

Méthodes de Simulation

JEAN-YVES TURNERET ET NICOLAS DOBIGEON

Institut de recherche en informatique de Toulouse (IRIT)
ENSEEIH, Toulouse, France

Pages personnelles :

<http://turneret.perso.enseeiht.fr/>

<http://dobigeon.perso.enseeiht.fr/>

Méthodes de simulation

- **Partie 1** : Bayes et Simulation
- **Partie 2** : Metropolis - Hastings
- **Partie 3** : L'échantillonneur de Gibbs
- **Demain** : Application à l'analyse d'images MRFM et hyperspectrales
- **Pour finir** : Livres, Sites Webs, Pages perso, ...

Partie 1 : Bayes et Simulation

- 1) Introduction : **modèles** statistiques
- 2) **Maximum de vraisemblance**
- 3) Méthodes **Bayésiennes**
- 4) Méthodes de base de **simulation**
- 5) **Méthodes de Monte Carlo** pour l'intégration
- 6) Méthodes numériques **déterministes**

Modèle Statistique (1)

Compromis entre

- un modèle **compliqué** proche de la réalité qui peut induire des méthodes d'estimation, de détection ou de classification non standards
- un modèle **simple** qui conduit à des hypothèses comme linéarité, Gaussianité, ... mais qui peut être trop éloigné du phénomène physique étudié

Avec le développement de la puissance de calcul, des méthodes comme **les méthodes de simulation** peuvent être envisagées plus facilement.

Modèle Statistique (2)

Parfois, on choisit un modèle simple mais la suppression de certaines informations rend le problème difficile :

- Modèles de **censure**

$$y_i = \min\{x_i, c\}$$

- Modèles de **mélanges**

$$y_i \sim p_1 f_1(x) + \dots + p_k f_k(x)$$

- Modèles **stationnaires par morceaux**

$$y_i \sim f_k(x) \text{ si } i \in [t_k, t_{k+1}[$$

Maximum de Vraisemblance

- **Définition** : Pour un échantillon $\mathbf{x} = (x_1, \dots, x_n)$ de densité $f(\mathbf{x}|\boldsymbol{\theta})$, la vraisemblance s'écrit :

$$L(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i|\boldsymbol{\theta})$$

- **Propriétés asymptotiques** : asymptotiquement sans biais, convergent et efficace
- Facile à comprendre et souvent facile à étudier

Mais **pose problème** pour de nombreux modèles statistiques

Exemple 1 : loi Gamma, α connu

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} \mathbb{I}_{\mathbb{R}^+}(x)$$

- **Log-vraisemblance :**

$$\begin{aligned} \ln L(\mathbf{x}|\alpha, \beta) &= -n \ln \Gamma(\alpha) - n\alpha \ln \beta \\ &\quad + (\alpha - 1) \sum_{i=1}^n \ln x_i - \frac{1}{\beta} \sum_{i=1}^n x_i \end{aligned}$$

- Estimateur du **maximum de vraisemblance** de β :

$$\hat{\beta} = \frac{1}{N\alpha} \sum_{i=1}^n x_i$$

Exemple 2 : loi Gamma, α inconnu

- Estimateur du **maximum de vraisemblance**

$$\frac{\partial}{\partial \alpha} \ln L(\mathbf{x}|\alpha, \beta) = 0$$

$$\frac{\partial}{\partial \beta} \ln L(\mathbf{x}|\alpha, \beta) = 0$$

Équations **non-linéaires** faisant intervenir la fonction digamma !!

Exemple 3 : loi de Student

$$f(x|\theta, p, \sigma) \propto \frac{1}{\sigma} \left(1 + \frac{(x - \theta)^2}{p\sigma^2} \right)^{-\frac{p+1}{2}}$$

- Log-vraisemblance :

$$\ln L(\mathbf{x}|\theta, p, \sigma) = - \left(\frac{p+1}{2} \right) \ln \left(\sigma^{\frac{2n}{p+1}} \prod_{i=1}^n \left(1 + \frac{(x_i - \theta)^2}{p\sigma^2} \right) \right)$$

possède n maxima locaux (p et σ^2 connus)

matlab: student

Modèles de mélange

● Définition

Supposons que x_i suive la loi de densité $f_j(x_i)$ avec la probabilité p_j :

$$f(x_i) = p_1 f_1(x_i) + \dots + p_k f_k(x_i)$$

● Vraisemblance

$$L(\mathbf{x}|\boldsymbol{\theta}, p) = \prod_{i=1}^n [p_1 f_1(x_i) + \dots + p_k f_k(x_i)]$$

comporte k^n termes. Donc les techniques classiques d'optimisation sont inappropriées à une telle fonction multimodale.

Méthodes Bayésiennes

- Vraisemblance

$$f(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n f(x_i|\boldsymbol{\theta})$$

- Loi a priori sur $\boldsymbol{\theta}$

$$\pi(\boldsymbol{\theta})$$

- Loi a posteriori

$$f(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

où $f(\mathbf{x}) = \int f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ est la loi marginale de \mathbf{x} .

Inférence Bayésienne

On rencontre deux types de problèmes avec les méthodes d'estimation Bayésiennes

$$E \left[C \left(\theta, \hat{\theta}(x) \right) \right] = \int \left[\int C(\theta, \hat{\theta}(x)) f(\theta, \mathbf{x}) d\mathbf{x} \right] d\theta$$

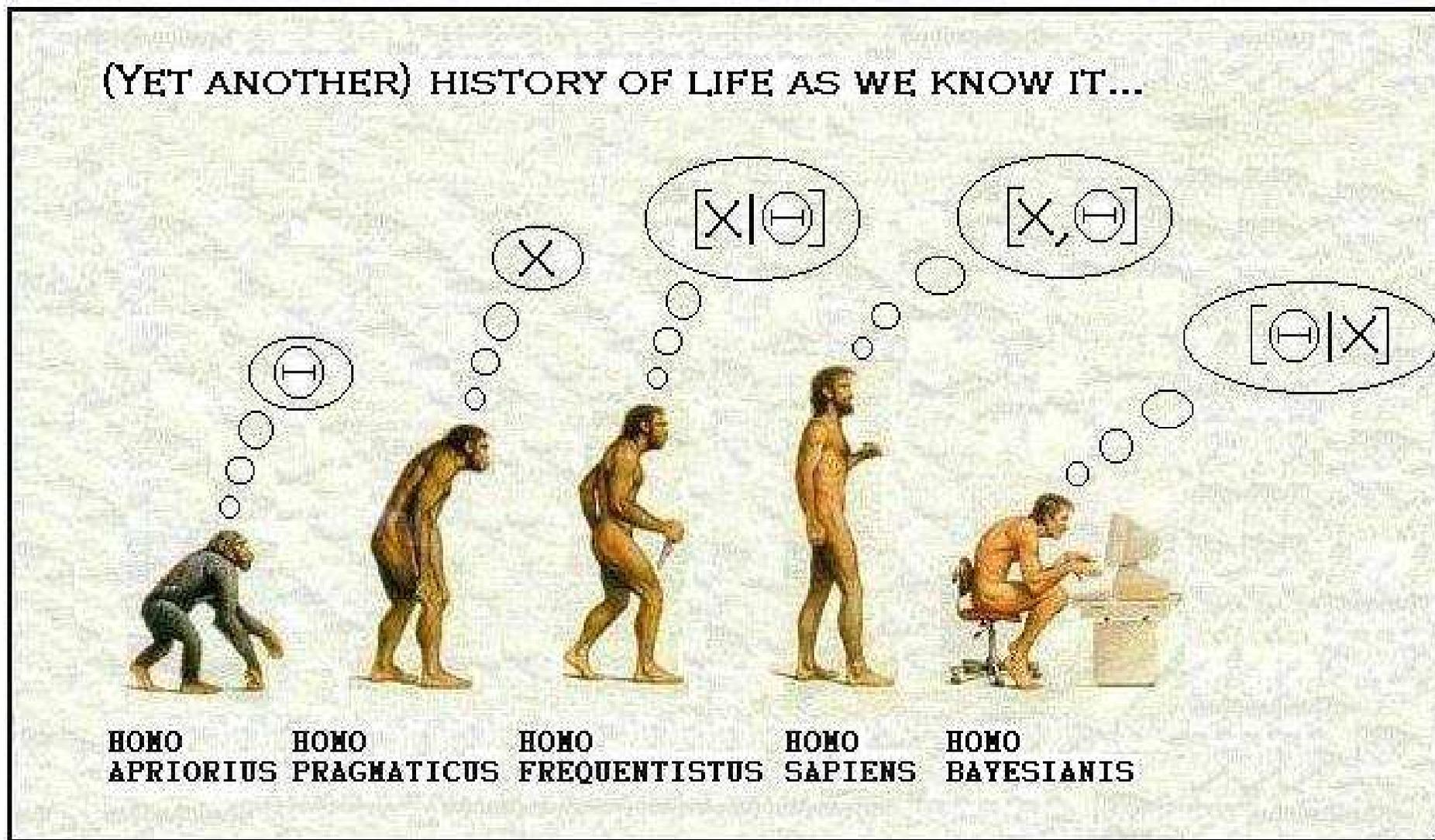
- Des problèmes d'**optimisation** (coût 0 – 1) : estimateur du **maximum a Posteriori**

$$\hat{\theta}_{\text{MAP}}(x) = \arg \max f(\theta|\mathbf{x}) = \arg \max f(\mathbf{x}|\theta)\pi(\theta)$$

- Des problèmes d'**intégration** (coût quadratique) : estimateur **MMSE**

$$\hat{\theta}_{\text{MMSE}}(x) = E[\theta|\mathbf{x}] = \int \theta f(\theta|\mathbf{x}) d\theta$$

Méthodes Bayésiennes



Exemple : le cas Gaussien

- Données

$$f(\mathbf{x}|\mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

- Loi a priori : $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$

$$\pi(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\sigma_0^2}\right)$$

- Loi a posteriori : $\mu|\mathbf{x} \sim \mathcal{N}(\mu_n, \sigma_n^2)$

$$\mu_n = \left(\frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right) \frac{1}{n} \sum_{i=1}^n x_i + \left(\frac{\sigma^2}{\sigma^2 + n\sigma_0^2}\right) \mu_0$$

matlab: Bayes

Lois conjuguées

- **Définition** : une loi a priori $\pi(\theta)$ est conjuguée si $f(\mathbf{x}|\theta)$ et $\pi(\theta)$ appartiennent à la même famille de lois.
- **Cas Gaussien**

$$f(\mathbf{x}|\mu, \sigma^2) \propto \frac{1}{(\sigma^2)^{n/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right]$$

☞ loi conjuguée pour μ : **loi normale**

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

☞ loi conjuguée pour σ^2 : **loi inverse gamma**

$$\pi(\sigma^2|\kappa, \gamma) \propto \frac{1}{(\sigma^2)^{\kappa+1}} \exp \left(-\frac{\gamma}{\sigma^2} \right)$$

Méthodes de simulation

- **Générateur uniforme**

Pour une fonction de répartition F définie sur \mathbb{R} , on définit son **inverse généralisée** par

$$F^{-1}(u) = \inf\{x; F(x) \geq u\}$$

Alors, si U est **uniforme** sur $[0, 1]$, la variable aléatoire $F^{-1}(U)$ est de fonction de répartition F car

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

- Cette méthode nécessite de **connaître l'inverse généralisée de la fonction de répartition.**

Méthodes de simulation

Certaines méthodes utilisent des propriétés spécifiques de la loi à simuler :

- **loi Exponentielle**

$$X = -\frac{1}{\lambda} \ln U, \quad U \sim \mathcal{U}([0, 1])$$

la méthode de l'inverse généralisée donne $X = -\frac{1}{\lambda} \ln(1 - U)$.

- **Loi Gamma et Beta**

$$Y = -b \sum_{j=1}^a \ln U_j \sim \mathcal{Ga}(a, b), \quad a \in \mathbb{N}^*$$

$$Y = \frac{\sum_{j=1}^a \ln U_j}{\sum_{j=1}^{a+b} \ln U_j} \sim \mathcal{Be}(a, b), \quad a, b \in \mathbb{N}^*$$

Méthodes de simulation

- **Méthode de Box Müller**

Si U_1 et U_2 sont deux variables indépendantes uniformes sur $[0, 1]$, alors

$$Y_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$

$$Y_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$$

sont des variables iid distribuées suivant une loi $\mathcal{N}(0, 1)$.

- **Loi de Poisson**

Si $X_i \sim \mathcal{E}(\lambda)$ et $N \sim \mathcal{P}(\lambda)$ alors

$$P[N = k] = P[X_1 + \dots + X_k \leq 1 < X_1 + \dots + X_{k+1}]$$

Méthodes d'acceptation-rejet

- Beaucoup de lois sont difficiles à simuler **directement** avec les méthodes précédentes
- Il y a certaines applications où la loi à simuler f est connue **à une constante multiplicative près** (méthodes Bayésiennes)
-  Une solution est de simuler à l'aide d'une loi de proposition g **plus simple** et d'utiliser un algorithme d'**acceptation-rejet**

Algorithme d'acceptation-rejet

Soit une loi d'intérêt de densité f et une **loi de proposition** de densité g telle que

$$f(x) \leq M g(x)$$

sur le support de f . Alors, on peut simuler suivant f avec l'algorithme suivant

- 1) **Générer** $X \sim g$ et $U \sim \mathcal{U}([0, 1])$
- 2) **Accepter** $Y = X$ si

$$U \leq \frac{f(X)}{M g(X)}$$

- 3) **Retourner en 1) si rejet**

Probabilité d'acceptation

$$\begin{aligned} P[X \text{ accepté}] &= P \left[U \leq \frac{f(X)}{Mg(X)} \right] = E \left[\mathbb{I}_{\{U \leq \frac{f(X)}{Mg(X)}\}} \right] \\ &= E \left[E \left[\mathbb{I}_{\{U \leq \frac{f(X)}{Mg(X)}\}} \mid X \right] \right] \\ &= E \left[\frac{f(X)}{Mg(X)} \right] \\ &= \int \frac{f(x)}{Mg(x)} g(x) dx = \frac{1}{M} \end{aligned}$$

loi de X

$$\begin{aligned} P[X < x | X \text{ accepté}] &= \frac{P[X < x, X \text{ accepté}]}{1/M} \\ &= MP \left[X < x, U < \frac{f(X)}{Mg(X)} \right] \\ &= ME \left[\mathbb{I}_{\{X < x, U \leq \frac{f(X)}{Mg(X)}\}} \right] \\ &= ME \left[E \left[\mathbb{I}_{\{X < x, U \leq \frac{f(X)}{Mg(X)}\}} \mid X \right] \right] \\ &= ME \left[\mathbb{I}_{\{X < x\}} \frac{f(X)}{Mg(X)} \right] \\ &= \int_{-\infty}^x \frac{f(x)}{g(x)} g(x) dx = F(x) \end{aligned}$$

Remarques

- Cet algorithme permet de simuler une densité **connue à une const. multiplicative près**, e.g. $f(\theta|x) \propto f(x|\theta)\pi(\theta)$
- La **probabilité d'acceptation est $1/M$** donc la valeur de M règle l'efficacité (vitesse) de l'algorithme
- Problème pour **des densités à queues lourdes**. Par exemple, on ne peut simuler une loi de Cauchy avec une loi de proposition normale (mais on peut faire l'inverse !)
- Utilisable pour **un grand nombre de lois** : $\mathcal{N}(0, 1)$, $\mathcal{G}a(a, b)$, lois normales tronquées, ...

Exemple : Cauchy \rightarrow Normale

- Loi cible

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

- Loi de proposition

$$g(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

- Choix de M

$$\frac{f(x)}{g(x)} = \sqrt{\frac{\pi}{2}} (1+x^2) e^{-x^2/2} \leq \sqrt{\frac{2\pi}{e}} = 1.52$$

valeur atteinte en ± 1 . Proba d'acceptation $1/M \simeq 0.66$.

matlab: accept-reject pour différentes valeurs de M

Intégration par la méthode de Monte Carlo

On cherche à évaluer

$$E[h(\Theta)] = \int_{\mathcal{P}} h(\theta) f(\theta) d\theta,$$

où \mathcal{P} est l'espace des paramètres, f est une densité connue et h est une fonction connue.

- **Solution** : générer un échantillon $(\theta_1, \dots, \theta_n)$ distribué suivant f pour approcher cette intégrale :

$$E[h(\Theta)] \simeq \bar{h}_n = \frac{1}{n} \sum_{j=1}^n h(\theta_j),$$

- **Justification** : loi forte des grands nombres

- **Erreur** : $O\left(\frac{1}{\sqrt{n}}\right)$ (remember, **curse of dimensionality!**) Cste

Intervalles de confiance

• **Variance :**

$$v_m = \frac{1}{m^2} \sum_{j=1}^m [h(\theta_j) - \bar{h}_m]^2$$

• **Loi asymptotique :**

$$\frac{\bar{h}_m - E[h(\Theta)]}{\sqrt{v_m}} \sim \mathcal{N}(0, 1)$$

☞ On peut déterminer des intervalles de confiance sur les paramètres inconnus !

Échantillonnage d'importance

- **Définition :**

$$E[h(\Theta)] = \int_{\mathcal{P}} \left[h(\theta) \frac{f(\theta)}{g(\theta)} \right] g(\theta) d\theta,$$

qui permet de simuler suivant g .

- **Estimation :** générer un échantillon $(\theta_1, \dots, \theta_n)$ distribué suivant g pour approcher cette intégrale :

$$E[h(\Theta)] \simeq \frac{1}{n} \sum_{j=1}^n \frac{f(\theta_j)}{g(\theta_j)} h(\theta_j),$$

Choix de la loi de proposition

- Loi g **simple** à simuler
- Si le **support** de g contient celui de f , l'estimateur converge vers

$$\int_{\mathcal{P}} h(\theta) f(\theta) d\theta$$

- La **variance** de l'estimateur est finie si

$$E \left[h^2(\Theta) \frac{f(\Theta)}{g(\Theta)} \right] < \infty$$

- Éviter les lois de proposition telles que $\sup_{\theta \in \mathcal{P}} \frac{f(\theta)}{g(\theta)} = \infty$
 - 1) pb si le support de f n'est pas inclus dans celui de g ,
 - 2) il existe une loi optimale minimisant la variance qui dépend de l'intégrale à calculer !

Exemple

Soit f la densité d'une **loi de Student** à ν degrés de liberté.
Calcul de

$$I = \int_a^\infty \theta^5 f(\theta) d\theta,$$

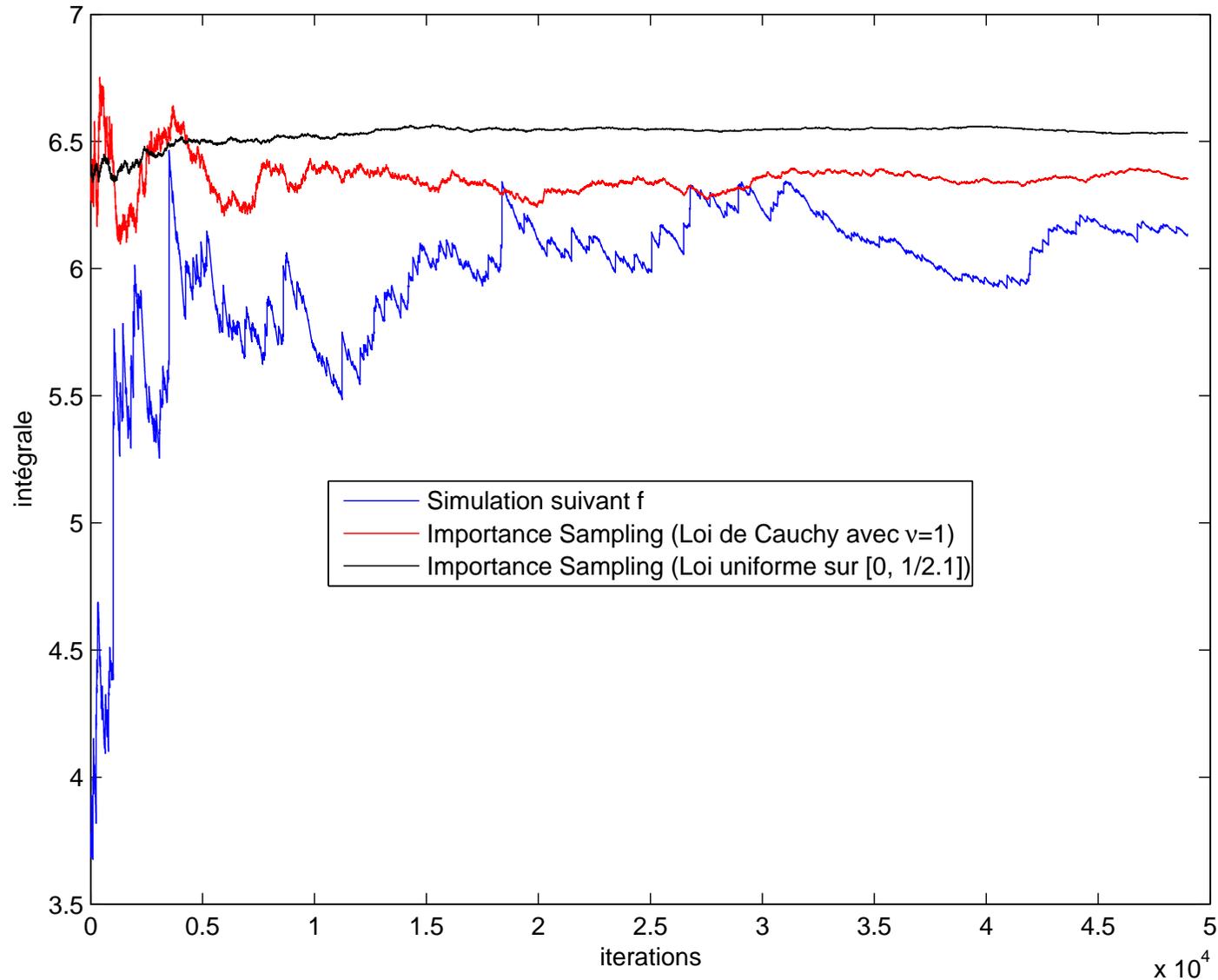
- Simulation **suivant f**
- Échantillonnage d'importance avec **loi de Cauchy**
- Un changement de variables $u = 1/\theta$ permet d'obtenir

$$I = \int_0^{\frac{1}{a}} a \frac{1}{au^7} f\left(\frac{1}{u}\right) du \simeq \frac{1}{a} \frac{1}{n} \sum_{i=1}^n \frac{1}{u_j^7} f\left(\frac{1}{u_j}\right),$$

où U suit une **loi uniforme sur $[0, \frac{1}{a}]$** .

matlab : integrale-student, $I = 6.54$, variance des estimées pour $n = 5000$

Exemple : $\nu = 12, a = 2.1$



Méthodes déterministes d'optimisation

Pour résoudre une équation de la forme

$$f(\theta) = 0,$$

on peut utiliser des algorithmes comme l'algorithme de

Newton-Raphson :

$$\theta_{n+1} = \theta_n + \left[\frac{\partial f}{\partial \theta}(\theta_n) \right]^{-1} f(\theta_n),$$

qui converge vers la solution $f(\theta) = 0$.

- **convergence lente** en $O(n^2)$ ou $O(n^3)$ alors que pour une méthode de simulation, on aura classiquement une convergence en $O(n)$!

Méthodes déterministes d'intégration

Pour calculer une intégrale de la forme

$$\int_a^b f(\theta) d\theta,$$

on peut utiliser des algorithmes basés sur les **sommes de Riemann** (méthode des trapèzes, méthode de Simpson, ...).

- On peut explorer des **zones de faibles probabilités**
- On a en général des problèmes pour des **fonctions multi-modales**.
- L'erreur est en $O\left(\frac{1}{n^{1/d}}\right)$, où d est la dimension de l'espace! (**curse of dimensionality**).

Pour les méthodes de Monte-Carlo, on aura une erreur en $O\left(\frac{1}{\sqrt{n}}\right)$!

Partie 2 : Metropolis - Hastings

- 1) **Introduction** : méthodes de Monte Carlo par chaînes de Markov (MCMC)
- 2) L'algorithme de Metropolis-Hastings **indépendant**
- 3) L'algorithme de Metropolis-Hastings **à marche aléatoire**
- 4) Algorithme de Green **à sauts réversibles**

Introduction

Pour approcher l'intégrale

$$\int_{\mathcal{P}} h(\theta) f(\theta) d\theta,$$

il n'est pas nécessaire de simuler suivant f (cf. échant. d'importance). Le principe des méthodes MCMC est de construire une **chaîne de Markov ergodique dont la loi stationnaire est f** :

- **Idée** : on part d'une valeur $\theta^{(0)}$ et on construit $\theta^{(t)}$ à l'aide d'un **noyau de transition** tel que la loi cible est f
- Pour t_0 "grand", $\theta^{(t_0)}$ est distribué suivant f
- **Remarque** : Les valeurs générées $\theta^{(t_0)}, \theta^{(t_0+1)}, \dots$ sont **dépendantes** car $\theta^{(t)}$ est une chaîne de Markov

Principes des méthodes MCMC

● Hypothèses

- On connaît la loi cible f à une constante multiplicative près
- On définit une **loi de proposition** (appelée aussi **loi instrumentale**) $q(y|\theta)$.

● Algorithme

- **Initialisation** : choix de $\theta^{(0)}$
- À partir de $\theta^{(t)}$, on génère $y^{(t)}$ à l'aide de la loi de proposition et on accepte ou rejette cette valeur de $y^{(t)}$ à l'aide d'une **procédure d'acceptation-rejet**. La valeur retenue est notée $\theta^{(t+1)}$.
- Les premières valeurs générées par l'algorithme ne seront pas utilisées pour l'inférence ("**burn-in**")

L'algorithme de Metropolis-Hastings

Étant donné $\theta^{(t)}$,

1. **Générer** $y_t \sim q(y|\theta^{(t)})$.
2. **Acceptation-Rejet**

$$\theta^{(t+1)} = \begin{cases} y_t & \text{avec prob. } \rho(\theta^{(t)}, y_t), \\ \theta^{(t)} & \text{avec prob. } 1 - \rho(\theta^{(t)}, y_t), \end{cases}$$

où

$$\rho(\theta, y) = \min \left\{ \frac{f(y)}{f(\theta)} \frac{q(\theta|y)}{q(y|\theta)}, 1 \right\} .$$

Propriétés et commentaires

- Cas **symétrique** :

$$\rho(\theta^{(t)}, y_t) = \min \left\{ \frac{f(y_t)}{f(\theta^{(t)})}, 1 \right\} .$$

- On accepte toujours les valeurs de y_t **augmentant** la "vraisemblance"
- La loi cible f peut être connue **à une constante multiplicative près**
- La chaîne $(\theta^{(t)})_t$ peut prendre plusieurs fois la même valeur \Rightarrow échantillon non iid

Convergence

- Hypothèses

- Probabilité d'acceptation

$$\mathbf{P} \left[\frac{f(y_t) q(\theta^{(t)} | y_t)}{f(\theta^{(t)}) q(y_t | \theta^{(t)})} \geq 1 \right] < 1. \quad (1)$$

i.e., l'événement $\{\theta^{(t+1)} = \theta^{(t)}\}$ est possible.

- Loi de proposition

$$q(y|\theta) > 0 \text{ pour tout } (\theta, y), \quad (2).$$

En particulier, le support de la loi de proposition doit inclure le support de la loi cible !

Convergence

- **Conclusions**

- **Ergodicité**

Pour h tel que $E_f[|h(\Theta)|] < \infty$,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T h(\theta^{(t)}) = \int h(\theta) f(\theta) d\theta$$

- **Convergence en variation totale**

$$\lim_{n \rightarrow \infty} \left\| \int K^n(\theta, \cdot) \mu(d\theta) - f \right\|_{TV} = 0$$

pour toute loi initiale μ , $K^n(\theta, \cdot)$ est le noyau de la chaîne après n transitions.

En particulier

$$\lim_{t \rightarrow \infty} P[\theta^{(t)} \in A] = \int_A f(\theta) d\theta$$

Metropolis-Hastings - Cas indépendant

La loi de proposition $q(y|\theta^{(t)})$ est **indépendante** de $\theta^{(t)}$

Étant donné $\theta^{(t)}$,

1. **Générer** $y_t \sim q(y)$.
2. **Acceptation-Rejet**

$$\theta^{(t+1)} = \begin{cases} y_t & \text{avec prob. } \min \left\{ \frac{f(y_t)}{f(\theta^{(t)})} \frac{q(\theta^{(t)})}{q(y_t)}, 1 \right\}, \\ \theta^{(t)} & \text{sinon} \end{cases}$$

Propriétés

- L'échantillon généré n'est pas iid
- Si $f(\theta) \leq Mq(\theta)$, $\forall \theta \in \text{supp } f$, alors $\|\cdot\|_{TV} \leq (1 - \frac{1}{M})^n$ (ergodicité uniforme)
- La probabilité d'acceptation est $\geq 1/M$ (i.e \geq proba acceptation-rejet)

Exemple : Loi Gamma

Soit f la densité d'une loi gamma $\mathcal{G}a(\alpha, \beta)$. Calcul de

$$I = \int_{-\infty}^{\infty} \theta^2 f(\theta) d\theta,$$

- **Acceptation rejet** avec $q(\theta) \sim \mathcal{G}a\left([\alpha], \frac{[\alpha]}{\alpha}\right)$, $f(\theta) < Mq(\theta)$

$$M = \exp\{\alpha(\ln(\alpha) - 1) - [\alpha](\ln([\alpha]) - 1)\}$$

- Algo de **Metropolis-Hastings** avec $q(\theta) \sim \mathcal{G}a\left([\alpha], \frac{[\alpha]}{\alpha}\right)$

$$\rho(\theta^{(t)}, y_t) = \min \left\{ \left(\frac{y_t}{\theta^{(t)}} \exp \left[\frac{\theta^{(t)} - y_t}{\alpha} \right] \right)^{\alpha - [\alpha]}, 1 \right\}$$

Matlab : lois-gamma, $I = 8.33$, TSVP pour exemples

nombre de données aléatoire avec acceptation-rejet

Acceptation-Rejet - Loi Gamma

1. **Générer** $y \sim \mathcal{Ga} \left([\alpha], \frac{[\alpha]}{\alpha} \right)$.
2. **Acceptation-Rejet**

$$\theta^{(t)} = y \text{ avec prob. } \left(\frac{ey \exp(-y/\alpha)}{\alpha} \right)^{\alpha - [\alpha]}$$

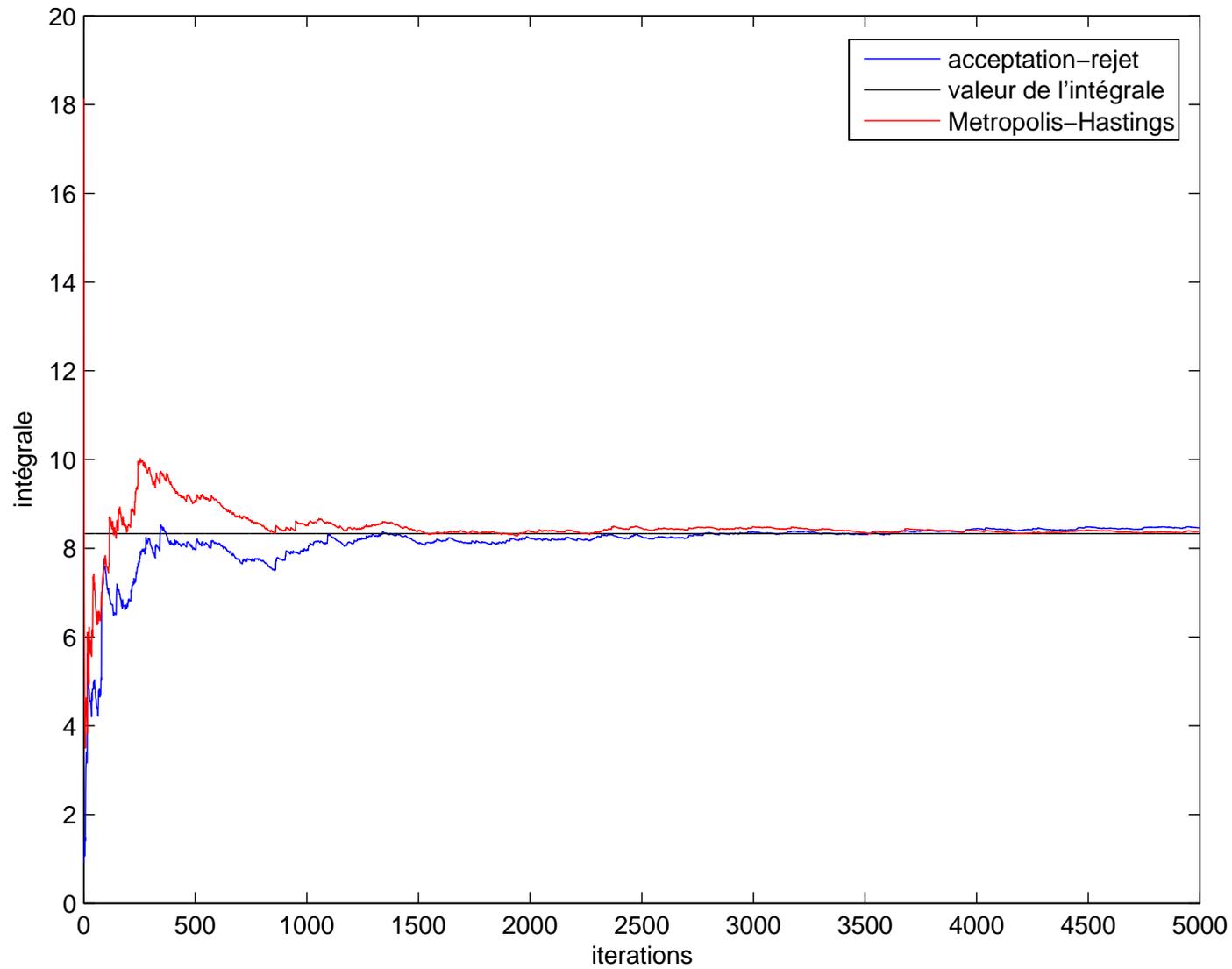
Metropolis-Hastings - Loi Gamma

Étant donné $\theta^{(t)}$,

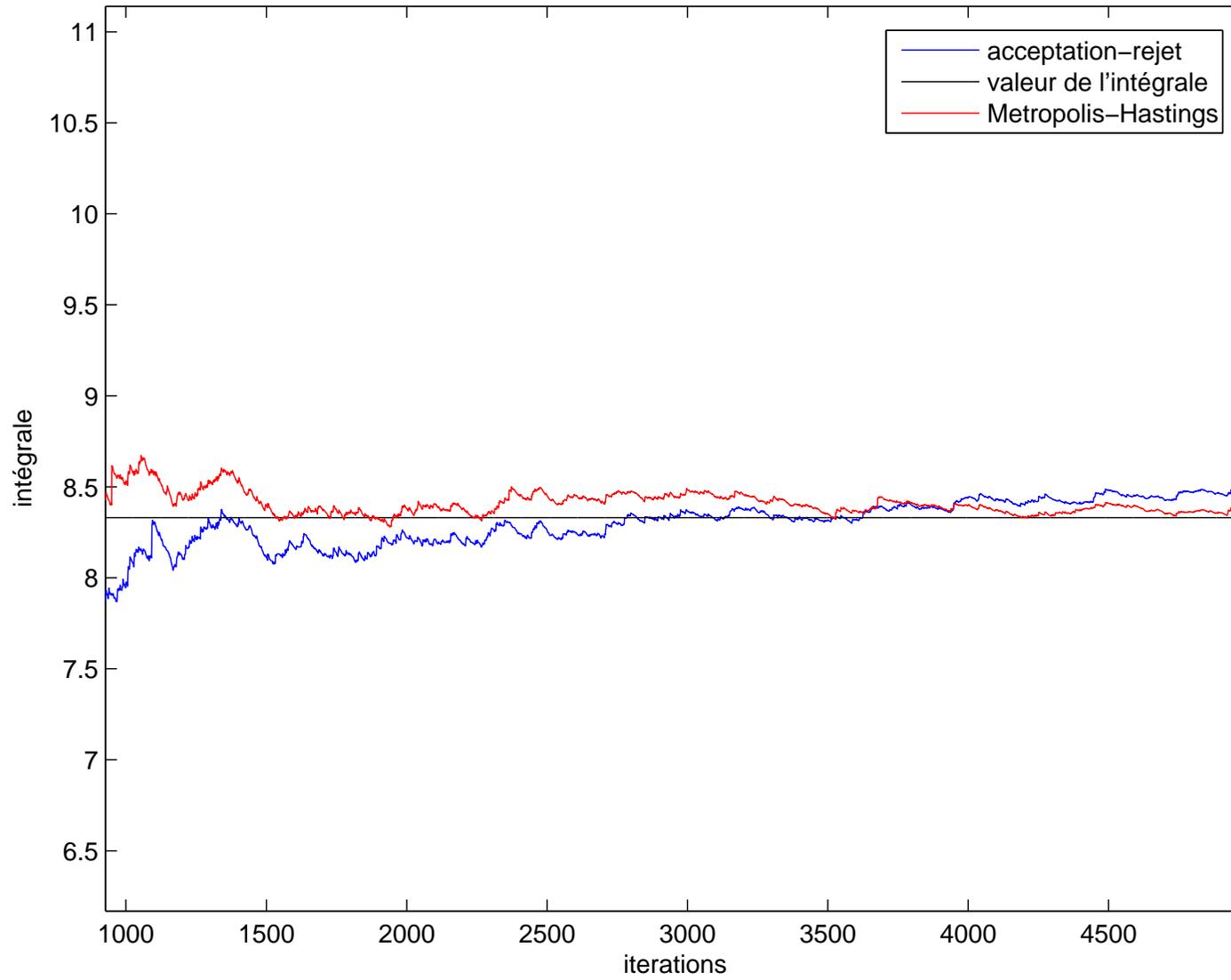
1. **Générer** $y_t \sim \mathcal{Ga} \left([\alpha], \frac{[\alpha]}{\alpha} \right)$.
2. **Acceptation-Rejet**

$$\theta^{(t+1)} = \begin{cases} y_t & \text{avec prob. } \min \left\{ \left(\frac{y_t}{\theta^{(t)}} \exp \left\{ \frac{\theta^{(t)} - y_t}{\alpha} \right\} \right)^{\alpha - [\alpha]}, 1 \right\} \\ \theta^{(t)} & \text{sinon} \end{cases}$$

Exemple : $\alpha = 2.43, \beta = 1$



Zoom



Metropolis-Hastings - Marche Aléatoire

La loi de proposition q est telle que

$$y_t = \theta^{(t)} + \epsilon_t,$$

où ϵ_t indépendant de $\theta^{(t)}$, i.e. $q(y|\theta) = q(y - \theta)$. Si q est **symétrique**, on obtient l'algorithme suivant :

Étant donné $\theta^{(t)}$,

1. **Générer** $y_t \sim q(y - \theta^{(t)})$.
2. **Acceptation-Rejet**

$$\theta^{(t+1)} = \begin{cases} y_t & \text{avec prob. } \min \left\{ \frac{f(y_t)}{f(\theta^{(t)})}, 1 \right\}, \\ \theta^{(t)} & \text{sinon} \end{cases}$$

Propriétés

- Pas d'ergodicité **uniforme**
- Conditions suffisantes d'ergodicité **géométrique** pour des densités symétriques log-concaves ... (Mengersen & Tweedie, 1996)

$$\forall \theta \in \mathcal{P}, \quad \left\| \int K^n(\theta, \cdot) \mu(d\theta) - f \right\|_{TV} \leq \frac{M}{r^n},$$

avec $M < \infty$ et $r > 1$.

Applet 1 : exemple d'algorithme de Metropolis-Hastings à marche aléatoire, Jeff Rosenthal (Thanks!)

Applet 2 : problème de la non-convergence uniforme, Jeff Rosenthal (Thanks!)

Extensions

- Adaptive Rejection Metropolis Sampling (ARMS)
- Algorithme de Metropolis-Hastings à sauts réversibles
- Algorithmes de Langevin
- ...

Metropolis-Hastings avec sauts réversibles

“One of the things we do not know is the number of things we do not know” - Peter Green

- **Dans quel cas ?**

Lorsque l'espace des paramètres inconnus est de taille inconnue

- mélanges de lois
- modèles de types ARMA
- modèles stationnaires par morceaux

- **Solution**

- utiliser une loi de proposition qui permet de se déplacer dans des espaces de différentes dimensions

Densités jointe et a posteriori

• Loi jointe

$$f(k, \theta^{(k)}, \mathbf{x}) = f(\mathbf{x}|k, \theta^{(k)})f(\theta^{(k)}|k)f(k), \quad k \in \mathcal{K}, \theta^{(k)} \in \Theta_k$$

• $f(k)$: a priori sur le nombre de paramètres ($k \sim \mathcal{P}(\lambda)$)

• $f(\theta^{(k)}|k)$: loi a priori sur les paramètres sachant k

• $f(\mathbf{x}|k, \theta^{(k)})$: vraisemblance

• Loi a posteriori

$$f(k, \theta^{(k)}|\mathbf{x}) = \frac{f(k, \theta^{(k)}, \mathbf{x})}{\int \int f(k, \theta^{(k)}, \mathbf{x})d\theta^{(k)}dk} \propto f(k, \theta^{(k)}, \mathbf{x})$$

$(k, \theta^{(k)}) \in \bigcup_{k \in \mathcal{K}} \mathcal{C}_k$, $\mathcal{C}_k = \{k\} \times \mathbb{R}^{n_k}$ espace de dimension variable.

Transition de \mathcal{M}_k vers $\mathcal{M}_{k'}$

Pour se déplacer de \mathbb{R}^{n_k} vers $\mathbb{R}^{n_{k'}}$, avec $k \neq k'$, on doit compléter ces espaces afin de définir un difféomorphisme $g_{kk'}$

• Transition de \mathcal{M}_k vers $\mathcal{M}_{k'}$

$$g_{kk'} = \begin{cases} g_{1kk'} \left\{ \begin{array}{l} \mathbb{R}^{n_k + n_{kk'}} \rightarrow \mathbb{R}^{n_{k'}} \\ (\theta^{(k)}, u) \mapsto \theta^{(k')} \end{array} \right. \\ g_{2kk'} \left\{ \begin{array}{l} \mathbb{R}^{n_k + n_{kk'}} \rightarrow \mathbb{R}^{n_{k'k}} \\ (\theta^{(k)}, u) \mapsto u' \end{array} \right. \end{cases}$$

avec $n_k + n_{kk'} = n_{k'} + n_{k'k}$.

Transition de $\mathcal{M}_{k'}$ vers \mathcal{M}_k

Afin d'assurer la réversibilité, il faut aussi définir un difféomorphisme $g_{k'k}$ allant de $\mathbb{R}^{n'_k}$ vers \mathbb{R}^{n_k}

• Transition de $\mathcal{M}_{k'}$ vers \mathcal{M}_k

$$g_{k'k} = \begin{cases} g_{1k'k} \begin{cases} \mathbb{R}^{n_{k'}+n_{k'k}} \rightarrow \mathbb{R}^{n_k} \\ (\theta^{(k')}, u') \mapsto \theta^{(k)} \end{cases} \\ g_{2k'k} \begin{cases} \mathbb{R}^{n_{k'}+n_{k'k}} \rightarrow \mathbb{R}^{n_{kk'}} \\ (\theta^{(k')}, u') \mapsto u \end{cases} \end{cases}$$

Remarque : on peut avoir $u = 0$ ou $u' = 0$!

Probabilité d'acceptation

Le nouvel état $\theta^{(k')} = g_{1k'k}(\theta^{(k)}, u)$ est accepté avec la probabilité

$$\rho_{kk'} = \min \left\{ \frac{\text{Posterior } \mathcal{M}_{k'}}{\text{Posterior } \mathcal{M}_k} \frac{p_{k'k}}{p_{kk'}} \frac{\text{Proposal } u'}{\text{Proposal } u} \left| \frac{\partial(\theta^{(k')}, u')}{\partial(\theta^{(k)}, u)} \right|, 1 \right\}$$

avec

- $p_{k'k}$: proba de tenter un déplacement de $\mathbb{R}^{n_{k'}}$ vers \mathbb{R}^{n_k}
- $p_{kk'}$: proba de tenter un déplacement de \mathbb{R}^{n_k} vers $\mathbb{R}^{n_{k'}}$
- $\left| \frac{\partial(\theta^{(k')}, u')}{\partial(\theta^{(k)}, u)} \right|$: Jacobien de la transformation

Exemple scolaire

● Modèle \mathcal{M}_1

● $x_i \sim \mathcal{N}(\theta_1, 1), i \leq 50, x_i \sim \mathcal{N}(\theta_2, 1), i > 50, \mathcal{C}_1 = \{2\} \times \mathbb{R}^2$

● Posterior

$$\propto \prod_{j=1}^2 \exp \left(-\frac{1}{2} \sum_{i=t_j}^{t_{j+1}-1} (x_i - \theta_j)^2 \right) \exp \left(-\frac{1}{2} (\theta_j - \mu)^2 \right)$$

● Modèle \mathcal{M}_2

● $x_i \sim \mathcal{N}(\theta, 1), i = 1, \dots, 100, \mathcal{C}_2 = \{1\} \times \mathbb{R}$

● Posterior

$$\left(\frac{1}{2\pi} \right)^{50} \exp \left(-\frac{1}{2} \sum_{i=1}^{100} (x_i - \theta)^2 \right) \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} (\theta - \mu)^2 \right) \frac{1}{2}$$

Difféomorphisme g_{12}

- Passage de \mathcal{M}_1 à \mathcal{M}_2

$$g_{12} \begin{cases} \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ (\theta_1, \theta_2) \mapsto (\theta = \frac{\theta_1 + \theta_2}{2}, u = \frac{\theta_1 - \theta_2}{2}) \end{cases}$$

- Probabilité d'acceptation

$$\frac{\text{Posterior}_{\mathcal{M}_2}}{\text{Posterior}_{\mathcal{M}_1}} \frac{1/2}{1/2} \frac{q(u)}{1} |\text{Jacobien}| = \frac{\pi_2\left(\frac{\theta_1 + \theta_2}{2}\right) q\left(\frac{\theta_1 - \theta_2}{2}\right) \frac{1}{2}}{\pi_1(\theta_1, \theta_2)}$$

- Proposal $u \sim \mathcal{N}(\mu, 1)$

Difféomorphisme g_{21}

- Passage de \mathcal{M}_2 à \mathcal{M}_1

$$g_{21} \begin{cases} \mathbb{R}^2 \rightarrow \mathbb{R}^2 \\ (\theta, u) \mapsto (\theta_1 = \theta + u, \theta_2 = \theta - u) \end{cases}$$

- Probabilité d'acceptation

$$\frac{\text{Posterior}_{\mathcal{M}_1}}{\text{Posterior}_{\mathcal{M}_2}} \frac{1/2}{1/2} \frac{1}{q(u)} |\text{Jacobien}| = \frac{\pi_1(\theta + u, \theta - u)}{\pi_2(\theta)q(u)} 2$$

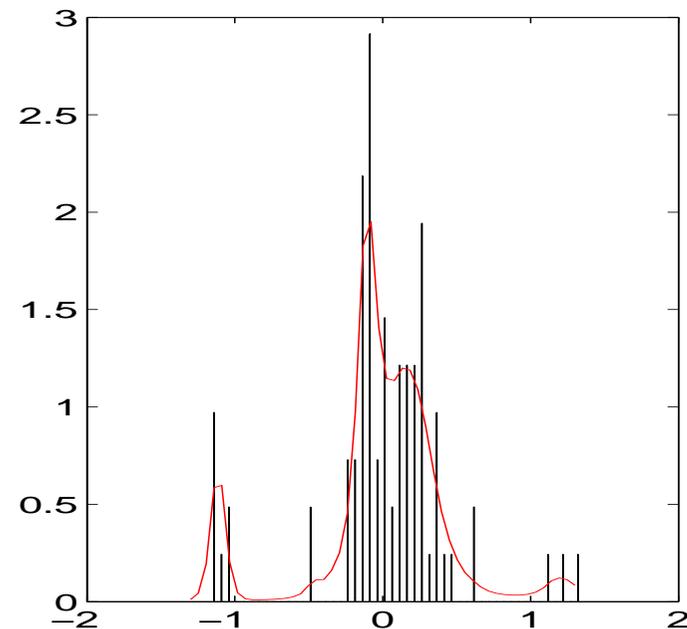
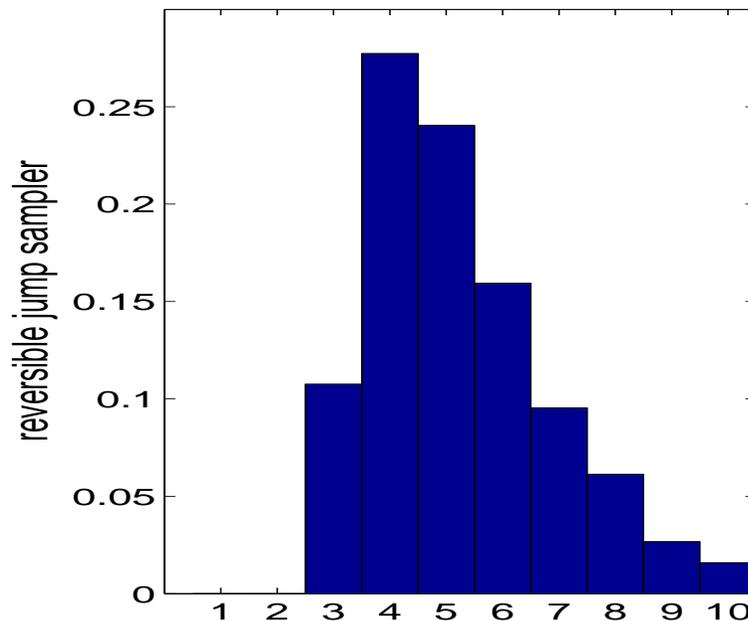
- Proposal $u \sim \mathcal{N}(\mu, 1)$

Matlab : `samplingGreen`

Mélange de Gaussiennes

- **Modèle** : y_1, \dots, y_n i.i.d., r **inconnu**

$$f(y|\theta_r) = \sum_{i=1}^r \frac{\omega_i}{\sqrt{2\pi\sigma_i^2}} \exp \left[-\frac{(y - m_i)^2}{2\sigma_i^2} \right]$$



Codes C disponibles sur la page d'Olivier Cappé,

http://www.tsi.enst.fr/~cappe/ctrj_mix

Partie 3 : L'échantillonneur de Gibbs

- 1) Principes généraux
- 2) Complétion
- 3) Convergence
- 4) Modèles hiérarchiques
- 5) Augmentation de données
- 6) Algorithme MCMC hybride

Principes généraux

Pour simuler suivant une loi $f(\boldsymbol{\theta})$ avec $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$, on peut utiliser l'idée suivante

- **Initialisation** : générer un vecteur $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ suivant une loi de proposition initiale π_0
- Simuler suivant **les lois conditionnelles**

$$\begin{aligned} \Theta_i | \theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p \\ \sim f_i(\theta_i | \theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p) \end{aligned}$$

for $i = 1, 2, \dots, p$.

L'échantillonneur de Gibbs

Étant donné $\theta^{(t)} = (\theta_1^{(t)}, \dots, \theta_p^{(t)})$,

1. **Générer** $\theta_1^{(t+1)} \sim f_1(\theta_1 | \theta_2^{(t)}, \dots, \theta_p^{(t)})$,

2. **Générer** $\theta_2^{(t+1)} \sim f_2(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)})$,

...

p. **Générer** $\theta_p^{(t+1)} \sim f_p(\theta_p | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_{p-1}^{(t+1)})$,

Seules les **lois conditionnelles** f_1, \dots, f_p sont utilisées pour la simulation. Donc, même pour un problème de grande dimension, toutes les simulations sont **univariées** !

Propriétés

- Taux d'acceptation égal à 1
- Choix de la loi de proposition imposé par la méthode
- Nécessite de connaître les lois conditionnelles de f
- Ne peut s'appliquer si le vecteur paramètre à simuler est de dimension variable
- Algorithme multi-dimensionnel par construction

Cas bidimensionnel

Pour simuler suivant

$$(X, Y) \sim f(x, y)$$

l'échantillonneur de Gibbs se réduit à

Simuler x_0 et pour $t = 1, 2, \dots$, générer (x_t, y_t) comme suit

1. $y_t \sim f_{y|x}(\cdot | x_{t-1})$,
2. $x_t \sim f_{x|y}(\cdot | y_t)$,

où $f_{y|x}$ et $f_{x|y}$ sont les lois conditionnelles du couple (X, Y) .

Remarque : $(x_t)_t$, $(y_t)_t$ et $(x_t, y_t)_t$ sont des **chaînes de Markov**.

Cas Gaussien : $X_i \sim \mathcal{N}(\mu, \sigma^2)$

- **Vraisemblance**

$$f(\mathbf{x}|\mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

- **Lois a priori**

- Moyenne

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

- Variance

$$\sigma^2 \sim \mathcal{IG}(\kappa, \gamma)$$

Lois conditionnelles

- **moyenne**

$$\mu | \sigma^2, \mathbf{x} \sim \mathcal{N}(\mu_n, \sigma_n^2)$$

avec

$$\mu_n = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n x_i \right) + \left(\frac{\sigma_n^2}{\sigma^2 + n\sigma_0^2} \right) \mu_0 \text{ et } \sigma_n^2 = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + n\sigma_0^2}$$

- **variance**

$$\sigma^2 | \mu, \mathbf{x} \sim \mathcal{IG} \left(\frac{n}{2} + \kappa, \frac{1}{2} \sum_{i=1}^n (x_i - m)^2 + \gamma \right)$$

Donc, on peut simuler des couples (μ, σ^2) avec l'échantillonneur de Gibbs

Complétion

- **Définition** : la densité g est une **complétion** de f si

$$\int_{\mathcal{Z}} g(\boldsymbol{\theta}, \boldsymbol{\eta}) d\boldsymbol{\eta} = f(\boldsymbol{\theta}),$$

i.e. si f est une loi marginale de g .

- **Intérêt** : les lois conditionnelles de g sont parfois plus simples à simuler que celles de f (e.g. **analyse Bayésienne hiérarchique**).
- **Notations** : pour $p > 1$, soit $\mathbf{Y} = (\boldsymbol{\theta}, \boldsymbol{\eta})$ de densité $g(\mathbf{y}) = g(y_1, \dots, y_p)$ et de lois conditionnelles

$$Y_i | y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p \sim g_i(y_i | y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p)$$

Echantillonneur de Gibbs après complétion

Étant donné $\mathbf{y}^{(t)} = (y_1^{(t)}, \dots, y_p^{(t)})$,

1. **Générer** $y_1^{(t+1)} \sim g_1(y_1 | y_2^{(t)}, \dots, y_p^{(t)})$,

2. **Générer** $y_2^{(t+1)} \sim g_2(y_2 | y_1^{(t+1)}, y_3^{(t)}, \dots, y_p^{(t)})$,

...

p. **Générer** $y_p^{(t+1)} \sim g_p(y_p | y_1^{(t+1)}, y_2^{(t+1)}, \dots, y_{p-1}^{(t+1)})$,

Exemple : loi Cauchy-Normale (1)

• Posterior

$$f(\theta|\theta_0) \propto \frac{e^{-\theta^2/2}}{[1 + (\theta - \theta_0)^2]^\nu}$$

• Complétion

On a

$$f(\theta|\theta_0) \propto \int_0^\infty e^{-\theta^2/2} e^{-[1+(\theta-\theta_0)^2] \eta/2} \eta^{\nu-1} d\eta$$

d'où

$$g(\theta, \eta) \propto e^{-\theta^2/2} e^{-[1+(\theta-\theta_0)^2] \eta/2} \eta^{\nu-1}$$

Exemple : loi Cauchy-Normale (2)

- Lois conditionnelles

$$g_1(\eta|\theta) = \mathcal{Ga}\left(\nu, \frac{1 + (\theta - \theta_0)^2}{2}\right),$$

$$g_2(\theta|\eta) = \mathcal{N}\left(\frac{\theta_0\eta}{1 + \eta}, \frac{1}{1 + \eta}\right).$$

Le paramètre η n'a pas d'intérêt physique et sert uniquement à **simplifier la simulation** d'un échantillon $\theta^{(t)}$.

Condition de positivité

- **Positivité**

$$g^{(i)}(y_i) > 0, \quad i = 1, \dots, p \Rightarrow g(y_1, \dots, y_p) > 0$$

où $g^{(i)}$ est la loi marginale de Y_i (ou support de la loi cible g égal au produit cartésien des supports des $g^{(i)}$)

- Pour montrer la **convergence** de l'échantillonneur de Gibbs, la loi cible doit vérifier la condition de positivité.

- **Contre-exemple**

$$g(y_1, y_2) = \frac{1}{2\pi} [\mathbb{I}_\epsilon(y_1, y_2) + \mathbb{I}_{\epsilon'}(y_1, y_2)],$$

où ϵ et ϵ' sont deux disques de rayons 1 centrés sur $(1, 1)$ et $(-1, -1)$

(autre exemple : vecteur non Gaussien dont les lois marginales sont Gaussiennes).

Illustration de la non-positivité

Initialisation Aléatoire

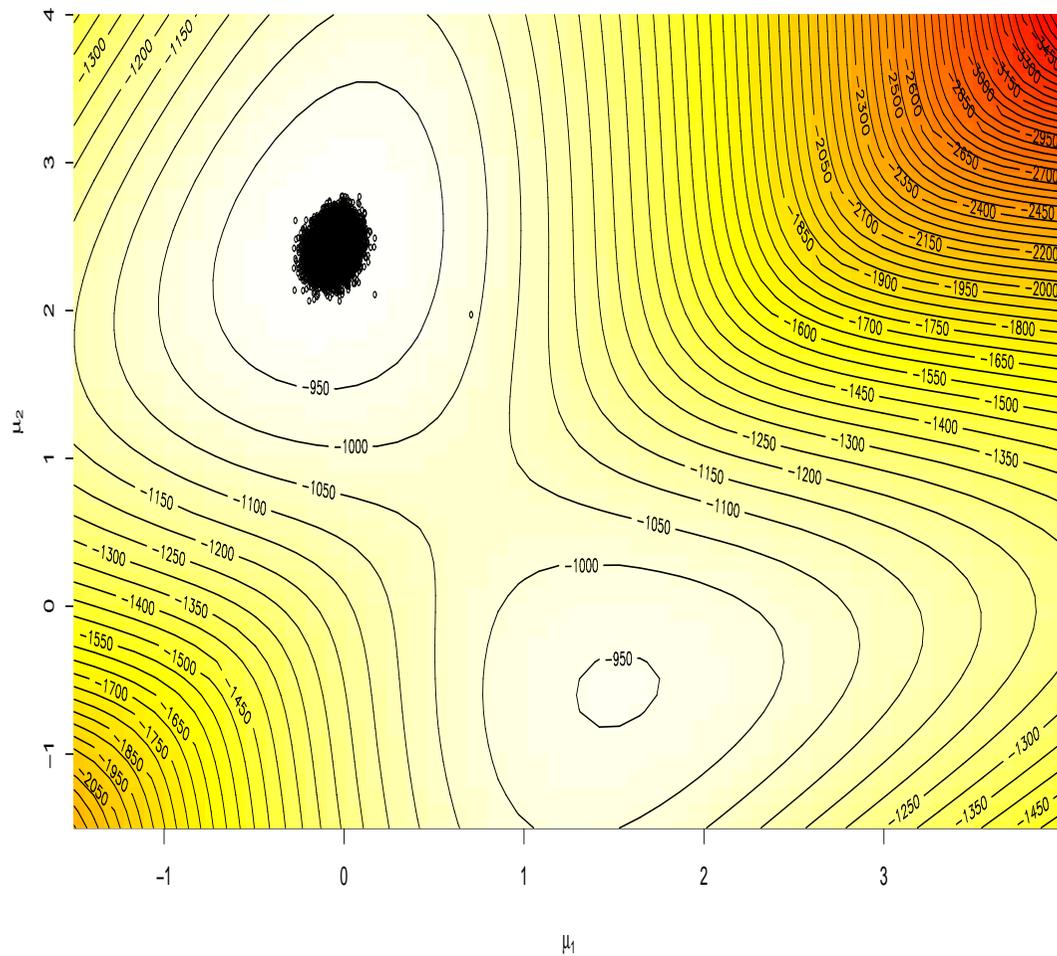
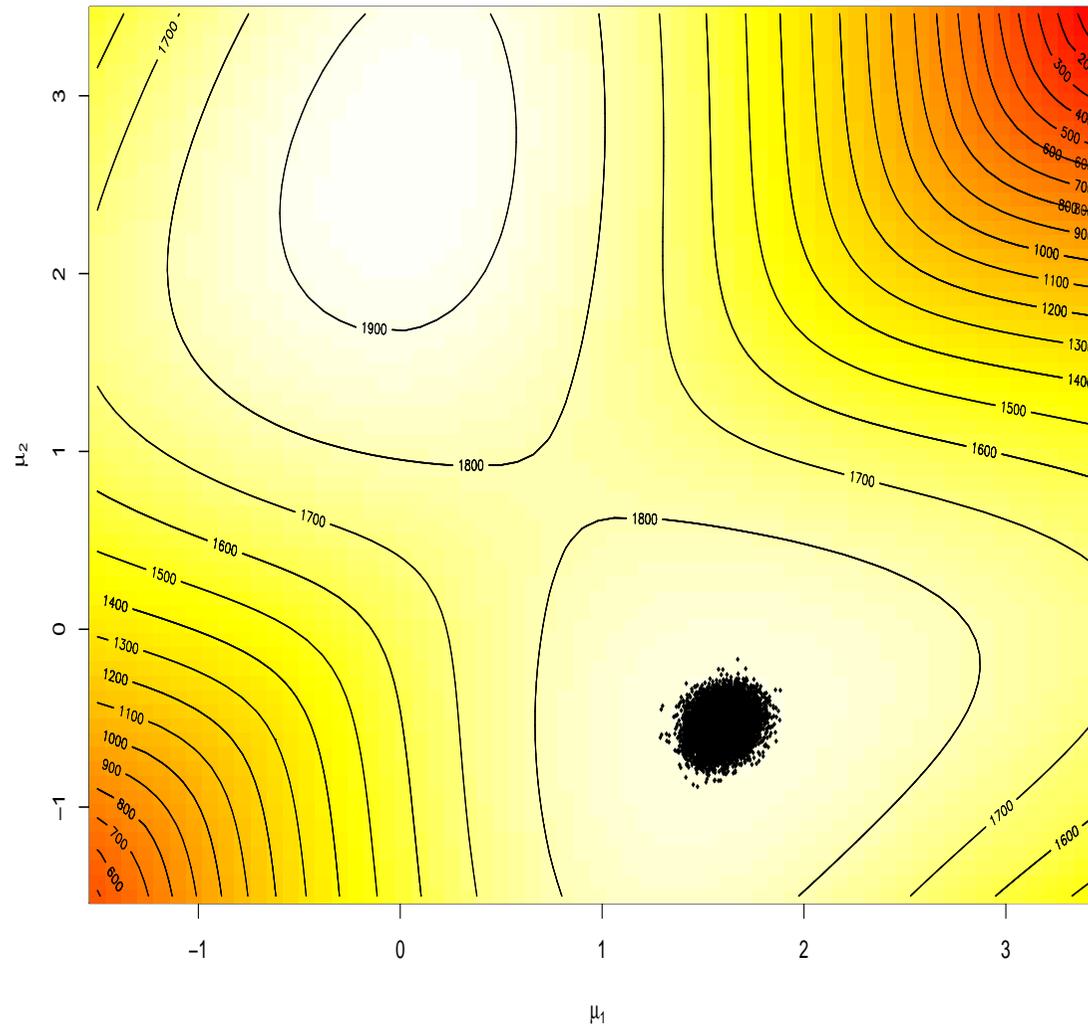


Illustration de la non-positivité

Gibbs coincé autour du mauvais mode



Convergence de l'échantillonneur de Gibbs

Si la **condition de positivité** est vérifiée et si le noyau de transition est absolument continu par rapport à g , on a

- **Ergodicité**

Si $\int |h(y)|g(y)dy < \infty$, alors

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T h(y^{(t)}) = \int h(y)g(y)dy$$

- **Convergence en variation totale**

$$\lim_{n \rightarrow \infty} \left\| \int K^n(y, \cdot) \mu(dy) - g \right\|_{TV} = 0$$

pour toute loi initiale μ .

Remarques

- L'échantillonneur de Gibbs est la **composition de p algorithmes de Metropolis-Hastings** avec des probabilités d'acceptation uniformément égales à 1.
- Échantillonneur de Gibbs à **balayage aléatoire**

Modèles hiérarchiques

L'échantillonneur de Gibbs est particulièrement bien adapté aux **modèles hiérarchiques** :

- Les paramètres inconnus sont munis de lois a priori ainsi que les **hyperparamètres** associés
- En général, on introduit des lois **non informatives** au dernier niveau de la hiérarchie

Exemple

- Données Poissonniennes

$$\begin{cases} X_i \sim \mathcal{P}(\lambda_1) & \text{pour } i = 1, \dots, l_1, \\ X_i \sim \mathcal{P}(\lambda_2) & \text{pour } i = l_1 + 1, \dots, n, \end{cases}$$

avec l_1 connu.

- Lois a priori sur les paramètres

$$\lambda_1 \sim \mathcal{Ga}(\alpha, \beta), \quad \lambda_2 \sim \mathcal{Ga}(\alpha, \beta), \quad \alpha = 2.$$

- Loi a priori sur les hyperparamètres

$$f(\beta) = \frac{1}{\beta} \mathbb{I}_{\mathbb{R}^+}(\beta)$$

- **Loi jointe**

$$f(\mathbf{x}, \boldsymbol{\lambda}, \beta) \propto \frac{1}{\beta} \prod_{i=1}^{l_1} \left[\frac{\lambda_1^{x_i}}{x_i!} e^{-\lambda_1} \right] \prod_{i=l_1+1}^n \left[\frac{\lambda_2^{x_i}}{x_i!} e^{-\lambda_2} \right] \prod_{i=1}^2 \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta \lambda_i}$$

- **Loi conditionnelles**

- pour les paramètres λ_i

$$\lambda_1 | \beta, \mathbf{x} \sim \mathcal{Ga} \left(\sum_{i=1}^{l_1} x_i + \alpha, \beta + l_1 \right)$$

$$\lambda_2 | \beta, \mathbf{x} \sim \mathcal{Ga} \left(\sum_{i=l_1+1}^n x_i + \alpha, \beta + n - l_1 \right),$$

- pour β

$$\beta | \mathbf{x}, \boldsymbol{\lambda} \sim \mathcal{Ga} (2\alpha, \lambda_1 + \lambda_2)$$

Algorithme MCMC hybride

● Motivations

- La convergence de l'échantillonneur de Gibbs peut être **lente** car on simule une seule composante à chaque itération
- Pas de problème avec **la loi de proposition** comme avec l'algorithme de Metropolis-Hastings
- Certaines lois conditionnelles peuvent être **impossibles à simuler**
- **Définition** : un algorithme MCMC **hybride** est une méthode MCMC utilisant simultanément des étapes d'échantillonneur de Gibbs et des étapes de Metropolis-Hastings

Algorithme MCMC hybride

Remplacer chaque étape i où une simulation suivant la loi conditionnelle $g_i(y_i | y_j, j \neq i)$ est impossible par

1. **Simuler** $\tilde{y}_i \sim q_i(y_i | y_1^{(t+1)}, \dots, y_i^{(t)}, y_{i+1}^{(t)}, \dots, y_p^{(t)})$,
2. **Prendre**

$$y_i^{(t+1)} = \begin{cases} \tilde{y}_i & \text{avec probabilité } \rho \\ y_i^{(t)} & \text{avec probabilité } 1 - \rho \end{cases}$$

$$\rho = 1 \wedge \left\{ \frac{g_i(\tilde{y}_i | y_1^{(t+1)}, \dots, y_i^{(t)}, y_{i+1}^{(t)}, \dots, y_p^{(t)})}{g_i(y_i^{(t)} | y_1^{(t+1)}, \dots, y_i^{(t)}, y_{i+1}^{(t)}, \dots, y_p^{(t)})} \frac{q_i(y_i^{(t)} | y_1^{(t+1)}, \dots, \tilde{y}_i, y_{i+1}^{(t)}, \dots, y_p^{(t)})}{q_i(\tilde{y}_i | y_1^{(t)}, \dots, y_i^{(t)}, y_{i+1}^{(t)}, \dots, y_p^{(t)})} \right\}$$

Remarque : l'étape de Metropolis-Hastings n'est utilisée **qu'une fois** (et la convergence est assurée).

Matlab : `metropolis_within_Gibbs`

Remerciements

- **Christian Robert** : pour ses excellents transparents
- **Éric Moulines** : pour ses excellents fichiers Latex
- **Olivier Cappé** : pour ses excellents conseils
- **Nicolas Dobigeon** : pour ses excellents programmes Matlab
- **Florent Chatelain** : pour ses excellents talents d'informaticien
- **Jeff Rosenthal and Laird Breyer** : for their excellent applets

Livres

- **Méthodes de Monte Carlo par Chaînes de Markov :**
C. Robert, Economica, 1996
- **Discretization and MCMC Convergence Assessment**
C. Robert, Springer, 1998
- **Monte Carlo Statistical Methods**
C. Robert et G. Casella, Springer, 2nd edition, Aout 2004
- **Markov Chain Monte Carlo in Practice**
W. R. Gilks, S. Richardson, and D. J. Spiegelhalter,
Chapman & Hall, 1996
- **Inference in Hidden Markov Models**
O. Cappé, E. Moulines, and T. Rydén, Springer, 2005

Tutoriaux

- Probabilistic Inference Using MCMC Methods, R. Neal, <http://www.cs.toronto.edu/%7Eradford/ftp/review.pdf>, 1993.
- Markov Chain Monte Carlo Method and Its Application by S. P. Brooks, *The Statistician*, vol. 47, no 1, pp. 69-100, 1998.
- Markov Chain Monte Carlo Method: Some Practical Implications of Theoretical Results by G. O. Roberts and J. S. Rosenthal, *Canadian J. Stat.*, vol. 26, pp. 5-31, 1998.
- Markov Chain Monte Carlo Methods with Applications to Signal Processing by W. J. Fitzgerald, *Signal Processing*, vol. 81, no 1, pp. 3-18, 2001.

Tutoriaux

- **Markov Chain Monte Carlo for Statistical Inference** by **J. Besag**, <http://www.statslab.cam.ac.uk/mcmc/pages/list.html>, April 2001.
- **Markov Chain Monte Carlo and Gibbs Sampling** by **B. Walsh**, <http://nitro.biosci.arizona.edu/courses/EEB581-2004/handouts/Gibbs.pdf>, 2004.
- **Introducing Markov chain Monte Carlo** by **W. R. Gilks, S. Richardson and D. J. Spiegelhalter**, in Markov Chain Monte Carlo in Practice. London, U.K.: Chapman & Hall, 1996..
- ...

Pages Personnelles

- **Georges Casella** : <http://www.stat.ufl.edu/casella/>
- **Petar Djuric** : <http://www.ee.sunysb.edu/djuric/>
- **Andrew Gelman** : <http://www.stat.columbia.edu/~gelman/>
- **John Geweke** : <http://www.biz.uiowa.edu/faculty/jgeweke/>
- **C. Geyer** : <http://www.stat.umn.edu/charlie/>
- **Simon Godsill** : www-sigproc.eng.cam.ac.uk/sjg/
- **Peter Green** : <http://www.maths.bris.ac.uk/peter/>
- **Per Mykland** : <http://galton.uchicago.edu/~mykland/>
- **A. Raftery** : <http://www.stat.washington.edu/raftery/>
- **Jeff. Rosenthal** : <http://probability.ca/jeff/>

Pages Personnelles (TS)

- **Olivier Cappé** : <http://www.tsi.enst.fr/cappe/>
- **Manuel Davy** : <http://www-lagis.univ-lille1.fr/davy/>
- **Arnaud Doucet** : <http://www.cs.ubc.ca/arnaud/>
- **Jérôme Idier** : <http://www.irccyn.ec-nantes.fr/idier/>
- **Eric Moulines** : <http://www.tsi.enst.fr/moulines/>
- **C. Robert** : <http://www.ceremade.dauphine.fr/~xian/>

Pages Personnelles (TI)

- **X. Descombes** : <http://www-sop.inria.fr/ariana/personnel/Xavier.Descombes/>
- **A. Djafari** : <http://djafari.free.fr/>
- **J. Idier** : <http://www.irccyn.ec-nantes.fr/idier/>
- **J.-F. Giovanelli** : <http://giovanelli.free.fr/>
- **P. Ciuciu** : <http://www.lnao.fr/spip.php?rubrique8>
- **C. Collet** : <http://lsiit-miv.u-strasbg.fr/lsiit/perso/collet/Perso/ColletPerso.htm>
- **H. Snoussi** : <http://h.snoussi.free.fr/>

Logiciels

- **Convergence Diagnosis and Output Analysis (CODA) :**
<http://www.mrc-bsu.cam.ac.uk/bugs/documentation/coda03/cdaman03.html>
- **Bayesian Analysis, Computation and Communication (BACC) Software (Matlab, S-plus, R) :**
<http://www2.cirano.qc.ca/bacc/>
- **Toolbox Matlab econometrics**
<http://www.spatial-econometrics.com/>
- **Bayesian inference Using Gibbs Sampling (BUGS)**
<http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>

Divers

- MCMC Preprint service

<http://www.statslab.cam.ac.uk/~mcmc/index.html>

- Special Issue on Markov Chain Monte Carlo (MCMC)
Methods for Signal Processing

Eurasip Signal Processing, vol. 81 (1), 2001.

- Special issue on Monte Carlo methods for statistical
signal processing

IEEE Transactions on Signal Processing, vol. 50 (5),
2002.