Min-wise independent families with respect to any linear order

Peter J. Cameron and Pablo Spiga[*][†]
School of Mathematical Sciences
Queen Mary, University of London
Mile End Road
London, E1 4NS
U.K.

Abstract

A set of permutations $\mathscr{S}$ on a finite linearly ordered set Ω is said to be k-min-wise independent, k-MWI for short, if $\Pr(\min\pi(X)=\pi(x))=1/|X|$ for every $X\subseteq\Omega$ such that $|X|\leq k$ and for every $x\in X$. (Here $\pi(x)$ and $\pi(X)$ denote the image of the element x or subset X of Ω under the permutation π, and Pr refers to a probability distribution on $\mathscr{S}$, which we take to be the uniform distribution.) We are concerned with sets of permutations which are k-MWI families for any linear order. Indeed, we characterize such families in a way that does not involve the underlying order. As an application of this result, and using the Classification of Finite Simple Groups, we deduce a complete classification of the k-MWI families that are groups, for $k\geq 3$.

1 Introduction

We let $\mathrm{Sym}\,\Omega$ and $\mathrm{Alt}\,\Omega$ denote the symmetric group and the alternating group on the set Ω respectively. If k is a natural number then $\mathrm{Sym}(k)$ will denote the

[*]Supported by a grant from the Istituto Nazionale di Alta Matemati ca, INdAM

[†]Current address: Department of Mathematics, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta, Canada T1K 3M4

symmetric group on the set $\{1,\ldots,k\}$. We denote by $\pi(x)$ or $\pi(X)$ the image of the element x or subset X under the permutation π. If G is a permutation group on the set Ω and X is a subset of Ω then G_X denotes the set stabilizer of X in G, i.e. $G_X = \{g \in G \mid g(X) = X\}$. If $\leq$ is a linear order in Ω and X is a subset of Ω then we shall denote by $\min_{\leq} X$ the minimal element of X in $(\Omega, \leq)$. Moreover, in the case that $\alpha \leq \beta$ and $\alpha \neq \beta$ we will write $\alpha < \beta$. If σ is a permutation in Ω then it defines a linear order $\leq_\sigma$, where $\alpha \leq_\sigma \beta$ if and only if $\sigma^{-1}(\alpha) \leq \sigma^{-1}(\beta)$. The minimum element of X with respect to $\leq_\sigma$ will be denoted by $\min_{\leq_\sigma}(X)$.

Let $\mathscr{S}$ be a set of permutations of Ω, Pr be a probability distribution on $\mathscr{S}$ and k be a natural number. $\mathscr{S}$ is called a *k-min-wise independent family*, *k-MWI* for short, if

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

for any $X \subseteq \Omega$ such that $|X| \leq k$ and for any $x \in X$. This definition was motivated by applications in computer science. In fact such a family is important in algorithms used in practice by software to find duplicate documents, see [3]. Later, such sets were applied in other contexts such as derandomization of algorithms. We say that G is a *k-MWI group* if G is a k-MWI family and G is a permutation group of Ω.

In this paper we consider exclusively k-MWI families $\mathscr{S}$ for the uniform distribution. In [1] Theorem 3.1, it has been proved that if G is a k-MWI group with respect to some probability distribution Pr then G is k-MWI with respect to the uniform distribution. Therefore dealing with k-MWI groups our assumption is not at all a restriction.

We begin with a definition.

Definition 1 We say that a set of permutations $\mathscr{S}$ is locally k-MWI, $k \geq 1$, if for every subset X of size at most k, $\tau \in \mathscr{S}$ and for every $x \in X$, $y \in \tau(X)$ we have that

$$\frac{|\{\pi \in \mathscr{S} \mid \pi(X) = \tau(X), \pi(x) = y\}|}{|\{\pi \in \mathscr{S} \mid \pi(X) = \tau(X)\}|} = \frac{1}{|X|}.$$

Our main result is the following:

Theorem 1 *Let $\mathscr{S}$ be a set of permutations of* $\operatorname{Sym}\Omega$ *and k be a natural number. $\mathscr{S}$ is a k-MWI family with respect to any linear order and with respect the uniform distribution if and only if $\mathscr{S}$ is locally k-MWI.*

As a consequence of this theorem we prove a complete classification of the k-MWI groups with respect to any linear order in the underlying set Ω, for $k \geq 3$.

In the next section we give the proof of Theorem 1. Then we outline the classification of groups with this property, and discuss some further directions.

2 Proof of Theorem 1

Let Ω and $\mathscr{S}$ be as in the statement of the theorem. First we prove the forward direction. So suppose that $\mathscr{S}$ is k-MWI with respect to any linear ordering of Ω. Without loss of generality we may assume that $\Omega = \{1, \dots, n\}$.

Choose $h \leq k$. Let $A = \{1, \dots, h\}$, $B = \{2, \dots, h\}$, and set $\mathscr{F} = \{X \subseteq \Omega \mid |X| = h-1\}$. Now define a non-simple bipartite graph Γ: the vertex set of Γ is $\Omega \cup \mathscr{F}$; for each $\pi \in \mathscr{S}$, there is an edge joining $\pi(1) \in \Omega$ to $\pi(B) \in \mathscr{F}$.

For $\tau \in \mathscr{S}$, $i \in A$, and $Y \subseteq A$, let us denote by $f_\tau(i,Y)$ the number of edges (i,X) of Γ such that $X \cap \tau(A) = Y$, where $i \in \tau(A)$ and $Y \subseteq \tau(A)$.

Fix τ in $\mathscr{S}$, and pick σ in $\mathrm{Sym}(h)$ (the subgroup of $\mathrm{Sym}(n)$ fixing $\{h+1, \dots, n\}$ pointwise) and $\tau\sigma(i)$ in $\tau(A)$. The number of permutations π in $\mathscr{S}$ having $\pi(1) = \tau\sigma(i)$ as $\leq_{\tau\sigma}$-minimum of the set $\pi(A)$ is the number of edges $(\tau\sigma(i), \tau\sigma(X))$ in Γ such that $\tau\sigma(i) <_{\tau\sigma} \tau\sigma(X)$. By definition of $\leq_{\tau\sigma}$, this means $i < X$, and, as $i \in A$, this is equivalent to $i < X \cap A$. Summing, we have

$$|\{\pi \in \mathscr{S} \mid \min_{\leq_{\tau\sigma}} \pi(A) = \pi(1) = \tau\sigma(i)\}| = \sum_{Y \subseteq \{i+1,\dots,h\}} f_\tau(\tau\sigma(i), \tau\sigma(Y)).$$

Now, $\mathscr{S}$ is a k-MWI family with respect to any linear order on Ω. Therefore we have

$$\begin{aligned} \frac{|\mathscr{S}|}{h} &= |\{\pi \in \mathscr{S} \mid \min_{\leq_{\tau\sigma}} \pi(A) = \pi(1)\}| \\ &= \sum_{i=1}^{h} \sum_{Y \subseteq \{i+1,\dots,h\}} f_\tau(\tau\sigma(i), \tau\sigma(Y)) \\ &\qquad + |\{\pi \in \mathscr{S} \mid \min_{\leq_{\tau\sigma}} \pi(A) = \pi(1), \pi(1) \notin \tau(A)\}|. \end{aligned} \tag{1}$$

We claim that the second summand in (1) does not depend on $\sigma \in \mathrm{Sym}(h)$. Indeed, let π be a permutation in $\mathscr{S}$ such that $\min_{\leq_{\tau\sigma}} \pi(A) = \pi(1)$ and $\pi(1) \notin \tau(A)$. We get $\min_{\leq_\sigma} \tau^{-1}\pi(A) = \tau^{-1}\pi(1)$ and $\tau^{-1}\pi(1) \notin A$. Now, σ is a permutation stabilizing the set A and acting trivially on $\Omega \setminus A$; therefore we have $\min_{\leq} \tau^{-1}\pi(A) = \tau^{-1}\pi(1)$. This proves our claim. In particular, from equation (1) we have that

$$Q(\sigma) = \sum_{i=1}^{h} \sum_{Y \subseteq \{i+1,\dots,h\}} f_\tau(\tau\sigma(i), \tau\sigma(Y)) \tag{2}$$

is a constant that does not depend on the choice of σ in $\mathrm{Sym}(h)$.

We claim that $f_\tau(\tau(i), \tau(Y)) = f_\tau(\tau\sigma(i), \tau\sigma(Y))$ for every $\sigma \in \mathrm{Sym}(h)$ such that $\sigma(Y \cup \{i\}) = Y \cup \{i\}$. We prove this by induction on $|Y|$. Assume $|Y| = 1$. Let $1 \le i < j \le h$ and σ be a permutation of $\mathrm{Sym}(h)$ mapping i into $h-1$ and j into h. Using the definition of Γ we get

$$0 = Q(\sigma) - Q((h-1,h)\sigma) = f_\tau(\tau(i), \{\tau(j)\}) - f_\tau(\tau(j), \{\tau(i)\}).$$

Therefore $f_\tau(\tau(i), \{\tau(j)\}) = f_\tau(\tau(j), \{\tau(i)\})$. Assume the result for $|Y| = l-1$ and let us prove it for $|Y| = l$. Let $1 \le i_{l+1} < \cdots < i_2 < i_1 \le h$ and σ be a permutation mapping i_j into $h-j+1$. Consider the permutation $\eta = (h-l, \dots, h-1, h)$. Now, using the inductive hypothesis we have

$$\begin{aligned} 0 &= Q(\sigma) - Q(\eta\sigma) \\ &= \sum_{i=h-l}^{h} \sum_{Y \subseteq \{h-l+1,\dots,h\}} (f_\tau(\tau\sigma(i), \tau\sigma(Y)) - f_\tau(\tau\eta\sigma(i), \tau\eta\sigma(Y))) \\ &= f_\tau(\tau(i_{l+1}), \tau(\{i_l, \dots, l_1\})) - f_\tau(\tau(i_l), \tau\{i_{l-1}, \dots, i_1, i_{l+1}\}). \end{aligned}$$

Similarly, using η^{l-j+1} rather than η, we have

$$f_\tau(\tau(i_j), \tau(Y - \{i_j\})) = f_\tau(\tau(i_{l+1}), \tau(Y - \{i_{l+1}\}))$$

for every j, where $Y = \{i_{l+1}, \dots, i_1\}$.

Now we are ready to prove the forward implication in the theorem. By the previous discussion, $f(\tau(1), \tau(B)) = f(\tau\sigma(1), \tau\sigma(B))$ for every $\sigma \in \mathrm{Sym}(h)$. This proves that, for every x in $\tau(A)$, the number of elements in $\mathscr{S}$ such that $\pi(1) = x$ and $\pi(A) = \tau(A)$ equals the number of elements such that $\pi(1) = \tau(1)$ and $\pi(A) = \tau(A)$. Therefore we are done.

For the reverse implication, assume that $\mathscr{S}$ is locally k-MWI. Let $h \le k$ and let X be an h-set of Ω and $x \in X$. Let us denote by Σ the set $\{\pi(X) \mid \pi \in \mathscr{S}\}$. We have

$$\begin{aligned} |\{\pi \in \mathscr{S} \mid \min \pi(X) = \pi(x)\}| &= \sum_{Y \in \Sigma} |\{\pi \in \mathscr{S} \mid \pi(X) = Y, \min Y = \pi(x)|\} \\ &= \sum_{Y \in \Sigma} \frac{|\{\pi \in \mathscr{S} \mid \pi(X) = Y\}|}{|X|} = \frac{|\mathscr{S}|}{|X|}, \end{aligned}$$

so the theorem has been proved. We note that this direction of the proof was given in [5], Lemma 2, in the case where $\mathscr{S}$ is a group.

3 A consequence of Theorem 1

Corollary 1 *Let G be a finite permutation group on the set Ω. Then G is a k-MWI group with respect to any linear order in Ω if and only if for every subset X of Ω of size at most k we have that G_X is transitive on X.*

Proof This is immediate from Theorem 1.

We note that if, for every subset X of Ω of size k, the group G_X is transitive on X, then G is $(k-1)$-homogeneous. In fact, let A and B be $(k-1)$-sets. Assume that $A \cap B$ is a $(k-2)$-set. Then A and B lie in the same G-orbit. For if $X = A \cup B$ then $A = X \setminus \{b\}$ and $B = X \setminus \{a\}$, for some $a \in A$ and $b \in B$. Now, X is a k-set, so by hypothesis, G_X contains an element mapping a into b, and so, A into B. With an easy induction on $|A \cap B|$ and with a connectedness argument we get that all $(k-1)$-sets are in the same orbit.

This remark allow us to get the following classification.

Theorem 2 *Let G be a finite permutation group on the set Ω and let k be a positive integer with $k \geq 3$. Then the following conditions are equivalent:*

(a) G is a k-MWI group with respect to any linear order on Ω;

(b) G_X is transitive on X for any subset X of Ω with $|X| \leq k$;

(c) G is one of the groups from Table 1.

Proof (Sketch) Corollary 1 shows that (a) and (b) are equivalent. We have to show that (b) and (c) are equivalent.

Assume that (b) holds. Then G is h-homogeneous for any $h < k$ (in particular G is 2-homogeneous). Now, apart known exceptions, if G is a h-homogeneous group with degree n, for $h \leq n/2$, then G is h-transitive. The list of all possible exceptions can be found in [4]. Thus the proof of Theorem 2 is a case-by-case analysis among the list of 2-transitive groups and the list of groups in [4].

In this analysis, the following remark is useful.

Suppose that G is a t-transitive permutation group on G and that all $G_{\alpha_1,\ldots,\alpha_t}$-orbits except $\{\alpha_1\},\ldots,\{\alpha_t\}$ have different size. Then G_X is transitive on X for any subset X of Ω with $|X| \le t+1$. In particular G is $(t+1)$-MWI.

Using this tool, we can deal with the almost simple groups. For instance, M_{22} is 3-transitive and the stabilizer of four distinct points has orbits of size $1,1,1,3,16$. Therefore, M_{22} is 4-MWI with respect to any linear order. Furthermore, M_{22} is not 4-homogeneous, therefore M_{22} can not be 5-MWI with respect to all linear orders.

The analysis of the affine 2-transitive groups requires other remarks. We present and prove the main ingredient of this classification.

Let G be an affine 2-transitive group on V, V an n-dimensional $\mathbb{F}_q$-vector space, $q = p^m$. If G is a 3-MWI group with respect to any linear order then $q = 2,3,4$ or $q = 8$. In particular, if $q = 8$ then G contains the Galois group of $\mathbb{F}_8$.

To prove this, assume that $q > 2$. By Corollary 1, G_X is transitive on X for any $X \subseteq V$ of size 3. Fix $(e_i)_i$ a basis of V, $a \in \mathbb{F}_q \backslash \{0,1\}$ and $X = \{0, e_1, ae_1\}$. The group G_X is transitive on X if and only if it contains an element $\varphi : \xi \mapsto A\xi^\sigma + v$ such that $\varphi(0) = e_1$, $\varphi(e_1) = ae_1$ and $\varphi(ae_1) = 0$. This proves that for all $a \in \mathbb{F}_q \backslash \{0,1\}$ there exists $\sigma \in \mathrm{Aut}(\mathbb{F}_q)$ such that $a^{\sigma+1} - a^\sigma + 1 = 0$. In particular any $a \in \mathbb{F}_q \backslash \{0,1\}$ is a root of $X^{p^i+1} - X^{p^i} + 1$ for some i. This yields that the characteristic of $\mathbb{F}_q$ is either 2 or 3.

Assume that $q = 3^m$. The equation $X^{3^i+1} - X^{3^i} + 1$ has at most $3^i + 1$ roots. Therefore summing on all the equations we have $\sum_{i=0}^{m-1} (3^i + 1) \ge 3^m - 2$. This happens if and only if $m = 1$.

Consider the case $q = 2^m$. Now, let us study the solutions of the equation $X^{2^{m-1}+1} + X^{2^{m-1}} + 1$ in $\mathbb{F}_q$. We have $0 = X^{2^m} + X = X^{-2}(X^{2^{m-1}+1})^2 + X = X^{-2}(X^{2^m} + 1) + X = X + X^{-1} + X^{-2}$, if and only if $X^3 + X + 1 = 0$. Therefore $X^{2^{m-1}+1} + X + 1$ has at most 3 solutions in $\mathbb{F}_q$. This yields $\sum_{i=0}^{m-2} (2^i + 1) + 3 \ge 2^m - 2$. This happens if and only if $q = 2,4$ or $q = 8$. Now the remaining part is easy to achieve.

Further details of the classification may be obtained from the second author. In Table 1, C denotes the Galois group of $\mathbb{F}_8$ over $\mathbb{F}_2$.

For $k = 2$, no complete classification exists. The groups which are 2-MWI for every linear order are just those transitive groups for which every pair of points is interchanged by some group element. These groups are sometimes referred to as *generously transitive*, and have the property that the permutation character is multiplicity-free (so they are examples of *Gelfand pairs*), in which all irreducible constituents are real. See Saxl [6], for example.

4 Concluding remarks

For practical purposes it is often necessary to get a small k-MWI family. In other words, for fixed Ω and k, the complexity of the algorithms using MWI families is strictly related to the size of the family. So clearly the problem consists in finding a compromise between k and the size of the family $\mathscr{S}$. From Theorem 1 we realize that if the family has to be k-MWI with respect to any linear order then the actual size has to be comparatively big. In particular, it is worth noting that if G is a k-MWI group with respect to any linear order and $k \geq 7$ then G has to contain the alternating group $\mathrm{Alt}(\Omega)$, see Theorem 2. Therefore it is reasonable to look at particular orders of the underlying set. Bargachev [2] has shown that there are 4-MWI groups of degree n and size $O(n^2)$. From Table 1 we see that the order of a 4-MWI group with respect to any linear order and degree n has to be at least $\Omega(n^3)$.

Next we present a variant of this problem. We say that the family $\mathscr{S}$ is (ε,k)-MWI if

$$\frac{1}{|X|(1+\varepsilon)} \leq \Pr(\min \pi(X) = \pi(x)) \leq \frac{1}{|X|(1-\varepsilon)}$$

for every subset X of Ω of size at most k and for every $x \in X$. Here k is a positive integer and $\varepsilon \geq 0$. One might hope that for "small" values of ε the variety of families that arise is considerably richer than the previous ones. Also, we remark that a group G is (ε,k)-MWI with respect to some probability distribution Pr then G is (ε,k)-MWI with respect to the uniform distribution. The proof of this result is exactly the same as Theorem 3.1 in [1].

Also, mimicking Definition 1 one can define a local approximated version: indeed, a set of permutations $\mathscr{S}$ is locally (ε,k)-MWI, $k \geq 1$, if for every subset X of size at most k, $\tau \in \mathscr{S}$ and for every $x \in X$, $y \in \tau(X)$ we have that

$$\frac{1}{|X|(1+\varepsilon)} \leq \frac{|\{\pi \in \mathscr{S} \mid \pi(X) = \tau(X), \pi(x) = y\}|}{|\{\pi \in \mathscr{S} \mid \pi(X) = \tau(X)\}|} \leq \frac{1}{|X|(1-\varepsilon)}.$$

Clearly, if $\mathscr{S}$ is a locally (ε,k)-MWI family then $\mathscr{S}$ is (ε,k)-MWI with respect to any linear order, see the last paragraph of the proof of Theorem 1. A permutation group G which is locally (ε, k)-MWI for any $\varepsilon < 1$ is k-MWI, by the equivalence of (a) and (b) in Theorem 2.

Finally, we remark that every elementary abelian 2-group G, acting regularly, is $(\frac{1}{3}, 3)$-MWI with respect to any order. For take a 3-set $X = \{\alpha, \beta, \gamma\}$, and let $\delta \in \Omega$ be the point such that the stabilizer G_Y of $Y = \{\alpha, \beta, \gamma, \delta\}$ has order 4. It is easy to prove that for every $\sigma \in G$ we have

$$|\{\pi \in G_Y \mid \min \sigma\pi(X) = \sigma\pi(\alpha)\}| \in \{1, 2\}.$$

Summing over a transversal of G_Y in G we have that G is $(\frac{1}{3}, 3)$-MWI with respect to any linear order. The size of G is $n = |\Omega|$; a group which is 3-MWI with respect to any order has size at least $n(n-1)/2$. On the other hand, this group is not locally $(\varepsilon, 3)$-MWI for any $\varepsilon < 1$, since the stabiliser of a 3-set acts trivially on it.

References

[1] V. Bargachev, Some properties of min-wise independent families and Groups of permutations, *Zap. Nauchn. Sen. S.-Peterburg. Otdel. Mat. Inst. Steklov* 316 (2004), 30–41, 224–225.

[2] V. Bargachev, personal communication.

[3] A. Z. Broder, M. Charikar, A. M. Frieze and M. Mitzenmacher, Min-wise independent permutations, *J. Comput. Syst. Sci.* **60**, (2000), 630–659.

[4] J.D.Dixon and B.Mortimer, Permutation Groups, *Springer*, (1996).

[5] C. Franchi and M. Vsemirnov, Min-wise independent groups, *European Journal of Combinatorics* **24**, (2003), 630–659.

[6] J. Saxl, On multiplicity-free permutation representations, in *Finite Geometries and Designs* (ed. P. J. Cameron, J, W. P. Hirschfeld and D. R. Hughes), pp. 337–353, London Math. Soc. Lecture Notes **49**, Cambridge Univ. Press, Cambridge, 1981.

Table 1: The k-MWI groups with respect to any order ($k \geq 3$)

G	Condition	$(\|\Omega\|, k)$
$\mathrm{Alt}\Omega \leq G \leq \mathrm{Sym}\Omega$	$\|\Omega\| \geq 4$	$(\|\Omega\|, \|\Omega\|)$
M_{12}		$(12,6)$
M_{24}		$(24,6)$
M_{11}		$(11,5)$ or $(12,4)$
M_{23}		$(23,5)$
$M_{22} \leq G \leq \mathrm{Aut}M_{22}$		$(22,4)$
$\mathrm{PSL}(n,q) \leq G \leq \mathrm{P\Gamma L}(n,q)$	$n \geq 3$	$((q^n-1)/(q-1),3)$
$\mathrm{PGL}(2,q) \leq G \leq \mathrm{P\Gamma L}(2,q)$	$q \neq 4,5,7$	$(q+1,4)$
$\mathrm{PSL}(2,q) \leq G \leq \mathrm{P\Sigma L}(2,q)$	$q \neq 4,7$	$(q+1,3)$
$\mathrm{PSL}(2,7) \leq G \leq \mathrm{PGL}(2,7)$		$(8,4)$
$\mathrm{PGL}(2,5)$		$(6,6)$
$\mathrm{PSL}(2,11)$		$(11,3)$
$\mathrm{Alt}(7)$		$(15,3)$
HS		$(176,3)$
Co_3		$(276,3)$
$\mathrm{Sp}(2d,2)$	$d \geq 3$	$(2^{2d-1}+2^{d-1},3)$
$\mathrm{Sp}(2d,2)$	$d \geq 3$	$(2^{2d-1}-2^{d-1},3)$
$\mathrm{PGU}(3,q) \leq G \leq \mathrm{P\Gamma U}(3,q)$		$(q^3+1,3)$
$\mathrm{A\Gamma L}(1,q)$	$q=3,8$	$(q,3)$
$\mathrm{ASL}(n,q) \leq G \leq \mathrm{A\Gamma L}(n,q)$	$q=3,4$; $n \geq 2$	$(q^n,3)$
$\mathrm{ASL}(n,2)$	$n \geq 2$	$(2^n,4)$
$\mathrm{A\Sigma L}(n,8) \leq G \leq \mathrm{A\Gamma L}(n,8)$	$n \geq 2$	$(8^n,3)$
$V \rtimes \mathrm{Alt}(6)$		$(16,3)$
$V \rtimes \mathrm{Alt}(7)$		$(16,4)$
$V \rtimes \mathrm{PSU}(3,3)$		$(64,3)$
$V \rtimes G_2(q) \trianglelefteq G$	$q=2,4$	$(q^6,3)$
$V \rtimes (G_2(8) \cdot C) \trianglelefteq G$		$(8^6,3)$
$V \rtimes \mathrm{Sp}(2d,q) \trianglelefteq G$	$q=2,3,4; d \geq 3$	$(q^{2d},3)$
$V \rtimes (\mathrm{Sp}(2d,8) \cdot C) \trianglelefteq G$	$d \geq 3$	$(8^{2d},3)$