Advanced Simulation

Problem Sheet 4

Exercise 1 (Kalman filter)

Consider a hidden Markov model defined as follows. $\{X_t\}_{t\geq 0}$ is a latent real-valued autoregressive Gaussian process defined by $X_0 \sim \mathcal{N}\left(m_0, \sigma_0^2\right)$ and

$$X_t = \phi X_{t-1} + V_t$$

where $V_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_V^2\right)$. We observe $\{Y_t\}_{t\geq 1}$ given by

$$Y_t = X_t + W_t$$

where $W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_W^2\right)$. For any generic sequence $\{z_i\}_{i\geq 1}$, we denote $(z_i, z_{i+1}, ..., z_j)$ by $z_{i:j}$ for $i < j$.

1. Give the expression of $f\left(x_t \mid x_{t-1}\right)$ and $g\left(y_t \mid x_t\right)$.
2. Show that if $p\left(x_t \mid y_{1:t}\right) = \mathcal{N}\left(x_t; m_{t|t}, \sigma_{t|t}^2\right)$ then $p\left(x_{t+1} \mid y_{1:t}\right) = \mathcal{N}\left(x_{t+1}; m_{t+1|t}, \sigma_{t+1|t}^2\right)$ and give the expressions of $m_{t+1|t}, \sigma_{t+1|t}^2$ as a function of $m_{t|t}, \sigma_{t|t}^2, \phi$ and σ_V^2.
3. Show that if $p\left(x_{t+1} \mid y_{1:t}\right) = \mathcal{N}\left(x_{t+1}; m_{t+1|t}, \sigma_{t+1|t}^2\right)$ then $p\left(x_{t+1} \mid y_{1:t+1}\right) = \mathcal{N}\left(x_t; m_{t+1|t+1}, \sigma_{t+1|t+1}^2\right)$ and give the expressions of $m_{t+1|t+1}, \sigma_{t+1|t+1}^2$ as a function of $m_{t+1|t}, \sigma_{t+1|t}^2$ and σ_W^2.
4. Show that if $p\left(x_{t+1} \mid y_{1:t}\right) = \mathcal{N}\left(x_{t+1}; m_{t+1|t}, \sigma_{t+1|t}^2\right)$ then $p\left(y_{t+1} \mid y_{1:t}\right) = \mathcal{N}\left(y_{t+1}; \mu_{t+1|t}, \beta_{t+1|t}^2\right)$ and give the expressions of $\mu_{t+1|t}, \beta_{t+1|t}^2$ as a function of $m_{t+1|t}, \sigma_{t+1|t}^2$ and σ_W^2.

Exercise 2 (SIS filter)

Consider a Hidden Markov Model (HMM) over 2 time steps,

$$\begin{aligned}
&X_0 \sim \mu(\cdot), \\
&Y_0 \mid (X_0 = x_0) \quad \sim g(\cdot|x_0), \\
&X_1 \mid (X_0 = x_0, Y_0 = y_0) \quad \sim f(\cdot|x_0), \\
&Y_1 \mid (X_{0:1} = x_{0:1}, Y_0 = y_0) \quad \sim g(\cdot|x_1).
\end{aligned}$$

We consider the following *sequential importance sampling* (SIS) algorithm. Let $N > 0$ be the number of particles.

- At time $t = 0$,
  - For $i = 1, ..., N$, sample $X_0^{(i)} \sim \mu$ and set $W_0^{(i)} = g(y_0|X_0^{(i)})$.
  - Normalize $\left\{W_0^{(i)}\right\}_{i=1}^n$ so they sum to 1.
  - Let $\hat{\nu}_0 = \sum_{i=1}^n W_0^{(i)} \delta_{X_0^{(i)}}$.
- At time $t = 1$,
  - For $i = 1, ..., n$, sample $X_1^{(i)} \sim f(\cdot|X_0^{(i)})$ and set $W_1^{(i)} = g(y_1|X_1^{(i)})W_0^{(i)}$.

- Normalize $\left\{W_1^{(i)}\right\}_{i=1}^{n}$ so they sum to 1.
- Let $\nu_t = \sum_{i=1}^{n} W_1^{(i)} \delta_{X_1^{(i)}}$.

1. What distribution does ν_t approximate? Why is this algorithm called a filter?
2. Express the marginal density $p_{Y_0}(y_0)$ and the joint density $p_{Y_0,Y_1}(y_0, y_1)$ as integrals involving μ, f, and g.
3. In terms of the particles generated in the SIS algorithm, suggest unbiased estimates of $p_{Y_0}(y_0)$ and $p_{Y_0,Y_1}(y_0, y_1)$.

Now consider an HMM over t time steps, in the degenerate case where $X_0, X_1, ..., X_{t-1}$ are independent and all have the same marginal distribution μ, i.e.,

$$X_0 \sim \mu(\cdot), \qquad X_k | \left(X_{0:k-1} = x_{0:k-1}, Y_{0:k-1} = y_{0:k-1}\right) \quad \sim \mu(\cdot), \quad k \geq 1.$$

As per a standard HMM, assume that observations are distributed according to

$$Y_0 | X_0 = x_0 \quad \sim g(\cdot | x_0), \qquad Y_k | \left(Y_{0:k-1} = y_{0:k-1}, X_{0:k} = x_{0:k}\right) \quad \sim g(\cdot | x_k), \quad k \geq 1.$$

4. Prove that

$$\begin{aligned} &\mathbb{V}\left[\frac{\prod_{k=0}^{t} g\left(y_k | X_k\right)}{p_{Y_{0:t}}\left(y_{0:t}\right)}\right] - \mathbb{V}\left[\frac{\prod_{k=0}^{t-1} g\left(y_k | X_k\right)}{p_{Y_{0:t-1}}\left(y_{0:t-1}\right)}\right] \\ &= \frac{\mathbb{E}\left\{\left(\prod_{k=0}^{t-1} g\left(y_k | X_k\right)\right)^2\right\}}{p^2_{Y_{0:t-1}}\left(y_{0:t-1}\right)} \left\{\frac{\mathbb{E}\left(g^2\left(y_t | X_t\right)\right)}{p^2_{Y_t}\left(y_t\right)} - 1\right\} \geq 0. \end{aligned} \tag{1}$$

5. Assuming that there exists $c > 1$ such that

$$\inf_{k \geq 0} \frac{\int g(y_k | x_k)^2 \mu(x_k) dx_k}{\left(\int g(y_k | x_k) \mu(x_k) dx_k\right)^2} \geq c, \tag{2}$$

   then show that

$$\mathbb{V}\left[\frac{\prod_{k=0}^{t-1} g(y_k | X_k)}{p_{Y_{0:t-1}}(y_{0:t-1})}\right] \geq c^t - 1, \tag{3}$$

   where the variance is with respect to the joint distribution of $X_0, ..., X_{t-1}$.

6. Briefly discuss the practical implications of the bound in part (5) for the efficiency of using Monte Carlo (for example the estimates you suggested in part (3) to approximate $p_{Y_0,...,Y_t}(y_{0:t})$ when t is large.

Simulation question (Reversible jump MCMC)

Consider two models. For model 1 the toy target distribution is given by

$$\pi(\theta \mid k = 1) = \exp\left(-\frac{1}{2}\theta^2\right)$$

whereas for model 2 it is given by

$$\pi(\theta \mid k = 2) = \exp\left(-\frac{1}{2}(\theta_1^2 + \theta_2^2)\right).$$

We want to design a trans-dimensional sampler to sample from the distribution of (k, θ).

- Implement standard Metropolis-Hastings kernels K_1 for model 1 and K_2 for model 2. Check that they work before going further.

- Implement trans-dimensional moves to go from model 1 to model 2. That is, for $\theta \in \mathbb{R}$, propose an auxiliary variable $u \in \mathbb{R}$ following the distribution of your choice and a deterministic mapping $G_{1\to 2}(\theta, u)$ to obtain a point in $\mathbb{R}^2$ which you will then accept or reject with the appropriate acceptance probability.
- Implement trans-dimensional moves to go from model 2 to model 1. That is, for $\theta \in \mathbb{R}^2$, propose a deterministic mapping $G_{2\to 1}(\theta)$ to obtain a point in $\mathbb{R}$ which you will then accept or reject with the appropriate acceptance probability.
- Put these kernels together to obtain a valid Reversible Jump algorithm. What is the proportion of visits to each model? What should it be in the limit of the number of iterations?

Simulation question (Parallel tempering)

Consider the following bimodal target distribution:

$$\pi(x) \propto \exp\left(-10(x^2-1)^2\right).$$

Introduce a sequence $0 < \gamma_1 < \ldots < \gamma_N = 1$ and

$$\forall k \in \{1,\ldots,N\} \quad \pi_k(x) = \pi(x)^{\gamma_k} \propto \exp\left(-10\gamma_k(x^2-1)^2\right).$$

- Run a standard Metropolis-Hastings targeting each π_k, using the same proposal distribution for each k. Does the chain mix equally well for each γ_k?
- We propose to make the N Metropolis-Hastings interact by adding swap moves. At each iteration, draw uniformly $k, l \in \{1,\ldots,N\}$. Then propose to exchange X_k for X_l, where X is the current value of the joint Markov chain targeting

  $$\pi^{\gamma_1} \otimes \pi^{\gamma_2} \otimes \ldots \otimes \pi^{\gamma_N}.$$

  Accept this move with probability

  $$\min\left(1, \frac{\pi^{\gamma_k}(x_l)\pi^{\gamma_l}(x_k)}{\pi^{\gamma_k}(x_k)\pi^{\gamma_l}(x_l)}\right).$$

  The resulting algorithm is called parallel tempering. Show that the swap move leaves the joint target invariant.
- Run the algorithm and show that it improves the mixing of each chain, especially for high values of γ.

Simulation question (linear Gaussian model – SIS and SIR)

1. Consider a simple linear Gaussian model as described in Chapter 8:

   $$\forall t \geq 1 \quad X_t = \phi X_{t-1} + \sigma_V V_t,$$
   $$\forall t \geq 1 \quad Y_t = X_t + \sigma_W W_t,$$

   with $X_0 \sim \mathcal{N}(0,1)$, $V_t, W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $\phi = 0.95$, $\sigma_V = 1$, $\sigma_W = 1$. Simulate $T = 100$ observations from this model.
2. Using the prior model as a proposal, implement a sequential importance sampling strategy to estimate $\mathbb{E}[x_t \mid y_{1:t}]$ for all $t \in \{1,\ldots,T\}$.
3. By computing the variance of the estimators for a fixed N and all $t \in \{1,\ldots,T\}$, comment on the empirical performance of SIS as t increases.
4. Conversely, visualize the performance of SIS when t is fixed and N increases.
5. Either implement a Kalman filter or use an existing package, to compare the Monte Carlo estimators with the exact values of $\mathbb{E}[x_t \mid y_{1:t}]$.
6. Add a resampling step to turn SIS into a particle filter (SIR, sequential importance resampling), and estimate $\mathbb{E}[x_t \mid y_{1:t}]$ again.