LOCAL TAIL BOUNDS
FOR FUNCTIONS OF INDEPENDENT RANDOM VARIABLES

Luc Devroye
School of Computer Science
McGill University
Montreal, Canada H3A 2A7

Gábor Lugosi
ICREA and Department of Economics
Pompeu Fabra University
Ramon Trias Fargas, 25-27
08005 Barcelona, Spain

ABSTRACT. It is shown that functions defined on $\{0, 1, \ldots, r-1\}^n$ satisfying certain conditions of bounded differences that guarantee subgaussian tail behavior also satisfy a much stronger "local" subgaussian property. For self-bounding and configuration functions we derive analogous locally subexponential behavior. The key tool is Talagrand's (1994) variance inequality for functions defined on the binary hypercube which we extend to functions of uniformly distributed random variables defined on $\{0, 1, \ldots, r-1\}^n$ for $r \geq 2$.

KEYWORDS AND PHRASES. Concentration inequalities, convex distance, configuration functions, hypercontractivity, Talagrand's inequality

1991 MATHEMATICS SUBJECT CLASSIFICATIONS: Primary 60F10.

RUNNING HEAD: LOCAL TAIL BOUNDS

Corresponding authors' address: Luc Devroye, School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The first author's research was sponsored by NSERC Grant A3456 and FQRNT Grant 90-ER-0291.

Introduction.

Concentration inequalities for functions of independent random variables establish upper bounds for the tail probabilities of such functions under general "smoothness" conditions, see, e.g., Talagrand (1995, 1996a, 1996b), Ledoux (1997, 2001), Boucheron, Lugosi, Massart (2000, 2003), McDiarmid (1998), etc. In this paper we take a closer look at the distribution of certain functions of independent random variables and show that the tail distribution exhibits a subgaussian (or subexponential) behavior in a stronger "local" sense in many cases when concentration inequalities predict a subgaussian (subexponential) tail.

First we consider real-valued functions defined on the binary hypercube $f : \{0,1\}^n \to \mathbb{R}$. If $X = (X_1, \ldots, X_n)$ is uniformly distributed on the hypercube, we are interested in the distribution of the random variable $f(X)$.

Our starting point is the following inequality, due to Talagrand (1994):

$$\mathbb{V}(f) \le \frac{9}{10} \sum_{i=1}^{n} \frac{\mathbb{E}\left(f(X) - f(X^{(i)})\right)^2}{1 + \log \frac{\sqrt{\mathbb{E}\left(f(X)-f(X^{(i)})\right)^2}}{\mathbb{E}\left|f(X)-f(X^{(i)})\right|}} \tag{1}$$

where $X^{(i)} = (X_1, \ldots, 1 - X_i, \ldots, X_n)$ is obtained by flipping the i-th bit of X and $\mathbb{V}(f)$ denotes the variance of the random variable $f(X)$. The constants shown here follow from a simple proof by Benjamini, Kalai, and Schramm (2003).

Note that Talagrand's inequality improves upon the well-known Efron-Stein inequality (see Efron and Stein, 1981; Rhee and Talagrand, 1986; Steele, 1986):

$$\mathbb{V}(f) \le \frac{1}{2} \sum_{i=1}^{n} \mathbb{E}\left(f(X) - f(X^{(i)})\right)^2 .$$

In Section 2 we show how to use Talagrand's inequality to prove "local" subgaussian concentration inequalities. As a simple example, we show that if $f : \{0,1\}^n \to \mathbb{R}$ is such that there exists a constant v such that $\sum_{i=1}^{n}(f(x) - f(x^{(i)}))_+^2 \le v$, then for all $k = 1, 2, 3, \ldots,$

$$a_{k+1} - a_k \le c\sqrt{v/k}$$

where a_k denotes a $1 - 2^{-k}$ quantile of $f(X)$ and c is a universal constant. The main argument is based on an observation of Benjamini, Kalai, and Schramm (2003) who show how Talagrand's inequality may be used to obtain exponential concentration inequalities. Even though Benjamini, Kalai, and Schramm do not mention the possibility of deriving local concentration inequalities, it is their argument which is at the basis of our proofs. The purpose of this paper is to elaborate on this argument and to derive local concentration inequalities under different conditions. In Sections 3, 4, and 5 various variants and extensions are introduced. In Section 3 local concentration inequalities are shown under different conditions that are satisfied for numerous natural examples such as *configuration functions* introduced by Talagrand (1995) of for *self-bounding functions*, see Boucheron, Lugosi, and Massart (2000), Maurer (2006), McDiarmid and Reed (2006).

In Section 4 Talagrand's inequality is extended from the binary hypercube to functions defined on $\{0, 1, \ldots, r-1\}^n$ under the uniform distribution. The main technical tool here is a suitable hypercontractive inequality proved by Alon, Dinur, Friedgut, and Sudakov (2004). This extension allows us to generalize the results of Sections 2 and 3 to functions defined on $\{0, 1, \ldots, r-1\}^n$.

In Section 5 we illustrate the use of the results of Section 4 by considering two classical, structurally similar, problems. We derive local concentration inequalities for the cost of the minimum weight spanning tree of a complete graph with random uniform weights on the edges and also for the assignment problem.

Functions with locally subgaussian behavior.

First we consider functions $f : \{0,1\}^n \to \mathbf{R}$ which satisfy the following properties: for all $x = (x_1, \ldots, x_n) \in \{0,1\}^n$

$$\sum_{i=1}^{n} (f(x) - f(x^{(i)}))_+^2 \le v \ , \qquad (*)$$

where v is a positive constant. (Here and throughout the paper, $a_+ = \max(a,0)$ and $a_- = \max(-a,0)$ denote the positive and negative parts of the real number a.) Clearly, if f is 1-Lipschitz under the Hamming distance then $v \le n$, but there are many interesting examples in which v is significantly smaller than n. It is well known (see Ledoux (1997), or Boucheron, Lugosi, and Massart (2003)) that for such functions

$$\mathbf{P}\{f(X) \ge \mathbf{E}f(X) + t\} \le e^{-t^2/4v} \ . \qquad (2)$$

Our basic result (Theorem 1) shows that tail quantiles of the random variable $f(X)$ are not far apart. In this sense, it is a local tail bound. For any $\alpha \in (0,1)$, define the α-quantile of f by

$$Q_\alpha = \inf\{z : \mathbf{P}\{f(X) \le z\} \ge \alpha\} \ .$$

In particular, we denote the median of $f(X)$ by $\mathbf{M}f = Q_{1/2}$.

Theorem 1. *Assume f satisfies (*) and let $B = \max_{x,i} |f(x) - f(x^{(i)})|$. Then for all $b > a \ge \mathbf{M}f$,*

$$b - a \le \sqrt{\frac{(72/5)v\mathbf{P}\{f(X) \in (a, b+B)\}}{\mathbf{P}\{f(X) \ge b\} \log \frac{e^2}{2\mathbf{P}\{f(X) \in (a,b+B)\}}}} \le \sqrt{\frac{(72/5)v\mathbf{P}\{f(X) > a\}}{\mathbf{P}\{f(X) \ge b\} \log \frac{e^2}{2\mathbf{P}\{f(X) > a\}}}} \ .$$

Proof. Define the function $g_{a,b} : \{0,1\}^n \to \mathbf{R}$ by

$$g_{a,b}(x) = \begin{cases} b & \text{if } f(x) \ge b \\ f(x) & \text{if } a < f(x) < b \\ a & \text{if } f(x) \le a \end{cases}$$

First observe that

$$\mathbf{V}(g_{a,b}(X)) \ge \frac{\mathbf{P}\{f(X) \ge b\}}{4}(b-a)^2 \ .$$

On the other hand, we may use Talagrand's inequality to obtain an upper bound for the variance of $g_{a,b}(X)$. To this end, observe that

$$\begin{aligned}
\mathbf{E}\left|g_{a,b}(X) - g_{a,b}(X^{(i)})\right| &= 2\mathbf{E}\left(g_{a,b}(X) - g_{a,b}(X^{(i)})\right)_+ \\
&= 2\mathbf{E}\left[\left(g_{a,b}(X) - g_{a,b}(X^{(i)})\right)_+ \mathbb{1}_{[f(X) \in (a,b+B)]}\right] \\
&\quad \text{(by the definition of } g_{a,b} \text{ and } B) \\
&\le 2\sqrt{\mathbf{E}\left(g_{a,b}(X) - g_{a,b}(X^{(i)})\right)_+^2}\sqrt{\mathbf{P}\{f(X) \in (a,b+B)\}} \\
&\quad \text{(by Cauchy-Schwarz)} \\
&= \sqrt{2\mathbf{E}\left(g_{a,b}(X) - g_{a,b}(X^{(i)})\right)^2}\sqrt{\mathbf{P}\{f(X) \in (a,b+B)\}} \ .
\end{aligned}$$

On the other hand,

$$\begin{aligned}\sum_{i=1}^n \mathbf{E}\left(g_{a,b}(X)-g_{a,b}(X^{(i)})\right)^2 &= 2\sum_{i=1}^n \mathbf{E}\left(g_{a,b}(X)-g_{a,b}(X^{(i)})\right)_+^2 \\ &= 2\mathbf{E}\left[\mathbb{1}_{[f(X)\in(a,b+B)]}\sum_{i=1}^n \left(g_{a,b}(X)-g_{a,b}(X^{(i)})\right)_+^2\right] \\ &\le 2v\mathbf{P}\{f(X)\in(a,b+B)\}\end{aligned}$$

where in the last step we used the fact that (*) implies that

$$\sum_{i=1}^n \left(g_{a,b}(X)-g_{a,b}(X^{(i)})\right)_+^2 \le \sum_{i=1}^n \left(f(X)-f(X^{(i)})\right)_+^2 \le v\ .$$

Combining the lower bound for the variance with the upper bound obtained by Talagrand's inequality yields the claim. □

To make Theorem 1 more transparent, we state a simple corollary for quantiles of $f(X)$. Using $\mathbf{P}\{f(X) > Q_{1-\gamma}\} \le \gamma$ and $\mathbf{P}\{f(X) \ge Q_{1-\delta}\} \ge \delta$, Theorem 1 implies the following bound for the distance between any two quantiles in the upper tail:

Theorem 2. *Assume f satisfies (*). Then for all $\delta < \gamma \le 1/2$,*

$$Q_{1-\delta} - Q_{1-\gamma} \le \sqrt{\frac{(72/5)v\gamma}{\delta\log\frac{e^2}{2\gamma}}}\ .$$

In particular, by choosing $\gamma = 2^{-k}$ and $\delta = 2^{-(k+1)}$ for some integer $k \ge 1$ and introducing

$$a_k = Q_{1-2^{-k}}\ ,$$

we get

$$a_{k+1} - a_k \le \frac{12}{\sqrt{5}}\sqrt{\frac{v}{(k-1)\log 2 + 2}} \le 7\sqrt{\frac{v}{k}}\ . \tag{3}$$

Summing over $k = 1, 2, \ldots, m-1$ and using $\sum_{k=1}^{m-1}(k-1)^{-1/2} \le \int_0^{m-1} x^{-1/2}dx = 2\sqrt{m-1}$, we obtain

$$a_m \le a_1 + 14\sqrt{v(m-1)}\ ,$$

recovering (up to a constant factor) the subgaussian concentration inequality (2) for f. However, Theorem 2 shows a subgaussian behavior in a significantly stronger sense. If $f(X)$ was a normal random variable with variance v, then one would have $a_k \sim \sqrt{2vk\log 2}$ and $a_{k+1} - a_k \sim \sqrt{v\log 2/k}$. This (up to a constant factor) is precisely of the form of the upper bound (3) for a general function f satisfying (*). Thus, the whole quantile sequence $\{a_k\}$ is a *contraction* of that of a normal random variable of variance a constant times v. (We say that a sequence $\{x_n\}$ is a contraction of another sequence $\{y_n\}$ if for every $n = 1, 2, \ldots$, $|x_{n+1} - x_n| \le |y_{n+1} - y_n|$.)

Remark 1. (constants.) Even though we offer explicit numerical constants in the inequalities derived throughout the paper, no optimality of these values is claimed. In fact, quite often we sacrifice better constants for convenience in the notation or for simpler arguments.

Example 1. (convex distance.) One of the main examples of a function satisfying (*) is Talagrand's *convex distance* (Talagrand, 1995) defined as follows. Let $A \subset \{0,1\}^n$ and define f as

$$f(x) = \sup_{\alpha \in [0,\infty)^n : \|\alpha\|=1} \inf_{y \in A} \sum_{i : x_i \neq y_i} |\alpha_i|$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Talagrand shows that for any set A with $\mathbb{P}\{X \in A\} \geq 1/2$,

$$\mathbb{P}\{f(X) \geq t\} \leq 2e^{-t^2/4} \ .$$

(Note that Talagrand's result is true in any product space with product measure.) It is shown by Boucheron, Lugosi, and Massart (2003) that f satisfies (*) with $v = 1$. This implies that for all $k = 1, 2, 3, \ldots,$,

$$a_{k+1} - a_k \leq \frac{7}{\sqrt{k}} \ .$$

Example 2. (largest eigenvalue of a random graph.) Let $f(X)$ denote the largest eigenvalue of the adjacency matrix of a random graph $G(m, 1/2)$ on m vertices such that each edge appears with probability 1/2. Thus, $n = \binom{m}{2}$ and $X_i = 1$ if and only if edge i is present in the graph. Füredi and Komlós (1981) showed that $f(X)$ is asymptotically normally distributed with expectation $m/2$ and variance 1/2. Alon, Krivelevich, and Vu (2002) show that $f(x)$ satisfies (*) with $v = 4$ (see also Maurer (2006)) and conclude that $a_k \leq \mathbf{M}f(X) + \sqrt{32(k+2)\log 2}$. Theorem 2 implies the nonasymptotic local subgaussian estimate

$$a_{k+1} - a_k \leq \frac{14}{\sqrt{k}}$$

for $k = 1, 2, \ldots$. Note that Alon, Krivelevich, and Vu also prove a concentration result for the r-th largest inequality of the form $a_k \leq \mathbf{M}f(X) + Cr\sqrt{k}$. Their argument may be combined with ours to obtain an analogous local concentration inequality.

Example 3. (rademacher averages.) Another example is a *Rademacher average* of the form

$$f(x) = \sup_{\alpha \in A} \sum_{i=1}^{n} \alpha_i (x_i - 1/2)$$

where $A \subset \mathbb{R}^n$ is a set of vectors α with $\|\alpha\| \leq 1$. It is easy to see that condition (*) is satisfied with $v = 1$.

Remark 2. (asymmetric distributions.) We note here that Talagrand proved his inequality (1) in a more general setup in which the components X_i of X are i.i.d. Bernoulli(p) random variables for some $p \in (0,1)$. In this more general case Theorem 2 becomes

$$Q_{1-\delta} - Q_{1-\gamma} \leq \sqrt{\frac{Cv\gamma}{\delta \log \frac{1}{2\gamma}} \log \frac{1}{p(1-p)}}$$

for some constant C.

One obtains a corollary of a slightly different flavor by choosing, in Theorem 1, $a = k$ and $b = k+1$ for some integer $k \geq \mathbf{M}f$, Theorem 1 implies the following local lower bound for the distribution of f:

COROLLARY 1. *Assume f satisfies (*). Then for all $k \geq \mathbb{E}f + \sqrt{4v \log 2}$,*

$$\frac{q_k}{\sum_{i \geq k+1} q_i} + 1 \geq \frac{5}{288} \frac{(k - \mathbb{E}f)^2}{v^2} + \frac{5}{72v} \log \frac{e^2}{2}$$

where $q_k = \mathbb{P}\{f(X) \in [k, k+1)\}$.

PROOF. This follows immediately by noting that, on the one hand by Theorem 1, for $k \geq \mathbf{M}f$,

$$\sum_{i \geq k} q_i \leq (72/5)v \frac{q_k + \sum_{k+1 \leq i \leq k+B+1} q_i}{\log \frac{e^2}{2(q_k + \sum_{k+1 \leq i \leq k+B+1} q_i)}} \leq (72/5)v \frac{q_k + \sum_{i \geq k+1} q_i}{\log \frac{e^2}{2(q_k + \sum_{i \geq k+1} q_i)}}$$

so that

$$q_k + \sum_{i \geq k+1} q_i \geq \frac{e^2}{2} \exp\left(-(72/5)v \left(\frac{q_k}{\sum_{i \geq k+1} q_i} + 1 \right) \right) .$$

By the concentration inequality (2), for all $k \geq \mathbb{E}f$,

$$q_k + \sum_{i \geq k+1} q_i = \mathbb{P}\{f(X) \geq k\} \leq e^{-(k - \mathbb{E}f)^2/4v} .$$

Since $\mathbf{M}f \leq \mathbb{E}f + \sqrt{4v \log 2}$, combining the upper and lower bounds implies the corollary. $\square$

REMARK 3. (MONOTONICITY OF THE TAIL.) An obvious corollary is that $q_{k+1} \leq q_k$ whenever $k \geq \mathbb{E}f + (25/\sqrt{5})v$.

In some applications, even though $(*)$ is not satisfied, the similar condition

$$\sum_{i=1}^{n} (f(x) - f(x^{(i)}))_-^2 \leq v , \qquad (*^-)$$

holds. For such cases the next analog of Theorem 1 is true. The proof is omitted as it is a straightforward modification. In Section 5 we present some applications of this result.

THEOREM 3. *Assume f satisfies $(*^-)$ and let $B = \max_{x,i} |f(x) - f(x^{(i)})|$. Then for all $b > a \geq \mathbf{M}f$,*

$$b - a \leq \sqrt{\frac{(72/5)v\mathbb{P}\{f(X) \in (a-B, b)\}}{\mathbb{P}\{f(X) \geq b\} \log \frac{e^2}{2\mathbb{P}\{f(X) \in (a-B,b)\}}}} .$$

In particular, for all $\delta < \gamma \leq 1/2$, by taking $a = Q_{1-\gamma} + B$ and $b = Q_{1-\delta}$, we have

$$Q_{1-\delta} - Q_{1-\gamma} \leq B + \sqrt{\frac{(72/5)v\gamma}{\delta \log \frac{e^2}{2\gamma}}} .$$

Configuration functions.

In this section we consider functions defined on the binary hypercube. Just as in Section 1, let $f : \{0,1\}^n \to \mathbb{R}$ and assume that X is uniformly distributed over $\{0,1\}^n$.

Often, the sum of the squared changes appearing in condition (*) cannot be bounded by a constant but it can be related to the value of the function itself. Consider the following conditions:

$$|f(x) - f(x^{(i)})| \le B \quad \text{for all } x \text{ and } i \text{ and } \quad \sum_{i=1}^{n} (f(x) - f(x^{(i)}))_+^2 \le \phi(f(x)) \qquad (**)$$

where ϕ is a fixed nonnegative nondecreasing function defined on the reals. In many applications, such as for configuration functions, one may take ϕ to be the identity and in some others it has the form $\phi(u) = au + b$ (see Talagrand (1995), Boucheron, Lugosi, and Massart (2000, 2003), Devroye (2002)). For example, it is shown by Boucheron, Lugosi, and Massart (2003) (for various extensions see also Maurer, 2006; Reed and McDiarmid, 2006) that if (**) is satisfied with $\phi(u) = u$ and $B \le 1$, then

$$\mathbb{P}\{f(X) \ge \mathbb{E}f(X) + t\} \le e^{-t^2/(2\mathbb{E}f(X)+2t/3)} .$$

Boucheron, Lugosi, and Massart (2003) offer concentration inequalities for the case when $\phi(u) = cu^\alpha$ for some $\alpha \in (0,2)$.

A straightforward modification of the proof of Theorem 1 yields the following:

Theorem 4. *Assume f satisfies (**) and let $B = \max_{x,i} |f(x) - f(x^{(i)})|$. Then for all $b > a \ge \mathbf{M}f$,*

$$b - a \le \sqrt{\frac{(72/5)\phi(b+B)\mathbb{P}\{f(X) > a\}}{\mathbb{P}\{f(X) \ge b\} \log \frac{e^2}{2\mathbb{P}\{f(X)>a\}}}} .$$

Also, for all $\delta < \gamma \le 1/2$,

$$Q_{1-\delta} - Q_{1-\gamma} \le \sqrt{\frac{(72/5)\phi(Q_{1-\delta} + B)\gamma}{\delta \log \frac{e^2}{2\gamma}}} .$$

In particular, recalling the notation $a_k = Q_{1-2^{-k}}$,

$$a_{k+1} - a_k \le 7\sqrt{\frac{\phi(a_{k+1} + B)}{k}} .$$

Example 4. (self-bounding functions.) In many interesting applications, ϕ may be taken to be the identity function and $B = 1$. These functions have been called *self-bounding*, see Boucheron, Lugosi, and Massart (2000), Maurer (2006), McDiarmid and Reed (2006). In general, if $\phi(u)$ is linear, then by the above-mentioned concentration inequality, for all $k \ge \mathbb{E}f(X)$, $a_k \le ck$, and therefore

$$a_{k+1} - a_k \le C$$

where c, C are constants. Thus, in this case the quantile sequence $\{a_k\}$ is a contraction of that corresponding to an *exponentially* distributed random variable with parameter $O(1)$, in a similar sense that functions satisfying (*) had a quantile sequence contracting a gaussian quantile sequence.

Example 5. (longest increasing subsequence.) Let now $x = (x_1, \ldots, x_n) \in \{0, 1, \ldots, r-1\}^n$ and define $f(x)$ to be the length of the longest increasing subsequence of $x_1, \ldots, x_n$, that is, the largest positive integer m for which there exist $1 \le i_1 < \cdots < i_m \le n$ such that $x_{i_1} \le x_{i_2} \le \ldots \le x_{i_m}$. Tracy and Widom (2001) and Johansson (2001) showed that if X is uniformly distributed over $\{0, 1, \ldots, r-1\}^n$, then $(f(X) - n/r)/\sqrt{2n/r}$ converges, in distribution, to a random variable whose distribution depends on r (see also Its, Tracy, and Widom, 2001). In the binary case (i.e., when $r = 2$), $f(x)$ is the longest subsequence of the form $000 \cdots 00111 \cdots 11$, and Theorem 4 may be readily used. It is immediate to see that f satisfies (**) with $B = 1$ and $\phi(u) = u$ and therefore Theorem 4 implies a nonasymptotic local subexponential concentration inequality. (To see why (**) is satisfied, fix a maximal increasing subsequence in x and observe that $(f(x) - f(x^{(i)}))_+ = 0$ whenever x_i is not in this maximal sequence.) The same inequality holds when $f(x) = \log_2 N(x)$ where $N(x)$ is the number of all increasing subsequences of x. The fact that $\log_2 N(x)$ satisfies (**) with $B = 1$ and $\phi(u) = u$ was observed by Boucheron, Lugosi, and Massart (2000). If $r > 2$, one may use the results of Section 4 below to obtain analogous bounds.

Remark 4. (concentration inequalities.) The recursion for the sequence $\{a_k\}$ given by Theorem 4 allows one to derive concentration inequalities for general functions ϕ. We illustrate this for the example when $\phi(u) \le cu^\alpha$ for some $c > 0$ and $\alpha \in [0, 2]$. Then Theorem 4 implies, after some work, that there exist constants C, t_0 such that for $t \ge t_0$,

$$\mathbb{P}\{f(X) \ge t\} \le \begin{cases} Ce^{-t^{2-\alpha}/C} & \text{if } 0 \le \alpha < 2 , \\ Ce^{-(\log t)^2/C} & \text{if } \alpha = 2 . \end{cases}$$

The case $\alpha < 2$ has already been dealt with by Boucheron, Lugosi, and Massart (2003), but the $\alpha = 2$ case seems to be new.

Functions defined on the r-ary hypercube.

The purpose of this section is to extend the results of Theorems 1, 2, and 3 to functions f defined on the r-ary cube $\{0, 1, \ldots, r-1\}^n$, equipped with the uniform distribution. In order to do this, we need to generalize Talagrand's variance inequality to this case. In particular, we prove the following:

Theorem 5. *Let $r \ge 2$ be a positive integer and let $f : \{0, 1, \ldots, r-1\}^n \to \mathbb{R}$ be a real-valued function. Suppose $X = (X_1, \ldots, X_n)$ is uniformly distributed on $\{0, 1, \ldots, r-1\}^n$. For $1 \le i \le n$, $0 \le j \le r-1$ and for each $x = (x_1, \ldots, x_n)$, denote $x_{i,j} = (x_1, \ldots, x_{i-1}, x_i \oplus j, x_{i+1}, \ldots, x_n)$ where $\oplus$ stands for addition modulo r. Writing*

$$\Delta_i f(x) = f(x) - \frac{1}{r} \sum_{j=0}^{r-1} f(x_{i,j}) ,$$

we have

$$\mathbb{V}(f) \le 10(\log C_r) \sum_{i=1}^{n} \frac{\mathbb{E}\,(\Delta_i f(X))^2}{1 + \log \frac{\sqrt{\mathbb{E}(\Delta_i f(X))^2}}{\mathbb{E}|\Delta_i f(X)|}} .$$

where $C_r = (9/2)r^3$ is the constant of Lemma 1 below.

As a consequence, Theorems 1,2,3, and 4 may now be extended to functions defined on $\{0, 1, \ldots, r-1\}^n$ with the only difference that in the conditions on f, $f(x) - f(x^{(i)})$ is replaced by $\Delta_i f(x)$ and the upper bounds in all four theorems are multiplied by $(10/3)\sqrt{\log C_r}$. For example, we will use the following result in Section 5:

COROLLARY 2. *Assume $f : \{0,1,\ldots,r-1\}^n \to \mathrm{R}$ is such that there exists $v > 0$ such that*

$$\sum_{i=1}^{n} (\Delta_i)_-^2 \le v$$

and let $B = \max_{x,i} |\Delta_i f(x)|$. Then for all $k = 1,2,3,\ldots,$

$$a_{k+1} - a_k \le B + 24\sqrt{\log C_r}\sqrt{\frac{v}{k}}\ .$$

The proof of Theorem 5 is analogous to Talagrand's (1994) original argument which was based on the Beckner-Bonami hypercontractive inequality (see Bonami, 1970; Beckner, 1975) of Fourier analysis on the binary hypercube. Here we use an extension of this inequality to functions defined on $\{0,1,\ldots,r-1\}^n$ due to Alon, Dinur, Friedgut, and Sudakov (2004) which we recall below.

For any $S = (S_1,\ldots,S_n) \in \{0,1,\ldots,r-1\}^n$, define the function

$$u_S(x) = \omega^{\langle S,x\rangle}$$

where $\omega = e^{2\pi i/r}$ and $\langle S,x\rangle = \sum_{i=1}^n S_i x_i \bmod r$. It is easy to see (see Alon, Dinur, Friedgut, and Sudakov (2004)) that the u_S form an orthonormal basis of the space of complex-valued functions defined over $\{0,1,\ldots,r-1\}^n$. To simplify notation, we will write

$$\int f = \frac{1}{r^n} \sum_{x\in\{0,1,\ldots,r-1\}^n} f(x) \quad \text{and} \quad \|f\|_q = \left(\int f^q\right)^{1/q}\ .$$

Denote by

$$\widehat{f}(S) = \int f\overline{u}_S$$

the Fourier coefficients of f where $\overline{u}_S$ stands for the complex conjugate of u_S. A key ingredient of the proof is the following hypercontractive inequality:

LEMMA 1. *(Alon, Dinur, Friedgut, and Sudakov (2004)). For any $f : \{0,1,\ldots,r-1\}^n \to \mathrm{R}$ and $k = 1,\ldots,n$,*

$$\left\| \sum_{S:|S|\le k} \widehat{f}(S) u_S \right\|_4 \le C_r^k \left(\sum_{S:|S|\le k} \widehat{f}(S)^2 \right)^{1/2}$$

where $C_r = (9/2)r^3$.

PROOF OF THEOREM 5. Writing $f_{i,j}(x) = f(x_{i,j})$, it is easy to see that $\widehat{f}_{i,j}(S) = \widehat{f}(S)\omega^{jS_i}$. Thus,

$$\frac{1}{r}\sum_{j=0}^{r-1} \widehat{f}_{i,j}(S) = \begin{cases} \widehat{f}(S) & \text{if } S_i = 0\ , \\ 0 & \text{if } S_i \ne 0\ , \end{cases}$$

and therefore the Fourier coefficients of $\Delta_i f$ satisfy

$$\widehat{\Delta_i f}(S) = \begin{cases} 0 & \text{if } S_i = 0\ , \\ \widehat{f}(S) & \text{if } S_i \ne 0\ . \end{cases}$$

This and Parseval's identity imply that

$$\mathrm{V}(f) = \|f\|_2^2 - \left(\int f\right)^2 = \sum_{S\neq\overline{0}} \widehat{f}(S)^2 = \sum_{i=1}^{n}\sum_{S\neq\overline{0}} \frac{\widehat{\Delta_i f}(S)^2}{|S|}$$

where $|S|$ denotes the number of nonzero components of S and $\overline{0}$ is the all-zero vector.

Thus, in order to prove the theorem, it suffices to show that for any $f : \{0,1,\ldots,r-1\}^n \to \mathrm{R}$,

$$\sum_{S\neq\overline{0}} \frac{\widehat{f}(S)^2}{|S|} \le 10\log C_r \frac{\|f\|_2^2}{1+\log\frac{\|f\|_2}{\|f\|_1}}$$

which is what we do in the remaining part of the proof. Fix $k \le n$ and observe that

$$\begin{aligned}
\sum_{S:|S|=k} \widehat{f}(S)^2 &= \int \left(\sum_{S:|S|=k} \widehat{f}(S)u_S\right) f \\
&\le \left\|\sum_{S:|S|=k} \widehat{f}(S)u_S\right\|_4 \cdot \|f\|_{4/3} \quad \text{(by Hölder)} \\
&\le C_r^k \left(\sum_{S:|S|=k} \widehat{f}(S)^2\right)^{1/2} \cdot \|f\|_{4/3} \quad \text{(by Lemma 1)}\ .
\end{aligned}$$

This implies

$$\sum_{S:|S|=k} \widehat{f}(S)^2 \le C_r^{2k}\|f\|_{4/3}^2$$

and we have, for all positive integers m,

$$\sum_{S:1\le|S|\le m} \frac{\widehat{f}(S)^2}{|S|} \le \|f\|_{4/3}^2 \sum_{k=1}^{m} \frac{C_r^{2k}}{k} \le K\frac{C_r^{2m}}{m}\|f\|_{4/3}^2$$

where $K = \frac{36^2/2}{36^2/2-1}$. At the last step we used the fact that $C_r \ge 36$ and therefore $C_r^{2(k+1)}/(k+1) \ge (36^2/2)C_r^{2k}/k$. Now we may write

$$\begin{aligned}
\sum_{S\neq\overline{0}} \frac{\widehat{f}(S)^2}{|S|} &= \sum_{S:1\le|S|\le m} \frac{\widehat{f}(S)^2}{|S|} + \sum_{S:|S|>m} \frac{\widehat{f}(S)^2}{|S|} \\
&\le K\frac{C_r^{2m}}{m}\|f\|_{4/3}^2 + \frac{1}{m+1}\sum_{S:|S|>m} \widehat{f}(S)^2 \\
&\le \frac{1}{m+1}\left(2KC_r^{2m}\|f\|_{4/3}^2 + \|f\|_2^2\right)\ .
\end{aligned}$$

Now we choose m as the largest integer such that $C_r^{2m}\|f\|_{4/3}^2 \le e^{1/3}\|f\|_2^2$ so that

$$m+1 \ge \frac{\log\left(e^{1/3}\|f\|_2/\|f\|_{4/3}\right)}{\log C_r}$$

and

$$\sum_{S\neq\overline{0}} \frac{\widehat{f}(S)^2}{|S|} \leq \frac{1}{m+1} \cdot (2K+1)\|f\|_2^2 \leq \frac{(2K+1)\|f\|_2^2 \log C_r}{\log\left(e^{1/3}\|f\|_2/\|f\|_{4/3}\right)} \ .$$

The proof is finished by observing that

$$\int f^{4/3} = \|f^{3/2}\|_{8/9}^{8/9} \leq \|f^{3/2}\|_1^{8/9} \leq \left(\|f\|_1 \cdot \|f\|_2^2\right)^{4/9}$$

by the Cauchy-Schwarz inequality, and therefore $\|f\|_{4/3}^3 \leq \|f\|_1 \cdot \|f\|_2^2$, which is equivalent to

$$\frac{e\|f\|_2}{\|f\|_1} \leq \left(\frac{e^{1/3}\|f\|_2}{\|f\|_{4/3}}\right)^3 \ .$$

□

Remark 5. (logarithmic sobolev inequalities). An alternative route, yielding better numerical constants than Lemma 1, would be to use a sharp logarithmic Sobolev inequality of Diaconis and Saloff-Coste (1996, Theorem A.1) which implies hypercontractivity by Gross' theorem, see Gross (1993).

Minimum weight spanning tree and the assignment problem.

In this section we derive local concentration bounds for two classical problems: the minimum weight spanning tree and the assignment problem. In these examples the random variables of interest are functions of independent random variables uniformly distributed in $[0,1]$. By simple discretization we may approximate them by functions defined over $\{0,1,\ldots,r-1\}^n$ and use the result of the previous section. Since in Corollary 2 the dependence on r is only logarithmic, we may take r to be quite large (proportional to n in these cases) and still obtain meaningful results.

Concentration inequalities for both cases may be derived, for example, by Talagrand's (1995) results. In fact, Talagrand works out the case of the assignment problem. In order to conveniently use general concentration inequalities, Talagrand uses a truncation argument, a technique we also adopt below. Interestingly, the proofs in both examples below are identical and use simple structural properties of the function at hand.

Example 6. (minimum weight spanning tree.) Consider the random variable T_m defined as the sum of weights on the minimum spanning tree of the complete graph K_m with independent uniformly distributed (on $[0,1]$) weights $Y_{i,j}$ $(1 \le i < j \le m)$ on the edges. A classical result of Frieze (1985) shows that $\lim_{m\to\infty} \mathbb{E}T_m = \zeta(3)$. Janson (1995) and Wästlund (2005c) prove that if the edge weights are exponentially distributed with parameter 1, then $\sqrt{m}(T_m - \zeta(3))$ converges, in distribution, to a centered normal random variable with variance $6\zeta(4) - 4\zeta(3)$. Here we study the related random variable $\overline{T}_m$ obtained when the $Y_{i,j}$ are replaced by $\min(Y_{i,j}, \delta_m)$ where $\delta_m > 0$ is a small positive number. Note that if $\delta_m = c \log m/m$ for some $c > 1$ then $T_m = \overline{T}_m$ with high probability. In order to see this just observe that $T_m \ne \overline{T}_m$ implies that the largest edge weight in the minimum spanning tree is greater than δ_m. But this is just the probability that the random graph $G(m, \delta_m)$ is not connected which is at most $2\left(e^{m^{(1-c)/2}} - 1\right) + 2^{m+1}m^{-(c-1)m/4}$ (see, Erdős and Rényi, 1960; Palmer, 1985), which is at most $4m^{-c/4}$, if $c \ge 2$.

To be able to use Corollary 2, we need to approximate T_m by a function defined on $\{0, 1, \dots, r-1\}^n$ under the uniform distribution where $n = \binom{m}{2}$. In order to do this, we replace the random variables $Y_{i,j}$ by their "discretized" approximation $\lfloor rY_{i,j}\rfloor/r$. If we denote the cost of the minimum spanning tree defined by the edge costs $\min\left(\lfloor rY_{i,j}\rfloor/r, \delta_m\right)$ by $\widetilde{T}_m$, then clearly $|\overline{T}_m - \widetilde{T}_m| \le m/r$. The random variable $\widetilde{T}_m$ may now be considered as a function of $n = \binom{m}{2}$ independent variables $X_{i,j}$, all uniformly distributed on $\{0, 1 \dots, r-1\}$, by defining $\lfloor rY_{i,j}\rfloor = X_{i,j}$. Now we may use Corollary 2. Clearly, we may take $B = \delta_m$. On the other hand,

$$\sum_{1\le i<j\le m} \left(\Delta_{i,j}\right)_-^2 \le m\delta_m^2$$

and therefore, denoting by $\widetilde{a}_k$ the $1 - 2^{-k}$-quantile of $\widetilde{T}_m$, we obtain

$$\widetilde{a}_{k+1} - \widetilde{a}_k \le \delta_m + 24\sqrt{\frac{m\delta_m^2}{k}}\sqrt{\log(9r^3/2)}\ .$$

This, in turn implies that if $\overline{a}_k$ denotes the $1 - 2^{-k}$-quantile of $\overline{T}_m$, then, for all $k = 1, 2, 3, \dots$,

$$\overline{a}_{k+1} - \overline{a}_k \le 2m/r + \delta_m + 24\sqrt{\frac{m\delta_m^2}{k}}\sqrt{\log(9r^3/2)}\ .$$

By choosing, say, $r = m^2$ and $\delta_m = c\log m/m$ for some constant $c > 1$, we obtain

$$\overline{a}_{k+1} - \overline{a}_k \le C\left(\sqrt{\frac{\log^3 m}{mk}} + \frac{\log m}{m}\right)$$

for a constant C depending on c only. This inequality shows a local subgaussian behavior whenever $k \le m\log m$. It may be regarded as a local nonasymptotic version of the limit theorem of Janson and Wästlund, up to the logarithmic factors we needed to give up for technical reasons. For larger values of k the second term dominates the first one, which corresponds to a subexponential behavior in the far tail. We do not know if this term is necessary. In order to convert this into a useful bound for the original problem T_m, one needs to choose c so large that the bound $\mathbb{P}\{T_m \ne \overline{T_m}\} \le 4m^{-c/4}$ does not dominate 2^{-k}. Choosing $c = \max(2, 4(k+2)\log 2/\log m)$, one obtains

$$a_{k+1} - a_{k-1} \le \begin{cases} \frac{2}{m} + \frac{4(k+2)\log 2}{m} + 96\log 2\sqrt{\frac{3(k+2)}{m}\log\frac{9m^6}{2}} & \text{if } k+2 > \frac{\log m}{2\log 2} \\ \frac{2}{m} + \frac{2\log m}{m} + 28\sqrt{\frac{\log^2 m}{km}\log\frac{9m^6}{2}} & \text{if } k+2 \le \frac{\log m}{2\log 2} \end{cases}$$

In order to compare this local bound to concentration inequalities, note that Theorem 7 of Boucheron, Lugosi, and Massart (2003) implies that $\mathbb{P}\{\overline{T}_m \geq \mathbb{E}\overline{T}_m + t\} \leq e^{-t^2 m/(4(e-1)c^2 \log^2 m)}$, or in other words, that

$$\overline{a}_k \leq \mathbb{E}\overline{T}_m + \sqrt{\frac{kc^2 \log^2 m \log 2}{4(e-1)m}} \ .$$

Again, choosing $c = \max(2, 4(k+2)\log 2/\log m)$, one obtains

$$a_{k-1} \leq \mathbb{E}\overline{T}_m + \sqrt{\frac{\log 2}{e-1}} \max\left(\sqrt{\frac{k\log^2 m}{m}}, \sqrt{\frac{4(k+2)^3}{m}}\right) \ .$$

By summing the "local" inequality in k, one obtains a concentration inequality that is only slightly weaker than the one derived here, as we get an extra factor of $\sqrt{\log m}$. This is due to the approximation by discretization, necessary to apply Corollary 2.

Example 7. (the assignment problem.) In the assignment problem, given an $m \times m$ array $\{Y_{i,j}\}_{m\times m}$ of independent random variables distributed uniformly on $[0,1]$, one considers the random quantity

$$Z_m = \min_{\pi} \sum_{i=1}^{m} Y_{i,\pi(i)}$$

where the minimum is taken over all permutations π of $\{1,\ldots,m\}$. Culminating a long series of partial results, Aldous (2001) showed that $\lim_{m\to\infty} \mathbb{E}Z_m = \zeta(2)$. In the case when the $Y_{i,j}$ are exponentially distributed with parameter 1, Linusson and Wästlund (2004) and Nair, Prabhakar, and Sharma (2005) independently proved that for all m, $\mathbb{E}Z_m = \sum_{i=1}^{m} i^{-2}$. See also Wästlund (2005a). Wästlund (2005b) also derives an explicit formula for the variance of Z_m. In particular, he proves that $\mathbb{V}(Z_m) = 4(\zeta(2) - \zeta(3))/m + O(m^{-1/2})$. Talagrand (1995) proves (in the uniform model) an exponential concentration inequality very similar to the one described for the minimum weight spanning tree above.

In fact, in order to get local concentration inequalities, one may proceed exactly as we did in the previous example: first one replaces the $Y_{i,j}$ by the truncated variables $\min(Y_{i,j}, \delta_m)$. If $\overline{Z}_m$ denotes the cost of the optimal assignment based on the truncated variables, then Proposition 10.3 of Talagrand (1995) implies that there exists a constant K such that $\mathbb{P}\{Z_m \neq \overline{Z}_m\} \leq e^{-m\delta_m/K}$, an inequality that is completely analogous to the one we used in the study of the minimum weight spanning tree. Second, we use the discretized approximation of the truncated variables. Then just as for the minimum weight spanning tree, we may take $B = \delta_m$ in Corollary 2 and observe that

$$\sum_{1\leq i<j\leq m} \left(\Delta_{i,j}\right)_-^2 \leq m\delta_m^2$$

which leads to inequalities completely analogous to those obtained for the minimum weight spanning tree example above. In particular, if a_k denotes the $1-2^{-k}$ quantile of Z_m, then there exists a constant C such that

$$a_{k+1} - a_{k-1} \leq C \max\left(\frac{k}{m} + \sqrt{\frac{k\log m}{m}}, \frac{\log m}{m} + \sqrt{\frac{\log^3 m}{km}}\right) \ .$$

References.

D. Aldous. The $\zeta(2)$ Limit in the Random Assignment Problem. *Random Structures and Algorithms*, vol. 18, pp. 381-418, 2001.

N. Alon, I. Dinur, E. Friedgut and B. Sudakov. Graph products, Fourier analysis and spectral techniques. *Geometric and Functional Analysis*, vol. 14, pp. 913-940, 2004.

N. Alon, M. Krivelevich, and V.H. Vu. On the concentration of eigenvalues of random symmetric matrices. *Israel Math. Journal*, vol. 131, pp. 259–267, 2002.

W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, vol. 102, pp. 159–182, 1975.

I. Benjamini, G. Kalai, and O. Schramm. First passage percolation has sublinear distance variance. *Annals of Probability* vol. 31, pp. 197–1978, 2003.

A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. Inst. Fourier (Grenoble)* vol. 20, pp. 335–402 (1970).

S. Boucheron, G. Lugosi, and P. Massart, "A sharp concentration inequality with applications in random combinatorics and learning," *Random Structures and Algorithms*, vol. 16, pp. 277–292, 2000.

S. Boucheron, G. Lugosi, and P. Massart, "Concentration inequalities using the entropy method," *Annals of Probability*, vol. 31, pp. 1583–1614, 2003.

L. Devroye, "Laws of large numbers and tail inequalities for random tries and Patricia trees," *Journal of Computational and Applied Mathematics*, vol. 142, pp. 27–37, 2002.

L. Devroye and G. Lugosi. *Combinatorial Methods in Density Estimation.* Springer, New York, 2000.

P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Annals of Applied Probability*, vol. 6, pp. 695–750, 1996.

B. Efron and C. Stein. The jackknife estimate of variance. *Annals of Statistics*, vol. 9, pp. 586–596, 1981.

P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17–61, 1960.

A. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Mathemathics*, vol. 10, pp. 47–56, 1985.

Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica* vol. 1, pp. 233–241, 1981.

L. Gross. Logarithmic Sobolev inequalities and contractivity properties of semigroups. *Lecture Notes in Mathematics* vol. 1563. Springer, Berlin, 1993.

A.R. Its, C.A. Tracy, and H. Widom, Random words, Toeplitz determinants and integrable systems I. in *Random Matrix Models and Their Applications*, pp. 245–258. Math. Sci. Res. Inst. Publ. 40, eds. P. Bleher and A. R. Its, Cambridge Univ. Press, Cambridge, 2001.

S. Janson. The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Structures Algorithms* vol. 7, pp. 337–355, 1995.

K. Johansson. Discrete orthogonal polynomial ensembles and the Plancherel measure. *Annals of Mathematics* vol. 153, pp. 259–296, 2001.

M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, vol. 1, pp. 63–87, 1997.

M. Ledoux. *The Concentration of Measure Phenomenon.* American Mathematical Society, Providence, RI, 2001.

S. Linusson and J. Wästlund. A proof of Parisi's conjecture on the random assignment problem, *Probability Theory and Related Fields* vol. 128, pp. 419–440, 2004.

A. Maurer. Concentration inequalities for functions of independent variables. *Random Structures and Algorithms*, 2006.

C. McDiarmid. Concentration. In M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–248. Springer, New York, 1998.

C. McDiarmid and B. Reed. Concentration of self-bounding functions and an inequality of Talagrand, 2006.

C. Nair, B. Prabhakar, and M. Sharma. Proofs of the Parisi and Coppersmith-Sorkin random assignment conjectures. *Random Structures Algorithms*, vol. 27, pp. 413–444, 2005.

E.M. Palmer. *Graphical Evolution*. John Wiley & Sons, New York, 1985.

W.-S. Rhee and M. Talagrand. Martingale inequalities and the jackknife estimate of variance. *Statistics and Probability Letters*, vol. 4, pp. 5–6. 1986.

J.M. Steele. An Efron-Stein inequality for nonsymmetric statistics. *Annals of Statistics*, vol. 14, pp. 753–758, 1986.

M. Talagrand. On Russo's approximate zero-one law. *Annals of Probability*, vol. 22, pp. 1576–1587, 1994.

M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S.*, vol. 81, pp. 73–205, 1995.

M. Talagrand. New concentration inequalities in product spaces. *Inventiones Mathematicae*, vol. 126, pp. 505–563, 1996a.

M. Talagrand. A new look at independence. *Annals of Probability*, vol. 24, pp. 1–34, 1996b.

C.A. Tracy and H. Widom. On the distributions of the lengths of the longest monotone subsequences in random words. *Probability Theory and Related Fields* vol. 119, pp. 350–380, 2001.

J. Wästlund. A simple proof of the Parisi and Coppersmith-Sorkin formulas for the random assignment problem. *Linköping Studies in Mathematics*, No. 6, 2005a.

J. Wästlund. The variance and higher moments in the random assignment problem. *Linköping Studies in Mathematics*, No. 8, 2005b.

J. Wästlund. Evaluation of Janson's constant for the variance in the random minimum spanning tree problem. *Linköping Studies in Mathematics*, No. 7, 2005c.