Sequential Monte Carlo methods

Lecture 12 – Particle Matropolis Hastings

Thomas Schön, Uppsala University

2017-08-28

Outline – Lecture 12

Aim: Describe how we can make use of the particle filter inside the Metropolis Hastings algorithm to produce exact samples from the parameter posterior distribution for a nonlinear state space model.

Outline:

1. Using unbiased estimates within Metropolis Hastings
2. Exact approximation – Particle Metropolis Hastings (PMH)
3. Examples

Recall – Auxiliary variables (from lecture 6)

Target distribution: $\pi(x)$, difficult to sample from

Idea: Introduce another variable U with conditional distribution $\pi(u \mid x)$

The joint distribution $\pi(x, u) = \pi(u \mid x)\pi(x)$ admits $\pi(x)$ as a marginal by construction, i.e., $\int \pi(x, u)\mathrm{d}u = \pi(x)$.

Sampling from the joint $\pi(x, u)$ may be easier than directly sampling from the marginal $\pi(x)$!

The variable U is an **auxiliary variable**. It may have some "physical" interpretation (an unobserved measurement, unknown temperature, ...) but this is not necessary.

Pseudo-marginal Metropolis Hastings

The use of a non-negative and unbiased likelihood estimate within Metropolis Hastings is called the **pseudo-marginal approach**.

Algorithm 1 Pseudo-marginal Metropolis Hastings

1. **Initialize ($m = 1$):** Set $\theta[1]$ and run a particle filter for $\hat{z}[1]$.
2. **For $m = 2$ to M, iterate:**
   a. Sample $\theta' \sim q(\theta \mid \theta[m-1])$.
   b. Sample $\hat{z}' \sim \psi(z \mid \theta', y_{1:T})$ (i.e. run a particle filter).
   c. With probability
   $$\alpha = \min\left(1, \frac{\hat{z}' p(\theta')}{\hat{z}[m-1]p(\theta[m-1])} \frac{q(\theta[m-1] \mid \theta')}{q(\theta' \mid \theta[m-1])}\right)$$
   set $\{\theta[m], \hat{z}[m]\} \leftarrow \{\theta', \hat{z}'\}$ (accept candidate sample) and with prob. $1-\alpha$ set $\{\theta[m], \hat{z}[m]\} \leftarrow \{\theta[m-1], \hat{z}[m-1]\}$ (reject candidate sample).

Exact approximation

The pseudo-marginal Metropolis Hastings algorithm is one member of the family of so-called **exact approximation** algorithms.

Explanations of this slightly awkward name:

- It is an **exact** Metropolis Hastings algorithm in the sense that the target distribution of interest is the stationary distribution of the Markov chain,
- despite the fact that it makes use of an **approximation** of the likelihood in evaluating the acceptance probability.

The variance of the estimator $\widehat{Z}$ will significantly impact the convergence speed.

Two reflections on the PMH algorithm

1. PMH is a **standard MH algorithm** sampling from the joint target $\pi(\theta, z)$, rather than the original target $\pi(\theta)$. We have used the auxiliary variables trick, where the marginal of the joint target $\pi(\theta, z)$ w.r.t. z is by construction the original target $\pi(\theta)$.

2. Using a likelihood estimator $\widehat{Z}$ means that the marginal of

$$\pi(\theta, z \mid y_{1:T}) = \frac{z p(\theta) \psi(z \mid \theta, y_{1:T})}{p(y_{1:T})}$$

w.r.t. $\widehat{Z}$ will **not equal** the marginal of

$$\psi(\theta, z \mid y_{1:T}) = \frac{p(y_{1:T} \mid \theta) p(\theta) \psi(z \mid \theta, y_{1:T})}{p(y_{1:T})}$$

w.r.t. $\widehat{Z}$. This is ok, since we are only interested in the marginal w.r.t. θ, which remains the same for both π and ψ, namely $p(\theta \mid y_{1:T})$.

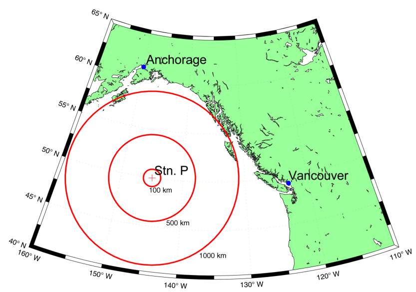
Examples

ex) Nonlinear marine biogeochemical model

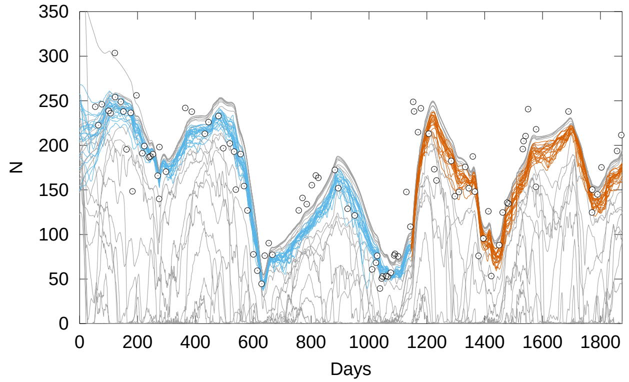
Studies nitrogen through an ecosystem in four compartments:

1. Nutrient (shown on next slide),
2. Phytoplankton,
3. Zooplankton and
4. Detritus.

The NPZD model used consists of nonlinear continuous-time diff. eq. with $x_t \in \mathbb{R}^{15}$, $\theta \in \mathbb{R}^{15}$, with discrete-time noise.



ex) Nonlinear marine biogeochemical model



Dissolved inorganic nitrogen concentration, circles are observations (very noisy), blue posterior paths from PMMH, red predictions and the grey lines are drawn from the prior.

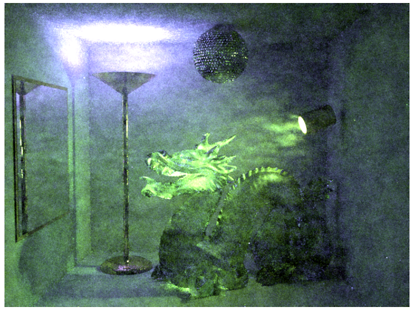
John Parslow, Noel Cressie, Edward P. Campbell, Emlyn Jones and Lawrence Murray. **Bayesian learning and predictability in a stochastic nonlinear dynamical model.** *Ecological Applications*, 23(4): 679–698, 2013.

ex) The pseudo-marginal idea is general

CG example (rendering images in heterogeneous media): An MH algorithms producing samples of the light paths connecting the sensor with light sources in the scene.

Results using equal time rendering

Our method that builds on MLT



Metropolis light transport (MLT)

Joel Kronander, Thomas B. Schön and Jonas Unger. **Pseudo-marginal Metropolis light transport.** *Proceedings of SIGGRAPH ASIA Technical Briefs*, Kobe, Japan, November, 2015.

Particle MCMC = SMC + MCMC

A systematic and correct way of combining SMC and MCMC.

Builds on an extended target construction.

Intuitively: SMC is used as a high-dimensional proposal mechanism on the space of state trajectories $\mathcal{X}^T$.

A bit more precise: Construct a Markov chain with $p(\theta, x_{1:T} \mid y_{1:T})$ (or one of its marginals) as its stationary distribution.

Very useful both for parameter and state learning.

Exact approximations

Further reading

Introducing the pseudo-marginal idea in a general setting:

Christophe Andrieu and Gareth O. Roberts. **The pseudo-marginal approach for efficient Monte Carlo computations.** *The Annals of Statistics*, 37(2):697–725, 2009.

Introducing PMCMC:

Christophe Andrieu, Arnaud Doucet and Roman Holenstein. **Particle Markov chain Monte Carlo methods.** *Journal of the Royal Statistical Society: Series B*, 72:269-342, 2010.

A (hopefully) pedagogical tutorial on PMH:

Thomas B. Schön, Andreas Svensson, Lawrence Murray and Fredrik Lindsten. **Probabilistic learning of nonlinear dynamical systems using sequential Monte Carlo.** *Pre-print arXiv:1703.02419*, 2017.

A few concepts to summarize lecture 12

Exact approximation: A family of MCMC algorithms that are exact in the sense that the target distribution of interest is the stationary distribution of the Markov chain, despite the fact that it makes use of an approximation of the likelihood in evaluating the acceptance probability.

Pseudo-marginal Metropolis Hastings makes use of a non-negative and unbiased likelihood estimate within the Metropolis Hastings algorithm.

Particle Metropolis Hastings makes use of a particle filter to guide an MCMC method through the parameter space. It provides a state-of-the-art solution for learning nonlinear SSMs.