Comment on Article by Wade and Ghahramani
Julyan Arbel, Riccardo Corradin, Michal Lewandowski

To cite this version:
Julyan Arbel, Riccardo Corradin, Michal Lewandowski. Comment on Article by Wade and Ghahramani. 2018. <hal-01950655>

HAL Id: hal-01950655
https://hal.archives-ouvertes.fr/hal-01950655
Submitted on 21 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Comment on Article by Wade and Ghahramani

Julyan Arbel∗, Riccardo Corradin† and Michal Lewandowski∗

Abstract. We propose a simulation study to emphasise the difference between Variation of Information and Binder’s loss functions in terms of number of clusters estimated by means of (1) the use of the MCMC output only and (2) a “greedy” method.

Wade and Ghahramani’s paper is a very neat contribution to Bayesian cluster analysis in at least two respects: (i) by formalizing cluster credible coverage via Hasse diagrams, and (ii) by recasting the problem in a decision theory framework, with tangible improvements brought by the Variation of Information (VI) loss function (Meilă, 2007) over Binder’s (Binder, 1978; Dahl, 2006).

We propose a simulation study implementing two algorithms provided by Wade and Ghahramani’s package mcclust.ext for finding the argument minimizing the posterior expected loss: (1) the draw algorithm, which restricts the minimization problem to the MCMC output, and (2) the greedy algorithm, which is more reliable as it also scans the neighbouring clusters of the MCMC output, but with a larger computational cost. While increasing the sample size, we point out the radically different behavior of the number of clusters estimated under VI and Binder, especially with the greedy algorithm.

Our simulation study is based on the same data generation as in the first example of Section 6.1 in Wade and Ghahramani (2017): a mixture of four Gaussian distributions equally weighted with means (±2, ±2) and identity covariance matrix. We estimated the model using a marginal approach provided by BNPmix1 R package. We synthesised the output with mcclust.ext package.2 The Dirichlet process mixture model was estimated with mass parameter fixed to 1, and by specifying an independent base measure on locations and scales, with a 0-vector prior mean for the location component and an identity matrix prior mean for the scale component (25,000 iterations with 5,000 burn-in period). We considered four different sample sizes \( n = \{20, 40, 100, 300\} \).

The results are shown in Figure 1. With the draw algorithm, the cluster estimates under both losses are quite close in terms of number of clusters. In contrast, the greedy algorithm leads to cluster estimates obtained via Binder’s loss function with excessive size, while that obtained via VI remains coherent with the number of components of the model (four).

Similarly to the authors’ finding, ours’ indicates that Binder’s loss function exhibits an undesirable property of overestimating the number of clusters (Miller and Harrison, 2013, 2014). Variation of Information tends to lessen this problem. As alluded to by the

∗ Univ. Grenoble Alpes, Inria, CNRS, LJK, 38000 Grenoble, France. julyan.arbel@inria.fr; michal.lewandowski@inria.fr
† DISMEQ, University of Milano Bicocca, 20126 Milano MI, Italy. riccardo.corradin@unimib.it
1Package available at https://github.com/rcorradin/BNPmix, can be installed via devtools.

© 0000 International Society for Bayesian Analysis DOI: 0000
Figure 1: Size of the cluster estimate under VI (yellow line) and Binder (green light).
Left: draw algorithm. Right: greedy algorithm.

authors, a theoretical study of the asymptotic behavior of the VI estimator would be very timely. Especially in light of the recent contribution by Rajkowski (2016) about the asymptotic behavior of the cluster estimator under the $0−1$ loss (MAP estimator).

References