

## Cours 5

## 5 Méthodes algorithmiques

Le calcul effectif des lois a posteriori peut s'avérer extrêmement difficile. En particulier, la prédictive nécessite des calculs d'intégrales parfois multiples qui peuvent être d'un lourdeur rédhibitoire ! On notera également qu'on s'intéresse très souvent à l'espérance mathématique de la loi a posteriori. On se doit donc de calculer des expressions de la forme :

$$\int_{\Theta} \theta \cdot \pi(\theta|x) d\theta.$$

Pour approximer cette quantité, une première idée consiste à simuler des valeurs de  $\theta$ ,  $(\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(M)})$  suivant la loi  $\pi(\theta|x)$  et d'approximer  $E^{\pi(\cdot|x)}(\theta)$  par :

$$\bar{\theta}_M = \frac{1}{M} \sum_{i=1}^M \theta^{(i)}.$$

On sait (loi faible des grands nombres) que  $\bar{\theta}_M$  converge en probabilité vers  $\theta$ . Sous certaines conditions (sans biais et variance asymptotique nulle), on a une convergence en moyenne quadratique.

Ce type de méthode appartient à la famille des *méthodes de Monte Carlo* qui regroupe toutes les méthodes qui visent à calculer une valeur numérique en utilisant des procédés aléatoires, des procédés de simulation probabiliste. La dénomination provient de ce que les roulettes de casino permettent d'obtenir d'excellentes suites de nombres aléatoires. Ces méthodes sont utilisées pour le calcul d'intégrales complexes mais également en optimisation.

Nous allons dans un premier temps, nous intéresser aux méthodes de simulations classiques avant d'aborder les méthodes de Monte Carlo par Chaînes de Markov (MCMC) dont on peut donner la définition suivante (Robert p. 126) :

**Définition 13** *On appelle Algorithme MCMC (pour Markov Chain Monte Carlo), toute méthode produisant une chaîne de Markov ergodique de loi stationnaire la distribution d'intérêt.*

Dans le cas d'une loi a posteriori, on fabrique une chaîne de Markov dont les états sont les valeurs du paramètre et qui lorsqu'elle entre en régime stationnaire fournit (produit) des réalisations de cette loi.

### 5.1 Méthode d'inversion

Elle s'appuie sur le résultat suivant :

**Proposition 7** – *Soit  $F$  une fonction de répartition. Soit  $U$  une variable aléatoire uniforme sur  $[0, 1]$ .*

*Alors :  $F^{-1}(U)$  est une variable aléatoire de fonction de répartition  $F$ .*

La preuve de cette proposition est directe puisque  $P(U \leq u) = u$ , on a :

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x),$$

ce qui montre que la variable aléatoire  $F^{-1}(U)$  a pour fonction de répartition  $F$ . Ainsi, pour générer des réalisations d'une variable aléatoire de loi  $F$ , il suffit de générer des réalisations d'une loi uniforme id est de tirer au hasard des nombres entre 0 et 1 et de calculer l'image par la fonction réciproque de  $F$  de ces nombres.

**Exemple :** voir TD 3, exercice 1.

## 5.2 Méthode d'Acceptation-Rejet

Supposons une densité  $f$  à simuler. Le principe de la méthode consiste à imiter  $f$  par une densité  $g$  facilement simulable. Soit  $g$  cette densité que l'on sait simuler. Supposons  $f(x) \leq Mg(x)$ ,  $M > 1$ .

*La procédure*

1. Générer  $x \sim g$ ,  $u \sim \mathcal{U}[0, 1]$ ;
2. Si  $u \leq f(x)/Mg(x)$ , accepter  $y = x$  ;  
sinon retourner en 1.

*fournit des variables aléatoires distribuées suivant  $f$ .*

On vérifie que cette procédure génère effectivement une réalisation d'une réalisation de loi  $f$  en écrivant la fonction de répartition de la variable aléatoire  $Y$  associée à la valeur acceptée. On remarque tout d'abord qu'en moyenne, la probabilité de rejeter un  $x$  suivant  $g$  sera égale à :

$$\int P\left(U > \frac{f(x)}{Mg(x)}\right) g(x) dx = \int \left(1 - \frac{f(x)}{Mg(x)}\right) g(x) dx = 1 - \frac{1}{M}$$

La valeur de  $x$  peut être accepté dès le premier tour ou à l'extrême, au bout d'une infinité de tours, on peut donc exprimer la densité de  $Y$  au point  $x$  par :

$$g(x) \sum_{i=0}^{+\infty} \left(1 - \frac{1}{M}\right)^i P\left(U \leq \frac{f(x)}{Mg(x)}\right)$$

qui, tout calcul fait, vaut  $f(x)$ .

Ce type algorithme est particulièrement intéressant en analyse bayésienne. En effet, il s'applique lorsqu'on a connaissance de la densité d'intérêt à une constante près. On pourra donc s'affranchir dans le cas d'une loi a posteriori du calcul de la predictive en cherchant une fonction  $g$  imitant  $f(x|\theta)\pi(\theta)$ .

### 5.3 L'algorithme d'augmentation des données

D'une manière générale, l'algorithme d'augmentation des données (algorithme TW pour "de Tanner et Wong") a pour objet le calcul d'une loi conditionnelle. On peut donc l'utiliser en particulier pour calculer la loi a posteriori. L'idée est d'obtenir une approximation de  $\pi(\theta|x)$  par *augmentation des données (data augmentation)* en décomposant la loi d'intérêt supposé «compliquée» en un mélange de lois plus «faciles». Pour ce faire, on essayera de mettre la loi sous la forme :

$$\pi(\theta|x) = \int_Z \pi(\theta|z, x) \pi(z|x) dz \quad (5)$$

où on a introduit une variable  $z$  *variables latentes* telle que les lois  $\pi(\theta|x; z)$  et  $\pi(z|x)$  sont accessibles, connues.

De sorte que, pour simuler suivant  $\pi(\theta | x)$ , on pourra choisir un composant de ce mélange suivant  $\pi(z | x)$ , puis simuler suivant  $\pi(\theta|x; z)$ .

En pratique, on approxime l'intégrale (5) par :

$$\frac{1}{m} \sum_{j=1}^m \pi(\theta|x; z_j)$$

où les  $z_j$  sont des réalisations de  $\pi(z | x)$ .

**Exemple** – Supposons que l'on observe  $x = (x_1, x_2, x_3, x_4) = (125, 18, 20, 34)$ , la répartition en quatre classes d'un groupe d'animaux de taille  $n = 197$ . Cette répartition suit une loi multinomiale pour laquelle on se donne les valeurs de paramètres suivantes :

$$\left( \frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4} \right), \quad \theta \in [0, 1].$$

Si l'on se donne une loi uniforme sur  $\mathbb{R}^+$  pour loi a priori sur  $\theta$ , la loi a posteriori est :

$$\pi(\theta | x) \propto \left( \frac{1}{2} + \frac{\theta}{4} \right)^{x_1} \left( \frac{1-\theta}{4} \right)^{x_2} \left( \frac{1-\theta}{4} \right)^{x_3} \left( \frac{\theta}{4} \right)^{x_4}$$

Augmentons les données en divisant la première cellule par 2 cellules de probabilités  $1/2$  et  $\theta/4$ . Posons  $z_1 + z_2 = x_1$ ,  $z_2 = x_3$ ,  $z_3 = x_4$  et  $z_4 = x_5$ . Ces données augmentées  $z = (z_1, z_2, z_3, z_4, z_5)$  suivent une loi multinomiale de paramètres :

$$\left( \frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4} \right),$$

et tout calcul fait la loi a posteriori prend une forme de loi Bêta :

$$\pi(\theta | x, z) \propto \theta^{z_2+z_5} (1-\theta)^{z_3+z_4}.$$

La loi de  $z$  conditionnellement à  $x$  est obtenue par le rapport des multinomiales qui est, en fait, la loi de  $(z_1, z_2)$  conditionnellement à  $x$  et qui revient à considérer une loi binomiale de paramètres

$(x_1, \theta/(2 + \theta))$  pour  $z_2$ .

L'expression (5) se met donc ici sous la forme :

$$\pi(\theta | x) = \sum_{z_2=0}^{125} \pi(\theta | z_2, x) \pi(z_2 | x) dz \quad (6)$$

où  $\pi(z_2 | x)$  est une loi binomiale de paramètres  $(x_1, \theta/(2 + \theta))$  et  $\pi(\theta | z_2, x)$  est une loi bêta de paramètres  $(z_2 + x_4 + 1, x_2 + x_3 + 1)$ .

La formule (6) permet donc de générer une suite de  $\{\theta^{(p)}, p \in \mathbb{N}\}$  à partir d'une valeur initiale  $\theta^{(0)}$ . En effet, on peut tirer un  $m$ -échantillon  $(z_2^{(1)}, \dots, z_2^{(m)})$  suivant une loi binomiale de paramètres  $(x_1, \theta^{(p)}/(2 + \theta^{(p)}))$  qui permet d'approximer  $\pi(\theta | z)$  en écrivant  $\frac{1}{m} \sum_{j=1}^m \pi(\theta | z_2^{(j)}, x)$ . Pour obtenir une réalisation de  $\theta$  suivant  $\pi(\theta | x)$ , on choisira une composante  $(*)$  du mélange et on tirera suivant la loi Bêta de paramètre  $(z_2^{(*)} + x_4 + 1, x_2 + x_3 + 1)$ , obtenant ainsi  $\theta^{(1)}$ . En recommençant, on peut fabriquer  $\theta^{(2)}$  et ainsi de suite.

□

La procédure générale est la suivante :

A l'issue de la  $p^{\text{ème}}$  étape, disposant de  $\theta^{(p)}$ , on effectue :

1. une étape « *imputation* »

Pour  $j$  allant de 1 à  $m$ ,

- Simuler  $z^{(j)}$  suivant  $\pi(z | \theta^{(p)}; x)$

2. une étape « *a posteriori* »

(a) Tirer une composante  $(*)$  de  $g^{(p+1)}(\theta) = \frac{1}{m} \sum_{j=1}^m \pi(\theta | z^{(j)}, x)$ ,

(b) Générer  $\theta^{(p+1)}$  suivant la loi  $\pi(\theta | z^{(*)}, x)$

La suite de valeurs générées par cette procédure correspond au mouvement d'une chaîne de Markov. En effet, la densité du *latent* sachant *l'observé* peut s'exprimer par :

$$\pi(z | x) = \int_{\Theta} \pi(z | \phi, x) \pi(\phi | x) d\phi. \quad (7)$$

De sorte qu'en remplaçant  $\pi(z | x)$  par (7) dans (5) et en changeant l'ordre d'intégration, il vient :

$$\begin{aligned} \pi(\theta | x) &= \int_Z \pi(\theta | z, x) \int_{\Theta} \pi(z | x, \phi) \pi(\phi | x) d\phi dz \\ &= \int_{\Theta} \left[ \int_Z \pi(\theta | z, x) \pi(z | x, \phi) dz \right] \pi(\phi | x) d\phi \end{aligned} \quad (8)$$

On pose :

$$\int_Z \pi(\theta | z, x) \pi(z | x, \phi) dz = Q(\theta, \phi), \quad (9)$$

$\pi(\cdot | x)$  est donc tel que :

$$\pi(\theta | x) = \int_Z Q(\theta, \phi) \pi(\phi | x) d\phi$$

Autrement dit, la loi a posteriori  $\pi(\cdot | x)$  est solution d'une équation de la forme :

$$f(\theta) = \int Q(\theta, \phi) f(\phi) d\phi.$$

L'algorithme itératif TW suggéré précédemment se propose de calculer successivement :

$$f^{(p+1)}(\cdot) = T f^{(p)}(\cdot)$$

partant d'une valeur initiale  $f^{(0)}$  de  $\pi(\theta | x)$ .

Ainsi, la loi a posteriori  $\pi(\theta | x)$  apparaît comme un point fixe de l'opérateur  $T$ . Si l'on souhaite étudier la convergence de l'algorithme, on est donc ramené à l'étude de la convergence d'une chaîne de Markov.

◇