Concentration inequalities:

basics and some new challenges

M. Ledoux

University of Toulouse, France

& Institut Universitaire de France

Measure concentration

geometric functional analysis,

probability theory, statistics, learning theory,

statistical mechanics, stochastic dynamics,

random matrix theory,

randomized algorithms, complexity,

and more...

Part I

milestones of measure concentration

major tools and results

Part II

super-concentration

some new challenges

Part I

milestones of measure concentration

major tools and results

isoperimetric ideas

Gaussian concentration

convex distance inequality

entropic method

P. Lévy (1919)

spherical isoperimetry

μ normalized uniform measure on $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ standard sphere

$f : \mathbb{S}^n \to \mathbb{R}$ continuous

$$\mu\big(|f - m| < \omega(\eta)\big) \ \geq \ 1 - 2\,e^{-(n-1)\,\eta^2/2}$$

m median of f for μ

$\omega(\eta)$ modulus of continuity of f

for n large, functions with small oscillations

are almost constant (high dimensional effect)

V. Milman (1970)

Dvoretzky's theorem on spherical sections of convex bodies

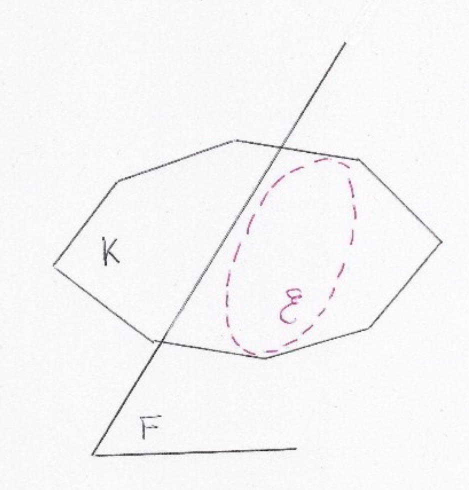
for all $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$

every convex body $K \subset \mathbb{R}^n$ (unit ball for $\| \cdot \|$)

there exist F subspace of $\mathbb{R}^n$, $\dim(F) \geq \delta(\varepsilon) \log n$

and $\mathcal{E} \subset F$ ellipsoid

$$(1 - \varepsilon)\,\mathcal{E} \subset K \cap F \subset (1 + \varepsilon)\,\mathcal{E}$$



$$(1-\varepsilon)\,\mathcal{E} \;\subset\; K\cap F \;\subset\; (1+\varepsilon)\,\mathcal{E}$$

find F **at random**

concentration of spherical measures in high dimension

Lévy's inequality

$$\mu\big(|f - m| < \omega(\eta)\big) \;\geq\; 1 - 2\,e^{-(n-1)\,\eta^2/2}$$

$$f : \mathbb{S}^n \to \mathbb{R}, \quad f(x) = \|x\|$$

$\|\cdot\|$ gauge of K

metric measure space (X, d, μ)

(X, d) metric space

μ Borel measure on X, $\mu(X) = 1$

concentration function

$$\alpha_\mu(r) = \alpha_{(X,d,\mu)}(r) = \sup\left\{1 - \mu(A_r)\,;\, \mu(A) \geq \tfrac{1}{2}\right\}, \quad r > 0$$

$$A_r = \left\{x \in X\,;\, d(x, A) < r\right\}$$

μ uniform on $\mathbb{S}^n \subset \mathbb{R}^{n+1}$: $\quad \alpha_\mu(r) \leq e^{-(n-1)\,r^2/2}$

if $\mu(A) \geq \frac{1}{2}$, then for $r \sim \frac{1}{\sqrt{n}}$, $\quad \mu(A_r) \approx 1$

$$F : X \to \mathbb{R} \quad 1\text{-Lipschitz} \quad (\|F\|_{\mathrm{Lip}} \leq 1)$$

m median of F for μ $\qquad \mu(F \leq m),\ \mu(F \geq m) \geq \frac{1}{2}$

$$A = \{F \leq m\} \implies A_r \subset \{F < m + r\}$$

deviation inequality

$$\mu(F \geq m + r) \leq \alpha_\mu(r), \quad r > 0$$

concentration inequality

$$\mu\big(|F - m| \geq r\big) \leq 2\,\alpha_\mu(r), \quad r > 0$$

median $\leftrightarrow$ **mean**

variance bound $\qquad \mathrm{Var}_\mu(F) \leq 4\int_0^\infty r\alpha_\mu(r)dr \quad (\leq C\,\|F\|^2_{\mathrm{Lip}})$

variety of examples and tools

- spectral methods
- probabilistic and combinatorial tools
- product measures
- coupling, correlation estimates
- geometric, functional, transportation inequalities

$$d\gamma(x) = e^{-|x|^2/2} \, \frac{dx}{(2\pi)^{n/2}}$$ on $\mathbb{R}^n$ (Euclidean metric)

$$\alpha_\gamma(r) \leq e^{-r^2/2}$$

if $\gamma(A) \geq \frac{1}{2}$ then $\gamma(A_r) \geq 1 - e^{-r^2/2}, \quad r > 0$

$r = 5$ or 10, $\qquad \gamma(A_r) \approx 1$

$F : \mathbb{R}^n \to \mathbb{R}$ 1- Lipschitz, m median or mean

$$\gamma\big(|F - m| \geq r\big) \leq 2\, e^{-r^2/2}, \quad r \geq 0$$

independent of the dimension n

V. Sudakov, B. Tsirel'son (1974), C. Borell (1975)

reference example

geometry of Gauss space

infinite dimensional analysis

Ornstein-Uhlenbeck semigroup, hypercontractivity

supremum of Gaussian processes

dimension reduction

$\mathcal{F}$ collection of functions $f : S \to \mathbb{R}$

$G(f), f \in \mathcal{F}$ centered Gaussian process

$M = \sup_{f \in \mathcal{F}} G(f)$, M Lipschitz

Gaussian concentration

$$\mathbb{P}\big(|M - m| \geq r\big) \leq 2\, e^{-r^2/2\sigma^2}, \quad r \geq 0$$

m mean or median, $\sigma^2 = \sup_{f \in \mathcal{F}} \mathbb{E}\big(G(f)^2\big)$

$$\mathrm{Var}\big(\sup_{f \in \mathcal{F}} G(f)\big) = \mathrm{Var}(M) \leq \sigma^2 = \sup_{f \in \mathcal{F}} \mathrm{Var}\big(G(f)\big)$$

Johnson-Lindenstrauss (1984)

metric geometry, theoretical computer science

compressed sensing, machine learning

N points $p_1, \ldots, p_N$ in $\mathbb{R}^n$, $\varepsilon > 0$

there exists (at random) $\pi : \mathbb{R}^n \to \mathbb{R}^k$ linear

$$k \geq \frac{c}{\varepsilon^2} \log N$$

$$(1-\varepsilon)|p_i - p_j| \leq \big|\pi(p_i) - \pi(p_j)\big| \leq (1+\varepsilon)|p_i - p_j|$$

π **quasi-isometry**

$$X = \{0,1\}^n$$

$$d(x,y) \ = \ \# \left\{ 1 \leq i \leq n;\, x_i \neq y_i \right\}$$ Hamming metric

μ uniform

$$\alpha_\mu(r) \ \leq \ e^{-r^2/2n}$$

$F : X = \{0,1\}^n \to \mathbb{R}$ $\quad 1$- Lipschitz

$$\big| F(x_1, \dots, x_i, \dots, x_n) - F(x_1, \dots, y_i, \dots, x_n) \big| \ \leq \ 1$$

$$\mu\big(|F - m| \geq r \big) \ \leq \ 2\, e^{-r^2/2n}, \quad r \geq 0,$$ (m mean or median)

$$X = \{0,1\}^n$$

$$d(x,y) = \#\{1 \le i \le n;\, x_i \ne y_i\}$$ Hamming metric

μ uniform: product measure

$$\alpha_\mu(r) \le e^{-r^2/2n}$$

same **on any** product space

$$X = X_1 \times \cdots \times X_n$$

$$d(x,y) = \#\{1 \le i \le n;\, x_i \ne y_i\}$$ Hamming metric

$$\mu = \mu_1 \otimes \cdots \otimes \mu_n$$

$$X = X_1 \times \cdots \times X_n$$

$$d(x,y) = \# \{1 \leq i \leq n;\, x_i \neq y_i\}$$ Hamming metric

$$\mu = \mu_1 \otimes \cdots \otimes \mu_n$$

$$\alpha_\mu(r) \leq e^{-r^2/2n}$$

weighted Hamming metric

$$d_w(x,y) = \sum_{i=1}^{n} w_i \, \mathbf{1}_{x_i \neq y_i}, \quad w_i \geq 0$$

$$\alpha_\mu(r) \leq e^{-r^2/2|w|^2}, \qquad |w|^2 = \sum_{i=1}^n w_i^2$$

$$\mu(A) \geq \frac{1}{2}, \qquad \mu\big(d_w(\cdot, A) \geq r\big) \leq e^{-r^2/2|w|^2}$$

M. Talagrand (1995)

$$\mu(A) \geq \tfrac{1}{2}\,, \qquad \mu\Big(d_w(\cdot, A) \geq r\Big) \;\leq\; e^{-r^2/2|w|^2}$$

uniform in the weight

$$\mu(A) \geq \tfrac{1}{2}\,, \qquad \mu\Big(\sup_{|w|=1} d_w(x, A) \geq r\Big) \;\leq\; 2\,e^{-r^2/4}$$

applications to empirical processes, learning,

combinatorial probability, convex geometry,

random matrices etc.

sample illustration

$\mu = \mu_1 \otimes \cdots \otimes \mu_n$ on $[0,1]^n$

$F : [0,1]^n \to \mathbb{R}$ 1-Lipschitz , **convex**

m median

$$\mu\big(|F - m| \geq r\big) \leq 4\, e^{-r^2/4}, \quad r \geq 0$$

same as for Gaussian

dimension free

$X_1, \ldots, X_n$ independent random variables in $(S, \mathcal{S})$

$\mathcal{F}$ collection of functions $f : S \to \mathbb{R}$

$$M = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i)$$

M Lipschitz and convex

concentration inequalities on

$$\mathbb{P}\Big(\big| M - \mathbb{E}(M) \big| \geq r \Big), \quad r \geq 0$$

quantify the asymptotics $n \to \infty$

$$M = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i)$$

$$|f| \leq 1, \quad \mathbb{E}(f(X_i)) = 0, \quad f \in \mathcal{F}$$

$$\mathbb{P}\big(|M - m| \geq r\big) \leq C \exp\left(-\frac{r}{C} \log\left(1 + \frac{r}{\sigma^2 + m}\right)\right), \quad r \geq 0$$

m mean or median, $C > 0$ numerical constant

$$\sigma^2 = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} \mathbb{E}(f^2(X_i))$$

M. Talagrand (1996)

I. Herbst (1975)

link between functional inequalities

and measure concentration

logarithmic Sobolev inequalities

connection with optimal transportation

model: logarithmic Sobolev inequality (γ standard normal)

$$\int_{\mathbb{R}^n} f \log f \, d\gamma \;\leq\; \frac{1}{2} \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} \, d\gamma, \qquad f > 0, \; \int_{\mathbb{R}^n} f \, d\gamma = 1$$

$F : \mathbb{R}^n \to \mathbb{R}$ 1- Lipschitz $\int_{\mathbb{R}^n} F \, d\gamma = 0$

$$Z(\lambda) = \int_{\mathbb{R}^n} e^{\lambda F} d\gamma, \quad \lambda \in \mathbb{R}$$

$$\lambda Z'(\lambda) - Z(\lambda) \log Z(\lambda) \;\leq\; \frac{\lambda^2}{2} Z(\lambda)$$

$$Z(\lambda) \;\leq\; e^{\lambda^2/2}$$

Gaussian concentration $\gamma\big(|F| \geq r\big) \;\leq\; 2\, e^{-r^2/2}, \quad r \geq 0$

most successful tool

- empirical processes (convex distance inequality)
- variety of functional inequalities
- models from statistical mechanics
- optimal transport

S. Boucheron, G. Lugosi, P. Massart (2013)

μ probability measure on $\mathbb{R}^n$

$$H(\nu \,|\, \mu) \, = \, \int \log \frac{d\nu}{d\gamma} \, d\nu, \quad \nu \ll \mu$$

relative entropy

$$\mathrm{W}_p(\nu, \mu)^p \, = \, \inf_{\nu \leftarrow \pi \rightarrow \mu} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |x - y|^p d\pi(x, y)$$

$$(1 \leq p < \infty)$$

Kantorovich-Wasserstein distance

Otto-Villani theorem (2000)

Fisher information $\quad \mathrm{I}(\nu \,|\, \mu) \,=\, \int_{\mathbb{R}^n} \Big| \nabla \Big(\log \frac{d\nu}{d\mu} \Big) \Big|^2 d\nu$

logarithmic Sobolev inequality for μ (on $\mathbb{R}^n$)

$$\mathrm{H}(\nu \,|\, \mu) \,\leq\, \frac{C}{2}\, \mathrm{I}(\nu \,|\, \mu), \qquad \nu \ll \mu$$

implies quadratic transportation cost inequality

$$\mathrm{W}_2^2(\nu, \mu) \,\leq\, 2C\, \mathrm{H}(\nu \,|\, \mu), \qquad \nu \ll \mu$$

$\mu = \gamma$ M. Talagrand (1996)

K. Marton (1996), S. Bobkov, F. Götze (1999)

$$W_1(\nu, \mu) \leq C \sqrt{H(\nu \,|\, \mu)}, \quad \nu \ll \mu$$

equivalent to Gaussian concentration $\alpha_\mu(r) \leq e^{-r^2/C}$

N. Gozlan (2009)

$$W_2(\nu, \mu) \leq C \sqrt{H(\nu \,|\, \mu)}, \quad \nu \ll \mu$$

equivalent to

dimension free Gaussian concentration $\alpha_{\mu^{\otimes n}}(r) \leq e^{-r^2/C}$

- analysis and geometry of metric measure spaces
- empirical processes for dependent variables
- interacting models, Gibbs measures
- stochastic dynamical systems
- large deviations
- exchangeable pairs, Stein's method

Part II

some new challenges of measure concentration

measure concentration property

general, valid for **every** Lipschitz function

specific functionals

with specific features

improved concentration bounds

S. Chatterjee (2013)

standard concentration

$$\mathrm{Var}\big(F(X)\big) = O\big(\|F\|_{\mathrm{Lip}}^2\big)$$

super-concentration

$$\mathrm{Var}\big(F(X)\big) = o\big(\|F\|_{\mathrm{Lip}}^2\big)$$

sub-linearity, sub-diffusivity

Gaussian fields, percolation models,

spin glasses, extreme eigenvalues etc.

$(X_1, \ldots, X_n)$ Gaussian sample

$$M_n = \max_{1 \leq i \leq n} X_i$$

standard concentration

$$\mathrm{Var}(M_n) = \mathrm{Var}\Big(\max_{1 \leq i \leq n} X_i\Big) \leq \max_{1 \leq i \leq n} \mathrm{Var}(X_i)$$

example: $X_1, \ldots, X_n$ iid standard normal

$$\mathrm{Var}(M_n) = O\Big(\frac{1}{\log n}\Big)$$

(Gumbel fluctuations)

binary tree V_n of depth n

$X_v, v \in V_n$ iid standard normal

$$M_n = \max_\tau \sum_{v \in \tau} X_v, \qquad \tau \text{ path in } V_n$$

standard concentration

$$\mathrm{Var}(M_n) \leq \max_\tau \mathrm{Var}\Big(\sum_{v \in \tau} X_v \Big) = n$$

however $\mathrm{Var}(M_n) = O(1)$

baby model of (2-dimensional) Gaussian free field

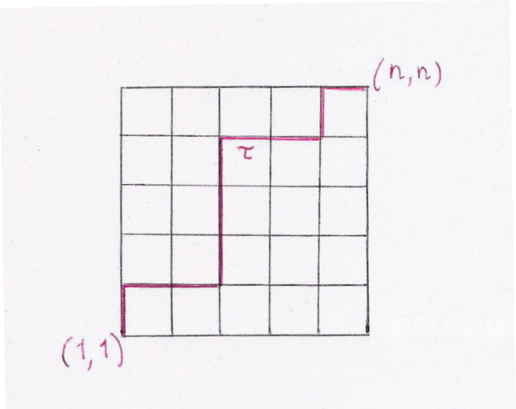
M. Bramson, J. Ding, O. Zeitouni (2012-16)

(branching random walks/Brownian motions)

$X_{ij}, \quad 1 \leq i,j \leq n,$ iid random variables

$$M_n = \max_{\tau} \sum_{(i,j) \in \tau} X_{ij}$$

τ up/right paths from $(1,1)$ to (n,n)



$$M_n = \max_{\tau} \sum_{(i,j)\in\tau} X_{ij} \qquad 2n\text{-Lipschitz (in the } X_{ij}\text{'s)}$$

X_{ij} Gaussian (or Bernoulli)

$\mathrm{Var}(M_n) = O(n)$ Gaussian behaviour

however (X_{ij} exponential)

$\mathrm{Var}(M_n) = O(n^{2/3})$ random matrix behaviour

best known $\mathrm{Var}(M_n) = O\left(\frac{n}{\log n}\right)$

I. Benjamini, G. Kalai, O. Schramm (2003)

B. Graham, S. Chatterjee (2008)

general framework

$X = (X_t)_{t \in T}$ Gaussian process

$$\mathbb{E}\Big(\sup_{t \in T} X_t\Big)$$

general chaining method

$$\mathrm{Var}\Big(\sup_{t \in T} X_t\Big) \quad ?$$

lecture 2: generic/general methods

towards super-concentration (sub-linear variance)

Poincaré inequalities, hypercontractivity

$X = X^N = (X_{ij})_{1 \leq i,j \leq N}$ symmetric matrix

Wigner matrices: $X_{ij}, i \leq j,$ independent (finite moments)

statistics of interest

$\lambda_1 \leq \cdots \leq \lambda_N$ eigenvalues of $\frac{X}{\sqrt{N}}$

spectral measure $\frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$

eigenvalues in the bulk $\lambda_j, \quad \eta N \leq j \leq (1-\eta)N$

eigenvalues at the edge $\lambda_N = \max_{1 \leq i \leq N} \lambda_i$

Wigner matrices

$X_{ij}, i \leq j$ iid, $\mathbb{E}(X_{ij}) = 0, \ \mathbb{E}(X_{ij}^2) = 1$

E. Wigner (1955)

limiting spectral measure

$$\frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i} \to \text{sc}$$

semi-circle law $\quad d\text{sc}(x) = \frac{1}{2\pi} \sqrt{4 - x^2}\, dx$

supported on $[-2, +2]$

γ_j theoretical locations of the eigenvalues

$$\int_{-2}^{\gamma_j} d\mathrm{sc}(x) = \frac{j}{N}, \qquad 1 \leq j \leq N$$

bulk $\quad \eta N \leq j \leq (1-\eta)N$

$$\frac{\sqrt{4-\gamma_j^2}\,N}{\sqrt{2\log N}}\,(\lambda_j - \gamma_j) \to G \qquad \text{standard normal}$$

GUE J. Gustavsson (2005)

Wigner matrices T. Tao, V. Vu (2011)

(four-moment theorem)

edge $\lambda_N \to 2$ (a.s.)

$$N^{2/3}\big[\lambda_N - 2\big] \to F_{TW}$$

F_{TW} Tracy-Widom distribution

GUE/GOE C. Tracy, H. Widom (1995)

Wigner matrices A. Soshnikov (1999),

L. Erdös, H. T. Yau and co., T. Tao, V. Vu (2010-12),

J.-O. Lee, J. Yin (2014)

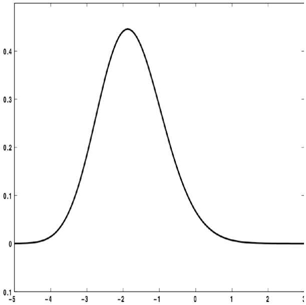
$$F_{\mathrm{TW}}(s) = \exp\bigg(-\int_s^\infty (x-s)u(x)^2 dx\bigg), \quad s \in \mathbb{R}$$

$u'' = 2u^3 + xu$ Painlevé II equation

mean $\simeq -1.77$

$$F_{\mathrm{TW}}(s) \sim e^{-s^3/12} \quad \text{as} \quad s \to -\infty$$

$$1 - F_{\mathrm{TW}}(s) \sim e^{-4s^{3/2}/3} \quad \text{as} \quad s \to +\infty$$



density (GUE)

variance bounds?

$X_{ij},\ i \leq j$ iid standard Gaussians

classical concentration

λ_j, λ_N 1-Lipschitz

$$\mathrm{Var}(\lambda_j),\ \mathrm{Var}(\lambda_N) = O\Big(\frac{1}{N}\Big),$$ Gaussian tails

asymptotics

$$\mathrm{Var}(\lambda_j) = O\Big(\frac{\log N}{N^2}\Big) \qquad (\eta N \leq j \leq (1-\eta)N)$$

$$\mathrm{Var}(\lambda_N) = O\Big(\frac{1}{N^{4/3}}\Big),$$ Tracy-Widom tail

tail bounds reflecting

Tracy-Widom asymptotics and F_{TW}

$$N^{2/3}\big[\lambda_N - 2\big] \,\to\, F_{TW}$$

$$\mathbb{P}\big(\lambda_N \geq 2+\varepsilon\big) \leq C\, e^{-N\,\varepsilon^{3/2}/C}, \quad 0 < \varepsilon \leq 1$$

$$\mathbb{P}\big(\lambda_N \leq 2-\varepsilon\big) \leq C\, e^{-N^2\,\varepsilon^{3}/C}, \quad 0 < \varepsilon \leq 1$$

$$\varepsilon \,=\, \frac{s}{N^{2/3}}$$

GUE/GOE models

$$\mathrm{Var}(\lambda_j) = O\Big(\frac{\log N}{N^2}\Big) \qquad (\eta N \leq j \leq (1-\eta)N)$$

$$\mathrm{Var}(\lambda_N) = O\Big(\frac{1}{N^{4/3}}\Big)$$

determinantal/tridiagonal structure

Wigner matrices S. Dallaporta, V. Vu (2011)

(localization and four-moment theorem)

GUE/GOE models

$$\mathbb{P}\big(\lambda_N \geq 2 + \varepsilon\big) \leq C\, e^{-N\,\varepsilon^{3/2}/C}, \quad 0 < \varepsilon \leq 1$$

$$\mathbb{P}\big(\lambda_N \leq 2 - \varepsilon\big) \leq C\, e^{-N^2\,\varepsilon^{3}/C}, \quad 0 < \varepsilon \leq 1$$

determinantal/tridiagonal structure

J. Ramirez, B. Rider, B. Virag (2011)

Wigner matrices: partly open

right of the mean: O. Feldheim, S. Sodin (2010) (moment method)

similar questions for β-ensembles

GUE model

bulk $\quad \eta N \leq j \leq (1-\eta)N$

$$\frac{1}{2\pi}\sqrt{4-\gamma_j^2}\,N\,(\lambda_{j+1}-\lambda_j) \to F_{\text{Gaudin}}$$

T. Tao (2013)

variance and (Gaussian) tail bounds?

Kannan-Lovasz-Simonovits conjecture

$d\mu = e^{-V}dx$ log-concave on $\mathbb{R}^n$ (V convex)

covariance $\mu = \mathrm{Id}$

$$\mathrm{Var}(f) \leq C \int_{\mathbb{R}^n} |\nabla f|^2 d\mu$$

C independent of n

E. Milman (2009)

enough $\mathrm{Var}(f) \leq C \, \|f\|_{\mathrm{Lip}}^2$

specific functional: $f(x) = |x|$

central limit theorem for log-concave measures

B. Klartag (2007)

conjecture

$$\mathrm{Var}\big(|x|\big) \le C$$

best known result

$$\mathrm{Var}\big(|x|\big) \le C\, n^{2/3}$$

O. Guédon, E. Milman (2011)

Plan of the lectures

lecture 2 (on the board)

general hypercontractive tools towards sub-linearity

(soft, non-optimal rates)

Talagrand's inequality and influences

applications to some Gaussian models

Plan of the lectures

lecture 3 (on the board)

(optimal) bounds on GUE/GOE eigenvalues

variance bounds for Wigner matrices

exponential bounds

Thank you for your attention