Stochastic Models (Lecture #4)

Thomas Verdebout

Université libre de Bruxelles (ULB)

Today

Today, our goal will be

- to discuss limits of sequences of rv, and
- to study famous limiting results.

Convergence of sequences of rv

Let $X_1, X_2, \dots$ be i.i.d. rv, that is, rv that are independent and identically distributed. Assume X_1 is square-integrable, and write $\mu := \mathbb{E}[X_1]$ and $\sigma^2 := \mathrm{Var}[X_1]$.

Let $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Then

- $\mathbb{E}[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X_1] = \mu$, and
- $\mathrm{Var}[\overline{X}_n] = \frac{1}{n^2}\mathrm{Var}[\sum_{i=1}^n X_i] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[X_i] = \frac{1}{n}\mathrm{Var}[X_1] = \frac{\sigma^2}{n}$,which converges to 0 as $n \to \infty$.

Consequently, we feel intuitively that $\overline{X}_n \to X$, where $X = \mu$.

How to make this convergence precise?

Convergence of sequences of r.v.'s

Consider a sequence of r.v.'s (X_n) and a r.v. X, defined on $(\Omega, \mathcal{A}, \mathbb{P})$.

How to define $X_n \to X$ (as $n \to \infty$)?

$X_n \stackrel{a.s.}{\to} X$ (almost surely) $\Leftrightarrow \mathbb{P}[\limsup_n |X_n - X| = 0] = 1$.

$X_n \stackrel{P}{\to} X$ (in probability) $\Leftrightarrow \lim_n \mathbb{P}[|X_n - X| > \varepsilon] = 0, \forall \varepsilon > 0$.

$X_n \stackrel{L^r}{\to} X$ (in L^r, $r > 0$) $\Leftrightarrow \mathbb{E}[|X_n - X|^r] \to 0$.

$X_n \stackrel{\mathcal{D}}{\to} X$ in distribution (or in law) $\Leftrightarrow F^{X_n}(x) \to F^X(x)$ for all x at which F^X is continuous.

Convergence of sequences of r.v.'s

Consider as example $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}, \lambda)$.

$X_n \stackrel{a.s.}{\to} X$: pointwise convergence (modulo null set).

$X_n \stackrel{P}{\to} X$: convergence in measure.

$X_n \stackrel{L^r}{\to} X$: $\int_0^1 |X_n(t) - X(t)|^r dt \to 0$.

Convergence of sequences of r.v.'s

A principal question: What is the relation among those 4 types of convergence?

Lemma. (Markov inequality). If $E|Y|^r < \infty, r > 0$, then

$$\mathbb{P}[|Y| > \varepsilon] \leq \frac{E|Y|^r}{\varepsilon^r}.$$

With $Y = X - \mathbb{E}[X]$ and $r = 2$ this becomes Chebyshev's inequality:

$$\mathbb{P}[|X - \mathbb{E}[X]| > \varepsilon] \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Proof. Since $|Y|^r \geq \varepsilon^r I\{|Y| > \varepsilon\}$, it follows that $\mathbb{E}|Y|^r \geq \varepsilon^r \mathbb{P}[|Y| > \varepsilon]$.

An easy consequence is that

$$X_n \stackrel{L^r}{\to} X \text{ implies } X_n \stackrel{P}{\to} X.$$

The other direction is not true, we give a counter example.

Example 1:

Let $Y_1, Y_2, \ldots$ be i.i.d. rv, with common distribution

distribution of Y_i		
values	0	2
probabilities	$\frac{1}{2}$	$\frac{1}{2}$

Define $X_n = \prod_{i=1}^n Y_i$.

The distribution of X_n is

distribution of X_n		
values	0	2^n
probabilities	$1 - \frac{1}{2^n}$	$\frac{1}{2^n}$

We feel that $X_n \to X$, where $X = 0$.

Convergence of sequences of rv

In probability:

For all $\varepsilon > 0$,
$\mathbb{P}[|X_n - X| > \varepsilon] = \mathbb{P}[X_n > \varepsilon] \leq \mathbb{P}[X_n > 0] = \frac{1}{2^n} \to 0$,
as $n \to \infty$.

$\Rightarrow X_n \xrightarrow{P} X$, as $n \to \infty$.

In L^1:

$\mathbb{E}[|X_n - X|] = \mathbb{E}[X_n] = 0 \times \left(1 - \frac{1}{2^n}\right) + 2^n \times \frac{1}{2^n} = 1$,
which does not go to zero, as $n \to \infty$.

$\Rightarrow$ the convergence does not hold in the L^1 sense.

Convergence of sequences of rv

One can further show that $X_n \stackrel{a.s.}{\to} X$ implies $X_n \stackrel{P}{\to} X$.

Again the other direction is not true:

Example 2:

$(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}((0,1]), m)$ (where m is the Lebesgue measure). Further let

$$\begin{array}{ll} X_1(\omega) = \mathbb{I}_{(0,1/2]}(\omega) & X_2(\omega) = \mathbb{I}_{(1/2,1]} \\ X_3(\omega) = \mathbb{I}_{(0,1/4]} & X_4(\omega) = \mathbb{I}_{(1/4,2/4]} \cdots \\ X_7(\omega) = \mathbb{I}_{(0,1/8]} & X_8(\omega) = \mathbb{I}_{(1/8,2/8]} \cdots \end{array}$$

For $k \in \{2^n - 1, 2^n, \ldots, 2^{n+1} - 2\}$ we have

$$\mathbb{P}(|X_k - 0| > 0) = 2^{-n}.$$

BUT $\mathbb{P}(\limsup_k |X_k - 0| = 0) = 0.$

Convergence of sequences of rv

Thus we have for the moment

$$X_n \stackrel{a.s.}{\to} X \quad \begin{array}{c}\Longrightarrow\\ \not\Longleftarrow\end{array} \quad X_n \stackrel{P}{\to} X \quad \begin{array}{c}\Longleftarrow\\ \not\Longrightarrow\end{array} X_n \stackrel{L^r}{\to} X$$

Convergence of sequences of rv

Assume $X_n \stackrel{P}{\to} X$. We wish to prove that then $X_n \stackrel{\mathcal{D}}{\to} X$.

For the following argument we use (exercise)

$$\mathbb{P}(A) - \mathbb{P}(B) \leq \mathbb{P}(A \cap B^c).$$

Set $A = [X \leq x - \varepsilon]$ and $B = [|X_n - X| \geq \varepsilon]$. Then we have

$$A \cap B^c = [X \leq x - \varepsilon] \cap [|X_n - X| < \varepsilon] \subset [X_n \leq x].$$

Thus

$$\mathbb{P}[X \leq x - \varepsilon] - \mathbb{P}[|X_n - X| \geq \varepsilon] \leq \mathbb{P}[X_n \leq x].$$

A similar argument (exercise) shows that

$$\mathbb{P}[X_n \leq x] \leq \mathbb{P}[|X_n - X| \geq \varepsilon] + \mathbb{P}[X \leq x + \varepsilon].$$

Convergence of sequences of rv

By the previous inequalities we infer that for any $\varepsilon > 0$

$$\begin{aligned} F^X(x-\varepsilon) &= \mathbb{P}[X \leq x-\varepsilon] \\ &\leq \liminf_n \mathbb{P}[X_n \leq x] \\ &\leq \limsup_n \mathbb{P}[X_n \leq x] \\ &\leq \mathbb{P}[X \leq x+\varepsilon] = F^X(x+\varepsilon). \end{aligned}$$

Now let ε tend to zero and use that F^X is continuous in x.

Hence $X_n \stackrel{P}{\to} X$ implies $X_n \stackrel{\mathcal{D}}{\to} X$.

Convergence of sequences of rv

Thus we have

$$\begin{array}{ccccc} X_n \stackrel{a.s.}{\to} X & \begin{array}{c}\Longrightarrow \\ \not\Longleftarrow\end{array} & X_n \stackrel{P}{\to} X & \begin{array}{c}\Longleftarrow \\ \not\Longrightarrow\end{array} & X_n \stackrel{L^r}{\to} X \\ & & \Downarrow \not\Uparrow & & \\ & & X_n \stackrel{\mathcal{D}}{\to} X & & \end{array}$$

A useful criterion:

Lemma (Borel-Cantelli-Lemma). If $\sum_{k\geq 1} \mathbb{P}(|X_k - X| > \varepsilon) < \infty$ holds for any $\varepsilon > 0$, then

$$X_k \stackrel{a.s.}{\to} X.$$

Hence in Example 1, $X_n \stackrel{a.s.}{\to} X$.

Also, if $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$, then there is a subsequence $\{n_k\}$ such that

$$\sum_{k\geq 1} \mathbb{P}(|X_{n_k} - X| > \varepsilon) < \infty.$$

Hence $X_n \stackrel{P}{\to} X$ implies $X_{n_k} \stackrel{\text{a.s.}}{\to} X$ along some properly chosen subsequence.

Convergence of sequences of rv

The previous example shows that arrows can sometimes be reverted. Here some other sufficient conditions.

- $X_n \xrightarrow{P} X \Rightarrow$ there exists a subsequence (X_{n_k}) for which $X_{n_k} \xrightarrow{a.s.} X$.
- $X_n \xrightarrow{P} X$ and the X_n's are uniformly integrable$^{*)}$ $\Rightarrow X_n \xrightarrow{L^1} X$.
- $X_n \xrightarrow{\mathcal{D}} \text{const} \Rightarrow X_n \xrightarrow{P} \text{const}$.

$^{*)}$Definition: the X_n's are uniformly integrable $\Leftrightarrow$ $\lim_{\alpha\to\infty} \sup_n \int_{\{|X_n|\geq\alpha\}} |X_n|\, dP = 0$.

The latter condition implies that $\sup_n \mathbb{E}[|X_n|] < \infty$, thus if this condition is violated, our sequence will not be uniformly integrable. (See Example 1).

Limiting theorems

Here are the two most famous limiting results in probability and statistics...

The law of large numbers (LLN):

Let $X_1, X_2, \ldots$ be i.i.d. integrable rv.
Write $\mu := \mathbb{E}[X_1]$.

Then

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \stackrel{a.s.}{\to} \mu.$$

Remark:

- Interpretation for favorable/fair/defavorable games of chance.

Limiting theorems

An example...

Let $X_1, X_2, \ldots$ be i.i.d., with $X_i \sim \mathrm{Bern}(p)$.
Then $\mu = \mathbb{E}[X_i] = p$, so that

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \stackrel{a.s.}{\to} p.$$

Remark: If $Y_1, Y_2, \ldots$ are i.i.d. and $X_i = I\{Y_i \leq x\}$ then $\bar{X}_n = F_n(x)$ is called the *empirical distribution function*. Note that $EX_i = F^Y(x)$. We see that

$$F_n^Y(x) \stackrel{a.s.}{\to} F^Y(x).$$

This result holds uniformly in x and is called *Fundamental Theorem of Statistics.*

Limiting theorems

Let us prove the following weak law of large numbers: if $X_1, X_2, X_3, \ldots$ are i.i.d. with $\mathbb{E}[X_1] = \mu$ and $\mathrm{Var}[X_n] = \sigma^2 < \infty$, then

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

Proof. We have for any $\varepsilon > 0$:

$$\begin{aligned}
\mathbb{P}[|\overline{X}_n - \mu| > \epsilon] &\leq \frac{\mathrm{Var}[\overline{X}_n]}{\varepsilon^2} \\
&\leq \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad (n \to \infty).
\end{aligned}$$

Limiting theorems

The central limit theorem (CLT):

Let $X_1, X_2, \ldots$ be i.i.d. square-integrable rv.
Write $\mu := \mathbb{E}[X_1]$ and $\sigma^2 := \mathrm{Var}[X_1]$.

Then

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} Z, \qquad \text{with } Z \sim \mathcal{N}(0,1).$$

Remarks:

- $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mathbb{E}[\bar{X}_n]}{\sqrt{\mathrm{Var}[\bar{X}_n]}} = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}}$, where $S_n = X_1 + \cdots + X_n$.
- It says sth about the speed of convergence in $\bar{X}_n \xrightarrow{a.s.} \mu$.
- It allows for computing $\mathbb{P}[\bar{X}_n \in B]$ for large n...
- It is valid whatever the distribution of the X_i's!

Two examples...

(A) Let $X_1, X_2, \dots$ be i.i.d., with $X_i \sim \mathrm{Bern}(p)$.
Then $\mu = \mathbb{E}[X_i] = p$ and $\sigma^2 = \mathrm{Var}[X_i] = p(1-p)$, so that

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} Z, \qquad \text{with } Z \sim \mathcal{N}(0,1).$$

Application:

This allows for getting an approximation of $\mathbb{P}[\bar{X}_n \in B]$ for large n, by using that $\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \approx \mathcal{N}(0,1)$ for large n.

(B) Let us assume that we play 30 games of roulette, each time we bet 1 Euro either on red or black. What is the probability that that in the end we made a gain?

Define $X_n = Y_1 + \ldots + Y_n$ where $Y_n = 1$ if we win, or -1 if we loose. Then X_n is the amount of money we won after n games. Here

$$\mathbb{E}[X_{30}] = 30 \times \mathbb{E}[Y_1] = -\frac{30}{37},$$
$$\mathrm{Var}[X_{30}] = 30 \times \mathrm{Var}[Y_1] = 30(1 - (1/37)^2).$$

Hence

$$\begin{aligned}
\mathbb{P}(X_{30} > 0) &= \mathbb{P}\left(X_{30} + \frac{30}{37} > \frac{30}{37}\right) \\
&= \mathbb{P}\left((X_{30} + \frac{30}{37})/\sqrt{30(1 - (1/37)^2)} > \frac{30}{37}/\sqrt{30(1 - (1/37)^2)}\right) \\
&\approx \mathbb{P}(\mathcal{N}(0,1) > 0.15) = 1 - \Phi(0.15) \approx 0.44.
\end{aligned}$$

Limiting theorems

This should be compared to 0.37 for the exact distribution.

The following table shows the normal approximation for the previous problem for sample sizes $n = 30 * k$, $k = 1, \ldots, 10$ and the corresponding exact probabilities.

k	approx	exact
[1,]	0.4411370	0.3701876
[2,]	0.4170575	0.3672352
[3,]	0.3987845	0.3584982
[4,]	0.3835484	0.3490209
[5,]	0.3702720	0.3397180
[6,]	0.3584003	0.3308089
[7,]	0.3476023	0.3223349
[8,]	0.3376615	0.3142836
[9,]	0.3284267	0.3066269
[10,]	0.3197876	0.2993330

Sketch of the proof

Let us assume that $X_1, X_2, \ldots$ are i.i.d.

Then we define $\varphi_X(t) := \mathbb{E}\exp(itX_1)$. This is the so-called *characteristic function* of X_1. (By i.i.d. assumption all X_i have the same characteristic function.)

One can show:

- For random variables X and Y, if $\varphi_X(t) = \varphi_Y(t)$ then $X \stackrel{\mathrm{D}}{=} Y$. (Uniqueness theorem)
- If $\varphi_{X_n}(t) \to \varphi_X(t)$ for all t, then $X_n \stackrel{\mathrm{D}}{\to} X$. (Continuity theorem)
- If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.
- $\varphi_{aX}(t) = \varphi_X(at)$.
- $\varphi_{X+\mu}(t) = e^{it\mu}\varphi_X(t)$.

Sketch of the proof

By assumption $X_1, X_2, \ldots$ are i.i.d., it follows that for $Z_n = \frac{1}{\sigma\sqrt{n}} S_n$, where $S_n = (X_1 - \mu) + (X_2 - \mu) + \ldots + (X_n - \mu)$.

$$\begin{aligned}
\varphi_{Z_n}(t) &= \varphi_{S_n}(t/(\sigma\sqrt{n})) \\
&= (\varphi_{X_1-\mu}(t/(\sigma\sqrt{n})))^n \\
&= \left(\mathbb{E} e^{it(X_1-\mu)/(\sigma\sqrt{n})}\right)^n \\
&\approx \left(\mathbb{E}\Big[\underbrace{1 + \frac{it}{\sigma\sqrt{n}}(X_1 - \mu) - \frac{t^2}{2\sigma^2 n}(X_1 - \mu)^2}_{\text{2 term Taylor expansion}}\Big]\right)^n \\
&= \left(1 - \frac{t^2}{2n}\right)^n \\
&\to e^{-t^2/2} = \mathbb{E} e^{itZ},
\end{aligned}$$

where $Z \sim \mathcal{N}(0, 1)$.