HOMEWORK 3
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. For this exercise you need to know that $\sum_{k=0}^{\infty} x^k/k! = \exp(x)$. We write $X \sim \text{Poisson}(\theta)$ to indicate that the random variable X has the Poisson distribution with parameter $\theta > 0$. The pmf of this distribution is

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

and $f(x) = 0$ elsewhere.

(a) Suppose that $X \sim \text{Poisson}(\theta)$, show that $\mathrm{E}\, X = \theta$ and $\mathrm{Var}\, X = \theta$. *Hint*: Establish that

$$\mathrm{E}\, X^2 = \sum_{x=0}^{\infty} x^2 \frac{1}{x!}\theta^x \exp(-\theta) = \sum_{x=0}^{\infty} (x+1)\frac{1}{x!}\theta^{x+1} \exp(-\theta).$$

(b) Suppose $X_1, \ldots, X_n$ are independent Poisson rv's with expectation θ. Find an expression for the log-likelihood function,

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

(c) Find an expression for the first derivative of $\ell_n(\theta)$, and use this expression to find the maximum likelihood estimator, say $\widehat{\theta}_n(X_1, \ldots, X_n)$ (which we'll often just write as $\widehat{\theta}_n$).

(d) Show that the maximum likelihood estimator is unbiased for θ, that is

$$\mathrm{E}\,[\,\widehat{\theta}_n(X_1, \ldots, X_n)] = \theta,$$

and find also the variance of $\widehat{\theta}_n(X_1, \ldots, X_n)$.

(e) Here is some data that we assume stem from $n = 10$ independent draws from a Poisson distribution with parameter $\theta > 0$.

$$(x_1, \ldots, x_{10}) = (2, 3, 4, 1, 4, 1, 1, 0, 0, 2).$$

Estimate θ, that is compute $\widehat{\theta}_n(x_1, \ldots, x_n)$.

(f) Here is Matlab code where we sample independently n times from a Poisson distribution with parameter $\theta > 0$. Implement the code and draw a histogram.

Date: September 8, 2020.

```
x = poissrnd(theta,1,n)
histogram(x,"Normalization","pdf")
```

Exercise 2. The Pareto distribution is often used to model the distribution of wealth in a society. The pdf of the Pareto distribution is

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

and $f(x) = 0$ for $x < x_{\min}$, with $\alpha > 0$ and $x_{\min} > 0$. Until exercise (i) we'll assume that $x_{\min}$ is a known number.

(a) Find the cdf $F_\alpha(x)$ of the Pareto distribution.
(b) Suppose $X \sim F_\alpha$. Assume that $\alpha > 1$, find an expression for $\mathrm{E}\, X$.
(c) Find also the variance $\mathrm{Var}\, X$, when $\alpha > 2$.
(d) Suppose that $X_1, \ldots, X_n$ are i.i.d. samples from the Pareto distribution. Write down an expression for the log-likelihood function

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

(e) Find an expression for the first derivative of $\ell_n(\alpha)$, and use this expression to find the maximum likelihood estimator for α, say $\widehat{\alpha}_n(X_1, \ldots, X_n)$.
(f) Here is a dataset on the wealth in millions of kroner for $n = 10$ individuals,

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0.58	1.44	1.03	23.75	0.59	2.13	3.39	0.80	1.28	3.89

We assume that these data are the outcomes of 10 independent draws from a Pareto distribution with $x_{\min} = 0.5$, and $\alpha > 0$ unknown. Use the maximum likelihood estimator you found in (e) to estimate α.
(g) The pth quantile (or $100p$th percentile) of a distribution with cdf F is the solution, say x_p, to $F(x) = p$. Thus $F^{-1}(p) = x_p$. Find the inverse function F_α^{-1} of the cdf you found in (a).
(h) Use the maximum likelihood estimate of α from (f) to estimate the 90th percentile $x_{0.9}$ of the wealth distribution in the population from which the data in (f) stem.
(i) Suppose that $x_{\min}$ is also unknown. Based on an i.i.d. sample $X_1, \ldots, X_n$ from the pdf in (1), try to find the maximum likelihood estimators for $x_{\min}$ and α.

Exercise 3. Suppose that $Y_1, \ldots, Y_n$ are i.i.d. random variables from the normal distribution with expectation μ and variance $\sigma^2 > 0$. In this exercise we take both μ and σ^2 to be unknown, and want to estimate these using the maximum likelihood estimator. Recall that the pdf of the normal distribution is

$$f(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(y-\mu)^2\},$$

for $y \in (-\infty, \infty)$.

(a) Write down an expression for the log-likelihood function

$$\ell_n(\mu, \sigma^2) = \sum_{i=1}^{n} \log f(Y_i; \mu, \sigma^2).$$

(b) Show that the partial derivaties are

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

and

$$\frac{\partial}{\partial \sigma^2} \ell_n(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (Y_i - \mu)^2,$$

and use this to find the maximum likelihood estimators $\widehat{\mu}_n$ and $\widehat{\sigma}_n^2$ for μ and σ^2.

(c) Show that $\widehat{\mu}_n$ is an unbiased estimator for μ.

(d) Show that $\widehat{\sigma}_n^2$ is biased for σ^2, i.e. that $\mathrm{E}\,\widehat{\sigma}_n^2 \neq \sigma^2$.

(e) Construct an estimator $\widetilde{\sigma}_n^2 = b(n)\widehat{\sigma}_n^2$ that is unbiased for σ^2.

Exercise 4. Assume that a test for Covid-19 is such that it gives the correct result in 99 percent of the cases when a person is infected, and the correct result in 96 percent of the cases when a person is not infected (these are called the *specificity* and *sensitivity* of a test, respectively). Assume also that 34 out of 100 000 people in Oslo are infected with Covid-19. Of all the people in Oslo, a person is chosen at random and tested.

(a) The test is positive. What is the probability that this person is truly infected?

(b) Check you answer from (a) by estimating the probability on simulated data. Understand, implement, and run the Matlab code below a few times.

```
sims = 10^5;
sick = binornd(1,34/10^5,1,sims);
positive = zeros(1,sims);
for i = 1:sims
    if sick(i) == 1
        positive(i) = binornd(1,0.99,1,1);
    else
        positive(i) = binornd(1,0.04,1,1);
    end
end

pr_hat = mean(sick.*positive)/mean(positive);
pr = 0.99*34/(0.99*34 + 0.04*99966);

fprintf("%f should be close to %f\n",[pr_hat,pr])
```

(c) In the 'real Oslo', why does your answer from (a) not mean that a person who tests positive is most probably healthy?

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