Exercises set III
PhD course on Sequential Monte Carlo methods 2019

Department of Information Technology, Uppsala University

August 21, 2019

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.

III.1 **Metropolis-Hastings.**
RECOMMENDED PROBLEM IF YOU HAVE NEVER IMPLEMENTED MCMC/METROPOLIS-HASTINGS BEFORE.

Assume that you are interested in samples from the following distribution:

$$\pi(x) \propto \sin^2(x) \exp(-|x|) \qquad (x \in \mathbb{R}) \tag{1}$$

Implement a Metropolis-Hastings sampler to generate samples from $\pi(x)$. Use a Gaussian random walk as proposal $q(x \mid x') = \mathcal{N}(x \mid x', \sigma^2)$, and plot your result as an histogram with π overlaid. Try different values of σ^2 (i.e., tune the proposal) and see how it affects the result.

Algorithm 1 Metropolis Hastings (MH)

---

1. **Initialize:** Set the initial state of the Markov chain $x[1]$.
2. **For** $i = 1$ **to** M**, iterate:**
   a. Sample $x' \sim q(x \mid x[i])$.
   b. Sample $u \sim \mathcal{U}[0, 1]$.
   c. Compute the acceptance probability
   $$\alpha = \min\left(1, \frac{\pi(x')}{\pi(x[i])} \frac{q(x[i] \mid x')}{q(x' \mid x[i])}\right)$$
   d. Set the next state $x[i+1]$ of the Markov chain according to
   $$x[i+1] = \begin{cases} x' & \text{if } u \le \alpha \\ x[i] & \text{otherwise} \end{cases}$$

III.2 **Gibbs sampling**

RECOMMENDED PROBLEM IF YOU HAVE NEVER IMPLEMENTED MCMC/GIBBS SAMPLING BEFORE.

Sample from the 2-dimensional Gaussian distribution

$$\pi(x) = \mathcal{N}\left(x \mid \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \begin{bmatrix} 0.3 & 0.1 \\ 0.1 & 1 \end{bmatrix}\right) \tag{2}$$

by using Gibbs sampling for each component. Start in $(0, 0)$, and plot your result.

Algorithm 2 Gibbs sampler for a 2-dimensional random vector $x \triangleq [x^1 \; x^2]$

Initialize: Set the initial state of the Markov chain $x[0]$.
For $i = 1$ **to** M**, iterate:**
Sample $x^1[i] \sim \pi(x^1 \mid x^2[i-1])$
Sample $x^2[i] \sim \pi(x^2 \mid x^1[i])$

Here, $\pi(x^1 \mid x^2)$ means the conditional distribution of x^1 given x^2 under the target distribution π. $x \triangleq [x^1 \; x^2]$.

Hint: Use that the if

$$p(x) = \mathcal{N}(x \mid \mu, \mathbf{\Sigma}) \tag{3a}$$

with

$$x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}, \qquad \mu = \begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \qquad \mathbf{\Sigma} = \begin{bmatrix} \sigma_{aa} & \sigma_{ab} \\ \sigma_{ab} & \sigma_{bb} \end{bmatrix}, \tag{3b}$$

then

$$p(x_a \mid x_b) = \mathcal{N}(x_a \mid, \mu_{a|b}, \sigma_{a|b}) \tag{3c}$$

where

$$\mu_{a|b} = \mu_a + \frac{\sigma_{ab}}{\sigma_{bb}}(x_b - \mu_b), \qquad \sigma_{a|b} = \sigma_{aa} - \frac{\sigma_{ab}^2}{\sigma_{bb}}. \tag{3d}$$

III.3 **Resampling**

Randomly generate 100 particles x^i from some distribution π of your choice, and 100 (positive) weights w^i. Normalize the weights such that $\sum_i w^i = 1$, and use the weighted samples $\{x^i, w^i\}$ to estimate the mean m of π, and denote this estimate by $\widehat{m}$.

(i) Resample the particles x^i (from the weights w^i) using multinomial resampling, and estimate the mean from the resampled (now equally weighted) samples. Denote this estimate $\widehat{m}_m$.

(ii) Repeat (i) for systematic resampling, and denote this estimate $\widehat{m}_s$.

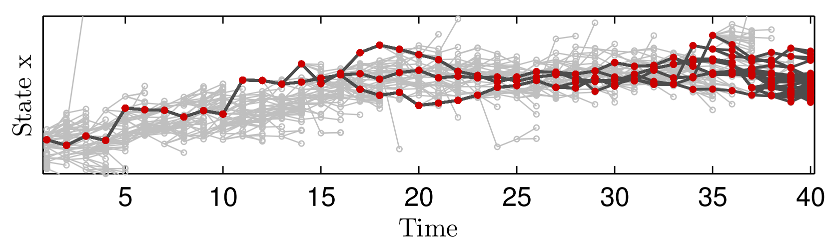
(iii) Repeat (i) for stratified resampling, and denote this estimate $\widehat{m}_t$.

Note that both $\widehat{m}$, $\widehat{m}_m$, $\widehat{m}_s$ and $\widehat{m}_t$ are unbiased estimates of the mean m. In particular is $\mathbb{E}[\widehat{m}] = m$ (where the expectation is over the randomness in the sample and weight generation), and $\mathbb{E}[\widehat{m}_m] = \mathbb{E}[\widehat{m}_s] = \mathbb{E}[\widehat{m}_t] = \widehat{m}$ (where the expectation is over the randomness in the resampling procedure). (Can you prove this?) But even though the resampling is unbiased, the variance of the estimators $\widehat{m}_m$, $\widehat{m}_s$ and $\widehat{m}_t$ is always[1] larger than (or possibly equal to) the variance of $\widehat{m}$. That is, the resampling 'adds' variance. We will now try to quantify this, for this example:

Repeat (i), (ii) and (iii) multiple times, and report an estimate of the variance for $\widehat{m} - \widehat{m}_m$, $\widehat{m} - \widehat{m}_s$, and $\widehat{m} - \widehat{m}_t$ respectively, conditionally on $\widehat{m}$ (that is, do not sample new particles from π, but only repeat the resampling step). Which resampling scheme appears to be the preferred one, in terms of variance?

III.4 **Path-space view**

Return to the stochastic volatility model in problem in I.4, and plot the genealogy of the particles at time T (i.e., the ancestral line to all particles x_T: the ancestor to a particle is determined by the resampling.), and confirm that degeneracy occurs. The plot could look something like this:



Try both multinomial and systematic resampling. Is there any difference in how quickly the paths degenerate? What happens if you add ESS-triggered resampling (i.e., perform the resampling only when ESS goes below a certain threshold)?

[1] If the Rao-Blackwell theorem is familiar to you, you may try to prove this.