

THE SUBGROUP ALGORITHM FOR GENERATING UNIFORM RANDOM VARIABLES

PERSI DIACONIS

*Department of Statistics
Stanford University
Stanford, California*

MEHRDAD SHAHSHAHANI

*Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California*

We suggest a simple algorithm for Monte Carlo generation of uniformly distributed variables on a compact group. Examples include random permutations, Rubik's cube positions, orthogonal, unitary, and symplectic matrices, and elements of GL_n over a finite field. The algorithm reduces to the "standard" fast algorithm when there is one, but many new examples are included.

1. INTRODUCTION

We begin with the problem of generating uniformly distributed random permutations of n objects. This arises in generating random bridge deals, in Monte Carlo investigation of rank tests, and elsewhere. Assume we have available a source of independent uniformly distributed random variables U_1, U_2, \dots , on $[0,1]$, and wish to convert them into a stream of permutations.

The most frequently used algorithm involves a factorization of the uniform distribution. Informally, begin with n registers containing the numbers 1 to n in order. At stage i , a random integer J_i is chosen between i and n and the i th

and J_i 's registers have their contents switched. This continues from $1 \leq i < n$. It will be argued below that this produces a random permutation. Further discussion of this algorithm is in Knuth [28, p. 139–141]. Knuth attributes the algorithm to Moses and Oakford [32]. Moses (personal communication) found a similar idea in Fisher and Yates [15, p. 20]. Other discussions of this subject appear in Nijenhuis and Wilf [33] and Devroye [6, Chapters 9 and 13].

The following algorithm abstracts the idea to any finite group. See Herstein [23] for any unexplained group theoretic terminology.

Subgroup Algorithm. Let G be a finite group. Let $G_0 = G \supset G_1 \supset \cdots \supset G_r$ be a nested chain of subgroups (not necessarily normal). Let C_i be coset representatives for G_{i+1} in G_i , $0 \leq i < r$. Clearly G can be represented as

$$G \cong C_0 \times C_1 \times \cdots \times C_{r-1} \times G_r$$

in the sense that each $g \in G$ has a unique representation as $g_0 g_1 \cdots g_r$ with $g_i \in C_i$ and $g_r \in G_r$. It follows that if the g_i are chosen uniformly at random in their respective domains and multiplied together, the resulting product element g will be uniformly distributed on G .

Example 1 (The symmetric group): Take $G = S_n$, the group of permutations of n letters. Consider the chain

$$S_n \supset S_{n-1} \supset S_{n-2} \supset \cdots \supset \{\text{id}\}, \quad (1.1)$$

with S_{n-i} the set of permutations of n letters that fix the first i letters. Coset representatives C_i for $S_{n-(i+1)}$ in S_{n-i} are the identity and the set of transpositions (i, j) , $i < j \leq n$. In this case, the subgroup algorithm becomes the usual algorithm. Choosing the cycles $(j, j-1, \dots, i+1, i)$, $i \leq j \leq n$ as coset representatives gives the Fisher-Yates algorithm.

It is instructive to note that many variants are possible. Consider replacing the chain in Eq. (1.1) by

$$S_n \supset S_{n-2} \supset S_{n-4} \supset \cdots \supset \{\text{id}\}. \quad (1.2)$$

Coset representations for $S_{n-2(i+1)}$ in S_{n-2i} are permutations of the form

$$(2i+1, j)(2i+2, k)$$

where $2i+1 \leq j \leq n$ and $2i+1 \leq k \neq j \leq n$. These coset representatives may be ordered setting up a 1-1 correspondence between them and the numbers $1, 2, \dots, (n-2i)(n-2i-1)$. This would have the advantage of making fewer calls to the random number generator and the disadvantage of more book-keeping.

It is instructive to interpret an extreme case—the trivial decomposition

$$S_n \supset \{\text{id}\}. \quad (1.3)$$

There are $n!$ coset representatives. The subgroup algorithm can be carried out by choosing a random integer less than $n!$ and setting up a 1-1 correspondence

One way to proceed uses the factorial number system: every number between 0 and $n! - 1$ can be uniquely represented as

$$a_1 + a_2 2! + \cdots + a_{n-1} (n-1)! \quad \text{with } 0 \leq a_i \leq i.$$

The "digits" a_i are determined as the remainders on successive division by 2, 3, \dots , $n-1$. Devroye [6] gives history and variations.

These digits can be used to represent a permutation in a variety of ways. One simple method uses the unique representation of a permutation as a product of transpositions through Eq. (1.1): $(1, n - a_{n-1})(2, n - a_{n-2}) \cdots (n-1, n - a_1)$.

There does not appear to be *any* practical difference between algorithms based on Eqs. (1.1), (1.2), or (1.3): the savings in calls to the random number generator are balanced by steps like determining the digits a_i .

Example 2 (The alternating group): Many combinatorial problems involve subgroups of the permutation group. For example, in the still popular 15 puzzle, square tiles bearing the numbers 1, 2, \dots , 15 are placed into a 4×4 grid, leaving one empty square. The tiles are to be slid around to bring them to a given order.

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	BLANK

It is well known (see, e.g., Gilbert [17, p. 81], or Rouse-Ball [36, p. 299]) that the group of possible arrangements of the puzzle pieces is the alternating group A_{15} —the even permutations in S_{15} .

One way to choose a random element in A_n uses the chain of subgroups

$$A_n \supset A_{n-1} \supset A_{n-2} \supseteq \cdots \supseteq A_3 \supseteq \{\text{id}\},$$

where, for example, A_{n-1} is the set of even permutations fixing 1. Coset representatives for A_{n-i-1} in A_{n-i} may be chosen as

$$\text{id}, \quad (i, j, j+1) \quad 1 < j < n, \quad \text{and } (i, n, n-1).$$

Another method of generating a random element of A_n is to run the subgroup algorithm based on Eq. (1.1), generating a random element of S_n . Keep track of how many times the identity is chosen. This determines if the chosen permutation is even or odd. If even, keep it, if odd, make one final transposition (say (1 2)).

These methods can be used with a real puzzle, provided the pieces are removable. With nonremovable pieces, the following algorithm works: Fill up the first row and column with randomly chosen pieces. This leaves a 3×3 puzzle. Fill its top row and first column with random choices. Finally, choose a random "cycle" of the remaining 3 pieces.

Example 3 (Rubik's cube, etc.): The popular puzzle Rubik's cube (see Singmaster [38] or Eidswick [14]) and its many variants offer other finite groups. Usually, one is given a set of basic generators (e.g., the allowable twists) and not even told the group. Fortunately, there is a polynomial time algorithm due to Sims [37] that takes a set of generators of a permutation group G on n letters and finds explicit coset representatives for the tower

$$G = G_n \supset G_{n-1} \supset G_{n-2} \cdots \supset \{\text{id}\},$$

where $G_{n-1} = G \cap S_{n-1}$. We have carried this computation out for the Rubik's cube puzzle. We are thus in a position to generate a random starting position, should anyone want to run a cube contest.

Sims' algorithm is nicely explained by Furst, Hopcroft, and Luks [16]. It is implemented in the algebraic program Cayley.

The remainder of this article treats further applications of the subgroup algorithm. Section 2 treats GL_n with coefficients in a finite field. Section 3 treats the orthogonal group, discussing and comparing a variety of algorithms. Section 4 develops some general theory and applies it to the unitary and symplectic groups.

2. $GL_n(\mathbb{F}_q)$

Let \mathbb{F}_q be a finite field of q elements. Let V_n be a vector space of dimension n over \mathbb{F}_q . Let GL_n be the set of $n \times n$ invertible matrices with coefficients in \mathbb{F}_q . Then GL_n is a finite group of order

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}). \quad (2.1)$$

A familiar example has $q = 2$, so V_n is the binary n -tuples and GL_n consists of $n \times n$ invertible binary matrices, all operations occurring mod 2.

One obvious approach to generating a random element $\alpha \in GL_n$ is to choose each of the n^2 entries at random, and then check if α is invertible. This last step can be done using Gaussian elimination. The algorithm we know requires an order of n^3 operations.

The chance that a randomly chosen matrix is in GL_n is the ratio of Eq. (2.1) to q^{n^2} —the total number of matrices. This is $\prod_{i=1}^n \left(1 - \frac{1}{q^i}\right)$.

When $q = 2$ and n is larger than 10, this is about .29. The probability of a successful first attempt increases to 1 with increasing q . This rejection algorithm may be judged satisfactory, even for $q = 2$, because of simplicity.

To apply the subgroup algorithm, a decreasing chain of subgroups is

$$GL_n \supset P_n \supset GL_{n-1}. \quad (2.2)$$

In Eq. (2.2) P_n consists of $\{\alpha \in GL_n: \alpha e_1^t = a_1 e_1^t\}$ with e_1 the row vector in V_n containing a 1 in the first coordinate and zeros elsewhere, and a_1 is a nonzero element of \mathbb{F}_q . Thus, a typical entry $\alpha \in P_n$ looks like

$$\alpha = \begin{pmatrix} a_1 & | & x & & \\ \hline 0 & & & & \\ \vdots & & & & \\ 0 & & & & \alpha_1 \end{pmatrix} \quad (2.3)$$

$x \in V_{n-1}$, and $\alpha_1 \in GL_{n-1}$.

The cardinality of P_n is thus

$$|P_n| = (q-1)q^{n-1}(q^{n-1}-1) \cdots (q^{n-1}-q^{n-2}). \quad (2.4)$$

To understand P_n , it is useful to observe that GL_n acts transitively on projective space \mathbb{P}^{n-1} —the lines in V_n . The subgroup of matrices fixing the line through e_1 is P_n . Thus

$$GL_n/P_n \cong \mathbb{P}^{n-1}. \quad (2.5)$$

Coset representatives for P_n in GL_n can be chosen by enumerating lines as:

	# representatives of this type
(00...1) - 1	
(0...1*) - q	
⋮	
(1*...*) - q^{n-1}	(2.6)

Here * denotes any element of \mathbb{F}_q . It follows that

$$|\mathbb{P}^{n-1}| = 1 + \dots + q^{n-1}.$$

This agrees with Eqs. (2.1), (2.4), and (2.5).

It remains to identify the lines in Eq. (2.6) with elements in GL_n . This requires choosing, for any x in Eq. (2.6), a matrix α_x such that

$$\alpha_x e_1^t = x^t.$$

Over the real field, α_x can be chosen as a Householder reflection (see Section 3). In the case of a finite field, one simple choice follows: suppose x can be written

$$(0 \quad 0 \quad 1 \quad \dots \quad x \quad \dots)$$

Then α_x can be taken as

$$\begin{pmatrix} O_1 & & & & I_1 \\ \hline 1 & & & & \\ x_{n-l+1} & & & & \\ \vdots & \ddots & & & \\ x_n & & & 1 & \\ \hline & & & & O_2 \end{pmatrix} \quad (2.7)$$

Here O_1 is an $n - (l + 2) \times (l + 2)$ matrix of zeros, I_1 is an $n - (l + 2) \times n - (l + 2)$ identity, the bottom left block is $(l + 1) \times (l + 1)$ with ones down the diagonal, the first column as specified, and zeros elsewhere. Finally O_2 is a $(l + 2) \times n - (l + 2)$ matrix of zeros.

The matrix α_x is nonsingular and satisfies $\alpha_x e_i^t = x^t$. Moreover, because of its sparse structure, it is possible to compute the product $\alpha_x y^t$ (for $y \in V_n$) in the order of n operations. Hence, for $\beta \in GL_n$, $\alpha_x \beta$ can be computed in n^2 operations.

To complete the description of a general stage of the algorithm, it is necessary to choose coset representatives for GL_{n-1} in P_n . Embed GL_{n-1} in P_n by

$$\alpha_1 \rightarrow = \begin{pmatrix} 1 & 0 & & 0 \\ & 0 & & \\ & & \alpha_1 & \\ & 0 & & 0 \end{pmatrix} \quad \text{for } \alpha_1 \in GL_{n-1} .$$

From Eq. (2.2), coset representatives are given by

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} \quad (2.8)$$

with a_{11} nonzero and a_{1j} arbitrary. Again, if α has the form of Eq. (2.8), and $\beta \in GL_n$, then $\alpha \beta$ can be computed in n^2 operations.

Combining the two stages gives a reasonably simple way of choosing a random element $\alpha \in GL_n$, given the choice of a random element $\alpha_1 \in GL_{n-1}$:

Algorithm. Given α_1 uniformly distributed in GL_{n-1} , this algorithm outputs α uniformly distributed in GL_n .

1. Choose a_{11} at random in $\mathbb{F}_q - \{0\}$, choose a_{1j} independently, at random in \mathbb{F}_q . Form the matrix A as in Eq. (2.8).
2. Choose an integer l , $0 \leq l \leq n-1$ from the distribution

$$P(l) = q^l(q-1)/(q^n-1).$$
3. Given l , choose $l+1$ elements x_{n-l}, \dots, x_n independently at random in \mathbb{F}_q . Form the matrix X as in Eq. (2.7).
4. Form the product $XA\alpha_1$.

This algorithm can be used inductively to generate a random element in GL_n in the order of n^3 operations.

Remarks: (1) As in the case of the symmetric group, it is possible to use other towers, taking the subgroup fixing a given k -dimensional subspace for example. We have not experimented with any of these variants. (2) Calabi and Wilf [4] give an extremely elegant way of choosing a random k -dimensional subspace of V_n .

3. THE ORTHOGONAL GROUP

The subgroup algorithm works, in essentially the same way, for any compact topological group G . The idea is to find a closed subgroup $H \subset G$, choose an element of H at random, choose a coset representative at random, and multiply. A careful description is slightly technical and put off until the next section. In this section we treat an example in some detail.

Let $O(n)$ be the group of $n \times n$ orthogonal matrices. $O(n)$ has a natural uniform distribution called Haar measure. In probabilistic notation, the random matrix X is uniformly distributed if

$$P\{X \in A\} = P\{X \in \Gamma A\}$$

for every $A \subset O(n)$ and $\Gamma \in O(n)$.

In two dimensions, a random X can be specified as

$$X = \begin{bmatrix} \cos \theta & \sin \theta \\ -b \sin \theta & b \cos \theta \end{bmatrix};$$

with θ uniform on $[0, 2\pi]$ and $b = \pm 1$ with probability $\frac{1}{2}$. Halmos [20] contains a clear discussion of Haar measure in general.

In some applications, uniformly distributed random matrices are required for large n . For example, statisticians often inspect high-dimensional data ($n = 10-20$) by looking at low-dimensional projections. In one approach, called the grand tour, a low-dimensional subspace is moved around at random in n -

dimensional space. This can be done by choosing random $n \times n$ rotations. See Asimov [2] for more details.

A second applied problem requiring random rotations occurs in cryptography. Sloane [39] has described problems of telephone encryption that require many random orthogonal matrices of high dimension.

With this motivation, we consider a host of methods for generating random orthogonal matrices. The ideas may be useful in related problems. The bottom line is this: the fastest practical algorithm is the subgroup algorithm specialized to this group.

Method A (The classical algorithm). Let X_{ij} be independent normal variables with mean 0 and variance 1, for $1 \leq i, j \leq n$. Treat the X_{ij} as an $n \times n$ matrix. Perform the Gram-Schmidt algorithm. This results in an $n \times n$ orthogonal matrix which is uniformly distributed. This is straightforward to show using the invariance of the normal distribution. See Eaton [13, p. 234] for a proof.

If the Gram-Schmidt algorithm is carried out in the usual way, the i th row has to be taken out of all rows above it. This requires $i - 1$ inner products. Each inner product requires n additions and multiplications. Thus the total number of operations is of order n^3 . For large n , e.g., 256, this is of order 1.6×10^7 which is simply too slow for large scale use. A theoretical improvement, due to Odlyzko [34] allows the Gram-Schmidt computation to be made by the fastest available matrix multiplication algorithm. This is $n^{2.388}$ at present.

Method B. The currently best algorithm for generating random orthogonal matrices is based on the tower

$$O(n) \supset O(n-1) \supset O(n-2) \supset \dots \supset O(2),$$

with $O(n-1)$ the subgroup of $O(n)$ fixing the vector e_1 . Versions of the algorithm have been given by Wedderburn [44], Heiberger [21] (correction by Tanner and Thisted [41]). Stewart [40] gives a very clear discussion.

We will give a new presentation aimed at showing the relation with the subgroup algorithm. Consider the top two terms in the tower. If we knew how to choose a random element of $O(n-1)$, and coset representatives for $O(n-1)$ in $O(n)$ at random, then the subgroup algorithm (and induction) finish the job.

Coset representatives for $O(n-1)$ in $O(n)$ can be specified by saying where e_1 goes. Thus the coset space $O(n)/O(n-1)$ can be identified with \mathbb{S}^{n-1} —the $(n-1)$ -dimensional sphere in \mathbb{R}^n . For $x \in \mathbb{S}^{n-1}$ (as a row vector), define the Householder reflection

$$I - 2x'x. \tag{3.1a}$$

For every $v \in \mathbb{S}^n$ the reflection with

$$x = (e_1 - v)/c \text{ and } c = \sqrt{(e_1 - v)(e_1 - v)'} \tag{3.1b}$$

takes e_1 into v . Choosing v at random results in a randomly chosen coset representative. (This is made more precise in Section 4.) This results in the following:

LEMMA 3.1: *Let V be chosen at random on the n -sphere. Let Γ_1 be chosen at random in $O(n-1)$. Then, with x defined by Eq. (3.1b)*

$$(I - 2x^t x) \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \Gamma_1 & \\ 0 & & & \end{pmatrix} \quad (3.2)$$

is uniformly distributed on $O(n)$.

The standard way of choosing V at random on S_n is to take $V = (Z_1, \dots, Z_n) / \sqrt{Z_1^2 + \dots + Z_n^2}$, with Z_i independent standard normal. Marsaglia [31] discusses refinements. From here, induction gives a simple algorithm for choosing a uniformly distributed element of $O(n)$. A usable version of the algorithm is detailed by Stewart [40]. The complete algorithm requires successively multiplying together $n-1$ reflections. For y a vector in \mathbb{R}^n , the product $(I - 2x^t x)y'$ can be computed in the order of n operations (just compute the inner product xy' , then multiply $2x'$ by this number and add to y'). Thus, for $\Gamma \in O(n)$ the product $(I - 2x^t x)\Gamma$ can be computed in the order of n^2 . It follows that Stewart's algorithm is also an n^3 algorithm. Preliminary computations show the constant is enough smaller than the corresponding constant for the classical algorithm to make a substantial difference. Further discussion of this algorithm is given in Section 4.

Other Methods. Here is a brief review of some other approaches to generating random orthogonal matrices. None of these approaches is particularly recommended, but the ideas may prove useful in related problems.

1. *Random Givens rotations.* As is well known, any orthogonal matrix can be written uniquely as a product of 2×2 rotations (Givens rotations) of the form $\begin{pmatrix} c & s \\ -s & c \end{pmatrix}$, see Golub and Van Loan [19]. Several authors have shown how to generate a random orthogonal matrix by multiplying together $\binom{n}{2}$ such rotations. The cosines are chosen independently from beta densities. It takes $4n$ operations to multiply by a basic rotation, so this is an order n^3 algorithm. Monte Carlo work carried out by Don Fraser (personal communication) suggests this algorithm works well for n up to size 50. A careful derivation of the algorithm and further comparisons are given by Anderson and Olkin [1]. Vilenkin [43, 9.1] gives the result in group theoretic language.

2. *Methods based on densities.* There are many approaches to generating real valued random variable with given density f . A valuable compendium is

given by Devroye [6]. Toward this end an expression for the density of Haar measure in a suitable coordinate system might be of some use.

There are several known expressions for the density; it can be given in coordinates of e^k , with k skew-symmetric, in Euler angles (Vilenkin [43, Chapter 9]), or in the language of Lie theory (Weyl [45, Chapter 7]), (Helgason [22, Chapter 1]).

We focus on one coordinate system to give the flavor of our approach. Let $O'(n)$ be the subset of $O(n)$ having no characteristic root equal to -1 . Let K_n be the set of skew-symmetric real $n \times n$ matrices. A 1-1 transformation τ from $O'(n)$ to K_n is defined by

$$\tau\Gamma = (I - \Gamma)(I + \Gamma)^{-1} = (I + \Gamma)^{-1}(I - \Gamma)$$

with τ^{-1} given by

$$\tau^{-1}k = (I - k)^{-1}(I + k) = (I + k)(I - k)^{-1}.$$

This allows the skew-symmetric matrices to serve as coordinates for $O(n)$, up to a set of measure zero. This is called Cayley's parametrization. See Weyl [45, p. 56] for further discussion.

The skew-symmetric matrices can be naturally identified with $\binom{n}{2}$ dimensional Euclidean space. It can be shown that Haar measure on $O(n)$ has density

$$\frac{cdk}{[\det(I + kk')]^{(n-1)/2}} \quad \text{for } c = \pi^{n(n-1)/4} \prod_{k=2}^n \frac{\Gamma\left(\frac{k-1}{2}\right)}{\Gamma(k-1)}. \quad (3.3)$$

Here is one possible use of Eq. (3.3): for Γ chosen from Haar measure on $O(n)$, define

$$K = (I - \Gamma)(I + \Gamma)^{-1}.$$

For large n , the coordinates of Γ are approximately independent normal variables having mean zero and variance $1/n$ (see the following paragraph). It appears that the coordinates of K are approximately independent normal variables, with mean zero, and variance $1/2n$. This suggests choosing K with such normal coordinates, and basing a rejection method on the density in Eq. (3.3).

3. Random reflections. In light of the difficulty of generating uniform variables in $O(n)$ when n is large, it is natural to seek schemes for generating approximately uniformly distributed variables. This is not as easy as it seems.

To make things precise, it must be specified how the approximation will be measured. Let X be a random $n \times n$ matrix. Let Γ be a Haar distributed random matrix. One approach to measuring closeness in distribution is to ask that any linear combination of the elements of X be close to the same linear combination of the elements of Γ in the weak star topology (on \mathbb{R}).

Diaconis and Newman [9] showed that in this sense Haar measure can be

approximated by a matrix X of independent normal random variables having mean zero and variance $1/n$:

$$\sup_{A:|A|^2=n} \text{Tr}(A\Gamma) \xrightarrow{wk^*} n(0,1)$$

where the sup is over all $n \times n$ matrices having $\text{Tr}(A'A) = n$, and convergence is as $n \rightarrow \infty$.

With X as described, many other distributional features match a Haar distributed matrix: most pairs of rows or columns are close to orthogonal (Diaconis and Freedman [7]). The first $2n$ moments of $\text{Tr}(X)$ equal the first $2n$ moments of $\text{Tr}(\Gamma)$ (Diaconis and Mallows [8]).

Of course, X is not orthogonal. For example, the length of the longest row in X is "off" by $\sqrt{\log n}$. It is natural to try to generate an approximately uniformly distributed random matrix, staying within the orthogonal group, by performing an easily computed random walk.

Neil Sloane and Morris Eaton independently suggested a product of random reflections: $I - 2x'x$ with x chosen uniformly on the n -sphere. Diaconis and Shahshahani [11] showed that a product of k random reflections converges to Haar measure in total variation distance if $k = \frac{1}{2}n \log n + cn$, the error being e^{-c} . Conversely, if $k = \frac{1}{2}n \log n - cn$ reflections are used, the two distributions are far apart. Since it takes of the order of n operations to multiply by a reflection, this algorithm requires of the order of $n^3 \log n$ steps, and will be slower than the subgroup algorithm for large n .

4. CONTINUOUS GROUPS, GENERAL THEORY

In this section we make precise notations like "choose a coset at random." We prove that the continuous analog of the subgroup algorithm works. As examples, we give algorithms for random unitary and symplectic matrices. Finally, we give results delineating which compact groups have closed subgroups.

Let G be a compact topological group. To simplify things, assume throughout that as a topological space, G is metrizable, separable, and complete (G is "Polish"). This includes all examples of practical interest. Any unexplained terminology may be found in Hewitt and Ross [24,25].

Let $H \subset G$ be a closed subgroup. The quotient space $X = G/H$ will be called the space of cosets. The map that assigns $g \in G$ to the coset containing g is denoted $\pi:G \rightarrow X$. To choose coset representatives, let $\phi: X \rightarrow G$ be a measurable inverse of π (so $\pi\phi(x) = x$). The existence of ϕ under our hypothesis follows from Theorem 1 in Bondar [3].

Define $T:G \rightarrow X \times H$ by

$$T(g) = (\pi(g), (\phi\pi(g))^{-1}g).$$

The map T is 1-1, onto, bimeasurable with inverse

$$T^{-1}(x, h) = \phi(x)h.$$

Let dP_G , dP_H , dP_X be invariant measures on G , H , X respectively, normalized so that each space has total mass 1. Then

LEMMA 4.1: *Haar measure on G admits of decomposition*

$$T(dP_G) = dP_X \times dP_H.$$

PROOF: From the definition of invariant measures

$$\int_G f(g) dP(g) = \int_X \int_H f(gh) dP(h) dP(x) \quad (4.1)$$

where $x = gH = \pi(g)$. The required result follows from the product decomposition defined by T .

The next corollary, which is simply a restatement of Eq. 4.1, is the rigorous version of the subgroup algorithm.

COROLLARY 1: *Denoting the image of dP_X under ϕ by dP_X again, we have*

$$dP_G = dP_X * dP_H,$$

with the $$ denoting convolution of probabilities on G .*

The condition that ϕ be a section can be substantially relaxed. In fact,

COROLLARY 2: *Let $\tilde{\phi}$ be a map from $X \rightarrow G$ with the property that is measure preserving transformation of X . Let $d\tilde{P}_X$ be the image of dP_X under $\tilde{\phi}$. Then*

$$dP_G = d\tilde{P}_X * dP_H.$$

PROOF: Let A and B be subsets of X and H respectively, and $P_G(C)$ denote the total mass of the set C relative to the measure dP_G , etc. The statement of the corollary is equivalent to

$$P\{g: \pi(g) \in A, \eta(g) \in B\} = \tilde{P}_X \times P_H\{(x, h) \pi\tilde{\phi}(x) \in A, h \in B\},$$

where $\eta(g) = \phi(\pi(g))^{-1}g$. The above equality follows from Eq. (4.1) and the fact that $\tilde{\phi}$ is measure preserving.

Remarks: (1) Lemma 4.1 is a special case of much more general theorems such as Theorem 2 in Bondar [3]. It is a fair amount of work to do the translation, so we have given the argument. (2) It is crucial that H is closed. Indeed, the classical construction of a non-measurable set (due to Vitali) uses $G = [0, 2\pi]$ under addition and H the subgroup of rationals in G . Then, there is no measurable choice of coset representatives. (3) We have assumed that G is Polish. Baire measurable representatives may be chosen for any compact group. See Kupka [29] or Kehlet [26].

Example (The orthogonal group): Consider $O(n)$ as in Section 4 with $O(n-1) = \{\Gamma: \Gamma e_1^i = e_1^i\}$ and $X = S_n$, the unit sphere. Then $\pi(\Gamma) = \Gamma e_1^i$. The map

$$\begin{cases} \phi(x) = I - 2V'V/C, \text{ with } V = -x + e_1, C = VV' \\ \phi(e_1) = I \end{cases}$$

is a measurable inverse of π that is continuous except at e_1 (there is no continuous choice of coset representatives). Corollary 1 gives a rigorous justification for Lemma 3.1.

Remarks: (1) Here is an application of Corollary 3 to this problem. Consider the map $\tilde{\phi}: S_n \rightarrow O(n)$ defined by $\tilde{\phi}(x) = I - 2x'x$. This does not serve as an inverse of π . But, at least in two dimensions, if x is uniformly distributed on S_n , $\tilde{\phi}(x)e_1^i$ is uniformly distributed on S_n . It follows that $\pm(I - 2x'x)$ is uniform on $O(2)$. Curiously, the image of e_1 under a random reflection is not uniform on S_n for $n \geq 3$. (2) The subgroup algorithm in $O(3)$ may be described in words as follows: make a random rotation fixing the north pole (the vector e_1^i) and then take the north pole to a random point on the sphere. This results in a random rotation in $O(3)$ (up to sign).

It is a classical fact that any rotation in $SO(3)$ fixes a line and is a two-dimensional rotation about that line. Under Haar measure, the line fixed is randomly situated. It is natural to try to generate a random rotation by reversing the order of the two steps above: pick a random point on the sphere, and rotate a random angle about this fixed point.

It is instructive to observe that this construction is badly wrong: Kendall and Moran [27, p. 93] show that the angle about the fixed point should have density proportional to $(\sin \theta)^2$, not $d\theta$.

Here are two new examples.

Example (Unitary matrices): Consider the group $U(n)$ of $n \times n$ matrices M with complex entries, satisfying

$$M^*M = I$$

where $*$ denotes the conjugate transpose. There is a natural tower

$$U(n) \supset U(n-1) \supset U(n-2) \supset \dots \supset U(2),$$

where $U(n-1)$ is regarded as the subgroup of $U(n)$ fixing e_1^i . Thus, a matrix $M \in U(n-1)$ can be represented as

$$M = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & M_1 & \\ 0 & & & \end{pmatrix}, \text{ with } M_1^*M_1 = I_{n-1, n-1},$$

Coset representations for $U(n-1)$ in $U(n)$ may be chosen as the images of e_1^i under $U(n)$. That is, as

$$S_n = \{x \in \mathbb{C}^n: xx^* = 1\}.$$

Here is one useful method of choosing a map $\phi: S_n \rightarrow U(n)$. For $x \in \mathbb{C}^n$, with $x_1 = |x_1|e^{i\theta}$, $0 \leq \theta < 2\pi$, define

$$v = x + e^{i\theta}e_1, \quad u = v/\sqrt{vv^*}.$$

Let $\phi(x) = -e^{i\theta}(I - 2u^i\bar{u})$. This ϕ is unitary and $\phi(x)e_1^i = x^i$.

It is easy to choose a random point on S_n by using the standard complex normal distribution. Let $Z = X + iY$ be a complex random variable with independent real and imaginary parts X and Y having normal distributions with mean 0 and variance $\frac{1}{2}$. If $\bar{Z} = (Z_1, Z_2, \dots, Z_n)$ has independent standard complex normal coordinates, then $\bar{Z}/\sqrt{\bar{Z}Z^*}$ is uniform on S_n .

With these ingredients, the algorithm proceeds inductively, just as for the orthogonal group.

Example (The symplectic group): There is only one other “natural” infinite nested family of compact groups—the symplectic groups $Sp(n)$. These can be described as the set of $2n \times 2n$ unitary matrices that preserve an alternating form. As such, they appear extensively in modern mechanics. A readable careful discussion of $Sp(n)$ may be found in Chevalley [5, Chapter 1].

For our purposes, it is useful to work with another representation. $Sp(n)$ may be represented as the $n \times n$ matrices M with entries in the quaternions \mathbb{H} , satisfying

$$M^*M = I$$

with $*$ denoting the conjugate transpose. The conjugate of the quaternion $(a + ib + jc + hd)^* = (a - ib - jc - hd)$. The groups nest as

$$Sp(n) \supset Sp(n-1) \supset \dots \supset Sp(2) > Sp(1)$$

with $Sp(n-1)$ contained in $Sp(n)$ as $\{M: Me_1^i = e_1^i\}$.

To use the subgroup algorithm, we need the analog of a reflection over the quaternions. One idea that works: Let $x \in \mathbb{H}^n$, as a row vector. Suppose $x_1 \neq 0$, and $xx^* = 1$. Write $x_1 = |x_1|q$ with $|x_1| = \sqrt{x_1^*x_1}$, and q a unit quaternion: $qq^* = 1$. Let $u = x + qe_1$.

A routine computation shows that the matrix

$$M = -q \left(I - \frac{2u^i\bar{u}}{uu^*} \right) \tag{4.2}$$

is in $Sp(n)$ and satisfies $Me_1^i = x^i$. Here, $\bar{u} = (u_1^*, \dots, u_n^*)$.

Now, the ideas explained for the orthogonal and unitary group work. Choose x at random on $xx^* = 1$ by using a standard quaternionic normal distribution.

Here is a classical example in the present language.

Example (Coin tossing and Lebesgue measure): The subgroup algorithm works whenever there is a nested increasing tower of groups. One familiar example where the tower is infinite: Take $G = \mathbb{Z}_2^\infty$ and consider the tower

$$G \supset G_1 \supset G_2 \supset \dots$$

where $G_i = \{g \in G: g = (0, 0, \dots, 0, *, *, *, \dots)\}$, with the first i coordinates zero, and all the remaining coordinates unrestricted. The subgroup algorithm gives the usual decomposition of the uniform distribution on $[0, 1] \cong G$ as the infinite convolution of $X_i/2^i$ with X_i independent Bernoulli $P(X_i = 1) = \frac{1}{2} = P(X_i = 0)$.

As the final topic of this section we consider the problem of the existence of closed subgroups. To begin with, we argue that every infinite compact group contains a closed nontrivial subgroup. A topological group is said to have no small subgroups if there exists a neighborhood U of the identity such that the only subgroup in U is $\{\text{id}\}$. Clearly a group with small subgroups contains nontrivial closed subgroups.

A famous theorem of Gleason [18] implies that a group with no small subgroups is a Lie group. The structure of compact Lie groups is well known; see Chapter 11 of Pontryagin [35]: If G is Abelian, then the connected component of the identity is a finite dimensional torus which certainly has nontrivial closed subgroups, hence G does. If G is not Abelian, then its maximal torus is a nontrivial closed subgroup.

For finite groups, observe that a finite group with no proper subgroups is \mathbb{Z}_p , the residues modulo a prime. This is the only case where the subgroup algorithm does not work to produce a nontrivial factorization of the uniform distribution. It is easy to see that the uniform distribution does not factor on \mathbb{Z}_2 or \mathbb{Z}_3 . One approach uses the result that $1 + z$ and $1 + z + z^2$ are irreducible over the reals. A nontrivial factorization of U would lead to a factorization of the associated polynomial.

To complete the discussion we prove:

LEMMA 4.2: *Let $p \geq 5$ be a prime. Let \mathbb{Z}_p be the integers mod p . Then, there exist nonuniform probabilities P_1, P_2 on \mathbb{Z}_p , such that*

$$P_1 * P_2 = U.$$

PROOF: For $i = 1, 2, \dots, \frac{p-1}{2}$, let a_i, b_i be determined by

$$a_i + 2b_i = 1, \quad a_i + 2b_i \cos\left(\frac{2\pi i^2}{p}\right) = 0.$$

Noting that $\cos\left(\frac{2\pi i^2}{p}\right) \neq 1$ for i in the indicated range;

$$b_i = \left\{ 2 \left(1 - \cos \frac{2\pi i^2}{p} \right) \right\}^{-1}, \quad a_i = -\cos \left(\frac{2\pi i^2}{p} \right) / \left(1 - \cos \frac{2\pi i^2}{p} \right).$$

Define signed measures Q_i on Z_p by

$$Q_i(0) = a_i, \quad Q_i(i) = Q_i(-i) = b_i, \quad Q_i(j) = 0 \text{ otherwise.}$$

The argument depends on the discrete Fourier transform of a measure. If P is a measure on Z_p and $k \in Z_p$, define

$$\rho_k(P) = \frac{1}{p} \sum_{j=0}^{p-1} P(j) e^{2\pi i j k / p}.$$

For the uniform distribution,

$$\rho_k(U) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that for $k \neq 0$, $\rho_{\pm k}(Q_k) = 0$, $\rho_0(Q_k) = 1$. Now let signed measures R_1 and R_2 be defined by

$$R_1 = \underset{i=1}{\overset{a}{*}} Q_i, \quad R_2 = \underset{i=a+1}{\overset{(p-1)/2}{*}} Q_i \quad \text{for fixed } 1 \leq a \leq (p-1)/2.$$

Finally, for sufficiently small ϵ the measures $U + \epsilon R_1$ and $U + \epsilon R_2$ are positive measures and can be normed to be probabilities, say P_1 and P_2 . We claim $U = P_1 * P_2$. Indeed, for $k \neq 0$, $\rho_k(P_1 * P_2) = \rho_k(P_1) \rho_k(P_2) = 0$. To show that the decomposition is nontrivial, it suffices to show that R_i are nonzero, $i = 1, 2$. This follows from the fact that for $k \neq j$, $\rho_j(Q_k) \neq 0$. Indeed,

$$\begin{aligned} \rho_j(Q_k) &= a_k + 2b_k \cos \left(\frac{2\pi j k}{p} \right) \\ &= \frac{1}{1 - \cos(2\pi k^2/p)} \left\{ -\cos \left(\frac{2\pi k^2}{p} \right) + \cos \left(\frac{2\pi j k}{p} \right) \right\}. \end{aligned}$$

This is zero if and only if $j = k$.

Remarks: (1) This type of factorization can be extended to noncommutative groups, see Diaconis and Shahshahani [10]. (2) Factoring the uniform distribution on Z_p is sufficiently close to some classical factorization results to warrant discussion. A well known elementary probability problem argues that it is impossible to load two dice so that the sum is uniform. More generally, Dudewicz and Dann [12] show that it is impossible to find probabilities P_1 and P_2 on the set $\{1, 2, \dots, n\}$ such that $P_1 * P_2$ is the uniform distribution on $\{2, \dots, 2n\}$. A related result asks for a decomposition of the uniform distribution on the set $0, 1, 2, \dots, n$. Lukacs [30, p. 182–183], reviews the literature on this problem. He shows factorization is possible when, and only when, n is prime. The difference between the three results is this: In Lemma 4.2, and

in the subgroup factorization, addition is (mod n). In the dice result, both factors must be supported on $\{1, \dots, n\}$ while the uniform distribution is on $\{2, \dots, n\}$. In the results reported in Lukacs, the factors are permitted to have arbitrary support.

The remarks before the lemma and Lemma 4.2 combine to give

THEOREM 1. *The uniform distribution on a compact Polish group G can be factored as $U = P_1 * P_2$, and P_i not uniform, if and only if $G \neq Z_2$ or Z_3 .*

NOTE: Further discussion of random rotations and their applications to cryptography can be found in S. P. Lloyd (1977), "Random Rotation Secrecy System," Technical Memo 77-1217-3, Bell Labs, and S. P. Lloyd (1978), "Choosing a Rotation at Random," Technical Memo 78-1217-4, Bell Labs.

References

1. Anderson, T. W. and Olkin, I. (1985). Generation of random orthogonal matrices. Technical Report No. 6, Department of Statistics, Stanford University. To appear, *SIAM Jour. Sci. Statist. Comput.*
2. Asimov, D. (1983). The grand tour. *SIAM Jour. Sci. Statist. Comput.* 6: 128-143.
3. Bondar, J. V. (1976). Borel cross-sections and maximal invariants. *Ann. Statist.* 4: 866-877.
4. Calabi, E. and Wilf, H. S. (1977). On the sequential and random selection of subspaces over a finite field. *Jour. Combin. Th.* 22: 107-109.
5. Chevalley, C. (1946). *Theory of Lie Groups*. Princeton University Press, Princeton, New Jersey, Chapter 1.
6. Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, Chapters 9 and 13.
7. Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *Ann. Stat.* 12: 793-815.
8. Diaconis, P. and Mallows, C. (1984). The trace of Haar measure. Unpublished manuscript.
9. Diaconis, P. and Newman, C. (1984). The elements of a random orthogonal matrix are approximately normal. Unpublished manuscript.
10. Diaconis, P. and Shahshahani, M. (1986a). On square roots of the uniform distribution on compact groups. *Proc. Amer. Math. Soc.* 98: 341-348.
11. Diaconis, P. and Shahshahani, M. (1986b). Products of random matrices as they arise in the study of random walks on groups. *Contemporary Math.* 50: 183-195.
12. Dudewicz, E. J. and Dann, R. E. (1972). Equally likely dice sums do not exist. *Amer. Statistician* 26: #4, 41-42.
13. Eaton, M. L. (1983). *Multivariate Statistics*. Wiley, New York, p. 234.
14. Eidswick, J. A. (1986). Cubelike puzzles—what are they and how do you solve them? *Amer. Math. Monthly* 93: 157-176.
15. Fisher, R. A. and Yates, F. (1938). *Statistical Tables for Biological Agricultural and Medical Research*. Oliver and Boyd, London, p. 20.
16. Furst, M., Hopcroft, J., and Luks, E. (1980). Polynomial time algorithms for permutation groups. *Proc. 21st FOCS I*, 36-41.
17. Gilbert, W. J. (1976). *Modern Algebra with Applications*. Wiley, New York, p. 81.
18. Gleason, A. M. (1952). Groups without small subgroups. *Ann. of Math. Stat.* 56: 193-212.
19. Golub, G. and Van Loan, C. (1983). *Matrix Computations*. Johns Hopkins, Baltimore, Maryland, Chaps. 11, 12.
20. Halmos, P. (1950). *Measure Theory*. Van Nostrand, New York.

21. Heiberger, R. M. (1978). Generation of random orthogonal matrices. *Applied Statistics* 27: 199–206.
22. Helgason, S. (1984). *Groups and Geometric Analysis*. Academic Press, New York, Chapter 1.
23. Herstein, I. N. (1964). *Topics in Algebra*. Blaisdell, Waltham, Massachusetts, Chap. 2.
24. Hewitt, E. and Ross, K. (1963). *Abstract Harmonic Analysis I*. Springer-Verlag, Berlin.
25. Hewitt, E. and Ross, K. (1970). *Abstract Harmonic Analysis II*. Springer-Verlag, Berlin.
26. Kehlet, E. T. (1984). Cross sections for quotient maps of locally compact groups. *Math. Scand.* 55: 152–160.
27. Kendall, M. G. and Moran, P. A. P. (1963). *Geometrical Probability*. Griffin, London, p. 93.
28. Knuth, D. (1981). *The Art of Computer Programming*, Vol. II, 2nd ed. Addison-Wesley, Reading, Massachusetts, p. 139–141.
29. Kupka, J. (1983). Strong liftings with application to measurable cross sections of locally compact groups. *Israel Jour. Math.* 44: 243–261.
30. Lukacs, E. (1970). *Characteristic Functions*, 2nd ed. Griffin, London, p. 182–183.
31. Marsaglia, G. (1972). Choosing a point from the surface of a sphere. *Ann. Math. Statist.* 43: 645–646.
32. Moses, L. E. and Oakford, R. A. (1963). *Tables of Random Permutations*. Stanford University Press, Stanford, California, p. 4–5.
33. Nijenhuis, A. and Wilf, H. S. (1978). *Combinatorial Algorithms*, 2nd ed. Academic Press, New York.
34. Odlyzko, A. (1986). Personal communication.
35. Pontryagin (1966). *Topological Groups*, 2nd ed. Gordon and Breach, New York, Chapter 11.
36. Rouse-Ball, W. W. (1962). *Mathematical Recreations and Essays*, 11th ed. Macmillan, New York, p. 299.
37. Sims, C. (1970), *Computation Problems in Abstract Algebra* (John Leech, ed.). Pergamon Press, New York, p. 176–177.
38. Singmaster, D. (1981). *Notes on Rubik's Magic Cube*. Enslow Publishers, Hillside, New Jersey.
39. Sloane, N. J. A. (1983). Encrypting by random relations. In *Cryptography: Lecture Notes in Computer Science*, Vol. 149, (T. Beth, ed.). Springer-Verlag, Berlin, p. 71–128.
40. Stewart, G. W. (1980). The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Jour. Numer. Anal.* 17: 403–409.
41. Tanner, M. A. and Thisted, R. (1982). A remark on AS127. Generation of random orthogonal matrices. *Applied Statistics* 31: 190–192.
42. Varadarajan, V. S. (1974). *Lie Groups, Lie Algebras and Their Representations*. Prentice Hall, Englewood Cliffs, New Jersey.
43. Vilenkin, N. J. (1968). *Special Functions and The Theory of Group Representations*. American Mathematical Society, Providence, Rhode Island, Chapter 9.
44. Wedderburn, R. W. M. (1975). Generating random rotations. Research Report, Rothamsted Experimental Station.
45. Weyl, H. (1946). *The Classical Groups*. Princeton University Press, Princeton, New Jersey, Chapter 7, p. 56.