On the Bias and Variance of the Sample Estimators

Phil Schniter

January 28, 2019

1 Sample Estimators

Say we want to estimate the mean and variance of a random variable x given the N realizations $\boldsymbol{x} = [x_1, \ldots, x_N]^\top$. The standard way to estimate the mean is using the "sample mean,"

$$\overline{x} \triangleq \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{1}$$

The standard way to estimate the variance is using the "sample variance," but there are two common definitions for it:

$$s_x^2 \triangleq \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{2}$$

$$\xi_x^2 \triangleq \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2, \tag{3}$$

It is said that s_x^2 is "biased" and ξ_x^2 is "unbiased." We clarify these claims below.

2 Bias of the Sample Estimators

Suppose that x has the true mean $\mu = \mathrm{E}\{x\}$ and the true variance $v = \mathrm{E}\{(x-\mu)^2\}$. Then a mean estimate $\widehat{\mu}(\boldsymbol{x})$ and a variance estimate $\widehat{v}(\boldsymbol{x})$ are said to be *unbiased* if

$$\mathrm{E}\{\widehat{\mu}(\boldsymbol{x})\} = \mu \quad \text{and} \quad \mathrm{E}\{\widehat{v}(\boldsymbol{x})\} = v. \tag{4}$$

Note that, in (4), we treat $\{x_i\}$ as random variables, so that $\widehat{\mu}(\boldsymbol{x})$ and $\widehat{v}(\boldsymbol{x})$ are also random. (In contrast, μ and v are deterministic.) In particular, we treat $\{x_i\}$ as independent and identically distributed (i.i.d.) random variables with the same distribution as x, implying that

$$\mathrm{E}\{x_i\} = \mu, \ \forall i \tag{5}$$

$$\mathrm{E}\{x_i^2\} = \mu^2 + v, \ \forall i \tag{6}$$

$$\mathrm{E}\{x_i x_j\} = \mu^2 + v\delta_{i-j} \quad \text{where} \quad \delta_{i-j} = \begin{cases} 1 & i = j \\ 0 & i \neq j, \end{cases} \tag{7}$$

which will be used below.

Let us first investigate whether $\overline{x}(\boldsymbol{x})$ is an unbiased estimator of the mean μ. To do this, we take its expectation (over $\boldsymbol{x}$):

$$\begin{aligned}
\mathrm{E}\{\overline{x}\} &= \mathrm{E}\left\{\frac{1}{N}\sum_{i=1}^{N} x_i\right\} && (8)\\
&= \frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\{x_i\} && \text{by linearity of expectation} \quad (9)\\
&= \frac{1}{N}\sum_{i=1}^{N}\mu && \text{since } x_i \text{ are i.i.d. with mean } \mu \quad (10)\\
&= \mu. && (11)
\end{aligned}$$

Thus we see that $\overline{x}$ is an unbiased estimator of μ.

Next we investigate whether $s_x^2(\boldsymbol{x})$ or $\xi_x^2(\boldsymbol{x})$ are unbiased estimators of the variance v. Both estimators are scalings of $\sum_{i=1}^{N}(x_i - \overline{x})^2$, and so we analyze that quantity first.

$$\begin{aligned}
\mathrm{E}\left\{\sum_{i=1}^{N}(x_i-\overline{x})^2\right\} &= \sum_{i=1}^{N}\mathrm{E}\{(x_i-\overline{x})^2\} && \text{by linearity of expectation} \quad (12)\\
&= \sum_{i=1}^{N}\mathrm{E}\{x_i^2 - 2x_i\overline{x} + \overline{x}^2\} && (13)\\
&= \sum_{i=1}^{N}\Big[\mathrm{E}\{x_i^2\} - 2\,\mathrm{E}\{x_i\overline{x}\} + \mathrm{E}\{\overline{x}^2\}\Big] && \text{by linearity of expectation.} \quad (14)
\end{aligned}$$

As for the three terms in the sum, (5)-(7) imply

$$\begin{aligned}
\mathrm{E}\{x_i^2\} &= \mu^2 + v,\ \forall i && (15)\\
2\,\mathrm{E}\{x_i\overline{x}\} &= 2\,\mathrm{E}\left\{x_i\frac{1}{N}\sum_j x_j\right\} = \frac{2}{N}\sum_j \mathrm{E}\{x_i x_j\} = \frac{2}{N}\sum_j[\mu^2 + v\delta_{i-j}] = 2\mu^2 + \frac{2v}{N}, \forall i && (16)\\
\mathrm{E}\{\overline{x}^2\} &= \mathrm{E}\left\{\left(\frac{1}{N}\sum_j x_j\right)^2\right\} = \frac{1}{N^2}\,\mathrm{E}\left\{\sum_{i,j} x_i x_j\right\} = \frac{1}{N^2}\sum_{i,j}[\mu^2 + v\delta_{i-j}] && (17)\\
&= \mu^2 + \frac{v}{N^2}\sum_{i,j}\delta_{i-j} = \mu^2 + \frac{v}{N} && (18)
\end{aligned}$$

Putting these together, we find

$$\begin{aligned}
\mathrm{E}\left\{\sum_{i=1}^{N}(x_i-\overline{x})^2\right\} &= \sum_{i=1}^{N}\big(\mu^2 + v - 2\mu^2 - \frac{2v}{N} + \mu^2 + \frac{v}{N}\big) && (19)\\
&= \sum_{i=1}^{N}\big(v - \frac{v}{N}\big) = N\big(v - \frac{v}{N}\big) = (N-1)v. && (20)
\end{aligned}$$

Thus, the bias of our two variance estimators becomes clear:

$$\mathrm{E}\{s_x^2\} = \frac{1}{N}\,\mathrm{E}\left\{\sum_{i=1}^{N}(x_i - \overline{x})^2\right\} = \frac{N-1}{N}v \qquad \text{biased estimate of } v! \tag{21}$$

$$\mathrm{E}\{\xi_x^2\} = \frac{1}{N-1}\,\mathrm{E}\left\{\sum_{i=1}^{N}(x_i - \overline{x})^2\right\} = v \qquad \text{unbiased estimate of } v! \tag{22}$$

3 Variance of the Sample Estimators

In addition to the bias on the sample estimators, we may also be interested in the variance. We first investigate the variance of the sample-mean estimator.

$$\begin{aligned}
\mathrm{var}\{\overline{x}\} &= \mathrm{E}\left\{(\overline{x} - \mathrm{E}\{\overline{x}\})^2\right\} && \text{by definition} && (23)\\
&= \mathrm{E}\left\{(\overline{x} - \mu)^2\right\} && \text{from results above} && (24)\\
&= \mathrm{E}\{\overline{x}^2 - 2\mu\overline{x} + \mu^2\} && && (25)\\
&= \mathrm{E}\{\overline{x}^2\} - 2\mu\,\mathrm{E}\{\overline{x}\} + \mu^2 && \text{by linearity of expectation} && (26)\\
&= \mu^2 + \frac{v}{N} - 2\mu^2 + \mu^2 && \text{from results above} && (27)\\
&= \frac{v}{N}. && && (28)
\end{aligned}$$

The variances of s_x^2 and ξ_x^2 are more difficult to analyze and depend on the particular distribution of x, not merely its mean μ and variance v. For the case where $x \sim \mathcal{N}(\mu, v)$, it is not too difficult to show that

$$\mathrm{var}\{\xi_x^2\} = \frac{2v^2}{N-1} \tag{29}$$

$$\mathrm{var}\{s_x^2\} = \mathrm{var}\{\frac{N-1}{N}\xi_x^2\} = \frac{2v^2}{N}. \tag{30}$$

4 Mean-squared error of the Sample Estimators

From the above bias and variance equations, we can compute the mean-squared errors of the sample estimators.

$$\begin{aligned}
\mathrm{E}\{(\overline{x} - \mu)^2\} &= \mathrm{E}\{\overline{x} - \mu\}^2 + \mathrm{var}\{\overline{x}\} && (31)\\
&= (\mathrm{E}\{\overline{x}\} - \mu)^2 + \mathrm{var}\{\overline{x}\} && (32)\\
&= (\mu - \mu)^2 + \frac{v}{N} && (33)\\
&= \frac{v}{N} && (34)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\{(s_x^2 - v)^2\} &= \mathrm{E}\{s_x^2 - v\}^2 + \mathrm{var}\{s_x^2\} && (35)\\
&= (\mathrm{E}\{s_x^2\} - v)^2 + \mathrm{var}\{s_x^2\} && (36)\\
&= \left(\frac{N-1}{N}v - v\right)^2 + \frac{2v^2}{N} && (37)\\
&= \left(\frac{1}{N^2} + \frac{2}{N}\right)v^2 = \frac{2N+1}{N^2}v^2 && (38)
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}\{(\xi_x^2 - v)^2\} &= \mathrm{E}\{\xi_x^2 - v\}^2 + \mathrm{var}\{\xi_x^2\} && (39)\\
&= (\mathrm{E}\{\xi_x^2\} - v)^2 + \mathrm{var}\{\xi_x^2\} && (40)\\
&= (v - v)^2 + \frac{2v^2}{N-1} && (41)\\
&= \frac{2v^2}{N-1} && (42)
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathrm{E}\{(s_x^2 - v)^2\} &= \frac{2N+1}{N^2} v^2 && (43)\\
&< \frac{2N+2}{N^2} v^2 && (44)\\
&< \frac{N^2}{N^2-1} \frac{2N+2}{N^2} v^2 && (45)\\
&= \frac{2}{N-1} v^2 = \mathrm{E}\{(\xi_x^2 - v)^2\}, && (46)
\end{aligned}
$$

and so s_x^2 actually has a lower MSE than ξ_x^2 even though s_x^2 is biased.