Réseaux Bayésiens

Exemples 1

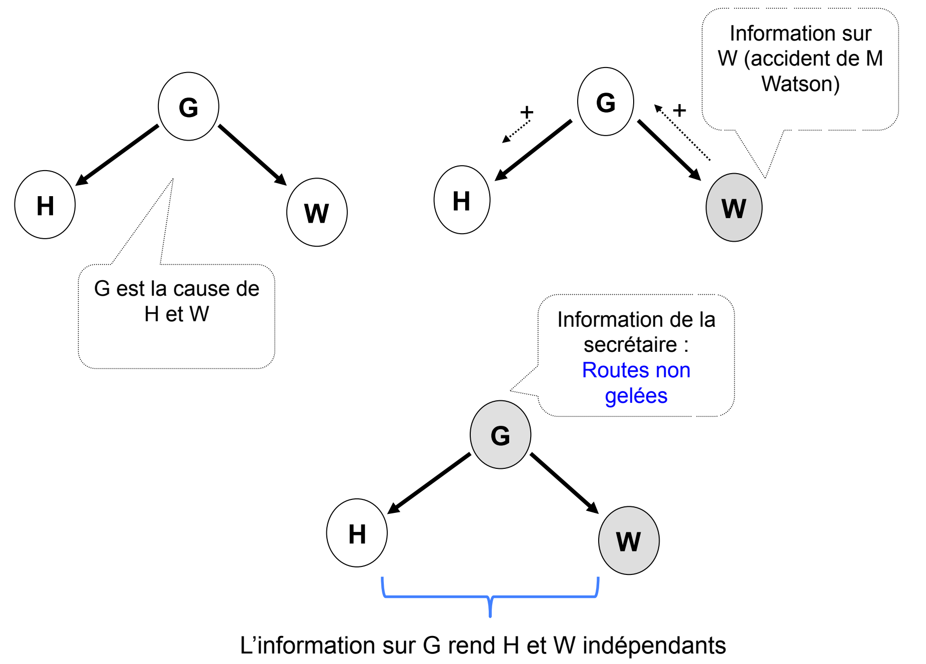
M. Smith attend l'arrivée de M. Holmz et Dr Watson. Ils sont en retard et M. Smith a un autre rendez-vous important juste après. Etant donné qu'il fait froid, M. Smith suppose alors que les routes sont gelées, il conclut alors qu'ils ont probablement eu un accident (ses deux visiteurs sont de mauvais conducteurs). Sa secrétaire entre et l'informe que Dr Watson a eu un accident, il annonce à sa secrétaire qu'il compte aller à son rendez-vous car étant donné que les routes sont gelées, il est fort probable que M. Holms a eu un accident lui aussi. Sa secrétaire l'informe qu'il se trompe et que les routes ne sont pas gelées. Tenant compte de cette nouvelle information, il décide d'attendre M. Holms encore dix minutes.

Soient les 3 variables aléatoires binaires :

G : routes Gelées

W : Dr Watson a eu un accident

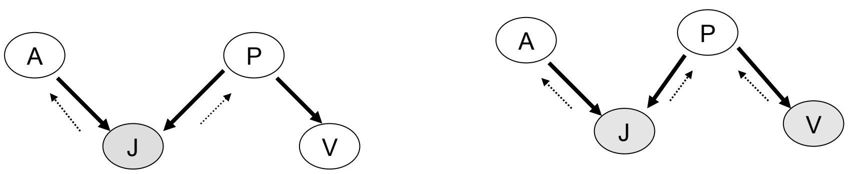
H : M. Holms a eu un accident



Exemple 2

Ce matin-là, alors que le temps est clair et sec, M. Holms sort de sa maison. Il s'aperçoit que la pelouse de son jardin est humide. Il se demande alors s'il a plu pendant la nuit, ou s'il a simplement oublié de débrancher son arroseur automatique. Il jette alors un coup d'œil à la pelouse de son voisin, et s'aperçoit qu'elle est également humide. Il en déduit alors qu'il a probablement plu, et il décide de partir au travail sans vérifier son arroseur automatique.

Soient les variables aléatoires suivantes :
A : M. Holms a oublié son arroseur automatique
P : Il a plu pendant la nuit
J : L'herbe de mon jardin est humide
V : L'herbe du jardin du voisin est humide

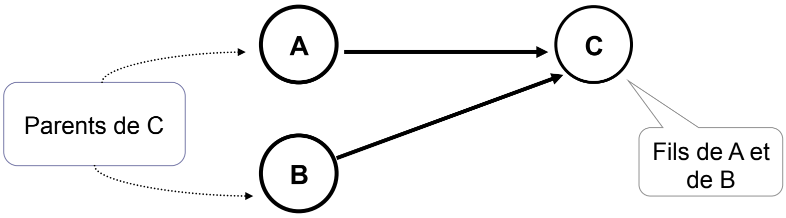


Réseaux causaux

Graphe sans circuits

Sommets : Un ensemble de variables aléatoires ayant chacune un ensemble d'états fini.

Arcs : entre les variables (traduisant une certaine causalité).



Réception d'information élémentaire

$\mathcal{E}$: ***information élémentaire*** sur un nœud A

- *Déterministe* si A prend une valeur bien déterminée a_i, elle sera dite « *instanciée* ».
- *Imprécise* si l'arrivée de l'information $\mathcal{E}$ modifie notre croyance sur A.

  Exemple d'une information élémentaire :

  « A ne peut prendre que deux valeurs a_i et a_j ».

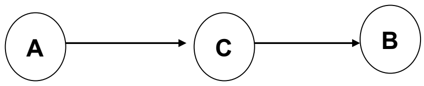
Prob(A) : probabilité de A avant l'arrivée de $\mathcal{E}$.

Prob(A/$\mathcal{E}$) : probabilité de A après l'arrivée de l'information $\mathcal{E}$.

En général $\text{Prob}(A/\mathcal{E}) \neq \text{Prob}(A)$

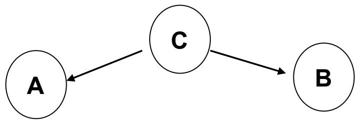
d-séparation

- Connexion en série :



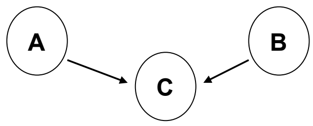
L'information circule de A à B uniquement lorsque C n'est pas instanciée.

- Connexion divergente :



L'information circule de A à B uniquement lorsque C n'est pas instanciée.

- Connexion convergente :



L'information circule de A à B lorsque soit C ou soit l'un de ses descendants a reçu une information élémentaire.

d-séparation

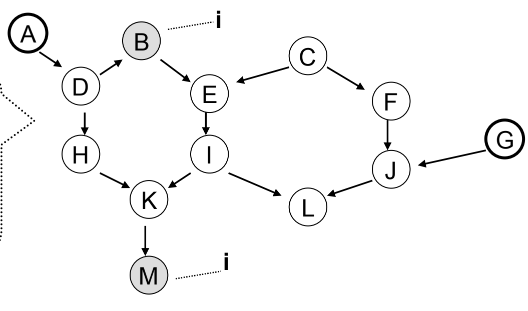
2 noeuds A et B, dans un réseau causal, sont dites d-séparés si sur toute chaîne entre A et B il existe un nœud intermédiaire V telle que l'une des deux conditions suivantes soit réalisée :

- La connexion, en V, est en série ou divergente et V est instancié.

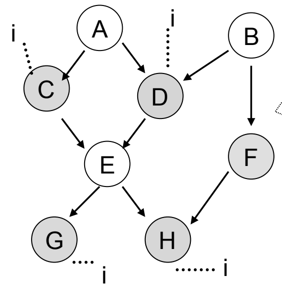
- La connexion, en V, est convergente et l'on ne dispose pas d'informations élémentaires ni sur V ni sur de ses descendants.

Exemple : Graphe causal avec B et M instanciés.

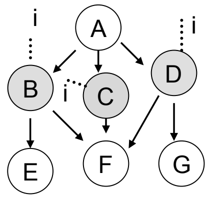
On a : A et G sont d-séparés



d-séparation : Autres exemples

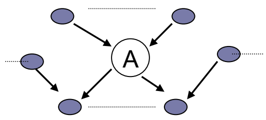


C, D, G, H , F sont instanciées. E est d-séparés de B et A



B, C, D sont instanciées. A est d-séparés de E, F et G

Markov Blanket (Couverture de Markov) d'une variable A est formée par les parents de A, des fils de A et des nœuds partageant un fils avec A.



Résultat Si tous les nœuds de la « couverture de Markov » de A sont instanciés alors A est d-séparé du reste des nœuds du graphe

Probabilités

A, B et C sont 3 variables aléatoires ayant pour ensembles d'états respectifs : $\{a_1,a_2\}$, $\{b_1 , b_2\}$ et $\{c_1, c_2, c_3\}$. On donne ci-dessous leur table de probabilité jointe P(A,B,C).

$p(a_2,b_1,c_2)$

	a_1		a_2		
	b_1	b_2	b_1	b_2	P(C)
c_1	0,01	0,04	0,1	0,278	0,428
c_2	0,004	0,03	0,108	0,152	0,294
c_3	0,006	0,01	0,062	0,2	0,278
P(A,B)	0,02	0,08	0,27	0,63	

Table de la loi marginale du couple (A,B)

Table de la loi marginale de la variable C

$$P(a_i,b_j) = P(a_i,b_j,c_1) + P(a_i,b_j,c_2) + P(a_i,b_j,c_3)$$

$$P(c_k) = \sum_{\substack{i=1,2\\ j=1,2}} P(a_i,b_j,c_k)$$

Lois marginales (suite)

$$p(A = a_i) = \sum_{k,j} p(a_i, b_j, c_k) = \sum_j p(a_i, b_j)$$

$$p(B = b_j) = \sum_{k,i} p(a_i, b_j, c_k) = \sum_i p(a_i, b_j)$$

	a_1	a_2
P(B)	0,29	0,71

	b_1	b_2
P(A)	0,1	0,9

Lois conditionnelles

$$p(A = a_i, B = b_j / C = c_k) = p(a_i, b_j, c_k) \big/ p(c_k)$$

P(A,B/C)

	a_1		a_2		
	b_1	b_2	b_1	b_2	
c_1	0,023	0,093	0,234	0,65	1
c_2	0,014	0,102	0,367	0,517	1
c_3	0,022	0,036	0,223	0,719	1

Indépendance conditionnelle

Etant donné 3 variables aléatoires A,B et C (discrètes).

A et C seront dites indépendantes « sachant » B si :

$p(A=a_i/B=b_j,C=c_k)= p(A=a_i/B=b_j)$

Remarque, cette définition entraîne aussi :

$p(C=c_k /B=b_j, A=a_i)= p(C=c_k/B=b_j)$

- **Résultat** : A et C sont indépendantes « sachant » B, si et seulement si :

  $p(A=a_i ,C=c_k/B=b_j) = p(A=a_i/B=b_j)\ p(C=c_k/B=b_j)$

Réseau Bayésien (RB)

- Un ensemble de n variables U,
- Chaque variable admet un nombre fini d'états,
- Un ensemble d'arcs entre les variables,
- Le graphe obtenu est sans circuit,
- À chaque variable X ayant $p_a(X)$ comme ensemble de parents, on suppose connu la table de probabilités conditionnelles $p(X / p_a(X))$,
- Si X n'a pas de parents, on suppose connu P(X) (probabilité à priori).

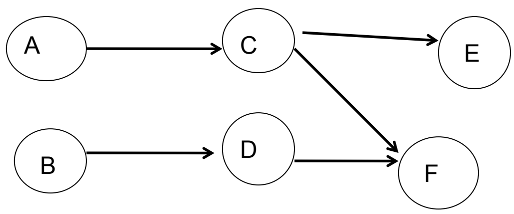
Théorème

Soit $\mathcal{R}$ un réseau bayésien sur U ={A_1,….,A_n} (l'ensemble des variables aléatoires). La probabilité jointe P(U) est définie par :

$$p(U) = \prod_{i=1,..,n} p\big(A_i / p_a(A_i)\big)$$

Où $p_a(A_i)$ représente l'ensemble des parents de A_i
(l'ensemble $p(A_i)$ est parfois vide)

Exemple d'un RB



On se donne :

Les probabilités à priori P(A), P(B)

Les probabilités conditionnelles P(C/A), P(D/B), P(E/C), P(F/C,D)

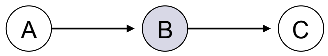
Réseaux Bayésiens et d-séparation

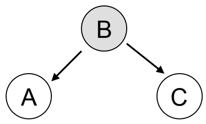
Résultat :

$\mathcal{R}$ un réseau Bayésien, e est une information sur des variables de U. Si l'information e rend les deux variables A et B (de U) d-séparées dans le graphe causal, alors elles sont indépendantes « sachant » l'information e : $\mathbf{p(A/B,}e\mathbf{) = p(A/}e\mathbf{)}$.

Remarque :

Ce résultat est facilement vérifiable pour les connexions en séries et divergentes avec e une instanciation de B.

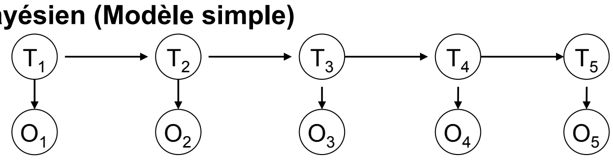




Exemple 3

Langage L sur une alphabet {a,b}. Lors de la transmission, les mots sont séparés par le symbole « c ». D'autre part, certains caractères peuvent être affectés par la transmission et être confondus avec d'autres. Un mot de 5 caractères est transmis, construire un modèle qui permet de calculer la probabilité du mot transmis $t=t_1t_2t_3t_4t_5$ sachant le mot observé est $o=o_1o_2o_3o_4o_5$.

Un réseau Bayésien (Modèle simple)



Ce modèle suppose que : $p(T_1,T_2,T_3,T_4,T_5,O_1,O_2,O_3,O_4,O_5)=$

$p(T_1)p(O_1/T_1)p(T_2/T_1)p(O_2/T_2)p(T_3/T_2)p(O_3/T_3)p(T_4/T_3)p(O_4/T_4)p(T_5/T_4)p(O_5/T_5)$

NB *Si les deux tables $p(O_i/T_i)$ et $p(T_{i+1}/T_i)$ sont constantes (indépendantes de i) ce modèle génère une séquence de 5 mots suivant un modèle d'une chaîne de Markov cachée..*

Exercices Réseaux Bayésiens

I - Modéliser par un réseau bayésien et suivre le raisonnement de l'histoire suivante :

« M. Holmes reçoit un coup de téléphone de son ami Dr Watson l'informant que la sonnerie d'alarme s'est déclanchée à sa maison de campagne. Il est alors convaincu qu'un voleur s'est introduit dans sa maison, il prend alors sa voiture et se dirige vers sa maison de campagne. Au cours de son chemin, il entend les informations à la radio qui annoncent qu'un petit tremblement de terre a eu lieu dans cette région. Or un tremblement de terre peut provoquer le déclenchement de la sonnerie d'alarme, M. Holmes décide alors retourner à son bureau ».

II- On considère la table de la probabilité jointe des 3 variables aléatoires A, B et C :

	b1	b2
a1	(0.006,0.054)	(0.048,0.432)
a2	(0.014,0.126)	(0.032,0.288)

- Calculer P(B,C) et P(B)
- A et C sont elles indépendantes sachant B ?

III- La table suivante représente un teste T pour un évènement A :

P(T/A)	A=oui	A=non
T=oui	0.99	0.001
T=non	0.01	0.999

Ce teste sera appliqué dans 2 contextes différents :

Contexte 1 : L'évènement A se réalise avec une probabilité de 0.2.

Contexte 2 : L'évènement A se réalise avec une fréquence de 0.001.

Calculer dans les deux cas la probabilité :

P(A=oui/T=oui)

Solution de I

Solution de II

Loi marginal

P(A,B,C)	b1 c1	b1 c2	b2 c1	b2 c2
a1	0,006	0,054	0,048	0,432
a2	0,014	0,126	0,032	0,288

	p(B)
b1	0,2
b2	0,8

Probabilité conditionnelle

P(B,C)	b1 c1	b1 c2	b2 c1	b2 c2
	0,02	0,18	0,08	0,72

P(C/B)	b1 c1	b1 c2	b2 c1	b2 c2
	0,1	0,9	0,1	0,9

Loi marginal

P(A,B)	b1 a1	b1 a2	b2 a1	b2 a2
	0,06	0,14	0,48	0,32

P(A/B)	b1 a1	b1 a2	b2 a1	b2 a2
	0,3	0,7	0,6	0,4

Loi marginal

Probabilité conditionnelle

Calculer P(A,C/B) et de p(A/B)P(C/B)

P(A,C/B)	b1 c1	b1 c2	b2 c1	b2 c2
a1	0,03	0,27	0,06	0,54
a2	0,07	0,63	0,04	0,36

P(A/B)P(C/B)	b1 c1	b1 c2	b2 c1	b2 c2
a1	0,03	0,27	0,06	0,54
a2	0,07	0,63	0,04	0,36

Ainsi A et C sont indépendantes « sachant B »

Solution de III

Contexte 1

P(A,T)=P(T/A)P(A)

1 : p(A)	0,2	0,8

p(A,T)	A=oui	A=non
T=oui	0,198	0,0008
T=non	0,002	0,7992

p(T)
0,1988
0,8012

P(A=o/T=o) = 0,996

P(A=oui/T=oui)= P(A=oui,T=oui)/P(T=oui)

Dans ce cas le teste est très bon

Solution de III (suite)

P(A,T)=P(T/A)P(A)

Contexte 2

2 : p(A)	0,001	0,999

p(A,T)	A=oui	A=non	p(T)
T=oui	0,00099	0,0008	0,00179
T=non	0,00001	0,998001	0,998011

P(A=o/T=o) = 0,553

P(A=oui/T=oui)= P(A=oui,T=oui)/P(T=oui)

Dans ce cas le teste est mauvais, car cette probabilité est proche de 0,5

Détermination des probabilités conditionnelles

Problème : Si A admet n modalités et si chaque variable Ai de p(A) admet n_i modalités alors la table P(A/p(A)) contient $n \times n_1 \times \ldots \times n_p$ éléments.

• Utilisation de la fonction sigmoïde logistique

Si les variables A_i sont binaires et codées en {0,1}, on associe à la variable A_i un paramètre w_i et on définit :

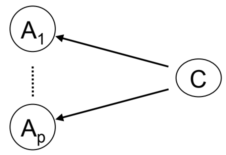
$$\mathbf{P}(\mathbf{A} = 1/\mathbf{p}(\mathbf{A})) = 1 \Big/ 1 + \mathbf{exp}-\left(\sum_{\mathbf{i}=}^{\mathbf{p}} \mathbf{w}_\mathbf{i}\mathbf{A}_\mathbf{i} + \mathbf{w}_0\right)$$

Dans ce cas il suffit de déterminer les n+1 paramètres w_i.

• <u>Naïf Bayes</u>

Dans le cas d'un problème de diagnostic, on a des symptômes et des causes. Une cause C dépend des symptômes $A_1, \ldots, A_p$, il s'agit de déterminer la probabilité de la cause C connaissant ses symptômes : $P(C/ A_1=a_1, \ldots, A_p=a_p)$.

Si on fait l'hypothèse que les symptômes sont indépendants connaissant la cause C :



$$p(a_1,\ldots,a_p/c) = \prod_{i=1..p} p(a_i/c)$$

Tenant compte de cette relation on tire que :

$$p(c/a_1,\ldots,a_p) = {p(a_1,\ldots,a_p/c)p(c)} \Big/ {p(a_1,\ldots,a_p)} = k\,p(c) \prod_{i=1\ldots p} p(a_i/c)$$

Où *k* regroupe les termes indépendants de c et constitue un terme de normalisation

Ou logique

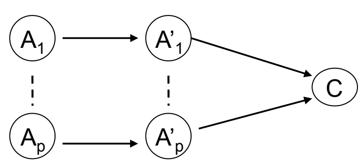
On suppose que C et tous les A_i sont des variable binaires. On considère la variable **A**$=(A_1,\ldots A_p)$. La table du **ou logique** est définie par :

$P_{ou}(C=1/ \mathbf{A_1=A_2=\ldots=A_p=0})=0$ et $P_{ou}(C=1/ \mathbf{A} \neq (0,0,\ldots,0))=1$.

Ou bien : $P_{ou}(C=0/ \mathbf{A_1=A_2=\ldots=A_p=0})=1$ et $P_{ou}(C=1/ \mathbf{A} \neq (0,0,\ldots,0))=0$.

Ou Bruité

On considère en plus la variable **A'**$=(A'_1,\ldots A'_p)$, où les variables A'_i sont binaires et 2 à 2 indépendantes, et le réseau bayésien suivant :



On suppose défini les probabilités :

$P_b(A'_i / A_i)$

Et $P_b(C/\mathbf{A'})=P_{ou}(C/\mathbf{A'})$

Ou bruité (suite 1)

Définition d'un réseau bayésien

La probabilité jointe :

$P_b(\boldsymbol{A}, \boldsymbol{A'}, C) = \quad P_{ou}(C / A'_1, \ldots, A'_p) \prod_{j=1}^{p} P_b(A'_j / A_j) P_b(A_j)$

Pour A=(a_1, …, a_j, ..., a_p)

On en déduit :

$$\begin{cases} p_b(C=0/A=a) = \prod_{j=1}^{p} P_b(A'_j = 0 / A_j = a_j) \\ p_b(C=1/A=a) = 1 - \prod_{j=1}^{p} P_b(A'_j = 0 / A_j = a_j) \end{cases}$$

Si pour tout j , on pose : $P_b(A_j'=0/A_j=1) = q_j$ et $P_b(A_j'=0/A_j=0) = 1$

On a alors :

$$\begin{cases} p_b(C=0/A=a) = \prod_{j/a_j=1} q_j \\ p_b(C=1/A=a) = 1 - \prod_{j/a_j=1} q_j \end{cases} \quad \text{(Rel 1)}$$

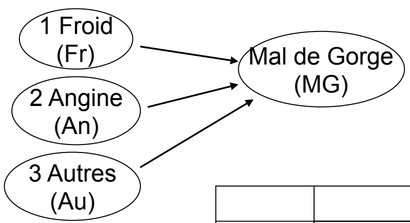
Ou bruité (suite 2)

Interprétation :

Si A_j = 1 alors C=1 sauf si un évènement ε_j se réalise et qui entraîne C=0. L'évènement ε_j se réalise avec la probabilité q_j.

On suppose en plus que les évènements q_j sont indépendants.

Exemple : Le mal de gorge du matin peut être causé par le froid, une angine ou autres causes. Ce qui correspond au graphe suivant :



On pose : q_1=0,4 , q_2=0,3 et q_3=0,1

Table de $P_{ou}(C=1/A_1,A_2,A_3)$

	A2	=1	A2	=0
	A3=1	A3=0	A3=1	A3=0
A1 =1	**$1-q_1q_2q_3$= 0,99**	**$1-q_1q_2$= 0,88**	**$1-q_1q_3$= 0,96**	**$1-q_1$=0,6**
A1=0	**$1-q_2q_3$ =0,97**	**$1-q_2$ = 0,7**	**$1-q_3$ = 0,90**	**0**

Suite exemple 3 : **Fréquences des mots et des transitions observés**

2 premières lettres \ 3 dernières lettres	aaa	aab	aba	abb	baa	bab	bba	bbb	
aa	0,017	0,021	0,019	0,019	0,045	0,068	0,045	0,068	0,302
ab	0,033	0,04	0,037	0,038	0,011	0,016	0,01	0,015	0,2
ba	0,011	0,014	0,01	0,01	0,031	0,046	0,031	0,045	0,198
bb	0,05	0,06	0,056	0,057	0,016	0,023	0,015	0,023	0,3
	0,111	0,135	0,122	0,124	0,103	0,153	0,101	0,151	

P(O/T)

	T=a	T=b
O=a	0,80	0,15
O=b	0,10	0,80
O=c	0,10	0,05

$p(T_i / T_{1-1})$

	a	b
a	0,6	0,4
b	0,4	0,6

$P(T_2 / T_1)$

	a	b
a	0,24	0,74
b	0,76	0,26

$P(T_3 / T_2)$

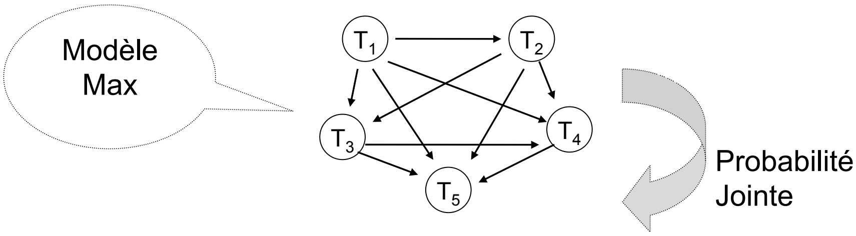
Si on retient le modèle précédent et on recalcule les probabilités des mots en utilisant la formule :

$$\mathbf{p(T_1,T_2,T_3,T_4,T_5)= p(T_1)p(T_2/T_1)p(T_3/T_2)p(T_4/T_3)p(T_5/T_4)}$$

On obtient la probabilité jointe suivante :

2 premières lettres \ 3 dernières lettres	aaa	aab	aba	abb	baa	bab	bba	bbb
aa	0,016	0,023	0,018	0,021	0,044	0,067	0,050	0,061
ab	0,030	0,044	0,033	0,041	0,011	0,015	0,012	0,014
ba	0,010	0,016	0,012	0,014	0,029	0,045	0,033	0,041
bb	0,044	0,067	0,057	0,061	0,016	0,023	0,017	0,021

Suite Exemple 3 : Apprentissage de la structure



$$P(T_1,T_2,T_3,T_4,T_5) = P(T_5 / T_1,T_2,T_3,T_4) \times P(T_4 / T_1,T_2,T_3) \times P(T_3 / T_1,T_2) \times P(T_2 / T_1) \times P(T_1)$$

Réduire ce graphe : consiste à choisir un graphe sans circuit parmi 2^{10} possibilités.

Distance entre deux lois de probabilités :

P étant la vraie probabilité et P* celle déduite d'un modèle graphique. On peur définir 2 mesures de distance :

Euclidienne $$Dist_E(P,P^*) = \sum_{\omega \in Mots} \left(P(\omega) - P^*(\omega)\right)^2$$

Kullback-leibler $$Dist_{K-L}(P,P^*) = \sum_{\omega \in Mots} p(\omega) \ln \frac{p(\omega)}{p^*(\omega)}$$

Reproduire au mieux P avec la structure du graphe la plus simple

Taille du graphe :

U **étant l'ensemble des variables aléatoires. Pour tout modèle de graphe M choisi, on défini :**

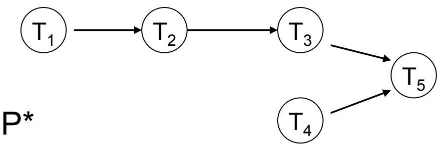
$$\text{Taille}(M) = \sum_{A \in U} card\left(\{A\} \cup P_a(A)\right)$$

P* étant la probabilité définie par le réseaux bayésien du modèle M. On défini la fonction coût suivante :

$\text{Taille}(M) + K\text{Dist}_E(P,P^*)$

K qui réalise le compromis entre les 2 termes.

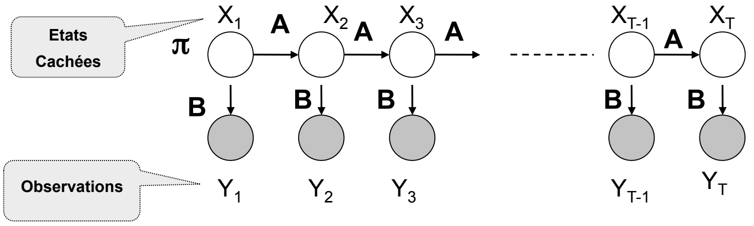
Si on prend K=10.000, le modèle M qui réalise le minimum de la fonction coût est défini par le graphe suivant :



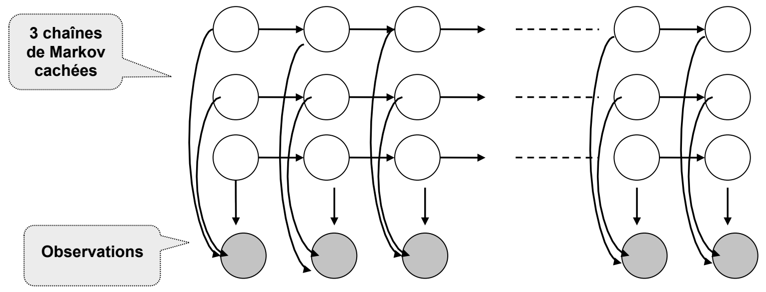
La table des probabilités P* correspondante :

				3 dernières lettres				
2 premières lettres	aaa	aab	aba	abb	baa	bab	bba	bbb
aa	0,017	0,021	0,019	0,019	0,045	0,068	0,045	0,068
ab	0,034	0,04	0,037	0,038	0,010	0,015	0,01	0,015
ba	0,011	0,014	0,01	0,01	0,031	0,045	0,030	0,045
bb	0,051	0,062	0,057	0,057	0,015	0,023	0,015	0,023

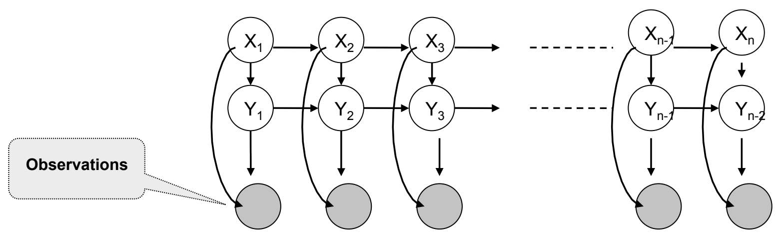
Modèle Chaîne de Markov Cachée (de taille T)



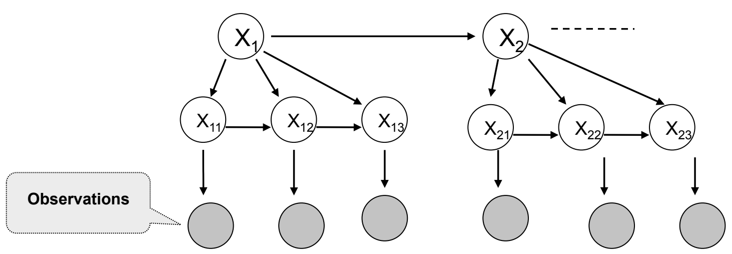
Factorial Hidden Markov Model



Switching Hidden Markov Models



Mixtures of Hidden Markov Models



Réseaux Bayésiens Dynamiques (RBD)

Un RBD est défini par un couple $(B_1, B_{\rightarrow})$ où B_1 est un réseau bayésien qui définit la propabilité à priori p(Z_1) où Z_1 désigne l'ensemble des variables de B_1.

$B_{\rightarrow}$ est formé d'une suite de deux tranches qui définissent les probabilité : p(Z_t/ p(Z_{t-1}) au sens d'un graphe sans circuit avec :

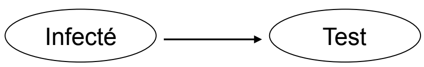
$P\left(Z_t / Z_{t-1}\right) = \prod_{i=1}^{N} P\left(Z_t^i / Pa(Z_t^i)\right)$ où Z_t^i est le ième nœud au temps t >1

$Pa(Z_t^i)$ est formée par des noeuds au temps t et des nœuds du temps t-1

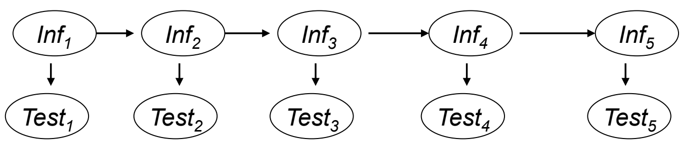
Exemples

Le lait d'une vache peut être infecté. Pour détecter si le lait est infecté, on dispose d'un test. Ce test peut donner soit un résultat *positif* soit un résultat *négatif*. Le test n'est pas parfait, il peut donner un résultat positif sur un lait non infecté ainsi qu'un résultat négatif sur un lait infecté.
Nous avons deux variables aléatoires :
Infecté : ayant deux valeurs *oui* ou *non*
Test : ayant deux valeurs *positif* ou *négatif*

Réseau Bayésien

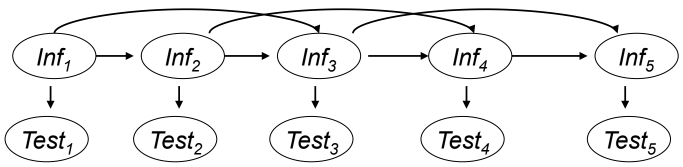


Modélisation de 5 jours avec la propriété 1-jour de mémoire de la variable « infecté »



Exemple (suite)

Modélisation de 5 jours avec la propriété 2-jours de mémoire de la variable « infecté »



Modélisation de 5 jours avec la propriété 2-jours de mémoire de la variable « infecté » et 1-jour de mémoire de la variable Test