Concentration Inequalities: An introduction & some recent results (five lectures)

Jean-René Chazottes

Centre National de la Recherche Scientifique & Ecole Polytechnique, Paris

May 22-24

Slogan

A random variable that *"smoothly"* depends on the influence of many *weakly dependent* random variables is, on an *appropriate scale*, *essentially* constant (= to its expected value).

Goal of these lectures:
Converting this slogan into maths for independent random variables, Markov chains, and Gibbs measures.

Plan of the lectures

1. Independent Random Variables
   - A random walk
   - Hoeffding's inequality
   - Three applications of Hoeffding's inequality
   - More on tail probabilities for sums
   - Azuma-Hoeffding inequality & Gaussian concentration bound
   - Three applications of the Gaussian concentration bound
2. Markov chains
3. Gibbs measures

Lecture I:
Independent random variables

Take independent random variables $X_1, \dots, X_n$ 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: $n = 30$, $u = 5$, $1/u^2 = 4.10^{-2}$, $2\,\mathrm{e}^{-u^2/2} \approx 7.5\,10^{-6}$)

Conclusion: though $|S_n| = \mathcal{O}(n)$, it is sharply concentrated in a much narrower interval of size $\mathcal{O}(\sqrt{n})$.

Glimpse into the Gaussian paradise

$Z_1, \ldots, Z_n$ 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

$$\begin{aligned}\mathbb{P}\left(|Z_1 + \cdots + Z_n| \geq u\sqrt{n}\right) &\leq \frac{2}{u\sqrt{2\pi}}\, \mathrm{e}^{-\frac{u^2}{2}} \\ &\geq \max\left(0, 1 - \frac{1}{u^2}\right) \frac{2}{u\sqrt{2\pi}}\, \mathrm{e}^{-\frac{u^2}{2}}\end{aligned}$$

for all $u > 0$ (the lower bound being non-trivial for $u > 1$).

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

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

$$\lim_{n\to\infty} \mathbb{P}(|S_n| \geq u\sqrt{n}) = \frac{2}{\sqrt{2\pi}} \int_u^\infty e^{-\frac{x^2}{2}}\, dx \quad \text{(CLT)}$$

and

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

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

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

Hence

$$\sum_{n\geq 1}\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq 2\sum_{n\geq 1}\exp\left(-\frac{n\,\varepsilon^2}{2}\right) < +\infty,\ \forall \varepsilon > 0.$$

$$\Downarrow$$

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon \text{ infinitely often}\right) = 0$$

$$\Downarrow$$

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

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

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

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

We start with a lemma:

Hoeffding (1963)

Let Z be a real-valued random variable with $\mathbb{E}[Z] = 0$ and $\alpha \leq Z \leq \beta$ for some reals $\alpha < \beta$. Then

$$\log \mathbb{E}\left[\, \mathrm{e}^{\lambda Z} \,\right] \leq \frac{\lambda^2 (\beta - \alpha)^2}{8}, \ \forall \lambda \in \mathbb{R}.$$

Proof of Hoeffding's lemma

Set

$$\psi(\lambda) = \log \mathbb{E}\left[\, e^{\lambda Z} \,\right],\ \lambda \in \mathbb{R}.$$

By Taylor's expansion

$$\psi(\lambda) = \underbrace{\psi(0)}_{=0} + \lambda \underbrace{\psi'(0)}_{=\mathbb{E}[Z]=0} + \frac{\lambda^2}{2}\psi''(\theta)$$

for some $\theta \in (0, \lambda)$.
By elementary computation

$$\psi''(\lambda) = \mathrm{Var}(Z_\lambda)$$

where $Z_\lambda \in [\alpha, \beta]$ is a r.v. with density $f(x) = e^{-\psi(\lambda)}\, e^{\lambda x}$ wrt $\mathbb{P}$.

Since $\left| Z_\lambda - \frac{\alpha+\beta}{2} \right| \leq \frac{\beta-\alpha}{2}$ then

$$\mathrm{Var}(Z_\lambda) = \mathrm{Var}\left(Z_\lambda - \frac{\alpha+\beta}{2}\right) \leq \frac{(\beta-\alpha)^2}{4}. \qquad \blacksquare$$

Hoeffding's inequality

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

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

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

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

Caveat:

Hoeffding's inequality is insensitive to the variance of the X_i's.

For **all** r.v. with a distribution concentrated on, say, $[-1, 1]$ (*i.e.*, $X_i \in [-1, 1]$ a.s.), we get **the same bound** as for $X_i = \pm 1$ with $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$, which are the most spread-out random variables in this class.

Proof of Hoeffding's inequality

Set $Y_i = X_i - \mathbb{E}[X_i]$. By independence of the X_i's we get for every $u \geq 0$ and every $\lambda > 0$

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq u\right) &= \mathbb{P}\left(\exp\left(\lambda \sum_{i=1}^{n} Y_i\right) \geq \exp(\lambda u)\right) \\
&\leq \exp(-\lambda u)\,\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} Y_i\right)\right] \\
&\leq \exp(-\lambda u) \prod_{i=1}^{n} \mathbb{E}\left[\exp(\lambda Y_i)\right].
\end{aligned}$$

For each $i \leq n$, Y_i is centered ($\mathbb{E}[Y_i] = 0$) and belongs to the interval

$$[a_i - \mathbb{E}[X_i], b_i - \mathbb{E}[X_i]]\,.$$

We apply Hoeffding's lemma to each Y_i with $\alpha = a_i - \mathbb{E}[X_i]$, $\beta = b_i - \mathbb{E}[X_i]$, which gives at once the 1st inequality and also

$$\mathbb{P}\left(\sum_{i=1}^{n} Y_i \geq u\right) \leq \exp\left(-\lambda u + \frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2\right).$$

The minimum value of the right hand side is attained for

$$\lambda = \frac{4u}{\sum_{i=1}^{n}(b_i - a_i)^2}.$$

■

Another look at Hoeffding's inequality

For concreteness, assume that $X_1, X_2, \ldots$ are i.i.d. with $0 \leq X_i \leq 1$. Then

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| \geq \varepsilon\right) \leq 2\,\mathrm{e}^{-2n\,\varepsilon^2}, \ \varepsilon > 0, \ n \geq 1.$$

Let

$$\delta := 2\,\mathrm{e}^{-2n\,\varepsilon^2} \qquad \text{(“confidence”)}$$

If

$$n \geq \frac{1}{2\varepsilon^2}\log\left(\frac{2}{\delta}\right),$$

then, with probability at least $1-\delta$, the difference between the empirical mean and the true mean is at most ε.

Now, imagine that we want 10 times more accuracy: one can check that for $\varepsilon' = \varepsilon/10$ we need $100n$ samples.

I. Binomial confidence intervals.

Goal: estimate the unknown parameter $p \in]0,1[$ of a Bernoulli r.v. X by observing the realizations of independent copies $X_1, X_2, \ldots$ of X.
Empirical estimator:

$$\overline{X}_n := \frac{X_1 + \cdots + X_n}{n}$$

which converges a.s., as $n \to +\infty$, to p, by the strong law of large numbers.
Confidence intervals for $\overline{X}_n$ at a confidence level $1-\alpha$, such as 95%: the resulting intervals bracket the parameter p with probability at least 0.95.

Usually: one applies the CLT... which is an asymptotic result! For every $\alpha \in \,]0,1[\,$,

$$\lim_{n \to +\infty} \mathbb{P}\left(p \in \left[\overline{X}_n - \frac{a}{2\sqrt{n}}, \overline{X}_n + \frac{a}{2\sqrt{n}}\right]\right) \geq 1 - \alpha$$

where a solves $1 - \alpha = \int_{-a}^{a} (2\pi)^{-\frac{1}{2}} \, \mathrm{e}^{-\frac{u^2}{2}} \, \mathrm{d}u$.

Let's apply Hoeffding's inequality: for all $a > 0$

$$\mathbb{P}\left(\left|\overline{X}_n - p\right| \geq a\right) \leq 2\,\mathrm{e}^{-2na^2}\ .$$

This ensures that, for all $n > (2a^2)^{-1}\ln 2$,

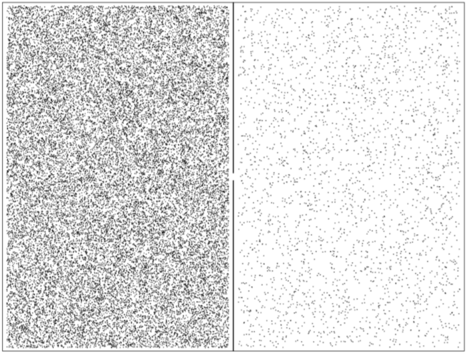
$$\mathbb{P}\left(\left|\overline{X}_n - p\right| \leq a\right) \geq 1 - 2\,\mathrm{e}^{-2na^2} = 1 - \alpha$$

where $\alpha = 2\,\mathrm{e}^{-2na^2}$ lies in $]0,1[$.

Therefore we obtain an exact (*i.e.*, non-asymptotic):

$$\mathbb{P}\left(p \in \left[\overline{X}_n - \sqrt{\frac{\ln\left(\frac{2}{\alpha}\right)}{2n}}, \overline{X}_n + \sqrt{\frac{\ln\left(\frac{2}{\alpha}\right)}{2n}}\right]\right) \geq 1 - \alpha\,.$$

II. The Ehrenfest model.



container A container B

Total of N molecules of gas

$Z_t^{(N)}$: number of molecules in container A at time $t \in \mathbb{N}$

Evolution rule: for $i \in \{1, \ldots, N-1\}$,

$$\mathbb{P}\big(Z_{t+1}^{(N)} = i+1 \big| Z_t^{(N)} = i\big) = \frac{N-i}{N}$$

and

$$\mathbb{P}\big(Z_{t+1}^{(N)} = i-1 \big| Z_t^{(N)} = i\big) = \frac{i}{N}\,.$$

Guess: after a long time, there will be, on average, $N/2$ molecules in each container (thermal equilibrium).

No time to wait? start with the probability distribution ensuring that: $\mathrm{B}(N, 1/2)$, so the number of molecules in container A will be i with probability

$$\pi(i) = \binom{N}{i} 2^{-N},\ 0 \le i \le N.$$

Look at the “temperature” in container A:

$$\frac{Z^{(N)}}{N}$$

where $Z^{(N)} \stackrel{\text{law}}{=} \mathrm{B}(N, 1/2) \stackrel{\text{law}}{=} \sum_{i=1}^{N} X_i$ with $X_i \stackrel{\text{law}}{=} \text{Bernoulli}(1/2)$

Now apply Hoeffding’s inequality:

$$\mathbb{P}\left(\left|\frac{Z^{(N)}}{N} - \frac{1}{2}\right| \geq u\right) \leq 2\,\mathrm{e}^{-2Nu^2}, \;\; u > 0.$$

Typically $N = 6.10^{23}$.
Take, *e.g.*, $u = 10^{-9}$: the probabilty of observing a fluctuation of more than one billionth is less than $2\,\mathrm{e}^{-1.2\times 10^6}$!!!

III. Empirical cumulative distribution function.

Suppose we observe the realizations of i.i.d. r.v. $X_1, \ldots, X_n$ with a common cumulative distribution function (CDF, for short) $\mathcal{F}(x) := \mathbb{P}(X \leq x)$, $x \in \mathbb{R}$.
Empirical CDF:

$$\mathcal{F}_n(x; X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.$$

The $\mathbb{1}_{\{X_i \leq x\}}$'s are independent Bernoulli r.v. with parameter $\mathbb{P}(X \leq x) = \mathcal{F}(x)$, so $\mathbb{E}[\mathbb{1}_{\{X_i \leq x\}}] = \mathcal{F}(x)$.

By the strong law of large numbers, for each $x \in \mathbb{R}$,

$$\lim_{n \to +\infty} \mathcal{F}_n(x; X_1, \ldots, X_n) = \mathcal{F}(x), \text{ a.s..}$$

By Hoeffding's inequality, for all $u \geq 0$ and for all $n \in \mathbb{N}$

$$(\clubsuit) \qquad \sup_{x \in \mathbb{R}} \mathbb{P}\left(|\mathcal{F}_n(x; X_1, \ldots, X_n) - \mathcal{F}(x)| \geq u\right) \leq 2\,\mathrm{e}^{-2nu^2}\,.$$

This is not really satisfactory in view of Glivenko-Cantelli theorem:

$$\lim_{n \to +\infty} \sup_{x \in \mathbb{R}} |\mathcal{F}_n(x; X_1, \ldots, X_n) - \mathcal{F}(x)| = 0, \text{ a. s..}$$

QUESTION: Can we put the supremum over x inside the probability in ($\clubsuit$)? This means that we want to consider the random variable

$$F(X_1, \ldots, X_n) := \sup_{x \in \mathbb{R}} |\mathcal{F}_n(x; X_1, \ldots, X_n) - \mathcal{F}(x)|$$

which is not a sum of random variables.

More on tail probabilities for sums: link with large deviation theory & non-Gaussian tails

Let $X_1, X_2, \ldots$ be independent random variables.
Using Cramér-Chernoff method: $\forall u > 0, \forall n \geq 1$,

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}[X_1] \geq u\right)$$
$$\leq \exp\left(-n \sup_{\lambda \geq 0} \left\{\lambda(u + \mathbb{E}[X_1]) - \psi_{X_1}(\lambda)\right\}\right)$$

where

$$\psi_{X_1}(\lambda) := \log \mathbb{E}\left[\, e^{\lambda X_1} \,\right].$$

The bound is non-trivial if there exists $b > 0$ such that $\psi_{X_1}(b) < +\infty$.

If X_1 is a *bounded* random variable, then by Hoeffding's lemma

$$\log \mathbb{E}\left[\, \mathrm{e}^{\lambda X_1} \,\right] \le \lambda \mathbb{E}[X_1] + \frac{\lambda^2}{8} \operatorname{osc}(X_1)^2$$

which gives Hoeffding's inequality.

Now take for instance $X_i \overset{\text{law}}{=} \text{Poisson}(\theta)$ ($\theta > 0$). One gets

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} - \theta \ge u \right) \le \exp\left[-n\theta h\left(\frac{u}{\theta}\right)\right] \,, u > 0,$$

where

$$h(u) := (1+u)\log(1+u) - u, \;\; u \ge -1$$

To compare more easily with Hoeffding's inequality, observe that for all $u \geq 0$

$$h(u) \geq \frac{u^2}{2 + 2u/3}\,.$$

Hence

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} - \theta \geq u\right) \leq \exp\left(-\frac{n}{2\theta}\frac{u^2}{1 + \frac{u}{3\theta}}\right), u > 0.$$

There are two regimes: if $u \ll 3\theta$ we recover a Gaussian bound of the form $\exp(-nu^2)$, whereas if $u \gg 3\theta$ we get a bound of the form $\exp(-c'nu)$.

Lecture II: Independent random variables

(continuation & ending)

FROM $X_1 + \cdots + X_n$ TO NONLINEAR FUNCTIONS OF $X_1, \ldots, X_n$

Take **independent** random variables $X_1, X_2, \ldots, X_n$.

AIM: generalize Hoeffding's inequality in replacing

$$X_1 + \cdots + X_n \quad (\textit{linear} \text{ function of } X_1, \ldots, X_n)$$

by

$$F(X_1, \ldots, X_n) \quad (\text{typically a } \textit{nonlinear} \text{ function of } X_1, \ldots, X_n)$$

under mild assumptions on F.

Let

- $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be an integrable random variable;
- $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}$ a filtration;
- $Y - \mathbb{E}[Y] = \sum_{k=1}^n \left(\mathbb{E}[Y|\mathcal{F}_k] - \mathbb{E}[Y|\mathcal{F}_{k-1}]\right) = \sum_{k=1}^n \Delta_k$.

Azuma-Hoeffding inequality

$$\forall \lambda \in \mathbb{R},\ \log \mathbb{E}\left[\exp(\lambda(Y - \mathbb{E}[Y])\right] \leq \frac{\lambda^2}{8} \sum_{i=1}^n \operatorname{osc}(\Delta_i)^2$$

whence

$$\forall u \geq 0,\ \mathbb{P}\left(|Y - \mathbb{E}[Y]| \geq u\right) \leq 2 \exp\left(-\frac{2u^2}{\sum_{i=1}^n \operatorname{osc}(\Delta_i)^2}\right).$$

Note: $\operatorname{osc}(\Delta_i) = \sup \Delta_i - \inf \Delta_i \leq 2\|\Delta_i\|_\infty$.

Proof of Azuma-Hoeffding inequality

Take $\lambda > 0$ (to be chosen later on):

$$\begin{aligned}\mathbb{P}\left(Y - \mathbb{E}[Y] \geq u\right) &= \mathbb{P}\Big(\exp\left(\lambda(Y - \mathbb{E}[Y])\right) \geq \exp(\lambda u)\Big)\\ &\leq \exp(-\lambda u)\, \underbrace{\mathbb{E}\left[\exp\left(\lambda(Y - \mathbb{E}[Y])\right)\right]}_{=\mathbb{E}[\exp(\lambda(\Delta_n + \cdots + \Delta_1))]}.\end{aligned}$$

Now

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\lambda(\Delta_1 + \cdots + \Delta_n)\right)\right]\\ &= \mathbb{E}\left[\mathbb{E}\Big[\exp\left(\lambda(\Delta_1 + \cdots + \Delta_{n-1})\right)\Big|\mathcal{F}_{n-1}\Big]\right]\\ &= \mathbb{E}\Big[\exp\left(\lambda(\Delta_1 + \cdots + \Delta_{n-1})\right)\, \mathbb{E}\big[\exp(\lambda\Delta_n)\big|\mathcal{F}_{n-1})\big]\Big].\end{aligned}$$

Proof of Azuma-Hoeffding inequality (continued)

By (a slightly modified version of) Hoeffding's Lemma

$$\mathbb{E}\left[\exp(\lambda\Delta_k)\middle|\mathcal{F}_{k-1})\right] \leq \exp\left(\frac{\lambda^2}{8}\operatorname{osc}(\Delta_k)^2\right)$$

Take $Z = \Delta_k$ *(so* $\beta - \alpha = \operatorname{osc}(\Delta_k)$*),* $\mathbb{E}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{k-1}]$ *and observe that* $\mathbb{E}[\Delta_k|\mathcal{F}_{k-1}] = 0$.
By induction one gets

$$\mathbb{E}\left[\exp\left(\lambda(\Delta_1 + \cdots + \Delta_n)\right)\right] \leq \exp\left(\frac{\lambda^2}{8}\sum_{i=1}^{n}\operatorname{osc}(\Delta_i)^2\right).$$

Hence, setting $c = \sum_{i=1}^{n}\operatorname{osc}(\Delta_i)^2$,

$$\begin{aligned}\mathbb{P}\left(Y - \mathbb{E}(Y) \geq u\right) &\leq \exp\left(-\lambda u + \frac{c\lambda^2}{8}\right)\\ &\leq \exp\left(\inf_{\lambda>0}\left(-\lambda u + \frac{c\lambda^2}{8}\right)\right) = \exp\left(-\frac{2u^2}{c}\right).\end{aligned}$$

The same holds for $-Y$. ■

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

$F : \mathscr{S}^n \to \mathbb{R}$ satisfies the **bounded differences property** if there are some positive constants $\ell_1(F), \dots, \ell_n(F)$ such that

$$\left|F(x_1, \dots, x_i, \dots, x_n) - F(x_1, \dots, x_i', \dots, x_n)\right| \leq \ell_i(F)$$

for all $x_1, \dots, x_i, x_i', \dots, x_n$ in $\mathscr{S}$.

Gaussian concentration bound (McDiarmid, 1989)

Let $X_1, \ldots, X_n$ be independent random variables taking values in a set $\mathscr{S}$. Then, for all functions with the bounded differences property, for all $\lambda \in \mathbb{R}$,

$$\log \mathbb{E}\left[\exp\left(\lambda(F(X_1, \ldots, X_n) - \mathbb{E}[F(X_1, \ldots, X_n)])\right)\right] \leq \frac{\lambda^2}{8} \sum_{i=1}^{n} \ell_i(F)^2 .$$

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

$$\mathbb{P}\left(F(X_1, \ldots, X_n) - \mathbb{E}[F(X_1, \ldots, X_n)] \geq u\right) \leq \exp\left(\frac{-2u^2}{\sum_{i=1}^{n} \ell_i(F)^2}\right) .$$

Hence

$$\mathbb{P}\left(|F(X_1, \ldots, X_n) - \mathbb{E}[F(X_1, \ldots, X_n)]| \geq u\right) \leq 2\exp\left(\frac{-2u^2}{\sum_{i=1}^{n} \ell_i(F)^2}\right)$$

Corollary

We have

$$\mathrm{Var}\big(F(X_1,\dots,X_n)\big) \le \frac{1}{4}\sum_{i=1}^{n} \ell_i(F)^2 \,.$$

Proof. Let Z a r.v. such that $\mathbb{E}[Z]=0$ and such that there exists $v>0$ such that

$$\mathbb{E}\left[\,\mathrm{e}^{\lambda Z}\,\right] \le \mathrm{e}^{v\lambda^2}, \forall \lambda \in \mathbb{R}.$$

Then

$$\frac{\mathbb{E}\left[\,\mathrm{e}^{\lambda Z}\,\right]-1}{\lambda^2} \le \frac{\mathbb{E}\left[\,\mathrm{e}^{v\lambda^2}\,\right]-1}{\lambda^2}.$$

Then write Taylor's expansion and $\lambda \downarrow 0$ to get
$\mathrm{Var}(Z)=\mathbb{E}[Z^2] \le 2v.$ ∎

Little checking with our random walk

Back to our random walk:

$$\begin{aligned}
&\mathscr{S} = \{-1, +1\} \\
&F(X_1, \ldots, X_n) = X_1 + \cdots + X_n = S_n \\
&\mathbb{E}[F(X_1, \ldots, X_n)] = \sum_{i=1}^{n} \mathbb{E}(X_i) = 0 \\
&\ell_i(F) = 2, \text{ hence} \sum_{i=1}^{n} \ell_i(F)^2 = 4n.
\end{aligned}$$

Hence we get back Chernoff's inequality: $\forall u \geq 0$

$$\mathbb{P}\big(|S_n| \geq nu\big) \leq 2 \exp\left(-\frac{u^2}{2n}\right) .$$

Two remarks

• The above concentration bound, and more generally, any concentration bound, is concerned with the **fluctuations** of $F(X_1, \ldots, X_n)$ around its expected value $\mathbb{E}[F(X_1, \ldots, X_n)]$.

In general, those bounds don't provide any information on the magnitude of $\mathbb{E}[F(X_1, \ldots, X_n)]$.

• We *do have* to normalize in some sense $F(X_1, \ldots, X_n)$ to control

$$\mathbb{E}\left[\mathrm{e}^{\lambda F(X_1,\ldots,X_n)}\right]$$

since otherwise one can make this quantity arbitrarily large by adding a large number to $F(X_1, \ldots, X_n)$, which does change the $\ell_i(F)$'s. Substracting $\mathbb{E}[F(X_1, \ldots, X_n)]$ to $F(X_1, \ldots, X_n)$ is a natural way to avoid thar.

Proof of the Gaussian concentration bound (without the optimal constant)

Apply Azuma-Hoeffding inequality with

$$
\begin{aligned}
&Y = F(X_1, \ldots, X_n) \\
&\mathcal{F}_k = \sigma(X_1, \ldots, X_k),\ \mathcal{F}_0 = \{\emptyset, \Omega\} \text{ (trivial sigma-field)} \\
&\mathbb{E}[Y|\mathcal{F}_0] = \mathbb{E}[Y] \quad \text{and} \quad \mathbb{E}(Y|\mathcal{F}_n) = Y.
\end{aligned}
$$

$\{\mathbb{E}[F(X_1, \ldots, X_n)|\mathcal{F}_k]\}_{k=0}^n$ is usually called the Doob martingale associated to $F(X_1, \ldots, X_n)$.

Observe that $\Delta_k = \mathbb{E}[F(X_1, \ldots, X_n)|\mathcal{F}_k] - \mathbb{E}[F(X_1, \ldots, X_n)|\mathcal{F}_{k-1}]$ is a random variable as a function of $X_1, \ldots, X_k$.

Now let $X_1', \ldots, X_n'$ be an independent copy of $X_1, \ldots, X_n$; then

$$\mathbb{E}[Y|\mathcal{F}_{k-1}] = \mathbb{E}[F(X_1, \ldots, X_k', \ldots, X_n)|\mathcal{F}_k]$$

Then

$$\begin{aligned}\Rightarrow \quad \Delta_k &= \mathbb{E}(Y|\mathcal{F}_k) - \mathbb{E}(Y|\mathcal{F}_{k-1}) \\ &= \mathbb{E}\big[F(X_1, \ldots, X_k, \ldots, X_n) - F(X_1, \ldots, X_k', \ldots, X_n)|\mathcal{F}_k\big]\end{aligned}$$

$$\Rightarrow \quad \|\Delta_k\|_\infty \le \ell_k(F). \qquad \blacksquare$$

Looking back at the proof of Azuma-Hoefding inequality we get at once

$$\begin{aligned}
&\log \mathbb{E}\Big[\exp\Big(\lambda(F(X_1,\ldots,X_n) - \mathbb{E}[F(X_1,\ldots,X_n)])\Big)\Big] \\
&= \log \mathbb{E}\left[\exp\left(\lambda(\Delta_1 + \cdots + \Delta_n)\right)\right] \\
&\le \frac{\lambda^2}{2} \sum_{i=1}^{n} \|\Delta_i\|^2 \\
&\le \frac{\lambda^2}{2} \sum_{i=1}^{n} \ell_i(F)^2.
\end{aligned}$$

Hence

$$\mathbb{P}\Big(|F(X_1,\ldots,X_n) - \mathbb{E}[F(X_1,\ldots,X_n)]| \ge u\Big) \le 2\exp\left(\frac{-u^2}{2\sum_{i=1}^{n} \ell_i(F)^2}\right)$$

Let $\mathscr{S}$ be a measurable set (*e.g.*, a finite set, like $\{0, 1\}$).
Fix $n \in \mathbb{N}$.
Let $X_1, \ldots, X_n$ be i.i.d. r.v. taking values in $\mathscr{S}$.
Let

$$\mathrm{d}_H(\underline{x}, \underline{y}) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq y_i\}} \quad \text{(Hamming distance)}$$

where $\underline{x} = (x_1, \ldots, x_n), \underline{y} = (y_1, \ldots, y_n) \in \mathscr{S}^n$.

Pick a set $A \subset \mathscr{S}^n$ with $\mathbb{P}(A) > 0$.

Concentration on very small subsets

For every $u > 0$, one has

$$\mathbb{P}\left\{\mathrm{d}_H\big((X_1, \ldots, X_n), A\big) \geq \left(u + \sqrt{\frac{1}{2} \ln \frac{1}{\mathbb{P}(A)}}\right)\sqrt{n}\right\} \leq \mathrm{e}^{-2u^2} .$$

Interpretation:

Define the r-fattening of A as

$$[A]_r := \{\underline{z} \in \mathscr{S}^n : \mathrm{d}_H(\underline{z}, A) \leq r\}$$

$$\mathbb{P}\left\{\mathrm{d}_H\big((X_1, \ldots, X_n), A\big) \geq \left(u + \sqrt{\frac{1}{2}\ln\frac{1}{\mathbb{P}(A)}}\right)\sqrt{n}\right\} = 1 - \mathbb{P}([A]_r)$$

with

$$r = \left(u + \sqrt{\frac{1}{2}\ln\frac{1}{\mathbb{P}(A)}}\right)\sqrt{n}.$$

Numerical example: take $\mathbb{P}(A) = 10^{-6}$, u such that $u + \sqrt{(1/2)\ln(1/\mathbb{P}(A))} = 10$, then $\mathrm{e}^{-2u^2} \approx \mathrm{e}^{-108}$.

Proof

Take $F(x_1,\ldots,x_n) = \mathrm{d}_H(\underline{x}, A)$. Check that $\ell_i(F) = 1, \forall i$.
Apply the Gaussian concentration bound to $Y = F(X_1,\ldots,X_n)$:

$$\mathbb{P}(Y \geq \mathbb{E}[Y] + u) \leq \exp\left(-\frac{2u^2}{n}\right) \quad (\forall u > 0).$$

Upper bound for $\mathbb{E}[Y]$? Apply again the Gaussian concentration bound to $-\lambda Y$ with $\lambda > 0$:

$$\exp(\lambda \mathbb{E}[Y])\, \mathbb{E}\left[\exp(-\lambda Y)\right] \leq \exp\left(\frac{n\lambda^2}{8}\right) \cdot$$

But $Y \equiv 0$ on A, hence

$$\mathbb{E}\left[\exp(-\lambda Y)\right] \geq \mathbb{E}\left[\mathbb{1}_A \exp(-\lambda Y)\right] = \mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A) \cdot$$

$$\Rightarrow \quad \mathbb{E}[Y] \leq \inf_{\lambda>0}\left\{\frac{n\lambda}{8} + \frac{1}{\lambda}\ln\frac{1}{\mathbb{P}(A)}\right\} = \sqrt{\frac{n}{2}\log\frac{1}{\mathbb{P}(A)}} \cdot \qquad \blacksquare$$

Take a finite set $\mathscr{S} = \{1, \dots, \mathrm{Card}(\mathscr{S})\}$ ("alphabet").
Let $X_1, X_2, \dots$ be i.i.d. r.v. taking values in $\mathscr{S}$.
Let $X \overset{\text{law}}{=} X_i$ with distribution $\{p(1), \dots, p(\mathrm{Card}(\mathscr{S}))\}$.

SHANNON ENTROPY of this distribution:

$$H(X) = -\sum_{s=1}^{\mathrm{Card}(\mathscr{S})} p(s) \log p(s) \in [0, \log \mathrm{Card}(\mathscr{S})].$$

Two extreme cases:
1. There is s^* such that $p(s^*) = 1$: no indeterminacy at all, the next symbol is always s^* ;
2. $p(s) = 1/\mathrm{Card}(\mathscr{S})$, $\forall s \in \mathscr{S}$: maximal indeterminacy.

Asymptotic equipartition property

By using the strong law of large numbers, one has

$$-\frac{1}{n}\log \mathbb{P}(X_1,\ldots,X_n) = H(X) \quad \text{almost surely.}$$

Consequences: Let $p(x_1,\ldots,x_n) := \mathbb{P}(X_1 = x_1,\ldots,X_n = x_n)$.
Given $\epsilon > 0$ and n large enough,

$$p(x_1,\ldots,x_n) \asymp \mathrm{e}^{-n(H(X)\pm\epsilon)}$$

for all $(x_1,\ldots,x_n) \in \mathcal{G}_{n,\epsilon} \subset \mathscr{S}^n$ with $\mathbb{P}(\mathcal{G}_{n,\epsilon}) \approx 1$.
Hence

$$\frac{\mathrm{Card}(\mathcal{G}_{n,\epsilon})}{\mathrm{Card}(\mathscr{S}^n)} \asymp \exp\Big(-n(\underbrace{\log\mathrm{Card}(\mathscr{S}) - H(X)}_{>0} \pm\epsilon\Big)\,.$$

Empirical entropy

Empirical distribution:

$$\hat{p}_n(s) = \hat{p}_n(s; X_1, \ldots, X_n) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j = s\}}, \ s \in \mathscr{S}.$$

Plug-in estimator:

$$\widehat{H}_n = \widehat{H}_n(X_1, \ldots, X_n) = - \sum_{s=1}^{\mathrm{Card}(\mathscr{S})} \hat{p}_n(s) \log \hat{p}_n(s).$$

By the strong law of large numbers, $p_n(s) \xrightarrow[n\to\infty]{} p(s)$, almost surely, for each $s \in \mathscr{S}$, thus

$$\widehat{H}_n \xrightarrow[n\to\infty]{} H(X), \text{ almost surely.}$$

One has $0 \leq \widehat{H}_n \leq \log n$ and $0 \leq \mathbb{E}[\widehat{H}_n] \leq H(X)$ for every $n \in \mathbb{N}$.

How does $\widehat{H}_n$ concentrate around $\mathbb{E}[\widehat{H}_n]$?

Fluctuation bounds for empirical entropy

For all $u \geq 0$

$$\mathbb{P}\left(|\widehat{H}_n - \mathbb{E}[\widehat{H}_n]| \geq u\right) \leq 2\exp\left(-\frac{nu^2}{2(1+\log n)^2}\right).$$

In particular

$$\mathrm{Var}(\widehat{H}_n) \leq \frac{(1+\log n)^2}{n}.$$

Proof

Let

$$F(x_1,\ldots,x_n) = -\sum_{s=1}^{\mathrm{Card}(\mathscr{S})} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{x_j=s\}} \log\left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{x_j=s\}}\right).$$

Claim (homework!):

$$\ell_i(F) \leq \frac{2(1+\log n)}{n},\ \ i = 1,\ldots,n.$$

Conclude by the Gaussian concentration bound

Setting (recap):
i.i.d. r.v. $(X_1, X_2, \ldots, X_n, \ldots)$, $X_i \overset{\text{law}}{=} X$, $\mathcal{F}(x) = \mathbb{P}(X \leq x)$.

Given $x \in \mathbb{R}$ and $X_1, \ldots, X_n$ define

$$\mathcal{F}_n(x) = \mathcal{F}_n(x; X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.$$

Can we get the following?

$$\mathbb{P}\left(\sup_{x \in \mathbb{R}} |\mathcal{F}_n(x; X_1, \ldots, X_n) - \mathcal{F}(x)| \geq u \right) \leq 2\,\mathrm{e}^{-2nu^2}, \;\; n \geq 1, u > 0.$$

We are interested in the r.v.

$$\mathcal{KS}_n = \mathcal{KS}_n(X_1, \ldots, X_n) = \sup_{x \in \mathbb{R}} |\mathcal{F}_n(x) - \mathcal{F}(x)|.$$

By Glivenko-Cantelli theorem

$$\mathcal{KS}_n \xrightarrow[n\to\infty]{} 0 \quad \text{almost surely,}$$

and for all $u > 0$

$$\mathbb{P}\left(\sqrt{n}\,\mathcal{KS}_n > u\right) \xrightarrow[n\to\infty]{} 2\sum_{r\geq 1}(-1)^{r-1}\exp(-2u^2r^2).$$

(Kolmogorov-Smirnov test)

The easy part

Consider

$$F(X_1, \dots, X_n) = \sup_x |\mathcal{F}_n(x) - \mathcal{F}(x)|.$$

Check that

$$\ell_i(F) = \frac{1}{n}, \ \ i = 1, \dots, n.$$

Thus, by the Gaussian concentration bound, for all $u > 0$, for all $n \in \mathbb{N}$,

$$\mathbb{P}\left(|\mathcal{KS}_n - \mathbb{E}[\mathcal{KS}_n]| \geq u\right) \leq 2\,\exp(-2nu^2)$$

and

$$\mathbb{P}\left(|\sqrt{n}\,\mathcal{KS}_n - \mathbb{E}[\sqrt{n}\,\mathcal{KS}_n]| \geq u\right) \leq 2\,\exp(-2u^2).$$

The tricky part: Getting rid of $\mathbb{E}[\sqrt{n}\,\mathcal{KS}_n]$

Dvoretsky-Kiefer-Wolfowitz inequality

$$\mathbb{P}\left(\sqrt{n}\,\mathcal{KS}_n \geq u\right) \leq 4\,\exp\left(-u^2/8\right), \quad \forall u > 0.$$

Clever proof only using elementary steps such that one boils down to

$$\mathbb{E}\left[\mathrm{e}^{n\lambda \mathcal{KS}_n}\right] \leq 4\mathbb{E}\left[\mathrm{e}^{2\lambda \sum_{i=1}^n \varepsilon_i}\right],\ \lambda > 0,$$

where $\mathbb{P}(\varepsilon_i = \pm 1) = 1/2$ (our initial example!).

Optimal bound (Massart, 1990):

$$\mathbb{P}\left(\sqrt{n}\,\mathcal{KS}_n \geq u\right) \leq 2\,\exp(-2u^2), \quad \forall u > 0.$$

Proof: very technical.

A remark

Given two probability distributions $\mathbb{P}_X$ and $\mathbb{P}_Y$ on $\mathbb{R}$ with cumulative distribution functions $\mathcal{F}_X$ and $\mathcal{F}_Y$,

$$d_{Kolmo}(\mathbb{P}_X, \mathbb{P}_Y) = \sup_{x \in \mathbb{R}} |\mathcal{F}_X(x) - \mathcal{F}_Y(x)|$$

is the Kolmogorov distance between them.

Another possible distance is the Kantorovich distance:

$$\begin{aligned} d_{Kanto}\big(\mathbb{P}_X, \mathbb{P}_Y\big) &= \sup \left\{ \int g \, \mathrm{d}\mathbb{P}_X - \int g \, \mathrm{d}\mathbb{P}_Y : |g(x) - g(y)| \leq |x - y| \right\} \\ &= \int |\mathcal{F}_X(x) - \mathcal{F}_Y(x)| \, \mathrm{d}x \leq d_{Kolmo}(\mathbb{P}_X, \mathbb{P}_Y)\,. \end{aligned}$$

Lecture III:
Markov chains

Plan of the lectures

SO FAR:

- $X_1, X_2, \ldots$ independent r.v. taking values in $\mathscr{S}$, *i.e.* **product measures** on $\mathscr{S}^{\mathbb{N}}$:

  $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \times \cdots \times \mathbb{P}(X_n = x_n).$

- **Martingale decomposition:**

$$F(X_1, \ldots, X_n) - \mathbb{E}[F(X_1, \ldots, X_n)] = \sum_{i=1}^{n} \Delta_i$$

  where the Δ_i's are the increments of the Doob martingale associated to $F(X_1, \ldots, X_n)$: **This is completely general.**

Now: $\{X_i\}$ form a Markov chain (non-product measure).

Preliminaries

We start in a rather abstract context: let $\mathscr{S}$ be a metric space with distance d.

Separately Lipschitz functions:
$F : \mathscr{S}^{\mathbb{Z}} \to \mathbb{R}$ such that

$$\ell_i(F) = \sup \left\{ \frac{|F(\underline{x}) - F(\underline{y})|}{d(x_i, y_i)} : x_j = y_j,\ \forall j \neq i,\ x_i \neq y_i \right\}$$

where $\underline{x} = (\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots) \in \mathscr{S}^{\mathbb{Z}}$

One can think of a function $F(x_1, \ldots, x_n)$ as a function on $\mathscr{S}^{\mathbb{Z}}$ with $\ell_i(F) = 0$ for $i > n$ and $i \leq 0$.

Preliminaries: continuation & ending

Let $\{X_i\}_{i\in\mathbb{Z}}$ be a stationary process where the X_i's take values in $\mathscr{S}$. Denote by $\mathbb{P}$ its joint distribution.

Let:

- $\mathcal{F}^i_{-\infty}$ be the sigma-field generated by $\{X_k : k \leq i\}$;
- $\mathcal{F} = \sigma\left(\cup_{i=-\infty}^{+\infty}\mathcal{F}^i_{-\infty}\right)$;
- $\mathcal{F}_{-\infty} = \bigcap_i \mathcal{F}^i_{-\infty}$ (tail sigma-field).

Assume that $\mathbb{P}$ is tail-trivial (*i.e.*, $\forall A \in \mathcal{F}_{-\infty}, \mathbb{P}(A) \in \{0,1\}$).

The basic telescoping

Recall that *in general*

$$F - \mathbb{E}(F) = \sum_{i \in \mathbb{Z}} \Delta_i$$

where

$$\Delta_i = \Delta_i(X_{-\infty}^i) = \mathbb{E}[F|\mathcal{F}_i] - \mathbb{E}[F|\mathcal{F}_{i-1}]$$

with $\mathcal{F}_i = \sigma(X_{-\infty}^i)$.

Technical remark: one needs $F \in L^1(\mathbb{P})$ to use Lévy Upward and Downward theorems.

Some notations

$\mathbb{P}_{X_{-\infty}^{i}}$: the joint distribution of $\{X_j, j \geq i+1\}$ given $X_{-\infty}^{i}$.

$\widehat{\mathbb{P}}_{X_{-\infty}^{i}, Y_{-\infty}^{i}}$: a coupling of $\mathbb{P}_{X_{-\infty}^{i}}$ and $\mathbb{P}_{Y_{-\infty}^{i}}$.

For $-\infty \leq i < j \leq +\infty$:
$X_i^j := X_i, X_{i+1}, \ldots, X_j$ and $x_i^j := x_i, x_{i+1}, \ldots, x_j$.

The second telescoping

$$\Delta_i = \Delta_i(X_{-\infty}^i) =$$

$$\int d\mathbb{P}_{X_{-\infty}^{i-1}}(z_i) \int d\widehat{\mathbb{P}}_{X_{-\infty}^i, X_{-\infty}^{i-1} z_i}(y_{i+1}^\infty, z_{i+1}^\infty)\left[F(X_{-\infty}^i y_{i+1}^\infty) - F(X_{-\infty}^{i-1} z_i z_{i+1}^\infty)\right].$$

Now insert the inequality

$$F(\underline{x}) - F(\underline{y}) \leq \sum_{k \in \mathbb{Z}} \ell_k(F)\, d(x_k, y_k)$$

to get

$$\Delta_i \leq \sum_{j=0}^{\infty} D_{i,i+j}\, \ell_{i+j}(F)$$

The coupling matrix D

We have introduced the upper-triangular random matrix

$$D_{i,i+j} = D_{i,i+j}^{X_{-\infty}^{i}} =$$

$$\int \mathrm{d}\mathbb{P}_{X_{-\infty}^{i-1}}(z_i) \int \mathrm{d}\widehat{\mathbb{P}}_{X_{-\infty}^{i},\, X_{-\infty}^{i-1} z_i}(y_{i+1}^{\infty}, z_{i+1}^{\infty})\, d(y_{i+j}, z_{i+j})$$

where $i \in \mathbb{Z}$, $j \in \mathbb{N}$, and $D_{i,i} = 1$ ($\forall i \in \mathbb{Z}$).

FOR THE SAKE OF CONCRETENESS:
consider a Markov chain $\{X_n\}_{n\in\mathbb{Z}}$ with *discrete* state space $\mathscr{S}$ equipped with the discrete distance $d(x, y) = \delta_{xy}$.

Moreover assume that the transition kernel $P = (p(x, y))_{(x,y)\in\mathscr{S}\times\mathscr{S}}$ is irreducible and aperiodic.

Finally, assume positive recurrence, *i.e.*, for some (hence, all) $x \in \mathscr{S}$, $\mathbb{E}_x[T_x] < +\infty$, where $T_x = \inf\{n \in \mathbb{N} : X_n = x\}$.

Therefore, there is a unique invariant probability distribution π for the chain, so

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \pi(x_0)p(x_0, x_1)\cdots p(x_{n-1}, x_n).$$

The Markovian case

Take a Markovian coupling and use stationarity:

$$D_{i,i+j}^{X_{-\infty}^{i}} = D_{i,i+j}^{X_{i-1},X_i} = \sum_{z\in\mathscr{S}} p(X_{i-1}, z) \int \mathrm{d}\widehat{\mathbb{P}}_{X_i,z}(u_0^{\infty}, v_0^{\infty}) d(u_j, v_j).$$

Defining the COUPLING TIME

$$T(u_0^{\infty}, v_0^{\infty}) = \inf\{k \geq 0 : u_i = v_i,\ \forall i \geq k\}$$

we have

$$d(u_j, v_j) \leq \mathbb{1}_{\{T(u_0^{\infty}, v_0^{\infty}) \geq j\}}$$

whence

$$D_{i,i+j}^{X_{i-1},X_i} \leq \sum_{z\in\mathscr{S}} p(X_{i-1}, z)\, \widehat{\mathbb{P}}_{X_i,z}(T \geq j).$$

Gaussian concentration bound

Recap:

$$F - \mathbb{E}(F) = \sum_{i \in \mathbb{Z}} \Delta_i \quad \text{and}$$

$$\Delta_i(X_{i-1}, X_i) \leq \sum_{z \in \mathscr{S}} p(X_{i-1}, z) \sum_{j=0}^{\infty} \widehat{\mathbb{P}}_{X_i, z}(T \geq j)\, \ell_{i+j}(F)$$

Now apply Azuma-Hoeffding inequality:

$$\log \mathbb{E}\big[\exp(F - E(F))\big] \leq \frac{1}{2} \sum_{i \in \mathbb{Z}} \|\Delta_i(X_{i-1}, X_i)\|_\infty^2 .$$

After some (uninteresting) work, one gets

$$\sum_{i \in \mathbb{Z}} \|\Delta_i(X_{i-1}, X_i)\|_\infty^2 \leq \frac{\zeta(1+\varepsilon)}{2} \left(\sup_{u,v \in \mathscr{S}} \widehat{\mathbb{E}}_{u,v}\left(T^{1+\varepsilon}\right) \right)^2 \times \sum_{i \in \mathbb{Z}} \ell_i(F)^2$$

where $\varepsilon > 0$ is arbitrary.

Gaussian concentration bound

Theorem

Let $\{X_n\}_{n\in\mathbb{Z}}$ be a Markov chain as above. There exists a constant $D > 0$ such that, for all separately bounded Lipschitz functions $F : \mathscr{S}^{\mathbb{Z}} \to \mathbb{R}$, we have

$$\log \mathbb{E}\Big[\exp(F - \mathbb{E}(F))\Big] \leq D \sum_{i\in\mathbb{Z}} \ell_i(F)^2$$

where

$$D = \frac{\zeta(1+\varepsilon)}{4}\left(\sup_{u,v\in\mathscr{S}} \widehat{\mathbb{E}}_{u,v}\big(T^{1+\varepsilon}\big)\right)^2 .$$

Back to H-of-C

The simplest example: $\mathscr{S}$ finite

Use Doeblin's coupling: let $\{(X_n, Y_n)\}_n$ be the Markov chain on the state space $\mathscr{S} \times \mathscr{S}$ which evolved in the following way:

- $\{X_n\}$ and $\{Y_n\}$ evolve independently according to the transition kernel P until the first (random) time $X_n = Y_n$ and
- they evolve together with the same transition kernel after that time.

Equivalently, define the transition kernel on $\mathscr{S} \times \mathscr{S}$ by

$$q((x,y),(x',y')) := \begin{cases} p(x,x')p(y,y') & \text{if} \quad x \neq y \\ p(x,x') & \text{if} \quad x = y \text{ and } x' = y' \\ 0 & \text{otherwise.} \end{cases}$$

The diagonal $\{(x,y) \in \mathscr{S} \times \mathscr{S} : x = y\}$ is an absorbing set.

Lemma (Doeblin)

$$\exists\, \rho \in (0,1), \exists c > 0 \quad \text{such that} \quad \sup_{u,v\in\mathscr{S}} \widehat{\mathbb{P}}_{u,v}(T \geq j) \leq c\,\rho^j,\ \forall j \in \mathbb{N}.$$

This is more than enough to get $\sup_{u,v\in\mathscr{S}} \widehat{\mathbb{E}}_{u,v}\left(T^{1+\varepsilon}\right) < +\infty$.

Proof of the lemma

Recall that T is the coupling time (“coalescence” time).
Irreducibility and aperiodicity of P mean that there exists $m \geq 1$ such that

$$\min_{x,y \in \mathscr{S}} p^{(m)}(x, y) =: \varepsilon > 0.$$

Then

$$\widehat{\mathbb{P}}_{(u,v)}(T \leq m) \geq \sum_{z \in \mathscr{S}} p^{(m)}(u, z) p^{(m)}(v, z) \geq \varepsilon \sum_{z \in \mathscr{S}} p^{(m)}(v, z) = \varepsilon$$

for all $(u, v) \in \mathscr{S} \times \mathscr{S}$. The Markov property yields

$$\widehat{\mathbb{P}}_{(u,v)}(T \leq km) \geq 1 - (1 - \varepsilon)^k .$$

Lecture IV: Markov chains

(continuation & ending)

Beyond the Gaussian concentration bound: moment concentration bounds

What happens if we don't get a uniform (in X_{i-1}, X_i) decay of $D_{i,i+j}^{X_{i-1},X_i}$ as a function of j ?
This means that $\|\Delta_i(X_{i-1}, X_i)\|_\infty = +\infty$, hence Azuma-Hoeffding inequality is not applicable, hence a Gaussian concentration bound is out of hope.

Answer: we may obtain only *moment bounds*. We have to replace Azuma-Hoeffding inequality by another martingale inequality, namely *Burkholder inequality*: for $p \in \mathbb{N}$,

$$\mathbb{E}\left[\big(F - \mathbb{E}(F)\big)^{2p}\right] \leq (2p-1)^{2p}\, \mathbb{E}\left[\left(\sum_{i\in\mathbb{Z}} \Delta_i^2\right)^p\right].$$

(The constant $(2p-1)^{2p}$ is optimal.)

Moment concentration bound of order p

Theorem

Let $p \in \mathbb{N}$ and let F be separately Lipschitz and $L^{2p}(\mathbb{P})$-integrable. Then, for every $\epsilon > 0$,

$$\mathbb{E}\left[\left(F - \mathbb{E}(F)\right)^{2p}\right] \leq C_p \left(\sum_{i \in \mathbb{Z}} \ell_i(F)^2\right)^p$$

where

$$C_p = (2p-1)^{2p} \left(\frac{\zeta(1+\epsilon)}{2}\right)^p \times \sum_{x,y \in \mathscr{S}} \pi(x) p(x,y) \left(\sum_{z \in \mathscr{S}} p(x,z) \widehat{\mathbb{E}}_{y,z}\left[(T+1)^{1+\frac{\epsilon}{2}}\right]\right)^{2p}.$$

If, for some $p \in \mathbb{N}$, the previous inequality holds then, by Markov's inequality,

$$\mathbb{P}\left(|F - \mathbb{E}[F]| \geq u\right) \leq C_p \, \frac{\left(\sum_{i\in\mathbb{Z}} \ell_i(F)^2\right)^p}{u^{2p}}, \; u > 0.$$

We get an algebraic decay.

An instructive example: the "house-of-cards" process

- $\mathscr{S} = \{0, 1, 2, \dots\}$;
- For all $k \in \mathbb{Z}_+$, $\mathbb{P}(X_{k+1} = x + 1 | X_k = x) = 1 - q_x$ and $\mathbb{P}(X_{k+1} = 0 | X_k = x) = q_x$, $x \in \mathscr{S}$;
- $0 < q_x < 1$, $x \in \mathscr{S}$.

The transition kernel is irreducible and aperiodic.

One can prove that it is a positive recurrent Markov chain if and only if

$$\sum_n \prod_{k=0}^{n} (1 - q_k) < +\infty \,.$$

The three different concentration regimes

Three cases:

(1) $q := \inf\{q_x : x \in \mathscr{S}\} > 0 \Rightarrow$ Gaussian concentration bound. GCB

(2) $\exists \alpha \in (0,1)$ s.t. $q_x = x^{-\alpha}$

$$\Rightarrow \quad \forall p \in \mathbb{N},\ \mathbb{E}\left[(F - \mathbb{E}(F))^{2p}\right] \leq C_p \left(\sum_i \ell_i(F)^2\right)^p$$

where C_p grows too fast with p to get a Gaussian concentration bound.

(3) $\exists \gamma > 0$ s.t. $q_x = \gamma / x \Rightarrow$ moment concentration bound up to some critical $p(\gamma)$.

One can construct a coupling such that

$$\widehat{\mathbb{P}}_{x,y}(T \geq j) \leq \prod_{k=0}^{j-1}(1 - q^*_{x+k}),\ x \geq y$$

where $q^*_n = \inf\{q_s : s \leq n\}$.

The coupling

It is based of the representation of the chain as a random recursion: take $U_1, U_2, \dots$ i.i.d. r.v. uniformly distributed on $[0, 1]$; then

$$X_{k+1} = (X_k + 1)\mathbb{1}_{\{U_{k+1} \geq q_{X_k}\}} \,.$$

The coupling works as follows: run two copies of the chain starting from different initial states; when they hit the ground (state 0) together for the first time, then they stay together forever.

A final remark on Markov chains

Theorem (Dedecker-Gouëzel, 2015)

For an irreducible aperiodic Markov chain with a *general* state space $\mathscr{S}$, the Gaussian concentration bound holds if, and only if, the chain is *geometrically ergodic*.

(Based on coupling ideas.)

Lecture V:
Gibbs measures

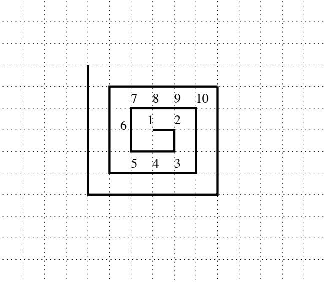
Plan of the lectures

Previously:
Markov chains with state space $\mathscr{S}$ $\rightsquigarrow$ non-product measures on $\mathscr{S}^{\mathbb{Z}}$.

Now:
Gibbs measures, which are non-product measures on $\mathscr{S}^{\mathbb{Z}^d}$, $d \geq 2$, where we take $\mathscr{S} = \{-1, +1\}$ (spins) for definiteness.

Strategy: same as for Markov chains, that is, introduce a "coupling matrix" $(D_{\boldsymbol{i},\boldsymbol{j}})_{\boldsymbol{i},\boldsymbol{j} \in \mathbb{Z}^d}$ indexed by d-dimensional integers.

The basic telescoping & the coupling "matrix"



Enumeration of $\mathbb{Z}^d$:

$$e : \mathbb{Z}^d \to \mathbb{N}$$

$$(\leq \boldsymbol{i}) := \{\boldsymbol{j} \in \mathbb{Z}^d : e(\boldsymbol{j}) \leq e(\boldsymbol{i})\}$$

$\mathcal{F}_{\leq \boldsymbol{i}}$: $\sigma-$field generated by $\omega_{\boldsymbol{j}}, \boldsymbol{j} \leq \boldsymbol{i}$

$$F : \mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R},\ \omega \in \mathscr{S}^{\mathbb{Z}^d} (\text{“configuration”})$$

$$F - \mathbb{E}[F] = \sum_{\boldsymbol{i} \in \mathbb{Z}^2} \Delta_{\boldsymbol{i}}, \quad \Delta_{\boldsymbol{i}} = \mathbb{E}[F|\mathcal{F}_{\leq \boldsymbol{i}}] - \mathbb{E}[F|\mathcal{F}_{< \boldsymbol{i}}]$$

$$D_{\boldsymbol{i},\boldsymbol{j}}^{\omega_{\leq i}} = \widehat{\mathbb{P}}_{\boldsymbol{i},+,-}\left(\omega_{\boldsymbol{j}}^{(1)} \neq \omega_{\boldsymbol{j}}^{(2)}\right)$$

where we couple

$$\mathbb{P}(\cdot|\omega_{<\boldsymbol{i},+_{\boldsymbol{i}}}) \quad \text{and} \quad \mathbb{P}(\cdot|\omega_{<\boldsymbol{i},-_{\boldsymbol{i}}}).$$

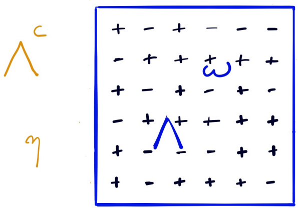
Boltzmann-Gibbs kernel

$$\gamma_\Lambda^{(\boldsymbol{\beta})}(\omega|\eta) = \frac{\exp\left(-\boldsymbol{\beta}\,\mathcal{H}_\Lambda(\omega|\eta)\right)}{Z_\Lambda^{(\boldsymbol{\beta})}(\eta)}, \quad \Lambda \Subset \mathbb{Z}^d, \omega, \eta \in \mathscr{S}^{\mathbb{Z}^d}.$$

$\rightsquigarrow$ Gibbs measures on $\mathscr{S}^{\mathbb{Z}^d}$ depending on η in general (DLR equation)

Parameter $\beta \geq 0$: inverse temperature

Special case: $\beta = 0$ (infinite temperature)
$\rightsquigarrow$ uniform **product measure** ($\rightsquigarrow$ Gaussian concentration bound).

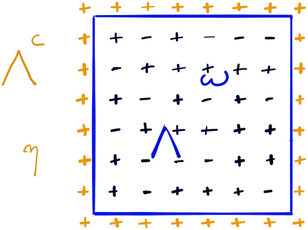


Ising model (Markov random field)

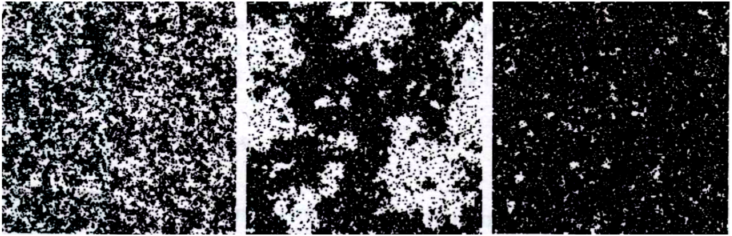
$$\mathcal{H}_\Lambda(\omega|\eta) = - \sum_{\substack{\boldsymbol{i},\boldsymbol{j}\in\Lambda \\ \|\boldsymbol{i}-\boldsymbol{j}\|_1=1}} \omega_{\boldsymbol{i}}\,\omega_{\boldsymbol{j}} - \sum_{\substack{\boldsymbol{i}\in\partial\Lambda,\, \boldsymbol{j}\notin\Lambda \\ \|\boldsymbol{i}-\boldsymbol{j}\|_1=1}} \omega_{\boldsymbol{i}}\,\eta_{\boldsymbol{j}}$$

$\eta_{\boldsymbol{j}} = +1, \forall \boldsymbol{j} \in \mathbb{Z}^d$ ("+-boundary condition"), gives rise to μ^+.

Fact: there exists a unique Gibbs measure μ for all $\beta < \beta_c$, whereas there are several ones for all $\beta > \beta_c$, depending on η, in fact, two extremal ones: μ^+ and μ^-.



Phase transition in the Ising model for $d = 2$



β increases from left to right

'$+$': black, '$-$': white

$\beta_c = (1/2)\sinh^{-1}(1) \approx 0.4407$

The magentization

Let $M_n(\omega) = \sum_{\boldsymbol{i} \in C_n} s_0(T_{\boldsymbol{i}}\,\omega)$, where $s_0(\omega) = \omega_0$, be the total magnetization in C_n, and where $(T_{\boldsymbol{i}}\,\omega)_{\boldsymbol{j}} = \omega_{\boldsymbol{j}-\boldsymbol{i}}$ (shift operator). Then

$$\frac{M_n(\omega)}{(2n+1)^d}$$

is the magnetization per spin in C_n. For any shift-invariant probability measure ν on $\mathscr{S}^{\mathbb{Z}^d}$,

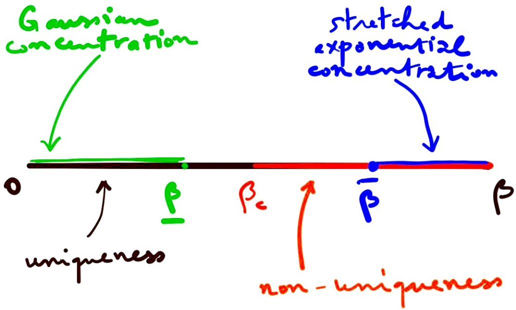
$$\mathbb{E}_\nu\left[\frac{M_n(\omega)}{(2n+1)^d}\right] = \mathbb{E}_\nu[s_0]$$

is the mean magnetization per site (magnetization, for short) wrt ν.

The following is well-known for the Ising model ($d \geq 2$):

- for $\beta < \beta_c$, $\mathbb{E}_\mu[s_0] = 0$;
- for $\beta > \beta_c$, $\mathbb{E}_{\mu^+}[s_0] \neq 0$.

Concentration for the Ising model



Let $F : \mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R}$ and

$$\ell_{\boldsymbol{i}}(F) = \sup_{\omega \in \mathscr{S}^{\mathbb{Z}^d}} |F(\omega^{(\boldsymbol{i})}) - F(\omega)|, \ \boldsymbol{i} \in \mathbb{Z}^d,$$

where $\omega^{(\boldsymbol{i})}$ is obtained from ω by flipping the spin at $\boldsymbol{i}$.

Theorem: Gaussian concentration bound ($\beta < \underline{\beta}$)

Let μ be the (unique) Gibbs measure of the Ising model. There exists a constant $D > 0$ such that, for all functions F with $\sum_{\boldsymbol{i} \in \mathbb{Z}^d} \ell_{\boldsymbol{i}}(F)^2 < +\infty$, one has

$$\mathbb{E}_\mu\left[\exp(F - \mathbb{E}_\mu(F))\right] \leq \exp\left(D \sum_{\boldsymbol{i} \in \mathbb{Z}^d} \ell_{\boldsymbol{i}}(F)^2\right).$$

Remark. As shown by C. Külske, the Gaussian concentration bounds holds in the Dobrushin uniqueness regime with $D = 2(1 - \mathfrak{c}(\gamma))^{-2}$, where $\mathfrak{c}(\gamma)$ is Dobrushin's contraction coefficient.

Recall that the Gaussian concentration implies that for all $u \geq 0$ one has

$$\mu\Big(\omega \in \mathscr{S}^{\mathbb{Z}^d} : |F(\omega) - \mathbb{E}_\mu[F]| \geq u\Big) \leq 2\exp\left(\frac{-u^2}{4D\sum_{\boldsymbol{i}\in\mathbb{Z}^d} \ell_{\boldsymbol{i}}(F)^2}\right).$$

At sufficiently low temperature, we can gather all moment bounds to obtain the following. We denote by μ^+ the Gibbs measure for the $+$-phase of the Ising model.

Theorem: Stretched-exponential concentration bound ($\beta > \overline{\beta}$)

There exists $\varrho = \varrho(\beta) \in (0,1)$ and $c_\varrho > 0$ such that for all functions F with $\sum_{\boldsymbol{i}\in\mathbb{Z}^d} \ell_{\boldsymbol{i}}(F)^2 < +\infty$, for all $u \geq 0$, one has

$$\mu^+\Big(\omega \in \mathscr{S}^{\mathbb{Z}^d} : |F(\omega) - \mathbb{E}_{\mu^+}[F]| \geq u\Big) \leq 4\exp\left(\frac{-\,c_\varrho\, u^\varrho}{\left(\sum_{\boldsymbol{i}\in\mathbb{Z}^d} \ell_{\boldsymbol{i}}(F)^2\right)^{\frac{\varrho}{2}}}\right).$$

Some applications

Other models besides the standard Ising model: Potts, long-range Ising, etc.

- Ergodic sums in *arbitrarily shaped* volumes;
- Fluctuations in the Shannon-McMillan-Breiman theorem;
- First occurrence of a pattern of configuration in another configuration;
- Bounding $\bar{d}$-distance by relative entropy;
- Fattening patterns;
- Almost-sure central limit theorems;
- Speed of convergence of the empirical measure.

Application 1: Almost-sure central limit theorems (only part of the story)

This application shows that one can also get *limit theorems* out of concentration inequalities.

Informal statement:
If you know that the central limit theorem holds for some function $f : \mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R}$ wrt to a shift-invariant probability measure, and if you can prove that this measure satisfies a *moment concentration bound of order* 2, then the almost-sure central limit theorem holds in the sense of Kantorovich distance.

(Cf. Chazottes-Collet-Redig 2016-paper for a precise statement.)

Given $f : \mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R}$ and ν a shift-invariant probability measure on $\mathscr{S}^{\mathbb{Z}^d}$, the usual form of the CLT is: for all $u \in \mathbb{R}$

$$\lim_{n\to\infty} \nu \left\{ \omega \in \mathscr{S}^{\mathbb{Z}^d} : \frac{\sum_{\boldsymbol{i}\in C_n} f(T_{\boldsymbol{i}}\omega)}{(2n+1)^{\frac{d}{2}}} \leq u \right\} = G_{0,\sigma_f}\big((-\infty, u]\big)$$

where

$$\sigma_f^2 = \sum_{\boldsymbol{i}\in\mathbb{Z}^d} \int f \cdot f \circ T_{\boldsymbol{i}} \,\mathrm{d}\nu \in (0, +\infty).$$

and where G_{0,σ_f} is the Gaussian measure with mean 0 and variance σ_f.

The CLT can be re-written as

$$\lim_{n\to\infty} \mathbb{E}_\nu \left[\mathbb{1}_{\left\{ \sum_{\boldsymbol{i}\in C_n} f(T_{\boldsymbol{i}}\cdot)/(2n+1)^{\frac{d}{2}} \leq u \right\}} \right] = G_{0,\sigma_f}\big((-\infty, u]\big).$$

The ASCLT consists in replacing $\mathbb{E}_\nu$ by a point-wise logarithmic average and get an almost-sure version of the CLT: for all $u \in \mathbb{R}$

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{n=1}^{N} \frac{1}{n} \mathbb{1}_{\left\{ \sum_{\boldsymbol{i}\in C_n} f(T_{\boldsymbol{i}}\,\omega)/(2n+1)^{\frac{d}{2}} \leq u \right\}} = G_{0,\sigma_f}\big((-\infty, u]\big)$$

for ν-a.e. ω.

We will only formulate two results for $f = s_0$ (magnetization).

To state the theorems, define

$$d_{Kanto}(\nu_1, \nu_2) = \sup \left(\mathbb{E}_{\nu_1}(g) - \mathbb{E}_{\nu_2}(g) \right)$$

where the sup is taken over all functions $g : \mathbb{R} \to \mathbb{R}$ that are 1-Lipschitz.

Metrizes the weak topology on the set of probability measures on $\mathbb{R}$ with a first moment.

High-temperature Ising model

Theorem

Let $\beta < \underline{\beta}$. Then

$$\lim_{N\to\infty} d_{Kanto}\left(\frac{1}{\log N}\sum_{n=1}^{N}\frac{1}{n}\,\delta_{M_n(\omega)/(2n+1)^{\frac{d}{2}}},\, G_{0,\sigma^2}\right) = 0$$

where

$$\sigma^2 = \sum_{\boldsymbol{i}\in\mathbb{Z}^d}\int s_0 \cdot s_0 \circ T_{\boldsymbol{i}}\,\mathrm{d}\mu \in (0,\infty).$$

Low-temperature Ising model

Theorem

Let $\beta > \overline{\beta}$. Then

$$\lim_{N\to\infty} d_{Kanto}\left(\frac{1}{\ln N}\sum_{n=1}^{N}\frac{1}{n}\,\delta_{(M_n(\omega)-\mathbb{E}_{\mu^+}[s_0])/(2n+1)^{\frac{d}{2}}}, G_{0,\sigma^2}\right) = 0$$

where

$$\sigma^2 = \sum_{\boldsymbol{i}\in\mathbb{Z}^d}\int s_0\cdot s_0\circ T_{\boldsymbol{i}}\,\mathrm{d}\mu^+ \in (0,\infty).$$

Application 2: "SPEED" OF CONVERGENCE OF THE EMPIRICAL MEASURE

Take $\Lambda \Subset \mathbb{Z}^d$ and $\omega \in \mathscr{S}^{\mathbb{Z}^d}$ and let

$$\mathcal{E}_\Lambda(\omega) = \frac{1}{|\Lambda|} \sum_{\boldsymbol{i} \in \Lambda} \delta_{T_{\boldsymbol{i}}\,\omega}$$

where $(T_{\boldsymbol{i}}\,\omega)_{\boldsymbol{j}} = \omega_{\boldsymbol{j}-\boldsymbol{i}}$ (shift operator).

Let μ be an ergodic measure on $\mathscr{S}^{\mathbb{Z}^d}$. If $(\Lambda_n)_n$ is a sequence of cube $\uparrow \mathbb{Z}^d$ (more generally, a van Hove sequence), then

$$\mathcal{E}_{\Lambda_n}(\omega) \xrightarrow[\text{weakly}]{n\to\infty} \mu.$$

Question: If μ is a Gibbs measure, what is the "speed" of this convergence?

KANTOROVICH DISTANCE on the set of probability measures on $\mathscr{S}^{\mathbb{Z}^d}$:

$$d_{Kanto}(\mu_1, \mu_2) = \sup_{\substack{G:\mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R} \\ G \text{ 1-Lipshitz}}} \left(\mathbb{E}_{\mu_1}(G) - \mathbb{E}_{\mu_2}(G)\right)$$

where $|G(\omega) - G(\omega')| \leq d(\omega, \omega') = 2^{-k}$, where k is the sidelength of the largest cube in which ω and ω' coincide.

Lemma. Let μ be a probability measure and

$$F(\omega) = \sup_{\substack{G:\mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R} \\ G \text{ 1-Lipshitz}}} \left(\sum_{\boldsymbol{i}\in\Lambda} G(T_{\boldsymbol{i}}\omega) - \mathbb{E}_\mu(G)\right).$$

Then

$$\sum_{\boldsymbol{i}\in\mathbb{Z}^d} \ell_{\boldsymbol{i}}(F)^2 \leq c_d\,|\Lambda|$$

where $c_d > 0$ depends only on d.

Proof

Ising model at high & low temperature

Gaussian concentration for the empirical measure ($\beta < \underline{\beta}$)

Let μ be the (unique) Gibbs measure of the Ising model. There exists a constant $C > 0$ such that, for all $\Lambda \Subset \mathbb{Z}^d$ and for all $u \geq 0$, one has

$$\mu\Big\{\omega \in \mathscr{S}^{\mathbb{Z}^d} : \Big| d_{Kanto}(\mathcal{E}_\Lambda(\omega), \mu) - \mathbb{E}_\mu\big[d_{Kanto}(\mathcal{E}_\Lambda(\cdot), \mu)\big]\Big| \geq u\Big\}$$

$$\leq 2 \exp\big(- C\,|\Lambda| u^2\big).$$

We denote by μ^+ the Gibbs measure for the $+$-phase of the Ising model.

Stretched-exponential concentration for the empirical measure ($\beta > \overline{\beta}$)

There exist $\varrho = \varrho(\beta) \in (0,1)$ and a constant $c_\varrho > 0$ such that, for all $\Lambda \Subset \mathbb{Z}^d$ and for all $u \geq 0$, one has

$$\mu^+\Big\{\omega \in \mathscr{S}^{\mathbb{Z}^d} : \Big| d_{\scriptscriptstyle Kanto}(\mathcal{E}_\Lambda(\omega), \mu^+) - \mathbb{E}_{\mu^+}\big[d_{\scriptscriptstyle Kanto}(\mathcal{E}_\Lambda(\cdot), \mu^+)\big]\Big| \geq u\Big\}$$
$$\leq 4 \exp\left(-c_\varrho |\Lambda|^{\frac{\varrho}{2}} u^\varrho\right).$$

Can we estimate $\mathbb{E}_\mu\left[d_{Kanto}(\mathcal{E}_\Lambda(\cdot),\mu)\right]$?

Let

$$\mathscr{L} = \left\{G : \mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R} : G \text{ 1-Lipschitz}\right\}$$

and

$$\mathcal{Z}_G^\Lambda := \frac{1}{|\Lambda|}\sum_{\boldsymbol{i}\in\Lambda}\left(G\circ T_{\boldsymbol{i}} - \mathbb{E}_\mu(G)\right),\ \Lambda \Subset \mathbb{Z}^d.$$

Then

$$\mathbb{E}_\mu\left[d_{Kanto}(\mathcal{E}_\Lambda(\cdot),\mu)\right] = \mathbb{E}_\mu\left(\sup_{G\in\mathscr{L}}\mathcal{Z}_G^\Lambda\right).$$

Notice that we have functions defined on a Cantor space, which is really different from the case of, say, $[0,1]^k \subset \mathbb{R}^k$.

Theorem

Let μ be a probability measure on $\mathscr{S}^{\mathbb{Z}^d}$ satisfying the Gaussian concentration bound. Then

$$\mathbb{E}_\mu\left[d_{\textit{Kanto}}\left(\mathcal{E}_\Lambda(\cdot),\mu\right)\right] \preceq \begin{cases} |\Lambda|^{-\frac{1}{2}(1+\log|\mathscr{S}|)^{-1}} & \text{if} \quad d=1 \\ \exp\left(-\frac{1}{2}\left(\frac{\log|\Lambda|}{\log|\mathscr{S}|}\right)^{1/d}\right) & \text{if} \quad d\geq 2. \end{cases}$$

For (a_Λ) and (b_Λ) indexed by finite subsets of $\mathbb{Z}^d$ we denote $a_\Lambda \preceq b_\Lambda$ if, for every sequence (Λ_n) such that $|\Lambda_n|\to+\infty$ as $n\to+\infty$, we have $\limsup_n \frac{\log a_{\Lambda_n}}{\log b_{\Lambda_n}} \leq 1$.

It is possible to get *bounds* but they are really messy.

Some references

By no means exhaustive!

- J.-R. Chazottes, P. Collet, F. Redig. *On concentration inequalities and their applications for Gibbs measures in lattice systems*. Preprint (2016). https://arxiv.org/abs/1610.06502
- J.-R. Chazottes, P. Collet, C. Külske, F. Redig. *Concentration inequalities for random fields via coupling*. Probab. Theory Relat. Fields (2007). https://arxiv.org/abs/math/0503483
- J.-R. Chazottes, F. Redig. *Concentration inequalities for Markov processes via coupling*. Electron. J. Probab. (2009). https://arxiv.org/abs/0810.0097
- D.P. Dubhashi, A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009.

- A. Kontorovich, M. Raginsky. Concentration of measure without independence: a unified approach via the martingale method. Preprint (2016)

  https://arxiv.org/abs/1602.00721
- C. Külske. *Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures.* Commun. Math. Phys. (2003).

  http://www.ruhr-uni-bochum.de/imperia/md/content/mathematik/lehrstuhl-ii/kuelske-pub/autocorrdobrushin-fulltext.pdf
- C. McDiarmid. *On the method of bounded differences.* Surveys in Combinatorics (1989). https://www.stats.ox.ac.uk/__data/assets/pdf_file/0009/4113/bdd-diffs89.pdf

- M. Ledoux. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89. American Mathematical Society, 2001.
- M. Raginsky, I. Sason. *Concentration of measure inequalities in information theory, communications and coding*. Foundations and Trends in Communications and Information Theory, vol. 10, issues 1 and 2, pp. 1-246, 2013

  `https://arxiv.org/pdf/1212.4663.pdf`

Geometrically ergodic irreducible aperiodic Markov chain

There exists a set $C \subset \mathscr{S}$ ("small set"), an integer $m > 0$, a probability measure ν, and $\delta \in (0, 1)$, $\kappa > 1$, such that

- For all $x \in C$ one has $P^m(x, \cdot) \geq \delta\nu$;
- The return time τ_C to C is such that $\sup_{x \in C} \mathbb{E}_x(\kappa^{\tau_C}) < \infty$.

If $\mathscr{S}$ is countable then this is equivalent to the fact that the return time to some (or equivalently any) point has an exponential moment.

◂ Dedecker-Gouëzel Theorem

DLR equation

μ is a Gibbs measure for a given potential Φ if, for all $\Lambda \Subset \mathbb{Z}^d$ and for all $A \in \mathfrak{B}(\mathscr{S}^{\mathbb{Z}^d})$

$$\mu(A) = \int \mathrm{d}\mu(\eta) \sum_{\omega' \in \Lambda} \gamma_\Lambda(\omega' | \eta) \, \mathbb{1}_A(\omega'_\Lambda \eta_{\Lambda^c})$$

where Φ is a real-valued function having two arguments: a finite subset of $\mathbb{Z}^d$ and a configuration $\omega \in \mathscr{S}^{\mathbb{Z}^d}$, and where

$$\mathcal{H}_\Lambda(\omega | \eta) = \sum_{\Lambda' \cap \Lambda \neq \emptyset} \Phi(\Lambda', \omega_\Lambda \eta_{\mathbb{Z}^d \setminus \Lambda})$$

where Λ' runs through the set of finite subsets of $\mathbb{Z}^d$.

◂ Boltzmann-Gibbs kernel

Dobrushin contraction coefficient

Let

$$C_{\boldsymbol{i},\boldsymbol{j}}(\gamma) = \sup_{\substack{\omega,\omega' \in \mathscr{S}^{\mathbb{Z}^d} \\ \omega_{\mathbb{Z}^d \setminus \boldsymbol{j}} = \omega'_{\mathbb{Z}^d \setminus \boldsymbol{j}}}} \|\gamma_{\{\boldsymbol{i}\}}(\cdot|\omega) - \gamma_{\{\boldsymbol{i}\}}(\cdot|\omega')\|_\infty .$$

Then in our context $C_{\boldsymbol{i},\boldsymbol{j}}$ only depends on $\boldsymbol{i} - \boldsymbol{j}$ and we define

$$\mathfrak{c}(\gamma) = \sum_{\boldsymbol{i} \in \mathbb{Z}^d} C_{0,\boldsymbol{i}}(\gamma).$$

Dobrushin's uniqueness regime: $\mathfrak{c}(\gamma) < 1$.

◂ Gaussian concentration bound

van Hove sequence

A sequence $(\Lambda_n)_n$ of nonempty finite subsets of $\mathbb{Z}^d$ is said to tend to infinity in the sense of van Hove if, for each $\boldsymbol{i} \in \mathbb{Z}^d$, one has

$$\lim_{n\to+\infty} |\Lambda_n| = +\infty \quad \text{and} \quad \lim_{n\to+\infty} \frac{|(\Lambda_n + \boldsymbol{i})\backslash\Lambda_n|}{|\Lambda_n|} = 0.$$

◂ Empirical measure

Proof of the Lemma

Let $\omega, \omega' \in \mathscr{S}^{\mathbb{Z}^d}$ and $G : \mathscr{S}^{\mathbb{Z}^d} \to \mathbb{R}$ be a 1-Lipschitz function. Without loss of generality, we can assume that $\mathbb{E}_\mu(G) = 0$. We have

$$\sum_{\boldsymbol{i}\in\Lambda} G(T_{\boldsymbol{i}}\,\omega) \leq \sum_{\boldsymbol{i}\in\Lambda} G(T_{\boldsymbol{i}}\,\omega') + \sum_{\boldsymbol{i}\in\Lambda} d(T_{\boldsymbol{i}}\,\omega, T_{\boldsymbol{i}}\,\omega').$$

Taking the supremum over 1-Lipschitz functions thus gives

$$F(\omega) - F(\omega') \leq \sum_{\boldsymbol{i}\in\Lambda} d(T_{\boldsymbol{i}}\,\omega, T_{\boldsymbol{i}}\omega').$$

We can interchange ω and ω' in this inequality, whence

$$|F(\omega) - F(\omega')| \leq \sum_{\boldsymbol{i}\in\Lambda} d(T_{\boldsymbol{i}}\,\omega, T_{\boldsymbol{i}}\,\omega').$$

Now we assume that there exists $\boldsymbol{k} \in \mathbb{Z}^d$ such that $\omega_{\boldsymbol{j}} = \omega'_{\boldsymbol{j}}$ for all $\boldsymbol{j} \neq \boldsymbol{k}$. This means that $d(T_{\boldsymbol{i}}\,\omega, T_{\boldsymbol{i}}\,\omega') \leq 2^{-\|\boldsymbol{k}-\boldsymbol{i}\|_\infty}$ for all $\boldsymbol{i} \in \mathbb{Z}^d$, whence

$$\ell_{\boldsymbol{k}}(F) \leq \sum_{\boldsymbol{i}\in\Lambda} 2^{-\|\boldsymbol{k}-\boldsymbol{i}\|_\infty}.$$

Therefore, using Young's inequality,

$$\begin{aligned}\sum_{\boldsymbol{i}\in\mathbb{Z}^d} \ell_{\boldsymbol{i}}(F)^2 &\leq \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \left(\sum_{\boldsymbol{i}\in\mathbb{Z}^d} \mathbb{1}_\Lambda(\boldsymbol{i})\, 2^{-\|\boldsymbol{k}-\boldsymbol{i}\|_\infty} \right)^2 \\ &\leq \sum_{\boldsymbol{i}\in\mathbb{Z}^d} \mathbb{1}_\Lambda(\boldsymbol{i}) \times \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^d} 2^{-\|\boldsymbol{k}\|_\infty} \right)^2.\end{aligned}$$

We thus obtain the desired estimate with $c_d = \left(\sum_{\boldsymbol{k}\in\mathbb{Z}^d} 2^{-\|\boldsymbol{k}\|_\infty} \right)^2$. ∎

◂ Kantorovich distance