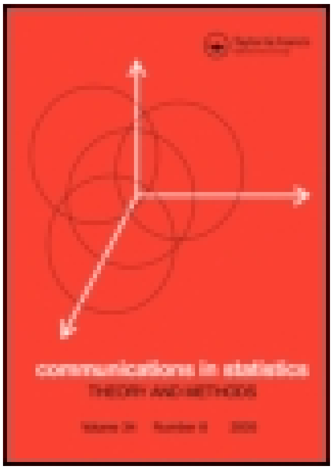

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

A solution to the multivariate behrens-fisher problem

D.G. Nel [a] & C.A. Van Der Merwe [a]
[a] Department of Statistics , University of the Orange Free State , P.O. Box 339, Bloemfontein, 9300, Republic of South Africa
Published online: 27 Jun 2007.

COMMUN. STATIST.-THEORY METH., 15(12), 3719-3735 (1986)

A SOLUTION TO THE MULTIVARIATE BEHRENS-FISHER PROBLEM

D.G. Nel C.A. van der Merwe

Department of Statistics
University of the Orange Free State
P.O. Box 339, Bloemfontein, 9300
Republic of South Africa

Key Words and Phrases: Behrens-Fisher problem, covariance matrix, degrees of freedom, elementary symmetrical functions, linear sum of Wishart matrices, patterned matrices, transition matrices, Wishart approximation.

ABSTRACT

Some new algebra on pattern and transition matrices is used to determine the degrees of freedom and the parameter matrix, if the distribution of a linear sum of Wishart matrices is approximated by a single Wishart distribution. This approximation is then used to find a solution to the multivariate Behrens-Fisher problem similar to the Welch (1947) solution in the univariate case.

1. INTRODUCTION

The classical multivariate Behrens-Fisher problem is the problem of testing whether the means of two independent multivariate normal distributions are the same, when the covariance matrices are not equal. We will not attempt to give reference to the many papers which appeared on this topic, but instead we will present a method which is in a sense similar to the method used by Welch (1947) in the univariate case.

Welch used Satterthwaite's (1946) approximation to the distribution of a linear sum of independent chi-square variates by a sing-

0361-0926/86/1512-3719$3.50/0

le σ^2-chi-square variate, to construct a variate which is approximately distributed as Student's t. Accordingly we will first discuss the "sum of Wisharts distribution" in section 3 and then approximate this distribution by a single Wishart distribution, by comparing the means and variance-covariance matrices of the variates involved. This necessitates some new results on the algebra of patterned matrices which are discussed in section 2. In section 4 we will apply our approximation to find a solution to the multivariate Behrens-Fisher problem. Some examples are discussed in section 5.

2. SOME PROPERTIES OF TRANSITION AND PATTERN MATRICES

A variable matrix $X:p\times q$ is said to be (linearly) patterned (Nel (1980), Wiens (1985) if:

(i) X has p^* functionally independent elements, and

(ii) the remaining $pq-p^*$ elements are linear combinations of the p^* functionally independent elements or are constants.

If the column vectors of X are stacked in order of appearance into a single column vector and the constant elements in X are replaced by zeroes, we will denote the resultant vector as $\text{vec } X{:}pq\times 1$. (Nel (1985)). If the functionally independent and variable elements of X are stacked columnwise in order of appearance, with a positive sign, into a single column vector, excluding the functionally dependent and constant elements, the resultant vector will be denoted as $\text{vecp } X{:}p^*\times 1$.

There exists a transition matrix $\bar{K}_{pq}{:}p^*\times pq$, which transforms vecp X onto vec X e.g.

(2.1) $\quad \bar{K}'_{pq} \text{ vecp } X=\text{vec } X$

The Moore-Penrose inverse K_{pq}, of $\bar{K}_{pq}$ now obviously transforms vec X onto vecp X e.g.:

(2.2) $\quad K'_{pq} \text{ vec } X=\text{vecp } X$

The symmetric and idempotent matrix $M_{pq}=K_{pq}\bar{K}_{pq}$, called the pattern matrix, now characterizes the pattern of the patterned matrix X in the sense that $M_{pq} \text{ vec } Y=\text{vec } Y$, if Y has the pattern of X and $M_{pq} \text{ vec } Y\neq\text{vec } Y$ if Y does not have the pattern of X.

(See Nel (1980), (1985) for more details).

If X is a pxp patterned matrix, we will denote the corresponding transition matrices and pattern matrix simply as: K_p^-, K_p and M_p. Henceforth we will assume these matrices to be the transition and pattern matrices for the symmetric, skew-symmetric or correlation patterns, unless stated otherwise. For these matrices we have that:

(2.3) $(K_p^-(A\otimes A)K_p)^{-1}=K_p^-(A^{-1}\otimes A^{-1})K_p$, and for the symmetric pattern matrix: $M_p(A\otimes A)=(A\otimes A)M_p$. (Browne (1974), Nel (1980))

Although the following results may be valid for arbitrary matrices, we need only the case where A and B are symmetric matrices which are simultaneously diagonalizable by a nonsingular matrix X. Thus:

$$X^{-1}AX=\Lambda=\mathrm{diag}(\lambda_1,\dots,\lambda_p)$$
$$X^{-1}BX=\Delta=\mathrm{diag}(\delta_1,\dots,\delta_p)\quad ,$$

where $AB=BA,\lambda_1,\dots,\lambda_p$ are the latent roots of A and $\delta_1,\dots,\delta_p$ are the latent roots of B. (Bellman (1970) p.56).

If tr_jX denotes the j-th elementary symmetric function of X then:

(2.4) $$tr_j(K_p^-(A\otimes B)K_p)=tr_j(K_p^-(\Lambda\otimes\Delta)K_p).$$

The proof is as follows:

$$tr_j(K_p^-(A\otimes B)K_p)$$
$$=tr_j(K_p^-(X\Lambda X^{-1})\otimes(X\Delta X^{-1})K_p)=tr_j(K_p^-(X\otimes X)^{-1}M_p(\Lambda\otimes\Delta)M_p(X\otimes X)K_p),$$
from (2.3)
$$=tr_j(K_p^-(X\otimes X)^{-1}K_pK_p^-(\Lambda\otimes\Delta)K_pK_p^-(X\otimes X)K_p)$$
$$=tr_j(K_p^-(\Lambda\otimes\Delta)K_p),$$ from (2.3) and since $tr_j\ XY=tr_j\ YX$.

In particular if:

$K_p^-=K_p^-(s)$ is the symmetric transition matrix, then

(2.5) $$|K_p^-(A\otimes B)K_p|=2^{-\frac{1}{2}p(p+1)}\prod_{i\leq j}(\lambda_i\delta_j+\lambda_j\delta_i)$$ (Nel (1978))

(2.6) $$|K_p^-(A\otimes A)K_p|=|A|^{p+1}$$ (Nel (1978), Henderson and Searle (1979))

and

(2.7) $$tr(K_p^-(A\otimes B)K_p)=\frac{1}{2}(trAB+trAtrB)\ .$$

We can prove (2.7) as follows:

Since $\Lambda\otimes\Delta=\mathrm{diag}(\lambda_1\delta_1,\dots,\lambda_1\delta_p,\dots,\lambda_p\delta_1,\dots,\lambda_p\delta_p)$ it follows from the definition of $K_p = K_p(s)$ that:

$$K_p'(\Lambda\otimes\Delta)K_p = \operatorname{diag}(\lambda_1\delta_1,\ \tfrac{1}{4}(\lambda_1\delta_2+\lambda_2\delta_1),\ldots,\ \tfrac{1}{4}(\lambda_1\delta_p+\lambda_p\delta_1),\ldots,\lambda_p\delta_p)$$

and thus

$$\begin{aligned} K_p^-(\Lambda\otimes\Delta)K_p &= (K_p'K_p)^{-1}K_p'(\Lambda\otimes\Delta)K_p \\ &= \operatorname{diag}(1,2,\ldots,2,\ldots,1)\ K_p'(\Lambda\otimes\Delta)K_p \\ &= \operatorname{diag}(\lambda_1\delta_1,\tfrac{1}{2}(\lambda_1\delta_2+\lambda_2\delta_1),\ldots,\tfrac{1}{2}(\lambda_1\delta_p+\lambda_p\delta_1),\ldots,\lambda_p\delta_p). \end{aligned}$$

Now $\operatorname{tr}(K_p^-(A\otimes B)K_p) = \operatorname{tr}(K_p^-(\Lambda\otimes\Delta)K_p)$

$$\begin{aligned} &= \sum_{i=1}^{p}\lambda_i\delta_i + \tfrac{1}{2}\sum_{i<j}(\lambda_i\delta_j+\lambda_j\delta_i) \\ &= \operatorname{tr}\Lambda\Delta + \tfrac{1}{2}(\operatorname{tr}\Lambda\operatorname{tr}\Delta - \operatorname{tr}\Lambda\Delta) \\ &= \tfrac{1}{2}(\operatorname{tr}AB+\operatorname{tr}A\operatorname{tr}B) \quad . \end{aligned}$$

If B=A then

$$\begin{aligned} (2.8)\qquad \operatorname{tr}(K_p^-(A\otimes A)K_p) &= \tfrac{1}{2}(\operatorname{tr}A^2+\operatorname{tr}^2A) \\ &= \operatorname{tr}^2A - \operatorname{tr}_2A. \end{aligned}$$

and if B=I then

$$\operatorname{tr}(K_p^-(A\otimes I_p)K_p) = \tfrac{1}{2}(p+1)\operatorname{tr}A. \qquad \text{(Nel (1981), Wiens (1985))}$$

If $K_p^- = K_p^-(ss)$ is the skew-symmetric transition matrix then:

$$(2.9)\qquad |K_p^-(A\otimes B)K_p| = 2^{\frac{1}{2}p(p-1)}\prod_{i<j}(\lambda_i\delta_j+\lambda_j\delta_i), \qquad \text{(Nel (1980))}$$

and

$$(2.10)\qquad \operatorname{tr}(K_p^-(A\otimes B)K_p) = \tfrac{1}{2}(\operatorname{tr}A\operatorname{tr}B - \operatorname{tr}AB) \quad .$$

The proof is similar to the proof of (2.7). Notice that now:

$$K_p^-(\Lambda\otimes\Delta)K_p = \operatorname{diag}(\tfrac{1}{2}(\lambda_1\delta_2+\lambda_2\delta_1),\ldots,\tfrac{1}{2}(\lambda_{p-1}\delta_p+\lambda_p\delta_{p-1})) \quad .$$

Thus

$$\begin{aligned} \operatorname{tr}(K_p^-(A\otimes B)K_p) &= \tfrac{1}{2}\sum_{i<j}(\lambda_i\delta_j+\lambda_j\delta_i) \\ &= \tfrac{1}{2}(\operatorname{tr}A\operatorname{tr}B - \operatorname{tr}AB) \quad . \end{aligned}$$

If B=A, then

$$\operatorname{tr}(K_p^-(A\otimes A)K_p) = \tfrac{1}{2}(\operatorname{tr}^2A - \operatorname{tr}A^2) = \operatorname{tr}_2A. \qquad \text{(Nel (1981))}$$

and if $B=I_p$, then

$$\operatorname{tr}(K_p^-(A\otimes I_p)K_p) = \tfrac{1}{2}(p-1)\operatorname{tr}A. \qquad \text{(Nel (1981))}$$

Other identities of this kind are given in Nel (1980).

If A and B are pxp matrices such that $fA=B$, then:

$$\mathrm{tr}_j(A^{-1}B)=\binom{p}{j}\frac{\mathrm{tr}_j\,B}{\mathrm{tr}_j\,A} \tag{2.11}$$

This follows by taking the j-th elementary symmetric functions of both sides in $fA=B$ and $fI_p=A^{-1}B$ and equating the values for f^j.

3. THE SUM OF WISHARTS DISTRIBUTION AND A WISHART APPROXIMATION

Suppose $U=\sum_{i=1}^{r} g_iA_i$, where $g_i>0$ and A_i:pxp distributed independently as Wishart $W_p(n_i,\Sigma_i)$. We denote the distribution of U as $U\sim \mathrm{SoW}(n_1,\ldots,n_r;g_1\Sigma_1,\ldots,g_r\Sigma_r)$, called the sum of Wisharts-distribution. Notice that since the coefficients $g_1,\ldots,g_r$ simply become part of the parameter matrices, we will, without loss of generality, consider only the sum of Wisharts variate: $U=\sum_{i=1}^{r} A_i$, where $U\sim \mathrm{SoW}(n_1,\ldots,n_r;\Sigma_1,\ldots,\Sigma_r)$.

The density function of U follows from the joint distribution of $A_1,\ldots,A_r$ by transforming to $U=\sum_{i=1}^{r} A_i$ and $T_i=U^{-\frac{1}{2}}A_iU^{-\frac{1}{2}}$,$i=1,\ldots,r-1$ with Jacobian $J(A_i\to U_i)=|U|^{-\frac{1}{2}(p+1)(r-1)}$. By transforming T_i to HT_iH', where H is an element of the group of orthogonal matrices $O(p)$, and integrating over $O(p)$ and T_i $(i=1,\ldots, r-1)$ we obtain.

(van der Merwe and Nel (1985), see also Chikuse (1980) for the non-central case)

$$f(U)=\frac{|U|^{\frac{1}{2}(N-p-1)}\mathrm{etr}(-\frac{1}{2}\Sigma_r^{-1}U)}{\Gamma_p(\frac{1}{2}N)\prod_{i=1}^{r}|2\Sigma_i|^{\frac{1}{2}n_i}}\prod_{i=1}^{r-1}\frac{(\frac{1}{2}n_i)_{\kappa_i}}{k_i!}\,. \tag{3.1}$$

$$\frac{\theta_{\phi_1}^{\kappa_1..\kappa_{r-1}}}{(\frac{1}{2}N)_{\phi_1}}\;C_{\phi_1}^{\kappa_1..\kappa_{r-1}}(-\tfrac{1}{2}\Lambda_1U,\ldots,-\tfrac{1}{2}\Lambda_rU)\quad\text{, where}$$

$$\Lambda_i=\Sigma_i^{-1}-\Sigma_r^{-1},\quad N=\sum_{i=1}^{r}N_i,$$

ϕ_1 is a partition of $f_1=\sum_{i=1}^{r-1} k_i$ into no more than p parts and

$C_{\phi}^{\kappa_1 \cdots \kappa_r}(X_1,\ldots,X_r)$ is the invariant polynomial with matrix arguments $X_1,\ldots,X_r$, as defined by Davis (1979, 1980, 1981) and Chikuse (1980).

$\sum_{\kappa_1 \ldots \kappa_{r-1};\phi}$ denotes the summation $\sum_{k_1=0}^{\infty} \cdots \sum_{k_{r-1}=0}^{\infty} \sum_{\kappa_1} \cdots \sum_{\kappa_r} \sum_{\phi \in \kappa_1 \cdot \ldots \cdot \kappa_{r-1}}$

Chikuse also derived the density of U but in the form:

$$(3.2) \qquad f(U) = \frac{|U|^{\frac{1}{2}(N-p-1)} \operatorname{etr}\left(-\frac{1}{4}\sum_{i=1}^{r}\Sigma_i^{-1}U\right)}{\Gamma_p(\frac{1}{2}N)\prod_{i=1}^{r}|2\Sigma_i|^{\frac{1}{2}n_i}} \sum_{\kappa_1,\ldots,\kappa_r;\phi} \prod_{i=1}^{r}\left[\frac{(\frac{1}{2}n_i)_{\kappa_i}}{k_i!}\right] \frac{\theta_{\phi_2}^{\kappa_1 \ldots \kappa_r}}{(\frac{1}{2}N)_{\phi_2}} C_{\phi_2}^{\kappa_1 \ldots \kappa_r}(\Delta_1 U_1,\ldots,\Delta_r U_r),$$

where $\Delta_i = -\frac{1}{4}(\Sigma_i - \sum_{\substack{j=1 \\ j\neq i}}^{r}\Sigma_j)$ and where ϕ_2 is a partition of $f_2 = \sum_{i=1}^{r} k_i$

By comparing the two density functions, we obtain the interesting identity:

$$(3.2) \qquad \operatorname{etr}(-A_r U) \sum_{\kappa_1,\ldots,\kappa_r;\phi_2} \prod_{i=1}^{r}\left[\frac{(\frac{1}{2}n_i)_{\kappa_i}}{k_i!}\right] \frac{\theta_{\phi_2}^{\kappa_1,\ldots,\kappa_r}}{(\frac{1}{2}N)_{\phi_2}} C_{\phi_2}^{\kappa_1,\ldots,\kappa_r}(A_1U,\ldots,A_rU) = \sum_{\kappa_1,\ldots,\kappa_{r-1};\phi_1} \prod_{i=1}^{r-1}\left[\frac{(\frac{1}{2}n_i)_{\kappa_i}}{k_i!}\right] \frac{\theta_{\phi_1}^{\kappa_1,\ldots,\kappa_{r-1}}}{(\frac{1}{2}N)_{\phi_1}} C_{\phi_1}^{\kappa_1,\ldots,\kappa_{r-1}}((A_1-A_r)U,\ldots,(A_{r-1}-A_r)U)$$

Applying this identity to (3.1), we obtain the density function of U in the form:

$$(3.3) \qquad f(U) = \frac{|U|^{\frac{1}{2}(N-p-1)}}{\Gamma_p(\frac{1}{2}N)\prod_{i=1}^{r}|2\Sigma_i|^{\frac{1}{2}n_i}} \sum_{\kappa_1,\ldots,\kappa_r;\phi_2} \prod_{i=1}^{r}\left[\frac{(\frac{1}{2}n_i)_{\kappa_i}}{k_i!}\right]$$

$$\frac{\theta_{\phi_2}^{\kappa_1,\ldots,\kappa_r}}{(\frac{1}{2}N)_{\phi_2}} \quad C_{\phi_2}^{\kappa_1,\ldots,\kappa_r}(-\tfrac{1}{2}\Sigma_1^{-1}U,\ldots,-\tfrac{1}{2}\Sigma_r^{-1}U),$$

which can be seen as the multivariate analogue of the sum of chi-squares density of Robbins (1946), Pitman and Robbins (1949) and Pachares (1955), which can be written as:

$$f(u)=\frac{u^{\frac{1}{2}N-1}}{\prod_{i=1}^{r}(2g_i)^{\frac{1}{2}n_i}\Gamma(\frac{1}{2}N)} \sum_{k_1=0}^{\infty}\cdots\sum_{k_r=0}^{\infty}\prod_{i=1}^{r}\left[\frac{(\frac{1}{2}n_i)_{k_i}}{(2g_i)^{k_i}k_i!}\right]. \tag{3.4}$$

$$\frac{(-u)^{f_2}}{(\frac{1}{2}N)_{f_2}}$$, where $u=g_1a_1+\ldots+g_ra_r$, $g_i>0$ and a_i distributed as chi-square with n_i degrees of freedom.

If $\Sigma_1=\Sigma_2=\ldots=\Sigma_r=\Sigma$ then $U\sim W_p(N,\Sigma)$, where $N=\sum_{i=1}^{r}n_i$ and we get a simplified version of (3.3) namely:

$$\operatorname{etr}(X)=\sum_{\kappa_1,\ldots,\kappa_r;\phi}\prod_{i=1}^{r}\left[\frac{(\frac{1}{2}n_i)_{\kappa_i}}{k_i!}\right]\frac{(\theta_{\phi_2}^{\kappa_1,\ldots,\kappa_r})^2}{(\frac{1}{2}N)_{\phi_2}}C_{\phi_2}(X). \tag{3.5}$$

This identity also follows by equating coefficients of $C_\phi(X)$ on both sides of $\prod_{i=1}^{r}|I-X|^{-\frac{1}{2}n_i}=|I-X|^{-\frac{1}{2}N}$ and using the relation

$$\sum_{k_1}\cdots\sum_{k_r}\sum_{\phi_2\in\kappa_1\cdots\kappa_r}\binom{f_2}{k_1\ldots k_r}\left(\theta_{\phi_2}^{\kappa_1,\ldots,\kappa_r}\right)^2\prod_{i=1}^{r}(\tfrac{1}{2}n_i)_{\kappa_i}=(\tfrac{1}{2}N)_{\phi_2}$$

(Davis (1985)).

Satterthwaite (1946) approximated the sum of chi-squares distribution by a σ^2-chi-square distribution:

$$z\sim\sigma^2\chi_f^2$$

where

$$\sigma^2=\frac{\sum_{i=1}^{r}n_ig_i^2}{\sum_{i=1}^{r}n_ig_i} \quad\text{and}\quad f=\frac{(\sum_{i=1}^{r}n_ig_i)^2}{\sum_{i=1}^{r}n_ig_i^2} \tag{3.6}$$

We will use a similar approach to approximate the distribution of $U \sim SoW(n_1,\dots,n_r;\Sigma_1,\dots,\Sigma_r)$ by $Z \sim W_p(f,\Sigma)$ (See also Tan and Gupta (1983)).

Now $E(Z)=f\Sigma$ and $E(U)=\sum_{i=1}^{r} n_i\Sigma_i$. By equating these expected values we get:

$$\Sigma = \frac{1}{f}\sum_{i=1}^{r} n_i\Sigma_i \tag{3.7}$$

The variance-covariance matrix of vecp Z is given by:

$$\mathrm{Var}(\mathrm{vecp}\ Z)=2f\ K_p'(\Sigma\otimes\Sigma)K_p \quad \text{(Browne (1974), Nel(1978))} \tag{3.8}$$

where K_p denotes the symmetric transition matrix. This matrix is also reported in the literature in the form:

$$\mathrm{Var}(\mathrm{vec}\ Z)=2f\ M_p(\Sigma\otimes\Sigma)M_p=2f\ M_p(\Sigma\otimes\Sigma)$$

where $M_p=\frac{1}{2}(I_{p^2}+I_{(p,p)})$ is the symmetric pattern matrix and

$I_{(p,p)}$ the $p^2 x p^2$ permuted identity or permutation matrix. (Nel (1980))

The variance-covariance matrix of vecp U follows as:

$$\mathrm{Var}(\mathrm{vecp}\ U = 2K_p'(\sum_{i=1}^{r} n_i(\Sigma_i\otimes\Sigma_i))K_p \tag{3.9}$$

Equating these matrices, using (3.7) and premultiplying by $(K_p'K_p)^{-1}$ gives:

$$K_p^-[(\sum_{i=1}^{r} n_i\Sigma_i)\otimes(\sum_{i=1}^{r} n_i\Sigma_i)]K_p=f.K_p^-[\sum_{i=1}^{r} n_i(\Sigma_i\otimes\Sigma_i)]K_p \tag{3.10}$$

which is of the form B=fA. Solving for f^j by applying the j-th elementary symmetric function to both sides and using (2.11) we get for $j=1,2,3,\dots,p^*=\frac{1}{2}p(p+1)$ that:

$$f \equiv g_j = \left[\frac{tr_j(K_p^-[(\sum_{i=1}^{r} n_i\Sigma_i)\otimes(\sum_{i=1}^{r} n_i\Sigma_i)]K_p)}{tr_j(K_p^-[\sum_{i=1}^{r} n_i(\Sigma_i\otimes\Sigma_i)]K_p)}\right]^{\frac{1}{j}} \tag{3.11}$$

A better approach is to apply the j-th elementary symmetric function to both sides of $fI_{p^*}=BA^{-1}$, which yields

$$f \equiv h_j = \left[\frac{tr_j(K_p^-[(\sum_{i=1}^{r} n_i\Sigma_i)\otimes(\sum_{i=1}^{r} n_i\Sigma_i)][\sum_{i=1}^{r} n_i(\Sigma_i\otimes\Sigma_i)]^{-1}K_p)}{\binom{p^*}{j}}\right]^{\frac{1}{j}} \tag{3.12}$$

$$=\left[\frac{tr_j(A^{-1}B)}{\binom{p^*}{j}}\right]^{\frac{1}{j}}$$

The latter approach has the advantage of an ordering relationship among the h_j's e.g.

$$h_1 \geq h_2 \geq \ldots\ldots\ldots\ldots \geq h_{p^*} \quad \text{(Archbold (1958),p.55)}$$

The smallest possible value for f, by using this procedure and (2.7), is given by:

$$\text{(3.13)} \qquad f=h_{p^*}=g_{p^*}=\left[\frac{\left|\sum_{i=1}^{r} n_i\Sigma_i\right|^{p+1}}{\left|K_p^-\left[\sum_{i=1}^{r} n_i(\Sigma_i\otimes\Sigma_i)\right]K_p\right|}\right]^{\frac{2}{p(p+1)}}$$

We suggest the use of this value for f as it is the most conservative among the h_j's. Tan and Gupta (1983) conjectured that f is given by an expression equivalent to this value. Thus the minimum value of the h_j's proves their conjecture.

The easiest value for f to compute is:

$$\text{(3.14)} \qquad f=g_1=\frac{tr(\sum_{i=1}^{r} n_i\Sigma_i)^2+tr^2(\sum_{i=1}^{r} n_i\Sigma_i)}{\sum_{i=1}^{r} n_i(tr\Sigma_i^2+tr^2\Sigma_i)}$$

which follows from (2.8). If p=1 then (3.11), (3.12), (3.13) and (3.14) reduce to Satterthwaite's value for f as given in (3.6). A measure of the goodness of the approximation is the matrix BA^{-1}. The closer this matrix is to a diagonal matrix of the form $fI_p{}^*$, the better is the approximation. Notice that the results (3.11) and (3.12) are equivalent for all $j=1,\ldots,p^*$ only if $B=fA$.

We will now apply these results to find a solution to the multivariate Behrens-Fisher problem which is equivalent to the Welch solution to the univariate problem.

4. A SOLUTION TO THE MULTIVARIATE BEHRENS-FISHER PROBLEM

Suppose $\bar{\underline{X}}_1$ and $\bar{\underline{X}}_2$ are the sample mean vectors and S_1 and S_2 the sample covariance matrices of random samples of sizes N_1 and N_2 respectively from two independent multivariate normal distributions $N_p(\underline{\mu}_1,\Sigma_1)$ and $N_p(\underline{\mu}_2,\Sigma_2)$, where $\Sigma_1 \neq \Sigma_2$. Then

(4.1) $$\underline{X} = \bar{\underline{X}}_1 - \bar{\underline{X}}_2 - (\underline{\mu}_1 - \underline{\mu}_2) \sim N(\underline{0}, \frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2)$$

independently of $U = \frac{1}{N_1}S_1 + \frac{1}{N_2}S_2$. The latter matrix U has a sum of Wisharts distribution:

$$U \sim \text{SoW}(N_1-1,\ N_2-1, g_1\Sigma_1, g_2\Sigma_2)$$

where

$$g_i = \frac{1}{N_i(N_i-1)}\ , i=1,2.$$

Using the approximation in section 3, we have that:

(4.2) $$U \sim W_p(f,\Sigma)\ ,\ \text{where } \Sigma = \frac{1}{f}(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2)$$

(4.3) $$f \equiv h_j = \left[\frac{tr_j(E^{-1}G)}{\binom{p^*}{j}}\right]^{\frac{1}{j}},\quad h_1 \geq h_2 \geq \ldots \geq h_{p^*}$$

or

(4.4) $$f \equiv g_j = \left[\frac{tr_j G}{tr_j E}\right]^{\frac{1}{j}},$$

where

(4.5) $$E = K_p^-[\frac{1}{N_1-1}\ (\frac{1}{N_1}\Sigma_1 \otimes \frac{1}{N_1}\Sigma_1) + \frac{1}{N_2-1}\ (\frac{1}{N_2}\Sigma_2 \otimes \frac{1}{N_2}\Sigma_2)]K_p\ ,$$

(4.6) $$G = K_p^-[(\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2) \otimes (\frac{1}{N_1}\Sigma_1 + \frac{1}{N_2}\Sigma_2)]K_p\ .$$

In practise we will replace Σ_i by S_i in (4.5) and (4.6). The easiest formula to use is:

(4.7) $$f = g_1 = \frac{tr(\frac{1}{N_1}S_1 + \frac{1}{N_2}S_2)^2 + tr^2(\frac{1}{N_1}S_1 + \frac{1}{N_2}S_2)}{\frac{1}{N_1-1}\ [tr(\frac{1}{N_1}S_1)^2 + tr^2(\frac{1}{N_1}S_1)] + \frac{1}{N_2-1}[tr(\frac{1}{N_2}S_2)^2 + tr^2(\frac{1}{N_2}S_2)]}$$

If p=1 these formulae (4.3), (4.4) and (4.7) reduce to the wellknown formula for the degrees of freedom as given by Welch (1947):

$$f = \frac{\left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}\right)^2}{\frac{\left(\frac{S_1^2}{N_1}\right)}{N_1 - 1} + \frac{\left(\frac{S_2^2}{N_2}\right)}{N_2 - 1}}$$

The approximation of Hotellings T^2 statistic is now:

$$T^2 = (\bar{\underline{X}}_1 - \bar{\underline{X}}_2 - (\underline{\mu}_1 - \underline{\mu}_2))' \left(\frac{1}{N_1} S_1 + \frac{1}{N_2} S_2\right)^{-1} (\bar{\underline{X}}_1 - \bar{\underline{X}}_2 - (\underline{\mu}_1 - \underline{\mu}_2))$$

If H_0: $\underline{\mu}_1 = \underline{\mu}_2$ is true then approximately:

$$F = \frac{T^2}{f} \cdot \frac{f-p+1}{p}$$

has an F distribution with p and f-p+1 degrees of freedom. We will reject H_0 against $H_1 : \underline{\mu}_1 \neq \underline{\mu}_2$ if $F > F_{p,f-p+1;1-\alpha}$, where $F_{p,f-p+1;1-\alpha}$ is the $100(1-\alpha)$th percentile of the F-distribution with p and f-p+1 degrees of freedom.

The following example illustrates the method:

Example 4.1

Consider the electrical consumption problem given on pages 243 and 246 of Johnson and Wichern (1982), where X_1= electrical consumption during peak hours, July 1977. X_2= electrical comsumption during off-peak hours, July 1977, both measured in kilowatt hours.

The two populations in question are:

I : people with and II : people without airconditioning. The following results were obtained:

$$\text{I: } \bar{\underline{X}}_1 = \begin{bmatrix} 204.4 \\ 556.6 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 13825.3 & 23823.4 \\ 23823.4 & 73107.4 \end{bmatrix}, \quad N_1 = 45$$

$$\text{II: } \bar{\underline{X}}_2 = \begin{bmatrix} 130.0 \\ 355.0 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 8632.0 & 19616.7 \\ 19616.7 & 55964.5 \end{bmatrix}, \quad N_2 = 55$$

Now:

$$K_2=\begin{bmatrix}1 & 0 & 0\\ 0 & \tfrac{1}{2} & 0\\ 0 & \tfrac{1}{2} & 0\\ 0 & 0 & 1\end{bmatrix}, \quad K_2^-=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \text{ and consequently}$$

$$E=\begin{bmatrix}2601.364 & 4709.157 & 8617.613\\ 9466.401 & 22972,141 & 52224.434\\ 8725.627 & 26268.102 & 79159.024\end{bmatrix}$$

$$G=\begin{bmatrix}215457.813 & 407453.816 & 770538.844\\ 822587.609 & 2004216.656 & 4638568.188\\ 785130.938 & 2341141.844 & 6980931.250\end{bmatrix} \quad \text{and}$$

$$E^{-1}G=\begin{bmatrix}69.213 & 4.371 & -0.558\\ 8.480 & 78.599 & 6.209\\ -0.525 & 3.011 & 86.190\end{bmatrix}.$$

Consequently:

$h_1=78.0004$, $h_2=77.726$, $h_3=77.4362$; $g_1=87.8486$, $g_2=83.4628$ and $g_3=77.4362$.

Now $T^2=15.66$, conservatively we choose $f=h_3=g_3=77.4362$ and consequently $F=7.7288$. The corresponding critical value at 5% significance level is:

$$F_{2,77.44,0.95}=3.1161 \text{ and we will reject } H_0.$$

If we choose the largest value of f namely $f=g_1=87.8486$, then $F=7.7408$ and $F_{2,87.8486,0.95}=3.103$. It is interesting to note that these critical values differ very slightly and that the particular choice of f among the candidates seems not to be too critical. Note that $E^{-1}G$ is relatively close to a diagonal matrix of the form fI_{n*}, where f can be approximated by the h_j's or g_j's.

Example 4.2

It is interesting to compare the different values of f which are obtained for different values of N_1, N_2, S_1 and S_2.

For p=2 we choose S_1 equal to $\begin{bmatrix} 25 & 10 \\ 10 & 53 \end{bmatrix}$ and then vary S_2, N_1 and N_2. Since $g_3=h_3$, while h_2 falls between h_1 and h_3, we do not report them.

Case	S_2	N_1	N_2	g_1	g_2	h_3	h_1
1	$\begin{bmatrix} 25 & 10 \\ 10 & 53 \end{bmatrix}$	20	30	40.875	40.9007	40.875	40.875
2	$\begin{bmatrix} 49 & 21 \\ 21 & 45 \end{bmatrix}$	20	30	42.78	42.832	43.09	43.39
3	$\begin{bmatrix} 49 & -21 \\ -21 & 45 \end{bmatrix}$	20	30	41.24	41.996	42.68	43.70
4	$\begin{bmatrix} 100 & -10 \\ -10 & 82 \end{bmatrix}$	20	30	46.74	46.789	46.81	46.96
5	$\begin{bmatrix} 169 & -26 \\ -26 & 200 \end{bmatrix}$	20	30	43.26	43.183	43.13	43.30
6.	$\begin{bmatrix} 10000 & -4900 \\ -4900 & 252401 \end{bmatrix}$	20	30	29.02	29.62	29.12	29.12
	$\begin{bmatrix} 10000 & -4900 \\ -4900 & 252401 \end{bmatrix}$	45	5	4.0002	4.006	4.001	4.001
7.	$\begin{bmatrix} 10000 & -4900 \\ -4900 & 252401 \end{bmatrix}$	30	20	19.006	19.182	19.0351	19.0352

For p=3 we choose S_1 equal to $\begin{bmatrix} 25 & -10 & 15 \\ -10 & 53 & 22 \\ 15 & 22 & 61 \end{bmatrix}$ and again vary S_2, N_1 and N_2.

Case	S_2	N_1	N_2	g_1	h_6	h_1
1	$\begin{bmatrix} 25 & -10 & 15 \\ -10 & 53 & 22 \\ 15 & 22 & 61 \end{bmatrix}$	20	30	40.875	40.875	40.875
2	$\begin{bmatrix} 49 & 21 & 14 \\ 21 & 45 & -12 \\ 14 & -12 & 38 \end{bmatrix}$	20	30	38.08	44.83	42.33
3	$\begin{bmatrix} 49 & -21 & -14 \\ -21 & 45 & 24 \\ -14 & 24 & 38 \end{bmatrix}$	20	30	39.14	42.37	41.20
4	$\begin{bmatrix} 100 & -10 & -30 \\ -10 & 82 & 21 \\ -30 & 21 & 62 \end{bmatrix}$	20	30	45.33	47.43	46.30
5	$\begin{bmatrix} 169 & -26 & 39 \\ -26 & 200 & 64 \\ 39 & 64 & 178 \end{bmatrix}$	20	30	45.28	44.14	44.00
6	$\begin{bmatrix} 10000 & -4900 & 3600 \\ -4900 & 252401 & -7764 \\ 3600 & -7764 & 41440 \end{bmatrix}$	20	30	29.03	29.12	29.12
7	$\begin{bmatrix} 10000 & -4900 & 3600 \\ -4900 & 252401 & -7764 \\ 3600 & -7764 & 41440 \end{bmatrix}$	45	5	4.002	4.001	4.001
8	$\begin{bmatrix} 10000 & -4900 & 3600 \\ -4900 & 252401 & -7764 \\ 3600 & -7764 & 41440 \end{bmatrix}$	30	20	19.006	19.035	19.035

When the elements of S_2 are much greater in magnitude than those of S_1, the degrees of freedom approach N_2-1. This also follows from (4.7) when S_1 is small in relation to S_2. We also note that the values of h_j and g_j are remarkably close for each case observed. If $S_1=S_2$, then of course $E^{-1}G$ is exactly equal to $f\,I_p^*$, as in case 1, where

$$f=\frac{\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}{\left(\frac{1}{(N_1-1)N_1^2}+\frac{1}{(N_2-1)N_2^2}\right)}=40.875.$$

It should be noted that when $\Sigma_1=\Sigma_2$ in (4.3), (4.4) and (4.7), we obtain the above expression for f. This expression, like Welch's expression for $\sigma_1^2=\sigma_2^2$, does not reduce to the degrees of freedom for the pooled variance case namely:

N_1+N_2-2. Only when $\Sigma_1=\Sigma_2$ and $N_1=N_2(=N)$, the degrees of freedom is $f=2N-2$.

The following table (Table 4.1) gives the values of the minimum h_j, namely h_6, for different N_1, N_2 and S_2. In the third column are the values of h_6 for $S_2=S_1$ and in the other columns are the values of h_6 where only the 1,1-element of S_2 differ from S_1.

Table 4.1

Values of h_6 for different matrices S_2

N_1	N_2	$S_2=S_1$	(1,1)=45	(1,1)=55	(1,1)=65	(1,1)=75	(1,1)=85	(1,1)=95	(1,1)=105	(1,1)=115
5	45	4.93	5.13	5.33	5.70	5.88	6.06	6.24	6.41	6.57
10	40	13.86	14.89	15.84	17.51	18.24	18.91	19.53	20.10	20.62
15	35	26.56	28.83	30.61	33.15	34.06	34.81	35.44	35.97	36.43
20	30	40.875	43.07	44.24	45.28	45.53	45.71	45.83	45.93	46.00
25	25	48.00	47.56	46.91	45.90	45.54	45.25	45.01	44.82	44.65
30	20	40.875	38.65	37.38	36.00	35.59	35.29	35.04	34.85	34.69
35	15	26.56	24.88	24.07	23.27	23.04	22.87	22.74	22.64	22.56
40	10	13.86	13.19	12.88	12.58	12.50	12.43	12.39	12.35	12.32
45	5	4.93	4.81	4.75	4.70	4.68	4.67	4.66	4.657	4.65

If $S_2=S_1$ the values of $f=h_6$ are symmetrical around the point $N_1=N_2=25$. The greater the difference between the 1,1-element of S_2 from the corresponding element in S_1, the less symmetrical the values of f are. Note that the lower values in the table are slowly decreasing towards the value of N_2-1 and the upper values are slowly increasing towards N_2-1 as the 1,1-element of S_2 increases. Similar situations can be considered for differences between the other diagonal elements and also when more elements of S_2 differ from those of S_1- the typical situation for which this theory is applicable.

Acknowledgement:

The authors are grateful to Drs. Y. Chikuse and A.W. Davis for comments regarding the use of the invariant polynomials. We also wish to express our gratitude to the CSIR for grants which made this research possible.

REFERENCES

ANDERSON, T.W. (1958) An introduction to Multivariate Statististical Analysis., John Wiley, New York.

ARCHBOLD, J.W. (1958) Algebra, Sir Isaac Pitman and Sons, Ltd., London.

BELLMAN, R. (1970 Introduction to matrix analysis (See Ed). McGraw Hill, New York.

BROWNE, M.W. (1974) Generalized least squares estimators in the analysis of covariance structures. S. Afr. Statist. J. 8, 1-24.

CHIKUSE, Y. (1980) Invariant polynomials with matrix arguments and their applications. Multivariate Statistical Analysis (V). 53-67.

DAVIS, A.W. (1979) Invariant polynomials with two matrix arguments extending the zonal polynomials: Applications to multivariate distribution theory. Ann. Inst. Statist. Math., Vol. 31, Part A, 465-485.

DAVIS, A.W. (1980) Invariant polynomials with two matrix arguments extending the zonal polynomials. Multivariate Analysis (V). Ed. P.R. Krishnaiah, 287-299.

DAVIS, A.W. (1981) On the construction of a class of invariant polynomials in several matrices, extending the zonal polynomials. Ann. Inst. Statist. Math., 33, A, 297-313.

DAVIS, A.W. (1985) Personal communication.

JOHNSON, R.A. and WICHERN, D.W. (1982) Applied multivariate statistical analysis. Prentice Hall, Englewood Cliffs, N.J.

NEL, D.G. (1978) On the symmetric multivariate normal distribution and the asymptotic expansion of a Wishart matrix. S.A. Statist. J., 12, 145-159.

NEL, D.G. (1980) On matrix differentiation in statistics. S. Afr. Statist. J., 14, 137-193.

NEL, D.G. (1981) On patterned matrices and their applications in multivariate statistics. Tech. Report. No. 67, Dept. of Math. Statist., University of the O.F.S.

NEL, D.G. (1985) A matrix derivation of the asymptotic covariance matrix of sample correlation coefficients. Linear Algebra and its Applications, 67, 137-145.

PACHARES, J. (1955) Note on the distribution of a definite quadratic form. Ann. Math. Statist., 26, 128-131.

ROBBINS, H. (1948) The distribution of a definite quadratic form. Ann. Math. Statist., 19, 266-270.

SATTERTHWAITE, F.E. (1946) An aaproximate distribution of estimates of variance components. Biometrics, 2, 110-114.

TAN, W.Y. and GUPTA, R.P. (1983) On approximating a linear combination of central Wishart matrices with positive coefficients. Comm. Statist. Theor. Meth.,12 (22), 2589-2600.

VAN DER MERWE, C.A. and NEL, D.G. (1985) On the distribution of a linear sum of Wishart matrices and related matric-t variates. Tech. Report No. 107, Dept. of Math. Statist., University of the O.F.S.

WELCH, B.L. (1947) The generalization of Student's problem when several different population variances are involved. Biometrika, 34, 28-35.

WIENS, D.P. (1985) On some Pattern-Reduction matrices which appear in Statistics. Linear Algebra and its Applications, 67, 233-258.

Received by Editorial Office January, 1986; Revised August, 1986.

Recommended Anonymously.

Refereed by George Yeh, Texas-New Mexico Power and Light Co. Fort Worth, TX, and Anonymously.