Maximum Likelihood Estimator for Variance is Biased: Proof

Dawen Liang
Carnegie Mellon University
dawenl@andrew.cmu.edu

1   Introduction

Maximum Likelihood Estimation (MLE) is a method of estimating the parameters of a statistical model. It is widely used in Machine Learning algorithm, as it is intuitive and easy to form given the data. The basic idea underlying MLE is to represent the likelihood over the data w.r.t the model parameters, then find the values of the parameters so that the likelihood is maximized.

For example, given N 1-dimensional data points x_i, where $i = 1, 2, \cdots, N$ and we assume the data points are drawn i.i.d. from a Gaussian distribution. Then we could estimate the mean μ and variance σ^2 of the true distribution via MLE. Per definition, $\mu = \mathbb{E}[x]$ and $\sigma^2 = \mathbb{E}[(x - \mu)^2]$. Thus, intuitively, the mean estimator $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$ and the variance estimator $s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2$ follow. It is easy to check that these estimators are derived from MLE setting. See Chapter 2.3.4 of Bishop (2006).

2   Biased/Unbiased Estimation

In statistics, we evaluate the "goodness" of the estimation by checking if the estimation is "unbiased". By saying "unbiased", it means the expectation of the estimator equals to the true value, e.g. if $\mathbb{E}[\overline{x}] = \mu$ then the mean estimator is unbiased. Now we will show that the equation actually holds for mean estimator.

$$\mathbb{E}[\overline{x}] = \mathbb{E}[\frac{1}{N} \sum_{i=1}^{N} x_i] = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[x]$$
$$= \frac{1}{N} \cdot N \cdot \mathbb{E}[x]$$
$$= \mathbb{E}[x] = \mu$$

The first line makes use of the assumption that the samples are drawn i.i.d from the true distribution, thus $\mathbb{E}[x_i]$ is actually $\mathbb{E}[x]$. From the proof above, it is shown that the mean estimator is unbiased.

Now we move to the variance estimator. At the first glance, the variance estimator $s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2$ should follow because mean estimator $\overline{x}$ is unbiased. However, it is not the

case:

$$\mathbb{E}[s^2] = \mathbb{E}[\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2]$$

$$= \frac{1}{N} \mathbb{E}[\sum_{i=1}^{N} x_i^2 - 2 \sum_{i=1}^{N} x_i \bar{x} + \sum_{i=1}^{N} \bar{x}^2]$$

We know $\sum_{i=1}^{N} x_i = N \cdot \bar{x}$ and $\sum_{i=1}^{N} \bar{x}^2 = N \cdot \bar{x}^2$. Plug these into the derivation:

$$\mathbb{E}[s^2] = \frac{1}{N} \mathbb{E}[\sum_{i=1}^{N} x_i^2 - 2N \cdot \bar{x}^2 + N \cdot \bar{x}^2]$$

$$= \frac{1}{N} \mathbb{E}[\sum_{i=1}^{N} x_i^2 - N \cdot \bar{x}^2]$$

$$= \frac{1}{N} \mathbb{E}[\sum_{i=1}^{N} x_i^2] - \mathbb{E}[\bar{x}^2]$$

$$= \mathbb{E}[x^2] - \mathbb{E}[\bar{x}^2]$$

According to the alternative definition of variance, $\sigma_x^2 = \mathbb{E}[x^2] - \mathbb{E}[x]^2$ and similarly, $\sigma_{\bar{x}}^2 = \mathbb{E}[\bar{x}^2] - \mathbb{E}[\bar{x}]^2$, where the random variable is $\bar{x}$. Note that $\mathbb{E}[x] = \mathbb{E}[\bar{x}] = \mu$. Plug the 2 equations to the derivation:

$$\mathbb{E}[s^2] = (\sigma_x^2 + \mu^2) - (\sigma_{\bar{x}}^2 + \mu^2)$$

$$= \sigma_x^2 - \sigma_{\bar{x}}^2$$

$$\sigma_{\bar{x}}^2 = \text{VAR}[\bar{x}] = \text{VAR}[\frac{1}{N} \sum_{i=1}^{N} x_i] = \frac{1}{N^2} \text{VAR}[\sum_{i=1}^{N} x_i]$$

Since the samples are drawn i.i.d.

$$\text{VAR}[\sum_{i=1}^{N} x_i] = \sum_{i=1}^{N} \text{VAR}[x] = N \cdot \text{VAR}[x]$$

Thus,

$$\sigma_{\bar{x}}^2 = \frac{1}{N} \text{VAR}[x] = \frac{1}{N} \sigma_x^2$$

Plug back to the $\mathbb{E}[s^2]$ derivation,

$$\mathbb{E}[s^2] = \frac{N-1}{N} \sigma_x^2$$

Therefore, $\mathbb{E}[s^2] \neq \sigma_x^2$ and it is shown that we tend to underestimate the variance. In order to overcome this biased problem, the maximum likelihood estimator for variance can be slightly modified to take this into account:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

It is easy to show that this modified variance estimator is unbiased.

References

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics).* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.