

Modélisation Stochastique et Approche Bayésienne

Introduction et principes généraux

Roland Donat

Université de Bretagne Sud

ENSIBS - Spécialité Cyber Data

<https://roland-donat.github.io/cours-rb/ensibs/>

Objectifs pédagogiques

- Se sensibiliser et pratiquer la modélisation probabiliste
- Replacer le développement des techniques bayésiennes dans l'Histoire scientifique
- Comprendre le principe de raisonnement bayésien
- Introduire le concept de réseau bayésien
- Savoir construire un réseau bayésien pour un problème donné

Plan de la présentation

- 1 Introduction
- 2 Rappels de probabilité
- 3 Approche bayésienne
- 4 Les réseaux bayésiens
- 5 Concepts fondamentaux
- 6 Conclusion, perspectives et outils
- 7 Annexes

Plan

- 1 Introduction
 - Objet du cours
 - Historique
- 2 Rappels de probabilité
- 3 Approche bayésienne
- 4 Les réseaux bayésiens
- 5 Concepts fondamentaux
- 6 Conclusion, perspectives et outils
- 7 Annexes

Quelques définitions avant de démarrer

Modélisation stochastique

- Modélisation : Action de représenter (en simplifiée) une entité, un phénomène, le monde, etc...
 - Stochastique : Synonyme du mot "aléatoire"
- ⇒ Représentation d'éléments possédant des comportements aléatoires

Historique

Comment tout a commencé...



Révérénd Thomas Bayes (1702–1761)

- 1763 : *An Essay towards solving a Problem in the Doctrine of Chances*
- ⇒ **Notion de probabilité conditionnelle** : *If there be two subsequent events, the probability of the second b/N and the probability of both together P/N , and it being first discovered that the second event has also happened, the probability I am right [i.e., the conditional probability of the first event being true given that the second has happened] is P/b .*
- ⇒ **Théorème de Bayes**

Historique

Comment tout a commencé...

Pierre-Simon Laplace (1749–1827)

- 1774 : Mémoire sur la Probabilité des Causes par les Événements
- 1812 : Reformulation et clarification des travaux de Bayes



Plan

- 1 Introduction
- 2 Rappels de probabilité
- 3 Approche bayésienne
- 4 Les réseaux bayésiens
- 5 Concepts fondamentaux
- 6 Conclusion, perspectives et outils
- 7 Annexes

Rappels de probabilité

Variable aléatoire discrète et finie

Variable aléatoire discrète et finie

- Intuition : Une variable aléatoire (v.a.) est une "variable" ^{α} dont les valeurs possibles sont les résultats d'un phénomène aléatoire.
- Un peu plus formellement, une variable aléatoire (v.a.) X est définie sur un ensemble \mathcal{X} représentant les résultats possibles d'une expérience aléatoire
- Si l'ensemble \mathcal{X} est discret et fini, on parle alors de variable aléatoire discrète et finie

α . En fait, une v.a. est une application entre deux espaces mesurables mais n'en parlons plus...

Exemples de variables discrètes et finies

- X : "intensité d'un séisme", $\mathcal{X} = \{[0, 3[, [3, 7[, [7, \infty[\}$
- Y : "l'alarme de ma voiture sonnera t-elle cette nuit?",
 $\mathcal{Y} = \{\text{non}, \text{oui}\}$

Rappels de probabilité

Variable aléatoire discrète et finie

Variable aléatoire discrète et finie

- Intuition : Une variable aléatoire (v.a.) est une "variable" ^{α} dont les valeurs possibles sont les résultats d'un phénomène aléatoire.
- Un peu plus formellement, une variable aléatoire (v.a.) X est définie sur un ensemble \mathcal{X} représentant les résultats possibles d'une expérience aléatoire
- Si l'ensemble \mathcal{X} est discret et fini, on parle alors de variable aléatoire discrète et finie

α . En fait, une v.a. est une application entre deux espaces mesurables mais n'en parlons plus...

Exemples de variables discrètes et finies

- X : "intensité d'un séisme", $\mathcal{X} = \{[0, 3[, [3, 7[, [7, \infty[\}$
- Y : "l'alarme de ma voiture sonnera t-elle cette nuit?",
 $\mathcal{Y} = \{\text{non}, \text{oui}\}$

Rappels de probabilité

Loi de probabilité discrète

Loi de probabilité discrète

- Soit une v.a. X définie sur le domaine discret et fini $\mathcal{X} = \{x_1, \dots, x_N\}$
- L'application P de \mathcal{X} dans $[0, 1]$ définit la loi de probabilité de la v.a. X si

$$\forall x_i \in \mathcal{X}, P(X = x_i) \in [0, 1] \text{ et } \sum_{n=1}^N P(X = x_n) = 1$$

⇒ On note alors $P(X)$ la loi de probabilité de la v.a. discrète et finie X à valeurs dans \mathcal{X}

- Interprétation : Une loi de probabilité permet de quantifier le caractère aléatoire d'une v.a. (i.e. d'un phénomène aléatoire)

Exemple de loi sur l'intensité d'un séisme représentée par la v.a. X

$P(X)$		
$[0, 3[$	$[3, 7[$	$[7, \infty[$
0.9999	0.00009	0.00001

Rappels de probabilité

Loi jointe - Définition

Loi jointe

- Soit X_1, \dots, X_N une suite de v.a. discrètes à valeurs dans les ensembles $\mathcal{X}_1, \dots, \mathcal{X}_N$ respectivement
- La loi jointe des v.a. X_1, \dots, X_N est caractérisée par les probabilités :

$$P(\{X_1 = x_1\} \text{ et } \{X_2 = x_2\} \text{ et } \dots \text{ et } \{X_N = x_N\})$$

pour tous $x_1 \in \mathcal{X}_1, \dots, x_N \in \mathcal{X}_N$.

- Hypothèse : Les v.a. sont considérées comme a priori interdépendantes
- Remarque : La loi jointe contient toute l'information sur le phénomène aléatoire associé aux v.a. X_1, \dots, X_N

Exemple de loi jointe sur l'intensité d'un séisme X et l'occurrence d'une alarme Y

$P(X, Y)$					
$[0, 3[$ non	$[3, 7[$ non	$[7, \infty[$ non	$[0, 3[$ oui	$[3, 7[$ oui	$[7, \infty[$ oui
0.9989001	$2.7 \cdot 10^{-5}$	$5 \cdot 10^{-7}$	$9.999 \cdot 10^{-4}$	$6.3 \cdot 10^{-5}$	$9.5 \cdot 10^{-6}$

Rappels de probabilité

Loi jointe - Notations

Conventions et simplifications d'écritures

Il est courant de simplifier l'écriture des expressions probabilistes pour gagner en lisibilité :

- Le terme "et" dans l'expression d'une probabilité jointe est remplacé par une virgule ou un point virgule
- Omission des accolades dans les expressions sauf si ces dernières permettent d'améliorer la lisibilité

$$\Rightarrow P(\{X_1 = x_1\} \text{ et } \dots \text{ et } \{X_N = x_N\}) = P(X_1 = x_1, \dots, X_N = x_N)$$

Attention

- La quantité $P(X_1 = x_1, \dots, X_N = x_N)$ avec $x_1 \in \mathcal{X}_1, \dots, x_N \in \mathcal{X}_N$ est une probabilité, c'est à dire un réel entre 0 et 1
- L'objet $P(X_1, \dots, X_N)$ est une loi de probabilité qui peut être représentée sous forme de tableau dans le cas où les v.a. X_1, \dots, X_N sont discrètes et finies

Rappels de probabilité

Loi jointe - Notations

Conventions et simplifications d'écritures

Il est courant de simplifier l'écriture des expressions probabilistes pour gagner en lisibilité :

- Le terme "et" dans l'expression d'une probabilité jointe est remplacé par une virgule ou un point virgule
- Omission des accolades dans les expressions sauf si ces dernières permettent d'améliorer la lisibilité

$$\Rightarrow P(\{X_1 = x_1\} \text{ et } \dots \text{ et } \{X_N = x_N\}) = P(X_1 = x_1, \dots, X_N = x_N)$$

Attention

- La quantité $P(X_1 = x_1, \dots, X_N = x_N)$ avec $x_1 \in \mathcal{X}_1, \dots, x_N \in \mathcal{X}_N$ est une probabilité, c'est à dire un réel entre 0 et 1
- L'objet $P(X_1, \dots, X_N)$ est une loi de probabilité qui peut être représentée sous forme de tableau dans le cas où les v.a. X_1, \dots, X_N sont discrètes et finies

Rappels de probabilité

Probabilité conditionnelle

Probabilité conditionnelle

- Lorsque deux phénomènes aléatoires représentés par des v.a. X et Y sont corrélés, les valeurs prises par Y influent sur les valeurs prises par X et inversement
- La loi $P(Y|X)$ représente l'information sur la loi de Y conditionnellement aux valeurs de X
- Par définition : $P(Y|X) = \frac{P(X, Y)}{P(X)}$

Exemple d'une loi d'occurrence d'une alarme Y conditionnellement à l'intensité d'un séisme X

X	$P(Y X)$	
	Y	
	non	oui
$[0, 3[$	0.999	0.001
$[3, 7[$	0.3	0.7
$[7, \infty[$	0.05	0.95

Rappels de probabilité

Indépendance

Indépendance

- Les v.a. X et Y sont indépendantes si et seulement si :

$$P(X|Y) = P(X) \iff P(Y|X) = P(Y) \iff P(X, Y) = P(X) \times P(Y)$$

- Notation : $X \perp\!\!\!\perp Y$

Exemple

- X : résultat pile ou face, $\mathcal{X} = \{\text{pile}, \text{face}\}$
- Y : valeur pièce, $\mathcal{Y} = \{0.50, 1, 2\}$

$P(X Y)$		
Y	X	
	pile	face
50 centimes	0.5	0.5
1 euros	0.5	0.5
2 euros	0.5	0.5

Rappels de probabilité

Indépendance

Indépendance conditionnelle

- Les v.a. X et Y sont indépendantes conditionnellement à Z si et seulement si :

$$\begin{aligned} P(X|Y, Z) = P(X|Z) &\iff P(Y|X, Z) = P(Y|Z) \\ &\iff P(X, Y|Z) = P(X|Z) \times P(Y|Z) \end{aligned}$$

- Notation : $X \perp\!\!\!\perp Y|Z$

Exemple

- X : historique accident, $\mathcal{X} = \{\text{aucun}, \text{accident(s)}\}$
- Y : sexe du conducteur, $\mathcal{Y} = \{\text{femme}, \text{homme}\}$
- Z : nombre de points au permis, $\mathcal{Z} = \{< 5, \geq 5\}$

		$P(X Y, Z)$	
Y	Z	X	
		aucun	accident(s)
femme	< 5	0.2	0.8
femme	≥ 5	0.45	0.55
homme	< 5	0.2	0.8
homme	≥ 5	0.45	0.55

Rappels de probabilité

Marginalisation

Marginalisation

- Soient X et Y deux v.a. discrètes à valeurs dans \mathcal{X} et \mathcal{Y} respectivement. L'opération de marginalisation sur la v.a. X (ou sommation sur la v.a. Y) est définie par :

$$P(X = x) = \sum_{y \in \mathcal{Y}} P(X = x, Y = y), \forall x \in \mathcal{X}$$

- Notation : $P(X) = \sum_Y P(X, Y)$.

Exemple

X à valeurs dans $\mathcal{X} = \{x_1, x_2, x_3\}$ et Y à valeurs dans $\mathcal{Y} = \{y_1, y_2\}$

$P(X, Y)$					
x_1	x_2	x_3	x_1	x_2	x_3
y_1	y_1	y_1	y_2	y_2	y_2
1/4	1/12	1/4	0	1/6	1/4

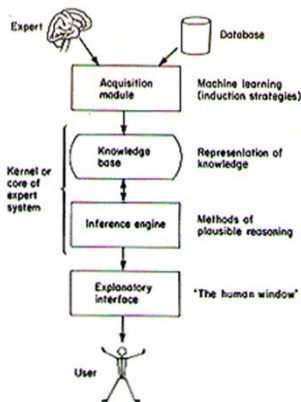
$P(Y) = \sum_X P(X, Y)$	
y_1	y_2
7/12	5/12

Plan

- 1 Introduction
- 2 Rappels de probabilité
- 3 Approche bayésienne
 - Retour sur l'historique
- 4 Les réseaux bayésiens
- 5 Concepts fondamentaux
- 6 Conclusion, perspectives et outils
- 7 Annexes

Historique

Deux siècles plus tard...

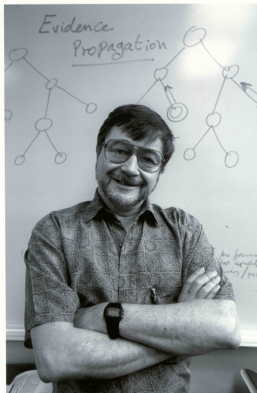


L'ère des systèmes experts (1970-1990)

- Formalisation des connaissances sur un système sous forme de règles déterministes
Exemple : SI $X = \text{VRAI}$ ET $Y = \text{FAUX}$ ALORS $Z = \text{VRAI}$
 - Moteur d'inférence reposant sur la logique booléenne
- ⇒ Dédution d'informations à partir d'une base de règles

Historique

Deux siècles plus tard...



Judea Pearl (1936-) : Père des réseaux bayésiens

- 1982 : *Reverend Bayes on inference engines : A distributed hierarchical approach*
 $P(X = \text{Ok}) = 0.3$ et
 $P(Z = \text{Défaillant}) = 0.2$ Alors
 $P(Y = \text{Dégradé}) = ?$
- (Pearl 1988) : *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*
- 2004 : MIT's Technology Review place l'apprentissage des réseaux bayésiens en quatrième position des dix technologies émergentes qui vont changer le monde
- 2011 : Judea Pearl reçoit le Prix Turing

Approche bayésienne

Principes

Principes de l'approche bayésienne

- L'approche bayésienne repose sur l'utilisation du langage des probabilités pour réaliser des raisonnements
- Cette approche de raisonnement généralise la logique booléenne
- Chaque raisonnement se matérialise par une suite de calculs probabilistes
- La conclusion d'un raisonnement bayésien est donnée sous la forme d'une loi de probabilité

Approche bayésienne

Exemple de raisonnement

Alarme et vol de voiture

- Situation : J'ai été réveillé par l'alarme de ma voiture en pleine nuit
- Question : Ma voiture est-elle en train de se faire voler?

Raisonnement de type bayésien

- 1 Modélisation : on liste les phénomènes qui pourraient expliquer l'alarme (e.g. fausse alarme, tremblement de terre, criminalité, etc...) et on tente d'établir leur probabilité
- 2 Inférence : on réalise des calculs de probabilité pour obtenir la réponse qui nous intéresse, i.e. ma voiture est-elle en train de se faire voler?
- 3 Décision : si, compte tenu du contexte, la probabilité d'un vol est inférieur à un seuil donné, je me recouche, sinon j'appelle la police

Approche bayésienne

Théorème de Bayes

Théorème de Bayes

- Le théorème (ou formule) de Bayes permet d'exprimer la loi conditionnelle d'un phénomène X sachant un phénomène Y en fonction de la loi du phénomène Y sachant X et de la loi marginale du phénomène X .
- Soient X et Y deux v.a., on a :

$$P(X|Y) = \frac{P(Y|X)P(X)}{\sum_x P(Y|X)P(X)}$$

- Dans le contexte de cette formule, la loi $P(X)$ est souvent appelée loi a priori du phénomène X
- ⇒ Le théorème de Bayes permet d'effectuer des raisonnements probabilistes impliquant deux phénomènes aléatoires

Approche bayésienne

Formule de Bayes généralisée

Formule de Bayes généralisée

- Soit X_1, \dots, X_N une suite de v.a. discrètes. La loi jointe des v.a. admet la factorisation suivante :

$$P(X_1, \dots, X_N) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2) \dots P(X_N|X_1, \dots, X_{N-1})$$

- Pour $N = 2$, on retrouve la définition d'une loi de probabilité conditionnelle
- Cette formule donne une factorisation **toujours valide** de la loi jointe sous forme de lois conditionnelles (plus pratiques à manipuler)
- Toutefois, cette factorisation reste difficile à utiliser en pratique **mais les réseaux bayésiens vont nous sauver!**

Plan

- 1 Introduction
- 2 Rappels de probabilité
- 3 Approche bayésienne
- 4 **Les réseaux bayésiens**
 - Les réseaux bayésiens en trois points
 - Domaines d'application
 - Définition formelle
 - Principe
 - Exemple
- 5 Concepts fondamentaux
- 6 Conclusion, perspectives et outils
- 7 Annexes

Les réseaux bayésiens en trois points

Qu'est ce que c'est?

- Les réseaux bayésiens désignent un formalisme de modélisation graphique
- Ce formalisme permet de représenter des connaissances incertaines et d'effectuer des raisonnements probabilistes sur ces connaissances

Sur quoi ça repose?

- Théorie des graphes (aspect qualitatif)
- Théorie des probabilités (aspect quantitatif)

Pourquoi est-ce intéressant?

- Formalisme graphique intuitif (facilite la communication)
- Puissance de modélisation

Domaines d'application

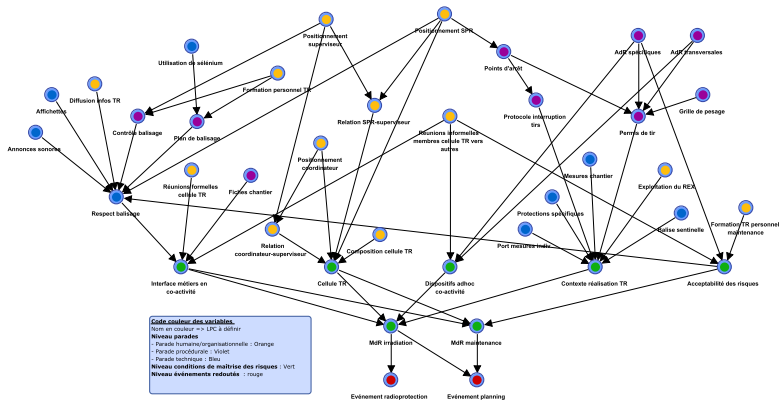
Quelques champs d'applications

- Intelligence artificielle
- Biologie, Médecine
- Marketing, Finance
- Sûreté de fonctionnement

Types d'analyses possibles

- Analyse prévisionnelle : Expliquer un phénomène par rapport à son contexte
 - Diagnostic : Comprendre le contexte/les causes associé à l'occurrence d'un événement
 - Simulation : Étudier un système complexe en générant des scénarios
- ⇒ Applications à l'aide à la décision en général

Un petit exemple...



Définition formelle

Definition (Réseau bayésien (RB))

Un RB, noté \mathcal{M} , représentant la distribution d'une suite de v.a. X_1, \dots, X_n est défini par :

- Un graphe orienté sans circuit \mathcal{G} où
 - ⇒ Les **nœuds** représentent les v.a.
 - ⇒ Les **arcs** indiquent des relations de dépendances
- Les **Lois de Probabilité Conditionnelles (LPC)** de chaque v.a. X_i sachant ses parents $pa(X_i)$, notées $P(X_i|pa(X_i))$

Définition formelle

Definition (Réseau bayésien (RB))

Un RB, noté \mathcal{M} , représentant la distribution d'une suite de v.a. X_1, \dots, X_n est défini par :

- Un graphe orienté sans circuit \mathcal{G} où
 - ⇒ Les **nœuds** représentent les v.a.
 - ⇒ Les **arcs** indiquent des relations de dépendances
 - Les **Lois de Probabilité Conditionnelles (LPC)** de chaque v.a. X_i sachant ses parents $pa(X_i)$, notées $P(X_i|pa(X_i))$

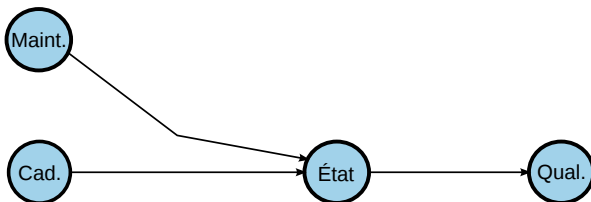


Définition formelle

Definition (Réseau bayésien (RB))

Un RB, noté \mathcal{M} , représentant la distribution d'une suite de v.a. X_1, \dots, X_n est défini par :

- Un graphe orienté sans circuit \mathcal{G} où
 - ⇒ Les **nœuds** représentent les v.a.
 - ⇒ Les **arcs** indiquent des relations de dépendances
- Les **Lois de Probabilité Conditionnelles (LPC)** de chaque v.a. X_i sachant ses parents $pa(X_i)$, notées $P(X_i|pa(X_i))$

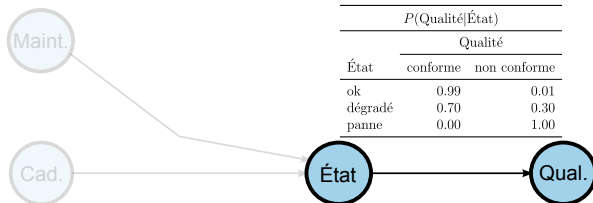


Définition formelle

Definition (Réseau bayésien (RB))

Un RB, noté \mathcal{M} , représentant la distribution d'une suite de v.a. X_1, \dots, X_n est défini par :

- Un graphe orienté sans circuit \mathcal{G} où
 - ⇒ Les **nœuds** représentent les v.a.
 - ⇒ Les **arcs** indiquent des relations de dépendances
- Les **Lois de Probabilité Conditionnelles (LPC)** de chaque v.a. X_i sachant ses parents $pa(X_i)$, notées $P(X_i|pa(X_i))$

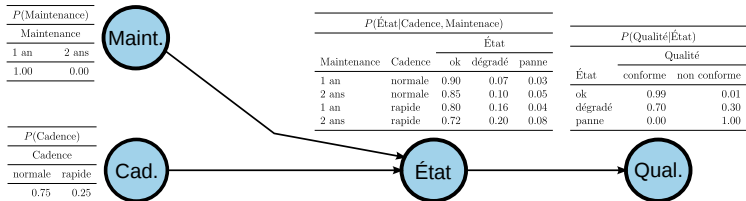


Définition formelle

Definition (Réseau bayésien (RB))

Un RB, noté \mathcal{M} , représentant la distribution d'une suite de v.a. X_1, \dots, X_n est défini par :

- Un graphe orienté sans circuit \mathcal{G} où
 - ⇒ Les **nœuds** représentent les v.a.
 - ⇒ Les **arcs** indiquent des relations de dépendances
- Les **Lois de Probabilité Conditionnelles (LPC)** de chaque v.a. X_i sachant ses parents $pa(X_i)$, notées $P(X_i|pa(X_i))$



Principe de modélisation avec un RB

Représentation qualitative à l'aide d'un graphe orienté sans circuit

- Représenter graphiquement la loi jointe d'un ensemble de variables aléatoires X_1, \dots, X_n , c-à-d. $P(X_1, \dots, X_n)$
- Spécifier les relations de dépendances locales de chaque variable pour en déduire le comportement global du système

Représentation quantitative à l'aide de LPC

- Spécifier la "force" des relations de dépendances locales à l'aide de probabilités conditionnelles
- Exploiter l'ensemble des relations d'indépendances entre les variables pour simplifier la loi jointe du processus

Exemple de l'étudiant

Contexte

- Un étudiant souhaite obtenir une lettre de recommandation

Comment modéliser les raisonnements intuitifs suivants ?

- L'écriture d'une lettre de recommandation dépend de la note de l'étudiant à l'examen
- La réussite d'un étudiant à un examen quelconque dépend de son niveau scolaire et de la difficulté de l'examen
- Les résultats au BAC dépendent du niveau scolaire

Questions

- Les résultats au BAC d'un étudiant peuvent-ils avoir une influence sur l'obtention d'une lettre de recommandation ?
- Connaître le niveau scolaire d'un étudiant peut-il nous aider à connaître la difficulté d'un examen ?
- Quelle est la meilleur stratégie pour obtenir une lettre ?

Exemple de l'étudiant

Contexte

- Un étudiant souhaite obtenir une lettre de recommandation

Comment modéliser les raisonnements intuitifs suivants ?

- L'écriture d'une lettre de recommandation dépend de la note de l'étudiant à l'examen
- La réussite d'un étudiant à un examen quelconque dépend de son niveau scolaire et de la difficulté de l'examen
- Les résultats au BAC dépendent du niveau scolaire

Questions

- Les résultats au BAC d'un étudiant peuvent-ils avoir une influence sur l'obtention d'une lettre de recommandation ?
- Connaître le niveau scolaire d'un étudiant peut-il nous aider à connaître la difficulté d'un examen ?
- Quelle est la meilleur stratégie pour obtenir une lettre ?

Exemple de l'étudiant

Contexte

- Un étudiant souhaite obtenir une lettre de recommandation

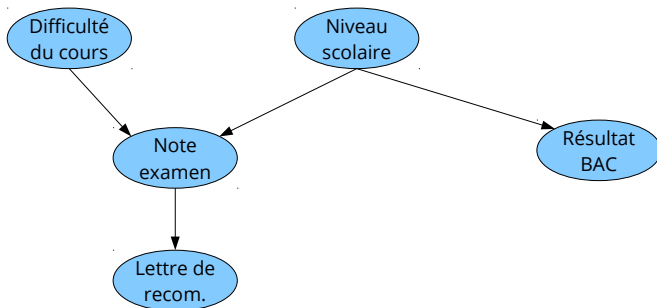
Comment modéliser les raisonnements intuitifs suivants ?

- L'écriture d'une lettre de recommandation dépend de la note de l'étudiant à l'examen
- La réussite d'un étudiant à un examen quelconque dépend de son niveau scolaire et de la difficulté de l'examen
- Les résultats au BAC dépendent du niveau scolaire

Questions

- Les résultats au BAC d'un étudiant peuvent-ils avoir une influence sur l'obtention d'une lettre de recommandation ?
- Connaître le niveau scolaire d'un étudiant peut-il nous aider à connaître la difficulté d'un examen ?
- Quelle est la meilleur stratégie pour obtenir une lettre ?

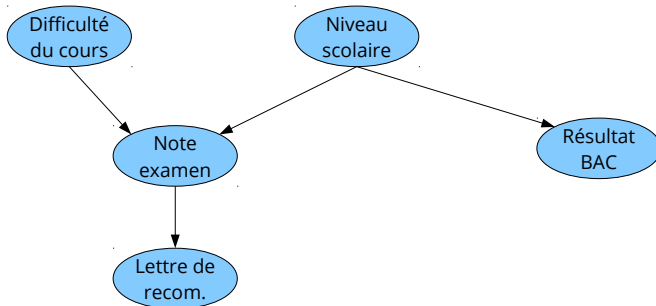
Exemple des étudiants



Exemple des étudiants

P(Difficulté)		
Difficulté		
<i>facile</i>	<i>moyenne</i>	<i>importante</i>
0,3	0,6	0,1

P(Niveau)	
Niveau	
<i>faible</i>	<i>fort</i>
0,2	0,8

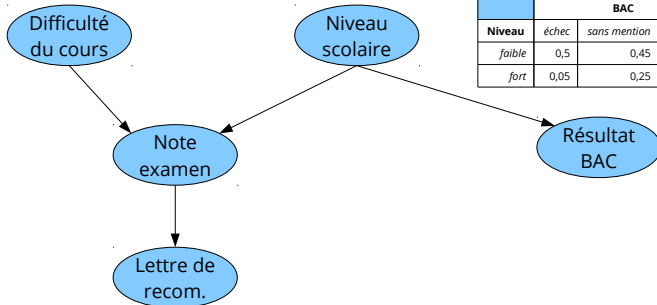


Exemple des étudiants

P(Difficulté)		
Difficulté		
<i>facile</i>	<i>moyenne</i>	<i>importante</i>
0,3	0,6	0,1

P(Niveau)	
Niveau	
<i>faible</i>	<i>fort</i>
0,2	0,8

P(BAC Niveau)			
	BAC		
Niveau	échec	sans mention	mention
<i>faible</i>	0,5	0,45	0,05
<i>fort</i>	0,05	0,25	0,7

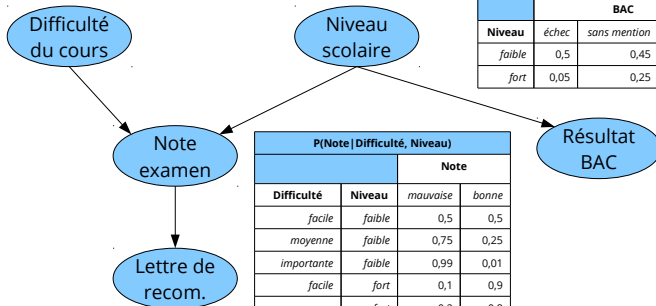


Exemple des étudiants

P(Difficulté)		
Difficulté		
<i>facile</i>	<i>moyenne</i>	<i>importante</i>
0,3	0,6	0,1

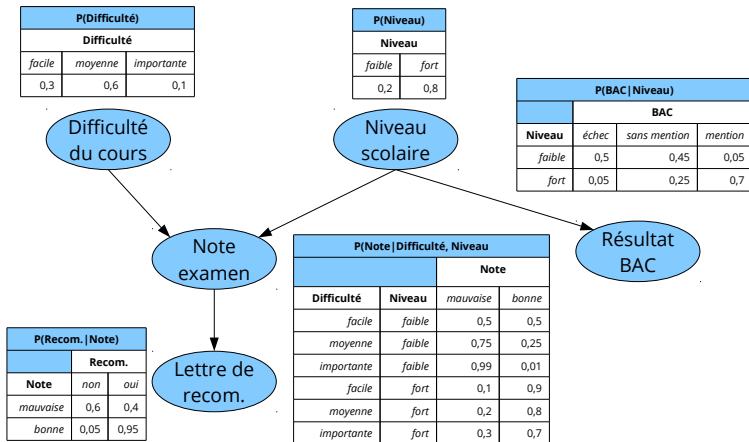
P(Niveau)	
Niveau	
<i>faible</i>	<i>fort</i>
0,2	0,8

P(BAC Niveau)			
	BAC		
Niveau	échec	sans mention	mention
<i>faible</i>	0,5	0,45	0,05
<i>fort</i>	0,05	0,25	0,7



P(Note Difficulté, Niveau)			
		Note	
Difficulté	Niveau	<i>mauvaise</i>	<i>bonne</i>
<i>facile</i>	<i>faible</i>	0,5	0,5
<i>moyenne</i>	<i>faible</i>	0,75	0,25
<i>importante</i>	<i>faible</i>	0,99	0,01
<i>facile</i>	<i>fort</i>	0,1	0,9
<i>moyenne</i>	<i>fort</i>	0,2	0,8
<i>importante</i>	<i>fort</i>	0,3	0,7

Exemple des étudiants



Plan

- 1 Introduction
- 2 Rappels de probabilité
- 3 Approche bayésienne
- 4 Les réseaux bayésiens
- 5 Concepts fondamentaux**
 - Factorisation de la loi jointe
 - Complexité
 - Relations d'indépendances
 - Fidélité
- 6 Conclusion, perspectives et outils
- 7 Annexes

Loi jointe et factorisation

Factorisation dans un RB

- La loi jointe d'une suite de v.a. X_1, \dots, X_n représentée dans un RB se factorise comme le produit des LPC de chacune des variables

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{pa}(X_i))$$

- Rappel de la formule de Bayes généralisée (toujours vraie) :

$$P(X_1, \dots, X_n) = P(X_1) \times P(X_2 | X_1) \times P(X_3 | X_1, X_2) \times \dots \times P(X_n | X_1, \dots, X_{n-1})$$

⇒ La factorisation dans un RB permet de simplifier la formule de Bayes

Exercice

Écrire la factorisation du RB dans l'exemple de l'étudiant

Loi jointe et factorisation

Factorisation dans un RB

- La loi jointe d'une suite de v.a. X_1, \dots, X_n représentée dans un RB se factorise comme le produit des LPC de chacune des variables

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{pa}(X_i))$$

- Rappel de la formule de Bayes généralisée (toujours vraie) :

$$P(X_1, \dots, X_n) = P(X_1) \times P(X_2 | X_1) \times P(X_3 | X_1, X_2) \times \dots \times P(X_n | X_1, \dots, X_{n-1})$$

⇒ La factorisation dans un RB permet de simplifier la formule de Bayes

Exercice

Écrire la factorisation du RB dans l'exemple de l'étudiant

Complexité spatiale d'une LPC

Complexité spatiale d'une LPC

- Soient $\mathbf{X} = X_1, \dots, X_n$ et $\mathbf{Y} = Y_1, \dots, Y_m$ deux suites de v.a. à valeurs dans les ensembles discrets et finis $\mathcal{X}_1, \dots, \mathcal{X}_n$ et $\mathcal{Y}_1, \dots, \mathcal{Y}_m$
- La complexité spatiale associée à la LPC $P(X_1, \dots, X_n | Y_1, \dots, Y_m) = P(\mathbf{X} | \mathbf{Y})$, notée $CS(P(\mathbf{X} | \mathbf{Y}))$, est définie par le nombre de configurations de valeurs différentes que peuvent prendre les v.a. $X_1, \dots, X_n, Y_1, \dots, Y_m$
- Autrement dit : $CS(P(\mathbf{X} | \mathbf{Y})) = \prod_{i=1}^n |\mathcal{X}_i| \prod_{j=1}^m |\mathcal{Y}_j|$ où $|\mathcal{X}_i|$ est le nombre d'éléments dans l'ensemble \mathcal{X}_i

Exemples de l'étudiant

- la complexité spatiale de la LPC $P(\text{Note} | \text{Niveau}, \text{Difficulté})$ est égale à $2 \times 2 \times 3 = 12$
- la complexité spatiale de la LPC $P(\text{Niveau})$ est égale à 3

Complexité probabiliste d'une LPC

Complexité probabiliste d'une LPC

- Soient $\mathbf{X} = X_1, \dots, X_n$ et $\mathbf{Y} = Y_1, \dots, Y_m$ deux suites de v.a. à valeurs dans les ensembles discrets et finis $\mathcal{X}_1, \dots, \mathcal{X}_n$ et $\mathcal{Y}_1, \dots, \mathcal{Y}_m$
- La complexité probabiliste associée à la LPC $P(X_1, \dots, X_n | Y_1, \dots, Y_m) = P(\mathbf{X} | \mathbf{Y})$, notée $CP(P(\mathbf{X} | \mathbf{Y}))$, est définie par le nombre de paramètres (probabilités) nécessaires pour définir la LPC
- Autrement dit : $CP(P(\mathbf{X} | \mathbf{Y})) = \prod_{i=1}^n (|\mathcal{X}_i| - 1) \prod_{j=1}^m |\mathcal{Y}_j|$
- La CP tient simplement compte du fait qu'une LPC doit sommer à 1 pour chaque configuration des variables de conditionnement
- La CP mesure le potentiel de modélisation d'une loi

Exemple de l'étudiant

- la complexité probabiliste de la LPC $P(\text{Note} | \text{Niveau}, \text{Difficulté})$ est égale à $1 \times 2 \times 3 = 6$
- la complexité probabiliste de la LPC $P(\text{Niveau})$ est égale à 2

Complexité d'une loi jointe naturelle

Complexité d'une loi jointe naturelle

- Soit X_1, \dots, X_n une suite de v.a. à valeurs dans les ensembles discrets et finis $\mathcal{X}_1, \dots, \mathcal{X}_n$
- La **complexité spatiale** d'une loi jointe naturelle $P(X_1, \dots, X_n)$ vaut

$$CS(P(X_1, \dots, X_n)) = \prod_{i=1}^n |\mathcal{X}_i|$$

- La **complexité probabiliste** d'une loi jointe naturelle $P(X_1, \dots, X_n)$ vaut

$$CP(P(X_1, \dots, X_n)) = CS(P(X_1, \dots, X_n)) - 1$$

Complexité d'une loi jointe dans un RB

Complexité d'une loi jointe factorisée dans un RB

- Soit X_1, \dots, X_n une suite de v.a. à valeurs dans les ensembles discrets et finis $\mathcal{X}_1, \dots, \mathcal{X}_n$
- La complexité spatiale (resp. probabiliste) d'une loi jointe factorisée dans un RB, notée $P^{RB}(X_1, \dots, X_n)$, est définie comme étant la somme des complexités spatiales (resp. probabilistes) associées à chacune des LPC $P(X_i | \text{pa}(X_i))$.
- La **complexité spatiale** a pour expression :

$$CS(P^{RB}(X_1, \dots, X_n)) = \sum_{i=1}^n CS(P(X_i | \text{pa}(X_i))) = \sum_{i=1}^n |\mathcal{X}_i| \times \prod_{\mathcal{X} \in \text{pa}(\mathcal{X}_i)} |\mathcal{X}|$$

- La **complexité probabiliste** a pour expression :

$$CP(P^{RB}(X_1, \dots, X_n)) = \sum_{i=1}^n CP(P(X_i | \text{pa}(X_i))) = \sum_{i=1}^n (|\mathcal{X}_i| - 1) \times \prod_{\mathcal{X} \in \text{pa}(\mathcal{X}_i)} |\mathcal{X}|$$

Conséquences pratiques

RB = Représentation parcimonieuse

- Un RB est une représentation compacte d'un processus aléatoire
- Moins il y a d'arcs dans le graphe :
 - ⇒ plus des hypothèses d'indépendances conditionnelles entre les variables sont posées
 - ⇒ plus le potentiel de modélisation, i.e. la capacité à représenter des phénomènes complexes, diminue (*CP* faible)
 - ⇒ plus la représentation par RB est avantageuse du point de vue du stockage, du paramétrage et de la complexité calculatoire (*CS* faible)

Complexité : Exemple

Complexité du modèle de l'étudiant

- Loi jointe naturelle :

$$\Rightarrow CS = [3(\text{Difficulté}) \times 2(\text{Niveau}) \times 2(\text{Note}) \times 3(\text{BAC}) \times 2(\text{Recom.})] = 72$$

$$\Rightarrow CP = CS - 1 = 71 \text{ paramètres}$$

- Représentation par RB :

$$\Rightarrow CS = 3(\text{Difficulté}) + 2(\text{Niveau}) + 2 \times 3 \times 2(\text{Note}) + 3 \times 2(\text{BAC}) + 2 \times 2(\text{Recom.}) = 27$$

$$\Rightarrow CP = 2(\text{Difficulté}) + 1(\text{Niveau}) + 1 \times 3 \times 2(\text{Note}) + 2 \times 2(\text{BAC}) + 1 \times 2(\text{Recom.}) = 15 \text{ paramètres}$$

Complexité : Exemple

Complexité du modèle de l'étudiant

- Loi jointe naturelle :
 - $\Rightarrow CS = [3(\text{Difficulté}) \times 2(\text{Niveau}) \times 2(\text{Note}) \times 3(\text{BAC}) \times 2(\text{Recom.})] = 72$
 - $\Rightarrow CP = CS - 1 = 71$ paramètres
- Représentation par RB :
 - $\Rightarrow CS = 3(\text{Difficulté}) + 2(\text{Niveau}) + 2 \times 3 \times 2(\text{Note}) + 3 \times 2(\text{BAC}) + 2 \times 2(\text{Recom.}) = 27$
 - $\Rightarrow CP = 2(\text{Difficulté}) + 1(\text{Niveau}) + 1 \times 3 \times 2(\text{Note}) + 2 \times 2(\text{BAC}) + 1 \times 2(\text{Recom.}) = 15$ paramètres

Relations d'indépendances

RB et indépendances conditionnelles

- Un RB peut être vu comme un codage de relations d'indépendances conditionnelles parmi un ensemble de variables aléatoires
- Attention :
 - L'absence d'arc entre deux variables ne signifie pas qu'elles sont indépendantes
 - Un arc entre deux variables ne signifie pas toujours qu'elles sont dépendantes (cf. notion de fidélité)

Structures fondamentales

Les relations d'indépendances dans un RB se déduisent des trois structures fondamentales suivantes :

- 1 Connexion série : $X \rightarrow Y \rightarrow Z$
- 2 Connexion divergente : $X \leftarrow Y \rightarrow Z$
- 3 Connexion convergente (V-structure) : $X \rightarrow Y \leftarrow Z$

Relations d'indépendances

RB et indépendances conditionnelles

- Un RB peut être vu comme un codage de relations d'indépendances conditionnelles parmi un ensemble de variables aléatoires
- Attention :
 - L'absence d'arc entre deux variables ne signifie pas qu'elles sont indépendantes
 - Un arc entre deux variables ne signifie pas toujours qu'elles sont dépendantes (cf. notion de fidélité)

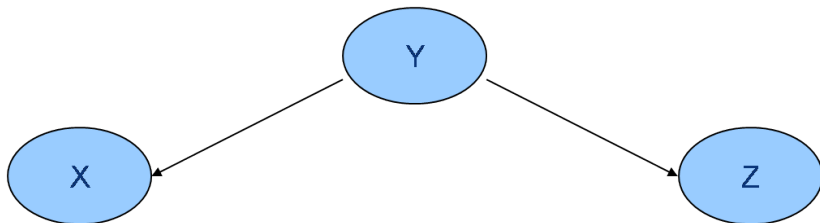
Structures fondamentales

Les relations d'indépendances dans un RB se déduisent des trois structures fondamentales suivantes :

- 1 Connexion série : $X \rightarrow Y \rightarrow Z$
- 2 Connexion divergente : $X \leftarrow Y \rightarrow Z$
- 3 Connexion convergente (V-structure) : $X \rightarrow Y \leftarrow Z$

Relations d'indépendances

Connexion divergente

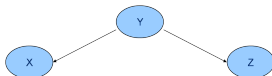


Propriétés

- X et Z sont dépendantes.
- Si Y est connue, alors X et Z sont **indépendantes**
- ⇔ $X \perp\!\!\!\perp Z | Y$: X et Z sont indépendantes **conditionnellement à Y**
- ⇔ Si la loi $P(Y)$ est connue, X n'intervient pas dans le calcul de $P(Z)$ (et inversement)
- ⇔ $P(Z|X, Y) = P(Z|Y) \Leftrightarrow P(X, Z|Y) = P(X|Y) \times P(Z|Y)$

Relations d'indépendances

Connexion divergente - Démonstration



Démonstration de la propriété $X \perp\!\!\!\perp Z|Y$

- La loi des v.a. X , Y et Z se factorise dans ce RB comme suit :

$$P(X, Y, Z) = P(X|Y) \times P(Y) \times P(Z|Y)$$

- Conditionnons par Y en divisant par $P(Y)$:

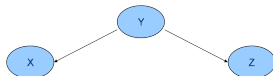
$$P(X, Y, Z) \times \frac{1}{P(Y)} = P(X|Y) \times P(Y) \times P(Z|Y) \times \frac{1}{P(Y)}$$

- À gauche, on a par définition $\frac{P(X, Y, Z)}{P(Y)} = P(X, Z|Y)$ et à droite le terme $P(Y)$ disparaît.

$$\Rightarrow P(X, Z|Y) = P(X|Y) \times P(Z|Y) \Rightarrow X \perp\!\!\!\perp Z|Y$$

Relations d'indépendances

Connexion divergente - Démonstration



Démonstration de la propriété $X \perp\!\!\!\perp Z | Y$

- La loi des v.a. X , Y et Z se factorise dans ce RB comme suit :

$$P(X, Y, Z) = P(X|Y) \times P(Y) \times P(Z|Y)$$

- Conditionnons par Y en divisant par $P(Y)$:

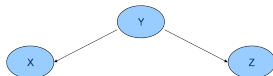
$$P(X, Y, Z) \times \frac{1}{P(Y)} = P(X|Y) \times P(Y) \times P(Z|Y) \times \frac{1}{P(Y)}$$

- À gauche, on a par définition $\frac{P(X, Y, Z)}{P(Y)} = P(X, Z|Y)$ et à droite le terme $P(Y)$ disparaît.

$$\Rightarrow P(X, Z|Y) = P(X|Y) \times P(Z|Y) \Rightarrow X \perp\!\!\!\perp Z | Y$$

Relations d'indépendances

Connexion divergente - Démonstration



Démonstration de la propriété $X \perp\!\!\!\perp Z | Y$

- La loi des v.a. X , Y et Z se factorise dans ce RB comme suit :

$$P(X, Y, Z) = P(X|Y) \times P(Y) \times P(Z|Y)$$

- Conditionnons par Y en divisant par $P(Y)$:

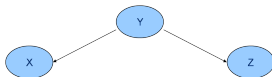
$$P(X, Y, Z) \times \frac{1}{P(Y)} = P(X|Y) \times P(Y) \times P(Z|Y) \times \frac{1}{P(Y)}$$

- À gauche, on a par définition $\frac{P(X, Y, Z)}{P(Y)} = P(X, Z|Y)$ et à droite le terme $P(Y)$ disparaît.

$$\Rightarrow P(X, Z|Y) = P(X|Y) \times P(Z|Y) \Rightarrow X \perp\!\!\!\perp Z | Y$$

Relations d'indépendances

Connexion divergente - Démonstration



Démonstration de la propriété $X \perp\!\!\!\perp Z | Y$

- La loi des v.a. X , Y et Z se factorise dans ce RB comme suit :

$$P(X, Y, Z) = P(X|Y) \times P(Y) \times P(Z|Y)$$

- Conditionnons par Y en divisant par $P(Y)$:

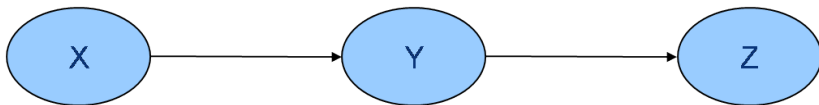
$$P(X, Y, Z) \times \frac{1}{P(Y)} = P(X|Y) \times P(Y) \times P(Z|Y) \times \frac{1}{P(Y)}$$

- À gauche, on a par définition $\frac{P(X, Y, Z)}{P(Y)} = P(X, Z|Y)$ et à droite le terme $P(Y)$ disparaît.

$$\Rightarrow P(X, Z|Y) = P(X|Y) \times P(Z|Y) \Rightarrow X \perp\!\!\!\perp Z | Y$$

Relations d'indépendances

Connexion série

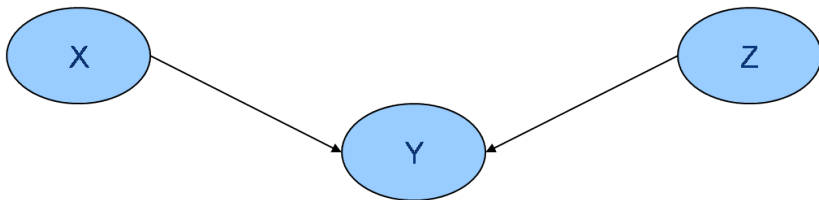


Propriétés

- X et Z sont dépendantes.
- Si Y est connue, alors X et Z sont **indépendantes**
- ⇔ $X \perp\!\!\!\perp Z | Y$: X et Z sont indépendantes **conditionnellement à Y**
- ⇔ Si la loi $P(Y)$ est connue, X n'intervient pas dans le calcul de $P(Z)$ (et inversement)
- ⇔ $P(Z|X, Y) = P(Z|Y) \Leftrightarrow P(X, Z|Y) = P(X|Y) \times P(Z|Y)$

Relations d'indépendances

Connexion convergente (V-structure)



Propriétés

- $X \perp\!\!\!\perp Z$: X et Z sont indépendantes.
- Si Y est connue, alors X et Z sont **dépendantes**
- ⇒ X et Z sont **dépendantes conditionnellement à Y**
- ⇒ Si la loi $P(Y)$ est connue, X intervient dans le calcul de $P(Z)$ (et inversement)
- ⇒ $P(X, Z) = P(X) \times P(Z)$

Notion de fidélité

Définition

Un RB \mathcal{M} représentant la loi jointe d'une suite de v.a. X_1, \dots, X_n de graphe \mathcal{G} et de LPC $P(X_i | \text{pa}(X_i))$ est qualifié de fidèle si toutes les relations d'indépendances conditionnelles induites par la loi jointe $P(X_1, \dots, X_n)$ peuvent être déduites du graphe \mathcal{G}

Exemple d'infidélité

- $\mathcal{G} = X_1 \rightarrow X_2$
 - $P(X_2 | X_1 = 0) = P(X_2 | X_1 = 1) = [p, 1 - p]$
- ⇒ Les variables X_1 et X_2 sont indépendantes d'un point de vue probabiliste mais semblent dépendantes à la lecture du graphe

Plan

- 1 Introduction
- 2 Rappels de probabilité
- 3 Approche bayésienne
- 4 Les réseaux bayésiens
- 5 Concepts fondamentaux
- 6 Conclusion, perspectives et outils**
- 7 Annexes

Conclusion

Points clés de ce cours

- Un réseau bayésien permet de structurer un phénomène aléatoire sous forme de graphe :
 - Noeuds du graphe = Variables aléatoires
 - Arcs entre noeuds = Relations de dépendance
- L'information apportée par la structure du graphe permet en général de simplifier la loi jointe du phénomène aléatoire étudié - notion de parcimonie
- Si les LPC du RB sont connues, des algorithmes d'inférence probabiliste permettent de déduire de nouvelles connaissances sur le phénomène aléatoire étudié

RB = formalisme séduisant

- Lisibilité des modèles
- Rigueur mathématique
- Fort potentiel de modélisation

Perspectives

Perspectives du cours

- Apprendre à faire des calculs dans un RB
- Renseigner automatiquement les LPC à partir de données
- Étudier quelques extensions des RB
- Appliquer les RB à un problème de classification automatique réelle

Outils pour manipuler les RB

Outils graphiques

- Genie : outil gratuit sous Windows et Linux (*via* Wine)
- Hugin : outil commercial sous Windows

Librairies informatiques

- Agrum : librairie opensource C++ avec une interface Python
PyAgrum
- gRain : librairie opensource en R
- Smile : librairie gratuite C++

Références



Pearl, J. (sept. 1988). *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. 1^{re} éd. Morgan Kaufmann.



Shachter, Ross D. (1998). "Bayes-ball : Rational Pastime (for Determining Irrelevance and Requisite Information in Belief Networks and Influence Diagrams)". In : *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. UAI'98. Madison, Wisconsin : Morgan Kaufmann Publishers Inc., p. 480-487.

Plan

- 1 Introduction
- 2 Rappels de probabilité
- 3 Approche bayésienne
- 4 Les réseaux bayésiens
- 5 Concepts fondamentaux
- 6 Conclusion, perspectives et outils
- 7 Annexes
 - Principe du Bayes Ball

Principe du Bayes Ball (Shachter 1998)

Objectif

Déterminer si deux ensembles de variables A et B dans un RB sont indépendantes (conditionnellement à un ensemble de variables C)

Principes

- Colorier les variables observées/instanciées de l'ensemble C
- Faire passer une balle d'une des variables de A vers une des variables de B en respectant les règles du Bayes Ball (cf. slide suivante)
- La balle peut voyager dans le sens opposé des liens du graphe
- Si la balle ne peut atteindre B à partir de A alors $A \perp\!\!\!\perp B | C$

Principe du Bayes Ball (Shachter 1998)

Objectif

Déterminer si deux ensembles de variables A et B dans un RB sont indépendantes (conditionnellement à un ensemble de variables C)

Principes

- Colorier les variables observées/instanciées de l'ensemble C
- Faire passer une balle d'une des variables de A vers une des variables de B en respectant les règles du Bayes Ball (cf. slide suivante)
- La balle peut voyager dans le sens opposé des liens du graphe
- Si la balle ne peut atteindre B à partir de A alors $A \perp\!\!\!\perp B | C$

Règles du Bayes Ball

→ Balle bloquée

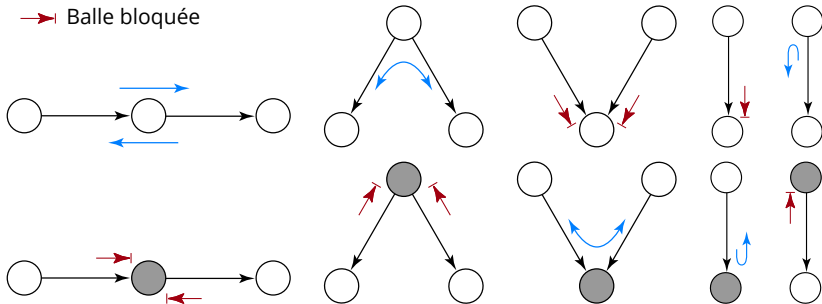


Image de M. Paskin issu d'un cours sur les modèles graphiques