

Hand-in assignments

PhD course on Sequential Monte Carlo methods 2019

Martin Hellkvist
Signals and Systems, Uppsala University

September 17, 2019

H.1 Importance sampling theory

(a) Show that the estimator \hat{Z} is unbiased.

Solution: Plug the estimator $\hat{Z} = \frac{1}{N} \sum_{i=1}^N \frac{\tilde{\pi}(x^i)}{q(x^i)}$ into the expectation:

$$\mathbb{E}[\hat{Z}] = \frac{1}{N} \sum_{i=1}^N \mathbb{E} \frac{\tilde{\pi}(x^i)}{q(x^i)} = \frac{1}{N} N \mathbb{E}_X \frac{\tilde{\pi}(x)}{q(x)}, \quad (1)$$

where the last equality holds because the expectation is constant w.r.t the samples x^i . Using the definition of expectation, and that $\tilde{\pi}(x) := Z\pi(x)$ the expression becomes:

$$\mathbb{E}[\hat{Z}] = \int \frac{\tilde{\pi}(x)}{q(x)} q(x) dx = \int \tilde{\pi}(x) dx = \int Z\pi(x) dx = Z \int \pi(x) dx = Z. \quad (2)$$

So the expectation of the estimate is exactly the true value, hence the estimator is unbiased.

(b) Using the Cauchy distribution as proposal, implement an IS that samples from the standard normal distribution. Find the gamma that minimizes the asymptotic variance of the IS and provide plots for the approximation of the normalizing constant $\sqrt{2\pi}$.

Solution: the function to minimize is $\int \pi^2(x)/q(x) dx - 1$:

$$\gamma^* = \arg \min_{\gamma > 0} \int \frac{\pi^2(x)}{q(x)} dx - 1 = \arg \min_{\gamma > 0} \int \frac{\pi^2(x)}{q(x)} dx = \arg \min_{\gamma > 0} f(\gamma). \quad (3)$$

Plugging in the expressions $\pi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}$ and $q(x) = \frac{\gamma}{\pi(\gamma^2+x^2)}$, the function $f(\gamma)$ becomes

$$f(\gamma) = \int \frac{\pi^2(x)}{q(x)} dx = \int \frac{1}{2\pi} \frac{\exp\{-x^2\}}{\frac{\gamma}{\pi(\gamma^2+x^2)}} dx = \int \frac{\exp\{-x^2\}}{2} \frac{\gamma^2+x^2}{\gamma} dx = \quad (4)$$

$$= \int \frac{\gamma^2}{2\gamma} \exp\{-x^2\} dx + \int \frac{1}{2\gamma} x^2 \exp\{-x^2\} dx. \quad (5)$$

The integration is over the whole real axis for x , and the expression becomes

$$f(\gamma) = \frac{\gamma^2}{4\gamma} \sqrt{\pi} \operatorname{erf}(x) + \frac{1}{8\gamma} \left(\sqrt{\pi} \operatorname{erf}(x) - 2 \exp\{-x^2\} x \right) \Bigg|_{x \rightarrow -\infty}^{x \rightarrow +\infty} = \quad (6)$$

$$= \frac{(2\gamma^2+1)\sqrt{\pi} \operatorname{erf}(x) - 2 \exp\{-x^2\} x}{8\gamma} \Bigg|_{x \rightarrow -\infty}^{x \rightarrow +\infty} = \quad (7)$$

$$= \frac{\left((2\gamma^2+1)\sqrt{\pi} \cdot 1 - 2 \cdot 0 \right) - \left((2\gamma^2+1)\sqrt{\pi} \cdot (-1) - 2 \cdot 0 \right)}{8\gamma} = \frac{(2\gamma^2+1)}{4\gamma} \sqrt{\pi}, \quad (8)$$

where $\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the *error function*.

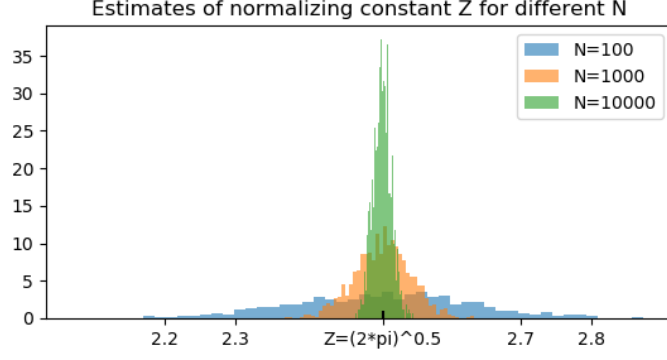


Figure 1: For three different values of N we see different variance in the estimate of Z . As N increases, the variance decreases. The average of each histogram is close to the true value $Z = \sqrt{2\pi}$. The histogram was generated with 1000 simulations for each N .

The function $f(\gamma)/\sqrt{\pi} = \frac{(2\gamma^2+1)}{4\gamma} = \frac{\gamma}{2} + \frac{1}{4\gamma}$ is convex for $\gamma > 0$ and has a unique minimum at $\gamma^* \rightarrow \frac{d}{d\gamma} \left(\frac{\gamma}{2} + \frac{1}{4\gamma} \right) = 0$ which is found as

$$\frac{1}{2} - \frac{1}{4\gamma^2} = 0 \Leftrightarrow \gamma = 1/\sqrt{2}. \quad (9)$$

The simulation results illustrated in figure 1 demonstrates how the implemented IS is unbiased.

(c) Now target the Cauchy distribution using the standard normal distribution.

Solution: The variance of the estimator is now

$$\int \frac{\pi^2(x)}{q(x)} dx - 1 = \int \frac{\gamma^2}{\pi^2(\gamma^2 + x^2)^2} \frac{1}{\frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}} dx - 1 = \int \frac{\gamma^2}{\pi^2(\gamma^2 + x^2)^2} \sqrt{2\pi} \exp\{x^2\} dx - 1, \quad (10)$$

which does not converge, i.e., the variance is infinite. Although the estimator is theoretically unbiased, it is numerically infeasible to attain an unbiased estimate.

Figure 2 illustrates the IS with the standard normal distribution as proposal, and figure 3 instead shows the histograms generated with the normal distribution with standard deviation 20 instead. An increased variance of the proposal decreases the bias that is actually implicitly introduced by this infinite variance.

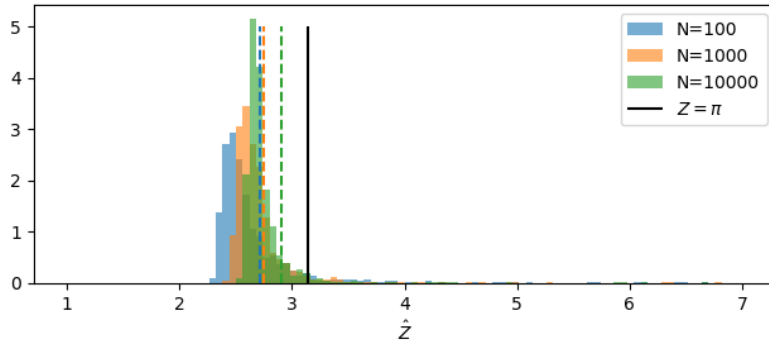


Figure 2: Task 1c. The distribution of the estimated normalization was skewed due to thin tails of the proposal.

H.2 Particle filter for a linear Gaussian state

(a) The pdf's are $p(x_t|x_{t-1}) = \mathcal{N}(x_t; 0.8x_{t-1}, 0.5)$ and $p(y_t|x_t) = \mathcal{N}(y_t; 2x_t, 0.1)$.

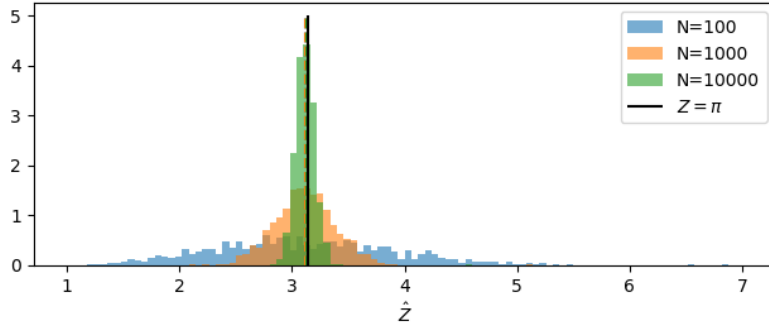


Figure 3: Task 1c. When the tails of the proposal were heavier, the estimates clearly clearly illustrated the unbiasedness of the IS.

N	MAD	Variance
10	0.1489	0.020
50	0.0412	0.023
100	0.0271	0.023
2000	0.0060	0.024
5000	0.0042	0.024
Kalman	0.0000	0.025

Table 1: 2c. Simulation results for the Bootstrap Particle filter

(b) Because we perfectly know the noise covariances, and that the model matches the simulated system, the Kalman filter is the optimal linear filter. The particle filter is on the other hand can model nonlinearities in the data. Therefore, we can expect the particle filter's estimated trajectory to be closer to the true one, if we number of particles is large enough.

(c) After implementing the bootstrap filter, the variance was computed for each N as $Var(\hat{x}_t) = \sum_{i=1}^N w_t^i(x_t^i)^2 + (\sum_{i=1}^N w_t^i x_t^i)^2$. The Kalman filter's variance converges after only a few iterations. The variances are plotted in figure 4.

Letting MAD denote the absolute difference between the BPF state estimate and the Kalman state estimate, averaged over time, the results together with average variance of the BPF is presented in table 1.

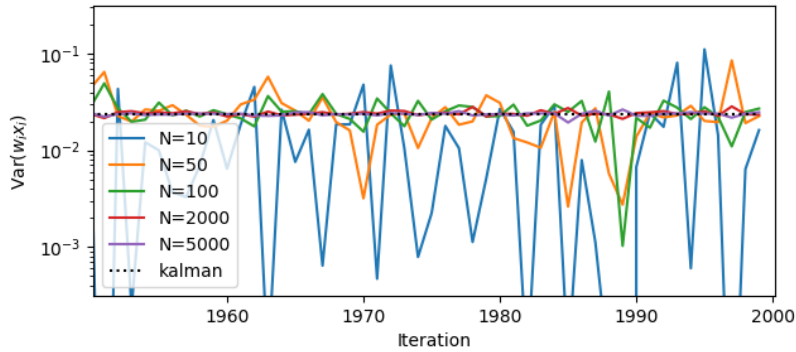


Figure 4: 2c. Variance over particles per iteration of the BPF state estimates compared to the Kalman variance.

(d) Derivation of the Fully Adapted Particle Filter pdf's follows. The first proposal, for the adjustment multipliers ν :

$$\nu_{t-1}^i = p(y_t | x_{t-1}^i) = \frac{p(y_t, x_{t-1}^i)}{p(x_{t-1}^i)} \propto p(y_t, x_{t-1}^i). \quad (11)$$

N	MAD
10	0.0388
50	0.0181
100	0.0122
2000	0.0028
5000	0.0019

Table 2: 2d. The mean average difference between the Fully Adapted Particle filter's state trajectory and the Kalman filter's state trajectory.

From the probabilistic model we have

$$\mathcal{N}(y_t; 2x_t^i, 0.1), \quad (12)$$

from which we can plug in the model for x_t^i :

$$\mathcal{N}(y_t; 2(0.8x_{t-1}^i + V_t), 0.1) = \mathcal{N}(y_t; 1.6x_{t-1}^i + 2V_t, 0.1) = \mathcal{N}(y_t; 1.6x_{t-1}^i, 4 \cdot 0.5 + 0.1), \quad (13)$$

thus, we have

$$p(y_t|x_{t-1}^i) = \mathcal{N}(y_t; 1.6x_{t-1}^i, 2.1). \quad (14)$$

For the optimal proposal q :

$$q(x_t|x_{t-1}, y_t) \propto p(y_t|x_t, x_{t-1})p(x_t|x_{t-1}) = \mathcal{N}(y_t; 2x_t, 0.1)\mathcal{N}(x_t; 0.8x_{t-1}, 0.5) \propto \quad (15)$$

$$\propto \exp\left\{-\frac{(y_t - 2x_t)^2}{2 \cdot 0.1}\right\} \exp\left\{-\frac{(x_t - 0.8x_{t-1})^2}{2 \cdot 0.5}\right\} \propto \quad (16)$$

$$\propto \exp\left\{-\frac{(x_t - \frac{1}{42}(20y_t - 1.6x_{t-1}))^2}{2 \cdot \frac{1}{42}}\right\}, \quad (17)$$

so we get the normal distribution q :

$$q(x_t|x_{t-1}, y_t) = \mathcal{N}(x_t; \frac{1}{42}(20y_t - 1.6x_{t-1}), \frac{1}{42}) \quad (18)$$

Table 2 presents the MAD between the FAPF and the Kalman state trajectories for one simulation. Compared to the results for the BPF, the MAD is now a lot lower for small N , while they are still lower for larger N 's but not as strikingly.

(e) Figure 5 illustrates how only a few particles at time 1950 are remaining as ancestors for the final particles.

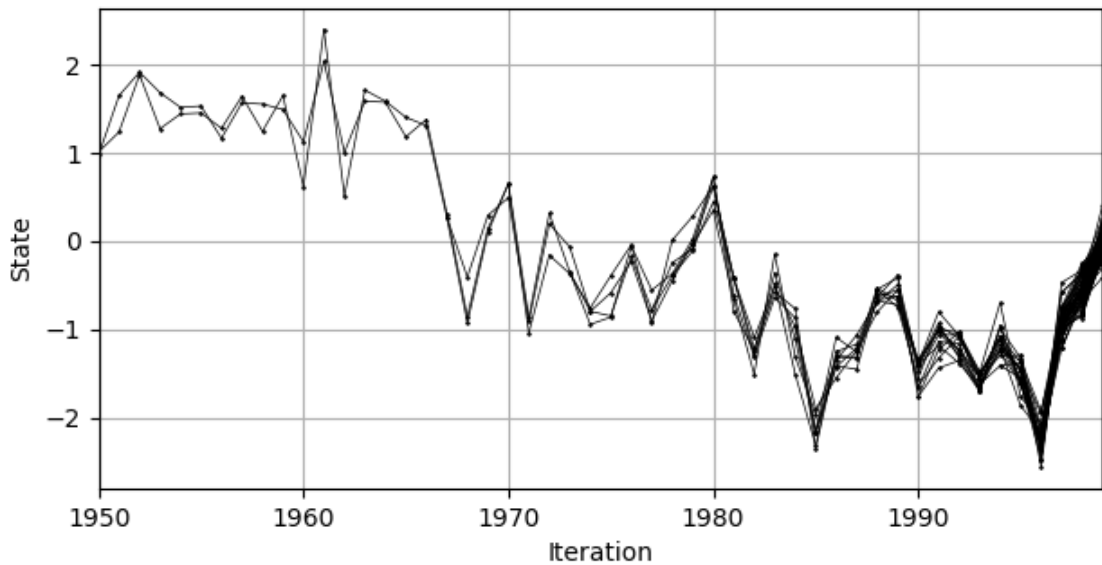


Figure 5: 2e. Particle genealogy from time 1950 to 1999 for the fully adaptive particle filter.

(f) With systematic resampling the FAPF obtains less degeneracy and the final particles have many more ancestors. As illustrated in figure 6, at iteration 2 there are around 30 particles who are ancestors to the final 100 particles.

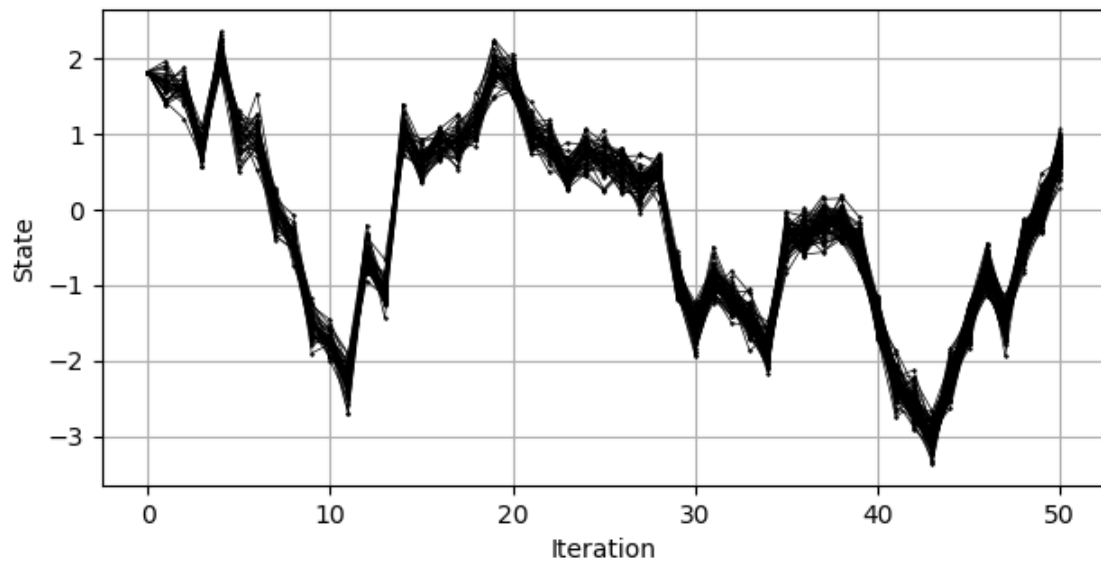


Figure 6: 2f. The early genealogy for the final 100 particles.