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

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

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, Gácis, and Körner, 1976; Marton 1986, 1996, 1997; Dembo 1997);

- **Talagrand’s induction method**, 1996;

chernoff bounds

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

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

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

Serguei Bernstein (1880-1968)

Herman Chernoff (1923–)
chernoff bounds

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

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

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

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 $\lambda > 0$,

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

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

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

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

We obtain

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

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 $X_i \leq 1$. Let

$$v = \sum_{i=1}^{n} \mathbb{E} \left[ X_i^2 \right].$$

Then

$$\mathbb{P} \left\{ \sum_{i=1}^{n} (X_i - \mathbb{E}X_i) \geq t \right\} \leq \exp \left( - \frac{t^2}{2(v + t/3)} \right).$$
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 \( E_i[\cdot] = E[\cdot | X_1, \ldots, X_i] \). Thus, \( E_0Z = EZ \) and \( E_nZ = Z \).
**martingale representation**

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

\[ Z = f(X_1, \ldots, X_n). \]

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 $E_i[\cdot] = E[\cdot | X_1, \ldots, X_i]$. Thus, $E_0Z = EZ$ and $E_nZ = Z$.

Writing

$$\Delta_i = E_iZ - E_{i-1}Z,$$

we have

$$Z - EZ = \sum_{i=1}^{n} \Delta_i.$$

This is the Doob martingale representation of $Z$. 

Joseph Leo Doob (1910–2004)
martingale representation: the variance

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

Now if \( j > i \), \( E_i \Delta_j = 0 \), so

\[ E_i \Delta_j \Delta_i = \Delta_i E_i \Delta_j = 0 , \]

We obtain

\[ \text{Var}(Z) = E \left[ \left( \sum_{i=1}^{n} \Delta_i \right)^2 \right] = \sum_{i=1}^{n} E \left[ \Delta_i^2 \right] . \]
martingale representation: the variance

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

Now if \(j > i\), \(E_i \Delta_j = 0\), so

\[
E_i \Delta_j \Delta_i = \Delta_i E_i \Delta_j = 0 ,
\]

We obtain

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

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 $f : \mathcal{X}^n \rightarrow \mathbb{R}$ and $Z = f(X_1, \ldots, X_n)$. Then

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

where $\mathbb{E}^{(i)}Z$ is expectation with respect to the $i$-th variable $X_i$ only.
efron-stein inequality (1981)

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

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

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

We obtain more useful forms by using that

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

for any constant $a$. 

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

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

$Z$ is concentrated if it doesn’t depend too much on any of its variables.
If $X'_1, \ldots, X'_n$ 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

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

$Z$ 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 $f_i$

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

then

$$\text{Var}(Z) \leq E \left[ \sum_{i=1}^{n} (Z - Z_i)^2 \right]$$
efron, stein, and steele

Bradley Efron

Charles Stein

Mike Steele
weakly self-bounding functions

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

\[ \sum_{i=1}^{n} \left( f(x) - f_i(x^{(i)}) \right)^2 \leq af(x) + b. \]
weakly self-bounding functions

\( f : \mathcal{X}^n \rightarrow [0, \infty) \) is weakly \((a, b)\)-self-bounding if there exist \( f_i : \mathcal{X}^{n-1} \rightarrow [0, \infty) \) 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

\[
\text{Var}(f(X)) \leq a\mathbb{E}f(X) + b .
\]
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 \( \text{Var}(f(X)) \leq \mathbb{E}f(X) \).
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 \( \text{Var}(f(X)) \leq \mathbb{E}f(X) \).

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

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

and

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

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

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

Configuration functions.
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

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

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

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

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

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

\[
\text{Var}(Z) \leq \frac{1}{2n}
\]

regardless of the distribution and the richness of \( \mathcal{A} \).
\(X_1, \ldots, X_n\) 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:

\[
Z - \mathbb{E}Z = \sum_{i=1}^{n} \Delta_i \quad \text{where} \quad \Delta_i = \mathbb{E}_iZ - \mathbb{E}_{i-1}Z,
\]

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

To get exponential inequalities, we bound the moment generating function \(\mathbb{E}e^{\lambda(Z - \mathbb{E}Z)}\).
Suppose that the martingale differences are bounded: $|\Delta_i| \leq c_i$. Then

$$
\mathbb{E}e^{\lambda(Z - \mathbb{E}Z)} = \mathbb{E}e^{\lambda(\sum_{i=1}^{n} \Delta_i)} = \mathbb{E}\mathbb{E}_ne^{\lambda(\sum_{i=1}^{n-1} \Delta_i) + \lambda \Delta_n}
$$

$$
= \mathbb{E}e^{\lambda(\sum_{i=1}^{n-1} \Delta_i)} \mathbb{E}_ne^{\lambda \Delta_n}
$$

$$
\leq \mathbb{E}e^{\lambda(\sum_{i=1}^{n-1} \Delta_i)} e^{\lambda^2 c_n^2/2} \quad \text{(by Hoeffding)}
$$

\ldots

$$
\leq e^{\lambda^2(\sum_{i=1}^{n} c_i^2)/2}.
$$

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

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

then the martingale differences are bounded.

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

$$\mathbb{P}\{|Z - \mathbb{E}Z| > t\} \leq 2e^{-2t^2 / \sum_{i=1}^n c_i^2}.$$
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 \( X_1, \ldots, X_n \) are independent, then

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

McDiarmid’s inequality.

Colin McDiarmid
hoeffding in a hilbert space

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

$$
P \left\{ \left\| \sum_{i=1}^{n} X_i \right\| > t \right\} \leq e^{-\frac{(t-\sqrt{v})^2}{2v}}.
$$
Let $X_1, \ldots, X_n$ be independent zero-mean random variables in a separable Hilbert space such that $\|X_i\| \leq c/2$ and denote $v = nc^2/4$. Then, for all $t \geq \sqrt{v}$,

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

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

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

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

Also,

$$
E \left\| \sum_{i=1}^{n} X_i \right\| \leq \sqrt{E \left\| \sum_{i=1}^{n} X_i \right\|^2} = \sqrt{\sum_{i=1}^{n} E \|X_i\|^2} \leq \sqrt{v}.
$$
bounded differences inequality

- Easy to use.
- Distribution free.
- Often close to optimal.
- Does not exploit “variance information.”
- Often too rigid.
- Other methods are necessary.
If $X, Y$ are random variables taking values in a set of size $N$,}

$$H(X) = - \sum_x p(x) \log p(x)$$

$$H(X|Y) = H(X, Y) - H(Y)$$

$$= - \sum_{x,y} p(x, y) \log p(x|y)$$

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

Claude Shannon (1916–2001)
han’s inequality

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

\[
\sum_{i=1}^{n} \left( H(X) - H(X^{(i)}) \right) \leq H(X)
\]

Proof:

\[
H(X) = H(X^{(i)}) + H(X_i | X^{(i)}) \\
\leq H(X^{(i)}) + H(X_i | X_1, \ldots, X_{i-1})
\]

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

\[
(n - 1)H(X) \leq \sum_{i=1}^{n} H(X^{(i)})
\]
number of increasing subsequences

Let $N$ be the number of increasing subsequences in a random permutation. Then

$$\text{Var}(\log_2 N) \leq \mathbb{E} \log_2 N.$$
number of increasing subsequences

Let $N$ be the number of increasing subsequences in a random permutation. Then

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

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 $f$ is self-bounding:

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

and

$$\sum_{i=1}^{n} (f_n(x) - f_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)) \leq f_n(x).$$
number of increasing subsequences

For fixed $x_1, \ldots, x_n$, draw an increasing sequence uniformly at random.

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

$\gamma_i : 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0$
number of increasing subsequences

\[ H(\gamma_1, \ldots, \gamma_n) = x^n(x) \]

\[ H(\gamma_1^{(i)}) \leq x_{n-1}(x^{(i)}) \]

(uniform distribution)

(uniform distribution maximizes entropy)

\[
\sum_{i=1}^{n} (x(x) - x_{n-1}(x^{(i)})) \leq \sum_{i=1}^{n} (H(Y) - H(Y^{(i)}))
\]

\leq H(Y) \quad \text{(by Han's inequality)}

= x^n(x).

\[ f \text{ is self-bounding.} \]
The entropy of a random variable $Z \geq 0$ is

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

where $\Phi(x) = x \log x$. By Jensen’s inequality, $\text{Ent}(Z) \geq 0$. 
The entropy of a random variable $Z \geq 0$ is

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

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

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

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

Then

$$
\text{Ent}(Z) \leq \mathbb{E} \sum_{i=1}^{n} \text{Ent}^{(i)}(Z).
$$
a logarithmic sobolev inequality on the hypercube

Let $X = (X_1, \ldots, X_n)$ be uniformly distributed over $\{-1, 1\}^n$. If $f : \{-1, 1\}^n \to \mathbb{R}$ and $Z = f(X)$,

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

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

Implies Efron-Stein.
Sergei Lvovich Sobolev
(1908–1989)
herbst’s argument: exponential concentration

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

\[
g(x) = e^{\frac{\lambda f(x)}{2}} \quad \text{where} \quad \lambda \in \mathbb{R}.
\]

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

\[
\text{Ent}(g(X)^2) = \lambda \mathbb{E} \left[ Ze^{\lambda Z} \right] - \mathbb{E} \left[ e^{\lambda Z} \right] \log \mathbb{E} \left[ Ze^{\lambda Z} \right] \\
= \lambda F'(\lambda) - F(\lambda) \log F(\lambda).
\]

Differential inequalities are obtained for \( F(\lambda) \).
As an example, suppose \( f \) 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 \( f \) 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!
Let $X = (X_1, \ldots, X_n)$ be a vector of i.i.d. standard normal if $f : \mathbb{R}^n \to \mathbb{R}$ and $Z = f(X)$,

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

(Gross, 1975).
gaussian log-sobolev inequality

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

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

(Gross, 1975).

Proof sketch: By the subadditivity of entropy, it suffices to prove it for $n = 1$.

Approximate $Z = f(X)$ by

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

where the $\varepsilon_i$ are i.i.d. Rademacher random variables.

Use the log-Sobolev inequality of the hypercube and the central limit theorem.
gaussian concentration inequality

Herbst’s argument may now be repeated: Suppose $f$ is Lipschitz: for all $x, y \in \mathbb{R}^n$,

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

Then, for all $t > 0$,

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

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

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

then

\[
P \{ |Z - \mathbb{E}Z| \geq u \} \leq 2e^{-u^2/(2\sigma^2)}
\]

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

\[
f(Y) = \max_{i=1, \ldots, n} (AY)_i \overset{\text{distr.}}{=} \max_{i=1, \ldots, n} X_i
\]

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 $X_1, \ldots, X_n$ are independent. Let $Z = f(X_1, \ldots, X_n)$ and $Z_i = f_i(X^{(i)}) = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

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

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

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 $t > 0$,

$$P \{ Z - \mathbb{E}Z > t \} \leq e^{-t^2/(2v)}.$$
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 $t > 0$,

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

This implies the bounded differences inequality and much more.
example: the largest eigenvalue of a symmetric matrix

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

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

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

$$(Z - Z_{i,j})_+ \leq \left(v^TAv - v^TA'_{i,j}v\right) \mathbb{1}_{Z > Z_{i,j}}$$

$$= \left(v^T(A - A'_{i,j})v\right) \mathbb{1}_{Z > Z_{i,j}} \leq 2 \left(v_iv_j(X_{i,j} - X'_{i,j})\right)_+$$

$$\leq 4|v_iv_j|.$$ 

Therefore,

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

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

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

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

and

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

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

Recall that $\text{Var}(Z) \leq E_Z$. We have much more:

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

and

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

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

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

Recall that $\text{Var}(Z) \leq \mathbb{E}Z$. 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.

Configuration functions.
exponential efron-stein inequality

Define

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

and

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

By Efron-Stein,

\[ \text{Var}(Z) \leq \mathbb{E}V^+ \quad \text{and} \quad \text{Var}(Z) \leq \mathbb{E}V^- . \]
Define
\[ V^+ = \sum_{i=1}^{n} E' \left[ (Z - Z_i')^2_+ \right] \]
and
\[ V^- = \sum_{i=1}^{n} E' \left[ (Z - Z_i')^2_- \right] . \]

By Efron-Stein,
\[ \text{Var}(Z) \leq E V^+ \quad \text{and} \quad \text{Var}(Z) \leq E V^- . \]

The following exponential versions hold for all \( \lambda, \theta > 0 \) with \( \lambda \theta < 1 \):
\[ \log E e^{\lambda(Z - E Z)} \leq \frac{\lambda \theta}{1 - \lambda \theta} \log E e^{\lambda V^+ / \theta} . \]

If also \( Z_i' - Z \leq 1 \) for every \( i \), then for all \( \lambda \in (0, 1/2) \),
\[ \log E e^{\lambda(Z - E Z)} \leq \frac{2\lambda}{1 - 2\lambda} \log E e^{\lambda V^-} . \]
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

\( f : \mathcal{X}^n \rightarrow [0, \infty) \) is weakly \((a, b)\)-self-bounding if there exist \( f_i : \mathcal{X}^{n-1} \rightarrow [0, \infty) \) 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

\[
P \{ Z \geq \mathbb{E}Z + t \} \leq \exp \left( -\frac{t^2}{2 \left( a\mathbb{E}Z + b + at/2 \right)} \right).
\]
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 \( x \in \mathcal{X}^n \),

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

Then

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

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

\[
P \{Z \leq \mathbb{E}Z - t\} \leq \exp \left( -\frac{t^2}{2 (a\mathbb{E}Z + b + c - t)} \right).
\]

where \( c = (3a - 1)/6 \).
Let \( X = (X_1, \ldots, X_n) \) have independent components, taking values in \( \mathcal{X}^n \). Let \( A \subset \mathcal{X}^n \).

The Hamming distance of \( X \) to \( A \) is

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

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

The **Hamming distance** of $X$ to $A$ is

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

**Concentration of measure!**

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

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

The Hamming distance of $X$ to $A$ is

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

Concentration of measure!

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

**Proof:** By the bounded differences inequality,

\[ P\{E_{d(X, A)} - d(X, A) \geq t\} \leq e^{-2t^2/n}. \]

Taking \( t = E_{d(X, A)} \), we get

\[ E_{d(X, A)} \leq \sqrt{n \log \frac{1}{P\{A\}}}. \]

By the bounded differences inequality again,

\[ P\left\{ d(X, A) \geq t + \sqrt{\frac{n}{2} \log \frac{1}{P\{A\}}} \right\} \leq e^{-2t^2/n} \]
The weighted Hamming distance is

\[ d_\alpha(x, A) = \inf_{y \in A} d_\alpha(x, y) = \inf_{y \in A} \sum_{i : x_i \neq y_i} |\alpha_i| \]

where \( \alpha = (\alpha_1, \ldots, \alpha_n) \). The same argument as before gives

\[
P \left\{ d_\alpha(X, A) \geq t + \sqrt{\frac{\|\alpha\|^2}{2} \log \frac{1}{P\{A\}}} \right\} \leq e^{-2t^2/\|\alpha\|^2},
\]

This implies

\[
\sup_{\alpha : \|\alpha\| = 1} \min (P\{A\}, P \{d_\alpha(X, A) \geq t\}) \leq e^{-t^2/2}.
\]
convex distance inequality

convex distance:

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

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

Talagrand’s convex distance inequality:

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

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

Talagrand’s original proof was different.
convex lipschitz functions

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

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

If $A$ is convex, then

$$D(x, A) \leq d_T(x, A) .$$
convex lipschitz functions

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

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

If $A$ is convex, then

$$D(x, A) \leq d_T(x, A).$$

Proof:

$$D(x, A) = \inf_{\nu \in \mathcal{M}(A)} \|x - E_\nu Y\| \quad \text{(since $A$ is convex)}$$

$$\leq \inf_{\nu \in \mathcal{M}(A)} \sqrt{\sum_{j=1}^n (E_\nu 1_{x_j \neq Y_j})^2} \quad \text{(since $x_j, Y_j \in [0, 1]$)}$$

$$= \inf_{\nu \in \mathcal{M}(A)} \sup_{\alpha: \|\alpha\| \leq 1} \sum_{j=1}^n \alpha_j E_\nu 1_{x_j \neq Y_j} \quad \text{(by Cauchy-Schwarz)}$$

$$= d_T(x, A) \quad \text{(by minimax theorem)}.$$
convex lipschitz functions

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

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

and

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

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

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

and

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

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

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

By the convex distance inequality,

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

Take $s = \mathbb{M}f(X)$ for the upper tail and $s = \mathbb{M}f(X) - t$ for the lower tail.
For a convex function $\phi$ on $[0, \infty)$, the $\phi$-entropy of $Z \geq 0$ is

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

$H_{\phi}$ is subadditive:

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

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

$\phi$ entropies
**φ** entropies

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

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

\( H_{\phi} \) 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) \( \phi \) is twice differentiable on \((0, \infty)\), and either \( \phi \) 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

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

\[ \nu^+ = \sum_{i=1}^{n} (Z - Z'_i)^2. \]
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 \( q/2 \leq \alpha \leq q - 1 \),
\[ \mathbb{E} [(Z - \mathbb{E}Z)^q] \]
\[ \leq \mathbb{E} [(Z - \mathbb{E}Z)^\alpha]^{q/\alpha} + \alpha (q - \alpha) \mathbb{E} [V^+ (Z - \mathbb{E}Z)^{q-2}] , \]
and similarly for \( \mathbb{E} [(Z - \mathbb{E}Z)^q] \).
moment inequalities

We may solve the recursions, for \( q \geq 2 \).
moment inequalities

We may solve the recursions, for \( q \geq 2 \).

If \( V^+ \leq c \) for some constant \( c \geq 0 \), then for all integers \( q \geq 2 \),

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

where \( K = 1 / (e - \sqrt{e}) < 0.935 \).
We may solve the recursions, for \( q \geq 2 \).

If \( V^+ \leq c \) for some constant \( c \geq 0 \), then for all integers \( q \geq 2 \),

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

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

More generally,

\[
(\mathbb{E} [(Z - \mathbb{E}Z)_+^q])^{1/q} \leq 1.6 \sqrt{q} \left( \mathbb{E} [V^{+q/2}] \right)^{1/q} .
\]
Let $X_1, \ldots, X_n$ be independent Rademacher variables and $Z = \sum_{i=1}^{n} a_i X_i$. For any integer $q \geq 2$,

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

$$V^+ = \sum_{i=1}^{n} \mathbb{E} \left[ \left( a_i (X_i - X_i') \right)_+^2 \mid X_i \right] = 2 \sum_{i=1}^{n} a_i^2 \mathbb{1}_{a_i X_i > 0} \leq 2 \sum_{i=1}^{n} a_i^2,$$
Let \( X_1, \ldots, X_n \) be independent real-valued random variables with \( \mathbb{E}X_i = 0 \). Define

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

Then for any integer \( q \geq 2 \),

\[
(\mathbb{E} [Z^q])^{1/q} \leq \sigma \sqrt{10q + 3q (\mathbb{E} [Y^q])^{1/q}}.
\]
