6 Convergence in $\mathbb{R}^d$ and in Metric Spaces

A sequence of elements $a_n$ of $\mathbb{R}^d$ converges to a limit $a$ if and only if, for each $\epsilon > 0$, the sequence $a_n$ eventually lies within a ball of radius $\epsilon$ centered at $a$. It’s okay if the first few (or few million) terms lie outside that ball—and the number of terms that do lie outside the ball may depend on how big $\epsilon$ is (if $\epsilon$ is small enough it will take millions of terms before the remaining sequence lies inside the ball). This can be made mathematically precise by introducing a letter (say, $N_\epsilon$) for how many initial terms we have to throw away, so that $a_n \rightarrow a$ if and only if there is an $N_\epsilon < \infty$ so that, for each $n \geq N_\epsilon$, $|a_n - a| < \epsilon$: only finitely many $a_n$ can be farther than $\epsilon$ from $a$.

The same notion of convergence really works in any metric space, where we require that some measure of the distance $d(a_n, a)$ from $a_n$ to $a$ tend to zero in the sense that it exceeds each number $\epsilon > 0$ for at most some finite number $N_\epsilon$ of terms.

Points $a_n$ in $d$-dimensional Euclidean space will converge to a limit $a \in \mathbb{R}^d$ if and only if each of their coordinates converges in $\mathbb{R}$; and, since there are only finitely many coordinates, if they all converge then they do so uniformly (i.e., for each $\epsilon$ we can take the same $N_\epsilon$ for all $d$ of the coordinate sequences), so all notions of convergence in $\mathbb{R}^d$ are equivalent; for example,

$$\max_{1 \leq i \leq d} |x_i - y_i| \leq \left[ \sum_{1 \leq i \leq d} (x_i - y_i)^2 \right]^{1/2} \leq \sum_{1 \leq i \leq d} |x_i - y_i| \leq d \max_{1 \leq i \leq d} |x_i - y_i|$$

Convergence is much more complex and interesting for random variables.

6.1 Convergence of Random Variables

For random variables $X_n$ the idea of convergence to a limiting random variable $X$ is more delicate, since each $X_n$ is a function of $\omega \in \Omega$ and usually there are infinitely many points $\omega \in \Omega$. What should we mean in saying that a sequence $X_n$ converges to a limit $X$? That $X_n(\omega)$ converges to $X(\omega)$ for each fixed $\omega$? Or that $X_n(\omega)$ converges uniformly in $\omega \in \Omega$? Or that some notion of the distance $d(X_n, X)$ between $X_n$ and the limit $X$ decreases to zero? Should the probability measure $P$ be involved in some way?
Here are a few different choices of what we might mean by the statement that \( X_n \) converges to \( X \), for a sequence of random variables \( X_n \) and a random variable \( X \), all defined on the same probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \):

\[\text{pw: The real numbers } X_n(\omega) \to X(\omega) \text{ for every } \omega \in \Omega \text{ (pointwise convergence):} \]
\[ (\forall \varepsilon > 0) \ (\forall \omega \in \Omega) \ (\exists N_{\varepsilon, \omega} < \infty) \ (\forall n \geq N_{\varepsilon, \omega}) \ |X_n(\omega) - X(\omega)| < \varepsilon. \]

\[\text{uni: The sequences of real numbers } X_n(\omega) \to X(\omega) \text{ uniformly for } \omega \in \Omega: \]
\[ (\forall \varepsilon > 0) \ (\exists N_{\varepsilon} < \infty) \ (\forall \omega \in \Omega) \ (\forall n \geq N_{\varepsilon}) \ |X_n(\omega) - X(\omega)| < \varepsilon. \]

\[\text{a.s.: Outside some null event } N \in \mathcal{F}, \text{ each sequence of real numbers } X_n(\omega) \to X(\omega) \text{ (Almost-Sure convergence, or convergence “almost everywhere” (a.e.):) for some } N \in \mathcal{F} \text{ with } \mathbb{P}[N] = 0, \]
\[ (\forall \varepsilon > 0) \ (\forall \omega \notin N) \ (\exists N_{\varepsilon, \omega} < \infty) \ (\forall n \geq N_{\varepsilon, \omega}) \ |X_n(\omega) - X(\omega)| < \varepsilon. \]
\[ \text{i.e., } \mathbb{P}\left\{ \bigcup_{\varepsilon > 0} \bigcap_{n < \infty} \bigcup_{n \geq N} |X_n(\omega) - X(\omega)| \geq \varepsilon \right\} = 0. \]

\[\text{L}_{\infty}: \text{ Outside some null event } N \in \mathcal{F}, \text{ the sequences of real numbers } X_n(\omega) \to X(\omega) \text{ converge uniformly (“almost-uniform” or “L}_{\infty} \text{” convergence): for some } N \in \mathcal{F} \text{ with } \mathbb{P}[N] = 0, \]
\[ (\forall \varepsilon > 0) \ (\exists N_{\varepsilon} < \infty) \ (\forall \omega \notin N) \ (\forall n \geq N_{\varepsilon}) \ |X_n(\omega) - X(\omega)| < \varepsilon. \]

\[\text{i.p.: For each } \varepsilon > 0, \text{ the probabilities } \mathbb{P}[|X_n - X| > \varepsilon] \to 0 \text{ (convergence “in probability”, or “in measure”):} \]
\[ (\forall \varepsilon > 0) \ (\forall \eta > 0) \ (\exists N_{\varepsilon, \eta} < \infty) \ (\forall n \geq N_{\varepsilon, \eta}) \ \mathbb{P}[|X_n - X| > \varepsilon] < \eta. \]

\[\text{L}_1: \text{ The expectation } \mathbb{E}[|X_n - X|] \text{ converges to zero (convergence “in } L_1\text{”):} \]
\[ (\forall \varepsilon > 0) \ (\exists N_{\varepsilon} < \infty) \ (\forall n \geq N_{\varepsilon}) \ \mathbb{E}[|X_n - X|] < \varepsilon. \]

\[\text{L}_p: \text{ For some fixed number } p > 0, \text{ the expectation of the } p^{th} \text{ absolute power } \mathbb{E}[|X_n - X|^p] \text{ converges to zero (convergence “in } L_p\text{,” sometimes called “in the } p^{th} \text{ mean”):} \]
\[ (\forall \varepsilon > 0) \ (\exists N_{\varepsilon} < \infty) \ (\forall n \geq N_{\varepsilon}) \ \mathbb{E}[|X_n - X|^p] < \varepsilon. \]
The distributions of \( X_n \) converge to the distribution of \( X \), i.e., the measures \( P \circ X_n^{-1} \) converge in some way to \( P \circ X^{-1} \) (“vague” or “weak” convergence, or “convergence in distribution”, sometimes written \( X_n \Rightarrow X \)):

\[
(\forall \epsilon > 0) \ (\forall \phi \in \mathcal{C}_b(\mathbb{R})) \ (\exists N_{\epsilon, \phi} < \infty) \ (\forall n \geq N_{\epsilon, \phi}) \quad \mathbb{E}[|\phi(X_n) - \phi(X)|] < \epsilon.
\]

Which of these eight notions of convergence is right for random variables? The answer is that all of them are useful in probability theory for one purpose or another. You will want to know which ones imply which other ones, under what conditions. All but the first two (pointwise, uniform) notions depend upon the measure \( P \); it is possible for a sequence \( X_n \) to converge to \( X \) in any of these senses for one probability measure \( P \), but to fail to converge for another \( P' \). Most of them can be phrased as metric convergence for some notion of distance between random variables:

\( i.p. \): \( X_n \rightarrow X \) in probability if and only if \( d_0(X, X_n) \rightarrow 0 \) as real numbers, where:

\[
d_0(X, Y) = \mathbb{E} \left( \frac{|X - Y|}{1 + |X - Y|} \right).
\]

\( L_1 \): \( X_n \rightarrow X \) in \( L_1 \) if and only if \( d_1(X, X_n) = \|X - X_n\|_1 \rightarrow 0 \) as real numbers, where:

\[
\|Z\|_1 = \mathbb{E}|Z|
\]

\( L_p \): \( X_n \rightarrow X \) in \( L_p \) if and only if \( d_p(X, X_n) = \|X - X_n\|_p \rightarrow 0 \) as real numbers, where:

\[
\|Z\|_p = (\mathbb{E}|Z|^p)^{1/p}
\]

\( L_\infty \): \( X_n \rightarrow X \) almost uniformly if and only if \( d_\infty(X, X_n) = \|X - X_n\|_\infty \rightarrow 0 \) as real numbers, where:

\[
\|Z\|_\infty = \sup\{r \geq 0 : \ P(|Z| > r) > 0\}
\]

As the notation suggests, convergence in probability and in \( L_\infty \) are in some sense limits of convergence in \( L_p \) as \( p \rightarrow 0 \) and \( p \rightarrow \infty \), respectively. Almost-sure convergence is an exception: there is no metric notion of distance \( d(X, Y) \) for which \( X_n \rightarrow X \) almost surely if and only if \( d(X, X_n) \rightarrow 0 \).
6.1.1 Almost-Sure Convergence

Let \( \{X_n\} \) and \( X \) be a collection of RV’s on some \((\Omega, \mathcal{F}, P)\). The set of points \( \omega \) for which \( X_n(\omega) \) does converge to \( X(\omega) \) is just

\[
\bigcap_{\epsilon > 0} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{ \omega : |X_n(\omega) - X(\omega)| \leq \epsilon \},
\]

the points which, for all \( \epsilon > 0 \), have \( |X_n(\omega) - X(\omega)| \) less than \( \epsilon \) for all but finitely-many \( n \). The sequence \( X_n \) is said to converge “almost everywhere” (a.e.) to \( X \), or to converge to \( X \) “almost surely” (a.s.), if this set of \( \omega \) has probability one, or (conversely) if its complement is a null set:

\[
P\left( \bigcup_{\epsilon > 0} \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{ \omega : |X_n(\omega) - X(\omega)| > \epsilon \} \right) = 0.
\]

Despite its appearance the union over \( \epsilon > 0 \) is only countable, since we need include only rational \( \epsilon \) (or, for that matter, any sequence \( \epsilon_k \) tending to zero, such as \( \epsilon_k = 1/k \)). Thus \( X_n \to X \) a.e. if and only if, for each \( \epsilon > 0 \),

\[
P\left( \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{ \omega : |X_n(\omega) - X(\omega)| > \epsilon \} \right) = 0. \quad \text{(a.e.)}
\]

This combination of intersection and union occurs frequently in probability, and has a name; for any sequence \( E_n \) of events, \( \bigcap_{n=m}^{\infty} \bigcup_{n=m}^{\infty} E_n \) is called the \textit{lim sup} of the \( \{E_n\} \), and is sometimes described more colorfully as \( E_n \ i.o. \), the set of points in \( E_n \) “infinitely often.” Its complement is the \textit{lim inf} of the sets \( F_n = E_n^c \cup \bigcup_{n=m}^{\infty} F_n \); the set of points in all but finitely many of the \( F_n \). Since \( P \) is countably additive, and since the intersection in the definition of \textit{lim sup} is \textit{decreasing} and the union in the definition of \textit{lim inf} is \textit{increasing}, always we have

\[
P\left[ \bigcup_{n=m}^{\infty} E_n \right] \geq P\left[ \bigcap_{n=m}^{\infty} E_n \right] \text{ and } P\left[ \bigcap_{n=m}^{\infty} F_n \right] \leq P\left[ \bigcup_{n=m}^{\infty} F_n \right]
\]
as \( m \to \infty \). Thus,

\[\textbf{Theorem 1} \quad X_n \to X \ P\text{-a.s. if and only if for every } \epsilon > 0, \]

\[
\lim_{m \to \infty} P[|X_n - X| > \epsilon \text{ for some } n \geq m] = 0.
\]

In particular, \( X_n \to X \ P\text{-a.s. if } \sum_{n=1}^{\infty} P[|X_n - X| > \epsilon] < \infty \text{ for each } \epsilon > 0 \) (why?).
6.1.2 Convergence In Probability

The sequence $X_n$ is said to converge to $X$ “in probability” (i.p.) if, for each $\epsilon > 0$,

$$P[\omega : |X_n(\omega) - X(\omega)| > \epsilon] \to 0. \quad (i.p.)$$

If we denote by $E_n$ the event $[\omega : |X_n(\omega) - X(\omega)| > \epsilon]$ we see that convergence 
almost surely requires that $P(\bigcup_{n>m} E_n) \to 0$ as $m \to \infty$, while convergence 
in probability requires only that $P(E_n) \to 0$. Thus:

**Theorem 2** If $X_n \to X$ a.s. then $X_n \to X$ i.p.

Here is a partial converse:

**Theorem 3** If $X_n \to X$ i.p., then there is a subsequence $n_k$ such that $X_{n_k} \to X$ a.s.

**Proof.** Set $n_0 = 0$ and, for each integer $k \geq 1$, set

$$n_k = \inf \left\{ n > n_{k-1} : P\left[ \omega : |X_n(\omega) - X(\omega)| > \frac{1}{k} \right] \leq 2^{-k} \right\}.$$ 

For any $\epsilon > 0$ we have $\frac{1}{k} \leq \epsilon$ eventually (namely, for $k \geq k_0 = \lceil \frac{1}{\epsilon} \rceil$) and for each $m \geq k_0$,

$$P\left[ \bigcup_{k=m}^{\infty} [\omega : |X_{n_k}(\omega) - X(\omega)| > \epsilon] \right] \leq P\left[ \bigcup_{k=m}^{\infty} [\omega : |X_{n_k}(\omega) - X(\omega)| > \frac{1}{k}] \right]$$

$$\leq \sum_{k=m}^{\infty} P[\omega : |X_{n_k}(\omega) - X(\omega)| > \frac{1}{k}]$$

$$\leq \sum_{k=m}^{\infty} 2^{-k} = 2^{1-m} \to 0 \text{ as } m \to \infty. \quad \Box$$

In fact, this characterizes convergence in probability:

**Theorem 4** Let $\{X_n\}$, $X$ be random variables on $(\Omega, \mathcal{F}, P)$. Then $X_n \to X$ 
i.p. if and only if every sequence $\mathbb{N} \ni n_k \nearrow \infty$ has a subsequence $n_{k_i}$ such that $X_{n_{k_i}} \to X$ a.s.
Proof. The “only if” (⇒) direction is just Theorem 3. Suppose (for contradiction) that \( X_n \not\to X \, i.p. \); then for some \( \epsilon > 0 \) and \( \delta > 0 \) there are infinitely-many \( n \) for which \( \Pr[|X_n - X| > \epsilon] > \delta \). Let \( n_k \) be an increasing sequence satisfying this bound. By hypothesis, there is a subsequence along which \( X_{n_{k_i}} \to X \) a.s.; but by Theorem 2, also \( X_{n_{k_i}} \to X \, i.p. \), a contradiction. \( \Box \)

Theorem 5 Let \( \{X_n\} \), \( X \) be real-valued random variables on \((\Omega, \mathcal{F}, \Pr)\) and let \( \phi : \mathbb{R} \to \mathbb{R} \) be continuous. Then \( Y_n \equiv \phi(X_n) \to Y \equiv \phi(X) \, i.p. \).

Proof. For an easy but indirect proof, simply apply Theorem 4. For a more direct approach, begin by selecting any \( \epsilon > 0 \) and \( \delta > 0 \). Find a compact set \( K_\epsilon \subset \mathbb{R} \) with \( \Pr[X \in K_\epsilon] \geq 1 - \epsilon/2 \); since \( \phi \) is uniformly continuous on \( K_\epsilon \), find \( \eta > 0 \) such that

\[
(\forall x \in K_\epsilon)(\forall y \in \mathbb{R}) \quad |x - y| \leq \eta \quad \Rightarrow \quad |\phi(x) - \phi(y)| < \delta.
\]

Now, since \( X_n \to X \, i.p. \), find \( N \in \mathbb{N} \) such that

\[
(\forall n \geq N) \Pr[|X_n - X| > \eta] \leq \frac{\epsilon}{2}.
\]

Then for \( n \geq N \),

\[
\begin{align*}
\Pr[|Y_n - Y| > \delta] &\leq \Pr[X \not\in K_\epsilon] + \Pr[X \in K_\epsilon, \, |\phi(X_n) - \phi(X)| > \delta] \\
&\leq \Pr[X \not\in K_\epsilon] + \Pr[X \in K_\epsilon, \, |X_n - X| > \eta] \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{align*}
\]

\( \Box \)

The same result (with the same proof) is true for random variables \( X_n, X \) taking values in any \( \sigma \)-compact complete separable metric space \( \mathcal{X} \) (and, in particular, for \( \mathcal{X} = \mathbb{R}^d \)).

### 6.1.3 A Counter-Example

If \( X_n \to X \, a.s. \) implies \( X_n \to X \, i.p. \), and if the converse holds at least along subsequences, are the two notions really identical? Or is it possible for RV’s \( X_n \) to converge to \( X \, i.p. \), but not \( a.s. \)? The answer is that the
two notions are different, and that a.s. convergence is strictly stronger than
convergence i.p. Here’s an example:

First notice that every integer \( n \in \mathbb{N} \) can be written uniquely in the form
\( n = i + 2^j \) for integers \( j \geq 0 \) and \( 0 \leq i < 2^j \) (set \( j := \lfloor \log_2 n \rfloor \) and
\( i := n - 2^j \)). Let \((\Omega, \mathcal{F}, \mathbb{P})\) be the unit interval with Borel sets and Lebesgue
measure. Define a sequence of random variables \( X_n : \Omega \rightarrow \mathbb{R} \) by

\[
X_n(\omega) = \begin{cases} 
1 & \text{if } \frac{i}{2^j} < \omega \leq \frac{i+1}{2^j} \\
0 & \text{otherwise}
\end{cases}
\]

where \( n = i + 2^j \), \( 0 \leq i < 2^j \).

SO,

\[
X_1 \text{ is uniform on 1 interval of length 1,}
\]

\[
X_2, X_3 \text{ are uniform on 2 intervals of length 1/2,}
\]

\[
X_4, \ldots, X_7 \text{ are uniform on 4 intervals of length 1/4,}
\]

\[
X_8, \ldots, X_{15} \text{ are uniform on 8 intervals of length 1/8,}
\]

and, in general, each \( X_n \) is one on an interval of length \( 2^{-j} \); since \( \frac{1}{n} \leq \frac{1}{2^n} < \frac{2}{n} \),

\[
P[|X_n| > \epsilon] = 2^{-j} < \frac{2}{n} \rightarrow 0
\]

for each \( 0 < \epsilon < 1 \) and \( X_n \rightarrow 0 \) i.p. On the other hand, for every \( j > 0 \) we
have each \( \omega \in \Omega \) in one of the \( 2^j \) intervals of length \( 2^{-j} \) where some \( X_n \) is
one,

\[
\Omega = \bigcup_{i=0}^{2^j-1} \left( \frac{i}{2^j}, \frac{i+1}{2^j} \right) = \bigcup_{n=2^j}^{2^{j+1}-1} [\omega : X_n(\omega) = 1]
\]

so for every \( \omega \in \Omega \) and \( j \in \mathbb{N} \) there is some \( n \geq 2^j \) with \( X_n(\omega) = 1 \) (and so
there are infinitely-many such \( X_n \)); thus \( [\omega : X_n(\omega) \rightarrow 0] \) is empty, not a set
of probability one! Obviously \( X_n \) does not converge a.s.

This example is a building-block for several examples to come, so getting to
know it well is worth while. Try to verify that \( X_n \rightarrow 0 \) in probability and
in \( L_p \) but not almost surely. What is \( \|X_n\|_p \)? Why doesn’t \( X_n \rightarrow 0 \) a.s.? What
would happen if we multiplied \( X_n \) by \( n \)? By \( n^2 \)? By \( j \)? What about
the subsequence \( Y_n = X_{2n} \)? Does \( X_n \) converge in \( L_\infty \)?
In summary, for $1 < p < q < \infty$ the convergence implications are:

\[ L_\infty \overset{a.s.}{\rightarrow} L_q \overset{\text{i.p.}}{\rightarrow} L_p \overset{\text{a.s.}}{\rightarrow} L_1 \]

with partial converses ($\text{i.p.} \rightarrow \text{a.s.}$ along subsequences, $\text{i.p.} \rightarrow L_1, L_p, L_q$ under UI). All of these imply convergence in distribution, which we'll consider later.

### 6.2 Uniform Integrability

Let $Y \geq 0$ be integrable on some probability space $(\Omega, \mathcal{F}, P)$,

\[ \mathbb{E}[Y] = \int_{\Omega} Y \, dP < \infty. \]

By Lebesgue's DCT it follows that

\[ \lim_{t \to \infty} \mathbb{E}[Y \mathbf{1}_{\{Y > t\}}] = \int_{\{\omega : Y(\omega) > t\}} Y \, dP = 0 \]

since $\{Y \mathbf{1}_{\{Y > t\}}\}$ is dominated by $Y \in L_1$ and converges to zero as $t \to \infty$.
and, consequently, that for any sequence of random variables \(X_n\) also dominated by \(Y\) in the sense that \(|X_n| \leq Y\) a.s.,

\[
E[|X_n| 1_{\{|X_n| > t\}}] \leq E[Y 1_{\{|Y| > t\}}] \to 0, \text{ uniformly in } n.
\]

Call the sequence \(X_n\) uniformly integrable (or simply UI) if \(E[|X_n| 1_{\{|X_n| > t\}}] \to 0\) uniformly in \(n\), even if it is not dominated by a single integrable random variable \(Y\). The big result is:

**Theorem 6** If \(X_n \to X\) i.p. and if \(X_n\) is UI then \(X_n \to X\) in \(L_1\).

**Proof.** Without loss of generality take \(X \equiv 0\). Fix any \(\epsilon > 0\); find (by UI) \(t_\epsilon > 0\) such that \(E[|X_n| 1_{\{|X_n| > t_\epsilon\}}] \leq \epsilon\) for all \(n\). Now (by \(X_n \to X\) i.p.) find \(N_\epsilon \in \mathbb{N}\) such that, for \(n \geq N_\epsilon\), \(P[|X_n| > \epsilon] < \epsilon / t_\epsilon\); then:

\[
E[|X_n|] = \int_{|X_n| \leq \epsilon} |X_n| dP + \int_{\epsilon < |X_n| \leq t_\epsilon} |X_n| dP + \int_{t_\epsilon < |X_n|} |X_n| dP
\]

\[
\leq \int_{|X_n| \leq \epsilon} \epsilon dP + \int_{\epsilon < |X_n| \leq t_\epsilon} t_\epsilon dP + \int_{t_\epsilon < |X_n|} |X_n| dP
\]

\[
\leq \epsilon + t_\epsilon \times P[|X_n| > \epsilon] + \epsilon
\]

\[
\leq 3\epsilon.
\]

\[\square\]

Theorem 6 has a partial converse— if \(\{X_n\} \subset L_1\) and \(X_n \to X\) i.p. and if \(E[X_n] \to E[X]\), then \(\{X_n\}\) is UI (see Theorem 9). For another proof, first note \(E[|X_n| 1_{\{|X_n| > t\}}] = E[|X_n|] - E[|X_n| 1_{\{|X_n| \leq t\}}]\). For \(\epsilon > 0\) first pick \(t < \infty\) s.t. \(E[|X| 1_{\{|X| > t\}}] < \epsilon\) (possible, since \(X \in L_1\)); then (by DCT) find \(N\) s.t. for \(n \geq N\), \(\|X_n 1_{\{|X_n| \leq t\}} - X 1_{\{|X| \leq t\}}\|_1 < \epsilon\) and (by assumption) \(E[|X_n|] - E[|X|] < \epsilon\). Then \(E[|X_n| 1_{\{|X_n| > t\}}] \leq 3\epsilon\) for \(n \geq N\); since \(X_j \in L_1\) for \(j < N\), the result follows (with \(t_\epsilon = \max(t, t_1, \ldots, t_{N-1})\)).

Similarly, for any \(p > 0\), \(X_n \to X\) (i.p.) and \(|X_n|^p\) UI (for example, \(|X_n| \leq Y \in L_p\), or \(||X_n||_q \leq B < \infty\) for some \(q > p\)) gives \(X_n \to X\) (\(L_p\)). In the special case of \(|X_n| \leq Y \in L_p\) this is just Lebesgue’s Dominated Convergence Theorem (DCT).

We have seen that \(\{X_n\}\) is UI whenever \(|X_n| \leq Y \in L_1\), but UI is more general than that. Here are two more criteria:

**Theorem 7** If \(\{X_n\}\) is uniformly bounded in \(L_p\) for some \(p > 1\) then \(\{X_n\}\) is UI.
Proof. Let $B \in \mathbb{R}_+$ be an upper bound for $E|X_n|^p$ and let $q = \frac{p}{p-1}$. By Hölder’s inequality

$$
\mathbb{E} \left[ |X_n| 1_{\{|X_n| > t\}} \right] \leq \|X_n\|_p \|1_{\{|X_n| > t\}}\|_q
$$
$$
\leq B \left\{ \mathbb{P}|X_n| > t \right\}^{1/q}
$$
$$
\leq B \left\{ \mathbb{E} \left[ |X_n|^p / t^p \right] \right\}^{1/q}
$$
$$
= B \left\{ \|X_n\|^p / t^p \right\}^{1/q}
$$
$$
= B \|X_n\|^{p-1} / t^{p-1}
$$
since $p/q = p-1$
$$
\leq B^p t^{1-p} \to 0
$$

uniformly in $\{n\}$.

Theorem 8 The random variables $\{X_n\}$ are UI if and only if they are uniformly bounded in $L_1$ and

$$(\forall \epsilon > 0)(\exists \delta > 0)(\forall A \in \mathcal{F} \ w/ \mathbb{P}(A) < \delta) \ \mathbb{E}[|X_n| 1_A] < \epsilon.$$ 

If $(\Omega, \mathcal{F}, \mathbb{P})$ is non-atomic, the condition “$\{X_n\}$ is uniformly bounded in $L_1$” is unnecessary.

Proof. Straightforward. If $(\Omega, \mathcal{F}, \mathbb{P})$ is non-atomic, for each $\epsilon > 0$ find $\delta > 0$ by hypothesis and then cover $\Omega$ with $[1/\delta]$ sets $A_i \in \mathcal{F}$ with $\mathbb{P}(A_i) < \delta$ to see $\mathbb{E}|X_n| < \epsilon [1/\delta]$ uniformly. Try to offer an example to illustrate what can go wrong if $\mathbb{P}$ has an atom $\omega^* \in \Omega$ with $c = \mathbb{P}([\omega^*]) > 0$.

Theorem 9 Let $X_n \to X$ i.p. on $(\Omega, \mathcal{F}, \mathbb{P})$. Then the following are equivalent:

1. $\{X_n\}$ is UI
2. $X_n \to X$ in $L_1$
3. $E|X_n| \to E|X|

Proof. Since 1 $\Rightarrow$ 2 by Theorem 6 and 2 $\Rightarrow$ 3 by the triangle inequality, it remains only to show that 3 $\Rightarrow$ 1. Let $X_n \to X$ i.p. and $E|X_n| \to E|X|$. Fix $\epsilon > 0$ and, for each $t > 0$, set

$$
fx(x) = \begin{cases} 
|\epsilon| & 0 \leq |\epsilon| \leq t \\
t(t + 1 - |\epsilon|) & t < |\epsilon| \leq t + 1 \\
t + 1 \leq |\epsilon| < \infty
\end{cases}
$$
and note that $0 \leq |x| |1_{|x| \leq t}| \leq f_t(x) \leq |x| |1_{|x| \leq t+1}|$ and $f_t(x) \leq t$. Since $f_t$ is continuous and bounded, $f_t(X_n) \rightarrow f_t(X)$ in probability by Theorem 5 and also in $L_1$ by DCT. For large enough $n$ (say, $n \geq N'$), $||f_t(X_n) - f_t(X)||_1 \leq \epsilon$ and so

$$E[|X_n|1_{|X_n| \leq t+1}] \geq E[f_t(X_n)] \geq E[f_t(X)] - \epsilon \geq E[|X|1_{|X| \leq t}] - \epsilon. \quad (1)$$

Since $E[|X_n|] \rightarrow E[|X|]$, for sufficiently large $n$ (say, $n \geq N \geq N'$)

$$E[|X_n|] \leq E[|X|] + \epsilon. \quad (2)$$

Upon subtracting (1) from (2), we have

$$E[|X_n|1_{|X_n| > t+1}] \leq E[|X|1_{|X| > t}] + 2\epsilon.$$

Since $X \in L_1$, $E[|X|1_{|X| > t}] \rightarrow 0$ as $t \rightarrow \infty$ by DCT so we can find $T$ sufficiently large that when $t > T$ it is smaller than $\epsilon$, whereupon $n \geq N$ and $t \geq T$ imply

$$E[|X_n|1_{|X_n| > t+1}] \leq 3\epsilon$$

completing the proof that $\{X_n\}$ is UI.

$\Box$

### 6.3 Cauchy Convergence

Every form of convergence we have considered for random variables except a.s. can be represented as convergence in a metric space. A sequence $X_n$ converges to a limit $X$ in a metric space if every ball centered at $X$ contains all but finitely-many of $\{X_n\}$. Sometimes we wish to consider a sequence $X_n$ that converges to some limit, without knowing the limit in advance. The concept of **Cauchy Convergence** is ideal for this—we insist that for each $\epsilon > 0$, all but finitely-many of the $\epsilon$-balls centered at points $X_m$ of the sequence contain all but finitely-many of the points. For any of distance measures $d_p$ of Section (6.1), with $0 \leq p \leq \infty$, say “$X_n$ is a Cauchy sequence in $L_p$” if

$$(\forall \epsilon > 0)(\exists N_\epsilon < \infty)(\forall m, n \geq N_\epsilon) \quad d_p(X_m, X_n) < \epsilon.$$ 

The spaces $L_p$ for $0 \leq p \leq \infty$ are all complete in the sense that if $X_n$ is Cauchy for $d_p$ then there exists $X \in L_p$ for which $d_p(X_n, X) \rightarrow 0$. To see this, take an increasing subsequence $N_k$ along which $d_p(X_m, X_n) < 2^{-k}$ for $n \geq m \geq N_k$, and set $X_0 \equiv 0$ and $N_0 = 0$; set $Y_k \equiv X_{N_k} - X_{N_{k-1}}$. Check to confirm that $X \equiv \sum_{k=1}^{\infty} Y_k$ is well-defined and $X \in L_p$ a.s., and that $d_p(X_n, X) \rightarrow 0$. 

Page 11 December 14, 2011
6.4 Summary: UI and Convergence Concepts

I. Uniform Integrability (UI)

A. \(|X_n| \leq Y \in L_r, r > 0\), implies \(\int_{[|X_n| \geq t]} |X_n| r \, dP \leq \int_{[Y \geq t]} Y^r \, dP \to 0 \) as \(t \to \infty\), uniformly in \(n\). This is the definition of UI for \(r = 1\).

B. If \((\Omega, \mathcal{F}, P)\) nonatomic, \(X_n\) UI iff \((\forall \epsilon)(\exists \delta) \supset P[|X_n| \geq \delta] \leq \epsilon \) for \(P[A] \leq \delta \) (take \(\delta = \frac{\epsilon}{\epsilon^p}\)). If \((\Omega, \mathcal{F}, P)\) has atoms, also need \(E|X_n| \leq B\).

C. \(E|X_n|^p \leq c^p < \infty\) implies \(|X_n|^p\) UI for each \(r < p\). \(\delta = (\epsilon/c)^q\), \(\frac{1}{p} + \frac{1}{q} = 1\), \(q = \frac{p}{p-1}\).

1. Remark: not for \(r = p\) (counterexample: \(X_n = n1_{(0,1/n]}\))

D. Main result: If \(X_n \to X\) i.p., then \(|X_n|^p\) UI \(\iff X_n \to X\text{ in }L_p \iff E|X_n|^p \to E|X|^p\).

II. Convergence in Distribution (aka Vague Convergence)

A. \(X_n \to X\) i.p. iff \((\forall n_k)(\exists n_k) \supset X_{n_k} \to X\) a.s. (by contradiction)

B. \(X_n \to X\text{ a.s. and }\phi(x)\text{ continuous implies }\phi(X_n) \to \phi(X)\text{ a.s.}

C. \(X_n \to X\) i.p. and \(\phi(x)\text{ continuous implies }\phi(X_n) \to \phi(X)\) i.p. (use A)

D. Definition: \(X_n \Rightarrow X\) if \((\forall \phi \in C_0(\mathbb{R}))\supset E\phi(X_n) \to E\phi(X)\)

1. Prop: \(X_n \Rightarrow X\text{ implies }X_n \Rightarrow X\) (use II.C)

2. Prop: \(X_n \Rightarrow X\text{ implies }F_n(r) \to F(r)\text{ where }F(r) = F(r-).

   a. Remark: Even if \(X_n \Rightarrow X\), \(F_n(r)\text{ may not converge where }F(r)\text{ jumps;}

   b. Remark: Even if \(X_n \Rightarrow X\), \(f_n(r) = F_n'(r)\text{ may not converge to }f(r) = F'(r);\text{ in fact, either may fail to exist.}

III. Implications among various notions: a.s., i.p., \(L_p, L_q, L_\infty\), dist. \((0 < p < q < \infty)\):

A. a.s. \(\Rightarrow\) i.p. (by Easy Borel-Cantelli)

   1. i.p. \(\Rightarrow\) a.s. along subsequences

   2. i.p. \(\neq\) a.s. (counteg: \(X_n(\omega) = 1_{(i/2^j,(i+1)/2^j]}(\omega)\text{, }n = i + 2^j\))

B. \(L_p \Rightarrow\) i.p. (by Markov’s inequality)

   1. i.p. \(\Rightarrow\) \(L_p\) under Uniform Integrability

   2. i.p. \(\neq\) \(L_p\) (counterexample: \(X_n = n^{1/p}1_{(0,1/n]}\))

C. \(L_q \Rightarrow\) \(L_p\) for \(p < q\) (by Jensen’s inequality)
1. \( L_p \not= L_q \) (counterexample: \( X_n = n^{1/q} 1_{[0,1/n]} \))

D. \( L_\infty \implies L_p \) (simple estimate, or \( \|X\|_p \uparrow \) as \( p \uparrow \) by Jensen)

1. \( L_p \not= L_\infty \) (counterexample: \( X_n = 1_{[0,1/n]} \))

E. \( L_\infty \implies a.s. \) (uniform cge implies pointwise cge)

F. \( i.p \implies dist. \) (II.D.1 above)

1. \( dist. \not= i.p. \) (counterexample: \( X_n, X \) on different spaces)

2. \( dist. \implies a.s. \) (\( \exists (\Omega, \mathcal{F}, P), X_n, X \ni X_n \rightarrow X \) a.s....)