

# Cours 3 : L'échantillonneur de Gibbs

- 1) Principes généraux
- 2) Complétion
- 3) Convergence
- 4) Le théorème de Hammersley-Clifford
- 5) Modèles hiérarchiques
- 6) Augmentation de données
- 7) Algorithme MCMC hybride
- 8) Dangers

# 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  $\pi_0$
- 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 : $X_i \sim \mathcal{N}(m, \sigma^2)$

- **Vraisemblance**

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

- **Lois a priori**

- **Moyenne**

$$m \sim \mathcal{N}(m_0, \sigma_0^2)$$

- **Variance**

$$\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$$

# Lois conditionnelles

## • moyenne

$$m|\sigma^2, \mathbf{x} \sim \mathcal{N}(M, \Sigma^2)$$

avec

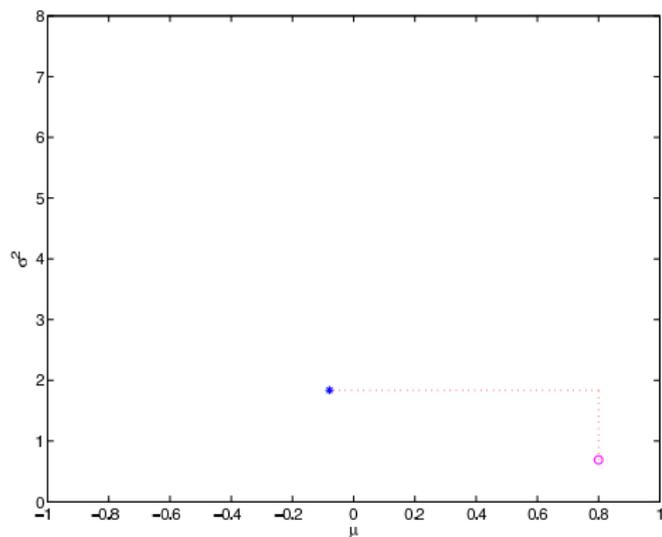
$$M = \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \left( \frac{1}{n} \sum_{i=1}^n x_i \right) + \left( \frac{\sigma^2}{\sigma^2 + n\sigma_0^2} \right) m_0 \text{ et } \Sigma^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2 + n\sigma_0^2}$$

## • variance

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

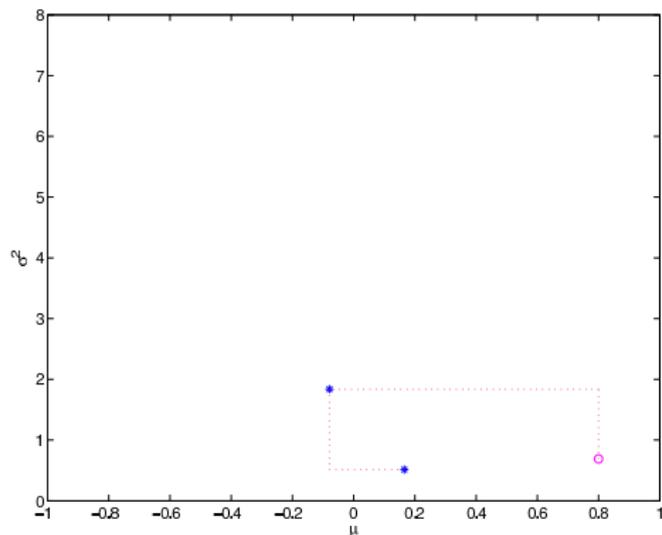
Donc, on peut simuler des couples  $(m, \sigma^2)$  avec l'échantillonneur de Gibbs

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



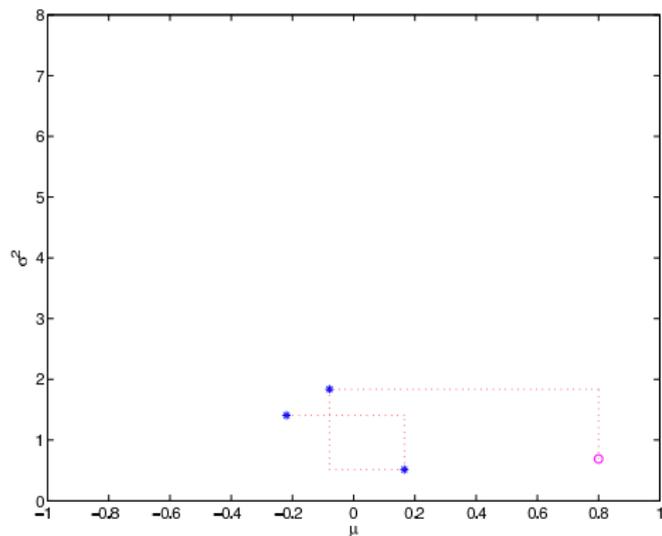
Number of Iterations 1

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



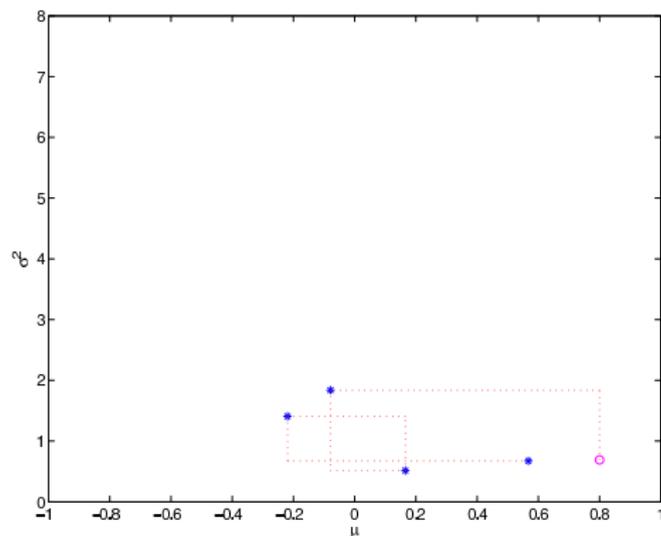
Number of Iterations 1, 2

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



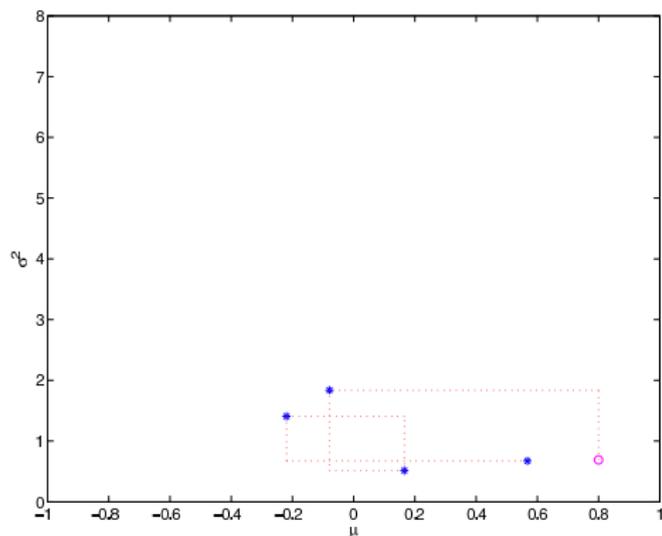
Number of Iterations 1, 2, 3

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



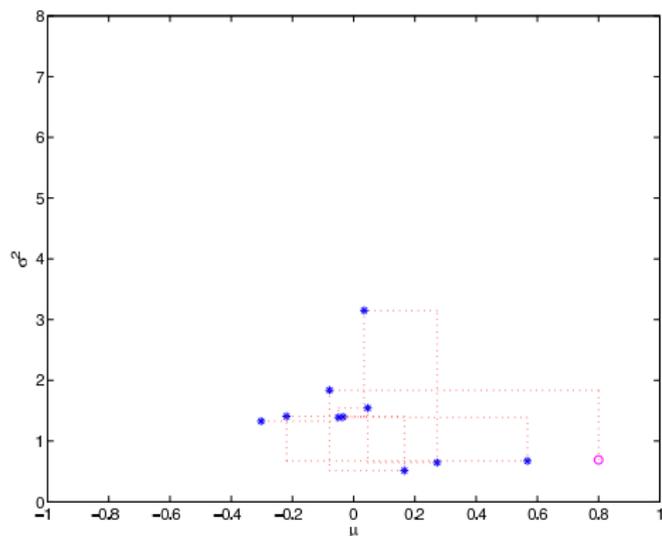
Number of Iterations 1, 2, 3, 4

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



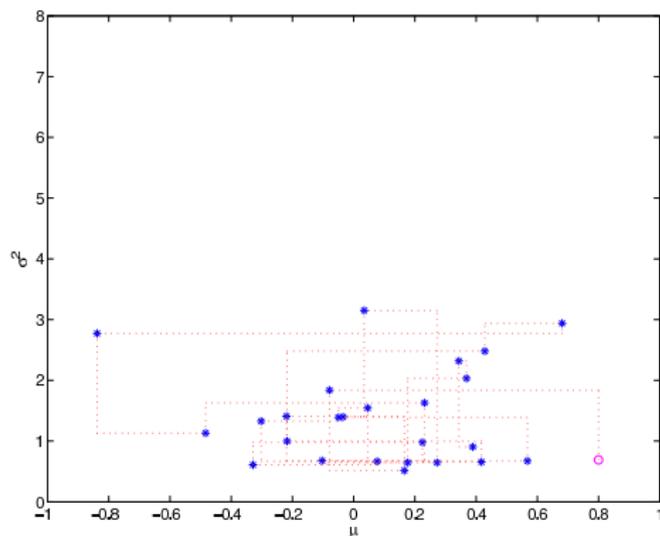
Number of Iterations 1, 2, 3, 4, 5

## Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



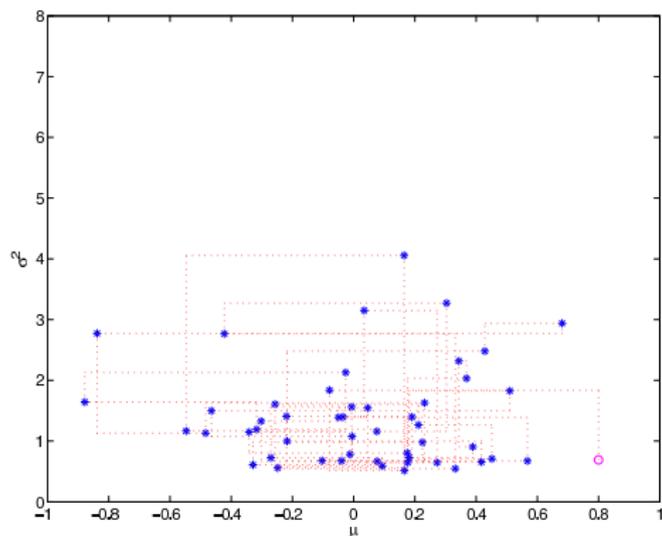
Number of Iterations 1, 2, 3, 4, 5, 10

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



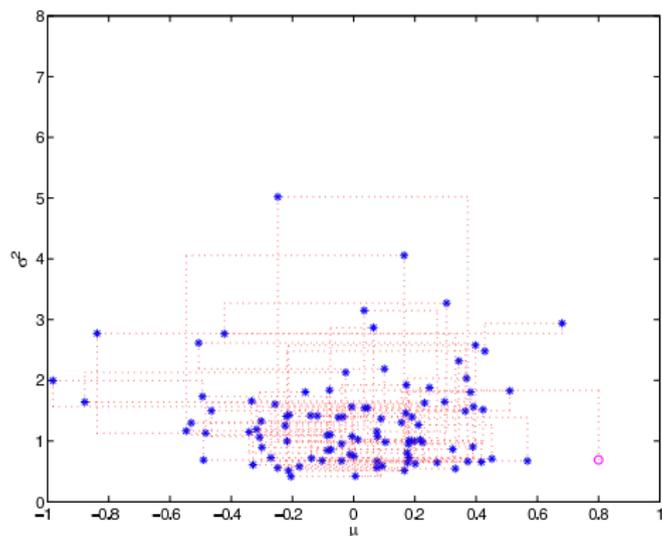
Number of Iterations 1, 2, 3, 4, 5, 10, 25

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



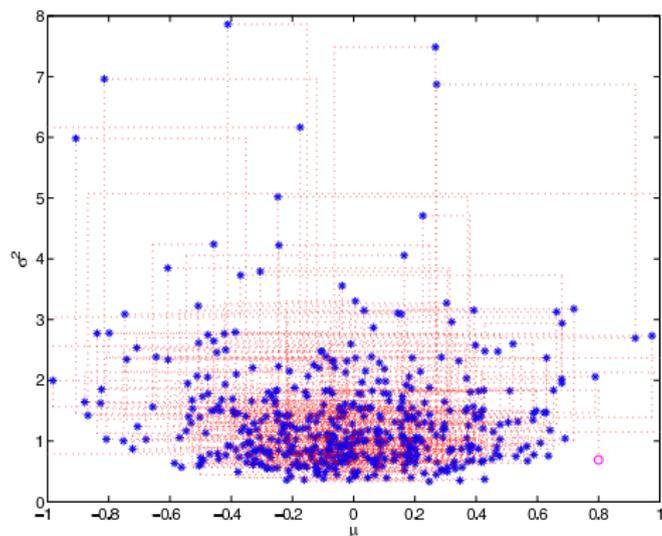
Number of Iterations 1, 2, 3, 4, 5, 10, 25, 50

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



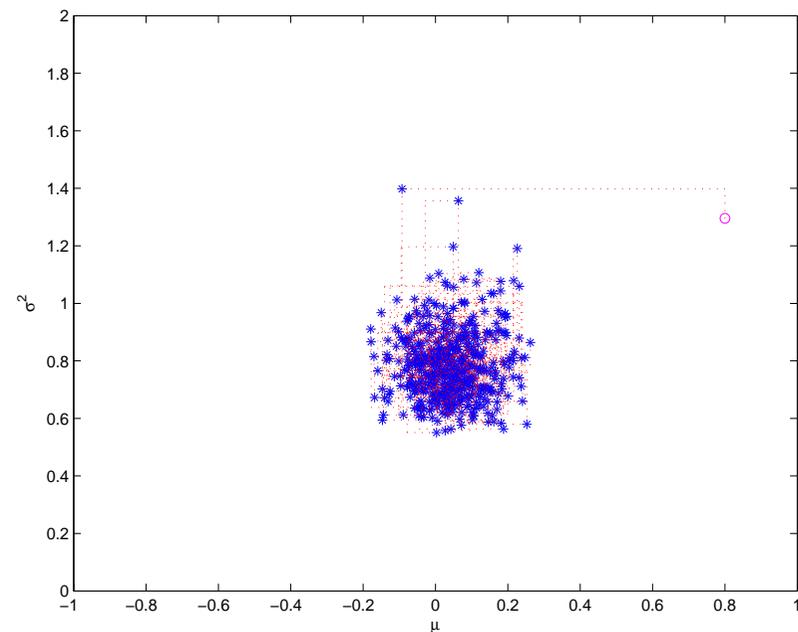
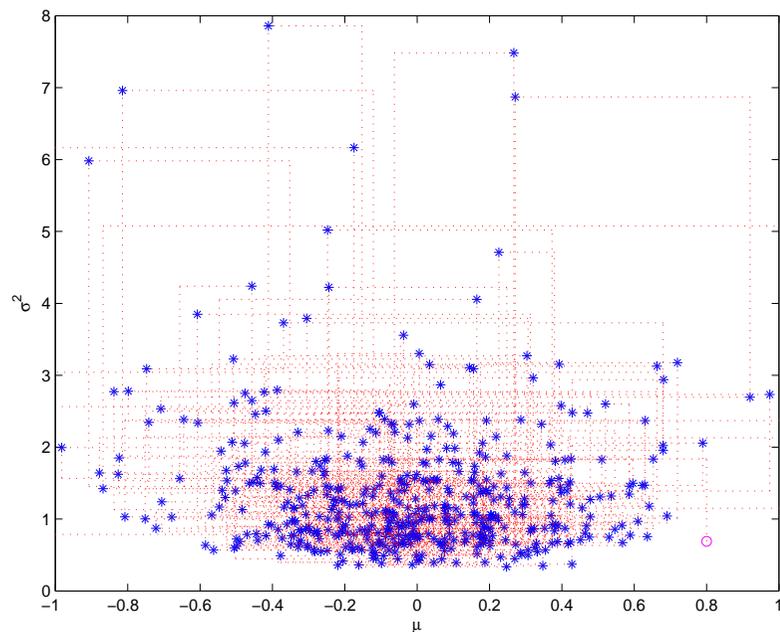
Number of Iterations 1, 2, 3, 4, 5, 10, 25, 50, 100

# Example of results with $n = 10$ observations from the $N(0, 1)$ distribution



Number of Iterations 1, 2, 3, 4, 5, 10, 25, 50, 100, 500

# Example of Results with, Left $n = 10$ Observations; Right, $n = 100$ Observations from the $N(0, 1)$ Distribution



# 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

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

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

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

# Condition de positivité

- **Positivité**

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

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.

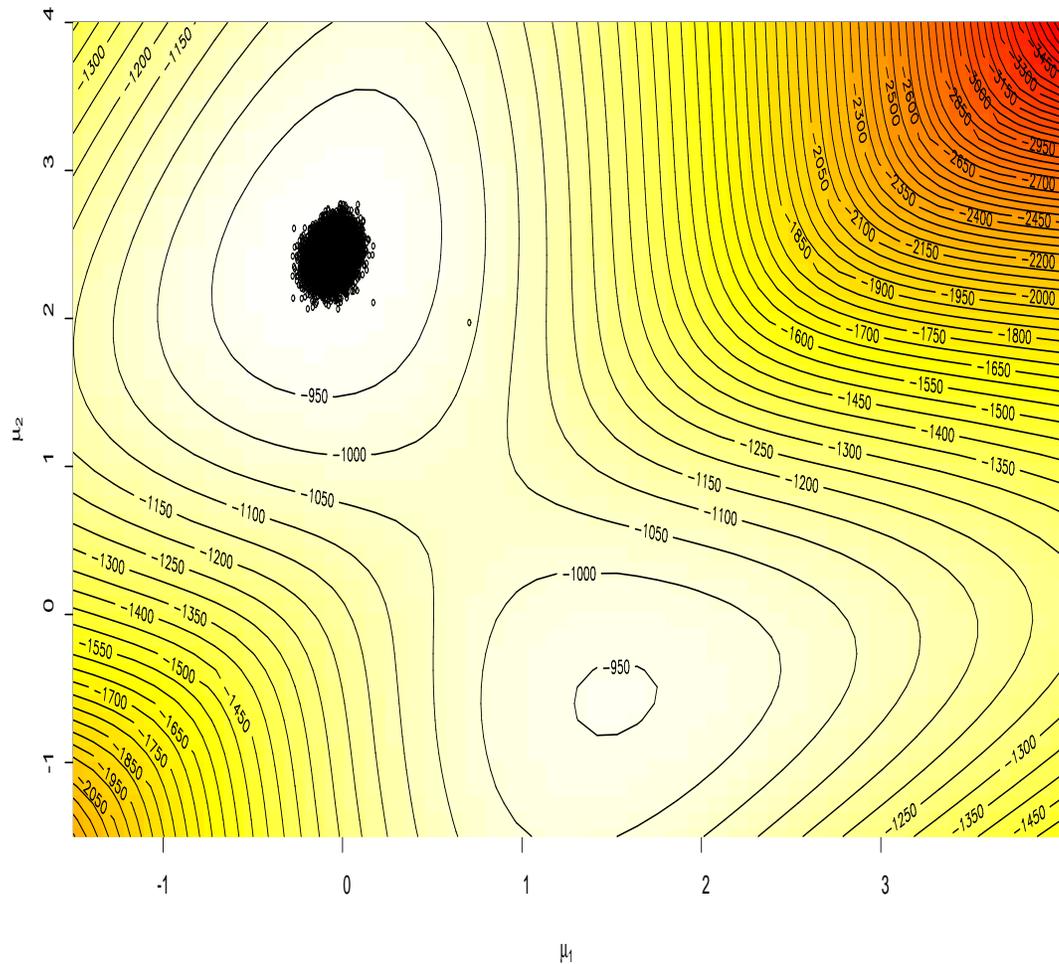
- **Contre-exemple**

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

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

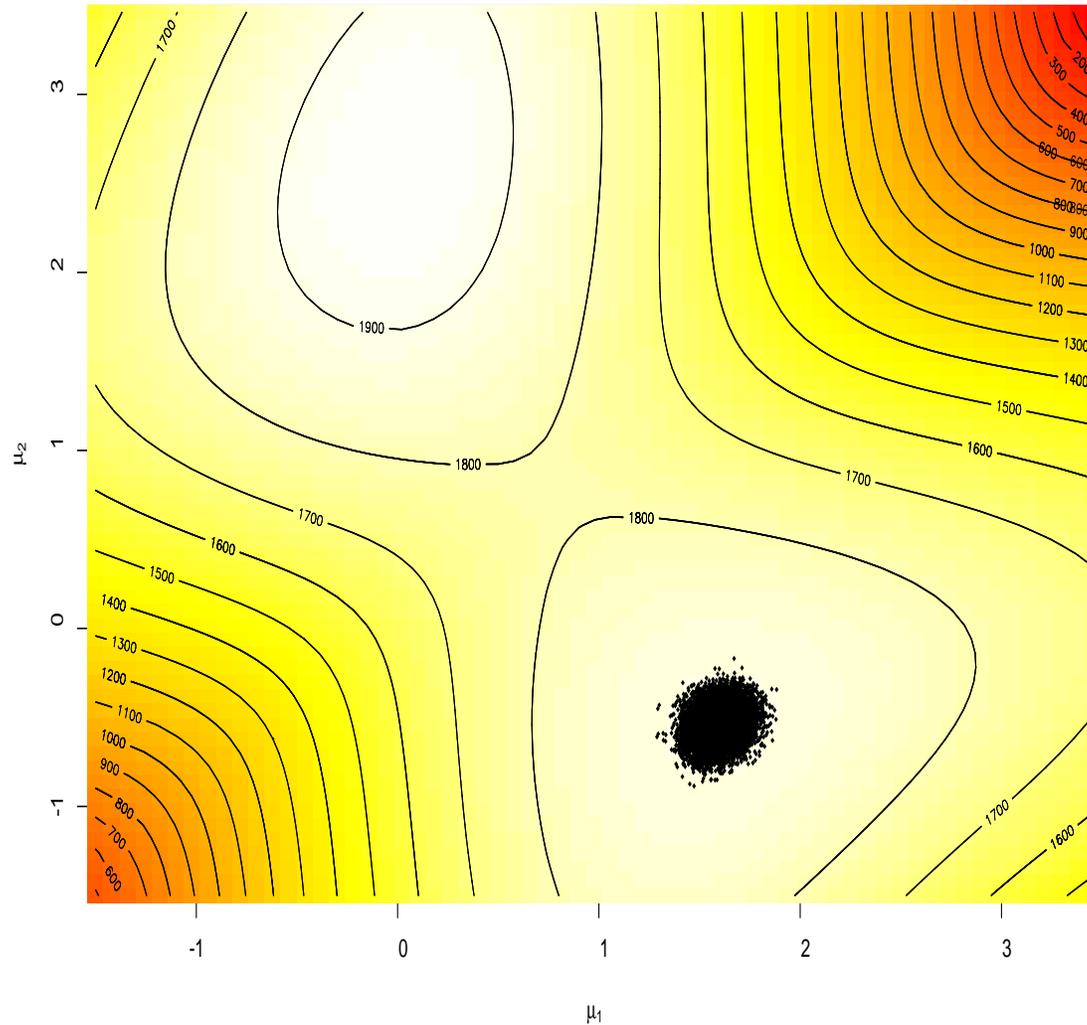
# Illustration de la non-positivité

## Initialisation Aléatoire



# Illustration de la non-positivité

Gibbs coïncé 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  $\mu$ .

# 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**

# Le théorème de Hammersley-Clifford

Une loi jointe est **caractérisée** par l'ensemble de ses lois conditionnelles.

- **Dimension 2**

Si la densité jointe  $g(y_1, y_2)$  a des lois conditionnelles notées  $g_1(y_1|y_2)$  et  $g_2(y_2|y_1)$ , alors (Hammersley and Clifford, 1970)

$$g(y_1, y_2) = \frac{g_2(y_2|y_1)}{\int g_2(v|y_1)/g_1(y_1|v) dv}.$$

# Généralisation

- Sous l'hypothèse de **positivité**, une loi jointe  $g$  peut s'écrire

$$g(y_1, \dots, y_p) \propto \prod_{j=1}^p \frac{g_{l_j}(y_{l_j} | y_{l_1}, \dots, y_{l_{j-1}}, y'_{l_{j+1}}, \dots, y'_{l_p})}{g_{l_j}(y'_{l_j} | y_{l_1}, \dots, y_{l_{j-1}}, y'_{l_{j+1}}, \dots, y'_{l_p})}$$

pour toute permutation  $l$  définie sur  $\{1, \dots, p\}$  et tout  $y' \in \mathcal{Y}$ .

# 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  $\lambda_i$

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

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

- pour  $\beta$

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

# Données Poissonniennes cachées

Observations	0	1	2	3	4 ou plus
Nombre	139	128	55	25	13

- **Données** : observations du nombre de données égales à 0, 1, 2, 3 et du nombre de données  $\geq 4$ .
- **Vraisemblance**

$$\ell(x_1, \dots, x_5; \lambda) \propto e^{-347\lambda} \lambda^{128+55 \times 2+25 \times 3} \left( 1 - e^{-\lambda} \sum_{i=0}^3 \frac{\lambda^i}{i!} \right)^{13},$$

- **Idée** : on munit  $\lambda$  d'une loi a priori  $\pi(\lambda) = 1/\lambda$  et on complète ce paramètre par  $\mathbf{y} = (y_1, \dots, y_{13})$ .

## Loi a posteriori

$$\ell(\lambda, y_1, \dots, y_{13} | x_1, \dots, x_5) \propto e^{-347\lambda} \lambda^{128+55 \times 2+25 \times 3} \left( \prod_{i=1}^{13} \lambda^{y_i} e^{-\lambda} \right) \frac{1}{\lambda},$$

## Lois conditionnelles

$$y_i | \lambda \sim \mathcal{P}(\lambda) \mathbb{I}_{y_i \geq 4}, i = 1, \dots, 13$$

$$\lambda | \mathbf{y} \sim \mathcal{Ga} \left( 313 + \sum_{i=1}^{13} y_i, 360 \right)$$

## Estimateur de $\lambda$

$$\hat{\lambda} = \frac{1}{360T} \sum_{t=1}^T \left( 313 + \sum_{i=1}^{13} y_i^{(t)} \right)$$

# Conditionnement - Rao-Blackwellization

- **Espérances conditionnelles**

$$E[h(\Lambda)] = E[E[h(\Lambda)|\mathbf{Y}]]$$

- **Estimateurs**

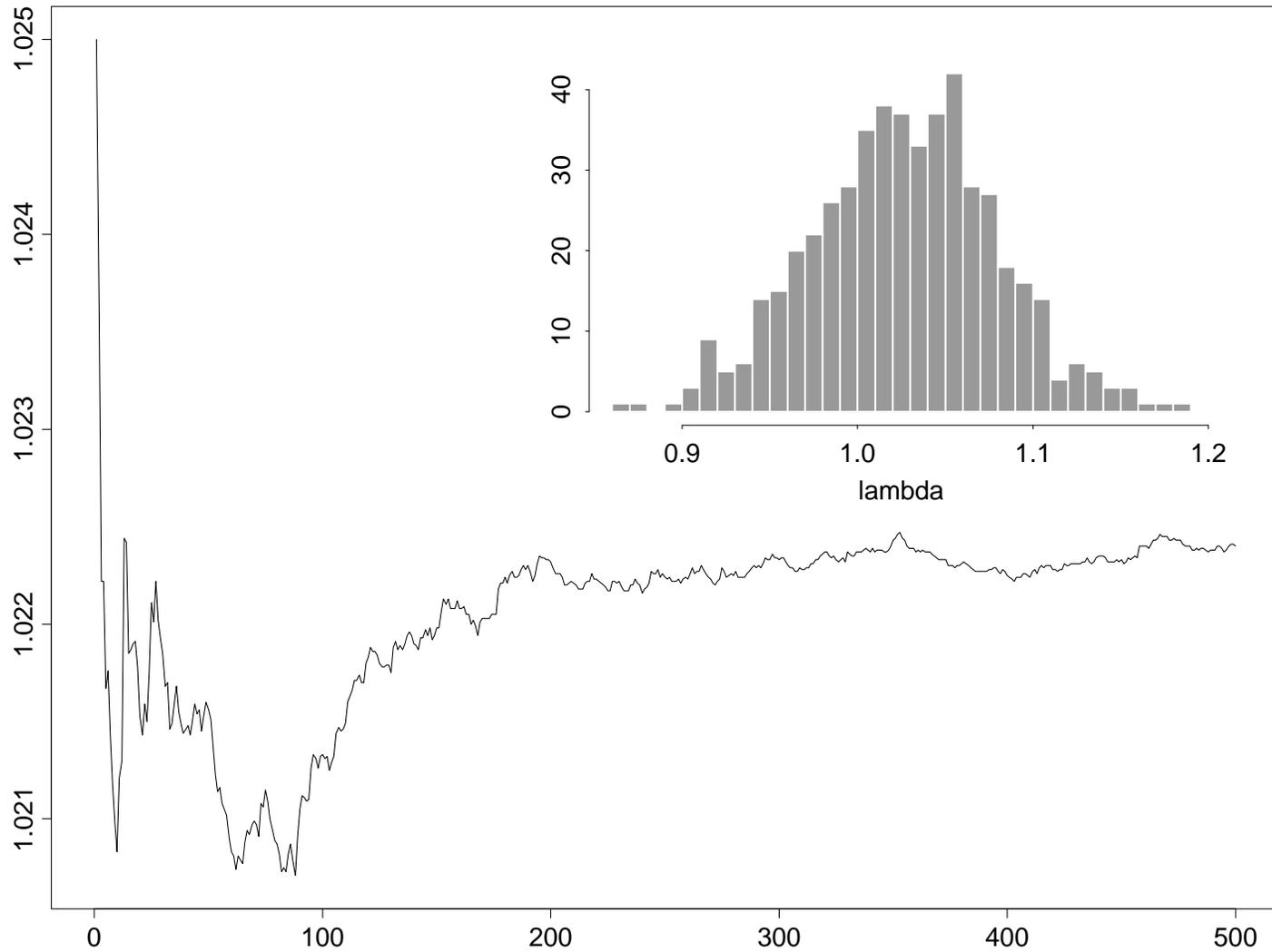
Ici on sait calculer  $g(\mathbf{Y}) = E[h(\Lambda)|\mathbf{Y}]$ . On en déduit deux estimateurs

$$\hat{I}_1 = \frac{1}{T} \sum_{t=1}^T h(\Lambda_t)$$

$$\hat{I}_2 = \frac{1}{T} \sum_{t=1}^T g(\mathbf{Y}_t) = \frac{1}{T} \sum_{t=1}^T E[h(\Lambda)|\mathbf{Y}_t]$$

**Réduction de variance**

# Résultats de simulation



# 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`

# Dangers

- **Modèle à effets aléatoires**

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, I, \quad j = 1, \dots, J,$$

avec

$$\alpha_i \sim \mathcal{N}(0, \sigma^2) \text{ et } \varepsilon_{ij} \sim \mathcal{N}(0, \tau^2),$$

- **Lois a priori**

La loi a priori de Jeffreys (impropre) pour les paramètres  $\mu$ ,  $\sigma$  et  $\tau$  est

$$\pi(\mu, \sigma^2, \tau^2) = \frac{1}{\sigma^2 \tau^2} .$$

# Lois conditionnelles

Les lois conditionnelles sont définies par

$$\alpha_i | y, \mu, \sigma^2, \tau^2 \sim \mathcal{N} \left( \frac{J(\bar{y}_i - \mu)}{J + \tau^2 \sigma^{-2}}, (J\tau^{-2} + \sigma^{-2})^{-1} \right),$$

$$\mu | \alpha, y, \sigma^2, \tau^2 \sim \mathcal{N}(\bar{y} - \bar{\alpha}, \tau^2 / JI),$$

$$\sigma^2 | \alpha, \mu, y, \tau^2 \sim \mathcal{IG} \left( \frac{I}{2}, \frac{1}{2} \sum_i \alpha_i^2 \right),$$

$$\tau^2 | \alpha, \mu, y, \sigma^2 \sim \mathcal{IG} \left( \frac{IJ}{2}, \frac{1}{2} \sum_{i,j} (y_{ij} - \alpha_i - \mu)^2 \right),$$

et sont faciles à simuler. Mais la loi jointe n'existe pas !

# Simulations

Évolution de  $\mu^{(t)}$  et histogramme pour 1000 itérations

