PROPOSED SOLUTIONS
HOMEWORK 3
GRA6039 ECONOMETRICS WITH PROGRAMMING
AUTUMN 2020

EMIL A. STOLTENBERG

Solutions to Ex. 1. The random variable X has the Poisson distribution with parameter $\theta > 0$. We write $X \sim \text{Poisson}(\theta)$. 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, with $\theta > 0$. **(a)** The expectation of X is

$$\begin{aligned}
\mathrm{E}\,[X] &= \sum_{x=0}^{\infty} x f_\theta(x) = \sum_{x=0}^{\infty} x \frac{1}{x!}\theta^x \exp(-\theta) = \sum_{x=1}^{\infty} x \frac{1}{x!}\theta^x \exp(-\theta) = \sum_{x=1}^{\infty} \frac{1}{(x-1)!}\theta^x \exp(-\theta) \\
&= \sum_{x=0}^{\infty} \frac{1}{x!}\theta^{x+1} \exp(-\theta) = \theta \sum_{x=0}^{\infty} \frac{1}{x!}\theta^x \exp(-\theta) = \theta,
\end{aligned}$$

where the last equality follows because $\sum_{x=0}^{\infty}(1/x!)\theta^x \exp(-\theta) = 1$ since $f_\theta(x)$ is a pmf. **(b)** To find the variance of X we'll use that $\mathrm{Var}(X) = \mathrm{E}\,[X^2] - (\mathrm{E}\,[X])^2$, so we need to find $\mathrm{E}\,[X^2]$:

$$\begin{aligned}
\mathrm{E}\,[X^2] &= \sum_{x=0}^{\infty} x^2 f_\theta(x) = \sum_{x=0}^{\infty} x^2 \frac{1}{x!}\theta^x \exp(-\theta) = \sum_{x=1}^{\infty} x \frac{1}{(x-1)!}\theta^x \exp(-\theta) \\
&= \sum_{x=0}^{\infty} (x+1) \frac{1}{x!}\theta^{x+1} \exp(-\theta) = \theta\Big\{\sum_{x=0}^{\infty} x \frac{1}{x!}\theta^x \exp(-\theta) + \sum_{x=0}^{\infty} \frac{1}{x!}\theta^x \exp(-\theta)\Big\} \\
&= \theta\Big\{\mathrm{E}\,[X] + \sum_{x=0}^{\infty} f_\theta(x)\Big\} = \theta(\theta + 1) = \theta^2 + \theta,
\end{aligned}$$

then

$$\mathrm{Var}(X) = \mathrm{E}\,[X^2] - (\mathrm{E}\,[X])^2 = \theta^2 + \theta - \theta^2 = \theta.$$

Date: September 14, 2020.

(b) Let $X_1, \dots, X_n$ be i.i.d. Poisson with expectation $\theta > 0$. The log-likelihood function is

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

(c) The first derivative of $\ell_n(\theta)$ is

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_n(\theta) = \frac{1}{\theta}\sum_{i=1}^n X_i - n,$$

and when we set this equal to zero and solve for θ we find the maximum likelihood estimator

$$\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n.$$

(d) The expectation of $\widehat{\theta}_n$ is (using Prop. 2.3 in the Lecture notes)

$$\mathrm{E}\,[\widehat{\theta}_n] = \mathrm{E}\,\big[\frac{1}{n}\sum_{i=1}^n X_i\big] = \frac{1}{n}\sum_{i=1}^n \mathrm{E}\,[X_i] = \frac{1}{n}\sum_{i=1}^n \theta = \frac{n}{n}\theta = \theta.$$

Since the $X_1, \dots, X_n$ are independent, $\mathrm{Cov}(X_i, X_j) = 0$ whenever $i \neq j$ (see HW2, Ex. 3(e)), so

$$\mathrm{Var}(\widehat{\theta}_n) = \mathrm{Var}\big(\frac{1}{n}\sum_{i=1}^n X_i\big) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) + \frac{2}{n^2}\sum_{1\le i<j\le n}\mathrm{Cov}(X_i, X_j)$$
$$= \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^n \theta = \frac{n}{n^2}\theta = \frac{\theta}{n}.$$

This is easier to see if $n = 2$. Then (see HW2, Ex. 3(f))

$$\mathrm{Var}\big(\frac{1}{2}\sum_{i=1}^2 X_i\big) = \mathrm{Var}\big(\frac{X_1}{2} + \frac{X_2}{2}\big) = \frac{1}{4}\mathrm{Var}(X_1) + \frac{1}{4}\mathrm{Var}(X_2) + \frac{2}{4}\mathrm{Cov}(X_1, X_2),$$

and $\mathrm{Cov}(X_1, X_2) = 1$ when X_1 and X_2 are independent. **(e)** Here is a Matlab script where we estimate θ

```
x = [2,3,4,1,4,1,1,0,0,2];
mean(x) % = 1.8
```

thus $\widehat{\theta}_n(x_1, \dots, x_n) = \widehat{\theta}_n(2,3,4,1,4,1,1,0,0,2) = 1.8$, this is our estimate for θ. **(f)** Use the following Matlab code to make the histogram in Figure 1. Here we set $\theta = 2.34$ and $n = 1000$.

```
x = poissrnd(2.34,1,1000)
histogram(x,"Normalization","pdf")
```

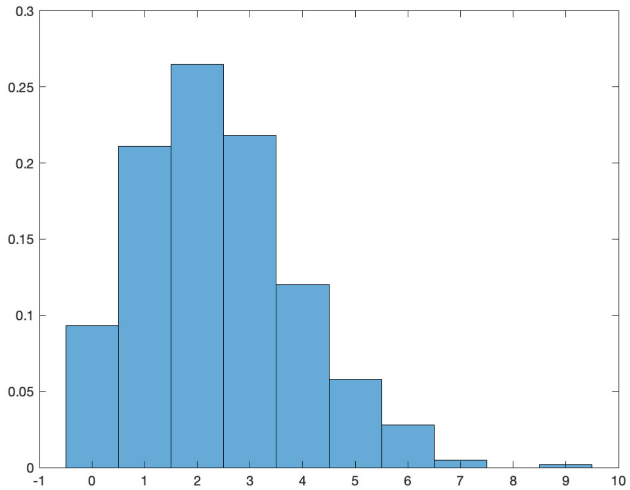

FIGURE 1. A density histogram of $n = 1000$ independent draws from a Poisson distribution with $\theta = 2.34$

Solutions to Ex. 2. 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)** For $x < x_{\min}$, the cdf $F_\alpha(x) = 0$. For $x > x_{\min}$, the cdf is

$$F_\alpha(x) = \int_{x_{\min}}^{x} \frac{\alpha x_{\min}^\alpha}{y^{\alpha+1}}\,\mathrm{d}y = -\Big|_{x_{\min}}^{x} \frac{x_{\min}^\alpha}{y^\alpha} = -\frac{x_{\min}^\alpha}{x^\alpha} + \frac{x_{\min}^\alpha}{x_{\min}^\alpha} = 1 - \frac{x_{\min}^\alpha}{x^\alpha}.$$

The cdf of the Pareto distribution is therefore,

$$F_\alpha(x) = \begin{cases} 1 - (x_{\min}/x)^\alpha, & x \geq x_{\min}, \\ 0, & x < x_{\min}. \end{cases}$$

(b) Assume that $\alpha > 1$. The expectation if X is

$$\begin{aligned} \mathrm{E}[X] &= \int_{x_{\min}}^{\infty} x \frac{\alpha x_{\min}^\alpha}{x^{\alpha+1}}\,\mathrm{d}x = \int_{x_{\min}}^{\infty} \frac{\alpha x_{\min}^\alpha}{x^\alpha}\,\mathrm{d}x \\ &= -\Big|_{x_{\min}}^{\infty} \frac{\alpha x_{\min}^\alpha}{\alpha - 1} \frac{1}{x^{\alpha-1}} = \frac{\alpha x_{\min}^\alpha}{\alpha - 1} \frac{1}{x_{\min}^{\alpha-1}} = \frac{\alpha x_{\min}}{\alpha - 1}. \end{aligned}$$

When $\alpha < 1$, the expectation is infinite. **(c)** Assume that $\alpha > 2$. Again we use $\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$, and compute

$$\begin{aligned} \mathrm{E}[X^2] &= \int_{x_{\min}}^{\infty} x^2 \frac{\alpha x_{\min}^\alpha}{x^{\alpha+1}}\,\mathrm{d}x = \int_{x_{\min}}^{\infty} \frac{\alpha x_{\min}^\alpha}{x^{\alpha-1}}\,\mathrm{d}x \\ &= -\Big|_{x_{\min}}^{\infty} \frac{\alpha x_{\min}^\alpha}{\alpha - 2} \frac{1}{x^{\alpha-2}} = \frac{\alpha x_{\min}^\alpha}{\alpha - 2} \frac{1}{x_{\min}^{\alpha-2}} = \frac{\alpha x_{\min}^2}{\alpha - 2}. \end{aligned}$$

Then

$$\mathrm{Var}(X) = \mathrm{E}\,[X^2] - (\mathrm{E}\,[X])^2 = \frac{\alpha x_{\min}^2}{\alpha - 2} - \frac{\alpha^2 x_{\min}^2}{(\alpha-1)^2} = \frac{\alpha x_{\min}^2}{(\alpha-2)(\alpha-1)^2}.$$

(d) Suppose that $X_1, \ldots, X_n$ are i.i.d. samples from the Pareto distribution. The log-likelihood function is

$$\begin{aligned}\ell_n(\alpha) &= \sum_{i=1}^n \log f_\alpha(X_i) = \sum_{i=1}^n \{\log(\alpha) + \alpha \log(x_{\min}) - (\alpha+1)\log(X_i)\} \\ &= \sum_{i=1}^n \{\log(\alpha) - \alpha \log(X_i/x_{\min}) - \log(X_i)\} \\ &= n\log(\alpha) - \alpha \sum_{i=1}^n \log(X_i/x_{\min}) - \sum_{i=1}^n \log(X_i).\end{aligned}$$

(e) Differentiate $\ell_n(\alpha)$ with respect to α and set this equal to zero,

$$\frac{\mathrm{d}}{\mathrm{d}\alpha}\ell_n(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^n \log(X_i/x_{\min}) = 0.$$

Solve for α and we find the maximum likelihood estimator

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

(f) Here is a Matlab script where we use the estimator $\widehat{\alpha}_n$ to estimate α

```
x = [0.58,1.44,1.03,23.75,0.59,2.13,3.39,0.80,1.28,3.89];
xmin = 0.5;
1/mean(log(x/xmin))
```

Our estimate of α is 0.7825. **(g)** The inverse of the cdf F_α that we found in (a) is

$$F_\alpha^{-1}(u) = \frac{x_{\min}}{(1-u)^{1/\alpha}}.$$

(h) A natural way of estimating the 90th percentile $x_{0.9}$ of the wealth distribution in the population from which the data in (f) stem, is to plug the maximum likelihood estimator $\widehat{\alpha}_n$ into $F_\alpha^{-1}(u)$, then

$$\widehat{x}_{0.9} = F_{\widehat{\alpha}_n}^{-1}(0.9) = \frac{x_{\min}}{(1-0.9)^{1/\widehat{\alpha}_n}} = \frac{0.5}{(0.1)^{1/0.7825}} = 9.4826.$$

This means that according to our estimates the 10 percent most wealthy have a wealth of 9.48 millions or more.

(i) Now suppose that $x_{\min}$ is also unknown. Looking at the likelihood function,

$$\begin{aligned}\ell_n(\alpha, x_{\min}) &= n\log(\alpha) - \alpha\sum_{i=1}^n \log(X_i/x_{\min}) - \sum_{i=1}^n \log(X_i) \\ &= n\log(\alpha) + n\alpha\log(x_{\min}) - (\alpha+1)\sum_{i=1}^n \log(X_i),\end{aligned}$$

we see that $\ell_n(\alpha, x_{\min})$ is increasing in $x_{\min}$. But since $X_i \geq x_{\min}$ for all i, $x_{\min}$ cannot be bigger than the smallest X_i. Therefore, the maximum likelihood estimators become

$$\widehat{x}_{\min} = \min_{i \leq n} X_i = \min\{X_1, \ldots, X_n\},$$
$$\widehat{\alpha}_n = \frac{n}{\sum_{i=1}^n \log(X_i/\widehat{x}_{\min})}.$$

Solutions to Ex. 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) The log-likelihood function is

$$\begin{aligned}\ell_n(\mu, \sigma^2) &= \sum_{i=1}^n \log f(y_i; \mu, \sigma^2) = \sum_{i=1}^n \{-\frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(Y_i - \mu)^2 - \log\sqrt{2\pi}\} \\ &= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \mu)^2 - n\log(\sqrt{2\pi}).\end{aligned}$$

(b) Differentiate with respect to μ and with respect to σ^2, and set both partial derivatives equal to zero,

$$\frac{\partial}{\partial\mu}\ell_n(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^n (Y_i - \mu) = 0,$$
$$\frac{\partial}{\partial\sigma^2}\ell_n(\mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (Y_i - \mu)^2 = 0.$$

This is a system of two equations in two unknowns, the unknowns being μ and σ^2. The solution gives the maximum likelihood estimators, they are

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n Y_i = \bar{Y}_n, \quad \text{and} \quad \widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

(c) An unbiased estimator is an estimator whose expectation equals what it is an estimator for. That is, if $\mathrm{E}[\widehat{\mu}_n] = \mu$, then we call $\widehat{\mu}_n$ unbiased for μ, or simply unbiased. Using Prop. 2.3 in the lecture notes we see that $\widehat{\mu}_n$ is unbiased, because.

$$\mathrm{E}[\widehat{\mu}_n] = \mathrm{E}\big(\frac{1}{n}\sum_{i=1}^n Y_i\big) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[Y_i] = \frac{1}{n}\sum_{i=1}^n \mu = \mu.$$

(d) To show that $\widehat{\sigma}_n^2$ is biased for σ^2, we must show that $\mathrm{E}[\widehat{\sigma}_n^2]$ does not equal σ^2. To compute the expectation of $\widehat{\sigma}_n^2$ let's first write

$$\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2 = \frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}_n^2.$$

By Prop. 2.3 (linearity of the expectation) we have that

$$\mathrm{E}\,[\widehat{\sigma}_n^2] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\,[Y_i^2] - \mathrm{E}\,[\bar{Y}_n^2].$$

We can compute these two expectations using the $\mathrm{Var}(Y) = \mathrm{E}\,[Y^2] - (\mathrm{E}\,[Y])^2$ formula. For each i

$$\mathrm{E}\,[Y_i^2] = \mathrm{Var}\,(Y_i) + (\mathrm{E}\,[Y_i])^2 = \sigma^2 + \mu^2.$$

Since the $Y_1, \ldots, Y_n$ are independent

$$\mathrm{Var}(\bar{Y}_n) = \frac{\sigma^2}{n}.$$

Therefore

$$\mathrm{E}\,[\bar{Y}_n^2] = \mathrm{Var}\,(\bar{Y}_n) + (\mathrm{E}\,[\bar{Y}_n])^2 = \frac{\sigma^2}{n} + \mu^2.$$

Inserting this is our expression for $\mathrm{E}\,[\widehat{\sigma}_n^2]$ we get

$$\mathrm{E}\,[\widehat{\sigma}_n^2] = \frac{1}{n}\sum_{i=1}^{n}(\sigma^2+\mu^2) - \frac{\sigma^2}{n} - \mu^2 = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2,$$

which shows that $\widehat{\sigma}_n^2$ is biased. **(e)** We now construct an estimator that is unbiased for σ^2. Since $\mathrm{E}\,[\widehat{\sigma}_n^2] = (n-1)\sigma^2/n$, we see that the estimator

$$\widetilde{\sigma}_n^2 = \frac{n}{n-1}\widehat{\sigma}_n^2,$$

is unbiased, because

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

Notice that in this exercise we only used that the $Y_1, \ldots, Y_n$ we i.i.d. with expectation μ and variance σ^2. We did not use that they are normally distributed. Our derivation of the estimator $\widetilde{\sigma}_n^2$ is the reason for the empirical variance of a sample $X_1, \ldots, X_n$ being defined as

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

The $n-1$ in the denominator makes s_X^2 unbiased for the true variance!

Solutions to Ex. 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) Let '+' indicate positive test, and 'sick' indicate that the person is truly infected. Then Bayes rule gives

$$\Pr(\text{sick} \mid +) = \frac{\Pr(+ \mid \text{sick})\Pr(\text{sick})}{\Pr(+)}.$$

There are two possibilities: a person is either sick or not sick, the law of total probability therefore gives,

$$\Pr(+) = \Pr(+ \mid \text{sick})\Pr(\text{sick}) + \Pr(+ \mid \text{not sick})\Pr(\text{not sick}),$$

so that

$$\Pr(\text{sick} \mid +) = \frac{\Pr(+ \mid \text{sick})\Pr(\text{sick})}{\Pr(+ \mid \text{sick})\Pr(\text{sick}) + \Pr(+ \mid \text{not sick})\Pr(\text{not sick})}.$$

The numbers given in the text are $\Pr(+ \mid \text{sick}) = 0.99$ (the specificity of the test), $\Pr(+ \mid \text{not sick}) = 0.04$ (which is 1 minus the sensitivity of the test), and $\Pr(\text{sick}) = 34/10^5$, so that $\Pr(\text{not sick}) = 1 - 34/10^5 = 99966/10^5$. Then

$$\Pr(\text{sick} \mid +) = \frac{0.99 \times \frac{34}{10^5}}{0.99 \times \frac{34}{10^5} + 0.04 \times \frac{99966}{10^5}} = \frac{0.99 \times 34}{0.99 \times 34 + 0.04 \times 99966} = 0.00835.$$

(b) Run the this Matlab script a few times to estimate $\Pr(\text{sick} \mid +) = 0.00835$ on simulated data.

```
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? The most important reason for this is that the people who get's tested are *not* randomly selected. They have symptoms. Thus, in the population of people who actually gets tested, the probability $\Pr(\text{sick})$ is much higher than $34/10^5$. This, in turn makes the probability $\Pr(\text{sick} \mid +)$ much higher than what we found in (a). Also, but less important, the numbers for the sensitivity and specificity of the test are just numbers I made up. Perhaps the test is better than what we postulated in this exercise?

Here is an article (in Norwegian) about the sensitivity and specificity of tests for Covid-19. https://www.faktisk.no/artikler/r8q/er-14-av-15-positive-koronaprover-falske

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