Modélisation Stochastique et Réseaux Bayésiens

Apprentissage automatique des lois de probabilité conditionnelles

Roland Donat

Université de Bretagne Sud

ENSIBS - Spécialité Cyber Data

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

Plan de la présentation

Objectifs pédagogiques

- Comprendre les problématiques liées à la construction pratique d'un réseau bayésien
- Estimer automatiquement les lois de probabilité conditionnelles (LPC) à partir de données observées sur le phénomène étudié

Plan

1. Introduction
2. Rappels statistiques
3. Apprentissage des LPC : Données complètes
4. Conclusion
5. Bonus : Apprentissage des LPC : Données incomplètes

Introduction

Objectif

Objectif

- Modéliser un phénomène aléatoire à partir d'un RB

$\Rightarrow$ Représenter la loi jointe d'une suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$

Problèmes

- Comment déterminer la structure du RB (i.e. le graphe) ?
- Comment estimer les lois de probabilité conditionnelles (LPC) : $P(X_d | \text{pa}(X_d))$, $d = 1, \ldots, D$?

Approches envisageables

- Approche par expertise : Utilisation d'avis d'experts et connaissances métiers
- **Approche statistique : Utilisation de bases de données (contenant éventuellement des informations incomplètes)**
- Approche mixte : expertise + bases de données

Introduction

Objectif

Objectif

- Modéliser un phénomène aléatoire à partir d'un RB

$\Rightarrow$ Représenter la loi jointe d'une suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$

Problèmes

- Comment déterminer la structure du RB (i.e. le graphe) ?
- Comment estimer les lois de probabilité conditionnelles (LPC) : $P(X_d | \mathrm{pa}(X_d))$, $d = 1, \ldots, D$?

Approches envisageables

- Approche par expertise : Utilisation d'avis d'experts et connaissances métiers
- **Approche statistique : Utilisation de bases de données (contenant éventuellement des informations incomplètes)**
- Approche mixte : expertise + bases de données

Introduction

Objectif

Objectif

- Modéliser un phénomène aléatoire à partir d'un RB

$\Rightarrow$ Représenter la loi jointe d'une suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$

Problèmes

- Comment déterminer la structure du RB (i.e. le graphe) ?
- Comment estimer les lois de probabilité conditionnelles (LPC) : $P(X_d | \text{pa}(X_d))$, $d = 1, \ldots, D$?

Approches envisageables

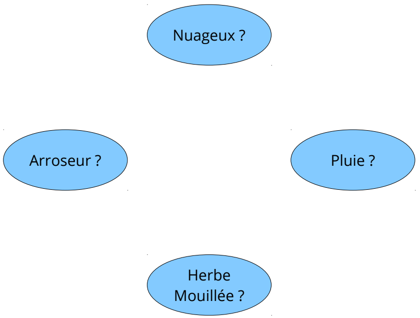
- Approche par expertise : Utilisation d'avis d'experts et connaissances métiers
- **Approche statistique : Utilisation de bases de données (contenant éventuellement des informations incomplètes)**
- Approche mixte : expertise + bases de données

Introduction

Exemple : Pourquoi l'herbe de mon jardin est-elle mouillée ?

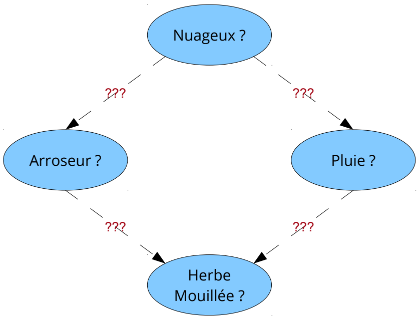
Introduction

Exemple : Pourquoi l'herbe de mon jardin est-elle mouillée ?



Introduction

Exemple : Pourquoi l'herbe de mon jardin est-elle mouillée ?

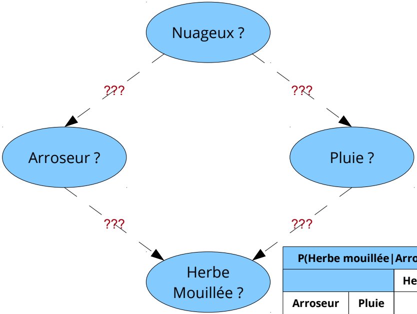


Introduction

Exemple : Pourquoi l'herbe de mon jardin est-elle mouillée ?

P(Nuageux)	
Nuageux	
non	*oui*
???	???

P(Arroseur|Nuageux)		
	Arroseur	
Nuageux	*non*	*oui*
non	???	???
oui	???	???

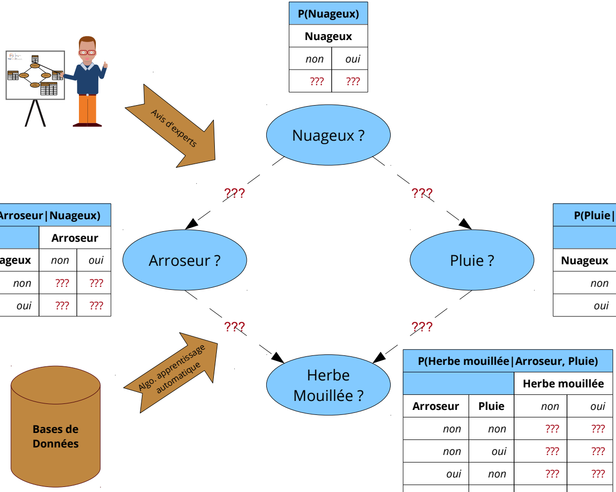


P(Pluie|Nuageux)		
	Pluie	
Nuageux	*non*	*oui*
non	???	???
oui	???	???

P(Herbe mouillée|Arroseur, Pluie)			
		Herbe mouillée	
Arroseur	**Pluie**	*non*	*oui*
non	*non*	???	???
non	*oui*	???	???
oui	*non*	???	???
oui	*oui*	???	???

Introduction

Exemple : Pourquoi l'herbe de mon jardin est-elle mouillée ?



P(Nuageux)	
Nuageux	
non	*oui*
???	???

P(Arroseur|Nuageux)		
	Arroseur	
Nuageux	*non*	*oui*
non	???	???
oui	???	???

P(Pluie|Nuageux)		
	Pluie	
Nuageux	*non*	*oui*
non	???	???
oui	???	???

P(Herbe mouillée|Arroseur, Pluie)			
		Herbe mouillée	
Arroseur	Pluie	*non*	*oui*
non	*non*	???	???
non	*oui*	???	???
oui	*non*	???	???
oui	*oui*	???	???

Plan

Rappels statistiques

Base de données - Exemple

Date	Éq. dom.	Éq. ext.	FT	HT
2013-08-09	Montpellier	Paris SG	1-1	1-0
2013-08-10	Bordeaux	Monaco	0-2	0-0
2013-08-10	Evian	Sochaux	1-1	1-0
2013-08-10	Lille	Lorient	1-0	1-0
2013-08-10	Lyon	Nice	4-0	1-0
2013-08-10	Nantes	Bastia	2-0	1-0
2013-08-10	Rennes	Reims	2-1	1-1
2013-08-10	Valenciennes	Toulouse	3-0	1-0
2013-08-11	Ajaccio	St Etienne	0-1	0-1
2013-08-11	Guingamp	Marseille	1-3	0-3
2013-08-16	Sochaux	Lyon	1-3	1-2
2013-08-17	Bastia	Valenciennes	2-0	0-0
2013-08-17	Marseille	Evian	2-0	1-0
2013-08-17	Nice	Rennes	2-1	1-1

Rappels statistiques

Base de données - Exemple

Date	Éq. dom.	Éq. ext.	FT	HT
2013-08-09	Montpellier	Paris SG	1-1	1-0
2013-08-10	Bordeaux	Monaco	0-2	0-0
2013-08-10	Evian	Sochaux	1-1	1-0
2013-08-10	Lille	Lorient	1-0	1-0
2013-08-10	Lyon	Nice	4-0	1-0
2013-08-10	Nantes	Bastia	2-0	1-0
2013-08-10	Rennes	Reims	2-1	1-1
2013-08-10	Valenciennes	Toulouse	3-0	1-0
2013-08-11	Ajaccio	St Etienne	0-1	0-1
2013-08-11	Guingamp	Marseille	1-3	0-3
2013-08-16	Sochaux	Lyon	1-3	1-2
2013-08-17	Bastia	Valenciennes	2-0	0-0
2013-08-17	Marseille	Evian	2-0	1-0
2013-08-17	Nice	Rennes	2-1	1-1

Caractéristiques des données

- Nombre de variables : 5
- Valeurs prises par la variable ”FT”
- Observation caractérisée par le vecteur (2013-08-10, Lille, Lorient, 1-0, 1-0)

Rappels statistiques

Base de données - Exemple

Date	Éq. dom.	Éq. ext.	FT	HT
2013-08-09	Montpellier	Paris SG	1-1	1-0
2013-08-10	Bordeaux	Monaco	0-2	0-0
2013-08-10	Evian	Sochaux	1-1	1-0
2013-08-10	Lille	Lorient	1-0	1-0
2013-08-10	Lyon	Nice	4-0	1-0
2013-08-10	Nantes	Bastia	2-0	1-0
2013-08-10	Rennes	Reims	2-1	1-1
2013-08-10	Valenciennes	Toulouse	3-0	1-0
2013-08-11	Ajaccio	St Etienne	0-1	0-1
2013-08-11	Guingamp	Marseille	1-3	0-3
2013-08-16	Sochaux	Lyon	1-3	1-2
2013-08-17	Bastia	Valenciennes	2-0	0-0
2013-08-17	Marseille	Evian	2-0	1-0
2013-08-17	Nice	Rennes	2-1	1-1

Caractéristiques des données

- Nombre de variables : 5
- Valeurs prises par la variable ”FT”
- Observation caractérisée par le vecteur (2013-08-10, Lille, Lorient, 1-0, 1-0)

Rappels statistiques

Base de données - Exemple

Date	Éq. dom.	Éq. ext.	FT	HT
2013-08-09	Montpellier	Paris SG	1-1	1-0
2013-08-10	Bordeaux	Monaco	0-2	0-0
2013-08-10	Evian	Sochaux	1-1	1-0
2013-08-10	Lille	Lorient	1-0	1-0
2013-08-10	Lyon	Nice	4-0	1-0
2013-08-10	Nantes	Bastia	2-0	1-0
2013-08-10	Rennes	Reims	2-1	1-1
2013-08-10	Valenciennes	Toulouse	3-0	1-0
2013-08-11	Ajaccio	St Etienne	0-1	0-1
2013-08-11	Guingamp	Marseille	1-3	0-3
2013-08-16	Sochaux	Lyon	1-3	1-2
2013-08-17	Bastia	Valenciennes	2-0	0-0
2013-08-17	Marseille	Evian	2-0	1-0
2013-08-17	Nice	Rennes	2-1	1-1

Caractéristiques des données

- Nombre de variables : 5
- Valeurs prises par la variable ”FT”
- Observation caractérisée par le vecteur (2013-08-10, Lille, Lorient, 1-0, 1-0)

Rappels statistiques

Base de données - Définition

Définition : Base de données

- Une base de données (BdD), notée $\mathcal{D}$, est un ensemble de N observations/individus/exemples caractérisés par D variables
- Formellement, une BdD peut se mettre sous la forme d'une matrice :

$$\mathcal{D} = \begin{bmatrix} x_{1,1} & \dots & x_{1,d} & \dots & x_{1,D} \\ \vdots & & \vdots & & \vdots \\ x_{n,1} & \dots & x_{n,d} & \dots & x_{n,D} \\ \vdots & & \vdots & & \vdots \\ x_{N,1} & \dots & x_{N,d} & \dots & x_{N,D} \end{bmatrix}$$

  - Le vecteur colonne $\boldsymbol{x}_{\cdot,d} = (x_{1,d}, \dots, x_{n,d}, \dots, x_{N,d})$ représente toutes les observations de la variable d
  - Le vecteur ligne $\boldsymbol{x}_{n,\cdot} = (x_{n,1}, \dots, x_{n,d}, \dots, x_{n,D})$ représente la n-ème observation de la BdD

Rappels statistiques

Modélisation statistique : objectifs et démarche

Objectifs

1. Résumer quantitativement l'information contenue dans une BdD en utilisant un modèle probabiliste
2. Exploiter le modèle pour déduire de nouvelles connaissances

Démarche

- Considérer les données comme des réalisations de variables aléatoires (v.a.) associées à une certaine loi jointe
- Formellement, cela signifie que chaque observation $\boldsymbol{x}_{n,\cdot} = (x_{n,1}, \ldots, x_{n,D})$ est supposée être une réalisation d'une suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ de loi jointe $\mathcal{L}(\boldsymbol{\theta})$ où $\boldsymbol{\theta}$ représente les paramètres de la loi
- Notation : $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{L}(\boldsymbol{\theta})$

Rappels statistiques

Modélisation statistique : objectifs et démarche

Objectifs

1. Résumer quantitativement l'information contenue dans une BdD en utilisant un modèle probabiliste
2. Exploiter le modèle pour déduire de nouvelles connaissances

Démarche

- Considérer les données comme des réalisations de variables aléatoires (v.a.) associées à une certaine loi jointe
- Formellement, cela signifie que chaque observation $\boldsymbol{x}_{n,\cdot} = (x_{n,1}, \ldots, x_{n,D})$ est supposée être une réalisation d'une suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ de loi jointe $\mathcal{L}(\boldsymbol{\theta})$ où $\boldsymbol{\theta}$ représente les paramètres de la loi
- Notation : $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{L}(\boldsymbol{\theta})$

Rappels statistiques

Modélisation statistique : remarques et exemple

Remarques

- Si les observations dans les données sont indépendants, on parle de données i.i.d. (Indépendantes et Identiquement Distribuées)
- Si les caractéristiques (variables) des observations sont indépendantes, alors chaque variable X_d suit une loi $\mathcal{L}_d(\boldsymbol{\theta}_d)$ (Notation : $X_d \sim \mathcal{L}_d(\boldsymbol{\theta}_d)$)

Modèle gaussien

- Données unidimensionnelles ($D = 1$)

$\Rightarrow$ $\boldsymbol{X} = X_1 \sim \mathcal{N}(\mu, \sigma)$

$\Rightarrow$ $\boldsymbol{\theta} = (\mu, \sigma)$: moyenne et écart-type

- Données multidimensionnelles ($D \geq 1$)

$\Rightarrow$ $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\Rightarrow$ $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$: vecteur des moyennes et matrice de variance-covariance

Rappels statistiques

Modélisation statistique : remarques et exemple

Remarques

- Si les observations dans les données sont indépendants, on parle de données i.i.d. (Indépendantes et Identiquement Distribuées)
- Si les caractéristiques (variables) des observations sont indépendantes, alors chaque variable X_d suit une loi $\mathcal{L}_d(\boldsymbol{\theta}_d)$ (Notation : $X_d \sim \mathcal{L}_d(\boldsymbol{\theta}_d)$)

Modèle gaussien

- Données unidimensionnelles ($D = 1$)

$\Rightarrow$ $\boldsymbol{X} = X_1 \sim \mathcal{N}(\mu, \sigma)$

$\Rightarrow$ $\boldsymbol{\theta} = (\mu, \sigma)$: moyenne et écart-type

- Données multidimensionnelles ($D \geq 1$)

$\Rightarrow$ $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$\Rightarrow$ $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$: vecteur des moyennes et matrice de variance-covariance

Rappels statistiques

Estimation des paramètres d'un modèle

Contexte

- Données : On dispose d'un jeu de données $\mathcal{D} = (\boldsymbol{x}_{1,\cdot}, \ldots, \boldsymbol{x}_{N,\cdot})$ où chaque observation $\boldsymbol{x}_{n,\cdot}$ est caractérisée par D variables $(x_{n,1}, \ldots, x_{n,D})$
- Modélisation : On suppose que $\mathcal{D}$ est une suite de N réalisations i.i.d. du vecteur aléatoire $\boldsymbol{X} = (X_1, \ldots, X_D)$ distribué selon la loi $\mathcal{L}(\boldsymbol{\theta})$
- Problématique : Comment estimer les paramètres $\boldsymbol{\theta}$ à partir des données $\mathcal{D}$?

Solution

- Construire un estimateur de $\boldsymbol{\theta}$!

$\Rightarrow$ Approche classique : Déterminer l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta}$, souvent noté $\boldsymbol{\theta}^{\mathrm{MV}}$

Rappels statistiques

Estimation des paramètres d'un modèle

Contexte

- Données : On dispose d'un jeu de données $\mathcal{D} = (\boldsymbol{x}_{1,\cdot}, \ldots, \boldsymbol{x}_{N,\cdot})$ où chaque observation $\boldsymbol{x}_{n,\cdot}$ est caractérisée par D variables $(x_{n,1}, \ldots, x_{n,D})$
- Modélisation : On suppose que $\mathcal{D}$ est une suite de N réalisations i.i.d. du vecteur aléatoire $\boldsymbol{X} = (X_1, \ldots, X_D)$ distribué selon la loi $\mathcal{L}(\boldsymbol{\theta})$
- Problématique : Comment estimer les paramètres $\boldsymbol{\theta}$ à partir des données $\mathcal{D}$?

Solution

- Construire un estimateur de $\boldsymbol{\theta}$!

$\Rightarrow$ Approche classique : Déterminer l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta}$, souvent noté $\boldsymbol{\theta}^{\text{MV}}$

Rappels statistiques

Estimation des paramètres d'un modèle - exemple de données

Données de clientèle bancaire

- Âge : âge de la personne
- Épargne : la personne a t-elle de l'épargne ?
- Vente livret A : a t-on réussi à vendre un livret A à cette personne ?

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Rappels statistiques

Définition de la vraisemblance d'un modèle par rapport à une observation

Vraisemblance d'un modèle par rapport à une observation

- Soit $\mathcal{D} = (\boldsymbol{x}_{1,\cdot}, \ldots, \boldsymbol{x}_{N,\cdot})$ un ensemble de données i.i.d. modélisées par les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{L}(\boldsymbol{\theta})$
- La vraisemblance mesure la plausibilité d'un modèle probabiliste (caractérisé par des paramètres $\boldsymbol{\theta}$) par rapport à l'observation $\boldsymbol{x}_{n,\cdot} \in \mathcal{D}$

⇒ La vraisemblance est notée $L(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot})$

- En pratique le logarithme de la vraisemblance est souvent utilisé, on parle alors de log-vraisemblance $\ell(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot}) = \ln L(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot})$

Interprétation

- une vraisemblance **élevée** signifie que le modèle choisi est **crédible** par rapport à la donnée observée
- une vraisemblance **faible** signifie que le modèle choisi est **peu crédible** par rapport à la donnée observée

Rappels statistiques

Définition de la vraisemblance d'un modèle par rapport à une observation

Vraisemblance d'un modèle par rapport à une observation

- Soit $\mathcal{D} = (\boldsymbol{x}_{1,\cdot}, \ldots, \boldsymbol{x}_{N,\cdot})$ un ensemble de données i.i.d. modélisées par les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{L}(\boldsymbol{\theta})$
- La vraisemblance mesure la plausibilité d'un modèle probabiliste (caractérisé par des paramètres $\boldsymbol{\theta}$) par rapport à l'observation $\boldsymbol{x}_{n,\cdot} \in \mathcal{D}$

$\Rightarrow$ La vraisemblance est notée $L(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot})$

- En pratique le logarithme de la vraisemblance est souvent utilisé, on parle alors de log-vraisemblance $\ell(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot}) = \ln L(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot})$

Interprétation

- une vraisemblance **élevée** signifie que le modèle choisi est **crédible** par rapport à la donnée observée
- une vraisemblance **faible** signifie que le modèle choisi est **peu crédible** par rapport à la donnée observée

Rappels statistiques

Vraisemblance d'un modèle discret et fini

Vraisemblance d'un modèle discret et fini

- Si les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ sont discrètes et finies, alors la vraisemblance du modèle $P(X_1, \ldots, X_D)$ sachant l'observation $\boldsymbol{x}_{n,\cdot}$ est donnée par :

$$L(P; \boldsymbol{x}_{n,\cdot}) = P(X_1 = x_{n,1}, \ldots, X_D = x_{n,D}), \quad \text{(ici } \boldsymbol{\theta} = P\text{)}$$

Exemple

- Observation caractérisée par
  - Âge : [26,59]
  - Épargne : non
  - Vente livret A : échec

$\Rightarrow$ Vraisemblance = 0.12

$\Rightarrow$ Log-vraisemblance $\simeq$ -2.12

Rappels statistiques

Vraisemblance d'un modèle discret et fini

Vraisemblance d'un modèle discret et fini

- Si les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ sont discrètes et finies, alors la vraisemblance du modèle $P(X_1, \ldots, X_D)$ sachant l'observation $\boldsymbol{x}_{n,\cdot}$ est donnée par :

$$L(P; \boldsymbol{x}_{n,\cdot}) = P(X_1 = x_{n,1}, \ldots, X_D = x_{n,D}), \quad (\text{ici } \boldsymbol{\theta} = P)$$

Exemple

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	*non*	*échec*	0,08
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	*non*	*échec*	0,12
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	*oui*	*succès*	0,046875

- Observation caractérisée par
  - Âge : [26,59]
  - Épargne : non
  - Vente livret A : échec

$\Rightarrow$ Vraisemblance = 0.12

$\Rightarrow$ Log-vraisemblance $\simeq$ -2.12

Rappels statistiques

Vraisemblance d'un modèle discret et fini

Vraisemblance d'un modèle discret et fini

- Si les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ sont discrètes et finies, alors la vraisemblance du modèle $P(X_1, \ldots, X_D)$ sachant l'observation $\boldsymbol{x}_{n,\cdot}$ est donnée par :

$$L(P; \boldsymbol{x}_{n,\cdot}) = P(X_1 = x_{n,1}, \ldots, X_D = x_{n,D}), \quad (\text{ici } \boldsymbol{\theta} = P)$$

Exemple

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	*non*	*échec*	0,08
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	***non***	***échec***	**0,12**
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	*oui*	*succès*	0,046875

- Observation caractérisée par
  - Âge : [26,59]
  - Épargne : non
  - Vente livret A : échec

$\Rightarrow$ Vraisemblance = 0.12

$\Rightarrow$ Log-vraisemblance $\simeq$ -2.12

Rappels statistiques

Vraisemblance d'un modèle discret et fini

Vraisemblance d'un modèle discret et fini

- Si les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ sont discrètes et finies, alors la vraisemblance du modèle $P(X_1, \ldots, X_D)$ sachant l'observation $\boldsymbol{x}_{n,\cdot}$ est donnée par :

$$L(P; \boldsymbol{x}_{n,\cdot}) = P(X_1 = x_{n,1}, \ldots, X_D = x_{n,D}), \quad (\text{ici } \boldsymbol{\theta} = P)$$

Exemple

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	*non*	*échec*	0,08
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	***non***	***échec***	**0,12**
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	*oui*	*succès*	0,046875

- Observation caractérisée par
  - Âge : [26,59]
  - Épargne : non
  - Vente livret A : échec

$\Rightarrow$ Vraisemblance = 0.12

$\Rightarrow$ Log-vraisemblance $\simeq$ -2.12

Rappels statistiques

Vraisemblance d'un modèle par rapport à un jeu de données

Vraisemblance d'un modèle par rapport à un jeu de données

- La vraisemblance d'un modèle (paramétré par $\boldsymbol{\theta}$) par rapport à un jeu de données $\mathcal{D}$, notée $L(\boldsymbol{\theta}; \mathcal{D})$, correspond à la plausibilité du modèle $\boldsymbol{\theta}$ par rapport à l'observation des données $\mathcal{D}$
- Lorsque les données sont supposées i.i.d., la vraisemblance est définie par

$$L(\boldsymbol{\theta}; \mathcal{D}) = \prod_{n=1}^{N} L(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot})$$

- Lorsque l'on s'intéresse à un jeu de données, on utilise souvent la log-vraisemblance donnée par

$$\ell(\boldsymbol{\theta}; \mathcal{D}) = \ln L(\boldsymbol{\theta}; \mathcal{D}) = \sum_{n=1}^{N} \ell(\boldsymbol{\theta}; \boldsymbol{x}_{n,\cdot})$$

Rappels statistiques

Vraisemblance d'un modèle par rapport à un jeu de données

Interprétation

- La vraisemblance est un indicateur permettant de comparer différents modèles probabilistes compte tenu de données oberservées
- Dans l'absolu plus la log-vraisemblance d'un modèle par rapport à des $\mathcal{D}$ est élevée, plus le modèle probabiliste considéré est adapté pour représenter le phénomène sous-jacent

Attention

- La vraisemblance décroît avec le nombre de données observées

$\Rightarrow$ Il n'est pas pertinent de comparer des vraisemblances obtenues à partir de jeux de données de taille différente

$\Rightarrow$ En revanche, il peut être intéressant de comparer des vraisemblances moyennes par donnée observée

Rappels statistiques

Vraisemblance d'un modèle par rapport à jeu de données - Exemple

Loi jointe

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	*non*	*échec*	0,08
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	*non*	*échec*	0,12
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	*oui*	*succès*	0,046875

Données

A : Classe âge	E : Épargne	V : Vente livret A	Vraisemblance	Log-vraisemblance
[26,59]	non	échec		
[18,25]	non	échec		
60+	oui	succès		
		Total données		

Rappels statistiques

Vraisemblance d'un modèle par rapport à jeu de données - Exemple

Loi jointe

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	*non*	*échec*	0,08
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	***non***	***échec***	**0,12**
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	*oui*	*succès*	0,046875

Données

A : Classe âge	E : Épargne	V : Vente livret A	Vraisemblance	Log-vraisemblance
[26,59]	non	échec	**0,12**	**-2.1203**
[18,25]	non	échec		
60+	oui	succès		
		Total données		

Rappels statistiques

Vraisemblance d'un modèle par rapport à jeu de données - Exemple

Loi jointe

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	***non***	***échec***	**0,08**
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	*non*	*échec*	0,12
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	*oui*	*succès*	0,046875

Données

A : Classe âge	E : Épargne	V : Vente livret A	Vraisemblance	Log-vraisemblance
[26,59]	non	échec	0,12	-2.1203
[18,25]	non	échec	**0,08**	**-2.5257**
60+	oui	succès		
		Total données		

Rappels statistiques

Vraisemblance d'un modèle par rapport à jeu de données - Exemple

Loi jointe

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	*non*	*échec*	0,08
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	*non*	*échec*	0,12
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	***oui***	***succès***	**0,046875**

Données

A : Classe âge	E : Épargne	V : Vente livret A	Vraisemblance	Log-vraisemblance
[26,59]	non	échec	0,12	-2.1203
[18,25]	non	échec	0,08	-2.5257
60+	oui	succès	**0,046875**	**-3.0603**
		Total données		

Rappels statistiques

Vraisemblance d'un modèle par rapport à jeu de données - Exemple

Loi jointe

P(Âge, Épargne, Vente livret A)			
Âge	**Épargne**	**Vente livret A**	Proba
[18,25]	*non*	*échec*	0,08
[18,25]	*non*	*succès*	0,12
[18,25]	*oui*	*échec*	0,0075
[18,25]	*oui*	*succès*	0,0425
[26,59]	*non*	*échec*	0,12
[26,59]	*non*	*succès*	0,08
[26,59]	*oui*	*échec*	0,06
[26,59]	*oui*	*succès*	0,24
60+	*non*	*échec*	0,05625
60+	*non*	*succès*	0,00625
60+	*oui*	*échec*	0,140625
60+	*oui*	*succès*	0,046875

Données

A : Classe âge	E : Épargne	V : Vente livret A	Vraisemblance	Log-vraisemblance
[26,59]	non	échec	0,12	-2.1203
[18,25]	non	échec	0,08	-2.5257
60+	oui	succès	0,046875	-3.0603
		Total données	**4,5E-4**	**-7.7063**

Rappels statistiques

Estimateur du maximum de vraisemblance

Estimateur du maximum de vraisemblance (Fisher 1922)

- Soit $\mathcal{D} = (\boldsymbol{x}_{1,\cdot}, \ldots, \boldsymbol{x}_{N,\cdot})$ un ensemble de données supposées i.i.d. modélisées par les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{L}(\boldsymbol{\theta})$
- On dit que $\boldsymbol{\theta}^{\mathrm{MV}}$ est un estimateur du maximum de vraisemblance (EMV) de $\boldsymbol{\theta}$ si $\boldsymbol{\theta}^{\mathrm{MV}}$ maximise la vraisemblance, c-à-d.

$$\boldsymbol{\theta}^{\mathrm{MV}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathcal{D})$$

Propriétés des EMV

Les EMV sont :

- Convergents en probabilité vers les paramètres à estimer
- Efficaces, i.e. ils convergent rapidement
- Asymptotiquement normaux, i.e. il est facile de construire des intervalles de confiance sur les estimations obtenues

Rappels statistiques

Estimateur du maximum de vraisemblance

Estimateur du maximum de vraisemblance (Fisher 1922)

- Soit $\mathcal{D} = (\boldsymbol{x}_{1,\cdot}, \ldots, \boldsymbol{x}_{N,\cdot})$ un ensemble de données supposées i.i.d. modélisées par les v.a. $\boldsymbol{X} = (X_1, \ldots, X_D) \sim \mathcal{L}(\boldsymbol{\theta})$
- On dit que $\boldsymbol{\theta}^{\text{MV}}$ est un estimateur du maximum de vraisemblance (EMV) de $\boldsymbol{\theta}$ si $\boldsymbol{\theta}^{\text{MV}}$ maximise la vraisemblance, c-à-d.

$$\boldsymbol{\theta}^{\text{MV}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathcal{D})$$

Propriétés des EMV

Les EMV sont :

- Convergents en probabilité vers les paramètres à estimer
- Efficaces, i.e. ils convergent rapidement
- Asymptotiquement normaux, i.e. il est facile de construire des intervalles de confiance sur les estimations obtenues

Plan

Apprentissage des LPC : Données complètes

Contexte

Contexte

- Données : On dispose d'un jeu de données $\mathcal{D} = (\boldsymbol{x}_{1,\cdot}, \ldots, \boldsymbol{x}_{N,\cdot})$ où chaque observation $\boldsymbol{x}_{n,\cdot}$ est caractérisée par D variables $(x_{n,1}, \ldots, x_{n,D})$
- Modélisation : On suppose que $\mathcal{D}$ est une suite de N réalisations i.i.d. du vecteur aléatoire discret et fini $\boldsymbol{X} = (X_1, \ldots, X_D)$ représenté par un RB $\mathcal{M}$ défini par ses LPC, notées ici $\theta_d = (P(X_d|\text{pa}(X_d)))_{d=1,\ldots,D}$
- Le graphe du RB est supposé connu

Apprentissage des LPC : Données complètes

Contexte - Exemple

Données

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

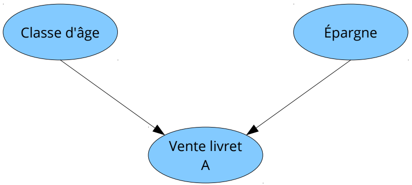
Apprentissage des LPC : Données complètes

Contexte - Exemple

Données

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Modélisation



Apprentissage des LPC : Données complètes

Vraisemblance dans un RB

Vraisemblance dans un RB

- La vraisemblance du RB $\mathcal{M}(\theta_1, \ldots, \theta_D)$ par rapport à une donnée $\boldsymbol{x}_{n,\cdot} \in \mathcal{D}$ est définie par

$$L(\theta_1, \ldots, \theta_D; \boldsymbol{x}_{n,\cdot}) = \prod_{d=1}^{D} P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$$

$\Rightarrow$ La vraisemblance du RB $\mathcal{M}(\theta_1, \ldots, \theta_D)$ par rapport aux données $\mathcal{D}$ i.i.d. est définie par

$$L(\theta_1, \ldots, \theta_D; \mathcal{D}) = \prod_{n=1}^{N} \prod_{d=1}^{D} P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$$

$\Rightarrow$ D'où la log-vraisemblance :
$\ell(\theta_1, \ldots, \theta_D; \mathcal{D}) = \sum_{n=1}^{N} \sum_{d=1}^{D} \ln P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$

Apprentissage des LPC : Données complètes

Vraisemblance dans un RB

Vraisemblance dans un RB

- La vraisemblance du RB $\mathcal{M}(\theta_1, \ldots, \theta_D)$ par rapport à une donnée $\boldsymbol{x}_{n,\cdot} \in \mathcal{D}$ est définie par

$$L(\theta_1, \ldots, \theta_D; \boldsymbol{x}_{n,\cdot}) = \prod_{d=1}^{D} P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$$

$\Rightarrow$ La vraisemblance du RB $\mathcal{M}(\theta_1, \ldots, \theta_D)$ par rapport aux données $\mathcal{D}$ i.i.d. est définie par

$$L(\theta_1, \ldots, \theta_D; \mathcal{D}) = \prod_{n=1}^{N} \prod_{d=1}^{D} P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$$

$\Rightarrow$ D'où la log-vraisemblance :
$\ell(\theta_1, \ldots, \theta_D; \mathcal{D}) = \sum_{n=1}^{N} \sum_{d=1}^{D} \ln P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$

Apprentissage des LPC : Données complètes

Vraisemblance dans un RB

Vraisemblance dans un RB

- La vraisemblance du RB $\mathcal{M}(\theta_1, \ldots, \theta_D)$ par rapport à une donnée $\boldsymbol{x}_{n,\cdot} \in \mathcal{D}$ est définie par

$$L(\theta_1, \ldots, \theta_D; \boldsymbol{x}_{n,\cdot}) = \prod_{d=1}^{D} P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$$

$\Rightarrow$ La vraisemblance du RB $\mathcal{M}(\theta_1, \ldots, \theta_D)$ par rapport aux données $\mathcal{D}$ i.i.d. est définie par

$$L(\theta_1, \ldots, \theta_D; \mathcal{D}) = \prod_{n=1}^{N} \prod_{d=1}^{D} P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$$

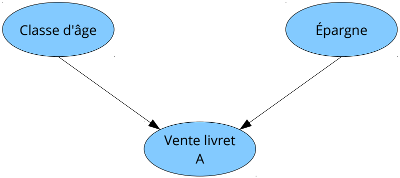
$\Rightarrow$ D'où la log-vraisemblance :
$\ell(\theta_1, \ldots, \theta_D; \mathcal{D}) = \sum_{n=1}^{N} \sum_{d=1}^{D} \ln P(X_d = x_{n,d} | \text{pa}(X_d) = \text{pa}(x_{n,d}))$

Apprentissage des LPC : Données complètes

Maximum de vraisemblance dans un RB - Exemple

$\hat{P}^{MV}$(Âge)		
Âge		
[18,25]	*[26,59]*	*60+*

$\hat{P}^{MV}$(Épargne)	
Épargne	
non	*oui*



$\hat{P}^{MV}$(Vente|Âge, Épargne)			
		Vente	
Âge	**Épargne**	*échec*	*succès*
[18,25]	*non*		
[18,25]	*oui*		
[26,59]	*non*		
[26,59]	*oui*		
60+	*non*		
60+	*oui*		

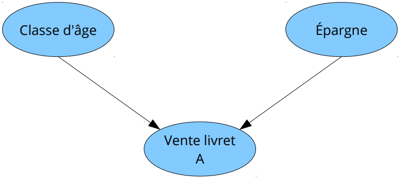
A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Apprentissage des LPC : Données complètes

Maximum de vraisemblance dans un RB - Exemple

^P^MV(Âge)		
Âge		
[18,25]	*[26,59]*	*60+*
3/14	7/14	4/14

^P^MV(Épargne)	
Épargne	
non	*oui*



^P^MV(Vente|Âge, Épargne)			
		Vente	
Âge	Épargne	*échec*	*succès*
[18,25]	*non*		
[18,25]	*oui*		
[26,59]	*non*		
[26,59]	*oui*		
60+	*non*		
60+	*oui*		

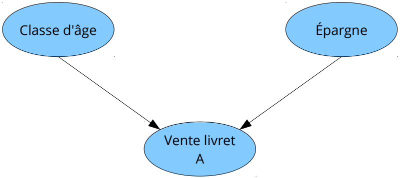
A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Apprentissage des LPC : Données complètes

Maximum de vraisemblance dans un RB - Exemple

$\hat{P}^{MV}$(Âge)		
Âge		
[18,25]	*[26,59]*	*60+*
3/14	7/14	4/14

$\hat{P}^{MV}$(Épargne)	
Épargne	
non	*oui*
9/14	5/14



$\hat{P}^{MV}$(Vente|Âge, Épargne)			
		Vente	
Âge	**Épargne**	*échec*	*succès*
[18,25]	*non*		
[18,25]	*oui*		
[26,59]	*non*		
[26,59]	*oui*		
60+	*non*		
60+	*oui*		

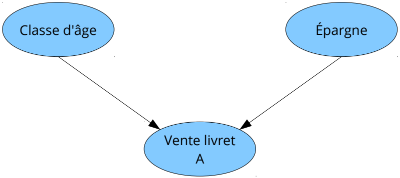
A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Apprentissage des LPC : Données complètes

Maximum de vraisemblance dans un RB - Exemple

^P^MV(Âge)		
Âge		
[18,25]	*[26,59]*	*60+*
3/14	7/14	4/14

^P^MV(Épargne)	
Épargne	
non	*oui*
9/14	5/14



^P^MV(Vente|Âge, Épargne)			
		Vente	
Âge	Épargne	*échec*	*succès*
[18,25]	*non*	1/2	1/2
[18,25]	*oui*		
[26,59]	*non*		
[26,59]	*oui*		
60+	*non*		
60+	*oui*		

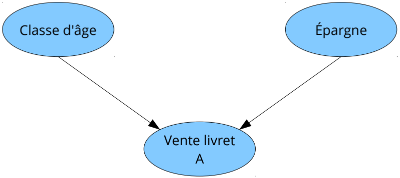
A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Apprentissage des LPC : Données complètes

Maximum de vraisemblance dans un RB - Exemple

$\hat{P}^{MV}$(Âge)		
Âge		
[18,25]	*[26,59]*	*60+*
3/14	7/14	4/14

$\hat{P}^{MV}$(Épargne)	
Épargne	
non	*oui*
9/14	5/14



$\hat{P}^{MV}$(Vente|Âge, Épargne)			
		Vente	
Âge	**Épargne**	*échec*	*succès*
[18,25]	*non*	1/2	1/2
[18,25]	*oui*	0/1	1/1
[26,59]	*non*		
[26,59]	*oui*		
60+	*non*		
60+	*oui*		

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Apprentissage des LPC : Données complètes

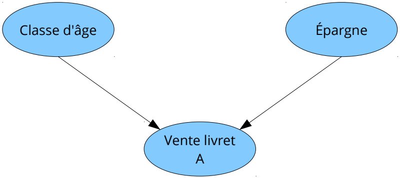
Maximum de vraisemblance dans un RB - Exemple

$\hat{P}^{MV}$(Âge)

Âge		
[18,25]	*[26,59]*	*60+*
3/14	7/14	4/14

$\hat{P}^{MV}$(Épargne)

Épargne	
non	*oui*
9/14	5/14



$\hat{P}^{MV}$(Vente|Âge, Épargne)

		Vente	
Âge	**Épargne**	*échec*	*succès*
[18,25]	*non*	1/2	1/2
[18,25]	*oui*	0/1	1/1
[26,59]	*non*	4/5	1/5
[26,59]	*oui*	1/2	1/2
60+	*non*	2/2	0/2
60+	*oui*	1/2	1/2

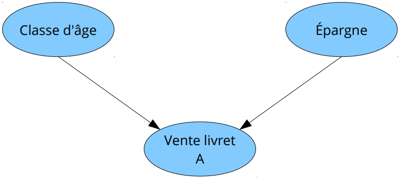
A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Apprentissage des LPC : Données complètes

Maximum de vraisemblance dans un RB - Exemple

$\hat{P}^{MV}$(Âge)		
Âge		
[18,25]	*[26,59]*	*60+*
3/14	7/14	4/14

$\hat{P}^{MV}$(Épargne)	
Épargne	
non	*oui*
9/14	5/14



$\hat{P}^{MV}$(Vente|Âge, Épargne)			
		Vente	
Âge	**Épargne**	*échec*	*succès*
[18,25]	*non*	1/2	1/2
[18,25]	*oui*	0/1	1/1
[26,59]	*non*	4/5	1/5
[26,59]	*oui*	1/2	1/2
60+	*non*	2/2	0/2
60+	*oui*	1/2	1/2

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Apprentissage des LPC : Données complètes

Maximum de vraisemblance dans un RB

Estimateurs du maximum de vraisemblance

- On cherche les estimateurs du maximum de vraisemblance (EMV) des LPC $\theta_1^{\mathrm{MV}}, \ldots, \theta_D^{\mathrm{MV}}$ vérifiant :

$$\left(\theta_1^{\mathrm{MV}}, \ldots, \theta_D^{\mathrm{MV}}\right) = \underset{\theta_1, \ldots, \theta_D}{\arg\max} \sum_{n=1}^{N} \sum_{d=1}^{D} \ln P\left(X_d = x_{n,d} | \mathrm{pa}\left(X_d\right) = \mathrm{pa}\left(x_{n,d}\right)\right)$$

$\Rightarrow$ Les estimations du maximum de vraisemblance pour chaque LPC ont pour expression :

$$\hat{P}^{\mathrm{MV}}\left(X_d = x_{d,k} | \mathrm{pa}\left(X_d\right) = \mathbf{x}'_{d,j}\right) = \hat{\theta}^{\mathrm{MV}}_{d,j,k} = \frac{N_{d,j,k}}{\sum_{k=1} N_{d,j,k}}$$

- $x_{d,k}$: k-ème valeur possible pour la v.a. X_d
- $\mathbf{x}'_{d,j}$: j-ème configuration de valeurs possibles pour les parents de la v.a. X_d
- $N_{d,j,k}$: Nombre d'occurrences de l'événement $\{X_d = x_{d,k}$ et $\mathrm{pa}\left(X_d\right) = \mathbf{x}'_{d,j}\}$ dans les données $\mathcal{D}$

Plan

Conclusion

Compétences acquises

- Rappels généraux sur l'apprentissage statistique
- Formalisation du problème d'apprentissage des LPC du point de vue statistique
- Mise en oeuvre de la méthode du maximum de vraisemblance pour estimer les LPC d'un RB

Pour aller plus loin

- Apprentissage des LPC dans le cas de données incomplètes
- Apprentissage automatique de la structure d'un RB

Plan

Apprentissage des LPC : Données incomplètes

Représentation disjonctive des données

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Représentation classique

- Les observations de chaque variable sont données explicitement

$\Rightarrow$ Représentation intuitive mais peu adaptée aux traitements des données incomplètes

Apprentissage des LPC : Données incomplètes

Représentation disjonctive des données

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**1**	0	**1**	0
0	**1**	0	**1**	0	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
0	0	**1**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**1**	0

Représentation disjonctive

- Chaque variable X possédant K modalités est décomposée en K sous-variables binaires où la modalité prise est associée à la valeur 1

$\Rightarrow$ Représentation plus complexe mais adaptée aux traitements des différents types de données incomplètes

Apprentissage des LPC : Données incomplètes

Types de données incomplètes

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
60+	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Types de données incomplètes

1. Données manquantes
2. Données partiellement observées (généralisation cas 1)
3. Données partiellement observées pondérées (généralisation cas 1-2)

Apprentissage des LPC : Données incomplètes

Types de données incomplètes

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
?	non	échec
[26,59]	non	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
?	non	échec
[26,59]	?	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	échec

Types de données incomplètes

1. Données manquantes
2. Données partiellement observées (généralisation cas 1)
3. Données partiellement observées pondérées (généralisation cas 1-2)

Apprentissage des LPC : Données incomplètes

Types de données incomplètes

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
?	?	?	**1**	0	**1**	0
0	**1**	0	**1**	0	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
?	?	?	**1**	0	**1**	0
0	**1**	0	?	?	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**1**	0

Types de données incomplètes

1. Données manquantes
2. Données partiellement observées (généralisation cas 1)
3. Données partiellement observées pondérées (généralisation cas 1-2)

Apprentissage des LPC : Données incomplètes

Types de données incomplètes

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
?	**?**	**0**	**1**	0	**1**	0
0	**1**	0	**1**	0	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
0	**?**	**?**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**1**	0

Types de données incomplètes

1. Données manquantes
2. Données partiellement observées (généralisation cas 1)
3. Données partiellement observées pondérées (généralisation cas 1-2)

Apprentissage des LPC : Données incomplètes

Types de données incomplètes

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	**1**	**0**	**1**	0	**1**	0
0	**1**	0	**1**	0	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
0	**1/3**	**2/3**	**1**	0	**1**	0
0	**1**	0	**1**	**1**	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**1**	0

Types de données incomplètes

1. Données manquantes
2. Données partiellement observées (généralisation cas 1)
3. Données partiellement observées pondérées (généralisation cas 1-2)

Apprentissage des LPC : Données incomplètes

Types de données incomplètes (Rubin 1976)

Hypothèse MCAR : *Missing Completely At Random*

- La perte de données est issue d'un phénomène aléatoire indépendant des variables observées

$\Rightarrow$ Filtrage des données manquantes pour l'estimation des paramètres

Hypothèse MAR : *Missing At Random*

- Les variables ayant des observations manquantes dépendent de variables observées

$\Rightarrow$ Estimer les paramètres en utilisant l'algorithme EM

Hypothèse NMAR : *Not Missing At Random*

- Les variables ayant des observations manquantes dépendent de variables inconnues non observées

$\Rightarrow$ Identifier et ajouter de nouvelles variables pertinentes dans le modèle

$\Rightarrow$ Retour à l'hypothèse MAR

Apprentissage des LPC : Données incomplètes

Algorithme Expectation-Maximisation (Dempster, Laird et Rubin 1977)

Principes de l'algorithme EM

- Classe d'algorithmes d'optimisation itératifs
- Décomposition d'un problème d'optimisation complexe en deux sous problèmes d'optimisation alternés plus simples

Principale application

- Calculer les estimations des paramètres d'un modèle probabiliste par la méthode du maximum de vraisemblance en présence de données incomplètes

Propriétés

- Méthode d'optimisation locale

$\Rightarrow$ La convergence vers l'optimum global n'est pas garantie

$\Rightarrow$ La qualité de la solution calculée dépend de l'initialisation de l'algorithme

Apprentissage des LPC : Données incomplètes

Algorithme Expectation-Maximisation (Dempster, Laird et Rubin 1977)

Principes de l'algorithme EM

- Classe d'algorithmes d'optimisation itératifs
- Décomposition d'un problème d'optimisation complexe en deux sous problèmes d'optimisation alternés plus simples

Principale application

- Calculer les estimations des paramètres d'un modèle probabiliste par la méthode du maximum de vraisemblance en présence de données incomplètes

Propriétés

- Méthode d'optimisation locale

$\Rightarrow$ La convergence vers l'optimum global n'est pas garantie

$\Rightarrow$ La qualité de la solution calculée dépend de l'initialisation de l'algorithme

Apprentissage des LPC : Données incomplètes

Algorithme Expectation-Maximisation (Dempster, Laird et Rubin 1977)

Principes de l'algorithme EM

- Classe d'algorithmes d'optimisation itératifs
- Décomposition d'un problème d'optimisation complexe en deux sous problèmes d'optimisation alternés plus simples

Principale application

- Calculer les estimations des paramètres d'un modèle probabiliste par la méthode du maximum de vraisemblance en présence de données incomplètes

Propriétés

- Méthode d'optimisation locale

$\Rightarrow$ La convergence vers l'optimum global n'est pas garantie

$\Rightarrow$ La qualité de la solution calculée dépend de l'initialisation de l'algorithme

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Contexte

Contexte

- Données disponibles $\mathcal{D} = \mathcal{D}_C \cup \mathcal{D}_I$:
  - $\mathcal{D}_C$: Observations complètement observées
  - $\mathcal{D}_I$: Observations partiellement observées
- Modélisation : $\mathcal{D}$ = réalisations i.i.d. de la suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ représenté par un RB $\mathcal{M}$ de structure supposée connue

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	?	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
?	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	?

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Contexte

Contexte

- Données disponibles $\mathcal{D} = \mathcal{D}_C \cup \mathcal{D}_I$:
  - $\mathcal{D}_C$: Observations complètement observées
  - $\mathcal{D}_I$: Observations partiellement observées
- Modélisation : $\mathcal{D}$ = réalisations i.i.d. de la suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ représenté par un RB $\mathcal{M}$ de structure supposée connue

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	?	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
?	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	?

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Contexte

Contexte

- Données disponibles $\mathcal{D} = \mathcal{D}_\mathsf{C} \cup \mathcal{D}_\mathsf{I}$:
  - $\mathcal{D}_\mathsf{C}$: Observations complètement observées
  - $\mathcal{D}_\mathsf{I}$: Observations partiellement observées
- Modélisation : $\mathcal{D}$ = réalisations i.i.d. de la suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ représenté par un RB $\mathcal{M}$ de structure supposée connue

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	?	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
?	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	?

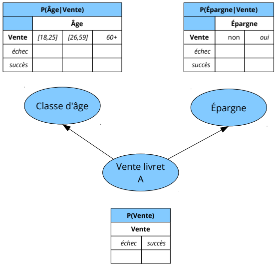
Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Contexte

Contexte

- Données disponibles $\mathcal{D} = \mathcal{D}_{\mathrm{C}} \cup \mathcal{D}_{\mathrm{I}}$:
  - $\mathcal{D}_{\mathrm{C}}$: Observations complètement observées
  - $\mathcal{D}_{\mathrm{I}}$: Observations partiellement observées
- Modélisation : $\mathcal{D} =$ réalisations i.i.d. de la suite de v.a. $\boldsymbol{X} = (X_1, \ldots, X_D)$ représenté par un RB $\mathcal{M}$ de structure supposée connue

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	?	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
?	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	?



Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Données d'entrée

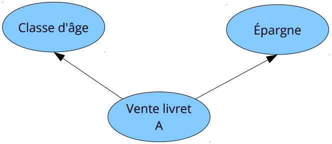
Données $\mathcal{D}$

A : Classe âge	E : Épargne	V : Vente livret A
[26,59]	non	échec
[18,25]	non	échec
[26,59]	non	échec
[26,59]	?	échec
60+	oui	succès
60+	oui	échec
[18,25]	oui	succès
60+	non	échec
?	non	échec
[26,59]	non	succès
[18,25]	non	succès
[26,59]	oui	succès
[26,59]	oui	échec
[26,59]	non	?

Modèle $\mathcal{M}$

P(Âge|Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec			
succès			

P(Épargne|Vente)		
	Épargne	
Vente	non	*oui*
échec		
succès		



P(Vente)	
Vente	
échec	*succès*

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Données d'entrée

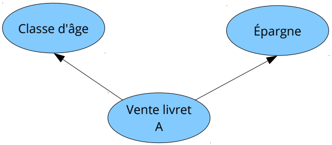
Données $\mathcal{D}$

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**?**	**?**	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
?	**?**	**?**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**?**	**?**

Modèle $\mathcal{M}$

P(Âge|Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec			
succès			

P(Épargne|Vente)		
	Épargne	
Vente	non	*oui*
échec		
succès		



P(Vente)	
Vente	
échec	*succès*

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Initialisation

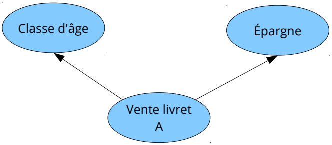
Données $\mathcal{D}^{(0)}$

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**?**	**?**	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
?	**?**	**?**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**?**	**?**

Modèle $\mathcal{M}^{(0)}$

P(Âge|Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec	**1/3**	**1/3**	**1/3**
succès	**1/3**	**1/3**	**1/3**

P(Épargne|Vente)		
	Épargne	
Vente	non	*oui*
échec	**1/2**	**1/2**
succès	**1/2**	**1/2**



P(Vente)	
Vente	
échec	*succès*
1/2	**1/2**

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Intération 1 - **Étape E**

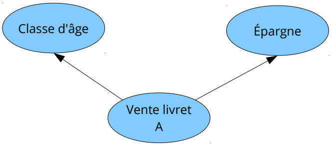
Données $\mathcal{D}^{(1)}$

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**1/2**	**1/2**	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
1/3	**1/3**	**1/3**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**1/2**	**1/2**

Modèle $\mathcal{M}^{(0)}$

P(Âge|Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec	1/3	1/3	1/3
succès	1/3	1/3	1/3

P(Épargne|Vente)		
	Épargne	
Vente	non	*oui*
échec	1/2	1/2
succès	1/2	1/2



P(Vente)	
Vente	
échec	*succès*
1/2	1/2

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Intération 1 - **Étape M**

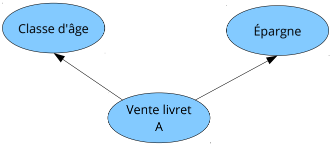
Données $\mathcal{D}^{(1)}$

A : Classe âge			**E : Épargne**		**V : Vente livret A**	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**1/2**	**1/2**	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
1/3	**1/3**	**1/3**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**1/2**	**1/2**

Modèle $\mathcal{M}^{(1)}$

$\hat{P}^{(1)}$(Âge | Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec	**7/24**	**10/24**	**7/24**
succès	**2/5**	**2/5**	**1/5**

$\hat{P}^{(1)}$(Épargne | Vente)		
	Épargne	
Vente	non	*oui*
échec	**5,5/8**	**2,5/8**
succès	**2/5**	**3/5**



$\hat{P}^{(1)}$(Vente)	
Vente	
échec	*succès*
8,5/14	**5,5/14**

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Intération 2 - **Étape E**

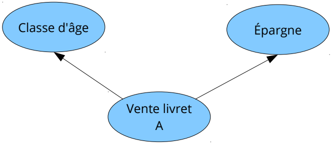
Données $\mathcal{D}^{(2)}$

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**0,6875**	**0,3125**	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
0,292	**0,417**	**0,292**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**0,735**	**0,265**

Modèle $\mathcal{M}^{(1)}$

$\hat{P}^{(1)}$(Âge|Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec	0,292	0,417	0,292
succès	0,4	0,4	0,2

$\hat{P}^{(1)}$(Épargne|Vente)		
	Épargne	
Vente	non	*oui*
échec	0,6875	0,3125
succès	0,4	0,6



$\hat{P}^{(1)}$(Vente)	
Vente	
échec	*succès*
0,607	0,393

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Intération 2 - **Étape M**

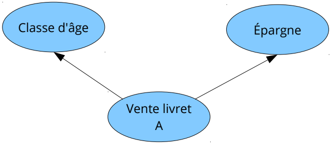
Données $\mathcal{D}^{(2)}$

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**0,6875**	**0,3125**	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
0,292	**0,417**	**0,292**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**0,735**	**0,265**

Modèle $\mathcal{M}^{(2)}$

$\hat{P}^{(2)}$(Âge | Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec	**0,286**	**0,428**	**0,286**
succès	0,4	0,4	0,2

$\hat{P}^{(2)}$(Épargne | Vente)		
	Épargne	
Vente	non	*oui*
échec	**0,711**	**0,289**
succès	0,4	0,6



$\hat{P}^{(2)}$(Vente)	
Vente	
échec	*succès*
0,624	**0,376**

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Illustration

Algorithme EM : Intération 3 - **Étape E**

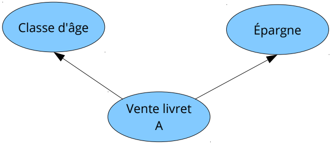
Données $\mathcal{D}^{(3)}$

A : Classe âge			E : Épargne		V : Vente livret A	
[18,25]	*[26,59]*	*60+*	*non*	*oui*	*échec*	*succès*
0	**1**	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
1	0	0	**1**	0	**1**	0
0	**1**	0	**0,711**	**0,289**	**1**	0
0	0	**1**	0	**1**	0	**1**
0	0	**1**	0	**1**	**1**	0
1	0	0	0	**1**	0	**1**
0	0	**1**	**1**	0	**1**	0
0,286	**0,428**	**0,286**	**1**	0	**1**	0
0	**1**	0	**1**	0	0	**1**
1	0	0	**1**	0	0	**1**
0	**1**	0	0	**1**	0	**1**
0	**1**	0	0	**1**	**1**	0
0	**1**	0	**1**	0	**0,759**	**0,241**

Modèle $\mathcal{M}^{(2)}$

$\hat{P}^{(2)}$(Âge|Vente)			
	Âge		
Vente	*[18,25]*	*[26,59]*	*60+*
échec	0,286	0,428	0,286
succès	0,4	0,4	0,2

$\hat{P}^{(2)}$(Épargne|Vente)		
	Épargne	
Vente	non	*oui*
échec	0,711	0,289
succès	0,4	0,6



$\hat{P}^{(2)}$(Vente)	
Vente	
échec	*succès*
0,624	0,376

Apprentissage des LPC : Données incomplètes

Algorithme EM dans les RB - Méthode

Algorithme EM dans les RB (structure connue)

- Initialisation du modèle RB $\mathcal{M}^{(0)}$ caractérisé par ses LPC
- **Étape E** : Estimer la loi des données manquantes à partir du RB courant, noté $\mathcal{M}^{(t)}$
  - Pour chaque donnée incomplète $\boldsymbol{x} = (\boldsymbol{x}_{\text{obs}}, \boldsymbol{x}_{\text{mqt}}) \in \mathcal{D}_{\text{I}}$
    - Calculer la distribution de ses variables manquantes $\boldsymbol{X}_{\text{mqt}}$ conditionnellement à ses variables observées $\boldsymbol{X}_{\text{obs}}$

    $\Rightarrow$ Calculer $P(\boldsymbol{X}_{\text{obs}} | \boldsymbol{X}_{\text{mqt}})$ en utilisant un algorithme d'inférence dans $\mathcal{M}^{(t)}$

  $\Rightarrow$ "Nouvelles" données courante $\mathcal{D}^{(t+1)} = \mathcal{D}_{\text{C}} \cup \mathcal{D}_{\text{I}}^{(t+1)}$ complétées par les distributions des variables manquantes
- **Étape M** : Estimer les LPC du modèle $\mathcal{M}^{(t+1)}$ à partir des données complétées $\mathcal{D}^{(t)}$

  $\Rightarrow$ Utilisation de la méthode du maximum de vraisemblance
- Répéter les étapes E et M tant que $\mathcal{M}^{(t)}$ et $\mathcal{M}^{(t+1)}$ sont significativement différents

Références

Dempster, A., N. Laird et D. B. Rubin (1977). “Maximum Likelihood from Incomplete Data Via the EM Algorithm”. In : *Journal of the Royal Statistical Society* B 39, p. 1-38.

Fisher, R. A. (1922). “On the Mathematical Foundations of Theoretical Statistics”. In : *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222, p. 309-368.

Rubin, D. B. (déc. 1976). “Inference and missing data”. In : *Biometrika* 63.3, p. 581-592.