CONCENTRATION INEQUALITIES

JOINT TELEMATICS GROUP
IEEE INFORMATION THEORY SOCIETY
SUMMER SCHOOL

June 27-July 1, 2016.
IISC, Bangalore

bvrao@cmi.ac.in

DAY 1

STEPHANE BOUCHERON
GABOR LUGOSI
PASCAL MASSART

CONCENTRATION INEQUALITIES

OXFORD UNIVERSITY PRESS (2013)
****
MICHEL LEDOUX

THE CONCENTRATION of MEASURE PHENOMENON

Amer. Math Soc. (2000)

DEVDATT P DUBHASHI
ALESSANDRO PANCONESI

CONCENTRATION of MEASURE for the
ANALYSIS of RANDOMIZED ALGORITHMS

CAMBRIDGE UNIV. PRESS (2009)
****

MAXIM RAGINSKY
IGAL SASON

CONCENTRATION of MEASURE INEQUALITIES
in
INFORMATION THEORY, COMMUNICATION and CODING

nowPublishers (2013)

SOURAV CHATTERJEE

CONCENTRATION INEQUALITIES
WITH EXCHANGEABLE PAIRS

STANFORD UNIV. THESIS (2006)
****

SHANKAR BHAMIDI
NOTES of COURSE at CHAPEL HILL

Organization::

Overview and Chebyshev

Cramer; Chernoff; Hoeffding; Azuma, McDiarmid

Effron-Stein

Stein and Chatterjee

Entropy Log-Sobolev

Talagrand.

Old adage

Do Not Leave Anything to CHANCE

Old adage

Do Not Leave Anything to CHANCE

New Philosophy

CHANCE WORKS WONDERS: MAKE IT WORK FOR YOU

Toss fair coin ONCE. Can you say anything?

NOT MUCH

Toss Fair Coin 1000 times. Can you say anything?

Toss fair coin ONCE. Can you say anything?

NOT MUCH

Toss Fair Coin 1000 times. Can you say anything?

YES, We get Approximately 500 heads.

More number of outcomes did NOT lead to more uncertainty!
(if you think of right attribute)

Collective behaviour of LARGE number of particles

Performance of a programme with LARGE number of components

Messages with LARGE number of signals.

A random variable that depends in a smooth way on influence of a large number of independent random variables, but not too much on any one of them,

is

essentially constant and satisfies Chernoff type bounds.

(M Talagrand/V D Milman)

($\mathfrak{a}$)
P L Chebyshev $X \geq 0 \quad t > 0$:: $\qquad P(X \geq t) \leq E(X)/t.$
Use $X \geq X\, I_{(X \geq t)} \geq t\, I_{(X \geq t)}$

$$P(X \geq t) \leq E(X^2)/t^2.$$

X a RV mean μ variance σ^2 (finite)

$$P(|X - \mu| > t) \leq \sigma^2/t^2.$$

!NOT CONCENTRATION INEQUALITIES!
Usually CONCENTRATION refers to EXPONENTIAL TAIL BOUNDS. (People differ)
THEN WHY AM I WASTING YOUR TIME ON THIS?

THIS ITSELF GIVES RESULTS if WE ARE CLEVER

ALWAYS THE SEED. YOU KEEP REFINING.

USUALLY X IS A FUNCTION OF SEVERAL INDEPENDENT RANDOM VARIABLES.

USE FUNCTIONS OTHER THAN SQUARE

GET BETTER BOUNDS FOR σ^2

OR DO BOTH

$(\mathfrak{b})$

WEAK LAW of LARGE NUMBERS (WLLN)

$X_1, X_2, \cdots$, INDEPENDENT IDENTICALLY DISTRIBUTED

mean μ variance σ^2

$$A_n = \frac{1}{n}\sum_1^n X_i$$

$$P(|A_n - \mu| \geq \epsilon) \leq \frac{1}{n}\,\frac{\sigma^2}{\epsilon^2} \to 0$$

WHY IS IT INTERESTING?

($\mathfrak{c}$)
WEIERSTRASS

f CONTINUOUS REAL VALUED FUNCTION on $[0,1]$

for $n \geq 1$

$$P_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}x^k(1-x)^{n-k}$$

GIVEN $\epsilon > 0$ THERE IS N SUCH THAT FOR ALL $n \geq N$

$$\sup_{x\in[0,1]} |f(x) - P_n(x)| \leq \epsilon.$$

WHY IS THIS INTERESTING?

Use:: probabilities add to one.

$$|f(x) - P_n(x)| \leq \sum_{k=0}^{n} \left| f\left(\frac{k}{n}\right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k}$$

Choose:: $\delta > 0$: $\quad |x - y| \leq \delta \Rightarrow |f(x) - f(y)| < \epsilon/2$.

$\leq \{$sum over k with $|\frac{k}{n} - x| \leq \delta$ $\} +$ $\{$sum where $|\frac{k}{n} - x| > \delta\}$.

$$\leq (\epsilon/2) \;+\; 2Cnx(1-x)/n^2\delta^2$$

N such that $2C/N\delta^2 < \epsilon/2$. Here C is bound for f.
Use Chebyshev. Note: N does not depend on x.

$(\mathfrak{d})$
ERDOS:

GIVEN INTEGERS $k > 2$ AND $l > 2$

THERE IS A GRAPH G WHICH

HAS CHROMATIC NUMBER AT LEAST k

HAS NO CYCLES OF LENGTH SMALLER THAN l.

Took more than a decade to actually construct!

$0 < p < 1$; $G(n,p)$ is Erdos-Renyi graph on n vertices, edge probability p. Choose θ such that $\theta l < 1$.
(for each n) Take $p = n^{\theta - 1}$ (depending on n).
X RV; number of cycles of length at most l in $G(n,p)$

$$E(X) = \sum_{3}^{l} \frac{(n)_i}{2i} p^i \leq \sum_{3}^{l} n^i \frac{n^{\theta i - i}}{2i}$$

$$\leq \sum_{3}^{l} \frac{n^{\theta i}}{2i} = o(n)$$

Now CONSIDER LARGE n so that

$$P(X > n/2) < 1/2$$

Let $I =$ cardinality of maximal independent set.

$$P(I \geq x) \leq \binom{n}{x}(1-p)^{\binom{x}{2}} \leq \left[ne^{-p(x-1)/2}\right]^x$$

Let $x = 1 + \text{Ceiling}(\frac{3}{p}\log n)$ (depends on n)
$e^{p(x-1)/2} \geq n^{3/2}$; $P(I \geq x) = o(1)$.

Now CONSIDER n so large that

$$P(I > x_n) < 1/2$$

Take large n so that both above hold. Pick a graph such that

$$I(G) < x_n \leq 1 + 3n^{1-\theta} \log n$$

$$X(G) < n/2.$$

REMOVE ONE VERTEX FROM EACH OF ITS CYCLES of LENGTH AT MOST l.
Have Graph G^* on at least $n/2$ vertices.
For any graph H

$$|H| \leq I(H)\chi(H)$$

WHY?

Colour with $\chi(H)$ colours. V_i vertices of colour i.
V_i is Independent set $\quad |V_i| \leq I(H)$. $\quad H = \cup V_i$

$$\chi(G^*) \geq \frac{|G^*|}{I(G^*)} \geq \frac{n/2}{1 + 3n^{1-\theta} \log n} \uparrow \infty.$$

For all large n right side exceeds k.

($\mathfrak{c}$)
Cramer-Chernoff::

$$X_1, X_2, \cdots$$

i.i.d values ± 1 probabilities $1/2$ each. $\qquad S_n = \sum_1^n X_i$
For any $\lambda > 0$

$$P(S_n > t) = P(e^{\lambda S_n} > e^{\lambda t}) \leq E\left(e^{\lambda S_n}\right) e^{-\lambda t}.$$

Minimized at $\lambda = t/n$ giving

$$P(|S_n| > t) \leq 2e^{-t^2/2n}$$

This is concentration inequality.

$$E(e^{\lambda S_n}) = \left(\frac{e^{\lambda} + e^{-\lambda}}{2}\right)^n$$

$$\frac{1}{2}(e^{\lambda} + e^{-\lambda}) = 1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \frac{\lambda^6}{6} + \cdots$$

$$\leq 1 + \frac{\lambda^2/2}{1} + \frac{(\lambda^2/2)^2}{2!} + \frac{(\lambda^2/2)^3}{3!} + \cdots$$

$$= e^{\lambda^2/2}.$$

$$E(e^{\lambda S_n})e^{-\lambda t} \leq \exp\{n\frac{\lambda^2}{2} - \lambda t\}$$

So minimize. $n\lambda - t = 0$ or $\lambda = t/n$

(f)
STRONG LAW OF LARGE NUMBERS (SLLN)

$$\sum_n P(|S_n/n| > t) \leq 2 \sum_n e^{-nt^2/2} < \infty.$$

Borel-Cantelli shows; Almost surely

$$S_n/n \to 0.$$

[A_n events $\sum P(A_n) < \infty$. Let A be the set of points which belong to infinitely many of these events. Then $P(A) = 0$.]
Why is it interesting?

Points that belong to infinitely many sets A_i is

$$\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i.$$

probability of this set, whatever n you take, is smaller than $\sum_{i=n}^{\infty} P(A_i)$ tail sum of a convergent series.

Hence probability of the set is zero.

($\mathfrak{g}$) Main Point:

Z RV. Assume $E\{e^{\lambda Z}\} < \infty$.
(at least for some $\lambda > 0$. We consider those positive λ below)

$$\Psi(\lambda) = \log E(e^{\lambda Z})$$

$$P(Z > t) \leq e^{\Psi(\lambda)} e^{-\lambda t} = e^{-[\lambda t - \Psi(\lambda)]}.$$

Set

$$\Psi^*(t) = \sup\{\lambda t - \Psi(\lambda) : \lambda \geq 0\}$$

$$P(Z > t) \leq e^{-\Psi^*(t)}.$$

($\mathfrak{h}$) An example:

$$Z \sim N(0, \sigma^2)$$
$$\Psi(\lambda) = \lambda^2\sigma^2/2$$
$$\Psi^*(t) = t^2/2\sigma^2$$

For $t > 0$

$$P(Z > t) \leq e^{-t^2/2\sigma^2}.$$

$$P(|Z| > t) \leq 2e^{-t^2/2\sigma^2}.$$

(i) JOHNSON-LINDENSTRAUSS:
Given: a set S of n points in R^D; $0 < \epsilon < 1$; $0 < \delta < 1$
Take

$$d \geq \frac{100}{\epsilon^2} \ \log(\frac{n}{\sqrt{\delta}})$$

Take $W = ((W_{ij}))_{d \times D}$; $\{W_{ij}\}$ i.i.d. $N(0,1)$.

$X = \frac{1}{\sqrt{d}} W$ transforms R^D to R^d.
CONCLUSION: With probability at least $1 - \delta$; for $v_1, v_2 \in S$

$$(1-\epsilon)_{|}|v_1 - v_2|| \leq ||Xv_1 - Xv_2|| \leq (1+\epsilon)||v_1 - v_2||.$$

Data compression. d did not depend on D, depended on n.

X is 'ISOMETRY' 'ON THE AVERAGE'.
$W_i v = \sum_j W_{ij} v_j; \quad X_i v = \frac{1}{\sqrt{d}} W_i v$
$Wv = (W_1 v, \cdots, W_d V)'.; \quad Xv = Wv/\sqrt{d}$
let $v \in R^D$. Then $E(Wv)$ is zero vector.

$$W_i v \sim N(0, ||v||^2); \qquad X_i v \sim N(0, ||v||^2/d)$$

$$Wv \sim N_d(0, ||v||^2\ I); \qquad Xv \sim N_d(0, ||v||^2 d^{-1}\ I)$$

$$E(\|Xv\|^2) = \|v\|^2$$

proceed to do only one inequality of the theorem.

Now take $v \in R^D$ and $||v|| = 1$.

$$Ee^{\lambda W_i v} = e^{\lambda^2/2}.$$

$$||Xv||^2 - 1 = \frac{1}{d}\sum_i [(W_i v)^2 - 1] = \frac{1}{d} Z \text{ say}$$

Have for $0 < \lambda < 1/2$.

$$\log Ee^{\lambda[(W_i v)^2 - 1]} = \log \frac{1}{\sqrt{1-2\lambda}} - \lambda \leq \frac{\lambda^2}{1-2\lambda}$$

$$\log Ee^{\lambda Z} \leq \frac{d\lambda^2}{1-2\lambda}$$

ξ standard normal. What is $E(e^{\lambda \xi^2})$

No mathematical issues; assume $0 < \lambda < 1/2$.

$$\int e^{\lambda x^2} \int \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$= \int \frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2}{2}(1-2\lambda)} dx$$

$$\frac{1}{\sqrt{(1-2\lambda)}} \frac{1}{\sqrt{2\pi}} \int e^{-u^2/2} du$$

$0 < x < 1$

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots$$

$$\geq -x - \frac{x^2}{2} - \frac{x^3}{2} - \frac{x^4}{2} - \cdots$$

$$\geq -x - \frac{x^2}{2}\frac{1}{1-x}$$

$$\log\frac{1}{\sqrt{(1-2\lambda)}} \leq \lambda + \lambda^2\frac{1}{1-2\lambda}$$

Accept: For $t > 0$

$$\sup[t\lambda - \frac{d\lambda^2}{1-2\lambda} : 0 < \lambda < 1/2] = \frac{d}{2}\left[1 + \frac{t}{d} - \sqrt{1 + \frac{2t}{d}}\right] \quad (\spadesuit)$$

$$P(Z > u) \le e^{-[d+u-\sqrt{d^2+2ud}]/2}$$

For $t > 0$

$$P(\frac{1}{d}Z > \sqrt{\frac{4t}{d}} + \frac{2t}{d}) = P(Z > \sqrt{4td} + 2t)$$

$$\le e^{-[d+2t+\sqrt{4td}-\sqrt{d^2+4td+2d\sqrt{4td}}]/2}$$

$$= e^{-t}$$

$$T = \{\frac{v - w}{||v - w||} : v \neq w; v, w \in S\}$$

$$P(||Wv||^2 - 1 > \sqrt{\frac{4t}{d}} + \frac{2t}{d} \text{ for some } v \in T) \leq n^2 e^{-t}$$

Take $t = \log(n^2/\delta)$

$$n^2 e^{-t} = \delta$$

Shall show

$$\sqrt{\frac{4t}{d}} + \frac{2t}{d} < \epsilon$$

$$\frac{4t}{d} \leq \frac{4\log(n^2/\delta)}{100\log(n/\sqrt{\delta})}\epsilon^2 = \frac{8\epsilon^2}{100} \leq 2\epsilon^2/25$$

$$\sqrt{\frac{4t}{d}} \leq 2\epsilon/5$$

$$\frac{2t}{d} \leq \frac{4\epsilon^2}{100} = \epsilon/5$$

ADD and DONE!

Now proof of (♠) For $t > 0$

$$\sup[t\lambda - d\frac{\lambda^2}{1-2\lambda} : 0 < \lambda < 1/2] = \frac{d}{2}\left[1 + \frac{t}{d} - \sqrt{1 + \frac{2t}{d}}\ \right]$$

$$t\lambda - d\frac{\lambda^2}{1-2\lambda} = t\lambda + \frac{d}{4}[1 + 2\lambda - \frac{1}{1-2\lambda}]$$

Derivative =0 gives

$$t + \frac{d}{4}[2 - \frac{2}{(1-2\lambda)^2}] = 0$$

$$\frac{2t+d}{2} = \frac{d}{2(1-2\lambda)^2}$$

$$1 - 2\lambda = \sqrt{\frac{d}{d+2t}}$$

$$\lambda = \frac{1}{2} - \frac{1}{2}\sqrt{\frac{d}{d+2t}} = \frac{1}{2} - \frac{1}{2\alpha}$$

$$\alpha = \sqrt{\frac{d+2t}{d}}$$

Also

$$1 + 2\lambda = 2 - \frac{1}{\alpha}; \quad 1 - 2\lambda = \frac{1}{\alpha}$$

Sup equals

$$\frac{t}{2} - \frac{t}{2\alpha} + \frac{d}{4}[2 - \frac{1}{\alpha} - \alpha]$$

$$\frac{1}{\alpha} + \alpha = \frac{1+\alpha^2}{\alpha} = \frac{1 + \frac{d+2t}{d}}{\alpha} = \frac{2d+2t}{d\alpha}$$

So sup equals

$$\frac{t}{2} - \frac{t}{2\alpha} + \frac{d}{2} - \frac{d+t}{2\alpha}$$

$$= \frac{1}{2}[d + t - \frac{d+2t}{\alpha}]$$

$$= \frac{1}{2}[d + t - \sqrt{d(d+2t)}]$$

$$= \frac{d}{2}\left[1 + \frac{t}{d} - \sqrt{1 + 2\frac{t}{d}}\right]$$

CONCENTRATION INEQUALITIES

JOINT TELEMATICS GROUP
IEEE INFORMATION THEORY SOCIETY
SUMMER SCHOOL

June 27-July 1, 2016.
IISC, Bangalore

bvrao@cmi.ac.in
DAY 2

(j)

(Digression)
A RV Z with mean zero is SUBGAUSSIAN if

$$\Psi(\lambda) \leq \frac{1}{2}\lambda^2\theta; \quad \lambda \in R$$

for some $\theta > 0$. If Z is so then $-Z$ is also so.
For such a RV Z we have for $t > 0$,

$$P(Z > t) \leq e^{-t^2/2\theta}$$

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

(Subgaussian with variance parameter θ)

$(\mathfrak{t})$

(Digression) Nomenclature:
CRAMER transform:

$$\Psi^*(t) = \sup\{\lambda t - \Psi(\lambda) : \lambda > 0\}$$

FENCHEL-LEGENDRE transform

$$\Psi^*(t) = \sup\{\lambda t - \Psi(\lambda) : \lambda \in R\}$$

Jensen tells $\Psi(\lambda) \geq \lambda E(Z)$.
So: If $t \geq E(Z)$ then for $\lambda \leq 0$

$$\lambda t - \Psi(\lambda) \leq \lambda[t - E(Z)]$$

$$\leq 0 \quad \text{if} \quad \lambda < 0.$$

F-L is same as C.

(l)

(Digression) WHAT IF COIN IS BIASED?

$X_1, X_2, \cdots$ values $1/0$ probabilities p and $1-p=q$;independent..

$$S_n = \sum_1^n X_i \sim B(n,p)$$

$$P(S_n > (p+t)n) \le (pe^\lambda + q)^n e^{-(p+t)n\lambda}.$$

Minimized at

$$e^\lambda = \frac{q(p+t)}{p(q-t)}.$$

$$P(S_n > (p+t)n) \le e^{-nH}$$

where

$$H = [(p+t)\log\frac{p+t}{p} + (q-t)\log\frac{q-t}{q}]$$

Should minmize

$$(pe^{\lambda}+q)^{n}e^{-n\lambda(p+t)}.$$

Minimize

$$(pe^{\lambda}+q)e^{-\lambda(p+t)}.$$

Or its logarithm

$$\log(pe^{\lambda}+q)-\lambda(p+t)$$

derivative equate to zero.

$$pe^{\lambda}=(pe^{\lambda}+q)(p+t)$$

$$p(q-t)e^{\lambda}=q(p+t); \qquad e^{\lambda}=\frac{q(p+t)}{p(q-t)}.$$

Then

$$(pe^{\lambda}+q)=\frac{q(p+t)}{q-t}+q=\frac{q}{q-t}$$

$$e^{-\lambda(p+t)} = \left[\frac{q(p+t)}{p(q-t)}\right]^{p+t}$$

$$(pe^{\lambda} + q)e^{-\lambda(p+t)} = \frac{q}{q-t} \; \frac{[p(q-t)]^{p+t}}{[q(p+t)]^{p+t}}.$$

$$= \left(\frac{p}{p+t}\right)^{p+t} \; \left(\frac{q}{q-t}\right)^{q-t}$$

$$= \exp\{-(p+t)\log\frac{p+t}{p} - (q-t)\log\frac{q-t}{q}\}$$

$$= e^{-H}$$

$$(pe^{\lambda} + q)^n e^{-n\lambda(p+t)} = e^{-nH}.$$

This is a glimpse of how ENTROPY enters!

$$H = [(p+t)\log\frac{p+t}{p} + (q-t)\log\frac{q-t}{q}]$$

is Entropy of probability $\{p+t, q-t\}$ w.r.t. $\{p, q\}$.
It is positive. More later.

$$P\left(\frac{S_n}{n} - p > t\right) \leq e^{-nH} \to 0$$

with a similar inequality leading to SLLN.

($\mathfrak{m}$) Load balancing:

n jobs and m processors.

Each job is allotted at random to one of the m processors. How balanced is the load. For Example take $n = m \log m$. (not integer! Do not worry, can make precise)

On the average each processor gets $\log m$ jobs. What are the chances that load of some processor exceeds $2 \log m$.

Fix ONE processor. Number of jobs allotted to this processor. is sum of n Bernoulli variables; $p = 1/m$. can show $P(S > 2 \log m) \leq \exp\{-m \log 2\} = 1/m^2$. Use union bound,

$$P(\text{load of at least one processor exceeds } 2 \log m) \leq 1/m$$

$(\mathfrak{n})$:

(Digression) unequal success probabilities.
$X_1, \cdots, X_n$ Bernoulli $1-0$ prob:$p_1, \cdots, p_n$. $X = \sum X_i$

$$E(e^{\lambda X}) = \prod (p_i e^{\lambda} + q_i) \leq (pe^{\lambda} + q)^n$$

$$p = \frac{1}{n} \sum p_i; \quad q = \frac{1}{n} \sum q_i$$

AM-GM inequality.

Sums but only of independent RV?
Independent RV but only sums?
($\mathfrak{a}$) AZUMA HOEFFDING:
$d_1, \cdots, d_n$ bounded RV; $E(d_i) = 0$
Expectation of distinct product is zero.

$$E(\prod_1^k d_{i_k}) = 0, \;\; 1 \leq d_{i_1} < \cdots < d_{i_k} \leq n$$

CONCLUSION: For $t > 0$

$$P(|\sum d_i| > t) \leq 2e^{-t^2/(2\sum ||d_i||^2)}.$$

Martingale differences are good examples. ($\|d_i\|$ is its bound)

Let $|d_i| \le c_i$ a.e. Fix $\lambda > 0$. Note $e^{\lambda x}$ convex in x

$$e^{\lambda x} \le \frac{e^{\lambda c_i} + e^{-\lambda c_i}}{2} + \frac{e^{\lambda c_i} - e^{-\lambda c_i}}{2}\,\frac{x}{c_i}; \quad -c_i \le x \le c_i$$

$$e^{\lambda d_i} \le \frac{e^{\lambda c_i} + e^{-\lambda c_i}}{2} + \frac{e^{\lambda c_i} - e^{-\lambda c_i}}{2}\,\frac{d_i}{c_i}.$$

$$E\left(\prod_1^m [\alpha_i d_i + \beta_i]\right) = \prod_1^m \beta_i.$$

$$E(\prod e^{\lambda d_i}) \le \prod \frac{e^{\lambda c_i} + e^{-\lambda c_i}}{2} \le e^{\lambda^2(\sum c_i^2/2)}.$$

$$P(\sum d_i > t) \le e^{\lambda^2(\sum c_i^2/2)} e^{-t\lambda}$$

Minimized at $\lambda = t/\sum c_i^2$.

($\mathfrak{b}$) Shamir-Spencer:

$G(n,p)$ Erdos-Renyi model $0 < p < 1$. $\mu_n = E(\chi)$.

$$P(|\chi(G) - \mu_n| > t\sqrt{n-1}) \leq 2e^{-t^2/2}$$

(μ_n is of the order $n/logn$)
n vertices: $\{1, 2, \cdots, n\}$.
I_{ij} one or zero edge ij present OR not.
$F_k = \{I_{ij} : 1 \leq i, j \leq k\}$ for $k = 2, 3, \cdots n$.
$X_1 = E(\chi) = \mu_n; \quad X_2 = E(\chi \| F_2) \cdots \quad X_n = E(X \| F_n) = \chi$
$d_i = X_{i+1} - X_i$ for $1 \leq i \leq n-1$.
Claim: $|d_i| \leq 1$
Since $\sum d_i = \chi - \mu_n$ we are done.

For want of a nail the shoe was lost
For want of a shoe the horse was lost
$\cdots$ $\cdots$ the kingdom was lost
And all for the want of a horseshoe nail

For want of a nail the shoe was lost
For want of a shoe the horse was lost
$\cdots$ $\cdots$ the kingdom was lost
And all for the want of a horseshoe nail

ignorant of a definition, computation was lost
ignorant of the computation, theorem was lost
ignorant of the theorem, beautiful application lost
And all for not knowing a little silly definition.

($\mathfrak{c}$) CONDITIONAL EXPECTATION

Have two random variables X and Y on same space.

X taking values $\{a_i\}$ and Y taking $\{b_j\}$

$P(X = a_i, Y = b_j) = p_{ij}$.

$P(X = a_i) = \sum_j p_{ij} = p_{i\bullet}$; $\qquad P(Y = b_j) = \sum_i p_{ij} = p_{\bullet j}$;

$E(X|Y)$ is a random variable. When Y takes the value b_j then this takes value $\sum_i a_i p_{ij}/p_{\bullet j}$
Conditional distribution of X given $Y = b_j$ is the following probability: value a_i with probability $p_{ij}/p_{\bullet j}$
Conditional expectation is nothing but expectation w.r.t. conditional distribution.
Similar: conditional expectation given hundred random variables.

X, Y have joint density $f(x, y)$

densities of X and Y are
$f(x, \bullet) = \int f(x, y)dy; \qquad f(\bullet, y) = \int f(x, y)dx$

$E(X|Y)$ is a random variable; when Y takes the value y then this takes the value $\int xf(x, y)dx/f(\bullet, y)$

Conditional distribution of X given $Y = y$ is the function $f_{X|Y}(x) = f(x, y)/f(\bullet, y)$ regarded as a function of x.

Conditional expectation is nothing but expectation w.r.t. conditional distribution.

ZEROS (frightening but harmless)

Similar: conditional expectation given hundred variables.

If ξ is a function of (X, Y) then $E(\xi \cdot Z \| X, Y) = \xi \cdot E(Z \| X)$.
smothing:: $E(Z \| X, Y) = W$ say $E(W \| X) = U$ say Then
$E(Z \| X) = U$ Straight forward verification.
In particular, $E[\ E(Y \| X)\ ] = E(Y)$. Effect of this for us is the following: Have $Z, X_1, \cdots, X_n$

$$d_i = E(Z \| X_1, \cdots, X_i) - E(Z \| X_1, \cdots X_{i-1})$$

Then this is a multiplicative family (H-Z hyp. holds). For example,

$$E(d_4 d_5) = E[\ E(d_4 d_5 \| X_1, \cdots, X_4)\ ]$$

inner thing $= d_4 E(d_5 \| X_1, \cdots, X_4)$
but $d_5 = E(Z \| X_1, \cdots, X_5) - E(Z \| X_1, \cdots, X_4)$, use smoothing.

($\mathfrak{d}$) Chvatal-Sankoff:

$X_i : i \geq 1; Y_i : i \geq 1$ i.i.d finite alphabet valued.

$$L_n = \max \left\{ k : \begin{array}{c} \exists 1 \leq i_1 < i_2 < \cdots < i_k \leq n; \\ \exists 1 \leq j_1 < j_2 < \cdots < j_k \leq n; \\ X_{i_1} = Y_{j_1}; X_{i_2} = Y_{j_2}, \cdots X_{i_k} = Y_{j_k}. \end{array} \right\}$$

Understanding DNA sequences/large programs. $a_n = E(L_n)$.

$$P(|L_n - a_n| \geq t) \leq 2e^{-t^2/8n}$$

$$d_k = E(L_n \| X_i, Y_i : i \leq k) - E(L_n \| X_i, Y_i : i \leq k-1).$$

$$\sum d_i = L_n - a_n; \qquad |d_i| \leq 2.$$

As in SLLN, L_n/n converges if a_n/n converges. YES, THEY DO.

($\mathfrak{e}$) Graphs again.

Cycle passing through all vertices is HAMILTONIAN CYCLE.

Deciding existence of such cycle is NP hard!

$G(n, 1/2)$ has Hamiltonian cycle with high probability.
Means: Probability of this, say p_n, converges to one.

(in fact, there is a polynomial algorithm to get the cycle!)
Shall not do but here is a step towards that.

$G(n, 1/2)$ is *TRACTABLE* with high probability.

Means THREE things:

(i) w.h.p. Every vertex has between $\frac{n}{2} - \frac{n}{50}$ and $\frac{n}{2} + \frac{n}{50}$ neighbors.

(ii) w.h.p. for every pair of vertices u, v;

$$\frac{3}{4}n - \frac{n}{50} \leq |N(u) \cup N(v)| \leq \frac{3}{4}n - \frac{n}{50}$$

(iii) w.h.p. For every triple u, v, w of vertices

$$\frac{7}{8}n - \frac{n}{50} \leq |N(u) \cup N(v) \cup N(w)| \leq \frac{7}{8}n - \frac{n}{50}$$

For example, for each pair u, v the number

$$|N(u) \cup N(v) - u - v|$$

is sum of $n-2$ independent Bernoulli;
1 w.p. $3/4$ and
0 w.p. $1/4$.
Azuma says (ii) fails with probability at most

$$2 \exp\left\{-\frac{(\frac{n}{50}-2)^2}{6(n-2)}\right\}.$$

Chances of (ii) failing for at least one pair is at most n^2 times earlier and goes to zero.
(Frieze and Bruce Reed)

Objection:
Typical input is a uniformly chosen random graph? Unrealistic.

Objection:
Typical input is a uniformly chosen random graph? Unrealistic.

Answer:
No more unrealistic than the belief that
studying the pathological examples
constructed in NP completeness
yields information about typical instances.

Also Helps in Understanding
WHAT IS IT THAT MAKES THE PROBLEM DIFFICULT?

Erdos and Wilson: w.h.p. $G(n, 1/2)$ graph has a unique vertex of max degree.

(f) HOEFFDING LEMMA:
$X_1, \cdots, X_n$ indep. $a_i \leq X_i \leq b_i$.

$$S = \sum (X_i - EX_i)$$

Then
(i) $P(S > t) \leq \exp\{-2t^2 / \sum (b_i - a_i)^2\}$ and
(ii) Variance $X_i \leq (b_i - a_i)^2/4$.

Just note $|X_i - (\frac{a_i + b_i}{2})| \leq (b_i - a_i)/2$

(i) is Azuma-Hoeffding . (ii) is immediate.

Can do much much better! Strengthen (ii) to give (i).

($\mathfrak{g}$) HOEFFDING LEMMA AGAIN:

$EY = 0$; $a \le Y \le b$; $\Psi(\lambda) = \log Ee^{\lambda Y}$

Then:

$$\Psi''(\lambda) \le (b-a)^2/4$$

Y is subgaussian variance parameter $(b-a)^2/4$.

$|Y - \frac{b+a}{2}| \le \frac{b-a}{2}$ variance$(Y) \le (b-a)^2/4$

$$\Psi''(\lambda) = e^{-\Psi(\lambda)}E(Y^2 e^{\lambda Y}) - e^{-2\Psi(\lambda)}(E[Ye^{\lambda Y}])^2.$$

P distribution of Y

Think Z with distribution $dQ = e^{-\Psi(\lambda)}e^{\lambda x}dP(x)$.

$$\Psi''(\lambda) = \mathrm{Var}(Z) \le (b-a)^2/4. \quad \forall\ \lambda$$

$$\Psi(\lambda) = \Psi(0) + \lambda\Psi'(0) + \frac{\lambda^2}{2}\Psi''(?) \le \lambda^2(b-a)^2/8.$$

Sum of INDEP SUBGAUSSIAN things is again so.

$$\Psi(\lambda) = \log Ee^{\lambda Y}$$

$$\Psi'(\lambda) = \frac{1}{Ee^{\lambda Y}} E(Ye^{\lambda Y}) = e^{-\Psi(\lambda)} E(Ye^{\lambda Y})$$

$$\Psi''(\lambda) = e^{-\Psi(\lambda)} E(Y^2 e^{\lambda Y}) - e^{-2\Psi(\lambda)} (E[Ye^{\lambda Y}])^2.$$

Y value y_i prob: p_i $i \geq 1$

Z value y_i prob: $e^{-\Psi(\lambda)} e^{\lambda y_i} p_i$ $i \geq 1$

$$E(Z^2) = e^{-\Psi(\lambda)} E(Y^2 e^{\lambda Y});\ (EZ)^2 = e^{-2\Psi(\lambda)} (E[Ye^{\lambda Y}])^2.$$

CONCENTRATION INEQUALITIES

JOINT TELEMATICS GROUP
IEEE INFORMATION THEORY SOCIETY
SUMMER SCHOOL

June 27-July 1, 2016.
IISC, Bangalore

bvrao@cmi.ac.in
DAY 3

($\mathfrak{h}$) YET ANOTHER LOOK at HOEFFDING.

we knew $\Psi''(\lambda) \le (b-a)^2/4$ So

$$\lambda\Psi'(\lambda) - \Psi(\lambda) = \int_0^\lambda \theta\Psi''(\theta)d\theta \le \lambda^2 v/2.$$

$v = (b-a)^2/4$

$$\frac{1}{\lambda}\Psi'(\lambda) - \frac{1}{\lambda^2}\Psi(\lambda) \le v/2$$

So $G(\lambda) = \Psi(\lambda)/\lambda$ satisfies $G'(\lambda) \le v/2$

Known $G \to 0$ as $\lambda \to 0$

$$G(\lambda) \le \lambda v/2; \quad \Psi(\lambda) \le \lambda^2 v/2.$$

Glimpse of Herbst argument!

(j) Better form of A-H popularized by McDIARMID

$f : \Omega^n \to R$ BOUNDED DIFFERENCE PROPERTY means
for each i; $1 \le i \le n$ there is a number c_i such that
if $a, b \in \Omega^n$ differ **only** i-th coordinate

$$|f(a) - f(b)| \le c_i$$

$\{X_i, i \le n\}$ indep RV values in Ω;

McDIARMID INEQUALITY.
f on Ω^n with Bounded Difference Property (with $\{c_i\}$)
$Z = f(X_1, \cdots, X_n)$. THEN

$$P(|Z - EZ| \ge t) \le 2e^{-2t^2/(\sum c_i^2)}.$$

$$d_i = E(Z \| X_1, \cdots, X_i) - E(Z \| X_1, \cdots, X_{i-1}); i \geq 1$$

$$\sum_1^n d_i = Z - E(Z); \quad E(d_i) = 0$$

Indeed $E(d_i \| X_j : j \leq i-1) = 0$

$$E(\prod_1^n e^{\lambda d_i})$$

condition on $\{X_i : i \leq n-1\}$

$$= E\left\{\prod_1^{n-1} e^{\lambda d_i} E(e^{\lambda d_n} \| X_i; i \leq n-1)\right\}$$

Use Hoeffding conditionally

$$E(e^{\lambda d_n} \| X_i; i \leq n-1) \leq e^{\lambda^2 c_n^2/8}$$

conditioned on $(X_i : i \leq n-1)|$ need to know: d_n has mean zero and lies in an interval of length c_n. Yes. THUS

$$E(\prod_1^n e^{\lambda d_i}) \leq e^{\lambda^2 c_n^2/8} E[\prod_1^{n-1} e^{\lambda d_i}]$$

Now condition on $\{X_i : i \leq n-2\}$ etc till you reach

$$E(\prod_1^n e^{\lambda d_i}) \leq e^{\sum \lambda^2 c_i^2/8}.$$

Conditionally on $X_1, \cdots, X_{i-1}$ the random variable d_i has mean zero noted already as consequence of smoothing.
Point I could not convince you in the lecture is:

Conditionally on $X_1, \cdots, X_{i-1}$ the random variable d_i takes values in an interval of length c_i :: ready for applying Hoeffding.

Setup: Product space Ω^n;
points denoted $x = (x_1, \cdots, x_n)$
product probability $P(x) = p_1(x_1) \cdots p_n(x_n)$ (independence)
X_i coordinate functions. $X = (X_1, \cdots, X_n)$;
$Z = f(X)$; Fix an i.
$d_i = E(Z \| X_1, \cdots, X_i) - E(Z \| X_1, \cdots, X_{i-1})$
Conditioned on $X_1, \cdots, X_{i-1}$ the second term in the above difference is a constant.

Need only show: conditioned on $X_1, \cdots, X_{i-1}$ the first term in the above difference takes values in an interval of length c_i.

Let $g(x_1, \cdots, x_i) = E(Z \| X_1 = x_1, , \cdots, X_i = x_i)$
(g is a function on the product space but depends on the first i coordinates.)

Only need to show that for fixed $x_1, \cdots, x_{i-1}$, as the variable x_i varies, the function g takes values in an interval of length c_i.
So fix $a_1, \cdots, a_{i-1}$. It suffices to show that for any a_i and a_i^* we have

$$g(a_1, \cdots, a_{i-1}, a_i) - g(a_1, \cdots, a_{i-1}, a_i^*) \leq c_i$$

Use the definition of conditional expectation,
remember the conditional distribution of $X_{i+1}, \cdots, X_n$ given $(X_1, \cdots, X_i)$ is just its usual product distribution by independence. So that

$$g(a_1, \cdots, a_{i-1}, a_i) =$$

$$\sum f(a_1, \cdots, a_{i-1}, a_i, u_{i+1}, \cdots, u_n) p_{i+1}(u_{i+1}) \cdots p_n(u_n)$$

and

$$g(a_1, \cdots, a_{i-1}, a_i^*) =$$

$$\sum f(a_1, \cdots, a_{i-1}, a_i^*, u_{i+1}, \cdots, u_n) p_{i+1}(u_{i+1}) \cdots p_n(u_n)$$

where sum is over all the u's in both the above.
Subtract and use hyp. on f. [without independence, the factor multiplying f in the two places may be different and may not be able to combine the two sums.]

($\mathfrak{k}$) random function

Pick a function g at random from the n^n functions of the set $\{1, 2, \cdots, n\}$ to itself.
$L(g)$ is the number of y such that $g(x) = y$ has no solution.
Complement of Range of g. Then:

$$P\left(\left|L(g) - \frac{n}{e}\right| > t\sqrt{n} + 1\right) \leq 2e^{-2t^2}.$$

Note, using indicators,

$$E(L) = n(1 - \frac{1}{n})^n$$

$$(1 - \frac{1}{n})^n \uparrow 1/e$$

so $E(L) \leq n/e$

$$\frac{L_n}{n-1} = \frac{n}{n-1}(1 - \frac{1}{n})^n = \frac{1}{(1 + \frac{1}{n-1})^{n-1}} \downarrow \frac{1}{e}$$

So $L(n) \geq (n-1)/e$. Think of L as a map from $\{1, 2, \cdots, n\}^n$ by identifying functions g as the point $(g(1), \cdots, g(n))$.

Bounded difference property with $c_i = 1$.
McDiarmid completes.

$$(1+\frac{1}{n})^n =$$

$$1+1+\frac{1}{2!}1.(1-\frac{1}{n})+\frac{1}{3!}1.(1-\frac{1}{n}).(1-\frac{1}{n})(1-\frac{2}{n})+\cdots$$

increases in n. Also for $x>1$;
$x\log(1-\frac{1}{x})$ has derivative $\log(1-\frac{1}{x})+\frac{1}{x-1}$

$$\log(1-\frac{1}{x}) = -\frac{1}{x}-\frac{1}{2x^2}-\frac{1}{3x^3}-\cdots$$

$$\geq -\frac{1}{x}-\frac{1}{x^2}-\frac{1}{x^3}-\cdots$$

$$= -\frac{1}{x-1}$$

positive derivative; so function increases in x.

(I) EFFRON-STEIN:

$X_1, \cdots X_n$ INDEP. $Z = f(X_1, X_2, \cdots, X_n)$

$E^{(i)}$ cond.exp. given all but X_i: $E^{(i)}Z = E(Z \| X_j : j \neq i)$. THEN

1. $\mathrm{Var}(Z) \leq \sum E[(Z - E^{(i)}Z)^2] = v$ (say)
2. Let $\{Y_i : i \leq n\}$ indep copy of X's and

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

$$\text{Then} \qquad v = \frac{1}{2} \sum E[(Z - Z_i')^2]$$

3. $v = \inf \sum E[(Z - \xi_i)^2]$

inf over all $\{\xi_i\}$ square integrable functions of $\{(X^{(i)})\}$.

1. $E[(Z - E^{(i)}(Z))^2] = EE^{(i)}[(Z - E^{(i)}(Z))^2]$

$$\text{Var }(Z) \leq E \text{ Var }^{(i)}(Z)$$

Total variation is smaller than average of 'local variations'

Understanding overall Fluctuations through Local fluctuations.

3. useful in calculations of v.

Set

$$Y_i = E_i Z - E_{i-1} Z, \quad E_i = E(\cdots \| X_j; j \leq i)$$

$$Z - EZ = \sum Y_i$$

$$E(Z - EZ)^2 = \sum EY_i^2$$

$$E_i[E^{(i)} Z] = E_{i-1} Z$$

$$Y_i = E_i[Z - E^{(i)} Z]$$

$$\text{(Jensen)} \qquad Y_i^2 \leq E_i[(Z - E^{(i)} Z)^2]$$

$$EY_i^2 \leq EE_i\{[Z - E^{(i)} Z]^2\} = E[(Z - E^{(i)} Z)^2]$$

$$\text{Var}(Z) \leq v.$$

For 2, observe X, Y i.i.d. then $var(X) = E(X - Y)^2/2$

If $\{Y_i\}$ independent copy of $\{X_i\}$ then

GIVEN $X^{(i)} = \{X_j : j \neq i\}$; Z_i' is independent of Z So

$$E^{(i)}[Z - E^{(i)}Z)^2] = var^{(i)}(Z) = E^{(i)}(Z - Z_i')^2/2$$

$$E[Z - E^{(i)}Z)^2] = EE^{(i)}[Z - E^{(i)}Z)^2/2$$

$$= EE^{(i)}(Z - Z_i')^2/2 \quad = E[(Z - Z_i')^2]/2$$

For 3; use $var(X) = \inf\{E(X - a)^2 : a \in R\}$.

$(\mathfrak{m})$
If f on Ω^n has bounded difference property with $\{c_i\}$ then
$var(Z) \leq \sum c_i^2/4 \qquad Z = f(X_1, \cdots, X_n)$
(Assumed $\{X_i\}$ independent)
To see this, Put

$$\xi_i = \frac{1}{2}[\sup_a f(X_1, \cdots, X_{i-1}, a, X_{i+1}, \cdots X_n)$$

$$+ \inf_b f(X_1, \cdots, X_{i-1}, b, X_{i+1}, \cdots X_n)]$$

then $(Z - \xi_i)^2 \leq c_i^2/4$.

Use part (3) of E-S.

$(\mathfrak{n})$
Binpacking:
$X_1, \cdots, X_n$ uniformly picked from $[0, 1]$.
$Z = f(X_1, \cdots, X_n)$ is the minimum number of bins of size one needed to pack them.

$\mathrm{Var}(Z) \leq n/4$.

Longest 'matching' subsequence
$Z = f(X_1, \cdots, X_n, Y_1, \cdots, Y_n)$

$\mathrm{Variance}(Z) \leq 2n/4 \quad = n/2$.

$(\mathfrak{p})$

SELFBOUNDING FUNCTION:

$f : S^n \to [0, \infty)$

There exist $f_i : S^{n-1} \to [0, \infty)$ satisfying:

$$0 \leq f(x_1, \cdots, x_n) - f_i(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n) \leq 1$$

$$\sum_1^n [f(x_1, \cdots, x_n) - f_i(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n]^2 \leq f(x_1, \cdots, x_n)$$

Equivalently

$$0 \leq f(x) - f_i(x^{(i)}) \leq 1$$

$$\sum [f(x) - f_i(x^{(i)})]^2 \leq f(x).$$

For self bounding function

$$\text{Var } f(X) \leq E[f(X)]$$

To see this,
Use Effron-Stein (part 3 and 1) taking

$$\xi_i = f_i(X^{(i)})$$

($\mathfrak{q}$) $\uparrow$-subsequences.
$X_1, \cdots, X_n$ i.i.d. uniform $[0,1]$.

$f(x_1, \cdots, x_n)$ is length of the largest increasing subsequence of $(x_1, \cdots, x_n)$

$L = f(X_1, \cdots, X_n)$.

$$var(L) \leq E(L).$$

Because, $f_i : [0,1]^{n-1}$ length of largest increasing subsequence of this $(n-1)$-tuple will serve our purpose.

notation:

$$x = (x_1, \cdots, x_n)$$

$$x^{(i)} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$$

for all x and i;

$$0 \leq f(x) - f(x^{(i)}) \leq 1$$

for all x

$$\sum [f(x) - f(x^{(i)})]^2 \leq f(x)$$

$(\mathfrak{r})$ Vapnik-Cervonenkis dimension.
$X_1, \cdots, X_n$ independent S-valued with continuous distribution.
$\mathcal{A}$ is a collection of subsets of S.

For $x = (x_1, \cdots, x_n)$, distinct points of S define

$$tr(x) = \{A \cap \{x_1, \cdots, x_n\} : A \in \mathcal{A}\}$$

$$T(x) = |tr(x)|$$

x is shattered if $T(x) = 2^{|x|}$

VC dimension $D(x)$ is the size of the largest shattered subset of x.

$D = D(X_1, \cdots, X_n)$ This is almost surely well defined.

Then $var(D) \leq E(D)$ useful in learning theory.

($\mathfrak{s}$) Configuration functions:

Have families of subsets: $\Pi_1 \subset S$; $\Pi_2 \subset S^2$; etc $\Pi_n \subset S^n$.

This is Hereditary, that is, if $(x_1, \cdots, x_m) \in \Pi_m$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq m$ then $(x_{i_1}, \cdots, x_{i_k}) \in \Pi_k$.

If $x \in S^n$ then length of the largest subsequence of x with property Π is denoted $f(x)$.

$X_1, \cdots, X_n$ are independent S valued and $Z = f(X_1, \cdots, X_n)$ then

$$Var(Z) \leq E(Z).$$

(t) concentration for self bounding functions:

Setup: Product space S^n; product probability,

coordinate variables (X_i); self bounding function f

$$Z = f(X_1, \cdots, X_n)$$

$$P(Z \geq EZ + t) \leq \exp\left\{-\frac{t^2}{2EZ + 2(t/3)}\right\}; \qquad t > 0$$

$$P(Z \leq EZ - t) \leq \exp\left\{-\frac{t^2}{2EZ}\right\} \qquad 0 < t < EZ.$$

Proof is via Entropy techniques.

CONCENTRATION INEQUALITIES

JOINT TELEMATICS GROUP
IEEE INFORMATION THEORY SOCIETY
SUMMER SCHOOL

June 27-July 1, 2016.
IISC, Bangalore

bvrao@cmi.ac.in

DAY 4

($\mathfrak{u}$) Sourav Chatterjee:

X Random variable values in S.
$f : S \to R \quad E[f(X)] = 0$.
Assume:
(i) Exists RV X' such that (X, X') EXCHANGEABLE.
(ii) Exists $F : S^2 \to R$ such that

$$F(a, b) = -F(b, a); \quad E(F(X, X') \| X) = f(X).$$

Exchangeable means: (X, X') and (X', X) have the same distribution.
PLAN: Learn f using F.

$$v(x) = \frac{1}{2} E\{|[f(X) - f(X')]F(X, X')| \ \| X = x\}$$

THEOREM:
$h : S \to R$ and $E|h(X)F(X, X')| < \infty$. Then

$$E[h(X)f(X)] = \frac{1}{2}E\{[h(X) - h(X')]F(X, X')\}$$

$$\mathrm{Var} f(X) = \frac{1}{2}E\{[f(X) - f(X')]F(X, X')\}$$

Proof: condition on (f, F) says

$$E[h(X)f(X)] = E[h(X)F(X, X')]$$

Exchangeability says $= E[h(X')F(X', X)]$

hypothesis on F says $= E[-h(X')F(X, X')]$

So each equals half of sum.
For last equality take $h = f$. Note $Ef(X) = 0$. Done.

ASSUME:

$$E[e^{\lambda f(X)}|F(X, X')|] < \infty; \quad |v(x)| \leq c$$

THEN CONCLUSION:

$$Ee^{\lambda f(X)} \leq e^{c\lambda^2/2}; \quad P(|f(X)| > t) \leq 2e^{-t^2/2c}.$$

Proof: Put $m(\lambda) = Ee^{\lambda f(X)}$, Use earlier theorem.

$$m'(\lambda) = \frac{1}{2}E\{[e^{\lambda f(X)} - e^{\lambda f(X')}]F(X, X')\}$$

$$\left|\frac{e^x - e^y}{x - y}\right| = \int_0^1 e^{tx+(1-t)y} dt$$

$$\leq \int_0^1 te^x + (1-t)e^y \;\; dt = \frac{1}{2}[e^x + e^y]$$

So $|e^x - e^y| \leq \frac{1}{2}|e^x + e^y| \;\; |x - y|$

$$|m'(\lambda)| \leq \frac{|\lambda|}{4} E\left\{[e^{\lambda f(X)} + e^{\lambda f(X')}] \;\; |[f(X) - f(X')] \;\; F(X, X')|\right\}$$

$$= \frac{|\lambda|}{2} E[e^{\lambda f(X)} v(X) + e^{\lambda f(X')} v(X')]$$

$$= |\lambda| E\{e^{\lambda f(X)} v(X)\} \leq c|\lambda| m(\lambda).$$

For $\lambda > 0$,

$$\{\log m(\lambda)\}' \leq c\lambda; \quad m(0) = 1$$

$$\log m(\lambda) \leq c\lambda^2/2$$

The main reason for doing this is the following: I did not see applications of this in the computer science, coding, communication literature. Of course has several applications, some mentioned in Sourav's thesis.

Briefly leaf through one application.

($\mathfrak{v}$) CURIE-WEISS
$c \in R$; $\beta > 0$ fixed.

$$H(\sigma) = -\frac{1}{n}\sum_{i<j}\sigma_i\sigma_j - c\sum\sigma_i; \quad \sigma \in \{-1,+1\}^n$$

Probability space: set of configurations$= \{-1,+1\}^n$
Probability: $p(\sigma) \propto e^{-\beta H(\sigma)}$. Gibbs probability.

Magnetization: $m(\sigma) = \frac{1}{n}\sum\sigma_i$.
Belief: concentrated at solution of $x = \tanh(\beta x + \beta c)$.

THEOREM:

$$E[m - \tanh(\beta m + \beta c)]^2 \leq \frac{2 + 2\beta}{n} + \frac{\beta^2}{n^2}.$$

$$P\{|m - \tanh(\beta m + \beta c)| > \frac{\beta}{n} + t\} \leq 2e^{-nt^2/(4+4\beta)}.$$

Reason:
Shall produce (σ, σ') exchangeable pair.
σ according to Gibbs distribution.
Pick σ, choose a coordinate I at random
Calculate conditional distribution of I-th coord. Given all others
Replace I-th coordinate of σ according to this distribution.
This is σ'.

$$F(\sigma,\sigma') = \sum(\sigma_i - \sigma'_i); \quad m_i(\sigma) = \frac{1}{n}\sum_{j\neq i}\sigma_j.$$

$$E(\sigma_i \| \sigma_j : j \neq i) = \tanh(\beta m_i + \beta c)$$

$$f(\sigma) = E\{F(\sigma,\sigma')\|\sigma) = \frac{1}{n}\sum[\sigma_i - E(\sigma_i\|\sigma_j : j \neq i)]$$

$$= m(\sigma) - \frac{1}{n}\sum_{1}^{n} tanh(\beta m_i + \beta c).$$

$|F(\sigma,\sigma')| \leq 2$; σ and σ' differ in only one coordinate. $\tanh x$ is 1-Lip function.

$$|f(\sigma) - f(\sigma')| \leq |m(\sigma) - m(\sigma')| + \frac{\beta}{n}\sum|m_i(\sigma) - m_i(\sigma')|$$

$$\leq \frac{2(1+\beta)}{n}$$

By earlier Theorems

$$\mathrm{Var}[m - \frac{1}{n}\sum_{1}^{n} \tanh(\beta m_i + \beta c)] \leq 2(1+\beta)/n$$

$$P\{|m - \frac{1}{n}\sum \tanh(\beta m_i + \beta c)| \geq t\} \leq 2e^{-nt^2/(4+4\beta)}.$$

Also by Lip nature of F,

$$|\tanh(\beta m_i + \beta c) - \tanh(\beta m + \beta c)| \leq \beta/n$$

Done.

($\mathfrak{a}$) Entropy

f non-negative function on a probability space.
Deal with finite sets. However makes sense in general with appropriate integrability conditions.

$$Ent(f) = E(f \log f) - E(f) \log E(f).$$

Compare

$$Var(f) = E(f^2) - (Ef)^2$$

Both are $E(\Phi(f)) - \Phi(Ef)$; in one case $\Phi(x) = x^2$ and in the other $\Phi(x) = x \log x$ (defined for $x \geq 0$)
For example Z is non-negative random variable $E(Z) = 1$ then

$$Ent(Z) = E(Z \log Z)$$

If on your space have a probability Q with $dQ/dP = Z$ then of course, usual divergence $D(Q|P)$ is just our $Ent(Z)$.

We consider $X_1, \cdots, X_n$ independent and $Z = f(X_1, \cdots, X_n)$
Loosely refer $Ent(f)$ for $Ent(Z) = Ent\{f(X_1, \cdots, X_n)\}$
Effron Stein:

$$var(Z) \leq \sum E[var^{(i)}(Z)]$$

Or

$$E(Z^2) - (EZ)^2 \leq \sum E[E^{(i)}(Z^2) - \{E^{(i)}(Z)\}^2].$$

Or if $\Phi(x) = x^2$ then

$$E(\Phi(Z)) - \Phi(EZ) \leq \sum E[E^{(i)}(\Phi(Z)) - \Phi\{E^{(i)}(Z)\}].$$

Suggests:

Theorem:(Subadditivity of entropy)

$$Ent(Z) \leq E \sum_{1}^{n} Ent^{(i)}(Z)$$

where

$$Ent^{(i)}(Z) = E^{(i)}(Z \log Z) - E^{(i)}(Z) \log E^{(i)}(Z).$$

Shall do modulo Han. First show Log Sobolev inequality for the cube

$\{-1,1\}^n$ uniform probability
coordinate variables $X_1, \cdots, X_n$ and $X = (X_1, \cdots X_n)$
f a real function on the cube.
Shall put

$$\epsilon(f) = \frac{1}{2} E[\sum \{f(X) - f(X'^i)\}^2]$$

X'^i is X with i-th coordinate replaced by a variable X'_i independent of everything you see and which has same distribution as X_i.
Thus takes values ± 1 with prob. $1/2$.
Observation: No need of independent copy.

$$\epsilon(f) = \frac{1}{4} E[\sum \{f(X) - f(X^{*i})\}^2]$$

X^{*i} is just X with i-th coordinate flipped.

First def:

$$\epsilon(f) = \frac{1}{2}\sum EE^{(i)}\{f(X) - f(X'^{i})\}^2$$

second def:

$$\epsilon(f) = \frac{1}{4}\sum EE^{(i)}\{f(X) - f(X^{*i})\}^2$$

Enough to show

$$E^{(i)}[\{f(X) - f(X^{*i})\}^2] = \frac{1}{2}E^{(i)}[\{f(X) - f(X^{*i})\}^2]$$

Fix one value of $X^{(i)}$ say $a^{(i)}$.

Then LHS equals

$$\frac{1}{4}[f(a^{(i)}, \pm 1) - f(a^{(i)}, \pm 1)]^2$$

$$\frac{1}{4}\{[f(a^{(i)}, +1) - f(a^{(i)}, -1)]^2 + [f(a^{(i)}, +1) - f(a^{(i)}, -1)]^2\}$$

$$= \frac{1}{2}[f(a^{(i)}, +1) - f(a^{(i)}, -1)]^2$$

$$= \frac{1}{2}E^{(i)}\{[\{f(X) - f(X^{*i})\}^2]\|X^{(i)} = a^{(i)}\}$$

Loosely, we are using $(a^{(i)}, 1)$ for the point with i-th coordinate one and others as in $a^{(i)}$.

LogSobolev for the cube

As above we have:

cube; uniform probability; X coordinate vector ; $Z = f(X)$

THEOREM:

$$Ent(Z^2) \leq 2\epsilon(Z); \quad \text{or} \quad Ent(f^2) \leq 2\epsilon(f)$$

Proof: Sub additivity of entropy says

$$Ent(Z^2) \leq E \sum Ent^{(i)}(Z^2)$$

Enough to show

$$Ent^{(i)}(Z^2) \leq \frac{1}{2} E^{(i)}[f(X) - f(X^{*i}]^2.$$

Given $X^{(i)}$ the RV Z takes two values a, b. So amounts to showing

$$\frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} \le \frac{1}{2}(a-b)^2$$

No loss to assume $0 < b < a$ use $(|a| - |b|)^2 \le (a-b)^2$
Fix b. Define on $[b, \infty)$

$$\varphi(a) = \frac{a^2}{2}\log a^2 + \frac{b^2}{2}\log b^2 - \frac{a^2+b^2}{2}\log\frac{a^2+b^2}{2} - \frac{1}{2}(a-b)^2$$

$$\varphi(b) = 0$$

$$\varphi'(a) = a \log \frac{2a^2}{a^2 + b^2} - (a - b)$$

$$\varphi'(b) = 0$$

$$\varphi''(a) = 1 + \log \frac{2a^2}{a^2 + b^2} - \frac{2a^2}{a^2 + b^2} \leq 0$$

($\log x \leq x - 1$) Enough to say $\varphi(a) \leq 0$ for all $a \geq b$.

special case: $f = I_A$. A subset of cube.

$$Ent(f^2) = -P(A)\log P(A)$$

$$4\epsilon(f) = Influence(A)$$

What is influence of A.?
Influence of i-th coordinate is
$Influence_i(A) = P[I_A(x) \neq I_A(x^{*i})]$
Total influence is sum of influences of all coordinates
$Influence(A) = \sum Influence_i(A)$
So

$$P(A)\log\frac{1}{P(A)} \leq Influence(A)/2$$

Before subadditivity, an observation:

$$Ent(Z) = E(Z \log Z) - (EZ)\log(EZ)$$

Let $c > 0$

$$Ent(cZ) = E[cZ\log(cZ)] - (EcZ)\log E(cZ)$$

$$= c\{E(Z)\log c + E(Z\log Z) - E(Z)\log(EZ) - (EZ)\log c\}$$
$$= cEnt(Z)$$

back to $Ent(Z) \leq E \sum_1^n Ent^{(i)}(Z)$ (discrete case)

If holds for Z then holds for cZ. So Assume $EZ = 1$

Need to fix up notation;

Ω^n is the space with product probability

points $x = (x_1, \cdots, x_n)$

$$p(x) = p_1(x_1)p_2(x_2) \cdots p_n(x_n)$$

(X_i are coordinate variables, etc)

Q is the probability $Q(x) = Z(x)p(x)$

then

$$Ent(Z) = D(Q\|P)$$

Recall

$$D(Q\|P) = \sum Q(x) \log[Q(x)/P(x)]$$
$$= E_P(Z \log Z)$$

By Han for relative entropy (P product measure; Q any)

$$D(Q\|P) \leq \sum [D(Q\|P) - D(Q^{(i)}\|P^{(i)})]$$

Shall show, to complete proof,

$$\sum [D(Q\|P) - D(Q^{(i)}\|P^{(i)})] = E \sum Ent^{(i)}(Z)$$

Recall, $Q^{(i)}$ is the marginal of Q on the space Ω^{n-1} which is product Ω^n with i-th coordinate space removed.
Or, marginal distribution of $X^{(i)}$ under Q.
Similarly $P^{(i)}$ is marginal distribution of $X^{(i)}$ under P

$$Q^{(i)}(u^{(i)}) = \sum_y Z(u^{(i)}, y)p(u^{(i)}, y)$$

$$= P^{(i)}(u^{(i)}) \sum_y Z(u^{(i)}, y)\frac{p(u^{(i)}, y)}{P^{(i)}(u^{(i)})}$$

$$= P^{(i)}(u^{(i)})E^{(i)}(Z) \text{ use def of } E^{(i)}(Z)$$

$$D(Q^{(i)} \| P^{(i)})$$

$$= \sum_{u^{(i)}} Q^{(i)}(u^{(i)}) \ \ \log[Q^{(i)}(u^{(i)})/P^{(i)}(u^{(i)})]$$

$$= \sum_{u^{(i)}} E^{(i)}(Z) P^{(i)}(u^{(i)}) \log E^{(i)}(Z) \ \text{ by above}$$

$$= \sum_{u^{(i)}} E^{(i)}(Z) \log E^{(i)}(Z) \ \ P^{(i)}(u^{(i)}) \ \text{ rearrange}$$

$$= E \ \ \{E^{(i)}(Z) \log E^{(i)}(Z)\}$$

by definition of expectation of function of $X^{(i)}$.

$$\text{Also} \qquad D(Q\|P) = E(Z\log Z) = E\ \ E^{(i)}(Z\log Z)$$

Subtract and see.
Han for relative entropy is Han for entropy plus algebra. Shall not do. Recall $H(X) = -\sum p(x)\log p(x)$ etc.
Here is Hahn for entropy: $X_1, \cdots, X_n$ random variables $X = (X_1, \cdots, X_n)$ and $X^{(i)}$ is X WITHOUT X_i.
Theorem:

$$H(X_1, \cdots, X_n) \le \frac{1}{n-1}\sum H(X^{(i)})$$

Proof:

$$H(X) = H(X^{(i)}) + H(X_i\|X^{(i)})$$
$$\le H(X^{(i)}) + H(X_i\|X_1, \cdots, X_{i-1})$$

Add over i etc.

CONCENTRATION INEQUALITIES

JOINT TELEMATICS GROUP
IEEE INFORMATION THEORY SOCIETY
SUMMER SCHOOL

June 27-July 1, 2016.
IISC, Bangalore

bvrao@cmi.ac.in
DAY 5

($\mathfrak{a}$) LogSobolev $\longrightarrow$ concentration: cube.

Cube; uniform probability, f a function. $Z = f(X)$
Put $g = e^{\lambda f(X)/2} = e^{\lambda Z/2}$

$$Ent(g^2) = Ent(e^{\lambda Z}) = \lambda E(Ze^{\lambda Z}) - Ee^{\lambda Z} \log Ee^{\lambda Z}$$

$$F(\lambda) = Ee^{\lambda Z}$$

$$Ent(g^2) = \lambda F'(\lambda) - F(\lambda) \log F(\lambda)$$

$$Ent(g^2) \leq \sum E[\{(e^{\lambda f(X)/2} - e^{\lambda f(X^{*i})/2})_+\}^2]$$

For $z > y$

$$e^{z/2} - e^{y/2} \leq \frac{(z-y)}{2} e^{z/2}$$

$$Ent(g^2) \leq \frac{\lambda^2}{4} E \sum [(f(X) - f(X^{*i}))_+^2 e^{\lambda f(X)}]$$

$$= \frac{\lambda^2}{4} E[e^{\lambda f(X)} \sum (f(X) - f(X^{*i}))_+^2]$$

$$v = \max \left\{ \sum_1^n [f(x) - f(x^{*i})]_+^2 : x \text{ in the cube} \right\}$$

$$Ent(e^{\lambda f}) \leq \frac{v\lambda^2}{4} E e^{\lambda f(X)}$$

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \leq \frac{v\lambda^2}{4} F(\lambda)$$

$$\frac{F'}{\lambda F} - \frac{\log F}{\lambda^2} \leq \frac{v}{4}$$

Set

$$G(\lambda) = \frac{\log F(\lambda)}{\lambda}$$

Then

$$G'(\lambda) \leq v/4; \qquad G(\lambda) \to EZ \text{ as } \lambda \to 0$$

Take $\lambda > 0$. Integrate from 0 to λ to see

$$G(\lambda) \leq EZ + \frac{v\lambda}{4}. \qquad F(\lambda) \leq e^{\lambda EZ + \frac{v\lambda^2}{4}}. \ \ \lambda > 0.$$

Thus

$$Ee^{\lambda(Z-EZ)} \leq e^{v\lambda^2/4}.$$

$$P(Z > EZ + t) \leq e^{-t^2/v}$$

Similarly (integrating from $-\lambda$ to 0 etc etc)

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

($\mathfrak{b}$) Talagrand:

Product space set up: S^n product probability P.
f bounded difference property; $c_1, c_2, \cdots, c_n > 0$
McDiarmid

$$P(|f(X) - Ef(X)| > t) \leq 2e^{-2t^2/\sum c_i^2}.$$

Keep in mind one sided inequality too.
In particular for Hamming distance and set A

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

because (one-sided) McD with $t = Ed_H(X, A)$; gives

$$P(Ed_H - d_H \geq t) \leq e^{-2t^2/n}$$

$$P(d_H(X, A) \leq 0) \leq e^{-2t^2/n}$$

$$P(A) \leq e^{-2t^2/n}; \quad Ed_H(X, A) \leq \sqrt{\frac{n}{2} \log \frac{1}{P(A).}}$$

The other side of McD is gives, for any $t > 0$;

$P(d_H > Ed_H + t) \leq e^{-2t^2/n}$ giving

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

if $P(A) \geq 1/2$; $\qquad P\{d_H \geq t + \sqrt{n(\log 2)/2}\} \leq e^{-2t^2/n}$

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

if $P(A) \geq 1/2$; $\qquad P\{d_H \geq t + \sqrt{n(\log 2)/2}\} \leq e^{-2t^2/n}$

For nearly almost all points
need to change only $\sqrt{n}$ (order) many coordinates
to bring the point inside A !!!

for general positive weights (c_i) and $d_c(x,y) = \sum c_i I_{(x_i \neq y_i)}$

$$P\left\{d_c(X,A) \geq t + \sqrt{\frac{||c||^2}{2}\log\frac{1}{P(A)}}\right\} \leq e^{-2t^2/||c||^2}.$$

If $||c|| = 1$ then $P\left\{d_c(X,A) \geq t + \sqrt{\frac{1}{2}\log\frac{1}{P(A)}}\right\} \leq e^{-2t^2}.(\spadesuit)$
Some algebra to put this in better form.

Put $u(A) = \sqrt{\frac{1}{2}\log\frac{1}{P(A)}}$

$$2u \geq t \quad \longrightarrow \quad P(A) \leq e^{-t^2/2}.$$

$$2u \leq t \quad \rightarrow \quad 2t - 2u \geq 2t - t = t \quad \rightarrow \quad 2(t-u)^2 \geq t^2/2.$$

So

$$P[d_c(X,A) \geq t] \leq e^{-t^2/2}$$

Use $t - u$ instead of t in $(\spadesuit)$

$$2u \geq t \to u \geq t/2 \to u^2 \geq t^2/4$$

$$\frac{1}{2} \log \frac{1}{P(A)} \geq t^2/4$$

$$\frac{1}{P(A)} \geq e^{t^2/2}$$

Either case

$$P(A)\; P[d_c(X,A) \geq t] \leq e^{-t^2/2}.$$

Talagrand: even if you increase the second set such a thing holds!

Talagrand's convex distance:

$$d_T(x,A) = \sup\{d_c(x,A) : ||c|| = 1; c \geq 0\}$$

Talagrand convex distance inequality: THEOREM

$$P(A)P(d_T(x,A) \geq t) \leq e^{-t^2/4}$$

Shall only make you believe $\leq e^{-t^2/10}$.
Need several ingradients.

First: $f : S^n \to [0, \infty)$; $a > 0$
Say: f is self bounding by factor of a or (a-self bounding) if

there are non-negative f_i on S^{n-1} such that for all $x \in S^n$

(i) $0 \leq f(x) - f_i(x^{(i)}) \leq 1$

(ii) $\sum [f(x) - f_i(x^{(i)})]^2 \leq af(x)$

Fact: for such f and $Z = f(X)$;

$$\log Ee^{\lambda(Z-EZ)} \leq \lambda^2 \frac{a(EZ)}{2 - a\lambda}; \qquad P(Z \leq EZ - t) \leq e^{\frac{t^2}{2aE(Z)}}$$

second ingradient: Take $f(x) = [d_T(x, A)]^2$.

Then f is (4)-self bounding; following functions witness.

$$f_i(x^{(i)}) = \inf\{f(x_1, \cdots, x_{i-1}, y, x_{i+1}, \cdots, x_n) : y \in S\}.$$

Accept!

Taking $Z = d_T^2(X, A)$ and $t = Ed_T^2$ above inequality tells

$$P(A) = (d_T^2 \leq Ed_T^2 - t) \leq e^{-t^2/(8Ed_T^2)} = e^{-Ed_T^2/8}$$

Taking $\lambda = 1/10$ we get

$$\log Ee^{d_T^2/10} \leq \frac{1}{40}Ed_T^2 + \frac{1}{10}Ed_T^2 = \frac{1}{8}Ed_T^2$$

So $\qquad Ee^{d_T^2/10} \leq e^{Ed_T^2/8}$

$$P(d_T \geq t) \leq Ee^{d_T^2/10}e^{-t^2/10} \leq e^{Ed_T^2/8}e^{-t^2/10}$$

$$P(A)P(d_T \geq t) \leq e^{-Ed_T^2/8}e^{Ed_T^2/8}e^{-t^2/10} = e^{-t^2/10}$$

Talagrand distance is NOT distance between points; it is not a metric on the space. It only defines distance between a point and a set A in a product space.

It is called convex distance due to the following reason: Let U be the set of vectors $\alpha = (\alpha_1, \cdots, \alpha_n) \in R^n$ with the following properties:
(i) for each i; α_i is either zero or one;
(ii) there is a $y \in \Omega^n$ such that $\{i : \alpha_i = 0\} \subset \{i : x_i = y_i\}$.

Convex hull of U be denoted V. Then distance of the origin from A, that is, $\inf\{\|v\| : v \in V\}$ is precisely $d_T(x, A)$.

($\mathfrak{c}$) Stochastic Travelling Salesman Given n distinct points

$$z = (z_1, \cdots, z_n)$$

of the unit square $[0,1] \times [0,1]$ at random Find tour (cycle) of least possible length. This length is denoted $L(z)$.

Theorem: There is a number $c > 0$ (not depending on n) such that

$$P\{|L - M(L)| > t\} \leq 2e^{-t^2/4c}$$

$M(L)$ is median of L.

For a random variable Z median

$$M(Z) = \sup\{t : P(Z \leq t) \leq 1/2\}$$

Several other ways of defining

Concentration around the median?

Shifted party from mean to median?

It does not matter

I was not clear about the constants, but tried to outline a deduction of the theorem above from Talagrand using Lipschitz map of $[0, 1]$ onto the square $[0, 1] \times [0, 1]$.

But see book of Dubhashi and Panconesi who outline a simpler argument.

Another example from Dubhashi and Panconesi
Product set up, have Ω^n product probability

Suppose $f : \Omega^n \to [0, \infty)$ and $r \geq 1$ integer.

Say f is r-certifiable if for each x there is a set $J(x) \subset \{1, 2, \cdots, n\}$ satisfying two conditions.

(i) $|J(x)| \leq rf(x)$
(ii) If y agrees with x on these coordinates then $f(y) \geq f(x)$

$J(x)$ is a certificate for $f(x)$.

Suppose f is r-certifiable. It is Hamming Lipschitz with constant c. That is changing one coordinate of x changes value of $f(x)$ by at most c.

Theorem:

$$P\{f > M(f) + t\} \leq 2\exp\left\{-\frac{t^2}{4c^2 r[M(f) + t]}\right\}$$

$$P\{f < M(f) - t\} \leq 2\exp\left\{-\frac{t^2}{4c^2 rM(f)}\right\}$$

isoperimetry:
Consider standard Gaussian probability P on R. For a given number $0 < c < 1$ which sets A with $P(A) = c$ have largest measure for their t-blowup ($= \{x : d(x, A) \leq t\}$)? Answer: half lines H with $P(H) = c$. Moreover as soon as $P(H) \geq 1/2$ the probability of the complement of blowup decreases rapidly.

Similar result holds for (normalized) surface area of a sphere and spherical caps instead of half spaces. Also as soon as the spherical cap has area more than half,then area of the complement of the blowup decreases rapidly.

Such results are at the heart of Talagrand inequality. Possibly this is the reason for calling it iso-perimetric inequality.

THANK YOU