Annexe 5

Les réseaux bayésiens

A.5.1. Rappels de probabilités

Dans un univers Ω (la plupart du temps fini) dont les éléments sont appelés «éventualités» et les parties «évèenements», une probabilité a été définie (chapitre 2) comme une mesure donnant à chaque partie (évènement) une valeur entre 0 et 1.
Les axiomes d'une telle mesure de probabilité sont, dans le cas fini, les données $p(\emptyset) = 0$, $p(\Omega) = 1$ (les évènements impossible et certain), p est croissante pour l'inclusion $A \subset B \Rightarrow p(A) \leq p(B)$, ainsi que l'additivité :

$$\forall A\ \forall B\ \ A \cap B = \emptyset \Rightarrow p(A \cup B) = p(A) + p(B)$$

Lorsqu'un évenement A est réalisé, on appelle probabilité conditionnelle (connaissant A) la mesure définie par $p_A(B) = p(B / A) = p(A \cap B) / p(A)$.
On peut vérifier qu'il s'agit d'une autre mesure de probabilité et on a la relation :

$$p(A \cap B) = p(A / B)p(B) \text{ assortie de la définition :}$$

$$\text{A et B évenements indépendants} \Leftrightarrow p(A \cap B) = p(A).p(B)$$
$$\Leftrightarrow p_A(B) = p(B) \Leftrightarrow p_B(A) = p(A)$$

Formule de Bayes, elle permet de relier une probabilité conditionnelle à la probabilité conditionnelle écrite en sens inverse :

$$p_A(B) = \frac{p_B(A)\, p(B)}{p_B(A)\, p(B) + p_{\overline{B}}(A)\, p(\overline{B})} \quad \text{et pour une partition de } \Omega \text{ en } B_1, \ldots, B_n,$$

$$\text{on a } \ p(B_k / A) = \frac{p(A / B_k)\, p(B_k)}{p(A / B_1)\, p(B_1) + p(A / B_2)\, p(B_2) + \ldots + p(A / B_n)\, p(B_n)}$$

Une variable aléatoire numérique X est une fonction de Ω dans R dont les probabilités associées aux différentes valeurs que peut prendre X, constituent la loi de probabilité.
La loi vérifie $\sum p(X = x) = 1$ pour toutes les valeurs x que peut prendre X, formule dans laquelle «X = x» désigne en fait l'ensemble des éventualités $\omega \in \Omega$ pour lesquelle $X(\omega) = x$.
Deux variables aléatoires X et Y définies sur le même univers sont dites :

X, Y indépendantes

$$\Leftrightarrow \forall x \text{ valeur de } X, \forall y \text{ valeur de } Y,\ p(X = x \text{ et } Y = y) = p(X = x).p(Y = y)$$

A.5.2. Définition d'un réseau bayésien

Un réseau bayésien est un graphe où les noeuds sont des faits (ces faits seront plus précisément associés à des variables aléatoires, par exemple au noeud associé à la variable X, les faits pourront être $X = x_1$ ou $X = x_2$, ... ou $X = x_n$, si X peut prendre n valeurs), on pourra donc confondre dans la suite les noeuds et les variables aléatoires, quant aux arcs, ce sont des relations de causalité. Sur l'arc reliant le fait A au fait B, il y aura donc une pondération qui sera la probabilité conditionnelle notée p(B / A).
Les réseaux bayésiens sont donc des graphes orientés généralement acycliques, c'est à dire sans boucles. Le problème est maintenant de décrire un algorithme de mise à jour des probabilités marginales reliant les noeuds en fonction de l'information disponible à chaque instant. En effet, lorsqu'une nouvelle information arrive en un noeud du graphe, les probabilités de chacun des états doivent être recalculés conformément à l'axiomatique des probabilités et en supposant toujours l'indépendance entre frères et soeurs dans l'arbre. C'est cette propagation de l'information que nous allons devoir détailler dans ce qui suit.

NOTATIONS

Le produit tensoriel (termes à termes) de deux vecteurs est noté $\otimes$, il est défini par :

$$(a, b, c) \otimes (a', b', c') = (aa', bb', cc').$$

On considère qu'en chaque noeud se trouve une variable aléatoire discrète et que l'état de cette variable X est décrit par le vecteur V_X des probabilités d'avoir différentes valeurs pour X. Pour le fait X = x, nous noterons p(x) sa probabilité.
On note e_X- toute l'information disponible chez les descendants de X, c'est à dire dans le sous-arbre de racine X, et e_X+, celle qui est disponible dans le reste du graphe.
Dans le cas simple d'une arborescence où chaque noeud X n'a qu'un parent U, l'état sera décrit par la matrice $M_{X/U}$ à n lignes et p colonnes dans la mesure où U possède p valeurs.
Chaque élément de cette matrice est donc de la forme $p(X = x_i / U = u_j)$

PROPAGATION DE LA CONFIANCE DANS UN RÉSEAU BAYÉSIEN

Les calculs vont se propager de telle façon qu'ils vont être traîtés localement en chaque noeud.

CALCUL A PRIORI DANS LE CAS D'UNE ARBORESCENCE

A la racine U, la données du vecteur V_U des probabilités $p(u_j)$ et la donnée de la matrice $M_{X/U}$ pour chaque fils X de U, permet de calculer le vecteur V_X des probabilités de X par simple produit matriciel $V_X = M_{X/U}.V_U$ c'est à dire pour chaque ligne $p(x_i) = \Sigma_{1 \leq j \leq p}\, p(x_i / u_j).p(u_j)$.
En chaque noeud X on peut donc stocker les valeurs $p(x_i)$. Si X et Y sont deux fils du même père U, on supposera toujours X et Y indépendantes.

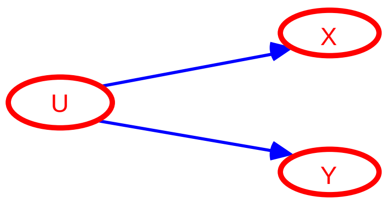

Figure A.5.1 Propagation d'un père U vers ses fils X et Y.

Ce type de calcul permet de faire descendre l'information dans toute l'arborescence.

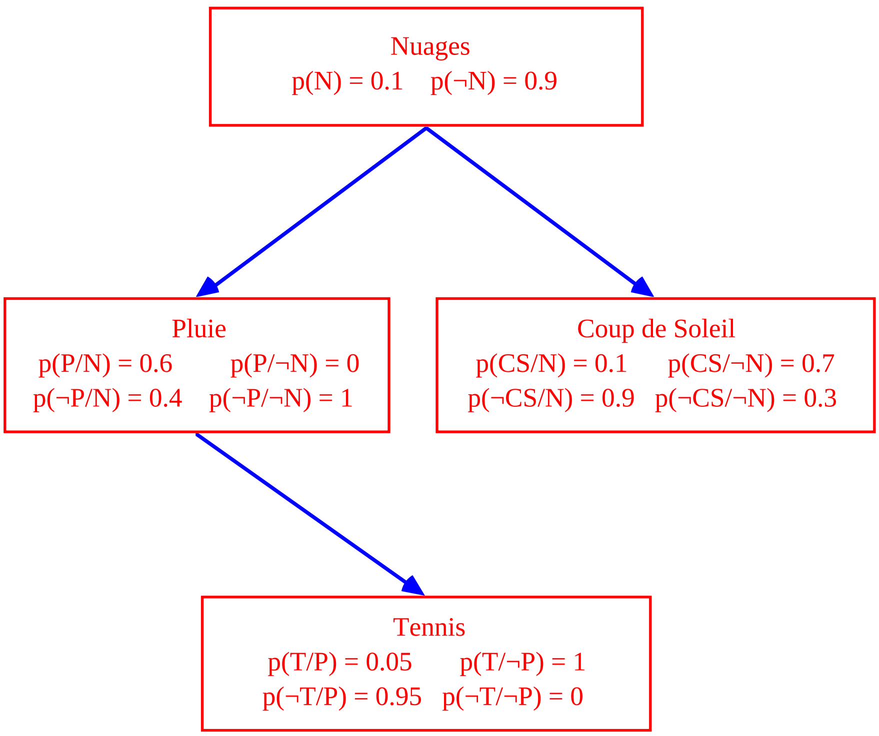

Figure A.5.2 Exemple de [Person 91], P = pluie, N = nuages, T = tennis, CS = coup de soleil.

CAS D'UN ARBRE QUELCONQUE

(X a par exemple deux parents U et V), on va toujours supposer l'indépendance de ces deux variables. Si U possède les valeurs u_1, ... u_p et V les valeurs v_1, ... v_k, alors la matrice des probabilités conditionnelles de X, notée $M_{X/U, V}$ comportera n lignes et pk colonnes, et le calcul descendant sera identique quoique le vecteur $V_{U, V}$ doit être de dimension pk (c'est le produit extérieur des messages des pères).

CALCUL A POSTERIORI

Considérons à présent le cas d'une arborescence où le père est unique et dans la chaîne U père de X, père de Y, au cas où on apprend que $Y = y_0$ pour le noeud Y, on va remonter l'information pour actualiser la loi de X puis celle de U.
La formule de Bayes donne $p(x / y_0) = \alpha p(y_0 / x).p(x)$ où α est un facteur multiplicatif tel que ait $\sum p(x / y_0) = 1$ pour toutes les valeurs x de X.

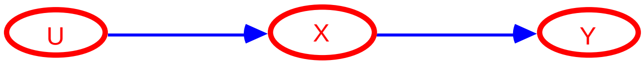

Figure A.5.3 Propagation le long d'une chaîne.

En termes vectoriels, on va définir le «message ascendant de Y vers X» qui est noté par $\Lambda_{y0\ X}$, comme le vecteur ligne formé par les $p(y_0 / x)$.
On a donc $V_{X/y0} = \alpha\Lambda_{y0\ X} \otimes V_X$.
Intéressons nous maintenant au père U de X, pour chaque valeur u possible pour U, nous avons : $p(u / y_0) = \alpha p(y_0 / u)p(u)$

$= \alpha[\sum_{x \text{ valeur de } X} p(y_0 / x, u)\ p(x / u)]\ p(u)$

$= \alpha[\sum p(y_0 / x)\ p(x / u)]\ p(u)$

car Y est indépendante de U connaissant X.

Nous avons donc plus généralement, pour une donnée de probabilités sur Y, une matrice Λ_{YX} formée par les lignes Λ_{yiX}, dite «message ascendant» contenant les probabilités conditionnelles $p(y / x)$ et qui permet le calcul $V_{X/Y} = \alpha\Lambda_{YX} \otimes V_Y$.
Le calcul précédent s'exprime alors par le produit matriciel $\Lambda_{XU} = \Lambda_{YX}.M_{X/U}$, c'est dire que le message ascendant de chaque étape ne nécessite que la matrice locale et le message ascendant qui précède.

CIRCULATIONS DESCENDANTE ET ASCENDANTE COMBINÉES DANS UNE CHAÎNE

Le message ascendant se propage jusqu'au bout de la chaîne. C'est seulement après cela que se propage le message descendant, et qu'en chaque sommet X, le vecteur V_X sera obtenu par produit tensoriel des deux messages reçus Π_X et Λ_X, puis normalisation.

CAS D'UNE ARBORESCENCE OU D'UN ARBRE QUELCONQUE

La mise à jour d'un noeud X doit se faire en consultant le père U et les fils Y_i.

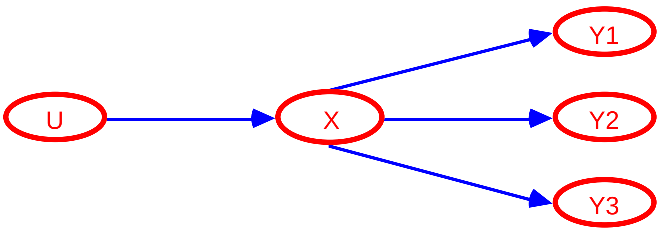

Figure A.5.4 Propagation dans une arborescence.

Posons Π_X le vecteur des probabilités p(x / e_X+) c'est à dire connaissant tout ce qui est situé au dessus de X il est calculé comme expliqué ci-dessus (même dans le cas d'un arbre quelconque) $M_{X/U}.V_U$

Posons d'autre part Λ_X celui des probabilités p(e_X- / x) de l'état de toute la descendance de X connaissant les différentes valeurs de X.

Le calcul est le même que précédemment, mais supposant les fils Y_i de X indépendants, on peut faire, grâce à cette hypothèse, le produit des probabilités pour avoir p(e_X- / x) = $\alpha\Pi$p(Y_i = y_{i0} / x).p(x) où α est un facteur multiplicatif tel que Σ_x valeur de X p(x / e_X-) = 1.

On donne donc à X le vecteur Λ_X des probabilités p(e_X- / x) calculé par le produit termes à termes des messages des fils :

$$\Lambda_X = \Lambda_{Y1X} \otimes \Lambda_{Y2X} \otimes ... \otimes\Lambda_{YiX} \otimes ...$$

On aura en définitive le vecteur d'état de X connaissant le reste du réseau :

$$[p(x / e_X+, e_X-)] = \alpha\ \Lambda_X \otimes \Pi_X$$

Donc encore une fois, le produit tensoriel des messages ascendant et descendant, suivi d'une normalisation.

REMARQUES : Si X est la racine de l'arbre, Π_X = V_X et si X est une feuille, Λ_X est le vecteur formé par des 1 uniquement. Lorsqu'une information est du type X = x_0, alors Λ_X est calculé comme le produit tensoriel de ses fils et d'un «fils fictif» formé par le vecteur nul sauf de composante 1 pour l'indice de x_0, donc, seul cet indice aura une valeur non nulle dans le résultat.

PROPAGATION DES INFORMATIONS

Un message ascendant est calculé à partir de la donnée locale et des messages ascendants antérieurs, par contre un message descendant est calculé à partir du ou des messages amont, qui eux recoivent d'autres messages ascendant. La règle de propagation des informations est donc la suivante : lorsque une information sur les probabilités marginales arrive en un ou plusieurs noeuds de l'arbre, ce sont les messages ascendants qui sont calculés et envoyés en premier. Les messages descendants sont ensuite propagés depuis la racine.

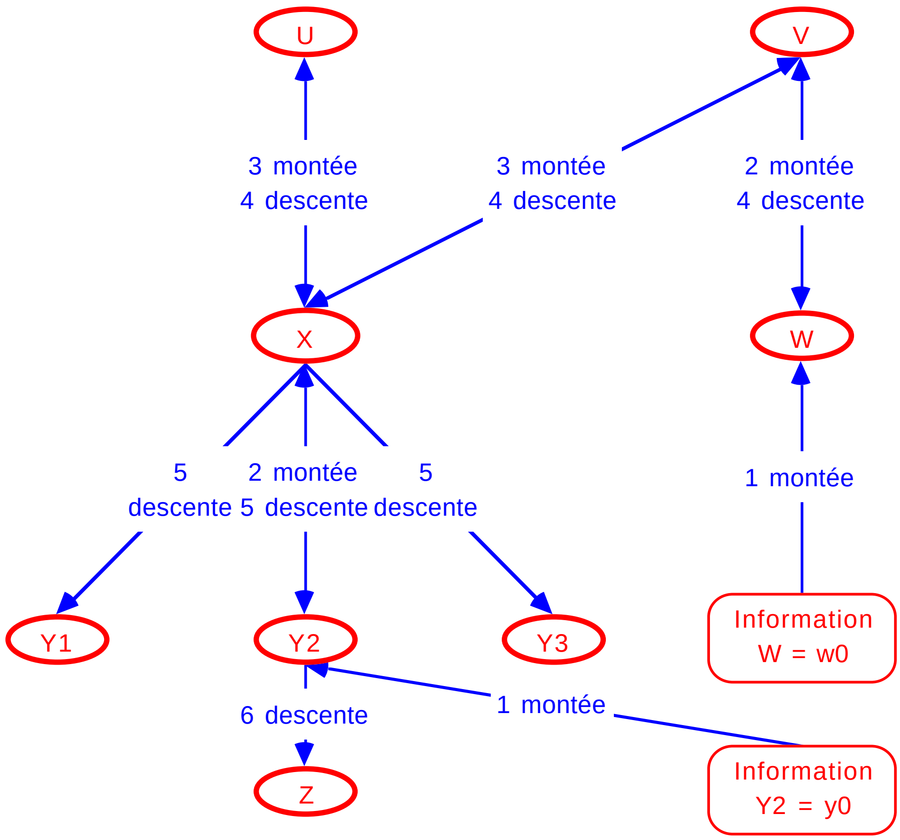

Figure A.5.5 Exemple de propagation des informations dans l'arbre suivant, deux informations portant sur Y_2 et W, vont se propager dans un ordre marqué par les étapes de 1 à 6.

EXEMPLE DE [Cooper 84], [Pearl 88]

On suppose établies les probabilités conditionnelles figurant dans l'arbre ci-dessous où la présence de tumeur au cerveau et d'un taux élevé de calcium dans le sang est causé par un cancer, ces deux faits pouvant entrainer le coma.
Dans l'hypothèse où les maux de tête sont observés, l'information remonte en T puis en C pour redescendre en Ca et en T, enfin en Co et en M.
Après circulation des messages et normalisation, les probabilités stabilisées sont p(Cancer) = 0.097, p(Calcium) = 0.094, p(Tumeur) = 0.031 et p(Coma) = 0.

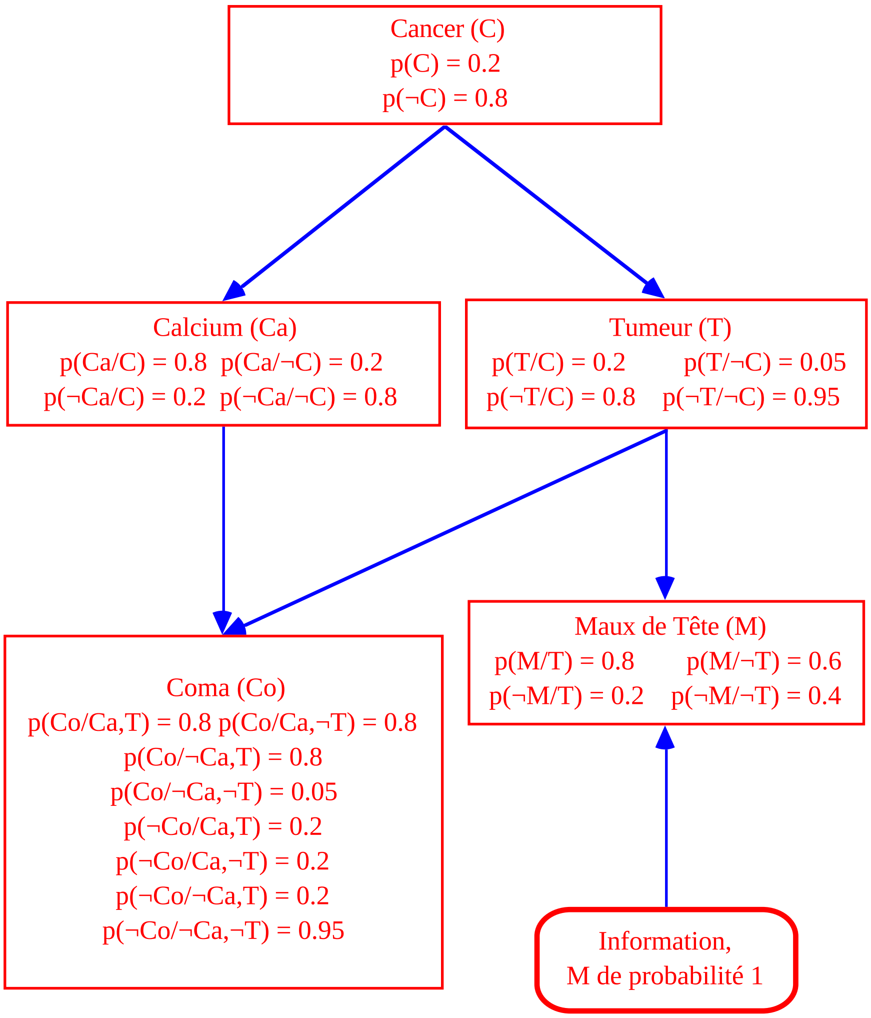

Figure A.5.6 Exemple de Cooper.

CAS DE GRAPHES AVEC CIRCUITS Dans le cas où il y a des boucles, le processus précédent peut être infini, le problème est donc bien plus compliqué, aussi a-t-il été proposé de partitionner le graphe en arbres de façon permanente (ce qui supprime le problème) ou de façon temporaire.