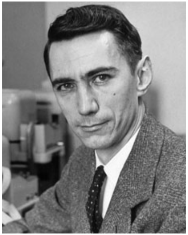
shannon entropy

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

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

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

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



Claude Shannon
(1916–2001)

han's inequality

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

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

Te Sun Han

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

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

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 $d_H(x, x') = 1$. Then

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

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

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

and therefore

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

Thus, by Han's inequality,

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

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

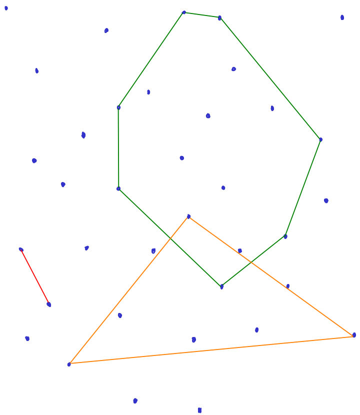
$$\partial_E(A) = \left\{(x, x') : x \in A, x' \in A^c, d_H(x, x') = 1\right\} \ .$$

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 $\mathbf{X_1, \ldots, X_n}$ be independent points in the plane (of arbitrary distribution!).
Let $\mathbf{N}$ be the number of subsets of points that are in convex position.
Then

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

proof

By Efron-Stein, it suffices to prove that $\mathbf{f}$ 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,

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

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

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

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

Denote

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

Then by Han's inequality,

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

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.

Sergei Lvovich Sobolev
(1908–1989)

herbst's argument: exponential concentration

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

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

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

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

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

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

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

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 $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}$ 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).

beyond bernoulli and gaussian: the entropy method

For general distributions, logarithmic Sobolev inequalities are not available.

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

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

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

Michel Ledoux

the entropy method

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

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

Then for all $t > 0$,

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

This implies the bounded differences inequality and much more.

self-bounding functions

Suppose $\mathbf{Z}$ satisfies

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

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

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

and

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

exponential efron-stein inequality

Define

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

and

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

By Efron-Stein,

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

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

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

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

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

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

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

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

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

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

the isoperimetric view

Let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ 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

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

the isoperimetric view

Proof: By the bounded differences inequality,

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

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

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

By the bounded differences inequality again,

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

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:

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

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.

It can also be recovered from Marton's transportation inequality.

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

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

Proof:

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

convex lipschitz functions

Let $\mathbf{X} = (X_1, \ldots, X_n)$ 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.

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

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

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,

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

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 $f : \{-1, 1\}^n \to [0, \infty)$,

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

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 $\mathbf{f : \{-1, 1\}^n \to \mathbb{R}}$,

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

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

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

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

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 $\mathbf{A} \subset \{-1, 1\}^n$ be a monotone set and let $\mathbf{X} = (X_1, \ldots, X_n)$ 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 $p \in [0, 1]$.

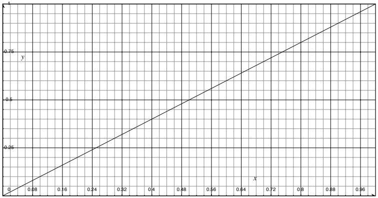
Let p_a be such that $P_{p_a}(A) = a$.

Critical value $= p_{1/2}$

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

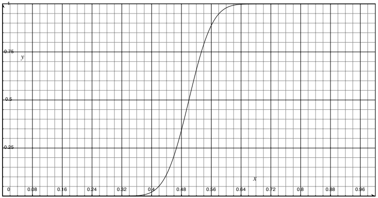
two (extreme) examples

dictatorship



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

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



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

In what cases do we have a quick transition?

russo's lemma

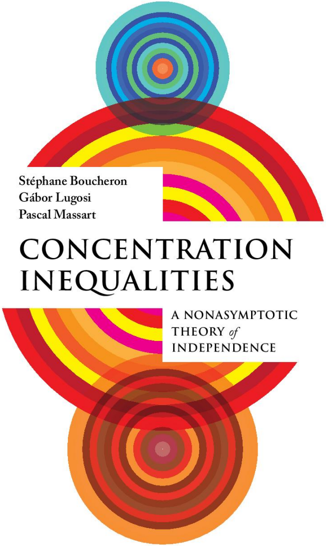
If $\mathbf{A}$ is monotone,

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

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.



Stéphane Boucheron
Gábor Lugosi
Pascal Massart

CONCENTRATION INEQUALITIES

A NONASYMPTOTIC THEORY *of* INDEPENDENCE

OXFORD