IEOR 290A – Lecture 3
Stochastic Convergence

1 Types of Stochastic Convergence

There are several types of stochastic convergence, and they can be thought of as direct analogs of *convergence of measures*. Here, we will only be concerned with two basic types of stochastic convergence that are commonly used in the theory of regression.

1.1 Convergence in Distribution

A sequence of random variables $X_1, X_2, \ldots$ converges in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(u) = F_X(u),$$

for every point u at which $F_X(u)$ is continuous, where $F_{X_n}(u)$ is the distribution function for X_n and $F_X(u)$ is the distribution function for X. This type of convergence is denoted $X_n \xrightarrow{d} X$.

1.2 Convergence in Probability

A sequence of random variables $X_1, X_2, \ldots$ converges in probability to a random variable X if for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0.$$

This type of convergence is denoted $X_n \xrightarrow{p} X$ or as $X_n - X \xrightarrow{p} 0$. There are a few additional notations for denoting convergence in distribution, and these are similar to big-O notation that is used in mathematics. We define the following little-o_p notation

$$X_n = o_p(a_n) \Leftrightarrow X_n/a_n \xrightarrow{p} 0.$$

There is a similar big-O_p notation that denotes stochastic boundedness. We say that $X_n = O_p(a_n)$ if for any $\epsilon > 0$ there finite $M > 0$ such that

$$\mathbb{P}(|X_n/a_n| > M) < \epsilon, \forall n.$$

1.3 Relationships Between Modes of Convergence

There are several important points to note:

- Convergence in probability implies convergence in distribution.
- Convergence in distribution does not always imply convergence in probability.
- If X_n converges in distribution to a constant x_0, then X_n also converges in probability to x_0.

2 Concentration About the Mean

Consider an infinite sequence of (mutually) independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$, and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample average. There are a number of results that show that the sample average $\overline{X}_n$ is "close" to the mean of the distribution. The intuition is that the sample average can only deviate "far" from the mean if the random variables act in concert to pull the average in one direction, but the probability that the random variables pull in the same direction is small because of their independence.

2.1 Weak Law of Large Numbers

If the random variables X_i have a finite first moment $\mathbb{E}|X_i| < \infty$, then $\overline{X}_n \xrightarrow{p} \mu$ where $\mu = \mathbb{E}(X_i)$. In words — the weak law of large numbers states that if i.i.d. random variables have a finite first moment, then their sample average converges in probability to the mean of the distribution. Note that we could also write this result as $\overline{X}_n - \mu = o_p(1)$.

2.2 Central Limit Theorem

A more precise statement of the convergence of sample averages is given by the following theorem: If the random variables X_i with mean $\mathbb{E}(X_i) = \mu$ have a finite variance $\mathrm{Var}(X_i) = \sigma^2 < \infty$, then

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where $\mathcal{N}(0, \sigma^2)$ is the distribution of a Gaussian random variable with mean 0 and variance σ^2. This is a more precise statement because it describes the distribution of the sample average when it is appropriately scaled by $\sqrt{n}$. This scaling is important because otherwise $\overline{X}_n \xrightarrow{d} \mu$ by the weak law of large numbers (and the fact that convergence in probability implies convergence in distribution).

3 Extensions of Central Limit Theorem

There are a number of extensions of the Central Limit Theorem, and these are important for understanding the convergence properties of regression estimators, which can be quite complicated.

3.1 Continuous Mapping Theorem

Suppose that $g(u)$ is a continuous function. The continuous mapping theorem states that convergence of a sequence of random variables $\{X_n\}$ to a limiting random variable X is preserved under continuous mappings. More specifically:

- If $X_n \xrightarrow{d} X$, then $g(X_n) \xrightarrow{d} g(X)$.
- If $X_n \xrightarrow{p} X$, then $g(X_n) \xrightarrow{p} g(X)$.

3.2 Slutsky's Theorem

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} y_0$, where y_0 is a constant, then

- $X_n + Y_n \xrightarrow{d} X + y_0$;
- $Y_n X_n \xrightarrow{d} y_0 X$;

There is a technical point to note about this theorem: Convergence of Y_n to a constant is a subtle but important feature for this result, because the theorem will not generally hold when Y_n converges to a non-constant random variable. Consider the example where $X_n \sim \mathcal{N}(0,1)$ and $Y_n = X_n$, then $X_n + Y_n = \mathcal{N}(0,4)$ which does not converge in distribution to $\mathcal{N}(0,1) + \mathcal{N}(0,1) = \mathcal{N}(0,2)$.

3.3 Delta Method

Using the Continuous Mapping Theorem and Slutsky's Theorem, we can now state an extension of the Central Limit Theorem. Consider an infinite sequence of random variables $\{X_n\}$ that satisfies

$$\sqrt{n}[X_n - \theta] \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where θ and σ^2 are finite constants. If $g(u)$ is continuously differentiable at θ and $g'(\theta) \neq 0$, then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{d} g'(\theta)\mathcal{N}(0, \sigma^2).$$

This result is derived using the Lagrange form of Taylor's Theorem, and can hence be thought of as a Taylor polynomial expansion version of the Central Limit Theorem. There are higher-order versions of the Delta Method that we do not discuss here. For instance, the Second-Order Delta Method applies when $g'(u) = 0$ and $g''(u) \neq 0$.