DISCUSSION OF

Bayesian Analysis of Single Molecule Experimental Data

by S.C. Kou, X.S. Xie and J.S. Liu

Mark Steel, University of Warwick

The authors are to be congratulated for a very interesting and stimulating paper which builds on an emerging literature in biochemistry on single-molecule experiments. I am afraid that I have no real expertise in bioinformatics, and, as a consequence, I will have to limit this discussion to rather more standard issues in Bayesian statistics.

Bayesian inference with latent processes. The authors propose models that depend on unobserved stochastic processes, and that unsurprisingly complicates the computational analysis. There is a growing literature (for example in finance) on Bayesian inference with latent stochastic processes, and it would be interesting, in my view, to compare the approach used here with some of the alternative proposals in the literature.

A first issue in this context is that the authors use component-wise updating of the diffusion chains. There is some mention of "blocking" the sampler on the diffusions in Section 3.2 and a promise of coming back to these block-wise moves in a later section, but that presumably refers to the group move, rather than a "real" blocking move, in the sense of *e.g.* Liu *et al.* (1994) or Shephard and Pitt (1997). Would the more usual types of blocking help in this setting?

A second, and possibly more critical, issue is to be how to deal with the overconditioning (*i.e.* the high correlation between the latent variables and the parameters). This occurs in the case of the $x(t)$ process introduced in the continuous diffusive model. The proposed solution is the use of a scale transformation which jointly updates the process and its diffusion parameter. One possible problem with such a scale transformation is that the whole process is scaled at once, and such moves could be hard to accept if the likelihood is particularly informative about certain parts of the process. Here it seems to work, possibly as a consequence of the "generalized Gibbs" step of Liu and Sabatti (2000), which brings the likelihood into play. This scaling is somewhat reminiscent of the approach of Roberts and Stramer (2001) who reparameterise the "missing data" in a partially observed diffusion model (although they use a slightly different transformation to achieve continuity of the sample paths at the observed points). This can be given an interpretation as a so-called "non-centred parameterisation" as explained in Papaspiliopoulos *et al.* (2003). Basically, the non-centring resolves the situation where the augmented data contain much more information about the parameters than the observed data. Barndorff-Nielsen and Shephard (2001) introduce a class of stochastic volatility models where the (latent) volatility is modelled as an Ornstein-Uhlenbeck process, driven by a positive Lévy

process without Gaussian component. In the context of these models non-centring is also proposed to solve the overconditioning problem in Roberts *et al.* (2004). For the same type of models, Griffin and Steel (2003) propose a sampler where the volatility process is thinned in line with the proposed parameter values. The latter paper implements so-called "dependent thinning", where the jumps added to or deleted from the process tend to be relatively small. This is achieved by the use of the Ferguson-Class representation of Lévy jump processes, which allows us to focus on small jumps when proposing changes to the process. Thus, the new process will be "close" to the previous process, while also being compatible with the new value of the parameters. In fact, Roberts *et al.* (2004) mention that the latter approach can also be given an interpretation as a non-centred parameterisation. Even though the models used in the papers mentioned are somewhat different, I wonder if similar ideas could be used in the context of the present paper.

Prior issues. The authors mention in the introduction that "we have a rather in-depth knowledge (prior) on various parameters involved". Nevertheless, there is little discussion of prior elicitation or prior issues in general in the paper. In the simulated data examples a flat prior is used throughout. The authors do not make explicit what "flat" means, but if it means uniform on (the logarithms of) the parameters defined on $\Re_+$ that makes the prior improper and raises the familiar spectre of posterior existence. Does the posterior exist with the chosen prior? In addition, why would uniformity on the particular parameterisation used (rather than some other parameterisation) be a reasonable choice to reflect a lack of prior information (which is presumably what the authors intend)? Is there perhaps an invariance argument that can motivate this choice? Also, it would be useful to get a feeling for the importance of the prior assumptions, through, for example, a prior sensitivity analysis. Whereas data sets of typical sizes could swamp the prior in certain directions, I am not convinced this is necessarily the case for all prior dimensions. It would be good to know where to concentrate the effort in eliciting a prior for practical problems, and some ideas as to how such elicitation might proceed could be quite valuable for applied users of these methods.

Estimation of the Bayes factor. The identity in (5.1) used for the estimation of the Bayes factor is an interesting one, which, as the authors state, has not been used a great deal in the literature. I wonder if the reason might be the same as for the "harmonic mean" estimator, which constitutes the special case if the simpler model has no unknown parameters. Newton and Raftery (1994) comment that the harmonic mean does not, generally, satisfy a Gaussian central limit theorem and can be very unstable as a consequence. One would expect that such problems will be somewhat less severe for a likelihood ratio, but I wonder if stability of the resulting estimates is still an issue, especially if the extra parameters in the larger model involve a latent process.

Finally, it gives me great pleasure to second the vote of thanks.

Additional References

Barndorff-Nielsen, O.E., and Shephard, N. (2001), "Non-Gaussian OU Based Models and Some of Their Uses in Financial Economics," *Journal of the Royal Statistical Society*, B, 63, 167-241 (with discussion).

Griffin, J.E. and Steel, M.F.J. (2003), "Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility," with J. Griffin, mimeo, University of Warwick.

Liu, J.S., Wong, W.H. and Kong, A. (1994), "Covariance structure of the Gibbs sampler with applications to the comparison of estimators and augmentation schemes," *Biometrika*, 81, 27-40.

Papaspiliopoulos, O., Roberts, G.O. and Sköld, M. (2003), "Non-centered parameterisations for hierarchical models and data augmentation," in *Bayesian Statistics 7*, J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith and M. West, eds., Oxford University Press: Oxford, pp. 307-326.

Roberts, G.O., Papaspiliopoulos, O. and Dellaportas, P. (2004), "Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes," *Journal of the Royal Statistical Society*, B, 66, 369-393.

Roberts, G.O. and Stramer, O. (1994), "On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm," *Biometrika*, 88, 603-621.

Shephard, N. and Pitt, M.K. (1997), "Likelihood analysis of non-Gaussian measurement time series," *Biometrika*, 84, 653-667.