Convergence of Sequences of Random Variables

Let \( X_1, X_2, \ldots \) be a sequence of random variables and let \( X \) be another random variable. We want to make precise what we mean by \( X_n \to X \) as \( n \to \infty \). There are several different ways we define this convergence. Unless otherwise stated we assume that \( X \) and all the \( X_i \)'s are defined on the same probability space \((\Omega, \mathcal{F}, \mathbb{P})\).

Why is knowing that \( X_n \to X \) in some sense useful? If we know this, then \( X \) is somehow "close" to \( X_n \) for "large" \( n \) then exact statements we can make about \( X \) are approximately true for \( X_n \) for large \( n \).

Modes of Convergence

1. We say that \( \{X_n\} \) converges to \( X \) almost surely, written \( X_n \xrightarrow{a.s.} X \) or \( X_n \to X \) a.s., if
   \[
   \mathbb{P}\left( \omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty \right) = 1
   \]
   We also say that \( X_n \) converges to \( X \) with probability 1, or \( X_n \to X \) w.p. 1.
   This mode of convergence explicitly depends on the functional form of the random variables \( X \) and \( X_n \).

2. We say that \( \{X_n\} \) converges to \( X \) in probability, written \( X_n \xrightarrow{p} X \), if for any given \( \varepsilon > 0 \),
   \[
   \mathbb{P}\left( |X_n - X| > \varepsilon \right) \to 0 \text{ as } n \to \infty.
   \]
   This mode of convergence depends on the joint distribution of \( X_n \) and \( X \) for every \( n \).

3. We say that \( \{X_n\} \) converges to \( X \) in the \( r \)-th mean, where \( r > 0 \), written \( X_n \xrightarrow{\text{m}^r} X \), if
   \[
   \mathbb{E}[|X_n - X|^r] \to 0 \text{ as } n \to \infty.
   \]
   Again, this mode of convergence depends on the joint distribution of \( X \) and \( X_n \) for every \( n \). If \( r = 2 \), this is also called mean squared convergence.
We say that \( \{X_n\} \) converges to \( X \) in distribution, written \( X_n \xrightarrow{d} X \), if
\[
F_n(x) \to F(x)
\]
for all \( x \in \mathbb{R} \) such that \( F(X) \) is continuous at \( x \), where \( F_n \) is the df of \( X_n \) and \( F \) is the df of \( X \).

This mode of convergence only depends on the marginal distributions of \( X_n \) and \( X \). This mode of convergence is also called weak convergence.

Before discussing the relationships between these modes of convergence, let us state and prove the Weak Law of Large Numbers.

**Theorem**  Weak Law of Large Numbers

Let \( X_1, X_2, \ldots \) be a sequence of independent and identically distributed random variables with finite mean \( \mu \) and finite variance \( \sigma^2 \). Let \( \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \). Then \( \bar{X}_n \xrightarrow{p} \mu \).

**Proof.** Let \( \varepsilon > 0 \) be given. We wish to show
\[
P(\left|\bar{X}_n - \mu\right| > \varepsilon) \to 0.
\]
But
\[
P(\left|\bar{X}_n - \mu\right| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2}
\]
by Chebyshev's inequality
\[
= \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{as} \quad n \to \infty.
\]

**Example** Let \( X_1, X_2, \ldots \) be i.i.d. with \( P(X_i = 1) = \frac{1}{2} \) and \( P(X_i = -1) = \frac{1}{2} \). Then \( \mu = E[X_i] = 0 \) for all \( i \). Then by the WLLN we have \( \bar{X}_n \xrightarrow{p} 0 \). Let us now show that \( \bar{X}_n \xrightarrow{d} 0 \).

The limit df is \( F(x) =\begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \)

Let \( F_n(x) \) be the df of \( \bar{X}_n \). For \( x > 0 \), we have
\[
F_n(x) = P(\bar{X}_n \leq x) = 1 - P(\bar{X}_n > x),
\]
but
\[
P(\bar{X}_n > x) \leq P(\bar{X}_n < -x \text{ or } \bar{X}_n > x) = P(1|\bar{X}_n| > x) \to 0 \quad \text{because} \quad \bar{X}_n \xrightarrow{p} 0.
\]
So \( F_n(x) = 1 - P(\bar{X}_n > x) \to 1 \) as \( n \to \infty \).

So \( F_n(x) \to F(x) \) as \( n \to \infty \) for \( x > 0 \).
For \( x < 0 \) we have
\[
F_n(x) = P(\overline{X}_n \leq x) \leq P(\overline{X}_n \leq x \text{ or } \overline{X}_n > -x) \\
\leq P(\frac{1}{n} |X| > x) \to 0 \text{ as } n \to \infty
\]
So \( F_n(x) \to F(x) \) as \( n \to \infty \) again, since \( \overline{X}_n \to 0 \).

For \( x < 0 \), \( F(x) \) is discontinuous. Therefore, \( \overline{X}_n \to 0 \).

Why don't we require \( F_n(0) \to F(0) \) in the definition of convergence in distribution? In this example, consider
\[
F_n(0) = P(\overline{X}_n \leq 0) = P(S_n \leq 0), \text{ where } S_n = \sum_{i=1}^{n} X_i.
\]
If \( n \) is odd, then \( S_n \) cannot equal 0, and by symmetry,
\[
P(S_n < 0) = P(S_n > 0) = \frac{1}{2}, \text{ so } P(S_n \leq 0) = \frac{1}{2}.
\]
If \( n \) is even, \( P(S_n = 0) > 0 \), but as \( n \to \infty \) \( P(S_n = 0) \to 0 \).
Also by symmetry, \( P(S_n < 0) = P(S_n > 0) \), so
\[
P(S_n \leq 0) = \frac{1}{2} + \epsilon(n), \text{ where } \epsilon(n) \to 0 \text{ as } n \to \infty.
\]
Therefore, \( F_n(0) = P(S_n \leq 0) \to \frac{1}{2} \text{ as } n \to \infty \).

So \( F_n(0) \not\to F(0) = 1 \)

If we were to require that \( F_n(0) \to F(0) \) in the definition of convergence in distribution, then we would be in the undesirable or non-useful situation that \( \overline{X}_n \) in this example would not converge to anything in distribution, yet it does converge to 0 in probability.
For the 4 modes of convergence we have the following implications:

\[ X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{d} X \quad \Rightarrow \quad X_n \xrightarrow{r.m.} X \]

In general no other implications hold.

**Theorem** \( X_n \xrightarrow{r.m.} X \implies X_n \xrightarrow{p} X \).

**Proof.** Let \( \varepsilon > 0 \) be given. Then

\[
P(X_n - X > \varepsilon) = P\left(\left|X_n - X\right| > \varepsilon\right) 
\leq \frac{E[|X_n - X|^r]}{\varepsilon^r} 
\xrightarrow{\text{by Markov's inequality}} 0
\]

since \( E[|X_n - X|^r] \to 0 \) by assumption.

**Theorem** \( X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{p} X \).

**Proof.** First, recall that \( X_n(\omega) \to X(\omega) \) as \( n \to \infty \) means that for any \( \varepsilon > 0 \) there exists \( N \) such that for \( n \geq N \),

\( |X_n(\omega) - X(\omega)| \leq \varepsilon \). Let \( \varepsilon > 0 \) be given.

Define \( A(E) = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon \text{ for infinitely many } n\} \)

\( A_n(E) = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon \} \)

\( B_n(E) = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon \text{ for some } m \geq n\} \)

Note that

(i) if \( \omega \in A(E) \) then \( X_n(\omega) \) does not converge to \( X(\omega) \). Since \( X_n \xrightarrow{a.s.} X \) we have \( P(A(E)) = 0 \) for any \( \varepsilon > 0 \).

(ii) To show that \( X_n \xrightarrow{p} X \), we want to show that \( P(A_n(E)) \to 0 \) as \( n \to \infty \).

(iii) \( A_n(E) \subseteq B_n(E) \) for every \( n \).

(iv) \( B_1(E) \supseteq B_2(E) \supseteq B_3(E) \supseteq \ldots \), i.e. \( \{B_n(E)\} \) is a decreasing sequence of sets.

Next, note that \( A(E) = \bigcap_{n=1}^{\infty} B_n(E) \) (check that if \( \omega \in A(E) \) then \( \omega \in B_n(E) \) for every \( n \), and if \( \omega \in B_n(E) \) for every \( n \) then \( \omega \in A(E) \)).
Then $0 = P(A(\epsilon)) = P(\bigcap_{n=1}^\infty B_n(\epsilon))$
\[= P(\lim_{n \to \infty} B_n(\epsilon)) \quad \text{since the } \{B_n(\epsilon)\} \text{ are decreasing} \]
\[= \lim_{n \to \infty} P(B_n(\epsilon)) \quad \text{by the continuity of probability} \]

But $A_n(\epsilon) \subseteq B_n(\epsilon)$, so $P(A_n(\epsilon)) \leq P(B_n(\epsilon))$. Therefore, if $\lim_{n \to \infty} P(B_n(\epsilon)) = 0$ then $\lim_{n \to \infty} P(A_n(\epsilon)) = 0$.

That is, $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$. So $X_n \xrightarrow{p} X$.

Example (showing that $X_n \xrightarrow{p} X$ does not imply that $X_n \xrightarrow{a.s.} X$).

Let $X_1, X_2, \ldots$ be independent random variables with
\[P(X_n = 0) = 1 - \frac{1}{n} \quad \text{Let } \epsilon > 0 \text{ be given}.\]

Then $P(|X_n - 0| > \epsilon) = P(X_n = 1) = \frac{1}{n} \to 0$ as $n \to \infty$.

Therefore, $X_n \xrightarrow{p} 0$. Let us now see that $X_n$ does not converge to 0 a.s. Referring to the set $B_n(\epsilon)$ from proof of last theorem, consider
\[B_n(\epsilon)^c = \{ \omega \in \Omega : |X_m(\omega) - 0| \leq \epsilon \text{ for all } m \geq n \}.\]

\[P(B_n(\epsilon)^c) = P(X_n \leq \epsilon, X_{n+1} \leq \epsilon, \ldots) = \lim_{M \to \infty} P(X_n \leq \epsilon, X_{n+1} \leq \epsilon, \ldots, X_M \leq \epsilon) \]
\[= \lim_{M \to \infty} P(X_n \leq \epsilon) \times P(X_{n+1} \leq \epsilon) \times \ldots \times P(X_M \leq \epsilon) \quad \text{by independence} \]
\[= \lim_{M \to \infty} P(X_n = 0) \times P(X_{n+1} = 0) \times \ldots \times P(X_M = 0) \]
\[= \lim_{M \to \infty} (1 - \frac{1}{n}) (1 - \frac{1}{n+1}) \times \ldots \times (1 - \frac{1}{M}) \]
\[= \lim_{M \to \infty} \frac{(n+1)(n+2) \ldots (n+M)}{n} \times \ldots \times \frac{(M+1) \ldots (M+M)}{M} \]
\[= \lim_{M \to \infty} \frac{n-1}{M} = 0 \]
Therefore, \( P(B_n(\varepsilon)) = 1 \). Then \( P(\bigcap_{n=1}^{\infty} B_n(\varepsilon)) = 1 \). Therefore, \( P(A(\varepsilon)) = 1 \) since \( A(\varepsilon) = \bigcap_{n=1}^{\infty} B_n(\varepsilon) \). Recall, if \( \omega \in A(\varepsilon) \) then \( X_n(\omega) \) does not converge to the limit. Therefore, with probability 1, \( X_n \) does not converge to 0.

Remark 1. The basic reason why \( X_n \to 0 \) a.s. in preceding example is that we constructed the \( X_n \)'s to be independent. Consider the following construction. Let \( \Omega = (0,1) \) and \( P \) be the Uniform \((0,1)\) distribution. Define \( X_n(\omega) = \begin{cases} 0 & \text{if } \omega \geq \frac{1}{n} \\ 1 & \text{if } \omega < \frac{1}{n} \end{cases} \)

Then \( P(X_n = 0) = 1 - \frac{1}{n} \) and \( P(X_n = 1) = \frac{1}{n} \), which is the same distribution for \( X_n \) as in the preceding example. Then \( X_n \) does exactly as in preceding example. Here, we can check that the \( X_n \) constructed as they are converge to 0 a.s. as well. Let \( \omega \in (0,1) \). We check if \( X_n(\omega) \to 0 \) as \( n \to \infty \). But eventually \( n \) will be big enough \( (n > \frac{1}{\omega}) \) such that \( X_n(\omega) = 0 \). Therefore, \( \lim_{n \to \infty} X_n(\omega) = 0 \).

Since \( P((0,1)) = 1 \), we see that \( X_n \to 0 \).

Remark 2. The sequence \( \{X_n^3\} \) from the example also converges to 0 in the \( r \)-th mean, since
\[
E[(X_n - 0)^r] = E[X_n^r] = (0^r)(1 - \frac{1}{n}) + (1^r)\frac{1}{n} = \frac{1}{n} \to 0 \quad \text{as} \quad n \to \infty.
\]
This illustrates that \( X_n \xrightarrow{r} X \not\Rightarrow X_n \xrightarrow{a.s.} X \).
We start today with a theorem giving a sufficient condition for convergence almost surely. Recall, $X_n \overset{a.s.}{\rightarrow} X$ means that for any $\varepsilon > 0$, $P(|X_n - X| > \varepsilon) \rightarrow 0$ as $n \rightarrow \infty$. If this probability converges to 0 at a fast enough rate , then $X_n$ will converge to $X$ a.s.

**Theorem** If $\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$ for any $\varepsilon > 0$, then $X_n \overset{a.s.}{\rightarrow} X$.

**Proof.** Recall the sets

$A_n(\varepsilon) = \{ \omega \in \Omega: |X_n(\omega) - X(\omega)| > \varepsilon \}$

$B_n(\varepsilon) = \{ \omega \in \Omega: |X_m(\omega) - X(\omega)| > \varepsilon \text{ for some } m \geq n \}$

$A(\varepsilon) = \{ \omega \in \Omega: |X_n(\omega) - X(\omega)| > \varepsilon \text{ for infinitely many } n \}$

First, we show that the given condition implies $\lim_{n \rightarrow \infty} P(B_n(\varepsilon)) = 0$.

Note that $B_n(\varepsilon) = \bigcup_{m=n}^{\infty} A_m(\varepsilon)$. Then

$$P(B_n(\varepsilon)) = P(\bigcup_{m=n}^{\infty} A_m(\varepsilon))$$

$$\leq \sum_{m=n}^{\infty} P(A_m(\varepsilon)) \quad \text{(by subadditivity of probability)}.$$ 

$$\rightarrow 0 \text{ as } n \rightarrow \infty \text{ since } \sum_{m=1}^{\infty} P(A_m(\varepsilon)) < \infty.$$

So $P(B_n(\varepsilon)) \rightarrow 0$ as $n \rightarrow \infty$ for any $\varepsilon > 0$. Then, as in proof of a.s. convergence implying convergence in probability, we have $P(A(\varepsilon)) = 0$. For $X_n(\omega) \rightarrow X(\omega)$ to hold, $\omega$ should belong in $A(\varepsilon)^c$ for every $\varepsilon > 0$. It is sufficient to show this for $\varepsilon$ of the form $\frac{1}{k}$. Take $\varepsilon = \frac{1}{k}$. Then $P(A(\frac{1}{k})) = 0$ for every $k \geq 1$.

Then $P(A(\frac{1}{k})^c) = 1$ for every $k \geq 1$.

Let $C = \bigcap_{k=1}^{\infty} A(\frac{1}{k})^c$. Then $P(C) = 1$.

If we choose for every $k \geq 1$, there exists an $N$ such that for $n \geq N$, $|X_n(\omega) - X(\omega)| \leq \frac{1}{k}$ for all $n \geq N$. This implies that $X_n(\omega) \rightarrow X(\omega)$. Since $P(C) = 1$, we conclude that $X_n \overset{a.s.}{\rightarrow} X$. 
Example (counterexample showing $X_n \overset{a.s.}{\to} X$ does not imply $X_n \overset{r,m}{\to} X$)

Take $\Omega = (0,1)$ and $P$ the Uniform $(0,1)$ distribution.

Take $X_n(\omega) = \begin{align*}
0 & \quad \text{if } \omega \geq \frac{1}{n} \\
\frac{1}{n^2} & \quad \text{if } \omega < \frac{1}{n}
\end{align*}$

Then $P(X_n = 0) = 1 - \frac{1}{n}$ and $P(X_n = n^2) = \frac{1}{n}$. If $\omega \in (0,1)$ then $X_n(\omega)$ will equal 0 for all $n$ such that $\frac{1}{n} \leq \omega$, or $n \geq \frac{1}{\omega}$. So $\lim_{n \to \infty} X_n(\omega) = 0$ for all $\omega \in (0,1)$. Since $P((0,1)) = 1$, we have that $X_n \overset{a.s.}{\to} 0$. But

$$E[|X_n - 0|^r] = E[X_n^r] = (0^r)(1 - \frac{1}{n}) + (n^2)^r \frac{1}{n} = n^{2r-1} \to \infty$$

So $X_n \not\overset{r,m}{\to} 0$ for $r > \frac{1}{2}$.

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

Therefore, $X_n \overset{p}{\to} 0$. So convergence in probability does not imply convergence in the $r$th mean.

Now let us show that $X_n \overset{p}{\to} X$ does imply $X_n \overset{d}{\to} X$.

**Theorem** $X_n \overset{p}{\to} X \implies X_n \overset{d}{\to} X$.

**Proof** Let $F_n$ be the df of $X_n$, $F$ the df of $X$, and let $x \in \mathbb{R}$ be such that $F$ is continuous at $x$. We wish to show $F_n(x) \to F(x)$ as $n \to \infty$. Our approach will be to lower bound and upper bound $F_n(x)$ by quantities that are close to $F(x)$.

(a) $F_n(x) = P(X_n \leq x)$

$$= P(X_n \leq x, X \leq x + \varepsilon) + P(X_n \leq x, X > x + \varepsilon)$$

$$\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon)$$

$$= F(x + \varepsilon) + P(|X_n - X| > \varepsilon)$$
(b) \( F(x - \varepsilon) = P(X \leq x - \varepsilon) \)
\[ = P(X \leq x - \varepsilon, X_n \leq x) + P(X \leq x - \varepsilon, X_n > x) \]
\[ \leq P(X_n \leq x) + P(\lvert X_n - x \rvert > \varepsilon) \]
\[ = F_n(x) + P(\lvert X_n - x \rvert > \varepsilon) \]

Combining (a) and (b), we have
\[ F(x - \varepsilon) - P(\lvert X_n - x \rvert > \varepsilon) \leq F_n(x) \leq F(x + \varepsilon) + P(\lvert X_n - x \rvert > \varepsilon) \]

First, let \( n \to \infty \) (note that we do not know at this point that \( \lim F_n(x) \) exists). Take \( \liminf \) of all quantities and \( \limsup \) of all quantities to get
\[ F(x - \varepsilon) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x + \varepsilon) \]

The above holds for every \( \varepsilon > 0 \) and so, since \( F \) is continuous at \( x \), taking \( \varepsilon \to 0 \), we get
\[ F(x) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x) \]

Therefore, \( \liminf_{n \to \infty} F_n(x) = \limsup_{n \to \infty} F_n(x) = F(x) \), and so
\[ \lim_{n \to \infty} F_n(x) = F(x) \]. Thus, \( X_n \overset{d}{\to} X \).

**Example** (Counterexample showing that \( X_n \overset{d}{\to} X \) does not imply \( X_n \overset{p}{\to} X \)).

Let \( X, X_1, X_2, \ldots \) be i.i.d. Uniform \((0,1)\) random variables.

Then \( F_n(x) = F(x) \) for all \( x \in \mathbb{R} \) and for every \( n \). So, trivially, \( X_n \overset{d}{\to} X \). But \( P(\lvert X_n - x \rvert > \varepsilon) \) is the same for every \( n \), and is not 0 (in fact it is \( (1-\varepsilon)^2 \)).

Thus, \( X_n \) does not converge to \( X \) in probability.