A Connection Between Random Variables and Latin k-cubes

Ruben Michel[*] Gadi Taubenfeld[†] Andrew Berman[‡]

September 28, 1993

Abstract

The subject of latin squares is about 200 years old, and it abounds with many solved and unsolved problems. In this paper we establish an interconnection between latin k-cubes and random variables. When combined with the rich theory of latin k-cubes, this connection yields new results about independent random variables, which generalize and extend other recent results. Our results are applicable for the construction of efficient algorithms.

Keywords: Random variables, latin squares, randomized algorithms.

1 Introduction

Randomized algorithms have been extensively studied in the literature, since they are often faster and easier to analyze than deterministic ones [17, 22, 24]. A problem that has recently attracted attention in the computer science community is how to derandomize polynomial time randomized algorithms while keeping their time complexity polynomial [1, 17, 20, 6].

The derandomization procedure used in most references is based on the following observation: Randomized algorithms typically use a sequence of *independent* random variables, e.g., several coin tosses. In many instances the algorithm will exhibit similar performance if the random variables are k*-wise independent* rather than fully independent (i.e., every k random variables are independent). The size of the sample space required to construct the k-wise independent sequence is exponentially smaller than the size of the space required for the fully independent sequence. This fact is exploited in the derandomization procedure as follows: Rather than applying the probabilistic algorithm on a random independent sequence, the probabilistic algorithm is run on all k-wise independent sequences, thereby yielding a deterministic algorithm with polynomial (albeit high degree) time complexity.

Recently, [5, 2, 21] have introduced a slightly different derandomization procedure using ϵ-biased random variables which is not examined here.

The motivation for this paper is the following elementary question: What is the *longest* sequence of ℓ-wise independent uniformly distributed random variables assuming values in Z_n that can be generated with just k such variables? Here we say that the random variables $W_1, \cdots, W_k$ assuming values in Z_n *generate* the random variables $X_1, \cdots, X_m$ assuming values in Z_n if there exist functions $f_1, \cdots, f_m$ of Z_n^k into Z_n such that $X_i = f_i(W_1, \cdots, W_k)$, for $1 \leq i \leq m$.

This naive question leads to a surprising connection, which is the main result reported here: The problem of constructing a set of independent uniformly distributed random variables out of few such variables is closely and constructively related to the problem of constructing a set of orthogonal latin hypercubes. Specifically, we prove that there exists a set of $t + 2$ pairwise independent uniformly distributed random variables assuming values in Z_n, which are generated from just 2 such variables,

[*] Computer Science and Information Science Department, Univ. of Massachusetts at Amherst, Amherst, MA 01003.
[†] AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974. Email: gadi@research.att.com
[‡] Computer Science Department, University of Washington, Seattle, WA 98195.

if and only if there exists a set of size t of orthogonal latin squares of order n. We prove a similar result also for higher dimensions.

The problem of constructing the biggest set of orthogonal k-cubes has been thoroughly studied, see [3, 4, 7, 8, 9, 13, 14, 18, 19, 25] and others. Thus, any of the known constructions of orthogonal latin k-cubes can be readily modified to yield a set of k-wise independent uniformly distributed random variables.

The case $k = 2$, i.e., constructing uniform pairwise independent random variables, is particularly salient. Here the problem of constructing orthogonal latin cubes, which in the two dimensional case are called latin squares, is a classic problem studied by the great mathematician Euler some 200 years ago and by many others thereafter. It is known that for prime p there exist $p - 1$ orthogonal latin squares and no more. Thus, using the connection reported in this paper, we conclude that two uniform random variables that assume values in Z_p can generate $p + 1$ such pairwise independent random variables and no more, thereby providing tight upper and lower bounds. This result improves on [15, 10] which construct just p such variables. Interestingly, their construction is a special case of a well-known construction of orthogonal latin squares (see, [16] page 366).

2 Independent random variables and latin k-cubes

2.1 Basic definitions

A set $\{X_1, ..., X_m\}$ of random variables is an *ℓ-wise independent set* if every subset of ℓ random variables is mutually independent. We address the following question: What is the biggest set of ℓ-wise independent uniformly distributed random variables assuming values in Z_n that can be generated with just k such variables, as a function of ℓ and k? We concentrate on the case $\ell = k$.

Let S be a set of n elements. A *latin square of order n based on S* is a square $A = [a_{ij}]$, where $i, j = 1, ..., n$, with the requirement that each row or column is an n-permutation of elements of S. We will use "latin square of order n" as a shorthand for "latin square of order n based on Z_n". Let $A_1 = [a^1_{ij}]$ and $A_2 = [a^2_{ij}]$ denote two latin squares of order $n \geq 3$. The squares A_1 and A_2 are called *orthogonal* provided that the n^2 2-samples (a^1_{ij}, a^2_{ij}), where $i, j = 1, ..., n$, are distinct.

The problem of finding orthogonal latin squares of a given order is a classical problem in combinatorial mathematics. Euler conjectured in 1782 that there exists no pair of orthogonal latin squares of order $n \equiv 2 \pmod 4$. Tarry around 1900 verified the validity of Euler's conjecture for $n = 6$ [26]. Bose, Shrikhande, and Parker in 1959 proved the falsity of Euler's conjecture [8]. They showed that for any n where $n > 2$ and $n \neq 6$, there exists a pair of orthogonal latin squares of order n.

Next we define the natural generalization of the notion of latin squares to higher dimensions. A *latin k-cube of order n based on S* is a hypercube $A = [a_{i_1,...,i_k}]$, where $i_j = 1, ..., n$ for all $1 \leq j \leq k$, with the requirement that the values that occupy positions $(i_1, ..., i_k)$ and $(i'_1, ..., i'_k)$ are different if $(i_1, ..., i_k)$ and $(i'_1, ..., i'_k)$ differ at exactly one coordinate. This definition is equivalent to saying that each element appears exactly once along each axis. As in the case of latin squares, we use "latin k-cube of order n" as a shorthand for "latin k-cube of order n based on Z_n". We note that the definition of latin k-cube is similar to that of a k-dimensional permutation cube [13].

Let $A_\ell = [a^\ell_{i_1,...,i_k}]$ where $1 \leq \ell \leq k$, denote k latin k-cubes of order $n \geq 3$. $A_1, ..., A_k$ are *orthogonal* provided that the n^k k-samples $(a^1_{i_1,...,i_k}, ..., a^k_{i_1,...,i_k})$, where $i_j = 1, ..., n$ for all $1 \leq j \leq k$, are distinct. A set of orthogonal latin k-cubes of order n is a set of latin k-cubes of order n where every k of its members are orthogonal.

2.2 Known results about latin k-cubes

We list below three useful results about latin k-cubes.

1. Let $n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_N^{\alpha_N}$ be the prime power decomposition of an arbitrary positive integer n, where the p_i are distinct primes and the α_i are positive integers. Let $\beta = \min(p_i^{\alpha_i} - 1)$ where

$i = 1, ..., N$. If $\beta \geq 2$, then for any $k \leq \beta$ there exists a set of β orthogonal latin k-cubes of order n.

2. There exist k orthogonal latin k-cubes of order n, for every $n > 2$, $n \neq 6$.

3. The size of any set of orthogonal latin k-cubes of order n, is at most $(n-1)(k-1)$.

The first result establishes a lower bound on the size of any maximal set of orthogonal latin k-cubes of order n. This result was first proved in [18] for the case of $k = 2$, and was then generalized for any $k \leq \beta$ in [13, 4]. The second result was proved by Arkin and Straus [4] establishing that if there exists two orthogonal latin squares of order n then there exists k orthogonal latin k-cubes of order n for each $k > 2$. Since, as mentioned, there are orthogonal latin squares of every order $n > 2$, $n \neq 6$, the second result follows. The third result is proved in [14]. For other results about latin k-cubes see [11, 12].

3 Main results

In this section we establish an interconnection between k-wise independent sets and sets of orthogonal latin k-cubes, and use it to prove new results about the maximal size of k-wise independent sets. Since there does not exist a pair of orthogonal latin squares of order 2, we will assume that $n \geq 3$.

3.1 The Connection Theorem

The Connection Theorem

1. *For $n \geq 3$ and $t \geq 2$, there exists a pairwise independent set of $t+2$ uniformly distributed random variables assuming values in Z_n which is generated by two independent uniformly distributed random variables assuming values in Z_n* **if and only if** *there exists a set of size t of orthogonal latin squares of order n.*

2. *For $n \geq 3$, $k > 2$ and $t \geq k$, there exists a k-wise independent set of $t+k$ uniformly distributed random variables assuming values in Z_n which is generated by k mutually independent uniformly distributed random variables assuming values in Z_n* **only if** *there exists a set of size t of orthogonal latin k-cubes of order n.*

3. *For $n \geq 3$, $k > 2$ and $t \geq k$, there exists a k-wise independent set of t uniformly distributed random variables assuming values in Z_n which is generated by k mutually independent uniformly distributed random variables assuming values in Z_n* **if** *there exists a set of size t of orthogonal latin k-cubes of order n.*

One would have hoped that the converse proposition of the second part of the theorem also hold. Unfortunately, it does not. To prove the connection Theorem we need the following lemmas.

Lemma 1 *For $n \geq 3$ and $t \geq k$, there exists a k-wise independent set of t uniformly distributed random variables assuming values in Z_n which is generated by k mutually independent uniformly distributed random variables assuming values in Z_n* **if and only if** *there exists an n^k by t array $A = [a_{ij}]$, where $i = 1, ..., n^k$ and $j = 1, ..., t$. The entries of A are elements in Z_n and in every k columns each k-tuple appears exactly once.*

Proof: Let the array A be given. For any $1 \leq \ell \leq t$, we define a function called f_ℓ, of Z_n^k into Z_n as follows. Let $r = (\sum_{j=1}^{k} i_j n^{k-j}) + 1$, then for any $(i_1, ..., i_k) \in Z_n^k$, $f_\ell(i_1, ..., i_k) = a_{r\ell}$. Starting with any k mutually independent uniformly distributed random variables $W_1, \cdots, W_k$ assuming values in Z_n, we can generate t random variables $X_\ell = f_\ell(W_1, \cdots, W_k)$ where $1 \leq \ell \leq t$. Because of the assumptions on A, $\{X_1, ..., X_t\}$ is a k-wise independent set of uniformly distributed random variables assuming values in Z_n.

To prove the converse proposition, let $\{X_1, \cdots, X_t\}$ be a k-wise independent set of t uniformly distributed random variables assuming values in Z_n which is generated by some k mutually independent uniformly distributed random variables $W_1, \cdots, W_k$ assuming values in Z_n. By definition, there must exist functions $f_1, ..., f_t$ of Z_n^k into Z_n where $X_\ell = f_\ell(W_1, \cdots, W_k)$ for all $1 \leq \ell \leq t$ (which are k-wise independent and uniformly distributed over Z_n).

We define the array A as follows. For any $(i_1, ..., i_k) \in Z_n^k$ and any $1 \leq \ell \leq t$, the value $f_\ell^k(i_1, ..., i_k)$ occupies position (r, ℓ) in A, where r and ℓ are given as above. It is not difficult to see that for every k numbers $b_1, ..., b_k \in Z_n$, the set of k equations $\{f_i^k(a_1, ..., a_k) = b_i\}_{i=1}^k$ has a unique solution. Hence, the array A is as required. ▮

Our next lemma is adapted, with minor changes, from [23] (Theorem 1.3, page 82).

Lemma 2 *For $n \geq 3$ and $t \geq 2$, there exists a set of size t of orthogonal latin squares of order n* **if and only if** *there exists an n^2 by $t + 2$ array $A = [a_{ij}]$, where $i = 1, ..., n^2$ and $j = 1, ..., t + 2$. The entries of A are elements in Z_n and in each two columns each pair appears exactly once.*

Proof: Let the array A be given. We permute the rows of A so that the entries in the first two columns are in lexicographic order $(0, 0)$,...,$(n - 1, n - 1)$. Let A' be the array resulting from permuting the rows of A as described above. For each $\ell = 3, ..., t+2$, we define an n by n array A_ℓ where the entry $A_\ell[i, j]$ equals the entry $A'[(i - 1)n + j, \ell]$. Column 1 of A' implies that A_ℓ does not have two equal entries in a row, and column 2 of A' implies that A_ℓ does not have two equal entries in a column, and hence A_ℓ is a latin square. Also any two latin squares are orthogonal because of the structure of the columns of A. Thus, $\{A_3, ..., A_{t+2}\}$ is a set of orthogonal latin squares of order n and size t.

The converse proposition is proved similarly. Let $\{A_3, ..., A_{t+2}\}$ be a set of orthogonal latin squares of order n and size t. We define the array A as follows. The first two columns are in lexicographic order $(0, 0)$,...,$(n - 1, n - 1)$, and for each $\ell = 3, ..., t + 2$, the entry $A[(i - 1)n + j, \ell]$ equals the entry $A_\ell[i, j]$. ▮

Lemma 3 *For $n \geq 3$ and $t \geq k$, there exists a set of size t of orthogonal latin k-cubes of order n* **if** *there exists an n^k by $t + k$ array $A = [a_{ij}]$, where $i = 1, ..., n^k$ and $j = 1, ..., t + k$. The entries of A are elements in Z_n and in each k columns each k-tuple appears exactly once.*

Proof: Let the array A be given. We permute the rows of A so that the entries in the first k columns are in lexicographic order $(0, ..., 0)$,...,$(n - 1, ..., n - 1)$. For each $l = k + 1, ..., k + t$, we define a hypercube A_l of dimension k and of order n as follows. Let $r = (\sum_{j=1}^k i_j n^{k-j}) + 1$, then for any $(i_1, ..., i_k) \in Z_n^k$, the value in position (r, l) in A after permuting the rows as described above, occupies position $(i_1, ..., i_k)$ in A_l. It is not difficult to see that $\{A_{k+1}, ..., A_{t+k}\}$ is a set of orthogonal latin k-cubes of order n and size t, since the assumptions on A are such that each hypercube is a latin k-cube of order n. ▮

Lemma 4 *For $n \geq 3$ and $t \geq k$, there exists a set of size t of orthogonal latin k-cubes of order n* **only if** *there exists an n^k by t array $A = [a_{ij}]$, where $i = 1, ..., n^k$ and $j = 1, ..., t$. The entries of A are elements in Z_n and in each k columns each k-tuple appears exactly once.*

Proof: Let $\{A_1, ..., A_t\}$ be a set of orthogonal latin k-cubes of order n and size t. The array A is defined as follows. Let $r = (\sum_{j=1}^k i_j n^{k-j}) + 1$. Then for any $(i_1, ..., i_k) \in Z_n^k$ and any $1 \leq \ell \leq t$, the value in position (r, ℓ) in A equals the value that occupies position $(i_1, ..., i_k)$ in A_ℓ. The requirements on A follows from the assumptions about $\{A_1, ..., A_t\}$. ▮

Unfortunately, adding k more columns, as is done for $k = 2$ in Lemma 2, does not work when $k > 2$. The first part of the Connection Theorem follows from Lemma 1 and Lemma 2. The second part follows from Lemma 1 and Lemma 3. The third part follows from Lemma 1 and Lemma 4.

We point out that a connection similar to that stated in the Connection Theorem exists also between sets of orthogonal linear latin k-cubes and sets of k-wise independent linear random variables.

3.2 Consequences of the Connection Theorem

The following corollaries are immediate consequences of The Connection Theorem and the observations of the previous section.

1. Let $n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_N^{\alpha_N}$ be the prime power decomposition of an arbitrary positive integer n, where the p_i are distinct primes and the α_i are positive integers. Let $\beta = \min(p_i^{\alpha_i} - 1)$ where $i = 1, ..., N$. If $\beta \geq 2$, then for any $k \leq \beta$ there exists a set of β k-wise independent uniformly distributed random variables assuming values in Z_n generated by k independent uniformly distributed random variables assuming values in Z_n. Furthermore, when $\beta = 2$, there exists a set of $\beta + 2$ pairwise independent uniformly distributed random variables assuming values in Z_n which is generated by two independent uniformly distributed random variables assuming values in Z_n.

2. The size of any k-wise independent set of uniformly distributed random variables assuming values in Z_n, which is generated by k independent uniformly distributed random variables assuming values in Z_n, is at most $n(k-1)+1$.

Notice that when $k = 2$ and n is a power of a prime the above upper and lower bounds are tight.

A *projective plane* consists of a set of elements called *points* and a set of elements called *lines*, with a relation connecting them such that: (1) each two points belong to exactly one line, (2) each two lines have in common exactly one point, and (3) there are at least four points no three of which belong to the same line. When such a plane has a finite number of $n + 1$ points on every line, it is called a *finite* projective plane of order n. In interesting result is that: Every projective plane of order $n \geq 3$ defines, and is defined by, a set of $n - 1$ orthogonal latin squares of order n [7, 25]. This result is a special case of general theorem which involves geometric k-nets ([11] page 270). This result and the Connection Theorem implies the following:

- Every finite projective plane of order $n \geq 3$ defines, and is defined by, a set of $n + 1$ pairwise independent uniformly distributed random variables assuming values in Z_n, which is generated by two pairwise independent uniformly distributed random variables assuming values in Z_n.

It should be possible to extend this last connection to higher dimensions as is done in the Connection Theorem. There exist several impossibility results for finite projective planes. For example, if n is a positive integer congruent to 1 or 2 modulo 4, there cannot exist any finite projective plane of order n unless n can be expressed as a sum of two integral squares, $n = a^2 + b^2$ [9]. Consequently, for such n there does not exist a set of $n + 1$ pairwise independent uniformly distributed random variables assuming values in Z_n, which is generated by two pairwise independent uniformly distributed random variables assuming values in Z_n.

Acknowledgments We thank Prabhakar Raghavan and Tuvi Ezion for helpful discussions.

References

[1] N. Alon, L. Babai, and A. Itai. A fast and simple randomized parallel algorithm for the independent set problem. *Journal of algorithms*, 7:567–583, 1986.

[2] N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3:289–304, 1992.

[3] J. Arkin, V.E. Hoggatt, and E.G. Straus. Systems of magic latin k-cubes. *Canad. Jour. Math.*, 28:1153–1161, 1976.

[4] J. Arkin and E.G. Straus. Latin k-cubes. *Fibonacci Quarterly*, 12:288–292, 1974.

[5] R. Ben-Nathan. On dependent random variables over small spaces. M.Sc. thesis, Hebrew University, Jerusalem, Israel, February 1990.

[6] B. Berger and J. Rompel. Simulating $(\log^c n)$-wise independence in NC. In *Proc. 30th IEEE Symp. on Foundations of Computer Science*, pages 2–7, 1989.

[7] C. Bose. On the application of properties of galois fields to the problem of construction of hyper-graeco latin squares. *Sankhyā*, 3:323–386, 1936.

[8] C. Bose, S. S. Shrikhande, and E.T. Parker. Further results on the construction of mutually orthogonal latin squares and the falsity of Euler's conjecture. *Canad. Jour. Math.*, 12:189–203, 1960.

[9] R.H. Bruck and H.J. Ryser. The nonexistence of certain finite projective planes. *Canad. Jour. Math.*, 1:88–93, 1949.

[10] B. Chor and O. Goldreich. On the power of two-point sampling. *Journal of Complexity*, 5:96–106, 1989.

[11] J. Denes and A.D. Keedwell. *Latin squares and their applications.* Academic Press, 1974.

[12] J. Denes and A.D. Keedwell. *Latin squares: New developments in the theory and applications.* Elsvier science publishers, 1991.

[13] A. Heppes and P. Révész. A new generalization of the method of latin squares and orthogonal latin squares and its application to the design of experiments. *Magyar Tud. Akad. Mat. Int. Közl.*, 1:376–390, 1956. Hungarian.

[14] L. Humblot. Sur une extension de la notion de carrés latin. *C. R. Acad. Sci. Paris*, 273:795–798, 1971. MR 44(1972)#3892.

[15] A. Joffe. On a sequence of almost deterministic pairwise independent random variables. *Proc. of the American Mathematical Society*, 29:381–382, 1971.

[16] C. Liu. *Introduction to Combinatorial Mathematics.* McGraw Hill, Inc., 1968.

[17] M. Luby. A simple parallel algorithm for maximal independent set problem. In *Proc. 17nd ACM Symp. on Theory of Computing*, pages 1–10, 1985.

[18] H.F. MacNeish. Euler squares. *Ann. Math.*, 23:221–227, 1922.

[19] H.B. Mann. The construction of orthogonal latin squares. *Ann. Math. Stat.*, 13:418–423, 1942.

[20] R. Motwani, J. Naor, and M. Naor. The probabilistic method yields deterministic parallel algorithms. *Proc. 30th IEEE Symp. on Foundations of Computer Science*, pages 8–13, 1989.

[21] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. In *Proc. 22nd ACM Symp. on Theory of Computing*, pages 213–223, 1990.

[22] M. Rabin. Probabilistic algorithms for testing primality. *Journal of Number Theory*, 12:128–138, 1980.

[23] Herbert J. Ryser. *Combinatorial Mathematics.* The Mathematical Association of America, 1963.

[24] R. Solovay and V. Strassen. Fast monte-carlo test for primality. *SIAM Journal of Computing*, 6:84–85, 1977.

[25] W.L. Stevens. The completely orthogonalized latin squares. *Ann. Eugen.*, 9:82–93, 1939.

[26] G. Tarry. Le problème de 36 officieurs. *Compte Rendu de l'Association Francaise pour l'Avancement de Science Naturel*, 1:122–123, 1900. **2** (1901), 170-203.