BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 53, Number 1, January 2016, Pages 173–177
http://dx.doi.org/10.1090/bull/1511
Article electronically published on August 26, 2015

Upper and lower bounds for stochastic processes, by M. Talagrand, Modern methods and classical problems, Ergebnisse der Mathematik und iher Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas, 3rd Series of Modern Surveys in Mathematics], Vol. 60, Springer, Heidelberg, 2014, xvi+626 pp., ISBN 978-3-642-54074-5/978-3-642-54075-2, US $149.00

I was at the beginning of my graduate studies at the Courant Institute when one of my professors, in his typical calm voice, advised in class: "Probability is all about inequalities. Knowing how to do upper bounds is essential, but the true art lies in handling the lower bounds." Somehow, I kept these words in the back of my mind. Now, they truly came alive while reading and reviewing this wonderful book by M. Talagrand. The quote certainly does not do full justice to my field, but it is too tempting not to recall.

To see how simple inequalities arising in probability may lead to ingenious results (and thus give some evidence of the above paragraph), let me try to motivate the reader with two classical, simple examples. The first one, almost 100 years old, comes from number theory. Let $\nu(n)$ denote the number of primes p dividing n without multiplicity (though counting multiplicity makes little difference). The following result roughly says that almost all integers n have very close to $\log\log n$ number of primes.

Theorem 1. *Let $f(n)$ be any function such that $\lim_n f(n) = \infty$. Then the number of integers x in $\{1, \dots, n\}$ such that*

$$|\nu(x) - \log\log n| > f(n)\sqrt{\log\log n}$$

is $o(n)$.

A quite complicated proof of the above theorem appeared in a 1920 paper of Hardy and Ramanujan [3]. But here is a sketch of a simple argument, given in [5] and beautifully presented in [6, Chapter 4], that establishes Theorem 1. It is a simple application of Chebychev's inequality, one of the simplest inequalities in probability theory:

$$\mathbb{P}(|X - \mathbb{E}X| > \delta) \leq \delta^{-2}\mathrm{Var}X.$$

The argument goes as follows. Let x be chosen uniformly at random from $\{1, \dots, n\}$. For p prime, we set

$$X_p = \begin{cases} 1 & \text{if } p \text{ divides } x, \\ 0 & \text{otherwise,} \end{cases}$$

and let $X = \sum_{p \leq n^{1/2}} X_p$. As no $x \leq n$ can have more than two prime factors larger than $n^{1/2}$, it suffices to study X to understand $\nu(x)$.

The average of X_p satisfies $\mathbb{E}X_p = \lfloor n/p \rfloor / n = 1/p + O(1/n)$. By linearity of expectation one gets

$$\mathbb{E}X = \mathbb{E}\sum_{p \leq n^{1/2}} X_p = \sum_{p \leq n^{1/2}} 1/p + O(1/n) = \log\log n + O(1), \tag{0.1}$$

2010 *Mathematics Subject Classification.* Primary 60-01, 60G05, 60G70.

where we used the fact that $\sum_{p\le x}(1/p) = \log\log x + O(1)$ (which comes from Abel summation and Stirling's formula). We now write

$$\mathrm{Var}X = \sum_{p\le n^{1/2}} \mathrm{Var}X_p + \sum_{p\ne q} \mathrm{Cov}[X_p, X_q].$$

The first sum can be handled as in (0.1): $\mathrm{Var}X_p = (1/p)(1-1/p) + O(1/n)$ and

$$\sum_{p\le n^{1/2}} \mathrm{Var}X_p = \log\log n + O(1). \tag{0.2}$$

It turns out that the sum of the covariances $\mathrm{Cov}[X_p, X_q] = \mathbb{E}X_pX_q - \mathbb{E}X_p\mathbb{E}X_q$ is negligible compared to (0.2). Indeed, for $p \ne q$, $X_pX_q = 1$ if and only if $(pq)|x$. Hence, $\mathbb{E}X_pX_q = \lfloor n/(pq)\rfloor/n$ and a direct calculation leads to

$$\mathrm{Cov}[X_p, X_q] \le \frac{1}{n}\left[\frac{1}{p} + \frac{1}{q}\right],$$

which implies

$$\sum_{p\ne q} \mathrm{Cov}[X_p, X_q] \le \frac{1}{n}\sum_{p\ne q}\left(\frac{1}{p} + \frac{1}{q}\right) \le 2n^{-1/2} \sum_{p\le n^{1/2}} \frac{1}{p} = o(1).$$

That is, $\mathrm{Var}X = \log\log n + O(1)$ and the Chebychev inequality gives

$$\mathbb{P}\Big(|X - \log\log n| > \lambda\sqrt{\log\log n}\Big) < \lambda^{-2} + o(1),$$

for any constant $\lambda > 0$. The same holds for ν, proving Theorem 1.

Another inequality that is quite useful in probability theory is the lower bound (0.3), known as "the second moment method", or the Payley–Zigmund inequality. For any nonnegative random variable X with $\mathbb{E}X^2 < \infty$ and any $\theta > 0$,

$$\mathbb{P}(X \ge \theta\mathbb{E}X) \ge \frac{(1-\theta)^2(\mathbb{E}X)^2}{\mathbb{E}X^2}. \tag{0.3}$$

(The reader will probably notice that (0.3) is essentially Chebychev's inequality disguised.)

In particular, if X takes only nonnegative integer values, we have the simple bound:

$$\mathbb{P}(X > 0) = \mathbb{P}(X \ge 1) \ge \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}. \tag{0.4}$$

Here is a nice use of (0.4) that I learned in Lawler and Limic's [4] gray book. It roughly says that if B is a Brownian motion in $\mathbb{R}^d$, then $d = 4$ is the critical dimension in which Brownian paths start to avoid each other. Precisely,

$$\mathbb{P}\big(B[0,1] \cap B[2,3] \ne \emptyset\big) \begin{cases} > 0, & d \le 3, \\ = 0, & d \ge 4. \end{cases} \tag{0.5}$$

We sketch the proof for a random walk S_n. Equation (0.5) then follows by taking the limit $n \to \infty$. The trick is to consider the following number of intersections of a random walk path S_n,

$$J_n = \sum_{j=0}^{n} \sum_{k=2n}^{3n} \mathbf{1}_{\{S_j = S_k\}}.$$

Note that J_n is integer valued and

$$\mathbb{P}(S[0,n] \cup S[2n,3n] \neq \emptyset) = \mathbb{P}(J_n \geq 1).$$

If we write $p(n) = \mathbb{P}(S_n = 0)$, then translation invariance implies

$$\mathbb{E}(J_n) = \sum_{j=0}^{n} \sum_{k=2n}^{3n} p(k-j) \sim \sum_{j=0}^{n} \sum_{k=2n}^{3n} \frac{1}{(k-j)^{d/2}} \sim \sum_{j=0}^{n} n^{1-(d/2)} \sim n^{2-(d/2)}, \tag{0.6}$$

where we used the fact that $p(n) \sim n^{-d/2}$. For $d \geq 5$, the "first moment bound"

$$\mathbb{P}(J_n > 0) \leq \mathbb{E}J_n$$

and (0.6) suffice to end the proof of (0.5). The cute part is when $d = 3, 4$. We now need to compute the second moment of J_n. This is done through

$$\mathbb{E}(J_n^2) = \sum_{0 \leq i,j \leq n} \sum_{2n \leq k,\ell \leq 3n} \mathbb{P}(S_i = S_k, S_j = S_\ell)$$

and, for our choices of indices,

$$\begin{aligned} \mathbb{P}(S_i = S_k, S_j = S_\ell) &\leq \left(\max_{m \geq n, x \in \mathbb{Z}^d} \mathbb{P}(S_m = x) \right) \left(\max_{x \in \mathbb{Z}^d} \mathbb{P}(S_{|\ell - k|} = x) \right) \\ &\leq \frac{c}{n^{d/2}(|\ell - k| + 1)^{d/2}}, \end{aligned}$$

where the last inequality follows from the local central limit theorem. Combining the last two displays, one arrives at

$$\mathbb{E}(J_n^2) \leq \begin{cases} cn, & d = 3, \\ c \log n, & d = 4, \\ cn^{(4-d)/2}, & d \geq 5, \end{cases} \tag{0.7}$$

which, combined with (0.6) and (0.4), ends the proof when $d = 3$. The critical case $d = 4$ needs an extra step, which can be found in [4, Section 10.1].

The examples and methods above belong to the "classical theory of stochastic processes". These methods have been widely used and generalized in several directions. They belong to the toolbox of almost every probabilist. Due to their importance, they are included in any classical first-year graduate course in probability. They appear in several textbooks in probability, combinatorics, statistics.... But not in Talagrand's book.

Although he writes a book about inequalities of stochastic processes, Talagrand focuses on modern abstract methods, completely abdicating the "classical approach". He describes problems on which the above strategies would not work and develops an abstract methodology to deal with some of these situations. The methods described in his book, many of them developed by its author, are much inspired by the idea of "chaining" that goes back to Kolmogorov. To try to explain this idea in the current review (it is very well explained in the book by the way!), let me illustrate one fundamental example that I believe is the starting and one of the selling points of Talagrand's textbook.

Consider a subset T of $\ell^2(\mathbb{N})$, and i.i.d. Bernoulli random variables $(\epsilon_i)_{i \geq 1}$. For $t \in T$, set $X_t = \sum_{i \geq 1} t_i \epsilon_i$, and let

$$b(T) := \mathbb{E} \sup_{t \in T} X_t.$$

The process $(X_t)_{t\in T}$ is called a Bernoulli process, and one wishes to understand the value of $b(T)$ from the geometry of the set T (as a subset of $\ell^2(\mathbb{N})$). The bound

$$b(T) \le \sup_{t\in T} \|t\|_1 \tag{0.8}$$

holds trivially, and it is also possible to show (see Chapter 5) that

$$b(T) \le \sqrt{\frac{\pi}{2}}\, g(T), \tag{0.9}$$

where $g(T) = \mathbb{E}\sup_{t\in T}\sum_{i\ge 1} t_i g_i$, with g_i independent, standard Gaussians. The ℓ^1 bound (0.8) and the Gaussian bound (0.9) have very different flavors. For instance, if we take $T = \{0, a\}$ for some $a \notin \ell^1$, the former is meaningless while the second is of constant order. In the case $T = \ell^1$, we have the opposite: the Gaussian bound is infinite while (0.8) is equal to 1.

Strikingly, these are the only two useful bounds for any Bernoulli process! In Talagrand's book, we learn that any other upper bound must come essentially from a mixture of (0.8) and (0.9). To be more precise, in 2013 Bednorz and Latala [1,2] proved that there exist universal constants c and C such that for any subset T of ℓ^2,

$$cb^*(T) \le b(T) \le Cb^*(T),$$

where

$$b^*(T) := \inf\left\{ g(T_1) + \sup_{t\in T_2}\|t\|_1 \ ;\ T \subseteq T_1 + T_2 \right\}.$$

This theorem, previously known as the Bernoulli conjecture (see Theorem 5.1.5), has several implications in the convergence of random Fourier series and lies at the core of the theory of suprema of random processes. One of the highlights of this textbook is the presentation of its proof and several generalizations, as well as open questions.

I hope the examples above give to the reader a taste of what this book is about (and not about). But they barely touch the full scope of the monograph. Talagrand goes to infinity and beyond and shows, for instance:

(1) how the chaining method allows one to derive sufficient and necessary conditions for convergence of Fourier series (Chapters 3 and 7);
(2) how to proceed when dealing with α-stable processes (Chapter 8) and infinitely divisible processes (Chapter 11); and
(3) a characterization of sequences (a_m) such that for each orthonormal sequence (ϕ_m) the series $\sum_{m\ge 1} a_m\phi_m$ converges almost surely.

In the last chapters, he spends a fair amount of time describing an ambitious and long range open research program to determine the limits and boundaries of the chaining method.

Talagrand's goal in this book is, without any doubt, very ambitious. It is not an introductory volume, but it contains marvelous ideas that should very likely be in the toolbox of anyone dealing with stochastic processes. It was my companion during many long (and happy) days of last summer. And I still feel I barely scratched its surface.

Acknowledgments

I thank Michael Damron for all his valuable input in a previous version of this review.

References

[1] Witold Bednorz and Rafał Latała, *On the suprema of Bernoulli processes* (English, with English and French summaries), C. R. Math. Acad. Sci. Paris **351** (2013), no. 3-4, 131–134, DOI 10.1016/j.crma.2013.02.013. MR3038002

[2] Witold Bednorz and Rafał Latała, *On the boundedness of Bernoulli processes*, Ann. of Math. (2) **180** (2014), no. 3, 1167–1203, DOI 10.4007/annals.2014.180.3.8. MR3245015

[3] Hardy, G. H. *Some Famous Problems of the Theory of Numbers.* Clarendon Press, 1920.

[4] Gregory F. Lawler and Vlada Limic, *Random walk: a modern introduction*, Cambridge Studies in Advanced Mathematics, vol. 123, Cambridge University Press, Cambridge, 2010. MR2677157 (2012a:60132)

[5] Paul Turán, *On a Theorem of Hardy and Ramanujan*, J. London Math. Soc. **S1-9**, no. 4, 274, DOI 10.1112/jlms/s1-9.4.274. MR1574877

[6] Noga Alon and Joel H. Spencer, *The probabilistic method*, 3rd ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2008. MR2437651 (2009j:60004)

Antonio Auffinger
Mathematics Department
Northwestern University
2033 Sheridan Road
Evanston, Illinois 60208
E-mail address: auffing@math.northwestern.edu