CentraleSupélec

Statistique bayésienne : méthodes de Monte-Carlo

Travaux pratiques

CentraleSupelec, Option Mathématiques Appliquées

Département Signaux et Statistique
v. 2019/01 · Emmanuel Vazquez

L'objectif principal de ces travaux pratiques est de mettre en œuvre un algorithme de Metropolis-Hastings et de l'utiliser sur un problème d'estimation. Le langage de programmation est libre.

1 Générateur pseudo-aléatoire

- *Choisissez une mise en œuvre de générateur pseudo-aléatoire dans votre langage de programmation afin de générer une suite de nombres pseudo-aléatoires selon une loi uniforme $\mathcal{U}([0,1])$. Quel est le générateur pseudo-aléatoire utilisé ?*
- *Trouver et mettre en œuvre une méthode pour générer deux fois la même suite de nombres pseudo-aléatoires selon une loi uniforme $\mathcal{U}([0,1])$.*

2 Méthode d'acceptation-rejet

- *Mettre en œuvre un algorithme de type acceptation-rejet pour simuler des variables aléatoires de densité*
$$f(x) \;\propto\; \exp\left(-1/2(x-1)^2\right)|\sin(\pi x)|\,, \tag{1}$$
où $\propto$ indique la proportionalité (écrire une fonction générant un vecteur aléatoire de longueur n prescrite).

3 Algorithme de Metropolis-Hastings

3.1 Simulation sur $\mathbb{R}$

- *Écrire une fonction nommée* `mhsim` *de sorte que le script présenté dans le tableau 1 (écrit en langage Matlab, pour référence) produise une réalisation d'une chaîne de Markov partant de $x_0 = 0$, de longueur $n_{\text{samples}} = 5 \cdot 10^4$, avec un noyau de transition de MH ayant la loi du $\chi^2(3)$ comme loi invariante, et construit à partir d'une densité instrumentale*
$$q(y \mid x) = \frac{1}{s}\,\varphi\left(\frac{y-x}{s}\right)\,,$$
où φ est la densité de la loi $\mathcal{N}(0,1)$. La fonction `mhsim` *renvoie les réalisations simulées ainsi qu'un taux d'acceptation moyen.*
- *Discuter des propriétés de convergence de la chaîne en fonction de la valeur du paramètre s.*
- *Proposer et mettre en œuvre une méthode permettant de régler automatiquement le taux d'acceptation.*
- *Proposer et mettre en œuvre une méthode pour diagnostiquer automatiquement un éventuel problème de convergence à partir de la simulation de plusieurs chaînes de Markov partant d'états initiaux différents.*

3.2 Un problème d'estimation

3.3 Problème

- Système à compartiments (système dynamique à temps continu d'après Walter and Pronzato (1997))
- Vecteur d'état $q = (q_1,\, q_2)^\mathsf{T} \to$ quantités de matière dans deux compartiments échangeant de la matière

```
%%% SCRIPT MATLAB METROPOLIS-HASTINGS
% --- INITIALISATION ---
clear all, close all
s = 1e-3;
x0 = 0.0;
nsamples = 50000;
logpdf      = @(x)    log(chi2pdf(x,3));
logproppdf  = @(x, y) log(normpdf(y, x, s));
proprnd     = @(x)    x + s*randn;
% --- SIMULATION MH ---
[x, taux_acc] = mhsim(x0, nsamples, logpdf, logproppdf, proprnd);
% --- ANALYSE DES RESULTATS ---
% ... à effectuer/compléter
```

TABLEAU 1 – Script runmh.m

— Équations d'évolution :

$$\begin{cases} \dot{q}_1 &= -(x_1 + x_3)q_1 + x_2 q_2, \\ \dot{q}_2 &= x_1 q_1 - x_2 q_2. \end{cases} \tag{2}$$

Formellement, q est une fonction $(x, t) \in \mathbb{R}^3 \times \mathbb{R} \mapsto q(x, t) \in \mathbb{R}^2$.

— À $t = 0$, injection de matière dans le compartiment (1), de sorte que pour x quelconque, $q(x, 0) = (1, 0)^\mathsf{T}$

— Mesures bruitées

$$y_i \sim \mathcal{N}\big(q_2(x, t_i), s^2\big), \quad s > 0,$$

de la quantité de matière dans le compartiment (2) aux instants $t_i = 1, 2, 3, ..., 15$.

— Objectif : estimer le vecteur de paramètres x à partir des observations, ainsi que le maximum de la quantité de matière dans le compartiment (2).

3.4 Travail

- *Construire le simulateur*

$$(x, t) \mapsto (q_2(x, t_1),\ q_2(x, t_2), \ldots,\ q_2(x, t_{15}))$$

- *Construire un vecteur d'observations $y = (y_1, \ldots, y_{15})^\mathsf{T}$ pour le paramètre*

$$x^\star = (0.6, 0.35, 0.15)^\mathsf{T}.$$

- *Représenter graphiquement la loi a posteriori de x en choisissant comme loi a priori la loi uniforme sur $[0, 1]^3$. Influence de la variance du bruit d'observation ?*
- *Utiliser l'algorithme de Metropolis-Hastings pour simuler des réalisations de x distribuées selon la densité a posteriori.*
- *Utiliser la chaîne générée pour obtenir une estimation du maximum de la quantité de matière dans le compartiment (2).*

Références

E. Walter and L. Pronzato. *Identification of parametric models*. Springer Verlag New-York, 1997.