3-4 Fondements des réseaux Bayésiens

Les réseaux Bayésiens constituent l'un des formalismes les plus complets et les plus cohérents pour l'acquisition, la représentation et l'utilisation de la connaissance par des ordinateurs. Ils doivent leur nom aux travaux de Thomas Bayes au 18$^{\text{ème}}$ siècle sur la théorie des probabilités. Ils sont le résultat de recherches effectuées dans les années 80, dues à J. Pearl à UCLA et à une équipe de recherche Danoise à l'université de Aalborg.
Les réseaux Bayésiens sont le résultat de la convergence de deux domaines de recherche, en l'occurrence l'intelligence artificielle et les statistiques.

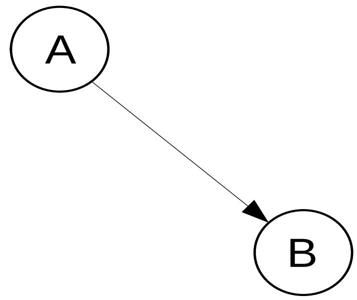
Du point de vue de l'intelligence artificielle, la construction d'un système expert nécessite la prise en compte de l'incertitude dans le raisonnement. En effet, un expert humain est capable de porter un jugement sur une situation même en l'absence de toutes les données nécessaires. Pour l'imiter, le système expert doit être capable de raisonner sur des faits et des règles incertains. Dans ce sens, les réseaux Bayésiens ainsi que d'autres méthodes permettent d'intégrer l'incertitude dans le raisonnement. Les réseaux Bayésiens se distinguent par son aspect quantitatif.

Du point de vue statistique, l'utilisation de réseaux Bayésiens permet d'extraire, à partir des tableaux de mesures, une représentation graphique compacte, sans perte d'information. Le statisticien peut l'utiliser pour raisonner, effectuer des interprétations et tirer des conclusions plus facilement.

3-4-1 Représentation graphique de la causalité

3-4-1-1 Définition

Intuitivement, la figure suivante permet de schématiser l'influence d'un événement, d'un fait, ou d'une variable (A) sur un événement, un fait, ou une variable (B) respectivement en les reliant par une flèche orientée de A vers B.



Supposons que A et B soient des faits, qui peuvent être vrais, ou faux. Logiquement, ce graphe peut se lire comme "la connaissance que je possède de A détermine la connaissance que j'ai de B".

Deux interprétations peuvent être distinguées :

- Détermination stricte : sachant avec certitude que A est vrai, je peux en déduire B avec certitude.
- Simple influence : si je connais A avec certitude, *mon opinion* sur B est modifiée, sans que je puisse toutefois affirmer si B est vrai ou faux.

Il est à noter que la flèche peut fonctionner dans les deux sens même si elle est orientée de A vers B. En effet, une explication détaillée dans **[32]** nous permet de conserver à l'esprit que l'information ne circule pas seulement dans le sens des flèches :

Supposons par exemple que la relation causale soit A⇨B. Cela signifie que si A est vrai, B l'est également. Si A est faux, B peut être vrai ou faux.

A	**B**
V	V
F	V
F	F

La table de vérité ci-dessus nous permet d'affirmer que si B est faux, A l'est également.

Du point de vue de la logique, il s'agit simplement de la contraposée de A⇨B. Du point de vue de la causalité, cela montre qu'une relation causale, même orientée, est inversible, de l'effet vers la cause, même si elle ne l'est que partiellement.

Ainsi :

S'il existe une relation causale de A vers B, toute information sur A peut modifier la connaissance que j'ai de B, et, réciproquement, toute information sur B peut modifier la connaissance que j'ai de A. **[32]**

3-4-1-2 Circulation de l'information dans un graphe causal

À l'aide d'un exemple extrêmement classique dans la littérature sur les réseaux Bayésiens extrait de **[33]**, et repris dans **[34]**, nous allons présenter la circulation de l'information dans un graphe causal :

" Ce matin-là, alors que le temps est clair et sec, M. Holmes sort de sa maison, et 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'oeil à la pelouse de son voisin, M. Watson, et s'aperçoit qu'elle est également humide. Il en déduit alors qu'il a probablement plu, et décide de partir au travail sans vérifier son arroseur automatique ".

Soient A, P, J et V des faits tels que :

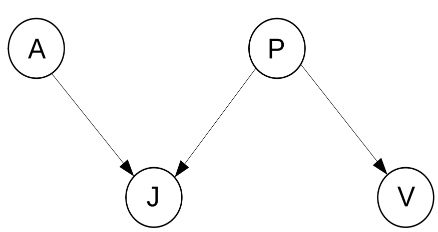
A : j'ai oublié de débrancher mon arroseur automatique

P : il a plu pendant cette nuit

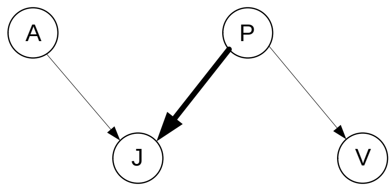
J : l'herbe de mon jardin est humide

V : l'herbe du jardin de mon voisin est humide

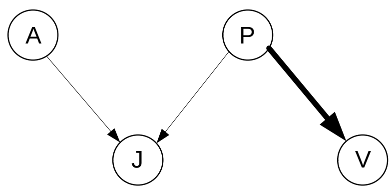
La représentation graphique du modèle causal utilisé par M. Holmes est la suivante :



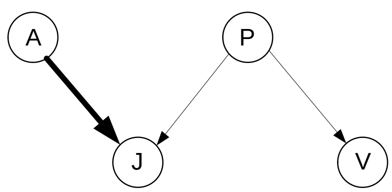
Le raisonnement de M. Holmes se résume dans l'explication suivante des causalités :



S’il a plu pendant la nuit, l’herbe de mon jardin est humide.



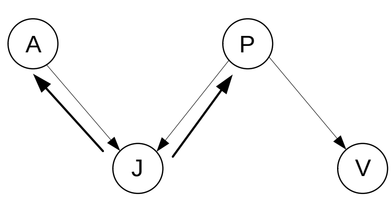
S’il a plu pendant la nuit, l’herbe du jardin de mon voisin est également humide.



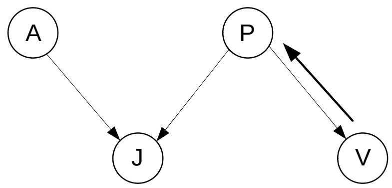
Si j’ai oublié de débrancher mon arroseur automatique, l’herbe de mon jardin est humide.

Étant donné que l’information J est vraie, le modèle nous indique que J a du être causé soit par A, soit par P. Les deux causes sont a priori plausibles (probables) par manque d’information complémentaire.

La véracité de V renforce la croyance en P, quoique le voisin pourrait lui aussi avoir oublié de débrancher son arroseur automatique. Dans cette première analyse, seulement le sens de circulation *Effet* → *Cause* a été considéré.



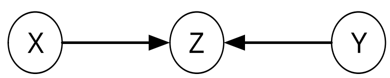
La connaissance de *J* renforce la croyance dans l'une des deux causes *A* ou *P*



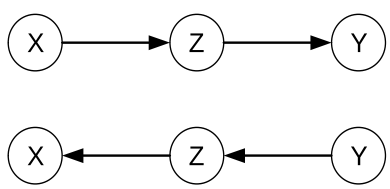
La connaissance de *V* augmente la croyance dans la cause P. La cause *A* devient moins plausible

M. Holmes a donc déduit que son arroseur automatique était à l'arrêt à partir du fait que la pelouse de son voisin était humide.

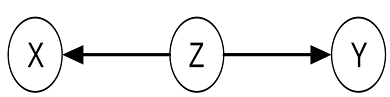
En général, pour expliquer la circulation de l'information dans un graphe causal, Patrick Naim dans **[32]** a considéré les 3 cas suivants :



Connexion convergente : X et Y causent Z.



Connexion en série : X cause Z, Z cause Y (ou le cas symétrique).



Connexion divergente : Z cause X et Y.

Dans le tableau suivant **[32]**, l'auteur a expliqué la circulation de l'information de X à Y et a donné des exemples pour chacun des cas :

Graphe	Propriété	Exemple
$X \rightarrow Z \leftarrow Y$	L'information ne peut circuler de X à Y que si Z est connu.	X = tremblement de terre Y = voleur Z = alarme Le fait qu'il y ait eu un tremblement de terre dans le voisinage n'a aucun lien a priori avec que le fait que ma maison ait été cambriolée. En revanche, si mon alarme s'est déclenchée, j'ai tendance à croire que je viens d'être cambriolé. Si maintenant j'apprends qu'il vient d'y avoir un tremblement de terre dans le voisinage, je suis rassuré sur l'éventualité d'un cambriolage.
$X \rightarrow Z \rightarrow Y$	L'information ne peut circuler de X à Y que si Z n'est pas connu.	X = ensoleillement Y = récolte Z = prix du blé. Si la saison a été ensoleillée, la récolte sera abondante. Si la récolte est abondante, le prix du blé est bas. Si je sais déjà que la récolte a été abondante, le fait de connaître l'ensoleillement ne m'apprend plus rien sur le prix du blé.
$X \leftarrow Z \rightarrow Y$	L'information ne peut circuler de X à Y que si Z n'est pas connu.	X = la pelouse de mon jardin est humide Y = la pelouse de mon voisin est humide Z = il a plu cette nuit Si la pelouse de mon jardin est humide (X), j'ai tendance à croire qu'il a plu cette nuit (Z), et donc que la pelouse de mon voisin sera aussi humide (Y). Si en revanche je sais qu'il a plu cette nuit (Z), je peux affirmer que la pelouse du jardin de mon voisin sera humide (Y), et l'information que je peux avoir sur l'état de ma propre pelouse (X) n'y change rien.

Tableau 1 : Circulation de l'information d'un nœud à un autre

3-4-2 Notion de probabilité associée

L’assimilation de la plausibilité d’un fait à une probabilité mathématique permet d’exprimer ce dernier quantitativement.

Reprenons l’exemple du graphe causal simple A cause B et supposons que la règle $A \Rightarrow B$ est vraie. Pour formaliser cette proposition en termes de probabilités, considérons deux variables A et B qui peuvent prendre toutes les deux les valeurs Vrai et Faux.

Probabilités a priori :

Du fait que nous n’avons aucune information sur les événements A et B, nous n’avons pas attribué *a priori* une croyance plus forte à un événement plutôt qu’à son contraire

Evénement	**Probabilité**	**Commentaire**
A = V	1/2	A priori, rien ne me permet de dire que A est plus certain que Ậ. J’attribue donc la probabilité 1/2 aux deux événements
A = F	1/2	

Probabilités conditionnelles

B = V | A = V	1	J’admets que la règle $A \Rightarrow B$ est vraie, donc, si *A* s’est réalisé, B l’est aussi.
B = F | A = V	0	
B = V | A = F	1/2	Par contre, si A ne s’est pas réalisé, je ne peux rien dire sur la réalisation de B.
B = F | A = F	1/2	

Pour connaître la valeur de P (A=V | B=F).

Considérons les événements A =V et A = F. Ils vérifient les conditions d’application du théorème de Bayes, puisque :

$(A = V) \cap (A= F) = \emptyset$

$(A = V) \cup (A = F) = E$

Le théorème de Bayes nous donne :

$$P(A = V \mid B = F) = \frac{P(B = F \mid A = V) * P(A = V)}{P(B = F \mid A = V) * P(A = V) + P(B = F \mid A = F) * P(A = F)}$$

Donc : P (A=V | B = F) = 0

3-4-3 Les réseaux Bayésiens

3-4-3-1 Définition :

Un réseau Bayésien est défini comme suit :

- Une suite de variables notée V et une suite d'arcs entre les variables, notée E,
- Chaque variable possède un nombre fini d'états exclusifs,
- Les variables et les arcs forment un graphe orienté acyclique (La définition des graphes orientés acyclique est présentées dans l'Annexe A), noté G=(V,E),
- À chaque variable A avec ses parents B_1, B_2,…., B_n est associée une probabilité conditionnelle $P(A \mid B_1, B_2, \ldots, B_n)$. Lorsque la variable ne possède pas de parent, la dernière quantité devient une probabilité marginale P(A) **[35]**

En d'autres termes **[36]** :

B = (G, θ) est un réseau Bayésien si G = (X, E) est un graphe acyclique dirigé dont les sommets représentent un ensemble de variables aléatoires $X = \{ X_1, X_2, \ldots, X_n \}$, et si $\theta_i = \left[P(X_i / X_{Pa(X_i)}) \right]$ est la matrice des probabilités conditionnelles du noeud i connaissant l'état de ses parents $Pa(X_i)$ dans G.

Une hypothèse imposée par la théorie des réseaux Bayésiens est que pour chaque variables X_i, l'ensemble de variables $Pa(X_i)$ doit être tel que X_i est conditionnellement indépendant à X_j ($j \neq i$) sachant $Pa(X_i)$ (noté $X_i \perp X_j | P_a(X_i)$).

Un réseau Bayésien B représente une distribution de probabilité sur X dont la loi jointe peut se simplifier de la manière suivante :

$$P(X_1, X_2, .., X_n) = \prod_{i=1}^{n} P(X_i / X_{p_a(X_i)})$$

3-4-3-2 Exemple de réseau Bayésien

Nous allons illustrer un exemple issu des travaux de Lauritzen et Spiegelhalter **[37]**; cité dans plusieurs références sur les réseaux Bayésiens **[34]**, **[38]**, et **[39]** ; il figure aussi dans la présentation de plusieurs logiciels de calcul sur les réseaux Bayésiens (Netica : ***http://www.norsys.com/netica.html***, Hugin : ***http://www.hugin.dk***/ et Bayesia : ***http://www.bayesia.com***) et il a pour domaine une base de connaissances médicales.

Énoncé de l'exemple :

La dyspnée (difficulté à respirer) peut être causée par une tuberculose, un cancer du poumon ou une bronchite (ou par aucune ou plusieurs d'entre ces causes). Un passage récent en Asie augmente les risques de tuberculose ; fumer est reconnu comme étant un facteur important de risque pour le cancer du poumon ou la bronchite. Une radiographie des poumons ne permet pas de différencier le cancer du poumon et la tuberculose, ni ne permet d'observer la présence ou non d'une dyspnée.

Considérons le cas d'un spécialiste des maladies des poumons désirant modéliser sa connaissance sur le diagnostic du cancer ou de la tuberculose afin de résoudre des situations telles que celle d'un patient atteint de dyspnée, ayant séjourné en Asie, sans connaissance sur son passé de fumeur, ni de radiographie des poumons. Le spécialiste voudrait pouvoir déterminer les probabilités d'existence des 3 maladies connues; ou encore connaître les causes après avoir donné les valeurs observées qui sont les symptômes.

Choix des variables aléatoires :

Fumer : Le patient consomme-t-il du tabac, qui prend les valeurs {Oui, Non}.

V. en Asie : Le patient a-t-il visité l'Asie, qui prend les valeurs {Oui, Non}.

Bronchite : Le patient a-t-il une bronchite, qui prend les valeurs {Présent, Absent}.

Cancer : Existence d'une tumeur au poumon, qui prend les valeurs {Présent, Absent}.

Tuber. : Le patient a-t-il la Tuberculose, qui prend les valeurs {Présent, Absent }.

Dyspnée : Le patient a-t-il des difficultés respiratoires, qui prend les valeurs {Présent, Absent}.

Radiologie : Résultat radiologique, qui prend les valeurs {Anormal, Normal}.

Hypothèses de modélisation :

- Les variables Fumer (consommation du tabac) et Visite en Asie sont les causes, supposées indépendantes; elle constituent donc les racines du réseau Bayésien.
- Les variables Bronchite, Cancer et Tuberculose sont les maladies, influencées par Visite en Asie pour la Tuberculose et consommation du tabac (Fumer) pour le Cancer et la Bronchite.
- Les variables Dyspnée et Radiologie constituent les symptômes ; on suppose que la Dyspnée est influencée par les 3 maladies, alors que la Radiologie est influencée par la Tuberculose et le Cancer. De plus, on n'est pas capable d'isoler les influences du Cancer et de la Tuberculose sur la Radiologie et sur la Dyspnée ; c'est pourquoi une nouvelle variable Tuber. ou Cancer (Tuberculose ou Cancer) a été introduite, elle exprime la présence de l'une au moins des maladies Cancer et Tuberculose.

Modélisation de la base de connaissances :

La base de connaissances peut être modélisée comme suit :

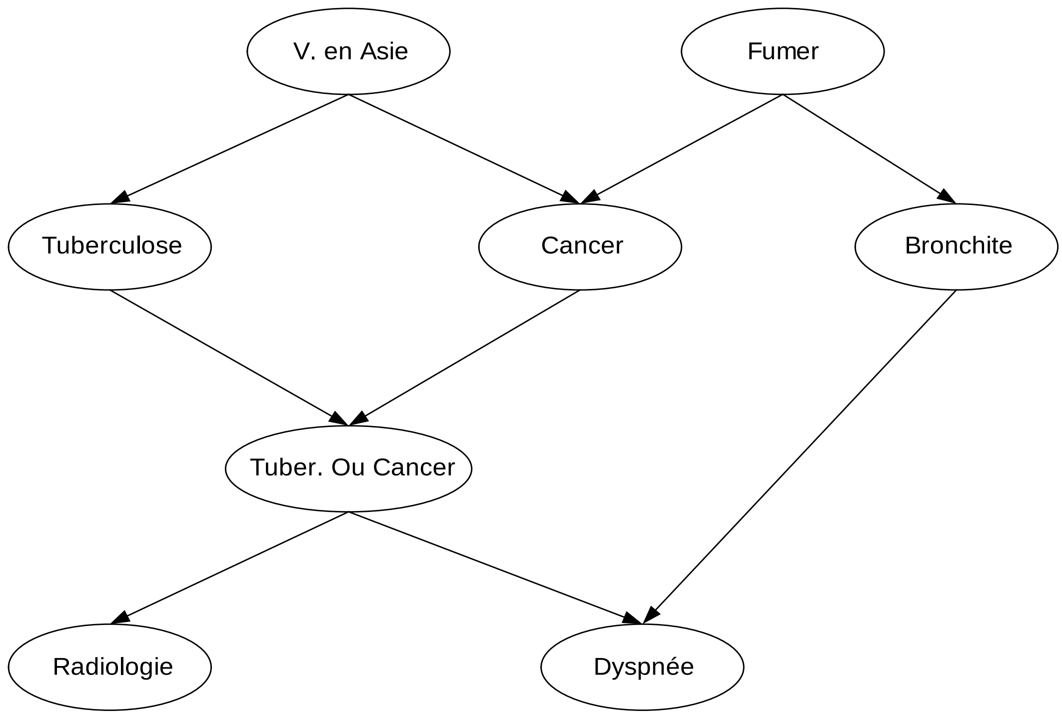


Figure 1 : Réseau Bayésien pour diagnostiquer la dyspnée

Lois des probabilités des variables aléatoires :

Les lois dans ce cas de variables discrètes sont des tables de probabilités; pour les deux variables sans parent (Fumer) et (V. en Asie) ce sont des probabilités absolues tandis que pour les autres variables ce sont des tables de probabilités conditionnelles, conditionnellement aux parents.

- *Pour (Fumer)* :

  P(Fumer = Oui) = 0.5.

- *Pour (V. en Asie)* :

  P(V.enAsie = Oui) = 0.01.

- *Pour (Bronchite)* :

  P(Bronchite = Présent/Fumer = Oui) = 0.6

  P(Bronchite = Présent/Fumer = Non) = 0.3

- *Pour (Tuberculose) :*

  P(Tuber. = Présent/V.enAsie = Oui) = 0.05

  P(Tuber. = Présent/V.enAsie = Non) = 0.01

- *Pour (Cancer) :*

  P(Cancer = Présent/Fumer = Oui) = 0.1

  P(Cancer = Présent/Fumer = Non) = 0.01

Pour ("Tuberculose ou Cancer"):

Tuber.	**Cancer**	**Tuber. Ou Cancer = vraie**
Présent	Présent	1
Présent	Absent	1
Absent	Présent	1
Absent	Absent	0

Tableau 2 : Loi de "Tuber. ou Cancer" conditionnellement à Tuber. et Cancer

Pour (Radiologie) :

Tuber. Ou Cancer	**Radiologie = Anormal**
Vrai	0.98
Faux	0.05

Tableau 3 : Loi de Radiologie conditionnellement à "Tuberculose ou Cancer"

Pour (Dyspnée) :

Tuber. Ou Cancer	**Bronchite**	**Dyspnée = Présent**
Vrai	Présent	0.9
Vrai	Absent	0.7
Faux	Présent	0.8
Faux	Absent	0.1

Tableau 4 : Loi de conditionnellement à "Tuber. ou Cancer" et Bronchite

3-4-3-3 Propriétés

Un Réseau Bayésien est caractérisé par une :

- Description qualitative des dépendances entre des variables (graphe causal) tel que :

- Le graphe ne doit pas être cyclique
- On peut numéroter les variables (ordre topologique)
- Description quantitative de ces dépendances – Probabilités conditionnelles-

Les réseaux Bayésiens possèdent de nombreuses appellations :

- Système expert Bayésien, Système expert probabiliste, réseau de croyance, réseau causal, etc.
- Belief network, Bayesian network, probabilistic independence networks.

3-5 Techniques de modélisation

3-5-1 Inférence

L'inférence dans un réseau Bayésien consiste à propager une ou plusieurs informations certaines au sens du réseau, (c'est à dire les valeurs établies par certaines variables), pour en déduire comment ceci intervient sur les probabilités d'évènements liés à d'autres variables.

Il existe plusieurs algorithmes d'inférence dans les réseaux Bayésiens. Ils sont classés en deux groupes :

- *Les méthodes d'inférence exactes* qui exploitent les indépendances conditionnelles contenues dans les réseaux et donnent à chaque inférence les probabilités *a posteriori* exactes. Par exemple l'algorithme Clustering effectue l'inférence en transformant le réseau en un arbre pour lequel chaque noeud regroupe plusieurs noeuds du réseau initial.
- *Les méthodes approchées* qui estiment les probabilités *a posteriori*. Pour ces méthodes, deux exécutions d'une inférence peuvent donner des probabilités *a posteriori* différentes. Par exemple les algorithmes d'échantillonnage stochastique qui estiment les probabilités en effectuant plusieurs tirages dans l'ensemble des combinaisons possibles des états des variables du réseau.

3-5-2 Apprentissage

L’apprentissage d’un réseau Bayésien consiste à estimer les lois de probabilités conditionnelles de chaque nœud du réseau (apprentissage des paramètres) et à trouver la meilleure structure du graphe représentant le problème à résoudre (apprentissage de la structure).

Les techniques d’apprentissage peuvent se regrouper en deux grandes familles, selon la disponibilité de données concernant le problème à traiter **[32]** :

- Apprentissage automatique à partir d’une base de données (données complètes ou non), par des approches statistiques ou Bayésiennes si la structure du réseau n'est pas fournie *a priori* par un expert.
- Acquisition de connaissance avec un expert du domaine.

La recherche scientifique dans le domaine des réseaux Bayésiens a permis l’élaboration d’algorithmes « d’apprentissage » performants, capables d’estimer automatiquement les paramètres des modèles et même d’élaborer leur structure à partir des connaissances *a priori* et des données disponibles. C’est en considérant que le nombre de modèles de causalités reliant un certain nombre de variables est fini, même s’il est grand, qu’on peut finalement envisager de se passer d’expert, et de construire un modèle uniquement à partir des données, en recherchant simplement parmi tous les modèles possibles celui qui représente le mieux la réalité.

Dans ce qui suit, et vu la disponibilité des données expérimentales de notre problématique et la possibilité de se passer de l’expert, on va se limiter à un apprentissage à partir des données pour élaborer notre structure du réseau et nos paramètres (tables de probabilités).

3-5-2-1 Apprentissage de paramètres

On entend par paramètres, les probabilités conditionnelles ou la partie quantitative du réseau Bayésien. Ce type d’apprentissage consiste à estimer les distributions de probabilités à partir de données disponibles complètes ou incomplètes :

Cas de données complètes :

- Apprentissage statistique :
  Cette méthode repose sur l’estimation d’un événement par sa fréquence d’apparition dans la base de données.
- Apprentissage Bayésien :

Cette méthode consiste à trouver les paramètres θ les plus probables en utilisant des a priori sur ces paramètres.

Cas de données incomplètes :

Généralement, les données sont souvent incomplètes. Certaines variables ne sont jamais observées ou ne le sont que partiellement. Les méthodes d'estimation les plus utilisées sont fondées sur l'algorithme itératif Expectation-Maximisation (EM).

- Apprentissage statistique :

  L'algorithme EM s'applique à la recherche des paramètres en répétant jusqu'à convergence de l'espérance et la maximisation **[32]**.
- Apprentissage Bayésien :

  Il suffit de remplacer le maximum de vraisemblance de l'étape de maximisation par un maximum (ou une espérance) a posteriori **[32]**.

3-2-2-2 Apprentissage de structure

L'objectif de ce type d'apprentissage est de trouver la meilleure structure d'un réseau Bayésien qui représente les données. Théoriquement, pour trouver une telle structure, tous les graphes possibles doivent être parcouru, en leur associant un score, puis en choisissant le graphe ayant le score le plus élevé. Mais pratiquement, il a été démontré dans **[41]** que r(n), le nombre de structures différentes pour un réseau Bayésien possédant n variables, est super-exponentielle. Il est donné par la formule de récurrence suivante :

$$r(n) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} 2^{i(n-1)} r(n-i) = n^{2^{O(n)}}$$

On trouve par exemple :

- r(1) = 1,
- r(2) = 3,
- r(3) = 25,
- r(4) = 543,
- r(5) = 29281,

- r(10) = $4,2 \times 10^{18}$.

Il est donc impossible d'effectuer un parcours exhaustif en un temps raisonnable dès que le nombre de noeuds dépasse 7 ou 8. La plupart des méthodes d'apprentissage de structure utilisent alors une heuristique de recherche dans l'espace des graphes acycliques dirigés.

De nombreuses méthodes sont proposées dans la littérature. On trouve à titre d'exemple :

L'algorithme PC, recherche de causalité :

Cet algorithme a été introduit par Spirtes, Glymour et Scheines en 1993 **[42]**. La structure du réseau Bayésien peut être construite à partir de l'ensemble des relations d'indépendances conditionnelles, en utilisant les tests statistiques d'indépendance conditionnelle entre deux variables quelconques conditionnellement à un ensemble de variables. Le principe de cet algorithme est le suivant :

- Construire un graphe non dirigé qui va contenir les relations entre les variables, en se basant sur les tests d'indépendance conditionnelle,
- Détecter les V-Structures, à ce niveau, le graphe obtenu est un graphe partiellement dirigé,
- Propager les orientations de certains arcs.

Un autre algorithme (IC) basé sur le même principe a été introduit par une équipe concurrente à la même époque **[43]**.Cet algorithme construit le graphe non orienté en rajoutant des arêtes au lieu d'en supprimer.

L'algorithme BN-PC :

Cette méthode de recherche de causalité nommée BN-PC-B **[44]** reprend le principe de l'algorithme PC en essayant de diminuer le nombre de tests d'indépendance conditionnelle. Elle est constituée de trois phases :

- Rechercher une structure initiale arborescente
- Rechercher des ensembles de conditionnement entre les variables pour décider s'il faut relier les nœuds correspondants
- Essayer de retirer des arcs superflus

Le graphe partiellement dirigé obtenu à la fin de la 3ème phase est alors orienté complètement de la même manière que l'algorithme PC.

L'arbre de poids maximal :

Cette méthode proposée par Chow et Liu **[45]** est dérivée de la recherche de l'arbre de recouvrement de poids maximal (maximal weight spanning tree ou MWST). Elle s'applique aussi à la recherche de structure d'un réseau Bayésien en fixant un poids à chaque arête potentielle A–B de l'arbre. Ce poids peut être :

- L'information mutuelle entre les variables A et B comme proposé par **[45]**,
- La variation du score local lorsqu'on choisit B comme parent de A **[46]**.

Une fois cette matrice de poids définie, il suffit d'utiliser un des algorithmes standard de résolution du problème de l'arbre de poids maximal comme l'algorithme de Kruskal **[47]**.

L'arbre non dirigé retourné par cet algorithme doit ensuite être dirigé en choisissant une racine puis en parcourant l'arbre par une recherche en profondeur. La racine peut être choisie aléatoirement ou à l'aide de connaissance a priori, ou encore en prenant la variable représentant la classe pour des problèmes de classification.

L'algorithme K2 : **[50]**

Cet algorithme tente de maximiser la probabilité de la structure à partir des données. En ajoutant un ordre sur les nœuds, de manière à ce qu'un noeud ne puisse être parent d'un autre que s'il précède celui-ci dans cet ordre, il est alors possible de réduire l'espace de recherche, qui devient alors l'ensemble des réseaux Bayésiens respectant cet ordre d'énumération. L'algorithme K2 teste l'ajout de parents en respectant cet ordre. Le premier noeud ne peut pas posséder de parents et pour les noeuds suivants, l'ensemble de parents choisi est celui qui augmente le plus la mesure Bayésienne.

La recherche gloutonne : **[48]**

L'algorithme de recherche gloutonne (Greedy Search) est un algorithme très répandu en optimisation. Son principe est le suivant :

- Partir d'un graphe de départ,
- Définir un voisinage de ce graphe,
- Associer un score à chaque graphe du voisinage, le meilleur graphe est alors choisi comme point de départ de l'itération suivante.

La notion de voisinage d'un graphe orienté acyclique (dag) est définie à l'aide de trois opérateurs :

- Ajout d'arc du graphe d'origine,
- Suppression d'arc du graphe d'origine,
- Inversion d'arc du graphe d'origine.

Si le graphe obtenu réalise un maximum (local) de la fonction de score, l'algorithme s'arrête. L'espace de recherche est alors l'espace complet des dag.

L'algorithme Structural-EM : **[49]**

Cette méthode a été introduite en se basant sur le principe de l'algorithme EM. Elle permet de traiter le cas des données incomplètes sans avoir à ajouter une nouvelle modalité (variable non mesurée) à chaque noeud.
Cette méthode itérative part d'une structure initiale pour estimer la distribution de probabilité des variables cachées ou manquantes grâce à l'algorithme EM classique. L'espérance d'un score par rapport à ces variables cachées est ensuite calculée pour tous les réseaux Bayésiens du voisinage afin de choisir la structure suivante.

3-6 Pourquoi utilise-t-on les réseaux Bayésiens?

3-6-1 Domaines d'application

Les réseaux Bayésiens possèdent plusieurs domaines d'applications :

Santé :
Le domaine du diagnostic médical fut parmi les premières applications des réseaux Bayésiens. Le projet Human Genome est une application récente des algorithmes issus des réseaux Bayésiens qui a permis d'améliorer considérablement la recherche de la localisation de certains gènes.

Industrie :

Les applications dans le domaine industriel sont :

- Diagnostic de pannes en fonction de mesures du système
- Fusion de données :

La société danoise HUGIN a développé pour le compte de la société LOCKHEED un système de contrôle d'un véhicule sous-marin autonome. Ce système évalue en permanence les capacités du véhicule a réagir à certains types d'événements.

Banque/Finance :

Plusieurs applications existent; on les trouve surtout dans les domaines suivant :

- Analyse financière,
- Scoring,
- Évaluation de risques, détection de fraudes.

Saurav Dutta et Lynford E. Graham de Rutgers University ont montré la faisabilité de l'application des réseaux Bayésiens à l'analyse financière, en particulier pour la prédiction de faillites.

Informatique :

Les réseaux Bayésiens peuvent être utilisés pour des applications d'accès intelligent à l'information, d'adaptation intelligente des systèmes aux utilisateurs.

3-6-2 Points forts et limites des réseaux Bayésiens

3-6-2-1 Avantages des réseaux Bayésiens:

Une représentation intuitive d'un domaine de connaissance :

La représentation des connaissances utilisées dans les réseaux Bayésiens est la plus intuitive possible, il s'agit simplement de relier des causes et des effets par des flèches.

Cette représentation des connaissances par liens entre causes et effets est souvent plus naturelle que la représentation par règles de production. Ainsi, il est plus facile pour un expert de formaliser ses connaissances sous forme de graphe causal que sous forme de système à base de règles.

Une représentation de connaissances lisible :

La représentation des connaissances est assez lisible (par opposition aux réseaux de neurones par exemple, ou même aux arbres de décision).

Des capacités d'apprentissage :

Les algorithmes actuels permettent l'apprentissage de façon très complète même en l'absence totale de connaissances et par la méthode du maximum de vraisemblance on peut aboutir à la structure du réseau la plus adaptée; aussi la détermination des matrices de probabilités conditionnelles peut être effectuée par simple calcul de fréquences ou par détermination du maximum de vraisemblance si on dispose de connaissances sur la structure de causalité, et d'une base d'exemples représentative.

Des capacités d'explication et de simulation très complète :

Le fonctionnement général d'un réseau Bayésien est celui d'un système de propagation d'informations. On peut propager des informations de différentes natures : certitude positive (on connaît avec certitude la valeur de l'un des noeuds du réseau), certitude négative (on peut éliminer avec certitude l'une des possibilités pour l'un des noeuds, sans pour autant connaître sa valeur), ou information incertaine.

Disponibilité des outils et qualité de l'offre en matière de logiciels :

Les algorithmes d'apprentissage de structures et de paramètres sont disponibles sous plusieurs plateformes.

Un formalisme unificateur :

Les réseaux Bayésiens permettent de traiter des applications intégrant des degrés d'incertitudes. Cette particularité est prise en compte par le formalisme probabiliste. Les techniques disponibles pour traiter ce genre de problème (modèles de regression, réseaux de neurones, arbres de décision) ne sont pas construites sur un tel formalisme.

Les réseaux Bayésiens permettent de considérer dans un même formalisme la représentation de modèles de causalité et les statistiques multivariées.

Les techniques les plus utilisées pour le data mining comme les arbres de décision ou les réseaux de neurones peuvent également être représentés au sein de ce formalisme, qui permet également de prendre en compte l'optimisation de certains des paramètres d'apprentissage du réseau.

3-3-2-2 Comparaison avec d'autres techniques

Le tableau suivant, extrait de **[32],** permet d'illustrer une comparaison des réseaux Bayésiens et des autres techniques de modélisation telle que les réseaux neuronaux, les arbres de décision et les systèmes experts.

Connaissances	**Analyse de données**	**Réseaux neuronaux**	**Arbres de décision**	**Systèmes experts**	**Réseaux Bayésiens**
ACQUISITION					
Expertise seulement				*	
Données seulement	+	*	+		+
Mixte	+	+	+		*
Incrémental		+			*
Généralisation	+	*	+		+
Données incomplètes		+			*
REPRÉSENTATION					
Incertitude				+	*
Lisibilité	+	+	+	+	*
Facilité			*		
Homogénéité					*
UTILISATION					
Requêtes élaborées	+			+	*
Utilité économique	+	+			*
Performances	+	*			

Tableau 5 : Comparaison des réseaux Bayésiens avec d'autres techniques

L'acquisition, la représentation et l'utilisation des connaissances. La représentation adoptée est la suivante :

- Á chaque ligne correspond une caractéristique, qui peut être un avantage, ou la prise en compte d'un problème spécifique.
- Si la technique considérée permet de prendre en compte ce problème, ou présente cet avantage, un signe + est placé dans la case correspondante.
- Un signe * est placé dans la case de la meilleure technique du point de vue de la caractéristique considérée

3-6-2-3 Limites des réseaux Bayésiens

Probabilités :

L'utilisation des graphes de causalité est une approche très intuitive et naturelle. L'utilisation des probabilités qui est essentielle pour rendre ces modèles quantitatifs reste cependant non intuitive par rapport à un décideur.

Complexité des algorithmes :

La généralité du formalisme des réseaux Bayésiens aussi bien en terme de représentation que d'utilisation les rend souvent difficile à partir d'une certaine taille. La complexité des réseaux Bayésiens ne se traduit donc pas seulement en termes de compréhension par les utilisateurs.

Apprentissage :

Du point de vue de l'apprentissage, les réseaux Bayésiens disposent d'un recul très inférieur à celui dont disposent maintenant les réseaux neuronaux.

Lisibilité des graphes :

Si le graphe représente un grand nombre de nœuds, il est moins lisible que s'il est représenté par un arbre de décision. En plus, les probabilités ne sont pas représentables et dans ce cas on ne peut pas distinguer si un arc et important ou non.

Variables continues :

Les algorithmes d'apprentissage et d'inférence ainsi que les outils disponibles pour la manipulation des réseaux Bayésiens utilisent jusqu'à présent seulement des variables discrètes.