A new Probability inequality using typical moments and Concentration Results

Ravindran Kannan
Microsoft Research Labs., India; kannan@microsoft.com

***Abstract*—We present two probability inequalities. The simpler first inequality weakens both hypotheses in Höffding-Azuma inequality. Using it, we generalize concentration results previously known for the uniform density for the TSP, MWST and Random Projections to long-tailed inhomogeneous distributions. The second more complicated inequality further weakens the moment requirements and using it, we prove the best possible concentration for the long-studied bin packing problem as well as some others.**

***Index Terms*—Probability Inequality; Concentration; Long-Tailed distributions**

I. Introduction

Probabilistic Analysis of Algorithms is now a decades-old subject with many beautiful results. Generally, one assumes a simple probability distribution on the input (often uniform, Poisson distribution - see TSP section below, Bernoulli trials or Gaussians) and proves results on the expected value and concentration about this value. With modern massive data problems, worst-case analysis may lead to prohibitively high running times and so probabilistic analysis has emerged again as an important approach. But, it has been realized that the well-studied simple distributions do not necessarily model data well. Some empirically observed features from data point to characteristics not shared by simple distributions - long tails, inhomogeneity (i.e., variables not being i.i.d.), lack of independence etc.. We provide here a basic tool which will be of use in developing concentration results for these new problems, as well as for existing ones.

For real-valued random variables $X_1, X_2, \ldots X_n$ satisfying absolute bounds and the Martingale (difference) condition:

$$|X_i| \leq 1 \quad ; \quad E(X_i | X_1, X_2, \ldots X_{i-1}) = 0,$$

the widely used Höffding-Azuma (H-A) inequality asserts the following tail bound: $\Pr\left(|\sum_{i=1}^n X_i| \geq t\right) \leq c_1 e^{-c_2 t^2/n}$, for some constants c_1, c_2. The main aim of this paper is to weaken the assumption of an absolute bound, while retaining the essential strength of the conclusion. We present two theorems which do this, both upper bounding $E(\sum_{i=1}^n X_i)^m$ (the m th moment of $\sum_{i=1}^n X_i$) for some even integer m; from this, it is simple to get tail bounds by Markov inequality.

Our Theorem 1 is simply stated. But both H-A inequality and Chernoff bounds are very special cases of it. For the Traveling Salesperson Problem (TSP), earlier hard concentration results for the uniform density were made easy by Talagrand's celebrated inequality; using Theorem 1, we are able to prove as strong concentration, but, for more general densities allowing both longer tails and inhomogeneity. We do the same for the minimum weight spanning tree problem as well. We also consider random graphs where edge probabilities are not equal (inhomogeneity). We show a concentration result for the chromatic number (which has been well-studied under the traditional model with equal edge probabilities.) Theorem 1 also weakens the Martingale difference condition to a condition we call Strong Negative Correlation; this weakening has several uses too. A notable one is when we pick a random vector(s) of unit length as in the well-known Johnson-Lindenstrauss (JL) Theorem on Random Projections. Using Theorem 1, we prove a more general theorem than JL where longer-tailed distributions are allowed.

The absolute bound of H-A is weakened in Theorem 1 to bounds on (even) moments of X_i conditioned on (any) value of $X_1 + X_2 + \ldots + X_{i-1}$. A further weakening is obtained in our Main Theorem - Theorem (6) whose proof is more complicated. In Theorem (6), we use information on conditional moments of X_i conditioned on "typical values" of $X_1 + X_2 + \ldots + X_{i-1}$ as well as the "worst-case" values. This is very useful in many contexts as we show. Using Theorem 2, we settle the (discrete case of the) stochastic bin-packing problem studied by Rhee and Talagrand and others by proving concentration results which we show are best possible. We also give a proof of concentration for the longest increasing subsequence problem. In the full paper, we discuss applications to the number of triangles in sparse random graphs and several others.

II. Theorem 1

In theorem (1), we weaken the absolute bound $|X_i| \leq 1$ of H-A to something weaker than $E(X_i^l | X_1 + X_2 + \ldots + X_{i-1}) \leq 1$ for all even l upto a certain m. We prove a bound on the m (which is even) th moment of $\sum_{i=1}^n X_i$. Note the m is the same - the higher the moment bounded by the hypothesis, the higher the moment bounded by the conclusion. Under H-A's absolute bound assumption, the Theorem below also yields H-A's conclusion.

Theorem 1. *Let $X_1, X_2, \ldots X_n$ be real valued random variables and m an even positive integer satisfying the following for $i = 1, 2, \ldots n$:*

$$E X_i (X_1 + X_2 + \ldots X_{i-1})^l \leq 0 \;, l < m, odd. \tag{1}$$

$$E(X_i^l | X_1 + X_2 + \ldots + X_{i-1}) \leq \left(\frac{n}{m}\right)^{(l-2)/2} \; l! \;, l \leq m, \;\; even. \tag{2}$$

Then, we have

$$E\left(\sum_{i=1}^{n} X_i\right)^m \leq (24nm)^{m/2}.$$

Proof Let $M_l = \text{MAX}_{i=1}^n E(X_i^l | X_1, X_2, \ldots X_{i-1})$ for even $l \leq m$. For $1 \leq i \leq n$ and $q \in \{0, 2, 4, \ldots m-2, m\}$, define

$$f(i, q) = E\left(\sum_{j=1}^{i} X_j\right)^q.$$

Using the two assumptions, we derive the following recursive inequality for $f(n, m)$, which we will later solve (much as one does in a Dynamic Programming algorithm):

$$f(n,m) \leq f(n-1,m) + \frac{11}{5} \sum_{t \in \{2,4,6,\ldots m\}} \frac{m^t}{t!} M_t f(n-1, m-t), \tag{3}$$

Proof of (3):[1] Let $A = X_1 + X_2 + \ldots X_{n-1}$. Let $a_l = \frac{m^l}{l!} E|X_n|^l |A|^{m-l}$. Expanding $(A + X_n)^m$, we get

$$E(A + X_n)^m \leq EA^m + mEX_n A^{m-1} + \sum_{l=2}^{m} a_l. \tag{4}$$

Now, we note that $EX_n A^{m-1} \leq 0$ by hypothesis (1) and so the second term may be dropped. [In fact, this would be the only use of the Martingale difference condition if we had assumed it; we use SNC instead, since it clearly suffices.] We will next bound the "odd terms" in terms of the two even terms on the two sides using a simple "log-convexity" of moments argument. For odd $l \geq 3$, we have

$$E|X_n|^l |A|^{m-l} \leq (E(X_n^{l+1} A^{m-l-1}))^{1/2} (E(X_n^{l-1} A^{m-l+1}))^{1/2}$$

Also, $\frac{1}{l!} \leq \frac{6}{5} \frac{1}{\sqrt{(l+1)!}} \frac{1}{\sqrt{(l-1)!}}$

So, a_l is at most 6/5 times the geometric mean of a_{l+1} and a_{l-1} and hence is at most 6/5 times their arithmetic mean. Plugging this into (4), we get

$$E(\sum_{i=1}^{n} X_i)^m \leq EA^m + \frac{11}{5}(a_2 + a_4 + \ldots + a_m) \tag{5}$$

Now, we use the standard trick of "integrating over" X_n first and then over A (which is also crucial for proving H-A) to get for even l: $EX_n^l A^{m-l} = E_A\left(A^{m-l} E_{X_n}(X_n^l | A)\right) \leq M_l EA^{m-l}$ which yields (3).

We view (3) as a recursive inequality for $f(n,m)$. We will use this same inequality for the proof of the Main theorem, but there we use an inductive proof; here, instead, we will now "unravel" the recursion to solve it. [But first, note that the dropping the $EX_n A^{m-1}$ ensured that the coefficient of EA^m is 1 instead of the 11/5 we have in front of the other terms. This is important: if we had 11/5 instead, since the term does not reduce m, but only n, we would get a $(11/5)^n$ when we unwind the recursion. This is no good; we can get m terms in the exponent in the final result, but not n.]

[1] E will denote the expectation of the entire expression which follows.

Now to solve the recursive inequality, imagine a (recursion) tree (it is really a directed graph with possibly more than one path between a pair of nodes) whose root is marked $f(n,m)$. A node of the tree marked $f(i,q)$ (for $i \geq 2$, $0 \leq q \leq m$, even) has $(q/2)+1$ edges going from it to nodes marked $f(i-1,q), f(i-1,q-2), \ldots f(i-1,0)$; these edges have "weights" respectively $1, \frac{11}{5}\frac{q^2}{2!}M_2, \frac{11}{5}\frac{q^4}{4!}M_4, \ldots \frac{11}{5}\frac{q^q}{q!}M_q$ which are respectively at most

$$1, \frac{11}{5}\frac{m^2}{2!}M_2, \frac{11}{5}\frac{m^4}{4!}M_4, \ldots \frac{11}{5}\frac{m^q}{q!}M_q.$$

A node marked $f(1,q)$ has one child - a leaf marked $f(0,0)$ connected by an edge of weight M_q. Define the weight of a path from a node to a leaf as the product of the weights of the edges along the path. It is easy to show by induction on the depth of a node that $f(i,q)$ is the sum of weights of all paths from node marked $f(i,q)$ to a leaf.

Now, there is a 1-1 correspondence between paths from $f(n,m)$ to a leaf and elements of the following set : $L = \{(l_1, l_2, \ldots l_n) : l_i \geq 0, \text{ even } ; \sum_{i=1}^n l_i = m\}$; l_i indicates that at level i we take the l_i th edge - i.e., we go from node $f(i, m - l_n - l_{n-1} - \ldots l_{i+1})$ to $f(i-1, m - l_n - l_{n-1} - \ldots l_i)$ on this path. For an $l = (l_1, l_2, \ldots l_n) \in L$ and $t \in \{0, 2, 4, \ldots m\}$, define

$$g_t(l) = \text{ number of } i \text{ with } l_i = t .$$

Clearly, the vector $g(l) = (g_0(l), g_2(l), \ldots, g_m(l))$ belongs to the set

$$H = \{h = (h_0, h_2, h_4, \ldots h_m) : \sum_t t h_t = m; h_t \geq 0; \sum_t h_t = n\}.$$

Since the weight of an edge corresponding to l_i at any level is at most $(\frac{11}{5})^z M_{l_i} \frac{m^{l_i}}{l_i!}$, where $z = 1$ iff $l_i \geq 2$, and the number of non-zero l_i along any path is at most $m/2$, we have

$$f(n,m) \quad \leq \textstyle\sum_{l \in L} (\frac{11}{5})^{m/2} \prod_t M_t^{g_t(l)} \frac{m^{t g_t(l)}}{(t!)^{g_t(l)}}$$

For an $h \in H$, the number of $l \in L$ with $g_t(l) = h_t \forall t$ is the number of ways of picking subsets of the n variables of cardinalities $h_0, h_2, h_4, \ldots h_m$, namely,

$$\binom{n}{h_0, h_2, h_4, \ldots h_m} = \frac{n!}{h_0! h_2! h_4! \ldots h_m!} \leq \frac{n^{h_2 + h_4 + \ldots h_m}}{h_2! h_4! \ldots h_m!}.$$

Thus, we have (using the assumed upper bound on conditional moments)

$$f(n,m) \leq (\tfrac{11}{5})^{m/2} \textstyle\sum_{h \in H} \frac{n^{h_2+h_4+\ldots h_m}}{h_2! h_4! \ldots h_m!} \prod_t m^{t h_t} \frac{n^{h_t((t/2)-1)}}{m^{h_t((t/2)-1)}}$$

$$\leq (\tfrac{11}{5})^{m/2} \textstyle\sum_h (nm)^{\sum_t t h_t/2} \frac{m^{\sum h_t}}{h_2! h_4! \ldots h_m!} \leq (\frac{11}{5} nm)^{m/2} |H| e^{m/2},$$

since $\frac{m^{\sum h_t}}{h_2! h_4! \ldots h_m!}$ is easily seen to be maximized when $h_2 = p/2$ and the rest are 0 : $\frac{m^{h_t} m^{h_2}}{h_2! h_t!} \leq \frac{m^{h_2 + (t h_t/2)}}{(h_2 + (t h_t/2))!}$ implies that we may "transfer" h_t into h_2. Now, we bound $|H|$: each element of H corresponds to a unique $\frac{m}{2}$-vector $(h_2, 2h_4, 4h_8, \ldots)$ with coordinates summing to $m/2$. Thus $|H|$ is at most the number of partitions of $m/2$ into $m/2$ parts which is $\binom{m}{m/2} \leq 2^m$. This finishes the proof of the Theorem. □

Remark 1. *The bound on m th moment of $\sum_i X_i$ in the theorem will be used in a standard fashion to get tail bounds. For any t, by Markov inequality, we get from the theorem $Pr(|\sum_i X_i| \geq t) \leq \frac{(24nm)^{m/2}}{t^m}$. The right hand side is minimized at $m = t^2/(cn)$. So if the hypothesis of the theorem holds for this m (as is the case if $|X_i| \leq 1$) we get the conclusion of H-A: $Pr(|\sum_i X_i| \geq t) \leq c_1 \exp\left(\frac{-t^2}{c_2 n}\right)$.*

The set-up for Chernoff bounds is: $X_1, X_2, \ldots X_n$ are i.i.d. Bernoulli random variables with $EX_i = \nu$. For any $t \leq n\nu$ Chernoff bounds assert: $Pr\left(|\sum_{i=1}^n (X_i - \nu)| > t\right) \leq e^{-t^2/(cn\nu)}$. We can get this by applying the theorem with $m = t^2/72n\nu$ to: $X_i' = (X_i - \nu)/\sqrt{\nu}$ to get $E\left(\sum_{i=1}^n (X_i - \nu)\right)^m \leq (cnm\nu)^{m/2}$ For $t \geq n\nu$, we use the theorem on $X_i' = \sqrt{\frac{n}{m}}(X_i - \nu)$ with $m = ct$ to get $E\left(\sum_{i=1}^n (X_i - \nu)\right)^m \leq (cm)^m$ and by Markov, $Pr\left(|\sum_{i=1}^n (X_i - \nu)| > t\right) \leq e^{-ct}$.

Comparisons with Burkholder type inequalities and Efron-Stein type inequalities are given in section (XII).

III. Notation, Concentration for functions of independent random variables

Theorem 1 and the Main Theorem (6) will often be applied to a real-valued function $f(Y_1, Y_2, \ldots Y_n)$ of independent (not necessarily real-valued) random variables $Y_1, Y_2, \ldots$ to show concentration of f. This is usually done using the Doob's Martingale construction which we recall in this section. While there is no new stuff in this section, we will introduce notation used throughout the paper.

Let $Y_1, Y_2, \ldots Y_n$ be independent random variables. Denote $Y = (Y_1, Y_2, \ldots Y_n)$. Let $f(Y)$ be a real-valued function of Y. One defines the classical Doob's Martingale:

$$X_i = E(f|Y_1, Y_2, \ldots Y_i) - E(f|Y_1, Y_2, \ldots Y_{i-1}).$$

It is a standard fact that the X_i form a Martingale difference sequence and so (1) is satisfied. We will use the short-hand $E^i f$ to denote $E(f|Y_1, Y_2, \ldots Y_i)$, so

$$X_i = E^i f - E^{i-1} f.$$

Let $Y^{(i)}$ denote the $n-1$-tuple of random variables $Y_1, Y_2, \ldots Y_{i-1}, Y_{i+1}, \ldots Y_n$ and suppose $f(Y^{(i)})$ is also defined. Let

$$Z_i = f(Y) - f(Y^{(i)}).$$

$$\text{Then, } X_i = E^i Z_i - E_{Y_i}\left(E^i Z_i\right), \tag{6}$$

since $Y^{(i)}$ does not involve Y_i. f, Y_i, X_i, Z_i will all be reserved for these quantities throughout the paper. We use c to denote a generic constant which can have different values.

IV. Random TSP with Inhomogeneous, long-tailed distributions

One of the earliest problems to be studied under Probabilistic Analysis [19] is the concentration of the length f of the shortest Hamilton cycle through a set of n points picked uniformly independently at random from a unit square. It is known that $Ef \in \Theta(\sqrt{n})$ and concentration in intervals of constant length with sub-Gaussian tails was proved after many earlier steps by Rhee and Talagrand [15] and Talagrand's inequality yielded a simpler proof of this. All of the proofs replace the uniform density by a Poisson distribution. Here, we will give a simple self-contained proof of the concentration result for more general distributions than the Poisson. Two important points of our more general distribution are

- Inhomogeneity (some areas of the unit square having greater probability than others) is allowed.
- Longer tails (for example with power-law distributions) than the Poisson are allowed.

Theorem 2. *Suppose the unit square is divided into n small squares, each of side $1/\sqrt{n}$. Let $Y_1, Y_2, \ldots Y_n$ 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 $f = f(Y_1, Y_2, \ldots Y_n)$ is the length of the shortest Hamilton tour through $Y_1 \cup Y_2 \cup \ldots Y_n$. We have [2]

$$E(f - Ef)^m \leq (cm)^{m/2}.$$

Proof Order the small squares in $\sqrt{n}$ layers - the first layer consists of all squares touching the bottom or left boundary; the second layer consists of all squares which are 1 square away from the bottom and left boundary etc. until the last layer is the top right square (order within each layer is arbitrary.) Let S_i be the i th square. Let $\tau = \tau(Y_{i+1}, \ldots Y_n)$ be the minimum distance from a point of S_i to a point in $Y_{i+1} \cup \ldots Y_n$ and $\tau_0 = \text{Min}(\tau, 2\sqrt{2})$. τ_0 depends only on $Y_{i+1}, \ldots Y_n$. Since we may take a detour from a tour of $Y^{(i)}$ at the nearest point to S_i, tour Y_i and then return to the tour of $Y^{(i)}$, we have (see notation in section (III)) $Z_i \leq 2\tau_0 + O(1/\sqrt{n}) + f(Y_i)$, Since $Z_i \geq 0$, we get using (6)for any even l:

$$-E_{Y_i}(E^i Z_i) \leq X_i \leq E^i Z_i \tag{7}$$

$$\implies |X_i| \leq 2E\tau_0 + O(1/\sqrt{n}) + f(Y_i)$$

$$\implies E^{i-1} X_i^l \leq 2^l (E\tau_0)^l + \frac{c^l}{n^{l/2}} + \frac{c^l}{n^{l/2}} E|Y_i|^{l/2} \tag{8}$$

where the last step uses the following well-known fact [19].

Claim 1. *For any square B of side α in the plane and any set of s points in B, there is a Hamilton tour through the points of length at most $c\alpha\sqrt{s}$.*

First focus on $i \leq n - 100 \ln n$. For any $\lambda \in [0, 5\sqrt{\ln n}/\sqrt{n}]$, there is a square region T_λ of side λ inside $S_{i+1} \ldots S_n$ (indeed, inside the later layers) which touches S_i. So, $\Pr(\tau \geq \sqrt{2}\lambda) \leq \Pr(T_\lambda \cap (Y_{i+1} \cup \ldots Y_n) = \emptyset) \leq e^{-cn\lambda^2}$ by the hypothesis that $\Pr(|Y_j| = 0) < c < 1$. This implies (by integration)

[2] If each Y_i is generated according to a Poisson of intensity 1 (=Area of small square times n), then $E|Y_i|^l \leq l^l$ and so the conditions of the theorem are satisfied for all m (with room to spare). Choosing m to be best value we get (see Remark (1)) for $t \in O(\sqrt{n})$, $\Pr(|f - Ef| \geq t) \leq c_1 e^{-\Omega(t^2)}$ matching Rhee and Talagrand's result (but for constants). Note that by Claim (1), there is nothing to prove for $t \geq c\sqrt{n}$.

that $E\tau_0 \le O(1/\sqrt{n})$. Plugging this and the assumption that $E|Y_i|^{l/2} \le l^l$ into (8), we get $E^{i-1}X_i^l \le \frac{l^l}{n^{l/2}}$. We now apply theorem (1) to $c_6\sqrt{n}X_i$, for $i = 1, 2, \dots n - 100\ln n$ to get

$$E\left(\sum_{i=1}^{n-100\ln n} X_i\right)^m \le (cm)^{m/2}. \tag{9}$$

We also have $|\sum_{i=n-100\ln n+1}^{n} X_i| \le 2\sqrt{2} + \frac{c\sqrt{\ln n}\sqrt{\sum_{i=n-100\ln n+1}^{n} |Y_i|}}{n^{1/2}}$, the last since all these small squares are inside a square of side $\sqrt{\ln n}/\sqrt{n}$. Now using $E\left(\sum_{i=n-100\ln n+1}^{n} |Y_i|\right)^{m/2} \le c(\ln n)^{m/2} m^{m-\epsilon m}$, the theorem follows. □

V. Minimum Weight Spanning tree

This problem is tackled similarly to the TSP in the previous section. We will get the same result as Talagrand's inequality is able to derive, the proof is more or less the same as our proof for the TSP, except that there is an added complication because adding points does not necessarily increase the weight of the minimum spanning tree. The standard example is when we already have the vertices of an equilateral triangle and add the center to it.

Theorem 3. *Under the same hypotheses and notation as in Theorem (2), suppose $f = f(Y_1, Y_2, \dots Y_n)$ is the length of the minimum weight spanning tree on $Y_1 \cup Y_2 \cup \dots Y_n$. We have*

$$E(f - Ef)^m \le (cm)^{m/2}.$$

Proof If we already have a MWST for $Y \setminus Y_i$, we can again connect the point in $Y_{i+1}, \dots Y_n$ closest to S_i to S_i, then add on a MWST on Y_i to get a spanning tree on Y. This implies again that $Z_i \le \tau_0 + \frac{c\sqrt{|Y_i|}}{\sqrt{n}}$. But now, we could have $f(Y) < f(\hat{Y})$. We show that

Claim 2. $Z_i \ge -c_{10}\tau_0 - \frac{c\sqrt{|Y_i|}}{\sqrt{n}}$.

Proof We may assume that $Y_i \ne \emptyset$. Consider the MWST T of Y. We call an edge of the form $(x, y) \in T : x \in Y_i, y \in Y \setminus Y_i$, with $|x - y| \ge c_9/\sqrt{n}$, a long edge and an edge $(x, y) \in T : x \in Y_i, y \in Y \setminus Y_i$, with $|x - y| < c_9/\sqrt{n}$ a short edge. It is well-known that the degree of each vertex in T is $O(1)$ (we prove a more complicated result in the next para), so there are at most $6|Y_i|$ short edges; we remove all of them and add a MWST on the non-Y_i ends of them at a cost of at most $O(\sqrt{|Y_i|}/\sqrt{n})$ by Claim (1).

We claim that there are at most $O(1)$ long edges - indeed if $(x, y), (w, z)$ are any two long edges with $x, w \in Y_i$, we have $|y - z| \ge |x - y| - \frac{\sqrt{2}}{\sqrt{n}}$, since otherwise, $T \setminus (x, y) \cup (y, z) \cup (x, w)$ would contain a better spanning tree than T. Similarly, $|y - z| \ge |w - z| - \frac{\sqrt{2}}{\sqrt{n}}$. Let x_0 be the center of square S_i. The above implies that in the triangle x_0, y, z, we have $|y - z| \ge |x_0 - y| - \frac{6}{\sqrt{n}}, |x_0 - z| - \frac{6}{\sqrt{n}}$. But $|y - z|^2 = |y - x_0|^2 + |z - x_0|^2 - 2|y - x_0||z - x_0| \cos(y, x_0, z)$. Assume without loss of generality that $|y - x_0| \ge |z - x_0|$. If the angle y, x_0, z were less than 10 degrees, then we would have $|y - z|^2 \le |y - x_0|^2 + |z - x_0|^2 - 1.8|y - x_0||z - x_0| < (|y - x_0| - 0.04|z - x_0|)^2$ a contradiction. So, we must have that the angle is at least 10 degrees which implies that there are at most 36 long edges.

Let a be the point in $Y_{i+1}, \dots Y_n$ closest to S_i if $Y_{i+1} \cup \dots \cup Y_n$ is non-empty; otherwise, let a be the point in $Y_1 \cup Y_2 \cup \dots Y_{i-1}$ closest to S_i. We finally replace each long edge $(x, y), x \in Y_i$ by edge (a, y). This clearly only costs us $O(\tau_0)$ extra, proving the claim.

Now the proof of the theorem is completed analogously to the TSP. □

VI. Chromatic Number of inhomogeneous random graphs

Martingale inequalities have been used in different (beautiful) ways on the chromatic number χ of an (ordinary) random graph $G(n, p)$, where each edge is chosen independently to be in with probability p ([16],[2], [3],[8]).

Here we study chromatic number in a more general model. An inhomogeneous random graph - denoted $G(n, P)$ - has vertex set $[n]$ and a $n \times n$ matrix $P = \{p_{ij}\}$ where p_{ij} is the probability that edge (i, j) is in the graph. Edges are in/out independently. Let

$$p = \frac{\sum_{i,j} p_{ij}}{\binom{n}{2}}$$

be the average edge probability. Let $\chi = \chi(G(n, P)$ be the chromatic number. Since each node can change the chromatic number by at most 1, it is easy to see that $\Pr(|\chi - E\chi| \ge t) \le c_1 e^{-c_2 t^2/n}$ by H-A. Here we prove a better result when the graph is sparse, i.e., when $p \in o(1)$.

Theorem 4. *For any $t \in (0, n\sqrt{p})$, we have*

$$Pr(|\chi - E\chi| \ge t) \le e^{\frac{-ct^2}{n\sqrt{p}\ln n}}.$$

Remark 2. *Given only p, note that χ could be as high as $\Omega(n\sqrt{p})$: for example, p_{ij} could be 1 for $i, j \in T$ for some T with $|T| = O(n\sqrt{p})$ and zero elsewhere.*

Proof Let $p_i = \sum_j p_{ij}$ be the expected degree of i. Let

$$S = \{i : p_i \ge n\sqrt{p}\}.$$

$|S| \le 2n\sqrt{p}$. Split the $n - |S|$ vertices of $[n] \setminus S$ into $k = (n - |S|)\sqrt{p}$ groups $G_1, G_2, \dots G_k$ by picking for each vertex a group uniformly at random independent of other vertices. It follows by routine application of Chernoff bounds that with probability at least 1/2, we have : (i) for each i, the sum of $p_{ij}, j \in$ (same group as i) $\le O(\ln n)$ and (ii) $|G_t| \in O(\ln n/\sqrt{p})$ for all t. We choose any partition of $[n] \setminus S$ into $G_1, G_2, \dots G_k$ satisfying (i) and (ii) at the outset and fix this partition. Then we make the random choices to choose $G(n, P)$.

Define Y_i for $i = 1, 2, \dots k + |S|$ as the set of edges (of $G(n, P)$) in $G_i \times (G_1 \cup G_2 \cup \dots G_{i-1})$. We can define the Doob's Martingale $X_i = E(\chi|Y_1, Y_2, \dots Y_i) - E(\chi|Y_1, Y_2, \dots Y_{i-1})$. First consider $i = 1, 2, \dots k$. Define Z_i as in section III. Let d_j be the degree of vertex j in G_i in the graph induced on G_i alone. Z_i is at most $\max_{j\in G_i} d_j$, since we can always color G_i with this many additional colors. d_j is the sum of independent Bernoulli random variables with $Ed_j = \sum_{l\in G_i} p_{jl} \leq O(\ln n)$. By Remark (1), we have that $E(d_j - Ed_j)^l \leq \text{MAX}((cl\ln n)^{l/2}, (cl)^l$. Hence, $E^{i-1}(Z_i^l) \leq (cl)^l + (cl\ln n)^{l/2}$.

We will apply Theorem (1) to the sum

$$\frac{c_7 X_1}{\ln n} + \frac{c_7 X_2}{\ln n} + \dots \frac{c_7 X_k}{\ln n}.$$

It follows from the above that these satisfy the hypothesis of the Theorem provided $m \leq k$. From this, we get that

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

For $i = k+1, \dots k+|S|$, Z_i are absolutely bounded by 1, so by the Theorem $E(X_{k+1}+X_{k+2}+\dots X_{k+|S|})^m \leq (c|S|m)^{m/2}$. Thus,

$$E\left(\sum_{i=1}^{k+|S|} X_i\right)^m \leq (cmk\ln n)^{m/2}.$$

Let $t\in(0, n\sqrt{p})$. We take $m =$ the even integer nearest to $t^2/(c_4 n\sqrt{p}\ln n)$ to get the theorem. □

Question In studying chromatic number of random graphs, the maximum average degree of any sub-graph (a quantity we call MAD) is very useful. Clearly, the chromatic number is at most MAD, since we can first remove a vertex with degree at most MAD, color the rest recursively and then put the vertex back in. For a inhomogeneous random graph $G(n, P)$, we define

$$\text{MAD}(P) = \text{MAX}_{U\subseteq[n]} \frac{\sum_{i,j\in U} p_{ij}}{|U|}.$$

Is the following statement true for $G(n, P)$:

$$\Pr\left(|\chi - E\chi)| \geq t\right) \leq \exp\left(-\frac{ct^2}{\text{MAD}(P)\ln n}\right)?$$

VII. Random Projections

A famous theorem of Johnson-Lindenstrauss [20] asserts that if v is picked uniformly at random from the surface of the unit ball in $\mathbf{R}^n$, then for $k\leq n$, and $\epsilon\in(0,1)$, [3]

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

The proofs of this exploit the details of the uniform density or the Gaussian in the equivalent way of picking v - pick each v_i according to a Gaussian and scale to length 1. Here, we will prove the same conclusion under weaker hypotheses which allows again longer tails (and so does not use any special property of the Gaussian). This is the first application which uses the Strong Negative Correlation condition rather than the Martingale Difference condition.

[3] A clearly equivalent statement talks about the length of the projection of a fixed unit length vector onto a random $k-$ dimensional sub-space.

Theorem 5. *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* [4]

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

Proof The theorem will be applied with $X_i = Y_i^2 - EY_i^2$. First, (i) implies for odd l: $EX_i(X_1+X_2+\dots X_{i-1})^l \leq 0$, by (an elementary version) of the FKG inequality. [If $X_1+X_2+\dots X_{i-1} = W$, then since W^l is an increasing function of W for odd l and $E(X_i|W)$ a non-increasing function of W, we have $EX_iW^l = E_W\left(E(X_i|W)W^l\right) \leq E_W(E(X_i|W))EW^l = EX_iEW^l = 0$.] Now, for even l, $E^{i-1}(X_i^l) \leq 2^l EY_i^{2l} + 2^l(EY_i^2)^l \leq (cl)^l/n^l$. So we may apply the theorem to the scaled variables $c_7 nX_i$, for $i=1,2,\dots k$ for $m\leq k$ to get the result. □

VIII. Main Probability Inequality

Now, we come to the main theorem. We will again assume Strong Negative Correlation (1) of the real-valued random variables $X_1, X_2, \dots X_n$. The first main point of departure from Theorem (1) is that we allow different variables to have different bounds on conditional moments. A more important point will be that we will use information on conditional moments conditioned on "typical" values of previous variables as well as the pessimistic "worst-case" values. More specifically, we assume the following bounds on moments for $i = 1, 2, \dots n$ (m again is an even positive integer):

$$E(X_i^l|X_1+X_2+\dots X_{i-1}) \leq M_{il} \quad \text{for } l = 2,4,6,8\dots m. \tag{10}$$

In some cases, the bound M_{il} may be very high for the "worst-case" $X_1+X_2+\dots X_{i-1}$. We will exploit the fact that for a "typical" $X_1+X_2+\dots X_{i-1}$, $E(X_i^l|X_1+X_2+\dots X_{i-1})$ may be much smaller. To this end, suppose

$$\mathcal{E}_{i,l}\ ,\ l = 2,4,6,\dots m\,; i = 1,2,\dots n$$

are events. $\mathcal{E}_{i,l}$ is to represent the "typical" case. $\mathcal{E}_{il}$ will be the whole sample space. In addition to (10), we assume that

$$E(X_i^l|X_1+X_2+\dots X_{i-1}, \mathcal{E}_{i,l}) \leq L_{il} \tag{11}$$

$$\Pr(\mathcal{E}_{i,l}) = 1 - \delta_{i,l} \tag{12}$$

Theorem 6 (Main Theorem)**.** *Let $X_1, X_2, \dots X_n$ be real valued random variables satisfying Strong Negative Correlation*

[4] As usual, it is simple to derive tail bounds from the moment bound in the theorem. For $\epsilon>0$, put $m = k\epsilon^2/(ec)$ if $\epsilon^2\leq ec$ and $m=k$ otherwise to get

$$\Pr\left(\left|\sum_{i=1}^k (Y_i^2 - EY_i^2)\right| \geq \epsilon\frac{k}{n}\right) \leq c_1 \text{Min}\left(\exp\left(\frac{-k\epsilon^2}{2ec}\right), \left(\frac{c}{\epsilon^2}\right)^{k/2}\right).$$

(1) and m be a positive even integer. Then for $X=\sum_{i=1}^n X_i$,

$$EX^m \le (cm)^{\frac{m}{2}+1}\left(\sum_{l=1}^{m/2}\frac{m^{1-\frac{1}{l}}}{l^2}\left(\sum_{i=1}^n L_{i,2l}\right)^{\frac{1}{l}}\right)^{m/2}$$

$$+(cm)^{m+2}\sum_{l=1}^{m/2}\frac{1}{nl^2}\sum_{i=1}^n\left(nM_{i,2l}\delta_{i,2l}^{2/(m-2l+2)}\right)^{m/2l}.$$

There are two central features of the Theorem. The first is the distinction between typical and worst case conditional moments which we have already discussed. Note that while the M_{il} may be much larger than L_{il}, the M_{il} get modulated by $\delta_{il}^{2/(p-l+2)}$ which can be made sufficiently small.

A second feature of the Theorem is similar to Theorem (**??**) in that the second moment term will often be the important one. If we have

$$\text{MAX}_i L_{i,2l}=L_{2l}, \tag{13}$$

then we get an upper bound of

$$(cm)^{m/2}\left(nL_2+\sqrt{np}L_4^{1/2}+\ldots\right)^{m/2},$$

where we note that for $m<<n$, the coefficients of higher moments decline fast, so that under reasonable conditions, the nL_2 term is what matters. In this case, it will not be difficult to see that we have qualitatively sub-Gaussian behavior with variance equal to the sum of the variances.

IX. Proof of the Main Theorem

Proof (of the Theorem) Let $A=X_1+X_2+\ldots X_{n-1}$. In what follows, l will run over even integers from 2 to m. As in the proof of Theorem (1), we get

$$E(A+X_n)^m\le EA^m+3\sum_l\binom{m}{l}Eb_l,$$

where, $b_l=X_n^lA^{m-l}$. Denote by ω points in the sample space; so $A(\omega)$ is the value of A at ω. We have

$$Eb_l=\Pr(\mathcal{E}_{n,l})E(b_l|\mathcal{E}_{n,l})+\Pr(\neg\mathcal{E}_n)E(b_l|\neg\mathcal{E}_{n,l})$$
$$\le L_{nl}\int_{\omega\in\mathcal{E}_{n,l}}A(\omega)^{m-l}d\omega+M_{nl}\int_{\omega\in\neg\mathcal{E}_{n,l}}A(\omega)^{m-l}d\omega.$$

We use Hölder's inequality to get that the second term is at most $M_{n,l}\left(EA^{m-l+2}\right)^{\frac{m-l}{m-l+2}}\left(\Pr(\neg\mathcal{E}_{n,l})\right)^{\frac{2}{m-l+2}}\le \hat{M}_{nl}\left(EA^{m-l+2}\right)^{\frac{m-l}{m-l+2}}$, where $\hat{M}_{i,l}=M_{i,l}\delta_{i,l}^{2/(m-l+2)}$.

We use Young's inequality which says that for any $a,b>0$ real and $q,r>0$ with $\frac{1}{q}+\frac{1}{r}=1$, we have $ab\le a^q+b^r$; we apply this below with $q=(m-l+2)/2$ and $r=(m-l+2)/(m-l)$ and λ_{nl} a positive real to be specified later :

$$\hat{M}_{nl}(EA^{m-l+2})^{\frac{m-l}{m-l+2}}$$
$$=\left(\hat{M}_{nl}^{2m/l(m-l+2)}\lambda_{nl}^{-\frac{m-l}{m-l+2}}\right)\left(\hat{M}_{nl}^{\frac{l-2}{l}}\lambda_{nl}EA^{m-l+2}\right)^{\frac{m-l}{m-l+2}}$$
$$\le\hat{M}_{nl}^{m/l}\lambda_{nl}^{-\frac{m-l}{2}}+\hat{M}_{nl}^{(l-2)/l}\lambda_{nl}EA^{m-l+2},$$

So, we get :

$$E\left(\sum_{i=1}^n X_i\right)^m\le\sum_{\substack{l\ge0\\ \text{even}}}^m a_{nl}EA^{m-l},$$

where

$$a_{nl}=1+3\lambda_{n2}\binom{m}{2}, \qquad l=0$$
$$a_{nl}=3\binom{m}{l}L_{nl}+3\binom{m}{l+2}\hat{M}_{n,l+2}^{l/(l+2)}\lambda_{n,l+2},\quad 2\le l\le m-2$$
$$a_{nl}=3L_{nm}+3\sum_{\substack{l_1\ge2\\ \text{even}}}^m\binom{m}{l_1}\frac{\hat{M}_{nl_1}^{m/l_1}}{\lambda_{nl_1}^{(m-l_1)/2}}, \qquad l=m.$$

An exactly similar argument yields for any $r\le n$ and any $q\le m$, even

$$E\left(\sum_{i=1}^r X_i\right)^q\le\sum_{\substack{l\ge0\\ \text{even}}}^q a_{rl}^{(q)}E(\sum_{i=1}^{r-1}X_i)^{q-l},$$

where (since $\delta_{r,l}^{1/(q-l+2)}\le\delta_{r,l}^{1/(m-l+2)}$)

$$a_{rl}^{(q)}=1+3\lambda_{r2}\binom{q}{2}, \qquad l=0$$
$$a_{rl}^{(q)}=3\binom{q}{l}L_{rl}+3\binom{q}{l+2}\hat{M}_{r,l+2}^{l/(l+2)}\lambda_{r,l+2},\quad 2\le l\le q-2$$
$$a_{rl}^{(q)}=3L_{rq}+3\sum_{\substack{l_1\ge2\\ \text{even}}}^q\binom{q}{l_1}\frac{\hat{M}_{rl_1}^{q/l_1}}{\lambda_{rl_1}^{(q-l_1)/2}}, \qquad l=q.$$

Now we set

$$\lambda_{rl}=\frac{1}{3m^2n^{2/l}}\text{for } \ l=2,4,6,8,\ldots.$$

Then we get

$$a_{rl}^{(q)}\le a_{rl}=1+\frac{1}{n},\quad l=0$$
$$a_{rl}^{(q)}\le a_{rl}=3\binom{m}{l}\left(L_{rl}+\hat{M}_{r,l+2}^{l/(l+2)}n^{-2/(l+2)}\right),\ 2\le l\le q-2$$
$$a_{rq}^{(q)}\le\hat{a}_{rq}=3L_{rq}+3\sum_{\substack{l_1\ge2\\ \text{even}}}^q\binom{q}{l_1}\hat{M}_{rl_1}^{q/l_1}(3m^2)^{(q-l_1)/2}n^{(q-l_1)/l_1}.$$

It is important to make $a_{r0}^{(q)}$ not be much greater than 1 because in this case only n is reduced and so in the recurrence, this could happen n times. Note that except for $l=q$, the other a_{rl} do not depend upon q; we have used $\hat{a}_{rq}$ to indicate that this extra dependence. With this, we have

$$E\left(\sum_{i=1}^r X_i\right)^q\le\hat{a}_{rq}+\sum_{\substack{l\ge0\\ \text{even}}}^{q-2}a_{rl}E(\sum_{i=1}^{r-1}X_i)^{q-l}.$$

We wish to solve these recurrences by induction on r,q. Intuitively, we can imagine a tree with root marked (r,q) (since we are bounding $E(\sum_{i=1}^r X_i)^q$. The root has $\frac{q}{2}+1$ children which are marked $(r-1,q-l)$ for $l=0,2,\ldots q/2$; the node marked $(r-1, q-l)$ is trying to bound $E(\sum_{i=1}^{r-1} X_i)^{q-l}$. There are also weights on the edges of a_{rl} respectively. The tree keeps going until we reach the leaves - which are marked $(1, q)$ or $(r, 0)$. It is intuitively easy to argue that the bound we are seeking at the root is the sum over all paths from the root to the leaves of the product of the edge weights on the path. We formalize this in a lemma.

For doing that, for $1 \le r \le n; 2 \le q \le m$, q even and $1 \le i \le r$ define $S(r, q, i)$ as the set of $s = (s_i, s_{i+1}, s_{i+2}, \dots s_r)$ with $s_i > 0; s_{i+1}, s_{i+2}, \dots s_r \ge 0$ and $\sum_{j=i}^{r} s_j = q; s_j$ even.

Lemma 1. *For any $1 \le r \le n$ and any $q \le p$ even, we have*

$$E(\sum_{i=1}^{r} X_i)^q \le \sum_{i=1}^{r} \sum_{s \in S(r,q,i)} \hat{a}_{i,s_i} \prod_{j=i+1}^{r} a_{j,s_j}.$$

Proof Indeed, the statement is easy to prove for the base case of the induction - $r = 1$ since $\mathcal{E}_{1l}$ is the whole sample space and $EX_1^q \le L_{1q}$. For the inductive step, we proceed as follows.

$$E(\sum_{i=1}^{r} X_i)^q \le \sum_{\substack{s_r \ge 0 \\ \text{even}}}^{q-2} a_{r,s_r} E(\sum_{i=1}^{r-1} X_i)^{q-s_r} + \hat{a}_{r,q}$$
$$\le \hat{a}_{r,q} + \sum_{i=1}^{r-1} \sum_{\substack{s_r \ge 0 \\ \text{even}}}^{q-2} a_{r,s_r} \sum_{s \in S(r-1,q-s_r,i)} \hat{a}_{i,s_i} \prod_{j=i+1}^{r-1} a_{j,s_j}.$$

We clearly have $S(m, q, m) = \{q\}$ and for each fixed $i, 1 \le i \le r-1$, there is a 1-1 map $S(r-1, q, i) \cup S(r-1, q-2, i) \cup \dots S(r-1, 2, i) \to S(r, q, i)$ given by $s = (s_i, s_{i+1}, \dots s_{r-1}) \to s' = (s_i, \dots s_{r-1}, q - \sum_{j=i}^{r-1} s_j)$ and it is easy to see from this that we have the inductive step, finishing the proof of the Lemma. The "sum of products" form in the lemma is not so convenient to work with. We will now get this to the "sum of moments" form stated in the Theorem. This will require a series of (mainly algebraic) manipulations with ample use of Young's inequality, the inequality asserting $(a_1 + a_2 + \dots a_r)^q \le r^{q-1}(a_1^q + a_2^q + \dots a_r^q)$ for positive reals $a_1, a_2, \dots$ and $q \ge 1$ and others.

So far, we have (moving the $l = 0$ terms separately in the first step)

$$E\left(\sum_{i=1}^{n} X_i\right)^m \le \left(\prod_{i=1}^{n} a_{i0}\right) \sum_{i=1}^{n} \sum_{s \in S(n,m,i)} \hat{a}_{i,s_i} \prod_{\substack{j=i+1 \\ s_j \ne 0}}^{n} a_{j,s_j}$$
$$\le 3 \sum_{i=1}^{n} \sum_{s \in S(n,m,i)} \hat{a}_{i,s_i} \prod_{\substack{j=i+1 \\ s_j \ne 0}}^{n} a_{j,s_j}$$
$$\le 3 \sum_{t \ge 1}^{m/2} \left(\sum_{i=1}^{n} \hat{a}_{i,2t}\right) \sum_{s \in Q(p-2t)} \prod_{\substack{j=1 \\ s_j \ne 0}}^{n} a_{j,s_j} \tag{14}$$

where, $Q(q) = \{s = (s_1, s_2, \dots s_n) : s_i \ge 0 \text{ even }; \sum_j s_j = q\}$

Fix q for now. For $s \in Q(q), l = 0, 1, 2, \dots p/2$, let $T_l(s) = \{j : s_j = 2l\}$ and $t_l(s) = |T_l(s)|$. Note that $\sum_{l=0}^{q/2} l t_l(s) = q/2$. Call $t(s) = (t_0(s), t_1(s), t_2(s), \dots t_{q/2}(s))$ the "signature" of s. In the special case when a_{il} is independent of i, the signature clearly determines the "s term" in the sum (14). For the general case too, it will be useful to group terms by their signature. Let (the set of possible signatures) be T. [T consists of all $t = (t_0, t_1, t_2, \dots t_{q/2})$ with $t_l \ge 0$ $\sum_{l=1}^{q/2} l t_l = q/2$ $t_0 \le n; \sum_{l=0}^{q/2} t_l = n$.

$$\text{Now, } \sum_{s \in Q(q)} \prod_{\substack{j=1 \\ s_j \ne 0}}^{n} a_{j,sj} = \sum_{t \in T} \overbrace{\sum_{T_0, T_1, T_2, \dots T_{q/2} : |T_l| = t_l}}^{T_l \text{ partition } [n]} \prod_{l=1}^{q/2} \prod_{i \in T_l} a_{i,2l}$$
$$\le \sum_{t \in T} \prod_{l=1}^{q/2} \frac{1}{t_l!} \left(\sum_{i=1}^{n} a_{i,2l}\right)^{t_l},$$

since the expansion of $(\sum_{i=1}^{n} a_{i,2l})^{t_l}$ contains $t_l!$ copies of $\prod_{i \in T_l} a_{i,2l}$ (as well other terms we do not need.) Now define $R = \{r = (r_1, r_2, \dots r_{q/2}) : r_l \ge 0; \sum_l r_l = q/2\}$. We have

$$\sum_{t \in T} \prod_{l=1}^{q/2} \frac{1}{t_l!} \left(\sum_{i=1}^{n} a_{i,2l}\right)^{t_l} \le \sum_{r \in R} \prod_{l} \frac{1}{(r_l/l)!} (\sum_{i=1}^{n} a_{i,2l})^{r_l/l}$$
$$\le \frac{1}{(q/2)!} \left(\sum_{l=1}^{q/2} m^{1-(1/l)} \left(\sum_i a_{i,2l}\right)^{1/l}\right)^{q/2}, \tag{15}$$

where the first inequality is seen by substituting $r_l = t_l l$ and noting that the terms corresponding to the r such that $l | r_l \forall l$ are sufficient to cover the previous expression and the other terms are non-negative. To see the second inequality, we just expand the last expression and note that the expansion contains $\prod_l (\sum_i a_{i,2l})^{r_l/l}$ with coefficient $\binom{q/2}{r_1, r_2, \dots r_{q/2}}$ for each $r \in R$. Now, it only remains to see that $m^{r_l(1-(1/l))} \ge \frac{r_l!}{(r_l/l)!}$, which is obvious. Thus, we have plugging in (15) into (14), we have that EX^m is at $\lesssim$[5] $\sum_{t=1}^{\frac{m}{2}} \frac{3}{(\frac{m}{2}-t)!} (\sum_{i=1}^{n} \hat{a}_{i,2t}) \left(\sum_{l=1}^{\frac{m}{2}-t} m^{1-\frac{1}{l}} (\sum_i a_{i,2l})^{\frac{1}{l}}\right)^{\frac{m}{2}-t}$.

$$\text{Now, } (\frac{m}{2} - t)! \ge (\frac{m}{2} - t)^{\frac{m}{2}-t} e^{-\frac{m}{2}} e^t$$
$$\ge m^{\frac{m}{2}-t} e^{-\frac{m}{2}} \text{Min}_t \left[\left(\frac{\frac{m}{2} - t}{m}\right)^{\frac{m}{2}-t} e^t\right] \ge m^{\frac{m}{2}-t} (2e)^{-\frac{m}{2}},$$

the last using Calculus to differentiate the log of the expression with respect to t to see that the min is at $t = 0$. Thus, EX^m is $\lesssim \sum_t \left[\left(\frac{3}{m} \sum_{l=1}^{\frac{m}{2}-t} m^{1-\frac{1}{l}} (\sum_i a_{i,2l})^{\frac{1}{l}}\right)^{\frac{m}{2}-t}\right] [\sum_{i=1}^{n} \hat{a}_{i,2t}]$. Let α, β denote the quantities in the 2 square brackets respectively. Young's inequality gives us: : $\alpha\beta \le \alpha^{m/(m-2t)} + \beta^{m/2t}$. Thus, EX^m is bounded by

$$\sum_{t=1}^{\frac{m}{2}} \left(\sum_i \hat{a}_{i,2t}\right)^{\frac{m}{2t}} + \left(\sum_{l=1}^{\frac{m}{2}-1} m^{-\frac{1}{l}} \left(\sum_i a_{i,2l}\right)^{\frac{1}{l}}\right)^{\frac{m}{2}} \tag{16}$$

[5]We use $\alpha \lesssim \beta$ to mean $\alpha \le c^m \beta$, since we freely allow c^m factors in the theorem. But we will explicitly keep track of $m^{O(m)}$ factors.

In what follows, let l_1 run over even values to m and i run from 1 to n.
$\sum_{t=1}^{\frac{m}{2}} \left(\sum_i \hat{a}_{i,2t}\right)^{\frac{m}{2t}} \lessapprox \sum_t \left(\sum_i L_{i,2t}\right)^{\frac{m}{2t}}$
$+m^m \sum_t \left(\frac{1}{n}\sum_i \sum_{l_1\leq 2t}\left(\frac{2t}{l_1 m}\right)^{l_1}(n\hat{M}_{i,l_1})^{2t/l_1}\right)^{\frac{m}{2t}} \lessapprox$
$\sum_t (\sum_i L_{i,2t})^{\frac{m}{2t}} + m^m \sum_{t,l_1} \frac{t^{\frac{m}{2t}}}{l_1^{l_1} n^{\frac{m}{2t}}}\left(\sum_i (n\hat{M}_{i,l_1})^{2t/l_1}\right)^{\frac{m}{2t}} \lessapprox$

$$\sum_t (\sum_i L_{i,2t})^{\frac{m}{2t}} + m^m \sum_{l_1} \frac{1}{nl_1^{l_1}} \sum_i (n\hat{M}_{i,l_1})^{m/l_1}. \qquad (17)$$

$\sum_{l=1}^{\frac{m}{2}-1} m^{-(1/l)} \left(\sum_i a_{i,2l}\right)^{\frac{1}{l}} \leq$
$\sum_{l=1}^{\frac{m}{2}-1} m^{-\frac{1}{l}} \left(\frac{m^{2l}}{(2l)!}\right)^{\frac{1}{l}} \left(\sum_i L_{i,2l} + \frac{\hat{M}_{i,2l+2}^{\frac{1}{l+1}}}{n^{1/(l+1)}}\right)^{\frac{1}{l}} \leq$
$m^2 \sum_{l=1}^{\frac{m}{2}-1} \frac{m^{-\frac{1}{l}}}{l^2}\left(\left(\sum_i L_{i,2l}\right)^{\frac{1}{l}} + \left(\sum_i \hat{M}_{i,2l+2}\right)^{\frac{1}{l+1}}\right) \leq$

$$m^2 \sum_{l=1}^{\frac{m}{2}} \frac{m^{-\frac{1}{l}}}{l^2}\left(\sum_i L_{i,2l}\right)^{1/l} + m^2 \sum_{l=2}^{\frac{m}{2}} \frac{1}{(l-1)^2}\left(\sum_i \hat{M}_{i,2l}\right)^{\frac{1}{l}}. \qquad (18)$$

We will further bound the last term using Hölder's inequality:
$\left(\sum_{l=2}^{\frac{m}{2}} \frac{\left(\sum_i \hat{M}_{i,2l}\right)^{1/l}}{(l-1)^2}\right)^{\frac{m}{2}} \leq$
$\left(\sum_{l=1}^{\infty}\frac{1}{l^2}\right)^{(m-2)/2}\left(\sum_l \frac{1}{(l-1)^2}\left(\sum_i \hat{M}_{i,2l}\right)^{\frac{m}{2l}}\right)$

$$\leq 2^m \sum_{l=1}^{\frac{m}{2}} \frac{1}{nl^2}\sum_i (n\hat{M}_{i,2l})^{\frac{m}{2l}}. \qquad (19)$$

Now plugging (18,17,19) into (16), we get the Theorem. □

X. Bin Packing

Now we tackle bin packing. The input consists of n i.i.d. items - $Y_1, Y_2, \ldots Y_n \in (0,1)$. Suppose $EY_1 = \mu$ and $\mathbf{Var}Y_1 = \sigma$. Let $f = f(Y_1, Y_2, \ldots Y_n)$ be the minimum number of capacity 1 bins into which the items $Y_1, Y_2, \ldots Y_n$ can be packed. It was shown (after many successive developments) using non-trivial bin-packing theory ([14]) that (with $c, c' > 0$ fixed constants) for $t \in (0, cn(\mu^2+\sigma^2))$,

$$\Pr(|f - Ef| \geq t) \leq c' e^{-ct^2/(n(\mu^2+\sigma^2))}.$$

Talagrand [18] gives a simple proof of this from his inequality (this is the first of the six or so examples in his paper.) [We can also give a simple proof of this from our theorem.] The "ideal" interval of length $O(\sqrt{n}\sigma)$ (as for sums of independent random variables) is impossible.[6]

Our main aim here is to prove that the best length of the interval of concentration is $O(\sqrt{n}(\mu^{3/2}+\sigma))$ when the items take on only one of a fixed finite set of values (discrete distributions - a case which has received much attention in the literature for example [13] and references therein).

[6]An example is when items are of size $1/k$ or $(1/k)+\epsilon$ (k a positive integer) with probability 1/2 each. σ is $O(\epsilon)$. But it is easy to see that interval of concentration has to be at least $\Omega(\sqrt{n}\mu^2)$.

Theorem 7. *Suppose $Y_1, Y_2, \ldots Y_n$ are i.i.d. drawn from a discrete distribution with r atoms each with probability at least $\frac{1}{\log n}$. Let $EY_1 = \mu \leq \frac{1}{r^2 \log n}$ and $\mathbf{Var}Y_i = \sigma^2$. Then for any $t \in (0, n(\mu^3+\sigma^2))$, we have*

$$Pr(|f - Ef| \geq t + r) \leq c_1 e^{-ct^2/(n(\mu^3+\sigma^2))}.$$

Proof Let item sizes be $\zeta_1, \zeta_2, \ldots \zeta_j \ldots \zeta_r$ and the probability of picking type j be p_j. We have : mean $\mu = \sum_j p_j \zeta_j$ and standard deviation $\sigma = (\sum_j p_j(\zeta_j - \mu)^2)^{1/2}$.

[While our proof of the upper bound here is only for problems with a fixed finite number of types, it would be nice to extend this to continuous distributions.] Note that if $\mu \leq r/\sqrt{n}$, then earlier results already give concentration in an interval of length $O(\sqrt{n}(\mu+\sigma)$ which is then $O(r+\sigma)$, so there is nothing to prove. So assume that $\mu \geq r/\sqrt{n}$.

Define a "bin Type" as an $r-$ vector of non-negative integers specifying number of items of each type which are together packable into one bin. If bin type i packs a_{ij} items of type j for $j = 1, 2, \ldots r$ we have $\sum_j a_{ij}\zeta_j \leq 1$. Note that s, the number of bin types depends only on ζ_j, not on n.

For any set of given items, we may write a Linear Programming relaxation of the bin packing problem whose answers are within additive error r of the integer solution. If there are n_j items of size ζ_j in the set, the Linear program is : Primal : (x_i number of bins of type i.)

$$\text{Min} \sum_{i=1}^{s} x_i \quad \text{subject to} \sum_{i=1}^{s} x_i a_{ij} \geq n_j \forall j\ ; x_i \geq 0.$$

Since an optimal basic feasible solution has at most r non-zero variables, we may just round these r up to integers to get an integer solution; thus the additive error is at most r as claimed. In what follows, we prove concentration not for the integer program's value, but for the value of the Linear Program. The Linear Program has the following dual : (y_j "imputed" size of item j)

$$\text{MAX} \sum_{j=1}^{r} n_j y_j \text{ s.t. } \sum_j a_{ij} y_j \leq 1 \text{ for } i = 1, 2, \ldots s;\ y_j \geq 0.$$

Suppose now, we have already chosen all but Y_i. Now, we pick Y_i at random; say $Y_i = \zeta_k$. Let $Y = (Y_1, Y_2, \ldots Y_n)$ and $Y' = (Y_1, Y_2, \ldots Y_{i-1}, Y_{i+1}, \ldots Y_n)$ We denote by $f(Y)$ the value of the Linear Program for the set of items Y. Let

$$Z_i = f(Y) - f(Y').$$

Suppose we have the optimal solution of the LP for Y'. Let i_0 be the index of the bin type which packs $\lfloor 1/\zeta_k \rfloor$ copies of item of type k. Clearly if we increase x_{i_0} by $\frac{1}{\lfloor 1/\zeta_k \rfloor}$, we get a feasible solution to the new primal LP. So

$$Z_i \leq \frac{1}{\lfloor 1/\zeta_k \rfloor} \leq \zeta_k + 2\zeta_k^2.$$

Now, we lower bound Z_i by looking at the dual. For this, let y be the dual optimal solution for Y'. (Note : Thus, $y = y(Y')$ is a function of Y'.) y is feasible to the new dual LP too (after adding in Y_i). So, we get:$Z_i \geq y_k$ and also $y_k \leq \zeta_k + 2\zeta_k^2$. $0 \leq Z_i \leq \zeta_k + 2\zeta_k^2$ gives us

$$E(Z_i^2|Y') \leq \sum_j p_j(\zeta_j + 2\zeta_j^2)^2 \leq \mu^2 + 65\sigma^2 + 64\mu^3. \quad (20)$$

$$E(Z_i|Y') \geq \sum_j p_j y_j(Y') = \mu - \delta(Y') \text{ (say)}. \quad (21)$$

Say the number of items of type j in Y' is $(n-1)p_j + \Delta_j$. Recall that ζ is a feasible dual solution.

$$\sum_j((n-1)p_j + \Delta_j)y_j \geq \sum_j((n-1)p_j + \Delta_j)\zeta_j$$

$$\delta(Y') \leq \frac{1}{n-1}(\sum_j(\Delta_j^2/p_j))^{1/2}(\sum_j p_j(y_j - \zeta_j)^2)^{1/2}$$

$$\leq \frac{32(\mu+\sigma)r}{n}\text{MAX}_j|\Delta_j/\sqrt{p_j}|, \quad (22)$$

where the first step is because y is an optimal dual solution and we have used the fact that $-\zeta_j \leq y_j - \zeta_j \leq 2\zeta_j^2 \leq 2\zeta_j$. Let $(i-1)p_j + \Delta_j'$ and $(n-i)p_j + \Delta_j''$ respectively be the number of items of size ζ_j among $Y_1, Y_2, \ldots Y_{i-1}$ and $Y_{i+1}, \ldots Y_n$. Since Δ_j'' is the sum of $n-i$ i.i.d. random variables, each taking on value $-p_j$ with probability $1-p_j$ and $1-p_j$ with probability p_j, we have $E(\Delta_j'')^2 = \mathbf{Var}(\Delta_j'') \leq np_j$. Consider the event

$$\mathcal{E}_i : |\Delta_j'| \leq 100\sqrt{p\ln(10p/\mu)p_j(i-1)} \quad \forall j.$$

p is to be specified later, but will satisfy $p \leq \frac{1}{10}n(\mu^3+\sigma^2)$. The expected number of "successes" in the $i-1$ Bernoulli trials is $p_j(i-1)$. By using Chernoff, we get $\Pr(\neg\mathcal{E}_i) = (\text{ say })\delta_i \leq \mu^{4p}p^{-4p}$. Using (21) and (22), we get

$$E(Z_i|Y_1, Y_2, \ldots Y_{i-1}; \mathcal{E}_i) \geq \mu - E(\delta|Y_1, Y_2, \ldots Y_{i-1}; \mathcal{E}_i)$$
$$\geq \mu - \frac{32\mu r}{n}E(\max_j \frac{1}{\sqrt{p_j}}(100\sqrt{p\ln(10p/\mu)p_j(i-1)}$$
$$+ (E(\Delta'')^2)^{1/2})) \geq \mu - c\mu^{5/2}r\sqrt{\ln(10p/\mu)} - \frac{c\mu r}{\sqrt{n}}$$

So, we get recalling (20) and using $\mathbf{Var}Z_i = EZ_i^2 + (EZ_i)^2$ $\mathbf{Var}(Z_i|Y_1, Y_2, \ldots Y_{i-1}; \mathcal{E}_i) \leq c(\mu^3 + \sigma^2)$, using $\frac{r}{\sqrt{n}} \leq \mu \leq \frac{1}{r^2 \log n}$. Also, we have $\mathbf{Var}(Z_i|Y_1, Y_2, \ldots Y_{i-1}) \leq E(Z_i^2|Y_1, Y_2, \ldots Y_{i-1}) \leq c\mu^2$. We now appeal to (6) to see that these also give upper bounds on $\mathbf{Var}(X_i)$. Note that $|Z_i| \leq 1$ implies that $L_{i,2l} \leq L_{i,2}$. Now to apply the Theorem, we have $L_{i,2l} \leq c(\mu^3 + \sigma^2)$. So the "$L$ terms" are bounded as follows :

$$\sum_{l=1}^{p/2} \frac{p^{1-(1/l)}}{l^2}\left(\sum_{i=1}^n L_{i,2l}\right)^{1/l} \leq cn\left(\mu^3 + \sigma^2\right)$$

noting that $p \leq n(\mu^3 + \sigma^2)$ implies that the maximum of $((n/p)(\mu^3 + \sigma^2)^{1/l}$ is attained at $l = 1$ and also that $\sum_l(1/l^2) \leq 2$. Now, we work on the M terms in the Theorem. $\max_i \delta_i \leq \mu^{4p}p^{-4p} = \delta^*$ (say).

$$\sum_{l=1}^{p/2}(1/n)\sum_{i=1}^n(n\hat{M}_{i,2l})^{p/2l} = \sum_{l=1}^{p/2} e^{h(l)},$$

where $h(l) = \frac{p}{2l}\log n + \frac{p}{l(p-2l+2)}\log\delta^*$. We have $h'(l) = -\frac{p}{2l^2}\log n - \log\delta^* \frac{p(p-4l+2)}{l^2(p-2l+2)^2}$. Thus for $l \geq (p/4) + (1/2)$, $h'(l) \leq 0$ and so $h(l)$ is decreasing. Now for $l < (p/4) + (1/2)$, we have $\frac{p}{2l^2}\log n \geq -(\log\delta^*)\frac{p(p-4l+2)}{l^2(p-2l+2)^2}$, so again $h'(l) \leq 0$. Thus, $h(l)$ attains its maximum at $l = 1$, so $(36p)^{p+2}\sum_{l=1}^{p/2} e^{h(l)} \leq p(36p)^{p+3}n^{p/2}\delta^*$ giving us $(36p)^{p+2}\sum_{l=1}^{p/2}(n\hat{M}_{2l}^*)^{p/2l} \leq (cnp(\mu^3+\sigma^2))^{p/2}$. Thus we get from the Main Theorem that $E(f - Ef)^m \leq (cmn(\mu^3 + \sigma^2))^{\frac{m}{2}}$, from which Theorem (7) follows by the choice of $m = \lfloor \frac{t^2}{c_5 n(\mu^3+\sigma^2)} \rfloor$.

A. Lower Bound on Spread for Bin Packing

Suppose again $Y_1, Y_2, \ldots Y_n$ are the i.i.d. items. Suppose the distribution is :

$$\Pr\left(Y_1 = \frac{k-1}{k(k-2)}\right) = \frac{k-2}{k-1}\ ;\ \Pr\left(Y_1 = \frac{1}{k}\right) = \frac{1}{k-1}.$$

This is a "perfectly packable distribution" (well-studied class of special distributions) ($k-2$ of the large items and 1 of the small one pack.) Also, σ is small. But we can have number of $1/k$ items equal to $\frac{n}{k-1} - c\sqrt{\frac{n}{k}}$. Number of bins required $\geq \sum_i X_i = \frac{n}{k} + \frac{n}{k(k-1)} + c\sqrt{\frac{n}{k}}\left(\frac{1}{k}\left(\frac{k-1}{k-2} - 1\right)\right) \geq \frac{n}{k-1}$. So at least $c\sqrt{\frac{n}{k}}$ bins contain only $(k-1)/k(k-2)$ sized items (the big items). The gap in each such bin is at least $1/k$ for a total gap of $\Omega(\sqrt{n}/k^{3/2})$. On the other hand, if the number of small items is at least $n/(k-1)$, then each bin except two is perfectly fillable.

XI. LONGEST INCREASING SUBSEQUENCE

Let $Y_1, Y_2, \ldots Y_n$ be i.i.d., each distributed uniformly in $[0,1]$. We consider here $f(Y) =$ the length of the longest increasing subsequence (LIS) of Y. This is a well-studied problem. It is known that $Ef = (2+o(1))\sqrt{n}$. Here, we supply a (fairly simple) proof from Theorem (6) that f is concentrated in an interval of length $O(n^{1/4})$ with sub-Gaussian tails. Talagrand [18] gave the first (very simple) proof of this. [But by now better intervals of concentration, namely $O(n^{1/6})$ are known, using detailed arguments specific to this problem [1].] Our argument follows from two claims below. Call Y_i essential for Y if Y_i belongs to every LIS of Y (equivalently, $f(Y \setminus Y_i) = f(Y) - 1$.) Fix $Y_1, Y_2, \ldots Y_{i-1}$ and for $j \geq i$, let $a_j = \Pr(Y_j \text{ is essential for } Y|Y_1, Y_2, \ldots Y_{i-1})$

Claim 3. *$a_i, a_{i+1}, \ldots a_n$ form a non-decreasing sequence.*

Proof Let $j \geq i$. Consider a point ω in the sample space where Y_j is essential, but Y_{j+1} is not. Map ω onto ω' by swapping the values of Y_j and Y_{j+1}; this is clearly a 1-1 measure preserving map. If θ is a LIS of ω with $j \in \theta, j+1 \notin \theta$, then $\theta \setminus j \cup j+1$ is an increasing sequence in ω'; so $f(\omega') \geq f(\omega)$. If $f(\omega') = f(\omega)+1$, then an LIS α of ω' must contain both j and $j+1$ and so contains no k such that Y_k is between Y_j, Y_{j+1}. Now $\alpha \setminus j$ is an LIS of ω contradicting the assumption that j is essential for ω. So $f(\omega') = f(\omega)$. So, $j+1$ is essential for ω' and j is not. So, $a_j \leq a_{j+1}$. □

Claim 4. *$a_i \leq c/\sqrt{n-i+1}$.*

Proof $a_i \leq \frac{1}{n-i+1}\sum_{j\geq i} a_j$. Now $\sum_{j\geq i} a_j = a$ (say) is the expected number of essential elements among $Y_i, \ldots Y_n$ which is clearly at most $Ef(Y_i, Y_{i+1}, \ldots Y_n) \leq 4\sqrt{n-i+1}$, so the claim follows. $\square$ Z_i is a 0-1 random variable with $E(Z_i|Y_1, Y_2, \ldots Y_{i-1}) \leq c/\sqrt{n-i+1}$. Thus it follows (using (6) of section (III)) that

$$E(X_i^2|Y_1, Y_2, \ldots Y_{i-1}) \leq c/\sqrt{n-i+1}.$$

Clearly, $E(X_i^l|Y_1, Y_2, \ldots Y_{i-1}) \leq E(X_i^2|Y_1, Y_2, \ldots Y_{i-1})$ for $l \geq 2$, even. Thus we may apply the main Theorem with $\mathcal{E}_{il}$ equal to the whole sample space. Assuming $p \leq \sqrt{n}$, we see that (using $\sum_l (1/l^2) = O(1)$)

$$E(f - Ef)^p \leq (c_1 p)^{(p/2)+2} n^{p/4},$$

from which one can derive the asserted sub-Gaussian bounds.

XII. Discussion

The "sub-Gaussian" behaviour - $e^{-t^2\cdots}$ with the "correct" variance in remark (1) needs that the exponent of m in the upper bound in Theorem (1) be $\frac{m}{2}$. The well-known Burkholder type inequalities [9] cannot give this because of known lower bounds. One point here is that we treat carefully the different moments. Rewriting the upper bound we required in Theorem (1) as

$$\left(E(X_i^l|X_1 + X_2 + \ldots + X_{i-1})\right)^{1/l} \leq \left(\frac{n}{m}\right)^{\frac{1}{2}-\frac{1}{l}} \frac{l}{e},$$

we see that (for $m \leq n$,) the right hand side is highest at $l = 2$, so we put weaker requirements on the higher moments.

Another class of inequalities are the Efron-Stein inequalities, where one takes a high moment of the sum of squared variations of the function on changing one variable at a time-see [4] for a recent result on these lines. This does get the correct exponent of m, and is very useful if one can show that for any point in the sample space, not too many variables change the function too much. In contrast we only consider changing one variable. But even for the classical Longest Increasing Subsequence (LIS) problem, where for example, Talagrand's crucial argument is that only a small number $O(\sqrt{n})$ of elements (namely those in the current LIS) cause a decrease in the length of the LIS by their deletion, we are able to bound individual variations (in essence arguing that EACH variable has roughly only a $O(1/\sqrt{n})$ probability of changing the length of the LIS) sufficiently to get a concentration result.

Besides the situation like JL theorem, the Strong Negative correlation condition is also satisfied by the so-called "negatively associated" random variables ([7] for example). Variables in occupancy (balls and bins) problems, 0-1 variables produced by a randomized rounding algorithm of Srinivasan [17] etc. are negatively associated [11].

An interesting open question is whether there are good algorithms under the more general distributions for the TSP and other problems.

Acknowledgements : Thanks to Alan Frieze, Svante Janson, Manjunath Krishnapur, Claire Mathieu, Assaf Naor, Yuval Peres for helpful discussions.

References

[1] J. Baik, P. Deift and K. Johansson, On the Distribution of the length of the longest increasing subsequence of random permutations, Journal of the American Mathematical Society 12 (1999), no. 4, 1119–1178.

[2] B. Bollobás, Random Graphs, Cambridge Studies in advanced mathematics, 73 (2001)

[3] B. Bollobás, Martingales, isoperimetric inequalities and random graphs, in *Combinatorics*, Proceedings Eger., (1987) (Hajnal, et. al Eds) Colloq. Math. Soc. Janós Bolyai, **52** North-Holland, Amsterdam, pp 113-139.

[4] S. Boucheron, O. Bousquet, G. Lugosi and P. Massart, Moment inequalities for functions of independent random variables, The Annals of Probability, 2005 V0l 33, No. 2, p 514-560.

[5] Coffman, E. G., Jr., and Lueker, G. S., Probabilistic Analysis of Packing and Partitioning Algorithms, Wiley & Sons, 1991.

[6] S. Dasgupta and A. Gupta, An elementary proof of the Johnosn-Lindenstrauss Lemma, International Computer Science Institute, TR-99-006, (1999).

[7] D. Dubhashi and D. Ranjan, Balls and bins: A study in negative dependence, Random Structures and Algorithms, 13, 99-124 (1998)

[8] A. M. Frieze, On the independence number of random graphs, Discrete Mathematics, **81** pp171-175 (*183*).

[9] P. Hitczenko, Best constants in martingale version of Rosenthal's inequality, The Annals of Probability, (1990) Vol. 18, No. 4, p. 1656-1668.

[10] S. Janson, T. Luczak and A. Rucinski, Random Graphs, Wiley- Interscience Series in Discrete Mathematics and Optimization (2000).

[11] K. Joag-Dev and F. Proschan, Negative-association of random variables with applications, Annals of Statistics, **11** 286-295 (1983).

[12] C. McDiarmid. Concentration. In Probabilistic Methods for Algorithmic Discrete Mathematics, edited by M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, pp. 195248, Algorithms and Combinatorics 16. Berlin: Springer, 1998

[13] J. Csirik, D. S. Johnson, C. Kenyon, J. B. Orlin, P. W. Shor, and R. R. Weber On the Sum-of-Squares Algorithm for Bin Packing, . Thirty-Second Annual ACM Symposium on Theory of Computing (STOC), 208-217, 2000. Journal version in JACM, 53(1), 1-65, 2006.

[14] W. Rhee, Inequalities for the bin packing problem III, Optimization, 29 (1994) p 381-385

[15] W. Rhee and M. Talagrand, A sharp deviation for the stochastic Traveling salesman problem, Annals of Probability, 17, pp 1-8 (1989).

[16] E. Shamir and J. Spencer, Sharp concentration of the chromatic number of random graphs, Combinatorica **7**, p 121-129.

[17] A. Srinivasan, Distributions on level sets with applications to approximation algorithms, in the Proc. of the 42 nd IEEE Symposium on Foundations of Computer Science, (FOCS) 2001.

[18] M. Talagrand,Concentration of measure and isoperimetric inequalities in product spaces, Publications mathématiques de l'I.H.É.S., tome 81 (1995) p. 73-205.

[19] J. M. Steele, Probability Theory and Combinatorial Optimization, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM (1997)

[20] S. Vempala, The random projection method, DIMACS Series in Discrete Mathemtics and TCS, volume 65 (2000)