Basic Concepts of Inference

Statistical Inference is the process of making conclusions using data that is subject to random variation.

Here are some basic definitions.

• $\text{Bias}(\hat{\theta}) := \mathbf{E}(\hat{\theta}) - \theta$, where θ is the true parameter value and $\hat{\theta}$ is an estimate of it computed from data.

An estimator whose bias is 0 is called *unbiased*. Contrast bias with:

• $\text{Var}(\hat{\theta}) = \mathbf{E}(\hat{\theta} - \mathbf{E}(\hat{\theta}))^2$. Variance measures "precision" or "reliability".

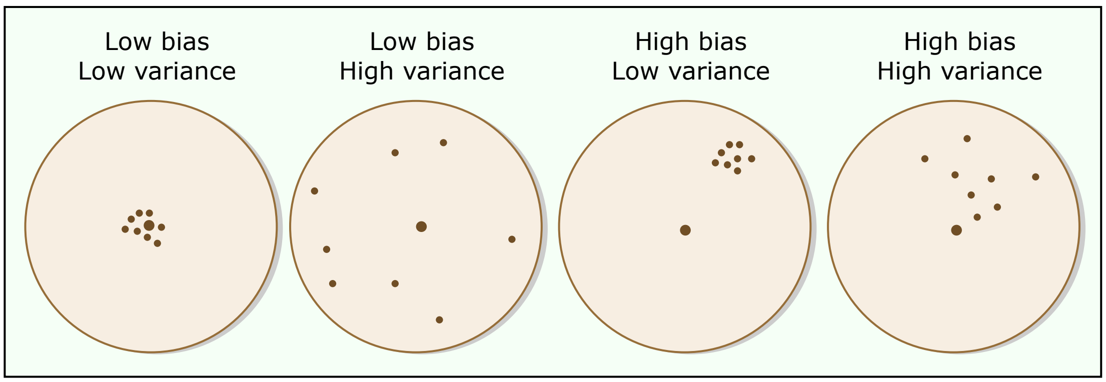


Image by MIT OpenCourseWare.

• Mean-Squared Error (MSE) - a way to measure the goodness of an estimator.

$$\begin{aligned} MSE(\hat{\theta}) &= \mathbf{E}(\hat{\theta} - \theta)^2 \\ &= \mathbf{E}[\hat{\theta} - \mathbf{E}(\hat{\theta}) + \mathbf{E}(\hat{\theta}) - \theta]^2 \\ &= \mathbf{E}[\hat{\theta} - \mathbf{E}(\hat{\theta})]^2 + \mathbf{E}[\mathbf{E}(\hat{\theta}) - \theta]^2 + 2\mathbf{E}\left([\hat{\theta} - \mathbf{E}(\hat{\theta})][\mathbf{E}(\hat{\theta}) - \theta]\right) \end{aligned}$$

The first term is $\text{Var}(\hat{\theta})$. In the second term, the outer expectation does nothing because the inside is a constant. The second term is just the bias squared. In the third term, the part $\mathbf{E}(\hat{\theta}) - \theta$ is a constant, so we can pull it out of the expectation. But then what's left inside the expectation is $\mathbf{E}[\hat{\theta} - \mathbf{E}(\hat{\theta})]$ which is zero, so the third term is zero.

$$MSE(\hat{\theta}) = \left(\text{Bias}(\hat{\theta})\right)^2 + \text{Var}(\hat{\theta}). \tag{1}$$

Perhaps you have heard of the "Bias-Variance" tradeoff. This has to do with statistical modeling and will be discussed when you hear about regression. It boils down to a tradeoff in how you create a statistical model. If you try to create a low bias model, you risk that your model might not explain the data well and have a high variance and thus a larger MSE. If you try to create a low variance model, it may do so at the expense of a larger bias and then still a larger MSE.

- We will now show why we use:

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 \text{ rather than } s^2_{\text{wrong}} = \frac{1}{n}\sum_i (x_i - \bar{x})^2.$$

The answer is that s^2 is an unbiased estimator for σ^2!

Let's show this. We have to calculate $\text{Bias}(S^2) = \mathbf{E}(S^2) - \sigma^2$ which means we need $\mathbf{E}(S^2)$. Remember that S^2 follows a (scaled) chi-square distribution, and if we go back and look in the notes for the chi-square distribution, we'll find that the expectation for S^2 is σ^2. (It's one of the last equations in the chi-square notes). So, $\text{Bias}(S^2) = \sigma^2 - \sigma^2 = 0$. This is why we use $n-1$ in the denominator of S^2.

However, it turns out that the mean square error is worse when we use $n-1$ in the denominator: $MSE(S^2) > MSE(S^2_{\text{wrong}})$.

Let's show this. Again going back to the notes on the chi-square distribution, we find that:

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

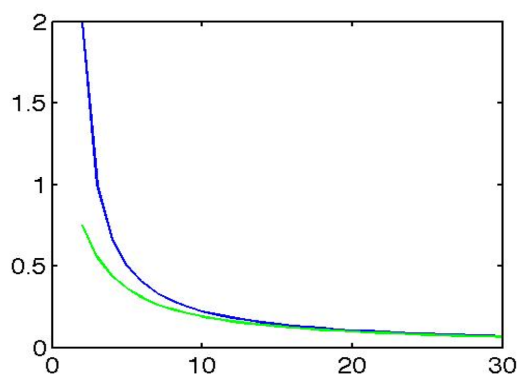
Plugging this in to equation (1) using S^2 as the estimator $\hat{\theta}$, we find:

$$MSE(S^2) = \text{Var}(S^2) + \big(\text{Bias}(S^2)\big)^2 = \frac{2\sigma^4}{n-1} + 0,$$

whereas

$$MSE(S^2_{\text{wrong}}) = \text{(skipping steps here)} = \frac{2n-1}{n^2}\sigma^4.$$

And if you plot those two on the same plot, you'll see that $MSE(S^2)$ is bigger than $MSE(S^2_{\text{wrong}})$.



$MSE(S^2)$ (top) and $MSE(S^2_{\text{wrong}})$ (bottom) versus n for $\sigma^2 = 1$.

So using S^2 rather than S^2_{wrong} actually hurts the mean squared error, but not by much and actually the difference between the two shrinks as n gets large.

- The standard deviation of $\hat{\theta}$ is called the standard error.

$$SE(\bar{x}) = s/\sqrt{n}$$

is the estimated standard error of the mean for for independent r.v. - this appears a lot.

15.075J / ESD.07J Statistical Thinking and Data Analysis
Fall 2011