**Definition: Converge almost surely**

A sequence of random variables \( \{X_n\} \) is said to converge almost surely (a.s.) to a random variable \( X \), denoted by \( X_n \stackrel{a.s.}{\rightarrow} X \), if

\[
\lim_{n \to \infty} \Pr \left( X_n = X \right) = 1
\]

An equivalent way of defining almost sure convergence is by

\[
\lim_{n \to \infty} \Pr \left( |X_m - X| < \varepsilon, \text{ all } m \geq n \right) = 1
\]

The almost sure convergence is the mode of convergence associated with the strong law of large numbers (SLLN).

**Definition: Converge in r'th mean**

Let \( \{X_n\} \) be a sequence of random variables such that \( E(|X_n|^r) < \infty \) for all \( n \in \mathbb{N} \) and \( E(|X|^r) < \infty \) for \( r > 0 \), then the sequence converges to \( X \) in \( r \)'th mean, denoted by \( X_n \rightarrow X \), if

\[
\lim_{n \to \infty} E \left( |X_n - X|^r \right) = 0
\]

Of particular interest in what follows is the convergence in mean \( (r = 1) \) and mean square \( (r = 2) \).

**Definition: Converge in probability**

A sequence of random variables \( \{X_n\} \) is said to converge in probability to a random variable \( X \), denoted by \( X_n \rightarrow^p X \), if

\[
\lim_{n \to \infty} \Pr \left( |X_n - X| < \varepsilon \right) = 1
\]
**Definition: Converge in distribution**

A sequence of random variables \( \{X_n\} \) with distribution functions \( \{F_n(x)\} \) is said to converge in distribution to \( X \), denoted by \( X_n \xrightarrow{D} X \) if

\[
\lim_{n \to \infty} F_n(x) = F(x)
\]

at every continuity point \( x \) of \( F(x) \).

The mode of convergence related to the central limit theorem (CLT) is that of convergence in distribution.

Another way of looking at the convergence in probability is:

Consider the sequence of stochastic variables

\[ X_1, X_2, \ldots, X_n, \ldots \]

and the sequence of their distribution functions

\[ F_1, F_2, \ldots, F_n, \ldots \]

\( X_n \) converges to \( X \) in probability if

\[
\lim_{n \to \infty} P(\{|X_n - X| > \varepsilon\}) = 0 \quad \text{for any } \varepsilon > 0
\]

For a given \( \varepsilon \) and the distribution function \( F_n(\cdot) \) we can find the probability

\[ P(X_n < X - \varepsilon) + P(X_n > X + \varepsilon). \]

The sequence of these probabilities is

\[ a_1, a_2, \ldots, a_n, \ldots \]

Convergence in probability requires that this sequence has the limit 0.

Convergence in probability is stated as

\[ X_n \xrightarrow{P} X \text{ or } \plim X_n = X \]

**Lemma**

Let \( \{X_n\} \) be a random vector sequence and \( g(\cdot): \mathbb{R}^k \to \mathbb{R} \) a continuous function at \( X \), then

1) \[ X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X) \]
2) \( X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X) \)
3) \( X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X) \)

This lemma can be used to prove the following results:

\[
\text{plim}(X_n + Y_n) = \text{plim} X_n + \text{plim} Y_n = \mu + ?
\]

\[
\text{plim}(X_n \cdot Y_n) = \text{plim} X_n \cdot \text{plim} Y_n = \mu \cdot ?
\]

\[
\begin{align*}
\text{plim} \left( \frac{X_n}{Y_n} \right) &= \frac{\text{plim} X_n}{\text{plim} Y_n} = \frac{\mu}{\xi} \\
\text{plim} g(X_n) &= g(\text{plim} X_n) = g(\mu) \\
\text{plim} X_n^p &= (\text{plim} X_n)^p = \mu^p \\
\end{align*}
\]

\[
\begin{align*}
\text{plim} A_T^{-1} &= (\text{plim} A_T)^{-1} \\
\text{plim} A_T \cdot B_T &= \text{plim} A_T \cdot \text{plim} B_T
\end{align*}
\]

**Definition: Consistent estimator**

\( \hat{\theta}_n \) is a consistent estimator of \( \theta \) if

\[
\text{plim} \hat{\theta}_n = \theta
\]

**Theorem. Sufficient conditions for convergence in probability**

If

\[
\lim_{{n \to \infty}} E(\hat{\theta}_n) = \theta
\]

and

\[
\lim_{{n \to \infty}} V(\hat{\theta}_n) = 0
\]

then \( \text{plim} \hat{\theta} = \theta \).

Otherwise convergence in probability is often proved using Chesbychef’s inequality

\[
\Pr \left( \left| X - \mu \right| \geq k \cdot \sigma \right) \leq \frac{1}{k^2}
\]

Convergence in distribution can be proved by finding the limit of the sequence of characteristic functions

\[
f_n = E(\exp(i \cdot n \cdot X_n))
\]
\( \phi_n \to \phi \) for \( n \to \infty \Rightarrow F_n \stackrel{D}{\to} F \\

For simple cases this is not necessary.

**Example**

Let \( X_i \sim \text{nid}(\mu, s^2) \)

Let \( \bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \)

Then the sequence \( \{X_n\} \)

\( \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n, \ldots \)

has the following sequence of distribution functions

\( N(\mu, \frac{\sigma}{\sqrt{1}}), N(\mu, \frac{\sigma}{\sqrt{2}}), \ldots, N(\mu, \frac{\sigma}{\sqrt{n}}) \)

This sequence of distribution functions degenerates to the non-stochastic constant \( \mu \) because the variance goes to zero.

Instead we can look at \( \sqrt{n} \bar{X} \), or usually we look at

\( \sqrt{n} Z_n = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \)

\( \{ \sqrt{n} Z_n \} \) has the following sequence of distribution functions

\( N(0,1), N(0,1), \ldots, N(0,1), \ldots \)

which obviously has the limit \( N(0,1) \).

Otherwise one of the following Lemmas, the Cramer’s Theorem, or the Mann-Wald Theorem, can be used to find the limiting distribution function.

**Lemma**

Let \( \{X_n, Y_n\} \) be a sequence of pair of random k×1 vectors.

Then:

1. If \( \mathcal{L}(X_n - Y_n) \to 0 \) and \( X_n \rightarrow^D X \Rightarrow Y_n \rightarrow^D Y \)
2. If \( X_n \rightarrow^D X \) and \( Y_n \stackrel{p}{\to} 0 \Rightarrow X_n \cdot Y_n \stackrel{p}{\to} 0 \)
(3) If \( X_n \xrightarrow{d} X \) and \( Y_n \xrightarrow{p} C \) (constant) \( \Rightarrow (X_n + Y_n) \xrightarrow{d} X + C \)
If \( X_n \xrightarrow{d} X \) and \( Y_n \xrightarrow{p} C \) (constant) \( \Rightarrow Y_nX_n \rightarrow CX \)

(4) (for \( Y_n \) and \( C \) kxk non-singular) \( Y_n^{-1}X_n \xrightarrow{d} C^{-1} X \)

**Cramér's Theorem**

Let \( H_n \) be kxk matrix, such that

(i) \( \lim_{n \to \infty} H_n = H \)

where \( H \) is a kxk matrix with rank kxk.

Let \( \xi_n \) be an m element stochastic vector such that

(ii) \( \sqrt{n} (\xi_n - \xi) \xrightarrow{d} N(0, \Sigma) \)

where \( \xi \) is an m element non-stochastic vector, and \( \Sigma \) is an mxm positive semidefinite matrix.

Define the k element stochastic vector \( \eta_n = H_n \xi_n \). Then

(1) \( \sqrt{n} (\eta_n - H \xi) \xrightarrow{d} N(0, H \Sigma H') \)

**Mann-Wald Theorem**

Let \( Z_n \) be a k element stochastic vector such that

(i) \( \lim \frac{1}{n} \sum_{i=1}^{n} Z_i Z_i = Q \) or \( \lim \frac{Z'Z}{n} = Q \)

where \( Q \) is a kxk positive definite matrix.

Assume

(ii) \( u_i \sim iid(0,s^2) \)

(iii) \( E(Z_i u_i) = 0 \)

Then

(1) \( \lim \frac{1}{n} \sum_{i=1}^{n} Z_i u_i = 0 \) or \( \lim \frac{Z'u}{n} = 0 \)

and
Using a Taylor series approximation and Cramer's Theorem, the following theorem can be proved.

**Theorem. Asymptotic Distribution of a Nonlinear Function**

If \( \sqrt{n} \left( \hat{\theta} - \theta \right) \xrightarrow{D} N(0, \Sigma) \)

and if \( g(\theta) \) is a continuous function and its first and second order derivatives exist, then

\[
\sqrt{n} \left( g(\hat{\theta}) - g(\theta) \right) \xrightarrow{D} N(0, G(\theta) \Sigma G(\theta)')
\]

where \( G(\theta) = \begin{pmatrix} \frac{\partial g(\theta)}{\partial \theta} \end{pmatrix} \)

**Lemma. plim of sample moment**

Let \( m_k \) be a sample moment of \( \mu_k = E(x^k) \).
Assume that \( \mu_{2k} \) exists then \( \text{plim} \ m_k = \mu_k \).

**Lindberg-Feller CLT (Greene Thm 4.14)**

Suppose that \( x_1, x_2, x_3, \ldots, x_n \) are a sample of random vectors such that \( E[x_i] = \mu_i, \text{Var}[x_i] = Q_i \),

and all mixed third moments of the multivariate distribution are finite. Let

\[
\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} \mu_i, \\
\bar{Q}_n = \frac{1}{n} \sum_{i=1}^{n} Q_i.
\]

We assume that

\[
\lim_{n \to \infty} \bar{Q}_n = Q,
\]

where \( Q \) is a finite, positive definite matrix, and that for every \( i \),
Then

\[ \lim_{n \to \infty} \left( \frac{Q_n}{n} \right)^{-1} Q_i = 0. \]

\[ \sqrt{n} \left( \bar{x}_n - \mu_n \right) \xrightarrow{d} N(0, Q) \]