Journal of Computational and Applied Mathematics 239 (2013) 37–49

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

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

Maximum likelihood estimation for the tensor normal distribution: Algorithm, minimum sample size, and empirical bias and dispersion

Ameur M. Manceur [a], Pierre Dutilleul [a,b,*]

[a] Department of Plant Science, McGill University, Macdonald Campus, 21,111 Chemin du bord du lac, Ste-Anne-de-Bellevue, Québec, H9X 3V9, Canada
[b] Department of Mathematics and Statistics, McGill University, 805 rue Sherbrooke Ouest, Montréal, Québec, H9X 3V9, Canada

ARTICLE INFO

Article history:
Received 15 November 2011

Keywords:
Empirical bias and dispersion
Maximum likelihood estimation
Minimum sample size
Multi-stage algorithm
Separable variance–covariance structure
Tensor normal distribution

ABSTRACT

Recently, there has been a growing interest in the analysis of multi-dimensional data arrays (e.g. when a univariate response is sampled in 3-D space or when a multivariate response is sampled in time and 2-D space). In this article, we scrutinize the problem of maximum likelihood estimation (MLE) for the tensor normal distribution of order 3 or more, which is characterized by the separability of its variance–covariance structure; there is one variance–covariance matrix per dimension. In the 3-D case, the system of likelihood equations for the three variance–covariance matrices has no analytical solution, and therefore needs to be solved iteratively. We studied the convergence of an iterative three-stage algorithm (MLE-3D) that we propose for this, determined the minimum sample size required for matrix estimates to exist, and computed by simulation the empirical bias and dispersion of the Kronecker product of the three variance–covariance matrix estimators in eight scenarios. We found that the standardized bias and a matrix measure of dispersion decrease monotonically and tend to vanish with increasing sample size, so the Kronecker product estimator is consistent. An example with 3-D spatial measures of glucose content in the brain is also presented. Finally, results are discussed and the 4-D case is presented with simulation results in an appendix. Software is available for interested users.

1. Introduction

There is growing literature on the analysis of two-dimensional (2-D) and three-dimensional (3-D) data arrays, also called "multi-way data" [1–5]. Such data present correlations and heterogeneity of the variance, both within and among dimensions, through multiple responses and space–time levels. The variance–covariance structure is then often modeled to reduce the number of parameters and ensure the existence of parameter estimates. In a separable model (sometimes called "factorized" or "Kronecker structured"), the variance–covariance matrix of the vectorized multi-dimensional array is the Kronecker (direct) product of a number of variance–covariance matrices equal to the number of dimensions. The variance–covariance matrices used as factors in the Kronecker product define the respective dependencies and variability among rows and columns in 2D and among rows, columns and edges (or slices) in 3D and beyond.

In 2D, Dutilleul [6–8] presented an iterative two-stage algorithm (MLE-2D) to estimate by maximum likelihood (ML) the variance–covariance parameters of the matrix normal distribution $\mathbf{X} \sim N_{n_1,n_2}(\mathbf{M}, \mathbf{U}_1, \mathbf{U}_2)$, where the random matrix $\mathbf{X}$ is $n_1 \times n_2$, $\mathbf{M} = E(\mathbf{X})$, $\mathbf{U}_1$ is the $n_1 \times n_1$ variance–covariance matrix for the rows of $\mathbf{X}$ (e.g. repeated measures in space), and $\mathbf{U}_2$ is

* Corresponding author at: Department of Plant Science, McGill University, Macdonald Campus, 21,111 Chemin du bord du lac, Ste-Anne-de-Bellevue, Québec, H9X 3V9, Canada. Tel.: +1 514 398 7870; fax: +1 514 398 7897.
E-mail address: pierre.dutilleul@mcgill.ca (P. Dutilleul).

0377-0427/$ – see front matter
doi:10.1016/j.cam.2012.09.017

the $n_2 \times n_2$ variance–covariance matrix for the columns of $\mathbf{X}$ (e.g. repeated measures in time). The matrix normal distribution model implies a separable variance–covariance structure, defined by $\mathbf{U}_2 \otimes \mathbf{U}_1$. Other authors also studied the MLE-2D algorithm, and later nicknamed it "flip-flop" [9]. Werner et al. [5] compared it to four alternative estimation procedures, and found it was providing estimators with the lowest normalized root-MSE (mean square error), starting with very small sample sizes. The MLE-2D algorithm was found to be useful in brain science [3], image analysis [1], biochemistry [4], electrical engineering [5], and the environmental sciences [10], for example. Two unstructured variance–covariance matrices (with no other assumption than positive definiteness) are then estimated using a small number K of replicates, with $K \geq \max(\frac{n_1}{n_2}, \frac{n_2}{n_1}) + 1$, in order to ensure that the estimated variance–covariance matrices are positive definite.

Three-dimensional data arrays are obtained when a single response is sampled in 3-D space or in 2-D space and time or when multiple responses are recorded in 2-D space or in 1-D space and time. In the natural and life sciences, such data are provided by the measurement of wood density in given growth rings and directions at several heights in a tree trunk [11], the recording of a vector of air pollutants at a number of field stations over months [2], and the monitoring of a vector of physiological variables in different organs over days [12]. These data have rarely been analyzed by using a 3-D statistical methodology, apparently because it was not easy to access and the computational tools were not available. As the collection of 3-D data arrays is rising, it has become timely to fill in the gap. Therefore, we present the MLE-3D algorithm, define its conditions of application, and study by simulation the properties of estimators in this article. In that non-trivial extension of the MLE-2D algorithm, the parameters are estimated by maximum likelihood under the relevant statistical distribution called a "tensor normal distribution" and characterized by multivariate normality and a separable three-way variance–covariance structure, with no need to specify a variance–covariance matrix model at each dimension. Below, we summarize the approach followed and the results obtained in five earlier studies where a three-way separable variance–covariance structure was used for data analysis, prior to inserting our contribution in the developing field and explaining how our article is organized.

In 1993, Barton and Fuhrman [13] explored the modeling of the variance–covariance structure of multi-dimensional data arrays that commonly arise in signal processing problems. They presented a notation system based on a "natural hierarchical block structure on the covariance data", and discussed the variance–covariance structures of block-circulant, block-Toeplitz type *vs*. unstructured type, with a limited discussion of estimation algorithms. Corrections were provided in [14].

Still in 1993, Mardia and Goodall [2, p. 358] presented an iterative three-stage estimation algorithm which resembles the MLE-3D algorithm that will be presented here, but the authors did not use it to analyze their multivariate spatio-temporal data due to insufficient replication. Eventually, they applied the MLE-2D algorithm by making the assumption of some temporal independence, and reported convergence in 10–14 iterations [2, p. 357]. Throughout, the authors assumed that the expected value of the random multi-dimensional array was constant along one of the three dimensions (i.e. time), and chose to use an isotropic variogram spherical model for the spatial variance–covariance matrix.

In the context of the analysis of doubly and triply repeated measures in the medical sciences, Galecki [15] tried various types of variance–covariance structures, including autoregressive, compound symmetric, spherical (i.e. independence and homoscedasticity), and unstructured. He also developed the concept of covariance profile, and presented one application of an estimation algorithm without detailing it.

In 2006, an iterative ML algorithm for 3-D data arrays was proposed in [12], together with a likelihood ratio test (LRT) aimed to assess the adequacy of a three-way separable model for the variance–covariance structure. The estimation algorithm assumed an autoregressive or compound symmetric structure for one variance–covariance matrix, and an intra-class correlation structure was assumed for the two others. The simulation study designed to verify the small-sample behavior of the LRT was limited to a reduced form of the tensor normal distribution of order 3, in which one of the three variance–covariance matrices was a scalar.

More recently, a Newton–Raphson type of algorithm (i.e. with no "flip-flop") was used in [16] to estimate by maximum likelihood the parameters of the tensor normal distribution model. This algorithm was developed in the context of tensor-valued signals in electrical engineering. In that Newton–Raphson ML algorithm, a score function is used and a variance parameter is estimated in addition to the three variance–covariance matrices. Furthermore, Richter et al. [16] used a block-diagonal approximation for the Hessian matrix required by the Newton–Raphson algorithm.

To our knowledge, Dutilleul [6] was the first to present the probability density function and the moment generating function of the tensor normal distribution, using tensor notations inspired from McCullagh [17]. Other notations for tensor operators such as the inner and outer products and the tensor multiplication appear more popular nowadays [18], so Dutilleul's original equations were re-written in Section 2 here. Still to our knowledge, Mardia and Goodall [2], followed by Dutilleul [7], first proposed the MLE-3D algorithm for the tensor normal distribution, which was recently presented in [19, p. 185, in 3-D or more] and [20, p. 15] in a theoretical framework and without numerical and simulation results concerning the convergence of the algorithm and the empirical properties of the estimators. This is where our contribution lies, i.e. in a detailed presentation of the algorithm with numerical and simulation results in addition to practical guidelines and software.

More specifically, we start by defining the tensor normal distribution of order 3 or more from the moment generating function in the general case and from the probability density function in the regular case (of particular interest for ML estimation) in Section 2; so doing, we will see that only the Kronecker product of variance–covariance matrices is defined uniquely. In Section 3, the system of likelihood equations for the mean and variance–covariance parameters of the tensor normal distribution of order 3 is derived; the complete MLE-3D algorithm is presented in details; and the minimum sample