188:1996.04

The Behrens Fisher problem by third order asymptotics

D.A.S. Fraser
Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

Q. Zeng
Department of Mathematics & Statistics
York University
North York, Canada, M3J 1P3

SUMMARY

The Behrens Fisher problem concerns inference for the difference between two means when samples are available for each and the population variance are not assumed equal; the distribution forms for the populations have typically been taken to be normal. For the normal case exact tests of singificance that use all the data information seem unavailable except perhaps in the equal sample sign case. Recent third order asymptotics is used to derive tests for the differences in means, covering both the normal and the nonnormal case. Simulations are used to assess the accuracy of the uniform distribution for the calculated p values.

1. INTRODUCTION

Let $y_{11}, \ldots, y_{1m}$ be a sample from the location scale model $\sigma_1^{-1} f_1\{\sigma_1^{-1}(y_1 - \mu_1)\}$ and $y_{21}, \ldots, y_{2n}$ similarly be from the model $\sigma_2^{-1} f_2\{\sigma_2^{-1}(y_2 - \mu_2)\}$. The generalized Behrens (1929) Fisher (1935) problem is to obtain inference for the difference $\delta = \mu_2 - \mu_1$ in the locations of the two distributions.

We do not assume that the location and scale parameters are means and standard deviations so as to allow distribution forms with long tails such as the Student with small degrees of freedom. This in turn points to a notational arbitrariness most easily noted using the notation $y_1 = \mu_1 + \sigma_1 z$, $y_2 = \mu_2 + \sigma_2 z_2$ with z_1, z_2 respectively from $f_1(z_1)$, $f_2(z_2)$, referred to as the error distributions. For example with the first population, an affine transformation affecting μ_1, σ_1 can be compensated by an inverse transformation affecting z_1. Thus in general contexts some attention needs to be given to a sensible definition for the location and scale parameters; for some discussion see Fraser (1976). A somewhat related concern has been voiced from the experimental side, that to speak of a difference, of means, $\delta = \mu_2 - \mu_1$ when the variances need not be equal is an unrealistic problem. We acknowledge these concerns but choose to address the basic problem as described in the preceding paragraph.

In this paper we discuss the application of recent third asymptotic theory to the generalized Behrens Fisher problem. An oncleve of needed asymptotic results is presented in Section 2. Then in Section 3 we examine the traditional Behrens Fisher problem with standard normal error distributions f_1 and f_2, and obtain expressions for third order p-values $p(\delta)$ for testing a difference δ; these calculations families ingredients, maximum likelihood values, information determinants, and reparameterizations. In Section 4 we describe simulations and compare the resulting p values with the expected uniform $(0, 1)$ distribution.

In Section 5 we examine briefly the generalized Behrens Fisher problem and illustrate this in Section 6 for Student (6) error distributions. The concluding Section 7 gives some background on the Behrens Fisher problem and presents some concluding remarks.

2. RECENT THIRD ORDER ASYMPTOTICS

Third order asymptotic methods have been long available, often in the form of Edgeworth type expansions; see, for example, Feller Vol. 2, pp. 504-525. The practical reliability of these methods, however, was often uncertain with p values sometimes outside the $[0, 1]$ range. The introduction to statistics of saddlepoint methods (Daniels, 1954; Barndorff-Nielsen & Cox, 1979; Lugannani & Rice, 1980) produced formulas of remarkable practical accuracy; the context for these methods focussed on sums and averages when a cumulant generating function was available, either directly or as a by product of an exponential model.

The extension of these more accurate methods to general contexts is relatively recent. Barndorff-Nielsen (1980,1983) established the existence of an approximate third order ancillary, and developed the p^* formula to approximate the conditional probability density given that ancillary. A third order formula for a tail probability for testing a scalar parameter was developed by Barndorff-Nielsen (1986) but its implementation involved location scale correcting a likelihood ratio and its accuracy was somewhat uncertain (Fraser, 1990). A subsequent r^* formula (Barndorff-Nielsen, 1990) was found to give more accurate results and third order accuracy was established later. An alternative approach (Fraser & Reid, 1989, 1993) used tangent exponential models to obtain third order formulas giving accurate approximations. All the preceding formulas for implementation require a full approximate ancillary that complements the maximum likelihood statistic but they do not provide a mechanism for developing such an ancillary; this in effect limits the application of the approximations to canonical parameters of an exponential or transformation model.

An exact first derivative ancillary (Fraser, 1964) was used to develop the tangent directions of a third order ancillary (Fraser & Reid, 1995); also it was shown that these direction at the data point are the only information needed concerning the ancillary, in order to obtain third order inference. We now describe the corresponding formulas for testing a scalar parameter.

Consider a continuous statistical model $f(y;\theta)$ which has asymptotic properties and suppose that $\theta = (\lambda, \psi)$ can be separated into a scalar interest parameter ψ and a nuisance parameter λ of dimension $p-1$. Two third order formulas are available for obtaining significance $p(\psi)$, the Lugannani and Rice (1980) type formula

$$p_1(\psi) = \Phi(r) + \varphi(r)\big(1/r - 1/q\big) \ , \tag{2.1}$$

and the Barndorff-Nielsen (1990) type formula

$$p_2(\psi) = \Phi\{(r) + r^{-1}\log(r/q)\} \ , \tag{2.2}$$

where r is the signed likelihood ratio

$$r = \mathrm{sgn}(\hat{\psi} - \psi) \cdot [2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}] \tag{2.3}$$

and q is a standardized maximum likelihood departure; $\hat{\theta}$ and $\hat{\theta}_\psi$ are the maximum likelihood estimates under the full model and under the restricted model with ψ fixed.

In the case of a scalar parameter θ, and thus no nuisance parameter, we have

$$q = (\hat{\varphi} - \varphi)\hat{\jmath}_{\varphi\varphi}^{1/2} \tag{2.4}$$

where φ is a nominal reparameterization

$$\varphi = \ell_{;v}(\theta : y^0) = \frac{d}{dv}\log f(y;\theta)|_{y^\circ}$$

obtained from the tangent direction v to the third order ancillary; the determination of v is discussed below.

For the more general vector case we need an overall nominal reparameterization

$$\varphi = \ell_{;V}(\theta; y^\circ) = \left\{\frac{d}{dv_1}\log f(y;\theta)|_{y^\circ}, \ldots, \frac{d}{dv_p}\log f(y;\theta)|_{y^\circ}\right\}' \tag{2.5}$$

obtained from tangent directions $V = (v_1, \ldots, v_p)$ to a third order ancillary, and then need a rotated scalar component $\bar{\varphi} = c_p(\hat{\theta}_\psi)\varphi$ that agrees with ψ at $\hat{\theta}_\psi$; for this let $c_p(\hat{\theta}_\psi) = k_p(\hat{\theta}_\psi)/|k_p(\hat{\theta}_\psi)|$ be the unit vector obtained from the pth row in $K(\hat{\theta}_\psi)$:

$$K(\theta) = J_p^{-1}(\theta) \ , \ J_p(\theta) = \frac{\partial \varphi}{\partial \theta} \ . \tag{2.6}$$

Then the standardized maximum likelihood departure for ψ is

$$q = M \cdot \frac{|\hat{\jmath}_{\varphi\varphi}|^{1/2}}{|\jmath_{(\lambda\lambda)}(\hat{\theta}_\psi)|^{1/2}} = (\hat{\bar{\varphi}} - \bar{\varphi}) \cdot \frac{|\hat{\jmath}_{\varphi\varphi}|^{1/2}}{|\jmath_{(\lambda\lambda)}(\hat{\theta}_\psi)|^{1/2}} \tag{2.7}$$

where

$$|\hat{\jmath}_{\varphi\varphi}| = |\hat{\jmath}_{\theta\theta}||J_p(\hat{\theta})|^{-2} \ , \ |\jmath_{(\lambda\lambda)}(\hat{\theta}_\psi)| = |\hat{\jmath}_{\lambda\lambda}(\hat{\theta}_\psi)||J_{p-1}(\hat{\theta}_\psi)|^{-2} \tag{2.8}$$

are overall and nuisance information determinants recalibrated on the φ scale. For this let $J_{p-1}(\theta)$ be the matrix formed by the first $p-1$ columns of $J_p(\theta)$, and let $|J_{p-1}(\theta)|$ be the corresponding matrix volume

$$|J_{p-1}(\theta)| = |J'_{p-1}(\theta) J_{p-1}(\theta)|^{1/2} \ . \tag{2.9}$$

If the denominator of (2.7) were evaluated at $\hat{\theta}$ we would have the ordinary standardization for φ; the modification with evaluation at $\hat{\theta}_\psi$ leads to a marginal p value averaged over conditional nuisance parameter distribution.

In a general context the vector v or the vectors $v_1, \ldots, v_p$ are tangent to an exact first derivative ancillary (Fraser & Reid, 1995). In the present Behrens Fisher context we have a transformation model most easily described in the structural notation (Fraser, 1979)

$$\begin{aligned} y_{1i} &= \mu_1 + \sigma_1 z_{1i} \\ y_{2i} &= \mu_2 + \sigma_2 z_{2i} \ . \end{aligned} \tag{2.10}$$

Location scale model theory gives location vectors; let

$$v_1(1, \ldots, 1, 0, \ldots, 0)m^{-1/2} \ , \ v_2(0, \ldots, 0, 1, \ldots, 1)n^{-1/2} \tag{2.11}$$

be the unit 1 vectors for the first and second sample. And it gives scale vectors: let v_3 and v_4 be the corresponding unit residual vectors for the two samples

$$\begin{aligned} v_3 &= (y_{11} - \bar{y}_1, \dots, y_{1m} - \bar{y}_1, 0, \dots 0)/\{\sum (y_{1i} - \bar{y}_1)^2\}^{1/2} \\ v_4 &= (0, \dots, 0, y_{21} - \bar{y}_2, \dots, y_{2n} - \bar{y}_2)/\{\sum (y_{2i} - \bar{y}_2)^2\}^{1/2} \end{aligned} \tag{2.12}$$

Then v_1, v_2, v_3, v_4 are tangent to the ordinary ancillary of the transformation model.

3. BEHRENS FISHER WITH NORMAL ERROR

The traditional Behrens Fisher problem has a first sample $y_{11}, \dots, y_{1m}$ from a normal (μ_1, σ_1) distribution and a second sample $y_{21}, \dots, y_{2n}$ from a normal (μ_2, σ_2) distribution, and interest lies in the parameter $\delta = \mu_2 - \mu_1$. The model has a likelihood statistic or minimal significient statistic $(\bar{y}_1, \bar{y}_2, S_1^2, S_2^2)$ where S_1^2 and S_2^2 are the sums of squares of residuals for the two samples; this statistic has the same dimension as the parameter. The full model and the reduced model each have both exponential form and transformation form, this contributes substantial simplicity. However the parameter of interest δ is neither a canonical exponential parameter nor a canonical transformation parameter, and this is central to the difficulties with the Behrens Fisher problem.

The likelihood function from the full or reduced model is

$$\begin{aligned} \ell(\theta) = &-\frac{m}{2}\log\sigma_1^2 - \frac{m}{2\sigma_1^2}(\bar{y}_1 - \mu)^2 - \frac{1}{2\sigma_1^2}S_1^2 \\ &-\frac{n}{2}\log\sigma_2^2 - \frac{n}{2\sigma_1^2}(\bar{y}_2 - \mu - \delta) - \frac{1}{2\sigma_2^2}S_2^2 \end{aligned} \tag{3.1}$$

where we use $\theta = (\sigma_1^2, \sigma_2^2, \mu, \delta)$ with $\mu = \mu_1$. From the score function we obtain the overall maximum likelihood estimate

$$\hat{\theta} = (S_1^2/m, S_2^2/n, \bar{y}_1, \bar{y}_2 - \bar{y}_1)' \ , \tag{3.2}$$

and the corresponding information matrix

$$\jmath = \begin{bmatrix} \frac{m}{2\hat{\sigma}_1^2} & 0 & 0 & 0 \\ 0 & \frac{n}{2\hat{\sigma}_2^2} & 0 & 0 \\ 0 & 0 & (\frac{m}{\hat{\sigma}_1^2} + \frac{n}{\hat{\sigma}_2^2}) & \frac{n}{\hat{\sigma}_2^2} \\ & & \frac{n}{\hat{\sigma}_2^2} & \frac{n}{\hat{\sigma}_2^2} \end{bmatrix} \tag{3.3}$$

with determinant $|\hat{\jmath}| = m^2 m^2 / 4\hat{\sigma}_1^4 \hat{\sigma}_2^4$.

In the present normal context, the issue of an ancillary does not arise if we work with the reduced model. Accordingly to obtain the nominal reparameterization, we differentiate with respect to $\bar{y}_1$, $\bar{y}_2$, S_1^2, S_2^2 and obtain

$$\begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{bmatrix} = \begin{bmatrix} m(\mu - \bar{y}_1)/\sigma_1^2 \\ n(\delta + \mu - \bar{y}_2)/\sigma_2^2 \\ -(m-1)/2\sigma_1^2 \\ -(n-1)/2\sigma_2^2 \end{bmatrix} \tag{3.4}$$

which records canonical parameters of the exponential model provided we consider the data point as fixed. An affine transformation would eliminate the data point but we keep the present form to better indicate the more general issues that arise in the nonnormal case. The Jacobian of the parameter change is

$$J_4(\theta) = \begin{bmatrix} m(\bar{y}_1 - \mu)/\sigma_1^4 & 0 & m/\sigma_1^2 & 0 \\ 0 & n(\bar{y}_2 - \delta - \mu)/\sigma_2^4 & n/\sigma_2^2 & n/\sigma_2^2 \\ (m-1)/2\sigma_1^4 & 0 & 0 & 0 \\ 0 & (n-1)/2\sigma_2^4 & 0 & 0 \end{bmatrix} \tag{3.5}$$

with determinant $mn(m-1)(n-1)/4\sigma_1^6\sigma_2^6$. The information determinant for the new parameterization is then $|\hat{\jmath}_{\theta\theta}||J_4(\hat{\theta})|^{-2}$ giving

$$|\hat{\jmath}_{\varphi\varphi}| = \frac{m^2 n^2}{4\sigma_1^4 \hat{\sigma}_2^4} \cdot \left(\frac{mn(n-1)(n-1)}{4\hat{\sigma}_1^6 \hat{\sigma}_2^6}\right)^{-2} = \frac{4\hat{\sigma}_1^8 \hat{\sigma}_2^8}{(m-1)(n-1)} \ . \tag{3.6}$$

Now consider the restricted model with δ fixed. ¿From the score equations we obtain

$$\hat{\mu}_\delta = \left(\frac{m\bar{y}_1}{\hat{\sigma}_{1\delta}^2} + \frac{n\bar{y}_{2\delta}}{\hat{\sigma}_{2\delta}^2}\right)\left(\frac{m}{\hat{\sigma}_{1\delta}^2} + \frac{n}{\hat{\sigma}_{2\delta}^2}\right) \tag{3.7}$$

where $\bar{y}_{2\delta}$ denotes $\bar{y}_2 - \delta$, and

$$\begin{aligned} \hat{\sigma}^2_{1\delta} &= (\bar{y}_1 - \hat{\mu}_\delta)^2 + S_1^2/m \\ \hat{\sigma}^2_{2\delta} &= (\bar{y}_{2\delta} - \bar{\mu}_\delta)^2 + S_2^2/n \ . \end{aligned} \tag{3.8}$$

These equations lend themselves to iterative solution, giving

$$\hat{\theta}_\delta = (\hat{\sigma}^2_{1\delta}, \hat{\sigma}^2_{2\delta}, \hat{\mu}_\delta, \delta) \ .$$

The nuisance parameter information matrix based on $\lambda = (\sigma_1^2\ ,\ 1\sigma_2^2\ ,\ \mu)$ comes easily from the calculations or the top left corner of (3.3),

$$\jmath_{\lambda\lambda}(\hat{\theta}_\delta) = \begin{bmatrix} \dfrac{m}{2\hat{\sigma}^2_{1\delta}} & 0 & \dfrac{m(\bar{y}_1 - \hat{\mu}_\delta)}{\hat{\sigma}^4_{1\delta}} \\ 0 & \dfrac{n}{2\hat{\sigma}^2_{2\sigma}} & \dfrac{n(\bar{y}_2 - \delta - \hat{\mu}_\delta)}{\sigma^4_{2\delta}} \\ \dfrac{m(\bar{y}_1 - \hat{\mu}_\delta)}{\hat{\sigma}^4_{1\delta}} & \dfrac{n(\bar{y}_2 - \delta - \hat{\mu}_\delta)}{\hat{\sigma}^4_{2\sigma}} & \dfrac{m}{\hat{\sigma}^2_{1\delta}} + \dfrac{n}{\hat{\sigma}^2_{2\delta}} \end{bmatrix} . \tag{3.9}$$

Using the matrix $J_3(\theta)$ formed from the first three columns of (3.5) we obtain $J_3'(\theta)\ J_3(\theta)$,

$$\begin{bmatrix} \dfrac{m^2(\bar{y}_1 - \mu)^2}{\hat{\sigma}_1^8} + \dfrac{(m-1)^2}{4\sigma_1^8} & 0 & \dfrac{m^2(\bar{y}_1 - \mu)}{\sigma_1^6} \\ 0 & \dfrac{n^2(\bar{y}_2 - \delta - \mu)^2}{\sigma_2^8} + \dfrac{(n-1)^2}{4\sigma_2^8} & \dfrac{n^2(\bar{y}_2 - \delta - \mu)}{\sigma_2^6} \\ \dfrac{m^2(\bar{y}_1 - \mu)}{\sigma_1^6} & \dfrac{n^2(\bar{y}_2 - \delta - \mu)}{\sigma_2^6} & \dfrac{m^2}{\sigma_1^4} + \dfrac{n^2}{\sigma_2^4} \end{bmatrix} . \tag{3.10}$$

The nuisance information determinant calibrated on the φ scale is then

$$|\jmath_{(\lambda\lambda)}(\hat{\theta}_\delta)| = |j_{\lambda\lambda}(\hat{\theta}_\delta)||J_3'(\hat{\theta}_\delta) J_3(\hat{\theta}_\delta)|^{-1} \tag{3.11}$$

which is easily evaluated from the maximum likelihood $\hat{\theta}_\delta$ and the matrices (3.9), (3.10).

The third order procedure described in Section 2 needs a maximum likelihood estimate in the φ scale. For this we need the scalar parameter $\bar{\varphi}$ that agrees with δ at $\hat{\theta}_\delta$, that is, $\bar{\varphi}$ = constant and δ = constant are tangent to each other at $\hat{\theta}_\delta$. To calculate $\bar{\varphi} = c_4(\hat{\theta}_\delta)\varphi$ we require the unit vector $c_4(\hat{\theta}_\delta) = k_4/|k_4|$ based on the 4th row k_4 in $K(\hat{\theta}_\delta) = J_4^{-1}(\hat{\theta}_\delta)$, which is of course orthogonal to the first three columns of $J_4(\hat\theta_\delta)$; the vector k_4 is easily elicited from (3.5),

$$k_4 = \Big(-\frac{\sigma_{1\delta}^2}{m}\ ,\ \frac{\sigma_{2\delta}^2}{n}\ ,\ \frac{2(\bar y_1 - \hat\mu_\delta)}{m-1}\hat\sigma_{1\delta}^2\ ,\ -\frac{2(\bar y_2 - \delta - \hat\mu_\delta)}{n-1}\hat\sigma_{2\delta}^2\Big)\ . \tag{3.12}$$

Then after further calculation we obtain the maximum likelihood departure

$$M = \hat{\hat\varphi} - \bar y = \{(\hat\delta - \delta + \hat\mu - \hat\mu_\delta)\frac{\hat\sigma_{2\delta}^2}{\hat\sigma_2} - (\hat\mu - \hat\mu_\delta)\frac{\hat\sigma_{1\delta}^2}{\hat\sigma_1^2}\}/|k_4|\ , \tag{3.13}$$

and from it with (3.6) and (3.11) obtain the standardized maximum likelihood departure q in (2.7).

Perhaps the most important ingredient in the formulas (2.1) and (2.2) is the signed likelihood ratio r given by (2.3). We substitute $\hat\theta$ from (3.2) and $\hat\sigma_\delta$ from (3.7) and (3.8) in the likelihood function (3.1) and obtain

$$r = \mathrm{sgn}(\bar y_2 - \bar y_1 - \delta)\Big\{m\log\frac{\hat\sigma_{1\delta}^2}{\hat\sigma_1^2} + n\log\frac{\hat\sigma_{2\delta}^2}{\hat\sigma_2^2}\Big\}^{1/2}\ .$$

The significance function for testing δ from third order asymptotic is then given by (2.1) or (2.2).

4. ASSESSMENT OF THIRD ORDER SIGNIFICANCE FUNCTION

The third order limiting distribution of $p_1(\delta)$ or $p_2(\delta)$ is uniform $(0,1)$, for sample sizes m, n large. For finite sample sizes the actual distribution may depart from this uniform distribution and in doing so may depend on the nuisance parameters. We first check what sort of dependence can exist on the nuisance parameters and then discuss some simulations that assess departure from the uniform distribution.

The signficance functions depend on r and q which in turn depend on $\hat\theta$, $\hat\theta_\delta$, M, $\hat\jmath_{\varphi\varphi}$, and $j_{(\lambda\lambda)}$. We have an interest parameter δ and nuisance parameters σ_1^2, σ_2^2, μ.

The basic quantities $r = r(\delta)$ and $q = q(\delta)$ are invariant under any one-one transformation that leaves the full model and the restricted hypothesis model unchanged; this can be routinely checked.

Now consider $p(\delta)$ obtained from data $\{y_{1i}, y_{2j}\}$ and $\tilde{p}(\delta)$ obtained from data $\{\tilde{y}_{1i}, \tilde{y}_{2j}\} = \{y_{1i} - a, y_{2j} - a\}$. For the modified data only the parameter μ is affected, $\tilde{\mu} = \mu - a$; the full and restricted models are unchanged. Simple calculation shows that r and q are the same for the original and for the modified data. It follows that the distribution of $p_1(\delta)$ or $p_2(\delta)$ does not depend on μ.

Now consider $p(\delta_0)$ obtained from data $\{y_{1i}, y_{2j}\}$ and $\tilde{p}(0)$ obtained from data $\{\tilde{y}_{1i}, \tilde{y}_{2j}\} = \{y_{1i}, y_{2j} - \delta_0\}$. For the modified data only the parameter δ is affected, $\tilde{\delta} = \delta - \delta_0$; thus the restricted model with $\delta = \delta_0$ for the initial data is the same as the restricted model $\tilde{\delta} = 0$ for the modified data. Simple calculations then show that $r(\delta_0)$, $q(\delta_0)$ for the first are equal to $r(0)$, $q(0)$ for the second. It follows that the null distribution of $p_1(\delta)$ or $p_2(\delta)$ does not depend on δ.

Now suppose that $\mu = 0$ and consider $p(0)$ obtained from data $\{y_{1i}, y_{2j}\}$ and $\tilde{p}(0)$ obtained from data $\{\tilde{y}_{1i}, \tilde{y}_{2j}\} = \{cy_{1i}, cy_{2i}\}$ with $c > 0$. For the modified data, the parameters $\tilde{\sigma}_1$, $\tilde{\sigma}_2$ are equal to $c\sigma_1$, $c\sigma_2$ respectively; the full and restricted model are unchanged. Again simple calculation shows that r and q are the same for the original and for the modified data. This together with preceding results shows that the null distribution of $p_1(\delta)$ or $p_2(\delta)$ can depend only on the variance ratio given say by the nuisance parameter $\rho^2 = (\sigma_2^2/n)/(\sigma_1^2/n)$ or $\lambda^2 = \sigma_2^2/\sigma_1^2$, and of course on the sample sizes m, n.

To check on the reliability of the uniform $(0,1)$ distribution for the test quantities $p_1(\delta)$ and $p_2(\delta)$ we sampled $N =$ for the following combinations.

$$\begin{pmatrix} m \\ n \\ p^2 \end{pmatrix} = (\qquad)$$

The sample distribution functions for p_1 and p_2 were very close to the straightline of the uniform distribution (0.1). In practice the tail values are of particularly interest. Accordingly we calculated $r^\circ = \Phi^{-1}(p_1)$, and $r^* = \Phi^{-1}(p_2)$ and plotted values against normal quantiles, examining in particular the values outside the $(-3, +3)$ range. See Figures , .

In conclusion we mention that $p_1(\delta)$ and $p_2(\delta)$ represent probability integral transformations of a marginal distribution recorded on the curve $\hat{\lambda}_\delta = \hat{\lambda}_\delta^\circ$ in the space $R^2 \times \{0, \infty\}^2$ of the variable $(\bar{y}_1, \bar{y}_2, S_1^2, S_2^2)$. The distribution on the curve is obtained by projecting the δ-fixed distribution along the ancillary surfaces for that δ-fixed distribution. Some of these calculations can be done exactly rather than asymptotically. In particular we can integrate the normal distribution on the orbits corresponding to $[a, 1]\{y_{1i}, y_{2i}\} = \{y_{1i} + a\ ,\ y_{2j} + a\}$ and then for $\delta = 0$ integrate the gamma distribution on the orbits corresponding to $[0, c]\{y_{1i}, y_{2i}\} = \{cy_{1i}, cy_{2j}\}$. The resulting distribution can be expressed in terms of

$$u^2 = \frac{s_2^2}{s_1^2} \qquad t = \frac{\bar{y}_2 - \bar{y}_1}{(\frac{s_1^2}{m} + \frac{s_2^2}{n})^{1/2}}$$

where $s_1^2 = S_1^2/(m-1)$ and $s_2^2 = S_2/(n-1)$ are the mean squares for residuals. We then have that

$$u_2 \sim \lambda^2\ F(n-1, m-1)$$

where F has the F distribution on $n-1$ over $m-1$ degrees of freedom, and

$$t|u^2 \sim (\ldots) \quad t(m+n-2, \alpha)$$

where t refers to the noncentral Student distribution with $m+n-2$ degrees of freedom and noncentrality $\alpha = \delta/(\sigma_1^2/m + \sigma_2^2/n)^{1/2}$

For the case $\delta = 0$ we record the reduced model for (t, r):

$$f(t|r^2) \cdot f(r^2) =$$

on $(-\infty, \infty) \times (0, 1)$. In this form we have a 2 dimensional distribution and a single real parameter p^2 and inherit third order similar regions defined by $p_1(0)$ or $p_2(0)$. The asymptotics may not be applicable directly to this model as of the scale transformation used in its derivation is specific to the value $\delta = 0$.

5. GENERALIZED BEHRENS FISHER PROBLEM

Consider the generalized Behrens Fisher problem. Let $y_{11}, \ldots, y_{1m}$ be a sample from the distribution $\sigma_1^{-1} f\{\sigma_1^{-1}(y_1 - \mu_1)\}$ and $y_{21}, \ldots, y_{2n}$ be a sample from the distribution $\sigma_2^{-1} f_2\{\sigma_2^{-1}(y_2 - \mu_2)\}$, and suppose that $\delta = \mu_2 - \mu_1$ is the parameter of interest. We assume that the location and scale parameter have been appropriately defined in relation to the error densities $f_1(z)$, $f_2(z)$.

Location-scale model theory gives an exact ancillary of dimension $m + n - 4$ with tangent vectors given by $V = (v_1, \ldots, v_4)$ recorded in (2.11), (2.12).

For the overall log-likelihood function it is convenient to let $\ell_1(z) = \log f_1(z)$ and $\ell_2(z) = \log f_2(z)$; we obtain

$$\ell(\theta) = \sum \ell_1\{\sigma_1^{-1}(y_{1i} - \mu)\} - m \log \sigma_1 + \sum \ell_2\{\sigma_2^{-1}(y_{2i} - \mu + \delta)\} - n \log \sigma_2$$

where $\theta = (\sigma_1, \sigma_2, \mu, \delta)$ and $\mu = \mu_1$, $\delta = \mu_2 - \mu_1$.

For the nominal reparameterization we let $\ell_1'(z) = d\ell_1(z)/dz$, $\ell_2'(z) = d\ell_2(z)/dz$; then

$$\begin{aligned}
\varphi_1 &= \sum \ell_1'\{\sigma_1^{-1}(y_{1i} - \mu_1)\}/\sqrt{m}\sigma_1 \\
\varphi_2 &= \sum \ell_2'\{\sigma_2^{-1}(y_{2i} - \mu_2)\}/\sqrt{n}\sigma_2 \\
\varphi_3 &= \sum \ell_1'\{\sigma_1^{-1}(y_{1i} - \mu_1)\}(y_{1i} - \bar{y}_1)/\{\sum (y_{1i} - \bar{y}_1)^2\}^{1/2}\sigma_1 \\
\varphi_4 &= \sum \ell_2'\{\sigma_2^{-1}(y_{2i} - \mu_2)\}(y_{2i} - \bar{y}_2)/\{\sum (y_{2i} - \bar{y}_2)^2\}^{1/2}\sigma_2
\end{aligned}$$

The maximum likelihood estimates $\hat{\theta}$ and $\hat{\theta}_\delta$ can be obtained by iterative calculations from the score equations. The Hessian $\ell_{\theta\theta}(\theta)$ then produces the informations $|\hat{\jmath}_{\theta\theta}|$ and $|\hat{\jmath}_{\lambda\lambda}(\hat{\theta}_\delta)|$ which can be recalibrated as $|\hat{\jmath}_{\varphi\varphi}|$ and $|\hat{\jmath}_{(\lambda\lambda)}(\hat{\theta}_\delta)|$ using the Jacobians $|J_4(\hat{\theta})|$ and $|J_3(\hat{\theta}_\delta)|$ obtained from $J_4(\theta) = \partial\varphi/\partial\theta'$. The scalar parameter component $\bar{\varphi} = c_p(\hat{\theta}_\psi)\varphi$ is obtained from the unit vector $c_p(\theta)$ derived from the fourth row of $J_4^{-1}(\hat{\theta}_\delta)$. We then substitute in (2.3) and (2.7) and obtain the third order significances $p_1(\delta)$, $p_2(\delta)$.

6. BEHRENS FISHER PROBLEM WITH STUDENT ERROR

Consider a first sample of $m = 4$ from $y = 5 + z$ and a second sample of $n = 7$ from $y = 10 + 2z$ where the z's come from the Student (6) distribution. The data are ...... for the first sample and are ...... for the second sample. We analyze as if all 4 parameters σ_1, σ_2, μ_1, μ_2 were unknown and calculate the significance function $p(\delta)$ for the difference $\delta = \mu_2 - \mu_1$.

First we use the appropriate analysis based on the procedure described in Section 5, assuming of course that $f_1(z)$ and $f_2(z)$ are Student (6). The significance function $p(\delta)$ is plotted as a solid curve in Figure 1.

Next for comparison purpose we use a normal analysis based on the procedure described in Sections 3 and 4. The corresponding significance function is plotted as a dotted curve in Figure 1. This can be viewed as an incorrect analysis for the data above.

We note that the normal analysis differs substantially from the correct analysis; this is consistent with the findings in Fraser (1976).

We now consider a first sample of $m = 4$ from the normal $(5, 1)$ and a second sample of $n = 7$ from the normal $(10, 2)$. The data are ...... for the first sample and are ...... for the second sample. Again we analyze as if all 4 parameters σ_1, σ_2, μ_1, μ_2 were unknown and calculate the significance function $p(\delta)$ for the difference $\delta = \mu_2 - \mu_1$.

First we do the appropriate analysis based on Section 3 and 4. The significance function is plotted as a solid curve in Figure 2. We also do the incorrect Student (6) analysis based on Section 5; the significance function is plotted as a dotted curve in Figure 2.

We note that the incorrect Student (6) Analysis is reasonable close to the correct normal analysis. Again this is consistent with the single sample findings in Fraser (1976).

7. CONCLUDING REMARKS

REFERENCES

Barndorff-Nielsen, O.E. (1980). Conditionality resolutions. *Biometrika* **67**, 293-310.

Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-365.

Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized signed likelihood ratio. *Biometrika* **70**, 343-365.

Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557-563.

Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical inference. *J.R. Statist. Soc. B* **41**, 279-312.

Behrens, W.V. (1929). Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtunger. Landw. Jb., 68, 807.

Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631-650.

Feller, W. (1971). An Introduction to Probability Theory and its Applications, Second Edition, Vol. 2; New York: Wiley.

Fisher, R.A. (1935). The feducial argument in statistical inference. *Ann. Eug. Lond.*, **6**, 391.

Fisher, R.A. and Yates, F. (1943). Statistical Tables for Agricultural, Biological and Medical Research, Second Edition, New York: Hafner.

Fraser, D.A.S. (1964). Local conditional sufficiency. *J.R. Statist. Soc. B* **26**, 52-62.

Fraser, D.A.S. (1976). Necessary analysis and adaptive inference. *J. Amer. Statist. Assoc.* **71**, 99-110.

Fraser, D.A.S. (1979). Inference and Linear Models, New York: McGraw-Hill.

Fraser, D.A.S. (1989). ¿From multiparameter likelihood to tail probability for a scalar parameter. Technical Report, University of Toronto, Department of Statistics.

Fraser, D.A.S. (1990). Tail probabilities from observed likelihoods, *Biometrika* **77**, 333-341.

Fraser, D.A.S. and Reid, N. (1993). *Statistica Sinica* **3**, 67-82.

Jeffreys, H. (1937). On the relation between direct and inverse methods in statistics. *Proc. Roy. Soc. A* **160**, 325.

Jeffreys, H. (1940). Note on the Behrens-Fisher formula. *Ann. Eug. Lond.*, **10**, 48.

Lugannani, R. and Rice, S.O. (1980). Saddlepoint approximation for the distribution of the sums of independent random variables. *Adv. Appl. Prob.* **12**, 475-490.