Technical Report No. 92

# MULTIVARIATE NORMAL DATA GENERATOR

L. J. Bledsoe and J. D. Gustafson

Grasslands Ecology Research Laboratory

Colorado State University

Fort Collins, Colorado

GRASSLAND BIOME

U.S. International Biological Program

April 1971

# TABLE OF CONTENTS

## ABSTRACT

This FORTRAN program is designed to generate vectors of data whose components are distributed according to the multivariate normal distribution function with a given mean vector and a given covariance matrix. The data generated can be either printed on the output unit or punched on Hollerith cards under program options.

## GENERAL DESCRIPTION

The program requires the following specifications: the number of data points to be punched (N); the number of components in each data point (M); the printing and punching options; a mean vector having M components; and an M by M upper triangular covariance matrix. The covariance matrix must be positive-definite (Graybill 1969) or the program will abort with a "negative argument in square root function" error. In addition to printing and/or punching the data, the estimated mean vector and estimated covariance matrix are calculated. The method is that of Naylor et al. (1968). The appendix contains program listing, sample data, and output.

## DATA FORMAT

| Cards | Description | Format |
|---|---|---|
| Card 1 | | |
| Column 1-5 | Number of components/data points | I5 |
| Column 6-10 | Number of data points | I5 |
| Column 11-15 | 5 digit odd number to initialize random number sequence | I5 |
| Card 2 | | |
| Column 1-3 | "YES" if printed results are desired, or else "NO" | R3 |
| Column 4-6 | "YES" if punched results are desired, or else "NO" | R3 |
| Card 3 | (Optional) format for punched data cards, omit if no punched output is called for (card 2) | 8A10 |
| Card 4 | Vector of means--continue on second card if necessary | 8F10.0 |

| Cards | Description | Format |
|---|---|---|
| Card 5-5+N | Covariance matrix. Upper triangular portion only. First number on each card is the diagonal term. Start a new row on each card. | 8F10.0 |

As many data sets as desired may be run at one time. End data deck with *two* blank cards for proper system exit.

## PROGRAM LIMITATIONS

There is no limitation to the number of data points that may be generated; however, the appropriate page and punch limits must be included in the system control cards if large amounts of printed and punched output are expected. The program dimensions are set up for a maximum of a ten-component data point. This can be modified by changing the dimension card and also modifying the limit on the DO loop which is the first card after statement 900. This DO loop initializes the sum and sum of square counters to zero for each run.

## PURPOSE AND APPLICATIONS

This program is designed to provide a test for statistical data analysis procedures by providing a data set whose statistical characteristics are exactly known. It can be used for familiarizing a user with a particular method of analysis, for testing a method of analysis for which the results (estimates of parameters) are only approximate, or for determining the robustness of statistical tests to failures of the data to meet the theoretical requirements of the test. For example, it can be used in a discriminant analysis procedure to determine if the program can properly distinguish

between components and data which have different means, but are correlated and have high variances.

## LITERATURE CITED

Graybill, F. A. 1969. Introduction to matrices with applications in statistics. Wadsworth Publ. Co., Belmont, California. 372 p.

Naylor, T. H., J. L. Balintfy, D. S. Burdick, and K. Chu. 1968. Computer simulation techniques. John Wiley & Sons, Inc., New York. 352 p.

## APPENDIX I

## PROGRAM LISTING

```
      PROGRAM NRMGEN (INPUT,OUTPUT,PUNCH,TAPE5=INPUT,TAPE6=OUTPUT,
     1TAPE7=PUNCH)
C     TESTS THE MULTIVARIATE NORMAL DEVIATE GENERATOR BY CALCULATING
C     AN ESTIMATE OF THE MEAN VECTOR AND VARIANCE-COVARIANCE MATRIX
C     M = NO. OF COMPONENTS IN VECTOR
C     N = NO. OF DATA POINTS TO BE PRODUCED
C     INIT = K DIGIT ODD NUMBER TO INITIALIZE RANDOM SEQUENCE
C     PRT= 3RYES MEANS PRINT GENERATED DATA ON OUTPUT FILE
C     PUN= 3RYES MEANS PUNCH GENERATED DATA ON UNIT 7
C     SUPPLY FORMAT FOR PUNCH CARDS IF PUNCHED OUTPUT REQUESTED
      DIMENSION V(10,10),VE(10,10),X(10),XMU(10),XE(10),FMT(8)
      INTEGER PRT,PUN
    1 READ (5,100) M,N,INIT,PRT,PUN
  100 FORMAT(3I5/2A3)
      IF (M.EQ.0) GO TO 99
      WRITE (6,200) M,N,INIT,PRT,PUN
      IF (PUN.EQ.3RYES) READ (5,900) FMT
  900 FORMAT(8A10)
      DO 90 I=1,10
      X(I)=0.
      XMU(I)=0.
      XE(I)=0.
      DO 90 J=1,10
      V(I,J)=0.
   90 VE(I,J)=0.
      READ (5,300) (XMU(I),I=1,M)
      WRITE (6,400) (XMU(I),I=1,M)
  300 FORMAT(10F8.0)
      DO 80 I=1,M
   80 READ (5,300) (V(I,J),J=1,M)
      DO 50 I=2,M
      K=I-1
      DO 50 J=1,K
   50 V(I,J)=V(J,I)
      WRITE (6,600)
      DO 70 I=1,M
   70 WRITE (6,700) (V(I,J),J=1,M)
      CALL NRMINIT(X40),V,M,INIT,10)
      IF (PRT.EQ.3RYES) WRITE (6,800)
      XXX=TIMEF(0.)
      DO 10 K=1,N
      CALL NRM(X)
      IF (PRT.EQ.3RYES) WRITE (6,700) (X(LL),LL=1,M)
      IF (PUN.EQ.3RYES) WRITE (7,FMT) (X(LL),LL=1,M)
      DO 10 I=1,M
      XE(I)=XE(I)+X(I)
      DO 10 J=I,M
```

```
   10 VE(I,J)=VE(I,J)+X(I)*X(J)
      C1=1./N
      DO 20 I=1,M
      DO 20 J=1,M
      VE(I,J)=(VE(I,J)-(XE(I)*XE(J))*C1)/(N-1)
      IF(I.GT.J) VE(I,J)=VE(J,I)
   20 CONTINUE
      DO 30 I=1,M
   30 XE(I)=XE(I)*C1
      WRITE (6,1100) (XE(I),I=1,M)
      WRITE (6,1200)
      DO 40 I=1,M
   40 WRITE (6,1300) (VE(I,J),J=1,M)
      XXX=(TIMEF(0)-XXX)
      WRITE(6,1400) XXX
      GO TO 1
   99 STOP
  200 FORMAT(*1MULTIVARIATE NORMAL DATA GENERATOR*/
     15X,I5,*-COMPONENT DATA, *,I5,* DATA POINTS*/
     25X,I5,* INITIALIZES RANDOM SEQUENCE*/
     35X,*PRINT -- *,R3/5X,*PUNCH -- *,R3)
  400 FORMAT(//* MEAN VECTOR*/10G12.5)
  600 FORMAT(//* COVARIANCE MATRIX - EXPECTED*)
  700 FORMAT(* *10G12.5/(* *12X,9G12.5))
  800 FORMAT(//* RANDOM VECTORS*//)
 1100 FORMAT(//* MEAN VECTOR - ESTIMATED*//10G12.5)
 1200 FORMAT(//* COVARIANCE MATRIX - ESTIMATED*)
 1300 FORMAT(10G12.5)
 1400 FORMAT(//* ELAPSED TIME = *,F8.0,* MILLISECONDS*)
      END
      SUBROUTINE NRMINIT(X,V,N,M,ND)
      DIMENSION V(ND,ND),X(ND)
      COMMON/L1/C(10,10),Z(10),K
      K=N
      CALL RANSET(M)
      CALL MATSQR(V,C,N,ND)
      DO 10 I=1,N
   10 Z(I)=X(I)
      RETURN
      END
      SUBROUTINE MATSQR(V,C,N,NDIM)
      DIMENSION V(NDIM,NDIM),C(NDIM,NDIM)
      DO 30 J=1,N
      DO 30 I=1,N
      C(I,J)=0.
      IF(I.LT.J) GO TO 25
      IF(J.NE.1) GO TO 5
      C(I,J)=V(I,J)/SQRT(V(1,1))
      GO TO 30
    5 IF(I.NE.J) GO TO 15
      SUM2=0.
      IK=I-1
      K=1
```

```
   10 SUM2=SUM2+C(I,K)*C(I,K)
      K=K+1
      IF(K.LE.IK) GO TO 10
      C(I,J)=SQRT(V(I,J)-SUM2)
      GO TO 30
   15 SUMPRO=0.
      IJ=J-1
      K=1
   20 SUMPRO=SUMPRO+C(I,K)*C(J,K)
      K=K+1
      IF(K.LE.IJ) GO TO 20
      C(I,J)=(V(I,J)-SUMPRO)/C(J,J)
      GO TO 30
   25 C(I,J)=0
   30 CONTINUE
      RETURN
      END
      SUBROUTINE NRM(X)
      DIMENSION X(1),Y(10)
      COMMON/LI/C(10,10),Z(10),N
      DO 10 I=1,N
   10 Y(I)=GAUSS(0.,1.)
      CALL VMULT(Y,C,X,N,10)
      CALL VADD(X,Z,X,N,10)
      RETURN
      END
      SUBROUTINE VMULT(X,Y,Z,N,ND)
      DIMENSION X(ND),Y(ND,ND),Z(ND)
      DO 10 I=1,N
      Z(I)=0.
      DO 10 J=1,N
   10 Z(I)=Z(I)+X(J)*Y(I,J)
      RETURN
      END
      SUBROUTINE VADD(X,Y,Z,N,ND)
      DIMENSION X(ND),Y(ND),Z(ND)
      DO 10 I=1,N
   10 Z(I)=X(I)+Y(I)
      RETURN
      END
      FUNCTION GAUSS(X,S)
      X1=RANF(0.)
      X2=RANF(0.)
      C1=SIN(6.283185*X1)*SQRT(-2.*ALOG(X2))
      GAUSS=C1*S+X
      RETURN
      END
```

## DATA

```
     5  [illegible]
NO  NO
30.       30.       30.       30.       30.
5.        2.        1.        0.        0.
4.        1.        0.        0.
3.        0.        0.
2.        0.
10.
     3    1057101
YESNO
0.        0.        0.
1.        0.        0.
1.        1.
1.
     5    10055073
YESNO
10.       20.       12.       20.       0.
4.        3.5       1.5       0.        0.
4.        3.        .         .
9.        0.        .
4.        0.
4.
```

## SAMPLE OUTPUT

```
MULTIVARIATE NORMAL DATA GENERATOR
        5-COMPONENT DATA.  1000 DATA POINTS
     23961 INITIALIZES RANDOM SEQUENCE
     PRINT -- NO
     PUNCH -- NO


 MEAN VECTOR
  30.000        30.000        30.000        30.000        30.000

 COVARIANCE MATRIX - EXPECTED
    5.0000        2.0000        1.0000        0.            0.
    2.0000        4.0000        1.0000        0.            0.
    1.0000        1.0000        3.0000        0.            0.
    0.            0.            0.            2.0000        0.
    0.            0.            0.            0.           10.000

 MEAN VECTOR - ESTIMATED

  29.974        29.970        29.939        30.028        30.100

 COVARIANCE MATRIX - ESTIMATED
    5.4836        1.9645        .91977        .29854E-01    .11399
    1.9645        4.1192        .95384        .63842E-01   -.15805E-02
    .91977        .95384       2.9331         .49101E-01    .19442
    .29854E-01    .63842E-01    .49101E-01    1.9404       -.76047E-02
    .11399       -.15805E-02    .19442       -.76047E-02    9.3424

 ELAPSED TIME =      3225. MILLISECONDS
```

```
MULTIVARIATE NORMAL DATA GENERATOR
        3-COMPONENT DATA.     15 DATA POINTS
     57101 INITIALIZES RANDOM SEQUENCE
     PRINT -- YES
     PUNCH -- NO


MEAN VECTOR
0.              0.              0.


COVARIANCE MATRIX - EXPECTED
  1.0000        0.              0.
 0.             1.0000          0.
 0.             0.              1.0000


RANDOM VECTORS


  .60436E-03    .97951          -.21783
 -.12885        -.74113         -.39497
  .14031        -1.1227          .55461E-01
  .82486        -.40891         -.55549
 -.62405        -.33310          .51870
 -.50320        -1.3895          .46441
 -.35518         .50771E-01     -.94018
  1.1991         1.8221           .75519
-1.3135         1.5476          -.85765
 1.2198          .61694           .13794


MEAN VECTOR - ESTIMATED

 .45986E-01    .10215          -.24949


COVARIANCE MATRIX - ESTIMATED
 .67948         .14977           .19042
 .14977        1.2191           -.23358
 .19042        -.23358           .77842


ELAPSED TIME =        57. MILLISECONDS
```

MULTIVARIATE NORMAL DATA GENERATOR
5-COMPONENT DATA. 100 DATA POINTS
55673 INITIALIZES RANDOM SEQUENCE
PRINT -- YES
PUNCH -- NO

MEAN VECTOR

| | | | | |
|---|---|---|---|---|
| 10.000 | 20.000 | 12.000 | 20.000 | 0. |

COVARIANCE MATRIX - EXPECTED

| | | | | |
|---|---|---|---|---|
| 4.0000 | 3.6000 | 1.8000 | 0. | 0. |
| 3.6000 | 4.0000 | 3.0000 | 0. | 0. |
| 1.8000 | 3.0000 | 4.0000 | 0. | 0. |
| 0. | 0. | 0. | 4.0000 | 0. |
| 0. | 0. | 0. | 0. | 4.0000 |

RANDOM VECTORS

| | | | | |
|---|---|---|---|---|
| 10.000 | 20.244 | 13.060 | 18.223 | -4.2822 |
| 9.0406 | 19.023 | 9.9320 | 17.911 | .91856 |
| 11.238 | 20.817 | 14.222 | 17.936 | .34224 |
| 8.9294 | 20.454 | 13.259 | 21.718 | .52603 |
| 9.7495 | 17.941 | 5.0838 | 22.538 | 2.3128 |
| 10.990 | 19.508 | 9.3405 | 18.959 | .50847 |
| 14.462 | 24.135 | 15.716 | 20.259 | -1.0074 |
| 12.212 | 22.907 | 14.041 | 23.708 | -1.2427 |
| 8.0610 | 19.979 | 12.996 | 17.137 | .34907 |
| 9.9426 | 20.563 | 13.603 | 21.447 | 1.3642 |
| 7.7730 | 18.284 | 13.355 | 23.180 | -.74730 |
| 11.807 | 22.608 | 16.184 | 24.326 | .80855 |
| 7.2722 | 17.911 | 11.844 | 19.852 | .86075 |
| 11.699 | 23.266 | 16.228 | 21.436 | -.22649 |
| 12.054 | 21.804 | 15.538 | 21.379 | -1.8562 |
| 11.134 | 20.341 | 10.003 | 24.299 | 2.5445 |
| 6.0880 | 17.157 | 7.5004 | 18.188 | -.23959 |
| 6.9913 | 17.619 | 12.231 | 15.277 | 2.8350 |
| 6.4370 | 16.553 | 11.361 | 20.570 | -.62791E-01 |
| 8.1990 | 18.536 | 15.633 | 22.173 | -.26889 |
| 7.4608 | 16.821 | 4.6409 | 23.542 | .24729 |
| 10.972 | 21.021 | 15.052 | 18.380 | 1.4206 |
| 12.034 | 22.480 | 11.360 | 23.361 | -1.9098 |
| 10.668 | 19.444 | 8.5041 | 18.376 | -.67788 |
| 12.797 | 22.523 | 9.9333 | 20.825 | -3.0232 |
| 7.6869 | 16.506 | 7.0525 | 23.038 | 1.1483 |

| | | | | |
|---|---|---|---|---|
| 10.369 | 20.509 | 10.009 | 20.379 | 2.3384 |
| 11.038 | 22.114 | 12.394 | 18.671 | 1.3965 |
| 8.9188 | 19.017 | 7.5853 | 17.797 | -1.7069 |
| 8.1820 | 17.124 | 7.6682 | 20.795 | 1.7553 |
| 8.5591 | 18.681 | 15.106 | 18.781 | 2.2629 |
| 8.6554 | 18.681 | 9.5009 | 21.866 | -.26838 |
| 7.3996 | 18.420 | 14.343 | 16.485 | -.18089E-01 |
| 6.6262 | 15.390 | 8.4782 | 21.122 | 2.6642 |
| 8.2850 | 17.681 | 11.057 | 18.556 | .66833 |
| 11.349 | 20.323 | 12.547 | 18.865 | -.57856 |
| 9.9932 | 19.803 | 10.061 | 25.061 | -.86536 |
| 11.428 | 20.916 | 10.640 | 21.509 | .87400 |
| 11.378 | 21.265 | 10.290 | 20.623 | -1.7008 |
| 9.9034 | 20.099 | 12.955 | 20.083 | -1.2442 |
| 6.2648 | 16.210 | 10.335 | 19.725 | -3.2394 |
| 6.9273 | 17.482 | 9.8792 | 20.831 | -2.8377 |
| 11.805 | 21.092 | 10.224 | 22.555 | -1.4260 |
| 10.851 | 19.154 | 9.0255 | 21.463 | 1.0666 |
| 9.1265 | 19.381 | 8.0955 | 24.077 | -.90897 |
| 7.1267 | 17.147 | 11.822 | 18.103 | -.42140 |
| 11.532 | 20.759 | 8.0765 | 19.427 | 2.5201 |
| 11.403 | 21.989 | 17.782 | 21.159 | 1.6575 |
| 13.711 | 24.821 | 17.285 | 17.481 | .20300 |
| 9.7524 | 19.596 | 13.626 | 22.740 | .76846 |
| 9.6282 | 18.100 | 11.158 | 19.421 | .38464 |
| 9.6012 | 20.713 | 15.375 | 23.213 | .59204 |
| 14.764 | 24.570 | 15.113 | 21.912 | 2.4535 |
| 10.109 | 20.065 | 10.724 | 18.706 | -.96706 |
| 13.257 | 22.767 | 14.019 | 18.238 | .32567 |
| 10.196 | 20.401 | 10.219 | 20.925 | -1.5984 |
| 13.902 | 22.227 | 12.293 | 17.709 | 4.6295 |
| 9.0632 | 19.802 | 14.580 | 23.542 | -.89748 |
| 9.0524 | 18.151 | 4.9454 | 21.944 | .29469 |
| 11.711 | 21.691 | 12.549 | 20.878 | -1.1900 |
| 12.535 | 23.315 | 14.714 | 18.886 | -.78790E-01 |
| 12.503 | 22.770 | 18.995 | 19.398 | .30587 |
| 14.402 | 23.214 | 15.381 | 22.990 | .88352 |
| 11.391 | 21.344 | 13.357 | 19.523 | .68266 |
| 10.268 | 20.992 | 13.587 | 19.170 | .23074 |
| 8.9464 | 17.515 | 8.0724 | 15.283 | 1.6237 |
| 8.1064 | 17.671 | 14.377 | 21.284 | -1.6798 |
| 10.268 | 21.505 | 16.394 | 20.007 | .59776 |
| 10.068 | 19.802 | 13.220 | 22.521 | .36788 |
| 10.953 | 21.147 | 12.155 | 19.048 | 1.3434 |
| 12.268 | 20.587 | 7.2030 | 17.221 | -.10411 |
| 7.3932 | 19.055 | 12.516 | 16.779 | .81302 |
| 10.890 | 22.204 | 18.285 | 19.136 | -.92692 |
| 9.3395 | 19.571 | 14.029 | 17.972 | 2.3115 |

| | | | | |
|---|---|---|---|---|
| 8.3816 | 20.133 | 13.597 | 19.035 | -.68239 |
| 11.045 | 20.724 | 12.725 | 20.137 | -1.1174 |
| 7.9872 | 18.254 | 10.469 | 18.792 | 1.1532 |
| 12.592 | 23.152 | 16.451 | 21.643 | -.39975 |
| 10.044 | 20.990 | 11.457 | 22.966 | 1.3898 |
| 13.096 | 24.466 | 13.447 | 21.703 | 1.0295 |
| 5.3873 | 14.240 | 5.0015 | 19.089 | -3.5706 |
| 11.555 | 21.966 | 13.464 | 19.956 | 2.8837 |
| 11.825 | 22.253 | 15.994 | 20.020 | -4.1237 |
| 8.0842 | 17.259 | 3.7731 | 18.706 | 1.9024 |
| 11.136 | 23.034 | 17.614 | 17.631 | -1.7101 |
| 10.768 | 21.659 | 15.962 | 25.051 | -2.2595 |
| 12.321 | 23.824 | 14.790 | 20.257 | -.40391E-01 |
| 9.4621 | 20.337 | 13.089 | 19.673 | 2.1420 |
| 10.140 | 19.045 | 9.2437 | 21.763 | -1.5463 |
| 10.321 | 20.225 | 12.944 | 17.601 | -2.1159 |
| 9.1846 | 19.900 | 12.319 | 19.480 | .63152 |
| 11.162 | 20.776 | 13.934 | 22.330 | -3.6309 |
| 9.9991 | 19.922 | 13.004 | 24.037 | .36132 |
| 8.4767 | 18.091 | 13.629 | 20.131 | 4.2771 |
| 9.4946 | 20.145 | 14.591 | 20.846 | -.56001 |
| 9.0704 | 19.147 | 13.694 | 22.746 | .24697E-01 |
| 12.574 | 21.858 | 6.9327 | 19.566 | 1.1461 |
| 5.7213 | 16.488 | 11.418 | 15.200 | -2.7220 |
| 10.494 | 18.357 | 10.242 | 20.414 | -2.6816 |
| 12.107 | 22.972 | 21.787 | 19.532 | 2.3260 |

MEAN VECTOR - ESTIMATED

| | | | | |
|---|---|---|---|---|
| 10.091 | 20.134 | 12.223 | 20.312 | .59855E-01 |

COVARIANCE MATRIX - ESTIMATED

| | | | | |
|---|---|---|---|---|
| 4.2153 | 4.1077 | 2.7991 | .71207 | .25734 |
| 4.1077 | 4.8346 | 4.5589 | .60659 | .14326 |
| 2.7991 | 4.5589 | 11.221 | .62639E-01 | -.87452E-01 |
| .71207 | .60659 | .62639E-01 | 4.7796 | -.27320 |
| .25734 | .14326 | -.87452E-01 | -.27320 | 3.0608 |

ELAPSED TIME = 670. MILLISECONDS