GENERATION OF RANDOM ORTHOGONAL MATRICES

by

T. W. Anderson, I. Olkin, L.G. Underhill

Technical Report No. 216
August 1985

Prepared under the Auspices
of
National Science Foundation
DMS 84-11411
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California

GENERATION OF RANDOM ORTHOGONAL MATRICES

T. W. Anderson, I. Olkin and L.G. Underhill

Stanford University, Stanford University, University of Cape Town

ABSTRACT

In order to generate a random orthogonal matrix distributed according to Haar measure over the orthogonal group it is natural to start with a matrix of normal random variables and then factor it by the singular value decomposition. A more efficient method is obtained by using Householder transformations. We propose another alternative based on the product of $n(n-1)/2$ orthogonal matrices, each of which represents an angle of rotation. Some numerical comparisons of alternative methods are made.

. Key words and phrases: Simulation, Monte Carlo, Haar measure, random normal numbers

1. Introduction.

A number of approaches have been used in generating random orthogonal matrices distributed according to the Haar measure over the orthogonal group.

As a starting point let $X = (x_{ij})$ be a $p \times p$ matrix consisting of p^2 independent standard normal random variables $x_{ij}, i, j = 1, \ldots, p$. An application of the Gram-Schmidt orthogonalization procedure amounts to finding an upper triangular matrix R with positive diagonal elements such that $X = QR$. This method is called the QR factorization of X, and the matrix Q is the desired orthogonal matrix.

From another point of view we can choose pn independent standard normal random variables $x_{ij}, i = 1, \ldots, p; j = 1, \ldots, n$ from which we form the $p \times n$ matrix $X = (x_{ij})$. An orthogonal matrix G is now obtained from the eigenvalue decomposition

$$XX' = GD_\ell G', \tag{1}$$

where $D_\ell = \mathrm{diag}(\ell_1, \ldots, \ell_p)$ is a diagonal matrix with elements $\ell_1, ..., \ell_p$ which are the characteristic roots of XX'. Alternatively, G can be obtained from the singular value decomposition

$$X = GD_\ell^{1/2}H, \tag{2}$$

where H is a $p \times n$ matrix satisfying $HH' = I_p$.

Another characterization of a random orthogonal matrix is obtained from a skew-symmetric matrix S:

$$G = D_\varepsilon(I + S)^{-1}(I - S), \tag{3}$$

where $D_\varepsilon = \mathrm{diag}(\varepsilon_1, ..., \varepsilon_n), \varepsilon_i = \pm 1$. (The matrix D_ε is added in order to guarantee that $\det(I + D_\varepsilon G) \neq 0$.) Although this approach could be used to generate random orthogonal matrices G, it requires first generating random skew-symmetric matrices S with joint density function proportional to

$$\det(I + S)^{-a}, \tag{4}$$

Olkin (1951). However, there does not seem to be a simple way to generate such random skew-symmetric matrices. Thus, it appears to be easier to generate random orthogonal matrices than random skew-symmetric matrices.

Another characterization is obtained directly from the definition: an $n \times n$ orthogonal matrix is an ordered set of n rows and of n-dimensional vectors, each of which is defined over the surface of a unit n-dimensional sphere, and each vector is orthogonal to each of the rows preceding it. This definition was the basis of an algorithm by Heiberger (1978). (See Tanner and Thisted (1982) for additional commentary.) The algorithm makes use of independent standard normal random variables.

A characterization based on Householder transformations was used by Stewart (1980). A symmetric orthogonal matrix is defined by

$$H = I - 2\frac{uu'}{\|u\|^2}, \tag{5}$$

where u is a nonzero vector and $\|\cdot\|$ is the Euclidean norm. For any x, there is a Householder transformation H_x such that

$$H_x x = r(1,0,...,0)', \quad r = \pm\|x\|. \tag{6}$$

Stewart's method is based on choosing independent $N(0,\sigma^2)$ random vectors $x_1 \in R^n, x_2 \in R^{n-1}, ..., x_{n-1} \in R^2$. For $j = 1, ..., n-1$, let $\bar{H}_{x_j}$ be the Householder transformation (6), and define the $n \times n$ matrix

$$H_j = \text{diag}(I_{j-1}, \bar{H}).$$

The product

$$Q = H_1 H_2 \cdots H_{n-1} D_\varepsilon \tag{7}$$

is a random orthogonal matrix, where $D_\varepsilon = \text{diag}(\varepsilon_1, \cdots, \varepsilon_n)$, and $\varepsilon_i = \pm 1, i = 1, \cdots, n$.

We propose yet another characterization of an orthogonal matrix in terms of $n(n-1)/2$ rotations plus reflections, which is formally equivalent to (7). The differences are centered on how to generate the random variables. More specifically, define the Givens matrix

$$(8) \qquad G_{ij} \equiv G_{ij}(\theta_{ij}) = \begin{array}{cc} & \begin{array}{ccccc} & i & & j & \end{array} \\ \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array} & \begin{pmatrix} I & 0 & 0 & 0 & 0 \\ 0 & \cos\theta_{ij} & 0 & -\sin\theta_{ij} & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & \sin\theta_{ij} & 0 & \cos\theta_{ij} & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix} \end{array},$$

where $-\pi/2 \leq \theta_{ij} \leq \pi/2$. Then every orthogonal matrix G can be written as

$$(9) \qquad G = (G_{12}G_{13} \cdots G_{1n})(G_{23} \cdots G_{2n}) \cdots (G_{n-1,n})D_\varepsilon,$$

where $D_\varepsilon = \text{diag}(\varepsilon_1, ..., \varepsilon_n)$, $\varepsilon_i = \pm 1, i = 1, ..., n$. The θ_{ij} represent the angles of the rotation and the matrix D_ε represents reflections.

Although the representation (9) is generally well-known, it has not been used very often in this form. Because of this we give a derivation in the Appendix. Tumura (1965) uses this representation to provide a derivation of the noncentral distribution of the characteristic roots of a Wishart matrix, thereby providing an elementary development of zonal polynomials. (There is an error in this paper in the definition of G; we have not checked whether the subsequent methodology is correct.)

To generate a random orthogonal matrix choose D_ε at random, i.e.

$$P\{\varepsilon_i = +1\} = P\{\varepsilon_i = -1\} = 1/2, \quad i = 1, ..., n,$$

and let the $n(n-1)/2$ angles θ_{ij} be mutually independent with joint density function which is proportional to

$$(10) \qquad \left(\prod_{j=2}^{n} \cos^{j-2}\theta_{1j}\right)\left(\prod_{j=3}^{n} \cos^{j-3}\theta_{2j}\right) \cdots \left(\prod_{j=n}^{n} \cos\theta_{n-1,j}^{j-n}\right).$$

Once the angles and ε's have been chosen, they are combined as in (8) and (9).

To carry out this program, we need to generate random angles according to the density (10). The density of an angle α is of the form

$$(11) \qquad 2\frac{cos^{a-1}\alpha \quad sin^{b-1}\alpha}{B(\frac{a}{2},\frac{b}{2})}, \quad 0 \leq \alpha \leq \pi/2,$$

from which the density of x defined by $\sqrt{x} = \cos\alpha$ ($\sin\alpha = \pm\sqrt{1-x}$) is a beta distribution

$$(12) \qquad \beta(x; \frac{a}{2}, \frac{b}{2}) = \frac{x^{\frac{1}{2}a-1}(1-x)^{\frac{1}{2}b-1}}{B(\frac{a}{2},\frac{b}{2})}, \; 0 \leq x \leq 1.$$

Thus, the procedure requires generation of random variables x with density function (12). But this is standard by virtue of the representation $x = u/(u+v)$, where u and v are independent, with u and v having chi-square distributions with a and b degrees of freedom respectively. Of course, chi-square random variables can be generated directly from random normal variables or from uniform random variables.

More specifically, from (12) for fixed $i = 1,\ldots,n-1$, we need to generatate $n-i$ independent random variables having respective $\beta(x;\frac{1}{2},\frac{1}{2}), \beta(x;\frac{2}{2},\frac{1}{2}), \cdots, \beta(x;\frac{n-i}{2},\frac{1}{2})$ densities. A critical simplification occurs from the fact that if x_1, x_2 and x_3, are independent chi-square random variables with a, b and c degrees of freedom respectively, then $x_1/(x_1+x_2)$ is independent of $(x_1+x_2)/(x_1+x_2+x_3)$. Consequently, we begin our process by generating $y_1 = x_1/(x_1+x_2)$ and then using the denominator to generate $y_3 = (x_1+x_2+x_3)/(x_1+x_2+x_3+x_4)$. The process then continues. This means that for each successive y_j we need only one additional chi-square variate. This method provides considerable simplification and rapidity in the generation of the required random angles.

Numerical Results.

To test this procedure we carried out a simulation using 1000 replications for each case, $n = 4(2)12$. Random orthogonal matrices were generated by the method of Householder transformations (HT) and the method of successive rotations (ROT). In each case the polar method (see Knuth, 1969, p. 104) was used to generate the random normal numbers. The execution times in CPU seconds were

n	4	6	8	10	12
HT	31.3	84.9	187.2	340.4	601.0
ROT	30.8	69.2	136.8	207.3	313.8

For each n and for each method we calculate the first four sample moments (over replications) of the entry in each of the n^2 positions, and compare these with the expectations

$Ev_i = 0, Ev_i^2 = 1/n, Ev_i^3 = 0, Ev_i^4 = 3/n(n+2)$. The number of entry positions within k standard errors is listed in Table 1. Although the n^2 elements of an orthogonal matrix are dependent, we do not take this fact into account in the tabulation.

Table 1. Number of Moment Entries in Standard Error Intervals

Householder Method

k	Ev				Ev^2				Ev^3				Ev^4			
n	0-1	1-2	2-3	> 3	0-1	1-2	2-3	> 3	0-1	1-2	2-3	> 3	0-1	1-2	2-3	>3
4	14	2	0	0	12	13	1	0	15	3	1	0	12	3	1	0
6	23	11	2	0	22	9	5	0	23	13	0	0	20	13	3	0
8	39	21	4	0	45	13	6	0	46	14	4	0	43	18	1	2
10	65	31	4	0	60	36	4	0	66	29	5	0	63	32	5	0
12	100	37	6	1	98	38	7	1	104	32	6	2	96	41	6	1

Successive Rotations Method

k	Ev				Ev^2				Ev^3				Ev^4			
n	0-1	1-2	2-3	> 3	0-1	1-2	2-3	> 3	0-1	1-2	2-3	> 3	0-1	1-2	2-3	>3
4	11	3	2	0	9	7	0	0	10	4	2	0	10	6	0	0
6	25	9	2	0	24	7	4	1	26	7	3	0	26	6	3	1
8	43	18	3	0	45	17	2	0	41	20	3	0	46	17	1	0
10	71	26	2	1	60	39	1	0	74	19	6	1	70	27	3	0
12	93	45	5	1	93	40	10	1	98	38	8	0	95	41	7	1

Appendix

A.1. Construction of an Orthogonal Matrix.

Let $G = (g_{ij})$ be an arbitrary orthogonal matrix. We shall show how G determines uniquely the representation (9). Let $\Gamma_{ij} = G'_{ij}$. First find θ_{12} $(-\pi/2 < \theta_{12} \le \pi/2)$ so

$$(\Gamma_{12}G)_{21} = -\sin\theta_{12}\, g_{11} + \cos\theta_{12}\, g_{21} = 0. \tag{14}$$

Then $\cos\theta_{12} \ge 0$ and $\sin\theta_{12}$ has the sign of g_{21}/g_{11}. (If $g_{11} = 0$, let $\theta_{12} = \pi/2$.) The matrix $\Gamma_{1j}, j = 3, \ldots, n$, is constructed $(-\pi/2 < \theta_{ij} \le \pi/2)$ so

$$\begin{aligned} &(\Gamma_{1j} \cdot \Gamma_{1,j-1} \cdot \ldots \cdot \Gamma_{12}G)_{j1} \\ &\quad = -\sin\theta_{1j}\,(\Gamma_{1,j-1} \cdot \ldots \cdot \Gamma_{12}G)_{11} + \cos\theta_{1j}\,(\Gamma_{1,j-1} \cdot \ldots \cdot \Gamma_{12}G)_{j1} = 0. \end{aligned} \tag{15}$$

Then $\cos\theta_{1j} \geq 0$. The matrix $\Gamma_{1n} \cdot \ldots \cdot \Gamma_{12} G$ has 0's in the first column, except the first element which is ± 1 (since the matrix is orthogonal). That fact implies that the elements in the first row are 0 except the first element. We define the scalar ε_1 and the $(n-1) \times (n-1)$ matrix $G^{(2)}$ by

$$\Gamma_{1n} \cdot \ldots \cdot \Gamma_{12} G = \begin{bmatrix} \varepsilon_1 & 0 \\ 0 & G^{(2)} \end{bmatrix}. \tag{16}$$

Then $G^{(2)}$ is orthogonal.

At the j-th stage

$$\Gamma_{j-1,n} \cdot \ldots \cdot \Gamma_{j-1,j} \cdot \Gamma_{j-2,n} \cdot \ldots \cdot \Gamma_{12} G = \begin{bmatrix} D^{(j-1)} & 0 \\ 0 & G^{(j)} \end{bmatrix}, \tag{17}$$

where D_1 is a $(j-1) \times (j-1)$ diagonal matrix with $\varepsilon_1, \ldots, \varepsilon_{j-1}$ as diagonal elements and $G^{(j)}$ is an $(n-j+1) \times (n-j+1)$ orthogonal matrix. Note that $\Gamma_{j,j+1}, \ldots, \Gamma_{n-1,n}$ are of the form

$$\begin{bmatrix} I_{j-1} & 0 \\ 0 & \Gamma^{(j)} \end{bmatrix}, \tag{18}$$

where the orthogonal $\Gamma^{(j)}$ has order $n-j+1$. Then $\theta_{j,j+1}, \ldots, \theta_{jn}$ are taken as at the first stage to make the first column of $\Gamma^{(j)}_{jn} \cdot \ldots \cdot \Gamma^{(j)}_{j,j+1} G^{(j)}$ consist of 0's. Thus

$$\Gamma^{(j)}_{jn} \cdot \ldots \cdot \Gamma^{(j)}_{j,j+1} G^{(j)} = \begin{bmatrix} \varepsilon_j & 0 \\ 0 & G^{(j+1)} \end{bmatrix}. \tag{19}$$

Finally, $G^{(n)} = \varepsilon_n$. Thus

$$\Gamma_{n-1,n} \cdot \ldots \cdot \Gamma_{12} G = D_\varepsilon. \tag{20}$$

The representation (9) is obtained from (20) with $G_{ij} = \Gamma'_{ij}$.

A.2. Representation of a Random Orthogonal Matrix.

The Haar measure of G is characterized by the property that ΓG has the same measure as G for every orthogonal Γ, or equivalently that $G\Gamma$ has the same measure as G for every orthogonal Γ. It follows that any column or any row of G has the uniform distribution over the n-sphere. If Γ_2 is an orthogonal matrix of order $n-1$, then from (16) we see that

$$\Gamma_{1n} \cdot \ldots \cdot \Gamma_{12} G \begin{pmatrix} 1 & 0 \\ 0 & \Gamma_2 \end{pmatrix} = \begin{bmatrix} \varepsilon_1 & 0 \\ 0 & G^{(2)}\Gamma_2 \end{bmatrix} \tag{21}$$

has the same measure as (16) because the first column of G (on which $\Gamma_{1n}, \ldots, \Gamma_{12}$ depend) is the same as the first column of $G \begin{pmatrix} 1 & 0 \\ 0 & \Gamma_2 \end{pmatrix}$. Hence $G^{(2)}\Gamma_2$ has the same measure as $G^{(2)}$ for every orthogonal Γ_2. Moreover, $\Gamma_{1n}, \ldots, \Gamma_{12}, \varepsilon_1$, and $G^{(2)}$ are independent.

The first column of $G_{1n} \cdot \ldots \cdot G_{12}$ is the transpose of

$$(22) \qquad \left(\prod_{j=2}^{n} \cos\theta_{1j},\ \sin\theta_{12} \prod_{j=3}^{n} \cos\theta_{1j}\ \sin\theta_{13} \prod_{j=4}^{n} \cos\theta_{1j}, \ldots, \sin\theta_{1,n-1}\cos\theta_{1n},\ \sin\theta_{1n}\right).$$

It consists of the polar coordinates of the first column of G. The density of $\theta_{12}, \ldots, \theta_{1n}$ is a constant times

$$\cos^{n-2}\theta_{1n} \cdot \cos\theta_{1,n-1}^{n-3} \cdot \ldots \cdot \cos\theta_{13}$$

(see e.g., Anderson, 1984, problem 4, Chapter 7). The scalar $\varepsilon_1 = -1$ and $+1$ with probability 1/2.

In turn the distribution of $\theta_{j,j+1}, \ldots, \theta_{jn}$, and ε_j are found.

$$(22) \qquad \left(\prod_{j=2}^{n} \cos\theta_{1j}, \sin\theta_{12} \prod_{j=3}^{n} \cos\theta_{1j} \sin\theta_{13}, \prod_{j=4}^{n} \cos\theta_{1j}, \ldots, \sin\theta_{1,n-1}\cos\theta_{1n}, \sin\theta_{1n}\right).$$

References

Anderson, T. W. (1984). **An Introduction to Multivariate Statistical Analysis.** (second edition). John Wiley & Sons, Inc.: New York.

Heiberger, R. M. (1978). Algorithm AS 127. Generation of random orthogonal matrices. **Applied Statistics**, 27, 199–206.

Heiberger, R. M., Velleman, P. F., and Ypelaar, M. A. (1983). Generating test data for independently controllable features for multivariate general linear forms. **Journal of the American Statistical Association**, 78, 585–595.

Knuth, D. (1981). **Seminumerical Algorithms: The Art of Comuter Programming** (Second Edition). Addison-Wesley: New York.

Olkin, I. (1951). On distribution problems in multivariate analysis. Doctoral Dissertation, University of North Carolina.

Stewart, G. W. (1980). The efficient generation of random orthogonal matrices with an application to condition estimators. **SIAM J. Numer. Anal.**, 17, 403–409.

Tanner, M. A. and Thisted, R. A. (1981). **Applied Statistics**, 31, 190–192.

Tumura, Y. (1965). The distributions of latent roots and vectors. **T.R.U. Math**, 1, 1–16.

Velleman, P. F. and Ypelaar, M. A. (1980). Constructing regressions with controlled features: a method for probing regression performance. **Journal of the American Statistical Association**, 75, 839–844.