cours 3
Apprentissage de structure de RB

Aix*Marseille université
Initiative d'excellence

Master SID — Raisonnement dans l'incertain

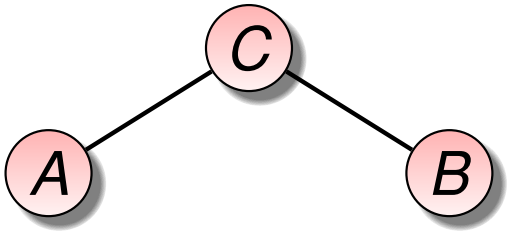
©CG(2026)

1. Indépendances et graphe

Rappel sur les réseaux bayésiens

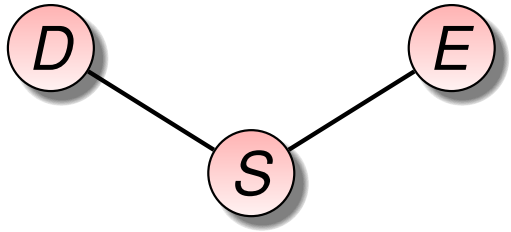
non($A \perp\!\!\!\perp B$)

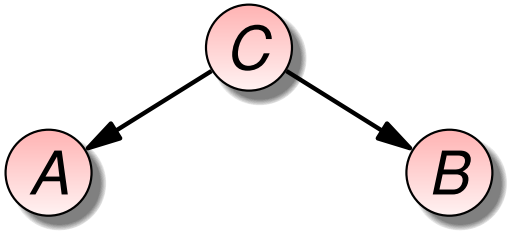
$A \perp\!\!\!\perp B \mid C$

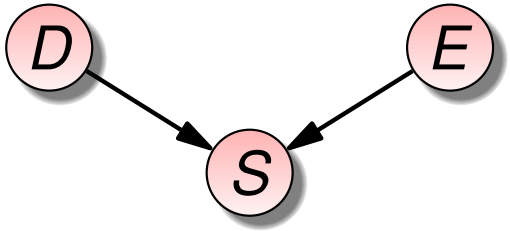


$D \perp\!\!\!\perp E$

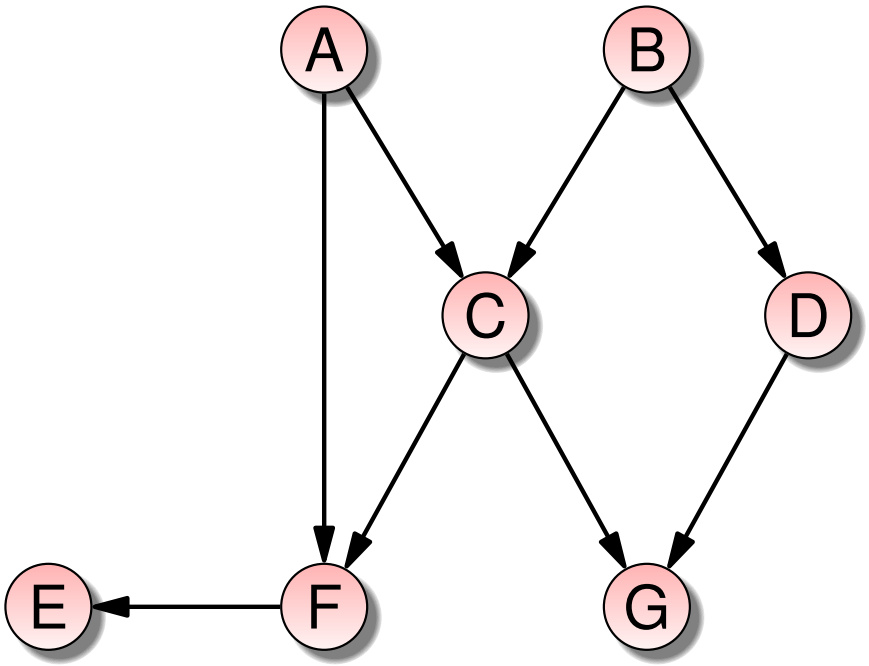
non($D \perp\!\!\!\perp E \mid S$)







Indépendances et modèle graphique



- $A \perp\!\!\!\perp G$?
- $E \perp\!\!\!\perp G$?
- $A \perp\!\!\!\perp D$?
- $A \perp\!\!\!\perp D | G$?

$\Longrightarrow$ Raisonnement sur la partie graphique du modèle

$\Longrightarrow$ Vérifications d'indépendances conditionnelles sans connaître les valeurs des probabilités !

d-séparation

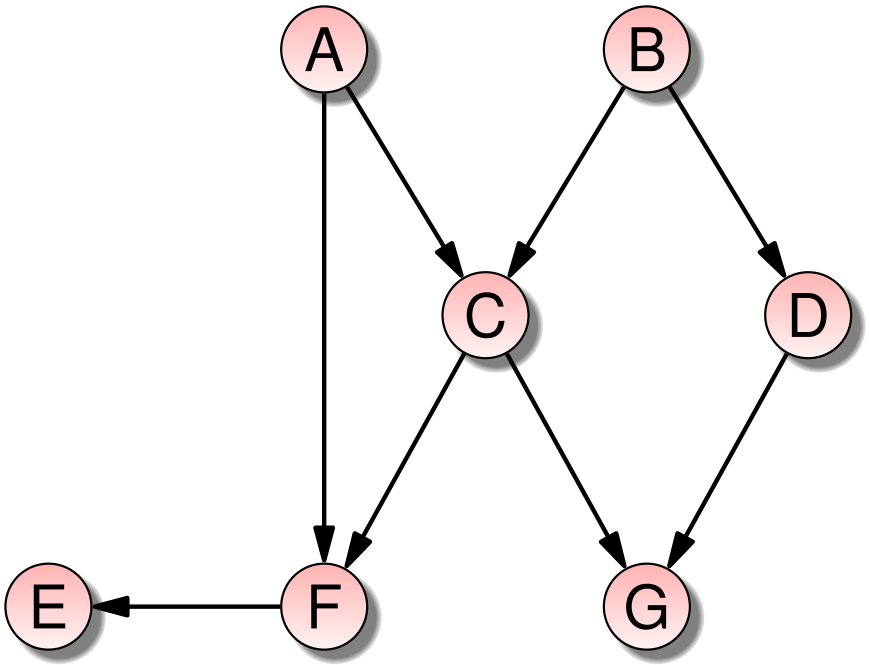
Chaîne $\langle X_1, \ldots, X_n \rangle$

- ensemble de nœuds $\{X_1, \ldots, X_n\}$
- pour tout $i \in \{1, \ldots, n-1\}$, le graphe contient l'arc $X_i \rightarrow X_{i+1}$ ou $X_{i+1} \rightarrow X_i$

Chaîne $\langle X_1, \ldots, X_n \rangle$ *bloquée par un ensemble* $\mathbf{Z}$

- Bloquée si et seulement si $\exists\, i \in \{2, \ldots, n-1\}$ tel que l'une des 2 propriétés ci-dessous est vérifiée :
  1. (X_{i-1}, X_i, X_{i+1}) est une V-structure : $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$, et ni X_i ni ses descendants ne sont dans $\mathbf{Z}$
  2. (X_{i-1}, X_i, X_{i+1}) n'est pas une V-structure et $X_i \in \mathbf{Z}$

d-séparation – bis



- Chaîne $\langle D, B, C, A \rangle$ bloquée par $\emptyset$?
- Chaîne $\langle D, B, C, A \rangle$ bloquée par $\{E\}$?
- Chaîne $\langle D, B, C, F, A \rangle$ bloquée par $\{E\}$?
- Chaîne $\langle D, G, C, A \rangle$ bloquée par $\{E\}$?
- Chaîne $\langle D, G, C, F, A \rangle$ bloquée par $\{E\}$?

d-séparation – ter

d-séparation

- A, B, C trois variables aléatoires ou groupes de variables disjoints
- A est d-séparé de B par C si toute chaîne entre A et B est bloquée par C.

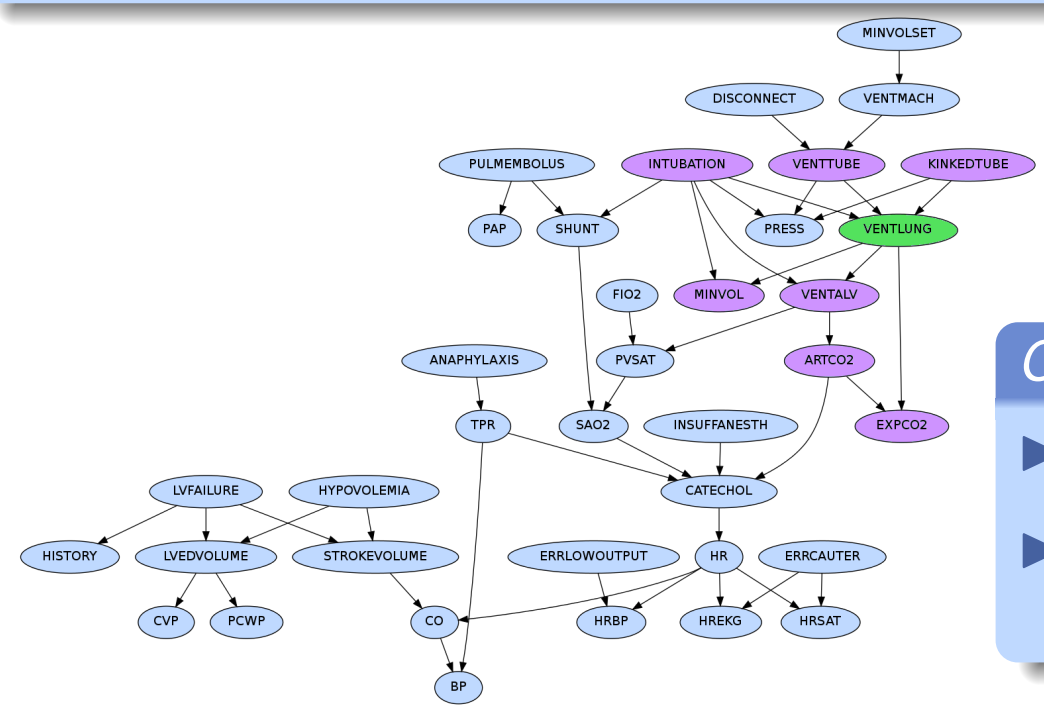
Théorème

- A, B, C trois variables aléatoires ou groupes de variables disjoints
- A est d-séparé de B par $C \Longrightarrow A \perp\!\!\!\perp B | C$

Conséquence de la *d*-séparation

Couverture de Markov d'un nœud X_i

- $\mathcal{X} = \{X_1, \ldots, X_n\}$: nœuds du réseau bayésien
- $MB(X_i)$ = couverture de Markov de X_i
  $= \text{ensemble de nœuds t.q. } X_i \perp\!\!\!\perp (\mathcal{X} \setminus (MB(X_i) \cup \{X_i\})) | MB(X_i)$
- $MB(X_i) = \{\text{parents de } X_i\} \cup \{\text{enfants de } X_i\} \cup \{\text{parents des enfants de } X_i\} \setminus \{X_i\}$



- X_i : nœud en vert
- $MB(X_i)$: nœuds en violet

Conséquences

- $P(X_i | \mathcal{X} \setminus \{X_i\}) = P(X_i | MB(X_i))$
- Classification : observer $MB(X_i)$ suffit pour « classifier » X_i

Séparation, indépendance et maps

- Notation : $\langle A \perp_{\mathcal{G}} B|C\rangle$: A d-séparé de B par C dans graphe $\mathcal{G}$
  $A \perp\!\!\!\perp_P B|C$: selon la distribution de probabilité P,
  A indépendant de B conditionnellement à C

Définitions

- Soit $\mathcal{G}$ un graphe et P une distribution de probabilité
- $\mathcal{G}$ I-map (independence map) de P ssi $\langle A \perp_{\mathcal{G}} B|C\rangle \Longrightarrow A \perp\!\!\!\perp_P B|C$
  Interprétation : absence d'arc dans $\mathcal{G} \Longrightarrow$ indépendance dans P
- $\mathcal{G}$ D-map (dependence map) de P ssi $A \perp\!\!\!\perp_P B|C \Longrightarrow \langle A \perp_{\mathcal{G}} B|C\rangle$
  Interprétation : présence d'arc dans $\mathcal{G} \Longrightarrow$ dépendance
- $\mathcal{G}$ P-map (perfect map) de P ssi $A \perp\!\!\!\perp_P B|C \Longleftrightarrow \langle A \perp_{\mathcal{G}} B|C\rangle$
  Interprétation : présence d'arc dans $\mathcal{G}$ ssi dépendance
  absence d'arc dans $\mathcal{G}$ ssi indépendance

Réseaux bayésiens et maps

Définition d'un réseau bayésien

- $\mathcal{G}$ un graphe, muni de la d-séparation
- $\mathcal{G}$ est une I-map d'une distribution P

$\Longrightarrow \mathcal{G}$ est une structure de réseau bayésien pour P

$\Longrightarrow P$ est factorisable selon le graphe $\mathcal{G}$

Propriété de Markov globale (PMG)

$\mathcal{G}$ vérifie la PMG pour P ssi $\mathcal{G}$ est une I-map pour P.

$\Longrightarrow$ les réseaux bayésiens vérifient la PMG

Propriété de Markov locale (PML)

$\mathcal{G}$ vérifie la PML pour P ssi pour toute variable X :

$$X \perp\!\!\!\perp_P \text{non descendants}(X) \mid \text{parents}(X).$$

$\Longrightarrow$ les réseaux bayésiens vérifient la PML

PMG $\Longleftrightarrow$ PML

2 Apprentissage de structure

Apprentissage de structure – une 1ère idée

- Objectif : déterminer la structure $\mathcal{G}$ à partir de données **D**
- Rappel : $\mathcal{G}$ P-map (perfect map) de P ssi

$$A \perp\!\!\!\perp_P B | C \iff \langle A \perp_{\mathcal{G}} B | C \rangle$$

- Algorithme « naïf » :
  - créer toutes les structures $\mathcal{G}$ possibles
  - $\forall\, \mathcal{G}$, calculer tous les triplets (A, B, C) t.q. $\langle A \perp_{\mathcal{G}} B | C \rangle$
  - tester si $A \perp\!\!\!\perp_P B | C$ (par exemple, test du χ^2 en utilisant **D**)
  - si vrai pour tout triplet (A, B, C), structure $\mathcal{G}$ trouvée

Problèmes de l'algorithme naïf (1/3)

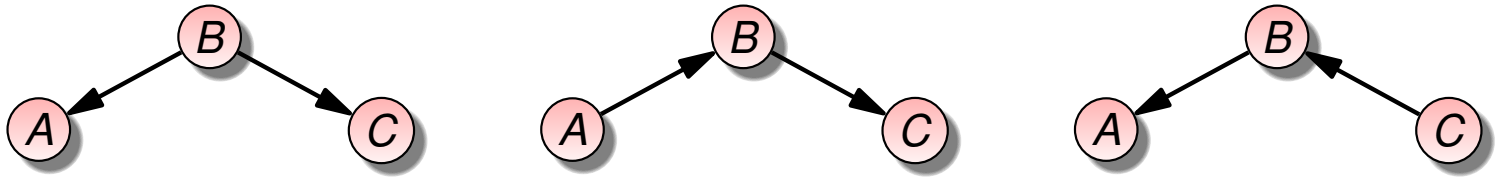
Théorème – Robinson(1977)

Le nombre de structures $\mathcal{G}$ à n nœuds est super-exponentiel :

$$\#(n) = \begin{cases} 1 & \text{si } n \leq 1 \\ \sum_{i=1}^{n} (-1)^{i+1} 2^{i(n-1)} C_n^i \times \#(n-1) & \text{si } n > 1 \end{cases}$$

$\Longrightarrow$ on ne peut pas tester toutes les structures

3 réseaux bayésiens équivalents (mêmes indep.) :



$\Longrightarrow$ ne les compter que pour 1 seul réseau !

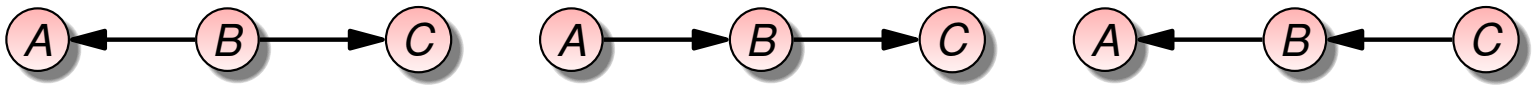
$\Longrightarrow$ Appliquer l'algorithme dans l'espace des classes d'équivalence de Markov

Classe d'équivalence de Markov

Définition : équivalence de Markov

- $\mathcal{G}_1, \mathcal{G}_2$ deux structures de RB contenant les mêmes nœuds/variables aléatoires
- $\mathcal{G}_1, \mathcal{G}_2 \in$ même classe d'équivalence de Markov ssi, pour tous ensembles de variables disjoints A, B, C :

$$\langle A \perp_{\mathcal{G}_1} B | C \rangle \iff \langle A \perp_{\mathcal{G}_2} B | C \rangle$$



Définitions

- $\mathcal{G}$ structure de RB. Squelette de $\mathcal{G}$ obtenu en remplaçant les arcs $X \to Y$ par des arêtes $X - Y$.
- $\{$nœuds $X, Y, Z\}$ = v-structure $\iff$ dans $\mathcal{G}$, $\exists\, X \to Y \leftarrow Z$ et $\nexists X \to Z$ et $\nexists Z \to X$

Théorème – Verma et Pearl (1991)

$\mathcal{G}_1, \mathcal{G}_2 \in$ même classe d'équivalence de Markov ssi même squelette et mêmes v-structures.

Problèmes de l'algorithme naïf (2/3)

Appliquer l'algo dans l'espace des classes d'équivalence

Propriété expérimentale – Gillispie et Perlman (2002)

$$\text{Ratio } \frac{\text{taille de l'espace des DAG}}{\text{taille de l'espace des classes d'équivalence}} \approx 3,7$$

$\Longrightarrow$ Pas d'avantage à utiliser les classes d'équivalence

Théorème – Chickering, Heckerman, Meek (2004)

L'apprentissage de structure de RB est NP-hard.

- 2 alternatives :
  - Apprentissage « exact » pour des « petits » RB
  - Apprentissage « approché » $\Longrightarrow$ heuristiques

Problèmes de l'algorithme naïf (3/3)

- 2ème problème : tester si $A \perp\!\!\!\perp_P B | C$ impossible si

  D petite ou $\Omega_{A \cup B \cup C}$ de grande taille

$A \perp_{\mathcal{G}} B|C \iff (X \perp_{\mathcal{G}} Y|C \; \forall X \in A, \forall Y \in B)$

- Perfect map $\implies A \perp\!\!\!\perp_P B|C \iff (X \perp\!\!\!\perp_P Y|C \;\; \forall X \in A, \forall Y \in B)$

Tests d'indépendance

- Hypothèse (**DAG-faithfulness**) : P représentable par une perfect map $\mathcal{G}$
- Ne tester que l'indépendance conditionnelle de couples de variables

Théorème d'Hammersley-clifford

P distribution strictement positive $\implies$ P représentable par une perfect map.

Relations déterministes entre variables $\implies$ P non strictement positive

Apprentissage fondé sur les contraintes

Idée générale de l'apprentissage sous contraintes

1. Apprendre le squelette via des tests d'indépendance
2a. Orienter les v-structures
2b. Propager ces orientations afin qu'elles ne créent pas de nouvelle v-structure
3. Orienter le reste des arêtes sans créer de nouvelle v-structure

- Algorithmes classiques :
  - Inductive causation (IC) – Verma et Pearl (1990)
  - PC – Spirtes, Glymour et Scheines (2000)

Bibliographie

- Chickering D., Heckerman D. et Meek C. (2004) « Large-Sample Learning of Bayesian Networks is NP-Hard », Journal of Machine Learning Research, 5 :1287–1330
- Gillispie S. et Perlman M. (2002) « The size distribution for Markov equivalence classes of acyclic digraph models », Artificial Intelligence, 141 :137–155
- Robinson, R. (1977) « Counting unlabeled acyclic digraphs », Combinatorial Mathematics V, 622 : :28–43
- Spirtes E., Glymour C. et Scheines R. (2000) Causation, Prediction and Search, 2nd edition, Springer-Verlag
- Verma T. et Pearl J. (1990) « Equivalence and synthesis of causal models », Proceedings of UAI, 220–227