



Covariance matrix estimation using repeated measurements when data are incomplete

Mohammad Y. Al-Rawwash

Department of Statistics, Yarmouk University, Irbid, Jordan

Abstract

Modeling repeated measurements data has been studied extensively lately in the parametric situation. However, it is significant to study the effect of various treatments over a period of time where the repeated measurements on the same subject are expected to be correlated. The correlation among the repeated measurements for all the subjects will be studied via the covariance matrix. The positive definite constraint is one of obstacles that encounters modeling the covariance structure, however the Cholesky decomposition removes this constraint and allows modeling the components of the covariance matrix [M. Pourahmadi, Joint mean-covariance models with applications to longitudinal data: unconstrained parameterization, *Biometrika* 86 (1999) 677–690]. In this paper, we adopt the estimation procedures introduced by Diggle and Verbyla [Nonparametric estimation of covariance structure in longitudinal data, *Biometrics* 54 (1998) 401–415] where the variogram cloud as well as the squared residuals are used to estimate the variogram and the variances via the kernel smoothing. Selecting the appropriate bandwidth value is one of the important steps in the estimation process, thus in our data analysis we choose the bandwidth using one of the most simple straight forward methods which is the cross-validation method developed by Rice and Silverman [Estimating the mean and covariance structure nonparametrically when the data are curves, *J. Roy. Statist. Soc. B* 53 (1991) 233–243] and adapted by others. Finally we apply these nonparametric

E-mail address: rawwash@yu.edu.jo

techniques as well as a graphical method [M. Pourahmadi, Joint mean-covariance models with applications to longitudinal data: unconstrained parameterization, *Biometrika* 86 (1999) 677–690] to a real life data and use the penalized likelihood criterion like AIC and BIC to compare models of our interest.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Cholesky decomposition; Covariance matrix; Cross-validation; Graphical methods; Kernel smoothing; Longitudinal data; Repeated measurements; Variogram cloud

1. Introduction

In longitudinal studies, modeling the covariance matrix is considered to be a vital issue, however a basic goal is to model and describe the mean response as a function of time as well as other related explanatory variables. In real life data set, missing values for different subject is commonly occurred and causing the data set to be incomplete. The sample covariance matrix for such data set could be negative definite when there is a severe loss of observation which creates another problem that needs to be addressed. The correlation among the repeated measurements has been studied in the last three decades in the parametric literature [20,26,1,13]. It is traditional to adopt some stationary covariance structures from a collection that includes autoregressive, compound symmetry, . . . , etc. On the other hand, these well-known stationary structure includes few parameters and may not be suitable for fitting certain data. Hence, this was a motivation for some substantial work that appeared in the nonparametric and Bayesian literature [23,29,4,5]. An alternative technique is to use a data-driven method where the data itself shapes the structure of the covariance matrix.

As an illustration, we will study the mice data [18] that consists of 30 mice each of them has been injected with mouse colon carcinoma cell and the size of the tumor has been measured on eleven consecutive occasions. The data set includes 28 missing values which could have caused the covariance matrix to be negative definite. These missing values occurred due to the death of 11 mice on different times. Had it been that the second mouse as well as the fourth mouse died before the end of the experiment then the outcome data set will provide a negative definite covariance matrix. This could be a motivation to find an alternative method to model the covariance matrix and override this constraint and provide an efficient estimate of Σ using a two stage process: a nonparametric approach is used in the first stage to estimate Σ , while in the second stage we adopt a parametric graphical method to model the components of the covariance matrix.

The general setup and notation will be as follows. Suppose that we have m different subject each of which has n_i different repeated measurements. Also, let

Download English Version:

<https://daneshyari.com/en/article/9506479>

Download Persian Version:

<https://daneshyari.com/article/9506479>

[Daneshyari.com](https://daneshyari.com)