les méthodes de Monte Carlo : exposé introductif

Fabien Campillo[1] Vivien Rossi[2]

[1]Projet ASPI
IRISA-INRIA Rennes

[2]IURC
Université Montpellier I

11 octobre 2006

plan

1. introduction
2. Monte Carlo
   - Exemples de base
   - Echantillonnage d'importance
3. MCMC
   - Rappels sur les chaînes de Markov
   - échantillonneur de Metropolis-Hastings
   - échantillonneur de Gibbs
   - Algorithme hybride

plan

Motivations : modélisation environementale

objectif : évaluation et gestion de ressources renouvelable

- peu de données
- données trés bruitées
- modèles incertains (paramètres inconnus)
- période d'acquisition des données élevée

$\longrightarrow$ cadre de travail bayésien bien adapté

- utiliser le savoir des experts
- beaucoup de temps pour faire des calculs

principe bayésien

$\longrightarrow$ estimer un paramètre θ avec les données $x_1, \ldots, x_n$

- loi a priori sur θ de densité de π_0
- calcul de la loi a posteriori

$$p(\theta|x_1, \ldots, x_n) = \frac{L(\theta, x_1, \ldots, x_n)\pi_0(\theta)}{\displaystyle\int L(\theta, x_1, \ldots, x_n)\pi_0(\theta) d\theta}$$

- en général on ne sait pas calcul analytiquement la loi a posteriori

$\longrightarrow$ approximation numérique par les méthodes de Monte Carlo par chaînes de Markov

Généralités sur les méthodes de Monte Carlo

méthodes de Monte Carlo : techniques d'estimation s'appuyant sur la simulation d'un grand nombre de variables aléatoires

avantages

- domaine d'application très vaste
- peu d'hypothèses de mise œuvre
- facile à implémenter

inconvéniants

- nécessite un **bon** générateur aléatoire
- grande variablilité ($\longrightarrow$ mal adapté aux pb d'optimisation)
- pas concurrentiel

plan

estimation de l'espérance $\theta = E[f(X)]$

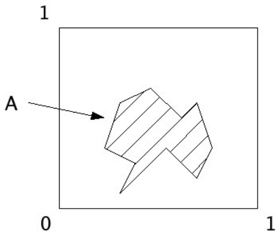
- X est une v.a. que l'on sait simuler

$\longrightarrow$ l'estimateur naturel de θ

$$\hat{\theta} = \frac{1}{N} \sum_{i=1}^{N} f(X_i) \text{ avec } X_1, \dots, X_N \;\; i.i.d. \text{ de même loi que } X$$

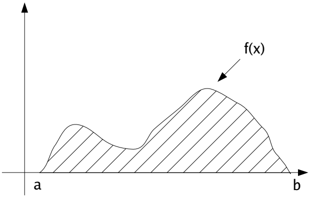
- propriétés théoriques (loi des grands nombres)
  - $E[\hat{\theta}] = \theta$
  - $V[\hat{\theta}] = \frac{1}{N} V[f(X)]$
  - intervalle de confiance par approximation gaussienne ...

estimation d'un volume



- objectif : estimer le volume A inclus dans $[0\ 1]^2$
- $\mathrm{vol}(A) = E[1_A(X)]$ avec $X \sim \mathcal{U}([0\ 1]^2)$
- méthode
  1. tirer N points uniformément dans $[0\ 1]^2$
  2. compter le nombre n de points dans A
  3. $\hat{\mathrm{vol}}(A) = \frac{n}{N}$

estimation d'une intégrale $I = \int_a^b f(x)dx$



- $I = \int f(x)1_{[a\ b]}(x)dx = (b-a)E[f(X)]$ avec $X \sim \mathcal{U}[a\ b]$

- méthode
  1. Tirer $x_1, \ldots, x_N$ uniformément dans $[a\ b]$
  2. $\hat{I} = \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i)$

- comparaison avec les méthodes d'intégration par quadrature
  - Vitesse Monte Carlo en $N^{-1/2}$
  - Vitesse Quadrature en $N^{-s/d}$ où d dimension de l'espace, s est tel que les dérivées d'ordre inférieur ou égal à s sont bornées.

importance sampling : estimation de $\theta = E[f(X)]$

- X v.a de densité π que l'on ne sait pas forcément simuler
- choisir une loi instrumentale de densité π^{prop} telle que $\pi^{prop}(x) > 0$ si $\pi(x)$
- un estimateur de θ est

$$\hat{\theta}' = \frac{1}{N}\sum_{i=1}^{N} f(Y_i)\frac{\pi(Y_i)}{\pi^{prop}(Y_i)} \text{ avec } Y_1, \ldots, Y_N \sim \pi^{prop}(y)dy$$

- permet de réduire la variance si π^{prop} est bien choisie
- exemple type : estimer la probabilité d'un événement rare

retour au cadre bayésien

- estimer un paramètre θ avec les données $x_1, \ldots, x_n$
- tirer un échantillon suivant $p(\theta | x_1, \ldots, x_n) \propto L(\theta, x_1, \ldots, x_n) \pi_0(\theta)$
- on ne peut pas utiliser la méthode Monte Carlo basique ni l'importance sampling

$\longrightarrow$ Monte Carlo par Chaîne de Markov

plan

MCMC : principe

- but : échantillonner selon une loi cible de densité $\pi(z)$ connue à une constante multiplicative près

$$\pi \stackrel{\text{déf}}{=} \text{loi}(\mathsf{Z})\,.$$

où $\mathsf{Z} = \mathsf{Z}_{1:n}$ est une variable réelle n-dimensionnelle

- principe
  - on construit une chaîne de Markov $(\mathsf{Z}^{(k)})_{k\geq 0}$ sur $\mathbb{R}^n$ dont la loi limite est π
  - lorsque la chaîne devient stationnaire, on extrait un échantillon de la chaîne $(\mathsf{Z}^{(k_1)}, \ldots, \mathsf{Z}^{(k_N)})$ pour estimer les quantités d'intérêt [théorème ergodique]

MCMC : cadre théorique

Soit une mesure de probabilité π définie sur un espace mesurable $(\mathbf{E}, \mathcal{E})$.

- le but des méthodes MCMC est de proposer une approximation de cette mesure cible :

$$\pi(f) \stackrel{\text{déf}}{=} \int_{\mathbf{E}} f(z)\, \pi(\mathrm{d}z)$$

  où f est une fonction réelle mesurable et π-intégrable

- si l'on sait construire une chaîne de Markov $\{\mathsf{Z}_n\}_{n\in\mathbb{N}}$ ergodique dont la mesure invariante est π alors :

$$\frac{1}{K} \sum_{k=1}^{K} f(\mathsf{Z}_k) \xrightarrow[K\uparrow\infty]{\mathbb{P}_z\text{-p.s.}} \pi(f)$$

  pour π-presque toute condition initiale z où $\mathbb{P}_z$ est la probabilité pour laquelle $\mathsf{Z}_0 = z$.

chaînes de Markov : réversibilité

- un noyau de transition $P(z, \mathrm{d}\tilde{z})$ sur $(\mathbf{E}, \mathcal{E})$ est dit **réversible** par rapport à la mesure $\pi(\mathrm{d}z)$ lorsque :

$$\int_A \pi(\mathrm{d}z)\, P(z, B) = \int_B \pi(\mathrm{d}z)\, P(z, A)\,, \quad \forall A,\, B \in \mathcal{E} \qquad (1)$$

avec $A = \mathbf{E}$: $\int \pi(\mathrm{d}z)\, P(z, B) = \int_B \pi(\mathrm{d}z) = \pi(B) \quad \forall B \in \mathcal{E}$
i.e. $\pi P = \pi$, «P préserve π» ou «π est invariant par P»

- **équation de bilan détaillé** : (1) s'écrit, pour tout $A,\, B \in \mathcal{E}$,

$$\int\int \mathbf{1}_A(z)\, \mathbf{1}_B(\tilde{z})\, \pi(\mathrm{d}z)\, P(z, \mathrm{d}\tilde{z}) = \int\int \mathbf{1}_A(z)\, \mathbf{1}_B(\tilde{z})\, \pi(\mathrm{d}\tilde{z})\, P(\tilde{z}, \mathrm{d}z)$$

$$\iff \pi(\mathrm{d}z)\ P(z, \mathrm{d}\tilde{z}) = \pi(\mathrm{d}\tilde{z})\ P(\tilde{z}, \mathrm{d}z)$$

Cette égalité entre mesures est définie sur $(\mathbf{E} \times \mathbf{E}, \mathcal{E} \otimes \mathcal{E})$

chaînes de Markov : réversibilité (suite)

- une chaîne de Markov $\{\mathsf{Z}_n\}_{n\in\mathbb{N}}$ définie sur $(\mathbf{E},\mathcal{E})$ de loi initiale $\pi(\mathrm{d}z)$ et de noyau de transition $P(z,\mathrm{d}\tilde{z})$ est dite *réversible* si P est réversible par rappport à π :

$$\mathbb{P}(\mathsf{Z}_n \in A\,,\ \mathsf{Z}_{n+1} \in B) = \mathbb{P}(\mathsf{Z}_{n+1} \in A\,,\ \mathsf{Z}_n \in B)\,, \quad \forall A,\, B \in \mathcal{E}\,.$$

- vérifier la réversibilité (1) $\iff$ vérifier l'équation de bilan détaillé. l'équation de bilan détaillé $\implies$ π est invariante par P : $\forall\ A \in \mathcal{E}$,

$$\begin{aligned}
\pi P(A) &= \int \pi(\mathrm{d}z)\, P(z,A) \\
&= \int\int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}z)\, P(z,\mathrm{d}\tilde{z}) \\
&= \int\int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}\tilde{z})\, P(\tilde{z},\mathrm{d}z) \\
&= \int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}\tilde{z}) \int P(\tilde{z},\mathrm{d}z) = \pi(A)\,.
\end{aligned}$$

Chaine de Markov : irréductibilité et périodicité

Soit une chaîne de Markov $\{Z_n\}_{n\in\mathbb{N}}$ de noyau de transition $P(z, \mathrm{d}\tilde{z})$

- une chaîne est π-**irréductible**, avec π mesure de probabilité, si

$$\forall z \in \mathbf{E}\,,\ \forall A \in \mathcal{E} \text{ avec } \pi(A) > 0\,,\ \exists n \in \mathbb{N} \text{ tel que } P^n(z, A) > 0$$

$$\text{où } P^n(z, A) \stackrel{\text{déf}}{=} \mathbb{P}(X_n \in A | X_0 = z)$$

- si une chaîne est π-irréductible où π est une mesure de probabilité invariante, alors cette mesure est unique aux négligeables près.

- une chaîne π-irréductible est **périodique** s'il existe une partition disjointe $\mathbf{E} = A_0 + \cdots + A_n$ (avec $n + 1 \geq 3$) telle que $\pi(A_n) = 0$ et

$$z \in A_{i-1} \Rightarrow P(z, A_i) = 1\,,\ 1 \leq i \leq n-1\,,$$
$$z \in A_{n-1} \Rightarrow P(z, A_0) = 1\,.$$

Dans le cas contraire elle est dite **apériodique**.

chaîne de Markov : comportement asymptotique

Si une chaîne de Markov

1. admet π comme mesure de probabilité invariante π
2. est π-irréductible
3. est apériodique

alors il existe $C \in \mathcal{E}$ tel que $\pi(C) = 1$ et pour tout $z \in C$ et $A \in \mathcal{E}$:

$$\mathbb{P}(\mathsf{Z}_n \in A | \mathsf{Z}_0 = z) \xrightarrow[n\uparrow\infty]{} \pi(A)$$

4. Si de plus la chaîne est Harris récurrente alors $C = \mathbf{E}$.

échantillonneur de Metropolis-Hastings (MH)

- objectif : construire un noyau de transition $P(z, \mathrm{d}\tilde{z})$ réversible et régulier sur $(\mathbf{E}, \mathcal{E})$ dont π est la mesure invariante.

- ingrédients :
  1. un noyau de transition $Q(z, \mathrm{d}\tilde{z})$ sur $(\mathbf{E}, \mathcal{E})$,
  2. une fonction mesurable $\alpha : \mathbf{E} \times \mathbf{E} \mapsto [0, 1]$.

- principe :
  1. partant d'un état $z \in \mathbf{E}$, le noyau $Q(z, \mathrm{d}\tilde{z})$ permet de générer un candidat $\tilde{z} \in \mathbf{E}$.
  2. ce candidat est accepté avec probabilité $\alpha(z, \tilde{z})$, sinon on conservera l'état z.

- mise en pratique
  - le noyau de proposition Q doit être donné
  - l'algorithme MH détermine l'expression de α

noyau de transition

le noyau de transition de Metropolis-Hastings $P(z, \mathrm{d}\tilde{z})$ est

$$P(z, \mathrm{d}\tilde{z}) \stackrel{\text{déf}}{=} \alpha(z, \tilde{z})\, Q(z, \mathrm{d}\tilde{z}) + \underbrace{\left\{ \int_{\mathbf{E}} (1 - \alpha(z, z'))\, Q(z, \mathrm{d}z') \right\}}_{r(z)} \delta_z(\mathrm{d}\tilde{z}) \tag{2}$$

où $\delta_z(\mathrm{d}\tilde{z})$ est la mesure de Dirac en z et $r(z)$ est la probabilité de rejet. .i.e. pour tout mesurable A

$$P(z, A) \stackrel{\text{déf}}{=} \int_A \alpha(z, \tilde{z})\, Q(z, \mathrm{d}\tilde{z}) + r(z)\, \mathbf{1}_A(z)\,, \quad \forall A \in \mathcal{E}\,.$$

l'algorithme MH perturbe un noyau de proposition $Q(z, \mathrm{d}\tilde{z})$ pour obtenir un noyau $P(z, \mathrm{d}\tilde{z})$ laissant invariant la mesure cible $\pi(\mathrm{d}z)$.

rapport de Metropolis-Hastings

- comment déterminer la probabilité d'acceptation $\alpha(z, \tilde{z})$ telle que le noyau associé (2) préserve la mesure cible π ?

- condition suffisante : la réversibilité est vérifie $\iff$ l'équation de bilan détaillé suivante est satisfaite :

$$\pi(\mathrm{d}z)\; P(z, \mathrm{d}\tilde{z}) = \pi(\mathrm{d}\tilde{z})\; P(\tilde{z}, \mathrm{d}z)\,. \tag{3}$$

- Le noyau de transition de MH (2) vérifie l'équation de bilan détaillé (3) si et seulement si :

$$\pi(\mathrm{d}z)\; Q(z, \mathrm{d}\tilde{z})\; \alpha(z, \tilde{z}) = \pi(\mathrm{d}\tilde{z})\; Q(\tilde{z}, \mathrm{d}z)\; \alpha(\tilde{z}, z)\,, \tag{4}$$

(la composante diagonale est négligeable dans cette équation)

cas mutuellement dominé

il existe une mesure ν telle que $\pi(\mathrm{d}z) = p(z)\,\nu(\mathrm{d}z)$ et $Q(z, \mathrm{d}\tilde{z}) = q(z, \tilde{z})\,\nu(\mathrm{d}\tilde{z})$.

- soit $R = \{(z, \tilde{z}) \in \mathbf{E} \times \mathbf{E} \,:\, p(z)\,q(z, \tilde{z}) > 0 \text{ et } p(\tilde{z})\,q(\tilde{z}, z) > 0\}$ et
$$r(z, \tilde{z}) = \frac{p(z)\,q(z, \tilde{z})}{p(\tilde{z})\,q(\tilde{z}, z)}\,.$$
- l'équation de bilan détaillé est vérifiée $\iff$
  1. $p(z)\,q(z, \tilde{z})\,\alpha(z, \tilde{z}) = 0$ pour tout $(z, \tilde{z}) \in R^c$ $\nu \times \nu$-p.p.
  2. $\alpha(z, \tilde{z})\,r(z, \tilde{z}) = \alpha(\tilde{z}, z)$ pour tout $(z, \tilde{z}) \in R$ $\nu \times \nu$-p.p.
- la probabilité d'acceptation :
$$\alpha(z, \tilde{z}) \stackrel{\text{déf}}{=} \begin{cases} 1 \wedge r(\tilde{z}, z)\,, & \text{si } (z, \tilde{z}) \in R\,, \\ 0\,, & \text{si } (z, \tilde{z}) \notin R\,. \end{cases}$$
  - la condition (1) est satisfaite par construction
  - pour la condition (2) : pour tout $(z, \tilde{z}) \in R$,
$$\alpha(z, \tilde{z})\,r(z, \tilde{z}) = \min\{r(z, \tilde{z})\,,\, r(\tilde{z}, z)\,r(z, \tilde{z})\} = \min\{r(z, \tilde{z})\,,\, 1\} = \alpha(\tilde{z}, z)$$

algorithme

Soit p la densité de la loi cible, q la densité de noyau markovien de proposition

- initialisation : choix arbitraire de $z^{(0)}$
- itération t
  1. Sachant $z^{(t-1)}$, on génère $\tilde{z} \sim q(z^{(t-1)}, z)$
  2. $\alpha = \min\left(\frac{p(\tilde{z})/q(z^{(t-1)}, \tilde{z})}{p(z)/q(\tilde{z}, z^{(t-1)})}, 1\right)$
  3. $z^{(t)} = \begin{cases} \tilde{z} & \text{avec proba } \alpha \\ z^{(t-1)} & \text{avec proba } 1-\alpha \end{cases}$

en pratique

quelques défauts :

- le choix de la loi de proposition est critique
  (si $supp(p) \subset supp(q) \implies \pi$-irréductibilité)
- l'initialisation est critique
- il est difficile de diagnostiquer la convergence
- le temps de chauffe peut être **très** long

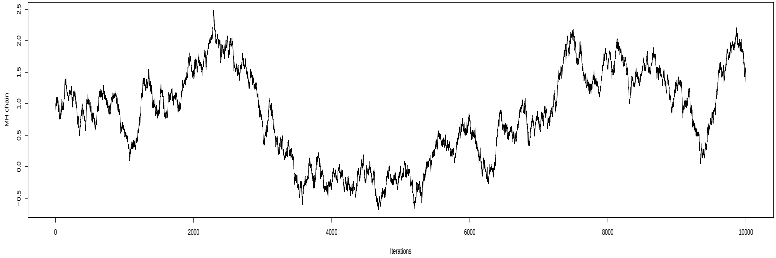
exemple

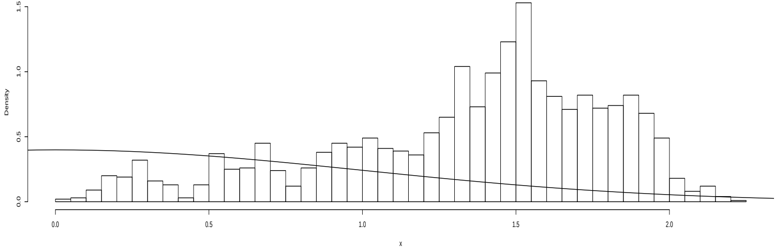
- soit la loi cible $\mathcal{N}(0,1)$
- soit la loi de proposition une marche aléatoire gaussienne :

$$q(z^{(t-1)},\tilde{z}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\{\frac{1}{2\sigma^2}(z^{(t-1)}-\tilde{z})^2\}$$

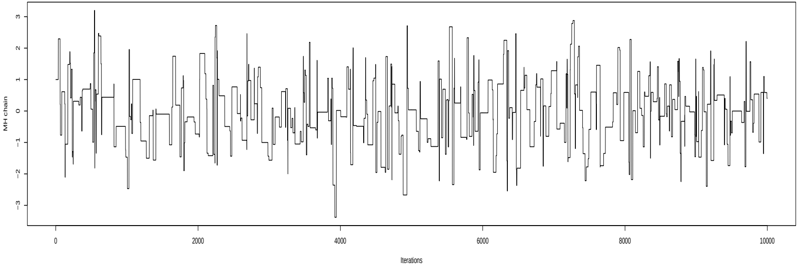
- la comportement de la chaîne varie beaucoup suivant le choix de σ^2

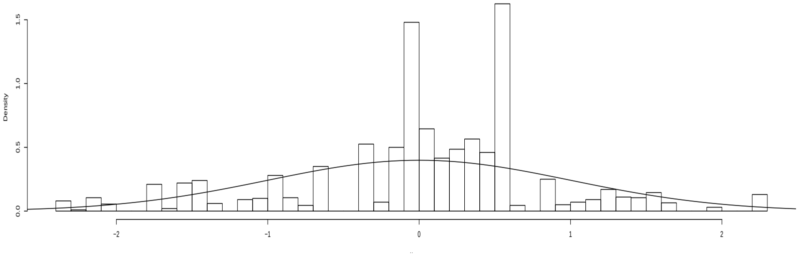
exemple : $\sigma^2 = 10^{-4}$



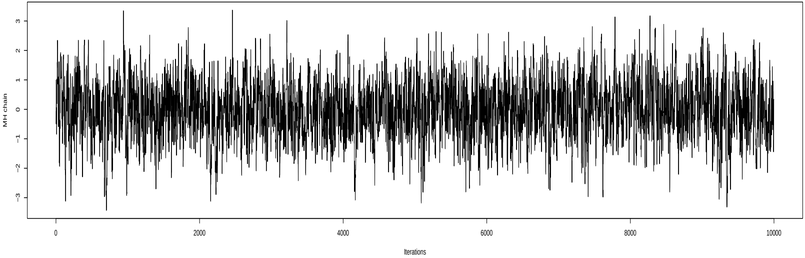


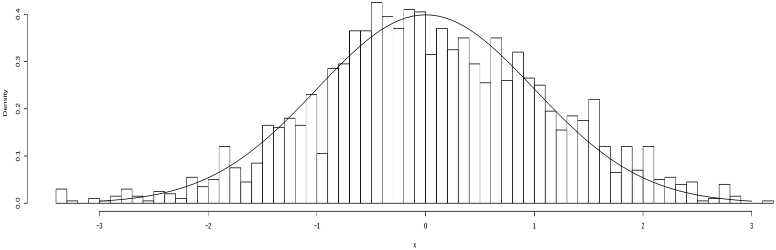
exemple : $\sigma^2 = 10^3$





exemple : $\sigma^2 = 1$





échantillonneur de Gibbs

soit la loi $\pi(\mathrm{d}z)$ d'une variable aléatoire $\mathsf{Z} = \mathsf{Z}_{1:d}$ définie sur l'espace mesurable produit

$$(\mathbf{E}, \mathcal{E}) = (\mathbf{E}_1 \times \cdots \times \mathbf{E}_d \,,\, \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_d) \,.$$

soit les lois marginales conditionnelles :

$$\pi_\ell(\mathrm{d}z_\ell | z_{\neg\ell}) \stackrel{\text{déf}}{=} \mathbb{P}(\mathsf{Z}_\ell \in \mathrm{d}z_\ell | \mathsf{Z}_{\neg\ell} = z_{\neg\ell}) \,, \quad \ell = 1 \cdots d$$

où $\neg\ell \stackrel{\text{déf}}{=} \{1, \ldots, d\} \setminus \{\ell\}$.

- principe de l'algorithme
  - construire une chaîne de Markov $\mathsf{Z}^{(k)}$ admettant π comme loi stationnaire
  - mise à jour séquentielle des composantes de $\mathsf{Z}_\ell^{(k)}$ de la chaîne

algorithme

- on part de $Z^{(0)}$ et l'itération $Z^{(k)} \to Z^{(k+1)}$ est :

$$
\begin{aligned}
Z_1^{(k+1)} &\sim \pi_1(\,\cdot\,|Z_{2:n}^{(k)}) \\
&\vdots \\
Z_\ell^{(k+1)} &\sim \pi_\ell(\,\cdot\,|Z_{1:\ell-1}^{(k+1)}, Z_{\ell+1:n}^{(k)}) \\
&\vdots \\
Z_n^{(k+1)} &\sim \pi_n(\,\cdot\,|Z_{1:n-1}^{(k+1)})
\end{aligned}
$$

- il des variantes avec sélection aléatoire de l'ordre des composantes

noyau de transition et invariance

- le noyau de transition de Gibbs $P(z, \mathrm{d}\tilde{z})$ est

$$P(z, \mathrm{d}\tilde{z}) \stackrel{\text{déf}}{=} \pi_1(\,\mathrm{d}\tilde{z}_1\,|z_{2:d}) \ldots \pi_\ell(\,\mathrm{d}\tilde{z}_\ell\,|\tilde{z}_{1:\ell-1}, z_{\ell+1:n}) \ldots \pi_d(\,\mathrm{d}\tilde{z}_d\,|\tilde{z}_{1:d-1})$$

- π est invariante par P : pour tout $A = \bigotimes_{\ell=1}^d A_\ell$

$$\pi P(A) = \int \pi(\mathrm{d}z)\, P(z, A) \;=\; \int\!\!\int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}z)\, P(z, \mathrm{d}\tilde{z})$$

$$= \int\!\!\int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}z)\, \pi_1(\,\mathrm{d}\tilde{z}_1\,|z_{2:d}) \prod_{\ell=2}^d \pi_\ell(\,\mathrm{d}\tilde{z}_\ell\,|\tilde{z}_{1:\ell-1}, z_{\ell+1:n})$$

$$= \int\!\!\int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}z_{2:d})\, \pi_1(\,\mathrm{d}\tilde{z}_1\,|z_{2:d}) \prod_{\ell=2}^d \pi_\ell(\,\mathrm{d}\tilde{z}_\ell\,|\tilde{z}_{1:\ell-1}, z_{\ell+1:n})$$

$$= \int\!\!\int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}\tilde{z}_1 \mathrm{d}z_{2:d}) \prod_{\ell=2}^d \pi_\ell(\,\mathrm{d}\tilde{z}_\ell\,|\tilde{z}_{1:\ell-1}, z_{\ell+1:n})$$

$$\ldots$$

$$= \int \mathbf{1}_A(\tilde{z})\, \pi(\mathrm{d}\tilde{z}) \;= \pi(A)\,.$$

Gibbs est un cas particulier de MH

$\exists\, \nu$: $\pi(\mathrm{d}z) = p(z)\,\nu(\mathrm{d}z)$ et $\pi_\ell(\,\mathrm{d}\tilde z_\ell\,|\tilde z_{1:\ell-1}, z_{\ell+1:n}) = p_\ell(z_\ell|z_{1:\ell-1}, \tilde z_{\ell+1:n})\,\nu(\mathrm{d}\tilde z)$
Soit $q(z,\tilde z) = p_1(z_1|\tilde z_{2:d}) \ldots p_\ell(z_\ell|z_{1:\ell-1}, \tilde z_{\ell+1:n}) \ldots p_d(z_d|z_{1:d-1})$

- formalisation type MH du noyau de transition de Gibbs $P(z, \mathrm{d}\tilde z)$:

$$P(z,\mathrm{d}\tilde z) = \alpha(z,\tilde z) q(z,\tilde z)\nu(\mathrm{d}\tilde z) + r(z)\ \delta_z(\mathrm{d}\tilde z)$$

  avec $\alpha(z,\tilde z) = \min(1, r(z,\tilde z))$ et $r(z,\tilde z) = \frac{p(z)\,q(z,\tilde z)}{p(\tilde z)\,q(\tilde z,z)}$

$$r(z,\tilde z) = \frac{p(\tilde z)}{p(z)}\ \frac{p_d(z_d|z_{1:d-1}) \ldots p_\ell(z_\ell|z_{1:\ell-1}, \tilde z_{\ell+1:n}) \ldots p_1(z_1|\tilde z_{2:d})}{p_d(\tilde z_d|\tilde z_{1:d-1}) \ldots p_\ell(\tilde z_\ell|\tilde z_{1:\ell-1}, z_{\ell+1:n}) \ldots p_1(\tilde z_1|z_{2:d})}$$

$$\vdots$$

$$= \frac{1}{r(z,\tilde z)} \Longrightarrow r(z,\tilde z) = 1$$

- $\Longrightarrow \alpha(z,\tilde z) = 1$: on accepte toujours la proposition

en pratique

- les avantages
  - pas de loi de proposition à choisir
  - pas de rejet des propositions
  - en général plus efficace que MH
- les inconvéniants
  - nécessite de savoir simuler suivant les lois conditionnelles
  - explore mal les «diagonales»

Metropolis-Hastings hybridé avec Gibbs (MHwG)

- soit la loi $\pi(\mathrm{d}z)$ d'une variable aléatoire $\mathsf{Z} = \mathsf{Z}_{1:d}$ définie sur l'espace mesurable produit

$$(\mathbf{E}, \mathcal{E}) = (\mathbf{E}_1 \times \cdots \times \mathbf{E}_d \,,\, \mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_d)\,.$$

- les lois marginales conditionnelles :

$$\pi_\ell(\mathrm{d}z_\ell | z_{\neg\ell}) \stackrel{\text{déf}}{=} \mathbb{P}(\mathsf{Z}_\ell \in \mathrm{d}z_\ell | \mathsf{Z}_{\neg\ell} = z_{\neg\ell})\,, \quad \ell = 1 \cdots d$$

  sont connue à une constante de normalisation près

- principe de l'algorithme
  - construire une chaîne de Markov $\mathsf{Z}^{(k)}$ admettant π comme loi stationnaire
  - choisir $Q_1, \ldots, Q_d$ une loi de proposition pour chaque composante
  - mise à jour séquentielle des composantes de $\mathsf{Z}_\ell^{(k)}$ de la chaîne

MHwG

partant de $Z_{1:n}^{(0)}$, la k-ième itération de l'échantillonneur $Z_{1:n}^{(k)} \to Z_{1:n}^{(k+1)}$ est la suivante : on pose $z_{1:n} \leftarrow Z_{1:n}^{(k)}$ et pour tout $\ell = 1:n$,

- on tire un candidat potentiel

$$\tilde{z}_\ell \sim q_\ell(\,\cdot\,|z_{\neg\ell})$$

- on pose

$$z_\ell \leftarrow \begin{cases} \tilde{z}_\ell & \text{avec probabilité } \alpha_\ell(z_\ell, \tilde{z}_\ell) \\ z_\ell & \text{avec probabilité } 1 - \alpha_\ell(z_\ell, \tilde{z}_\ell) \end{cases}$$

  avec

$$\alpha_\ell(z_\ell, \tilde{z}_\ell) \stackrel{\text{déf}}{=} 1 \wedge \frac{\pi_\ell(\tilde{z}_\ell|z_{\neg\ell})}{\pi_\ell(z_\ell|z_{\neg\ell})} \frac{q_\ell(z_\ell)}{q_\ell(\tilde{z}_\ell)}$$

alors $Z_{1:n}^{(k+1)} \leftarrow z_{1:n}$

algorithme

choisir $z_{1:n} \in \mathbb{R}^n$
for $k = 1, 2, \dots$ **do**
  for $\ell = 1 : n$ **do**
    $\tilde{z}_\ell \sim q_\ell(\cdot \,|z_{\neg \ell})$ {candidat potentiel}
    $u \sim \mathcal{U}[0,1]$
    if $u \le \alpha(z_\ell, \tilde{z}_\ell)$ **then**
      $z_\ell \leftarrow \tilde{z}_\ell$
    end if
  end for
end for

noyau de transition

- le noyau de transition de MHwG $K(z,\mathrm{d}\tilde z)$ est

$$K(z,\mathrm{d}\tilde z) \stackrel{\text{déf}}{=} K_1(\,\mathrm{d}\tilde z_1\,|z_{2:d})\dots K_\ell(\,\mathrm{d}\tilde z_\ell\,|\tilde z_{1:\ell-1},z_{\ell+1:n})\dots K_d(\,\mathrm{d}\tilde z_d\,|\tilde z_{1:d-1})$$

avec

$$K_\ell(\,\mathrm{d}\tilde z_\ell\,|\tilde z_{1:\ell-1},z_{\ell+1:n}) = \alpha(z_\ell,\tilde z_\ell)q_\ell(z_\ell,\tilde z_\ell)\nu(\mathrm{d}\tilde z_\ell) + r(z)\;\delta_z(\mathrm{d}\tilde z_\ell)$$

- propriété d'invariance, irréductibilité $\dots$

en pratique

même profil que MH

- le choix des lois de proposition est critique
  (si $supp(p) \subset supp(q) \implies \pi$-irréductibilité)
- l'initialisation est critique
- il est difficile de diagnostiquer la convergence
- le temps de chauffe peut être **très** long

conclusion

- MCMC
  - domaine d'application très vaste
  - simple à implémenter
  - lent à converger
  - diagnostique de la convergence difficile
- Modélisation environnementale
  - cadre bayésien
  - structure temporelle $\longrightarrow$ modèle de Markov cachés
  - MHwG utilisé en pratique mais pas toujours satisfaisant
  - passage aux filtres particulaires mais outil mal adapté
- besoin de développer des outils spécifiques
  - tirer parti des MCMC et des méthodes particulaires
  - faire interagir des chaînes parallèles (population Monte Carlo)
  - ...