Notes 3 : Modes of convergence

Math 733 - Fall 2013 *Lecturer: Sebastien Roch*

References: [Wil91, Chapters 2.6-2.8], [Dur10, Sections 2.2, 2.3].

1 Modes of convergence

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We will encounter various modes of convergence for sequences of RVs on $(\Omega, \mathcal{F}, \mathbb{P})$.

DEF 3.1 (Modes of convergence) *Let $\{X_n\}_n$ be a sequence of (not necessarily independent) RVs and let X be a RV. Then we have the following definitions.*

- **Convergence in probability:** $\forall \varepsilon > 0$, $\mathbb{P}[|X_n - X| > \varepsilon] \to 0$ *(as $n \to +\infty$); which we denote by $X_n \to_P X$.*
- **Convergence almost sure:** $\mathbb{P}[X_n \to X] = 1$.
- **Convergence in L^p ($p \geq 1$):** $\mathbb{E}|X_n - X|^p \to 0$.

To better understand the relationship between these different modes of convergence, we will need Markov's inequality as well as the Borel-Cantelli lemmas. We first state these, then come back to applications of independent interest below.

1.1 Markov's inequality

LEM 3.2 (Markov's inequality) *Let $Z \geq 0$ be a RV on $(\Omega, \mathcal{F}, \mathbb{P})$. Then for all $a > 0$*

$$\mathbb{P}[Z \geq a] \leq \frac{\mathbb{E}[Z]}{a}.$$

Proof: We have

$$\mathbb{E}[Z] \geq \mathbb{E}[Z \mathbb{1}_{\{Z \geq a\}}] \geq a\mathbb{E}[\mathbb{1}_{\{Z \geq a\}}] = a\mathbb{P}[Z \geq a],$$

where note that the first inequality uses nonnegativity. ■

Recall that (assuming the first and second moments exist):

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

LEM 3.3 (Chebyshev's inequality) *Let X be a RV on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathrm{Var}[X] < +\infty$. Then for all $a > 0$*

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

Proof: Apply Markov's inequality to $Z = (X - \mathbb{E}[X])^2$. ■

An immediate application of Chebyshev's inequality is the following.

THM 3.4 *Let $(S_n)_n$ be a sequence of RVs with $\mu_n = \mathbb{E}[S_n]$ and $\sigma_n^2 = \mathrm{Var}[S_n]$. If $\sigma_n^2/b_n^2 \to 0$, then*

$$\frac{S_n - \mu_n}{b_n} \to_P 0.$$

1.2 Borel-Cantelli lemmas

DEF 3.5 (Almost surely) *Event A occurs* almost surely (a.s.) *if $\mathbb{P}[A] = 1$.*

DEF 3.6 (Infinitely often, eventually) *Let $(A_n)_n$ be a sequence of events. Then we define*

$$A_n \textit{ infinitely often (i.o.)} \equiv \{\omega : \omega \textit{ is in infinitely many } A_n\} \equiv \limsup_n A_n \equiv \bigcap_m \bigcup_{n=m}^{+\infty} A_n.$$

Note that

$$\mathbb{1}_{A_n \textit{ i.o.}} = \limsup_n \mathbb{1}_{A_n}.$$

Similarly,

$$A_n \textit{ eventually (ev.)} \equiv \{\omega : \omega \textit{ is in } A_n \textit{ for all large } n\} \equiv \liminf_n A_n \equiv \bigcup_m \bigcap_{n=m}^{+\infty} A_n.$$

Note that

$$\mathbb{1}_{A_n \textit{ ev.}} = \liminf_n \mathbb{1}_{A_n}.$$

Also we have $(A_n \textit{ ev.})^c = (A_n^c \textit{ i.o.})$.

LEM 3.7 (First Borel-Cantelli lemma (BC1)) *Let $(A_n)_n$ be as above. If*

$$\sum_n \mathbb{P}[A_n] < +\infty,$$

then

$$\mathbb{P}[A_n \textit{ i.o.}] = 0.$$

Proof: This follows trivially from the monotone-convergence theorem (or Fubini's theorem). Indeed let $N = \sum_n \mathbb{1}_{A_n}$. Then

$$\mathbb{E}[N] = \sum_n \mathbb{P}[A_n] < +\infty,$$

and therefore $N < +\infty$ a.s. ∎

EX 3.8 *Let $X_1, X_2, \ldots$ be independent with $\mathbb{P}[X_n = f_n] = p_n$ and $\mathbb{P}[X_n = 0] = 1 - p_n$ for nondecreasing $f_n > 0$ and nonincreasing $p_n > 0$. By (BC1), if $\sum_n p_n < +\infty$ then $X_n \to 0$ a.s.*

The converse is only true in general for IID sequences.

LEM 3.9 (Second Borel-Cantelli lemma (BC2)) *If the events $(A_n)_n$ are independent, then $\sum_n \mathbb{P}[A_n] = +\infty$ implies $\mathbb{P}[A_n$ i.o.$] = 1$.*

Proof: Take $M < N < +\infty$. Then by independence

$$\begin{aligned}\mathbb{P}[\cap_{n=M}^N A_n^c] &= \prod_{n=M}^N (1 - \mathbb{P}[A_n]) \\ &\leq \exp\left(-\sum_{n=M}^N \mathbb{P}[A_n]\right) \\ &\to 0,\end{aligned}$$

as $N \to +\infty$. So $\mathbb{P}[\cup_{n=M}^{+\infty} A_n] = 1$ and further

$$\mathbb{P}\left[\cap_M \cup_{n=M}^{+\infty} A_n\right] = 1,$$

by monotonicity. ∎

EX 3.10 *Let $X_1, X_2, \ldots$ be independent with $\mathbb{P}[X_n = f_n] = p_n$ and $\mathbb{P}[X_n = 0] = 1 - p_n$ for nondecreasing $f_n > 0$ and nonincreasing $p_n > 0$. By (BC1) and (BC2), $X_n \to 0$ a.s. if and only if $\sum_n p_n < +\infty$.*

1.3 Returning to convergence modes

We return to our example.

EX 3.11 *Let $X_1, X_2, \ldots$ be independent with $\mathbb{P}[X_n = f_n] = p_n$ and $\mathbb{P}[X_n = 0] = 1 - p_n$ for nondecreasing $f_n > 0$ and nonincreasing $p_n > 0$. The cases $f_n = 1$, $f_n = \sqrt{n}$, and $f_n = n^2$ are interesting. In the first one, convergence in* *probability (which is equivalent to $p_n \to 0$) and in L^r ($1 \cdot p_n \to 0$) are identical, but a.s. convergence follows from a stronger condition ($\sum_n p_n < +\infty$). In the second one, convergence in L^1 ($\sqrt{n} p_n \to 0$) can happen without convergence a.s. ($\sum_n p_n < +\infty$) or in L^2 ($n p_n \to 0$). Take for instance $p_n = 1/n$. In the last one, convergence a.s. ($\sum_n p_n < +\infty$) can happen without convergence in L^1 ($n^2 p_n \to 0$) or in L^2 ($n^4 p_n \to 0$). Take for instance $p_n = 1/n^2$.*

In general we have:

THM 3.12 (Implications)

- *a.s. $\implies$ in prob (Hint: Fatou's lemma)*
- *$L^p \implies$ in prob (Hint: Markov's inequality)*
- *for $r \geq p \geq 1$, $L^r \implies L^p$ (Hint: Jensen's inequality)*
- *in prob if and only if every subsequence contains a further subsequence that convergence a.s. (Hint: (BC1) for $\implies$ direction)*

Proof: We prove the first, second and (one direction of the) fourth one. For the first one, we need the following lemma.

LEM 3.13 (Reverse Fatou lemma) *Let (S, Σ, μ) be a measure space. Let $(f_n)_n \in (\mathrm{m}\Sigma)^+$ such that there is $g \in (\mathrm{m}\Sigma)^+$ with $f_n \leq g$ for all n and $\mu(g) < +\infty$. Then*

$$\mu(\limsup_n f_n) \geq \limsup_n \mu(f_n).$$

(This follows from applying (FATOU) to $g - f_n$.)

Using the previous lemma on $\mathbb{1}\{|X_n - X| > \varepsilon\}$ gives the result.

For the second claim, note that by Markov's inequality

$$\mathbb{P}[|X_n - X| > \varepsilon] = \mathbb{P}[|X_n - X|^p > \varepsilon^p] \leq \frac{\mathbb{E}|X_n - X|^p}{\varepsilon^p}.$$

One direction of the fourth claim follows from (BC1). Indeed let $(X_{n(m)})_m$ be a subsequence of $(X_n)_n$. Take $\varepsilon_k \downarrow 0$ and let m_k be such that $n(m_k) > n(m_{k-1})$ and

$$\mathbb{P}[|X_{n(m_k)} - X| > \varepsilon_k] \leq 2^{-k},$$

which is summable. Therefore by (BC1), $\mathbb{P}[|X_{n(m_k)} - X| > \varepsilon_k \text{ i.o.}] = 0$, i.e., $X_{n(m_k)} \to X$ a.s. For the other direction, see [D]. ■

As a consequence of the last implication we get the following.

THM 3.14 *If f is continuous and $X_n \to X$ in prob then $f(X_n) \to f(X)$ in probability.*

Proof: For every subsequence $(X_{n(m)})_m$ there is a further subsequence $(X_{n(m_k)})_k$ which converges a.s. and hence $f(X_{n(m_k)}) \to f(X)$ a.s. But this implies that $f(X_n) \to f(X)$ in probability. ∎

Our example and theorem show that a.s. convergence does not come from a topology (or in particular from a metric). In contrast, it is possible to show that convergence in probability corresponds to the Ky Fan metric

$$\alpha(X,Y) = \inf\{\varepsilon \geq 0 \,:\, \mathbb{P}[|X-Y| > \varepsilon] \leq \varepsilon\}.$$

See [D].

1.4 Statement of laws of large numbers

Our first goal will be to prove the following.

THM 3.15 (Strong law of large numbers) *Let $X_1, X_2, \ldots$ be IID with $\mathbb{E}|X_1| < +\infty$. (In fact, pairwise independence suffices.) Let $S_n = \sum_{k\leq n} X_k$ and $\mu = \mathbb{E}[X_1]$. Then*

$$\frac{S_n}{n} \to \mu, \quad a.s.$$

If instead $\mathbb{E}|X_1| = +\infty$ then

$$\mathbb{P}\left[\lim_n \frac{S_n}{n} \text{ exists} \in (-\infty, +\infty)\right] = 0.$$

and

THM 3.16 (Weak law of large numbers) *Let $(X_n)_n$ be IID and $S_n = \sum_{k\leq n} X_k$. A necessary and sufficient condition for the existence of constants $(\mu_n)_n$ such that*

$$\frac{S_n}{n} - \mu_n \to_P 0,$$

is

$$n\,\mathbb{P}[|X_1| > n] \to 0.$$

In that case, the choice

$$\mu_n = \mathbb{E}[X_1 \mathbb{1}_{|X_1|\leq n}],$$

works.

Before we give the proofs of these theorems, we discuss further applications of Markov's inequality and the Borel-Cantelli lemmas.

2 Further applications...

2.1 ...of Chebyshev's inequality

Chebyshev's inequality and Theorem 3.4 can be used to derive limit laws in some cases where sequences are not necessarily IID. We give several important examples from [D].

EX 3.17 (Occupancy problem) *Suppose we throw r balls into n bins independently uniformly at random. Let N_n be the number of empty boxes. If A_i is the event that the i-th bin is empty, we have*

$$\mathbb{P}[A_i] = \left(1 - \frac{1}{n}\right)^r,$$

so that $N_n = \sum_{k \leq n} \mathbb{1}_{A_k}$ (not independent) and

$$E[N_n] = n\left(1 - \frac{1}{n}\right)^r.$$

In particular, if $r/n \to \rho$ we have

$$\frac{\mathbb{E}[N_n]}{n} \to e^{-\rho}.$$

Because there is no independence, the variance calculation is trickier. Note that

$$\mathbb{E}[N_n^2] = \mathbb{E}\left[\left(\sum_{m=1}^{n} \mathbb{1}_{A_m}\right)^2\right] = \sum_{1 \leq m, m' \leq n} \mathbb{P}[A_m \cap A_{m'}],$$

and

$$\begin{aligned}
\mathrm{Var}[N_n] &= \mathbb{E}[N_n^2] - (\mathbb{E}[N_n])^2 \\
&= \sum_{1 \leq m, m' \leq n} [\mathbb{P}[A_m \cap A_{m'}] - \mathbb{P}[A_m]\mathbb{P}[A_{m'}]] \\
&= n(n-1)[(1-2/n)^r - (1-1/n)^{2r}] + n[(1-1/n)^r - (1-1/n)^{2r}] \\
&= o(n^2) + O(n),
\end{aligned}$$

where we divided the sum into cases $m \neq m'$ and $m = m'$. Taking $b_n = n$ in Theorem 3.4, we have

$$\frac{N_n}{n} \to_P e^{-\rho}.$$

EX 3.18 (Coupon's collector problem) *Let $X_1, X_2, \ldots$ be IID uniform in $[n] = \{1, \ldots, n\}$. We are interested in the time it takes to see every element in $[n]$ at least once. Let*

$$\tau_k^n = \inf\{m \,:\, |\{X_1, \ldots, X_m\}| = k\},$$

be the first time we collect k different items, with the convention $\tau_0^n = 0$. Let $T_n = \tau_n^n$. Define $X_{n,k} = \tau_k^n - \tau_{k-1}^n$ and note that the $X_{n,k}$'s are independent (but not identically distributed) with geometric distribution with parameter $1 - (k-1)/n$. Recall that a geometric RV N with parameter p has law

$$\mathbb{P}[N = i] = p(1-p)^{i-1},$$

and moments

$$\mathbb{E}[N] = \frac{1}{p},$$

and

$$\mathrm{Var}[N] = \frac{1-p}{p^2} \left(\leq \frac{1}{p^2}\right).$$

Hence

$$\mathbb{E}[T_n] = \sum_{k=1}^{n} \left(1 - \frac{k-1}{n}\right)^{-1} = n \sum_{m=1}^{n} \frac{1}{m} \sim n \log n,$$

and

$$\mathrm{Var}[T_n] \leq \sum_{k=1}^{n} \left(1 - \frac{k-1}{n}\right)^{-2} = n^2 \sum_{m=1}^{n} \frac{1}{m^2} \leq Cn^2,$$

for some $C > 0$ not depending on n.

Taking $b_n = n \log n$ in Theorem 3.4 gives

$$\frac{T_n - n\sum_{m=1}^{n} m^{-1}}{n \log n} \to_P 0,$$

or

$$\frac{T_n}{n \log n} \to_P 1.$$

The previous example involved a so-called triangular array $\{X_{n,k}\}_{n \geq 1, 1 \leq k \leq n}$.

EX 3.19 (Random permutations) *Any permutation can be decomposed into cycles. E.g., if $\pi = [3, 9, 6, 8, 2, 1, 5, 4, 7]$, then $\pi = (136)(2975)(48)$. In fact, a uniform permutation can be generated by following a cycle until it closes, then starting over from the smallest unassigned element, and so on. Let $X_{n,k}$ be the* *indicator that the k-th element in this construction precedes the closure of a cycle. E.g., we have $X_{9,3} = X_{9,7} = X_{9,9} = 1$. The construction above implies that the $X_{n,k}$'s are independent and*

$$\mathbb{P}[X_{n,j} = 1] = \frac{1}{n-j+1}.$$

That is because only one of the remaining elements closes the cycle. Letting $S_n = \sum_{k\leq n} X_{n,k}$ be the number of cycles in π we have

$$\mathbb{E}[S_n] = \sum_{j=1}^{n} \frac{1}{n-j+1} \sim \log n,$$

and

$$\mathrm{Var}[S_n] = \sum_{j=1}^{n} \mathrm{Var}[X_{n,j}] \leq \sum_{j=1}^{n} \mathbb{E}[X_{n,j}^2] = \sum_{j=1}^{n} \mathbb{E}[X_{n,j}] = \mathbb{E}[S_n].$$

Taking $b_n = \log n$ in Theorem 3.4 we have

$$\frac{S_n}{\log n} \to_P 1.$$

2.2 ...of (BC1)

EX 3.20 (Head runs) *Let $(X_n)_{n\in\mathbb{Z}}$ be IID with $\mathbb{P}[X_n = 1] = \mathbb{P}[X_n = -1] = 1/2$. Let*

$$\ell_n = \max\{m \geq 1 \,:\, X_{n-m+1} = \cdots = X_n = 1\},$$

(with $\ell_n = 0$ if $X_n = -1$) and

$$L_n = \max_{1\leq m\leq n} \ell_m.$$

Note that $\mathbb{P}[\ell_n = k] = (1/2)^{k+1}$ for all n, k. (The $+1$ in the exponent is for the first -1.) We will prove

$$\frac{L_n}{\log_2 n} \to 1, \quad \textit{a.s.}$$

For the lower bound, it suffices to divide the sequence into disjoint blocks to use independence. Take blocks of size $[(1-\varepsilon)\log_2 n] + 1$ so that a block is all-1 with probability at least

$$2^{-[(1-\varepsilon)\log_2 n]-1} \geq n^{-(1-\varepsilon)}/2.$$

For n large enough

$$\mathbb{P}[L_n \leq (1-\varepsilon)\log_2 n] \leq \left(1 - n^{-(1-\varepsilon)}/2\right)^{n/\log_2 n} \leq \exp\left(-\frac{n^\varepsilon}{\log_2 n}\right),$$

which is summable. By (BC1),

$$\liminf_n \frac{L_n}{\log_2 n} \geq 1 - \varepsilon, \quad \textit{a.s.}$$

The upper bound follows from (BC1). Indeed note that, for any $\varepsilon > 0$,

$$\mathbb{P}[\ell_n \geq (1+\varepsilon)\log_2 n] = \sum_{k \geq (1+\varepsilon)\log_2 n} \left(\frac{1}{2}\right)^{k+1} \leq n^{-(1+\varepsilon)},$$

so that

$$\mathbb{P}[\ell_n \geq (1+\varepsilon)\log_2 n \ \textit{i.o.}] = 0,$$

Hence, there is N_ε (random) such that $\ell_n \leq (1+\varepsilon)\log_2 n$ for all $n \geq N_\varepsilon$ and note that the ℓ_n's with $n < N_\varepsilon$ are finite a.s. as they have a finite expectation. Therefore

$$\limsup_n \frac{L_n}{\log_2 n} \leq 1 + \varepsilon, \quad \textit{a.s.}$$

Since ε is arbitrary, we get the upper bound.

2.3 ...of (BC2)

We will need a more refined version of (BC2).

THM 3.21 *If $A_1, A_2, \ldots$ are pairwise independent and $\sum_n \mathbb{P}[A_n] = +\infty$ then*

$$\frac{\sum_{m=1}^n \mathbb{1}_{A_m}}{\sum_{m=1}^n \mathbb{P}[A_m]} \to 1, \quad \textit{a.s.}$$

Proof: Convergence in probability follows from Chebyshev's inequality. Let $X_k = \mathbb{1}_{A_k}$ and $S_n = \sum_{k \leq n} X_k$. Then by pairwise independence

$$\mathrm{Var}[S_n] = \sum_{k\leq n} \mathrm{Var}[X_k] \leq \sum_{k\leq n} \mathbb{E}[X_k^2] = \sum_{k\leq n} \mathbb{E}[X_k] = \sum_{k\leq n} \mathbb{P}[A_k] = \mathbb{E}[S_n],$$

using $X_k \in \{0, 1\}$. Then

$$\mathbb{P}[|S_n - \mathbb{E}[S_n]| > \delta \mathbb{E}[S_n]] \leq \frac{\mathrm{Var}[S_n]}{\delta^2 \mathbb{E}[S_n]^2} \leq \frac{1}{\delta^2 \mathbb{E}[S_n]} \to 0,$$

by assumption. In particular,

$$\frac{S_n}{\mathbb{E}[S_n]} \to_P 1.$$

We use a standard trick to obtain almost sure convergence. The idea is to take subsequences, use (BC1), and sandwich the original sequence.)

1. Take
$$n_k = \inf\{n \,:\, \mathbb{E}[S_n] \geq k^2\},$$
and let $T_k = S_{n_k}$. Since $\mathbb{E}[X_n] \leq 1$ we have in particular $k^2 \leq \mathbb{E}[T_k] \leq k^2 + 1$. Using Chebyshev again,
$$\mathbb{P}[|T_k - \mathbb{E}[T_k]| > \delta \mathbb{E}[T_k]] \leq \frac{1}{\delta^2 k^2},$$
which is summable so that, using (BC1) and the fact that δ is arbitrary,
$$\frac{T_k}{\mathbb{E}[T_k]} \to 1, \quad \text{a.s.}$$

2. For $n_k \leq n < n_{k+1}$, we have by monotonicity
$$\frac{T_k}{\mathbb{E}[T_{k+1}]} \leq \frac{S_n}{\mathbb{E}[S_n]} \leq \frac{T_{k+1}}{\mathbb{E}[T_k]}$$
Finally, note that
$$\frac{\mathbb{E}[T_k]}{\mathbb{E}[T_{k+1}]} \frac{T_k}{\mathbb{E}[T_k]} \leq \frac{S_n}{\mathbb{E}[S_n]} \leq \frac{T_{k+1}}{\mathbb{E}[T_k]} \frac{\mathbb{E}[T_{k+1}]}{\mathbb{E}[T_k]},$$
and
$$k^2 \leq \mathbb{E}[T_k] \leq \mathbb{E}[T_{k+1}] \leq (k+1)^2 + 1.$$
Since the ratio of the two extremes terms goes to 1, the ratio of the expectations goes to 1 and we are done.

■

We will see this argument again when we prove the strong law of large numbers.

EX 3.22 (Record values) *Let $X_1, X_2, \ldots$ be a sequence of IID RVs with a continuous DF F corresponding to, say, an individual's times in a race. Let*

$$A_k = \left\{ X_k > \sup_{j<k} X_j \right\},$$

that is, that time k is a new record. Let $R_n = \sum_{m \leq n} \mathbb{1}_{A_m}$, we will prove that

$$\frac{R_n}{\log n} \to 1, \quad \textit{a.s.}$$

Because F is continuous, there is no atom and $\mathbb{P}[X_j = X_k] = 0$ for $j \neq k$. Let $Y_1^n > \cdots > Y_n^n$ be the sequence $X_1, \ldots, X_n$ in decreasing order. By the IID assumption, the permutation $\pi_n(i) = j$ if $X_i = Y_j^n$ is clearly uniform by symmetry. In particular,

$$\mathbb{P}[A_n] = \mathbb{P}[\pi_n(n) = 1] = \frac{1}{n}.$$

Moreover, for any $m_1 < m_2$, note that on A_{m_2} the distribution of the relative ordering of the X_is for $i < m_2$ is unchanged by symmetry and therefore

$$\frac{\mathbb{P}[A_{m_1} \cap A_{m_2}]}{\mathbb{P}[A_{m_2}]} = \mathbb{P}[A_{m_1}] = \frac{1}{m_1}.$$

We have proved that the A_k's are pairwise independent and that $\mathbb{P}[A_k] = 1/k$. Now use the fact that

$$\sum_{i=1}^{n} \frac{1}{i} \sim \log n,$$

and the previous theorem. This proves the claim.

References

[Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.