HOMEWORK 5
GRA6039 ECONOMETRICS WITH PROGRAMMING
AUTUMN 2020

EMIL A. STOLTENBERG

Exercise 1. In this exercise we prove Property PLIM.2 in Wooldridge (2019, p. 723). Recall that $X_n \to_p a$ if for any $\varepsilon > 0$,

$$\Pr(|X_n - a| \geq \varepsilon) \to 0, \quad \text{as } n \to \infty.$$

Wooldridge (2019) writes $\text{plim}(X_n) = a$ where I write $X_n \to_p a$. Useful tools when proving the following claims are the Law of total probability (see hw1, Ex. 4(b)): If A and B are two events, then $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap B^c) = \Pr(A \mid B)\Pr(B) + \Pr(A \mid B^c)\Pr(B^c)$; and the triangle inequality: For any real numbers x and y

$$|x + y| \leq |x| + |y|.$$

Assume that $X_n \to_p a$ and $Y_n \to_p b$ as $n \to \infty$.

(a) Let c be a constant not equal to zero (or else it is trivial). Show that

$$cX_n \xrightarrow{p} ca.$$

(b) Show that

$$X_n + Y_n \xrightarrow{p} a + b.$$

(c) Show that $X_n Y_n \to_p ab$. *Hint*: Add zero a couple of times and use the triangle inequality to show that

$$|X_n Y_n - ab| \leq |(X_n - a)(Y_n - b)| + |b||X_n - a| + |a||Y_n - b|,$$

and use the results from (a) and (b).

(d) Show that $X_n/Y_n \to_p a/b$ provided $b \neq 0$. *Hint*: Use that if $Y_n \to b$, and $g(y)$ is a continuous function, then $g(Y_n) \to g(b)$, combined with the result from (c).

Exercise 2. A random variable Y has the Poisson distribution with parameter θ if its pmf is

$$f_\theta(y) = \frac{1}{y!}\theta^y \exp(-\theta), \quad \text{for } y = 0, 1, 2, \ldots,$$

and $f_\theta(y) = 0$ otherwise. Let $Y_1, \ldots, Y_n$ be i.i.d. Poisson with parameter θ. If you have not done so already, find the maximum likelihood estimator for θ (see hw3 Ex. 1). In this exercise we want to estimate the probability of $Y = 0$. You might need the following fact about maximum likelihood estimators: If $\widehat{\theta}_n$ is the MLE for θ, and $h(x)$ is some real valued function, then $h(\widehat{\theta}_n)$ is the MLE for $h(\theta)$.

Date: October 6, 2020.

(a) Let $\widehat{\theta}_n$ be the MLE for θ. Explain why $\widehat{\theta}_n \to_p \theta$ (we say that $\widehat{\theta}_n$ is a *consistent* estimator for θ).
(b) Find the MLE, say $\widehat{\alpha}_n$, for $\alpha = \Pr_\theta(Y = 0)$.
(c) Explain why $\widehat{\alpha}_n$ is consistent for α, i.e. that $\widehat{\alpha}_n \to_p \alpha$.

Exercise 3. Two random variables with the same cdf have the same distribution. In a previous homework we met the Pareto distribution. Recall that the pdf of the Pareto distribution is

$$f_\alpha(x) = \alpha \frac{x_{\min}^\alpha}{x^{\alpha+1}}, \quad \text{for } x \geq x_{\min},$$

and $f_\alpha(x) = 0$ when $x < x_{\min}$, where $x_{\min} > 0$ and $\alpha > 0$. In this exercise, we assume that $x_{\min}$ is known, while α is a parameter we want to estimate. Suppose that $X_1, \ldots, X_n$ are i.i.d. from $f_\alpha(x)$. If you have not already done so, please show that the maximum likelihood estimator for α is

$$\widehat{\alpha}_n = \frac{n}{\sum_{i=1}^n \log(X_i/x_{\min})}.$$

We will now try to show that $\widehat{\alpha}_n$ is consistent for α, i.e. that $\widehat{\alpha}_n \to_p \alpha$.

(a) Let $F_\alpha(x)$ be the cdf of the Pareto distribution. Find an expression for it (you did this in a previous homework).
(b) Suppose $X \sim F_\alpha$, i.e. X has the Pareto distribution. Define,

$$Y = \log(X/x_{\min}).$$

Show that the cdf of Y is

$$\widetilde{F}_\alpha(y) = 1 - \exp(-\alpha y), \quad \text{for } y \geq 0,$$

and $\widetilde{F}_\alpha(y) = 0$ for $y < 0$. *Hint*: Start with $\Pr(Y \leq y)$ an rearrange things so that you isolate X on the left hand side of the inequality, then use (a).
(c) Compute the expectation and variance of Y. *Hint*: Look back at hw2 Ex. 5.
(d) Let $Y_i = \log(X_i/x_{\min})$ for $i = 1, \ldots, n$. Explain why $\bar{Y}_n = (1/n)\sum_{i=1}^n Y_i \to_p 1/\alpha$.
(e) Explain why $\widehat{\alpha}_n$ is consistent for α.

Exercise 4. (This exercise is a continuation of Ex. 7 in hw2. For completeness, that exercise is restated here, in slightly rewritten form.) 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 $M_n = \max_{i \leq n} X_i = \max\{X_1, \ldots, X_n\}$ be the largest of the observations. In this exercise you need to know that

$$\lim_{n \to \infty} (1 + x/n)^n = \exp(x),$$

for any real number x.

(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
$$\Pr(M_n \le x) = F(x)^n.$$
(c) Consider the sequence $(Z_n)_{n\ge1}$ of random variables defined by
$$Z_n = n(\theta - M_n), \quad \text{for } n = 1, 2, \ldots.$$
Show that
$$\Pr(Z_n \le z) = 1 - \big(1 - \frac{z}{\theta n}\big)^n,$$
for $z \ge 0$ and $\Pr(Z_n \le z) = 0$ for $z < 0$.

(d) Show that Z_n converges in distribution, and find the expectation and variance of the limiting distribution. That is, if $Z_n \to_d Z$, say, find $\mathrm{E}[Z]$ and $\mathrm{Var}(Z)$.

Exercise 5. Let $Y_1, \ldots, Y_n$ be i.i.d. random variables with $\mathrm{E}[Y_1] = \theta$ and $\mathrm{Var}(Y_1) = \tau^2 = 2.34$, which means that the expectation is unknown while the variance is known. A natural estimator for θ is $\widehat{\theta}_n = \bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$ (why?).

(a) Identify what distribution $\sqrt{n}(\widehat{\theta}_n - \theta)/\tau$ converges to.

(b) Let $L_n = L(Y_1, \ldots, Y_n)$ and $U_n = U(Y_1, \ldots, Y_n)$ be the lower and upper bound of a random interval $[L_n, U_n]$ that contains θ with probability approximately equal to 90%. Find expressions for L_n and U_n. *Hint*: To find the 0.95 quantile of the standard normal distribution in Matlab, run `norminv(0.95)`.

(c) Read the dataset `hw5_data.txt` into Matlab, and compute a realisation of the interval you found in (b). Here is some code that you can use.

```
cd("~/your path");
data = readmatrix("hw5_data.txt");
dimdata = size(data);
n = dimdata(1);
y = data(:,1);
```

(d) In this exercise we check by way of simulation how accurate the approximation in (b) is for $n = 50$. To do so we sample 50 random numbers from the Gamma distribution, compute the interval, and check whether our interval contains the expectation of the Gamma distribution. This we do 1 000 times, thus producing 1 000 intervals, then we count how many of these intervals contain the expectation of the Gamma distribution. The pdf of the Gamma distribution with parameters $a > 0$ and $b > 0$ is
$$f_{a,b}(y) = \frac{1}{b^a\Gamma(a)} y^{a-1}\exp(-y/b), \quad \text{for } y > 0,$$
and $f_{a,b}(y) = 0$ otherwise. Here $\Gamma(x)$ is the Gamma-function. If $Y \sim f_{a,b}(y)$, then $\mathrm{E}[Y] = ab$ and $\mathrm{Var}(Y) = ab^2$. Here is code where you have to fill in some details.

```
a = 2*2.34;
b = 1/sqrt(2);
n = 50;
sims = 1000
contains = zeros(1,sims);
for i = 1:sims
```

```
    y = gamrnd(a,b,1,n);
    Ln = % fill in
    Un = % fill in

    contains(i) = 1*((Ln <= a*b)&(a*b <= Un));
end

mean(contains)
```

Exercise 6. Simulate a conditional probability. Consider a fair die: Its possible outcomes are $\{1, 2, 3, 4, 5, 6\}$, and it being fair means that $\Pr(X = j) = 1/6$ for $j = 1, 2, 3, 4, 5, 6$, where X is the random variable we associate with a roll of a die. Let $A = \{1, 2, 3\}$, and $B = \{1, 3, 5\}$ be two events.

(a) Compute $\Pr(A \mid B)$.

(b) In the script below we use simulations to check the answer from (a). In other words, we *estimate* $\Pr(A \mid B)$ on simulated data.

```
x = datasample(1:6,100,"Replace",true);
inA = zeros(1,length(x)); inB = zeros(1,length(x));
for i = 1:length(x)
    if x(i) <= 3
        inA(i) = 1;
    end
    if (x(i) == 1)|(x(i) == 3)|(x(i) == 5)
        inB(i) = 1;
    end
end

mean(inA.*inB)/mean(inB)
```

The function `datasample` samples from the numbers $\{1, 2, 3, 4, 5, 6\}$ with uniform probability. The symbol | reads 'or', as in the union of two sets. Understand the code, run it a few times, and check you answer from (a).

References

Wooldridge, J. M. (2019). *Introductory Econometrics: A Modern Approach. Seventh Edition.* Cengage Learning, Boston, MA.

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