Monte Carlo Inference, Example Sheet 2

The associated example class will be held on **Friday 22 April, 9-10am**.

Exercise 1. Devise and implement a simulation experiment to approximate the probability $P(Z \in (0,\infty)^6)$ when $Z \sim N_6(0,\Sigma)$ and

$$\Sigma^{-1/2} = diag(0,1,2,3,4,5) + e.e^t,$$

with $e^t = (1,1,1,1,1,1)$, when using a distribution restricted to $(0,\infty)^6$ and importance sampling.

Exercise 2. Evaluate $\int_0^1 e^{-x^2}$ using stratified sampling with the programming language of your choice. You may want to try different number of strata. What is the magnitude of variance reduction compare to simple Monte Carlo integration?

Exercise 3. Consider a building that is located in a region where the strongest wind W in 50 years has an extreme value distribution:

$$P(W \le w) = \exp(-\exp((w-52)/4)), \quad 0 < w < \infty.$$

Suppose that the load (pressure) this wind puts on the building is $L = CW^2$, where the value of C depends on some properties of the building, the density of air, direction of the wind relative to the building geometry, positions of nearby buildings and other factors. By using Monte Carlo find the estimated 99th percentile of L and its 99% confidence interval.

Exercise 4. Show that if $f(x) \le Mg(x)$ for all $x \in supp(f)$ then the expected acceptance probability associated with the Metropolis Hastings algorithm is at least $1/M$ when the chain in stationary.

Exercise 5. (i) Consider a probability density g on $[0,1]$ and a function $0 < \rho < 1$ such that $\int_0^1 g(x)/(1-\rho(x))dx < \infty$. Prove that the Markov chain with transition kernel

$$K(x,x') = \rho(x)\delta_x(x') + (1-\rho(x))g(x'),$$

where δ_x is the Dirac mass at x, has stationary distribution $f(x) \propto g(x)/(1-\rho(x))$.

(ii) Provide the algorithm for generating the Markov chain associated with (i), and discuss the similarity with the accept-reject algorithm.

(iii) Implement the algorithm when g is the density of $Beta(\alpha+1,1)$ distribution and $\rho(x) = 1-x$. Give the expression of the stationary distribution f.

Exercise 6 (Gibbs as composition of MH steps)**.** Show that the transition kernel of the Gibbs sampler (for a density π on $\mathbb{R}^p$) is the composition of p Metropolis-Hastings transition kernels, each of which have acceptance probability 1.

Exercise 7 (Reversible Gibbs sampling)**.** Consider the following version of the Gibbs sampling algorithm, called reversible Gibbs, where in step II, and for $t = 0,\dots,T$ it has:
1) $Y_1 \sim \pi_1(\cdot|x_2^{(t)},\dots,x_p^{(t)})$
$\vdots$
$p-1)$ $Y_{p-1} \sim \pi_{p-1}(\cdot|Y_1,\dots,Y_{p-2},x_p^{(t)})$
$p)$ $X_p^{(t+1)} \sim \pi_p(\cdot|Y_1,\dots,Y_{p-1})$
$p+1)$ $X_{p-1}^{(t+1)} \sim \pi_{p-1}(\cdot|Y_1,\dots,Y_{p-2},x_p^{(t+1)})$
$\vdots$
$2p-1)$ $X_1^{(t+1)} \sim \pi_1(\cdot|x_2^{(t+1)},\dots,x_p^{(t+1)})$

Prove that the reversible Gibbs sampling algorithm induces a Markov chain that satisfies detailed balance (and hence has a stationary distribution π).

Exercise 8 (Gaussian variables on a lattice, with Ising prior)**.** Let $D = \{1,\ldots,K\}^2, K \geq 1$, and let $S \in \{-1,+1\}^D$ be random, $(S)_i = s_i, i \in D$, with distribution

$$\pi(S) \propto \exp\left\{-J \sum_{\{i,j\}\in\mathcal{N}} s_i s_j - H \sum_{i\in D} s_i\right\},$$

where $\mathcal{N}$ is the neighbour relationship set on D, i.e. it is the subset of the set of unordered pairs $\{\{i,j\} : i,j \in D, i \neq j\}$ containing $\{i,j\}$ if and only if i and j are vertical or horizontal neighbours.

Let $X = (X_i)_{i\in D}$ be random, with

$$X_i|S \overset{\text{i.i.d.}}{\sim} N(s_i, 1), \quad i \in D.$$

Given a realization of X, write the posterior distribution of $S|X$, compute the full conditionals $s_i|s_{-i}, X$ and give the Gibbs sampling algorithm for generating a Markov with stationary distribution $\pi(S|X)$. What can you say about irreducibility of the resulting Markov chain?

Exercise 9. Suppose that

$$\begin{aligned} Y_1,\ldots,Y_n|\mu,\tau &\overset{\text{i.i.d.}}{\sim} N(\mu,\tau^{-1}) \\ \mu &\sim N(0,1) \\ \tau &\sim \text{Gamma}(2,1), \end{aligned}$$

where $\tau > 0$, μ and τ are independent. Assuming that we have observed a realization $y_1,\ldots,y_n$ of $Y_1,\ldots,Y_n$, give the Gibbs sampler for simulating a Markov chain with stationary distribution $\pi(\mu,\tau|y_1,\ldots,y_n)$. (Recall that the density of the Gamma(α,β) is $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\exp(-\beta x)$, $x,\alpha,\beta > 0$).

Exercise 10. The beta model defines a distribution over a network whose nodes are labeled by $\{1,\ldots,n\}$ by assuming that the occurrence of edges is independent, and the probability p_{ij} of observing an edge between nodes i and j is parameterized as follows:

$$p_{i,j} = P(X_{ij} = 1) = \frac{e^{\beta_i+\beta_j}}{1+e^{\beta_i+\beta_j}}, \quad \forall i \neq j.$$

β_i can be interpreted as unknown constants that determine the propensity of node i to have edges.

(i) Show that the probability of observing a network $x = (x_{ij})_{1\leq i<j\leq n}$ can be written in exponential family form with sufficient statistics $d(x) = (d_1(x),\ldots,d_n(x))$, where d is the degree sequence of x, i.e., $d_i(x)$ is the number of edges adjacent to i in x.

(ii) We want to study the well-known network of friendship among monks, available in the package ERGM in `R` under the name "sampson" (to be loaded by `data(sampson)` and to be plotted by `plot(samplike)`). Explain how one can use MCMC methods to estimate the parameters for the beta model for this network data.

(iii) Use the command `ergm` to conduct the parameter estimation (with userterm `nodefactor(base=0)` for the degree sequence), and use the command `mcmc.diagnostics` to test the convergence of the chain. Discuss the results.