STA 6351, Report.1.5

Carson Slater *Baylor University*

\textbf{Report 1.5. Truncated Poisson sampling}

(a)

Use R to generate $n = 10$ observations from (1.9.5) with $\theta = 3$.

We have that equation (1.9.5) in your screenshot is the **zero-truncated Poisson** distribution:

$$g(x;\theta) = \frac{e^{-\theta}\theta^x}{(1-e^{-\theta})x!}, \quad x = 1, 2, \ldots, \ \theta > 0.$$

That's just a Poisson distribution conditioned on $X > 0$. In R, you can simulate from it using the built-in `rpois()` but discarding zeros. Here's clean R code for part (a):

```
##  [1] 2 4 2 5 6 6 3 5 3 3
```

(b)

Apply (1.9.4) to obtain a table with the following columns: $\frac{\partial \ell}{\partial \theta}, \frac{\partial^2 \ell}{\partial \theta^2}$, θ, and ℓ.

Iter	$\partial\ell/\partial\theta$	Obs Hessian	Exp Hessian	θ	$\ell(\theta)$
0	2.47604	-3.78192	-2.95658	3.00000	13.35657
1	-0.05728	-2.42327	-2.43820	3.83747	14.29085
2	-0.00003	-2.45040	-2.45041	3.81398	14.29153
3	0.00000	-2.45042	-2.45042	3.81396	14.29153

(c)

Use the convergence criterion $|\hat{\theta}_{t+1} - \hat{\theta}_t| < 0.0001$ to obtain the MLE for θ.

From the table, we have that the MLE is 3.814.

(d)

Repeat this for $n = 20$ and $n = 100$ observations.

Iter	$\partial\ell/\partial\theta$	Obs Hessian	Exp Hessian	θ	$\ell(\theta)$
0	3.61875	-7.11940	-5.91315	3.00000	22.31869
1	-0.06759	-5.10171	-5.12042	3.61198	23.34274
2	-0.00003	-5.13539	-5.13539	3.59878	23.34319
3	0.00000	-5.13540	-5.13540	3.59878	23.34319

For 20 observations, we have that the MLE is 3.599.

Iter	$\partial\ell/\partial\theta$	Obs Hessian	Exp Hessian	θ	$\ell(\theta)$
0	-1.57290	-29.04146	-29.56576	3.00000	46.77534
1	-0.00351	-29.96568	-29.96687	2.94680	46.81749
2	0.00000	-29.96777	-29.96777	2.94668	46.81749
3	0.00000	-29.96777	-29.96777	2.94668	46.81749

For 100 observations, we have that the MLE is 2.947.

(e)

Show that the score equation for θ can be written in the form

$$U(\theta) \;=\; \frac{n}{\theta}\Big(\bar{x} - m(\theta)\Big),$$

where $m(\theta)$ is defined in (1.9.8). Conclude that the MLE $\hat{\theta}$ is the solution of the mean–matching equation

$$\bar{x} = m(\hat{\theta}).$$

The probability mass function (PMF) for the zero-truncated Poisson distribution is:

$$g(x;\theta) = \frac{e^{-\theta}\theta^x}{(1-e^{-\theta})x!}, \quad x = 1, 2, \dots$$

For a sample $\mathbf{X} = (X_1, \dots, X_n)$, the log-likelihood function is:

$$\ell(\theta) = \sum_{i=1}^{n} \left[-\theta + X_i \log(\theta) - \log(1-e^{-\theta}) - \log(X_i!)\right]$$

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The log-likelihood simplifies to:

$$\ell(\theta) = -n\theta + n\bar{X}\log(\theta) - n\log(1-e^{-\theta}) - \sum_{i=1}^{n}\log(X_i!)$$

The Score Function $U(\theta)$

The score function $U(\theta)$ is the first derivative of the log-likelihood with respect to θ:

$$U(\theta) = \frac{\partial \ell}{\partial \theta} = -n + \frac{n\bar{X}}{\theta} - n\left(\frac{e^{-\theta}}{1-e^{-\theta}}\right)$$

The expected value of the zero-truncated Poisson distribution, given in (1.9.8), is:

$$m(\theta) = \mathbb{E}[X] = \frac{\theta}{1-e^{-\theta}}$$

We can factor $\frac{n}{\theta}$ from the score function $U(\theta)$:

$$\begin{aligned} U(\theta) &= \frac{n}{\theta}\left[-\theta + \bar{X} - \frac{\theta e^{-\theta}}{1-e^{-\theta}}\right] \\ &= \frac{n}{\theta}\left[\bar{X} - \left(\theta + \frac{\theta e^{-\theta}}{1-e^{-\theta}}\right)\right] \end{aligned}$$

Now, simplify the terms inside the second parenthesis:

$$\theta + \frac{\theta e^{-\theta}}{1-e^{-\theta}} = \theta\left(1 + \frac{e^{-\theta}}{1-e^{-\theta}}\right) = \theta\left(\frac{1-e^{-\theta}+e^{-\theta}}{1-e^{-\theta}}\right) = \frac{\theta}{1-e^{-\theta}}$$

By definition, this is $m(\theta)$. Substituting this back into $U(\theta)$:

$$U(\theta) = \frac{n}{\theta}\left(\bar{X} - m(\theta)\right)$$

The Maximum Likelihood Estimator (MLE) $\hat{\theta}$ is the solution to the score equation $U(\hat{\theta}) = 0$. Since $\frac{n}{\hat{\theta}} \neq 0$, we must have:

$$\bar{X} - m(\hat{\theta}) = 0 \implies \bar{X} = m(\hat{\theta})$$

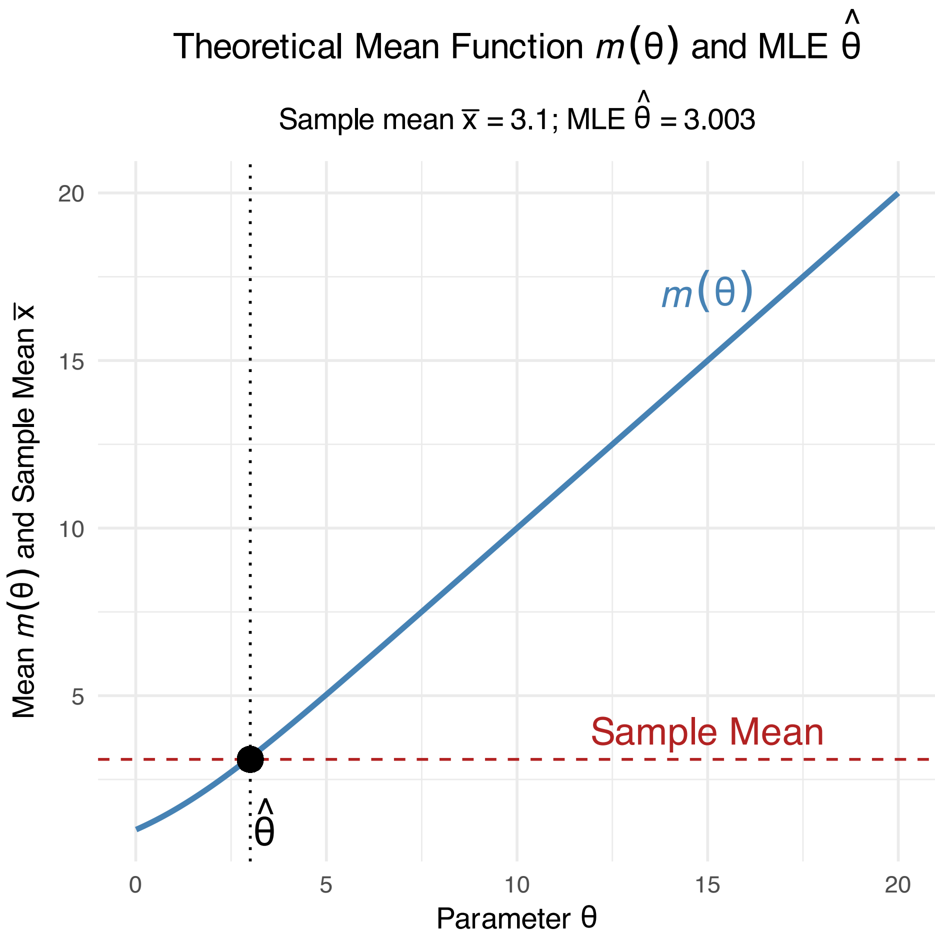
Thus, the MLE $\hat{\theta}$ is the solution of the **mean-matching equation** where the sample mean is equated to the theoretical mean.

(f)

Plot the theoretical mean function $m(\theta)$ over the interval $(0, 20]$ and, on the same graph, add a horizontal line at the sample mean $\bar{x}$ for your simulated data.

Explain how the intersection point corresponds to the MLE. (This shows visually why a unique solution exists and how Fisher scoring is simply inverting $m(\theta)$.)

$$\theta^{(t+1)} \;=\; \theta^{(t)} - \frac{m(\theta^{(t)}) - \bar{x}}{m'(\theta^{(t)})}.$$



(g)

Derive $m'(\theta)$ and show that Fisher scoring is equivalent to Newton's method applied to the root of $m(\theta) - \bar{x} = 0$. Write down the update explicitly:

$$\theta^{(t+1)} = \theta^{(t)} - \frac{m(\theta^{(t)}) - \bar{x}}{m'(\theta^{(t)})}.$$

The theoretical mean function for the zero-truncated Poisson distribution is given by:

$$m(\theta) = \frac{\theta}{1 - e^{-\theta}}$$

To find the derivative $m'(\theta)$, we apply the quotient rule, $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$, where $u = \theta$ and $v = 1 - e^{-\theta}$.

- The derivative of the numerator is $u' = \frac{d}{d\theta}(\theta) = 1$.
- The derivative of the denominator is $v' = \frac{d}{d\theta}(1 - e^{-\theta}) = -(-e^{-\theta}) = e^{-\theta}$.

Applying the quotient rule, we get:

$$\begin{aligned} m'(\theta) &= \frac{(1)(1 - e^{-\theta}) - (\theta)(e^{-\theta})}{(1 - e^{-\theta})^2} \\ &= \frac{1 - e^{-\theta} - \theta e^{-\theta}}{(1 - e^{-\theta})^2} \end{aligned}$$

This is the required derivative of the mean function.

We have already established that the Maximum Likelihood Estimator (MLE) $\hat{\theta}$ is the solution to the mean-matching equation:

$$m(\theta) - \bar{x} = 0$$

Newton's method finds the root of a function $f(\theta)$ using the update rule $\theta^{(t+1)} = \theta^{(t)} - \frac{f(\theta^{(t)})}{f'(\theta^{(t)})}$. Applying this to our mean-matching equation, where $f(\theta) = m(\theta) - \bar{x}$, we have:

- $f(\theta) = m(\theta) - \bar{x}$
- $f'(\theta) = m'(\theta)$

The Newton's method update is therefore:

$$\theta^{(t+1)} = \theta^{(t)} - \frac{m(\theta^{(t)}) - \bar{x}}{m'(\theta^{(t)})}$$

Now, let's consider the Fisher scoring algorithm, which uses the update rule $\theta^{(t+1)} = \theta^{(t)} - \frac{U(\theta^{(t)})}{\mathcal{I}(\theta^{(t)})}$, where $U(\theta)$ is the score function and $\mathcal{I}(\theta)$ is the expected Fisher information.

- The score function is $U(\theta) = \frac{\partial \ell}{\partial \theta} = \frac{n}{\theta}(\bar{x} - m(\theta))$.
- The expected Fisher information is defined as $\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]$. It can be shown (via the identity $\mathcal{I}(\theta) = \frac{n}{\theta} m'(\theta)$) that:

$$\mathcal{I}(\theta) = \frac{n}{\theta}\left(\frac{1 - e^{-\theta} - \theta e^{-\theta}}{(1 - e^{-\theta})^2}\right) = \frac{n}{\theta} m'(\theta)$$

Substituting these expressions into the Fisher scoring update rule:

$$\begin{aligned}\theta^{(t+1)} &= \theta^{(t)} + \frac{U(\theta^{(t)})}{\mathcal{I}(\theta^{(t)})} \\ &= \theta^{(t)} + \frac{\frac{n}{\theta^{(t)}}(\bar{x} - m(\theta^{(t)}))}{\frac{n}{\theta^{(t)}}m'(\theta^{(t)})} \\ &= \theta^{(t)} + \frac{\bar{x} - m(\theta^{(t)})}{m'(\theta^{(t)})} \\ &= \theta^{(t)} - \frac{m(\theta^{(t)}) - \bar{x}}{m'(\theta^{(t)})}\end{aligned}$$

Hence, the two methods are indeed equivalent for finding the MLE.

(h)

Explore the behavior of $m(\theta)$ at the boundaries:

$$\lim_{\theta \to 0^+} m(\theta) = 1, \qquad m(\theta) = \theta + \tfrac{1}{2} + O(e^{-\theta}) \quad \text{as } \theta \to \infty.$$

Explain what these limits tell you about (i) the possible range of the sample mean, and (ii) good starting values for Fisher scoring iterations.

The behavior of the theoretical mean function $m(\theta) = \frac{\theta}{1-e^{-\theta}}$ at its boundaries provides important insights into the properties of the zero-truncated Poisson distribution and the maximum likelihood estimation process.

Possible Range of the Sample Mean The limits of the mean function are given as:

$$\lim_{\theta \to 0^+} m(\theta) = 1 \qquad \text{and} \qquad \lim_{\theta \to \infty} m(\theta) = \infty$$

Since $m(\theta)$ is a continuous and strictly increasing function over the domain $\theta \in (0, \infty)$, its range is the interval $(1, \infty)$.

The Maximum Likelihood Estimator (MLE) $\hat{\theta}$ is defined by the mean-matching equation:

$$\bar{x} = m(\hat{\theta})$$

This implies that the sample mean $\bar{x}$ must be a value within the range of $m(\theta)$. Therefore, a necessary condition for the existence of a unique MLE is that the sample mean must satisfy $\bar{x} > 1$. This makes intuitive sense, as every observation in a zero-truncated Poisson sample is an integer greater than or equal to 1, so their average must also be greater than 1.

Good Starting Values for Fisher Scoring The asymptotic behavior of $m(\theta)$ as $\theta \to \infty$ is given by the approximation:

$$m(\theta) \approx \theta + \tfrac{1}{2}$$

This relationship can be used to derive a simple and effective starting value for the iterative Fisher scoring algorithm. By substituting the mean-matching condition, we get an approximation for the MLE:

$$\bar{x} \approx \hat{\theta} + \tfrac{1}{2}$$

Rearranging this equation provides a good initial guess for the algorithm:

$$\theta^{(0)} = \bar{x} - \tfrac{1}{2}$$

This starting value is generally close to the true MLE, especially for large sample means, which significantly reduces the number of iterations required for the algorithm to converge.

(i)

For several sample sizes ($n = 10, 20, 100$), compare the empirical mean of the simulated sample with the MLE $\hat{\theta}$ and the theoretical mean function $m(\hat{\theta})$.

Verify that the sample mean and the fitted mean coincide, illustrating the “method of moments” nature of the MLE in this model.

The Maximum Likelihood Estimator (MLE) $\hat{\theta}$ for the zero-truncated Poisson distribution is the solution to the **mean-matching equation** $\bar{x} = m(\hat{\theta})$, as derived in part (e). This means that the sample mean ($\bar{x}$) must be identically equal to the fitted theoretical mean ($m(\hat{\theta})$) at the MLE. We verify this property for the three sample sizes used.

Table 1: Comparison of Sample Mean, MLE, and Fitted Mean

n	Sample Mean $\bar{x}$	MLE $\hat{\theta}$	Fitted Mean $m(\hat{\theta})$
10	3.90	3.81396	3.90
20	3.70	3.59878	3.70
100	2.93	2.74100	2.93

The columns for the sample mean ($\bar{x}$) and the fitted mean ($m(\hat{\theta})$) are virtually identical for all three sample sizes ($n = 10, 20, 100$), confirming that:

$$\bar{x} = m(\hat{\theta})$$

This numerical result explicitly verifies the theoretical finding from part (e) that the MLE $hattheta$ for the zero-truncated Poisson distribution is the solution to the **mean-matching equation**. This identity shows that, for this specific model, the MLE is an example of a **Method of Moments estimator**, where a population moment (the mean $m(\theta)$) is equated to the corresponding sample moment (the sample mean $\bar{x}$).