Concentration inequalities

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.

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

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

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 $t > 0$.

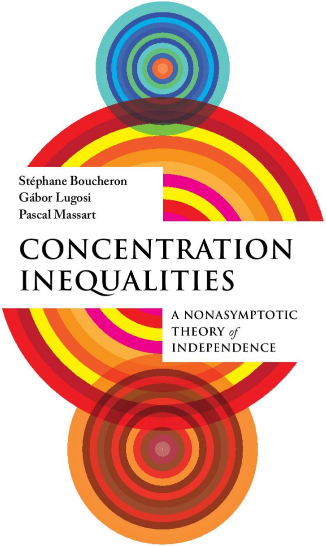
various approaches

- 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

markov's inequality

If $\mathbf{Z} \geq \mathbf{0}$, then

$$\mathbb{P}\{\mathbf{Z} > \mathbf{t}\} \leq \frac{\mathbb{E}\mathbf{Z}}{\mathbf{t}} \; .$$

markov's inequality

If $\mathbf{Z} \geq \mathbf{0}$, then

$$\mathbb{P}\{\mathbf{Z} > \mathbf{t}\} \leq \frac{\mathbb{E}\mathbf{Z}}{\mathbf{t}} \, .$$

This implies Chebyshev's inequality: if $\mathbf{Z}$ has a finite variance $\mathbf{Var}(\mathbf{Z}) = \mathbb{E}(\mathbf{Z} - \mathbb{E}\mathbf{Z})^2$, then

$$\mathbb{P}\{|\mathbf{Z} - \mathbb{E}\mathbf{Z}| > \mathbf{t}\} = \mathbb{P}\{(\mathbf{Z} - \mathbb{E}\mathbf{Z})^2 > \mathbf{t}^2\} \leq \frac{\mathbf{Var}(\mathbf{Z})}{\mathbf{t}^2} \, .$$

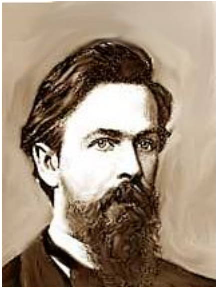
markov's inequality

If $\mathbf{Z} \geq \mathbf{0}$, then

$$\mathbb{P}\{\mathbf{Z} > \mathbf{t}\} \leq \frac{\mathbb{E}\mathbf{Z}}{\mathbf{t}} .$$

This implies Chebyshev's inequality: if $\mathbf{Z}$ has a finite variance $\mathbf{Var(Z)} = \mathbb{E}(\mathbf{Z} - \mathbb{E}\mathbf{Z})^2$, then

$$\mathbb{P}\{|\mathbf{Z} - \mathbb{E}\mathbf{Z}| > \mathbf{t}\} = \mathbb{P}\{(\mathbf{Z} - \mathbb{E}\mathbf{Z})^2 > \mathbf{t}^2\} \leq \frac{\mathbf{Var(Z)}}{\mathbf{t}^2} .$$



Andrey Markov (1856–1922)

sums of independent random variables

Let $X_1, \ldots, X_n$ be independent real-valued and let $Z = \sum_{i=1}^n X_i$.
By independence, $\mathrm{Var}(Z) = \sum_{i=1}^n \mathrm{Var}(X_i)$. If they are identically distributed, $\mathrm{Var}(Z) = n\mathrm{Var}(X_1)$, so

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i - n\mathbb{E}X_1\right| > t\right\} \le \frac{n\mathrm{Var}(X_1)}{t^2} \ .$$

Equivalently,

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i - n\mathbb{E}X_1\right| > t\sqrt{n}\right\} \le \frac{\mathrm{Var}(X_1)}{t^2} \ .$$

Typical deviations are at most of the order $\sqrt{n}$.

sums of independent random variables

Let $X_1, \ldots, X_n$ be independent real-valued and let $Z = \sum_{i=1}^n X_i$.
By independence, $\mathrm{Var}(Z) = \sum_{i=1}^n \mathrm{Var}(X_i)$. If they are identically distributed, $\mathrm{Var}(Z) = n\mathrm{Var}(X_1)$, so

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i - n\mathbb{E}X_1\right| > t\right\} \le \frac{n\mathrm{Var}(X_1)}{t^2} .$$

Equivalently,

$$\mathbb{P}\left\{\left|\sum_{i=1}^n X_i - n\mathbb{E}X_1\right| > t\sqrt{n}\right\} \le \frac{\mathrm{Var}(X_1)}{t^2} .$$

Typical deviations are at most of the order $\sqrt{n}$.

Pafnuty Chebyshev (1821–1894)

chernoff bounds

By the central limit theorem,

$$\lim_{n\to\infty} \mathbb{P}\left\{\sum_{i=1}^{n} X_i - n\mathbb{E}X_1 > t\sqrt{n}\right\} = 1 - \Psi(t/\sqrt{\mathrm{Var}(X_1)})$$

$$\leq e^{-t^2/(2\mathrm{Var}(X_1))}$$

so we expect an exponential decrease in $t^2/\mathrm{Var}(X_1)$.

chernoff bounds

By the central limit theorem,

$$\lim_{n\to\infty} \mathbb{P}\left\{\sum_{i=1}^{n} X_i - n\mathbb{E}X_1 > t\sqrt{n}\right\} = 1 - \Psi(t/\sqrt{\mathrm{Var}(X_1)})$$
$$\leq e^{-t^2/(2\mathrm{Var}(X_1))}$$

so we expect an exponential decrease in $t^2/\mathrm{Var}(X_1)$.
Trick: use Markov's inequality in a more clever way: if $\lambda > 0$,

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

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

chernoff bounds

If $\mathbf{Z} = \sum_{i=1}^{n} \mathbf{X}_i$ 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. Now it suffices to find bounds for $\mathbb{E}e^{\lambda X_i}$.

chernoff bounds

If $\mathbf{Z} = \sum_{i=1}^{n} \mathbf{X_i}$ 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. Now 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 $\mathbf{X_1, \ldots, X_n} \in [0,1]$, 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

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

a maximal inequality

Suppose $\mathbf{Y_1, \ldots, Y_N}$ are sub-Gaussian in the sense that

$$\mathbb{E} e^{\lambda Y_i} \le e^{\lambda^2 \sigma^2 / 2} \ .$$

Then

$$\mathbb{E} \max_{i=1,\ldots,N} Y_i \le \sigma \sqrt{2 \log N} \ .$$

a maximal inequality

Suppose $\mathbf{Y_1, \ldots, Y_N}$ are sub-Gaussian in the sense that

$$\mathbb{E}\mathbf{e}^{\lambda \mathbf{Y_i}} \le \mathbf{e}^{\lambda^2 \sigma^2/2} .$$

Then

$$\mathbb{E} \max_{\mathbf{i=1,\ldots,N}} \mathbf{Y_i} \le \sigma \sqrt{\mathbf{2 \log N}} .$$

Proof:

$$\mathbf{e}^{\lambda \mathbb{E} \max_{\mathbf{i=1,\ldots,N}} \mathbf{Y_i}} \le \mathbb{E}\mathbf{e}^{\lambda \max_{\mathbf{i=1,\ldots,N}} \mathbf{Y_i}} \le \sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbb{E}\mathbf{e}^{\lambda \mathbf{Y_i}} \le \mathbf{N}\mathbf{e}^{\lambda^2\sigma^2/2}$$

Take logarithms, and optimize in λ.

an application

Let $A_1, \ldots, A_N \subset \mathcal{X}$ and let $X_1, \ldots, X_n$ be i.i.d. random points in $\mathcal{X}$. Let

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

By Hoeffding's inequality, for each A,

$$\begin{aligned}\mathbb{E}e^{\lambda(P(A)-P_n(A))} &= \mathbb{E}e^{(\lambda/n)\sum_{i=1}^{n}(P(A)-\mathbb{1}_{X_i\in A})} \\ &= \prod_{i=1}^{n}\mathbb{E}e^{(\lambda/n)(P(A)-\mathbb{1}_{X_i\in A})} \le e^{\lambda^2/(8n)} \,.\end{aligned}$$

By the maximal inequality,

$$\mathbb{E}\max_{j=1,\ldots,N}(P(A_j) - P_n(A_j)) \le \sqrt{\frac{\log N}{2n}} \,.$$

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 $\mathbb{E}_n Z = Z$.

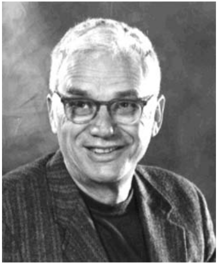
Writing

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

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

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

Now if $\mathbf{j} > \mathbf{i}$, $\mathbb{E}_i\boldsymbol{\Delta}_j = \mathbf{0}$, so

$$\mathbb{E}_i\boldsymbol{\Delta}_j\boldsymbol{\Delta}_i = \boldsymbol{\Delta}_i\mathbb{E}_i\boldsymbol{\Delta}_j = \mathbf{0}\ ,$$

We obtain

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

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 $X_1, \ldots, X_n$ 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.

We obtain more useful forms by using that

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

for any constant 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

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

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

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

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

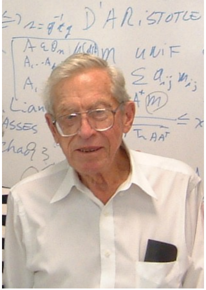
$$Z_i = f_i(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) ,$$

then

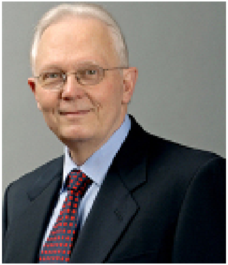
$$\mathbf{Var}(Z) \le \mathbb{E}\left[\sum_{i=1}^{n}(Z - Z_i)^2\right]$$

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 ϕ. 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 ϕ. 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

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

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

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

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

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

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

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

bounding the expectation

Let $\mathbf{P}_n'(A) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i' \in A}$ and let $\mathbb{E}'$ denote expectation only with respect to $X_1', \ldots, X_n'$.

$$\mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| = \mathbb{E} \sup_{A \in \mathcal{A}} |\mathbb{E}'[P_n(A) - P_n'(A)]|$$

$$\leq \mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P_n'(A)| = \frac{1}{n}\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_{i=1}^n (\mathbb{1}_{X_i \in A} - \mathbb{1}_{X_i' \in A}) \right|$$

bounding the expectation

Let $P'_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X'_i \in A}$ and let $\mathbb{E}'$ denote expectation only with respect to $X'_1, \ldots, X'_n$.

$$\mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| = \mathbb{E} \sup_{A \in \mathcal{A}} |\mathbb{E}'[P_n(A) - P'_n(A)]|$$

$$\leq \mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P'_n(A)| = \frac{1}{n}\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_{i=1}^{n} (\mathbb{1}_{X_i \in A} - \mathbb{1}_{X'_i \in A}) \right|$$

Second symmetrization: if $\varepsilon_1, \ldots, \varepsilon_n$ are independent Rademacher variables, then

$$= \frac{1}{n}\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_{i=1}^{n} \varepsilon_i (\mathbb{1}_{X_i \in A} - \mathbb{1}_{X'_i \in A}) \right| \leq \frac{2}{n}\mathbb{E} \sup_{A \in \mathcal{A}} \left| \sum_{i=1}^{n} \varepsilon_i \mathbb{1}_{X_i \in A} \right|$$

conditional rademacher average

If

$$R_n = \mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_{i=1}^{n} \varepsilon_i \mathbb{1}_{X_i \in A} \right|$$

then

$$\mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \le \frac{2}{n} \mathbb{E} R_n .$$

conditional rademacher average

If

$$\mathbf{R_n} = \mathbb{E}_{\varepsilon} \sup_{A \in \mathcal{A}} \left| \sum_{i=1}^{n} \varepsilon_i \mathbb{1}_{X_i \in A} \right|$$

then

$$\mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \leq \frac{2}{n} \mathbb{E} R_n \ .$$

$\mathbf{R_n}$ is a data-dependent quantity!

concentration of conditional rademacher average

Define

$$\mathbf{R}_n^{(i)} = \mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_{j \neq i} \varepsilon_j \mathbb{1}_{X_j \in A} \right|$$

One can show easily that

$$0 \leq \mathbf{R}_n - \mathbf{R}_n^{(i)} \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} (\mathbf{R}_n - \mathbf{R}_n^{(i)}) \leq \mathbf{R}_n \ .$$

By the Efron-Stein inequality,

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

concentration of conditional rademacher average

Define

$$\mathbf{R_n^{(i)}} = \mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_{j \neq i} \varepsilon_j \mathbb{1}_{X_j \in A} \right|$$

One can show easily that

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

By the Efron-Stein inequality,

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

Standard deviation is at most $\sqrt{\mathbb{E}\mathbf{R_n}}$!

concentration of conditional rademacher average

Define

$$\mathbf{R_n^{(i)}} = \mathbb{E}_\varepsilon \sup_{A \in \mathcal{A}} \left| \sum_{j \neq i} \varepsilon_j \mathbb{1}_{X_j \in A} \right|$$

One can show easily that

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

By the Efron-Stein inequality,

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

Standard deviation is at most $\sqrt{\mathbb{E} R_n}$!

Such functions are called self-bounding.

bounding the conditional rademacher average

If $S(X_1^n, \mathcal{A})$ is the number of different sets of form

$$\{X_1, \ldots, X_n\} \cap A : A \in \mathcal{A}$$

then R_n is the maximum of $S(X_1^n, \mathcal{A})$ sub-Gaussian random variables. By the maximal inequality,

$$\frac{1}{2} R_n \leq \sqrt{\frac{\log S(X_1^n, \mathcal{A})}{2n}} \; .$$

bounding the conditional rademacher average

If $\mathbf{S(X_1^n, \mathcal{A})}$ is the number of different sets of form

$$\{X_1, \ldots, X_n\} \cap A : A \in \mathcal{A}$$

then $\mathbf{R_n}$ is the maximum of $\mathbf{S(X_1^n, \mathcal{A})}$ sub-Gaussian random variables. By the maximal inequality,

$$\frac{1}{2} R_n \leq \sqrt{\frac{\log S(X_1^n, \mathcal{A})}{2n}} \ .$$

In particular,

$$\mathbb{E} \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| \leq 2\mathbb{E}\sqrt{\frac{\log S(X_1^n, \mathcal{A})}{2n}} \ .$$

random VC dimension

Let $\mathbf{V} = \mathbf{V}(x_1^n, \mathcal{A})$ be the size of the largest subset of $\{x_1, \ldots, x_n\}$ shattered by $\mathcal{A}$.
By Sauer's lemma,

$$\log S(X_1^n, \mathcal{A}) \leq V(X_1^n, \mathcal{A}) \log(n + 1) \ .$$

random VC dimension

Let $\mathbf{V} = \mathbf{V}(\mathbf{x}_1^n, \mathcal{A})$ be the size of the largest subset of $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ shattered by $\mathcal{A}$.
By Sauer's lemma,

$$\log \mathbf{S}(\mathbf{X}_1^n, \mathcal{A}) \leq \mathbf{V}(\mathbf{X}_1^n, \mathcal{A}) \log(\mathbf{n}+1) \ .$$

$\mathbf{V}$ is also self-bounding:

$$\sum_{i=1}^{n} (\mathbf{V} - \mathbf{V}^{(i)})^2 \leq \mathbf{V}$$

so by Efron-Stein,

$$\mathbf{Var}(\mathbf{V}) \leq \mathbb{E}\mathbf{V}$$

vapnik and chervonenkis

Vladimir Vapnik

Alexey Chervonenkis

beyond the variance

$X_1, \dots, X_n$ are independent random variables taking values in some set $\mathcal{X}$. Let $f : \mathcal{X}^n \to \mathbb{R}$ and $Z = f(X_1, \dots, X_n)$. 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 $\mathbb{E}_i[\cdot] = \mathbb{E}[\cdot | X_1, \dots, X_i]$.

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 $\mathbf{Z} = \mathbf{f}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\mathbf{f}$ is such that

$$|\mathbf{f}(\mathbf{x}_1, \ldots, \mathbf{x}_n) - \mathbf{f}(\mathbf{x}_1, \ldots, \mathbf{x}_i', \ldots, \mathbf{x}_n)| \leq \mathbf{c}_i$$

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 $\mathbf{Z} = \mathbf{f}(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ 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 $\mathbf{X}_1, \ldots, \mathbf{X}_n$ 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 $\mathbf{X_1, \dots, X_n}$ 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 (e.g., $\mathbf{L_1}$ error of kernel density estimate).

✳Does not exploit “variance information.”

✳Often too rigid.

✳Other methods are necessary.

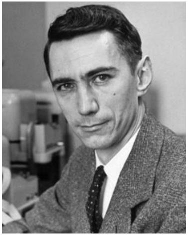
shannon entropy

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

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

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

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



Claude Shannon
(1916–2001)

han's inequality

Te Sun Han

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

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

Proof:

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

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

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

edge isoperimetric inequality on the hypercube

Let $\mathbf{A} \subset \{-1,1\}^n$. Let $\mathbf{E(A)}$ be the collection of pairs $\mathbf{x}, \mathbf{x}' \in \mathbf{A}$ such that $\mathbf{d_H(x, x') = 1}$. Then

$$|\mathbf{E(A)}| \leq \frac{|\mathbf{A}|}{2} \times \log_2 |\mathbf{A}| \ .$$

Proof: Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ be uniformly distributed over $\mathbf{A}$.
Then $\mathbf{p(x)} = \mathbb{1}_{\mathbf{x} \in \mathbf{A}} / |\mathbf{A}|$.
Clearly, $\mathbf{H(X)} = \log |\mathbf{A}|$. Also,

$$\mathbf{H(X)} - \mathbf{H(X^{(i)})} = \mathbf{H(X_i | X^{(i)})} = -\sum_{\mathbf{x} \in \mathbf{A}} \mathbf{p(x)} \log \mathbf{p(x_i | x^{(i)})} \ .$$

For $\mathbf{x} \in \mathbf{A}$,

$$\mathbf{p(x_i | x^{(i)})} = \begin{cases} 1/2 & \text{if } \overline{\mathbf{x}}^{(i)} \in \mathbf{A} \\ 1 & \text{otherwise} \end{cases}$$

where $\overline{\mathbf{x}}^{(i)} = (\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, -\mathbf{x}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$.

$$\mathbf{H(X) - H(X^{(i)}) = \frac{\log 2}{|A|} \sum_{x \in A} \mathbb{1}_{x, \bar{x}^{(i)} \in A}}$$

and therefore

$$\sum_{i=1}^{n} \left(\mathbf{H(X) - H(X^{(i)})} \right) = \frac{\log 2}{|A|} \sum_{x \in A} \sum_{i=1}^{n} \mathbb{1}_{x, \bar{x}^{(i)} \in A} = \frac{|E(A)|}{|A|} 2 \log 2 \, .$$

Thus, by Han's inequality,

$$\frac{|E(A)|}{|A|} 2 \log 2 = \sum_{i=1}^{n} \left(\mathbf{H(X) - H(X^{(i)})} \right) \leq \mathbf{H(X)} = \log |A| \, .$$

This is equivalent to the edge isoperimetric inequality on the hypercube: if

$$\partial_E(A) = \left\{(x, x') : x \in A, x' \in A^c, d_H(x, x') = 1\right\} .$$

is the edge boundary of A, then

$$|\partial_E(A)| \geq \log_2 \frac{2^n}{|A|} \times |A|$$

Equality is achieved for sub-cubes.

VC entropy is self-bounding

Let $\mathcal{A}$ is a class of subsets of $\mathbf{X}$ and $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$.
Recall that $S(x, \mathcal{A})$ is the number of different sets of form

$$\{x_1, \ldots, x_n\} \cap A : A \in \mathcal{A}$$

Let $f_n(x) = \log_2 S(x, \mathcal{A})$ be the VC entropy.
Then $0 \leq f_n(x) - f_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_n) \leq 1$ and

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

Proof: Put the uniform distribution on the class of sets $\{x_1, \ldots, x_n\} \cap A$ and use Han's inequality.

VC entropy is self-bounding

Let $\mathcal{A}$ is a class of subsets of $\mathbf{X}$ and $x = (x_1, \ldots, x_n) \in \mathcal{X}^n$.
Recall that $S(x, \mathcal{A})$ is the number of different sets of form

$$\{x_1, \ldots, x_n\} \cap A : A \in \mathcal{A}$$

Let $f_n(x) = \log_2 S(x, \mathcal{A})$ be the VC entropy.
Then $0 \leq f_n(x) - f_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1} \ldots, x_n) \leq 1$ and

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

Proof: Put the uniform distribution on the class of sets $\{x_1, \ldots, x_n\} \cap A$ and use Han's inequality.
Corollary: if $X_1, \ldots, X_n$ are independent, then

$$\mathrm{Var}(\log_2 S(X, \mathcal{A})) \leq \mathbb{E} \log_2 S(X, \mathcal{A}) \ .$$

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.

herbst's argument: exponential concentration

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

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

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

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

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

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 $\mathbf{n} = 1$.

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

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

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

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

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 $\mathbf{Z_i = \inf_{x_i'} f(X_1, \ldots, x_i', \ldots, X_n)}$ 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 $\mathbf{Z_i = \inf_{x_i'} f(X_1, \ldots, x_i', \ldots, X_n)}$ 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 $A = (X_{i,j})_{n\times n}$ be symmetric, the $X_{i,j}$ independent ($i \le j$) with $|X_{i,j}| \le 1$. 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

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

Therefore,

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

example: convex lipschitz functions

Let $\mathbf{f} : [0,1]^n \to \mathbb{R}$ be a convex function. Let $Z_i = \inf_{x_i'} f(X_1, \ldots, x_i', \ldots, X_n)$ and let X_i' be the value of x_i' for which the minimum is achieved. Then, writing $\overline{X}^{(i)} = (X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)$,

$$\sum_{i=1}^n (Z - Z_i)^2 = \sum_{i=1}^n (f(X) - f(\overline{X}^{(i)}))^2$$

$$\leq \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(X) \right)^2 (X_i - X_i')^2$$

(by convexity)

$$\leq \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(X) \right)^2$$

$$= \|\nabla f(X)\|^2 \leq L^2 \ .$$

convex lipschitz functions

If $\mathbf{f : [0,1]^n \to \mathbb{R}}$ is a convex Lipschitz function and $\mathbf{X_1, \ldots, X_n}$ are independent taking values in $\mathbf{[0,1]}$, $\mathbf{Z = f(X_1, \ldots, X_n)}$ satisfies

$$\mathbb{P}\{\mathbf{Z > \mathbb{E}Z + t}\} \leq \mathbf{e^{-t^2/(2L^2)}} \; .$$

convex lipschitz functions

If $\mathbf{f : [0,1]^n \to \mathbb{R}}$ is a convex Lipschitz function and $\mathbf{X_1, \ldots, X_n}$ are independent taking values in $\mathbf{[0,1]}$, $\mathbf{Z = f(X_1, \ldots, X_n)}$ satisfies

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

A similar lower tail bound also holds.

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

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

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,

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

exponential efron-stein inequality

Define

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

and

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

By Efron-Stein,

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

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

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

If also $\mathbf{Z}_i' - \mathbf{Z} \le 1$ for every $\mathbf{i}$, fhen for all $\lambda \in (0, 1/2)$,

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

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

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

weakly self-bounding functions

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

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

Then

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

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

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

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

the isoperimetric view

Let $X = (X_1, \ldots, X_n)$ 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 $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ have independent components, taking values in $\mathcal{X}^n$. Let $\mathbf{A} \subset \mathcal{X}^n$.

The Hamming distance of $\mathbf{X}$ to $\mathbf{A}$ is

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

Michel Talagrand

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

the isoperimetric view

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ have independent components, taking values in $\mathcal{X}^n$. Let $\mathbf{A} \subset \mathcal{X}^n$.

The Hamming distance of $\mathbf{X}$ to $\mathbf{A}$ is

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

Michel Talagrand

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

Concentration of measure!

the isoperimetric view

Proof: By the bounded differences inequality,

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

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

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

By the bounded differences inequality again,

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

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

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

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:

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

Talagrand's convex distance inequality:

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

convex distance inequality

convex distance:

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

Talagrand's convex distance inequality:

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

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

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

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

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

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:

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

John von Neumann (1903–1957)

Sergei Lvovich Sobolev
(1908–1989)

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

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

and

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

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

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

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

$$\mathbf{f}(\mathbf{x}) \leq \mathbf{s} + \mathbf{D}(\mathbf{x}, \mathbf{A}_s) \leq \mathbf{s} + \mathbf{d}_T(\mathbf{x}, \mathbf{A}_s) ,$$

By the convex distance inequality,

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

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

empirical processes

Let $\mathcal{T}$ be a countable index set.
For $i = 1, \ldots, n$, let $X_i = (X_{i,s})_{s \in \mathcal{T}}$ be vectors of real-valued random variables. Assume that $X_1, \ldots, X_n$ are independent.

The empirical process is $\sum_{i=1}^n X_{i,s}$, $s \in \mathcal{T}$.

We study concentration of the supremum:

$$Z = \sup_{s \in \mathcal{T}} \sum_{i=1}^n X_{i,s} .$$

empirical processes–the variance

We may use Efron-Stein: let

$$Z_i = \sup_{s \in \mathcal{T}} \sum_{j:j \neq i} X_{j,s}$$

and $\widehat{s} \in \mathcal{T}$ be such that $Z = \sum_{i=1}^{n} X_{i,\widehat{s}}$. Then

$$(Z - Z_i)_+ \leq (X_{i,\widehat{s}})_+ \leq \sup_{s \in \mathcal{T}} |X_{i,s}|$$

so

$$\mathbf{Var}(Z) \leq \mathbb{E} \sum_{i=1}^{n} (Z - Z_i)^2 \leq \mathbb{E} \sum_{i=1}^{n} \sup_{s \in \mathcal{T}} X_{i,s}^2 \, .$$

empirical processes–the variance

A more clever use of Efron-Stein: suppose $\mathbb{E}X_{i,s} = 0$.

Let $Z_i' = \sup_{s \in \mathcal{T}} \left(\sum_{j \neq i} X_{j,s} + X_{i,s}' \right)$. Note that

$$(Z - Z_i')_+^2 \le \left(X_{i,\widehat{s}} - X_{i,\widehat{s}}' \right)^2 .$$

By Efron-Stein,

$$\begin{aligned}
\mathbf{Var}(Z) &\le \mathbb{E} \sum_{i=1}^{n} (Z - Z_i')_+^2 \\
&\le \mathbb{E} \sum_{i=1}^{n} \mathbb{E}' \left[\left(X_{i,\widehat{s}} - X_{i,\widehat{s}}' \right)^2 \right] \\
&\le \mathbb{E} \sum_{i=1}^{n} \left(X_{i,\widehat{s}}^2 + \mathbb{E}' \left[X_{i,\widehat{s}}'^2 \right] \right) \\
&\le \mathbb{E} \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} X_{i,s}^2 + \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} \mathbb{E} X_{i,s}^2 .
\end{aligned}$$

weak and strong variance

We have proved that

$$\mathbf{Var}(\mathbf{Z}) \leq \mathbf{V} \quad \text{and} \quad \mathbf{Var}(\mathbf{Z}) \leq \mathbf{\Sigma}^2 + \boldsymbol{\sigma}^2$$

where

weak and strong variance

We have proved that

$$\mathbf{Var(Z)} \leq \mathbf{V} \quad \text{and} \quad \mathbf{Var(Z)} \leq \boldsymbol{\Sigma}^2 + \boldsymbol{\sigma}^2$$

where

$$\mathbf{V} = \sum_{i=1}^{n} \mathbb{E} \sup_{s \in \mathcal{T}} X_{i,s}^2 \qquad \text{strong variance}$$

weak and strong variance

We have proved that

$$\mathbf{Var(Z)} \leq \mathbf{V} \quad \text{and} \quad \mathbf{Var(Z)} \leq \mathbf{\Sigma}^2 + \boldsymbol{\sigma}^2$$

where

$$\mathbf{V} = \sum_{i=1}^{n} \mathbb{E} \sup_{s \in \mathcal{T}} X_{i,s}^2 \qquad \text{strong variance}$$

$$\mathbf{\Sigma}^2 = \mathbb{E} \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} X_{i,s}^2 \qquad \text{weak variance}$$

weak and strong variance

We have proved that

$$\mathrm{Var}(Z) \leq V \quad \text{and} \quad \mathrm{Var}(Z) \leq \Sigma^2 + \sigma^2$$

where

$$V = \sum_{i=1}^{n} \mathbb{E} \sup_{s \in \mathcal{T}} X_{i,s}^2 \qquad \text{strong variance}$$

$$\Sigma^2 = \mathbb{E} \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} X_{i,s}^2 \qquad \text{weak variance}$$

$$\sigma^2 = \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} \mathbb{E} X_{i,s}^2 \qquad \text{wimpy variance}$$

weak and strong variance

We have proved that

$$\mathrm{Var}(Z) \leq V \quad \text{and} \quad \mathrm{Var}(Z) \leq \Sigma^2 + \sigma^2$$

where

$$V = \sum_{i=1}^{n} \mathbb{E} \sup_{s \in \mathcal{T}} X_{i,s}^2 \qquad \text{strong variance}$$

$$\Sigma^2 = \mathbb{E} \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} X_{i,s}^2 \qquad \text{weak variance}$$

$$\sigma^2 = \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} \mathbb{E} X_{i,s}^2 \qquad \text{wimpy variance}$$

$$\sigma^2 \leq \Sigma^2 \leq V .$$

weak and strong variance

If $\mathbb{E}X_{i,s} = 0$ and $|X_{i,s}| \leq 1$, we also have, by symmetrization and contraction arguments,

$$\Sigma^2 \leq 8\mathbb{E}Z + \sigma^2$$

and therefore

$$\mathrm{Var}(Z) \leq 8\mathbb{E}Z + 2\sigma^2 \ .$$

weak and strong variance

If $\mathbb{E}X_{i,s} = 0$ and $|X_{i,s}| \leq 1$, we also have, by symmetrization and contraction arguments,

$$\Sigma^2 \leq 8\mathbb{E}Z + \sigma^2$$

and therefore

$$\mathrm{Var}(Z) \leq 8\mathbb{E}Z + 2\sigma^2 \ .$$

If the X_i are also identicaly distributed, then

$$\mathrm{Var}(Z) \leq 2\mathbb{E}Z + \sigma^2 \ .$$

empirical processes–exponential inequalities

A Bernstein type inequality. "Talagrand's inequality".

empirical processes–exponential inequalities

A Bernstein type inequality. "Talagrand's inequality".

Assume $\mathbb{E}\mathbf{X}_{i,s} = \mathbf{0}$, and $|\mathbf{X}_{i,s}| \leq \mathbf{1}$. For $\mathbf{t} \geq \mathbf{0}$,

$$\mathbb{P}\left\{\mathbf{Z} \geq \mathbb{E}\mathbf{Z} + \mathbf{t}\right\} \leq \exp\left(-\frac{\mathbf{t}^2}{2\left(2(\boldsymbol{\Sigma}^2 + \sigma^2) + \mathbf{t}\right)}\right) .$$

proof.

For each $i = 1, \ldots, n$, let $Z'_i = \sup_{s \in \mathcal{T}} (X'_{i,s} + \sum_{j \neq i} X_{j,s})$.
We already proved that

$$\sum_{i=1}^{n} \mathbb{E}'(Z - Z'_i)_+^2 \leq \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} X_{i,s}^2 + \sigma^2 \stackrel{\text{def.}}{=} W + \sigma^2 \ .$$

By the exponential Efron-Stein inequality, for $\lambda \in [0, 1)$,

$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda}{1 - \lambda} \log \mathbb{E} e^{\lambda(W + \sigma^2)} \ .$$

proof.

For each $\mathbf{i = 1, \ldots, n}$, let $\mathbf{Z'_i = \sup_{s \in \mathcal{T}} (X'_{i,s} + \sum_{j \neq i} X_{j,s})}$.
We already proved that

$$\sum_{i=1}^{n} \mathbb{E}'(Z - Z'_i)_+^2 \leq \sup_{s \in \mathcal{T}} \sum_{i=1}^{n} X_{i,s}^2 + \sigma^2 \stackrel{\text{def.}}{=} W + \sigma^2 \; .$$

By the exponential Efron-Stein inequality, for $\lambda \in [0, 1)$,

$$\log \mathbb{E} e^{\lambda(Z - \mathbb{E}Z)} \leq \frac{\lambda}{1 - \lambda} \log \mathbb{E} e^{\lambda(W + \sigma^2)} \; .$$

$\mathbf{W}$ is a self-bounding function, so

$$\log \mathbb{E} e^{\lambda W} \leq \Sigma^2 \left(e^{\lambda} - 1 \right) \; .$$

Putting things together implies the inequality.

bousquet's inequality

A Bennett type inequality with the right constant.
Assume $\mathbf{X_1, \ldots, X_n}$ are i.i.d. with $\mathbb{E}\mathbf{X_{i,s} = 0}$ and $\mathbf{X_{i,s} \leq 1}$.
For all $\mathbf{t \geq 0}$,

$$\mathbb{P}\{\mathbf{Z} \geq \mathbb{E}\mathbf{Z} + \mathbf{t}\} \leq \mathbf{e}^{-\mathbf{v}\mathbf{h}(\mathbf{t}/\mathbf{v})} .$$

where $\mathbf{v} = 2\mathbb{E}\mathbf{Z} + \sigma^2$ and $\mathbf{h(u) = (1+u)\log(1+u) - u}$.
In particular,

$$\mathbb{P}\{\mathbf{Z} \geq \mathbb{E}\mathbf{Z} + \mathbf{t}\} \leq \exp\left(-\frac{\mathbf{t}^2}{2(\mathbf{v} + \mathbf{t}/3)}\right) .$$

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 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 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 $\mathbf{q} \geq \mathbf{2}$.

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

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

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

moment inequalities

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

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

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

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

More generally,

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

sums: khinchine's inequality

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

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

sums: khinchine's inequality

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

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

Proof:

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

Aleksandr Khinchin
(1894–1959)

sums: rosenthal's inequality

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

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

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

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

influences

If $\mathbf{A} \subset \{-1,1\}^n$ and $\mathbf{X} = (X_1, \ldots, X_n)$ is uniform, the influence of the $\mathbf{i}$-th variable is

$$I_i(A) = \mathbb{P}\left\{\mathbb{1}_{X \in A} \neq \mathbb{1}_{X^{(i)} \in A}\right\}$$

where $X^{(i)} = (X_1, \ldots, X_{i-1}, 1 - X_i, X_{i+1}, \ldots, X_n)$.

The total influence is

$$I(A) = \sum_{i=1}^{n} I_i(A) .$$

influences

If $A \subset \{-1,1\}^n$ and $X = (X_1, \ldots, X_n)$ is uniform, the influence of the i-th variable is

$$I_i(A) = \mathbb{P}\left\{\mathbb{1}_{X \in A} \neq \mathbb{1}_{X^{(i)} \in A}\right\}$$

where $X^{(i)} = (X_1, \ldots, X_{i-1}, 1 - X_i, X_{i+1}, \ldots, X_n)$.

The total influence is

$$I(A) = \sum_{i=1}^{n} I_i(A) \ .$$

Note that

$$I(A) = 2^{-(n-1)} |\partial_E(A)| \ .$$

influences: examples

dictatorship: $\mathbf{A = \{x : x_1 = 1\}}$. $\mathbf{I(A) = 1}$.

parity: $\mathbf{A = \{x : \sum_i \mathbb{1}_{x_i=1} \text{ is even}\}}$. $\mathbf{I(A) = n}$.

majority: $\mathbf{A = \{x : \sum_i x_i > 0\}}$. $\mathbf{I(A) \approx \sqrt{2n/\pi}}$.

$$\text{by Efron-Stein,} \quad \mathbf{P(A)(1 - P(A)) \leq \frac{I(A)}{4}}$$

so dictatorship has smallest total influence (if $\mathbf{P(A) = 1/2}$).

improved efron-stein on the hypercube

Recall that for any $\mathbf{f} : \{-1, 1\}^n \to \mathbb{R}$ under the uniform distribution,

$$\mathrm{Ent}(f^2) \leq 2\mathcal{E}(f)$$

where $\mathrm{Ent}(f^2) = \mathbf{E}\left[f^2 \log(f^2)\right] - \mathbf{E}\left[f^2\right] \log \mathbf{E}\left[f^2\right]$ and

$$\mathcal{E}(f) = \frac{1}{4}\mathbb{E}\left[\sum_{i=1}^{n}\left(f(X) - f(\overline{X}^{(i)})\right)^2\right]$$

This implies, for any non-negative $f : \{-1, 1\}^n \to [0, \infty)$,

$$\mathbf{E}\left[f^2\right] \log \frac{\mathbf{E}\left[f^2\right]}{\mathbf{E}\left[f\right]^2} \leq 2\mathcal{E}(f) \ .$$

improved efron-stein on the hypercube

Recall the Doob-martingale representation $\mathbf{f(X) - Ef = \sum_{i=1}^n \Delta_i}$.
One easily sees that

$$\mathcal{E}(\mathbf{f}) = \sum_{i=1}^{n} \mathcal{E}(\Delta_i) .$$

But then, by the previous lemma,

$$\begin{aligned}
\mathcal{E}(\mathbf{f}) &\geq \sum_{j=1}^{n} \mathcal{E}(|\Delta_j|) \geq \frac{1}{2} \sum_{j=1}^{n} \mathbf{E}\left[\Delta_j^2\right] \log \frac{\mathbf{E}\left[\Delta_j^2\right]}{(\mathbf{E}|\Delta_j|)^2} \\
&= -\frac{1}{2} \mathrm{Var}(\mathbf{f}) \sum_{j=1}^{n} \frac{\mathbf{E}\left[\Delta_j^2\right]}{\mathrm{Var}(\mathbf{f})} \log \frac{(\mathbf{E}|\Delta_j|)^2}{\mathbf{E}\left[\Delta_j^2\right]} \\
&\geq -\frac{1}{2} \mathrm{Var}(\mathbf{f}) \log \frac{\sum_{j=1}^{n} (\mathbf{E}|\Delta_j|)^2}{\mathrm{Var}(\mathbf{f})}
\end{aligned}$$

improved efron-stein on the hypercube

We obtained that for any $\mathbf{f : \{-1,1\}^n \to \mathbb{R}}$,

$$\mathbf{Var(f)} \log \frac{\mathbf{Var(f)}}{\sum_{j=1}^{n} (\mathbf{E}|\Delta_j|)^2} \le 2\mathcal{E}(\mathbf{f}) \ .$$

(Falik and Samorodnitsky, 2007; Rossignol, 2006).

improved efron-stein on the hypercube

We obtained that for any $\mathbf{f : \{-1, 1\}^n \to \mathbb{R}}$,

$$\mathbf{Var(f) \log \frac{Var(f)}{\sum_{j=1}^{n} (E|\Delta_j|)^2} \leq 2\mathcal{E}(f)} \ .$$

(Falik and Samorodnitsky, 2007; Rossignol, 2006).
"Slightly" better than Efron-Stein.

improved efron-stein on the hypercube

We obtained that for any $\mathbf{f : \{-1, 1\}^n \to \mathbb{R}}$,

$$\mathrm{Var}(f) \log \frac{\mathrm{Var}(f)}{\sum_{j=1}^{n} (\mathbf{E}|\Delta_j|)^2} \le 2\mathcal{E}(f) \ .$$

(Falik and Samorodnitsky, 2007; Rossignol, 2006).
"Slightly" better than Efron-Stein.

Use this for $f(x) = \mathbb{1}_{x \in A}$ for $A \subset \{-1, 1\}^n$:

$$P(A)(1 - P(A)) \log \frac{4P(A)(1 - P(A))}{\sum_i I_i(A)^2} \le \frac{I(A)}{4}$$

kahn, kalai, linial

Corollary: (Kahn, Kalai, Linial, 1988).

$$\max_i \mathbf{I_i(A)} \geq \frac{\mathbf{P(A)(1 - P(A)) \log n}}{\mathbf{n}}$$

If the influences are equal,

$$\mathbf{I(A) \geq P(A)(1 - P(A)) \log n}$$

Another corollary: (Friedgut, 1998).
If $\mathbf{I(A) \leq c}$, $\mathbf{A}$ (basically) depends on a bounded number of variables. $\mathbf{A}$ is a "junta."

threshold phenomena

Let $\mathbf{A} \subset \{-1, 1\}^{\mathbf{n}}$ be a monotone set and let $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_n})$ be such that

$$\mathbb{P}\{\mathbf{X_i} = 1\} = \mathbf{p} \qquad \mathbb{P}\{\mathbf{X_i} = -1\} = 1 - \mathbf{p}$$

$$\mathbf{P_p(A)} = \sum_{\mathbf{x} \in \mathbf{A}} \mathbf{p}^{\|\mathbf{x}\|}(1 - \mathbf{p})^{\mathbf{n} - \|\mathbf{x}\|}$$

is an increasing function of $\mathbf{p} \in [0, 1]$.

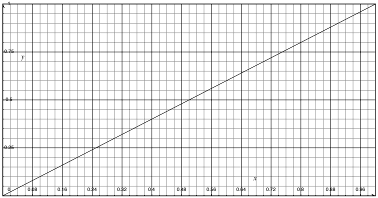
Let $\mathbf{p_a}$ be such that $\mathbf{P_{p_a}(A) = a}$.

Critical value $= \mathbf{p_{1/2}}$

Threshold width: $\mathbf{p_{1-\varepsilon} - p_\varepsilon}$

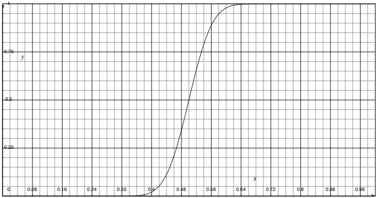
two (extreme) examples

dictatorship



threshold width $= \mathbf{1 - 2}\varepsilon$

majority (with $\mathbf{n = 101}$)



$\leq \sqrt{\mathbf{log(1/}\varepsilon\mathbf{)/(2n)}}$

In what cases do we have a quick transition?

russo's lemma

If $\mathbf{A}$ is monotone,

$$\frac{dP_p(A)}{dp} = I^{(p)}(A)$$

The Kahn, Kalai, Linial result, generalized for $p \neq 1/2$, implies that
if $\mathbf{A}$ is such that $I_1^{(p)} = I_2^{(p)} = \cdots = I_n^{(p)}$, then

$$p_{1-\varepsilon} - p_\varepsilon = O\left(\frac{\log\frac{1}{\varepsilon}}{\log n}\right)$$

On the other hand, if $p_{3/4} - p_{1/4} \geq c$ then $\mathbf{A}$ is (basically) a junta.

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.

thank you for the organization!

thank you for the organization!

Markus Reiß

thank you for the organization!

Markus Reiß