Concentration inequalities and the entropy method

Gábor Lugosi

ICREA and Pompeu Fabra University

Barcelona

what is concentration?

We are interested in bounding random fluctuations of functions of many independent random variables.

what is concentration?

We are interested in bounding random fluctuations of functions of many independent random variables.

$X_1, \ldots, X_n$ are independent random variables taking values in some set $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and

$$Z = f(X_1, \ldots, X_n) .$$

How large are "typical" deviations of Z from $\mathbb{E}Z$?
In particular, we seek upper bounds for

$$\mathbb{P}\{Z > \mathbb{E}Z + t\} \quad \text{and} \quad \mathbb{P}\{Z < \mathbb{E}Z - t\}$$

for $t > 0$.

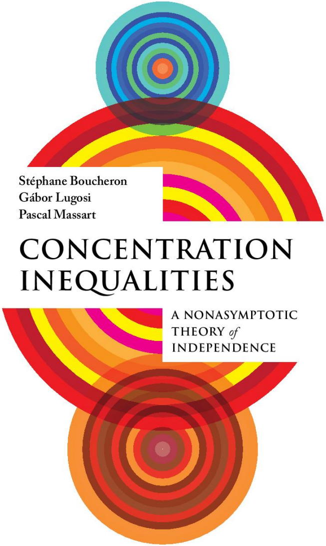
various approaches

- martingales (Yurinskii, 1974; Milman and Schechtman, 1986; Shamir and Spencer, 1987; McDiarmid, 1989,1998);

- information theoretic and transportation methods (Alhswede, Gács, and Körner, 1976; Marton 1986, 1996, 1997; Dembo 1997);

- Talagrand's induction method, 1996;

- logarithmic Sobolev inequalities (Ledoux 1996, Massart 1998, Boucheron, Lugosi, Massart 1999, 2001).



Stéphane Boucheron
Gábor Lugosi
Pascal Massart

CONCENTRATION INEQUALITIES

A NONASYMPTOTIC THEORY *of* INDEPENDENCE

OXFORD

chernoff bounds

By Markov's inequality, if $\lambda > 0$,

$$\mathbb{P}\{Z - \mathbb{E}Z > t\} = \mathbb{P}\left\{e^{\lambda(Z-\mathbb{E}Z)} > e^{\lambda t}\right\} \le \frac{\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}}{e^{\lambda t}}$$

Next derive bounds for the moment generating function $\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}$ and optimize λ.

chernoff bounds

By Markov's inequality, if $\lambda > 0$,

$$\mathbb{P}\{Z - \mathbb{E}Z > t\} = \mathbb{P}\left\{e^{\lambda(Z-\mathbb{E}Z)} > e^{\lambda t}\right\} \le \frac{\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}}{e^{\lambda t}}$$

Next derive bounds for the moment generating function $\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}$ and optimize λ.

If $Z = \sum_{i=1}^{n} X_i$ is a sum of independent random variables,

$$\mathbb{E}e^{\lambda Z} = \mathbb{E}\prod_{i=1}^{n} e^{\lambda X_i} = \prod_{i=1}^{n} \mathbb{E}e^{\lambda X_i}$$

by independence. It suffices to find bounds for $\mathbb{E}e^{\lambda X_i}$.

chernoff bounds

By Markov's inequality, if $\lambda > 0$,

$$\mathbb{P}\{Z - \mathbb{E}Z > t\} = \mathbb{P}\left\{e^{\lambda(Z-\mathbb{E}Z)} > e^{\lambda t}\right\} \leq \frac{\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}}{e^{\lambda t}}$$

Next derive bounds for the moment generating function $\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)}$ and optimize λ.

If $Z = \sum_{i=1}^n X_i$ is a sum of independent random variables,

$$\mathbb{E}e^{\lambda Z} = \mathbb{E}\prod_{i=1}^n e^{\lambda X_i} = \prod_{i=1}^n \mathbb{E}e^{\lambda X_i}$$

by independence. It suffices to find bounds for $\mathbb{E}e^{\lambda X_i}$.

Serguei Bernstein (1880-1968)

Herman Chernoff (1923–)

hoeffding's inequality

If $X_1, \ldots, X_n \in [0, 1]$, then

$$\mathbb{E} e^{\lambda(X_i - \mathbb{E} X_i)} \leq e^{\lambda^2/8} .$$

hoeffding's inequality

If $X_1, \ldots, X_n \in [0,1]$, then

$$\mathbb{E}e^{\lambda(X_i - \mathbb{E}X_i)} \le e^{\lambda^2/8} .$$

We obtain

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]\right| > t\right\} \le 2e^{-2nt^2}$$

Wassily Hoeffding (1914–1991)

bernstein's inequality

Hoeffding's inequality is distribution free. It does not take variance information into account.
Bernstein's inequality is an often useful variant:
Let $X_1, \ldots, X_n$ be independent such that $X_i \le 1$. Let $v = \sum_{i=1}^n \mathbb{E}\left[X_i^2\right]$. Then

$$\mathbb{P}\left\{\sum_{i=1}^n (X_i - \mathbb{E}X_i) \ge t\right\} \le \exp\left(-\frac{t^2}{2(v + t/3)}\right) .$$

martingale representation

$X_1, \ldots, X_n$ are independent random variables taking values in some set $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and

$$Z = f(X_1, \ldots, X_n) .$$

Denote $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot | X_1, \ldots, X_i]$. Thus, $\mathbb{E}_0 Z = \mathbb{E} Z$ and $\mathbb{E}_n Z = Z$.

martingale representation

$X_1, \ldots, X_n$ are independent random variables taking values in some set $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and

$$Z = f(X_1, \ldots, X_n) \ .$$

Denote $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot | X_1, \ldots, X_i]$. Thus, $\mathbb{E}_0 Z = \mathbb{E} Z$ and $\mathbb{E}_n Z = Z$.

Writing

$$\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z \ ,$$

we have

$$Z - \mathbb{E} Z = \sum_{i=1}^{n} \Delta_i$$

This is the Doob martingale representation of Z.

martingale representation

$X_1, \ldots, X_n$ are independent random variables taking values in some set $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and

$$Z = f(X_1, \ldots, X_n) \ .$$

Denote $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot | X_1, \ldots, X_i]$. Thus, $\mathbb{E}_0 Z = \mathbb{E} Z$ and $\mathbb{E}_n Z = Z$.

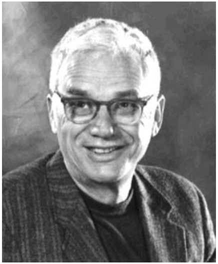
Writing

$$\Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z \ ,$$

we have

$$Z - \mathbb{E} Z = \sum_{i=1}^{n} \Delta_i$$

This is the Doob martingale representation of Z.



Joseph Leo Doob (1910–2004)

martingale representation: the variance

$$\mathbf{Var}\,(\mathbf{Z}) = \mathbb{E}\left[\left(\sum_{i=1}^{n} \Delta_i\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[\Delta_i^2\right] + 2\sum_{j>i} \mathbb{E}\Delta_i\Delta_j\ .$$

Now if $j > i$, $\mathbb{E}_i\Delta_j = 0$, so

$$\mathbb{E}_i\Delta_j\Delta_i = \Delta_i\mathbb{E}_i\Delta_j = 0\ ,$$

We obtain

$$\mathbf{Var}\,(\mathbf{Z}) = \mathbb{E}\left[\left(\sum_{i=1}^{n} \Delta_i\right)^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[\Delta_i^2\right]\ .$$

martingale representation: the variance

$$\mathbf{Var}(\mathbf{Z}) = \mathbb{E}\left[\left(\sum_{i=1}^{n}\Delta_i\right)^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[\Delta_i^2\right] + 2\sum_{j>i}\mathbb{E}\Delta_i\Delta_j \ .$$

Now if $j > i$, $\mathbb{E}_i\Delta_j = 0$, so

$$\mathbb{E}_i\Delta_j\Delta_i = \Delta_i\mathbb{E}_i\Delta_j = 0 \ ,$$

We obtain

$$\mathbf{Var}(\mathbf{Z}) = \mathbb{E}\left[\left(\sum_{i=1}^{n}\Delta_i\right)^2\right] = \sum_{i=1}^{n}\mathbb{E}\left[\Delta_i^2\right] \ .$$

From this, using independence, it is easy derive the Efron-Stein inequality.

efron-stein inequality (1981)

Let $X_1, \ldots, X_n$ be independent random variables taking values in $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and $Z = f(X_1, \ldots, X_n)$.
Then

$$\mathbf{Var}(Z) \leq \mathbb{E} \sum_{i=1}^{n} (Z - \mathbb{E}^{(i)} Z)^2 = \mathbb{E} \sum_{i=1}^{n} \mathbf{Var}^{(i)}(Z) \ .$$

where $\mathbb{E}^{(i)} Z$ is expectation with respect to the i-th variable X_i only.

efron-stein inequality (1981)

Let $\mathbf{X_1, \ldots, X_n}$ be independent random variables taking values in $\mathcal{X}$. Let $\mathbf{f} : \mathcal{X}^n \to \mathbb{R}$ and $\mathbf{Z = f(X_1, \ldots, X_n)}$.
Then

$$\mathbf{Var}(\mathbf{Z}) \le \mathbb{E} \sum_{i=1}^{n} (\mathbf{Z} - \mathbb{E}^{(i)}\mathbf{Z})^2 = \mathbb{E} \sum_{i=1}^{n} \mathbf{Var}^{(i)}(\mathbf{Z}) \; .$$

where $\mathbb{E}^{(i)}\mathbf{Z}$ is expectation with respect to the $\mathbf{i}$-th variable $\mathbf{X_i}$ only.

We obtain more useful forms by using that

$$\mathbf{Var}(\mathbf{X}) = \frac{1}{2}\mathbb{E}(\mathbf{X} - \mathbf{X}')^2 \quad \text{and} \quad \mathbf{Var}(\mathbf{X}) \le \mathbb{E}(\mathbf{X} - \mathbf{a})^2$$

for any constant $\mathbf{a}$.

efron-stein inequality (1981)

If $X'_1, \ldots, X'_n$ are independent copies of $X_1, \ldots, X_n$, and

$$Z'_i = f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n),$$

then

$$\mathrm{Var}(Z) \leq \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{n}(Z - Z'_i)^2\right]$$

Z is concentrated if it doesn't depend too much on any of its variables.

efron-stein inequality (1981)

If $X'_1, \ldots, X'_n$ are independent copies of $X_1, \ldots, X_n$, and

$$Z'_i = f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n),$$

then

$$\mathrm{Var}(Z) \leq \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{n}(Z - Z'_i)^2\right]$$

Z is concentrated if it doesn't depend too much on any of its variables.
If $Z = \sum_{i=1}^{n} X_i$ then we have an equality. Sums are the "least concentrated" of all functions!

efron-stein inequality (1981)

If for some arbitrary functions $\mathbf{f_i}$

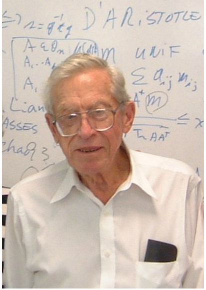
$$\mathbf{Z_i = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)} \ ,$$

then

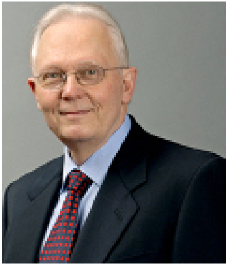
$$\mathbf{Var(Z) \leq \mathbb{E}\left[\sum_{i=1}^{n} (Z - Z_i)^2\right]}$$

efron, stein, and steele

Bradley Efron



Charles Stein



Mike Steele

weakly self-bounding functions

$f : \mathcal{X}^n \to [0, \infty)$ is weakly (a, b)-self-bounding if there exist $f_i : \mathcal{X}^{n-1} \to [0, \infty)$ such that for all $x \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right)^2 \leq af(x) + b \, .$$

weakly self-bounding functions

$\mathbf{f} : \mathcal{X}^n \to [0, \infty)$ is weakly (a, b)-self-bounding if there exist $f_i : \mathcal{X}^{n-1} \to [0, \infty)$ such that for all $x \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right)^2 \le af(x) + b\,.$$

Then

$$\mathrm{Var}(f(X)) \le a\mathbb{E}f(X) + b\ .$$

self-bounding functions

If

$$0 \leq f(x) - f_i(x^{(i)}) \leq 1$$

and

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right) \leq f(x) \ ,$$

then f is self-bounding and $\mathbf{Var}(f(X)) \leq \mathbb{E}f(X)$.

self-bounding functions

If

$$0 \leq f(x) - f_i(x^{(i)}) \leq 1$$

and

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right) \leq f(x) ,$$

then f is self-bounding and $\mathbf{Var}(f(X)) \leq \mathbb{E}f(X)$.

Rademacher averages, random VC dimension, random VC entropy, longest increasing subsequence in a random permutation, are all examples of self bounding functions.

self-bounding functions

If

$$0 \le f(x) - f_i(x^{(i)}) \le 1$$

and

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right) \le f(x) ,$$

then f is self-bounding and $\mathbf{Var}(f(X)) \le \mathbb{E}f(X)$.

Rademacher averages, random VC dimension, random VC entropy, longest increasing subsequence in a random permutation, are all examples of self bounding functions.

Configuration functions.

example: uniform deviations

Let $\mathcal{A}$ be a collection of subsets of $\mathcal{X}$, and let $X_1, \dots, X_n$ be n random points in $\mathcal{X}$ drawn i.i.d.
Let

$$P(A) = \mathbb{P}\{X_1 \in A\} \quad \text{and} \quad P_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \in A}$$

If $Z = \sup_{A \in \mathcal{A}} |P(A) - P_n(A)|$,

$$\mathrm{Var}(Z) \leq \frac{1}{2n}$$

example: uniform deviations

Let $\mathcal{A}$ be a collection of subsets of $\mathcal{X}$, and let $X_1, \ldots, X_n$ be n random points in $\mathcal{X}$ drawn i.i.d.
Let

$$P(A) = \mathbb{P}\{X_1 \in A\} \quad \text{and} \quad P_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \in A}$$

If $Z = \sup_{A \in \mathcal{A}} |P(A) - P_n(A)|$,

$$\mathrm{Var}(Z) \le \frac{1}{2n}$$

regardless of the distribution and the richness of $\mathcal{A}$.

beyond the variance

$X_1, \ldots, X_n$ are independent random variables taking values in some set $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and $Z = f(X_1, \ldots, X_n)$. Recall the Doob martingale representation:

$$Z - \mathbb{E}Z = \sum_{i=1}^{n} \Delta_i \quad \text{where} \quad \Delta_i = \mathbb{E}_i Z - \mathbb{E}_{i-1} Z ,$$

with $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot | X_1, \ldots, X_i]$.

To get exponential inequalities, we bound the moment generating function $\mathbb{E}e^{\lambda(Z - \mathbb{E}Z)}$.

azuma's inequality

Suppose that the martingale differences are bounded: $|\Delta_i| \leq c_i$. Then

$$\begin{aligned}
\mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} &= \mathbb{E}e^{\lambda\left(\sum_{i=1}^{n}\Delta_i\right)} = \mathbb{E}\mathbb{E}_n e^{\lambda\left(\sum_{i=1}^{n-1}\Delta_i\right)+\lambda\Delta_n} \\
&= \mathbb{E}e^{\lambda\left(\sum_{i=1}^{n-1}\Delta_i\right)}\mathbb{E}_n e^{\lambda\Delta_n} \\
&\leq \mathbb{E}e^{\lambda\left(\sum_{i=1}^{n-1}\Delta_i\right)} e^{\lambda^2 c_n^2/2} \text{ (by Hoeffding)} \\
&\cdots \\
&\leq e^{\lambda^2\left(\sum_{i=1}^{n} c_i^2\right)/2} .
\end{aligned}$$

This is the Azuma-Hoeffding inequality for sums of bounded martingale differences.

bounded differences inequality

If $Z = f(X_1, \ldots, X_n)$ and f is such that

$$|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \leq c_i$$

then the martingale differences are bounded.

bounded differences inequality

If $Z = f(X_1, \ldots, X_n)$ and f is such that

$$|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \le c_i$$

then the martingale differences are bounded.

Bounded differences inequality: if $X_1, \ldots, X_n$ are independent, then

$$\mathbb{P}\{|Z - \mathbb{E}Z| > t\} \le 2e^{-2t^2/\sum_{i=1}^n c_i^2} .$$

bounded differences inequality

If $\mathbf{Z = f(X_1, \ldots, X_n)}$ and $\mathbf{f}$ is such that

$$|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \le c_i$$

then the martingale differences are bounded.

Bounded differences inequality: if $\mathbf{X_1, \ldots, X_n}$ are independent, then

$$\mathbb{P}\{|Z - \mathbb{E}Z| > t\} \le 2e^{-2t^2 / \sum_{i=1}^n c_i^2} .$$

McDiarmid's inequality.

Colin McDiarmid

hoeffding in a hilbert space

Let $\mathbf{X_1, \ldots, X_n}$ be independent zero-mean random variables in a separable Hilbert space such that $\|\mathbf{X_i}\| \le \mathbf{c/2}$ and denote $\mathbf{v = nc^2/4}$. Then, for all $\mathbf{t} \ge \sqrt{\mathbf{v}}$,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^{n} \mathbf{X_i}\right\| > \mathbf{t}\right\} \le \mathbf{e}^{-(\mathbf{t}-\sqrt{\mathbf{v}})^2/(2\mathbf{v})} \ .$$

hoeffding in a hilbert space

Let $X_1, \ldots, X_n$ be independent zero-mean random variables in a separable Hilbert space such that $\|X_i\| \le c/2$ and denote $v = nc^2/4$. Then, for all $t \ge \sqrt{v}$,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^n X_i\right\| > t\right\} \le e^{-(t-\sqrt{v})^2/(2v)} .$$

Proof: By the triangle inequality, $\left\|\sum_{i=1}^n X_i\right\|$ has the bounded differences property with constants c, so

$$\mathbb{P}\left\{\left\|\sum_{i=1}^n X_i\right\| > t\right\} = \mathbb{P}\left\{\left\|\sum_{i=1}^n X_i\right\| - \mathbb{E}\left\|\sum_{i=1}^n X_i\right\| > t - \mathbb{E}\left\|\sum_{i=1}^n X_i\right\|\right\}$$

$$\le \exp\left(-\frac{\left(t - \mathbb{E}\left\|\sum_{i=1}^n X_i\right\|\right)^2}{2v}\right) .$$

Also,

$$\mathbb{E}\left\|\sum_{i=1}^n X_i\right\| \le \sqrt{\mathbb{E}\left\|\sum_{i=1}^n X_i\right\|^2} = \sqrt{\sum_{i=1}^n \mathbb{E}\|X_i\|^2} \le \sqrt{v} .$$

bounded differences inequality

✳Easy to use.

✳Distribution free.

✳Often close to optimal.

✳Does not exploit “variance information.”

✳Often too rigid.

✳Other methods are necessary.

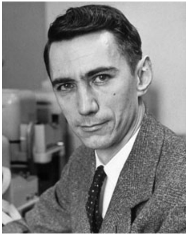
shannon entropy

If $\mathbf{X}, \mathbf{Y}$ are random variables taking values in a set of size $\mathbf{N}$,

$$\mathbf{H(X) = -\sum_{x} p(x) \log p(x)}$$

$$\mathbf{H(X|Y) = H(X, Y) - H(Y)}$$
$$\mathbf{= -\sum_{x,y} p(x, y) \log p(x|y)}$$

$\mathbf{H(X) \leq \log N}$ and $\mathbf{H(X|Y) \leq H(X)}$



Claude Shannon
(1916–2001)

han's inequality

If $\mathbf{X} = (X_1, \ldots, X_n)$ and
$X^{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, then

$$\sum_{i=1}^{n} \left(H(X) - H(X^{(i)}) \right) \le H(X)$$

Te Sun Han

Proof:

$$\begin{aligned} H(X) &= H(X^{(i)}) + H(X_i | X^{(i)}) \\ &\le H(X^{(i)}) + H(X_i | X_1, \ldots, X_{i-1}) \end{aligned}$$

Since $\sum_{i=1}^{n} H(X_i | X_1, \ldots, X_{i-1}) = H(X)$, summing the inequality, we get

$$(n-1) H(X) \le \sum_{i=1}^{n} H(X^{(i)}) .$$

number of increasing subsequences

Let $\mathbf{N}$ be the number of increasing subsequences in a random permutation. Then

$$\mathrm{Var}(\log_2 \mathbf{N}) \leq \mathbb{E} \log_2 \mathbf{N} \ .$$

number of increasing subsequences

Let $\mathbf{N}$ be the number of increasing subsequences in a random permutation. Then

$$\mathrm{Var}(\log_2 N) \le \mathbb{E} \log_2 N \ .$$

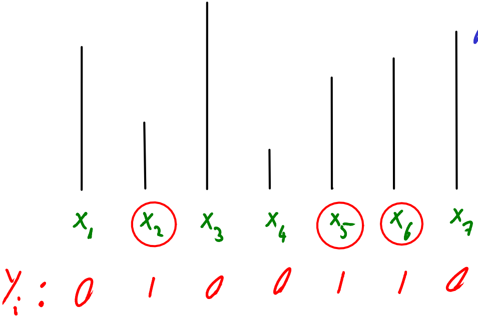
Proof: Let $X = (X_1, \ldots, X_n)$ be i.i.d. uniform $[0, 1]$.
$f_n(X) = \log_2 N$ is now a function of independent random variables. It suffices to prove that $\mathbf{f}$ is self-bounding:

$$0 \le f_n(x) - f_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_n) \le 1$$

and

$$\sum_{i=1}^{n} (f_n(x) - f_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_n)) \le f_n(x) \ .$$

number of increasing subsequences



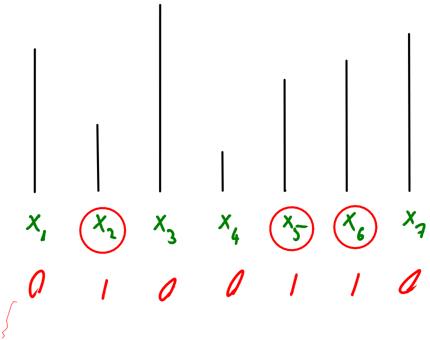
for fixed $x_1, \dots, x_n$, draw an increasing sequence uniformly at random.

$$y_i = \begin{cases} 1 & \text{if } x_i \text{ is in the increasing sequence} \\ 0 & \text{otherwise} \end{cases}$$

number of increasing subsequences

$$H(Y_1,\ldots,Y_n) = f_n(x)$$ (uniform distribution)

$$H(Y^{(i)}) \le f_{n-1}(x^{(i)})$$ (uniform distribution maximizes entropy)



$$\sum_{i=1}^{n}\left(f_n(x) - f_{n-1}(x^{(i)})\right) \le \sum_{i=1}^{n}\left(H(Y) - H(Y^{(i)})\right)$$

$$\le H(Y)$$ (by Han's inequality)

$$= f_n(x).$$

$\Downarrow$

f is <u>self-bounding</u>.

subadditivity of entropy

The entropy of a random variable $\mathbf{Z} \geq \mathbf{0}$ is

$$\mathbf{Ent}(\mathbf{Z}) = \mathbb{E}\mathbf{\Phi}(\mathbf{Z}) - \mathbf{\Phi}(\mathbb{E}\mathbf{Z})$$

where $\mathbf{\Phi}(\mathbf{x}) = \mathbf{x} \log \mathbf{x}$. By Jensen's inequality, $\mathbf{Ent}(\mathbf{Z}) \geq \mathbf{0}$.

subadditivity of entropy

The entropy of a random variable $Z \geq 0$ is

$$\mathrm{Ent}(Z) = \mathbb{E}\Phi(Z) - \Phi(\mathbb{E}Z)$$

where $\Phi(x) = x \log x$. By Jensen's inequality, $\mathrm{Ent}(Z) \geq 0$.

Han's inequality implies the following sub-additivity property.
Let $X_1, \ldots, X_n$ be independent and let $Z = f(X_1, \ldots, X_n)$,
where $f \geq 0$.
Denote

$$\mathrm{Ent}^{(i)}(Z) = \mathbb{E}^{(i)}\Phi(Z) - \Phi(\mathbb{E}^{(i)}Z)$$

Then

$$\mathrm{Ent}(Z) \leq \mathbb{E}\sum_{i=1}^{n} \mathrm{Ent}^{(i)}(Z) .$$

a logarithmic sobolev inequality on the hypercube

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be uniformly distributed over $\{-1, 1\}^n$. If $\mathbf{f} : \{-1, 1\}^n \to \mathbb{R}$ and $\mathbf{Z} = \mathbf{f}(\mathbf{X})$,

$$\mathbf{Ent}(\mathbf{Z}^2) \leq \frac{1}{2}\mathbb{E}\sum_{i=1}^{n}(\mathbf{Z} - \mathbf{Z}_i')^2$$

The proof uses subadditivity of the entropy and calculus for the case $\mathbf{n = 1}$.

Implies Efron-Stein.

Sergei Lvovich Sobolev
(1908–1989)

herbst's argument: exponential concentration

If $\mathbf{f} : \{-1,1\}^n \to \mathbb{R}$, the log-Sobolev inequality may be used with

$$\mathbf{g(x)} = e^{\lambda f(x)/2} \quad \text{where} \quad \lambda \in \mathbb{R} \ .$$

If $F(\lambda) = \mathbb{E}e^{\lambda Z}$ is the moment generating function of $Z = f(X)$,

$$\begin{aligned} \mathrm{Ent}(g(X)^2) &= \lambda \mathbb{E}\left[Ze^{\lambda Z}\right] - \mathbb{E}\left[e^{\lambda Z}\right] \log \mathbb{E}\left[Ze^{\lambda Z}\right] \\ &= \lambda F'(\lambda) - F(\lambda) \log F(\lambda) \ . \end{aligned}$$

Differential inequalities are obtained for $F(\lambda)$.

herbst's argument

As an example, suppose $\mathbf{f}$ is such that $\sum_{i=1}^{n}(Z - Z_i')_+^2 \leq v$. Then by the log-Sobolev inequality,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq \frac{v\lambda^2}{4}F(\lambda)$$

If $G(\lambda) = \log F(\lambda)$, this becomes

$$\left(\frac{G(\lambda)}{\lambda}\right)' \leq \frac{v}{4} \; .$$

This can be integrated: $G(\lambda) \leq \lambda \mathbb{E}Z + \lambda v/4$, so

$$F(\lambda) \leq e^{\lambda \mathbb{E}Z - \lambda^2 v/4}$$

This implies

$$\mathbb{P}\{Z > \mathbb{E}Z + t\} \leq e^{-t^2/v}$$

herbst's argument

As an example, suppose $\mathbf{f}$ is such that $\sum_{i=1}^n (Z - Z_i')_+^2 \leq v$. Then by the log-Sobolev inequality,

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq \frac{v\lambda^2}{4} F(\lambda)$$

If $G(\lambda) = \log F(\lambda)$, this becomes

$$\left(\frac{G(\lambda)}{\lambda}\right)' \leq \frac{v}{4} \; .$$

This can be integrated: $G(\lambda) \leq \lambda \mathbb{E}Z + \lambda v/4$, so

$$F(\lambda) \leq e^{\lambda \mathbb{E}Z - \lambda^2 v/4}$$

This implies

$$\mathbb{P}\{Z > \mathbb{E}Z + t\} \leq e^{-t^2/v}$$

Stronger than the bounded differences inequality!

gaussian log-sobolev inequality

Let $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_n})$ be a vector of i.i.d. standard normal If $\mathbf{f} : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}$ and $\mathbf{Z} = \mathbf{f}(\mathbf{X})$,

$$\mathbf{Ent}(\mathbf{Z}^2) \leq 2\mathbb{E}\left[\|\nabla \mathbf{f}(\mathbf{X})\|^2\right]$$

(Gross, 1975).

gaussian log-sobolev inequality

Let $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_n})$ be a vector of i.i.d. standard normal If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{Z} = \mathbf{f}(\mathbf{X})$,

$$\mathbf{Ent}(\mathbf{Z}^2) \leq 2\mathbb{E}\left[\|\nabla \mathbf{f}(\mathbf{X})\|^2\right]$$

(Gross, 1975).
Proof sketch: By the subadditivity of entropy, it suffices to prove it for $\mathbf{n} = \mathbf{1}$.
Approximate $\mathbf{Z} = \mathbf{f}(\mathbf{X})$ by

$$\mathbf{f}\left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \varepsilon_i\right)$$

where the ε_i are i.i.d. Rademacher random variables.
Use the log-Sobolev inequality of the hypercube and the central limit theorem.

gaussian concentration inequality

Herbst't argument may now be repeated:
Suppose $\mathbf{f}$ is Lipschitz: for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathbf{n}}$,

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq \mathbf{L}\|\mathbf{x} - \mathbf{y}\| \ .$$

Then, for all $\mathbf{t} > \mathbf{0}$,

$$\mathbb{P}\left\{\mathbf{f}(\mathbf{X}) - \mathbb{E}\mathbf{f}(\mathbf{X}) \geq \mathbf{t}\right\} \leq \mathbf{e}^{-\mathbf{t}^2/(2\mathbf{L}^2)} \ .$$

(Tsirelson, Ibragimov, and Sudakov, 1976).

an application: supremum of a gaussian process

Let $(X_t)_{t\in\mathcal{T}}$ be an almost surely continuous centered Gaussian process. Let $Z = \sup_{t\in\mathcal{T}} X_t$. If

$$\sigma^2 = \sup_{t\in\mathcal{T}} \left(\mathbb{E}\left[X_t^2\right]\right) ,$$

then

$$\mathbb{P}\left\{|Z - \mathbb{E}Z| \geq u\right\} \leq 2e^{-u^2/(2\sigma^2)}$$

an application: supremum of a gaussian process

Let $(X_t)_{t\in\mathcal{T}}$ be an almost surely continuous centered Gaussian process. Let $Z = \sup_{t\in\mathcal{T}} X_t$. If

$$\sigma^2 = \sup_{t\in\mathcal{T}} \left(\mathbb{E}\left[X_t^2\right]\right) \ ,$$

then

$$\mathbb{P}\left\{|Z - \mathbb{E}Z| \geq u\right\} \leq 2e^{-u^2/(2\sigma^2)}$$

Proof: We may assume $\mathcal{T} = \{1, ..., n\}$. Let Γ be the covariance matrix of $X = (X_1, \ldots, X_n)$. Let $A = \Gamma^{1/2}$. If Y is a standard normal vector, then

$$f(Y) = \max_{i=1,...,n} (AY)_i \stackrel{\text{distr.}}{=} \max_{i=1,...,n} X_i$$

By Cauchy-Schwarz,

$$|(Au)_i - (Av)_i| = \left|\sum_j A_{i,j}(u_j - v_j)\right| \leq \left(\sum_j A_{i,j}^2\right)^{1/2} \|u - v\|$$

$$\leq \sigma\|u - v\|$$

beyond bernoulli and gaussian: the entropy method

For general distributions, logarithmic Sobolev inequalities are not available.

Solution: modified logarithmic Sobolev inequalities.
Suppose $\mathbf{X_1, \ldots, X_n}$ are independent. Let $\mathbf{Z = f(X_1, \ldots, X_n)}$ and $\mathbf{Z_i = f_i(X^{(i)}) = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)}$.

Let $\phi(\mathbf{x}) = \mathbf{e^x - x - 1}$. Then for all $\lambda \in \mathbb{R}$,

$$\lambda \mathbb{E}\left[\mathbf{Z}e^{\lambda \mathbf{Z}}\right] - \mathbb{E}\left[e^{\lambda \mathbf{Z}}\right] \log \mathbb{E}\left[e^{\lambda \mathbf{Z}}\right] \leq \sum_{i=1}^{n} \mathbb{E}\left[e^{\lambda \mathbf{Z}} \phi\left(-\lambda(\mathbf{Z} - \mathbf{Z_i})\right)\right].$$

Michel Ledoux

the entropy method

Define $Z_i = \inf_{x_i'} f(X_1, \ldots, x_i', \ldots, X_n)$ and suppose

$$\sum_{i=1}^{n} (Z - Z_i)^2 \leq v \ .$$

Then for all $t > 0$,

$$\mathbb{P}\{Z - \mathbb{E}Z > t\} \leq e^{-t^2/(2v)} \ .$$

the entropy method

Define $Z_i = \inf_{x_i'} f(X_1, \ldots, x_i', \ldots, X_n)$ and suppose

$$\sum_{i=1}^{n} (Z - Z_i)^2 \leq v \ .$$

Then for all $t > 0$,

$$\mathbb{P}\{Z - \mathbb{E}Z > t\} \leq e^{-t^2/(2v)} \ .$$

This implies the bounded differences inequality and much more.

example: the largest eigenvalue of a symmetric matrix

Let $\mathbf{A} = (\mathbf{X}_{i,j})_{n\times n}$ be symmetric, the $\mathbf{X}_{i,j}$ independent ($\mathbf{i} \le \mathbf{j}$) with $|\mathbf{X}_{i,j}| \le 1$. Let

$$Z = \lambda_1 = \sup_{u:\|u\|=1} u^T A u \ .$$

and suppose $\mathbf{v}$ is such that $\mathbf{Z} = \mathbf{v}^T \mathbf{A} \mathbf{v}$.
$\mathbf{A}'_{i,j}$ is obtained by replacing $\mathbf{X}_{i,j}$ by $\mathbf{x}'_{i,j}$. Then

$$\begin{aligned}(Z - Z_{i,j})_+ &\le \left(v^T A v - v^T A'_{i,j} v\right) \mathbb{1}_{Z > Z_{i,j}} \\ &= \left(v^T (A - A'_{i,j}) v\right) \mathbb{1}_{Z > Z_{i,j}} \le 2 \left(v_i v_j (X_{i,j} - X'_{i,j})\right)_+ \\ &\le 4 |v_i v_j| \ .\end{aligned}$$

Therefore,

$$\sum_{1\le i\le j\le n} (Z - Z'_{i,j})_+^2 \le \sum_{1\le i\le j\le n} 16 |v_i v_j|^2 \le 16 \left(\sum_{i=1}^n v_i^2\right)^2 = 16 \ .$$

self-bounding functions

Suppose $\mathbf{Z}$ satisfies

$$0 \leq Z - Z_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} (Z - Z_i) \leq Z \; .$$

Recall that $\mathbf{Var}(Z) \leq \mathbb{E}Z$. We have much more:

$$\mathbb{P}\{Z > \mathbb{E}Z + t\} \leq e^{-t^2/(2\mathbb{E}Z + 2t/3)}$$

and

$$\mathbb{P}\{Z < \mathbb{E}Z - t\} \leq e^{-t^2/(2\mathbb{E}Z)}$$

self-bounding functions

Suppose $\mathbf{Z}$ satisfies

$$0 \leq Z - Z_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} (Z - Z_i) \leq Z .$$

Recall that $\mathbf{Var}(Z) \leq \mathbb{E}Z$. We have much more:

$$\mathbb{P}\{Z > \mathbb{E}Z + t\} \leq e^{-t^2/(2\mathbb{E}Z + 2t/3)}$$

and

$$\mathbb{P}\{Z < \mathbb{E}Z - t\} \leq e^{-t^2/(2\mathbb{E}Z)}$$

Rademacher averages, random VC dimension, random VC entropy, longest increasing subsequence in a random permutation, are all examples of self bounding functions.

self-bounding functions

Suppose $\mathbf{Z}$ satisfies

$$0 \le \mathbf{Z} - \mathbf{Z_i} \le 1 \quad \text{and} \quad \sum_{i=1}^{n} (\mathbf{Z} - \mathbf{Z_i}) \le \mathbf{Z} \ .$$

Recall that $\mathbf{Var(Z)} \le \mathbb{E}\mathbf{Z}$. We have much more:

$$\mathbb{P}\{\mathbf{Z} > \mathbb{E}\mathbf{Z} + \mathbf{t}\} \le \mathbf{e}^{-\mathbf{t}^2/(2\mathbb{E}\mathbf{Z} + 2\mathbf{t}/3)}$$

and

$$\mathbb{P}\{\mathbf{Z} < \mathbb{E}\mathbf{Z} - \mathbf{t}\} \le \mathbf{e}^{-\mathbf{t}^2/(2\mathbb{E}\mathbf{Z})}$$

Rademacher averages, random VC dimension, random VC entropy, longest increasing subsequence in a random permutation, are all examples of self bounding functions.

Configuration functions.

exponential efron-stein inequality

Define

$$V^+ = \sum_{i=1}^n \mathbb{E}' \left[(Z - Z_i')_+^2 \right]$$

and

$$V^- = \sum_{i=1}^n \mathbb{E}' \left[(Z - Z_i')_-^2 \right] \ .$$

By Efron-Stein,

$$\mathbf{Var}(Z) \le \mathbb{E}V^+ \quad \text{and} \quad \mathbf{Var}(Z) \le \mathbb{E}V^- \ .$$

exponential efron-stein inequality

Define

$$V^+ = \sum_{i=1}^{n} \mathbb{E}' \left[(Z - Z_i')_+^2 \right]$$

and

$$V^- = \sum_{i=1}^{n} \mathbb{E}' \left[(Z - Z_i')_-^2 \right] .$$

By Efron-Stein,

$$\mathbf{Var}(Z) \le \mathbb{E}V^+ \quad \text{and} \quad \mathbf{Var}(Z) \le \mathbb{E}V^- .$$

The following exponential versions hold for all $\lambda, \theta > 0$ with $\lambda\theta < 1$:

$$\log \mathbb{E}e^{\lambda(Z - \mathbb{E}Z)} \le \frac{\lambda\theta}{1 - \lambda\theta} \log \mathbb{E}e^{\lambda V^+/\theta} .$$

If also $Z_i' - Z \le 1$ for every i, then for all $\lambda \in (0, 1/2)$,

$$\log \mathbb{E}e^{\lambda(Z - \mathbb{E}Z)} \le \frac{2\lambda}{1 - 2\lambda} \log \mathbb{E}e^{\lambda V^-} .$$

weakly self-bounding functions

$f : \mathcal{X}^n \to [0, \infty)$ is weakly (a, b)-self-bounding if there exist $f_i : \mathcal{X}^{n-1} \to [0, \infty)$ such that for all $x \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right)^2 \leq af(x) + b \, .$$

weakly self-bounding functions

$f : \mathcal{X}^n \to [0, \infty)$ is weakly (a, b)-self-bounding if there exist $f_i : \mathcal{X}^{n-1} \to [0, \infty)$ such that for all $x \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right)^2 \leq af(x) + b \,.$$

Then

$$\mathbb{P}\{Z \geq \mathbb{E}Z + t\} \leq \exp\left(-\frac{t^2}{2\,(a\mathbb{E}Z + b + at/2)} \right) \,.$$

weakly self-bounding functions

$f : \mathcal{X}^n \to [0,\infty)$ is weakly (a,b)-self-bounding if there exist $f_i : \mathcal{X}^{n-1} \to [0,\infty)$ such that for all $x \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \left(f(x) - f_i(x^{(i)}) \right)^2 \le af(x) + b \,.$$

Then

$$\mathbb{P}\left\{Z \ge \mathbb{E}Z + t\right\} \le \exp\left(-\frac{t^2}{2\left(a\mathbb{E}Z + b + at/2\right)} \right) \;.$$

If, in addition, $f(x) - f_i(x^{(i)}) \le 1$, then for $0 < t \le \mathbb{E}Z$,

$$\mathbb{P}\left\{Z \le \mathbb{E}Z - t\right\} \le \exp\left(-\frac{t^2}{2\left(a\mathbb{E}Z + b + c_- t\right)} \right) \;.$$

where $c = (3a-1)/6$.

the isoperimetric view

Let $\mathbf{X} = (X_1, \ldots, X_n)$ have independent components, taking values in $\mathcal{X}^n$. Let $A \subset \mathcal{X}^n$.

The Hamming distance of X to A is

$$d(X, A) = \min_{y \in A} d(X, y) = \min_{y \in A} \sum_{i=1}^{n} \mathbb{1}_{X_i \neq y_i} .$$

Michel Talagrand

the isoperimetric view

Let $\mathbf{X} = (X_1, \ldots, X_n)$ have independent components, taking values in $\mathcal{X}^n$. Let $A \subset \mathcal{X}^n$.
The Hamming distance of $\mathbf{X}$ to A is

$$d(X, A) = \min_{y \in A} d(X, y) = \min_{y \in A} \sum_{i=1}^{n} \mathbb{1}_{X_i \neq y_i} \, .$$

Michel Talagrand

$$\mathbb{P}\left\{ d(X, A) \geq t + \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}[A]}} \right\} \leq e^{-2t^2/n} \, .$$

the isoperimetric view

Let $\mathbf{X} = (X_1, \ldots, X_n)$ have independent components, taking values in $\mathcal{X}^n$. Let $A \subset \mathcal{X}^n$.

The Hamming distance of X to A is

$$d(X, A) = \min_{y \in A} d(X, y) = \min_{y \in A} \sum_{i=1}^{n} \mathbb{1}_{X_i \neq y_i} .$$

Michel Talagrand

$$\mathbb{P}\left\{ d(X, A) \geq t + \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}[A]}} \right\} \leq e^{-2t^2/n} .$$

Concentration of measure!

the isoperimetric view

Proof: By the bounded differences inequality,

$$\mathbb{P}\{\mathbb{E}\mathbf{d}(\mathbf{X},\mathbf{A}) - \mathbf{d}(\mathbf{X},\mathbf{A}) \geq \mathbf{t}\} \leq \mathbf{e}^{-2\mathbf{t}^2/\mathbf{n}}.$$

Taking $\mathbf{t} = \mathbb{E}\mathbf{d}(\mathbf{X},\mathbf{A})$, we get

$$\mathbb{E}\mathbf{d}(\mathbf{X},\mathbf{A}) \leq \sqrt{\frac{\mathbf{n}}{2}\log\frac{1}{\mathbb{P}\{\mathbf{A}\}}}.$$

By the bounded differences inequality again,

$$\mathbb{P}\left\{\mathbf{d}(\mathbf{X},\mathbf{A}) \geq \mathbf{t} + \sqrt{\frac{\mathbf{n}}{2}\log\frac{1}{\mathbb{P}\{\mathbf{A}\}}}\right\} \leq \mathbf{e}^{-2\mathbf{t}^2/\mathbf{n}}$$

talagrand's convex distance

The weighted Hamming distance is

$$\mathbf{d}_\alpha(\mathbf{x}, \mathbf{A}) = \inf_{\mathbf{y} \in \mathbf{A}} \mathbf{d}_\alpha(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{y} \in \mathbf{A}} \sum_{\mathbf{i}:\mathbf{x}_\mathbf{i} \neq \mathbf{y}_\mathbf{i}} |\alpha_\mathbf{i}|$$

where $\alpha = (\alpha_1, \ldots, \alpha_\mathbf{n})$. The same argument as before gives

$$\mathbb{P}\left\{\mathbf{d}_\alpha(\mathbf{X}, \mathbf{A}) \geq \mathbf{t} + \sqrt{\frac{\|\alpha\|^2}{2} \log \frac{1}{\mathbb{P}\{\mathbf{A}\}}}\right\} \leq \mathbf{e}^{-2\mathbf{t}^2/\|\alpha\|^2} ,$$

This implies

$$\sup_{\alpha:\|\alpha\|=1} \min\left(\mathbb{P}\{\mathbf{A}\}, \mathbb{P}\left\{\mathbf{d}_\alpha(\mathbf{X}, \mathbf{A}) \geq \mathbf{t}\right\}\right) \leq \mathbf{e}^{-\mathbf{t}^2/2} .$$

convex distance inequality

convex distance:

$$\mathbf{d_T(x, A) = \sup_{\alpha \in [0,\infty)^n : \|\alpha\|=1} d_\alpha(x, A)} \ .$$

convex distance inequality

convex distance:

$$d_T(x, A) = \sup_{\alpha \in [0,\infty)^n : \|\alpha\|=1} d_\alpha(x, A) \ .$$

Talagrand's convex distance inequality:

$$\mathbb{P}\{A\}\mathbb{P}\{d_T(X, A) \geq t\} \leq e^{-t^2/4} \ .$$

convex distance inequality

convex distance:

$$\mathbf{d_T(x, A)} = \sup_{\alpha \in [0,\infty)^n : \|\alpha\| = 1} \mathbf{d_\alpha(x, A)} \ .$$

Talagrand's convex distance inequality:

$$\mathbb{P}\{\mathbf{A}\}\mathbb{P}\left\{\mathbf{d_T(X, A)} \geq \mathbf{t}\right\} \leq \mathbf{e^{-t^2/4}} \ .$$

Follows from the fact that $\mathbf{d_T(X, A)^2}$ is $\mathbf{(4, 0)}$ weakly self bounding (by a saddle point representation of $\mathbf{d_T}$).

Talagrand's original proof was different.

convex lipschitz functions

For $\mathbf{A} \subset [0,1]^n$ and $\mathbf{x} \in [0,1]^n$, define

$$\mathbf{D(x, A) = \inf_{y \in A} \|x - y\|} \ .$$

If $\mathbf{A}$ is convex, then

$$\mathbf{D(x, A) \leq d_T(x, A)} \ .$$

convex lipschitz functions

For $\mathbf{A} \subset [0,1]^n$ and $\mathbf{x} \in [0,1]^n$, define

$$\mathbf{D}(\mathbf{x}, \mathbf{A}) = \inf_{\mathbf{y} \in \mathbf{A}} \|\mathbf{x} - \mathbf{y}\| .$$

If $\mathbf{A}$ is convex, then

$$\mathbf{D}(\mathbf{x}, \mathbf{A}) \leq \mathbf{d_T}(\mathbf{x}, \mathbf{A}) .$$

Proof:

$$\begin{aligned}
\mathbf{D}(\mathbf{x}, \mathbf{A}) &= \inf_{\nu \in \mathcal{M}(\mathbf{A})} \|\mathbf{x} - \mathbb{E}_\nu \mathbf{Y}\| \quad \text{(since } \mathbf{A} \text{ is convex)} \\
&\leq \inf_{\nu \in \mathcal{M}(\mathbf{A})} \sqrt{\sum_{j=1}^{n} \left(\mathbb{E}_\nu \mathbb{1}_{x_j \neq Y_j}\right)^2} \quad \text{(since } x_j, Y_j \in [0,1]\text{)} \\
&= \inf_{\nu \in \mathcal{M}(\mathbf{A})} \sup_{\alpha : \|\alpha\| \leq 1} \sum_{j=1}^{n} \alpha_j \mathbb{E}_\nu \mathbb{1}_{x_j \neq Y_j} \quad \text{(by Cauchy-Schwarz)} \\
&= \mathbf{d_T}(\mathbf{x}, \mathbf{A}) \quad \text{(by minimax theorem)} .
\end{aligned}$$

convex lipschitz functions

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ have independent components taking values in $[0, 1]$. Let $\mathbf{f} : [0, 1]^n \to \mathbb{R}$ be quasi-convex such that $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq \|\mathbf{x} - \mathbf{y}\|$. Then

$$\mathbb{P}\{\mathbf{f}(\mathbf{X}) > \mathbb{M}\mathbf{f}(\mathbf{X}) + \mathbf{t}\} \leq 2e^{-t^2/4}$$

and

$$\mathbb{P}\{\mathbf{f}(\mathbf{X}) < \mathbb{M}\mathbf{f}(\mathbf{X}) - \mathbf{t}\} \leq 2e^{-t^2/4} .$$

convex lipschitz functions

Let $\mathbf{X} = (X_1, \ldots, X_n)$ have independent components taking values in $[0,1]$. Let $f : [0,1]^n \to \mathbb{R}$ be quasi-convex such that $|f(x) - f(y)| \le \|x - y\|$. Then

$$\mathbb{P}\{f(X) > \mathbb{M}f(X) + t\} \le 2e^{-t^2/4}$$

and

$$\mathbb{P}\{f(X) < \mathbb{M}f(X) - t\} \le 2e^{-t^2/4} .$$

Proof: Let $A_s = \{x : f(x) \le s\} \subset [0,1]^n$. A_s is convex. Since f is Lipschitz,

$$f(x) \le s + D(x, A_s) \le s + d_T(x, A_s) ,$$

By the convex distance inequality,

$$\mathbb{P}\{f(X) \ge s + t\}\mathbb{P}\{f(X) \le s\} \le e^{-t^2/4} .$$

Take $s = \mathbb{M}f(X)$ for the upper tail and $s = \mathbb{M}f(X) - t$ for the lower tail.

ϕ entropies

For a convex function ϕ on $[0, \infty)$, the ϕ-entropy of $Z \geq 0$ is

$$H_\phi(Z) = \mathbb{E}[\phi(Z)] - \phi(\mathbb{E}[Z]) \ .$$

H_ϕ is subadditive:

$$H_\phi(Z) \leq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[\phi(Z) \mid X^{(i)}\right] - \phi\left(\mathbb{E}\left[Z \mid X^{(i)}\right]\right)\right]$$

if (and only if) ϕ is twice differentiable on $(0, \infty)$, and either ϕ is affine or strictly positive and $1/\phi''$ is concave.

ϕ entropies

For a convex function ϕ on $[0, \infty)$, the ϕ-entropy of $Z \geq 0$ is

$$H_\phi(Z) = \mathbb{E}[\phi(Z)] - \phi(\mathbb{E}[Z]) \; .$$

H_ϕ is subadditive:

$$H_\phi(Z) \leq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left[\phi(Z) \mid X^{(i)}\right] - \phi\left(\mathbb{E}\left[Z \mid X^{(i)}\right]\right)\right]$$

if (and only if) ϕ is twice differentiable on $(0, \infty)$, and either ϕ is affine or strictly positive and $1/\phi''$ is concave.

$\phi(x) = x^2$ corresponds to Efron-Stein.

$x \log x$ is subadditivity of entropy.

We may consider $\phi(x) = x^p$ for $p \in (1, 2]$.

generalized efron-stein

Define

$$Z_i' = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n) ,$$

$$V^+ = \sum_{i=1}^{n} (Z - Z_i')_+^2 .$$

generalized efron-stein

Define

$$Z_i' = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n) \ ,$$

$$V^+ = \sum_{i=1}^n (Z - Z_i')_+^2 \ .$$

For $q \geq 2$ and $q/2 \leq \alpha \leq q - 1$,

$$\mathbb{E}\left[(Z - \mathbb{E}Z)_+^q\right]$$
$$\leq \mathbb{E}\left[(Z - \mathbb{E}Z)_+^\alpha\right]^{q/\alpha} + \alpha\,(q - \alpha)\,\mathbb{E}\left[V^+\,(Z - \mathbb{E}Z)_+^{q-2}\right] \ ,$$

and similarly for $\mathbb{E}\left[(Z - \mathbb{E}Z)_-^q\right]$.

moment inequalities

We may solve the recursions, for $\mathbf{q} \geq \mathbf{2}$.

moment inequalities

We may solve the recursions, for $\mathbf{q \geq 2}$.

If $\mathbf{V^+ \leq c}$ for some constant $\mathbf{c \geq 0}$, then for all integers $\mathbf{q \geq 2}$,

$$\left(\mathbb{E}\left[(\mathbf{Z} - \mathbb{E}\mathbf{Z})_+^{\mathbf{q}}\right]\right)^{\mathbf{1/q}} \leq \sqrt{\mathbf{Kqc}} \ ,$$

where $\mathbf{K = 1/\left(e - \sqrt{e}\right) < 0.935}$.

moment inequalities

We may solve the recursions, for $\mathbf{q \geq 2}$.

If $\mathbf{V^+ \leq c}$ for some constant $\mathbf{c \geq 0}$, then for all integers $\mathbf{q \geq 2}$,

$$\left(\mathbb{E}\left[(\mathbf{Z} - \mathbb{E}\mathbf{Z})_+^{\mathbf{q}}\right]\right)^{\mathbf{1/q}} \leq \sqrt{\mathbf{Kqc}}\ ,$$

where $\mathbf{K = 1/\left(e - \sqrt{e}\right) < 0.935}$.

More generally,

$$\left(\mathbb{E}\left[(\mathbf{Z} - \mathbb{E}\mathbf{Z})_+^{\mathbf{q}}\right]\right)^{\mathbf{1/q}} \leq \mathbf{1.6}\sqrt{\mathbf{q}}\left(\mathbb{E}\left[\mathbf{V^{+q/2}}\right]\right)^{\mathbf{1/q}}\ .$$

sums: khinchine's inequality

Let $X_1, \ldots, X_n$ be independent Rademacher variables and $Z = \sum_{i=1}^{n} a_i X_i$. For any integer $q \geq 2$,

$$\left(\mathbb{E}\left[Z_+^q\right]\right)^{1/q} \leq \sqrt{2Kq}\sqrt{\sum_{i=1}^{n} a_i^2}$$

sums: khinchine's inequality

Let $X_1, \ldots, X_n$ be independent Rademacher variables and $Z = \sum_{i=1}^n a_i X_i$. For any integer $q \geq 2$,

$$\left(\mathbb{E}\left[Z_+^q\right]\right)^{1/q} \leq \sqrt{2Kq}\sqrt{\sum_{i=1}^n a_i^2}$$

Proof:

$$V^+ = \sum_{i=1}^n \mathbb{E}\left[(a_i(X_i - X_i'))_+^2 \mid X_i\right] = 2\sum_{i=1}^n a_i^2 \mathbb{1}_{a_i X_i > 0} \leq 2\sum_{i=1}^n a_i^2 \ ,$$

Aleksandr Khinchin
(1894–1959)

sums: rosenthal's inequality

Let $\mathbf{X_1, \ldots, X_n}$ be independent real-valued random variables with $\mathbb{E}\mathbf{X_i = 0}$. Define

$$\mathbf{Z} = \sum_{i=1}^{n} \mathbf{X_i}\,, \quad \sigma^2 = \sum_{i=1}^{n} \mathbb{E}\mathbf{X_i^2}\,, \quad \mathbf{Y} = \max_{i=1,\ldots,n} |\mathbf{X_i}|\,.$$

Then for any integer $\mathbf{q \geq 2}$,

$$\left(\mathbb{E}\left[\mathbf{Z_+^q}\right]\right)^{1/q} \leq \sigma\sqrt{10q} + 3q\left(\mathbb{E}\left[\mathbf{Y_+^q}\right]\right)^{1/q}\,.$$

books

M. Ledoux. The concentration of measure phenomenon. American Mathematical Society, 2001.

D. Dubhashi and A. Panconesi. Concentration of measure for the analysis of randomized algorithms. Cambridge University Press, 2009.

S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities: a nonasymptotic theory of independence. Oxford University Press, 2013.