Full Information Maximum Likelihood (FIML) for Continuous Variables

Suppose that $\mathbf{y} = (y_1, y_2, \cdots, y_p)'$ has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and that $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$ is a random sample of the vector $\mathbf{y}$.

Specific elements of the vectors $\mathbf{y}_k$, $k = 1, 2, \ldots, n$ may be unobserved so that the data set comprising of n rows (the different cases) and p columns (variables 1, 2, . . . , p) have missing values.

Let $\mathbf{y}_k$ denote a vector with incomplete observations, then this vector can be replaced by $\mathbf{y}_k^* = \mathbf{X}_k \mathbf{y}_k$ where $\mathbf{X}_k$ is a selection matrix, and $\mathbf{y}_k$ has typical elements $(y_{k1}, y_{k2}, \ldots, y_{kp})$ with one or more of the y_{kj}s missing, $j = 1, 2, \ldots, p$.

Example:

Suppose $p = 3$, and that variable 2 is unobserved, then

$$\mathbf{y}_k^* = \begin{bmatrix} y_{k1} \\ y_{k3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{k1} \\ y_{k2} \\ y_{k3} \end{bmatrix}.$$

From the above example it can easily be seen that $\mathbf{X}_k$ is based on an identity matrix with rows deleted according to missing elements of $\mathbf{y}_k$.

If an observed vector $\mathbf{y}_k$ contains no unobserved values, then $\mathbf{X}_k$ is equal to the identity matrix and hence $\mathbf{y}_k^* = \mathbf{y}_k$.

Without loss in generality, $(\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_n)$ can be replaced with $(\mathbf{y}_1^*, \mathbf{y}_2^*, \ldots \mathbf{y}_n^*)$ where $\mathbf{y}_k^*$, $k = 1, 2, \ldots, n$ has a normal distribution with mean $\mathbf{X}_k \boldsymbol{\mu}$ and covariance matrix $\mathbf{X}_k \boldsymbol{\Sigma} \mathbf{X}_k'$. The log-likelihood for the non-missing data is $\sum_{k=1}^{n} \log f(\mathbf{y}_k^*, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, where $f(\mathbf{y}_k^*, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the pdf of $\mathbf{X}_k \mathbf{y}_k$ given the parameters $\boldsymbol{\mu}_k = \mathbf{X}_k \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_k = \mathbf{X}_k \boldsymbol{\Sigma} \mathbf{X}_k'$.

In practice, when data are missing at random, there are usually M patterns of missingness, where $M < n$. When this is the case, the computational burden of evaluating n likelihood functions is considerably decreased.

It is customary to define the chi-square statistic as $\chi^2 = F_0 - F_1$, where $F_0 = -2\ln L_0$, $F_1 = -2\ln L_1$, and where $\ln L_1$ denotes the log-likelihood (at convergence) when no restrictions are imposed on the parameters ($\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$). The quantity $\ln L_0$ denotes the log-likelihood value (at convergence) when parameters are restricted according to a postulated model. The degrees of freedom equals ν**,** where $\nu = p + p(p+1)/2 - k$ and k is the number of parameters in the model.