Mean-Squared Error, Bias, and Variance

Prof. Dennis Sun

Data 401
Where We Left Off

We determined which estimator was best by choosing the estimator $\hat{\lambda}$ whose sampling distribution “looks” most tightly concentrated around the true value of $\lambda$.

Can we make this procedure more quantitative?
We simulate $\hat{\lambda}_1, ..., \hat{\lambda}_B$. We can calculate

$$\frac{1}{B} \sum_{i=1}^{B} (\hat{\lambda}_i - \lambda)^2.$$ 

As $B \to \infty$, this converges to $\text{E}[(\hat{\lambda} - \lambda)^2]$. This is called the **mean-squared error** (MSE) of the estimator $\hat{\lambda}$.

For the three estimators, the MSEs when $\lambda = 0.5$ are:

$$0.44, 0.24, 1.31$$

So **Estimator 2** is the best.
Bias and Variance

- The bias of an estimator $\hat{\theta}$ tells us how much its expected value deviates from $\theta$:
  \[
  \text{Bias}[\hat{\theta}] = E[\hat{\theta}] - \theta.
  \]

- The variance of an estimator $\hat{\theta}$ tells us how much it deviates from its expected value (on average).
  \[
  \text{Var}[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2].
  \]
Bias-Variance Decomposition

For any estimator, the MSE of the estimator can be decomposed into two components:

\[
\text{MSE}[\hat{\theta}] = \text{Bias}[\hat{\theta}]^2 + \text{Var}[\hat{\theta}].
\]
Trading Bias for Variance?

Is it possible for a biased estimator to be better than an unbiased one?

Yes, but only if the reduction in variance is more than the increase in bias. In the example below, the blue estimator has the smaller MSE, even though it is biased and the orange estimator is not.