# Modélisation bayésienne hiérarchique et inférence numérique


Fabien Campillo

projet MERE

Inria Méditerranée, Montpellier

Vivien Rossi

Dynamique des forêts naturelles

Cirad Montpellier

# pourquoi cet exposé

- nombreux domaines : environnement, écologie, biologie, sciences sociales, robotique, traitement d'image, etc.
- "Bayesian Computation : A Statistical revolution", "The Bayesian revolution in genetics", "A Bayesian revolution in spectral analysis"...
- livres récents en environnement :
  - Clark. Models for Ecological Data, 2007
  - McCarthy. Bayesian Methods for Ecology, 2007
  - Clark, Gelfand (eds). Hierarchical Modelling for the Environmental Sciences, 2006
  - Le, Zidek. Statistical Analysis of Environmental Space-Time Processes, 2006
- explosion en 15 ans

# pourquoi cet exposé (suite)

- James S. Clark – Why environmental scientists are becoming Bayesians – Ecology Letters, 2005
- Tout le monde, notamment en environnement, devient-il "bayésien" ? Si oui, pourquoi et comment ?

Pour s'intéresser à cette question, on visite les points suivants :

- inférence bayésienne
  - formule de Bayes
  - a priori / a posteriori
- modélisation (adaptée à l'analyse) bayésienne
  - modèles bayésiens hiérarchiques (HBM)
  - modèles de Markov cachés (HMM)
- approximation numérique (adaptée à l'analyse) bayésienne
  - Monte Carlo (MC)
  - Monte Carlo par chaîne de Markov (MCMC)
  - Monte Carlo séquentiel (SMC), i.e. filtrage particulaire
- Exemples approfondis
  - MCMC
  - SMC

# théorie des probabilités et Bayes

- origines de la théorie des probabilités : jeux des hasard (prédiction des gains) – formalisée par Pascal et Fermat (1660)
- Bayes, "Problem in the Doctrine of Chances", (1763)

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

# théorie des probabilités et Bayes

Le révérend Thomas Bayes ( ?...)

- indépendamment démontré par Laplace (1774)
- Bayes n'aurait peut-être pas été bayésien !

# problem in the doctrine of chances

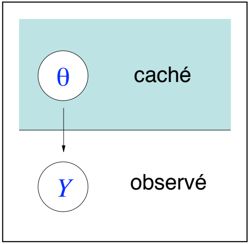
- Comment mettre à jour (update) une croyance reconnue (a priori) lorsque de nouvelles informations sont disponibles ? (par exemple au travers d'une nouvelle expérience)

- une formule qui est au centre du paradigme bayésien
- représentation des connaissances probabiliste

# inférence bayésienne

- soient $Y$ et $\theta$ deux quantités aléatoires

| | | | |
|---|---|---|---|
| $Y$ | : un effet | : une observation | : valeur connue $Y = y$ |
| $\theta$ | : une cause | : un état latent | : inconnu |

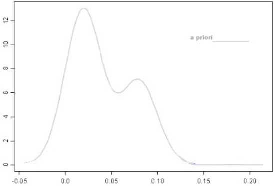



le plus simple réseau bayésien hiérarchique !

- inférer $\theta$ à partir de l'observation $Y = y$

# représentation de la connaissance probabiliste

- représentation de la connaissance sur $\theta$ :
  - avant observation $r \mapsto \pi_\theta(r)$ (a priori)




  - après observation $r \mapsto \pi_{\theta|Y}(r|y)$ (a posteriori)

?

- but : calculer la densité a posteriori

# formule de Bayes

## formule de Bayes

$$\pi_{\theta|Y}(r|y) = \frac{\pi_{Y|\theta}(y|r)\ \pi_\theta(r)}{\pi_Y(y)} = \frac{\pi_{Y|\theta}(y|r)\ \pi_\theta(r)}{\int \pi_{Y|\theta=r'}(y)\ \pi_\theta(r')\,dr'}$$

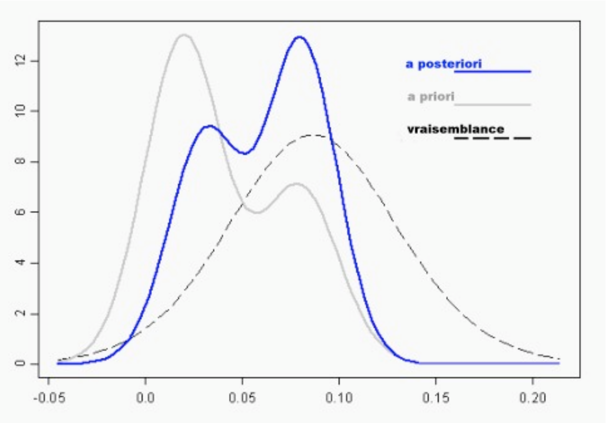
où encore :

$$\pi_{\theta|Y}(r|y) \propto \pi_{Y|\theta}(y|r) \times \pi_\theta(r)$$

$$\text{“a posteriori”} \propto \text{“vraisemblance”} \times \text{“a priori”}$$

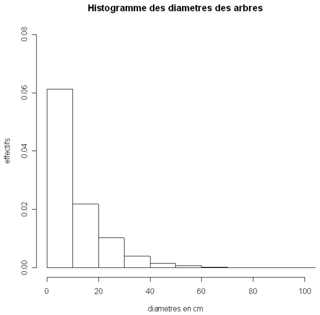
comme fonction de $r$ à $y$ fixé ($\propto$ signe de proportionnalité)

# formule de Bayes



# exemple

- répartition diamétrique des arbres d'une forêt naturelle




on mesure le diamètre de $n$ arbres $Y = (Y_1, \dots, Y_n)$

- $Y_i$ sont i.i.d. et suivent une loi exponentielle de paramètre $\theta$ inconnu, i.e. $\pi_{Y_i|\theta}(y_i|r) = r\,\exp(-r\,y_i)$
- a priori $\pi_\theta(r)$

# exemple (suite)

- vraisemblance

$$\pi_{Y_1,\ldots,Y_n|\theta}(y_1,\ldots,y_n|r)=\prod_{i=1}^{n}\pi_{Y_i|\theta}(y_i|r)=r\,\exp\big(-r\sum_{i=1}^{n}y_i\big)$$

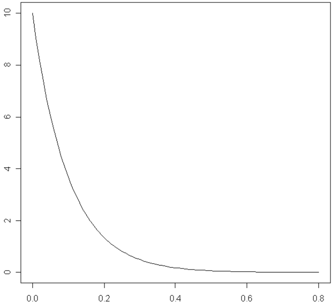
- loi a posteriori

$$\pi_{\theta|Y_1,\ldots,Y_n}(r|y_1,\ldots,y_n)\propto r\,\exp\big(-r\sum_{i=1}^{n}y_i\big)\,\pi_\theta(r)$$

# exemple (suite)

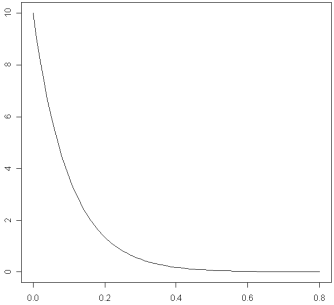
- il faut choisir une loi a priori pour $\theta$ :
  - loi quelconque $\longrightarrow$ loi a posteriori inaccessible
  - loi conjuguée $\longrightarrow$ loi a posteriori connue

- la loi Gamma est conjuguée à la loi exponentielle



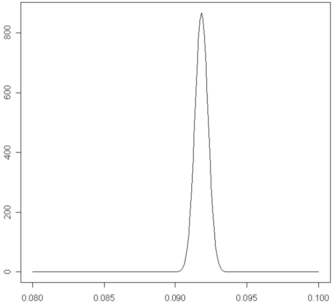

$$\pi_\theta = Ga(1, 10)$$

# exemple (suite)




densité a priori

$\pi_\theta = Ga(1, 10)$




densité a posteriori

$\pi_{\theta|Y_1,\ldots,Y_n} = Ga(1 + 39858, 10 + 434131.1)$

# remarques

- La logique inverse de l'approche bayésienne : conditionner par les résultats et évaluer les causes à partir des effets.
- Le processus d'intégration séquentiel des connaissances dans l'approche bayésienne (a priori $\rightarrow$ a posteriori) est très adapté aux sciences expérimentales ainsi qu'à l'informatique (tâches séquentielles).
- fournit un cadre conceptuel pour la prise en compte de l'aléa dans la prise de décision et l'inférence
- va au delà des statistiques : c'est aussi une théorie de l'apprentissage

# difficultés

- Double contrainte : recherche des lois a priori qui soient pertinentes pour le problème et conjuguées (i.e. telle que les lois a posteriori aient une forme explicite).
- La plupart des lois a priori pertinentes ne sont pas conjuguées.
- La plupart du temps, l'analyse bayésienne explicite n'est pas possible

$\longrightarrow$ approximation numérique

# subjectivité *vs* objectivité ?

- Cette approche implique une relation dialectique entre le statisticien et le spécialiste du domaine qui détermine la pertinence des a priori utilisées.
- L'approche bayésienne est souvent décrite comme (ou accusée d'être) subjective dans la mesure où elle s'opposerait à une analyse scientifique qui devrait nécessairement être objective.
- Le problème est que deux personnes analysant les mêmes nouvelles informations mais partant d'a priori différents peuvent aboutir à des résultats divergents.
  - La prise en compte d'a priori nuit à l'objectivité de l'analyse ?
  - Interprétation subjective des données ?
- le spécialiste peut influencer en sa faveur les expérimentations, l'approche bayésienne ne peut pas le détecter...

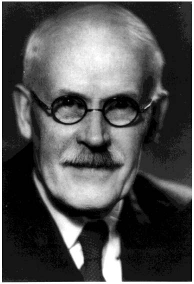
# début 20ème : l'approche subjective

Frank P. Ramsey

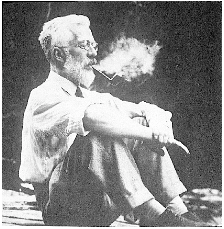
Bruno de Finetti

- Frank P. Ramsey «Truth and Probability» (1926) écrit en opposition au traité de probabilité de John Maynard Keynes.
- Bruno de Finetti, «La prévision : ses lois logiques, ses sources subjectives» (1937)

# bayésiens vs fréquentistes



Harold Jeffreys



Ronald A. Fisher

# bayésiens vs fréquentistes (suite)

- Dans les années 30, ce « risque de manque d'objectivité » a conduit une école de statisticiens (Fisher-Pearson-Neyman) a proposer de nouveaux outils dits non-bayésiens ou fréquentistes.
- Avec cette approche, des données issues de nombreuses expérimentations (réalisées dans un cadre strict) donneront les mêmes résultats quel que soit le statisticien qui traitera ces données.
- Ces méthodes sont-elles plus objectives ? Peut-on réellement les mettre en œuvre sans faire d'hypothèse a priori ?
- À cette époque (sans ordinateur), ces méthodes étaient plus facile à mettre en œuvre en pratique.

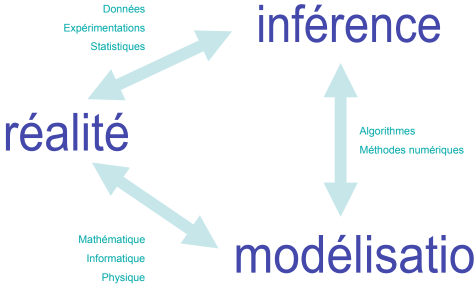
# bayésiens vs fréquentistes (suite)

- Les probabilités se fondent sur une vision expérimentale classique : il est possible de répéter indéfiniment et indépendamment la même expérience.
  - Les théorèmes fondamentaux des probabilités (loi des grands nombres, théorème central limite) s'appliquent
  - l'analyse statistique fréquentiste aussi : un estimateur fondé sur une infinité d'expériences indépendantes et identiques possède toutes les bonnes propriétés de convergence.
- Mais que faire face aux nombreuses situations n'entrant pas dans ce cadre : le changement global, les épidémies, la pauvreté (et beaucoup d'exemples de biologie) ? Souhaite-t-on (ou même peut-on) seulement renouveler les expériences correspondantes ? Que faire lorsqu'on dispose de peu de données ?

# bayésiens vs fréquentistes (suite)

- années 30 : axiomatisation des probabilités par Kolmogorov et avènement des « statistiques mathématiques »
- Les non bayésiens ont accaparé les termes «mathématique» et «objectivité» sous-entendant que les bayésiens...
- Les bayésiens pointent les faiblesses de l'approche fréquentiste : quel sens donner à la probabilité d'une expérience non renouvelable ?
- Argument "vachard" des non-bayésiens : admettons quand même que l'approche bayésienne est vraie, de toute façon on ne peut pas l'utiliser en pratique.
- L'inférence bayésienne a été en grande partie ignorée en faveur des approches fréquentistes en raison de faiblesses conceptuelles ( ?) et de difficultés de mise en œuvre.
- Ces faiblesses n'amenuisent. Pragmatisme : Elles s'amenuisent surtout grâce au succès rencontré par ces méthodes en pratique.

- Il est illusoire de vouloir faire de l'inférence sans modélisation. Les problèmes complexes nécessitent l'élaboration de modèles a priori (i.e. avant expérience) fondés sur la connaissance des spécialistes.




- De plus il faut de modèles pour gérer les données qui arrivent...

- Buts de la modélisation :
  (i) pour la compréhension
  (ii) pour la décision
- Il est toujours préférable d'étudier les deux de front :
  - Travailler sur (ii) sans travailler sur (i) : risques de toujours manquer de recul et de ne pas engager de réflexions scientifiques de profondeur.
  - Travailler sur (i) sans travailler sur (ii) : risques de perdre contact avec la réalité (et de ne plus voir de données. . .).
- Les méthodes bayésiennes
  - ont du succès dans les domaines « sans lois » : la biologie et les sciences sociales (il n'y pas d'équivalent des lois de la thermodynamique ou de la mécanique).
  - sont également adaptées aux domaines ne permettant d'accéder qu'à peu de données (c'est le cas de l'écologie et de l'environnement), il est alors nécessaire de proposer des modèles afin de guider les méthodes d'inférence.
  - peuvent aider dans les deux cas (i) et (ii).

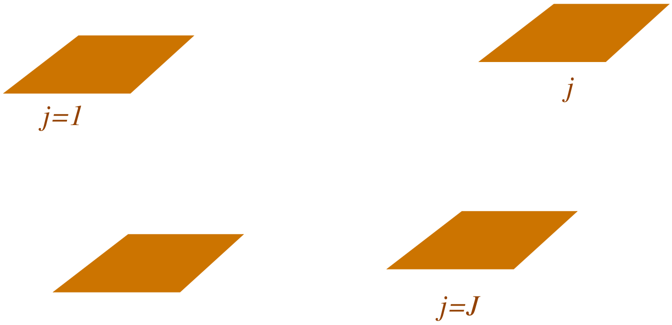
- Pourquoi ces modèles sont dits « bayésiens » ? Parce ce qu'ils sont adaptés à l'inférence bayésienne numérique.
- Deux axes :
  - Statique : DAGS (directed acyclic graphs)
  - Dynamique : HMM (hidden Markov models)
- Cela correspond également aux "Highly Structured Stochastic Systems" (HSSS)
- Les structures hiérarchiques sont fréquemment rencontrées en écologie.

# exemple en temps fixe

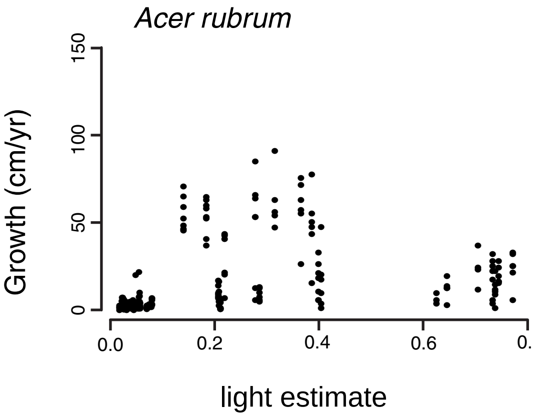
*Clark et al – Coexistence : How to identify trophic trade-offs – Ecology, 2002*

- on souhaite étudier la réponse de la croissance d'une espèce de plante à la lumière
- $m$ parcelles $(j = 1, \ldots, J)$
- sur parcelle $j$ : $n_j$ individus
- on mesure
  - $Y_{ij}$ croissance de l'individu $i$ sur la parcelle $j$ sur 2 ans
  - $X_j$ ensoleillement associé à chaque parcelle $j$

# exemple en temps fixe (suite)



# exemple en temps fixe (suite)



# exemple en temps fixe (suite)

- premier modèle :

$$Y_{ij} = \mu(\theta, X_j) + \varepsilon_{ij}$$

où

$$\mu(r, x_j) = r_1 + r_2 \times \frac{x_j}{r_3 + x_j} \qquad \text{fonction de Monod}$$

bruit i.i.d. $\pi_{\varepsilon_{ij}|\sigma^2}(e_{ij}|s^2) = N(0, s^2)$ $\quad$ ($\sigma^2$ de loi donnée)

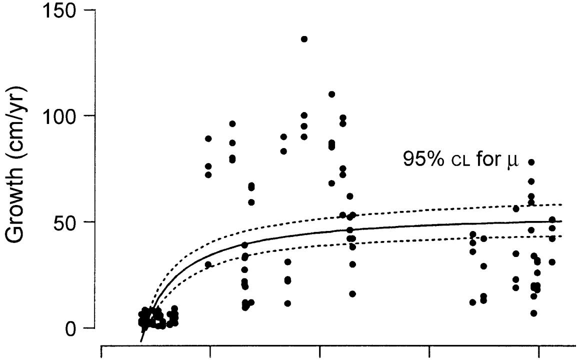
- vraisemblance :

$$\pi_{Y|X,r,s^2}(y|x,\theta,\sigma^2) = \prod_{j=1}^{J}\prod_{i=1}^{n_j} \underbrace{N(y_{ij}|\mu(r,x_j),s^2)}_{\substack{\text{densité gaussienne}\\ \text{moyenne } \mu(r,x_j)\\ \text{variance } s^2}}$$

notation $Y = (Y_{ij})$, $X = (X_j)$, $y = (y_{ij})$, $x = (x_j)$

# exemple en temps fixe (suite)

- on peut utiliser :
  - non bayésien : maximiser la vraisemblance
  - bayésienne : lois a priori sur $(\theta, \sigma^2)$
- on plonge ces estimées dans la fonction de croissance $\mu$



# exemple en temps fixe (suite)

- d'où proviennent les erreurs : mesures ? erreurs de modélisation ?
- affinement sur les mesures : les $Y$ sont OK, en revanche les $X$ sont douteux... supposons que l'on mesure $X_j^{\text{obs}}$

$$
\begin{aligned}
&Y_{ij} = \mu(\theta, X_j) + \varepsilon_{ij} && \\
&\pi_{X_j^{\text{obs}}|X_j} \text{ donnée} && (X_j \text{ de loi donnée}) \\
&\pi_{\varepsilon_{ij}|\sigma^2} \text{ donnée} && (\sigma^2 \text{ de loi donnée})
\end{aligned}
$$

# exemple en temps fixe (suite)

- affinement sur le modèle : la réponse de chaque individu $(ij)$ peut fluctuer

$$
\begin{array}{ll}
Y_{ij} = \mu(\theta_{ij}, X_j) + \varepsilon_{ij} & \\
\pi_{\theta_{ij}|\gamma} \text{ donnée} & (\gamma \text{ de loi donnée}) \\
\pi_{X_j^{\text{obs}}|X_j} \text{ donnée} & (X_j \text{ de loi donnée}) \\
\pi_{\varepsilon_{ij}|\sigma^2} \text{ donnée} & (\sigma^2 \text{ de loi donnée})
\end{array}
$$

# DAGS (directed acyclic graphs)

$Y_{ij} = \mu(\theta_{ij}, X_j) + \varepsilon_{ij}$

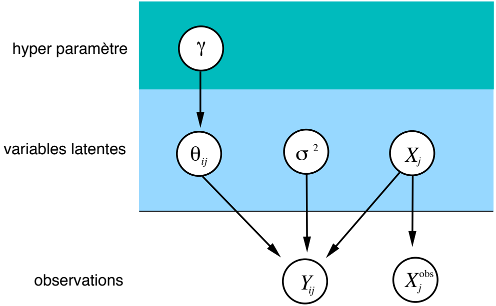
avec la donnée de

$\pi_{\theta_{ij}|\gamma}$

$\pi_{X_j^{\text{obs}}|X_j}$

$\pi_{\varepsilon_{ij}|\sigma^2}$

et de lois (a priori) sur $\gamma$, $X_j$, $\sigma^2$



# exemple en temps fixe (suite)

- loi du modèle

$$\pi_{Y,X^{\text{obs}},\theta,\sigma^2,X,\gamma}$$

$$= \pi_{Y|\theta,\sigma^2,X} \times \pi_{X^{\text{obs}}|X} \times \pi_{\theta|\gamma} \times \pi_{\sigma^2} \times \pi_X \times \pi_\gamma$$

$$= \prod_{j=1}^{J}\prod_{i=1}^{n_j} \pi_{Y_{ij}|\theta,\sigma^2,X} \prod_{j=1}^{J} \pi_{X_j^{\text{obs}}|X} \prod_{j=1}^{J}\prod_{i=1}^{n_j} \pi_{\theta_{ij}|\gamma}\, \pi_{\sigma^2}\, \pi_X\, \pi_\gamma$$

vraisemblance / lois a priori / hyper-loi a priori

# exemple en temps fixe (suite)

- loi a posteriori

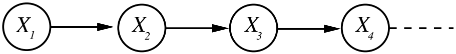
$$\pi_{\theta,\sigma^2,X,\gamma|Y,X^{\text{obs}}} \propto \pi_{Y|\theta,\sigma^2,X}\ \pi_{X^{\text{obs}}|X}\ \pi_{\theta|\gamma}\ \pi_{\sigma^2}\ \pi_X\ \pi_\gamma$$

- mais on ne sait pas intégrer ça....

# chaîne de Markov

- processus stochastique $X = (X_1, X_2, X_3 \dots)$ dont les évolutions futurs ne dépendent du passé que par l'intermédiaire du présent, i.e.

$$\pi_{X_{n+1}|X_n, X_{n-1}, \dots, X_1} = \pi_{X_{n+1}|X_n}$$




- extension des systèmes récurrents au cas stochastique
- utilisé en modélisation (HMM) comme en algorithmique (recuit simulé, MCMC)

# identification en temps mobile

- à l'instant $t$
  - on dispose d'observations

$$Y_{1:t} = (Y_1, Y_2, \ldots, Y_t)$$

  - on cherche à estimer des variables cachées

$$X_{1:t} = (X_1, X_2, \ldots, X_t)$$

  - on cherche éventuellement à estimer des paramètres inconnus $\theta$

# temps mobile

- approche séquentielle nécessaire si :
  - contrainte de temps réel
  - beaucoup d'observations à traiter / fouille de données
  - assimilation de données

- l'approche séquentielle n'est pas nécessaire si :
  - horizon fini
  - beaucoup de temps entre 2 observations

# séquentiel

- temps réel $\longrightarrow$ approche séquentielle, i.e. filtrage

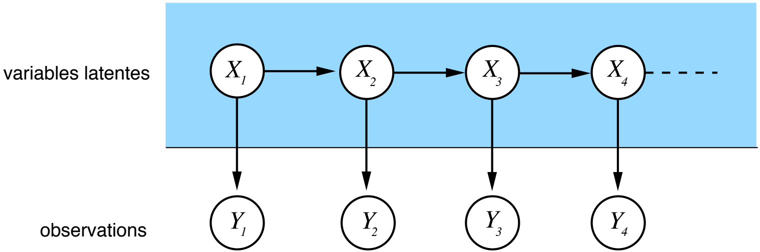
$$\pi_{X_{t-1},\theta|Y_{1:t-1}} \xrightarrow{\;\;\overset{Y_t}{\downarrow}\;\;} \pi_{X_t,\theta|Y_{1:t}}$$

(cela ne donne que les marginales)

- il faut des hypothèses
  - $X_t$ markovien
  - modèle d'observation simple

  $\Rightarrow$ modèle de Markov caché

# exemple en temps mobile (suite)



# modèle à espace d'état

- c'est équivalent à un modèle à espace d'état :

$$X_t = f(\theta, X_{t-1}, W_t)$$
$$Y_t = h(\theta, X_t, V_t)$$

  - $W_t$ et $V_t$ sont des bruits blancs (variables iid et centrées)
  - $W_t$, $V_t$, $X_1$, $\theta$ indépendants

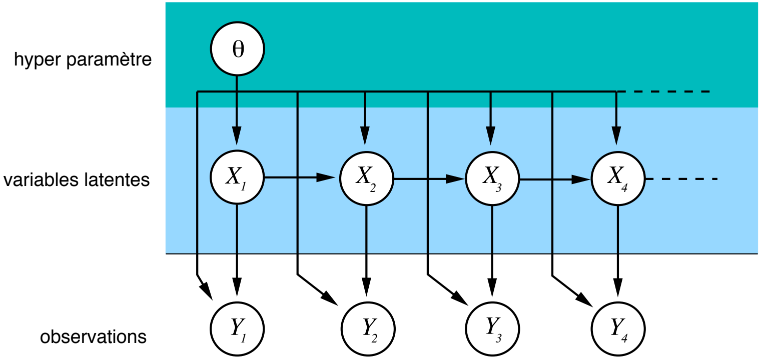
# exemple 1 : pêcherie

- modèle de croissance de Ricker

$$X_{t+1} = X_t \, e^{r - b\, X_t} \, e^{W_t} \qquad \text{évolution de la biomasse}$$
$$Y_t = h \, X_t \, e^{V_t}$$

- $W_t$ et $V_t$ bruits blancs gaussiens indépendants
- $\theta = (r, b, \sigma^2_W, \sigma^2_V)$

# exemple 1 : pêcherie (suite)



# exemple 2 : pêcherie

- modèle de différence avec retard (Deriso–Schnute)

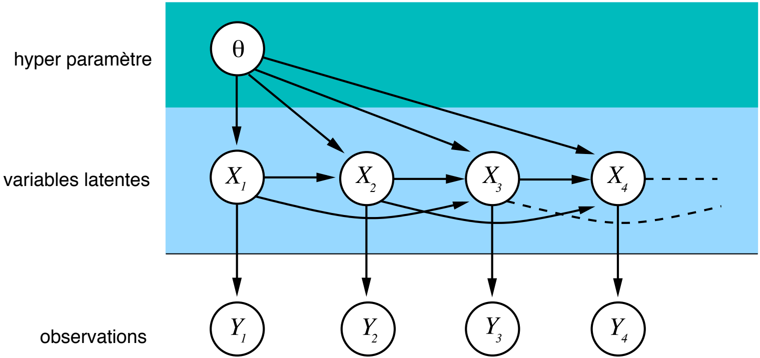
$$X_t = \Big[(1+\rho)\, e^{-M}\, \tfrac{X_{t-1}-C_{t-1}}{X_{t-1}}\, X_{t-1} \\ - \rho\, e^{-2M}\, \tfrac{X_{t-1}-C_{t-1}}{X_{t-1}}\, \tfrac{X_{t-2}-C_{t-2}}{X_{t-2}}\, X_{t-2} \\ + R\, \big(1 - \rho\, e^{-M}\, \omega\, \tfrac{X_{t-1}-C_{t-1}}{X_{t-1}}\big)\Big] \times e^{W_t}$$

évolution de la biomasse

$$Y_t = h\, X_t\, e^{V_t}$$

- $w_t$ et $v_t$ bruits blancs gaussiens indépendants
- $\theta = (K, R, q, \sigma^2_W, \sigma^2_V)$

# exemple 2 : pêcherie (suite)



# exemple 3 : forêt

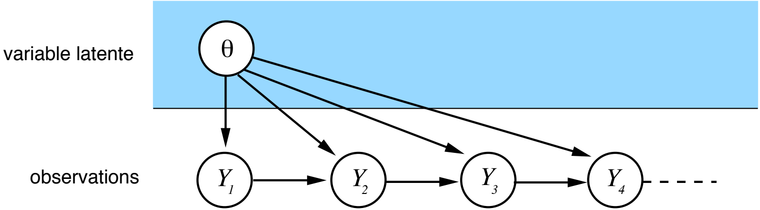
- modèle agrégé de dynamique de forêt :

$$Y_t = \begin{pmatrix} Y_t^1 \\ Y_t^2 \\ Y_t^3 \\ Y_t^4 \end{pmatrix} \qquad \begin{matrix} \#\ \text{individus } 10\text{cm} < \varnothing < 20\text{cm} \\ \#\ \text{individus } 20\text{cm} < \varnothing < 30\text{cm} \\ \#\ \text{individus } 30\text{cm} < \varnothing < 40\text{cm} \\ \#\ \text{individus } 40\text{cm} < \varnothing \end{matrix}$$

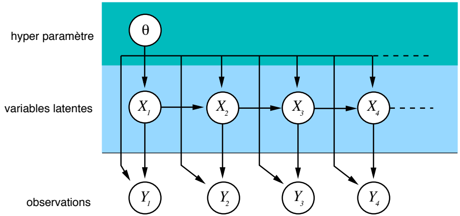
$$\pi_{Y_{t+1}|Y_t,\theta}(y_{t+1}|y_t, r) = q(y_{t+1}|y_t, r) \quad \text{(donnée)}$$

  - but : estimer $\theta$

# exemple 3 : forêt (suite)



# loi du modèle : cas général




$$\pi_{Y_{1:t},X_{1:t},\theta} = \pi_{Y_{1:t}|X_{1:t},\theta}\ \pi_{X_{1:t}|\theta}\ \pi_\theta$$

$$= \prod_{s=1}^{t} \pi_{Y_s|X_s,\theta} \prod_{s=2}^{t} \pi_{X_s|X_{s-1},\theta}\ \pi_{X_1|\theta}\ \pi_\theta$$

vraisemblance / loi du système (a priori) / hyper-loi a priori

# exemple en temps fixe (suite)

- loi a posteriori

$$\pi_{X_{1:t},\theta|Y_{1:t}} \propto \prod_{s=1}^{t} \pi_{Y_s|X_s,\theta} \prod_{s=2}^{t} \pi_{X_s|X_{s-1},\theta} \, \pi_{X_1|\theta} \, \pi_\theta$$

# formule de Bayes séquentielle (filtrage)

(oublions $\theta$)

$$\pi_{X_{t-1}|Y_{1:t-1}} \longrightarrow \pi_{X_t|Y_{1:t}}$$

- prédiction (Chapman-Kolmogorov)

$$\pi_{X_t|Y_{1:t-1}}(x_t|y_{1:t-1}) = \int \pi_{X_t|X_{t-1}}(x_t|x_{t-1}) \; \pi_{X_{t-1}|Y_{1:t-1}}(x_{t-1}|y_{1:t-1}) \, dx_{t-1}$$

- correction (Bayes)

$$\pi_{X_t|Y_{1:t}} \propto \pi_{Y_t|X_t} \, \pi_{X_t|Y_{1:t-1}}$$

- à part dans le cas linéaire/gaussien, on ne peut (presque) rien en faire...

- Monte Carlo : simuler le hasard pour évaluer des quantités déterministes (une intégrale, la solution d'une EDP etc.)
- préhistoire : aiguille de Buffon
- Ulam réalise que cela correspond parfaitement aux calculateurs

Manhattan Project in Los Alamos : Ulam, Richard Feynman, and John Von Neumann

# Monte Carlo

- c'est LA méthode numérique pour la statistique bayésienne : approximation de la loi conditionnelle (a posteriori) $\pi_{\theta|Y}$
- principe : si on dispose d'une échantillon

$$\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(N)} \stackrel{\text{iid}}{\sim} \pi_{\theta|Y}$$

alors

$$\widehat{\phi(\theta)} = \mathbb{E}[\phi(\theta)|Y] \simeq \frac{1}{N}\sum_{i=1}^{N} \phi(\theta^{(i)})$$

i.e.

$$\pi_{\theta|Y} \simeq \pi^N_{\theta|Y} \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{\theta^{(i)}}$$

- mais le plus souvent on ne sait pas échantillonner selon $\pi_{\theta|Y}$

$$\alpha^N = \frac{1}{N}\sum_{i=1}^{N}\phi(\theta^{(i)}) \xrightarrow[N\uparrow\infty]{} \alpha = \mathbb{E}[\phi(\theta)|Y=y]$$

- convergence : loi des grands nombres
- vitesse de convergence : théorème central limite

$$\frac{\alpha-\alpha^N}{\sqrt{N}} \simeq \mathcal{N}(0,\sigma^2) \qquad \text{où } \sigma^2 = \mathrm{var}(\phi(\theta)|Y)$$

# MCMC

*Metropolis, Rosenbluth, Rosenbluth, Teller, Teller – Equation of state calculations by fast computing machines – Journal The Journal of Chemical Physics, 1953*

- but : échantillonner selon une densité $\pi(z)$ (dite cible) connue analytiquement (à une constante multiplicative près)
- principe : construire une chaîne de Markov $(Z^{(k)})_{k\geq 0}$ dont la densité limite est $\pi(z)$ (cv : théorème erodique)

# Metropolis-Hastings

- comment atteindre $\pi$ à partir de (presque) n'importe quelle densité de transition $q^{\text{prop}}(z'|z)$ ?

- i.e. comment perturber $q^{\text{prop}}$ afin d'obtenir un noyau qui admet $\pi$ comme mesure invariante ?

# Metropolis-Hastings (suite)

- transition $Z^{(k)} \to Z^{(k+1)}$ en deux étapes :
  - proposer un candidat :

$$\tilde{Z} \sim q^{\text{prop}}(z'|Z^{(k)})$$

  - acceptation/rejet

$$Z^{(k+1)} \leftarrow \begin{cases} \tilde{Z} & \text{avec proba } \alpha & \text{(acceptation)} \\ Z^{(k)} & \text{avec proba } 1-\alpha & \text{(rejet)} \end{cases}$$

- prendre $\alpha$ t.q. ce noyau de transition laisse $\pi$ invariant, i.e.

$$\alpha = \frac{\pi(\tilde{Z})\; q^{\text{prop}}(\tilde{Z}|Z^{(k)})}{\pi(Z^{(k)})\; q^{\text{prop}}(Z^{(k)}|\tilde{Z})} \wedge 1$$

# Metropolis-Hastings (suite)

- convergence : sous des hypothèses faibles
- maîtrise de la vitesse de convergence : plus difficile, notamment en pratique

# $+/-$

- raisons du succès
  - simple/souple/variantes
  - s'applique dans beaucoup de cas
  - s'articule avec Monte Carlo, SMC, HMM, réseaux bayésiens, apprentissage...
  - interfacé avec `R` via `WinBugs`
- problèmes
  - vitesse de convergence
  - en pratique : "temps de chauffe" + test d'arrêt
  - risque des systèmes boîtes noires
  - la méthode est souvent utilisée sans comparaison
- efforts
  - populations de Monte Carlo
  - chaînes en parallèles

*Gordon, Salmond, Smith – Novel approach to nonlinear/non–Gaussian Bayesian state estimation – IEE Proceedings, 1993*

- approximation particulaire :

$$p(X_t|Y_{1:t}) \simeq p^N(X_t|Y_{1:t}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_t^{(i)}}(X_t)$$

- itération $p^N(X_{t-1}|Y_{1:t-1}) \to p^N(X_t|Y_{1:t})$ en deux étapes :
  - prédiction (mutation)
    $\xi_{t-}^{(1)}, \xi_{t-}^{(2)}, \dots, \xi_{t-}^{(N)} \sim p(X_t|X_{t-1} = \xi_{t-1}^{(i)})$
  - correction (sélection)
    $\omega_{t-}^{(i)} \propto p(Y_t|X_t = \xi_{t-}^{(i)})$ (pondération)
    $\xi_{t-}^{(1)}, \xi_{t-}^{(2)}, \dots, \xi_{t-}^{(N)} \sim \sum_{i=1}^{N} \omega_{t-}^{(i)}\, \delta_{\xi_{t-}^{(i)}}$ (rééchantillonnage)

- on a besoin :
  - singer le système d'état (simuler)
  - calculer les vraisemblances
  - une procédure efficace de rééchantillonnage

- intérêts :
  - simple/souple
  - facile à appréhender
  - variantes
  - Kalman d'ensemble en assimilation de données séquentielle
- succès sur des applications difficiles : suivi de mouvement dans des séquences d'images, robotique (SLAM), localisation de portables, GPS
- ce n'est pas une technique itérative : c'est pourquoi on l'utilise parfois à la place de MCMC
- idées pour les applications écologie/environnement :
  - combiner MCMC et SMC
  - spatio-temporel

- mélanger modélisation et analyse statistique
  - un modèle "crétin" avec des bonnes méthodes d'inférence...
  - prisonnier des données
  - développer (aussi) des modèles simples pour les aspects numériques
- L'aspect décisionnel est embryonnaire, mais les algorithmes sont (presque) à portée : POMPD (partially observed Marov decesion process)
- Il ne faut pas rêver : se méfier de l'effet « résultat ». MCMC et SMC donnent toujours des résultats... parfois ils sont mauvais.

- succès aussi dû à la diffusion par R
- il n'y a pas que le bayésien
- Concernant l'écologie, la question pourrait aller jusqu'à envisager de placer son analyse expérimentale dans un cadre bayésien.

- il y a percolation dans ce domaine entre application, modélisation probabiliste et inférence numérique
  - communauté très active
  - cela permet aussi d'intégrer des maths plus rigoureuses et sophistiquées
  - "Why environmental scientists are becoming ~~Bayesians~~ Monte Carlo'ists"
- Il est (presque) toujours possible de traiter les problèmes rencontrées en environnement/écologie
- (presque) une méthodologie simple et accessible
- qualité des algorithmes : simples, modulables, beaucoup de variantes