

Homework 4

Spring 2023

Jaiden Atterbury

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, \dots, 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 $\hat{\theta}_0^{mom}$, the method of moments estimator of θ_0 .

In order to find the method of moments estimator of θ_0 , we must find $E[X_i]$ and set it equal to \bar{X} . Where \bar{X} represents the random variable version of the sample average. We will first calculate $E[X_i]$ as shown below.

$$\begin{aligned} E[X_i] &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^1 (\theta_0 + 1)x^{\theta_0+1}dx \\ &= (\theta_0 + 1) \int_0^1 x^{\theta_0+1}dx \\ &= \frac{\theta_0 + 1}{\theta_0 + 2} [x^{\theta_0+2}]_0^1 \\ &= \frac{\theta_0 + 1}{\theta_0 + 2} \end{aligned}$$

As calculated above, $E[X_i] = \frac{\theta_0+1}{\theta_0+2}$. Setting this equal to \bar{X} we obtain $\frac{\theta_0+1}{\theta_0+2} = \bar{X}$. We will now isolate θ_0 as shown below.

$$\begin{aligned}\frac{\theta_0 + 1}{\theta_0 + 2} &= \bar{X} \\ \theta_0 + 1 &= (\theta_0 + 2)\bar{X} \\ \theta_0 + 1 &= \theta_0\bar{X} + 2\bar{X} \\ (1 - \bar{X})\theta_0 &= 2\bar{X} - 1 \\ \theta_0 &= \frac{2\bar{X} - 1}{1 - \bar{X}}\end{aligned}$$

As calculated above, $\hat{\theta}_0^{mom} = \frac{2\bar{X}-1}{1-\bar{X}}$.

b. Derive $\hat{\theta}_0^{mle}$, the maximum likelihood estimator of θ_0 . Be sure to show

- likelihood function
- log-likelihood function
- first derivative condition
- second derivative test

To find the maximum likelihood estimator of θ_0 . We must find the likelihood function, take the natural log of the likelihood function, find the critical point of the log-likelihood function, and lastly run the second derivative test to show the critical point is a maximum. First off, we will compute the likelihood function as shown below.

$$\begin{aligned}L(\theta) &= f(x_1) \times f(x_2) \times \cdots \times f(x_n) \\ &= (\theta + 1)x_1^\theta \times (\theta + 1)x_2^\theta \times \cdots \times (\theta + 1)x_n^\theta \\ &= (\theta + 1)^n(x_1 \times x_2 \times \cdots \times x_n)^\theta, \theta > -1\end{aligned}$$

Thus, as computed above, the likelihood function of θ is $L(\theta) = (\theta + 1)^n(x_1 \times x_2 \times \cdots \times x_n)^\theta, \theta > -1$. However, notice that taking the derivative of this function will not be easy, thus we will take the natural log of the likelihood function to turn multiplication into addition. This process is shown below.

$$\begin{aligned}\ell(\theta) &= \ln(L(\theta)) \\ &= \ln((\theta + 1)^n(x_1 \times x_2 \times \cdots \times x_n)^\theta) \\ &= \ln((\theta + 1)^n) + \ln((x_1 \times x_2 \times \cdots \times x_n)^\theta) \\ &= n \ln(\theta + 1) + \theta \ln(x_1 \times x_2 \times \cdots \times x_n), \theta > -1\end{aligned}$$

As computed above, the log-likelihood function is $\ell(\theta) = n \ln(\theta + 1) + \theta \ln(x_1 \times x_2 \times \cdots \times x_n), \theta > -1$. Now we must find the critical points of this function in order to find the candidates for the maximum likelihood estimator of θ_0 . This derivative calculation is shown below.

$$\begin{aligned}\frac{d}{d\theta}\ell(\theta) &= \frac{d}{d\theta}(n \ln(\theta + 1) + \theta \ln(x_1 \times x_2 \times \cdots \times x_n)) \\ &= \frac{n}{\theta + 1} + \ln(x_1 \times x_2 \times \cdots \times x_n), \theta > -1\end{aligned}$$

In order to find the critical points, we must find the values of θ such that this derivative is equal to zero.

$$\begin{aligned}
0 &= \frac{n}{\theta + 1} + \ln(x_1 \times x_2 \times \cdots \times x_n) \\
\frac{-n}{\theta + 1} &= \ln(x_1 \times x_2 \times \cdots \times x_n) \\
-n &= (\theta + 1) \ln(x_1 \times x_2 \times \cdots \times x_n) \\
-n &= \theta \ln(x_1 \times x_2 \times \cdots \times x_n) + \ln(x_1 \times x_2 \times \cdots \times x_n) \\
-n - \ln(x_1 \times x_2 \times \cdots \times x_n) &= \theta \ln(x_1 \times x_2 \times \cdots \times x_n) \\
\theta &= \frac{-n - \ln(x_1 \times x_2 \times \cdots \times x_n)}{\ln(x_1 \times x_2 \times \cdots \times x_n)}
\end{aligned}$$

As computed above, the critical point $\theta = \frac{-n - \ln(x_1 \times x_2 \times \cdots \times x_n)}{\ln(x_1 \times x_2 \times \cdots \times x_n)}$ is a candidate for the maximum likelihood estimator for θ_0 we will now compute the second derivative test below to show that this critical point is in fact a maximum.

$$\begin{aligned}
\frac{d^2}{d\theta^2} \ell(\theta) &= \frac{d^2}{d\theta^2} \left(\frac{n}{\theta + 1} + \ln(x_1 \times x_2 \times \cdots \times x_n) \right) \\
&= \frac{-n}{(\theta + 1)^2}, \theta > -1
\end{aligned}$$

Since $n > 0$ and $\theta > -1$ implies that $(\theta + 1)^2 > 0$, it follows that $\frac{-n}{(\theta + 1)^2} < 0$, $\forall \theta > -1$. Hence $f'' \left(\frac{-n - \ln(x_1 \times x_2 \times \cdots \times x_n)}{\ln(x_1 \times x_2 \times \cdots \times x_n)} \right) < 0$ and the critical point is a global maximum. Thus it follows that $\hat{\theta}_0^{mle} = \frac{-n - \ln(x_1 \times x_2 \times \cdots \times x_n)}{\ln(x_1 \times x_2 \times \cdots \times x_n)}$.

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

Calculating $\hat{\theta}_0^{mom}$:

```

# Sample of 10 data points from f(x):
x <- c(0.90, 0.78, 0.93, 0.64, 0.45, 0.85, 0.75, 0.93, 0.98, 0.78)

# Sample mean:
xbar <- mean(x)

# Method of moments estimator:
mom_est <- (2 * xbar - 1) / (1 - xbar)

```

As calculated above, the method of moments estimate of θ_0 is $\hat{\theta}_0^{mom} = 2.9751244$.

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

Plot of the log-likelihood function:

```

# Sample size:
n <- length(x)

# Product of the 10 sample values:
prod_x <- prod(x)

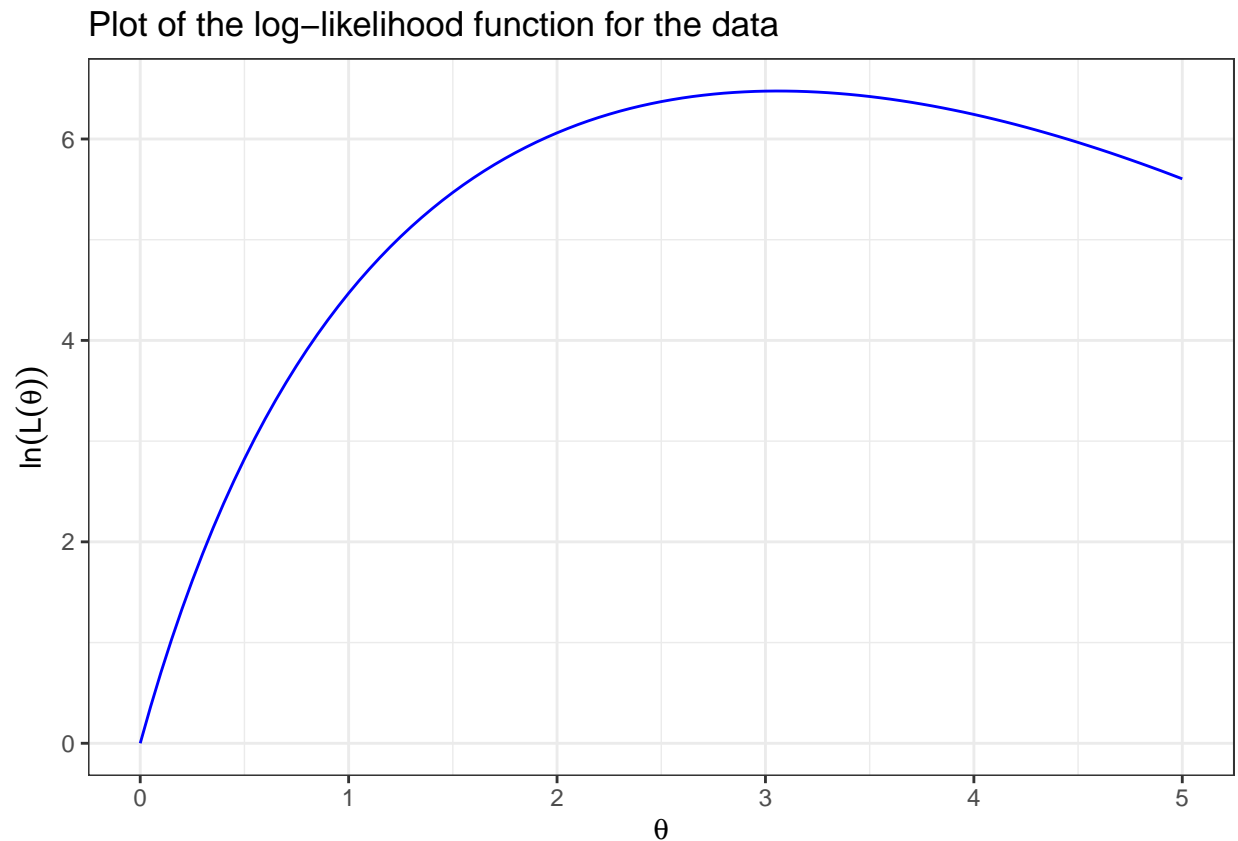
```

```

# Log-likelihood function:
loglikelihood <- function(theta) {
  n * log(theta + 1) + theta * log(prod_x)
}

# Plot the log-likelihood function:
ggplot(data = data.frame(values = c(0, 5)),
       mapping = aes(x = values)) +
  geom_function(fun = loglikelihood, color = "blue") +
  labs(x = expression(theta),
       y = expression(ln(L(theta))),
       title = "Plot of the log-likelihood function for the data") +
  theme_bw()

```



As can be seen from the plot of $\ell(\theta)$, the maximum of the log-likelihood function appears to be somewhere around 3. We will confirm this claim in the next part.

Calculating $\hat{\theta}_0^{mle}$:

```

# Maximum likelihood estimator:
mle_est <- (-n - log(prod_x)) / log(prod_x)

```

As calculated above, the sample maximum likelihood estimate of θ_0 is $\hat{\theta}_0^{mle} = 3.0607107$.

- (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, \dots, 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, \dots, 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.)

Suppose a set of cheap light bulbs have a lifetime, in hours, which are exponentially distributed with an unknown rate of λ_0 . Suppose also that we choose a random sample of ten light bulbs and turn them all on simultaneously and leave them on for 48 hours. During this 48 hour period, six bulbs went out at times x_1, x_2, \dots, x_6 . Furthermore, at the end of the experiment, four light bulbs were still working.

In order to derive the likelihood function $L(\lambda)$, we can model this event as observing values for X_1, X_2, \dots, X_6 which are exponential random variables and Y_1, Y_2, Y_3, Y_4 which are Bernoulli random variables. Y_1, Y_2, Y_3, Y_4 are 1 if the lifetime of their corresponding light bulb X is larger than 48, and 0 otherwise. The likelihood function is thus product of the six exponential density functions and the four Bernoulli PMFs.

With that said, we know the density function of our exponential random variables is $f(x) = \lambda_0 e^{-\lambda_0 x}$, $0 < x$, and the mass function of our Bernoulli random variables is $g(y) = \pi_0^y \cdot (1 - \pi_0)^{1-y}$, $y = 0, 1$.

Thus our likelihood function takes the form:

$$L(\lambda) = f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \cdot f(x_5) \cdot f(x_6) \cdot g(y_1) \cdot g(y_2) \cdot g(y_3) \cdot g(y_4)$$

However, since we already know there are 4 light bulbs that didn't burn out after 48 hours, it follows that Y_1, Y_2, Y_3, Y_4 all have observed values of 1. Thus our likelihood function becomes:

$$\begin{aligned} L(\lambda) &= f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \cdot f(x_5) \cdot f(x_6) \cdot g(1) \cdot g(1) \cdot g(1) \cdot g(1) \\ &= \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdot \lambda e^{-\lambda x_3} \cdot \lambda e^{-\lambda x_4} \cdot \lambda e^{-\lambda x_5} \cdot \lambda e^{-\lambda x_6} \cdot \pi \cdot \pi \cdot \pi \cdot \pi \\ &= \lambda^6 e^{-\lambda(x_1+x_2+x_3+x_4+x_5+x_6)} \pi^4, \lambda > 0 \end{aligned}$$

In the above likelihood, π represents the probability that the lifespan of a given light bulb, X , is greater than 48 hours. This is equivalent to the probability statement $P(X > 48)$. It turns out that this can actually be solved in terms of λ using the fact that the non-zero part of the CDF of an exponential random variable is $F(x) = 1 - e^{-\lambda x}$, $x > 0$.

$$\begin{aligned} \pi &= P(X > 48) \\ &= 1 - P(X \leq 48) \\ &= 1 - (1 - e^{-48\lambda}) \\ &= e^{-48\lambda} \end{aligned}$$

We can now plug in this expression for π into our likelihood function and obtain the final form of the likelihood function of λ which is: $L(\lambda) = \lambda^6 e^{-\lambda(x_1+x_2+x_3+x_4+x_5+x_6)} e^{-192\lambda} = \lambda^6 e^{-6\lambda\bar{x}} e^{-192\lambda} = \lambda^6 e^{\lambda(-192-6\bar{x})}$, $\lambda > 0$.

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

To find the maximum likelihood estimator of λ_0 . We must find the likelihood function, take the natural log of the likelihood function, find the critical point of the log-likelihood function, and lastly run the second derivative test to show the critical point is a maximum. As computed above, the likelihood function of λ is $L(\lambda) = \lambda^6 e^{\lambda(-192-6\bar{x})}$, $\lambda > 0$. However, notice that taking the derivative of this function will not be easy,

thus we will take the natural log of the likelihood function to turn multiplication into addition. This process is shown below.

$$\begin{aligned}\ell(\lambda) &= \ln(L(\lambda)) \\ &= \ln(\lambda^6 e^{\lambda(-192-6\bar{x})}) \\ &= \ln(\lambda^6) + \ln(e^{\lambda(-192-6\bar{x})}) \\ &= 6 \ln(\lambda) + \lambda(-192 - 6\bar{x}), \lambda > 0\end{aligned}$$

As computed above, the log-likelihood function is $\ell(\lambda) = 6 \ln(\lambda) + \lambda(-192 - 6\bar{x})$, $\lambda > 0$. Now we must find the critical points of this function in order to find the candidates for the maximum likelihood estimator of λ_0 . This derivative calculation is shown below.

$$\begin{aligned}\frac{d}{d\lambda}\ell(\lambda) &= \frac{d}{d\lambda}(6 \ln(\lambda) + \lambda(-192 - 6\bar{x})) \\ &= \frac{6}{\lambda} - 192 - 6\bar{x}, \lambda > 0\end{aligned}$$

In order to find the critical points, we must find the values of λ such that this derivative is equal to zero.

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

As computed above, the critical point $\lambda = \frac{6}{6\bar{x}+192}$ is a candidate for the maximum likelihood estimator for λ_0 we will now compute the second derivative test below to show that this critical point is in fact a maximum.

$$\begin{aligned}\frac{d^2}{d\lambda^2}\ell(\lambda) &= \frac{d^2}{d\lambda^2}\left(\frac{6}{\lambda} + 192 - 6\bar{x}\right) \\ &= \frac{-6}{\lambda^2}, \lambda > 0\end{aligned}$$

Since $\lambda^2 > 0$, it follows that $\frac{-6}{\lambda^2} < 0$, $\forall \lambda > 0$. Hence $f''\left(\frac{6}{6\bar{x}+192}\right) < 0$ and the critical point is a global maximum. Thus it follows that $\hat{\lambda}_0^{mle} = \frac{6}{6\bar{x}+192}$.

3. (SLR through origin) Suppose $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} Norm(a_i \mu_0, 1)$ 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 $\ell(\mu)$.

Since $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} Norm(a_i \mu_0, 1)$ where the a_i are known constants, it follows that the PDF of X_i

is $f(x_i) = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_i - a_i\mu)^2}$, $-\infty < x_i < \infty$. Hence the likelihood function of $L(\mu)$ is:

$$\begin{aligned} L(\mu) &= f(x_1) \times f(x_2) \times \cdots \times f(x_n) \\ &= \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_1 - a_1\mu)^2} \times \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_2 - a_2\mu)^2} \times \cdots \times \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}(x_n - a_n\mu)^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(x_1 - a_1\mu)^2 - \frac{1}{2}(x_2 - a_2\mu)^2 - \cdots - \frac{1}{2}(x_n - a_n\mu)^2} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}[(x_1 - a_1\mu)^2 + (x_2 - a_2\mu)^2 + \cdots + (x_n - a_n\mu)^2]} \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2}, \quad -\infty < \mu < \infty \end{aligned}$$

Hence, as computed above, the likelihood function of μ is $L(\mu) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2}$, $-\infty < \mu < \infty$. Now we will take the natural log of the likelihood function to find the log-likelihood function $\ell(\mu)$.

$$\begin{aligned} \ell(\mu) &= \ln(L(\mu)) \\ &= \ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2}\right) \\ &= \ln\left(\frac{1}{(2\pi)^{\frac{n}{2}}}\right) + \ln\left(e^{-\frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2}\right) \\ &= \ln(1) - \ln((2\pi)^{\frac{n}{2}}) - \frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2 \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2, \quad -\infty < \mu < \infty \end{aligned}$$

Hence, as computed above, the log-likelihood function of μ is $\ell(\mu) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2$, $-\infty < \mu < \infty$.

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

To derive an expression for $\hat{\mu}_0^{mle}$, the maximum likelihood estimator of μ_0 , we will first take the derivative of the log-likelihood function and find the critical points. Then we will run the second derivative test and show that critical point is a maximum. First we will compute the first derivative below.

$$\begin{aligned} \frac{d}{d\mu} \ell(\mu) &= \frac{d}{d\mu} \left(-\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - a_i\mu)^2 \right) \\ &= -2a_i \cdot \frac{-1}{2} \sum_{i=1}^n x_i - a_i\mu \\ &= \sum_{i=1}^n a_i x_i - \sum_{i=1}^n a_i^2 \mu, \quad -\infty < \mu < \infty \end{aligned}$$

Now that we have obtained the first derivative, we will set the expression to zero and solve for μ .

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

The critical point/candidate for the maximum likelihood estimator for μ is $\hat{\mu} = \frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n a_i^2}$. We will now compute the second derivative test below to show that this critical point is in fact a maximum.

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

Since $\sum_{i=1}^n a_i^2 > 0$, it follows that $-\sum_{i=1}^n a_i^2 < 0$, $\forall \mu$. Hence $f'' \left(\frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n a_i^2} \right) < 0$ and the critical point is a global maximum. Thus it follows that $\hat{\mu}_0^{mle} = \frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n a_i^2}$.

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

To see if $\hat{\mu}_0^{mle}$ is an unbiased estimator of μ_0 , we must take the expected value of $\hat{\mu}_0^{mle}$, and see if it equals μ_0 .

$$\begin{aligned}
E[\hat{\mu}_0^{mle}] &= E \left[\frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n a_i^2} \right] \\
&= \frac{E[\sum_{i=1}^n a_i X_i]}{\sum_{i=1}^n a_i^2} \quad (\text{Linearity of Expectation}) \\
&= \frac{\sum_{i=1}^n E[a_i X_i]}{\sum_{i=1}^n a_i^2} \quad (\text{Linearity of Expectation}) \\
&= \frac{\sum_{i=1}^n a_i E[X_i]}{\sum_{i=1}^n a_i^2} \quad (\text{Linearity of Expectation}) \\
&= \frac{\sum_{i=1}^n a_i a_i \mu}{\sum_{i=1}^n a_i^2} \quad (\text{Since } E[X_i] = a_i \mu) \\
&= \mu
\end{aligned}$$

Since $E[\hat{\mu}_0^{mle}] = \mu$, it follows that $\hat{\mu}_0^{mle}$ is an unbiased estimator of μ_0 .

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

To derive the standard error of $\hat{\mu}_0^{mle}$, we will find $SD[\hat{\mu}_0^{mle}]$. To start off, we will find $Var[\hat{\mu}_0^{mle}]$ then take the positive square root of this to come up with our final answer.

$$\begin{aligned}
Var[\hat{\mu}_0^{mle}] &= Var\left[\frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n a_i^2}\right] \\
&= \frac{Var[\sum_{i=1}^n a_i X_i]}{(\sum_{i=1}^n a_i^2)^2} \quad (\text{Non-linearity of Variance}) \\
&= \frac{\sum_{i=1}^n Var[a_i X_i]}{(\sum_{i=1}^n a_i^2)^2} \quad (\text{Independence of the } X_i \text{'s}) \\
&= \frac{\sum_{i=1}^n a_i^2 Var[X_i]}{(\sum_{i=1}^n a_i^2)^2} \quad (\text{Non-linearity of Variance}) \\
&= \frac{1}{\sum_{i=1}^n a_i^2} \quad (\text{Since } SD[X_i] = 1 \implies Var[X_i] = 1)
\end{aligned}$$

As computed above, $Var[\hat{\mu}_0^{mle}] = \frac{1}{\sum_{i=1}^n a_i^2}$ hence we can see that the standard error of $\hat{\mu}_0^{mle}$ is $SD[\hat{\mu}_0^{mle}] = \sqrt{\frac{1}{\sum_{i=1}^n a_i^2}}$.

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 $\hat{\theta}_0^{mle1}$ and $\hat{\theta}_0^{mle2}$ 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 $a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2}$, 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 $\hat{\theta}_0^{mle1}$ and $\hat{\theta}_0^{mle2}$
- clearly show your steps (for example, for finding a)
- clearly highlight your final estimators in each case by stating them.

As explained above, a scientist has obtained two random sample 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. From this it follows that sample 1 was drawn from the PDF $f(x) = \frac{1}{\theta_0} e^{-\frac{1}{\theta_0}x}$, $x > 0$ and sample 2 was drawn from the PDF $g(y) = \frac{1}{k\theta_0} e^{-\frac{1}{k\theta_0}y}$, $y > 0$. We will use this information to derive $\hat{\theta}_0^{mle1}$ and $\hat{\theta}_0^{mle2}$ from each of the samples, as well as the MLE from the combined sample, and the linear combination $a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2}$, with the smallest variance.

Deriving $\hat{\theta}_0^{mle1}$:

To find the maximum likelihood estimator of θ_0 for the first sample, we must find the likelihood function, take the natural log of the likelihood function, find the critical point of the log-likelihood function, and lastly run the second derivative test to show the critical point is a maximum. First off, we will compute the likelihood function as shown below.

$$\begin{aligned}
L(\theta) &= f(x_1) \times f(x_2) \times \cdots \times f(x_{n_1}) \\
&= \frac{1}{\theta} e^{-\frac{1}{\theta}x_1} \times \frac{1}{\theta} e^{-\frac{1}{\theta}x_2} \times \cdots \times \frac{1}{\theta} e^{-\frac{1}{\theta}x_{n_1}} \\
&= \left(\frac{1}{\theta}\right)^{n_1} e^{-\frac{1}{\theta}(x_1+x_2+\cdots+x_{n_1})} \\
&= \left(\frac{1}{\theta}\right)^{n_1} e^{-\frac{1}{\theta}n_1\bar{x}}, \theta > 0
\end{aligned}$$

Thus, as computed above, the likelihood function of θ is $L(\theta) = \left(\frac{1}{\theta}\right)^{n_1} e^{\frac{-1}{\theta} n_1 \bar{x}}$, $\theta > 0$. However, notice that taking the derivative of this function will not be easy, thus we will take the natural log of the likelihood function to turn multiplication into addition. This process is shown below.

$$\begin{aligned}\ell(\theta) &= \ln(L(\theta)) \\ &= \ln\left(\left(\frac{1}{\theta}\right)^{n_1} e^{\frac{-1}{\theta} n_1 \bar{x}}\right) \\ &= \ln\left(\left(\frac{1}{\theta}\right)^{n_1}\right) + \ln(e^{\frac{-1}{\theta} n_1 \bar{x}}) \\ &= n_1 \ln\left(\frac{1}{\theta}\right) + \frac{-n_1 \bar{x}}{\theta}, \theta > 0\end{aligned}$$

As computed above, the log-likelihood function is $\ell(\theta) = n_1 \ln\left(\frac{1}{\theta}\right) + \frac{-n_1 \bar{x}}{\theta}$, $\theta > 0$. Now we must find the critical points of this function in order to find the candidates for the maximum likelihood estimator of θ_0 . This derivative calculation is shown below.

$$\begin{aligned}\frac{d}{d\theta}\ell(\theta) &= \frac{d}{d\theta}\left(n_1 \ln\left(\frac{1}{\theta}\right) + \frac{-n_1 \bar{x}}{\theta}\right) \\ &= \frac{\frac{-n_1}{\theta^2}}{\frac{1}{\theta}} + \frac{n_1 \bar{x}}{\theta^2} \\ &= \frac{-n_1}{\theta} + \frac{n_1 \bar{x}}{\theta^2}, \theta > 0\end{aligned}$$

In order to find the critical points, we must find the values of θ such that this derivative is equal to zero.

$$\begin{aligned}0 &= \frac{-n_1}{\theta} + \frac{n_1 \bar{x}}{\theta^2} \\ \frac{n_1}{\theta} &= \frac{n_1 \bar{x}}{\theta^2} \\ n_1 \theta &= \frac{n_1}{\bar{x}} \\ \theta &= \bar{x}\end{aligned}$$

As computed above, the critical point $\theta = \bar{x}$ is a candidate for the maximum likelihood estimator for θ_0 we will now compute the second derivative test below to show that this critical point is in fact a maximum.

$$\begin{aligned}\frac{d^2}{d\theta^2}\ell(\theta) &= \frac{d^2}{d\theta^2}\left(\frac{-n_1}{\theta} + \frac{n_1 \bar{x}}{\theta^2}\right) \\ &= \frac{n_1}{\theta^2} - \frac{2n_1 \bar{x}}{\theta^3} \\ &= \frac{n_1}{\theta^2} \left[1 - \frac{2\bar{x}}{\theta}\right], \theta > 0\end{aligned}$$

Thus $f''(\bar{x}) = \frac{n_1}{\bar{x}^2} \left[1 - \frac{2\bar{x}}{\bar{x}}\right] = \frac{-n_1}{\bar{x}^2}$. Since $n_1 > 0$ and $\bar{x}^2 > 0$, it follows that $f''(\bar{x}) < 0$, and hence $\theta = \bar{x}$ is a local maximum. Thus it follows that $\hat{\theta}_0^{mle1} = \bar{x}$.

Deriving $\hat{\theta}_0^{mle2}$:

To find the maximum likelihood estimator of θ_0 for the second sample, we must find the likelihood function, take the natural log of the likelihood function, find the critical point of the log-likelihood function, and lastly run the second derivative test to show the critical point is a maximum. First off, we will compute the

likelihood function as shown below.

$$\begin{aligned}
L(\theta) &= g(y_1) \times g(y_2) \times \cdots \times g(y_{n_2}) \\
&= \frac{1}{k\theta} e^{\frac{-1}{k\theta} y_1} \times \frac{1}{k\theta} e^{\frac{-1}{k\theta} y_2} \times \cdots \times \frac{1}{k\theta} e^{\frac{-1}{k\theta} y_{n_2}} \\
&= \left(\frac{1}{k\theta} \right)^{n_2} e^{\frac{-1}{k\theta} (y_1 + y_2 + \cdots + y_{n_2})} \\
&= \left(\frac{1}{k\theta} \right)^{n_2} e^{\frac{-1}{k\theta} n_2 \bar{y}}, \theta > 0
\end{aligned}$$

Thus, as computed above, the likelihood function of θ is $L(\theta) = \left(\frac{1}{k\theta} \right)^{n_2} e^{\frac{-1}{k\theta} n_2 \bar{y}}, \theta > 0$. However, notice that taking the derivative of this function will not be easy, thus we will take the natural log of the likelihood function to turn multiplication into addition. This process is shown below.

$$\begin{aligned}
\ell(\theta) &= \ln(L(\theta)) \\
&= \ln \left(\left(\frac{1}{k\theta} \right)^{n_2} e^{\frac{-1}{k\theta} n_2 \bar{y}} \right) \\
&= \ln \left(\left(\frac{1}{k\theta} \right)^{n_2} \right) + \ln \left(e^{\frac{-1}{k\theta} n_2 \bar{y}} \right) \\
&= n_2 \ln \left(\frac{1}{k\theta} \right) + \frac{-n_2 \bar{y}}{k\theta}, \theta > 0
\end{aligned}$$

As computed above, the log-likelihood function is $\ell(\theta) = n_2 \ln \left(\frac{1}{k\theta} \right) + \frac{-n_2 \bar{y}}{k\theta}, \theta > 0$. Now we must find the critical points of this function in order to find the candidates for the maximum likelihood estimator of θ_0 . This derivative calculation is shown below.

$$\begin{aligned}
\frac{d}{d\theta} \ell(\theta) &= \frac{d}{d\theta} \left(n_2 \ln \left(\frac{1}{k\theta} \right) + \frac{-1}{k\theta} (n_2 \bar{y}) \right) \\
&= \frac{\frac{-n_2 k}{k^2 \theta^2}}{\frac{1}{k\theta}} + \frac{k n_2 \bar{y}}{k^2 \theta^2} \\
&= \frac{-n_2}{\theta} + \frac{n_2 \bar{y}}{k\theta^2}, \theta > 0
\end{aligned}$$

In order to find the critical points, we must find the values of θ such that this derivative is equal to zero.

$$\begin{aligned}
0 &= \frac{-n_2}{\theta} + \frac{n_2 \bar{y}}{k\theta^2} \\
\frac{n_2}{\theta} &= \frac{n_2 \bar{y}}{k\theta^2} \\
k n_2 \theta &= n_2 \bar{y} \\
\theta &= \frac{\bar{y}}{k}
\end{aligned}$$

As computed above, the critical point $\theta = \frac{\bar{y}}{k}$ is a candidate for the maximum likelihood estimator for θ_0 we will now compute the second derivative test below to show that this critical point is in fact a maximum.

$$\begin{aligned}
\frac{d^2}{d\theta^2} \ell(\theta) &= \frac{d^2}{d\theta^2} \left(\frac{-n_2}{\theta} + \frac{n_2 \bar{y}}{k\theta^2} \right) \\
&= \frac{n_2}{\theta^2} - \frac{2n_2 \bar{y}}{k\theta^3} \\
&= \frac{n_2}{\theta^2} \left(1 - \frac{2\bar{y}}{k\theta} \right), \theta > 0
\end{aligned}$$

Thus $f''\left(\frac{\bar{y}}{k}\right) = \frac{k^2 n_2}{\bar{y}^2} (1 - 2) = \frac{-k^2 n_2}{\bar{y}^2}$. Since $\frac{-k^2 n_2}{\bar{y}^2} < 0$, it follows that $f''\left(\frac{\bar{y}}{k}\right) < 0$, and hence $\theta = \frac{\bar{y}}{k}$ is a local maximum. Thus it follows that $\hat{\theta}_0^{mle1} = \frac{\bar{y}}{k}$.

Deriving the linear combination:

As explained above the scientist wants a single estimate of θ_0 , so they ask two statisticians for advice. One suggests finding the linear combination $a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2}$, with the smallest variance. To find this estimate we will take the variance of the linear combination and find the value of a that minimizes the variance, that will then be our estimate for θ_0 .

To find the a that minimizes the variance of $a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2}$, we will find an expression for $Var\left[a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2}\right]$ and take the first derivative to find the critical point, then we will use the second derivative test to show that the critical point is a minimum. In the following derivation we will use the fact that if $X \sim Exp(\theta_0)$, then $Var[\bar{X}] = \frac{\theta_0^2}{n}$. Furthermore since the two samples are independent, we can say that $Var\left[a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2}\right] = Var\left[a\hat{\theta}_0^{mle1}\right] + Var\left[(1-a)\hat{\theta}_0^{mle2}\right]$. With these facts we will derive an expression below.

$$\begin{aligned} Var\left[a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2}\right] &= a^2 Var\left[\hat{\theta}_0^{mle1}\right] + (1-a)^2 Var\left[\hat{\theta}_0^{mle2}\right] \quad (\text{Non-linearity of Variance}) \\ &= a^2 Var\left[\bar{X}\right] + (1-a)^2 Var\left[\frac{\bar{Y}}{k}\right] \\ &= a^2 Var\left[\bar{X}\right] + \frac{(1-a)^2}{k^2} Var\left[\bar{Y}\right] \quad (\text{Non-linearity of Variance}) \\ &= \frac{a^2 \theta_0^2}{n_1} + \frac{(1-a)^2 (k\theta_0)^2}{n_2 k^2} \\ &= \frac{a^2 \theta_0^2}{n_1} + \frac{(1-a)^2 \theta_0^2}{n_2} \end{aligned}$$

If we now let $f(a) = \frac{a^2 \theta_0^2}{n_1} + \frac{(1-a)^2 \theta_0^2}{n_2}$, then we can take the derivative with respect to a and obtain:

$$f'(a) = \frac{2a\theta_0^2}{n_1} - \frac{2(1-a)\theta_0^2}{n_2}$$

Setting this to zero and solving for θ_0 we obtain:

$$\begin{aligned} 0 &= \frac{2a\theta_0^2}{n_1} - \frac{2(1-a)\theta_0^2}{n_2} \\ \frac{2(1-a)\theta_0^2}{n_2} &= \frac{2a\theta_0^2}{n_1} \\ \frac{1-a}{n_2} &= \frac{a}{n_1} \\ (1-a)n_1 &= an_2 \\ (n_1 + n_2)a &= n_1 \\ a &= \frac{n_1}{n_1 + n_2} \end{aligned}$$

Now we will run the second derivative test to show this value of a is a minimum.

$$\begin{aligned} f'(a) &= \frac{2a\theta_0^2}{n_1} - \frac{2(1-a)\theta_0^2}{n_2} \\ f''(a) &= \frac{2\theta_0^2}{n_1} + \frac{2\theta_0^2}{n_2} \end{aligned}$$

Since $f''(a) = \frac{2\theta_0^2}{n_1} + \frac{2\theta_0^2}{n_2} > 0 \quad \forall \theta$ it follows that $f''(\frac{n_1}{n_1+n_2}) > 0$ and hence $a = \frac{n_1}{n_1+n_2}$ is the argmin of $Var \left[a\hat{\theta}_0^{mle1} + (1-a)\hat{\theta}_0^{mle2} \right]$. Therefore our linear combination estimate of θ_0 is $\hat{\theta} = \frac{n_1}{n_1+n_2}\bar{x} + (1 - \frac{n_1}{n_1+n_2})\frac{\bar{y}}{k} = \frac{n_1\bar{x} + n_2\frac{\bar{y}}{k}}{n_1+n_2}$.

Deriving the combined sample MLE:

As explained above the scientist wants a single estimate of θ_0 , so they ask two statisticians for advice. The second suggests finding the MLE from the combined sample.

To find the maximum likelihood estimator of θ_0 for the combined sample, we must find the combined likelihood function, take the natural log of the combined likelihood function, find the critical point of the combined log-likelihood function, and lastly run the second derivative test to show the critical point is a maximum. First off, we will compute the combined likelihood function as shown below.

$$\begin{aligned}
L(\theta) &= f(x_1) \times f(x_2) \times \cdots \times f(x_{n_1}) \times g(y_1) \times g(y_2) \times \cdots \times g(y_{n_2}) \\
&= \frac{1}{\theta} e^{-\frac{1}{\theta}x_1} \times \frac{1}{\theta} e^{-\frac{1}{\theta}x_2} \times \cdots \times \frac{1}{\theta} e^{-\frac{1}{\theta}x_{n_1}} \times \frac{1}{k\theta} e^{-\frac{1}{k\theta}y_1} \times \frac{1}{k\theta} e^{-\frac{1}{k\theta}y_2} \times \cdots \times \frac{1}{k\theta} e^{-\frac{1}{k\theta}y_{n_2}} \\
&= \left(\frac{1}{\theta}\right)^{n_1} \left(\frac{1}{k\theta}\right)^{n_2} e^{-\frac{1}{\theta} \sum_{i=1}^{n_1} x_i} e^{-\frac{1}{k\theta} \sum_{i=1}^{n_2} y_i} \\
&= \frac{1}{k^{n_2} \theta^{n_1+n_2}} e^{-\frac{1}{\theta} \sum_{i=1}^{n_1} x_i - \frac{1}{k\theta} \sum_{i=1}^{n_2} y_i} \\
&= \frac{1}{k^{n_2} \theta^{n_1+n_2}} e^{-\frac{1}{\theta} [\sum_{i=1}^{n_1} x_i + \frac{1}{k} \sum_{i=1}^{n_2} y_i]} \\
&= \frac{1}{k^{n_2} \theta^{n_1+n_2}} e^{-\frac{1}{\theta} [n_1\bar{x} + \frac{1}{k} n_2\bar{y}]}, \theta > 0
\end{aligned}$$

Thus, as computed above, the likelihood function of θ is $L(\theta) = \frac{1}{k^{n_2} \theta^{n_1+n_2}} e^{-\frac{1}{\theta} [n_1\bar{x} + \frac{1}{k} n_2\bar{y}]}$, $\theta > 0$. However, notice that taking the derivative of this function will not be easy, thus we will take the natural log of the likelihood function to turn multiplication into addition. This process is shown below.

$$\begin{aligned}
\ell(\theta) &= \ln(L(\theta)) \\
&= \ln \left(\frac{1}{k^{n_2} \theta^{n_1+n_2}} e^{-\frac{1}{\theta} [n_1\bar{x} + \frac{1}{k} n_2\bar{y}]} \right) \\
&= \ln \left(\frac{1}{k^{n_2} \theta^{n_1+n_2}} \right) + \ln \left(e^{-\frac{1}{\theta} [n_1\bar{x} + \frac{1}{k} n_2\bar{y}]} \right) \\
&= \ln(1) - \ln(k^{n_2} \theta^{n_1+n_2}) - \frac{1}{\theta} \left[n_1\bar{x} + \frac{1}{k} n_2\bar{y} \right] \\
&= -\ln(k^{n_2} \theta^{n_1+n_2}) - \frac{1}{\theta} \left[n_1\bar{x} + \frac{1}{k} n_2\bar{y} \right] \\
&= -\ln(k^{n_2}) - \ln(\theta^{n_1+n_2}) - \frac{1}{\theta} \left[n_1\bar{x} + \frac{1}{k} n_2\bar{y} \right] \\
&= -n_2 \ln(k) - (n_1 + n_2) \ln(\theta) - \frac{1}{\theta} \left[n_1\bar{x} + \frac{1}{k} n_2\bar{y} \right], \theta > 0
\end{aligned}$$

As computed above, the log-likelihood function is $\ell(\theta) = -n_2 \ln(k) - (n_1 + n_2) \ln(\theta) - \frac{1}{\theta} [n_1\bar{x} + \frac{1}{k} n_2\bar{y}]$, $\theta > 0$. Now we must find the critical points of this function in order to find the candidates for the maximum likelihood estimator of θ_0 . This derivative calculation is shown below.

$$\begin{aligned}
\frac{d}{d\theta} \ell(\theta) &= \frac{d}{d\theta} \left(-n_2 \ln(k) - (n_1 + n_2) \ln(\theta) - \frac{1}{\theta} \left[n_1\bar{x} + \frac{1}{k} n_2\bar{y} \right] \right) \\
&= \frac{-(n_1 + n_2)}{\theta} + \frac{n_1\bar{x} - \frac{1}{k} n_2\bar{y}}{\theta^2}, \theta > 0
\end{aligned}$$

In order to find the critical points, we must find the values of θ such that this derivative is equal to zero.

$$\begin{aligned}
0 &= \frac{-(n_1 + n_2)}{\theta} + \frac{n_1\bar{x} - \frac{1}{k}n_2\bar{y}}{\theta^2} \\
\frac{n_1 + n_2}{\theta} &= \frac{n_1\bar{x} - \frac{1}{k}n_2\bar{y}}{\theta^2} \\
\theta(n_1 + n_2) &= n_1\bar{x} + \frac{n_2}{k}\bar{y} \\
\theta &= \frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2}
\end{aligned}$$

As computed above, the critical point $\theta = \frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2}$ is a candidate for the maximum likelihood estimator for θ_0 we will now compute the second derivative test below to show that this critical point is in fact a maximum.

$$\begin{aligned}
\frac{d^2}{d\theta^2}\ell(\theta) &= \frac{d^2}{d\theta^2} \left(\frac{-(n_1 + n_2)}{\theta} + \frac{n_1\bar{x} - \frac{1}{k}n_2\bar{y}}{\theta^2} \right) \\
&= \frac{n_1 + n_2}{\theta^2} - \frac{2(n_1\bar{x} + \frac{1}{k}n_2\bar{y})}{\theta^3}, \theta > 0
\end{aligned}$$

Thus $f'' \left(\frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2} \right) = \frac{n_1 + n_2}{\theta^2} - \frac{2(n_1\bar{x} + \frac{1}{k}n_2\bar{y})}{\theta^3} = \frac{n_1 + n_2}{\left(\frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2} \right)^2} - \frac{2(n_1\bar{x} + \frac{1}{k}n_2\bar{y})}{\left(\frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2} \right)^3} = \frac{-(n_1 + n_2)^2}{n_1\bar{x} + \frac{n_2}{k}\bar{y}}$. Since $\frac{-(n_1 + n_2)^2}{n_1\bar{x} + \frac{n_2}{k}\bar{y}} < 0$,

it follows that $f'' \left(\frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2} \right) < 0$, and hence $\theta = \frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2}$ is a local maximum. Thus it follows that $\hat{\theta}_0^{mle1} = \frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2}$.

Conclusion:

As has been computed in the last two parts, the two methods yield the same estimate for θ_0 , namely $\hat{\theta} = \frac{n_1\bar{x} + \frac{n_2}{k}\bar{y}}{n_1 + n_2}$.