Asymptotics: Stochastic convergence

Probability calculus / Adv Stat I

Prof. Dr. Matei Demetrescu

Why bother?

We often encounter **random quantities** of the form

$$Y_n = g(X_1, ..., X_n), \quad \text{where} \quad n = 1, 2, 3, ...$$

A simple – but ubiquitous – case is the average of n random variables

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

The objective of asymptotic theory is to establish results relating to the stochastic behavior of such sequences Y_n when $n \to \infty$.

- Y_n may converge to a constant in various ways,
- or the distribution of Y_n may converge to some 'limit distribution'.

We do this to obtain an approximation of the behavior of Y_n!

Today's outline

Asymptotics: Stochastic convergence

1. Convergence of number and function sequences
2. Shorthand notation: Landau symbols
3. Convergence of sequences of random variables
4. Up next

Outline

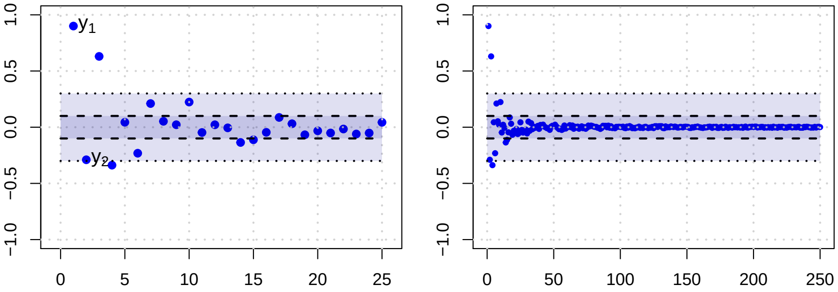
Real sequences and limits

Definition (Convergence of real number sequences)

A sequence of real numbers $\{y_n\}$ converges to $y \in \mathbb{R}^1$ iff for every real $\epsilon > 0$ there exists an integer $N(\epsilon)$ such that

$$|y_n - y| < \epsilon \quad \forall n \geq N(\epsilon).$$

The existence of the limit is denoted by $y_n \to y$ or $\lim_{n\to\infty} y_n = y$.



Relation to boundedness

For the limit of a sequence of numbers to exist, it is necessary (but not sufficient) that the sequence be bounded.

Example

The sequence $\quad y_n = 3 + n^{-2}, \quad n \in \mathbb{N}$

- is bounded, since $|y_n| \leq 4 \quad \forall n \in \mathbb{N}$,
- and has a limit $y_n \to 3$.

Example

The sequence $\quad y_n = \sin n, \quad n \in \mathbb{N}$

- is bounded, since $|\sin x| \leq 1 \quad \forall x$,
- but does not have a limit, since $\sin x$ cycles between $+1$ and -1.

Extensions

Convergence of sequences of vectors and matrices is defined elementwise.

Definition (Convergence of function sequences)

Let $\{f_n(x)\}$, $n \in \mathbb{N}$, be a sequence of functions having a common domain $D \subset \mathbb{R}^m$. The function sequence $\{f_n(x)\}$ converges to a function $f(x)$ with domain $D_0 \subset D$ iff for $n \to \infty$

$$f_n(x) \to f(x) \quad \forall x \in D_0.$$

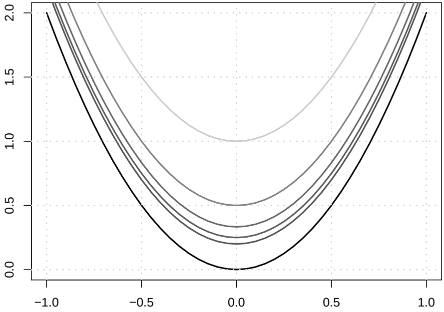
f is called the limiting function of $\{f_n\}$.

The definition implies that the values of the functions $f_n(x)$, $n = 1, 2, 3, ..$ converge to $f(x)$ pointwise for each single $x \in D_0$.

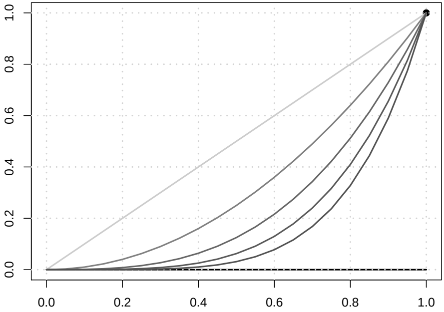
Hence, $f(x)$ can be viewed as an approximation of $f_n(x)$ when n is large for a given x.

Two examples

$$f_n(x) = n^{-1} + 2x^2,\ x \in [-1, 1]$$



$$f_n(x) = x^n,\ x \in [0, 1]$$



In fact, the latter is an example of non-uniform convergence.

Uniform convergence

Uniformity of the convergence requires $\sup_x |f_n(x) - f(x)| \to 0$ as $n \to \infty$ and is (much) stricter.

- In fact, $\sup_x |f_n(x) - f(x)|$ may be seen as a distance between two functions.
- Distances (between elements of a vector space) may be characterized formally:
  - Should be zero iff the two elements are the same
  - Should obey the triangle inequality.
  - Should be symmetric and nonnegative.
- The $\sup$ distance does fulfil these, but there are other metrics, e.g. the so-called L_2 distance, $\sqrt{\int (f_n(x) - f(x))^2 \mathrm{d}x}$.

Outline

Big-Oh, Small-oh (Landau) notations

Definition (Order of magnitude of a sequence)

Let $\{y_n\}$ be a real number sequence.

- $\{y_n\}$ is said to be **at most of order** n^k, denoted by $y_n = O(n^k)$, if there exists a finite constant c such that

$$\left| \frac{y_n}{n^k} \right| \leq c \quad \forall n \in \mathbb{N}.$$

- $\{y_n\}$ is said to be **of order smaller than** n^k, denoted by $y_n = o(n^k)$, if

$$\frac{y_n}{n^k} \to 0.$$

More generally, one may replace n^k by some sequence $\tilde{y}_n$.

Revisiting convergence and boundedness

- Write $O(n^0)$ and $o(n^0)$ as $O(1)$ and $o(1)$.
- They have a special interpretation...

Example

The sequence $$y_n = 3 + n^{-2}, \quad n \in \mathbb{N}$$

- is $O(1)$, since $|y_n| \leq 4 \quad \forall n \in \mathbb{N}$,
- and has a limit, $y_n - 3 = o(1)$.

Example

The sequence $$y_n = \sin n, \quad n \in \mathbb{N}$$

- is just $O(1)$, since $|\sin x| \leq 1 \quad \forall x$.

Typical relations

Note e.g. that

if $\{y_n\}$ is $O(n^k)$, then $\{y_n\}$ is $o(n^{k+\epsilon})\ \forall\epsilon > 0$;
if $\{y_n\}$ is $o(n^k)$, then $\{y_n\}$ is $O(n^k)$.

Example

Let $\{y_n\}$ be defined by $y_n = 3n^3 - n^2 + 2$, $n \in \mathbb{N}$. Since

$$\frac{y_n}{n^3} = 3 - \frac{1}{n} + \frac{2}{n^3} \quad \to \quad 3 < \infty, \qquad \text{we have} \qquad y_n = O(n^3);$$

Since for a positive ϵ

$$\frac{y_n}{n^{3+\epsilon}} = \frac{3}{n^\epsilon} - \frac{1}{n^{1+\epsilon}} + \frac{2}{n^{3+\epsilon}} \quad \to \quad 0, \qquad \text{we have} \qquad y_n = o(n^{3+\epsilon}).$$

But note that $y_n = O(n^\alpha)$ does not imply that $\frac{1}{y_n} = O(n^{-\alpha})$!

And some rules

If $x_n = O(\tilde{x}_n)$ and $y_n = O(\tilde{y}_n)$, then

$$\begin{aligned} x_n + y_n &= O(\max\{\tilde{x}_n; \tilde{y}_n\}), \\ x_n - y_n &= O(\max\{\tilde{x}_n; \tilde{y}_n\}), \\ x_n y_n &= O(\tilde{x}_n \tilde{y}_n). \end{aligned}$$

If $x_n = O(\tilde{x}_n)$ and $y_n = o(\tilde{y}_n)$, then

$$\begin{aligned} x_n + y_n &= O(\max\{\tilde{x}_n; \tilde{y}_n\}), \\ x_n - y_n &= O(\max\{\tilde{x}_n; \tilde{y}_n\}), \\ x_n y_n &= o(\tilde{x}_n \tilde{y}_n). \end{aligned}$$

Outline

Convergence depends on distance

In this section we extend the converge concepts for real number sequences to sequences of random variables.

For sequences of random variables, we distinguish among the following types/modes of convergence:

1.) *convergence in distribution;*
2.) *convergence in probability;*
3.) *convergence in mean square;*
4.) *almost-sure convergence.*

They express different ways in which the sequence may be close to its limit.

Convergence in Distribution

Definition (Convergence in Distribution)

Let $\{Y_n\}$ be a sequence of random variables with an associated sequence of cdfs $\{F_n\}$. If there exists a cdf F such that as $n \to \infty$

$$F_n(y) \to F(y) \quad \forall y \quad \text{at which } F \text{ is continuous,}$$

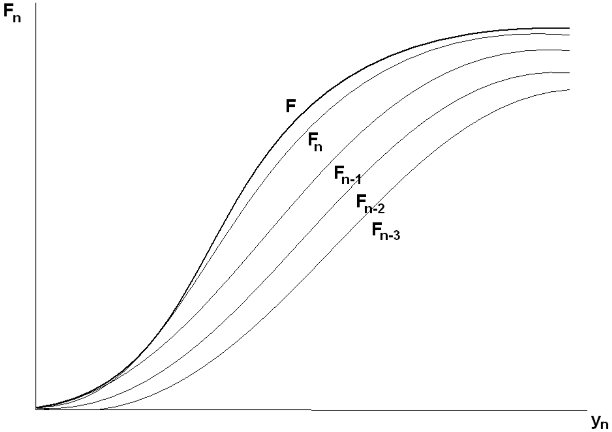
then Y_n converges in distribution to the random variable Y with cdf F.

We denote this by $Y_n \xrightarrow{d} Y$ or $Y_n \xrightarrow{d} F$. The function F is called the limiting cdf/limiting distribution of $\{Y_n\}$.

- The limiting cdf F can be the cdf of a *degenerate random variable* with $Y = c$, where c is a constant.
- In this case, we say that Y_n **converges in distribution to a constant**, and we denote this by $Y_n \xrightarrow{d} c$.

Convergence of distribution functions

- If the limit is continuous, convergence is also uniform.
- If $Y_n \xrightarrow{d} Y$, then as n becomes large, the actual cdf of Y_n can be approximated by the cdf F of the random variable Y.



... more concretely

Example

Let $\{Y_n\}$ be a sequence of random variables with an associated sequence of cdfs $\{F_n\}$ given by

$$F_n(y) = \begin{cases} 0 & \text{for} \quad y < 0 \\ (\frac{y}{\theta})^n & \text{for} \quad 0 \leq y < \theta \\ 1 & \text{for} \quad y \geq \theta \end{cases} .$$

We see that as $n \to \infty$,

$$F_n(y) \to F(x) = \begin{cases} 0 & \text{for} \quad y < \theta \\ 1 & \text{for} \quad y \geq \theta \end{cases} ,$$

which is the cdf a *degenerate random variable*, and we have $Y_n \xrightarrow{d} \theta$.

Establishing convergence in distribution

The following theorem is based upon the uniqueness of MGFs, and is very useful for identifying limiting distributions.

Theorem (5.1)

Let $\{Y_n\}$ be a sequence of random variables having an associated sequence of MGFs $\{M_{Y_n}(t)\}$. Let $M_Y(t)$ be the MGF of Y. Then

$$Y_n \xrightarrow{d} Y \qquad \textit{iff} \qquad M_{Y_n}(t) \to M_Y(t) \quad \forall t \in (-h, h), \; \textit{for some } h > 0.$$

What about densities?

- Convergence of cdfs does not imply convergence of pdfs
- ... but the converse holds; see Mittelhammer (1996, Theorem 5.1).

The χ^2 case

Example

Let $X_n \sim \chi^2_{(n)}$ with MGF $M_{X_n}(t) = (1-2t)^{-\frac{n}{2}}$, $t < 1/2$ and

$$Z_n \;=\; \frac{X_n - n}{\sqrt{2n}} \;=\; -\sqrt{\frac{n}{2}} + \frac{1}{\sqrt{2n}} X_n.$$

Therefore

$$M_{Z_n}(t) \;=\; e^{-\sqrt{\frac{n}{2}}t} \cdot M_{X_n}\left(\frac{1}{\sqrt{2n}} \cdot t\right) \;=\; e^{-\sqrt{\frac{n}{2}}t - \frac{n}{2}\ln\left(1-\sqrt{\frac{2}{n}}t\right)}.$$

But $\ln(1-x) = -x + x^2/2 + o(x^2)$ so

$$M_{Z_n}(t) \;=\; e^{\frac{t^2}{2}+o(1)} \;=\; e^{\frac{t^2}{2}} + o(1)$$

which is the same as $M_{Z_n}(t) \to e^{\frac{t^2}{2}}$, the MGF of the standard normal.

Asymptotic Distributions

The asymptotic distribution for a random variable Z_n is any distribution that provides an approximation to the true distribution of Z_n for large n.

If $\{Z_n\}$ has a limiting distribution,

- this may be considered as an asymptotic distribution, since
- it provides an approximation to the distribution of Z_n for large n.

The following definition of the asymptotic distribution

- generalizes the concept of approximating distributions for large n and
- includes cases where Z_n has no limiting distribution or a degenerate limiting distribution.

A bit of a misnomer

Definition (Asymptotic Distribution)

Let $\{Z_n\}$ be a sequence of random variables defined by

$$Z_n = h(X_n, a_n), \quad \text{where} \quad \begin{array}{l} X_n \xrightarrow{d} X \text{ (nondegenerate)}, \\ a_n : \text{sequence of numbers/parameters.} \end{array}$$

An *asymptotic distribution for* Z_n is the distribution of $h(X, a_n)$,

$$Z_n \overset{a}{\sim} h\left(X, a_n\right) \qquad \text{“}Z_n \text{ is asymptotically distributed as } h\left(X, a_n\right)\text{”}.$$

Example

We know e.g. that $W_n = \frac{X_n - n}{\sqrt{2n}} \xrightarrow{d} W \sim \mathcal{N}(0,1)$ if $X_n \sim \chi^2_{(n)}$.

Consider now $Y_n = h(W_n, n) = \sqrt{2n} \cdot W_n + n \ (= X_n)$.

Then $Y_n = h(W_n, n) \overset{a}{\sim} h(W, n) = \sqrt{2n} \cdot W + n \sim \mathcal{N}(n, 2n)$.

The continuous mapping theorem

Theorem (5.2 (CMT))

Let $X_n \xrightarrow{d} X$, and let $Y_n = g(X_n)$ with g a continuous function which depends on n only via X_n . Then $g(X_n) \xrightarrow{d} g(X)$.

This helps working with limiting/asymptotic distributions

Example

Consider e.g. $Z_n \xrightarrow{d} Z \sim \mathcal{N}(0,1)$. Then

- $g(Z_n) = 2Z_n + 5 \quad \xrightarrow{d} \quad 2Z + 5 \sim \mathcal{N}(5,4)$;
- $g(Z_n) = Z_n^2 \quad \xrightarrow{d} \quad Z^2 \sim \chi^2_{(1)}$.

Convergence in Probability

Focus on more than just distributions converging...

Definition (Convergence in probability)

The sequence of random variables $\{Y_n\}$ *converges in probability* to the random variable Y iff

$$\lim_{n\to\infty} \mathrm{P}\left(|Y_n - Y| < \epsilon\right) = 1 \quad \forall\, \epsilon > 0.$$

We denote this by $Y_n \xrightarrow{p} Y$, or $\operatorname{plim} Y_n = Y$, where Y is called the probability limit of Y_n.

The definition implies that if n is large enough, observing outcomes of Y_n is essentially equivalent to observing outcomes of Y.

Also note that the probability limit Y can be a *degenerate random variable* with $Y = c$, where c is a constant. We denote this by $Y_n \xrightarrow{p} c$.

Playing with normals

Example

Consider the random variable Y_n with pdf

$$f_n(y) = \tfrac{1}{n}\mathbb{I}_{\{0\}}(y) + \left(1 - \tfrac{1}{n}\right)\mathbb{I}_{\{1\}}(y) \quad \to_{n\to\infty} \quad \mathbb{I}_{\{1\}}(y).$$

Hence we have $\mathrm{P}\left(|Y_n - 1| = 0\right) \to 1$ as $n \to \infty$, so that

$$\lim_{n\to\infty} \mathrm{P}\left(|Y_n - 1| < \epsilon\right) = 1 \quad \forall\, \epsilon > 0, \qquad \text{and} \qquad \operatorname{plim} Y_n = 1.$$

Example

Let $Y \sim \mathcal{N}(0,1)$ and $Z_n \sim \mathcal{N}(0, \frac{1}{n})$, independent. Let $Y_n = Z_n + Y$ such that $Y_n \sim \mathcal{N}\left(0\ ,\ [1 + \frac{1}{n}]\right)$. Then, $\operatorname{plim} Y_n = Y$ since

$$\lim_{n\to\infty} \mathrm{P}\left(|Y_n - Y| < \epsilon\right) = \underbrace{\lim_{n\to\infty} \mathrm{P}\left(|Z_n| < \epsilon\right) \geq \lim_{n\to\infty}\left(1 - \frac{\operatorname{Var}(Z_n)}{\epsilon^2}\right)}_{\text{Chebyshev's Ineq.}} = 1.$$

Probabilistic Landau symbols

Definition

A (stochastic) sequence Y_n is said to be $o_p\left(n^k\right)$ if $\frac{Y_n}{n^k} \xrightarrow{p} 0$.

Things are slightly different for the big-Oh case, though:

Definition

1. A stochastic sequence Y_n is said to be **uniformly bounded in probability** if $\forall \epsilon > 0\ \exists C\left(\epsilon\right)$ such that $\sup_n \mathrm{P}\left(|Y_n| > C(\epsilon)\right) < \epsilon$.
2. A stochastic sequence Y_n is said to be $O_p\left(n^k\right)$ if $\frac{Y_n}{n^k}$ is uniformly bounded in probability.

The same rules as for deterministic Landau symbols apply for finite sums, products, etc.

The continuous mapping theorem (again?)

Theorem (5.3)

Let $X_n \xrightarrow{p} X$*, and let* $Y_n = g(X_n)$ *with* g *a continuous function which depends on* n *only via* X_n*. Then,*

$$\operatorname{plim} Y_n = \operatorname{plim} g(X_n) = g(\operatorname{plim} X_n) = g(X).$$

The theorem implies that the plim operator acts analogously to the standard $\lim$ operator of real analysis. You may in fact write

$$g(X + o_p(1)) = g(X) + o_p(1).$$

Example

Let $X_n \xrightarrow{p} 3$. Then the probability limit of $Y_n = \ln(X_n) + \sqrt{X_n}$ is

$$\operatorname{plim} Y_n = \ln(\operatorname{plim} X_n) + \sqrt{\operatorname{plim} X_n} = \ln(3) + \sqrt{3}.$$

Special cases

Theorem (5.4)

For the sequences of random variables X_n, Y_n, and the constant a.

a. $\operatorname{plim}(aX_n) = a(\operatorname{plim} X_n)$;
b. $\operatorname{plim}(X_n + Y_n) = \operatorname{plim} X_n + \operatorname{plim} Y_n$ *(the* plim *of a sum = the sum of the* plims*)*;
c. $\operatorname{plim}(X_n Y_n) = \operatorname{plim} X_n \operatorname{plim} Y_n$ *(the* plim *of a product = the product of the* plims*)*;
d. $\operatorname{plim}(X_n / Y_n) = (\operatorname{plim} X_n)/(\operatorname{plim} Y_n)$ *(if denominators are nonzero).*

The results of Theorem 5.4 extend to matrices by applying them to matrices element-by-element – see Mittelhammer (1996, p. 244-245).

Relation to convergence in distribution

Theorem (5.5)

$Y_n \xrightarrow{p} Y \Rightarrow Y_n \xrightarrow{d} Y.$

Example

Let e.g. $Y_n = (2 + \frac{1}{n})X + 3$, where $X \sim \mathcal{N}(1, 2)$. Then the plim of Y_n is

$$\text{plim}\, Y_n = \text{plim}\left(2 + \frac{1}{n}\right) \text{plim}\, X + \text{plim}\, 3 = 2X + 3 = Y \sim \mathcal{N}(5, 8).$$

Hence $Y_n \xrightarrow{d} Y \sim \mathcal{N}(5, 8)$.

More limits

Theorem (5.6)

$Y_n \xrightarrow{d} c \Rightarrow Y_n \xrightarrow{p} c.$

Theorem (5.7 (Slutsky's Theorem(s)))

Let $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} c$. *Then,*

a. $X_n + Y_n \xrightarrow{d} X + c$;
b. $X_n \cdot Y_n \xrightarrow{d} X \cdot c$;
c. $X_n / Y_n \xrightarrow{d} X / c.$

Convergence in mean square

Definition (Convergence in mean square)

The sequence of random variables $\{Y_n\}$ **converges in mean square** to the random variable Y, iff

$$\lim_{n \to \infty} \mathrm{E}\left((Y_n - Y)^2\right) = 0.$$

We denote this by $Y_n \stackrel{m}{\to} Y$.

First- and second order moments of Y_n and Y converge to one another; see below.

The basic tool

Theorem (5.8)

$Y_n \stackrel{m}{\to} Y$ *iff*

a. $\mathrm{E}(Y_n) \to \mathrm{E}(Y)$,
b. $\mathrm{Var}(Y_n) \to \mathrm{Var}(Y)$,
c. $\mathrm{Cov}(Y_n, Y) \to \mathrm{Var}(Y)$.

The necessary and sufficient conditions in Theorem 5.8 simplify, when Y is a constant, as stated in the following corollary.

Corollary (5.1)

$Y_n \stackrel{m}{\to} c$ *iff* $\mathrm{E}(Y_n) \to c$ *and* $\mathrm{Var}(Y_n) \to 0$.

Thank Markov/Chebychev for this shortcut

Theorem (5.9)

$Y_n \xrightarrow{m} Y \Rightarrow Y_n \xrightarrow{p} Y.$

- Often convergence in mean square is relatively easy to demonstrate.
- The sample average of $iid(0, \sigma^2)$ variables is the leading example.
- But the converse is not true!

Example

Let e.g. the pdf of Y_n be given by $f_n(y) = \begin{cases} 1 - \frac{1}{n^2} & \text{for} \quad y_n = 0 \\ \frac{1}{n^2} & \text{for} \quad y_n = n \end{cases}.$

It immediately follows that $\mathrm{P}(y_n = 0) \to 1$ so that $\operatorname{plim} Y_n = 0$. However,

$$\mathrm{E}\left((Y_n - 0)^2\right) \ = \ \mathrm{E}(Y_n^2) \ = \ 0^2 \cdot (1 - \tfrac{1}{n^2}) + n^2 \cdot \tfrac{1}{n^2} \ = \ 1 \quad \forall n,$$

so that $Y_n \overset{m}{\nrightarrow} 0$.

Outline

Coming up

Limiting theorems