Title
Multinormal Maximum Likelihood with a Covariance Bound

Permalink
https://escholarship.org/uc/item/5d77r3dx

Author
Jan de Leeuw

Publication Date
2011-10-25
MULTINORMAL MAXIMUM LIKELIHOOD WITH A COVARIANCE BOUND

JAN DE LEEUW

In some consulting work the problem came up to find the maximum likelihood estimate of the covariance matrix of a multivariate normal distribution in \( n \) dimensions, under the constraint that the estimate is bounded below by a known matrix.

This means we have to solve

\[
\begin{align*}
\min_{\Sigma \succeq \Sigma_0} & \quad \log |\Sigma| + \text{tr} \, \Sigma^{-1} S,
\end{align*}
\]

where \( S \) is an observed covariance matrix and \( \Sigma_0 \) is a known positive definite bound. We use the symbol \( \succeq \) for the Loewner order, i.e. \( A \succeq B \) means that \( A - B \) is positive semi-definite.

We can use an initial change of variables to simplify \((1)\). Since \( \Sigma_0 \) is positive definite, we can define the new variable \( \Theta = \Sigma^{-\frac{1}{2}} \Sigma_0^{-\frac{1}{2}} \) and the new target \( T = \Sigma_0^{-\frac{1}{2}} S \Sigma_0^{-\frac{1}{2}} \). Problem \((1)\) becomes

\[
\begin{align*}
\min_{\Theta \succeq I} & \quad \log |\Theta| + \text{tr} \, \Theta^{-1} T.
\end{align*}
\]

Observe that \( \Theta \succeq I \) simply means that the smallest eigenvalue of \( \Theta \) must be at least one.

Make an additional change of variables. Suppose \( T = K \Lambda K' \) is any eigen-decomposition of \( T \). Define the new variables \( \Xi = K \Theta K' \), and Problem \((2)\) now is

\[
\begin{align*}
\min_{\Xi \succeq I} & \quad \log |\Xi| + \text{tr} \, \Xi^{-1} \Lambda.
\end{align*}
\]

Date: April 5, 2006.

2000 Mathematics Subject Classification. 62H25.

Key words and phrases. Multivariate Analysis.
For a final change of variables, suppose $\Xi = L\Omega L'$ is any eigen-decomposition of $\Xi$. Then

$$\min_{L L' = L' L = I, \Omega \geq I} \min \log |\Omega| + \text{tr} L\Omega^{-1}L'\Lambda.$$ (4)

In scalar notation we can write for (4)

$$\min_{L L' = L' L = I} \sum_{i=1}^{n} \log \omega_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \omega_j^{-1} \ell_{ij}^2.$$ (5)

The matrix $\Pi$ with elements $\pi_{ij} = \ell_{ij}^2$ is doubly stochastic, i.e. it is non-negative and its rows and columns add up to one. We write this as $\Pi \in \mathcal{D}$. Since not every doubly stochastic matrix in $\mathcal{D}$ is the elementwise square of a rotation matrix we have

$$\min_{L L' = L' L = I} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \omega_j^{-1} \ell_{ij}^2 \geq \min_{\Pi \in \mathcal{D}} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \omega_j^{-1} \pi_{ij}.$$ (6)

In other words minimizing over $\mathcal{D}$ is a convex relaxation of the original problem. We now show that actually the relaxed and the original problem have the same solution.

The set $\mathcal{D}$ of doubly stochastic matrices is the convex hull of the set $\mathcal{P}$ of permutation matrices. This is the famous Birkhoff-Von Neumann Theorem [Berge, 1997, page 182]. And, by an equally famous theorem due to Hardy, Littlewood, and Polya [Berge, 1997, page 184],

$$\min_{\Pi \in \mathcal{P}} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \omega_j^{-1} \pi_{ij} = \sum_{i=1}^{n} \frac{\lambda_{[i]}}{\omega_{[i]}}$$ (7)

where the minimum is attained if the permutation puts $\omega$ in the same order as $\lambda$. Indices in square brackets are used for ordered vectors. Thus

$$\lambda_{[1]} \geq \cdots \geq \lambda_{[n]},$$ (8a)

$$\omega_{[1]} \geq \cdots \geq \omega_{[n]}.$$ (8b)

Now it suffices to solve

$$\min_{\omega \geq 0,\omega \geq 1} \sum_{i=1}^{n} \left\{ \log \omega_{[i]} + \frac{\lambda_{[i]}}{\omega_{[i]}} \right\},$$ (9)

for which the solution is simply $\omega_{[i]} = \max(\lambda_{[i]}, 1)$. 
Thus if we define \( \hat{\Omega} \) with diagonal elements \( \max(\lambda_i, 1) \), then the solution to Problem (1) is \( \hat{\Sigma} = \Sigma_0^{\frac{1}{2}} K'\hat{\Omega}K\Sigma_0^{\frac{1}{2}} \).

\( \hat{\Sigma} \) is a continuous function of \( S \), and consequently a consistent estimate of \( \Sigma \). But it is not differentiable at locations where \( S \) does not satisfy the constraint, and thus it is not asymptotically normal and the likelihood ratio test for the constraint is not asymptotically chi-squared. In fact, the results of [Chernoff [1954] show the asymptotic distribution is a mixture of chi-squares, with mixture probabilities given by the asymptotic probabilities that one or more of the smaller eigenvalues violate the constraints. If the true \( \Sigma \) is strictly larger than \( \Sigma_0 \), then \( \hat{\Sigma} \) is almost surely equal to \( S \), and the log likelihood ratio is almost surely equal to zero.

References


Department of Statistics, University of California, Los Angeles, CA 90095-1554

E-mail address, Jan de Leeuw: deleeuw@stat.ucla.edu

URL, Jan de Leeuw: http://gifi.stat.ucla.edu