

Adaptive Sequential Monte Carlo for High-Dimensional Bayesian Inference

Yure Santos de Oliveira

School of Applied Mathematics, Getulio Vargas Foundation

July 11, 2025

Abstract

Bayesian inference in high-dimensional models poses significant computational challenges, particularly due to the inefficiency of standard sampling techniques such as Importance Sampling and Markov Chain Monte Carlo. In this work, we explore the use of Adaptive Sequential Monte Carlo (SMC) methods for scalable and robust posterior inference and model selection. We review the theoretical foundations of SMC, including the use of backward kernels and adaptive proposals, and discuss strategies for estimating normalizing constants. The method is applied to a high-dimensional Gaussian Process regression problem, and its performance is compared against standard Importance Sampling and Annealed Importance Sampling in terms of posterior exploration and evidence estimation.

1 Introduction

Bayesian inference provides a principled framework for quantifying uncertainty and updating beliefs in light of new evidence. Its application to complex, high-dimensional models, however, presents formidable computational challenges. As the number of parameters grows, the posterior distributions of interest often become concentrated in a small volume of the parameter space, making them difficult to explore efficiently. Standard simulation-based techniques, such as Markov Chain Monte Carlo (MCMC), can struggle in these settings, suffering from slow convergence and poor mixing, while simpler methods like Importance Sampling (IS) often fail due to an inability to generate proposals that adequately cover the target posterior ([Robert and Casella, 2004](#)). These limitations hinder both parameter estimation and model selection, the latter of which relies on the computation of the marginal likelihood (or evidence), a notoriously difficult high-dimensional integral.

To overcome these hurdles, a more sophisticated class of algorithms is required. *Sequential Monte Carlo* (SMC) methods, also known as particle filters, have emerged as a powerful and flexible method for this purpose ([Doucet et al., 2001](#)). Originally developed for dynamic state-space models, SMC techniques have been successfully adapted for static Bayesian inference problems. By constructing a sequence of intermediate distributions that bridge a simple initial distribution (like the prior) with the complex target posterior, SMC methods are able to gradually guide a population of weighted samples, or “particles,” toward the regions of high probability.

This work focuses on *Adaptive Sequential Monte Carlo* samplers, which enhance the robustness and efficiency of the core SMC algorithm. We explore how dynamically tuning the proposal mechanisms using information from the current particle set can improve performance, particularly in high-dimensional settings where the geometry of the posterior is unknown. We will review the theoretical underpinnings of SMC, including the role of resampling steps to prevent weight degeneracy and the use of backward kernels to construct valid and efficient incremental weights. Furthermore, we will discuss how the SMC framework provides a natural and direct approach for estimating the normalizing constants required for model comparison.

To demonstrate the practical advantages of this approach, we apply an Adaptive SMC sampler to a challenging high-dimensional Gaussian Process regression problem. We compare its ability to explore the posterior distribution and estimate the marginal likelihood against more conventional methods like standard Importance Sampling and Annealed Importance Sampling (AIS). Through this application, we illustrate how Adaptive SMC offers a scalable and reliable solution for modern Bayesian computation.

2 Adaptive Sequential Monte Carlo

2.1 Sequential Monte Carlo

Let $\{P_t\}_{t \in \mathbb{T}}$, where $\mathbb{T} = \{1, 2, \dots, T\}$, be a sequence of probability measures defined on a common measurable space (E, \mathcal{E}) . We denote by $\pi_t(x) = \gamma_t(x)/Z_t$ the density of P_t relative to a common dominating measure.

Sequential Monte Carlo methods aim to sample from the sequence $\{P_t\}_{t \in \mathbb{T}}$ in a sequential manner: first sampling from P_1 , then from P_2 and so on. This is done by propagating and reweighting the set of particles used in the previous step.

In the context of Bayesian inference, one possibility is to define P_t as the posterior distribution given the first t observations. This setting naturally arises in online learning or state-space models, where information arrives sequentially. Alternatively, in static models where the posterior is defined from a fixed dataset, we may define a sequence of artificial intermediate steps to bridge a simple distribution and the target π_T . One common strategy is to define $\pi_t(x) \propto \pi_T(x)^{\phi_t} \pi_0(x)$, where $\{\phi_t\}_{t=1}^T$ is an increasing sequence such that $\phi_1 = 0$ and $\phi_T = 1$.

At time $t = 1$, we start with the target P_1 which is assumed to be easy to approximate efficiently by Importance Sampling with a proposal Q_1 that has density q_1 . At the time t , the target is P_t and the proposal distribution is built moving the particles by using a Markov kernel $K_t : E \times \mathcal{E} \rightarrow [0, 1]$. The particles obtained this way have distribution Q_t with density

$$q_t(x') = \int_E q_{t-1}(x) K_t(x, x') dx,$$

that is then used as proposal for Importance Sampling with target P_t .

As t increases, the discrepancy between the proposal Q_t and the target P_t is likely to grow. This causes the variance of the importance weights to increase over time. Eventually, this leads to a phenomenon known as *weight degeneracy*, where one or a very small number of particles will have significant weights, while the weights of all other particles will be nearly zero.

When this occurs, the particle set provides a very poor approximation of the target distribution. To combat this issue, a *resampling* step is introduced. When the Effective Sample Size

$$ESS = \frac{1}{\sum_{i=1}^N (W_t^{(i)})^2}$$

falls below a predefined threshold, the particles are resampled with probabilities given by their current weights and the weights are set to $1/N$. This step discards particles with low weights and multiplies those with high weights.

Beskos et al. (2014) study the stability of this algorithm and show that, in high-dimensional static problems, it can efficiently control the variability of the importance sampling weights. They also show that the resampling step leads to a reduction of the Monte Carlo error and increase in the ESS.

The major drawback of this algorithm is that it requires to evaluate q_t pointwise for each t , what is typically not possible. An alternative strategy (Del Moral et al., 2006) is to introduce artificial backward Markov kernels $L_{t-1} : E \times \mathcal{E} \rightarrow [0, 1]$ and perform IS between the joint importance distribution $q_{1:t}(x_{1:t}) = q_1(x_1) \prod_{k=1}^{t-1} K_k(x_k, x_{k+1})$ and the artificial joint target $\tilde{\pi}_{1:t}(x_{1:t}) = \tilde{\gamma}_{1:t}(x_{1:t}) / \tilde{Z}_{1:t}$ where

$$\tilde{\gamma}_{1:t}(x_{1:t}) = \gamma_t(x_t) \prod_{k=1}^{t-1} L_k(x_{k+1}, x_k).$$

The weight associated with each trajectory $x_{1:t}$ is then given by the ratio

$$w_{1:t}(x_{1:t}) = \frac{\tilde{\gamma}_{1:t}(x_{1:t})}{q_{1:t}(x_{1:t})} = w_{1:t-1}(x_{1:t-1}) \tilde{w}_t(x_{t-1}, x_t),$$

where \tilde{w}_t is the unnormalized incremental weight

$$\tilde{w}_t(x_{t-1}, x_t) = \frac{\gamma_t(x_t) L_{t-1}(x_t, x_{t-1})}{\gamma_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t)}.$$

This formulation avoids the need to evaluate the marginal proposal density $q_t(x_t)$ directly, while still yielding a valid importance sampling estimator for expectations with respect to P_t (Del Moral et al., 2006).

The choice of backward kernels L_t does not affect the correctness of the algorithm, but it significantly impacts its efficiency. A theoretically optimal choice is the backward kernel that minimizes the variance of the importance weights:

$$L_{t-1}^{\text{opt}}(x_t, x_{t-1}) = \frac{q_{t-1}(x_{t-1}) K_t(x_{t-1}, x_t)}{q_t(x_t)},$$

which, however, is intractable in practice. As a result, approximations or heuristic alternatives are often used. When K_t is an MCMC kernel of invariant distribution P_t , a generic approximation is given by

$$L_{t-1}(x_t, x_{t-1}) = \frac{\pi_t(x_{t-1}) K_t(x_{t-1}, x_t)}{\pi_t(x_t)},$$

and in this case, we have $\tilde{w}_t(x, x') = \gamma_t(x) / \gamma_{t-1}(x')$.

This general method encompasses a broad class of algorithms used in Bayesian computation, such as Particle Filters, and Annealed Importance Sampling (Neal, 2001). Its flexibility makes it particularly suited to high-dimensional or multimodal inference problems, where standard importance sampling or MCMC methods often fail.

Algorithm 1 Sequential Monte Carlo Sampler

```
1: for  $i = 1, \dots, N$  do
2:   Draw  $X_1^{(i)} \sim Q_1$ .
3:   Calculate the weight:  $w_1(X_1^{(i)}) = y_1(X_1^{(i)})/q(X_1^{(i)})$ .
4: end for
5: Normalize the weights:  $W_1^{(i)} = w_1^{(i)} / \sum_{j=1}^N w_1^{(j)}$ .
6: for  $t = 2, \dots, T$  do
7:   Calculate the Effective Sample Size:  $ESS = 1 / \sum_{i=1}^N (W_{t-1}^{(i)})^2$ .
8:   if  $ESS < \text{Threshold}$  then
9:     Resample  $\{X_{t-1}^{(i)}, W_{t-1}^{(i)}\}_{i=1}^N$  to obtain a new set of particles  $\{\hat{X}_{t-1}^{(i)}\}_{i=1}^N$ .
10:    Reset particles:  $X_{t-1}^{(i)} = \hat{X}_{t-1}^{(i)}$ .
11:    Reset weights:  $W_{t-1}^{(i)} = 1/N$ .
12:   end if
13:   for  $i = 1, \dots, N$  do
14:     Draw  $X_t^{(i)} \sim K_t(\cdot | X_{t-1}^{(i)})$ .
15:     Calculate  $\tilde{w}_t(X_{t-1}^{(i)}, X_t^{(i)}) = \frac{y_t(X_t^{(i)})L_{t-1}(X_t^{(i)}, X_{t-1}^{(i)})}{y_{t-1}(X_{t-1}^{(i)})K_t(X_{t-1}^{(i)}, X_t^{(i)})}$ .
16:   end for
17:   Normalize the weights:  $W_t^{(i)} = \frac{W_{t-1}^{(i)} \tilde{w}_t(X_{t-1}^{(i)}, X_t^{(i)})}{\sum_{j=1}^N W_{t-1}^{(j)} \tilde{w}_t(X_{t-1}^{(j)}, X_t^{(j)})}$ .
18: end for
19: Output: The final weighted sample set  $\{X_T^{(i)}, W_T^{(i)}\}_{i=1}^N$ .
```

2.2 Adaptive proposals

A critical factor influencing the performance of SMC samplers is the choice of forward Markov kernels K_t . Poorly chosen kernels lead to high variance in incremental weights, accelerating weight degeneracy and necessitating frequent resampling. In high-dimensional spaces, fixed proposal kernels often fail to adequately explore the target distribution.

Adaptive SMC methods dynamically tune proposal kernels using information from the particle population. The core idea of adapting a proposal mechanism during a simulation builds on foundational work in the adaptive MCMC literature ([Haario et al., 2001](#); [Andrieu and Atchadé, 2006](#)). A simple way to implement this within SMC is to set K_t as a random walk Metropolis kernel with empirical covariance estimates ([Jasra et al., 2011](#)). Specifically, at iteration t , we generate a proposal for each particle $X_{t-1}^{(i)}$ from a Normal distribution:

$$X^* \sim \mathcal{N}(X_{t-1}^{(i)}, \Sigma_{t-1}),$$

where Σ_{t-1} is the empirical covariance matrix of the particle population at time $t - 1$:

$$\Sigma_{t-1} = \frac{1}{N-1} \sum_{i=1}^N (X_{t-1}^{(i)} - \bar{X}_{t-1})(X_{t-1}^{(i)} - \bar{X}_{t-1})^\top.$$

The proposal is then accepted with probability

$$\alpha = \min \left(1, \frac{y_t(X^*)}{y_t(X_{t-1}^{(i)})} \right),$$

where γ_t is the unnormalized target density. If the proposal is accepted, we set $X_t^{(i)} = X^*$; otherwise, we retain the current particle: $X_t^{(i)} = X_{t-1}^{(i)}$.

This yields a Markov kernel K_t that satisfies detailed balance with respect to P_t , ensuring that P_t is invariant. The covariance Σ_{t-1} is updated at each step, allowing the random walk to adapt to local scaling and correlation in the target distribution. This adaptive scheme is particularly useful in high-dimensional settings, where isotropic or static proposals tend to perform poorly.

2.3 Estimating Normalizing Constants

In Bayesian model selection, a central quantity of interest is the marginal likelihood, or *normalizing constant*, defined as

$$Z_T = \int \gamma_T(\theta) d\theta = \int L(y | \theta) \pi(\theta) d\theta.$$

This integral is generally intractable in high dimensions, but its value is crucial for comparing models via Bayes factors.

Sequential Monte Carlo methods provide a natural estimator for Z_T , as the algorithm implicitly constructs an importance sampling estimate of each intermediate normalization constant Z_t , and hence of the final Z_T (Del Moral et al., 2006). Specifically, given the incremental weights $\tilde{w}_t(x_{t-1}, x_t)$, the marginal likelihood can be estimated by

$$\hat{Z}_T = \prod_{t=1}^T \left(\sum_{i=1}^N W_{t-1}^{(i)} \tilde{w}_t^{(i)} \right),$$

where $\tilde{w}_t^{(i)}$ is the incremental weight associated with the i -th particle at iteration t .

Another option is to use *path sampling*, also known as *thermodynamic integration* (Gelman and Meng, 1998). This method considers a continuous sequence of distributions $\pi_t(\theta) \propto \gamma_t(\theta)$, where the unnormalized density is defined as

$$\gamma_t(\theta) = L(y | \theta)^{\phi_t} \pi(\theta),$$

with $\phi_t \in [0, 1]$ interpolating between the prior ($\phi = 0$) and the posterior ($\phi = 1$). Under mild regularity conditions, the log-normalizing constant satisfies

$$\log Z_T - \log Z_0 = \int_0^1 \mathbb{E}_{\pi_t} [\log L(y | \theta)] d\phi_t.$$

This integral can be approximated numerically using evaluations at discrete values $\{\phi_t\}_{t=1}^T$, typically via the trapezoidal rule:

$$\log \hat{Z}_T = \sum_{t=2}^T (\phi_t - \phi_{t-1}) \cdot \frac{1}{2} (\bar{\ell}_t + \bar{\ell}_{t-1}),$$

where $\bar{\ell}_t = \frac{1}{N} \sum_{i=1}^N \log L(y | \theta_t^{(i)})$ is the empirical mean of the log-likelihood under π_t .

Both estimators can be computed within the same SMC run, offering complementary diagnostics and robustness in the estimation of marginal likelihoods.

3 Application: Gaussian Process Regression

We illustrate the application of Adaptive Sequential Monte Carlo in the context of Bayesian inference and model selection by sampling from the posterior and estimating the marginal likelihood in Gaussian Process (GP) regression. Specifically, we compare the performance of three methods: Importance Sampling (IS), Annealed Importance Sampling (AIS), and Adaptive SMC.

3.1 Model and setup

We consider a standard Gaussian Process regression model for a response vector $y \in \mathbb{R}^n$ given covariates $X \in \mathbb{R}^{n \times d}$. Under the GP prior, the latent function $f \sim \mathcal{GP}(0, k_\theta)$ is evaluated at the inputs X , yielding the model

$$y \mid X, \theta \sim \mathcal{N}(0, K_\theta + \sigma^2 I_n),$$

where K_θ is the $n \times n$ covariance matrix defined by a kernel function $k_\theta(x, x')$. To allow for input dimension-specific scaling, we employ a squared exponential kernel with Automatic Relevance Determination (ARD):

$$k_\theta(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{j=1}^d \frac{(x_j - x'_j)^2}{\ell_j^2}\right).$$

The full parameter vector is $\theta = (\ell_1, \dots, \ell_d, \sigma_f, \sigma)$, giving a total of $d + 2$ parameters to infer. We assign a LogNormal(0, 1) prior to each of these parameters independently.

For this example, we construct a synthetic dataset with $n = 300$ datapoints and an input space of $d = 50$. The design matrix X is generated by sampling each coordinate from a standard normal distribution. The responses y are then generated from the GP model using true parameters set to $\sigma_f = 1$ (signal variance) and $\sigma = 1$ (noise variance). The true lengthscales, $\{\ell_j\}_{j=1}^{50}$, were drawn uniformly over the interval $[100, 200]$.

Each of the methods being compared starts by sampling $N = 500$ particles from the prior distribution. For the Adaptive SMC and AIS algorithms, we use a sequence of tempered distributions

$$\pi_t(\theta) \propto L(y \mid \theta)^{\phi_t} \pi(\theta),$$

where the temperature ϕ_t increases linearly from 0 to 1 over $T = 5$ steps. Typically, finer schedules with more intermediate steps are used to ensure a smoother transition between distributions. However, a coarser schedule is employed here as it yielded the best results in terms of computational time. At each iteration, a Metropolis kernel with adaptively tuned Gaussian proposals is used to propagate particles, as described in Section 2.2. The resampling step in the Adaptive SMC algorithm is triggered whenever the ESS falls below the threshold of $N/2 = 250$.

3.2 Results

To evaluate the performance of the three methods, we executed 100 independent runs for each sampler. We analyzed the stability of the marginal likelihood estimates and the overall computational efficiency of each algorithm. The summarized results,

including 90% sample intervals calculated from the empirical 5th and 95th percentiles, are presented below.

The primary goal of these methods is often to compute the log marginal likelihood, $\log Z$. A reliable method should produce estimates with low variance across independent runs. Table 1 reports the mean estimate and the 90% sample interval for each sampler.

Method	Mean($\log Z$)	90% Sample Interval
Importance Sampling	-511.04	[-511.43, -510.72]
Annealed IS	-515.59	[-516.10, -515.11]
Adaptive SMC	-511.19	[-511.46, -510.90]

Table 1: Log Marginal Likelihood ($\log Z$) Estimation Results.

Note that the Adaptive SMC sampler achieves the tightest 90% sample interval, indicating it is the most stable among the three for evidence estimation. Notably, the interval for AIS is significantly different from the other two, suggesting that the sampler produced a biased result. With its coarse annealing schedule, the particles likely failed to transition correctly to the final posterior, leading to an inaccurate estimate of the normalizing constant. While the interval for IS is reasonably narrow, its practical utility is undermined by the poor particle quality, as shown next.

Beyond the evidence estimate, it is critical to assess the quality of the final particle set and the computational cost. We use the Effective Sample Size (ESS) as a measure of particle quality and the ESS generated per second of runtime (ESS/sec) as a holistic measure of efficiency.

Method	Time (s)	ESS	ESS / s
Importance Sampling	6.35	33.8	5.33
	[6.11, 6.92]	[25.6, 43.7]	[4.04, 6.89]
Annealed IS	22.09	41.2	1.87
	[21.30, 23.36]	[33.0, 48.7]	[1.47, 2.20]
Adaptive SMC	22.70	328.5	14.49
	[21.67, 24.37]	[301.4, 355.5]	[12.81, 15.79]

Table 2: Efficiency and Sample Quality Metrics, showing mean and (90% *interval*).

Importance Sampling was the fastest method but suffered from catastrophic weight degeneracy. An average ESS of only 33 out of 500 particles means the final sample provides a very poor approximation of the posterior distribution, making it unreliable for inference.

Annealed Importance Sampling did not resolve this issue. Despite being as computationally expensive as SMC, it produced a final particle set of equally poor quality to IS. Its extremely low efficiency (1.87 ESS/s) confirms it struggled with the high-dimensional posterior.

Adaptive SMC stands out as the most effective method. Although it required a similar runtime to AIS, it generated a high-quality particle set, retaining an average ESS of over 320. This combination of high-quality samples and reasonable runtime gives it the

highest overall efficiency, making it nearly 3 times more efficient than IS and 8 times more efficient than AIS.

In conclusion, the adaptive resampling and proposal mechanisms of the SMC sampler were essential for navigating the high-dimensional parameter space, preventing the weight collapse that rendered IS and AIS ineffective and yielding both stable evidence estimates and a useful posterior approximation.

4 Conclusion

In this work, we explored the application of Adaptive Sequential Monte Carlo for robust Bayesian inference and model selection in a high-dimensional setting. The computational experiment, centered on a Gaussian Process regression model, illustrated the limitations of conventional sampling methods when faced with a complex distribution.

The results demonstrate that standard Importance Sampling, while fast, is impractical due to weight degeneracy, yielding a particle set that fails to approximate the posterior. More notably, Annealed Importance Sampling, often considered a more robust alternative, also proved inadequate. It was not only computationally inefficient but also produced a significantly biased estimate of the marginal likelihood.

In contrast, the Adaptive SMC sampler proved to be a powerful and reliable tool. By dynamically tuning its proposal kernels using the empirical covariance of the particle set and rejuvenating the sample through resampling, it successfully navigated the complex posterior. This resulted in stable and accurate estimates of the marginal likelihood and, critically, a high-quality particle approximation of the posterior, as evidenced by its high Effective Sample Size.

This underscores a crucial point: for many high-dimensional static problems, merely introducing a temperature schedule is not enough. The success of an algorithm hinges on its ability to adapt. The mechanisms within the Adaptive SMC method provide this capability, making it a powerful tool for modern computational statistics on a wide variety of problems.

References

- Andrieu, C. and Atchadé, Y. F. (2006). On the efficiency of adaptive mcmc algorithms. In *Proceedings of the 1st International Conference on Performance Evaluation Methodologies and Tools*, valuetools '06, page 43–es, New York, NY, USA.
- Beskos, A., Crisan, D., and Jasra, A. (2014). Stability of sequential Monte Carlo methods in high dimensions. *The Annals of Applied Probability*, 24(4):1396–1445.
- Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.
- Doucet, A., De Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. Springer New York, NY.
- Gelman, A. and Meng, X.-L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical science*, pages 163–185.

- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.
- Jasra, A., Stephens, D. A., Doucet, A., and Tsagaris, T. (2011). Inference for Lévy-driven stochastic differential equations: a sequential Monte Carlo approach. *Scandinavian Journal of Statistics*, 38(1):1–21.
- Neal, R. M. (2001). Annealed importance sampling. In *Statistics and computing*, volume 11, pages 125–139. Springer.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer Science & Business Media.