Inférence/calibration (?) bayésienne

Julien Papaïx (avec l'aide de Samuel Soubeyrand et d'Emily Walker)

Introduction à la modélisation hiérarchique bayésienne

Modélisation paramétrique

Observations $y_1, \ldots, y_n$, avec

$$\mathbf{y} = (y_1, \ldots, y_n) \sim [\mathbf{y}|\theta],\ \theta \text{ paramètre inconnu.}$$

Objectif : estimer le paramètre θ à partir de l'échantillon $y_1, \ldots, y_n$.

Exemple : régression linéaire classique, $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 I)$, c'est à dire

$$y_i = \beta_0 + \beta_1 * x_{1i} + \ldots + \varepsilon_i, \text{ avec } \varepsilon_i \sim N(0, \sigma^2).$$

Une approche classique : le maximum de vraisemblance

Vraisemblance : mesure de l'adéquation entre la distribution observée sur un échantillon aléatoire et une loi de probabilité supposée décrire une réalité sur la population dont l'échantillon est issu.

On cherche la valeur de θ qui maximise la vraisemblance, c'est à dire on cherche la valeur de θ qui rend l'observation de $\mathbf{y}$ la plus probable.

Approche bayésienne : formule de Bayes

Le paramètre inconnu θ devient une variable aléatoire au même titre que les observations.

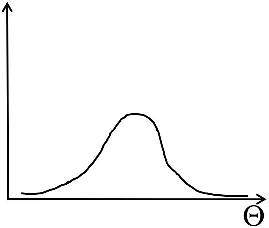
$\Rightarrow$ on cherche donc à caractériser la distribution *a posteriori* des paramètres, $[\theta|\mathbf{y}]$, à l'aide de la formule de Bayes.

$$[\theta|\mathbf{y}] = \frac{[\theta][\mathbf{y}|\theta]}{[\mathbf{y}]}$$

Approche bayésienne : principe général

Connaissance a priori

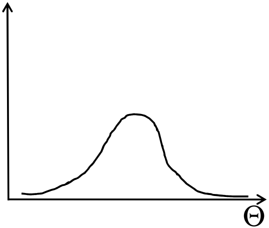
$$[\Theta]$$



Approche bayésienne : principe général

Connaissance a priori

$[\Theta]$



Données
+
Vraisemblance

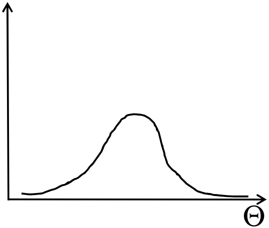
$[\mathbf{y}|\Theta]$

$$\begin{cases} \mu_i = \beta_0 + \beta_1 x_{1i} + \dots \\ y_i \sim N(\mu_i, \sigma^2) \end{cases}$$

Approche bayésienne : principe général

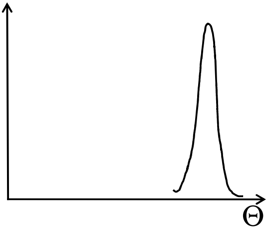
Connaissance a priori

$[\Theta]$



Connaissance après
mise à jour

$[\Theta|\mathbf{y}]$



Données
+
Vraisemblance

$[\mathbf{y}|\Theta]$

$$\begin{cases} \mu_i = \beta_0 + \beta_1 x_{1i} + \dots \\ y_i \sim N(\mu_i, \sigma^2) \end{cases}$$

Approche bayésienne : estimation

Problème $\rightarrow$ multiples intégrales à calculer :

- $[\theta|\mathbf{y}] = [\theta][\mathbf{y}|\theta]/[\mathbf{y}]$, avec $[\mathbf{y}] = \int_\Theta([\mathbf{y}|\alpha][\alpha])d\alpha$,
- calcul des distributions marginales si θ est multivarié,
- modèles à variables latentes...

Approche bayésienne : estimation

Problème $\rightarrow$ multiples intégrales à calculer :

- $[\theta|\mathbf{y}] = [\theta][\mathbf{y}|\theta]/[\mathbf{y}]$, avec $[\mathbf{y}] = \int_{\Theta}([\mathbf{y}|\alpha][\alpha])d\alpha$,
- calcul des distributions marginales si θ est multivarié,
- modèles à variables latentes...

Approche bayésienne : estimation

Problème $\rightarrow$ multiples intégrales à calculer :

- $[\theta|\mathbf{y}] = [\theta][\mathbf{y}|\theta]/[\mathbf{y}]$, avec $[\mathbf{y}] = \int_{\Theta}([\mathbf{y}|\alpha][\alpha])d\alpha$,
- calcul des distributions marginales si θ est multivarié,
- modèles à variables latentes...

Approche bayésienne : estimation

Problème $\rightarrow$ multiples intégrales à calculer :

- $[\theta|\mathbf{y}] = [\theta][\mathbf{y}|\theta]/[\mathbf{y}]$, avec $[\mathbf{y}] = \int_\Theta([\mathbf{y}|\alpha][\alpha])d\alpha$,
- calcul des distributions marginales si θ est multivarié,
- modèles à variables latentes...

Approche bayésienne : estimation

Problème $\rightarrow$ multiples intégrales à calculer :

- $[\theta|\mathbf{y}] = [\theta][\mathbf{y}|\theta]/[\mathbf{y}]$, avec $[\mathbf{y}] = \int_\Theta([\mathbf{y}|\alpha][\alpha])d\alpha$,
- calcul des distributions marginales si θ est multivarié,
- modèles à variables latentes...

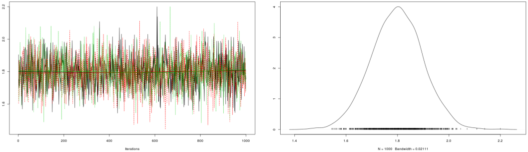
Utilisation de **méthodes numériques stochastiques** permettant de **générer un échantillon issu de la distribution *a posteriori***.

Approche bayésienne : estimation

Problème $\rightarrow$ multiples intégrales à calculer :

- $[\theta|\mathbf{y}] = [\theta][\mathbf{y}|\theta]/[\mathbf{y}]$, avec $[\mathbf{y}] = \int_{\Theta}([\mathbf{y}|\alpha][\alpha])d\alpha$,
- calcul des distributions marginales si θ est multivarié,
- modèles à variables latentes...

Utilisation de **méthodes numériques stochastiques** permettant de **générer un échantillon issu de la distribution *a posteriori***.



Modèle hiérarchique

Processus d'observation

Processus biologique

Paramètres

Hyper Paramètres

Exemple non hiérarchique

$$\underbrace{[\theta|y]}_{\text{a posteriori}} \propto \underbrace{[\theta]}_{\text{a priori}} \quad \underbrace{[y|\theta]}_{\text{vraisemblance}}$$

Exemple non hiérarchique

$$\underbrace{[\theta_1, \theta_2 | x, y]}_{\text{a posteriori}} \propto \underbrace{[\theta_1][\theta_2]}_{\text{a priori}} \quad \underbrace{[y | \theta_1, \theta_2, x]}_{\text{vraisemblance}}$$

Exemple non hiérarchique

$$\underbrace{[\theta_1, \theta_2|x, y]}_{\text{a posteriori}} \propto \underbrace{[\theta_1][\theta_2]}_{\text{a priori}} \quad \underbrace{[y|\theta_1, \theta_2, x]}_{\text{vraisemblance}}$$

Exemple : régression linéaire classique

$$\begin{cases} \textbf{A priori} \\ \theta_2 \sim \mathcal{U}(0, 10), \\ \theta_1 \sim \mathcal{N}(0, 1000), \\ \textbf{Vraisemblance} \\ y_i \sim \mathcal{N}(\theta_1 x_i, \theta_2^2). \end{cases}$$

Exemple non hiérarchique

Processus d'observation

Processus biologique

Paramètres

Hyper Paramètres

Exemple non hiérarchique

Modèle non hiérarchique

Processus d'observation

Processus biologique

$$\left[y \middle| \theta_1, \theta_2, x \right]$$

Paramètres

Hyper Paramètres

Un premier cas hiérarchique : la variabilité individuelle

Variabilité additionnelle : le paramètre θ_1 varie entre individus ou sites.

Un premier cas hiérarchique : la variabilité individuelle

Variabilité additionnelle : le paramètre θ_1 varie entre individus ou sites.

$$\begin{array}{rcll} [\theta_1, \theta_2, \theta_3 | x, y] & \propto & [y|\theta_1, \theta_2, x] & \text{vraisemblance} \\ & \times & [\theta_1|\theta_3, \theta_4][\theta_2] & \text{a priori} \\ & \times & [\theta_3][\theta_4] & \text{hyper a priori} \end{array}$$

Un premier cas hiérarchique : la variabilité individuelle

$$[\theta_1, \theta_2, \theta_3 | x, y] \propto [y|\theta_1, \theta_2, x][\theta_1|\theta_3, \theta_4][\theta_2][\theta_3][\theta_4]$$

Exemple : régression linéaire mixte

$$\begin{cases} \textbf{hyper a priori} \\ \theta_4 \sim \mathcal{U}(0, 10), \\ \theta_3 \sim \mathcal{N}(0, 1000), \\ \textbf{A priori} \\ \theta_2 \sim \mathcal{U}(0, 10), \\ \theta_{1,i} \sim \mathcal{N}(\theta_3, \theta_4^2), \\ \textbf{Vraisemblance} \\ y_i \sim \mathcal{N}(\theta_{1,i} x_i, \theta_2^2). \end{cases}$$

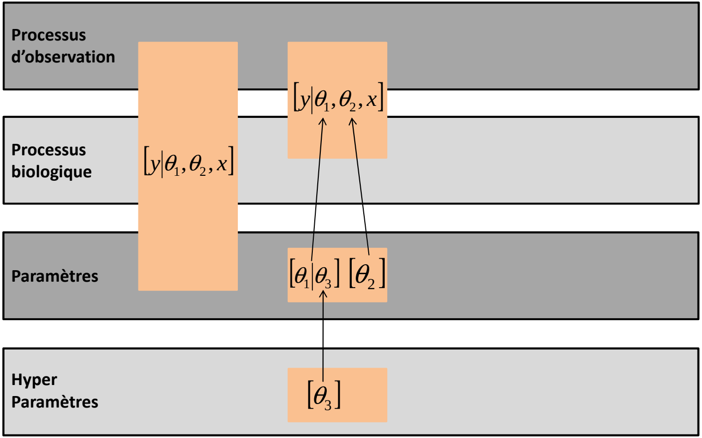
Un premier cas hiérarchique : la variabilité individuelle

Modèle non hiérarchique

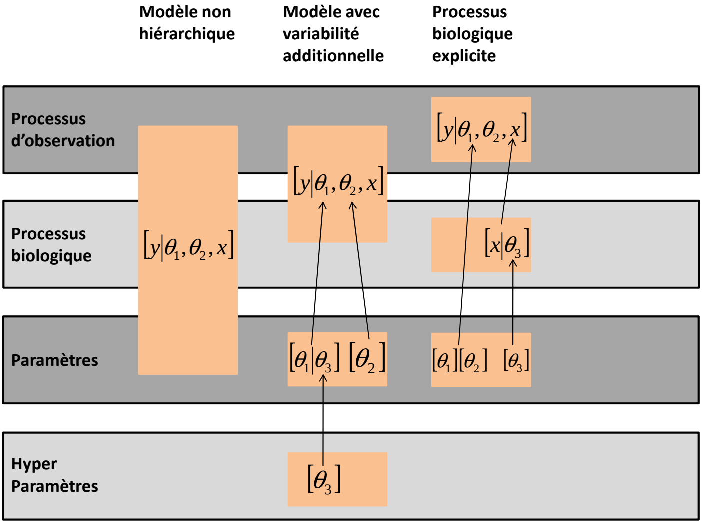
Processus d'observation	
Processus biologique	$\left[y \mid \theta_1, \theta_2, x\right]$
Paramètres	
Hyper Paramètres	

Un premier cas hiérarchique : la variabilité individuelle

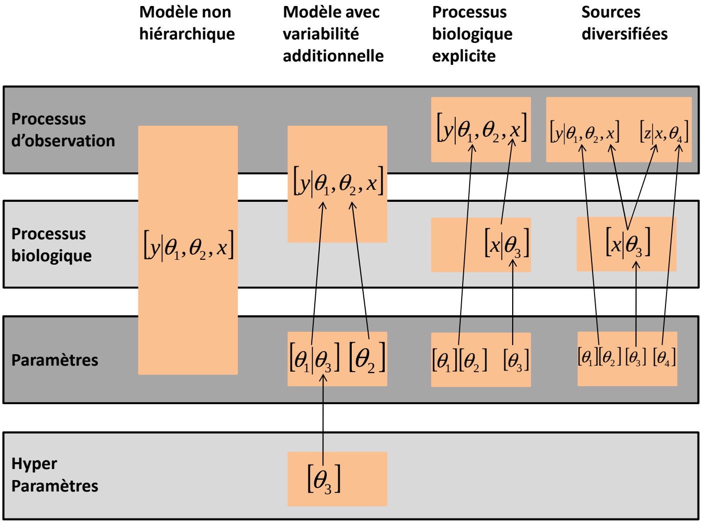
Modèle non hiérarchique | **Modèle avec variabilité additionnelle**



Modèle avec processus identifié



Intégration de différentes sources de données



D'après Clark, 2005

Estimation des distributions a posteriori à l'aide de méthodes numériques

Pourquoi utiliser des méthodes numériques ?

- Ce que l'on cherche à quantifier en bayésien :
  - la loi a posteriori

$$f(\theta \mid Y) = \frac{f(Y \mid \theta)\pi(\theta)}{\int_\Theta f(Y \mid \alpha)\pi(\alpha)d\alpha}$$

  - et ses caractéristiques : moments a posteriori, maximum a posteriori, quantiles a posteriori, intervalles de crédibilité...

- Exemple : Nombre de succès sur n essais indépendants
  - modèle binomial : $Y \mid \theta \sim \text{Binomiale}(n, \theta)$
  - prior beta : $\theta \sim \text{Beta}(a, b)$
  - posterior [1] : $\theta \mid Y \sim \text{Beta}(a + Y, b + n - Y)$

1. Détail du calcul :

$$f(\theta \mid Y) = \frac{C_n^Y \theta^Y (1-\theta)^{n-Y} \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}}{\int_0^1 C_n^Y \alpha^Y (1-\alpha)^{n-Y} \frac{\alpha^{a-1}(1-\alpha)^{b-1}}{B(a,b)} d\alpha}$$

$$= \frac{\theta^{(a+Y)-1}(1-\theta)^{(b+n-Y)-1}}{B(a+Y, b+n-Y)}$$

Pourquoi utiliser des méthodes numériques ?

- Ce que l'on cherche à quantifier en bayésien :
  - la loi a posteriori

$$f(\theta \mid Y) = \frac{f(Y \mid \theta)\pi(\theta)}{\int_\Theta f(Y \mid \alpha)\pi(\alpha)d\alpha}$$

  - et ses caractéristiques : moments a posteriori, maximum a posteriori, quantiles a posteriori, intervalles de crédibilité...

- Mais la loi a posteriori peut être difficilement calculable
  - Modèle avec nombreux paramètres et variables latentes :

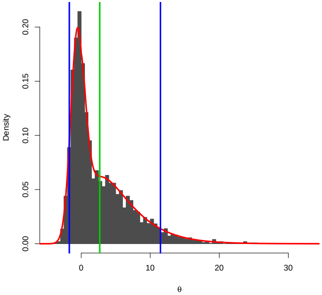
$$f(\theta_1, \ldots, \theta_K \mid Y) = \frac{f(Y \mid \theta_1, \ldots, \theta_K)\pi(\theta_1, \ldots, \theta_K)}{\int_{\Theta_1} \cdots \int_{\Theta_K} f(Y \mid \theta_1, \ldots, \theta_K)\pi(\alpha_1, \ldots, \alpha_K)d\alpha_1 \cdots d\alpha_K}$$

  - Les intégrales multiples (grande dimension) rendent difficile le calcul de la posterior jointe $f(\theta_1, \ldots, \theta_K \mid Y)$, des posteriors marginales $f(\theta_k \mid Y)$, des moments a posteriori $E(\theta_i^q \mid Y)$...

Pourquoi utiliser des méthodes numériques ?

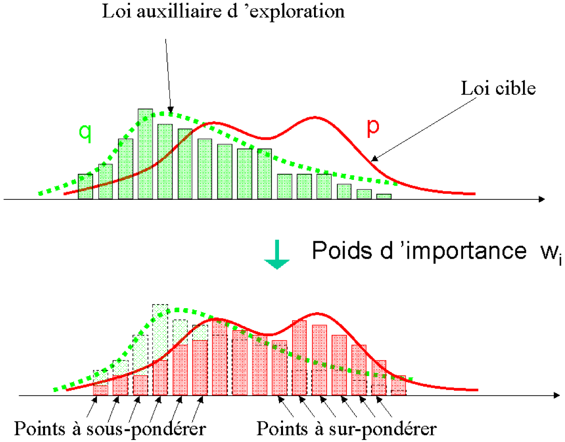
- Ce que peuvent les méthodes numériques :
  - générer un échantillon issu de la loi a posteriori
  - sans passer par le calcul d'intégrales multiples

- A quoi sert cet échantillon ?
  - connaître intimement la loi a posteriori
  - estimer la densité a posteriori
  - estimer les moments a posteriori ex : $E(\theta \mid Y) \approx \frac{1}{I}\sum_{i=1}^{I} \theta^{(i)}$
  - estimer des intervalles de crédibilité
  - ...



Ré-échantillonnage (sampling importance resampling)

Méthode numérique basé sur l'échantillonnage d'importance :



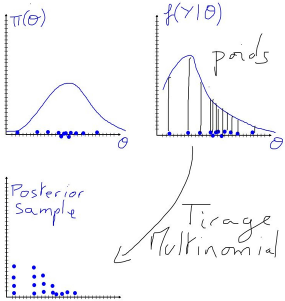
Extrait de Parent et Bernier (2007)

Ré-échantillonnage (sampling importance resampling)

Algorithme

1. Tirer un I-échantillon $\{\theta^{(1)}, \ldots, \theta^{(I)}\}$ dans la prior π
2. Attribuer à chaque $\theta^{(i)}$ le poids $w_i = \frac{f(Y|\theta^{(i)})}{\sum_{j=1}^{I} f(Y|\theta^{(j)})}$
3. Tirer avec remise (tirage multinomial) J valeurs dans le I-échantillon $\{\theta^{(1)}, \ldots, \theta^{(I)}\}$ avec les probabilités $w_1, \ldots, w_I$

Illustration des contributions de la prior et de la vraisemblance dans la méthode de ré-échantillonnage :



Ré-échantillonnage (sampling importance resampling)

Algorithme

1. Tirer un I-échantillon $\{\theta^{(1)}, \dots, \theta^{(I)}\}$ dans la prior π
2. Attribuer à chaque $\theta^{(i)}$ le poids $w_i = \frac{f(Y|\theta^{(i)})}{\sum_{j=1}^{I} f(Y|\theta^{(j)})}$
3. Tirer avec remise (tirage multinomial) J valeurs dans le I-échantillon $\{\theta^{(1)}, \dots, \theta^{(I)}\}$ avec les probabilités $w_1, \dots, w_I$

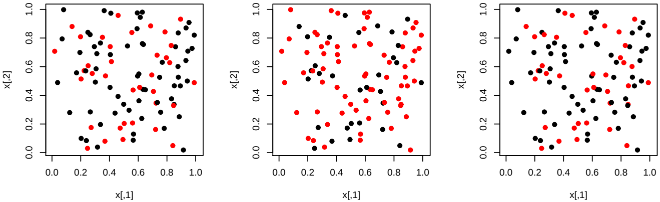
Amélioration possible de l'algorithme

- **en tirant l'I-échantillon dans une loi auxiliaire g**
- **et en corrigeant les poids w_i**

Ré-échantillonnage : Appli à un modèle métapopulationnel

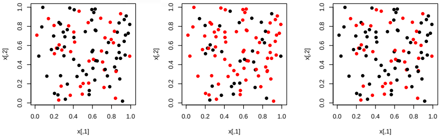
Données de présence/absence d'un pathogène

- sur un ensemble de sites hôtes $x_1, \ldots, x_n$
- à trois dates successives successives $t = 1$, $t = 2$ et $t = 3$



Noir : site vide, $Y(x,t) = 0$
Rouge : site occupé, $Y(x,t) = 1$

Ré-échantillonnage : Appli à un modèle métapopulationnel (suite)



Noir : site vide, $Y(x,t) = 0$
Rouge : site occupé, $Y(x,t) = 1$

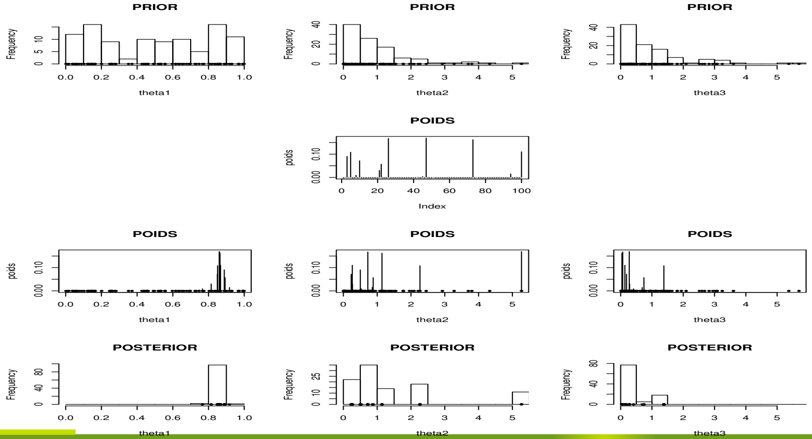
- Probabilités de transitions :
  - $1 \to 0$: θ_1 (probabilité d'extinction)
  - $1 \to 1$: $1 - \theta_1$
  - $0 \to 1$: $1 - \exp\{-S(x,t)\}$ (probabilité de colonisation)
  - $0 \to 0$: $\exp\{-S(x,t)\}$

où $S(x,t) = \sum_{j=1}^{n} Y(x_j, t-1) \frac{\theta_2}{2\pi\theta_3^2} \exp\left(-\frac{||x - x_j||}{\theta_3}\right)$

- Lois a priori :
  - $\theta_1 \sim Uniforme(0,1)$
  - $\theta_2 \sim Exponentielle(1)$
  - $\theta_3 \sim Exponentielle(1)$

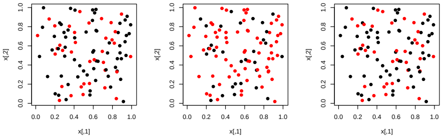
Ré-échantillonnage : Appli à un modèle métapopulationnel (suite)

Etapes de l'algorithme SIR (100 jeux de paramètres simulés) :

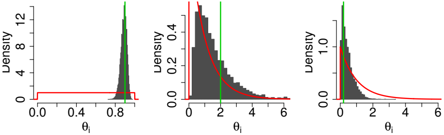


Ré-échantillonnage : Appli à un modèle métapopulationnel (suite)

- Données (noir : vide, $Y(x,t) = 0$; rouge : occupé, $Y(x,t) = 1$) :

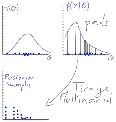


- Distributions a posteriori obtenues par la méthode de ré-échantillonnage (10^5 jeux de paramètres simulés) :

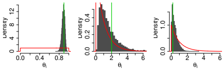


Ré-échantillonnage : Limites

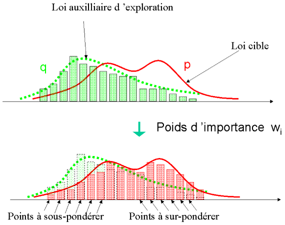
- Doublons dans l'échantillon a posteriori



- Beaucoup de vecteurs de paramètres θ générés peuvent ne pas être retenus dans l'échantillon a posteriori ("rendement" faible)

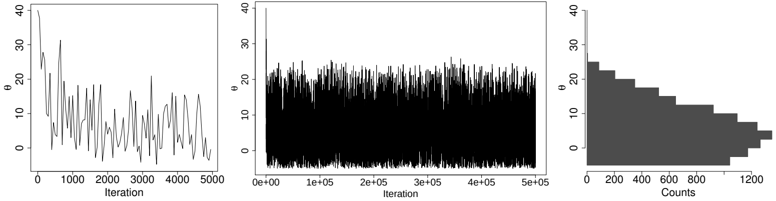


- Problème de la loi auxilliaire d'exploration...



MCMC : Présentation

- Méthodes de Monte Carlo par Chaînes de Markov
- Algorithmes utilisés dans WinBugs, OpenBUGS, Jags...
- Algorithmes séquentiels : une séquence de réalisations dépendantes (i.e. une chaîne) de θ est générée
- Exploration ciblée de l'espace des paramètres (et des variables latentes)
- Qu'est-ce qu'une chaîne ?

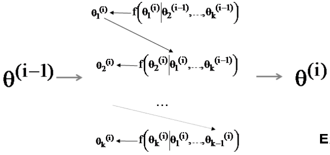


Algorithme

0. Initialisation : donner des valeurs initiales aux K composantes de θ : $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_K^{(0)})$

i. A l'itération $i \in \{1, \ldots, I\}$:

- Générer $\theta_1^{(i)}$ selon la loi $f(\theta_1 \mid Y, \theta_2^{(i-1)}, \ldots, \theta_K^{(i-1)})$
- Générer $\theta_2^{(i)}$ selon la loi $f(\theta_2 \mid Y, \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_K^{(i-1)})$
- Générer $\theta_3^{(i)}$ selon la loi $f(\theta_3 \mid Y, \theta_1^{(i)}, \theta_2^{(i)}, \theta_4^{(i-1)}, \ldots, \theta_K^{(i-1)})$
- $\ldots$
- Générer $\theta_K^{(i)}$ selon la loi $f(\theta_K \mid Y, \theta_1^{(i)}, \ldots, \theta_{K-1}^{(i)})$



Extrait de Parent et Bernier (2007)

MCMC : Algorithme de Metropolis-Hastings

- Généralisation du MCMC–Gibbs
- Une loi de proposition arbitraire remplace la loi conditionnelle
- La mise à jour des paramètres n'est pas systématique (étape d'acceptation-rejet)
- La loi de proposition intervient dans la proba de mise à jour

MCMC : Algorithme de Metropolis-Hastings (suite)

Algorithme

0. Initialisation : donner des valeurs initiales aux K composantes de θ : $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_K^{(0)})$

$i, k.$ A l'itération $i \in \{1, \dots, I\}$, pour chaque k :

- Générer θ_k^{cand} selon la loi de proposition $g(\theta_k \mid \theta_k^{(i-1)})$
- Mettre à jour le paramètre $(\theta_k^{(i)} = \theta_k^{cand})$ avec la probabilité[2] :

$$\min\left[1, \frac{Posterior(\theta_k^{cand})g(\theta_k^{(i-1)} \mid \theta_k^{cand})}{Posterior(\theta_k^{(i-1)})g(\theta_k^{cand} \mid \theta_k^{(i-1)})}\right]$$

$$= \min\left[1, \frac{(Vraisemblance \times Prior)(\theta_k^{cand})g(\theta_k^{(i-1)} \mid \theta_k^{cand})}{(Vraisemblance \times Prior)(\theta_k^{(i-1)})g(\theta_k^{cand} \mid \theta_k^{(i-1)})}\right]$$

- Ne pas mettre à jour $(\theta_k^{(i)} = \theta_k^{(i-1)})$ sinon

---

2. Expression de la probabilité de mise à jour :

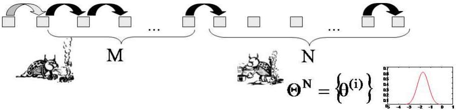
$$\min\left[1, \frac{f(Y \mid \theta_{1:k-1}^{(i)}, \theta_k^{cand}, \theta_{k+1:K}^{(i-1)})\pi(\theta_{1:k-1}^{(i)}, \theta_k^{cand}, \theta_{k+1:K}^{(i-1)})g(\theta_k^{(i-1)} \mid \theta_k^{cand})}{f(Y \mid \theta_{1:k-1}^{(i)}, \theta_k^{(i-1)}, \theta_{k+1:K}^{(i-1)})\pi(\theta_{1:k-1}^{(i)}, \theta_k^{(i-1)}, \theta_{k+1:K}^{(i-1)})g(\theta_k^{cand} \mid \theta_k^{(i-1)})}\right]$$

MCMC : Algorithme de Metropolis-Hastings (suite)

- Connaissance nécessaire de la loi cible (la loi a posteriori) qu'à une constante près

$$f(\theta_1, \ldots, \theta_K \mid Y) \propto f(Y \mid \theta_1, \ldots, \theta_K)\pi(\theta_1, \ldots, \theta_K)$$

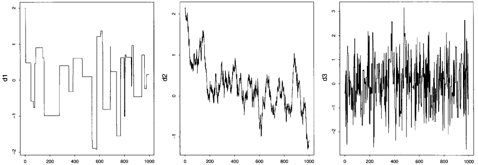
- Mise à jour par bloc
- Algorithmes hybrides combinant Gibbs et M-H
- Comme dans le MCMC–Gibbs, période de chauffe (burn-in) et régime de croisière (stationnaire)



- Contrairement au MCMC–Gibbs, choix des lois de proposition (paramètres de réglages, tuning)
- Question de la convergence de la chaîne vers son régime de croisière (vrai aussi pour le MCMC–Gibbs)

MCMC : Choix des lois de proposition g

- La rapidité de convergence de la chaîne dépend du choix des lois de proposition (formes et valeurs des paramètres)



(a) Proposal variance too large (b) Proposal variance too small (c) Proposal variance approximately optimised

Extrait de Roberts and Rosenthal (2001)

- Tirage indépendant de la valeur courante du paramètre

  ex1 : $g(\theta \mid \theta^{(i-1)})$ est la densité d'une $Normale(\mu, \sigma^2)$

  ex2 : $g(\theta \mid \theta^{(i-1)})$ est la densité d'une $Gamma(\alpha, \beta)$

- Marche aléatoire homogène

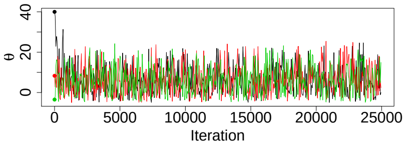
  ex1 : $g(\theta \mid \theta^{(i-1)})$ est la densité d'une $Normale(\theta^{(i-1)}, \sigma^2)$

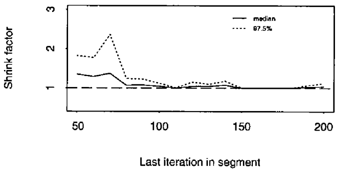
  ex2 : $g(\theta \mid \theta^{(i-1)})$ est la densité d'une $Gamma((\frac{\theta^{(i-1)}}{\sigma})^2, \frac{\sigma^2}{\theta^{(i-1)}})$ tq $E(\theta \mid \theta^{(i-1)}) = \theta^{(i-1)}$ et $Var(\theta \mid \theta^{(i-1)}) = \sigma^2$

- Taux de mise à jour à viser (par essai–erreur) : 25%
- Méthodes adaptatives

MCMC : Diagnostics de convergence des chaînes

- Quelle convergence ?
  - Convergence vers la stationarité
  - Convergence des moyennes empiriques
    La chaîne a-t-elle exploré toutes les facettes de la distribution cible ?
  - Convergence vers un échantillonnage i.i.d.
- Un exemple de diagnostic : la méthode de Gelman–Rubin qui est basée
  - sur plusieurs chaînes initialisées différemment
  - sur les variances inter-chaîne et intra-chaîne



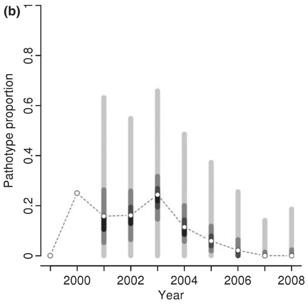


MCMC : Qualité de l'ajustement

répliquer/simuler les données

Principe :
A chaque itération, une donnée est simulée selon le modèle supposé et les valeurs courantes des paramètres

$\Rightarrow$ simulation d'un jeux de données conforme aux hypothèses faites.

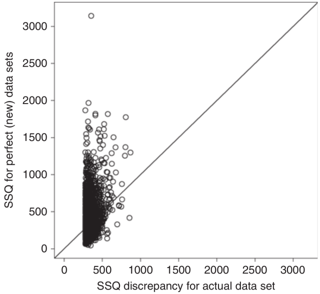


MCMC : Qualité de l'ajustement

Bayesian p-Values

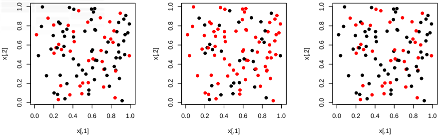
L'idée : comparer le défault d'ajustement du modèle pour les vraies données avec celui du modèle pour les données répétées.

- Choix de la mesure de distance en fonction de ce que l'on veut quantifier (valeurs extrèmes...).
- Problème : les données sont utilisées deux fois, pour générer les données répliquées et dans la comparaison.
- Méthode peu discréminante, non conseillée pour faire du choix de modèle (facteurs de Bayes).



MCMC : Appli à un modèle métapopulationnel (suite)

- Données (noir : vide, $Y(x,t) = 0$; rouge : occupé, $Y(x,t) = 1$) :



- MCMC hybride :
  - Gibbs pour le paramètre de survie θ_1 :

$$\theta_1 \mid Y, \theta_2, \theta_3 \sim Beta(1 + \sum_{x,t}[Y_{xt} : 1 \to 0], 1 + \sum_{x,t}[Y_{xt} : 1 \to 1])$$

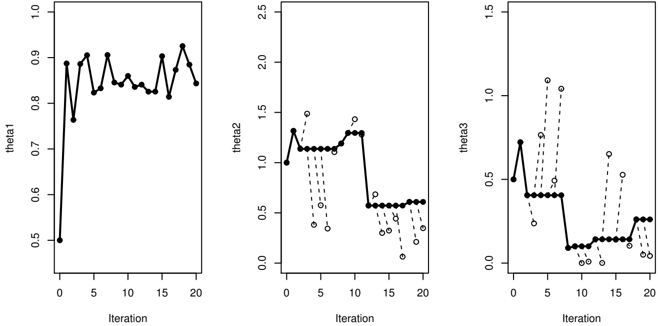
  - Metropolis-Hastings par bloc pour les paramètres de colonisation (θ_2, θ_3) avec pour lois de propositions :

$$\theta_2 \mid \theta_2^{(i-1)} \sim Gamma\left(\left(\frac{\theta_2^{(i-1)}}{0.5}\right)^2, \frac{0.5^2}{\theta_2^{(i-1)}}\right)$$

$$\theta_3 \mid \theta_3^{(i-1)} \sim Gamma\left(\left(\frac{\theta_3^{(i-1)}}{0.2}\right)^2, \frac{0.2^2}{\theta_3^{(i-1)}}\right)$$

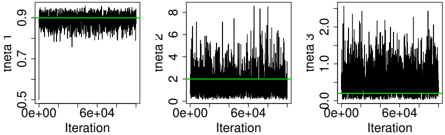
MCMC : Appli à un modèle métapopulationnel (suite)

20 premières itérations de l'algorithm MCMC hybride :

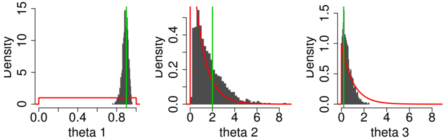


MCMC : Appli à un modèle métapopulationnel (suite)

- Chaînes :



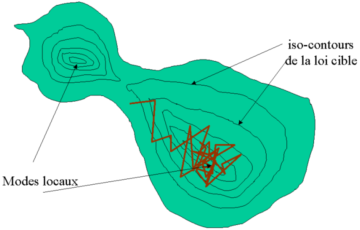
- Distributions a posteriori :



Conclusions

MCMC : Difficultés d'utilisation

- Paramètres de réglage
  - longueur de la période de chauffe (Gibbs et M–H)
  - longueur de la chaîne (Gibbs et M–H)
  - sous-échantillonnage de la chaîne (Gibbs et M–H)
  - formes et paramètres des lois de propositions (M–H)
- Dimension élevée
- Difficulté d'explorer toute la loi a posteriori
  - Ex. des lois a posteriori multimodales avec faibles probas entre les modes
  - Problème de capture autour des modes locaux
  - Comment améliorer l'exploration ?



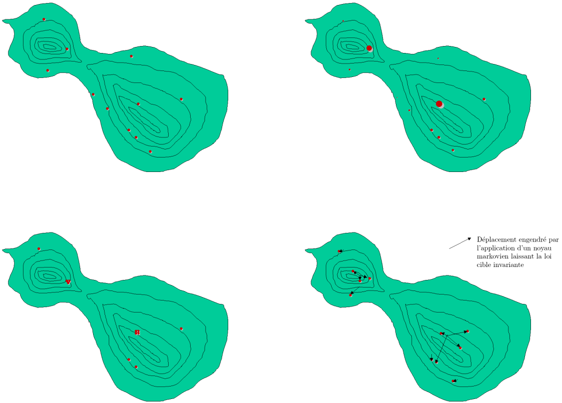
Extrait de Parent & Bernier (2007)

Les outils et logiciels

- INLA (Integrated Nested Laplace Approximations) : R-INLA
- MCMC (Markov Chain Monte Carlo) : WinBUGS, OpenBUGS, JAGS, Nimble
- HMC (Hamiltonian Monte Carlo) : R-Stan
- PMC (Population Monte Carlo) : BIIPS, Nimble

Méthode particulaire : Principe

Simuler N chaînes de Markov en parallèle, en éliminant celles loin du mode et en multipliant celles qui sont proches



Extrait de Parent & Bernier (2007)

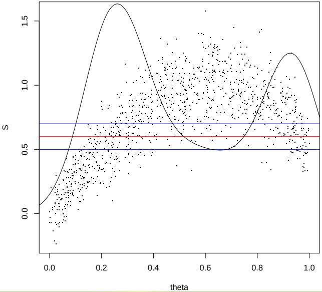
ABC : Principe

- Difficulté d'évaluer la distribution a posteriori

$$f(\theta_1, \ldots, \theta_K \mid Y) = \frac{f(Y \mid \theta_1, \ldots, \theta_K)\pi(\theta_1, \ldots, \theta_K)}{\int_{\Theta_1} \cdots \int_{\Theta_K} f(Y \mid \theta_1, \ldots, \theta_K)\pi(\alpha_1, \ldots, \alpha_K) d\alpha_1 \cdots d\alpha_K}$$

- Ré-échantillonnage, MCMC Gibbs ou M–H, Méthode particulaire : il suffit de pouvoir calculer
  - la vraisemblance $f(Y \mid \theta_1, \ldots, \theta_K)$
  - les conditionnelles $f(\theta_k \mid Y, \theta_{-k})$
- Mais si ces calculs ne sont pas possibles ?
  (exemple des modèles de coalescence)

Avec le calcul bayésien approché (ABC) : il suffit de pouvoir simuler Y sachant $\theta_1, \ldots, \theta_K$



Merci !