Concentration Inequalities for Dynamical Systems

Jean-René Chazottes

Centre National de la Recherche Scientifique & Ecole Polytechnique, Paris

DinAmicI V, Rome, June 5–9

Introduction

Warming-up: a random walk

Take independent random variables $X_1, \ldots, X_n$ such that

$$X_i = \begin{cases} +1 & \text{(right) with probability } \frac{1}{2} \\ -1 & \text{(left) with probability } \frac{1}{2}. \end{cases}$$

Position at 0: $S_0 := 0$.

Position at time n:

$$S_n = X_1 + \cdots + X_n.$$

One has

$$\mathbb{E}(S_n) = \sum_{i=1}^{n} \mathbb{E}[X_i] = 0 \quad \text{and} \quad \sqrt{\mathrm{Var}(S_n)} = \sqrt{\sum_{i=1}^{n} \mathrm{Var}(X_i)} = \sqrt{n}\,.$$

Basic idea: use Chebyshev's inequality to get

$$\mathbb{P}\left(|S_n| \geq u\sqrt{n}\right) \leq \frac{\mathrm{Var}(S_n)}{(u\sqrt{n})^2} = \frac{1}{u^2}, \ \forall\, 0 < u < \sqrt{n}.$$

Much better bound (Chernoff, 1952):

$$\mathbb{P}\left(|S_n| \geq u\sqrt{n}\right) \leq 2\exp\left(-\frac{u^2}{2}\right), \ \forall\, 0 < u < \sqrt{n}.$$

(Example: $n = 30$, $u = 5$, $\frac{1}{u^2} = 4.10^{-2}$, $2\,\mathrm{e}^{-u^2/2} \approx 7.5\,10^{-6}$)

$|S_n|$ varies in an interval of size $\mathcal{O}(n)$ but it sharply concentrates in a much narrower interval of size $\mathcal{O}(\sqrt{n})$.

Glimpse into the Gaussian paradise

Take $Z_1, \ldots, Z_n$ i.i.d. with $Z_i \overset{\text{law}}{=} \mathcal{N}(0,1)$

Since $\frac{Z_1+\cdots+Z_n}{\sqrt{n}} \overset{\text{law}}{=} \mathcal{N}(0,1)$, one gets

$$\frac{2}{\sqrt{2\pi}} \frac{u}{u^2+1} \, \mathrm{e}^{-\frac{u^2}{2}} \leq \mathbb{P}\left(|Z_1 + \cdots + Z_n| \geq u\sqrt{n}\right) \leq \frac{2}{\sqrt{2\pi}} \frac{1}{u} \, \mathrm{e}^{-\frac{u^2}{2}}$$

for all $u > 0$.

Comparison with the central limit theorem & the Berry-Esseen estimate

Back to the random walk. Take $u > 0$.

$$\textsc{Central Limit Theorem}: \lim_{n\to\infty} \mathbb{P}(|S_n| \geq u\sqrt{n}) = \frac{2}{\sqrt{2\pi}} \int_u^\infty e^{-\frac{x^2}{2}}\, dx$$

$$\textsc{Berry-Esseen bound}: \mathbb{P}(|S_n| \geq u\sqrt{n}) \leq \frac{2}{\sqrt{2\pi}} \underbrace{\int_u^\infty e^{-\frac{x^2}{2}}\, dx}_{\leq \frac{1}{u} e^{-\frac{u^2}{2}}} + \frac{2C}{\sqrt{n}}$$

where $C =$ absolute constant ≈ 0.5.

Hence, one has to take $n \approx e^{u^2}$ to get back Chernoff's inequality! (Example: $n \approx 7.10^{10}$ for $u = 5$)

Looking at the scale of the law of large numbers

Rescaling Chernoff's bound one gets

$$\mathbb{P}\left(|S_n| \geq nu\right) \leq 2\exp\left(-\frac{nu^2}{2}\right), \quad \forall 0 < u < 1.$$

$$\Downarrow$$

$$\frac{S_n}{n} \xrightarrow[n \to +\infty]{\text{a.s.}} 0\,(= \mathbb{E}(X_1)).$$

Large deviations: asymptotic & non-asymptotic

Take $0 < u < 1$. One has

$$\mathbb{P}(S_n \geq un) \leq \exp(-nI(u)), \quad \forall n \geq 1,$$

where

$$I(u) = \begin{cases} \ln 2 + \frac{1+u}{2} \ln\left(\frac{1+u}{2}\right) + \frac{1-u}{2} \ln\left(\frac{1-u}{2}\right) & \text{if } u \in [-1, 1] \\ +\infty & \text{otherwise.} \end{cases}$$

and

$$I(u) \geq \frac{u^2}{2}.$$

Moreover:

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}(S_n \geq un) = -I(u) \quad \text{(large deviation asymptotics)}$$

A first upgrade of Chernoff's inequality: Hoeffding's inequality

Let X_i, $i = 0, \ldots, n-1$ be independent real-valued random variables and assume that $X_i \in [a_i, b_i]$ (a.s.) for some real numbers a_i, b_i, $i = 0, \ldots, n-1$. Then for all $\lambda \in \mathbb{R}$

$$\log \mathbb{E}\left[e^{\lambda\left(\sum_{i=0}^{n-1} (X_i - \mathbb{E}[X_i]) \right)} \right] \leq \sum_{i=0}^{n-1} \frac{(b_i - a_i)^2}{8} \lambda^2 .$$

In particular, for all $u > 0$,

$$\mathbb{P}\left(\left| \sum_{i=0}^{n-1} (X_i - \mathbb{E}[X_i]) \right| \geq u \right) \leq 2 \exp\left(-\frac{2u^2}{\sum_{i=0}^{n-1} (b_i - a_i)^2} \right) .$$

Gaussian concentration bound (McDiarmid, 1989)

Take **independent** random variables $X_0, X_1, \ldots, X_{n-1}$.

Aim: generalize Hoeffding's inequality in replacing

$$X_0 + \cdots + X_{n-1} \quad (\textit{linear} \text{ function of } X_0, \ldots, X_{n-1})$$

by

$F(X_0, \ldots, X_{n-1}) \quad$ (typically a *nonlinear* function of $X_0, \ldots, X_{n-1}$)

under mild assumptions on F.

Why going beyond partial sums?

Examples of functions of $X_0, \dots, X_{n-1}$ which are not sums:

- $F(X_0, \dots, X_{n-1}) = \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{X_i \leq x\}} - \mathbb{P}(X \leq x) \right|$
- Fattening Hamming balls ;
- Kernel density estimation ;
- Plug-in estimator for Shannon entropy ;
- Etc

The bounded difference property

Let Ω be a set (think of a subset of $\mathbb{R}$).

$F : \Omega^n \to \mathbb{R}$ satisfies the **bounded difference property** if there are some positive constants $\ell_0(F), \ldots, \ell_{n-1}(F)$ such that

$$\left|F(x_0, \ldots, x_i, \ldots, x_{n-1}) - F(x_0, \ldots, x_i', \ldots, x_{n-1})\right| \leq \ell_i(F)$$

for all $x_0, \ldots, x_i, \ldots, x_{n-1}$ and x_i' in Ω.

Gaussian concentration bound (McDiarmid, 1989)

Let $X_0, \ldots, X_{n-1}$ be independent random variables taking values in a set Ω. Then, for all functions with the bounded difference property, for all $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}\Big[\exp\big(\lambda(F(X_0,\ldots,X_{n-1}) - \mathbb{E}[F(X_0,\ldots,X_{n-1})])\big)\Big] \le \frac{\lambda^2}{8}\sum_{i=0}^{n-1} \ell_i(F)^2 .$$

In particular, for all $u \ge 0$,

$$\mathbb{P}\big(F(X_0,\ldots,X_{n-1}) - \mathbb{E}[F(X_0,\ldots,X_n)] \ge u\big) \le \exp\left(\frac{-2u^2}{\sum_{i=0}^{n-1}\ell_i(F)^2}\right).$$

Hence

$$\mathbb{P}\big(|F(X_0,\ldots,X_n) - \mathbb{E}[F(X_0,\ldots,X_n)]| \ge u\big) \le 2\exp\left(\frac{-2u^2}{\sum_{i=1}^{n}\ell_i(F)^2}\right)$$

Dynamical Systems

Set-up

(Ω, d): metric space, $T :\circlearrowleft$, $\mu \circ T^{-1} = \mu$

For $x_0 \in \Omega$, let $X_k(x_0) = T^k x_0$, $k \geq 0$.

If x_0 is distributed according to μ then $\{X_k\}_{k\geq 0}$ is a stationary process on $\Omega^{\mathbb{N}}$ with distribution

$$\mathrm{d}\tilde{\mu}(x_0, x_1, \ldots) := \mathrm{d}\mu(x_0) \otimes \delta_{x_1 = Tx_0} \otimes \delta_{x_2 = T^2 x_0} \otimes \cdots$$

which is not a product measure.

For $F : \Omega^{\mathbb{N}} \to \mathbb{R}$ let

$$\mathbb{E}[F] := \int_\Omega F(x, Tx, \ldots)\, \mathrm{d}\mu(x) = \int_{\Omega^{\mathbb{N}}} F\, \mathrm{d}\tilde{\mu}.$$

Separately Lipschitz functions

$F : \Omega^{\mathbb{N}} \to \mathbb{R}$ is separately Lipschitz if for all $i = 0, 1, \ldots$ one has

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

$$\leq \mathbf{Lip}_i(\mathbf{F})\, d(x_i, x_i')$$

for all $x_0, \ldots, x_i, \ldots$ and x_i' in Ω.

Often $F(x_0, x_1, \ldots) = F(x_0, x_1, \ldots, x_{n-1})$, *i.e.*, $\mathrm{Lip}_i(F) = 0$ for all $i \geq n$.

The simplest example to keep in mind

$f : \Omega \to \mathbb{R}$ Lipschitz, $F(x_0, x_1, \dots, x_{n-1}) = \sum_{i=0}^{n-1} f(x_i)$, hence

$$F(x, Tx, \dots, T^{n-1}x) = \sum_{i=0}^{n-1} f(T^i x).$$

So

$$\mathrm{Lip}_i(F) = \mathrm{Lip}(f) \quad \text{for all } i = 0, 1, \dots, n-1.$$

Gaussian concentration bound for uniform Young towers with exponential tails

Theorem (JRC-Gouëzel)

Let $(\Delta, \widehat{T}, \hat{\mu})$ be a uniform Young tower with exponential tails. Then there exists $C > 0$ such that for any separately Lipschitz function $F : \Delta^{\mathbb{N}} \to \mathbb{R}$,

$$\log \int \mathrm{e}^{F(x, \widehat{T}x, \dots) - \widehat{\mathbb{E}}[F]} \, \mathrm{d}\hat{\mu}(x) \leq C \sum_{i=0}^{\infty} \mathrm{Lip}_i(F)^2.$$

If $F(x_0, x_1, \dots) = F(x_0, x_1, \dots, x_{n-1})$ for some n, $\sum_{i=0}^{\infty} \mathrm{Lip}_i(F)^2$ is a finite sum.

Corollary

For any separately Lipschitz function $F : \Delta^{\mathbb{N}} \to \mathbb{R}$ and for any $u > 0$,

$$\hat{\mu}\left\{x \in \Delta : \left|F(x, \widehat{T}x, \dots) - \widehat{\mathbb{E}}[F]\right| \geq u\right\} \leq 2\,\exp\left(-\frac{u^2}{4C\sum_{i=0}^{\infty}\mathrm{Lip}_i(F)^2}\right).$$

For the simplest F, *i.e.*, Birkhoff sum of a Lipschitz function $f : \Omega \to \mathbb{R}$, with $\int f \mathrm{d}\hat{\mu} = 0$: for any $u > 0$ and for any n

$$\hat{\mu}\left\{x \in \Delta : \left|S_n f(x)\right| \geq u\sqrt{n}\right\} \leq 2\,\exp\left(-\frac{u^2}{4C\mathrm{Lip}(f)^2}\right).$$

Examples

Dynamical sytems (Ω, T, μ) "modeled" by a uniform Young tower $(\Delta, \widehat{T}, \hat{\mu})$ with exponential tails, μ being the SRB measure:

1. Subshifts of finite type with a Gibbs measure for a Hölder potential ;
2. Uniformly piecewise expanding map of the interval ;
3. Axiom A attractors ;
4. The quadratic family for Benedicks-Carleson parameters ;
5. The Hénon attractor for Benedicks-Carleson parameters ;
6. Piecewise hyperbolic maps like the Lozi attractor ;
7. Sinai billiard maps with finite horizon ;
8. Etc

Exponential decay of correlations for Lipschitz functions.

Application to the tracing of orbits

Consider (Ω, T, μ) modeled by a Young tower with exponential tails. Suppose that $\mathrm{diam}(\Omega) < \infty$.

Let $A \subset \Omega$ such that $\mu(A) > 0$ and

$$\mathcal{S}_A(x,n) = \inf_{y \in A} \sum_{j=0}^{n-1} d(T^j x, T^j y), \ \ n \in \mathbb{N}.$$

$\mathcal{S}_A(x,n)/n$ measures how well we can trace orbits starting off A by an orbit starting in A.
A priori $\mathcal{S}_A(x,n)/n$ varies in an interval of size $\mathcal{O}(1)$.

Theorem

For any $n \in \mathbb{N}$ and for any $u > 0$

$$\mu\left\{x \in \Omega \,:\, \frac{\mathcal{S}_A(x,n)}{n} \geq \left(2\sqrt{C \log \frac{1}{\mu(A)}} + u\right)\frac{1}{\sqrt{n}}\right\} \leq 2\,\mathrm{e}^{-\frac{u^2}{4C}}$$

Numerical illustration

Take $C = 1/8$, $\mu(A) = 10^{-6}$, and $u > 0$ such that $2\sqrt{C \ln \frac{1}{\mu(A)}} + u = 5$.

Then

$$\mu\left\{x \in \Omega \,:\, \frac{\mathcal{S}_A(x,n)}{n} \geq \frac{5}{\sqrt{n}}\right\} \leq 3.10^{-17}.$$

Take, *e.g.*, $n = 10^4$:

$$\mu\left\{x \in \Omega \,:\, \frac{\mathcal{S}_A(x,10^4)}{10^4} \geq 0.05\right\} \leq 3.10^{-17}.$$

Think of $T(x) = 2x \bmod 1$: $\mu(A) = |A| = 10^{-6}$.

Proof

Let $F(x_0, \ldots, x_{n-1}) = \inf_{y \in A} \sum_{j=0}^{n-1} d(x_j, T^j y)$.

Check that $\mathrm{Lip}_i(F) \leq 1$, $i = 0, \ldots, n-1$.

One has $F(x, Tx, \ldots, T^{n-1}x) = \mathcal{S}_A(x, n)$.

By the (corollary of the) Gaussian concentration bound one has for all $n \geq 1$ and $u > 0$

$$\mu \left\{ x \in \Omega \, : \, \frac{\mathcal{S}_A(x, n)}{n} \geq \mathbb{E}\left[\mathcal{S}_A(\cdot, n)\right] + u\sqrt{n} \right\} \leq 2\, \mathrm{e}^{-\frac{u^2}{4C}} \; .$$

Upper bound for $\mathbb{E}\left[\mathcal{S}_A(\cdot, n)\right]$?

Upper bound for $\mathbb{E}\left[\mathcal{S}_A(\cdot, n)\right]$:

Write $\mathcal{S}_A$ for $\mathcal{S}_A(\cdot, n)$.

Using again the Gaussian concentration bound ($\mathcal{S}_A \equiv 0$ on A):

$$\mu(A) = \int \mathrm{e}^{-\lambda \mathcal{S}_A} \mathbb{1}_A \mathrm{d}\mu \leq \int \mathrm{e}^{-\lambda \mathcal{S}_A} \mathrm{d}\mu \leq \mathrm{e}^{-\lambda \mathbb{E}[\mathcal{S}_A]} \mathrm{e}^{Cn\lambda^2} .$$

Hence

$$\mathbb{E}[\mathcal{S}_A] \leq Cn\lambda + \frac{1}{\lambda} \log \frac{1}{\mu(A)} .$$

True for any $\lambda > 0$, so one can optimize:

$$\mathbb{E}[\mathcal{S}_A] \leq 2\sqrt{Cn \log \frac{1}{\mu(A)}} \qquad \blacksquare$$

Application to the empirical measure

Let

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

For μ ergodic, it converges weakly for μ-a.e. x to μ.

We are interested in

$$d_{\mathrm{K}}\big(\mathcal{E}_n(x), \mu\big) = \sup\left\{\frac{1}{n}\sum_{j=0}^{n-1} g(T^j x) - \int g \mathrm{d}\mu : g : \Omega \to \mathbb{R} \text{ is } 1 - \text{Lipschitz}\right\},$$

the Kantorovich distance between the empirical measure and μ.

Theorem

Let (Ω, T, μ) be modeled by a Young tower with exponential tails. Then for all $n \geq 1$ and for all $u > 0$ one has

$$\mu\left\{x \in \Omega : \left|d_{\mathrm{K}}\big(\mathcal{E}_n(x), \mu\big) - \mathbb{E}\big[d_{\mathrm{K}}\big(\mathcal{E}_n(\cdot), \mu\big)\big]\right| \geq \frac{u}{\sqrt{n}}\right\} \leq 2\,\mathrm{e}^{-\frac{u^2}{4C}}$$

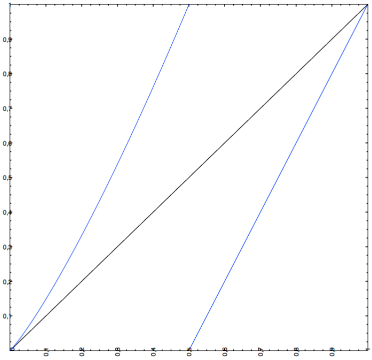
Upper bound for $\mathbb{E}\big[d_{\mathrm{K}}\big(\mathcal{E}_n(\cdot), \mu\big)\big]$? Difficult to get a good one in general.

Special case: one-dimensional map with summable decay of correlations for functions with bounded variation. Then

$$\mathbb{E}\big[d_{\mathrm{K}}\big(\mathcal{E}_n(\cdot), \mu\big)\big] \leq \frac{\text{cte}}{\sqrt{n}}\,,\ n \geq 1.$$

When Gaussian concentration fails: The basic example

Map of the interval with an indifferent fixed point:



$x + 2^{\alpha} x^{1+\alpha}$ $\quad$ $2x - 1$

For $\alpha \in (0,1)$, there is a unique absolutely continuous invariant probability:

$$d\mu(x) = h(x)dx, \; h(x) \sim x^{-\alpha} \text{ as } x \downarrow 0.$$

Decay of correlations for Lipschitz functions: $\mathcal{O}(1/n^{\frac{1}{\alpha}-1})$.

Theorem (JRC-Gouëzel)

Let $\alpha \in (0, 1/2)$. Then there exists $C > 0$ such that, for any $n \in \mathbb{N}$, for any separately Lipschitz function $F(x_0, \ldots, x_{n-1})$

$$\mu\left\{\left|F(x, Tx, \ldots, T^{n-1}x) - \mathbb{E}(F)\right| \geq u\right\}$$

$$\leq C\frac{\left(\sum_{i=0}^{n-1} \mathrm{Lip}_i(K)^2\right)^{\frac{1}{\alpha}-1}}{u^{\frac{2}{\alpha}-2}}, \ \forall u > 0.$$

Remarks:

- The previous bound is almost optimal.
- There is a general theorem for nonuniform Young towers with polynomial tails.
- It applies *e.g.* to billiard maps studied by Chernov and Zhang for which the decay of correlations behaves like $O((\log n)^c/n^{1/\alpha-1})$, for some parameter α that can be chosen freely in $(0, 1/2]$ and some $c > 0$.

Supplementary Material

Uniform Young tower

It is a space Δ such that

1. This space is partitioned into subsets $\Delta_{\alpha,\ell}$ (for $\alpha \in \mathbb{N}$ and $\ell \in [0, \phi(\alpha) - 1]$, where ϕ is an integer-valued return time function). The dynamics sends bijectively $\Delta_{\alpha,\ell}$ on $\Delta_{\alpha,\ell+1}$ if $\ell < \phi(\alpha) - 1$, and $\Delta_{\alpha,\phi(\alpha)-1}$ on $\Delta_0 := \bigcup_\alpha \Delta_{\alpha,0}$.
2. The distance is given by $d_\beta(x, y) := \beta^{s(x,y)}$ where $\beta < 1$ and $s(x, y)$ is the separation time for the whole dynamics, i.e., the first n such that $\widehat{T}^n x$ and $\widehat{T}^n y$ are not in the same element of the partition.
3. There is an invariant probability measure $\hat{\mu}$ such that the inverse g of its Jacobian satisfies $|g(x)/g(y) - 1| \leq C\, d_\beta(\widehat{T}x, \widehat{T}y)$ for any x and y in the same element of the partition.
4. We have $\gcd(\phi(\alpha) \,:\, \alpha \in \mathbb{N}) = 1$ (i.e., the tower is aperiodic).

◂ Gaussian concentration bound for uniform Young towers

Uniform Young tower with expoential tails: there exists $c_0 > 0$ such that

$$\int_{\Delta_0} \mathrm{e}^{c_0 \phi} \, \mathrm{d}\hat{\mu} < \infty.$$

Dynamical systems modeled by a uniform Young tower

They admit an SRB measure μ and there is a uniform Young tower $(\Delta, \widehat{T}, \hat{\mu})$ and a projection map $\pi : \Delta \to \Omega$ such that $T \circ \pi = \pi \circ \widehat{T}$ and $\mu = \hat{\mu} \circ \pi^{-1}$.
It can also be ensured that the projection map is contracting, i.e., $d(\pi x, \pi y) \leq d_\beta(x, y)$ for every x, y in the same partition element.
If f is a bounded Lipschitz function on X, it lifts to a function $f \circ \pi$ which is Lipschitz in the tower.

◂ Examples of dynamical systems modeled by uniform Young towers

Some references

- J.-R. Chazottes, S. Gouëzel. Optimal concentration inequalities for dynamical systems with , Commun. Math. Phys. 316 (2012), 47 pp.
- J.-R. Chazottes, P. Collet, B. Schmitt. Statistical consequences of the Devroye inequality for processes. Applications to a class of non-uniformly hyperbolic dynamical systems. Nonlinearity 18 (2005), no. 5, 2341–2364.
- P. Collet, S. Martínez, B. Schmitt. Exponential inequalities for dynamical measures of expanding maps of the interval. Probab. Theory Related Fields 123 (2002), no. 3, 301–322.
- S. Gouëzel, I. Melbourne. Moment bounds and concentration inequalities for slowly mixing dynamical systems. Electron. J. Probab. 19 (2014), no. 93, 30 pp.
- M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89. American Mathematical Society, 2001.