

# ON THE METHOD OF BOUNDED DIFFERENCES

Colin McDiarmid

## §1 Introduction

In the beginning Maurey (1979) used an inequality for bounded martingale difference sequences to prove an isoperimetric inequality for the symmetric group  $S_n$ ; this isoperimetric inequality was then used in investigating normed spaces, for related work see Milman & Schechtman (1986). Then Shamir & Spencer (1987) and Rhee & Talagrand (1987) introduced this 'bounded differences' method to a wide public of researchers in combinatorics and the mathematics of operational research and computer science, with dramatic impact. The purpose of this paper is to discuss the method and some of these applications. The underlying martingale result is due to Hoeffding (1963), Azuma (1967) and has often been referred to as 'Azuma's inequality'.

In this introductory section we first present a simple version of a bounded differences inequality which does not require us to introduce any concepts like martingales. We then sketch a plan of the rest of the paper. First recall the Chernoff bound on the tails of the binomial distribution.

(1.1) Lemma (Chernoff (1952)): Let  $X_1, \dots, X_n$  be independent random variables, with  $P[X_k=1] = p$  and  $P[X_k=0] = 1-p$  for each  $k$ . Then for any  $t > 0$

$$P[|\sum X_k - np| \geq t] \leq 2\exp[-2t^2/n].$$

The sum here is of course over  $k$  from 1 to  $n$ . This will always be the

case when we write an unadorned sum  $\Sigma$  or product  $\Pi$ .

The above inequality has proved its usefulness many times in combinatorics and elsewhere (see for example Erdős & Spencer (1974)). There are other times when one would like to use it but the corresponding object of interest will not quite separate into independent parts. What can we do? Perhaps we can use the following 'independent bounded differences inequality'.

(1.2) Lemma: Let  $X_1, \dots, X_n$  be independent random variables, with  $X_k$  taking values in a set  $A_k$  for each  $k$ . Suppose that the (measurable) function  $f: \Pi A_k \rightarrow \mathbb{R}$  satisfies

$$(1.3) \quad |f(\underline{x}) - f(\underline{x}')| \leq c_k$$

whenever the vectors  $\underline{x}$  and  $\underline{x}'$  differ only in the  $k$ th co-ordinate. Let  $Y$  be the random variable  $f[X_1, \dots, X_n]$ . Then for any  $t > 0$ ,

$$P(|Y - E(Y)| \geq t) \leq 2 \exp \left[ -2t^2 / \Sigma c_k^2 \right].$$

If we take each  $A_k = \{0,1\}$  and  $f(\underline{x}) = \Sigma x_k$  we may obtain lemma (1.1). Often we shall take  $A_k$  as a set of edges in a graph, as for example in lemmas (3.1) to (3.3) below.

The plan of this paper is as follows. In the next section we show part of the recent impact of these inequalities by sketching 'before' and 'after' pictures of our knowledge about colouring random graphs. Then we sketch proofs of the 'after' theorems using lemmas derived from lemma (1.2) above.

In the next section, section 4, we introduce martingales and present the basic 'Azuma's inequality' of Hoeffding (1963), Azuma (1967). (This does not quite yield lemma (1.2) above.)

In section 5 we present the 'Hoeffding' family of inequalities for sums of independent bounded random variables, which extend the Chernoff bound. These results generalise to inequalities involving martingales, and in section 6 we give these generalisations (and prove lemma (1.2)). These results in section 6 extend

Azuma's inequality.

In sections 7 and 8 we briefly sketch several further applications of these inequalities concerning isoperimetric inequalities in graphs and certain problems in operational research and computer science, and we conclude in section 9.

## §2 Colouring random graphs – before and after

In this section we show the dramatic effect of the bounded differences method on our knowledge about the stability number and the chromatic number of a random graph. (We cannot talk about colourings without talking about stable sets.)

For a splendid introduction to the theory of random graphs see Bollobás (1985). Recall that the random graph  $G_{n,p}$  has vertex set  $\{1, \dots, n\}$  and the  $\binom{n}{2}$  possible edges occur independently with probability  $p = p(n)$ . For simplicity we shall focus mainly on the case  $p$  constant, with  $0 < p < 1$ . Let  $q = 1-p$  and  $b = 1/q$ . Recall also that a set of vertices in a graph  $G$  is stable if no two are adjacent; and the stability number  $\alpha(G)$  is the largest size of such a set. A colouring of  $G$  is a partition of the vertices into stable sets; and the chromatic number  $\chi(G)$  is the least number of blocks in such a partition.

The stability number  $\alpha[G_{n,p}]$  is concentrated on only a few values (see Matula (1970), (1972), (1976) Grimmett & McDiarmid (1975), Bollobás & Erdős (1976)). Indeed (when  $p$  is constant) there is a function  $r = r(n,p)$  (see section 3(c) below) such that

$$(2.1) \quad P[r-1 \leq \alpha[G_{n,p}] \leq r] \rightarrow 1 \text{ as } n \rightarrow \infty.$$

This result was proved by 'classical' first and second moment methods (see Bollobás (1985)). Our knowledge about the stability number of sparse random graphs was patchy (again see Bollobás (1985)).

Now for colouring. Grimmett and McDiarmid (1975) (see also Bollobás & Erdős (1976) showed that (when  $p$  is constant) the chromatic number  $\chi[G_{n,p}]$

satisfies

$$(2.2) \quad \frac{n}{2 \log_b n} (1+o(1)) \leq \chi[G_{n,p}] \leq \frac{n}{\log_b n} (1+o(1))$$

for almost all graph  $G_{n,p}$  (that is, with probability  $\rightarrow 1$  as  $n \rightarrow \infty$ ). Also they conjectured that the lower bound was the 'correct' value. For results about sparse random graphs and related results see McDiarmid (1984). There was no significant improvement on the 1975 result (2.2) until Matula (1987) introduced a factor  $2/3$  into the upper bound, by using an elegant 'expose-and-merge' method for generating a random graph.

That was the 'before' picture, now for the 'after'.

The first breakthrough was when Shamir and Spencer in 1987 showed that the chromatic number  $\chi[G_{n,p}]$  is highly concentrated. Let us say that the graph function  $X$  is concentrated in width  $s = s(n,p)$  if there exists a function  $u = u(n,p)$  so that

$$P[u \leq X[G_{n,p}] \leq u+s] \rightarrow 1 \text{ as } n \rightarrow \infty.$$

Also let us use  $\omega(n)$  to denote a function which tends to  $\infty$  arbitrarily slowly as  $n \rightarrow \infty$ .

(2.3) Theorem (Shamir & Spencer (1987)): For any probability function  $p = p(n)$ ,  $\chi[G_{n,p}]$  is concentrated in width  $n^{1/2} \omega(n)$ .

(2.4) Theorem (Shamir & Spencer (1987)): Let  $p = n^{-\alpha}$ , where  $0 < \alpha < 1$ .

- i) For  $0 < \alpha < 1/2$ ,  $\chi[G_{n,p}]$  is concentrated in width  $s = n^{(1/2)-\alpha} \omega(n)$ .
- ii) For  $1/2 < \alpha < 1$ ,  $\chi[G_{n,p}]$  is concentrated in width  $s = \lceil (2\alpha+1)/(2\alpha-1) \rceil$ .

Shamir and Spencer comment that it is both a strength and a weakness of the method that  $u$  is not given explicitly. Luczak (1988a) reports considerable strengthenings of (2.3), (2.4) again based on the bounded differences method.

A year later Bollobás (1988a) gave the outstanding application of the bounded differences method. He showed roughly that for the function  $r$  in (2.1)

above,

$$(2.5) \quad P[\alpha[G_{n,p}] < r-2] \text{ is tiny compared to } 2^{-n}.$$

Given (2.1) and a suitable form of (2.5) it will be easy to handle colourings, and obtain

$$(2.6) \quad \text{\underline{Theorem}: For almost all graphs } G_{n,p} \\ \chi[G_{n,p}] = (1+o(1))n/2\log_b n.$$

This of course establishes the conjecture in Grimmett & McDiarmid (1975). (Korsunov (1980) put forward a proof for  $p = 1/2$  but this was incomplete and hard to follow.) Bollobás actually proved a more precise theorem; and reworking his arguments carefully (for this paper!) shows the following even more precise result.

$$(2.7) \quad \text{\underline{Theorem}: For almost all graphs } G_{n,p} \\ \chi[G_{n,p}] = n/[2\log_b n - 2\log_b \log_b n + o(1)].$$

Similar methods may be used to handle the chromatic number of (i) moderately sparse random graphs  $G_{n,p}$  (Bollobás (1988a), Luczak (1988b)); (ii)  $k$ -out graphs  $G_{n,k\text{-out}}$  where  $k$  is about  $cn$  (Bollobás (1988c)); and (iii) random  $m$  hypergraphs (Bollobás (1988c), Shamir (1988b), (1988c)). (To form the  $m$  graph  $G_{n,k\text{-out}}$  each of the vertices  $1, \dots, n$  independently directs edges to  $k$  vertices at random, then directions are ignored and any multiple edges combined.)

Now let us turn back to stable sets. Frieze (1989) uses the bounded differences method to pin down the stability number of sparse random graphs. A corollary of his theorem is the following.

$$(2.8) \quad \text{\underline{Theorem}: Let } p = p(n) \text{ be given, write } d = d(n) = np, \text{ and suppose that}$$

$d(n) \rightarrow \infty$  but  $d(n) = o(n)$  as  $n \rightarrow \infty$ . Then for almost all graphs  $G_{n,p}$

$$\alpha[G_{n,p}] = \frac{2n}{d} (\log d - \log \log d - \log 2 + 1 + o(1)).$$

### §3 Colouring random graphs – proofs

#### (a) General lemmas

The following useful results are all special cases of lemma (1.2). They are implicit in Shamir & Spencer (1987) (see also Bollobás (1988c)) though we have improved the bounds here.

(3.1) Lemma: Let  $[A_1, \dots, A_m]$  be a partition of the edge-set of the complete graph  $K_n$  into  $m$  blocks; and suppose that the graph function  $f$  satisfies  $|f(G) - f(G')| \leq 1$  whenever the symmetric difference  $E(G) \Delta E(G')$  is contained in a single block  $A_k$ . Then the random variable  $Y = f[G_{n,p}]$  satisfies

$$P(|Y - E(Y)| \geq t) \leq 2 \exp[-2t^2/m] \quad \text{for } t > 0.$$

The above result lemma (3.1) follows directly from lemma (1.2) with each  $c_k = 1$ . The next two results follow directly from lemma (3.1). For the former let  $A_k$  be the set of edges  $\{j, k\}$  where  $j < k$ ; and for the latter let the blocks  $A_k$  be singletons.

(3.2) Lemma: Suppose that the graph function  $f$  satisfies  $|f(G) - f(G')| \leq 1$  whenever  $G'$  can be obtained from  $G$  by changing edges incident with a single vertex. Then the corresponding random variable  $Y = f[G_{n,p}]$  satisfies

$$P(|Y - E(Y)| \geq t) \leq 2 \exp[-2t^2/n] \quad \text{for } t > 0.$$

(3.3) Lemma: Suppose that the graph function  $f$  satisfies  $|f(G) - f(G')| \leq 1$  whenever  $G$  and  $G'$  differ in only one edge. Then the corresponding random variable  $Y = f[G_{n,p}]$  satisfies

$$P(|Y - E(Y)| \geq t) \leq 2 \exp \left[ -2t^2 / \left\lceil \frac{n}{2} \right\rceil \right] \quad \text{for } t > 0.$$

**(b) Concentration of  $\chi[G_{n,p}]$**

By lemma (3.2) above we deduce immediately that if  $Y$  is the random variable  $\chi[G_{n,p}]$  then for any  $t > 0$

$$(3.4) \quad P(|Y - E(Y)| \geq t) \leq 2 \exp \left[ -2t^2/n \right].$$

This proves theorem (2.3) (and is a sharpening of (28) in [SS]).

Clearly the same inequality will hold if  $Y$  is  $\Psi[G_{n,p}]$  where  $\Psi(G)$  is the achromatic number (see McDiarmid (1982) for the asymptotic behaviour of  $\Psi[G_{n,p}]$ ); or  $Y$  is  $i[G_{n,p}]$  where  $i(G)$  is the interval number of  $G$  (Scheinerman (1989)); or in many other examples.

Now let us consider a second proof of theorem (2.3) which will also prove theorem (2.4). This proof method is due to Alan Frieze, see also Luczak (1988a). We thus avoid the complications of the proof in Shamir & Spencer (1987) of theorem (2.4) (though the inequality theorem 7 in that paper may be of some interest).

Let  $\omega(n) \rightarrow \infty$  arbitrarily slowly, and let  $u = u(n)$  be the least integer such that

$$(3.5) \quad P[\chi[G_{n,p}] \leq u] \geq 1/\omega(n).$$

For a graph with  $n$  vertices, let  $f(G)$  be the least size of a subset  $W$  of vertices of  $G$  such that  $\chi(G \setminus W) \leq u(n)$ . Let  $Y$  be the random variable  $f[G_{n,p}]$ . Then by lemma (3.2)

$$(3.6) \quad P(|Y - E(Y)| \geq t) \leq 2 \exp \left[ -2t^2/n \right] \quad \text{for } t > 0.$$

But by (3.5),  $P(Y = 0) \geq 1/\omega(n)$ , so we must have  $E(Y) \leq n^{1/2} \omega(n)$  for large  $n$ ; and now by (3.6) we have

$$(3.7) \quad P\left[ Y < 2n^{1/2} \omega(n) \right] \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Theorem (2.3) follows easily from (3.5) and (3.7). However, we can also now prove theorem (2.4) (which is a slight sharpening of the original result). We need only show the following lemma.

(3.8) Lemma: With probability tending to 1, every set of at most  $n^{1/2}\omega(n)$  vertices in  $G_{n,p}$  may be  $s$ -coloured, where  $s$  is given in theorem (2.4).

We may prove lemma (3.8) by showing that the expected number of sets as above with each degree at least  $s-1$  tends to 0 as  $n \rightarrow \infty$ , as in Shamir & Spencer (1987).

### (c) Stable sets in $G_{n,p}$

Here we present a suitable explicit form of (2.5), namely lemma (3.11) below, following the ideas in Bollobás (1988a).

Given an integer  $r$ ,  $1 \leq r \leq n$ , let  $E(n,r)$  be the expected number of stable sets of size  $r$  in  $G_{n,p}$ . Thus  $E(n,r) = \binom{n}{r} q^{\binom{r}{2}}$ . Next for real  $r$ ,  $1 \leq r \leq n$ , let

$$\hat{E}(n,r) = (2\pi)^{-1/2} n^{n+(1/2)} (n-r)^{-(n-r+(1/2))} r^{-(r+(1/2))} q^{r(r-1)/2}.$$

By Stirling's formula,  $E(n,r) = (1+o(1))\hat{E}(n,r)$  if  $r=r(n)$  is an integer,  $r(n) \rightarrow \infty$  but  $r(n)=o(\log n)$ . Now define

$$(3.9) \quad r = r(n) = 2\log_b n - 2\log_b \log_b n + 2\log_b \left[ \frac{e}{2} \right].$$

Then straightforward calculations show that for any  $t = t(n) = o(1)$ ,

$$(3.10) \quad \hat{E}(n, r+t) = n^{1-t+o(\log \log n / \log n)}.$$

It follows that

$$P[\alpha[G_{n,p}] < r + 1 + 1/\log \log n] \rightarrow 1 \text{ as } n \rightarrow \infty.$$

We shall show that

$$(3.11) \text{ Lemma: } P[\alpha[G_{n,p}] < \lfloor r - 1/2 - 1/\log \log n \rfloor] = o[\exp[-n^{1+1/\log \log n}]].$$

Proof: Set  $t(n) = -(1/2) - 1/\log \log n$ . Then by (3.10)

$$\hat{E}(n, r+t) = n^{3/2+(1+o(1))/\log \log n}.$$

Let  $s = s(n) = \lfloor r(n) + t(n) \rfloor$ . Further calculations now show that

$$(3.12) \quad \hat{E}(n', s(n)) = n^{3/2 + (1+o(1))/\log \log n}$$

for some integer  $n' = n'(n)$  with  $n[q^{1/2} + o(1)] \leq n' \leq n$ .

We are going to use lemma (3.3), so we shall need an appropriate function  $f$ . This is the clever trick. For a graph  $G$  of order  $n$  we define  $f(G)$  to be the maximum number of sets in a collection  $S_1, S_2, \dots$  of stable sets of size  $s(n)$  with  $|S_i \cap S_j| \leq 1$  for  $i \neq j$ . Of course  $|f(G) - f(G')| \leq 1$  if  $G$  and  $G'$  differ only in one edge, so we can indeed apply lemma (3.3). Let  $Y$  be the random variable  $f[G_{n,p}]$ . Then

$$(3.13) \quad P(|Y - E(Y)| \geq t) \leq 2 \exp[-4t^2/n^2] \quad \text{for } t > 0.$$

But  $P[\alpha[G_{n,p}] < s] = P(Y = 0)$ , so

$$(3.14) \quad P[\alpha[G_{n,p}] < s] \leq 2 \exp[-4E(Y)^2/n^2].$$

To use this inequality we want a good lower bound on  $E(Y)$ . For graphs  $G$  on  $\{1, \dots, n\}$  define  $f'(G)$  to be the number of stable sets  $S \subseteq \{1, \dots, n'\}$  of size  $s(n)$  which are such that for any other such set  $S'$  we have  $|S \cap S'| \leq 1$ . Of course  $E(Y) \geq E[f'(G_{n,p})]$ . Further calculations show that  $E[f'(G_{n,p})] = (1+o(1))E(n', s(n))$ , and hence lemma (3.11) follows from (3.12) and (3.14). □

#### (d) Colouring $G_{n,p}$

Let us see here how lemma (3.11) on large stable sets allows us to pin down  $\chi[G_{n,p}]$ . The approach was sketched out in Bollobás and Erdős (1976) (section 5(i)), but it was not until Bollobás used the bounded differences method to prove a result like lemma (3.11) that it could be made to work. We shall sketch a proof of the following result, which of course yields theorem (2.7). A detailed proof is given in McDiarmid (1989).

(3.15) Theorem: Let  $0 < p = 1 - q < 1$  be fixed, let  $b = 1/q$ , and let

$$r = r(n) = 2 \log_b n - 2 \log_b \log_b n + 2 \log_b \left[ \frac{e}{2} \right],$$

as before. Then for almost all graphs  $G_{n,p}$

$$n/\{r+o(1)\} \leq \chi[G_{n,p}] \leq n/\left\{r - \frac{1}{2} - \frac{1}{1-q} \frac{1}{2} + o(1)\right\}.$$

The lower bound follows from an easy first moment calculation which need not concern us here. The upper bound follows from the two lemmas below, one concerning random graphs and one deterministic.

Let us say that a graph  $G$  has property  $Q_n$  if it has  $n$  nodes and for all subsets  $W$  of at least  $n/\log^3 n$  nodes we have  $\alpha(G[W]) \geq s(|W|)$ . Here  $G[W]$  denotes the subgraph of  $G$  induced by  $W$ , and the function  $s$  is defined in the proof of lemma (3.11).

(3.16) Lemma: Almost all graphs  $G_{n,p}$  have property  $Q_n$ .

(3.17) Lemma: Consider (deterministic) graphs  $G_n$  with property  $Q_n$  for  $n = 1, 2, \dots$

Then

$$\chi[G_n] \leq n/\left\{r(n) - \frac{1}{2} - \frac{1}{1-q} \frac{1}{2} + o(1)\right\}.$$

The deterministic lemma (3.17) follows by considering repeatedly picking out large stable sets until less than  $n/\log^3 n$  nodes remain, and then colouring the remaining nodes all with different colours.

Proof of lemma (3.16):

By lemma (3.11), for  $n$  sufficiently large

$$\begin{aligned} & P\{G_{n,p} \text{ does not have } Q_n\} \\ & \leq 2^n \exp\left[-\left[n/\log^3 n\right]^{1+(1/2\log\log n)}\right] \\ & \leq 2^n e^{-n} \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

□

### (e) Stability number of sparse random graphs

The upper bound in theorem (2.8) from Frieze (1989) is straightforward (see Bollobás (1985) lemma XI.21). Let us sketch the lines of the lower bound

proof.

For a suitable function  $n' = n'(n)$  let  $\mathcal{P}_n$  be a partition of  $\{1, \dots, n\}$  into  $n'$  almost equal-sized sets. (In Frieze (1989)  $n'$  is about  $n(\log d)^2/d$ .) In a graph  $G$  with  $n$  vertices, call a set  $S$  of vertices  $\mathcal{P}$ -stable if it is stable and  $|S \cap B| \leq 1$  for each block  $B$  of  $\mathcal{P}_n$ . Let  $\beta(G)$  be the largest size of a  $\mathcal{P}$ -stable set. Of course  $\beta(G) \leq \alpha(G)$ . (Frieze attributes the idea of considering  $\mathcal{P}$ -stable sets rather than just stable sets to Luczak.)

Let  $Y_n$  be the random variable  $\beta[G_{n,p}]$ . Lemma (3.1) shows that for any  $t > 0$

$$(3.18) \quad P[|Y_n - E[Y_n]| \geq t] \leq 2 \exp[-2t^2/n'].$$

But now let

$$k = k(n) = \frac{2n}{d}(\log d - \log \log d - \log 2 + 1 - \epsilon).$$

Second moment calculations on the number of  $\mathcal{P}$ -stable sets of size  $k$  show that  $P[Y_n \geq k]$  does not tend to zero very quickly. It then follows from (3.18) that  $E[Y_n] \geq k - (\epsilon n/d)$ , and the lower bound in theorem (2.8) follows by using (3.18) again.

#### §4 Martingales

Let  $(\Omega, \mathcal{F}, P)$  be a probability triple. Given a random variable  $X$  and a sub- $\sigma$ -field  $\mathcal{F}$  of  $\mathcal{F}$ , we shall use the notation  $E(X | \mathcal{F})$  to denote the expectation of  $X$  conditional on  $\mathcal{F}$ . In most of our applications  $\Omega$  will be a finite set. There is then a partition of  $\Omega$  such that the  $\sigma$ -field  $\mathcal{F}$  is the collection of sets which are unions of blocks of the partition. A random variable is a real-valued function  $X$  defined on  $\Omega$  such that  $X$  is constant on the blocks. Also  $E(X | \mathcal{F})$  is simply the function  $f$  defined on  $\Omega$  which is constant on the blocks, the constant on each block being the average value of  $X$  on the block.

A nested sequence  $\mathcal{F}_0 = \{\phi, \Omega\} \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}$  of sub- $\sigma$ -fields of  $\mathcal{F}$  is called a filter. In the finite case this corresponds to a sequence of increasingly refined partitions of  $\Omega$ , starting with the trivial partition with one block  $\Omega$ . Given

a filter, a sequence  $X_0, X_1, X_2, \dots$  of integrable random variables is called a martingale if  $E[X_{n+1} | \mathcal{F}_n] = X_n$  for each  $n \geq 0$ . Also, given a filter, a sequence  $Y_1, Y_2, \dots$  of integrable random variables is called a martingale difference sequence (mds) if for each  $n \geq 1$ ,  $Y_n$  is  $\mathcal{F}_n$ -measurable and  $E[Y_n | \mathcal{F}_{n-1}] = 0$ .

From a martingale  $X_0, X_1, X_2, \dots$  we obtain a martingale difference sequence by setting  $Y_k = X_k - X_{k-1}$ ; and conversely from  $X_0$  and a martingale difference sequence  $Y_1, Y_2, \dots$  we obtain a martingale  $X_0, X_1, X_2, \dots$  by setting  $X_k = X_0 + \sum_{i=1}^k Y_i$ . Thus we may focus on either sequence.

We shall be interested here only in finite filters  $\mathcal{F} = \{\phi, \Omega\} \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ . All corresponding martingales  $X_0, X_1, \dots, X_n$  are then obtained by Doob's martingale process: let  $X$  be an integrable random variable and define  $X_k = E[X | \mathcal{F}_k]$  for  $k = 0, \dots, n$ . Thus  $X_0 = E(X)$  and  $X_n = X$  if  $X$  is  $\mathcal{F}_n$ -measurable. Then if  $Y_1, \dots, Y_n$  is the corresponding martingale difference sequence we of course have  $X - EX = \sum Y_k$ .

The basic inequality for a bounded martingale difference sequence is the following lemma of Hoeffding (1963), Azuma (1967) (see also Freedman (1975)), which has often been referred to as 'Azuma's inequality'.

(4.1) Lemma: Let  $Y_1, \dots, Y_n$  be a martingale difference sequence with  $|Y_k| \leq c_k$  for each  $k$ , for suitable constants  $c_k$ . Then for any  $t > 0$ ,

$$P[|\sum Y_k| \geq t] \leq 2 \exp\left[-t^2/2\sum c_k^2\right].$$

Suppose as in (1.1) that  $X_1, \dots, X_n$  are independent, with  $P[X_k=1] = p$  and  $P[X_k=0] = 1-p$ . Set  $Y_k = X_k - p$  and  $c_k = \max(p, 1-p)$ , and apply lemma (4.1) to obtain the Chernoff bound (1.1), except that the bound is weakened slightly (if  $p \neq 1/2$ ). We may also deduce a similar weakened version of lemma (1.2). All our applications here will in fact be based on less symmetrical forms of lemma (4.1) and will thus avoid a gratuitous factor  $1/4$  in the exponent in

previously presented bounds.

We shall prove stronger results in section 6 below (see corollary (6.9)) but it is convenient to prove lemma (4.1) here. First we prove one preliminary lemma.

(4.2) Lemma: Let the random variable  $Y$  satisfy  $E(Y) = 0$  and  $|Y| \leq 1$ . Then for any  $h > 0$

$$E[e^{hY}] \leq e^{h^2/2}.$$

Proof: As  $e^{hx}$  is a convex function of  $x$ ,

$$e^{hx} \leq \frac{1-x}{2} e^{-h} + \frac{1+x}{2} e^h \quad \text{for } -1 \leq x \leq 1,$$

and so

$$E[e^{hY}] \leq \frac{1}{2}[e^{-h} + e^h] \leq e^{h^2/2},$$

by considering series expansions. □

Proof of lemma (4.1):

Let  $S_k = \sum_{i=1}^k Y_i$ . By the Markov or Bernstein inequality, for any  $h > 0$

$$\text{Prob}[S_n \geq t] \leq e^{-ht} E[e^{hS_n}].$$

But

$$\begin{aligned} E[e^{hS_n}] &= E[e^{hS_{n-1}} e^{hY_n}] \\ &= E\left[e^{hS_{n-1}} E[e^{hY_n} | \mathcal{F}_{n-1}]\right] \\ &\leq E\left[e^{hS_{n-1}} \exp\left[\frac{1}{2}(hc_n)^2\right]\right] \quad \text{by lemma (4.2)} \\ &\leq \exp\left[\frac{1}{2}h^2 \sum_{k=1}^n c_k^2\right] \quad \text{on iterating.} \end{aligned}$$

Now set  $h = t/\sum_{k=1}^n c_k^2$  to obtain

$$\text{Prob}[S_n \geq t] \leq \exp\left[-t^2/2\sum_{k=1}^n c_k^2\right].$$

By symmetry the absolute value gives at most a factor two. □

For a similar simple proof of a weaker version of the inequality see Milman & Schechtman (1986). Above we used our bounds on  $|Y_k|$  to obtain good bounds on  $E\left[e^{hY_k} \mid \mathcal{F}_{k-1}\right]$ . Other assumptions which yield bounds on this latter quantity will yield related inequalities — see for example Johnson et al (1985), Bollobás (1988c) theorem 7.

## §5 Inequalities for bounded independent summands

In this section we present the 'Hoeffding' family of inequalities for sums of independent bounded random variables. In the next section we extend these to inequalities involving martingales.

### (a) Results

We shall take the following inequality as our starting point.

(5.1) Theorem (Hoeffding (1963)): Let the random variables  $X_1, \dots, X_n$  be independent, with  $0 \leq X_k \leq 1$  for each  $k$ . Let  $X = \frac{1}{n}\sum X_k$ ,  $p = E[X]$ , and  $q = 1-p$ . Then for  $0 \leq t < q$ ,

$$P(X - p \geq t) \leq \left[ \left[ \frac{p}{p+t} \right]^{p+t} \left[ \frac{q}{q-t} \right]^{q-t} \right]^n.$$

A special case of interest is when each random variable  $X_k$  is 1 with probability  $p$  and 0 with probability  $q$ . The theorem then reduces to a bound on a tail of the binomial distribution due to Chernoff (1952) though the methods involved date back to Bernstein (see Chvátal (1979)). This bound is good for large deviations (see Chernoff (1952), Bahadur & Ranga Rao (1960), Bahadur (1971)) though inequalities closer to the normal approximation of DeMoivre – Laplace are naturally better for small deviations (see Feller (1968), Bollobás (1985)). (For related inequalities concerned with variations of the  $p_i$  subject to  $\sum p_i = p$  see Hoeffding (1956), Gleser (1975). The bound applies also to the hypergeometric

distribution — see Hoeffding (1963), Chvátal (1979).) It is straightforward to obtain from theorem (5.1) weaker but more useful bounds.

(5.2) Corollary: As in theorem (5.1), let the random variables  $X_1, \dots, X_n$  be independent, with  $0 \leq X_k \leq 1$  for each  $k$ , let  $\bar{X} = \frac{1}{n} \sum X_k$  and let  $p = E[\bar{X}]$ .

(a) For  $t > 0$ ,

$$(5.3) \quad P(\bar{X} - p \geq t) \leq \exp[-2nt^2],$$

$$(5.4) \quad P(\bar{X} - p \leq -t) \leq \exp[-2nt^2].$$

(b) For  $0 < \epsilon < 1$

$$(5.5) \quad P(\bar{X} - p \geq \epsilon p) \leq \exp\left[-\frac{1}{3}\epsilon^2 np\right],$$

$$(5.6) \quad P(\bar{X} - p \leq -\epsilon p) \leq \exp\left[-\frac{1}{2}\epsilon^2 np\right].$$

The inequalities in (a) are perhaps the basic workhorses, but often  $p$  is small in applications and then the inequalities in (b) may be better. Part (a) is due to Hoeffding (1963) who also discusses relationships between theorem (5.1), the inequalities in (a) and other similar inequalities. Part (b) appears in Angluin & Valiant (1979) (in the binomial case). In Karp (1979) the inequality (5.3) is attributed to Angluin. For similar results (in the binomial case) based on Stirling's approximation see Bollobás (1985) Chapter 1, corollary 4 and theorem 7(i).

Hoeffding also gives a powerful extension of corollary (5.2) (a) to the case when the ranges of the summands need not be the same.

(5.7) Theorem: Let the random variables  $X_1, \dots, X_n$  be independent, with  $a_k \leq X_k \leq b_k$  for each  $k$ , for suitable constants  $a_k, b_k$ . As before let  $\bar{X} = \frac{1}{n} \sum X_k$  and  $p = E[\bar{X}]$ . Then for  $t > 0$

$$P(\bar{X} - p \geq t) \leq \exp\left[-2n^2 t^2 / \sum [b_k - a_k]^2\right],$$

$$P(\bar{X} - p \leq -t) \leq \exp\left[-2n^2 t^2 / \sum [b_k - a_k]^2\right].$$

## (b) Proofs

Let us prove the above results here, as the proofs will be useful also for the extensions of these results in the next section (and proofs of the much used inequalities (5.5) and (5.6) do not seem to be easily available in the literature).

Proof of theorem (5.1):

Let  $m = (p + t)n$ . For  $s \geq 1$ ,

$$\begin{aligned} P\left[\sum X_k \geq m\right] &\leq E\left[s^{\sum X_k - m}\right] \\ &= s^{-m} \prod E\left[s^{X_k}\right] \\ &= s^{-m} \prod [q_k + p_k s] \\ &\leq s^{-m} (q + ps)^n, \end{aligned}$$

since geometric means are at most arithmetic means. Now set  $s = \frac{(p+t)q}{p(q-t)}$  to obtain the desired inequality.  $\square$

Proof of corollary (5.2):

(a) Let  $f(t) = (p+t)\ln \frac{p}{p+t} + (q-t)\ln \frac{q}{q-t}$  for  $-p < t < q$ . Then

$$f'(t) = \ln \frac{p(q-t)}{(p+t)q},$$

and

$$f''(t) = -\frac{1}{(p+t)(q-t)} \leq -4$$

since

$$(p+t)(1-(p+t)) \leq \frac{1}{4}.$$

But  $f(0) = f'(0) = 0$ , and so it follows from Taylor's theorem that for  $0 \leq t < q$

$$\begin{aligned} f(t) &= \left[t^2/2\right]f''(s) \quad \text{for some } s, 0 \leq s \leq t \\ &\leq -2t^2. \end{aligned}$$

The inequality (5.3) now follows from theorem (5.1), and by considering  $1-X_k$  we obtain (5.4).

(b) Now let  $g(x) = f(xp)$ , for  $0 \leq x \leq 1$ ,  $xp < q$ . Then  $g'(x) = pf'(xp)$  and

$$g''(x) = p^2 f''(xp) = -\frac{p}{(1+x)(q-xp)} \leq -\frac{p}{1+x}.$$

Thus

$$\begin{aligned} g'(x) &\leq -p \ln(1+x) \\ &\leq -(2p/3)x \quad \text{since } 0 \leq x \leq 1, \end{aligned}$$

and so

$$g(x) \leq -(p/3)x^2.$$

Together with theorem (5.1) this yields (5.5).

Finally let  $h(x) = g(-x)$  for  $0 \leq x < 1$ . Then  $h'(x) = -g'(-x)$  and  $h''(x) = g''(-x)$ . Thus  $h(0) = h'(0) = 0$  and

$$h''(x) = -\frac{p}{(1-x)(q+xp)} \leq -p,$$

so  $h(x) \leq -px^2/2$  and (5.6) follows from theorem (5.1).



It remains here only to prove theorem (5.7). But once we have proved lemma (5.8) below which extends lemma (4.2), then theorem (5.7) will follow by a proof like that of lemma (4.1).

(5.8) Lemma: Let  $X$  be a random variable with  $E(X) = 0$ ,  $a \leq X \leq b$ . Then for

$$E[e^{hX}] \leq \exp\left[\frac{1}{8}h^2(b-a)^2\right].$$

Arguing as for lemma (4.2)

$$e^{hx} \leq \frac{x-a}{b-a} e^{hb} + \frac{b-x}{b-a} e^{ha} \quad \text{for } a \leq x \leq b,$$

and so

$$(5.9) \quad E[e^{hX}] \leq \frac{b}{b-a} e^{ha} - \frac{a}{b-a} e^{hb} = e^{f(\hat{h})}$$

where  $\hat{h} = h(b-a)$  and  $f(x) = -px + \log[1-p+pe^x]$ .

Now we may argue as in the proof of corollary (5.2). We have

$$f'(x) = -p + \frac{p}{p+(1-p)e^{-x}}$$

and

$$f''(x) = \frac{p(1-p)e^{-x}}{[p+(1-p)e^{-x}]^2} \leq \frac{1}{4}.$$

Also  $f(0) = f'(0) = 0$ , so

$$f(\hat{h}) \leq \frac{1}{8} \hat{h}^2 = \frac{1}{8} h^2 (b-a)^2. \quad \square$$

## §6 Inequalities for bounded martingale difference sequences

### (a) Results

Here we present the 'Hoeffding' family of inequalities for bounded martingale difference sequences. These extend the basic inequality (4.1) ('Azuma's inequality') and the inequalities of section 5 for the independent case. The main thrust of these results was pointed out in Hoeffding (1963) and we largely follow his treatment. We shall also give an extension of lemma (1.2). The use of any of these inequalities in combinatorics is what we call the 'bounded differences' method.

(6.1) Theorem: Let  $Y_1, \dots, Y_n$  be a martingale difference sequence with  $-a_k \leq Y_k \leq 1-a_k$  for each  $k$ , for suitable constants  $a_k$ . Let  $a = \frac{1}{n} \sum a_k$  and  $\bar{a} = 1 - a$ . Then for any  $t > 0$

$$P[\sum Y_k \geq nt] \leq \left[ \left( \frac{a}{a+t} \right)^{a+t} \left( \frac{\bar{a}}{\bar{a}-t} \right)^{\bar{a}-t} \right]^n.$$

To obtain theorem (5.1) from theorem (6.1) set  $a_k = E[X_k]$  and  $Y_k = X_k - a_k$ . We shall prove theorem (6.1) shortly. As before we want weaker but more useful bounds. The same proof as for the corollary (5.2) to theorem (5.1) will yield

(6.2) Corollary: As above, let  $Y_1, \dots, Y_n$  be a martingale difference sequence with  $-a_k \leq Y_k \leq 1-a_k$  for each  $k$ , for suitable constants  $a_k$ ; and let  $a = \frac{1}{n} \sum a_k$ .

(a) For any  $t > 0$ ,

$$(6.3) \quad P[\sum Y_k \geq t] \leq \exp[-2t^2/n],$$

$$(6.4) \quad P\left[\sum Y_k \leq -t\right] \leq \exp\left[-2t^2/n\right].$$

(b) For any  $0 \leq \epsilon \leq 1$ ,

$$(6.5) \quad P\left[\sum Y_k \geq \epsilon a\right] \leq \exp\left[-\frac{1}{3}\epsilon^2 a/n\right],$$

$$(6.6) \quad P\left[\sum Y_k \leq -\epsilon a\right] \leq \exp\left[-\frac{1}{2}\epsilon^2 a/n\right].$$

To obtain corollary (5.2) from corollary (6.2) set  $a_k = E[X_k]$  and  $Y_k = X_k - a_k$  as before.

The next result extends inequalities (6.3) and (6.4) above and theorem (5.7) for the independent case, and Azuma's inequality lemma (4.1). It seems to be the most useful inequality for the bounded differences method. We shall state it in terms of martingales, with a corollary in terms of martingales difference sequences and a corollary extending lemma (1.2).

(6.7) Theorem: Let  $(\phi, \Omega) = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_n$  be a filter. Let the integrable random variable  $X$  be  $\mathcal{F}_n$ -measurable, and let  $X_0, X_1, \dots, X_n$  be the martingale obtained by setting  $X_k = E[X | \mathcal{F}_k]$ . Suppose that for each  $k=1, \dots, n$  there is a constant  $c_k$  and an  $\mathcal{F}_{k-1}$ -measurable function  $a_k$  such that

$$(6.8) \quad a_k \leq X_k \leq a_k + c_k.$$

Then for any  $t > 0$ ,

$$P(X - EX \geq t) \leq \exp\left[-2t^2/\sum c_k^2\right],$$

$$P(X - EX \leq -t) \leq \exp\left[-2t^2/\sum c_k^2\right].$$

The next result is essentially the special case with each function  $a_k$  constant. It extends the basic inequality (4.1) (Azuma's inequality).

(6.9) Corollary: Let  $Y_1, \dots, Y_n$  be a martingale difference sequence with  $a_k \leq Y_k \leq b_k$  for each  $k$ , for suitable constants  $a_k, b_k$ . Then for any  $t > 0$

$$P\left[\sum Y_k \geq t\right] \leq \exp\left[-2t^2/\sum (b_k - a_k)^2\right],$$

$$P\left[\sum Y_k \leq -t\right] \leq \exp\left[-2t^2/\sum (b_k - a_k)^2\right].$$

The last result here will also follow easily from theorem (6.7). It is an extension of lemma (1.2) in which the condition (1.3) is replaced by a weaker but less attractive condition.

(6.10) Corollary: Let  $Z_1, \dots, Z_n$  be random variables with  $Z_k$  taking values in a set  $A_k$ , and let  $\underline{Z}$  denote the random vector  $[Z_1, \dots, Z_n]$ . Let  $f: \prod A_k \rightarrow \mathbb{R}$  be an appropriately measurable function. Suppose that there are constants  $c_1, \dots, c_n$  so that

$$(6.11) \quad \begin{aligned} & |E\{f(\underline{Z}) \mid [Z_1, \dots, Z_{k-1}] = [z_1, \dots, z_{k-1}], Z_k = z_k\} \\ & - E\{f(\underline{Z}) \mid [Z_1, \dots, Z_{k-1}] = [z_1, \dots, z_{k-1}], Z_k = z'_k\}| \leq c_k \end{aligned}$$

for each  $k = 1, \dots, n$  and  $z_i \in A_i$  ( $i = 1, \dots, k-1$ ) and  $z_k, z'_k \in A_k$ . Then for any  $t > 0$ ,

$$P(|f(\underline{Z}) - Ef(\underline{Z})| \geq t) \leq 2 \exp\left[-2t^2 / \sum c_k^2\right].$$

## (b) Proofs

### Proof of theorem (6.1)

We combine the proofs of theorems (4.1) and (5.1). Let  $S_k = \sum_{i=1}^k Y_i$ .

For any  $h > 0$ ,

$$\begin{aligned} P\{S_n \geq nt\} &\leq e^{-hnt} E\{\exp hS_n\} \\ &= e^{-hnt} E\{\exp[hS_{n-1}] E\{\exp hY_n \mid \mathcal{F}_{n-1}\}\} \\ &\leq e^{-hnt} E\{\exp[hS_{n-1}]\} \left[ (1-a_n) e^{-ha_n + a_n e^{h[1-a_n]}} \right] \quad \text{by (6.9)} \\ &\leq e^{-hnt} \prod_{k=1}^n \left[ (1-a_k) e^{-ha_k + a_k e^{h[1-a_k]}} \right] \quad \text{on iterating} \\ &= e^{-hnt} e^{-h\sum a_k} \prod_{k=1}^n \left[ 1 - a_k + a_k e^h \right] \\ &\leq e^{-hnt} e^{-hna} \left[ 1 - a + a e^h \right]^n. \end{aligned}$$

Now set  $e^h = \frac{(a+t)\bar{a}}{a(\bar{a}-t)}$  to obtain the desired inequality.



Proof of theorem (6.7)

Let  $Y_k = X_k - X_{k-1}$  and  $S_k = \sum_{i=1}^k Y_i = X_k - X_0$ . We argue initially as in the proof of theorem (6.1). For any  $h > 0$ ,

$$\begin{aligned} P(X - EX \geq t) &= P[S_n \geq t] \\ &\leq e^{-ht} E[\exp h S_n] \\ &= e^{-ht} E[\exp h S_{n-1} E[\exp h Y_n | \mathcal{F}_{n-1}]] \\ &\leq e^{-ht} E[\exp h S_{n-1}] \exp\left[\frac{1}{8} h^2 c_n^2\right] \quad \text{by lemma (5.8)} \\ &\leq e^{-ht} \exp\left[\frac{1}{8} h^2 \Sigma c_k^2\right] \quad \text{on iterating.} \end{aligned}$$

Now set  $h = 4t/\Sigma c_k^2$  to obtain the former inequality in theorem (6.7). To deduce the latter replace  $X$  by  $-X$ . □

Proof of corollary (6.10):

Let  $\mathcal{F}_k$  be the  $\sigma$ -field generated by  $Z_1, \dots, Z_k$ . Let  $X_k = E[f(\underline{Z}) | \mathcal{F}_k]$  and let

$$a_k = \text{ess inf } [X_k | \mathcal{F}_{k-1}], \quad b_k = \text{ess sup } [X_k | \mathcal{F}_{k-1}].$$

Then the condition (6.11) says that  $b_k - a_k \leq c_k$ . So corollary (6.10) now follows from theorem (6.7). □

**(c) Inequalities for maxima**

All the inequalities we have met here are based on the Bernstein inequality

$$P(X \geq t) \leq e^{-ht} E[e^{hX}] \quad \text{for } h > 0.$$

Thus they can all be strengthened in the following way, as noted in Hoeffding (1963). See also Steiger (1967), (1969), (1970).

Let  $Y_1, \dots, Y_n$  be a martingale difference sequence, let  $h > 0$ , let  $S_k = Y_1 + \dots + Y_k$  and let  $T_k = \exp h S_k$ . As long as the  $T_k$  are integrable,  $T_1, \dots, T_n$  forms a submartingale and so we can use Doob's maximal inequality (see for example Chung (1974) pages 320 and 330) to deduce that for any  $t > 0$ ,

$$\begin{aligned} P\left[\max_k S_k \geq t\right] &= P\left[\max_k T_k \geq e^{ht}\right] \\ &\leq e^{-ht} E[T_n] = e^{-ht} E[e^{hS_n}]. \end{aligned}$$

It follows that in all our inequalities in this section we may replace  $\Sigma Y_k$  by  $\max S_k$ , and similarly for independent summands. We do not have applications here for these extensions.

We have already discussed some applications of the bounded differences method to random graphs. See Bollobás (1988b), (1988c) for further applications in this area in particular concerning the first cycle problem and the probability of containing a given small subgraph (and see section 8(d), (e) below). In the last two sections before our concluding remarks we shall discuss two further areas of applications.

### §7 Isoperimetric inequalities for graphs

Consider a finite metric space  $(V, d)$ . We shall be interested in particular in the case when  $V$  is the vertex set of a graph  $G$  and  $d$  measures distance in the graph.

For  $A \subseteq V$  and  $t > 0$  the  $t$ -neighbourhood  $A_t$  of  $A$  is the set  $\{v \in V: d(v, A) \leq t\}$ . Here  $d(v, A)$  is of course  $\min\{d(v, x): x \in A\}$ . An 'isoperimetric inequality' means a lower bound on  $|A_t|$  depending on  $|A|$  and  $t$ . For an introduction to discrete isoperimetric inequalities see Bollobás (1986).

#### (a) General results

Following Schechtman (1982), Milman & Schechtman (1986) we make a fussy but useful definition. In the finite metric space  $(V, d)$  a partition sequence  $[[\mathcal{P}_k, c_k]: k=0, \dots, n]$  consists of a sequence  $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_n$  of increasingly refined partitions of  $V$ , starting with the trivial partition  $\mathcal{P}_0$  (with a single block  $V$ ) and ending with the discrete partition  $\mathcal{P}_n$  (into singleton blocks), and a sequence

$c_0, c_1, \dots, c_n$  of numbers with the following property: for each  $k=1, \dots, n$ , whenever  $A, B \in \mathcal{P}_k$  and  $A, B \subseteq C \in \mathcal{P}_{k-1}$  for some  $C$  then there is a bijection  $\varphi: A \rightarrow B$  with  $d(x, \varphi(x)) \leq c_k$  for all  $x \in A$ .

If the space has such a partition sequence we shall say that it has partition size at most  $\sum c_k^2$ . Observe that here  $\sum c_k$  is always at least the diameter  $\max \{d(x, y) : x, y \in V\}$ . Usually we shall have each  $c_k = 1$ , and indeed we do so in each of the example below.

(7.1) Example: The  $n$ -cube  $Q_n$  has vertex set  $V = \{0, 1\}^n$  and two vectors are adjacent if they differ in just one coordinate. Then  $Q_n$  has diameter and partition size equal to  $n$ . Indeed we may take  $\mathcal{P}_k$  as the partition into equivalence classes where two vectors are equivalent if they agree in the first  $k$  digits. The appropriate function  $\varphi$  just switches the  $k$ th digit.

(7.2) Example: The 'permutation graph'  $S_n$  has vertices the permutations of  $\{1, \dots, n\}$  and two vertices  $g$  and  $h$  are adjacent if  $g^{-1}h$  is a transposition. Then  $S_n$  has diameter and partition size equal to  $n-1$ . Indeed we may think of a permutation  $g$  as a sequence  $(g(1), \dots, g(n))$ , and take  $\mathcal{P}_k$  as above for  $k=0, \dots, n-1$ . If  $g \in A$ ,  $h \in B$  then the appropriate function  $\varphi$  acts on  $x \in A$  by swapping the numbers  $x(k) = g(k)$  and  $h(k)$ .

(7.3) Example: We may generalise the last example as follows (see Schechtman (1982)). The  $k$ -tuple graph  $S_{n,k}$  has vertices the  $k$ -tuples of distinct members of  $\{1, \dots, n\}$ , and two vertices  $g$  and  $h$  are adjacent if and only if either they differ in exactly one co-ordinate or they differ in exactly two co-ordinates, say  $j$  and  $k$ , and  $g(j) = h(k)$ ,  $g(k) = h(j)$ . Thus distinct vertices  $g$  and  $h$  are adjacent if and only if they may be extended to permutations which are adjacent in the permutation graph  $S_n$ . The graph  $S_{n,k}$  has diameter and partition size equal to  $k$  if  $k \leq n-1$ .

The basic result here follows easily from theorem (6.7) or corollary (6.10). It is similar to lemma (1.2), which we took as our workhorse earlier.

(7.4) Theorem: Suppose that the finite metric space  $(V, d)$  has a partition sequence  $\left[ \left[ \mathcal{A}_k, c_k \right] : k=1, \dots, n \right]$ . Let the function  $f$  on  $V$  satisfy  $|f(x) - f(y)| \leq d(x, y)$  for all  $x, y \in V$ . Let  $X$  be uniformly distributed over  $V$ . Then for any  $t > 0$ ,

$$\begin{aligned} P(f(X) - Ef(X) \geq t) &\leq \exp \left[ -2t^2 / \sum c_k^2 \right], \\ P(f(X) - Ef(X) \leq -t) &\leq \exp \left[ -2t^2 / \sum c_k^2 \right]. \end{aligned}$$

(7.5) Corollary: Suppose that the finite metric space  $(V, d)$  has a partition sequence  $\left[ \left[ \mathcal{A}_k, c_k \right] : k=1, \dots, n \right]$ . Let  $A \subseteq V$  with  $|A|/|V| = \alpha$ ,  $0 < \alpha < 1$ . Then for any  $t \geq t_0 = \left[ \frac{1}{2} \left[ \log \frac{1}{\alpha} \right] \sum c_k^2 \right]^{1/2}$ ,

$$|A_t|/|V| \geq 1 - \exp \left[ -2[t - t_0]^2 / \sum c_k^2 \right].$$

Thus for any  $t > 0$  and any  $\gamma > 0$ ,

$$(7.6) \quad |A_t|/|V| \geq 1 - \left[ \frac{1}{\alpha} \right]^{\gamma^2} \exp \left[ -2 \left[ \frac{\gamma}{1+\gamma} \right]^2 t^2 / \sum c_k^2 \right].$$

The above results follow the lines of the extension in Schechtman (1982) of the argument in Maurey (1979) for example (7.2) above — see also Milman & Schechtman (1986), Bollobás (1988c). For applications to the analysis of biased random sources see Shamir (1988a).

Proof of corollary (7.5):

Let  $t_1 = Ef(X)$ . Then by theorem (7.4)

$$\alpha = P(f(X) = 0) = P[f(X) \leq t_1 - t_1] \leq \exp \left[ -2t_1^2 / \sum c_k^2 \right],$$

and it follows that  $t_1 \leq t_0$ . Now by theorem (7.4) again, for  $t \geq t_0$ ,

$$\begin{aligned} 1 - |A_t|/|V| &= P(f(X) > t) \\ &\leq P[f(X) > t_1 + [t - t_0]] \\ &\leq \exp \left[ -2[t - t_0]^2 / \sum c_k^2 \right]. \end{aligned}$$

To prove the inequality (7.6) consider separately the cases  $t \leq (1+\gamma)t_0$  and  $t \geq (1+\gamma)t_0$ . □

Corollary (7.5) yields remarkable concentration results for the examples (7.1)–(7.3) above. Let us state these results for the  $n$ -cube  $Q_n$  and the permutation graph  $S_n$ .

(7.7) **Proposition:** Let  $A$  be a subset of the  $2^n$  vertices of the  $n$ -cube  $Q_n$ , with  $|A|/2^n = \alpha$ ,  $0 < \alpha < 1$ . Then for  $t \geq t_0 = \left[ \frac{1}{2} \left[ \log \frac{1}{\alpha} \right] n \right]^{1/2}$ ,

$$|A_t|/2^n \geq 1 - \exp \left[ -2 \left[ t - t_0 \right]^2 / n \right].$$

Thus for any  $t > 0$  and  $\gamma > 0$

$$|A_t|/2^n \geq 1 - \left[ \frac{1}{\alpha} \right]^{\gamma^2} \exp \left[ -2 \left[ \frac{\gamma}{1+\gamma} \right]^2 t^2 / n \right].$$

(7.8) **Proposition:** Let  $A$  be a subset of the  $n!$  vertices of the permutation graph  $S_n$ , with  $|A|/n! = \alpha$ ,  $0 < \alpha < 1$ . Then for any  $t \geq t_0 = \left[ \frac{1}{2} \left[ \log \frac{1}{\alpha} \right] (n-1) \right]^{1/2}$ ,

$$|A_t|/n! \geq 1 - \exp \left[ -2 \left[ t - t_0 \right]^2 / (n-1) \right].$$

Thus for any  $t > 0$  and  $\gamma > 0$

$$|A_t|/n! \geq 1 - \left[ \frac{1}{\alpha} \right]^{\gamma^2} \exp \left[ -2 \left[ \frac{\gamma}{1+\gamma} \right]^2 t^2 / n \right].$$

Proposition (7.8) is a tightening of the original result of Maurey (1979), though it is still not clear how tight it is – see Bollobás (1987). In order to set these last results in a more general framework let us introduce some definitions. See Milman & Schechtman (1986) for more background. For a graph  $G$  with vertex set  $V$  and diameter  $D$ , and for any  $0 < \epsilon < 1$  let

$$\alpha(G, \epsilon) = \min \{ 1 - |A_{\epsilon D}| / |V| : A \subseteq V, |A| / |V| \geq 1/2 \}.$$

A sequence  $G_1, G_2, \dots$  of graphs is a Lévy family if  $\alpha[G_n, \epsilon] \rightarrow 0$  as  $n \rightarrow \infty$ , for every  $\epsilon$ . It is a concentrated Lévy family if there are  $c_1, c_2 > 0$  such that

$\alpha[G_n, \epsilon] \leq c_1 \exp[-c_2 \epsilon n^{1/2}]$  for all  $n$  and  $\epsilon$ . It is a normal Lévy family if there are  $c_1, c_2 > 0$  such that  $\alpha[G_n, \epsilon] \leq c_1 \exp[-c_2 \epsilon^2 n]$  for all  $n$  and  $\epsilon$ .

The propositions above show that both the family  $[Q_n]$  of cubes and the family  $[S_n]$  of permutation graphs form normal Lévy families with parameter  $c_2$  arbitrarily close to 2. In fact for the  $n$ -cube  $Q_n$  we can do slightly better — see the next subsection.

The notion of partition sequences above is quite a natural framework within which to apply the martingale inequalities (6.7) or (6.11), but it is not obvious that we have hit the 'right' level of generality. Milman & Schechtman (1986) give an attractive intermediate level which neatly contains examples (7.1) to (7.3). We shall consider only the finite case here.

Let  $G$  be a (finite) group with a translation invariant metric  $d$ ; that is,  $d(g, h) = d(rg, rh) = d(gr, hr)$  for all  $g, h, r \in G$ . Given a subgroup  $H$  we have a natural metric  $\bar{d}$  defined on the set  $G/H$  of left cosets  $rH$  by setting  $\bar{d}(rH, sH) = d(r, sH)$  where of course  $d(r, sH)$  is the minimum value of  $d(r, sh)$  over  $h \in H$ .

(7.9) Theorem: Let  $G$  be a group with a translation invariant metric  $d$ . Let  $G = G_0 \supseteq G_1 \supseteq \dots \supseteq G_n = \{e\}$  be a decreasing sequence of subgroups, and let  $c_k$  be the diameter of the space  $G_{k-1}/G_k$  for each  $k = 1, \dots, n$ .

(a) Let the real-valued function  $f$  on  $G$  satisfy  $|f(x) - f(y)| \leq d(x, y)$  for all  $x, y \in G$ .

Let  $X$  be uniformly distributed over  $G$ . Then for any  $t > 0$

$$P(|f(X) - Ef(X)| \geq t) \leq 2 \exp\left[-2t^2 / \sum c_k^2\right].$$

(b) Let  $A$  be a subset of  $G$  with  $|A|/|G| = \alpha$ ,  $0 < \alpha < 1$ . Then for any

$$t \geq t_0 = \left[ \frac{1}{2} \left[ \log \frac{1}{\alpha} \right] \sum c_k^2 \right]^{1/2},$$

$$|A_t|/|G| \geq 1 - \exp\left[-2[t - t_0]^2 / \sum c_k^2\right];$$

and so for any  $t > 0$  and  $\gamma > 0$

$$|A_t|/|G| \geq 1 - \left[ \frac{1}{\alpha} \right]^{\gamma^2} \exp\left[-2 \left[ \frac{\gamma}{1+\gamma} \right]^2 t^2 / \sum c_k^2\right].$$

**(b) Exact isoperimetric inequalities**

A Hamming ball centred at a vertex  $v$  of the cube  $Q_n$  consists of all vertices at distance less than  $d$  from  $v$  for some  $d$ , together with some vertices at distance  $d$ .

Let  $A$  and  $C$  be two sets of vertices of  $Q_n$  with distance  $\rho$  between them. Then a seminal theorem of Harper (1966) (see also Bollobás (1986) page 129) asserts that there exist Hamming balls  $B_0$  centred at the all-zero vector and  $B_1$  centred at the all-one vector such that  $|B_0| = |A|$ ,  $|B_1| = |C|$  and the distance between  $B_0$  and  $B_1$  is at least  $\rho$ . This gives the exact solution to the isoperimetric inequality for the  $n$ -cube. From this result and a one-sided version of the Chernoff bounds (1.1) we may obtain the following result (see Amir & Milman (1980), Alon & Milman (1985)).

(7.10) Proposition: Let  $A$  be a set of  $2^{n-1}$  vertices in the cube  $Q_n$ . Then for any  $t \geq 0$ ,

$$|A_t|/2^n \geq 1 - \exp[-2t^2/n].$$

This result is 'cleaner' than the corresponding case of proposition (7.7). We now see that the graphs  $Q_n$  form a normal Lévy family with parameters  $c_1 = 1$ ,  $c_2 = 2$ .

In the last subsection we obtained our isoperimetric inequality (7.7) for the cube from a result on concentration of measure. We may also reverse this process.

(7.11) Proposition: Let  $f$  be a function defined on the vertex set  $V$  of the  $n$ -cube  $Q_n$  such that if  $x$  and  $y$  are adjacent then  $|f(x) - f(y)| \leq 1$ . Let the random variable  $X$  be uniformly distributed over  $V$ . Let  $L$  be a median of  $f$ ; that is,  $P(f(X) \leq L) \geq \frac{1}{2}$  and  $P(f(X) \geq L) \geq \frac{1}{2}$ . Then for any  $t > 0$

$$P(|f(X) - L| \geq t) \leq 2 \exp[-2t^2/n].$$

Proof: Let  $A = \{x \in V: f(x) \leq L\}$  and  $B = \{x \in V: f(x) \geq L\}$ . Then

$$\begin{aligned} P(|f(X) - L| \leq t) &\geq P[X \in A_t \cap B_t] \\ &= -1 + P[X \in A_t] + P[X \in B_t] \\ &\geq 1 - 2\exp[-2t^2/n] \end{aligned}$$

by proposition (7.10). □

(We already knew this inequality with  $L$  replaced by  $Ef(X)$ , for example by lemma (1.2) with each  $c_k = 1$ . See also Milman & Schechtman (1986), page 142, for the relationship between deviations from the mean and from the median.)

Recently several further exact discrete isoperimetric inequalities have been obtained — see Bollobás & Leader (1988a), (1988b), (1988c) and subsection (d) below.

### (c) Two results of Alon and Milman

Let  $H_k$  be a graph with vertex set  $V_k$ , for  $k=1, \dots, n$ . The cartesian product  $G = \Pi H_k$  has vertex set  $\Pi V_k$ , and two vertices  $x = [x_1, \dots, x_n]$  and  $y = [y_1, \dots, y_n]$  are adjacent if and only if they differ in exactly one coordinate and if this is the  $k$ th co-ordinate then  $x_k$  and  $y_k$  are adjacent in  $H_k$ . Note that  $G$  has diameter the sum of the diameters of the graphs  $H_k$ . From corollary (7.5) we deduce the following extension of proposition (7.7) about the  $n$ -cube.

(7.12) Proposition: For  $k = 1, \dots, n$  let  $H_k$  be a graph with diameter  $D_k$ , and let the graph  $G_n$  with vertex set  $V$  be the cartesian product  $\Pi H_k$ . Let  $A \subseteq V$  with

$$|A|/|V| = \alpha, \quad 0 < \alpha < 1. \quad \text{Then for any } t \geq t_0 = \left[ \frac{1}{2} \log \left[ \frac{1}{\alpha} \right] \Sigma D_k^2 \right]^{1/2},$$

$$|A_t|/|V| \geq 1 - \exp \left[ -2 \left[ t - t_0 \right]^2 / \Sigma D_k^2 \right];$$

and for any  $t > 0$  and any  $\gamma > 0$

$$|A_t|/|V| \geq 1 - \left[ \frac{1}{\alpha} \right]^{\gamma^2} \exp \left[ -2 \left[ \frac{\gamma}{1+\gamma} \right]^2 t^2 / \Sigma D_k^2 \right].$$

This result shows that if  $G_n$  is the Cartesian product of  $n$  copies of a fixed graph  $H$  then the  $G_n$ 's form a normal Lévy family with parameter  $c_2$  arbitrarily close to 2. In a remarkable paper, Alon & Milman (1985) used quite different methods to show that the family  $[G_n]$  is concentrated, and commented that the methods of Maurey and Schechtman which we are using here show that  $[G_n]$  is a normal Lévy family (though with  $c_2 = 1/64$ ). Bollobás and Leader (1988c) show that we may also take  $c_2 = 6D^2/[k^2-1]$  where  $k$  is the order of  $H$ .

Let us consider a second interesting result from Alon & Milman (1985). For  $k \geq 2$ , the odd graph  $O_k$  has vertices corresponding to the  $(k-1)$ -subsets of a  $(2k-1)$ -set, and two vertices are adjacent if and only if the corresponding sets are disjoint. Thus  $O_2$  is the triangle  $K_3$  and  $O_3$  is the Petersen graph; and the graph  $O_k$  has diameter  $k-1$  (see Biggs (1974)). It is shown in Alon & Milman (1985) that the family  $[O_k]$  is concentrated.

Now for  $1 \leq k \leq n$  let  $\hat{S}_{n,k}$  be the graph with vertices the  $k$ -subsets of an  $n$ -set, and with two vertices adjacent if and only if the corresponding sets  $A, B$  satisfy  $|A \setminus B| = |B \setminus A| = 1$ . Thus the graph  $\hat{S}_{2k-1, k-1}$  has the same vertices as  $O_k$ , and if two vertices are adjacent in this graph then they are at distance 2 in  $O_k$ . Observe that  $\hat{S}_{n,k}$  has diameter  $k$  if  $k \leq n/2$ .

(7.13) Proposition: Consider the graph  $\hat{G} = \hat{S}_{n,k}$ , with vertex  $\hat{V}$  say,  $|\hat{V}| = \binom{n}{k}$ . Let  $\hat{A} \subseteq \hat{V}$  with  $|\hat{A}|/|\hat{V}| = \alpha$ ,  $0 < \alpha < 1$ . Then for any  $t \geq t_0 = \left[ \frac{1}{2} \left[ \log \frac{1}{\alpha} \right] k \right]^{1/2}$ ,  

$$|\hat{A}_t|/|\hat{V}| \geq 1 - \exp \left[ -2 \left[ t - t_0 \right]^2 / k \right].$$

We may of course think of the vertices of the graph  $\hat{S}_{n,k}$  as members of the vertex set  $\{0,1\}^n$  of the  $n$ -cube  $Q_n$ . Then two vertices of  $\hat{S}_{n,k}$  are adjacent in  $\hat{S}_{n,k}$  if and only if they are at distance 2 in  $Q_n$ . Thus we see that the concentration phenomenon observed for the  $n$ -cube  $Q_n$  in proposition (7.7) holds also for the slice  $\sum x_i = k$ . From the above we find that the graphs  $O_k$  form a normal Lévy family with parameter  $c_2$  about  $1/2$ .

Proof: Consider the graph  $G = S_{n,k}$  in example (7.3), with vertex set  $V$ . For each vertex  $g$  in  $V$  let  $\hat{g}$  be the  $k$ -set of elements listed in  $g$ . Let  $A = \{g \in V: \hat{g} \in \hat{A}\}$ . Then  $|A|/|V| = |\hat{A}|/|\hat{V}|$ . Also, for each vertex  $g$  in  $V$

$$d_G(g, A) = d_{\hat{G}}(\hat{g}, \hat{A}).$$

Hence, for any  $t > 0$ ,  $|A_t|/|V| = |\hat{A}_t|/|\hat{V}|$ . Now we may complete the proof by applying corollary (7.5) to example (7.3).  $\square$

#### (d) Monotonic functions

We considered in proposition (7.7) above the uniform distribution over the vertex set  $V = \{0,1\}^n$  of the  $n$ -cube  $Q_n$ . Now consider a more general distribution, of particular interest in the theory of random graphs (with  $n$  as the number of edges in a complete graph). Let  $0 < p < 1$  and let the random variable  $X$  take values in  $V$ , with  $P(X=x) = p^s(1-p)^{n-s}$  where  $s = \sum x_k$ .

Bollobás and Leader (1986b) give a beautiful exact isoperimetric inequality for down-sets (hereditary properties) in  $Q_n$ . From this they deduce various estimates which they note are much better than can be obtained from Azuma's inequality lemma (4.1), for the case when  $p$  (or  $1-p$ ) is very small. The martingale inequalities (6.5), (6.6) give an alternative approach.

(7.14) Lemma: Let  $f$  be a non-decreasing function defined on the vertex set  $V$  of the  $n$ -cube  $Q_n$  such that if  $x$  and  $y$  are adjacent then  $|f(x) - f(y)| \leq 1$ . Let the random variable  $X$  be distributed over  $V$  as above. Then, for any  $0 \leq \epsilon \leq 1$ ,

$$P(f(X) - Ef(X) \geq \epsilon np) \leq \exp\left[-\frac{1}{3}\epsilon^2 np\right],$$

$$P(f(X) - Ef(X) \leq -\epsilon np) \leq \exp\left[-\frac{1}{2}\epsilon^2 np\right].$$

Proof: Let  $\mathcal{F}_k$  be the  $\sigma$ -field generated by the natural partition  $\mathcal{P}_k$  given in example (7.1), and let  $Y_k$  be  $E[f(X) | \mathcal{F}_k]$ . Then

$$-p \leq Y_k - Y_{k-1} \leq 1 - p,$$

and so we may use (6.5) and (6.6).  $\square$

(7.15) Proposition: Let the random variable  $X$  be distributed over the vertex set  $V$  of the  $n$ -cube  $Q_n$  as above; that is,  $P(X=x) = p^s(1-p)^{n-s}$  where  $s = \sum x_k$ . Let  $A \subseteq V$  be decreasing (that is,  $x \in A$  and  $y \leq x$  implies  $y \in A$ ) with  $P(X \in A) = \alpha$ ,  $0 < \alpha < 1$ .

Let  $t_0 = \left[ 2 \log \left[ \frac{1}{\alpha} \right] np \right]^{1/2}$ . Then for  $t_0 \leq t \leq t_0 + np$ ,  

$$P[X \in A_t] \geq 1 - \exp \left[ -\frac{1}{3} [t - t_0]^2 / np \right].$$

Proof: Take  $f(x) = d(x, A)$  in the lemma above. Note that  $d(x, A) \leq \sum x_i$  since the all-zero vector is in  $A$ . Thus  $t_1 = Ef(X)$  satisfies  $t_1 \leq np$ . By lemma (7.14)

$$\alpha = P(f(X) = 0) = P(f(X) \leq t_1 - t_1) \leq \exp \left[ -\frac{1}{2} t_1^2 / np \right],$$

and it follows that  $t_1 \leq t_0$ . But now by lemma (7.14) again, for  $t_0 \leq t \leq t_0 + np$

$$\begin{aligned} 1 - P[X \in A_t] &= P(f(X) > t) \\ &\leq P[f(X) > t_1 + [t - t_0]] \\ &\leq \exp \left[ -\frac{1}{3} [t - t_0]^2 / np \right]. \end{aligned}$$

□

## §8 Applications in operational research and computer science

### (a) Bin packing

Given an  $n$ -vector  $\underline{x} = [x_1, \dots, x_n]$  where each  $x_i \in [0, 1]$ , let  $B(\underline{x})$  be the least number of unit size bins needed to store  $n$  items with these sizes. Let  $X_1, \dots, X_n$  be independent random variables each taking values in  $[0, 1]$ . The bounded differences method lets us much strengthen previous work on the behaviour of the random variable  $B = B[X_1, \dots, X_n]$  (Grimmett (1985), Rhee (1985)).

(8.1) Theorem: For any  $t > 0$

$$P(|B - E(B)| \geq t) \leq 2 \exp \left[ -2t^2 / n \right].$$

This is the first application in Rhee & Talagrand (1987a) (with an improved bound). To prove it from lemma (1.2) we need only note that for  $\underline{x}, \underline{x}' \in [0, 1]^n$  we

have  $|B(\underline{x}) - B(\underline{x}')| \leq 1$  whenever  $\underline{x}$  and  $\underline{x}'$  differ in only one coordinate. Similar results hold for the number of bins used by certain heuristics — see Rhee & Talagrand (1987a).

Now let us use  $B_n$  in place of  $B$  above. From the subadditive inequality

$$E[B_{m+n}] \leq E[B_m] + E[B_n]$$

it follows that  $\frac{1}{n}E[B_n] \rightarrow \beta$  as  $n \rightarrow \infty$ , where  $\beta = \inf \frac{1}{n}E[B_n]$  ( $0 \leq \beta \leq 1$ ).

(8.2) Corollary: Let  $\epsilon > 0$ . Then

$$P\left[\left|\frac{1}{n}B_n - \beta\right| > \epsilon\right] = O\left[\exp\left\{-(2-o(1))\epsilon^2 n\right\}\right].$$

### (b) Knapsack problems

Let  $\underline{b}$  be a fixed non-negative  $m$ -vector. Consider a list  $\underline{x}_1 = [c_1, \underline{a}_1]$ , ...,  $\underline{x}_n = [c_n, \underline{a}_n]$  of  $(1+m)$ -vectors, where each  $c_k \in [0,1]$  and each  $m$ -vector  $\underline{a}_k \geq \underline{0}$ . Denote the list  $[\underline{x}_1, \dots, \underline{x}_n]$  by  $\underline{x}$ , and let  $K(\underline{x})$  be the value of the corresponding 'multi-knapsack' problem

$$\max \sum_k c_k z_k$$

subject to

$$\sum_k \underline{a}_k z_k \leq \underline{b}$$

$$z_k = 0 \text{ or } 1 \quad (k = 1, \dots, n).$$

Now let  $\underline{X}_1, \dots, \underline{X}_n$  be independent random variables, where each  $\underline{X}_k$  is a  $(1+m)$ -vector  $[C_k, \underline{A}_k]$ , with  $C_k \in [0,1]$  and the  $m$ -vector  $\underline{A}_k \geq \underline{0}$ . The behaviour of the corresponding random knapsack value  $K = K[\underline{X}_1, \dots, \underline{X}_n]$  has been investigated in Frieze & Clarke (1984), Meante et al (1984), Mamer & Schilling (1988), Schilling (1988).

As with bin packing we obtain a strong concentration result immediately from lemma (1.2). We need only note that if the lists  $\underline{x}$ ,  $\underline{x}'$  differ in only one coordinate vector  $[c_k, \underline{a}_k]$  then  $|K(\underline{x}) - K(\underline{x}')| \leq 1$ , to obtain

(8.3) Theorem: For any  $t > 0$ ,

$$P(|K - E(K)| \geq t) \leq 2\exp[-2t^2/n].$$

**(c) Travelling salesman problem**

Given  $n$  points  $\underline{x}_1, \dots, \underline{x}_n$  in the unit square  $[0,1]^2$  let  $T[\underline{x}_1, \dots, \underline{x}_n]$  be the shortest length of a tour through them. Now let  $\underline{X}_1, \dots, \underline{X}_n$  be independent random variables, each uniformly distributed on the unit square. Marvellous results are known related to the asymptotic behaviour of the random variable  $T = T[\underline{X}_1, \dots, \underline{X}_n]$ , starting with the seminal paper of Beardwood, Halton and Hammersley (1959) – see the survey article by Karp & Steele (1985). We are interested here in the concentration of  $T$ .

Given a fixed point  $\underline{y}$  in the unit square let  $Y$  be the random shortest distance from  $\underline{y}$  to one of the points  $\underline{X}_{k+1}, \dots, \underline{X}_n$ . Then, as observed in Steele (1981), for some constant  $c_1 > 0$

$$P(Y > t) \leq [1 - c_1 t^2]^{n-k-1} \quad (\text{for } t > 0).$$

Hence  $E(Y) \leq c_2(n-k)^{-1/2}$  for some constant  $c_2 > 0$ . It follows by considering first  $\underline{y} = \underline{x}_k$  then  $\underline{y} = \underline{x}'_k$  that

$$\begin{aligned} |E[T | [\underline{X}_1, \dots, \underline{X}_k] = [\underline{x}_1, \dots, \underline{x}_k]] - E[T | [\underline{X}_1, \dots, \underline{X}_{k-1}] = [\underline{x}_1, \dots, \underline{x}_{k-1}], \underline{X}_k = \underline{x}'_k]| \\ \leq 2c_2(n-k)^{-1/2}. \end{aligned}$$

So by Azuma's inequality lemma (4.1) we find that for  $t > 0$

$$(8.4) \quad P(|T - E(T)| \geq t) \leq \exp[-t^2/c \log n]$$

for a suitable constant  $c > 0$ . This result of Rhee & Talagrand (1987a) improves on work of Steele (1981) for large deviations  $t$ .

The inequality (8.4) is of no use if we are interested in small  $t$ , say  $t = o[(\log n)^{1/2}]$ . For such small deviations direct application of the bounded difference method fails – see Rhee & Talagrand (1987a), (1987b), (1989), Rhee (1988), Steele (1989).

**(d) Minimum spanning trees**

In the complete graph  $K_n$  with independent random edge lengths each uniformly distributed on  $[0,1]$ , the expected length of a minimum spanning tree tends to  $\zeta(3) \simeq 1.2$  as  $n \rightarrow \infty$ . This remarkable result is due to Frieze (1985). In Frieze & McDiarmid (1989) the result is extended and strengthened. This was possible partly because we could replace a messy second moment calculation by a simple use of the bounded differences method, and obtain a far stronger result.

Let us see how this goes. Let  $G$  be a fixed graph, with say  $n$  edges. Given a list  $\underline{x} = [x_1, \dots, x_n]$  of these edges, let  $G_j(\underline{x})$  be the subgraph of  $G$  with edges  $x_1, \dots, x_j$ , and let  $\kappa_j(\underline{x})$  be the number of components of  $G_j(\underline{x})$ . Now suppose that the edges of  $G$  have independent random edge lengths each uniformly distributed on  $[0,1]$ . Let  $\underline{X} = [X_1, \dots, X_n]$  be the list of edges rearranged in increasing order. Thus all  $n!$  possible orders are equally likely. Knowing how fast  $\kappa_j(\underline{X})$  usually decreases with  $j$  tells us how much of the minimum spanning tree we can expect to build with short edges.

It turns out then that we are interested in the random variable  $Z = Z(\underline{X})$ , where  $Z(\underline{x}) = \sum \left\{ \kappa_j(\underline{x}) : j=1, \dots, m \right\}$  and  $m \leq n$ . Here we focus on the concentration of  $Z$ . But as in example (7.3) we see that

$$\begin{aligned} & |E[\kappa_j(\underline{X}) | [X_1, \dots, X_k] = [e_1, \dots, e_k]] \\ & \quad - E[\kappa_j(\underline{X}) | [X_1, \dots, X_{k-1}] = [e_1, \dots, e_{k-1}], X_k = e'_k] | \leq 1. \end{aligned}$$

It now follows that

$$\begin{aligned} & |E(Z | [X_1, \dots, X_{k-1}] = [e_1, \dots, e_{k-1}]) \\ & \quad - E(Z | [X_1, \dots, X_{k-1}] = [e_1, \dots, e_{k-1}], X_k = e'_k) | \leq m - k + 1. \end{aligned}$$

Hence by corollary (6.10), for any  $t > 0$

$$(8.5) \quad P(|Z - E(Z)| \geq t) \leq 2 \exp[-12t^2/m(m+1)(2m+1)].$$

**(e) Second eigenvalue of random regular graphs**

Let  $G$  be an  $r$ -regular graph, possibly with loops or multiple edges. The adjacency matrix (with loops counted twice) has  $r$  as the eigenvalue with

largest absolute value. Let  $\lambda_2(G)$  be the next largest eigenvalue in absolute value. If  $|\lambda_2(G)|$  is much less than  $r$  then  $G$  has useful 'expansion' properties and the natural random walk on the vertices is 'rapidly mixing' — see for example Broder & Shamir (1987) and the references there. Both these properties are much sought in computer science.

Consider random  $2d$ -regular graphs constructed as follows on the vertex set  $V = \{1, \dots, n\}$ . Let  $A$  be the set of all permutations  $\sigma$  on  $V$ . Pick  $\sigma_1, \dots, \sigma_d$  independently at random from  $A$ , and let the graph  $G$  have the  $dn$  edges  $\{v, \sigma_k v\}$  for  $k = 1, \dots, d$  and  $v \in V$ . Let  $Y$  be the corresponding random variable  $|\lambda_2(G)|$ . We are interested in large  $n$  and moderate  $d$ . Broder & Shamir (1987) show essentially that  $E[Y] \leq (c + o(1))d^{3/4}$  as  $n \rightarrow \infty$ ; and for any  $t > 0$

$$(8.6) \quad P(|Y - E(Y)| \geq t) \leq 2 \exp[-t^2/8d].$$

The concentration result (8.6) may be proved as follows. Standard manipulations show that

$$2d - \lambda_2 = \inf \sum_k \sum_v [f(\sigma_k v) - f(v)]^2,$$

where the infimum is over all real-valued functions  $f$  on  $V$  satisfying  $\sum_v f(v) = 0$  and  $\sum_v f(v)^2 = 1$ . But if  $\sum_v f(v)^2 = 1$  the triangle inequality gives

$$\sum_v [f(\sigma_k v) - f(v)]^2 \leq 4.$$

Hence we may apply lemma (1.2) with each  $c_k = 4$ .

This neat application of the bounded differences method has been rather eclipsed by recent work of Friedman (1988) and Kahn & Szemerédi (1988), who show that  $Y$  concentrates near the lower bound of about  $2(2d-1)^{1/2}$ .

### (f) Heap building

Suppose that we wish to build a heap on  $n$  elements, when all  $n!$  initial orders are equally likely (see for example Knuth (1973)). In McDiarmid & Reed (1989) a new algorithm is introduced, and it is shown that the random number  $B_n$  of comparisons required satisfies  $\frac{1}{n}E[B_n] \rightarrow \alpha$  as  $n \rightarrow \infty$ , where  $\alpha \simeq 1.52$ . (This is the

best known average case behaviour.) The result is given extra weight by the fact that  $B_n$  concentrates strongly around the mean. This concentration is our interest here.

The algorithm is a variant of Floyd's method (see Knuth (1973)) and consists of a sequence of 'merge' operations moving up a binary tree. We can use corollary (6.10), since the average effect on  $B_n$  of learning the history of one more merge is small. The reason for this is that as we move up the binary tree the expected number of times we revisit this old working is at most 1. We find the following result, and similar results for other variants of Floyd's method.

(8.7) Theorem For any  $\epsilon > 0$ , if  $n$  is sufficiently large

$$P\left[\left|\frac{1}{n}B_n - \alpha\right| \geq \epsilon\right] < \exp\left[-\left[\epsilon^2/9\right]n\right].$$

### §9 Concluding Remarks

The bounded differences method is just the application of certain inequalities related to bounded martingale difference sequences, but we have seen that it is wonderfully useful in combinatorics and the mathematics of Operational Research. As long as there are small bounds and large deviations the inequalities really bite, and allow us to handle problems that seemed intractable only a short time ago.

How many more interesting applications will have appeared by the time of the conference in July?

**Acknowledgement** I would like to thank Martin Dyer for helpful comments.

**References**

- Alon, N. & Milman, V.D. (1985).  $\lambda_1$ , isoperimetric inequalities for graphs, and superconcentrators. *J. Comb. Th. B* 38, 73–88.
- Amir, D. & Milman, V.D. (1980). Unconditional and symmetric sets in  $n$ -dimensional normed spaces. *Israel J. Math.* 37, 3–20.
- Angluin, D. & Valiant, L.G. (1979). Fast probabilistic algorithms for Hamiltonian circuits and matchings. *J. Computer and System Sciences* 18, 155–193.
- Azuma, K. (1967). Weighted sums of certain dependent random variables. *Tôkoku Math. J.* 19, 357–367.
- Bahadur, R.R. (1971). Some limit theorems in Statistics. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics 4, SIAM, Philadelphia, USA.
- Bahadur, R.R. & Ranga Rao, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* 31, 1015–1027.
- Beardwood, J., Halton, J.H. and Hammersley, J. (1959). The shortest path through many points. *Proc. Camb. Phil. Soc.* 55, 299–327.
- Biggs, N. (1974). *Algebraic Graph Theory*. Cambridge: Cambridge University Press.
- Bollobás, B. (1985). *Random Graphs*. London: Academic Press.
- Bollobás, B. (1986). *Combinatorics*. Cambridge: Cambridge University Press.
- Bollobás, B. (1987). Martingales, isoperimetric inequalities and random graphs. *Colloq. Math. Soc. János Bolyai* 52, 113–139.
- Bollobás, B. (1988a). The chromatic number of random graphs. *Combinatorica* 8, 49–55.
- Bollobás, B. (1988b). Sharp concentration of measure phenomena in the theory of random graphs. manuscript.
- Bollobás, B. & Erdős, P. (1976). Cliques in random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society* 80, 419–427.
- Bollobás, B. & Leader, I. (1988a). An isoperimetric inequality on the discrete torus. manuscript.
- Bollobás, B. & Leader, I. (1988b). Isoperimetric inequalities and fractional set systems. manuscript.
- Bollobás, B. & Leader, I. (1988c). Compressions and isoperimetric inequalities. manuscript.
- Broder, A. & Shamir, E. (1987). On the second eigenvalue of random regular graphs. *In* 28th Annual Symposium on Foundations of Computer Science, 286–294.

- Burkholder, D.L. (1973). Distribution function inequalities. *Ann. Probab.* 1, 19–42.
- Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypotheses based on the sum of observations. *Ann. Math. Statist.* 23, 493–507.
- Chung, K.L. (1974). A course in probability theory, Second edition. New York and London: Academic Press.
- Chvátal, V. (1979). The tail of the hypergeometric distribution. *Discrete Math.* 25, 285–287.
- Doob, J.L. (1953). *Stochastic Processes*. New York: John Wiley and Sons.
- Erdős, P. & Spencer, J. (1974). *Probabilistic Methods in Combinatorics*. New York: Academic Press.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Volume 1, Third Edition*. New York: John Wiley and Sons.
- Freedman, D.A. (1975). On tail probabilities for martingales. *Ann. Probab.* 3, 100–118.
- Friedman, J. (1988). On the second eigenvalue and random walks in random  $d$ -regular graphs. manuscript.
- Frieze, A.M. (1985). On the value of a random minimum spanning tree problem. *Discrete Applied Math.* 10, 47–56.
- Frieze, A.M. (1989). On the independence number of random graphs. *Discrete Math.*, to appear.
- Frieze, A.M. & Clarke, M.R.B. (1984). Approximation algorithms for the  $m$ -dimensional 0–1 knapsack problem: worst case and probabilistic analysis. *Europ. J. O. R.* 15, 100–109.
- Frieze, A.M. & Luczak, T. (1988). On the independence number of random regular graphs. manuscript.
- Frieze, A.M. & McDiarmid, C.J.H. (1989). On random minimum length spanning trees. *Combinatorica*, to appear.
- Garsia, A.M. (1973). *Martingale inequalities, Seminar Notes on Recent Progresses*. New York: W.A. Benjamin.
- Gleser, L.J. (1975). On the distribution of the number of successes in independent trials. *Ann. Probab.* 3, 182–188.
- Grimmett, G.R. (1985). Large deviations in subadditive processes and first-passage percolation. *Contemporary Mathematics* 41, 175–194.
- Grimmett, G.R. & McDiarmid, C.J.H. (1975). On colouring random graphs. *Mathematical Proceedings of the Cambridge Philosophical Society* 77, 313–324.
- Harper, L.H. (1966). Optimal numberings and isoperimetric problems on graphs. *J. Comb. Th.* 1, 385–393.

- Hoeffding, W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.* 27, 713–721.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58, 13–30.
- Johnson, W.B., Schechtman, G. & Zinn, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *Ann. Prob.* 13, 234–253.
- Kahn, J. & Szemerèdi, E. (1988). manuscript.
- Karp, R.M. (1979). A patching algorithm for the nonsymmetric travelling-salesman problem. *SIAM J. Comput.* 8, 561–573.
- Karp, R.M., Lenstra, J.K., McDiarmid, C.J.H. & Rinnooy Kan, A.H.G. (1985). Probabilistic analysis of combinatorial algorithms: an annotated bibliography. *In* *Combinatorial Optimisation: Annotated Bibliographies*, edited by M. O'hEigeartaigh, J.K. Lenstra and A.H.G. Rinnooy Kan. Chichester, England: John Wiley and Sons.
- Karp, R.M. & Steele, J.M. (1985). Probabilistic analysis of heuristics. *In* *The Travelling Salesman Problem: A Guided Tour of Combinatorial Optimization*, eds. E.L. Lawler et al. New York: John Wiley and Sons.
- Kingman, J.F.C. (1976). Subadditive processes. *Lecture Notes in Math.* 539, 168–222. Berlin and New York: Springer-Verlag.
- Knuth, D.E. (1973). *The Art of Computer Programming, Volume III: Sorting and Searching*. Reading, Mass: Addison-Wesley.
- Korsunov, A.D. (1980). The chromatic number of  $n$ -vertex graphs. *Metody Diskret. Analiz.* No.35, 14–44, 104 (in Russian).
- Luczak, T. (1988a). Note on the sharp concentration of the chromatic number of random graphs. manuscript.
- Luczak, T. (1988b). The chromatic number of random graphs. manuscript.
- McDiarmid, C.J.H. (1982). Achromatic numbers of random graphs. *Math. Proc. Camb. Phil. Soc.* 92, 21–28.
- McDiarmid, C.J.H. (1984). Colouring random graphs. *Ann. Oper. Res.* 1, 183–200.
- McDiarmid, C.J.H. (1989). On the chromatic number of random graphs. manuscript.
- McDiarmid, C.J.H. (1989). Probabilistic analysis of tree search. manuscript.
- McDiarmid, C.J.H. & Reed, B.A. (1989). Building heaps fast. *J. Algorithms*, to appear.
- Mamer, J.W. & Schilling, K.E. (1988). On the growth of random knapsacks. manuscript.

- Matula, D.W. (1970). On the complete subgraphs of a random graph. *Combinatorial Mathematics and its Applications*, Chapel Hill, 356–369.
- Matula, D.W. (1972). The employee party problem. *Notices A.M.S.* 19, A–382.
- Matula, D.W. (1976). The largest clique size in a random graph. Technical Report, Department of Computer Science, Southern Methodist University, Dallas, Texas.
- Matula, D.W. (1987). Expose and merge exploration and the chromatic number of a random graph. *Combinatorica* 7, 275–284.
- Maurey, B. (1979). Construction de suites symétriques. *Compt. Rend. Acad. Sci. Paris* 288, 679–681.
- Meante, M., Rinnooy Kan, A.H.G., Stougie, L. & Vercellis, C. (1984). A probabilistic analysis of the multiknapsack value function. manuscript.
- Milman, V. & Schechtman, G. (1986). Asymptotic theory of finite dimensional normed spaces. *Lecture Notes in Math.* 1200. Berlin and New York: Springer-Verlag.
- Rhee, W.T. (1985). Convergence of optimal stochastic bin packing. *Operations Research Letters* 4, 121–123.
- Rhee, W.T. (1988). On the fluctuations of the stochastic travelling salesperson problem. manuscript, Ohio State University.
- Rhee, W.T. & Talagrand, M. (1987a). Martingale inequalities and NP-complete problems. *Math. of O. R.* 12, 177–181.
- Rhee, W.T. & Talagrand, M. (1987b). Martingale inequalities, interpolation and NP-complete problems. manuscript.
- Rhee, W.T. & Talagrand, M. (1987c). Martingale inequalities and the jackknife estimate of variance. manuscript.
- Rhee, W.T. & Talagrand, M. (1989). A sharp deviation inequality for the stochastic travelling salesman problem. *Ann. Probab.* 17, 1–8.
- Schechtman, G. (1982). Lévy type inequality for a class of finite metric spaces. *Lecture Notes in Math.* 939, 211–215. Berlin and New York: Springer-Verlag.
- Scheinerman, E.R. (1989). On the interval number of random graphs. *Discrete Math.*, to appear.
- Schilling, K.E. (1988). On the growth of m-dimensional random knapsacks. manuscript.
- Shamir, E. (1988a). A slightly random source confronts a random witness-set. manuscript.
- Shamir, E. (1988b). Chromatic numbers of random hypergraphs and associated graphs. *In Randomness and Computation*, ed. S. Micali. Greenwich, Connecticut: JAI Press.

- Shamir, E. (1988c). Generalised stability and chromatic numbers of random graphs. manuscript.
- Shamir, E. & Spencer, J. (1987). Sharp concentration of the chromatic number on random graphs  $G_{n,p}$ . *Combinatorica* 7, 121–129.
- Steele, J.M. (1981). Complete convergence of short paths and Karp's algorithm for the TSP. *Math. Oper. Res.* 6, 374–378.
- Steele, J.M. (1989). Seedlings in the theory of shortest paths. manuscript.
- Steiger, W.L. (1967). Some Kolmogoroff-type inequalities for bounded random variables. *Biometrika* 54, 641–648.
- Steiger, W.L. (1969). A best possible Kolmogoroff-type inequality for martingales and a characteristic property. *Ann. Math. Statistics* 40, 764–769.
- Steiger, W.L. (1970). Bernstein's Inequality for Martingales. *Z. Wahrscheinlichkeitstheorie verw. Geb.* 16, 104–106.
- Stout, W.F. (1974). *Almost Sure Convergence*. New York: Academic Press (lemma 4–2–3 and exercise 4–2–2).

Institute of Economics and Statistics,  
Oxford University,  
St. Cross Building,  
Manor Road,  
Oxford,  
OX1 3UL.