Bayesian Analysis (2011) **6**, Number 4, pp. 541–546

Rejoinder

Jason Wyse*, Nial Friel† and Håvard Rue‡

We would like to begin by expressing our gratitude to the discussants for their insightful comments on the work presented in this paper and possible further improvements which could be made to the methodology. Both discussants highlight strengths and weaknesses associated with the approaches developed in the paper which have given us much food for thought. We'll reply to the discussants in turn.

Fearnhead

Motivation for RFRs

One of the main outcomes of the paper is to widen the class of changepoint models which can be analyzed using recursive computing approaches (Yao (1984); Barry and Hartigan (1992); Liu and Lawrence (1999); Fearnhead (2006)). However, it is does not seem sensible to fit these classes of models, i.e. GMRFs modelling dependence, to a very small amount of data. Thus the recursive methods of the aforementioned articles can not be used in their raw form- there needs to be some mechanism to allow GMRF models to be fitted to at least reasonable sized segments. Dependency cannot be sensibly modelled using just a few data points. Fearnhead notes that the motivation for the RFRs in the paper is purely computational. We agree that there are distinct computational advantages to using RFRs, however we point out that the motivation for RFRs is two-fold; they give a reduction in computation and additionally enforce larger segments (with at least, say $g = 5$ data points) so that GMRFs can be used to model data dependence.

Asymptotics

Consideration of asymptotics would be a promising direction to proceed in developing the approaches proposed in this paper further. Hypothetically, such results would give

*Department of Statistical Science, University College London, London, UK, jason@stats.ucl.ac.uk
†School of Mathematical Sciences, University College Dublin, Dublin, Ireland, nial.friel@ucd.ie
†Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway, Havard.Rue@math.ntnu.no

 DOI:10.1214/11-BA620REJ

a more formal validation of the proposed methods. If $n \gg K$, RFRs should introduce a negligible error as noted by Fearnhead, and would be a sensible choice for a practitioner. Moreover $n \gg K$ would imply that a large value of g may be reasonable, with less fear of missing changepoints, hence making segments larger and any approximations using INLA more accurate. However, if the number of changepoints increases approximately linearly with the amount of data then it may be more difficult to make an intuitive choice of g. If it is too large, changepoints may be missed and having this value too small adds considerably to the computation, especially when n is large.

It may be possible to consider the case of modelling segment lengths as iid in the framework considered in the paper, although we have not explored this. Potentially, one could take the popular geometric prior on segment lengths (or some other member of the negative binomial family as considered in Fearnhead (2006)), but adjust for the situation where we only allow changepoints to occur in the reduced time index set by choosing the geometric parameter p_g taking g into consideration. This does not avoid the problem of a large g missing changepoints in the case where the number of changepoints increases approximately linearly as more data are collected. However, if we are considering classes of changepoint models where data are iid within a segment and the segment marginal likelihood may be computed exactly, even using a small value of g could reduce the computational effort for the recursions mentioned by Fearnhead which will ordinarily scale quadratically with n. On a rough inspection it appears that the computational demands of such a recursive scheme would be about $O(n_r^2)$ where $n_r = \lfloor n/g + 1 - \mathrm{I}(g = 1) \rfloor$ since the recursions will only be computed at n_r points in the data, where $\lfloor a \rfloor$ denotes the greatest integer less than a. The storage requirements would be $O(n_r)$. Choosing p_g could be difficult, but if there is a value already considered reasonable for $g = 1$, say, p, then one possibility may be to use $p_g = pn/n_r$. It has been documented by Wyse and Friel (2010) that analyses may be sensitive to the choice of p however, so this is still an issue.

MAP estimation

Fearnhead outlines some subtle issues associated with the MAP estimation of the number of changepoints and their positions as used in the paper. Fearnhead notes that if you compute the joint MAP estimate of all changepoints using a Viterbi algorithm it may give different results to first computing the MAP number of changepoints and then computing their MAP positions recursively. This is a subtlety to bear in mind when proposing possible extensions to this work in the future. While the results given in the paper act as a useful summary of the datasets analyzed, the possibility that their accuracy may be compromised by a more poorly defined MAP approximation than that given by the Viterbi algorithm should be kept in mind. One way of overcoming these potential problems while still maintaining the advantages of using RFRs and the wider class of segment models given by GMRFs may be to use the type of $O(n_r^2)$ recursive algorithm outlined at the end of the previous section on asymptotics. It would then be possible to extend the Viterbi algorithm in Fearnhead (2005) to this situation where recursions are computed instead at the time points in the reduced time index set.

Koop

Priors for segment parameters

It is certainly the case that priors on the parameters in each segment will have a big impact on forecasts from changepoint models because of the issues outlined by Koop. If there is only a small amount of data collected after a changepoint a forecast from any model which regards segments as being entirely independent will be dominated by the prior assumptions on the segment parameters. It is thus imperative for such situations to use an approach which allows the prior on the parameters for segment j to depend on the parameters in segment $j-1$, as in time-varying state-space models. It would be a sensible next step to try to extend the approaches developed in this paper for this situation, drawing on the work of Fearnhead and Liu (2011). This brings up another important extension, which is to the online analysis of changepoint models (Fearnhead and Liu (2007)), which would be particularly important in the forecasting setting. For the examples considered in the paper, segment lengths are of a reasonable duration, and forecasting has not been considered. But for similar reasons outlined in our reply to Fearnhead (Motivation for RFRs), forecasting with a small amount of data may still be an issue because of the use of an elaborate data model within segments. It will be interesting to see the further challenges that arise through extensions of the current methodology in these directions.

Priors for the changepoints

Koop's point about the misleading nature of uniform distributions in $p(\tau_1, \tau_2) = p(\tau_1)p(\tau_2|\tau_1)$ is very interesting, especially from the perspective of the forecaster. Here, our reason for using such a prior is for computational benefit. In fact the particular prior we use has very good computational properties as spotted by Fearnhead (2006). Not all priors of the above form will have these good computational properties, and so a general recommendation based on Koop's argument would be not to use these priors in settings where recursions are the tool of analysis, since you lose in two ways– computation and prior objectivity.

A more difficult issue is the desire which Economists would often have in incorporating dependence between priors on the durations $d_j = \tau_j - \tau_{j-1}$. This will be a difficult issue in general for approaches involving recursions, not just the simulation free approach outlined in this paper. This is since computing the recursions for all possible segmentations of the data and hence possible priors on the durations would lead to a formidable computing and storage task. Monte Carlo simulation is still the most promising option for these types of priors on the durations.

Priors for the number of changepoints

Following the comments by Fearnhead on the subtleties of MAP estimation for changepoint models and the argument given by Koop on the advantages of allowing in and out-of-sample changepoints, a very useful avenue appears to be in pursuing the $O(n_r^2)$ recursions discussed in the reply to Fearnhead (MAP estimation). This would mimic the approaches advocated by Koop, since the number of changepoints is not set in advance. There may still be situations when placing a prior on the number of changepoints is sensible if there is existing considerable knowledge that a certain number should occur. For example, in the coal mining data the identified changes can be linked to changes in the coal mining industry at those times in history. Other settings may have similar "big events" that have been known to occur, with the desire to detect these events in series of data retrospectively– possibly in the search for markers of these events in the future. In such situations the approaches outlined in this paper will still be very useful, especially if the series are very long and need to be scanned on a large scale to find potential changepoints.

References

Barry, D. and Hartigan, J. A. (1992). "Product Partition Models for Change Point Problems." *The Annals of Statistics*, 20: 260–279.

Fearnhead, P. (2005). "Exact Bayesian Curve Fitting and Signal Segmentation." *IEEE*

Transactions on Signal Processing, 53: 2160–2166.

— (2006). "Exact and efficient Bayesian inference for multiple changepoint problems." *Statistics and Computing*, 16: 203–213.

Fearnhead, P. and Liu, Z. (2007). "On-line inference for multiple changepoint problems." *Journal of the Royal Statistical Society, Series B*, 69: 589–605.

— (2011). "Efficient Bayesian analysis of multiple changepoint models with dependence across segments." *Statistics and Computing*, 21: 217–229.

Liu, J. S. and Lawrence, C. E. (1999). "Bayesian inference on biopolymer models." *Bioinformatics*, 15: 38–52.

Wyse, J. and Friel, N. (2010). "Simulation-based Bayesian analysis for multiple change-points." *arXiv:1011.2932v1 [stat.CO]*.

Yao, Y.-C. (1984). "Estimation of a Noisy Discrete-Time Step Function: Bayes and Empirical Bayes Approaches." *The Annals of Statistics*, 12: 1434–1447.