Université Paris Dauphine
Département MIDO

Master 2 MASH – Computational methods and MCMC

Practical n°2

stoehr@ceremade.dauphine.fr

Exercise 1 *(Gibbs Sampler).*

The Gibbs sampler, first introduced by Geman and Geman [1984] and generalised by Gelfand and Smith [1990], is an essential element in Markov chain Monte Carlo methods [Robert and Casella, 2004]. It produces a Markov chain associated with a given target joint distribution, denoted π, by alternatively sampling from each of its conditionals.

The Ising model is a well known model used in various applications such as electromagnetism or image processing. It corresponds to a discrete random process $\mathbf{X}$ defined on an undirected graph $\mathscr{G}$ which induces a topology on a set of sites $\mathscr{S} = \{1, \dots, n\}$ and taking values in $\mathscr{X} = \{-1, 1\}^n$.

In this exercise, we look at the model defined on a regular grid $h \times w$ for the four closest neighbours graph. $\mathbf{X} = (X_{i,j})$ can be represented by a matrix with h rows and w columns and the density writes as

$$f(\mathbf{x} \,|\, \theta) \propto \exp\left[\theta \left(\sum_{j=1}^{w} \sum_{i=1}^{h-1} x_{i,j} x_{i+1,j} + \sum_{j=1}^{w-1} \sum_{i=1}^{h} x_{i,j} x_{i,j+1} \right)\right].$$

1. Compute $\mathbb{P}\left[X_{i,j} = x_{i,j} \,\middle|\, X_{-i,-j} = x_{-i,-j}, \theta\right]$, $i \in [\![1, h]\!]$, $j \in [\![1, w]\!]$.
2. Write a function `gibbs_sampling(n_iter, theta)` which samples from the Ising model with parameter `theta` using the Gibbs sampler. The sampler can be initialised using $\mathbf{x} = (1, \dots, 1)$.
3. Sample a realisation from an Ising model defined on a 30×30 grid with parameter $\theta = 1$ and $\theta = 0.4$ (`n_iter` = 100 iterations). What do you observe?

Exercise 2 *(ABC).*

ABC algorithm is a computational method which stemmed from population genetics to deal with intractable likelihoods, that is models whose likelihood cannot be easily computed, or cannot be computed at all,but which can be simulated from [Tavaré et al., 1997, Beaumont et al., 2002]. Consider a likelihood $f(\cdot \mid \theta)$, $\theta \in \Theta$, and a prior distribution $\pi(\cdot)$ on Θ. Given an observed data set x^{obs}, the principle of the method to get a sample (approximately) from $\pi(x^{\text{obs}} \mid \theta) \propto f(x^{\text{obs}} \mid \theta)\pi(\theta)$ is

(a) to simulate pairs of parameters and pseudo-data from the prior predictive, *i.e.* $\theta \sim \pi(\cdot)$ and $x \sim f(\cdot \mid \theta)$,

(b) to keep only the parameters that bring the pseudo-data x close enough to the observed data x^{obs}, *i.e.*, given $\varepsilon > 0$, a distance d and a projection S of the data, called summary statistic, keep $\{\theta\,;\, d(S(x^{\text{obs}}), S(x)) < \varepsilon\}$.

Let consider the Hardy Weinberg model introduced in the previous practicle with the three different possible configurations of the haptoglobine: AA, aa, aA. Given $\theta \in]0,1[$ the frequence of allele a, we have

$$\mathbb{P}[X = \mathrm{AA}] = (1-\theta)^2, \quad \mathbb{P}[X = \mathrm{aa}] = \theta^2, \quad \text{and} \quad \mathbb{P}[X = \mathrm{aA}] = 2\theta(1-\theta).$$

We use the observation from Plato, *et al.* (1964) on a sample of $n = 190$ people:

Genotype	AA	aa	aA
Count	10	112	68

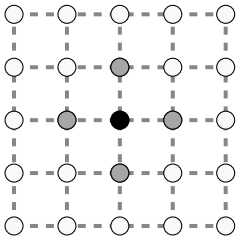
1. **(a)** Asumming that the prior distribution on θ is uniform on $[0,1]$, sample from the posterior distribution of θ using ABC.

(b) Compare your result with the theoretical posterior distribution.

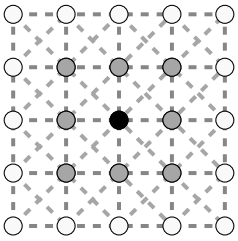
(c) Compare the empirical posterior mean you got with Metropolis-Hastings and with ABC.

2. Asumming that the prior distribution on θ is a Beta distribution, sample from the posterior distribution of θ using ABC.

Exercise 3 *(ABC model choice).* We are interested in the Ising model for two adjacency structure represented below: the four closest neighbours graph (a) and the eight closest neighbours graph (b) – neighbours of the vertex in black are represented by vertices in gray.



(a)



(b)

Given a parameter $\beta \in \mathbb{R}$, the joint distribution of the model is given by

$$f(\mathbf{x} \mid \beta) = \frac{1}{Z(\beta)} \exp\left(\beta \sum_{i \sim j} x_i x_j\right),$$

where $i \sim j$ means that $(i,j) \in \mathscr{S}$ are neighbors (*i.e.*, linked by an edge) in $\mathscr{G}$.

A question of interest when dealing with the Ising model is to select the dependency structure $\mathscr{G}$ that best fits an observed dataset $\mathbf{x}^{\mathrm{obs}}$, *i.e.*, we aim at finding the maximum *a posteriori*

$$\widehat{m} = \underset{m \in \mathscr{M}}{\arg\max}\, \pi(m \mid \mathbf{x}^{\mathrm{obs}}) \propto \pi(m) \int_{\Theta_m} f_m(\mathbf{x}^{\mathrm{obs}} \mid \theta_m) \pi_m(\theta_m) \mathrm{d}\theta_m,$$

$\mathscr{M} = \{m : 1, \ldots, M\}$ is a set of model, $\pi(\cdot)$ a prior on the model space $\mathscr{M}$, π_m a prior on parameter space Θ_m associated to model m and $f_m(\cdot \mid \theta_m)$ the likelihood of model m.

To approximate $\widehat{m}$, ABC model choice starts by simulating numerous triplets $(m, \beta_m, \mathbf{x})$ from the joint Bayesian model. Afterwards, it approximates the posterior probabilities by the frequency of each model number associated with simulated $\mathbf{x}$'s in a neighbourhood of $\mathbf{x}^{\text{obs}}$, that is defined as simulations whose distances to the observation measured in terms of summary statistics, *i.e.* $d\left(S(\mathbf{x}), S(\mathbf{x}^{\text{obs}})\right)$, fall below a threshold ϵ.

In this example, we set the prior on model index $\pi(\cdot)$ to the uniform distribution on $\{1,2\}$, the prior on parameter space for model $m = 1$ to the uniform distribution on $[0,1]$ and the one for model $m = 2$ to the uniform distribution on $[0,0.5]$. The distance d is chosen to be the L^2-standardised distance. Finally we use the vector of summary statistics composed of the summary statistic $S(\mathbf{x}) = \sum_{i\sim j} x_i x_j$ under each model which is sufficient for the model choice problem Grelaud et al. [2009], and that can be computed using the function `sum_stat(x)` (*c.f.*, source_Ising.R)

1. Write a function `abc_ref_table(n, m, h, w)` which draws `n` particles from the joint Bayesian model using `m` iterations of the Gibbs sampler and returns a table containing model indices, parameter values and summaries of the generated data.

2. We set to 100 the number of iterations of the Gibbs sampler.

(a) Draw a pseudo-observation $\mathbf{x}^{\text{obs}}$ from an Ising model defined on a 20×20 grid with a four closest neighborhood structure at $\beta = 0.5$.

(b) Perform ABC model choice for the pseudo-observation $\mathbf{x}^{\text{obs}}$ by keeping the 100 closest simulations with respect to the L^2-standardised distance on a reference table of size 10000.

References

M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian Computation in Population Genetics. *Genetics*, 162(4):2025–2035, 2002.

A. E. Gelfand and A. F. M. Smith. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

A. Grelaud, C. P. Robert, J.-M. Marin, F. Rodolphe, and J.-F. Taly. ABC likelihood-free methods for model choice in Gibbs random fields. *Bayesian Analysis*, 4(2):317–336, 2009.

C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics, second edition edition, 2004.

S. Tavaré, D. J. Balding, R. C. Griffiths, and P. Donnelly. Inferring Coalescence Times From DNA Sequence Data. *Genetics*, 145(2):505–518, 1997.