Sequential Monte Carlo for static Bayesian models

Associate Professor Chris Drovandi
Queensland University of Technology and
Australian Centre of Excellence for Mathematical and Statistical Frontiers

Collaborators: Leah South and others

30th November, 2018

Outline

Bayesian Posterior Sampling

Interest is in sampling from the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, where

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{Z}$$

and $Z = \int_{\Theta} f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$.

Z is referred to as the evidence and is useful for model choice.

Markov Chain Monte Carlo

MCMC methods construct an ergodic Markov chain with the posterior as its limiting distribution.

A common MCMC algorithm is Metropolis Hastings (MH-) MCMC, where proposals $\boldsymbol{\theta}^*$ are accepted with probability

$$\min\left(1, \frac{f(\mathbf{y}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}\right),$$

where $q(\cdot)$ is the proposal density.

Markov Chain Monte Carlo

Some Limitations:

- Difficult to automate and adapt the method
- Need to tune the proposal distribution for good performance.
- Convergence can be difficult to assess
- Can have difficulty exploring irregular posteriors (eg multi-modality)
- Standard MCMC is a serial algorithm

Sequential Monte Carlo (Chopin et al 2002)

SMC methods can be a useful alternative to MCMC in some applications as they are...

- Naturally adaptive
- Easily parallelisable
- More capable of dealing with multimodal or complex distributions
- Able to produce an estimate of the unknown normalising constant

Sequential Monte Carlo

Basic idea:

- Moving a population of N particles through a sequence of distributions (starting with one easy to sample from and finishing at the target posterior)
- Can introduce the effect of either the data (data annealing) or the likelihood (likelihood annealing) sequentially

Sequential Monte Carlo

In likelihood annealing, the power posteriors are defined by

$$\pi_t(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})^{\gamma_t}\pi(\boldsymbol{\theta}),$$

where $0 = \gamma_0 < \gamma_t < \gamma_T = 1$.

At each iteration, the following steps are applied

- reweighting
- resampling
- moving to avoid particle degeneration, for example by several runs of an MCMC kernel (largest impact on estimates, highest cost). We can make use of population of particles.

Sequential Monte Carlo

Denote the collection of particles representing target t as $\{\boldsymbol{\theta}_t^i, W_t^i\}_{i=1}^N$.
Re-weight step:

$$w_{t+1}^i = W_t^i f(\boldsymbol{y}|\boldsymbol{\theta}_t^i)^{\gamma_{t+1}-\gamma_t}$$

Effective sample size (ESS) can be estimated as $\text{ESS} = 1/\sum_i (W_{t+1}^i)^2$.
The sequence of γ_t can be selected adaptively to maintain a particular ESS, e.g. $N/2$.

Boost ESS back up to N via resampling. e.g. Multinomial Re-sampling.

(Diversify Particles) Apply MCMC kernel with π_{t+1}-invariant distribution R_{t+1} times. Determine R_{t+1} adaptively by performing 1 iteration on each particle and inspecting acceptance rate.

Sequential Monte Carlo - Estimating the Evidence

- SMC provides convenient estimate of evidence, Z
- We can write Z as:

$$Z = \frac{Z_T}{Z_0} = \prod_{t=0}^{T-1} \frac{Z_{t+1}}{Z_t},$$

with $Z_0 = 1$. It is easy to show that

$$Z_{t+1}/Z_t \approx \sum_{i=1}^{N} w_{t+1}^i.$$

Call this the "standard" SMC estimator.

Independent Proposals within SMC

We explore the use of independent MCMC proposals in SMC.

Benefits of independent proposals in SMC

- Takes advantage of the population of particles to form efficient independent proposals
- Take advantage of the parallelisable nature of SMC
- Re-use all information generated in the SMC process (better estimates of posterior quantities and evidence compared with standard SMC)

Our Independent Proposal

Our efficient independent proposals are based on

- Copula models (Sklar 1959) which model dependence between parameters while also modelling their marginals separately
- Mixture models for marginals and dependence

Fitted using SMC particles. Denote independent proposal for target π_t as $q^{\phi_t}(\cdot)$.

Recycling all Proposals

We can use independent proposals $q^{\phi_t}(\cdot)$ for $t = 0, \ldots, T$ as importance distributions.

- T+1 separate estimators of the evidence

$$\hat{Z} = \frac{1}{N_t} \sum_{i=1}^{N_t} \omega_t^i, \text{ where } \omega_t^i = \frac{f(\mathbf{y}|\boldsymbol{\theta}_t^i)\pi(\boldsymbol{\theta}_t^i)}{q^{\phi_t}(\boldsymbol{\theta}_t^i)}$$

- A single evidence estimator

$$\hat{Z} = \frac{1}{\sum_{k=0}^{T} N_k} \sum_{t=0}^{T} \sum_{i=1}^{N_t} \omega_t^i$$

This estimator doesn't take into account the different ESS's from different temperatures.

Recycling all Proposals

A novel importance sampling estimator (similar to Nguyen et al 2014)

- Define $\lambda_t = \frac{ESS_t}{\sum_{t=0}^{T} ESS_t}$
- Weights are scaled by λ_t
- The evidence estimator is

$$\widehat{Z} = \sum_{t=0}^{T} \frac{\lambda_t}{R_t N} \sum_{m=1}^{R_t N} \omega_t^m.$$

We refer to this method as combined importance sampling CIS.

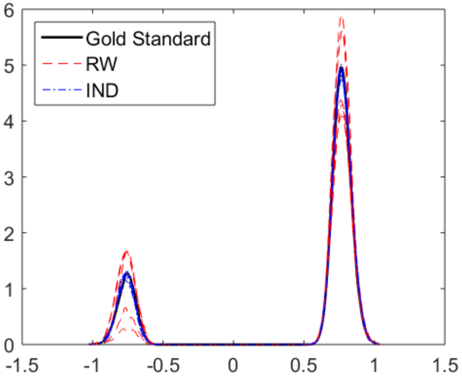
Example

Factor analysis example

- Model choice example with 1, 2 or 3 factors (12-21 parameters)
- Well separated modes in the 2 factor model
- Will demonstrate some of the improvements achieved in this example, based on 100 runs with 5000 particles

Example

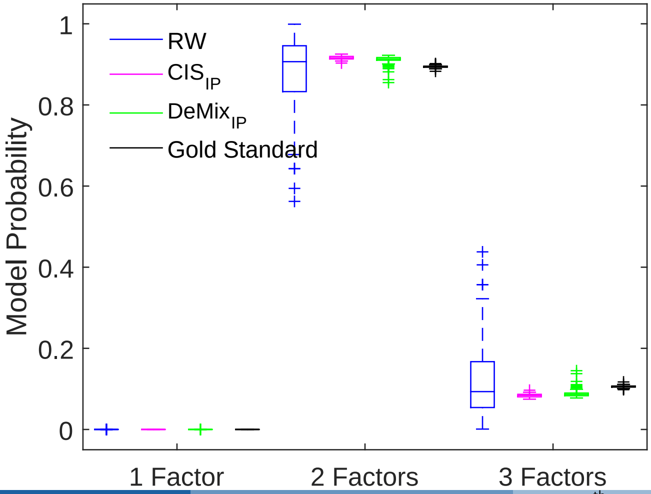
An example posterior marginal in the two factor model



ESS of 5000 (standard SMC) and 29000 (recycling SMC) on average.

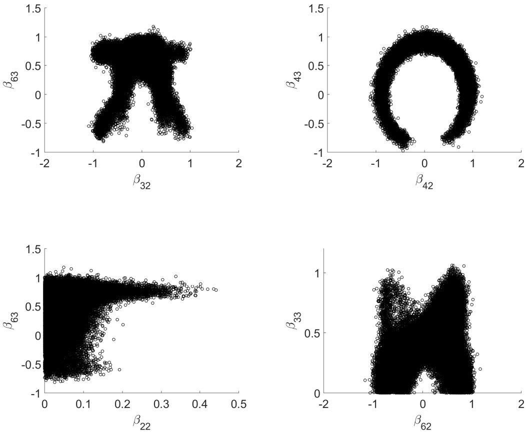
Example

Model choice



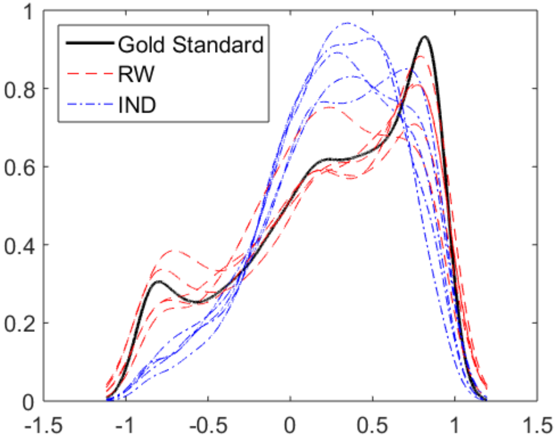
Example

Challenges: some bivariate scatterplots for the three factor model



Example

Challenges: an example marginal in the three factor model



Independent SMC Summary

Benefits:

- Efficient MCMC proposal within SMC
- Recycle all proposals in estimating the posterior and evidence
  - Significant improvement in posterior inference compared to no recycling or accepted particle recycling, *if* independent proposals cover the tails of the target
  - Precise estimates of the evidence via IS identities

Challenges:

- Difficulty achieving tail coverage with current independent proposals

Control Variates

Determine an auxiliary function $\tilde{\varphi}(\boldsymbol{\theta}) = \varphi(\boldsymbol{\theta}) + h(\boldsymbol{\theta})$ such that $\mathbb{E}_\pi[\tilde{\varphi}(\boldsymbol{\theta})] = \mathbb{E}_\pi[\varphi(\boldsymbol{\theta})]$ and $\mathbb{V}_\pi[\tilde{\varphi}(\boldsymbol{\theta})] < \mathbb{V}_\pi[\varphi(\boldsymbol{\theta})]$, where $\mathbb{V}_\pi$ denotes the variance with respect to $\pi(\boldsymbol{\theta}|\boldsymbol{y})$.

The standard Monte Carlo estimator for the expectation is replaced with the unbiased, reduced variance estimator,

$$\widehat{\mathbb{E}_\pi[\tilde{\varphi}(\boldsymbol{\theta})]} = \frac{1}{N}\sum_{i=1}^{N}\left[\varphi(\boldsymbol{\theta}_i) + h(\boldsymbol{\theta}_i)\right],$$

where $\{\boldsymbol{\theta}^i\}_{i=1}^N \sim \pi(\cdot|\boldsymbol{y})$.

Zero-variance Control Variates

So called zero-variance control variates are defined as:

$$\begin{aligned} h(\boldsymbol{\theta}) &= \Delta_{\boldsymbol{\theta}} P(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} P(\boldsymbol{\theta}) \cdot \nabla_{\boldsymbol{\theta}} \log \pi(\boldsymbol{\theta}|\boldsymbol{y}) \\ &= \boldsymbol{\beta}^T \boldsymbol{x}, \end{aligned}$$

where $P(\boldsymbol{\theta})$ is a Q-order polynomial, $\boldsymbol{\beta}$ is the polynomial coefficients and $\boldsymbol{x}$ is a vector of terms involving $\boldsymbol{\theta}$ and $\nabla_{\boldsymbol{\theta}} \log \pi(\boldsymbol{\theta}|\boldsymbol{y})$.

Obtain $\boldsymbol{\beta}$ via ordinary least squares (OLS) and then $\widehat{\varphi(\boldsymbol{\theta})} = \widehat{\alpha} - \widehat{\boldsymbol{\beta}}^T \boldsymbol{x}$:

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = \underset{\alpha, \boldsymbol{\beta}}{\arg\min} \frac{1}{N} \sum_{i=1}^{N} \left[\varphi(\boldsymbol{\theta}_i) - \alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i \right]^2 .$$

Suffers from curse of dimensionality since regression involves $\binom{d+Q}{d}$ terms for Qth order polynomial.

Regularised Control Variates

- Making the polynomial $P(\boldsymbol{\theta})$ a function of a subset of $\boldsymbol{\theta}$ based on prior information
- Least absolute shrinkage and selection operator (lasso) so that we replace OLS with

$$\arg\min_{\hat{\alpha},\hat{\boldsymbol{\beta}}} \frac{1}{M}\sum_{i=1}^{M}\left[\varphi(\boldsymbol{\theta}_i) - \hat{\alpha} + \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i\right]^2 + \lambda||\hat{\boldsymbol{\beta}}||_1.$$

Application to SMC

Applications to SMC:

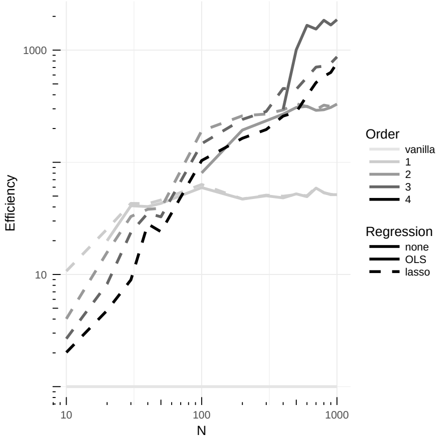
- Posterior expectations
- SMC evidence estimator

$$Z = \prod_{t=1}^{T} \mathbb{E}_{\pi_{t-1}} \left[\frac{\eta_t(\boldsymbol{\theta})}{\eta_{t-1}(\boldsymbol{\theta})} \right].$$

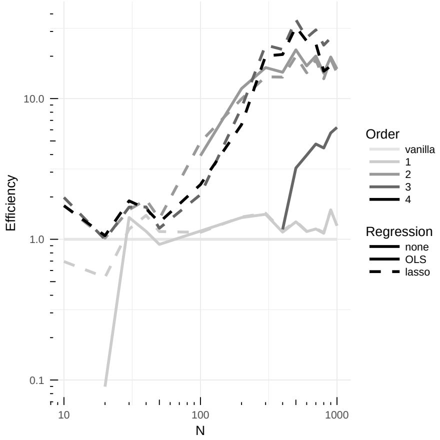
- Comparisons to alternative evidence estimators

Example

An 11-dimensional capture-recapture example:



(e) Posterior expectations



(f) Evidence estimation

Future Work:

- Best way to use cheap likelihood surrogate in SMC.
- Mixing of different MCMC proposals.
- SMC for Hierarchical models.

Open Problems:

- How to the estimate the ‘true’ ESS from output of SMC (ESS of N under assumption particles are independent).

Key References

Chopin, N. (2002). A sequential particle filter method for static models. Biometrika, 89(3):539-552.

Nguyen, T. L. T., Septier, F., Peters, G. W., and Delignon, Y. (2014). Improving SMC sampler estimate by recycling all past simulated particles. In 2014 IEEE Workshop. In Statistical Signal Processing (SSP), pages p117-120.

South, L. F., Pettitt, A. N. and Drovandi, C. C. (2018). Sequential Monte Carlo for static models with independent MCMC proposals. To appear in Bayesian Analysis.

South, L. F., Oates, C., Mira, A. and Drovandi, C. (2018). Regularised Zero-Variance Control Variates. https://arxiv.org/abs/1811.05073.

Email: c.drovandi@qut.edu.au
Web: chrisdrovandi.weebly.com
Twitter: @chris_drovandi