Sequential Self-Normalized Importance Sampling

Particle Filtering in Dynamical Systems

Sai Sampath Kedari
sampath@umich.edu

Contents

1 Objective

The objective of this report is to derive particle filtering from first principles as a *sequential self-normalized importance sampling* (SNIS) approximation to the exact Bayesian filtering recursion.

The derivation is carried out at the level of probability measures. No algorithmic shortcuts, model assumptions, or implementation choices are introduced.

Primary Goal

The primary goal is to demonstrate that the Bayesian filtering equations, which define a recursive update of probability distributions, admit a natural empirical approximation when self-normalized importance sampling is applied sequentially in time.

Under this viewpoint, particle filtering is not introduced as an algorithm, but as a *representation-level approximation* of Bayesian filtering.

Scope of the Derivation

This report focuses exclusively on the probabilistic foundation of particle filtering. Specifically, the derivation:

- Begins with the exact Bayesian filtering recursion,
- Introduces an empirical measure representation of the filtering distribution,
- Applies self-normalized importance sampling at each time step,
- Derives the resulting recursive update of normalized particle weights.

The end result is the construction of

Sequential Self-Normalized Importance Sampling for Bayesian Filtering.

What Is *Not* Addressed

To preserve conceptual clarity, this report intentionally excludes all algorithmic refinements and implementation-level considerations, including:

- Linear or nonlinear system specializations,
- Gaussian or non-Gaussian noise assumptions,
- Design of proposal distributions,
- Weight degeneracy diagnostics,
- Effective sample size (ESS),
- Resampling strategies,
- Bootstrap and optimal particle filters,
- Particle smoothing methods.

These topics rely on the derivation presented here and are treated in subsequent reports.

Perspective

This report establishes the first and irreducible step of particle filtering:

> The Bayesian filtering recursion remains exact. Only the representation of the filtering distribution is approximated.

Gaussian filtering approximates distributions using finite-dimensional parametric families. Particle filtering approximates distributions using finite empirical measures.

This report formalizes that transition.

2 Empirical Distributions via Self-Normalized Importance Sampling (Recap)

This section provides a concise recap of empirical distributions and self-normalized importance sampling (SNIS), with the sole purpose of establishing a practical and unified viewpoint:

> *A probability distribution can be approximated by a finite set of samples and normalized weights.*

This viewpoint underlies particle filtering.

2.1 Problem Setup

Let $f(x)$ denote a target probability density on a space $\mathcal{X}$. We are interested in evaluating expectations of the form

$$\mathbb{E}_{X\sim f}[h(X)] = \int h(x)\, f(x)\, dx,$$

for a measurable function $h : \mathcal{X} \to \mathbb{R}$.

Direct sampling from f is assumed to be infeasible. Instead, we introduce a proposal distribution $g(x)$ satisfying

$$f(x) > 0 \;\Rightarrow\; g(x) > 0.$$

2.2 Sampling and Importance Weights

We draw N samples from the proposal distribution:

$$x_1, x_2, \dots, x_N \sim g.$$

The mismatch between the proposal g and the target f is corrected using importance weights

$$w(x) = \frac{f(x)}{g(x)}.$$

When f is known only up to a normalizing constant, we work with unnormalized weights

$$\hat{w}_i = \frac{\hat{f}(x_i)}{g(x_i)}, \qquad i = 1, \dots, N,$$

and define normalized weights

$$\tilde{w}_i = \frac{\hat{w}_i}{\sum_{j=1}^N \hat{w}_j}, \qquad \sum_{i=1}^N \tilde{w}_i = 1.$$

2.3 SNIS as an Empirical Distribution

The samples $\{x_i\}_{i=1}^N$ together with the normalized weights $\{\tilde{w}_i\}_{i=1}^N$ define a discrete empirical distribution

$$\boxed{\hat{P}_N(dx) = \sum_{i=1}^{N} \tilde{w}_i \, \delta_{x_i}(dx).}$$

This empirical distribution assigns probability mass $\tilde{w}_i$ to each sampled point x_i, and zero probability elsewhere.

2.4 Expectations under the Empirical Distribution

For any function h, expectations under the empirical distribution are computed exactly:

$$\mathbb{E}_{X\sim\hat{P}_N}[h(X)] = \sum_{i=1}^{N} \tilde{w}_i \, h(x_i).$$

This yields the self-normalized importance sampling approximation

$$\boxed{\mathbb{E}_{X\sim f}[h(X)] \;\approx\; \mathbb{E}_{X\sim\hat{P}_N}[h(X)].}$$

The approximation occurs at the *distributional level*:

$$f(x) \;\approx\; \hat{P}_N(dx).$$

Once this approximation is made, all expectations under f are replaced by exact expectations under the empirical distribution.

Key Takeaway

At this stage, the essential idea is simple:

- draw N samples from a proposal distribution,
- compute importance weights,
- normalize the weights,
- obtain an empirical approximation of the target distribution.

This is the only mechanism needed to approximate a probability distribution.

Particle filtering applies this same construction *sequentially in time* to approximate filtering and smoothing distributions in dynamical systems.

Notation

N	number of samples (particles)
$\mathcal{X}_n = (X_0, X_1, \ldots, X_n)$	state trajectory up to time step n
$\mathcal{Y}_n = (y_1, y_2, \ldots, y_n)$	measurements up to time step n
$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n)$	smoothing distribution (target distribution)
$\pi(\mathcal{X}_n)$	proposal distribution on state trajectories
$\hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n)$	empirical approximation of the smoothing distribution
$\mathcal{X}_n^{(i)}$	i-th sampled state trajectory
$w_n^{(i)}$	unnormalized importance weight of sample i
$\bar{w}_n^{(i)}$	normalized importance weight of sample i
$\sum_{i=1}^{N} \bar{w}_n^{(i)} = 1$	weight normalization constraint

The unnormalized and normalized importance weights are defined as

$$w_n^{(i)} = \frac{\mathbb{P}(\mathcal{X}_n^{(i)} \mid \mathcal{Y}_n)}{\pi(\mathcal{X}_n^{(i)})}, \qquad \bar{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

The empirical approximation is represented by the weighted sample set

$$\{\mathcal{X}_n^{(i)}; \bar{w}_n^{(i)}\}_{i=1}^{N}.$$

This collection can be written explicitly, row by row, as

$$\begin{aligned}
\{\mathcal{X}_n^{(1)}; \bar{w}_n^{(1)}\} &= \big((x_0^{(1)}, x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})\ ;\ \bar{w}_n^{(1)}\big), \\
\{\mathcal{X}_n^{(2)}; \bar{w}_n^{(2)}\} &= \big((x_0^{(2)}, x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)})\ ;\ \bar{w}_n^{(2)}\big), \\
&\vdots \\
\{\mathcal{X}_n^{(N)}; \bar{w}_n^{(N)}\} &= \big((x_0^{(N)}, x_1^{(N)}, x_2^{(N)}, \ldots, x_n^{(N)})\ ;\ \bar{w}_n^{(N)}\big).
\end{aligned}$$

Equivalently,

$\{\mathcal{X}_n^{(1)}, \mathcal{X}_n^{(2)}, \ldots, \mathcal{X}_n^{(N)}\}$ are the N samples, and $\{\bar{w}_n^{(1)}, \bar{w}_n^{(2)}, \ldots, \bar{w}_n^{(N)}\}$ are the corresponding normalized weights.

3 Empirical Approximation via SNIS on Trajectory Space

Having fixed the notation, we now consider the most immediate approximation strategy suggested by the availability of a proposal distribution on trajectory space.

Suppose that we are able to draw samples from the proposal distribution $\pi(\mathcal{X}_n)$. A natural first step is to generate N trajectory samples

$$\{\mathcal{X}_n^{(1)}, \mathcal{X}_n^{(2)}, \ldots, \mathcal{X}_n^{(N)}\}, \qquad \mathcal{X}_n^{(i)} \sim \pi(\mathcal{X}_n).$$

Given the target smoothing distribution $\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n)$, each sampled trajectory is assigned an unnormalized importance weight

$$w_n^{(i)} = \frac{\mathbb{P}(\mathcal{X}_n^{(i)} \mid \mathcal{Y}_n)}{\pi(\mathcal{X}_n^{(i)})}, \qquad i = 1, \ldots, N.$$

The corresponding normalized weights are defined as

$$\bar{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^N w_n^{(j)}}, \qquad \sum_{i=1}^N \bar{w}_n^{(i)} = 1.$$

Using the weighted sample set $\{\mathcal{X}_n^{(i)}; \bar{w}_n^{(i)}\}_{i=1}^N$, we define the empirical approximation of the smoothing distribution as

$$\hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \sum_{i=1}^N \bar{w}_n^{(i)} \, \delta_{\mathcal{X}_n^{(i)}}(\mathcal{X}_n).$$

This empirical distribution serves as a discrete approximation of the true smoothing distribution,

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n) \;\approx\; \hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n).$$

As a consequence, expectations of any measurable function $h(\mathcal{X}_n)$ under the smoothing distribution admit the approximation

$$\mathbb{E}_{\mathcal{X}_n \sim \mathbb{P}(\cdot \mid \mathcal{Y}_n)}[h(\mathcal{X}_n)] \;\approx\; \mathbb{E}_{\mathcal{X}_n \sim \hat{P}(\cdot \mid \mathcal{Y}_n)}[h(\mathcal{X}_n)] = \sum_{i=1}^N \bar{w}_n^{(i)} \, h\Big(\mathcal{X}_n^{(i)}\Big).$$

4 Why Direct Importance Sampling on Trajectory Space Breaks Down

At time step n, we have constructed an empirical approximation of the smoothing distribution

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n) \;\approx\; \hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n),$$

which allows expectations of any function $h(\mathcal{X}_n)$ to be approximated using the empirical distribution. In this sense, all inference questions associated with the smoothing distribution at time n can be answered approximately.

The question is how to proceed when a new measurement becomes available.

Upon receiving a new observation y_{n+1}, the object of interest shifts to the updated smoothing distribution

$$\mathbb{P}(\mathcal{X}_{n+1} \mid \mathcal{Y}_{n+1}), \qquad \mathcal{X}_{n+1} = (X_0, X_1, \ldots, X_n, X_{n+1}), \quad \mathcal{Y}_{n+1} = (y_1, y_2, \ldots, y_{n+1}).$$

A direct extension of the previous construction would attempt to approximate this new distribution by applying importance sampling on the enlarged trajectory space. That is, one would introduce a proposal distribution $\pi(\mathcal{X}_{n+1})$, draw N trajectory samples, and construct a new empirical approximation from scratch.

This approach is conceptually straightforward. However, it breaks down for structural reasons.

Unbounded growth of the trajectory dimension. Each time step appends one additional state variable to the trajectory. As a result, the smoothing distribution at time $n+1$ is defined on a strictly higher-dimensional space than at time n. Both the target distribution $\mathbb{P}(\mathcal{X}_{n+1} \mid \mathcal{Y}_{n+1})$ and the proposal distribution $\pi(\mathcal{X}_{n+1})$ therefore live in an ever-expanding space.

As the dimension increases, drawing representative samples from a proposal on the full trajectory space rapidly becomes intractable. Even for moderate time horizons, importance sampling in such high dimensions becomes computationally prohibitive.

Lack of reuse of previously computed information. At time n, substantial effort has already been invested in constructing an empirical approximation of $\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n)$. If, at time $n+1$, we discard this approximation and restart importance sampling on $\mathcal{X}_{n+1}$, then none of the information encoded in the previous empirical distribution is reused.

This contradicts the structure of the inference problem. The smoothing distribution at time $n+1$ is not independent of the smoothing distribution at time n; it is directly built upon it. Restarting the approximation from scratch at each time step is therefore fundamentally inefficient.

Infeasibility of proposal construction on long trajectories. Effective importance sampling requires a proposal distribution that closely resembles the target. As the trajectory length grows, the target smoothing distribution encodes increasingly complex dependencies across time. Designing a proposal $\pi(\mathcal{X}_{n+1})$ that adequately captures this structure on the full trajectory space becomes increasingly difficult, and in practice infeasible.

Taken together, these observations lead to a clear conclusion. Approximating the smoothing distribution independently at each time step by applying importance sampling on the full trajectory space is not a viable strategy. The dimensionality grows without bound, sampling becomes intractable, and previously computed information is discarded rather than exploited.

To make sequential inference possible, we require a method that propagates samples and weights forward in time, reusing the empirical approximation at time n to construct the approximation at time $n+1$. This necessity motivates the development of a recursive, sequential approximation of the smoothing distribution.

5 Toward a Recursive Empirical Approximation

The previous section established that importance sampling on trajectory space provides a principled empirical approximation of the smoothing distribution at a fixed time. However, it also revealed that recomputing this approximation from scratch at each time step is not viable.

Despite this, the empirical viewpoint itself remains sound. The limitation lies not in the approximation, but in how it is applied over time.

Our objective is therefore to construct an empirical approximation of the smoothing distribution

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n)$$

that evolves coherently as new observations arrive, without discarding previously computed information and without sampling anew from an ever-growing trajectory space.

Suppose that at time $n-1$ we already possess an empirical approximation

$$\hat{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1}),$$

represented by a finite set of weighted trajectories. Upon receiving a new measurement y_n, the target distribution changes to

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n), \qquad \mathcal{X}_n = (\mathcal{X}_{n-1}, X_n).$$

Rather than restarting the approximation process from scratch, the structure of this expression suggests a different strategy: the new distribution differs from the previous one only by the addition of a single state variable and a single observation.

This observation motivates the search for a recursive transformation that updates the empirical approximation incrementally. Specifically, we seek a mapping

$$g\Big(\hat{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1}),\, y_n\Big) \;=\; \hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n),$$

which propagates both samples and weights forward in time.

Such a transformation must:

- extend existing trajectories by one additional state,
- update importance weights using only newly available information,
- preserve a finite empirical representation at every time step.

Deriving this transformation requires combining the factorization of the Bayesian smoothing distribution with the principles of importance sampling. When these elements are assembled carefully, the result is a sequential construction of empirical distributions that evolves in time.

This sequential application of self-normalized importance sampling is the foundation of what is known as *particle filtering*.

6 Factorization of the Smoothing Distribution and Recursive Weight Update

We consider the smoothing distribution at time step n:

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n), \qquad \mathcal{X}_n = (X_0, \ldots, X_n), \quad \mathcal{Y}_n = (y_1, \ldots, y_n).$$

Since $\mathcal{X}_n = (X_n, \mathcal{X}_{n-1})$ and $\mathcal{Y}_n = (y_n, \mathcal{Y}_{n-1})$, we write

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \mathbb{P}(X_n, \mathcal{X}_{n-1} \mid y_n, \mathcal{Y}_{n-1}).$$

Step 1: Conditional Bayes Rule

Apply the conditional Bayes rule

$$\mathbb{P}(A \mid B, C) = \frac{\mathbb{P}(B \mid A, C)\,\mathbb{P}(A \mid C)}{\mathbb{P}(B \mid C)},$$

with $A = (X_n, \mathcal{X}_{n-1})$, $B = y_n$, and $C = \mathcal{Y}_{n-1}$:

$$\mathbb{P}(X_n, \mathcal{X}_{n-1} \mid y_n, \mathcal{Y}_{n-1}) = \frac{\mathbb{P}(y_n \mid X_n, \mathcal{X}_{n-1}, \mathcal{Y}_{n-1})\,\mathbb{P}(X_n, \mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}.$$

Therefore,

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \frac{\mathbb{P}(y_n \mid X_n, \mathcal{X}_{n-1}, \mathcal{Y}_{n-1})\,\mathbb{P}(X_n, \mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}.$$

Step 2: Conditional Independence of the Observation

By conditional independence implied by the Bayesian network (d-separation),

$$y_n \perp \mathcal{X}_{n-1} \mid X_n, \qquad y_n \perp \mathcal{Y}_{n-1} \mid X_n,$$

and hence

$$\mathbb{P}(y_n \mid X_n, \mathcal{X}_{n-1}, \mathcal{Y}_{n-1}) = \mathbb{P}(y_n \mid X_n).$$

Substituting,

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \frac{\mathbb{P}(y_n \mid X_n)\,\mathbb{P}(X_n, \mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}.$$

Step 3: Factor the Joint Predictive Distribution

Next, factor the joint predictive distribution:

$$\mathbb{P}(X_n, \mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1}) = \mathbb{P}(X_n \mid \mathcal{X}_{n-1}, \mathcal{Y}_{n-1})\,\mathbb{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1}).$$

Substituting,

$$\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \frac{\mathbb{P}(y_n \mid X_n)\,\mathbb{P}(X_n \mid \mathcal{X}_{n-1}, \mathcal{Y}_{n-1})\,\mathbb{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}.$$

Step 4: Conditional Independence of the Dynamics

Using the Markov property of the state process (d-separation),

$$X_n \perp \mathcal{X}_{n-1} \mid X_{n-1}, \qquad X_n \perp \mathcal{Y}_{n-1} \mid X_{n-1},$$

and therefore

$$\mathbb{P}(X_n \mid \mathcal{X}_{n-1}, \mathcal{Y}_{n-1}) = \mathbb{P}(X_n \mid X_{n-1}).$$

Substituting,

$$\boxed{\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \frac{\mathbb{P}(y_n \mid X_n)\,\mathbb{P}(X_n \mid X_{n-1})\,\mathbb{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}.}$$

Step 5: Proposal Factorization

Introduce a proposal distribution on trajectory space:

$$\pi(\mathcal{X}_n) = \pi(X_n \mid \mathcal{X}_{n-1})\, \pi(\mathcal{X}_{n-1}).$$

Step 6: Importance Weights

Define the unnormalized importance weights:

$$w_n(\mathcal{X}_n) = \frac{\mathbb{P}(\mathcal{X}_n \mid \mathcal{Y}_n)}{\pi(\mathcal{X}_n)}.$$

Substitute the factorized smoothing distribution and proposal:

$$w_n(\mathcal{X}_n) = \frac{\frac{\mathbb{P}(y_n|X_n)\,\mathbb{P}(X_n|X_{n-1})\,\mathbb{P}(\mathcal{X}_{n-1}|\mathcal{Y}_{n-1})}{\mathbb{P}(y_n|\mathcal{Y}_{n-1})}}{\pi(X_n \mid \mathcal{X}_{n-1})\, \pi(\mathcal{X}_{n-1})}.$$

Rewrite by separating factors:

$$w_n(\mathcal{X}_n) = \left[\frac{1}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}\right] \left[\frac{\mathbb{P}(y_n \mid X_n)\,\mathbb{P}(X_n \mid X_{n-1})}{\pi(X_n \mid \mathcal{X}_{n-1})}\right] \left[\frac{\mathbb{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{\pi(\mathcal{X}_{n-1})}\right].$$

Recognize the final factor as the previous weight:

$$w_{n-1}(\mathcal{X}_{n-1}) = \frac{\mathbb{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{\pi(\mathcal{X}_{n-1})}.$$

Therefore, the recursive update is

$$\boxed{w_n(\mathcal{X}_n) = \left[\frac{1}{\mathbb{P}(y_n \mid \mathcal{Y}_{n-1})}\right] \left[\frac{\mathbb{P}(y_n \mid X_n)\,\mathbb{P}(X_n \mid X_{n-1})}{\pi(X_n \mid \mathcal{X}_{n-1})}\right] w_{n-1}(\mathcal{X}_{n-1}).}$$

7 Sequential Construction of the Empirical Smoothing Distribution

The derivation so far establishes a crucial result.

Starting from an empirical approximation of the smoothing distribution at time $n-1$,

$$\hat{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1}) = \sum_{i=1}^{N} \bar{w}_{n-1}^{(i)}\, \delta_{\mathcal{X}_{n-1}^{(i)}},$$

we have shown that the arrival of a new observation y_n admits a *recursive update* of both particle trajectories and weights.

Each particle trajectory is extended incrementally by sampling a new state

$$x_n^{(i)} \sim \pi(\,\cdot \mid \mathcal{X}_{n-1}^{(i)}), \qquad \mathcal{X}_n^{(i)} = (\mathcal{X}_{n-1}^{(i)}, x_n^{(i)}),$$

and the corresponding importance weights evolve according to

$$w_n^{(i)} = w_{n-1}^{(i)} \frac{\mathbb{P}(y_n \mid x_n^{(i)})\, \mathbb{P}(x_n^{(i)} \mid x_{n-1}^{(i)})}{\pi(x_n^{(i)} \mid \mathcal{X}_{n-1}^{(i)})}, \qquad \bar{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

As a result, the empirical approximation of the smoothing distribution at time n is obtained directly from the approximation at time $n-1$:

$$\hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \sum_{i=1}^{N} \bar{w}_n^{(i)}\, \delta_{\mathcal{X}_n^{(i)}}.$$

This construction avoids resampling full trajectories from scratch, prevents unbounded growth of proposal dimensionality, and respects the recursive structure of Bayesian inference. Particle filtering emerges as the sequential application of self-normalized importance sampling on trajectory space

8 Sequential Importance Sampling: Updating Trajectories and Weights

This section explains, in full detail, how an empirical approximation of the smoothing distribution is updated when a new measurement arrives.

The goal is to answer the following question:

> Given a weighted empirical approximation of the smoothing distribution at time $n-1$, how do we construct the corresponding approximation at time n when a new observation y_n is observed?

Throughout this section, all approximations are performed on *trajectory space*. No state-wise shortcuts are taken.

Particles are trajectories

A **particle** is not a single state. A particle is an entire **state trajectory**.

For particle i at time n, the associated trajectory is

$$\mathcal{X}_n^{(i)} = (x_0^{(i)}, x_1^{(i)}, \ldots, x_n^{(i)}),$$

together with an associated importance weight.

From this point onward, we use the term **trajectory** exclusively.

Step 0: Initialization at time 0

At time 0, the smoothing distribution reduces to the prior:

$$p(\mathcal{X}_0 \mid \mathcal{Y}_0) = p(X_0).$$

We introduce a proposal distribution $\pi_0(X_0)$ and draw N i.i.d. samples:

$$x_0^{(i)} \sim \pi_0(X_0), \qquad i = 1, \dots, N.$$

Each sample defines an initial trajectory

$$\mathcal{X}_0^{(i)} = (x_0^{(i)}).$$

The corresponding unnormalized importance weights are

$$w_0^{(i)} = \frac{p(x_0^{(i)})}{\pi_0(x_0^{(i)})},$$

and the normalized weights are

$$\bar{w}_0^{(i)} = \frac{w_0^{(i)}}{\sum_{j=1}^{N} w_0^{(j)}}.$$

This produces the initial empirical approximation:

$$\hat{P}(\mathcal{X}_0 \mid \mathcal{Y}_0) = \sum_{i=1}^{N} \bar{w}_0^{(i)} \, \delta_{\mathcal{X}_0^{(i)}}.$$

What we have at time $n-1$

Assume that at time $n-1$ we have already constructed an empirical approximation of the smoothing distribution:

$$\hat{P}(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1}) = \sum_{i=1}^{N} \bar{w}_{n-1}^{(i)} \, \delta_{\mathcal{X}_{n-1}^{(i)}}.$$

This means we have exactly:

- N **deterministic trajectories**

$$\mathcal{X}_{n-1}^{(i)} = (x_0^{(i)}, x_1^{(i)}, \dots, x_{n-1}^{(i)}),$$

- and N **normalized weights**

$$\bar{w}_{n-1}^{(1)}, \dots, \bar{w}_{n-1}^{(N)}, \qquad \sum_{i=1}^{N} \bar{w}_{n-1}^{(i)} = 1.$$

Compactly, the entire approximation is the weighted set

$$\{\mathcal{X}_{n-1}^{(i)}; \bar{w}_{n-1}^{(i)}\}_{i=1}^{N}.$$

There is no randomness left in this set. All randomness at the next step enters only through sampling new states.

What changes at time n: extending trajectories

A new measurement y_n arrives, and the target becomes the new smoothing distribution

$$p(\mathcal{X}_n \mid \mathcal{Y}_n), \qquad \mathcal{X}_n = (\mathcal{X}_{n-1}, X_n).$$

To construct an empirical approximation at time n, we do **not** resample full trajectories. Instead, we extend each existing trajectory by **exactly one** new state.

For each trajectory $i = 1, \dots, N$, we sample

$$x_n^{(i)} \sim \pi(x_n \mid \mathcal{X}_{n-1}^{(i)}).$$

Expanding the conditioning makes the dependence explicit:

$$x_n^{(i)} \sim \pi\big(x_n \mid x_0^{(i)}, x_1^{(i)}, \dots, x_{n-1}^{(i)}\big).$$

Crucial point. Because each trajectory $\mathcal{X}_{n-1}^{(i)}$ is different, each conditional proposal $\pi(x_n \mid \mathcal{X}_{n-1}^{(i)})$ is, in general, different. Thus, we are not drawing N samples from one proposal, but one sample from each of N trajectory-dependent proposals.

Each sampled state is appended to form a new trajectory:

$$\mathcal{X}_n^{(i)} = (\mathcal{X}_{n-1}^{(i)}, x_n^{(i)}).$$

How the weight update imitates Bayes' rule

The Bayesian smoothing recursion factorizes as

$$p(\mathcal{X}_n \mid \mathcal{Y}_n) = \frac{p(y_n \mid X_n)\, p(X_n \mid X_{n-1})\, p(\mathcal{X}_{n-1} \mid \mathcal{Y}_{n-1})}{p(y_n \mid \mathcal{Y}_{n-1})}.$$

This expresses a Bayesian update on trajectory space: the previous posterior is multiplied by new information (likelihood and dynamics) and renormalized.

Sequential importance sampling performs the same update, but on a finite weighted set of trajectories.

Unnormalized weight update

The importance weight of a trajectory is

$$w_n(\mathcal{X}_n) = \frac{p(\mathcal{X}_n \mid \mathcal{Y}_n)}{\pi(\mathcal{X}_n)}.$$

Using the proposal factorization

$$\pi(\mathcal{X}_n) = \pi(x_n \mid \mathcal{X}_{n-1})\, \pi(\mathcal{X}_{n-1}),$$

we obtain the recursive update

$$w_n(\mathcal{X}_n) = \left[\frac{1}{p(y_n \mid \mathcal{Y}_{n-1})}\right] \left[\frac{p(y_n \mid X_n)\, p(X_n \mid X_{n-1})}{\pi(X_n \mid \mathcal{X}_{n-1})}\right] w_{n-1}(\mathcal{X}_{n-1}).$$

The evidence term $p(y_n \mid \mathcal{Y}_{n-1})$ is constant across trajectories and is dropped in practice since weights are normalized.

Thus, for each trajectory i,

$$\boxed{w_n^{(i)} = w_{n-1}^{(i)} \frac{p(y_n \mid x_n^{(i)})\, p(x_n^{(i)} \mid x_{n-1}^{(i)})}{\pi(x_n^{(i)} \mid \mathcal{X}_{n-1}^{(i)})}}.$$

Normalization and new empirical approximation

The normalized weights are

$$\bar{w}_n^{(i)} = \frac{w_n^{(i)}}{\sum_{j=1}^{N} w_n^{(j)}}.$$

The updated empirical approximation is

$$\hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \sum_{i=1}^{N} \bar{w}_n^{(i)}\, \delta_{\mathcal{X}_n^{(i)}}.$$

Once this approximation is constructed, expectations under the smoothing distribution are computed exactly under the empirical measure.

Algorithm 1: Sequential Importance Sampling on Trajectory Space (No Resampling)

Require: N; measurements $y_{1:n}$; initial proposal $\pi_0(x_0)$; trajectory proposals $\pi_k(x_k \mid \mathcal{X}_{k-1})$ for $k = 1, \ldots, n$;
prior $p(x_0)$; transition model $p(x_k \mid x_{k-1})$; likelihood $p(y_k \mid x_k)$.
Ensure: $\hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n) = \sum_{i=1}^{N} \bar{w}_n^{(i)} \, \delta_{\mathcal{X}_n^{(i)}}$.

Step 0 (initialization on X_0):;
Draw N i.i.d. samples $\{x_0^{(i)}\}_{i=1}^{N} \overset{\text{i.i.d.}}{\sim} \pi_0(x_0)$;
for $i = 1$ **to** N **do**
  $\mathcal{X}_0^{(i)} \leftarrow (x_0^{(i)})$;
  $w_0^{(i)} \leftarrow \dfrac{p(x_0^{(i)})}{\pi_0(x_0^{(i)})}$;

Normalize: $\bar{w}_0^{(i)} \leftarrow \dfrac{w_0^{(i)}}{\sum_{j=1}^{N} w_0^{(j)}}$ for $i = 1, \ldots, N$;

For $k = 1, \ldots, n$ **(extend trajectories and update weights):**;
for $k = 1$ **to** n **do**
  for $i = 1$ **to** N **do**
    (1) Sample one new state per existing trajectory:;
    $x_k^{(i)} \sim \pi_k\Big(x_k \mid \mathcal{X}_{k-1}^{(i)}\Big)$;
    (2) Append to form the new trajectory (particle):;
    $\mathcal{X}_k^{(i)} \leftarrow \Big(\mathcal{X}_{k-1}^{(i)}, x_k^{(i)}\Big)$;
    (3) Incremental importance ratio:;
    $v_k^{(i)} \leftarrow \dfrac{p\Big(y_k \mid x_k^{(i)}\Big) \; p\Big(x_k^{(i)} \mid x_{k-1}^{(i)}\Big)}{\pi_k\Big(x_k^{(i)} \mid \mathcal{X}_{k-1}^{(i)}\Big)}$;
    (4) Unnormalized weight update (evidence omitted):;
    $w_k^{(i)} \leftarrow v_k^{(i)} \, \bar{w}_{k-1}^{(i)}$;
  (5) Normalize weights:;
  $\bar{w}_k^{(i)} \leftarrow \dfrac{w_k^{(i)}}{\sum_{j=1}^{N} w_k^{(j)}} \quad$ for $i = 1, \ldots, N$;

9 Limitations of Pure Sequential Importance Sampling and Roadmap

The preceding derivation establishes the exact recursive form of *sequential importance sampling* (SIS) on trajectory space. At each time step, the Bayesian smoothing distribution is approximated by a finite weighted empirical measure whose support consists of extended state trajectories.

Importantly, this construction is exact at the level of probability measures. No approximation is made in the Bayesian recursion itself. The only approximation introduced is representational: the replacement of the true smoothing distribution by a finite empirical distribution.

From a theoretical standpoint, this completes the derivation of particle filtering as *sequential self-normalized importance sampling*. From a practical standpoint, however, this construction is not yet sufficient.

Why Pure SIS Is Statistically Unstable

Although the SIS recursion is mathematically valid, its statistical behavior over time is problematic.

As time progresses, the recursive multiplication of likelihood and transition terms causes the importance weights to become increasingly imbalanced. In typical dynamical systems, probability mass concentrates onto a vanishingly small subset of trajectories, while the majority of trajectories carry negligible weight.

Viewed at the level of empirical measures, this manifests as:

- collapse of the effective support of the empirical distribution,
- loss of representational diversity in trajectory space,
- deterioration of approximation quality despite increasing computational effort.

This phenomenon is not an implementation artifact. It is a structural consequence of applying importance sampling recursively on an expanding trajectory space.

What This Report Deliberately Does Not Address

The present report intentionally stops at the exact SIS formulation. It does not attempt to resolve the statistical pathologies described above.

In particular, it does not address:

- quantitative measures of weight concentration,
- diagnostics for assessing empirical approximation quality,
- mechanisms for redistributing probability mass across trajectories,
- proposal design strategies aimed at mitigating weight collapse.

These issues are not secondary details. They are essential for making particle filtering viable in practice. However, addressing them meaningfully requires first understanding the exact SIS structure from which they arise.

Roadmap

Having established particle filtering as *sequential self-normalized importance sampling on trajectory space*, subsequent developments will build directly on this foundation.

The next steps proceed in the following order:

1. Quantifying weight concentration via the notion of *effective sample size* (ESS),
2. Interpreting degeneracy as a failure of the empirical measure, not merely of individual weights,
3. Introducing resampling as a measure-theoretic operation that redistributes probability mass while preserving expectations,

4. Transitioning from full trajectory representations to filtering distributions on the current state,
5. Recovering practical particle filtering algorithms as structured approximations of the exact SIS recursion.

Each of these steps modifies the empirical representation while leaving the underlying Bayesian recursion unchanged.

Perspective

The central message of this report is therefore the following:

> Sequential importance sampling is the correct starting point for particle filtering, but not its endpoint.

Only by first deriving SIS at the level of probability measures can the necessity, role, and limitations of later refinements be understood without ambiguity.

This report establishes that foundation.