Sequential Monte Carlo methods

Lecture 14 – Particle Gibbs

Fredrik Lindsten, Linköping University

2021-08-19

Outline – Lecture 14

Aim: Present the particle Gibbs algorithm; a systematic method for combining particle filters and MCMC within a Gibbs sampling framework.

Outline:

1. The conditional importance sampling kernel
2. The particle Gibbs kernel
3. Ancestor sampling

Gibbs sampling for nonlinear dynamical systems

Gibbs sampling for dynamical system:

- Draw $\theta^\star \sim p(\theta \mid x_{0:T}, y_{1:T})$, **OK!**
- Draw $x_{0:T}^\star \sim p(x_{0:T} \mid \theta^\star, y_{1:T})$. **Hard!**

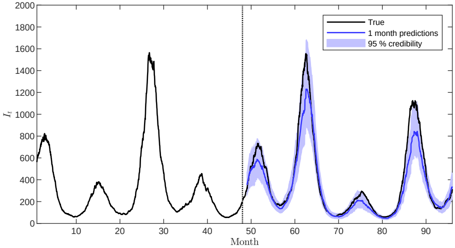
Problem: $p(x_{0:T} \mid \theta, y_{1:T})$ not available!

Idea: Approximate $p(x_{0:T} \mid \theta, y_{1:T})$ using a particle filter?

Better idea: Sample $x_{0:T}^\star$ from a Markov kernel $\kappa_{N,\theta}$, constructed using a particle filter, which has $p(x_{0:T} \mid \theta, y_{1:T})$ as a stationary distribution!

Ex) Epidemiological modelling

Compute $p(\theta, x_{0:T} \mid y_{1:T})$, where $y_{1:T} = (y_1, y_2, \dots, y_T)$ and use it to compute the predictive distribution.



Disease activity (number of infected individuals I_t) over an eight year period.

Fredrik Lindsten, Michael I. Jordan and Thomas B. Schön. **Particle Gibbs with ancestor sampling.** *Journal of Machine Learning Research (JMLR)*, 15:2145-2184, June 2014.

The conditional importance sampling kernel

Conditional importance sampling

Simplified setting: Given a target distribution $\pi(x)$, construct a Markov kernel $\kappa_N(x, x^\star)$—**using importance sampling**—which has stationary distribution $\pi(x)$.

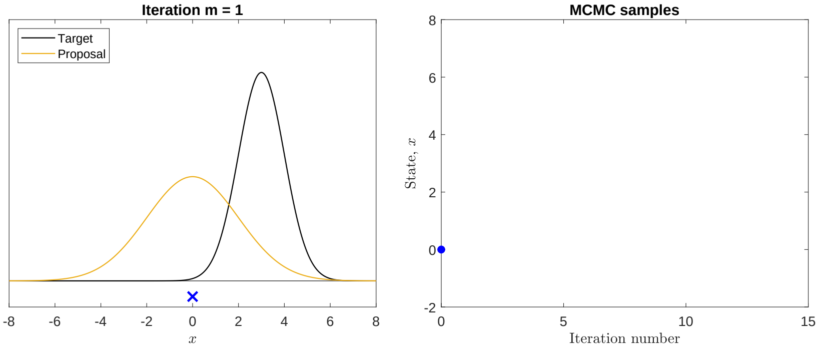
Markov kernel = stochastic procedure with input x and output $x^\star$.

We define $\kappa_N(x, x^\star)$ by the following "conditional importance sampler":

Input: x

1. Draw $x^i \sim q(x)$, $i = 1, \ldots, N-1$
2. Set $x^N = x$
3. Compute $\widetilde{w}^i = \frac{\widetilde{\pi}(x^i)}{q(x^i)}$, $i = 1, \ldots, N$ and normalize: $w^i = \frac{\widetilde{w}^i}{\sum_{j=1}^N \widetilde{w}^j}$
4. Draw $b \sim \mathcal{C}(\{w^i\}_{i=1}^N)$
5. **Output:** $x^\star = x^b$

ex) Conditional importance sampling illustration



Validity of the CIS kernel

Thm (stationarity): $\kappa_N(x, x^\star)$ defined by the conditional importance sampling procedure has $\pi(x)$ as stationary distribution for any $N \geq 1$.

Thm (ergodicity): Under weak conditions, samples generated from $\kappa_N(x, x^\star)$ converge (at a geometric rate) towards $\pi(x)$ from any starting point.

(Proofs as exercises in the practical session.)

Particle Gibbs

The particle Gibbs kernel

Let θ be fixed and let $x_{0:T} \in \mathcal{X}^{T+1}$ be a given state trajectory, denoted the **reference trajectory**.

We want to construct a Markov kernel $\kappa_{N,\theta}(x_{0:T}, x^\star_{0:T})$ on $\mathcal{X}^{T+1}$.

Particle Gibbs: Run a particle filter, but at each time step

- sample only $N-1$ particles in the standard way.
- set the Nth particle deterministically: $x_t^N = x_t$ and $a_t^N = N$.
- At final time $t = T$, output $x^\star_{0:T} = x^b_{0:T}$ with $b \sim \mathcal{C}(\{w_T^i\}_{i=1}^N)$

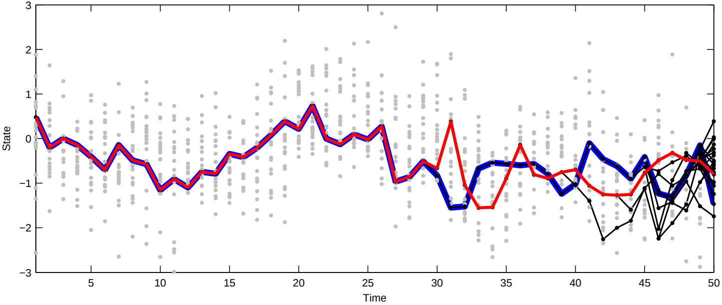
The particle Gibbs algorithm

Algorithm Bootstrap particle filter

1. **Initialize ($t = 0$):**
   - Draw $x_0^i \sim p(x_0)$, $i = 1, \ldots, N$.
     Set $x_0^N = x_0$.
   - Set $w_0^i = \frac{1}{N}$, $i = 1, \ldots, N$.
2. **for $t = 1$ to T:**
   - Draw $a_t^i \sim \mathcal{C}(\{w_{t-1}^j\}_{j=1}^N)$, $i = 1, \ldots, N-1$.
   - Draw $x_t^i \sim p(x_t \mid x_{t-1}^{a_t^i}, \theta)$, $i = 1, \ldots, N-1$.
     Set $x_t^N = x_t$ and $a_t^N = N$.
   - Set $\widetilde{w}_t^i = p(y_t \mid x_t^i)$, $i = 1, \ldots, N$.
   - Normalize the weights: $w_t^i = \frac{\widetilde{w}_t^i}{\sum_{\ell=1}^N \widetilde{w}_t^\ell}$, $i = 1, \ldots, N$.

Draw $b \sim \mathcal{C}(\{w_T^i\}_{i=1}^N)$ and output $x_{0:T}^\star = x_{0:T}^b$.

The particle Gibbs kernel



- The algorithm stochastically "maps" $x_{0:T}$ into $x^{\star}_{0:T}$.
- Implicitly defines a Markov kernel $\kappa_{N,\theta}(x_{0:T}, x^{\star}_{0:T})$ on $\mathcal{X}^{T+1}$ — **the particle Gibbs kernel**.

Particle Gibbs sampler for SSMs

Algorithm Particle Gibbs for parameter inference in SSMs

1. **Initialize:** Set $x_{0:T}[1]$ and $\theta[1]$ arbitrarily
2. **For** $m = 2$ **to** M**, iterate:**
   a. Draw $\theta[m] \sim p(\theta \mid x_{0:T}[m-1], y_{1:T})$
   b. Draw $x_{0:T}[m] \sim \kappa_{N,\theta[m]}(x_{0:T}[m-1], x^{\star}_{0:T})$

Some properties of the particle Gibbs kernel

Validity

Stationary distribution: $p(x_{0:T} \mid \theta, y_{1:T})$ is a stationary distribution of the particle Gibbs kernel, for any number of particles $N \geq 1$,

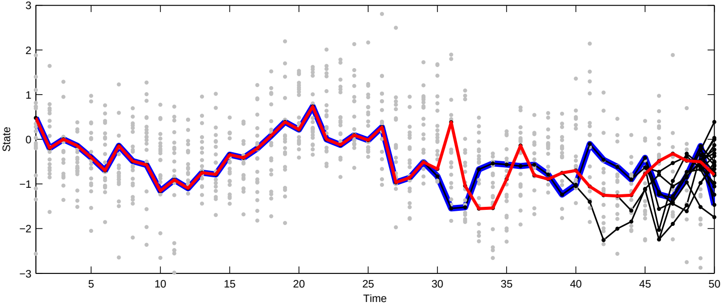
$$\int p(x_{0:T} \mid \theta, y_{1:T}) \kappa_{N,\theta}(x_{0:T}, x^\star_{0:T}) dx_{0:T} = p(x^\star_{0:T} \mid \theta, y_{1:T}).$$

Ergodicity: The particle Gibbs kernel is ergodic for any $N \geq 2$ under weak conditions.

...what's the catch?!?

Scaling: For the mixing rate **not to deteriorate** as T becomes large, we need (at least) $N \propto T$.

Path degeneracy for particle Gibbs



The scaling $N \propto T$ is required to tackle the path degeneracy, since otherwise

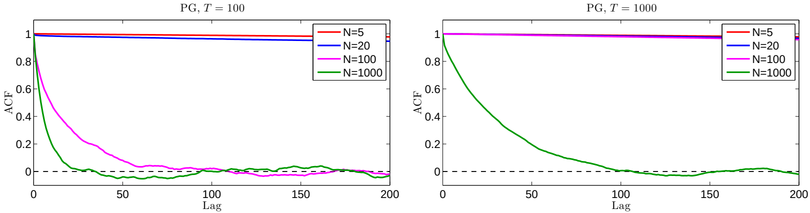
$$\mathbb{P}(X_t^\star \neq x_t) \to 0, \qquad \text{as } T - t \to \infty.$$

ex) Particle Gibbs

Stochastic volatility model,

$$X_{t+1} = 0.9X_t + V_t, \qquad V_t \sim \mathcal{N}(0, \Theta),$$
$$Y_t = E_t \exp\left(\tfrac{1}{2}X_t\right), \qquad E_t \sim \mathcal{N}(0, 1).$$

Consider the ACF of $\theta[m] - \mathbb{E}[\Theta \mid y_{1:T}]$.



Ancestor sampling

Ancestor sampling

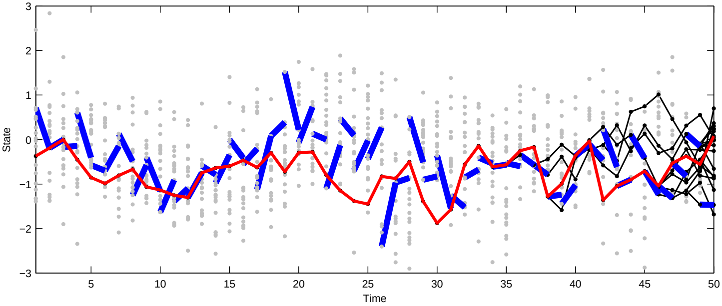
Particle Gibbs:

Let $x_{1:T} = (x_1, \ldots, x_T)$ be a fixed *reference trajectory*.

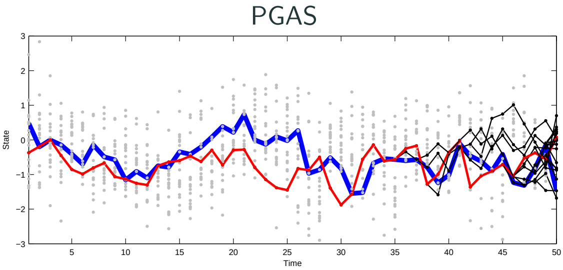
- Sample only $N-1$ particles in the standard way.
- Set the Nth particle deterministically: $x_t^N = x_t$.
- ~~Set $a_t^N = N$.~~
- Sample $a_t^N \in \{1, \ldots, N\}$ with

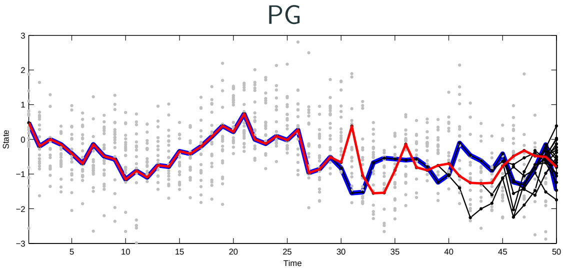
$$\mathbb{P}\left(A_t^N = j\right) \propto w_{t-1}^j p(x_t \mid x_{t-1}^j, \theta).$$

Particle Gibbs with ancestor sampling



PGAS vs. PG


PGAS
3
2
1
0
−1
−2
−3
State
5
10
15
20
25
30
35
40
45
50
Time

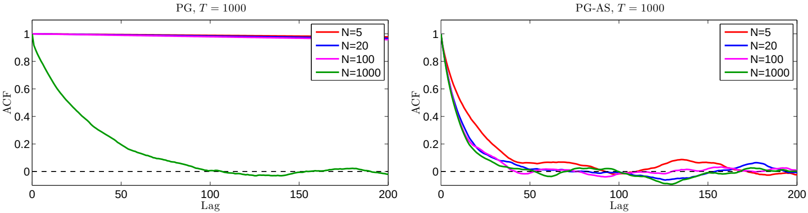

PG
3
2
1
0
−1
−2
−3
State
5
10
15
20
25
30
35
40
45
50
Time

ex) Stochastic volatility model, cont'd

Stochastic volatility model,

$$X_{t+1} = 0.9 X_t + V_t, \qquad V_t \sim \mathcal{N}(0, \Theta),$$
$$Y_t = E_t \exp\left(\tfrac{1}{2} X_t\right), \qquad E_t \sim \mathcal{N}(0, 1).$$

Consider the ACF of $\theta[m] - \mathbb{E}[\Theta \mid y_{1:T}]$.



Practical considerations

Resampling in the conditional particle filter

Care needs to be taken when implementing the resampling step in the conditional particle filter!

Common implementation:

```
a[t,:] <- resampling(w[t-1,:])
a[t,N] <- N
a[t,N] <- N
```

Low-variance resampling (stratified, systematic, ...) can be used, but require special implementations to maintain the correct stationary distribution:

Nicolas Chopin and Sumeetpal S. Singh. **On Particle Gibbs Sampling.** *Bernoulli*, 21:1855–1883, 2015.

Modularity of particle Gibbs

The **"kernel view"** on particle Gibbs is useful for constructing composite MCMC schemes.

ex) Multiple dynamical systems

Multiple time series/state space models with shared parameter θ.

Θ

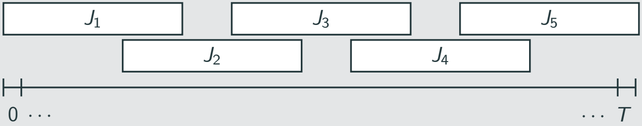
Model 1:	$\cdots$	Model n:
$X_t = f_1(X_{t-1}, V_t, \Theta)$, $Y_t = g_1(X_t, E_t, \Theta)$.	$\cdots$	$X_t = f_n(X_{t-1}, V_t, \Theta)$, $Y_t = g_n(X_t, E_t, \Theta)$.

1. Use a separate particle Gibbs kernel for each model.
2. Update the parameter using all the models' state variables.

Modularity of particle Gibbs

The **"kernel view"** on particle Gibbs is useful for constructing composite MCMC schemes.

ex) Temporal blocking



- Consider a sub-block of states $x_{s:u}$
- Sample $x^{\star}_{s:u} \sim \kappa^{s:u}_{N,\theta}(x_{0:T}, x^{\star}_{s:u})$ where $\kappa^{s:u}_{N,\theta}$ is a particle Gibbs kernel targeting $p(x_{s:u} \mid \theta, x_{s-1}, x_{u+1}, y_{s:u})$.
- Only need $N \propto |u - s|$

Sumeetpal S. Singh, Fredrik Lindsten, and Eric Moulines. **Blocking Strategies and Stability of Particle Gibbs Samplers.** *Biometrika*, 104:953–969, 2017.

A few concepts to summarize lecture 14

Particle Gibbs kernel: A Markov kernel on the space of state trajectories, constructed using a particle filter, which has the exact joint smoothing distribution as its stationary distribution.

Modularity of particle Gibbs: The particle Gibbs kernel can be used as a plug-and-play component in other MCMC schemes.

Ancestor sampling: A simple modification of the particle Gibbs construction, in which the ancestor indices of the input particles are sampled anew at each time step of the underlying particle filter. This mitigates the effect of path degeneracy and can therefore (significantly) improve the ergodicity of the particle Gibbs kernel.