Sequential Monte Carlo

Sequential Monte Carlo (SMC) methods [1] can be powerful alternatives to Markov chain Monte Carlo (MCMC) methods [7] for performing inference on static Bayesian models. SMC methods are adaptive, parallelisable and are more capable of dealing with multimodal or complex targets.

Assume there is data y and interest is in a statistical model parameterised by \( \theta \). Likelihood annealing SMC traverses a population of \( N \) particles through a sequence of distributions defined by the power posteriors

\[
\pi_t(\theta | y) \propto f(y | \theta)^{1/t} \pi(\theta),
\]

where \( 0 < \gamma_t < \gamma_{t+1} < 1 \) and \( 0 < t < T \). A weighted particle set targeting \( \pi_t \) is represented by \( \{ W_t^i, \theta_t^i \}_{i=1}^N \), where \( W_t^i \) is a normalised weight.

The effect of the likelihood is introduced smoothly through the following steps:

- Reweighting the particle set to target \( \pi_{t+1} \). The new unnormalised weights are
  \[
  w_{t+1} = W_t^i f(y | \theta_t^i)^{1 - 1/\gamma_t}, \quad \text{for } i = 1, \ldots, N.
  \]
- Resampling.
- Diversifying the particles, often via several iterations of an MCMC kernel with a multivariate normal walk (RW) proposal.

Evidence Estimation in SMC

The log evidence can be estimated using the stepping stone (SS) identity

\[
\log Z = \sum_{i=1}^N \log E_{\gamma_i}(y | \theta) f(y | \theta)^{1/\gamma_i},
\]

or with the thermodynamic identity (TI)

\[
\log Z = \int_0^1 \mathbb{E}_\gamma \log f(y | \theta) d\gamma,
\]

which gives the log evidence as an integral with respect to the temperature \( \gamma \) [8]. We use a 2nd order quadrature approximation [2] for the integral in (2).

Using Derivative Information

The motivation for using the derivatives \( \nabla_y \log \pi_t(\theta | y) \) is that we would like to achieve the same level of precision with fewer likelihood evaluations.

Choice of Move Kernel

Using efficient move kernels leads to a higher acceptance rate and therefore fewer log likelihood evaluations. The Metropolis-adjusted Langevin algorithm (MALA, [3]) is an alternative to the popular RW proposal and it uses the log posterior as the stepping stone:

\[
q^\phi(\theta' | \theta) = \mathcal{N}(\theta'; \theta, \frac{\tau^2}{2} \nabla^2 \log \pi(\theta | y)), \quad \tau^2 = \beta^{-2} \nabla^2 \log \pi(\theta | y),
\]

where the \( \nabla^2 \log \pi(\theta | y) \) is a local measure of curvature for the log posterior and is referred to as the metric tensor. We learn \( \beta \) from the population of particles. The results in this paper are based on using the empirical covariance matrix for \( \hat{\Sigma} \), but if the second derivatives, \( \nabla^2 \log \pi(\theta | y) \in \mathbb{R}^{d \times d} \), are available, these can be used to compute the observed or expected Fisher-Rao metric tensor \( \hat{\Sigma} \) at \( \theta \) (MMALA,[3]).

Post-hoc Adjustment

Control variates can be used to get lower variance estimators of expectations \( \mathbb{E}_\pi(\phi(\theta)) \). The general framework for control variates is to determine an auxiliary function \( \phi(\theta) = \phi(\theta) + h(\theta) \) such that \( \mathbb{E}_\pi[\phi(\theta)] = \mathbb{E}_\pi[h(\theta)] \) and \( \nabla_y \mathbb{E}_\pi[\phi(\theta)] < \nabla_y \mathbb{E}_\pi[h(\theta)] \), where \( \phi(\theta) \) denotes the variance with respect to target \( \theta \). This can be achieved by choosing some random variable which is correlated with \( \phi(\theta) \) and has a known expectation. Here we use zero-variance control variables (ZV-CV, [6]), which require only the derivative of the log target or some unbiased estimator of this quantity.

We apply ZV-CV to all expectations in (1) and (2).

Recapture Example

Here we estimate the parameters of a Cormack-Jolly-Seber model based on the capture and recapture of a bird species [5]. The parameters are \( \theta = (\phi_1, \ldots, \phi_5, \beta_0, \ldots, \beta_5, \gamma_0, \gamma_1) \), where \( \gamma_0 \) represents the probability of survival from year \( i \) to year \( i + 1 \) and \( \beta_0 \) represents the probability of being captured in year \( k \). The likelihood for the model is

\[
\ell(y | \theta) = \prod_{i=1}^T \prod_{k=1}^{N_y} \left( 1 - \prod_{j=1}^{p_k} (1 - \gamma_0 \beta_j) \right) \left( \phi(t_{ij}, \beta) \right) \left( \phi(t_{ik}, \beta) \right),
\]

where \( D_i \) is the number of birds released in year \( i \) and \( y_k \) is the number of animals caught in year \( k \) out of the number released in year \( i \). \( U(0, 1) \) priors are used and all parameters are transformed to the real line for the move step. Results: 100 SMC runs with \( N = 1000 \) particles are performed. The figure below shows the improvement in posterior and evidence estimation that can be achieved with derivative variates. RW uses between 1.5 and 25 times the number of log likelihood calculations that MALA uses.

Factor Analysis Example

[4] use factor analysis models for data \( Y \) related to the exchange rate of 6 currencies relative to the British pound. The factor analysis models assume that \( Y \sim \mathcal{N}(0, \Omega) \). To reduce model complexity, \( \Omega \) is parameterised by

\[
\Omega = \beta \Sigma + \Sigma,
\]

where \( \beta \) is a \( 6 \times k \) lower triangular matrix with positive diagonal elements and \( \Sigma \) is a \( 6 \times 6 \) diagonal matrix with positive elements. Here \( k \) is the number of factors. The prior and further details on this model can be found in [4].

Ongoing and Future Work

- MMALA - making use of the second derivatives
- control functionals for improved convergence
- an example with a nonlinear ordinary differential equation

References