Explicit and Near-Optimal Construction of t-Rankwise Independent Permutations

Nicholas Harvey
Department of Computer Science and Department of Mathematics,
University of British Columbia, Vancouver, Canada

Arvin Sahami
Department of Computer Science and Department of Mathematics,
University of British Columbia, Vancouver, Canada

Abstract

Letting $t \leq n$, a family of permutations of $[n] = \{1, 2, \ldots, n\}$ is called t-rankwise independent if for any t distinct entries in $[n]$, when a permutation π is sampled uniformly at random from the family, the order of the t entries in π is uniform among the $t!$ possibilities.

Itoh et al. show a lower bound of $(n/2)^{\lfloor \frac{t}{4} \rfloor}$ for the number of members in such a family, and provide a construction of a t-rankwise independent permutation family of size $n^{O\left(t^2 / \ln(t)\right)}$.

We provide an explicit, deterministic construction of a t-rankwise independent family of size $n^{O(t)}$ for arbitrary parameters $t \leq n$. Our main ingredient is a way to make the elements of a t-independent family "more injective", which might be of independent interest.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Rankwise independent permutations

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2024.67

Category RANDOM

1 Introduction

An important topic in the area of pseudorandomness is the construction of random variables such that any t of them are independent (for some parameter $t \in \mathbb{N}$), given a small source of purely random bits. A fundamental notion introduced by Wegman and Carter in 1979 [2] is that of a t-independent family[1], defined as follows (see also [9, Definition 3.31]).

▶ **Definition 1** (t-independent family). *Let m, n, t be positive integers with $t \leq m$. A family $\mathcal{H}$ of functions mapping $[m] \to [n]$ is called t-independent if, when $h \in \mathcal{H}$ is chosen uniformly at random, for any t distinct $x_1, \ldots, x_t \in [m]$ and t elements $y_1, \ldots, y_t \in [n]$,*

$$\mathbb{P}\left(h(x_i) = y_i \text{ for } i = 1, \ldots, t\right) = \frac{1}{n^t},$$

or equivalently, that the t random variables $h(x_1), \ldots, h(x_t)$ are independently and uniformly distributed in $[n]$.

These t-independent families are well-studied, and have found various applications. One example is to derandomize a randomized algorithm that uses certain independent random variables, but one can relax the assumption of being *mutually* independent to any t of them being independent. Then often one can derandomize the algorithm by iterating over the elements of $\mathcal{H}$ to find a function for which the algorithm succeeds. See [9, section 3.5] for such

[1] Throughout this paper, we use the term "family" to refer to a multiset, meaning that the members need not be distinct.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024).
Editors: Amit Kumar and Noga Ron-Zewi; Article No. 67; pp. 67:1–67:12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

an application for the MaxCut problem. It is then desirable to have explicit constructions of such a family $\mathcal{H}$ with small size. In fact, explicit constructions of such families of near-optimal size are known [4] for any parameters $1 \leq t \leq m$ and any n.

One can also define analogous families when restricting to permutations of $[n]$ instead of general functions. This natural restriction yields the notion of t-independent permutations.

▶ **Definition 2** (t-independent permutation). *A family Π is called t-independent if it contains permutations of $[n]$ such that, for any t distinct $x_1, \ldots, x_t \in [n]$ and any t distinct elements $y_1, \ldots, y_t \in [n]$,*

$$\mathbb{P}\left(\pi(x_i) = y_i \text{ for } i = 1, \ldots, t\right) = \prod_{i=0}^{t-1} \frac{1}{n-i}$$

when $\pi \in \Pi$ is chosen uniformly at random.

Explicit construction of such families with a small size, namely such that $|\Pi| \leq n^{O(t)}$, remains an open problem. This bound is near-optimal, since there is an obvious lower bound of $|\Pi| \geq \prod_{i=0}^{t-1}(n-i)$, which follows from the definition.

In fact, there are few non-trivial constructions of such families for any $t \geq 4$. Perhaps the closest result in this direction is a probabilistic proof for the existence of small (i.e., with $|\Pi| \leq n^{O(t)}$) t-independent permutations for any $1 \leq t \leq n$ due to Kuperberg, Lovett and Peled [7]. However, their proof does not seem to yield an efficient deterministic or randomized construction of the family, as it has a tiny success probability.

Many relaxed notions related to t-independence have been proposed for permutation families, including "t-restricted min-wise independent" [1] and "t-rankwise independent" families [5]. The latter is the focus of this paper.

▶ **Definition 3** (t-rankwise independent permutation). *A family Π of permutations over $[n]$ is called t-rankwise independent if for any t distinct points $x_1, \ldots, x_t \in [n]$,*

$$\mathbb{P}\left(\pi(x_1) < \pi(x_2) < \ldots < \pi(x_t)\right) = \frac{1}{t!}$$

when $\pi \in \Pi$ is chosen uniformly at random.

Another interesting type of permutation families has recently been proposed in the cryptography community. This is the notion of a *perfect sequence covering array* (PSCA).

▶ **Definition 4** (Yuster [10]). *Let $t \leq n$. The family Π of permutations of $[n]$ is called a* $\mathrm{PSCA}(n, t)$ *if there exists a fixed $\lambda \in \mathbb{N}$ such that for any t distinct indices $i_1, \ldots, i_t \in [n]$, there are exactly λ permutations $\pi \in \Pi$ such that*

$$(i_1, i_2, \ldots, i_t) \text{ is a subsequence of } (\pi(1), \pi(2), \ldots, \pi(n)).$$

(The notation and wording have been adapted to match ours.)

Let $g^*(n, t)$ denote the smallest size of a PSCA family Π. Naturally, researchers in this field are interested in the value of $g^*(n, t)$, and in the construction of families that asymptotically achieve this minimum size.

It was observed in [6] that t-rankwise independent families and PSCAs are isomorphic. Specifically, Π is a $\mathrm{PSCA}(n, t)$ family if and only if $\Pi^{-1} = \left\{\pi^{-1} \colon \pi \in \Pi\right\}$ is a t-rankwise independent family of permutations over $[n]$. Consequently, our construction of t-rankwise independent permutations can immediately be translated into a construction of PSCAs. Henceforth we will only use the terminology of t-rankwise independent families, and will no longer refer to PSCAs.

Itoh et al. [5] show a lower bound of $(n/2)^{\lfloor \frac{t}{4} \rfloor} \leq |\Pi|$ for the size of a t-rankwise independent family Π. They also construct a family Π with $|\Pi| \leq n^{O(t^2/\ln(t))}$, which does not asymptotically match the lower bound.

We present a deterministic algorithm for constructing a t-rankwise independent family Π of permutations over $[n]$, with $|\Pi| \leq n^{O(t)}$. This asymptotically matches the known lower bound. Formally, the following is our main result.

▶ **Theorem 5** (Main). *There exists a constant $C > 0$ such that the following is true. Let n, t be positive integers with $t \leq n$. Then there exists a t-rankwise independent family Π consisting of permutations of $[n]$ such that $|\Pi| \leq (Cn)^{35t}$. Furthermore, the whole family can be constructed by a deterministic algorithm in $n^{O(t)}$ time. (The implied constant in the $O(.)$ notation does not depend on either n or t).*

Our construction starts in Section 2.2 with a t-independent family $\mathcal{H}$, based on Reed-Solomon codes. The next step, appearing in Section 2.3, modifies it to obtain another t-independent family $\mathcal{G}$ whose members, roughly speaking, look "more injective". This step is the main technical contribution of the paper, and might be of independent interest. (Note that, since $\mathcal{G}$ is a t-independent family, not all the maps in $\mathcal{G}$ can be injective). Finally, in Section 2.4, we use this t-independent family $\mathcal{G}$ to construct permutations of $[n]$, yielding the t-rankwise independent family Π.

2 The construction

2.1 Overview

Our construction involves three steps, which build upon each other.

1. Construct $\mathcal{H}$, a t-independent family of $[n] \to \mathbb{Z}_N$ maps, where $N = \Theta(n^3)$.
2. Construct $\mathcal{G}$, a t-independent family of $[n] \to \mathbb{Z}_N$ maps, such that each map's image has size at least $n - 16t$. Intuitively, this condition says that each map has very few collisions, or is almost injective. (Being injective is equivalent to the image having size exactly n).
3. Construct Π, a t-rankwise independent family of permutations on $[n]$.

The most substantial of these steps is the construction of $\mathcal{G}$, whereas the construction of $\mathcal{H}$ is the most trivial. We explain these steps in the following sections.

2.2 Construction of $\mathcal{H}$

The construction of $\mathcal{H}$ is standard. The first step is to find a prime p in the interval $[n^3, 2n^3]$. This must exist, by Bertrand's postulate, and can be found in $\tilde{O}(n^3)$ time using exhaustive search and a deterministic primality test. We set $N = p$, and therefore

$$n^3 \;\leq\; N \;\leq\; 2n^3. \tag{1}$$

Let $\mathcal{H}$ be the family of $[n] \to \mathbb{F}_N$ maps defined by polynomials over $\mathbb{F}_N$ of degree less than t, namely

$$\mathcal{H} = \left\{ \sum_{0 \leq i \leq t-1} a_i x^i : a_i \in \mathbb{F}_N \right\}.$$

This family is well-known to be t-independent; see, e.g., [3, Exercise 5.8]. Note that the size of the family is $|\mathcal{H}| = p^t = N^t$.

2.3 Construction of $\mathcal{G}$

The next step is to use the family $\mathcal{H}$ to build a family $\mathcal{G}$. Each map in $\mathcal{H}$ will yield exactly one map in $\mathcal{G}$. The family $\mathcal{G}$ will retain $\mathcal{H}$'s property of being t-independent. In addition, we will be able to guarantee that every map in $\mathcal{G}$ has image size at least $n - 16t$. Thus each map has few collisions (although this is an informal term that we have not yet defined).

The family $\mathcal{G}$ has a simple form, and it is constructed by the pseudocode shown in Algorithm 1. This algorithm computes a single, specific map $\alpha : [n] \to \mathbb{Z}_N$, then it constructs

$$\mathcal{G} \;=\; \{\; h + \alpha \;:\; h \in \mathcal{H} \;\}.$$

▷ Claim 6. For any map α, the resulting family $\mathcal{G}$ will be t-independent.

Proof. Suppose that h is chosen uniformly at random from $\mathcal{H}$. For any t distinct entries $x_1, \ldots, x_t \in [n]$, $\{h(x_i)\}_{i\in[t]}$ are independent, and hence $\{f_i(h(x_i))\}_{i\in[t]}$ are independent for any deterministic functions f_i. In particular, since α is not random, letting $f_i(z) = z + \alpha(x_i)$, we have that $\{h(x_i) + \alpha(x_i)\}_{i\in[t]}$ remain independent. Lastly, for any $k \in [n]$, $h(k) + \alpha(k)$ is uniformly distributed since $h(k)$ is uniform in $\mathbb{Z}_N$, and α is not random. Thus $\{(h + \alpha)(x_i)\}_{i\in[t]}$ are independent and uniform in $\mathbb{Z}_N$, as desired. ◁

We will prove that there is a specific choice of α such that *every* $h \in \mathcal{H}$ satisfies

$$|(h + \alpha)([n])| \;=\; |\{\; h(x) + \alpha(x) \;:\; x \in [n] \;\}| \;\geq\; n - 16t,$$

which is the desired property of the family $\mathcal{G}$. In fact, it is possible to show that a random choice of α will satisfy this property with positive probability. However, this would not quite achieve the goals of this paper, since ultimately we want an explicit, deterministic construction of a t-rankwise independent family of permutations. Instead, we will obtain a deterministic construction by derandomizing the randomized construction of α.

Algorithm 1 contains pseudocode for this procedure, which we now briefly explain. The algorithm computes the values $\alpha(1), \alpha(2), \ldots, \alpha(n)$ one-by-one, in that order. Thinking of $h + \alpha$ as mapping the "balls" $[n]$ to the "bins" $\mathbb{Z}_N$, then S_k^h is the set of bins that have already received balls (for this particular function h). In order to be as injective as possible, we want to avoid a collision (for every h) between the k^{th} ball and these bins – that is, we want $(h + \alpha)(k) \notin S_k^h \;\; \forall h \in \mathcal{H}$. To do so, the algorithm uses a potential function (shown in (2)) in which the variable x corresponds to the value that will be used for $\alpha(k)$. This function penalizes any value x which would cause any further collision among any function $h \in \mathcal{H}$. This potential function is essentially a pessimistic estimator, as explained in Section 2.3.1 below.

▶ **Lemma 7.** *Algorithm 1 returns a t-independent family $\mathcal{G}$ satisfying the following.*

$$|g([n])| \geq n - 16t \quad \forall g \in \mathcal{G}$$

The subset of the codomain that experienced a "collision" is defined to be

$$\mathcal{Y} = \left\{\; y \in \mathbb{Z}_N \;:\; |g^{-1}(y)| \geq 2 \;\right\},$$

and the subset of the domain involved in these collisions is defined to be

$$\mathcal{X} \;=\; \bigcup_{y\in\mathcal{Y}} g^{-1}(y) \;=\; g^{-1}(\mathcal{Y}).$$

▶ **Corollary 8.** *The family $\mathcal{G}$ produced by Lemma 7 satisfies $|\mathcal{X}| \leq 32t$.*

Algorithm 1 Main Algorithm.

Input: t-independent family $\mathcal{H}$ of $[n] \to \mathbb{Z}_N$ maps s.t. $|\mathcal{H}| = N^t$.
Output: t-independent family $\mathcal{G}$ of $[n] \to \mathbb{Z}_N$ maps s.t. $|\mathcal{G}| = N^t$, $|g([n])| \geq n - 16t \ \forall g \in \mathcal{G}$.

$\lambda \leftarrow \ln(16tN/n^2)$
$\mathcal{G} \leftarrow \emptyset$
for $k = 1, \ldots, n$ **do**
  ▷ *Compute the value* $\alpha(k)$
  for $h \in \mathcal{H}$ **do**
    Let $S_k^h = \{\, h(i) + \alpha(i) \;:\; 1 \leq i \leq k-1 \,\} \subseteq \mathbb{Z}_N$, and note that $S_1^h = \emptyset$.
    This is $(h+\alpha)([k-1])$, the set of values that already appear in the image of $h+\alpha$.
    Define

$$\beta_k^h\big(\alpha(1), \alpha(2), \ldots, \alpha(k-1), x\big) \;=\; \begin{cases} 1 & \text{if } h(k) + x \in S_k^h \\ 0 & \text{otherwise} \end{cases}$$

    To ease notation, we will use the shorthand
    $\beta_k^h(x) = \beta_k^h\big(\alpha(1), \alpha(2), \ldots, \alpha(k-1), x\big)$.
  end for
  Pick

$$a \in \operatorname{argmin}_{x \in \mathbb{Z}_N} \sum_{h \in \mathcal{H}} \exp\left(\lambda\Big(\beta_k^h(x) + \sum_{1 \leq i \leq k-1} \beta_i^h\big(\alpha(1), \ldots, \alpha(i)\big)\Big)\right) \tag{2}$$

  Let $\alpha(k) \leftarrow a$
end for
return the family $\mathcal{G} = \{\, h + \alpha \;:\; h \in \mathcal{H} \,\}$.

A formal proof is in Appendix A, and here we present only a sketch.

Proof (Sketch). The size of $\mathcal{X}$ is maximized by having exactly $16t$ bins containing exactly 2 balls, and $n - 32t$ bins containing exactly 1 ball. ◀

2.3.1 Proof of Lemma 7

For each function $h \in \mathcal{H}$ and integer $k \in [n]$, there is a function $\beta_k^h \colon \mathbb{Z}_N^k \to \{0,1\}$ that is defined in Algorithm 1, and which we define equivalently here as

$$\beta_k^h(x_1, \ldots, x_k) \;=\; \begin{cases} 1 & \text{if } \exists 1 \leq i \leq k-1 \text{ s.t. } h(k) + x_k = h(i) + x_i \pmod N \\ 0 & \text{otherwise.} \end{cases}$$

We will use the notation $\beta_k^h(x_k)$ for $\beta_k^h(x_1, \ldots, x_k)$ when $x_1, \ldots, x_{k-1}$ are clear from context.

The scalar $\lambda > 0$ is as defined as in Algorithm 1. Additionally, define the scalar $c_\lambda > 0$ and the function $\psi_k \colon \mathbb{Z}_N^k \to \mathbb{R}^+$ by

$$\begin{aligned} c_\lambda \;&=\; \mathbb{E}\exp(\lambda Y) > 0 \\ \psi_k(x_1, \ldots, x_k) \;&=\; \sum_{h \in \mathcal{H}} \exp\left(\lambda \sum_{i=1}^{k} \beta_i^h(x_1, \ldots, x_i)\right) \cdot c_\lambda^{n-k}, \end{aligned} \tag{3}$$

where Y is a random variable having the Bernoulli distribution with parameter n/N, which we write as $\mathrm{Bern}\,(n/N)$. We will often write $\psi_k(x_k)$ instead of $\psi_k(x_1,\ldots,x_k)$ for notational convenience.

Intuitively, $\psi_k(x_1,\ldots,x_k)$ is a pessimistic estimator of the expected number of functions $h \in \mathcal{H}$ which would have $|(h+\alpha)([n])| > n - 16t$ given that $\alpha(i) = x_i\ \forall i \in [k]$, and that the rest of the entries $\alpha(k+1),\ldots,\alpha(n)$ are chosen uniformly at random from $\mathbb{Z}_N$.

Let $\alpha\colon [n] \to \mathbb{Z}_N$ be the mapping constructed by Algorithm 1.

▷ Claim 9. $\psi_0 \geq \psi_1(\alpha(1)) \geq \psi_2(\alpha(2)) \geq \ldots \geq \psi_n(\alpha(n))$, where here we use the notation $\psi_i(\alpha(i))$ to denote $\psi_i\,(\alpha(1),\alpha(2),\ldots,\alpha(i))$.

▷ Claim 10. $1 > \psi_0 = \exp(-16\lambda t)\cdot|\mathcal{H}|\cdot[\mathbb{E}\exp(\lambda Y)]^n$.

Together, Claims 9 and 10 imply that

$$1 \;>\; \psi_n(\alpha(n)) \;=\; \sum_{h\in\mathcal{H}} \exp\left(\lambda\Big(\sum_{i=1}^{k}\beta_i^h(\alpha(i))\Big) - 16\lambda t\right).$$

Since all summands are non-negative, it follows that, for every $h \in \mathcal{H}$, we have

$$\exp\left(\lambda\Big(\sum_{i=1}^{k}\beta_i^h(\alpha(i))\Big) - 16\lambda t\right) \;<\; 1.$$

Observe that $\sum_{i\leq k}\beta_i^h(\alpha(i)) = k - |S_k^h|\ \forall k,h$. Taking the log and rearranging, we obtain that

$$n - |S_n^h| \;=\; \sum_{i=1}^{n}\beta_i^h(\alpha(i)) \;<\; 16t \qquad \forall h \in \mathcal{H}.$$

Let $g = h + \alpha$. Since $|g([n])| = |S_n^h|$, we have $|g([n])| > n - 16t$ for all $h \in \mathcal{H}$. This completes the proof of Lemma 7.

Proof of Claim 9. We will show that $\psi_k(\alpha(k)) \leq \psi_k(\alpha(k-1))\ \forall 1 \leq k \leq n$. So let $k \in [n]$ be arbitrary.

Our first observation is that, in the algorithm's iteration k, it chooses the value $a = \alpha(k)$ to minimize $\psi_k(\alpha(1),\ldots,\alpha(k-1),a)$. This holds because the functions

$$\sum_{h\in\mathcal{H}}\exp\left(\lambda\beta_k^h(x) + \lambda\sum_{i=1}^{k-1}\beta_i^h(\alpha(i))\right) \qquad \text{and} \qquad \psi_k(\alpha(1),\ \alpha(2),\ldots,\alpha(k-1),x)$$

are positive multiples of each other.

Since $\alpha(k)$ minimizes ψ_k, we clearly have

$$\psi_k(\alpha(1),\ldots,\alpha(k)) \;\leq\; \mathbb{E}_{U\sim\mathrm{Unif}(\mathbb{Z}_N)}\psi_k(\alpha(1),\ldots,\alpha(k-1),U),$$

where $\mathrm{Unif}(S)$ denotes the uniform distribution on the set S. Hence in order to show that $\psi_k(\alpha(k)) \leq \psi_{k-1}(\alpha(k-1))$, it suffices to prove that

$$\mathbb{E}_{U\sim\mathrm{Unif}(\mathbb{Z}_N)}\,\psi_k(\alpha(1)\ldots,\alpha(k-1),U) \;\leq\; \psi_{k-1}(\alpha(k-1)). \tag{4}$$

Since ψ_k and ψ_{k-1} are both sums over $h \in \mathcal{H}$, it will suffice to prove this inequality for each summand. More specifically, we will ignore the $e^{-16\lambda t}$ constant and define

$$\psi_k^h(x) \;=\; \exp\left(\lambda\sum_{i=1}^{k-1}\beta_i^h(\alpha(i)) + \lambda\beta_k^h(x)\right)\cdot c_\lambda^{n-k},$$

where, as above, $c_\lambda = \mathbb{E}\exp(\lambda Y)$, and Y is Bern(n/N). Towards our inductive proof, we may rewrite this as

$$\psi_k^h(x) \;=\; \psi_{k-1}^h\big(\alpha(k-1)\big) \cdot \frac{1}{c} \cdot \exp\big(\lambda\beta_k^h(\alpha(1),\ldots,\alpha(k-1),x)\big).$$

Plugging this into our goal (4), it suffices to prove that

$$\mathbb{E}_{U\sim\mathrm{Unif}(\mathbb{Z}_N)}\; \psi_{k-1}^h\big(\alpha(k-1)\big) \cdot \frac{1}{c} \cdot \exp\big(\lambda\beta_k^h(\alpha(1),\ldots,\alpha(k-1),U)\big) \;\le\; \psi_{k-1}\big(\alpha(k-1)\big),$$

or equivalently (observing that $\psi_{k-1}^h(\alpha(k-1)) > 0$),

$$\mathbb{E}_{U\sim\mathrm{Unif}(\mathbb{Z}_N)} \exp\big(\lambda\beta_k^h(\alpha(1),\ldots,\alpha(k-1),U)\big) \;\le\; c_\lambda = \mathbb{E}\exp(\lambda Y). \tag{5}$$

Note that there are exactly $|S_k^h|$ values of U that result in $\beta_k^h(\alpha(1),\alpha(2),\ldots,\alpha(k-1),U)$ taking the value 1, whereas the rest result in the value 0. Since U is uniformly distributed on $\mathbb{Z}_N$ and $|S_k^h| \le n$ for all $k \in [n]$, $h \in \mathcal{H}$, it follows that $\beta_k^h(\alpha(1),\ldots,\alpha(k-1),U)$ has a Bernoulli distribution Bern(p) where $p \le n/N$. Since Y has the distribution Bern(n/N), the desired inequality (5) follows. ◁

For the next proof, we will require the following statement of the Chernoff bound. A proof is given in Appendix A.

▶ **Theorem 11** (Poisson tail of Chernoff bound)**.** *Let $Y_1,\ldots,Y_n$ be independent random variables supported on $[0,1]$. Let $\mu = \mathbb{E}\sum_{i=1}^n Y_i$. Then, for any $\delta \ge 1$, if $\lambda = \ln(1+\delta)$ then*

$$\mathbb{P}\left(\sum_{i=1}^n Y_i \ge (1+\delta)\mu\right) \;\le\; \mathbb{E}\exp\left(\lambda\sum_{i=1}^n Y_i - \lambda(1+\delta)\mu\right) \;\le\; (1+\delta)^{-(1+\delta)\mu/4}.$$

Proof of Claim 10. Let $Y_1,\ldots,Y_n$ be i.i.d. Bern$(\frac{n}{N})$ random variables. We may rewrite the definition of ψ_0 from (3) using these Y_i random variables as

$$\psi_0 \;=\; |\mathcal{H}| \cdot \mathbb{E}\exp\left(\lambda\sum_{i=1}^n Y_i - 16\lambda t\right).$$

To prove the claim, we must show that this is less than 1.

To do so, consider any fixed $h \in \mathcal{H}$. We will use the Chernoff bound as stated in Theorem 11, with $1+\delta = 16tN/n^2$. (Note that $\delta \ge 1$, as required, since $N \ge n^3$.) The value of λ required by the theorem is $\ln(1+\delta) = \ln(16tN/n^2)$, which matches the definition in Algorithm 1. Lastly, note that

$$\mu = \mathbb{E}\sum_{i=1}^n Y_i = n^2/N,$$

since each Y_i is Bern(n/N). Thus $\lambda(1+\delta)\mu = 16\lambda t$. Applying the theorem, we obtain

$$\mathbb{E}\exp\left(\lambda\sum_{k=1}^n Y_i - 16\lambda t\right) \;\le\; (1+\delta)^{-(1+\delta)\mu/4} \;=\; \big(16tN/n^2\big)^{-4t} \;<\; n^{-4t} \;\le\; N^{-t},$$

since $n^3 \le N \le 2n^3$ by (1), and also using $n \ge 2$. Thus, in conclusion

$$\psi_0 \;<\; |\mathcal{H}| \cdot N^{-t} \;=\; 1.$$ ◁

Algorithm 2 Construction of Π from $\mathcal{G}$.

Input: t-independent family $\mathcal{G}$ of $[n] \to \mathbb{Z}_N$ maps.
Output: t-rankwise independent family of permutations on $[n]$.

1: Let $\Pi \leftarrow \emptyset$
2: Let $\tau \leftarrow 32t$
3: **for** $g \in \mathcal{G}$ **do**
4:     Let $\Sigma = \{ (\sigma_1, \dots, \sigma_N) : \sigma_i \text{ is a permutation of } g^{-1}(i) \}$
5:     Let $s \leftarrow \tau!/|\Sigma|$
6:     **for** $(\sigma_1, \dots, \sigma_N) \in \Sigma$ **do**
7:         Let $L \leftarrow [\,]$ be an empty list
8:         **for** $i = 1, \dots, N$ **do**
9:             Append to L the elements of $g^{-1}(i)$ in the order given by σ_i
10:         **end for**
11:         Add s copies of the permutation $\pi : [n] \to [n]$, where $\pi(i) = L[i]$, to the set Π
12:     **end for**
13: **end for**
14: **return** Π

2.4 Construction of Π

The last step is to use the family $\mathcal{G}$ of maps to build the t-rankwise independent family Π of permutations on $[n]$. Pseudocode for this process is shown in Algorithm 2. Roughly speaking, the algorithm first sorts the elements of $[n]$ according to the order induced by the functions in $\mathcal{G}$ and then "breaks ties" using permutations in Σ (see line 4); also note that the number of new permutations will hence depend on $|\Sigma|$ which is not necessarily fixed for all $g \in \mathcal{G}$. The algorithm finally inserts the new permutations in Π. Note that in the algorithm, we view integers $i \in [N]$ as elements of $\mathbb{Z}_N$ in the natural manner.

In order for line 11 to make sense, we must establish the following claim.

▷ Claim 12. The value $s = \tau!/|\Sigma|$ is a positive integer.

Proof. As above, define

$$\mathcal{Y} = \{ y \in \mathbb{Z}_N : |g^{-1}(y)| \geq 2 \}$$
$$\mathcal{X} = \bigcup_{y \in \mathcal{Y}} g^{-1}(y) = g^{-1}(\mathcal{Y}).$$

Informally, $\mathcal{Y}$ is the set of bins containing multiple balls, and $\mathcal{X}$ is the set of balls that are not alone in their bin. By Lemma 7, we know that $|\mathcal{X}| \leq 32t = \tau$.

Let S_K denote the symmetric group on the set K. Observe that Σ is simply the direct product $\prod_{y \in \mathbb{Z}_N} S_{g^{-1}(y)}$, which has an obvious isomorphism to $\prod_{y \in \mathcal{Y}} S_{g^{-1}(y)}$, since we can ignore y with $|g^{-1}(y)| \in \{0, 1\}$. In turn, this is isomorphic to a subgroup of $S_{\mathcal{X}}$. It follows that $|\Sigma|$ divides $|S_{\mathcal{X}}|$, which divides $\tau!$ since $|\mathcal{X}| \leq \tau$. ◁

▷ Claim 13. The family Π is t-rankwise independent.

Proof. We want to show

$$\mathbb{P}\left(\pi(x_1) < \ldots < \pi(x_t)\right) = \frac{1}{t!} \tag{6}$$

for any t distinct indices $x_1, \dots, x_t$. For notational convenience, let us assume $x_1 = 1, x_2 = 2, \dots, x_t = t$. It can be seen that our proof does not use the indices $x_1, \dots, x_t$.

To generate π, we will first pick $g \in \mathcal{G}$ uniformly at random, then pick $(\sigma_1, \ldots, \sigma_N) \in \Sigma$ uniformly at random. Since each $g \in \mathcal{G}$ produces exactly $\tau!$ elements in Π, this is equivalent to picking π uniformly. Note that, since Σ is a Cartesian product, the distribution on the σ_i is equivalent to picking $\sigma_i \in S_{g^{-1}(i)}$ uniformly and independently at random.

For $i \in [t]$ define

$$R_i = \text{rank of } \pi(i) \text{ among } \pi(1), \ldots, \pi(t) = |\{j \in [t] \colon \pi(j) \le \pi(i)\}|.$$

Let $\overline{R} = (R_1, \ldots, R_t)$. Let us view $\overline{R}$ as an element of the symmetric group S_t (with $\overline{R}(i) = R_i$). In the remainder of the proof, we will establish that

$$\mathbb{P}\left(\overline{R} = r\right) \;=\; \mathbb{P}\left(\overline{R} = r\rho\right) \quad \forall r, \rho \in S_t. \tag{7}$$

Together with the fact that $1 = \sum_{\rho \in S_t} \mathbb{P}\left(\overline{R} = r\rho\right)$, we obtain $\mathbb{P}\left(\overline{R} = r\right) = \frac{1}{t!}\ \forall r \in S_t$. Thus, when r is the identity permutation, this establishes (6), for the case $x_i = i\ \forall i \in [t]$.

In order to prove (7), let us introduce some notation for convenience. Throughout the proof, let $\overline{X}$ denote the random vector $(X_1, X_2, \ldots, X_t)$ where $X_i = g(i)$. Let $\bar{i}$ denote the t-tuple $\bar{i} = (i_1, \ldots, i_t) \in \mathbb{Z}_N^t$. Intuitively, X gives the random locations of the first t balls, and $\bar{i}$ gives a specific list of locations that might be the outcome for those balls.

By the law of total probability

$$\mathbb{P}\left(\overline{R} = r\right) \;=\; \sum_{\bar{i} \in \mathbb{Z}_N^t} \mathbb{P}\left(\overline{R} = r \mid \overline{X} = \bar{i}\right) \cdot \mathbb{P}\left(\overline{X} = \bar{i}\right) \tag{8}$$

$$\mathbb{P}\left(\overline{R} = r\rho\right) \;=\; \sum_{\bar{i} \in \mathbb{Z}_N^t} \mathbb{P}\left(\overline{R} = r\rho \mid \overline{X} = \bar{i}\right) \cdot \mathbb{P}\left(\overline{X} = \bar{i}\right) \tag{9}$$

Since ρ is a permutation, one can write the second equation as

$$\mathbb{P}\left(\overline{R} = r\rho\right) \;=\; \sum_{\bar{i} \in \mathbb{Z}_N^t} \mathbb{P}\left(\overline{R} = r\rho \mid \overline{X} = \bar{i}\rho\right) \cdot \mathbb{P}\left(\overline{X} = \bar{i}\rho\right), \tag{10}$$

where, for a t-tuple v and permutation $\rho \in S_t$, the notation $v\rho$ denotes the t-tuple whose coordinates are permuted according to ρ, i.e., $(v\rho)_i = v_{\rho(i)}$.

Observe that by the t-independence of $X_1, \ldots, X_t$, we have

$$\mathbb{P}\left(\overline{X} = \bar{i}\right) = \mathbb{P}\left(\overline{X} = \bar{i}\rho\right) \;=\; \frac{1}{N^t}.$$

Thus to show (8) equals (10), it suffices to show that

$$\mathbb{P}\left(\overline{R} = r \mid \overline{X} = \bar{i}\right) \;=\; \mathbb{P}\left(\overline{R} = r\rho \mid \overline{X} = \bar{i}\rho\right).$$

Call the permutation $r \in S_t$ "feasible" w.r.t. the sequence $i_1, \ldots, i_t$ if for any $p, q \in [t]$, if $i_p < i_q$ then $r(p) < r(q)$. In words, this means that the order of $i_1, \ldots, i_t$ is given by the permutation r. It is possible that several indices in $[t]$ have the same value in the sequence $i_1, \ldots, i_t$, in which case r is allowed to induce any ordering among them.

We observe that $\mathbb{P}\left(\overline{R} = r \mid \overline{X} = \bar{i}\right) = 0 \iff r$ is not feasible w.r.t $\bar{i}$. We also note that r is feasible w.r.t $\bar{i}$ iff $r\rho$ is feasible w.r.t $\bar{i}\rho$, and hence

$$\mathbb{P}\left(\overline{R} = r \mid \overline{X} = \bar{i}\right) = 0 \iff \mathbb{P}\left(\overline{R} = r\rho \mid \overline{X} = \bar{i}\rho\right) = 0.$$

So it remains to check the equality of the conditional probabilities for a permutation r feasible to the t-tuple $\bar{i}$. In fact we can calculate the conditional probability explicitly.

Let $S = \{i_1, \ldots, i_t\}$ and for $s \in \mathbb{Z}_N$, let $B_s = \{k \in [t] \colon i_k = s\} \subseteq g^{-1}(s)$ (observe that $B_s = \emptyset \ \forall s \notin S$). If one views the indices $[t]$ as balls being thrown into the bins $\mathbb{Z}_N$, then S would be the set of bins occupied by $[t]$ and B_s represents balls among $[t]$ falling into bin s. For $s \in \mathbb{Z}_N$ define the event

$$E_s \;=\; \{\, \forall i, j \in B_s,\ \sigma_s(i) < \sigma_s(j) \iff r(i) < r(j) \,\} \;=\; \{\, \sigma_s \text{ permutes } B_s \text{ according to } r \,\}.$$

Note that the permutation σ_s is chosen uniformly at random from $S_{g^{-1}(s)}$, and hence there is $\frac{1}{|B_s|!}$ probability that the rank induced over the indices appearing in B_s is the same rank as the one induced by r. That is,

$$\mathbb{P}\left(E_s \mid \overline{X} = \bar{i}\right) = \frac{1}{|B_s|!}.$$

Note that assuming r is feasible w.r.t $\bar{i}$, we have $\overline{R} = r$ iff $\overline{R}$ and r induce the same order over all the entries of B_s for all $s \in S$. That is,

$$\left\{\overline{R} = r\right\} = \bigcap_{s \in S} E_s$$

conditioned on $\overline{X} = \bar{i}$.

Note that the permutations $\{\sigma_s \colon s \in S\}$ are chosen independently when conditioned on $\overline{X} = \bar{i}$ so $\{E_s\}_{s \in S}$ are independent and hence

$$\mathbb{P}\left(\overline{R} = r \mid \overline{X} = \bar{i}\right) = \mathbb{P}\left(\bigcap_s E_s \mid \overline{X} = \bar{i}\right) = \prod_{s \in S} \mathbb{P}\left(E_s \mid \overline{X} = \bar{i}\right) = \prod_{s \in S} \frac{1}{|B_s|!}.$$

Finally, we verify that the analogous computation for $\mathbb{P}\left(\overline{R} = r \mid \overline{X} = \bar{i}\rho\right)$ yields the same result. Let $S' = \left\{(\bar{i}\rho)_k \colon k \in [t]\right\}$; since ρ is a permutation, it follows that $S' = S$. Similarly letting $B'_s = \left\{k \in [t] \colon (\bar{i}\rho)_k = s\right\}$, this time we have

$$\mathbb{P}\left(\overline{R} = r \mid \overline{X} = \bar{i}\rho\right) = \prod_{s \in S' = S} \frac{1}{|B'_s|!}.$$

However it is clear that $|B_s| = |B'_s| \ \forall s \in \mathbb{Z}_N$, as $B'_s = (\rho^{-1})(B_s)$ (since ρ^{-1} is a bijection between the two sets). Therefore

$$\prod_{s \in S'} \frac{1}{|B'_s|!} = \prod_{s \in S} \frac{1}{|B_s|!}$$

which we argued earlier is sufficient to prove (7). ◁

▷ Claim 14. There is a constant $C > 0$ such that $|\Pi| \leq (Cn)^{35t}$.

Proof. It is clear that each map $g \in \mathcal{G}$ contributes exactly $|\Sigma| \cdot s = \tau!$ permutations to Π. Thus,

$$|\Pi| \;=\; \tau! \cdot |\mathcal{G}| \;\leq\; (32t)^{32t} \cdot |\mathcal{H}| \;\leq\; (32n)^{32t} \cdot N^t \;\leq\; (32n)^{32t} \cdot (2n^3)^t,$$

by (1). ◁

3 Conclusion and Future Work

Our algorithm for constructing Π runs in time $n^{O(t)}$, which is quite efficient size $|\Pi| = n^{O(t)}$. However, in applications often one is interested in sampling only a single permutation from Π. In this case, it may be unnecessary to construct the whole family. It is natural to ask if one can give a more explicit construction of t-rankwise independent families. That is, can a t-rankwise independent family Π of permutations of $[n]$ be constructed such that

- $|\Pi| \leq n^{O(t)}$, and
- sampling a single permutation from Π can be done in time $O(n)$?

We also re-emphasize that the problem of explicitly constructing a t-independent permutation family Π over $[n]$ with $|\Pi| \leq n^{O(t)}$ remains open. Such a construction would strengthen the results of this paper, as it would be a t-rankwise independent permutation family as well.

References

1. Andrei Z. Broder, Moses Charikar, Alan M. Frieze, and Michael Mitzenmacher. Min-wise independent permutations. *Journal of Computer and System Sciences*, 60(3):630–659, 2000. `doi:10.1006/jcss.1999.1690`.
2. J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18:143–154, 1979. `doi:10.1016/0022-0000(79)90044-8`.
3. Venkat Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory*, 2018. Manuscript.
4. Nicholas Harvey and Arvin Sahami. Explicit orthogonal arrays and universal hashing with arbitrary parameters. In *Proceedings of the ACM Symposium on Theory of Computation (STOC)*, 2024.
5. Toshiya Itoh, Yoshinori Takei, and Jun Tarui. On permutations with limited independence. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '00, pages 137–146, USA, 2000. Society for Industrial and Applied Mathematics.
6. Enrico Iurlano. Growth of the perfect sequence covering array number. *Des. Codes Cryptography*, 91(4):1487–1494, December 2022. `doi:10.1007/s10623-022-01168-3`.
7. Greg Kuperberg, Shachar Lovett, and Ron Peled. Probabilistic existence of rigid combinatorial structures. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 1091–1106, 2012. `doi:10.1145/2213977.2214075`.
8. Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.
9. Salil P. Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012. `doi:10.1561/0400000010`.
10. Raphael Yuster. Perfect sequence covering arrays. *Des. Codes Cryptography*, 88(3):585–593, March 2020. `doi:10.1007/s10623-019-00698-7`.

A Omitted proofs

Proof of Corollary 8. For notational convenience, let $X_i = |g^{-1}(i)|$ for $i \in [N]$. Observe that $n = \sum_{i \in [N]} X_i$ and $|g([n])| = \sum_{i \in [N]} 1_{\{X_i \geq 1\}}$. Then we may write

$$2 \cdot \big(n - |g([n])|\big) \;=\; 2 \sum_{i \in [N]} \Big(\underbrace{X_i - 1_{\{X_i \geq 1\}}}_{=0 \text{ if } X_i \in \{0,1\}} \Big) \;=\; \sum_{i \in [N]} \underbrace{1_{\{X_i \geq 2\}} \cdot 2(X_i - 1)}_{\geq 1_{\{X_i \geq 2\}} \cdot X_i}$$

$$\geq \sum_{i \in [N]} 1_{\{X_i \geq 2\}} \cdot X_i \;=\; |\mathcal{X}|.$$

Thus, by Lemma 7, $|\mathcal{X}| \leq 2 \cdot \big(n - |g([n])|\big) \leq 2 \cdot (16t) = 32t$. ◀

Proof of Theorem 11. Observe that

$$1_{\{\sum_{i=1}^n Y_i \geq (1+\delta)\mu\}} \;\leq\; \exp\Big(\lambda \sum_{i=1}^n Y_i - \lambda(1+\delta)\mu\Big)$$

and hence taking expectations implies

$$\mathbb{E}1_{\{\sum_{i=1}^n Y_i \geq (1+\delta)\mu\}} \;=\; \mathbb{P}\left(\sum_{i=1}^n Y_i \geq (1+\delta)\mu\right) \;\leq\; \mathbb{E}\exp\Big(\lambda \sum_{i=1}^n Y_i - \lambda(1+\delta)\mu\Big).$$

Next, as shown in [8, Theorem 4.1 and its proof], letting $\lambda = \ln(1+\delta)$, we have the inequality

$$\mathbb{E}\exp\Big(\lambda \sum_{i=1}^n Y_i - \lambda(1+\delta)\mu\Big) \;\leq\; \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

It remains to prove that

$$\left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \;\leq\; (1+\delta)^{-(1+\delta)\mu/4} \quad \forall \delta \geq 1.$$

As $0 \leq \mu$, it suffices to show

$$\frac{e^\delta}{(1+\delta)^{1+\delta}} \;\leq\; (1+\delta)^{-(1+\delta)/4} \quad \forall \delta \geq 1.$$

After taking logs and performing simple algebraic manipulations, we arrive at another equivalent inequality

$$\frac{4}{3} \leq (1+\tfrac{1}{\delta})\ln(1+\delta) \quad \forall \delta \geq 1.$$

For $x \geq 0$, let $f(x) = (1+\frac{1}{x})\ln(1+x)$. We note that

$$f'(x) = \frac{x - \ln(1+x)}{x^2} \geq 0 \quad \forall x > 0$$

since $\ln(x+1) \leq x \quad \forall x > 0$. Thus in particular f is non-decreasing over $[1,\infty)$ and hence

$$(1+\frac{1}{\delta})\ln(1+\delta) = f(\delta) \geq f(1) = 2\ln(2) > \frac{4}{3} \quad \forall \delta \geq 1$$

as desired. ◀