Convergence of Random Variables

* We will often wish to examine sequences of random variables.

* In some cases these sequences may converge to a limiting random variable in some sense.

* We have already seen this in the sense of convergence of moment generating functions implying convergence of cumulative distribution functions.

* This is one form of convergence of random variables.

Convergence in Distribution

Definition 7.1

Suppose that $X_1, X_2, \ldots$ *is a sequence of random variables. We say that this sequence converges in distribution to a random variable* X *if*

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

at all points x *at which* $F_X(x) = P(X \leqslant x)$ *is continuous.*

We will denote this type of convergence by $X_n \xrightarrow{d} X$.

Examples of Convergence in Distribution

* If $X_n \sim \text{Binomial}(n, p)$ then

$$F_{Z_n}(z) \;=\; \mathsf{P}\left(\frac{X_n - np}{\sqrt{np(1-p)}} \leqslant z\right) \;\to\; \Phi(z) \quad \text{as } n \to \infty$$

* Suppose $X_n \sim \text{Binomial}(n, \lambda/n)$ then

$$\mathsf{P}(X_n = x) \;\to\; \frac{\lambda^n e^{-\lambda}}{x!}$$

Hence $X_n \xrightarrow{d} X$ where $X \sim \text{Poisson}(\lambda)$.

* Suppose $U_i \overset{iid}{\sim} \text{Uniform}(0,1)$ and $X_n = \max\{U_1, \ldots, U_n\}$ then

$$\mathsf{P}(X_n \leqslant x) \;\to\; \begin{cases} 0 & x < 1 \\ 1 & x \geqslant 1 \end{cases}$$

In this sense we can say that $X_n \xrightarrow{d} 1$.

Central Limit Theorem

Theorem 7.1 (Central Limit Theorem)

Let $X_1, X_2, \ldots$ *be a sequence of* iid *random variables with* $\mathsf{E}(X_i) = \mu$ *and* $\mathsf{Var}(X_i) = \sigma^2 < \infty$. *Define the sample mean* $\overline{X}_n = n^{-1} \sum_1^n X_i$ *and let* $F_n(x)$ *denote the cdf of the random variable*

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

Then for any $x \in \mathbb{R}$,

$$\lim_{n \to \infty} F_n(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \, dz$$

Convergence in Distribution

* This is quite a weak form of convergence in general.

* $X_n \xrightarrow{d} X$ only means that the distribution functions converge but doesn't say anything about whether the random variables are close in any sense.

* Note that, a sequence of continuous cdf's can converge but the corresponding sequence of pdf's may not.

Convergence in Probability

Definition 7.2

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

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

We will generally write this as

$$X_n \xrightarrow{p} X$$

Some Mathematical Concepts

Definition 7.3

Suppose that $\{a_n\}$ is a sequence of real numbers. Define

$$
\begin{aligned}
b_n &= \sup\{a_n, a_{n+1}, a_{n+2}, \ldots\} \\
c_n &= \inf\{a_n, a_{n+1}, a_{n+2}, \ldots\}
\end{aligned}
$$

We define the Limit Superior ($\limsup$) and Limit Inferior ($\liminf$) of the sequence $\{a_n\}$ to be

$$
\begin{aligned}
\limsup_{n\to\infty} a_n &= \lim_{n\to\infty} b_n \\
\liminf_{n\to\infty} a_n &= \lim_{n\to\infty} c_n
\end{aligned}
$$

The sequence $\{a_n\}$ converges to a limit if, and only if,

$$
\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n = \lim_{n\to\infty} a_n
$$

Properties of Convergence in Probability

Theorem 7.2

Suppose that $X_1, X_2, \ldots$ converges in probability to a random variable X and that h is a continuous function. Then

$$h(X_n) \xrightarrow{p} h(X)$$

Theorem 7.3

Suppose that $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ are two sequences of random variables such that $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$. Then

1. $X_n + Y_n \xrightarrow{p} X + Y$.

2. $X_n Y_n \xrightarrow{p} XY$.

Relationship between Convergence in Probability and Distribution

Theorem 7.4

If a sequence of random variables $\{X_n\}$ converges in probability to a random variable X then the sequence also converges in distribution to X.

In general the converse of this theorem is not true but it is true in one special case.

Theorem 7.5

If a sequence of random variables $\{X_n\}$ converges in distribution to a constant μ then random variable X then the sequence also converges in probability to μ.

Weak Law of Large Numbers

Theorem 7.6 (Weak Law of Large Numbers)
Let $X_1, X_2, \ldots$ *be* iid *random variables with* $\mathrm{E}(X_i) = \mu$ *and* $\mathrm{Var}(X_i) = \sigma^2 < \infty$. *Define the sample mean* $\overline{X}_n = n^{-1} \sum_1^n X_i$. *Then the sequence of random variables* $\overline{X}_1, \overline{X}_2, \ldots$ *converges in probability to the constant* μ. *That is for every* $\varepsilon > 0$

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| < \varepsilon) = 1$$

Slutsky's Theorem

Theorem 7.7 (Slutsky's Theorem)
If $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} a$ *where* a *is a constant then*

1. $X_n + Y_n \xrightarrow{d} X + a.$

2. $X_n Y_n \xrightarrow{d} aX.$

Almost Sure Convergence

Definition 7.4

A sequence of random variables $X_1, X_2, \ldots$ *converges almost surely to a random variable* X *if, for every* $\varepsilon > 0$,

$$P(\lim_{n\to\infty} |X_n - X| < \varepsilon) \;=\; 1$$

Theorem 7.8

If a sequence of random variables $X_1, X_2, \ldots$ *converges almost surely to a random variable* X*, the sequence also converges in probability to* X*.*

Theorem 7.9

Suppose that the sequence $X_1, X_2, \ldots$ *converges in probability to a random variable* X *then there exists a subsequence of* $X_1, X_2, \ldots$ *which converges almost surely to* X*.*

Strong Law of Large Numbers

Theorem 7.10 (Strong Law of Large Numbers)
Let $X_1, X_2, \ldots$ *be* iid *random variables with finite mean* $\mathsf{E}(X_i) = \mu$ *and finite variance* $\mathsf{Var}(X_i) = \sigma^2$. *Define the sample mean* $\overline{X}_n = n^{-1}\sum_1^n X_i$. *Then the sequence of random variables* $\overline{X}_1, \overline{X}_2, \ldots$ *converges almost surely to the constant* μ. *That is for every* $\varepsilon > 0$

$$P(\lim_{n\to\infty} |\overline{X}_n - \mu| < \varepsilon) = 1$$

* In fact the requirement for finite variance is a stronger condition than required. Both Laws of Large Numbers are true even when only the mean is finite.

Bounded in Probability

Definition 7.5

Suppose that $X_1, X_2, \ldots$ is a sequence of random variables. We say that the sequence is bounded in probability if for all $\varepsilon > 0$, there exists a constant B_ε and an integer N_ε such that

$$n \geqslant N_\varepsilon \quad \Rightarrow \quad \mathsf{P}\Big(|X_n| \leqslant B_\varepsilon\Big) \geqslant 1 - \varepsilon.$$

Theorem 7.11

Suppose that $X_1, X_2, \ldots$ is a sequence of random variables and that there is a random variable X such that $X_n \xrightarrow{d} X$ then $\{X_n\}$ is bounded in probability.

Theorem 7.12

Suppose that the sequence of random variables $\{X_n\}$ is bounded in probability and $Y_n \xrightarrow{p} 0$ then $X_n Y_n \xrightarrow{p} 0$.

O_p and o_p Notation

Definition 7.6

Suppose that $\{X_n\}$ is a sequence of random variables and $\{a_n\}$ is a sequence of constants. Then we say that

$$X_n = o_p(a_n) \iff \frac{X_n}{a_n} \xrightarrow{p} 0$$

We say that

$$X_n = O_p(a_n) \iff \frac{X_n}{a_n} \text{ is bounded in probability}$$

We can also replace the constants $\{a_n\}$ by a sequence of random variables $\{Y_n\}$.

Theorem 7.13

Suppose that $\{Y_n\}$ is a sequence of random variables which are bounded in probability and that $\{X_n\}$ is another sequence of random variables such that $X_n = o_p(Y_n)$. Then $X_n \xrightarrow{p} 0$.

Delta Method

Theorem 7.14 (Delta Method)

Let $Y_1, Y_2, \ldots$ *be a sequence of random variables such that* $\sqrt{n}(Y_n - \theta)$ *converges in distribution to a normal*$(0, \sigma^2)$ *random variable. Suppose that* g *is a function such that* $g'(\theta)$ *exists and is not 0, then*

$$\sqrt{n}\Big(g(Y_n) - g(\theta)\Big) \xrightarrow{d} \textit{normal}\left(0, \Big(g'(\theta)\sigma\Big)^2\right)$$

Theorem 7.15 (Second Order Delta Method)

Let $Y_1, Y_2, \ldots$ *be a sequence of random variables such that* $\sqrt{n}(Y_n - \theta)$ *converges in distribution to a normal*$(0, \sigma^2)$ *random variable. Suppose that* g *is a function such that* $g'(\theta) = 0$ *and* $g''(\theta)$ *exists and is not 0, then*

$$n\Big(g(Y_n) - g(\theta)\Big) \xrightarrow{d} \frac{g''(\theta)\sigma^2}{2}\chi_1^2$$