Rosenthal's inequality

Theorem 1 (Rosenthal's inequality)**.** *Let* $X_1, \ldots, X_n$ *be independent real valued centered random variables on some probability space* $(\Omega, \mathcal{A}, \mathbb{P})$ *with* $\mathbb{E}(|X_i|^p) < \infty$, $i = 1, \ldots, n$, *for some* $p > 2$. *Then*

$$\mathbb{E}\bigg(\bigg| \sum_{i=1}^{n} X_i \bigg|^p \bigg) \leq C_p \bigg(\sum_{i=1}^{n} \mathbb{E}\big(|X_i|^p\big) + \bigg(\sum_{i=1}^{n} \mathbb{E}(X_i^2) \bigg)^{p/2} \bigg) \tag{1}$$

where $C_p > 0$ *only depends on* p.

The inequality of the above theorem was emphasized by H. P. Rosenthal in his 1970 article [27]. It is surprising that this important and useful probabilistic inequality actually appeared first in a functional analytic context, as a tool to study subspaces of L^p-spaces, and historically rather late in the studies on sums of independent random variables throughout the 20th century.

The note is a brief exposition of the proof of this inequality, and discussion of some related results. In particular, the exposition emphasizes the comparison with a stronger inequality due to J. Hoffmann-Jørgensen for Banach space valued random variables. A short bibliography of developments of Rosenthal's inequality in wider contexts and settings is addressed in the last section.

1 Proof and discussion

The proof of Rosenthal's inequality presented in this section is the one developed in the original contribution [27].

Proof. If X is a centered random variable with a finite q-th moment, $q \geq 1$, and if X' is an independent copy of X, then

$$\mathbb{E}\big(|X|^q\big) \leq \mathbb{E}\big(|X - X'|^q\big) \leq 2^{q-1}\, \mathbb{E}\big(|X|^q\big). \tag{2}$$

Since $X - X'$ has a symmetric law, it is enough to prove Rosenthal's inequality (1) for (independent) symmetric random variables $X_1, \ldots, X_n$.

The sample $(X_1, \ldots, X_n)$ of independent symmetric random variables has the same distribution as $(\varepsilon_1 X_1, \ldots, \varepsilon_n X_n)$ where $\varepsilon_1, \ldots, \varepsilon_n$ are independent Bernoulli random variables with $\mathbb{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$, $i = 1, \ldots, n$, independent from the X_i's. By Khintchine's inequality conditionally on $(X_1, \ldots, X_n)$,

$$\mathbb{E}\left(\left| \sum_{i=1}^{n} X_i \right|^p \right) = \mathbb{E}\left(\left| \sum_{i=1}^{n} \varepsilon_i X_i \right|^p \right) \leq K_p \, \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i^2 \right)^{p/2} \right) \tag{3}$$

for some constant $K_p > 0$.

Set $Z_i = X_i^2$, $i = 1, \ldots, n$, $q = \frac{p}{2} > 1$, and, for $0 < r \leq q$,

$$A_r = \left(\sum_{i=1}^{n} \mathbb{E}(Z_i^r) \right)^{1/r}, \qquad B_r = \left(\mathbb{E}\left(\left(\sum_{i=1}^{n} Z_i \right)^r \right) \right)^{1/r}.$$

A basic step in the argument is as follows:

$$\begin{aligned} \mathbb{E}\left(\left(\sum_{i=1}^{n} Z_i \right)^{q-1} Z_1 \right) &\leq 2^{q-1} \, \mathbb{E}\left(\left(Z_1^{q-1} + \left(\sum_{i=2}^{n} Z_i \right)^{q-1} \right) Z_1 \right) \\ &= 2^{q-1} \, \mathbb{E}(Z_1^q) + 2^{q-1} \, \mathbb{E}\left(\left(\sum_{i=2}^{n} Z_i \right)^{q-1} \right) \mathbb{E}(Z_1) \end{aligned}$$

by independence between Z_1 and $\sum_{i=2}^{n} Z_i$. Since the Z_i's are non-negative, by definition of B_r,

$$\mathbb{E}\left(\left(\sum_{i=1}^{n} Z_i \right)^{q-1} Z_1 \right) \leq 2^{q-1} \, \mathbb{E}(Z_1^q) + 2^{q-1} B_{q-1}^{q-1} \, \mathbb{E}(Z_1).$$

The same inequality holds true with Z_1 replaced by any Z_j, $j = 1, \ldots, n$. Summing the resulting inequalities over j, it follows that

$$B_q^q \leq 2^{q-1} \left(A_q^q + B_{q-1}^{q-1} A_1 \right).$$

By Jensen's inequality $B_{q-1} \leq B_q$, so that

$$B_q^q \leq 2^{q-1} \left(A_q^q + B_q^{q-1} A_1 \right) \leq 2^q \max\left(A_q^q, B_q^{q-1} A_1 \right).$$

Re-interpretating this inequality case by case expresses that

$$B_q \leq 2^q \max\left(A_q, A_1 \right). \tag{4}$$

In terms of the variables $X_i^2 = Z_i$, $i = 1, \ldots, n$, and $p = 2q$, the latter inequality (4) yields

$$\mathbb{E}\bigg(\Big(\sum_{i=1}^n X_i^2 \Big)^{p/2} \bigg) \, \leq \, 2^{p^2/4} \max \bigg(\sum_{i=1}^n \mathbb{E}\big(|X_i|^p\big), \Big(\sum_{i=1}^n \mathbb{E}(X_i^2) \Big)^{p/2} \bigg).$$

Inserted into (3), the conclusion follows. □

When $1 \leq p \leq 2$, the inequality (sometimes called Marcinkiewicz-Zygmund inequality) takes the form

$$\mathbb{E}\bigg(\Big| \sum_{i=1}^n X_i \Big|^p \bigg) \, \leq \, C_p \sum_{i=1}^n \mathbb{E}\big(|X_i|^p\big)$$

and immediately follows from (3).

Rosenthal's inequality (1) may be reversed. Clearly, by Jensen's inequality,

$$\mathbb{E}\bigg(\Big| \sum_{i=1}^n X_i \Big|^p \bigg) \, \geq \, \bigg(\mathbb{E}\bigg(\Big| \sum_{i=1}^n X_i \Big|^2 \bigg) \bigg)^{p/2} \, = \, \bigg(\sum_{i=1}^n \mathbb{E}(X_i^2) \bigg)^{p/2}.$$

On the other hand, for symmetric random variables $X_1, \ldots, X_n$,

$$\mathbb{E}\bigg(\Big| \sum_{i=1}^n X_i \Big|^p \bigg) \, = \, \mathbb{E}\bigg(\Big| \sum_{i=1}^n \varepsilon_i X_i \Big|^p \bigg) \, \geq \, k_p \, \mathbb{E}\bigg(\Big(\sum_{i=1}^n |X_i|^2 \Big)^{p/2} \bigg) \, \geq \, k_p \, \mathbb{E}\bigg(\sum_{i=1}^n |X_i|^p \bigg) \tag{5}$$

by the lower-bound in Khintchine's inequality and $p > 2$.

The above proof of Rosenthal's inequality also entails an inequality for non-negative independent random variables $Z_1, \ldots, Z_n$ in the form

$$\mathbb{E}\bigg(\Big(\sum_{i=1}^n Z_i \Big)^q \bigg) \, \leq \, C'_q \bigg(\sum_{i=1}^n \mathbb{E}(Z_i^q) + \Big(\sum_{i=1}^n \mathbb{E}(Z_i) \Big)^q \bigg) \tag{6}$$

where $C'_q > 0$ only depends on $q > 1$.

Applied to variables of the form $X_i = \varepsilon_i x_i$, $x_i \in \mathbb{R}$, $i = 1, \ldots, n$, Rosenthal's inequality is a generalization of the Khintchine inequality (for $p > 2$).

The issue about the optimal growth in $p \to \infty$ of the constant C_p in Rosenthal's inequality (1) has been addressed and solved in [13] where it is shown $(C_p)^{1/p} \sim \frac{p}{\log p}$, in constrast with the $\sqrt{p}$-behavior of the constant $(K_p)^{1/p}$ in Khintchine's inequality (3). (Clearly, the above original proof only produces exponential dependence in p.)

2 Hoffmann-Jørgensen's inequality

Clearly, under the hypotheses of Theorem 1,

$$\sum_{i=1}^{n} \mathbb{E}(X_i^2) = \mathbb{E}\bigg(\Big|\sum_{i=1}^{n} X_i\Big|^2\bigg)$$

so that Rosenthal's inequality may be seen as a moment comparison of the sum $\sum_{i=1}^{n} X_i$ (in particular, the inequality is trivial by Jensen's inequality for $p \leq 2$). With this observation in mind, it is natural to try to compare the p-th moment with another q-th moment for $q < p$, in the same way the Kahane inequality [18]

$$\mathbb{E}\bigg(\Big\|\sum_{i=1}^{n} \varepsilon_i x_i\Big\|^p\bigg) \leq K_{p,q}\bigg(\mathbb{E}\bigg(\Big\|\sum_{i=1}^{n} \varepsilon_i x_i\Big\|^q\bigg)\bigg)^{p/q}, \tag{7}$$

for vectors $x_1, \ldots, x_n$ in a normed space, extends the Khintchine inequality.

In [11], J. Hoffmann-Jørgensen answers this question with radically new methods in a Banach space valued random variable context.

Theorem 2 (Hoffmann-Jørgensen's inequality)**.** *Let $X_1, \ldots, X_n$ be independent centered random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a (real separable) Banach space $(B, \|\cdot\|)$ such that $\mathbb{E}(\|X_i\|^p) < \infty$, $i = 1, \ldots, n$, for some $p > 0$. Then, for $0 < q < p$,*

$$\mathbb{E}\bigg(\Big\|\sum_{i=1}^{n} X_i\Big\|^p\bigg) \leq C_{p,q}\bigg(\mathbb{E}\big(\max_{1\leq i\leq n} \|X_i\|^p\big) + \bigg(\mathbb{E}\bigg(\Big\|\sum_{i=1}^{n} X_i\Big\|^q\bigg)\bigg)^{p/q}\bigg) \tag{8}$$

where $C_{p,q} > 0$ only depends on p, q.

Proof. As in the real case, it is enough to prove the inequality for independent symmetric random variables $X_1, \ldots, X_n$ (with an analogue of (2) for $q < 1$).

Hoffmann-Jørgensen's inequality is more precisely the following statement. Only the symmetric case is considered there but arbitrary independent random variables may be studied similarly (cf. [11]).

Proposition 3. *Let $X_1, \ldots, X_n$ be independent symmetric random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a (real separable) Banach space $(B, \|\cdot\|)$, and set $S_n = X_1 + \cdots + X_n$. For any $s, t > 0$,*

$$\mathbb{P}\big(\|S_n\| \geq s + 2t\big) \leq \mathbb{P}\big(\max_{1\leq i\leq n} \|X_i\| \geq s\big) + 4\,\big[\mathbb{P}\big(\|S_n\| \geq t\big)\big]^2. \tag{9}$$

Proof. Let $\tau = \min\{j = 1, \dots, n; \|S_j\| \geq t\}$, where $S_j = X_1 + \dots + X_j$, $j = 1, \dots, n$. The event $\{\tau = j\}$ only depends on the random variables $X_1, \dots, X_j$, and

$$\bigcup_{j=1}^{n} \{\tau = j\} \; = \; \Big\{ \max_{1 \leq j \leq n} \|S_j\| \geq t \Big\}$$

as a disjoint union. For every $j = 1, \dots, n$ (with the convention $S_0 = 0$),

$$\|S_n\| \; \leq \; \|S_{j-1}\| + \|X_j\| + \|S_n - S_j\|$$

so that

$$\begin{aligned} \mathbb{P}\big(\tau = j \, ; \|S_n\| \geq s + 2t\big) \; &\leq \; \mathbb{P}\big(\tau = j \, ; \max_{1 \leq i \leq n} \|X_i\| \geq s\big) \\ &\quad + \mathbb{P}(\tau = j) \, \mathbb{P}\big(\|S_n - S_j\| \geq t\big) \end{aligned}$$

since $\|S_{j-1}\| < t$ on $\{\tau = j\}$. Summing over $j = 1, \dots, n$,

$$\begin{aligned} \mathbb{P}\big(\|S_n\| \geq s + 2t\big) \; &= \; \mathbb{P}\big(\max_{1 \leq j \leq n} \|S_j\| \geq t \, ; \|S_n\| \geq s + 2t\big) \\ &= \; \sum_{j=1}^{n} \mathbb{P}\big(\tau = j \, ; \|S_n\| \geq s + 2t\big) \\ &\leq \; \mathbb{P}\big(\max_{1 \leq i \leq n} \|X_i\| \geq s\big) + \mathbb{P}\big(\max_{1 \leq j \leq n} \|S_j\| \geq t\big) \, \mathbb{P}\big(\max_{1 \leq j \leq n} \|S_n - S_j\| \geq t\big). \end{aligned}$$

It remains to recall the classical Lévy inequality for sums of symmetric samples $T_k = Y_1 + \dots + Y_k$, $k = 1, \dots, n$,

$$\mathbb{P}\big(\max_{1 \leq k \leq n} \|T_k\| \geq t\big) \; \leq \; 2 \, \mathbb{P}\big(\|T_n\| \geq t\big), \tag{10}$$

to conclude the proof.

For completeness, the proof of (10) relies on the same strategy. With τ defined in the same way as above,

$$\mathbb{P}\big(\|T_n\| \geq t\big) \; = \; \sum_{k=1}^{n} \mathbb{P}\big(\tau = k \, ; \|T_n\| \geq t\big). \tag{11}$$

By symmetry of the sample $(Y_1, \dots, Y_n)$ (for example $Y_1, \dots, Y_n$ independent and symmetric), for every $k = 1, \dots, n$, the sample $(Y_1, \dots, Y_k, -Y_{k+1}, \dots, -Y_n)$ has the same distribution as $(Y_1, \dots, Y_n)$, hence it also holds true that

$$\mathbb{P}\big(\|T_n\| \geq t\big) \; = \; \sum_{k=1}^{n} \mathbb{P}\big(\tau = k \, ; \|T_k - R_k\| \geq t\big) \tag{12}$$

where $R_k = T_n - T_k$, $k = 1, \dots, n$. By the triangle inequality,

$$2\|T_k\| \; \leq \; \|T_k + R_k\| + \|T_k - R_k\| \; = \; \|T_n\| + \|T_k - R_k\|$$

so that, on $\{\tau = k\}$, either $\|T_n\| \geq t$ or $\|T_k - R_k\| \geq t$. Adding therefore (11) and (12),

$$\mathbb{P}\big(\max_{1\leq k\leq n} \|T_k\| \geq t\big) \;=\; \sum_{k=1}^{n} \mathbb{P}\big(\tau = k\big) \;\leq\; 2\,\mathbb{P}\big(\|T_n\| \geq t\big).$$

□

On the basis of (9), Theorem 2 easily follows. Namely

$$\mathbb{E}\big(\|S_n\|^p\big) \;=\; 3^p \int_0^\infty \mathbb{P}\big(\|S_n\| \geq 3t\big) d(t^p).$$

Choose then

$$t_0 \;=\; \Big[2\cdot 3^p\,\mathbb{E}\big(\|S_n\|^q\big)\Big]^{1/q}$$

so that $\mathbb{P}(\|S_n\| \geq t) \leq \frac{1}{2\cdot 3^p}$ for every $t \geq t_0$. Then, by (9) with $s = t$,

$$\begin{aligned}
\mathbb{E}\big(\|S_n\|^p\big) &\leq\; (3t_0)^p + 3^p \int_{t_0}^\infty \mathbb{P}\big(\|S_n\| \geq 3t\big) d(t^p) \\
&\leq\; (3t_0)^p + 3^p \int_{t_0}^\infty \mathbb{P}\big(\max_{1\leq i\leq n} \|X_i\| \geq t\big) d(t^p) + 3^p \int_{t_0}^\infty \big[\mathbb{P}\big(\|S_n\| \geq t\big)\big]^2 d(t^p) \\
&\leq\; (3t_0)^p + 3^p\,\mathbb{E}\big(\max_{1\leq i\leq n} \|X_i\|^p\big) + \frac{1}{2}\int_{t_0}^\infty \mathbb{P}\big(\|S_n\| \geq t\big) d(t^p).
\end{aligned}$$

Hence

$$\mathbb{E}\big(\|S_n\|^p\big) \;\leq\; 2(3t_0)^p + 2\cdot 3^p\,\mathbb{E}\big(\max_{1\leq i\leq n} \|X_i\|^p\big)$$

which is the announced result by the choice of t_0. □

Hoffmann-Jørgensen's inequality (8) may also be reversed. The point is that the proof of Lévy's inequality (10) also gives rise to the inequality

$$\mathbb{P}\big(\max_{1\leq i\leq n} \|X_i\| \geq t\big) \;\leq\; 2\,\mathbb{P}\big(\|S_n\| \geq t\big), \quad t \geq 0.$$

The comparison between the real valued Rosenthal's inequality (1) and the vector valued Hoffmann-Jørgensen's inequality (8) requires some care, linking the conclusions to the theory of type and cotype of Banach spaces (cf. [20]).

Clearly

$$\mathbb{E}\big(\max_{1\leq i\leq n} \|X_i\|^p\big) \;\leq\; \sum_{i=1}^{n} \mathbb{E}\big(\|X_i\|^p\big).$$

When $p > q = 2$, in Hilbert spaces or more generally type 2 Banach spaces,

$$\mathbb{E}\bigg(\bigg\|\sum_{i=1}^n X_i\bigg\|^2\bigg) \leq C\sum_{i=1}^n \mathbb{E}\big(\|X_i\|^2\big)$$

which directly compares to, and therefore together with the preceding inequality improves upon, (1).

In the reverse form, it is not true in general that

$$\mathbb{E}\bigg(\bigg\|\sum_{i=1}^n X_i\bigg\|^p\bigg) \geq \frac{1}{C}\sum_{i=1}^n \mathbb{E}\big(\|X_i\|^p\big)$$

which only holds true in cotype p, $p \geq 2$, Banach spaces (in particular Hilbert spaces).

3 Legacy

The legacy of Rosenthal's inequality is wide and significant. It has been used as a most convenient and useful tool to bound sums of (independent) random variables in numerous contexts and applications, mostly of probabilistic nature.

At the same time, the form itself of Rosenthal's inequality has been extended and generalized to various contexts and settings. The following mentions, in an incomplete and partial selection of a few pointers to the literature, some of these generalizations.

Rosenthal's inequality has been a most valuable tool in its original functional analytic setting, with developments and applications to symmetric structures, cf. e.g. [12, 14, 1]...

Rosenthal's inequality was extended to discrete martingales as part of the Burkholder-Davis-Gundy inequalities [4]. Continuous martingales have been considered later in [8] e.g. Dependence conditions and stationarity have been examined too [9, 7, 22]..., as well as exchangeable structures [12], quadratic forms [5], negatively associated sequences [28], matrix ensembles [21] etc.

In parallel to the Hoffmann-Jørgensen inequality for Banach space valued random variables, extensions of Rosenthal's inequality for (infinite-dimensional) random vectors have been widely considered [1, 29, 25]... The study of Rosenthal's inequality in non commutative functional analytic and probabilistic frameworks has been in particular most fruitful [15, 16, 6, 24, 17]...

In its original formulation, Rosenthal's inequality was also improved in various ways [23, 26, 3]... In addition to [13], the study of the best constants in (1) gave rise to numerous developments, in the original setting, and in the context of the various generalizations (martingales, non commutative spaces etc.), see in particular [10, 19, 2, 24]...

References

[1] S. V. Astashkin. Martingale Rosenthal inequalities in symmetric spaces. *Mat. Sb.* 205, 41–62 (2014); translation in *Sb. Math.* 205, 1720–1740 (2014).

[2] S. V. Astashkin, F. Sukochev. Best constants in Rosenthal-type inequalities and the Kruglov operator. *Ann. Probab.* 38, 1986–2008 (2010).

[3] S. V. Astashkin, K. Tikhomirov. A probabilistic version of Rosenthal's inequality. *Proc. Amer. Math. Soc.* 141, 3539–3547 (2013).

[4] D. Burkholder. Distribution function inequalities for martingales. *Ann. Probability* 1, 19–42 (1973).

[5] V. de la Peña, R. Ibragimov, S. Sharakhmetov. On extremal distributions and sharp L_p-bounds for sums of multilinear forms. *Ann. Probab.* 31, 630–675 (2003).

[6] S. Dirksen, B. de Pagter, D. Potapov, F. Sukochev. Rosenthal inequalities in noncommutative symmetric spaces. *J. Funct. Anal.* 261, 2890–2925 (2011).

[7] P. Doukhan, M. Neumann. Probability and moment inequalities for sums of weakly dependent random variables, *Stochastic Process. Appl.* 117, 878–903 (2007).

[8] K. Dzhaparidze, E. Valkeila. On the Hellinger type distances for filtered experiments. *Probab. Theory Related Fields* 85, 105–117 (1990).

[9] I. Fazekas, A. Kukush, T. Tómács. On the Rosenthal inequality for mixing fields. *Ukraïn. Mat. Zh.* 52, 266–276 (2000); translation in *Ukrainian Math. J.* 52, 305–318 (2000).

[10] P. Hitczenko. Best constants in martingale version of Rosenthal's inequality. *Ann. Probab.* 18, 1656–1668 (1990).

[11] J. Hoffmann-Jørgensen. Sums of independent Banach space valued random variables. *Studia Math.* 52, 159–186 (1974).

[12] W. B. Johnson, B. Maurey, G. Schechtman, L. Tzafriri. *Symmetric structures in Banach spaces.* Mem. Amer. Math. Soc. 19 (1979).

[13] W. B. Johnson, G. Schechtman, J. Zinn. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Probab.* 13, 234–253 (1985).

[14] W. Johnson, G. Schechtman. Sums of independent random variables in rearrangement invariant function spaces. *Ann. Probab.* 17, 789–808 (1989).

[15] M. Junge, Q. Xu. Noncommutative Burkholder/Rosenthal inequalities. *Ann. Probab.* 31, 948–995 (2003).

[16] M. Junge, X. Parcet, Q. Xu. Rosenthal type inequalities for free chaos. *Ann. Probab.* 35, 1374–1437 (2007).

[17] M. Junge, Q. Zeng. Noncommutative Bennett and Rosenthal inequalities. *Ann. Probab.* 41, 4287–4316 (2013).

[18] J.-P. Kahane. *Some random series of functions.* Heath Math. Monographs, 1968. Cambridge Univ. Press, 1985, 2nd edn.

[19] R. Latała. Estimation of moments of sums of independent real random variables. *Ann. Probab.* 25, 1502–1513 (1997).

[20] M. Ledoux, M. Talagrand. *Probability in Banach spaces (Isoperimetry and processes).* Classics in Mathematics. Springer (2011).

[21] L. Mackey, M. Jordan, R. Y. Chen, B. Farrell, J. Tropp. Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.* 42, 906–945 (2014).

[22] F. Merlevède, M. Peligrad. Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples. *Ann. Probab.* 41, 914–960 (2013).

[23] S. V. Nagaev, I. Pinelis. Some inequalities for the distributions of sums of independent random variables. *Teor. Verojatnost. i Primenen.* 22, 254–263 (1977).

[24] A. Osękowski. A note on the Burkholder-Rosenthal inequality. *Bull. Pol. Acad. Sci. Math.* 60, 177–185 (2012).

[25] I. Pinelis. Rosenthal-type inequalities for martingales in 2-smooth Banach spaces. *Theory Probab. Appl.* 59, 699–706 (2015).

[26] I. Pinelis, S. Utev. Estimates of moments of sums of independent random variables. *Teor. Veroyatnost. i Primenen.* 29, 554–557 (1984).

[27] H. P. Rosenthal. On the subspaces of L^p $(p > 2)$ spanned by sequences of independent random variables. *Israel J. Math* 8, 273–303 (1970).

[28] Q.-M. Shao. A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.* 13, 343–356 (2000).

[29] L. Yu. Generalized Rosenthal's inequality for Banach-space-valued martingales. *Acta Math. Sci. Ser. B* 29, 305–312 (2009).