Probabilistic Graphical Models, pyAgrum & Causality

Pierre-Henri WUILLEMIN & Gaspard DUCAMP

LIP6 & ex-LIP6

{17|22}/06/2021

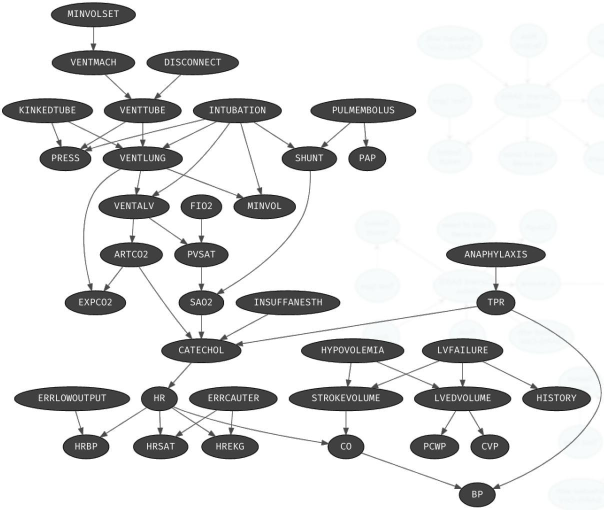
D-séparation

Classification et BNs

Bayesian classification

Causalité

Réseaux bayésien et information



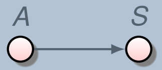
Qui communiquent ? Dans quel contexte ?

Communications & contextes

Indépendance garantie

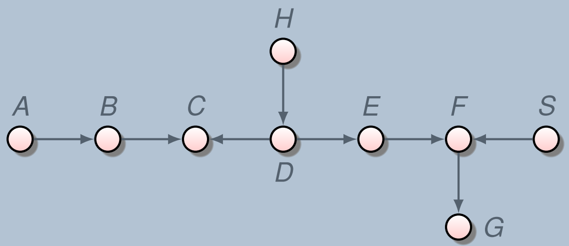
$\Rightarrow A \perp\!\!\!\perp S$

Dépendance garantie



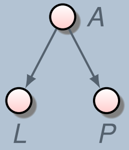
$\Rightarrow A \not\perp\!\!\!\perp S$

cas (plus) général



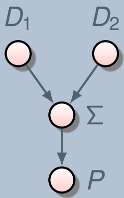
$\exists K \subset \mathbb{X}, A \perp\!\!\!\perp S \mid K$ et $\exists K' \subset \mathbb{X}, A \not\perp\!\!\!\perp S \mid K'$

Lecture & Pointure



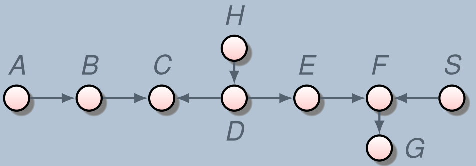
$$\Rightarrow \begin{cases} L \not\perp\!\!\!\perp P \\ L \perp\!\!\!\perp P \mid A \end{cases}$$

Dés & Parité



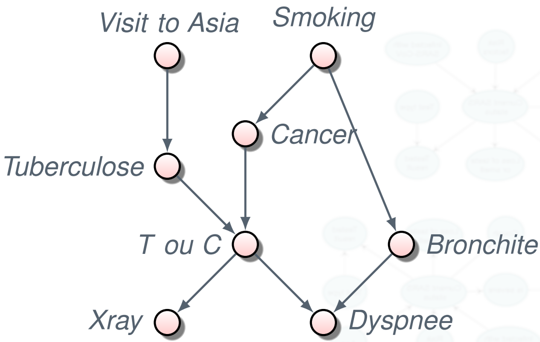
$$\Rightarrow \begin{cases} D_1 \perp\!\!\!\perp D_2 \\ D_1 \not\perp\!\!\!\perp D_2 \mid \Sigma \\ D_1 \not\perp\!\!\!\perp D_2 \mid P \end{cases}$$

D-séparation



- $A \perp\!\!\!\perp S$?
- $A \perp\!\!\!\perp S \mid C$?
- $A \perp\!\!\!\perp S \mid C, G$?
- $A \perp\!\!\!\perp S \mid C, G, E$?

Du modèle statistique aux connaissances expertes



- Savoir si le patient fume n'est pas pertinent si il n'a ni cancer ni bronchite.
- Si un patient a une dyspnée, savoir qu'il n'a pas de tuberculose augmente les risques de cancer.
- Les risques de Tuberculose et de Bronchite sont indépendants, mais ne le sont plus si la Radiographie est positive.

Pause notebook

Sur la d-séparation

D-séparation

Classification et BNs

Bayesian classification

Causalité

X (dimension d, features) and Y (dimension 1)

Using a database $\Pi_a = (x^{(k)}, y^{(k)})_{k \in \{1,\dots,N\}}$ (supervised learning), one can estimate the joint distribution $P(X, Y)$.

Classification

For a vector x, values of X, the goal is to predict the class : $\widehat{y}$.

1. Maximum of the likelihood (ML) $$\widehat{y}_{ML} = \arg\max_y P(x|y)$$
2. Maximum a posteriori (MAP) $$\widehat{y}_{MAP} = \arg\max_y P(y) \cdot P(x|y)$$

Those distributions may be hard to estimate.

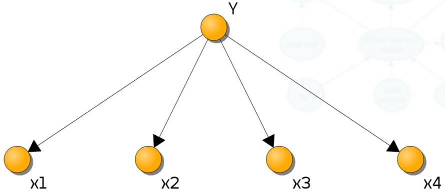
$P(X|Y)$ may induce more parameters than $|\Pi_a|$!!

How to compute $P(X|Y)$?

Naive Bayes classifier

if we assume $\forall k \neq l, X_k \perp\!\!\!\perp X_l \mid Y$ then $P(x,y) = P(y) \cdot \prod_{k=1}^{d} P(x_k|y)$

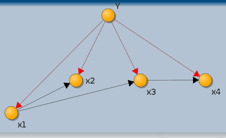
Very strong assumption ! In most cases, it is an approximation.
However, this approximation often gives good results.



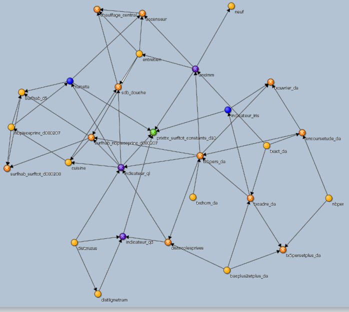
- Parameters estimation : trivial (if no missing values)
- ML : $\prod_{k=1}^{d} P(x_k|y)$ ...
- MAP : $P(y|x_1, \cdots, x_d)$: inference in the BN !

TAN : Tree-Augmented Naive Models

Every variable X_i can have Y and another parent among X (only one !).



Complete Bayesian network



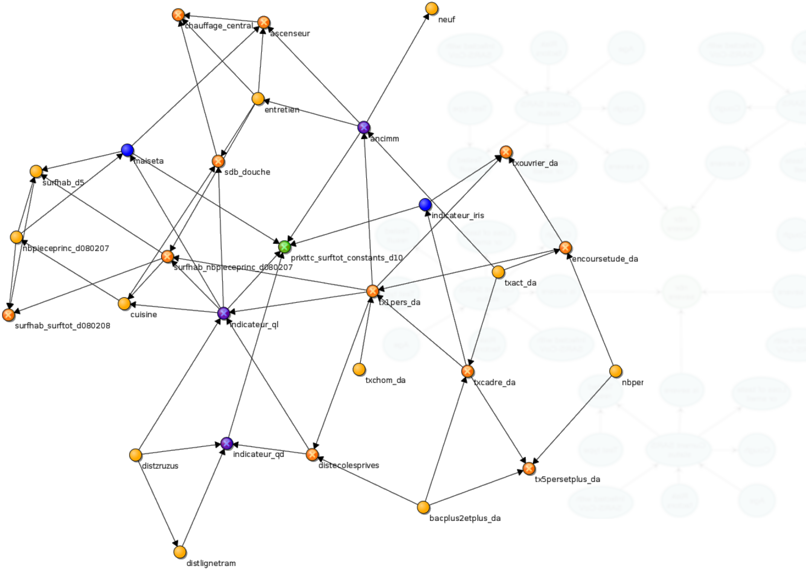
In a BN including Y and (X_i), infer $P(Y|X_1, \cdots, X_n)$.

Note : There is no need to all X_i :

Markov Blanket $MB(.)$

$$P(Y|X) = P(Y|MB(Y))$$

Zoom



chauffage_central
ascenseur
neuf
entretien
ancimm
maiseta
sdb_douche
surfhab_d5
txouvrier_da
indicateur_iris
nbpieceprinc_d080207
prixttc_surftot_constants_d10
encoursetude_da
surfhab_nbpieceprinc_d080207
txact_da
tx1pers_da
cuisine
surfhab_surftot_d080208
indicateur_ql
txchom_da
txcadre_da
nbper
distzruzus
indicateur_qd
distecolesprives
tx5persetplus_da
bacplus2etplus_da
distlignetram

Pause notebook

Sur la classification

D-séparation

Classification et BNs
Bayesian classification

Causalité

La posture des statistiques

“We often see the terms cause, effect, and causal modeling used in the research literature. We do not endorse this practice and therefore do not use these terms here !”

(Schumaker and Lomax, 1996)

“It would be very healthy if more researchers abandoned thinking of and using terms such as cause and effect”

(Muthen, 1987)

Observation versus intervention

⚠ Application directe du théorème de Bayes (si les lois sont >0).

$P(R, W) = P(R) \cdot P(W \mid R)$

$R \longrightarrow W$

$P(R, W) = P(W) \cdot P(R \mid W)$

$R \longleftarrow W$

⚠ Les relations symétriques ne sont pas capables de représenter la causalité. Il faut définir une nouvelle notion non-symétrique liées aux notions de probabilités observationnelles (marginales et conditionnelles).

Proposition : intervention et sémantique causale

$Rain \Longrightarrow Wet$

$$\begin{cases} P(W|do(R)) = P(W|R) \\ P(R|do(W)) = P(R) \end{cases}$$

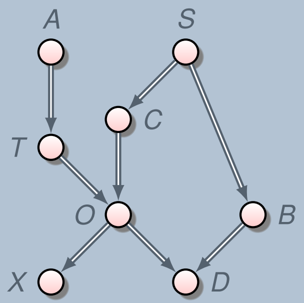
Intervention dans un réseau bayésien causal *parfait*

Dans un modèle bayésien causal, une intervention sur une variable

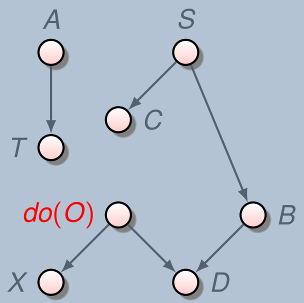
- se propage probabilistiquement vers ses enfants,
- ne se propage par vers ses parents.

⚠ Les calculs sont probabilistes mais dans un BN *mutilé*.

Modèle causal : M

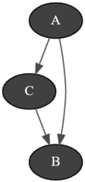


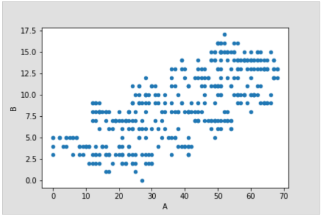
$BN_{do(O)}$

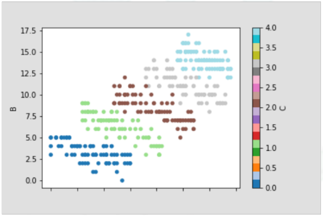


$$P_M(X|do(O)) = P_{BN_{do(O)}}(X|O)$$

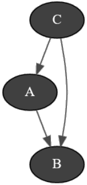
Le modèle observationnel :

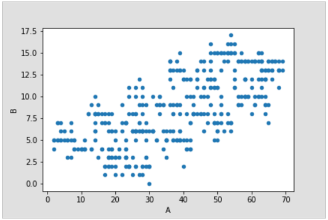


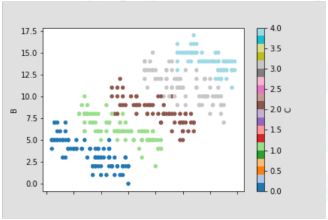




Un équivalent de Markov :

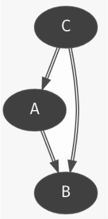


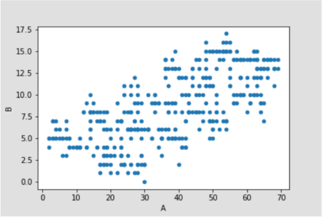


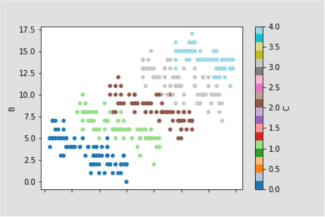


Retour sur Simpson

- Le modèle causal :

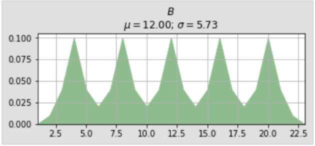




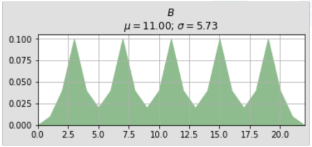


Sémantique : A =médicament, B =niveau de santé, C =âge

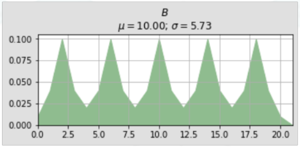
- P(B|do(A)) :



Doing $A = 10$



Doing $A = 20$

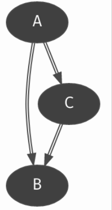


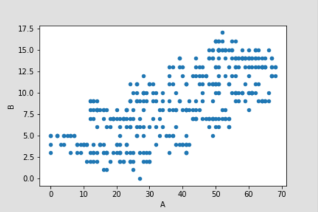
Doing $A = 30$

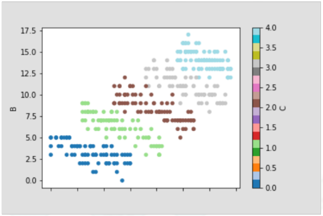
Observer (forcer) C apporte une information pertinente.

Retour sur Simpson

- Le modèle causal :

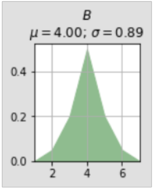




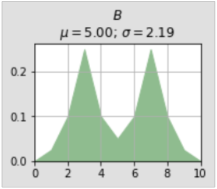


Sémantique : A =médicament, B =niveau de santé, C =effet secondaire

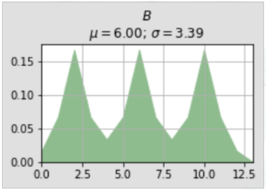
- P(B|do(A)) :



Doing $A = 10$



Doing $A = 20$

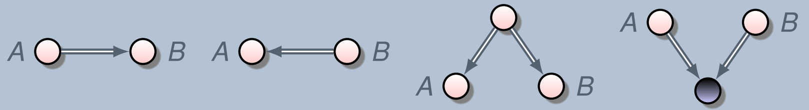


Doing $A = 30$

Observer C apporte une information non pertinente.

- ⚠ Impossible de construire un modèle causal parfait à partir des données uniquement !

A et B sont corrélés donc ... ?

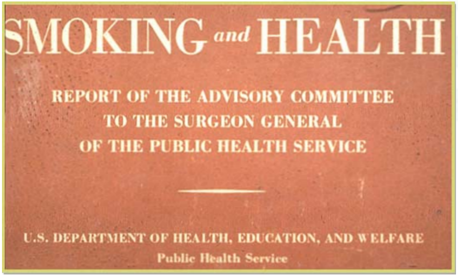


- Par contre, les algorithmes constraint-based fournissent un graphe partiellement orienté *raisonnablement causal*.
- Ces algorithmes peuvent également repérer des causes latentes.

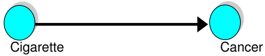
- ⚠ Nécessité d'informations supplémentaires (hors data).
- ⚠ Variables latentes

Calculs dans des modèles causaux non complets

- La présence de variables latentes rend caduque la méthode du BN mutilé pour calculer les interventions.
- Le modèle observationnel n'est pas le modèle causal
- La question : est-ce que le modèle observationnel est suffisant pour faire des calculs causaux ?

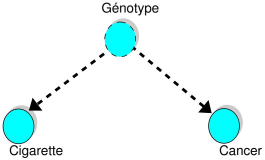


- Surgeon General (1964) : La très forte corrélation entre Cigarette et Cancer (du poumon) est causale !



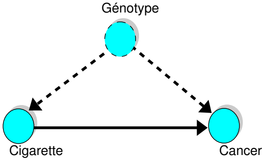
$$P(Ca \mid do(Ci)) = P(Ca \mid Ci)$$

- L'industrie du tabac rétorque qu'il est tout à fait possible qu'un autre modèle explique cette corrélation.



$$P(Ca \mid do(Ci)) = P(Ca)$$

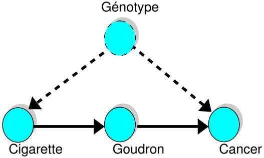
- [Pearl, 95] Dans un effort diplomatique, les 2 parties concèdent que l'autre modèle doit être un peu vrai.



$P(Ca \mid do(Ci))$ incalculable !
on ne sait pas discerner l'influence causale des 2 causes à partir des observations.

Tabac et cancer (2)

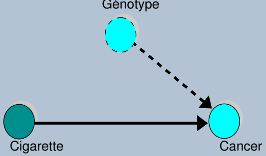
- Avant d'abandonner, on essaye l'introduction de facteurs auxiliaires : le tabac provoque le cancer à cause du goudron déposé dans les poumons.



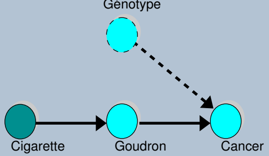
$P(Ca \mid do(Ci))$ calculable !

Énigmatique, non ?

a. Calcul impossible dans le sous-modèle



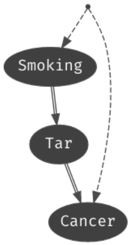
b. Calcul possible dans le sous-modèle



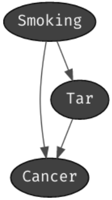
$$P(Can|do(Smo)) = \sum_{t} P(tar|Smo) \sum_{smo'} P(smo')P(Can|smo', tar)$$

Definition (Effet causal identifiable)

Un effet causal est identifiable dans un modèle causal G si $P(y \mid do(x))$ peut être calculé à partir de la loi jointe (positive) des variables observables dans G.



$$P(C|do(S)) = \sum_t P(t|S) \sum_{s'} P(s')P(C|s', t)$$



- Un effet causal identifiable est calculable uniquement avec des données "passives" et avec un graphe spécifiant qualitativement les causalités.
- La positivité de la loi jointe assure que tous les contextes sont observables.

Une algèbre interventionniste

- Il existe une algèbre du conditionnement classique : **règle de Bayes**

Quelle est la probabilité qu'il pleuve, si (on voit que) l'herbe est mouillée ?

$$P(\text{Pluie} \mid \text{Humide}) = P(\text{Humide} \mid \text{Pluie}) \cdot \frac{P(\text{Pluie})}{P(\text{Humide})}$$

etc.

- Il nous faut une algèbre du conditionnement par intervention

Le but de cet algèbre serait de fournir les outils nécessaires pour obtenir une formule de probabilité classique (sans *do*) à partir d'une formulation causale.

Quelle est la probabilité qu'il pleuve, si on mouille l'herbe ?

$$P(\text{Pluie} \mid do(\text{Humide})) = P(\text{Pluie})$$

Règles du calcul des conditionnements par intervention

Soit $G_{\overline{X}}$ et $G_{\underline{X}}$ les graphes G moins les arcs resp. arrivant en et issus de X.
$\forall X, Y, Z, W$ sous-ensembles disjoints de variables de G,

Règle 1 : élimination d'observations

$$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}} \Rightarrow P(y \mid do(x), z, w) = P(y \mid do(x), w)$$

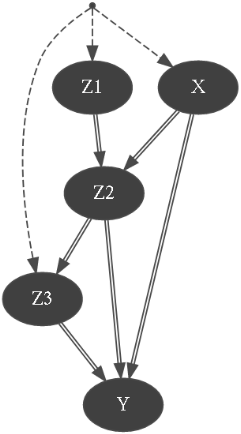
Règle 2 : échange interventions/observations

$$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}\underline{Z}}} \Rightarrow P(y \mid do(x), do(z), w) = P(y \mid do(x), z, w)$$

Règle 3 : suppression d'interventions

$$(Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X},\overline{Z_W}}} \Rightarrow P(y \mid do(x), do(z), w) = P(y \mid do(x), w)$$

NB- Z_W est l'ensemble des nœuds de Z qui n'ont pas d'ancêtres dans W dans $G_{\overline{X}}$...



$$P(Y \mid \hookrightarrow X) = \sum_{Z1,Z2,Z3} P\left(Y \mid X, Z2, Z3\right) \cdot P\left(Z2 \mid X, Z1\right) \cdot \left(\sum_{X'} P\left(Z1\right) \cdot P\left(X' \mid Z1\right) \cdot P\left(Z3 \mid X', Z1, Z2\right) \right)$$

	Y	
X	0	1
0	0.5236	0.4764
1	0.4866	0.5134

HomeWork

Sur la causalité