Computing MLE Bias Empirically

Kar Wai Lim
Australian National University

January 31, 2017

Abstract

This note studies the bias arises from the MLE estimate of the rate parameter and the mean parameter of an exponential distribution.

1 Motivation

Although maximum likelihood estimation (MLE) methods provide estimates that are useful, the estimates themselves are not guaranteed to be unbiased. Nevertheless, MLE methods are still highly regarded in practice due to several of their properties, notably, the estimates are consistent and asymptotically normal (Casella and Berger, 2002; Panchenko, 2006).

The most popular example that illustrates the bias of the MLE methods is the MLE estimate of the variance parameter σ^2 of a normal distribution $\mathrm{N}(\mu, \sigma^2)$, we refer the readers to Liang (2012) for details. Another example that is of interest is that of an exponential distribution. In this case, the MLE estimate of the rate parameter λ of an exponential distribution $\mathrm{Exp}(\lambda)$ is biased, however, the MLE estimate for the mean parameter $\mu = 1/\lambda$ is unbiased. Thus, the exponential distribution makes a good case study for understanding the MLE bias.

In this note, we attempt to quantify the bias of the MLE estimates empirically through simulations. For this purpose, we will use the exponential distribution as example.

2 MLE for Exponential Distribution

In this section, we provide a brief derivation of the MLE estimate of the rate parameter λ and the mean parameter μ of an exponential distribution. We note that MLE estimates are values that maximise the likelihood (probability density function) or loglikelihood of the observed data.

Let $\{x_i\}$ be *i.i.d.* random variables that are exponentially distributed, written as

$$x_i \sim \mathrm{Exp}(\lambda)\,. \tag{1}$$

The likelihood function associated with $X = \{x_i\}$ can be written as

$$p(X) = \prod_{i=1}^{N} \lambda\, e^{-\lambda\, x_i}\,, \tag{2}$$

with the following log likelihood:

$$L(\lambda, X) = N \log \lambda - \lambda \sum_{i=1}^{N} x_i\,. \tag{3}$$

Solving for the MLE estimate $\hat{\lambda} = \arg\min_{\lambda} L(\lambda, X)$ gives us

$$\hat{\lambda} = \frac{1}{\bar{x}} \tag{4}$$

where $\bar{x}$ is the mean of $\{x_i\}$:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{5}$$

We note that by using the invariant property of the MLE, the MLE estimate $\hat{\mu}$ is simply

$$\hat{\mu} = \bar{x} \,. \tag{6}$$

2.1 Bias of the MLE Estimates

An estimate is unbiased if the expectation of the estimate equals to its true value.

We first note that the MLE estimate $\hat{\mu}$ is unbiased, evidenced by

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}[\bar{x}] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[x_i] = \mu \,. \tag{7}$$

For the MLE estimate $\hat{\lambda}$, its expectation is

$$\mathbb{E}[\hat{\lambda}] = \mathbb{E}\left[\frac{1}{\bar{x}}\right] \geq \frac{1}{\mathbb{E}[\bar{x}]} = \lambda \,. \tag{8}$$

The inequality follows from Jensen's inequality with the convex function $f(x) = 1/x$.

To quantify the bias of $\hat{\lambda}$, we can derive the bias $\mathcal{B}(\hat{\lambda})$ directly using the properties of gamma distribution. First, we note that the exponential distribution is a special case of the gamma distribution with shape parameter 1, that is,

$$x_i \sim \text{Gamma}(1, \lambda) \,. \tag{9}$$

Since the summation of gamma random variables is also gamma distributed (see Proposition 3.1, Taylor, 2009), we have

$$\sum_{i=1}^{N} x_i \sim \text{Gamma}(N, \lambda) \,. \tag{10}$$

Notice that the inverse of a gamma random variable is inverse-gamma distributed (Wikipedia, 2016):

$$\frac{1}{\sum_{i=1}^{N} x_i} \sim \text{InvGamma}(N, \lambda) \,, \tag{11}$$

with expectation

$$\mathbb{E}\left[\frac{1}{\sum_{i=1}^{N} x_i}\right] = \frac{\lambda}{N-1} \,, \qquad \text{for } N > 1 \,. \tag{12}$$

With this established, we can derive the bias of the MLE estimate $\hat{\lambda}$ as follows:

$$\mathcal{B}(\hat{\lambda}) = \mathbb{E}\left[\frac{1}{\bar{x}}\right] - \lambda = \mathbb{E}\left[\frac{N}{\sum_{i=1}^{N} x_i}\right] - \lambda = \frac{N\,\lambda}{N-1} - \lambda = \frac{1}{N-1}\,\lambda \,. \tag{13}$$

We can see that the bias approaches zero as N increases.

3 Empirical Bias for Exponential Distribution

In this section, we perform experiments to evaluate the bias of the MLE estimates empirically through Monte Carlo method. The empirically bias of an estimate $\hat{\lambda}$ can be computed as

$$\hat{\mathcal{B}}(\hat{\lambda}) = \frac{1}{M}\sum_{j=1}^{M}\hat{\lambda}^{(j)} - \lambda\,, \tag{14}$$

where $\hat{\lambda}^{(j)}$ is the MLE estimate for λ in the j-th simulation experiment. To be precise,

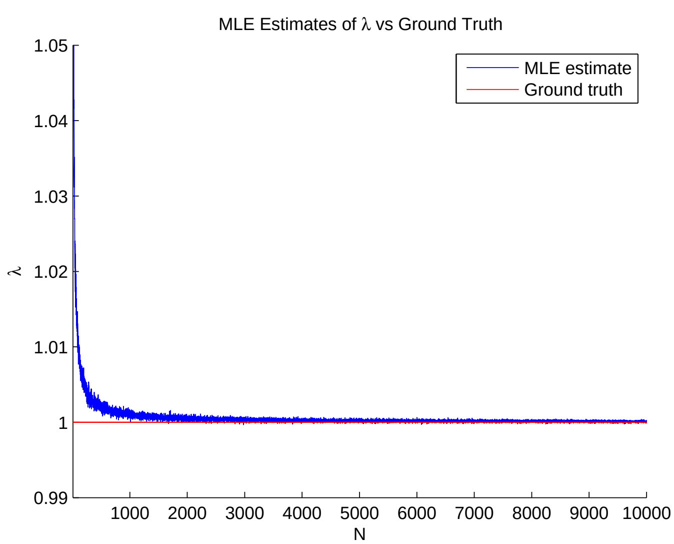
$$\hat{\lambda}^{(j)} = \frac{N}{\sum_{i=1}^{N} x_i^{(j)}}\,, \tag{15}$$

noting that $x_i^{(j)}$ is the i-th sample of the j-th simulation experiment.

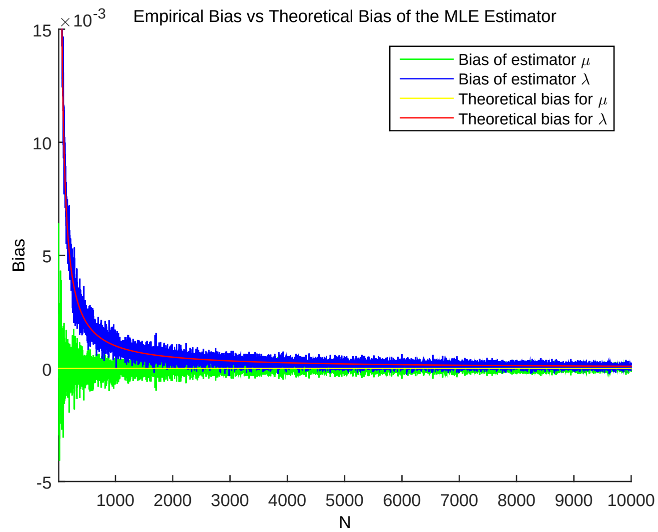
In the following, we compute $\hat{\mathcal{B}}(\hat{\lambda})$ by varying N from 1 to 10000 and setting $M = 10000$. For simplicity, we let $\lambda = \mu = 1$. The experiments are performed using Matlab. Figure below plots the mean of the MLE estimates $\bar{\hat{\lambda}}$ against the number of samples N used in the experiments. The mean $\bar{\hat{\lambda}}$ is computed as follows:

$$\bar{\hat{\lambda}} = \frac{1}{M}\sum_{j=1}^{M}\hat{\lambda}^{(j)}\,, \tag{16}$$

with varying N. From the following plot, we can see that the mean of the MLE estimates deviate from its true value. However, the deviation decreases as the number of samples N increases.



Next, we plot the empirical bias of the MLE estimator $\hat{\lambda}$ (blue) and its theoretical counterpart (red) *vs* the number of samples in the figure below. Here, we can see that the empirical bias fluctuates around its theoretical values. Further, we also superimpose the empirical bias of the MLE estimate $\hat{\mu}$ (green) for illustrative purpose. Since the MLE estimator for μ is unbiased, the empirical bias fluctuates around 0 (yellow).

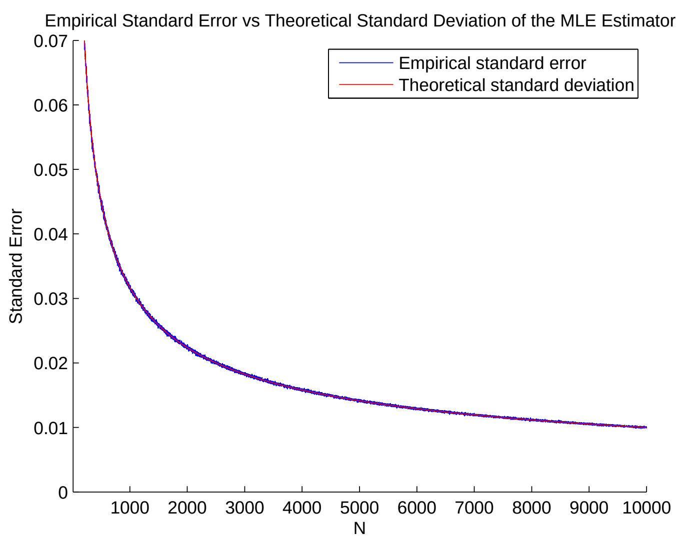


We note that the fluctuations come from two independent sources. Firstly, the fluctuation is due to variability within the MLE estimates, that is, $\mathbb{V}[\hat{\lambda}] \neq 0$. The second source of variability comes from the Monte Carlo simulations, which approaches zero only when $M \to \infty$. To illustrate, in the figure below, we compare the standard errors of $\hat{\mathcal{B}}(\hat{\lambda})$ against their theoretical values, given as

$$\mathbb{V}[\hat{\lambda}] = \mathbb{V}\left[\frac{1}{\bar{x}}\right] = \mathbb{V}\left[\frac{N}{\sum_{i=1}^{N} x_i}\right] = \frac{N^2\,\lambda^2}{(N-1)^2(N-2)}\,, \qquad \text{for } N > 2\,. \tag{17}$$

This corresponds to the first source of variability. The variability from the Monte Carlo simulation is in fact smaller, which is given by

$$\mathbb{V}[\hat{\mathcal{B}}(\hat{\lambda})] = \frac{1}{M}\mathbb{V}[\hat{\lambda}] = \frac{N^2\,\lambda^2}{M(N-1)^2(N-2)}\,, \qquad \text{for } N > 2\,. \tag{18}$$



We note that the standard errors computed for the estimator $\hat{\mu}$ have similar values as those of $\hat{\lambda}$ (thus not plotted). This is because their variances are very close to one another (due to $\lambda = \mu$) for large N:

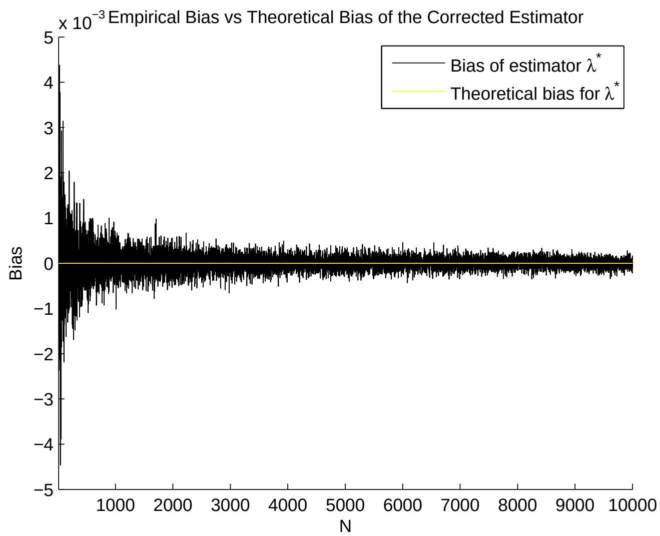
$$\mathbb{V}[\hat{\mu}] = \mathbb{V}[\bar{x}] = \mathbb{V}\left[\frac{\sum_{i=1}^{N} x_i}{N}\right] = \frac{1}{N\lambda^2} = \frac{\mu^2}{N}\,. \tag{19}$$

3.1 Bias Correction

To recap, the expected value of the MLE estimator $\hat{\lambda}$ is $\frac{N}{N-1}\,\lambda$. We can simply multiply a correction factor $\frac{N-1}{N}$ to the MLE estimator to eliminate the bias, this gives us an unbiased estimator

$$\hat{\lambda}^* = \frac{N-1}{N\,\bar{x}} = \frac{N-1}{\sum_{i=1}^{N} x_i}\,. \tag{20}$$

In the following graph, we compute the empirical bias of the adjusted estimator.



As expected, the MLE estimates centred around 0, verifying that the estimator $\hat{\lambda}^*$ is unbiased. We note that the variance of the estimator is

$$\mathbb{V}[\hat{\lambda}^*] = \frac{(N-1)^2}{N^2}\mathbb{V}[\hat{\lambda}] = \frac{\lambda^2}{(N-2)}\,, \qquad \text{for } N > 2\,, \tag{21}$$

which is slightly lower than that of the MLE estimator $\hat{\lambda}$, this corrected estimator is thus better in terms of both bias and variability.

References

Casella, G. and Berger, R. L. (2002). *Statistical Inference*, volume 2. Duxbury Pacific Grove, CA.

Liang, D. (2012). Maximum likelihood estimator for variance is biased: Proof. Retrieved from http://dawenl.github.io/files/mle_biased.pdf.

Panchenko, D. (2006). Lecture 3 Properties of MLE: consistency, asymptotic normality. Fisher information. Retrieved from http://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture3.pdf.

Taylor, J. (2009). Lecture 20: Sums of independent random variables. Retrieved from https://math.la.asu.edu/~jtaylor/teaching/Fall2010/STP421/lectures/lecture20.pdf.

Wikipedia (2016). Inverse-gamma distribution. Retrieved from https://en.wikipedia.org/wiki/Inverse-gamma_distribution.