Outline – Lecture 15

Aim: Show how SMC can be used for a much wider class of problems than inference in state-space models.

Outline:
1. Summary of day 3
2. Examples of probabilistic models
3. General SMC formulation
4. Locally optimal proposals

Summary of day 3

Simulate a Markov chain which is designed in such a way that its stationary distribution coincides with the target distribution.

An MCMC sampler generates the Markov chain \( \{x[m]\}_{m=1}^M \) by:

- **Initialize:** set \( x[1] \) arbitrarily.

- **For** \( m = 2 \) **to** \( M \): sample \( x[m] \sim \kappa(x[m-1], x^*) \).

\( \kappa(x, x^*) \) is a **Markov kernel** on \( \mathcal{X} \), i.e. a conditional distribution for the next state \( x^* \) given the current state \( x \).
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 $\theta' \sim q(\theta | \theta[m-1])$.
   b. Sample $\hat{z}' \sim \psi(z | \theta', y_{1:T})$ (i.e. run a particle filter).
   c. With probability
      $$\alpha = \min \left(1, \frac{\hat{z}' p(\theta')}{\hat{z}[m-1] p(\theta[m-1])} \frac{q(\theta[m-1] | \theta')}{q(\theta' | \theta[m-1])}\right)$$
      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).

Gibbs kernel: This procedure defines a Markov kernel $\kappa(x, x^*)$ with stationary distribution $\pi(x)$.

Particle Gibbs kernel: A Markov kernel $\kappa_{N, \theta}(x_{0:T}, \hat{x}_{0:T})$ 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 $N$th particle deterministically: $x^N_t = x_t$ and $a^N_t = N$.
- At final time $t = T$, output $x^*_{0:T} = x^0_{0:T}$ with
  $b \sim C\left(\{w^T_i\}_{i=1}^N\right)$

Examples of probabilistic models

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

A phylogenetic (evolutionary) tree shows the inferred evolutionary relationships among various species based upon similarities and differences in their physical or genetic characteristics.

Probabilistic graphical models

A probabilistic graphical model (PGM) is a probabilistic model where a graph $G = (V, E)$ represents the conditional independency structure between random variables,

1. a set of vertices $V$ (nodes) represents the random variables
2. a set of edges $E$ containing elements $(i, j) \in E$ connecting a pair of nodes $(i, j) \in V \times V$

Gaussian process state-space model

The Gaussian process (GP) is a non-parametric and probabilistic model for nonlinear functions.

Non-parametric means that it does not rely on any particular parametric functional form to be postulated.

$$X_t = f(X_{t-1}) + V_t, \quad \text{s.t.} \quad f(X) \sim GP(0, \kappa_\eta, f(x, x')),$$
$$Y_t = g(X_t) + E_t, \quad \text{s.t.} \quad g(X) \sim GP(0, \kappa_\eta, g(x, x')).$$

The model functions $f$ and $g$ are assumed to be realizations from Gaussian process priors and $V_t \sim \mathcal{N}(0, Q)$, $E_t \sim \mathcal{N}(0, R)$.

Task: Compute the posterior $p(f, g, Q, R, \eta, x_0:T | y_1:T)$. 

General SMC formulation
Model specification

SMC can be used to approximate a sequence of probability distributions on a sequence of probability spaces of increasing dimension.

Let $\{\pi_k(x_{1:k})\}_{k \geq 1}$ be an arbitrary sequence of target distributions

$$\pi_k(x_{1:k}) = \frac{\tilde{\pi}_k(x_{1:k})}{Z_k}$$

- The domain of $x_k$ is $\mathcal{X}_k$, and $\mathcal{X}_{1:k} = \mathcal{X}_k \times \mathcal{X}_{1:k-1}$ for all $k$.
- $\tilde{\pi}_k(x_{1:k})$ can be evaluated pointwise.
- The normalizing constant $Z_k$ may be unknown.

Common tasks:
1. Approximate the normalization constant $Z_k$.
2. Approximate $\pi_k(x_k)$ and compute $\int \phi(x_k)\pi_k(x_k)dx_k$.

General Sequential Monte Carlo

Sequential Monte Carlo approximates

$$\pi_k(x_{0:k}) \approx \frac{1}{N} \sum_{i=1}^{N} w^i_k \delta_{x^i_{0:k}}(x_{0:k}).$$

The weighted particle populations $\{x^i_{0:k}, w^i_k\}_{i=1}^{N}$ are generated sequentially for $k = 1, 2, \ldots$

ex) State space model

The sequence of target distributions $\{\pi_k(x_{0:k})\}_{k=1}^{n}$ can be constructed in many different ways.

The most basic construction arises from chain-structured graphs, such as the state space model (SSM).

$$
\begin{align*}
\pi_{1:t}(x_{0:t}) &= \frac{\pi_{1:t}(x_{0:t})}{p(x_{0:t} | y_{1:t})} = \frac{p(x_{0:t}, y_{1:t})}{p(y_{1:t})} \\
Z_t &= \int \tilde{\pi}(x_{0:t})dx_{0:t}
\end{align*}
$$

General Sequential Monte Carlo

Assume that we have obtained $\{x^i_{0:k-1}, w^i_{k-1}\}_{i=1}^{N}$.

**Resampling:** Sample $a^i_k$ with $\mathbb{P}(a^i_k = j) = \nu^i_k$, $j = 1, \ldots, N$.

**Propagation:** $x^i_k \sim q_k(x_k | x^i_{1:k-1})$ and $x^i_{0:k} = (x^i_{0:k-1}, x^i_k)$.

**Weighting:** $w^i_k \propto \frac{w^i_{k-1}}{\nu^i_{k-1}} \frac{\tilde{\pi}_k(x^i_k)}{\tilde{\pi}_k(x^i_{0:k-1})}$.

The result is a new weighted set of particles $\{x^i_{0:k}, w^i_k\}_{i=1}^{N}$.
**SMC for probabilistic graphical models**

Recall – Probabilistic graphical models

A **probabilistic graphical model** (PGM) is a probabilistic model where a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents the conditional independency structure between random variables,

1. a set of vertices $\mathcal{V}$ (nodes) represents the random variables
2. a set of edges $\mathcal{E}$ containing elements $(i, j) \in \mathcal{E}$ connecting a pair of nodes $(i, j) \in \mathcal{V} \times \mathcal{V}$

Key idea

SMC methods are used to approximate a sequence of probability distributions on a sequence of spaces of increasing dimension.

**Key idea:**

1. Introduce a sequential decomposition of the PGM.
2. Each subgraph induces an intermediate target dist.
3. Apply SMC to the sequence of intermediate target dist.

Using an artificial sequence of intermediate target distributions for an SMC method is a powerful (quite possibly underutilized) idea.

**Key question:** Exactly how do we define $\tilde{\pi}_k(x_{1:k})$?

**ex) Illustrating possible graph decomposition**

Using a 2D lattice model from statistical physics, $x \in (-\pi, \pi]$. The intermediate sequence of target distributions can be chosen

$$
\tilde{\pi}_k(x_{L_k}) \propto \tilde{\pi}_{k-1}(x_{L_{k-1}}) e^{K(x_{L_{k-1}}) \cos(x_k - \mu(x_{L_{k-1}}))}.
$$

$L_k$ – index to the nodes in the $k^{\text{th}}$ intermediate target $\tilde{\pi}_k(x_{L_k})$. 

\[ p(x_{1:k}) \propto e^{-\beta H(x_{1:k})}, \quad H(x_{1:k}) = - \sum_{(i,j) \in \mathcal{E}} J_{ij} \cos(x_i - x_j), \]
**SMC for graphical models – algorithm**

**Algorithm SMC for graphical models**

1. **Initialize** \((k = 1)\):
   (a) Draw \(x_{L1}^{i} \sim q_1(\cdot)\).
   (b) Set \(w_i^1 = W_1(x_{L1}^i)\).

2. **For** \(k = 2\) to \(K\) **do**:
   (a) **Resampling**: Draw \(a_i^k, P(a_i^k = j) = \tilde{w}_j^{k-1}/\sum_l \tilde{w}_l^{k-1}\).
   (b) **Propagation**: Draw \(\xi_i^k \sim q_k(\cdot| x_{L1}^{i-1}, x_{Lk}^i = x_{Lk-1}^i \cup \xi_i^k)\).
   (c) **Weighting**: Set \(\tilde{w}_i^k = W_k(x_{Lk}^i)\).

3. **End**

\(L_k\) – index to the nodes in the \(k^{th}\) intermediate target \(\tilde{\pi}_k(x_{L_k})\).

\(\xi_k^i\) – nodes added at step \(k\).

Also provides an unbiased estimate of the **normalizing constant**! 16/19

---

**What about stability?**

**Recall**: for a state-space model we need exponential forgetting for the particle filter to be stable.

**The same is true in the general case!**

If there are **strong** and **long-ranging** dependencies among the variables \(X_{1:k}\) under the distribution \(\pi_k\), then the asymptotic variance of SMC may be exponential in \(k\).

However,
- In many applications we do have fast enough forgetting (though, it can be difficult to verify theoretically)
- Even if this is not the case, SMC can give good results for moderate values of \(k\)

---

**Further reading**

**SMC for phylogenetic trees:**

**SMC for graphical models:**

---

**ex) Classical XY-model**

This model is borrowed from

17/19

---

**Further reading**

**SMC for phylogenetic trees:**

**SMC for graphical models:**

19/19