How-to: Multivariate MCMC

Maxime Garnier
ENS Cachan, June 30^{th} 2014

Introduction

A *Markov Chain Monte-Carlo* (MCMC) algorithm is a powerful tool to explore a given space of parameters. The aim of this "tutorial" is to give a ready-to-use recipe for rapidly writing such an algorithm.

MCMC algorithms are useful in many situations but it is especially recommended to use them for solving optimisation problems such as the minimisation of a function.

Basic MCMC algorithm: Metropolis

Consider a m-dimensional parameter space and a vector $\vec{\theta}$ of this space.

The first idea is to randomly sample the vectors of this space. Let k denote the k^{th} iteration and i the i^{th} component of the $\vec{\theta}$ vector. The sampling then reads, using a multidimensional Central-Limit Theorem argument:

$$\theta_{k+1}^{(i)} = \theta_k^{(i)} + \delta\theta_i \mathcal{N}(0,1).$$

where $\mathcal{N}(0,1)$ denotes the normal (Gaussian) law of mean 0 and variance 1.

Then one has to consider the *likelihood* ratio $\gamma = \frac{\mathcal{L}_{k+1}}{\mathcal{L}_k}$ [1] to choose whether to accept the move.

The conditions are as follows:

- If $\gamma \geq 1$: Accept the move.
- If $\gamma < 1$: Choose a random number r between 0 and 1.
  - If $r < \gamma$: accept the move.
  - Else: reject it.

This is the simplest case of MCMC algorithm.

The *best-fit parameters* may be determined by taking the mean value (or the median) of all visited values. But **be careful**, there is is a so-called *burn-in* period where the algorithm really searches its way through the parameter space and it must not be taken into account in the determination of the *best-fit parameters*.

The acceptance rate η is defined as:

$$\eta = \frac{\#\text{ accepted moves}}{\#\text{ total moves}}$$

[1] We consider the case of the minimisation of the χ^2 with Gaussian errors in a fitting problem, $\mathcal{L} = \exp(-\chi^2/2)$.

Roberts, Gelman and Gilks [2] have proved that provided some assumptions the ideal acceptance rate is $\eta = \mathbf{0.234}$. The hypotheses required by the Roberts, Gelman and Gilks theorem concern the proposal distribution f (we chose a Gaussian):

- Independance: $F(\vec{\theta}) = \prod_{i=1}^{m} f(\theta_i)$.
- Regularity: $f \in \mathcal{C}^2(\mathbb{R})$.
- Generalisation to the case: $f = f_i$.

It is easily understandable that the difficulty lies in the choice of the $\delta\theta_i$ steps values. Too small $\delta\theta_i$ would cause the algorithm to explore only a small zone of the parameter space whereas too high values would lead to a inaccurate exploration.

Parameters correlations

In real-life cases one often encounters correlated parameters. These correlations have a disastrous effect on the algorithm presented above. Indeed, the exploration of the space is no longer efficient. At least two approaches may be chosen.

- Variance-Covariance matrix approach.
- Preliminary orthogonalisation of the parameter space.

Variance-Covariance matrix

The variance-covariance matrix V is defined as:

$$\begin{aligned} V_{i,j} &= \mathrm{cov}(\theta_i, \theta_j),\ \forall (i,j) \in [\![0; m]\!]^2 \\ V_{i,i} &= \mathrm{var}(\theta_i) \end{aligned}$$

with:

$$\begin{aligned} \mathrm{cov}(\theta_i, \theta_j) &= \sum_{k=1}^{N} \frac{(\theta_k^{(i)} - \overline{\theta^{(i)}})(\theta_k^{(j)} - \overline{\theta^{(j)}})}{\sigma_k^2} \\ \mathrm{var}(\theta_i) &= \sum_{k=1}^{N} \frac{(\theta_k^{(i)} - \overline{\theta^{(i)}})^2}{\sigma_k^2} \end{aligned}$$

This matrix can be built after a full preliminary MCMC run and its diagonalisation leads to the construction of a space of uncorrelated parameters. It is merely a change-of-basis procedure.

However, a full MCMC run takes a lot of time, especially if the parameter space is complicated.

[2] Roberts, J. O., Gelman A and Gilks, W. R. *Weak Convergence And Optimal Scaling of Random Walk Metropolis Algorithms*, The Annals of Applied Probability, 1997

Preliminary orthogonalisation

To avoid losing time, one can think of orthogonalising the parameter space **before** running the MCMC algorithm. This can be done *via* a Gram-Schmidt procedure as proposed by N. Kains in his PhD thesis (2010) for gravitational microlensing light-curve fitting.

Consider a model μ depending on time and on the parameters θ_i [3] and consider the following vectors:

$$P_i = {}^t\left(\frac{\partial \mu}{\partial \theta_i}(t_j)\right)_{j=1\cdots N_{\text{mes}}}$$

The general Gram-Schmidt procedure taims at creating an orthonormalised basis $(\mathbf{e}_k)$ from an initial basis $(\mathbf{v}_k)$:

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} P_{\mathbf{u}_j}(\mathbf{v}_k)$$

and

$$\mathbf{e}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$$

with:

$$P_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$$

the projection operator of the vector $\mathbf{v}$ on the vector $\mathbf{u}$.

with the dot product:

$$A \cdot B = \sum_{i=1}^{N_{\text{mes}}} \frac{A(t_i)B(t_i)}{\sigma_i^2}$$

Applying this procedure to the P_i vectors, one is able to compute the corresponding B_i vectors of the uncorrelated parameter space:

$$B_i = {}^t\left(\frac{\partial \mu}{\partial \beta_i}(t_j)\right)_{j=1\cdots N_{\text{mes}}}$$

Where the β_i parameters are the uncorrelated parameters one sought.

The change-of-basis is done *via* the matrices T and S:

$$T_{ij} = \begin{cases} \frac{P_i \cdot B_j}{B_j \cdot B_j} & i \geq j \\ 0 & i < j \end{cases}$$

$$\vec{\beta} = S\vec{\theta},\ S = T^{-1}$$

and

$$B_i = \sum_{j=1}^{i} S_{ij} P_j$$

[3] This may be applied to other model such as spectra.