Multivariate exponential dispersion models

Bent Jørgensen

University of Southern Denmark, Denmark, email: bentj@stat.sdu.dk

Keywords: convolution method, exponential dispersion model, multivariate gamma distribution, multivariate generalized linear model, multivariate Poisson distribution.

In order to develop a general approach for analysis of non-normal multivariate data, it would be desirable to obtain a simple-minded framework that can accommodate a wide variety of different types of data, much like generalized linear models do in the univariate case. There is no shortage of multivariate distributions available, but the main stumbling block so far has been the lack of a suitable multivariate form of exponential dispersion model.

In the univariate case, an exponential dispersion model $\mathrm{ED}(\mu, \sigma^2)$ is a two-parameter family parametrized by the mean μ and dispersion parameter σ^2, with variance $\sigma^2 V(\mu)$, where V denotes the unit variance function. The generalized linear models paradigm is based on combining a link function with a suitable linear model. Estimation uses quasi-likelihood for the regression parameters, and the Pearson statistic for estimating the dispersion parameter.

We consider a new k-variate exponential dispersion model $\mathrm{ED}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ aimed at providing a fully flexible covariance structure corresponding to a mean vector $\boldsymbol{\mu}$ and a positive-definite dispersion matrix $\boldsymbol{\Sigma}$. The covariance matrix is of the form $\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma} \odot \boldsymbol{V}(\boldsymbol{\mu})$, where $\odot$ denotes the Hadamard (elementwise) product between two matrices, and $\boldsymbol{V}(\boldsymbol{\mu})$ denotes the (matrix) unit variance function. We consider a multivariate generalized linear model for independent response vectors $\boldsymbol{Y}_i \sim \mathrm{ED}_k(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ defined by $g(\boldsymbol{\mu}_i^\top) = \boldsymbol{x}_i \boldsymbol{B}$, where the link function g is applied coordinatewise to $\boldsymbol{\mu}_i^\top$, $\boldsymbol{x}_i$ is an m-vector of covariates, and $\boldsymbol{B}$ is an $m \times k$ matrix of regression coefficients. We estimate the regression matrix $\boldsymbol{B}$ using a quasi-score function, and we estimate the dispersion matrix $\boldsymbol{\Sigma}$ using a multivariate Pearson statistic defined as a weighted sum of squares and cross-products matrix of residuals. This model specializes to the classical multivariate multiple regression model when g is the identity function and $\mathrm{ED}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the multivariate normal distribution.

The construction of the multivariate exponential dispersion model $\mathrm{ED}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is based on an extended convolution method, which makes the marginal distributions follow a given univariate exponential dispersion model. We illustrate the method by considering multivariate versions of the Poisson and gamma distributions, and discuss some of the challenges faced in the implementation of the method.