Maria DeIorio[1](University of Oxford) and Christian P. Robert (CEREMADE, Université Paris Dauphine)

Amidst the wide scope of possible extensions of their paper, the authors mention the case of mixtures

$$\sum_{j=1}^{k} p_j f(x \mid \theta_j)\,,$$

which is quite interesting, as it illustrates the versatility of the DIC criterion under different representations of the same model.

In this setup, if the p_j's are known, the associated *completed* likelihood is

$$L(\boldsymbol{\theta} \mid (x_1, z_1), \ldots, (x_n, z_n)) \propto \prod_{i=1}^{n} f(x_i \mid \theta_{z_i}) = \prod_{j=1}^{k} \prod_{i:z_i=j} f(x_i \mid \theta_j)\,. \tag{1}$$

Thefore, conditional on the latent variables $\boldsymbol{z} = (z_1, \ldots, z_n)$, and setting the saturated deviance $f(x)$ to 1, define

$$[\mathrm{DIC} \mid \boldsymbol{z}] = \sum_{j=1}^{k} \sum_{i:z_i=j} \left[-4\mathbb{E}\left\{\log\left(f(x_i \mid \theta_j)\right) \mid \boldsymbol{x}, \boldsymbol{z}\right\} + 2\log\left(f(x_i \mid \widehat{\theta}_j)\right)\right]$$

where $\widehat{\theta}_j = \mathbb{E}(\theta_j \mid \boldsymbol{x}, \boldsymbol{z})$ (under proper identifiability constraints, see Celeux *et al.*, 2000). The *integrated* DIC is then

$$\mathrm{DIC}_1 = \sum_{\boldsymbol{z} \in \mathcal{Z}} [\mathrm{DIC} \mid \boldsymbol{z}] \Pr(\boldsymbol{z} \mid \boldsymbol{x})\,,$$

where $\Pr(\boldsymbol{z} \mid \boldsymbol{x})$ can be approximated (Casella *et al.*, 1999).

A second possibility is the *observed* DIC, DIC_2, based on the observed likelihood, which does not use the latent variables $\boldsymbol{z}$. (We notice the strong dependence of the DIC criterion on the choice of the saturated function f and the corresponding lack of clear guidance outside exponential families. For instance, if $f(x_i)$ goes from the marginal density to the extreme alternative where both θ_1 and θ_2 are set equal to x_i, DIC_2 goes from -31.71 to 166.6 in the following example.)

A third possibility is the *full* DIC, DIC_3, based on the completed likelihood (1) when it incorporates $\boldsymbol{z}$ as a additional parameter, in which case the saturated deviance could be the normal standardized deviance, although we still use $f(x) = 1$ for comparison purposes.

The three above possibilities lead to rather different figures, as shown by Table 1 for the simulated dataset in Figure 1; this table exhibits in addition a lack of clear domination of the mixture ($k = 2$) versus the normal distribution ($k = 1$) (first column), except when $\boldsymbol{z}$ is set to its true value (second column) or estimated (last column). Note that, for the full DIC, p_D is far from 102; this may be due to the fact that, for some combinations of $\boldsymbol{z}$, the likelihood is the same. (This also relates to the fact that $\boldsymbol{z}$ is not a parameter in the classical sense.)

[1]Work partially supported by EU TMR network ERB–FMRX–CT96–0095.

	normal ($k=1$)	complete [DIC $\mid \boldsymbol{z}$]	integrated DIC_1	observed DIC_2	full DIC_3
DIC	465.1	413.5	462.6	457.6	447.4
ΔDIC	–	-51.6	-2.5	-7.5	-17.6
p_D	0.99	1.96	2.27	1.98	28.06

Table 1: Comparison of the three different criteria *[Columns 3–5]* for a simulated sample of 100 observations from $0.5\mathcal{N}(5,1.5)+0.5\mathcal{N}(7.5,8)$ with a conjugate prior $\theta_1 \sim \mathcal{N}(4,5)$ and $\theta_2 \sim \mathcal{N}(8,5)$, and of the DIC criterion based on the true complete sample $(\boldsymbol{x},\boldsymbol{z})$ *[Column 2]* and the DIC criterion for the single component normal model (with a $\mathcal{N}(6,5)$ prior and a variance set of 6.07) *[Column 1]*.

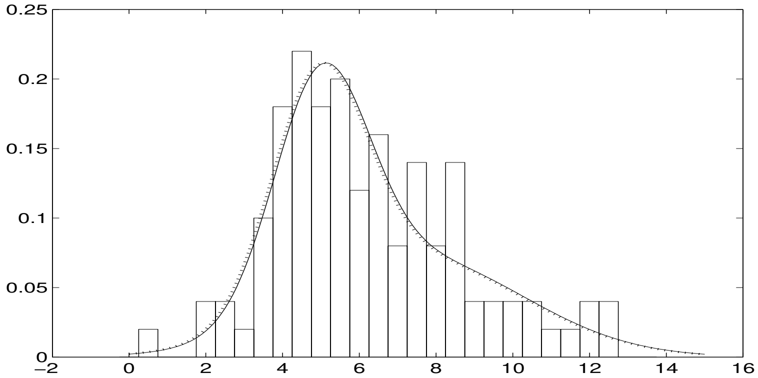

Figure 1: Histogram of the simulated dataset and true density.

References

Casella, G., Robert, C.P., and Wells, M.T. (2000) Mixture Models, Latent Variables and Partitioned Importance Sampling. Tech. report, CREST, Paris.

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.