Chapter 5

Convergence

5.1 Types of Convergence

5.1.1 Almost Sure Convergence

Definition 5.1.1. A sequence of random variables X_n converges **almost surely** to another random variable X if

$$P(\lim_{n\to\infty} X_n = X) = 1$$

Or equivalently

$$P(s : X_n(s) \to X(s)) = 1$$

and denoted as $X_n \overset{as}{\to} X$

5.1.2 Convergence in Mean

Definition 5.1.2. Let X_n be a sequence of random variable. X_n converges to X **in the** r^{th} **mean** or **in the** L^r **norm** to a random variable X if

$$\lim_{n\to\infty} E[|X_n - X|^r] = 0$$

and denoted by $X_n \overset{L^r}{\to} X$. We frequently use L^1 and L^2 convergence.

5.1.3 Convergence in Probability

Definition 5.1.3 (Converge in Probability)**.** Let X_n be a sequence of random variables and let X be another random variable. We say that X_n **converges in probability to X** if, for all $\epsilon > 0$

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

or equivalently

$$\lim_{n\to\infty} P(|X_n - X| \leq \epsilon) = 1$$

We write as

$$X_n \overset{P}{\to} X$$

Theorem 5.1.1. *Suppose $X_n \overset{P}{\to} X$, and a is a constant, then $aX_n \overset{P}{\to} aX$*

Proof. Let $\epsilon > 0$, then

$$P(|aX_n - aX| \geq \epsilon) = P(|a||X_n - X| \geq \epsilon) = P(|X_n - X| \geq \epsilon/|a|)$$

Since X_n converges in probability to X, the above equation goes to 0 as $n \to \infty$. □

Theorem 5.1.2. *Suppose* $X_n \xrightarrow{P} X$ *and* $Y_n \xrightarrow{P} Y$*, then* $X_n + Y_n \xrightarrow{P} X + Y$

Proof. Fix some $\epsilon > 0$, we want to find

$$\begin{aligned} &\lim_{n\to\infty} P(|X_n + Y_n - (X + Y)| \geq \epsilon) \\ =&\lim_{n\to\infty} P(|X_n + Y_n - X - Y| \geq \epsilon) \\ =&\lim_{n\to\infty} P(|X_n - X + Y_n - Y| \geq \epsilon) \end{aligned}$$

Notice that the event $X_n - X + Y_n - Y \geq \epsilon$ is a subset of event $X_n - X \geq \epsilon/2 \cup Y_n - Y \geq \epsilon/2$, hence:

$$\leq \lim_{n\to\infty} P(|X_n - X| \geq \epsilon/2 \quad \cup \quad |Y_n - Y| \geq \epsilon/2)$$

By probability axiom that $P(A \cup B) \leq P(A) + P(B)$

$$\leq \lim_{n\to\infty} P(|X_n - X| \geq \epsilon/2) + P(|Y_n - Y| \geq \epsilon/2)$$

Since $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon/2) = 0 \text{ and } \lim_{n\to\infty} P(|Y_n - Y| \geq \epsilon/2) = 0$$

Therefore

$$P(|X_n - X| \geq \epsilon/2) + P(|Y_n - Y| \geq \epsilon/2) = 0 + 0 = 0$$

Hence:

$$\lim_{n\to\infty} P(|X_n + Y_n - (X + Y)| \geq \epsilon) \leq \lim_{n\to\infty} P(|X_n - X| \geq \epsilon/2) + P(|Y_n - Y| \geq \epsilon/2) = 0$$

Thus $X_n + Y_n \xrightarrow{P} X + Y$ by definition of convergence. □

Theorem 5.1.3. *Suppose* $X_n \xrightarrow{P} a$ *and the real function* g *is continuous at* a*. Then* $g(X_n) \xrightarrow{P} g(a)$

Proof. Let $\epsilon > 0$, since g is continuous at a, by the definition of continuity, there exist $\delta > 0$ such that if $|x - a| < \delta$, then $|g(x) - g(a)| < \epsilon$, thus

$$|g(x) - g(a)| \geq \epsilon \Rightarrow |x - a| \geq \delta$$

Substituting X_n for x we obtain

$$|g(X_n) - g(a)| \geq \epsilon \Rightarrow |X_n - a| \geq \delta$$

and hence

$$P(|g(X_n) - g(a)| \geq \epsilon) \leq P(|X_n - a| \geq \delta)$$

we know that $P(|X_n - a| \geq \delta)$ converges to 0 as $n \to \infty$ by definition, hence $P(|g(X_n) - g(a)| \geq \epsilon)$ converges to 0 as $n \to \infty$ as well. □

Theorem 5.1.4. *Suppose* $X_n \xrightarrow{P} X$ *and* $Y_n \xrightarrow{P} Y$*, then* $X_nY_n \xrightarrow{P} XY$

Proof. We know that

$$\begin{aligned}(X_n - Y_n)^2 &= X_n^2 - 2X_nY_n + Y_n^2 \\ 2X_nY_n &= X_n^2 + Y_n^2 - (X_n - Y_n)^2 \\ X_nY_n &= \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2\end{aligned}$$

Using the last theorem, we know that

$$\begin{aligned}&\xrightarrow{P} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}(X - Y)^2 \\ &= XY\end{aligned}$$

□

5.1.4 Convergence in Distribution

Definition 5.1.4 (Converge in Distribution)**.** Let X_n be a sequence of random variables and let X be another random variable. Let F_n and F_X be, respectively, the cdfs of X_n and X. We say that X_n **converges in distribution to X** if

$$\lim_{n\to\infty} F_n(x) = F_X(x) \quad \text{at all x for which } F \text{ is continuous}$$

We write as $X_n \xrightarrow{D} X$ or $X_n \rightsquigarrow X$

5.2 Relationship between Convergence

So far we have introduced four types of convergence. But do realize that some of these convergence types are "stronger" than others and some are "weak". By this, we mean the following: If Type A convergence is stronger than Type B convergence, it means that Type A convergence implies Type B convergence. Figure 5.1 summarizes how these types of convergence are related. In this figure, the stronger types of convergence are on top and, as we move to the bottom, the convergence becomes weaker. For example, using the figure, we conclude that if a sequence of random variables converges in probability to a random variable X , then the sequence converges in distribution to X as well.

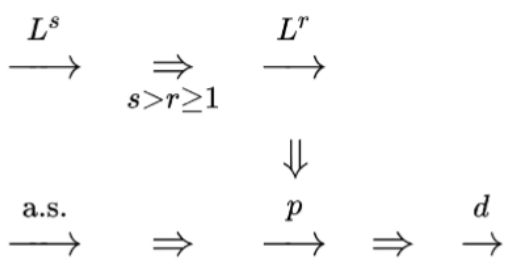

Figure 5.1: Forms of convergence that imply other forms of convergence (source: wikipedia)

Theorem 5.2.1. *If $X_n \xrightarrow{a.s} X$, then $X_n \xrightarrow{P} X$*

Proof. This is almost obvious:

$$X_n \xrightarrow{a.s} X \Rightarrow P(\lim_{n\to\infty} X_n = X) \Rightarrow P(\lim_{n\to\infty} X_n - X = 0 < \epsilon) \Rightarrow X_n \xrightarrow{P} X$$

□

Theorem 5.2.2. *If* $X_n \xrightarrow{L^r} X$ *for some* $r \geq 1$*, then* $X_n \xrightarrow{P} X$

Proof. For any $\epsilon > 0$, we have

$$P(|X_n - X| > \epsilon) = P(|X_n - X|^r > \epsilon^r)$$

Using Markov's inequality

$$P(|X_n - X|^r > \epsilon^r) \leq \frac{E[|X_n - X|^r]}{\epsilon^2}$$

Since $X_n \xrightarrow{L^r} X$, then $\lim_{n\to\infty} E[|X_n - X|^r] = 0$, hence

$$\lim_{n\to\infty} P(|X_n - X|^r > \epsilon^r) \leq 0$$

which is exactly what we need to prove. □

Now we are going to show that convergence in probability implies convergence in distribution, but before proving this theorem, we need to first prove a lemma

Lemma 5.2.3. *Let* X, Y *be random variables, let* a *be a real number and* $\epsilon > 0$*, then*

$$P(Y \leq a) \leq P(X \leq a + \epsilon) + P(|Y - X| > \epsilon)$$

Proof.

$$\begin{aligned}
P(Y \leq a) &= P(Y \leq a \cap X \leq a + \epsilon) + P(Y \leq a \cap X > a + \epsilon) && \text{By the law of total probability} \\
&\leq P(X \leq a + \epsilon) + P(Y \leq a \cap X > a + \epsilon) && P(A \cap B) \leq P(A) \\
&\leq P(X \leq a + \epsilon) + P(Y - X \leq a - X \cap X > a + \epsilon) && Y \leq a \Rightarrow Y - X \leq a - X \\
&\leq P(X \leq a + \epsilon) + P(Y - X \leq a - X \cap a - X < -\epsilon) && X > a + \epsilon \Rightarrow a - X < -\epsilon \\
&\leq P(X \leq a + \epsilon) + P(Y - X < -\epsilon) && \text{Since } Y - X \leq a - X \cap a - X < -\epsilon \\
&\leq P(X \leq a + \epsilon) + P(Y - X < -\epsilon) + P(Y - X > \epsilon) && \forall \omega \in \Omega, P(\omega) \in [0, 1] \\
&= P(X \leq a + \epsilon) + P(|Y - X| > \epsilon)
\end{aligned}$$

□

With the help of this lemma, we can now prove the following theorem easily

Theorem 5.2.4. *If* $X_n \xrightarrow{P} X$*, then* $X_n \xrightarrow{D} X$

Proof. Let $\epsilon > 0$, and let $Y = X_n$, using the last lemma we get

$$P(X_n \leq a) \leq P(X \leq a + \epsilon) + P(|X_n - X| > \epsilon)$$

Let $Y = X, X = X_n, a = a - \epsilon$, we get

$$P(X \leq a - \epsilon) \leq P(X_n \leq a) + P(|X_n - X| > \epsilon)$$

$$\Rightarrow P(X \leq a - \epsilon) - P(|X_n - X| > \epsilon) \leq P(X_n \leq a)$$

Combining these two parts we get

$$P(X \leq a - \epsilon) - P(|X_n - X| > \epsilon) \leq P(X_n \leq a) \leq P(X \leq a + \epsilon) + P(|X_n - X| > \epsilon)$$

Taking $\lim_{n\to\infty}$ we obtain

$$F_X(a - \epsilon) \leq \lim_{n\to\infty} F_X(X_n \leq a) \leq F_X(a + \epsilon)$$

where F_X is the CDF of X. As $\epsilon \to 0^+$, we get

$$\lim_{n\to\infty} F_X(X_n \leq a) = P(X \leq a)$$

which means X_n converges to X in distribution. □

Theorem 5.2.5. *Let X_n, X, Y_n, Y be random variables. Let g be a continuous function.*

1. *If $X_n \xrightarrow{\text{P}} X$ and $Y_n \xrightarrow{\text{P}} Y$, then $X_n + Y_n \xrightarrow{\text{P}} X + Y$.*
2. *If $X_n \xrightarrow{\text{qm}} X$ and $Y_n \xrightarrow{\text{qm}} Y$, then $X_n + Y_n \xrightarrow{\text{qm}} X + Y$.*
3. *If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, then $X_n + Y_n \rightsquigarrow X + c$.*
4. *If $X_n \xrightarrow{\text{P}} X$ and $Y_n \xrightarrow{\text{P}} Y$, then $X_n Y_n \xrightarrow{\text{P}} XY$.*
5. *If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$, then $X_n Y_n \rightsquigarrow cX$.*
6. *If $X_n \xrightarrow{\text{P}} X$, then $g(X_n) \xrightarrow{\text{P}} g(X)$.*
7. *If $X_n \rightsquigarrow X$, then $g(X_n) \rightsquigarrow g(X)$.*

*Part (3) and (5) combined are known as **Slutzky's theorem**. It is worth noting that $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ does not in general imply that $X_n + Y_n \rightsquigarrow X + Y$. .*

5.3 The Law of Large Numbers

5.3.1 The Weak Law of Large Numbers (WLLN)

The **weak law of large numbers** (also called Khinchin's law) states that the sample average converges in probability towards the expected value

Theorem 5.3.1 (Weak Law of Large Numbers)**.** *If $X_1, \cdots, X_n$ are I.I.D random variables with common mean μ and variance $\sigma^2 < \infty$, Let $\overline{X}_n = (\sum_{i=1}^n X_i)/n$ then*

$$\overline{X}_n \xrightarrow{P} \mu$$

or

$$\forall \epsilon > 0, P(|\overline{X}_n - \mu| > \epsilon) \to 0$$

Proof. Using Chebyshev's inequality, for any $\epsilon > 0$

$$P(|\overline{X}_n - \mu| > \epsilon) \leq \frac{V[\overline{X}_n - \mu]}{\epsilon^2} = \frac{V[\overline{X}_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

which tends to be 0 as $n \to \infty$ □

5.3.2 The Kolmogorov's Strong Law of Large Number

The strong law of large numbers (also called Kolmogorov's law) states that the sample average converges **almost surely** to the expected value.

Theorem 5.3.2 (Strong Law of Large Numbers)**.** *If $X_1, \cdots, X_n$ are I.I.D random variables with common mean μ and variance $\sigma^2 < \infty$, Let $\overline{X}_n = (\sum_{i=1}^n X_i)/n$ then*

$$\overline{X}_n \xrightarrow{a.s} \mu$$

Proof. □

5.4 The Central Limit Theorem (CLT)

The law of large numbers says that the distribution of $\overline{X}_n$ centers near μ, but that does not give us enough information about $\overline{X}_n$. Hence we introduce central limit theorem (CLT). CLT states that, $\overline{X}_n$ has a distribution which is approximately Normal $\mathcal{N}(\mu, \sigma^2/n)$.

Theorem 5.4.1 (Central Limit Theorem)**.** *Let $X_1, \cdots, X_n$ be IID random variable with mean μ and variance $\sigma^2 < \infty$. Let $\overline{X}_n = (\sum_{i=1}^n X_i)/n$. Then, the random variable*

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$$

converges in distribution to the standard normal random variable as n goes to infinity, that is

$$Z_n \xrightarrow{D} Z \text{ or } \lim_{n\to\infty} P(Z_n \leq x) = \Phi(x)$$

where $\Phi(x)$ is the standard normal CDF.

Proof. □

5.4.1 Lyapunov CLT

Theorem 5.4.2 (Lyapunov CLT)**.** *Suppose $X_1, \ldots, X_n$ are independent but **not** necessarily identically distributed. Let $\mu_i = \mathbb{E}[X_i]$, let $\sigma_i = \text{Var}(X_i)$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Then if we satisfy the Lyapunov condition:*

$$\lim_{n\to\infty} \frac{1}{s_n^3} \sum_{i=1}^n \mathbb{E}|X_i - \mu|^3 = 0$$

then

$$\frac{1}{s_n} \sum_{i=1}^n [X_i - \mu_i] \xrightarrow{d} N(0,1).$$

First notice that if we are dealing with IID case, then we have usual CLT.

Consider the case when the Lyapunov condition is violated. When all the random variables are deterministic except X_1 which has ($\mu_1 \text{and} \sigma_1 > 0$). Then $\sigma_n^3 = \sigma_1^3$ with the third moment $\mathbb{E}|X_1 - \mu|^3 > 0$ so the condition fails.

Roughly, what can happen in the non-identically distributed case is that only one random variable can dominate the sum in which case you are not really averaging many things so you do not have a CLT.

5.4.2 CLT with Estimated Variance

We saw that in our typical use case of the CLT (constructing confidence intervals) we needed to know the variance σ. In practice, we most often do not know this. However, we can estimate this quantity in the usual way,

$$\widehat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \widehat{\mu})^2$$

It turns out that we can replace the standard deviation in the CLT by $\widehat{\sigma}$ and still have the same convergence in distribution, i.e.

$$\frac{\sqrt{n}(\widehat{\mu} - \mu)}{\widehat{\sigma}_n} \xrightarrow{d} N(0,1)$$

Proof. First observe that if we can show that $\frac{\sigma}{\widehat{\sigma}_n} \xrightarrow{d} 1$, then an application of Slutsky's theorem and the CLT gives us the desired result.

Since square-root is a continuous map, by the continuous mapping theorem, it suffices to show that $\frac{\sigma^2}{\widehat{\sigma}_n^2} \xrightarrow{d} 1$. We will instead show the stronger statement that,

$$\widehat{\sigma}_n^2 \xrightarrow{p} \sigma^2,$$

which implies the desired statement via the continuous mapping theorem (see Larry's notes for more details). Note that,

$$\begin{aligned}\widehat{\sigma}_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2 \\ &\xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2\end{aligned}$$

using the fact that $\frac{n-1}{n} \to 1$. Now,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)^2 \xrightarrow{p} \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$$

using the WLLN. This concludes the proof. □

5.4.3 Rate of Convergence in CLT

While the central limit theorem is an asymptotic result (i.e. a statement about $n \to \infty$) it turns out that under fairly general conditions we can say how close to a standard normal the average is, in distribution, for finite values n. Such results are known as Berry Esseen bounds. Roughly, they are proved by carefully tracking the remainder terms in our Taylor series proof.

Theorem 5.4.3 (Berry-Esseen)**.** *Suppose that* $X_1, \ldots, X_n \sim P$*. Let* $\mu = \mathbb{E}[X_1], \sigma^2 = \mathbb{E}\left[(X_1 - \mu)^2\right]$*, and* $\mu_3 = \mathbb{E}\left[|X_1 - \mu|^3\right]$*. Let*

$$F_n(x) = \mathbb{P}\left(\frac{\sqrt{n}(\widehat{\mu} - \mu)}{\sigma} \leq x\right)$$

denote the CDF of the normalized sample average. If $\mu_3 < \infty$ *then,*

$$\sup_x |F_n(x) - \Phi(x)| \leq \frac{9\mu_3}{\sigma^3\sqrt{n}}$$

This bound is roughly saying that if μ_3/σ^3 is small then the convergence to normality in distribution happens quite fast.

5.5 The Delta Method

Theorem 5.5.1 (The Delta Method)**.** *Suppose that*

$$\frac{\sqrt{n}\,(Y_n - \mu)}{\sigma} \rightsquigarrow N(0,1)$$

and that g is a differentiable function such that $g'(\mu) \neq 0$. Then

$$\frac{\sqrt{n}\left(g\left(Y_n\right)-g(\mu)\right)}{\left|g'(\mu)\right| \sigma} \rightsquigarrow N(0,1).$$

In other words,

$$Y_n \approx N\left(\mu, \frac{\sigma^2}{n}\right) \quad \textit{implies that} \quad g\left(Y_n\right) \approx N\left(g(\mu),\left(g'(\mu)\right)^2 \frac{\sigma^2}{n}\right)$$

Proof. The basic idea is simply to use Taylor's approximation. We know that,

$$g\left(X_n\right) \approx g(\mu)+g'(\mu)\left(X_n-\mu\right)$$

so that,

$$\frac{\sqrt{n}\left(g\left(X_n\right)-g(\mu)\right)}{\sigma} \approx g'(\mu) \frac{\sqrt{n}\left(X_n-\mu\right)}{\sigma} \xrightarrow{d} N\left(0,\left[g'(\mu)\right]^2\right).$$

To be rigorous however we need to take care of the remainder terms. Here is a more formal proof.

By a rigorous application of Taylor's theorem we obtain,

$$\frac{\sqrt{n}\left(g\left(X_n\right)-g(\mu)\right)}{\sigma}=g'(\widetilde{\mu}) \frac{\sqrt{n}\left(X_n-\mu\right)}{\sigma}$$

where $\tilde{\mu}$ is on the line joining μ to $\widehat{\mu}$. We know by the WLLN that $\widehat{\mu} \xrightarrow{p} \mu$ and so $\widetilde{\mu} \xrightarrow{p} \mu$. Since g is continuously differentiable, we can use the continuous mapping theorem to conclude that,

$$g'(\widetilde{\mu}) \xrightarrow{p} g'(\mu).$$

Now, we apply Slutsky's theorem to obtain that,

$$g'(\widetilde{\mu}) \frac{\sqrt{n}\left(X_n-\mu\right)}{\sigma} \xrightarrow{d} g'(\mu) N(0,1) \stackrel{d}{=} N\left(0,\left[g'(\mu)\right]^2\right).$$

□

Example 5.5.1. Let $X_1, \ldots, X_n$ be IID with finite mean μ and finite variance σ^2. By the central limit theorem, $\sqrt{n}\left(\bar{X}_n-\mu\right) / \sigma \rightsquigarrow N(0,1)$. Let $W_n=e^{\bar{X}_n}$. Thus, $W_n=g\left(\bar{X}_n\right)$ where $g(s)=e^s$. Since $g'(s)=e^s$, the delta method implies that $W_n \approx N\left(e^\mu, e^{2 \mu} \sigma^2 / n\right)$.