Limit theorems & concentration inequalities for Gibbs measures on shifts of finite type

Jean-René Chazottes

Centre de Physique Théorique de l'Ecole polytechnique

Porquerolles, September 26-30, 2021

cnrs
dépasser les frontières

ÉCOLE POLYTECHNIQUE
UNIVERSITÉ PARIS-SACLAY

INSTITUT POLYTECHNIQUE DE PARIS

The purpose of these two mini-courses is to introduce, in a very comfortable setting, the basic ideas to deal with limit theorems and concentration inequalities in dynamical systems.

The good news is that most of the ideas we introduce survive for much more general dynamical systems.

Tentative plan

Lecture 1 : The basic theory and limit theorems for Birkhoff sums

- Ruelle's Perron-Frobenius theorem
- Gibbs measures and their basic properties
- Equilibrium states and the variational principle
- Large deviation asymptotics and central limit asymptotics for Birkhoff sums

Lecture 2 : Gaussian concentration bound and some applications

- Gaussian concentration
- A kind of shadowing
- Empirical measure
- Beyond subshifts of finite type and Gibbs measures, and beyong Gaussian concentration

Let Ω be a non-empty finite set. Points $x \in \Omega$ are called “configurations”.

Given a probability measure ν on Ω, which is simply a probability vector here, we define its entropy

$$s(\nu) = -\sum_{x \in \Omega} \nu(x) \log \nu(x)$$

where it is understood that $u \log u = 0$ if $u = 0$.

Given a function $\varphi : \Omega \to \mathbb{R}$ ("potential"), we define a real number $Z(\varphi)$ called the partition function and a probability measure μ_φ on Ω, called a Gibbs measure, by

$$Z(\varphi) \stackrel{\text{def}}{=} \sum_{x\in\Omega} \mathrm{e}^{\varphi(x)} \quad \text{and} \quad \mu_\varphi(x) \stackrel{\text{def}}{=} \frac{\mathrm{e}^{\varphi(x)}}{Z(\varphi)}.$$

We have the following "variational principle". The maximum of the expression

$$s(\nu) + \int \varphi \mathrm{d}\nu$$

over all probability measures ν on Ω is $P(\varphi) \stackrel{\text{def}}{=} \log Z(\varphi)$, and is reached precisely for $\nu = \mu_\varphi$. Of course

$$\int \varphi \mathrm{d}\nu \stackrel{\text{def}}{=} \sum_{x\in\Omega} \varphi(x)\nu(x).$$

Consider the one-parameter family of Gibbs measures $(\mu_{\beta\varphi})_{\beta\in\mathbb{R}}$:

$$\mu_{\beta\varphi}(x) = \frac{\mathrm{e}^{\beta\varphi(x)}}{Z(\beta\varphi)}.$$

As $\frac{\mu_{\beta\varphi}(x)}{\mu_{\beta\varphi}(x')} \to 0$ as $\beta \to +\infty$ if $\varphi(x) < \varphi(x')$, the measure $\mu_{\beta\varphi}$ converges to the equidistribution on $\Omega_{\max} \stackrel{\text{def}}{=} \{x : \varphi(x) = \max_\Omega \varphi\}$ if $\beta \to +\infty$. An analogous statements holds for $\beta \to -\infty$ with max replaced with min. It follows in particular that

$$\lim_{\beta\to+\infty} \int \varphi \, \mathrm{d}\mu_{\beta\varphi} = \max_\Omega \varphi \quad \text{and} \quad \lim_{\beta\to-\infty} \int \varphi \, \mathrm{d}\mu_{\beta\varphi} = \min_\Omega \varphi.$$

We also have that $\beta \mapsto P(\beta\varphi) \stackrel{\text{def}}{=} \log Z(\beta\varphi)$ is a real analytic map, and

$$P'(\beta\varphi) = \int \varphi \, \mathrm{d}\mu_{\beta\varphi}$$

and

$$P''(\beta\varphi) = \mathsf{Var}_{\mu_{\beta\varphi}}(\varphi) \stackrel{\text{def}}{=} \int \varphi^2 \mathrm{d}\mu_{\beta\varphi} - \left(\int \varphi \, \mathrm{d}\mu_{\beta\varphi} \right)^2 \geq 0.$$

Hence $\beta \mapsto P(\beta\varphi) = \log Z(\beta\varphi)$ is convex.

We have equality if and only if φ is constant.

A consequence: If $u^* \in \mathbb{R}$ such that $\min_\Omega \varphi < u^* < \max_\Omega \varphi$, then there exists a unique value $\beta^* \in \mathbb{R}$ such that

$$\int \varphi \, d\mu_{\beta^* \varphi} = u^*$$

and $\mu_{\beta^*\varphi}$ maximizes entropy among all the probability measures ν such that $\int \varphi \, d\nu = u^*$.

All the previous quantities and identities, appropriately defined in the context of $\Omega = A^{\mathbb{N}}$ will show up.

The goal will be to construct Gibbs measures which are left invariant by the shift map.

Lecture 1 : The basic theory and limit theorems for Birkhoff sums

A: a finite set
M: a $|A| \times |A|$ matrix of zeroes and ones where the (i, j)th entry is zero precisely when it is a forbidden word of length two.
Define

$$\Omega = \Omega_A = \left\{x = (x^i)_{i=0}^{\infty} : x^i \in A, i \geq 0, M(x^i, x^{i+1}) = 1\right\} \subseteq A^{\mathbb{N}}.$$

Example : $A = \{a, b\}$, $M(a, a) = M(a, b) = M(b, a) = 1$ and $M(b, b) = 0$, hence the word bb is forbidden in the configurations.

Shift map T: $(Tx)^n = x^{n+1}$, $n = 0, 1, \ldots$.

Then (Ω, T) is shift of finite type.

We give A the discrete topology, making Ω a compact space with the corresponding product topology which is generated by the corresponding cylinder sets

$$[a^0 a^1 \cdots a^n] = \{x \in \Omega : x^k = a^k, 0 \leq k \leq n\}$$

where $a^0, \ldots, a^n \in A$, $n = 0, 1, \ldots$.

A distance metrizing Ω:

$$\mathsf{d}_\theta(x, y) = \theta^{\inf\{k \geq 0 \,:\, x^k \neq y^k\}}$$

where $\theta \in (0, 1)$ is some fixed number.

Assumption: there exists $m \geq 1$ such that $M^m(s, s') > 0 \ \forall (s, s') \in A^2$. This is equivalent to the fact that (Ω, T) is topologically mixing.

Probability measures are defined on the Borel sigma-algebra which is generated by cylinder sets.

A probability measure μ is shift-invariant if $\mu \circ T^{-1} = \mu$.
Equivalently:

$$\int f \circ T \, \mathrm{d}\mu = \int f \mathrm{d}\mu \quad \text{for all continuous functions } f : \Omega \to \mathbb{R}.$$

The set of shift-invariant probability measures is compact in the weak topology.

Goal

Given $\varphi : \Omega \to \mathbb{R}$ "sufficiently regular", how can one construct the corresponding Gibbs measures and equilibrium states?

What about uniqueness?

What about their statistical properties?

What we mean by a Gibbs measure

Definition

A probability measure μ on Ω is called a Gibbs measure for the potential $\varphi \in \mathcal{C}(\Omega)$ if there are constants $c_\varphi \geq 1$ and $P(\varphi) \in \mathbb{R}$ such that

$$c_\varphi^{-1} \leq \frac{\mu\big([x^0 \cdots x^{n-1}]\big)}{\exp\big(-nP(\varphi) + S_n\varphi(x)\big)} \leq c_\varphi$$

for any $x = (x^i)_{i=0}^{\infty} \in \Omega$ and for any $n \geq 1$. We do not require that μ should be shift-invariant.

As usual, $S_n\varphi(x) = \sum_{j=0}^{n-1} \varphi(T^j x)$ (nth Birkhoff sum of φ under the shift).

If μ is a Gibbs measure then

$$P(\varphi) = \lim_{n\to+\infty} \frac{1}{n} \log \sum_{\substack{a^0,\dots,\, a^{n-1}\in A \\ M(a^i,a^{i+1})=1}} e^{\sup\{S_n\varphi(x)\,:\, x^i=a^i,\ i=0,\dots,n-1\}} .$$

Regular = Lipschitz

For $f \in \mathcal{C}(\Omega)$ let

$$\mathrm{var}_n(f) \stackrel{\text{def}}{=} \sup\{|f(x) - f(y)| : x^i = y^i, 0 \leq i \leq n-1\}.$$

Then $\mathrm{var}_n(f) \to 0$.

Now consider

$$\{f \in \mathcal{C}(\Omega) : \exists L > 0 \text{ such that } \mathrm{var}_n(f) \leq L\theta^n, n = 1, 2, \ldots\}$$

and

$$\mathrm{lip}(f) = \sup\left\{\frac{|f(x) - f(y)|}{\mathrm{d}_\theta(x, y)} : x \neq y\right\} = \sup\left\{\frac{\mathrm{var}_n(f)}{\theta^n} : n \in \mathbb{N}\right\}.$$

A norm making this space a Banach space is

$$\|f\| = \|f\|_\infty + \mathrm{lip}(f).$$

Given $\varphi : \Omega \to \mathbb{R}$ continuous, define Ruelle's Perron-Frobenius operator, or transfer operator, $\mathsf{R}_\varphi : \mathcal{C}(\Omega) \to \mathcal{C}(\Omega)$ as

$$\begin{aligned}\mathsf{R}_\varphi f(x) &= \sum_{y \in T^{-1}\{x\}} f(y)\, \mathrm{e}^{\varphi(y)} = \sum_{Ty=x} f(y)\, \mathrm{e}^{\varphi(y)} \\ &= \sum_{a \in A} f(ax)\, \mathrm{e}^{\varphi(ax)},\ x \in \Omega\end{aligned}$$

where $ax = ax^0x^1\ldots$.
By induction one checks that

$$\begin{aligned}\mathsf{R}_\varphi^k f(x) &= \sum_{T^k y=x} f(y)\, \mathrm{e}^{S_k\varphi(y)} \\ &= \sum_{a^0,\ldots,a^{k-1} \in A} f(a^0 \ldots a^{k-1}x)\, \mathrm{e}^{\sum_{i=0}^{k-1} \varphi(T^i(a^0\ldots a^{k-1}x))}\end{aligned}$$

where $\mathsf{R}_\varphi^1 = \mathsf{R}_\varphi$, $\mathsf{R}_\varphi^2 = \mathsf{R}_\varphi \circ \mathsf{R}_\varphi$, and so forth.

Ruelle's Perron-Frobenius theorem

Let φ be Lipschitz. Then there are $\lambda_\varphi > 0$, $K_\varphi > 1$, h_φ Lipschitz, and a measure ν_φ such that $K_\varphi^{-1} \leq h_\varphi \leq K_\varphi$ and

$$\mathrm{R}_\varphi\, h_\varphi = \lambda_\varphi h_\varphi, \quad \mathrm{R}_\varphi^*\, \nu_\varphi = \lambda_\varphi \nu_\varphi, \quad \int h_\varphi \mathrm{d}\nu_\varphi = 1.$$

Moreover, there exists a constant $c = c(\varphi)$ and $\rho = \rho(\varphi) < 1$ such that for all f Lipschitz and for all $k \geq 1$ we have

$$\left\| \lambda_\varphi^{-k}\, \mathrm{R}_\varphi^k f - \Big(\int f \mathrm{d}\nu_\varphi \Big) h_\varphi \right\| \leq c\rho^k \|f\|$$

where $\|\cdot\| := \|\cdot\|_\infty + \mathrm{lip}(\cdot)$.

Theorem.

Let φ be Lipschitz. Then

- The probability measure $\mu_\varphi \stackrel{\text{def}}{=} h_\varphi \nu_\varphi$ is shift-invariant.
- It is a Gibbs measure with $P(\varphi) = \log \lambda_\varphi$.
- It is mixing (hence ergodic), and it is the unique Gibbs measure for φ.
- It has exponential decay of correlations: there exist $D > 0$, $\gamma \in (0, 1)$ such that for f, g Lipschitz

$$\left| \int f \cdot g \circ T^n \, \mathrm{d}\mu_\varphi - \int f \, \mathrm{d}\mu_\varphi \int g \, \mathrm{d}\mu_\varphi \right| \leq D \|f\| \|g\| \gamma^n, \;\; n \geq 0.$$

Proof that μ_φ is shift-invariant

Let $f \in \mathcal{C}(\Omega)$.
Notice that for $f_1, f_2 \in \mathcal{C}(\Omega)$

$$\big((\mathrm{R}_\varphi f_1) \cdot f_2\big)(x) = \sum_{Ty=x} f_1(y)\ \mathrm{e}^{\varphi(y)} f_2(x) = \sum_{Ty=x} f_1(y)\ \mathrm{e}^{\varphi(y)} f_2(Ty)$$
$$= \mathrm{R}_\varphi\left(f_1 \cdot (f_2 \circ T)\right)(x).$$

Hence

$$\int f \mathrm{d}\mu_\varphi = \int f\, h_\varphi \mathrm{d}\nu_\varphi = \int \lambda_\varphi^{-1}\, \mathrm{R}_\varphi\, h_\varphi \cdot f\, \mathrm{d}\nu_\varphi = \lambda_\varphi^{-1} \int \mathrm{R}_\varphi \left(h_\varphi \cdot (f \circ T)\right) \mathrm{d}\nu_\varphi$$
$$= \int \left(h_\varphi \cdot (f \circ T)\right) \lambda_\varphi^{-1} \mathrm{d}(\mathrm{R}_\varphi^* \nu_\varphi) = \int \left(h_\varphi \cdot (f \circ T)\right) \mathrm{d}\nu_\varphi$$
$$= \int f \circ T\, \mathrm{d}\mu_\varphi.$$

Proof that μ_φ is a Gibbs measure (sketch)

Fix $x \in \Omega$, $n \geq 1$ and let $E \stackrel{\text{def}}{=} [x^0 \cdots x^{n-1}]$.

Then

$$\mu_\varphi(E) = \int \mathbb{1}_E h_\varphi \, d\nu_\varphi = \lambda_\varphi^{-n} \int \mathrm{R}_\varphi^n \left(\mathbb{1}_E h_\varphi\right) d\nu_\varphi$$

Now, get an upper bound and a lower bound for $\mathrm{R}_\varphi^n \left(\mathbb{1}_E h_\varphi\right)$.
(We omit the details, see Bowen's book.)

Normalization of potentials and probabilistic interpretation thereof

One can normalize φ: For $f \in \mathcal{C}(\Omega)$ let

$$\mathrm{Q}_\varphi f = \frac{\mathrm{R}_\varphi(f h_\varphi)}{\lambda_\varphi h_\varphi}.$$

Thus

$$\mathrm{Q}_\varphi 1 = 1 \quad \text{and} \quad \mathrm{Q}_\varphi^* \mu_\varphi = \mu_\varphi.$$

Let g denote the inverse of the "Jacobian" of T, and $g^{(k)}$ the inverse of the "Jacobian" of T^k, that is,

$$g = \frac{h_\varphi}{\lambda_\varphi\, h_\varphi \circ T} \exp(\varphi) \quad \text{and} \quad g^{(k)} = \frac{h_\varphi}{\lambda_\varphi^k\, h_\varphi \circ T^k} \exp\left(S_k \varphi\right) . \qquad (1)$$

Therefore

$$\mathrm{Q}_\varphi f(x) = \sum_{Ty=x} g(y) f(y) \quad \text{and} \quad \mathrm{Q}_\varphi^k f(x) = \sum_{T^k y=x} g^{(k)}(y) f(y)\,.$$

We have a Markov chain with state space Ω and the probability to jump from x to ax is $g(ax)$ (so we are looking "backward in time").

Let ν be a T-invariant probability measure. Its entropy is

$$s(\nu) = \lim_{n\to+\infty} -\frac{1}{n} \sum_{a^{0:n-1}\in A^n} \nu([a^{0:n-1}]) \log \nu([a^{0:n-1}]).$$

Definition

A shift-invariant probability measure μ is an equilibrium state for $\varphi \in \mathcal{C}(\Omega)$ if

$$s(\mu) + \int \varphi \, \mathrm{d}\mu = \sup_{\nu \text{ is } T-\text{invariant}} \left(s(\nu) + \int \varphi \, \mathrm{d}\nu \right).$$

Equilibrium states always exist.

Theorem (Variational principle for Lipschitz potentials)

Let φ be a Lipschitz potential. Then its Gibbs measure μ_φ is the unique equilibrium state for φ.

There exists some measurable set $\mathcal{T}_{\mu_\varphi} \subset \Omega$ with $\mu_\varphi(\mathcal{T}_{\mu_\varphi}) = 1$ (the set of "typical points" for μ_φ) such that

$$\frac{S_n f(x)}{n} = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \xrightarrow[n \to +\infty]{} \int f \mathrm{d}\mu_\varphi$$

for every $x \in \mathcal{T}_{\mu_\varphi}$ and every continuous function $f : \Omega \to \mathbb{R}$. This statement can reformulated by saying that

$$\mathcal{E}_n(x) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{j=0}^{n-1} \delta_{T^j x} \xrightarrow[n \to +\infty]{} \mu_\varphi$$

for every $x \in \mathcal{T}_{\mu_\varphi}$ in the weak topology sense.

Two basic questions

Take $u > 0$. At which speed does

$$\mu_\varphi\left(x \in \Omega : \frac{S_n f(x)}{n} \geq \int f \mathrm{d}\mu_\varphi + u\right)$$

decays to 0?

Does $S_n f/\sqrt{n}$ converge in law to Gaussian random variable with mean 0 and with a certain variance to be determined?

Take a continuous function f such that $\int f \mathrm{d}\mu = 0$.
We are interested in computing the exponential rate at which the μ-probability of the set of points x such that $S_n f(x)/n$ is, say, greater than $u > 0$.
We have

$$\begin{aligned}
\mu\left(x \in \Omega : \frac{S_n f(x)}{n} \geq u\right) &= \mu\big(x \in \Omega : \beta S_n f(x) \geq n\beta u\big) \quad \text{for any } \beta > 0 \\
&\leq \mathrm{e}^{-n\beta u} \int \mathrm{e}^{\beta S_n f} \, \mathrm{d}\mu \qquad \text{(by Markov's inequality)} \\
&= \exp\left(-n\left(\beta u - \frac{1}{n}\log \int \mathrm{e}^{\beta S_n f} \, \mathrm{d}\mu\right)\right).
\end{aligned}$$

Let us make a leap of faith and *assume* that the following limit exists for every $\beta \in \mathbb{R}$:

$$\kappa_f(\beta) \stackrel{\text{def}}{=} \lim_{n\to+\infty} \frac{1}{n} \log \int \mathrm{e}^{\beta S_n f}\, \mathrm{d}\mu.$$

We get

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mu \left(x \in \Omega : \frac{S_n f(x)}{n} \geq u \right) \leq -\sup_{\beta>0} \big(\beta u - \kappa_f(\beta)\big) \stackrel{\text{def}}{=} -s_f(u)$$

where s_f is thus the Legendre-Fenchel transform of κ_f.

Theorem

Let φ be a Lipschitz potential. Then,

$$\kappa_f(\beta) = P(\varphi + \beta f) - P(\varphi),\ \ \beta \in \mathbb{R}.$$

Proof:

Using Ruelle's PF theorem (first to the potential φ, and then to the potential $f + \varphi$), we have for any $n \geq 1$

$$\begin{aligned}
&\int \mathrm{e}^{\beta S_n f}\, \mathrm{d}\mu_\varphi \\
&= \int \mathrm{e}^{\beta S_n f}\, h_\varphi \mathrm{d}\nu_\varphi = \int \mathrm{e}^{\beta S_n f}\, h_\varphi \mathrm{d}\big(\lambda_\varphi^{-n}\, \mathrm{R}_\varphi^{*n}\, \nu_\varphi\big) \\
&= \lambda_\varphi^{-n} \int \mathrm{R}_\varphi^{n}\, \big(h_\varphi\, \mathrm{e}^{\beta S_n f}\,\big) \mathrm{d}\nu_\varphi = \lambda_\varphi^{-n} \int \sum_{T^n y = x} h_\varphi(y)\, \mathrm{e}^{S_n(\varphi+\beta f)(y)}\,\big) \mathrm{d}\nu_\varphi(x) \\
&= \lambda_\varphi^{-n} \int \mathrm{R}_{\varphi+\beta f}^{n}(h_\varphi) \mathrm{d}\nu_\varphi \\
&= \lambda_\varphi^{-n} \int \Big(\lambda_{\varphi+\beta f}^{n} h_{\varphi+\beta f} \Big(\int h_\varphi \mathrm{d}\nu_{\varphi+\beta f}\Big) \mathrm{d}\nu_\varphi + \mathcal{O}\big((\rho_{\varphi+\beta f} \lambda_{\varphi+\beta f})^n\big)\Big)\, \mathrm{d}\nu_\varphi.
\end{aligned}$$

Hence $\kappa_f(\beta) = \log \lambda_{\varphi+\beta f} - \log \lambda_\varphi = P(\varphi + \beta f) - P(\varphi)$. □

Theorem

Let f be Lipzchitz. Assume that f is not cohomologous to a constant, that is, *there is no* b Lipschitz and $c \in \mathbb{R}$ such that $f = c + b \circ T - b$. Then for any interval I with $I \cap (\underline{p}_f, \overline{p}_f) \neq \emptyset$ we have

$$\lim_{n \to +\infty} \frac{1}{n} \log \mu_\varphi \left(x \in \Omega : \frac{S_n f(x)}{n} \in I \right) = - \inf_{u \in I \cap (\underline{p}_f, \overline{p}_f)} s_f(u)$$

where

$$\underline{p}_f \stackrel{\text{def}}{=} \lim_{\beta \to -\infty} \frac{\mathrm{d}}{\mathrm{d}\beta} P(\varphi + \beta f) = \inf_{\nu \ T\text{-invariant}} \int f \mathrm{d}\nu$$

and

$$\overline{p}_f \stackrel{\text{def}}{=} \lim_{\beta \to +\infty} \frac{\mathrm{d}}{\mathrm{d}\beta} P(\varphi + \beta f) = \sup_{\nu \ T\text{-invariant}} \int f \mathrm{d}\nu.$$

(The case $f = c + b \circ T - b$ is special because $\|S_n f / n\|_\infty \leq c + (2\|b\|_\infty)/n$ which becomes close to c when $n \gg 1$, so $S_n f / n$ almost doesn't fluctuate.)

Central limit asymptotics of Birkhoff sums

Let f be Lipschitz such that $\int f \mathrm{d}\mu_\varphi = 0$.
One can prove that

$$\frac{\mathrm{d}P(\varphi + sf)}{\mathrm{d}s}\Big|_{s=0} = \int f \mathrm{d}\mu_\varphi = 0$$

and

$$\frac{\mathrm{d}^2 P(\varphi + sf)}{\mathrm{d}s^2}\Big|_{s=0} = \lim_{n\to+\infty} \frac{1}{n} \int \left(S_n f\right)^2 \mathrm{d}\mu_\varphi \stackrel{\mathrm{def}}{=} \sigma_f^2$$

and

$$\sigma_f^2 = \int f^2 \mathrm{d}\mu_\varphi + 2 \sum_{j\geq 1} \int f \cdot f \circ T^j \mathrm{d}\mu_\varphi < \infty$$

where $\int f \cdot f \circ T^j \mathrm{d}\mu_\varphi$ decays expoentially fast to 0.

Theorem

The variance σ_f^2 is equal to 0 if and only f is cohomologous to a constant, that is, there exist b Lipschitz, $c \in \mathbb{R}$ such that $f = c + b \circ T - b$. The function $s \mapsto P(\varphi + sf)$ $(s \in \mathbb{R})$ is convex, and strictly convex if $\sigma_f^2 \neq 0$.

Theorem (Berry-Esseen inequality for Gibbs measures)
Let f be Lipschitz such that $\int f \mathrm{d}\mu_\varphi = 0$. Assume that $\sigma_f^2 \neq 0$. Then, uniformly in $u \in \mathbb{R}$, we have

$$\mu_\varphi\left(x \in \Omega : \frac{S_n f(x)}{\sqrt{n}} \leq u\right) = \frac{1}{\sigma_f\sqrt{2\pi}} \int_{-\infty}^{u} \mathrm{e}^{-\frac{v^2}{2\sigma_f^2}} \,\mathrm{d}v + O\left(\frac{1}{\sqrt{n}}\right).$$

In particular, the central limit theorem holds for Lipschitz functions:

$$\frac{S_n f(x)}{\sqrt{n}} \xrightarrow{\text{law}} \mathcal{N}(0, \sigma_f^2).$$

The strategy to prove the central limit theorem for Birkhoff sums relies on characteristic functions and the identity

$$\int \mathrm{e}^{\frac{\mathrm{i}\beta S_n f}{\sqrt{n}}} \,\mathrm{d}\mu_\varphi = \int \mathrm{Q}^n_{\varphi + \frac{\mathrm{i}\beta}{\sqrt{n}} f} 1 \,\mathrm{d}\mu_\varphi, \ \beta \in \mathbb{R}$$

where 1 denotes the function which is constantly equal to 1.

LECTURE 2 :
Gaussian concentration bound, and some applications

Recap

(Ω, T) is a subshift of finite type of the full shift $(A^{\mathbb{N}}, T)$ where A is a finite set (alphabet).

Take a Lipschitz potential $\varphi : \Omega \to \mathbb{R}$ (with respect to the distance $\mathrm{d}_\theta(x,y) = \theta^{\inf\{k \geq 0 \,:\, x^k \neq y^k\}}$).

Then there exists a unique Gibbs measure μ_φ which is shift-invariant:
$\exists\ c_\varphi \geq 1$ such that

$$c_\varphi^{-1} \leq \frac{\mu_\varphi\big([x^0 \cdots x^{n-1}]\big)}{\exp\big(-nP(\varphi) + S_n\varphi(x)\big)} \leq c_\varphi$$

for any $x = (x^i)_{i=0}^{\infty} \in \Omega$ and for any $n \geq 1$, where $P(\varphi) = \log \lambda_\varphi$.
(Remember that $\mathrm{R}_\varphi\, h_\varphi = \lambda_\varphi h_\varphi$, etc.)
μ_φ is also the unique equilibrium state for φ and in particular

$$P(\varphi) = s(\mu_\varphi) + \int \varphi \, \mathrm{d}\mu_\varphi.$$

We saw two basic **limit theorems** for Bikhoff sums of Lipschitz functions (observables):

Large deviations:
For f Lipschitz with $\int f \, \mathrm{d}\mu_\varphi = 0$

$$\mu_\varphi\left(x \in \Omega : \frac{S_n f(x)}{n} \geq u\right)$$

decays exponentially fast with n with rate function given by the Legendre transform of

$$\beta \mapsto P(\varphi + \beta f) - P(\phi).$$

Central limit theorem:

$$\frac{S_n f(x)}{\sqrt{n}} \xrightarrow{\text{law}} \mathcal{N}(0, \sigma_f^2)$$

where

$$\sigma_f^2 = \lim_{n \to +\infty} \frac{1}{n} \int \left(S_n f\right)^2 \mathrm{d}\mu_\varphi = \int f^2 \mathrm{d}\mu_\varphi + 2 \sum_{j \geq 1} \int f \cdot f \circ T^j \mathrm{d}\mu_\varphi < \infty$$

Going beyond Birkhoff sums and non-asymptotic results

What can we say for general observables of the form $F(x, Tx, \ldots, T^{n-1}x)$ which are sufficiently regular but otherwise can be non-additive or implicitly defined? (Birkhoff sums are an example of an *additive* F.)

Can be obtain an upper *bound* for

$$\mu_\varphi\left(x \in \Omega : \left|F(x, Tx, \ldots, T^{n-1}x) - \int F(y, Ty, \ldots, T^{n-1}y)\, \mathrm{d}\mu_\varphi(y)\right| \geq u\right)$$

which decays fast in $u > 0$ and in n (after an appropriate rescaling)?

This is the purpose of "concentration inequalities".

A MOTIVATING EXAMPLE

Consider the empirical measure $\mathcal{E}_n(x) := \frac{1}{n}\sum_{j=0}^{n-1}\delta_{T^j x}$.

We saw that there exist $\mathcal{T}_{\mu_\varphi} \subset \Omega$ with $\mu_\varphi(\mathcal{T}_{\mu_\varphi}) = 1$ such that

$$\mathcal{E}_n(x) \xrightarrow[n\to+\infty]{} \mu_\varphi$$

for every $x \in \mathcal{T}_{\mu_\varphi}$ in the weak topology sense.

Consider the Kantorovich distance d_{K} on the space of probability measures.

At which speed $\mathrm{d}_{\mathrm{K}}(\mathcal{E}_n(x), \mu_\varphi)$ goes to 0?

For two probability measures μ_1, μ_2 on Ω

$$\mathsf{d}_{\mathsf{K}}(\mu_1, \mu_2) = \sup \left(\int g \, \mathrm{d}\mu_1 - \int g \, \mathrm{d}\mu_2 : g : \Omega \to \mathbb{R} \text{ is 1-Lipschitz} \right)$$

By Kantorovich-Rubinstein theorem one has the dual representation

$$\mathsf{d}_{\mathsf{K}}(\mu_1, \mu_2) = \inf \left(\int \int \mathsf{d}_\theta(x, y) \, \mathrm{d}\pi(x, y) : \pi \text{ is a coupling of } \mu_1 \text{ and } \mu_2 \right)$$

A class of functions

Let $n \in \mathbb{N}$.

$F : \Omega^n \to \mathbb{R}$ is **separately Lipschitz** if

$$\left|F(x_0, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_{n-1}) - F(x_0, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_{n-1})\right|$$

$$\leq \mathrm{lip}_i(F)\, d_\theta(x_i, x'_i)$$

for all $x_1, \ldots, x_i, \ldots, x_n, x'_i$ in Ω et $\forall i = 1, \ldots, n$.

Basic but important example:

$f : \Omega \to \mathbb{R}$ Lipschitz and $F(x_0, \ldots, x_{n-1}) = f(x_0) + \cdots + f(x_{n-1})$
whence $F(x, Tx, \ldots, T^{n-1}x) = S_n f(x)$.
One has $\mathrm{lip}_i(F) = \mathrm{lip}(f)$, $i = 0, \ldots, n-1$.

Gaussian concentration bound (GCB)

Theorem

Let φ be a Lipschitz potential.
Then there exists $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $F : \Omega^n \to \mathbb{R}$, we have

$$\int \mathrm{d}\mu_\varphi(x)\, \mathrm{e}^{F(x,\dots,T^{n-1}x) - \int F(y,\dots,T^{n-1}y)\mathrm{d}\mu_\varphi(y)} \leq \mathrm{e}^{\frac{C}{2}\sum_{i=0}^{n-1} \mathrm{lip}_i(F)^2}$$

Crucial point : C neither depends on n nor on F.

Remarks:

- One can get a (ugly) explicit expression for C in terms of $|A|$, $\|h_\varphi\|_\infty$, $\|\varphi\|$, m, etc, by using a result by Stoyanov.
- Centering $F(x, \dots, T^{n-1}x)$ in some way of another is necessary because the right-hand side is invariant to constant offsets of the function.

Two corollaries of GCB(C)

First corollary:

$$\mu_\varphi\left(x\in\Omega : F(x,\dots,T^{n-1}x)\geq \int F(y,\dots,T^{n-1}y)\,\mathrm{d}\mu_\varphi(y)+u\right)$$
$$\leq \exp\left(-\frac{u^2}{2C\sum_{i=0}^{n-1}\mathrm{lip}_i(F)^2}\right),\forall n\in\mathbb{N},\forall u>0.$$

Remark:

GCB(C) tells us about $F(x,\dots,T^{n-1}x)-\int F(y,\dots,T^{n-1}y)\,\mathrm{d}\mu_\varphi(y)$ but very often we are interested in $F(x,\dots,T^{n-1}x)$, so we have to find a “good” upper bound for $\int F(y,\dots,T^{n-1}y)\mathrm{d}\mu_\varphi(y)$.

PROOF

To alleviate notation set

$$F = F(x, \ldots, T^{n-1}x), \int F = \int F(y, \ldots, T^{n-1}y)\mathrm{d}\mu_\varphi(y), \textit{etc.}$$

Then, for any $\eta > 0$, one has by Markov's inequality

$$\mu_\varphi\left(F - \int F \geq u\right) = \mu_\varphi\left(\mathrm{e}^{\eta\left(F - \int F\right)} \geq \mathrm{e}^{\eta u}\right) \leq \mathrm{e}^{-\eta u}\int \mathrm{e}^{\eta(F-\int F)}$$

$$\leq \mathrm{e}^{-\eta u}\,\mathrm{e}^{\frac{C\eta^2}{2}\sum_{i=0}^{n-1}\mathrm{lip}_i(F)^2} \quad \text{(GCB(C) applied to } \eta F\text{)}.$$

Then minimize the r.h.s. over $\eta > 0$. □

Applying the previous bound to $-F$ we get by a union bound

$$\mu_\varphi\left(x\in\Omega : \left|F(x,\ldots,T^{n-1}x)-\int F(y,\ldots,T^{n-1}y)\mathrm{d}\mu_\varphi(y)\right|\geq u\right)$$
$$\leq 2\exp\left(-\frac{u^2}{2C\sum_{i=0}^{n-1}\mathrm{lip}_i(F)^2}\right),\forall n\in\mathbb{N},\forall u>0.$$

Second corollary:

$$\int F^2(x,\ldots,T^{n-1}x)\,\mathrm{d}\mu_\varphi(x) - \Big(\int F(y,\ldots,T^{n-1}y)\,\mathrm{d}\mu_\varphi(y)\Big)^2$$
$$\leq C\sum_{i=0}^{n-1} \mathrm{lip}_i(F)^2.$$

Hence $C\sum_{i=0}^{n-1} \mathrm{lip}_i(F)^2$ is a proxy for the variance of the separately Lipschitz function $F:\Omega^n\to\mathbb{R}$.

Proof

For every $\eta > 0$

$$\frac{1}{\eta^2}\Big(\int e^{\eta(F-\int F)}\,d\mu_\varphi - 1\Big) \le \frac{1}{\eta^2}\Big(e^{\frac{C\eta^2}{2}\sum_{i=0}^{n-1}\mathrm{lip}_i(F)^2} - 1\Big)$$

By Taylor expansion

$$\int e^{\eta(F-\int F)}\,d\mu_\varphi - 1 = \eta\Big(\underbrace{\int\big(F-\int F\big)}_{=0}\Big) + \frac{\eta^2}{2}\underbrace{\int\big(F-\int F\big)^2}_{=\mathrm{Var}(F)} + o(\eta^2)$$

and

$$e^{\frac{C\eta^2}{2}\sum_{i=0}^{n-1}\mathrm{lip}_i(F)^2} - 1 = \frac{C\eta^2}{2}\sum_{i=0}^{n-1}\mathrm{lip}_i(F)^2 + o(\eta^2). \qquad \square$$

Comparison with large deviations and central limit asymptotics in the case of Birkhoff sums

$f : \Omega \to \mathbb{R}$ Lipschitz and $F(x_0, \ldots, x_{n-1}) = f(x_0) + \cdots + f(x_{n-1})$
whence $F(x, Tx, \ldots, T^{n-1}x) = S_n f(x)$.
One has $\mathrm{lip}_i(F) = \mathrm{lip}(f)$, hence $\sum_{i=0}^{n-1} \mathrm{lip}_i(F)^2 = n\,\mathrm{lip}(f)^2$.

We get

$$\mu_\varphi\Big(x \in \Omega : \Big|S_n f(x) - n\int f \,\mathrm{d}\mu_\varphi\Big| \geq u\Big) \leq 2\,\mathrm{e}^{-\frac{u^2}{2Cn\,\mathrm{lip}(f)^2}},\ \forall u > 0, n \geq 1.$$

Scale of large deviations: replace u by un to get

$$\mu_\varphi\left(x \in \Omega : \left|\frac{S_n f(x)}{n} - \int f \, d\mu_\varphi\right| \geq u\right) \leq 2 \underbrace{e^{-\frac{nu^2}{2C \operatorname{lip}(f)^2}}}_{\text{exponentially decaying in n}}, \forall u > 0, n \geq 1.$$

Scale of the central limit theorem: replace u by $u\sqrt{n}$ to get

$$\mu_\varphi\left(x \in \Omega : \left|\frac{S_n f(x) - \int f \, d\mu_\varphi}{\sqrt{n}}\right| \geq u\right) \leq 2 \underbrace{e^{-\frac{u^2}{2C \operatorname{lip}(f)^2}}}_{\text{Gaussian tail}}, \forall u > 0, n \geq 1.$$

For Birkhoff sums appropriately normalized we get the right dependences in u and n wrt to large deviations and central limit asymptotics.

Two applications of GCB(C)

(among many others)

Shadowing orbits using orbits started from a subset of Ω

Soit $B \subset \Omega$ tel que $\mu_\varphi(B) > 0$:

$$\mathcal{S}_B(x, n) = \frac{1}{n} \inf_{y \in B} \sum_{i=0}^{n-1} \mathrm{d}_\theta(T^i x, T^i y) \in [0, 1] \, .$$

Theorem

For all $u > \sqrt{2C \ln(\mu_\varphi(B)^{-1})}$ and for all $n \in \mathbb{N}$, we have

$$\mu_\varphi \left\{ x \in \Omega : \mathcal{S}_B(x, n) \geq \frac{u}{\sqrt{n}} \right\} \leq \mathrm{e}^{-\frac{u^2}{8C}} \, .$$

Proof

Let $F(x_0, \ldots, x_{n-1}) = \frac{1}{n} \inf_{y \in B} \sum_{j=0}^{n-1} \mathrm{d}_\theta(x_j, T^j y)$ so that

$$F(x, \ldots, T^{n-1}x) = \mathcal{S}_B(x, n).$$

You can check that $\mathrm{lip}_i(F) = \frac{1}{n}$, $i = 0, \ldots, n-1$.

Using the above corollary we have

$$\mu_\varphi \left\{ x \in \Omega : \mathcal{S}_B(x, n) \geq \int \mathcal{S}_B(y, n) \, \mathrm{d}\mu_\varphi(y) + \frac{u}{\sqrt{n}} \right\} \leq \mathrm{e}^{-\frac{u^2}{2C}}, \forall n \geq 1, u > 0.$$

Now we want to obtain an upper bound for $\int \mathcal{S}_B(y, n) \, \mathrm{d}\mu_\varphi(y)$.

Upper bound for $\int \mathcal{S}_B(y, n)\, \mathrm{d}\mu_\varphi(y)$

We have for every $\eta > 0$

$$\mu_\varphi(B) = \int \mathrm{e}^{-\eta \mathcal{S}_B(x,n)}\, \mathbb{1}_B(x)\, \mathrm{d}\mu_\varphi(x) \leq \int \mathrm{e}^{-\eta \mathcal{S}_B(x,n)}\, \mathrm{d}\mu_\varphi(x)$$

$$\underbrace{\leq}_{\text{by GBC(C)}} \mathrm{e}^{-\eta \int \mathcal{S}_B(y,n) \mathrm{d}\mu_\varphi(y)}\, \mathrm{e}^{\frac{C\eta^2}{2n}}\,.$$

Hence

$$\int \mathcal{S}_B(y, n) \mathrm{d}\mu_\varphi(y) \leq \frac{C\eta}{2n} + \frac{\log\left(\mu_\varphi(B)^{-1}\right)}{\eta} \quad \forall \eta > 0.$$

Optimizing over $\eta > 0$ yields

$$\int \mathcal{S}_B(y, n) \mathrm{d}\mu_\varphi(y) \leq \sqrt{\frac{2C \log(\mu_\varphi(B)^{-1})}{n}}.$$

Remember that $\mathcal{E}_n(x) = \frac{1}{n}\sum_{j=0}^{n-1}\delta_{T^j x}$ and there exist $\mathcal{T}_{\mu_\varphi} \subset \Omega$ with $\mu_\varphi(\mathcal{T}_{\mu_\varphi}) = 1$ such that

$$\mathcal{E}_n(x) \xrightarrow[n\to+\infty]{} \mu_\varphi$$

for every $x \in \mathcal{T}_{\mu_\varphi}$ in the weak topology sense.

At which speed $\mathrm{d}_\mathrm{K}(\mathcal{E}_n(x), \mu_\varphi)$ goes to 0?

Theorem

There exists u_0 and $C' > 0$ such that for any $u > u_0$ and any $n \geq 1$

$$\mu_\varphi\left(x \in \Omega : \mathrm{d}_\mathrm{K}(\mathcal{E}_n(x), \mu_\varphi) \geq \frac{u}{\sqrt{n}}\right) \leq \mathrm{e}^{-C' u^2}.$$

Define the function

$$F(x_0,\dots,x_{n-1}) = \sup\left\{\frac{1}{n}\sum_{j=0}^{n-1} g(x_j) - \int g\,\mathrm{d}\mu_\varphi : g:\Omega\to\mathbb{R} \text{ is 1-Lipschitz}\right\}.$$

It is pretty clear that $\mathrm{lip}_i(F) \leq 1/n$ for all $i = 1,\dots,n-1$.

The hard part is to find a good upper bound for $\int \mathrm{d}_K(\mathcal{E}_n(y),\mu)\,\mathrm{d}\mu_\varphi(y)$.
We omit the proof.

More applications

In the context of shifts of finite type, there are other applications:

- Plug-in estimator for entropy
- Return times (another entropy estimator)
- Speed of Markov approximation in $\overline{d}$-distance.
- Etc.

Beyond shifts of finite type with a Gibbs measure, and beyond Gaussian concentration (very sketchy)

There is a large class of nonuniformly hyperbolic dynamical systems modelled by Young towers with return-time functions with exponential tails for which GCB(C) holds. The proof is almost the same as the one for shifts of finite type with a Gibbs measure.

GCB(C) breaks down for nonuniformly hyperbolic dynamical systems modelled by Young towers with return-time functions with polynomial tails.
The prototype of such systems is the map $T : [0, 1] \to [0, 1]$ given by

$$Tx = \begin{cases} x(1 + 2^\alpha x^\alpha) & 0 \leq x < \frac{1}{2} \\ 2x - 1 & \frac{1}{2} \leq x < 1 \end{cases}$$

where $\alpha \in (0, 1)$ is a parameter. The trouble (only) comes from the indiferrent fixed point at 0.

A very few references

(I should finish to type my lecture notes soon and they of course contain a proper list of references)

- R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Second revised edition. With a preface by David Ruelle. Edited by Jean-René Chazottes. Lecture Notes in Mathematics, 470. Springer-Verlag, Berlin, 2008.
- J.-R. Chazottes, Fluctuations of observables in dynamical systems: from limit theorems to concentration inequalities, 42 pages, chapitre du livre Nonlinear Dynamics: New Directions (H. González-Aguilar & E. Ugalde, eds., Nonlinear Systems and Complexity, vol. 11, Springer, 2015).
- J.-R. Chazottes, S. Gouëzel, Optimal concentration inequalities for dynamical systems, Commun. Math. Phys. **316** (2012).