Evaluating Methods: Consistency

Prof. Dennis Sun

Data 401

Statistical Framework for Evaluating Methods

Given data, we assume a probability model for the data and estimate the parameters in that model.

\[
\theta_0 \xrightarrow{p_{\theta_0}(x)} \text{Data} \xrightarrow{L(\theta) = p_{\theta}(x)} \hat{\theta}
\]

How do we know if the estimate is any good or not?

Statisticians assume that there is some true underlying model that generated the data and study how \( \hat{\theta} \) compares to \( \theta_0 \).

Notice that there are 3 “\( \theta \)”s floating around:

- \( \theta_0 \) is the true value of the parameter
- \( \theta \) is the variable that we optimize over to get the MLE
- \( \hat{\theta} \) is the estimated value of the parameter
The Problem with Bias

One metric that we can calculate is the bias:

\[ \text{bias} = E[\hat{\theta}] - \theta_0. \]

What’s the problem with bias?

- The bias looks at how far \( \hat{\theta} \) is from \( \theta_0 \), on average. Averages do not take into account variability.
- **Basu’s elephant**: A circus owner plans to ship 50 elephants and needs a rough estimate of their total weight. He decides to weigh one randomly selected elephant.

One elephant, Sambo, is given a \( 99/100 \) probability of being selected. The “only” unbiased estimator of the total weight is:

- If Sambo is selected, we multiply its weight by \( 100/99 \).
- If one of the other elephants is selected, we multiply its weight by \( 4900 \).

This is clearly a ridiculous estimate of the total weight of 50 elephants, but it is unbiased.

Another Criterion: Consistency

Another desirable property of a method is consistency. As we get more and more data (as \( n \to \infty \)),

\[ \hat{\theta}_n \to \theta_0 \]

This seems odd. \( \hat{\theta}_n \) is random, since it depends on the random data. But \( \theta_0 \) is fixed. How can something random converge to something fixed?

Consider the MLE of the Poisson, \( \hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \).

We see that \( \hat{\mu}_n \to \mu_0 \) in this particular case.
Law of Large Numbers

The Law of Large Numbers says that if $X_1, \ldots, X_n$ are i.i.d. from any distribution with mean $\mu \overset{\text{def}}{=} \mathbb{E}[X_1]$, then the sample mean $\overline{X}$ is consistent for $\mu$.

$$\overline{X} \overset{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} X_i \to \mu.$$ 

The proof essentially follows from two observations:

1. $\mathbb{E}[\overline{X}] = \mu$
2. $\text{SD} [\overline{X}] = \frac{\text{SD}[X_i]}{\sqrt{n}} \to 0$ as $n \to \infty$.

But what about other estimators that are not the sample mean?

The MLE is Consistent

Remember that $\hat{\theta}_n \overset{\text{def}}{=} \arg\max_{\theta} \log p_{\theta}(X_1, \ldots, X_n)$.

If $X_1, \ldots, X_n$ are i.i.d., then:

$$\hat{\theta}_n = \arg\max_{\theta} \sum_{i=1}^{n} \log p_{\theta}(X_i)$$