Likelihood Estimation with Incomplete Array Variate Observations

Deniz Akdemir

September 18, 2012

Abstract

Missing data estimation is an important challenge with high-dimensional data arranged in the form of an array.In this paper we propose a probability model for partially observed multi-way array data. Fisher scoring and expectation maximization are used for estimation of the parameters of this distribution. The main application is to missing data imputation for multi way data.

1 Introduction

A vector is a one way array, a matrix is a two way array, by stacking matrices we obtain three way arrays, etc, ... Array variate random variables up to two dimensions has been studied intensively in Gupta and Nagar [2000] and by many others. For arrays observations of 3, 4 or in general i dimensions probability models have been proposed very recently in (Akdemir and Gupta [2011], Srivastava et al. [2008a] and Ohlson et al. [2011]).

Incomplete data are a major concern for the analysis of array variate random variables. The purpose of this article is to develop likelihood based methods for estimation and inference for a class of array random variables when we only have partially observed arrays.

In Section 2, we introduce a normal model for array variables. In Section 3, we introduce the Full EM and the Hybrid FS-EM algorithms for parameter estimation and missing data imputation. Two examples illustrating the use of these algorithms are in Section 4.

2 Array Normal Random Variable

The family of normal densities with Kronecker delta covariance structure are given by

$$\phi(\widetilde{X};\widetilde{\mathcal{M}},\mathcal{A}_1,\mathcal{A}_2,\ldots\mathcal{A}_i)=\frac{\exp\left(-\frac{1}{2}\|(\mathcal{A}_1^{-1})^1(\mathcal{A}_2^{-1})^2\ldots(\mathcal{A}_i^{-1})^i(\widetilde{X}-\widetilde{\mathcal{M}})\|^2\right)}{(2\pi)^{(\prod_j m_j)/2}|\mathcal{A}_1|^{\prod_{j\neq 1}m_j}|\mathcal{A}_2|^{\prod_{j\neq 2}m_j}\ldots|\mathcal{A}_i|^{\prod_{j\neq i}m_j}} \quad (1)$$

where $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_i$ are nonsingular matrices of orders $m_1, m_2, \ldots, m_i$; the R-Matrix multiplication (Rauhala [2002]) which generalizes the matrix multiplication (array multiplication in two dimensions) to the case of k-dimensional arrays is defined element wise as

$$((\mathcal{A}_1)^1(\mathcal{A}_2)^2 \ldots (\mathcal{A}_i)^i \widetilde{X}_{m_1\times m_2\times\ldots\times m_i})_{q_1q_2\ldots q_i}$$
$$= \sum_{r_1=1}^{m_1}(\mathcal{A}_1)_{q_1r_1}\sum_{r_2=1}^{m_2}(\mathcal{A}_2)_{q_2r_2}\sum_{r_3=1}^{m_3}(\mathcal{A}_3)_{q_3r_3}\ldots\sum_{r_i=1}^{m_i}(\mathcal{A}_i)_{q_ir_i}(\widetilde{X})_{r_1r_2\ldots r_i}$$

and the square norm of $\widetilde{X}_{m_1\times m_2\times\ldots m_i}$ is defined as

$$\|\widetilde{X}\|^2 = \sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2}\ldots\sum_{j_i=1}^{m_i}((\widetilde{X})_{j_1j_2\ldots j_i})^2.$$

Note that R-Matrix multiplication is sometimes referred to as the Tucker product (Kolda [2006]).

The main advantage in choosing a Kronecker structure is the decrease in the number of parameters. The estimation and inference for the parameters of the array normal distribution with Kronecker delta covariance structure, based on a random sample of fully observed arrays $\{\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_N\}$, can been accomplished by maximum likelihood estimation (Srivastava et al. [2008b], Akdemir and Gupta [2011], Srivastava et al. [2008a] and Ohlson et al. [2011]) or by Bayesian estimation (Hoff [2011]).

The operator *rvec* describes the relationship between $\widetilde{X}_{m_1\times m_2\times\ldots m_i}$ and its monolinear form $\boldsymbol{x}_{m_1m_2\ldots m_i\times 1}$. $rvec(\widetilde{X}_{m_1\times m_2\times\ldots m_i}) = \boldsymbol{x}_{m_1m_2\ldots m_i\times 1}$ where $\boldsymbol{x}$ is the column vector obtained by stacking the elements of the array $\widetilde{X}$ in the order of its dimensions; i.e., $(\widetilde{X})_{j_1j_2\ldots j_i} = (\boldsymbol{x})_j$ where $j = (j_i - 1)m_{i-1}m_{i-2}\ldots m_1 + (j_i - 2)m_{i-2}m_{i-3}\ldots m_1 + \ldots + (j_2 - 1)m_1 + j_1$.

The following are very useful properties of the array normal variable with Kronecker Delta covariance structure.

Property 2.1 *If* $\widetilde{X} \sim \phi(\widetilde{X};\ \widetilde{\mathcal{M}},\ \mathcal{A}_1,\ \mathcal{A}_2,\ \ldots\ \mathcal{A}_i)$ *then* $rvec(\widetilde{X}) \sim \phi(rvec(\widetilde{X});\ rvec(\widetilde{\mathcal{M}}),\ \mathcal{A}_i \otimes \ldots \otimes \mathcal{A}_2 \otimes \mathcal{A}_1)$.

Property 2.2 *If* $\widetilde{X} \sim \phi(\widetilde{X}; \widetilde{\mathcal{M}}, \mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_i)$ *then* $E(rvec(\widetilde{X})) = rvec(\widetilde{\mathcal{M}})$ *and* $cov(rvec(\widetilde{X})) = (\mathcal{A}_i \otimes \ldots \otimes \mathcal{A}_2 \otimes \mathcal{A}_1)(\mathcal{A}_i \otimes \ldots \otimes \mathcal{A}_2 \otimes \mathcal{A}_1)'$.

In the remaining of this paper we will assume that the matrices $\mathcal{A}_i$ are square root of the positive definite matrices $\boldsymbol{\Sigma}_i$ for $i = 1, 2, \ldots, i$ and we will put $\Lambda = \boldsymbol{\Sigma}_i \otimes \ldots \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1$.

3 Updating Equations for the Parameters

Using linear predictors for the purpose of imputing missing values in multivariate normal data dates back at least as far as (Anderson [1957]). The EM algorithm

(Dempster et al. [1977]) is usually utilized for multivariate normal distribution with missing data. The EM method goes back to (Orchard and Woodbury [1972]) and (Beale and Little [1975]). Trawinski and Bargmann [1964] and Hartley and Hocking [1971] developed the Fisher scoring algorithm for incomplete multivariate normal data.

Let $\boldsymbol{x}$ be a k dimensional observation vector which is partitioned as

$$\begin{bmatrix} R \\ M \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_r \\ \boldsymbol{x}_m \end{bmatrix}$$

where $\boldsymbol{x}_r$ and $\boldsymbol{x}_m$ represent the vector of observed values and the missing observations correspondingly. Here

$$\begin{bmatrix} R \\ M \end{bmatrix}$$

is an orthogonal permutation matrix of zeros and ones and

$$\boldsymbol{x} = \begin{bmatrix} R \\ M \end{bmatrix}' \begin{bmatrix} \boldsymbol{x}_r \\ \boldsymbol{x}_m \end{bmatrix}.$$

The covariance of $\begin{bmatrix} \boldsymbol{x}_r \\ \boldsymbol{x}_m \end{bmatrix}$ is given by

$$\begin{bmatrix} R \\ M \end{bmatrix} cov(\boldsymbol{x}) \begin{bmatrix} R \\ M \end{bmatrix}' = \begin{bmatrix} \boldsymbol{\Sigma}_{rr} & \boldsymbol{\Sigma}_{rm} \\ \boldsymbol{\Sigma}_{mr} & \boldsymbol{\Sigma}_{mm} \end{bmatrix}.$$

3.1 Fisher Scoring Algorithm

3.1.1 Score Function for $\widetilde{\mathcal{M}}$

Let $\widetilde{X}_1, \widetilde{X}_2, \dots, \widetilde{X}_N$ be a random sample of array observations from the distribution with density $\phi(\widetilde{X}; \widetilde{\mathcal{M}}, \mathcal{A}_1, \mathcal{A}_2, \dots \mathcal{A}_i)$. When the covariance parameters $\mathcal{A}_1, \mathcal{A}_2, \dots \mathcal{A}_i$ are known the score function for $\widetilde{\mathcal{M}}$ is readily available by using the array-monolinear form relationship in Property 2.1 and the corresponding theory for the multivariate normal variable with missing observations.

Let $\boldsymbol{x}_l = rvec(\widetilde{X}_l)$ and

$$\begin{bmatrix} R_l \\ M_l \end{bmatrix} \boldsymbol{x}_l = \begin{bmatrix} \boldsymbol{x}_{rl} \\ \boldsymbol{x}_{ml} \end{bmatrix}$$

for $l = 1, 2, \dots, N$. The score function for $\widetilde{\mathcal{M}}$ is given by

$$\Psi(\widetilde{\mathcal{M}}) = \sum_{l=1}^{N} R_l'(R_l \Lambda R_l')^{-1}(\boldsymbol{x}_{rl} - R_l rvec(\widetilde{\mathcal{M}})).$$

The estimating equation $\Psi(\widetilde{\mathcal{M}}) = 0$ gives the explicit solution

$$rvec(\widehat{\widetilde{\mathcal{M}}}) = J^{-1} \sum_{l=1}^{N} R_l'(R_l \Lambda R_l')^{-1} \boldsymbol{x}_{rl} \tag{2}$$

where J is the information matrix for $rvec(\widetilde{\mathcal{M}})$ and is given by

$$J = \sum_{l=1}^{N} R_l'(R_l \Lambda R_l')^{-1} R_l.$$

The asymptotic covariance for $rvec(\widehat{\widetilde{\mathcal{M}}})$ is therefore J^{-1}.

3.1.2 Score Function for $\widetilde{\mathcal{A}_k}$

Let $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_N$ be a random sample of array observations from the distribution with density $\phi(\widetilde{X}; \widetilde{\mathcal{M}}, \mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_i)$.

Assume $\widetilde{\mathcal{M}}$ and all of $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_i$ are known except for $\mathcal{A}_k$. In this case, the variable

$$\widetilde{Z} = (\mathcal{A}_1^{-1})^1 (\mathcal{A}_2^{-1})^2 \ldots (\mathcal{A}_{k-1})^{-1})^{k-1} (I_{m_k})^k (\mathcal{A}_{k+1}^{-1})^{k+1} \ldots (\mathcal{A}_i^{-1})^i (\widetilde{X} - \widetilde{\mathcal{M}})$$

has density $\phi(\widetilde{Z}; \widetilde{0}, I_{m_1}, I_{m_2}, \ldots I_{m_{k-1}}, \mathcal{A}_k, I_{m_{k-1}} I_{m_i})$.

Now, let $Z_{(k)}$ denote the $m_k \times \prod_{j \neq k} m_j$ matrix obtained by stacking the elements of $\widetilde{Z}$ along the kth dimension. Hence, we can write $Z_{(k)} \sim \phi(Z_{(k)};$ $\mathbf{0}_{m_k \times \prod_{j \neq k} m_j}$, $\mathcal{A}_k$, $I_{\prod_{j \neq k} m_j})$. therefore the corresponding random sample ($Z_{(k)1}$, $Z_{(k)2}$, ..., $Z_{(k)N}$) = ($\boldsymbol{z}_1$, $\boldsymbol{z}_2$, ... $\boldsymbol{z}_{N \prod_{j \neq k} m_j}$) provides a random sample of size $N \prod_{j \neq k} m_j$ from the m_k-variate normal distribution with mean zero and covariance $\boldsymbol{\Sigma}_k = \mathcal{A}_k A_k'$.

Let $\sigma_{k_{lm}}$ denote the lmth element of $\boldsymbol{\Sigma}_k$ for $1 \leq l \leq m \leq m_k$. The corresponding elements of the score function for $\boldsymbol{\Sigma}_k$ under multivariate normality are given by ()

$$\Psi(\boldsymbol{\Sigma}_k)_{lm} = \sum_{q=1}^{N \prod_{j \neq k} m_j} tr\{W_{k_{lmq}} (\boldsymbol{z}_q \boldsymbol{z}_q' - \boldsymbol{\Sigma}_{k_{rrq}})\}$$

where

$$W_{k_{lmq}} = \boldsymbol{\Sigma}_{k_{rrq}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{k_{rrq}}}{\partial \sigma_{k_{lm}}} \boldsymbol{\Sigma}_{k_{rrq}}^{-1}.$$

The sensitivity matrix S_k for $\boldsymbol{\Sigma}_k$, defined as the expected derivative of the estimating function $\Psi(\boldsymbol{\Sigma}_k)_{lm}$ with respect to the entries $\boldsymbol{\Sigma}$, has elements given by

$$S(\boldsymbol{\Sigma}_k)_{(lm)(l'm')} = - \sum_{q=1}^{N \prod_{j \neq k} m_j} tr(\boldsymbol{\Sigma}_{k_{rrq}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{k_{rrq}}}{\partial \sigma_{k_{lm}}} \boldsymbol{\Sigma}_{k_{rrq}}^{-1} \frac{\partial \boldsymbol{\Sigma}_{k_{rrq}}}{\partial \sigma_{k_{l'm'}}}).$$

and dimension $(m_k(m_k+1)/2)^2$. The Newton scoring algorithm for $\boldsymbol{\Sigma}_k$ is hence given by means of the update

$$\boldsymbol{\Sigma}_k^{t+1} = \boldsymbol{\Sigma}_k^t - S(\boldsymbol{\Sigma}_k^t)^{-1} \Psi(\boldsymbol{\Sigma}_k^t) \tag{3}$$

where the result of the matrix product $S(\boldsymbol{\Sigma}_k)^{-1} \Psi(\boldsymbol{\Sigma}_k)$ is understood as a m_k^2 symmetric matrix with lower triangle defined by symmetry.

3.2 The EM Algorithm

Let $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_N$ be a random sample of array observations from the distribution with density $\phi(\widetilde{X}; \widetilde{\mathcal{M}}, \mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_i)$. Let the current values of the parameters be $\widetilde{\mathcal{M}}^t, \mathcal{A}_1^t, \mathcal{A}_2^t, \ldots \mathcal{A}_i^t$.

3.2.1 The updating equation for $\widetilde{\mathcal{M}}$

The updating equation of the parameter $\widetilde{\mathcal{M}}$ is given by

$$\begin{aligned} rvec(M^{t+1}) &= \frac{1}{N}\sum_{l=1}^{N} rvec(\widehat{\widetilde{X}_l}) \\ &= rvec\widetilde{\mathcal{M}}^t + \sum_{l=1}^{N} \Lambda^t R_l'(R_l \Lambda^t R_l')^{-1}(\boldsymbol{x}_{rl} - R_l rvec(\widetilde{\mathcal{M}}^t)) \end{aligned} \tag{4}$$

3.2.2 The updating equation for $\mathbf{\Sigma}_k$

Let

$$\widetilde{Z} = (\mathcal{A}_1^{t\,-1})^1 (\mathcal{A}_2^{t\,-1})^2 \ldots (\mathcal{A}_{k-1}^{t\,-1})^{k-1} (I_{m_k})^k (\mathcal{A}_{k+1}^{t\,-1})^{k+1} \ldots (\mathcal{A}_i^{t\,-1})^i (\widetilde{X} - \widetilde{M}^t).$$

Let $Z_{(k)}$ denote the $m_k \times \prod_{j\neq k} m_j$ matrix obtained by stacking the elements of $\widetilde{Z}$ along the kth dimension with the qth column represented by $\boldsymbol{z}_q$. The updating equation for $\mathbf{\Sigma}_k$ is given by

$$\mathbf{\Sigma}_k^{t+1} = \frac{1}{N\prod_{j\neq k} m_j} \sum_{q=1}^{N\prod_{j\neq k} m_j} [\hat{\boldsymbol{z}}_q \hat{\boldsymbol{z}}_q' + M_q'(\mathbf{\Sigma}_{k_{mmq}}^t - \mathbf{\Sigma}_{k_{mrq}}^t \mathbf{\Sigma}^{t\,-1}_{k_{rrq}} \mathbf{\Sigma}_{k_{rmq}}^t) M_q]. \tag{5}$$

4 Flip-Flop Algorithm for Incomplete Arrays

Inference about the parameters of the model in (1) for the matrix variate case has been considered in the statistical literature (Roy and Khattree [2003], Roy and Leiva [2008], Lu and Zimmerman [2005], Srivastava et al. [2008b], etc.). The Flip-Flop Algorithm Srivastava et al. [2008b] is proven to attain maximum likelihood estimators of the parameters of two dimensional array variate normal distribution. In (Akdemir and Gupta [2011], Ohlson et al. [2011] and Hoff [2011]), the flip flop algorithm was extended to general array variate case.

For the incomplete matrix variate observations with Kronecker delta covariance structure parameter estimation and missing data imputation methods have been developed in Allen and Tibshirani [2010].

The following is a modification of the Flip-Flop algorithm for the incomplete array variable observations:

Algorithm for estimation:

Given the current values of the parameters, repeat steps 1 and 2 until convergence:

1. Update $\widetilde{\mathcal{M}}$ using (2) or (4),
2. For $k = 1, 2, \ldots, i$ update $\mathbf{\Sigma}_k$ using (3) or (5).

Note that at each step of this algorithm we can choose the EM or Fisher Scoring updating equations. Therefore there are four modifications possible:

1. Full FS: Both steps of the estimation algorithm uses the Fisher scoring updating equations.
2. Full EM: Both steps of the estimation algorithm uses the EM updating equations.
3. Hybrid FS-EM: First step uses the Fisher scoring update and second step uses the EM update.
4. Hybrid EM-FS: First step uses the EM update and second step uses the Fisher scoring update.

In the following we have only implemented the Full EM and the Hybrid FS-EM algorithms.

5 Illustrations

Example 5.1 *In this first example we have simulated data from a 2×2 array normal distribution with differing number of observations. For each sample size, we have repeated the experiment 10 times. The convergence of the estimator of Λ is checked by reporting the mean $L = \|\Lambda - \widehat{\Lambda}\|^2$ over 10 trials at each sample size. True covariance components were $\begin{bmatrix} 2 & .6 \\ .6 & 3 \end{bmatrix}$ and $\begin{bmatrix} 4 & -.6 \\ -.6 & 1 \end{bmatrix}$. Sample sizes 50,100, 200 and 500 were used. Missing data intensity defined as the proportion of the number of randomly selected (with replacement) data points that were set to missing to the total number of data points, in the experiments this was set to $\frac{1}{4}$. Figure 1 display the results from the Hybrid and EM algorithms. As the number of observations increase, L decreases towards zero.*

Example 5.2 *In this example, we use a subset of the data previously analyzed by Basford et al. [1991]. The most comprehensive analyses of these data as well as experimental details can be found in Basford and Tukey [1999]. The data set involved measurements on 58 different soybean lines observed on 6 traits and in 8 environments. Because of the low number of replications, we have only included the first 20 lines in our analysis. We have assumed that the 20×6 matrices of observations from different environments (4 locations, 2 times) were independent and identically generated from a two way array (matrix) normal distribution. We have deleted all the observations for the lines 1 through 10 for the last environment and estimated these using the Full EM algorithm. The average correlation between the true and the estimated values over the 6 variables was 0.57. We have also used applied imputation using 2-nearest neighbors regression*

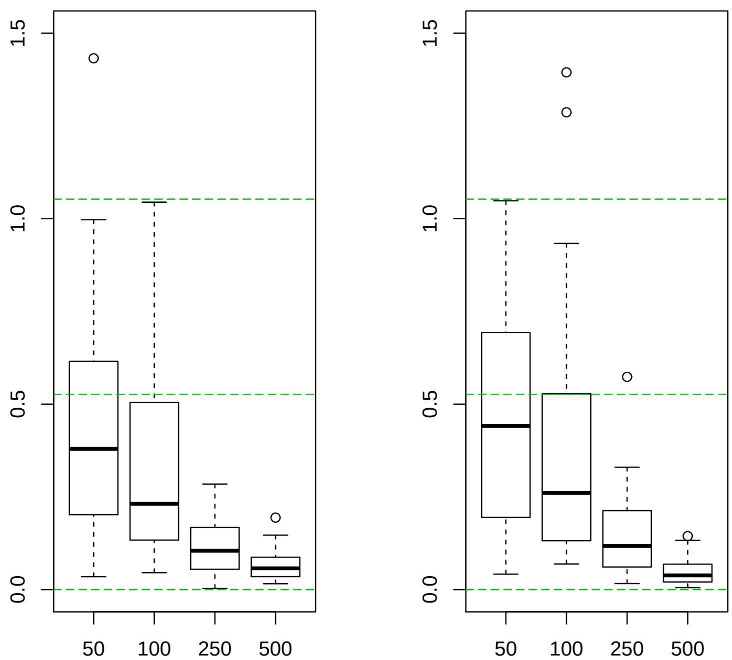

Figure 1: The convergence of the Full EM (Left) and the Hybrid FS-EM (Right) algorithms. As the number of observations increase, L decreases towards zero.

(Hastie et al. [2001]) and random forest regression (Stekhoven and Bühlmann [2012]) using the 120×8 *data matrix representing the 120 variable-location pairs and 8 replications. The corresponding correlation values were 0.55 and 0.57.*

6 Conclusions

We have formulated a parametric model for array variate data and developed suitable estimation methods for the parameters of this distribution with possibly incomplete observations. The main application of this paper has been to multi-way regression (missing data imputation), once the model parameters are given we are able to estimate the unobserved components of any array from the observed parts of the array. We have assumed no structure on the missingness pattern other than assuming that it is fixed.

The methods developed here use the assumption that the data is generated from a distribution with Kronecker delta covariance structure. The suitability of this model to any data set is questionable. The choice of model and and determination of its order could be accomplished using a model selection criteria based on the likelihood function which is available through the results in this paper.

References

D. Akdemir and A. K. Gupta. Array variate random variables with multiway kronecker delta covariance matrix structure. *Journal of Algebraic Statistics*, 2(1):98–113, 2011.

G.I. Allen and R. Tibshirani. Transposable regularized covariance models with an application to missing data imputation. *The Annals of Applied Statistics*, 4(2):764–790, 2010.

T.W. Anderson. Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the american Statistical Association*, 52(278):200–203, 1957.

K.E. Basford and J.W. Tukey. *Graphical analysis of multiresponse data: Illustrated with a plant breeding trial*. CRC Press, 1999.

KE Basford, PM Kroonenberg, and IH DeLacy. Three-way methods for multiattribute genotype ũ environment data: an illustrated partial survey. *Field Crops Research*, 27(1-2):131–157, 1991.

E.M.L. Beale and R.J.A. Little. Missing values in multivariate analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 129–145, 1975.

A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC Monographs and Surveys in Pure and Applied Mathematics. Chapman and Hall, 2000.

HO Hartley and RR Hocking. The analysis of incomplete data. *Biometrics*, pages 783–823, 1971.

T. Hastie, R. Tibshirani, B. Narasimhan, G. Chu, T. Hastie, R. Tibshirani, B. Narasimhan, and G. Chu. impute: Imputation for microarray data. *Bioinformatics*, 17(6):520–525, 2001.

P.D. Hoff. Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, 55(1):530–543, 2011.

T.G. Kolda. *Multilinear operators for higher-order decompositions*. United States. Department of Energy, 2006.

N. Lu and D.L. Zimmerman. The Likelihood Ratio Test for a Separable Covariance Matrix. *Statistics & Probability Letters*, 73(4):449–457, 2005.

M. Ohlson, M. Rauf Ahmad, and D. von Rosen. The multilinear normal distribution: Introduction and some basic properties. *Journal of Multivariate Analysis*, 2011.

T. Orchard and M.A. Woodbury. A missing information principle: theory and applications. In *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 697–715, 1972.

U.A. Rauhala. Array Algebra Expansion of Matrix and Tensor Calculus: Part 1. *SIAM Journal on Matrix Analysis and Applications*, 24:490, 2002.

A. Roy and R. Khattree. Tests for Mean and Covariance Structures Relevant in Repeated Measures Based Discriminant Analysis. *Journal of Applied Statistical Science*, 12(2):91–104, 2003.

A. Roy and R. Leiva. Likelihood Ratio Tests for Triply Multivariate Data with Structured Correlation on Spatial Repeated Measurements. *Statistics & Probability Letters*, 78(13):1971–1980, 2008.

MS Srivastava, T. Nahtman, and D. von Rosen. Estimation in General Multivariate Linear Models with Kronecker Product Covariance Structure. *Research Report Centre of Biostochastics, Swedish University of Agriculture science. Report*, 1, 2008a.

M.S. Srivastava, T. von Rosen, and D. Von Rosen. Models with a Kronecker Product Covariance Structure: Estimation and Testing. *Mathematical Methods of Statistics*, 17(4):357–370, 2008b.

D.J. Stekhoven and P. Bühlmann. Missforest nonparametric missing value imputation for mixed type data. *Bioinformatics*, 28(1):112–118, 2012.

I.M. Trawinski and RE Bargmann. Maximum likelihood estimation with incomplete multivariate data. *The Annals of Mathematical Statistics*, 35(2): 647–657, 1964.