FINAL REPORT : WORKSHOP ON MIXTURE ESTIMATION AND APPLICATIONS

International Centre for Mathematical Sciences (ICMS),
Edinburgh, March 3–5, 2010

Organisers: K.L. Mengersen (Queensland University of Technology, Brisbane), C.P. Robert, Université Paris Dauphine) and D.M. Titterington (University of Glasgow)

1 Substantial deviations from the original proposal

So far as the scientific content and programme-structure were concerned, few changes were made, the most substantial being the replacement of a section about 'social-science applications' by one about 'financial applications' and the substitution of a few of the named speakers by others, mainly because of availability. The other major change was to increase the level of participation from the original number of about 35 to the eventual number of 70. This change was implemented because of the obviously high level of interest in the topic and the fact that the new ICMS facilities would comfortably cope with such numbers. The extra participants were either pre-selected by the organisers or identified from an email trawl for expressions of interest, followed by vetting by the organisers to make sure that they would be likely to benefit from their participation; almost all such applications were successful. However, the number of talks was not increased.

2 Short description of the meeting

Statistical mixture distributions are used to model scenarios in which certain variables are measured but a categorical variable is missing. For example, although clinical data on a patient may be available their disease category may not be, and this adds significant degrees of complication to the statistical analysis. The above situation characterises the simplest mixture-type scenario; variations include, among others, hidden Markov models, in which the missing variable follows a Markov chain model, and latent structure models, in which the missing variable or variables represent model-enriching devices rather than real physical entities. In the planning of the workshop the term 'mixture' was taken to include these and other variations along with the simple mixture. The motivating factors for the workshop were that research on inference and computational techniques for mixture-type models is currently experiencing major advances and that simultaneously the application of mixture modelling to many fields in science and elsewhere has never been so rich. We thus assembled top players, from statistics and machine learning, in both methodological research and applied inference at this fertile interface. The methodological component involved both Bayesian and non-Bayesian contributions, and biology and finance featured strongly among the application areas that were covered.

3 Workshop report

3.1 Structure of the workshop

As mentioned in Section 1, the workshop largely followed the structure proposed originally.

It opened with a lively Bayesian 'debate', with two main speakers (Sylvia Frühwirth-Schnatter discussing the nagging problem of label-switching and John Geweke raising the issue of intrepretability of mixture models) followed by prepared discussion contributions from Gilles Celeux, Agostino Nobile and Christian Robert. After lunch there was a session on general methodology, with Geoff McLachlan looking at the highly topical

case of high-dimensional data and Murray Aitkin exploring another perennial problem, that of identifying the number of components in a mixture model. This led neatly into a session about asymptotic theory, with a more theoretical talk by Jiahua Chen on the number-of-components problem and a presentation by Bruce Linsday about the issue of high dimension. The day closed with a poster session, a well-received and highly-educational public lecture by Kerry Mengersen and a wine reception in the Chapterhouse.

Much of the second day was devoted to presentations by machine-learning speakers, i.e. researchers doing the same sort of research as other participants but based in computer science/informatics environments rather than in statistics groups. First came a special lecture on Bayesian nonparametrics by Michael Jordan (UC Berkeley), one of the first holders of a joint appointment in Statistics and Computer Science. Further presentations came from Yee Whye Teh (on hierarchical clustering), Iain Murray (on ways of sampling latent variable models), Katherine Heller (on Dirichlet process models), Chris Williams (on binary latent trees) and Mark Girolami (on appliactions in Raman resonance spectroscopy). There was then a session on Bayesian computation with talks by statisticians Christophe Andrieu and Paul Fearnhead, following which participants migrated to the Informatics Forum where Michael Jordan presented a Distinguished Lecture to a huge, standing-room-only audience.

The third day was devoted mainly to applications, starting off with Claire Alston's talk about animal husbandry and Brendan Murphy's work on social network data. These talks sandwiched Olivier Cappé's methodological presentation about online estimation. The final two sessions concerned financial applications (talks by Robert Kohn and Richard Gerlach) and biological applications (talks by Kim-Anh Do, Michael Newton, Peter Müller and Chris Holmes).

As mentioned in the original proposal, a volume of papers by the workshop speakers, edited by the organisers, is currently under development for Wiley. The volume is likely to comprise 15 chapters.

3.2 Highlights

Highly successful general features of the workshop were the mix of Bayesian/non-Bayesian theory, of mainstream statistics with machine learning and of theory with varied applications. Judging from our own perceptions and the feedback summary (see below), particular highlights were the initial Bayesian debate, and the special lecture by Michael Jordan.

3.3 Involvement of participants and new collaborations

The talks generally stimulated lively discussion and there was clear evidence of considerable interaction in the breaks. We are not aware of any specific new collaborations but the feedback summary, again, suggests that this is indeed developing in quite a few cases.

3.4 Funding

We understand that the meeting kept to its budget, as awarded by the ICMS. This was supplemented by a workshop grant of £900 from the Royal Statistical Society which provided partial support for a number of young researchers.

3.5 Feedback from participants

As organisers we were sent a summary of the responses from 33 (i.e. approximately 50%) of the participants. The feedback was, gratifyingly, almost uniformly positive and it is hard to identify any substantial criticism;

any isolated lukewarm comment seemed to be cancelled out by several enthusiastic responses about the same issue.

3.6 Concluding comments

Overall, the organisers believe that the workshop was highly successful, their belief being reinforced by the above-mentioned written feedback and also by the verbal remarks of departing participants. This success resulted from the preponderance of very good and distinguished speakers, the good (chilly but bright and dry!) weather and, certainly not least, the excellence of the ICMS location and organisation. The meeting had an inauspicious beginning thanks to a power cut resulting from a problem outwith the ICMS but the calm and positive attitude of the ICMS staff meant that this had virtually no impact on the proceedings.