APMA 1650 Consistency

Rebecca Santorella

December 2, 2018

To understand consistency, we first need to understand convergence in probability.

Definition 1.1. θ_n *converges to* θ *in probability* if, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\theta_n - \theta| \leq \epsilon) = 1,$$

or, equivalently,

$$\lim_{n\to\infty} P(|\theta_n - \theta| > \epsilon) = 0.$$

We denote this by $\theta_n \xrightarrow{P} \theta$.

Sometimes we have estimators that are sums or functions of other estimators, so it's useful to know how convergence in probability is affected by these operators. We have the following theorem.

Theorem 1.1. Suppose that $\hat{\theta}_n$ converges in probability to θ and that $\hat{\theta}'_n$ converges in probability to θ'. Then,

1. $\hat{\theta}_n + \hat{\theta}'_n \xrightarrow{P} \theta + \theta'$
2. $\hat{\theta}_n \cdot \hat{\theta}'_n \xrightarrow{P} \theta \cdot \theta'$
3. If $\hat{\theta}'_n \neq 0$, then $\frac{\hat{\theta}_n}{\hat{\theta}'_n} \xrightarrow{P} \frac{\theta}{\theta'}$
4. If g is a real valued function that is continuous at θ, then $g(\hat{\theta}_n) \xrightarrow{P} g(\theta)$.

Now that we understand convergence in probability, we can restate the weak law of large numbers. Convince yourself that the following statement of the weak law of large numbers is equivalent to the one presented in the lecture notes.

Theorem 1.2 (Weak Law of Large Numbers)**.** Let $X_1, \ldots, X_n$ be a sequence of independent identically distributed (iid) random variables, each having a mean $E[X_i] = \mu$ and variance $\sigma^2 < \infty$. Let

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}.$$

Then for any $\epsilon > 0$,

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

By taking $n \to \infty$, it follows that $\bar{X} \xrightarrow{P} \mu$.

Now we've set the stage so that we can understand what it means for an estimator to be consistent.

Definition 1.2. Let $\hat{\theta}_n$ be an estimator for θ by using a sample of size n. The estimator $\hat{\theta}_n$ is *consistent* if $\hat{\theta}_n \xrightarrow{P} \theta$.

There are a few ways to show an estimator is consistent. The most common way will be to apply the weak law of large numbers. In particular, we will often use that an application of the weak law shows that the sample moments converge to the population moments (if $E[X^k] < \infty$, then $\frac{1}{n}\sum_{i=1}^n X_i^k \xrightarrow{P} E[X^k]$.) If the estimator in question is not made up of sample moments, we will need to take a different approach to prove consistency. If the estimator is unbiased, one approach is to show that the variance converges to 0:

Theorem 1.3. An unbiased estimator $\hat{\theta}_n$ for θ is a consistent estimator of θ if

$$\lim_{n\to\infty} V(\hat{\theta}_n) = 0.$$

The proof of this is a simple application of Chebyshev's inequality. Note that this theorem does require the estimator to be unbiased. In general, there is no relationship between consistency and bias of an estimator. Here are a few examples to show that one does not imply the other.

Example 1.1 (Unbiased and consistent)**.** In problem 1 on homework 8, $\hat{\mu} = \bar{X}$ is unbiased and consistent. We've shown previously that $E[\bar{X}] = \mu$, so $\hat{\mu}$ is unbiased. By the weak law of large numbers, $\bar{X} \xrightarrow{P} \mu$, so $\hat{\mu}$ is also consistent.

Example 1.2 (Unbiased but not consistent)**.** Consider an estimator of μ given by $\hat{\mu}' = X_1$. Then $E[X_1] = \mu$, so $\hat{\mu}'$ is unbiased. However, since we only have one sample, $\hat{\mu}'$ does not converge to any value and is not consistent.

Example 1.3 (Biased but consistent)**.** In problem 1 on homework 8, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$ is a biased estimator but is consistent. We showed on a previous homework that $E[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$, so $\hat{\sigma}^2$ is biased. However, by the weak law of large numbers, $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{P} E[X^2] = \mu^2 + \sigma^2$, and by Theorem 1.1 part 4, $\bar{X}^2 \xrightarrow{P} \mu^2$. Applying Theorem 1.1 part 1, we see that $\hat{\sigma}^2 \xrightarrow{P} \mu^2 + \sigma^2 - \mu^2 = \sigma^2$. Thus, $\hat{\sigma}^2$ is consistent.