Cours 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 $\theta^{(t_0)}, \theta^{(t_0+1)}, \ldots$ sont dépendantes

Principes des méthodes MCMC

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

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

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

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

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

Convergence

- Conclusions
  - Ergodicité

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

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

  - Convergence en variation totale

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

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

En particulier

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

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

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

Propriétés

- L'échantillon généré n'est pas iid
- Si $f(\theta) \leq Mq(\theta),\ \forall \theta \in \text{supp } f$, alors $\|.\|_{TV} \leq \left(1 - \frac{1}{M}\right)^n$ (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 $q(\theta) \sim \mathcal{G}a\left([\alpha], \frac{[\alpha]}{\alpha}\right)$, $f(\theta) < Mq(\theta)$

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

- 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 $y \sim \mathcal{G}a\left([\alpha], \frac{[\alpha]}{\alpha}\right)$.
2. Acceptation-Rejet

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

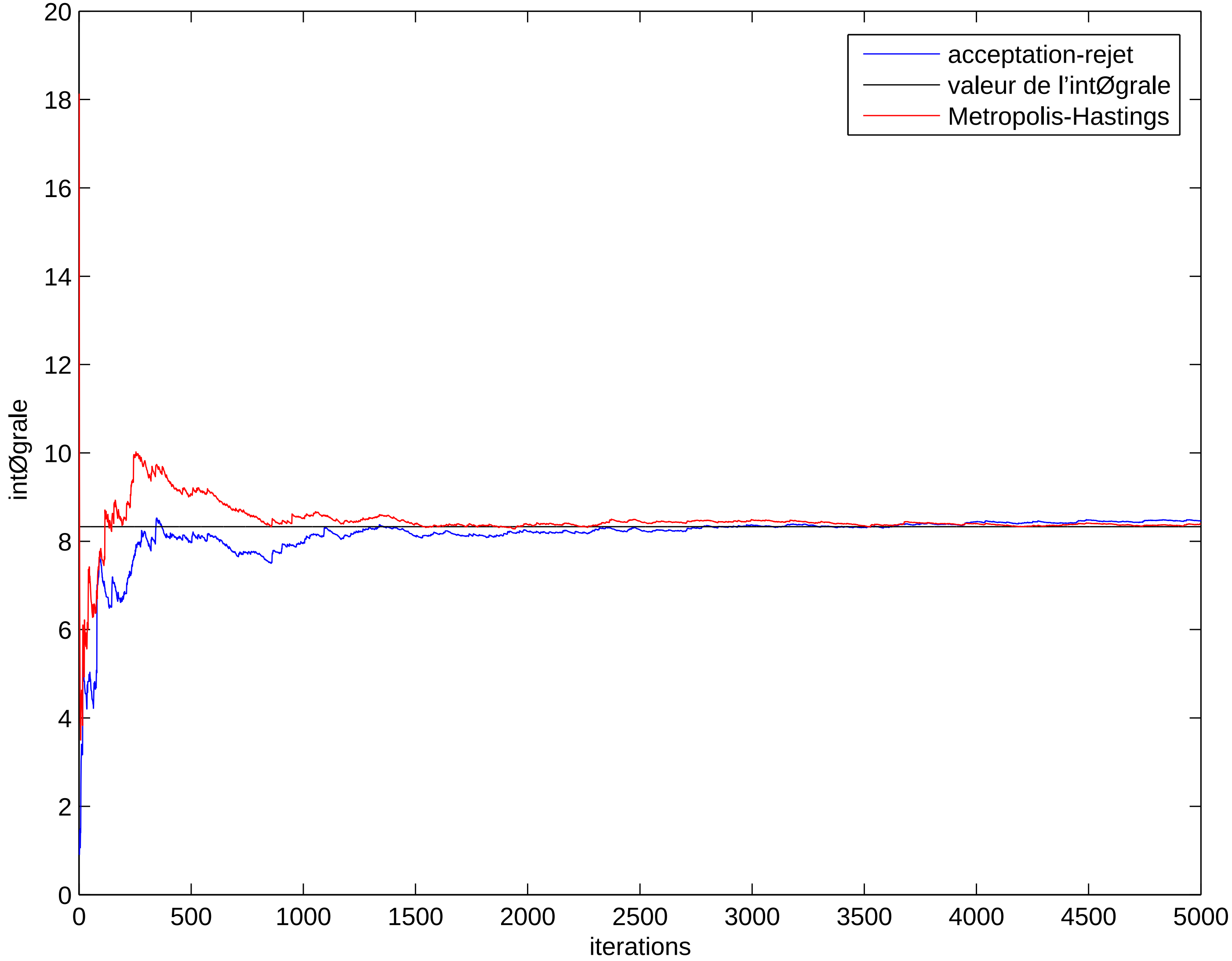
Metropolis-Hastings - Loi Gamma

Étant donné $\theta^{(t)}$,

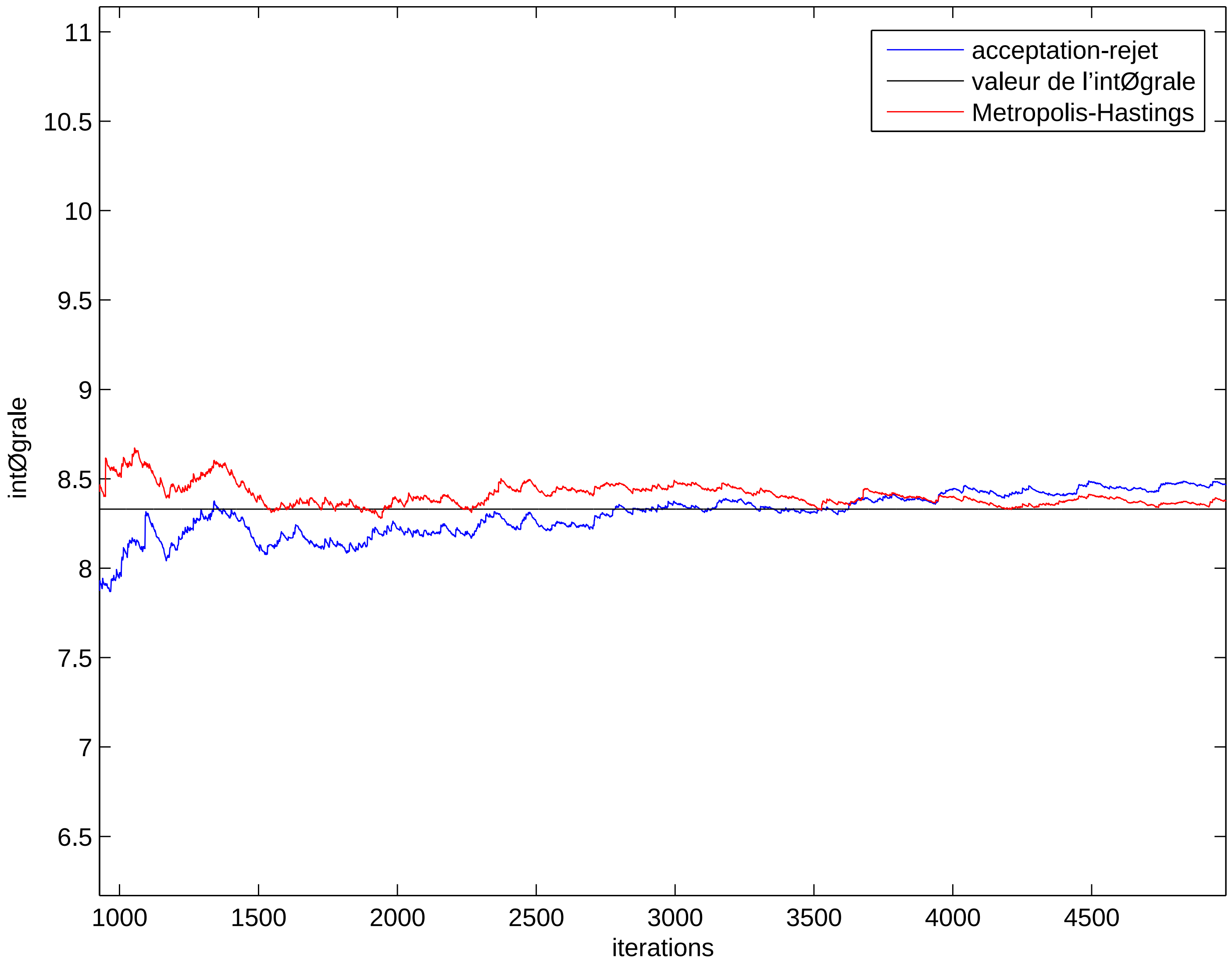
1. Générer $y_t \sim \mathcal{G}a\left([\alpha], \frac{[\alpha]}{\alpha}\right)$.
2. Acceptation-Rejet

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

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



Zoom



acceptation-rejet
valeur de l'intØgrale
Metropolis-Hastings
intØgrale
iterations
11
10.5
10
9.5
9
8.5
8
7.5
7
6.5
1000
1500
2000
2500
3000
3500
4000
4500

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!)

Exemple : Loi Normale

Simulation de données suivant la loi normale $\mathcal{N}(0,1)$.

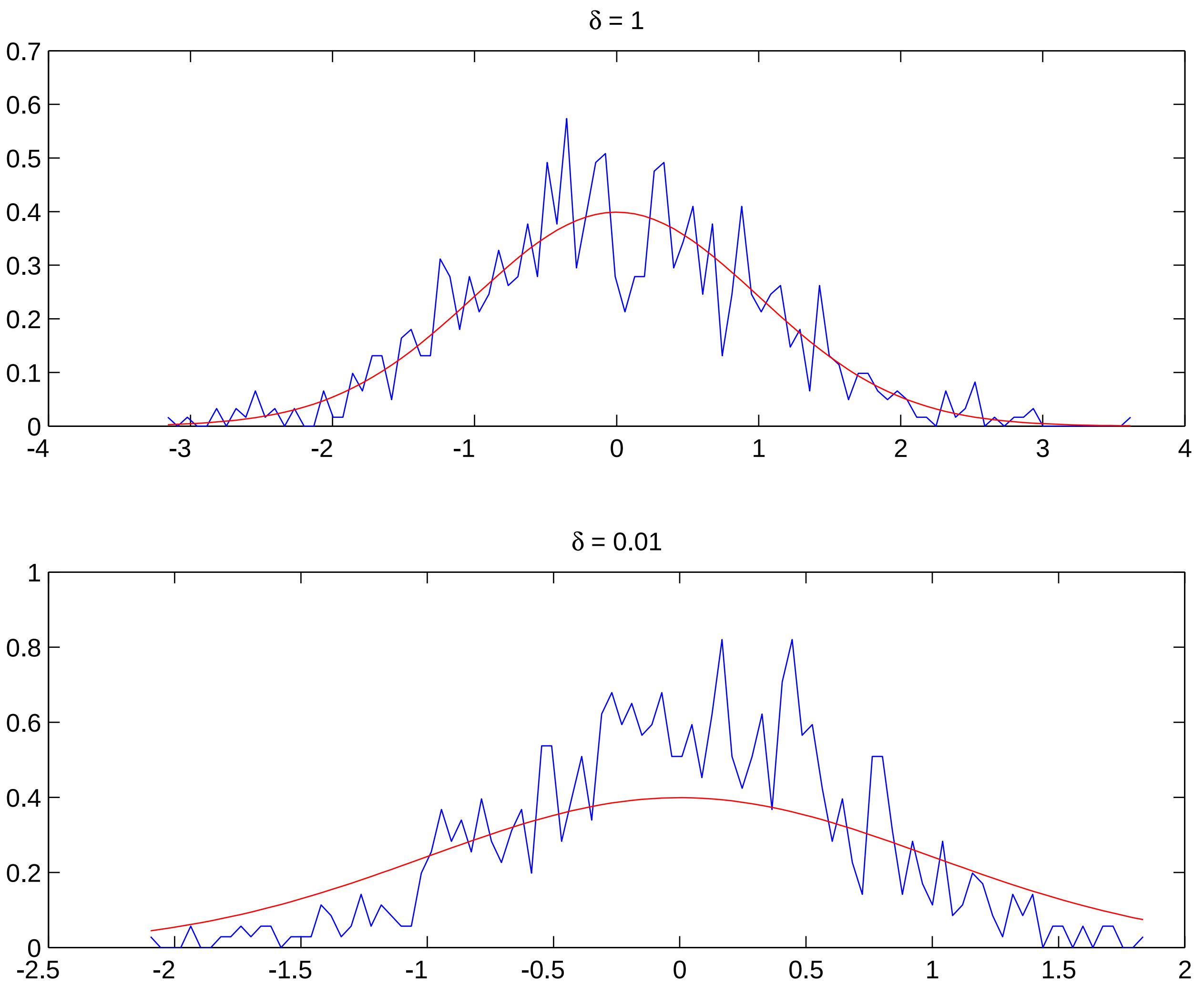
- Metropolis-Hastings - Indépendant avec $q(y) \sim \mathcal{U}[-3,+3]$
- Algo de Metropolis-Hastings - Marche Aléatoire avec $q(\epsilon_t) \sim \mathcal{U}[-\delta,+\delta]$ (Hastings, 1970)

Probabilité d'acceptation

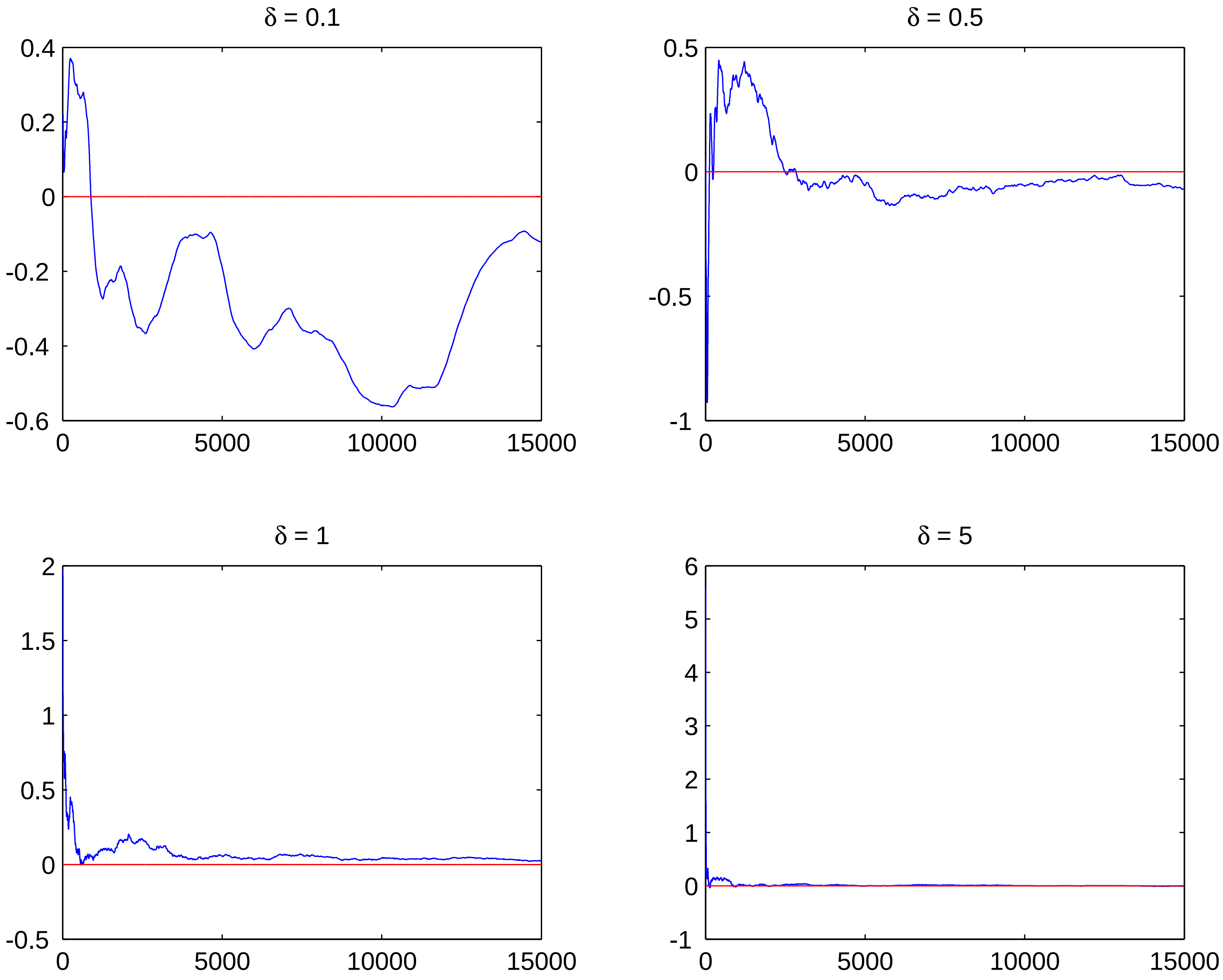
$$\min\left\{\exp\left\{(\theta_{(t)}^2 - y_t^2)/2\right\}\ ,1\right\}$$

Matlab : loi-gauss et loi-gauss-delta pour $d=1$ et $d=0.01$

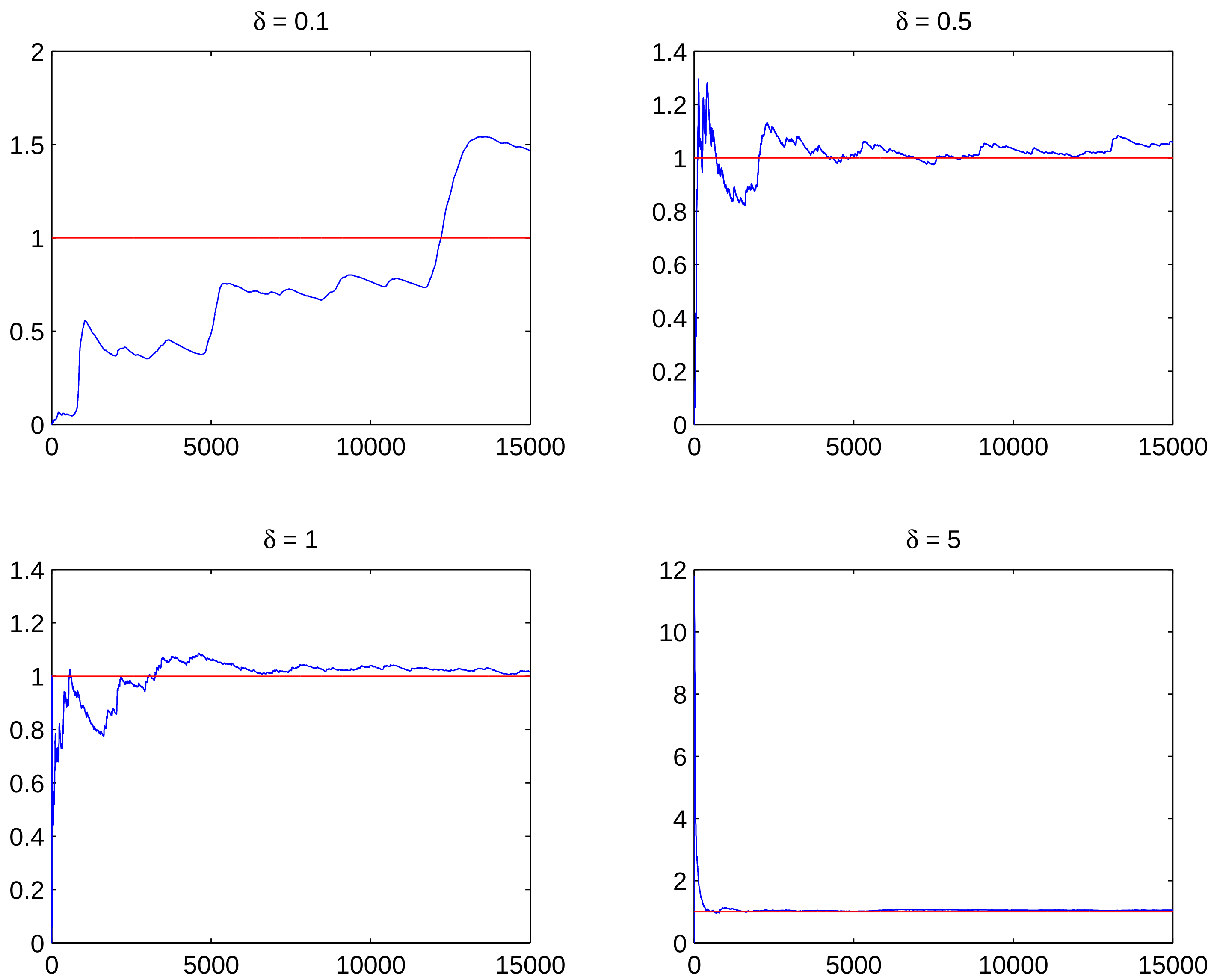
Lois cibles pour $\delta = 0.01$ et $\delta = 1$



Calcul des moyennes pour $\delta \in \{0.1, 0.5, 1, 5\}$



Calcul des variances pour $\delta \in \{0.1, 0.5, 1, 5\}$



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

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

  - $f(k)$: a priori sur le nombre de paramètres ($k \sim \mathcal{P}(\lambda)$)
  - $f(\theta^{(k)}|k)$: loi a priori sur les paramètres sachant k
  - $f(\boldsymbol{x}|k,\theta^{(k)})$: vraisemblance

- Loi a posteriori

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

$\left(k,\theta^{(k)}\right) \in \bigcup_{k\in\mathcal{K}} \mathcal{C}_k$, $\mathcal{C}_k = \{k\} \times \mathbb{R}^{n_k}$ 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 $\mathcal{M}_{k'}$

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

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 $\mathcal{M}_k$

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

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

Probabilité d'acceptation

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

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

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 $\mathbb{R}^{n_{k'}}$
- $\left| \frac{\partial\left(\theta^{(k')}, u'\right)}{\partial\left(\theta^{(k)}, u\right)} \right|$: 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 $\mathcal{M}_2$
  - $x_i \sim \mathcal{N}(\theta, 1), i = 1, ..., 100, \quad \mathcal{C}_2 = \{1\} \times \mathbb{R}$
  - 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 $\mathcal{M}_2$

$$g_{12} \left\{ \begin{array}{c} \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \\ (\theta_1, \theta_2) \mapsto (\theta = \frac{\theta_1+\theta_2}{2}, u = \frac{\theta_1-\theta_2}{2}) \end{array} \right.$$

- Probabilité d'acceptation

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

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

Difféomorphisme g_{21}

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

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

- Probabilité d'acceptation

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

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

Matlab : samplingGreen

Optimisation du taux d'acceptation

- Un algorithm générique "Adaptive rejection Metropolis sampling (ARMS)"
- choix d'une loi instrumentale q qui approche f de façon à ce que le rapport f/q soit borné, de façon à avoir l'ergodicité uniforme
- Algorithme à marche aléatoire

Dans les deux derniers cas, le choix de q est critique !

Metropolis-Hastings Indépendant

$$
\begin{aligned}
\rho &= E\left[\min\left\{\frac{f(Y)\,q(\Theta)}{f(\Theta)\,q(Y)},1\right\}\right] \\
&= 2P\left(\frac{f(Y)}{q(Y)} \geq \frac{f(\Theta)}{q(\Theta)}\right), \qquad\qquad \Theta \sim f,\ Y \sim q,
\end{aligned}
$$

Loi de proposition q paramètrée par η et on cherche η qui maximise le taux d'acceptation moyen

$$
\hat{\rho}(\eta) = \frac{2}{m} \sum_{i=1}^{m} \mathbb{I}_{\{f(y_i)q(\theta_i) > f(\theta_i)q(y_i)\}} \ ,
$$

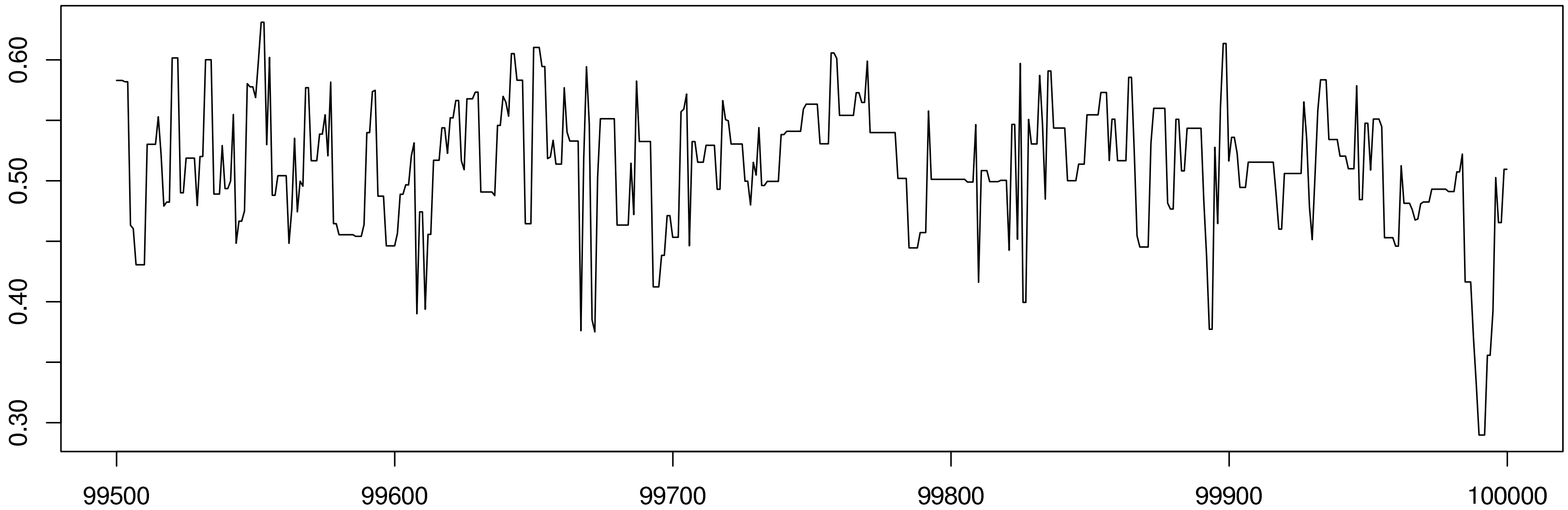
où $\theta_1, \ldots, \theta_m$ échantillon de densité f et $y_1, \ldots, y_m$ échantillon iid de densité q.

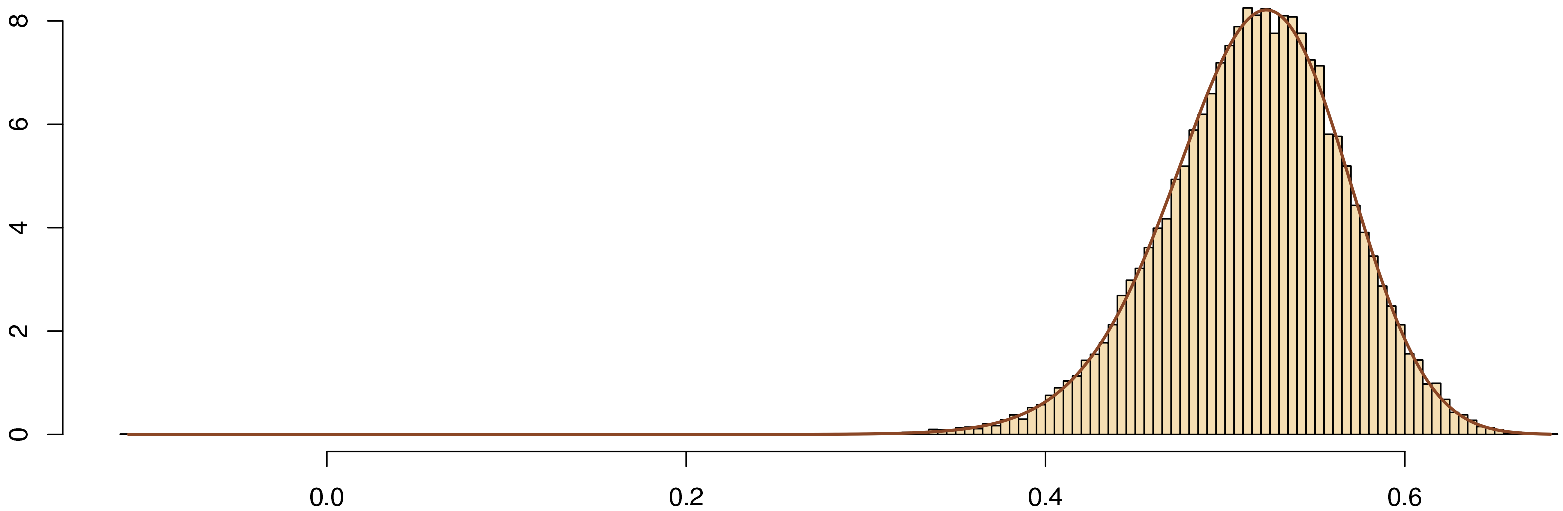
Metropolis-Hastings à marche aléatoire

- Un taux d'acceptation moyen élevé n'indique pas nécessairement que l'algorithme évolue correctement car la marche aléatoire peut évoluer trop lentement (exemple typique des densités multi-modales)
- Un taux d'acceptation moyen faible signifie que le déplacement entre y_t et $\theta^{(t)}$ est rapide
- Règle empirique (Gelman, Gilks et Robert, 1995) : taux d'acceptation de 50% pour les modèles de dimension 1 et 2, et de 25% pour les modèles de dimension supérieure

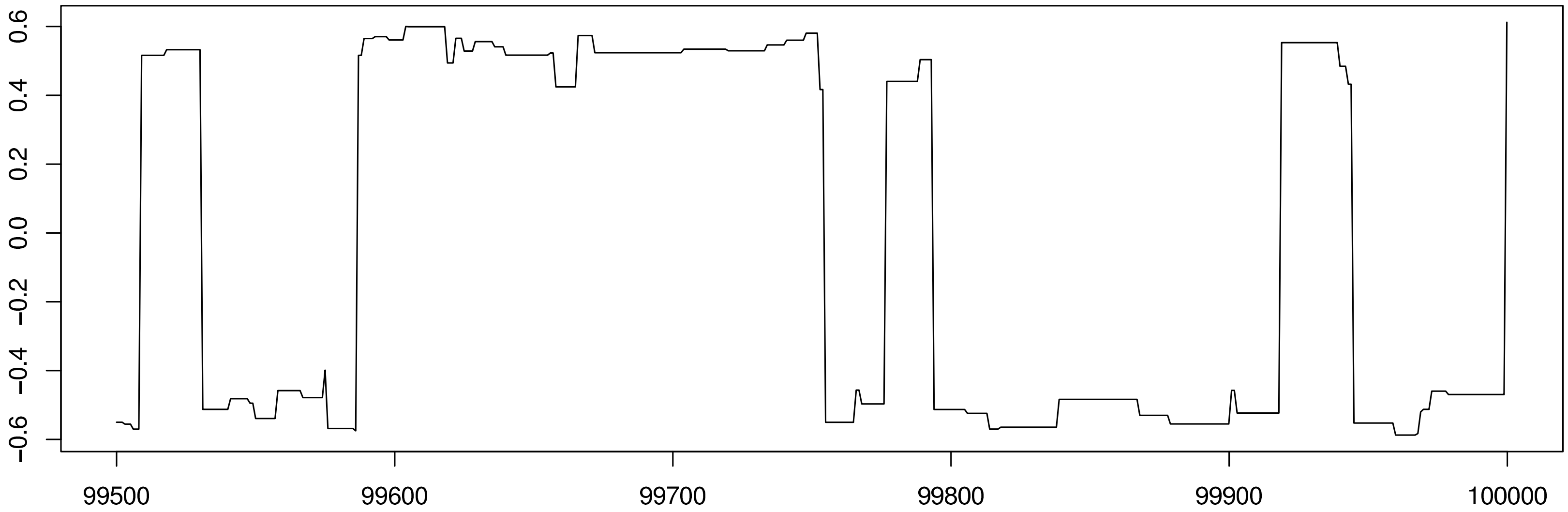
Applets Laird Breyer + exemples 2 derniers slides

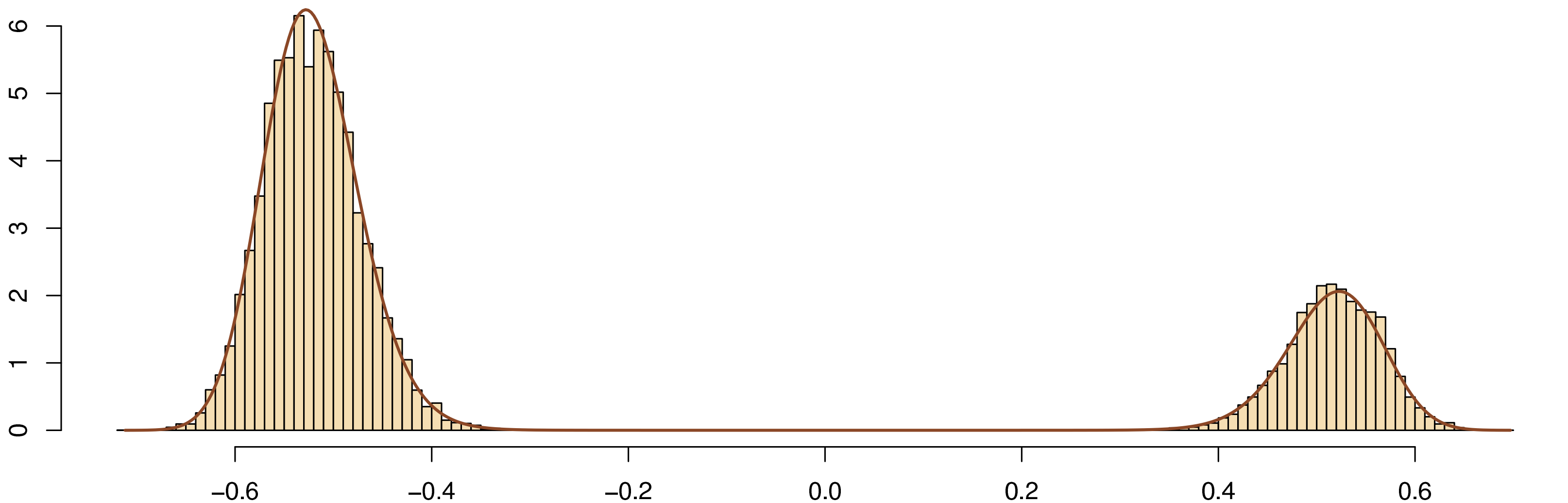
Exemple d'une loi bimodale





Exemple d'une loi bimodale

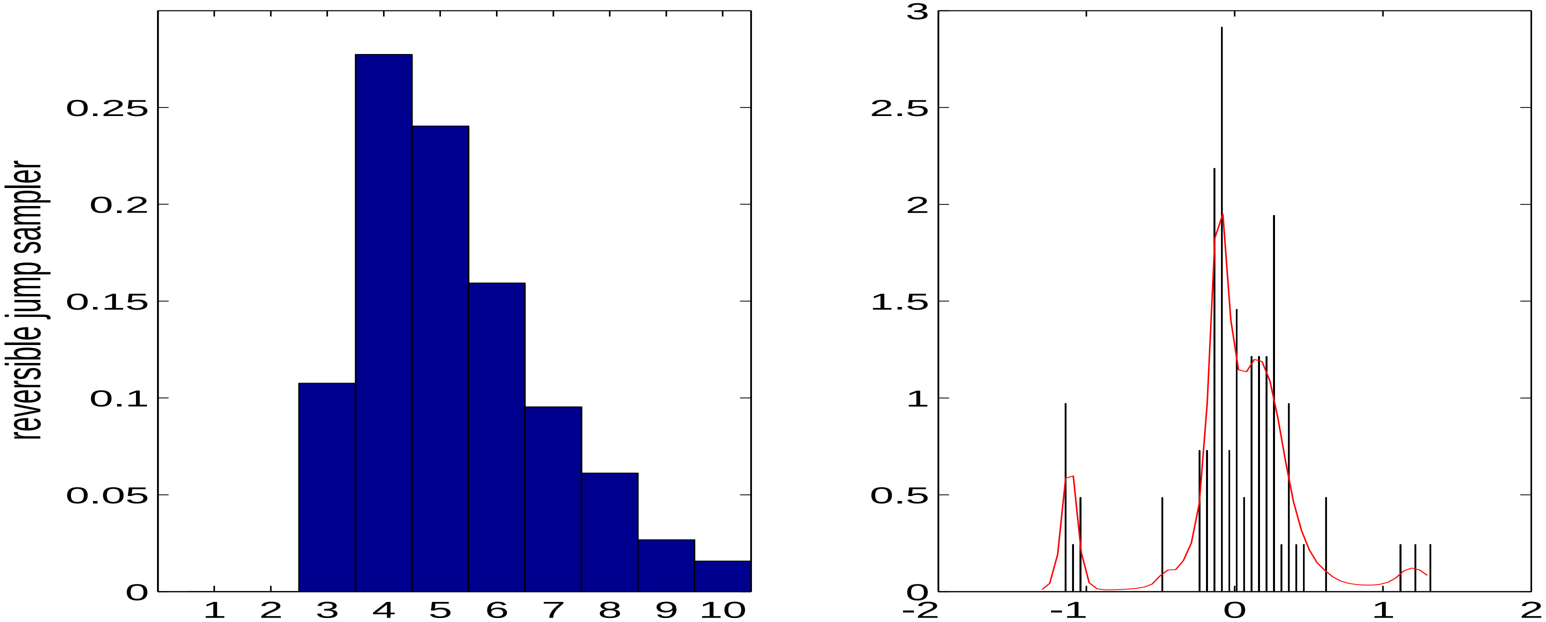




Mélange de Gaussiennes

- Modèle : $y_1, \ldots, y_n$ 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`