Weight Degeneracy, Effective Sample Size, and Resampling in Particle Filtering

Particle Filtering in Dynamical Systems

Sai Sampath Kedari
sampath@umich.edu

Contents

1 Weight Degeneracy in Sequential Importance Sampling

Sequential importance sampling represents the smoothing distribution at time n by a finite weighted empirical measure

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

where each particle corresponds to a complete state trajectory $\mathcal{X}_n^{(i)} = (x_0^{(i)}, \ldots, x_n^{(i)})$ and $\bar{w}_n^{(i)}$ denotes its normalized importance weight.

At each time step, the unnormalized weights are updated multiplicatively by incremental importance ratios involving the likelihood and transition model. As a result, small differences between trajectories are repeatedly amplified over time. Trajectories that happen to align slightly better with early observations accumulate increasingly large weights, while others are exponentially downweighted.

The typical outcome is that, after a moderate number of time steps, the normalized weight vector becomes extremely imbalanced:

$$\bar{w}_n^{(i^\star)} \approx 1, \qquad \bar{w}_n^{(i)} \approx 0 \quad \text{for } i \neq i^\star,$$

for some index $i^\star$.

Although N trajectories are present, almost all probability mass concentrates on a single trajectory. From the perspective of the empirical measure, the approximation effectively collapses to a point mass. Despite having sampled many trajectories, the resulting distribution behaves as if it were supported on only one.

This collapse has severe consequences. Expectations under the empirical distribution are dominated by a single trajectory, uncertainty is no longer represented, and alternative explanations of the data are lost. Increasing N does not resolve the problem once this collapse has occurred, because additional particles carry negligible weight.

This phenomenon is known as *weight degeneracy*.

Weight degeneracy is not a numerical artifact or an implementation flaw. It is a structural consequence of repeatedly applying importance sampling on an expanding trajectory space. As time progresses, the effective number of trajectories contributing to the approximation inevitably decreases.

To make this collapse explicit and quantifiable, and to decide when corrective action is required, we introduce the notion of *effective sample size*.

2 Effective Sample Size for Self-Normalized Importance Sampling

The collapse described in the previous section raises a natural quantitative question:

> *How many independent, equally weighted samples would give the same Monte Carlo accuracy as a given weighted empirical distribution?*

This question leads directly to the notion of *effective sample size* (ESS).

Independent Sampling as the Reference

Let $f(x)$ denote a target distribution and let h be a measurable test function. Suppose we can sample directly from f:

$$X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} f.$$

The standard Monte Carlo estimator

$$\hat{\mu}_{\text{MC}} = \frac{1}{N} \sum_{i=1}^{N} h(X_i)$$

satisfies

$$\text{Var}(\hat{\mu}_{\text{MC}}) = \frac{1}{N} \text{Var}_f(h(X)) \,.$$

Each sample contributes one full unit of information. This variance scaling serves as the benchmark for efficiency.

Self-Normalized Importance Sampling

Now suppose direct sampling from f is not possible. Instead, we draw samples from a proposal distribution g:

$$X_1, \ldots, X_N \sim g, \qquad w(X) = \frac{f(X)}{g(X)}.$$

The self-normalized importance sampling (SNIS) estimator is

$$\hat{\mu}_{\text{SNIS}} = \sum_{i=1}^{N} \bar{w}_i \, h(X_i), \qquad \bar{w}_i = \frac{w(X_i)}{\sum_{j=1}^{N} w(X_j)}.$$

Although N samples are generated, their contributions are unequal. Samples with large weights dominate the estimator, while many samples contribute negligibly. Consequently, the estimator behaves as if it were based on fewer independent samples.

Variance-Based Definition of ESS

The effective sample size N_{eff} is defined by equating the variance of the SNIS estimator to that of an ideal Monte Carlo estimator with N_{eff} independent samples:

$$\text{Var}(\hat{\mu}_{\text{SNIS}}) \;\approx\; \frac{1}{N_{\text{eff}}} \text{Var}_f(h(X)) \,.$$

For large N, the variance of the SNIS estimator admits the approximation

$$\text{Var}(\hat{\mu}_{\text{SNIS}}) \;\approx\; \frac{1}{N} \mathbb{E}_g\Big[w(X)^2 \big(h(X) - \mu\big)^2 \Big], \qquad \mu = \mathbb{E}_f[h(X)].$$

Equating variances yields

$$\frac{1}{N_{\text{eff}}} \text{Var}_f(h) = \frac{1}{N} \mathbb{E}_g\big[w^2 (h - \mu)^2 \big] \,.$$

Solving for N_{eff},

$$N_{\text{eff}} = N \frac{\text{Var}_f(h)}{\mathbb{E}_g[w^2(h-\mu)^2]}.$$

To obtain a diagnostic independent of the particular test function h, we use the bound

$$\mathbb{E}_g\big[w^2(h-\mu)^2\big] \leq \mathbb{E}_g[w^2]\ \text{Var}_f(h),$$

which leads to the approximation

$$N_{\text{eff}} \ \approx\ \frac{N}{\mathbb{E}_g[w^2]}.$$

Empirical ESS Formula

Replacing expectations by empirical averages and expressing the result in terms of normalized weights gives the standard ESS estimator:

$$\boxed{N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} (\bar{w}_i)^2}.}$$

Interpretation

- If all weights are equal, $\bar{w}_i = 1/N$, then $N_{\text{eff}} = N$.
- If a single weight dominates, $\bar{w}_k \approx 1$, then $N_{\text{eff}} \approx 1$.

Thus, ESS measures how many *independent, equally weighted samples* the weighted empirical distribution is effectively worth.

In sequential importance sampling, ESS typically decreases over time, reflecting the progressive concentration of probability mass onto fewer trajectories. When ESS becomes small, the empirical approximation no longer provides a reliable representation of the target distribution.

3 Resampling

At time step k, sequential importance sampling produces the empirical approximation

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

When the effective sample size

$$N_{\text{eff}}(k) = \left(\sum_{i=1}^{N} (\bar{w}_k^{(i)})^2 \right)^{-1}$$

falls below a prescribed threshold N_{th}, the empirical distribution is considered degenerate and resampling is performed.

Definition of Resampling

Resampling is a stochastic transformation of the empirical measure

$$\hat{P}(\mathcal{X}_k \mid \mathcal{Y}_k)$$

into a new empirical measure with *uniform weights*.

Specifically, resampling consists of drawing N trajectories

$$\tilde{\mathcal{X}}_k^{(1)}, \ldots, \tilde{\mathcal{X}}_k^{(N)}$$

independently according to the discrete distribution defined by the current weights:

$$\mathbb{P}\Big(\tilde{\mathcal{X}}_k^{(j)} = \mathcal{X}_k^{(i)}\Big) = \bar{w}_k^{(i)}.$$

The resulting empirical approximation is

$$\boxed{\hat{P}^{\text{res}}(\mathcal{X}_k \mid \mathcal{Y}_k) = \frac{1}{N} \sum_{j=1}^{N} \delta_{\tilde{\mathcal{X}}_k^{(j)}}.}$$

All resampled trajectories carry equal weight $1/N$.

What Resampling Does

Resampling replaces a weighted empirical measure by an unweighted one whose support consists of repeated copies of high-probability trajectories.

As a consequence:

- Trajectories with large weights are replicated.
- Trajectories with negligible weights are discarded.
- Duplicate trajectories are unavoidable.
- The cardinality of the support remains N.

What Resampling Does *Not* Do

Resampling does *not* introduce new state values. It does *not* modify the Bayesian recursion. It does *not* change the target distribution.

It is purely a redistribution of probability mass within the existing empirical representation.

Preservation of Expectations

Resampling preserves expectations in expectation. For any measurable function h,

$$\mathbb{E}\left[\frac{1}{N} \sum_{j=1}^{N} h(\tilde{\mathcal{X}}_k^{(j)})\right] = \sum_{i=1}^{N} \bar{w}_k^{(i)}\, h(\mathcal{X}_k^{(i)}).$$

Thus, resampling does not bias Monte Carlo estimates. It trades variance in the weights for variance introduced by random replication.

Role in Sequential Inference

Resampling restores a balanced empirical representation by eliminating weight concentration. However, it does not restore diversity by itself.

Diversity is reintroduced only at the next propagation step, when trajectories are extended through the state transition model and proposal distribution.

Resampling therefore acts as a *reset of the empirical measure*, enabling sequential importance sampling to continue without collapse.

Summary

Resampling is a measure-theoretic operation that:

- acts only on the empirical representation,
- preserves expectations,
- resets weights to uniform values,
- enables long-term sequential inference.

With ESS-based resampling, sequential importance sampling becomes a practical particle filtering algorithm.

Algorithm 1: Sequential Importance Sampling on Trajectory Space with ESS-Based Resampling

Require: N; threshold N_{th}; measurements $y_{1:n}$; initial proposal $\pi_0(x_0)$; trajectory proposals $\pi_k(x_k \mid \mathcal{X}_{k-1})$;
prior $p(x_0)$; transition model $p(x_k \mid x_{k-1})$; likelihood $p(y_k \mid x_k)$.
Ensure: $\hat{P}(\mathcal{X}_n \mid \mathcal{Y}_n)$.

Step 0 (initialization):;
Draw $\{x_0^{(i)}\}_{i=1}^N \sim \pi_0(x_0)$;
$\mathcal{X}_0^{(i)} \leftarrow (x_0^{(i)})$;
$w_0^{(i)} \leftarrow \dfrac{p(x_0^{(i)})}{\pi_0(x_0^{(i)})}$;
Normalize weights.

for $k = 1$ **to** n **do**
  for $i = 1$ **to** N **do**
    $x_k^{(i)} \sim \pi_k(x_k \mid \mathcal{X}_{k-1}^{(i)})$;
    $\mathcal{X}_k^{(i)} \leftarrow (\mathcal{X}_{k-1}^{(i)}, x_k^{(i)})$;
    $v_k^{(i)} \leftarrow \dfrac{p(y_k \mid x_k^{(i)})\, p(x_k^{(i)} \mid x_{k-1}^{(i)})}{\pi_k(x_k^{(i)} \mid \mathcal{X}_{k-1}^{(i)})}$;
    $w_k^{(i)} \leftarrow v_k^{(i)}\, \bar{w}_{k-1}^{(i)}$;
  Normalize $\{\bar{w}_k^{(i)}\}$;
  Compute $N_{\text{eff}} \leftarrow \left(\sum_{i=1}^N (\bar{w}_k^{(i)})^2 \right)^{-1}$;
  if $N_{\text{eff}} < N_{\text{th}}$ **then**
    Resample trajectories $\{\mathcal{X}_k^{(i)}\}$ according to $\{\bar{w}_k^{(i)}\}$;
    Set $\bar{w}_k^{(i)} \leftarrow \frac{1}{N}$ for all i;