#44

Statistische Hefte 28, 217–231 (1987)

Statistische Hefte
Statistical Papers

Algorithms

An adjusted likelihood-ratio algorithm for the Behrens-Fisher problem

Hamparsum Bozdogan and Donald E. Ramirez

Received: Aug. 13, 1986; revised version: Apr. 1, 1987

Key Words and Phrases: Adjusted likelihood; Scaled likelihood; Behrens-Fisher problem; Nonunique maximum likelihood estimators

We present an algorithm which will incorporate an adjusted likelihood and a scaled likelihood procedure for the Behrens-Fisher problem. Nonuniqueness of the maximum likelihood estimates for the common mean are demonstrated on a simple example showing how ill-conditioned the maximum likelihood equations can be. Our results, however, show that the significance levels are stable even when the means are sensitive to perturbations.

1. INTRODUCTION AND THE BEHRENS-FISHER PROBLEM

In many practical situations it is necessary to test the equality of the means when neither is known, and we cannot reasonably assume that the unknown variances σ_1^2 and σ_2^2 are equal. This is the case in many experimental situations where the researcher wishes to compare the means, but he is unwilling to assume that $\sigma_1^2 = \sigma_2^2$, or it is inappropriate to make such an assumption.

The comparison of two normal means, when the variances are unknown and unequal, is called the *Behrens-Fisher problem*.

More formally, suppose we have a random sample of size n_1, say $X_{11}, X_{12}, \cdots, X_{1n_1}$, from $N(\mu_1, \sigma_1^2)$ and a second independent sample of size n_2, say $X_{21}, X_{22}, \cdots, X_{2n_2}$, from $N(\mu_2, \sigma_2^2)$. We are interested in testing

$$H_0: \mu_1 = \mu_2 \ (= \mu) \tag{1.1}$$

against

$$H_a: \mu_1 \neq \mu_2 \tag{1.2}$$

when σ_1^2 and σ_2^2 are unknown and not assumed to be equal.

In an earlier paper, Bozdogan and Ramirez (1986) studied the Behrens-Fisher problem via an adjusted likelihood-ratio test using the maximum likelihood estimates (MLEs) of the parameters under both the null and the alternative models. This procedure allows the significance levels to be adjusted in accordance with the degrees of freedom to balance the risk due to the bias in using the maximum likelihood estimates and the risk due to the increase of variance.

In a recent paper, Sugiura and Gupta (1985) showed that the maximum likelihood estimate of the common mean μ may not be unique. In this paper, we present an algorithm which will incorporate our adjusted likelihood procedure, as well as a scaled likelihood procedure. This algorithm checks for the rare event of nonunique MLE for the common mean. In this paper, we further demonstrate these procedures on data sets which will show the nonuniqueness of the MLE for the common mean. The examples show that the maximum likelihood equations can be quite ill-conditioned. Fortunately in this case, the significance levels are not sensitive to the perturbations in the means.

2. NOTATION

In this paper we assume that

$$X_{1i} \sim N(\mu_1, \sigma_1^2), \quad i = 1,2,\cdots,n_1$$

and (2.1)

$$X_{2i} \sim N(\mu_2, \sigma_2^2), \quad i = 1,2,\cdots,n_2$$

are independent random samples of sizes n_1 and n_2, respectively. We let $\overline{X}_1$ and $\overline{X}_2$ denote the sample means, S_{u1}^2 and S_{u2}^2 denote the unbiased sample variances, and S_1^2 and S_2^2 denote the biased sample variances.

3. THE WELCH PROCEDURE

When the population variances are not equal $(\sigma_1^2 \neq \sigma_2^2)$ and are unknown, then the most common procedure for testing the equality of means is the Welch procedure [see, for example, Welch (1937)]. For the Welch procedure, one computes

$$t_w = \frac{\overline{X}_1 - \overline{X}_2}{\left[\frac{S_{u1}^2}{n_1} + \frac{S_{u2}^2}{n_2}\right]^{1/2}}, \tag{3.1}$$

The Welch procedure assumes that $S_{u1}^2/n_1 + S_{u2}^2/n_2$ is approximately distributed as $(\sigma^2/\nu)\chi_\nu^2$, for some $\nu > 0$ and $\sigma^2 > 0$. By matching first and second moments, Welch showed

$$t_w \mathrel{\dot\sim} t_{\nu_w} \text{ (approximately)} \tag{3.2}$$

with t_{ν_w} a t-distribution with degrees of freedom

$$\nu_w = \frac{\left[\frac{S_{u1}^2}{n_1} + \frac{S_{u2}^2}{n_2}\right]^2}{\frac{(S_{u1}^2/n_1)^2}{n_1-1} + \frac{(S_{u2}^2/n_2)^2}{n_2-1}} \tag{3.3}$$

4. THE BEHRENS-FISHER MLE

Under the null hypothesis H_0: $\mu_1 = \mu_2$ $(=\mu)$, the MLEs are the roots of the set of three following equations:

$$\left.\begin{aligned} &\frac{n_1(\bar{X}_1-\mu)}{\sigma_1^2} + \frac{n_2(\bar{X}_2-\mu)}{\sigma_2^2} = 0 \\ &\sigma_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1} (X_{1i}-\mu)^2 = S_1^2 + (\bar{X}_1-\mu)^2 \\ &\sigma_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2} (X_{2i}-\mu)^2 = S_2^2 + (\bar{X}_2-\mu)^2 \end{aligned}\right\} . \tag{4.1}$$

From the first equation in (4.1), we obtain very readily the equation for $\hat{\mu}$. Thus, under H_0, the maximum likelihood equations are given by

$$\hat{\mu} = \frac{A\bar{X}_1 + B\bar{X}_2}{A + B}, \quad A = (\hat{\sigma}_1^2/n_1)^{-1}, \quad B = (\hat{\sigma}_2^2/n_2)^{-1}, \tag{4.2}$$

and

$$\hat{\sigma}_1^2 = S_1^2 + (\bar{X}_1-\hat{\mu})^2,$$

$$\hat{\sigma}_2^2 = S_2^2 + (\bar{X}_2-\hat{\mu})^2.$$

We note that S_1^2 and S_2^2 are the biased sample variances and they need to be distinguished from the unbiased sample variances S_{u1}^2 and S_{u2}^2. Also from (4.1), we see that $\hat{\mu}$ solves the cubic polynomial in μ:

$$n_1(S_2^2 + \overline{X}_2^2 - 2\overline{X}_2\mu + \mu^2)(\overline{X}_1-\mu)$$

$$+ n_2(S_1^2 + \overline{X}_1^2 - 2\overline{X}_1\mu + \mu^2)(\overline{X}_2-\mu) = 0 \qquad (4.3)$$

Sugiura and Gupta (1985) have shown that the cubic polynomial in (4.3) has either a unique solution with large probability, or three solutions with small positive probability.

Expanding the cubic polynomial in (4.3) as

$A\mu^3 + B\mu^2 + C\mu + D = 0$ yields:

$$A = \frac{1}{n_1} + \frac{1}{n_2}$$

$$B = - \left[(\frac{2}{n_1} + \frac{1}{n_2})\ \overline{X}_1 + (\frac{1}{n_1} + \frac{2}{n_2})\ \overline{X}_2 \right] \qquad (4.4)$$

$$C = \frac{S_1^{\ 2}}{n_1} + \frac{S_2^{\ 2}}{n_2} + \frac{\overline{X}_1^{\ 2}}{n_1} + \frac{\overline{X}_2^{\ 2}}{n_2} + 2\ (\frac{1}{n_1} + \frac{1}{n_2})\ \overline{X}_1\ \overline{X}_2$$

$$D = - \left[\frac{\overline{X}_1^{\ 2}\ \overline{X}_2}{n_1} + \frac{\overline{X}_1\ \ \overline{X}_2^{\ 2}}{n_2} + \frac{S_1^{\ 2}}{n_1}\ \overline{X}_2 + \frac{S_2^{\ 2}}{n_2}\ \overline{X}_1 \right]$$

The usual procedure for analytically solving a cubic is as follows. We rewrite the cubic first as:

$$\mu^3 + p\mu^2 + q\mu + r = 0, \qquad (4.5)$$

and, then as

$$(\mu + \frac{p}{3})^3 + a\ (\mu + \frac{p}{3}) + b = 0 \qquad (4.6)$$

with

$$a = \frac{1}{3}\ (3q - p^2) \qquad (4.7)$$

$$b = \frac{1}{27}(2p^3 - 9pq + 27r).$$

Let

$$\text{DISC} = \frac{b^2}{4} + \frac{a^3}{27}. \qquad (4.8)$$

The roots of the cubic are unique provided that either DISC > 0 or DISC = a = b = 0. If the roots are not unique, then trigonometric solutions can be found [e.g., Hodgman (1957)].

5. NUMERICAL EXAMPLES.

We now examine the following three examples.

Example 1

X_1	X_2
0	6
0	6
2	8
-2	4

$\bar{X}_1 = 0$

$\bar{X}_2 = 6$

$s_1^2 = 2$

$s_2^2 = 2$

$\hat{\mu} = 0.35425$ or

$\hat{\mu} = 5.6458$

Example 2

X_1	X_2
.01	6
0	6
2	8
-2	4

$\bar{X}_1 = .0025$

$\bar{X}_2 = 6$

$s_1^2 = 2.00002$

$s_2^2 = 2$

$\hat{\mu} = 5.6456$

Example 3

X_1	X_2
0	5.99
0	6
2	8
-2	4

$\bar{X}_1 = 0$

$\bar{X}_2 = 5.9975$

$s_1^2 = 2$

$s_2^2 = 2.00002$

$\hat{\mu} = 0.35442$

The first example is contrived to produce a nonunique MLE for the common mean $\hat{\mu}$. The second and third examples show how ill-conditioned the data set is. A slight perturbation in the data shifts $\hat{\mu}$ from 5.6456 in Example 2 to 0.35442 in Example 3. Example 1 provides a concrete example to the phenomenon observed by Sugiura and Gupta (1985).

The Behrens-Fisher maximum likelihood equations will have ...ble solutions when the sample means are close to each ...her. The numerical problems occur when the means are very far apart. In this case the MLE for the common mean is close to one of the sample means. The third root is between the sample means and is not a local maximum but rather a local minimum for the likelihood equation. A heuristic argument is as follows. The distribution with mean near the middle of the sample means can have $n = n_1 + n_2$ data points as outliers while the distribution centered at one of the sample means will have only n_1 or n_2 data points as outliers.

In the next section, we will see for testing the null hypothesis $H_0 : \mu_1 = \mu_2 \ (= \mu)$ of equal means that the significance levels are stable on these data sets.

6. THE ADJUSTED LIKELIHOOD PROCEDURE

In a previous paper, Bozdogan and Ramirez (1986) showed that the likelihood ratio test can be improved by replacing the approximation

$$- 2 \ell n \lambda \sim \chi_1^2 \tag{6.1}$$

by

$$- 2 \ell n \lambda \sim \chi^2_{\nu_w} \tag{6.2}$$

where λ denotes the likelihood ratio test statistic for H_0 and H_0 and H_a given by (1.1) and (1.2), and where ν_w is given by (3.3).

7. THE SCALED LIKELIHOOD PROCEDURE

Our earlier work yields the approximation

$$E\,(-\,2\ell n\ \lambda) \doteq \frac{\nu_w}{\nu_w-2}\ . \tag{7.1}$$

Consequently, the likelihood ratio test can also be improved by replacing the approximation (6.1) by

$$-\,2\,\left(\frac{\nu_w-2}{\nu_w}\right)\,\ell n\ \lambda \sim \chi_1^2\ . \tag{7.2}$$

Table 7.1 shows our results on the three examples we constructed.

Looking at Table 7.1, we see that although the solutions of the maximum likelihood equations are quite ill-conditioned, the two-sided significance levels α are indeed stable across the the three examples. In each case, one would reject the null hypothesis H_0 of common means with confidence greater than 99%. We note further that since the two groups have nearly equal variances, the significance levels for the pooled t-test and the Welch procedure will be about the same value.

Table 7.1 Results of the Three Numerical Examples

Example 1	Example 2	Example 3
$\hat{\mu} = 0.35425$ $\hat{\sigma}_1^2 = 2.1255$ $\hat{\sigma}_2^2 = 33.874$ or $\hat{\mu} = 5.6458$ $\hat{\sigma}_1^2 = 33.874$ $\hat{\sigma}_2^2 = 2.1255$	$\hat{\mu} = 5.6458$ $\hat{\sigma}_1^2 = 33.844$ $\hat{\sigma}_2^2 = 2.1256$	$\hat{\mu} = 0.35442$ $\hat{\sigma}_1^2 = 2.1256$ $\hat{\sigma}_2^2 = 33.844$
$t_w = -5.196$	$t_w = -5.194$	$t_w = -5.194$
$\nu_w = 6.000$	$\nu_w = 6.000$	$\nu_w = 6.000$
$-2 \ln \lambda = 11.56$	$-2 \ln \lambda = 11.56$	$-2 \ln \lambda = 11.56$
$\alpha_{Welch} = .00202$	$\alpha_{Welch} = .00203$	$\alpha_{Welch} = .00203$
$\alpha_{Adjusted} = .00550$	$\alpha_{Adjusted} = .00551$	$\alpha_{Adjusted} = .00551$
$\alpha_{Scaled} = .00157$	$\alpha_{Scaled} = .00157$	$\alpha_{Scaled} = .00157$

8. THE ALGORITHM BFSOLVE

The subroutine BFSOLVE requires as input the sample sizes N1, N2 for the two groups and the vectors of data X1, X2. For output, the subroutine returns the global means MEAN1, MEAN2 and the global biased sample variances VAR1, VAR2. Under the hypothesis H_0, the subroutine returns the common mean as BFMEAN(1). In case of nonuniqueness, the other mean is returned as BFMEAN(2). All computations are with BFMEAN(1). The MLE for the variances are BFVAR1 and BFVAR2. The two-sided significance levels for the Welch, adjusted-likelihood, and scaled-likelihood procedures are returned as ALPHAW, ALADJ, and ALSCAL. In the case of nonuniqueness, all significance levels should be quite small. Two logical variables UNIQUE, ILLCND are used to flag whether the roots of the maximum likelihood equations are unique and whether DISC from (4.8) satisfies

$$\text{DISC} \leq 0. \qquad (8.1)$$

If (8.1) is satisfied, ILLCND is set equal to .TRUE. and the user is warned of a potential ill-conditioned problem with the common mean μ.

The subroutine BFSOLV is written in FORTRAN 77 and has been used successfully by the authors on a PRIME 550/750/9955 and a CDC cyber 180/855.

```
*************************************************************
      SUBROUTINE BFSOLV(N1,N2,X1,X2,MEAN1,MEAN2,VAR1,VAR2,BFMEAN,
     *             BFVAR1,BFVAR2,M2LNLK,BIAS,ALSCAL,TVALW,DFW,ALPHAW,
     *             ALADJ,UNIQUE,ILLCND)
C     DEVELOPED BY HAMPARSUM BOZDOGAN AND DONALD E. RAMIREZ
C     MATHEMATICS DEPARTMENT, MATH-ASTRO BUILDING
C     UNIVERSITY OF VIRGINIA, CHARLOTTESVILLE, VA. 22903, U.S.A.
C     COPYRIGHT © 02/03/87
C     BFSOLV COMPUTES THE SIGNIFICANCE LEVELS FOR THE WELCH PROCEDURE
C     FOR THE SCALED LIKELIHOOD PROCEDURE AND
C     FOR THE ADJUSTED LIKELIHOOD PROCEDURE
C     N1 INPUT: THE SAMPLE SIZE FOR THE FIRST GROUP
C     N2 INPUT: THE SAMPLE SIZE FOR THE SECOND GROUP
C     X1 INPUT: THE VECTOR OF DATA FOR THE FIRST GROUP
C     X2 INPUT: THE VECTOR OF DATA FOR THE SECOND GROUP
C     MEAN1 OUTPUT: THE SAMPLE MEAN FOR THE FIRST GROUP
C     MEAN2 OUTPUT: THE SAMPLE MEAN FOR THE SECOND GROUP
C     VAR1 OUTPUT: THE BIASED VARIANCE FOR THE FIRST GROUP
C     VAR2 OUTPUT: THE BIASED VARIANCE FOR THE SECOND GROUP
C     BFMEAN OUTPUT: VECTOR OF LENGTH 2 FOR THE MLE FOR THE COMMON MEAN
C     BFVAR1 OUTPUT: THE MLE FOR THE VARIANCE FOR THE FIRST GROUP
C     BFVAR2 OUTPUT: THE MLE FOR THE VARIANCE FOR THE SECOND GROUP
C     M2LNLK OUTPUT: MINUS TWICE THE LOG-LIKELIHOOD RATIO
C     BIAS OUTPUT: EXPECTED VALUE OF M2LNLK
C     ALSCAL OUTPUT: THE SIGNIFICANCE LEVEL FOR THE SCALED LIKELIHOOD
C                  PROCEDURE
C     TVALW OUTPUT: THE WELCH T-STATISTIC
C     DFW OUTPUT: THE DF'S FOR THE WELCH PROCEDURE
C     ALPHAW OUTPUT: THE SIGNIFICANCE LEVEL FOR THE WELCH PROCEDURE
C     ALADJ OUTPUT: THE SIGNIFICANCE LEVEL FOR THE ADJUSTED LIKELIHOOD
C                  PROCEDURE
C     UNIQUE OUTPUT: LOGICAL VARIABLE FOR WHETHER BFMEAN IS UNIQUE
C     ILLCND OUTPUT: LOGICAL VARIABLE FOR WHETHER THE B-F EQUATION
C                  MAY BE ILL-CONDITIONED
      INTEGER N1,N2,IER,
     *        LEVOLD,LEVNEW
      REAL X1,X2,MEAN1,MEAN2,VAR1,VAR2,
     *     BFMEAN(2),BFVAR1,BFVAR2,ALSCAL,
     *     TWELCH,TVALW,ALPHAW,DFW,
     *     M2LNLK,ALADJ,BIAS,ONE
      DIMENSION X1(*),X2(*)
C     UERSET,MDTD,MDCH ARE IMSL ROUTINES
C     UERSET: SETS MESSAGE LEVEL FOR IMSL ROUTINES
C     MDTD: COMPUTES TWO-SIDED TAIL REGIONS FOR THE T-DISTRIBUTION
C     MDCH: COMPUTES LEFT-HANDED TAIL REGIONS FOR THE CHI-SQUARE
C           DISTRIBUTION
      EXTERNAL STATS,UERSET,BFVAR,MDTD,
     *         MDCH,MLUNIQ,TWELCH
      LOGICAL UNIQUE,ILLCND
      INTRINSIC LOG
C     STATS: COMPUTES MEAN AND BIASED VARIANCE
      CALL STATS(X1,N1,MEAN1,VAR1)
      CALL STATS(X2,N2,MEAN2,VAR2)
      LEVNEW = 1
      CALL UERSET(LEVNEW,LEVOLD)
C     MLUNIQ: COMPUTES MLE FOR THE COMMON MEAN
      CALL MLUNIQ(UNIQUE,N1,N2,MEAN1,MEAN2,VAR1,VAR2,BFMEAN,ILLCND)
C     BFVAR: COMPUTES MLE'S FOR SEPARATE VARIANCES
      CALL BFVAR(X1,N1,X2,N2,BFMEAN(1),BFVAR1,BFVAR2)
      M2LNLK=N1*LOG(1.0+(BFMEAN(1)-MEAN1)**2/VAR1)
     *         +N2*LOG(1.0+(BFMEAN(1)-MEAN2)**2/VAR2)
```

```
      TVALW=TWELCH(MEAN1,MEAN2,VAR1,VAR2,N1,N2,DFW)
C     MDTD: COMPUTES TWO-SIDED TAIL REGIONS FOR THE T-DISTRIBUTION
      CALL MDTD(TVALW,DFW,ALPHAW,IER)
      BIAS=DFW/(DFW-2.0)
C     MDCH: COMPUTES LEFT-HANDED TAIL REGIONS FOR THE CHI-SQUARE
C           DISTRIBUTION
      CALL MDCH(M2LNLK,BIAS,ALADJ,IER)
      ALADJ=1.0-ALADJ
      ONE=1.0
C     MDCH: COMPUTES LEFT-HANDED TAIL REGIONS FOR THE CHI-SQUARE
C           DISTRIBUTION
      CALL MDCH(M2LNLK/BIAS,ONE,ALSCAL,IER)
      ALSCAL=1.0-ALSCAL
      RETURN
      END
************************************************************
      SUBROUTINE STATS(A,N,MEAN,VAR)
      REAL A,MEAN,VAR,SUM,SUMSQ
      INTEGER N,ROW
      DIMENSION A(*)
      INTRINSIC REAL
      SUM=0.0
      SUMSQ=0.0
      DO 10, ROW=1,N
         SUM=SUM+A(ROW)
         SUMSQ=SUMSQ+A(ROW)*A(ROW)
10    CONTINUE
      MEAN=SUM/REAL(N)
      VAR=(SUMSQ-SUM*SUM/REAL(N))/REAL(N)
      RETURN
      END
***********************************************************************
      SUBROUTINE MLUNIQ(UNIQUE,N1,N2,MEAN1,MEAN2,VAR1,VAR2,BFMEAN,
     +                  ILLCND)
      INTEGER N1,N2
      REAL MEAN1,MEAN2,VAR1,VAR2,A,B,C,D,DISC
      REAL P,Q,R,AA,BB,SMALL,BFMEAN(2),W,Z,PHI,PI
      REAL LNLIKE,ROOT1,ROOT2,ROOT3,LTROOT,RTROOT,LTLNLK,RTLNLK
      REAL AADIV3,BBDIV2
      INTRINSIC ABS,SQRT,SIGN,ACOS,COS,MIN,MAX
      EXTERNAL LNLIKE
      LOGICAL UNIQUE,ILLCND
      DATA SMALL /1E-7/
      PI=ACOS(-1.0)
C     CUBIC IS A*X**3 + B*X**2 + C*X + D
      A=1.0/N1 + 1.0/N2
      B=-((2.0/N1 + 1.0/N2)*MEAN1 + (1.0/N1 + 2.0/N2)*MEAN2)
      C=VAR1/N1 + VAR2/N2 + MEAN1*MEAN1/N1 + MEAN2*MEAN2/N2
     +          + 2.0*(1.0/N1 + 1.0/N2)*MEAN1*MEAN2
      D=-(MEAN1*MEAN1*MEAN2/N1 + MEAN1*MEAN2*MEAN2/N2
     +  + VAR1*MEAN2/N1 + VAR2*MEAN1/N2)
C     TESTING FOR DISCRIMINANT=0
C     CUBIC REWRITTEN AS X**3 + P*X**2 + Q*X + R
      P=B/A
      Q=C/A
      R=D/A
      AA=(3.0*Q-P*P)/3.0
      AADIV3=AA/3.0
      BB=(2.0*P*P*P-9.0*P*Q+27.0*R)/27.0
      BBDIV2=BB/2.0
      DISC=BB*BB/4.0 +AA*AA*AA/27.0
C     REAL ROOTS OF A CUBIC ARE UNIQUE IF AND ONLY IF DISC > 0
C     OR IF DISC = AA = BB = 0
C     REFERENCE: HODGMAN, CHARLES D. (EDITOR), STANDARD MATHEMATICAL
```