CONVERGENCE OF A SEQUENCE OF RANDOM VARIABLES

Consider the problem of interpreting the limit in the relative frequency definition of probability

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

Clearly, this limit can be interpreted only in the probabilistic sense. The aim of this module of the lectures is to introduce the concepts behind such limits. More formally, it will deal with the convergence of a sequence of random variables. This convergence concept is the starting point for the advanced concepts of continuity, differentiation and integration of random functions of time.. We start with an outline of the convergence of a sequence of real numbers

Convergence of a sequence of real numbers

Consider a sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ The sequence converges to a *limit* x if corresponding to every $\varepsilon > 0$, we can find a positive integer N such that $|x - x_n| < \varepsilon$ for all $n > N$. For example, the sequence $1, \frac{1}{2}, \cdots, \frac{1}{n}, \cdots$ converges to the number 0. Because, for any $\varepsilon > 0$, we can choose a positive integer $N > \frac{1}{\varepsilon}$ such that $|0 - x_n| = \frac{1}{n} < \varepsilon$ for $n > N$.

The Cauchy criterion gives the condition for convergence of a sequence without actually finding the limit. The sequence $\{x_n\}$ converges if and only if for every $\varepsilon > 0$ there exists a positive integer N such that $|x_{n+m} - x_n| < \varepsilon$ for all $n > N$ and all $m > 0$.

Example 1: Apply Cauchy criterion to check the convergence of the sequence given by $x_n = \frac{1}{n}$.

Solution: Here

$|x_{n+m} - x_n| = \frac{m}{n(n+m)}$ According to Cauchy criterion, we require

$$\frac{m}{n(n+m)} < \varepsilon$$

$$\Rightarrow n(n+m) > \frac{m}{\varepsilon}$$

$$\Rightarrow (n+m)^2 > \frac{m}{\varepsilon}$$

$$\Rightarrow n > \sqrt{\frac{m}{\varepsilon}} - m$$

If we choose any positive integer $N > \sqrt{\frac{1}{\varepsilon}} - 1$, then any $n > N$ will satisfy the above inequality.

Limit Superior and Limit Inferior of a sequence of real numbers

Recall that a function $g(.)$ is continuous in $\mathbb{R}$ if and only if for any convergent sequence $\{x_n\}_{n=1}^{\infty}$,

$\lim_{n\to\infty} g(x_n) = g(\lim_{n\to\infty} x_n)$. The probability $P(.)$ is a function defined on events which are members of the sigma field $\mathbb{F}$. The continuity of the probability $P(.)$ can be similarly defined.

If the sequence $\{x_n\}_{n=1}^{\infty}$ is not convergent, we can study the limiting behavior of the sequence in terms of the *limit superior* and the *limit inferior* as defined below.

Definition:

The limit superior of $\{x_n\}_{n=1}^{\infty}$ is defined as

$$\lim_{n\to\infty}\sup(x_n) = \inf_{n\ge 1}(\sup_{j\ge n} x_j)$$

where $\sup_{j\ge n} x_j$ = *least upper bound of* $\{x_j \mid j \ge n\}$ and

$\inf_{j\ge n} x_j$ = *greatest lower bound of* $\{x_j \mid j \ge n\}$.

Similarly, the limit inferior of $\{x_n\}_{n=1}^{\infty}$ is defined as

$$\lim_{n\to\infty}\inf(x_n) = \sup_{n\ge 1}(\inf_{j\ge n} x_j)$$

These limits defined in the case of a sequence of real numbers can be extended to a sequence of events as follows.

Limits of Events in a Probability space

Continuity theorem: Suppose $A_1, A_2, \ldots \in \mathbb{F}$.

(a) If $A_1 \subset A_2 \subset \ldots$, then

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(b) If $A_1 \supset A_2 \supset \ldots$, then

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right)$$

Proof: Clearly $A_n = \bigcup_{i=1}^{n} A_i$ so that

$$A = \lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n .$$

As shown in Figure 5 , A_n can be expressed as a union of disjoint subsets given by

$$A_n = \bigcup_{i=1}^{n} E_i$$

where $E_i = A_i - A_{i-1}$ with $E_0 = \Phi$.

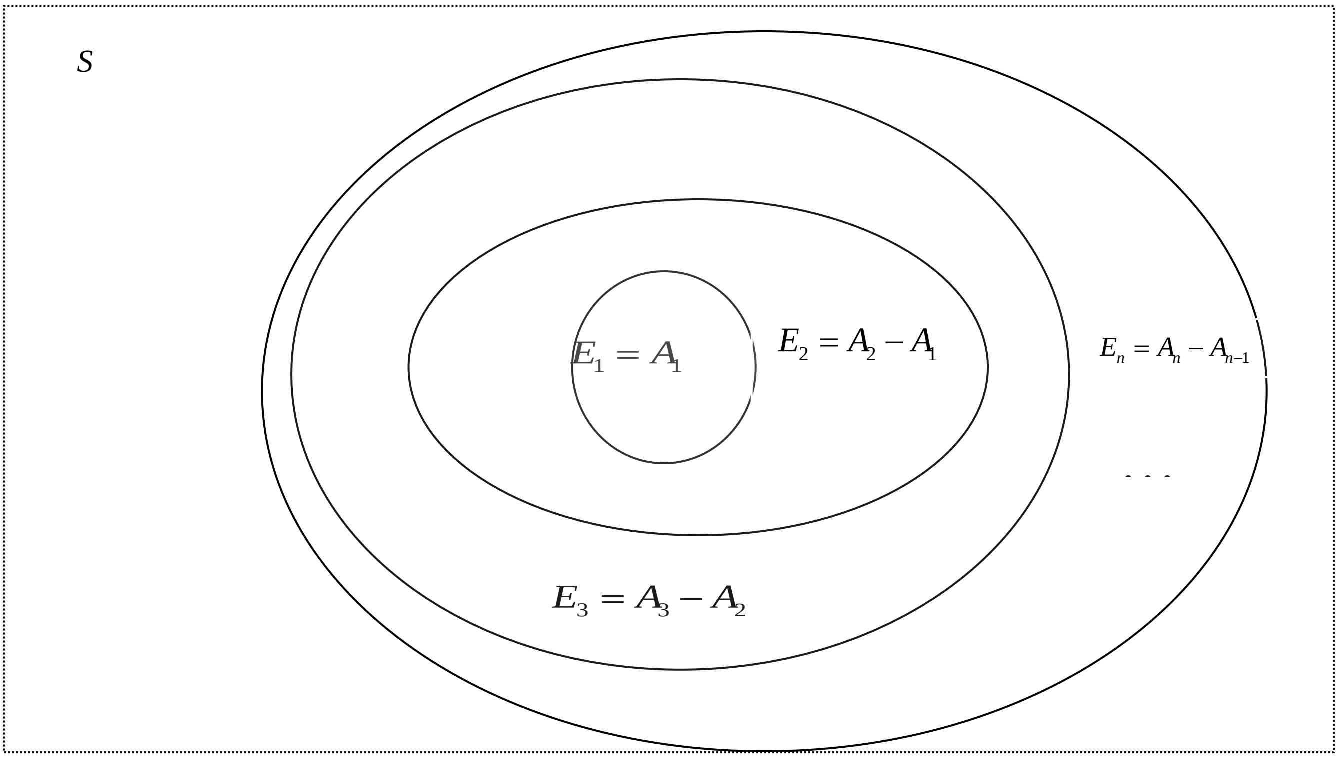


Figure 5: Representation of A_n as a union of disjoint subsets

Then

$$A = \bigcup_{i=1}^{\infty} E_i$$

$$\therefore P(A) = P\left(\bigcup_{i=1}^{\infty} E_i\right)$$

$$= \sum_{i=1}^{\infty} P(E_i) \qquad \text{(Using countable additivity)}$$

Now from

$A_n = \bigcup_{i=1}^{n} E_i$ we get,

$$P(A_n) = \sum_{i=1}^{n} P(E_i)$$

$$\therefore \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(E_i)$$

$$= \sum_{i=1}^{\infty} P(E_i)$$

$$= P(A)$$

$$\therefore \lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

(b) Given $A_1 \supset A_2 \supset \ldots$, we get

$A_1^c \subset A_2^c \subset \ldots,$

Using the results in (a), we get

$$\therefore \lim_{n\to\infty} P(A_n^c) = P\left(\bigcup_{i=1}^{\infty} A_i^c\right)$$

$$= 1 - P\left(\bigcup_{i=1}^{\infty} A_i^c\right)^c$$

$$= 1 - P\left(\bigcap_{i=1}^{\infty} A_i\right)$$

$$\therefore \lim_{n\to\infty} (1 - P(A_n)) = 1 - P\left(\bigcap_{i=1}^{\infty} A_i\right)$$

Therefore,

$$\therefore \lim_{n\to\infty} P(A_n) = P\left(\bigcap_{i=1}^{\infty} A_i\right)$$

The continuity theorem establishes the continuity of the probability function analogous to that of a function of a real variable and is of fundamental importance.

Limit Superior and Limit Inferior of a sequence of events

These limits defined in the case of a sequence of real numbers can be extended to a sequence of events as follows.

Consider an arbitrary sequence $A_1, A_2, \ldots \in \mathbb{F}$ which is not a monotone. We cannot define the limit of this sequence like that of a monotonic sequence.

Suppose $B_n = \bigcup_{k=n}^{\infty} A_k$ and $C_n = \bigcap_{k=n}^{\infty} A_k$. Observe that $\{B_n\}$ and $\{C_n\}$ monotone sequences. $\{B_n\}$ is a monotonically decreasing sequence while $\{C_n\}$ is a monotonically increasing sequence.

Definition: Suppose $\left\{A_n\right\}_{n=1}^{\infty}$ is a sequence of events in $\mathbb{F}$. Then the *limit superior* of the sequence is defined by

$$\lim_{n\to\infty}\sup A_n = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$$

Note that

$$s \in \lim_{n\to\infty}\sup A_n \Leftrightarrow s \in \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$$

$$\Leftrightarrow \forall n, \exists k \geq n \text{ such that } s \in A_k$$

$$\Leftrightarrow s \in A_n \text{ infinitely many times}$$

Therefore, $\lim\limits_{n\to\infty} \sup A_n = \{ s |\, s \in \text{infinitely many } A_n \}$

Definition: The *limit inferior of a sequence* of events. $A_1, A_2, \ldots$ in $\mathbb{F}$ is defined as

$$\lim_{n\to\infty}\inf A_n = \bigcup_{n=1}^{\infty}\bigcap_{n=n}^{\infty} A_k$$

Clearly,

$$s \in \lim_{n\to\infty}\inf A_n$$

$$\Leftrightarrow s \in \bigcup_{n=1}^{\infty}\bigcap_{n=n}^{\infty} A_k$$

$$\Leftrightarrow \exists n \geq 1 \text{ such that } \forall k \geq n, s \in A_k$$

$$\Leftrightarrow \exists n \geq 1 \text{ such that } \forall k \geq n, s \in A_k$$

$$\Leftrightarrow s \in A_n \text{ for all but finitely many } n$$

Theorem: $\lim\limits_{n\to\infty} \inf A_n \subseteq \lim\limits_{n\to\infty} \sup A_n$

Proof: Suppose $s \in \lim\limits_{n\to\infty} \inf A_n$

Then $s \in A_n$ for all but finitely many n

$\Rightarrow$ s belongs to infinitely many of A_n

$\Rightarrow s \in \lim\limits_{n\to\infty} \sup A_n$

$\Rightarrow \lim\limits_{n\to\infty} \inf A_n \subseteq \lim\limits_{n\to\infty} \sup A_n$

The sequence $\left\{A_n\right\}_{n=1}^{\infty}$ is *said to converge to the limit A* if $\lim\limits_{n\to\infty} \inf A_n = \lim\limits_{n\to\infty} \sup A_n = A$

Suppose $B_n = \bigcup\limits_{k=n}^{\infty} A_k$ and $C_n = \bigcap\limits_{k=n}^{\infty} A_k$. Observe that $\left\{B_n\right\}$ and $\left\{C_n\right\}$ monotone sequences. Note that $\left\{B_n\right\}$ is a monotonically decreasing sequence while $\left\{C_n\right\}$ is a monotonically increasing sequence.

Clearly,

$$\lim_{n\to\infty}\sup A_n = \bigcap_{n=1}^{\infty} B_n$$

and

$$\lim_{n\to\infty}\inf\ A_n = \bigcup_{n=1}^{\infty} C_n\ .$$

Example 1

Suppose the sequence of events with $\{A_n\}$ is given by

$$A_n = \left(1+\frac{1}{n},\ \ 2+\frac{1}{n}\right)$$

$$\lim_{n\to\infty}\ \sup A_n = \bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty} A_j$$

$$= \bigcap_{n=1}^{\infty} B_n \text{ , where } B_n = \bigcup_{j=n}^{\infty} A_j$$

$$B_n = \bigcup_{j=n}^{\infty}\left(1+\frac{1}{j},\ \ 2+\frac{1}{j}\right) = \left(1,\ \ 2+\frac{1}{n}\right)$$

$$\therefore \lim_{n\to\infty}\ \sup A_n = \bigcap_{n=1}^{\infty}\left(1,\ \ 2+\frac{1}{n}\right)$$

$$= (1,\ 2]$$

$$\lim_{n\to\infty}\ \inf A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k = \bigcup_{n=1}^{\infty} C_n \text{ , where}$$

$$C_n = \bigcap_{j=n}^{\infty} A_j = \bigcap_{j=n}^{\infty}\left(1+\frac{1}{j},\ \ 2+\frac{1}{j}\right)$$

$$= \left(1+\frac{1}{n},\ 2\right]$$

$$\therefore \lim_{n\to\infty}\ \inf A_n = \bigcup_{n=1}^{\infty}\left(1+\frac{1}{n},\ \ 2\right]$$

$$= [1, 2]$$

$$\boxed{\therefore \lim_{n\to\infty}\ \inf A_n = \lim_{n\to\infty}\ \sup A_n}$$

Example 2

$$A_n = \begin{cases} \left(1-\frac{1}{n},\ \ 3+\frac{1}{n}\right) & n \text{ odd} \\ \left(2-\frac{1}{n},\ \ 5+\frac{1}{n}\right) & n \text{ even} \end{cases}$$

Then,

$$\bigcup_{j=n}^{\infty} A_j = \begin{cases} \left(1-\frac{1}{n},\ 5+\frac{1}{n+1}\right) & n \text{ odd} \\ \left(1-\frac{1}{n+1},\ 5+\frac{1}{n}\right) & n \text{ even} \end{cases}$$

$$\begin{aligned} \therefore \lim_{n\to\infty} \sup A_n &= \bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty} A_j \\ &= [1,\ 5] \end{aligned}$$

Again,

$$\bigcap_{j=n}^{\infty} A_j = [2,\ \ 3]$$

$$\therefore \lim_{n\to\infty} \inf A_n = [2,\ \ 3]$$

We can apply the continuity theorem to find the probabilities $P(\lim_{n\to\infty}\sup A_n)$ and $P(\lim_{n\to\infty}\inf A_n)$.

Thus

$$\begin{aligned} P(\lim_{n\to\infty}\sup A_n) &= P(\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty} A_j) \\ &= \lim_{n\to\infty} P(\bigcup_{j=n}^{\infty} A_j) \end{aligned}$$

and

$$\begin{aligned} P(\lim_{n\to\infty}\inf A_n) &= P\left(\bigcup_{n=1}^{\infty}\bigcap_{n=n}^{\infty} A_k\right) \\ &= \lim_{n\to\infty} P\left(\bigcap_{n=n}^{\infty} A_k\right) \end{aligned}$$

These probabilities will be used to determine the limiting probabilities involving a sequence of random variables. Particularly, the following theorems are important.

Theorem $P\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} P(A_k)$

Proof:

Introduce the event $\qquad B_n = \bigcup_{k=1}^{n} A_k$

Clearly $\{B_n\}$ is an increasing sequence of events

Also $\quad \bigcup_{k=1}^{n} A_k = \bigcup_{k=1}^{n} B_k$

Now

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^{\infty} B_k\right)$$

$$= \lim_{n\to\infty} P(B_n) \qquad \because \{\mathrm{B}_n\} \text{ is increasing}$$

$$= \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} A_k\right)$$

$$\le \lim_{n\to\infty} \sum_{k=1}^{n} P(A_k)$$

$$\le \sum_{k=1}^{\infty} P(A_k)$$

So $P\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} P(A_k)$

Remark

We can similarly show that

$$P\left(\bigcup_{k=n}^{\infty} A_k\right) \le \sum_{k=n}^{\infty} P(A_k)$$

Theorem

$$P(\liminf_{n\to\infty} A_n) \le \liminf_{n\to\infty} P(A_n) \le \limsup_{n\to\infty} P(A_n) \le P(\limsup_{n\to\infty} A_n)$$

By continuity theorem.

$$P(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} P(\bigcup_{j=n}^{\infty} A_j)$$

Note that the right hand side of above is a limit of a real sequence.

Also, $\mathrm{P}(A_n) \le P(\bigcup_{j=n}^{\infty} A_j) \qquad (\because A_n \subseteq \bigcup_{j=n}^{\infty} A_j))$

Using the property of the limit inferior and the limit superior of a real sequence, we get

$$\liminf_{n\to\infty} P(A_n) \le \limsup_{n\to\infty} P(A_n) \le P(\limsup_{n\to\infty} A_n)$$

Again

$$P(\liminf_{n\to\infty} A_n) = \lim_{n\to\infty} P(\bigcap_{j=n}^{\infty} A_j)$$

and

$$P(\bigcap_{j=n}^{\infty} A_j) \le P(A_n)$$

$$\therefore P(\liminf_{n\to\infty} A_n) = \lim_{n\to\infty} P(\bigcap_{j=n}^{\infty} A_j) \le \liminf_{n\to\infty} P(A_n)$$

Thus

$$P(\liminf_{n\to\infty} A_n) \le \liminf_{n\to\infty} P(A_n) \le \limsup_{n\to\infty} P(A_n) \le P(\limsup_{n\to\infty} A_n)$$

Borel-Cantelli Lemma

Recall that the probability is a continuous set function and we have established that

$$P\left(\lim_{n\to\infty}\sup A_n\right) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\right)$$

$$= \lim_{n\to\infty} P\left(\bigcup_{k=n}^{\infty} A_k\right)$$

Theorem: Let $\{A_n\}$ be a sequence of events in $(S,\mathbb{F},P)$.

(a) If $\sum_{n=1}^{\infty} P(A_n) < \infty$ then

$$P\left(\lim_{n\to\infty}\sup A_n\right) = P\left(A_n i.o\right) = 0$$

(b) If the events in $\{A_n\}$ are mutually independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$, then

$$P\left(\lim_{n\to\infty}\sup A_n\right) = P\left(A_n i.o\right) = 1$$

Proof (a): $P\left(\lim_{n\to\infty}\sup A_n\right) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty} A_j\right) = \lim_{n\to\infty} P\left(\bigcup_{j=n}^{\infty} A_j\right) \le \lim_{n\to\infty}\sum_{j=n}^{\infty} P(A_j) = 0$

whenever $\sum_{n=1}^{\infty} P(A_n) < \infty$

Proof (b): Suppose $A = \lim_{n\to\infty}\sup A_n = \bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty} A_j$

Then, $A^C = \left(\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty} A_k\right)^C = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_j^C$

By the continuity theorem,

$$P\left(A^C\right) = P\left(\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty} A_j^C\right)$$

$$= \lim_{n\to\infty} P\left(\bigcap_{j=n}^{\infty} A_j^C\right)$$

$$= \lim_{n\to\infty}\prod_{j=n} P(A_j^C)$$

$$= \lim_{n\to\infty}\prod_{k=n}(1 - P(A_j)$$

$$\le \prod_{j=n} \mathrm{e}^{-P(A_j)} \qquad (\because 1 - x \le e^{-x}$$

$$= \mathrm{e}^{-\sum_{j=n}^{\infty} P(A_j)} = 0$$

Because $\sum_{j=n}^{\infty} P(A_j)=\infty$ whenever $\sum_{j=1}^{\infty} P(A_j)=\infty$.

$\therefore P\left(A\right) = 1$

Convergence of a Sequence of Random Variables

The convergence of a random sequence $\{X_n\}$ cannot be defined as above. Note that for each $s \in S$, $X_1(s), X_2(s), \cdots, X_n(s), \cdots$ represent a sequence of numbers. Thus $\{X_n\}$ represents a family of sequences of numbers. Convergence of a random sequence is to be defined using different criteria.

There are several criteria for defining the convergence of a random sequence. Five of these criteria are explained below.

Convergence Everywhere

A sequence of random variables, $\{X_n\}$ is said to converge everywhere to X if

$|X(s) - X_n(s)| \to 0$ for $n > N$ and $\forall s \in S$.

This case is the simple extension of convergence of a sequence of real numbers and the sequence of numbers for each sample point is convergent.

Almost sure (a.s.) convergence or convergence with probability 1

A random sequence $\{X_n\}$ may not converge for every $s \in S$.

Consider the event $\{s \mid X_n(s) \to X\}$ defined on the sample space S. The probability of this event is determined to define the convergence with probability 1 or almost sure (a.s.) convergence.

Definition: The sequence $\{X_n\}$ is said to converge to X *almost sure* or *with probability 1* if $P(\{s \mid X_n(s) \to X(s)\}) = 1$ *as* $n \to \infty$, or equivalently, for every $\varepsilon > 0$, there exists N such that $P\{s \left| X_n(s) - X(s) \right| < \varepsilon \text{ for all } n \ge N \} = 1$. We write $\{X_n\} \xrightarrow{a.s.} X$ in this case

Example 1

Suppose $S = \{s_1, s_2, s_3\}$ and $\{X_n\}$ be a sequence of random variables with

$$X_n(s_1) = 1, X_n(s_2) = -1$$

and

$$X_n(s_3) = n$$

We see that $X_n(s_3)$ is a diverging sequence.

Define a random variable X suc h that

$$X(s_1) = 1, X(s_2) = -1$$

and

$$X(s_3) = 1$$

$$\therefore \{s \mid X_n(s) \to X(s)\} = \{s_1, s_2\}$$

$$\Rightarrow P(\{s \mid X_n(s) \to X(s)\}) = P(\{s_1, s_2\})$$

Therefore $\{X_n\} \xrightarrow{a.s.} X$ if $P(\{s_1, s_2\}) = 1$

Test for a.s. Convergence

Consider the convergence condition in terms of ε and N. For each , $\varepsilon > 0$, we can find a positive integer m such that $\frac{1}{m+1} \le \varepsilon \le \frac{1}{m}$. The condition for the convergence $X_n(s) \to X(s)$ can now be written in terms of m and N as follows:

The sequence $\{X_n(s)\}$ converges to $X(s)$ if for each positive integer $m > 0$, there exists a number N such that $\left| X_n(s) - X(s) \right| < \frac{1}{m}$ for $n > N$.

For each positive integer *m*, the event $D_m = \left\{ s \left\| X_n(s) - X(s) \right| \ge \frac{1}{m} \text{ for infinitely many times} \right\}$ is the set all sample points for which the corresponding sequence $\{X_n(s)\}$ diverges. Thus,

$$D_m = \limsup_{n \to \infty} \left\{ s \left\| X_n(s) - X(s) \right| \ge \frac{1}{m} \right\}$$

$$D_m = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \left\{ s \left\| X_n(s) - X(s) \right| \ge \frac{1}{m} \right\}$$

We have to consider each positive integer *m*, and therefore the event of sample points representing the divergent sequences is given by

$$D = \bigcup_{m=1}^{\infty} D_m = \bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} \left\{ s \left\| X_j(s) - X(s) \right| \ge \frac{1}{m} \right\}$$

Clearly, $D^c = \{s \mid X_n(s) \to X\}$

$\{X_n\} \xrightarrow{a.s.} X$ if $P(D^c) = 1$ or equivalently, $P(D) = P\left(\bigcup_{m=1}^{\infty} D_m\right) = 0$

Theorem: $\{X_n\} \xrightarrow{a.s} X$ if and only if $P(D_m) = 0$ for each $m \geq 1$.

Proof: Suppose $\{X_n\} \xrightarrow{a.s} X$

then $P\left(\bigcup_{m=1}^{\infty} D_m\right) = 0$

but $D_m \subseteq \bigcup_{i=1}^{\infty} D_i$

thus $P(D_m) \leq P\left(\bigcup_{i=1}^{\infty} D_i\right) = 0$

Next, if $P(D_m) = 0 \ \forall m \geq 1$, then

$$P\left(\bigcup_{m=1}^{\infty} D_m\right) \leq \sum_{m=1}^{\infty} P(D_m) = 0$$

The above theorem gives a necessary and sufficient condition for a.s. convergence. The following theorem, which is a consequence of the first Borel-Cantelli lemma, gives a sufficient test for a.s. convergence.

Theorem: If for each $m \geq 1$, $\sum_{n=1}^{\infty} P\left(|X_n - X| \geq \frac{1}{m}\right) < \infty$, then $\{X_n\} \xrightarrow{a.s.} X$

Proof:

Suppose for each $m \geq 1$,

$$\sum_{n=1}^{\infty} P\left(|X_n - X| \geq \frac{1}{m}\right) < \infty$$

$$\Rightarrow P\left(\limsup_{n\to\infty}\left\{s \middle| |X_n(s) - X| > \frac{1}{m}\right\}\right) = 0 \quad \text{(Using Borel Cantelli Lemma 1)}$$

$$\Rightarrow P(D_m) = 0$$

$$\Rightarrow \{X_n\} \xrightarrow{a.s} X$$

Example: Suppose the random sequence $\{X_n\}$ is given by

$$X_n = \begin{cases} 0 & \text{with probability } 1 - \frac{1}{2^n} \\ n & \text{with probability } \frac{1}{2^n} \end{cases}$$

Then,

$$P\left(\left\{|X_n - 0| \geq \frac{1}{m}\right\}\right) = P(X_n = n) = \frac{1}{2^n},$$

$$\therefore \sum_{n=1}^{\infty} P\left(\left\{|X_n - 0| \geq \frac{1}{m}\right\}\right) = \sum_{n=1}^{\infty} \frac{1}{2^n}$$

$$= \frac{\frac{1}{2}}{1-\frac{1}{2}}$$

$$= 1 < \infty$$

$$\therefore \{X_n\} \xrightarrow{as} \{X = 0\}$$

Convergence in mean square sense

Definition A random sequence $\{X_n\}$ is said to converge in the mean-square sense (m.s) to a random variable X if $E(X_n - X)^2 \to 0 \quad as \quad n \to \infty$

X is called the mean-square limit of the sequence and we write

$$l.i.m.\ X_n = X$$

where $l.i.m.$ means limit in mean-square.

We also write $X_n \xrightarrow{m.s.} X$

Theorem: If $\{X_n\} \xrightarrow{m.s.} X$, then as n $\to \infty, E(X_n) \to E(X)$ and $\mathrm{E}(X_n)^2 \to E(X)^2$

We have, $|E(X_n) - E(X)| = |E(X_n - X)|$

$$= \left|\sqrt{E^2(X_n - X)}\right|$$

$$\leq \sqrt{E(X_n - X)^2}$$

$$\therefore \lim_{n\to\infty} |E(X_n) - E(X)| = 0$$

Again

$$\left|E(X_n)^2 - E(X)^2\right| = \left|E\left(X^2{}_n - X^2\right)\right|$$

$$= \left|E(X_n - X)(X_n + X)\right| \qquad \text{(Using Cauchy Schwartz}$$

$$\leq \sqrt{E(X_n - X)^2 E(X_n + X)^2}$$

Inequality)

$$\therefore \lim_{n\to\infty} \left|E\left(X_n{}^2\right) - E\left(X^2\right)\right| = 0$$

$$\therefore E\left(X_n^2\right) \to E\left(X^2\right)$$

Remark

Theorem: $\{X_n\} \xrightarrow{m..s.} X \Rightarrow \{X_n\} \xrightarrow{P} X$

Proof: Suppose $\{X_n\} \xrightarrow{m.s.} X$

Then $\lim_{n\to\infty} E(X_n - X)^2 = 0$

Now, for $\varepsilon > 0$,

$$P\left(\left|X_n - X\right| > \varepsilon\right) = P\left(\left|X_n - X\right|^2 > \varepsilon^2\right)$$

$$\le E(X_n - X)^2 / \varepsilon^2 \qquad \text{(Using Markov Inequality)}$$

$$\therefore \lim_{n\to\infty} P\left(\left|X_n - X\right| > \varepsilon\right) \le \lim_{n\to\infty} E(X_n - X)^2 / \varepsilon^2$$

$$= 0$$

$$\therefore \{X_n\} \xrightarrow{p} X$$

The converse is not generally true. $\{X_n\} \xrightarrow{P} \{X\}$ does not necessarily imly $\{X_n\} \xrightarrow{m.s.} X$

- The following **Cauchy criterion** gives the condition for m.s. convergence of a random sequence without actually finding the limit. The sequence $\{X_n\}$ converges in m.s. if and only if, for every $\varepsilon > 0$ there exists a positive integer N such that

  $E\left(\left|x_{n+m} - x_n\right|^2\right) \to 0$ as $n \to \infty$ for all $m > 0$.

Convergence in probability 11

Associated with the sequence of random variables $\{X_n\}$ we can define a sequence of probabilities $P\left(\left|X_n - X\right| > \varepsilon\right), n = 1, 2, \ldots$ for every $\varepsilon > 0$. The sequence $\{X_n\}$ is said to convergent to X in probability if this sequence of probability is convergent that is

$$P\left(\left|X_n - X\right| > \varepsilon\right) \to 0 \quad as \quad n \to \infty \text{ for every } \varepsilon > 0.$$

We write $X_n \xrightarrow{P} X$ to denote convergence in probability of the sequence of random variables to the random variable X.

Example 2

Suppose $\{X_n\}$ be a sequence of random variables with

$$P\{X_n = 0\} = 1 - \frac{1}{n^2}$$

and

$$P\{X_n = n\} = \frac{1}{n^2}$$

Clearly

$$P\{|X_n - 0| > \varepsilon\} = P\{X_n = n\} = \frac{1}{n^2} \to 0$$
$$as\ n \to \infty.$$

Therefore $\{X_n\} \xrightarrow{P} \{X = 0\}$. Thus the above sequence converges to a constant in probability.

Theorem: $\{X_n\} \xrightarrow{a.s} X \Rightarrow \{X_n\} \xrightarrow{P} X$

Proof: Suppose $\{X_n\} \xrightarrow{a.s} X$

Then for any $m \geq 1$

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty}\left\{s \left| \left|X_j(s) - X(s)\right| \geq \frac{1}{m}\right.\right\}\right) = 0$$

$$\begin{aligned}\therefore \lim_{n\to\infty} P\left(\left\{s \left| \left|X_n(s) - X(s)\right| \geq \frac{1}{m}\right.\right\}\right) &\leq \lim_{n\to\infty} P\left(\bigcup_{j=n}^{\infty}\left\{s \left| \left|X_j(s) - X(s)\right| \geq \frac{1}{m}\right.\right\}\right)\\ &= P\left(\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty}\left\{s \left| \left|X_j(s) - X(s)\right| \geq \frac{1}{m}\right.\right\}\right)\\ &= 0\end{aligned}$$

In the above, we get the expression in second line by applying the Continuity Theorem for a sequence of decreasing events.

$\therefore \{X_n\} \xrightarrow{p} X$

The converse of the theorem is not true as illustrated by the following example.

<u>Example 3</u>

Suppose $\{X_n\}$ be a sequence of independent random variables defined by

$$X_n = \begin{cases} 1 \text{ with probability } \dfrac{1}{n} \\ 0 \text{ with probability } 1\text{-}\dfrac{1}{n} \end{cases}$$

Then for any $m \geq 1$,

$$P\left(\left\{|X_n - 0| \geq \frac{1}{m}\right\}\right) = P\left(\{X_n = 1\}\right) = \frac{1}{n}$$

$$\therefore \lim_{n\to\infty} P\left(\left\{|X_n - 0| \geq \frac{1}{m}\right\}\right) = 0$$

so that $\{X_n\} \xrightarrow{P} \{X = 0\}$

But $\sum_{n=1}^{\infty} P\left(\left\{|X_n - X| \geq \frac{1}{m}\right\}\right) = \sum \frac{1}{n} \to \infty$

Then events $\{X_n = 1\}$,n=1,2,… are independent

Therefore, by Borel Cantelli second lemma,

$$P\left(\bigcap_{n=1}^{\infty}\bigcup_{j=n}^{\infty}\left\{s\left||X_j(s)-X(s)\right|\geq\frac{1}{m}\right\}\right)=1$$

$$\therefore X_n \overset{a.s.}{\not\longrightarrow} \{X=0\}$$

Theorem: $\{X_n\}\xrightarrow{m..s.}X \Rightarrow \{X_n\}\xrightarrow{P}X$

Proof: Suppose $\{X_n\}\xrightarrow{m.s.}X$

Then $\lim_{n\to\infty} E(X_n - X)^2 = 0$

Now, for $\varepsilon > 0$,

$$P\left(|X_n - X| > \varepsilon\right) = P\left(|X_n - X|^2 > \varepsilon^2\right)$$

$$\leq E(X_n - X)^2 / \varepsilon^2 \qquad \text{(Using Markov Inequality)}$$

$$\therefore \lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) \leq \lim_{n\to\infty} E(X_n - X)^2 / \varepsilon^2$$

$$= 0$$

$$\therefore \{X_n\}\xrightarrow{p}X$$

The converse is not generally true. $\{X_n\}\xrightarrow{p}X$ does not necessarily imply $\{X_n\}\xrightarrow{m.s.}X$

Example 4 Consider the random sequence in **Example 2**.

Here

$$E(X_n - X)^2 = E(X_n - 0)^2$$

$$= EX_n^2$$

$$= 0\times(1-\frac{1}{n^2}) + n^2\times\frac{1}{n^2}$$

$$= 1$$

$$\therefore \{X_n\} \overset{ms}{\not\longrightarrow} X$$

Convergence in distribution

Consider the random sequence $\{X_n\}$ and a random variable X. Suppose $F_{X_n}(x)$ and $F_X(x)$ are the distribution functions of X_n and X respectively. The sequence is said to converge to X in distribution if $F_{X_n}(x) \to F_X(x) \quad as \quad n\to\infty.$ for all x at which $F_X(x)$ is continuous. Here the two distribution functions eventually coincide. We write $X_n \xrightarrow{d} X$ to denote convergence in distribution of the random sequence $\{X_n\}$ to the random variable X.

Example 3

Suppose $\{X_n\}$ is a sequence of independent RVs with each random variable X_i having the uniform density

$$f_{X_i}(x) = \begin{cases} \frac{1}{a} & 0 \le x \le a \\ 0 & \text{other wise} \end{cases}$$

Define $Z_n = \max(X_1, X_2, \ldots. X_n)$

We can show that

$$F_{Z_n}(z) = \begin{cases} 0, & z < 0 \\ \frac{z^n}{a^n}, & 0 \le z < a \\ 1 & \text{otherwise} \end{cases}$$

Clearly,

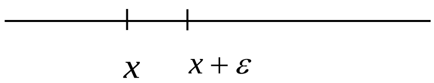
$$\lim_{n\to\infty} F_{Z_n}(z) = F_Z(z) = \begin{cases} 0, & z < a \\ 1 & z \ge a \end{cases}$$

$\therefore \{Z_n\}$ converges to $Z = a$ in *distribution.*

Theorem: $\{X_n\} \xrightarrow{P} X \Rightarrow \{X_n\} \xrightarrow{d} X$

Proof: $\{X_n\} \xrightarrow{P} X$

$$\Rightarrow \lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0 \qquad \text{for } \varepsilon > 0$$



Now consider the subsets $\{X \le x+\varepsilon\}$ and $\{X > x+\varepsilon\}$ partitioning the real line at $x+\varepsilon$.

Then,

$$\{X_n \le x\} = \{X_n \le x, X \le x+\varepsilon\} \bigcup \{X_n \le x, X > x+\varepsilon\}$$

$$\begin{aligned} \therefore \quad F_{X_n}(x) &= P(\{X_n \le x\}) \\ &= P\{X_n \le x, X \le x+\varepsilon\} + P\{X_n \le x, X > x+\varepsilon\} \\ &\le P(\{X \le x+\varepsilon\}) + P(\{X_n \le x, X > x+\varepsilon\}) \\ &\le P(\{X \le x+\varepsilon\}) + P(\{|X_n - X| > \varepsilon\}) \\ &= F_X(x+\varepsilon) + P(\{|X_n - X| > \varepsilon\}) \end{aligned}$$

Now $\lim_{n\to\infty} P(\{|X_n - X| > \varepsilon\}) = 0$

$$\therefore \lim_{n\to\infty} F_{X_n}(x) \le F_X(x+\varepsilon) \qquad (1)$$

Similarly by changing the role of X_n and X we can show that

$$F_X(x-\varepsilon) \le \lim_{n\to\infty} F_{X_n}(x) \qquad (2)$$

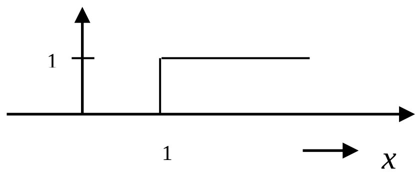
From (1) and (2), we get

$$F_X(x-\varepsilon) \leq \lim_{n\to\infty} F_{X_n}(x) \leq F_X(x+\varepsilon)$$
$$\therefore \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

The converse of the theorem is not true. For this consider a sequence of RVs defined on $S = \{H, T\}$ by $X_n(H) = 1$, $X_n(T) = 0$. Assume $P(\{T\}) = P(\{H\}) = \frac{1}{2}$

Suppose X is another RV defined on the same sample space by $X(H) = 0$, $X(T) = 1$.

Clearly $F_{X_n}(x) \to F_X(x) \qquad \forall x$



$F_{Xn}(x) = F_X(x)$

Now for any $0 < \varepsilon < 1$,

$\{s \,|\, |X_n - X| > \varepsilon\}$

$= \{H, T\}$

$\therefore P(\{s \,|\, |X_n - X| > \varepsilon\}) = 1$

$\therefore \{X_n\} \not\xrightarrow{P} X$.

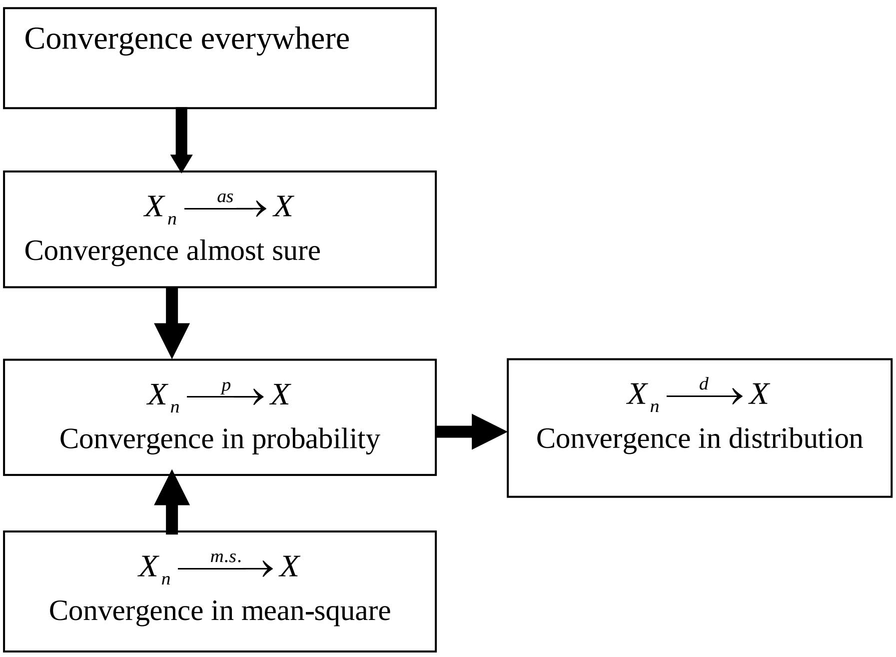
We state without proof the following theorem:

Continuity theorem of Convergence

Suppose $\psi_X(\omega) = Ee^{j\omega X}$ and $\psi_{X_n}(\omega) = Ee^{j\omega X_n}$

$\lim_{n\to\infty} F_{X_n}(x) \to F_X(x)$ iff $\lim_{n\to\infty} \psi_{X_n}(\omega) = \psi_X(\omega)$ for each ω and $\psi_X(\omega)$ is continuous at $\omega = 0$.

Relation between Types of Convergence



Laws of Large numbers

Consider a sequence of random variables $\{X_n\}$ with a common mean μ. It is common practice to determine μ on the basis of the sample mean defined by the relation

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

where $S_n = \sum_{i=1}^{n} X_i$. We assert that $\frac{S_n}{n} \to \mu$ as $n \to \infty$. Note that μ is a deterministic constant whereas $\frac{S_n}{n}$ is a function of n random variables. The laws of large numbers are the basis of such assertions.

More generally, suppose $\{X_n\}$ is a sequence of random variables with $\mu_i = EX_i,\ i = 1, 2, ..., n$.

Then,

$$E\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} EX_i$$
$$= \frac{1}{n}\sum_{i=1}^{n} \mu_i$$

The sequence $\{X_n\}$ is said to obey the *strong law of large numbers* if

$$\frac{S_n}{n} \xrightarrow{a.s.} \frac{1}{n}\sum_{i=1}^{n} \mu_i .$$

Similarly, $\{X_n\}$ is said to obey the *weak law of large numbers* if

$$\frac{S_n}{n} \xrightarrow{P.} \frac{1}{n}\sum_{i=1}^{n} \mu_i .$$

We consider weak law for a more general case and the strong law for the special case when $\{X_n\}$ is a sequence of iid random variables.

Theorem 1 Weak law of large numbers(WLLN): Suppose $\{X_n\}$ is a sequence of random variables defined on a probability space $(S, \mathbb{F}, P)$ with finite mean $\mu_i = EX_i,\ i = 1, 2, ..., n$ and finite second moments. If

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} cov(X_i, X_j) = 0,$$

then $\frac{S_n}{n} \xrightarrow{P.} \frac{1}{n}\sum_{i=1}^{n} \mu_i$.

Proof: We have

$$E(\frac{S_n}{n}-\frac{1}{n}\sum_{i=1}^{n}\mu_i)^2 = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}\mu_i\right)^2$$

$$=\frac{1}{n^2}E\left(\sum_{i=1}^{n}(X_i-\mu_{i_i})\right)^2$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}E(X_i-\mu_{i_i})^2+\frac{1}{n^2}\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{j=1,j\neq i}}^{\mathrm{n}}E(X_i-\mu_{i_i})(X_j-\mu_{j_j})$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\sigma_{i_i}^2+\frac{1}{n^2}\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{j=1,j\neq i}}^{\mathrm{n}}cov(X_i,X_j)$$

$$\therefore \lim_{n\to\infty}E(\frac{S_n}{n}-\frac{1}{n}\sum_{i=1}^{n}\mu_i)^2=\lim_{n\to\infty}\left(\frac{1}{n^2}\sum_{i=1}^{n}\sigma_{i_i}^2+\frac{1}{n^2}\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{j=1,j\neq i}}^{\mathrm{n}}cov(X_i,X_j)\right)$$

Now $\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sigma_{i_i}^2=0$, as each $\sigma_{i_i}^2$ is finite. Also,

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{\mathrm{i=1}}^{\mathrm{n}}\sum_{\mathrm{j=1,j\neq i}}^{\mathrm{n}}cov(X_i,X_j)=0$$

$$\lim_{n\to\infty}E(\frac{S_n}{n}-\frac{1}{n}\sum_{i=1}^{n}\mu_i)^2=0$$

$$\Rightarrow \frac{S_n}{n}\xrightarrow{m.s.}\frac{1}{n}\sum_{i=1}^{n}\mu_i$$

$$\Rightarrow \frac{S_n}{n}\xrightarrow{P.}\frac{1}{n}\sum_{i=1}^{n}\mu_i$$

Special Case of the WLLN

(a) Suppose $\{X_n\}$ is a sequence of independent and identically distributed random variables defined on a probability space $(S,\mathbb{F},P)$

Then we have

$EX_i=\text{constant}=\mu(say)$

$\mathrm{var}(X_i)=\text{constant}=\sigma^2(say)$ *and*

$\mathrm{cov}(X_i,X_j)=0$

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sigma_{i_i}^2=\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=0$$

$$\therefore \frac{S_n}{n}\xrightarrow{P.}\mu$$

(b) Suppose $\{X_n\}$ is a sequence of independent random variables defined on a probability space $(S,\mathbb{F},P)$ with the mean $\mu_i=EX_i,\ i=1,2,...,n$ and finite second moments .

Then we have

X_i and X_j are independent

$\therefore \operatorname{cov}(X_i, X_j) = 0$

Again

each $\sigma_{i_i}^2$ is finite

$\therefore \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma_{i_i}^2 = 0$

$\therefore \frac{S_n}{n} \xrightarrow{P.} \frac{1}{n} \sum_{i=1}^{n} \mu_i$

(c) Suppose $\{X_n\}$ is a sequence of uncorrelated random variables with the mean $\mu_i = EX_i$, $i = 1, 2, ..., n$ and finite second moments defined on a probability space $(S, \mathbb{F}, P)$.

Then we have $\operatorname{cov}(X_i, X_j) = 0$ by definition.

We can proceed as in case (b) to establish the result

$\frac{S_n}{n} \xrightarrow{P.} \frac{1}{n} \sum_{i=1}^{n} \mu_i$

The strong law of large number is based on the concept of almost sure convergence and stated in the following theorem:

Interpretation of relative frequency definition of probability

The relative frequency definition of probability can be interpreted using the *weak law of large number*. Suppose an experiment is repeated n times and a particular event A occurs n_A times. During these repetitions, then,

$$\frac{n_A}{n} \xrightarrow{P} P(A)$$

which is the interpretation.

To prove the above result consider a sequence of random variables $\{I_{A_n}\}$ given by

$$I_{A_n} = 1 \quad \text{if A occurs in } n^{th} \text{ trial}$$
$$=0 \quad \text{otherwise}$$

Then,

$$EI_{A_n} = 1 \times P(A) + 0 \times (1 - P(A))$$
$$= P(A)$$

Now,

$$\frac{1}{n}\sum_{i=1}^{n} I_{A_i} = \frac{\text{Number of occurences of } A}{n}$$
$$= \frac{n_A}{n}$$

So, using the *weak law of large number*,

$$\frac{n_A}{n} \xrightarrow{P} P(A)$$

Strong Law of Large Numbers

One of the important applications of the a.s. convergence is the *Strong Law of Large Numbers*. The Kolmogorov's strong law of large numbers is stated in the following theorem.

Theorm 1: Suppose $\{X_n\}$ is a sequence of iid random variables defined on a probability space $(S, \mathbb{F}, P)$ with common mean μ and $S_n = \sum_{i=1}^{n} X_i$. Then

$$\frac{S_n}{n} \xrightarrow{a.s} \mu$$

Though the theorem is general, we will prove the following restricted version:

Theorem 2: Suppose $\{X_n\}_{n=1}^{\infty}$ is a sequence of iid random variables defined on a probability space $(S, \mathbb{F}, P)$ with common mean μ and finite fourth central moment ($E(X_n - EX_n)^4 < \infty$). Then

$$\frac{S_n}{n} \xrightarrow{a.s} \mu$$

Proof: We have to prove that

$$P\left(\limsup_{n\to\infty}\left\{s\left|\left|\frac{S_n}{n} - \mu\right| \geq \frac{1}{m}\right.\right\}\right) = 0 \qquad \forall \text{ positive integer } m.$$

Let us examine the fourth moment

$$E\left(\frac{S_n}{n} - \mu\right)^4$$

$$= E\left(\frac{\sum_{i=1}^{n}(X_i - \mu)}{n}\right)^4$$

$$= E\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right)^4$$

where $Y_i = X_i - \mu$

In the expansion of $\left(\sum_{i=1}^{n} Y_i\right)^4$, there will be terms of the form $Y_i^4, Y_i^3 Y_j, Y_i^2 Y_j^2, Y_i Y_j Y_k^2, Y_i Y_j Y_k Y_l$.

We note that $EY_i^3 Y_j = EY_i Y_j Y_k^2 = EY_i Y_j Y_k Y_l = 0$ as Y_i is of zero mean and the sequence is independent. Therefore, the term of the form EY_i^4 and $EY_i^2 EY_j^2$ contribute to the fourth central moment. There are n terms of the form EY_i^4 and $^nC_2 \times {}^4C_2 = \frac{n(n-1)}{2} \times 6 = 3n(n-1)$ terms of the form $EY_i^2 EY_j^2$.

$$\therefore E\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right)^4 = \frac{1}{n^4}\left[nEY_i^4 + {}^nC_2 \times {}^4C_2\left(EY_i\right)^2\right]$$

$$= \frac{EY_i^4}{n^3} + \frac{3n(n-1)}{n^4}\left(EY_i^2\right)^2$$

$$\leq \frac{K}{n^3} + \frac{3}{n^2}K$$

(Assume $EY_i^4 = K < \infty$ and $(EY_i^2)^2 \leq EY_i^4 = K$)

$$\therefore E\left(\frac{\sum_{i=1}^{n}(X_i - \mu)}{n}\right)^4 \leq \frac{K}{n^3} + \frac{3K}{n^2}$$

Now $P\left(\left\{s \middle| \left|\frac{1}{n}\sum_{i=1}^{n} X_i(s) - \mu\right| \geq \frac{1}{m}\right\}\right)$

$$= P\left(\left\{s \middle| \left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i(s) - \mu\right)\right)^4 \geq \frac{1}{m^4}\right\}\right)$$

$$\leq \frac{E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^4}{\frac{1}{m^4}}$$ (By Markov inequality)

$$\leq \frac{m^4 K}{n^3} + \frac{3m^4}{n^2}K$$

$$\therefore \sum_{n=1}^{\infty} P\left(\left\{s \middle| \left|\frac{1}{n}\sum_{i=1}^{n} X_i(s) - \mu\right| \geq \frac{1}{m}\right\}\right)$$

$$\leq \sum_{n=1}^{\infty} \frac{m^4 K}{n^3} + \frac{3m^4}{n^2}K$$

$$< \infty$$

Hence according to the Borel Cantelli Lemma,

$$P\left(\limsup_{n\to\infty}\left\{s \middle| \left|\sum_{i=1}^{\infty}\left(X_i(s) - \mu\right)\right| \geq \frac{1}{m}\right\}\right) = 0$$

Thus, $\frac{S_n}{n} \xrightarrow{a.s} \mu$

Central Limit Theorem

Suppose $\{X_n\}$ is a sequence of independent and identically distributed random variables each with mean μ and variance σ^2 and $S_n = \sum_{i=1}^{n} X_i$ By the weak law of large numbers, $\frac{S_n}{n} \xrightarrow{P} \mu$. Note that the convergence in probability implies the convergence in distribution. Therefore,

$$\frac{S_n}{n} \xrightarrow{d} \mu$$

From the WLLN, we may conclude that for large *n,* $S_n \simeq n\mu$. The *central limit theorem* (CLT) gives the asymptotic distribution of the difference $S_n - n\mu$. The CLT is stated in terms of the standardized average $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Clearly, $\mathrm{E}Z_n = 0$ and $\mathrm{Var}(Z_n) = 1$

There are several special cases of the CLT. Here we state and prove the celebrated Lindeberg – Levy central limit theorem.

Theorem: Suppose $\{X_n\}$ is a sequence of i.i.d. random variables with mean μ and variance $\sigma^2 < \infty$. Let $S_n = \sum_{i=1}^{n} X_i$ and $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then $Z_n \xrightarrow{d} Z \sim N(0,1)$ in the sense that

$$\lim_{n\to\infty} F_{z_n}(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Proof: We shall prove the theorem using the characteristic function of Z_n and the continuity theorem of convergence.

Suppose $\{Z_n\}$ is a sequence of random variables with distribution function $\{F_{Z_n}\}$ and the corresponding characteristic function function $\phi_{Z_n}(\omega) = Ee^{j\omega Z_n}$. According to the continuity theorem of convergence $\lim_{n\to\infty} F_{Z_n}(z) = F_Z(z)$ if and only if $\lim_{n\to\infty} \phi_{Z_n}(\omega) = \phi_Z(\omega)$ for each ω and $\phi_Z(\omega)$ is continuous at $\omega = 0$.

Now,

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

$$\therefore \phi_{Z_n}(\omega) = \mathrm{E}e^{j\omega\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)}$$

$$= \mathrm{E}e^{j\omega\frac{\sum_{i=1}^{n}(X_i-\mu)}{\sigma\sqrt{n}}}$$

$$= \mathrm{E}\prod_{i=1}^{n} e^{\frac{j\omega(X_i-\mu)}{\sigma\sqrt{n}}} \qquad \text{(Using idependence of } X_i\text{s)}$$

$$= \prod_{i=1}^{n} Ee^{\frac{j\omega(X_i-\mu)}{\sigma\sqrt{n}}}$$

$$= \left(Ee^{\frac{j\omega(X_i-\mu)}{\sigma\sqrt{n}}}\right)^n$$

$$= \left(Ee^{\frac{j\omega Y_i}{\sqrt{n}}}\right)^n \qquad \text{where } \mathrm{Y}_i = \frac{X_i-\mu}{\sigma}$$

$$= \left[1 + \frac{j\omega}{\sigma\sqrt{n}} EY_i + \frac{(j\omega)^2}{2!n} EY_i^2 + o\left(\frac{\omega^2}{n}\right)\right]^n$$

$$= \left[1 - \frac{\omega^2}{2n} + o\left(\frac{\omega^2}{n}\right)\right]^n$$

$$\therefore \lim_{n\to\infty}\phi_{Z_n}(\omega) = \lim_{n\to\infty}\left[1 - \frac{\omega^2}{2n} + o\left(\frac{\omega^2}{n}\right)\right]^n$$

$$= \lim_{n\to\infty}\left[1 - \frac{\omega^2}{2n}\right]^n$$

$$= e^{-\frac{\omega^2}{2}}$$

which is the characteristic function of $Z \sim N(0,1)$

Applying continuity theorem of convergence in distribution,

$F_{Z_n}(z) \to F_Z(x)$

Suppose the MGF of the random variable in the sequence exists near s=0.

$$M_{Z_n}(s) = Ee^{s\frac{\sum_{i=1}^{n}(X_i-\mu)}{\sigma\sqrt{n}}}$$

$$= \prod_{i=1}^{n} M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)$$

$$= \left(M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)\right)^n$$

$$\ln M_{Z_n}(s) = n \ln M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)$$

$$= \frac{\ln M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)}{\frac{1}{n}}$$

$$\therefore \lim_{n\to\infty} \ln M_{Z_n}(s) = \lim_{n\to\infty} \frac{\ln M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)}{\frac{1}{n}}$$

Noting that $M_{Y_i}(0) = 1$ the condition for the L'Hospital rule applies. Applying the rule twice, we get

$$\lim_{n\to\infty} \ln M_{Z_n}(s) = \lim_{n\to\infty} \frac{\ln M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)}{\frac{1}{n}}$$

$$= \lim_{n\to\infty} \frac{\frac{M'_{Y_i}\left(\frac{s}{\sqrt{n}}\right)s}{\frac{1}{2}n^{-\frac{3}{2}}M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)}}{\frac{1}{n^2}}$$

$$= \lim_{n\to\infty} \frac{M'_{Y_i}\left(\frac{s}{\sqrt{n}}\right)s}{2n^{-\frac{1}{2}}M_{Y_i}\left(\frac{s}{\sqrt{n}}\right)}$$

$$= \lim_{n\to\infty} \frac{sM'_{Y_i}\left(\frac{s}{\sqrt{n}}\right)}{2n^{-\frac{1}{2}}}$$

$$= \lim_{n\to\infty} \frac{s^2 M''_{Y_i}\left(\frac{s}{\sqrt{n}}\right)}{2}$$

$$= \frac{s^2}{2}$$

Thus z_n is approximately Gaussian for large *n*. This in turn implies that $S_n \sim N(n\mu, n\sigma^2)$ for large *n*.

The CLT is true under more general conditions. The i.i.d. part of the Lindeberg – Levy theorem need not be satisfied. We state two of these conditions without proof:

- **Liapounov theorem:** Suppose $\{X_n\}_{n=1}^{\infty}$ is a sequence of independent random variables with mean $\mu_n = EX_n$ and variance $\sigma_n^2 = E(X_n - \mu_n)^2$ and $S_n = \sum_{i=1}^{n} X_i$.

  Clearly $\mu_{S_n} = \sum_{i=1}^{n} \mu_i$ and $\sigma_{S_n}^2 = \sum_{i=1}^{n} \sigma_i^2$. If for some $\delta > 0$,

  $$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} E(X_k - \mu_k)^{2+\delta}}{(\sigma_{S_n})^{2+\delta}} = 0,$$

  then $\dfrac{S_n - \mu_{S_n}}{\sigma_{S_n}} \xrightarrow{d} Z \sim N(0,1)$

- A corollary of the Liapounov theorem is as follows:

  Suppose each of X_ns are uniformly bounded, that is, $|X_n| < C, \ \forall n$ and $\lim_{n\to\infty} \sigma_{S_n}^2 \to \infty$. Then it is easy to verify that the conditions of the Liapounov theorem are satisfied. Thus $\dfrac{S_n - \mu_{S_n}}{\sigma_{S_n}} \xrightarrow{d} Z \sim N(0,1)$.

Remark

- The CLT states that the distribution function $F_{Z_n}(z)$ converges to a Gaussian distribution function. The theorem does not say that the pdf $f_{Z_n}(z)$ is a Gaussian pdf in the limit. For example, suppose each X_i has a Bernoulli distribution. Then the pdf of Z_n consists of impulses and can never approach the Gaussian pdf.

- The Cauchy distribution does not meet the conditions for the central limit theorem to hold. As we have noted earlier, this distribution does not have a finite mean or a variance.

Suppose a random variable X_i has the Cauchy distribution

$$f_{X_i}(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < \infty.$$

The characteristic function of X_i is given by

$$\phi_{X_i}(\omega) = e^{-|\omega|}$$

The sample mean S_n will have the characteristic function

$$\phi_{\mu_n}(\omega) = e^{-n|\omega|}$$

Thus the sum of large number of Cauchy random variables will not follow a Gaussian distribution.

The central-limit theorem is one of the most widely used results of probability.

- a random variable is result of several independent causes, then the random variable can be considered to be Gaussian. For example, the thermal noise in a resistor is the result of the independent motion of billions of electrons and is modeled as Gaussian.
- The observation error/ measurement error of any process is modeled as a Gaussian.
- The CLT can be used to simulate a Gaussian distribution given a routine to simulate a particular random variable

Normal approximation of the Binomial distribution

One of the applications of the CLT is in approximation of the Binomial coefficients.

Suppose $X_1, X_2, X_3, \ldots, X_n \ldots$ is a sequence of Bernoulli(p) random variables with $P(\{X_i = 1\}) = p$ and $P(\{X_i = 0\}) = 1 - p$.

Then $S_n = \sum_{i=1}^{n} X_i$ is a Binomial distribution with $\mu_{S_n} = np$ and $\sigma_{S_n}^2 = n\,p\,(1-p)$.

$$\text{Thus, } \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

$$\therefore \quad P\left(k-1 < S_n \le k\right) = \int_{k-1}^{k} \frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-\frac{1}{2}\cdot\frac{(y-np)^2}{np(1-p)}}\, dy$$

Assuming the integrand interval = 1,

$$P\left(S_n = k\right) \simeq \frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-\frac{1}{2}\cdot\frac{(y-np)^2}{np(1-p)}} \times 1$$

$$= \frac{1}{\sqrt{2\pi np(1-p)}}\, e^{-\frac{1}{2}\cdot\frac{(y-np)^2}{np(1-p)}}$$

This is normal approximation to the Binomial coefficients and is known as the ***DeMoirre-Laplace approximation.***