PREPRINT PDMI — 21/2003

On Some Properties of Min-Wise Independent Families and Groups

Victor Bargachev

St.Petersburg State University,
Department of Mathematics and Mechanics,
Universitetski prospekt, 28, St.Peterhof,
St.Petersburg, 198504, Russia

December 30, 2003

Abstract

Min-wise independent families of permutations are studied. The connection of min-wise indepenence and independence on l-th minimum is obtained in the second section. In the third section min-wise indepenent families, which are groups also, are considered.

1 Introduction

The notion of min-wise independent permutations was introduced and studied in [1], [2] and [3]. It was motivated by the purpose of efficient filtering "almost identical" documents among huge collection of Web documents.

Similar concepts in a slightly different context also appeared in other papers. For references, see [3]. Later min-wise independent permutations were applied to derandomization of algorithms for NP-complete problems ([4], [5]).

Let us recall main definitions from [2]. Throughout this paper we use the following notation. Let $[n]$ denote a set $\{1, \ldots, n\}$. Let S_n denote a symmetric group of $[n]$. For $\pi \in S_n$ and $X, Y \subseteq [n]$ we write $\pi(X) < \pi(Y)$ if $\pi(x) < \pi(y)$ for any $x \in X$ and any $y \in Y$. For $X \subseteq [n]$ we define the rank of element $x \in X$ as follows: $\operatorname{rank}(x, X) := |\{x' \in X : \ x' \leqslant x\}|$. For an event A, let $\operatorname{ind}(A)$ be its indicator, i.e., $\operatorname{ind}(A) = 1$ if A is true and $\operatorname{ind}(A) = 0$ otherwise.

Definition 1.1 ([2], [3]). *A family of permutations $\mathcal{F} \subseteq S_n$ is called k*-restricted min-wise independent *(abbreviated k-MWI) if for every non-empty $X \subseteq [n]$ with $|X| \le k$ and for any $x \in X$, when a permutation π is chosen at random from $\mathcal{F}$ with some specified probability distribution, we have*

$$\Pr_{\mathcal{F}}(\min(\pi(X)) = \pi(x)) = \frac{1}{|X|}. \tag{1}$$

If the distribution on $\mathcal{F}$ is uniform, the family is called unbiased*, otherwise it is called* biased.

Clearly, the condition $\min(\pi(X)) = \pi(x)$ can be rewritten as

$$\operatorname{rank}(\pi(x), \pi(X)) = 1.$$

A natural problem is to study families $\mathcal{F}$ that satisfy

$$\operatorname{rank}(\pi(x), \pi(X)) = l$$

for some l. We study the connections between this general and the original notion of min-wise independence in more detail in section 2. The case of k-MWI families, which in addition are groups, attracted a special interest, e.g., see [6], [7], [8]. In section 3 we study further properties of min-wise independent groups and improve results from [7].

2 Min-wise independence and the l-th minimum

The following lemma appeared in [9] and independently in [10]:

Lemma 2.1. *Let $\mathcal{F} \subseteq S_n$ be a k-MWI family. Let $X, Y \subseteq [n]$, $X \cap Y = \emptyset$, $|X| + |Y| \leqslant k - 1$ and $z \in [n] \setminus (X \cup Y)$. Then*

$$\Pr_{\mathcal{F}}(\pi(X) < \pi(z) < \pi(Y)) = \frac{|X|! \cdot |Y|!}{(|X| + |Y| + 1)!}.$$

An easy consequence of this lemma is that min-wise independence is equivalent to max-wise independence. Actually, the more general result is true:

Lemma 2.2. *Let $\mathcal{F} \subseteq S_n$ be a k-MWI family, $X \subseteq [n]$, $|X| \leqslant k$, $x \in X$ and $l \in [|X|]$. Then*

$$\Pr_{\mathcal{F}}(\operatorname{rank}(\pi(x), \pi(X)) = l) = \frac{1}{|X|}. \tag{2}$$

Proof. Lemma 2.1 implies immediately

$$\begin{aligned}
&\Pr_{\mathcal{F}}(\operatorname{rank}(\pi(x), \pi(X)) = l) \\
&\qquad = \sum_{Y \subseteq X \setminus \{x\}, |Y| = l-1} \Pr_{\mathcal{F}}(\pi(Y) < \pi(x) < \pi((X \setminus Y) \setminus \{x\})) \\
&\qquad = \sum_{Y \subseteq X \setminus \{x\}, |Y| = l-1} \frac{(l-1)!(|X|-l)!}{|X|!} = \frac{1}{|X|}.
\end{aligned}$$

□

In particular, taking $l = |X|$ we get that min-wise independence implies max-wise independence. The special case of this lemma for $k = n$ was discovered in [3].

Now we show that lemma 2.2 has the following interesting converse:

Theorem 2.1. *Let $\mathcal{F} \subseteq S_n$ and $k \in [n]$. Suppose that for every non-empty $X \subseteq [n]$ with $|X| \leqslant k$ and for every $x \in X$, there exists $l = l(x, X) \leqslant |X|$ such that (2) is true. Then $\mathcal{F}$ is a k-MWI family.*

We will proof this theorem by induction on k. The case when $k = 1$ is trivial since any family is 1-MWI. The inductive step is given by the following lemma.

Lemma 2.3. *Let $k > 1$ and $\mathcal{F} \subseteq S_n$ be a $(k-1)$-MWI family. Assume that for every $X \subseteq [n]$ with $|X| = k$ and for every $x \in X$ there exists $l = l(x, X) \leqslant k$ such that (2) is true. Then $\mathcal{F}$ is a k-MWI family.*

Proof. First of all, we fix arbitrary $X \subseteq [n]$ and $x \in X$. For every $i \in [k]$, define a random variable $M_i : \mathcal{F} \to \{0, 1\}$ as follows

$$M_i(\pi) \ := \ \operatorname{ind}(\operatorname{rank}(\pi(x), \pi(X)) = i).$$

Let p_i denote the expectation $\mathrm{E}(M_i)$. We must show that $p_1 = \frac{1}{k}$. By the assumption of the lemma $p_l = \frac{1}{k}$, where $l = l(x, X)$. For every $t \in [k]$ we define another random variable $S_t : \mathcal{F} \to \mathbb{R}$:

$$S_t(\pi) \ := \sum_{\substack{Y \subseteq X \setminus \{x\}, \\ |Y| = t-1}} \operatorname{ind}(\pi(x) = \min \pi(X \setminus Y)).$$

Fix $\pi \in \mathcal{F}$ and choose $j \in [k]$ such that $j = \operatorname{rank}(\pi(x), \pi(X))$, i.e.

$$M_j(\pi) = 1 \text{ and } M_i(\pi) = 0 \text{ for } i \neq j. \tag{3}$$

Let $Z = \{z \in X : \ \pi(z) < \pi(x)\}$, in particular $|Z| = j - 1$. For a given $Y \subseteq X \setminus \{x\}$, one can show that $\pi(x) = \min(\pi(X \setminus Y))$ if and only if $Z \subseteq Y$. It implies that for every t

$$S_t(\pi) = \sum_{\substack{Y \subseteq X \setminus \{x\}, \\ |Y| = t-1}} \operatorname{ind}(Z \subseteq Y) = \binom{k-j}{t-j}, \tag{4}$$

where $\binom{k-j}{t-j} = 0$ when $t < j$. Using (3) we can rewrite equation (4) in the following form:

$$S_t(\pi) = \sum_{i=1}^{k} \binom{k-i}{t-i} M_i(\pi).$$

Hence

$$\mathrm{E}(S_t) = \sum_{i=1}^{k} \binom{k-i}{t-i} p_i. \tag{5}$$

On the other hand, when $t > 1$, by the definition of S_t we have

$$\begin{aligned}\mathrm{E}(S_t) = &\sum_{\substack{Y \subseteq X \setminus \{x\},\\ |Y|=t-1}} \mathrm{E}(\mathrm{ind}(\pi(x) = \min \pi(X \setminus Y)))\\ &= \sum_{\substack{Y \subseteq X \setminus \{x\},\\ |Y|=t-1}} \Pr(\pi(x) = \min \pi(X \setminus Y)) = \binom{k-1}{t-1} \frac{1}{k-t+1} \end{aligned} \tag{6}$$

since $|X \setminus Y| = k - t + 1 \leqslant k - 1$ and $\mathcal{F}$ is a $(k-1)$-MWI family. Combining (5) and (6) we obtain the following linear system of $(k-1)$ equations in k (real) unknowns

$$\sum_{i=1}^{k} \binom{k-i}{t-i} p_i = \binom{k-1}{t-1} \frac{1}{k-t+1}, \quad t = 2, \dots, k \tag{7}$$

One can verify that taking $p_i = \frac{1}{k}$ we get a solution of system (7) and taking

$$p_i = (-1)^{i+1} \binom{k-1}{i-1}$$

we get a solution of the homogeneous system corresponding to (7).

Deleting the first column from the matrix of coefficients of system (7) we obtain a lower triangular matrix with 1's on its main diagonal. Therefore, the space of solutions is one-dimensional, i.e. the most general solution has the form $p_i = \frac{1}{k} + c(-1)^{i+1}\binom{k-1}{i-1}$ for some $c \in \mathbb{R}$. Recall that by assumptions of the lemma, $p_l = \frac{1}{k}$. Hence, $c = 0$ and in particular $p_1 = \frac{1}{k}$. This completes the proof of lemma 2.3 and theorem 2.1. □

3 Min-wise independent groups

The case of k-MWI families, which in addition are groups, attracted a special interest, e.g., see [6], [7], [8]. In this section we study further properties of min-wise independent groups and improve results from [7]. In particular, we establish connections between unbiased and biased k-MWI groups.

Given a family $\mathcal{F} \subseteq S_n$ and $\sigma \in S_n$ we define

$$\mathcal{F}\sigma = \{\pi \in S_n : \exists \tau \in \mathcal{F} \ \pi = \tau\sigma\}.$$

For a given probability distribution $\Pr_{\mathcal{F}}$, let us define the distribution on $\mathcal{F}\sigma$ as follows

$$\Pr_{\mathcal{F}\sigma}(\pi) = \Pr_{\mathcal{F}}(\pi\sigma^{-1}). \tag{8}$$

The proofs of the next two lemmas are straightforward and follow immediately from the definition of k-MWI families.

Lemma 3.1. *Let $\mathcal{F} \subseteq S_n$ be a k-MWI family with respect to the distribution $\Pr_{\mathcal{F}}$. Then for every $\sigma \in S_n$, $\mathcal{F}\sigma$ is a k-MWI family with respect to the induced distribution (8).*

Lemma 3.2. *Let $\mathcal{F} \subseteq S_n$ and $\Pr_i$, $i = 1, \ldots, l$, be probability distributions on $\mathcal{F}$. If $\mathcal{F}$ is a k-MWI family with respect to each $\Pr_i$, $i = 1, \ldots, l$, then $\mathcal{F}$ is also a k-MWI family with respect to any convex combination of distributions $\Pr_1, \ldots, \Pr_l$. In particular, $\mathcal{F}$ is a k-MWI family with respect to $\Pr^*$, where*

$$\Pr^*(\pi) = \frac{1}{l}\sum_{i=1}^{l} \Pr_i(\pi). \tag{9}$$

The next theorem states that every biased k-MWI group is also unbiased.

Theorem 3.1. *Let $\mathcal{F} \subseteq S_n$ be a biased k-MWI group with respect to some distribution $\Pr$. Then $\mathcal{F}$ is also an unbiased k-MWI group.*

Proof. Let $\mathcal{F} = \{\tau_1, \ldots, \tau_m\}$. Since $\mathcal{F}$ is a group, for each $i = 1, \ldots, m$ we can define the following distribution on $\mathcal{F}$:

$$\Pr_i(\pi) = \Pr(\pi\tau_i^{-1}).$$

By lemma 3.1, $\mathcal{F}$ is a k-MWI group with respect to every $\Pr_i$. By lemma 3.2, $\mathcal{F}$ is a k-MWI group with respect to $\Pr^*$ defined in (9). But for every $\pi \in \mathcal{F}$

$$\Pr^*(\pi) = \frac{1}{m}\sum_{i=1}^{m} \Pr_i(\pi) = \frac{1}{m}\sum_{i=1}^{m} \Pr(\pi\tau_i^{-1}) = \frac{1}{m}\sum_{\tau\in\mathcal{F}} \Pr(\tau) = \frac{1}{m},$$

i.e. the distribution $\Pr^*$ is uniform. □

Corollary 3.1. *Let $\mathcal{F} \subseteq S_n$ be a biased k-MWI family and $\mathcal{G} \subseteq S_n$ be a group containing $\mathcal{F}$. Then $\mathcal{G}$ is an unbiased k-MWI group.*

Proof. Note that we may regard $\mathcal{G}$ as a biased k-MWI group with respect to the distribution $\Pr_{\mathcal{G}}$, where $\Pr_{\mathcal{G}}(\pi) = \Pr_{\mathcal{F}}(\pi)$ if $\pi \in \mathcal{F}$ and $\Pr_{\mathcal{G}}(\pi) = 0$ if $\pi \in \mathcal{G} \setminus \mathcal{F}$. Now the claim follows from theorem 3.1. □

In [7] some relations between min-wise independence and transitivity was studied. In particular, the following theorem was proved:

Theorem 3.2 ([7]). *Let $\mathcal{F} \subseteq S_n$ be an unbiased k-MWI group. If either $n \geqslant 6$ and $k \geqslant \lfloor \frac{n}{2} \rfloor$ or $n \neq 3$ and $k > \frac{n}{2}$ then $\mathcal{F}$ is transitive.*

For biased k-MWI groups a significantly weaker result was known. Namely in [7], an analog of theorem (3.2) was proved under assumption $k \geqslant n - \sqrt{n} + 1$. Now we can improve this result.

Theorem 3.3. *Let $\mathcal{F} \subseteq S_n$ be a biased k-MWI group. If either $n \geqslant 6$ and $k \geqslant \lfloor \frac{n}{2} \rfloor$ or $n \neq 3$ and $k > \frac{n}{2}$ then $\mathcal{F}$ is transitive.*

Indeed, if $\mathcal{F}$ is a biased k-MWI group, then it is also unbiased by theorem 3.1 and hence theorem 3.2 can be applied. We show that a similar result holds for families of permutations.

Definition 3.1. *Let $\mathcal{F} \subseteq S_n$ be a family of permutations. A set $X \subseteq [n]$ is called $\mathcal{F}$-invariant if $\pi(X) = X$ for every $\pi \in \mathcal{F}$.*

There are always at least two invariant sets, namely $\emptyset$ and $[n]$, which are called trivial. Note, that if $\mathcal{F}$ is group then $\mathcal{F}$ is transitive if and only if there is no non-trivial $\mathcal{F}$-invariant set.

Theorem 3.4. *Let $\mathcal{F} \subseteq S_n$ be a biased k-MWI family. If either $n \geqslant 6$ and $k \geqslant \lfloor \frac{n}{2} \rfloor$ or $n \neq 3$ and $k > \frac{n}{2}$ then there is no non-trivial $\mathcal{F}$-invariant set.*

Proof. Let $X \subseteq [n]$ be a $\mathcal{F}$-invariant set. Define a group $\mathcal{G} = \{\pi \in S_n :\ \pi(X) = X\}$. Clearly, $\mathcal{F} \subseteq \mathcal{G}$ and hence $\mathcal{G}$ is an unbiased k-MWI group by corollary 3.1. By theorem 3.2 the group $\mathcal{G}$ is transitive on $[n]$. This implies that either $X = [n]$ or $X = \emptyset$ since X is a $\mathcal{G}$-invariant set. $\square$

References

[1] A. Z. Broder, *On the Resemblance and Containment of Documents.* Proc. of Compression and Complexity of Sequences (1998), 21–29.

[2] A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher, *Min-Wise Independent Permutations.* Proc. of the 30th Annual ACM Symposium on Theory of Computing (1998), 327–336.

[3] A. Z. Broder, M. Mitzenmacher, *Completeness and robustness properties of min-wise independent permutations*, Random Structures and Algorithms, (2001) **18**, 18–30.

[4] A. Z. Broder, M. Charikar, M. Mitzenmacher, *A Derandomization Using Min-Wise Independent Permutations.* Proc. of RANDOM'98, Lecture Notes in Computer Science 1518, Springer (1998), pp.15-24.

[5] M. A. Vsemirnov, E. A. Hirsch, E. Ya. Dantsin, S. Ivanov, *Algorithms for SAT and Upper Bounds on Their Complexity.* J. Math. Sci., 2003, vol.118, issue 2, pp.4948–4962.

[6] M. Vsemirnov. *Automorphisms of projective spaces and min-wise independent sets of permutations.* Quaderni del Seminario Mathematico di Brescia, 2002, N53, pp. 1–16. Submitted to SIAM J. Discrete Math.

[7] C. Franchi, M. Vsemirnov, Min-wise independent groups. Quaderni del Seminario Mathematico di Brescia, 2002, N58, pp. 1–23. Submitted to European J. Combinatorics.

[8] C. Franchi. *Multiple transitivity and min-wise indendence in permutaion groups.* Quaderni del Seminario Mathematico di Brescia, 2003, N7, pp. 1–9.

[9] S. A. Norin *A Polynomial Lower Bound on the Size of a k-restricted Min-Wise Independent Set of Permutations.* Journal of Mathematical Sciences, 2003, vol.118, issue 2, pp.4994-5000.

[10] J. Matoušek, M. Stojaković, *On restricted min-wise independence of permutations*. Preprint.