A Bayesian Approach for detecting Climate Shifts

Dr Clare McGrory
University of Queensland

Centre for Applications in Natural Resource Mathematics
University of Queensland
Brisbane
Queensland, Australia

October 5, 2017

Outline

- Brief background on Sequential Monte Carlo (SMC) and motivation for this research.
- Time-efficient variational Bayesian (VB) approach for finding posterior estimates of hidden Markov models.
- A novel hybrid scheme: Transdimensional SMC with VB proposals - SMCVB - for hidden Markov modelling.
- Example of results and application to regime shift modelling.

Sequential Monte Carlo (SMC)

- Sequential Monte Carlo (SMC) techniques provide a means of sampling from the posterior distribution of interest in Bayesian inference.
- In SMC, a weighted sample of 'particles' is generated from a sequence of probability distributions which 'converge' to the target distribution of interest, in this case a Bayesian posterior distribution.
- SMC methods are based on the idea of sampling from the resulting related sequence of target posterior distributions.

Sequential Monte Carlo (SMC)

- Early research in this area focused on the use of the sequential approach to analyse data that truly arose sequentially over time (see Doucet et al. (2001) for an overview).

- This was done by proposing an initial population of samples from the initial target posterior - these are referred to as ‘particles’.

- These currrent particles are then reweighted via importance sampling and resampled to approximate the next target posterior density in the sequence.

Doucet, A., De Freitas, J.F.G., Gordon, N.J.: Sequential Monte Carlo Methods in Practice. Springer, NewYork (2001)

Sequential Monte Carlo

- SMC has also been applied to static problems where the observed data are treated as a sequence by reading the dataset in batches. This concept has also been explored extensively with various schemes having been proposed.

- In particular, Chopin (2002) proposed the data-tempering SMC algorithm.

  Chopin, N.: A sequential particle filter method for static models. Biometrika 89, 539–551 (2002)

- While SMC schemes are faster than many MCMC-based approaches, there is still scope for exploring ways to more efficiently target the posterior with better proposals for generating particles.

Sequential Monte Carlo (SMC): Motivation for this work

- VB is a very fast alternative to MCMC which has been shown to often provide a very good approximation to the true posterior.

- A hybrid SMCVB scheme based on this was outlined in the context of finite mixture modelling in

  McGrory C.A. et al.: Transdimensional Sequential Monte Carlo using Variational Bayes SMCVB. Under revision. (2014).

- By using VB to generate the proposal distributions for new particles, we aim to make the SMC scheme more efficient.

Sequential Monte Carlo (SMC): Motivation for this work

- Most existing static SMC approaches are restricted to fixed-dimensional space which can be restrictive for practical application since estimating a suitable dimension for the model is usually an important part of the analysis.

- A transdimensional SMC algorithm was provided in Del Moral et al. (2006) for the changepoint problem with the number of changepoints being unknown.
  - A birth move was used to generate a new changepoint and the algorithm made use of reversible jump Markov chain Monte Carlo (RJMCMC) kernels to maintain particle diversity.
  - Disadvantage of using RJMCMC is that it is very computationally intensive.

VB for Fitting a Hidden Markov Model with Gaussian Noise

- Assume a Gaussian hidden Markov model (HMM) where the system can be in any one of K states at any time-point i, but the actual state sequence is hidden.

- Observations correspond to a noisy realisation of the actual state sequence. We assume a discrete first-order Markovian dependence structure, therefore the current state depends only on the state occupied at the last time-point.

McGrory, C.A. and Titterington, D.M. (2009). Bayesian analysis of hidden Markov models using variational approximations. Australian and New Zealand Journal of Statistics, vol. 51(2), pp 227–244.

VB for Fitting a Hidden Markov Model with Gaussian Noise

- Given that the system is in state j_1 at time-point i, the transition matrix π represents the probability of moving to state j_2 at time-point $i+1$.

- Transition matrix is defined as $\pi = \{\pi_{j_1 j_2}\}$ where $\pi_{j_1 j_2} = p(z_{i+1} = j_2 | z_i = j_1)$ and z_i is the latent variable representing the state at time i;

- Observed data is denoted by $\{y_i; i = 1, \ldots, n\}$, and the emission probabilities, i.e., the conditional probabilities of state membership at each time-point, are denoted by $p(y_i | z_i = j) = p_j(y_i | \phi_j)$.

VB for Fitting a Hidden Markov Model with Gaussian Noise

$$
\begin{aligned}
p(y, z, \theta) \quad = \quad & \prod_{i=1}^{n} \prod_{j=1}^{K} (p_j(y_i|\phi_j))^{z_{ij}} \times \prod_{i=1}^{n-1} \prod_{j_1=1}^{K} \prod_{j_2=1}^{K} (\pi_{j_1 j_2})^{z_{ij_1} z_{i+1 j_2}} \\
& \times \prod_{j=1}^{K} p_j(\phi_j) \prod_{j_1=1}^{K} p(\pi_{j_1}),
\end{aligned}
$$

where z_{ij} is a latent indicator variable such that $z_{ij} = 1$, if $z_i = j$, and $z_{ij} = 0$, if $z_i \neq j$.

Standard conjugate priors are used.

Hidden Markov Modelling Using Variational Bayes (VB): Approach Outline

- VB approach is non-simulation based and as a result it provides a highly time-efficient way of performing inference.

- Particularly useful for analysing large datasets.

- The VB approach provides an approximation to the posterior distribution of interest; this is referred to as the variational posterior.

McGrory, C. A., Titterington, D. M.:Variational approximations in Bayesian model selection for finite mixture distributions. Compututational Statistics and Data Analysis, 51, 5352–5367 (2007)

Hidden Markov Modelling Using Variational Bayes (VB): Approach Outline

The variational approximation for the posterior distribution $p(\theta|y)$ is found as the appropriate marginal distribution of the approximation to the joint conditional density $p(\theta, z|y)$; it is this joint conditional density which is approximated in the VB approach.

- In order to approximate $p(\theta, z|y)$ introduce a more easily computed distribution: $q(\theta, z)$.

- The variational distribution is chosen to minimise the Kullback-Leibler divergence between $q(\theta, z)$ and $p(\theta, z|y)$

- Equivalently, this amounts to choosing $q(\theta, z)$ to maximise the lower bound for $p(\mathbf{y})$.

Hidden Markov Modelling Using Variational Bayes (VB): Approach Outline

- In order to make this maximisation tractable, the standard VB approach is to assume that the variational distribution takes the factorised form
$$q(\theta, z) = q_\theta(\theta) q_z(z)$$

- The lower bound can then be maximised to obtain the algebraic forms of the variational update expressions for each of the model parameters and for the hidden indicator variables.

- These update equations can then be solved iteratively to obtain the estimated variational posterior.

Hidden Markov Modelling Using Variational Bayes (VB): Variational Posteriors

$$
\begin{aligned}
q_{j_1}(\pi_{j_1}) &= \mathrm{Dir}(\pi_{j_1}|\{\alpha_{j_1 j_2}\}), \\
q(\mu_j|\tau_j) &= \mathrm{N}(\mu_j|m_j, (\beta_j\tau_j)^{-1}), \\
q(\tau_j) &= \mathrm{Ga}\left(\tau_j|\frac{1}{2}\eta_j, \frac{1}{2}\delta_j\right).
\end{aligned}
$$

- The forward-backward algorithm is used to find the $a^*_{j_1 j_2}$ which are the estimates of the probabilities of transition from states j_1 to state j_2, and the b^*_{ij}'s are estimates of the emission probabilities given that the system is in state j at time point i.

- These are then used in the update equation for $q_{ij} = q_z(z_i = j) = p(z_i = j_1|y_1, \ldots, y_n)$ and $q_z(z_i = j_1, z_{i+1} = j_2)$.

Hidden Markov Modelling Using Variational Bayes (VB): Approach Outline

- We can iteratively solve these update expressions to find the variational posterior estimates.
- In the standard approach for VB fitting of hidden Markov models, the algorithm is initialised with a sufficiently large number of components and, as the algorithm converges, redundant components are eliminated through the approximation.
- This means that the VB approach estimates a suitable number of components for the model, and this estimated K will be less than or equal to the initial number proposed.
- This property is an intrinsic feature of the VB approach.

A novel hybrid scheme: Transdimensional SMC with VB proposals - SMCVB

- Within the context of hidden Markov modelling, we propose a new transdimensional SMC algorithm based on the idea of using the variational Bayes (VB) approach to generate proposal distributions.
- In other words, the algorithm uses particles drawn from a VB approximation to the posterior rather than from the prior.
- Priors can be quite diverse leading to inefficiency in the SMC.

A novel hybrid scheme: Transdimensional SMC with VB proposals - SMCVB

- The complete-data target posterior is

$$\pi(\theta) = \pi(\theta | y_1, \cdots, y_n)$$

- The target posterior at iteration t ($t = 1, \cdots, T$) is

$$\pi_t(\theta) = \pi_t(\theta | y_1, \cdots, y_{n_t}),$$

  where $n_1 \leq n_2 \leq \cdots \leq n_T = n$ is an increasing set of sample sizes.

- By separating the data into batches in this way we form a sequence of target posteriors which on average smoothly converge to the complete data target posterior.

Transdimensional SMC with VB proposals - SMCVB: Algorithm Outline

0. Initialise:
   - We generate a set of R particles$(\theta_r^{(0)}, W_r^{(0)})_{r=1,\cdots,R}$ with associated weights $\{W_r^{(0)}\}$ to target the initial posterior $\pi_{t_0}(\theta)$.
   - We do this by estimating the VB partial posterior $\pi_{VB}(\theta|y_1,\cdots,y_{n_0})$ to obtain the posterior estimates
   - We can then draw R particles from these estimated posteriors, which results in vectors of the form $\{\theta_R^{(0)} = (\mu_r^{(0)}, \tau_r^{(0)}, \rho_r^{(0)})\}$
   - The weights are then obtained as

$$W_r^{(0)} \propto \frac{p(y_1,\cdots,y_{n_0}|\theta_r^{(0)})p(\theta_r^{(0)})}{\pi_{VB}(\theta_r^{(0)}|y_1,\cdots,y_{n_0})}$$

We then normalise the weights to obtain $W_r^{(0)}$.

Transdimensional SMC with VB proposals - SMCVB: Algorithm Outline

1. Reweight:
   - We update the weights at iteration t using the n_tth batch of data giving
   $$W_r^{(t)} \propto W_r^{(t-1)} \times p(y_{n_{t-1}+1}, \cdots, y_{n_t} | \theta_r^{(t-1)}),$$
   where $r = 1, \cdots, R$.

Transdimensional SMC with VB proposals - SMCVB: Algorithm Outline

2. Resample if the ESS is not large enough:
   - We resample R values from the current set of particles using a suitable selection scheme such as multinomial, residual or stratified sampling. We use multinomial sampling.
   - We resample the $\{(\theta_r^{(t-1)}, W_r^{(t-1)})\}_{r=1,\cdots,R}$ to get $\{(\theta_r'^{(t)}, 1/R)_{r=1,\cdots,R}\}$.

This means that the set of resampled particles may contain more than one copy of some of the particles from the previous set $\{\theta_r^{(t-1)}\}$.

Transdimensional SMC with VB proposals - SMCVB: Algorithm Outline

3. Move:
   - We move to a new set of particles, that will become the $\{\theta_r^{(t)}\}$ to be carried forward, using a Metropolis–Hastings (MH) update, where the proposal distribution is obtained from the VB posterior mean of the parameters based on data $y_1, \cdots, y_{n_t}$.

   - For each r we propose a parameter $\theta_r^{*(t)}$ from $\pi_{VB}(\theta|y_1, \cdots, y_{n_t})$ and accept or reject, in favour of $\theta_r^{\prime(t)}$, according to the ratio

$$MH_r = \frac{p(y_1, \cdots, y_{n_t}|\theta_r^{*(t)})}{p(y_1, \cdots, y_{n_t}|\theta_r^{\prime(t)})} \times \frac{\pi_{VB}(\theta_r^{\prime(t)}|y_1, \cdots, y_{n_t})}{\pi_{VB}(\theta_r^{*(t)}|y_1, \cdots, y_{n_t})}$$

$$\times \frac{p(\theta_r^{*(t)})}{p(\theta_r^{\prime(t)})} \times \frac{p(*->')}{p('->*)}.$$

Transdimensional SMC with VB proposals - SMCVB: Algorithm Outline

4. Iterate: repeat steps 1-3 until $n_t = n$.

- In this way we eventually reach the target which is the posterior for the full dataset.

- The step that makes our approach novel in comparison to other SMC algorithms is step 3 where we use a VB posterior mean estimate of the model parameters in order to generate proposal particles.

Examples of Some Results Obtained for a Simulated Dataset of 1000 Data Points

Parameters of the Gaussian noise distributions		
State	Mean	Standard Deviation
1	1.00	0.50
2	2.00	0.15
3	2.50	0.30

Post. Means of the mean		
SMCVB	VB	MCMC
1.01	1.01	1.01
2.00	2.00	2.00
2.56	2.56	2.55

Post. Means of the Std Dev		
SMCVB	VB	MCMC
0.53	0.53	0.53
0.15	0.15	0.15
0.26	0.26	0.27

Climate Regime Shift Detection

- A regime shift is a term commonly used to describe an abrupt change in some aspect of the characteristic behaviors associated with a natural phenomenon.

- Climate variability is one such example of a natural phenomenon for which there is much interest in understanding, and if possible, predicting when changes in patterns might occur.

Rodionov, S. N. (2004), A sequential algorithm for testing climate regime shifts, Geophys. Res. Lett., 31.

Climate Regime Shift Detection

- Climate events can have a large impact on the environment, therefore there has been interest in research investigating other such events that have occurred over the years.

- The observed natural fluctuations in climate we will look at are referred to as Pacific Decadal Oscillation (PDO).

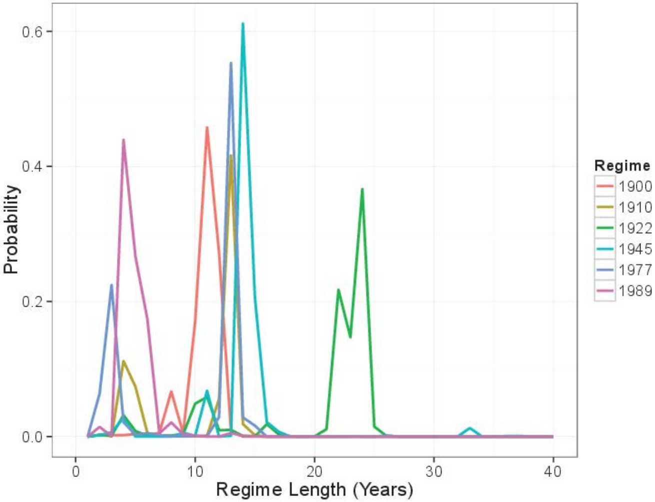
- Extreme observations in the PDO correspond to large fluctuations in the climate of the Pacific Basin and North American region.

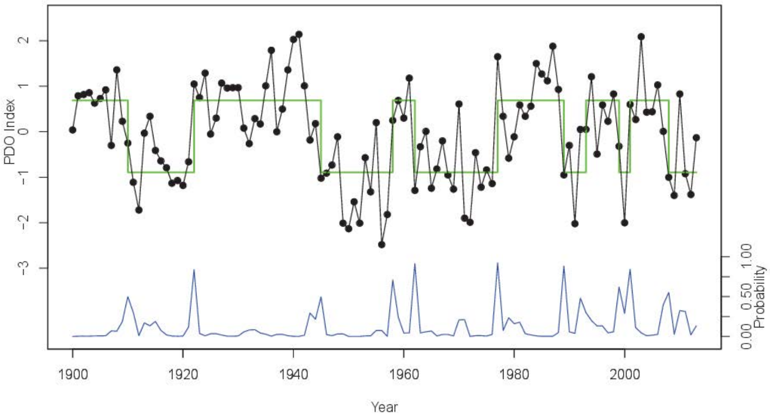
- There is much scope for improving upon existing analytical approaches.

Climate Regime Shift Detection

- The majority of approaches proposed in the early literature use basic statistical techniques such as tests of significant differences from one time point to the next in the series.
- Some slightly more involved approaches were proposed more recently but a difficulty of these was that they could not be used for time points lying close to either end of the time series.
- Sequential analyses have been proposed in the literature but the drawback of these is that although it is known that fluctuations in climate can take place over time periods of varied lengths, a regime is rigidly defined as spanning a defined time-period.

Some Illustrative Graphs





PDO Index
2
1
0
−1
−2
−3
Probability
0.00 0.50 1.00
1900
1920
1940
1960
1980
2000
Year

Summary

- A new hybrid Bayesian algorithm has been presented.
- It appears that this hybrid algorithm might lead to a better fit to the data than can be achieved using a standard variational Bayes approach in some cases.
- There is much scope for further development and useful application of the ideas presented here.