

Quelques mots sur l'inégalité de concentration de Azuma-Hoeffding

Djalil Chafaï



20 mars 2012*

Résumé

L'inégalité de concentration de Azuma-Hoeffding fait partie de ces beaux outils mathématiques à la fois simples et utiles. Elle mérite d'être connue par tous ceux qui s'intéressent aux probabilités. Dans cet exposé, on en donne une version due à McDiarmid, ainsi qu'une application aux mesures spectrales empiriques de matrices aléatoires, puis au problème du voyageur de commerce en optimisation combinatoire randomisée.

1 Inégalité de Azuma-Hoeffding

On dit qu'une variable est concentrée lorsqu'elle reste proche d'une quantité déterministe (sa moyenne ou sa médiane par exemple) avec grande probabilité. Une telle propriété peut être obtenue en contrôlant la queue de distribution de la variable, par exemple au moyen de moments, comme dans l'inégalité quadratique de Tchebychev basée sur la variance ou l'inégalité exponentielle de Chernoff basée sur la transformée de Laplace (moments exponentiels). L'inégalité de Azuma-Hoeffding¹ exploite une information sur l'oscillation de la variable (diamètre du support). Des inégalités de concentration plus sophistiquées mais moins belles sont également disponibles, comme celles de Bennett et de Bernstein, exploitant la variance ou plus généralement les moments. Il existe également une inégalité très belle due à Talagrand, à base de convexité, que Paul-Marie Samson peut vous expliquer. Ces dernières années ont vu le développement de tout un arsenal fonctionnel autour des inégalités de concentration, que cet exposé n'aborde pas (avis aux amateurs!).

Théorème 1.1 (Inégalité de concentration de Azuma-Hoeffding). *Soit $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$ une variable aléatoire intégrable. Soit $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n = \mathcal{F}$ une filtration d'interpolation quelconque, et d_1, \dots, d_n les accroissements de la martingale de Doob associée*

$$Y - \mathbb{E}(Y) = \sum_{k=1}^n \mathbb{E}(Y | \mathcal{F}_k) - \mathbb{E}(Y | \mathcal{F}_{k-1}) = \sum_{k=1}^n d_k.$$

*Notes d'exposé du mardi 20 mars 2012 au groupe de travail Analyse, Statistique, et Probabilités (ASPro) du Laboratoire LAMA UMR CNRS 8050 de l'Université Paris-Est Marne-la-Vallée.

1. Bernard Maurey, cofondateur du LAMA, a beaucoup œuvré à la popularisation de cette inégalité.

On a alors, pour tout $r \geq 0$,

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq r) \leq 2 \exp\left(-\frac{2r^2}{\text{osc}(d_1)^2 + \dots + \text{osc}(d_n)^2}\right).$$

Si par exemple $Y = n^{-1/2}(X_1 + \dots + X_n)$ avec $(X_k)_{k \geq 1}$ i.i.d. centrées à valeurs dans $[-1, 1]$ alors en prenant $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$ on a $d_k = n^{-1/2}X_k$. Comme $\text{osc}(d_k) = 2n^{-1/2}$, on obtient la borne gaussienne en $2 \exp(-\frac{1}{2}r^2)$ suggérée par le théorème central limite pour la variance la plus grande possible² (soit 1, atteinte si les X_k valent ± 1 avec probabilité $1/2$).

On a toujours $\text{osc}(d_k) = \sup d_k - \inf d_k = \text{diamètre}(\text{support}(d_k)) \leq 2\|d_k\|_\infty$ mais cette dernière majoration peut causer une perte. Dans le cas où $Y = F(X_1, \dots, X_n)$ avec $(X_k)_{k \geq 1}$ i.i.d. alors on prend $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$. De plus, si X'_1, \dots, X'_n est une copie indépendante de X_1, \dots, X_n alors $\mathbb{E}(Y | \mathcal{F}_{k-1}) = \mathbb{E}(F(X_1, \dots, X_k, \dots, X_n) | \mathcal{F}_k)$ ce qui donne

$$d_k = \mathbb{E}(Y | \mathcal{F}_k) - \mathbb{E}(Y | \mathcal{F}_{k-1}) = \mathbb{E}(F(X_1, \dots, X_k, \dots, X_n) - F(X_1, \dots, X'_k, \dots, X_n) | \mathcal{F}_k).$$

Cette identité ramène le problème au contrôle de l'oscillation partielle conditionnelle de F .

Démonstration. Soit U une v.a.r. telle que $\mathbb{E}(U) = 0$ et $a \leq U \leq b$. La convexité de $u \mapsto e^u$ donne, pour tout $t \geq 0$ et tout $a \leq x \leq b$,

$$e^{tx} \leq \frac{x-a}{b-a}e^{tb} + \frac{b-x}{b-a}e^{ta}.$$

En posant $p = -a/(b-a)$ et $f(u) = -pu + \log(1-p+pe^u)$ il vient donc

$$\mathbb{E}(e^{tU}) \leq \frac{b}{b-a}e^{ta} - \frac{a}{b-a}e^{tb} = e^{ta}((1-p) + pe^{t(b-a)}) = e^{f(t(b-a))}.$$

À présent on a

$$f'(u) = -p + \frac{pe^u}{1-p+pe^u} \quad \text{et} \quad f''(u) = p(1-p)\frac{e^u}{(1-p+e^u)^2} \leq \frac{1}{4}.$$

Comme $f(0) = f'(0) = 0$, on en déduit que $f(u) \leq u^2/8$ et donc (c'est la partie Hoeffding !)

$$\mathbb{E}(e^{tU}) \leq e^{\frac{t^2}{8}(b-a)^2}.$$

Appliquée à $U = d_k = \mathbb{E}(Y | \mathcal{F}_k) - \mathbb{E}(Y | \mathcal{F}_{k-1})$ sachant \mathcal{F}_{k-1} , cela donne

$$\mathbb{E}(e^{td_k} | \mathcal{F}_{k-1}) \leq e^{\frac{t^2}{8}\text{osc}(d_k)^2}.$$

Ensuite, en écrivant la somme télescopique $Y - \mathbb{E}(Y) = d_n + \dots + d_1$ on obtient

$$\mathbb{E}(e^{t(Y-\mathbb{E}(Y))}) = \mathbb{E}(e^{t(d_{n-1}+\dots+d_1)}\mathbb{E}(e^{td_n} | \mathcal{F}_{n-1})) \leq \dots \leq e^{\frac{t^2}{8}(\text{osc}(d_1)^2+\dots+\text{osc}(d_n)^2)}.$$

À présent, pour tout $t, r > 0$, avec $c := \text{osc}(d_1)^2 + \dots + \text{osc}(d_n)^2$, par l'inégalité de Markov,

$$\begin{aligned} \mathbb{P}(Y - \mathbb{E}(Y) \geq r) &= \mathbb{P}(e^{t(Y-\mathbb{E}(Y))} \geq e^{tr}) \leq e^{-tr} \mathbb{E}(e^{t(Y-\mathbb{E}(Y))}) \\ &\leq e^{-tr + \frac{ct^2}{8}} \leq e^{\inf_{t>0}(-tr + \frac{ct^2}{8})} = e^{-\frac{2r^2}{c}}. \end{aligned}$$

En utilisant cela pour Y et $-Y$, on obtient le résultat souhaité pour $\mathbb{P}(|Y - \mathbb{E}(Y)| \geq r)$. \square

2. À ce propos, l'inégalité dite de Bhatia-Davis affirme que si X est une variable aléatoire de support $[a, b]$ et de moyenne $m \in [a, b]$ alors $\text{Var}(X) \leq (b-m)(m-a)$. Elle est plus fine que la borne $\text{Var}(X) \leq \frac{1}{4}(b-a)^2$.

2 Une application au spectre de matrices aléatoires

Les *valeurs singulières*³ d'une matrice $A \in \mathcal{M}_n(\mathbb{C})$ sont les valeurs propres de la matrice hermitienne semi-définie positive $\sqrt{AA^*}$ où $A^* = \overline{A}^T$. On les note $s_1(A) \geq \dots \geq s_n(A)$. La mesure spectrale empirique associée est $\nu_A = \frac{1}{n} \sum_{k=1}^n \delta_{s_k(A)}$. Lorsque A est aléatoire, il est classique de s'interroger sur le comportement asymptotique de ν_A quand $n \rightarrow \infty$, comme dans les théorèmes de Wigner et de Marchenko-Pastur. Nous allons aborder cette question au moyen de fonctions tests particulières, bien adaptées à l'usage de l'inégalité de Azuma-Hoeffding. Plus précisément, on appelle *variation* d'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ la quantité

$$V(f) = \sup_{x \in \mathcal{S}} \sum_{k \in \mathbb{Z}} |f(x_{k+1}) - f(x_k)| \in [0, +\infty]$$

où \mathcal{S} désigne l'ensemble des suites croissantes de \mathbb{R} indexées par \mathbb{Z} . Si f est dérivable avec $f' \in L^1(\mathbb{R})$ alors $V(f) = \|f'\|_1$, tandis que si $f = \mathbf{1}_{]-\infty, s]}$ avec $s \in \mathbb{R}$ alors $V(f) = 1$.

Théorème 2.1 (Concentration pour les mesures spectrales empiriques). *Si M est une matrice aléatoire à valeurs dans $\mathcal{M}_n(\mathbb{C})$ avec des lignes indépendantes alors pour toute fonction mesurable $f : \mathbb{R} \rightarrow \mathbb{R}$ bornée et pour tout réel $r \geq 0$,*

$$\mathbb{P} \left(\left| \int f d\nu_M - \mathbb{E} \int f d\nu_M \right| \geq r \right) \leq 2 \exp \left(-\frac{nr^2}{2V(f)^2} \right).$$

Il est tout à fait remarquable que ce résultat de concentration ne requière aucune hypothèse d'intégrabilité sur les coefficients de la matrice! La preuve s'adapte sans difficulté au cas des matrices rectangulaires ainsi qu'aux matrices carrées hermitiennes, à coefficients réels ou complexes. Combiné au lemme de Borel-Cantelli, le résultat permet de ramener l'analyse asymptotique de ν_M à sa celle de sa moyenne $\mathbb{E}\nu_M$.

Démonstration. La formule variationnelle de Courant-Fischer s'écrit, pour tout $1 \leq k \leq n$,

$$s_k(A) = \max_{V \in \mathcal{G}_{n,k}} \min_{\substack{x \in V \\ \|x\|_2=1}} \|Ax\|_2 = \min_{V \in \mathcal{G}_{n,n-k+1}} \max_{\substack{x \in V \\ \|x\|_2=1}} \|Ax\|_2$$

où $\mathcal{G}_{n,k}$ est l'ensemble des sous-espaces vectoriels⁴ de dimension k de \mathbb{C}^n . De ces formules on déduit sans trop de difficulté que pour tous $A, B \in \mathcal{M}_n(\mathbb{C})$ avec $\text{rang}(A - B) = 1$, pour tout $1 \leq k \leq n$, avec la convention $s_0(A) = \infty$ et $s_{n+1}(A) = 0$, on a l'entrelacement

$$s_{k+1}(A) \leq s_k(B) \leq s_{k-1}(A).$$

Soient à présent $F_A : \mathbb{R} \rightarrow [0, 1]$ et $F_B : \mathbb{R} \rightarrow [0, 1]$ les fonctions de répartitions des mesures de probabilités ν_A et ν_B définies pour tout $t \in \mathbb{R}$ par

$$F_A(t) = \frac{\text{card}\{1 \leq k \leq n : s_k(A) \leq t\}}{n} \quad \text{and} \quad F_B(t) = \frac{\text{card}\{1 \leq k \leq n : s_k(B) \leq t\}}{n}.$$

Si $\text{rang}(A - B) \leq 1$ alors l'entrelacement précédent donne

$$\sup_{t \in \mathbb{R}} |F_A(t) - F_B(t)| \leq \frac{1}{n}.$$

À présent, si $f : \mathbb{R} \rightarrow \mathbb{R}$ est dérivable avec $f' \in L^1(\mathbb{R})$ alors, par intégration par parties,

$$\left| \int f d\nu_A - \int f d\nu_B \right| = \left| \int_{\mathbb{R}} f'(t)(F_A(t) - F_B(t)) dt \right| \leq \frac{1}{n} \int_{\mathbb{R}} |f'(t)| dt.$$

3. D'une grande importance en analyse numérique et en statistique, cf. [CGLP, Chapitre 4].

4. La notation $\mathcal{G}_{n,k}$ vient du fait qu'en géométrie, l'ensemble $\mathcal{G}_{n,k}$ est une variété grassmannienne.

Comme le membre de gauche ne dépend que d'au plus $2n$ points, on obtient par approximation que pour toute fonction mesurable et bornée $f : \mathbb{R} \rightarrow \mathbb{R}$,

$$\left| \int f d\nu_A - \int f d\nu_B \right| \leq \frac{V(f)}{n}.$$

À partir de maintenant, on fixe $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurable et bornée. Notons $A(x_1, \dots, x_n)$ la matrice $n \times n$ dont les lignes sont x_1, \dots, x_n . On définit $F : (\mathbb{C}^n)^n \rightarrow \mathbb{R}$ par

$$F(x_1, \dots, x_n) = \int f d\nu_{A(x_1, \dots, x_n)}.$$

Pour tout $1 \leq k \leq n$ et tous vecteurs x_1, \dots, x_n, x'_k de \mathbb{C}^n , on a

$$\text{rang} \{ A(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) - A(x_1, \dots, x_{k-1}, x'_k, x_{k+1}, \dots, x_n) \} \leq 1$$

et donc

$$|F(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n) - F(x_1, \dots, x_{k-1}, x'_k, x_{k+1}, \dots, x_n)| \leq \frac{V(f)}{n}.$$

Le résultat découle alors de l'inégalité de Azuma-Hoeffding pour $F(R_1, \dots, R_n)$ où R_1, \dots, R_n sont les lignes de M , avec la filtration $\mathcal{F}_k = \sigma(R_1, \dots, R_k)$ (ici $\|d_k\|_\infty \leq V(f)/n$). \square

3 Une application au problème du voyageur de commerce

Le problème du voyageur de commerce⁵ est un grand classique de l'optimisation combinatoire (randomisée ou non). Il s'agit de trouver une tournée, c'est-à-dire un chemin circulaire qualifié parfois de circuit hamiltonien, de longueur minimale, passant par des points prescrits $X_1, \dots, X_n \in \mathbb{R}^d$, avec $n, d \geq 2$. Cela revient à résoudre sur le groupe symétrique

$$L_n = \min_{\sigma \in \Sigma_n} \sum_{k=1}^n \|X_{\sigma(k)} - X_{\sigma(k+1)}\|_2$$

où $\sigma(n+1) = \sigma(1)$. On peut chercher à déterminer la valeur du minimum ou un point où il est atteint⁶. Concernant la valeur du minimum, le fameux théorème de Bearwood-Halton-Hammersley affirme qu'il existe une constante $0 < \gamma_d < \infty$ qui dépend de $d \geq 2$ telle que si $(X_k)_{k \geq 1}$ sont des variables aléatoires i.i.d. bornées de densité f sur \mathbb{R}^d alors

$$\frac{L_n}{n^{(d-1)/d}} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \gamma_d \int_{\mathbb{R}^d} f(x)^{(d-1)/d} dx.$$

En particulier L_n est d'ordre \sqrt{n} en dimension $d = 2$. La preuve de ce théorème se trouve par exemple dans [S] et [Y]. Nous allons nous contenter d'établir, au moyen de l'inégalité de Azuma-Hoeffding, que la variable aléatoire L_n est concentrée autour de son espérance $\mathbb{E}(L_n)$ à l'échelle $n^{(d-1)/d}$, quand les variables aléatoires $(X_k)_{k \geq 1}$ sont i.i.d. de loi uniforme sur le cube $[0, 1]^d$. La méthode est classique en optimisation combinatoire randomisée. Elle permet, dans ce cas précis, au moyen du lemme de Borel-Cantelli, de réduire la preuve de la convergence presque sûre à la preuve de la convergence de l'espérance.

5. En anglais, on dit Traveling Salesman Problem (TSP). Parmi les problèmes classiques en optimisation combinatoire randomisée, on compte également le problème de la plus longue sous-suite croissante, le problème de l'arbre couvrant minimal, et le problème de l'appariement euclidien minimal. Cf. [S] et [Y].

6. Par exemple avec un algorithme stochastique comme le recuit simulé (simulated annealing).

Théorème 3.1 (Concentration pour le problème du voyageur de commerce randomisé).
 Pour tout $d \geq 2$, si les $(X_k)_{k \geq 1}$ sont i.i.d. de loi uniforme sur le cube $[0, 1]^d$ alors il existe une constante $c_d > 0$ telle que pour tout entier $n \geq 2$ et tout réel $t \geq 0$,

$$\mathbb{P}(|L_n - \mathbb{E}(L_n)| \geq n^{(d-1)/d} t) \leq 2 \exp \begin{cases} -c_d t^2 \frac{n}{\log(n)} & \text{si } d = 2; \\ -c_d t^2 n & \text{si } d > 2. \end{cases}$$

Démonstration. On dispose du résultat géométrique suivant : pour tout $x \in [0, 1]^d$,

$$g_k(x) := \mathbb{E} \left(\min_{1 \leq i \leq k} |X_i - x| \right) \leq c_d k^{-1/d}.$$

En effet, si $B(x, r)$ désigne la boule de centre x et de rayon $r > 0$ dans \mathbb{R}^d , le volume minimal de $B(x, r) \cap [0, 1]^d$ quand x parcourt $[0, 1]^d$ est atteint lorsque x est un coin du cube $[0, 1]^d$. Lorsque $r \leq 1$, la valeur du minimum est $2^{-d} |B(0, r)| = 2^{-d} |B(0, 1)| r^d$ (dessin). Si $1 < r \leq \sqrt{d}$ alors le volume minimal est atteint quand x est un coin du cube, mais la valeur du minimum est difficile à calculer. Elle reste supérieure ou égale à celle du cas $r = 1$. Ainsi, pour tout $0 < r \leq \sqrt{d}$, ce volume minimal est $\geq 2^{-d} |B(0, 1)| (r/\sqrt{d})^d = a_d r^d$. Donc pour tout $x \in [0, 1]^d$ et tout $0 < r \leq \sqrt{d}$, en utilisant à la fin $(1 - u)^k \leq e^{-ku}$,

$$\begin{aligned} \mathbb{P} \left(\min_{1 \leq i \leq k} |X_i - x| \geq r \right) &= \prod_{i=1}^k \mathbb{P}(X_i \in B(x, r)^c) \\ &= \left(1 - |B(x, r) \cap [0, 1]^d| \right)^k \\ &\leq \left(1 - a_d r^d \right)^k \\ &\leq \exp \left(-a_d k r^d \right). \end{aligned}$$

Ceci reste valable si $r > \sqrt{d}$ car dans ce cas $\mathbb{P}(\min_{1 \leq i \leq k} |X_i - x| \geq r) = 0$. À présent,

$$\mathbb{E} \left(\min_{1 \leq i \leq k} |X_i - x| \right) = \int_0^\infty \mathbb{P} \left(\min_{1 \leq i \leq k} |X_i - x| \geq r \right) dr \quad \text{et} \quad \int_0^\infty e^{-br^d} dr = \frac{\Gamma(1/d)}{db^{1/d}}.$$

Ceci achève la preuve du résultat géométrique annoncé. Utilisons à présent l'inégalité de Azuma-Hoeffding pour $F(X_1, \dots, X_n) = L_n$ et $\mathcal{F}_k = \sigma(X_1, \dots, X_k)$. Il nous faut majorer $\text{osc}(d_k)$. Soit L la fonction qui associe à un ensemble fini de \mathbb{R}^d la longueur minimale de la tournée. Pour tout $S \subset \mathbb{R}^d$ fini et tout $x \in \mathbb{R}^d$,

$$L(S) \leq L(S \cup \{x\}) \leq L(S) + 2 \min_{y \in S} |x - y|.$$

En appliquant cette inégalité à $S = \{x_1, \dots, x_n\} \setminus \{x_k\}$ et à $x = x_k$ et $x = x'_k$ on obtient

$$\begin{aligned} |F(x_1, \dots, x_k, \dots, x_n) - F(x_1, \dots, x'_k, \dots, x_n)| &\leq 2 \min_{i \neq k} |x'_k - x_i| + 2 \min_{i \neq k} |x_k - x_i| \\ &\leq 2 \min_{i > k} |x'_k - x_i| + 2 \min_{i > k} |x_k - x_i|. \end{aligned}$$

En posant $g_k(x) := \mathbb{E} \left(\min_{1 \leq i \leq k} |X_i - x| \right)$ on obtient pour $1 \leq k \leq n - 1$,

$$|d_k| \leq 2 \mathbb{E} \left(\min_{i > k} |X_k - X_i| + \min_{i > k} |X'_k - X_i| \middle| \mathcal{F}_k \right) = 2g_{n-k}(X_k) + 2\mathbb{E}(g_{n-k}(X'_k)).$$

Le résultat géométrique précédent donne, pour $d \geq 2$ et $1 \leq k \leq n - 1$,

$$\|d_k\|_\infty \leq c_d(n - k)^{-1/d}.$$

Comme $\|d_n\|_\infty \leq c_d$ pour une autre constante car les X_i sont bornées, on obtient

$$\sum_{k=1}^n \|d_k\|_\infty^2 \leq \begin{cases} c_d \log(n) & \text{pour } d = 2, \\ c_d n^{(d-2)/d} & \text{pour } d > 2. \end{cases}$$

Cela donne enfin, pour tout réel $t \geq 0$,

$$\mathbb{P}(|L_n - \mathbb{E}(L_n)| \geq t) \leq 2 \exp \begin{cases} -\frac{c_d t^2}{\log(n)} & \text{si } d = 2; \\ -\frac{c_d t^2}{n^{(d-2)/d}} & \text{si } d > 2. \end{cases}$$

□

Notes et commentaires. La preuve de l'inégalité de Azuma-Hoeffding donnée ici est tirée de l'article de survol de McDiarmid [D]. L'application aux mesures spectrales de matrices se trouve dans [BCC] et [GL]. Dans le même esprit, on pourra consulter le travail récent de Tropp [Tr]. La borne log-linéaire en n peut être renforcée en n^2 pour les matrices à coefficients i.i.d. gaussiens, cf. [GZ]. L'application à la preuve du théorème de Marchenko-Pastur classique se trouve par exemple dans [CGLP]. L'usage de l'inégalité de Azuma-Hoeffding en optimisation combinatoire randomisée – en particulier au voyageur de commerce – se trouve par exemple dans [S] et [Y]. Pour en savoir un peu plus sur les inégalités de concentration en général, on pourra consulter [L] et [BLM].

Références

- [A] K. Azuma, *Weighted sums of certain dependent random variables*, Tôhoku Math. J. (2) 19 (1967) 357–367, MR-0221571
- [BCC] Ch. Bordenave, P. Caputo, D. Chafaï, *Spectrum of non-Hermitian heavy tailed random matrices*, Communications in Mathematical Physics 307, 513–560 (2011) MR-2837123 arXiv :1006.1713
- [BLM] S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities with applications*, proto-livre (2012)
- [CGLP] D. Chafaï, O. Guédon, G. Lecué, A. Pajor, *Interactions between compressed sensing, random matrices, and high dimensional geometry*, proto-livre Panoramas et Synthèses (2012)
- [D] C. McDiarmid, *On the method of bounded differences*, Surveys in combinatorics (1989) London Math. Soc. Lecture Note Ser. 141, 148–188, Cambridge Univ. Press (1989) MR-1036755
- [GL] A. Guntuboyina and H. Leeb, *Concentration of the spectral measure of large Wishart matrices with dependent entries*, Electron. Commun. Probab. 14 (2009) 334–342, MR-2535081
- [GZ] A. Guionnet, O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab. 5 (2000), 119–136, MR-1781846
- [H] W. Hoeffding *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58 (1963) 13–30 MR-0144363
- [L] M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs 89, American Mathematical Society (2001) MR-1849347
- [M] B. Maurey, *Construction de suites symétriques*, CRAS Paris Sér. A-B 288 (1979), no. 14, MR-0533901
- [S] M. J. Steele, *Probability theory and combinatorial optimization*, SIAM (1997) viii+159, MR-1422018
- [T] M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, IHES Publ. Math. No. 81 (1995), 73–205, MR-1361756
- [Tr] J. Tropp, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math. (2011)
- [Y] J. E. Yukich, *Probability theory of classical Euclidean optimization problems*, Lecture Notes in Mathematics, 1675. Springer-Verlag (1998) x+152 pp. MR-1632875