Association Models for a Multivariate Binary Response

Anders Ekholm
University of Helsinki, Department of Statistics,
P.O.Box 54, 00014 University of Helsinki, Finland
E-mail: anders.ekholm@helsinki.fi

Abstract

Let $Y_i = (Y_{i1}, \ldots, Y_{iq})$ denote the multivariate binary response of unit $i$ with $q_i$ subunits and $i = 1, \ldots, n$. The units are for example patients and the subunits are repeated measurements on the patient, or the units are families and the subunits are the members of a family. Associated with each response $Y_i$ is a vector $x_i$ of $p$ explanatory variables. The $Y^{(1)}, \ldots, Y^{(n)}$ are assumed independent. Any statistical model, purporting to be useful, should consist of two sets of equations:

(i) marginal regression equations for $\mu_{it} = E(Y_{it} | x_{it}) = h(x_{it}, \beta)$, where $t = 1, \ldots, q_i$, $h(\cdot)$ is a known function and $\beta$ is a vector of regression parameters, and

(ii) equations describing, in terms of a vector $\alpha$ of association parameters, the mechanism which generates association between $Y_{i1}, \ldots, Y_{iq}$, that is, between the subunits of a unit.

The solution to this twofold problem is well known when $Y^{(i)}$ is multivariate normal and $h(x_{it}, \beta) = x_{it} \beta^T$, denoted $Y^{(i)} \sim N_q(\mu_i(\beta), \Sigma_i(\alpha))$, where $\mu_i = (\mu_{i1}, \ldots, \mu_{iq})$. The association structure is then specified by deriving the elements of the $q_i \times q_i$ matrix $\Sigma_i$ from the association mechanism described in (ii). When $Y_{it}$ is a Bernoulli trial, then $Y^{(i)}$ is a $B_q$ distribution, that is, $Y^{(i)}$ follows a $q_i$-dimensional Bernoulli distribution, with $\pi_{i}(\beta, \alpha)$ a $2^{q_i}$ vector of cell probabilities. For the $B_q$ distribution no uniformly adopted approach exists. The association structure of the $B_q$ distribution is, in fact, much richer than that of the $N_q$ distribution and more entangled with the univariate means. There is no natural separation of $\pi$ into mean and association parameters like there is for the $N_q$ distribution. We parametrise the $B_q$ distribution by the univariate marginal probabilities and dependence ratios of all orders. This parametrisation supports likelihood inference for both $\beta$ of model (i) and $\alpha$ of model (ii). Five types of association models are proposed and illustrated by reanalysing three empirical data sets. The proposed approach is compared and contrasted with association models for multivariate normal responses, and with other models for multivariate binary responses.

REFERENCES

