A DATA BASED RANDOM NUMBER GENERATOR
FOR A MULTIVARIATE DISTRIBUTION -
A USERS' MANUAL

Barry A. Bodt
Malcolm S. Taylor

November 1982

US ARMY ARMAMENT RESEARCH AND DEVELOPMENT COMMAND
BALLISTIC RESEARCH LABORATORY
ABERDEEN PROVING GROUND, MARYLAND

Approved for public release; distribution unlimited.
Destroy this report when it is no longer needed. Do not return it to the originator.

Secondary distribution of this report is prohibited.

Additional copies of this report may be obtained from the National Technical Information Service, U. S. Department of Commerce, Springfield, Virginia 22161.

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

The use of trade names or manufacturer's names in this report does not constitute endorsement of any commercial product.
A DATA BASED RANDOM NUMBER GENERATOR FOR A MULTIVARIATE DISTRIBUTION - A USERS' MANUAL

Barry A. Bodt
Malcolm S. Taylor

US Army Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005

US Army Armament Research & Development Command
US Army Ballistic Research Laboratory (DRDAR-BL)
Aberdeen Proving Ground, MD 21005

November 1982

Approved for public release; distribution unlimited.

Let X be a k-dimensional random variable serving as input for a system with output Y (not necessarily of dimension k). Given X, an outcome Y or a distribution of outcomes G(Y|X) may be obtained either explicitly or implicitly. We consider here the situation in which we have a real world data set \( \{x_j\}_{j=1}^n \) and a means of simulating an outcome Y. A method for empirical random number generation based on the sample of observations of the random variable X without estimating the underlying density is discussed.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF ILLUSTRATIONS.</td>
<td>5</td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>7</td>
</tr>
<tr>
<td>II. THE ALGORITHM</td>
<td>7</td>
</tr>
<tr>
<td>III. EXAMPLES</td>
<td>9</td>
</tr>
<tr>
<td>IV. CONCLUSIONS</td>
<td>17</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>17</td>
</tr>
<tr>
<td>APPENDIX A. COMPUTER PROGRAM.</td>
<td>19</td>
</tr>
<tr>
<td>DISTRIBUTION LIST.</td>
<td>31</td>
</tr>
</tbody>
</table>
### LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Data base for a mixture of three bivariate normal distributions.</td>
<td>10</td>
</tr>
<tr>
<td>2.</td>
<td>Combined sample: Data base and pseudo-observations.</td>
<td>11</td>
</tr>
<tr>
<td>3.</td>
<td>Marginal data for four-dimensional measurements.</td>
<td>12</td>
</tr>
<tr>
<td>4.</td>
<td>Simulated behind armor debris.</td>
<td>13</td>
</tr>
<tr>
<td>5.</td>
<td>Data base for MLRS bomblets.</td>
<td>14</td>
</tr>
<tr>
<td>6.</td>
<td>Simulated MLRS bomblet data.</td>
<td>15</td>
</tr>
<tr>
<td>7.</td>
<td>Combined sample: Data base and pseudo-observations.</td>
<td>16</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

An axiom of many simulation studies is that an outcome \( Y \), or distribution of outcomes \( G(Y|X) \), of interest can be computer generated using as input experimentally derived data \( \{X_j\}_{j=1}^n \).

A commonly encountered procedure is one in which a set of experimental data is considered to be a random sample from some underlying but unknown distribution; this data is then modeled by a common statistical distribution to provide a convenient representation (with a coincident loss of information), and is then used as the basis for generating additional pseudo-observations (Monte Carlo values). The intent inherent in this procedure is that the pseudo-observations maintain the statistical structure of the original data set.

The intermediate step of modeling or "fitting" the data as a statistical distribution is sometimes unnecessary and sometimes nearly impossible. For example, for multimodal data or multivariate data, it is usually difficult and often unrealistic to attempt to characterize the data analytically. Because of this fact, there exists little, if any, guidance for the practitioner who is confronted with data of this type. Notice, however, that to serve as input for simulation, all that may actually be required is to provide pseudo-observations that exhibit the same statistical characteristics as the original data set, with no real necessity to formally characterize the underlying distribution.

It is in response to this observation that the following research was initiated and algorithm developed.

II. THE ALGORITHM

Let us consider the following situation addressed by Thompson and Taylor: \(^1\) We have a random sample \( \{X_j\}_{j=1}^n \) of size \( n \) from a multivariate distribution of dimension \( k \), and we want to generate pseudorandom vectors from the underlying, but unknown, distribution that gave rise to the random sample. Since we do not know, and usually will never know, the form of this distribution, our attack should be empirical. We shall endeavor to see to it that our pseudorandom vectors look very much like those in the original data set. In so doing, we will maintain the essential structural integrity of the problem.

We now direct our attention to the mechanics of the algorithm. After carrying out a rough rescaling to account for differing variances that may exist among the \( k \) variates, we select at random one of the \( n \) data points, say \( X_1 \), from the data base and then proceed to determine its \( m-1 \) nearest neighbors. The nearest neighbors are determined under the ordinary Euclidean

---

metric. The value of \( m \), which can best be determined after perusal of the data, will depend upon the sample size \( n \) and the characteristics of the data. A conservative estimate would be to choose \( m = n/20 \).

The vectors \( \{X_j\}_{j=1}^m \) are now coded about the sample mean \( \bar{X} = \frac{1}{m} \sum X_i \) to yield \( \{X'_j\} = \{X_j - \bar{X}\}_{j=1}^m \), and an independent random sample of size \( m \) is generated from the uniform distribution \( U(1/m - \sqrt{\frac{3(m-1)}{m^2}}, 1/m + \sqrt{\frac{3(m-1)}{m^2}}) \).

Now the linear combination

\[
X' = \sum_{k=1}^m u_k X'_k
\]

is formed, where \( \{u_k\}_{k=1}^m \) is the random sample from the \( U(1/m - \sqrt{\cdot}, 1/m + \sqrt{\cdot}) \).

Finally the translation

\[
X = X' + \bar{X}
\]

restores the relative magnitude, and \( X \) is a pseudorandom vector which we propose to be representative of the multivariate distribution that provided the \( \{X_j\}_{j=1}^n \).

To obtain the next pseudorandom vector we randomly select another of the \( n \) data points and proceed as above.

We will now attempt to advance the algorithm by considering the mathematics that suggests the mechanics that we have just outlined. Consider the distribution of \( X_1 \) and its \( m-1 \) nearest neighbors:

\[
\{(x_{1,k}, x_{2,k}, \ldots, x_{k,k})'\}_{k=1}^m = \{X_k\}_{k=1}^m.
\]

Let us suppose that this "truncated set" of random observations has mean vector \( \mu \) and covariance matrix \( \sigma \). Let \( \{u_k\}_{k=1}^m \) be an independent random sample from the uniform distribution \( U(1/m - \sqrt{\cdot}, 1/m + \sqrt{\cdot}) \). Then, \( E(u_k) = 1/m \), \( \text{Var}(u_k) = (m-1)/m^2 \), and \( \text{Cov}(u_i, u_j) = 0 \), for \( i \neq j \).

Forming the linear combination

\[
Z = \sum_{k=1}^m u_k X_k
\]
we have, for the \( r \)th component \( z_r = u_1 x_{r1} + u_2 x_{r2} + \ldots + u_m x_{rm} \), the following relations

\[
E(z_r) = m \cdot \frac{1}{m} \cdot \mu_r = \mu_r,
\]

\[
\text{Var}(z_r) = \sigma_r^2 + \frac{(m-1)}{m} \cdot \mu_r^2,
\]

\[
\text{Cov}(z_r, z_s) = \sigma_{rs} + \frac{(m-1)}{m} \cdot \mu_r \cdot \mu_s.
\]

Clearly, if the mean vector of \( X \) was \( \mu = (0, 0, \ldots, 0)' \), then the mean vector and covariance matrix of \( Z \) would be identical to those of \( X \). In the less idealized situation with which we are confronted, the translation to the sample mean of the nearest neighbor cloud should result in the pseudo-observation having very nearly the same mean and covariance structure as that of the (truncated) distribution of the points in the nearest neighbor cloud, a conjecture substantiated in many actual cases that have been considered. For \( m \) moderately large, our algorithm essentially samples from \( n \) Gaussian distributions with the means and covariance matrices corresponding to those of the \( n m \)-nearest-neighbor clouds.

III. EXAMPLES

For a substantial test case, we considered a mixture of three bivariate normal distributions. The first (\( N_1 \)) has mean vector \((\mu_1, 1/2)\) and covariance matrix \(\begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}\); the second (\( N_2 \)) has mean vector \((\mu_2, 1/2)\) and covariance matrix \(\begin{pmatrix} 1/2 & 1 \\ 1 & 1/2 \end{pmatrix}\); and the third (\( N_3 \)) has mean vector \((\mu_3, 5/2)\) and covariance matrix \(\begin{pmatrix} 1 & 1/10 \\ 1/10 & 1 \end{pmatrix}\). The corresponding mixing scalars are \(\alpha_1 = 1/2\), \(\alpha_2 = 1/3\), and \(\alpha_3 = 1/6\), respectively. To establish a data base, a sample of eighty-five points was generated from this distribution via Monte Carlo simulation, and appears in Figure 1; a sample of eighty-five pseudorandom values was then produced by the algorithm, and the combined sample is shown in Figure 2.

Notice that the structure of the data is maintained in that the modes are preserved; the algorithm has not attempted to fill in gaps where gaps belong; the algorithm has, however, generated some points outside the boundary of the convex hull of the data base, all of which are desirable properties. These observations lend credence to the term "structural integrity" mentioned previously.

An application of the algorithm to a real world data set is summarized in Figures 3 and 4. In Figure 3, a two-dimensional marginal of a set of 973 four-dimensional behind armor debris measurements is portrayed; in Figure 4, 973 simulated data points are produced by our procedure. Once again, the salient features of the data set are preserved.
Figure 1. Data base for a mixture of three bivariate normal distributions.
Figure 2. Combined sample: Data base and pseudo-observations.
Figure 3. Marginal data for four-dimensional measurements.
Figure 4. Simulated behind armor debris.
Figure 5. Data base for MLRS bomblets.
Figure 6. Simulated MLRS bomblet data.
Figure 7. Combined sample: Data base and pseudo-observations.
Figures 5 through 7 display the results of applying the algorithm to a data base of modest size. Here a set of 49 bivariate measurements on MLRS bomblets shown in Figure 5 was supplemented by an additional 160 pseudo-observations (Figure 6), with the results portrayed in Figure 7.

The FORTRAN program of the algorithm appears in Appendix A. This program as listed produces plots using the IMSL Library; Figures 1-7 are plots produced by DISSPLA.

IV. CONCLUSIONS

We have demonstrated a means of empirical random number generation based on a sample of observations of a random variable X. No estimation of the underlying density is required. And, because of the local nature of the generation scheme, it is essentially free of assumptions on the underlying density of X. Naturally, any attempt to use this algorithm for generating bona fide new observations using the computer rather than producing real world data would be unwise. Rather, the algorithm operates somewhat like a smooth interpolator---highly dependent on the quality of the data points on which it is based. It gives us a means of avoiding nonrobust conclusions due to "holes" in the data set at important points of the simulation model.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to Csaba K. Zoltani for bringing this problem to their attention and to William E. Baker for suggestions and assistance provided during its pursuit.
APPENDIX A. COMPUTER PROGRAM

STEP 1. CREATION OF THE DATA FILE

Data should be created and stored as an MFA permanent file. This file should contain the following information. Note: Let pl denote program limitations.

CARD 1. 5 INPUTS

1. NUMDAT - number of input data points
   I4 1-4
   pl. 5 ≤ NUMDAT ≤ 1000

2. NUMGEN - number of pseudopoints generated
   I4 6-9
   pl. 5 ≤ NUMGEN ≤ 1000

3. IDIM - dimension n of n-space data
   I1 11
   pl. 1 ≤ IDIM ≤ 8

4. NRANDM - random number seed
   I3 16-18
   pl. 1 ≤ NRANDM ≤ 999

5. NUMPLT - number of plots requested
   I2 21-22
   Note - all possible 2-d plots (NUMPLT = -1)
   - no plots (NUMPLT = 0)
   - r plots (NUMPLT = r)
   - NUMPLT plots will be generated for both data and pseudo-observations.
   pl. -1 ≤ NUMPLT ≤ IDIM(IDIM-1)/2

CARD 2 & CARD 3

If NUMPLT = r, request plots of variables $X_i$ vs. $Y_j$ by indicating $j$ in Card 2 and the corresponding $i$ in Card 3.

Column 1 2 ... $r$
Card(2) $Y_1$ $Y_2$ ... $Y_r$
Card(3) $X_1$ $X_2$ ... $X_r$

If NUMPLT = 0 or -1, cards 2 and 3 should not be included, otherwise information on these cards would be used as data.
Card(4), Card(5) ... Card(NUMDAT)

These cards should contain data to be read in with F10.3 format. Data may consist of a maximum of 8 variables $X_1 - X_8$.

<table>
<thead>
<tr>
<th>Columns</th>
<th>1 - 10</th>
<th>11 - 20</th>
<th>...</th>
<th>71 - 80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card(4)</td>
<td>$X_1$ Data</td>
<td>$X_2$ Data</td>
<td></td>
<td>$X_8$ Data</td>
</tr>
<tr>
<td>Card(5)</td>
<td>$X_1$ Data</td>
<td>$X_2$ Data</td>
<td></td>
<td>$X_8$ Data</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Card(NUMDAT)</td>
<td>$X_1$ Data</td>
<td>$X_2$ Data</td>
<td></td>
<td>$X_8$ Data</td>
</tr>
</tbody>
</table>

**STEP 2**

Before the program can be run, the data file must be made accessible to the MFZ with (PERMIT, PFN, MFZ).

**STEP 3**

To run the program create and submit the following 3 card MFZ job.

JOBNAME, STMFZ, T100.
ACCOUNT, XXXXXXX.
BEGIN, DBRNG, DBRNG, PFI = ____ , PFO = ____ , UN = ____ , RJE = RJEXXXX.

Where

PFI = file name under which the input data file is stored,
PFO = file name under which the pseudo data is to be stored,
UN = user name identification for above two files,
XXXX = a 4 digit code designating a particular RJE, as the output device.

If omitted, then the central site will serve as the output destination.
C THIS PROGRAM PRODUCES PSEUDO-RANDOM OBSERVATIONS FROM REAL
C DATA FOR UNIVARIATE AND MULTIVARIATE CASES. THE PSEUDO-
C RANDOM OBSERVATIONS WILL MAINTAIN THE CHARACTERISTICS OF
C THE REAL DATA WITHOUT ANY DISTRIBUTION ASSUMPTION ON THE
C POPULATION FROM WHICH THE REAL DATA CAME. AN EXAMPLE OF
C PROPER USE OF THIS PROGRAM WOULD BE IN THE CREATION OF PSEUDO-
C RANDOM OBSERVATIONS FOR INPUT TO A COMPUTER SIMULATION
C MODEL.
C
C CAUTION: THIS PROGRAM DOES NOT PRODUCE REAL DATA AND PRODUCED
C PSEUDORANDOM OBSERVATIONS SHOULD NOT BE USED AS SUCH.
C
C NUMDAT - NUMBER OF INPUT DATA POINTS
C NUMGEN - NUMBER OF PSEUDO-POINTS TO BE GENERATED
C IDIM - NUMBER OF VARIABLES IN INPUT DATA SET
C RANDOM - RANDOM NUMBER SEED
C NUMPLT - NUMBER OF PLOTS REQUESTED
C DATA - MATRIX HOLDING INPUT DATA SET
C NUMPLT - MATRIX HOLDING PLOT REQUESTS
C
C PROGRAM DRNG(INPUT, OUTPUT, TAPE6 = OUTPUT, TAPE8 = TAPE9)
C
C DIMENSION DATA(1000, 10), PSEUDO(1000, 10), NUMPLT(2, 40), DIM(10)
C DIMENSION MAX(10)
C
C READ AND WRITE INITIAL INPUT VARIABLES
C
C READ(8, 1000)NUMDAT, NUMGEN, IDIM, RANDOM, NUMPLT
C WRITE(3, 2030)
C WRITE(3, 2040)NUMDAT, NUMGEN, IDIM, RANDOM, NUMPLT
C
C CHECK FOR INVALID VALUES IF INITIAL INPUT VARIABLES
C
C CALL CHECK(NUMDAT, NUMGEN, IDIM, RANDOM, NUMPLT, NCHECK)
C IF(NCHECK.EQ.16) GO TO 90
C IF(NNUMPLT.GT.2) GO TO 10
C
C ESTABLISH NUMPLT FOR ALL POSSIBLE IDIM(IDIM-1)/2 PLOTS
C
C CALL SETPLT(NUMPLT, IDIM)
C GO TO 40
C
C USER ESTABLISHES DESIRED PLOTS
C F VALUES FIRST CARD, CORRESPONDING X VALUES SECOND CARD
C
C 10 DO 20 K = 1, 2
C READ(8, 1010)(NUMPLT(K, L), L = 1, NUMPLT)
C 20 CONTINUE
C
C IF USER DEFINED, WRITE PLT REQUESTS REQUESTED
C
C WRITE(6, 2050)
C DO 30 K = 1, 2
C WRITE(5, 2070)(NUMPLT(K, L), L = 1, NUMPLT)
C 30 CONTINUE
C
C READ REAL DATA AND WRITE FIRST FIVE POINTS
C
C 40 DO 50 I = 1, NUMDAT
C READ(8, 1020)(DATA(I, J), J = 1, IDIM)
C 50 CONTINUE
C
C WRITE(6, 2060)
C DO 50 I = 1, 5
C WRITE(5, 2020)(DATA(I, J), J = 1, IDIM)
C 50 CONTINUE
C VARY THE RANDOM NUMBER SEQUENCE USING INPUT VARIABLE NRANDOM
DO 70 =1,NRANDOM
RN=RNANF(0)
70 CONTINUE
C WRITE HEADER FOR OUTPUT
WRITE(5,2000)
C DETERMINE THE CORRELATION MATRIX, MEAN AND VARIANCES FOR REAL DATA
CALL CORREL(DATA,IDIM,NUMDAT)
C SCALE DATA SO THAT EACH VARIABLE WILL CARRY EQUAL WEIGHT IN
C THE NEIGHBORHOOD SELECTION PROCESS
CALL SCALE(DATA,NUMDAT,IDIM,ZMIN,ZMAX)
C GENERATE NUMGEN PSEUDO DATA POINTS
CALL GERATIDATA,NUMDAT,IDIM,PSEJDD)
C RESCALE THE DATA AND THE CORRESPONDING PSEUDO DATA TO THEIR
C ORIGINAL MAGNITUDES
CALL RESCAL DATA,NUMDAT,IDIM,ZMIN,ZMAX)
CALL RESCAL(PSEUDO,NUMGEN,IDIM,ZMIN,ZMAX)
C WRITE HEADER FOR OUTPUT
WRITE(5,2010)
C DETERMINE THE CORRELATION MATRIX, MEAN AND VARIANCES FOR PSEUDO-
C DATA
CALL CORREL(PSEUDO,IDIM,NUMGEN)
C WRITE THE PSEUDORANDOM OBSERVATIONS INTO A PERMANENT
C FILE (PSEUDO)
DO 80 J=1,NUMGEN
WRITE(4,2020)PSEUDO(J,L),L=1,IDIM
80 CONTINUE
C CALL PLOT ROUTINE IF REQUESTED.
C
1000 FORMAT(14,1X,I4,1X,I4,1X,4X,13,2X,12)
1100 FORMAT(3611)
1020 FORMAT(8F10.3)
2000 FORMAT(14I1,2X,'CORRELATIONS, MEAN AND VARIANCES OF INPUT DATA SE
TXY//)
2010 FORMAT///23X,'CORRELATIONS, MEANS AND VARIANCES OF PSEUDO DATA S
ETXY//)
2020 FORMAT(4X,8(F12.3X))
2030 FORMAT(14I1,4X,NUMDAT*,2X,NMGGEN*,3X,IDIM*,3X,'NRANDOM*,
+2X,NUMPLT*)
2440 FORMAT///,8X,*REQUESTED Y OVER X*//
2540 FORMAT///,5X,*REQUESTED Y OVER X*//
2640 FORMAT///,5X,*REQUESTED LAST FIVE DATA POINTS*//
2740 FORMAT(5X,3612)
2840 FORMAT(4X,141,1*)
END
C THIS SUBROUTINE initializes the initial input data for values
C which will cause the program to fail. For example, numdat
C can be a maximum of 1000 as the dimension statement only
C allows for 1000 data points. If an incorrect value is
C detected, ncheck will be set to 1. When returned as 1
C the program will stop. If returned as 0 the program will
C continue normally.

SUBROUTINE CHECK(numdat,numgen,idim,nrandom,numplt,nccheck)
    ------------------------------------------------------------------
    IF(numdat.GT.1000,JS.numdat.LT.5)GO TO 100
    IF(numgen.LT.5,JS.numgen.GT.1000)GO TO 200
    IF(idim.LT.1,JS.idim.GT.1000)GO TO 300
    IF(nrandom.LE.0)GO TO 400
    K=IDIM*IDIM-1/2
    IF(NUMPLT.GT.K)GO TO 500
    IF(NUMPLT.LT.-1)GO TO 500
    IF(idim.EQ.1.AND.numplt.EQ.-1)GO TO 500
    NCHECK=0
    RETURN
100 WRITE(6,2000)
    NCHECK=1
    RETURN
200 WRITE(6,2010)
    NCHECK=1
    RETURN
300 WRITE(6,2020)
    NCHECK=1
    RETURN
400 WRITE(6,2030)
    NCHECK=1
    RETURN
500 WRITE(6,2040)
    NCHECK=1
    RETURN
2000 FORMAT(/,5X,'(INVALID NUMBER OF INPUT DATA POINTS)')
2010 FORMAT(/,5X,'(INVALID NUMBER OF PSEUDO DATA POINTS)')
2020 FORMAT(/,5X,'(INVALID DIMENSION N OF V-SPACE DATA)')
2030 FORMAT(/,5X,'(INVALID SEED)')
2040 FORMAT(/,5X,'(INVALID NUMBER OF 20 PLOTS FOR DIMENSION SPECIFIED)')
END

C C This subroutine initializes the nplt matrix so that all possible
C 2-D plot combinations are considered. There is a total of
C IDIM*(IDIM-1)/2 plots which could be made. In main if numplt=0, 40
C plots will be made. If numplt=1, all plots will be made.
C
SUBROUTINE SETPLT(nplt,idim)
    DIMENSION NPLT(2,40)
    K=1
    II=IDIM-1
    DO 20 I=1,II
      JJ=I+1
      DO 10 J=JJ,IDIM
        NPLT(1*K)=I
        NPLT(2*K)=J
        K=K+1
      10 CONTINUE
    20 CONTINUE
    RETURN
END
C THIS SUBROUTINE SCALES THE DATA SO THAT EACH VARIABLE WILL
C CARRY EQUAL WEIGHT IN THE NEIGHBORHOOD SELECTION PROCESS. THE
C SCALED DATA WILL THEN BE RETURNED TO MAIN. THE PROCESS USED
C IS ( X(I)-MIN(X(I)) ) / RANGE(X(I)) FOR EACH VARIABLE.
C
SUBROUTINE SCALE(TDATA,NSORT,IDIM,ZMIN,ZMAX)
  
  DIMENSION TDATA(1000+10),ZMIN(10),ZMAX(10)

C INPUT FOR BUBBLE SORT
C  
NTOP=NSORT-1
C
C LOOP WHICH SORTS ON EACH VARIABLE ( NRANK ) AND THEN
C ESTABLISHES ITS MINIMUM AND MAXIMUM ( ZMIN ) AND ( ZMAX )
C RESPECTIVELY
C
     DO 10 I=1,IDIM
       NRANK=I
       CALL SORT(TDATA,NSORT,NTOP,NRANK,IDIM)
       ZMIN(I)=DATA(1,I)
       ZMAX(I)=DATA(NSORT,I)
   10 CONTINUE
C
C LOOP WHICH PERFORMS THE ABOVE MENTIONED TRANSFORMATION
C
     DO 20 J=1,NSORT
       DO 30 K=1,NTOP
       TDATA(J,K)=(DATA(J,K)-ZMIN(K))/(ZMAX(K)-ZMIN(K))
   30 CONTINUE
   20 CONTINUE
RETURN
END

C THIS ROUTINE SORTS THE DATA MATRIX ON POSITION NRANK
C THE SORT USED IS A COMMON BUBBLE SORT WHICH WILL ESTABLISH
C THE FIRST NTOP POINTS FROM SMALLEST TO LARGEST - WHERE
C SMALLEST T OR LARGEST IS DETERMINED BY POSITION NRANK. NOTE
C THAT WHEN AN EXCHANGE TAKES PLACE THE ENTIRE ROW VECTOR;
C SOME POINT (W,X,Y,...) IS EXCHANGED. NOTE ALSO THAT 0
C REPRESENTS THE DISTANCE (SQUARED) COMPUTED IN EUCLEDIAN AND STORED
C IN POSITION IDIST.
C
C
C SUBROUTINE SORT(TDATA,NSORT,NTOP,NRANK,IDIM)
  
  DIMENSION TDATA(1000+10)

  IDIST=IDIM+1
C
TAKES THE FIRST ITH VALUE AND COMPARE IT TO THE I+1TH
C VALUE. IF THE ITH VALUE IS SMALLER, EXCHANGE IT WITH
C THE I+1TH VALUE SO THAT THE ITH VALUE IS THEN SMALLER.
C THEN COMPARE THE I+1TH VALUE WITH THE I+2TH VALUE AND
C SO ON.
C
     DO 30 I=1,NTOP
       L=I+1
       DO 20 J=1,NSORT
       IF(TDATA(I,NRANK).LT.TDATA(J,NRANK)) GO TO 20
   20 CONTINUE
     30 CONTINUE
     20 CONTINUE
RETURN
END

24
THIS SUBROUTINE DOES THE ACTUAL GENERATION OF THE PSEUDORANDOM OBSERVATIONS, AND RETURNS THEM IN A MATRIX ( PSEUDO ). THE ALGORITHM USED WAS DEVELOPED BY DR. JIM THOMPSON OF RICE UNIVERSITY AND DR. MALCOLY TAYLOR OF BRL.

SUBROUTINE GENERATE(DAT,A,NUMDAT,NUMGEN,IDI,M,PSEUDO)

DIMENSION DATA(1000,10),PSEUDO(1000,10), AVERAGE(10)
DIMENSION TRANS(25,10)

CALL INITMAT(PSEUDO,0)
DO 20 L=1,NUMGEN
DO 10 J=1,IDI
PSEUDO(L,J)=0.
10 CONTINUE
20 CONTINUE

CALL INITMAT(T,0)

CALL ESTBULATE THE SIZE OF THE NEIGHBORHOOD OF NEAREST POINTS TO BE USED IN A LINEAR COMBINATION.

NEIGHB=INT(FLOAT(NUMDAT)/20.)
IF(NEIGHB.LT.5)NEIGHB=5
IF(NEIGHB.GT.20)NEIGHB=20

CALL INITMAT(DIST,0)

CALL INITMAT(IDIST,1)

CALL INITMAT(DIST,0)

DO 30 J=1,NUMDAT
DATA(J,IDIST)=0.
30 CONTINUE

WEIGHT IS THE WEIGHTING FACTOR TO BE USED IN CALCULATING THE MEAN OF THE NEIGHB NEAREST NEIGHBORS. IT ALSO Serves AS THE MEAN OF THE SPECIAL UNIFORM DISTRIBUTION USED IN THE LINEAR COMBINATION.

WEIGHT=1./FLOAT(NEIGHB)

CALL UNADJ1 HELPS DEFINE THE UNIFORM DISTRIBUTION WITH MEAN WEIGHT.

UNADJ1=(3.-(FLOAT(NEIGHB)-1.))/(FLOAT(NEIGHB)**2.)**.5

THE FOLLOWING LDOO GENERATES NUMGEN PSEUDORANDOM OBSERVATIONS.

DO 120 JJJ=1,NUMGEN

CALL INITMAT(AVERAGE,0)

DO 40 YSET=1,IDI
AVERAGE(YSET)=0.
40 CONTINUE
RANDOMLY PICK A DATA POINT ( KCENTR ) AROUND WHICH A NEIGHBORHOOD WILL BE FORMED.

RN=RANF(0)
KCENTR=INT(RN*FL3AT(NUMDAT))+1

ESTABLISH THE EUCLIDEAN DISTANCE SQUARED OF ALL POINTS FROM KCENTR.
CALL EUCLD(Data, NUMDAT, IDIM, KCENTR)

SORT THE POINTS IN THEIR EUCLIDEAN DISTANCE FROM SMALLEST TO LARGEST. IN THIS FASHION THE NEIGHBMOREST NEIGHBORS WILL BE CHOSEN.
CALL SORT(Data, NUMDAT, NEIGHB, IDIST, IDIM)

COMPUTE THE AVERAGE OF X, Y, Z, ..., IN (X, Y, Z, ..., ) OF KCENTR AND ITS NEAREST NEIGHBORS

DO 66 I=1, NEIGHB
DO 65 J=1, IDIM
AVERAGE(J)=AVERAGE(J)+DATA(I, J)*WEIGHT
50 CONTINUE
60 CONTINUE

CREATE A TRANSLATED DATA SET ( TRANSO ) TO BE USED IN THE CREATION OF ONE POINT.

DO 80 M=1, NEIGHB
DO 70 L=1, IDIM
TRANSO(M, L)=DATA(M, L)-AVERAGE(L)
70 CONTINUE
80 CONTINUE

BEGIN THE LOOP WHICH CREATES A NEW POINT BY TAKING A LINEAR COMBINATION OF THE TRANSLATED DATA.
DO 100 I=1, NEIGHB

ESTABLISH A RANDOM NUMBER FROM THE SPECIAL JNIFJR\# DISTRIBUTION TO BE MULTIPLIED BY ONE DATA VECTOR (X, Y, Z, ...).

RN=RANF(0)

RN=RN*2*UNADJ+1(DIST-UNADJ)

L33 WHICH ADDS THE TRANSFORMED VECTORS TO CREATE ONE NEW POINT

DO 90 J=1, IDIM
PSEUDO(JJJ, J)=PSEUDO(JJJ, J)*RN*TRANSO(I, J)
90 CONTINUE
100 CONTINUE

LOOP WHICH ADDS BACK IN THE AVERAGE OF THE NEIGHBORHOOD WHICH WAS TAKEN AWAY FROM ( TRANSO )

DO 110 L=1, IDIM
PSEUDO(JJJ, L)=PSEUDO(JJJ, L)+AVERAGE(L)
110 CONTINUE
120 CONTINUE
RETURN
C
C THIS SUBROUTINE COMPUTES THE MEAN VECTOR, VARIANCE VECTOR
C AND CORRELATION MATRIX OF ANY MATRIX SUBMITTED TO IT. IF THEN
C WRITES THIS INFORMATION IN HARD COPY.
C
C SUBROUTINE CORREL(CDATA,ID1,NUMC3R)
C ---------------------------------------------------------------
DIMENSION CDATA(100,10),SUMX(10),SUMXY(10,10),VAR(10),LINE(8)
DIMENSION AVR(10),CORR(10,10),LINE1(40),LINE2(18)
C INSURES OUTPUT NEATNESS
C
DATA LINE/*X1,*,X2,*,X3,*,X4,*,X5,*,X6,*,X7,*,X8,*/
DATA LINE2/*E*,*,E*,*,E*,*,E*,*,E*,*,E*,*,E*,*,E*,*,E*,*,E*/
DO 10 I=1,40
LINE1(I)=1
10 CONTINUE
C
C INITIALIZE THE SUM OF X(I) AND THE SUM OF X(I)*X(J) TO ZERO.
C
DO 30 I=1,ID1
SUMX(I)=0.
DO 20 M=1,ID1
SUMXY(M,M)=0.
20 CONTINUE
30 CONTINUE
C
C COMPUTE THE SUM OF X(I) AND THE SUM OF X(I)*X(J).
C
DO 60 I=1,NUMC3R
DO 50 JJ=1,ID1
SUMX(JJ)=SUMX(JJ)+CDATA(JJ)
50 CONTINUE
60 CONTINUE
C
C GET THE REAL VALUE OF THE NUMBER OF POINTS CONSIDERED.
C
W=NUMC3R
C
C COMPUTE THE MEANS AND SAMPLE VARIANCES OF X,Y,Z,... OF
C (X,Y,Z,....).
C
DO 70 J=1,ID1
AVR(J)=SUMX(J)/W
VAR(J)=(SUMXY(J,J)/W-AVR(J)**2.)*W/(W-1.)
70 CONTINUE
C
C COMPUTE THE CORRELATION OF X(L) AND X(K).
C
DO 90 K=1,ID1
K=L
DO 80 K=L,ID1
CORR(K,K)=SUMXY(L,K)-SUMXY(L)*SUMX(K)/AVR(K)
CORR(K,K)=CORR(K,K)*AVR(L)/AVR(K)**2.*W
80 CONTINUE
90 CONTINUE
THE FOLLOWING ARE ALL AIDS IN WRITING THE INFORMATION IN AN ACCEPTABLE FASHION.

```
TEMP=FLOAT(IDIM)
NDQ=INT(TEMP/2.*9.)-5
WRITE(6,2010)(LINE1(J),J=1,NDQ),(LINE2(I),I=1,IDIM)
WRITE(6,2010)(LINE4(K),K=1,IDIM)
DO 100 I=1,IDIM
 IF(I.EQ.IDIM)GOTO 110
 K=I+1
 WRITE(6,2220)(CCCRR(J,I),J=1,1),(CCCRR(I,J),J=K,IDIM)
 100 CONTINUE
 WRITE(6,2020)(IDJCORR(J,IDIM),J=1,IDIM)
 WRITE(6,2030)(AVR(I),I=1,IDIM)
 WRITE(6,2040)(VAR(I),I=1,IDIM)
 WRITE(6,2050)(VAR(J),J=1,IDIM)
2000 FORMAT(14D,58AL)
2010 FORMAT(/,9X,K,7(F8.5,1X))
2020 FORMAT(/,1X,'P',3X,B8.5,1X))
2030 FORMAT(/,7X,B8(13X,A2))
2040 FORMAT(/,2X,'MEAN',3X,B9(13X,F11.3))
2050 FORMAT(/,2X,'VAR',4X,3(1X,F11.3))
RETURN
END
```

THIS SUBROUTINE CALCULATES THE EUCLIDEAN DISTANCE BETWEEN KCENTR AND ALL OTHER POINTS.

```
SUBROUTINE EUCLID(DDATA, NUMDAT, IDIM, KCENTR)

DIMENSION DDATA(1000,16)
IDIST=IDIM+1
DO 20 J=1,NUMDAT
DSQUR=0.
DO 10 L=1,10
 DSQUR=DSQUR+(DDATA(J,L)-DDATA(KCENTR,L))**2.
10 CONTINUE
DDATA(J,IDIST)=DSQUR
20 CONTINUE
RETURN
END
```

THEM SUBROUTINE RESCALES THE DATA AND PSEUDO DATA BACK TO ITS ORIGINAL MAGNITUDE.

```
SUBROUTINE RESCALES(RDATA, NUMBER, IDIM, ZMIN, ZMAX)

DIMENSION RDATA(1000,10), ZMIN(10), ZMAX(10)
DO 20 J=1,NUMBER
DO 10 K=1,IDIM
 RDATA(J,K)=RDATA(J,K)*(ZMAX(K)-ZMIN(K)+ZMIN(K))
10 CONTINUE
20 CONTINUE
RETURN
END
```

28
THIS ROUTINE PREPARES FOR THE CALLING OF AN IMSL PLOT ROUTINE SO THAT ALL REQUESTED 2-D PLOTS CAN BE MADE.

SUBROUTINE PLOT(PSEUDO, DATA, ZMIN, ZMAX, NPLT, NUMOAT, NUMGEN, NUMPLT)

DIMENSION PSEUDO(100, 10), DATA(100, 10), ZMIN(10), ZMAX(10)
DIMENSION NUM(10), IHEAD(22), ITITLE(144), JTITLE(144)
DIMENSION X(1000), Y(1000), IMAX(515), ICHAR(10), RANGE(4)

C ESTABLISH SOME HOLLERITH STRINGS FOR CLARITY OF OUTPUT.

DATA N1M/1,2,3,4,5,6,7,8/
DATA IHEAD/1,2,3,4,5,6,7,8/
DATA IHEAD/1,2,3,4,5,6,7,8/
DATA IHEAD/1,2,3,4,5,6,7,8/
DATA IHEAD/1,2,3,4,5,6,7,8/
DATA IHEAD/1,2,3,4,5,6,7,8/
DO 10 J=1,44
IITLE(J)='*
JTITLE(J)='*
10 CONTINUE
DO 20 J=1,24
L=I(J)
IITLE(L)=IHEAD(I)
20 CONTINUE
DO 30 J=1,22
L=J+33
JTITLE(L)=IHEAD(I)
30 CONTINUE
IITLE(97)='X'
IITLE(97)='X'
IITLE(126)='X'
IITLE(126)='X'

C SET VARIABLE VALUES ASKED FOR BY IMSL ROUTINE

I=1
IID=1

C BEGIN LOOP WHICH PLOTS FIRST THE DATA AND THEN THE PSEUDO DATA IN TWO
C SEPARATE PLOTS FOR A GIVEN X(I) AND X(J)

DO 60 J=1,NPLT
C K AND L ARE THE X(Y) AND X(K) COORDINATES RESPECTIVELY

K=NPLT(J,J)
L=NPLT(2,J)

C ESTABLISH THE ENDS POINTS OF THE X AND Y AXES.

RANGE(1)=ZMIN(L)-10*(ZMAX(L)-ZMIN(L))
RANGE(2)=ZMAX(L)-10*(ZMAX(L)-ZMIN(L))
RANGE(3)=ZMIN(K)-10*(ZMAX(K)-ZMIN(K))
RANGE(4)=ZMAX(K)-10*(ZMAX(K)-ZMIN(K))

C ASSIGN THE X AND Y VECTORS VALUES FOR PLOTTING

DO 40 J=1,NUMOAT
X(J)=DATA(J,J)
Y(J)=DATA(J,J)
40 CONTINUE
DOCTOR THE HEADING FOR A DATA PLOT AND SET THE OUTPUT CHARACTER

ITITLE(*)=NUM(K)
ITITLE(55)=NUM(L)
ITITLE(98)=NUM(L)
ITITLE(127)=NUM(K)
IC=AR(I)="D"

CALL THE IMSL ROUTINE WHICH WILL GIVE BACK A 2-D PLOT

IA=NUM)AT
CALL USPLT(X,Y,IA,IA,Y,INC,ITITLE,RANGE,ICHR,IOPT,IAG4,IER)

ASSIGN THE X AND Y VECTORS VALUES FOR PLOTTING

07 50 J2=L,NUMGEN
X(J2)=PSEUDO(J2,L)
Y(J2)=PSEUDO(J2,K)
50 CONTINUE

DOCTOR THE HEADING FOR A PSEUDO DATA PLOT AND SET THE OUTPUT CHARACTER

JTITLE(*9)=NUM(K)
JTITLE(56)=NUM(L)
JTITLE(98)=NUM(L)
JTITLE(127)=NUM(K)
IC=AR(I)="P"
IA=NUMGEN

CALL THE IMSL ROUTINE WHICH WILL GIVE BACK A 2-D PLOT

CALL USPLT(X,Y,IA,IA,Y,INC,JTITLE,RANGE,ICHR,IOPT,IAG4,IER)
60 CONTINUE
RETURN
END
<table>
<thead>
<tr>
<th>No. of Copies</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Administrator Defense Technical Info Center ATTN: DTIC-DDA Cameron Station Alexandria, VA 22314</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Materiel Development and Readiness Command ATTN: DRCDMD-ST 5001 Eisenhower Avenue Alexandria, VA 22333</td>
</tr>
<tr>
<td>2</td>
<td>Commander US Army Armament Research and Development Command ATTN: DRDAR-TDC Dover, NJ 07801</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Armament Research and Development Command ATTN: DRDAR-TSS (2 cys) Dover, NJ 07801</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Armament Materiel Readiness Command ATTN: DRSAR-LEP-L, Tech Lib Rock Island, IL 61299</td>
</tr>
<tr>
<td>1</td>
<td>Director US Army ARRADCOM Benet Weapons Laboratory ATTN: DRDAR-LCB-TL Watervliet, NY 12189</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Aviation Research and Development Command ATTN: DRDAV-E 4300 Goodfellow Blvd. St. Louis, MO 63120</td>
</tr>
<tr>
<td>1</td>
<td>Director US Army Air Mobility Research and Development Laboratory Ames Research Center Moffett Field, CA 94035</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Communications Rsch and Development Command ATTN: DRDCO-PPA-SA Fort Monmouth, NJ 07703</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Electronics Research and Development Command Technical Support Activity ATTN: DELSD-L Fort Monmouth, NJ 07703</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Missile Command ATTN: DRSMI-R Redstone Arsenal, AL 35898</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Missile Command ATTN: DRSMI-YDL Redstone Arsenal, AL 35898</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Tank Automotive Rsch and Development Command ATTN: DRDTA-UL Warren, MI 48090</td>
</tr>
<tr>
<td>1</td>
<td>Director US Army TRADOC Systems Analysis Activity ATTN: ATAA-SL, Tech Lib White Sands Missile Range NM 88002</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Research Office ATTN: DRXRO-MA, Dr. Robert Launer P.O. Box 12211 Research Triangle Park, NC 27709</td>
</tr>
<tr>
<td>1</td>
<td>Commander US Army Combat Development Experimentation Command ATTN: ATEC-SA, Dr. Marion R. Bryson Fort Ord, CA 93941</td>
</tr>
</tbody>
</table>
DISTRIBUTION LIST

Aberdeen Proving Ground

Dir, USAMSAA
ATTN: DRXSY-D
DRXSY-MP, H. Cohen

Cdr, USATECOM
ATTN: DRSTE-TO-F
DRSTE-AD, Dr. Ted Trybul

Dir, USACSL, Bldg. E3516, EA
ATTN: DRDAR-CLB-PA
USER EVALUATION OF REPORT

Please take a few minutes to answer the questions below; tear out this sheet, fold as indicated, staple or tape closed, and place in the mail. Your comments will provide us with information for improving future reports.

1. BRL Report Number_____________________________________

2. Does this report satisfy a need? (Comment on purpose, related project, or other area of interest for which report will be used.)
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

3. How, specifically, is the report being used? (Information source, design data or procedure, management procedure, source of ideas, etc.)
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

4. Has the information in this report led to any quantitative savings as far as man-hours/contract dollars saved, operating costs avoided, efficiencies achieved, etc.? If so, please elaborate.
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

5. General Comments (Indicate what you think should be changed to make this report and future reports of this type more responsive to your needs, more usable, improve readability, etc.)
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

6. If you would like to be contacted by the personnel who prepared this report to raise specific questions or discuss the topic, please fill in the following information.

   Name:_______________________________________________________________

   Telephone Number:___________________________________________________

   Organization Address:_________________________________________________
Director
US Army Ballistic Research Laboratory
Aberdeen Proving Ground, MD 21005

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE $300

BUSINESS REPLY MAIL
FIRST CLASS PERMIT NO 12062 WASHINGTON, DC
POSTAGE WILL BE PAID BY DEPARTMENT OF THE ARMY

Director
US Army Ballistic Research Laboratory
ATTN: DRDAR-TSB -S
Aberdeen Proving Ground, MD 21005