Journal of Statistical Planning and Inference 142 (2012) 1215–1224

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Efficient estimation for incomplete multivariate data

Bent Jørgensen*, Hans Chr. Petersen

Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

ARTICLE INFO

Article history:
Received 8 August 2011
Accepted 30 November 2011
Available online 16 December 2011

Keywords:
EM algorithm
Estimating function
Fisher scoring algorithm
Godambe information matrix
Missing data
REML estimation

ABSTRACT

We review the Fisher scoring and EM algorithms for incomplete multivariate data from an estimating function point of view, and examine the corresponding quasi-score functions under second-moment assumptions. A bias-corrected REML-type estimator for the covariance matrix is derived, and the Fisher, Godambe and empirical sandwich information matrices are compared. We make a numerical investigation of the two algorithms, and compare with a hybrid algorithm, where Fisher scoring is used for the mean vector and the EM algorithm for the covariance matrix.

1. Introduction

Incomplete multivariate data are a major concern in many applied areas such as for example osteology and paleontology, where the proportion of missing values may be large, and it is clearly important to use incomplete data methods that are both statistically and computationally efficient, see e.g. Holt and Benfer (2000), Stefan (2004) or Petersen (2007). We shall hence compare the two main algorithms for incomplete multivariate normal data, namely the EM algorithm of Dempster et al. (1977), and the Fisher scoring algorithm developed by Trawinski and Bargmann (1964) and Hartley and Hocking (1971). However, we discuss these algorithms in the more general setting of estimating functions under second-moment assumptions, drawing inspiration from the generalized estimating equations (GEE) of Liang and Zeger (1986), and developing suitable matrix representations for the results. The use of estimating functions under second-moment assumptions is a well-established technique, especially in biostatistics, see for example Diggle et al. (2002).

A further motivation comes from the need to obtain a bias-corrected estimator for the covariance matrix $\mathbf{\Sigma}$, which we achieve using the REML-type estimation method developed by Jørgensen and Knudsen (2004), based on a bias-corrected estimating function for $\mathbf{\Sigma}$. The estimate of $\mathbf{\Sigma}$ often serves as input to further multivariate analysis procedures, such as principal components analysis, classification, and clustering, requiring an estimator for $\mathbf{\Sigma}$ that is as accurate as possible. This point is particularly delicate for data with arbitrary missing-data patterns, where the effective degrees of freedom may vary across the different variable pairs in the data. We concentrate on the simple case where the full data are assumed to be i.i.d. from a multivariate distribution with finite second moments, which helps bring out the main points in the discussion without the complications of more general sampling schemes.

The study of estimation based on incomplete samples from the multivariate normal distribution has a long history, dating back to Wilks (1932), Matthai (1951) and Lord (1955), who considered the bivariate and trivariate normal cases.

* Corresponding author. Tel.: +45 6550 3397.
E-mail addresses: bentj@stat.sdu.dk (B. Jørgensen), hcpetersen@stat.sdu.dk (H.C. Petersen).

0378-3758/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2011.11.024

Further special cases were considered by Nicholson (1957) and Buck (1960), after which Trawinski and Bargmann (1964) and Hartley and Hocking (1971) developed the Fisher scoring algorithm in full generality.

A separate development based on techniques for imputing missing values began with Anderson (1957), followed by authors such as Orchard and Woodbury (1972) and Beale and Little (1975), who developed what is now known as the EM algorithm for multivariate normal data. When Dempster et al. (1977) introduced the EM algorithm as a simple and reasonably efficient estimation technique for incomplete data, this algorithm quickly became the standard for incomplete multivariate normal data, see e.g. Liski (1985), Liski and Nummi (1988), Kleinbaum (1973), Schafer (1997, Chapter 5), Johnson and Wichern (2007, pp. 251–256), and Little and Rubin (2002, Chapter 11). The popularity of the EM algorithm all but arrested the further development of the Fisher scoring algorithm in this setting.

The EM algorithm generally requires more iterations for convergence than the Fisher scoring algorithm, although the latter takes more computing time per iteration. However, the EM algorithm does not easily produce the information matrix required for the asymptotic variance of the estimators, and the well-known bias of the maximum likelihood estimator for $\boldsymbol{\Sigma}$ is not easily removed in this context. This motivates a review of the two algorithms along with a discussion of the proper calculation of the Fisher and Godambe information matrices and the derivation of a bias-corrected estimator for $\boldsymbol{\Sigma}$.

2. Incomplete data

We first establish a suitable notation for missing data, following Hartley and Hocking (1971), in order to facilitate the theoretical development and practical implementation of the methods.

Consider a k-vector of data $\boldsymbol{Y}$, partitioned into observed data $\boldsymbol{Y}_r$ and missing data $\boldsymbol{Y}_m$,

$$\begin{bmatrix}\boldsymbol{Y}_r\\ \boldsymbol{Y}_m\end{bmatrix}=\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}\boldsymbol{Y}. \tag{2.1}$$

Here the $k\times k$ matrix $[\boldsymbol{R}^\top,\boldsymbol{M}^\top]^\top$ appearing on the right-hand side of (2.1) is an orthogonal permutation matrix of zeroes and a single one in each row and each column ($\boldsymbol{R}=$ retain; $\boldsymbol{M}=$ missing). It follows from the orthogonality that the inverse of the mapping (2.1) is

$$\boldsymbol{Y}=\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}^\top\begin{bmatrix}\boldsymbol{Y}_r\\ \boldsymbol{Y}_m\end{bmatrix}=\boldsymbol{R}^\top\boldsymbol{Y}_r+\boldsymbol{M}^\top\boldsymbol{Y}_m. \tag{2.2}$$

This matrix representation of the missing data structure is very useful both theoretically and practically, and we may think of (2.1) and (2.2) as giving the relation between the rectangular ($\boldsymbol{Y}$) and ragged ($\boldsymbol{Y}_r$) representation of the data. The orthogonality relation:

$$\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}^\top=\boldsymbol{I}$$

implies the useful relations:

$$\boldsymbol{R}\boldsymbol{R}^\top=\boldsymbol{I},\quad \boldsymbol{M}\boldsymbol{M}^\top=\boldsymbol{I},\quad \boldsymbol{R}\boldsymbol{M}^\top=\boldsymbol{0}. \tag{2.3}$$

Similarly, the relation

$$\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}^\top\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}=\boldsymbol{I}$$

implies

$$\boldsymbol{R}^\top\boldsymbol{R}+\boldsymbol{M}^\top\boldsymbol{M}=\boldsymbol{I}.$$

Let us introduce the notation

$$\boldsymbol{Y}\sim[\boldsymbol{\mu};\boldsymbol{\Sigma}],$$

which means that $\boldsymbol{Y}$ follows a (k-variate) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We consider estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ based on this second-moment assumption. Let $\boldsymbol{\mu}_r$ and $\boldsymbol{\mu}_m$ denote the mean vectors of $\boldsymbol{Y}_r$ and $\boldsymbol{Y}_m$, respectively. Then (2.1) and (2.2) imply

$$\begin{bmatrix}\boldsymbol{\mu}_r\\ \boldsymbol{\mu}_m\end{bmatrix}=\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}\boldsymbol{\mu}\quad\text{and}\quad\boldsymbol{\mu}=\boldsymbol{R}^\top\boldsymbol{\mu}_r+\boldsymbol{M}^\top\boldsymbol{\mu}_m,$$

respectively. Also, (2.1) implies that

$$\mathrm{Var}\begin{bmatrix}\boldsymbol{Y}_r\\ \boldsymbol{Y}_m\end{bmatrix}=\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}\boldsymbol{\Sigma}\begin{bmatrix}\boldsymbol{R}\\ \boldsymbol{M}\end{bmatrix}^\top=\begin{bmatrix}\boldsymbol{R\Sigma R}^\top & \boldsymbol{R\Sigma M}^\top\\ \boldsymbol{M\Sigma R}^\top & \boldsymbol{M\Sigma M}^\top\end{bmatrix}=\begin{bmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{\Sigma}_{rm}\\ \boldsymbol{\Sigma}_{mr} & \boldsymbol{\Sigma}_m\end{bmatrix}, \tag{2.4}$$