UPPSALA
UNIVERSITET

Sequential Monte Carlo methods

Lecture 12 – Particle Matropolis Hastings

Thomas Schön, Uppsala University

2021-08-18

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

Bayesian parameter inference in SSMs

Full probabilistic model of a nonlinear parametric SSM:

$$p(x_{1:T}, \theta, y_{1:T}) = p(y_{1:T} \mid x_{1:T}, \theta) \underbrace{p(x_{1:T}, \theta)}_{\text{prior}}$$

$$= \underbrace{\prod_{t=1}^{T} p(y_t \mid x_t, \theta)}_{\text{observation}} \underbrace{\prod_{t=1}^{T-1} \underbrace{p(x_{t+1} \mid x_t, \theta)}_{\text{dynamics}} \underbrace{p(x_1 \mid \theta)}_{\text{state}} \underbrace{p(\theta)}_{\text{param.}}}_{\text{prior}}$$

Bayesian **parameter** inference amounts to computing

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

or more commonly some integral of the form

$$\mathbb{E}[\varphi(\theta) \mid y_{1:T}] = \int \varphi(\theta) p(\theta \mid y_{1:T}) \mathrm{d}\theta.$$

Using MH for parameter inference in a dynamical system

Algorithm 1 Metropolis Hastings (MH)

1. **Initialize:** Set the initial state of the Markov chain $\theta[1]$.
2. **For** $m = 1$ **to** M**, iterate:**
   a. Sample $\theta' \sim q(\theta \mid \theta[m])$.
   b. Sample $u \sim \mathcal{U}[0, 1]$.
   c. Compute the acceptance probability
   $$\alpha = \min\left(1, \frac{p(y_{1:T} \mid \theta')p(\theta')}{p(y_{1:T} \mid \theta[m])p(\theta[m])} \frac{q(\theta[m] \mid \theta')}{q(\theta' \mid \theta[m])}\right)$$
   d. Set the next state $\theta[m+1]$ of the Markov chain according to
   $$\theta[m+1] = \begin{cases} \theta' & u \leq \alpha \\ \theta[m] & \text{otherwise} \end{cases}$$

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.

Acceptance probability

General:

$$\alpha = \min\left(1, \frac{\pi(\theta')}{\pi(\theta[m])} \frac{q(\theta[m] \mid \theta')}{q(\theta' \mid \theta[m])}\right)$$

For our current auxiliary variable construction:

$$\begin{aligned}
\alpha &= \min\left(1, \frac{\pi(\theta', z' \mid y_{1:T})}{\pi(\theta[m], z[m] \mid y_{1:T})} \frac{q(\theta[m] \mid \theta')\psi(z[m] \mid \theta[m], y_{1:T})}{q(\theta' \mid \theta[m])\psi(z' \mid \theta', y_{1:T})}\right) \\
&= \min\left(1, \frac{z' p(\theta')\psi(z' \mid \theta', y_{1:T})}{z[m], p(\theta[m])\psi(z[m], \mid \theta[m], y_{1:T})} \frac{q(\theta[m] \mid \theta')\psi(z[m] \mid \theta[m], y_{1:T})}{q(\theta' \mid \theta[m])\psi(z' \mid \theta', y_{1:T})}\right) \\
&= \min\left(1, \frac{z' p(\theta')}{z[m], p(\theta[m])} \frac{q(\theta[m] \mid \theta')}{q(\theta' \mid \theta[m])}\right)
\end{aligned}$$

Particle Metropolis Hastings

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

Algorithm 2 Particle 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.

Explanation for 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 $p(\theta \mid y_{1:T})$.

Using the PMH for smoothing

A possibly under-appreciated fact is that the PMH algorithm provides a **solution to the smoothing problem as well**!

In step 2b, when we run the PF, select one of the state trajectories according to its weights and store that together with the unbiased likelihood estimate.

We then accept or reject the new parameter, the likelihood estimate and the state trajectory is step 2c.

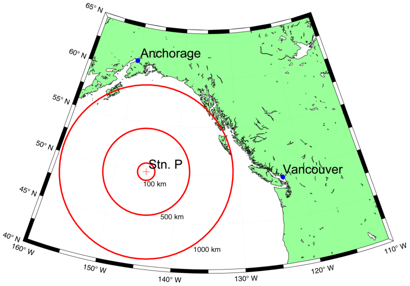
Examples

ex) Nonlinear marine biogeochemical model

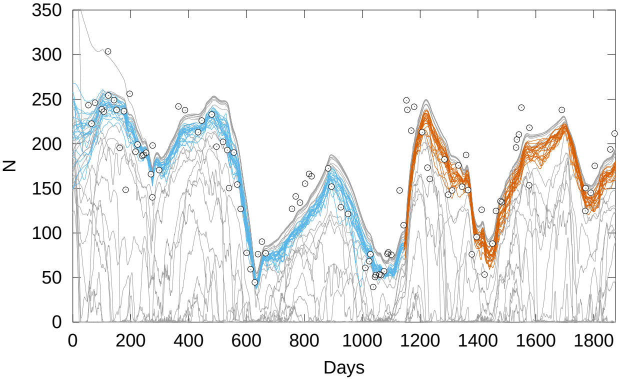
Studies nitrogen through an ecosystem in four compartments:

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.

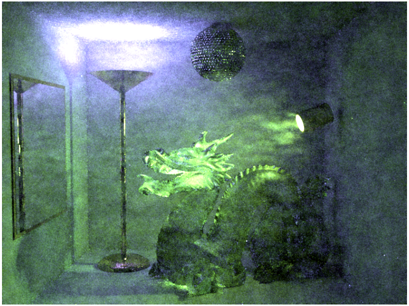
John Parslow, Noel Cressie, Edward P. Campbell, Emlyn Jones and Lawrence Murray. **Bayesian learning and predictability in a stochastic nonlinear dynamical model.** *Ecological Applications*, 23(4): 679–698, 2013.

ex) The pseudo-marginal idea is general

CG example (rendering images in heterogeneous media): An MH algorithm 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 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.** *Mechanical Systems and Signal Processing (MSSP)*, 104:866–883, 2018.

An implementation-focused tutorial on PMH:

Johan Dahlin and Thomas B. Schön. **Getting started with particle Metropolis-Hastings for inference in nonlinear dynamical models.** *Journal of Statistical Software (JSS)*, 88:1–41, 2019.

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.

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.