FACTOR ANALYSIS USING THE BHATTACHARYYA DISCREPANCY FUNCTION

KHALED ALYANI*, MAHER MOAKHER*

* *Laboratory for Mathematical and Numerical Modeling in Engineering Science, National Engineering School at Tunis, University of Tunis El Manar, ENIT-LAMSIN, B.P. 37, 1002 Tunis Belvédère, Tunisia.*

Abstract. Maximum likelihood factor analysis method has been considered for many years an effective method for the estimation of factor matrices. In this work, based on a Bhattacharyya discrepancy function, we propose new approach for the solution of problems involving factor analysis. We will compare this approach with the method of Maximum likelihood.

1. Introduction and problem formulation

Factor analysis is a statistical method that was initially developed by psychologists and pioneered by the works of Spearman and Thomson. It is a branch of multivariate analysis which deals with the internal structure of matrices of covariances and correlations. It is used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. Even though factor analysis was originated in psychometrics, it has been used in behavioral sciences, social sciences, marketing, product management, operations research and other fields that deal with data sets where there are large numbers of observed variables that are thought to reflect a smaller number of latent variables.

In factor analysis it is assumed that the covariance matrix $\mathbf{C}$ has the form:

$$\mathbf{C} = \mathbf{L}\mathbf{L}^{\mathbf{T}} + \mathbf{D},$$

where $\mathbf{L}$ is the matrix of factor loadings and $\mathbf{D}$ is a diagonal matrix. The most interesting and important studies dealing with factor analysis problem were concerned with obtaining efficient estimates of the factor loadings $\mathbf{L}$ and of the residual variances by the method of maximum likelihood [1, 2] . In this work we give a new approach to the estimation of the problem of factor analysis, based on the Bhattacharyya discrepancy function, resulting from the Bhattacharyya distance (also called log-determinant 0-distance) recently defined in [3]:

$$d_B(\mathbf{X}, \mathbf{Y}) := 2\sqrt{\log \frac{\det \frac{1}{2}(\mathbf{X} + \mathbf{Y})}{\sqrt{\det(\mathbf{X}) \det(\mathbf{Y})}}},$$

where $\mathbf{X}$ and $\mathbf{Y}$ are two symmetric positive-definite matrices. This approach will be illustrated through applications in psychology and in geoscience.

2. Numerical results

In a comparison with the maximum likelihood method, primary computations has shown a better performance of our approach. Further results will be provided.

References

[1] K. G. Jöreskog. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4):443–482, 1967.

[2] J. C. De Winter and D. Dodou. Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, 39(4):695–710, 2012.

[3] Z. Chebbi and M. Moakher. Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function. *Linear Algebra and its Applications*, 436(7):1872–1889, 2012.