# Probability and Measure

Robert L. Wolpert
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA

Convergence of Random Variables

## 1 Convergence Concepts

### 1.1 Convergence of Real Numbers

A sequence of real numbers $a_n$ 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 $a_n \to a$ 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 (complete) *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; and, since there are only finitely many of them, 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).

## 2 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 asking about the convergence of a sequence $X_n$ of random variables to a limit $X$? Should we mean that $X_n(\omega)$ converges to $X(\omega)$ for each fixed $\omega$? Or that these sequences converge *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 $\mathsf{P}$ 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 $(\Omega, \mathcal{F}, \mathsf{P})$:

*pw:* The sequence of real numbers $X_n(\omega) \to X(\omega)$ for every $\omega \in \Omega$ (pointwise convergence):

$$(\forall \epsilon > 0)\,(\forall \omega \in \Omega)\,(\exists N_{\epsilon,\omega} < \infty)\,(\forall n \geq N_{\epsilon,\omega}) \quad |X_n(\omega) - X(\omega)| < \epsilon.$$

*uni:* The sequences of real numbers $X_n(\omega) \to X(\omega)$ *uniformly* for $\omega \in \Omega$:

$$(\forall \epsilon > 0)\,(\exists N_{\epsilon} < \infty)\,(\forall \omega \in \Omega)\,(\forall n \geq N_{\epsilon}) \quad |X_n(\omega) - X(\omega)| < \epsilon.$$

*a.s.*.: Outside some null event $\mathcal{N} \in \mathcal{F}$, each sequence of real numbers $X_n(\omega) \to X(\omega)$ (Almost-Sure convergence, or "almost everywhere" (*a.e.*)): for some $\mathcal{N} \in \mathcal{F}$ with $\mathsf{P}[\mathcal{N}] = 0$,

$$(\forall \epsilon > 0)\,(\forall \omega \notin \mathcal{N})\,(\exists N_{\epsilon,\omega} < \infty)\,(\forall n \geq N_{\epsilon,\omega}) \quad |X_n(\omega) - X(\omega)| < \epsilon,$$

*i.e.*,

$$\mathsf{P}\big\{ \cup_{\epsilon>0} \cap_{N<\infty} \cup_{n \geq N} \quad |X_n(\omega) - X(\omega)| \geq \epsilon \big\} = 0.$$

$L_\infty$: Outside some null event $\mathcal{N} \in \mathcal{F}$, the sequences of real numbers $X_n(\omega) \to X(\omega)$ converge *uniformly* ("almost-uniform" or "$L_\infty$" convergence): for some $\mathcal{N} \in \mathcal{F}$ with $\mathsf{P}[\mathcal{N}] = 0$,

$$(\forall \epsilon > 0)\,(\exists N_{\epsilon} < \infty)\,(\forall \omega \notin \mathcal{N})\,(\forall n \geq N_{\epsilon}) \quad |X_n(\omega) - X(\omega)| < \epsilon.$$

*i.p.*: For each $\epsilon > 0$, the probabilities $\mathsf{P}[|X_n - X| > \epsilon] \to 0$ (convergence "in probability", or "in measure"):

$$(\forall \epsilon > 0)\,(\forall \eta > 0)\,(\exists N_{\epsilon,\eta} < \infty)\,(\forall n \geq N_{\epsilon,\eta}) \quad \mathsf{P}[|X_n - X| > \epsilon] < \eta.$$

$L_1$: The expectation $\mathsf{E}[|X_n - X|]$ converges to zero (convergence "in $L_1$"):

$$(\forall \epsilon > 0)\,(\exists N_\epsilon < \infty)\,(\forall n \geq N_\epsilon) \quad \mathsf{E}[|X_n - X|] < \epsilon.$$

$L_p$: For some fixed number $p > 0$, the expectation of the $p^{th}$ power $\mathsf{E}[|X_n - X|^p]$ converges to zero (convergence "in $L_p$," sometimes called "in the $p^{th}$ mean"):

$$(\forall \epsilon > 0)\,(\exists N_\epsilon < \infty)\,(\forall n \geq N_\epsilon) \quad \mathsf{E}[|X_n - X|^p] < \epsilon.$$

*i.d.*: The *distributions* of $X_n$ converge to the distribution of $X$, *i.e.*, the measures $\mathsf{P} \circ X_n^{-1}$ converge in some way to $\mathsf{P} \circ X^{-1}$ ("vague" or "weak" convergence, or "convergence in distribution", sometimes written $X_n \Rightarrow X$):

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

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 $\mathsf{P}$; it is possible for a sequence $X_n$ to converge to $X$ in any of these senses for one probability measure $\mathsf{P}$, but to fail to converge for another $\mathsf{P}'$. Most of them can be phrased as metric convergence for some notion of distance between random variables:

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

$$d_0(X, Y) \equiv \mathsf{E}\left(\frac{|X - Y|}{1 + |X - Y|}\right)$$

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

$$||X - Y||_1 \equiv \mathsf{E}|X - Y|$$

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

$$||X - Y||_p \equiv (\mathsf{E}|X - Y|^p)^{1/p}$$

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

$$||X - Y||_\infty = \text{l.u.b.}\{r < \infty : \mathsf{P}[|X - Y| > r] > 0\}$$

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

## 2.1 Almost-Sure Convergence

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

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

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:

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

The union over $\epsilon > 0$ is only a countable one, 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 $\epsilon > 0$,

$$\mathsf{P}\Big[\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} [\omega : |X_n(\omega) - X(\omega)| > \epsilon]\Big] = 0. \qquad (a.e.)$$

This combination of intersection and union occurs frequently in probability, and has a name; for any sequence $E_n$ of events, $[\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n]$ is called the *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 *lim inf* of the sets $F_n = E_n^c$, $[\bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} F_n]$: the set of points in all but finitely many of the $F_n$. Since $\mathsf{P}$ is countably additive, and since the intersection in the definition of *lim sup* is *decreasing* and the union in the definition of *lim inf*

is *increasing*, always we have
$\mathsf{P}[\bigcup_{n=m}^{\infty} E_n] \searrow \mathsf{P}[\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} E_n]$ and $\mathsf{P}[\bigcap_{n=m}^{\infty} F_n] \nearrow \mathsf{P}[\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty} F_n]$ as $m \to \infty$. Thus,

**Theorem 1** *$X_n \to X$ $\mathsf{P}$-a.s.. if and only if for every $\epsilon > 0$,*

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

In particular, $X_n \to X$ $\mathsf{P}$-*a.s.*. if $\sum \mathsf{P}[|X_n - X| > \epsilon] < \infty$ for each $\epsilon > 0$ (why?).

## 2.2 Convergence In Probability

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

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

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

**Theorem 2** *If $X_n \to X$ a.e. 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.e.*

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

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

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

$$\begin{aligned}
\mathsf{P}\Big[\bigcup_{k=m}^{\infty}[\omega : |X_{n_k}(\omega) - X(\omega)| > \epsilon]\Big] &\le \mathsf{P}\Big[\bigcup_{k=m}^{\infty}[\omega : |X_{n_k}(\omega) - X(\omega)| > \frac{1}{k}]\Big] \\
&\le \sum_{k=m}^{\infty}\mathsf{P}[\omega : |X_{n_k}(\omega) - X(\omega)| > \frac{1}{k}] \\
&\le \sum_{k=m}^{\infty} 2^{-k} = 2^{1-m} \to 0 \text{ as } m \to \infty.
\end{aligned}$$

□

### 2.2.1 A Counter-Example

If $X_n \to X$ *a.e.* 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.e.*? The answer is that the two notions *are* different, and that *a.e.* convergence is strictly stronger than convergence *i.p.* Here's an example:

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be the unit interval with Borel sets and Lebesgue measure. Define a sequence of random variables $X_n : \Omega \to \mathbb{R}$ by

$$X_n(\omega) = \begin{cases} 1 & \text{if } \frac{i}{2^j} < \omega \le \frac{i+1}{2^j} \\ 0 & \text{otherwise} \end{cases} \qquad \text{where } n = i + 2^j,\ 0 \le i < 2^j.$$

Each $X_n$ is one on an interval of length $2^{-j}$, where $j = \lfloor \log_2(n) \rfloor$; since $\frac{1}{n} \le \frac{1}{2^j} < \frac{2}{n}$,

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

for each $0 < \epsilon < 1$ and $X_n \to 0$ *i.p.* On the other hand, for every $j > 0$ we have

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

so $[\omega : X_n(\omega) \to 0]$ is *empty*, not a set of probability one! Obviously $X_n$ does not converge *a.e.* This example is a building-block for several examples to come, so getting to know it well is worth while. Try to verify that $X_n \to 0$ in probability and in $L_p$ but *not* almost surely. What *is* $||X_n||_p$? Why *doesn't* $X_n \to 0$ *a.s.*? What would happen if we multiplied $X_n$ by $n$? By $n^2$? What about the subsequence $Y_n = X_{2^n}$? Does $X_n$ converge in $L_\infty$?

## 3 Cauchy Convergence

Sometimes we wish to consider a sequence $X_n$ that converges to *some* limit $X$, perhaps without knowing $X$ in advance; the concept of *Cauchy Convergence* is ideal for this. For any of the distance measures $d_p$ above, with $0 \le p \le \infty$, say "$X_n$ is a Cauchy sequence in $L_p$" if

$$(\forall \epsilon > 0)(\exists N < \infty)(\forall n \ge m \ge N) \qquad d_p(X_m, X_n) < \epsilon.$$

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

## 4 Uniform Integrability

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

$$\mathsf{E}[Y] = \int_\Omega Y\, dP < \infty;$$

it follows (from DCT or MCT, for example) that

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

and, consequently, that for any sequence of random variables $X_n$ "dominated" by $Y$ in the sense that $|X_n| \le Y$ *a.s.*,

$$\begin{aligned}\lim_{t\to\infty} \mathsf{E}[X_n\, 1_{[X_n>t]}] &= \int_{[\omega: X_n(\omega)>t]} X_n\, dP \\ &\le \int_{[\omega: Y(\omega)>t]} Y\, dP \\ &= 0, \text{ uniformly in } n.\end{aligned}$$

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

**Theorem 4** *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 $\mathsf{E}[|X_n|\, 1_{[X_n>t_\epsilon]}] \le \epsilon$ for all $n$. Now find (by $X_n \to X$ *i.p.*) $N_\epsilon \in \mathbb{N}$ such that, for $n \ge N_\epsilon$, $\mathsf{P}[|X_n| > \epsilon] < \epsilon/t_\epsilon$; then:

$$
\begin{aligned}
E[|X_n|] &= \int_{[|X_n|\le\epsilon]} |X_n|\, d\mathsf{P} + \int_{[\epsilon<|X_n|\le t_\epsilon]} |X_n|\, d\mathsf{P} + \int_{[t_\epsilon<|X_n|]} |X_n|\, d\mathsf{P} \\
&\le \int_{[|X_n|\le\epsilon]} \epsilon\, d\mathsf{P} + \int_{[\epsilon<|X_n|\le t_\epsilon]} t_\epsilon\, d\mathsf{P} + \int_{[t_\epsilon<|X_n|]} |X_n|\, d\mathsf{P} \\
&\le \epsilon + (t_\epsilon)\mathsf{P}[|X_n| > \epsilon] + \epsilon \\
&\le 3\epsilon.
\end{aligned}
$$

□

Similarly, for any $p > 0$, $X_n \to X$ (*i.p.*) and $|X_n|^p$ UI (for example, $|X_n| \le Y \in L_p$) gives $X_n \to X$ $(L_p)$. In the special case of $|X_n| \le Y \in L_p$ this is just Lebesgue's **Dominated Convergence Theorem** (DCT).

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

**Theorem 5** *If $\{X_n\}$ is uniformly bounded in $L_p$ for some $p > 1$ then $\{X_n\}$ is UI.*

**Proof.** Let $c \in \mathbb{R}_+$ be an upper bound for $\mathsf{E}|X_n|^p$. First recall that, by Fubini's Theorem, any random variable $X$ satisfies for any $q > 0$

$$\mathsf{E}|X|^q = \int_0^\infty q\, x^{q-1}\, \mathsf{P}[|X| > x]\, dx.$$

We apply this for $q = 1$ and $q = p$ to the random variables $X_n$ and, for $t > 0$, to $X_n 1_{|X_n|>t}$. Fix any $t > 0$; then

$$
\begin{aligned}
\mathsf{E}\big[|X_n|\,1_{|X_n|>t}\big] &= \int_0^\infty \mathsf{P}[|X_n|\,1_{|X_n|>t} > x]\,dx \\
&= \int_0^t \mathsf{P}[|X_n| > t]\,dx + \int_t^\infty \mathsf{P}[|X_n| > x]\,dx \\
&\le t\,\mathsf{P}[|X_n|^p > t^p] + \int_t^\infty \frac{p\,x^{p-1}}{p\,t^{p-1}} \mathsf{P}[|X_n| > x]\,dx \\
&\le t\,\frac{\mathsf{E}|X_n|^p}{t^p} + \frac{1}{p\,t^{p-1}} \int_0^\infty p\,x^{p-1} \mathsf{P}[|X_n| > x]\,dx \\
&= t^{1-p}\,(1 + p^{-1})\mathsf{E}|X_n|^p \\
&\le c\,t^{1-p}\,(1 + p^{-1}) \\
&\to 0 \text{ as } t \to \infty, \text{ uniformly in } n.
\end{aligned}
$$

□

**Theorem 6** *If $\{X_n\}$ is UI, then*

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

*Conversely, if $\{X_n\}$ is uniformly bounded in $L_1$ and if $(\forall \epsilon > 0)(\exists \delta > 0)$ such that $\mathsf{E}[|X_n| 1_A] < \epsilon$ whenever $\mathsf{P}[A] < \delta$, then $\{X_n\}$ is UI.*

**Proof.** Straightforward. The condition "$\{X_n\}$ is uniformly bounded in $L_1$" is unnecessary if $(\Omega, \mathcal{F}, \mathsf{P})$ is non-atomic.

# 5 Summary: Uniform Integrability and Convergence Concepts

I. Uniform Integrability (UI)

   A. $|X_n| < Y \in L_r$, $r > 0$, implies $\int_{[|X_n|>t]} |X_n|^r \, d\mathsf{P} \le \int_{[Y>t]} Y^r \, d\mathsf{P} \to 0$ as $t \to \infty$, uniformly in $n$. This is the *definition* of UI.

   B. If $(\Omega, \mathcal{F}, \mathsf{P})$ nonatomic, $X_n$ UI iff $\forall \epsilon \exists \delta \ni \int_\Lambda |X_n| \, d\mathsf{P} < \epsilon$ for $\mathsf{P}[\Lambda] \le \delta$ (take $\delta = \epsilon/2t$)
   [If $(\Omega, \mathcal{F}, \mathsf{P})$ has atoms must also require $\mathsf{E}|X_n| \le B$]

   C. $\mathsf{E}|X_n|^p \le c^p < \infty$ implies $|X_n|^r$ 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. (Chung Thm 4.5.4≈SIR Thm 6.6.2): If $X_n \to X$ *i.p.* then $|X_n|^r$ UI iff $X_n \to X$ in $L_r$ iff $\mathsf{E}|X_n|^r \to \mathsf{E}|X|^r$.

II. Vague Convergence

   A. $X_n \to X$ *i.p.* iff $\forall n_k \exists n_{k_i} \ni X_{n_{k_i}} \to X$ *a.e.* (by contradiction)

   B. $X_n \to X$ *a.s.*. and $\phi(x)$ continuous implies $\phi(X_n) \to \phi(X)$ *a.s.*.

   C. $X_n \to X$ *i.p.* and $\phi(x)$ continuous implies $\phi(X_n) \to \phi(X)$ *i.p.* (use A)

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

      1. Prop: $X_n \to X$ *i.p.* implies $X_n \Rightarrow X$ (use II.C)

      2. Prop: $X_n \Rightarrow X$ implies $F_n(r) \to F(r)$ wherever $F(r) = F(r-)$.

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

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

III. Implications among these notions: *a.e.*, *i.p.*, $L_r$, $L_p$, $L_\infty$, *i.d.* ($0<r<p<\infty$):

   A. *a.e.* $\Longrightarrow$ *i.p.* (by Easy Borel-Cantelli)

      1. *i.p.* $\Longrightarrow$ *a.e.* along subsequences

      2. *i.p.* $\not\Longrightarrow$ *a.e.* (counterexample: $X_n(\omega) = 1_{(i/2^j,(i+1)/2^j]}(\omega)$, $n = i + 2^j$)

   B. $L_p \Longrightarrow$ *i.p.* (by Chebychev's inequality)

      1. *i.p.* $\Longrightarrow L_p$ under Uniform Integrability

2. $i.p. \not\Rightarrow L_p$ (counterexample: $X_n = n^{1/p} 1_{(0,1/n]}$)

C. $L_p \Longrightarrow L_r$ (by Jensen's inequality)

1. $L_r \not\Rightarrow L_p$ (counterexample: $X_n = n^{1/p} 1_{(0,1/n]}$)

D. $L_\infty \Longrightarrow L_p$ (simple estimate)

1. $L_p \not\Rightarrow L_\infty$ (counterexample: $X_n = n^{1/2p} 1_{(0,1/n]}$)

E. $L_\infty \Longrightarrow a.e.$ (uniform cgce implies pointwise cgce)

F. $i.p. \Longrightarrow i.d.$ (II.D.1 above)

1. $i.d. \not\Rightarrow i.p.$ (counterexample: $X_n$, $X$ on different spaces)
2. $i.d. \Longrightarrow a.s..$ ($\exists (\Omega, \mathcal{F}, \mathsf{P}),\ X_n, X \ni X_n \to X\ \ a.e.$...)

# 6 Infinite Coin-Toss and the Laws of Large Numbers

The traditional interpretation of the *probability* of an event $E$ is its *asymptotic frequency*: the limit as $n \to \infty$ of the fraction of $n$ repeated, similar, and independent trials in which $E$ occurs. Similarly the "expectation" of a random variable $X$ is taken to be its *asymptotic average*, the limit as $n \to \infty$ of the average of $n$ repeated, similar, and independent replications of $X$. As statisticians trying to make inference about the underlying probability distribution $f(x|\theta)$ governing observed random variables $X_i$, this suggests that we should be interested in the probability distribution for large $n$ of quantities like the average of the RV's, $\frac{1}{n}\sum_{i=1}^{n} X_i$.

Three of the most celebrated theorems of probability theory concern this sum. For independent random variables $X_i$, all with the same probability distribution satisfying $\mathsf{E}|X_i|^3 < \infty$, set $\mu = \mathsf{E}X_i$, $\sigma^2 = \mathsf{E}|X_i - \mu|^2$, and $S_n = \sum_{i=1}^{n} X_i$. The three main results are:

**Laws of Large Numbers:**

$$\frac{S_n - n\mu}{\sigma n} \longrightarrow 0 \qquad (i.p. \text{ and } a.s..)$$

**Central Limit Theorem:**

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \Longrightarrow N(0,1) \qquad (i.d.)$$

**Law of the Iterated Logarithm:**

$$\limsup \pm \frac{S_n - n\mu}{\sigma\sqrt{2n\log\log n}} = 1.0 \qquad (a.s..)$$

Together these three give a clear picture of how quickly and in what sense $\frac{1}{n}S_n$ tends to $\mu$. We begin with the Law of Large Numbers (LLN), in its "weak" form (asserting convergence *i.p.*) and in its "strong" form (convergence *a.s.*.). There are several versions of both theorems. The simplest requires the $X_i$ to be IID and $L_2$; stronger results allow us to weaken (but not eliminate) the independence requirement, permit non-identical distributions, and consider what happens if the RV's are only $L_1$ (or worse!) instead of $L_2$.

The text covers these things well; to complement it I am going to: (1) Prove the simplest version, and with it the Borel-Cantelli theorems; and (2)

Show what happens with Cauchy random variables, which don't satisfy the requirements (the LLN fails).

I. Weak version, non-iid, $L_2$: $\mu_i = \mathsf{E}X_i$, $\sigma_{ij} = \mathsf{E}[X_i - \mu_i][X_j - \mu_j]$
   A. $Y_n = (S_n - \Sigma\mu_i)/n$ satisfies $\mathsf{E}Y_n = 0$, $\mathsf{E}Y_n^2 = \frac{1}{n^2}\Sigma_i\sigma_{ii} + \frac{2}{n^2}\Sigma_{i<j}\sigma_{ij}$;
      1. If $\sigma_{ii} \le M$ and $\sigma_{ij} \le 0$, Chebychev $\Longrightarrow Y_n \to 0$, *i.p.*
      2. (pairwise) IID $L_2$ is OK

II. Strong version, non-iid, $L_2$: $\mathsf{E}X_i = 0$, $\mathsf{E}X_i^2 \le M$, $\mathsf{E}X_iX_j \le 0$.
   A. $\mathsf{P}[|S_n| > n\epsilon] < \frac{Mn}{n^2\epsilon^2} = \frac{M}{n\epsilon^2}$
      1. $\mathsf{P}[|S_{n^2}| > n^2\epsilon] < \frac{M}{n^2\epsilon^2}$, $\Sigma_n\mathsf{P}[|S_{n^2}| > n^2\epsilon] < \frac{M\pi^2}{6\epsilon^2}$
      2. Borel-Cantelli: $\mathsf{P}[|S_{n^2}| > n^2\epsilon \ i.o.] = 0$, $\frac{1}{n^2}S_{n^2} \to 0 \ a.s.$.
      3. $D_n = \max_{n^2 \le k < (n+1)^2} |S_k - S_{n^2}|$, $\mathsf{E}D_n^2 \le 2n\mathsf{E}|S_{(n+1)^2} - S_{n^2}| \le 4n^2M$
      4. Chebychev: $\mathsf{P}[D_n > n^2\epsilon] < \frac{4n^2M}{n^4\epsilon^2}$, $D_n \to 0 \ a.s.$.
   B. $|S_k/k| \le \frac{|S_{n^2}| + D_n}{n^2} \to 0 \ a.s.$., QED
      1. Bernoulli RV's, normal number theorem, Monte Carlo.

III. Weak version, pairwise-iid, $L_1$
   A. Equivalent sequences: $\sum_n \mathsf{P}[X_n \ne Y_n] < \infty$
      1. $\sum_n [X_n - Y_n] < \infty \ a.s.$.
      2. $\sum_{i=1}^n [X_i]$, $a_n \sum_{i=1}^n [X_i]$ converge iff $\sum_{i=1}^n [Y_i]$, $a_n \sum_{i=1}^n [Y_i]$ do
      3. $Y_n = X_n 1_{[|X_n| \le n]}$

IV. Counterexamples: Cauchy,
   A. $X_i \sim \frac{dx}{\pi[1+x^2]} \Longrightarrow \mathsf{P}[|S_n|/n \le \epsilon] \equiv \frac{2}{\pi}\tan^{-1}(\epsilon) \not\to 1$, WLLN fails.
   B. $\mathsf{P}[X_i = n] = \frac{\pm c}{n^2}$, $n \ge 1$; $X_i \notin L_1$, and $S_n/n \not\to 0$ *i.p.* or *a.s.*.
   C. $\mathsf{P}[X_i = n] = \frac{\pm c}{n^2 \log n}$, $n \ge 3$; $X_i \notin L_1$, but $S_n/n \to 0$ *i.p.* and not *a.s.*.

   D. Medians: for ANY RV's $X_n \to X_\infty$ *i.p.*, then $m_n \to m_\infty$ if $m_\infty$ is unique. Let $X_i$ be iid standard Cauchy RV's, with

$$\mathsf{P}[X_1 \le t] = \int_{-\infty}^{t} \frac{dx}{\pi[1+x^2]} = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$$

and characteristic function

$$\mathsf{E}\, e^{i\lambda X_1} = \int_{-\infty}^{\infty} e^{i\lambda x}\frac{dx}{\pi[1+x^2]} = e^{-|\lambda|},$$

so $S_n/n$ has characteristic function

$$\mathsf{E}\, e^{i\lambda S_n/n} = \mathsf{E}\, e^{i\frac{\lambda}{n}[X_1+\cdots+X_n]} = \left(\mathsf{E}\, e^{i\frac{\lambda}{n}X_1}\right)^n = (e^{-|\frac{\lambda}{n}|})^n = e^{-|\lambda|}$$

and $S_n/n$ also has the standard Cauchy distribution with $\mathsf{P}[S_n/n \le t] = \frac{1}{2} + \frac{1}{\pi}\arctan(t)$; in particular, $S_n/n$ does not converge almost surely, or even in probability.