THE GENERATION OF A RANDOM SAMPLE-COVARIANCE MATRIX

by Alan H. Feiveson
Manned Spacecraft Center
Houston, Texas

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JANUARY 1966
THE GENERATION OF A RANDOM SAMPLE-COVARIANCE MATRIX

By Alan H. Feiveson

Manned Spacecraft Center
Houston, Texas
ABSTRACT

In simulating trajectory estimation problems, a rapid procedure is desirable for generating random sample-covariance matrices based on large numbers of observations. By using existing random-number generators, an economical method is developed that yields a matrix $S^*$ whose elements have the same joint distribution as the elements of the sample-covariance matrix $S$. 
THE GENERATION OF A RANDOM SAMPLE-COVARIANCE MATRIX

By Alan H. Feiveson
Manned Spacecraft Center

SUMMARY

Trajectory estimation simulation problems make desirable a rapid procedure for generating random sample-covariance matrices based on large numbers of observations. This paper first presents an algorithm for such a procedure and then shows its derivation from the Cochran-Fisher Theorem concerning quadratic forms. Finally, an example is given.

INTRODUCTION

In trajectory analysis, the "best" estimate of the state is a function of the covariance matrices $R_i$ associated with the observation stations. For practical use, estimates must be substituted for the unknown exact $R_i$. In some cases, estimating the $R_i$ directly from the observations may be desirable.

The well-known "best", or unbiased-maximum-likelihood-based (u.m.l.b.), estimator of a covariance matrix $R_i$ is given by

$$ S = \frac{1}{n-1} \sum_{i=1}^{n} \left( X_i - \bar{X} \right) \left( X_i - \bar{X}^T \right) $$

where the $X_i$ are the observation vectors and $n$ is the sample size. To simulate a procedure where u.m.l.b. estimates are used, random matrices must be generated that have the same distribution as these estimates.

The obvious method of generating a matrix $S^*$, having the same distribution as $S$, is to generate the $n$ observation vectors $\{X_i; \ i = 1, \ldots, n\}$. But if each vector $X_i$ has $p$ components, generating $n$ observation vectors necessitates generating at least $np$ random numbers.
This paper presents an alternate method of generating $S^*$ which requires using only $p(p + 1)/2$ random numbers - usually a much smaller quantity than $np$.

**SYMBOLS**

- $A_i$: matrices in Cochran's Theorem
- $b_{ij}$: $i,j$th element of $B$
- $b^*_{ij}$: $i,j$th element of $B^*$
- $C^T$: transpose of the matrix $C$
- $I$: identity matrix
- $i, j, k$: indices of summation
- $N(\emptyset, R)$: normally distributed with mean $\emptyset$ and covariance matrix $R$
- $N_i, N_{ij}$: standardized normal random variates
- $n$: sample size
- $p$: size of covariance matrix (number of variables in one observation)
- $Q$: matrix equal to $I - \sum_{k=1}^{j-1} Q_k$
- $Q_i$: matrix equal to $v_i^T v_i / v_i v_i^T$
- $r_j$: $j$th row of matrix $W$
- $r_j^T$: transpose of $r_j$
\( t_k^T \)  
transpose of \( t_k \)

\( v_j \)  
random variable

\( w_{ij} \)  
\( ij^{th} \) element of \( W \)

\( x \)  
1 x (n - 1) random vector in Cochran's Theorem

\( y_j \)  
\( j^{th} \) of a set of orthogonal 1 x (n - 1) vectors

\( y_j^T \)  
transpose of \( y_j \)

\( z_k, t_k \)  
p x 1 vectors

\( \chi^2 (n - j) \)  
chi-square with \( n - j \) degrees of freedom

\( \nu_i \)  
rank of \( A_i \)

\( \emptyset \)  
p x 1 null vector

\( \sim \)  
is distributed as

**METHOD**

Let \( S = A/(n - 1) \) be the u.m.l.b. estimator of a p x p covariance matrix \( R \) from an independent normally distributed sample of size \( n \). It can be shown (ref. 1) that

\[
A = \sum_{k=1}^{n-1} z_k z_k^T \tag{2}
\]

where the p x 1 vectors \( \{z_k; k = 1, 2, \ldots n - 1\} \) are independent and normally distributed with zero mean and covariance matrix \( R \).
Since $R$ is a covariance matrix, it is semipositive definite. Therefore, a matrix $C$ exists such that

$$CC^T = R \quad (3)$$

It follows that the vector $z_k$ can be written

$$z_k = Ct_k \quad (4)$$

where

$$t_k \sim N(0, I)$$

Let

$$B = \{b_{ij}\} = \sum_{k=1}^{n-1} t_k t_k^T \quad (5)$$

Then,

$$CBC^T = C \sum_{k=1}^{n-1} t_k t_k^T C^T = A \quad (6)$$

**Generation of $A^*$**

Let $A^*$ be a generated matrix whose elements have the same joint distribution as those of $A$. To obtain $S^* = A^*/(n - 1)$, it is necessary only to generate a matrix $B^*$ whose elements are distributed as the elements of $B$. Then, $A^*$ is computed so that

$$A^* = CB^* C^T \quad (7)$$
Hence, the problem is reduced to generating the random symmetric matrix $B^*$. An algorithm for generating $B^*$ is given below. For a justification of this procedure, refer to the Analysis.

**Generation of $B^*$**

1. Generate $p$ independent $\chi^2$ variables $v_j$, $j = 1, \ldots, p$, having $n - j$ degrees of freedom. One method of obtaining $v_j$ is to generate a standard normal variate $N_j$ and substitute it into the Wilson-Hilferty $\chi^2$ approximation (ref. 2). The approximation can be written

$$v_j \approx (n - j) \left[ 1 - \frac{2}{9(n - j)} + N_j \sqrt{\frac{2}{9(n - j)}} \right]^3$$

2. Generate $p(p - 1)/2$ independent standard normal variates $N_{ij}$, $i < j$, and $j = 1, 2, \ldots, p$.

3. Form the diagonal elements of $B^*$ \( \left\{ b^*_{jj}, j = 1, \ldots, p \right\} \) as follows:

$$b^*_{11} = v_1$$

$$b^*_{jj} = v_j + \sum_{i=1}^{j-1} N_{ij}^2 (j > 1)$$

4. Form the off-diagonal elements of $B^*$ as follows:

$$b^*_{lj} = b^*_{jl} = N_{lj} \sqrt{v_l}$$

$$b^*_{ij} = b^*_{ji} = N_{ij} \sqrt{v_i} + \sum_{k=1}^{i-1} N_{ki} N_{kj} (i > 1)$$

Once $B^*$ has been generated, $A^*$ follows from equation (7).
ANALYSIS

Using the notation of the Method section and noting that by joining the vectors $t_k$ and $k = 1, 2, ..., n - 1$ as columns, a $p \times (n-1)$ matrix $W$ can be formed

$$W = \{w_{ij}\} = \begin{pmatrix} t_1 & t_2 & \cdots & t_{n-1} \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix}$$

where $r_j$ is the $j^{th}$ $1 \times (n - 1)$ row vector of $W$. Thus, the $ij^{th}$ element of $B$, $b_{ij}$, is equal to $r_i r_j^T$.

By using the Schmidt orthogonalization process, a set of orthogonal vectors $\{v_j, j = 1, 2, \ldots, p\}$ can be generated where

$$v_j = r_j - r_j v_1^T y_1 / y_1 y_1^T - \cdots - r_j v_{j-1}^T y_{j-1} / y_{j-1} y_{j-1}^T$$

$$= r_j \left(I - Q_1 - Q_2 - \cdots - Q_{j-1}\right)$$

$$= r_j Q_j$$

where $Q_j = y_j^T y_j / y_1 y_1^T$, $Q = I - \sum_{k=1}^{j-1} Q_k$ and $I$ is the $(n - 1) \times (n - 1)$ identity matrix.

The matrices $Q, Q_1, ..., Q_{j-1}$ have the following significant properties:

1. $Q_1, Q_2, ..., Q_{j-1}$ have a rank of one.
2. $Q_i Q_j = 0$ for $i \neq j$.
3. $Q, Q_1, ..., Q_{j-1}$ are symmetric idempotents.
4. $Q$ has rank $n - j$. 
Proof

1. The vector \( y_i \) clearly spans the entire range space of \( Q_i \).

2. \( Q_i Q_j = \frac{y_i^T y_i y_j^T y_j}{y_i^T y_i (y_j^T y_j)} = 0 \) because \( y_i^T y_j^T = 0 \) for \( i \neq j \).

3. Clearly \( Q_i \) is symmetric. To show idempotence,

\[
Q_i Q_i = \frac{y_i^T \begin{pmatrix} y_i y_i^T \\ y_i^T y_i \end{pmatrix} y_i}{y_i^T y_i (y_i^T y_i)} = \frac{y_i^T y_i}{y_i^T y_i} = Q_i
\]

and

\[
QQ = \begin{pmatrix} I - Q_1 - \ldots - Q_{j-1} \\ I - Q_1 - \ldots - Q_{j-1} \end{pmatrix} \begin{pmatrix} I - Q_1 - \ldots - Q_{j-1} \\ I - Q_1 - \ldots - Q_{j-1} \end{pmatrix}
= I - 2 \left( Q_1 + \ldots + Q_{j-1} \right) + \left( Q_1 + \ldots + Q_{j-1} \right)
= I - \left( Q_1 + \ldots + Q_{j+1} \right) = Q
\]

4. This follows from elementary theorems on idempotent matrices (ref. 3). Consider the following form of the Cochran-Fisher Theorem.

Theorem

If \( x \) is a \( 1 \times (n - 1) \) random vector distributed \( N(\varnothing, I) \), and if

\[
xx^T = \sum_{i=1}^{k} xA_i x^T
\]

the rank of the sum of the \( A_i \)'s equalling the sum of the ranks of the separate \( A_i \)'s is a necessary and sufficient condition for \( xA_1 x^T \) to be distributed as central \( \chi^2 \) with \( \nu_1 \) degrees of freedom (where \( \nu_1 \) is the rank of \( A_1 \)), and for \( xA_1 x^T, xA_2 x^T, \ldots, xA_k x^T \) to be jointly independent (ref. 4).
Note that the inner product $r_j r_j^T$ can be written

$$
\begin{align*}
r_j r_j^T &= r_j^T r_j = r_j^T \left( Q + Q_1 + \ldots + Q_{j-1} \right) r_j^T \\
&= r_j^T Q r_j + \sum_{k=1}^{j-1} r_j^T Q_k r_j^T
\end{align*}
(9)
$$

Equation (9) satisfies the condition of the Theorem where the matrices $Q, Q_1, \ldots, Q_{j-1}$ play the role of the $A_i$. It therefore follows that

$$
r_j^T Q r_j = r_j^T Q r_j = r_j^T \left( r_j^T Q \right)^T = y_j y_j^T \sim \chi^2(n - j)
$$

Since the $y_j$ are mutually orthogonal and normally distributed, the quantities $y_j y_j^T$, $(j = 1, 2, \ldots, p)$, are mutually independent. They can be generated independently using random variables $v_j$, having the $\chi^2$ distribution with $n - j$ degrees of freedom.

Once the set $\{y_j y_j^T, j = 1 \ldots p\}$ is given, the quantities

$$
\sigma_{ij} = \left( r_j^T Q r_j \right)^{1/2} = \left( \frac{r_j y_i^T y_i r_j^T}{y_i y_i^T} \right)^{1/2} = \frac{r_j y_i^T}{y_i y_i^T}^{1/2}
(10)
$$

being normalized linear combinations of $N(0,1)$ variates, are themselves, $N(0,1)$ variates.

Since all the elements of the matrix $W$ are mutually independent, $\sigma_{ij}$ is independent of $\sigma_{i'j'}$ for $j \neq j'$, $i < j$, $i' < j'$. Furthermore, as a consequence of the Theorem, it is known that for $i \neq i'$, $\sigma_{ij}$ is independent.
of \( \sigma_{ij} \). Therefore, the \( p(p + 1)/2 \) quantities, \( y_jy_j^T \) and \( \sigma_{ij}(j = 1, p; i < j) \), can be generated independently, using the \( \chi^2 \) random variable \( v_j \) for \( y_jy_j^T \) and standardized normal variates \( N_{ij} \) for \( \sigma_{ij} \).

The diagonal elements of \( B^* \) are easily computed from equation (9). Let

\[
b_{ii}^* = v_i
\]

\[
b_{jj}^* = v_j + \sum_{i=1}^{j-1} N_{ij}^2 (j > 1)
\]

Since \( \sigma_{ij} \sqrt{y_iy_i^T} = r_jy_i^T \), it follows that

\[
N_{ij} \sqrt{v_i} \sim r_jy_i^T
\]

From equation (7) for \( i < j \),

\[
r_jy_i^T = r_j \left[ r_i^T - \left( \frac{r_iy_1^T}{y_1y_1^T} y_1^T \right) - \left( \frac{r_iy_2^T}{y_2y_2^T} y_2^T \right) - \ldots - \left( \frac{r_iy_{i-1}^T}{y_{i-1}y_{i-1}^T} y_{i-1}^T \right) \right]
\]

\[
\sim r_jr_i^T - \left[ \frac{N_{1i}}{\sqrt{v_1}} \left( \frac{r_jy_2^T}{y_2y_2^T} \right) + \frac{N_{2i}}{\sqrt{v_2}} \left( \frac{r_jy_2^T}{y_2y_2^T} \right) + \ldots + \frac{N_{i-1i}}{\sqrt{v_{i-1}}} \left( \frac{r_jy_{i-1}^T}{y_{i-1}y_{i-1}^T} \right) \right]
\]

\[
\sim b_{ji} = \left( \frac{N_{1i}N_{1j}}{\sqrt{v_1}} + \frac{N_{2i}N_{2j}}{\sqrt{v_2}} + \ldots + \frac{N_{i-1i}N_{i-1j}}{\sqrt{v_{i-1}}} \right)
\]
Therefore, \( b_{ij}^* = b_{ji}^* \) can be generated by

\[
\begin{align*}
  b_{ij}^* &= N_{ij} \sqrt{v_1} \\
  b_{ij}^* &= N_{ij} \sqrt{v_1} + \sum_{k=1}^{i-1} N_{ki} N_{kj}(i - 1).
\end{align*}
\]

Example

Consider the generation of \( S^* \) based on 101 observations when \( R \) is given to be

\[
\begin{bmatrix}
  .45 & -.21 & 0 \\
  -.21 & .50 & .05 \\
  0 & .05 & .25
\end{bmatrix}
\]

Then \( n = 101, p = 3, \) and \( C = \)

\[
\begin{bmatrix}
  .6 & -.3 & 0 \\
  0 & .7 & .1 \\
  0 & 0 & .5
\end{bmatrix}
\]

It is necessary to generate only 6 (instead of 606) random numbers from an \( N(0,1) \) population. They are:

\[
\begin{align*}
  N_1 &= -0.258 & N_{12} &= -0.585 \\
  N_2 &= -0.882 & N_{13} &= 0.332 \\
  N_3 &= 1.869 & N_{23} &= -0.110
\end{align*}
\]
The Wilson-Hilferty $\chi^2$ approximation gives:

\[
\begin{align*}
\chi_1^2 &= 100 \left[ 1 - \frac{2}{(9)(100)} + \frac{(-0.238) \sqrt{2}}{\sqrt{300}} \right]^3 = 96.027 \\
\chi_2^2 &= 99 \left[ 1 - \frac{2}{(9)(99)} + \frac{(-0.882) \sqrt{2}}{\sqrt{891}} \right]^3 = 86.492 \\
\chi_3^2 &= 98 \left[ 1 - \frac{2}{(9)(98)} + \frac{(-1.869) \sqrt{2}}{\sqrt{882}} \right]^3 = 125.769 
\end{align*}
\]

Finally, the procedure given in the Method section yields

\[
\begin{align*}
b_{11}^* &= 96.027 \\
b_{22}^* &= 86.492 + (-0.585)^2 = 86.835 \\
b_{33}^* &= 125.769 + (0.332)^2 + (-0.110)^2 = 125.891 \\
b_{12}^* &= -0.585 \sqrt{96.027} = -5.734 \\
b_{13}^* &= 0.332 \sqrt{96.027} = 3.250 \\
b_{23}^* &= -0.110 \sqrt{86.492} + (-0.585)(0.332) = -1.216
\end{align*}
\]

Thus,

\[
A^* = c^TBc
\]

\[
= \begin{bmatrix}
44.449 & -20.412 & 1.157 \\
-20.412 & 43.638 & 5.869 \\
1.157 & 5.869 & 31.473
\end{bmatrix}
\]
and

\[
S^* = A^*(n - 1) = \begin{bmatrix}
0.444 & -0.204 & 0.012 \\
-0.204 & 0.436 & 0.059 \\
0.012 & 0.059 & 0.315
\end{bmatrix}
\]

CONCLUDING REMARKS

This report has presented an economical method of generating a \( p \times p \) sample covariance matrix based on \( n \) observations. The method requires the generation of only \( p(p + 1)/2 \) random numbers instead of the usually much larger quantity \( np \). The matrix \( C \) referred to in the Method section may be obtained by methods readily adaptable to computers.

Manned Spacecraft Center  
National Aeronautics and Space Administration  
Houston, Texas, October 18, 1965

REFERENCES


"The aeronautical and space activities of the United States shall be conducted so as to contribute ... to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—National Aeronautics and Space Act of 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Technical information generated in connection with a NASA contract or grant and released under NASA auspices.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

TECHNICAL REPRINTS: Information derived from NASA activities and initially published in the form of journal articles.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities but not necessarily reporting the results of individual NASA-programmed scientific efforts. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C. 20546