bounded differences inequality–examples

Recall the L_1 error of the kernel density estimator

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

where

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

We saw that Z satisfies the boubded differences property with $c_i = 2/n$ for all i.
We obtain

$$\mathbb{P}\{|Z - \mathbb{E}Z| > t\} \leq 2e^{-nt^2/2} \ .$$

hoeffding in a hilbert space

Let $X_1, \ldots, X_n$ be independent zero-mean random variables in a separable Hilbert space such that $\|X_i\| \le 1$. Then, for all $t \ge 2\sqrt{n}$,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^{n} X_i\right\| > t\right\} \le e^{-t^2/(8n)} .$$

hoeffding in a hilbert space–proof

By the triangle inequality, $\left\|\sum_{i=1}^{n} \boldsymbol{X}_i\right\|$ has the bounded differences property with constants 2, so

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

$$\leq \exp\left(-\frac{\left(\boldsymbol{t} - \mathbb{E}\left\|\sum_{i=1}^{n} \boldsymbol{X}_i\right\|\right)^2}{2\boldsymbol{n}}\right) .$$

Also,

$$\mathbb{E}\left\|\sum_{i=1}^{n} \boldsymbol{X}_i\right\| \leq \sqrt{\mathbb{E}\left\|\sum_{i=1}^{n} \boldsymbol{X}_i\right\|^2} = \sqrt{\sum_{i=1}^{n} \mathbb{E}\left\|\boldsymbol{X}_i\right\|^2} \leq \sqrt{\boldsymbol{n}} .$$

bounded differences inequality

✳Easy to use.

✳Distribution free.

✳Often close to optimal (e.g., L_1 error of kernel density estimate).

✳Does not exploit "variance information."

✳Often too rigid.

✳Other methods are necessary.

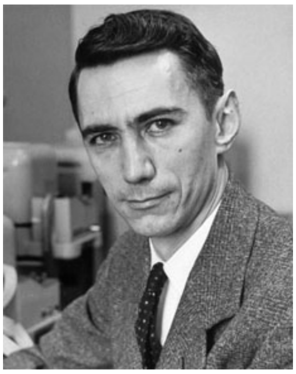
shannon entropy

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

Te Sun Han

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:

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

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

edge isoperimetric inequality on the hypercube

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

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

Proof: Let $X = (X_1, \ldots, X_n)$ be uniformly distributed over A.
Then $p(x) = \mathbb{1}_{x \in A}/|A|$.
Clearly, $H(X) = \log |A|$. Also,

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

For $x \in A$,

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

where $\overline{x}^{(i)} = (x_1, \ldots, x_{i-1}, -x_i, x_{i+1}, \ldots, x_n)$.

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

and therefore

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

Thus, by Han's inequality,

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

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

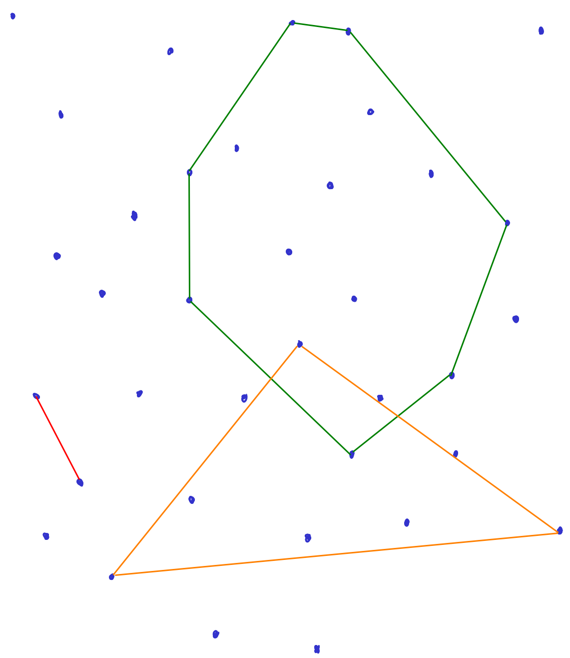
$$\partial_E(A) = \left\{(x, x') : x \in A, x' \in A^c, d_H(x, x') = 1\right\} \ .$$

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.

combinatorial entropies–an example



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

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

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.

This is a deterministic property so fix the points.

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,

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

VC entropy is self-bounding

Let $\mathcal{A}$ is a class of subsets of 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

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

vapnik and chervonenkis


Vladimir Vapnik


Alexey Chervonenkis

subadditivity of entropy

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

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

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

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

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

Then

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

a logarithmic sobolev inequality on the hypercube

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

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

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

Implies Efron-Stein.

Sergei Lvovich Sobolev
(1908–1989)

herbst's argument: exponential concentration

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

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

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

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

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

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

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 $Z = f(X)$,

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

(Gross, 1975).

Proof sketch: Similar to how we proved the Gaussian Poincaré inequality from Efron-Stein.
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 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 f is Lipschitz: for all $x, y \in \mathbb{R}^n$,

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

Then, for all $t > 0$,

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

(Tsirelson, Ibragimov, and Sudakov, 1976).

an application: supremum of a gaussian process

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

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

then

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

Proof: We have already seen that Z can be written as a σ-Lipschitz function of a standard normal vector.

beyond bernoulli and gaussian: the entropy method

For general distributions, logarithmic Sobolev inequalities are not available.

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

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

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

Michel Ledoux

the entropy method

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

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

Then for all $t > 0$,

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

This implies the bounded differences inequality and much more.

example: convex lipschitz functions

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

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

self-bounding functions

Suppose Z satisfies

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

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

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

and

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

combinatorial entropies, configuration functions are examples of self bounding functions.

conditional rademacher average

Let $X_1, \ldots, X_n$ be independent, taking values in a set $\mathcal{X}$ and let $\mathcal{A}$ be a class of subsets of $\mathcal{X}$.
The conditional Rademacher average is

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

concentration of conditional rademacher average

Define

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

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,

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

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

The exponential inequalities also apply.

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,

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

For all $\lambda, \theta > 0$ with $\lambda\theta < 1$,

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

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

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

weakly self-bounding functions

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

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

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

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

"An isoperimetric inequality".

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,

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

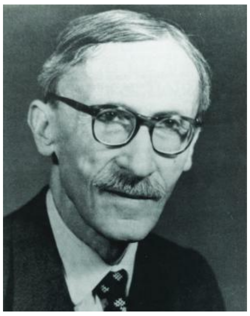
isoperimetry implies concentration

Suppose

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

Let $f : \mathcal{X}^n \to \mathbb{R}$ satisfy the bounded differences property (with $c_i = 1$).
Then, taking $A = \{x \in \mathcal{X}^n : f(x) \leq \mathbb{M}f(X)\}$, we have $\mathbb{P}[A] \geq 1/2$ and

$$\begin{aligned} \mathbb{P}\{f(X) > \mathbb{M}f(X) + t\} &\leq \mathbb{P}\{d(X,A) \geq t\} \\ &\leq e^{-2(t-\sqrt{n}/(2/\log 2))^2/n} \end{aligned}$$



Paul Lévy (1886–1971)