Bayesian Analysis (2009) **4**, Number 3, pp. 457–464

Rejoinder

Stefano Monni[1] and Mahlet G. Tadesse[2]

We thank the discussants for their remarks and insights. We will organize our replies by topics, as some questions were raised by more than one discussant.

Nature of the Correlation and Effective Sample Size

Some of the observations made by the discussants are on the nature of the correlation captured by our model. We agree with Professors Chipman, George and McCulloch (hereafter referred to as CGM) that the correlation captured by our method is not the same as that captured by the method of Brown et al. (1998) (BVF). In our model, the error terms ε_j in a component are assumed to be independent. Nevertheless, the outcomes Y_j are correlated because they have the same dependence on the predictor variables $\sum_r X_{ri}\beta_r$ (Breiman and Friedman 1997). We recognize that the totality of the correlation among outcomes may not be captured by assuming independent errors, and that ignoring a potential dependence among the error terms biases the posterior variance of the model parameters (Gelman et al. 1995). However, we believe that this bias is somehow mitigated in that we are not drawing inference on β, but simply identifying associations between X and Y variables. In addition, this assumption allows us to gain in (algorithmic) simplicity and efficacy. If we were to allow for correlation among the error terms and specify $\varepsilon_j \sim N(0, \Sigma)$ as in BVF, it would not be possible to integrate out the regression coefficients, for the prior covariance of the β could not be related to Σ (unlike BVF, where instead $B_{p\times q} \sim \mathcal{N}(B_0, H_{p\times p} \otimes \Sigma_{q\times q})$). Accordingly, updating of the regression coefficients would be required at each MCMC iteration and an appropriate reallocation scheme for these parameters would need to be defined when splitting and merging components, with a consequent complication of the algorithm. Furthermore, by taking the noise terms among the outcome variables to be independent, we are able to circumvent the high-dimensionality problem and convert the situation into one with an effective sample size equal to $N \cdot n_k$ in each component k, where N is the true sample size and n_k the number of outcomes in that component, as noted by CGM and Professor Li.

CGM noted that BVF have to estimate many more regression coefficients than we do when assessing variables in a component. This is true and it is exactly what we are avoiding by exploiting the cluster structure in the data. Outcomes are allocated to the same component because of their identical dependence on the same set of covariates, thus a single m_k-vector β is used rather than an $m_k \times n_k$ matrix of regression coefficients. On a similar note, Li wonders about the possibility of clustering response variables affected by the same predictors with different regression coefficients. In our current formulation,

[1]Department of Public Health, Weill Cornell Medical College, New York, NY, mailto:stm2013@med.cornell.edu

[2]Department of Mathematics, Georgetown University, Washington, DC, mailto:mgt26@georgetown.edu

 DOI:10.1214/09-BA416REJ