Comment on Article by Celeux et al.

Bradley P. Carlin*
University of Minnesota

Congratulations to Drs. Celeux, Forbes, Robert, and Titterington (henceforth CFRT) on a stimulating, important, and much-awaited paper. At least one of the authors' interest in the problem dates to the discussion of the original DIC paper (Spiegelhalter et al., 2002) by DeIorio and Robert (2002), where it was shown that the effective sample size p_D as originally defined can behave badly (in particular, it can be negative, a clearly nonsensical result) for classes of missing data models, especially mixture models. CFRT suggest that the root problem is poor identifiability of the model parameters, leading to their posterior mean being a poor estimate, which in turn hurts p_D. In the context of the issues brought about by the missing data $\mathbf{Z}$ (which can be treated symmetrically with the parameters θ, asymmetrically, or integrated out entirely), the authors devlop a large number of new p_D and DIC possibilities, which they classify as "observed," "complete," and "conditional." After defining these new DICs, the authors investigate them in two settings (a simple random effects setting and a much more challenging mixture setting) and comparing them in two data examples, one real (the classic "galaxy data" often used to illustrate mixture modeling) and one simulated.

I like the authors' cleverness in coming up with alternate DICs that solve certain problems. The justification for DIC in the original paper, as well as its subsequent validation in practice, is essentially only within the exponential family, so calling attention to defects and proposing remedies outside this family is important work. However, the approach is fairly ad hoc, and also leads us into "casework" (comparing a large number of competitors in a potentially large number of problem settings) which may not be a tenable strategy in the long run. DIC is not derivation free, but the authors' work does not refer at all to any derivation, nor to any subsequent interpretation of model complexity. As a result, it's sometimes hard to get a good feel for why a certain version works or doesn't work except through case-specific exemplification.

Still, the approach pays clear dividends. The authors note that p_{D_3} does not perform well in Table 1, and it also gets the "wrong answer" $(1 - \log 2)$ in Section 4.1, a setting where an effective model size statistic that works by counting degrees of freedom (DF) should obtain 1, since the univariate grand mean θ is the only unknown parameter in this simple model. While every reader will have his or her favorite, my own preference is for p_{D_7} since it treats $\mathbf{Z}$ and θ symmetrically (though maybe that's just because I am not a mixture modeler). Indeed, this is essentially the approach adopted by the `WinBUGS` DIC tool, except of course for the means replacing CFRT's modes. Thus the performance of this approach here may thus inform about the appropriateness of its use more broadly, as has occurred with the rise of `WinBUGS` as a tool for routine Bayesian data analysis, and hence its seductively easy-to-use DIC tool for model choice as well.

---

*Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, http://www.biostat.umn.edu/~brad