THE SUBGROUP ALGORITHM FOR GENERATING UNIFORM RANDOM VARIABLES

BY

PERSI DIACONIS AND MEHRDAD SHAHSHAHANI

TECHNICAL REPORT NO. 257
SEPTEMBER 1986

PREPARED UNDER THE AUSPICES OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE SUBGROUP ALGORITHM FOR GENERATING UNIFORM RANDOM VARIABLES

BY

PERSI DIACONIS AND MEHRDAD SHAHSHAHANI

TECHNICAL REPORT NO. 257
SEPTEMBER 1986

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS86-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
THE SUBGROUP ALGORITHM FOR GENERATING UNIFORM RANDOM VARIABLES

Persi Diaconis
Department of Statistics
Stanford University
Stanford, CA 94305

Mehrdad Shahshahani
Jet Propulsion Laboratory
Pasadena, CA 91105

Abstract

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 $\text{GL}_n$ over a finite field. The algorithm reduces to the "standard" fast algorithm when there is one, but many new examples are included.

I. 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 $U_1, U_2, \ldots$, 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 ith and $J_i$'s registers have their contents switched. This continues from $1 \leq i \leq n$. It will be argued below that this produces a random permutation. Further discussion of this algorithm is on pp. 139-141 of Knuth (1981).
Knuth attributes the algorithm to Moses and Oakford (1963). Moses (personal communication) attributes the idea to Fisher and Yates (1938, p. 20). Other discussions of this subject appear in Nijenhuis and Wilf (1978) and Devroye (1986, Chapters 9, 13).

The following algorithm abstracts the idea to any finite group. See Herstein (1964) for any unexplained group theoretic terminology.

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

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

in the sense that each $g \in G$ has a unique representation as $g_0 g_1 \ldots g_r$ 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

$$S_n \supset S_{n-1} \supset S_{n-2} \supset \ldots \supset \{\text{id}\},$$

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.

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

$$S_n \supset S_{n-2} \supset S_{n-4} \supset \ldots \supset \{\text{id}\}.$$

2
Coset representations for $S_{n-2(i+1)}$ in $S_{n-2i}$ are permutations bringing a pair of elements between $i+1$ and $n$ into $i+1$ and $i+2$. These may be taken as the set of products $(i,j)(i+1,k)$ with $i \leq j \leq n$, $i+1 \leq k \leq n$. These coset representatives may be ordered lexicographically, setting up a 1-1 correspondence between them and the numbers 1, 2, ..., $(n-i)(n-i-1)$. This would have the advantage of making fewer calls to the random number generator and the disadvantage of more bookkeeping.

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

\[(1.3) \quad S_n = \{ \text{id} \}. \]

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 with the permutations. This is a useful approach if $n$ is small (e.g., $\leq 10$).

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

$$a_1 + a_22! + \ldots + a_{n-1}(n-1)! \quad \text{with} \quad 0 \leq a_i \leq i.$$  

The "digits" $a_i$ are determined as the remainders on successive division by $2$, $3$, ..., $n-1$.

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 (1.1): $(1,n-a_{n-1})(2,n-a_{n-2})\ldots(n-1,n-a_1)$.

There doesn't appear to be any practical difference between algorithms based on (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, ..., 15 are placed into a $4 \times 4$ grid, leaving one empty square. The tiles are to be slid around to bring them to a given order.

\[
\begin{array}{cccc}
1 & 2 & 3 & 4 \\
5 & 6 & 7 & 8 \\
9 & 10 & 11 & 12 \\
13 & 14 & 15 & \text{BLANK} \\
\end{array}
\]

It is well known (see, e.g., Gilbert (1976, pg. 81), or Rouse-Ball (1962, pg. 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

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

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

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

Another method of generating a random element of $A_n$ is to run the subgroup algorithm based on (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)).

Example 3. (Rubik's cube, etc.). The popular puzzle Rubick's cube (see Singmaster (1981) or Eidswick (1986)) 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 (1970) that takes a set of generators of a permutation group \( G \) on \( n \) letters, and finds explicit coset representatives for the tower

\[
G = G_n \triangleright G_{n-1} \triangleright G_{n-2} \ldots \triangleright \{\text{id}\},
\]

where \( G_{n-i} = G \cap S_{n-i} \). We have carried this computation out for the Rubick's cube puzzle which involves a subgroup of \( S_{27} \). 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 (1980). It is implemented in the algebraic program Caley.

The remainder of this paper treats further applications of the subgroup algorithm. Section 2 treats \( \text{GL}_n(\mathbb{F}_q) \) 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.

Section 2. \( \text{GL}_n(\mathbb{F}_q) \).

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 \( \text{GL}_n \) be the set of \( n \times n \) invertible matrices with coefficients in \( \mathbb{F}_q \). Then \( \text{GL}_n \) is a finite group of order

\[
(2.1) \quad (q^n - 1)(q^n - q) \ldots (q^n - q^{n-1}).
\]

A familiar example has \( q = 2 \), so \( V_n \) is the binary \( n \)-tuples and \( \text{GL}_n \) consists of \( n \times n \) invertible binary matrices, all operations occurring mod 2.

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

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

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 needed. To simplify the explanation, consider the first steps of the chain:

(2.2) \[ GL_n \supset P_n \supset GL_{n-1}. \]

In (2.2) $P_n$ consists of $\{\alpha \in GL_n : \alpha = e_1 \}$ with $e_1$ the row vector in $V_n$ containing a 1 in the first coordinate and zeros elsewhere. Thus a typical entry $\alpha \in P_n$ looks like

(2.3) \[
\alpha = \begin{pmatrix}
a_1 & x \\
0 & \\
\vdots & \\
0 & \alpha_1
\end{pmatrix}
\]

with $a_1$ a non-zero element of $F_q$, $x \in V_{n-1}$, and $\alpha_1 \in GL_{n-1}$.

The cardinality of $P_n$ is thus

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

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

(2.5) \[ GL_n/P_n \approx \mathbb{F}^{n-1}. \]
Coset representatives for $P_n$ in $GL_n$ can be chosen by enumerating lines as:

$$
\begin{align*}
\text{# representatives} & \quad \text{of this type} \\
(0 \ 0 \ \ldots \ 1) & \quad - 1 \\
(0 \ \ldots \ 1\ast) & \quad - q^2 \\
\vdots & \\
(1 \ \ast\ldots\ast) & \quad - q^{n-1}
\end{align*}
$$

Here $\ast$ denotes any element of $F_q$. It follows that

$$|F_q^{n-1}| = 1 + \ldots + q^{n-1}.$$ 

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

It remains to identify the lines in (2.4) with elements in $GL_n$. This requires choosing, for any $x$ in (2.6), a matrix $\alpha_x$ such that

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

Over the real field, $\alpha_x$ 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

$$(0, \ldots, 0, 1, x_{n-2}, \ldots, x_n).$$

Then $\alpha_x$ can be taken as

$$
\begin{pmatrix}
0_1 & I_1 \\
1 & \\
x_{n-2} & \vdots & \ddots & 0_2 \\
x_n & & & 1
\end{pmatrix}
$$

(2.7)

Here $0_1$ is a $n-(\ell+1) \times (\ell+1)$ matrix of zeros, $I_1$ is an $n-(\ell+1) \times n-(\ell+1)$ identity, the bottom left block is $(\ell+1) \times (\ell+1)$ with ones down the diagonal,
the first column as specified, and zeros elsewhere. Finally $O_2$ is a $(\ell+1) \times n-(\ell+1)$ matrix of zeros.

The matrix $\alpha_\chi$ is non-singular and satisfies $\alpha_\chi e_1^t = x^t$. Moreover, because of its sparse structure, it is possible to compute the product $\alpha_\chi y^t$ (for $y \in V_n$) in order $n$ operations. Hence, for $\beta \in \text{GL}_n$, $\alpha_\chi \beta$ 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 $\text{GL}_{n-1}$ in $P_n$. Embed $\text{GL}_{n-1}$ in $P_n$ by

$$\alpha_1 \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha_1 \\ 0 \end{pmatrix} \quad \text{for } \alpha_1 \in \text{GL}_{n-1}.$$ 

From (2.2), coset representatives are given by

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{13} \\ 0 & 1 & \ddots & \vdots \\ 0 & & & 1 \end{pmatrix} \quad (2.8)$$

with $a_{11}$ non-zero and $a_{1j}$ arbitrary. Again, if $\alpha$ has the form (2.8), and $\beta \in \text{GL}_n$, then $\alpha \beta$ can be computed in $n^2$ operations.

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

**Algorithm.** Given $\alpha_1$ uniformly distributed in $\text{GL}_{n-1}$, this algorithm outputs $\alpha$ uniformly distributed in $\text{GL}_n$.

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 (2.8).
2) Choose an integer \( l, 1 \leq l \leq n \) from the distribution

\[ P(l) = q^{2-l}(q-1)/(q^n-1). \]

3) Given \( l \), choose \( l-1 \) elements \( x_{n-l}, \ldots, x_n \) independently at random in \( \mathbb{F}_q \). Form the matrix \( X \) as in (2.7).

4) Form the product \( XA\mathbf{e}_l \).

This algorithm can be used inductively to generate a random element in \( \text{GL}_n \) in order \( n^3 \) 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 (1977) 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

$$X = \begin{pmatrix} \cos \theta & \sin \theta \\ b \sin \theta & b \cos \theta \end{pmatrix}.$$ 

with $\theta$ uniform on $[0, 2\pi)$ and $b = \pm 1$ with probability $1/2$. Halmos (1950) 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 (1983) for more details.

A second applied problem requiring random rotations occurs in cryptography. Neil Sloane (1983) 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, due to G. W. Stewart, 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 (1983, pg. 234) for a proof.

If the Gram-Schmidt algorithm is carried out in the usual way, the ith row has to be taken out of all rows above it. This requires $i-1$ inner products.

10
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 $1.6 \times 10^7$ which is simply too slow for large scale use. A theoretical improvement, due to Andrew Odlyzko (1986) allows the Gram-Schmidt computation to be made by the fastest available matrix multiplication algorithm. This is $n^{2.388}$ at present.

**Method B (Stewart's algorithm).** The currently best algorithm for generating random orthogonal matrices is due to G. W. Stewart (1980). It is based on the tower

$$O(n) \supseteq O(n-1) \supseteq O(n-2) \supseteq \ldots \supseteq O(2),$$

with $O(n-1)$ the subgroup of $O(n)$ fixing the vector $e_1$.

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 elements 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 $S_n$ - the sphere in $n$-dimensions. For $x \in S_n$ (as a row vector), define the Householder reflection

$$(3.1a) \quad I - 2x^tx.$$

For every $v \in S_n$ the reflection with

$$(3.1b) \quad x = (e_1 + v)/c \quad \text{and} \quad c = \sqrt{(e_1 + v)(e_1 + v)^t}$$

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 \( \Gamma_1 \) be chosen at random in \( O(n-1) \). Then, with \( x \) defined by (3.1b)

\[
(I - 2x^tx) \begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & & & \\
\vdots & & & \Gamma_1 \\
0 & & & 
\end{pmatrix}
\]

is uniformly distributed on \( O(n) \).

The standard way of choosing \( V \) at random on \( S_n \) is to take
\( V = (Z_1, \ldots, Z_n)/\sqrt{Z_1^2 + \ldots + Z_n^2} \), with \( Z_i \) independent standard normal. Marsaglia (1972) discusses refinements. From here, induction gives a simple algorithm for choosing a uniformly distributed element of \( O(n) \). A useable version of the algorithm is detailed by Stewart (1980). The complete algorithm requires successively multiplying together \( n-1 \) reflections. For \( y \) a vector in \( \mathbb{R}^n \), the product \( (I - 2x^tx)y^t \) can be computed in order \( n \) operations (just compute the inner product \( xy^t \), then multiply \( x^t \) by this number and add to \( y^t \)). Thus for \( \Gamma \in O(n) \) the product \( (I - 2x^tx)\Gamma \) can be computed in order \( n^2 \). 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.

C. 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.

Random Givens rotations. As is well known, any orthogonal matrix can be written uniquely as a product of \( 2 \times 2 \) rotations (Givens rotations) of the form \( \begin{pmatrix} c & s \\ -s & c \end{pmatrix} \), see Golub and Van Loan (1983). Several authors have shown how
to generate a random orthogonal matrix by multiplying together \(^n\choose 2\) such rotations. The cosines are chosen independently from beta densities. Preliminary computations show this is an order \(n^4\) 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 (1985). Vilenkin (1968, 9.1) gives the result in group theoretic language.

**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 (1986). 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 (1968, Chapter 9)), or in the language of Lie theory (Weyl 1946, Chapter 7)).

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 anti-symmetric real \(n \times n\) matrices. A 1-1 transformation \(\tau\) from \(O'(n)\) to \(K_n\) is defined by

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

with \(\tau^{-1}\) 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 Caley's parametrization. See Weyl (1946, pg. 56) for further discussion.

The skew-symmetric matrices can be naturally identified with \(^n\choose 2\) dimensional Euclidean space. It can be shown that Haar measure on \(O(n)\) has density
\[(3.3) \quad cdk \frac{n-1}{\det(I+kk')}^{1/2} \]

For an explicit constant \(c\). Some formulas can be found in Olkin (1951).

Here is one possible use of formula (3.3): for \(\Gamma\) chosen from Haar measure on \(O(n)\), define

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

For large \(n\), the coordinates of \(\Gamma\) are approximately independent mean zero normal variables with variance \(1/n\). It follows that the coordinates of \(K\) are approximately independent normal variables, with mean zero, and variance \(4/n\).

This suggests choosing \(K\) with such normal coordinates, and basing a rejection method on the density (3.2).

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 \(\Gamma\) 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 \(\Gamma\) in the weak star topology (on \(R\)).

Diaconis and Newman (1984) showed that in this sense, Haar measure can be approximated by a matrix \(X\) of independent normal random variables having mean zero and variance \(1/n:\)

\[\sup_{A: \|A\|^2=n} \text{Tr}(A\Gamma) \rightarrow n(0,1)\]

where the \(\sup\) is over all \(n \times n\) matrices having \(\text{Tr}(A^TA) = n\), 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 (1984)). The first $2n$ moments of $\text{Tr}(X)$ equal the first $2n$ moments of $\text{Tr}(\Gamma)$ (Diaconis and Mallows (1984)).

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: $I - 2x^t x$ with $x$ chosen uniformly on the $n$-sphere. Diaconis and Shahshahani (1986b) 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 order $n$ operations to multiply by a reflection, this algorithm requires of order $n^3 \log n$ steps, and will be slower than Stewart's algorithm for large $n$.


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 extend Stewart's algorithm to the unitary and symplectic groups. Finally, we give results delimiting 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 metrisable, separable, and complete ($G$ is "Polish"). This includes all examples of practical interest. Any unexplained terminology may be found in Hewitt and Ross (1963, 1970).
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 $\pi : G \to X$. To choose coset representatives, let $\phi : X \to G$ be a measurable inverse of $\pi$ (so $\pi \phi(x) = x$). The existence of $\phi$ under our hypothesis follows from Theorem 1 in Bondar (1976).

Define $T : G \to (X,H)$ by $T(g) = (\pi(g), \eta(g))$ where $g$ has the unique representation $g = \phi(\pi(g)) \eta(g)$ (so $\eta(g) = \phi(\pi(g))^{-1} g$). The map $T$ is a 1-1, onto, bi-measurable map with inverse $T^{-1}(x,h) = \phi(x)h$. The basic fact may be stated as follows:

**Lemma 4.1.** Let $P_G$ be Haar measure on $G$. The image measure under $T$ is a product measure

$$P_G T^{-1} = Q \times P_H$$

with

$$P_H = \text{Haar measure on } H \quad Q = P_G \pi^{-1} \text{ on } X.$$

**Proof:** For $A \subset H$, $B \subset H$, define

$$\mu(A \times B) = P_G T^{-1}(A \times B) = P_G \{ g : \pi(g) \in A, \eta(g) \in B \}.$$  

From the definitions, $T^{-1}(A \times B)h = T^{-1}(A \times Bh)$ for $h \in H$. This implies that for fixed $A$, $\mu(A \times \cdot)$ is a right $H$ invariant measure on $H$. By uniqueness of Haar measure, it equals $Q(A)P_H(\cdot)$. For some real constant $Q(A)$. This may be shown to be a $\sigma$-additive probability with is evidently equal to $P_G \pi^{-1}$. □

**Corollary 1.** For every continuous function $f$ on $G$,

$$\int_G f(g) \, P_G(dg) = \int_X \int_H f(\phi(x)h) \, Q(dx) \, P_H(dh)$$
Proof: Consider $\mathbf{f}$ as the indicator of a set of the form $T^{-1}(A \times B)$. The left side becomes $P_G\{g: \pi(g) \in A, \eta(g) \in B\}$. The right side becomes $Q \times P_H\{(x, h): \pi \phi(x) \in A, h \in B\}$. These are evidently equal. 

The next corollary is the rigorous version of the subgroup algorithm:

Corollary 2. Let $P_X$ be the image of $Q$ under $\phi$. Regard $P_H$ as a probability on $G$. Then

$$P_G = P_X \ast P_H.$$ 

Above, we used the fact that $\phi$ was an inverse of $\pi$. This is not crucial, as shown by the next result.

Corollary 3. Let $\tilde{\phi}$ be a map from $X \to G$ with the property that $\pi \tilde{\phi}$ is a measure preserving transformation from $X \times X$, where $X$ has the uniform distribution $Q$ defined above. Let $\tilde{P}_X$ be the image of $Q$ under $\tilde{\phi}$. Then

$$P_G = \tilde{P}_X \ast P_H.$$ 

Proof: What must be verified is

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

This follows from Corollary 1. 

Remarks. (1) Lemma 4.1 is a special case of much more general theorems such as Theorem 2 in Bondar (1976). 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 (1983) or Kehlet (1984).

Example (The orthogonal group). Consider O(n) as in Section 4 with O(n-1) = \{T: Te_1^t = e_1\} and X = \mathbb{S}_{n}, the unit sphere. Then \( \pi(T) = Te_1^t \). The map
\[
\begin{cases}
\phi(x) = I - 2V^tV/C , \text{ with } V = x + e_1, \ c = VV^T \\
\phi(-e_1) = -I
\end{cases}
\]
is a measurable inverse of \( \pi \) that is continuous except at \(-e_1\) (there is no continuous choice of coset representatives). Corollary 2 gives a rigorous justification for Lemma 3.1.

Remarks. (1) Here is an application of Corollary 3 to this problem.
Consider the map \( \tilde{\phi}: \mathbb{S}_{n} \to O(n) \) defined by \( \tilde{\phi}(x) = I - 2x^tx \). This does not serve as an inverse of \( \pi \). But, at least in two-dimensions, if \( x \) is uniformly distributed on \( \mathbb{S}_{n} \), \( \tilde{\phi}(x)e_1^t \) is uniformly distributed on \( \mathbb{S}_{n} \). It follows that \( \pm(I - 2x^tx) \) is uniform on \( O(2) \). We do not know if this goes through for higher values of \( n \).

(2) The subgroup algorithm in \( O(3) \) may be described in words as follows: make a random rotation fixing the north pole (the vector \( e_1^t \)) 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 this construction is badly wrong: Kendall and Moran (1963, pg. 95) show that the angle about the fixed point should have density proportional to \( \sin^2 \theta \), 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 conjugate transpose. There is a natural tower

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

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

$$ M = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & & & \\
& & \ddots & \\
& & & M_{n-1}
\end{pmatrix}, \quad \text{with} \quad M_{n-1}^* M_{n-1} = I_{n-1,n-1}. $$

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

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

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

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

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

It is easy to choose a random point on $\mathcal{A}_n$ 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 $1/2$. If $\tilde{Z} = (Z_1, Z_2, \ldots, Z_n)$ has independent standard complex normal coordinates, then $\tilde{Z}/\tilde{Z}^*$ is uniform on $\mathcal{A}_n$. 

19
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 $\mathcal{Sp}(n)$. 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 $\mathcal{Sp}(n)$ may be found in Chevalley (1946, Chapter 1).

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

$$M^*M = I$$

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

$$\mathcal{Sp}(n) \supset \mathcal{Sp}(n-1) \supset \ldots \supset \mathcal{Sp}(2)$$

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

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 $u = x^*q e_1$.

A routine computation shows that the matrix

$$(4.1) \quad M = -q\left(I - \frac{2u^t\bar{u}}{uu^*}\right)$$

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

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:
\[ x = (Z_1, \ldots, Z_n)/\mathbb{Z}^* \]

\[ Z_i = A_i + iB_i + jC_i + hD_i, \quad A_i, B_i, C_i, D_i \text{ independent, mean zero, variance } 1/4. \]

Then form \( M \) as in (4.1). Multiply this by

\[
\begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & & & \\
& & \ddots & \\
0 & & & M_1
\end{pmatrix}
\]

with \( M_1^* M_1 = I_{n-1, n-1} \).

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

\[ G \supset G_1 \supset G_2 \supset \ldots \]

where \( G_i = \{ g \in G : g = (0, 0, \ldots, 0, *, *, *, \ldots) \} \), 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 \( P(X_i = 1) = 1/2 = P(X_i = 0) \).

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 non-trivial 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 \{id\}. Clearly a group with small subgroups contains non-trivial closed subgroups.

A famous theorem of Gleason (1952) 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 (1966): If \( G \) is Abelian, then the connected component of the identity is a finite dimensional torus which certainly has non-trivial
closed subgroups, hence $G$ does. If $G$ is not Abelian, then its maximal torus is a non-trivial closed subgroup.

For finite groups, observe that a finite group with no proper subgroups is $Z_p^*$, the residues modulo a prime. This is the only case where the subgroup algorithm does not work to produce a non-trivial factorization of the uniform distribution. It is easy to see that the uniform distribution does not factor on $Z_2$ or $Z_3$. One approach uses the result that $1 + z$ and $1 + z + z^2$ are irreducible over the reals. A non-trivial factorization of $U$ would lead to a factorization of the associated polynomial.

To complete the discussion we prove:

Lemma 4.2. Let $p > 5$ be a prime. Let $Z_p$ be the integers mod $p$. Then, there exist non-uniform probabilities $P_1, P_2$ on $Z_p$, such that

$$P_1 * P_2 = U.$$  

Proof: For $i = 1, 2, \ldots, \frac{p-1}{2}$, 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

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

22
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 \( k \neq 0 \), \( \rho_{\pm k}(Q_k) = 0 \), \( c_0(Q_k) = 1 \). Now let signed measures \( R_1 \) and \( R_2 \) be defined by
\[
R_1 = \sum_{i=1}^{a} Q_i, \quad R_2 = \sum_{i=a+1}^{(p-1)/2} Q_i \quad \text{for fixed } 1 \leq a \leq (p-1)/2.
\]

Finally, for sufficiently small \( \epsilon \) 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 \( U = P_1 \ast P_2 \). Indeed, for \( k \neq 0 \), \( \rho_k(P_1 \ast P_2) = \rho_k(P_1) \rho_k(P_2) = 0 \). To show that the decomposition is non-trivial, it suffices to show that \( R_i \) are non-zero, \( i = 1, 2 \). This follows from the fact that for \( k \neq j \), \( \rho_j(Q_k) \neq 0 \). Indeed,
\[
\rho_j(Q_k) = a_k + 2b_k \cos\left(\frac{2\pi kj}{p}\right) = \frac{1}{1 - \cos(2\pi k^2)} \left\{ -\cos\left(\frac{2\pi k^2}{p}\right) + \cos\left(\frac{2\pi jk}{p}\right) \right\}.
\]

This is zero if and only if \( j = k \).

\( \square \)

Remarks. (1) This type of factorization can be extended to non-commutative groups, see Diaconis and Shahshahani (1986a).

(2) Factoring the uniform distribution on \( \mathbb{Z}_p \) 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 (1972) show that it is impossible to find probabilities \( P_1 \) and \( P_2 \) on the set \( \{1,2,\ldots,n\} \) such that \( P_1 \ast P_2 \) is the uniform distribution on \( \{2,\ldots,2n\} \). A related result asks for a decomposition of the uniform distribution on the set \( 0, 1, 2, \ldots, n \). Lukacs
(1970), pp. 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 \((\text{mod } n)\). In the dice result, both factors must be supported on \( \{1,\ldots,n\} \) while the uniform distribution is on \( \{2,\ldots,n\} \). 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 \( U = P_1 * P_2 \), with \( P_1 \) not uniform, if and only if \( G \neq Z_2 \) or \( Z_3 \).

**References**


