

Two new probability inequalities and concentration results

Ravindran Kannan

September 2, 2010

Random Variables associated with Combinatorial Problems

Random Variables associated with Combinatorial Problems

- ▶ Minimum length tour visiting each of n i.i.d. points in the unit square. [Travelling Saleman Problem](#)

Random Variables associated with Combinatorial Problems

- ▶ Minimum length tour visiting each of n i.i.d. points in the unit square. [Travelling Saleman Problem](#)
- ▶ Minimum length of edges so that there is a path between each pair of n i.i.d. points in unit square. [Minimum Spanning Tree](#)

Random Variables associated with Combinatorial Problems

- ▶ Minimum length tour visiting each of n i.i.d. points in the unit square. [Travelling Saleman Problem](#)
- ▶ Minimum length of edges so that there is a path between each pair of n i.i.d. points in unit square. [Minimum Spanning Tree](#)
- ▶ Minimum number of capacity 1 bins into which we can pack n i.i.d. reals in $[0, 1]$. [Bin Packing](#)

Random Variables associated with Combinatorial Problems

- ▶ Minimum length tour visiting each of n i.i.d. points in the unit square. [Travelling Saleman Problem](#)
- ▶ Minimum length of edges so that there is a path between each pair of n i.i.d. points in unit square. [Minimum Spanning Tree](#)
- ▶ Minimum number of capacity 1 bins into which we can pack n i.i.d. reals in $[0, 1]$. [Bin Packing](#)
- ▶ Minimum number of colours to be assigned to vertices of a random graph [Graph Colouring](#) $G_{n,p}$ so that no two adjacent vertices get the same colour.

Random Variables associated with Combinatorial Problems

- ▶ Minimum length tour visiting each of n i.i.d. points in the unit square. [Travelling Saleman Problem](#)
- ▶ Minimum length of edges so that there is a path between each pair of n i.i.d. points in unit square. [Minimum Spanning Tree](#)
- ▶ Minimum number of capacity 1 bins into which we can pack n i.i.d. reals in $[0, 1]$. [Bin Packing](#)
- ▶ Minimum number of colours to be assigned to vertices of a random graph [Graph Colouring](#) $G_{n,p}$ so that no two adjacent vertices get the same colour.

[Concentration](#) Prove probability of deviation from mean is small.
Are tail probabilities upper bounded by normal density of same variance (“sub-Gaussian”?)

Study of such random variables started by Physicists -
Beardwood, Halton and Hammersley....(A more basic question we don't deal with: What is the expected value (in the limit?))



The standard tools for proving concentration can deal with:

- ▶ Uniform density in the unit square for TSP, MWST, other geometric problems – (essentially) Equivalently Poisson
- ▶ Homogeneous (all equal) edge probabilities in Random Graph $G(n, p)$.
- ▶ Gaussians for random vectors/projections
- ▶ i.i.d. uniform $[0, 1]$ for longest increasing sub-sequence.....

But modern data seems to have

- ▶ Heavy Tails
- ▶ Inhomogeneity (not i.i.d.)
- ▶ Not fully independent ...

A half-baked discussion of heavy-tails

Example: Communication/queueing Network: One extreme: n independent users or jobs. Could allow each user/job to have heavy-tailed processing time/number of packets, but still maintain independence among users. Then, say, the number of arrivals in any fixed period still has exponential (in n) tails.

Opposite extreme used in Algorithm Analysis : Worst-case : “Adversary” arranges arrivals/processing times (but still subject to global totals being right- eg. total of n jobs) so as to force the worst-case on the algorithm.

A more general set-up: Multiple, say, still, $O(n)$ adversaries. Each controlling, say, a “cogent part” of the Network. Adversaries are independent. Now we may not have exponential tails for “local variables” (number of packets which must pass through an edge.) Here: General Tool for these and also some more classical examples.

Hoeffding-Azuma inequality

- ▶ X_1, X_2, \dots, X_n real random variables with
 - ▶ $E(X_i | X_1, X_2, \dots, X_{i-1}) = 0$ **Martingale Differences**
 - ▶ $|X_i| \leq 1$ **ABSOLUTE BOUND**
- ▶ Then,

$$\Pr \left(\left| \sum_{i=1}^n X_i \right| \geq t \right) \leq \exp \left(-\frac{ct^2}{n} \right).$$

Tails of $\sum_{i=1}^n X_i$ are “at most” tails of Gaussian - $N(0, n)$. If the X_i were independent with variance 1 each, then Central Limit Theorem gives us such a result (at least in the limit). H-A yields similar result (but for constant c) even though X_i are not (necessarily) independent.

But penalty: Assumption of absolute bound.

Power-law and other heavy tails

But, even for the simple example, with **i.i.d.** X_i with

$$\Pr(X_i = s) = \frac{c}{s^{10}} \quad \text{for } s = 1, 2, \dots$$

H-A says nothing about concentration of $\sum X_i$.

Simple starting thought: In this example, 8 th moment of X_i exists (9 th does not), so can we bound the 8 th moment of $\sum X_i$?

What (minimal ?) conditions on X_i will allow us to conclude tails of $\sum_i X_i$ are sub-Gaussian as in H-A ?

Theorem 1

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.
 - ▶ (\Leftarrow **Martingale**)

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.
 - ▶ (\Leftarrow **Martingale**)
- ▶ **Bounds on Conditional Moments**
 $E(X_i^k | X_1 + X_2 + \dots X_{i-1}) \leq k! \left(\frac{n}{m}\right)^{(k/2)-1}$, even $k \leq m$.

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.
 - ▶ (\Leftarrow Martingale)
- ▶ **Bounds on Conditional Moments**
 $E(X_i^k | X_1 + X_2 + \dots X_{i-1}) \leq k! \left(\frac{n}{m}\right)^{(k/2)-1}$, even $k \leq m$.
 - ▶ ($\Leftarrow |X_i| \leq 1$.)

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.
 - ▶ (\Leftarrow Martingale)
- ▶ **Bounds on Conditional Moments**
 $E(X_i^k | X_1 + X_2 + \dots X_{i-1}) \leq k! \left(\frac{n}{m}\right)^{(k/2)-1}$, even $k \leq m$.
 - ▶ ($\Leftarrow |X_i| \leq 1$.)
- ▶ **Conclusion** $E(\sum_{i=1}^n X_i)^m \leq (100nm)^{m/2}$. (\sim tails of $N(0, n)$)

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.
 - ▶ (\Leftarrow Martingale)
- ▶ **Bounds on Conditional Moments**
 $E(X_i^k | X_1 + X_2 + \dots X_{i-1}) \leq k! \left(\frac{n}{m}\right)^{(k/2)-1}$, even $k \leq m$.
 - ▶ ($\Leftarrow |X_i| \leq 1$.)
- ▶ **Conclusion** $E(\sum_{i=1}^n X_i)^m \leq (100nm)^{m/2}$. (\sim tails of $N(0, n)$)
- ▶ H-A and Chernoff are very special cases.

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.
 - ▶ (\Leftarrow Martingale)
- ▶ **Bounds on Conditional Moments**
 $E(X_i^k | X_1 + X_2 + \dots X_{i-1}) \leq k! \left(\frac{n}{m}\right)^{(k/2)-1}$, even $k \leq m$.
 - ▶ ($\Leftarrow |X_i| \leq 1$.)
- ▶ **Conclusion** $E(\sum_{i=1}^n X_i)^m \leq (100nm)^{m/2}$. (\sim tails of $N(0, n)$)
- ▶ H-A and Chernoff are very special cases.
- ▶ **Adversary** “sets” previous variables; later variables can be integrated out.

Theorem 1

- ▶ X_1, X_2, \dots, X_n real r.v.'s. $m \leq n$ even positive integer.
- ▶ **Strong Negative Correlation** $EX_i(X_1 + X_2 + \dots X_{i-1})^k \leq 0$
odd $k < m$.
 - ▶ (\Leftarrow Martingale)
- ▶ **Bounds on Conditional Moments**
 $E(X_i^k | X_1 + X_2 + \dots X_{i-1}) \leq k! \left(\frac{n}{m}\right)^{(k/2)-1}$, even $k \leq m$.
 - ▶ ($\Leftarrow |X_i| \leq 1$.)
- ▶ **Conclusion** $E(\sum_{i=1}^n X_i)^m \leq (100nm)^{m/2}$. (\sim tails of $N(0, n)$)
- ▶ H-A and Chernoff are very special cases.
- ▶ **Adversary** “sets” previous variables; later variables can be integrated out.
- ▶ Hypothesis: Conditions on moments up to m th. Conclusion: Same m !!

Theorem 1 in Words

Theorem 1 in Words

- ▶ X_i negatively correlated with any **odd** power of $X_1 + X_2 + \dots X_{i-1}$. (For example, $X_i = \text{budget for period } i$ decreasing function of total budget for previous periods will do, as seen from a mild version of FKG inequality.)

Theorem 1 in Words

- ▶ X_i negatively correlated with any **odd** power of $X_1 + X_2 + \dots X_{i-1}$. (For example, $X_i = \text{budget for period } i$ decreasing function of total budget for previous periods will do, as seen from a mild version of FKG inequality.)
- ▶ Much Weaker Condition than :

Theorem 1 in Words

- ▶ X_i negatively correlated with any **odd** power of $X_1 + X_2 + \dots X_{i-1}$. (For example, $X_i = \text{budget for period } i$ decreasing function of total budget for previous periods will do, as seen from a mild version of FKG inequality.)
- ▶ Much Weaker Condition than :

$$E(X_i^k | X_1 + X_2 + \dots X_{i-1} = a) \leq 1 \text{ for all } a, k \leq m \text{ even. } (1)$$

(For any adversarial setting of previous variables, even moments of X_i are bounded by 1. (1) in turn weaker than $|X_i| \leq 1$.

Theorem 1 in Words

- ▶ X_i negatively correlated with any **odd** power of $X_1 + X_2 + \dots X_{i-1}$. (For example, X_i = budget for period i decreasing function of total budget for previous periods will do, as seen from a mild version of FKG inequality.)
- ▶ Much Weaker Condition than :

$$E(X_i^k | X_1 + X_2 + \dots X_{i-1} = a) \leq 1 \text{ for all } a, k \leq m \text{ even. } (1)$$

(For any adversarial setting of previous variables, even moments of X_i are bounded by 1. (1) in turn weaker than $|X_i| \leq 1$.

- ▶ **Conclusion** $\sum_{i=1}^n X_i$ has sub-Gaussian ($N(0, n)$) tails.

- ▶ Proof of Theorem 1 : Computer Sciency - Dynamic Programming like recursion.

- ▶ Proof of Theorem 1 : Computer Sciency - Dynamic Programming like recursion.
- ▶ Implies concentration for TSP, MWST, Johnson-Lindenstrauss under more general distributions (heavier tails, inhomogeneous) which is as strong as what was known for only uniform distributions.

- ▶ Proof of Theorem 1 : Computer Sciency - Dynamic Programming like recursion.
- ▶ Implies concentration for TSP, MWST, Johnson-Lindenstrauss under more general distributions (heavier tails, inhomogeneous) which is as strong as what was known for only uniform distributions.
- ▶ Also concentration for chromatic number of “inhomogeneous” random graphs (different edge probabilities)

- ▶ Proof of Theorem 1 : Computer Sciency - Dynamic Programming like recursion.
- ▶ Implies concentration for TSP, MWST, Johnson-Lindenstrauss under more general distributions (heavier tails, inhomogeneous) which is as strong as what was known for only uniform distributions.
- ▶ Also concentration for chromatic number of “inhomogeneous” random graphs (different edge probabilities)
- ▶ Theorem 2 to come later is generally much stronger, more complicated proof, but starting on the same lines as Theorem 1.

- ▶ Proof of Theorem 1 : Computer Sciency - Dynamic Programming like recursion.
- ▶ Implies concentration for TSP, MWST, Johnson-Lindenstrauss under more general distributions (heavier tails, inhomogeneous) which is as strong as what was known for only uniform distributions.
- ▶ Also concentration for chromatic number of “inhomogeneous” random graphs (different edge probabilities)
- ▶ Theorem 2 to come later is generally much stronger, more complicated proof, but starting on the same lines as Theorem 1.
- ▶ Using Theorem 2, will settle bin-packing by proving best possible concentration.

- ▶ Proof of Theorem 1 : Computer Sciency - Dynamic Programming like recursion.
- ▶ Implies concentration for TSP, MWST, Johnson-Lindenstrauss under more general distributions (heavier tails, inhomogeneous) which is as strong as what was known for only uniform distributions.
- ▶ Also concentration for chromatic number of “inhomogeneous” random graphs (different edge probabilities)
- ▶ Theorem 2 to come later is generally much stronger, more complicated proof, but starting on the same lines as Theorem 1.
- ▶ Using Theorem 2, will settle bin-packing by proving best possible concentration.
- ▶ Other Applications ?

Classical TSP

n iid, uniform points in the unit square. f length of TSP tour through the points. One of the earliest problems in Probabilistic Analysis. Known $Ef \in O(\sqrt{n})$. First introduced and studied by Hammersely and others.

Classical TSP

n iid, uniform points in the unit square. f length of TSP tour through the points. One of the earliest problems in Probabilistic Analysis. Known $Ef \in O(\sqrt{n})$. First introduced and studied by Hammersely and others.

After much effort, Rhee and Talagrand proved concentration of f in $O(1)$ size interval with Gaussian tails - ce^{-ct^2} . Talagrand's inequality gave a simpler proof. Steele simplified the proof further.

Summary: TSP tour - mean $\theta(\sqrt{n})$; sub-Gaussian tails with Var $O(1)$.

All proofs start with uniform is well approximated by Poisson process :

Classical TSP

n iid, uniform points in the unit square. f length of TSP tour through the points. One of the earliest problems in Probabilistic Analysis. Known $Ef \in O(\sqrt{n})$. First introduced and studied by Hammersely and others.

After much effort, Rhee and Talagrand proved concentration of f in $O(1)$ size interval with Gaussian tails - ce^{-ct^2} . Talagrand's inequality gave a simpler proof. Steele simplified the proof further.

Summary: TSP tour - mean $\theta(\sqrt{n})$; sub-Gaussian tails with Var $O(1)$.

All proofs start with uniform is well approximated by Poisson process :

- ▶ Divide unit square into n small squares (say), each of side $1/\sqrt{n}$.

Classical TSP

n iid, uniform points in the unit square. f length of TSP tour through the points. One of the earliest problems in Probabilistic Analysis. Known $Ef \in O(\sqrt{n})$. First introduced and studied by Hammersely and others.

After much effort, Rhee and Talagrand proved concentration of f in $O(1)$ size interval with Gaussian tails - ce^{-ct^2} . Talagrand's inequality gave a simpler proof. Steele simplified the proof further.

Summary: TSP tour - mean $\theta(\sqrt{n})$; sub-Gaussian tails with Var $O(1)$.

All proofs start with uniform is well approximated by Poisson process :

- ▶ Divide unit square into n small squares (say), each of side $1/\sqrt{n}$.
- ▶ Y_i is a Poisson of intensity 1 (=Area of square times n) in the i th square. So, $E|Y_i|^k \leq k!$.

Our Theorem

In words : Unit Square divided into n small squares, each of side $1/\sqrt{n}$. Y_i set of points in i th square. Y_i independent. [Not i.i.d., internal correlations allowed.] With (roughly)

$$E|Y_i|^k < k^{2k}, \quad k \text{ even}, k \leq m$$

(instead of $E|Y_i|^k \leq k^k$ for all k)

we again get sub-Gaussian with Var. $O(1)$.

Our Theorem

In words : Unit Square divided into n small squares, each of side $1/\sqrt{n}$. Y_i set of points in i th square. Y_i independent. [Not i.i.d., internal correlations allowed.] With (roughly)

$$E|Y_i|^k < k^{2k}, \quad k \text{ even}, k \leq m$$

(instead of $E|Y_i|^k \leq k^k$ for all k)

we again get sub-Gaussian with Var. $O(1)$.

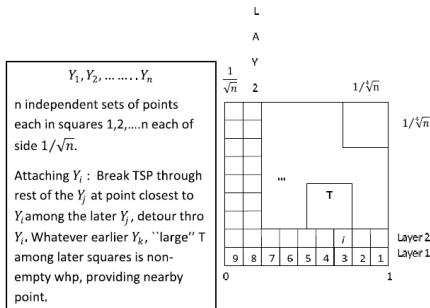
Theorem Let Y_1, Y_2, \dots, Y_n be independent sets of points generated in each small square respectively such that for a fixed constant $c \in (0, 1)$, an even positive integer $m \leq n$, and an $\epsilon > 0$, we have for $1 \leq i \leq n$ and $1 \leq l \leq m/2$,

$$\Pr(|Y_i| = 0) \leq c \quad ; \quad E|Y_i|^l \leq (O(l))^{(2-\epsilon)l}.$$

Suppose $f = f(Y_1, Y_2, \dots, Y_n)$ is the length of the shortest Hamilton tour through $Y_1 \cup Y_2 \cup \dots \cup Y_n$. We have

$$E(f - Ef)^m \leq (cm)^{m/2} \left(\implies e^{-ct^2} \dots \right).$$

TSP-contd.



Δ_i = Increase in TSP on adding Y_i to rest. Can break tour point closest to Y_i , detour thro Y_i . Whp, $T \subseteq$ later square: point (no conditioning!!). $TSP(Y_i) \leq c\sqrt{|Y_i|/\sqrt{n}}$. So, mo of Δ_i reduced to moments of $|Y_i|$.

Minimum conditions for $O(1)$ variance, higher moments

Stepping back, traditional results do prove that variance of tour length is $O(1)$ for n uniform random points in the unit square. Several such results - saying - variance is bounded even as the number of points goes to infinity. Rough high level reason: to each point, there are two other points within distance $O(1/\sqrt{n})$, so **sum of squared incremental costs** is only $O(1)$... Much empirical evidence ([Applegate, Johnson](#)) too.

But, we are making very strong assumptions - n i.i.d points...

[Aditya Bhaskar, K.](#) Mildest assumptions under which such results can be proved ?

Suppose Y_i are independent; $E|Y_i| = 1 \forall i$ and

$\Pr(|Y_i| = 0) < c < 1$. Then, the variance of the tour length is $O(1)$. Also if $E|Y_i|^l \leq l^{2l}$ for l up to $m/2$, then $E(f - Ef)^m$ bounded by moments of $N(0, 1)$. Other geometric problems.

Minimum weight spanning tree

Exactly Analogous Result. Complication : Not Monotonic. I.e., Δ_i not always positive.

Chromatic Number χ of Inhomogeneous random Graphs

First, traditional Random Graph $G_{n,p}$:

The MAXIMUM change in the chromatic number on adding one vertex (and all its edges) is 1. This + Hoeffding-Azuma easily implies sub-Gaussian tails with variance $O(n)$.

$$\text{Prob}(|\chi - E\chi| \geq t) \leq e^{-ct^2/n}.$$

But now, add groups of $1/p$ vertices at a time. Internal to the group, each vertex is adjacent to $O(1)$ vertices (in expectation); if so, $O(1)$ new colors suffice. HIGH MOMENTS OF additional number of colors can also be shown to be so bounded.... Implies

$$\text{Prob}(|\chi - E\chi| \geq t) \leq e^{-ct^2/(np)}.$$

For traditional $G(n, p)$, much tighter concentration (± 1 - Shamir, Spencer; Bollobas; Frieze; Luczak; Alon, Krivelevich; Achlioptas, Naor) known for almost surely.

Focus has been on tighter (almost sure) concentration, not necessarily exponential tails.

Different Model - $G(n, \text{capital } P)$

Random graph with edge probabilities $P = \{p_{ij}\}$ - inhomogeneous.

Average edge probability $p = \frac{\sum_{ij} p_{ij}}{\binom{n}{2}}$. We prove:

Theorem For $t \leq n\sqrt{p}$, $\text{Prob}(|\chi - E\chi| \geq t) \leq e^{-c^* t^2 / (n\sqrt{p})}$.

Question: Better results?

Random Projections

Johnson-Lindenstrauss (JL) Theorem : For v u.a.r. unit n -vector.
 $k \leq n$. $\epsilon \in (0, 1)$.

$$\Pr\left(\left|\sum_{i=1}^k v_i^2 - \frac{k}{n}\right| \geq \epsilon \frac{k}{n}\right) \leq c_1 e^{-c_2 k \epsilon^2}.$$

Proofs exploit uniform density or equivalently Gaussian. Here same result for more general distributions (long-tailed, inhomogeneous).

Theorem Suppose $Y = (Y_1, Y_2, \dots, Y_n)$ is a random vector picked from a distribution such that (for a $k \leq n$) (i)

$E(Y_i^2 | Y_1^2 + Y_2^2 + \dots + Y_{i-1}^2)$ is a non-increasing function of $Y_1^2 + Y_2^2 + \dots + Y_{i-1}^2$ for $i = 1, 2, \dots, k$ and (ii) for even $l \leq k$,

$E(Y_i^l | Y_1^2 + Y_2^2 + \dots + Y_{i-1}^2) \leq (cl)^{l/2} / n^{l/2}$. Then for any even integer $m \leq k$, we have

$$E\left(\sum_{i=1}^k (Y_i^2 - EY_i^2)\right)^m \leq (cmk)^{m/2} / n^m.$$

First use of Strong Negative Correlation instead of Martingale condition

Again : Higher the k in Hypothesis, higher the k in conclusion.

Longest Increasing Sequence (LIS)

Y_1, Y_2, \dots, Y_n i.i.d. uniform on $[0, 1]$. $f(Y_1, Y_2, \dots, Y_n)$ = length of the longest increasing sequence. Much study of f . $Ef \approx \sqrt{2n}$.

Concentration: In interval of length $O(n^{1/3})$: **Frieze; Bollobás and Brightwell. Talagrand** Interval of length $O(n^{1/4})$ by a simple proof.

Here, same. [By now better results are known, but using very specialized (beautiful) techniques.]

Δ_i = increase in f caused by adding Y_i to

$Y_1, Y_2, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n$. 0 or 1.

While Δ_i may be 1 in the worst-case (when we only get concentration in interval of length $O(\sqrt{n})$), one can simply show that $E\Delta_i \leq \frac{c}{\sqrt{n-i+1}}$ with adversarial choice of Y_1, Y_2, \dots, Y_{i-1} .

So, $E(\Delta_i^k | Y_1, Y_2, \dots, Y_{i-1}) \leq \frac{1}{\sqrt{n-i+1}}$ for all $k \geq 2$.

Question Can this be improved - i.e., to $O(n^{0.25-\epsilon})$. Will be interesting to do from general principles like here.

Random Matrices

Open Question Upper bounds on eigen/singular values of random matrices.

Wigner; Füredi, Komlos; Vu: A symmetric matrix with random above-diagonal entries, each with mean 0, variance σ^2 and absolutely bounded by 1. Then largest eigen-value of A is whp $O(\text{length of one row of } A)$. **Almost no correlation between rows.**

For consumer-product OR document-term matrices, cannot assume total independence of all entries. May however assume rows (consumers, documents) are independent vector-valued random variables.

Dasgupta, Hopcroft, Kannan, Mitra based on a Functional Analysis results of Rudelson and Lust-Picard: If A has independent vector-valued rows of mean 0, maximum variance σ^2 , length $O(\sqrt{n}\sigma)$, then maximum singular value of A is at most $O(\sqrt{n \ln n} \sigma)$. Clustering applications.

Essentially Wigner with log factors. Can this be derived (without all the Functional Analysis) using results here ?? Can this be

Question

TSP : what are the heaviest tails for which such concentration can be proved ??

Question Do poly time algorithms work for these heavy-tailed distributions ?

Question Other Geometric problems.

Question Configuration model for a random **regular** graph : A random Matching. Edge probability is a decreasing function of number of previous edges. Strong Negative Correlation (SNC)...
More generally : Degree Constraint graphs..

Question What is the chromatic number of $G(n, P)$? Is it within a log factor of $n(p + \sigma)$, where p, σ are mean, standard deviation of p_{ij} .

More Speculative

Heavy-tailed Queueing Theory Arrival and/or Service times may be heavy-tailed rather than Poisson in many modern settings.

Queueing Theorists have already studied this, but generally in the steady-state. In Networking and other applications, would be nice to get confidence guarantees that always hold.

Contingency Tables Given $m, n, a_1, a_2, \dots, a_m, b_1, b_2, \dots, b_n$ consider all $m \times n$ arrays with non-negative integer entries with row sums a_1, a_2, \dots, a_m respy and column sums b_1, b_2, \dots, b_n . Important Statistics Problem : I have a particular table. Is it random ?

Diaconis and Efron have argued the correct precise question is : what percentage of tables have some parameter greater than my table ? Concentration results for some parameters (like the sum of entries in a subset of positions) would be useful in this context.

Contingency Tables have some negative correlations. Row and column sums being fixed should imply the greater some entries are the lesser others are...

Major Issue with Thm. 1: Typical Vs Worst-case conditional moments

Bin-Packing $Y_1, Y_2, \dots, Y_n \in [0, 1]$ i.i.d. (arbitrary distribution) – “items”. $f(Y_1, Y_2, \dots, Y_n)$ = Minimum number of capacity 1 bins needed to pack the items.

Δ_i = increase in the number of bins when we add Y_i alone to $Y_1, Y_2, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n$. Again pretend Δ_i form a Martingale difference sequence.

Best **absolute** bound : $\Delta_i \leq 1$ only yields concentration in an interval of length \sqrt{n} .

Want to deal with $EY_1 = \mu \ll 1$. Let $\text{Var}(Y_1) = \sigma^2$.

Bin-Packing - contd.

If $\mu = EY_1 \ll 1$, then “typically” Δ_i (the increase in number of bins needed because of adding Y_i) is not 1.... If we were filling the bins one item at a time, there will be an overflow into the next bin only μ th of the time !! Even then, not a whole new bin, but only μ fraction of it is used up !! So, the variance of Δ_i should only be at most

$$\sigma^2 + \mu\mu^2$$

where $\sigma^2 = \text{var}(Y_i)$ is for when we do not overflow into next bin and (very roughly) μ^2 is for when we do, but we do only with probability μ .

Earlier best Concentration in an interval of length $O(\sqrt{n}(\mu + \sigma))$ - Rhee, Talagrand.

Here For discrete distributions : concentration (with sub-Gaussian tails) in an interval - $O(\sqrt{n}(\sigma + \mu^{3/2}))$. Also proof that this is **best possible** for discrete distributions. One other important point about Main thm: **variance matters more than higher moments.**

Bin Packing Contd.

The actual proof uses a Linear Programming formulation with one variable per “bin type”. The LP is used to prove that Δ_i is not too high. Its dual is used to show (by a more complicated argument) that Δ_i is not too low. Together, one gets a bound on variance... Stochastic bin-Packing is well-studied in CS ([Coffman](#); [Kenyon](#), [Sinclair](#)) because of applications. Of special interest in Concentration because the number of bins should be close to the sum of fractions, so the question had been : Is it as concentrated as the sum of independents ?

Different type of example - Polynomial r.v.

$f_s(G)$ is the number of s cliques in a graph G . Much study of the random variable $f = f(G_{n,p})$ for **FIXED** s : asymptotically normal **Rucinski** . For p small, difficult to get concentration results (by Azuma and even Talagrand) - because of small Ef . **Kim and Vu**; **Janson, Rucinski** get good (sub-Gaussian) tail bounds on

$$\Pr(f \geq 2Ef).$$

Kim and Vu More generally, concentration inequalities for random variables which are **POLYNOMIAL** functions of independent random variables. For this f , degree of polynomial is $\binom{s}{2}$. For fixed s , many results giving best tail bounds for **very large** deviations – $O(Ef)$. We get exponential tail bounds for $s = 3$, but for “small” deviations of the order of S.D.

Also we get : first results for s almost upto $O(\log n)$. [One reason for this large s : Max clique size in $G_{n,p}$ with p constant is $O(\log n)$.] $\log n$ degree too high for KV theorems.

Number of triangles in $G_{n,p}$

$$f(Y) = \sum_{k < j < i} Y_{ij} Y_{jk} Y_{ik}.$$

Easy : $\text{Var } f = O(n^3 p^3 + n^4 p^5)$.

Result Here : For deviations up to $(np)^{7/4}$, sub-Gaussian Tails (with correct variance - up to a constant.)

Remark : Beyond $(np)^{9/4}$ tails are provably not sub-Gaussian.

Remark: The case of “large” p ($p \geq 1/\sqrt{n}$) is easy and earlier result of Kim, VU; Janson and Rucinski prove sub-Gaussian tails.

Here, the more difficult $p \leq 1/\sqrt{n}$ is also tackled.

Open: Sub-Gaussian Tails for number of copies of other graphs.

Main Theorem

X_1, X_2, \dots, X_n satisfy strong negative correlation. Typical conditional moments : L_{im} . Worst-case conditional moments : M_{im} and Probability of being “atypical” : δ_i .

$$E \left(\sum_{i=1}^n X_i \right)^p \leq (cp)^{p/2} \left(\sum_{m=1}^{p/2} \frac{p^{1-(1/m)}}{m^2} \left(\sum_{i=1}^n L_{i,2m} \right)^{1/m} \right)^{p/2} \\ + (36p)^{p+2} \sum_{m=1}^{p/2} \frac{1}{n} \sum_{i=1}^n (nM_{i,2m})^{p/2m} \delta_i^{1/m}.$$

The $p/2$ is crucial to get sub-Gaussian bounds (obtained by setting p to the optimal value). If $\delta_i = 0$ and $L_{i,2m} \leq L_{2m}$, we get

$$(cp)^{p/2} \left[nL_2 + \sqrt{np}L_4^{1/2} + n^{1/3}p^{2/3}L_6^{1/3} + \dots \right]^{p/2} \approx (cpnL_2)^{p/2},$$

if higher moments do not grow too fast. Loosely : **sub-Gaussian tails with variance nL_2 as if X_i are independent.**

Martingale condition replaced by negative “correlation”

As a by-product (of the proof), main theorem replaces Martingale condition $E(X_i | X_1, X_2, \dots, X_{i-1}) = 0$ by a **strong negative correlation** condition :

$$EX_i(X_1 + X_2 + \dots + X_{i-1})^m \leq 0, \forall \text{ odd } m.$$

Examples : “Negatively associated” variables.

- ▶ Many “occupancy variables” - X_i number of balls in bin i .
- ▶ Uniform random $y = (y_1, y_2, \dots, y_n)$ subject to $|y| = 1$.
Random projections. [Condition on Aggregates – Degree-Constrained random graphs.]
- ▶ A randomized rounding scheme based on looking at pairs of variables. (Srinivasan)
- ▶ (A vague example) X_i : Deficit spending in period i . Penalty for over-spending in previous periods : X_i has mean $-\alpha(X_1 + X_2 + \dots + X_{i-1})$, for some $\alpha > 0$.

Other Questions

: Bound on the largest singular value of a random matrix.

Concentration of eigen-values of random matrices, [Alon](#), [Krivelevich](#), [Vu](#); functions of eigenvalues...

First-Passage Percolation (i.e., shortest path lengths in grid with i.i.d. edge lengths) : Concentration for normal, other edge lengths..