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 \(X\). Let \(f : X^n \to \mathbb{R}\) and

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

How large are “typical” deviations of \(Z\) from \(\mathbb{E}Z\)?
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;

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–)
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 $\mathbb{E}e^{\lambda X_i}$.
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 $\mathbb{E}e^{\lambda X_i}$.

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)
martingale representation

\(X_1, \ldots, X_n\) are independent random variables taking values in some set \(X\). Let \(f : X^n \rightarrow \mathbb{R}\) 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 \(E_i[\cdot] = E[\cdot|X_1, \ldots, X_i]\). 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 \(Z\).
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 \( Z \).
martingale representation: the variance

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

Now if \( j > i \), \( \mathbb{E} \Delta_j = 0 \), so

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

We obtain

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

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

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

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

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

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

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

We obtain more useful forms by using that

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

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.
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

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

$Z$ 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!
If for some arbitrary functions $f_i$

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

then

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

Bradley Efron

Charles Stein

Mike Steele
example: kernel density estimation

Let $X_1, \ldots, X_n$ be i.i.d. real samples drawn according to some density $\phi$. The kernel density estimate is

$$
\phi_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),
$$

where $h > 0$, and $K$ is a nonnegative “kernel” $\int K = 1$. The $L_1$ error is

$$
Z = f(X_1, \ldots, X_n) = \int |\phi(x) - \phi_n(x)| \, dx.
$$
example: kernel density estimation

Let $X_1, \ldots, X_n$ be i.i.d. real samples drawn according to some density $\phi$. The kernel density estimate is

$$
\phi_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right),
$$

where $h > 0$, and $K$ is a nonnegative “kernel” $\int K = 1$. The $L_1$ error is

$$
Z = f(X_1, \ldots, X_n) = \int |\phi(x) - \phi_n(x)| \, dx.
$$

It is easy to see that

$$
|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \leq \frac{1}{nh} \int \left| K \left( \frac{x - x_i}{h} \right) - K \left( \frac{x - x_i'}{h} \right) \right| \, dx \leq \frac{2}{n},
$$

so we get $\text{Var}(Z) \leq \frac{2}{n}$.

(Devroye, 1991.)
Luc Devroye on Taksim square (2013).
weakly self-bounding functions

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

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

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

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

Then

\[ \text{Var}(f(X)) \leq a\mathbb{E}f(X) + b. \]
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 \( \text{Var}(f(X)) \leq \mathbb{E}f(X) \).
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.
If
\[ 0 \leq f(x) - f_i(x^{(i)}) \leq 1 \]
and
\[ \sum_{i=1}^{n} (f(x) - f_i(x^{(i)})) \leq f(x) , \]
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.
$X_1, \ldots, X_n$ 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:

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

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

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

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

beyond the variance
azuma’s inequality

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

$$
E e^{\lambda (Z - EZ)} = E e^{\lambda \left( \sum_{i=1}^{n} \Delta_i \right)} = E E_n e^{\lambda \left( \sum_{i=1}^{n-1} \Delta_i \right) + \lambda \Delta_n}
$$

$$
= E e^{\lambda \left( \sum_{i=1}^{n-1} \Delta_i \right)} E_n e^{\lambda \Delta_n}
$$

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

$$
\cdots
$$

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

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

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

McDiarmid's inequality. Colin McDiarmid
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

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

McDiarmid’s inequality.
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 $X, Y$ are random variables taking values in a set of size $N$,

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

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

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

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

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

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

$$(n - 1)H(X) \leq \sum_{i=1}^{n} H(X^{(i)})$$
Let $X_1, \ldots, X_n$ be independent points in the plane (of arbitrary distribution!). Let $N$ be the number of subsets of points that are in convex position. Then

$$\text{Var}(\log_2 N) \leq \mathbb{E} \log_2 N.$$
By Efron-Stein, it suffices to prove that $f$ is self-bounding:

\[ 0 \leq f_n(x) - f_{n-1}(x^{(i)}) \leq 1 \]

and

\[ \sum_{i=1}^{n} \left( f_n(x) - f_{n-1}(x^{(i)}) \right) \leq f_n(x) . \]
proof

By Efron-Stein, it suffices to prove that $f$ is self-bounding:

$$0 \leq f_n(x) - f_{n-1}(x^{(i)}) \leq 1$$

and

$$\sum_{i=1}^{n} \left( f_n(x) - f_{n-1}(x^{(i)}) \right) \leq f_n(x).$$

The first property is obvious, only need to prove the second.
By Efron-Stein, it suffices to prove that $f$ is self-bounding:

$$0 \leq f_n(x) - f_{n-1}(x^{(i)}) \leq 1$$

and

$$\sum_{i=1}^{n} \left( f_n(x) - f_{n-1}(x^{(i)}) \right) \leq f_n(x).$$

The first property is obvious, only need to prove the second.

This is a deterministic property so fix the points.
Among all sets in convex position, draw one uniformly at random. Define $Y_i$ as the indicator that $x_i$ is in the chosen set.
proof

Among all sets in convex position, draw one uniformly at random. Define $Y_i$ as the indicator that $x_i$ is in the chosen set.

$$H(Y) = H(Y_1, \ldots, Y_n) = \log_2 N = f_n(x)$$

Also,

$$H(Y^{(i)}) \leq f_{n-1}(x^{(i)})$$
proof

Among all sets in convex position, draw one uniformly at random. Define $Y_i$ as the indicator that $x_i$ is in the chosen set.

$$H(Y) = H(Y_1, \ldots, Y_n) = \log_2 N = f_n(x)$$

Also,

$$H(Y^{(i)}) \leq f_{n-1}(x^{(i)})$$

so by Han’s inequality,

$$\sum_{i=1}^{n} \left( f_n(x) - f_{n-1}(x^{(i)}) \right) \leq \sum_{i=1}^{n} \left( H(Y) - H(Y^{(i)}) \right) \leq H(Y) = f_n(x)$$
Han’s inequality for relative entropies

Let \( P = P_1 \otimes \cdots \otimes P_n \) be a product distribution and \( Q \) an arbitrary distribution on \( \mathcal{X}^n \).
Han’s inequality for relative entropies

Let $\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n$ be a product distribution and $\mathbf{Q}$ an arbitrary distribution on $\mathcal{X}^n$.

By Han’s inequality,

$$D(\mathbf{Q} \parallel \mathbf{P}) \geq \frac{1}{n-1} \sum_{i=1}^{n} D(\mathbf{Q}^{(i)} \parallel \mathbf{P}^{(i)})$$
subadditivity of entropy

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, $\text{Ent}(Z) \geq 0$. 

Let $X_1, \ldots, X_n$ be independent and let $Z = f(X_1, \ldots, X_n)$, where $f \geq 0$. $\text{Ent}(Z)$ is the relative entropy between the distribution induced by $Z$ on $X_n$ and the distribution of $X = (X_1, \ldots, X_n)$. 

Denote $\text{Ent}^{(i)}(Z) = \mathbb{E}^{(i)}\Phi(Z) - \Phi(\mathbb{E}^{(i)}Z)$ then by Han’s inequality, $\text{Ent}(Z) \leq \sum_{i=1}^{n} \text{Ent}^{(i)}(Z)$. 
subadditivity of entropy

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, $\text{Ent}(Z) \geq 0$.

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 by Han’s inequality, $\text{Ent}(Z) \leq \mathbb{E} \sum_{i=1}^{n} \text{Ent}^{(i)}(Z)$.
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, $\text{Ent}(Z) \geq 0$.

Let $X_1, \ldots, X_n$ be independent and let $Z = f(X_1, \ldots, X_n)$, where $f \geq 0$.

$\text{Ent}(Z)$ is the relative entropy between the distribution induced by $Z$ on $\mathcal{X}^n$ and the distribution of $X = (X_1, \ldots, X_n)$. 

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

Then by Han’s inequality,$$\text{Ent}(Z) \leq \mathbb{E}n\sum_{i=1}^{n}\text{Ent}^{(i)}(Z)$$
subadditivity of entropy

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, $\text{Ent}(Z) \geq 0$.

Let $X_1, \ldots, X_n$ be independent and let $Z = f(X_1, \ldots, X_n)$, where $f \geq 0$.

$\text{Ent}(Z)$ is the relative entropy between the distribution induced by $Z$ on $\mathcal{X}^n$ and the distribution of $X = (X_1, \ldots, X_n)$.

Denote

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

Then by Han’s inequality,

$$\text{Ent}(Z) \leq \mathbb{E} \sum_{i=1}^{n} \text{Ent}^{(i)}(Z).$$
Let $X = (X_1, \ldots, X_n)$ be uniformly distributed over $\{-1, 1\}^n$. If $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ 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)
If $f : \{-1, 1\}^n \to \mathbb{R}$, the log-Sobolev inequality may be used with

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

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

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

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

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

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

Stronger than the bounded differences inequality!
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).
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.
gau
csian concentration inequality

Herbst’
t 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 \( t > 0 \),

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

(Tsirelson, Ibragimov, and Sudakov, 1976).
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 $\lambda \in \mathbb{R}$,

$$\lambda \mathbb{E} \left[ Z e^{\lambda Z} \right] - \mathbb{E} \left[ e^{\lambda Z} \right] \log \mathbb{E} \left[ e^{\lambda Z} \right] \leq \sum_{i=1}^{n} \mathbb{E} \left[ e^{\lambda Z} \phi (-\lambda (Z - Z_i)) \right].$$
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 $t > 0$,

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

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

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

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

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

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

Therefore,

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

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

\[ \text{Var}(Z) \leq \mathbb{E} V^+ \quad \text{and} \quad \text{Var}(Z) \leq \mathbb{E} V^- . \]
exponential efron-stein inequality

Define

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

and

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

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

Let $f : \mathcal{X}^n \to [0, \infty)$ be 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 \( x \in \mathcal{X}^n \),

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

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

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

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

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

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

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

\[ \mathbb{E} d(X, A) \leq \sqrt{\frac{n}{2} \log \frac{1}{\Pr \{ A \}}}. \]

By the bounded differences inequality again,

\[ \Pr \left\{ d(X, A) \geq t + \sqrt{\frac{n}{2} \log \frac{1}{\Pr \{ A \}}} \right\} \leq e^{-2t^2/n} \]
talagrand’s convex distance

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

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

This implies

\[ \sup_{\alpha: \|\alpha\| = 1} \min (\mathbb{P}\{A\}, \mathbb{P}\{d_\alpha(X, A) \geq t\}) \leq e^{-t^2/2} . \]
Talagrand's convex distance inequality:

\[ P\{A\} \leq e^{-t^2/4}. \]

Follows from the fact that \( d_T(x, A) \) is \((4, 0)\) weakly self-bounding (by a saddle point representation of \( d_T \)). Talagrand's original proof was different. It can also be recovered from Marton's transportation inequality.
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.

It can also be recovered from Marton’s transportation inequality.
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:

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

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

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

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

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

and

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

Let $X = (X_1, \ldots, X_n)$ have independent components taking values in $[0, 1]$. Let $f : [0, 1]^n \to \mathbb{R}$ 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

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

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

$$H_\Phi(Z) = \mathbb{E} \Phi(Z) - \Phi(\mathbb{E}Z).$$
\( \Phi \) 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:

\[
H_\Phi(Z) \leq \sum_{i=1}^{n} \mathbb{E}H_\Phi^{(i)}(Z)
\]

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 \( p \in (1, 2] \).
**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:

$$H_\Phi(Z) \leq \sum_{i=1}^{n} \mathbb{E} H_\Phi^{(i)}(Z)$$

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

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

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

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

More generally, 

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

where $K = \frac{1}{e - \sqrt{e}} < 0.935$.

Additionally, 

$$\left( \mathbb{E} \left[ (Z - \mathbb{E} Z)^q \right] \right)^{1/q} \leq 1.6 \sqrt{q \left( \mathbb{E} \left[ V + \frac{q}{2} \right] \right)^{1/q}}.$$
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 $K = 1/(e - \sqrt{e}) < 0.935$. 

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 $q \geq 2$,

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

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

More generally,

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