

Méthodes de Monte Carlo par chaîne de Markov

Pierre Gloaguen

Avril 2020

Rappels des cours précédents

- ▶ Méthodes de Monte Carlo pour le calcul d'espérances
- ▶ Approche par simulation de variables aléatoires i.i.d.
- ▶ Méthodes de simulation de loi (échantillons i.i.d.)
- ▶ Inférence bayésienne, technique nécessitant des algos de simulations de lois

Modèle probit

On veut simuler selon une loi $\pi(\theta|y_{1:n}, \mathbf{x}_{1:n})$ telle que:

$$\pi(\theta|y_{1:n}, \mathbf{x}_{1:n}) \propto e^{-\frac{1}{8}\theta^T\theta} \prod_{k=1}^n \phi(\mathbf{x}_k^T\theta)^{y_k} (1 - \phi(\mathbf{x}_k^T\theta))^{1-y_k}$$

- ▶ Possible par acceptation rejet si n n'est pas trop grand;
- ▶ Ensuite, ne fonctionne plus **en pratique** (probabilité d'acceptation devient trop faible).
- ▶ Nécessité de définir un autre algorithme.

Objectif du cours

- ▶ Présentation des méthodes de Monte Carlo par chaîne de Markov
- ▶ Rappel sur les chaînes de Markov (définitions)
- ▶ Théorème ergodique
- ▶ Algorithme de Metropolis Hastings
- ▶ Algorithme de Gibbs

Rappel sur les chaînes de Markov

Chaîne de Markov (à espace d'états fini)

Soit X_0 une variable aléatoire sur $\{1, \dots, K\}$ de loi π_0 .

- ▶ La suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans $\mathcal{K} = \{1, \dots, K\}$ est une chaîne de Markov si pour tout $n \geq 1$ est pour tout suite (k_0, \dots, k_n) d'éléments de \mathcal{K} , on a :

$$\mathbb{P}(X_n = k_n | X_0 = k_0, \dots, X_{n-1} = k_{n-1}) = \mathbb{P}(X_n = k_n | X_{n-1} = k_{n-1})$$

- ▶ Cette chaîne est *homogène* si, pour (i, j) dans $\mathcal{K} \times \mathcal{K}$:
 $\mathbb{P}(X_n = j | X_{n-1} = i) = \mathbb{P}(X_1 = j | X_0 = i) = P_{ij}$
- ▶ La matrice $P = (P_{ij})$ est la **matrice de transition** de la chaîne de Markov.
- ▶ Une chaîne de Markov homogène est entièrement caractérisée par π_0 et P .

Loi de la chaîne

Pour $n \geq 0$, on note π_n , la loi de l'état X_n , c'est à dire le vecteur ligne

$$\pi_n = (\pi_{n,1} = \mathbb{P}(X_n = 1), \dots, \pi_{n,K} = \mathbb{P}(X_n = K)).$$

On a:

- ▶ $\mathbb{P}(X_1 = j) = \sum_{i=1}^k \mathbb{P}(X_0 = i) \times \mathbb{P}(X_1 = j | X_0 = i) = \sum_{i=1}^k \pi_{0,i} P_{ij}$ Cette relation est résumée par l'équation $\pi_1 = \pi_0 P$
- ▶ Par récurrence, on montre que

$$P_{ij}^{(n)} := \mathbb{P}(X_n = j | X_0 = i) = (P^n)_{ij}$$

où P^n est la puissance n -ième de la matrice P .

- ▶ Ainsi:

$$\pi_n = \pi_0 P^n$$

Mesure invariante pour P

Soit π un vecteur (ligne) de probabilité sur \mathcal{K} .

- ▶ π est une **mesure invariante pour la chaîne de Markov de transition P** si:

$$\pi P = \pi$$

Mesure invariante pour P

Soit π un vecteur (ligne) de probabilité sur \mathcal{K} .

- ▶ π est une **mesure invariante pour la chaîne de Markov de transition P** si:

$$\pi P = \pi$$

- ▶ Si π_0 est une mesure invariante pour P , alors, pour tout n , $\pi_n = \pi_0$.
- ▶ Dans ce cas, les V.A. X_0, \dots, X_n sont identiquement distribuées (mais pas indépendantes!).

Irréductibilité

Une chaîne de Markov homogène sur \mathcal{K} , de transition P est **irréductible** si

$$\forall i, j \in \mathcal{K} \times \mathcal{K}, \exists n \text{ tel que } P_{i,j}^{(n)} > 0$$

- Pour deux états de la chaîne, il est possible d'accéder de l'un à l'autre en un temps fini.

Apériodicité

Soit $(X_n)_{n \geq 1}$ une chaîne de Markov homogène sur \mathcal{K} . Pour $k \in \mathcal{K}$,

- ▶ La *période* de l'état k , notée $d(k)$, est le P.G.C.D. de tous les entiers n tels que $P_{kk}^{(n)} > 0$ (avec la convention $\text{pgcd}(\emptyset) = +\infty$):

$$d(j) = \text{pgcd} \left\{ n \geq 1, P_{kk}^{(n)} > 0 \right\}$$

Une chaîne est dite apériodique si pour tout k dans \mathcal{K} , $d(k) = 1$.

Apériodicité

Soit $(X_n)_{n \geq 1}$ une chaîne de Markov homogène sur \mathcal{K} . Pour $k \in \mathcal{K}$,

- La *période* de l'état k , notée $d(k)$, est le P.G.C.D. de tous les entiers n tels que $P_{kk}^{(n)} > 0$ (avec la convention $\text{pgcd}(\emptyset) = +\infty$):

$$d(j) = \text{pgcd} \left\{ n \geq 1, P_{kk}^{(n)} > 0 \right\}$$

Une chaîne est dite apériodique si pour tout k dans \mathcal{K} , $d(k) = 1$.

Pour une chaîne irréductible, une condition suffisante pour être apériodique est qu'il existe un $k \in \mathcal{K}$ tel que $P_{kk} > 0$.

Théorème ergodique

Théorème ergodique

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur \mathcal{K} de loi initiale π_0 et de matrice de transition P . On suppose que cette chaîne est irréductible et apériodique. Alors:

Théorème ergodique

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur \mathcal{K} de loi initiale π_0 et de matrice de transition P . On suppose que cette chaîne est irréductible et apériodique. Alors:

1. Cette chaîne de Markov admet une unique mesure de probabilité invariante π .

Théorème ergodique

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur \mathcal{K} de loi initiale π_0 et de matrice de transition P . On suppose que cette chaîne est irréductible et apériodique. Alors:

1. Cette chaîne de Markov admet une unique mesure de probabilité invariante π .
2. $X_n \xrightarrow{\text{loi}} X$ où X est une v.a. de loi π .

Théorème ergodique

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur \mathcal{K} de loi initiale π_0 et de matrice de transition P . On suppose que cette chaîne est irréductible et apériodique. Alors:

1. Cette chaîne de Markov admet une unique mesure de probabilité invariante π .
2. $X_n \xrightarrow{\text{loi}} X$ où X est une v.a. de loi π .
3. Pour toute fonction φ intégrable par rapport à π , on a :

$$\frac{1}{M+1} \sum_{k=0}^M \varphi(X_k) \xrightarrow[M \rightarrow +\infty]{p.s.} \mathbb{E}_\pi[\varphi(X)].$$

Théorème ergodique

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur \mathcal{K} de loi initiale π_0 et de matrice de transition P . On suppose que cette chaîne est irréductible et apériodique. Alors:

1. Cette chaîne de Markov admet une unique mesure de probabilité invariante π .
2. $X_n \xrightarrow{\text{loi}} X$ où X est une v.a. de loi π .
3. Pour toute fonction φ intégrable par rapport à π , on a :

$$\frac{1}{M+1} \sum_{k=0}^M \varphi(X_k) \xrightarrow[M \rightarrow +\infty]{p.s.} \mathbb{E}_\pi[\varphi(X)].$$

4. Si $\varphi(X)$ admet un moment d'ordre supérieur à 2, on a

$$\sqrt{M} \left(\frac{1}{M+1} \sum_{k=0}^n \varphi(X_k) - \mathbb{E}_\pi[\varphi(X)] \right) \xrightarrow[M \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma^2)$$

Théorème ergodique

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur \mathcal{K} de loi initiale π_0 et de matrice de transition P . On suppose que cette chaîne est irréductible et apériodique. Alors:

1. Cette chaîne de Markov admet une unique mesure de probabilité invariante π .
2. $X_n \xrightarrow{\text{loi}} X$ où X est une v.a. de loi π .
3. Pour toute fonction φ intégrable par rapport à π , on a :

$$\frac{1}{M+1} \sum_{k=0}^M \varphi(X_k) \xrightarrow[M \rightarrow +\infty]{p.s.} \mathbb{E}_\pi[\varphi(X)].$$

4. Si $\varphi(X)$ admet un moment d'ordre supérieur à 2, on a

$$\sqrt{M} \left(\frac{1}{M+1} \sum_{k=0}^n \varphi(X_k) - \mathbb{E}_\pi[\varphi(X)] \right) \xrightarrow[M \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0, \sigma^2)$$

Une propriété analogue reste vraie quand la chaîne de Markov est à valeurs dans un ensemble continu (typiquement, \mathbb{R}^d).

Conséquence et intérêt pratique du théorème ergodique

- Pour estimer $\mathbb{E}_\pi[\varphi(X)]$, il suffit d'être capable de simuler une chaîne de Markov apériodique et irréductible de mesure de probabilité invariante π .

Conséquence et intérêt pratique du théorème ergodique

- ▶ Pour estimer $\mathbb{E}_\pi[\varphi(X)]$, il suffit d'être capable de simuler une chaîne de Markov apériodique et irréductible de mesure de probabilité invariante π .
- ▶ Il n'est pas nécessaire de savoir tirer selon π directement!

Conséquence et intérêt pratique du théorème ergodique

- ▶ Pour estimer $\mathbb{E}_\pi[\varphi(X)]$, il suffit d'être capable de simuler une chaîne de Markov apériodique et irréductible de mesure de probabilité invariante π .
- ▶ Il n'est pas nécessaire de savoir tirer selon π directement!
- ▶ Le point 2. dit qu'*au bout d'un certain temps*, les X_n simulés pourront être considérés comme de loi π (mais pas indépendants!)

Conséquence et intérêt pratique du théorème ergodique

- ▶ Pour estimer $\mathbb{E}_\pi[\varphi(X)]$, il suffit d'être capable de simuler une chaîne de Markov apériodique et irréductible de mesure de probabilité invariante π .
- ▶ Il n'est pas nécessaire de savoir tirer selon π directement!
- ▶ Le point 2. dit qu'*au bout d'un certain temps*, les X_n simulés pourront être considérés comme de loi π (mais pas indépendants!)
- ▶ Encore faut il être capable de construire une chaîne de Markov apériodique, irréductible, de loi invariante donnée par π !

Conséquence et intérêt pratique du théorème ergodique

- ▶ Pour estimer $\mathbb{E}_\pi[\varphi(X)]$, il suffit d'être capable de simuler une chaîne de Markov apériodique et irréductible de mesure de probabilité invariante π .
- ▶ Il n'est pas nécessaire de savoir tirer selon π directement!
- ▶ Le point 2. dit qu'*au bout d'un certain temps*, les X_n simulés pourront être considérés comme de loi π (mais pas indépendants!)
- ▶ Encore faut il être capable de construire une chaîne de Markov apériodique, irréductible, de loi invariante donnée par π !
- ▶ \longrightarrow Algorithme de Metropolis Hastings

Remarque sur le Théorème Central Limite

- ▶ Les V.A. dans l'estimateur Monte Carlo ne sont plus indépendantes.
- ▶ \Rightarrow la variance σ^2 n'est absolument pas triviale (il ne s'agit pas de $\mathbb{V}[\varphi(X)]$)!
- ▶ Pas nécessairement facile à estimer!
- ▶ Ainsi, avoir un IC asymptotique sur $\mathbb{E}_\pi[\varphi(X)]$ n'est plus du tout immédiat.

Algorithme de Metropolis Hastings

Réversibilité

Soit $\pi = (\pi_1, \dots, \pi_K)$ une mesure de probabilité sur \mathcal{K} et $(X_n)_{n \geq 0}$ une chaîne de Markov homogène de matrice de transition P et de loi initiale π_0 .

Réversibilité

Soit $\pi = (\pi_1, \dots, \pi_K)$ une mesure de probabilité sur \mathcal{K} et $(X_n)_{n \geq 0}$ une chaîne de Markov homogène de matrice de transition P et de loi initiale π_0 .

- ▶ π est **réversible** pour P si elle vérifie la condition d'équilibre:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, \pi_i \times P_{ij} = \pi_j \times P_{ji}$$

- ▶ *Propriété:* Si π est réversible pour une chaîne de Markov de transition P , alors, π est une mesure de probabilité invariante pour P .

Réversibilité

Soit $\pi = (\pi_1, \dots, \pi_K)$ une mesure de probabilité sur \mathcal{K} et $(X_n)_{n \geq 0}$ une chaîne de Markov homogène de matrice de transition P et de loi initiale π_0 .

- ▶ π est **réversible** pour P si elle vérifie la condition d'équilibre:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, \pi_i \times P_{ij} = \pi_j \times P_{ji}$$

- ▶ *Propriété:* Si π est réversible pour une chaîne de Markov de transition P , alors, π est une mesure de probabilité invariante pour P .
- ▶ *Preuve* Soit π une mesure de probabilité réversible pour P . On a tout de suite que

$$\begin{aligned} \forall j \in \mathcal{K} \quad (\pi P)_j &= \sum_{i=1}^K \pi_i P_{ij} \\ &= \sum_{i=1}^K \pi_j P_{ji} && \text{par réversibilité} \\ &= \pi_j && \text{par propriété de } P \\ \Rightarrow \pi P &= \pi \end{aligned}$$

Objectif de l'algorithme

- ▶ On veut simuler selon la loi π .
- ▶ Construire une chaîne de Markov irréductible et apériodique, de loi initiale π_0 et de transition P , **réversible pour P**
- ▶ On va se servir pour ça d'une chaîne de Markov de transition Q (réversible et apériodique) **parcourant le même espace que P** (le support de π).

Algorithme de Metropolis Hastings (formulation discrete)

- Soit Q une matrice stochastique $K \times K$ satisfaisant la condition suivante:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, Q_{ij} > 0 \Leftrightarrow Q_{ji} > 0$$

- Soit $(X_n)_{n \geq 0}$ la suite de variables aléatoires construite ainsi:

Algorithme de Metropolis Hastings (formulation discrete)

- Soit Q une matrice stochastique $K \times K$ satisfaisant la condition suivante:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, Q_{ij} > 0 \Leftrightarrow Q_{ji} > 0$$

- Soit $(X_n)_{n \geq 0}$ la suite de variables aléatoires construite ainsi:

1. On simule X_0 selon π_0 .

Algorithme de Metropolis Hastings (formulation discrete)

- Soit Q une matrice stochastique $K \times K$ satisfaisant la condition suivante:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, Q_{ij} > 0 \Leftrightarrow Q_{ji} > 0$$

- Soit $(X_n)_{n \geq 0}$ la suite de variables aléatoires construite ainsi:

1. On simule X_0 selon π_0 .
2. Pour $n \geq 1$:
 - a. On tire Y_n selon la loi $Q_{X_{n-1} \bullet}$ (la ligne de Q donnée par X_{n-1}).

Algorithme de Metropolis Hastings (formulation discrete)

- Soit Q une matrice stochastique $K \times K$ satisfaisant la condition suivante:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, Q_{ij} > 0 \Leftrightarrow Q_{ji} > 0$$

- Soit $(X_n)_{n \geq 0}$ la suite de variables aléatoires construite ainsi:

1. On simule X_0 selon π_0 .
2. Pour $n \geq 1$:
 - a. On tire Y_n selon la loi $Q_{X_{n-1} \bullet}$ (la ligne de Q donnée par X_{n-1}).
 - b. On tire une loi uniforme U indépendante de Y_n .

Algorithme de Metropolis Hastings (formulation discrete)

- Soit Q une matrice stochastique $K \times K$ satisfaisant la condition suivante:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, Q_{ij} > 0 \Leftrightarrow Q_{ji} > 0$$

- Soit $(X_n)_{n \geq 0}$ la suite de variables aléatoires construite ainsi:

1. On simule X_0 selon π_0 .
2. Pour $n \geq 1$:
 - a. On tire Y_n selon la loi $Q_{X_{n-1} \bullet}$ (la ligne de Q donnée par X_{n-1}).
 - b. On tire une loi uniforme U indépendante de Y_n .
 - c. On calcule la quantité

$$\alpha(X_{n-1}, Y_n) = \min \left(1, \frac{\pi_{Y_n} Q_{Y_n X_{n-1}}}{\pi_{X_{n-1}} Q_{X_{n-1} Y_n}} \right)$$

Algorithme de Metropolis Hastings (formulation discrete)

- Soit Q une matrice stochastique $K \times K$ satisfaisant la condition suivante:

$$\forall (i, j) \in K \times K, Q_{ij} > 0 \Leftrightarrow Q_{ji} > 0$$

- Soit $(X_n)_{n \geq 0}$ la suite de variables aléatoires construite ainsi:

1. On simule X_0 selon π_0 .
2. Pour $n \geq 1$:
 - a. On tire Y_n selon la loi $Q_{X_{n-1} \bullet}$ (la ligne de Q donnée par X_{n-1}).
 - b. On tire une loi uniforme U indépendante de Y_n .
 - c. On calcule la quantité

$$\alpha(X_{n-1}, Y_n) = \min \left(1, \frac{\pi_{Y_n} Q_{Y_n X_{n-1}}}{\pi_{X_{n-1}} Q_{X_{n-1} Y_n}} \right)$$

- d. On pose:

$$X_n = \begin{cases} Y_n & \text{si } U \leq \alpha(X_{n-1}, Y_n) \\ X_{n-1} & \text{sinon} \end{cases}$$

Algorithme de Metropolis Hastings (formulation discrete)

- Soit Q une matrice stochastique $K \times K$ satisfaisant la condition suivante:

$$\forall (i, j) \in \mathcal{K} \times \mathcal{K}, Q_{ij} > 0 \Leftrightarrow Q_{ji} > 0$$

- Soit $(X_n)_{n \geq 0}$ la suite de variables aléatoires construite ainsi:

1. On simule X_0 selon π_0 .
2. Pour $n \geq 1$:
 - a. On tire Y_n selon la loi $Q_{X_{n-1} \bullet}$ (la ligne de Q donnée par X_{n-1}).
 - b. On tire une loi uniforme U indépendante de Y_n .
 - c. On calcule la quantité

$$\alpha(X_{n-1}, Y_n) = \min \left(1, \frac{\pi_{Y_n} Q_{Y_n X_{n-1}}}{\pi_{X_{n-1}} Q_{X_{n-1} Y_n}} \right)$$

- d. On pose:

$$X_n = \begin{cases} Y_n & \text{si } U \leq \alpha(X_{n-1}, Y_n) \\ X_{n-1} & \text{sinon} \end{cases}$$

- **Propriété 1:** $(X_n)_{n \geq 1}$ est une chaîne de Markov de transition P où

$$P_{ij} = Q_{ij} \alpha(i, j) \text{ si } i \neq j, \quad P_{jj} = 1 - \sum_{i \neq j} P_{ij}$$

- **Propriété 2:** De plus π est invariante pour P .

Preuve

Matrice de transition P

On veut montrer que:

$$P_{ij} = Q_{ij}\alpha(i,j) \text{ si } i \neq j, \quad P_{jj} = 1 - \sum_{j \neq i} P_{ij}$$

Preuve

Matrice de transition P

On veut montrer que:

$$P_{ij} = Q_{ij}\alpha(i, j) \text{ si } i \neq j, \quad P_{jj} = 1 - \sum_{j \neq i} P_{ij}$$

Soit $i \neq j$:

$$\begin{aligned}\mathbb{P}(X_n = j | X_{n-1} = i) &= \mathbb{P}(Y_n = j, U \leq \alpha(X_{n-1}, Y_n) | X_{n-1} = i) \\ &= \mathbb{P}(Y_n = j, U \leq \alpha(i, j) | X_{n-1} = i) \\ &= \mathbb{P}(U \leq \alpha(i, j) | X_{n-1} = i, Y_n = j) \mathbb{P}(Y_n = j | X_{n-1} = i) \\ &= Q_{ij}\alpha(i, j)\end{aligned}$$

Preuve que π est mesure invariante

Il suffit de montrer que π est réversible pour P .

Soient $i \neq j \in \mathcal{K}$:

Preuve que π est mesure invariante

Il suffit de montrer que π est réversible pour P .

Soient $i \neq j \in \mathcal{K}$:

$$\begin{aligned}\pi_i P_{ij} &= \pi_i Q_{ij} \alpha(i, j) \\ &= \pi_i Q_{ij} \min \left(1, \frac{\pi_j Q_{ji}}{\pi_i Q_{ij}} \right) \\ &= \min (\pi_i Q_{ij}, \pi_j Q_{ji}) \\ &= \pi_j Q_{ji} \min \left(\frac{\pi_i Q_{ij}}{\pi_j Q_{ji}}, 1 \right) \\ &= \pi_j Q_{ji} \alpha(j, i) \\ &= \pi_j P_{ji}\end{aligned}$$

Algorithme dans le cas continu

Supposons qu'on veuille simuler dans \mathbb{R}^d selon une densité π , éventuellement connue à une constante près, c'est à dire que

$$\forall x \in \mathbb{R}^d, \pi(x) = \frac{\tilde{\pi}(x)}{\int_{\mathbb{R}^d} \tilde{\pi}(z) dz}$$

On remplace alors la matrice de transition par un *noyau de transition* sur \mathbb{R}^d , à savoir une fonction

$$\begin{aligned} q : \quad \mathbb{R}^d \times \mathbb{R}^d &\mapsto \mathbb{R}_+ \\ (x, y) &\mapsto q(x, y) \geq 0 \end{aligned}$$

telle que $\int_{\mathbb{R}^d} q(x, y) dy = 1$ (typiquement, la loi d'une marche aléatoire centrée en x).

Algorithme dans le cas continu

Supposons qu'on veuille simuler dans \mathbb{R}^d selon une densité π , éventuellement connue à une constante près, c'est à dire que

$$\forall x \in \mathbb{R}^d, \pi(x) = \frac{\tilde{\pi}(x)}{\int_{\mathbb{R}^d} \tilde{\pi}(z) dz}$$

On remplace alors la matrice de transition par un *noyau de transition* sur \mathbb{R}^d , à savoir une fonction

$$\begin{aligned} q : \mathbb{R}^d \times \mathbb{R}^d &\mapsto \mathbb{R}_+ \\ (x, y) &\mapsto q(x, y) \geq 0 \end{aligned}$$

telle que $\int_{\mathbb{R}^d} q(x, y) dy = 1$ (typiquement, la loi d'une marche aléatoire centrée en x).

Si on sait simuler, pour x fixé, selon q , et qu'on a $q(x, y) > 0 \Leftrightarrow q(y, x) > 0$, alors, l'algorithme de Metropolis reste valide en remplaçant π par $\tilde{\pi}$ et Q par q .

Algorithme dans le cas continu

Supposons qu'on veuille simuler dans \mathbb{R}^d selon une densité π , éventuellement connue à une constante près, c'est à dire que

$$\forall x \in \mathbb{R}^d, \pi(x) = \frac{\tilde{\pi}(x)}{\int_{\mathbb{R}^d} \tilde{\pi}(z) dz}$$

On remplace alors la matrice de transition par un *noyau de transition* sur \mathbb{R}^d , à savoir une fonction

$$\begin{aligned} q : \mathbb{R}^d \times \mathbb{R}^d &\mapsto \mathbb{R}_+ \\ (x, y) &\mapsto q(x, y) \geq 0 \end{aligned}$$

telle que $\int_{\mathbb{R}^d} q(x, y) dy = 1$ (typiquement, la loi d'une marche aléatoire centrée en x).

Si on sait simuler, pour x fixé, selon q , et qu'on a $q(x, y) > 0 \Leftrightarrow q(y, x) > 0$, alors, l'algorithme de Metropolis reste valide en remplaçant π par $\tilde{\pi}$ et Q par q .

- Le ratio ne nécessite pas la constante de normalisation car

$$\frac{\tilde{\pi}(y)}{\tilde{\pi}(x)} = \frac{\pi(y)}{\pi(x)}$$

Exemple 2: cas non conjugué

Exemple: Prédiction de présence d'oiseaux



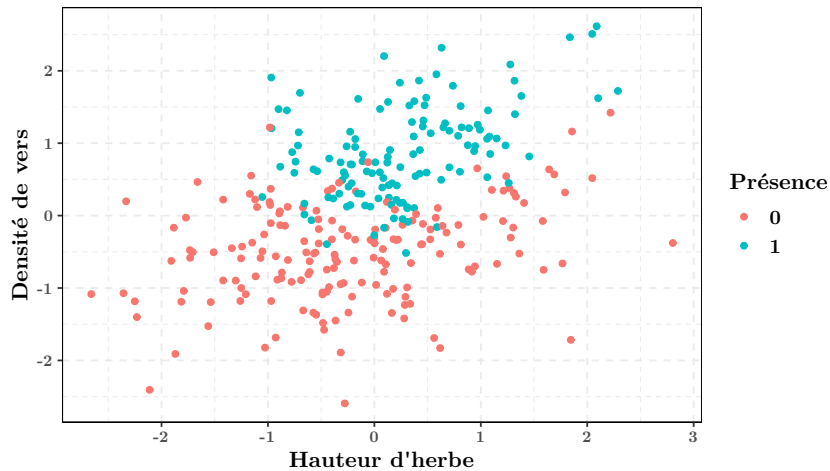
Une étude consiste en l'observation de la présence ou non de la linotte mélodieuse sur différents sites échantillonnés.

Caractéristiques des sites

Sur ces 300 sites sont mesurées différentes caractéristiques:

- ▶ Le nombre de vers moyens sur une surface au sol de $1m^2$. (Covariable 1)
- ▶ La hauteur d'herbe moyenne sur une surface au sol de $1m^2$. (Covariable 2)
- ▶ On calcule cette hauteur d'herbe au carré. (Covariable 3).

Données



Notations et modèle de régression probit

On note y_1, \dots, y_n les observations de présence (1 si on observe un oiseau, 0 sinon) sur les sites 1 à n .

On note

$$\mathbf{x}_k = \begin{pmatrix} \text{Nb. vers} & \text{Haut. herbe} & \text{Haut. herbe}^2 \\ x_{k,1} & x_{k,2} & x_{k,3} \end{pmatrix}^T$$

le vecteur des covariables sur le k -ème site ($1 \leq k \leq n$).

Notations et modèle de régression probit

On note y_1, \dots, y_n les observations de présence (1 si on observe un oiseau, 0 sinon) sur les sites 1 à n .

On note

$$\mathbf{x}_k = \begin{pmatrix} \text{Nb. vers} \\ X_{k,1} \\ \text{Haut. herbe} \\ X_{k,2} \\ \text{Haut. herbe}^2 \\ X_{k,3} \end{pmatrix}^T$$

le vecteur des covariables sur le k -ème site ($1 \leq k \leq n$).

On pose le modèle suivant:

$Y_k \sim \text{Bern}(p_k)$ où

$$p_k = \phi(\beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3}) = \phi(\mathbf{x}_k^T \theta),$$

où

- ▶ ϕ est la fonction de répartition d'une $\mathcal{N}(0, 1)$, i.e.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{u^2}{2}} du$$

- ▶ $\theta = \{\beta_0, \beta_1, \beta_2, \beta_3\}$ est le vecteur des paramètres à estimer.

Modèle Bayésien

Prior sur θ

Comme a priori sur θ , on choisit une normale avec une grande variance $\theta \overset{\text{prior}}{\sim} \mathcal{N}(0, 4I)$, donc

$$\pi(\theta) = \frac{1}{\sqrt{2\pi \times 4}^4} e^{-\frac{1}{8} \theta^T \theta}$$

où I est la matrice Identité (ici 4×4)

Modèle Bayésien

Prior sur θ

Comme a priori sur θ , on choisit une normale avec une grande variance $\theta \overset{\text{prior}}{\sim} \mathcal{N}(0, 4I)$, donc

$$\pi(\theta) = \frac{1}{\sqrt{2\pi \times 4^4}} e^{-\frac{1}{8} \theta^T \theta}$$

où I est la matrice Identité (ici 4×4)

Vraisemblance

Pour un vecteur d'observations $y_{1:k}$, la vraisemblance

$$L(y_{1:k}|\theta) = \prod_{k=1}^n \underbrace{\phi(\mathbf{x}_k^T \theta)^{y_k}}_{\text{Proba. présence}} \times \underbrace{(1 - \phi(\mathbf{x}_k^T \theta))^{1-y_k}}_{\text{Proba. absence}}$$

Modèle Bayésien

Prior sur θ

Comme a priori sur θ , on choisit une normale avec une grande variance $\theta \overset{\text{prior}}{\sim} \mathcal{N}(0, 4I)$, donc

$$\pi(\theta) = \frac{1}{\sqrt{2\pi \times 4^4}} e^{-\frac{1}{8}\theta^T \theta}$$

où I est la matrice Identité (ici 4×4)

Vraisemblance

Pour un vecteur d'observations $y_{1:k}$, la vraisemblance

$$L(y_{1:k}|\theta) = \prod_{k=1}^n \underbrace{\phi(\mathbf{x}_k^T \theta)^{y_k}}_{\text{Proba. présence}} \times \underbrace{(1 - \phi(\mathbf{x}_k^T \theta))^{1-y_k}}_{\text{Proba. absence}}$$

Posterior

Le posterior est donc donné par:

$$\pi(\theta|y_{1:n}) \propto \pi(\theta)L(y_{1:n}|\theta) \propto e^{-\frac{1}{8}\theta^T \theta} \prod_{k=1}^n \phi(\mathbf{x}_k^T \theta)^{y_k} (1 - \phi(\mathbf{x}_k^T \theta))^{1-y_k}$$

Algorithme de Metropolis Hastings

La loi stationnaire cible est $\pi(\mathbf{y}|\theta)$. Pour $n = 300$, l'acceptation rejet vu au cours précédent fonctionnera très mal en pratique.

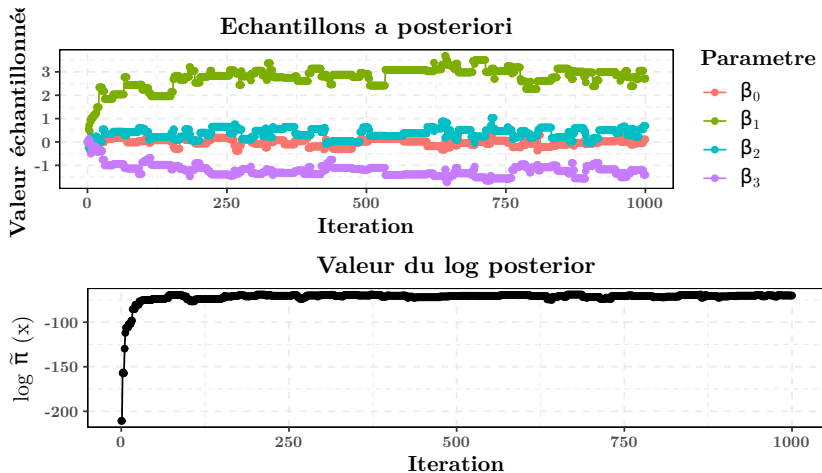
Algorithme de Metropolis Hastings

La loi stationnaire cible est $\pi(\mathbf{y}|\theta)$. Pour $n = 300$, l'acceptation rejet vu au cours précédent fonctionnera très mal en pratique.

On fait un algorithme de Metropolis Hastings avec comme loi de proposition une marche aléatoire dans \mathbb{R}^4 , de matrice de covariance $\tau^2 \times I_4$.

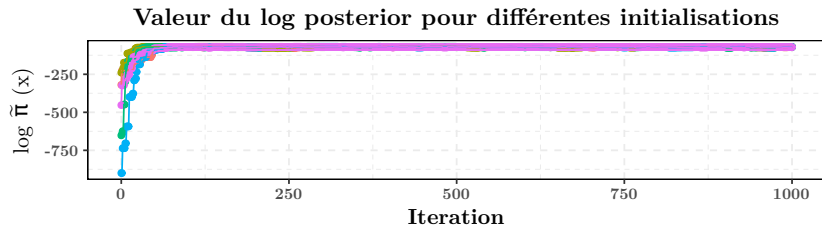
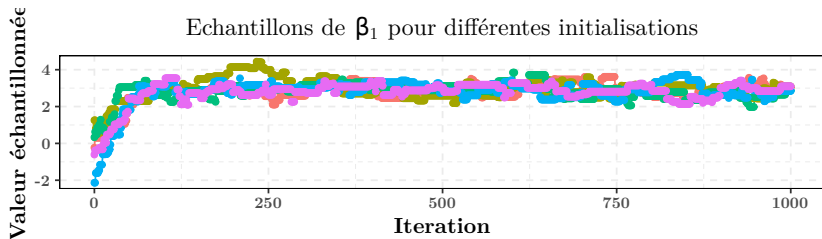
Résultat d'un algorithme lancé depuis un point de départ

- On choisit $\beta^{(0)} = (0, 0, 0, 0)$ et $\tau^2 = 0.1$, on lance 1000 itérations.

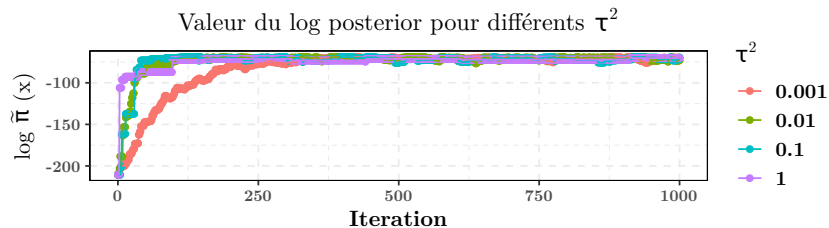
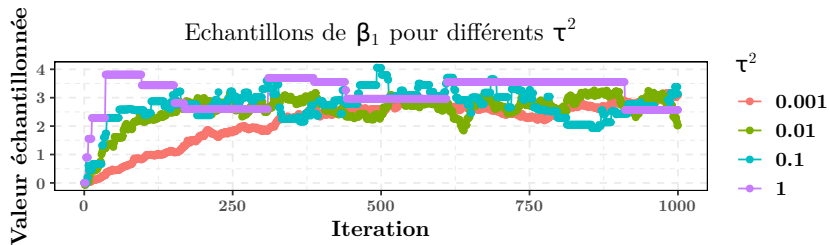


Sensibilité au point de départ

Il faut toujours vérifier la sensibilité au point de départ!



Influence de τ^2



Influence de τ^2

Taux d'acceptation dans l'algorithme

τ^2	Taux d'acceptation
0.001	0.781
0.010	0.515
0.100	0.147
1.000	0.013

Influence de τ^2

Taux d'acceptation dans l'algorithme

τ^2	Taux d'acceptation
0.001	0.781
0.010	0.515
0.100	0.147
1.000	0.013

Autocorrelation dans les chaînes

Correlation entre empirique entre β_1^n et $\beta_1^{(n+1)}$

τ^2	Autocorrelation
0.001	0.9993751
0.010	0.9917581
0.100	0.9835038
1.000	0.9864433

Reduction de l'autocorrelation

En pratique, on choisira une fraction des points. On appelle cela le **thinning**.

Reduction de l'autocorrelation

En pratique, on choisira une fraction des points. On appelle cela le **thinning**.

Autocorrélation en prenant un point sur 100.

τ^2	Autocorrelation
0.001	0.8092081
0.010	0.2395796
0.100	0.1502224
1.000	0.3861150

Estimation de la loi

Les premières valeurs n'ont aucune raison d'être tirées selon la loi cible.

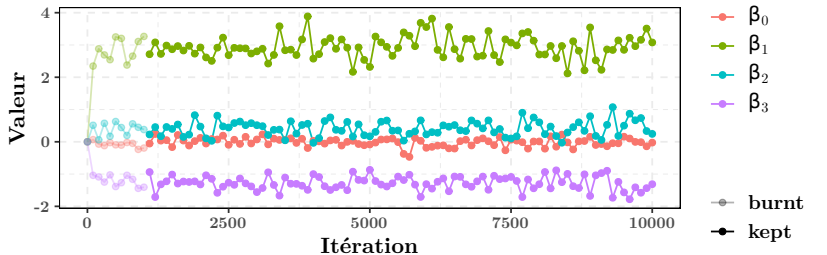
En pratique, on les supprimera. On appelle cela le **burn-in**.

Estimation de la loi

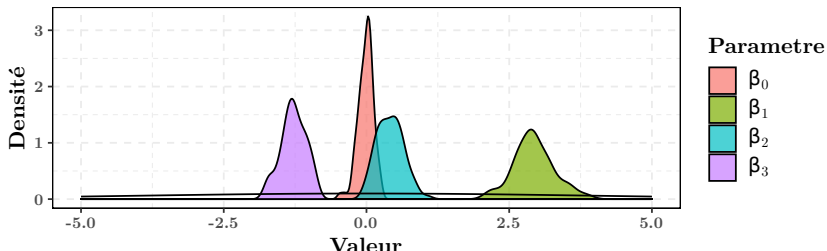
Les premières valeurs n'ont aucune raison d'être tirées selon la loi cible.

En pratique, on les supprimera. On appelle cela le **burn-in**.

Echantillon conservé (thin + burnin)



Loi a posteriori estimées



Autres algorithmes MCMC

Echantillonneur de Gibbs

- ▶ Utile quand θ est en grande dimension;
- ▶ On suppose qu'on sait simuler selon les loi conditionnelles de θ

Echantillonneur de Gibbs

- ▶ Utile quand θ est en grande dimension;
- ▶ On suppose qu'on sait simuler selon les loi conditionnelles de θ
- ▶ Soit X un vecteur aléatoire en dimension d $X = (X^{(1)}, \dots, X^{(d)})$.
- ▶ On note $X^{-(\ell)} = (X^{(1)}, \dots, X^{(\ell-1)}, X^{(\ell+1)}, X^{(d)})$,
- ▶ Si on sait simuler la variable aléatoire $X^{(\ell)} | X^{-(\ell)}$, l'algo est le suivant:
 1. Prendre $X_0 = (X_0^{(1)}, \dots, X_0^{(d)})$ tiré selon une loi initiale.
 2. Pour $k \geq 1$:
 - 2.1 Tirer ℓ uniformément dans $\{1, \dots, d\}$;
 - 2.2 Simuler Y selon la loi $X^{(\ell)} | \{X^{-(\ell)} = X_{k-1}^{-(\ell)}\}$
 - 2.3 Poser $X_k = (X_{k-1}^{(1)}, \dots, X_{k-1}^{(\ell-1)}, Y, X_{k-1}^{(\ell+1)}, X_{k-1}^{(d)})$

Propriété de l'échantillonneur de Gibbs

- ▶ L'échantillonneur de Gibbs est équivalent à un algorithme de Metropolis Hastings où la quantité α est toujours égale à 1,
- ▶ C'est à dire un Metropolis Hastings où on n'accepte tous les candidats!
- ▶ Algorithme utile dès que la simulation des lois conditionnelles est faisable.
- ▶ Si les lois conditionnelles induisent une matrice de transition (ou un noyau) de Markov irréductible et apériodique, alors le théorème ergodique s'applique.