Statistique Bayésienne

Anne Philippe

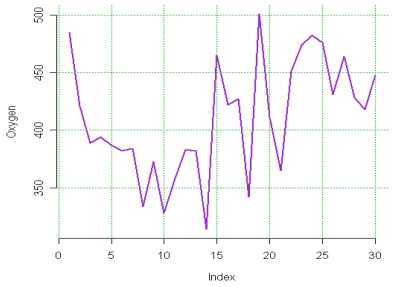
Laboratoire de Mathématiques Jean Leray
Université de Nantes

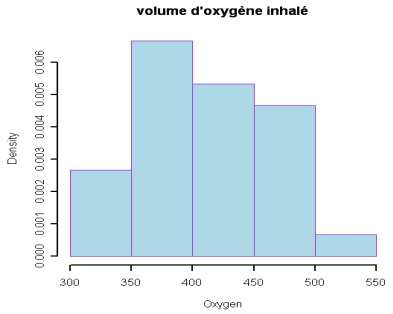
Automne 2007

Idée générale

$$x \sim P(\theta)$$

1. x est l'observation $\rightsquigarrow$ Connue
2. θ le paramètre inconnu, à estimer





Idée générale

- D'où ça vient ?
- Fondement des probabilités (début du 20°)
  - Fréquentiste
  - Subjectiviste
  - Logiciste
- Kolmogorov : espérance conditionnelle

Quelques références

1. Congdon, Peter Applied Bayesian modelling. Wiley Series in Probability and Statistics.
2. Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. "Bayesian Data Analysis" Chapman and Hall Texts in Statistical Science Series.
3. C.P. Robert The Bayesian Choice : from Decision-Theoretic Motivations to Computational Implementation (2001) Springer-Verlag, New York
4. C.P. Robert et G. Casella Monte Carlo Statistical Methods (1999) Springer-Verlag, New York.

Modèle paramétrique

Observations $x_1, \ldots, x_n$

$$x = (x_1, \ldots, x_n) \sim f_\theta(x), \qquad \theta \in \Theta \text{ est inconnu}$$

Objectif

on veut estimer le paramètre θ à partir de l'échantillon $x_1, \ldots x_n$.

Exemple

Observations suivant la loi normale $\mathcal{N}(m, \sigma^2)$ avec $\theta = (m, \sigma^2)$

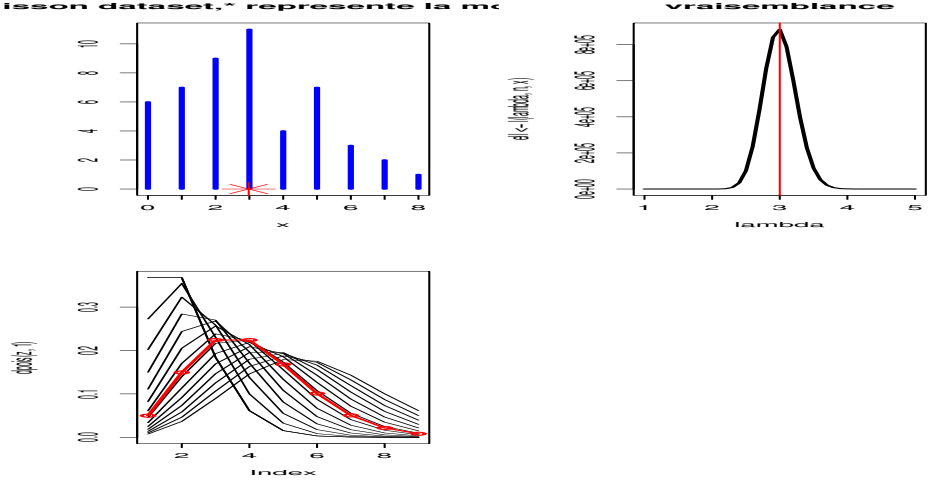
Une approche classique : le maximum de vraisemblance

la vraisemblance : c'est une fonction de Θ dans $\mathbb{R}^+$

$$\ell(\theta) \propto f_\theta(x)$$

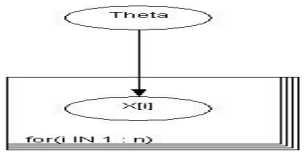
On cherche la valeur de θ qui maximise la vraisemblance.
c'est à dire on cherche la valeur de θ qui rend l'observation de x la plus probable.

Modèle de Poisson



Approche bayésienne

- Incertitude sur le paramètre θ est représentée par une probabilité π sur Θ.
- Le paramètre inconnu devient une variable aléatoire comme les observations



Définition

π est la loi a priori sur θ.

On interprète la loi des observations f_θ comme la loi conditionnelle des observations sachant θ

$$f(x|\theta) = f_\theta(x)$$

Théorème de Bayes

A et E des évènements $P(E) \neq 0$,
$P(A|E)$ et $P(E|A)$ sont liées par la relation

$$P(A|E) = P(E|A)\frac{P(A)}{P(E)}$$

Inversion des probabilités

Thomas Bayes, 1764

Inférence Bayésienne

La loi a priori sur θ : π
$+$
Observations suivant une loi $f(x|\theta)$
$\Downarrow$
On extrait des observations une information sur θ
On actualise la loi sur θ à partir des observations

$$\pi(\theta|x) = f(x|\theta)\frac{\pi(\theta)}{m(x)}.$$

Définition

La loi conditionnelle de θ sachant les observations x est appelée loi a posteriori

Modèle a priori

$$(\Theta, \pi(\theta))$$

$\downarrow$
modèle sur les observations

$$(\mathcal{X}, f(x|\theta))$$

$\downarrow$
Modèle a posteriori

$$(\Theta, \pi(\theta|x))$$

Pièces conformes

- X représente le nombre de pièces non-conformes dans un lot de taille n.
- La proportion p de pièces non conformes est inconnue

Question

Étant donné X, que peut on dire de p ?

Traduction Bayésienne

Loi *a priori* sur p : $p \sim \mathcal{U}([0,1])$

$$\pi(p) = \mathbb{I}_{[0,1]}(p)$$

Observation X : $X \sim \mathcal{B}(n,p)$

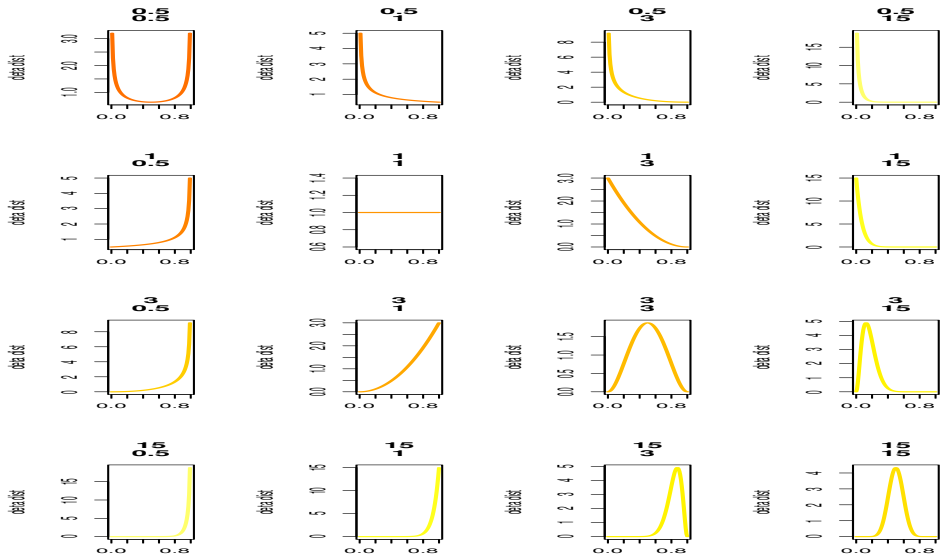
$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Loi *a posteriori* sur p : $p|x \sim \mathcal{B}e(x+1, n-x+1)$

loi Beta

$$\pi(p|X = x) \propto P(X = x|p)\pi(p) = p^x(1-p)^{n-x}\,\mathbb{I}_{[0,1]}(p)$$

Loi Beta $x \sim \mathcal{B}e(a,b)$, $\mathbb{E}(x) = \frac{a}{a+b}$ et $\mathrm{Var}(x) = \frac{ab}{(a+b)^2(a+b+1)}$



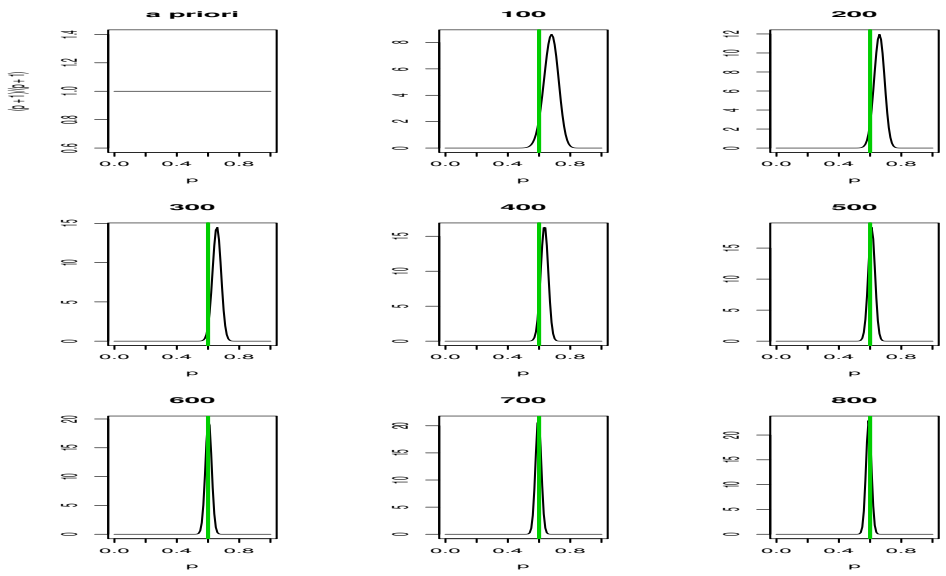
1. loi a priori sur p : loi uniforme
   la moyenne de p vaut $\frac{1}{2}$
2. On observe x nombre de pièces défectueuses

$$\Downarrow$$

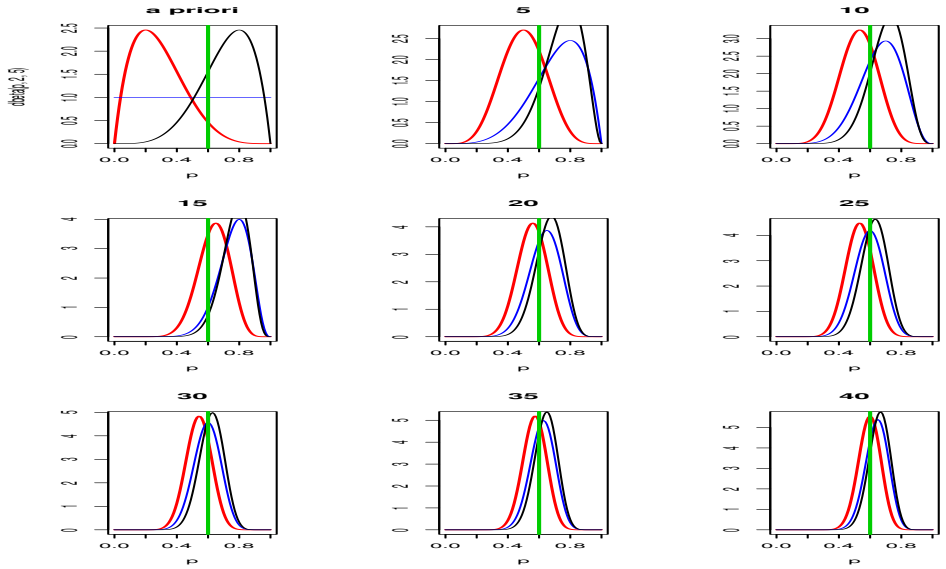
3. loi a posteriori sur p : loi beta
   la moyenne de p sachant x vaut

$$\mathbb{E}(p|x) = \frac{x+1}{n+2} = \frac{1}{2} - \frac{n}{2(n+1)} + \frac{x}{n+2}$$

la loi a priori uniforme $\rightsquigarrow$ suite des lois a posteriori quand le nb observations (n) varie



loi a priori favorisant $p < 1/2$ ou $p > 1/2$



Les lois qui interviennent ...

On se donne $f(x|\theta)$ et $\pi(\theta)$

- la loi jointe de (θ, x),

$$\varphi(\theta, x) = f(x|\theta)\pi(\theta)\,;$$

- la loi marginale de x,

$$m(x) = \int \varphi(\theta, x)\, d\theta = \int f(x|\theta)\pi(\theta)\, d\theta\,;$$

- la loi a posteriori de θ,

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}\,;$$

Évolution de l'information sur θ

$$\theta \sim \pi_0(\theta) \text{ [a priori]} + x_1, \ldots x_n \sim f(x|\theta) \text{ [}n\text{ mesures]}$$
$$\Downarrow$$
$$\theta|x \sim \pi_n(\theta|x_1, \ldots x_n) \text{ [a posteriori]}$$

Updater

$$\theta \sim \pi_n(\theta|x_1, \ldots x_n) \text{ [a priori]}$$
La loi a posteriori à l'étape n devient la loi a priori
$$+$$
$$x_{n+1} \sim f(x|\theta) \text{ [nouvelle observation]}$$
$$\Downarrow$$
$$\theta|x \sim \pi_n(\theta|x_1, \ldots x_n, x_{n+1}) \text{ [a posteriori]}$$

Choix de la loi a priori

On dispose d'informations sur θ

Question

Comment traduire cette information en loi a priori ?

Question

Comment traduire la qualité de cette information ?

!!! cas limite !!! : la loi a priori est concentrée sur $\{\theta_0\}$

$$\rightsquigarrow \pi(\theta|x) \equiv \pi(\theta)$$

Absence d'information : Approche non informative
On minimise le rôle de la loi a priori sur l'inférence

Détermination Subjective

modèle

X_t le nombre de pièces défectueuses dans un lot issu de la machine numéro t : $X_t \sim \mathcal{B}(n, p_t)$

Information a priori sur p_t : la proportion de pièces défectueuses.

	machine	1	2	3	4	5
p_t	Mean	0.3	0.4	0.5	0.2	0.2
	95% cred. int.	[0.1,0.5]	[0.2,0.6]	[0.3,0.7]	[0.05,0.4]	[0.05,0.4]

Si p_t suit une loi beta, on ajuste les paramètres pour que la moyenne et les quartiles coïncident avec nos informations

Time	1	2	3	4	5
Dist.	$\mathcal{B}e(6,14)$	$\mathcal{B}e(8,12)$	$\mathcal{B}e(12,12)$	$\mathcal{B}e(3.5,14)$	$\mathcal{B}e(3.5,14)$

Stratégie

On restreint le choix de π à une famille de lois paramétriques

$$\pi(\theta|\lambda) \qquad \lambda \in \Lambda$$

Définition

λ est appelé un hyper-paramètre

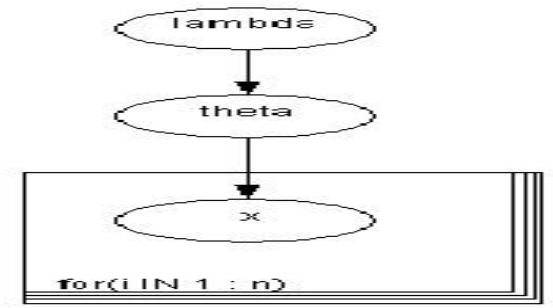
On fixe l'hyper-paramètre à partir de l'information que l'on possède sur les moments ou/et les quartiles

$$\lambda = \lambda_0$$

Alternative : Structure hiérarchique

On met une loi sur l'hyper paramètre λ :

- $\pi_{HP}(\lambda)$ de moyenne λ_0 et de variance τ
- le choix de τ traduit la confiance que l'on accorde à l'information contenue dans λ_0.



Lois conjuguées

$\mathcal{F}$ une famille de lois sur Θ

Définition

$\mathcal{F}$ est une famille conjuguée pour la vraisemblance $f(x|\theta)$
Si pour toute loi a priori $\pi \in \mathcal{F}$, la loi a posteriori $\pi(\theta|x) \in \mathcal{F}$.

- Préserve la structure sur la loi de θ
- l'information apportée par les observations se traduit uniquement par un changement de paramètres.

Famille Exponentielle

Cas particuliers : lois gaussiennes, betas binomiales ...

Définition

la densité est de la forme

$$f(x|\theta) = h(x)\exp\{\theta \cdot x - \psi(\theta)\},$$

Construction de la famille des lois a priori conjuguées :

$$\left\{\pi(\theta|\mu,\lambda) = K(\mu,\lambda)\, e^{\theta.\mu - \lambda\psi(\theta)}, \quad \lambda,\mu\right\}$$

A priori $(\mu,\lambda) \rightsquigarrow$ A posteriori $(\mu + x, \lambda + 1)$

les lois classiques

$f(x\|\theta)$ vraisemblance	$\pi(\theta)$ a priori	$\pi(\theta\|x)$ a posteriori
Normal $\mathcal{N}(\theta,\sigma^2)$	Normal $\mathcal{N}(\mu,\tau^2)$	Normal $\mathcal{N}(\rho(\sigma^2\mu+\tau^2x),\rho\sigma^2\tau^2)$ $\rho^{-1}=\sigma^2+\tau^2$
Binomial $\mathcal{B}(n,\theta)$	Beta $\mathcal{B}e(\alpha,\beta)$	Beta $\mathcal{B}e(\alpha+x,\beta+n-x)$
Poisson $\mathcal{P}(\theta)$	Gamma $\mathcal{G}(\alpha,\beta)$	Gamma $\mathcal{G}(\alpha+x,\beta+1)$
Normal $\mathcal{N}(\mu,1/\theta)$	Gamma $\mathcal{G}a(\alpha,\beta)$	Gamma $\mathcal{G}(\alpha+0.5,\beta+(\mu-x)^2/2)$

les lois classiques suite ...

$f(x\|\theta)$ vraisemblance	$\pi(\theta)$ a priori	$\pi(\theta\|x)$ a posteriori
Gamma $\mathcal{G}(\nu,\theta)$	Gamma $\mathcal{G}(\alpha,\beta)$	Gamma $\mathcal{G}(\alpha+\nu,\beta+x)$
Negative Binomial $\mathcal{N}eg(m,\theta)$	Beta $\mathcal{B}e(\alpha,\beta)$	Beta $\mathcal{B}e(\alpha+m,\beta+x)$
Multinomial $\mathcal{M}_k(\theta_1,\ldots,\theta_k)$	Dirichlet $\mathcal{D}(\alpha_1,\ldots,\alpha_k)$	Dirichlet $\mathcal{D}(\alpha_1+x_1,\ldots,\alpha_k+x_k)$

Lois non informatives

Question

Comment choisir la loi a priori lorsque l'on ne dispose pas d'information ?

On distingue trois grandes familles de lois

1. la loi uniforme (loi de Laplace)
2. maximisation d'un critère d'information (loi de Jeffrey)
3. argument fréquentiste (loi de concordance)

choix uniforme

$$\Theta = \{\theta_1, \cdots, \theta_p\} \qquad \pi(\theta_i) = 1/p$$

Extension au continu $\pi(\theta) \propto 1$

- La loi a priori n'est pas une probabilité
  mais si
  $$\int f(x|\theta)\ \mathrm{d}\theta < \infty$$
- on peut définir la loi a posteriori qui est bien une probabilité
- le choix dépend de la paramétrisation du modèle

Construction basée sur l'information

Principe : on maximise l'information apportée par les données
c'est-à-dire
on maximise la distance entre la loi priori et la loi a posteriori

$$\mathbb{E}^n\left[\int \pi(\theta|x_n)\log(\pi(\theta|x_n)/\pi(\theta))d\theta\right]$$

on obtient π_n, puis on prend la limite quand $n \to \infty$

Résultat

La loi dite de Jeffrey

$$\pi^*(\theta) \propto |I(\theta)|^{1/2}$$

où

$$I(\theta) = \mathbb{E}_\theta \left[\frac{\partial \ell}{\partial \theta^t} \frac{\partial \ell}{\partial \theta} \right]$$

Information de Fisher

Exemples

modèle Gaussien

- La variance est connue $\pi(\mu) \propto 1$
- La moyenne est connue $\pi(\sigma) \sim \sigma^{-1}$
- Les deux sont inconnues $\pi(\mu, \sigma) \sim \sigma^{-2}$

modèle binomial

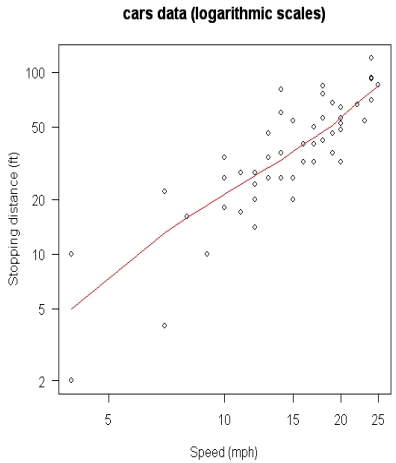
$x \sim \mathcal{B}(n, \theta)$ $\mathcal{B}e(1/2, 1/2)$

Un problème classique : la régression

On observe $x = (\text{vitesse}, \text{distance})$

$\log(\text{distance}) = a + b \log(\text{vitesse}) + \text{erreur}$

- $\theta = (a, b, \sigma^2)$
- $\log(\text{distance}) \sim \mathcal{N}(a + b \log(\text{vitesse}), \sigma^2)$



la régression : estimateurs classiques

On estime les paramètres par la méthode des moindres carrés
Voici le code R

```
> lm(log(dist) ~ log(speed), data = cars)

Call:
lm(formula = log(dist) ~ log(speed), data = cars)

Coefficients:
(Intercept)   log(speed)
    -0.7297       1.6024
```

la régression : approche bayésienne

Approche bayésienne non informative
voici le code R :

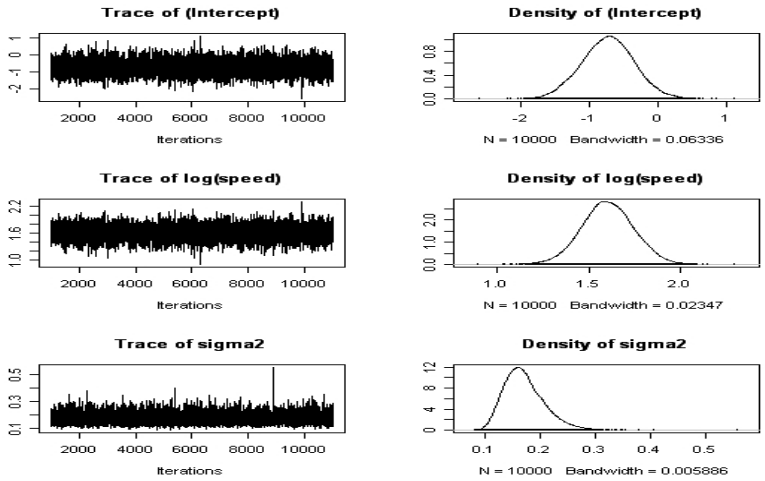
```
library(MCMCpack)
posterior  <- MCMCregress(log(dist) ~ log(speed), data = cars)
plot(posterior)

 Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean      SD  Naive SE Time-series SE
(Intercept) -0.7262 0.38441 0.0038441      0.0035905
log(speed)   1.6010 0.14294 0.0014294      0.0013524
sigma2       0.1719 0.03700 0.0003700      0.0004516
```


Le graphique suivant donne

- dans la colonne de gauche les chaînes de Markov simulées par un algorithme de Gibbs
- dans la colonne de droite la loi a posteriori marginale des différents paramètres



Argument fréquentiste

Concordance des régions de confiance :

On part d'une région de confiance fréquentiste $\{\theta \in C_x\}$ de niveau $1-\alpha$ c'est à dire

$$P_\theta(\theta \in \mathbb{C}_x) = \int_{C_x} f(x|\theta)\, dx = 1-\alpha$$

On cherche une loi a priori telle que la loi a posteriori vérifie

$$P(\theta \in C_x|x) \xrightarrow[n\to\infty]{} 1-\alpha$$

Loi de concordance en dimension 1

la loi de Jeffrey vérifie

$$P(\theta \le k_\alpha(x)|x) = 1-\alpha+O(n^{-1})$$

En dimension supérieure

on doit résoudre une équation de la forme

$$[I''(\theta)]^{-1/2} I'(\theta) \nabla \log \pi(\theta) + \nabla^t \{I'(\theta)[I''(\theta)]^{-1/2}\} = 0\,.$$

Performance des procédures d'estimation

Modèle : : $X \sim f(X|\theta)$ avec $X \in \mathcal{X}$ [Observation] et $\theta \in \Theta$ [inconnu]

Estimation de $g(\theta) \in \mathcal{D}$

Problèmes : :

- Évaluation du risque des procédures employées
- Comparaison des procédures employées

Question

Quelle est la meilleure procédure ?

Question

Existence ? Unicité ?

Risque / Coût

Elle repose sur l'existence d'une fonction de coût

$$\begin{array}{rccc} L: & \Theta \times \mathcal{D} & \to & \mathbb{R} \\ & (\theta, \delta) & \to & L(\theta, \delta) \end{array}$$

- L mesure l'erreur/la pénalité résultant de l'emploi de δ pour estimer $g(\theta)$

Définitions du risque d'un estimateur $\delta(x)$

1. Risque (fréquentiste) :

$$R(\theta, \delta) = \mathbb{E}_\theta(L[\theta, \delta(x)] = \int_{\mathcal{X}} L(\theta, \delta(x)) f(x|\theta) dx$$

2. Risque de Bayes :

$$\begin{aligned} r(\pi, \delta) &= \mathbb{E}^\pi[R(\theta, \delta)] = \int_\Theta R(\theta, \delta)\pi(\theta) d\theta \\ &= \int_\Theta \int_{\mathcal{X}} L(\theta, \delta(x)) f(x|\theta) dx \pi(\theta) d\theta \end{aligned}$$

Procédure de Bayes optimale

Étant donné :

- la loi des observations $x \sim f(x|\theta)$,
- la loi a priori π
- une fonction de coût L

On cherche l'estimateur qui minimise le risque bayésien

$$\delta^\pi(x) = \arg\min_d \mathbb{E}^\pi[L(\theta, d)|x] \ .$$

Définition

δ^π est l'estimateur de Bayes de θ associé à π

Remarque

L'estimateur de Bayes n'est pas nécessairement unique (p.s.)

Quelques exemples

Sous coût L^2

$$L(\delta, \theta) = |\theta - \delta|^2$$

l'estimateur de Bayes est égal à

$$\delta^\pi(x) = \mathbb{E}(\theta|x)$$

Sous coût L^1

$$L(\delta, \theta) = |\theta - \delta|$$

l'estimateur de Bayes est la médiane de loi *a posteriori*

Deux critères fréquentistes

1. Admissibilité
2. Minimaxité

Admissibilité

Définition

Un estimateur δ est admissible s'il n'existe pas δ' tq

$$R(\theta, \delta') \leq R(\theta, \delta) \ \forall\, \theta \qquad et \qquad R(\theta_0, \delta') < R(\theta_0, \delta)$$

Théorème

Si un estimateur de Bayes associé à une loi a priori π est unique, alors δ^π est admissible.

Minimaxité

Définition

Le risque minimax est donné par $\underline{R} = \inf_{\delta \in \mathcal{D}^*} \sup_{\theta} R(\theta, \delta)$

Définition

δ_0 *est un estimateur minimax si* $R(\theta, \delta_0) \leq \underline{R}$

Théorème

Un estimateur de Bayes qui a un risque fréquentiste constant est minimax.

Modèlisation par mélanges

Motivations

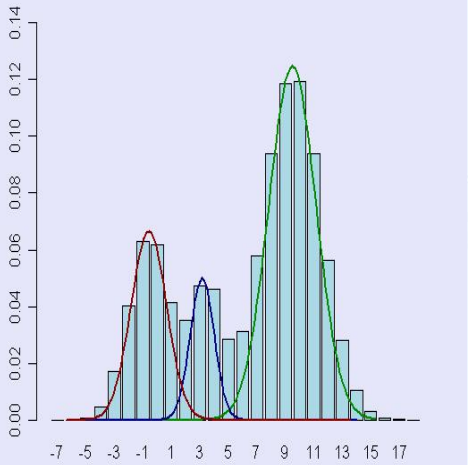
1. Phénomènes complexes // Structures multimodales
2. Populations hétérogènes et classes homogènes
3. Discrimination/Classification

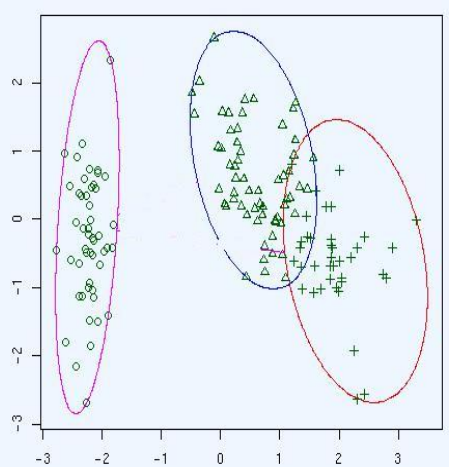
Définition

Le modèle admet une densité de la forme

$$g(x) = \sum_{i=1}^{k} p_i f(x|\theta_i) \ ,$$

avec la contrainte $p_1 + \ldots + p_k = 1$





Difficulté

Évaluation de la vraisemblance [k^n termes]

$$\mathrm{L}(\theta, \sigma, p|x) = \prod_{j=1}^{n} \left(\sum_{i=1}^{k} p_i f(x_j|\theta_i) \right),$$

- L'estimateur du maximum de vraisemblance ne peut pas être calculé facilement
- la loi a posteriori est difficile à évaluer

Données manquantes

$$x_1, \ldots, x_n \sim \sum_{i=1}^{k} p_i f(x|\theta_i),$$

On introduit les variables d'allocation :

z_i indicateur de la composante d'origine de x_i.

Réécriture du modèle :

$$x|z \sim f(x|\theta_z)$$

et

$$z \sim p_1 \mathbb{I}_{(z=1)} + \ldots + p_k \mathbb{I}_{(z=k)},$$

Choix de la loi a priori

Paramètres :

$$\{p_1, \ldots, p_k, \theta_1, \ldots, \theta_k, z_1, \ldots, z_n\}$$

On décompose la loi a priori de la forme suivante

$$\pi(p, \theta, z) = \pi(z|p, \theta_1, \ldots \theta_k)\pi(\theta_1, \ldots \theta_k, p) = \pi(z|p)\pi(\theta_1, \ldots, \theta_k, p)$$

où $\pi(z|p) \sim p_1 \mathbb{I}_{(z=1)} + \ldots + p_k \mathbb{I}_{(z=k)}$

- La loi de z sachant $p, \theta_1, \ldots \theta_k$ est indépendante de $\theta_1, \ldots \theta_k$
- p, $\theta_1, \ldots \theta_k$ sont indépendants

Choix de la loi a priori [suite]

Lorsque les composantes sont dans la famille exponentielle

$$f(x|\theta) = h(x)e^{\theta \cdot x - \psi(\theta)}, \qquad \theta \in \mathbb{R}^p,$$

on peut prendre pour chaque composante une loi a priori conjuguée

$$\pi(\theta|y_0, \lambda) \propto e^{\theta \cdot y_0 - \lambda\psi(\theta)}$$

et

$$(p_1, \ldots, p_k) \sim \mathrm{Dirichet}(\alpha_1, \ldots, \alpha_k)$$

de densité

$$\pi^D(p_1, \ldots, p_k) \propto p_1^{\alpha_1 - 1} \ldots p_k^{\alpha_k - 1} \mathbb{I}_{(p_1 + \ldots + p_k = 1)}.$$

Classification

On estime à partir de la loi a posteriori de z_i la composante d'origine de l'observation x_i.
Le critère est le suivant

On décide que l'observation x_i est issue de $f_{J(i)}$ où

$$J(i) = \text{argmax}_{\ell=1,\dots k} P(z_i = \ell | x_1, .., x_n)$$

Cas particulier : population à deux composantes

Il suffit de calculer $P(z_i = 1|x_1, .., x_n)$.
Si $P(z_i = 1|x_1, .., x_n) > 1/2$ alors on décide que la composante x_i est issue de la première composante.

Exemple du mélange de deux populations gaussiennes

le modèle s'écrit

$$p\mathcal{N}(m_1, \sigma_1^2) + (1-p)\mathcal{N}(m_2, \sigma_2^2) \qquad \text{ou} \qquad pf_1 + (1-p)f_2$$

On introduit des variables latentes

$$z = \begin{cases} 1 & \text{si } x \sim \mathcal{N}(m_1, \sigma_1^2) \\ 0 & \text{si } x \sim \mathcal{N}(m_2, \sigma_2^2) \end{cases}$$

Le choix de la loi a priori sur p est une loi beta de paramètres $(1, 1)$ [c'est aussi la loi uniforme].
Le choix des lois a priori sur μ_i et σ_i sont les lois conjuguées. [loi gaussienne sur μ_i et loi gamma sur σ_i^2]
Pour les z_i on prend

$$P(z_i = 1|p) = p, \qquad i = 1, ..., n$$

.

L'algorithme de Gibbs

On cherche à approcher la loi a posteriori des paramètres

$$p,\ \theta_1, \theta_2,\ z_i,\ i = 1...n$$

Les lois conditionnelles sont facilement simulables

$$z_i|x_1, ...x_n, p, \theta_1, \theta_2 \sim \text{bernoulli}\left(\frac{pf_1(x_i)}{pf_1(x_i) + (1-p)f_2(x_i)}\right)$$

$$p|x_1, ...x_n, z_1, ...z_n, \theta_1, \theta_2 \sim \text{beta}(1 + \sum_i z_i, 1 + n - \sum_i z_i)$$

$$\text{etc}$$

Approximation de la loi a posteriori par MCMC

A partir des N valeurs simulées par l'algorithme de Gibbs :

$$(z^{(1)}, p^{(1)}, m_1^{(1)}, m_2^{(1)}\sigma_1^{(1)}, \sigma_2^{(1)}) \dots (z^{(N)}, p^{(N)}, m_1^{(N)}, m_2^{(N)}\sigma_1^{(N)}, \sigma_2^{(N)})$$

Remarque

$z^{(j)}$ est un vecteur de taille n dont les coordonnées sont égales à 0 ou 1.

On peut faire une approximation de $\mathbb{E}(p|x_1 \dots, x_n)$, $\mathbb{E}(\mu_1|x_1 \dots, x_n)$ etc .... en prenant

$$\frac{1}{N}\sum_{j=1}^{N} p^{(j)} \qquad \text{etc}$$

Classification

On peut estimer à partir des valeurs $z_i^{(1)}, \dots z_i^{(N)}$ la probabilité $P(z_i = 1|x_1, .., x_n)$ en prenant

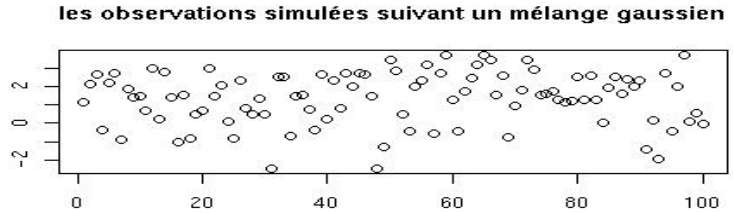
$$\hat{P}(z_i = 1|x_1, .., x_n) = \frac{1}{N} \sum_{j=1}^{N} \mathbb{I}_{z_i^{(j)}=1}$$

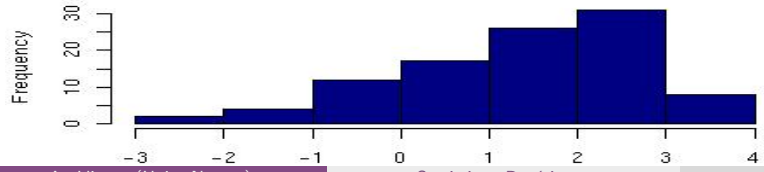
si $\hat{P}(z_i = 1|x_1, .., x_n) > 1/2$ *alors on décide que l'observation* x_i *est issue de la première composante* f_1

Les données simulées

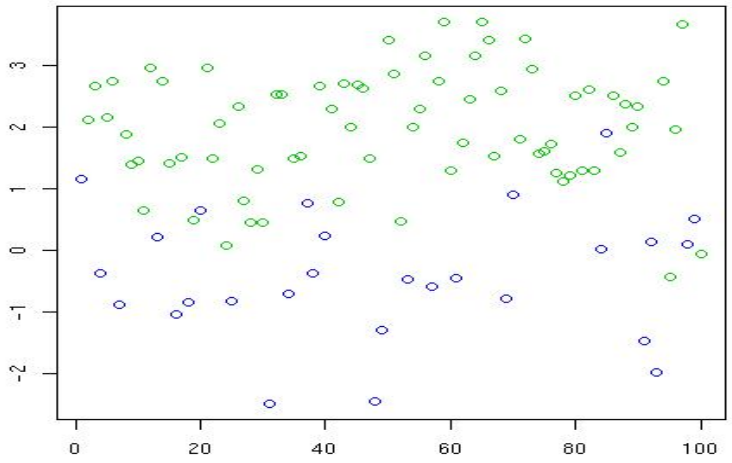
On simule suivant un mélange de deux lois gaussiennes

- la composante 1 est centrée et de variance 1
- la composante 2 est de moyenne 2 et de variance 1



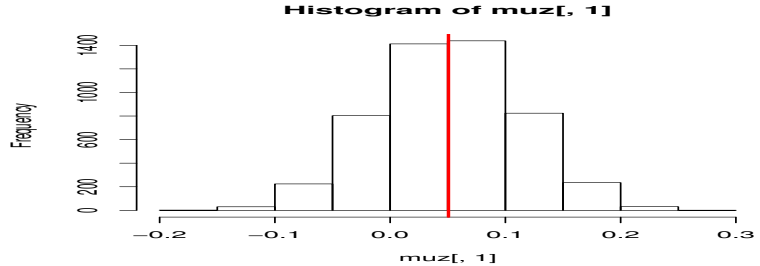


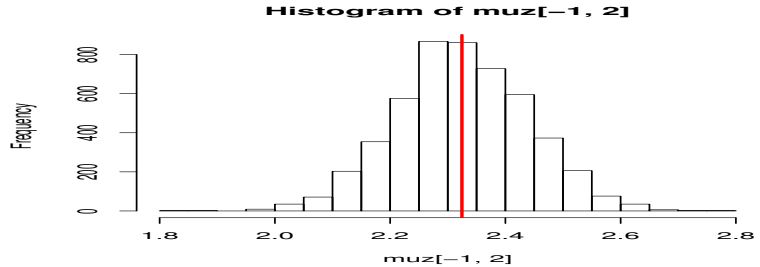
Les données manquantes ....



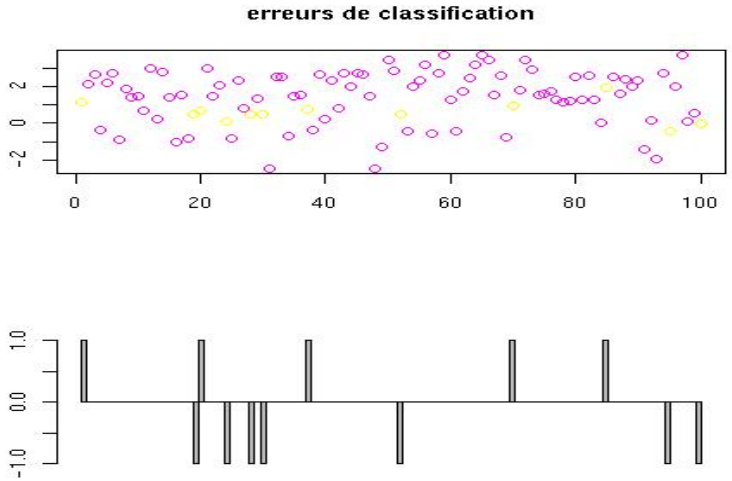
estimation des paramètres

Loi a posteriori pour les moyennes des composantes





Qualité de la classification



Estimation du nombre de composantes

Problème de sélection de modèles

Passage à k inconnu avec

$$k \sim Poi(\lambda)....$$

Le contexte

- On dispose d'une famille de modèles $\{M_i; i \in K\}$
- Pour chaque modèle, on dispose d'une structure paramétrique

Question

1. *Choix d'une variable discrète k correspondant à un modèle $M_k \in \{M_i; i \in K\}$, K fini.*
2. *Estimation du vecteur des paramètres $\theta^{(k)} \in \Theta_k \subset \mathbb{R}^{n_k}$ pour le modèle sélectionné M_k.*

Un problème de sélection de modèles ...

Soit $\mathbf{x}$ un processus ARMA d'ordre (p, q)

$$(x_t - \mu) - a_1(x_{t-1} - \mu) + \cdots - a_p(x_{t-p} - \mu) = \varepsilon_t - b_1\varepsilon_{t-1} + \cdots - b_q\varepsilon_{t-q}$$

où $(\varepsilon_t)_t$ est un bruit blanc gaussien centré.

1. estimation de p, q
2. estimation des coefficients, variance du bruit, moyenne

Les approches usuelles

elles utilisent un critère par exemple AIC, BIC ...
La démarche est la suivante

1. Estimation des paramètres pour chaque modèle.
2. Calcul du critère pour chaque modèle

On sélectionne le modèle qui minimise le critère

Approche bayésienne

k est inclus dans l'ensemble des paramètres.

$$p(\mathbf{x}, k, \theta^{(k)}) = \underbrace{\pi_0(k)\pi_1(\theta^{(k)}|k)}_{\text{loi a priori}}\underbrace{p(\mathbf{x}|\theta^{(k)}, k)}_{\text{vraisemblance}}$$

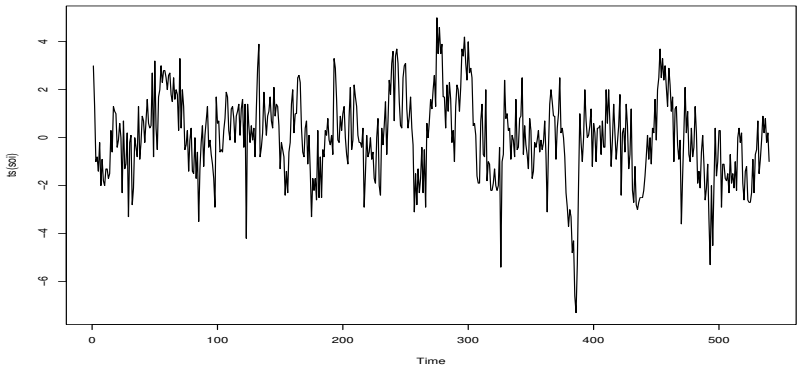
Estimateurs

1. Le paramètre discret est estimé par $\hat{k} = \underset{k \in K}{\operatorname{argmax}} P(k = k_0|\mathbf{x})$
2. pour chaque modèle M_k : son vecteur des paramètres $\theta^{(k)}$ est estimé par $\mathbb{E}(\theta^{(k)}|\mathbf{x}, k)$

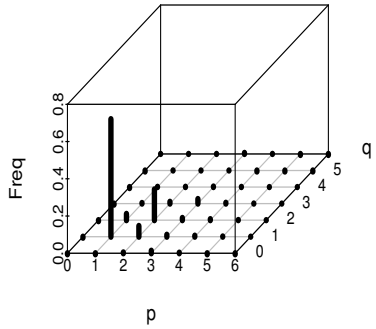
Cette approche nécessite en général la simulation de variables aléatoires en dimension variable.

- Algorithme d'Hasting Métropolis à sauts réversibles Green, 95
- Processus markovien de vie-et-mort Stephens, 00

This series consists of 540 observations on the Southern Oscillation Index (SOI), computed as the difference of the departure from the long-term monthly mean sea level pressures monthly mean sea level pressures. The index is one measure of the so-called "El Niño-Southern Oscillation", an event of critical importance and interest in climatological studies in recent decades.



Loi a posteriori des ordres



Pour cet exemple, on sélectionne le modèle ARMA (1,1)

La prévision

Tous les modèles contribuent au calcul de la prévision
Contrairement à l'approche classique où l'on calcule la prévision dans le modèle sélectionné.
Le prédicteur bayésien :

c'est un mélange de prédicteurs

Ayant observé $x_1, \ldots, x_n$, la densité prédictive de x_{n+1} est

$$f(y|x_1,\ldots,x_n) = \sum_k \pi(k|x_1,\ldots,x_n) \int f(y|x_1,\ldots,x_n,k\theta)\pi(\theta|x_1,\ldots,x_n,k)\,\mathrm{d}\theta$$

où

$$\int f(y|x_1,\ldots,x_n,k\theta)\pi(\theta|x_1,\ldots,x_n,k)\,\mathrm{d}\theta$$

correspond à la loi prédictive pour le modèle M_k