Homework 4

Spring 2023

Tanner Huck

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

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

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 can find the method of moments estimator of θ_0 using $E[X] = \bar{x}$.

$$\begin{aligned} E[X] &= \int_0^1 x \cdot f(x) dx \\ &= \int_0^1 x(\theta+1)x^{\theta} dx \\ &= \frac{\theta+1}{\theta+2} x^{\theta+2}|_0^1 \\ &= \frac{\theta+1}{\theta+2} \end{aligned}$$

Then setting this equal to $\bar{x}$ and solving for θ.

$$\begin{aligned} \frac{\theta+1}{\theta+2} &= \bar{x} \\ \theta + 1 &= \bar{x}(\theta+2) \\ \theta - \bar{x}\theta &= \bar{x}2 - 1 \\ \theta &= \frac{2\bar{x}-1}{1-\bar{x}} \end{aligned}$$

Hence $\widehat{\theta}_0^{mom} = \frac{2\bar{x}-1}{1-\bar{x}}$.

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

Using Def 24.2, first finding the liklihood function,

$$\begin{aligned} L(\theta) &= f(x_1) \times f(x_2) \times \cdots \times f(x_n) \\ &= (\theta+1)x_1^\theta \times \cdots \times (\theta+1)x_n^\theta \\ &= (\theta+1)^n \prod_{i=1}^n x_i^\theta \end{aligned}$$

Next, finding the log-liklihood function,

$$\begin{aligned} Ln(L(\theta)) &= Ln((\theta+1)^n \prod_{i=1}^n x_i^\theta) \\ &= nLn(\theta+1) + \theta Ln(\prod_{i=1}^n x_i) \\ &= nLn(\theta+1) + \theta \prod_{i=1}^n Ln(x_i) \end{aligned}$$

Then finding the first derivative of the log-liklihood function,

$$\begin{aligned} \frac{d}{d\theta} Ln(L(\theta)) &= \frac{d}{d\theta}[nLn(\theta+1) + \theta \prod_{i=1}^n Ln(x_i)] \\ &= \frac{n}{\theta+1} + \prod_{i=1}^n Ln(x_i) \end{aligned}$$

Next, setting this equal to zero and solving for θ, a critical point,

$$\begin{aligned} \frac{n}{\theta+1} + \prod_{i=1}^n Ln(x_i) &= 0 \\ -\frac{n}{\prod_{i=1}^n Ln(x_i)} - 1 &= \theta \end{aligned}$$

Then finding the second derivative to show our critical point is a maximizer,

$$\begin{aligned} \frac{d}{d\theta}[\frac{n}{\theta+1} + \prod_{i=1}^n Ln(x_i)] &= -\frac{n}{(\theta+1)^2} + 0 \\ &= \frac{-n}{(\theta+1)^2} \end{aligned}$$

Since $\frac{-n}{(\theta+1)^2}$ is less than 0 for any value of θ, than our $\widehat{\theta}_0^{mle} = -\frac{n}{\prod_{i=1}^n Ln(x_i)} - 1$ is a global maximum of the likelihood function, for $\theta > -1$.

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)
x_bar <- mean(x)

mom <- (2 * x_bar - 1) / (1 - x_bar)
```

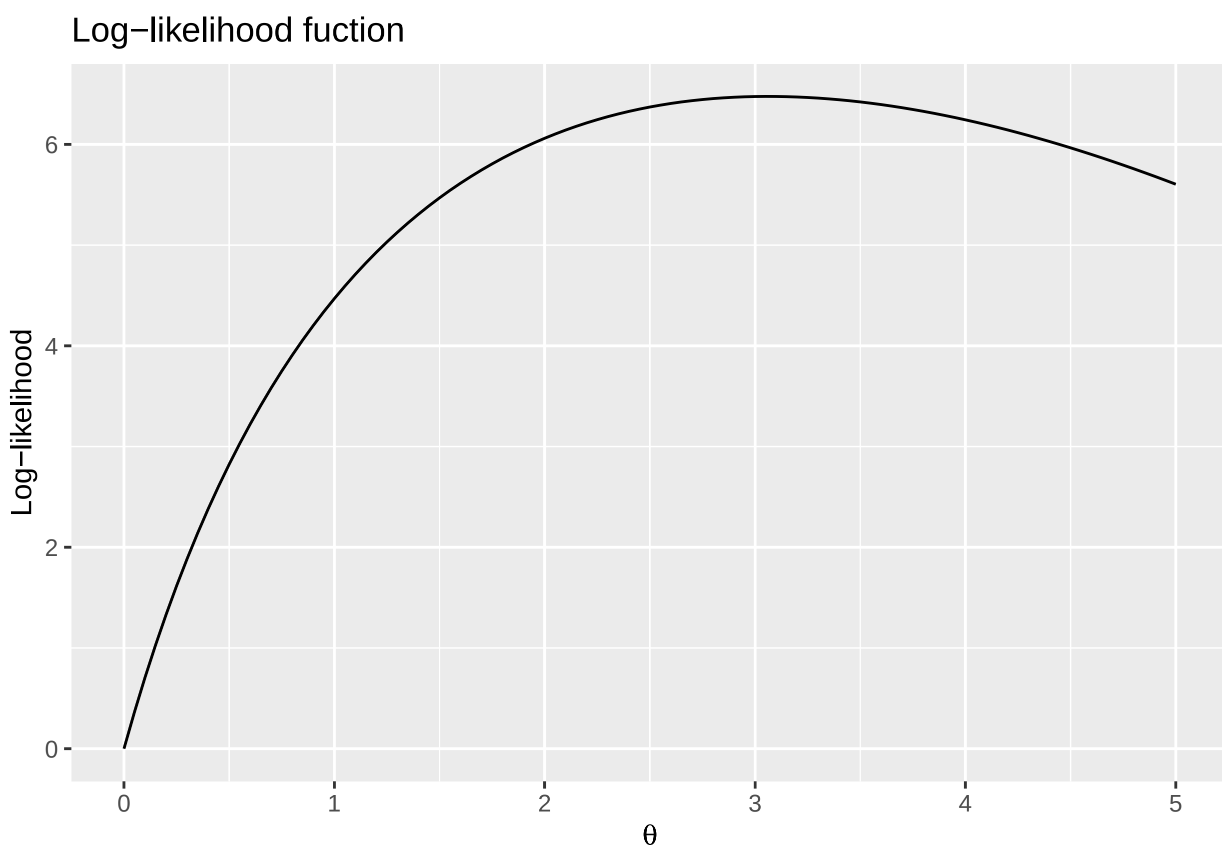
Using the $\widehat{\theta}_0^{mom}$ we found in part a, the $\widehat{\theta}_0^{mom}$ for this sample is about 2.98.

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

```
df <- data.frame(x = seq(0, 5, 1))
n <- length(x)

log_likelihood <- function(theta){
  n*log(theta + 1) + theta * log(prod(x))
}

ggplot(data=df, aes(x=x)) +
  geom_function(fun = log_likelihood) +
  labs(x = expression(theta),
       y = "Log-likelihood",
       title = "Log-likelihood fuction")
```




```
mle <- ( -n / log(prod(x))) -1
mle
```

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

Here we can see a plot of the log-likelihood function for the data from part c. From the graph we can see the max appears to be somewhere just above 3. From our calculation, the exact θ that will max the log-likelihood function is about 3.06.

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.)

Assume that $X_1, X_2, \ldots, X_6$ are the observed lifetimes of the six light bulbs that die within 48 hours and Y_1, Y_2, Y_3, Y_4 the lifetime of four bulbs that are still alive after 48 hours. Each X_i is an exponential random variable with PDF $f(x, \lambda_0) = \lambda_0 e^{-\lambda x_i}, x_i = 1, 2, \ldots, 6$. Each Y_i is a Bernoulli random variable where the output is 1 if the bulbs lifetime is larger than 48 hours and 0 otherwise, with PMF $f(y, \lambda_0) = (e^{-48\lambda_0})^y \cdot (1 - e^{-48\lambda_0})^{1-y}$. Thus, finding the likelihood function,

$$\begin{aligned}
L(\lambda) &= f(x_1) \times \cdots \times f(x_6) \times f(y_1) \times \cdots \times f(y_4) \\
&= \prod_{i=1}^{6} \lambda_0 e^{-\lambda x_i} \cdot \prod_{i=1}^{4} (e^{-48\lambda_0})^{y_i} \cdot (1 - e^{-48\lambda})^{1-y_1} \\
&= \lambda^6 e^{-\lambda \sum_{i=1}^{6} x_i} \cdot \prod_{i=1}^{4} (e^{-48\lambda}) \\
&= \lambda^6 e^{-\lambda \sum_{i=1}^{6} x_i} \cdot (e^{-4(48)\lambda}) \\
&= \lambda^6 e^{-\lambda \cdot (6\bar{x} + 192)} \\
&\text{for } \lambda > 0
\end{aligned}$$

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

$$\begin{aligned}
\frac{d}{d\lambda} LnL(\lambda) &= \frac{d}{d\lambda} Ln(\lambda^6 e^{-\lambda \cdot (6\bar{x}+192)}) \\
&= \frac{d}{d\lambda} (6Ln(\lambda) - \lambda \cdot (6\bar{x} + 192) \\
&= \frac{6}{\lambda} - (6\bar{x} + 192)
\end{aligned}$$

Then setting this equal to 0 and solving for λ,

$$\frac{6}{\lambda} - (6\bar{x} + 192) = 0$$

$$\lambda = \frac{6}{6\bar{x} + 192}$$

Then finding the second derivative to show our critical point is a maximizer,

$$\begin{aligned}
\frac{d^2}{d\lambda_0^2} \log L(\lambda_0) &= \frac{d}{d\lambda_0} \frac{6}{\lambda} - (6\bar{x} + 192) \\
&= -\frac{6}{\lambda_0^2}
\end{aligned}$$

Since $-\frac{6}{\lambda_0^2}$ is less than 0 for any value of λ, than our $\widehat{\lambda}_0^{mle} = \frac{6}{6\bar{x}+192}$ is a global maximum of the likelihood function.

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 $\ell(\mu)$.

Assume that $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Norm(a_i\mu_0, 1)$ where a_i are known constants. Then the PDF for each X_i is $f(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{(-1/2)(x_i - a_i\mu)^2}$ for $-\infty < x_i < \infty$. Then finding the likelihood function,

$$\begin{aligned} L(\mu) &= f(x_1) \times \cdots \times f(x_n) \\ &= \frac{1}{(2\pi)^{\frac{1}{2}}} e^{(-1/2)(x_1 - a_1\mu)^2} \times \cdots \times \frac{1}{(2\pi)^{\frac{1}{2}}} e^{(-1/2)(x_n - a_n\mu)^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{(-1/2)\sum_{i=1}^{n}(x_i - a_i\mu)^2} \end{aligned}$$

And finding the log-likelihood function,

$$\begin{aligned} \ell(\mu) = Ln(L(\mu)) &= Ln(\frac{1}{(2\pi)^{\frac{n}{2}}} e^{(-1/2)\sum_{i=1}^{n}(x_i - a_i\mu)^2}) \\ &= -\frac{n}{2}Ln(2\pi) - (1/2)\sum_{i=1}^{n}(x_i - a_i\mu)^2 \\ &\text{for } -\infty < \mu < \infty \end{aligned}$$

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

To find $\widehat{\mu}_0^{mle}$ we can follow the same steps as in problem 1. First finding the critical points,

$$\begin{aligned} \frac{d}{d\mu}\ell(\mu) &= -\frac{n}{2}Ln(2\pi) - (1/2)\sum_{i=1}^{n}(x_i - a_i\mu)^2 \\ &= \frac{1}{2}\sum_{i=1}^{n} 2a_i^2(\mu - \frac{x_i}{a_i}) \\ &= -\sum_{i=1}^{n} a_i^2(\mu - \frac{x_i}{a_i}) \\ &= -\sum_{i=1}^{n} a_i^2\mu + \sum_{i=1}^{n} a_i x_i \\ &= \sum_{i=1}^{n} a_i x_i - \mu\sum_{i=1}^{n} a_i^2 \end{aligned}$$

Then setting equal to 0 and solving for μ,

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

Then finding the second derivative,

$$\frac{d^2}{d\mu^2}\ell(\mu) = -\sum_{i=1}^{n} a_i^2$$

Since $-\sum_{i=1}^{n} a_i^2$ is less than zero for any μ, then our critical point is a global maximum. Thus $\widehat{\mu}_0^{mle} = \frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i^2}$.

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

To check if $\widehat{\mu}_0^{mle}$ an unbiased estimator of μ_0 we can check if its expected value is equal to μ.

$$
\begin{aligned}
E[\widehat{\mu}_0^{mle}] &= E[\frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n a_i^2}] \\
&\text{by linearity of expectation} \\
&= \frac{\sum_{i=1}^n a_i E(X_i)}{\sum_{i=1}^n a_i^2} \\
&\text{since } E[X_i] = a_i \mu \\
&= \frac{\sum_{i=1}^n a_i a_i \mu}{\sum_{i=1}^n a_i^2} \\
&= \frac{\sum_{i=1}^n a_i^2 \mu}{\sum_{i=1}^n a_i^2} \\
&= \mu
\end{aligned}
$$

Hence $\widehat{\mu}_0^{mle}$ is an unbiased estimator of μ_0.

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

To find the standard error of $\widehat{\mu}_0^{mle}$ need to computer the variance and take the square root.

$$
\begin{aligned}
Var[\widehat{\mu}_0^{mle}] &= Var(\frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n a_i^2}) \\
&\text{by non linearity of varience} \\
&= \frac{1}{(\sum_{i=1}^n a_i^2)^2} Var(\sum_{i=1}^n a_i X_i) \\
&\text{because each } x_i \text{ is independent} \\
&= \frac{1}{(\sum_{i=1}^n a_i^2)^2} \sum_{i=1}^n a_i^2 Var(X_i) \\
&\text{since } Var(X_i) = 1 \text{ for each } X_i \\
&= \frac{1}{(\sum_{i=1}^n a_i^2)^2} \sum_{i=1}^n a_i^2 \\
&= \frac{1}{\sum_{i=1}^n a_i^2}
\end{aligned}
$$

Hence the standard error of $\widehat{\mu}_0^{mle}$ is $\sqrt(\frac{1}{\sum_{i=1}^n a_i^2})$.

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.

Intro Assume that we have two random samples from an exponential distribution. The first sample is of size n_1 with mean θ_0 and a second random sample of size n_2 with mean $k\theta_0$ where k is a known number, and θ is unknown. Then we know that the first sample is drawn from the PDF $f_1(x) = \frac{1}{\theta_0}e^{\frac{-x}{\theta_0}}$ where x is greater than 0 and the second sample is drawn from the PDF $f_2(x) = \frac{1}{k\theta_0}e^{\frac{-x}{k\theta_0}}$ where x is greater than 0. Then following similar steps are earlier problems, we can find our MLE's.

Derivation of $\widehat{\theta}_0^{mle1}$:

$$\begin{aligned}
L(\theta|x_1) &= f_1(x_1) \times \cdots \times f_1(x_n) \\
&= \frac{1}{\theta}e^{\frac{-x_1}{\theta}} \times \cdots \times \frac{1}{\theta}e^{\frac{-x_n}{\theta}} \\
&= \prod_{i=1}^{n_1} \frac{1}{\theta}e^{\frac{-x_i}{\theta}} \\
&= \frac{1}{\theta^{n_1}}e^{\frac{-\sum_{i=1}^{n_1} x_i}{\theta}} \\
&\text{for } \theta > 0
\end{aligned}$$

$$\begin{aligned}
\ell(\theta) = Ln(L(\theta)) &= Ln(\frac{1}{\theta^{n_1}}e^{\frac{-\sum_{i=1}^{n_1} x_i}{\theta}}) \\
&= -n_1 Ln(\theta) - \frac{\sum_{i=1}^{n_1} x_i}{\theta}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{d\theta}\ell(\theta) &= -\frac{n_1}{\theta} + \frac{\sum_{i=1}^{n_1} x_i}{\theta^2} \\
&= \frac{-n_1\theta + \sum_{i=1}^{n_1} x_i}{\theta^2}
\end{aligned}$$

$$\begin{aligned}
\frac{d}{d\theta}\ell(\theta) &= 0 \\
\frac{-n_1\theta + \sum_{i=1}^{n_1} x_i}{\theta^2} &= 0 \\
-n_1\theta + \sum_{i=1}^{n_1} x_i &= 0 \\
\theta &= \frac{\sum_{i=1}^{n_1} x_i}{n_1} \\
\theta &= \frac{n_1\bar{x}}{n_1} \\
\theta &= \bar{x}
\end{aligned}$$

$$\begin{aligned}
\frac{d^2}{d\theta^2}\ell(\theta) &= \frac{d}{d\theta}(-\frac{n_1}{\theta} + \frac{\sum_{i=1}^{n_1} x_i}{\theta^2}) \\
&= \frac{n_1}{\theta^2} - \frac{2\sum_{i=1}^{n_1} x_i}{\theta^3} \\
&= \frac{n_1\theta - 2n_1\bar{x}}{\theta^3}
\end{aligned}$$

$$\frac{n_1\theta - 2n_1\bar{x}}{\theta^3} \Rightarrow \theta = \bar{x} \Rightarrow \frac{n_1\bar{x} - 2n_1\bar{x}}{\bar{x}^3}$$
$$\Rightarrow \frac{-n_1}{\bar{x}^2}$$

Note that we found only one critical point and when we plug it in to the second derivative we get $\frac{-n_1}{\bar{x}^2}$ which is always negative. Hence $\theta = \bar{x}$ is a global maximum for the log-likelihood function and $\widehat{\theta}_0^{mle1} = \bar{x}$.

Derivation of $\widehat{\theta}_0^{mle2}$:

$$\begin{aligned} L(\theta) &= f_2(x_1) \times \cdots \times f_2(x_{n_2}) \\ &= \frac{1}{k\theta}e^{\frac{-x_1}{k\theta}} \times \cdots \times \frac{1}{k\theta}e^{\frac{-x_{n_2}}{k\theta}} \\ &= \prod_{i=1}^{n_2} \frac{1}{k\theta}e^{\frac{-x_i}{k\theta}} \\ &= \frac{1}{(k\theta)}^{n_2} e^{\frac{-\sum_{i=1}^{n_2} x_i}{k\theta}} \\ &\text{for } \theta > 0 \end{aligned}$$

$$\begin{aligned} \ell(\theta) = Ln(L(\theta)) &= Ln(\frac{1}{(k\theta)}^{n_2} e^{\frac{-\sum_{i=1}^{n_2} x_i}{k\theta}}) \\ &= n_2 Ln(\frac{1}{(k\theta)})) - \frac{\sum_{i=1}^{n_2} x_i}{k\theta} \end{aligned}$$

$$\frac{d}{d\theta}\ell(\theta) = -\frac{n_2}{\theta} + \frac{\sum_{i=1}^{n_2} x_i}{k\theta^2}$$

$$\begin{aligned} \frac{d}{d\theta}\ell(\theta) &= 0 \\ -\frac{n_2}{\theta} + \frac{\sum_{i=1}^{n_2} x_i}{k\theta^2} &= 0 \\ \frac{n_2\bar{x} - kn_2\theta}{k\theta^2} &= 0 \\ \bar{x} - k\theta &= 0 \\ \frac{\bar{x}}{k} &= \theta \end{aligned}$$

$$\begin{aligned} \frac{d^2}{d\theta^2}\ell(\theta) &= \frac{d}{d\theta}(-\frac{n_2}{\theta} + \frac{\sum_{i=1}^{n_2} x_i}{k\theta^2}) \\ &= \frac{n_2}{\theta^2} - \frac{2\sum_{i=1}^{n_2} x_i}{k\theta^3} \\ &= \frac{n_2\theta - 2n_2\bar{x}}{k\theta^3} \end{aligned}$$

$$\frac{n_2\theta - 2n_2\bar{x}}{k\theta^3} \Rightarrow \theta = \frac{\bar{x}}{k} \Rightarrow \frac{n_2\frac{\bar{x}}{k} - 2n_2\frac{\bar{x}}{k}}{k\frac{\bar{x}}{k}^3}$$
$$\Rightarrow \frac{-n_2}{k\frac{\bar{x}}{k}^2}$$

Note that we found only one critical point and when we plug it in to the second derivative we get $\frac{-n_2}{k\frac{\bar{x}}{k}^2}$ which is always negative. Hence $\theta = \frac{\bar{x}}{k}$ is a global maximum for the log-likelihood function and $\widehat{\theta}_0^{mle2} = \frac{\bar{x}}{k}$.

Linear combination:

Now we want to find a linear combination $a\widehat{\theta}_0^{mle1} + (1-a)\widehat{\theta}_0^{mle2}$, with the smallest variance.

$$\begin{aligned}
Var[a\widehat{\theta}_0^{mle1} + (1-a)\widehat{\theta}_0^{mle2}] &\text{ by non linearity of varience, and plugging in our found } \widehat{\theta}_0^{mle1} \text{ and } \widehat{\theta}_0^{mle2} \\
&= a^2 Var[\bar{x_1}] + (1-a)^2 Var[\frac{\bar{x_2}}{k}] \\
&\text{where } \bar{x_1} \text{ is from the first experiment and } x_2 \text{ is from the second experiment} \\
&\text{then by non linearity of expectation again,} \\
&= a^2 Var[\bar{x_1}] + \frac{(1-a)^2}{k^2} Var[\bar{x_2}] \\
&\text{since each } x \sim Exp(\theta_0) \\
&= \frac{a^2\theta_0^2}{n_1} + \frac{(1-a)^2(k\theta_0)^2}{n_2k^2} \\
&= \frac{a^2\theta_0^2}{n_1} + \frac{(1-a)^2\theta_0^2}{n_2} \\
&\text{then taking the derivative and setting it equal to 0} \\
\frac{d}{da}\frac{a^2\theta_0^2}{n_1} + \frac{(1-a)^2\theta_0^2}{n_2} &= \frac{2a\theta_0^2}{n_1} - \frac{2(1-a)\theta_0^2}{n_2} \\
\frac{2a\theta_0^2}{n_1} - \frac{2(1-a)\theta_0^2}{n_2} &= 0 \\
2a\theta_0^2 n_2 - 2(1-a)\theta_0^2 n_1 &= 0 \\
2a\theta_0^2 n_2 &= 2(1-a)\theta_0^2 n_1 \\
an_2 &= (1-a)n_1 \\
an_2 + an_1 &= n_1 \\
a &= \frac{n_1}{n_1+n_2} \\
&\text{then completeing the second derivative test to show that this is a minimum} \\
\frac{d}{da}\frac{2a\theta_0^2}{n_1} - \frac{2(1-a)\theta_0^2}{n_2} &= \frac{2\theta_0^2}{n_1} + \frac{2\theta_0^2}{n_2}
\end{aligned}$$

As we can see, the second derivative $\frac{2\theta_0^2}{n_1} + \frac{2\theta_0^2}{n_2}$ is greater than 0 for any value of θ, thus our critical point $\frac{n_1}{n_1+n_2}$ will minimize the variance. Hence the linear combination will be,

$$a\widehat{\theta}_0^{mle1} + (1-a)\widehat{\theta}_0^{mle2} = \frac{n_1}{n_1+n_2}\bar{x_1} + (1 - \frac{n_1}{n_1+n_2})\frac{\bar{x_2}}{k}$$

Combining sample MLE:

Now if we were to calculate an MLE from a combined sample.

$$
\begin{aligned}
L(\theta|x_1) &= f_1(x_1) \times \cdots \times f_1(x_n) \times f_2(x_1) \times \cdots \times f_2(x_{n_2}) \\
&= \frac{1}{\theta} e^{\frac{-x_1}{\theta}} \times \cdots \times \frac{1}{\theta} e^{\frac{-x_n}{\theta}} \times \frac{1}{k\theta} e^{\frac{-x_1}{k\theta}} \times \cdots \times \frac{1}{k\theta} e^{\frac{-x_{n_2}}{k\theta}} \\
&= \prod_{i=1}^{n_1} \frac{1}{\theta} e^{\frac{-x_i}{\theta}} \times \prod_{i=1}^{n_2} \frac{1}{k\theta} e^{\frac{-x_i}{k\theta}} \\
&= \frac{1}{\theta^{n_1}} e^{\frac{-\sum_{i=1}^{n_1} x_i}{\theta}} \times \frac{1}{(k\theta)}^{n_2} e^{\frac{-\sum_{i=1}^{n_2} x_i}{k\theta}} \\
&= \frac{1}{\theta^{n_1+n_2} k^{n_2}} e^{-\frac{1}{\theta} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)}
\end{aligned}
$$

$$
\begin{aligned}
\ell(\theta) = Ln(L(\theta)) &= Ln(\frac{1}{\theta^{n_1+n_2} k^{n_2}} e^{-\frac{1}{\theta} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)}) \\
&= -Ln(\theta^{n_1+n_2} k^{n_2}) - \frac{1}{\theta} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k) \\
&= -n_2 Ln(k) - (n_1 + n_2) Ln(\theta) - \frac{1}{\theta} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{d\theta} \ell(\theta) &= \frac{d}{d\theta} (-n_2 Ln(k) - (n_1 + n_2) Ln(\theta) - \frac{1}{\theta} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)) \\
&= \frac{-(n_1 + n_2)}{\theta} + \frac{1}{\theta^2} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k))
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{d\theta} \ell(\theta) &= 0 \\
\frac{-(n_1 + n_2)}{\theta} + \frac{1}{\theta^2} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)) &= 0 \\
-(n_1 + n_2)\theta + n_1 \bar{x_1} + n_2 \bar{x_2}/k &= 0 \\
\theta &= \frac{n_1 \bar{x_1} + n_2 \bar{x_2}/k}{n_1 + n_2}
\end{aligned}
$$

$$
\begin{aligned}
\frac{d^2}{d\theta^2} \ell(\theta) &= \frac{d}{d\theta} \frac{-(n_1 + n_2)}{\theta} + \frac{1}{\theta^2} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)) \\
&= \frac{n_1 + n_2}{\theta^2} - \frac{2}{\theta^3} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k))
\end{aligned}
$$

$$
\begin{aligned}
\frac{n_1 + n_2}{\theta^2} - \frac{2}{\theta^3} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)) \Rightarrow \theta = \frac{n_1 \bar{x_1} + n_2 \bar{x_2}/k}{n_1 + n_2} &\Rightarrow \frac{n_1 + n_2}{(\frac{n_1 \bar{x_1} + n_2 \bar{x_2}/k}{n_1 + n_2})^2} - \frac{2}{(\frac{n_1 \bar{x_1} + n_2 \bar{x_2}/k}{n_1 + n_2})^3} \cdot (n_1 \bar{x_1} + n_2 \bar{x_2}/k)) \\
&= -\frac{(n_1 + n_2)^2}{(n_1 \bar{x_1} + n_2 \bar{x_2}/k)^2}
\end{aligned}
$$

Note that we found only one critical point and when we plug it in to the second derivative we get $-\frac{(n_1+n_2)^2}{(n_1\bar{x_1}+n_2\bar{x_2}/k)^2}$ which is always negative. Hence $\theta = \frac{n_1\bar{x_1}+n_2\bar{x_2}/k}{n_1+n_2}$ is a global maximum for the log-likelihood function and $\widehat{\theta}_0^{mle} = \frac{n_1\bar{x_1}+n_2\bar{x_2}/k}{n_1+n_2}$.

Conclusion After calculating the Linear combination of MLE's and a MLE from a combined sample, we can see that these give the same estimate of $\widehat{\theta}_0^{mle} = \frac{n_1\bar{x_1}+n_2\bar{x_2}/k}{n_1+n_2}$.