# PS2

Samuel Olson

## Problem 1

7.11, Casella & Berger

Let $X_1, \ldots, X_n$ be iid with pdf

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 \leq x \leq 1, \quad 0 < \theta < \infty.$$

Hint: In part (a), you can assume each observation lies in $X_i \in (0,1)$ for finding the MLE (since there is zero probability of "some $X_i = 0$ or 1 for $i = 1, \ldots, n$"). To find the variance in part (a), you should be able to show that $Y_i = -\log(X_i)$ has an exponential distribution with scale parameter $\beta = 1/\theta > 0$ so that

$$W = \sum_{i=1}^{n} Y_i$$

has a gamma ($\alpha = n, \beta$) distribution; then, you can compute the variance by finding moments $E_\theta(W^{-1})$ and $E_\theta(W^{-2})$.

### a)

Find the MLE of $\theta$, and show that its variance $\to 0$ as $n \to \infty$.

Given the probability density function pdf provided, the likelihood function for is given by:

$$L(\theta) = \prod_{i=1}^{n} \theta X_i^{\theta-1}$$

To make the evaluation easier, consider the log likelihood:

$$\log(L(\theta)) = \sum_{i=1}^{n} \log(\theta) + (\theta - 1)\sum_{i=1}^{n} \log(X_i) = n\log(\theta) + (\theta - 1)\sum_{i=1}^{n} \log(X_i)$$

We attempt to find a maximum by differentiating with respect to $\theta$ and setting the expression to 0:

$$\frac{d\log(L(\theta))}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \log(X_i) = 0 \to \hat{\theta}_{\text{MLE}} = -\frac{n}{\sum_{i=1}^{n} \log(X_i)}$$

To confirm this is a maximum, we then also check that the second derivative is negative:

$$\frac{d^2\log(L(\theta))}{d\theta} = -\frac{n}{\theta^2} < 0 \text{ given n} > 0$$

So our calculation is indeed an MLE.

We then must identify the variance of our MLE. To that end, define $Y_i = -\log(X_i)$, which follows an exponential distribution:

$$Y_i \sim \text{Exp}(1/\theta)$$

As parametrized, the sum $W = \sum_{i=1}^{n} Y_i$ follows a Gamma distribution:

$$W \sim \text{Gamma}(n, 1/\theta)$$

Utilizing some useful properties of the Gamma, we then know:

$$E(W^{-1}) = \frac{\theta}{n-1}, \quad E(W^{-2}) = \frac{\theta^2}{(n-1)(n-2)}$$

Since $\hat{\theta}_{\text{MLE}} = \frac{n}{W}$, the variance is:

$$\text{Var}(\hat{\theta}_{\text{MLE}}) = n^2 E(W^{-2}) - (nE(W^{-1}))^2 = \frac{n^2\theta^2}{(n-1)(n-2)} - \frac{n^2\theta^2}{(n-1)^2} = \frac{n^2\theta^2}{(n-1)^2(n-2)}$$

Consider the behavior of this function as n increases. Namely, since the denominator grows faster than the numerator, noting the expression in the denominator is n to a degree 3 and for the numerator n to a degree 2, we conclude:

$$\text{Var}(\hat{\theta}_{\text{MLE}}) \to 0 \quad \text{as} \quad n \to \infty.$$

## b)

Find the method of moments estimator of $\theta$.

We start by considering the first moment of $X$ given by:

$$E(X) = \int_0^1 xf(x|\theta)\,dx = \int_0^1 x\theta x^{\theta-1}\,dx = \theta \int_0^1 x^\theta\,dx$$

For the sake of space, breaking this up and evaluating by:

$$E(X) = \theta \left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1 = \theta\frac{1}{\theta+1} = \frac{\theta}{\theta+1}$$

By definition, the sample mean is:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then, by equating the sample mean equal to the population mean, we have:

$$\frac{\theta}{\theta+1} = \bar{X} \to \theta = \bar{X}(\theta+1) = \bar{X}\theta + \bar{X}$$

Our goal is to identify a formula with just $\theta$ on one side, so to that end, we have:

$$\theta - \bar{X}\theta = \bar{X} \to \theta(1 - \bar{X}) = \bar{X} \to \theta = \frac{\bar{X}}{1 - \bar{X}}$$

I'd be remiss not to note one potential issue though. We can't divide by zero, so we cannot have $\bar{X} < 1$. Under this condition, the method of moments estimator is not valid.

Bearing in mind that condition then, we say the method of moments estimator of $\theta$ is:

$$\hat{\theta}_{\text{MM}} = \frac{\bar{X}}{1 - \bar{X}}$$

# Problem 2

7.12(a), Casella & Berger

Let $X_1, \dots, X_n$ be a random sample from a population with pmf

$$P_\theta(X = x) = \theta^x (1-\theta)^{1-x}, \quad x = 0 \text{ or } 1, \quad 0 \le \theta \le \frac{1}{2}.$$

Hint: Note that the parameter space is $\Theta \equiv [0, 1/2]$. In maximizing the likelihood, it might be clearest to consider three data cases:

1. $\sum_{i=1}^n X_i = 0$;
2. $\sum_{i=1}^n X_i = n$; or
3. $0 < \sum_{i=1}^n X_i < n$.

In the last case, the derivative of log-likelihood $L(\theta)$ indicates that $L(\theta)$ is increasing on $(0, \bar{X}_n)$ and decreasing on $(\bar{X}_n, 1)$.

## a)

Find the method of moments estimator and MLE of $\theta$.

I am starting this problem in the order of what was specified. To that end:

### Method of Moments

As specified, the information above corresponds to a Bernoulli distributed random variable variable. Treating this as a given, the population mean is the parameter, i.e.:

$$E(X) = \theta$$

Equating this with the sample mean, which by definition is given by:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Solving for $\theta$:

$$\hat{\theta} = \bar{X}$$

Note that $\theta$ is restricted to $[0, 1/2]$, we need to make sure this condition is met, therefore we write our method of moments estimator as:

$$\hat{\theta} = \min\left(\bar{X}, \frac{1}{2}\right)$$

**Maximum Likelihood Estimator (MLE)**

The likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} P_\theta(X_i) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i}$$

Define $S = \sum_{i=1}^{n} X_i$, which allows us to simplify the likelihood function as:

$$L(\theta) = \theta^S(1-\theta)^{n-S}$$

We then take the log of the likelihood function to make evaluation easier:

$$\log(L(\theta)) = S\log\theta + (n-S)\log(1-\theta)$$

Differentiate and set the expression equal to 0, giving us:

$$\frac{\log(L(\theta))}{d\theta} = \frac{S}{\theta} - \frac{n-S}{1-\theta} = 0 \to \frac{S}{\theta} = \frac{n-S}{1-\theta}$$

We isolate and solve for $\theta$:

$$\theta = \frac{S}{n} = \bar{X}$$

To ensure this is the maximum, we evaluate the second derivative:

$$\frac{d^2\log(L(\theta))}{d\theta^2} = -\frac{S}{\theta^2} - \frac{n-S}{(1-\theta)^2}$$

Since $S \geq 0$ and $n - S \geq 0$, and $\theta \in (0,1)$, the second derivative is always negative. This ensures that $\theta = \bar{X}$ is a maximum.

Then, note that the parameter space is $\Theta \equiv [0, 1/2]$. If $\bar{X} \leq 1/2$, the MLE is $\hat{\theta}_{\text{MLE}} = \bar{X}$. If $\bar{X} > 1/2$, the likelihood function is decreasing for $\theta > 1/2$, so, similar to the method of moments estimator, we need to incorporate these conditions when giving the MLE. To that end, the MLE is given by:

$$\hat{\theta} = \min\left(\bar{X}, \frac{1}{2}\right)$$

# Problem 3

7.14, Casella & Berger

Let $X$ and $Y$ be independent exponential random variables, with

$$f(x|\lambda) = \frac{1}{\lambda}e^{-x/\lambda}, \quad x > 0, \quad f(y|\mu) = \frac{1}{\mu}e^{-y/\mu}, \quad y > 0.$$

We observe $Z$ and $W$ with

$$Z = \min(X, Y) \quad \text{and} \quad W = \begin{cases} 1 & \text{if } Z = X \\ 0 & \text{if } Z = Y. \end{cases}$$

In Exercise 4.26, the joint distribution of $Z$ and $W$ was obtained. Now assume that $(Z_i, W_i), i = 1, \ldots, n$, are $n$ iid observations. Find the MLEs of $\lambda$ and $\mu$.

Hint: You may use that the joint density of $(Z, W)$ is

$$f(z, w|\lambda, \mu) = \frac{dF(z, w)}{dz} = \begin{cases} \mu^{-1}e^{-z(\lambda+\mu^{-1})}, & z > 0, w = 0 \\ \lambda^{-1}e^{-z(\lambda+\mu^{-1})}, & z > 0, w = 1 \end{cases}$$

where

$$F(z, w|\lambda, \mu) = P(Z \leq z, W = w|\lambda, \mu).$$

Then, based on a random sample $(Z_i, W_i), i = 1, \ldots, n$ of pairs, this problem involves using calculus with two variables to find the MLE.

## Answer/Proof, Whatchumacallit

**Finding the joint density (joint pdf)**

$$P(Z \leq z, W = 0) = P(\min(X, Y) \leq z, Y \leq X).$$

By conditioning on $Y$, we integrate over the possible values of $Y$:

$$P(Z \leq z, W = 0) = \int_0^z \int_y^\infty f_X(x)f_Y(y)dxdy = \int_0^z \int_y^\infty \frac{1}{\lambda}e^{-x/\lambda}\frac{1}{\mu}e^{-y/\mu}dxdy = \int_y^\infty \frac{1}{\lambda}e^{-x/\lambda}dx = e^{-y/\lambda}$$

Continuing to simplify:

$$P(Z \leq z, W = 0) = \int_0^z \frac{1}{\mu}e^{-y/\mu}e^{-y/\lambda}dy = \int_0^z \frac{1}{\mu}e^{-y(1/\lambda+1/\mu)}dy$$

Leaving us with the expression:

$$P(Z \leq z, W = 0) = \frac{\lambda}{\lambda + \mu}\left(1 - e^{-z(1/\lambda+1/\mu)}\right)$$

Differentiating wrt $z$, we find the joint pdf to be:

$$f(z,0) = \frac{\lambda}{\lambda+\mu}\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)e^{-z(1/\lambda+1/\mu)}, \quad z>0$$

Now, for W = 0, we similarly have (via symmetry argument):

$$P(Z \leq z, W=1) = P(\min(X,Y) \leq z, X \leq Y) = \frac{\mu}{\lambda+\mu}\left(1-e^{-z(1/\lambda+1/\mu)}\right)$$

Differentiating to derive the pdf gives us:

$$f(z,1) = \frac{\mu}{\lambda+\mu}\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)e^{-z(1/\lambda+1/\mu)}, \quad z>0$$

So, the joint pdf is then defined by the expression:

$$f(z,w|\lambda,\mu) = \begin{cases} \frac{\lambda}{\lambda+\mu}\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)e^{-z(1/\lambda+1/\mu)}, & w=0, \\ \frac{\mu}{\lambda+\mu}\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)e^{-z(1/\lambda+1/\mu)}, & w=1 \end{cases}$$

**Likelihood**

Given $n$ i.i.d. observations $(Z_i, W_i), i = 1, \ldots, n$, the likelihood function is:

$$L(\lambda,\mu) = \prod_{i=1}^{n} f(z_i, w_i|\lambda,\mu)$$

Using the derived joint pdf, we may write:

$$L(\lambda,\mu) = \prod_{i=1}^{n}\left(\frac{\lambda}{\lambda+\mu}\right)^{1-w_i}\left(\frac{\mu}{\lambda+\mu}\right)^{w_i}\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)e^{-z_i(1/\lambda+1/\mu)}$$

Noting the support of W determines just the “front” part of the joint pdf, and all other terms remain the same.

Now, define:

$$S_z = \sum_{i=1}^{n} Z_i, \quad S_w = \sum_{i=1}^{n} W_i$$

This allows us to rewrite our above likelihood function as:

$$L(\lambda,\mu) = \left(\frac{\lambda}{\lambda+\mu}\right)^{n-S_w}\left(\frac{\mu}{\lambda+\mu}\right)^{S_w}\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)^{n}e^{-S_z(\lambda^{-1}+\mu^{-1})}$$

Taking the log-likelihood to make maximization easier gives us:

$$\log L(\lambda,\mu) = (n-S_w)\log\frac{\lambda}{\lambda+\mu} + S_w\log\frac{\mu}{\lambda+\mu} + n\log\left(\frac{1}{\lambda}+\frac{1}{\mu}\right) - S_z\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)$$

We will then derive wrt the parameters of interest to determine the MLEs (of each parameter, resp.)

**MLE for $\lambda$** Taking the partial derivative:

$$\frac{\partial \log L}{\partial \lambda} = \frac{n - S_w}{\lambda} - \frac{n}{\lambda(\lambda + \mu)} - S_z \left(-\frac{1}{\lambda^2}\right) = 0$$

Solving for $\lambda$:

$$\hat{\lambda} = \frac{S_z}{S_w}$$

To ensure this is a maximum, consider the second derivative:

$$\frac{\partial^2 \log L}{\partial \lambda^2} = \frac{S_w}{\lambda^2} - \frac{2S_z}{\lambda^3}.$$

For $\hat{\lambda} = \frac{S_z}{S_w}$:

$$\frac{\partial^2 \log L}{\partial \lambda^2}\Big|_{\lambda=\hat{\lambda}} = \frac{S_w}{(S_z/S_w)^2} - \frac{2S_z}{(S_z/S_w)^3} = -\frac{S_w^3}{S_z^2} < 0$$

So we have verified this is a maximum.

**MLE for $\mu$** Taking the partial derivative:

$$\frac{\partial \log L}{\partial \mu} = \frac{S_w}{\mu} - \frac{n}{\mu(\lambda + \mu)} - S_z \left(-\frac{1}{\mu^2}\right) = 0$$

Solving for $\mu$:

$$\hat{\mu} = \frac{S_z}{n - S_w}$$

To ensure this is a maximum, consider the second derivative:

$$\frac{\partial^2 \log L}{\partial \mu^2} = \frac{n - S_w}{\mu^2} - \frac{2S_z}{\mu^3}$$

For $\hat{\mu} = \frac{S_z}{(n-S_w)}$:

$$\frac{\partial^2 \log L}{\partial \mu^2}\Big|_{\mu=\hat{\mu}} = -\frac{(n - S_w)^3}{S_z^2} < 0$$

So we have verified this is a maximum.

**Conclusion** The maximum likelihood estimators are:

$$\hat{\lambda} = \frac{S_z}{S_w} = \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n W_i}, \quad \hat{\mu} = \frac{S_z}{n - S_w} = \frac{\sum_{i=1}^n Z_i}{n - \sum_{i=1}^n W_i}$$

# Problem 4

7.49, Casella & Berger

Let $X_1, \ldots, X_n$ be iid exponential $(\lambda)$.

## a)

Find an unbiased estimator of $\lambda$ based only on $Y = \min\{X_1, \ldots, X_n\}$.

$Y = X_{(1)}$ has pdf:

$$f_Y(y) = \frac{n!}{(n-1)!}\frac{1}{\lambda}e^{-y/\lambda}\left[1-(1-e^{-y/\lambda})\right]^{n-1} = \frac{n}{\lambda}e^{-ny/\lambda}$$

Thus, $Y \sim \text{Exponential}(\lambda/n)$, so $E[Y] = \lambda/n$ and $nY$ is an unbiased estimator of $\lambda$.

## b)

Find a better estimator than the one in part (a). Prove that it is better.

Since $f_X(x)$ is in the exponential family, $\sum_i X_i$ is a complete sufficient statistic. The expectation $E[nX_{(1)}|\sum_i X_i]$ provides the best unbiased estimator of $\lambda$. Since $E[\sum_i X_i] = n\lambda$, we must have $E[nX_{(1)}|\sum_i X_i] = \sum_i X_i/n$ by completeness.

Since any function of $\sum_i X_i$ that is unbiased for $\lambda$ is the best unbiased estimator, we conclude that:

$$\hat{\lambda} = \frac{\sum_i^n X_i}{n}$$

is the best unbiased estimator of $\lambda$.

## c)

The following data are high-stress failure times (in hours) of Kevlar/epoxy spherical vessels used in a sustained pressure environment on the space shuttle:

$$50.1, \quad 70.1, \quad 137.0, \quad 166.9, \quad 170.5, \quad 152.8, \quad 80.5, \quad 123.5, \quad 112.6, \quad 148.5, \quad 160.0, \quad 125.4.$$

Failure times are often modeled with the exponential distribution. Estimate the mean failure time using the estimators from parts (a) and (b).

Some computations:

```r
data <- c(50.1, 70.1, 137, 166.9, 170.5, 152.8, 80.5, 123.5, 112.6, 148.5, 160, 125.4)
min(data) * 12
```

```
## [1] 601.2
```

```r
sum(data)
```

```
## [1] 1497.9
```

```r
mean(data)
```

```
## [1] 124.825
```

From part (a):

$$\hat{\lambda}_Y = n \min(X_i) = 12(50.1) = 601.2$$

From part (b):

$$\hat{\lambda} = \frac{\sum_i X_i}{n} = \frac{1497.9}{12} = 124.825$$

Welp, those are pretty different estimates.

# Problem 5

Suppose someone collects a random sample $X_1, X_2, \ldots, X_n$ from an exponential $\beta = 1/\theta$ distribution with pdf

$$f(x|\theta) = \theta e^{-\theta x}, \quad x > 0,$$

and a parameter $\theta > 0$. However, due to a recording mistake, only truncated integer data $Y_1, Y_2, \ldots, Y_n$ are available for analysis, where $Y_i$ represents the integer part of $X_i$ after dropping all digits after the decimal place in $X_i$'s representation. (For example, if $x_1 = 4.9854$ in reality, we would have only $y_1 = 4$ available.) Then, $Y_1, \ldots, Y_n$ represent a random sample of iid (discrete) random variables with pmf

$$f(y|\theta) = P_\theta(Y_i = y) = e^{-\theta y} - e^{-\theta(1+y)}, \quad y = 0, 1, 2, 3, \ldots.$$

## a)

Show that the likelihood equals

$$L(\theta) = \left[e^{-\theta \bar{Y}_n}(1 - e^{-\theta})\right]^n,$$

where $\bar{Y}_n$ is the sample average.

Given the pmf of $Y_i$ is:

$$f(y|\theta) = P_\theta(Y_i = y) = e^{-\theta y} - e^{-\theta(1+y)}, \quad y = 0, 1, 2, 3, \ldots.$$

We note that the likelihood function for the sample $Y_1, Y_2, \ldots, Y_n$ is:

$$L(\theta) = \prod_{i=1}^{n} f(y_i|\theta) = \prod_{i=1}^{n} \left(e^{-\theta y_i} - e^{-\theta(1+y_i)}\right)$$

Simplifying a bit gives us:

$$L(\theta) = \prod_{i=1}^{n} e^{-\theta y_i}\left(1 - e^{-\theta}\right) = \left(1 - e^{-\theta}\right)^n \prod_{i=1}^{n} e^{-\theta y_i} = \left(1 - e^{-\theta}\right)^n e^{-\theta \sum_{i=1}^{n} y_i}$$

By definition, the sample average, $\bar{Y}_n$, is defined as:

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \to \sum_{i=1}^{n} y_i = n\bar{Y}_n$$

Using the above, we can incorporate this expression into the simplified likelihood function, and have:

$$L(\theta) = \left(1 - e^{-\theta}\right)^n e^{-\theta n \bar{Y}_n} = \left[e^{-\theta \bar{Y}_n}(1 - e^{-\theta})\right]^n$$

Achieving our desired result. Wa-hoo!

## b)

If $Y_n = \sum_{i=1}^{n} Y_i/n = 0$, show that an MLE for $\theta$ does not exist on the parameter space $(0, \infty)$.

(Recall: $Y_i$ is discrete and this corresponds to a pathological MLE case mentioned in class: $Y_1 = \cdots = Y_n = 0$. This event can happen but typically with small probability for large $n$.)

Note again, the likelihood function as given at the end of part a):

$$L(\theta) = \left[e^{-\theta \bar{Y}_n}(1 - e^{-\theta})\right]^n = \left[e^{-\theta(0)}(1 - e^{-\theta})\right]^n = \left[(1 - e^{-\theta})\right]^n$$

The goal is to find the MLE by maximizing $L(\theta)$ over $\theta > 0$.

The function $1 - e^{-\theta}$ is increasing in $\theta$, approaching 1 as $\theta \to \infty$. Since it is raised to the power $n$, we have:

$$\lim_{\theta \to 0} L(\theta) = (1 - 1)^n = 0$$

And

$$\lim_{\theta \to \infty} L(\theta) = (1 - 0)^n = 1$$

Since $L(\theta)$ is strictly increasing in $\theta$, it attains its supremum at $\theta \to \infty$. An MLE must be a finite value $\theta^*$ that maximizes $L(\theta)$ in the domain $(0, \infty)$. However, $L(\theta)$ is maximized at $\theta \to \infty$, meaning the supremum is not attained at any finite value. Therefore, no finite $\theta$ maximizes $L(\theta)$, implying the MLE does not exist.

## c)

If $0 < \bar{Y}_n$, show that the MLE $\hat{\theta}$ is

$$\hat{\theta} = \log(\bar{Y}_n^{-1} + 1)$$

Given the likelihood function of part a):

$$L(\theta) = \left[e^{-\theta \bar{Y}_n}(1 - e^{-\theta})\right]^n$$

To find the MLE of $\theta$, we use our typical approach of maximizing the log-likelihood function:

$$\log(L(\theta)) = n\left(-\theta \bar{Y}_n + \log(1 - e^{-\theta})\right)$$

Differentiating and setting equal to zero gives us:

$$\frac{d\log(L(\theta))}{d\theta} = n\left(-\bar{Y}_n + \frac{e^{-\theta}}{1 - e^{-\theta}}\right) = 0 \to \frac{e^{-\theta}}{1 - e^{-\theta}} = \bar{Y}_n \to e^{-\theta} = \bar{Y}_n(1 - e^{-\theta})$$

Simplifying some more, we have:

$$e^{-\theta} + \bar{Y}_n e^{-\theta} = \bar{Y}_n \to e^{-\theta}(1 + \bar{Y}_n) = \bar{Y}_n$$

Solving for $\theta$, with note of the monotonic transformation given from the log function, gives us:

$$e^{-\theta} = \frac{\bar{Y}_n}{1+\bar{Y}_n} \to -\theta = \log\left(\frac{\bar{Y}_n}{1+\bar{Y}_n}\right) \to \theta = \log\left(\frac{1+\bar{Y}_n}{\bar{Y}_n}\right)$$

Giving us:

$$\hat{\theta} = \log\left(\bar{Y}_n^{-1} + 1\right)$$

However, we need to do one validation! To confirm that $\hat{\theta}$ is a maximum, we compute the second derivative of the log-likelihood function:

$$\frac{d^2 \log(L(\theta))}{d\theta^2} = n\left(\frac{-e^{-\theta}(1-e^{-\theta}) + e^{-2\theta}}{(1-e^{-\theta})^2}\right) = n\left(\frac{-e^{-\theta} + e^{-2\theta} - e^{-2\theta}}{(1-e^{-\theta})^2}\right) = n\left(\frac{-e^{-\theta}}{(1-e^{-\theta})^2}\right)$$

Since $e^{-\theta} > 0$ and $(1-e^{-\theta})^2 > 0$, it follows that:

$$\frac{d^2 \log(L(\theta))}{d\theta^2} = n\left(\frac{-e^{-\theta}}{(1-e^{-\theta})^2}\right) < 0$$

So we have verified that the MLE we calculated is in fact a maximum. This leads us to conclude that the MLE of $\theta$ is:

$$\hat{\theta} = \log\left(\bar{Y}_n^{-1} + 1\right)$$