Bayesian Analysis (2006) **1**, Number 4, pp. 699–700

Comment on Article by Celeux et al.

Angelika van der Linde*,

The authors point to several problems occurring if $DIC = 2\overline{D(\theta)} - D(\overline{\theta})$ is applied to missing data models: (i) the focus of the model depends on whether the missing data $\mathbf{z}$ are considered as observations or as parameters or integrated out; (ii) the parameter θ may be non-identifiable; (iii) $\overline{\theta}$ can be a poor estimator of θ. They then investigate modifications of DIC mainly based on estimates other than $\overline{\theta}$ in $D(\overline{\theta})$ respectively $p_D = \overline{D(\theta)} - D(\overline{\theta})$ and/or on integration w.r.t. the missing data.

Comment (1). In (Spiegelhalter et al. (2002)) DIC was derived as an approximation to the posterior predictive target $-2E_{\mathbf{X}_{rep}}[\log f(\mathbf{X}_{rep}|\overline{\theta})|\mathbf{x}]$ where $\mathbf{X}_{rep}$ denotes replicate observations from the same experiment. It depends on the posterior mean $\overline{\theta}$ which may not universally be a good choice. Using this or a similar target as an expected loss in model comparison a specification of variables to be predicted and of parameters is presumed. The choice of the loss function then defines the purpose of a model independently of the sampling scheme, although the sampling scheme affects and possibly limits the evaluation of the expected loss. From this perspective the performances of predictive criteria for $\mathbf{X} = \mathbf{Y}$ only or for $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$ jointly are not comparable: if the purpose of a model is set to be predictive, should not the performance of the criterion be evaluated in terms of prediction rather than the detection of the 'true' (for example mixture) model? Furthermore, often, as in the derivation of DIC, the target is expanded and the resulting terms approximated. Modifications of the approximations as tried in the paper may reversely induce different targets; the inconsistency of DIC_5 is particularly worrying from this point of view. More generally, it should be examined if the new criteria $DIC_2 - DIC_8$, some of which look quite sensible, correspond to well justified losses and thus turn out to be valid from a decision theoretic point of view.

Comment (2). Loss functions similar to the one underlying DIC were analyzed in (van der Linde (2005)), where it was shown that the decomposition of the predictive target into terms of model fit and model complexity arises as a variant of the decomposition of (posterior predictive, that is) marginal entropy into conditional entropy and mutual information. It might therefore be argued that the divergence between future (replicate) observations $\mathbf{X}_{rep}$ and posterior parameters θ reasonably describes model complexity independently of any parameter estimate. If the sampling density $f(\cdot|\theta)$ belongs to an exponential family, this divergence can be represented using the posterior mean $\overline{\theta}$ and p_D be justified as an appropriate estimator (for details see (van der Linde (2004)). Clearly, mixture models do not fit into this framework and p_D fails as an estimator although model complexity is still well defined. In contrast, it is not clear what the 'complexity of a predictive density' (referred to at the end of section 4.1) means.

Comment (3). A major achievement initiated with the introduction of DIC and p_D

*University of Bremen, Germany, http://www.math.uni-bremen.de/~avdl/

 ba0004

is the formal quantification of the reduction of model complexity due to the information in a prior. Further analysis of model complexity in terms of mutual information reveals that model complexity is a symmetric (dual) concept of variables $\mathbf{X}_{rep}$ and random parameters θ. Hence, also restrictions inherent in the sampling distribution can reduce model complexity. For instance, in a linear regression model (with known variance) which is not of full rank p, say, p_D with more and more diffuse priors converges to the reduced rank $q < p$. Similarly, in the example of Scottish lip cancer discussed by Spiegelhalter et al. (2002) two out of 56 observations happened to be non-informative. This was correctly reflected by $p_D \approx 54$, indicating only 54 identifiable parameters under a vague prior (in the plots presented by Brooks (2002) in the discussion). Thus lack of identifiability of parameters as a feature of a (mixture) model shows up in diminishing model complexity. This is to be distinguished from a poor performance of an estimator of model complexity (like that of p_D for mixture models). However, even under lack of identifiability of parameters, an estimator of model complexity should not take negative values (as p_D does for mixture models where therefore it is inappropriate).

Concluding remarks. Although in my comments I mainly addressed some views presented in the paper by Celeux et al. which I do not share, I appreciate the paper as well. Already in the discussion of (Spiegelhalter et al. (2002)) some of the authors correctly pointed out that an unmodified DIC may be tied to exponential families, and this objection is elaborated in the present article. The authors also set up a challenge in insisting on a posterior predictive target based on variables that cannot be observed, thus turning the evaluation of the target into a particularly difficult problem. Posterior predictive assessment of missing data models requires further research, and in this paper stimulating arguments and proposals are contributed to the discussion.

References

Brooks, S. (2002). "Discussion of Spiegelhalter et al." *Journal of the Royal Statistical Society, Series B*, 64: 617. 700

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). "Bayesian measures of model complexity and fit." *Journal of the Royal Statistical Society, Series B*, 64: 583–639. With discussion. 699, 700

van der Linde, A. (2004). "On the association between a random parameter and an observable." *Test*, 13: 85–111. 699

— (2005). "DIC in variable selection." *statistica neerlandica*, 59: 45–56. 699