Mini-course on LSI

2024

Contents

1 Emergence of entropy in combinatorics, probability, statistics, analysis

1.1 Combinatorics

Number of microstates compatible with a macrostate, degree of freedom, disorder, volume, Stirling :

$$\frac{1}{n}\log\binom{n}{n_1,\ldots,n_r}\xrightarrow[n=n_1+\cdots+n_r\to\infty]{\nu_i=\frac{n_i}{n}\to p_i}\mathrm{S}(p):=-\sum_{i=1}^r p_i\log(p_i)\quad\text{also}\quad\binom{n}{n_1,\ldots,n_r}\approx\mathrm{e}^{n\mathrm{S}(\nu_1,\ldots,\nu_r)}.$$

If $A=\{1,\ldots,r\}$ and $n=n_1+\cdots+n_r$ then $\mathrm{Card}\{(x_1,\ldots,x_n)\in A^n:\forall 1\leq i\leq r:\sum_{k=1}^n\mathbf{1}_{x_k=i}=n_i\}=\binom{n}{n_1,\ldots,n_r}$.
Boltzmann observation in kinetic gas theory at the start of statistical physics.
Shannon information theory : average length per symbol with code of optimal length.

1.2 Probability

If $X_1,\ldots,X_n$ are i.i.d. of law μ on a finite set $A=\{a_1,\ldots,a_r\}$ then for all $x_1,\ldots,x_n\in A$,

$$\mathbb{P}((X_1,\ldots,X_n)=(x_1,\ldots,x_n))=\prod_{i=1}^r\mu_i^{\sum_{k=1}^n\mathbb{1}_{x_k=a_i}}=\prod_{i=1}^r\mu_i^{n\nu_i}=\mathrm{e}^{n\sum_{i=1}^n\nu_i\log\mu_i}=\mathrm{e}^{-n(\mathrm{S}(\nu)+\mathrm{H}(\nu\mid\mu))}.$$

where we have used the Boltzmann – Shannon entropy and Kullback – Leibler divergence or relative entropy :

$$\mathrm{S}(\nu):=-\sum_{i=1}^r\nu_i\log\nu_i\quad\text{and}\quad\mathrm{H}(\nu\mid\mu):=\sum_{i=1}^r\nu_i\log\frac{\nu_i}{\mu_i}=\sum_{i=1}^r\frac{\nu_i}{\mu_i}\log\frac{\nu_i}{\mu_i}\mu_i.$$

Boltzmann – Gibbsfication. At the heart of the Sanov large deviations principle via Laplace method.

1.3 Statistics

If $Y_1,\dots,Y_n$ are i.i.d. of law μ on a finite set $A=\{1,\dots,r\}$, then Fisher likelihood of data $(x_1,\dots,x_n)\in A^n$ is

$$\ell_{x_1,\dots,x_n}(\mu) := \mathbb{P}(Y_1=x_1,\dots,Y_n=x_n) = \prod_{k=1}^n \mu_{x_k}.$$

If $X_1,\dots,X_n$ are observed i.i.d. of law μ^* unknown then maximum likelihood estimator is

$$\widehat{\mu}_n := \arg\max_{\mu} \ell_{X_1,\dots,X_n}(\mu) = \arg\max_{\mu}\Big(\frac{1}{n}\log \ell_{X_1,\dots,X_n}(\mu)\Big).$$

Asymptotic analysis via law of large numbers and entropy as asymptotic contrast

$$\frac{1}{n}\log\ell_{X_1,\dots,X_n}(\mu) = \frac{1}{n}\sum_{k=1}^n \log\mu_{X_k} \xrightarrow[n\to\infty]{\text{a.s.}} \sum_{i=1}^r \mu_i^*\log\mu_i = \underbrace{-\mathrm{S}(\mu^*)}_{\text{const}} - \mathrm{H}(\mu^*\mid\mu).$$

1.4 Analysis

Deep elementary relation to powers : $\partial_{p=1} u^p = \partial_{p=1}\mathrm{e}^{p\log(u)} = u\log(u)$ and thus to L^p norms:

$$f\geq 0,\quad \partial_p\|f\|_p^p = \partial_p\int f^p\mathrm{d}\mu = \partial_p\int \mathrm{e}^{p\log(f)}\mathrm{d}\mu = \int f^p\log(f)\mathrm{d}\mu = \frac{1}{p}\int f^p\log(f^p)\mathrm{d}\mu.$$

1.5 Axioms

The entropy $\mathrm{S}=\mathrm{S}^{(r)}$ is characterized by the following three natural properties or axioms[1]:

(i) for all $r\geq 1$, $p\in\{(p_1,\dots,p_r)\in[0,1]^r : p_1+\cdots+p_r=1\}\mapsto \mathrm{S}^{(r)}(p)$ is continuous;

(ii) for all $r\geq 1$, $\mathrm{S}^{(r)}(\frac{1}{r},\dots,\frac{1}{r}) < \mathrm{S}^{(r+1)}(\frac{1}{r+1},\dots,\frac{1}{r+1})$

(iii) for all $r=r_1+\cdots+r_k\geq 1$; $\mathrm{S}^{(r)}(\frac{1}{r},\dots,\frac{1}{r}) = \mathrm{S}^{(k)}(\frac{r_1}{r},\dots,\frac{r_k}{r}) + \sum_{i=1}^k \frac{r_i}{r}\mathrm{S}^{(r_k)}(\frac{1}{r_k},\dots,\frac{1}{r_k})$.

Another characterization when $r=2$: $p\mapsto s(p)=S(p,1-p)=-p\log(p)-(1-p)\log(1-p)$ solves the ODE[2]

$$p(1-p)s''(p) = -1 \text{ on } [0,1] \text{ with boundary conditions } s(0)=s(1)=0.$$

2 Boltzmann – Gibbs measures and free energy

Boltzmann – Shannon differential or continuous entropy of a probability measure μ on $\mathbb{R}^n$:

$$\mathrm{S}(\mu) := \begin{cases} -\displaystyle\int f(x)\log(f(x))\mathrm{d}x & \text{if } \mathrm{d}\mu(x)=f(x)\mathrm{d}x \text{ and } f\log f\in\mathrm{L}^1(\mathrm{d}x)\\ +\infty & \text{otherwise}\end{cases}$$

Gaussian case and Shannon exponential entropy (volume again!)

$$\mathrm{S}(\mathcal{N}(m,K)) = \log\sqrt{(2\pi\mathrm{e})^n\det K} \quad\text{and}\quad \mathrm{N}(\mathcal{N}(m,K)) := \frac{\mathrm{e}^{\frac{2}{n}\mathrm{S}(\mathcal{N}(m,K))}}{2\pi\mathrm{e}} = (\det K)^{\frac{1}{n}}.$$

Kullback – Leibler relative entropy between two probability measures on the same space :

$$H(\nu\mid\mu) := \int\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\log\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\mathrm{d}\mu \quad \geq 0 \text{ with = iff } \mu=\nu.$$

Jensen inequality (and its equality case) for strictly convex function $x\in\mathbb{R}_+\mapsto x\log(x)$.

We take $V:\mathbb{R}^n$ or $A\to\mathbb{R}$, interpreted as an energy, such that $Z_\beta := \int \mathrm{e}^{-\beta V(x)}\mathrm{d}x<\infty$ for all $\beta>0$. Maximizing $\mu\mapsto\mathrm{S}(\mu)$ over the constraint of average energy $\int V\mathrm{d}\mu = v$ gives the maximizer

$$\mu_\beta := \frac{1}{Z_\beta}\mathrm{e}^{-\beta V}\mathrm{d}x,$$

provided that v is an admissive energy, typically $v\geq\min V$ when V has a unique global minimum. In the Gaussian case, V is quadratic while $1/\beta = v$ is the variance. Here $\mathrm{d}x$ is the Lebesgue or counting measure.

[1] As far as I know, this is not useful in practice to cook up the analogue of an entropy for a given problem. The best for such a quest is to consider a desirable property and to study the proof in the classical case to identify the key basic properties of entropy to imitate.

[2] This is useful for the mean time of fixation in the Wright-Fisher model in mathematical biology in the large population asymptotics.

Theorem 2.1. Variational characterization : maximum entropy at fixed average energy.

$$\int V\mathrm{d}\mu = \int V\mathrm{d}\mu_\beta \quad \Rightarrow \quad \mathrm{S}(\mu_\beta) - \mathrm{S}(\mu) = \mathrm{H}(\mu \mid \mu_\beta).$$

Dual point of view : instead of fixing the average energy, let us fix the inverse temperature β and introduce

$$\mathrm{F}(\mu) := \int V\mathrm{d}\mu - \frac{1}{\beta}\mathrm{S}(\mu)$$

which is the Helmholtz free energy. Lagrangian point of view, the constraint is added to the functional.

$$\mathrm{F}(\mu_\beta) = -\frac{1}{\beta}\log(Z_\beta) \quad \text{since} \quad \mathrm{S}(\mu_\beta) = \beta\int V\mathrm{d}\mu_\beta + \log Z_\beta.$$

Theorem 2.2. Variational characterization : minimum free energy at fixed temperature.

$$\mathrm{F}(\mu) - \mathrm{F}(\mu_\beta) = \frac{1}{\beta}\mathrm{H}(\mu \mid \mu_\beta).$$

This explains why H is often called free energy instead of relative entropy or Kulback – Leibler divergence.

3 Emergence of log-Sobolev : Markov diffusion processes

For $V : \mathbb{R}^n \to \mathbb{R}$ $\mathscr{C}^2$ with $\inf_x \mathrm{Hess}V(x) > -\infty$, the Boltzmann – Gibbs measure $\mu = \frac{1}{Z}\mathrm{e}^{-V}\mathrm{d}x$ is invariant and reversible for the non-explosive Markov diffusion process $(X_t)_{t\geq 0}$ solving the stochastic differential equation

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t.$$

This is an ergodic overdamped Langevin process. The associated Markov semigroup $(P_t)_{t\geq 0}$ defined by

$$P_t(f)(x) = \mathbb{E}(f(X_t) \mid X_0 = x) \quad \text{has infinitesimal generator} \quad L = \partial_{t=0}P_t = \Delta - \langle \nabla V, \nabla \rangle.$$

The reversibility of μ translates into an integration by parts

$$\int fLg\mathrm{d}\mu = \int gLf\mathrm{d}\mu = -\int \langle \nabla f, \nabla g\rangle \mathrm{d}\mu.$$

Moreover, if $\mathrm{d}\mu_0 = f_0\mathrm{d}\mu$, $f_0 \geq 0$, $\int f_0\mathrm{d}\mu = 1$, then $X_t \sim \mu_t$ with $\mathrm{d}\mu_t := f_t\mathrm{d}\mu$ where $f_t = P_t(f_0)$.

3.1 Boltzmann H-theorem, Bakry – Émery criterion, Gross hypercontractivity

Boltzmann H-theorem, via integration by parts :

$$\partial_t \mathrm{H}(\mu_t \mid \mu) = \partial_t \int f_t \log(f_t)\mathrm{d}\mu = \int (1 + \log(f_t))Lf_t\mathrm{d}\mu = \int \log(f_t)Lf_t\mathrm{d}\mu = -\int \frac{|\nabla f_t|^2}{f_t}\mathrm{d}\mu = -\mathrm{I}(\mu_t \mid \mu) \leq 0.$$

Fisher information :

$$\mathrm{I}(\nu \mid \mu) = -\int \log(f)Lf\mathrm{d}\mu = \int \frac{|\nabla f|^2}{f}\mathrm{d}\mu, \quad f := \frac{\mathrm{d}\nu}{\mathrm{d}\mu}.$$

Theorem 3.1. Exponential decay of relative entropy ⇔ LSI.

For all constant $c > 0$, the following properties are equivalent :

(i) Exponential decay of relative entropy : $\forall \mu_0$, $\forall t \geq 0$, $\mathrm{H}(\mu_t \mid \mu) \leq \mathrm{e}^{-\frac{4}{c}t}\mathrm{H}(\mu_0 \mid \mu)$.

(ii) Logarithmic Sobolev Inequality (LSI) : $\forall \nu$, $\mathrm{H}(\nu \mid \mu) \leq \frac{c}{4}\mathrm{I}(\nu \mid \mu)$.

This type of equivalence is not specific to this relative entropy and this Markov process[3].

[3] For any convex Φ such that $\Phi(1) = 0$, if we consider the Φ relative entropy or divergence $\mathrm{H}^\Phi(\nu \mid \mu) = \int \Phi(f)\mathrm{d}\mu$, $f = \mathrm{d}\nu/\mathrm{d}\mu$, then the exponential decay $\mathrm{H}^\Phi(\mu_t \mid \mu) \leq \mathrm{e}^{-\frac{4}{c}t}\mathrm{H}^\Phi(\mu_0 \mid \mu)$ for all t and μ_0 is equivalent to the Φ Sobolev functional inequality $\mathrm{H}^\Phi(\nu \mid \mu) \leq \frac{c}{4}\mathrm{I}^\Phi(\nu \mid \mu)$ for all ν, where $\mathrm{I}^\Phi(\nu \mid \mu) = -\int \Phi'(f)Lf\mathrm{d}\mu = \int \Phi''(f)|\nabla f|^2\mathrm{d}\mu$. In particular, for $\Phi(u) = u^2$, we get the Poincaré inequality $\mathrm{Var}_\mu(f) \leq \frac{c}{2}\mathbb{E}_\mu(|\nabla f|^2)$.

Proof. We get (i) from (ii) by Grönwall since $\partial_t \mathrm{H}(\mu_t \mid \mu) = -\mathrm{I}(\mu_t \mid \mu) \geq -\frac{4}{c}\mathrm{H}(\mu_t \mid \mu)$ (we used LSI for $\nu = \mu_t$). Conversely, if $\alpha(t) := \mathrm{e}^{-\frac{4}{c}t}\mathrm{H}(\mu_0 \mid \mu) - \mathrm{H}(\mu_t \mid \mu)$ then $\alpha(0) = 0$, $\alpha(t) \geq 0$ for all $t \geq 0$ thus $\alpha'(0) \geq 0$ (LSI for μ_0!). □

Dyson–Ornstein–Uhlenbeck example :

$$V(x) = \|x\|^2 - \beta \sum_{i<j} \log(x_i - x_j) \quad \text{with} \quad V(x) = +\infty \quad \text{outside} \quad \{x_1 > \cdots > x_n\}.$$

Theorem 3.2. Bakry – Émery criterion for LSI (1984) : role of convexity.

If $V = \frac{1}{2\sigma^2}\|\cdot\|^2 + C$ with C convex, then μ satisfies LSI : $\forall \nu,\ \mathrm{H}(\nu \mid \mu) \leq \frac{\sigma^2}{2}\mathrm{I}(\nu \mid \mu)$.

In particular by taking $C \equiv 0$ we get that the Gaussian $\mathcal{N}(0, \sigma^2 I_n)$ satisfies LSI. Moreover, a multivariate Gaussian $\mathcal{N}(m, K)$ satisfies the same LSI as $\mathcal{N}(0, \|K\|_{\mathrm{op}}^2 I_n)$ (just use the the Lipschitz map $x \mapsto \sqrt{K}(x - m)$).

Proof. By integration by parts and the Bochner formla[4], we get, after some algebra, an H-theorem for Fisher :

$$\partial_t \mathrm{I}(\mu_t \mid \mu) = -2\int \Gamma_2(\log(f_t))\mathrm{d}\mu_t \quad \text{where} \quad \Gamma_2(f) := \|\nabla^2 f\|_{\mathrm{HS}}^2 + \langle \nabla^2 V \nabla f, \nabla f\rangle \geq \frac{1}{\sigma^2}\|\nabla f\|^2.$$

By Grönwall this gives an exponential decay of Fisher information :

$$\forall \mu_0, \forall t \geq 0, \quad \mathrm{I}(\mu_t \mid \mu) \leq \mathrm{e}^{-\frac{2}{\sigma^2}t}\mathrm{I}(\mu_0 \mid \mu).$$

Finally, the LSI for $\nu = \mu_0$ comes from

$$\mathrm{H}(\mu_0 \mid \mu) = -\int_0^\infty \partial_t \mathrm{H}(\mu_t \mid \mu)\mathrm{d}t = \int_0^\infty \mathrm{I}(\mu_t \mid \mu)\mathrm{d}t \leq \mathrm{I}(\mu_0 \mid \mu)\int_0^\infty \mathrm{e}^{-\frac{2}{\sigma^2}t}\mathrm{d}t = \frac{\sigma^2}{2}\mathrm{I}(\mu_0 \mid \mu).$$

We also get the convexity of $t \mapsto \mathrm{H}(\mu_t \mid \mu)$ via $\partial_t^2 \mathrm{H}(\mu_t \mid \mu) \geq 0$, a refinement of the Boltzmann H-theorem. □

The proof above can be adapted to discrete univariate Markov processes (birth and death), using discrete log-concavity, despite the lack of chain rule for discrete derivatives. However this fails in higher dimensions due to the lack of a good notion of discrete convexity. Yet there are semi-efficient notions of discrete Ricci curvature, such as the Olivier – Joulin curvature, based on the idea of comparing $\mathrm{dist}(P_t(\cdot)(x), P_t(\cdot)(y))$ with $\mathrm{dist}(x, y)$.

For all $f : E \to \mathbb{R}_+$ we define (analogy with the variance, replacing x^2 with $x\log(x)$)

$$\mathrm{Ent}_\mu(f) := \int f\log(f)\mathrm{d}\mu - \int f\mathrm{d}\mu \log\int f\mathrm{d}\mu = \mathrm{H}(\nu \mid \mu)\int f\mathrm{d}\mu \quad \text{where} \quad \mathrm{d}\nu := \frac{f}{\int f\mathrm{d}\mu}\mathrm{d}\mu.$$

Theorem 3.3. Leonard Gross (1975) : Hypercontractivity of Markov semigroup ⇔ LSI.

For all constant $c > 0$ the following properties are equivalent (the norms are with respect to μ) :

(i) Hypercontractivity of semigroup : $\forall t \geq 0, \forall f, \forall p \geq 1, \|f_t\|_{p(t)} \leq \|f\|_p$ where $p(t) := 1 + (p-1)\mathrm{e}^{\frac{4}{c}t}$

(ii) Logarithmic Sobolev inequality (LSI) : $\forall f, \mathrm{Ent}_\mu(f^2) \leq c\int |\nabla f|^2 \mathrm{d}\mu$.

The term comes from the fact that $p(t) > p = p(0)$ for all $t > 0$.

Proof. Same idea as for exponential decay with this time $\alpha(t) := \log\|f_t\|_{p(t)}$.
Involves crucially the fact that $\partial_p \int f^p \mathrm{d}\mu = \int f^p \log(f)\mathrm{d}\mu$ for $f \geq 0$.
We can assume that $f \geq 0$ since $|f_t| \leq |f|_t$. For all $t > 0$, we find

$$\alpha'(t) = \left(\frac{1}{p(t)}\log\int f_t^{p(t)}\mathrm{d}\mu\right)' = \frac{p'(t)}{p(t)^2}\frac{1}{\int f_t^{p(t)}\mathrm{d}\mu}\left(\mathrm{Ent}_\mu(f_t^{p(t)}) + \frac{p(t)^2}{p'(t)}\int (Lf_t)f_t^{p(t)-1}\mathrm{d}\mu\right).$$

Now $p(t) - 1 = \frac{c}{4}p'(t)$, while by LSI and integration by parts we get

$$\mathrm{Ent}_\mu(g^p) \leq \frac{c}{4}\int \frac{|\nabla g^p|^2}{g^p}\mathrm{d}\mu = -\frac{c}{4}\frac{p^2}{(p-1)}\int (Lg)g^{p-1}\mathrm{d}\mu.$$

Therefore $\alpha'(t) \leq 0$ for any $t \geq 0$ is equivalent to LSI. □

[4]This is the commutation $\nabla L = L\nabla - \nabla^2 V\nabla$. On a Riemannian manifold, there is an additional Ricci curvature term $-\mathrm{Ric}(\nabla, \nabla)$.

LSI can also be deduced geometrically from Sobolev inequality on spheres (Beckner).
LSI can also be deduced from isoperimetric inequality (Ledoux, Bobkov).
LSI is inspiring for Hamilton Ricci flow for Poincaré conjecture (Perelman).
LSI is related to transportation of measure (Talagrand, Otto–Villani, Bobkov–Gentil–Ledoux).
LSI gives Poincaré by linearization. $\mathrm{Ent}_\mu((1+\varepsilon f)^2) = \frac{\varepsilon^2}{2}\mathrm{Var}_\mu(f) + o(\varepsilon^2)$ for f bounded.

3.2 Variational formula and tenrosization

Lemma 3.4. Variational formula.

For all probability measure μ on E and all $f : E \to \mathbb{R}_+$, $f \in L^1(\mu)$, we have the linearization

$$\mathrm{Ent}_\mu(f) = \sup\Big\{\int fg\mathrm{d}\mu : \int \mathrm{e}^g \mathrm{d}\mu \le 1\Big\}, \text{ supremum achieved for } g = \log(f) - \log\int f\mathrm{d}\mu.$$

In particular, the inequality ≤ 1 can be replaced by the equality $= 1$.

Proof. Follows from the convexity $uv \le u\log(u) - u + \mathrm{e}^v$, $u \ge 0$, $v \in \mathbb{R}$.
Alternatively, reduce by homogeneity to $\int f\mathrm{d}\mu = 1$, and then use Jensen for the concave log and the law $f\mathrm{d}\mu$:

$$\int fg\mathrm{d}\mu = \int f\log(f)\mathrm{d}\mu + \int \log\Big(\frac{\mathrm{e}^g}{f}\Big) f\mathrm{d}\mu \le \int f\log(f)\mathrm{d}\mu + \log\int \frac{\mathrm{e}^g}{f} f\mathrm{d}\mu \le \int f\log(f)\mathrm{d}\mu.$$

□

Replacing g such that $\int \mathrm{e}^g\mathrm{d}\mu = 1$ by $g - \log\int \mathrm{e}^g\mathrm{d}\mu$ without constraint on g, we get that relative entropy and log-Laplace transform are the Legendre transform of each other :

$$\mathrm{H}(\nu \mid \mu) = \sup_g\Big\{\int g\mathrm{d}\nu - \log\int \mathrm{e}^g\mathrm{d}\mu\Big\} \quad \text{and} \quad \sup_\nu\Big\{\int g\mathrm{d}\nu - \mathrm{H}(\nu \mid \mu)\Big\} = \log\int \mathrm{e}^g\mathrm{d}\mu.$$

Lemma 3.5. Tensorisation.

If $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ is a product probability measure on a product space $E = E_1 \times \cdots \times E_n$ then

$$\mathrm{Ent}_\mu(f) \le \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i}(f)\mathrm{d}\mu \quad \text{for all} \quad f \in L^1(\mu, E \to \mathbb{R}_+).$$

This is not specific to this type of relative entropy[5].

Proof. By induction on n, it suffices to consider the case $n = 2$. Let $g : E \to \mathbb{R}$ be such that $\int \mathrm{e}^g\mathrm{d}\mu = 1$. Then

$$g = g_1 + g_2 \quad \text{with} \quad g_1 := g - \log\int \mathrm{e}^g\mathrm{d}\mu_1 \quad \text{and} \quad g_2 := \log\int \mathrm{e}^g\mathrm{d}\mu_1,$$

in such a way that $\int \mathrm{e}^{g_1}\mathrm{d}\mu_1 = 1$ and $\int \mathrm{e}^{g_2}\mathrm{d}\mu_2 = 1$. The variational formula of Lemma 3.4 for μ_i and g_i gives

$$\int fg_1\mathrm{d}\mu_1 + \int fg_2\mathrm{d}\mu_2 \le \mathrm{Ent}_{\mu_1}(f) + \mathrm{Ent}_{\mu_2}(f), \quad \text{hence} \quad \int fg\mathrm{d}\mu \le \int \mathrm{Ent}_{\mu_1}(f)\mathrm{d}\mu_1 + \int \mathrm{Ent}_{\mu_2}(f)\mathrm{d}\mu_2,$$

and it remains to use the variational formula of Lemma 3.4 this time for μ and g. □

3.3 Log-Sobolev for Gaussian from two-points space via tensorization and CLT

Theorem 3.6. Logarithmic Sobolev inequality (LSI) for the Gaussian.

For all $n \ge 1$, denoting $\gamma_n := \mathcal{N}(0, I_n) = \gamma_1^{\otimes n}$, we have, for all $f \in L^2(\gamma^n) \cap \mathscr{C}^2(\mathbb{R}^n, \mathbb{R})$, in $[0, +\infty]$:

$$\mathrm{Ent}_{\gamma^n}(f^2) \le 2\int |\nabla f|^2\mathrm{d}\gamma^n.$$

[5] More generally, for Φ convex with $\Phi(1) = 0$, such a tensorization remains valid for the Φ relative entropy or divergence $\mathrm{Ent}_\mu^\Phi(f) = \int \Phi(f)\mathrm{d}\mu - \Phi(\int f\mathrm{d}\mu)$, iff $(u, v) \mapsto \Phi''(u)v^2$ is convex. In particular, it works for $\Phi(u) = u^2$ (variance) and $\Phi(u) = u\log(u)$ (relative entropy).

Moreover the constant 2 is optimal in the sense that equality is achieved for $f^2(x) = \mathrm{e}^{\langle \lambda, x\rangle}$, $\lambda \in \mathbb{R}^n$.

- By analogy with classical Sobolev inequalities we can rewrite LSI as
$$\int f^2 \log(f^2)\mathrm{d}\gamma^n \le \int f^2 \mathrm{d}\gamma^n \log \int f^2 \mathrm{d}\gamma^n + 2\int |\nabla f|^2 \mathrm{d}\gamma^n,$$
stating that $f^2\log(f^2) \in L^1(\gamma^n)$ as soon as $f^2 \in L^1(\gamma^n)$ and $|\nabla f|^2 \in L^1(\gamma^n)$.
- By an affine change of variable we get that $\mathcal{N}(m,\Sigma)$ satisfies an ISL with constant $\|\Sigma\|^2_{\mathrm{op}}$.
- The linearization of LSI via $f^2 = (1+\varepsilon g)^2$ gives a Poincaré inequality of constant 1 :
$$\mathrm{Var}_{\gamma^n}(f) := \int f^2 \mathrm{d}\gamma^n - \Big(\int f \mathrm{d}\gamma^n\Big)^2 \le \int |\nabla f|^2 \mathrm{d}\gamma^n.$$
- Infinite dimensional nature, tensorization gives LSI for Wiener measure with Mallianin derivative
- L^1, L^2, L^p, versions via chain rule, not available in discrete spaces (leads to modified LSI)

Proof following Gross and Bobkov. The idea is to start from the two-points space, forge a discret LSI, tensorize to the cube, and then use the CLT. Namely, let us consider the uniform law $\nu = \frac{1}{2}(\delta_{-1} + \delta_1)$ on $\{-1,1\}$. Then
$$\mathrm{Ent}_\nu(g^2) \le \frac{(g(1)-g(-1))^2}{2} \quad \text{for all } g : \{-1,1\} \to \mathbb{R}.$$
We can assume without loss of generality that $g \ge 0$, and by homogeneity that $g(1)^2 + g(-1)^2 = 2$, which reduces the inequality to the optimal univariate inequality (checkable by direct calculus, equality achieved for $u = 1$)
$$u\log(u) + (2-u)\log(2-u) \le (\sqrt{u} - \sqrt{2-u})^2, \quad 0 \le u \le 2.$$
Now, let us take $f \in \mathscr{C}^2_c(\mathbb{R},\mathbb{R})$ and let us define $g : \{-1,1\}^n \to \mathbb{R}$ as $g(x_1,\dots,x_n) := f(\frac{1}{\sqrt{n}}(x_1 + \cdots + x_n))$. Let $\mu := \nu^{\otimes n}$ be the uniform law on the cube $\{-1,1\}^n$. By tensorisation (Lemma 3.5) and the inequality on $\{-1,1\}$,
$$\mathrm{Ent}_\mu(g^2) \le \frac{1}{2}\int \sum_{i=1}^n (g(x^{i,+}) - g(x^{i,-}))^2 \mathrm{d}\mu$$
where $x^{i,\pm}_j := x_j$ if $j \ne i$ and $:= \pm 1$ if $j = i$. A Taylor formula at order 1 for f at $\frac{x_1+\cdots+x_n}{\sqrt{n}}$ gives
$$g(x^{i,+}) - g(x^{i,-}) = \frac{2}{\sqrt{n}} f'\Big(\frac{x_1+\cdots+x_n}{\sqrt{n}}\Big) + o\Big(\frac{1}{\sqrt{n}}\Big)$$
with an o uniform in x since f is $\mathscr{C}^2_c$ and thus with bounded second derivative. Therefore, thanks to the CLT,
$$\mathrm{Ent}_{\gamma_1}(f^2) \le 2\int f'^2 \mathrm{d}\gamma^1.$$
We can weaken the conditions on f by approximation arguments.
We can generalize to $\gamma^n = (\gamma^1)^{\otimes n}$ for all $n \ge 1$ by using tensorization again ! □

- Stability of LSI by tensorization or dimension free statements : if μ, ν satisfy to LSI with constants c_μ and c_ν then $\mu \otimes \nu$ satisfies to LSI with constant $\max(c_\mu, c_\nu)$. In particular if μ satisfies LSI with constant c then $\mu^{\otimes N}$ satisfies to LSI with same constant c for all N. The constant depend on the class of test functions. The tensorization works if the class of test functions as well as the LHS are both stable by tensorization.
- Stability by Lipschitz deformation. If μ satisfies LSI with constant c and then its image with a map F satisfies LSI with constant $c\|F\|^2_{\mathrm{Lip}}$. In particular Uniform([0,1]) satisfies LSI, and LSI is stable by convolution.
- Optimal transportation. Caffarelli showed using the Monge–Ampère equation and the maximum principle that the Bakry–Émery condition implies that μ is the image of γ_n with F such that $\|F\|_{\mathrm{Lip}} \le \sigma$, leading to LSI via Lipschitz deformation from the Gaussian case. On the other hand, Cordero–Erausquin used Monge–Ampere to get LSI directly in this case, still via Monge–Ampère and an exploit of convexity.
- There is also a stability by bounded perturbation on V, due to Holley–Stroock.
This was generalized by Bodineau–Helffer to V convex + bounded.
Generalized by Zegarlinski to spin systems with exponential decay of correlations.
Generalized by Bauerschmidt–Bodineau recently, in the spirit of high dimentional convexification…
- Tails beyond Gaussians. The probability measure $\frac{1}{Z_\alpha}\mathrm{e}^{-|x|^\alpha}\mathrm{d}x$ on $\mathbb{R}$, $\alpha > 0$, $Z_\alpha := \int_{\mathbb{R}} \mathrm{e}^{-|x|^\alpha}\mathrm{d}x < \infty$, satisfies LSI iff $\alpha \ge 2$, and a Poincaré inequality iff $\alpha \ge 1$. The Gaussian corresponds to the critical case $\alpha = 2$.

4 Log-Sobolev and concentration of measure

Theorem 4.1. LSI ⇒ sub-Gaussian Laplace transform of Lipschitz functions (Herbst 1998).

If $\mu \in \mathscr{P}(\mathbb{R}^n)$ satisfies to LSI with constant c :

$$\exists c \in \mathbb{R}_+,\ \ \forall f \in L^2(\mu) \cap \mathscr{C}^2(\mathbb{R}^n,\mathbb{R}),\ \ \mathrm{Ent}_\mu(f^2) \le c \int |\nabla f|^2 \mathrm{d}\mu.$$

then Lipschitz functions have sub-Gaussian Laplace transform :

$$\forall f : \mathbb{R}^n \to \mathbb{R} \text{ Lipschitz and in } L^1(\mu), \forall \theta \in \mathbb{R},\ L(\theta) := \int \exp(\theta f)\mathrm{d}\mu \le \exp\Big(\theta^2 \frac{c}{4}\|f\|_{\mathrm{Lip}}^2 + \theta \int f \mathrm{d}\mu\Big).$$

The same method with the Poincaré inequality produces a sub-exponential concentration inequality.

Proof. First of all we reduce to f bounded, $\mathscr{C}^\infty$, centered for μ, $\|f\|_{\mathrm{Lip}} = 1$, and $\theta > 0$.
Now, for all $\theta > 0$, the LSI with $\mathrm{e}^{\theta f}$ instead of f^2 gives, via $|\nabla \mathrm{e}^{\theta f}| = |\theta \nabla f| \mathrm{e}^{\theta f}$ and $\||\nabla f|\|_\infty = \|f\|_{\mathrm{Lip}} \le 1$, that

$$\theta L'(\theta) - L(\theta)\log L(\theta) \le \frac{c}{4}\theta^2 L(\theta), \quad \text{in other words} \quad K' \le \frac{c}{4} \text{ where } K(\theta) := \tfrac{1}{\theta}\log L(\theta).$$

The result follows from $K(0) = (\log L)'(0) = L'(0)/L(0) = \mu(f)$, which comes from $L(0) = 1$ and $L'(0) = \mu(f)$. □

Corollary 4.2. LSI ⇒ Sub-Gaussian concentration for Lipschitz functions.

If $\mu \in \mathscr{P}(\mathbb{R}^n)$ satisfies to LSI of constant c as in Theorem 4.1,
then for all $X \sim \mu$, $r \ge 0$, and $f : \mathbb{R}^n \to \mathbb{R}$ in $L^1(\mu)$,

$$\mathbb{P}\Big(\Big|f(X) - \mathbb{E}(f(X))\Big| \ge r\Big) \le 2\exp\Big(-\frac{r^2}{c\|f\|_{\mathrm{Lip}}^2}\Big).$$

More generally, if $X_1,\ldots,X_N$, $N \ge 1$, are i.i.d. of law μ, then

$$\mathbb{P}\Big(\Big|\frac{f(X_1)+\cdots+f(X_N)}{N} - \mathbb{E}(f(X_1))\Big| \ge r\Big) \le 2\exp\Big(-\frac{Nr^2}{c\|f\|_{\mathrm{Lip}}^2}\Big).$$

- Dimension free rewrite :

$$\mathbb{P}\Big(\sqrt{N}\Big|\frac{f(X_1)+\cdots+f(X_N)}{N} - \mathbb{E}(f(X_1))\Big| \ge r\Big) \le 2\exp\Big(-\frac{r^2}{c\|f\|_{\mathrm{Lip}}^2}\Big).$$

- A consequence is the exponential integrability for the square of $Y := f(X) - \mathbb{E}(Y)$:

$$\mathbb{E}(\mathrm{e}^{\theta Y^2}) = \theta \int_0^\infty r\mathrm{e}^{\theta r^2}\mathbb{P}(|Y| \ge r)\mathrm{d}r < \infty \quad \text{as soon as } \theta < \frac{1}{c\|f\|_{\mathrm{Lip}}^2}.$$

Proof. For the first, we reduce to $\|f\|_{\mathrm{Lip}} = 1$ and $\mu(f) = \int f \,\mathrm{d}\mu = 0$ by translation and dilation, then for all $r \ge 0$ and $\theta > 0$, the Markov inequality and Theorem 4.1 give

$$\mu(f \ge r) = \mu\Big(\mathrm{e}^{\theta f} \ge \mathrm{e}^{\theta r}\Big) \le \mathrm{e}^{-\theta r}\int \mathrm{e}^{\theta f}\,\mathrm{d}\mu \le \mathrm{e}^{-\theta r + \frac{c}{4}\theta^2} \le \mathrm{e}^{-\frac{r^2}{c}},$$

where the last inequality comes from the optimal choice $\theta = 2r/c$. By using the result on $\pm f$ we get

$$\mu\Big(\Big|f - \int f \mathrm{d}\mu\Big| \ge r\Big) \le 2\exp\Big(-\frac{r^2}{2\|f\|_{\mathrm{Lip}}^2}\Big).$$

For the second inequality, we observe that $x \in (\mathbb{R}^n)^N \mapsto F(x) := \frac{1}{N}(f(x_1)+\cdots+f(x_N))$ is Lipschitz with

$$\|F\|_{\mathrm{Lip}} \le \frac{\|f\|_{\mathrm{Lip}}}{N}\sup_{x\neq y}\frac{\sum_{i=1}^N |x_i - y_i|}{\sqrt{\sum_{i=1}^N |x_i - y_i|^2}} \le \frac{\|f\|_{\mathrm{Lip}}}{\sqrt{N}}.$$

Moreover $\mathbb{E}(F(X_1,\dots,X_N)) = \mathbb{E}(f(X_1))$. Furthermore $(X_1,\dots,X_N)\sim\mu^{\otimes N}$ satisfies LSI with same constant $2c$ (dimension free : does not depend on N), thanks to the tensorization method used for proving Theorem 3.6. □

- Unstability by tensor product of sub-Gaussiannity of Laplace transform of Lipschitz functions and sub-Gaussian concentration, hence the usefulness of LSI when it holds!

4.1 Wigner Ensembles

Let $S := (S_{ij})_{1\le i,j\le n}$ be an $n\times n$ real symmetric random matrix, $n\ge 1$.
Let $c_{ij}\in[0,+\infty]$ be the LSI constant of the law of S_{ij} (sparsity: take $c_{ij}=0$ if S_{ij} is constant (possibly $\equiv 0$).
Then for all $f:\mathbb{R}\to\mathbb{R}$ and all $r\ge 0$,

$$\mathbb{P}\Big(\Big|\mathrm{Tr}_n f\Big(\frac{S}{\sqrt{n}}\Big)-\mathbb{E}\mathrm{Tr}_n f\Big(\frac{S}{\sqrt{n}}\Big)\Big|\ge r\Big)\le 2\exp\Big(-\frac{n^2r^2}{\|f\|_{\mathrm{Lip}}^2\max_{ij}c_{ij}}\Big).$$

LSI tensorization and spectrum of a symmetric matrix is a Lipschitz wrt its entries (Weyl inequalities) :

$$|\lambda_i(A)-\lambda_i(B)|\le\|A-B\|_{\mathrm{op}}.$$

Special case : if S is Gaussian, say GOE, we can use the Gaussian LSI and the Lipschitz stability.

4.2 Beta-Ensembles

Let us consider the probability measure μ on $\mathbb{R}^n$ given by

$$\frac{1}{Z_n}\prod_{i=1}^n \mathrm{e}^{-\sum_{i=1}^n U(x_i)}\prod_{i<j}(x_i-x_j)^\beta\mathbf{1}_{x_1\le\cdots\le x_n}=\frac{1}{Z_n}\mathrm{e}^{-\big(\sum_{i=1}^n U(x_i)+\beta\sum_{i<j}\log\frac{1}{x_i-x_j}\big)}\mathbf{1}_{x_1\le\cdots\le x_n}$$

where $U:\mathbb{R}\to\mathbb{R}$ is such that $U(x)=C(x)+\frac{1}{2\sigma^2}\|x\|^2$, C is $\mathscr{C}^2$ and convex and where $\sigma,\beta>0$.
Satisfies LSI with constant $2\sigma^2$, by Bakry – Émery or Caffarelli thanks to the convexity of

$$(x_1,\dots,x_n)\mapsto\sum_{i=1}^n C(x_i)-\beta\sum_{i<j}\log(x_i-x_j).$$

References (standard and personal)

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panor. Synth.* Société Mathématique de France, 2000.

[2] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348. Springer, 2014.

[3] R. Bauerschmidt, T. Bodineau, and B. Dagallier. Stochastic dynamics and the Polchinski equation: an introduction. *Probab. Surv.*, 21:200–290, 2024.

[4] D. Chafaï. From Boltzmann to random matrices and beyond. *Ann. Fac. Sci. Toulouse, Math. (6)*, 24(4):641–689, 2015.

[5] D. Chafaï and J. Lehec. Logarithmic Sobolev Inequalities Essentials. lecture notes , 2018.

[6] D. Cordero-Erausquin. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.*, 161(3):257–269, 2002.

[7] L. Gross. Invariance of intrinsic hypercontractivity under perturbation of schrödinger operators. arXiv:2412.20282v1, 2024.

[8] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 1–134. Springer, Berlin, 2003.

[9] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Math. Surv. Monogr.* American Mathematical Society, 2001.

[10] M. Ledoux. More than fifteen proofs of the logarithmic Sobolev inequality. `http://perso.math.univ-toulouse.fr/ledoux/files/2023/03/15.pdf`, 2015.

[11] G. Royer. *An initiation to logarithmic Sobolev inequalities*, volume 14 of *SMF/AMS Texts and Monographs*. American Mathematical Society, Société Mathématique de France, 2007. Translated from the 1999 French original by Donald Babbitt.

[12] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. AMS, 2003.