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:

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

Talagrand's convex distance inequality:

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

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

Talagrand's original proof was different.

convex lipschitz functions

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

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

If A is convex, then

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

Proof:

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

John von Neumann (1903–1957)

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.

empirical processes

Let $\mathcal{T}$ be a countable index set.
For $i = 1, \ldots, n$, let $X_i = \left(X_{i,s}\right)_{s \in \mathcal{T}}$ 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

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

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

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

and therefore

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

If the $\boldsymbol{X_i}$ are also identicaly distributed, then

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

empirical processes–exponential inequalities

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

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

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

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 $\lambda \in [0, 1)$,

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

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 $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}X_{i,s} = 0$ and $X_{i,s} \leq 1$.
For all $t \geq 0$,

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

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

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

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.

$\phi(x) = x^2$ 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

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

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

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 $q \geq 2$.

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

More generally,

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

sums: khinchine's inequality

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

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

Proof:

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

Aleksandr Khinchin
(1894–1959)

sums: rosenthal's inequality

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

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

Then for any integer $q \geq 2$,

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

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: $\boldsymbol{A} = \{\boldsymbol{x} : \boldsymbol{x}_1 = 1\}$. $\boldsymbol{I}(\boldsymbol{A}) = 1$.

parity: $\boldsymbol{A} = \{\boldsymbol{x} : \sum_{\boldsymbol{i}} \mathbb{1}_{\boldsymbol{x}_{\boldsymbol{i}}=1} \text{ is even}\}$. $\boldsymbol{I}(\boldsymbol{A}) = \boldsymbol{n}$.

majority: $\boldsymbol{A} = \{\boldsymbol{x} : \sum_{\boldsymbol{i}} \boldsymbol{x}_{\boldsymbol{i}} > 0\}$. $\boldsymbol{I}(\boldsymbol{A}) \approx \sqrt{2\boldsymbol{n}/\pi}$.

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

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

improved efron-stein on the hypercube

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

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

where $\mathbf{Ent}(f^2) = E\left[f^2 \log(f^2)\right] - E\left[f^2\right] \log E\left[f^2\right]$ 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 $f : \{-1,1\}^n \to [0,\infty)$,

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

improved efron-stein on the hypercube

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

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

But then, by the previous lemma,

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

improved efron-stein on the hypercube

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

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

(Falik and Samorodnitsky, 2007; Rossignol, 2006; Talagrand (1994)).

"Slightly" better than Efron-Stein.

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

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

kahn, kalai, linial

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

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

If the influences are equal,

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

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

threshold phenomena

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

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

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

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

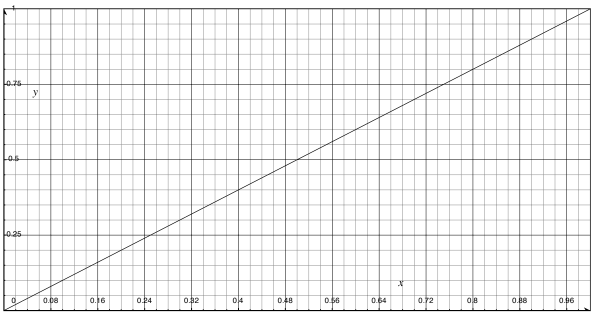
Let p_a be such that $P_{p_a}(A) = a$.

Critical value $= p_{1/2}$

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

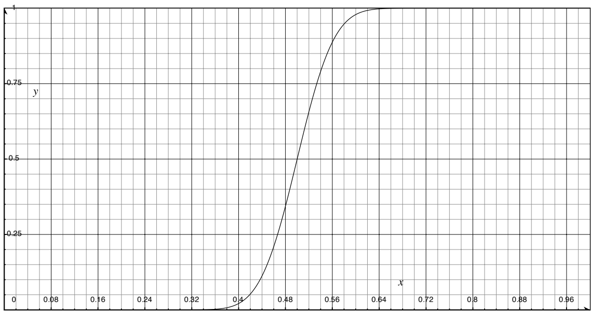
two (extreme) examples

dictatorship



threshold width $= 1 - 2\varepsilon$

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



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

In what cases do we have a quick transition?

russo's lemma

If 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 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 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.