Selected topics around the concentration of measure

DC

Learning (and) statistics with Talagrand
Journées de Statistique Mathématique 2026
January 7–9, 2026 – Orsay

Contents

1 High-dimensional convex bodies and concentration

- Lindeberg CLT principle
  Historically via coupling, today sums of martingale increments
  Condition is that total variance well spread for CLT convolutional effect
- Milman–Talagrand concentration principle = quantitative and nonlinear version
- Actually the concentration phenomenon is at the basis of statistical physics!
- The concentration phenomenon is important for high-dimensional stochastic models, in particular
  - statistical mechanics
  - randomized algorithms and randomized combinatorial optimization: TSP !
  - high-dimensional geometry including convex bodies and vectors/matrices/tensors
- Simplest instance is second moment method

1.1 Archimedes on sphere and cylinder

Theorem 1.1. Maxwell: geometric characterization of isotropic Gaussians.

In $\mathbb{R}^n$, $n \geq 2$, a law is at the same time product and rotationally invariant iff it is $\mathcal{N}(0,\sigma^2 I_n)$, $\sigma \geq 0$.

At the origin of kinetic gas theory in statistical physics, before Boltzmann, maybe known to Herschel.

Proof. We reduce to smooth density $f > 0$ by regularizing by convolution with $\mathcal{N}(0,\varepsilon I_n)$. Then $\log f(x) = h_1(x_1) + \cdots + h_n(x_n) = g(|x|^2)$ gives $h_i'(x_i) = 2g'(|x|^2)x_i$, and since $n \geq 2$, using $i \neq j$ forces g' to be constant. □

Theorem 1.2. Mehler: Euclidean spheres and Gaussians.

(i) $Z \sim \mathcal{N}(0,I_n)$ iff $Z/|Z|$ and $|Z|$ are independent with $Z/|Z| \sim \mathrm{Unif}(\mathbb{S}^{n-1})$ and $|Z|^2 \sim \chi^2(n)$.

(ii) If $X \sim \mathrm{Unif}(\sqrt{n}\mathbb{S}^{n-1})$, then $\mathrm{proj}_{\mathbb{R}^k}(X) = (X_1,\ldots,X_k) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,I_k)$, for all fixed $k \geq 1$.

(iii) If $X \sim \mathrm{Unif}(\sqrt{n}\mathbb{S}^{n-1})$, then $\langle X,\theta\rangle \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$ for all choices of $\theta \in \mathbb{S}^{n-1}$.

Diaconis and Freedman, and Stroock, discovered that this was due to Mehler (1866), not to Poincaré or Borel.

Proof.

(i) Spherical coordinates: $\mathrm{e}^{-\frac{|x|^2}{2}}\mathrm{d}x = r^{n-1}\mathrm{e}^{-\frac{r^2}{2}}\mathrm{d}r\mathrm{d}\sigma$.

(ii) We have $X \overset{\mathrm{d}}{=} \sqrt{n}Z/|Z|$, $Z \sim \mathcal{N}(0, I_n)$ by (i), but $(Z_1,\dots,Z_k) \sim \mathcal{N}(0, I_k)$ while $|Z|/\sqrt{n} \xrightarrow[n\to\infty]{} 1$ in probability by the LLN. Therefore $\sqrt{n}(Z_1,\dots,Z_k)/|Z| \xrightarrow[n\to\infty]{} \mathcal{N}(0, I_k)$ in law by the Slutsky lemma. Alternatively, we can couple the uniform laws on spheres by an infinite sequence $Z_1, Z_2, \dots$ of iid $\mathcal{N}(0,1)$, and the SLLN to get $|Z|/\sqrt{n} \xrightarrow[n\to\infty]{} 1$ a.s. therefore $\sqrt{n}(Z_1,\dots,Z_k)/|Z| \xrightarrow[n\to\infty]{} (Z_1,\dots,Z_k)$ a.s. hence in law.

(iii) By rotational invariance, we can take $\theta = e_1$, and the result follows then from (ii) with $k=1$.

□

Corollary 1.3. High-dimensional concentration around equators and orthogonality.

(i) If $X \sim \mathrm{Unif}(\sqrt{n}\mathbb{S}^{n-1})$, then for all $r \geq 0$ and all choices of $\theta \in \mathbb{S}^{n-1}$, denoting $H_\theta := (\mathbb{R}\theta)^\perp$,

$$\mathbb{P}(\mathrm{dist}_{\mathbb{R}^n}(X, H_\theta) \geq r) \xrightarrow[n\to\infty]{} \mathbb{P}(|Z| \geq r) \leq \mathrm{e}^{-\frac{r^2}{2}}, \quad Z \sim \mathcal{N}(0,1).$$

(ii) If $X, Y \sim \mathrm{Unif}(\mathbb{S}^{n-1})$ are independent, then for all $r \geq 0$,

$$\mathbb{P}(\sqrt{n}|\langle X, Y\rangle| \geq r) \xrightarrow[n\to\infty]{} \mathbb{P}(|Z| \geq r) \leq \mathrm{e}^{-\frac{r^2}{2}}, \quad Z \sim \mathcal{N}(0,1).$$

Proof.

(i) $\mathrm{dist}_{\mathbb{R}^n}(X, H_\theta) = |\langle X, \theta\rangle|$, the law does not depend on θ. The equator is $E_\theta := H_\theta \cap \sqrt{n}\mathbb{S}^{n-1}$.

(ii) By Fubini–Tonelli, independence of X and Y, and rotational invariance of the uniform law on the sphere

$$\mathbb{P}(|\langle\sqrt{n}X, Y\rangle| \geq r) = \int\Big(\int \mathbf{1}_{|\langle\sqrt{n}x, y\rangle| \geq r}\mathbb{P}_X(\mathrm{d}x)\Big)\mathbb{P}_Y(\mathrm{d}y) = \int \mathbb{P}(|\langle\sqrt{n}X, y\rangle| \geq r)\mathbb{P}_Y(\mathrm{d}y) = \mathbb{P}(|\langle\sqrt{n}X, e_1\rangle| \geq r).$$

Alternatively, $(X, Y) = (\frac{Z}{|Z|}, \frac{Z'}{|Z'|})$, Z and Z' independent $\sim \mathcal{N}(0, I_n)$, and by CLT, LLN, and Slutsky lemma,

$$\sqrt{n}\langle X, Y\rangle = \sqrt{n}\frac{\sum_{i=1}^n Z_i Z_i'}{(\sqrt{n}(1+o(1)))^2} = \frac{\sum_{i=1}^n Z_i Z_i'}{\sqrt{n}}(1+o(1)) \xrightarrow[n\to\infty]{} \mathcal{N}(0,1).$$

□

- Isoperimetric inequality for uniform distribution μ on $\mathbb{S}^{n-1}$
  Discovered by Paul Lévy, Erhard Schmidt independently, inspired by isoperimetry for Lebesgue
  For all $A \subset \mathbb{S}^{n-1}$, $\mu(A + B(0,r)) \geq \mu(C + B(0,r))$ for all $r \geq 0$, where C is spherical cap with $\mu(C) = \mu(A)$.
  Generalized to positively curved Riemannian manifolds by Mikhaïl Gromov
  Revisited and generalized to infinite dimension by Dominique Bakry and Michel Ledoux
- Gaussian isoperimetry for $\mu = \mathcal{N}(0, I_k)$ on $\mathbb{R}^k$
  For all $A \subset \mathbb{R}^k$, $\mu(A + B(0,r)) \geq \mu(H + B(0,r))$ for all $r \geq 0$, where H is a half-space with $\mu(H) = \mu(A)$.
  Formulated by Vladimir Sudakov and Boris Tsirelson, Christer Borell independently
  Functional form by Sergey Bobkov, revisited and generalized by Dominique Bakry and Michel Ledoux

Actually the law of the projections can be computed explicitly.

Lemma 1.4. Spherical projections.

(i) If $X = (X_1, \dots, X_n) \sim \mathrm{Unif}(\mathbb{S}^{n-1})$, then $|(X_1,\dots,X_k)|^2 = X_1^2 + \cdots + X_k^2 \sim \mathrm{Beta}\big(\frac{k}{2}, \frac{n-k}{2}\big)$ for all $1 \leq k \leq n-1$.

(ii) If $X \sim \mathrm{Unif}(\mathbb{S}^{n-1})$, then $\langle X, \theta\rangle \sim \mathrm{Beta}_{[-1,1]}(\frac{n-1}{2}, \frac{n-1}{2})$ with density $\propto (1-x^2)_+^{\frac{n-3}{2}}$, for all $\theta \in \mathbb{S}^{n-1}$.

(iii) If $X \sim \mathrm{Unif}(\mathbb{S}^{n-1})$, then $\mathrm{proj}_{\mathbb{R}^k}(X)$ follows the multivariate Beta with density $\propto (1-|x|^2)_+^{\frac{n-k-2}{2}}$.

- (ii) : arcsine if $n=2$, uniform if $n=3$, semicircle if $n=4$.
  (ii) : we recover $\sqrt{n}\langle X,\theta\rangle \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$ via $(1-\frac{x^2}{n})^{\frac{n-3}{2}} \xrightarrow[n\to\infty]{} \mathrm{e}^{-\frac{1}{2}x^2}$.
  (ii) : the interpretation with spherical harmonics is known as the Funk–Hecke formula.

- (iii) : case $k=1$ (projection on a diameter) gives (ii) by rotational invariance
  (iii) : we recover $\mathrm{proj}_{\mathbb{R}^k}(X) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,I_k)$ via $(1-\frac{|x|^2}{n})^{\frac{n-k-2}{2}} \xrightarrow[n\to\infty]{} \mathrm{e}^{-\frac{1}{2}|x|^2}$.
  (iii) : multivariate Beta or Barenblatt profile (nonlinear PDE : porous media equation $\partial_t u = \Delta(u^m)$, $m>1$)

Proof.

(i) From $X \overset{\mathrm{d}}{=} \frac{Z}{|Z|}$, $Z\sim\mathcal{N}(0,I_n)$, we get

$$X_1^2+\cdots+X_k^2 \overset{\mathrm{d}}{=} \frac{Z_1^2+\cdots+Z_k^2}{(Z_1^2+\cdots+Z_k^2)+(Z_{k+1}^2+\cdots+Z_n^2)} = \frac{\chi^2(k)}{\chi^2(k)+\chi^2(n-k)} = \mathrm{Beta}\Big(\frac{k}{2},\frac{n-k}{2}\Big)$$

More generally, marginal of Dirichlet distribution : $\frac{\mathrm{Gamma}(a,1)}{\mathrm{Gamma}(a,1)+\mathrm{Gamma}(b,1)} = \mathrm{Beta}(a,b)$.

(ii) By rotational invariance and (i) $k=1$, $\langle X,\theta\rangle^2 \overset{\mathrm{d}}{=} \langle X,e_1\rangle^2 = X_1^2 \sim \mathrm{Beta}(\frac{1}{2},\frac{n-1}{2})$, then $\sqrt{\cdot}$ and symmetrize.

(iii) Same as for (ii) using rotational symmetrization.

□

Theorem 1.5. Archimedes sphere and cylinder : direct and reverse form.

(i) If $X\sim\mathrm{Unif}(\mathbb{S}^{n-1})$, $n\geq 3$, then $\mathrm{proj}_{\mathbb{R}^{n-2}}(X)\sim\mathrm{Unif}(\mathbb{B}^{n-2})$.

(ii) If $Z\sim\mathcal{N}(0,I_n)$, $n\geq 1$, and $E\sim\mathrm{Exp}(1)$ are independent, then $\frac{Z}{\sqrt{|Z|^2+2E}}\sim\mathrm{Unif}(\mathbb{B}^n)$.

- (i) we cannot replace $n-2$ by $n-k$ for $k\neq 2$.
  (i) $n=3$ says that the projection on a diameter is uniform. Geometrically, it is equivalent to Archimedes historical result : if we place a sphere in a tight cylinder then the surfaces are the same, and this remains the case for the surface between arbitrary parallel planes orthogonal to the cylinder. This allowed Archimedes to compute the surface of the sphere. His method is a precursor of infinitesimal calculus.

- Do not confuse with Archimedes principle and Eurêka.

- Archimedes of Syracuse (-287 – -212) was so proud of this discovery that the picture of it was engraved on his tombstone. This helped his admirer Cicero (-106 – -43) to identify it in -75, 150 years after his murder by a Roman soldier during the siege of Syracuse.

- Both sides of the Fields Medal are devoted to this theorem! Taught in high schools before modern maths.

- (ii) Useful for simulation and proofs, extends to $\mathbb{B}_p^n$ via $\propto \mathrm{e}^{-t^p}$ (Barthe–Guédon–Mendelson–Naor 2005).

Proof.

(i) The law of $Y := (X_1,\ldots,X_{n-2})$ is supported in $\mathbb{B}^{n-2}$ and is rotationally invariant. Next, the lemma gives $|Y|^2\sim\mathrm{Beta}(\frac{n-2}{2},1)$, with density $\propto r^{\frac{n-4}{2}}$, hence $|Y|$ has density $\propto r(r^2)^{\frac{n-4}{2}} = r^{n-3}$.

(ii) Follows from (i) used for dimension $n+2$ and $Z_{n+1}^2+Z_{n+2}^2\sim(\chi^2(1))^{*2} = \mathrm{Gamma}(\frac{2}{2},\frac{1}{2}) = \mathrm{Exp}(\frac{1}{2})$.

□

Corollary 1.6. CLT for the uniform distribution on the ball.

If $X\sim\mathrm{Unif}(\sqrt{n}\mathbb{B}^n)$ then $\langle X,\theta\rangle \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$ for all $\theta\in\mathbb{S}^{n-1}$.

- The analogue of the CLT for spheres.
  It is probably the most accessible CLT for a non-product convex body.
  Enriches the CLT already obtained: spheres (non-convex), and cube (product).

- All directions are the same due to rotational invariance. This CLT for the convex body $\mathbb{B}^n$ is valid for all directions, while for certain convex bodies such as the cube, certain directions are impossible for the CLT.
- In particular, for $\theta = e_1$, we get $X_1 \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$, while for $\theta = (\frac{1}{\sqrt{n}},\dots,\frac{1}{\sqrt{n}})$, we get $\frac{X_1+\cdots+X_n}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1)$.

Proof. By the reverse Archimedes theorem, $X \stackrel{\mathrm{d}}{=} \sqrt{n}Z/\sqrt{|Z|^2+2E}$ where $Z \sim \mathcal{N}(0,I_n)$ and $E \sim \mathrm{Exp}(1)$ are independent, but $\langle Z,\theta\rangle \sim \mathcal{N}(0,|\theta|^2) = \mathcal{N}(0,1)$ while $|Z|^2+2E = n(1+o(1))$ by the LLN. □

$$\boxed{\mathrm{Unif}(\sqrt{n}\mathbb{B}^n) \approx \mathrm{Unif}(\sqrt{n}\mathbb{S}^{n-1}) \approx \mathcal{N}(0,I_n) \approx \mathrm{Unif}(\{-1,1\}^n)}$$

Figure 1: Equivalences in high dimension n. Useful for modelling and in particular spin systems in statistical mechanics. Geometrically, the sphere with the uniform law behaves like a convex set, and a discrete cube.

Log-concave probability measure = Boltzmann–Gibbs measures with convex energy $\propto \mathrm{e}^{-V}$
Functional generalization or relaxation of uniform distribution on convex bodies $\frac{\mathbf{1}_K}{|K|} \propto \mathrm{e}^{-\infty \mathbf{1}_{K^c}}$

Theorem 1.7. Klartag : CLT for convex bodies or log-concave measures.

If X is log-concave with $\mathbb{E}(X) = 0$ and $\mathrm{Cov}(X) = I_n$, then there exists $\varepsilon_n \searrow 0$ and $\delta_n \searrow 0$ and a measurable subset $\Theta_n \subset \mathbb{S}^{n-1}$ such that $|\Theta_n| \geq (1-\delta_n)|\mathbb{S}^{n-1}|$ and $\sup_{\theta\in\Theta_n} \|\mathrm{Law}(\langle X,\theta\rangle) - \mathcal{N}(0,1)\|_{\mathrm{TV}} \leq \varepsilon_n$.

For the cube $X \sim \mathrm{Unif}([-\sqrt{3},\sqrt{3}]^n)$, the directions $\theta \in \{\pm e_1,\dots,\pm e_n\}$ are obviously impossible.

About the proof. Major achievement of probabilistic and geometric functional analysis. Show that $|X|$ is concentrated around $\sqrt{n}$ (thin-shell phenomenon) and then use the CLT for $\sqrt{n}\mathbb{S}^{n-1}$. The equivalence between thin-shell bounds and the Gaussian approximation property of typical marginal distributions goes back to Vladimir Sudakov, Persi Diaconis and David Freedman, and Keith Ball, among others. □

1.2 Thin-shell phenomenon

Theorem 1.8. Thin-shell phenomenon.

If $X_1,\dots,X_n$ are iid real random variables with $m_1 = \mathbb{E}(X_1) = 0$, $m_2 = \mathbb{E}(X_1^2) = 1$, and $m_4 = \mathbb{E}(X_1^4) < \infty$, then

$$|X| - \sqrt{n} = \sqrt{n}\Big(\frac{|X|}{\sqrt{n}} - 1\Big) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0, \tfrac{m_4-1}{4}) \quad \text{where} \quad X := (X_1,\dots,X_n).$$

Proof. By the LLN, $\frac{|X|^2}{n} \to m_2 = 1$ a.s. hence $|\frac{|X|}{\sqrt{n}} - 1| \to 0$ a.s. By the CLT for $(X_i^2)_{i\geq 1}$, $\sqrt{n}\big(\frac{|X|^2}{n} - 1\big) \to \mathcal{N}(0, m_4-1)$ in law, and then, by the delta method for $\sqrt{\cdot}$ at point 1, we get $\sqrt{n}\big(\frac{|X|}{\sqrt{n}} - 1\big) \to \mathcal{N}(0, \frac{m_4-1}{4})$ in law. □

- How about $X \sim \mathrm{Unif}(K)$ for a convex body K, $\mathbb{E}X = 0$ and $\mathrm{Cov}(X) = \mathbb{E}XX^\top = I_n$ (aka isotropic position).
- For all $1 \leq p \leq \infty$, $X \sim \mathrm{Unif}(\mathbb{B}_p^n)$ has the symmetries of the cube, so $\mathbb{E}(X) = 0$, $\mathrm{Cov}(X) = \mathbb{E}(X_1^2)I_n = \frac{\mathbb{E}(|X|^2)}{n} I_n$.
  The convex body $sB_p^n = B_p^n(s)$ with $s = 1/\sqrt{\mathbb{E}(X_1^2)} = \sqrt{n/\mathbb{E}(|X|^2)}$ is isotropic.
- The cube $\mathbb{B}_\infty(\sqrt{3}) = (\sqrt{3}[-1,1])^n$ is isotropic.
  Product convex set. Components of X are iid $\mathrm{Unif}(\sqrt{3}[-1,1])$, $m_1 = 0$, $m_2 = 1$, $m_4 = \frac{9}{5}$, $\sqrt{\frac{m_4-1}{4}} = \frac{1}{\sqrt{5}} \approx 0.45$.
  The extremal points $\{\pm\sqrt{3}\}^n$ have norm $\sqrt{3n}$.
- The ball $\mathbb{B}^n = \mathbb{B}_2^n(1)$, non-product convex body, the components of X are dependent. The previous theorem does not apply but we can explore the phenomenon. Since $|X|$ has density $r \in [0,1] \mapsto nr^{n-1}$, with second moment $\frac{n}{n+2}$, we get $\mathbb{E}(|X|^2) = \frac{n}{n+2} \xrightarrow[n\to\infty]{} 1$, and $\sqrt{n}X$ is isotropic in high-dimension. Moreover

$$\mathbb{E}(|X|) = \frac{n}{n+1} \quad \text{and} \quad \mathrm{Var}(|X|) = \frac{n}{n+2} - \Big(\frac{n}{n+1}\Big)^2 = \frac{n}{(n+2)(n+1)^2} = O\Big(\frac{1}{n^2}\Big),$$

  thus the theorem above extends : in high-dimension $\mathrm{Unif}(\sqrt{n}\mathbb{B}^n)$ is concentrated around $\sqrt{n}\mathbb{S}^{n-1}$.

Theorem 1.9. Thin-shell : Klartag-Lehec 2025.

If X is log-concave with $\mathbb{E}(X)=0$ and $\mathrm{Cov}(X)=I_n$ then $\mathbb{E}((|X|-\sqrt{n})^2)\le C$ where C is universal.

Boaz Klartag → Vitali Milman, proof of the Dvoretzky theorem by concentration of measure.
Lehec → Bernard Maurey, proofs based on stochastic calculus in convex geometry and functional analysis.

About the proof. Major achievement of probabilistic and geometric functional analysis. Relies crucially on stochastic calculus for interpolation (Eldan stochastic localization), and on coupling, among other tools. □

If $\mathrm{Var}(|X|^2)\le Cn$ for a universal constant $C>0$ then $\mathbb{E}((|X|-\sqrt{n})^2)\le C$. Indeed,

$$\mathbb{E}((|X|-\sqrt{n})^2)\le \mathbb{E}\Big((|X|-\sqrt{n})^2\Big(\frac{|X|+\sqrt{n}}{\sqrt{n}}\Big)^2\Big)\le \frac{\mathbb{E}((|X|^2-n)^2)}{n}=\frac{\mathrm{Var}(|X|^2)}{n}\le C.$$

If X satisfies Poincaré for $|\cdot|^2$ then by isotropy $\mathrm{Var}(|X|^2)\le C\mathbb{E}(|X|^2)=Cn$. KLS conjecture on Poincaré.

1.3 Sudakov–Tsirelson or Borell theorem

Theorem 1.10. Sudakov–Tsirelson or Borell : concentration for Lipschitz functions.

If $X\sim\mathcal{N}(0,I_n)$ and $f:\mathbb{R}^n\to\mathbb{R}$, then $\log\mathbb{E}\mathrm{e}^{\theta f}\le\frac{\theta^2}{2}\|f\|_{\mathrm{Lip}}^2+\theta\mathbb{E}f(X)$ for all $\theta\in\mathbb{R}$, in particular

$$\mathbb{P}(f(X)\le\mathbb{E}f(X)-r\|f\|_{\mathrm{Lip}})\le\mathrm{e}^{-\frac{r^2}{2}}\quad\text{and}\quad\mathbb{P}(f(X)\ge\mathbb{E}f(X)+r\|f\|_{\mathrm{Lip}})\le\mathrm{e}^{-\frac{r^2}{2}},\quad r\ge0.$$

Proof. We reduce first to $\|f\|_{\mathrm{Lip}}=1$ and $\mathbb{E}f(X)=0$ by dilation and translation. The concentration inequalities come then from the exponential Markov inequality $\mathbb{P}(\pm f(X)\ge r)\le\inf_{\theta>0}\mathrm{e}^{-\theta r}\mathbb{E}\mathrm{e}^{\pm\theta f}\le\mathrm{e}^{-\sup_{\theta\ge0}(\theta r-\frac{\theta^2}{2})}=\mathrm{e}^{-\frac{r^2}{2}}$.

To prove the Laplace bound $\log L(\theta):=\log\mathbb{E}\mathrm{e}^{\theta f(X)}\le\frac{\theta^2}{2}$, we can assume that $\theta>0$, reduce to smooth f by approximation and Rademacher theorem, so $\|f\|_{\mathrm{Lip}}:=\sup_{x\neq y}\frac{|f(x)-f(y)|}{|x-y|}=\||\nabla f|\|_\infty$, and then rely on the covariance representation with f and $g=\mathrm{e}^{\theta f}$, $\theta\ge0$, using the fact that Y_α has the law of X for all α, namely

$$\begin{aligned}\mathbb{E}f(X)\mathrm{e}^{\theta f(X)}&=\theta\int_0^1\mathbb{E}\langle\nabla f(X_\alpha),\nabla f(Y_\alpha)\rangle\mathrm{e}^{\theta f(Y_\alpha)}\mathrm{d}\alpha\\&\le\theta\int_0^1\mathbb{E}|\nabla f(X_\alpha)||\nabla f(Y_\alpha)|\mathrm{e}^{\theta f(Y_\alpha)}\mathrm{d}\alpha\\&\le\theta\int_0^1\mathbb{E}\mathrm{e}^{\theta f(Y_\alpha)}\mathrm{d}\alpha=\theta\mathbb{E}\mathrm{e}^{\theta f(X)},\end{aligned}$$

which gives the differential inequality $L'(\theta)\le\theta L(\theta)$ hence $\log\mathbb{E}\mathrm{e}^{\theta f}\le\frac{\theta^2}{2}$. □

Theorem 1.11. Houdré : covariance representation.

If $X\sim\mathcal{N}(0,I_n)$, $f,g:\mathbb{R}^n\to\mathbb{R}$ with $\||\nabla f|\|_\infty<\infty$ and $\||\nabla g|\|_\infty<\infty$, then

$$\mathbb{E}f(X)g(X)-\mathbb{E}f(X)\mathbb{E}g(X)=\int_0^1\mathbb{E}\langle(\nabla f)(X_\alpha),(\nabla g)(Y_\alpha)\rangle\mathrm{d}\alpha\quad\text{where}\quad\begin{pmatrix}X_\alpha\\Y_\alpha\end{pmatrix}\sim\mathcal{N}\Big(0,\begin{pmatrix}I_n&\alpha I_n\\\alpha I_n&I_n\end{pmatrix}\Big).$$

Following Dominique Bakry and Michel Ledoux, we could use alternatively $(X_\alpha,Y_\alpha):=(X,\alpha X+\sqrt{1-\alpha^2}Y)$, Y independent copy of X, $\alpha=\mathrm{e}^{-t}$, in relation with the Mehler formula for Ornstein–Uhlenbeck process, which suggests extensions to non-negatively curved manifolds and infinite dimensional Markov diffusion operators.

Proof. The vectors $X_1(=Y_1)$ and X have same law, while X_0 and Y_0 are independent with same law as X. Thus

$$\mathbb{E}f(X)g(X)-\mathbb{E}f(X)\mathbb{E}g(X)=\mathbb{E}f(X_1)g(Y_1)-\mathbb{E}f(X_0)g(Y_0)=\int_0^1\partial_\alpha\mathbb{E}f(X_\alpha)g(Y_\alpha)\mathrm{d}\alpha.$$

By approximation and bilinearity, it suffices to consider the case of trigonometric monomials, namely characteristic functions : $f(x)=\mathrm{e}^{\mathrm{i}\langle u,x\rangle}$ and $g(x)=\mathrm{e}^{\mathrm{i}\langle v,x\rangle}$, $u,v\in\mathbb{R}^n$. In this case

$$\mathbb{E}f(X_\alpha)g(Y_\alpha)=\exp\Big(-\frac{1}{2}\Big(|u|^2+2\alpha\langle u,v\rangle+|v|^2\Big)\Big).$$

Now it simply remains to note, using $\nabla f(x) = \mathrm{i}u\mathrm{e}^{\mathrm{i}\langle u,x\rangle}$ and $\nabla g(x) = \mathrm{i}v\mathrm{e}^{\mathrm{i}\langle v,x\rangle}$, that

$$\partial_\alpha \mathbb{E} f(X_\alpha)g(Y_\alpha) = -\langle u,v\rangle \mathbb{E} f(X_\alpha)g(Y_\alpha) = \mathbb{E}\langle \nabla f(X_\alpha), \nabla g(Y_\alpha)\rangle.$$

□

Theorem 1.12. Ibragimov, Sudakov, and Tsirelson : concentration via stochastic calculus.

If $X \sim \mathcal{N}(0, I_n)$ and $f : \mathbb{R}^n \to \mathbb{R}$ with $\||\nabla f|\|_\infty \le 1$ then $f(X) - \mathbb{E}f(X) \overset{\mathrm{d}}{=} B_T$ with B standard real Brownian motion and T a stopping time such that $T \le 1$. In particular, by the reflection principle,

$$\mathbb{P}(f(X) - \mathbb{E}f(X) \ge r) = \mathbb{P}(B_T \ge r) \le \mathbb{P}\Big(\sup_{0\le t\le 1} B_t \ge r\Big) = \mathbb{P}(|B_1| \ge r), \quad r \ge 0.$$

The affine case $f(x) = \langle a, x\rangle + b$, $|a| = 1$, shows that this bound is in fact optimal.

Proof. Reminds the Skorokhod embedding theorem, may provide T, but how to get $T \le 1$? Let $(W_s)_{s\in[0,1]}$ be a standard BM on $\mathbb{R}^n$, and let us consider the martingale $M_s = \mathbb{E}(f(W_1) \mid \mathcal{F}_s)$, $s \in [0,1]$. Then $M_0 = \mathbb{E}f(X)$, while M_1 has the law of $f(X)$. By the Dambis–Dubins–Schwarz theorem, there exists a real BM B such that $(B_{\langle M\rangle_s})_{0\le s\le 1}$ and $(M_s - M_0)_{0\le s\le 1}$ have same law, in particular $M_1 - M_0 = f(X) - \mathbb{E}f(X)$ has the law of B_T with $T := \langle M\rangle_1$.

It remains to show that $\langle M\rangle_1 \le 1$. Let $(P_s)_{s\in[0,1]}$ be the heat semigroup $P_s(f)(x) = \mathbb{E}(f(W_s) \mid W_0 = x)$. Then, by the Markov property, we get $M_s = \mathbb{E}(f(W_1) \mid W_s) = P_{1-s}(f)(W_s)$. Next, the Itô formula gives

$$M_t = M_0 + \int_0^t \nabla P_{1-s}(f)(W_s)\cdot \mathrm{d}W_s, \quad \text{hence} \quad \langle M\rangle_t = \int_0^t |\nabla P_{1-s}(f)|^2(W_s)\mathrm{d}s.$$

Now $\nabla P_{1-s}(f) = P_{1-s}\nabla f$, thus $|\nabla P_{1-s}(f)| \le P_{1-s}|\nabla f|$, hence $\||\nabla P_{1-s}(f)|\|_\infty \le 1$. □

2 Talagrand transportation inequality

The Kullback–Leibler divergence or relative entropy on $\mathcal{P}(E)$ is defined by

$$\mathrm{H}(\nu \mid \mu) := \begin{cases} \int f \log f \mathrm{d}\mu \in [0,+\infty] & \text{if } \nu \ll \mu,\ f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \\ +\infty & \text{otherwise} \end{cases}.$$

Makes sense since $u \in \mathbb{R}_+ \mapsto u\log(u)$ is convex, and since it is strictly convex, we get $\mathrm{H}(\nu \mid \mu) = 0$ iff $\nu = \mu$.

Lemma 2.1. Legendre duality for H as a convex function of f.

$$\mathrm{H}(\nu \mid \mu) = \sup_{\substack{g \\ \mathrm{e}^g \in L^1(\mu)}} \Big(\int g\mathrm{d}\nu - \log\int \mathrm{e}^g \mathrm{d}\mu\Big) \quad \text{and} \quad \log\int \mathrm{e}^g\mathrm{d}\mu = \sup_{\substack{\nu \\ \mathrm{H}(\nu|\mu)<\infty}} \Big(\int g\mathrm{d}\nu - \mathrm{H}(\nu \mid \mu)\Big)$$

and these suprema are achieved for $g = \log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ and $\mathrm{d}\nu = \dfrac{\mathrm{e}^g}{\int \mathrm{e}^g \mathrm{d}\mu}\mathrm{d}\mu$ respectively.

Also known as Donsker–Varadhan (large deviations) or Gibbs (statistical mechanics) variational formula.

2.1 Bobkov–Götze theorem

Let $\mathcal{P}_p(E)$ be the set of probability measures μ on the Polish space (E,d) with finite moment of order $p \ge 1$:

$$\int d(x_0,x)^p \mathrm{d}\mu(x) < \infty \quad \text{for some and thus all } x_0 \in E.$$

The Wasserstein or Monge–Kantorovich or coupling or transportation cost distance on $\mathcal{P}_p(E)$ is[1]

$$\mathrm{W}_p(\nu,\mu) := \Big(\inf_{\pi\in\Pi(\mu,\nu)} \iint_{E\times E} d(x,y)^p \pi(\mathrm{d}x,\mathrm{d}y)\Big)^{1/p} = \Big(\inf_{\substack{(X,Y) \\ X\sim\mu, Y\sim\nu}} \mathbb{E}(d(X,Y)^p)\Big)^{1/p}, \quad \mu,\nu \in \mathcal{P}_p(E),$$

[1] It can be shown that W_p is indeed a distance on $\mathcal{P}_p(E)$, and that $\mathrm{W}_p(\mu_n,\mu) \to 0$ iff $\mu_n \to \mu$ narrowly and in the sense of moments up to order p. Studied by Leonid Vitalyevich Kantorovich (1912 – 1986), Nobel prize in Economics 1975, remarkable for a Soviet mathematician, but also Cédric Villani (1973 –), Fields medalist, as a relaxation of the optimal transport problem of Gaspard Monge (1746 – 1818). Major contributions by Yann Brenier (1957 –), Luis Caffarelli (1948 –), Alessio Figalli (1984 –).

where $\Pi(\mu,\nu)$ is the set of probability measures on $E\times E$ with marginals μ and ν. It is convex and contains $\mu\otimes\nu$.

The following lemma makes use of the Polish assumption.

Lemma 2.2. Kantorovich–Rubinstein duality.

$$\mathrm{W}_1(\nu,\mu)=\sup_{\|f\|_{\mathrm{Lip}}\leq 1}\Big(\int f\mathrm{d}\nu-\int f\mathrm{d}\mu\Big)\quad\text{where}\quad \|f\|_{\mathrm{Lip}}:=\sup_{x\neq y}\frac{|f(x)-f(y)|}{d(x,y)}.$$

Theorem 2.3. Bobkov–Götze: Talagrand transportation inequality for W_1.

For all $\mu\in\mathscr{P}_1(E)$ and constant $c>0$, the following properties are equivalent:

(i) Sub-Gaussian bound on Laplace transform of Lipschitz functions:

$$\log\int \mathrm{e}^{\theta f}\mathrm{d}\mu\leq\theta^2\frac{c}{4}\|f\|^2_{\mathrm{Lip}}+\theta\int f\mathrm{d}\mu\quad\text{for all }\theta\in\mathbb{R}\text{ and }f:E\to\mathbb{R}\text{ Lipschitz}$$

(ii) Talagrand transportation inequality $T_1:\mathrm{W}_1(\nu,\mu)\leq\sqrt{c\mathrm{H}(\nu\mid\mu)}$ for all $\nu\in\mathscr{P}_1(E)$

- Works for $\mathcal{N}(0,I_n)$ with $c=2$ thanks to the Sudakov–Tsirelson–Borell theorem
- Gozlan–Léonard extension to arbitrary tails: $\log L_{\frac{f-\mu f}{\|f\|_{\mathrm{Lip}}}}\leq\alpha\Leftrightarrow\alpha^*(\mathrm{W}_1)\leq\mathrm{H}$, for any convex α with $\alpha(0)=0$. Same proof! Covers Weibull-type $\exp(-t^p)$ and Bernstein-type $\exp(-\min(t^2/\sigma^2,t))$ tails.

Proof. Let us prove that (i)⇒(ii). We can assume in (i) that $\theta>0$ by replacing f by $-f$, and assume additionally that $\int f\mathrm{d}\mu=0$ and $\|f\|_{\mathrm{Lip}}=1$ by translation and dilation. Now (i) rewrites for such a function $f:E\to\mathbb{R}$ as

$$\int \mathrm{e}^g\,\mathrm{d}\mu\leq 1\quad\text{where}\quad g:=\theta f-\theta^2\tfrac{c}{4}.$$

The variational formula for the relative entropy gives, for all $\nu\in\mathscr{P}_1(E)$ and this g,

$$\int\Big(\theta f-\theta^2\frac{c}{4}\Big)\mathrm{d}\nu=\int g\mathrm{d}\nu\leq\mathrm{H}(\nu\mid\mu).$$

We can still recover (i) by taking $\mathrm{d}\nu\propto\mathrm{e}^g\mathrm{d}\mu$ which gives $\big(\int\mathrm{e}^g\mathrm{d}\mu\big)\log\int\mathrm{e}^g\mathrm{d}\mu\leq 0$ since $u\log u\leq 0$ implies $u\leq 1$.

Now, since $\int f\mathrm{d}\mu=0$, the previous property involving H rewrites as

$$\int f\mathrm{d}\nu-\int f\mathrm{d}\mu=\int f\mathrm{d}\nu\leq\frac{c}{4}\theta+\frac{1}{\theta}\mathrm{H}(\nu\mid\mu).$$

By taking the infimum over $\theta>0$ we get

$$\int f\mathrm{d}\nu-\int f\mathrm{d}\mu\leq\sqrt{c\mathrm{H}(\nu\mid\mu)}.$$

Taking $\sup_f$ gives (ii) by Kantorovich–Rubinstein duality. Finally, the arguments are reversible. □

Remark 2.4. Total variation as a singular case.

If d is the atomic distance $d(x,y)=\mathbf{1}_{x\neq y}$, then $\mathbb{E}(d(X,Y))=\mathbb{E}(\mathbf{1}_{X\neq Y})=\mathbb{P}(X\neq Y)$, $\|f\|_{\mathrm{Lip}}=\mathrm{osc}(f)$, and the Kantorovich–Rubinstein duality expresses that W_1 is then the total variation distance:

$$\mathrm{W}_1(\mu,\nu)=\inf_{\substack{(X,Y)\\X\sim\mu,Y\sim\nu}}\mathbb{P}(X\neq Y)=\sup_{\|f\|_\infty\leq\frac{1}{2}}\int f\mathrm{d}(\mu-\nu).$$

In this case, in the Bobkov–Götze equivalence, (i) is Hoeffding while (ii) is Pinsker.

2.2 Infimum convolution and quadratic cost

We set $\mathrm{BL}(E):=\{f:E\to\mathbb{R}:\|f\|_\infty<\infty,\|f\|_{\mathrm{Lip}}<\infty\}$.

Lemma 2.5. Kantorovich and infimum convolution.

$$\mathrm{W}_p(\nu,\mu)^p = \sup_{f\in \mathrm{BL}(E)}\left(\int Q(f)\mathrm{d}\nu - \int f\mathrm{d}\mu\right) \quad \text{where}^a \quad Q(f)(x) := \inf_{y\in E}\big(f(y)+d(x,y)^p\big).$$

[a] In other words $\mathrm{W}_p(\nu,\mu)^p = \sup\{\int g\mathrm{d}\nu - \int f\mathrm{d}\mu : f,g\in \mathrm{BL}(E), g(x)-f(y)\le d(x,y)^p, x,y\in E\}$.

Theorem 2.6. Bobkov–Götze: Talagrand transportation inequality for W_p.

For all $p\ge 1$, $\mu\in\mathscr{P}_p(E)$, and $c>0$, the following statements are equivalent:

(i) Talagrand transportation inequality T'_p: $\mathrm{W}_p(\nu,\mu)^p \le c\mathrm{H}(\nu\mid\mu)$ for all $\nu\in\mathscr{P}_p(E)$

(ii) Exponential integrability for infimum convolution: for all $f\in\mathrm{BL}(E)$,

$$\log\int \exp\Big(\frac{1}{c}\Big(Q(f)-\int f\mathrm{d}\mu\Big)\Big)\mathrm{d}\mu \le 0.$$

- (i) $T_p \equiv T'_p$ iff $p=2$, T_2 implies T_1 since $\mathrm{W}_1(\mu,\nu)^2\le \mathrm{W}_2(\mu,\nu)^2$ by Jensen inequality.
- (ii) for $p=2$ implies sub-Gaussian concentration for Lipschitz functions $\log\int \mathrm{e}^{\theta f}\mathrm{d}\mu \le \frac{c}{4}\theta^2\|f\|_{\mathrm{Lip}}^2+\theta\int f\mathrm{d}\mu$. Indeed, after reduction to $\int f\mathrm{d}\mu=0$, $\theta=1$, and f bounded, this follows from (ii) together with

$$\frac{1}{c}Q(cf)(x) \ge f(x)+\inf_{y\in E}\left(-\|f\|_{\mathrm{Lip}}d(x,y)+\frac{d(x,y)^2}{c}\right)\ge f(x)-\frac{c}{4}\|f\|_{\mathrm{Lip}}^2.$$

Proof. (ii)⇒(i). For any $f\in\mathrm{BL}(E)$ the property (ii) gives

$$\int \exp\left(\frac{1}{c}\left(Q(f)-\int f\mathrm{d}\mu\right)\right)\mathrm{d}\mu\le 1.$$

It follows then by the variational formula for H with an arbitrary $\nu\in\mathscr{P}_p(E)$ and $g:=\frac{1}{c}\left(Q(f)-\int f\mathrm{d}\mu\right)$,

$$\int\frac{1}{c}\left(Q(f)-\int f\mathrm{d}\mu\right)\mathrm{d}\nu = \int g\mathrm{d}\nu \le \mathrm{H}(\nu\mid\mu).$$

Taking the supremum over f gives (i) by the Kantorovich duality.

(i)⇒(ii). From the Kantorovich duality and then (i), for all $f\in\mathrm{BL}(E)$ and $\nu\in\mathscr{P}_p(E)$,

$$\int\left(Q(f)-\int f\mathrm{d}\mu\right)\mathrm{d}\nu = \int Q(f)\mathrm{d}\nu - \int f\mathrm{d}\mu = \mathrm{W}_p(\nu,\mu)^p\le c\mathrm{H}(\nu\mid\mu).$$

Since f and $Q(f)$ are bounded, we can take $\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}\propto\exp\left(\dfrac{1}{c}\left(Q(f)-\int f\mathrm{d}\mu\right)\right)$ to get (ii). □

Theorem 2.7. Talagrand W_2 inequality for Gaussian measures.

If $(E,d)=(\mathbb{R}^n,|\cdot|_2)$, $n\ge 1$, then $\mathrm{W}_2(\nu,\mathcal{N}(0,I_n))\le\sqrt{2\mathrm{H}(\nu\mid\mathcal{N}(0,I_n))}$ for all $\nu\in\mathscr{P}_2(\mathbb{R}^n)$.

- Still works for $1\le p<2$ but fails for $p>2$.
- Brenier theorem: $\mathrm{W}_2(\mu,\nu)^2=\mathbb{E}(|T_{\mu\to\nu}(X)-X|^2)$, $X\sim\mu$.

About the proof. Thanks to the additivity of $|\cdot|_2^2$, the functional inequality $\mathrm{W}_2(\cdot,\mu)\le\sqrt{c\mathrm{H}(\cdot\mid\mu)}$ tensorizes. The first proof of the theorem, due to Michel Talagrand, is by tensorization, the case $n=1$ being obtained by monotone rearrangement $T:=F_\nu^{-1}\circ F_\mu$. There is an alternative proof by Sergey Bobkov, Ivan Gentil, and Michel Ledoux, a follow-up of the works of Felix Otto and Cédric Villani, based on the logarithmic Sobolev inequality and a Hamilton–Jacobi equation, which can be seen as the analogue of the heat equation for infimum convolutions. This requires to control the gradient of $Q(f)$. This can be done using the fact that for all $f\in\mathrm{BL}(\mathbb{R}^n)$, the infimum-convolution $u(t,x)=Q_t(f)(x):=t^{-1}Q(tf)(x)$, solves the Hamilton–Jacobi equation (nonlinear PDE) $\partial_t u+\frac{1}{4}|\nabla_x u|_2^2=0$ on $(0,\infty)\times\mathbb{R}^n$, and $u(0,\cdot)=f$. This is known as the Hopf–Lax solution. □

Furthermore, it was proved by Nathaël Gozlan and Michel Ledoux that the Talagrand W_2 inequality is actually equivalent to a dimension-free sub-Gaussian concentration inequality for Lipschitz functions.