Homework 4

Spring 2023

KEY

Instructions

- This homework is due in Gradescope on Wednesday May 3 by midnight PST.
- Please answer the following questions in the order in which they are posed.
- Don't forget to knit the document frequently to make sure there are no compilation errors.
- When you are done, download the PDF file as instructed in section and submit it in Gradescope.

Exercises

1. (MOM vs.MLE) Suppose $X_1, X_2, \ldots, X_n$ are independently drawn from the PDF

$$f(x) = (\theta_0 + 1)x^{\theta_0} \quad 0 \leq x \leq 1$$

where the parameter $\theta_0 > -1$ so the PDF is non-zero and integrates to 1.

a. Derive $\widehat{\theta}_0^{mom}$, the method of moments estimator of θ_0.

We have one parameter so we must take a single expectation to set up our MoM system of equations. We may first find $E[X]$:

$$E[X] = \int_0^1 xf(x)dx = \int_0^1 x(\theta_0+1)x^{\theta_0}dx = \frac{(\theta_0+1)x^{\theta_0+2}}{\theta_0+2}|_0^1 = \frac{\theta_0+1}{\theta_0+2}$$

So we are solving the equation:

$$\bar{X} = \frac{\theta_0+1}{\theta_0+2} \Rightarrow \bar{X}(\theta_0+2) = \theta_0+1$$

Doing some algebra:

$$\begin{aligned}
\bar{X} &= \frac{\theta_0+1}{\theta_0+2} \\
\bar{X}(\theta_0+2) &= \theta_0+1 \\
\bar{X}\theta_0 - \theta_0 &= -2\bar{X}+1 \\
\hat{\theta}_0^{MoM} &= \frac{-2\bar{X}+1}{\bar{X}-1}
\end{aligned}$$

Thus we have $\hat{\theta}_0^{MoM} = \frac{-2\bar{X}+1}{\bar{X}-1}$

b. Derive $\widehat{\theta}_0^{mle}$, the maximum likelihood estimator of θ_0. Be sure to show
   - likelihood function
   - log-likelihood function
   - first derivative condition
   - second derivative test

In this case if we have $X_1, \ldots, X_n \overset{iid}{\sim} f(x)$ then our likelihood function will take the form:

$$L(\theta_0) = \Pi_{i=1}^n f(x_i) = \Pi_{i=1}^n (\theta_0 + 1) x_i^{\theta_0} = (\theta_0 + 1)^n \Pi_{i=1}^n x_i^{\theta_0}$$

From here we see our log-likelihood will take the form:

$$\ell(\theta_0) = ln\left[(\theta_0 + 1)^n \Pi_{i=1}^n x_i^{\theta_0}\right] = nln(\theta_0 + 1) + \theta_0 \sum_{i=1}^n log(x_i)$$

Now that we have our log-likelihood we may begin our optimization steps. First we may take the first derivative of our log-likelihood and set it to zero.

$$\ell'(\theta_0) = \frac{n}{\theta_0 + 1} + \sum_{i=1}^n log(x_i) = 0$$

Solving for θ_0 we yield:

$$\begin{aligned}
\frac{n}{\theta_0 + 1} + \sum_{i=1}^n log(x_i) &= 0 \\
n + (\theta_0 + 1)\sum_{i=1}^n log(x_i) &= 0 \\
(\theta_0 + 1)\sum_{i=1}^n log(x_i) &= -n \\
\hat{\theta}_0^{MLE} &= \frac{-n - \sum_{i=1}^n log(x_i)}{\sum_{i=1}^n log(x_i)}
\end{aligned}$$

Thus we have our MLE as $\hat{\theta}_0^{MLE} = \frac{-n - \sum_{i=1}^n log(x_i)}{\sum_{i=1}^n log(x_i)}$

Checking our second derivative to make sure this is a maximum we yield:

$$\ell''(\theta_0) = \frac{-n}{(\theta_0 + 1)^2} < 0$$

Thus this is a maximum. Usually we would also test the boundary condition ($\theta_0 > -1$), however since this is an open boundary, and $\lambda \neq -1$, testing the second derivative is sufficient (think about why).

c. Find $\widehat{\theta}_0^{mom}$ based on the sample below.

```
x <- c(0.90, 0.78, 0.93, 0.64, 0.45, 0.85, 0.75, 0.93, 0.98, 0.78)
```

Plugging in these values into our MoM equation we yield our estimate as:

```
x_bar <- mean(x)
mom <- (-2*x_bar+1)/(x_bar-1)
mom
```

[1] 2.975124

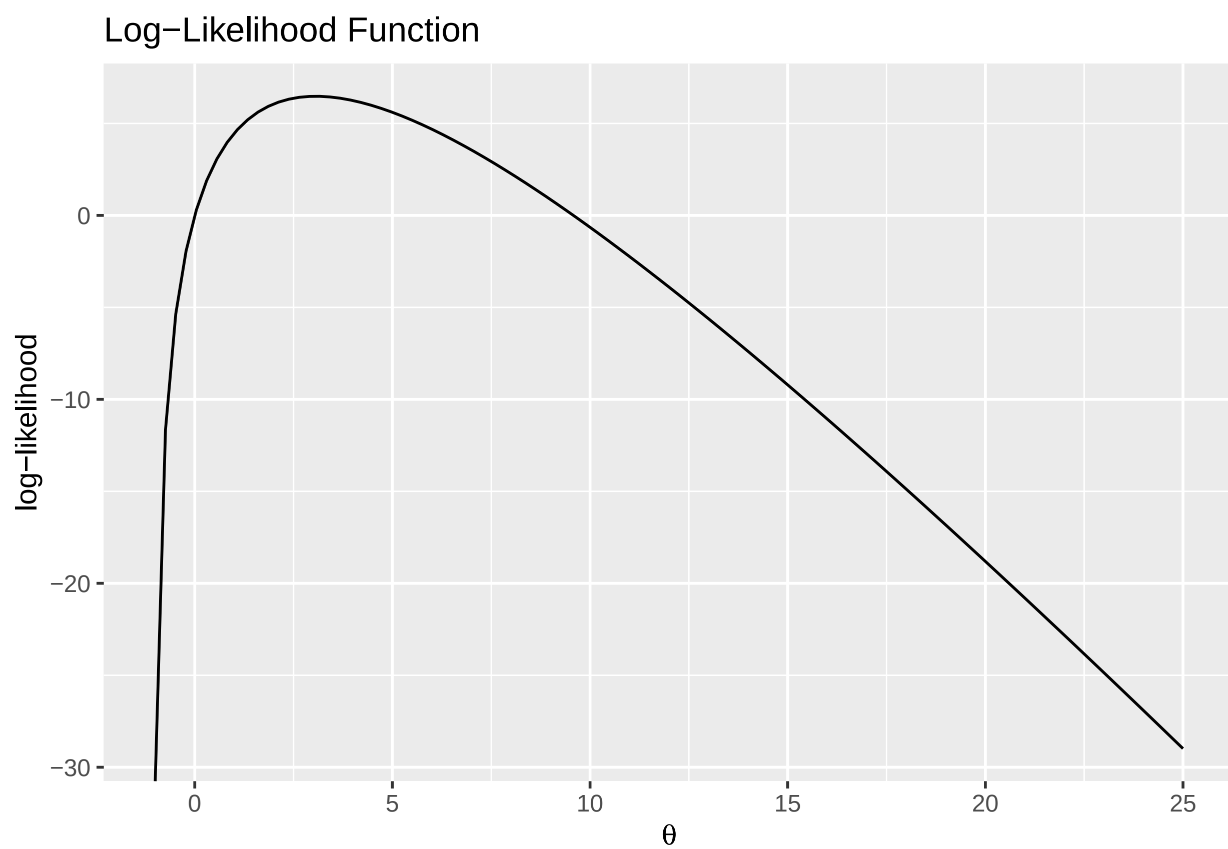
d. Make a plot of the log-likelihood function for the data from part c.and calculate $\widehat{\theta}_0^{mle}$.

First we may plot the log-likelihood function:

$$nln(\theta_0+1)+\theta_0\sum_{i=1}^{n}log(x_i)$$

```
n <- length(x)
log_like <- function(theta){

  n*log(theta+1) + theta*sum(log(x))
}
ggplot()+geom_function(fun = log_like)+
  xlim(c(-1,25))+
  labs(x = expression(theta),
       y = "log-likelihood",
       title = "Log-Likelihood Function")
```



We will have our MLE will be for this data:

```
(-n-sum(log(x)))/sum(log(x))
```

```
## [1] 3.060711
```

2. (Light bulbs) A set of cheap light bulbs have a lifetime (in hours) which is exponentially distributed with unknown rate λ_0:

$$f(x)=\lambda_0 exp(-\lambda_0\, x),\quad 0<x$$

Choosing a random sample of ten light bulbs, they are turned on simultaneously and observed for 48 hours. During this period, six bulbs went out, at times $x_1, x_2, \ldots, x_6$. At the end of the experiment, four light bulbs were still working.

a. Derive the likelihood function $L(\lambda)$. (Hint: we can model this as observing values for $X_1, X_2, \ldots, X_6$ which are exponential random variables and Y_1, Y_2, Y_3, Y_4 which are Bernoulli random variables which are 1 or 0 depending on whether the lifetime X is larger than 48 or not. The likelihood function is the product of the six exponential density functions and the four Bernoulli PMF.)

We have that for the 6 lightbulbs that went out before 48 hours, we observed times $X_1, \ldots, X_6$ that are $exp(\lambda_0)$ RV. For the lightbulbs that did not go out, we simply know the information that they have a failure time of greater than 48.

Per the hint we can model the 4 lightbulbs that did not go out as bernoulli RV, where a success is staying on, and a failure is turning off. In this case we viewed 4 successes, each with probability $P(X \geq 48)$. By properties of exponential RV and their CDF this is equal to $exp(-48\lambda_0)$.

Thus we have the likelihood:

$$L(\lambda_0) = [\prod_{i=1}^{6} \lambda_0 exp(-\lambda_0 x_i)] \prod_{i=1}^{4}(exp(-48\lambda_0)) = \lambda_0^6 exp(-\lambda_0 \sum_{i=1}^{6} x_i) \times (exp(4 \times -48\lambda_0))$$

b. Derive an expression for the MLE of λ_0 showing your work. Verify it is the global maximum of the likelihood function.

For the MLE we may first write the log-likelihood. We have this as:

$$\ell(\lambda_0) = 6ln(\lambda_0) - \lambda_0 \sum_{i=1}^{6} x_i + (4 \times -48\lambda_0))$$

Now we may take the first derivative and set it to 0

$$\ell'(\lambda_0) = \frac{6}{\lambda_0} - \sum_{i=1}^{6} x_i - 192 = 0$$

Solving this we see we yield $\hat{\lambda}_0^{MLE} = \frac{6}{\sum_{i=1}^{6} x_i + 192}$

Checking the second derivative we have:

$$\ell''(\lambda_0) = \frac{-6}{\lambda_0^2} < 0$$

Thus it is a local maximum. Since the boundary is open ($\lambda_0 > 0$), this is sufficient to also show that it is a global maximum.

3. (SLR through origin) Suppose $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(a_i\mu_0, 1)$ where the a_i are known constants. (FYI: This is a Simple Linear Regression (SLR) model which is forced to go through the origin since there is no intercept)

a. Write the likelihood function $L(\mu)$ and also the log-likelihood function $\ell(\mu)$.

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(\frac{-1}{2}(x_i - a_i\mu)) = (2\pi)^{-n/2} exp(\frac{-1}{2\sigma^2} \sum_{i=1}^{n}(x_i - a_i\mu)^2)$$

We have our log-likelihood will thus take the form:

$$\ell(\mu) = (-n/2)log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i - a_i\mu)^2$$

b. Derive an expression for $\widehat{\mu}_0^{mle}$, the MLE of μ_0. (Please show your steps clearly, including the second derivative test)

First we may take the first derivative of our log-likelihood and set it to 0.

$$\ell'(\mu) = \sum_{i=1}^{n} a_i(x_i - a_i\mu) = 0$$

Doing some algebra on the summation:

$$\begin{aligned}
\sum_{i=1}^{n} a_i(x_i - a_i\mu) &= 0 \\
\sum_{i=1}^{n}(a_i x_i - a_i^2\mu) &= 0 \\
\sum_{i=1}^{n}(a_i x_i) - \mu\sum_{i=1}^{n} a_i^2 &= 0 \\
\hat{\mu}^{MLE} &= \frac{\sum_{i=1}^{n}(a_i x_i)}{\sum_{i=1}^{n} a_i^2}
\end{aligned}$$

Checking the second derivative we have:

$$\ell''(\mu) = \sum_{i=1}^{n} -a_i^2 < 0$$

Thus we have that $\widehat{\mu}_0^{mle} = \frac{\sum_{i=1}^{n}(a_i x_i)}{\sum_{i=1}^{n} a_i^2}$

c. Is $\widehat{\mu}_0^{mle}$ an unbiased estimator of μ_0? Show your work.

We may take the expectation of our estimator to check if it is unbiased:

$$E\left[\frac{\sum_{i=1}^{n}(a_i x_i)}{\sum_{i=1}^{n} a_i^2}\right] = \frac{\sum_{i=1}^{n}(a_i E[x_i])}{\sum_{i=1}^{n} a_i^2} = \frac{\sum_{i=1}^{n}(a_i a_i \mu)}{\sum_{i=1}^{n} a_i^2} = \mu\frac{\sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i^2} = u$$

Thus it is an unbiased estimator.

d. Derive the standard error of $\widehat{\mu}_0^{mle}$.

First we may find the variance of our estimator using the fact that we have independent observations.

$$
\begin{aligned}
Var(\frac{\sum_{i=1}^{n}(a_i x_i)}{\sum_{i=1}^{n} a_i^2}) &= \frac{1}{(\sum_{i=1}^{n} a_i^2)^2} Var(\sum_{i=1}^{n}(a_i x_i)) \\
&= \frac{1}{(\sum_{i=1}^{n} a_i^2)^2} \sum_{i=1}^{n} Var(a_i x_i) \\
&= \frac{1}{(\sum_{i=1}^{n} a_i^2)^2} \sum_{i=1}^{n} a_i^2 Var(x_i) \\
&= \frac{1}{(\sum_{i=1}^{n} a_i^2)^2} \sum_{i=1}^{n} a_i^2 \times 1 \\
&= \frac{1}{\sum_{i=1}^{n} a_i^2}
\end{aligned}
$$

Thus we have that:

$$
SE(\widehat{\mu}_0^{mle}) = \sqrt{\frac{1}{\sum_{i=1}^{n} a_i^2}}
$$

4. (Two scientists) A scientist has obtained two random samples: one of size n_1 from an exponential distribution with mean θ_0 and another of size n_2 from an exponential distribution with mean $k\theta_0$, where k is a known number, but θ_0 is unknown.
   The scientist has computed the MLEs for θ_0 - let's call them $\widehat{\theta}_0^{mle1}$ and $\widehat{\theta}_0^{mle2}$ from each of the samples. Now they want a single estimate of θ_0, so they ask two statisticians for advice. One suggests finding the linear combination $a\widehat{\theta}_0^{mle1} + (1-a)\widehat{\theta}_0^{mle2}$, with the smallest variance.
   The other suggests finding the MLE from the combined sample. Show that both methods yield the same answer.

   To help us with the grading, please

- clearly separate the work pertaining to derivation of $\widehat{\theta}_0^{mle1}$ and $\widehat{\theta}_0^{mle2}$
- clearly show your steps (for example, for finding a)
- clearly highlight your final estimators in each case by stating them.

For simplicity, let us consider the rate parametrization form of the exponential distribution. That is let $\lambda_0 = \frac{1}{\theta_0}$, and let $\lambda_1 = \frac{1}{k\theta_0}$. Note that this also means that $\lambda_1 = \frac{\lambda_0}{k} = \frac{1}{k\theta_0}$.

First let us derive $\widehat{\theta}_0^{mle1}$ by finding $\widehat{\lambda}_0^{mle1}$

Letting $X_1, \ldots, X_{n_1} \sim exp(\lambda_0)$, we have our likelihood will take the form

$$
L(\lambda_0) = \prod_{i=1}^{n_1} \lambda_0 exp(-\lambda_0 x_i) = \lambda_0^{n_1} exp(-\lambda_0 \sum_{i=1}^{n_1} x_i)
$$

So we have our log-likelihood will take the form:

$$
\ell(\lambda_0) = n_1 ln(\lambda_0) - \lambda_0 \sum_{i=1}^{n_1} x_i
$$

Taking our first derivative and setting to zero we have:

$$
\ell'(\lambda_0) = \frac{n_1}{\lambda_0} - \sum_{i=1}^{n} x_i = 0
$$

This yield the estimator:

$$\widehat{\lambda}_0^{mle1} = \frac{n_1}{\sum_{i=1}^{n_1} x_i} = \frac{1}{\bar{X}}$$

Taking our second derivative we have:

$$\ell''(\lambda_0) = \frac{-n_1}{\lambda_0^2} < 0$$

Thus we have that $\widehat{\lambda}_0^{mle1}$ is a maximum. From here since we have that $\lambda_0 = \frac{1}{\theta_0}$ we have by the invariance property of MLEs that $\widehat{\theta}_0^{mle1} = \bar{X}$

By a similar process, if we let $Y_1, \ldots, Y_{n_2} \sim exp(\lambda_1)$, we have that $\widehat{\lambda}_1^{mle2} = \frac{1}{\bar{Y}}$. Using the fact that $\lambda_1 = \frac{1}{k\theta_0}$ this yields that $\widehat{\theta}_0^{mle2} = \frac{\bar{Y}}{k}$

Now let us proceed with the two scientists methods:

Scientist 1 (minimize variance):

We have that:

$$Var(a\widehat{\theta}_0^{mle1} + (1-a)\widehat{\theta}_0^{mle2}) = a^2 Var(\widehat{\theta}_0^{mle1}) + (1-a)^2 Var(\widehat{\theta}_0^{mle2})$$

Taking the derivative with respect to a and setting to 0 we yield:

$$2aVar(\widehat{\theta}_0^{mle1}) - 2(1-a)Var(\widehat{\theta}_0^{mle2}) = 0$$

Solving for a we have:

$$\hat{a} = \frac{Var(\widehat{\theta}_0^{mle2})}{Var(\widehat{\theta}_0^{mle1}) + Var(\widehat{\theta}_0^{mle2})}$$

From here lets derive what these variances are. Using that $\lambda_1 = \frac{\lambda_0}{k} = \frac{1}{k\theta_0}$:

$$Var(\widehat{\theta}_0^{mle2}) = Var(\frac{\bar{Y}}{k}) = \frac{1}{k^2} Var(\bar{Y}) = \frac{1}{k^2}\frac{Var(Y)}{n_2} = \frac{1}{n_2 k^2}\frac{1}{\lambda_1^2} = \frac{\theta_0^2 k^2}{n_2 k^2} = \frac{\theta_0^2}{n_2}$$

and

$$Var(\widehat{\theta}_0^{mle1}) = Var(\bar{X}) = \frac{Var(X)}{n_1} = \frac{1}{n_1 \lambda_0^2} = \frac{\theta_0^2}{n_1}$$

Plugging these values in we see that a is equal to:

$$\hat{a} = \frac{\frac{\theta_0^2}{n_2}}{\frac{\theta_0^2}{n_1} + \frac{\theta_0^2}{n_2}} = \frac{n_1}{n_2 + n_1}$$

From here we see the new weighted estimator would take the form:

$$a\widehat{\theta}_0^{mle1} + (1-a)\widehat{\theta}_0^{mle2} = \frac{n_1}{n_2 + n_1}\bar{X} + \frac{n_2}{n_2 + n_1}\frac{\bar{Y}}{k} = \frac{n_1\bar{X} + n_2\frac{\bar{Y}}{k}}{n_2 + n_1}$$

This should make sense as we see that when n_1 is incredible large, our combined MLE tends towards $\bar{X}$ and vice versa for when n_2 large.

Now for scientist 2.

We see that our combined likelihood, if we use $\lambda_0 = \frac{1}{\theta_0}$ (thus giving us that $\lambda_1 = \frac{\lambda_0}{k} = \frac{1}{k\theta_0}$) will take the form:

$$L(\lambda_0) = \left[\prod_{i=1}^{n_1} \lambda_0 exp(-\lambda_0 x_i)\right] \left[\prod_{i=1}^{n_2} \frac{\lambda_0}{k} exp(-\frac{\lambda_0}{k} y_i)\right] = \frac{\lambda_0^{n_1+n_2}}{k^{n_2}} exp(-\lambda_0 \sum_{i=1}^{n_1} x_i) exp(-\frac{\lambda_0}{k} \sum_{i=1}^{n_2} y_i)$$

Thus we have that the log-likelihood will be:

$$\ell(\lambda_0) = (n_1 + n_2) ln(\lambda_0) - (n_2) ln(k) - \lambda_0 \sum_{i=1}^{n_1} x_i - \frac{\lambda_0}{k} \sum_{i=1}^{n_2} y_i$$

Taking the first derivative and setting to zero:

$$\ell'(\lambda_0) = \frac{n_1 + n_2}{\lambda_0} - \sum_{i=1}^{n_1} x_i - \frac{1}{k} \sum_{i=1}^{n_2} y_i$$

This yields that:

$$\widehat{\lambda}_0^{mle} = \frac{n_1 + n_2}{\sum_{i=1}^{n_1} x_i + \frac{1}{k} \sum_{i=1}^{n_2} y_i}$$

Thus we have that:

$$\widehat{\theta}_0^{mle} = \frac{\sum_{i=1}^{n_1} x_i + \frac{1}{k} \sum_{i=1}^{n_2} y_i}{n_1 + n_2} = \frac{n_1 \bar{X} + n_2 \frac{\bar{Y}}{k}}{n_2 + n_1}$$

Which matches scientists one's claim.