PREPRINT PDMI — 12/2004

Constructing a $(k+1)$-restricted Min-Wise Independent Family from a k-restricted Min-Wise Independent Family when k is Odd

Victor Bargachev [1]

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

September 7, 2004

Abstract

Informally, a family $\mathcal{F} \subseteq S_n$ of permutations is k-restricted min-wise independent if, for any $X \subseteq [n]$ with $|X| \leqslant k$, each $x \in X$ has an equal chance of being mapped to the minimum among $\pi(X)$. In this paper we present a way to construct a $(k+1)$-restricted min-wise independent family from a k-restricted min-wise independent family when k is odd. As a corollary, we improve the existing upper bounds on the minimal size of k-restricted min-wise independent families for even $k \geqslant 4$.

[1]Supported in part by INTAS (grant no. 2000-447)

1 Introduction

The concept of min-wise independent permutations was introduced and studied in [1], [2] and [4]. Its appearance was motivated by the need for an efficient filtering of "almost identical" documents among huge collection of Web documents.

Similar concepts have also appeared in other papers in a slightly different context (see [4] for references). Later, min-wise independent permutations were applied to derandomization of algorithms for NP-complete problems ([3], [9]).

Let us recall the main definitions from [2]. Throughout this paper we use the following notation. Let $[n]$ denote the set $\{1, \ldots, n\}$. Let S_n denote the symmetric group on $[n]$. 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 $y \in Y$.

Definition 1 ([2], [4]). 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 $|X| \leq k$ and for any $x \in X$, when a permutation π is chosen at random from $\mathcal{F}$ with some specified probability distribution, we have

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

The family is called *unbiased* if the distribution on $\mathcal{F}$ is uniform, and *biased* otherwise.

2 From a k-MWI set to a $(k+1)$-MWI set

Given a k-MWI family $\mathcal{F}$ (either biased or unbiased), where k is odd, we construct a $(k+1)$-MWI family (biased in general) of size at most $2 \cdot |\mathcal{F}|$.

Lemma 1. *Let k be odd, $n > k$ and $\mathcal{F} \subseteq S_n$ be a k-MWI family. Assume further that for any $X \subseteq [n]$ with $|X| = k+1$ and for every $x \in X$ we have*

$$\Pr(\min \pi(X) = \pi(x)) = \Pr(\max \pi(X) = \pi(x)). \tag{2}$$

Then $\mathcal{F}$ is a $(k+1)$-MWI family.

Proof. Let us fix $X = \{x_0, \ldots, x_k\}$ and $x = x_0$. For any set $Z \subseteq [k]$ we define

$$P_Z = \{\pi \in \mathcal{F} : \ \pi(x_0) < \pi(x_j) \text{ for all } j \in Z\}$$

and

$$P = \bigcup_{z=1}^{k} P_{\{z\}} = \mathcal{F} \setminus \{\pi \in \mathcal{F} :\ \max(\pi(X)) = \pi(x_0)\}.$$

Let I_A denote the characteristic function of the set A, i.e. $I_A(\pi) = 1$ if $\pi \in A$ and $I_A(\pi) = 0$ otherwise. We have by the well-known inclusion-exclusion principle (recall here that k is odd)

$$\begin{aligned} I_P &= \sum_{l=1}^{k} (-1)^{l+1} \sum_{Z \subseteq [k], |Z|=l} I_{P_Z} \\ &= I_{P_{[k]}} + \sum_{l=1}^{k-1} (-1)^{l+1} \sum_{Z \subseteq [k], |Z|=l} I_{P_Z}. \end{aligned}$$

Evaluating expectations and taking into account (1) we obtain

$$\begin{aligned} \Pr(\min(\pi(X)) = \pi(x)) &= \mathrm{E}(I_{P_{[k]}}) \\ &= \mathrm{E}(I_P) + \sum_{l=1}^{k-1} (-1)^{l} \sum_{Z \subseteq [k], |Z|=l} \mathrm{E}(I_{P_Z}) \\ &= \mathrm{E}(I_P) + \sum_{l=1}^{k-1} (-1)^{l} \frac{1}{l+1} \binom{k}{l} \\ &= \mathrm{E}(I_P) + \frac{1-k}{1+k}. \end{aligned} \tag{3}$$

On the other hand, assumption (2) yields

$$\begin{aligned} \mathrm{E}(I_P) &= 1 - \Pr(\max(\pi(X)) = \pi(x_0)) \\ &= 1 - \Pr(\min(\pi(X)) = \pi(x_0)), \end{aligned}$$

which together with (3) implies

$$\Pr(\min(\pi(X)) = \pi(x_0)) = \frac{1}{k+1}.$$

□

The following lemma was independently discovered by several authors, e.g., see [7], [5]

Lemma 2 ([7], [5]). *Let $\mathcal{F}$ be a k-MWI set. Then for any $X \subseteq [n]$, $|X| \leq k$, and for any partition $X = A \cup \{z\} \cup B$, when a permutation π is chosen from $\mathcal{F}$ at random we have*

$$\Pr(\pi(A) < \pi(z) < \pi(B)) = \frac{|A|! \cdot |B|!}{(|A| + |B| + 1)!}.$$

In particular, for any such X and for any $x \in X$,

$$\Pr(\max \pi(X) = \pi(x)) = \frac{1}{|X|}.$$

For any $\mathcal{F} \subseteq S_n$ with some probability distribution on it and for any $\sigma \in S_n$ we define the set

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

and induce the probability distribution on it

$$\Pr_{\sigma\mathcal{F}}(\sigma\pi) = \Pr_{\mathcal{F}}(\pi).$$

In a similar way we define the set $\mathcal{F}\sigma$. Clearly, if $\mathcal{F}$ is a k-MWI family then the same is true for $\mathcal{F}\sigma$ for any permutation σ. But $\sigma\mathcal{F}$, in general, may be not a k-MWI set. However, in one special case multiplication on the left still preserves min-wise independence.

Lemma 3. *Let σ^* be the permutation, which reverts the order, i.e.,*

$$\sigma^* = (1\ n)(2\ n-1)\cdots$$

If $\mathcal{F}$ is a k-MWI set, then $\sigma^*\mathcal{F}$ is also a k-MWI set.

Proof. Note that $\Pr_{\sigma^*\mathcal{F}}(\min \pi(X) = \pi(x)) = \Pr_{\mathcal{F}}(\max \pi(X) = \pi(x))$. Now the claim follows from Lemma 2. □

Now we are able to prove the main result of this paper. Take $\mathcal{F}' = \mathcal{F} \cup \sigma^*\mathcal{F}$ and define the probability distribution as follows:

$$\Pr_{\mathcal{F}'}(\pi) = \begin{cases} \frac{1}{2}\Pr_{\mathcal{F}}(\pi), \text{ if } \pi \in \mathcal{F} \setminus \sigma^*\mathcal{F}; \\ \frac{1}{2}\Pr_{\sigma^*\mathcal{F}}(\pi), \text{ if } \pi \in \sigma^*\mathcal{F} \setminus \mathcal{F}; \\ \frac{1}{2}(\Pr_{\mathcal{F}}(\pi) + \Pr_{\sigma^*\mathcal{F}}(\pi)), \text{ if } \pi \in \mathcal{F} \cap \sigma^*\mathcal{F}. \end{cases}$$

Theorem 1. *Let $\mathcal{F}$ be a k-MWI set, where k is odd. Then the set $\mathcal{F}'$ defined above is a $(k+1)$-MWI set. In particular $|\mathcal{F}'| \leq 2 \cdot |\mathcal{F}|$.*

Proof. By Lemma 3, $\sigma^*\mathcal{F}$ is a k-MWI set and so is $\mathcal{F}' = \mathcal{F} \cup \sigma^*\mathcal{F}$. Note that for any $\pi \in \mathcal{F}$ and for any $X \subseteq \Omega$ and $x \in X$ we have

$$\begin{aligned} \min \pi(X) = \pi(x) \quad &\text{iff} \quad \max \sigma^*\pi(X) = \sigma^*\pi(x), \\ \max \pi(X) = \pi(x) \quad &\text{iff} \quad \min \sigma^*\pi(X) = \sigma^*\pi(x). \end{aligned}$$

In particular, (2) is true for $\mathcal{F}'$. Applying Lemma 1 we complete the proof. □

Remark 1. Even in the case when $\mathcal{F}$ is an unbiased k-MWI set, the distribution on $\mathcal{F}'$ can be nonuniform. However, if under conditions of Theorem 1 we have in addition that either $\sigma^*\mathcal{F} = \mathcal{F}$ or $\sigma^*\mathcal{F} \cap \mathcal{F} = \emptyset$, then $\mathcal{F}'$ is also an unbiased $(k+1)$-MWI set.

Itoh, Takei and Tarui [6] constructed unbiased k-MWI sets of permutations on $[n]$ of size $O(n^k)$, where the constant in the symbol O depended on k. Later [8] they constructed 3-MWI sets of size $O(n \log^2(n))$. Combining their results with Theorem 1 we obtain the following corollary.

Corollary 1. *Let k be even and $n \geq k$. There exists a k-MWI set on $[n]$ (probably, biased) of size $O(n^{k-1})$. There exists a 4-MWI set on $[n]$ (probably, biased) of size $O(n \log^2(n))$.*

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, M. Mitzenmacher, *Min-wise independent permutations.* J. Comput. Syst. Sci. 60(3), (2000), 630–659.

[3] 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), 15–24.

[4] A. Z. Broder, M. Mitzenmacher, *Completeness and robustness properties of min-wise independent permutations.* Random Struct. Algorithms 18(1), (2001), 18–30.

[5] J. Matoušek, M. Stojaković, *On restricted min-wise independence of permutations.* Random Struct. Algorithms 23, (2003), 397–408.

[6] T. Itoh, Y. Takei, J. Tarui *On permutations with limited independence.* Proc. of SODA'2000, (2000), 137–146.

[7] S. A. Norin *A polynomial lower bound for the size of a k-restricted min-wise independent set of permutations.* J. Math. Sci. 118(2), (2003), 4994-5000.

[8] J. Tarui, T. Itoh, Y. Takei, *A nearly linear size 4-min-wise independent permutation family by finite geometries.* RANDOM-APPROX (2003), 396–408.

[9] M. A. Vsemirnov, E. A. Hirsch, E. Ya. Dantsin, S. V. Ivanov, *Algorithms for SAT and upper bounds on their complexity.* J. Math. Sci. 118(2), (2003), 4948–4962.