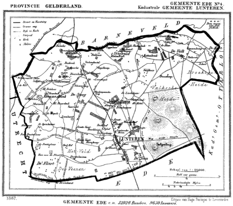
CONCENTRATION INEQUALITIES: AN INTRODUCTION & SOME RECENT RESULTS

Jean-René CHAZOTTES
CNRS & Ecole Polytechnique, Paris

Lunteren, November 13-16



Concentration-of-measure phenomenon

Shows up in PROBABILITY THEORY and STATISTICS, as well as in various areas such as GEOMETRY, FUNCTIONAL ANALYSIS (local theory of Banach spaces), DISCRETE MATHEMATICS (randomized algorithms, random graphs, etc).

Underlying principle

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 two talks:
Making sense out of this, for independent random variables, Markov chains, and Gibbs measures.

Plan

1. Independent Random Variables
   - Toy model
   - Azuma, Hoeffding & McDiarmid
   - Three applications

2. Markov chains & Gibbs measures

1. Independent Random Variables

Toy model

$S_n = X_1 + \cdots + X_n$, $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$, X_i independent ($\mathbb{E}(S_n) = 0$) One can prove (see below)

$$\mathbb{P}\left(S_n \geq u\right) \leq \exp\left(-\frac{u^2}{2n}\right), \quad \forall u > 0 \quad \text{(Chernov, 1952).}$$

A. Scale of "large deviations": $u \rightsquigarrow un$

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

The law of large numbers

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

$$\Downarrow$$

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

$$\Downarrow$$

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

Large deviations: asymptotic & non-asymptotic

Take $u > 0$. One has

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

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

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

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

Recall that

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

B. SCALE OF THE CENTRAL LIMIT THEOREM: $u \rightsquigarrow u\sqrt{n}$

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

(Numerical example: $n = 100$, $u = 5$, the above probability is $\leq 7.5\,10^{-6}$)

Interlude : the Gaussian paradise

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

Denote $S_n = Z_1 + \cdots + Z_n$.

Since $S_n/\sqrt{n} \stackrel{\text{law}}{=} \mathcal{N}(0,1)$, one gets

$$\max\left(0, 1 - \frac{1}{u^2}\right) \frac{2}{u\sqrt{2\pi}}\, e^{-\frac{u^2}{2}} \le \mathbb{P}\left(|S_n| \ge u\sqrt{n}\right) \le \frac{2}{u\sqrt{2\pi}}\, e^{-\frac{u^2}{2}},$$

for all $u > 0$.

Central limit theorem: asymptotic & non-asymptotic

Back to $X_i = \pm 1$, $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$, X_i independent.

Take $u > 0$. One has

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

and

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

where C = absolute constant > 0.
So one has to take $n \approx e^{u^2}$ to get back the previous inequality!

A FIRST GENERALIZATION

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

AIM: replace

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

by

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

under mild assumptions on F.

A key abstract result

No independence needed!

Let

- $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ 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

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

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

Proof of Azuma-Hoeffding inequality

Take $\lambda > 0$:

$$\mathbb{P}\left(Y - \mathbb{E}(Y) \geq u\right) = \mathbb{P}\Big(\exp\left(\lambda(Y - \mathbb{E}(Y))\right) \geq \exp(\lambda u)\Big)$$

$$\underset{\substack{\text{Markov}\\ \text{ineq.}}}{\leq} \exp(-\lambda u)\ \underbrace{\mathbb{E}\left[\exp\left(\lambda(Y - \mathbb{E}(Y))\right)\right]}_{=\mathbb{E}[\exp(\lambda(\Delta_n + \cdots + \Delta_1))]}.$$

Now

$$\mathbb{E}\left[\exp\left(\lambda(\Delta_n + \cdots + \Delta_1)\right)\right] =$$

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

But

$$\mathbb{E}\left(\exp(\lambda\Delta_k \big| \mathcal{F}_{k-1})\right) \leq \exp\left(\frac{\lambda^2}{8}\operatorname{osc}(\Delta_k)^2\right) \quad \text{(Hoeffding Lemma)}$$

Proof of Azuma-Hoeffding inequality (continued)

By induction one gets

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

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

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

The same holds for $-Y$. $\blacksquare$

Hoeffding lemma (1963)

Let Z be a random variable with $\mathbb{E}(Z) = 0$ and $a \leq Z \leq b$, and set

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

Then

$$\psi(\lambda) \leq \frac{\lambda^2 (b-a)^2}{8},\ \forall \lambda \in \mathbb{R}.$$

In the above proof, we took $Z = \Delta_k$ *(so* $b - a = \operatorname{osc}(\Delta_k)$*) and* $\mathbb{E} = \mathbb{E}(\cdot | \mathcal{F}_{k-1})$*.*

A (nice) proof of Hoeffding lemma

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

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

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

Let $X_1, \ldots, X_n$ be independent random variables, each taking values in a set $\mathscr{S}$.

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

$$|F(x_1, \ldots, x_n) - F(y_1, \ldots, y_n)| \leq \sum_{i=1}^{n} \ell_i(F)\, \mathbb{1}_{\{x_i \neq y_i\}}$$

for all $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$.

In other words:
if $x_j = y_j$, $j \neq i$ and $x_i \neq y_i$ then maximal oscillation of F is $\ell_i(F)$.

Gaussian concentration bound (McDiarmid, 1989)

For all functions with the bounded differences property,

$$\forall \lambda \in \mathbb{R},\ \mathbb{E}\Big[\exp\big(\lambda(F(X_1,\dots,X_n)-\mathbb{E}[F(X_1,\dots,X_n)])\big)\Big] \leq \exp\left(\frac{\lambda^2}{8}\sum_{i=1}^{n}\ell_i(F)^2\right).$$

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

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

Hence

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

Illustration (back to our toy model)

Back to $X_i = \pm 1$, $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$, X_i independent:

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

Hence $\forall u \geq 0$

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

Remarks

A drawback of the gaussian concentration bound : based on worst case changes of F!

Insensitivity to the variance of the X_i's:
take $X_1, \ldots, X_n$ i.i.d. r.v. taking values in $[-1,1]$ with $\mathbb{E}(X_i) = 0$; we get the same inequality as for $X_i = \pm 1$ with $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$ which has the largest possible variance among such r.v..

Possible cure: Bernstein inequality.

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

Apply Azuma-Hoeffding inequality with

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

Now let $X'_1, \ldots, X'_n$ be an independent copy of $X_1, \ldots, X_n$; then

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

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

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

Consider a finite set $\mathscr{S}$, fix $n \in \mathbb{N}$. Consider $X_1, \dots, X_n$ i.i.d. r.v. taking values in $\mathscr{S}$.
Let

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

where $\underline{x} = (x_1, \dots, x_n), \underline{y} = (y_1, \dots, y_n) \in \mathscr{S}^n$.
Now, pick a set $A \subset \mathscr{S}^n$ with (say) $\mathbb{P}(A) = 1/2$, and for $r \in \{0, 1, \dots, n\}$, let

$$[A]_r = \{\underline{z} \in \mathscr{S}^n : d_H(\underline{z}, A) \leq r\} \quad (r - \text{fattening})$$

where

$$d_H(\underline{z}, A) = \inf_{\underline{y} \in A} d_H(\underline{y}, \underline{z}).$$

Natural choice: $r = \lfloor n\epsilon \rfloor$, with $\epsilon \in (0, 1)$.

Sets of measure 1/2 are big

One has

$$\mathbb{P}\big([A]_{n\epsilon}\big) \geq 1 - \exp\left(-2n\left(\epsilon - \sqrt{\frac{\log 2}{2n}}\right)_+^2\right)$$

where $(u)_+ := \max(0, u)$.

Proof

Take $F(x_1, \ldots, x_n) = d_H(\underline{x}, A)$. Check that $\ell_i(F) = 1$, $i = 1, \ldots, n$.
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

$$\mathbb{E}\left[\exp(-\lambda Y)\right] \geq \mathbb{E}\left[\mathbb{1}_A \exp(-\lambda Y)\right] = \mathbb{E}[\mathbb{1}_A] = \frac{1}{2} \cdot$$

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

We conclude that, for every $u > 0$,

$$\mathbb{P}\left(Y \geq \underbrace{\sqrt{\frac{n\log 2}{2}} + u}_{=:\, v}\right) \leq \mathbb{P}(Y \geq \mathbb{E}[Y] + u) \leq \exp\left(-\frac{2u^2}{n}\right)$$

where $v > \sqrt{\frac{n\log 2}{2}}$.
Finally, take $v = n\epsilon$ and use that by definition

$$\mathbb{P}(Y \geq n\epsilon) = 1 - \mathbb{P}([A]_{n\epsilon}).$$

$\blacksquare$

Take a finite set $\mathscr{S}$ ("alphabet") and consider $X_1, X_2, \ldots$ i.i.d. r.v. taking values in $\mathscr{S}$.
Let $X \overset{\text{law}}{=} X_i$ with distribution $\mathbb{P} = \{p(s), s \in \mathscr{S}\}$.

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

Empirical distribution:

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

Plug-in estimator:

$$\widehat{H}_n = \widehat{H}_n(X_1, \ldots, X_n) = -\sum_{s\in\mathscr{S}} p_n(s)\log p_n(s).$$

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 $\mathbb{E}[\widehat{H}_n]$?

Theorem

Let

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

Claim:

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

Hence, by the Gaussian concentration bound, for all $u \ge 0$

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

In particular

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

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

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

$$\begin{aligned}\mathcal{F}_n(x) &= \mathcal{F}_n(x; X_1, \dots, X_n)\\ &= \frac{1}{n} \operatorname{Card}\left(\{1 \leq i \leq n : X_i \leq x\}\right)\\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}.\end{aligned}$$

One has $\mathbb{1}_{\{X_i \leq x\}} \stackrel{\text{law}}{=} \text{Bernoulli}(\mathcal{F}(x))$.

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 $u > 0$

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

(Kolmogorov-Smirnov test)

The easy part

Consider

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

Check that

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

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 considerations.)

Optimal bound (Massart, 1990):

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

SO FAR:

- $X_1, X_2, \ldots$ independent r.v. taking values in $\mathscr{S}$, *i.e.* **product measures** on $\mathscr{S}^{\mathbb{N}}$;
- **Martingale approach**.
  Various other approaches are available (MARTON, TALAGRAND, LEDOUX, BOBKOV & GÖTZE, and many others);

TALK 2: non-product measures such as Markov chains and Gibbs measures.

2. Markov chains & Gibbs measures

I will present some results from joint works with P. Collet, C. Külske and F. Redig.

(A glimpse of) Markov chains

Several approaches (Marton, Samson, Kontorovich, Paulin, and many others).

Here: combination of the martingale method and coupling.

For the sake of simplicity:
Markov chain $(X_n)_{n\in\mathbb{Z}}$ with *discrete* state space $\mathscr{S}$ equipped with the discrete distance $d(x,y)=\delta_{xy}$.

Separately Lipschitz functions

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

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

where $\underline{x} = (\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots)$

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 $i \leq 0$.

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

Some notation

$$\mathbb{P}_{X_{-\infty}^i}:$$

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 $\mathbb{P}_{Y_{-\infty}^i}$.

The second telescoping (still in the general case)

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

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

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

$$D_{i,i+j} = D^{X^i_{-\infty}}_{i,i+j} =$$

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

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

The Markovian case

If $(X_n)_{n\in\mathbb{Z}}$ is a **Markov chain** with discrete state space $\mathscr{S}$ with the discrete distance and transition kernel $P = (p(x,y))_{(x,y)\in\mathscr{S}\times\mathscr{S}}$, then, taking a Markovian coupling,

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

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:

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

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

Recall that Azuma-Hoeffding inequality is

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

After some work, one gets

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

where $\epsilon > 0$.

Gaussian concentration bound

There exists a constant $D > 0$ such that, for all separately Lipshitz functions $F : \mathscr{S}^{\mathbb{Z}} \to \mathbb{R}$, we have

$$\mathbb{E}\big[\exp(F - \mathbb{E}(F))\big] \leq \exp\left(D \sum_{i \in \mathbb{Z}} \ell_i(F)^2\right).$$

Back to H-of-C

The simplest example: aperiodic irreducible Markov chain with $\mathscr{S}$ finite

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

Beyond the Gaussian case

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?

Answer: we may obtain only *moment bounds*.

Illustration with the House-of-cards process (Berbee):

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

One can construct (explicitly) 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 $q^*_n = \inf\{q_s : s \leq n\}$.

Three cases:

(1) $q := \inf\{q_x : x \in \mathscr{S}\} > 0$;

(2) $q_x = x^{-\alpha}$ with $0 < \alpha < 1$;

(3) $q_x = \gamma/x$ where $\gamma > 0$.

(1) Gaussian concentration bound; ◂ GCB

(2) Moments of all orders:

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

where C_{2p} is independent of F, but grows too fast with p to get a Gaussian concentration bound;

(3) Moments up to some $p(\gamma)$.

Moment inequalities rely on *Burkholder inequality*:

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

A final remark on Markov chains

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*.

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

GIBBS MEASURES: non-product measures on $\mathscr{S}^{\mathbb{Z}^d}$ where we take $\mathscr{S} = \{-1, +1\}$ (spins) for definiteness.

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

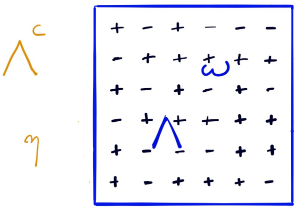
Boltzmann-Gibbs kernel

$$\gamma_\Lambda(\omega|\eta) = \frac{\exp\big(-\beta\mathcal{H}_\Lambda(\omega|\eta)\big)}{Z_\Lambda(\eta)}, \ \Lambda \Subset \mathbb{Z}^d.$$

$\rightsquigarrow$ Gibbs measures (DLR equation)

Parameter $\beta \geq 0$: inverse temperature

One extreme case: $\beta = 0 \rightsquigarrow$ uniform product measure (for which one has the Gaussian concentration bound).

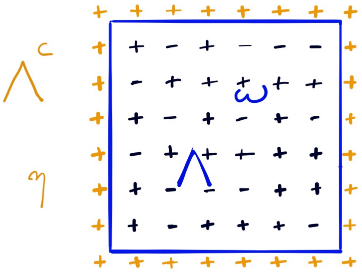


Ising model (Markov random field)

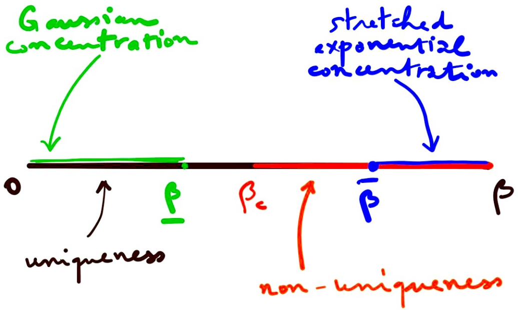
$$\mathcal{H}_\Lambda(\omega|\eta) = -\sum_{\substack{\boldsymbol{i},\boldsymbol{j}\in\Lambda \\ \|\boldsymbol{i}-\boldsymbol{j}\|_1=1}} \omega_{\boldsymbol{i}}\times\omega_{\boldsymbol{j}} - \sum_{\substack{\boldsymbol{i}\in\partial\Lambda,\, \boldsymbol{j}\notin\Lambda \\ \|\boldsymbol{i}-\boldsymbol{j}\|_1=1}}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

$\eta_{\boldsymbol{i}} = +1, \forall \boldsymbol{i} \in \mathbb{Z}^d$ ("+-boundary condition")

Fact: there exists a unique Gibbs measure for all $\beta < \beta_c$, whereas there are several ones for all $\beta > \beta_c$.



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 $\boldsymbol{i}$.

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

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

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 $u \geq u$ 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.

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

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

Applications

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

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

Application to 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 $\mathscr{S}^{\mathbb{Z}^d}$:

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

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

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

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

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

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

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

Let

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

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

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

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

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

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 ugly.

References *directly related* to these two talks

This is by no means a bibliography on the subject!

- K. Azuma. *Weighted Sums of Certain Dependent Random Variables*. T̂hoku Mathematical Journal (1967). https://www.jstage.jst.go.jp/article/tmj1949/19/3/19_3_357/_pdf
- 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

- R. van Handel. *Probability in high dimension*. Lecture notes (2014).
https://www.princeton.edu/~rvan/ORF570.pdf
- W. Hoeffding. *Probability inequalities for sums of bounded random variables*. Journal of the American Statistical Association (1963). http://www.csee.umbc.edu/~lomonaco/f08/643/hwk643/Hoeffding.pdf
- 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

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 the given potential if, for all $\Lambda \Subset \mathbb{Z}^d$ and for all $A \in \mathfrak{B}(\mathscr{S}^{\mathbb{Z}^d})$

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

◂ 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