Lecture 18. Glivenko–Cantelli theorem

Paul Delatte
delatte@usc.edu
University of Southern California

Last updated: 31 December, 2022

Theorem 1 ((Generalized) **Glivenko–Cantelli Theorem** (with Bracketing))**.** *Let $(X, \mathcal{A}, P)$ be a probability space and $\mathcal{F}$ a collection of measurable functions $f : X \to \mathbb{R}$. If $N_{[\,]}(\varepsilon, \mathcal{F}, L_1(P)) < \infty$ for all $\varepsilon > 0$, then $\mathcal{F}$ is P-Glivenko–Cantelli, that is,*

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} := \left(\sup_{f \in \mathcal{F}} \left| \mathbb{P}_n f - Pf \right| \right) \xrightarrow[n\to\infty]{a.s.^*} 0$$

where $\mathbb{P}_n f := \int f \, d\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n f(X_i)$ with $X_i \sim_{i.i.d.} P$ and $\mathbb{P} f := \int f \, d\mathbb{P}$, the superscript $*$ denotes outer probabilities defined for any subset B of X as $P^*(B) = \inf\{P(A) : A \supseteq B, A \in \mathcal{A}\}$, and $N_{[\,]}(\varepsilon, \mathcal{F}, L_1(P))$ is the L_1 bracketing number of $\mathcal{F}$ defined as the minimum number of L_1-ε-brackets $[l, u]_\varepsilon = \{f \in \mathcal{F} : l \le f \le u$ for some integrable functions $l, u : X \to \mathbb{R}$ with $\|u - l\|_1 \le \varepsilon\}$ to cover $\mathcal{F}$.

Proof. Let $\varepsilon > 0$. Pick finitely many L_1-ε-brackets $([l_i, u_i]_\varepsilon)_{i=1,\dots,m}$ of $\mathcal{F}$ such that $m = N_{[\,]}(\varepsilon, \mathcal{F}, L_1(P))$. In particular, the union of the $[l_i, u_i]_\varepsilon$ contains $\mathcal{F}$ and $\|u_i - l_i\|_1 = \int |u_i - f_i| \, dP = \int (u_i - f_i) \, dP = P(u_i - f_i) \le \varepsilon$. Then for all $f \in \mathcal{F}$, there is a bracket $[l_i, u_i]$ such that $(\mathbb{P}_n - P)f \le (\mathbb{P}_n - P)u_i + P(u_i - f) \le (\mathbb{P}_n - P)u_i + \varepsilon$ and $(P - \mathbb{P}_n)f \le (P - \mathbb{P}_n)l_i + \varepsilon$. Thus $\sup_{f \in \mathcal{F}} |(\mathbb{P}_n - P)f| \le \varepsilon + \max_{1 \le i \le m} (\mathbb{P}_n - P)u_i \vee \max_{1 \le i \le m} (P - \mathbb{P}_n) l_i$. By the strong law of large numbers for real random variables applied $2m$ times, the right side converges to ε a.s. as $n \to \infty$. Thus $\limsup_n (\sup_{f \in \mathcal{F}} |(\mathbb{P}_n - P)f|) \le \varepsilon$ a.s.* for every $\varepsilon > 0$. By taking a sequence $\varepsilon_k \downarrow 0$, the lim sup is 0 a.s.*. This concludes the proof. $\square$

Reference. T.2.4.1. in Wellner&vdV WCEP p.122 or T.6.1. in Wellner's EP notes p.34 or L.3.1. in van de Geer AEP p.25.

Remark. There are measurability issues with sup of uncountable families, hence the need to consider outer probability and outer expectation, which were introduced by Hoffmann-Jørgensen and Dudley. A similar issue exists for stochastic processes in exotic spaces.

Lemma 2 (**Chebyshev's Inequality**)**.** *For any real random variable X and $t > 0$,*

$$P(|X| \ge t) \le \frac{EX^2}{t^2}.$$

Proof. We have $EX^2 \ge E(X^2 \mathbb{1}_{|X| \ge t}) \ge t^2 P(|X| \ge t)$. $\square$

Reference. T.8.3.1. in Dudley RAP p.261.

Lemma 3 (**Borel–Cantelli Lemma**). *Let $(X, \mathcal{A}, P)$ be a probability space and $(A_n)_{n\in\mathbb{N}}$ a sequence of measurable sets. Define the event $\limsup_n A_n := \bigcap_{m\geq 1}\bigcup_{n\geq m} A_n (=: A_n$ i.o.$)$. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(\limsup_n A_n) = 0$.*

Proof. For each $m \in \mathbb{N}$, we have $P(\limsup_n A_n) \leq P(\bigcup_{n\geq m} A_n) \leq \sum_{n\geq m} P(A_n)$, where the second inequality follows from the union bound. Since $\sum_{n=1}^{\infty} P(A_n) < \infty$, we have $\sum_{n\geq m} P(A_n) \to 0$ as $m \to \infty$. $\square$

Reference. T.8.3.4. in Dudley RAP p.262.

Lemma 4. *For any nonnegative random variable Y, we have $EY \leq \sum_{n=0}^{\infty} P(Y > n) \leq EY + 1$. Thus $EY < \infty$ if and only if $\sum_{n=0}^{\infty} P(Y > n) < \infty$.*

Proof. Define $A_k := \{k < Y \leq k+1\}$ for $k = 0, 1, \ldots$. Then $\sum_{n=0}^{\infty} P(Y > n) = \sum_{n=0}^{\infty}\sum_{k\geq n} P(A_k) = \sum_{k=0}^{\infty}(k+1)P(A_k)$, by rearranging sums of positive terms. Define $U := \sum_{k=0}^{\infty} k\mathbb{1}_{A_k}$. Then $U \leq Y \leq U + 1$, so $EU \leq EY \leq EU + 1 \leq EY + 1$, from which the result follows. $\square$

Reference. L.8.3.6. in Dudley RAP p.263.

Lemma 5 ((Kolmogorov–Etemadi) **Strong Law of Large Numbers**). *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of identically distributed pairwise independent real-valued random variables such that $\mathbb{E}\,|X_1| < +\infty$. Let $S_n = \sum_{i=1}^{n} X_n$ and $\mathbb{E}\,X_1 = m$. Then*

$$\frac{S_n}{n} \xrightarrow{a.s.} m.$$

Proof. Since pairwise independence is preserved under a Borel transformation and $X_n = X_n^+ - X_n^-$ where $X_n^+ := \max(X_n, 0)$ and $X_n^- := -\min(X_n, 0)$, we can prove the results for $X_n \geq 0$ without loss of generality.

Define $Y_n := X_n\mathbb{1}_{\{X_n\leq n\}}$ and $T_n := \sum_{i=1}^{n} Y_i$. The Y_n form a sequence of pairwise independent random variables with moments of all orders (note that the Y_n are not identically distributed). Given $\alpha > 1$, define the fast time scale $k(n) := [\alpha^n]$ where $[x]$ denotes the greatest integer less than or equal to x.

Let $\varepsilon > 0$. By Chebyshev's inequality and pairwise independence, we have $P(|T_{k(n)} - ET_{k(n)}| > \varepsilon k(n)) \leq \mathrm{Var}(T_{k(n)})/k(n)^2\varepsilon^2 = \sum_{i=1}^{k(n)} \mathrm{Var}(Y_i)/k(n)^2\varepsilon^2 \leq \sum_{i=1}^{k(n)} EY_i^2/k(n)^2\varepsilon^2 = \sum_{i=1}^{k(n)} E(X_i^2\mathbb{1}_{\{X_i\leq i\}})/k(n)^2\varepsilon^2 = \sum_{i=1}^{k(n)} E(X_1^2\mathbb{1}_{\{X_1\leq i\}})/k(n)^2\varepsilon^2 \leq \sum_{i=1}^{k(n)} E(X_1^2\mathbb{1}_{\{X_1\leq k(n)\}})/k(n)^2\varepsilon^2 = E(X_1^2\mathbb{1}_{\{X_1\leq k(n)\}})/k(n)\varepsilon^2$. Thus $\sum := \sum_{n=1}^{\infty} P(|T_{k(n)} - ET_{k(n)}| > \varepsilon k(n)) \leq \sum_{n=1}^{\infty} E(X_1^2\mathbb{1}_{\{X_1\leq k(n)\}})/k(n)\varepsilon^2 = \varepsilon^{-2}\mathbb{E}\,(X_1^2\sum_{n=1}^{\infty}\mathbb{1}_{\{X_1\leq k(n)\}}/k(n))$ where the equality follows from Fubini–Tonelli.

Let $x > 0$ and define $N := \min\{n \geq 1 : k(n) \geq x\}$. Then $\alpha^N \geq x$, and since $y \leq 2[y]$ for any $y \geq 1$, we have $\sum_{n=1}^{\infty}\mathbb{1}_{\{x\leq k(n)\}}/k(n) = \sum_{k(n)\geq x} 1/k(n) \leq 2\sum_{n\geq N}\alpha^{-n} = c\alpha^{-N} \leq c/x$ where $c = 2\alpha/(\alpha - 1)$. Thus $\sum_{n=1}^{\infty}\mathbb{1}_{\{X_1\leq k(n)\}}/k(n) \leq a/X_1$ for $X_1 > 0$. Therefore $\sum \leq a\varepsilon^{-2}EX_1 < \infty$. By the Borel–Cantelli lemma, we have $|T_{k(n)} - ET_{k(n)}|/k(n) \to 0$ a.s.. By the dominated convergence theorem, $EY_i = \mathbb{E}\,(X_i\mathbb{1}_{\{X_1\leq i\}}) \to EX_1$ as $i \to \infty$. It follows that $ET_{k(n)}/k(n) \to EX_1$ as $n \to \infty$. Thus $T_{k(n)}/k(n) \to EX_1$ a.s. as $n \to \infty$.

Furthermore, since $EX_n < \infty$ by hypothesis, $\sum_n P(X_n \neq Y_n) = \sum_n P(X_n > n) < \infty$ by Lemma 4. By the Borel–Cantelli lemma, $X_n - Y_n \to 0$ a.s. as $n \to \infty$, hence $(S_n - T_n)/n \to 0$ a.s. as $n \to \infty$. By the previous result, $S_{k(n)}/k(n) \to EX_1$ a.s. as

$n \to \infty$. If $k(n) \le m \le k(n+1)$, then $\frac{k(n)}{k(n+1)} \frac{S_{k(n)}}{k(n)} \le \frac{S_m}{m} \le \frac{k(n+1)}{k(n)} \frac{S_{k(n+1)}}{k(n+1)}$ since $X_n \ge 0$. But $k(n+1)/k(n) \to \alpha$ as $n \to \infty$, hence $\frac{1}{\alpha} EX_1 \le \liminf_{m\to\infty} \frac{S_m}{m} \le \limsup_{m\to\infty} \frac{S_m}{m} \le \alpha EX_1$ a.s.. Letting $\alpha \to 1$ via rational $\alpha > 1$, we get $\lim_{m\to\infty} S_m/m = EX_1$ a.s.. This concludes the proof. $\square$

Reference. The proof is from Etemadi (1981). See T.5.4. in Bhattacharya&Waymire p.89 or T.2.4.1. in Durrett PTE p.76 or T.8.3.5. in Dudley RAP p.263 (which is stated for independent RVs but without change for the proof). For another proof based on the convergence of random series (using Kolmogorov's maximal inequality), see Kallenberg FMP (T.5.23. p.113) or Dembo's notes (S.2.3.2. p.91).

References

Bhattacharya, R., and E. C. Waymire (2016): *A basic course in probability theory*. Springer.

Dembo, A. (2019): "Probability theory," *Lecture notes for STAT310/MATH230 at Stanford*.

Dudley, R. M. (2004): *Real analysis and probability*. Cambridge University Press.

Durrett, R. (2019): *Probability: theory and examples*, vol. 49. Cambridge University Press.

Etemadi, N. (1981): "An elementary proof of the strong law of large numbers," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 55(1), 119–122.

Kallenberg, O. (2021): *Foundations of modern probability*. Springer.

van der Geer, S. A. (2000): *Applications of empirical process theory*. Cambridge University Press.

van der Vaart, A., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*. Springer.

Wellner, J. A. (2005): "Empirical processes," *Lecture notes*.