Christian P. Robert (CEREMADE, Université Paris Dauphine) and D.M. Titterington (University of Glasgow)
A question that arises regarding this thought-challenging paper was actually raised in the discussion of the (read) paper of Aitkin (1991), namely that the data seems to be used *twice* in the construction of p_D. Indeed, y is used a first time to produce the posterior distribution, $\pi(\theta|y)$, and the associated estimate, $\tilde{\theta}(y)$. The (Bayesian) deviance criterion then computes the posterior expectation of the *observed* likelihood $p(y|\theta)$,

$$\int \log p(y|\theta)\pi(d\theta|y) \propto \int \log p(y|\theta)p(y|\theta)\pi(d\theta)\,,$$

and thus uses y again, similarly to Aitkin's posterior Bayes factor

$$\int p(y|\theta)\pi(d\theta|y)\,.$$

This repeated use of y would appear to be a potential factor for overfitting.

It thus seems more pertinent [within the Bayesian paradigm] to follow an integrated approach along the lines of the posterior *expected* deviance of §6.2,

$$\int \mathbb{E}_{Y|\theta}\left[-2\,\log p(Y|\theta) + 2\,\log f(Y)\right]\pi(d\theta|y)$$

because this quantity would be strongly related to the posterior *expected* loss defined by the logarithmic deviance,

$$d(\theta,\tilde{\theta}) = \mathbb{E}_{Y|\theta}\left[\log p(Y|\theta) - \log p(Y|\tilde{\theta})\right]\,,$$

advocated in Robert (1996) and Dupuis and Robert (2002) as an intrinsic loss adequate for model fitting. In fact, the connection between p_D, the DIC criterion, and the logarithmic deviance would suggest to use this loss $d(\theta,\tilde{\theta})$ to compute the estimate plugged in p_D as the intrinsic Bayes estimator

$$\begin{aligned}\theta^\pi(y) &= \arg\min_{\tilde{\theta}} \mathbb{E}_{\theta|y}\left[\mathbb{E}_{Y|\theta}\left[\log p(Y|\theta) - \log p(Y|\tilde{\theta})\right]\right] \\ &= \arg\max \mathbb{E}_{Y|y}\left[p(Y|\tilde{\theta})\right]\end{aligned}$$

where the last expectation is computed under the predictive distribution. Not only does this make sense because of the connection aforementioned, but it also provides an estimator that is completely invariant to reparameterisation, and thus avoids the possibly difficult choice of the parameterisation of the problem. (See Celeux, Hurn and Robert, 2000, for an illustration in the setup of mixtures.)

References

Aitkin, M. (1991) Posterior Bayes factors (with discussion). *J. Royal Statistical Society* Series B **53**, 111–142.

Celeux, G., Hurn, M. and Robert, C.P. (2000) Computational and inferential difficulties with mixtures posterior distribution *J. American Statistical Society* **95**(3), 957–979.

Dupuis, J. and Robert, C.P. (2002) Model choice in qualitative regression models. *J. Statistical Planning and Inference* (to appear).

Robert, C.P. (1996b) Intrinsic loss functions. *Theory and Decision* **40** (2), 191–214.