For the Love of Gaussians

Celebrating the Work of Michel Talagrand

Ronan Herry

Société mathématique de Belgique, Recent Breakthroughs 2024

Université Libre de Bruxelles

Who is Michel Talagrand?

Quotes by Talagrand

- "Ce que j'aime, c'est de couper des intervalles en rondelles."
- "La théorie de la mesure m'a graduellement amené à apprendre des probabilités, même si ce que j'appelle par ce nom n'est pas toujours reconnu comme tel par les *vrais* probabilistes."
- "The first advice I received from my advisor Gustave Choquet was as follows: always consider a problem under the minimum structure in which it makes sense. By following it, one is naturally led to study problems with a kind of minimal and intrinsic structure. Not so many structures are really basic, and one may hope that these will remain of interest for a very long time."
- "The success of the approach of studying minimal structures has ultimately to be judged by its results."

Context of Talagrand mathematical youth

- PhD thesis with Choquet on functional analysis, "but the field of his mathematics was ending and not beginning".
- Pisier showed him links between the geometry of Banach spaces, probability (mostly Gaussian measures), random Fourier series.
- Pisier introduced him to the field of Gaussian processes around 1983. Few years later, Talagrand had solved the major conjecture in the field and drew far-reaching consequences in the above fields.
- Around the same time, Milman convinced him of the importance of concentration of measure.
- He also revolutionized the field by introducing "elementary enough" yet powerful ideas.

Talagrand’s philosophy

- Les variables gaussienses sont des compas qui arpentent le globe terrestre en tous sens, lui donnant son équilibre et son harmonie.
- Gaussian variables are a minimal structure that arise everywhere: in Probability, in Geometry, in Functional Analysis, in Statistical Physics.
- Gaussian inequalities serve as an idealised model, from which other models can be studied by “perturbation”.
- Quantifying deviations from or similarities to the Gaussian case allows to classify other models / space.

Crash course on Gaussian analysis

The Gaussian space

- The standard Gaussian measure on $\mathbb{R}$ is the probability measure

$$\gamma(\mathrm{d}x) := \exp\left(-\frac{x^2}{2}\right)(2\pi)^{-1/2}.$$

- The Gaussian space is $(\Omega, \mathfrak{W}, \mathbf{P}) := (\mathbb{R}^{\mathbb{N}}, \mathfrak{B}(\mathbb{R}^{\mathbb{N}}), \gamma^{\otimes \mathbb{N}})$.
- We consider the coordinates

$$g_i(\omega) := \omega_i, \qquad \omega = (\omega_i) \in \Omega.$$

- (g_i) is a sequence of independent standard Gaussian variables.

Why are Gaussians important

- $\hat{\gamma}(t) \coloneqq \int \mathrm{e}^{\mathrm{i}tx}\gamma(\mathrm{d}x) = \mathrm{e}^{-t^2/2} \simeq \gamma(t)$.
- Central limit theorem, statistical tests.
- Free case (no interaction) of infinite-dimensional models (statistical mechanics, SPDEs).
- Gaussian variables represent *all* the metric spaces embedded in the Hilbert space.

Theorem
$\ell^2(\mathbb{N})$ embeds bi-Lipschitz in $L^q(0,1)$, $1 \leq q < \infty$.

Theorem

$\ell^2(\mathbb{N})$ *embeds bi-Lipschitz in* $L^q(0,1)$, $1 \le q < \infty$.

Proof.

- $X : \ell^2(\mathbb{N}) \ni t \mapsto X_t := \sum_i t_i g_i$.
- Converging in $L^2(\mathbf{P})$, and $\mathbf{Law}[X_t] = \mathbf{Law}\big[\|t\|_{\ell^2} g_1\big]$.
- Thus, $\|X_t\|_{L^q} = \|t\|_{\ell^2} \Gamma\big(\frac{q+1}{2}\big)^{1/q} \pi^{-q/2} 2^{-1/2}$.

□

Gaussian processes

- For $T \subset \ell^2(\mathbb{N})$, the Gaussian process over T is $(X_t := \sum_i t_i g_i \,:\, t \in T)$.
- $d(t,s) := \|s-t\|_{\ell^2(\mathbb{N})} = \mathbf{E}\big[(X_t - X_s)^2\big]^{1/2}$.
- In general, one is interested in process indexed not by $\ell^2(\mathbb{N})$ but rather by:
  - the time $\mathbb{R}_+ \rightsquigarrow$ Brownian motion,
  - the complex plane $\mathbb{C} \rightsquigarrow$ Bargman–Fock,
  - or even a function space $\rightsquigarrow$ the Gaussian free field.

  It is known all Gaussian processes arise from the Hilbert space representation.
- Important to forget any Euclidean or metric structure on the initial index set and work in the Hilbert setting. Otherwise, the ideas are *polluted* by the extra structure that one would want to use inevitably.

Supremum of Gaussian processes

Supremum of Gaussian processes

- What is the order of $S(T) \coloneqq \mathbf{E} \sup_{t \in T} X_t - X_{t_0}$?
- $S(T) = \int_0^\infty \mathbf{P}\big[\sup_{t \in T}(X_t - X_{t_0}) \geq u\big] \mathrm{d}u$.
- $\mathbf{P}\big[\sup_{t \in T}(X_t - X_{t_0}) \geq u\big] \leq \sum_{t \in T} \mathbf{P}\big[(X_t - X_{t_0}) \geq u\big]$.

Key Fact

$$\mathbf{P}[|X_t - X_s| \geq u] \leq 2 \exp\left(-\frac{u^2}{2d(s,t)^2}\right),$$

with $d(s,t) = \mathbf{E}\big[(X_t - X_s)^2\big]^{1/2} = \|t - s\|_{\ell^2(\mathbb{N})}$.

- Before Talagrand, results by Kolmogorov, Dudley, Sudakov, Fernique and more based on the distance d and the chaining.

Chaining 101

- T finite.
- Consider an increasing sequence (T_n) of *approximating sets*.
- For all n, choose $\pi_n(t) \in T$.
- For n large enough $T_n = T$ thus $\pi_n(t) = t$.
- Trivially $X_t - X_{t_0} = \sum_n (X_{\pi_n(t)} - X_{\pi_{n-1}(t)})$.
- Different choices of (T_n) and (π_n) together with the key fact

$$\mathbf{P}[|X_t - X_s| \geq u] \leq 2\exp\left(-\frac{u^2}{2d(s,t)^2}\right),$$

yield different bound.

Supremum of Gaussian processes before Talagrand

Theorem (Dudley)
Choose (T_n) *with cardinality as small as possible such that* $\forall t \in T, \exists s \in T_n, d(t,u) \leq 2^{-n}$, *then*

$$S(T) := \mathbf{E} \sup_{t \in T} X_t \leq C \sum_n 2^{-n} \sqrt{\log |T_n|}.$$

Theorem (Fernique)
If $|T_n| \leq 2^{2^n}$, *then* $S(T) \leq C \sup_{t \in T} \sum_n 2^{n/2} d(t, T_n)$. *Equivalently, with*

$$\gamma_p(T,d) := \inf_{(T_n)} \sup_{t \in T} \sum_n 2^{n/p} d(t, T_n),$$

$$S(T) \leq C\gamma_2(T,d).$$

Talagrand's main contribution

Theorem (Talagrand)

$$\frac{1}{C}\gamma_2(T,d) \leq S(T) \leq C\gamma_2(T,d).$$

- γ_2 is the correct object to capture the size of a Gaussian process.
- Despite being very naive, the chaining when done optimal completely settles the question of the size of a Gaussian process.

Consequences of Talagrand's result and approach

- Comparison theorems: if a non-Gaussian process (Y_t) satisfies $\mathbf{P}[|Y_s - Y_t| \geq u] \leq 2\exp\left(-\frac{u^2}{2d^2(s,t)}\right)$, then $\mathbf{E}\sup_{t\in T} Y_t \leq C\gamma_2$.
- *Stable* processes: same as Gaussian but $\ell^2 \leftrightarrow \ell^p$ $(1 < p < 2)$, then similar results but $\gamma_2 \leftrightarrow \gamma_q$ with $\frac{1}{p} + \frac{1}{q} = 1$.
- *Deterministic* orthogonal series and random Fourier series.
- Geometry of metric spaces: $\gamma_2(T,d)$ for other spaces.
- The *Bernoulli conjecture* about all the possible ways to bound the process $\sum \varepsilon_i t_i$ (solved by Bednorz and Latała in 2014).
- See Talagrand's book *Lower Bounds for Stochastic Processes*.

Concentration of measure

The philosophy of the concentration of measure

- In a metric measure space (E, d, μ) trying to quantify how fast the measure of a set increases as the set is enlarged metrically.
- Similarly control of the deviation of Lipschitz functions $\mu(|f - \mu(f)| \geq t)$.
- In high dimension Lipschitz functions are almost constant.
- Similarly, if $\mu(A) \geq 1/2$, then $A_t := \{x \in E : d(x, A) \leq t\}$ is almost the whole space.

Examples

- On the round sphere $\mathbb{S}^{n-1}$, with normalised Haar measure σ_{n-1}

$$\sigma_{n-1}(|f - \sigma_{n-1}(f)| \geq t) \leq \left(\frac{\pi}{8}\right)^{1/2} \exp\bigl(-(n-2)t^2/2\bigr).$$

- On $\mathbb{R}^n$, with the standard Gaussian measure γ_n

$$\gamma_n(|f - \gamma_n(f)| \geq t) \leq \frac{1}{2}\mathrm{e}^{-t^2/2}.$$

Some issues with concentration of measure

- Dimension dependence: (E, d, μ) might satisfy the concentration but not the product space (E^n, d^n, μ^n).
- Distance dependent: need a distance on E which is not always here when studying probability, especially discrete models.
- No necessary and sufficient criterion: can be deduced from isoperimetric inequalities and other geometric functional inequalities, such as the Poincaré / spectral gap, but in general it is much weaker.

Talagrand's main contributions

- Talagrand inequality: characterizing dimension-free Gaussian concentration of measures in terms of relative entropy and optimal transport.
- Concentration in product spaces: on a arbitrary probability space (E, μ) construct some distances on the product space and quantifies the associated concentration of measure phenomenon.

Concentration in product space

- Too technical to give details but the distance looks like

$$d_n((x_i),(y_i)) = \sum_{i=1}^{n} h(x_i, y_i) 1_{x_i \neq y_i}.$$

- Applied by Talagrand to solve major questions regarding:
  - Percolation.
  - Spin glass theory.
  - Random graphs.
- Still highly used today. Still being understood and revisited.

Talagrand inequality

- Relative entropy: $\mathbf{Ent}(\nu \mid \mu) := \int \log \frac{\mathtt{d}\nu}{\mathtt{d}\mu} \mathtt{d}\nu$.
- Transportation cost: $\mathcal{T}_2(\nu_1, \nu_2) := \inf \int d^2(x, y) \pi(\mathtt{d}x \mathtt{d}y)$, where the infimum over all the couplings π of μ and ν.
- μ satisfies the Talagrand inequality provided

$$\mathcal{T}_2(\nu_1, \nu_2) \leq 2\,\mathbf{Ent}(\nu_1 \mid \mu) + 2\,\mathbf{Ent}(\nu_2 \mid \mu), \qquad \forall \nu_1, \nu_2.$$

Talagrand inequality

Theorem (Talagrand)

- *If μ satisfies the Talagrand inequality, then so does μ^n on the product space.*
- *If μ satisfies the Talagrand inequality, then it satisfies Gaussian concentration of measure*

$$\mu(|f - \mu(f)| \geq t) \leq \frac{1}{2}\mathrm{e}^{-t^2/2}.$$

- *The standard Gaussian measure satisfies the Talagrand inequality.*

Legacy of Talagrand's inequality

- In my opinion, the most important inequality in analysis and probability together with the logarithmic Sobolev inequality.
- Shaped the theory of optimal transport, functional inequalities, and more up to today.
  - Ricci curvature in non-smooth continuous spaces.
  - Relationship with coercive inequalities.
  - Transport-entropy inequalities, mostly for discrete spaces.
  - Theory of boolean functions.
- What happens when one changes the transportation or the entropy.
- Non-product, that is non-independent, models.
  - Random matrix.
  - Statistical physics.
  - Point processes.

Sketch of proof for the Gaussian

- Inequality on the two-point space $\{-1, +1\}$ with $\mu = \frac{1}{2}(\delta_{-1} + \delta_1)$.
- Use the stability by products to write an inequality for μ^n.
- Use an easily-established stability by Lipschitz maps to write an inequality for pushforward of μ^n by $\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$.
- Take $n \to \infty$ and use the central limit theorem.

Proof of the concentration of measure

- Take $A \subset E$ and $B = E \setminus A_t$ for $t > 0$.
- Take $\nu_1 = \frac{1_A}{\mu(A)}\mu$ and $\nu_2 = \frac{1_B}{\mu(B)}\mu$.
- Then $\mathcal{T}_2(\nu_1, \nu_2) \geq t^2$.
- $\mathbf{Ent}(\nu_1 \mid \mu) = -\log \mu(A)$,
  $\mathbf{Ent}(\nu_2 \mid \mu) = -\log \mu(B) = -\log(1 - \mu(A_t))$.
- $t^2 \leq -2\log \mu(A) - 2\log(1 - \mu(A_t))$.

Spin glasses and the Parisi formula

Statistical mechanics 101

- System of spins $\sigma = (\sigma_i)$ with $\sigma_i \in \{\pm 1\}$.
- Interacting with only pairwise interaction.
- Undergoing a thermal agitation at temperature $1/\beta$.
- With an external force h.
- The Hamiltonian for n spins is

$$H_n(\sigma) \coloneqq -\frac{1}{\sqrt{n}} \sum_{i<j} J_{ij} \sigma_i \sigma_j - h \sum \sigma_i.$$

- The partition function

$$Z_{n,\beta} \coloneqq \sum_{\sigma} \exp(-\beta H_n(\sigma)),$$

- The free energy, if it exists

$$F(\beta) \coloneqq \lim_{n \to \infty} \frac{1}{n} \log Z_{n,\beta}.$$

The Sherrington–Kirkpatrick Hamiltonian

- Here the interactions are *random*, more precisely with g_{ij} independent standard Gaussian variables

$$H_n(\sigma) = -\frac{1}{\sqrt{n}} \sum_{i<j} g_{ij}\sigma_i\sigma_j - h\sum_i \sigma_i.$$

- This can be seen as a Ising model in a disordered, that is random, environment.
- Thus the partition function is a priori random and one is interested by the average

$$\frac{1}{n}\,\mathbf{E}\left[\log\sum_\sigma \exp(-\beta H_n(\sigma))\right].$$

- It was known that the large n limit $f(\beta)$ exists, although the non average (random) limit $F(\beta)$ might not exist.

The Parisi ansatz

- For reasons I don't understand, the physicist had conjectured that

$$f(\beta) = \inf_{q \geq 0} \left\{ \frac{\beta^2}{4}(1-q)^2 + \int \log\cosh(x\beta\sqrt{q} + h)\gamma(\mathrm{d}x) \right\} + \log 2.$$

- Talagrand managed to rigorously prove this formula.
- Don't really understand the proof. Some relations with the supremum of Gaussian processes: for $h = 0$, $\lim_{\beta\to\infty} \frac{1}{\beta} \log Z_{n,\beta} = \sup_\sigma H_n(\sigma)$.
- Behind the mere proof brought in new ideas to study spin glasses.

Concluding thoughts

Things I did not talk about

- Lots of works in functional analysis in the 1970s
- Work on Maharam's conjecture.
    - Around 75s suggested that the conjecture was false and proposed an alternative.
    - His conjecture proved by others in the 80s.
    - Disproved himself the initial conjecture in 2008!
- The matching problem.
- Interest for the theory of computation and information.
- Probably more I am not even aware of!

What Talagrand teaches us about mathematics

- Gaussians serve as a fundamental and unifying element in mathematics.
- The boundary of mathematical fields are porous and ambiguous. Probability could show up when you least expect it!
- Maths is about hard-work, patience , perseverance, starting from simple concept, and a bit of luck.
- Next time you are stuck on a problem:
    - Look for that minimalistic structure in which your problem makes sense.
    - See if you can then formulate it, and hopefully solve it, for Gaussians.
    - Go back to your initial problem with the extra knowledge you gain from the understanding of the Gaussian case.

Thank you for your attention