Bayesian Analysis (2011) **6**, Number 2, pp. 205–208

Rejoinder

Peter D. Hoff*

Abstract. I thank the editor for the opportunity to expand upon the paper, and I thank the discussants for their insightful comments. In this rejoinder I elaborate on some of the topics from the discussion: the appropriateness of separable covariance models for array-valued data, the role of priors and penalties on estimation and the limiting nature of array valued data.

The article and discussion provide several examples of array-valued and matrix-valued data, including relational (network) data, space-time and imaging data. Additionally, in many statistical models the parameters themselves are arrays, even though the data are not. For example, consider a three-factor experiment or study in which the levels of the factors are indexed by the sets $\mathcal{I}, \mathcal{J}, \mathcal{K}$. Letting $y_{i,j,k,l}$ be the measurement on the lth subject with levels i, j and k of the three factors, the data itself may not be an array as the number of subjects per factor combination may vary, but the unknown cell means $\{\mu_{i,j,k} : i \in \mathcal{I}, j \in \mathcal{J}, k \in \mathcal{K}\}$ constitute an array of dimension $|\mathcal{I}| \times |\mathcal{J}| \times |\mathcal{K}|$.

It is often desirable to estimate or account for patterns of dependence or similarity among objects in the index sets of such arrays. The article provides some computational tools for doing so, by relating the multilinear Tucker product to a class of multivariate normal distributions with separable covariance structure. A separable covariance structure is a "reduced model", in the sense that not all covariance matrices are separable. In what situations is such a model restriction justifiable? What are the alternatives?

Lopes expresses some concern that separability might not be an appropriate assumption for space-time data. Indeed, Stein (2005) makes a convincing argument against using separable covariance matrices for such applications. For space-time data, however, we can often rely on some degree of smoothness or continuity. Smoothness in time and space allows us to build rich but relatively parsimonious dependence models based on a small number of parameters that describe spatial or temporal correlation functions. Thus in the space-time domain, there are a large number of non-separable alternatives to modeling dependence patterns. Even so, in some situations separability may still be useful: Genton (2007) argues that separable approximations to non-separable covariance matrices can be useful for some inferential tasks. Additionally, judicious combinations of separable and non-separable structure lead to flexible models as in Lopes et al. (2008). Another way to use separable covariance structure for non-separable covariance estimation is in prior specification: Consider a non-separable covariance matrix $\Sigma = \text{Cov}[\text{vec}(\mathbf{Y})]$, where $\mathbf{Y}$ is a multiway array. A hierarchical prior for Σ could be of the form $\Sigma^{-1} \sim \text{Wishart}(\nu_0, c \times \Sigma_K \otimes \cdots \otimes \Sigma_1)$, with the Σ_k's also having inverse-Wishart priors. This centers the prior for Σ around a separable value, but Σ is non-separable with probability one.

*Department of Statistics, University of Washington, Seattle, WA, mailto:pdhoff@uw.edu

 DOI:10.1214/11-BA606REJ

Outside the domain of space-time data, alternatives to separable covariance structure are few. In many situations the elements of the index sets of a data array are categorical, such as commodity type or country, as in the example in the article. In these cases there may be no natural ordering to the indices, and no natural embedding of them in a Euclidean space. How then should we model the covariance among the elements of the array? An unrestricted covariance model requires a separate covariance parameter for every pair of elements of the array. Even for moderately sized arrays, this may lead to an unreasonably large number of parameters to estimate. For example, if $\mathbf{Y}$ is an $m_1 \times m_2 \times m_3$ random array, then $\text{vec}(\mathbf{Y})$ is a random vector of length $m_1 m_2 m_3$, and so a sample size of $n \geq m_1 m_2 m_3$ replications is required in order to obtain a full-rank MLE of $\text{Cov}[\text{vec}(\mathbf{Y})]$. This is a problem for the trade data considered in the article, and for the matrix-variate data considered in Allen and Tibshirani (2010). Both articles consider datasets for which the sample size is essentially 1, in the sense that there are no independent replications of the matrices or arrays.

In such cases, a simpler, restricted covariance model may be desirable. A separable covariance model can be viewed as somewhat analogous to mean modeling in factorial designs: The classical additive-effects model obtained from an ANOVA decomposition represents the mean in a particular cell as $\text{E}[Y_{i,j,k}] = \mu + a_i + b_j + c_k$. Similarly, the separable covariance model is log-additive,

$$\log \text{Cov}[Y_{i_1,j_1,k_1}, Y_{i_2,j_2,k_2}] = \log \sigma^{(1)}_{i_1,i_2} + \log \sigma^{(2)}_{j_1,j_2} + \log \sigma^{(3)}_{k_1,k_2},$$

making the separable model seem like a natural candidate as a "simple" covariance model when data are spare. Unfortunately, even for this greatly restricted covariance model, the sample size is frequently insufficient to estimate an MLE. In the matrix case with one replication, even without estimating a mean, the likelihood for the matrix normal model is generally unbounded, and no unique estimator exists (although strangely, if the row dimension is equal to the column dimension then the likelihood is bounded, but the MLE is not unique).

One way around the issue of unbounded likelihoods is to take a Bayesian approach to estimation, thereby penalizing certain values of the covariance matrices. In the matrix case, Carvalho and West (2007) penalize complexity by using a prior which allows for zeros in the precision matrices, and Allen and Tibshirani (2010) penalize complexity via the L^1 or L^2 norms on the precision matrices. In the current article regarding the array normal model, I have discussed the use of standard inverse-Wishart priors for the covariance matrices, although priors analogous to those used by Carvalho and West (2007) and Allen and Tibshirani (2010) in the matrix case would be straightforward to implement. In particular, the L^2 penalty and the inverse-Wishart prior (with expected value proportional to the identity) both give equivariant estimates. The former is equivalent to using independent half-normal prior distributions for the eigenvalues of the precision matrix, and the latter corresponds to independent gamma priors. It would certainly be interesting to explore the L^2 penalty further, and see if the relationship between the SVD and the penalized estimate in the matrix case extends to the array case, as conjectured by Allen.

Prior distributions (or penalties) can provide a unique estimator in situations where data alone do not. However, in such cases, any differentiation between parameter values that are likelihood-equivalent is a reflection purely of the prior. Ideally, the prior reflects some knowledge about the actual covariance parameters or data generating mechanism. In the absence of such knowledge the choice of prior or penalty becomes (ironically) more subjective: Often a certain type of prior is chosen because it makes the posterior mode have an aesthetically pleasing (e.g. sparse) form. It might be better to express such preferences for certain types of parameter estimates via loss functions, in the context of a decision problem regarding the summarization or reporting of results.

References

Allen, G. I. and Tibshirani, R. (2010). "Transposable Regularized Covariance Models with an Application to Missing Data Imputation." *The Annals of Applied Statistics*, 4(2): 764–790. 206

Carvalho, C. M. and West, M. (2007). "Dynamic Matrix-variate Graphical Models." *Bayesian Analysis*, 2(1): 69–98. 206

Genton, M. G. (2007). "Separable approximations of space-time covariance matrices." *Environmetrics*, 18(7): 681–695.
URL http://dx.doi.org/10.1002/env.854 205

Lopes, H. F., Salazar, E., and Gamerman, D. (2008). "Spatial Dynamic Factor Analysis." *Bayesian Analysis*, 3(4): 759–792. 205

Stein, M. L. (2005). "Space-time covariance functions." *J. Amer. Statist. Assoc.*, 100(469): 310–321.
URL http://dx.doi.org/10.1198/016214504000000854 205