Exercises set IV
PhD course on Sequential Monte Carlo methods 2021

Linköping University and Uppsala University

August 17, 2021

This document contains exercises to make you familiar with the content of the course. *The exercises in this document are not mandatory, and you do not need to hand in your solutions.* The mandatory assignment is found in a separate document named "Hand-in". We strongly recommend that you carefully work through these exercises before starting with the mandatory assignments.

IV.1 Particle Metropolis-Hastings

Consider the following state-space model

$$x_t = \cos(\theta x_{t-1}) + v_t, \qquad v_t \sim \mathcal{N}(0,1) \tag{1a}$$
$$y_t = x_t + e_t, \qquad e_t \sim \mathcal{N}(0,1) \tag{1b}$$
$$x_0 \sim \mathcal{N}(0,1). \tag{1c}$$

Generate $T = 50$ data points $y_{1:T}$ from this model with $\theta = 1$, and then pretend that you forgot the true value of θ, but assume that you drew it from $\mathcal{N}(0,1)$. Use particle Metropolis-Hastings to infer the posterior distribution of θ, $p(\theta \mid y_{1:T})$, as follows:

(a) Implement a particle filter of your choice to deliver an estimate $\widehat{z}_\theta$ of the likelihood $p(y_{1:T} \mid \theta)$.

(b) Design a proposal $q(\theta' \mid \theta[k-1])$ for θ' as, e.g., a Gaussian random walk on θ-space.

(c) Implement a Metropolis-Hastings sampler, where the acceptance ratio is computed as

$$\min\left(1, \frac{\widehat{z}_{\theta'}}{\widehat{z}_{\theta[k-1]}} \frac{p(\theta')}{p(\theta[k-1])} \frac{q(\theta[k-1] \mid \theta')}{q(\theta' \mid \theta[k-1])}\right) \tag{2}$$

($p(\theta)$ is the prior).

Note: the standard Metropolis-Hastings algorithm requires the target, in this case $p(\theta \mid y_{1:T}) \propto p(y_{1:T} \mid \theta)p(\theta)$, to be evaluated exactly (up to proportionality). Here, however, we only have access to estimates $\widehat{z}_\theta$ of $p(y_{1:T} \mid \theta)$, and the fact that this still works is depending on the so-called pseudo-marginal Metropolis-Hastings. The key here is the unbiasedness of $\widehat{z}_\theta$.

IV.2 **Conditional particle filter**

Return to any of the particle filter implementations you have done earlier, and turn it into a *conditional* particle filter as follows:

- Introduce the concept of a reference state trajectory $x_{1:T}[k-1]$
- In each propagation step, use multinomial resampling but replace the propagation of the N:th particle x_{t-1}^N (after resampling) to x_t^N by just taking $x_t[k-1]$ from the reference state trajectory (instead of sampling from $p(x_t \,|\, x_{t-1}^N)$).
- After the final time step, $t = T$, sample a new reference trajectory $x_{1:T}[k]$ by sampling j from the categorical distribution defined by $\{w_T^i\}_{j=1}^N$, and take x_T^j and its ancestors $x_{T-1}^{a_t^j}$ etc.

This is the particle Gibbs Markov kernel, which maps one reference state trajectory $x_{1:T}[k-1]$ onto another $x_{1:T}[k]$, which can be combined with sampling the unknown parameters in a Gibbs fashion (i.e., sample θ conditional on $x_{1:T}[k-1]$ as $\theta[k] \sim p(\theta \,|\, x_{1:T}[k-1])$).

IV.3 **Conditional importance sampling**

(a) Implement an MCMC procedure to sample from $\pi(x) = \mathcal{N}(x \,|\, 1, 1)$ by using a conditional importance sampling kernel with proposal $q(x) = \mathcal{N}(x \,|\, 0, 1)$. Verify that you get samples from the target in the long run, even using $N = 2$ particles in the importance sampler.

(b) Consider a general conditional importance sampling kernel $\kappa_N(x, x^\star)$ with target $\pi(x)$, proposal $q(x)$, and with $N = 2$ samples. The method is then very similar to a Metropolis–Hastings sampler, with a proposal distribution which is independent of the current state of the Markov chain. The reason is that each iteration of the sampler involves generating a single sample from the proposal $x^1 \sim q(x)$ (since $N - 1 = 1$). The next state of the Markov chain will either be equal to this proposed sample (acceptance, $x^\star = x^1$), or the previous state of the Markov chain is retained (rejection, $x^\star = x$).

Compute the "acceptance probability" $\alpha(x, x^1) = \mathbb{P}\left(x^\star = x^1 | x, x^1\right)$ of this conditional importance sampler, and show that the kernel satisfies detailed balance,

$$\alpha(x, x^1) q(x^1) \pi(x) = \alpha(x^1, x) q(x) \pi(x^1).$$

(c) Show that the acceptance probability of conditional importance sampling with $N = 2$ (computed in the previous exercise) is bounded from above by the acceptance probability of the independent Metropolis–Hastings sampler using the same proposal.

Note: This means that Metropolis–Hastings is more efficient (accepts more frequently) than conditional importance sampling with $N = 2$. The benefit of conditional importance sampling is that it enables us to use $N > 2$ and still obtain a valid MCMC kernel. This means that we can propose many ($N - 1$, to be specific) new samples in parallel, and the probability of moving away from the current state of the Markov chain can be made arbitrarily large by increasing N.

IV.4 **Conditional importance sampling: proof of validity** The conditional importance sampling kernel $\kappa_N(x, x^\star)$ has $\pi(x)$ as its stationary distribution for any $N \geq 1$. For $N = 1$ the result is trivial. For $N = 2$ it follows from the detailed balance condition derived above. However, proving the result for arbitrary N requires a different approach that we will investigate in this exercise.

We will carry out the proof for a slightly different algorithm than the one presented in the lecture, which however is probabilistically equivalent. This modification makes the proof simpler.

The modified algorithm is as follows: Given x,
Step 1:

- Draw $b \sim \mathcal{U}(\{1, \ldots, N\})$
- Set $x^b = x$ (input state)
- Draw $x^i \sim q(x)$, $i \in \{1, \ldots, N\} \setminus \{b\}$

- Compute $\widetilde{w}^i = \widetilde{\pi}(x^i)/q(x^i)$, $i = 1, \ldots, N$ and normalize

Step 2:

- Draw $b^\star \sim \mathcal{C}(\{w^i\}_{i=1}^N)$
- Output $x^\star = x^{b^\star}$

(a) Make sure that you understand that the procedure above generates a draw from the conditional importance sampling kernel $\kappa_N(x, x^\star)$!

(b) To prove that $\pi(x)$ is a stationary distribution of $\kappa_N(x, x^\star)$ we will treat the "internal" random variables $(b, \{x^i\}_{i\neq b})$, generated by the procedure above, as auxiliary variables. Assume that the input state x is distributed according to $\pi(x)$. Then, what is the (marginal) distribution of the random variables $(x^{1:N}, b)$ after completing Step 1 of the procedure above? This distribution, which we can denote by $\pi(x^{1:N}, b)$, is often referred to as the *extended target*.

(c) Verify that Step 2 of the sampling procedure corresponds to a Gibbs step for the extended target, $b^\star \sim \pi(b \,|\, x^{1:N})$.

(d) Still assuming that $x \sim \pi(x)$, what is the (marginal) distribution of the random variables $(x^{1:N}, b^\star)$ after completing Steps 1–2 of the sampling procedure above? Specifically, what is the marginal distribution of the output variable $x^\star = x^{b^\star}$?

(e) The conclusion from the exercises above is (hopefully!) that the conditional importance sampling kernel $\kappa^N(x, x^\star)$ has $\pi(x)$ as a stationary distribution. However, for it to be a valid MCMC kernel, we also need it to be *ergodic*. In this and the next exercise we look into the ergodicity of the kernel.

Let $A \subset \mathcal{X}$. Note that we can write $\int_A \kappa_N(x, x^\star)\mathrm{d}x^\star$, i.e. the probability that a draw from the conditional importance sampling kernel falls in A, as:

$$\int_A \kappa_N(x, x^\star)\mathrm{d}x^\star = \mathbb{E}\left[\frac{\sum_{i=1}^N \mathbb{1}(X^i \in A)\omega(X^i)}{\sum_{j=1}^N \omega(X^j)}\right]$$

Assume that $\omega(x) \leq c$ and verify that

$$\mathbb{E}\left[\frac{\sum_{i=1}^N \mathbb{1}(X^i \in A)\omega(X^i)}{\sum_{j=1}^N \omega(X^j)}\right] \geq \left(1 - \frac{1}{1 + d(N-1)}\right)\int_A \pi(x)\mathrm{d}x$$

for some constant d.

This is referred to as a minorization of the conditional importance sampling kernel.

(f) Using the minorization condition, show that we can decompose

$$\kappa_N(x, x^\star) = (1 - \varepsilon_N)\pi(x^\star) + \varepsilon_N r_N(x, x^\star) \tag{3}$$

for some constant $\varepsilon_N < 1$ and Markov kernel $r_N(x, x^\star)$. Furthermore, let x_m denote the m'th step of a Markov chain with kernel $\kappa_N(x, x^\star)$, now prove that

$$|\mathbb{P}(X_m \in A) - \pi(A)| \leq \varepsilon_N^m,$$

for arbitrary initial state x.

Hint: Consider runing two copies of the Markov chain with kernel κ_N. By using (3) one joint step of these two Markov chains can be done in the following way:

(1) Flip an ε_N coin, a coin showing head w.pr. ε_N.

(2) If head, draw independently for each chain a new position using r_N.
If tails, draw a position from π and couple the two chains, i.e. in the future the two chains will always be at the same position.

Study the individual chains, do they evolve according to κ_N? What is the probability that the two chains are different? Use this to prove the bound.

IV.5 An SMC sampler for localization

We want to localize an object positioned at x_0 in the world $[-12, 12]^2$, as illustrated in Figure 1. We have access to a bunch of measurements $y_{1:M}$ of the position, corrupted by heavy-tail noise. As the measurements are corrupted by heavy-tailed noise from the exponential distribution, we are not really interested in just a point estimate of x_0, but instead we want the entire posterior distribution $p(x_0|y_{1:M})$ of its position, reflecting the uncertainty inherent in the problem.

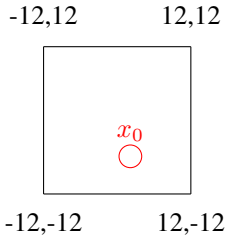


Figure 1: Illustration of our world $[-12, 12]^2$, with the true position x_0 indicated using a red circle.

(a) Prepare code for simulating M independent measurements from the following model

$$y_t^1 = x_0^1 + n_m^1 b_m^1, \tag{4a}$$
$$y_t^2 = x_0^2 + n_m^2 b_m^2, \tag{4b}$$

for $m = 1, \ldots, M$, where x_0^1 and x_0^2 are the components of x_0, and n_m^1 and n_m^2 are exponentially distributed with scale parameter 2, and $\mathbb{P}\left(b_m^1 = 1\right) = \mathbb{P}\left(b_m^1 = -1\right) = \frac{1}{2}$ and similarly for b_m^2.

(b) Prepare code for evaluating the density $p(y_m \,|\, x_0)$ based on the definitions above. This defines the likelihood in your problem, $p(y_{1:m} \,|\, x_0) = \prod_{m=1}^{M} p(y_m \,|\, x_0)$.

(c) Based on our background knowledge on the problem, you know that a reasonable prior for the position x_0 is the following,

$$p(x_0) = \mathcal{N}\left(x_0 \,|\, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix}\right). \tag{5}$$

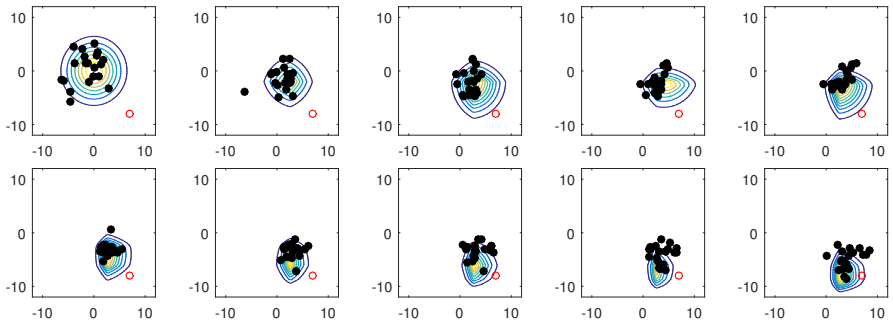
Design a likelihood tempered transition $\pi_0, \ldots, \pi_P$ from the prior $\pi_0(x_0) = p(x_0)$ to the posterior $\pi_P(x_0) = p(x_0|Y) \propto p(x_0) \prod_{t=1}^{T} p(y_t|x_0)$.

(d) Implement a π_n-invariant Metropolis-Hastings (MH) kernel based on a random walk proposal.

(e) Put everything together in an SMC sampler .

*There are alternative ways to update the particle weights in an SMC sampler. The easiest is probably to set the weight of particle x_{n-1}^i as $\pi_n(x_{n-1}^i)/\pi_{n-1}(x_{n-1}^i)$ [1, eq. (31)].

(f) Test your algorithm by making sure that it converges to something close to x_0 when, say, $N = 100$ particles and $M = 50$ measurements are used. Make plots to follow the evolution of the particles in the SMC sampler, such as



The black dots are the particles, the red circle is the true position of x_0 and the contours are proportional to π_n, for $n = 1, \ldots, 10$.

(g) Make a comparison between the SMC sampler and simply using the Metropolis-Hastings sampler to sample from π.

(h) OPTIONAL. To make the problem we are solving more interesting, assume another measurement model. Instead of measuring the objects absolute position, we are now measuring its relative distance to some sensors s_j,

$$y_j = \|x_0 - s_j\| + n, \tag{6}$$

where n still is exponentially distributed, and s_j denotes the coordinates for sensor j. That is, the distance between the sensor and the object is known (except for the noise), but not the angle.

Update your code using this measurement model instead. What does the posterior look like for the case of only one sensor in the origin? What about multiple sensors with different locations? How well is the SMC sampler doing in each scenario, respectively?

References

[1] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. "Sequential Monte Carlo samplers". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.3 (2006), pp. 411–436.