

Initiation aux modèles graphiques

Éric PARENT

(¹)Equipe MODélisation, Risque, Statistique, Environnement de l'UMR *MIA 518* INRA/
AgroParisTech (Math. Info. App. 518)

19 Mars 2019
Ecole Chercheur BIOBAYES
Statistique bayésienne

Initiation aux modèles graphiques

- Le langage naturel de conditionnement
- Qu'est ce qu'une inférence bayésienne?
- Qu'est ce qu'une modélisation bayésienne?
- Qu'est ce qu'un modèle hierarchique bayésien?

- Equation de Bayes locale
- Intérêt des « Gaussian Random Markov Fields »

- Perspectives

Comprendre et communiquer

- ...Bayesian statistics is difficult in the sense that thinking is difficult, Berry (1997)
- But drawing some little graphs may help thinking and improving communication...



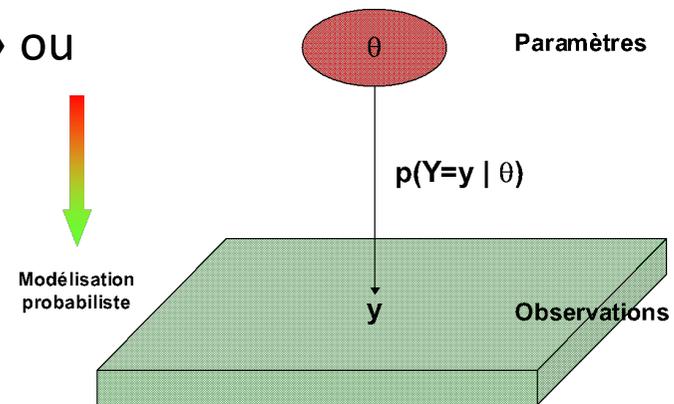
RAPPEL - Probabilités

- Concept fondamental

Mesure angulaire de la quantification de l'incertitude (toute inconnue en « Bayésien » mais seulement les grandeurs observables en « Fréquentiste »)

- La vraisemblance (cadres « Bayésien » ou « Fréquentiste »)

faire une hyp. sur le modèle probabiliste qui a généré les données



- Multivarié : Faire un modèle c'est décrire les covariations entre plusieurs grandeurs aléatoires: [A,B,C,D...]

- Concept mathématique: Loi conditionnelle et son interprétation

$$[A | B] \times [B] = [A, B] = [B | A] \times [A]$$

$$[A | B, C] \times [B | C] \times [C] = [A, B, C] \dots \text{etc.}$$

- Indépendance conditionnelle : A B sachant C

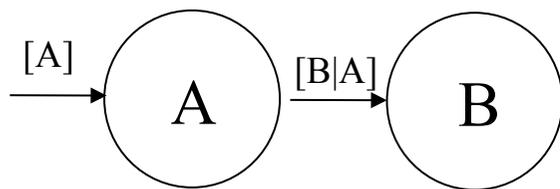
$$[A, B | C] = [A | C] \times [B | C]$$

Modélisation des dépendances probabilistes

Suivre la flèche

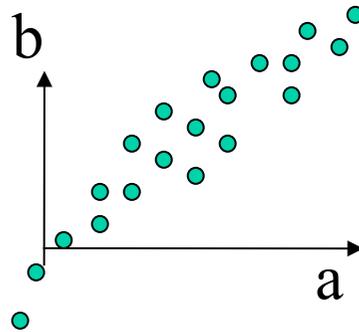
Arnolds, B., Castillo, E., and Sarabia, J. (1999).

Conditional Specification of Statistical Models. Springer Verlag, New York.

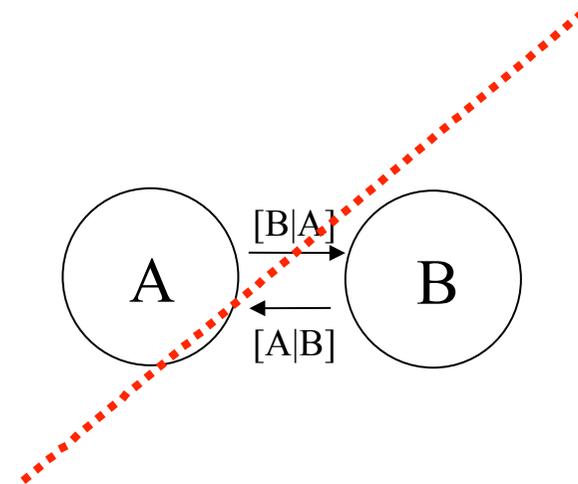


Idée suivie

$$[A,B]=[A][B|A]$$



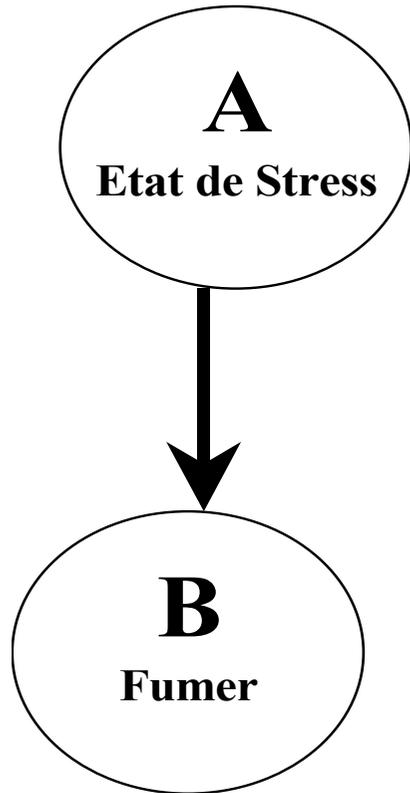
Modèle [A,B]



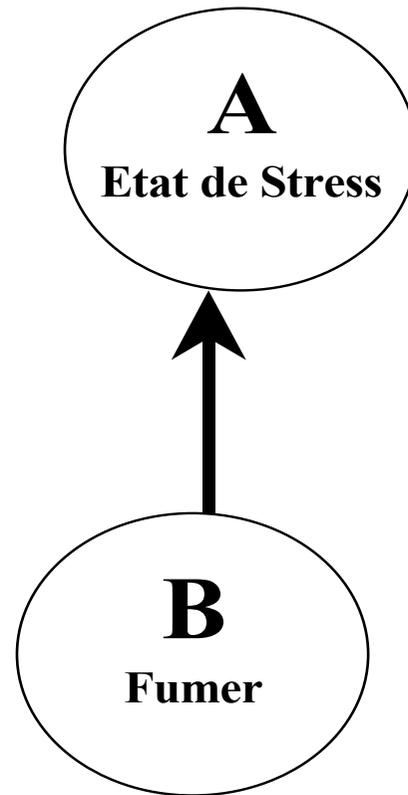
Idée abandonnée

...pour la construction

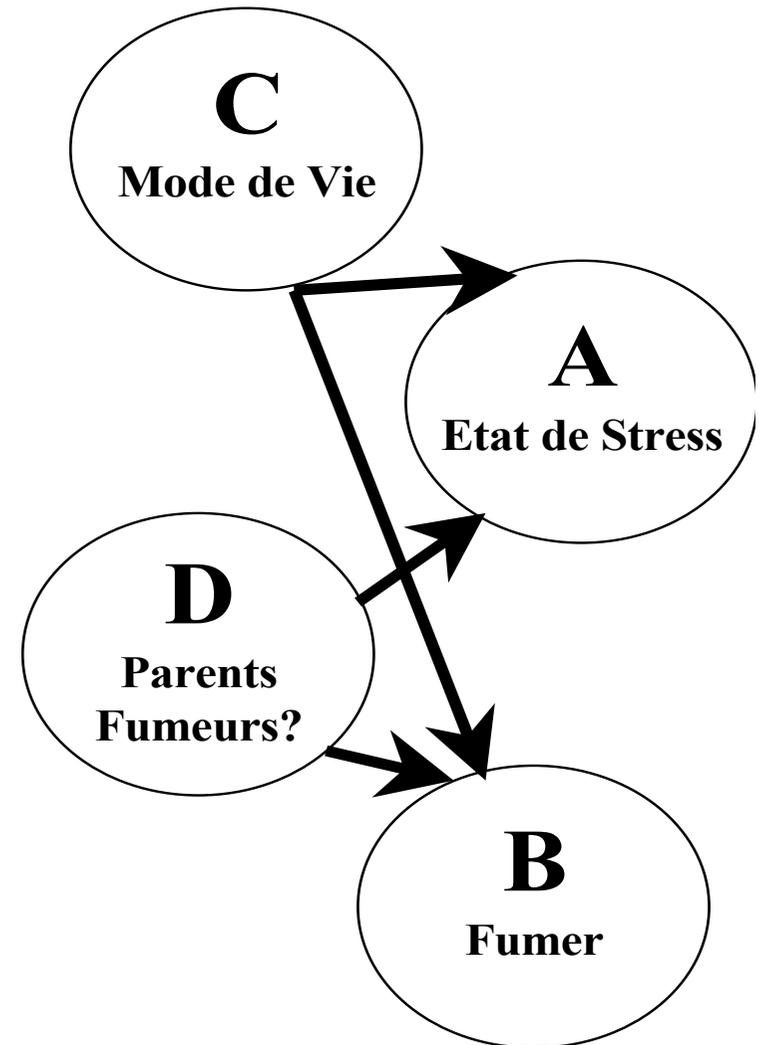
Causalité et dépendance probabiliste



Si je suis stressé,
il se peut
que je fume

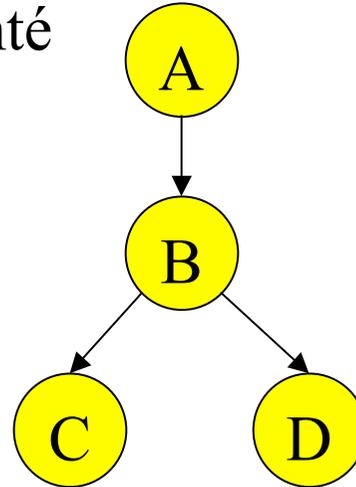


Fumer peut
éventuellement
provoquer
un état de stress



Modèle graphique ou réseau bayésien: un langage naturel

1. Un graphe acyclique orienté



2. Les lois conditionnelles pour chaque noeud

A	P(A)
Faux	0.6
Vrai	0.4

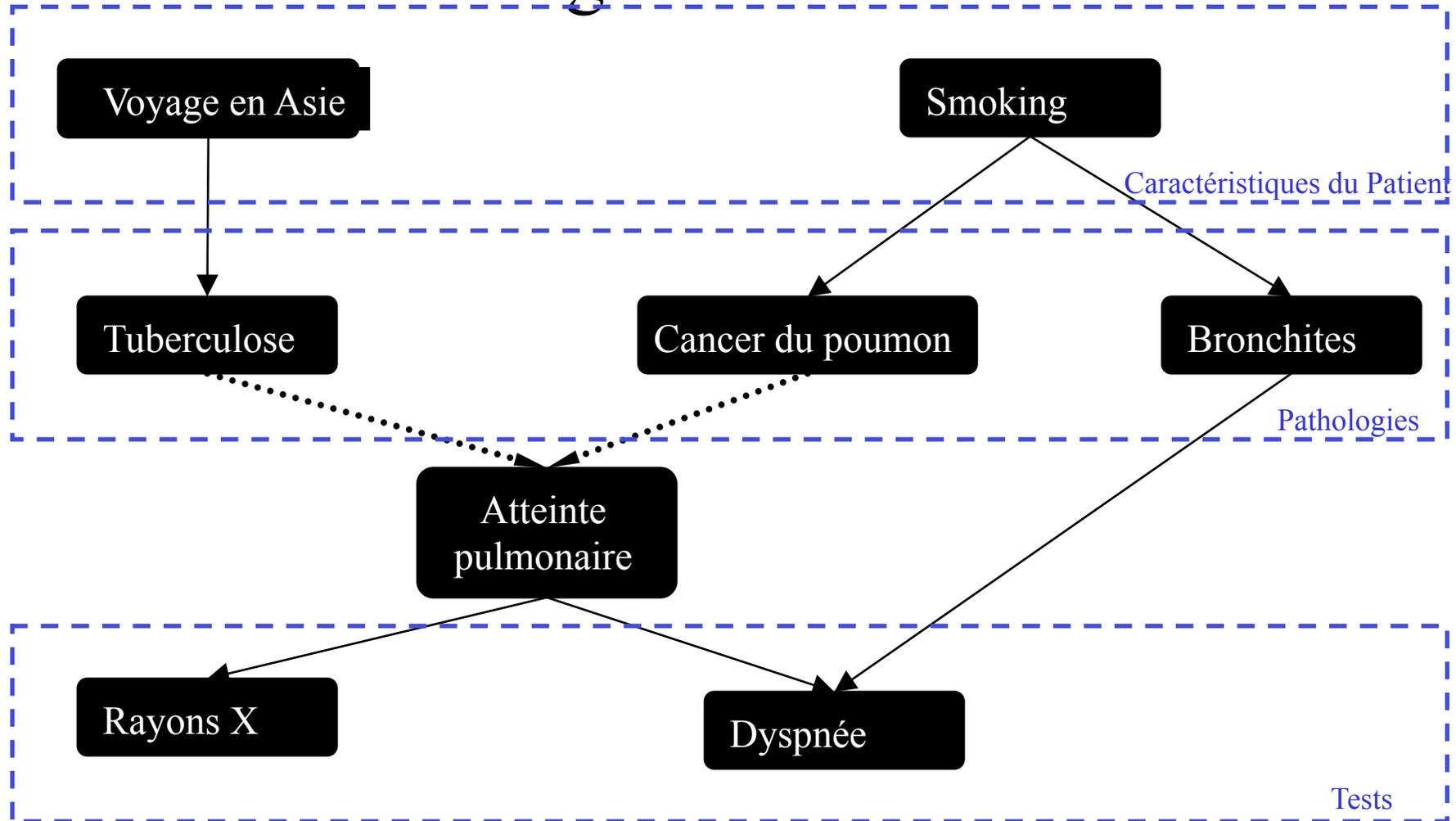
A	B	P(B A)
Faux	Faux	0.01
Faux	Vrai	0.99
Vrai	Faux	0.7
Vrai	Vrai	0.3

B	D	P(D B)
Faux	Faux	0.02
Faux	Vrai	0.98
Vrai	Faux	0.05
Vrai	Vrai	0.95

B	C	P(C B)
Faux	Faux	0.3
Faux	Vrai	0.7
Vrai	Faux	0.9
Vrai	Vrai	0.1

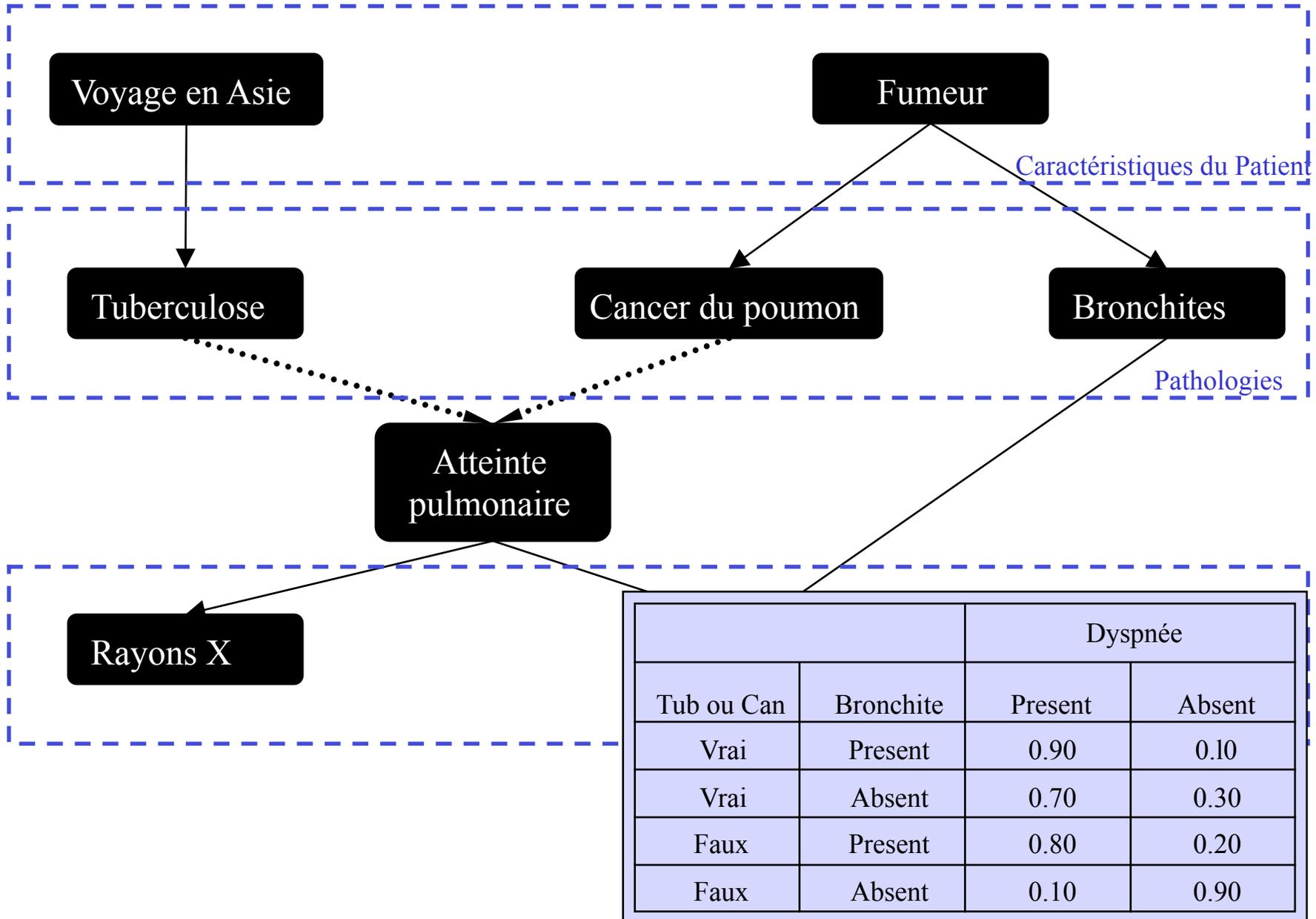
Exemple de réseau bayésien

Diagnostic médical

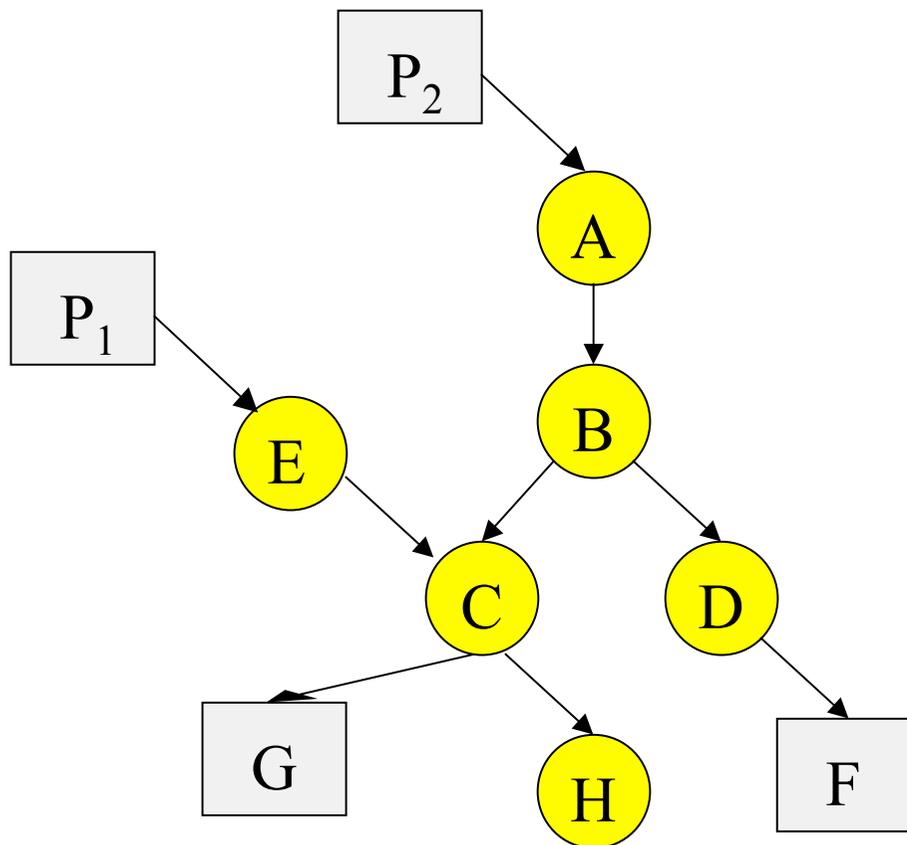


Exemple de réseau bayésien

Diagnostic médical



Qu'est-ce que l'inférence bayésienne?

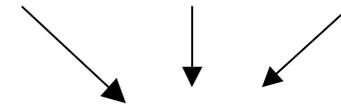
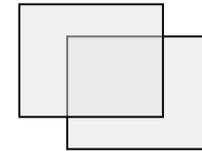
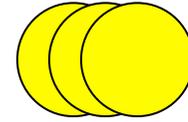
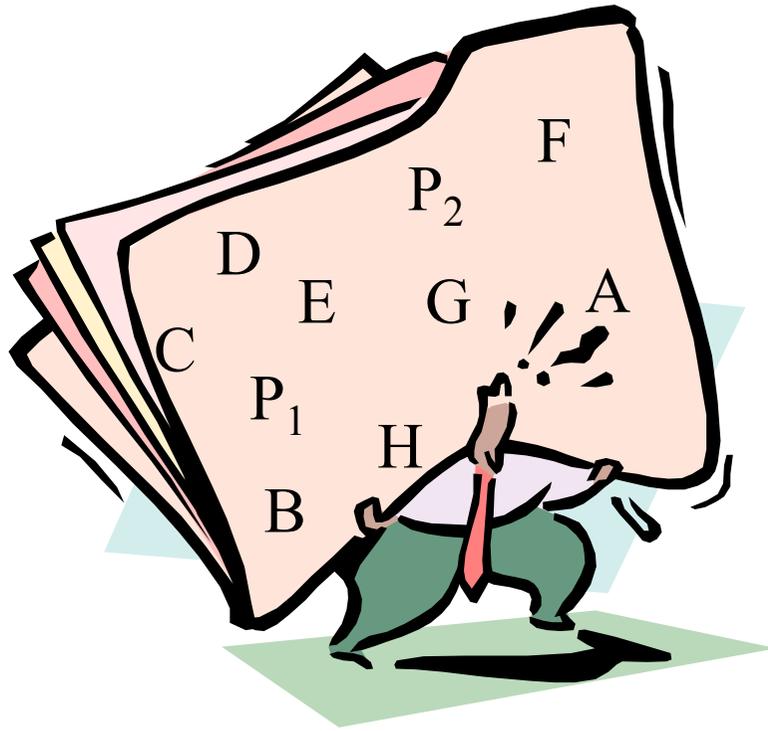


Chercher la loi CONDITIONNELLE (jointe) de certains nœuds sachant d'autres nœuds

Inférence = apprentissage statistique via les probabilités conditionnelles

Une connaissance (prior ou observation) « bloque » un nœud

Qu'est-ce que la modélisation bayésienne?



Nœuds: nommer les grandeurs concernées (art de la découpe)

Mettre en place les relations entre les nœuds

Identifier les inconnues

Identifier les quantités connues

Simplicité, subjectivité et parcimonie...



Exemple de modélisation bayésienne

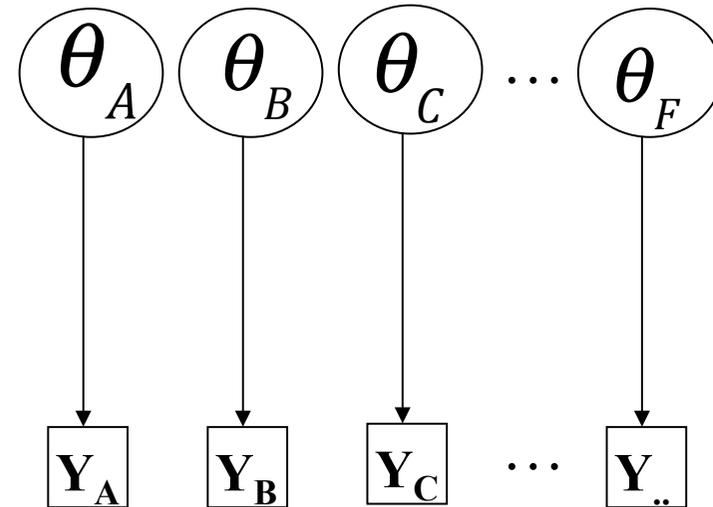
Décès dans les hopitaux

(Thanks to WinBugs)

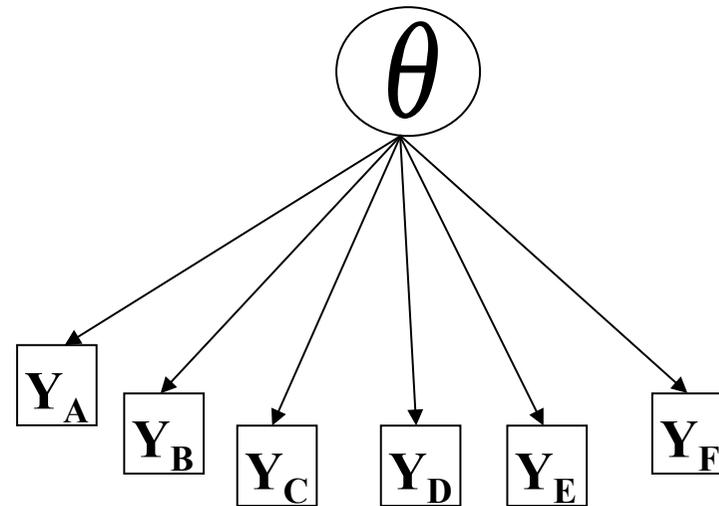
	Hopital	Opérations	Décès
• Quelle est la probabilité de mourir durant une intervention chirurgicale dans un hôpital en UK?	A	47	0
	B	148	18
	C	119	8
• Le nombre y_i de décès suit une loi binomiale de paramètre q_i et N_i essais (i indice de l'hôpital)	D	810	46
	E	211	8
	F	196	13
• Le prior sur q_i est pris dans une loi beta	G	148	9
	H	215	31
	I	207	14
	J	97	8
	K	256	29
L	360	24	

Décès dans les hopitaux UK

- Répéter l'analyse?

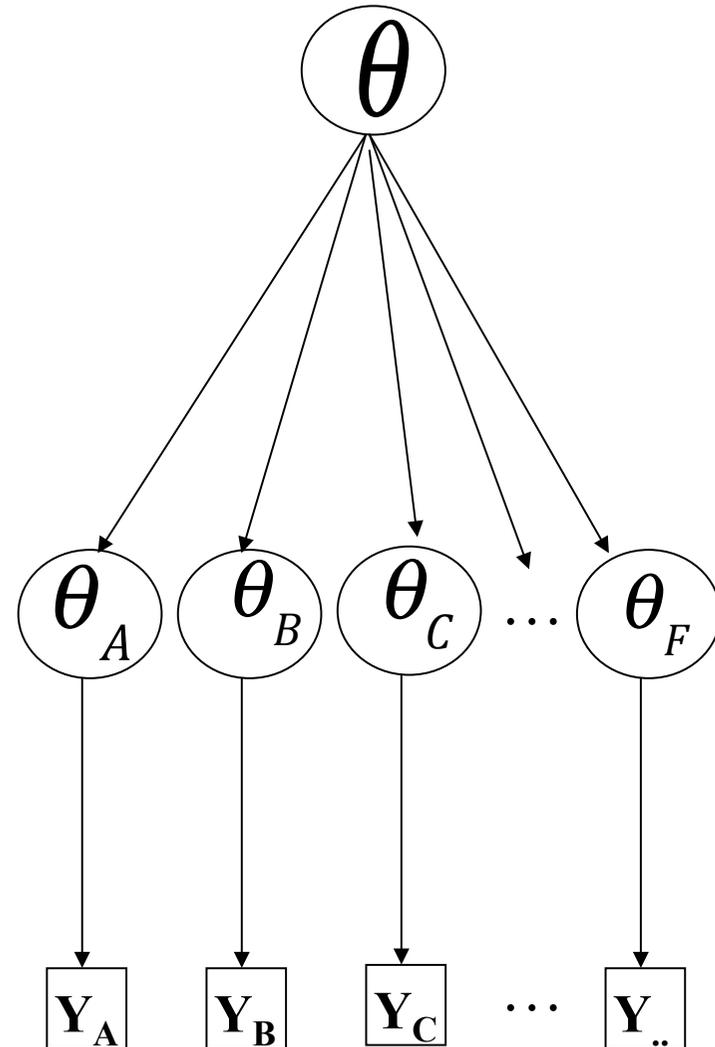


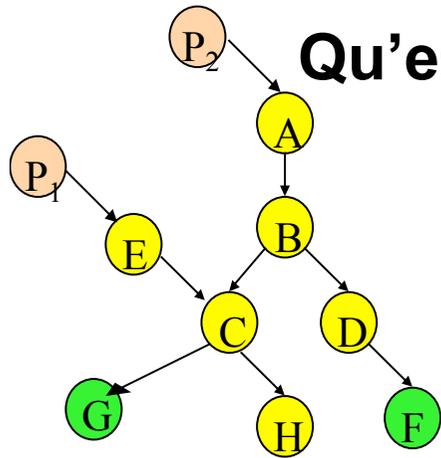
- Mélanger les données?



Décès dans les hopitaux UK Modèle structuré

- Exprimer la + ou –grande ressemblance entre établissements
- Autorise le transfert d'information entre établissements
- Permet de tempérer les estimations optimistes ou pessimistes quand n petit

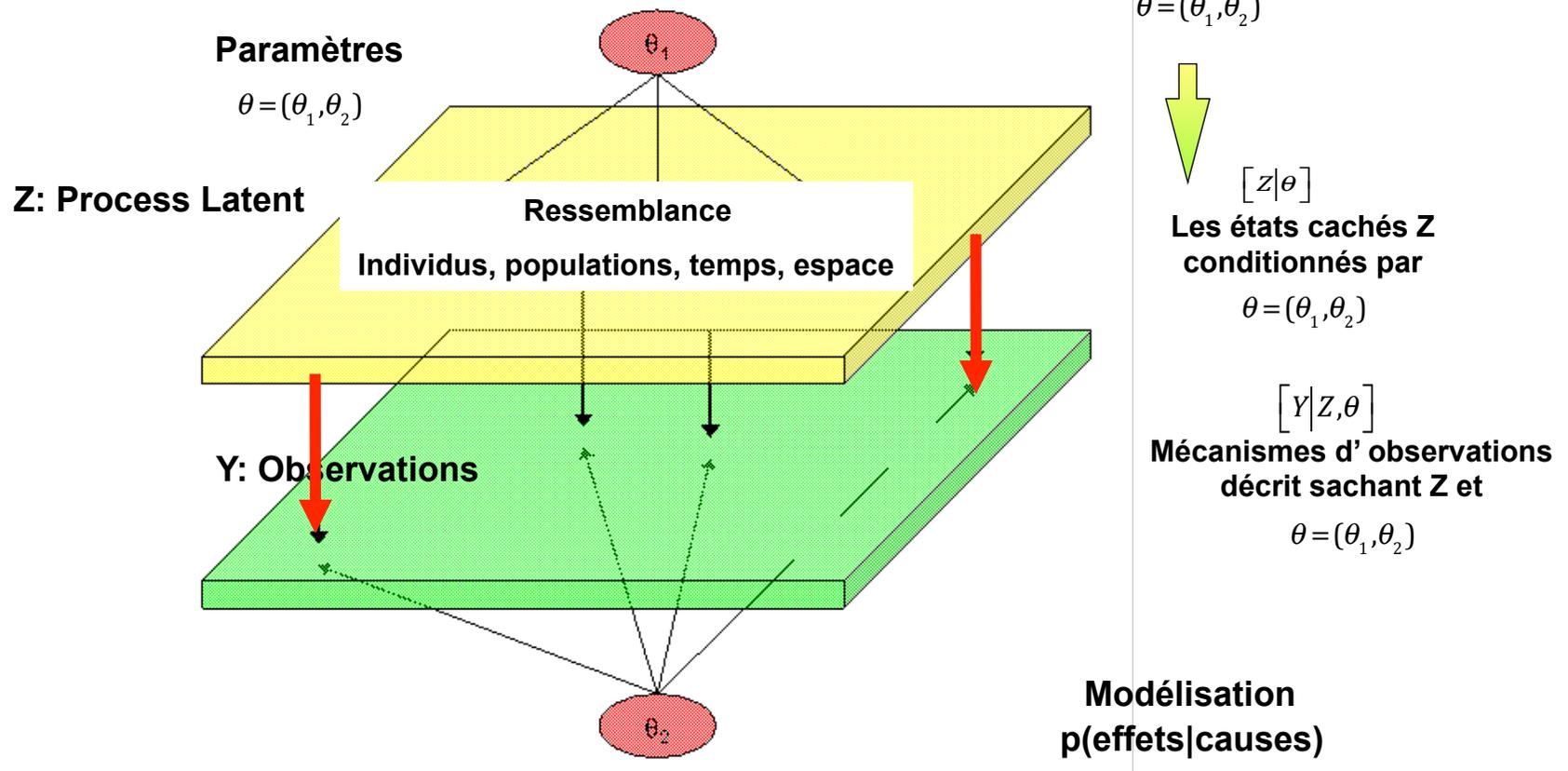




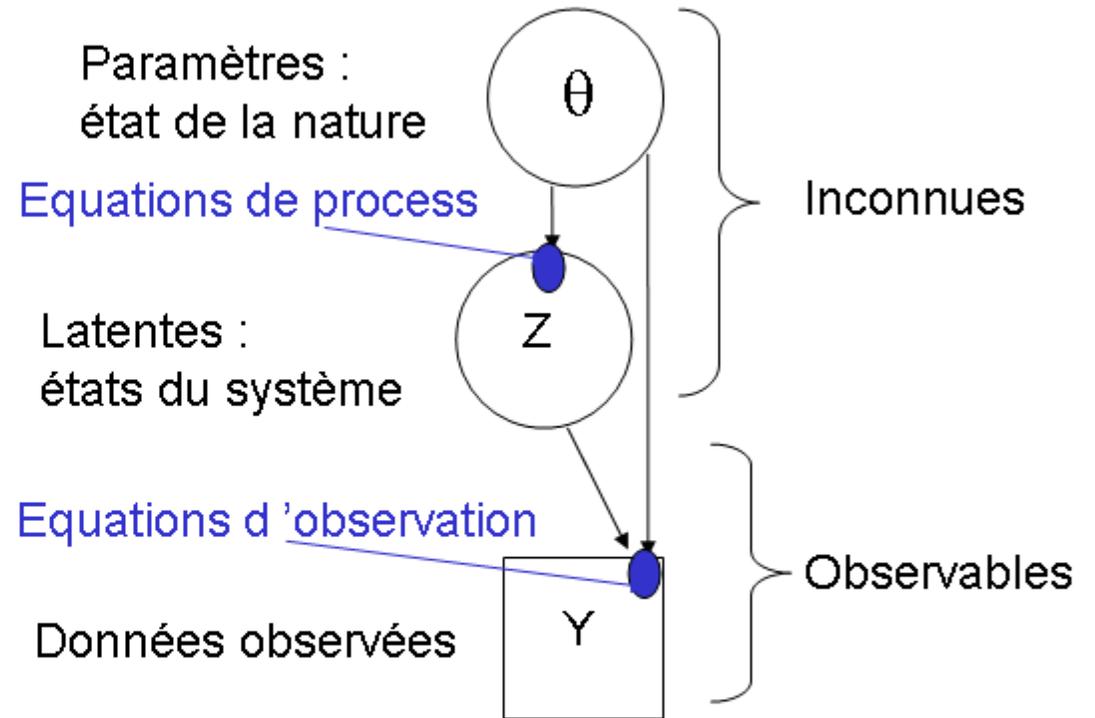
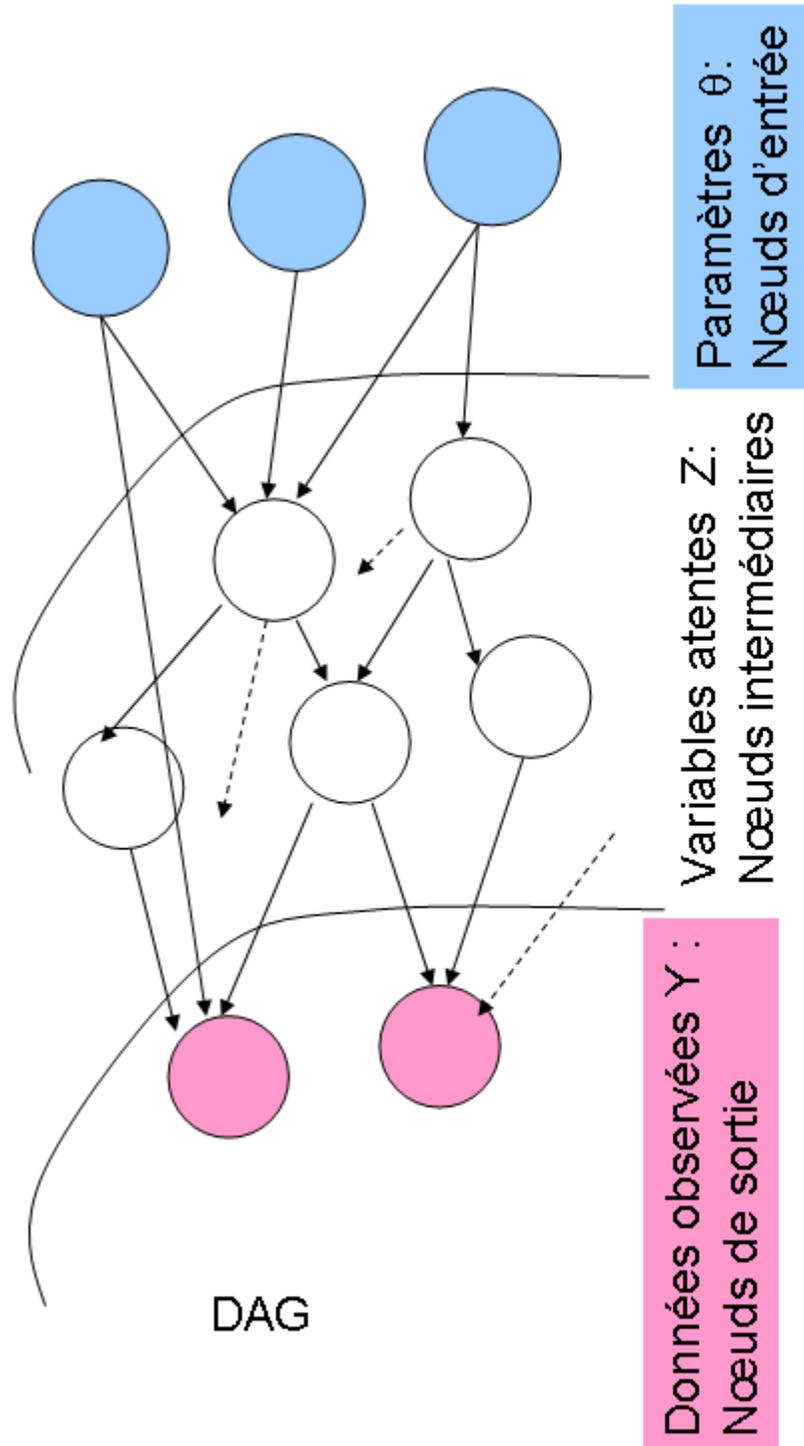
Qu'est ce que la modélisation hiérarchique?

- **Modéliser : Conditionner pour traduire la complexité**

Paramètres, Etats du processus et mécanismes d'Observation sont modélisés indépendemment



Modélisation hiérarchique



Restructuration hiérarchique du DAG

Dynamique spatio-temporelle de 5 espèces de papillons du Parc National des Ecrins

Travail de stage M1 de Clara NAVARRO

Modélisation spatio-temporelle de la dynamique d'émergence de lépidoptères univoltins à partir de données opportunistes (observations notées par les agents du Parc des Ecrins)

Objectif : concevoir une estimation conjointe de l'effort d'échantillonnage et de la dynamique spatio-temporelle de l'émergence de papillons



Data

- occurrences géolocalisées
- sélection des années 2009 à 2020 : 24 814 occurrences
- 5 espèces de papillons univoltins diurnes patrimoniaux



Aurore

Anthocharis cardamines



Citron

Gonepteryx rhamni



Gazé

Aporia crataegi



Demi-deuil

Melanargia galathea



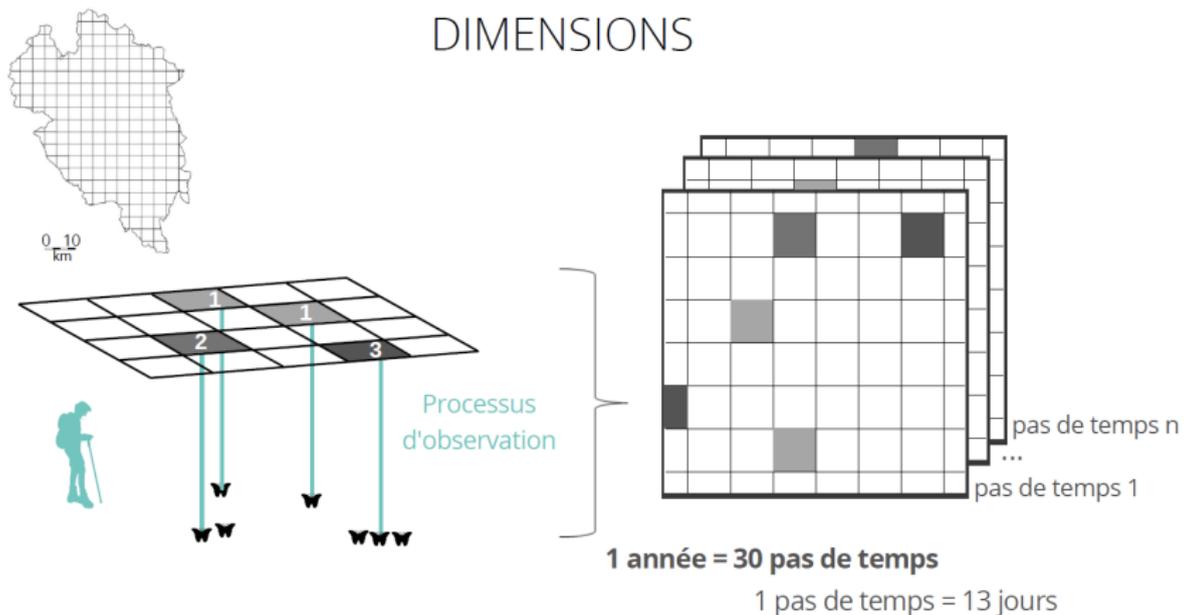
Apollon

Parnassius apollo

Parc national des Écrins

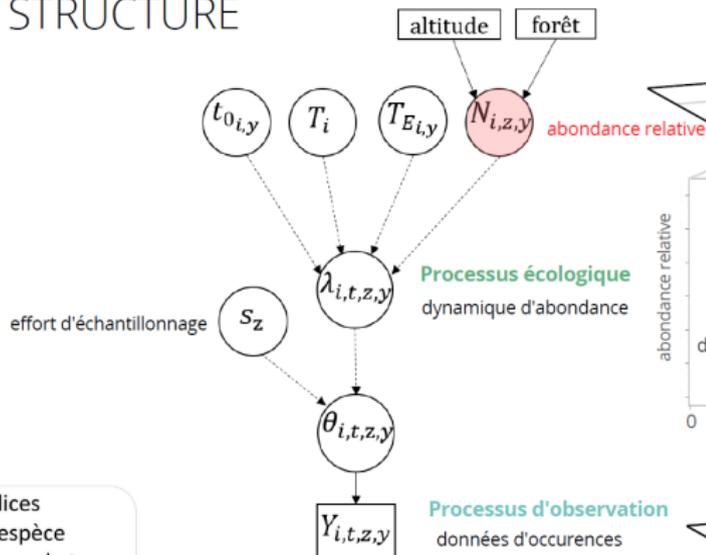


Processus d'observation



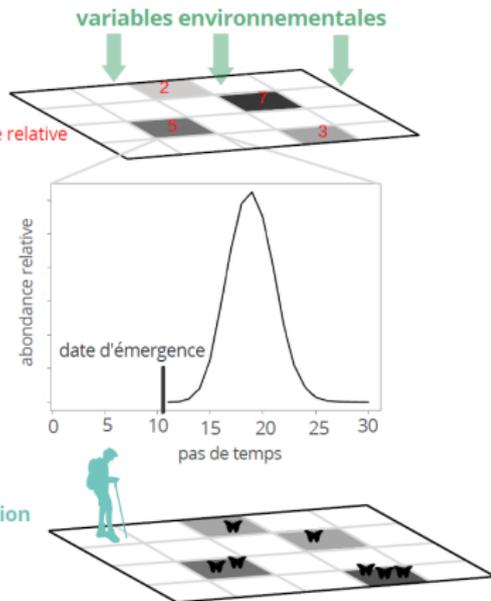
Modèle - représentation graphique (DAG)

STRUCTURE



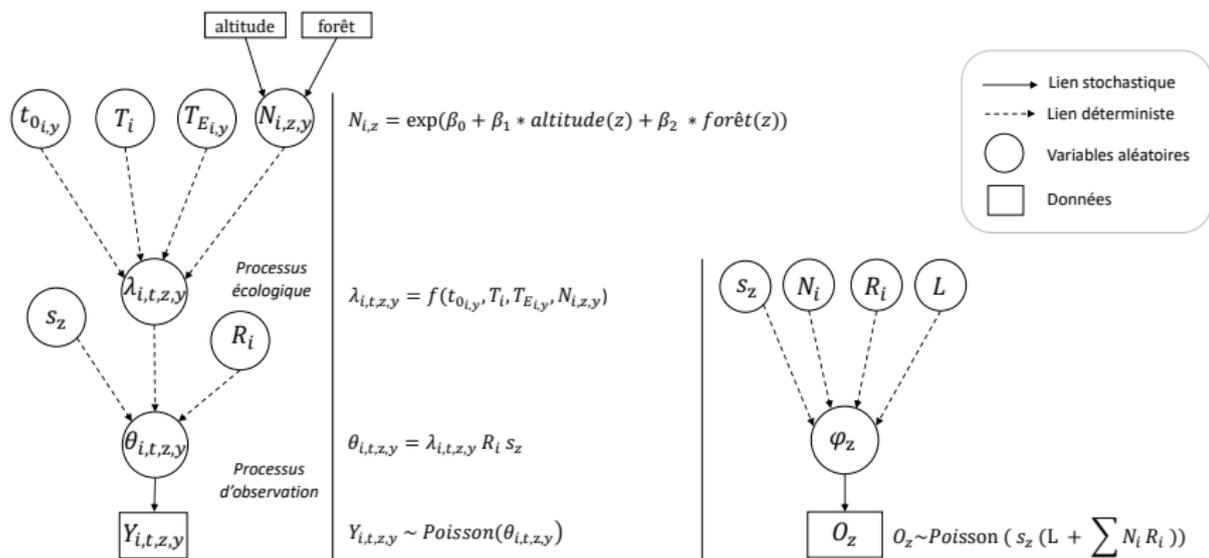
Indices

i : espèce
 t : pas de temps
 z : cellules
 y : années

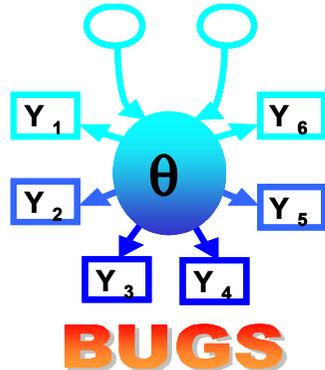


Model

Model overview



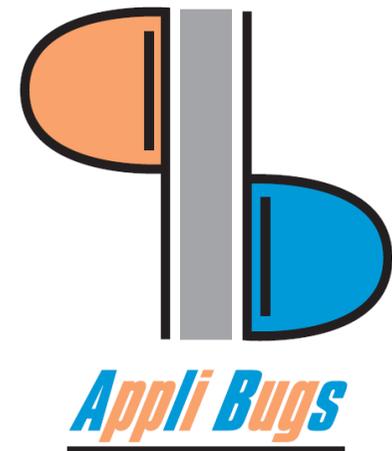
Friendly softwares, e.g. WinBUGS²



“Bayesian inference Using Gibbs Sampler”

Medical Research Council, Biostatistics Unit, Cambridge, UK
Imperial College, London, UK

- Bayesian statistical modeling
- Very popular
- Sampling based methods using Gibbs sampling (eventually hybrid)
- **Models can be described graphically**
- Freeware, available at : www.mrc-bsu.cam.ac.uk/bugs/
- Extensions :
 - Spatial models : GeoBUGS²
 - Convergence : Coda², Boa² ( Packages)
- Coupling with 



<http://w3.jouy.inra.fr/unites/miaj/public/matrisq/Contacts/applibugs.welcome.html>

Depuis Tom B. jusqu'aux nouvelles vagues MC(MC)



Reverend
Thomas Bayes
(1702-1761)

Bayes formula

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B)}$$

1762

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

- Conjugate prior
- Non informative prior
- Bayesian decision theory

1950's

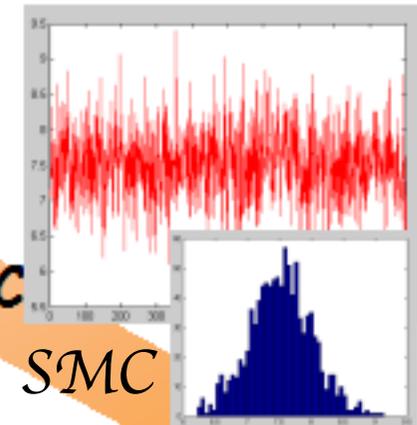
$$[\theta_1|y] = \int \int \int [\theta_1, \theta_2, \dots, \theta_n|y] d\theta_2 \dots d\theta_n$$

1990's

MCMC

- Gibbs sampling
- Hasting-Metropolis

SMC



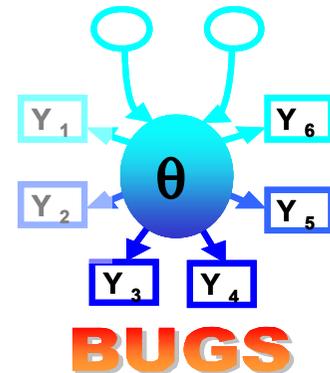
HMC

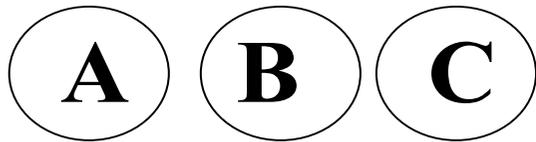
- Interactive MCMC's
- Approximate Variational Bayes
- ABC

Quelques propriétés mathématiques...

- Random Markov Fields : Voisinage et conditionnement
- Recours à la loi Normale

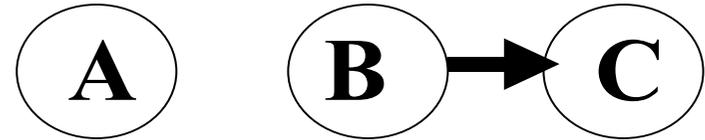
... très utiles pour l'inférence par échantillonneur de Gibbs





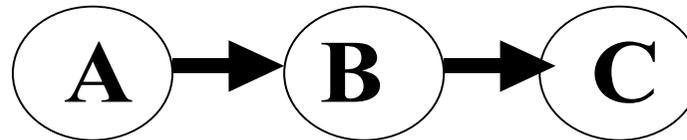
Indépendance

$$[A, B, C] = [C] * [B] * [A]$$



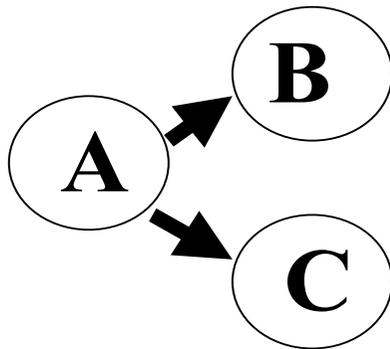
Un contre Deux

$$[A, B, C] = [C | B] * [B] * [A]$$



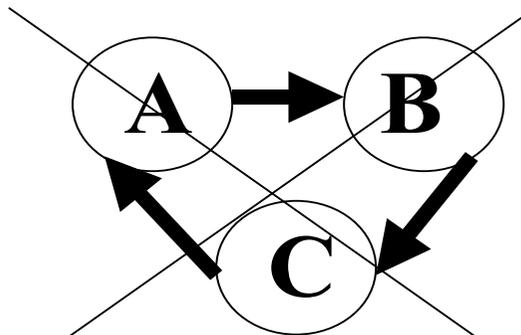
Chaîne

$$[A, B, C] = [C | B] * [B | A] * [A]$$



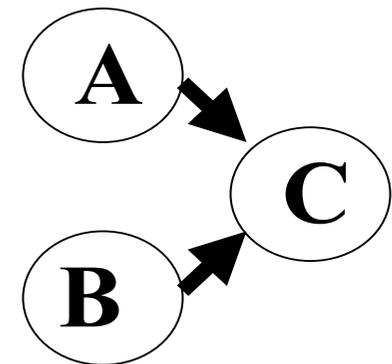
Héritage

$$[A, B, C] = [C | A] * [B | A] * [A]$$



~~Cycle~~

~~$[A, B, C] = ???$~~



Naissance

$$[A, B, C] = [C | A, B] * [B] * [A]$$

Les relations de dépendances conditionnelles possibles entre trois variables

Conditionnelles complètes et moralisation d'un DAG

- Sur un DAG la loi conjointe de tous les nœuds s'exprime facilement

$$[V_1, V_2, \dots, V_i, \dots, V_n] = \prod_{i=1}^n [V_i | Pa(V_i)]$$

avec la convention $[X | Pa(X)] = [X]$ quand le nœud X n'a pas de parents

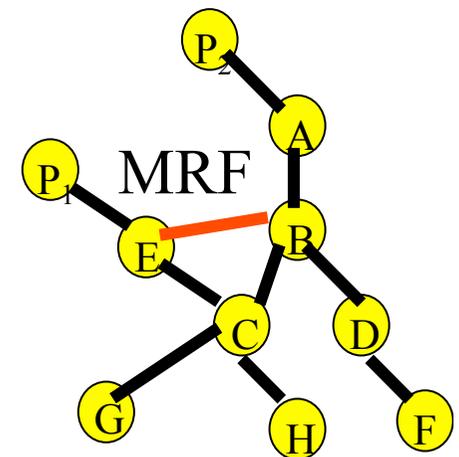
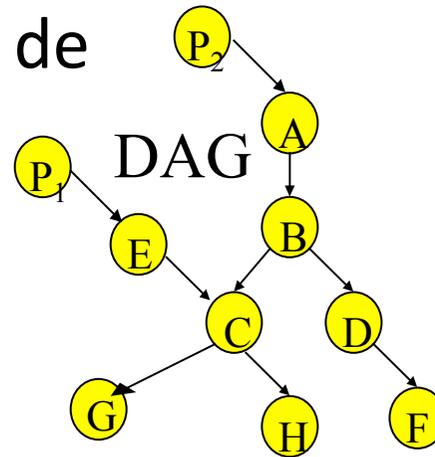
- La conditionnelle complète de X fait intervenir les parents de X , les enfants de X et les co-parents des enfants de X

$$[V_i | V_{-i}] \propto [V_1, V_2, \dots, V_i, \dots, V_n]$$

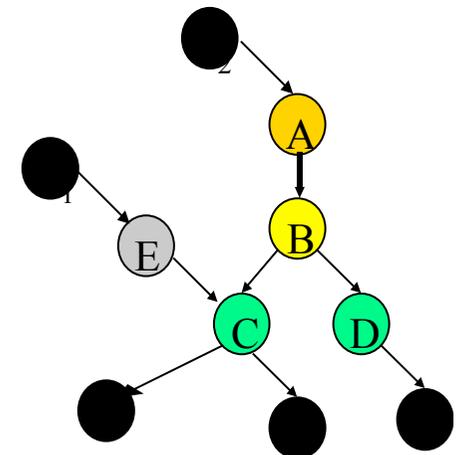
$$[V_i | V_{-i}] = K \times [V_i | Pa(V_i)] \times \prod_{j, Pa(V_j) \ni V_i} [V_j | V_i, V_i', \dots]$$

Couverture markovienne et moralisation d'un DAG

- La conditionnelle complète de X regroupe les membres de sa « famille » : le voisinage de conditionnement



- La formule de Bayes s'exprime localement sur les arêtes d'un DAG: $[B | A, C, D, E, F, G, H, P_1, P_2]$
 prior: $[B | A]$
 vraisemblance: $[D | B] \times [C | B, C]$



Rappel: Vecteur normal multivarié

- La densité de probabilité d'un vecteur normal de moyenne μ et de matrice de variance covariance Σ (définie positive) est:

$$[X|\mu,\Sigma]=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\frac{1}{\sqrt{|\Sigma|}}\exp\left\{-\frac{1}{2}(X-\mu)\Sigma^{-1}(X-\mu)\right\}$$

- Si on appelle $Q=\Sigma^{-1}$ la matrice de précision

$$[X|\mu,Q]=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\sqrt{|Q|}\exp\left\{-\frac{1}{2}(X-\mu)'Q(X-\mu)\right\}$$

GRMF

- Théorème : Soit X vecteur normal de précision Q alors

$$X_i \perp X_j \mid X_{-ij} \Leftrightarrow Q_{ij} = 0$$

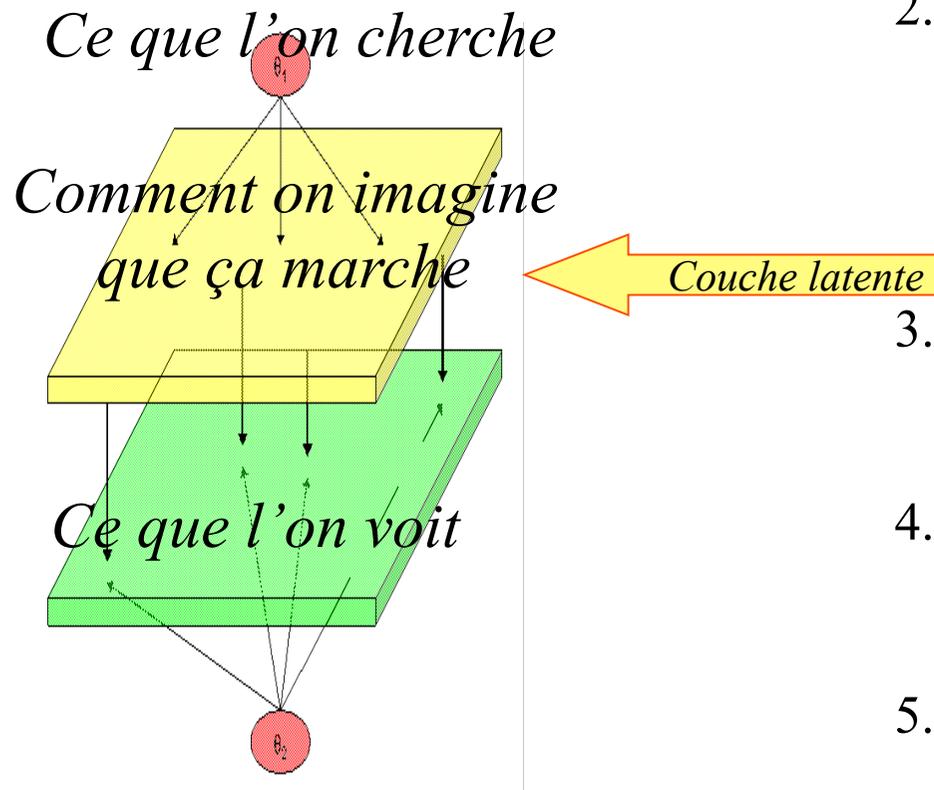
- Les valeurs non nulles de la matrice de précision définissent le graphe de voisinage de conditionnement
- Les lois conditionnelles s'expriment facilement:

$$E(X_i \mid X_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j \leftrightarrow i} Q_{ij} (X_j - \mu_j)$$

$$V(X_i \mid X_{-i}) = Q_{ii}^{-1}$$

$$\text{Corr}(X_i, X_j \mid X_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}$$

Intérêts du raisonnement bayésien



1. Modélisation =Raisonnement Conditionnel. Assemblage de briques conceptuelles en LEGO
2. Construire un modèle ? facile!..Identifier : paramètres, observables, V. Latentes. => Refuser les modèles had hoc.
3. Des couches latentes pour l'encodage du phénomène
4. Détective statistique =Renversement du raisonnement conditionnel
5. Une communauté dotée de nombreux outils
6.Lien avec vision prédictive et théorie de la décision statistique



Collection
Statistique
et probabilités
appliquées

Dirigée par
Yadolah Dodge

COMITÉ EDITORIAL :
Christian Genest
Université Laval, Québec

Marc Hallin
Université libre de Bruxelles,
Belgique

Ludovic Lebart
ENST, Paris

Stephan Morgenthaler
EPFL, Lausanne

Gilbert Saporta
CNAM, Paris

Cette collection met à la disposition du public intéressé par la statistique (étudiants, enseignants, chercheurs) des ouvrages qui concilient effort pédagogique et travail permanent de mise à jour. Cette démarche implique de prendre en compte de façon sélective et critique les renouvellements des concepts, des champs d'application et des outils de traitement. Seules une compréhension profonde et une appropriation des connaissances permettront de s'adapter aux évolutions qui n'ont pas fini de bouleverser cette discipline.



springer.com

Éric Parent, Jacques Bernier

Raisonnement bayésien

Modélisation et inférence

Cet ouvrage expose de façon détaillée la pratique de l'approche statistique bayésienne à l'aide de nombreux exemples choisis pour leur intérêt pédagogique. Il insiste particulièrement sur l'emploi du raisonnement conditionnel qui fonde la cohérence profonde des méthodes de la modélisation et de l'inférence statistique sous le paradigme bayésien.

La première partie donne les principes généraux de modélisation statistique permettant d'encadrer mais aussi de venir en secours de l'imagination de l'apprenti modélisateur. En examinant des exemples de difficulté croissante, le lecteur forge les clés pour construire son propre modèle. La seconde partie présente les algorithmes de calcul les plus utiles pour estimer les inconnues du modèle. Chaque méthode d'inférence est présentée et illustrée par de nombreux cas d'applications. Si nécessaire, un rappel de leur cadre théorique essentiel est présenté sans démonstration.

Le livre cherche ainsi à dégager les éléments clés de la statistique bayésienne, en faisant l'hypothèse que le lecteur possède les bases de la théorie des probabilités et s'est déjà trouvé confronté à des problèmes ordinaires d'analyse statistique classique. Il peut servir de support à un cours de modélisation ou de statistique appliquée dans un programme de Master ou d'École d'Ingénieur. Il s'adresse également aux chercheurs et utilisateurs désireux de s'assurer de la pertinence des méthodes qu'ils emploient. Le débutant, au prix d'un investissement intellectuel acceptable, ainsi bien que le spécialiste, pourront y trouver les informations fondamentales pour comprendre et mettre en œuvre des modèles répondant à leurs besoins spécifiques.

Raisonnement bayésien
Modélisation et inférence

É. Parent, J. Bernier



Collection
Statistique
et probabilités
appliquées

Éric Parent, Jacques Bernier

Raisonnement bayésien

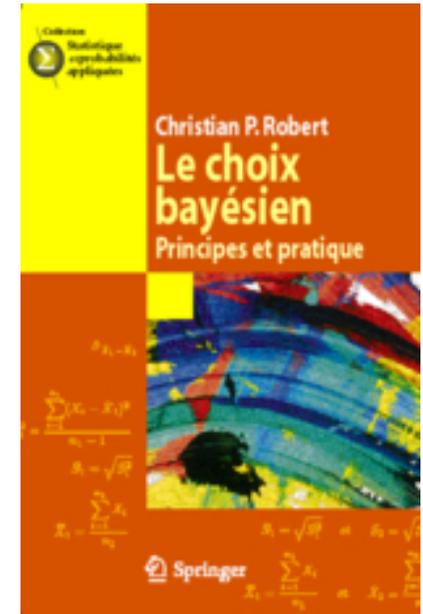
Modélisation et inférence



$$VPP = \frac{P \times Ss}{(P \times Ss) + [(1-P) \times (1-Sp)]}$$

$$VPP = \frac{P \times Ss}{(P \times Ss) + [(1-P) \times (1-Sp)]}$$

Springer



Christian P. Robert
Le choix bayésien
Principes et pratique

Springer



Collection
Statistique
et probabilités
appliquées

Dirigée par
Yadolah Dodge

COMITÉ EDITORIAL :
Christian Genest
Université Laval, Québec

Marc Hallin
Université libre de Bruxelles,
Belgique

Ludovic Lebart
Télécom ParisTech, Paris

Stephan Morgenthaler
EPFL, Lausanne

Gilbert Saporta
CNAM, Paris

Cette collection met à la disposition du public intéressé par la statistique (étudiants, enseignants, chercheurs) des ouvrages qui concilient effort pédagogique et travail permanent de mise à jour. Cette démarche implique de prendre en compte de façon sélective et critique les renouvellements des concepts, des champs d'application et des outils de traitement. Seules une compréhension profonde et une appropriation des connaissances permettront de s'adapter aux évolutions qui n'ont pas fini de bouleverser cette discipline.



springer.com

Jean-Jacques Boreux,
Éric Parent, Jacques Bernier

Pratique du calcul bayésien

Pratique du calcul bayésien est né de l'expérience acquise lors des cours donnés en sciences de l'environnement, tant à l'université de Liège (Jefon), qu'à la grande école AgroparTech (Paris). Son fil conducteur peut se résumer par la locution « de la plume à la souris », tournure empruntée à un opuscle retravaillé la vie d'une école fréquentée jadis par le premier auteur. La première partie privilégie les modèles statistiques paramétriques calculables « à la plume » et est particulièrement riche, tant du point de vue de la présentation des concepts fondateurs du paradigme bayésien, que de leurs applications opérationnelles, notamment en matière d'aide à la décision. Dès le premier chapitre, la représentation du modèle par un graphique spécifique orienté permet de distinguer clairement la phase où la créativité du chercheur s'exprime de celle où il calcule. À cette fin, le logiciel libre WINBUG[®] sera très utile à l'apprenti modélisateur. La seconde partie présente des applications réelles, plus sophistiquées, qui nécessitent souvent d'introduire une couche de variables latentes entre les observables et les paramètres. Conclure une inférence bayésienne sur ces modèles hiérarchiques implique un recours intensif aux méthodes modernes de calcul et mobilise donc « la souris » de l'ordinateur.

Cet ouvrage est dédié aux étudiants et chercheurs qui souhaitent apprendre le calcul bayésien avec des visées opérationnelles. Le lecteur est invité à l'utiliser comme un tremplin lui permettant d'aller aussi loin que son intérêt et/ou ses besoins l'exigent. C'est pourquoi, les treize chapitres offrent un compromis entre la rigueur du langage mathématique et la souplesse de la langue de Mallin. Le côté opérationnel est mis en avant. De nombreux exemples, le plus souvent réels, justifient les efforts et illustrent les raisonnements sous-jacents. Les développements théoriques sont donc volontairement limités à l'essentiel et le lecteur désireux de les poursuivre trouvera d'excellents ouvrages de référence publiés dans la même collection.

J.-J. Boreux, É. Parent, J. Bernier

Pratique du calcul bayésien



Collection
Statistique
et probabilités
appliquées

Jean-Jacques Boreux,
Éric Parent, Jacques Bernier

Pratique du calcul bayésien



Springer



Méthodes
BAYÉSIENNES
EN STATISTIQUE

Jean-Jacques Boreux, Éric Parent, Jacques Bernier



Springer

L'analyse statistique Bayésienne

Le cadre statistique Bayésien :

Un cadre de référence pour la modélisation stochastique et statistique, l'inférence et la prédiction, ...

... dans de nombreux domaines (sciences sociales, économie, finance, bio-médical, environnement, écologie, ...)



Efron B.1986. Why isn't every one a Bayesian ? (with comments and reply).
American Statistician, 40:1-11.

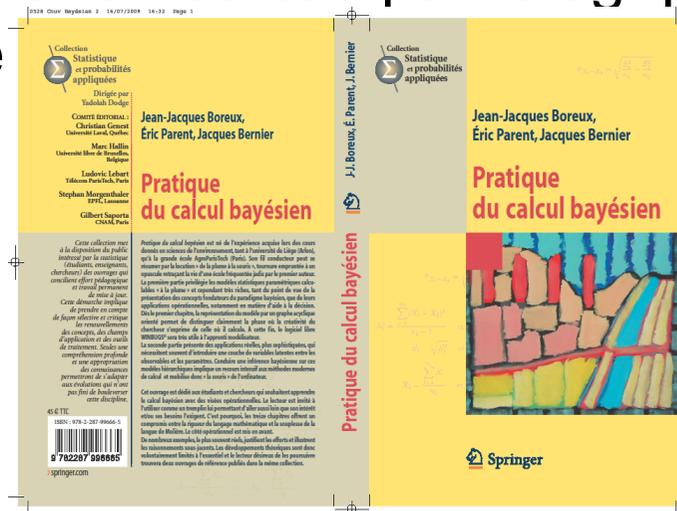
Ellison 2004. Bayesian inference in Ecology. Ecology Letter. 7: 509-520

Clark, J.S. 2005. Why environmental scientists are becoming Bayesians ?
Ecology Letters, 8: 2-14

Jeter des ponts

Le monde des expérimentateurs est actuellement imperméable, pour ne pas dire rebelle aux discours tenus par les statisticiens. En gros, c'est un monde de gens qui ont des besoins méthodologiques mais peu d'outils !

A contrario, il existe un monde monodisciplinaire, celui des mathématiciens. Dans ce monde, tous partagent le même intérêt et parfois le même enthousiasme pour dissenter sur telle technique mathématique astucieuse, améliorer d'un fifrelin tel estimateur, découvrir et répertorier les cas pathologiques, etc. Bref, souvent le monde des outils et peu de besoins





Réf. - Analyse bayésienne (général)

Bernier, J., Parent, E., Boreux, J.J. 2000. Statistiques pour l'environnement. Traitement Bayésien de l'incertitude. Éditions Tech. et Doc.

Box G.E.P. et Tiao G.C. 1992. Bayesian inference in statistical analysis. John Wiley and Sons, N-Y and London. Col. Wiley Classics Library. 588p.

Gelman, A., Carlin J.B., Stern H. and Rubin D.B. 2006. Bayesian data analysis. Second Edition. Chapman & Hall, London, New-York. Col. Texts in Statistical Science, 668 pp.

Hilborn, R., and Mangel, M. 1997. The ecological detective. Confronting models with data. Monographs in population biology, 28. Princeton University Press. Princeton, New Jersey.

Congdon P. 2001. Bayesian Statistical Modelling. Wiley. Wiley series in probability and statistics.

Berger, J.O. 1985. Statistical decision theory and bayesian analysis. 2nd edition. Springer. Springer series in statistics.

Kadane, J. B. 2011. Principles of Uncertainty. CRC Press. Chapman & Hall

Bernardo, J. M. and Smith, A. F. M. 1994. Bayesian theory. Wiley, London, 1 edition

Makowski, D., Monod, H. 2011. Analyse statistique des risques agro-environnementaux. Springer

Foreword de Kadane, 2011

With respect to the mathematical parts of this book, I can offer no better advice than

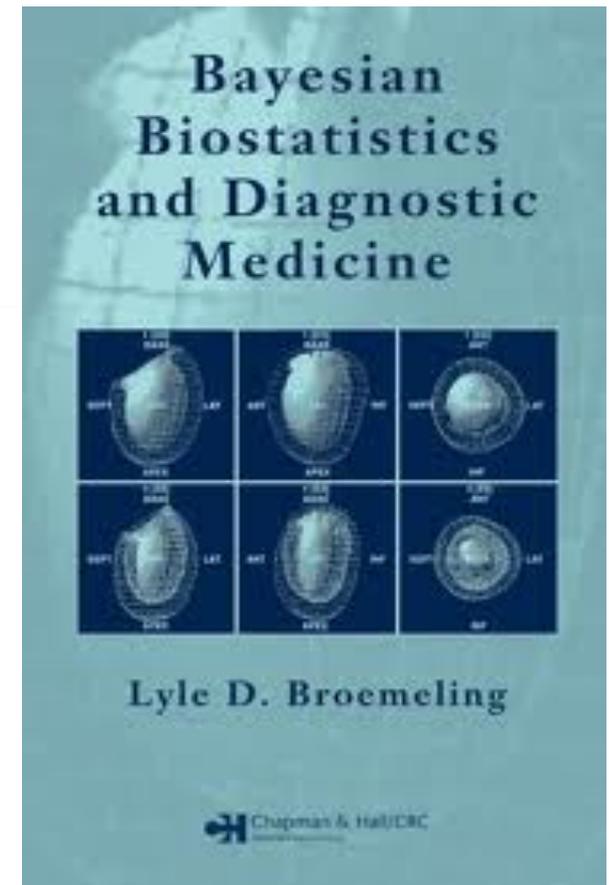
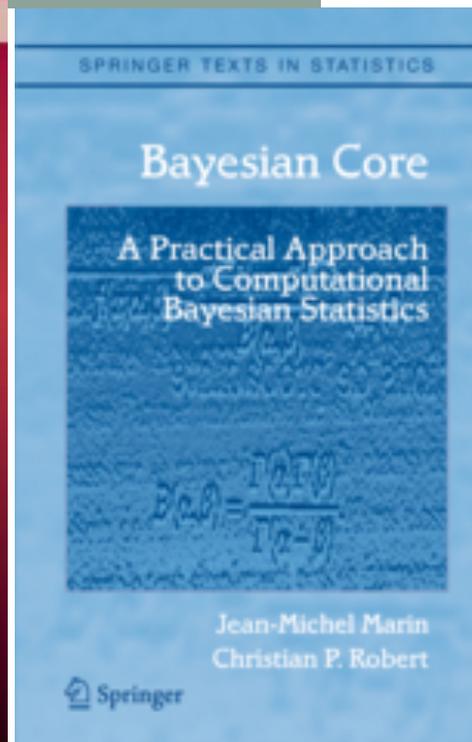
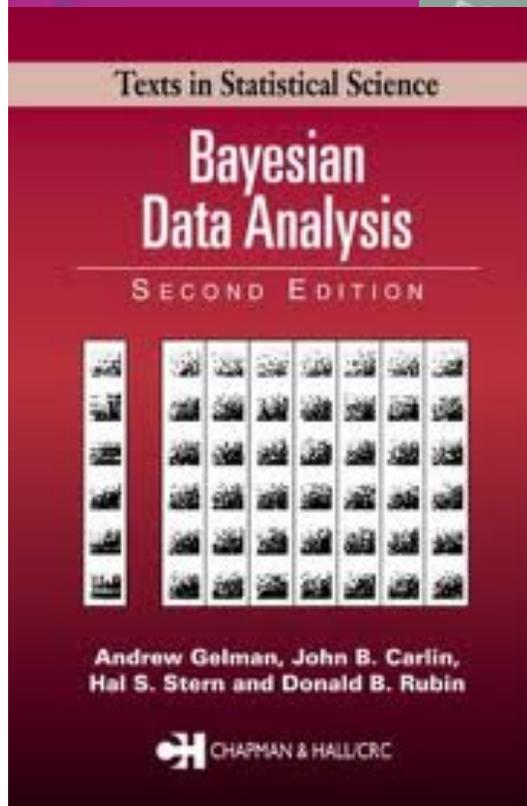
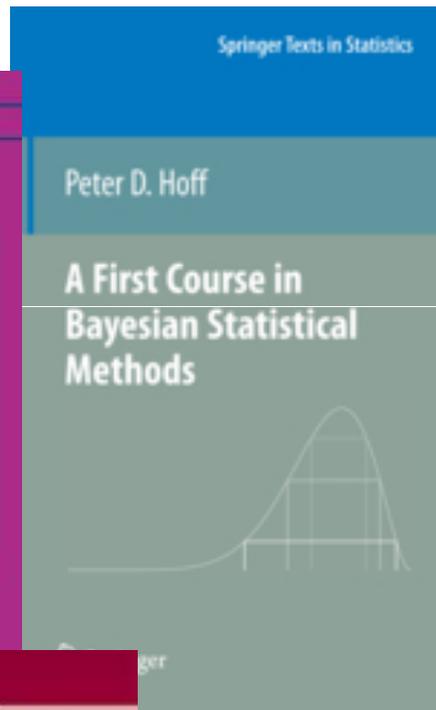
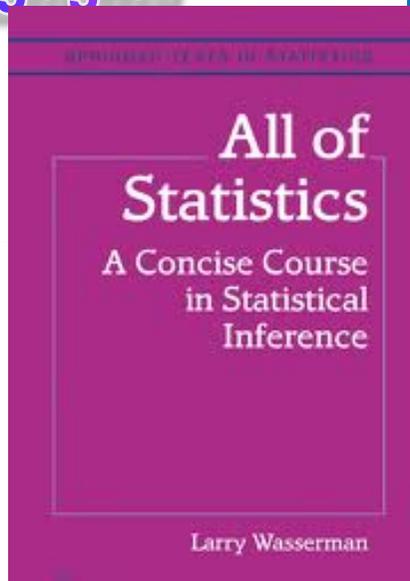
(Halmos, 1985, p. 69):

...study actively. Don't just read it: fight it! Ask your own questions, look for your own examples, discover your own proofs. Is the hypothesis necessary? Is the converse true? What happens in the classical special case? What about the degenerate cases?

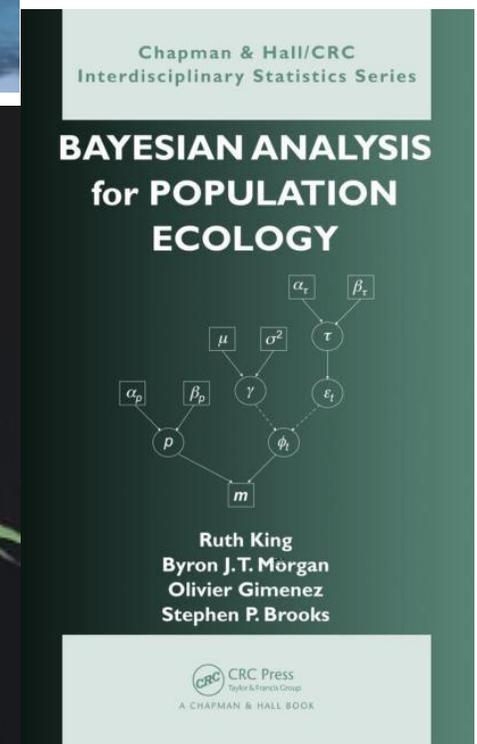
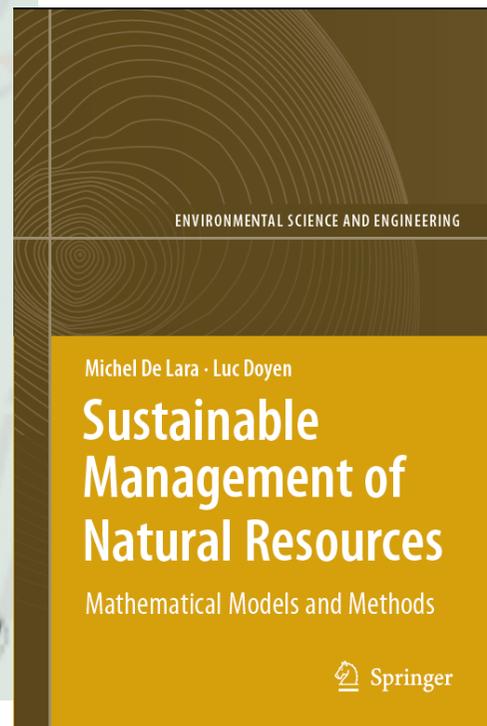
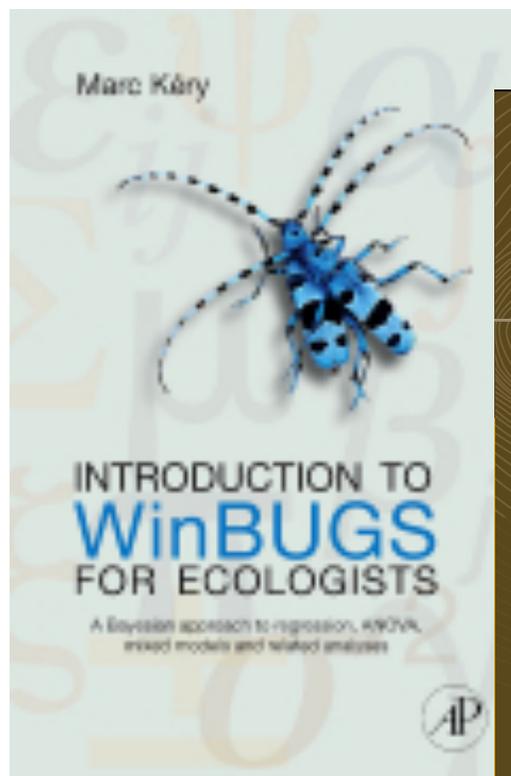
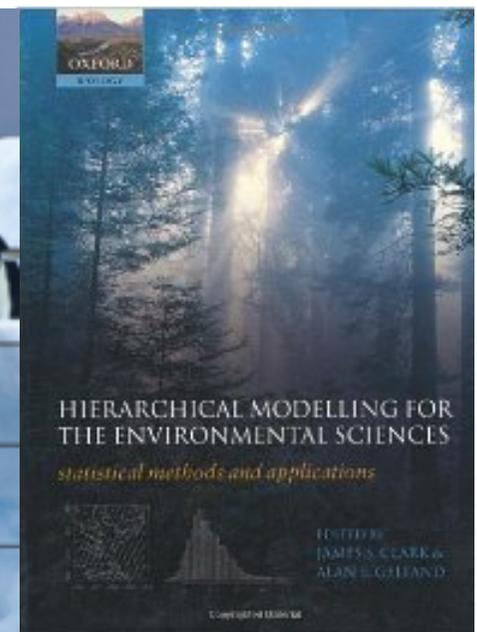
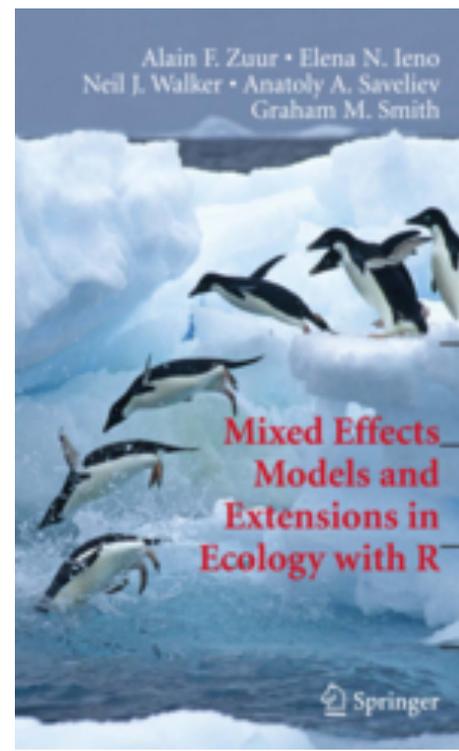
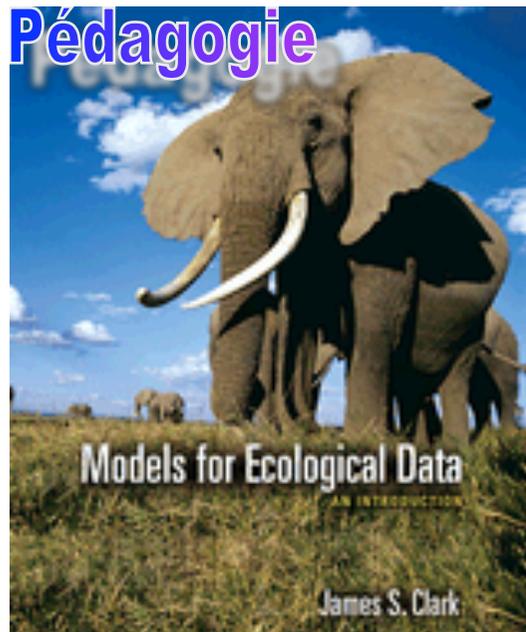
Where does the proof use the hypothesis?

In addition, for this book, it is relevant to ask "what does this result mean for understanding uncertainty? If it is a stepping stone, toward what is it a stepping stone? If this result were false, what consequences would that have?"

Pédagogie

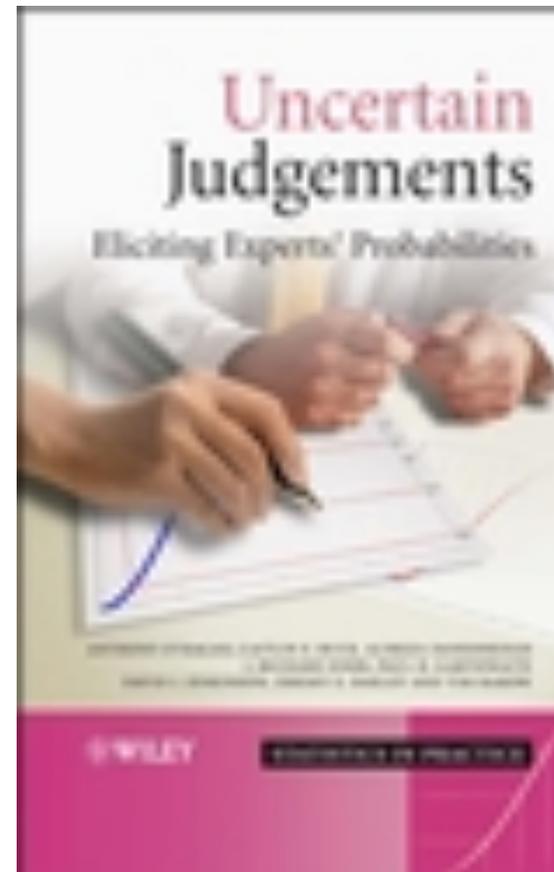


Pédagogie

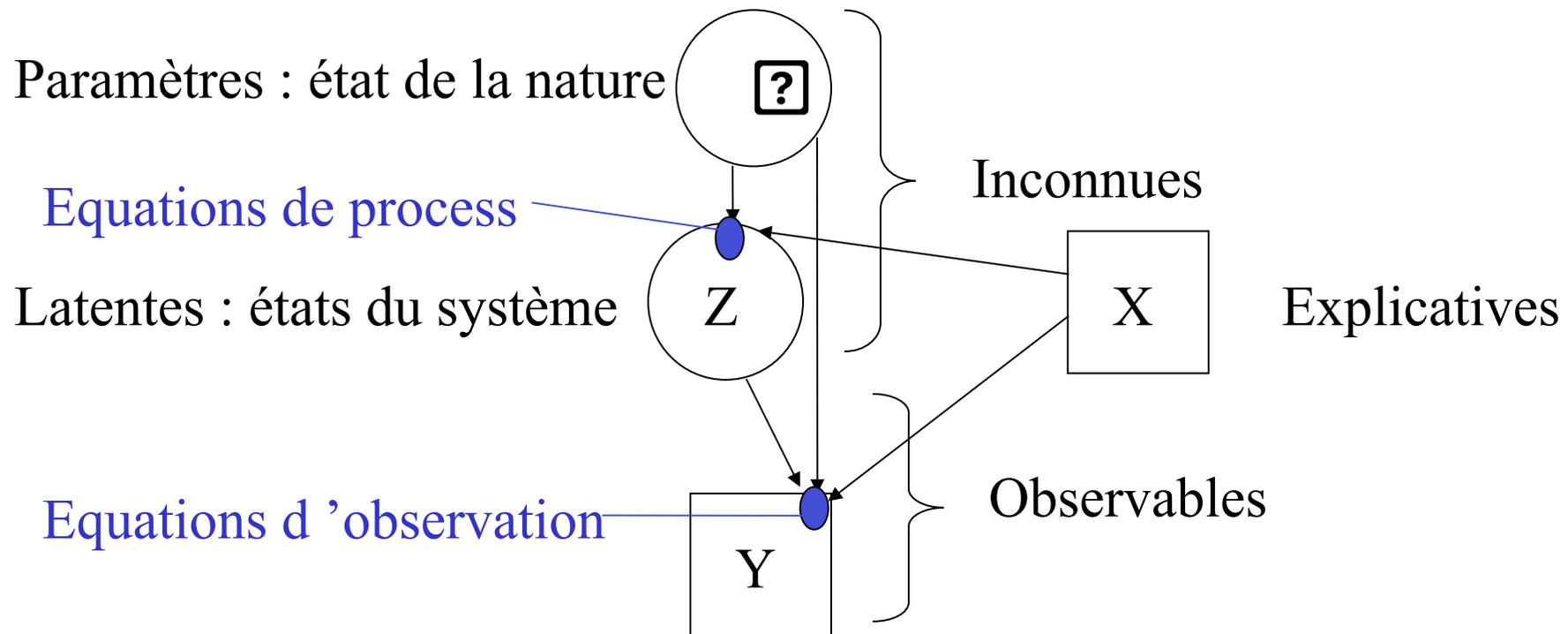


Encodage probabiliste d'expertise ou *élicitation*

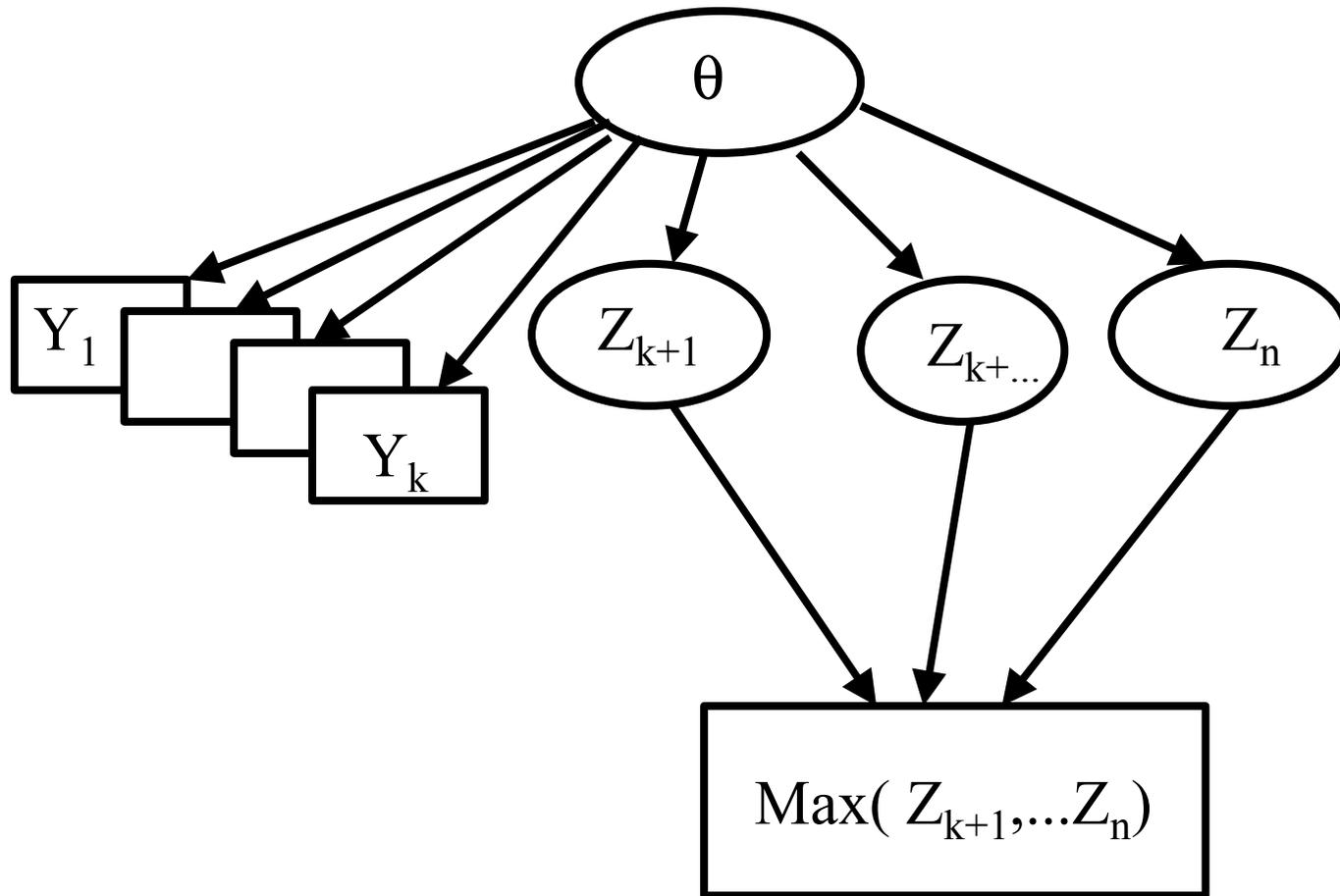
- Prior informatif
- Prior vague
- Convenance mathématique
- Principe d'entropie
- ...
- Refus par principe en stat classique.



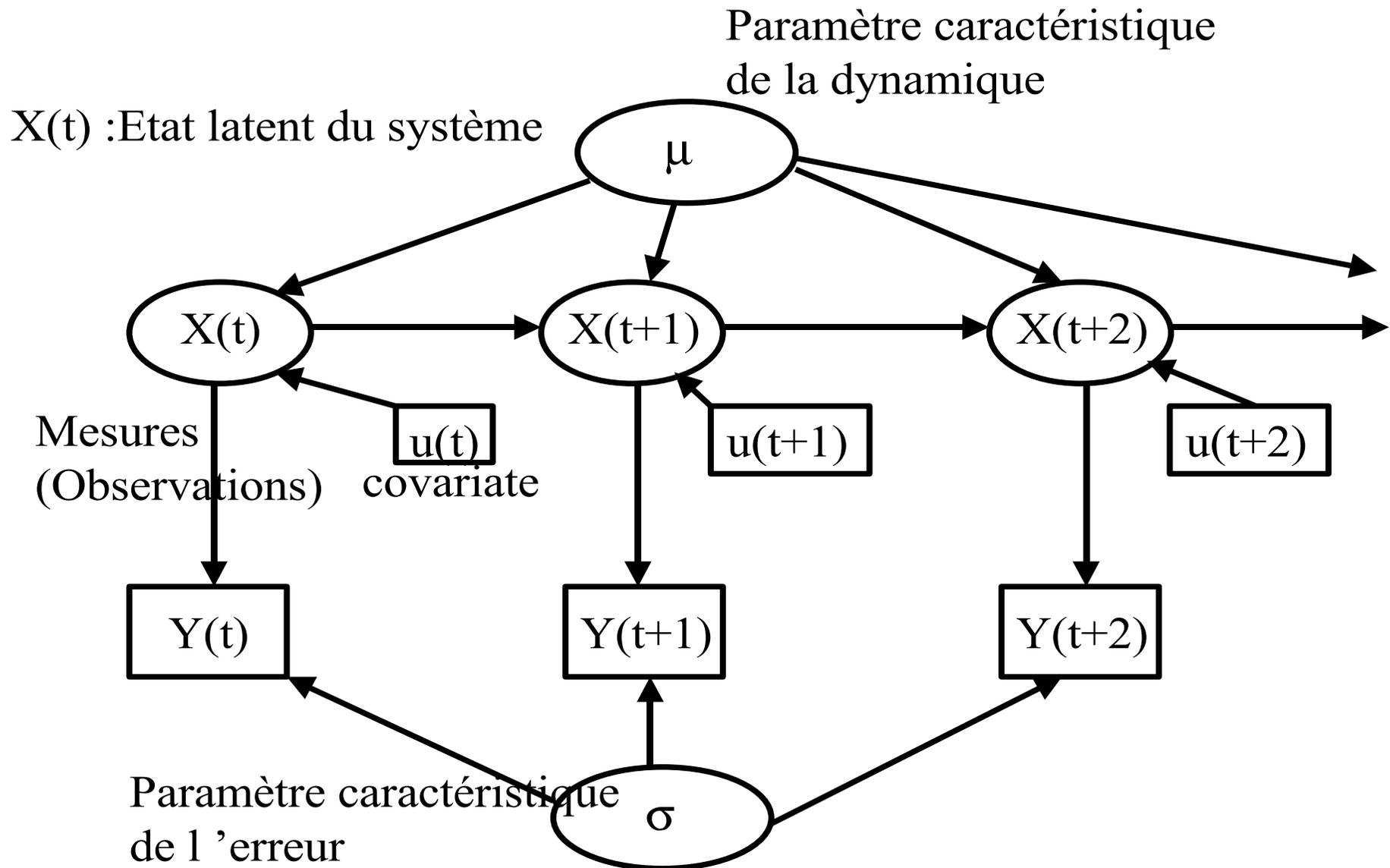
Modélisation hiérarchique



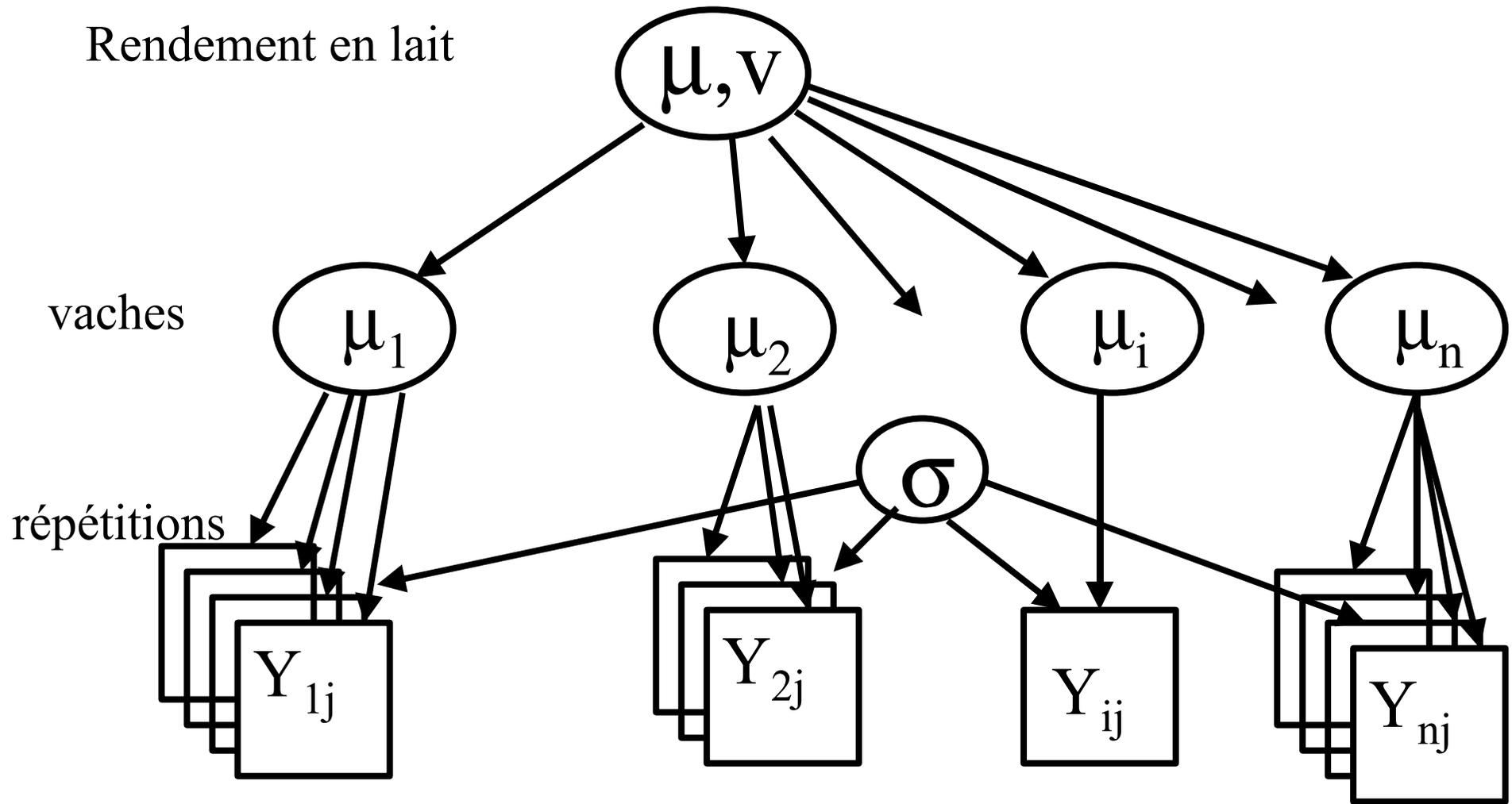
NB: Inférence bayésienne = Renverser le conditionnement



Structure graphique de modèle à données manquantes

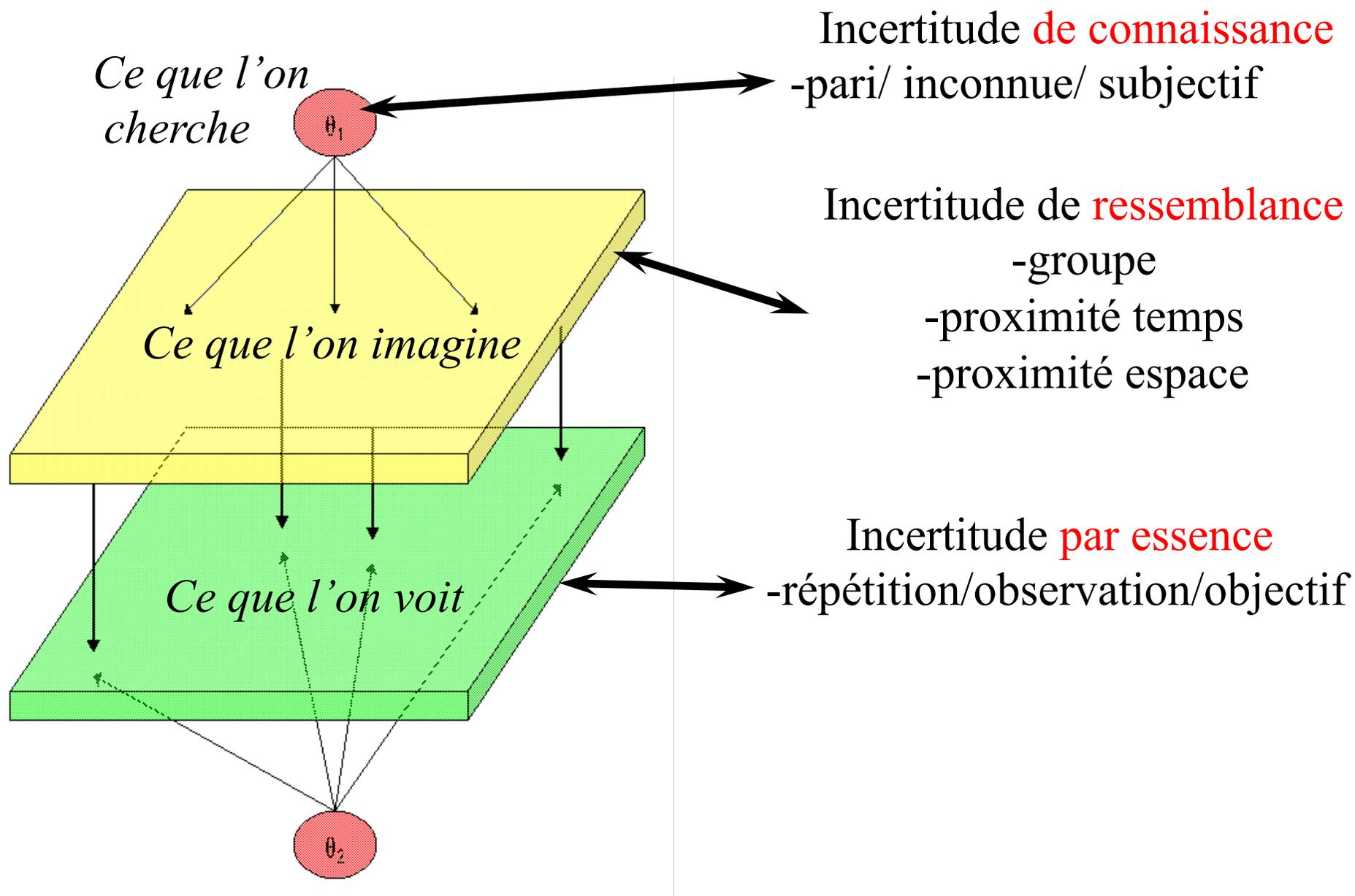


Modèle dynamique à temps discret



Graphe acyclique orienté d'un modèle hiérarchique de population

Philosophie : Encodage probabiliste de l'incertitude



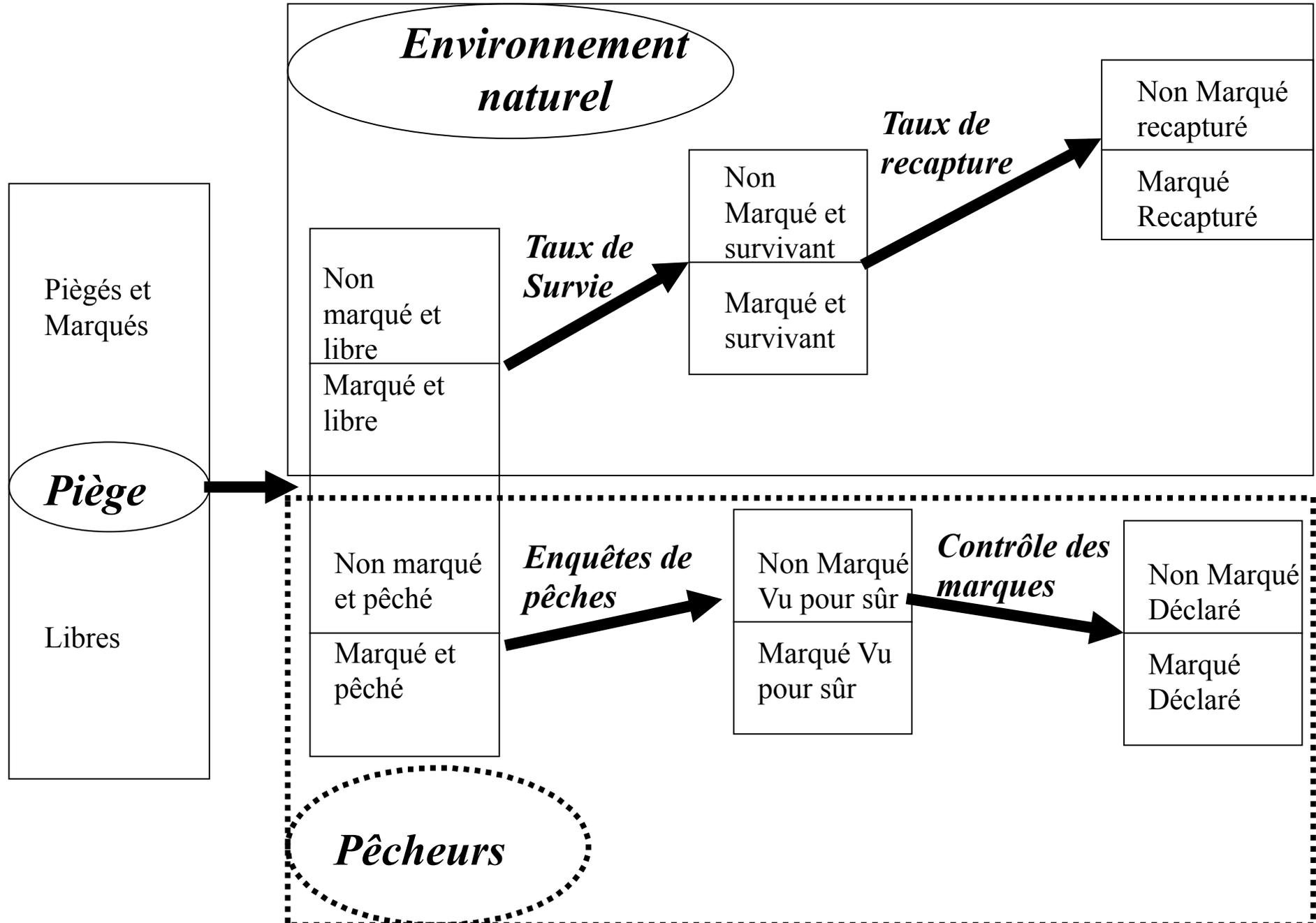
Exemple de modélisation hiérarchique bayésienne



Exemple de modélisation hiérarchique bayésienne



Exemple de modélisation hiérarchique bayésienne



DONNEES

	marqués	Déclarés marqués	Déclarés non marqués	Pêchés pour sûr	Capturés marqués	Capturés non marqués
année	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6
1994	156	3	14	42	4	14
1995	500	39	10	75	31	28
1996	502	25	8	87	45	14
1997	320	17	7	33	19	9
1998	442	50	5	66	56	13
1999	167	16	4	24	16	11

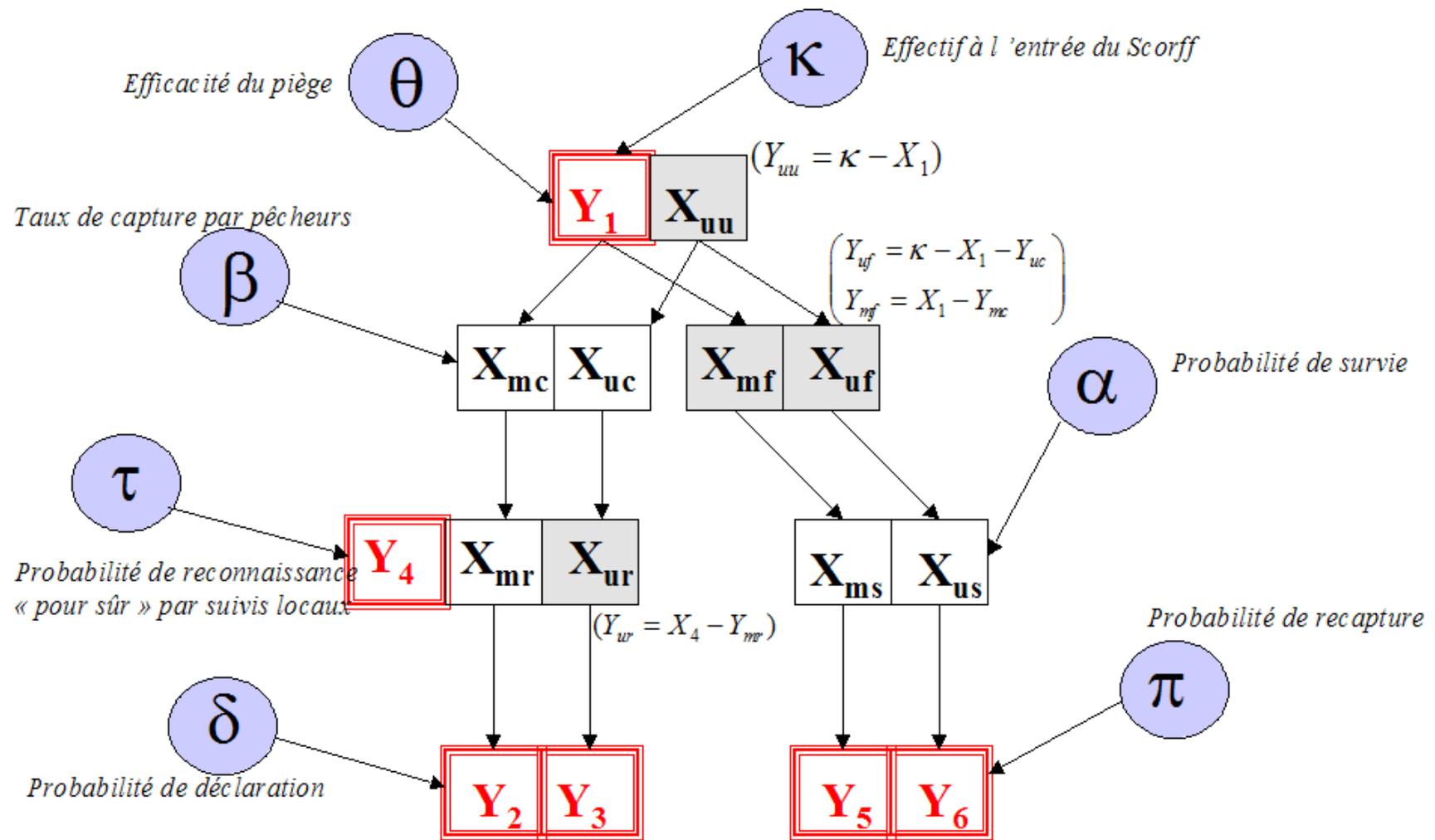
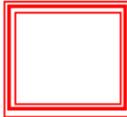
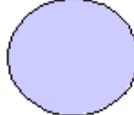


DIAGRAMME D'INFLUENCE


 Variable latente


 Variable observée


 Paramètres de comportement


 Intermédiaire déterministe