1 Review of non-random convergence concepts

By convergence, we mean the limiting behavior of a sequence of numbers (A sequence is just a list of numbers, ordered according to the integers.)

Exs: (i) $X_1 = 10,\ X_2 = 5.6,\ x_3 = 7, \ldots$ (no pattern); (ii) $X_n = \log(n),\ n = 1, 2, 3, \ldots$; (iii) $X_n = 2$; (iv) $X_n = \frac{1}{n}$; etc.

We say that a sequence converges if the limit $\lim_{n\to\infty} X_n$ exists and is finite (if the limit is $+/-\infty$, the we say the sequence diverges).

Recall: the statement $\lim_{n\to\infty} X_n = a$ means that for all $\epsilon > 0$ there exists some integer $n^*(\epsilon)$ such that, for all $n > n^*(\epsilon)$, $|X_n - a| < \epsilon$.

As n gets large, the "tail" of the sequence X_n lies in an ϵ-neighborhood of a, for all $\epsilon > 0$.

■■■

When the limit doesn't exist, then we use:

- $\text{limsup}_{n\to\infty} X_n \equiv \lim_{n\to\infty}(\sup_{m\geq n} X_m)$: least upper bound. (The sequence $\sup_{m\geq n} X_m$ is the maximum attained by the "tail" of the X sequence, excluding the first $n-1$ elements. The limsup is the usual limit of this sequence.)
- $\text{liminf}_{n\to\infty} X_n \equiv \lim_{n\to\infty}(\inf_{m\geq n} X_m)$: greatest lower bound
- The limsup and liminf always exists. If the limit exists, then lim=limsup=liminf.
- Consider: $X_n = (-1)^n$. The limit doesn't exist, but limsup is 1 and liminf is -1.
- Consider $X_n = (-1)^n \cdot 1/n$:

n	X_n	$\sup_{m\geq n} X_m$	$\inf_{m\geq n} X_m$
1	-1	$\frac{1}{2}$	-1
2	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{3}$
3	$-\frac{1}{3}$	$\frac{1}{4}$	$-\frac{1}{3}$
4	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{5}$
5	$-\frac{1}{5}$	$\frac{1}{6}$	$-\frac{1}{5}$
6	$\frac{1}{6}$	$\frac{1}{6}$	$-\frac{1}{7}$
$\vdots$			
∞	0	0	0

Note: for scalar sequences X_n, the definition of limsup implies that, for all ϵ, and all n, there exists an $n^*(\epsilon, n) > n$ (indeed, an infinite number of them) such that $X_{n^*(\epsilon,n)}$ lies within an ϵ-neighborhood of the limsup; that is, this occurs *infinitely often.* X_n also attains any small neighborhood of its liminf infinitely often.

■■■

2 Big-O and little-o notation

(Taken from White, *Asymptotic Theory for Econometricians*, chap. II)

- The sequence $\{b_n\}$ is *at most of order* n^λ, denoted $O(n^\lambda)$, if and only if for *some* real number Δ, $0 < \Delta < \infty$, there exists a finite integer N such that for all $n \geq N$, $|n^{-\lambda} b_n| < \Delta$.
- The sequence $\{b_n\}$ is of *order smaller than* n^λ, denoted $o(n^\lambda)$, if and only if for *every* real number $\delta > 0$, there exists a finite integer $N(\delta)$ such that for all $n \geq N(\delta)$, $|n^{-\lambda} b_n| < \delta$.

Remarks:

- $O(n^\lambda)$: $n^{-\lambda} b_n$ is eventually bounded. $O(1)$: b_n eventually bounded.
- $o(n^\lambda)$: $\lim_{n\to\infty} n^{-\lambda} b_n = 0$. $o(1)$: $b_n \to 0$.
- $b_n = o(n^\lambda) \longrightarrow b_n = O(n^\lambda)$
- $b_n = O(n^\lambda) \longrightarrow b_n = o(n^{\lambda+\delta})$ for $\delta > 0$

Let a_n and b_n be scalars. Then:

- If $a_n = O(n^\lambda)$ and $b_n = O(n^\mu)$, then $(a_n b_n) = O(n^{\lambda+\mu})$ and $(a_n + b_n) = O(n^{\max(\lambda,\mu)})$.
- If $a_n = o(n^\lambda)$ and $b_n = o(n^\mu)$, then $(a_n b_n) = o(n^{\lambda+\mu})$ and $(a_n + b_n) = o(n^{\max(\lambda,\mu)})$.
- If $a_n = O(n^\lambda)$ and $b_n = o(n^\mu)$, then $(a_n b_n) = o(n^{\lambda+\mu})$ and $(a_n + b_n) = O(n^{\max(\lambda,\mu)})$.

3 Convergence concepts for random sequences

For random sequences, how do we define convergence? By way of an example, we illustrate two important convergence concepts: **convergence in probability** and **convergence almost surely**.

- Consider this class as a population. What is μ, the population mean of hours slept last night?
- As a thought experiment, consider estimating μ by randomly sampling your classmates, and using the sample average.

n	value X_n	$\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

  So now each students has his/her own sequence of $\bar{X}_n$, which depends on the order in which she sampled her classmates.

- Intuitively, the sequence $\bar{X}_n$ "converges" to μ. But in what precise sense?
- First, *every* sequence compiled by a student in the class should converge (as $n \to 25$) to μ. This is what is meant by **almost sure convergence**: all (possible) realizations of the random sequence should converge.

  Formally: let $Z_1, Z_2, Z_3, \ldots, Z_n, \ldots$ and Z^* be defined on the same probability space $(\Omega, \mathbb{B}, P)$. Then for each $\omega \in \Omega$, let $Z_1(\omega), Z_2(\omega), \ldots, Z_n(\omega)$ and $Z^*(\omega)$ denote the corresponding realizations for the random variables. That is, for a fixed ω, the collection $\{Z_1(\omega), Z_2(\omega), \ldots, Z_n(\omega)\}$ is just a *non-random* sequence. Almost-sure convergence is the statement that these sequences converge (in the usual nonstochastic sense) for almost-all ω:

$$P(\omega : \lim_{n\to} Z_n(\omega) = Z^*(\omega)) = 1.$$

When $Z^n = \bar{X}_n$, and $Z^* = \mu$, then the result is the *strong law of large numbers*:

$$\bar{X}_n \xrightarrow{a.s.} \mu.$$

These points will be elaborated in the next set of lecture notes.

- On the other hand, let's consider the sequences compiled by everyone in the class. At each n, consider the *fraction* of sequences which are close to μ, in the sense that $\bar{X}_n$ lies within some neighborhood of μ.

n	$\#\left(\lvert\bar{X}_n^i - \mu\rvert > 1\right)$	$\#\left(\lvert\bar{X}_n^i - \mu\rvert > 0.5\right)$	$\#\left(\lvert\bar{X}_n^i - \mu\rvert > 0.25\right)$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

You expect that as $n \uparrow$, the faction of sequences for which $\bar{X}_n$ lies outside some neighborhood of μ should go down. This is formalized in the concept of **convergence in probability**.

Formally: as before, let $Z_1, Z_2, Z_3, \ldots, Z_n, \ldots$ and Z^* be defined on the same probability space $(\Omega, \mathbb{B}, P)$. Then for each $\omega \in \Omega$, let $Z_1(\omega), Z_2(\omega), \ldots, Z_n(\omega)$ and $Z^*(\omega)$ denote the corresponding values for the random variables. Convergence in probability is the statement that:

$$\forall \epsilon > 0 : \quad \lim_{n\to\infty} P(\omega : |Z_n(\omega) - Z^*(\omega)| < \epsilon) = 1.$$

Expanding out the definition of limit, we have:

$$\forall \epsilon, \delta > 0,\ \exists n^*(\delta;\epsilon) : \quad \forall n > n^*(\delta;\epsilon),\ P\left(\omega : |Z_n(\omega) - Z^*(\omega)| < \epsilon\right) > 1 - \delta.$$

For all $\epsilon > 0$, the sequence $P_n^\epsilon \equiv \text{Prob}\left(|Z_n - Z^*| < \epsilon\right)$ is a *non-random sequence*. The idea of convergence in probability is that the sequence P_n^ϵ converges to 1 in the usual nonstochastic sense, for all $\epsilon > 0$.

When $Z^n = \bar{X}_n$, and $Z^* = \mu$, then the result is the *weak law of large numbers*:

$$\bar{X}_n \xrightarrow{p} \mu.$$

In the last set of lecture notes, we proved this using Chebyshev's inequality.

- Convergence in probability is weaker than (implied by) almost-sure convergence. Essentially (as we will show in examples), convergence in probability allows for aberrant ω's for which $Z_n(\omega)$ bounces in and out of a neighborhood of $Z^*(\omega)$ infinitely often, as long as the measure of this set is decreasing in n. Convergence a.s. prohibits this.

  Theorem: $Z_n \overset{a.s.}{\to} Z^* \Longrightarrow Z_n \overset{p}{\to} Z^*$.

3.1 Examples

- Prob space: $([0,1], \mathbb{B}_{[0,1]}, \mu)$
- Consider: for $\omega \in [0,1]$

$$Z_n(\omega) = \begin{cases} 1 & \text{if } n \text{ odd and } \omega \in [0, \frac{1}{n}] \\ 1 & \text{if } n \text{ even and } \omega \in [1-\frac{1}{n}, 1] \\ \omega & \text{otherwise} \end{cases}$$

$$Z^*(\omega) = \omega$$

- Example: if $\omega = \frac{1}{4}$, then $Z_1 = 1$, $Z_2 = \frac{1}{4}$, $Z_3 = 1$, $Z_4 = Z_5 = Z_6 = \cdots = \frac{1}{4}$.
- Does $Z_n \overset{p}{\to} Z^*$? YES. For given ϵ small enough, and given n, the probability that $|Z_n(\omega) - Z^*(\omega)| > \epsilon$ is $\frac{1}{n}$. Therefore, for given δ, choose n^* such that $\frac{1}{n^*} < \delta$.
- Does $Z_n \overset{as}{\to} Z^*$? Yes We have $Z_n(\omega) \to Z^*(\omega) = \omega$, except for the measure-zero set $\{0\}$.

  More formally, for all $\epsilon > 0$ and small, and all $\omega \in (0,1]$, we have that, for $n > \max\left(\frac{1}{\omega}, \frac{1}{1-\omega}\right) \equiv n^*$, $|Z_n(\omega) - Z^*(\omega)| < \epsilon$. (Actually it is stronger: we have $Z_n(\omega) = Z^*(\omega) = \omega$ for $n > n^*$.)

- What about

$$Z_n(\omega) = \begin{cases} 1 & \text{if } \omega \in [0, 0.00001 + \frac{1}{n}] \\ \omega & \text{otherwise} \end{cases}$$

  Does $Z_n \overset{p}{\to} Z^*$? NO. Here $P(\omega : |Z_n(\omega) - Z^*(\omega)| > \epsilon) \to 0.00001$, so that for any $\delta < 0.00001$, we cannot find a n^*.

- Here is an example which does not converge almost surely, but converges in probability.

  Again, consider the probability space $([0,1], \mathbb{B}_{[0,1]}, \mu)$. Define the random variables $Z_1, Z_2, \ldots$ as:

$$\begin{aligned} Z_1 &= \mathbb{1}_{[0,1]} \\ Z_2 &= \mathbb{1}_{[0,\frac{1}{2}]} \\ Z_3 &= \mathbb{1}_{[\frac{1}{2},1]} \\ Z_4 &= \mathbb{1}_{[0,\frac{1}{3}]} \\ Z_5 &= \mathbb{1}_{[\frac{1}{3},\frac{2}{3}]} \\ Z_6 &= \mathbb{1}_{[\frac{2}{3},1]} \\ Z_7 &= \mathbb{1}_{[0,\frac{1}{4}]} \\ Z_8 &= \mathbb{1}_{[\frac{1}{4},\frac{2}{4}]} \\ Z_9 &= \mathbb{1}_{[\frac{2}{4},\frac{3}{4}]} \\ &[....] \\ Z^* &= 0 \end{aligned} \tag{1}$$

  We have $Z_n \stackrel{p}{\to} Z^*$: for every $\epsilon, \delta > 0$, we have $P(|Z_n - Z^*| < \epsilon) > 1 - \delta$ for $n > \frac{1}{\delta}$.

  We *do not* have $Z_n \stackrel{as}{\to} Z^*$: for all $\omega \in [0,1]$, limsup $Z_n(\omega) = 1$ and liminf $Z_n(\omega) = 0$.

Previous examples are somewhat "exotic"; for real-valued random variables it is not easy to find examples satisfying convergence in probability but not almost surely. Indeed:

Theorem: If $X_n \stackrel{p}{\to} X$, then there exists a sequence n_k of integers increasing to infinity such that $X_{n^k} \stackrel{as}{\to} X$. Briefly stated: convergence in probability implies a.s. convergence along a subsequence.

Proof: see Chung, A Course in Probability Theory, pp. 73-74.

For the example above, consider the subsequence $Z_1, Z_2, Z_4, Z_7, ...$ That is, for $k = 1, 2, 3, 4, \ldots$, set $n^k = n^{k-1} + (k-1)$ with initial value $n^1 = 1$. (We "pick out" only the random variables which are equal to one for ω in the left end of the unit interval.)