A measure of independence for a multivariate normal distribution and some connections with factor analysis

Martin Knott
The London School of Economics and Political Science

November 13, 2001

Abstract

This paper gives results for the population value of a measure of the goodness-of-fit of a general multivariate normal distribution to the simpler hypothesis of independent normal variables. The measure was introduced by Rudas, Clogg and Lindsay in 1994, who gave the value for the bivariate normal distribution. Connections with factor analysis are briefly discussed.

1 Introduction

Rudas, Clogg, and Lindsay (1994) introduced a new index of fit for contingency table models. Their idea was to measure the population goodness-of-fit of a model by writing the population under investigation as a mixture of the population under the model and an arbitrary population. The index of fit that they proposed is the largest mixing probability that can be given to the model so that the mixture representation is feasible. Their measure is the maximum proportion of the population under investigation that can be thought of as coming from the population under the assumptions of the model. It lies between 0 and 1, and is close to 1 for a model that fits well. To use the index in practice one must estimate it from a sample, but this paper concentrates on the population quantities.

This attractive index of fit (which is applicable much more widely than to models for contingency tables) has been further investigated in Clogg, Rudas, and Xi (1995), Rudas (1999).

In order to check if the measure is as attractive as it seems, it is useful to have its value worked out in a few simple cases. Rudas, Clogg and Lindsay gave the value of the index of fit, say ζ, for a general bivariate normal distribution with correlation coefficient ρ to the simpler model of independent normal distributions.

$$\zeta = \left[\frac{1-|\rho|}{1+|\rho|}\right]^{\frac{1}{2}}. \tag{1}$$

This is a very appealing measure of independence for a bivariate normal distribution, but it would be good to see the index of fit to independence for a general q-dimensional multivariate normal distribution. In Section 2 the problem of finding the index of fit for a general multivariate normal distribution is reduced to a one of maximising a concave function over a convex set. A dual problem is given and also extremality relations that characterise the solution. Section 3 applies the approach to obtain explicit solutions for equicorrelated normal variables, and gives an example of the use of a simple algorithm to obtain the index of fit for a general multivariate normal distribution. Section 4 contains a discussion and interpretation of the results in the context of factor analysis, in particular the method of minimum trace estimation of the linear factor model.

2 Optimisation Results

The multivariate normal distribution will be assumed to be non-degenerate. If it has a singular covariance matrix, then one could reduce the dimension by a rotation and continue. The arguments of Rudas, Clogg, and Lindsay (1994), see equation (12) in their Appendix A, show that ζ is the largest number such that for some choice of the univariate normal density functions $g_{X_i}(x_i)$ for $i = 1, \ldots, p$, and for all $\mathbf{x}$,

$$f_{\mathbf{X}}(\mathbf{x}) \geq \zeta \prod_{i=1}^{q} g_{X_i}(x_i), \tag{2}$$

where $f_{\mathbf{X}}(\mathbf{x})$ is the density function of the multivariate normal distribution of interest for the random variables $\mathbf{X} = (X_1, \ldots, X_q)$ taking values $\mathbf{x} = (x_1, \ldots, x_q)$.

It has been assumed implicitly that for the right of (2) no degenerate distributions are allowed for $\mathbf{X}$, which is necessary to include the joint distribution of $\mathbf{X}$ under independence in the class of distributions allowed for the left of (2).

One may therefore, without loss of generality, by making changes in location and scale on both sides of (2), take the random variables $\mathbf{X}$ on the left of (2) to have mean 0 and variance 1. Their correlation matrix will be written as $\mathbf{C}$. It is not clear at this stage what choice of the means, say $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_q)$, and the diagonal matrix of variances, say $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{qq})$, of the normal distributions on the right of (2) will allow ζ to be attained.

Inequality (2) may be written

$$-\frac{1}{2}\ln\det\mathbf{C} - \frac{1}{2}\mathbf{x}'\mathbf{C}^{-1}\mathbf{x} \geq \ln\zeta - \frac{1}{2}\ln\det\boldsymbol{\Sigma} - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}), \qquad (3)$$

which simplifies to

$$\frac{1}{2}\ln\frac{\det\boldsymbol{\Sigma}}{\zeta\det\mathbf{C}} \geq \frac{1}{2}\mathbf{x}'(\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1})\mathbf{x} + \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \qquad (4)$$

From (4) following the same argument as in Rudas, Clogg, and Lindsay (1994), we obtain ζ only if $\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1}$ is negative semidefinite.

Let us now decide what value for $\boldsymbol{\mu}$ gives the largest ζ. If $\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1}$ is singular, then for every $\mathbf{x}$ in the kernel space of $\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1}$ it must be required that $\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\mathbf{x} = 0$, otherwise by scaling up either $\mathbf{x}$ or $-\mathbf{x}$ by a large enough factor, ζ would be forced to be indefinitely small. So we can assume that $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ is in the image space of $\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1}$, and define $\mathbf{a}$ as an inverse image of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, so that

$$(\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1})\mathbf{a} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \qquad (5)$$

Then (4) becomes

$$\frac{1}{2}\ln\frac{\det\boldsymbol{\Sigma}}{\zeta\det\mathbf{C}} \geq \frac{1}{2}(\mathbf{x}-\mathbf{a})'(\mathbf{C}^{-1}-\boldsymbol{\Sigma}^{-1})(\mathbf{x}-\mathbf{a}) - \frac{1}{2}\mathbf{a}'(\mathbf{C}^{-1}-\boldsymbol{\Sigma}^{-1})\mathbf{a} - \frac{1}{2}\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \qquad (6)$$

The right side of (6) can be written

$$\frac{1}{2}(\mathbf{x}-\mathbf{a})'(\mathbf{C}^{-1}-\boldsymbol{\Sigma}^{-1})(\mathbf{x}-\mathbf{a}) - \frac{1}{2}\mathbf{a}'(\mathbf{C}^{-1}-\boldsymbol{\Sigma}^{-1})\mathbf{a} - \frac{1}{2}\mathbf{a}'(\mathbf{C}^{-1}-\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}(\mathbf{C}^{-1}-\boldsymbol{\Sigma}^{-1})\mathbf{a}$$

which simplifies to

$$\frac{1}{2}(\mathbf{x}-\mathbf{a})'(\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1})(\mathbf{x}-\mathbf{a}) + \mathbf{a}'(\mathbf{C}^{-1} + \mathbf{C}^{-1}\boldsymbol{\Sigma}\mathbf{C}^{-1})\mathbf{a}.$$

Since $\mathbf{a}'(\mathbf{C}^{-1} + \mathbf{C}^{-1}\boldsymbol{\Sigma}\mathbf{C}^{-1})\mathbf{a}$ is non-negative for all $\mathbf{a}$, the choice $\mathbf{a} = \mathbf{0}$ allows the largest ζ. Putting $\mathbf{a} = \mathbf{0}$ implies choosing $\boldsymbol{\mu} = \mathbf{0}$, as can be seen from (5). Note that one is free to choose $\mathbf{a} = \mathbf{0}$ without consideration of $(\mathbf{x}-\mathbf{a})'(\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1})(\mathbf{x}-\mathbf{a})$ because the inequality (6) must be true for all choices of $\mathbf{x}$, which is the same as for all choices of $\mathbf{x} - \mathbf{a}$.

Returning to (4), but taking $\boldsymbol{\mu} = \mathbf{0}$,

$$\frac{1}{2}\ln\frac{\det\boldsymbol{\Sigma}}{\zeta\det\mathbf{C}} \geq \frac{1}{2}\mathbf{x}'(\mathbf{C}^{-1} - \boldsymbol{\Sigma}^{-1})\mathbf{x}. \qquad (7)$$

Since $\mathbf{C}^{-1} - \mathbf{\Sigma}^{-1}$ is negative semidefinite, the inequality is at its most stringent when $\mathbf{x} = \mathbf{0}$. So the index ζ satisfies

$$\frac{1}{2} \ln \frac{\det \mathbf{\Sigma}}{\zeta \det \mathbf{C}} \geq 0, \tag{8}$$

for all $\mathbf{\Sigma}$ for which $\mathbf{C}^{-1} - \mathbf{\Sigma}^{-1}$ is negative semidefinite. That is, ζ is the largest value of

$$\left[\frac{\det \mathbf{\Sigma}}{\det C}\right]^{\frac{1}{2}} \tag{9}$$

for which $\mathbf{\Sigma} = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{qq})$ satisfies the condition that $\mathbf{C}^{-1} - \mathbf{\Sigma}^{-1}$ is negative semidefinite.

To find ζ one needs to maximise $\det \mathbf{\Sigma}$ subject to $\mathbf{C}^{-1} \leq \mathbf{\Sigma}^{-1}$ (using standard conventions for writing the Loewner partial ordering of matrices), see Halmos (1958) page 140. The final stage of the proof will proceed by reformulating the last statement of the problem, and then using a procedure based on convex optimisation methods.

It is true that $\mathbf{C}^{-1} \leq \mathbf{\Sigma}^{-1}$ if and only if $\mathbf{\Sigma} \leq \mathbf{C}$ (see Halmos (1958) page 168 Exercise 13). So the problem can be reformulated as that of maximising $\ln \det \mathbf{\Sigma}$ subject to $\mathbf{\Sigma} \leq \mathbf{C}$ and $\mathbf{\Sigma}$ diagonal. More precisely, we want to find

$$\sup \ln \det[\mathbf{\Sigma}] \tag{10}$$

over all diagonal positive semidefinite matrices $\mathbf{\Sigma}$ satisfying

$$\mathbf{\Sigma} \leq \mathbf{C}. \tag{11}$$

This form of the problem is seen as the maximisation of a concave function of the matrix $\mathbf{\Sigma}$ over a convex set of such matrices. It is easy to see that $\ln \det \mathbf{\Sigma}$ is concave and strictly decreasing on the set of positive semidefinite diagonal matrices $\mathbf{\Sigma}$. It is obvious that the constraint $\mathbf{\Sigma} \leq \mathbf{C}$ restricts $\mathbf{\Sigma}$ to a convex subset of all matrices $\mathbf{\Sigma}$.

So, all the well-known results from convex optimisation can be applied. When considering a dual problem, one needs to use the norm $\mathrm{trace}(\mathbf{A}'\mathbf{A})$ for the matrix $\mathbf{A}$, and the corresponding inner product. Results from, for instance, Ekeland and Temam (1976) pages 62 to 68 can be used to show that the problem in (10) and (11) has a solution, and that the solution is the same as that for the dual problem of finding

$$\inf[-\ln \det \mathrm{diag}\, \mathbf{D} + \mathrm{trace}\, \mathbf{CD} - q] \tag{12}$$

over all positive semidefinite matrices $\mathbf{D}$.

The dual problem is easier because there is no restriction on $\mathbf{D}$ except for positive semidefiniteness. Rather than justify in detail the application of the general convex optimisation theory, it seems better to prove directly Theorem 1. The notation $\mathrm{diag}\, \mathbf{D}$ means the diagonal matrix with diagonal elements equal to those of $\mathbf{D}$. To avoid trivialities any matrix inverse will be interpreted as a generalised inverse where necessary.

Theorem 1 *If the positive definite matrix $\hat{\mathbf{\Sigma}}$ and the positive semidefinite matrix $\hat{\mathbf{D}}$ are such that*

$$\hat{\mathbf{\Sigma}} = [\mathrm{diag}\, \hat{\mathbf{D}}]^{-1},$$

$$\hat{\mathbf{\Sigma}} \leq \mathbf{C}$$

and satisfy the extremality relation

$$\operatorname{trace}[(\hat{\mathbf{\Sigma}} - \mathbf{C})\hat{\mathbf{D}}] = 0$$

then $\hat{\mathbf{\Sigma}}$ *gives a solution to (10), (11) and* $\hat{\mathbf{D}}$ *gives a solution to (12), and the solutions are equal.*

The proof of Theorem 1 is as follows: From the extremality relation,

$$\begin{aligned}\ln\det\hat{\mathbf{\Sigma}} &= -\operatorname{trace}[(\hat{\mathbf{\Sigma}} - \mathbf{C})\hat{\mathbf{D}}] + \ln\det\hat{\mathbf{\Sigma}} \\ &\geq -\operatorname{trace}[(\mathbf{\Sigma} - \mathbf{C})\hat{\mathbf{D}}] + \ln\det\mathbf{\Sigma},\end{aligned}$$

for every diagonal positive definite matrix $\mathbf{\Sigma}$. The last inequality comes from considering maximum likelihood estimation of the variances $\mathbf{\Sigma}^{-1}$ for independent normally distributed random variables with means zero and sample covariance matrix $\hat{\mathbf{D}}$. It follows that for all positive definite diagonal $\mathbf{\Sigma}$ with $\mathbf{\Sigma} \leq \mathbf{C}$,

$$\ln\det\hat{\mathbf{\Sigma}} \geq \ln\det\mathbf{\Sigma}.$$

This is enough to prove that $\hat{\mathbf{\Sigma}}$ gives a solution to (10), (11).

On the other hand, for the dual problem,

$$\begin{aligned}-\ln\det\operatorname{diag}\hat{\mathbf{D}} + \operatorname{trace}\mathbf{C}\hat{\mathbf{D}} - q &= -\ln\det\operatorname{diag}\hat{\mathbf{D}} - \operatorname{trace}[((\operatorname{diag}\hat{\mathbf{D}})^{-1} - \mathbf{C})\hat{\mathbf{D}}] \\ &= -\ln\det\operatorname{diag}\hat{\mathbf{D}} \\ &\leq -\ln\det\operatorname{diag}\hat{\mathbf{D}} - \operatorname{trace}[((\operatorname{diag}\hat{\mathbf{D}})^{-1} - \mathbf{C})\mathbf{D}] \\ &\leq -\ln\det\operatorname{diag}\mathbf{D} - \operatorname{trace}[((\operatorname{diag}\mathbf{D})^{-1} - \mathbf{C})\mathbf{D}] \\ &= -\ln\det\operatorname{diag}\mathbf{D} + \operatorname{trace}\mathbf{C}\mathbf{D} - q,\end{aligned}$$

where $\mathbf{D}$ is any positive semidefinite matrix for which $\operatorname{diag}\mathbf{D}$ is invertible. The last inequality is again using maximum likelihood estimation of the variances of independent normal random variables with mean zero and covariance matrix $\mathbf{D}$, and shows that $\hat{\mathbf{D}}$ gives a solution to (12). It is also clear that the solutions for primary and dual problems are equal.

One other immediate consequence of Theorem 1 is that for every diagonal positive semidefinite $\mathbf{\Sigma}$ satisfying $\mathbf{\Sigma} \leq \mathbf{C}$, and for every positive semidefinite $\mathbf{D}$ it is true that

$$\ln\det\mathbf{\Sigma} \leq -\ln\det\operatorname{diag}\mathbf{D} + \operatorname{trace}\mathbf{C}\mathbf{D} - q, \tag{13}$$

so it is easy to generate upper bounds for the measure of independence. The left-hand side and right-hand side of this last inequality are also seen to be equal when $\mathbf{\Sigma} = \hat{\mathbf{\Sigma}}$ and $\mathbf{D} = \hat{\mathbf{D}}$.

3 Applications

As a first application the measure of independence for equally correlated multivariate normal variables will be displayed. The measure would be expected to change its form according to whether the common correlation coefficient ρ is

positive or negative, since the range of ρ for q equally correlated normal variables is between $-\frac{1}{q-1}$ and 1.

Supposing that $-\frac{1}{q-1} < \rho \leq 0$, it is easily verified that taking

$$\hat{\mathbf{D}} = \frac{1}{1+(q-1)\rho}\mathbf{1}\mathbf{1}',$$

where $\mathbf{1}$ is a $q \times 1$ vector with all elements equal to 1, and so

$$\operatorname{diag} \hat{\mathbf{\Sigma}} = (1+(q-1)\rho)\mathbf{I}_q$$

gives a solution. The difference $\mathbf{C} - \mathbf{\Sigma}$ is equal to $-q\rho(\mathbf{I} - \mathbf{1}\mathbf{1}'/q)$ which is positive semidefinite (and of rank $q-1$). The measure of independence is, from (9),

$$\zeta = \left[\frac{1+(q-1)\rho}{1-\rho}\right]^{\frac{(q-1)}{2}}. \tag{14}$$

If $0 \leq \rho < 1$, then taking

$$\hat{\mathbf{D}} = \frac{q}{(q-1)(1-\rho)}[\mathbf{I} - \mathbf{1}\mathbf{1}'/q],$$

gives a solution. The difference $\mathbf{C}-\mathbf{\Sigma}$ is equal to $\rho\mathbf{1}\mathbf{1}'$ which is positive semidefinite (and of rank 1). The measure of independence is, from (9),

$$\zeta = \left[\frac{1-\rho}{1+(q-1)\rho}\right]^{\frac{(q-1)}{2}}. \tag{15}$$

The results (14), (15) generalise the case $q = 2$ in Rudas, Clogg, and Lindsay (1994), and retain the curious reciprocal property of that special case.

The results for equally correlated normal random variables show that as $q \to \infty$, the measure of independence tends to 0 for fixed ρ. It is hard to say whether that is an attractive property or not.

Note that the rank of $\mathbf{C} - \mathbf{\Sigma}$ and also of $\hat{\mathbf{D}}$ for $\rho \leq 0$ and for $\rho \geq 0$ show examples of the extremes of its possible values. It would not be enough simply to look always for rank 1 or rank $q-1$ matrices.

A second application will show how easy the index is to calculate for a general correlation matrix. The correlation matrix that appears in Table 1 is from a study by Smith and Stanley (1983) on the relation between reaction times and intelligence test scores. Factor analysis results for these data appear in the source paper, and in Bartholomew and Knott (1999) pages 69 to 72.

The dual problem is very well behaved, so it is easy to write a simple algorithm in SPLUS to solve it. The following unsophisticated SPLUS instructions worked well in SPLUS5 on an old Sparcstation 20:

Table 1: Correlation coefficients for Smith and Stanley's data

1	0.466	0.552	0.340	0.576	0.510
0.466	1	0.572	0.193	0.263	0.239
0.552	0.572	1	0.445	0.354	0.356
0.340	0.193	0.445	1	0.184	0.219
0.576	0.263	0.354	0.184	1	0.794
0.510	0.239	0.356	0.219	0.794	1

```
corr<-c(1,0.466,0.552,0.340,0.576,0.510,
0.466,1,0.572,0.193,0.263,0.239,
0.552,0.572,1,0.445,0.354,0.356,
0.340,0.193,0.445,1,0.184,0.219,
0.576,0.263,0.354,0.184,1,0.794,
0.510,0.239,0.356,0.219,0.794,1
);
q<-6;
dim(corr)<-c(q,q);

e<-c(rnorm(q*q));dim(e)<-c(q,q);
theta<-0.01;

for (ii in 1:400) {
obj<- -sum(log(diag(e%*%t(e))))+sum(diag(corr%*%e%*%t(e)))-q;
diff<- solve(diag(diag(e%*%t(e))))-corr;
e1<-e+theta*diff%*%e;
objnew<--sum(log(diag(e1%*%t(e1))))+sum(diag(corr%*%e1%*%t(e1)))-q;
if (obj > objnew) {
e<-e1; theta<-2*theta} else {theta<-theta/2}
};
test<-eigen(solve(diag(diag(e%*%t(e))))-corr)$values;
zeta<-exp(-sum(log(diag(e%*%t(e))))/2-sum(log(eigen(corr)$values))/2)
```

The algorithm seeks to increase the value of the objective function for the dual problem. The correlation matrix is written as `corr`. The algorithm uses a matrix $\mathbf{E}$ such that $\mathbf{D} = \mathbf{E}\mathbf{E}'$ to make sure that $\mathbf{D}$ is positive semidefinite, and uses the derivatives $[(\operatorname{diag} \mathbf{E}\mathbf{E}')^{-1} - \mathbf{C}]\mathbf{E}$ of the objective function with respect to the elements of $\mathbf{E}$ to decide on the right direction to move $\mathbf{E}$. Though it is possible that this algorithm will stick at a local optimum, one has an automatic check on global optimality through testing $\hat{\mathbf{\Sigma}} = (\operatorname{diag} \hat{\mathbf{D}})^{-1} \leq \mathbf{C}$, and can move in a random direction to break the stalemate.

The measure of independence is given by this algorithm as $\zeta = 0.08970137$, so about 9% of the observations could be considered to be from independent normal distributions. The eigenvalues in `test` are 0.0001800282, -0.0007531514, -0.0683073036, -0.3883489215, -0.8617405225, -2.7810610728 showing that to the accuracy expected $(\operatorname{diag} \hat{\mathbf{D}})^{-1} \leq \mathbf{C}$, but that with two eigenvalues close to 0, the rank is effectively 4. This ties in with the presence of a Heywood case that

is discovered by maximum likelihood fitting of a normal factor model to this correlation matrix, see the discussion in Bartholomew and Knott (1999).

4 Factor Analysis

There are similarities between the problem in (10), (11) and the minimum trace method of fitting factor analysis models introduced to psychometrics by Bentler (1972). Given an observed covariance matrix $\mathbf{C}$, the minimum trace method finds a diagonal positive semidefinite matrix $\mathbf{\Sigma}$ such that $\mathbf{\Sigma} \leq \mathbf{C}$ and trace$[\mathbf{C}-\mathbf{\Sigma}]$ is minimised. The method thus seeks to maximise trace $\mathbf{\Sigma}$ while keeping $\mathbf{\Sigma} \leq \mathbf{C}$. It is parallel to the problem in (11) and (12), but maximises trace rather than determinant. The maximisation of trace $\mathbf{\Sigma}$ was considered using optimisation theory by Riccia and Shapiro (1982) and Shapiro (1982a), and in a manner closer to that used in this paper by ten Berge, Snijders, and Zegers (1981). Much of the interest has been in the algorithms to find the solutions, see Jamshidian and Bentler (1998).

The matrices $\hat{\mathbf{\Sigma}}$, $\hat{\mathbf{D}}$ giving the solution for the minimum trace problem are not in general the same as for the problem considered in this paper. One could propose a 'maximum determinant' method for fitting factor analysis models, and use the methods developed in this paper carry out the fitting, but there are perhaps too many ways to fit factor analysis models already. As it happens, the matrices $\hat{\mathbf{\Sigma}}$, $\hat{\mathbf{D}}$ leading to the solutions for the maximum determinant for equally correlated normal variables given in Section 3 also provide a solution for the minimum trace problem.

It is possible to construct a family of criteria for fitting factor analysis models that incudes the maximum determinant method and the minimum trace method as special cases. All that is necessary is to maximise over $\mathbf{\Sigma} = \mathrm{diag}(\sigma_{11}, \sigma_{22}, \ldots, \sigma_{qq})$

$$\frac{1}{\alpha} \ln \sum \sigma_{ii}^{\alpha}/q$$

for some α, $0 < \alpha \leq 1$, where as before $\mathbf{\Sigma} \leq \mathbf{C}$. The case $\alpha = 1$ gives the minimum trace method, while letting $\alpha \to 0$ gives the maximum determinant method. It is even possible to generalise the weighted minimum trace method introduced by Shapiro (1982b) by using the family of criteria

$$\frac{1}{\alpha} \ln \sum p_i \sigma_{ii}^{\alpha}, \tag{16}$$

where p_i are non-zero probabilities that sum to 1. Results for this extension corresponding to Theorem 1 are given in Theorems 2 and 3, which are offered without proof, but which can be proved in a manner not too dissimilar to Theorem 1.

Theorem 2 *Primal generalised weighted minimum trace For* $0 < \alpha < 1$, $\beta = \frac{\alpha}{\alpha-1}$ *and* $\mathbf{W} = \mathrm{diag}(\sqrt{p_1}, \ldots, \sqrt{p_q})$, *where* p_i *are non-zero probabilities adding to 1, (18), (19), (20) are a set of sufficient conditions for a diagonal positive semidefinite matrix* $\hat{\mathbf{\Sigma}} = \mathrm{diag}(\hat{\sigma}_{11}, \hat{\sigma}_{22}, \ldots, \hat{\sigma}_{qq})$ *to give a solution for*

$$\frac{1}{\alpha} \sup \ln[\mathrm{trace}\, \mathbf{W}\mathbf{\Sigma}^{\alpha}\mathbf{W}] \tag{17}$$

over all positive semidefinite diagonal $\mathbf{\Sigma} \leq \mathbf{C}$ *are*

$$\hat{\mathbf{\Sigma}} \leq \mathbf{C}, \tag{18}$$

for $s = 1, \ldots, q$

$$\hat{\mathbf{\Sigma}} = \frac{(\operatorname{diag}\hat{\mathbf{D}})^{\beta-1}}{\operatorname{trace}\mathbf{W}(\operatorname{diag}(\mathbf{D}))^{\beta}\mathbf{W}}, \tag{19}$$

$$\operatorname{trace}[\mathbf{W}(\hat{\mathbf{\Sigma}} - \mathbf{C})\mathbf{W}\hat{\mathbf{D}}] = 0, \tag{20}$$

where $\hat{\mathbf{D}}$ *is a positive semidefinite matrix.*

Theorem 3 *Dual generalised weighted minimum trace For* $0 < \alpha < 1$, $\beta = \alpha/(\alpha - 1)$ *and* $\mathbf{W} = \operatorname{diag}(\sqrt{p_1}, \ldots, \sqrt{p_q})$, *where* p_i *are non-zero probabilities adding to 1, (18), (19), (20) are a set of sufficient conditions for a positive semidefinite matrix with at least one non-zero diagonal element to give a solution for*

$$\inf[\operatorname{trace}[\mathbf{WCWD}] - 1 - \frac{1}{\beta}\ln[\operatorname{trace}[\mathbf{W}(\operatorname{diag}(\mathbf{D}))^{\beta}\mathbf{W}]] \tag{21}$$

over all positive semidefinite $\mathbf{D}$. *Condition (19) can also be written:*

$$\hat{\mathbf{D}} = \frac{(\operatorname{diag}(\hat{\mathbf{\Sigma}}))^{\alpha-1}}{\operatorname{trace}\mathbf{W}\mathbf{\Sigma}^{\alpha}\mathbf{W}}. \tag{22}$$

References

Bartholomew, D. J. and M. Knott (1999). *Latent Variable Models and Factor Analysis* (Second Edition ed.), Volume 7 of *Kendall Library of Statistics.* London: Arnold.

Bentler, P. M. (1972). A lower bound method for the dimension-free measurement of internal consistency. *Social Science and Research 1*, 343–357.

Clogg, C. C., T. Rudas, and L. Xi (1995). A new index of structure for the analysis of models for mobility tables and other cross-classifications. *Sociological Methodology 25*, 197–222.

Ekeland, I. and R. Temam (1976). *Convex Analysis and Variational Problems.* Amsterdam: North Holland Publishing Company.

Halmos, P. R. (1958). *Finite-dimensional Vector Spaces* (Second ed.). Princeton: D. Van Nostrand Company Inc.

Jamshidian, M. and P. M. Bentler (1998). A quasi-Newton method for minimum trace factor analysis. *Journal of Statistical Computing and Simulation 62*, 73–89.

Marshall, A. W. and I. Olkin (1979). *Inequalities: Theory of Majorization and Its Applications.* New York: Academic Press.

Riccia, G. D. and A. Shapiro (1982). Minimum rank and minimum trace of covariance matrices. *Psychometrika 47*, 443–448.

Rudas, T. (1999). The mixture index of fit and minimax regression. *Metrika 50*, 163–172.

Rudas, T., C. C. Clogg, and B. G. Lindsay (1994). A new index of fit based on mixture models for the analysis of contingency tables. *Journal of the Royal Statistical Society, Series B 56*, 623–639.

Shapiro, A. (1982a). Rank reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. *Psychometrika 47*, 187–199.

Shapiro, A. (1982b). Weighted minimum trace factor analysis. *Psychometrika 47*, 243–264.

Smith, G. A. and G. Stanley (1983). Clocking g: Relating intelligence and measures of timed performance. *Intelligence 7*, 353–368.

ten Berge, J. M. F., T. A. B. Snijders, and F. E. Zegers (1981). Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis. *Psychometrika 46*, 201–213.