Les Réseaux Bayésiens

Introduction Intuitive

(Jean-Baptiste.Denis@jouy.inra.fr)

PLAN

(1) A partir d'exemples
- botanique
- médical
- classique

(2) Réseau Bayésien

(3) Réseaux bayésiens et statistique bayésienne

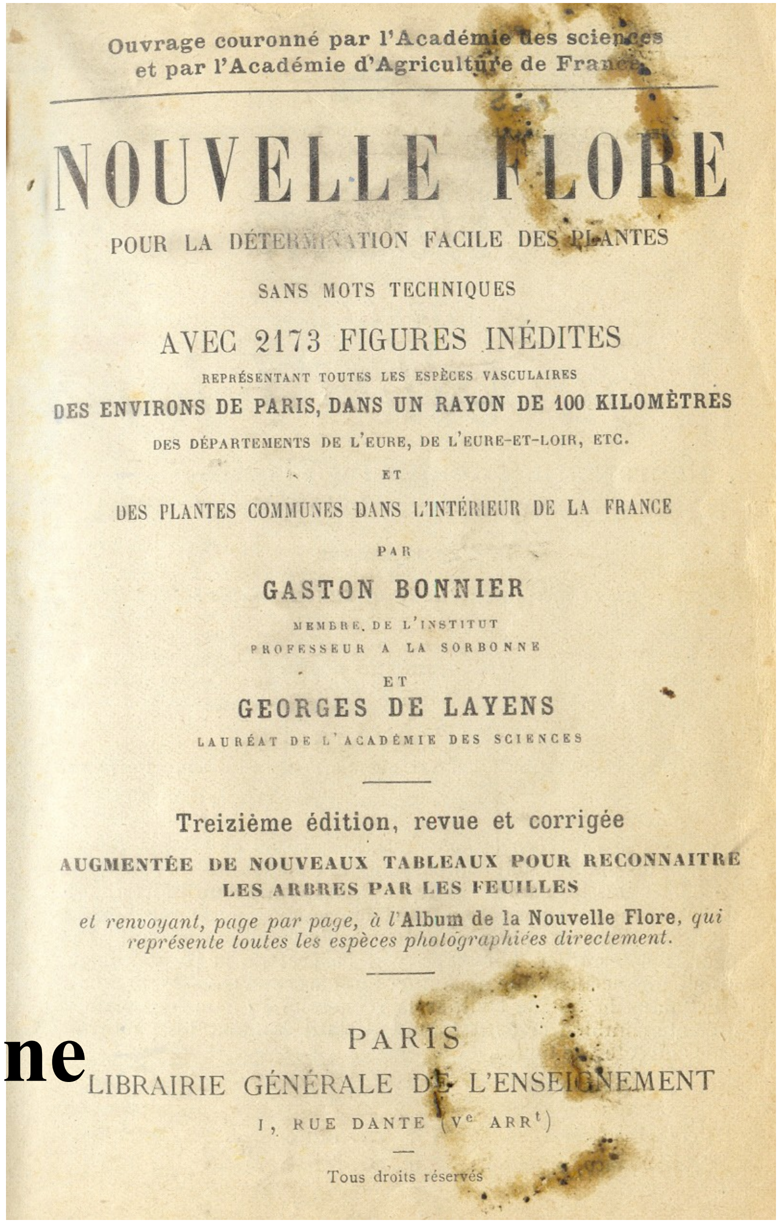
(4) Résaux bayésiens et apprentissage

(5) Réseaux bayésiens et réseaux de Markov

EXEMPLE BOTANIQUE

Gaston Bonnier

1853 - 1922



Ouvrage couronné par l'Académie des sciences
et par l'Académie d'Agriculture de France

NOUVELLE FLORE

POUR LA DÉTERMINATION FACILE DES PLANTES

SANS MOTS TECHNIQUES

AVEC 2173 FIGURES INÉDITES

REPRÉSENTANT TOUTES LES ESPÈCES VASCULAIRES

DES ENVIRONS DE PARIS, DANS UN RAYON DE 100 KILOMÈTRES

DES DÉPARTEMENTS DE L'EURE, DE L'EURE-ET-LOIR, ETC.

ET

DES PLANTES COMMUNES DANS L'INTÉRIEUR DE LA FRANCE

PAR

GASTON BONNIER

MEMBRE DE L'INSTITUT
PROFESSEUR A LA SORBONNE

ET

GEORGES DE LAYENS

LAURÉAT DE L'ACADÉMIE DES SCIENCES

Treizième édition, revue et corrigée

AUGMENTÉE DE NOUVEAUX TABLEAUX POUR RECONNAITRE
LES ARBRES PAR LES FEUILLES

*et renvoyant, page par page, à l'*Album de la Nouvelle Flore, *qui représente toutes les espèces photographiées directement.*

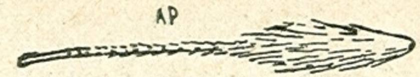
PARIS
LIBRAIRIE GÉNÉRALE DE L'ENSEIGNEMENT
1, RUE DANTE (Ve ARRt)

Tous droits réservés

Détermination d'espèces : **l'anémone**

EXEMPLE BOTANIQUE

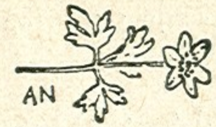
3. Anémone. *Anemone.* —
Styles *plumeux* AP ; fleurs *violettes*, velues, PL ; 1-4 d.



A. Pulsatille AC.
A. Pulsatilla L. ✠
prés découverts ; av.-j. ; *v.*

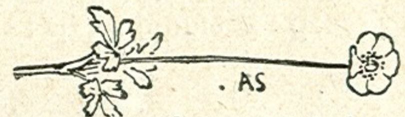
Styles *non plumeux*

- Fleurs *blanches ou roses*
  - Fleurs *sans poils* : feuilles de la base développées avec les fleurs ; 1-3 d. fig. AN.



A. Sylvie TC. α
A. nemorosa L.
bois ; ms.-av. ; *v.*

  - Fleurs *velues* ; feuilles de la base développées après les fleurs ; 2-5 d. fig. AS.

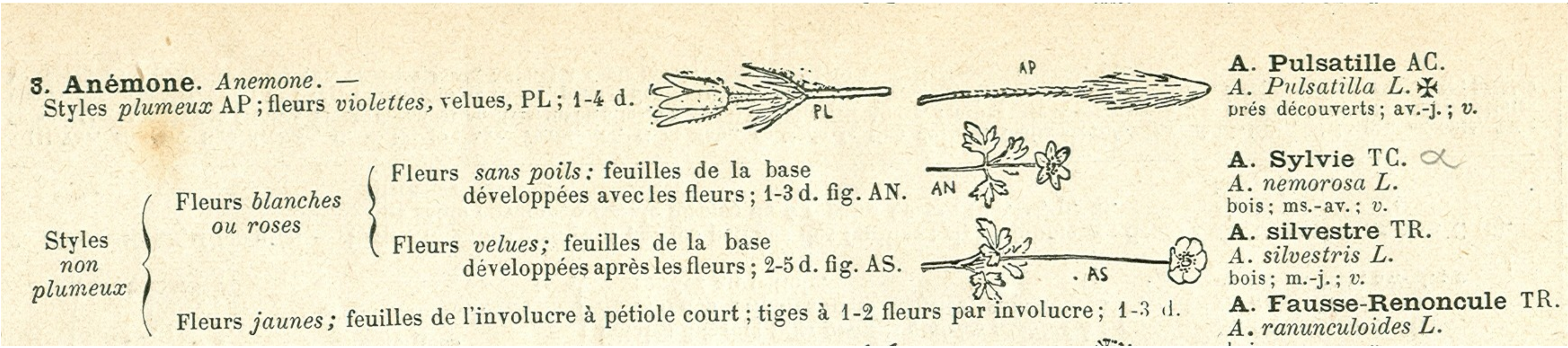


A. silvestre TR.
A. silvestris L.
bois ; m.-j. ; *v.*

- Fleurs *jaunes* ; feuilles de l'involucre à pétiole court ; tiges à 1-2 fleurs par involucre ; 1-3 d.

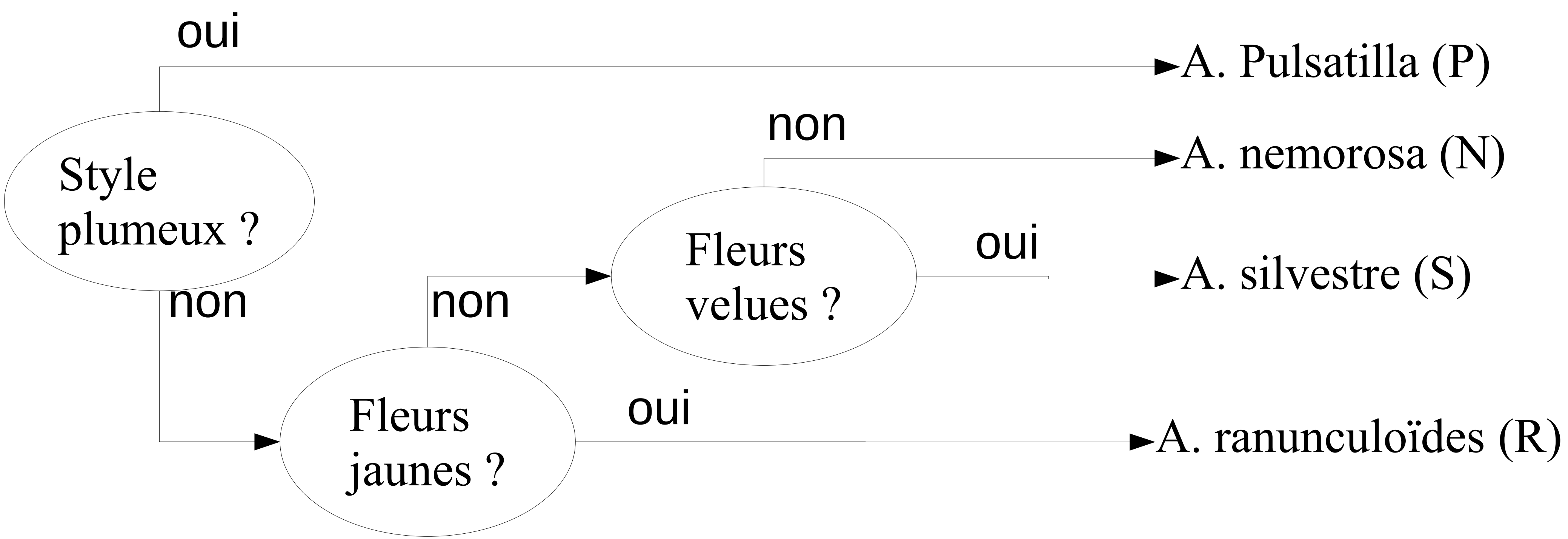
A. Fausse-Renoncule TR.
A. ranunculoides L.

EXEMPLE BOTANIQUE



3. **Anémone.** *Anemone.* — Styles *plumeux* AP ; fleurs *violettes*, velues, PL ; 1-4 d. — **A. Pulsatille** AC. *A. Pulsatilla L.* ✠ prés découverts ; av.-j. ; *v.*

Styles *non plumeux*
- Fleurs *blanches ou roses*
  - Fleurs *sans poils* : feuilles de la base développées avec les fleurs ; 1-3 d. fig. AN. — **A. Sylvie** TC. *A. nemorosa L.* bois ; ms.-av. ; *v.*
  - Fleurs *velues* ; feuilles de la base développées après les fleurs ; 2-5 d. fig. AS. — **A. silvestre** TR. *A. silvestris L.* bois ; m.-j. ; *v.*
- Fleurs *jaunes* ; feuilles de l'involucre à pétiole court ; tiges à 1-2 fleurs par involucre ; 1-3 d. — **A. Fausse-Renoncule** TR. *A. ranunculoides L.*



Unité MIA Jouy-en-Josas

EXEMPLE BOTANIQUE

Cette simplification mathématique n'est **pas raisonnable** !

- Si la question est mal comprise
- Si l'observation est mal faite
- Si la flore de la Région Parisienne a évolué (ou que l'exemplaire n'en provienne pas)
- Si ce n'était pas une anémone

La redondance n'est pas toujours inutile !

- MAIS AUSSI, si les choses n'étaient pas si tranchées ?

EXEMPLE BOTANIQUE

Recensons l'information disponible

Nom	Fleurs velues	Style	Couleur	Fréquence
P	oui	plumeux	violette	assez commune
N	non	non plumeux	blanche/rose	très commune
S	oui	non plumeux	blanche/rose	très rare
R	?	non plumeux	jaune	très rare

EXEMPLE BOTANIQUE

Traduisons la en probabilités

Nom	Fleurs velues	Style	Couleur	Fréquence
P	oui	plumeux	violette	assez commune
N	non	non plumeux	blanche/rose	très commune
S	oui	non plumeux	blanche/rose	très rare
R	?	non plumeux	jaune	très rare

Probibilité de la couleur sachant le nom

	violette	rose	blanche	jaune
Pulsatilla	0.9	0.05	0.05	0.0
nemorosa	0.0	0.50	0.50	0.0
silvesris	0.0	0.60	0.40	0.0
ranuncoloïdes	0.1	0.00	0.00	0.9

Probabilité conditionnelle : p(c|n)

EXEMPLE BOTANIQUE

Nom	Fleurs velues	Style	Couleur	Fréquence
P	oui	plumeux	violette	assez commune
N	non	non plumeux	blanche/rose	très commune
S	oui	non plumeux	blanche/rose	très rare
R	?	non plumeux	jaune	très rare

Probibilité du style sachant le nom

	S_plumeux	S_sans_plume
Pulsatilla	0.99	0.01
nemorosa	0.20	0.80
silvesris	0.10	0.90
ranuncoloïdes	0.05	0.95

Probabilité conditionnelle : p(s|n)

EXEMPLE BOTANIQUE

Nom	Fleurs velues	Style	Couleur	Fréquence
P	oui	plumeux	violette	assez commune
N	non	non plumeux	blanche/rose	très commune
S	oui	non plumeux	blanche/rose	très rare
R	?	non plumeux	jaune	très rare

Probibilité de la feuille sachant le nom

	F_poilues	F_sans_poils
Pulsatilla	0.95	0.05
nemorosa	0.01	0.99
silvesris	0.80	0.20
ranuncoloïdes	0.40	0.60

Probabilité conditionnelle : p(f|n)

EXEMPLE BOTANIQUE

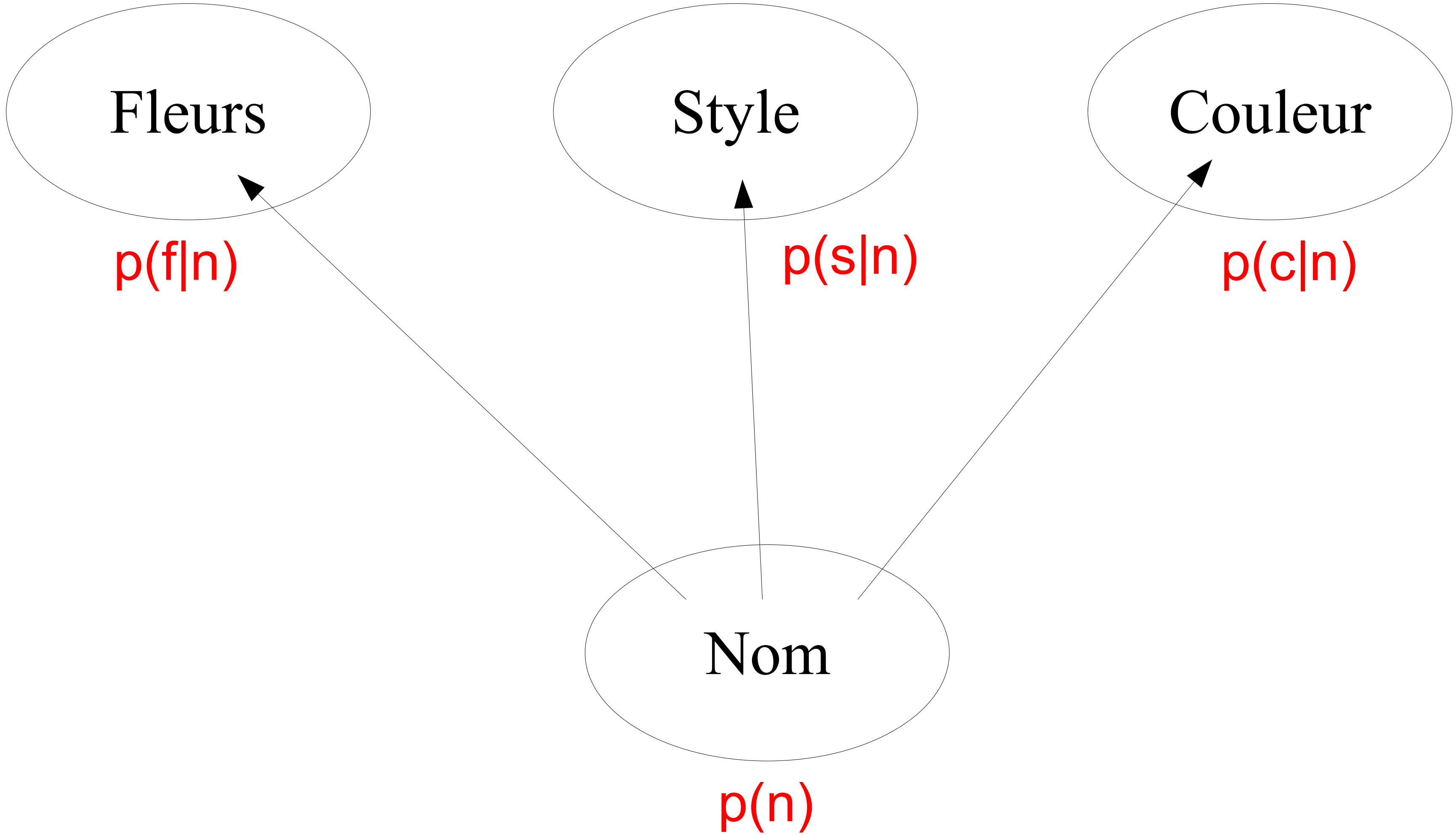
Nom	Fleurs velues	Style	Couleur	Fréquence
P	oui	plumeux	violette	assez commune
N	non	non plumeux	blanche/rose	très commune
S	oui	non plumeux	blanche/rose	très rare
R	?	non plumeux	jaune	très rare

	freq
Pulsatilla	0.10
nemorosa	0.88
silvesris	0.01
ranuncoloïdes	0.01

Probabilité marginale : p(n)

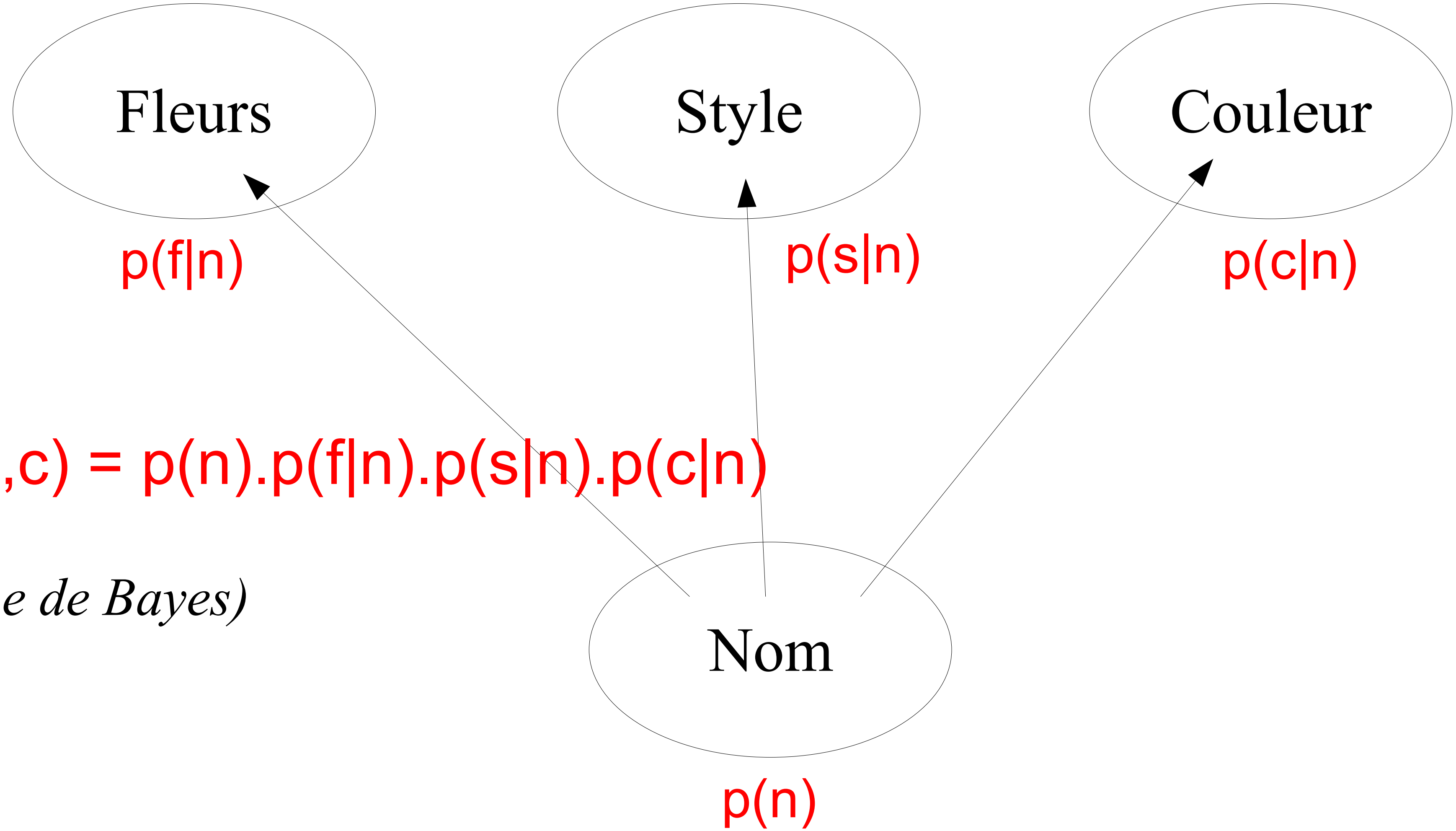
EXEMPLE BOTANIQUE

Le Réseau Bayésien associé



EXEMPLE BOTANIQUE

Le Réseau Bayésien associé exhibe une probabilité conjointe



$$p(n,f,s,c) = p(n).p(f|n).p(s|n).p(c|n)$$

(théorème de Bayes)

EXEMPLE BOTANIQUE

Le Réseau Bayésien peut servir à l'identification p(n,f,s,c) -> p(n|c)

Probibilité du nom sachant la couleur

	violette	rose	blanche	jaune
Pulsatilla	<u>0.989</u>	0.011	0.011	0
nemorosa	0.000	<u>0.976</u>	<u>0.980</u>	0
silvesris	0.000	0.013	0.009	0
ranuncoloïdes	0.011	0.000	0.000	<u>1</u>

EXEMPLE BOTANIQUE

Le Réseau Bayésien peut servir à l'identification p(n,f,s,c) -> p(n|s)

Probibilité du nom sachant le style

	S_plumeux	S_sans_plume
Pulsatilla	<u>0.820</u>	0.006
nemorosa	0.076	<u>0.985</u>
silvesris	0.069	0.002
ranuncoloïdes	0.035	0.007

EXEMPLE BOTANIQUE

Le Réseau Bayésien peut servir à l'identification p(n,f,s,c) -> p(n|f,s)

Probibilité du nom sachant la feuille et le style

, , S_plumeux

	F_poilues	F_sans_poils
Pulsatilla	0.918	0.272
nemorosa	0.001	0.499
silvesris	0.065	0.092
ranuncoloïdes	0.016	0.137

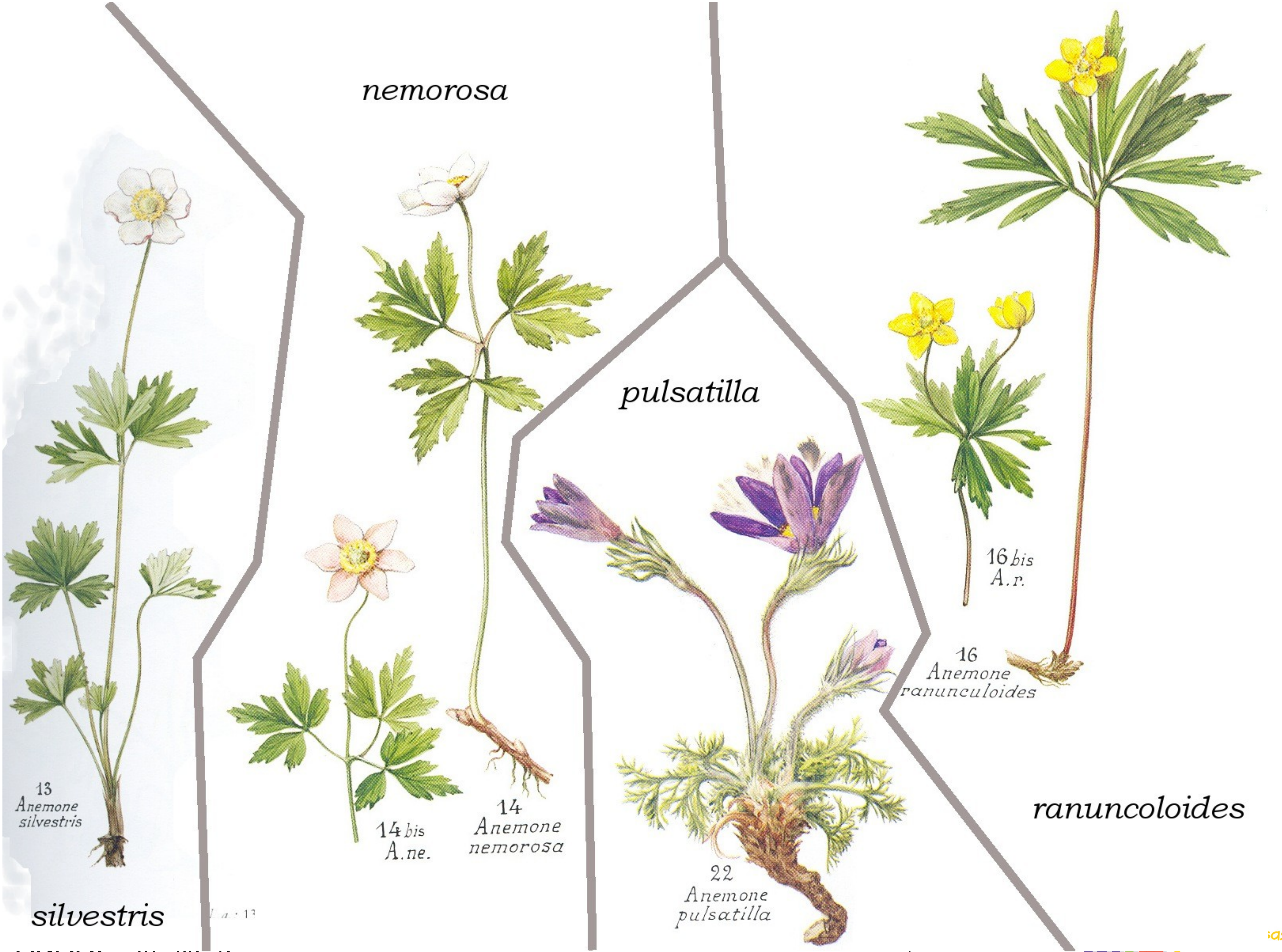
, , S_sans_plume

	F_poilues	F_sans_poils
Pulsatilla	0.272	0.000
nemorosa	0.499	0.995
silvesris	0.092	0.000
ranuncoloïdes	0.137	0.004

EXEMPLE BOTANIQUE

Le Réseau Bayésien peut servir à l'identification $p(n,f,s,c) \rightarrow p(n|f,s,c)$

```
                 , , F_poilues, S_plumeux
                violette  rose blanche jaune
Pulsatilla         0.998 0.537   0.634     0
nemorosa           0.000 0.005   0.006     0
silvesris          0.000 0.457   0.360     0
ranuncoloïdes      0.002 0.000   0.000     1
                 , , F_sans_poils, S_plumeux
                violette  rose blanche jaune
Pulsatilla         0.947 0.043   0.045     0
nemorosa           0.000 0.784   0.832     0
silvesris          0.000 0.173   0.122     0
ranuncoloïdes      0.053 0.000   0.000     1
                 , , F_poilues, S_sans_plume
                violette  rose blanche jaune
Pulsatilla         0.947 0.043   0.045     0
nemorosa           0.000 0.784   0.832     0
silvesris          0.000 0.173   0.122     0
ranuncoloïdes      0.053 0.000   0.000     1
  ......................
```


nemorosa
pulsatilla
ranuncoloides
silvestris
13 Anemone silvestris
14 bis A. ne.
14 Anemone nemorosa
22 Anemone pulsatilla
16 bis A. r.
16 Anemone ranunculoides

EXEMPLE MEDICAL

Diagnostic Médical : infection à virus

On définit deux variables aléatoires :

- V : l'individu considéré est infecté à valeurs dans *{0,1}*
- T : le résultat d'un test de détection à valeurs dans *{-,+}*

Des études précédentes ont permis de préciser la sensibilité et la spécificité de la détection :

- $P(T=+ \mid V=1) = 0.98$ (= spécificité)
- $P(T=+ \mid V=0) = 0.01$ (= 1 - sensibilité)

Question pratique : le résultat du test est positif, quelle est la probabilité que l'individu considéré soit infecté ?

EXEMPLE MEDICAL

Réponse : on ne peut pas répondre ! En effet

Si je dis «*Dans une classe la moitié des garçons portent lunettes mais seulement le tiers des filles*», on ne peut rien inférer sur le sexe d'un porteur de lunettes pris au hasard dans la classe !

CLASSE 1	avec	sans	TOTAUX
Filles	4	8	12
Garçons	4	4	8
TOTAUX	8	12	20

CLASSE 2	avec	sans	TOTAUX
Filles	1	2	3
Garçons	9	9	18
TOTAUX	10	11	21

Classe 1 : 8 garçons et 12 filles => 0.5 que ce soit une fille

Classe 2 : 18 garçons et 3 filles => 0.1 que ce soit une fille

EXEMPLE MEDICAL

Pour répondre : il faut savoir aussi que $P(V=1) = 0.002$
alors :

$$P(V=1 \mid T=+) = \left[P(T=+ \mid V=1) * P(V=1)\right] / P(T=+)$$

or $$P(T=+) = \left[P(T=+ \mid V=1) * P(V=1)\right] + \left[P(T=+ \mid V=0) * P(V=0)\right]$$
$$= (0.98 * 0.002) + (0.01 * 0.998) = 0.01194$$

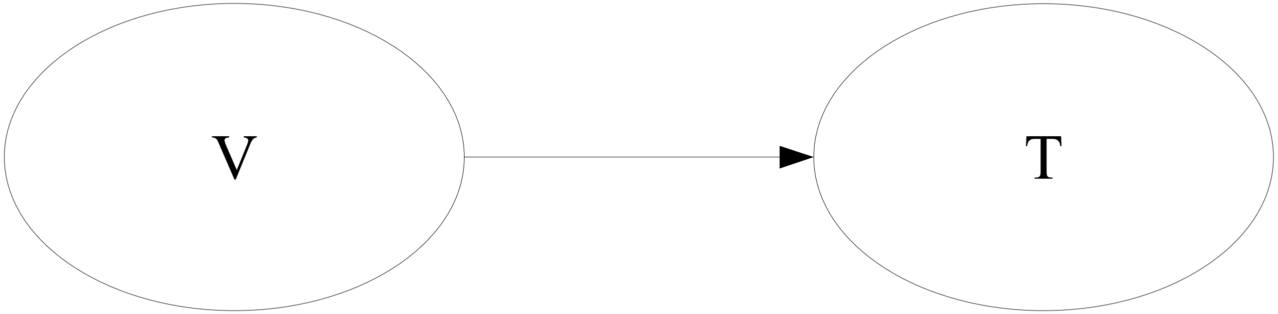
et $$P(V=1 \mid T=+) = (0.98 * 0.002) / 0.01194 = \mathbf{0.164}$$

(théorème de Bayes)

EXEMPLE MEDICAL

Diagnostic Médical : infection à virus

Sous forme de réseau bayésien :

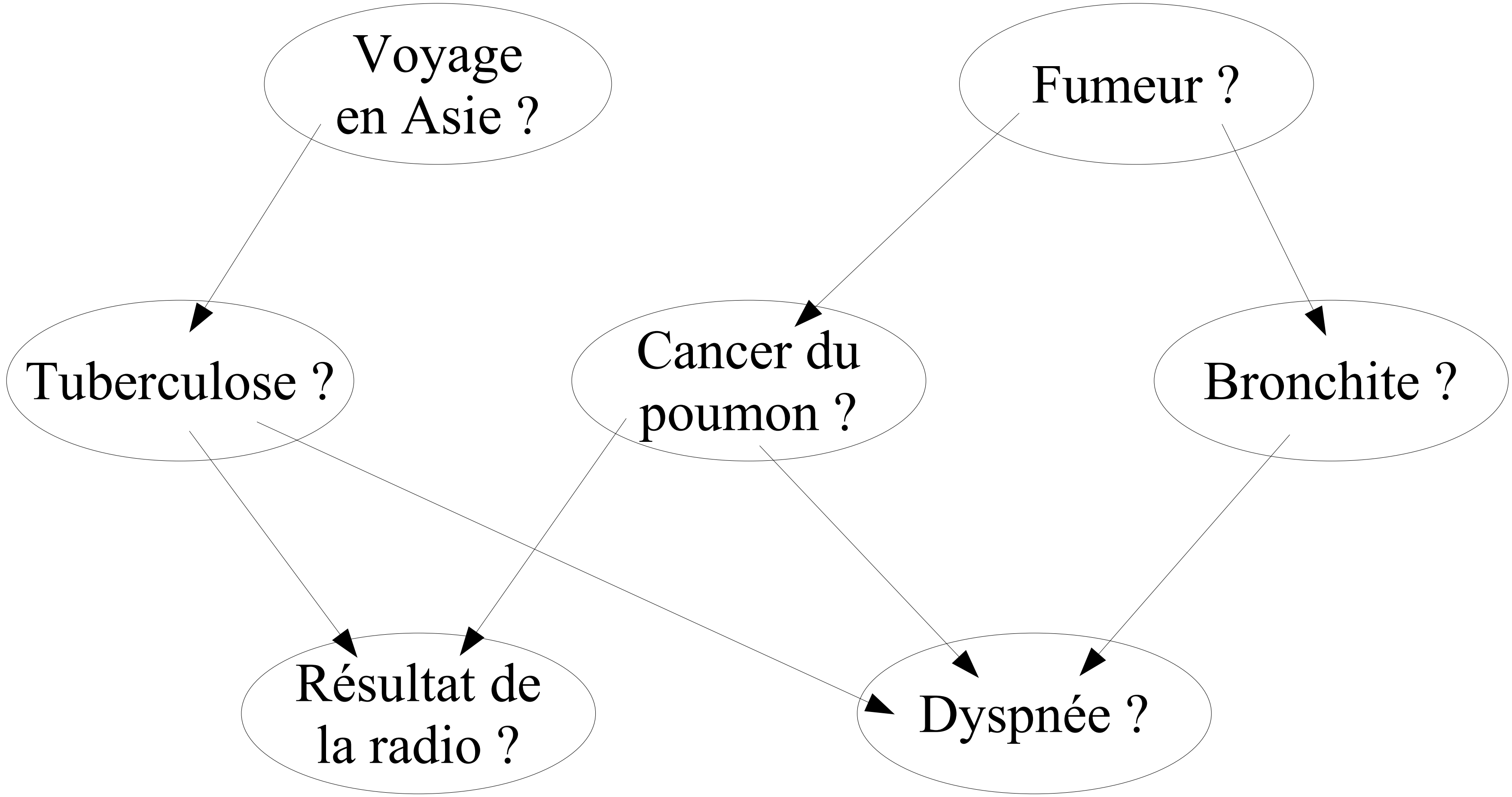


P(V) P(T | V)

Et la réponse est donnée par P(V|T)

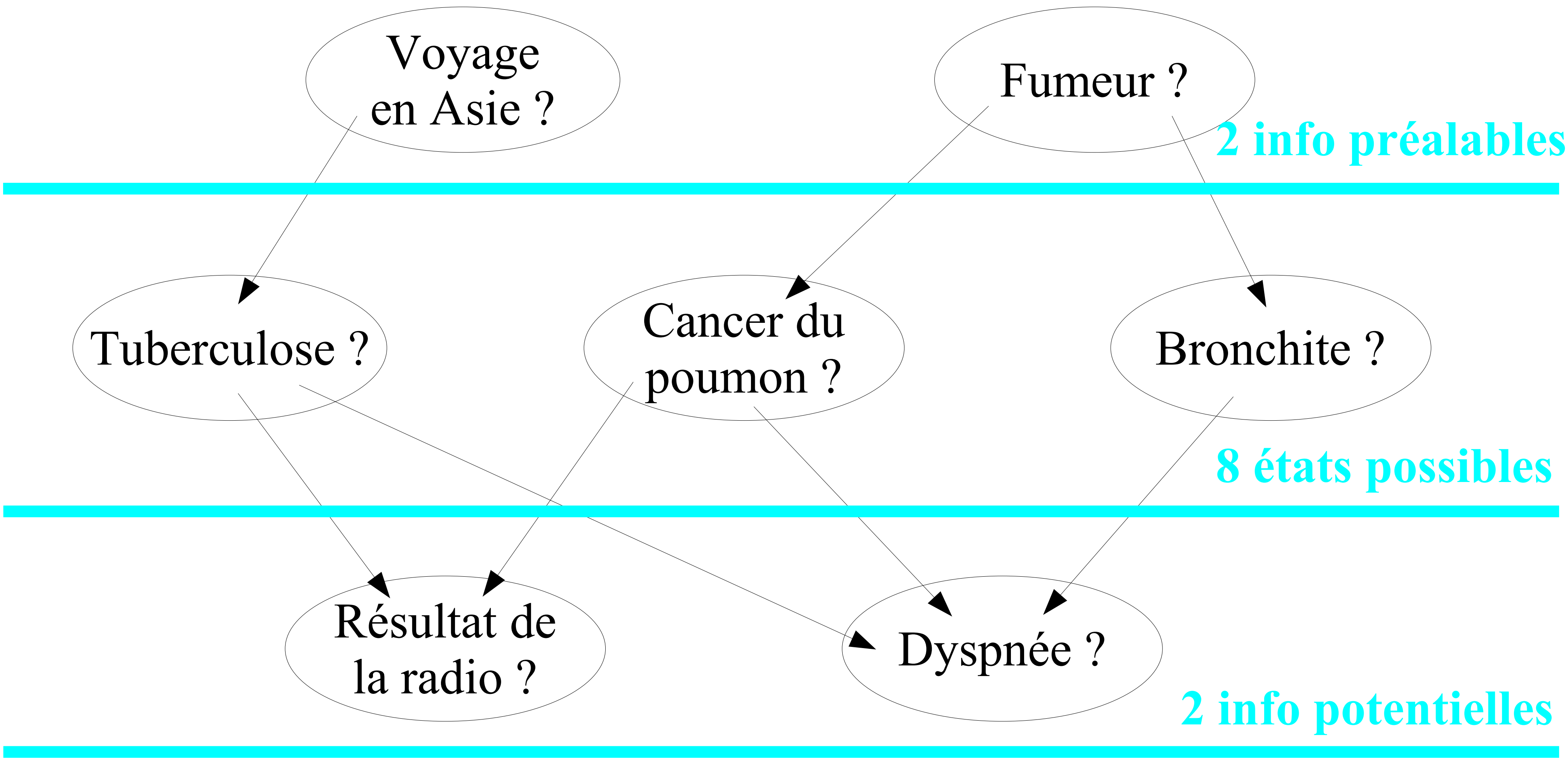
EXEMPLE CLASSIQUE

The Asian network (Lauritzen *et al.*, 1988)



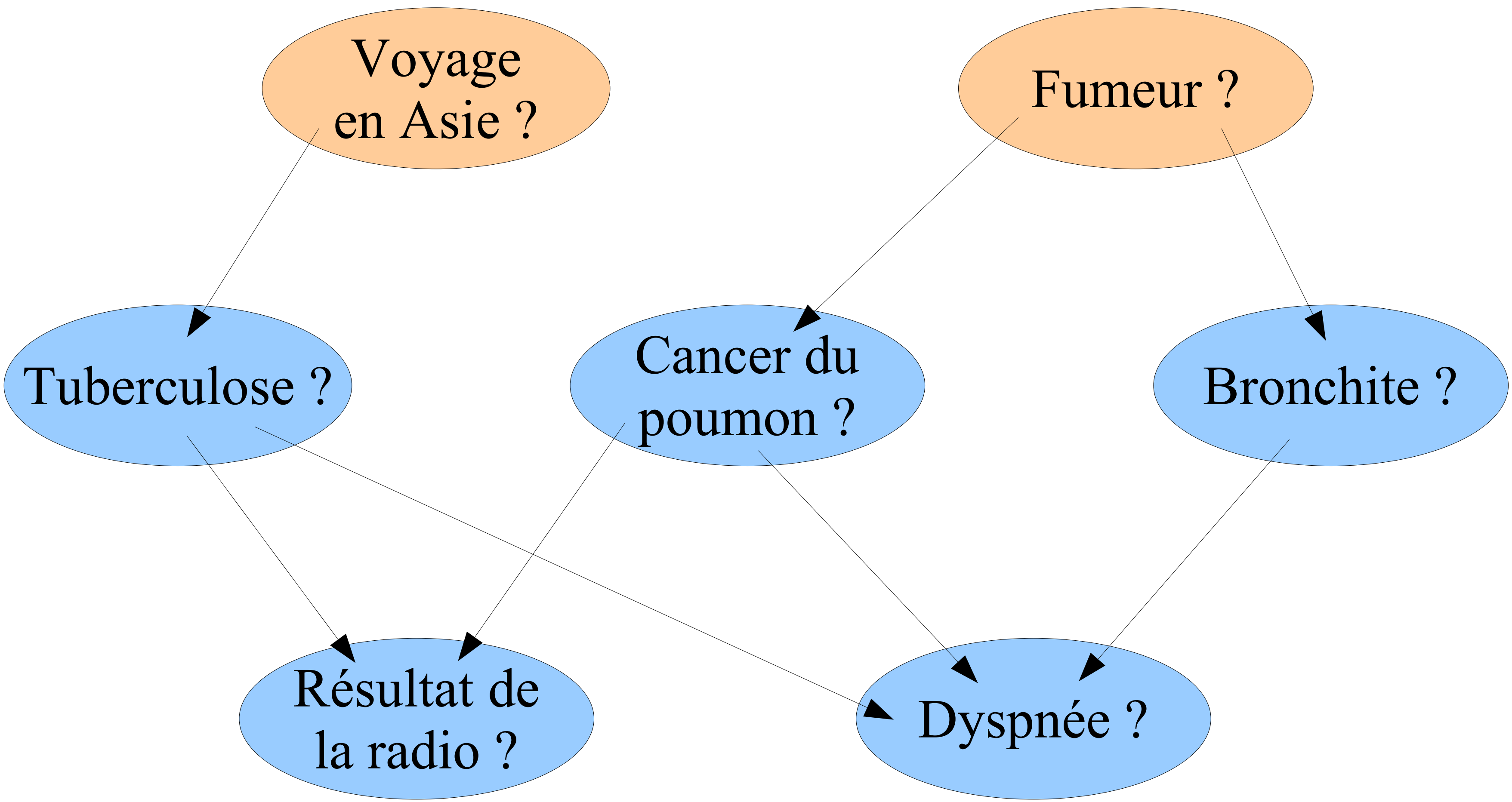
EXEMPLE CLASSIQUE

Variables binaires (simplification non obligée)



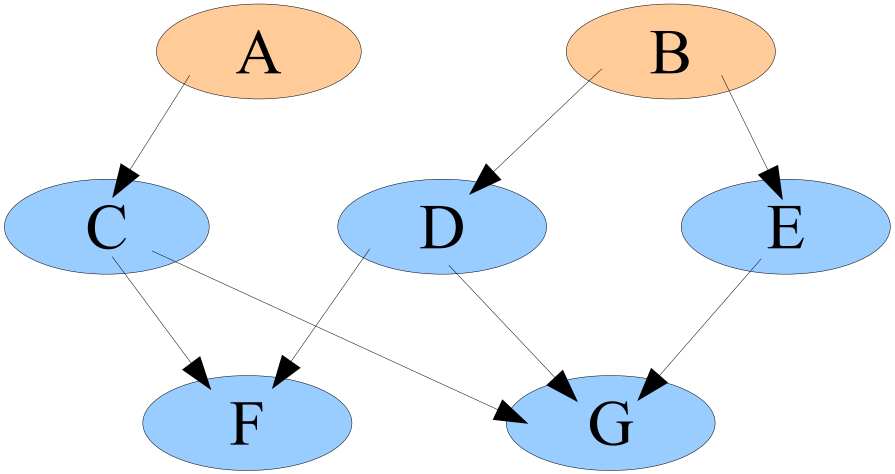
EXEMPLE CLASSIQUE

Définir les lois : (ancêtres : marginales) (autres : conditionnelles)



EXEMPLE CLASSIQUE

Le résultat est la loi conjointe qui donne rôle équivalent à chacune des 7 variables.
La causalité est perdue ! En modélisation probabiliste (en sciences ?), seule l'interprétation permet d'affirmer que V1 influence V2 et pas l'inverse.



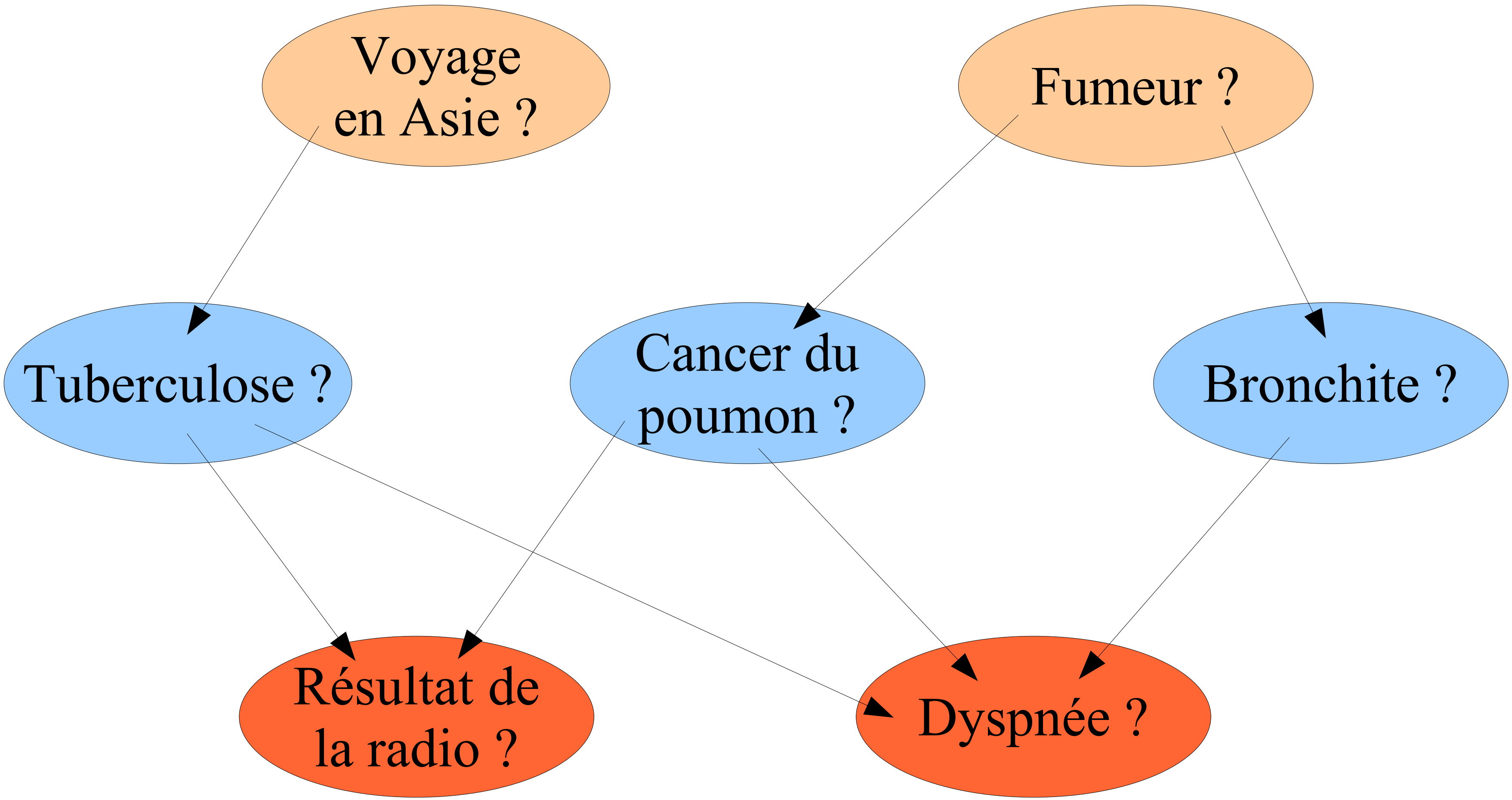
$$[A,B,C,D,E,F,G] = [A][B]\,[C|A][D|B][E|B]\,[F|C,D][G|C,D,E]$$

EXEMPLE CLASSIQUE

- On dispose d'une loi de probabilité d'un ensemble de variables. Les 7 variables de notre réseau.
- Si on s'intéresse à une variable particulière (cible), on va calculer sa loi marginale (on intègre sur tous les cas possibles des autres). On veut déterminer si le patient présente un cancer du poumon.
- Mais si certaines des autres sont connues, alors on va conditionner pour celles-là (ce n'est pas la peine de prendre en compte des cas qui ne sont pas possibles)
- Si certaines autres sont accessibles (moyennant un certain coût), on va s'interroger sur l'intérêt de les obtenir.

EXEMPLE CLASSIQUE

Le réseau est recalculé en fonction des informations nouvelles : les probabilités du diagnostic sont affinées.



RESEAUX BAYESIENS

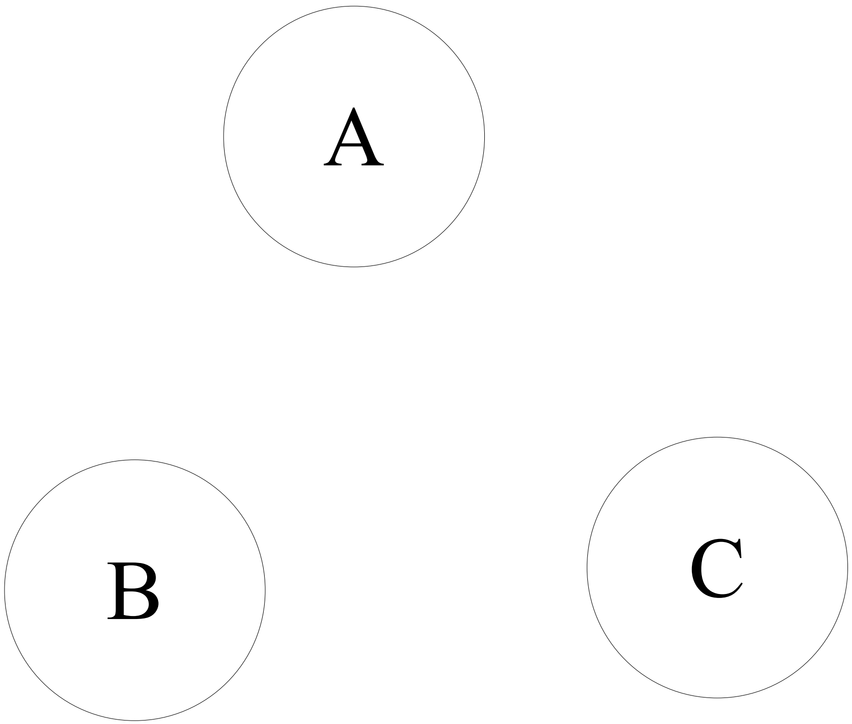
Un réseau bayésien définit la loi conjointe d'un ensemble de variables aléatoires. En général, la loi conjointe est restreinte dans une structure d'indépendance conditionnelle particulière qui est décrite par un graphe acyclique orienté construit à l'aide de distributions locales.

Aux noeuds ancêtres, on doit fournir la loi marginale ; aux autres noeuds on doit fournir la loi conditionnelle à leurs parents directs. La loi conjointe est le simple produit des lois associées à chaque noeud *(théorème de Bayes)*.

Les variables aléatoires associées aux noeuds peuvent être de tout type, en particulier discrètes ou continues.

Il peut y avoir différentes manières de représenter une même loi conjointe par un réseau bayésien.

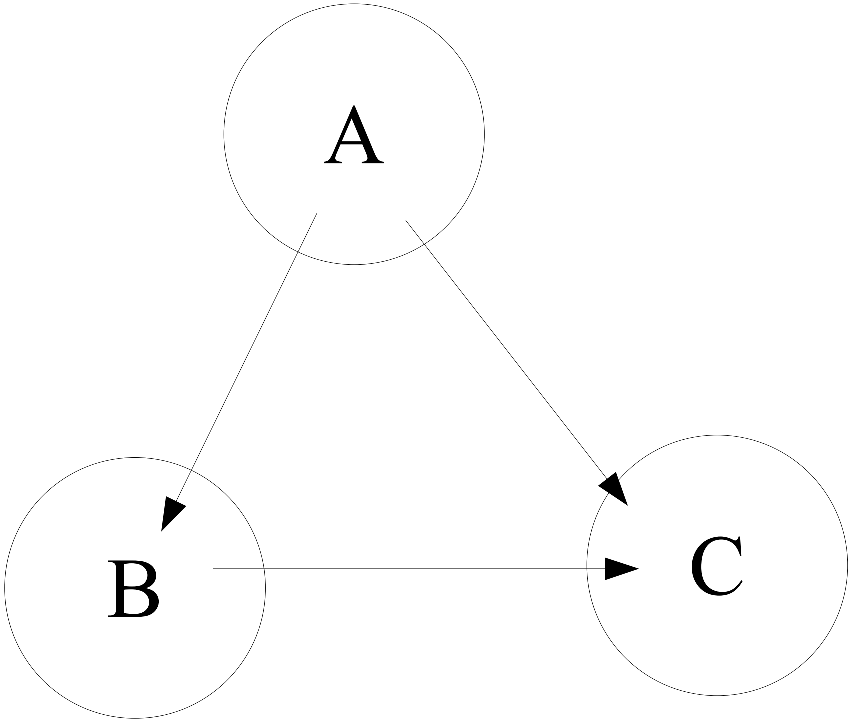
RESEAUX BAYESIENS



Que des ancêtres : indépendance

[A,B,C] = [A].[B].[C]

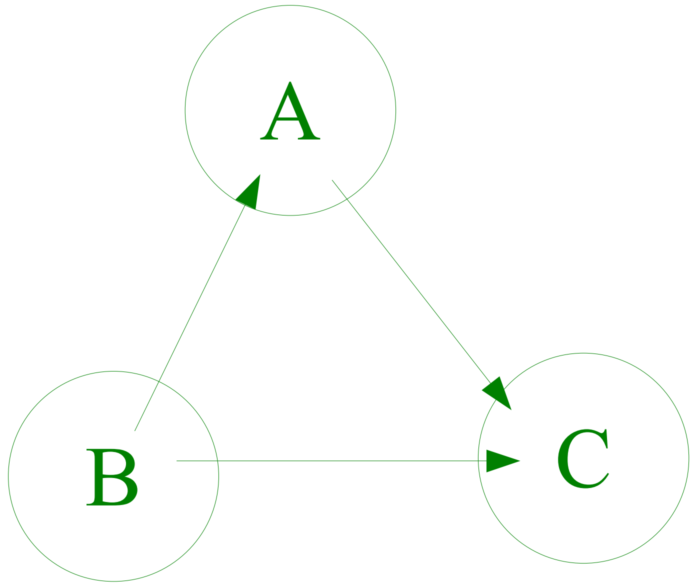
RESEAUX BAYESIENS



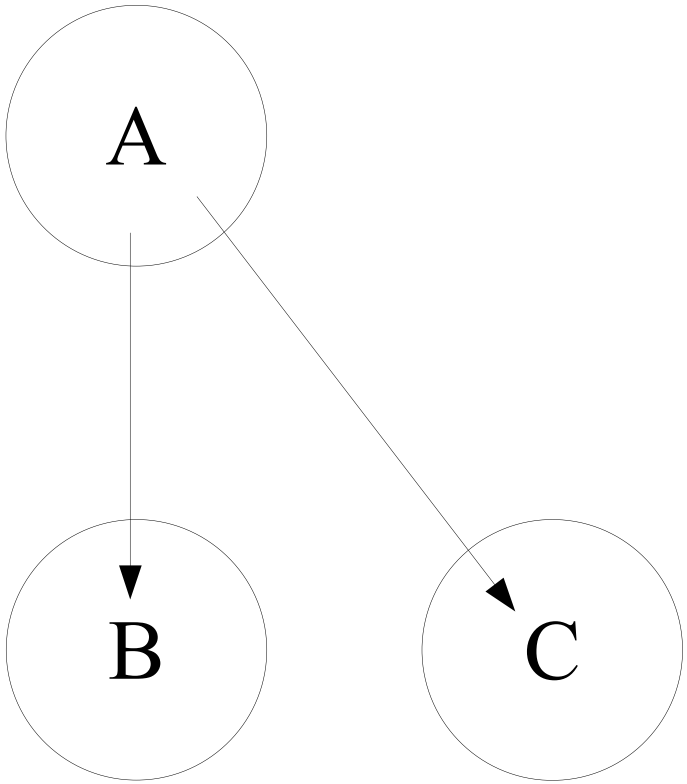
Toutes les flèches :

[A,B,C] = [A].[B|A].[C|A,B]
= [B].[A|B].[C|A,B]

Toutes conjointe peut se décomposer ainsi.



RESEAUX BAYESIENS

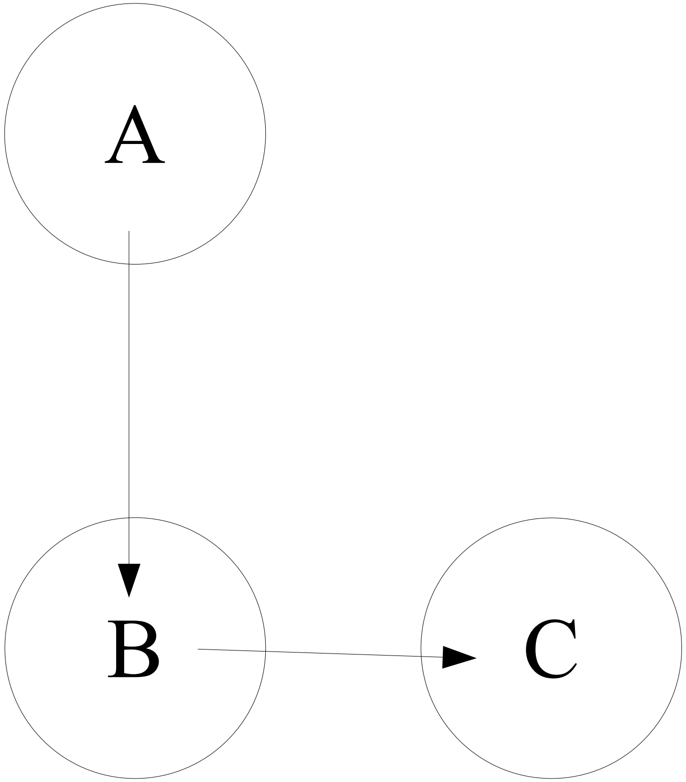


(B,C) « corrélés » ou mieux
(B|C) dépend de C

mais

((B,C) | A) indépendants !

RESEAUX BAYESIENS



(C|A) dépend de A

mais

(C |(A,B)) ne dépend pas de A !

(A|B) et (C|B) sont indépendants !

RESEAUX BAYESIENS

Causalité

Il est commode d'associer les flèches à des relations de causalité mais... (pour moi) c'est une décision subjective.

La meilleure preuve c'est qu'on peut parfois inverser des flèches en gardant la même probabilité conjointe.

idem que l'interprétation d'une corrélation statistique.

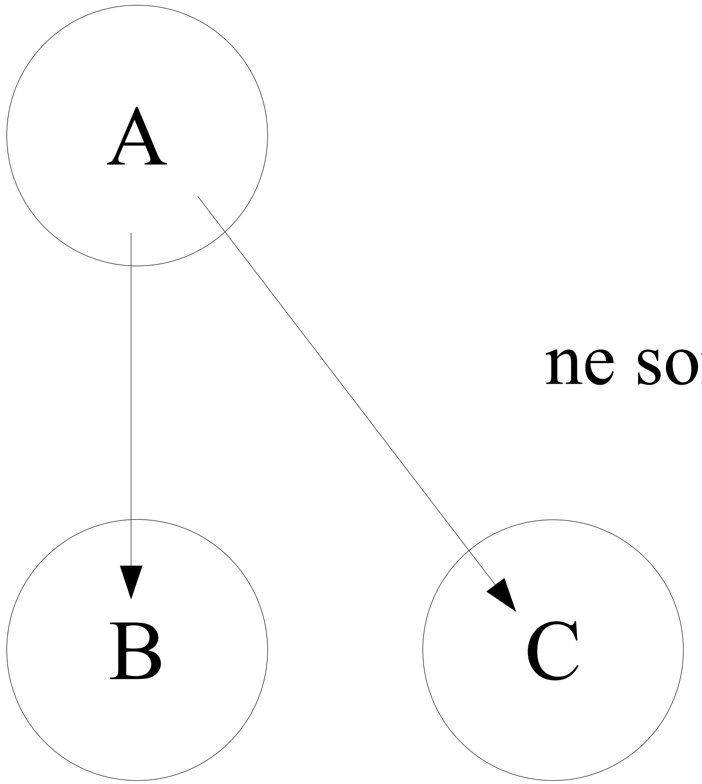
RESEAUX BAYESIENS

$$A \rightarrow B \rightarrow C$$

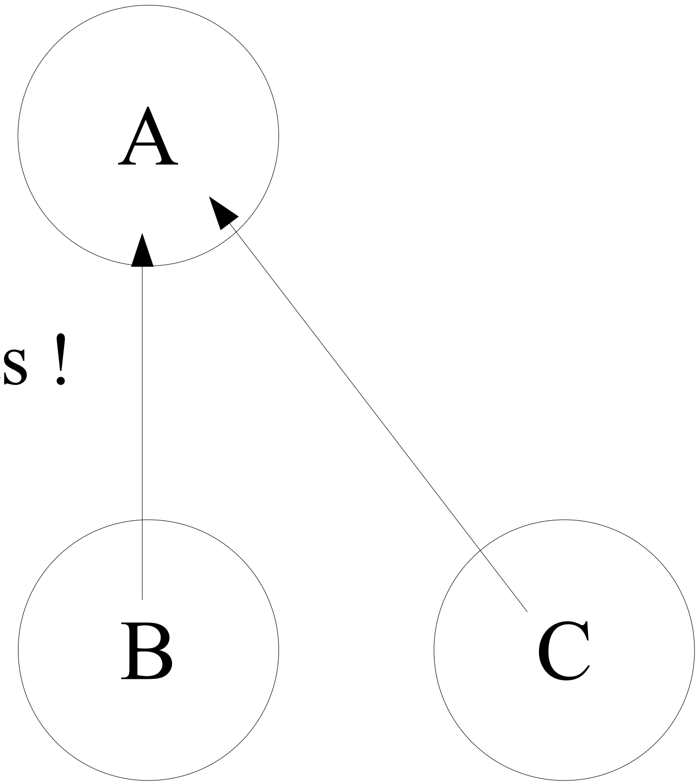
est équivalent à

$$A \leftarrow B \leftarrow C$$

RESEAUX BAYESIENS

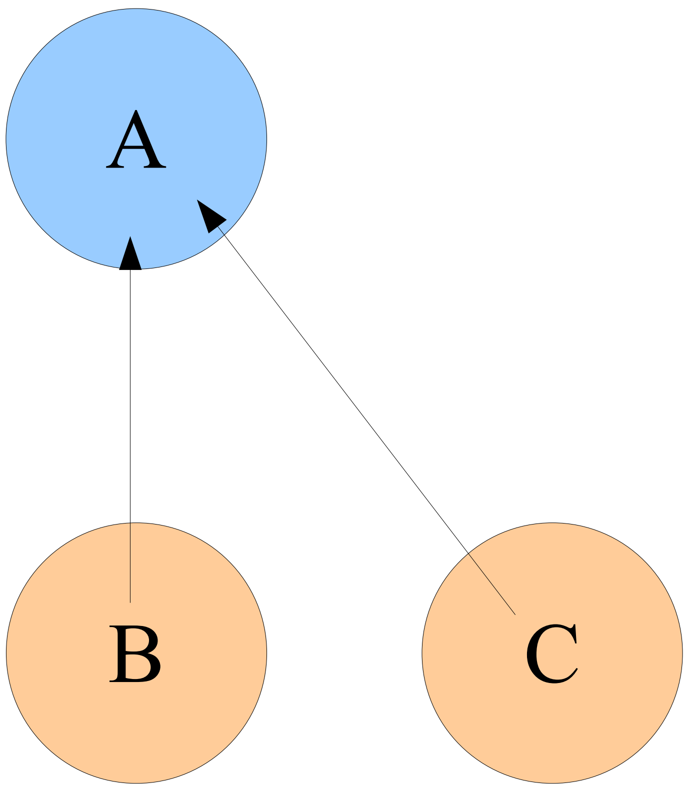


ne sont pas équivalents !



RESEAUX BAYESIENS

La plupart des exemples concernent des variables discrètes mais les **variables** peuvent aussi être **continues** !



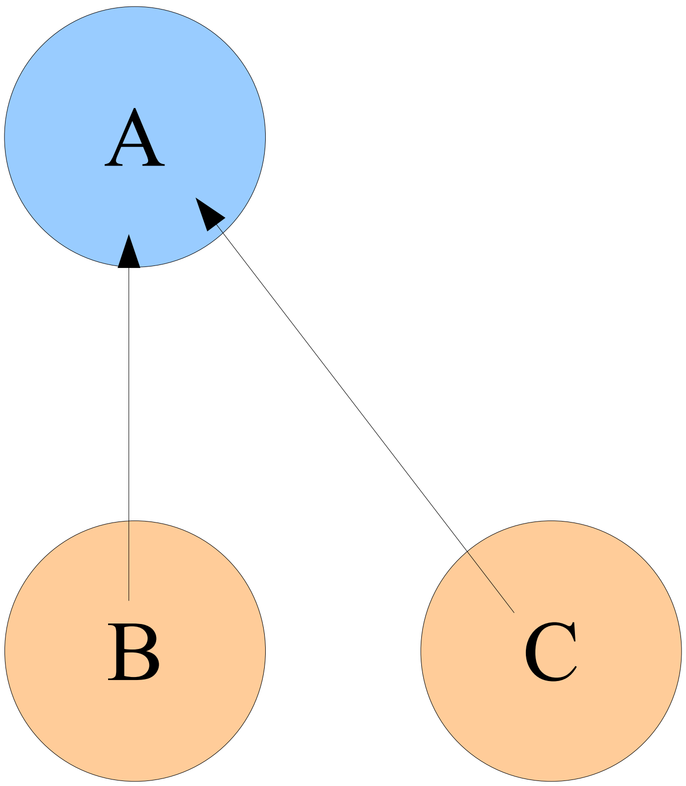
$A|(B,C) \sim N(B,C)$

$B \sim N(100,10)$

$C^2 \sim s^2\,Khi^2(5)$

RESEAUX BAYESIENS

Ou discrètes et continues...



$A|(B,C) \sim \text{Bin}(N_C, B)$

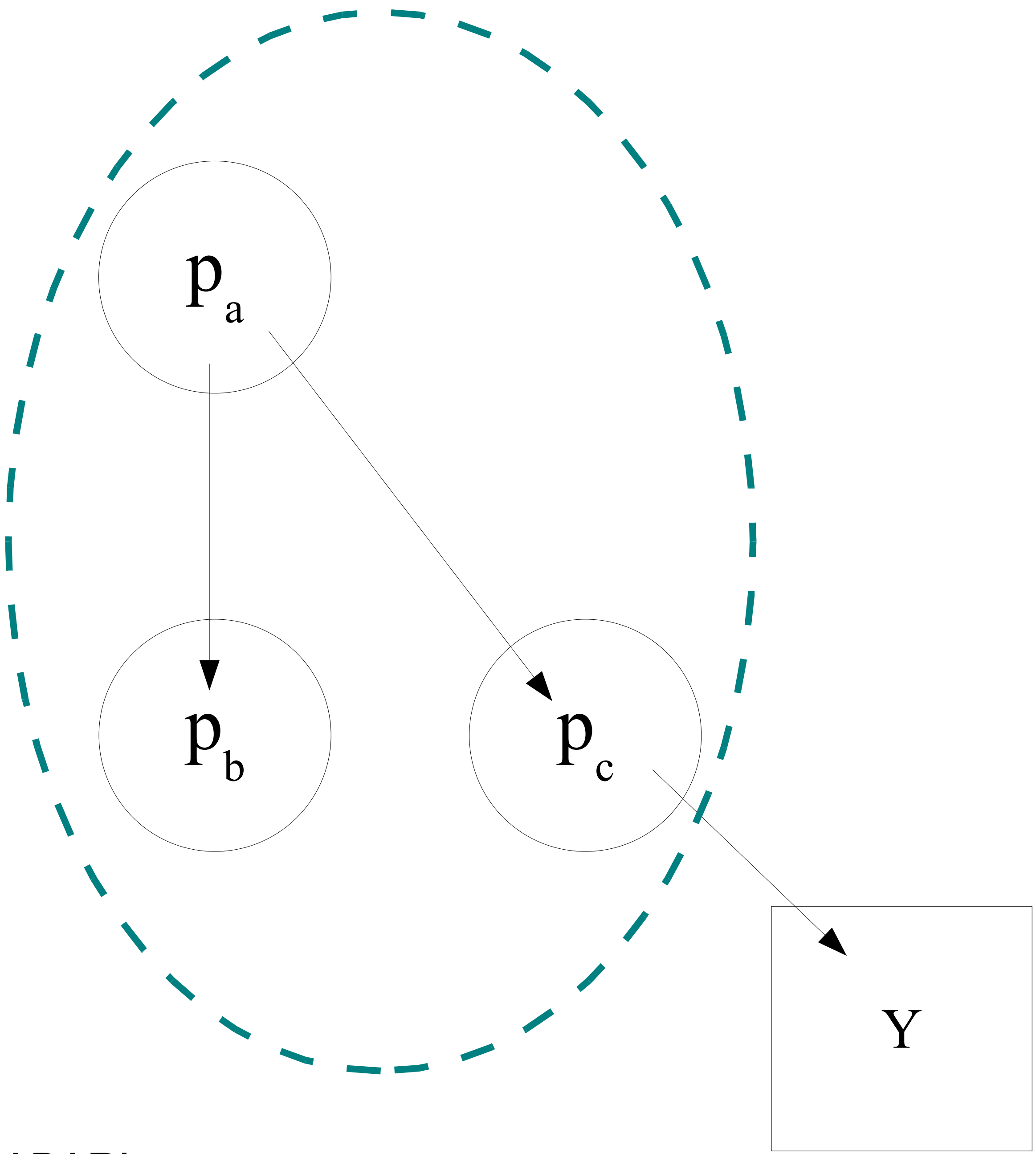
$B \sim \text{Bêta}(1,10)$

$C \sim \text{Poisson}(10)$

Réseaux B. et Statistique B.

- Le réseau bayésien définit un modèle probabiliste. Les observations ne sont pas forcément utiles.
- La statistique bayésienne permet l'estimation, l'inférence sur des modèles probabilistes à partir d'observations.
- Les deux reposent sur le **théorème de Bayes**.
- On peut faire de la statistique bayésienne sur un réseau bayésien, c'est ce que fait l'excellent logiciel **WinBUGS** ou son ersatz **Jags**.

Réseaux B. et Statistique B.



MODELE sur les paramètres auquel on raccroche les données (Y)

le modèle sans les données correspond à la priore ; le modèle conditionné par les données fournit la postériore

Réseaux B. et Statistique B.

WinBUGS

BUGS = Bayesian inference Using Gibbs Sampling : :+)

Win for *ms*Win*dows* : :+(

Langage de description de modèle similaire de R

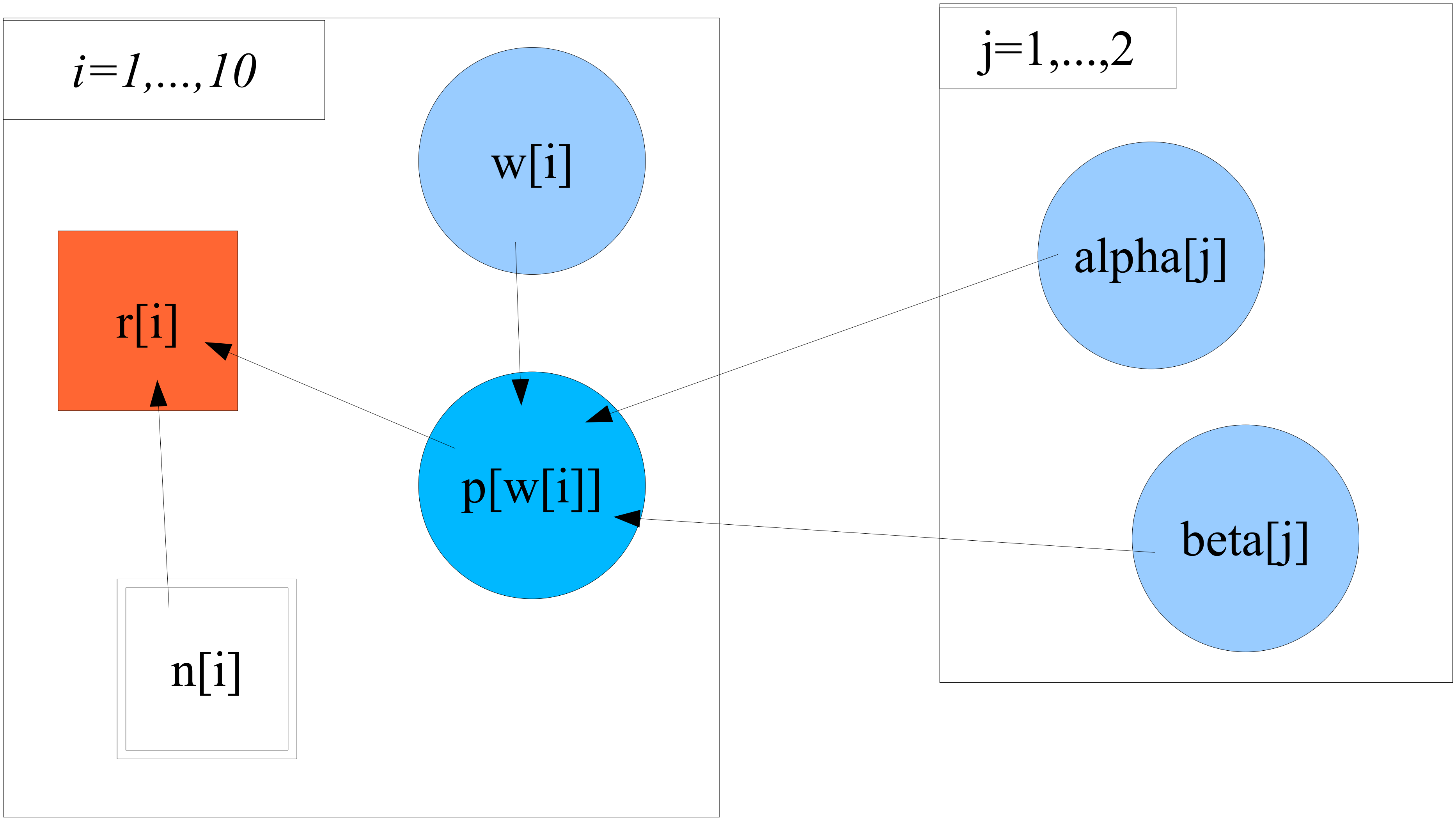
Interface graphique de construction de modèle

possibilité de lancer des scripts

Les modèles sont en fait des réseaux bayésiens et bâtis tels quels !

Réseaux B. et Statistique B.

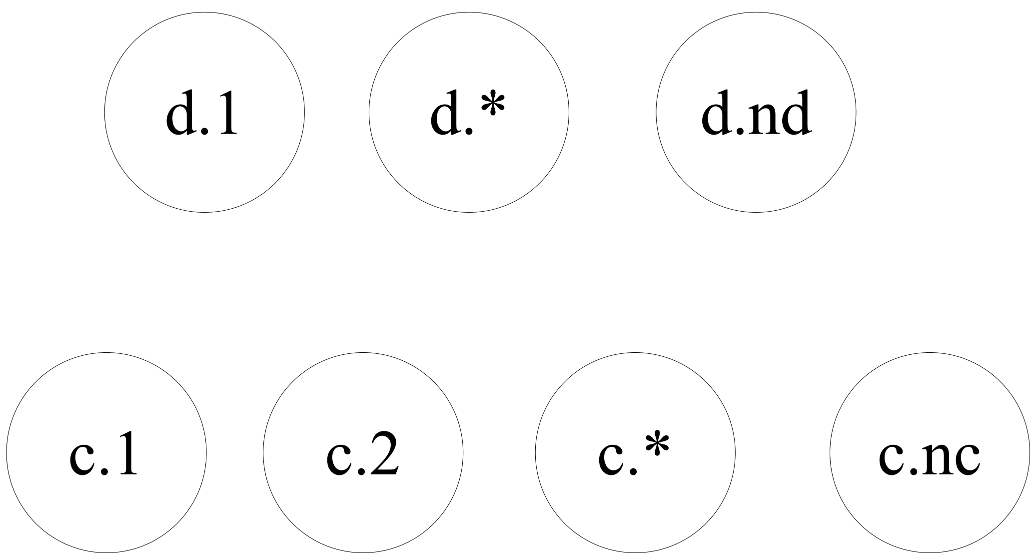
WinBUGS



R. B. et Apprentissage (DEAL)

Learning Bayesian Networks with Mixed Variables

Apprentissage de Réseaux Bayésiens à partir de Données pour des variables discrètes et/ou continues



A partir d'un tableau de données : trouver le réseau !

- **graphe**
- **proba**

Il suffit de définir les variables et leur type = colonnes du tab. de données

R. B. et Apprentissage (DEAL)

Formidable Ambition !

	c.0	c.1	c.2	c.3	c.4	c.5
d.0	1	1	3	25	5.43e+02	2.92e+04
d.1	1	2	12	200	8.68e+03	9.36e+05
d.2	3	12	144	4800	4.17e+05	8.99e+07
d.3	25	200	4800	320000	5.56e+07	2.39e+10
d.4	543	8688	417024	55603200	1.93e+10	1.66e+13
d.5	29281	936992	89951232	23986995200	1.66e+13	2.87e+16

nombre de réseaux en fonction du nombre de variables

même si bien restreinte.

R. B. et Apprentissage (DEAL)

- Les variables continues sont multinormales
- Les variables discrètes sont multinomiales
- Les parentés sont limitées à (**d -> d**), (**c->c**) et (**d->c**).
- Si (**d->c**), alors espérances et variances dépendent des combinaison de toutes les **d** !

R. B. et Apprentissage (DEAL)

Pour choisir un « meilleur » réseau bayésien, il faut définir un critère !

Le critère retenu est basé sur une approche de statistique bayésienne qui nécessite :

- la définition de priores sur les paramètres du modèle (arcs, probas déduites des arcs),
- des données pour calculer les postériores

$$p(D,d) = p(d|D).p(D)$$

où D est le réseau bayésien qui est donc probabilisé (par une priore) et d les données dont on a déjà expliqué le modèle.

Comme il y a beaucoup de calculs de postériores à mener, on prend des priores conjuguées des modélisations retenues (Dirichlet et Normal/Gamma inverse).

R. B. et Apprentissage (DEAL)

Utilisation d'un algorithme heuristique (amélioration de proche en proche) parmi l'ensemble des possibles :

- Ajoût de la relation qui maximise le score,
- Retrait de la relation qui maximise le score,
- Inversion d'une relation qui maximise le score,
- Perturbation par une action similaire effectuée au hasard.

Deux bons points :

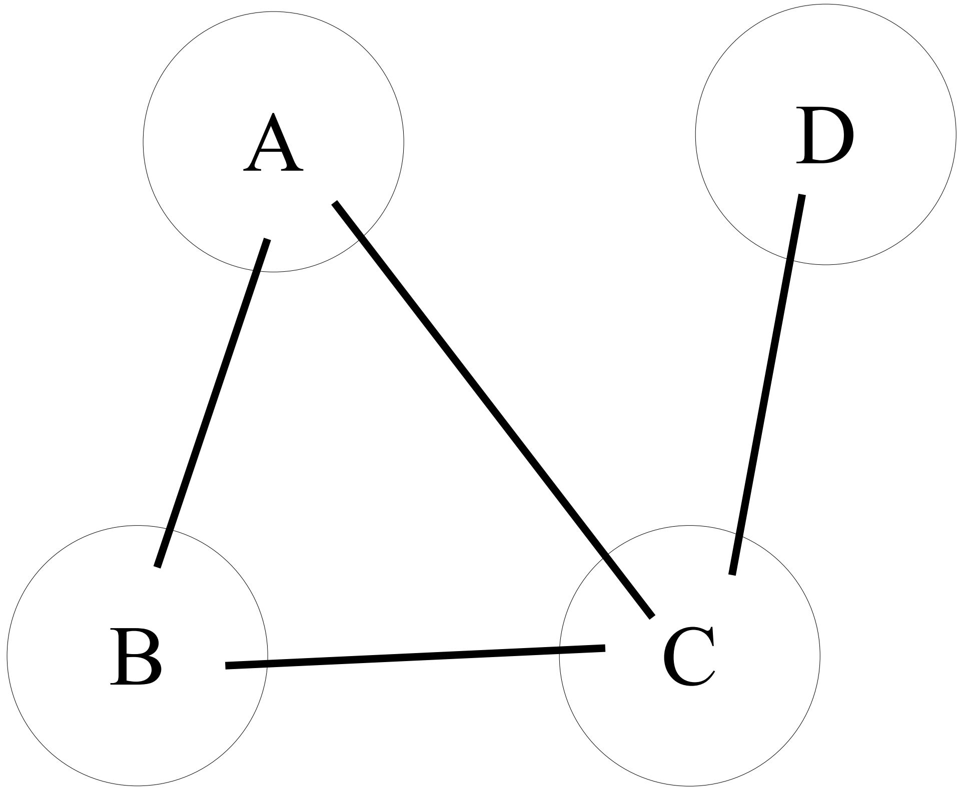
- Possibilité de choisir son point de départ,
- Possibilité d'interdire des relations (qu'on sait impossible)

Par contre :

- Il ne semble pas possible d'imposer un certain nombre de relations !

RESEAUX [Bayésiens | de Markov]

Réseaux de Markov souvent assimilés aux *modèles graphique*

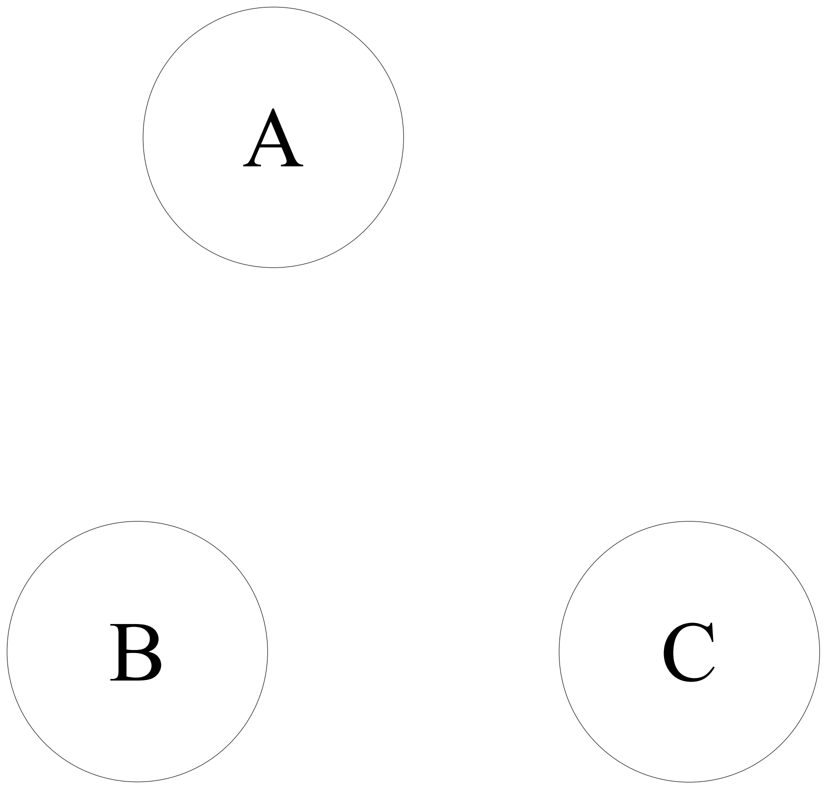


La densité conjointe s'écrit comme le produit de fonctions des « cliques » maximales :

[A,B,C,D] = f(A,B,C).g(C,D)

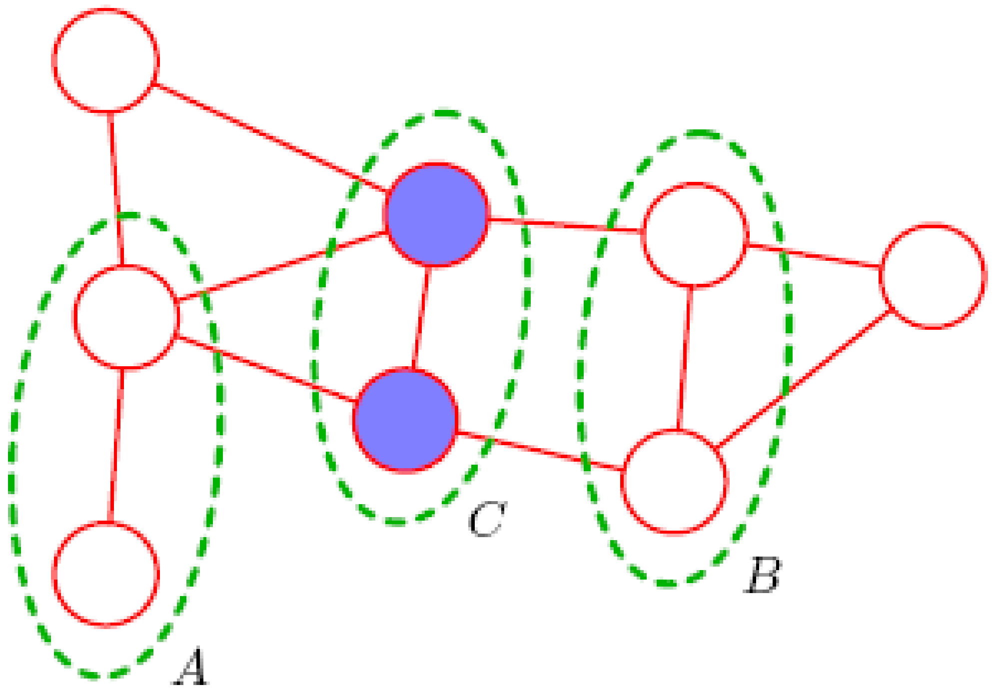
RESEAUX [Bayésiens | de Markov]

On retrouve que pas de liaison équivaut à l'indépendance...



$$[A,B,C] = f(A).g(B).h(C)$$

RESEAUX [Bayésiens | de Markov]

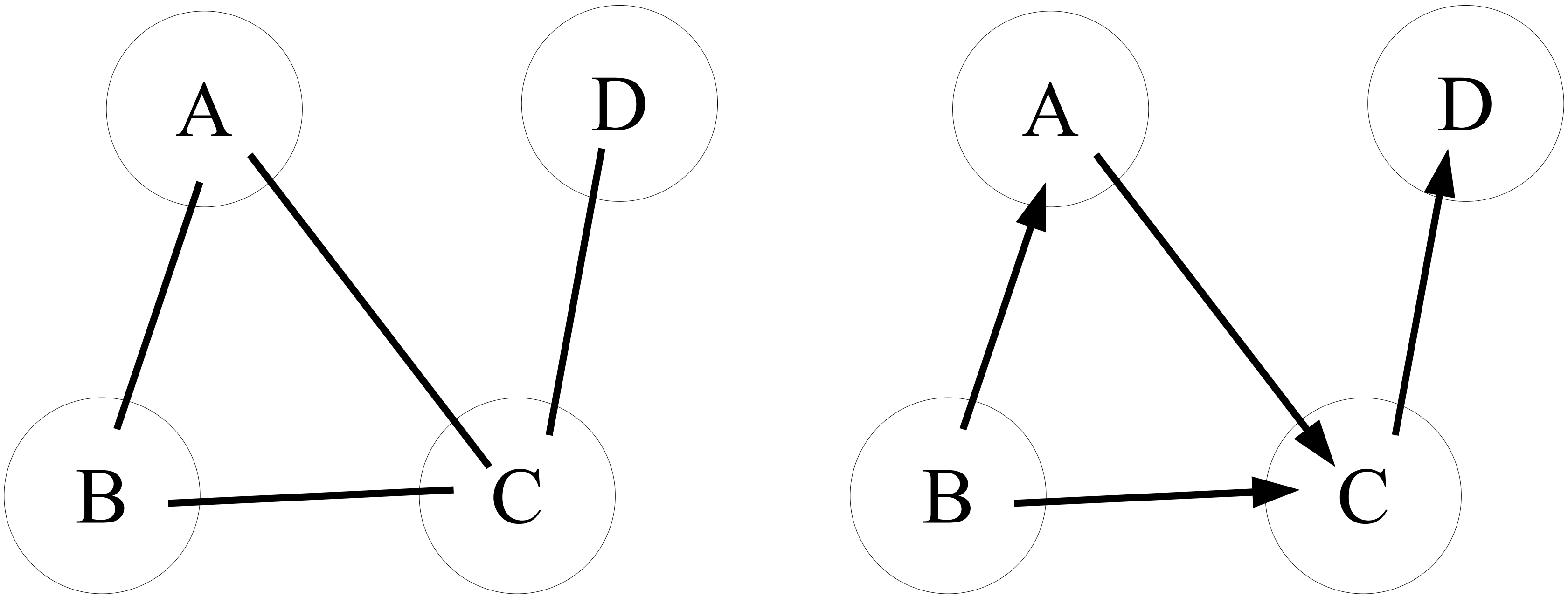


(+)

La D-séparabilité (?) est plus facile à vérifier que dans le cas des réseaux bayésiens.

{A} et {B} sont indépendants conditionnellement à {C}

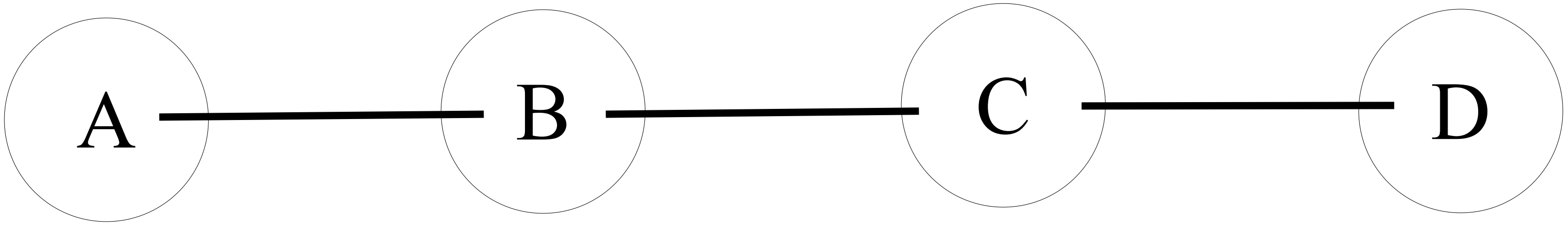
RESEAUX [Bayésiens | de Markov]



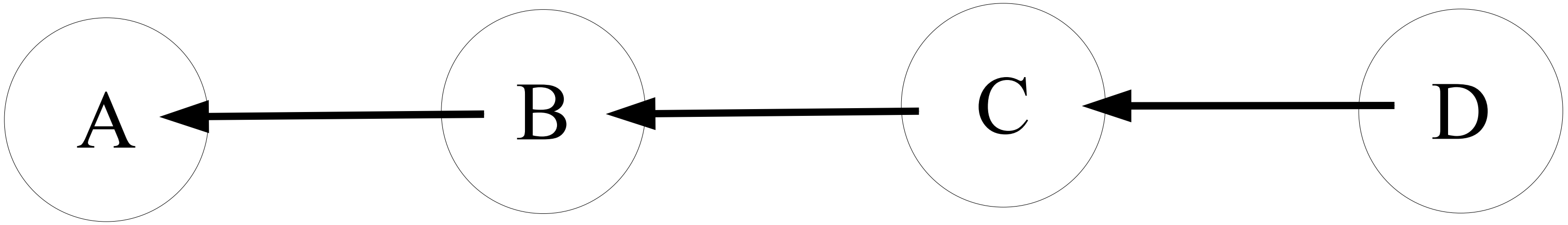
(-)

Pas d'interprétation locale similaire : plus difficile à comprendre et à construire...

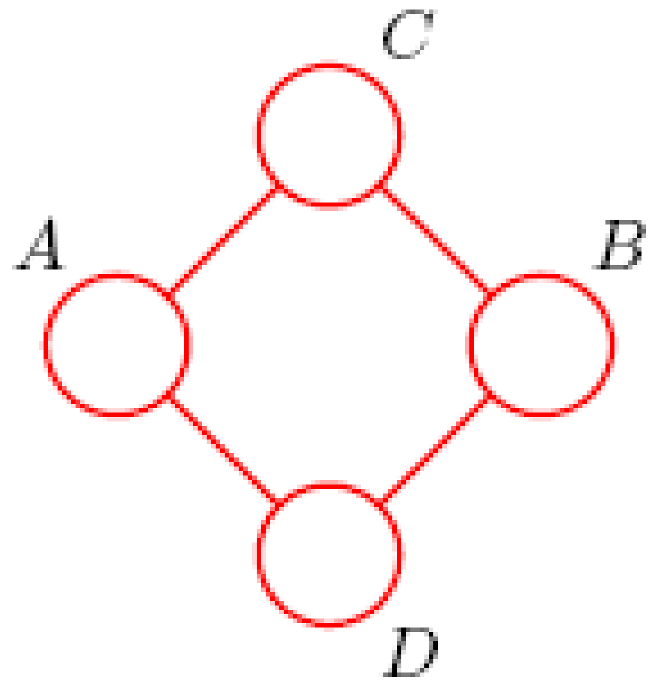
RESEAUX [Bayésiens | de Markov]



Certains réseaux sont équivalents

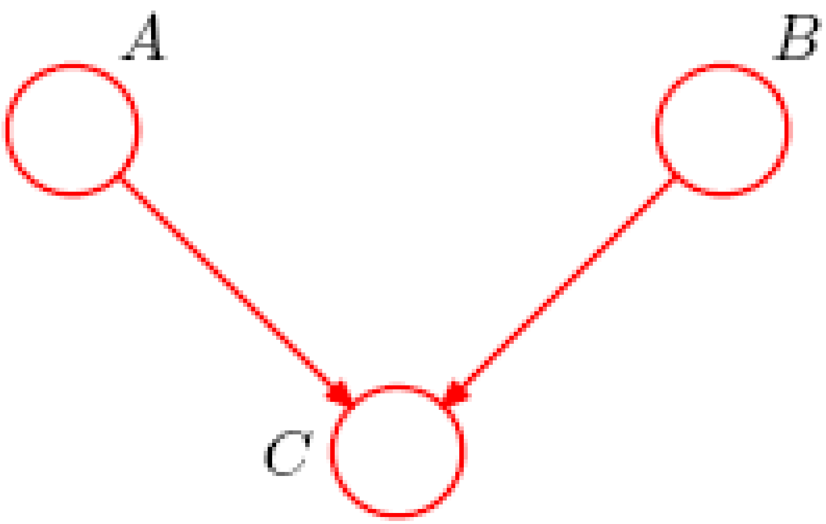


RESEAUX [Bayésiens | de Markov]



Des réseaux de Markov n'ont pas leur équivalent en réseau bayésien dans le sens où les propriétés d'indépendance conditionnelle dont ils jouissent ne peuvent pas être exprimées par un réseau de Markov sur les mêmes variables.

RESEAUX [Bayésiens | de Markov]



Mais des réseaux bayésiens n'ont pas leur équivalent en réseau de Markov dans le sens où les propriétés d'indépendance conditionnelle dont ils jouissent ne peuvent pas être exprimées par un réseau de Markov sur les mêmes variables.

Pour Conclure

Les réseaux bayésiens représentent un outil fort puissant :

- Pour réfléchir à un niveau intuitif
- Pour échanger avec des non spécialistes de la formalisation (*raisonner efficacement dans l'incertain*)
- Pour construire une modélisation complexe

Mais ils sont insuffisants (merci Judith) car limités à la structure, il faut leur ajouter les densités :

- Pour obtenir certaines simplifications
- Pour mettre en évidence certaines difficultés
- Pour aller jusqu'au bout

Pour Conclure

Ch. M. Bishop (2006) Pattern Recognition and Machine Learning, /chapter 8/, Springer

J.-M. Marin & F. Rossi (avril 2004) Découvrez les réseaux bayésiens, Linux Magazine, 60, 56-65.

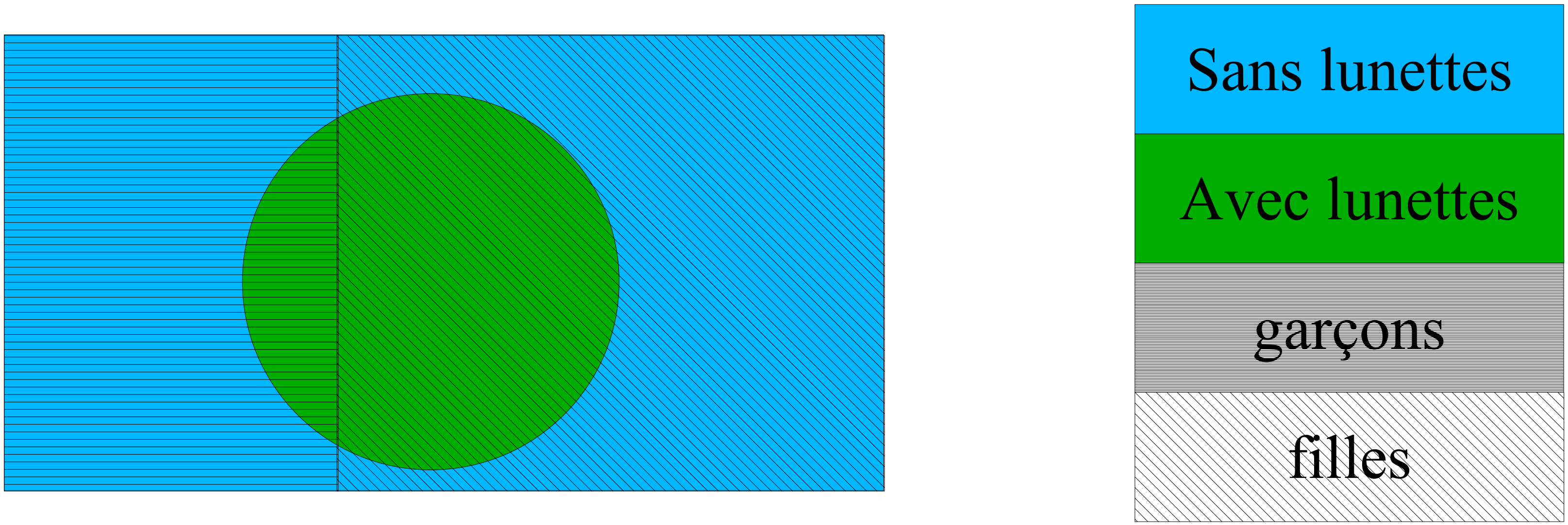
http://asi.insa-rouen.fr/~pleray/bibPL.html

Merci de votre attention !

Le Théorème de Bayes

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

- La proportion de filles parmi les enfants à lunettes est égale au nombre de filles à lunettes divisé par le nombre d'enfants à lunettes.



$$P(A|B)\cdot P(B)=P(A,B)=P(B|A)\cdot P(A)$$