Robust Auxiliary Particle Filters using Multiple Importance Sampling

Joel Kronander* and Thomas B. Schön†

UPPSALA UNIVERSITET

Summary

- We derive robust and efficient particle filters using deterministic mixture sampling via multiple importance sampling.
- The filters are easy to implement and in general perform on par with, or better than, the best of the standard proposal densities.
- For inference in state space models with heavy tailed transition and observational densities, the new filters can significantly outperform standard proposal distributions.

Auxiliary Particle Filters

Particle filters constitute a popular class of models for **inference** in **nonlinear/non-gaussian state space models (SSM)**

$$x_{t+1}|x_t \sim f(x_{t+1}|x_t),$$
$$y_t|x_t \sim g(y_t|x_t),$$

The **auxiliary particle filter (APF)** approximates the filtering distributions $\{p(x_t|y_{1:t})\}_{t=1}^T$ using a set of weighted particles that are propagated forward in time using resampling and sequential importance sampling. The choice of importance density is a crucial design choice for practical applications of particle filters. In general SSMs the optimal proposal function is intractable. In practice, a common choice is instead to **sample proportional** to either the **transition density** $f(\cdot|x_{t-1})$ or the **observation density** $g(y_t|\cdot)$.

Random mixture proposals

When the proposal function is a poor match for the target distribution, the resulting approximation can be extremely poor (infinite variance). Instead of sampling exclusively from either the transition or observation density, consider sampling from the **random mixture proposal**

$$q(x_t|x_{t-1}^{a_t}, y_t) = \alpha f(x_t|x_{t-1}^{a_t^i}) + (1-\alpha) q_g(x_t|y_t) \qquad (1)$$

where $\alpha \in [0, 1]$ is the mixture coefficient. This proposal function results in a finite variance of the importance weights if any of the mixture components provide a finite variance. Sampling from the random mixture distribution thus represents a robust choice when the optimal proposal function is intractable.

Main idea

Improve upon random mixture sampling by drawing a fixed number of samples from each proposal and combine them using unbiased weighting strategies via multiple importance sampling (MIS).

Multiple Importance Sampling

Let $\{q_p(x)\}_{p=1}^M$ denote a set of proposal distributions. Proposing n_p samples from $q_p(x)$ the MIS estimator of a target density $\pi(x)$ is given by

$$\widehat{\pi}(x) = \sum_{p=1}^{P} \frac{1}{n_p} \sum_{i=1}^{n_p} \beta_p(x^{p,i}) \frac{\pi(x^{p,i})}{q_p(x^{p,i})} \delta_{x^{p,i}}(x), \quad x^{p,i} \sim q_p(x), \qquad (2)$$

where $x^{p,i}$ denotes the i^{th} sample drawn from proposal q_p and $\beta_p(x) \geq 0$ are the **mixture weights** satisfying $\sum_{p=1}^M \beta_p(x) = 1$ for all x.

An APF proposing N_f particles from the transition density and N_g particles from the observation density can be derived by using a self-normalized MIS estimator.

Alg. 1 : MIS-APF, balance heuristic weights

1 - Sample $x_1^i \sim \mu(x_1)$ for $i = 1, \ldots, N = N_f + N_g$
2 - Set $w_t^i = \frac{1}{N}$ for $i = 1, \ldots, N$
3 For $t = 2, \ldots, T$
4 - Sample $a_t^i \sim Cat(\{v_{t-1}^j\}_{j=1}^N)$ for $i = 1, \ldots, N$
5 - Sample $x_t^i \sim f(x_t^i|x_{t-1}^{a_t^i})$ for $i = 1, \ldots, N_f$
6 - Sample $x_t^i \sim q_g(x_t^i|y_t)$ for $i = N_f + 1, \ldots, N$
7 - Set $\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} f(x_t^i|x_{t-1}^{a_t^i}) g(y_t|x_t^i)}{8} v_{t-1}^i (N_f f(x_t^i|x_{t-1}^{a_t^i}) + N_g q_g(x_t^i|y_t))$ for $i = 1, \ldots, N$
9 - Set $w_t^i = \frac{\widetilde{w}_t^i}{\sum_{i=1}^N \widetilde{w}_t^i}$ for $i = 1, \ldots, N$
10 End For

Alg. 2 :MIS-APF, low complexity weights

As Alg. 1 but with line 7 replaced with:

7a - Set $\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} g(y_t|x_t^i)}{v^i}$ for $i = 1, \ldots, N_f$

7b - Set $\widetilde{w}_t^i = \frac{w_{t-1}^{a_t^i} f(x_t^i|x_{t-1}^{a_t^i})}{v^i}$ for $i = N_f + 1, \ldots, N$

Numerical Illustration

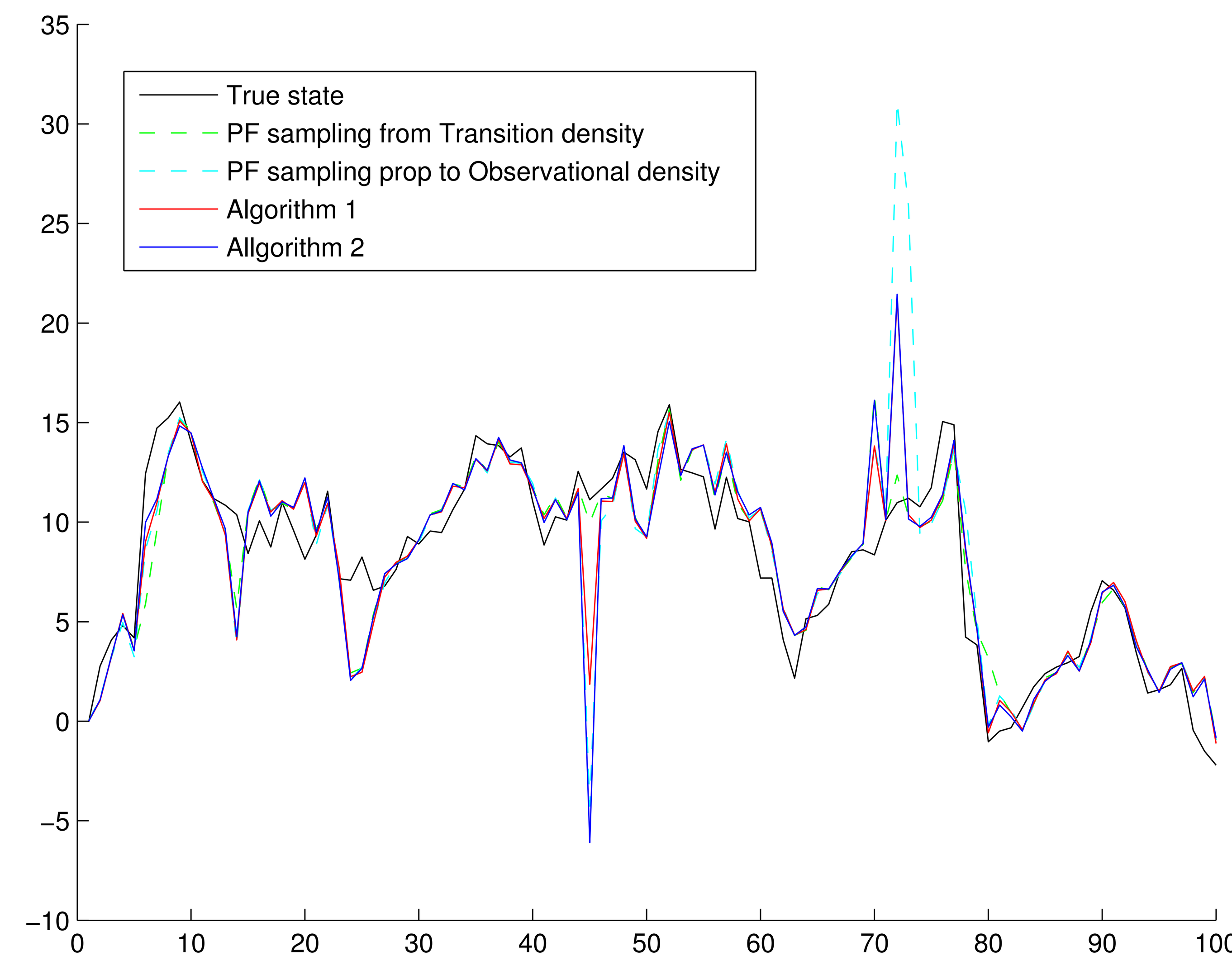
Consider the following state space model model

$$x_t = x_{t-1} + v_t, \qquad y_t = x_t + e_t$$

where $x_1 \sim \mathcal{N}(0, 0.1)$, $v_t \sim \text{St}(\nu_v)$ and $e_t \sim \text{St}(\nu_e)$, where $\text{St}(\nu)$ denotes the Student's t-distribution with ν degrees of freedom. The APF given in Algorithm 1 performs best for all considered parameter settings. The APF given in Algorithm 2 performs almost as good, and is for all parameter settings better than using an APF proposing from either the transition or proportional to the observation density for an equal computational cost.

σ_v	σ_e	Transition density	Observation density	Alg. 1	Alg. 2
2	2	56.1	9.1	**5.7**	5.9
2	3	135.4	2.3	**2.1**	**2.1**
3	2	4.8	8.8	**2.4**	2.5

Averaged MSE for the linear model (3) with student's t-distributed noise.



True states and inferred states for an example trajectory generated using the linear model (3) with $\sigma_v = 2$ and $\sigma_e = 2$.

More information and source code

vcl.itn.liu.se/publications/2014/KB14/

*joel.kronander@liu.se and †thomas.schon@it.uu.se