Réseaux Bayésiens

Bruno Bouzy

bruno.bouzy@u-paris.fr

Janvier 2023

Revoici les corneilles et les corbeaux !

Observations :

	gris	noir	total
corneille	5	45	50
corbeau	30	70	100
total	35	115	150

Probabilités joi

P(Corneille, Gris) = 5/150 = 0.033 P(Corneille, Noir) = 45/150 = 0.3

P(Corbeau, Gris) = 30/150 = 0.2 P(Corbeau, Noir) = 70/150 = 0.467

	gris	noir	
corneille	0.033	0.3	0.333
corbeau	0.2	0.467	0.667
oiseau	0.233	0.767	1

De l'inférence bayésienne vers les réseaux bayésiens

- Vocabulaire
  - Caractéristiques et classes → Variables
  - Valeurs → Evénements
  - Valeurs booléennes (vrai / faux)
  - N est le nombre de variables
  - P(A) (respectivement P(~A)) signifie probabilité que A soit vrai (resp. faux)
- Deux bonnes nouvelles :
  - Avec la table de probabilités jointes, on peut retrouver n'importe quelle probabilité (proba à priori, proba conditionnelle, etc.) (cf séance sur l'inférence bayésienne)
  - 2^N-1 valeurs de probabilités jointes sont nécessaires et suffisantes.
    - (S'il manque une seule proba jointe dans la table, on peut la retrouver)
- Une mauvaise nouvelle cependant :
  - La taille de la table vaut 2^N est exponentielle en fonction du nombre de variables
- Comment représenter l'information autrement et sans perte ?
  - Réponse : avec des « réseaux bayésiens » en utilisant des indépendances de variables

Un cas d'indépendance

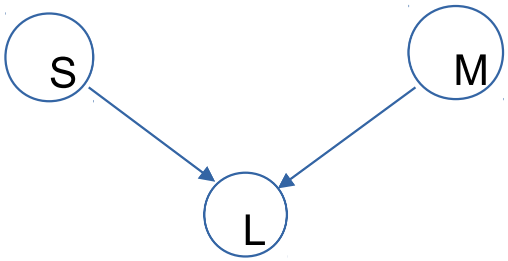
- Exemple avec deux variables
  - M : Manuela enseigne (sinon c'est Andrew)
  - S : il fait beau (Sunny)
- Hypothèse :
  - M et S sont indépendants (le fait que ce soit Manuela ou Andrew qui enseigne n'influence pas le fait qu'il fasse beau ou pas)
  - $P(S|M) = P(S)$ (7)
- Equations équivalentes:
  - $P(\sim S|M) = P(\sim S)$ (8) $P(M|S) = P(M)$ (9)
  - $P(M, S) = P(M)\ P(S)$ (10) $P(M, \sim S) = P(M)\ P(\sim S)$ (11)
  - $P(\sim M, S) = P(\sim M)\ P(S)$ (12) $P(\sim M, \sim S) = P(\sim M)\ P(\sim S)$ (13)
- Numériquement, on suppose 2 informations connues :
  - $P(S) = 0.3$ (14)
  - $P(M) = 0.6$ (15)
- Dans ce cas d'indépendance, avec 2 infos, on retrouve n'importe quelle autre probabilité.
- Or, dans le cas général, il fallait 2^2-1 = 3 infos
- Intuition : une indépendance de variables semble diminuer le nombre d'infos nécessaires pour tout retrouver.

Un cas de dépendance

- Exemple enrichi avec une 3ème variable
  - L : L'enseignant arrive en retard (Late = L)
- Hypothèse :
  - L dépend de S et M (L'enseignant est retardé par le mauvais temps et Andrew arrive plus souvent en retard que Manuela)
  - On connait cette dépendance par des probabilités conditionnelles
  - $P(L \mid M, S) = 0.05$ (16) $P(L \mid M, \sim S) = 0.1$ (17)
  - $P(L \mid \sim M, S) = 0.1$ (18) $P(L \mid \sim M, \sim S) = 0.2$ (19)
- Avec les 4 informations ci-dessus et les 2 informations P(M) et P(S), peut-on retrouver toutes les probabilités conditionnelles ?
  - OUI par exemple : $P(L, M, S) = P(L \mid M, S)\, P(M, S) = P(L \mid M, S)\, P(M)\, P(S)$
- Dans ce cas, avec une indépendance entre M et S et la dépendance conditionnelle de L à M et S, avec 4+2=6 infos, on retrouve n'importe quelle autre probabilité.
- Or, dans le cas général, il fallait $2^3-1 = 7$ infos
- Intuition : une dépendance de variables semble diminuer le nombre d'infos nécessaires pour tout retrouver.

Notation graphique

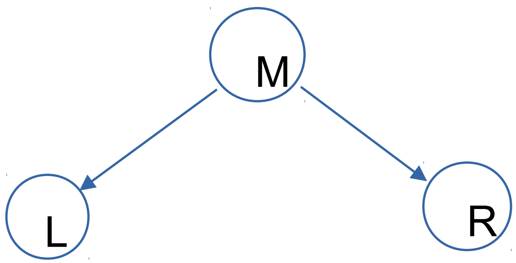
- A chaque variable est associé un nœud.
- A chaque relation de dépendance est associée une flèche entre nœuds.
- A chaque nœud est (sont) associée(s) une (des) information(s) permettant de calculer la probabilité que la variable du nœud soit vraie.



- L'équation (14) est associée au nœud S
- L'équation (15) est associée au nœud M
- Les équations (16) (17) (18) (19) sont associées au nœud L
- Les deux flèches indiquent que S et M « influencent » L.

Indépendance conditionnelle

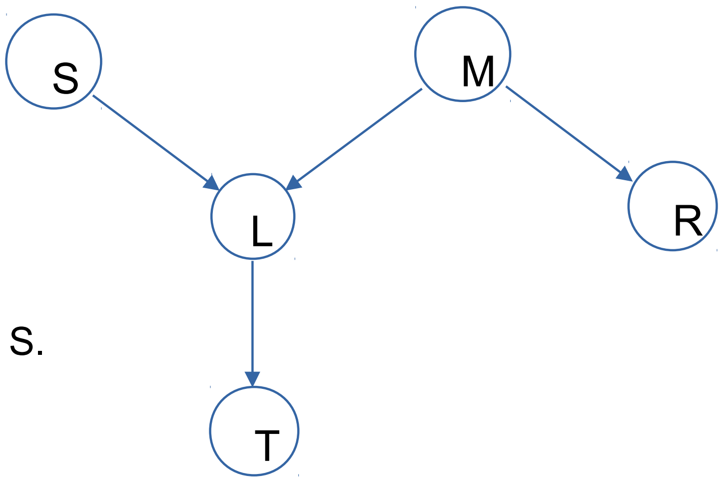
- Hypothèses : exemple avec 3 variables :
  - M : Manuela enseigne (sinon c'est Andrew)
  - L : L'enseignant arrive en retard (Late = L)
  - R : le sujet de l'enseignement porte sur les Robots (R)
  - L dépend de M et R dépend de M
  - M étant connu, L et R sont indépendants :
  - P(R | M, L) = P(R | M) et P(R | ~M, L) = P(R | ~M)
- Graphiquement :



- On suppose donc connus :
  - P(L | M) = 0.085 (20) P(L | ~M) = 0.17 (21)
  - P(R | M) = 0.3 (22) P(R| ~M) = 0.6 (23)
- Avec les 4 informations ci-dessus et l'information P(M), peut-on retrouver toutes les probabilités ?
  - OUI par exemple : P(L, M, R) = P(L | M, R) P(M, R) = P(L | M) P(R | M) P(M)
- Dans ce cas, avec 4+1= 5 infos, on retrouve n'importe quelle autre probabilité.
- Or, dans le cas général, il fallait 2^3-1 = 7 infos
- Intuition : une indépendance conditionnelle de variables semble à nouveau diminuer le nombre d'infos nécessaires pour tout retrouver.

Exemple avec 5 variables (1/2)

- Hypothèses :
  - T : Le cours commence à 8:35 (Time = T)
  - L : L'enseignant arrive en retard (Late = L)
  - R : le sujet de l'enseignement porte sur les Robots (R)
  - M : Manuela enseigne (sinon c'est Andrew)
  - S : il fait beau (Sunny)

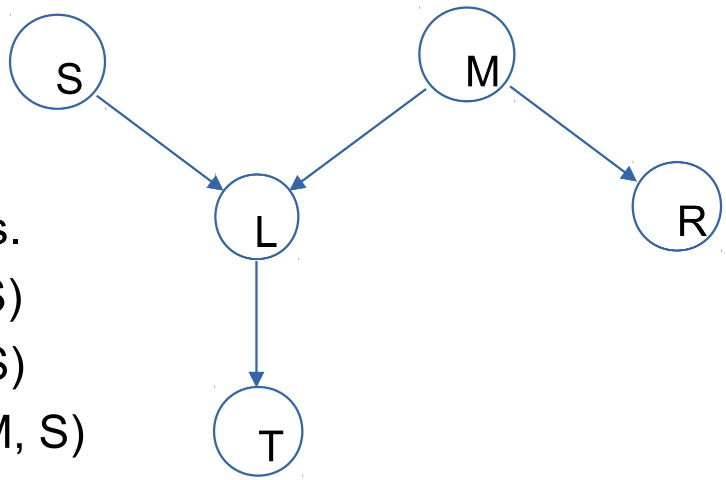


- Influences et indépendances conditionnelles :
  - T influencé par L et indépendant de R, M, S sachant L.
  - L influencé par M et S et indépendant de R sachant M et S.
  - R influencé par M et indépendant de L et S sachant M.
  - M et S indépendants.

- Graphiquement : 5 nœuds et 4 flèches

Exemple avec 5 variables (2/2)

- On suppose connues 10 informations :
  - P(M)   P(S)
  - P(L | M, S)   P(L | M, ~S)   P(L | ~M, S)   P(L | ~M, ~S)
  - P(R | M)   P(R| ~M)
  - P(T | L)   P(T | ~L)



- On peut retrouver toutes les probabilités souhaitées.
  - P(T, L, R, M, S) = P(T | L, R, M, S) P(L, R, M, S)
  - = P(T | L)   P(L, R, M, S)
  - = P(T | L) P(L | R, M, S) P(R, M, S)
  - = P(T | L) P(L | M, S)   P(R, M, S)
  - = P(T | L) P(L | M, S) P(R | M, S) P(M, S)
  - = P(T | L) P(L | M, S) P(R | M)   P(M) P(S)
- Dans ce cas, avec 10 infos au lieu de 31 infos, on retrouve n'importe quelle autre probabilité.

« Formalisation »

- Réseau bayésien = graphe orienté acyclique
- Un nœud contient :
  - une variable
  - Une table de probabilités de cette variable en fonction des valeurs des variables « parents » dont elle dépend.
- Construction avec m variables Xi 1<=i <=m
- Pour i allant de 1 à m
  - Ajouter Xi au graphe
  - Parents(Xi) : { nœuds Y du graphe influençant Xi }
  - Mettre à jour la table des probabilités conditionnelles P(Xi | Parents(Xi))

Calcul d'une probabilité jointe

Dans le cas général, étant donné un réseau bayésien, on a :

$P(X1, X2, \ldots, Xn) = \prod_i P(Xi \mid \text{affectations de Parents}(Xi))$

Dans l'exemple à 5 variables ci-avant :

$P(T, L, R, M, S)$	$= P(T \mid L, R, M, S)\, P(L, R, M, S)$	bayes
	$= P(T \mid L)\, P(L, R, M, S)$	indépendance conditionnelle
	$= P(T \mid L)\, P(L \mid R, M, S)\, P(R, M, S)$	bayes
	$= P(T \mid L)\, P(L \mid M, S)\, P(R, M, S)$	indép. condition.
	$= P(T \mid L)\, P(L \mid M, S)\, P(R \mid M, S)\, P(M, S)$	bayes
	$= P(T \mid L)\, P(L \mid M, S)\, P(R \mid M)\, P(M, S)$	indép. condition.
	$= P(T \mid L)\, P(L \mid M, S)\, P(R \mid M)\, P(M)\, P(S)$	indépendance

Calcul (exact) d'une probabilité conditionnelle arbitraire

Dans chacun de ses nœuds, le réseau contient des probabilités :

- « Tout court » dans les nœuds sans parent
- Conditionnelles dans les nœuds avec parent

Toutes les probabilités conditionnelles possibles ne sont pas explicitement présentes dans le réseau.

Comment calculer une probabilité conditionnelle non présente dans le réseau ?

Avec la formule (8) utilisant les probas jointes (cf diapo précédente et cf séance précédente)

Si 2 variables : P(A | B) = P(A, B) / (P(A, B) + P(~A, B))

Si 3 variables :

P(A | B, C) = P(A, B, C) / (P(A, B, C) + P(~A, B, C))

P(A, B | C) = P(A, B, C) / (P(A, B, C) + P(~A, B, C) + P(A, ~B, C) + P(~A, ~B, C))

Avec N variables à gauche du | :

Cela peut supposer le calcul (exponentiel) de 2^N probabilités jointes.

Mauvaise nouvelle.

Calcul approximatif d'une probabilité conditionnelle arbitraire

Pour éviter le calcul exponentiel en fonction de N, on peut utiliser un algorithme Monte-Carlo.

Idée : Effectuer M fois le parcours du réseau en tirant au hasard la valeur de la variable associé à un nœud.

```
Ms = 0    // nombre de fois où le parcours se termine par un succès (variables a gauche du | valorisées correctement)
Faire M fois
    Ms += parcours_réseau()
parcours_réseau() :
    Pour chaque nœud : noeud_traité = faux
    Tant que nœud non traité sans parent non traité,
        Si la variable n'est pas conditionnée
            proba_locale = lire_proba_dans _reseau(affectation des parents)
            valeur = tirer la valeur de la variable du nœud au hasard  suivant proba_locale
        Sinon valeur = valeur_conditionnée(variable)
        nœud_traité = vrai
    Si les variables à gauche du | sont correcte retourner 1
    Sinon retourner 0
Proba_conditionnelle = Ms/M
```

Le calcul est linéaire en fonction de M et de N. Bonne nouvelle.

Le calcul est approximatif. Plus M est grand plus le résultat tend vers la probabilité conditionnelle exacte.

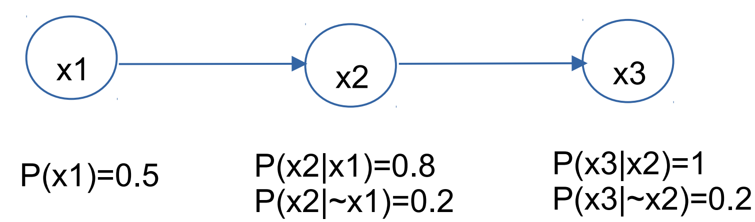
Induction de réseau bayésien à partir de données (1/3)

	x1	x2	x3
1	+	-	-
2	+	+	+
3	-	-	+
4	+	+	+
5	-	-	-
6	-	+	+
7	+	+	+
8	-	-	-
9	+	+	+
10	-	-	-

Données

Induction

Un réseau bayésien induit
(réseau bayésien 1)



Pour un jeu de données, il existe plusieurs réseaux bayésiens corrects.
Le jeu de données ne décide pas des relations entre variables.
Plusieurs jeux de relations entre variables sont possibles.
En construisant ce réseau, on a décidé (un peu arbitrairement) que x2 est influencé explicitement par x1 et que x3 est influencé explicitement par x2, mais pas par x1 si x2 est connu.

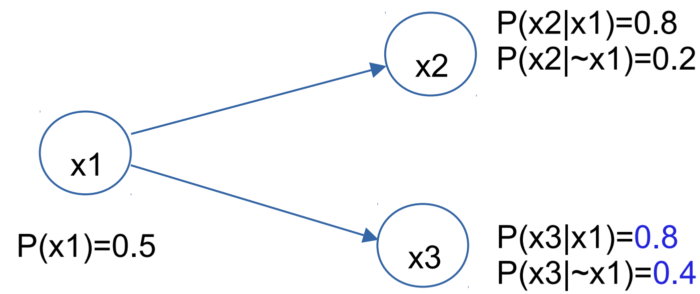
Induction de réseau bayésien à partir de données (2/3)

	x1	x2	x3
1	+	-	-
2	+	+	+
3	-	-	+
4	+	+	+
5	-	-	-
6	-	+	+
7	+	+	+
8	-	-	-
9	+	+	+
10	-	-	-

Données

Induction

Un autre réseau bayésien induit (réseau bayésien 2)



Si on fixe les relations du réseau, alors les probas conditionnelles sont fixées.
En construisant ce réseau, on a décidé (un peu arbitrairement) que x2 et x3 était indépendants étant donné x1.

Réseaux bayésiens

Induction de réseau bayésien à partir de données (3/3)

Si on calcule $P(x3|x1)$ et $P(x3|\sim x1)$ issues du réseau bayésien 1 avec les formules bayésiennes en passant par x2 et ~x2, on a :

$$P(x3|x1) = P(x3|x2,x1)P(x2|x1) + P(x3|\sim x2,x1)P(\sim x2|x1)$$
$$= P(x3|x2)P(x2|x1) + P(x3|\sim x2)P(\sim x2|x1)$$
$$P(x3|\sim x1) = P(x3|x2,\sim x1)P(x2|\sim x1) + P(x3|\sim x2,\sim x1)P(\sim x2|\sim x1)$$
$$= P(x3|x2)P(x2|\sim x1) + P(x3|\sim x2)P(\sim x2|\sim x1)$$

Numériquement on trouve :

$P(x3|x1) = 1\ 0.8 + 0.2\ 0.2 = 0.84$ et $P(x3|\sim x1) = 1\ 0.2 + 0.2\ 0.8 = 0.36$

Si on induit $P(x3|x1)$ et $P(x3|\sim x1)$ directement avec le jeu de données pour obtenir le réseau bayésien 2 alors :

$P(x3|x1) = 0.8$ et $P(x3|\sim x1) = 0.4$

D'où vient la différence ?

- Les relations décidées par l'expérimentateur sans connaître le domaine duquel sont extraites les données, introduit un biais.
- Dans les réseaux 1 et 2, les jeux de relations sont distincts et introduisent des biais cachés distincts.
- D'où la différence.
- Les experts d'un domaine sont les seuls capables de donner les indépendances (conditionnelles ou pas), les influences, c'est-à-dire les relations entre variables (les flèches du graphe)

Indépendance de variables et influence de variables

Dans un réseau bayésien, il faut distinguer indépendance de variables et influence de variables (visibles sur le graphe)

Avec un réseau simple où V (voleur) influence A (Alarme), V et A sont dépendants l'un de l'autre.

Avec le réseau bayésien 1, x1 influence x2 qui influence x3. Cependant, si x2 et connu, alors x3 n'est pas influencé par x1.

Avec le réseau bayésien 2, x3 et x2 sont indépendants conditionnellement à la connaissance de x1.

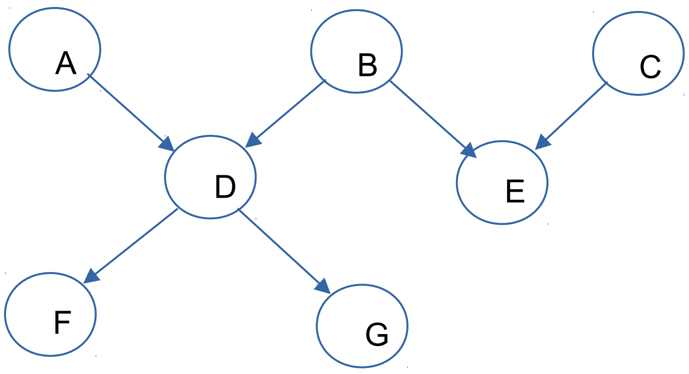
Existe-t-il un critère pour dire si deux variables sont indépendantes l'une de l'autre ?

Oui : la d-séparation.

La d-séparation

La règle de d-séparation :

Un nœud est conditionnellement indépendant de ses non descendants connaissant les valeurs des noeuds parents.



Sur cet exemple,

A, B, C sont indépendants inconditionnellement

A et B étant connus, D est indépendant de C et E

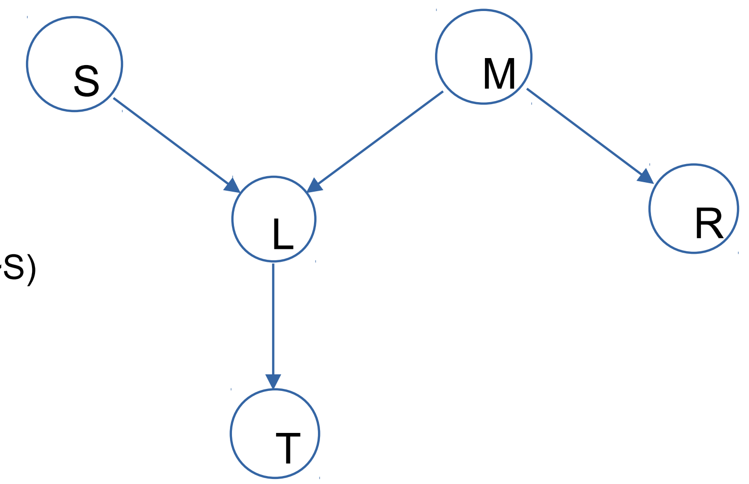
B et C étant connus, E est indépendant de A, D, F, G

D étant connu,

- F est indépendant de A, B, C, E, G
- G est indépendant de A, B, C, E, F

Retour sur l'exemple à 5 variables

On veut calculer P(R | T, ~S)



Méthode brute : avec les probabilités jointes :

P(R | T, ~S) = P(R, T, ~S) / (P(R, T, ~S) + P(~R, T, ~S))

(c'est long, difficile et source d'erreur mais c'est la bonne méthode)

Y a-t-il une méthode en passant par M ?

P(R | T, ~S) = P(R | T, ~S, M) P(M | T, ~S) + P(R | T, ~S, ~M) P(~M | T, ~S)

= P(R | M) P(M | T, ~S) + P(R | ~M) P(~M | T, ~S)

Désormais il faut calculer P(M | T, ~S)...

P(M | T, ~S) = P(M | T, ~S, L) P(L | T, ~S) + P(M | T, ~S, ~L)P(~L | T, ~S)

On n'avance pas : la méthode passant par M ne marche pas.

Avec la règle de d-séparation :

R est indépendant de S, L, T connaissant M : mais on ne connaît pas M...

Message : les notions d'influence, indépendance, conditionnelle ou pas, sont subtiles.

Conclusion, résumé

- Réseau bayésien (RB) :
  - Définition, exemples, indépendance
  - Stockage compact
- Calcul
  - de probabilité jointe
    - linéaire en fonction de N
  - exact de probabilité conditionnelle arbitraire
    - exponentiel en fonction de N
  - approximatif de proba conditionnelle arbitraire
    - linéaire en fonction de N
- Induction de RB
  - Les relations d'influence sont difficiles à déterminer
  - Vraisemblance d'un RB par rapport à un jeu de données (non abordé)
- D-séparation, subtilités.