Bias and Variance of Estimators

Recall that a statistic is a quantity that may be calculated from sample data. For instance, the sample mean and standard deviation , regression coefficient estimates , and quantities like , , , and even are all sample statistics.

These statistics vary from one sample to another, and if we use different samples (e.g., different *training* data sets), we are likely to see different values of these statistics. Because these statistics can take on many values across the different samples, they each have a distribution, a mean (i.e.,expected value) and variance of its own.

For instance, we’ve seen that under certain conditions, the sample mean of variable , which we denote as , will have the expected value of , where is the *true, actual value of the population mean of X*. Because the expected value of the statistic equals to the true value of the population parameter , we can say that is an unbiased estimator of .

Similarly, when we run OLS regression, it turns out that the we get with least squares estimation are unbiased estimators of the true beta coefficients in the population, .That is, . Likewise, with OLS, is an unbiased estimator of .

Specifically, the bias of an estimator is the difference between the expected value of the estimator and the true (population) value of the parameter being estimated. In the examples above, the bias is 0.

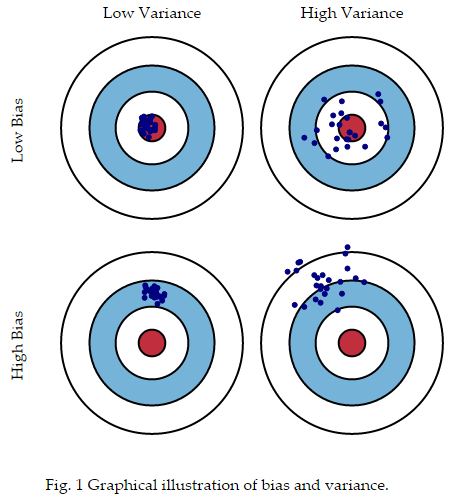
What might be some examples of biased estimators? Recall that when we calculate the sample standard deviation of some variable , we use the formula – that is, we divide the sum of the squared deviations from the mean not by the number of observations , but rather, by . The reason for this is that dividing by in the formula for results in an unbiased estimator of the true population standard deviation – that is, ; on the other hand, dividing by results in a biased estimator of – that is, we are consistently *underestimating* the true standard deviation when we divide by . See the accompanying Excel file for a demonstration of this.

An illustration of this may be seen [here](https://www.khanacademy.org/computer-programming/unbiased-variance-visualization/1167453164) and a video explaining the simulation may be seen [here](https://www.khanacademy.org/math/probability/descriptive-statistics/variance_std_deviation/v/another-simulation-giving-evidence-that-n-1-gives-us-an-unbiased-estimate-of-variance). Another illustration is [here](https://www.youtube.com/watch?v=vEZPbrqrt2M).

The variance of the estimator itself is also of interest. That is, ideally, we want an estimator which yields values that don’t vary too much from sample to sample. For instance, if the variable *X* has a symmetric distribution (e.g., normal, uniform, -distribution, etc), the sample mean equals the sample median , and both are unbiased estimators of , the true average value of . However, it can be shown that the sample mean has a lower variance\* than the sample median , and is usually the preferred estimator of , because traditionally, statisticians prefer to use the unbiased estimator with the lowest variance.

\*That is, there’s much less sample-to-sample variability in the values of the sample mean than in the values of the sample median.

Bias and variance of estimators may be better understood with the aid of the following diagram, where each point represents the value of a statistic (e.g., coefficient) resulting from different samples that come from the same population. The true value of the population parameter that the statistic is trying to estimate is the bullseye section of the target.



Source: <http://scott.fortmann-roe.com/docs/BiasVariance.html>

Going back to regression analysis, OLS seems to be a good model because it yields unbiased estimates of with the smallest variance. However, even though OLS estimators are unbiased, they still have higher variance than some biased estimators!

On the other hand, Ridge (and Lasso) regression predictions are biased, but have lower variance than OLS predictions for certain values of ! Because the (test data) MSE depends on both the bias and variance components, ridge regression will yield lower MSE compared to OLS regression *if* the relative increase in bias is smaller than the decrease in variance!