**TITLE (Main and Subtitle)**

SELF-CRITICAL, AND ROBUST, ESTIMATES FOR THE PARAMETERS OF THE MULTIVARIATE NORMAL DISTRIBUTION

**AUTHOR(s)**

N. J. Delaney
A. S. Paulson

**PERFORMING ORGANIZATION NAME AND ADDRESS**

Rensselaer Polytechnic Institute
Troy, New York 12181

**CONTRIBUTING OFFICE NAME AND ADDRESS**

Approved for public release; distribution unlimited.

**REPORT DATE**

1 June 1982

**NUMBER OF PAGES**

15

**ABSTRACT**

This algorithm yields joint robust estimates of the location vector and the variance-covariance matrix for samples from the multivariate normal distribution. The degree of robustness depends on a single filtering parameter, c, set by the user. The algorithm provides, for each observation, an internally determined weight which may be used to identify potential outliers.
SELF-CRITICAL, AND ROBUST, ESTIMATES FOR THE PARAMETERS
OF THE MULTIVARIATE NORMAL DISTRIBUTION

by

N. J. Delaney
Union College

and

A. S. Paulson*
Rensselaer Polytechnic Institute

*Research supported in part by U.S. Army Research Office under contract
DAA G29-81-K-0110.
Self-Critical, and Robust, Estimates for the Parameters of the Multivariate Normal Distribution

Keywords: Robust estimation; multivariate normal distribution; M-estimators; outlier identification.

Language: ISO Fortran

Purpose: This algorithm yields joint robust estimates of the location vector and the variance-covariance matrix for samples from the multivariate normal distribution. The degree of robustness depends on a single filtering parameter, \( c \), set by the user. The algorithm provides, for each observation, an internally determined weight which may be used to identify potential outliers.
Theory:

The problems inherent in multivariate parameter estimation and the identification of potential outliers have been documented in Gnanadesikan (1977, Chapter 5) and Barnett and Lewis (1979, Chapter 6). Procedures for the simultaneous estimation of the location vector $\mu$ and the covariance matrix $\Sigma$ of the $p$-variate Gaussian distribution

$$f(x|\mu, \Sigma) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{ -\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right\}, \quad (1)$$

given a random sample $x_1, x_2, \ldots, x_n$, have been discussed by Maronna (1976) and Huber (1977, Chapter 5). Our procedure stems from a single underlying primitive principle which reduces to likelihood as a special case and has desirable properties not possessed by competitors. Let $c$ be a real number. The self-critical estimators $\overline{\mu}$ and $\overline{\Sigma}$ are determined from the zeros of

$$\sum_{i=1}^{n} \frac{f_i^c}{Q} \left\{ (1+c) \frac{\partial \log f_i}{\partial \mu} - \frac{1}{Q} \frac{\partial Q}{\partial \mu} \right\} = 0, \quad (2)$$

and

$$\sum_{i=1}^{n} \frac{f_i^c}{Q} \left\{ (1+c) \frac{\partial \log f_i}{\partial \Sigma} - \frac{1}{Q} \frac{\partial Q}{\partial \Sigma} \right\} = 0, \quad (3)$$

where

$$Q(\mu, \Sigma; c) = \int_{R_p} f^{1+c}(x|\mu, \Sigma) dx$$

$$= \{(1+c)^{p/2}(2\pi)^{p} |\Sigma|^{-1/2} \} . \quad (4)$$

The arguments of $f$ and $Q$ have been suppressed in (2) and (3) for notational convenience. Equations (2) and (3) may also be obtained from the maximization of
with respect to $\bar{u}$ and $\bar{v}$ (Paulson and Delaney, 1982). It may be shown that

$$\lim_{c \to 0} L_c = \sum_{i=1}^{n} \log f_i$$

so that the self-critical procedure reduces to maximum likelihood as a special case. These procedures may also be developed from consideration of the generalized mean. We have termed the procedure self-critical because the information component supplied by the expressions in brackets {.} in (2) and (3) are "fed back" through the assumed density $f$ with degree of criticism determined by $c$. Observations which receive relatively low final weights $f_i^{c}/Q$, a scale-free expression, are candidates for special examination as potential outliers.

Equations (2) and (3) reduce, on simplification, to the joint iterative forms

$$\frac{\bar{u}}{m+1} = \frac{n}{i=1} w_{mi} x_i$$

and

$$\frac{\bar{v}}{m+1} = (1+c) \sum_{i=1}^{n} w_{mi} (x_i - \bar{u}) (x_i - \bar{u})^T,$$

where

$$w_{mi} = \frac{\exp\left(-\frac{c}{2} (x_i - \bar{u})^T \frac{1}{\bar{v}} (x_i - \bar{u})\right)}{\sum_{i=1}^{n} \exp\left(-\frac{c}{2} (x_i - \bar{u})^T \frac{1}{\bar{v}} (x_i - \bar{u})\right)}$$

for $m = 0, 1, ..., m^*$. Initial estimates $\bar{u}_0$ and $\bar{v}_0$ are required to set the
iterative procedure embodied in (7), (8), and (9) in motion. We have frequently used \( w_{m^*i} \) instead of the final \( f_i^C/Q \) to assess patterns in the data with respect to the multivariate Gaussian assumption and the internal consistency of the data.

It is recommended that the user experiment with the value \( c \) in examining a given data set. The estimators \( \tilde{\mu} \) and \( \tilde{\nu} \) are increasingly robust with increasing \( c \) because of the increasingly self-critical nature of the procedure. If for several values of \( c \), similar estimates of \( \mu \) and \( \nu \) are obtained and no weight \( w_{m^*i} \) is relatively small, then the data and Gaussianity are self-consistent. If the estimates of \( \mu \) and \( \nu \) vary substantially with \( c \) or at least one value of \( w_{m^*i} \) is small relative to the remainder, then those observations so labeled should be set aside for special examination vis-à-vis the remainder. Since the \( w_{m^*i} \) induce a virtually automatic ranking of the observations, they are very useful in determining patterns in the data. We have typically used \( 0 < c < 1 \) for the multivariate Gaussian distribution although Paulson, Presser, and Nicklin (1982) have found use for negative values of \( c \), as well as values of \( c \) in excess of unity. The magnitude to which \( c \) may be set is to a large extent dependent on the sample size.

The estimators \( \tilde{\mu} \) and \( \tilde{\nu} \), for all \( c > 0 \), are location and scale invariant, are M-estimators, are jointly asymptotic normal, have closed, bounded and redescendent to zero influence functions, and are, for moderate values of \( c \), relatively efficient (Paulson and Delaney, 1982).
Numerical Method:

The algorithm begins by computing initial estimates $\mu_0 = \hat{\mu}$, the usual vector of sample means, and $\Sigma_0 = \hat{\Sigma}$, the maximum likelihood estimator of the dispersion matrix. Formulae (7), (8), and (9) are the basis for the self-critical iterative procedure. Successive estimates of $\mu$ and $\Sigma$ are generated until the norms of consecutive estimates are less than a user specified small amount, ETA, for three successive iterations. Generally, we have used $ETA = 10^{-4}$ or $10^{-5}$. Of course, the number of iterations required for convergence will depend on the value of ETA specified.
Structure:

SUBROUTINE SCEST(X,M,N,MDIM,NDIM,C,ETA,IMAX,MU,V,WT,MUZ,VZ,IT,IFAULT)

Formal parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Real array (M,N)</td>
<td>input: data, M dimensions, N points</td>
</tr>
<tr>
<td>M</td>
<td>Integer</td>
<td>input: the number of dimensions</td>
</tr>
<tr>
<td>N</td>
<td>Integer</td>
<td>input: the number of sample points</td>
</tr>
<tr>
<td>MDIM</td>
<td>Integer</td>
<td>input: the row dimension of X in calling program</td>
</tr>
<tr>
<td>NDIM</td>
<td>Integer</td>
<td>input: the column dimension of X in calling program</td>
</tr>
<tr>
<td>C</td>
<td>Real</td>
<td>input: filtering parameter</td>
</tr>
<tr>
<td>ETA</td>
<td>Real</td>
<td>input: convergence criterion for norms of consecutive estimates</td>
</tr>
<tr>
<td>IMAX</td>
<td>Real</td>
<td>input: maximum number of iterations desired</td>
</tr>
<tr>
<td>MU</td>
<td>Real array (M)</td>
<td>output: final location vector estimate, M dimensions</td>
</tr>
<tr>
<td>V</td>
<td>Real array (M,M)</td>
<td>output: final var-cov matrix estimate</td>
</tr>
<tr>
<td>WT</td>
<td>Real array (N)</td>
<td>output: final weight assigned to each sample point in computation of estimates</td>
</tr>
<tr>
<td>MUZ</td>
<td>Real array (M)</td>
<td>output: initial location estimate; vector of means</td>
</tr>
<tr>
<td>VZ</td>
<td>Real array (M,M)</td>
<td>output: initial var-cov estimate; MLE</td>
</tr>
<tr>
<td>IT</td>
<td>Integer</td>
<td>output: the number of iterations used</td>
</tr>
<tr>
<td>IFAULT</td>
<td>Integer</td>
<td>output: 0 if no errors in computation, 1 if estimate of var-cov matrix is not positive definite, 2 if accuracy test failed on inverse calculation, 3 if convergence criteria not met in max desired iterations</td>
</tr>
</tbody>
</table>
Restrictions:

The number of dimensions must not exceed 10, unless the dimensions of dummy arrays VINV, TMP, MUSAV, VSAV, and VNORM are redeclared (see subroutines WEIGHT and CKNRM). This procedure requires a matrix inversion subroutine. As written, subroutine WEIGHT calls on the IMSL (International Mathematical and Statistical Library) subroutine LINV2F which inverts a matrix with specifiable accuracy = IDGT. The use of alternate matrix inversion routines would require rewriting this portion of the subroutine WEIGHT.

Precision:

The DOUBLE PRECISION version is recommended due to the iterative nature of the algorithm and the matrix inverse calculation.

Time:

Experience with an IBM3033 computer has shown that the procedure averaged 0.03 seconds for a bivariate sample of size 20 and 0.08 seconds for a trivariate sample of size 25. These computations were done using DOUBLE PRECISION and a convergence criterion, ETA = 10^{-5}. 
References:


APPENDIX

FORTRAN PROGRAM
AND
SUBROUTINES
REAL*8 X(5,150),MU(5),V(5,5),WT(150),MUZ(5),VZ(5,5)
REAL*8 ETA,C,CSAVE(5),TMP(5),TMP2(5)
DIMENSION VP(30)
M=5
MDIM=150
C DRIVER INPUT: N SAMPLE PTS<= 150, M DIMENSIONS <= 5
C IMAX MAX ITERs ALLOWED, NC VALUES OF C INPUT
C ETA CONVERG. CRITERION FOR FOMS
READ(5,900) N,M,IMAX,NC,ETA
900 FORMAT(I3,E10.5)
C DRIVER INPUT: CSAVE ARRAY OF C VALUES
READ(5,901)(CSAVE(I),I=1,NC)
901 FORMAT(5F5.3)
C DRIVER INPUT: VP ARRAY FOR FORMAT OF DATA CANDS PUT IN
READ(5,902) VP
902 FORMAT(30A&)
C DRIVER INPUT: Y(I,J) DATA ONE SAMPLE POINT PER CARD
DO 10 J=1,N
READ(5,VP)(Y(I,J),I=1,M)
10 CONTINUE
DO 100 KC=1,NC
C=CSAVE(KC)
CALL SCEST(X,M,N,MDIM,MDIM,C,ETA,IMAX,MU,V,WT,MUZ,VZ,IT,IFault)
WRITE(6,910)
910 FORMAT(1H1,17X,20H******************************)
IF(IFault.EQ.0)GO TO 30
C NONNORM. TERMINATION MESSAGES
GO TO (11,12,13),IFault
11 WRITE(6,911)IT
911 FORMAT(1HO,46HESTIMATE OF VAR-COV MATRIX ALGORITHMICALLY NOT
& 19H POSITIVE DEFINITE,4X,10HITER NO. = ,I5)
GO TO 30
12 WRITE(6,912)IT
912 FORMAT(1HO,45HACCURACY TEST FAILED ON INVERSE CALCULATION
& 10HITER NO. = ,I5)
GO TO 30
13 WRITE(6,913)IMAX
913 FORMAT(1HO,32HCONVERGENCE CRITERIA NOT MET IN I5,2X,10HITERATIONS)
C NORMAL TERMINATION AND OUTPUT
30 WRITE(6,930)C,ETA
930 FORMAT(1HO,10X,34H*** SELF-CRITICAL ESTIMATES ***/
& 28HFILTERING PARAMETER, C = F7.4,3X,
& 20HCRITERION = E12.5)
C COMPUTE INITIAL + FINAL CORRELATION ESTIMATES
C STORE IN LOWER OFF-DIAG OF VAR-COV FOR OUTPUT
DO 35 I=1,M
TMP2(I)=DSQRT(VZ(I,I))
35 CONTINUE
K=M-1
DO 45 I=1,K
L=I+1
DO 40 J=L,M
VZ(J,I)=VZ(I,J)/(TMP2(I)*TMP2(J))
40 CONTINUE
45 CONTINUE
C PRINT OUT INITIAL ESTIMATES
WRITE(6,931)
931 FORMAT(1H0,10X,17HINITIAL ESTIMATES)
DO 50 I=1,M
WRITE(6,932)MUZ(I),(VZ(T,J),J=1,M)
932 FORMAT(1H0,F10.5,8X,F10.5,3X))
50 CONTINUE
IF(IT.EQ.1.AND.IFAULT.GT.0)GO TO 100
C PRINT OUT FINAL SC ESTIMATES
WRITE(6,933)IT
933 FORMAT(1H0,10X,26HFINAL ESTIMATES ITEF NO. =,I5)
DO 60 I=1,M
WRITE(6,932)MU(I),(V(I,J),J=1,M)
60 CONTINUE
C PRINT OUT FINAL SC WEIGHTS PER EACH SAMPLE POINT
WRITE(6,934)
934 FORMAT(1H0,3X,3HNO.,3X,6HWEIGHT,10X,12HSAMPLE POINT)
DO 70 J=1,N
WRITE(6,935)J,WT(J),(X(I,J),I=1,M)
935 FORMAT(1H0,I5,3X,E12.5,3X,E10.5)
70 CONTINUE
100 CONTINUE
STOP
END
SUBROUTINE Scest(X, N, N, NDim, MDIM, ETA, IMAX, MU, V, WT, MUZ, VZ, IT, IFAULT)

CALCULATES SELF-CRITICAL ESTIMATES

REAL*8 X(N, N, N, NDim, NDim), MU(NDim), V(NDim, NDim), WT(NDim)
REAL*8 MUZ(NDim), VZ(NDim, NDim), ZERO, ONE
DATA ZERO, ONE/O.0DO, 1.0DO/
IFault=0

OBTAIN INITIAL ESTIMATES- MLE
CALL INITL(X, N, N, MUZ, VZ, NDim, NDim)

DO 10 I=1, N
  MU(I)=MUZ(I)
DO 5 J=1, N
  V(I, J)=VZ(I, J)
10 CONTINUE

ITERATIVE FORMATION OF NEW ESTIMATES
DO 100 IT=1, IMAX
CALL WEIGHT(X, N, N, NDim, NDim, C, MU, V, WT, IFAULT)
IF (IFault .NE. 0) RETURN

LOCATION ESTIMATE
DO 20 I=1, N
  MU(I)=ZERO
DO 15 J=1, N
  MU(I)=MU(I) + X(I, J) * WT(J)
20 CONTINUE

VAR-COV ESTIMATE
DO 40 I=1, N
  DO 35 J=1, N
    V(I, J)=ZERO
  30 DO 30 F=1, N
    V(I, J)=V(I, J) + (X(I, K) - MU(I)) * (X(J, K) - MU(J)) * WT(K)
30    V(I, J)=V(I, J) * (C+ONE)
    IF (I .NE. J) V(J, I)=V(I, J)
35 CONTINUE
40 CONTINUE

CHECK CONVERGENCE CRITERIA
CALL CKNPM(M, NDim, ETA, MU, V, IT, IFLAG)
IF (IFLAG .EQ. 1) RETURN
100 CONTINUE

CONVERGE. CRIT. NOT MET IN MAX ITERATIONS DESIRED
IFault=3
RETURN
END

SUBROUTINE INITL(X, N, N, MUZ, VZ, NDim, NDim)

FORMS INITIAL ESTIMATES- MLE

REAL*8 X(NDim, NDim), MUZ(NDim), VZ(NDim, NDim)
REAL*8 DENOM, ZERO, ONE, SUM
DATA ZERO, ONE/O.0DO, 1.0DO/
DENOM=CFL0AT(N)
DO 10 I=1, N
  MUZ(I)=ZERO
DO 5 J=1, N
SUBROUTINE WEIGHT(X,M,N,MDIM,KDIM,MU,V,WT,IFault)
FORMS WEIGHTS FOR EACH SAMPLE POINT
REAL*8 X(MDIM,MDIM),MU(MDIM),V(MDIM,MDIM),WT(MDIM)
REAL*8 VINV(10,10),TMP(10,10),WKAPEA(50)
REAL*8 R,SUM,ZERO,TWO
DATA ZERO,TWO/0.0D0,2.0D0/
DO 10 I=1,M
DO 5 J=1,M
5 TMP(I,J)=V(I,J)
10 CONTINUE

OBTAIN INVERSE OF CURRENT VAR-COV MATRIX ESTIMATE
LINV2F IS IMSL SUBROUTINE WHICH INVERTS MATRIX WITH
SPECIFIED ACCURACY=IDGT
IDGT=5
IER=0
CALL LINV2F(TMP,M,10,VINV,IDGT,WKAPEA,IER)
IF(IER.EQ.0) GO TO 20

ERRORS IN MATRIX INVERSION
IF(IER.GE.129) IFault=1
IF(IER.EQ.341) IFault=2
RETURN

END OF SECTION REQUIRED WITH IMSL SUBROUTINE

20 SUM=ZERO
DO 35 J=1,N
R=ZERO
DO 30 I=1,M
DO 25 K=1,M
25 R=R+(X(I,J)-MU(I))*VINV(I,K)*(X(K,J)-MU(K))
30 CONTINUE
R=R*TWO
IF(R.LT.ZERO) IFault=1
IF(IFault.EQ.1) RETURN
IF(R.LT.174.) WT(J)=DEYP(-R)
IF(R.GE.174.) WT(J)=ZERO
SUM=SUM+WT(J)
35 CONTINUE
DO 40 J=1,N
40 WT(J)=WT(J)/SUM
RETURN
END
SUBROUTINE CKNNRM (M, MDIM, ETA, MU, V, IT, IFLAG)

CHECKS CONVERGENCE CRITERIA

REAL*8 MU(MDIM), V(MDIM, MDIM), MUSAV(10), VSAV(10, 10), VNF(M)
REAL*8 MUNRM, ZERO
DATA ZERO/0.0000/, MCTMU/0/, MCTV/0/
IFLAG=0
IF (IT.EQ.1) GO TO 20

LOCATION NORM = SUM ABSOLUTE DIFFERENCE OF

CCNSEC. ESTIMATE ELEMENTS

VAR-COV. NORM = MAX ROW SUM ABS. DIFFERENCE

CCNSEC. ESTIMATE ELEMENTS

MUNRM=ZERO
DO 10 I=1,M
MUNRM=MUNRM+DABS (MUSAV (I) - MU(I))
VNRM (I)=ZERO
DO 5 J=1,M
5 VNRM (I)=VNRM (I) + DABS (VSAV(I,J) - V(I,J))
10 CONTINUE

MAX =1
DO 15 I=1,M
15 CONTINUE
IF (VNRM (I).GT.VNRM(MAX)) MAX=I

IF (VNRM(MAX).LT.ETA) MCTV=MCTV+1
IF (VNRM(MAX).GE.ETA) MCTV=0
IF (MUNRM.LT.ETA) MCTMU=MCTMU+1
IF (MUNRM.GE.ETA) MCTMU=0

CONVERGENCE CRITERIA MET ON 3 CONSEC. ESTIMATES
IF (MCTV.GT.2. AND. MCTMU.GT.2) IFLAG=1

SAVE CURRENT ESTIMATES

20 DO 30 I=1,M
25 VSAV(I,J)=V(I,J)
30 CONTINUE
RETURN
END