begin{cases} \theta x^{\theta-1}, & 0 \leq x \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

for some $\theta > 0$.

(a) Verify that $f(x)$ is a probability density function.
(b) Find an expression for the cdf. $F(x) = \int_{-\infty}^{x} f(y)\, \mathrm{d}y$.
(c) Suppose X has this distribution, we write $X \sim F$ or $X \sim f$. Find an expression for the probability $\Pr(X > 1/2)$.
(d) If $X \sim F$, find expressions for the expectation $\mathrm{E}\, X$ and the variance $\mathrm{Var}\, X$.

Exercise 3. To solve this exercise, use Proposition 2.3 in the Lecture notes. For random variables X and Y:

(a) Show that $\mathrm{Var}\, X = \mathrm{E}\,[X^2] - (\mathrm{E}\,[X])^2$.
(b) For some constant a, show that $\mathrm{Var}\,(aX) = a^2 \mathrm{Var}(X)$.
(c) For some constant a, show that $\mathrm{Var}\,(a + X) = \mathrm{Var}(X)$.
(d) Recall that $\mathrm{Cov}(X, Y) = \mathrm{E}\,[(X - \mathrm{E}\,[X])(Y - \mathrm{E}\,[Y])]$. Show that

$$\mathrm{Cov}(X, Y) = \mathrm{E}\,[XY] - \mathrm{E}\,[X]\,\mathrm{E}\,[Y].$$

(e) Show that if X and Y are independent, then $\mathrm{Cov}\,(X, Y) = 0$.

Date: August 31, 2020.

(f) Show that for constants a and b,

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y).$$

Exercise 4. Let $X_1, \ldots, X_n$ be i.i.d. random variables with expectation μ and variance σ^2. As usual, $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ is the empirical mean.

(a) Show that $\operatorname{E} \bar{X}_n = \mu$.
(b) Show that $\operatorname{Var} \bar{X}_n = \sigma^2/n$.
(c) Why do we often say that big sample sizes give more precise results?

Exercise 5. Suppose that the random variable X has the exponential distribution, that is, its pdf. is

$$f(x) = \begin{cases} \theta \exp(-\theta x), & x \geq 0, \\ 0, & x < 0, \end{cases}$$

for some $\theta > 0$.

(a) Verify that $f(x)$ is a probability density function.
(b) Show that the cdf. of X is $F(x) = \int_{-\infty}^x f(y)\,\mathrm{d}y = 1 - \exp(-\theta x)$ for $x \geq 0$ and zero for $x < 0$.
(c) Show that

$$\operatorname{E} X = \frac{1}{\theta}, \quad \text{and} \quad \operatorname{Var} X = \frac{1}{\theta^2}.$$

(d) Define a new random variable Z by

$$Z = \begin{cases} 1, & \text{if } X \geq \log(2)/\theta, \\ 0, & \text{if } X < \log(2)/\theta. \end{cases}$$

Show that

$$\operatorname{E} Z = \frac{1}{2}, \quad \text{and} \quad \operatorname{Var} Z = \frac{1}{4}.$$

Hint: Look at Ex. 1.

Exercise 6. When $Z \sim \mathrm{N}(0, 1)$, we say that Z has the standard normal distribution, the pdf. of which is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \quad \text{for } z \in \mathbb{R}.$$

The expectation of Z is zero, and its variance is one. Let Z_1 and Z_2 be independent standard normal random variables. Set

$$X = \sigma_X Z_1 + \mu_X$$
$$Y = \sigma_Y(\rho Z_1 + \sqrt{1 - \rho^2}\, Z_2) + \mu_Y,$$

where $\sigma_X, \sigma_Y > 0$, the correlation coefficient $\rho \in (-1, 1)$, and μ_X and μ_Y are real numbers.

(a) Show that $\operatorname{E} X = \mu_X$ and that $\operatorname{E} Y = \mu_Y$.
(b) Show that $\operatorname{Var} X = \sigma_X^2$ and that $\operatorname{Var} Y = \sigma_Y^2$.
(c) Show that $\operatorname{Cov}(X, Y) = \rho \sigma_X \sigma_Y$.
(d) Try to understand the Matlab script below, and run it a few times for different values of ρ, `rho` in the script.

```
n = 1000;
muX = 0; muY = 0;
sigmaX = 1; sigmaY = 1;
rho = 0.54321;

Z1 = normrnd(0,1,[1,n]); # simulate standard normals
Z2 = normrnd(0,1,[1,n]);

X = sigmaX*Z1 + muX;
Y = sigmaY*(rho*Z1 + sqrt(1 - rho^2)*Z2) + muY;

scatter(X,Y)
```

Exercise 7. Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with the uniform distribution on $[0, \theta]$, i.e. its pdf is $f(x) = 1/\theta$ for $0 \leq x \leq \theta$ and zero otherwise. Let $M_n = \max_{i \leq n} X_i$ be the largest of the observations.

(a) Show that the cdf. of one uniform X on $[0, \theta]$ is

$$F(x) = \begin{cases} 0, & \text{for } x < 0, \\ x/\theta, & \text{for } 0 \leq x \leq \theta, \\ 1 & \text{for } \theta \leq x. \end{cases} \tag{1}$$

(b) Explain why

$$\Pr(M_n \leq x) = F(x)^n.$$

Exercise 8. In this exercise we'll work with the very basics of Matlab. Suppose you have a small sample consisting of the height (measured in centimeters) of four people (Tom, Robert, Lisa, and Ann). Tom's height is 167 cm, Robert's is 185 cm, Lisa's is 177 cm, and Ann's is 159 cm.

(a) Assign each person's height to a variable. Name the variable the first letter of each person's name (e.g. the height of Tom is assigned to a variable named T). Use the following code (copy the code and paste it in the Matlab Editor):

```
T = 167;
R = 185;
L = 177;
A = 159;
```

Note the ";" behind each line of code. By including ";" at the end of a line, we *suppress the output*, meaning that the output is not printed to the command window. Also note that the equal sign "=" should be read *is made equal to*, rather than *is equal to*.

(b) We want to find the sum of the heights. Calculate this sum by using the variables defined in (a) rather than the numbers directly, and assign the output to a new variable named "sum".

(c) Calculate the sample mean, i.e. the average height. Use the variable named "sum" from (b) rather than the numbers directly.

(d) Let us improve the code to compute the sum of the heights and sample mean more efficiently. We begin by creating an array in which out height data is stored. Use the following code:

```
X = [T R L A];
```

As Matlab is designed to operate on whole arrays and matrices rather than to work with numbers one at a time, it is a good idea to always store data in arrays or matrices. Actually, Matlab is an abbreviation for "matrix laboratory".

(e) Now, calculate the sum of the heights and the sample mean using the following built-in Matlab functions: `sum(X)` and `mean(X)`.

Exercise 9. Matlab is useful for doing simulations. We will now use simulations to illustrate concepts we have gone through in class.

(a) Generate a vector of n random numbers from the normal distribution with expectation $\mu = 3$ (*mu*) and standard deviation parameter $\sigma = 2$ (*sigma*), and save the output in a variable named X. Calculate the empirical mean and variance of X. Use the following code to generate the numbers (copy the code and paste it in the Matlab Editor):

```
clear % This clears the workspace.
rng(6039); % This is a so-called random seed, needed for reproducibility.
n = 1000;
mu = 3;
sigma = 2;
X = normrnd(mu, sigma, n, 1)
```

(b) Define a new variable $\widetilde{X} = X - \bar{X}_n$. Calculate the empirical mean of $\widetilde{X}$. Was the result as expected given what you found in HW1 Exercise 9? (Note: Computers cannot (with some exceptions) do exact calculations. This average is in fact exactly zero, but this is as close to zero as the accuracy of the computer gets with the current setup.)

(c) Define a new variable $Y = aX$ and set $a = 4$. Calculate the empirical mean and variance of Y. Then, check if the average of Y is $a = 4$ times the mean of X, and that the variance of Y is $a^2 = 16$ times the variance of X. Is this result expected given what you went through in class?

(d) Generate a vector of n random numbers from the normal distribution with expectation $\mu = 3$ (*mu*) and standard deviation parameter $\sigma = 2$ (*sigma*), as in (a). Begin with $n = 10$ and create a histogram plot of X. In the same figure, also make a plot of the normal probability density function for $\mu = 3$ and $\sigma = 2$. Repeat this exercise for $n = 100$, $n = 1\,000$ and $n = 10\,000$. What do you notice about the the histogram of X as n gets larger?

```
clear
rng(6039);
mu = 3;
sigma = 2;
X = normrnd(mu, sigma, n, 1);
```

```
X_range = linspace(min(X), max(X));
histogram(X,'Normalization','pdf')
hold on % Retain current plot when adding new plots
plot(X_range, normpdf(X_range, mu, sigma))
```

(e) Now we are going to repeat what we did in (a) 1 000 times and save the empirical mean of X each time. In this way, we end up with 1 000 different averages. We are then going to make a histogram plot of these averages. Luckily, we have something called a “for loop” which automates this process - it would be tedious to run the code and save the average 1 000 times manually. Use the following code and begin with $n = 10$:

```
clear
rng(6039);
n = 10;
mu = 3;
sigma = 2;
N = 1000; % We are going to generate n numbers 1000 times
x_bar = zeros(1,N);

for i = 1:N % This is a so-called for loop
    X = normrnd(mu, sigma, n, 1);
    x_bar(i) = mean(X);
end
histogram(x_bar,'Normalization','pdf')
hold on
```

Now, re-run the for loop for $n = 100$ (i.e. by pasting the following piece of code below the “hold on”-line from the code above):

```
n = 100;
for i = 1:N
    X = normrnd(mu, sigma, n, 1);
    x_bar(i) = mean(X);
end
histogram(x_bar,'Normalization','pdf')
hold on
```

Lastly, re-run the for loop for $n = 1\,000$. What do you notice about the histograms for the different values of the sample size n? What seems to happen with the center of mass when n gets large?

Department of Economics, BI Norwegian Business School
Email address: emil.a.stoltenberg@bi.no
Email address: fredrik.piene@gmail.com