Combinatorics, Probability and Computing (2002) 11, 163–178. © 2002 Cambridge University Press
DOI: 10.1017/S0963548301005089 Printed in the United Kingdom

Concentration for Independent Permutations

COLIN McDIARMID

Department of Statistics, University of Oxford,
1 South Parks Road, Oxford OX1 3TG, UK
(e-mail: cmcd@stats.ox.ac.uk)

Received 27 May 2001; revised 6 September 2001

We give a concentration inequality involving a family of independent random permutations, which is useful for analysing certain randomized methods for graph colouring.

1. Introduction and main result

In this paper we present an extended version of a concentration inequality based on the work of Talagrand [7], which concerns a family of independent random permutations. One particular use is for analysing randomized methods for graph colouring that involve randomly relabelling the colours used in different parts of the graph: see [5], and see Section 3 below, where we discuss part of the analysis of such a method.

We now introduce the main result, Theorem 1.1 (though, for example, Theorem 2.1 may be of independent interest). Theorem 1.1 concerns the concentration of a real-valued random variable $Z = h(\mathbf{Y})$ about its median, where the random vector $\mathbf{Y}$ has independent coordinates that are either real-valued random variables or random permutations, and the function h satisfies certain conditions, for example that swapping any two elements in one of the permutations can change the value of h by at most c. Now for the details.

Given a finite non-empty set S, we may specify a permutation σ on S by the vector $(\sigma(i) : i \in S)$. Let $\mathrm{Sym}(S)$ denote the set of all $|S|!$ permutations σ of S. Let $(B_1, B_2, \ldots)$ be a finite family of finite non-empty sets, and let G denote the product $\prod_k \mathrm{Sym}(B_k)$. (We should perhaps say explicitly that the family is $(B_1, \ldots, B_n)$ for some n, and $G = \prod_{k=1}^{n} \mathrm{Sym}(B_k)$: however, we choose to avoid introducing this extra piece of notation, since the results do not depend on n.) Let $\pi = (\pi_1, \pi_2, \ldots)$ be a family of independent random permutations, where $\pi_k \in_U \mathrm{Sym}(B_k)$ for each k. This notation is used to mean that the random permutation π_k is uniformly distributed over the set $\mathrm{Sym}(B_k)$. Thus $\pi \in_U G$.

Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a finite family of independent random variables, where the random variable X_j takes values in a set Ω_j. Thus $\mathbf{X}$ takes values in the set $\Omega' = \prod_j \Omega_j$.

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

Assume that $\mathbf{X}$ and π are independent. Let $\Omega = \Omega' \times G$, so that the random variable $\mathbf{Y} = (\mathbf{X}, \pi)$ takes values in Ω.

Let c and r be positive constants, and suppose that the nonnegative real-valued function h on Ω satisfies the following three conditions for each $(\mathbf{x}, \sigma) \in \Omega$.

- Changing the value of a coordinate x_j can change the value of $h(\mathbf{x}, \sigma)$ by at most $2c$.
- Swapping any two elements in any σ_k can change the value of $h(\mathbf{x}, \sigma)$ by at most c.
- If $h(\mathbf{x}, \sigma) = s$, then, in order to show that $h(\mathbf{x}, \sigma) \geqslant s$, we need specify only at most rs coordinates of $(\mathbf{x}, \sigma)$. In other words, if $h(\mathbf{x}, \sigma) = s$, then there is a set of at most rs coordinates, such that $h(\mathbf{x}', \sigma') \geqslant s$ for any $(\mathbf{x}', \sigma') \in \Omega$, which agrees with $(\mathbf{x}, \sigma)$ on these coordinates.

Finally, let m be a median of the random variable $Z = h(\mathbf{Y})$. The following result is typically used with c and r small, say 1 or 2, to show that Z is concentrated around its median m: see Section 3 below.

Theorem 1.1. *For each $t \geqslant 0$,*

$$P(Z \geqslant m + t) \leqslant 2\exp\left(-\frac{t^2}{16rc^2(m+t)}\right), \tag{1.1}$$

and

$$P(Z \leqslant m - t) \leqslant 2\exp\left(-\frac{t^2}{16rc^2 m}\right). \tag{1.2}$$

The relationship between this result and earlier work is discussed briefly in the first two paragraphs of the next section, when we sketch the main steps in the proof. The two inequalities (1.1) and (1.2) combine immediately to show that, for each $0 \leqslant t \leqslant m$,

$$P(|Z - m| \geqslant t) \leqslant 4\exp\left(-\frac{t^2}{32rc^2 m}\right). \tag{1.3}$$

It follows from these results that the mean $\mathbf{E}[Z]$ is close to the median m. Indeed, if c and r are constants then $|\mathbf{E}[Z] - m|$ is $O(\sqrt{m})$ as $m \to \infty$, for example by [4, Lemma 4.6], and thus Z is also concentrated around its mean.

The comments above may suggest that, for typical problems, the bounds for deviations above and below the median are fairly similar, and that the mean and the median tend to be fairly interchangeable. We need to be more careful if we allow a 'bad set' or 'failure set', where the conditions on the function h may fail: see Section 4 below.

2. Further results and proofs

We prove Theorem 1.1 above in three main steps. Theorem 5.1 of Talagrand [7] concerns a single random permutation uniformly distributed over the set of all permutations of a set. Our first step is to give an extension, Theorem 2.1 below, of this result. The proof is similar to the proof of Theorem 5.1 in [7], but avoids the awkward 'double induction' there.

Theorem 4.1.1 of Talagrand [7] concerns independent random variables, and has proved to be the one of Talagrand's inequalities which is the most useful in discrete mathematics

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

and theoretical computer science: see, for example, [4, 6]. Our second step is to use this theorem and the method of proof of Theorem 2.1 to yield a common extension of these two theorems, Theorem 2.7 below. Then, from a corollary to this last result we deduce Theorem 1.1 above, following the treatment of a similar result (Theorem 4.3) in [4] (see also [6]).

In order to prove these results, we first need to recall and adapt some definitions and notation from [7]. Let I be a finite non-empty set. Consider a vector $\mathbf{x} = (x_i : i \in I)$ and a non-empty set A in a product space $\Omega = \prod_{i \in I} \Omega_i$. We let $U(A, \mathbf{x})$ be the set of all binary vectors $\mathbf{u}$ such that, starting from $\mathbf{x}$, we may reach a vector $\mathbf{y} \in A$ by changing only coordinates x_i such that $u_i = 1$ (and not necessarily changing all of them). Thus $\mathbf{1}$ is always in $U(A, \mathbf{x})$, and $\mathbf{0} \in U(A, \mathbf{x})$ if and only if $\mathbf{x} \in A$. Let $V(A, \mathbf{x})$ be the convex hull of the set $U(A, \mathbf{x})$. The *norm* $\|\mathbf{x}\|$ of $\mathbf{x}$ is $\left(\sum_i x_i^2\right)^{1/2}$. Talagrand's *convex distance* $d_T(A, \mathbf{x})$ between A and $\mathbf{x}$ is given by

$$d_T(A, \mathbf{x}) = \min\{\|\mathbf{v}\| : \mathbf{v} \in V(A, \mathbf{x})\}. \tag{2.1}$$

There is an alternative equivalent definition of the convex distance. For a nonnegative unit n-vector $\alpha = (\alpha_i : i \in I)$, the *$\alpha$-Hamming distance* $d_\alpha(A, \mathbf{x})$ is the minimum over all vectors $\mathbf{y} \in A$ of $\sum\{\alpha_i : y_i \neq x_i\}$. Then ([7], or see [4, Lemma 4.10]), for any non-empty set $A \subseteq \Omega$,

$$d_T(A, \mathbf{x}) = \sup_\alpha d_\alpha(A, \mathbf{x}), \tag{2.2}$$

where the supremum (maximum) is over all nonnegative unit vectors α. We shall also use the function f defined by

$$f(A, \mathbf{x}) = d_T(A, \mathbf{x})^2. \tag{2.3}$$

Let $B_1 \cup B_2 \cup \cdots$ be a partition of a finite non-empty set S. Let us call the direct product G of the symmetric groups $\mathrm{Sym}(B_1), \mathrm{Sym}(B_2), \ldots$ a *product group* of permutations on S (acting in the obvious way). (In Representation Theory, such a group is called a Young subgroup of permutations.) Recall that a permutation σ on S is specified by the vector $(\sigma(i) : i \in S)$. Thus (as sets of vectors) we have $G = \prod_k \mathrm{Sym}(B_k)$. Given $\pi \in G$, let π_k denote the restriction $(\pi(i) : i \in B_k)$ of π to B_k. Then $\pi \in_U G$ if and only if $\pi_1, \pi_2, \ldots$ are independent and each $\pi_k \in_U \mathrm{Sym}(B_k)$. The next result extends Theorem 5.1 of Talagrand [7], as discussed above.

Theorem 2.1. *Let G be a product group of permutations on the set S, let $A \subseteq G$ be non-empty, and let $\pi \in_U G$. Then*

$$\mathbf{P}(\pi \in A)\, \mathbf{E}\left[\exp \frac{1}{16} f(A, \pi)\right] \leqslant 1.$$

The proof works by induction on the order of G. We first sketch the rough idea and introduce some notation we shall use later. From now on we shall not use the notation π_i to refer to a restriction of π: in the next paragraph we introduce a new meaning for the notation.

Suppose that S is a set of integers and that the orbit $\{\sigma(1) : \sigma \in G\}$ of the element 1

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

in S is $\{1,\ldots,m\}$ for some $m \geqslant 2$. Let H denote the stabilizer of this element, that is, H is the subgroup of G consisting of the permutations σ such that $\sigma(1) = 1$. We write $\sigma\tau$ for the permutation with $\sigma\tau(k) = \sigma(\tau(k))$. Let $H_1 = H$, and for $i = 2,\ldots,m$ let H_i denote the coset Ht_{1i} of H, where t_{ij} denotes the transposition of i and j. Thus G is partitioned into the m cosets H_i, where H_i consists of the permutations σ with $\sigma(i) = 1$, and each H_i has size $|G|/m$. Let $A_i = A \cap H_i$. For each distinct i and j, we can upper-bound $f(A,\sigma)$ (with coordinate i counted twice) in terms of $f(A_i,\sigma)$ (with coordinate j counted twice) and $f(At_{ij},\sigma)$. When $\pi_i \in_U H_i$, then $\pi_i t_{1i} \in_U H$, and $\pi_i t_{1i}$ is the 'same distance' from At_{1i} as π_i is from A. Such results allow us to use an induction hypothesis applied to $\pi_H \in_U H$ to handle $\pi \in_U G$.

For the detailed proof we need a number of lemmas. We first check that the set $V(A,\sigma)$ of vectors indexed by S transforms in the natural way under composition of permutations and taking inverses. Let $\tau \in \mathrm{Sym}(S)$: given a vector $\mathbf{x} = (x_i : i \in S)$ we let $\mathbf{x}\tau$ denote the vector with $(\mathbf{x}\tau)_k = x_{\tau(k)}$, and, given a set V of such vectors, we let $V\tau$ denote the corresponding set of vectors $\mathbf{x}\tau$ for $\mathbf{x} \in V$.

Lemma 2.2. *For non-empty $A \subseteq \mathrm{Sym}(S)$ and $\sigma,\tau \in \mathrm{Sym}(S)$,*

$$\begin{aligned} V(A\tau,\sigma\tau) &= V(A,\sigma)\tau, \\ V(A^{-1},\sigma^{-1}) &= V(A,\sigma)\sigma^{-1}. \end{aligned}$$

Proof. Let $\mathbf{u} \in U(A,\sigma)$, so that there exists $\rho \in A$ such that $\rho(k) = \sigma(k)$ whenever $u(k) = 0$. We use the notation $u(k)$ here rather than u_k to avoid lengthy suffices below. Let $\bar{\mathbf{u}} = \mathbf{u}\tau$, that is, $\bar{u}(k) = u(\tau(k))$ for each k. If $\bar{u}(k) = 0$, then $u(\tau(k)) = 0$, so

$$\rho\tau(k) = \rho(\tau(k)) = \sigma(\tau(k)) = \sigma\tau(k),$$

and so $\bar{\mathbf{u}} \in U(A\tau,\sigma\tau)$. It follows that $U(A,\sigma)\tau \subseteq U(A\tau,\sigma\tau)$, and by symmetry we obtain $U(A\tau,\sigma\tau) = U(A,\sigma)\tau$. Hence $V(A\tau,\sigma\tau) = V(A,\sigma)\tau$.

We prove the latter inequality in a similar way. Let $\mathbf{u} \in U(A,\sigma)$, so that there exists $\rho \in A$ such that $\rho(k) = \sigma(k)$ whenever $u(k) = 0$. Let $\bar{\mathbf{u}} = \mathbf{u}\sigma^{-1}$. If $\bar{u}(k) = 0$, then $u(\sigma^{-1}(k)) = 0$ so $\rho\sigma^{-1}(k) = \sigma\sigma^{-1}(k) = k$, and so $\rho^{-1}(k) = \sigma^{-1}(k)$. Since $\rho^{-1} \in A^{-1}$, we have $\bar{\mathbf{u}} \in U(A^{-1},\sigma^{-1})$. It follows that $U(A^{-1},\sigma^{-1}) \supseteq U(A,\sigma)\sigma^{-1}$, and by symmetry we obtain $U(A^{-1},\sigma^{-1}) = U(A,\sigma)\sigma^{-1}$. Hence $V(A^{-1},\sigma^{-1}) = V(A,\sigma)\sigma^{-1}$. □

Recall that

$$f(A,\sigma) = d_T(A,\sigma)^2 = \min\left\{\sum_{i\in S} v_i^2 : \mathbf{v} \in V(A,\sigma)\right\}.$$

For $j \in S$, let

$$f(A,\sigma,j) = \min\left\{v_j^2 + \sum_{i\in S} v_i^2 : \mathbf{v} \in V(A,\sigma)\right\}.$$

Thus coordinate j is 'counted twice' here. Observe that $f(A,\sigma,j) = f(A,\sigma)$ if $\tau(j) = \sigma(j)$ for each $\tau \in A$, and in particular if $\tau(j) = j$ for each $\tau \in A \cup \{\sigma\}$. The above lemma yields the following.

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

Lemma 2.3. *For non-empty* $A \subseteq \mathrm{Sym}(S)$, $\sigma, \tau \in \mathrm{Sym}(S)$ *and* $j \in S$,

$$\begin{aligned} f(A,\sigma) &= f(A\tau, \sigma\tau), \\ f(A,\sigma,j) &= f(A\tau, \sigma\tau, \tau^{-1}(j)), \\ f(A,\sigma,j) &= f(A^{-1}, \sigma^{-1}, \sigma(j)). \end{aligned}$$

□

The next lemma is similar to Lemma 5.3 in [7], and is crucial to the induction.

Lemma 2.4. *Let* $A \subseteq \mathrm{Sym}(S)$, *let* $\sigma \in \mathrm{Sym}(S)$, *let* $i, j \in S$ *be distinct, and let* $0 \leqslant \lambda \leqslant 1$. *Let* $A_i = \{\tau \in A : \tau(i) = \sigma(i)\}$ *and* $A_j = \{\tau \in A : \tau(j) = \sigma(j)\}$, *and suppose that* A_i *and* A_j *are nonempty. Then*

$$f(A,\sigma,i) \leqslant 4(1-\lambda)^2 + \lambda f(A_i,\sigma,j) + (1-\lambda) f(A_j t_{ij}, \sigma).$$

Proof. We need one more piece of notation: given distinct $i, j \in S$, let

$$g(A,\sigma,i,j) = \min\left\{ \sum_{l \neq i,j} v_l^2 : \mathbf{v} \in V(A,\sigma) \right\}.$$

Let $\mathbf{s} \in V(A,\sigma)$ satisfy $g(A,\sigma,i,j) = \sum_{l\neq i,j} s_l^2$, and let $\mathbf{t} \in V(A,\sigma)$ satisfy $t_i = 0$ and $f(A_i,\sigma,j) = \sum_{l\neq j} t_l^2 + 2t_j^2$. Since $V(A,\sigma)$ is convex, the vector $\mathbf{v} = (1-\lambda)\mathbf{s} + \lambda\mathbf{t}$ is in $V(A,\sigma)$. Thus

$$f(A,\sigma,i) \leqslant \sum_{l\neq i,j} v_l^2 + 2v_i^2 + v_j^2.$$

Now, since the function $x \to x^2$ is convex,

$$v_l^2 \leqslant (1-\lambda)s_l^2 + \lambda t_l^2.$$

Also

$$2v_i^2 = 2((1-\lambda)s_i)^2 \leqslant 2(1-\lambda)^2,$$

and

$$v_j^2 = ((1-\lambda)s_j + \lambda t_j)^2 \leqslant 2(1-\lambda)^2 s_j^2 + 2\lambda^2 t_j^2 \leqslant 2(1-\lambda)^2 + 2\lambda t_j^2.$$

Hence

$$\begin{aligned} f(A,\sigma,i) &\leqslant \sum_{l\neq i,j} ((1-\lambda)s_l^2 + \lambda t_l^2) + 4(1-\lambda)^2 + 2\lambda t_j^2 \\ &= 4(1-\lambda)^2 + (1-\lambda)\sum_{l\neq i,j} s_l^2 + \lambda\left(\sum_{l\neq j} t_l^2 + 2t_j^2\right) \\ &= 4(1-\lambda)^2 + (1-\lambda)g(A,\sigma,i,j) + \lambda f(A_i,\sigma,j). \end{aligned}$$

Finally note that

$$g(A,\sigma,i,j) \leqslant f(At_{ij},\sigma) \leqslant f(A_j t_{ij},\sigma).$$

□

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

It is convenient to state two further lemmas before we start the main proof of Theorem 2.1. The first is a form of Hölder's inequality (see, for example, [1, p. 465] or [4, Lemma 4.12]), which we state here in a form useful for us.

Lemma 2.5. *For any (appropriately integrable) functions f and g and random variables X and Y, and any $0 \leqslant \lambda \leqslant 1$,*

$$\mathbf{E}\big(e^{\lambda f(X)} e^{(1-\lambda)g(X)}\big) \leqslant \big(\mathbf{E}(e^{f(X)})\big)^{\lambda} \big(\mathbf{E}(e^{g(X)})\big)^{1-\lambda}.$$

The last preliminary result we need is from [2, 7] (or see [4, Lemma 4.11]).

Lemma 2.6. *For all $0 < r \leqslant 1$,*

$$\inf_{0 \leqslant \lambda \leqslant 1} r^{-\lambda} e^{\frac{1}{4}(1-\lambda)^2} \leqslant 2 - r.$$

Proof of Theorem 2.1. We use induction on the order of G to prove a slightly stronger result. We show that, when $\pi \in_U G$, for each non-empty $A \subseteq G$ and each $s \in S$,

$$\mathbf{P}(\pi \in A)\, \mathbf{E}\left[\exp \frac{1}{16} f(A, \pi, s)\right] \leqslant 1. \tag{2.4}$$

Note that this inequality is trivial if G is trivial (*i.e.*, if G contains only the identity permutation) since then $A = G$.

For a non-empty subset K of $\mathrm{Sym}(S)$, let $\mathrm{supp}(K)$ denote the set of points $s \in S$ such that $\sigma(s) \neq s$ for some $\sigma \in K$. Note that $\mathrm{supp}(G)$ is empty if and only if G is trivial.

To prove (2.4) for nontrivial G it suffices to consider only $s \in \mathrm{supp}(G)$. For if $s \in \mathrm{supp}(G)$ then, for each $s' \in S \setminus \mathrm{supp}(G)$ and each $\sigma \in G$,

$$f(A, \sigma, s') = f(A, \sigma) \leqslant f(A, \sigma, s).$$

Now let $|G| \geqslant 2$ and suppose that the result (2.4) holds for any product group H which is a proper subgroup of G, together with any non-empty subset of H. We may assume without loss of generality that S is a set of integers, that $s = 1$ and that $s \in \mathrm{supp}(G)$. We now recall the notation introduced in the sketch of the proof. Let H denote the stabilizer of the element 1. We may assume without loss of generality that the orbit of the element 1 is $\{1, \ldots, m\}$ for some $m \geqslant 2$. Then H is a product group of permutations on S, where the block $\{1, \ldots, m\}$ has been split into the two blocks $\{1\}$ and $\{2, \ldots, m\}$. Let $H_1 = H$, and for $i = 2, \ldots, m$ let H_i denote the coset Ht_{1i} of H. Thus G is partitioned into the m cosets H_i, where H_i consists of the permutations σ with $\sigma(i) = 1$, and each H_i has size $|G|/m$.

Let $A \subseteq G$ be non-empty, and let p denote $\mathbf{P}(\pi \in A)$ so that $p = |A|/|G|$. For $i = 1, \ldots, m$ let $A_i = A \cap H_i$ and let $q_i = |A_i|/|H|$. Thus

$$p = \frac{1}{m} \sum_{i=1}^{m} q_i.$$

Observe that if $\sigma \in H_i$ then the present notation A_i agrees with that in Lemma 2.4. Choose j such that q_j is a maximum, and keep j fixed. The main part of the proof of (2.4) will consist in establishing the following claim.

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

Claim. Let $i \in \{1, \ldots, m\}$ and let $\pi_i \in_U H_i$. Then

$$\mathbf{E}\left[\exp \frac{1}{16} f(A, \pi_i, i)\right] \leqslant q_j^{-1}(2 - q_i/q_j). \tag{2.5}$$

Proof of Claim. For notational convenience let t_{11} denote the identity element e. If A_i is nonempty, then by Lemma 2.3

$$\mathbf{E}\left[\exp \frac{1}{16} f(A_i, \pi_i, j)\right] = \mathbf{E}\left[\exp \frac{1}{16} f(A_i t_{1i}, \pi_i t_{1i}, t_{1i}(j))\right],$$

and hence

$$\mathbf{E}\left[\exp \frac{1}{16} f(A_i, \pi_i, j)\right] \leqslant q_i^{-1} \tag{2.6}$$

by the induction hypothesis, since $A_i t_{1i} \subseteq H$, $\pi_i t_{1i} \in_U H$, and $|A_i t_{1i}|/|H| = q_i$. By Lemma 2.3 again,

$$\mathbf{E}\left[\exp \frac{1}{16} f(A_j t_{ij}, \pi_i)\right] = \mathbf{E}\left[\exp \frac{1}{16} f(A_j t_{ij} t_{1i}, \pi_i t_{1i})\right],$$

and hence

$$\mathbf{E}\left[\exp \frac{1}{16} f(A_j t_{ij}, \pi_i)\right] \leqslant q_j^{-1} \tag{2.7}$$

by the induction hypothesis, since $A_j t_{ij} t_{1i} \subseteq H$, $\pi_i t_{1i} \in_U H$, and $|A_j t_{ij} t_{1i}|/|H| = q_j$.

We shall use the results (2.6) and (2.7) to complete the proof of the Claim. Suppose first that $i \neq j$. If A_i is non-empty, then by Lemmas 2.4 and 2.5, for each $0 \leqslant \lambda \leqslant 1$,

$$\begin{aligned}
&\mathbf{E}\left[\exp \frac{1}{16} f(A, \pi_i, i)\right] \\
&\leqslant e^{\frac{1}{4}(1-\lambda)^2} \mathbf{E}\left[\exp \frac{\lambda}{16} f(A_i, \pi_i, j) \quad \exp \frac{1-\lambda}{16} f(A_j t_{ij}, \pi_i)\right] \\
&\leqslant e^{\frac{1}{4}(1-\lambda)^2} \left(\mathbf{E}\left[\exp \frac{1}{16} f(A_i, \pi_i, j)\right]\right)^{\lambda} \left(\mathbf{E}\left[\exp \frac{1}{16} f(A_j t_{ij}, \pi_i)\right]\right)^{1-\lambda} \\
&\leqslant e^{\frac{1}{4}(1-\lambda)^2} q_i^{-\lambda} q_j^{-(1-\lambda)} \\
&= e^{\frac{1}{4}(1-\lambda)^2} q_j^{-1} (q_i/q_j)^{-\lambda}.
\end{aligned}$$

By minimizing over λ using Lemma 2.6, we obtain

$$\mathbf{E}\left[\exp \frac{1}{16} f(A, \pi_i, i)\right] \leqslant q_j^{-1}(2 - q_i/q_j).$$

If A_i is empty, then, since

$$f(A, \sigma, i) \leqslant 3 + f(A t_{ij}, \sigma) \leqslant 3 + f(A_j t_{ij}, \sigma),$$

it follows by (2.7) that

$$\mathbf{E}\left[\exp \frac{1}{16} f(A, \pi_i, i)\right] \leqslant e^{\frac{3}{16}} \mathbf{E}\left[\exp \frac{1}{16} f(A_j t_{ij}, \pi_i)\right] \leqslant e^{\frac{3}{16}} q_j^{-1} \leqslant q_j^{-1}(2 - q_i/q_j).$$

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

Finally consider the case $i = j$. Then, by (2.6),

$$\mathbf{E}\left[\exp\frac{1}{16}f(A,\pi_i,i)\right] \leqslant q_i^{-1} = q_j^{-1}(2 - q_i/q_j).$$

This completes the proof of the Claim. □

We now resume the main proof. Recall that $s = 1$. When $\pi \in_U G$,

$$\mathbf{E}\left[\exp\frac{1}{16}f(A,\pi,\pi^{-1}(s))\right] = \frac{1}{m}\sum_{i=1}^{m}\mathbf{E}\left[\exp\frac{1}{16}f(A,\pi_i,i)\right],$$

where $\pi_i \in_U H_i$. Hence, when $\pi \in_U G$, by the Claim,

$$\begin{aligned}\mathbf{E}\left[\exp\frac{1}{16}f(A,\pi,\pi^{-1}(s))\right] &\leqslant \frac{1}{m}\sum_{i=1}^{m} q_j^{-1}(2 - q_i/q_j)\\ &= q_j^{-1}\left(2 - p/q_j\right)\\ &= p^{-1}x(2-x),\end{aligned}$$

where $x = p/q_j$. But $x(2-x) \leqslant 1$, which shows that

$$p\,\mathbf{E}\left[\exp\frac{1}{16}f(A,\pi,\pi^{-1}(s))\right] \leqslant 1. \tag{2.8}$$

This last result holds for any non-empty set $A \subseteq G$. In fact we shall use it with A replaced by $A^{-1} = \{\sigma^{-1} : \sigma \in A\}$. When $\pi \in_U G$, by Lemma 2.3 the random variables $f(A,\pi,s)$ and $f(A^{-1},\pi^{-1},\pi(s))$ have the same distribution, and thus so also does $f(A^{-1},\pi,\pi^{-1}(s))$. Hence

$$p\,\mathbf{E}\left[\exp\frac{1}{16}f(A,\pi,s)\right] = (|A^{-1}|/|G|)\mathbf{E}\left[\exp\frac{1}{16}f(A^{-1},\pi,\pi^{-1}(s))\right] \leqslant 1,$$

by (2.8). This completes the induction step, and thus the proof. □

We now extend Theorem 2.1 above. We need first to extend the definitions of $f(A,\mathbf{x})$ and $d_T(A,\mathbf{x})$ given at the start of Section 2. For a nonsingular matrix M with rows and columns indexed by I, we let

$$d_M(A,\mathbf{x}) = \min\{\|M\mathbf{v}\| : \mathbf{v} \in V(A,\mathbf{x})\},$$

and let

$$f_M(A,\mathbf{x}) = d_M(A,\mathbf{x})^2.$$

It may be checked, much as with (2.2), that

$$d_M(A,\mathbf{x}) = \sup_{\alpha} d_{\alpha M}(A,\mathbf{x}), \tag{2.9}$$

where the sup is over all unit vectors α indexed by I.

Let $\mathbf{X} = (X_1,\ldots,X_n)$ be a family of random variables, where X_j takes values in the set Ω_j. Let $\Omega' = \prod_{j=1}^{n}\Omega_j$. Assume that, for each non-empty $A' \subseteq \Omega'$,

$$\mathbf{P}(\mathbf{X} \in A')\,\mathbf{E}\left[\exp\frac{1}{4}f(A',\mathbf{X})\right] \leqslant 1. \tag{2.10}$$

Theorem 4.1.1 of Talagrand [7] (or see Theorem 4.9 of [4]) states that the inequality (2.10)

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

holds if the X_j are independent. Let G be a product group of permutations on a set S as before, let $\Omega = \Omega' \times G$, and let $\pi \in_U G$. Assume that $\mathbf{X}$ and π are independent. Let $\mathbf{Y}$ denote $(\mathbf{X}, \pi)$, so that $\mathbf{Y}$ takes values in Ω. We may assume that $\{1, \ldots, n\}$ and S are disjoint, and let $I = \{1, \ldots, n\} \cup S$. Let $M = (m_{ij})$ be the diagonal matrix with rows and columns indexed by the set I, where $m_{ii} = 2$ for $i \in \{1, \ldots, n\}$ and $m_{ii} = 1$ for $i \in S$. Thus

$$\|M\mathbf{v}\|^2 = 4\sum_{i=1}^{n} v_i^2 + \sum_{i \in S} v_i^2.$$

For $j \in S$, let

$$f_M(A, \mathbf{y}, j) = \min\{v_j^2 + \|M\mathbf{v}\|^2 : \mathbf{v} \in V(A, \mathbf{y})\}.$$

(Thus, if Mj is the diagonal matrix obtained from M by resetting the jj-entry to $\sqrt{2}$, then $f_M(A, \mathbf{y}, j) = f_{Mj}(A, \mathbf{y})$, but we shall not use this notation.)

Theorem 2.7. *Let $A \subseteq \Omega$ be non-empty. Then*

$$\mathbf{P}(\mathbf{Y} \in A)\, \mathbf{E}\left[\exp \frac{1}{16} f_M(A, \mathbf{Y})\right] \leqslant 1. \tag{2.11}$$

Markov's inequality immediately yields the following.

Corollary 2.8. *For any $t \geqslant 0$,*

$$\mathbf{P}(\mathbf{Y} \in A)\mathbf{P}(d_M(A, \mathbf{Y}) \geqslant t) \leqslant e^{-\frac{1}{16}t^2}. \qquad \square$$

Proof of Theorem 2.7. Just as with Theorem 2.1, we use induction on the order of G to prove a slightly stronger result. We show that, for each $s \in S$,

$$\mathbf{P}(\mathbf{Y} \in A)\, \mathbf{E}\left[\exp \frac{1}{16} f_M(A, \mathbf{Y}, s)\right] \leqslant 1. \tag{2.12}$$

Note that Lemmas 2.3 and 2.4 each hold with f replaced by f_M: let us refer to these results as Lemmas 2.3_M and 2.4_M respectively.

Suppose first that G contains only the identity element e. For $A \subseteq \Omega$, let $A' = \{\mathbf{x} \in \Omega' : (\mathbf{x}, e) \in A\}$. Then the left-hand side of (2.12) equals

$$\mathbf{P}(\mathbf{X} \in A')\, \mathbf{E}\left[\exp \frac{1}{4} f(A', \mathbf{X})\right],$$

which is at most 1, by assumption (2.10) above.

Now let $|G| \geqslant 2$ and suppose that the result (2.12) holds for any product group H that is a proper subgroup of G, together with any non-empty subset of $\Omega' \times H$ and $s \in S$. We shall establish the induction step much as for Theorem 2.1, by replacing σ by $(\mathbf{x}, \sigma)$, *etc.* In order to be able to keep the proof as similar as possible to the earlier proof, let us change the notation here, and assume that $S = \{1, \ldots, |S|\}$ and $\mathbf{X}$ is the vector $(X_{|S|+1}, \ldots, X_{|S|+n})$. We may now, as before, assume without loss of generality that $s = 1$ and that the orbit of the element 1 is $\{1, \ldots, m\}$ for some $m \geqslant 2$. Let H denote the stabilizer of the element 1,

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

so that H is a product group of permutations on S. Let $H_1 = H$, and for $i = 2, \ldots, m$ let H_i denote the coset Ht_{1i} of H. Let $A \subseteq \Omega$ be such that

$$p = \mathbf{P}(\mathbf{Y} \in A) > 0.$$

For $i = 1, \ldots, m$ let $A_i = \{(\mathbf{x}, \sigma) \in A : \sigma \in H_i\}$, and let

$$q_i = \mathbf{P}((\mathbf{X}, \pi) \in A | \pi(i) = 1),$$

so that

$$p = \frac{1}{m} \sum_{i=1}^{m} q_i.$$

Now, if $\pi_i \in_U H_i$, $\pi_H \in_U H$ and $A_i t_{1i}$ denotes $\{(\mathbf{x}, \sigma t_{1i}) : (\mathbf{x}, \sigma) \in A_i\}$, then

$$q_i = \mathbf{P}((\mathbf{X}, \pi_i) \in A_i) = \mathbf{P}((\mathbf{X}, \pi_i t_{1i}) \in A_i t_{1i}) = \mathbf{P}((\mathbf{X}, \pi_H) \in A_i t_{1i}),$$

since $\pi_i t_{1i} \in_U H$. Choose j such that q_j is a maximum, and keep j fixed. The main part of the proof of (2.12) will consist in establishing the following claim.

Claim. Let $i \in \{1, \ldots, m\}$ and let $\pi_i \in_U H_i$. Then

$$\mathbf{E}\left[\exp \frac{1}{16} f_M(A, (\mathbf{X}, \pi_i), i)\right] \leqslant q_j^{-1}(2 - q_i/q_j). \tag{2.13}$$

Proof of Claim. If $q_i > 0$, then, by Lemma 2.3_M,

$$\mathbf{E}\left[\exp \frac{1}{16} f_M(A_i, (\mathbf{X}, \pi_i), j)\right] = \mathbf{E}\left[\exp \frac{1}{16} f_M(A_i t_{1i}, (\mathbf{X}, \pi_i t_{1i}), t_{1i}(j))\right];$$

and hence

$$\mathbf{E}\left[\exp \frac{1}{16} f_M(A_i, (\mathbf{X}, \pi_i), j)\right] \leqslant q_i^{-1}, \tag{2.14}$$

by the induction hypothesis since $\pi_i t_{1i} \in_U H$. Similarly, by Lemma 2.3_M again,

$$\mathbf{E}\left[\exp \frac{1}{16} f_M(A_j t_{ij}, (\mathbf{X}, \pi_i))\right] = \mathbf{E}\left[\exp \frac{1}{16} f_M(A_j t_{ij} t_{1i}, (\mathbf{X}, \pi_i t_{1i}))\right];$$

and hence

$$\mathbf{E}\left[\exp \frac{1}{16} f_M(A_j t_{ij}, (\mathbf{X}, \pi_i))\right] \leqslant q_j^{-1}, \tag{2.15}$$

by the induction hypothesis since $\pi_i t_{1i} \in_U H$. We shall use the above results (2.14) and (2.15) to complete the proof of the Claim.

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

Suppose first that $i \neq j$. If $q_i > 0$, then by Lemmas 2.4_M and 2.5, for each $0 \leqslant \lambda \leqslant 1$,

$$\begin{aligned}
&\mathbf{E}\left[\exp\frac{1}{16}f_M(A,(\mathbf{X},\pi_i),i)\right] \\
&\leqslant\ e^{\frac{1}{4}(1-\lambda)^2}\mathbf{E}\left[\exp\left(\frac{\lambda}{16}f_M(A_i,(\mathbf{X},\pi_i),j)\right)\ \exp\left(\frac{1-\lambda}{16}f_M(A_jt_{ij},(\mathbf{X},\pi_i),i,j)\right)\right] \\
&\leqslant\ e^{\frac{1}{4}(1-\lambda)^2}\left(\mathbf{E}\left[\exp\frac{1}{16}f_M(A_i,(\mathbf{X},\pi_i),j)\right]\right)^{\lambda}\left(\mathbf{E}\left[\exp\frac{1}{16}f_M(At_{ij},(\mathbf{X},\pi_i),i,j)\right]\right)^{1-\lambda} \\
&\leqslant\ e^{\frac{1}{4}(1-\lambda)^2}q_i^{-\lambda}q_j^{-(1-\lambda)} \\
&=\ e^{\frac{1}{4}(1-\lambda)^2}q_j^{-1}(q_i/q_j)^{-\lambda}.
\end{aligned}$$

By minimizing over λ using Lemma 2.6, we obtain

$$\mathbf{E}\left[\exp\frac{1}{16}f_M(A,(\mathbf{X},\pi_i),i)\right] \leqslant q_j^{-1}(2-q_i/q_j).$$

If $q_i = 0$, then since $f_M(A,\mathbf{y},i) \leqslant 3 + f_M(A_jt_{ij},\mathbf{y})$, it follows from (2.15) that

$$\mathbf{E}\left[\exp\frac{1}{16}f_M(A,(\mathbf{X},\pi_i),i)\right] \leqslant e^{\frac{3}{16}}q_j^{-1} \leqslant q_j^{-1}(2-q_i/q_j).$$

Finally, if $i = j$, then by (2.14),

$$\mathbf{E}\left[\exp\frac{1}{16}f_M(A,(\mathbf{X},\pi_i),i)\right] \leqslant \mathbf{E}\left[\exp\frac{1}{16}f_M(A_i,(\mathbf{X},\pi_i),i)\right] \leqslant q_i^{-1} = q_j^{-1}(2-q_i/q_j).$$

This completes the proof of the Claim. □

We now resume the main proof. Recall that $s = 1$. When $\pi \in_U G$,

$$\mathbf{E}\left[\exp\frac{1}{16}f_M(A,(\mathbf{X},\pi),\pi^{-1}(s))\right] = \frac{1}{m}\sum_{i=1}^{m}\mathbf{E}\left[\exp\frac{1}{16}f_M(A,(\mathbf{X},\pi_i),i)\right],$$

where $\pi_i \in_U H_i$. Hence, by the Claim,

$$\begin{aligned}
\mathbf{E}\left[\exp\frac{1}{16}f_M(A,(\mathbf{X},\pi),\pi^{-1}(s))\right] &\leqslant \frac{1}{m}\sum_{i=1}^{m}q_j^{-1}(2-q_i/q_j) \\
&= q_j^{-1}\left(2-p/q_j\right) \\
&= p^{-1}(p/q_j)(2-(p/q_j)) \\
&\leqslant p^{-1},
\end{aligned}$$

since $x(2-x) \leqslant 1$. We use this result with A replaced by $A^{-1} = \{(\mathbf{x},\sigma^{-1}) : (\mathbf{x},\sigma) \in A\}$. By Lemma 2.3_M, the random variables $f_M(A,(\mathbf{X},\pi),s)$ and $f_M(A^{-1},(\mathbf{X},\pi),\pi^{-1}(s))$ have the same distribution. Hence

$$p\,\mathbf{E}\left[\exp\frac{1}{16}f_M(A,\mathbf{Y},s)\right] = p\,\mathbf{E}\left[\exp\frac{1}{16}f_M(A^{-1},(\mathbf{X},\pi),\pi^{-1}(s))\right] \leqslant 1.$$

This completes the induction step, and thus the proof. □

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

Proof of Theorem 1.1. We use Corollary 2.8 and argue much as in [4]. Note that it makes no difference in Theorem 1.1 if we assume that the sets B_k are pairwise disjoint, so let us do so. Observe first that if two permutations σ and τ disagree in at most k coordinates, then there is a product of at most $k-1$ transpositions taking σ to τ. Let $\omega = (\mathbf{x}, \sigma) \in \Omega$, and suppose that $h(\omega) = s$. Let J be a set of at most rs coordinates such that if $\omega' \in \Omega$ satisfies $\omega'_j = \omega_j$ for each $j \in J$ then $h(\omega') \geqslant s$. Let β be the indicator vector $\mathbf{1}_J$, and let α be the unit vector $|J|^{-1/2}\beta$. Then, for each $\omega' \in \Omega$,

$$h(\omega') \geqslant h(\omega) - cd_{\beta M}(\omega, \omega') = h(\omega) - c|J|^{\frac{1}{2}} d_{\alpha M}(\omega, \omega'),$$

where M is the matrix in Theorem 2.7, and so

$$h(\omega') \geqslant h(\omega) - \sqrt{\gamma h(\omega)} d_{\alpha M}(\omega, \omega'),$$

where $\gamma = rc^2$. (Thus h is like a 'γ-configuration function' in the language of [4].)

Let $a \geqslant 0$, let $A_a = \{\omega' \in \Omega : h(\omega') \leqslant a\}$, and suppose that A_a is non-empty. Then

$$h(\omega) \leqslant a + \sqrt{\gamma h(\omega)} d_{\alpha M}(\omega, \omega')$$

for each $\omega' \in A_a$, and so by minimizing over such ω' we have

$$h(\omega) \leqslant a + \sqrt{\gamma h(\omega)} d_{\alpha M}(A_a, \omega) \leqslant a + \sqrt{\gamma h(\omega)} d_M(A_a, \omega)$$

by (2.9). Thus, if $t > 0$, $\omega \in \Omega$ and $h(\omega) \geqslant a + t$, then

$$d_M(A_a, \omega) \geqslant \frac{h(\omega) - a}{\sqrt{\gamma h(\omega)}} \geqslant \frac{t}{\sqrt{\gamma}\sqrt{a+t}},$$

since the function $g(x) = (x-a)/\sqrt{x}$ is increasing for $x \geqslant a$. Thus, for each $t > 0$,

$$\mathbf{P}(h(\mathbf{Y}) \geqslant a + t) \leqslant \mathbf{P}\left(d_M(A_a, \mathbf{Y}) \geqslant \frac{t}{\sqrt{\gamma(a+t)}} \right).$$

Hence, by Corollary 2.8, for each $t > 0$,

$$\begin{aligned} &\mathbf{P}(h(\mathbf{Y}) \leqslant a)\mathbf{P}(h(\mathbf{Y}) \geqslant a + t) \\ &\leqslant \mathbf{P}(\mathbf{Y} \in A_a)\mathbf{P}\left(d_M(A_a, \mathbf{Y}) \geqslant \frac{t}{\sqrt{\gamma(a+t)}} \right) \\ &\leqslant \exp\left(-\frac{t^2}{16\gamma(a+t)} \right). \end{aligned}$$

Now we may complete the proof by appropriate choices of a in this last inequality. If we let $a = m$, then, since $\mathbf{P}(h(\mathbf{Y}) \leqslant m) \geqslant \frac{1}{2}$, we obtain (1.1); and if we let $a = m - t$ then, since $\mathbf{P}(h(\mathbf{Y}) \geqslant m) \geqslant \frac{1}{2}$, we obtain (1.2). □

3. An application to graph colouring

In a series of papers and a book (see [5]), Molloy and Reed prove many deep results on graph colouring by probabilistic methods. One of the tools they use is Theorem 1.1 above, and indeed such a result was 'commissioned' by them for use in the book [5]. In

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

particular, they use Theorem 1.1 to prove upper bounds on the chromatic number $\chi(G)$ of a graph G in terms of the clique number $\omega(G)$ and the maximum degree $\Delta(G)$.

By Brooks' theorem, (a) $\chi(G) \leqslant \Delta(G) + 1$; and (b) for any graph G with $\Delta(G) \geqslant 3$, if $\omega(G) \leqslant \Delta(G)$ then $\chi(G) \leqslant \Delta(G)$. The following theorem from [5] says roughly that if $\omega(G)$ is much less than $\Delta(G)$ then so is $\chi(G)$.

Theorem 3.1. *For any $\varepsilon > 0$, there exist Δ_0 and $\delta > 0$ such that, for any graph G with $\Delta(G) \geqslant \Delta_0$, if $\omega(G) \leqslant (1-\varepsilon)\Delta(G)$ then $\chi(G) \leqslant (1-\delta)\Delta(G)$.*

To prove this theorem, Molloy and Reed consider a colouring procedure along the following lines. First the nodes of G are partitioned into a number of 'dense' sets $D_1, D_2, \ldots$ and a 'sparse' set S. For a suitable choice of $\delta > 0$, each dense set D_i has a colouring using colours $1, 2, \ldots, c$, where $c = (1-\delta)\Delta$ and we write Δ for $\Delta(G)$. We make a first attempt to colour G as follows.

(1) For each 'sparse' node $v \in S$, independently assign a colour $X_v \in_U \{1, \ldots, c\}$.
(2) For each dense set D_i, independently pick $\pi_i \in_U \mathrm{Sym}(\{1, \ldots, c\})$, and use π_i to permute the given colours on D_i; that is, if node $v \in D_i$ originally has colour j we now assign it colour $\pi_i(j)$. (The idea is to make the colouring of D_i 'look random' to the rest of the graph.)
(3) For each 'sparse' node $v \in S$, if any neighbour of v is assigned the same colour as v, we uncolour v.
(4) For each dense set D_i and each colour γ, if any node in D_i coloured γ is adjacent to another node coloured γ, then each node in D_i coloured γ is uncoloured.

If a node is not uncoloured in step (3) or step (4) above we say that it *retains* its colour. We want to show that, with positive probability, the partial colouring formed by the retained colours has certain properties that allow us to extend it to a c-colouring of all of G. The property we focus on here is as follows.

For each node $v \in S$ let Y_v be the number of neighbours of v that retain their colour less the number of distinct colours retained on the neighbours of v. Let $y = \delta\Delta + 1$. If $Y_v \geqslant y$ for each $v \in S$ then we may greedily extend the partial colouring to all of S. For, when we come to colour a node in S, the number of forbidden colours must be at most $\Delta - y = c - 1$. A similar approach works for the dense sets D_i, but we do not consider that here.

We then want to show that, with positive probability, we have $Y_v \geqslant y$ for each $v \in S$. If we can prove that, for each $v \in S$, the probability that $Y_v < y$ is very small, then we can apply the Lovász Local Lemma to complete the proof. In fact, it is sufficient to prove that $\mathbf{P}(Y_v < y)$ is $o(\Delta^{-5})$. Our aim from now on is to establish the following claim.

Claim. $\mathbf{P}(Y_v \leqslant y) = e^{-\Omega(\Delta)}$.

Proof of Claim. Let Z_v be the number of colours γ which are assigned to exactly two neighbours of v and are retained by both. Observe that $Y_v \geqslant Z_v$. It may be shown that for a suitable choice of $\delta > 0$ we have $E(Z_v) \geqslant 2y$. At last we come to the concentration part: we show that Z_v is very unlikely to be far below its mean.

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

Let Z_v' be the number of colours assigned to at least two neighbours of v, and let Z_v'' be the number of colours either assigned to at least three neighbours of v, or assigned to at least two but retained by at most one. Then $Z_v = Z_v' - Z_v''$. We use Theorem 1.1 to show that both Z_v' and Z_v'' are concentrated.

First consider Z_v'. Changing some value x_v or swapping two elements in some π_i affects only two colours and so can change Z_v' by at most 2: hence we may take $c = 2$. (In fact the reader may easily check that Z_v' can change by only 1, and so we could take $c = 1$.) If a node is assigned colour γ, then to certify this we need reveal only one coordinate. So if $Z_v' \geqslant s$, then to certify this we need reveal only $2s$ coordinates: for each of s colours γ, we reveal the coordinates corresponding to the colour γ on two neighbours of v. Thus we may take $r = 2$. Also $Z_v' \leqslant \Delta/2$ and so its median $m_v' \leqslant \Delta/2$. Hence, by (1.3),

$$\mathbf{P}\big(|Z_v' - m_v'| \geqslant y/6\big) \leqslant 4\exp\left(-\frac{(y/6)^2}{2^8 m_v'}\right) = e^{-\Omega(\Delta)}.$$

We may handle Z_v'' similarly. Changing a value x_v or swapping two elements in some π_i affects only two colours, and so we may take $c = 2$. If $Z_v'' \geqslant s$, then to certify this we need reveal only $3s$ coordinates: for each of s colours γ, we exhibit coordinates corresponding to either three neighbours of v coloured γ, or two adjacent neighbours of v coloured γ, or two non-adjacent neighbours of v coloured γ together with one further node coloured γ, which causes at least one of them to lose its colour. Thus we may take $r = 3$. Hence, if m_v'' denotes a median of Z_v'', then much as before we obtain

$$\mathbf{P}\big(|Z_v'' - m_v''| \geqslant y/6\big) = e^{-\Omega(\Delta)}.$$

Now let $\mu = m_v' - m_v''$. Then, by the last two inequalities,

$$\begin{aligned}\mathbf{P}\big(|Z_v - \mu| \geqslant y/3\big) &\leqslant \mathbf{P}\big(|Z_v' - m_v'| \geqslant y/6\big) + \mathbf{P}\big(|Z_v'' - m_v''| \geqslant y/6\big)\\ &= e^{-\Omega(\Delta)}.\end{aligned}$$

It follows that if Δ is sufficiently large then $\mu \geqslant (4/3)y$, for otherwise

$$2y \leqslant E(Z_v) \leqslant (5/3)y + \Delta\,\mathbf{P}\big(Z_v - \mu \geqslant y/3\big) < 2y.$$

Then

$$\begin{aligned}\mathbf{P}(Y_v \leqslant y) &\leqslant \mathbf{P}(Z_v \leqslant y)\\ &\leqslant \mathbf{P}\big(|Z_v - \mu| \geqslant y/3\big)\\ &= e^{-\Omega(\Delta)},\end{aligned}$$

and we have established the claim. □

4. Bad sets

The first two conditions on the function h in Theorem 1.1 imply that, for each $(\mathbf{x}, \sigma)$ and $(\mathbf{x}', \sigma')$ in Ω,

$$h(\mathbf{x}, \sigma) - h(\mathbf{x}', \sigma') \leqslant 2c\, d(\mathbf{x}, \mathbf{x}') + c\, d(\sigma, \sigma'),$$

and indeed the proof of Theorem 1.1 uses exactly this property. (Here d denotes Hamming distance, and as usual the permutation σ on S is specified by the vector $(\sigma(i) : i \in S)$.) Of course, by symmetry we could replace the left-hand side above by its absolute value.

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

We now consider weakened versions of these conditions. Let $B \subseteq \Omega$ be a 'bad' or 'failure' set on which conditions may fail. Let c and r be positive constants, and suppose that the nonnegative real-valued function h on Ω satisfies the following two conditions for each $(\mathbf{x}, \sigma) \in \Omega \setminus B$.

- For each $(\mathbf{x}', \sigma')$ in Ω,

$$h(\mathbf{x}, \sigma) - h(\mathbf{x}', \sigma') \leqslant 2c\, d(\mathbf{x}, \mathbf{x}') + c\, d(\sigma, \sigma').$$

- If $h(\mathbf{x}, \sigma) = s$, then in order to show that $h(\mathbf{x}, \sigma) \geqslant s$ we need specify only at most rs coordinates of $(\mathbf{x}, \sigma)$.

The first condition above ensures that, as we move away from any vector $(\mathbf{x}, \sigma)$ not in the 'bad set' B, the value of the function h cannot decrease too suddenly.

Suppose that the random variable $\mathbf{Y}$ is as in Theorem 1.1, and assume that the bad set B satisfies $P(\mathbf{Y} \in B) < \frac{1}{2}$. Let m be a median of the random variable $Z = h(\mathbf{Y})$.

Theorem 4.1. *For each $t \geqslant 0$,*

$$P(Z \geqslant m + t) \leqslant 2 \exp\left(-\frac{t^2}{16rc^2(m+t)}\right) + P(B), \tag{4.1}$$

and

$$P(Z \leqslant m - t) \leqslant \left(\tfrac{1}{2} - P(B)\right)^{-1} \exp\left(-\frac{t^2}{16rc^2 m}\right) \tag{4.2}$$

$$\leqslant 4 \exp\left(-\frac{t^2}{16rc^2 m}\right) \tag{4.3}$$

if $P(B) \leqslant \frac{1}{4}$.

The latter inequalities here are perhaps at first sight surprising: even when the bad set B has probability as large as $\frac{1}{4}$, the probability of a large deviation below the median is small. In order to illustrate this behaviour with bad sets, let us return to the basic case without random permutations.

Example. Let $\Omega = \{0,1\}^n$ and let $\mathbf{X} = (X_1, \ldots, X_n)$ be uniformly distributed over Ω. Let γ be a large number, say $100n$. For $\mathbf{x} = (x_1, \ldots, x_n) \in \Omega$, let $h(\mathbf{x}) = \gamma$ if $x_1 = x_2 = 1$, and $h(\mathbf{x}) = \sum_i x_i$ otherwise. Observe that the median m of $h(\mathbf{X})$ is about $n/2$ and the mean is $n/2 + \frac{1}{4}(\gamma - 1 - n/2)$.

Let $B = \{\mathbf{x} \in \Omega : x_1 = x_2 = 1\}$ so $P(\mathbf{X} \in B) = \frac{1}{4}$. Let $\mathbf{x} \in \Omega \setminus B$. Then for all $\mathbf{x}' \in \Omega$ we have $h(\mathbf{x}) - h(\mathbf{x}') \leqslant d(\mathbf{x}, \mathbf{x}')$. Also, if $h(\mathbf{x}) = s$, then in order to show that $h(\mathbf{x}) \geqslant s$ we need to specify at most s coordinates. Thus Theorem 4.1 (without the random permutations) applies to this situation. Hence, by this result (or directly from bounds on the tails of the binomial distribution), large deviations below the median are unlikely: we find from (4.3) above that, for $t > 0$,

$$P(h(\mathbf{X}) \leqslant m - t) \leqslant 4 \exp\left(-\frac{t^2}{4m}\right).$$

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press

But for deviations above the median the picture is quite different: if $0 \leqslant t \leqslant \gamma - m$ then $P(h(\mathbf{X}) \geqslant m + t) \geqslant \frac{1}{4}$.

Proof of Theorem 4.1. Just as in the proof of Theorem 1.1, we find that if $t > 0$, $\omega \in \Omega \setminus B$, and $h(\omega) \geqslant a + t$, then

$$d_M(A_a, \omega) \geqslant \frac{h(\omega) - a}{\sqrt{\gamma h(\omega)}} \geqslant \frac{t}{\sqrt{\gamma(a+t)}}.$$

Thus, for each $t > 0$,

$$\mathbf{P}(h(\mathbf{Y}) \geqslant a + t \text{ and } (\Omega \setminus B)) \leqslant \mathbf{P}\left(d_M(A_a, \mathbf{Y}) \geqslant \frac{t}{\sqrt{\gamma(a+t)}}\right).$$

Hence, by Corollary 2.8, for each $t > 0$,

$$\begin{aligned}
&\mathbf{P}(h(\mathbf{Y}) \leqslant a)\mathbf{P}(h(\mathbf{Y}) \geqslant a + t) \\
&\leqslant \mathbf{P}(\mathbf{Y} \in A_a)\left(\mathbf{P}\left(d_M(A_a, \mathbf{Y}) \geqslant \frac{t}{\sqrt{\gamma(a+t)}}\right) + \mathbf{P}(B)\right) \\
&\leqslant \exp\left(-\frac{t^2}{16\gamma(a+t)}\right) + \mathbf{P}(h(\mathbf{Y}) \leqslant a)\mathbf{P}(B).
\end{aligned}$$

Hence

$$\mathbf{P}(h(\mathbf{Y}) \leqslant a)(\mathbf{P}(h(\mathbf{Y}) \geqslant a + t) - \mathbf{P}(B)) \leqslant \exp\left(-\frac{t^2}{16\gamma(a+t)}\right).$$

Now we may complete the proof by appropriate choices of a in this last inequality. If we let $a = m$ we obtain (4.1) and if we let $a = m - t$ we obtain (4.2). □

Extensions of Theorem 2.1 above to more general 'locally acting' groups of permutations are considered in [3].

Acknowledgements

I would like to thank Malwina Luczak, Mike Molloy, Bruce Reed and the referee for helpful comments.

References

[1] Durrett, R. (1996) *Probability: Theory and Examples*, 2nd edn, Duxbury Press.
[2] Johnson, W. and Schechtman, G. (1991) Remarks on Talagrand's deviation inequality for Rademacher's functions. In Vol. 1470 of *Lecture Notes in Mathematics*, Springer, 72–77.
[3] Luczak, M. and McDiarmid, C. (2001) Concentration for locally acting permutations. Manuscript.
[4] McDiarmid, C. (1998) Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics* (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, eds), Springer, 195–248.
[5] Molloy, M. and Reed, B. A. (2001) *Graph Colouring and the Probabilistic Method*, Springer.
[6] Steele, J. M. (1997) *Probability Theory and Combinatorial Optimization*, SIAM CBMS **69**.
[7] Talagrand, M. (1995) Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Institut des Hautes Études Scientifiques* **81** 73–205.
[8] Talagrand, M. (1996) A new look at independence. *Ann. Probab.* **24** 1–31.

https://doi.org/10.1017/S0963548301005089 Published online by Cambridge University Press