

ECE531 Screencast 4.5: Asymptotic Properties of Maximum Likelihood Estimation

D. Richard Brown III

Worcester Polytechnic Institute

Preliminaries: Stochastic Convergence

There are many forms of stochastic convergence including

- ▶ Convergence in **distribution** (weak convergence). A series of random variables X_1, X_2, \dots converges in distribution to another random variable X if

$$\lim_{n \rightarrow \infty} p_{X_n}(x) = p_X(x)$$

- ▶ Convergence in **probability**. A series of random variables X_1, X_2, \dots converges in probability to another random variable X if

$$\lim_{n \rightarrow \infty} \text{Prob} [|X_n - X| \geq \epsilon] = 0$$

- ▶ Convergence in **mean**. A series of random variables X_1, X_2, \dots converges in mean to another random variable X if

$$\lim_{n \rightarrow \infty} \text{E} [|X_n - X|] = 0$$

These are not equivalent.

Stochastic Convergence Example

Suppose X_n is a random variable with probability mass function

$$p_{X_n}(x) = \frac{n-1}{n}\delta(x) + \frac{1}{n}\delta(x-n)$$

First consider convergence in distribution. We see that, as $n \rightarrow \infty$,

$$p_{X_n}(x) \rightarrow \delta(x).$$

Hence $p_X(x) = \delta(x)$ is the limiting distribution.

Now consider convergence in mean. We can write

$$\mathbb{E}[X_n] = 0 \times \frac{n-1}{n} + n \times \frac{1}{n} = 1$$

for all n . Note $\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = 1 \neq 0 = \mathbb{E}[X]$.

Bottom line: Convergence in distribution doesn't necessarily imply that the limiting mean/variance of a sequence of random variables has the same mean/variance as the limiting distribution.

Asymptotic Normality of ML Estimators

Main idea: For i.i.d. observations each distributed as $p_Z(z; \theta)$ and under some regularity conditions similar to those you've seen before,

$$\sqrt{n}(\hat{\theta}_{\text{ML},n}(Y) - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{i(\theta)}\right)$$

where \xrightarrow{d} means convergence in distribution and

$$i(\theta) := \mathbb{E} \left\{ \left[\frac{\partial}{\partial \theta} \ln p_Z(Z; \theta) \right]^2 \right\}$$

is the Fisher information of a single observation Y_k about the parameter θ .

See Appendix 7B in Kay vI or Proposition IV.D.2 in the Poor textbook for the details.

A similar version of this theorem holds for vector parameters as well (Kay vI, Theorem 7.3).

Asymptotic Unbiasedness/Efficiency of ML Estimators

Technically, convergence in distribution is not sufficient to imply

$$E[\sqrt{n}(\hat{\theta}_n(Y) - \theta)] \rightarrow 0 \text{ (asymptotic unbiasedness)}$$

or

$$\text{var}[\sqrt{n}(\hat{\theta}_n(Y) - \theta)] \rightarrow \frac{1}{i(\theta)} \text{ (asymptotic efficiency)}.$$

Both of these properties are based on convergence in mean. Our theorem only guarantees convergence in distribution (under regularity).

Nevertheless, these asymptotic properties of the MLE are true in most cases of interest (as seen in our examples). When an estimator is asymptotically unbiased and asymptotically efficient, it is asymptotically MVU. This is almost always the case for MLE.