Stochastic Simulation
Professor: Prof. Fabio Nobile
Assistant: Eliott Van Dieren

Week of 2025-11-24 to 2025-11-28

Lab 11 of Thursday 27th November 2025

Exercice 1.

In many applications of interest, it is not uncommon to encounter the need for sampling from a multi-modal distribution f. The theory developed so far is directly applicable to these types of distributions. However, in practice, sampling from these distributions using MCMC can be computationally challenging, as we will investigate in this problem. Throughout this exercise, we will consider the bi-modal distribution on $\mathbb{R}$

$$f(x;\gamma,x_0)=\frac{e^{-\gamma(x^2-x_0)^2}}{Z},\quad \gamma>0, \tag{1.1}$$

where Z is some normalizing constant. Depending on the values of γ and x_0, designing a sampling strategy to properly sample from (1.1) can become challenging.

To solve this exercise, we will use the random walk Metropolis (RWM) algorithm which is a Metropolis-Hastings algorithm that uses a proposal distribution of the form $q(x,y)=\mathcal{N}(y;x,\sigma^2)$, where

$$\mathcal{N}(y;x,\sigma^2):=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-x)^2}{2\sigma^2}\right)$$

is the Gaussian density with mean x and variance σ^2. The parameter σ is a tuning parameter that controls the size of the steps taken by the Markov chain.

Intuitively, if the two modes of (1.1) are too far apart, using a RWM might not work, if the step-size is too small since the chain might get stuck in one of the modes and not be able to explore the other mode with acceptable probability. Conversely, a RWM with very large *steps* might tend to reject quite often, thus rendering the whole sampling procedure inefficient. We begin by verifying this. Implement the RWM algorithm using as proposal distribution $q(x,y)=\mathcal{N}(y;x,\sigma^2)$ and target distribution $f(x;\gamma,x_0)$ for $\gamma=1$, $x_0=1,4,9,25$. Try different choices of σ. Discuss the quality of your samples by analyzing the trace-plots (one realization of the chain), autocorrelation functions and histograms of the chains obtained.

Exercice 2.

Ideally, we would like to obtain (approximately) i.i.d samples from a target distribution f using Markov Chain Monte Carlo (MCMC) algorithms. One practical way of doing so is via *sub-sampling* (also called *batch sampling*), which is implemented to reduce or eliminate correlation between the successive values in the Markov chain. That is, instead of considering the entire chain $\{X_n : n\geqslant 0\}$, say, this technique sub-samples the chain with a batch size $k>1$, so that only the values $\{X_{kn} : n\geqslant 0\}$ are considered. If the covariance $\mathrm{Cov}_f(X_0,X_n)$ vanishes as $n\to\infty$, then the idea of sub-sampling is quite natural since X_{kn} and $X_{k(n+1)}$ can be considered to be

approximately independent for k sufficiently big; estimating such a k may be difficult in practice though. While sub-sampling provides a way of generating (approx.) i.i.d. samples from f and may thus be useful to assess the convergence of a MCMC method, it necessarily leads to an efficiency loss. Let $\{X_n \in \mathbb{R}^d \colon n \geqslant 0\}$ be a Markov chain with a unique stationary distribution f, and $X_0 \sim f$ (i.e., the chain is at equilibrium). Take $\phi \colon \mathbb{R}^d \to \mathbb{R}$ such that $\mathbb{E}_f(|\phi|^2) < \infty$ and consider two estimators for $\mu = \mathbb{E}_f(\phi)$, namely one that uses the entire Markov chain ($\widehat{\mu}$) and one based on sub-sampling ($\widehat{\mu}_k$) using only every k-th value:

$$\widehat{\mu} = \frac{1}{Nk} \sum_{n=1}^{Nk} \phi(X_n) \ , \quad \text{and} \quad \widehat{\mu}_k = \frac{1}{N} \sum_{n=1}^{N} \phi(X_{nk}) \ .$$

Show that the variance of $\widehat{\mu}$ satisfies $\mathrm{Var}_f(\widehat{\mu}) \leqslant \mathrm{Var}_f(\widehat{\mu}_k)$ for every $k > 1$.

Exercice 3.

Let $X \subset \mathbb{R}^d$ and $P_i : X \times \mathcal{B}(X) \to [0,1]$, $i = 1 \dots, m$ be Markov transition kernels on X with $\mathcal{B}(X)$ the associated $\sigma-$algebra.

(a) Given $a_1, \dots, a_m \in \mathbb{R}^+$, such that $\sum_{i=1}^m a_i = 1$, show that $P(x, A) = \sum_{i=1}^m a_i P_i(x, A)$ is a Markov transition kernel.

(b) Suppose that a measure $\pi : \mathcal{B}(X) \to [0,1]$ is invariant for each kernel P_i. Show that it is also invariant for $P = \sum_{i=1}^m a_i P_i$, where $a_1, \dots, a_m \in \mathbb{R}^+$, such that $\sum_{i=1}^m a_i = 1$. If each P_i is reversible, is P reversible?

(c) Under the same assumptions of point (b), define the Markov operator $\mathcal{P}_i$ associated to P_i (i.e., $\pi\mathcal{P}_i(A) = \int P(x, A) d\pi(x), \forall A \in \mathcal{B}(X)$). Then, show that π is also invariant for $\mathcal{P} = \mathcal{P}_{i_1} \circ \dots \circ \mathcal{P}_{i_k}$, for any choice of $i_1, \dots, i_k$. If each P_i is reversible, for which choice of $i_1, \dots, i_k$ is $\mathcal{P}$ reversible?