

New challenges of measure concentration

M. Ledoux

Institut de Mathématiques de Toulouse, France

Measure concentration

widely used

geometric functional analysis, convex geometry,

probability theory, statistics, learning theory,

statistical mechanics, stochastic dynamics,

random matrix theory, quantum information theory,

randomized algorithms, complexity,

graphs, topological infinite group theory,

and more...

Part I

Milestones

of **measure concentration**

some of the major tools

and methods

P. Lévy's inequality 1919

spherical isoperimetry

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

$F : S^n \rightarrow \mathbb{R}$ continuous

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

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

Milestone 1

real birth : new proof by **V. Milman 1970**

Dvoretzky's theorem on spherical sections of convex bodies

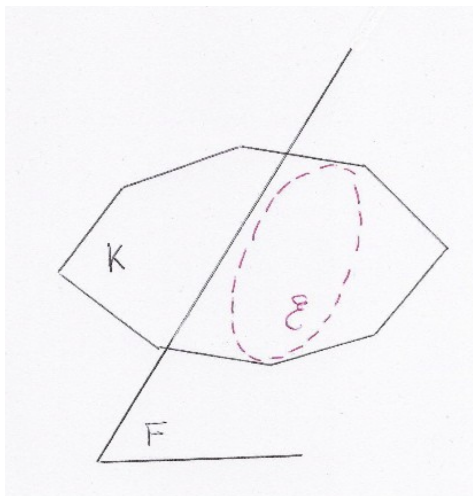
for all $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$

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

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



$$(1 - \varepsilon)E \subset K \cap F \subset (1 + \varepsilon)E$$

Milman's (Dvoretzky) idea :

find F at random

concentration of spherical measures in high dimension

Lévy's inequality

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

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

$\|\cdot\|$ gauge of K

new proof by **V. Milman 1970**

asymptotic geometric analysis



Vitali Milman

new proof by **V. Milman 1970**

asymptotic geometric analysis

“The concentration of measure phenomenon, ubiquitous in probability theory and statistical mechanics, was brought to geometry (starting from Banach spaces) by Vitali Milman, following the earlier work by Paul Lévy”

(M. Gromov 1999)

“The idea of concentration of measure, which was discovered by Vitali Milman, is arguably one of the great ideas of analysis in our times”

(M. Talagrand 1996)

framework for measure concentration

metric measure space (X, d, μ)

(X, d) metric space

μ Borel measure on X , $\mu(X) = 1$

concentration function

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

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

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

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

equivalent formulation on functions (Lévy's inequality)

$F : X \rightarrow \mathbb{R}$ 1-Lipschitz

m median of F for μ

$$\mu(F \leq m), \mu(F \geq m) \geq \frac{1}{2}$$

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

deviation inequality

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

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

median \leftrightarrow mean

variance bound $\text{Var}_\mu(F) \leq 4 \int_0^\infty r \alpha_\mu(r) dr \quad (\leq C)$

measure concentration property

less restrictive than isoperimetry ($r \rightarrow 0$)

easier to establish, widely shared

variety of examples and tools

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

some basic examples

illustrative applications

Gaussian concentration

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

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

$$r = 5 \quad \text{or} \quad 10, \quad \gamma(A_r) \approx 1$$

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

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

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

independent of the dimension n

Gaussian concentration

reference example

geometry of Gauss space

infinite dimensional analysis

Ornstein-Uhlenbeck semigroup, hypercontractivity

supremum of Gaussian processes

dimension reduction

example of **application**

the **Johnson-Lindenstrauss 1984 flatening lemma**

(metric geometry, theoretical computer science

compressed sensing, machine learning)

D. Dubhashi lectures

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

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

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

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

π **quasi-isometry**

concentration on the discrete cube

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

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

μ uniform

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

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

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

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

m mean or median

concentration on the discrete cube

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

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

μ uniform

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

concentration on the discrete cube

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

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

μ uniform : product measure

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

concentration on the discrete cube

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

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

μ uniform : product measure

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

same **on any** product space

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

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

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

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

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

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

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

$$F : X \rightarrow \mathbb{R} \quad \text{1-Lipschitz (bounded difference)}$$

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

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

martingale (induction) method

V. Milman, G. Schechtman 1986, C. McDiarmid 1989

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

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

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

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

weighted Hamming metric

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

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

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

Milestone 2

Talagrand's convex distance inequality 1995

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

$$\mu(A) \geq \frac{1}{2}, \quad \mu\left(\sup_{|w|=1} d_w(x, A) \geq r\right) \leq 2e^{-r^2/4}$$

(geometric) induction over the dimension n

typical application

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

$$F : [0, 1]^n \rightarrow \mathbb{R} \quad \text{1-Lipschitz, convex}$$

m median

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

same as for Gaussian

dimension free



Michel Talagrand

“A random variable that depends (in a “smooth” way)
on the influence of many independent random
variables (but not too much on any of them)
is essentially constant”

“A random variable that smoothly depends
on the influence of many independent
random variables satisfies Chernoff-type bounds”

(M. Talagrand 1996)

application to empirical processes

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

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

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

Z Lipschitz and convex

concentration inequalities on

$$\mathbb{P}\left(|Z - \mathbb{E}(Z)| \geq r\right), \quad r \geq 0$$

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

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

$$\mathbb{P}(|Z - M| \geq t) \leq K \exp\left(-\frac{t}{K} \log\left(1 + \frac{t}{\sigma^2 + M}\right)\right), \quad t \geq 0$$

$K > 0$ numerical constant, M mean or median

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

M. Talagrand 1996

q -point inequality

statistical applications : **P. Massart** lectures

numerous applications of Gaussian concentration
and Talagrand's convex distance inequality

**randomized algorithms, complexity,
theoretical computer science**

A. Panconesi lectures

Milestone 3

the **Herbst argument**

(**entropic method**)

link between functional inequalities

and measure concentration

connection with optimal transportation

the Herbst argument 1975

model : logarithmic Sobolev inequality

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

$$f > 0, \quad \int_{\mathbb{R}^n} f \, d\gamma = 1$$

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

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

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

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

Gaussian concentration



Ira Herbst

entropic method

- empirical processes
- variety of functional inequalities
- models from statistical mechanics
- optimal transport

entropic method : empirical processes

new proof of **Talagrand's** inequality
for empirical processes

P. Massart 2000

S. Boucheron, G. Lugosi, P. Massart 2005

moment methods

new proof by **S. Boucheron, G. Lugosi, P. Massart 2003**

of Talagrand's convex distance inequality **itself**

entropic method : functional inequalities

spectral or Poincaré inequalities
and exponential concentration

M. Gromov, V. Milman 1983

A. Borovkov, S. Utev 1983

families of functional inequalities

A. Guillin lectures

new developments with Lyapunov function method

entropic method : optimal transportation

μ probability measure on \mathbb{R}^n

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

relative entropy

$$W_p(\nu, \mu)^p = \inf_{\nu \leftarrow \pi \rightarrow \mu} \iint |x - y|^p d\pi(x, y)$$
$$(1 \leq p < \infty)$$

Kantorovich-Rubinstein-Wasserstein distance

transportation cost methods

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

coupling and transportation methods

- 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

new challenges of measure concentration

specific functionals

rather than inequalities

valid for **every** Lipschitz function

variance bounds

dependence structures

discrete and combinatorial models

geometric and combinatorial features

Example 1 : maxima of Gaussian variables

example : X_1, \dots, X_n iid standard Gaussian

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

$$F = \max_{1 \leq i \leq n} x_i \quad 1\text{-Lipschitz}$$

standard concentration argument

equivalent formulation on functions (Lévy's inequality)

$F : X \rightarrow \mathbb{R}$ 1-Lipschitz

m median of F for μ

$$\mu(F \leq m), \mu(F \geq m) \geq \frac{1}{2}$$

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

deviation inequality

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

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

median \leftrightarrow mean

variance bound $\text{Var}_\mu(F) \leq 4 \int_0^\infty r \alpha_\mu(r) dr \quad (\leq C)$

Example 1 : maxima of Gaussian variables

example : X_1, \dots, X_n iid standard Gaussian

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

$$F = \max_{1 \leq i \leq n} x_i \quad \text{1-Lipschitz}$$

standard concentration argument

$$\text{Var}(M_n) \leq 1$$

however

$$\text{Var}(M_n) = O\left(\frac{1}{\log n}\right)$$

general framework

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

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

general chaining method

$$\text{Var} \left(\sup_{t \in T} X_t \right) \quad ?$$

specific processes

Gaussian free field, percolation models,
spin glasses, extreme eigenvalues etc

S. Chatterjee lectures (super-concentration)

Example 2 : random matrices

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

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

statistics of interest

$\lambda_1 \leq \dots \leq \lambda_n$ eigenvalues of $\frac{X}{\sqrt{n}}$

eigenvalue counting function, $I \subset \mathbb{R}$ interval

$$\mathcal{N}_I = \sum_{i=1}^n \mathbf{1}_{\{\lambda_i \in I\}}$$

largest eigenvalue $\lambda_{\max} = \max_{1 \leq i \leq n} \lambda_i$

asymptotics

$$\frac{1}{n} \mathcal{N}_I = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\lambda_i \in I\}} \rightarrow \sigma(I)$$

σ limiting spectral measure compactly supported $(-a, +a)$

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

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

central limit theorem : I interval $\subset (-a, +a)$

$$\frac{\mathcal{N}_I - n\sigma(I)}{\sqrt{\log n}} \rightarrow Z \quad \text{centered Gaussian}$$

extreme eigenvalue statistics

$$\lambda_{\max} \rightarrow a$$

asymptotics

$$\frac{1}{n} \mathcal{N}_I = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\lambda_i \in I\}} \rightarrow \sigma(I)$$

σ limiting spectral measure compactly supported $(-a, +a)$

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

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

central limit theorem : I interval $\subset (-a, +a)$

$$\frac{\mathcal{N}_I - n\sigma(I)}{\sqrt{\log n}} \rightarrow Z \text{ centered Gaussian}$$

extreme eigenvalue statistics

$$\lambda_{\max} \rightarrow a$$

$$n^{2/3}[\lambda_{\max} - a] \rightarrow F_{TW}$$

F_{TW} **Tracy-Widom** distribution

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

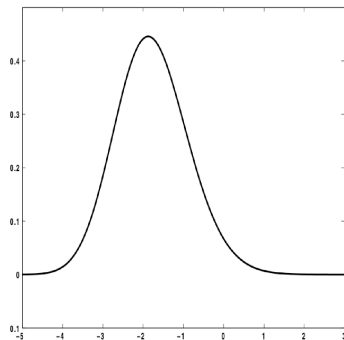
$$u'' = 2u^3 + xu \quad \text{Painlevé II equation}$$

(Gaussian Unitary Ensemble model)

mean $\simeq -1.77$

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

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



density (GUE)

finite range (exponential) inequalities ?

variance bounds ?

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

$$\lambda_{\max} \text{ Lipschitz : } \text{Var}(\lambda_{\max}) = O\left(\frac{1}{n}\right)$$

Tracy-Widom asymptotics

$$\text{Var}(\lambda_{\max}) = O\left(\frac{1}{n^{4/3}}\right)$$

known for Gaussian models (GUE determinantal)

families of Wigner matrices

L. Erdős, H.-T. Yau and **T. Tao, V. Vu 2011**

localization and four moment theorem

exponential tail inequalities : in progress

finite range (exponential) inequalities?

variance bounds?

eigenvalue counting function $\mathcal{N}_I = \sum_{i=1}^n \mathbf{1}_{\{\lambda_i \in I\}}$

I interval $\subset (-a, +a)$

$$\text{Var}(\mathcal{N}_I) = O(\log n)$$

asymptotics

$$\frac{1}{n} \mathcal{N}_I = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\lambda_i \in I\}} \rightarrow \sigma(I)$$

σ limiting spectral measure compactly supported $(-a, +a)$

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

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

central limit theorem : I interval $\subset (-a, +a)$

$$\frac{\mathcal{N}_I - n\sigma(I)}{\sqrt{\log n}} \rightarrow Z \text{ centered Gaussian}$$

finite range inequalities ?

variance bounds ?

eigenvalue counting function $\mathcal{N}_I = \sum_{i=1}^n \mathbf{1}_{\{\lambda_i \in I\}}$

I interval $\subset (-a, +a)$

$$\text{Var}(\mathcal{N}_I) = O(\log n)$$

known for Gaussian models (GUE determinantal)

families of Wigner matrices **S. Dallaporta, V. Vu 2011**

L. Erdős, H.-T. Yau and **T. Tao, V. Vu 2011**

localization and four moment theorem

exponential tail inequalities : in progress

invertibility of random matrices

concentration of singular values

M. Rudelson, R. Vershynin 2007-11

T. Tao, V. Vu 2009

Littlewood-Offord problem

circular law

compressed sensing

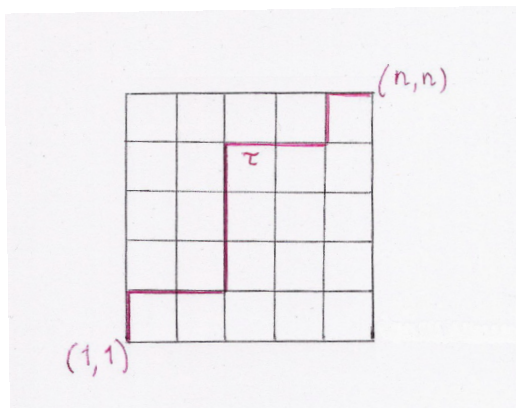
restricted isometric property

Example 3 : oriented last passage percolation

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

$$W_n = \max_{\tau} \sum_{(i,j) \in \tau} w_{ij}$$

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



$$W_n = \max_{\tau} \sum_{(i,j) \in \tau} w_{ij}$$

W_n $2n$ -Lipschitz (in the w_{ij} 's)

w_{ij} Gaussian or Bernoulli

$$\mathbb{P}\left(|W_n - \mathbb{E}(W_n)| \geq r\right) \leq C e^{-r^2/Cn}, \quad r \geq 0$$

$\text{Var}(W_n) = O(n)$ Gaussian behavior

however (w_{ij} exponential)

$\text{Var}(W_n) = O(n^{2/3})$ random matrix behavior

$$\text{Var}(W_n) = O\left(\frac{n}{\log n}\right)$$

I. Benjamini, G. Kalai, O. Schramm 2003, S. Chatterjee 2008

(Gaussian, hypercontractivity)

Example 4 : combinatorial models

from the cube $\{0, 1\}^n$ to non-product models

S^n symmetric group over n objects

$$d(x, y) = \# \{1 \leq i \leq n; x(i) \neq y(i)\}, \quad x, y \in S^n$$

$$\mu \text{ uniform} \quad \mu(\{x\}) = \frac{1}{n!}$$

B. Maurey 1979

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

martingale method , modified logarithmic Sobolev inequality

M. Talagrand 1996 convex distance version

what about more general

Cayley graphs, discrete models ?

probabilistic and combinatorial graph structures

martingale, Markov chain methods

couplings

functional inequalities

notions of curvature on discrete spaces

Y. Ollivier 2008, J. Maas 2011

optimal transportation tools

Example 5 : log-concave measures

geometric analysis

Kannan-Lovasz-Simonovits conjecture

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

covariance $\mu = \text{Id}$

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

C independent of n

E. Milman 2009

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

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

central limit theorem for log-concave measures

B. Klartag 2007

conjecture :

$$\text{Var}(|x|) \leq C$$

best known result :

$$\text{Var}(|x|) \leq C n^{2/3}$$

O. Guédon, E. Milman 2010

large deviation estimate **G. Paouris 2006**

better results for unconditional models or ψ_α