Introduction to Sequential Monte Carlo Methods

Haozhe Sun

August 30, 2023

Agenda

Introduction

What is Sequential Monte Carlo?

- The Sequential Monte Carlo Method, often known as particle filtering, is a computational technique used for estimating the state of a dynamic system by simulating a set of particles and recursively updating their weights and positions based on observational data.

Why is it Important?

- Widely used in fields such as robotics, signal processing, and machine learning.
- Essential for real-time tracking and prediction tasks.

Objective of This Presentation

- To provide an understanding of the fundamental principles of SMC.

Model Formulation

State and Observation Sequences:
Let $\{x_t\}_{t=0}^{\infty}$ denote the sequence of unobserved states, with $x_t \in \mathcal{X}$. Let $\{y_t\}_{t=1}^{\infty}$ denote a sequence of observations with $y_t \in \mathcal{Y}$.

Notation:
Define $x_{0:t} \equiv \{x_0, \ldots, x_t\}$ and $y_{1:t} \equiv \{y_1, \ldots, y_t\}$.

Model Assumptions:

- $\{x_t\}$ is a Markov process with initial distribution $p(x_0)$.
- Given $\{x_t\}$, the observations are conditionally independent.

Model Formulation

Equations:

$$
\begin{aligned}
p(x_t|x_{t-1}) &: \text{Transition equation} \\
p(y_t|x_t) &: \text{Observation equation}
\end{aligned}
$$

Objective:
Estimate the posterior distribution $p(x_{0:t}|y_{1:t})$ recursively. We may also care about $p(x_t|y_{1:t})$ or expectations such as

$$
I(h_t) \equiv \int h_t(x_{0:t})p(x_{0:t}|y_{1:t})dx_{0:t}
$$

for some $h_t : \mathcal{X}^t \to \mathbb{R}^n$ that is integrable with respect to $p(x_{0:t}|y_{1:t})$.

Model Formulation

Bayes' Theorem:

$$p(x_{0:t}|y_{1:t}) = \frac{p(y_{1:t}|x_{0:t})p(x_{0:t})}{\int p(y_{1:t}|x_{0:t})p(x_{0:t})dx_{0:t}}$$

Recursive Relationship:

The posterior distribution at any $t+1$ can be expressed in terms of $p(x_{0:t}|y_{1:t})$:

$$\begin{aligned} p(x_{0:t+1}|y_{1:t+1}) &= \frac{p(x_{0:t+1}, y_{1:t+1})}{p(y_{1:t+1})} \\ &= \frac{p(x_{t+1}, y_{t+1}|x_{0:t}, y_{1:t})p(x_{0:t}, y_{1:t})}{p(y_{t+1}|y_{1:t})p(y_{1:t})} \\ &= \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{p(y_{t+1}|y_{1:t})}p(x_{0:t}|y_{1:t}) \end{aligned}$$

The expression in the denominator is constant with respect to $x_{0:t+1}$.

Model Formulation

Joint Distribution Recursion:

$$p(x_{0:t+1}|y_{1:t+1}) = p(x_{0:t}|y_{1:t})\frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{p(y_{t+1}|y_{1:t})}$$

Marginal Distribution Recursion:

$$\text{Prediction:} \quad p(x_t|y_{1:t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1})dx_{t-1}$$

$$\text{Updating:} \quad p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})dx_t}$$

Model Formulation

Computational Complexity:

These expressions and recursions are deceptively simple because one cannot typically compute:

- The normalising constant $p(y_{1:t})$
- The marginals of the posterior $p(x_{0:t}|y_{1:t})$, in particular $p(x_t|y_t)$
- $I(f_t)$ as they require the evaluation of complex high-dimensional integrals.

Perfect Monte Carlo Sampling

- Assume N independent and identically distributed (i.i.d.) random samples (particles) $\left\{x_{0:t}^{(i)}; i = 1, \ldots, N\right\}$ according to $p(x_{0:t} \mid y_{1:t})$.
- Empirical estimate:

$$P_N(dx_{0:t} \mid y_{1:t}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{0t}^{(i)}}(dx_{0:t})$$

- Estimate of $I(f_t)$:

$$\begin{aligned} I_N(f_t) &= \int f_t(x_{0:t}) P_N(dx_{0:t} \mid y_{1:t}) \\ &= \frac{1}{N} \sum_{i=1}^{N} f_t(x_{0:t}^{(i)}) \end{aligned}$$

Perfect Monte Carlo Sampling

- Variance of $I_N(f_t)$ if posterior variance satisfies $\sigma_{f_t}^2 \triangleq \mathbb{E}_{p(x_{0t}|y_{1t})}\left[f_t^2\left(x_{0:t}\right)\right] - I^2\left(f_t\right) < +\infty$:

$$\text{var}(I_N(f_t)) = \frac{\sigma_{f_t}^2}{N}$$

- Strong Law of Large Numbers:

$$I_N(f_t) \underset{N \to +\infty}{\overset{a.s}{\to}} I(f_t)$$

- Central Limit Theorem:

$$\sqrt{N}[I_N(f_t) - I(f_t)] \underset{N \to +\infty}{\Longrightarrow} \mathcal{N}(0, \sigma_{f_t}^2)$$

- Advantage: Rate of convergence is independent of the dimension of the integrand.

Limitations of Perfect Monte Carlo Sampling

- Usually impossible to sample efficiently from the posterior distribution $p(x_{0:t} \mid y_{1:t})$.
- $p(x_{0:t} \mid y_{1:t})$ is multivariate, non-standard, and only known up to a proportionality constant.
- Markov Chain Monte Carlo (MCMC) methods are often used in applied statistics for complex distributions.
- MCMC methods are iterative and not well-suited for recursive estimation problems.

Importance Sampling

- We can't draw directly from $p(x_{0:t}|y_{1:t})$.
- Use importance sampling with a density $\pi(x_{0:t}|y_{1:t})$ and corresponding importance weight.

$$
\begin{aligned}
I(h_t) &\equiv E[h_t(x_{0:t})|y_{1:t}] \\
&= \frac{\int h_t(x_{0:t})p(x_{0:t}|y_{1:t})dx_{0:t}}{\int p(x_{0:t}|y_{1:t})dx_{0:t}} \\
&= \frac{\int h_t(x_{0:t})w(x_{0:t})\pi(x_{0:t}|y_{1:t})dx_{0:t}}{\int w(x_{0:t})\pi(x_{0:t}|y_{1:t})dx_{0:t}}
\end{aligned}
$$

where

$$
w(x_{0:t}) \equiv \frac{p(x_{0:t}|y_{1:t})}{\pi(x_{0:t}|y_{1:t})}
$$

is the importance weight.

Importance Sampling

- Draw a sample $\{x_{0:t}^{(i)}\}_{i=1}^{N}$ from $\pi(x_{0:t}|y_{1:t})$.

$$
\begin{aligned}
I(h_t) &= \frac{\frac{1}{N}\sum_i h_t(x_{0:t}) w(x_{0:t}^{(i)})}{\frac{1}{N}\sum_i w(x_{0:t}^{(i)})} \\
&= \sum_i h_t(x_{0:t}) \widetilde{w}_t^{(i)}
\end{aligned}
$$

where

$$
\widetilde{w}_t^{(i)} \equiv \frac{w(x_{0:t})}{\sum_j w(x_{0:t}^{(j)})}
$$

are the normalized importance weights.

Importance Sampling

- The approximate posterior is given by:

$$\hat{P}_N(dx_{0:t}|y_{1:t}) = \sum_i w_t^{(i)} \delta_{x_{0:t}}(dx_{0:t})$$

- Approximation of $I(h_t)$ can be done as:

$$\hat{I}(h_t) = \int h_t(x_{0:t}) \hat{P}_N(dx_{0:t}|y_{1:t})$$

- Limitation:
  - This method is not well-suited for recursive problems.

Sequential Importance Sampling

Objective: Modify the importance sampling method to compute an estimate $\hat{P}_N(dx_{0:t}|y_{1:t})$ without altering past simulated trajectories $\{x_{0:t-1}^{(i)}; i = 1, \ldots, N\}$.

Importance Function

The importance function $\pi(x_{0:t}|y_{1:t})$ at time t has as its marginal distribution at $t-1$ the importance function $\pi(x_{0:t-1}|y_{1:t-1})$:

$$\pi(x_{0:t}|y_{1:t}) = \pi(x_{0:t-1}|y_{1:t-1})\pi(x_t|x_{0:t-1}, y_{1:t})$$

Iterative Formula

Iteratively, one obtains:

$$\pi(x_{0:t}|y_{1:t}) = \pi(x_0) \prod_{k=1}^{t} \pi(x_k|x_{0:k-1}, y_{1:k})$$

Sequential Importance Sampling

- Calculate the importance weights recursively:

$$\widetilde{w}_t^{(i)} \propto \widetilde{w}_{t-1}^{(i)} \frac{p\left(\mathbf{y}_t \mid \mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)} \mid \mathbf{x}_{t-1}^{(i)}\right)}{\pi\left(\mathbf{x}_t^{(i)} \mid \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t}\right)}.$$

Sequential Importance Sampling

- A special case occurs when:

$$\pi(x_{0:t}|y_{1:t}) = p(x_{0:t}) = p(x_0)\prod_{k=1}^{t} p(x_k|x_{k-1})$$

- Resulting in:

$$w(x_{0:t}) \propto w(x_{0:t-1})p(y_t|x_t)$$

$$\tilde{w}_t^{(i)} \propto \tilde{w}_{t-1}^{(i)} p(y_t|x_t^{(i)})$$

- Efficiency issues:

SIS is usually inefficient for high-dimensional integrals because as $t \to \infty$, the importance weights for some particles quickly approach zero.

The Bootstrap Filter

- To prevent particle degeneracy, the bootstrap filter introduces a resampling step.
- Eliminates particles with low importance weights.

Uniformly-Weighted Distribution

$$P_N(dx_{0:t}|y_{1:t}) = N^{-1} \sum_i N(i)_t \delta_{x(i)_{0:t}}(dx_{0:t}),$$

where $N(i)_t$ is the number of offspring of the particle $x(i)_{0:t}$.

The Bootstrap Filter

- Most common mechanism involves resampling N times from $\hat{P}_N$ (Gordon et al., 1993).
- $\sum_i N(i)_t = N$ for all t.
- If $N(j)_t = 0$, the particle $x(j)_{0:t}$ dies.

Objective

$$\int h_t(x_{0:t}) P_N(dx_{0:t}|y_{1:t}) \approx \int h_t(x_{0:t}) \hat{P}_N(dx_{0:t}|y_{1:t}).$$

Surviving particles are approximately distributed according to $p(x_{0:t}|y_{1:t})$.

The Bootstrap Filter

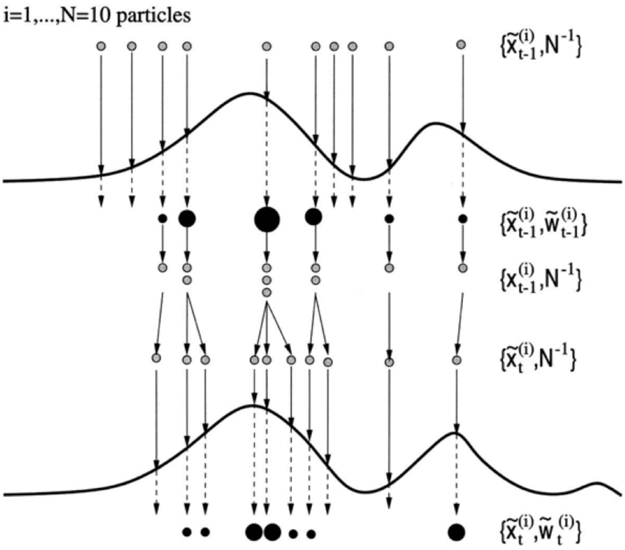
- After the selection step, the surviving particles $\tilde{x}$, that is the ones with $N_t C_t > 0$, are approximately distributed according to $p(x_{0:t}|y_{1:t})$.
- There are many different ways to select $N_t C_t$, the most popular being the one introduced in (Gordon et al. 1993).
- Here, one obtains the surviving particles by sampling N times from the (discrete) distribution $P_N(dx_{0:t}|y_{1:t})$.
- This is equivalent to sampling the number of offspring $N_t C_t$ according to a multinomial distribution of parameters $\tilde{w}_t$.

Algorithm

Algorithm 1 Bootstrap Filter Algorithm

Initialization
for $i = 1, \ldots, N$ **do**
    Draw $x_0^{(i)} \sim p(x_0)$
end for
Importance Sampling Step
for $i = 1, \ldots, N$ **do**
    Draw $\tilde{x}_t^{(i)} \sim p(x_t | x_{t-1}^{(i)})$
    Set $\tilde{x}_{0:t}^{(i)} \leftarrow (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)})$
    Calculate $w_t^{(i)} = p(y_t | \tilde{x}_t^{(i)})$ and normalize
end for
Resampling Step
Take N draws $\{x_{0:t}^{(i)}\}_{i=1}^N$ with replacement from $\{\tilde{x}_{0:t}^{(i)}\}_{i=1}^N$ with weights
Set $t \leftarrow t + 1$ and go to Importance Sampling step.

Algorithm



i=1,...,N=10 particles
$\{\tilde{x}^{(i)}_{t-1}, N^{-1}\}$
$\{\tilde{x}^{(i)}_{t-1}, \tilde{w}^{(i)}_{t-1}\}$
$\{x^{(i)}_{t-1}, N^{-1}\}$
$\{\tilde{x}^{(i)}_{t}, N^{-1}\}$
$\{\tilde{x}^{(i)}_{t}, \tilde{w}^{(i)}_{t}\}$

References

Doucet, A., de Freitas, N., & Gordon, N. (2001).
An Introduction to Sequential Monte Carlo Methods.
In *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

JASON BLEVINS.
Introduction to Sequential Monte Carlo Methods.
https://jblevins.org/notes/smc-intro

N. Kantas, A. Doucet, S.S. Singh, J.M. Maciejowski (2009).
An Overview of Sequential Monte Carlo Methods for Parameter Estimation in General State-Space Models.
IFAC Proceedings Volumes, 42(10), 774-785.
https://doi.org/10.3182/20090706-3-FR-2004.00129

Thank you!