A large deviation principle for empirical measures

David García-Zelada

Aix-Marseille University

February 3, 2021
Point Distributions Webinar

Outline

Table of Contents

On the sphere

In the two-dimensional sphere $S^2 = \{x \in \mathbb{R}^3 : \|x\| = 1\}$, define

$$G : S^2 \times S^2 \to (-\infty, \infty] \qquad \text{by} \qquad G(x,y) = -\log \|x-y\|_{\mathbb{R}^3}$$

and let σ be the uniform probability measure on S^2. Choose $\beta > 0$.

Coulomb gas measure

$$\mathrm{d}\mathbb{P}_n^\beta(x_1,\ldots,x_n) = \frac{1}{\mathcal{Z}} e^{-\beta \sum_{i<j}^n G(x_i,x_j)} \mathrm{d}\sigma^{\otimes n}(x_1,\ldots,x_n).$$

Let $(X_1^{(n)},\ldots,X_n^{(n)}) \sim \mathbb{P}_n^\beta$. Then, for every $f : S^2 \to \mathbb{R}$ continuous

$$\frac{1}{n}\sum_{i=1}^n f(X_i^{(n)}) \xrightarrow[n\to\infty]{\text{a.s.}} \int_{S^2} f \mathrm{d}\sigma.$$

On general manifolds

Interesting property:

$$\Delta_{S^2} G(x, \cdot) = 2\pi(\sigma - \delta_x).$$

More explicitly, for every smooth $f : S^2 \to \mathbb{R}$,

$$\int_{S^2} G(x,y)\Delta_{S^2} f(y)\mathrm{d}\sigma(y) = 2\pi \left(\int_{S^2} f \mathrm{d}\sigma - f(x) \right).$$

'Physically',

> $G(x, \cdot) =$ *electrostatic potential* generated by a point charge at x and by $-\sigma$.

We could have chosen any other 'background charge'...
in any other manifold...

Coulomb gas system

M : compact Riemannian manifold of volume one,

σ : Riemannian volume measure on M,

Λ : signed differentiable measure on M s.t. $\Lambda(M) = 1$.

Definition (Green function)

$G : M \times M \to (-\infty, \infty]$ continuous symmetric s.t.

$$\Delta_M G(x, \cdot) = \Lambda - \delta_x.$$

Coulomb gas measure

$$\mathrm{d}\mathbb{P}_n^\beta(x_1, \ldots, x_n) = \frac{1}{\mathcal{Z}} e^{-\beta \sum_{i<j}^n G(x_i, x_j)} \mathrm{d}\sigma^{\otimes n}(x_1, \ldots, x_n).$$

What can we say about $\mathbb{P}_n^\beta$?

- Let $(X_1^{(n)}, \dots, X_n^{(n)}) \sim \mathbb{P}_n^\beta$ for $\beta > 0$ and $\Lambda \geq 0$. Then,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i^{(n)}) \xrightarrow[n\to\infty]{\text{a.s.}} \int_M f \mathrm{d}\Lambda$$

for every $f : M \to \mathbb{R}$ continuous.

- If $\beta = 0$ then $(X_1^{(n)}, \dots, X_n^{(n)})$ are independent with law σ and

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i^{(n)}) \xrightarrow[n\to\infty]{\text{a.s.}} \int_M f \mathrm{d}\sigma,$$

which is the law of large numbers.

There must be a regime interpolating between Λ and σ.

Indeed! Define

$$H_n(x_1, \ldots, x_n) = \frac{1}{n^2} \sum_{i<j}^{n} G(x_i, x_j)$$

and redefine the Coulomb gas measure.

Coulomb gas measure

$$\mathrm{d}\mathbb{P}_n^{\beta_n} = \frac{1}{\mathcal{Z}} e^{-n\beta_n H_n} \mathrm{d}\sigma^{\otimes n}.$$

Suppose $\lim_{n\to\infty} \beta_n = \beta \in (0, \infty)$.

Then, for every $f : M \to \mathbb{R}$ continuous,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i^{(n)}) \xrightarrow[n\to\infty]{\text{a.s.}} \int_M f \mathrm{d}\mu_\beta,$$

where μ_β has a smooth density ρ_β with respect to σ that satisfies

$$\Delta_M \log \rho_\beta = \beta\mu_\beta - \beta\Lambda.$$

Table of Contents

General setting

M : Polish space (i.e., separable and completely metrizable),

σ : probability measure on M,

H_n : measurable function from M^n to $(-\infty, \infty]$.

Non-normalized Gibbs measure

$$\mathrm{d}\gamma_n^{\beta_n} = e^{-n\beta_n H_n} \mathrm{d}\sigma^{\otimes n}.$$

Endow $\mathcal{P}(M)$, the set of probability measures on M, with the *weak topology*, i.e., the **smallest topology** s.t.

$$\mu \in \mathcal{P}(M) \mapsto \int_M f \mathrm{d}\mu \in \mathbb{R}$$

is continuous for every $f : M \to \mathbb{R}$ bounded continuous.

Let $(X_1^{(n)}, \dots, X_n^{(n)}) \sim \frac{\gamma_n^{\beta_n}}{\gamma_n^{\beta_n}(M^n)}$. We want to find ν such that

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i^{(n)}) \xrightarrow[n\to\infty]{\text{a.s.}} \int_M f \, \mathrm{d}\nu$$

for every $f : M \to \mathbb{R}$ bounded continuous.

Equivalently,

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i^{(n)}} \xrightarrow[n\to\infty]{\text{a.s.}} \nu.$$

For this, H_n cannot be arbitrary, we need a notion of limit.

Notion of limit

- Sequence $\{H_n\}_{n\in\mathbb{N}}$ **uniformly bounded from below**,
- $H : \mathcal{P}(M) \to (-\infty, \infty]$.

Definition (Macroscopic limit)

H is the *macroscopic limit* of $\{H_n\}_{n\in\mathbb{N}}$ if

- $\forall \mu \in \mathcal{P}(M)$

$$\lim_{n\to\infty} \int_{M^n} H_n \mathrm{d}\mu^{\otimes n} = H(\mu) \quad \text{and}$$

- whenever $\frac{1}{n}\sum_{i=1}^n \delta_{x_i} \to \mu$

$$\liminf_{n\to\infty} H_n(x_1, \ldots, x_n) \geq H(\mu).$$

Theorem (Laplace principle, G-Z (2019))

Suppose

- H *is the macroscopic limit of* $\{H_n\}_{n\in\mathbb{N}}$ *and*
- $\{\beta_n\}_{n\in\mathbb{N}}$ *converges to some* $\beta \in (0,\infty)$.

Then, for every bounded continuous $f : \mathcal{P}(M) \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n\beta_n} \log \int_{M^n} e^{-n\beta_n f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right)} \mathrm{d}\gamma_n^{\beta_n}(x_1,\ldots,x_n)$$

$$= -\inf_{\substack{\mu\in\mathcal{P}(M)\\ \mathrm{d}\mu=\rho\,\mathrm{d}\sigma}} \left\{ f(\mu) + H(\mu) + \frac{1}{\beta}\int_M \rho\log\rho\,\mathrm{d}\sigma \right\}.$$

Free energy

Define the *relative entropy* of μ with respect to σ as

$$D(\mu\|\sigma) = \begin{cases} \int_M \rho \log \rho \, \mathrm{d}\sigma & \text{if } \mathrm{d}\mu = \rho \, \mathrm{d}\sigma \\ \infty & \text{if there is no such } \rho \end{cases}$$

and the *free energy* as

$$F_\beta(\mu) = H(\mu) + \frac{1}{\beta} D(\mu\|\sigma).$$

Large deviation principle

Using the free energy definition,

$$\lim_{n\to\infty} \frac{1}{n\beta_n} \log \int_{M^n} e^{-n\beta_n f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right)} \mathrm{d}\gamma_n^{\beta_n}(x_1,\ldots,x_n)$$

$$= -\inf_{\mu\in\mathcal{P}(M)} \{f(\mu) + F_\beta(\mu)\}.$$

For $f = 0$ we get $\lim_{n\to\infty} \frac{1}{n\beta_n} \log \gamma_n^{\beta_n}(M^n) = -\inf F_\beta$.

$$\lim_{n\to\infty} \frac{1}{n\beta_n} \log \int_{M^n} e^{-n\beta_n f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right)} \frac{\mathrm{d}\gamma_n^{\beta_n}(x_1,\ldots,x_n)}{\gamma_n^{\beta_n}(M^n)}$$

$$= -\inf_{\mu\in\mathcal{P}(M)} \{f(\mu) + F_\beta(\mu) - \inf F_\beta\}.$$

Suppose $(X_1^{(n)}, \dots, X_n^{(n)}) \sim \frac{\gamma_n^{\beta_n}}{\gamma_n^{\beta_n}(M^n)}$ and $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^{(n)}}$.

Theorem (Large deviation principle, G-Z (2019))

Suppose

- H *is the macroscopic limit of* $\{H_n\}_{n\in\mathbb{N}}$,
- $\{\beta_n\}_{n\in\mathbb{N}}$ *converges to some* $\beta \in (0,\infty)$ *and*
- $\inf F_\beta < \infty$.

Then, for every bounded continuous $f : \mathcal{P}(M) \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n\beta_n} \log \mathbb{E}\left[e^{-n\beta_n f(\hat{\mu}_n)}\right] = -\inf\left\{f + F_\beta - \inf F_\beta\right\}.$$

Why does this imply an almost sure convergence?

Almost sure convergence

Suppose $(X_1^{(n)}, \dots, X_n^{(n)}) \sim \frac{\gamma_n^{\beta_n}}{\gamma_n^{\beta_n}(M^n)}$ and $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i^{(n)}}$.

Corollary

Suppose

- H *is the macroscopic limit of* $\{H_n\}_{n\in\mathbb{N}}$*,*
- $\{\beta_n\}_{n\in\mathbb{N}}$ *converges to some* $\beta \in (0, \infty)$ *and*
- F_β *has a unique minimizer* ν*.*

Then,

$$\hat{\mu}_n \xrightarrow[n\to\infty]{\text{a.s.}} \nu.$$

Yes, but why?

Important:

- $\mathcal{P}(M)$ is metrizable and
- $\{\mu \in \mathcal{P}(M) : F_\beta(\mu) \leq a\}$ is compact for every $a \in \mathbb{R}$.

Suppose F_β has a unique minimizer ν. We want to show that ...

> almost surely, for every open neighborhood U of ν,
> $\hat{\mu}_n \notin U$ only for a finite number of n.

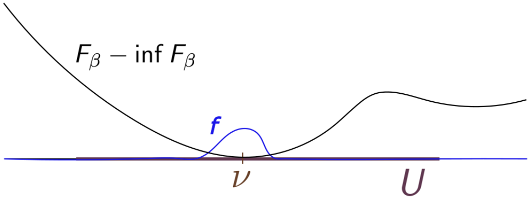
If $C = \inf\{f + F_\beta - \inf F_\beta\}$,

$$\lim_{n\to\infty} \frac{1}{n\beta_n} \log \mathbb{E}\left[e^{-n\beta_n f(\hat{\mu}_n)}\right] = -C$$

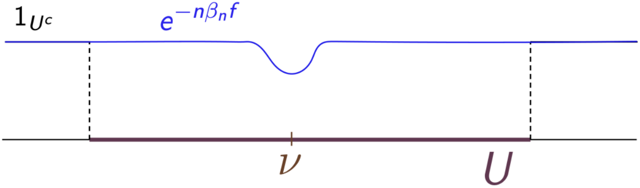
or, equivalently,

$$\mathbb{E}\left[e^{-n\beta_n f(\hat{\mu}_n)}\right] = e^{-Cn\beta_n(1+o(1))}.$$

- Take $U \subset \mathcal{P}(M)$ open that contains ν
- Take $f : \mathcal{P}(M) \to [0, \infty)$ s.t. $f|_{U^c} = 0$ and $f(\nu) > 0$.
- Notice that $C := \inf \{f + F_\beta - \inf F_\beta\} > 0$.



- Notice that $1_{\hat{\mu}_n \notin U} \leq e^{-n\beta_n f(\hat{\mu}_n)}$.



Then,

$$\mathbb{E}\left[1_{\hat{\mu}_n \notin U}\right] \leq \mathbb{E}\left[e^{-n\beta_n f(\hat{\mu}_n)}\right] = e^{-Cn\beta_n(1+o(1))}$$

so that $\mathbb{E}\left[\sum_{n=1}^{\infty} 1_{\hat{\mu}_n \notin U}\right] < \infty$, which implies $\sum_{n=1}^{\infty} 1_{\hat{\mu}_n \notin U} < \infty$ a.s. or, equivalently, almost surely:

$\hat{\mu}_n \notin U$ only for a finite number of n.

Conclude by taking a countable neighbourhood basis of ν.

Coulomb gas case

Recall that $G : M \times M \to (-\infty, \infty]$ continuous symmetric s.t.

$$\Delta_M G(x, \cdot) = \Lambda - \delta_x$$

and $$H_n(x_1, \ldots, x_n) = \frac{1}{n^2} \sum_{i<j} G(x_i, x_j).$$

Integrating, $$\int_{M^n} H_n \mathrm{d}\mu^{\otimes n} = \frac{n(n-1)}{2n^2} \int_{M \times M} G(x, y) \mathrm{d}\mu(x) \mathrm{d}\mu(y).$$

Macroscopic energy

$$H(\mu) = \frac{1}{2} \int_{M \times M} G(x, y) \mathrm{d}\mu(x) \mathrm{d}\mu(y).$$

Free energy

$$F_\beta(\mu) = \frac{1}{2}\int_{M\times M} G(x,y)\mathrm{d}\mu(x)\mathrm{d}\mu(y) + \frac{1}{\beta}\int_M \rho \log \rho \, \mathrm{d}\sigma.$$

What happens is that F_β is strictly convex and

$$\Delta_M \left(\nabla_\mu F_\beta\right) = \Lambda - \mu + \frac{1}{\beta}\Delta_M \log \rho$$

so that, if μ_β is the minimizer of F_β,

$$0 = \Lambda - \mu_\beta + \frac{1}{\beta}\Delta_M \log \rho_\beta.$$

Infinite β

What happens when $\beta_n \to \infty$? Two more conditions.

- $\{H_n\}_{n\in\mathbb{N}}$ **confining**: If $\liminf_{n\to\infty} H_n(x_1, \dots, x_n) < \infty$ then

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \right\}_{n\in\mathbb{N}} \quad \text{has a convergent subsequence.}$$

- H **regular**: If $H(\mu) < \infty$ then
  $\exists \, \{\mu_n\}_{n\in\mathbb{N}}$ s.t. $\mu_n \to \mu$,

$$\forall n, D(\mu_n \| \sigma) < \infty \text{ and } \lim_{n\to\infty} H(\mu_n) = H(\mu).$$

Theorem (Laplace principle, G-Z (2019))

Suppose

- H *is the macroscopic limit of* $\{H_n\}_{n\in\mathbb{N}}$,
- $\{H_n\}_{n\in\mathbb{N}}$ *is confining,*
- H *is regular and*
- $\beta_n \to \infty$.

Then, for every bounded continuous $f : \mathcal{P}(M) \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n\beta_n} \log \int_{M^n} e^{-n\beta_n f(\frac{1}{n}\sum_{i=1}^n \delta_{x_i})} \mathrm{d}\gamma_n^{\beta_n}(x_1, \ldots, x_n)$$

$$= - \inf_{\mu\in\mathcal{P}(M)} \{f(\mu) + H(\mu)\} .$$

Suppose $(X_1^{(n)}, \ldots, X_n^{(n)}) \sim \frac{\gamma_n^{\beta_n}}{\gamma_n^{\beta_n}(M^n)}$ and $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^{(n)}}$.

Corollary

Suppose

- H *is the macroscopic limit of* $\{H_n\}_{n\in\mathbb{N}}$,
- $\{H_n\}_{n\in\mathbb{N}}$ *is confining,*
- H *is regular,*
- $\beta_n \to \infty$ *and*
- H *has a unique minimizer* ν.

Then,

$$\hat{\mu}_n \xrightarrow[n\to\infty]{\text{a.s.}} \nu.$$

Coulomb gas case

Macroscopic energy

$$H(\mu) = \frac{1}{2}\int_{M\times M} G(x,y)\mathrm{d}\mu(x)\mathrm{d}\mu(y).$$

What happens is that H is strictly convex and

$$\Delta_M(\nabla_\mu H) = \Lambda - \mu$$

so that, if $\Lambda \geq 0$, the minimizer of H is Λ.

Table of Contents

Deterministic results

Theorem (Γ-convergence)

Suppose

- H *is the macroscopic limit of* $\{H_n\}_{n\in\mathbb{N}}$ *and*
- $\{H_n\}_{n\in\mathbb{N}}$ *is confining.*

Then, for every bounded continuous $f : \mathcal{P}(M) \to \mathbb{R}$*,*

$$\lim_{n\to\infty} \inf_{(x_1,\dots,x_n)\in M^n} \left\{ H_n(x_1,\dots,x_n) + f\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}\right)\right\} = \inf_{\mu\in\mathcal{P}(M)} \{H(\mu) + f(\mu)\}.$$

Let $(X_1^{(n)}, \ldots, X_n^{(n)})$ satisfy

$$\lim_{n\to\infty} \left| H_n(X_1^{(n)}, \ldots, X_n^{(n)}) - \inf H_n \right| = 0$$

and define $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^{(n)}}$.

Corollary (Convergence of almost minimizers)

Suppose

- H *is the macroscopic limit of* $\{H_n\}_{n\in\mathbb{N}}$,
- $\{H_n\}_{n\in\mathbb{N}}$ *is confining and*
- H *has a unique minimizer* ν.

Then,

$$\hat{\mu}_n \xrightarrow[n\to\infty]{} \nu.$$

Deterministic vs. probabilistic version

$$\inf_{(x_1,\dots,x_n)\in M^n}\left\{H_n(x_1,\dots,x_n)+f\left(\frac{1}{n}\sum_{i=1}^n\delta_{x_i}\right)\right\}$$

$$=\inf_{\nu\in\mathcal{P}(M^n)}\left\{\int_{M^n}H_n\mathrm{d}\nu+\int_{M^n}f\left(\frac{1}{n}\sum_{i=1}^n\delta_{x_i}\right)\mathrm{d}\nu(x_1,\dots,x_n)\right\}$$

while. . .

$$
\begin{aligned}
&-\frac{1}{n\beta_n}\log\int_{M^n} e^{-n\beta_n f(\frac{1}{n}\sum_{i=1}^n \delta_{x_i})}\mathrm{d}\gamma_n^{\beta_n}(x_1,\dots,x_n) \\
&= \inf_{\nu\in\mathcal{P}(M^n)}\Bigg\{\int_{M^n} H_n \mathrm{d}\nu + \int_{M^n} f\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right)\mathrm{d}\nu(x_1,\dots,x_n) \\
&\qquad + \frac{1}{n\beta_n} D(\nu\|\sigma^{\otimes n})\Bigg\}.
\end{aligned}
$$

You may find more in

G-Z. *A large deviation principle for empirical measures on Polish spaces: Application to singular Gibbs measures on manifolds.* Annales de l'IHP, Probabilités et Statistiques **55** (2019).

Inspiring article

Dupuis, Laschos, and Ramanan. *Large deviations for configurations generated by Gibbs distributions with energy functionals consisting of singular interaction and weakly confining potentials.* Electronic Journal of Probability **25** (2020).

Inspiring book

Dupuis and Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley series in probability and statistics (1997).

Thank you for your attention!