Søren Tolver Jensen
Søren Johansen

Estimation of Proportional Covariances

Preprint
January
1987

1

Institute of Mathematical Statistics
University of Copenhagen

Søren Tolver Jensen

and

Søren Johansen

ESTIMATION OF PROPORTIONAL COVARIANCES

Preprint 1987 No.1

INSTITUTE OF MATHEMATICAL STATISTICS

UNIVERSITY OF COPENHAGEN

January 1987

Abstract.

In the model for proportional covariance matrices of p-dimensional normally distributed random variables, it is shown that when maximising over the covariance matrix, the profile likelihood is strictly concave. From this result follows the existence and uniqueness of the maximum likelihood estimator.

A simple result on the global convergence of the Newton-Raphson algorithm is given for one dimensional exponential families.

Key words and phrases. Proportional covariances,convexity.

1. INTRODUCTION AND SUMMARY.

Let $X_1,\ldots,X_n$ be independent and p-dimensional normally distributed random variables with mean zero and covariance matrices given by

$$V(X_i) = \lambda_i\Sigma, \ i = 1,\ldots,n.$$

We want to discuss the maximum likelihood estimation of the parameters $(\lambda_1,\ldots,\lambda_n,\Sigma)$. Note first that the parameters are only identified if they are restricted by a constraint like $\prod_{i=1}^{n} \lambda_i = 1$.

The main result is that if the likelihood function is maximised over Σ, then the logarithm of the likelihood profile is strictly concave as a function of $\beta_i = \log(\lambda_i)$, $i = 1,\ldots,n$, under the constraint mentioned above if $n > p$. Further the likelihood profile goes to zero in all directions which guarantees the existence of a unique maximum likelihood estimator.

This problem has been treated recently by Eriksen (1987) and Flury (1986), where the history of the problem is given. The starting point for the recent interest is the report by Guttman, Kim and Olkin (1983)

We shall show that the results follow simply from the following algebraic identity.

LEMMA 1. Let $Y = (Y_1,\ldots,Y_n)$ be a pxn matrix and define for any ordered subset $I = (i_1,\ldots,i_p)$ of $(1,\ldots,n)$ the matrix $Y_I = (Y_{i_1},\ldots,Y_{i_p})$. Then

$$\det(YY^*) = \det(\sum_{i=1}^{n} Y_iY_i^*) = \sum_{|I|=p} \det(Y_IY_I^*).$$

Proof. This result can be derived from the formula for the expansion of a determinant in terms of minors, see Karlin (1968). A proof can be given from first principles as follows: By the definition of the determinant we get

$$\det(YY^*) = \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{j=1}^{p} \left(\sum_{i=1}^{n} Y_{ji} Y_{\sigma(j)i}\right)$$

where the summation is over all permutations σ of $(1,\ldots,p)$.

If we interchange Π and Σ we get

$$\det(YY^*) = \sum_{i_1=1}^{n} \cdots \sum_{i_p=1}^{n} \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{j=1}^{p} Y_{ji_j} Y_{\sigma(j)i_j}.$$

If, for $r \neq s$, we have $i_r = i_s$, then we define τ_{rs} as the transposition of r and s. Then clearly σ and $\sigma \circ \tau_{rs}$ have different sign but the same coefficient, thus the terms cancel each other. Hence we can assume that $(i_1,\ldots,i_p)$ is a p-subset of $(1,\ldots,n)$ and we have

$$\det(YY^*) = \sum_{|I|=p} \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{j=1}^{p} \left(\sum_{i\in I} Y_{ji} Y_{\sigma(j)i}\right)$$

$$= \sum_{|I|=p} \det(Y_I Y_I^*).$$

2. THE EXISTENCE AND UNIQUENESS OF THE MAXIMUM LIKELIHOOD ESTIMATOR.

The likelihood function corresponding to the observations $X_1,\ldots,X_n$ is given by

(2.1)

$$L(\lambda,\Sigma) = (2\pi)^{-np/2} \prod_{i=1}^{n} \lambda_i^{-p/2} (\det(\Sigma))^{-n/2} \exp\left(- \frac{1}{2}\sum_{i=1}^{n} X_i^{*}\Sigma^{-1}X_i \lambda_i^{-1}\right)$$

or

(2.2)

$$-2\log L(\lambda,\Sigma)/n =$$

$$p\log(2\pi) + \log(\det(\Sigma)) + \frac{p}{n}\sum_{i=1}^{n} \log(\lambda_i) + \frac{1}{n}\sum_{i=1}^{n} X_i^{*}\Sigma^{-1}X_i\lambda_i^{-1}$$

THEOREM 1. If $n > p$ then the log likelihood profile is strictly concave as a function of $\beta=\log\lambda$ if β is restricted to any subspace that does not contain the vector $(1,\ldots,1)$. Further the likelihood function goes to zero in all directions,which shows that the maximum likelihood estimator exists uniquely and can be found by optimising a concave funstion.

Proof. It follows from the theory of exponential families, that for fixed $\lambda_1,\ldots,\lambda_n$ the likelihood equation becomes

$$n\Sigma = \sum_{i=1}^{n} X_iX_i^{*}\lambda_i^{-1} . \tag{2.3}$$

And for fixed Σ the equation for λ_i is

$$p\lambda_i = X_i^{*}\Sigma^{-1}X_i . \tag{2.4}$$

If (2.3) is inserted into (2.2) we find the likelihood profile

$$-2\log L(\lambda,\Sigma(\lambda))/n = p\log(2\pi) + \log\det(\Sigma(\lambda)) + \frac{p}{n}\sum_{i=1}^{n}\log(\lambda_i) + p$$

$$= p\log(2\pi) + \log(g(\lambda)) + p.$$

Now define $Y_i = X_i\lambda_i^{-1/2}$, then since

$$n\Sigma(\lambda) = \sum_{i=1}^{n} Y_iY_i^{*}$$

we find from Lemma 1 that

$$\det(\Sigma(\lambda)) = n^{-p}\sum_{|I|=p}\det(Y_I Y_I^*) = n^{-p}\sum_{|I|=p}\prod_{i\in I}\lambda_i^{-1}\det(X_I X_I^*)$$

Hence we get the representation for the function g:

$$(2.5)\qquad g(\lambda) = n^{-p}\Big(\sum_{|I|=p}\Big(\prod_{i=1}^{n}\lambda_i^{p/n-1_I(i)}\Big)\det(X_I X_I^{*})\Big).$$

We can then identify $g(\lambda)$ with the Laplace transform of a measure μ on the set $\{I : |I|=p\}$ with mass $\mu(I) = \det(X_I X_I^*)$.

If $n > p$ then (as long as $\det(\Sigma) > 0$) we have $\mu(I) > 0$ (with probability 1). It is then well known, from the theory of exponential families, that $h(\beta) = \log(g(e^{\beta}))$ is a convex function. Moreover, if we consider $h(\beta_0+t(\beta_1-\beta_0))$, $t\epsilon R$, then this function is strictly convex, unless

$$\sum_{i=1}^{n}(\beta_1-\beta_o)_i(p/n - 1_I(i)) = 0 \text{ for all } I : |I| = p.$$

It follows, by choosing I suitably, that we must have $(\beta_o-\beta_1)_i=c$, but because of the constraint on β we must have $c = 0$, which proves that h is strictly convex.

We shall now show that $g(\lambda)$ tends to infinity if $(\lambda_{max}/\lambda_{min})$ tends to infinity.For any choice of $I(\lambda)$ we have

$$g(\lambda) \geq n^{-p}\prod_{j=1}^{n}\lambda_i^{p/n-1_{I(\lambda)}(i)}\min_{|I|=p}\det(X_I X_I^*).$$

Now we choose $I(\lambda)$ to contain the index of the p smallest values of λ. Without loss of generality we can let $\lambda_1\leq\ldots\leq\lambda_n$ and let $I(\lambda) = (1,\ldots,p)$ then

$$\prod_{i=1}^{n}\lambda_i^{p/n-1_{I(\lambda)}(i)} = \Big(\prod_{i=p+1}^{n}\lambda_i^{p} \Big/ \prod_{i=1}^{p}\lambda_i^{n-p}\Big)^{1/n} \geq (\lambda_n/\lambda_1)^{1/n}$$

Hence, since $\det(X_I X_I^*) > 0$ for all $I : |I| = p$, we find that for λ_n/λ_1 going to infinity we have $g(\lambda)$ going to infinity. This completes the proof of the theorem.

COROLLARY 1. Let $S_1,\ldots,S_k$ be independently distributed such that S_j has a Wishart distribution $W_p(n_j,\tau_j\Sigma)$. Then the maximum likelihood estimator exists uniquely if $\sum_{j=1}^{k} n_j > p$, under the restriction $\prod_{j=1}^{k} \tau_j = 1$.

Proof. This follows by letting $n = \sum_{j=1}^{k} n_j$ and restricting $\lambda_1,\ldots,\lambda_n$ to the subspace

$$\lambda_1=\ldots\lambda_{n_1}=\tau_1, \lambda_{n_1+1}= \cdots = \lambda_{n_1+n_2} = \tau_2, \text{ etc.}$$

Eriksen (1987) proves this result only in the case where $n_j \geq p$, $j=1,\ldots,k$. Flury(1986) has conjectured that the results hold for any positive n_j, $j=1,\ldots,k$. However, for $\sum_{j=1}^{k} n_j = p$ it is seen from (2.5) that the likelihood profile is constant, and for $\sum_{j=1}^{k} n_j < p$ equation (2.3) has no solution.

3. CALCULATION OF THE MAXMIMUM LIKELIHOOD ESTIMATOR.

Since the log-likelihood function is strictly concave a natural proce dure for calculating the estimate is to use a Newton-Raphson procedure with some modification of the step length to guarantee an increasing function. For this purpose we give the derivatives of the likelihood profile.

Using the matrix differentials

$$d(\log(\det(A))) = tr(A^{-1}d(A))$$

and

$$d(A^{-1}) = - A^{-1}(d(A))A^{-1}$$

we find

$$\delta(-2/n\log L(\lambda,\Sigma(\lambda)))/\delta\beta_j = p/n - tr(\Sigma^{-1}(X_jX_j^{*}e^{-\beta_j}))/n$$

$$\delta^2(-2/n\log L(\lambda,\Sigma(\lambda)))/\delta\beta_j\delta\beta_i = -tr(\Sigma^{-1}(X_iX_i^{*}e^{-\beta_i})\Sigma^{-1}(X_jX_j^{*}e^{-\beta_j}))/n^2 + \delta_{ij}tr(\Sigma^{-1}(X_jX_j^{*}e^{-\beta_j}))/n$$

where $\Sigma = \Sigma(\lambda) = (\sum_{i=1}^{n} X_iX_i^{*}e^{-\beta_i})/n$.

One can of course use these to show the concavity of the log likelihood profile, and with some extra work also the strict concavity, but the representation in Lemma 1 is useful for getting the behaviour at infinity.

Another algorithm which is simply implemented is given by the equations (2.3) and (2.4). The m'th step of the algorithm consists in calculating

$$\lambda_i^{(m+1)} = X_i^{*} (\Sigma^{(m)})^{-1}X_i/p$$

and

$$\Sigma^{(m+1)} = n^{-1} \sum_{i=1}^{n} X_iX_i^{*}(\lambda_i^{(m+1)})^{-1}.$$

This algorithm was studied by Eriksen (1987), who proved convergence from any starting value.

An apparently different algorithm is given by Flury(1986). It consists of a reparametrisation of Σ by its eigenvalues and eigenvectors but his PCM-algorithm is in fact the same.

4. MAXIMUM LIKELIHOOD ESTIMATION IN EXPONENTIAL FAMILIES.

In this section we shall show a very simple result about the Newton-Raphson algorithm for one dimensional exponential families, and conjecture that the same result holds for p-dimensional exponential families. Consider a real valued function f defined on R with a continuous positive derivative. We want to discuss the Newton-Raphson algorithm given by

(4.1) $$x_{n+1} = x_n - f(x_n)/f'(x_n) \quad ,n= 0,1,...$$

for some initial value x_0. We define

$$x^* = \inf\{x| \ f(x) \geq 0\}.$$

Since f is increasing we have that f is positive on the interval from x^* to ∞.

THEOREM 2. If

(4.2) f is convex on the interval from x^* to ∞

then the Newton-Raphson algorithm converges for any choice of x_0 to x^*.

Proof. Note that if f has a zero x' then $x' = x^*$, if $f > 0$ then $x^* = -\infty$, and if $f < 0$ then $x^* = +\infty$.

Let us distinguish three cases

1. For all $n = 0,1,\ldots$ we have $f(x_n) < 0$. In this case it is easily seen from (4.1) that x_n is increasing. If the limit is finite then (4.1) implies that the limit is a zero for f, and hence equal to x^*. If the limit is infinite then $f < 0$ and then also $x^* = \infty$.

2. By the same type of argument it follows that if $f(x_n) > 0$ for all n the result holds.

3. The last case is clearly that the sequence $f(x_n)$ contains negative and positive elements, and hence that f has a zero at a finite point x^*. Let n be such that $f(x_n) \geq 0$. Now the convexity of f on the interval from x^* to infinity gives

$$0 = f(x^*) \geq f(x_n) + f'(x_n)(x^* - x_n) = f'(x_n)(x^* - x_{n+1})$$

which shows that $x_{n+1} \geq x^*$, but from (4.1) it follows that $x_{n+1} \leq x_n$. Hence if $f(x_n) \geq 0$ then also $f(x_{n+1}) \geq 0$, and $x_{n+1} \leq x_n$. By induction one gets that x_{n+k} is decreasing as k goes to infinity. The limit point must be finite and it follows from (4.1) that the limit point is a zero for f.

COROLLARY 2. Let f have a continuous positive derivative and define

$$x^{**} = \sup\{x \mid f(x) \leq 0\} .$$

If

$$f \text{ is concave on the interval from } -\infty \text{ to } x^{**}$$

then the Newton-Raphson algorithm converges for any starting point to x^{**}.

Proof. We can apply Theorem 1 to the function $g(x) = -f(-x)$.

Let now μ be a non-negative measure on R with a finite Laplace transform φ. We assume to avoid a trivial case, that μ is not a one point measure. We define the exponential family generated by μ by the densities

$$f(x,\vartheta) = \exp(x\vartheta)/\varphi(\vartheta).$$

It is well known that if X is in the interior of the convex support of μ then there exists a unique maximum likelihood estimator. If X is in either end of the convex supprt of μ then the likelihoood function is monotone giving an infinite value of the estimator. This defines an extended real valued maximum likelihood estimator.

COROLLARY 3. In a one-parameter exponential family with canonical parameter set R, the Newton-Raphson algorithm applied to the reciprocal likelihood function converges for every choice of starting value to the maximum likelihood estimator.

Proof. The reciprocal likelihood function can be written as

$$K(\vartheta) = \int \exp(\vartheta(x-X))\mu(dx).$$

It is well known that K is convex, and that $k(\vartheta) = K'(\vartheta)$ is increasing since

$$k'(\vartheta) = \int (x-X)^2 \exp(\vartheta(x-X))\mu(dx) > 0.$$

We also find that

$$k'''(\vartheta) = \int (x-X)^4 \exp(\vartheta(x-X))\mu(dx) > 0$$

which shows that k'' is strictly increasing. If k'' is positive then k is convex and Theorem 2 can be directly applied. If k''is negative then Corollary 2 can be applied. If k'' changes sign then there is a point ϑ^{**} such that k is concave for $\vartheta < \vartheta^{**}$ and convex for $\vartheta > \vartheta^{**}$. If now k is positive then Corollary 2 applies and if k is negative then Theorem 2 applies. Finally if k has a zero ϑ^{*} then if ϑ^{*} is greater than ϑ^{**} then Theorem 2 can be applied, and if ϑ^{*} is less than ϑ^{**} then Corollary 2 can be applied. This completes the proof.

We have not been able to prove a similar result for p-dimensional exponential families. Such a result would be a useful addition to the applicability of these models.

REFERENCES

FLURY, B.K.(1986), Proportionality of k covariance matrices, Statistics and Probability Letters 4, 29-33.

ERIKSEN, P.S. (1987), Proportionality of covariance matrices, Ann. Statist. (in press).

GUTTMAN, I., KIM, D.Y. and OLKIN, I. (1983), Statistical inference for constants of proportionality between covariance matrices, Technical Report No. 59 Stanford University, Dept. of Statistics.

KARLIN, S. (1968). Total positivity. Stanford University Press, Stanford

PREPRINTS 1986

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jespersen, N.C.B.: On the Structure of Simple Transformation Models.

No. 2 Dalgaard, Peter and Johansen, Søren: The Asymptotic Properties of the Cornish-Bowden-Eisenthal Median Estimator.

No. 3 Jespersen, N.C.B.: Dichotomizing a Continuous Covariate in the Cox Regression Model.

No. 4 Asmussen, Søren: On Ruin Problems and Queues of Markov-Modulated M/G/1 Type.

No. 5 Asmussen, Søren: The Heavy Traffic Limit of a Class of Markovian Queueing Models.

No. 6 Jacobsen, Martin: Right Censoring and the Kaplan - Meier and Nelson - Aalen Estimators.

PREPRINTS 1987

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Jensen, Søren Tolver and Johansen, Søren: Estimation of Proportional Covariances.