Bayesian Analysis (2010) **5**, Number 3, pp. 461–464

Rejoinder

Ioanna Manolopoulou*, Cliburn Chan† and Mike West‡

We thank the discussants, Fabio Rigat and Nick Whiteley, for their insightful and positive comments. They suggest a number of potential directions for extension of the work and raise connections with other research. We address the points they raise in connection with broader modeling and communication considerations, followed by specific aspects and details of computational strategy.

1 Modeling and Communication

Discussion comments on general questions of applied statistical modeling relate to the need for attention to a balance between contextual/applied interests and statistical modeling refinements motivated by an application. A good deal of time and effort in collaborations and applied work is spent on communication of the relevance and roles of complex Bayesian models to non-statistical disciplinary scientists.

The specific setting here is that of non-parametric Bayesian mixture models. These models are nowadays standard and widely accepted by statistical and machine learning communities. Their demonstrated success in applications in many areas in the last decade or so has done much to foster understanding and appreciation among disciplinary scientists. In our current applied context of cell subtype characterisation in flow cytometry studies, mixture models are established (e.g. Chan et al. 2008; Pyne et al. 2009). For the purposes of communication we have promoted non-parametric DP mixtures as really just direct extensions of standard mixtures that allow for uncertainty about the (practically effective) number of components. That is easily communicated and the remaining technical aspect of note is just the use of effectively standard class of priors over component parameters. A substantial practical modeling bridge in our work in these applications is the clustering of subsets of Gaussian DP mixture components according to concentration around inferred locals modes in the distribution, and putative interpretation of some of these clustered components as defining (resulting non-Gaussian) subpopulations of biological interest.

On the specific question of inference on the concentration hyperparameter α, which again has been standard in the literature since the early 1990s, we note that this hierarchical model specification has the usual goals and attributes of inducing a degree of robustness while incurring negligible additional computational (Escobar and West 1995; Ishwaran and James 2002). Although the number of components has no immediate biological interpretation (other than as a gross upper bound on the number of subtypes) and so the role of α in its impact on the number of components is only of technical

*Department of Statistical Science, Duke University, Durham, NC, mailto:im30@stat.duke.edu
†Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, mailto:cliburn.chan@duke.edu
‡Department of Statistical Science, Duke University, Durham, NC, mailto:mw@stat.duke.edu

 DOI:10.1214/10-BA517REJ