We now look at methods for evaluating the quality of estimators. Ultimately, we will show that MLEs have many desirable properties (even though they may be biased).

The following refers to 7.3.1, page 330–331.

To be able to compare MLEs with other estimators, we need some general measures of quality that we may apply to many estimators. Here is possibly the most useful one:

Definition 7.3.1. Suppose we have a sample $X_1, \ldots, X_n$ from a $f_X(x|\theta)$ (where we imagine θ to be unknown). Let $W = W(X_1, \ldots, X_n)$ be a statistic (which we want to use as an estimator for θ). Then, the mean squared error (MSE) of W is $\mathrm{E}_\theta(W - \theta)^2$.

The subscript on the E reminds us that the expectation is dependent on the value of the parameter. The MSE can be viewed as a (number-valued) function of the parameter. (When calculating the MSE, always use the same value of θ in both places.) Note that

$$\mathrm{E}_\theta(W-\theta)^2 = \mathrm{Var}_\theta W + (\mathrm{E}_\theta W - \theta)^2. \tag{$*$}$$

The quantity $\mathrm{E}_\theta - \theta$ is called the *bias of W as an estimate of θ.*

$$\mathrm{Bias}_\theta W := \mathrm{E}_\theta W - \theta.$$

This is also a number-valued function of θ, and

$$\mathrm{MSE}_\theta W = \mathrm{Var}_\theta W + (\mathrm{Bias}_\theta W)^2.$$

Example. Let X be a random variable with $\mathrm{E}X = \mu$ and $\mathrm{Var}X = \sigma^2$. Let $X_1, \ldots, X_n$ be iid X.

In Theorem 5.2.6, we showed that the sample mean $\overline{X}$ is an unbiased estimator of μ, i.e., $\mathrm{E}\overline{X} = \mu$. We can easily calculate the MSE. From ($*$), it is $\mathrm{Var}\overline{X}$, and Theorem 5.2.6 tells us that this is $\frac{\sigma^2}{n}$. If X is normal (or binomial or Poisson or any of numerous distributions) the sample mean is also the MLE of μ.

We also showed in 5.2.6 that the sample variance S^2 is an unbiased estimator of σ^2, i.e., $\mathrm{E}S^2 = \sigma^2$. We have seen that when X is normal, S^2 is *not* the MLE; the MLE is $\frac{n-1}{n}S^2$. *So, which of these two estimators has smaller MSE?*

To answer this, we need to know $\mathrm{Var}S^2$. Since we are assuming that X is normal, $\frac{n-1}{\sigma^2}S^2$ is χ^2_{n-1}. Now, $\mathrm{Var}\chi^2_{n-1} = 2(n-1)$ (see pages 100-101). Thus:

$$2(n-1) = \mathrm{Var}\frac{n-1}{\sigma^2}S^2 = \frac{(n-1)^2}{\sigma^4}\mathrm{Var}S^2,$$

so

$$\mathrm{Var}S^2 = \frac{2\sigma^4}{n-1}.$$

(See the rest of this example worked out on page 331.)

The following refers to 7.3.2, page 334–339.

If two estimators W_1 and W_2 have the same bias, then the difference of their MSEs is

$$E_\theta(W_1 - \theta) - E_\theta(W_2 - \theta) = \mathrm{Var}_\theta W_1 - \mathrm{Var}_\theta W_2.$$

Thus, within the class of estimators with the same bias, estimators with smaller variance have smaller MSE, and if there is an estimator with least variance, it has the least MSE. (Of course, unbiased estimators get most attention.)

(See quiz from Friday for an example involving the Poisson distribution.)

This raises the question of whether there *is* an estimator with least variance? First, we address an easier question: is there a lower bound for the variance?

See **Theorem 7.3.9 (Cramér-Rao Inequality)** Assuming $f_{\vec{X}}(\vec{x}|\theta)$ is "well-behaved,"

$$\mathrm{Var}_\theta W(\vec{X}) \geq \frac{\left(\frac{d}{d\theta}\mathrm{E}_\theta W(\vec{X})\right)^2}{\mathrm{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\vec{X}|\theta)\right)^2\right)}$$

We will examine the details next time.