Why is the law of the iterated logarithm in Banach space at the origin of the development of new concentration inequalities for independent random variables

The classical law of the iterated logarithm (LIL in the following) is, with the law of large numbers and the central limit theorem, one of the main asymptotic statements in probability theory. The LIL is however almost no longer being taught in standard probability courses today (2020).

Going back to works by A. Khinchin [1] and A. Kolmogorov [2] in the 1920s (before actually the foundations of probability theory were laid by the latter!), the tools behind the LIL, namely exponential inequalities for sums of independent random variables, are perhaps more important than the asymptotic result itself. This observation has to be put in perspective with the nowadays strong interest into concentration inequalities, classical exponential inequalities for sums of independent random variables being a specific instance. In this regard, it might be worthwhile emphasizing why and how the study, in the decade 1975–1985, of the LIL for Banach space valued random variables motivated the elaboration and discovery, by M. Talagrand, of new concentration principles for product measures and independent random variables, major tools of modern probability theory.

The discussion of the classical limit theorems in probability theory is restricted here to the simple instance of independent and identically distributed summands. Let X be a random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and denote by X_n, $n \geq 1$, a sequence of independent copies of X. Set $S_n = X_1 + \cdots + X_n$, $n \geq 1$.

Kolmogorov's law of large numbers expresses that the sequence $\frac{S_n}{n}$, $n \geq 1$, converges almost surely to $\mathbb{E}(X)$ if and only if $\mathbb{E}(|X|) < \infty$ (the converse has to be understood as: if the sequence $\frac{S_n}{n}$, $n \geq 1$, is almost surely bounded, then $\mathbb{E}(|X|) < \infty$).

The classical central limit theorem states that if $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) < \infty$, then the sequence $\frac{S_n}{\sqrt{n}}$, $n \geq 1$, converges weakly to a centered Gaussian random variable G with variance $\sigma^2 = \mathbb{E}(X^2)$. The converse, somewhat less popular, is that whenever the sequence $\frac{S_n}{\sqrt{n}}$, $n \geq 1$, is uniformly tight, that is, for every $\varepsilon > 0$, there exists $M > 0$ such that

$$\sup_{n\geq 1} \mathbb{P}\bigg(\bigg|\frac{S_n}{\sqrt{n}}\bigg| > M\bigg) \leq \varepsilon,$$

then $\mathbb{E}(X^2) < \infty$, and $\mathbb{E}(X) = 0$ (the standard argument makes use of the Fourier transform, and one $0 < \varepsilon < 1$ is actually enough).

The law of the iterated logarithm (LIL) is the statement, due in this form to P. Hartman and A. Wintner [3], strongly relying on [2], that, whenever $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = \sigma^2 < \infty$, then

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = \sigma$$

almost surely. Hence, under a second moment, the sequence $\sqrt{2n\log\log n}$, $n \geq 3$, is the smallest one stabilizing almost surely the sums S_n. There is a more precise result, due to V. Strassen [4], emphasizing the interval $[-\sigma, +\sigma]$ as the almost sure limiting points and cluster set of the sequence $\frac{S_n}{\sqrt{2n\log\log n}}$, $n \geq 3$. As for the central limit theorem, there is a converse implication expressing that if the sequence $\frac{S_n}{\sqrt{2n\log\log n}}$, $n \geq 3$, is almost surely bounded, then $\mathbb{E}(X^2) < \infty$, and $\mathbb{E}(X) = 0$. As aforementioned, exponential inequalities for sums of independent random variables are at the root of this asymptotic result. A classical example of such exponential bounds is the Bennett inequality [5] which indicates that for real random variables $Y_1, \ldots, Y_n$ with a finite second moment such that $Y_i \leq 1$, $i = 1, \ldots, n$, almost surely, letting $S = Y_1 + \cdots + Y_n$, for any $t \geq 0$,

$$\mathbb{P}\Big(S \geq \mathbb{E}(S) + t\Big) \leq \exp\bigg(-(t+\sigma^2)\log\Big(1+\frac{t}{\sigma^2}\Big) - t\bigg)$$

where $\sigma^2 = \sum_{i=1}^n \mathbb{E}(Y_i^2)$. This evolved form suitably captures the respective Gaussian, when $t << \sigma^2$, and Poisson, when $t >> \sigma^2$, tails of sums of independent random variables. Only Gaussian tails are necessary towards the LIL.

The notation X, X_n, S_n are similar for vector valued random variables. In the following, it will be implicitly assumed, in the context of the central limit theorem and the LIL, that the given random variable X is centered, avoiding elementary additional statements in this regard. In particular, for real valued random variables, the existence of the second moment $\mathbb{E}(X^2) < \infty$ is necessary and sufficient for the central limit theorem and the LIL to hold.

If X is a random vector in $\mathbb{R}^d$, the preceding characterizations hold similarly, the conditions $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(X^2) < \infty$ being replaced by $\mathbb{E}(\|X\|) < \infty$ and $\mathbb{E}(\|X\|^2) < \infty$ for an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$. The limiting (centered) Gaussian vector in the central limit theorem has the covariance structure of X, whereas the cluster set of the sequence $\frac{S_n}{\sqrt{2n\log\log n}}$, $n \geq 3$, in the LIL is given by the unit ball of the Euclidean structure induced by the covariance of X.

What is now the picture when X takes its values in a (real separable) infinite dimensional Banach space B with norm $\|\cdot\|$?

The law of large numbers holds similarly if and only if $\mathbb{E}(\|X\|) < \infty$ [6] (cf. [8, Chapter 7]).

For the central limit theorem, the picture is radically different, and, in a general Banach space, the moment condition $\mathbb{E}(\|X\|^2) < \infty$ (actually even almost surely boundedness) is not sufficient anymore to ensure weak convergence of the sequence $\frac{S_n}{\sqrt{n}}$, $n \geq 1$, and is neither necessary. If however B is a Hilbert space, it has been shown by S. Varadhan [7] that the moment condition $\mathbb{E}(\|X\|^2) < \infty$ is necessary and sufficient for the central limit theorem, but Hilbert spaces are essentially the only infinite dimensional spaces with this property (see [8, Chapter 10] for an extensive discussion).

To discuss the LIL, let us agree that X satisfies (a bounded form of) the LIL if

$$\limsup_{n\to\infty} \frac{\|S_n\|}{\sqrt{2n \log\log n}} < \infty$$

almost surely (there is a compact form including the description of the limiting set but it will not be necessary to go into this more precise statement, cf. [8, Chapter 8] for details). As for the central limit theorem, the moment condition $\mathbb{E}(\|X\|^2) < \infty$ is no more sufficient for the (bounded) LIL to hold true in general, and neither necessary. But one of the first striking statements, obtained by V. Goodman, J. Kuebs and J. Zinn in [9], is that in a (infinite dimensional) Hilbert space H, the necessary and sufficient conditions for the LIL to be satisfied are that

$$\mathbb{E}\bigg(\frac{\|X\|^2}{\log\log(\|X\|+3)}\bigg) < \infty,$$

a level of integrability therefore just below $\mathbb{E}(\|X\|^2) < \infty$, equivalent actually to the almost sure boundedness of the sequence $\frac{X_n}{\sqrt{2n\log\log n}}$, $n \geq 3$, *and*

$$\mathbb{E}\big(\langle y, X\rangle^2\big) < \infty \quad \text{for all } y \in H,$$

that is, the one-dimensional LIL holds true for the real random variables $\langle y, X\rangle$, $y \in H$. *The classical (second) moment condition therefore splits into a strong one concerning the norm and a weak one along linear functionals.* The characterization extends to the class of type 2 Banach spaces, covering for example L^p, $p \geq 2$, spaces. Type 2 Banach spaces are actually those for which the second moment condition $\mathbb{E}(\|X\|^2) < \infty$ is sufficient for the central limit theorem to be satisfied (cf. [8, Chapter 10]).

While moment conditions are not sufficient in general to ensure the validity of the LIL, on the basis of the Hilbert space example, there is actually a full characterization in an arbitrary Banach space $(B, \|\cdot\|)$ which reduces the almost sure boundedness statement of the LIL to a boundedness in probability (somehow similar to the central limit theorem) [10]. Namely, the (bounded) LIL holds true for a random variable X with values in a Banach space $(B, \|\cdot\|)$ with dual space B^* if and only if

$$\mathbb{E}\bigg(\frac{\|X\|^2}{\log\log(\|X\|+3)}\bigg) < \infty,$$

$$\mathbb{E}\big(\langle y, X\rangle^2\big) < \infty \quad \text{for all } y \in B^*$$

and

$$\frac{S_n}{\sqrt{2n\log\log n}},\ n \geq 3, \quad \text{is stochastically bounded,}$$

meaning that for each $\varepsilon > 0$, there is $M > 0$ such that

$$\sup_{n \geq 3} \mathbb{P}\bigg(\bigg\| \frac{S_n}{\sqrt{2n \log\log n}} \bigg\| > M \bigg) \leq \varepsilon$$

(equivalently $\sup_{n\geq 3} \mathbb{E}\big(\frac{\|S_n\|}{\sqrt{2n\log\log n}}\big) < \infty$, see [10], [8, Chapter 8]).

It is not difficult to see that in a Hilbert space (or more generally a type 2 Banach space), the stochastic boundedness of the sequence $\frac{S_n}{\sqrt{2n\log\log n}}$, $n \geq 3$, is a consequence of the strong moment condition $\mathbb{E}\big(\frac{\|X\|^2}{\log\log(\|X\|+3)}\big) < \infty$. It is furthermore an interesting corollary that if X satisfies the central limit theorem, then the LIL holds true if and only if $\mathbb{E}\big(\frac{\|X\|^2}{\log\log(\|X\|+3)}\big) < \infty$ (see [8, Chapter 8]).

The fundamental relevance of weak moments in the description of the LIL in infinite dimensional Hilbert and Banach spaces reflects the basic concentration property of Gaussian distributions, expressed in the tail bound

$$\mathbb{P}\Big(\big| \|G\| - \mathbb{E}(\|G\|) \big| \geq t \Big) \leq 2\, e^{-t^2/2\sigma^2}, \quad t \geq 0,$$

for a (centered) Gaussian vector G with values in $(B, \|\cdot\|)$, where $\sigma^2 = \sup_{\|y\|\leq 1} \mathbb{E}(\langle y, G\rangle^2)$, y ranging over the unit ball of the dual space B^* of B [11, 12, 13] (see [8, Chapter 3]). The essential feature of such tail inequalities lies in the fact that the strong moment $\mathbb{E}(\|G\|)$ occurs as an additive parameter while the weak moment σ^2 regulates the exponential decay. The proof of the characterization of the LIL in Banach space in [10] is actually based on a Gaussian randomisation procedure making use of the preceding concentration bound (the boundedness in probability of $\frac{S_n}{\sqrt{2n\log\log n}}$, $n \geq 3$, reflecting the deviation by $\mathbb{E}(\|G\|)$ in the Gaussian inequality, while the action of the weak moments corresponds to σ^2).

These results and phenomena prompted the study, and discovery, by M. Talagrand, of more general concentration inequalities for independent random variables and vectors extending the Gaussian example, and emphasizing the respective roles of strong means and weak variances.

A first major contribution in this regard [14] is the analogue of Gaussian concentration for series $Z = \sum_i \varepsilon_i x_i$, with ε_i independent ± 1 symmetric Bernoulli random variables, x_i elements of a Banach space B, in the form of

$$\mathbb{P}\Big(\big| \|Z\| - \mathbb{E}(\|Z\|) \big| \geq t \Big) \leq 4\, e^{-t^2/16\sigma^2}, \quad t \geq 0,$$

with the same meaning for $\sigma^2 = \sup_{\|y\|\leq 1} \mathbb{E}(\langle y, Z\rangle^2)$. The essential feature here is the independence in the bound on the number of variables. Note furthermore, for the matter of comparison, that any (centered) Gaussian vector may be represented in distribution as a series $\sum_i g_i x_i$, where the g_i's are independent standard normal random variables.

The concentration inequality for series $\sum_i \varepsilon_i x_i$ was extended in [15] to arbitrary independent uniformly bounded random variables ε_i, then elaborated in [16] as a general convexity principle based on induction on the dimension (number of coordinates), in a form which has taken the name of "Talagrand's convex distance inequality".

At the same time, M. Talagrand introduced another family of abstract inequalities for product measures, called "control by a finite number of points", relying on rearrangements methods and induction on the number of coordinates, with which he was able to tackle the issue of Poisson integrability of series of independent bounded vector valued random variables [17]. It is as a combination of this tool and of the concentration property of Bernoulli series that a direct, concise proof of the LIL is presented in [8, Chapter 8].

After these landmarks contributions, the Talagrand concentration inequalities for product measures and independent random variables have been deepened and synthesized in the celebrated publications [18] and [19], with a wide range of illustrations and applications. In a further refinement, involving mass transportation ideas, M. Talagrand established in [20] a sharp concentration inequality for the supremum of empirical processes. In the context of this note, the latter expresses that if $Y_1, \ldots, Y_n$ are independent random vectors in a Banach space $(B, \|\cdot\|)$ with $\|Y_i\| \leq 1$, $i = 1, \ldots, n$, almost surely, letting $S = Y_1 + \cdots + Y_n$, for any $t \geq 0$,

$$\mathbb{P}\Big(\big|\|S\| - \mathbb{E}(\|S\|)\big| \geq t\Big) \leq C \exp\bigg(-\frac{t}{C} \log\Big(1 + \frac{t}{\Sigma^2}\Big)\bigg)$$

where $\Sigma^2 = \mathbb{E}\big(\sup_{\|y\| \leq 1} \sum_{i=1}^n \langle y, Y_i \rangle^2\big)$ and $C > 0$ is a numerical constant. The comparison with the real Bennett inequality clearly puts forward the role of the deviation from the mean $\mathbb{E}(\|S\|)$ and the exponential rate quantified by weak moments (it may be shown when the Y_i's are centered, $\mathbb{E}(\Sigma^2) \leq \mathbb{E}(\|S\|) + 8 \sup_{\|y\| \leq 1} \sum_{i=1}^n \mathbb{E}(\langle y, Y_i \rangle^2)$ [8, Chapter 6]). This inequality may then also be used to recover the LIL in Banach space.

References

[1] A. Khinchin. Über einen Satz der Wahrscheinlichkeitsrechnung. Fund. Math. 6, 9–20 (1924).

[2] A. Kolmogorov. Über das Gesetz des iterierten Logarithmus. Math. Ann. 101, 126–135 (1929).

[3] P. Hartman, A. Wintner. On the law of the iterated logarithm. Amer. J. Math. 63, 169–176 (1941).

[4] V. Strassen. An invariance principle for the law of the iterated logarithm. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 3, 211–226 (1964).

[5] G. Bennett. Probability inequalities for the sum of independent random variables. J. Amer. Stat. Assoc. 57, 33–45 (1962).

[6] É. Mourier. Éléments aléatoires dans un espace de Banach. Ann. Inst. H. Poincaré 13, 161–244 (1953).

[7] S. R. S. Varadhan. Limit theorems for sums of independent random variables with values in a Hilbert space. Sankhya. A24, 213–238 (1962).

[8] M. Ledoux, M. Talagrand. Probability in Banach spaces (Isoperimetry and processes). Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer (1991).

[9] V. Goodman, J. Kuelbs, J. Zinn. Some results on the LIL in Banach space with applications to weighted empirical processes. Ann. Probab. 9, 713–752 (1981).

[10] M. Ledoux, M. Talagrand. Characterization of the law of the iterated logarithm in Banach spaces. Ann. Probab. 16,1242–1264 (1988).

[11] C Borell. The Brunn-Minskowski inequality in Gauss space. Invent. Math. 30, 207–216 (1975).

[12] V. N. Sudakov, B. S. Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. J. Soviet. Math. 9, 9–18 (1978); translated from Zap. Nauch. Sem. L.O.M.I. 41, 14–24 (1974).

[13] I. A. Ibragimov, V. N. Sudakov, B. S. Tsirel'son. Norms of Gaussian sample functions. Proceedings of the third Japan-USSR Symposium on Probability Theory. Lecture Notes in Math. 550, 20–41 (1976). Springer.

[14] M. Talagrand. An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities. Proc. Amer. Math. Soc. 104, 905–909 (1988).

[15] W. Johnson, G. Schechtman. Remarks on Talagrand's deviation inequality for Rademacher's functions. Functional Analysis – Longhorn Notes, Lecture Notes in Math. 1470, 72–77. Springer (1991).

[16] M. Talagrand. A new isoperimetric inequality for product measure and the tails of sums of independent random variables. Geom. Funct. Anal. 1, 211–223 (1991).

[17] M. Talagrand. Isoperimetry and integrability of the sum of independent Banach-space valued random variables. Ann. Probab. 17, 1546–1570 (1989).

[18] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. Inst. Hautes Études Sci. Publ. Math. 81, 73–205 (1995).

[19] M. Talagrand. A new look at independence. Ann. Probab. 24, 1–34 (1996).

[20] M. Talagrand. New concentration inequalities in product spaces. Invent. Math. 126, 505–563 (1996).