CONVERGENCE CONCEPTS: *a.e.*, *in pr.*, L^1, L^P

Convergence of Real Numbers

A sequence of real numbers a_n converges to a limit a if and only if, for each $\epsilon > 0$, 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 ϵ 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.

CONVERGENCE OF RANDOM VARIABLES

For *random variables* X_n the idea of convergence to a limiting random variable X isn't so clear, since each X_n is a *function* of $\omega \in \Omega$. We might consider convergence of the numerical sequences $X_n(\omega)$, for each $\omega \in \Omega$, or we might think of convergence of the random variables themselves in some sense— of their expectations or distributions, for example. 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 $\mathsf{P}[\mathcal{N}] = 0$,

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

L^∞: 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^∞" convergence): for some $\mathcal{N} \in \mathcal{F}$ with $\mathsf{P}[\mathcal{N}] = 0$,

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

pr.: 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"):

$$(\forall \epsilon > 0)\,(\exists N_\epsilon < \infty)\,(\forall n \geq N_\epsilon) \quad \mathsf{E}[|X_n - X|^p]^{\frac{1}{p}} < \epsilon.$$

vg.: 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 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'.

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 ω 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 ϵ 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 ω 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 ϵ (or, for that matter, any sequence ϵ_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 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. *$X_n \to X$ P-a.e. if and only if for every $\epsilon > 0$,*

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

Convergence In Probability

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

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

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

Theorem. *If $X_n \to X$ a.e. then $X_n \to X$ in pr.* Here is a partial converse:

Theorem. *If $X_n \to X$ in pr., 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 $k \ge 1$, set

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

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

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

A Counter-Example

If $X_n \to X$ *a.e.* implies $X_n \to X$ *in pr.*, 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 *in pr.*, but not *a.e.*? The answer is that the two notions *are* different, and that *a.e.* convergence is strictly stronger than convergence *in pr.* 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 $\frac{1}{n} \le \frac{1}{2^j} < \frac{2}{n}$,

$$\mathsf{P}[|X_n| > \epsilon] = \frac{1}{2^j} < \frac{2}{n} \to 0$$

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

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

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.

UNIFORM INTEGRABILITY AND CONVERGENCE CONCEPTS

I. Uniform Integrability (UI)
  - A. $|X_n| < Y \in L^p$, $p > 0$, implies $\int_{[|X_n|>t]} |X_n|^p \, d\mathsf{P} \le \int_{[Y>t]} Y^p \, d\mathsf{P} \to 0$ as $t \to \infty$, uniformly in n. This is one *definition* of UI.
  - B. X_n UI iff $\forall \epsilon \exists \delta \ni \int_\Lambda |X_n| \, d\mathsf{P} < \epsilon$ for $\mathsf{P}[\Lambda] \le \delta$ (take $\delta = \epsilon/2t$)
  - C. $\mathsf{E}|X_n|^r \le c < \infty$ implies $|X_n|^p$ UI for each $p < r$ (take $t = (c/\epsilon)^{1/(r-p)}$).
    1. Remark: *not* for $p = r$ (counterexample: $X_n = 2^n 1_{(0, 2^{-n}]}$)
  - D. Main result: If $X_n \to X$ *in pr.* then $|X_n|^p$ UI iff $X_n \to X$ in L^p iff $\mathsf{E}|X_n|^p \to \mathsf{E}|X|^p$.

II. Vague Convergence
  - A. $X_n \to X$ *in pr.* 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. $X_n \to X$ *in pr.* and $\phi(x)$ continuous implies $\phi(X_n) \to \phi(X)$ *in pr.* (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: $X_n \to X$ *in pr.* 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.*, *in pr.*, L^r, L^p, L^∞, *vg.* ($0<p<r<\infty$):

A. *a.e.* $\Longrightarrow$ *in pr.* (by Easy Borel-Cantelli)
   1. *in pr.* $\Longrightarrow$ *a.e.* along subsequences
   2. *in pr.* $\not\Longrightarrow$ *a.e.* (counterexample: $X_n(\omega) = 1_{(\frac{i}{2^j}, \frac{i+1}{2^j}]}(\omega)$, $n = i + 2^j$)

B. $L^p \Longrightarrow$ *in pr.* (by Chebychev's inequality)
   1. *in pr.* $\Longrightarrow L^p$ under Uniform Integrability
   2. *in pr.* $\not\Longrightarrow L^p$ (counterexample: $X_n = 2^{n/p} 1_{(0,2^{-n}]}$)

C. $L^r \Longrightarrow L^p$ (by Jensen's inequality)
   1. $L^p \not\Longrightarrow L^r$ (counterexample: $X_n = 2^{n/r} 1_{(0,2^{-n}]}$)

D. $L^\infty \Longrightarrow L^p$ (simple estimate)
   1. $L^p \not\Longrightarrow L^\infty$ (counterexample: $X_n = 2^{n/2p} 1_{(0,2^{-n}]}$)

E. $L^\infty \Longrightarrow$ *a.e.* (uniform cgce implies pointwise cgce)

F. *in pr.* $\Longrightarrow$ *vg.* (II.D.1 above)
   1. *vg.* $\not\Longrightarrow$ *in pr.* (counterexample: X_n, X on different spaces)
   2. *vg.* $\Longrightarrow$ *a.s.* ($\exists(\Omega, \mathcal{F}, \mathsf{P})$, $X_n, X \ni X_n \to X$ *a.e.*: $\Omega = \mathbb{R}$, $\mathsf{P} = \lambda$, $X_n(\omega) = {F_n}^{-1}(\omega)$).

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:

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

Central Limit Theorem:

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

Law of the Iterated Logarithm:

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

Together these three give a clear picture of how quickly and in what sense $\frac{1}{n}S_n$ tends to μ. We begin with the Law of Large Numbers (LLN), in its "weak" form (asserting convergence *in pr.*) 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).

IV. 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 $\Longrightarrow Y_n \to 0$, *in pr.*
    2. (pairwise) IID L^2 is OK

V. 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: $\mathsf{P}[|S_{n^2}| > n^2\epsilon \text{ } i.o.] = 0$, $\therefore \frac{1}{n^2}S_{n^2} \to 0$ *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: $\mathsf{P}[D_n > n^2\epsilon] < \frac{4n^2M}{n^4\epsilon^2}$, $\therefore D_n \to 0$ *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.

VI. 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. $Y_n = X_n 1_{[|X_n| \le n]}$

VII. 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 $S_n/n \not\to 0$ *in pr.* 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 $S_n/n \to 0$ *in pr.* and not *a.s.*
  - D. Medians: for ANY RV's $X_n \to X_\infty$ *in pr.*, then $m_n \to m_\infty$ if m_∞ 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, in probability, or even in distribution.

A LAW OF LARGE NUMBERS FOR CORRELATED SEQUENCES

In many applications we would like a Law of Large Numbers for sequences of random variables that are *not* independent; for example, in Markov Chain Monte Carlo integration, we have a stationary Markov chain $\{X_t\}$ (this means that the distribution of X_t is the same for all t and that the conditional distribution of X_u for $u > t$, given $\{X_s | s \le t\}$, depends only on X_t) and want to estimate the population mean $\mathsf{E}[\phi(X_t)]$ for some function $\phi(\cdot)$ by the sample mean

$$\mathsf{E}[\phi(X_t)] \approx \frac{1}{T}\sum_{t=1}^{T} \phi(X_t).$$

Even though they are identically distributed, the random variables $Y_t \equiv \phi(X_t)$ won't be independent if the X_t aren't independent, so the LLN we already have won't quite work.

A sequence of random variables Y_t is called *stationary* if each Y_t has the same probability distribution and, moreover, each finite set $(Y_{t_1+h}, Y_{t_2+h}, ..., Y_{t_k+h})$ has a joint distribution that doesn't depend on h. The sequence is called "L^2" if each Y_t has a finite variance σ^2 (and hence also a well-defined mean μ); by stationarity it also follows that the *covariance*

$$\gamma_{st} = \mathsf{E}[(Y_s - \mu)(Y_t - \mu)]$$

is finite and depends only on the absolute difference $|t - s|$.

Theorem. If a stationary L^2 sequence has a summable covariance, *i.e.*, satisfies $\sum_{t=-\infty}^{\infty} |\gamma_{st}| \le c < \infty$, then

$$\mathsf{E}[Y_t] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} Y_t.$$

Proof. Let S_T be the sum of the first T Y_t's and set (as usual) $\bar{Y}_T \equiv S_T/T$. The variance of S_T is

$$\begin{aligned}
\mathsf{E}[(S_T - T\mu)^2] &= \sum_{s=1}^{T}\sum_{t=1}^{T} \mathsf{E}[(X_s - \mu)(X_t - \mu)] \\
&\le \sum_{s=1}^{T}\sum_{t=-\infty}^{\infty} |\gamma_{st}| \\
&\le T\,c,
\end{aligned}$$

so $\bar{Y}_T$ had variance $\mathsf{V}[\bar{Y}_T] \le c/T$ and by Chebychev's inequality

$$\begin{aligned}
\mathsf{P}[|\bar{Y}_T - \mu| > \epsilon] &\le \frac{\mathsf{E}[(\bar{Y}_T - \mu)^2]}{\epsilon^2} \\
&= \frac{\mathsf{E}[(S_T - T\mu)^2]}{T^2\epsilon^2} \\
&\le \frac{T\,c}{T^2\epsilon^2} \\
&= \frac{c}{T\epsilon^2} \to 0
\end{aligned}$$

A strong LLN follows with a bit more work, just as for *iid* random variables.

Examples

1. **IID:** If X_t are independent and identically distributed, and if $Y_t = \phi(X_t)$ has finite variance σ^2, then Y_t has a well-defined finite mean μ and $\bar{Y}_T \to \mu$. Here $\gamma_{st} = \begin{cases} \sigma^2 & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases}$, so $c = \sigma^2 < \infty$.
2. **AR$_1$:** If Z_t are *iid* $N(0,1)$ for $-\infty < t < \infty$, $\mu \in \mathbb{R}$, $\sigma > 0$, $-1 < \rho < 1$, and

$$\begin{aligned} X_t &\equiv \mu + \sigma \sum_{s=0}^{\infty} \rho^s Z_{t-s} \\ &= \rho X_{t-1} + \alpha + \sigma Z_t, \end{aligned} \tag{$*$}$$

where $\alpha = (1-\rho)\mu$, then the X_t are identically distributed (all with the $N(\mu, \frac{\sigma^2}{1-\rho^2})$ distribution) but not independent ($\gamma_{st} = \frac{\sigma^2}{1-\rho^2}\rho^{|s-t|} \neq 0$); this is called an "autoregressive process" (because of equation (*), expressing X_t as a regression of previous X_s's) of order one (because only one earlier X_s appears in (*)), and is about the simplest non-*iid* sequence occuring in applications. Since the covariance is summable,

$$\sum_{t=-\infty}^{\infty} |\gamma_{st}| = \frac{\sigma^2}{1-\rho^2}\frac{1+|\rho|}{1-|\rho|} = \frac{\sigma^2}{(1-|\rho|)^2} < \infty,$$

we have $\bar{X}_T \to \mu$ as before.
2. **Geometric Ergodicity:** If for some $0 < \rho < 1$ and $c > 0$ we have $\gamma_{st} < c\rho^{|s-t|}$ for a Markov chain Y_t the chain is called *Geometrically Ergodic* (because $c\rho^t$ is a geometric sequence), and the same argument as for AR$_1$ shows that $\bar{Y}_t$ converges; Meyn & Tweedie (1993), Tierney (1994), and others have given conditions for MCMC chains to be Geometric Ergodic, and hence for the almost-sure convergence of sample averages to population means.
3. **General Ergodicity:** Consider the sequences of random variables:
   1. $X_{n+1} \equiv \lceil 2^n \omega \rceil \pmod 1$, $0 < \omega \le 1$;
   2. $X_{n+1} \equiv X_n + \alpha \pmod 1$ (Does it matter if α is rational?);
   3. $X_{n+1} \equiv 4X_n(1 - X_n)$.

   For each, find a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ and a distribution of X_0 such that the X_n are all identically distributed; the sequence is *ergodic* if each $E \in \mathcal{F}$ left invariant by the transformation T that takes X_n to X_{n+1}, $E = T^{-1}(E)$, is trivial in the sense that $\mathsf{P}[E] = 0$ or $\mathsf{P}[E] = 1$. The *Ergodic Theorem* asserts that $\bar{X}_n$ converges to a T-invariant limit X_∞ as $n \to \infty$; since only constants are T-invariant for ergodic sequences, it follows that $\bar{X}_n \to \mu$. The conditions here are weaker than those for the usual CLT; in cases 2. and 3. above, for example, each X_n is completely determined by X_1 so there is no independence at all!

Stable Limit Laws

Let $S_n = X_1 + ... + X_n$ be the partial sum of iid random variables. IF the random variables are all square integrable, then the Central Limit Theorem applies and necessarily $\frac{S_n}{n\sigma^2} - \mu \Longrightarrow \mathsf{No}(0,1)$. But what if each X_n is *not* square integrable? Denote by $F(x) = \mathsf{P}[X_n \le x]$ the common CDF of the $\{X_n\}$.

Theorem (Stable Limit Law).

There exist constants $A_n > 0$ and $B_n \in \mathbb{R}$ and a distribution μ for which the

$$\frac{S_n}{A_n} - B_n \Longrightarrow \mu$$

if and only if there are constants $0 < \alpha \le 2$, $M^- \ge 0$, and $M^+ \ge 0$, with $M^- + M^+ > 0$, such that as $x \to \infty$ the following limits hold for every $\xi > 0$:

1. $\dfrac{F(-x)}{1 - F(x)} \to \dfrac{M^-}{M^+}$;
2. $M^+ > 0 \Rightarrow \dfrac{1 - F(x\xi)}{1 - F(x)} \to \xi^{-\alpha}$ $\qquad M^- > 0 \Rightarrow \dfrac{F(-x\xi)}{F(-x)} \to \xi^{-\alpha}$.

In this case the limit is the **Stable Distribution** with index α, with characteristic function

$$\mathsf{E}[e^{itY}] = e^{-|\gamma t|^\alpha [1 - i\beta \tan(\frac{\pi\alpha}{2}) \operatorname{sgn}(t)]},$$

where $\beta = \frac{M^+}{M^- + M^+}$ and $\gamma = (M^- + M^+)$. Condition 2. says that each tail must be fall off like a power (sometimes called *Pareto tails*), and the powers must be identical; Condition 1. gives the tail ratio. In a common special case, $M^- = 0$; for example, random variables X_n with the Pareto distribution (often used to model income) given by $P[X_n > t] = (k/t)^\alpha$ for $t \ge k$ will have a stable limit for their partial sums if $\alpha < 2$, and (by CLT) a normal limit if $\alpha \ge 2$. You can find out more details reading Chapter 9 of Breiman's book.