MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2018
Lecture 19

Uniform integrability, convergence of series

Contents

1. L_1 convergence (aka convergence in mean), L_1 LLN.
2. Uniform integrability
3. Convergence of series of independent summands

1 CONVERGENCE IN L_1

Definition 1 (Convergence in mean)**.** *A sequence of integrable random varibles X_j is said to converge in L_1 to X (also known as "convergence in mean"), denoted $X_j \overset{L_1}{\to} X$ if*

$$\mathbb{E}[|X_j - X|] \to 0 \qquad j \to \infty\,.$$

For $p > 1$ we define $X_j \overset{L_p}{\to} X$ if $\mathbb{E}[|X_j|^p] < \infty$ and $\mathbb{E}[|X_j - X|^p] \to 0$.

Some simple properties are given below:

Proposition 1. *(i) $X_n \overset{L_1}{\to} X$ implies $\mathbb{E}[|X|] < \infty$.*

(ii) $X_j \overset{L_1}{\to} X$ implies $X_j \overset{\text{i.p.}}{\to} X$

(iii) $X_j \overset{L_1}{\to} X$ does not imply and is not implied by $X_j \overset{\text{a.s.}}{\to} X$.

(iv) The space of integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ modulo almost-sure equivalence is a Banach space, denoted as $L_1(\Omega, \mathcal{F}, \mathbb{P})$ with norm $\|X\|_1 \triangleq \mathbb{E}[|X|]$. Similarly for $L_p(\Omega, \mathcal{F}, \mathbb{P})$.

Proof: (i) Follows from taking expectation in the triangle inequality:

$$|X| \leq |X_n - X| + |X_n| \tag{1}$$

(ii) and (iii) is an exercise. (iv) is outside the scope of this class. □

Our goal is to show the following the following (third!) variation of the LLN:

Proposition 2 (L_1-LLN)**.** *Let X_j be iid random variables with finite expectation, then*

$$\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{L_1} \mathbb{E}[X]$$

The proof of this proposition follows by Theorem 1 and Corollary 1 below.

2 UNIFORM INTEGRABILITY

Definition 2. *A collection of random variables $X_\alpha, \alpha \in S$ is uniformly integrable if*

$$k(b) \triangleq \sup_\alpha \mathbb{E}[|X_\alpha| 1\{|X_\alpha| > b\}] \to 0 \quad b \to \infty\,. \tag{2}$$

Some useful criteria for checking u.i.:

Proposition 3. *The following hold:*

(i) If $\mathbb{E}[|X|] < \infty$ then $\{X\}$ is u.i.

(ii) When $X_\alpha \stackrel{d}{=} Y_\alpha$ then $\{X_\alpha\}$–u.i. iff $\{Y_\alpha\}$–u.i.

(iii) $\{X_\alpha\}$-u.i. iff $\{X_\alpha\}$ is L_1-bounded and uniformly continuous:

$$\sup_\alpha \mathbb{E}[|X_\alpha|] < \infty \tag{3}$$

$$\sup_\alpha \mathbb{E}[|X_\alpha| 1_E] \to 0 \quad \text{as } \mathbb{P}[E] \to 0 \tag{4}$$

(iv) If $G : \mathbb{R}_+ \to \mathbb{R}_+$ is such that $\frac{G(t)}{t} \to \infty$ as t grows without bound[1] then

$$\sup \mathbb{E}[G(|X_\alpha|)] < \infty \quad \Rightarrow \quad \{X_\alpha\}\text{–u.i.}$$

[1] Some typical choices are $G(t) = t^2$, $|t|^{1+\epsilon}$ and $|t \log t|$.

(v) If $X_n \stackrel{\text{a.s.}}{\to} X$ and $\mathbb{E}[|X_n|] \to \mathbb{E}[|X|]$ then $\{X_n\}$ is u.i.

Proof: (i) follows from the MCT, (ii) is obvious from the definition, (v) is part of the homework.

For (iii), first notice that $\mathbb{E}[|X_\alpha|] \le k(b) + b$ for every $b > 0$ and thus (3) holds. Similarly, notice that for every b:

$$\mathbb{E}[|X_\alpha| 1_E] \le \mathbb{E}[|X_\alpha| 1\{|X_\alpha| > b\}] + b\mathbb{P}[E] \le k(b) + b\mathbb{P}[E]$$

and thus by taking $\mathbb{P}[E] \to 0$ and $b \to \infty$ we prove (4). Conversely, if (3) and (4) hold, but $\{X_\alpha\}$ is not uniformly integrable then for some sequence α_k and $\epsilon_0 > 0$ we have

$$\mathbb{E}[|X_{\alpha_k}| 1\{|X_{\alpha_k}| > k\}] \ge \epsilon_0 > 0 \tag{5}$$

On the other hand, by (3) and Markov inequality $\mathbb{P}[|X_{\alpha_k}| > k] \to 0$. Consequently, (5) contradicts (4).

Finally, to see (iv) just notice that for every $a > 0$ there exists $b > 0$ such that $\frac{G(t)}{t} \ge a$ for all $t > b$. Then,

$$G(|X_\alpha|) \ge a|X_\alpha| 1\{|X_\alpha| > b\}$$

and taking expectation here we obtain:

$$k(b) \le \frac{1}{a} \sup_\alpha \mathbb{E}[G(|X_\alpha|)]$$

from which (2) follows by taking $a \to \infty$. □

As a simple consequence of the Proposition we get:

Corollary 1. *Let X_j be identically distributed (not necessarily independent!) and integrable. Then collection of normalized sums $\left\{\frac{1}{n}\sum_{j=1}^n X_j, n = 1, \ldots\right\}$ is uniformly integrable.*

Proof: Indeed, by Proposition 3(i) and (ii) we get that $\{X_j, j = 1, \ldots\}$ is uniformly integrable. Now defining $Y_n = \frac{1}{n}\sum_{j=1}^n X_j$ we have

$$\sup_n \mathbb{E}[|Y_n|] \le \mathbb{E}[|X|] < \infty$$

and on the other hand

$$\sup_n \mathbb{E}[|Y_n| 1_E] \le \sup_n \mathbb{E}[|X_n| 1_E] \to 0 \quad \text{as } \mathbb{P}[E] \to 0 \,,$$

where the last step follows by (4) applied to $\{X_j\}$. Uniform integrability of $\{Y_j\}$ then follows from Proposition 3 (iii). □

The main value of studying uniform integrability is the following:

Theorem 1. *We have*

$$X_n \stackrel{L_1}{\to} X \qquad \Longleftrightarrow \qquad X_n \stackrel{\text{i.p.}}{\to} X \text{ and } \{X_n\}\text{– u.i.}$$

Proof: The $\Rightarrow$ direction follows from Markov's inequality and Proposition 3(iii). Indeed, by (1) we have the inequality

$$\mathbb{E}[|X_n|1_E] \le \mathbb{E}[|X|1_E] + \mathbb{E}[|X_n - X|1_E] \tag{6}$$

For very large $n \ge n_0$ the second term is smaller than ϵ and hence

$$\limsup_{\mathbb{P}[E]\to 0} \left(\sup_n \mathbb{E}[|X_n|1_E]\right) \le \epsilon + \limsup_{\mathbb{P}[E]\to 0} \left(\mathbb{E}[|X|1_E] + \max_{1\le n\le n_0} \mathbb{E}[|X_n - X|1_E]\right)$$

where the second term is zero by (4) because $\{|X|, |X_1 - X|, \ldots |X_{n_0} - X|\}$ is a uniformly integrable collection. Consequently, taking $\epsilon \to 0$ we have shown

$$\sup_n \mathbb{E}[X_n 1_E] \to 0 \quad \mathbb{P}[E] \to 0$$

Setting $E =$ in (6) we verify (4). Thus Proposition 3(iii) implies that the infinite collection $\{X_n\}$ is also u.i.

For the converse direction, we first notice that by characterization of convergence in probability there must exist a subsequence $X_{n_k} \stackrel{\text{a.s.}}{\to} X$. Then by Fatou's lemma and (3) we have

$$\mathbb{E}[|X|] = \mathbb{E}[\liminf_k |X_{n_k}|] \le \liminf_k \mathbb{E}[|X_{n_k}|] < \infty$$

Thus, the limit random variable is integrable and consequently (Exercise!) collection of nonnegative random variables $\{Y_n, n = 1, \ldots\}$ is u.i. and $Y_n \stackrel{\text{i.p.}}{\to} 0$, where

$$Y_n \triangleq |X_n - X|\,.$$

Then, we have for every $\epsilon > 0$

$$\mathbb{E}[Y_n] = \mathbb{E}[Y_n 1\{Y_n > \epsilon\}] + \mathbb{E}[Y_n 1\{Y_n \le \epsilon\}] \tag{7}$$

$$\le \epsilon + \mathbb{E}[Y_n 1\{Y_n > \epsilon\}]\,. \tag{8}$$

Since $Y_n \stackrel{\text{i.p.}}{\to} 0$ we have $\mathbb{P}[Y_n > \epsilon] \to 0$. Then by (4) the second term converges to zero as $n \to \infty$. Hence, for all $\epsilon > 0$

$$\limsup_{n\to\infty} \mathbb{E}[Y_n] \le \epsilon\,,$$

which shows $Y_n \stackrel{L_1}{\to} 0$. □

As a corollary we obtain a result we assumed before (in proving that convergence of characteristic functions implies convergence in distribution).

Corollary 2. *Let $f_n(x) \to f(x) \quad \forall x \in \mathbb{R}$ be a pointwise convergent sequence of pdfs. If $X_n \sim f_n$ and $X \sim f$ then $X_n \stackrel{\text{d}}{\to} X$.*

Proof. Let $\phi(x) = \frac{1}{2}f(x) + \sum_{n=1}^{\infty} 2^{-n-1} f_n(x)$. It is clear that $\phi(x)$ is another pdf. Let $(\phi, \mathcal{F}, \mathbb{P})$ be defined as $\Omega = \mathbb{R}, \mathcal{F} = \mathcal{B}, \mathbb{P}(dx) = \phi(x)dx$ and define random variables $Y_n(x) = f_n(x)/\phi(x)$. Note that as a consequence of our definition of ϕ this ratio is well-defined almost everywhere: $\phi(x) = 0$ implies $f_n(x) = 0$ and $\mathbb{P}(\{x : \phi(x) = 0\}) = 0$. Similarly, define $Y(x) = f(x)/\phi(x)$. We have $Y_n \stackrel{\text{a.s.}}{\to} Y$.

Furthermore, by construction $\mathbb{E}[|Y_n|] = 1 = \mathbb{E}[|Y|]$. Thus, by Prop. 3(v) the collection $\{Y_n\}$ is u.i.. Consequently, from the previous Theorem we have $\mathbb{E}[|Y_n - Y|] \to 0$. Rewriting this last statement explicitly, we have shown

$$\int_{\mathbb{R}} |f_n(x) - f(x)|dx \to 0 \qquad n \to \infty\,. \tag{9}$$

In particular, for any $E \in \mathcal{B}$ we have

$$P_{X_n}[E] = \int_E f_n(x) \to \int_E f(x) = P_X[E]\,,$$

and taking $E = (-\infty, a]$ shows convergence of CDFs of X_n to the CDF of X at every point a.

(In fact, (9) is usually stated as "distribution of X_n converges to the distribution of X in total-variation". This is a stronger mode of convergence than convergence in distribution.) □

3 SUMS OF INDEPENDENT RANDOM VARIABLES

A classical topic tightly related to the SLLN is convergence of sums

$$S_n = \sum_{j=1}^{n} X_j$$

when X_j are independent (and not identically distributed, of course). There is a great deal results about properties of S_n, and here we will only mention a core principle: *Convergence behavior of S_n, its central moments, concentration properties, etc are largely encoded in the behavior of* $\sum_{j=1}^n \mathrm{var}[X_j]$.

We start with an example. Consider two independent sequences:

$$\mathbb{P}[X_n = \pm 1] = \frac{1}{2n}, \quad \mathbb{P}[X_n = 0] = 1 - \frac{1}{n}\,, \tag{10}$$

$$\mathbb{P}[Y_n = \pm\frac{1}{n}] = \frac{1}{2}\,. \tag{11}$$

First, we notice that $\mathbb{E}[|X_n|] = \mathbb{E}[|Y_n|] = \frac{1}{2n}$, so that sum of first moments diverges at the same speed. Furthermore, both series $\sum X_n$ and $\sum Y_n$ do not absolutely converge:

$$\mathbb{P}\left[\sum_n |X_n| = +\infty\right] = \mathbb{P}\left[\sum_n |Y_n| = +\infty\right] = 1\,.$$

Indeed, for Y_n this is obvious as $|Y_n| = \frac{1}{n}$, while for X_n it follows from Borel-Cantelli that: $\mathbb{P}[|X_n| = 1\text{–i.o}] = 1$.

So far, we see that $\sum X_n$ and $\sum Y_n$ behave quite similarly. However, as we will see next it turns out that

$$\mathbb{P}\left[\sum_n X_n \text{ converges}\right] = 0 \qquad \sum_n \mathrm{var}[X_n] = +\infty \tag{12}$$

$$\mathbb{P}\left[\sum_n Y_n \text{ converges}\right] = 1 \qquad \sum_n \mathrm{var}[Y_n] < +\infty \tag{13}$$

The explanation of this phenomena is the following: While both series diverge absolutely, the rapidly decreasing variances of terms in Y_n allows for "sign cancellation" effect to kick in making the series $\sum Y_n$ converge (similar to convergence of $\sum_n \frac{(-1)^n}{n}$).

Theorem 2 (Kolmogorov, Khintchine)**.** *Let X_j be independent and*

$$\mu \triangleq \sum_{j=1}^{\infty} \mathbb{E}[X_j]\,, \quad |\mu| < \infty\,, \tag{14}$$

$$\sigma^2 \triangleq \sum_{j=1}^{\infty} \mathrm{var}[X_j] < \infty \tag{15}$$

then

$$S_n = \sum_{j=1}^{n} X_j$$

converges almost surely and in L_2 to a limit S with $\mathbb{E}[S] = \mu$, $\mathrm{var}[S] = \sigma^2$.

Conversely, if $|X_j| \le c$ for some constant c and $S_n \stackrel{\text{a.s.}}{\to} S$ with real-valued S, then conditions (14)-(15) hold.

Proof: We prove the direct part. Without loss of generality we assume $\mathbb{E}[X_j] = 0$. As we have shown in the homework (Cauchy criterion of almost sure convergence) it is sufficient to show that

$$\mathbb{P}[\sup_{k\ge 1} |S_{n+k} - S_n| > \epsilon] \to 0 \quad n \to \infty \tag{16}$$

Kolmogorov's inequality (see Theorem following the proof) shows that

$$\mathbb{E}[\sup_{k\ge 1} |S_{n+k} - S_n|^2] \le 2 \sum_{n}^{\infty} \mathrm{var}[X_j]\,. \tag{17}$$

By Chebyshev's inequality we obtain then

$$\mathbb{P}[\sup_{k\ge 1} |S_{n+k} - S_n| > \epsilon] \le \frac{2}{\epsilon^2} \sum_{n}^{\infty} \mathrm{var}[X_j] \tag{18}$$

Since sum of variances converges, the left-hand side of (18) decreases to 0 as $n \to \infty$ and thus (16) is shown. The proof of $S_n \stackrel{\text{a.s.}}{\to} S$ is complete.

Notice that (17) with $n = 0$ shows that "life-time maximum"

$$M_\infty \triangleq \sup_{n\ge 1} |S_n|$$

has finite second moment. Since

$$|S_n - S| \le 2M_\infty\,,$$

by the DCT it follows that

$$\mathbb{E}[|S_n - S|^2] \to 0$$

and similarly for $\mathbb{E}[S_n] \to \mathbb{E}[S]$, $\mathbb{E}[S_n^2] \to \mathbb{E}[S]$.

We proceed to proving the converse. First, assume $\mathbb{E}[X_j] = 0$ and suppose $S_n \overset{\text{a.s.}}{\to} S$ but

$$D_n = \sum_{j=1}^{n} \text{var}[X_j] \nearrow \infty$$

Then, notice that

$$\mathbb{E}[|X_j|^3] \leq \mathbb{E}[|X_j|^2]c$$

and thus

$$\sum_{j=1}^{n} \mathbb{E}[|X_j|^3] \leq D_n c$$

Consequently, by the CLT for non-identically distributed random variables we have

$$\frac{1}{\sqrt{D_n}} S_n \overset{\text{d}}{\to} Z \sim \mathcal{N}(0,1)\,.$$

On the other hand, we have for every t and $s > 0$ and for all n large enough

$$\mathbb{P}[S_n > t] \geq \mathbb{P}[S_n > s\sqrt{D_n}]\,. \tag{19}$$

Since $S_n \overset{\text{d}}{\to} S$ we also have for all t such that $\mathbb{P}[S = t] = 0$ that

$$\mathbb{P}[S_n > t] \to \mathbb{P}[S > t]\,.$$

However, upon taking the limit in (19) as $n \to \infty$ we get for all t and $s > 0$:

$$\mathbb{P}[S > t] \geq \mathbb{P}[Z > s]$$

Taking $s \to 0$ we get

$$\mathbb{P}[S > t] \geq \frac{1}{2} \qquad \forall t \in \mathbb{R}$$

which is a contradiction, as no distribution of S can satisfy such inequality.

Next, if $\mathbb{E}[X_j] = \mu_j$, then let $Y_j = X_j - X_j'$, where X_j' is an independent copy of X_j. In this way $\mathbb{E}[Y_j] = 0$, $\text{var}[Y_j] = 2\text{var}[X_j]$ and

$$\sum_{j=1}^{n} Y_j = S_n - S_n' \overset{\text{a.s.}}{\to} S - S'\,.$$

Hence, by the previous argument we have

$$\sum_{j=1}^{n} \mathrm{var}[X_j] < \infty$$

and by the direct part of the theorem

$$\sum_{j=1}^{n} (X_j - \mu_j) = S_n - \sum_{j=1}^{n} \mu_j$$

converges almost surely. Since S_n converges by assumption, so must do $\sum \mu_j$. $\square$

Remark: Conditions (14) are necessary for convergence $S_n \overset{\text{a.s.}}{\to} S$ only under assumption of the boundedness of X_j (see homework). We also mention that instead of relying on the CLT in the proof of the converse direction, we may have followed a more conventional route, based on the inequality

$$\mathbb{P}[\max_{1 \le k \le n} |S_k| > a] \ge 1 - \frac{(a+c)^2}{\mathrm{var}[S_n]}\,.$$

The proof of this inequality, however, would appear rather unnatural without mentioning stopping times. Either method, however, really just shows that condition $|X_j| \le c$ guarantees the width of the distribution of S_n has the same order as $\sqrt{\mathrm{var}[S_n]}$. (For unbounded X_j, rare large jumps may significantly increase the variance, while having very little effect on the bulk of the distribution of S_n).

Theorem 3 (Kolmogorov)**.** *Let X_j be independent, zero-mean with finite second moments and let*

$$M_n = \sup_{1 \le k \le n} \left| \sum_{j=1}^{k} X_j \right|, 1 \le n \le \infty\,.$$

Then we have for any $1 \le n \le \infty$

$$\mathbb{E}\left[|M_n|^2\right] \le 2 \sum_{j=1}^{n} \mathbb{E}[X_j^2]\,.$$

Proof: The case of $n = \infty$ follows from the case of finite n by the MCT. Let

$$S_k = X_1 + \cdots + X_k\,, \tag{20}$$

$$A_n = \max_{1 \le k \le n} S_k\,, \tag{21}$$

Note that for $n = 1$ we clearly have

$$\mathbb{E}[A_n^2] \leq \mathbb{E}[S_n^2]\,. \tag{22}$$

Assume (by induction) that (22) is shown for all sums of upto $n-1$ random variables. Then, notice that

$$A_n = X_1 + \max(0, X_2, X_2 + X_3, \ldots, \sum_2^n X_j)\,.$$

Since first and second terms are independent and $\mathbb{E}[X_1] = 0$ we get

$$\mathbb{E}[A_n^2] = \mathbb{E}[X_1^2] + \mathbb{E}[\max(0, X_2, X_2 + X_3, \ldots, \sum_2^n X_j)^2] \tag{23}$$

$$\leq \mathbb{E}[X_1^2] + \mathbb{E}[\max(X_2, X_2 + X_3, \ldots, \sum_2^n X_j)^2] \tag{24}$$

$$\leq \sum_{j=1}^n \mathbb{E}[X_j^2]\,, \tag{25}$$

where the first inequality follows from $(x^+)^2 \leq x^2$ and the second one is by the inductive assumption. Thus (22) holds for all n.

By symmetry, we also must have

$$\mathbb{E}[B_n^2] \leq \mathbb{E}[S_n^2]\,, \qquad B_n = \max_{1 \leq k \leq n} -S_k\,.$$

Finally, since $M_n^2 = \max(A_n^2, B_n^2)$ and using $\max(a,b) \leq a + b$ we complete the proof. □