Statistique Bayésienne

Anne Philippe

Laboratoire de Mathématiques Jean Leray
Université de Nantes

Décembre 2017

anne.philippe@univ-nantes.fr

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Contexte

Modèle paramétrique

On observe une réalisation d'un vecteur aléatoire $x_1, \ldots, x_n$

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

On suppose que la famille de lois $\left\{f_\theta^{(n)};\ \theta \in \Theta\right\}$ est connue

Objectif :

l'estimation du paramètre θ à partir

1. des observations $x_1, \ldots x_n$
2. des informations complémentaires $\rightsquigarrow$ information a priori

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Quelques références

1. Congdon, Peter Applied Bayesian modelling. Wiley Series in Probability and Statistics. item Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. "Bayesian Data Analysis" Chapman and Hall Texts in Statistical Science Series.
2. Ghosh, J. K., Delampady, M., and Samanta, T. (2006). An Introduction to Bayesian Analysis, Theory and Methods. Springer.
3. Marin, J.-M. and Robert, C. (2007). Bayesian Core : A Practical Approach to Computational Bayesian Statistics. Springer.
4. C.P. Robert The Bayesian Choice : from Decision-Theoretic Motivations to Computational Implementation (2001) Springer-Verlag, New York
5. C.P. Robert et G. Casella Monte Carlo Statistical Methods (1999) Springer-Verlag, New York.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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

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

$\theta \sim \pi$

$X_1, \dots X_n$ observations

Définition

π est la loi a priori sur θ.

Inférence Bayésienne

loi a priori
$\theta \sim \pi$

Observations
$x_1 \dots x_n \sim f(x|\theta)$

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

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 séquentielle de la loi a posteriori

n observations

loi a priori
$\theta \sim \pi_0(\theta)$

Observations
$(x_1, \dots x_n) \sim f(x|\theta)$

Loi a posteriori
$\theta|x \sim \pi_n(\theta|x_1, \dots x_n)$

Mise à jour : on observe x_{n+1}

loi a priori
$\theta \sim \pi_n(\theta|x_1, \dots x_n)$

Observation
$x_{n+1} \sim f(x|\theta)$

Loi a posteriori
$\pi_{n+1}(\theta|x_1, \dots x_n, x_{n+1})$
$\propto \pi_n(\theta|x_1, \dots, x_n) f(x_{n+1}|\theta)$

Modèle binomial : Pièces conformes

- X représente le nombre de pièces non-conformes dans un lot de n pièces.
- La proportion p de pièces non conformes est inconnue :
  "toutes les valeurs sont équiprobables."

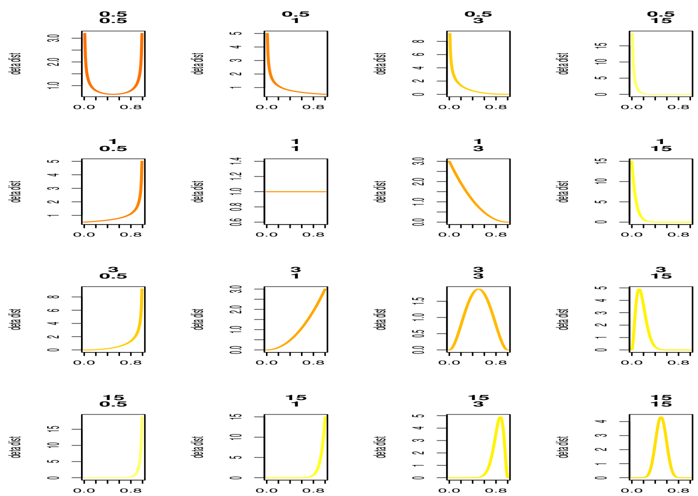
Traduction bayésienne

- La loi a priori : la loi uniforme $\pi(p) = \mathbb{I}_{[0,1]}(p)$
- 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 : c'est une loi beta
  $\mathcal{B}e(x+1, n-x+1)$

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

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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

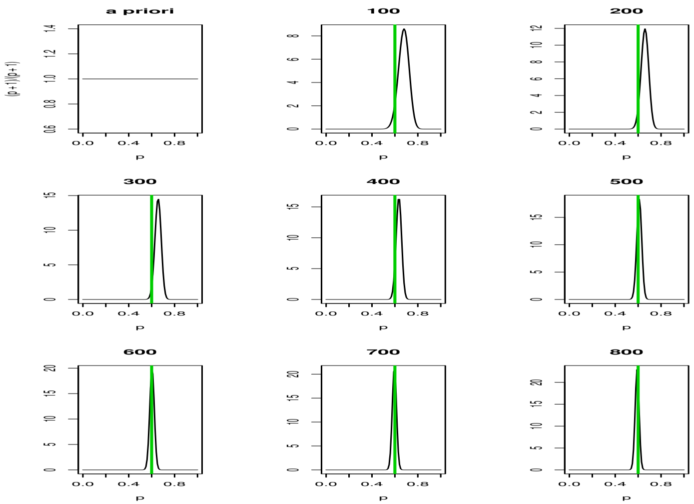
$$\Downarrow$$

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

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

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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



Modification de la loi a priori

On envisage deux situations : une loi a priori sur p

- favorisant $p < 1/2$
- favorisant $p > 1/2$

On suppose que p suit a priori une loi beta

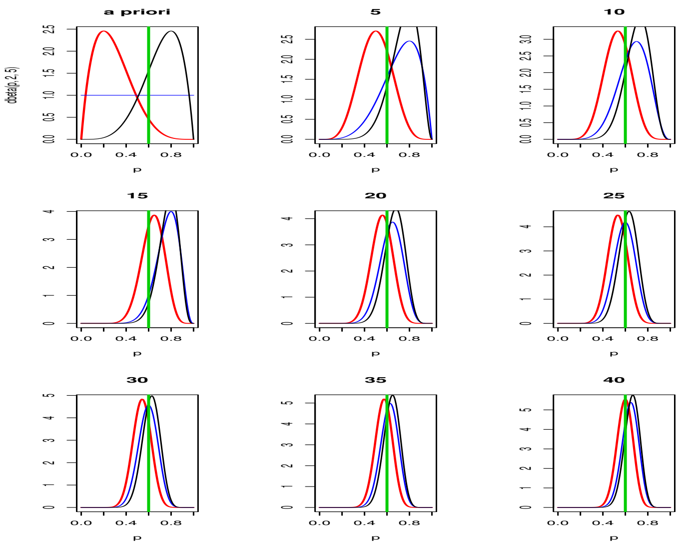
- loi a priori $p \sim Be(a, b)$
- loi a posteriori sur $p \sim Be(a + x, b + n - x)$

Choix de (a, b) :

- $a << b$ favorise les valeurs de $p < 1/2$
- $a >> b$ favorise les valeurs de $p > 1/2$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Comportement asymptotique de la loi a posteriori

Soit $X_1, ..., X_n$ iid suivant la loi f_θ et θ_0 la vraie valeur du paramètre.

Théorème

Soit $\pi(\cdot)$ la densité de la loi a priori sur $\theta \in \Theta \subset \mathbb{R}$, continue et positive en θ_0, $k_0 \in \mathbb{N}$. Soit U un voisinage de θ_0.
Pour un modèle régulier, on a

$$\int_U \pi(\theta|X_{1:n})\, \mathrm{d}\theta \to 1.$$

quand $n \to +\infty$,

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Théorème de Bernstein-von Mises

Théorème

Soit $\pi(\cdot)$ la densité de la loi a priori sur $\theta \in \Theta \subset \mathbb{R}$, continue et positive en θ_0, $k_0 \in \mathbb{N}$. Soit $\widehat{\theta}_n$ l'estimateur du maximum de vraisemblance et notons

$$t = n^{\frac{1}{2}}(\theta - \widehat{\theta}_n)$$

et $\widetilde{\pi}_n(\cdot|X_{1:n})$ la densité de la loi de t conditionnellement aux observations $X_{1:n}$, alors sous les hypothèses de régularité du modèle, nous avons

$$\int_{\mathbb{R}} \left| \widetilde{\pi}_n(t|X_{1:n}) - (2\pi)^{-\frac{1}{2}} |I(\theta_0)|^{\frac{1}{2}} e^{-\frac{1}{2}t' I(\theta_0) t} \right| \, \mathrm{d}t \to 0 \ ps,$$

où $I(\cdot)$ désigne l'information de Fisher du modèle et θ_0 la vraie valeur du paramètre.

Loi a priori impropre

Soit π une mesure positive et non une loi de probabilité sur Θ, i.e. $\pi(\theta) > 0$ pour tout $\theta \in \Theta$ et

$$\int \pi(\theta) \, \mathrm{d}\theta = +\infty.$$

Le cadre bayésien s'étend à un tel choix de loi a priori, dite impropre, dès lors que

$$m(x) := \int f(x|\theta)\pi(\theta) \, \mathrm{d}\theta < +\infty.$$

La loi a posteriori est définie par

$$\pi(\theta|x) = \pi(\theta) f(x|\theta) \frac{1}{m(x)}$$

C'est bien une loi de probabilité

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Approximation de Monte Carlo

L'inférence est calculée à partir de la loi a posteriori.
Elle utilise

- sa fonction de répartition
- sa densité
- sa fonction quantile
- sa constante de normalisation
- ses moments
- son mode

Ces quantités ne sont généralement pas connues explicitement et une approximation de Monte Carlo est nécessaire pour calculer

- les régions de confiance,
- les estimateurs de Bayes de $h(\theta)$
- les lois prédictives

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Description de la méthode

Simuler

$$\theta_1, \ldots, \theta_m \sim \pi(\theta|\, x) \text{ ou une loi qui approche } \pi(\theta|x)$$

Approche exacte ou méthode (MCMC)
A partir de l'échantillon simulé, on estime

- la densité $\pi(\theta|x)$ par un estimateur classique (histogramme / estimateur à noyau) calculé sur $\theta_1, \ldots, \theta_m$
- la fonction de répartition par le processus empirique

$$F_m(\theta|x) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{I}_{]-\infty,\theta]}(\theta_i)$$

- la fonction quantile par les quantiles empiriques de l'échantillon

$$Q_m(\alpha|x) = F_m^{-}(\alpha|x) = \inf\{t : \; F_m(t|x) \geq \alpha\}$$

Construction d'un estimateur de Bayes

Question

Quelle est la meilleure procédure pour estimer $g(\theta)$ à partir de la loi a posteriori ?

Elle repose sur la notion de risque

1. Risque (fréquentiste) :
   - quadratique

$$R(\theta,\delta) = \mathbb{E}_\theta((\theta - \delta(x))^2) = \int_{\mathcal{X}} (\theta - \delta(x))^2 f(x|\theta)dx$$

   - errreur absolue

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

2. Version Bayesienne du risque :

$$r(\pi,\delta) = \mathbb{E}^\pi[R(\theta,\delta)] = \int_\Theta R(\theta,\delta)\pi(\theta)d\theta$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Construction de l'estimateur de Bayes

étant donné :

- la loi des observations $x \sim f(x|\theta)$,
- la loi a priori π

On cherche l'estimateur δ^π qui minimise le risque bayésien

Risque quadratique

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

Risque absolu

$\delta^\pi(x)$ est la médiane de loi *a posteriori*

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Approximation des estimateurs par Monte Carlo

- A partir de l'échantillon simulé suivant la loi a posteriori

$$\theta_1, \ldots, \theta_m$$

- On approche $\mathbb{E}(h(\theta)|x)$ par

$$I_m = \underbrace{\frac{1}{m}\sum_{i=1}^{m} h(\theta_i)}_{\textit{Moyenne empirique}} \xrightarrow{m\to\infty} \mathbb{E}(h(\theta)|x)$$

Propriétés :

- $\mathbb{E}(I_m) = \mathbb{E}(\theta|x)$
- $\mathrm{Var}(I_m) = \frac{1}{m}\mathrm{Var}(\theta_1|x)$
  la vitesse $1/m$ est indépendante de la dimension

Remarque

Pour le risque absolu on approche l'estimateur par $Q_m(1/2|x)$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Famille exponentielle

La densité est de la forme : $f(x|\theta) = h(x)\exp\{\theta \cdot x - \psi(\theta)\}$
Cette famille de lois contient les lois gaussiennes, β, binomiales, Poisson....
Choix d'une loi a priori

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

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

On estime la moyenne de la loi $\mathbb{E}_\theta(X_1) = \psi(\theta)$

1. a priori
$$\mathbb{E}(\psi(\theta)) = \mu/\lambda$$
2. a posteriori
$$\mathbb{E}(\psi(\theta)|x_1,\ldots x_n) = \frac{\mu + \sum x_i}{n+\lambda}$$
3. quand n tend vers l'infini, $\mathbb{E}(\xi(\theta)|x_1,\ldots x_n)$ est équivalent à la moyenne empirique

Propriété asymptotique de l'estimateur de Bayes

Théorème

Sous les hypothèses du théorème de Théorème de Bernstein-von Mises.
De plus on suppose que

$$\int_\Theta \theta\pi(\theta)\, \mathrm{d}\theta < \infty$$

Nous avons

$$\sqrt{n}(\mathbb{E}(\theta|x_1,..x_n) - \widehat{\theta}_n) \to 0$$

presque sûrement

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Estimation d'un support de loi

On considère le modèle uniforme ; $X_1,\ldots,X_n$ des variables aléatoires iid suivant la loi uniforme sur $[0,\theta]$.
On veut estimer θ.
On suppose que θ suit une loi de Pareto (famille conjuguée)

$$\pi_{\alpha,\beta}(\theta) = \alpha\frac{\beta^\alpha}{\theta^{\alpha+1}}\mathbb{I}_{[\beta,+\infty[},\ \alpha > 1, \beta > 0.$$

Le choix de β est important

- Si $P_\theta(X_1 > \beta) = 0$ alors l'estimateur de Bayes est presque sûrement égal à $\beta\frac{\alpha+n}{\alpha+n-1}$.
  L'estimateur de Bayes n'est donc pas un estimateur consistant
- Si $P_\theta(X_1 > \beta) > 0$ alors l'estimateur de Bayes et l'estimateur du maximum de vraisemblance sont presque sûrement équivalents quand $n \to \infty$.
  Ils convergent vers la vraie valeur du paramètre.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

L'estimateur du maximum a posteriori (MAP)

S'il existe l'estimateur MAP est la valeur de θ qui maximise la densité de la loi a posteriori

$$\begin{aligned}\tilde{\theta}_n &= \arg\max \pi(\theta|x_1,..,x_n)\\ &= \arg\max \pi(\theta) f(x_1,..,x_n|\theta)\\ &= \arg\max \pi(\theta)\ell(\theta)\end{aligned}$$

- Si la loi a priori est la loi uniforme (ou loi impropre $\mathbb{I}_\Theta$) alors on retrouve l'EMV
- Pour un modèle régulier, si EMV est consistant alors le MAP est aussi consistant.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Intervalle de crédibilité

Soit θ unidimensionnel, on fixe $1-\alpha$ un niveau de confiance.

- Un intervalle de crédibilté $[a(x), b(x)]$ de niveau $1-\alpha$ vérifie la propriété

$$P(\theta \in [a(x), b(x)] | x) = \int_{a(x)}^{b(x)} \pi(\theta|x) = 1-\alpha$$

En pratique on utilise souvent les intervalles de crédibilité bilatéraux symétriques de niveau $1-\alpha$. Ils sont de la forme

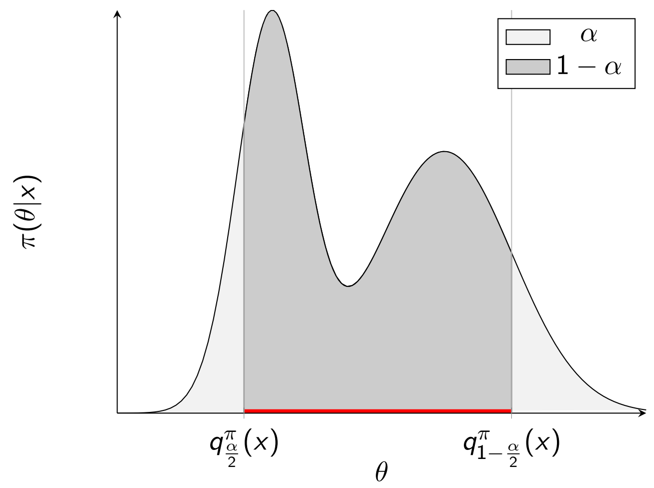
$$\theta \in \left[q^{\pi}_{\frac{\alpha}{2}}(x) \; ; \; q^{\pi}_{1-\frac{\alpha}{2}}(x)\right]$$

où $q^{\pi}_{\alpha}(x)$ est le quantile d'ordre α de la loi a posteriori.

$$\int_{-\infty}^{q^{\pi}_{\alpha}(x)} \pi(\theta|x) \, \mathrm{d}\theta = \alpha.$$

Illustration

Intervalle de crédibilité de niveau $1-\alpha$ sur θ



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Optimisation des intervalles de crédibilité

Il existe une infinité d'intervalles qui vérifient la condition

$$P(\theta \in [a(x), b(x)] | x) = \int_{a(x)}^{b(x)} \pi(\theta|x) = 1-\alpha$$

Ils sont de la forme

$$\theta \in \left[q^{\pi}_{\beta}(x) \; ; \; q^{\pi}_{1-\alpha+\beta}(x)\right]$$

pour tout $\beta \in [0, \alpha]$
Le plus court intervalle de crédibilité de niveau $1-\alpha$ est obtenu en prenant la valeur de β qui minimise

$$q^{\pi}_{1-\alpha+\beta}(x) - q^{\pi}_{\beta}(x)$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Approximation de Monte Carlo

A partir de l'échantillon simulé

$$\theta_1, \dots, \theta_m \sim \pi(\theta|x)$$

1. On estime les quantiles $q^{\pi}_{\beta}(x)$ par les quantiles empiriques de l'échantillon $Q_m(\beta)$ pour tout $\beta \in [0, \alpha]$
2. on cherche la valeur de β^* qui minimise $Q_m(1-\alpha+\beta) - Q_m(\beta)$
3. L'intervalle de crédibilité optimal approché est

$$\theta \in [Q_m(\beta^*), Q_m(1-\alpha+\beta^*)]$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Région HPD [1]

C'est une autre forme de régions de crédibilité construites sur les zones de plus haute densité a posteriori.
Ces régions sont de la forme

$$Q^{\pi}_{1-\alpha}(x) = \{\theta;\ \pi(\theta|x) \geq k^{\pi}_{1-\alpha}(x)\}\,,$$

où $k^{\pi}_{1-\alpha}(x)$ vérifie la relation

$$\int_{\{\{\theta;\,\pi(\theta|x)\geq k^{\pi}_{1-\alpha}(x)\}\}} \pi(\theta|x)\ \mathrm{d}\theta = 1-\alpha.$$

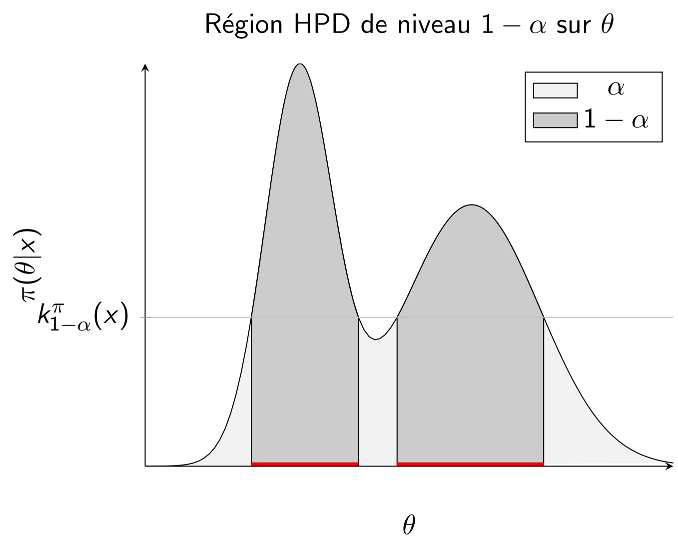
C'est la plus petit région de niveau $1-\alpha$

- Elles ne sont pas nécessairement connexes.
- Le paramètre θ peut être multidimensionnel.

1. highest posterior density

Illustration

Statistique Bayésienne
Anne Philippe



Approximation de Monte Carlo

Statistique Bayésienne
Anne Philippe

A partir de l'échantillon simulé

$$\theta_1, \dots, \theta_m \sim \pi(\theta|x)$$

1. On caclule les valeurs de $\eta_j = \pi(\theta_j|x)$ ou $\hat{\pi}(\theta_j|x)$ où $\hat{\pi}$ est l'estimateur de Monte Carlo de la densité.
2. On calcule K le quantile empirique d'ordre α de l'échantillon $\eta_j, j = 1, \dots, m$.
3. On prend comme région HPD, la région qui recouvre l'ensemble
$$\{\theta_i | i = 1, \dots, m,\ \eta_i > K\}$$

Couverture fréquentiste d'une région de crédibilité

Statistique Bayésienne
Anne Philippe

- Région de confiance fréquentiste de niveau $1-\alpha$:
  - $f_\theta(x)$ la loi des observations
  - on construit une statistique pivotale $T(x,\theta)$
    [sa loi de dépend pas de θ]
  - On construit I telle que
    $$P_\theta(T(x,\theta)\in I) = \int_{\{x:\ T(x,\theta)\in I\}} f_\theta(x)\ \mathrm{d}x = 1-\alpha$$
  - La région de confiance $\{\theta:\ \ T(x,\theta)\in I\}$
- Soit R_x une région de crédibilité bayésienne de niveau $1-\alpha$.
  - la couverture fréquentiste est donnée par
    $$P_\theta(\{x\ :\ \theta \in R_x\}) = \int_{\{x:\ \theta\in R_x\}} f_\theta(x)\ \mathrm{d}x = \beta(n,\theta)$$

Propriété asymptotique

La probabilité de couverture fréquentiste $\beta(n, \theta)$ diffère en général de $1 - \alpha$

Théorème

Sous des hypothèses générales de régularité, on montre que les intervalles de crédibilité unilatéraux et les régions HPD vérifient

$$\beta(n, \theta) = \int_{\{x:\ \theta \in R_x\}} f_\theta(x)\, \mathrm{d}x \to 1 - \alpha$$

quand le nombre n d'observations tend vers $+\infty$.

modèle de Poisson

- On suppose que les observations $x_1, \ldots, x_n$ sont iid suivant la loi de Poisson de paramètre θ
- On veut estimer le paramètre θ
- On suppose que la loi a priori est la loi Gamma de paramètre (a, b).

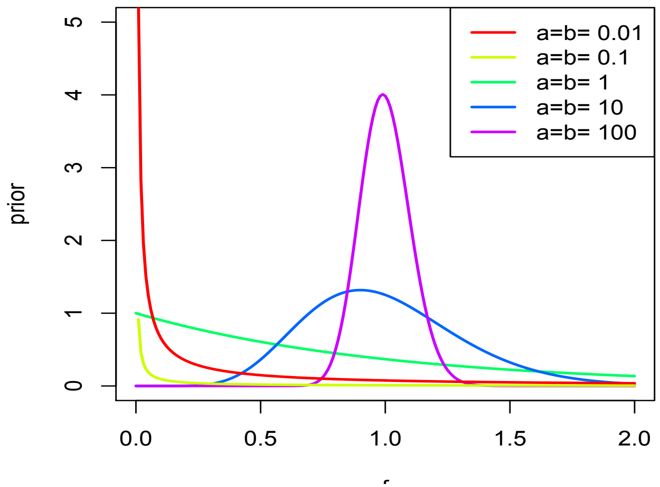
$[\mathbb{E}(\theta) = a/b$ et $\mathrm{Var}(\theta) = a/b^2$]

Information a priori :

- θ est autour de 1 $\rightsquigarrow$ $a = b$.
- On teste plusieurs choix de a. La loi se concentre autour de 1 quand a augmente

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Représentation de la loi a priori pour différents choix de a



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Les régions de crédibilité [$q_{2.5\%}(x); q_{97.5\%}(x)$]

On simule un échantillon suivant la loi de Poisson de paramètre 1 [cohérent avec notre information a priori].

- $n = 50$

a	0.010	0.100	1	10	100
$q_{2.5\%}(x)$	1.021	1.021	1.018	0.998	0.945
$q_{97.5\%}(x)$	1.657	1.656	1.646	1.567	1.281
$q_{97.5\%}(x) - q_{2.5\%}(x)$	0.636	0.635	0.628	0.569	0.337

- $n = 5000$

a	0.010	0.100	1	10	100
$q_{2.5\%}(x)$	1.002	1.002	1.002	1.002	1.002
$q_{97.5\%}(x)$	1.058	1.058	1.058	1.058	1.057
$q_{97.5\%}(x) - q_{2.5\%}(x)$	0.056	0.056	0.056	0.056	0.056

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Les régions de crédibilité [$q_{2.5\%}(x)$; $q_{97.5\%}(x)$]

On simule un échantillon suivant la loi de Poisson de paramètre 3 [mais la moyenne de la loi a priori est 1 et la variance tend vers 0 quand $a \to \infty$].

- $n = 50$

a	0.010	0.100	1	10	100
$q_{2.5\%}(x)$	2.760	2.757	2.724	2.454	1.542
$q_{97.5\%}(x)$	3.757	3.752	3.705	3.311	1.964

- $n = 5000$

a	0.010	0.100	1	10	100
$q_{2.5\%}(x)$	2.916	2.916	2.915	2.912	2.878
$q_{97.5\%}(x)$	3.011	3.011	3.011	3.007	2.972

Evaluation des probabilités de couverture fréquentiste

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Démarche :

- On simule $N = 5000$ échantillons suivant la loi de Poisson de paramètre θ_0
- Pour chaque échantillon on calcule la région de crédibilité
- on estime $\beta(n, \theta_0)$ par la proportion d'intervalles qui contiennent la valeur θ_0

Résultats pour $\theta_0 = 1$

a	0.010	0.100	1	10	100
$n = 10$	0.953	0.952	0.960	0.994	1.000
$n = 50$	0.941	0.944	0.945	0.969	1.000
$n = 500$	0.948	0.943	0.946	0.955	0.968

[suite]

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Résultats pour $\theta_0 = 3$

a	0.010	0.100	1	10	100
$n = 10$	0.941	0.946	0.942	0.160	0.000
$n = 500$	0.945	0.947	0.953	0.918	0.001
$n = 5000$	0.950	0.954	0.953	0.946	0.635
$n = 50000$	0.950	0.951	0.951	0.951	0.919

Modèle Gaussien

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

On dispose de n observations $X_1, \ldots, X_n$ iid suivant une loi gaussienne $\mathcal{N}(\theta, 1)$.

- On choisit comme loi a priori sur θ la loi gaussienne $\mathcal{N}(0, \tau^{-2})$, $\tau > 0$
- loi a posteriori est une loi Gaussienne

$$\mathcal{N}\left(\frac{\bar{X}_n}{1+\tau^2/n}, \frac{1}{n+\tau^2}\right)$$

- les régions HPD de niveau $1-\alpha(= .95)$ sont de la forme

$$\theta \in \left[\frac{\bar{X}_n}{1+\tau^2/n} - \frac{u_{1-\alpha/2}}{\sqrt{n+\tau^2}}; \frac{\bar{X}_n}{1+\tau^2/n} + \frac{u_{1-\alpha/2}}{\sqrt{n+\tau^2}}\right]$$

où u_α est le quantile d'ordre α de la loi gaussienne standard.

Modèle Gaussien (suite)

- Couverture fréquentiste de la région HPD

$$P_\theta(\theta \in I^{HPD}(\tau, \bar{X}_n)) =$$
$$F\left(\frac{\theta\tau^2}{\sqrt{n}} + u_{1-\alpha/2}\sqrt{\frac{n+\tau^2}{n}}\right) - F\left(\frac{\theta\tau^2}{\sqrt{n}} - u_{1-\alpha/2}\sqrt{\frac{n+\tau^2}{n}}\right)$$

  où F est la fonction de répartition de la loi gaussienne standard.
- Comportement asymptotique :
  - cette probabilité tend vers $1-\alpha$ quand $n \to \infty$.
  - cette probabilité tend vers $1-\alpha$ quand $\tau \to 0$.

Prédicteur ponctuel optimal

Le modèle : :

- observations : $(x_1, \ldots, x_n) \sim f^{(n)}(x|\theta)$
- π : la loi a priori
- la loi a posteriori

$$\pi(\theta|x_1, \ldots, x_n) \propto \pi(\theta) f^{(n)}(x|\theta)$$

Objectif : :

- On veut prévoir x_{n+1} à partir des observations passées $(x_1, \ldots, x_n)$.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Prédicteur en loi

Loi prédictive

c'est la loi conditionnelle de x_{n+1} sachant $(x_1, \ldots, x_n)$ c'est à dire

$$p(x_{n+1}|(x_1, \ldots, x_n)) = \int_\Theta f(x_{n+1}|\theta, (x_1, \ldots, x_n))\pi(\theta|x_1, \ldots, x_n)\, \mathrm{d}\theta$$

- On effectue de la prévision en loi
- On obtient des intervalles de prévision.
- Pour tout niveau $1-\alpha$, on calcule l'intervalle $[a_{n+1}, b_{n+1}]$ tel que

$$P(X_{n+1} \in [a_{n+1}, b_{n+1}]|x_1, .., x_n) = 1-\alpha$$

- solution 1 : a_{n+1} b_{n+1} sont respectivement les quantiles d'ordre $\alpha/2$, $1-\alpha/2$ de la loi predictive
- solution 2 : on optimise en cherchant le plus court intervalle

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Prédicteur ponctuel bayésien

Pour l'erreur quadratique, la meilleure approximation de x_{n+1} à partir d'une fonction de $(x_1, \ldots, x_n)$ est l'espérance conditionnelle

$$\hat{x}_{n+1} = \mathbb{E}(x_{n+1}|(x_1, \ldots, x_n))$$

On a donc le prédicteur ponctuel bayésien suivant

$$\begin{aligned}\hat{x}_{n+1} &= \int x_{n+1} \int_\theta f(x_{n+1}|\theta, (x_1, \ldots, x_n))\pi(\theta|x_1, \ldots, x_n)\, \mathrm{d}\theta\, \mathrm{d}x_{n+1} \\ &= \int_\Theta \pi(\theta|x_1, \ldots, x_n) \int x_{n+1} f(x_{n+1}|\theta, x_1, \ldots, x_n)\, \mathrm{d}x_{n+1}\, \mathrm{d}\theta\end{aligned}$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Prédicteur ponctuel fréquentiste

- On observe $x_1, \ldots, x_n$ de loi $f(x_1, \ldots, x_n|\theta)$ où θ est inconnu.
- Le meilleur prédicteur (au sens de l'erreur quadratique) de x_{n+1} est l'espérance conditionnelle

$$\begin{aligned}\hat{x}^f_{n+1}(\theta) &= \mathbb{E}(x_{n+1}|\theta, x_1, \ldots, x_n)\\ &= \int x_{n+1} f(x_{n+1}|\theta, x_1, \ldots, x_n)\, \mathrm{d}x_{n+1}\end{aligned}$$

- Le prédicteur optimal dépend de θ inconnu.
- En pratique
  - on commence par estimer θ par exemple par le maximum de vraisemblance $\hat{\theta}$
  - le prédicteur retenu est $\hat{x}^f_{n+1}(\hat{\theta})$.

Comparaison bayesien // fréquentiste

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

1. Prédicteur optimal si θ est connu :

$$\hat{x}^f_{n+1}(\theta) = \int x_{n+1} f(x_{n+1}|\theta, x_1, \ldots, x_n)\, \mathrm{d}x_{n+1}$$

2. Le prédicteur fréquentiste s'écrit

$$\hat{x}^f_{n+1}(\hat{\theta})$$

3. Le prédicteur bayésien s'écrit

$$\hat{x}_{n+1} = \int \hat{x}^f_{n+1}(\theta)\pi(\theta|x_1, \ldots, x_n)\, \mathrm{d}\theta$$

C'est un mélange de prédicteur où les poids sont donnés par la densité a posteriori

Approximation de Monte Carlo

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

A partir de l'échantillon simulé

$$\theta_1, \ldots, \theta_m \sim \pi(\theta|x)$$

1. On simule un échantillon suivant la loi prédictive en prenant

$$x_{n+1}(i) \sim f(x_{n+1}|\theta_i, x_1, \ldots, x_n), \text{ pour tout } i = 1, \ldots, m$$

2. On approche le predicteur ponctuel par la moyenne de l'échantillon $\frac{1}{m}\sum_{i=1}^m x_{n+1}(i)$
3. On approche la densité de la loi prédictive en calculant l'estimateur à noyau sur l'échantillon $x_{n+1}(i),\ i = 1, \ldots, m$
4. Les intervalles prédictifs sont approchés à l'aide des quantiles empiriques de l'échantillon $x_{n+1}(i),\ i = 1, \ldots, m$

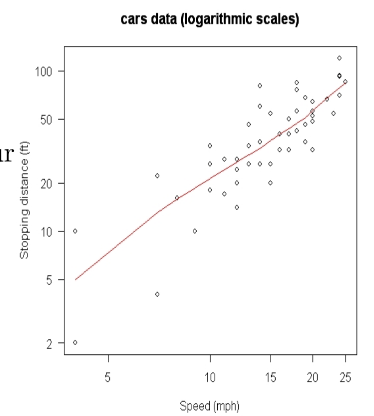
Un problème classique : la régression

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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

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

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



la régression : approche bayésienne

Approche bayésienne non informative
La loi a priori de Jeffreys est $\pi(a, b, \sigma) \propto \frac{1}{\sigma^2}$
La loi a priori est impropre mais la loi a posteriori est bien définie.

```
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
```


la régression : estimateurs classiques

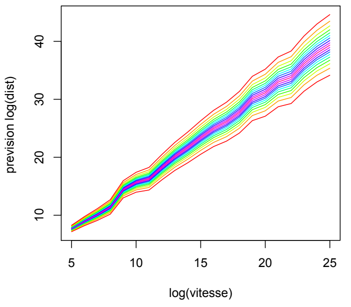
On estime les paramètres par la méthode des moindres carrés
Voici le code R

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

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

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

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

prévision bayésienne

$\rightsquigarrow$ prévision en loi.
On représente les quantiles d'ordre 5% à 95% de la loi prédictive



- On a représenté les intervalles de prévision de niveau 90%, 80% etc
- Rouge : intervalle à 90%

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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 ?

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

Approche informative

On utilise de l'information provenant

- de la connaissance des experts
- d'autres études statistiques menées dans un contexte similaire.
- de données historiques non utilisées dans l'étude.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Utilisation de plusieurs échantillons historiques

On veut estimer θ à partir d'un petit échantillon $x \in f(x|\theta)$
On dispose de K échantillons historiques x_i^H, i=1,...,K

- On estime le paramètre θ sur chacun des échantillons $x_i^H \rightsquigarrow \hat\theta_1, \dots, \hat\theta_K$.
- On suppose que $\hat\theta_1, \dots, \hat\theta_K$ sont iid suivant la loi a priori.
- On construit la loi a priori à partir de $\hat\theta_1, \dots, \hat\theta_K$ en prenant
  - l'histogramme des $\hat\theta_1, \dots, \hat\theta_K$.
  - l'estimateur à noyau de la densité calculée sur $\hat\theta_1, \dots, \hat\theta_K$.
  - $\pi \in \{\pi_\lambda, \lambda \in \Lambda\}$ on estime λ à partir des données $\hat\theta_1, \dots, \hat\theta_K$ par un estimateur paramétrique (EMV..)

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Utilisation d'un historique long

On veut estimer θ à partir d'un petit échantillon $x \sim f(x|\theta)$
On dispose d'un long échantillon historique $x^H \sim f(x^H|\theta)$
On suppose que les deux échantillons ont la même loi

- On estime θ à partir de l'échantillon (x, x^H)

$$\pi(\theta|x, x^H) \propto f(x|\theta) f(x^H|\theta) \pi_0(\theta)$$

- On réduit la contribution de l'historique en prenant comme loi a priori

$$\pi(\theta) \propto f(x^H|\theta)^\alpha \pi_0(\theta)$$

  où α est un niveau de confiance accordé à l'historique
- On prend une loi paramétrée dont le paramètre est fixé à partir de x^H. Par exemple la loi a priori est de moyenne l'estimation de θ à partir de l'échantillon x^H.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Estimating a Heart Transplant Mortality Rate

The number of deaths y within 30 days after the Heart transplant.

$$y \sim Poisson(\lambda E)$$

where E is the number of patients
and where

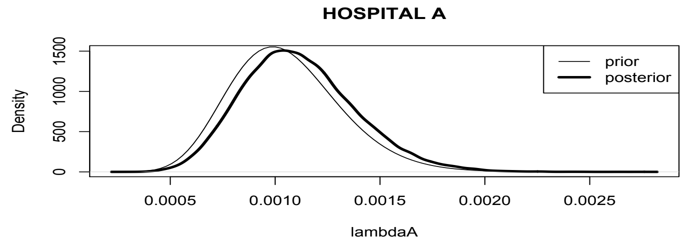
$$\lambda \text{ is the mortality rate per unit.}$$

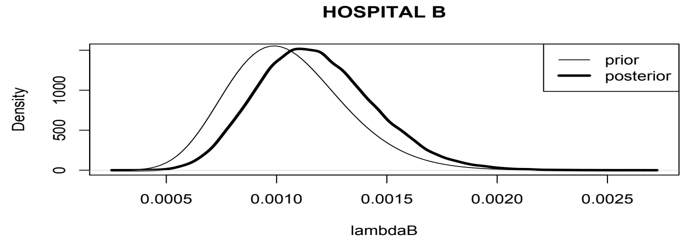
Prior distribution on λ : Gamma distribution
We fix the parameter from the observed data in 10 hospitals

- $a = \sum y_i = 16$ (number of deaths)
- $b = \sum e_i = 15174$ (number of patients)

$$\pi(\lambda) = b^a \frac{1}{\Gamma(a)} e^{-b\lambda} \lambda^{a-1} \qquad \mathbb{E}(\lambda) = a/b \qquad \mathrm{Var}(\lambda) = a/b^2$$

Hospital A : y = 1 and E= 66.
Standard estimate of mortality rate 1/66
Hospital B : y = 4 and E = 1767.
standard estimate of mortality rate ≈1/450.





Proportion of heavy sleepers

We want to estimate the proportion p of population who sleep at least 8 hours.

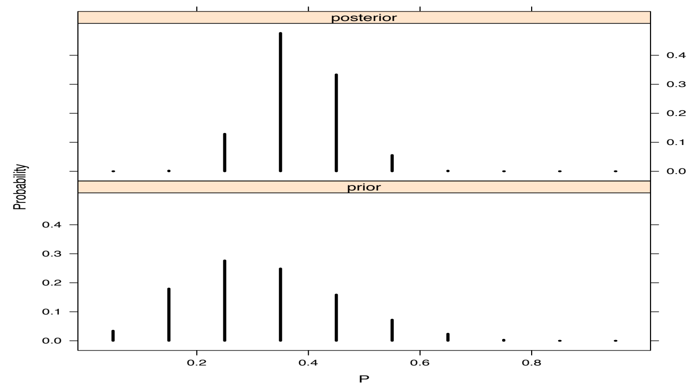
- The observations : 27 students such that
  s=11 : ≥ 8 hours
  f=16 : < 8 hours
- The likelihood function is given by

$$L(p) = p^s(1-p)^f$$

- Choice of the prior for p from expert information :
  1. Discrete prior
  2. Histogram prior
  3. Continuous prior

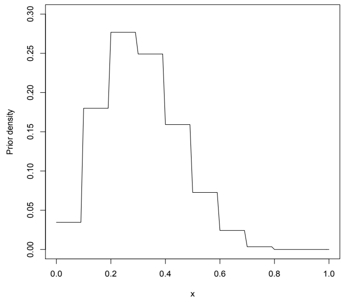
Discrete prior

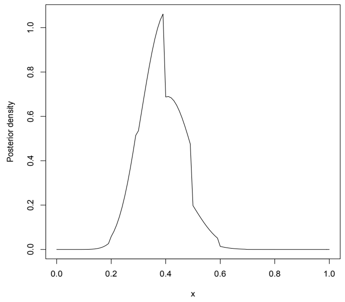
We fix a
list of plausible proportion values and then assign probabilities



Histogram prior distribution

The prior beliefs : the range of p is divided into intervals and we assign probabilities to the intervals.





PRIOR POSTERIOR

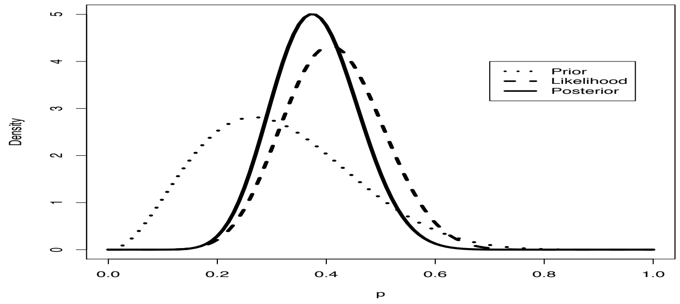
Continuous prior for p

Initial beliefs :

- the median is around 0.3
- p is less than 0.5 with probability .90

We assume than the prior distribution is a beta distribution with parameter

$$a = 3.4 \text{ and } b = 7.4.$$



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

predictive distributions

- Comparison of predictive distributions for two prior distributions
  - beta distribution
  - discrete distribution
- Goal : The distribution the number of heavy sleepers Y in a future sample of size $m = 27$.
- The observations : $(s, f) = (11, 16)$

The predictive distribution is of the form, for all $y = 0, \dots, m$

- Continuous :

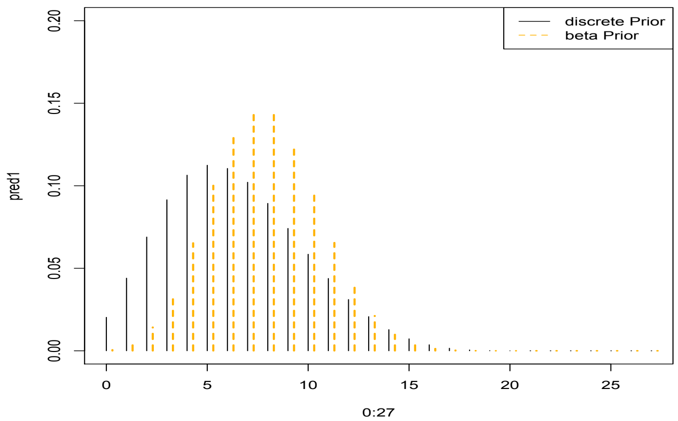
$$\mathbb{P}(Y = y|(f,s)) = \int f(y|p)\pi(p|(f,s))\,\mathrm{d}p$$

- Discrete :

$$\mathbb{P}(Y = y|(f,s)) = \sum_i f(y|p_i)\mathbb{P}(p = p_i|(f,s))$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Comparison of predictive distributions



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

A priori paramétrique

On restreint le choix de π à une famille paramétrique

$$\pi \in \{\pi_\lambda,\ \lambda \in \Lambda\}$$

Définition

λ est appelé un hyper-paramètre

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

$$\lambda = \lambda_0 \to \theta \sim \pi_{\lambda_0} \to X_1, \ldots X_n \text{ observations } f(x|\theta)$$

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.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

modèle de Poisson

- $x = (x_1, \ldots, x_n)$ iid suivant la loi de Poisson

$$f(x|\theta) = e^{-n\theta}\theta^{\sum x_i}\frac{1}{\prod x_i!}$$

- la loi a posteriori

$$\pi(\theta|x) \propto \pi(\theta)e^{-n\theta}\theta^{\sum x_i}$$

$$\rightsquigarrow \quad \pi(\theta) \propto e^{-b\theta}\theta^{a-1}$$

- On reconnait la densité d'une loi gamma

a priori	a posteriori
a	$a + \sum x_i$
b	$b + n$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

modèle uniforme

- Soit $X_1, \ldots, X_n$ des variables aléatoires iid suivant la loi uniforme sur $[0, \theta]$.

$$f(x|\theta) \propto \theta^{-n}\,\mathbb{I}_{[\max(x_1,\ldots,x_n),\infty[}(\theta)$$

- La famille des lois de Pareto est une famille de lois conjuguées

$$\pi_{\alpha,\beta}(\theta) = \alpha\frac{\beta^\alpha}{\theta^{\alpha+1}}\,\mathbb{I}_{[\beta,+\infty[}(\theta)$$

avec $\alpha > 1$ et $\beta > 0$

- évolution des paramètres :

a priori	a posteriori
α	$\alpha + n$
β	$\max(\beta, X_1, \ldots, X_n)$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

modèle gaussien

- $x = (x_1, \dots, x_n)$ iid suivant une loi gaussienne de variance connue.

$$\pi(\mu|x) \propto \pi(\mu) e^{\frac{-1}{2\sigma^2}(n\mu^2 - 2\mu \sum x_i)}$$

$\rightsquigarrow$ μ suit une loi gaussienne

- $x = (x_1, \dots, x_n)$ iid suivant une loi gaussienne de moyenne connue.

$$\pi(\sigma^2|x) \propto \pi(\sigma^2) e^{\frac{-1}{2\sigma^2} S_n} \frac{1}{\sigma^n}$$

$\rightsquigarrow$ σ^2 suit une loi inverse gamma

La densité de la loi inverse gamma de paramètre (a, b) est

$$f(x) = b^a \frac{1}{\Gamma(a)} e^{-b/x} x^{-a-1} \mathbb{I}_{\mathbb{R}_+^*}(x)$$

modèle gaussien (suite)

- $x = (x_1, \dots, x_n)$ iid suivant la loi gaussienne $\mathcal{N}(\theta_1, \theta_2)$.
- Une famille conjuguée :
  soit $(\lambda, \tau, a, b) \in \mathbb{R} \times \mathbb{R}_+^{*3}$. La loi est définie par
- la loi conditionnelle de θ_1 sachant θ_2 est la loi gaussienne de moyenne λ et de variance θ_2/τ
- la loi de θ_2 est la loi inverse Gamma de paramètres (a, b).

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Mélange de loi

On suppose que K experts donnent des avis différents.
On construit K loi a priori $\pi_j(\theta)$, $j = 1, \dots, K$.
La loi π_j traduit l'information fournit par l'expert j et $\pi_j(\theta|x)$ la loi a posteriori associée.
Pour prendre en compte l'information des K experts on prend un mélange de loi

$$\pi(\theta) = \sum_{i=1}^{K} q_i \pi_i(\theta)$$

avec $\sum q_i = 1$, q_i représente le poids du i ème expert c'est à dire la confiance accordée à son avis.
La loi a posteriori s'écrit comme un mélange des lois $\pi_i(\theta|x)$

$$\pi(\theta|x) = \sum_{i=1}^{K} Q_i \pi_i(\theta|x) \text{ et } Q_i = \frac{q_i \int \pi_i(\theta) f(x|\theta)\, \mathrm{d}\theta}{\sum_{j=1}^{K} q_j \int \pi_j(\theta) f(x\theta)\, \mathrm{d}\theta}$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

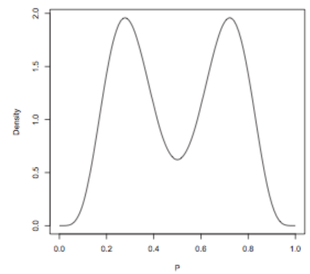
Application du mélange d'avis d'experts

On lance n fois une pièce que l'on suppose truquée.

- Observation : $x \sim B(n, p)$
- info a priori : p est proche de .3 (expert 1) et p est proche de .7 (expert 2).
- Choix de la loi a priori : un mélange de deux lois Beta

$$\pi(p) = q g_1(p) + (1-q) g_2(p)$$

- $q \in (0; 1)$ le poids attribué a chaque expert.
- g_1 est la densité d'une loi beta qui favorise les valeurs de p autour de .3
- g_2 est la densité d'une loi beta qui favorise les valeurs de p autour de .7



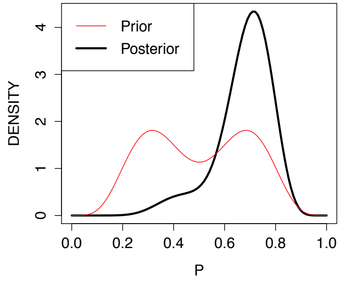
- La loi a posteriori est aussi un mélange de deux lois beta.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

résultats numériques

- On lance $n = 10$ pièces et on observe $x = 7$
- évolution des paramètres

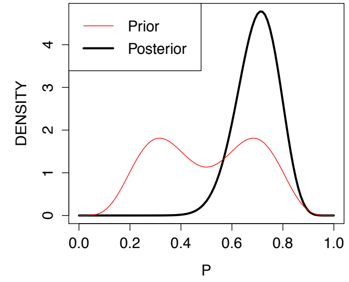
	proba	1 er composante	2 nd composante
a priori	q= 1/2	B(6,14)	B(14,6)
a posteriori	Q = .1	B(13,17)	B(21,9)



Résultats numériques (suite)

- On lance $n = 100$ pièces et on observe $x = 78$
- évolution des paramètres

		1 er composante	2 nd composante
a priori	q= 1/2	B(6,14)	B(14,6)
a posteriori	$Q = \approx 0$	B(84,36)	B(92,28)



Alternative : Structure hiérarchique

On inclut l'hyper paramètre λ à l'ensemble des paramètres

$$\theta \in \Theta \to (\theta, \lambda) \in \Theta x \Lambda$$

- la loi π_λ est interprétée comme la loi conditionnelle de θ sachant λ
- on choisit une loi a priori sur λ

$$\pi(\theta, \lambda) = \pi(\theta|\lambda)\pi_0(\lambda).$$

$\lambda \sim \pi_0$ → $\theta \sim \pi(\cdot|\lambda)$ → $X_1, \ldots X_n$ observations $f(x|\theta)$

Modèle hierarchique suite

- Le modèle hierarchique peut être réécrit comme un modèle bayésien dont la loi a priori sur θ est

$$\pi(\theta) = \int \pi(\theta|\lambda)\pi_0(\lambda)\, \mathrm{d}\lambda$$

- Le principe peut aussi s'étendre à λ lui-même dont la loi a priori peut dépendre d'un nouvel hyperparamètre, etc.
- Une loi a priori hiérarchique conduit à des estimateurs plus robustes, au sens où l'inférence est moins sensible au choix de la loi a priori

Exemple : X_i $i = 1 \dots n$ iid $X_i \sim Pois(\tau)$

Modèle 1
$\tau \sim Exponential(a)$
a fixé (par ex a=1)

$\tau \sim expo(1)$
↓
$X_1, \dots X_n$ iid Pois(τ)

Modèle 2 : hiérarchique
$\tau \sim Exponential(a)$
$a \sim Exponential(1)$

$a \sim expo(1)$
↓
$\tau \sim expo(a)$
↓
$X_1, \dots X_n$ iid Pois(τ)

Calcul des lois a posteriori

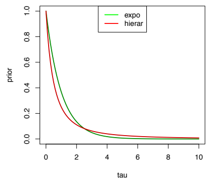
Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

A priori

Modèle 1 $E(\tau) = 1$

Modèle 2 $E(\tau) = \infty$ car $\pi(\tau) = \frac{1}{(1+\tau)^2}$

Lois a priori sur τ



A posteriori

Modèle 1 La loi a posteriori est la loi gamma de paramètres $(\sum_{i=1}^n X_i + 1, n + a)$

Modèle 2 On ne retrouve pas une loi classique

$$\pi(\tau|x) \propto e^{-n\tau} \tau^n \frac{1}{(\tau+1)^2}$$

.

Lois non informatives

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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 Jeffreys)
3. argument fréquentiste (loi de concordance)

Loi a priori de Laplace

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

La loi a priori de Laplace correspond au choix

$$\pi(\theta) \propto \mathbb{I}_\Theta(\theta).$$

En fonction de l'ensemble Θ on obtient

- une loi uniforme Θ
- une loi impropre.
  Il faut alors vérifier la condition $\int f(x|\theta)\, \mathrm{d}\theta < \infty$

Remarques

- La loi a posteriori n'est pas toujours définie
- la loi n'est pas invariante par reparamétrisation
  Reparamétrisation : $\eta = g(\theta)$ avec (g une bijection)

$$\pi(\theta) \propto 1 \Longrightarrow \tilde{\pi}(\eta) \propto \left| \frac{\mathrm{d}}{\mathrm{d}\eta} g^{-1}(\eta) \right|,$$

Le choix de loi a priori sur η n'est donc plus (en général) la loi de Laplace

Loi de Jeffreys

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 $n \to \infty$

Expression de la loi de Jeffreys

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Sa construction repose sur l'information de Fisher $I(\theta)$: la matrice défine par

$$I_{ij}(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial \log f(X|\theta)}{\partial\theta_i}\right)\left(\frac{\partial \log f(X|\theta)}{\partial\theta_j}\right)\right]$$

pour $\theta \in \Theta \in \mathbb{R}^d$

La loi non informative de Jeffreys est définie par

$$\pi^*(\theta) \propto \det^{\frac{1}{2}} I(\theta).$$

Cette loi est invariante par reparamétrisation.

Exemples

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

1. modèle binomial : $x \sim \mathcal{B}(n,p)$
   - L'information de Fisher s'écrit
     $$I(p) = \frac{n}{p(1-p)}.$$
   - La loi de Jeffrey est la loi beta $\mathcal{B}e(1/2, 1/2)$
2. modèle gaussien : $x \sim \mathcal{N}(\mu, \sigma^2)$ La loi de Jeffrey est
   - $\pi(\mu) \propto 1$ si la variance est connue
   - $\pi(\sigma) \propto \sigma^{-1}$ si la moyenne est connue
   - $\pi(\mu,\sigma) \propto \sigma^{-2}$ si les deux sont inconnues
   - Pour les trois modèles : la loi a priori est impropre mais la loi a posteriori est bien définie.
     Il suffit de vérifier que
     $$\int\int f(x_i|\mu,\sigma)\pi(\mu,\sigma)\ \mathrm{d}\mu\ \mathrm{d}\sigma < \infty$$

Loi de Référence

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Généralisation de la loi de Jeffreys pour $\Theta \subset \mathbb{R}^d$

- Les coordonnées sont regroupées par blocs.
- La loi a priori de référence est construite de façon conditionnelle.

Exemple : 2 blocs $\theta = (\theta_1, \theta_2)$

- θ_1 est le paramètre d'intérêt
- θ_2 est le paramètre de nuisance,

alors la loi a priori de référence est calculée à partir de

1. $\pi(\theta_2|\theta_1)$ la loi de Jeffreys associée à $f(x|\theta)$ conditionnellement à θ_1,
2. $\pi(\theta_1)$ comme la loi Jeffreys associée à

$$\widetilde{f}(x|\theta_1) = \int f(x|\theta_1,\theta_2)\pi(\theta_2|\theta_1)\ \mathrm{d}\theta_2.$$

lois de concordance

Argument fréquentiste sur les régions de crédibilité

Sous des hypothèses générales de régularité, on a

$$P_\theta(\theta \leq q^\pi_\alpha(x)) = \int_{\{x|\ \theta \leq q^\pi_\alpha(x)\}} f(x|\theta)\ \mathrm{d}x = \alpha + 0\left(n^{-\frac{1}{2}}\right),$$

pour tout $\alpha \in (0,1)$. [$q^\pi_\alpha(x)$ quantile de la loi a posteriori]

L'objectif est de trouver des lois avec une meilleur vitesse de convergence

Définition : Loi de concordance (matching prior)

On cherche π telle que

$$P_\theta(\theta \leq q^{\pi^*}_\alpha(x)) = \alpha + 0\left(n^{-1}\right),$$

pour tout $\alpha \in (0,1)$.

θ unidimensionnel

Proposition

Pour un modèle régulier, la loi a priori est une loi de concordance (matching prior) si et seulement si

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left\{\pi(\theta) I(\theta)^{-\frac{1}{2}}\right\} = 0.$$

où I est l'information de Fisher.

En dimension 1 :

- la loi de Jeffrey est une loi de concordance
- c'est l'unique solution

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative

Modèles Hiérarchiques

Paramètres multi-variés et données historiques.

Effet individuel

Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Motivation

Dans de nombreux modèles multi-variés, les paramètres peuvent être supposés

- liés/dépendants
- connectés à d'autres variables exogènes/explicatives.

$\rightsquigarrow$ la loi a priori doit refléter la dépendance entre les paramètres.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Modèle hierarchique

On veut estimer $\theta \in \mathbb{R}^p$ à partir d'un petit échantillon $x \in f(x|\theta)$

La difficulté est souvent d'intégrer de la dépendance entre les coordonnées du paramètre θ.

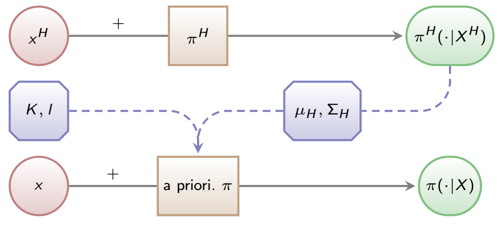
Contexte on dispose d'un échantillon historique x^H pour construire la loi a priori

Démarche

- On suppose que la loi a priori de θ est une loi gaussienne multivariée
- On suppose que les deux échantillons sont issus de la même famille de lois avec des paramètres similaires (mais pas nécessairement égaux)

Description du modèle

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters



La loi a priori π est une loi gaussienne multivariée

$$\mu = KE(\theta|x^H) = K\mu_H \text{ où } K \text{ est une matrice diagonale}$$

$$\Sigma = l^{-1}\mathrm{Var}(\theta|x^H) = l^{-1}\Sigma_H$$

1. la loi a priori sur θ a la même matrice de corrélation que la loi a posteriori calculée sur les données historiques.
2. K mesure la similarité entre les deux échantillons.
3. K, l sont des hyper-paramètres

loi des hyperparamètres

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

1. La loi a priori du facteur l est choisie " non informative" : loi gamma avec une grande variance
2. la loi sur K : on suppose que les $k_1, ..., k_p$ sont iid de moyenne 1 (car les paramètres des deux modèles sont supposés proches)
3. On peut ajouter un niveau hiérarchique sur K

$$\boxed{q \sim \pi_0 \qquad \mathbb{E}(q) = 1} \longrightarrow \boxed{\text{Sachant } q,\ k_i \text{ iid et } \mathbb{E}(k_i|q) = q}$$

Pour tout i, on a

$$E(k_i) = E(E(k_i|q)) = E(q) = 1$$

Application à la régression polynomiale

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Modèle : pour tout $t = 1...n$

$$y_t = \sum_{i=0}^{p} \theta_i x_t^i + \varepsilon_t$$

où

- ε_n sont i.i.d. $N(0, \sigma^2)$,
- $x_1, \ldots, x_n$ sont régulièrement espacés sur [-1,1]

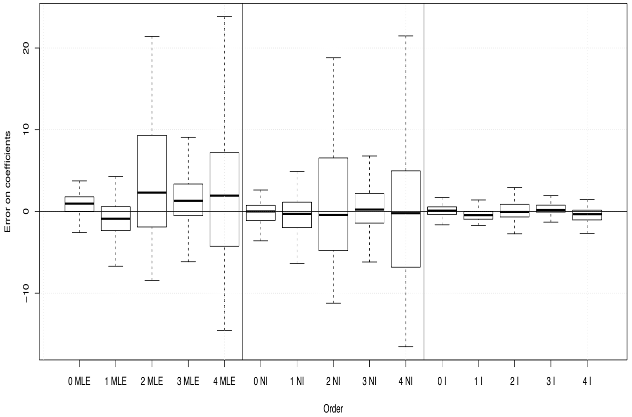
Description des données simulées :

échantillon y^H : $n_H = 200$ observations avec $p = 4$; coefficients $\theta^H = (2, -1, 3, 1, 2)$ et $\sigma^2 = 1$

échantillon y : 10 observations : $\sigma^2 = 4$ et

$$\theta_0 = \rho\theta_0^H \quad \theta_1 = \rho\theta_1^H \quad \theta_2 = \theta_2^H \quad \theta_3 = \theta_3^H \quad \theta_4 = \theta_4^H$$

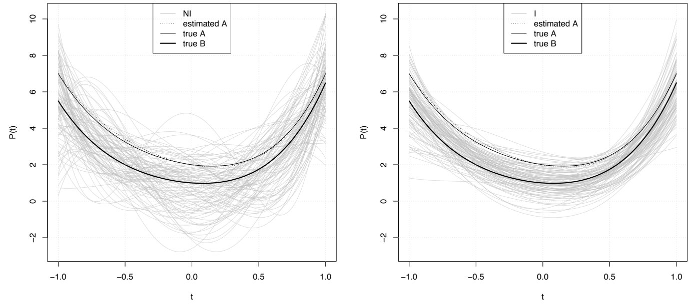
résultats numériques : $\rho = 0.5$



Erreurs d'estimation pour les 5 coefficients du modèle, MLE, loi non-informative (NI) and loi informative (I)

(cont.)

A -> historique et B -> l'échantillon



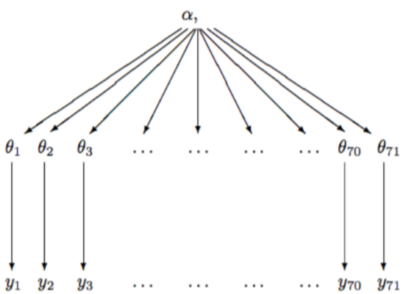
Polynomes estimés : comparaison de l'approche non-informative (left) et de l'approche informative (right).

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Modèle

On dispose de n observations $y_1, \ldots, y_n$

1. Chaque observation y_i est issue de la loi $f(\cdot|\theta_i)$
2. tous les θ_i sont distribués suivant une même loi de paramètre α
3. Le paramètre α est fixé ou supposé inconnu.



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Les lois du modèle

1. La loi des observations :
$$f(x_1, \ldots, x_n|\theta_1, \ldots, \theta_n)$$
2. La loi a priori
$$\pi(\theta_1, \ldots, \theta_n|\alpha)$$
3. loi de l'hyper-paramètre
$$\pi(\alpha)$$

La loi a posteriori est

$$\pi(\theta_1, \ldots, \theta_n, \alpha|x_1, \ldots, x_n) \propto f(x_1, \ldots, x_n|\theta_1, \ldots, \theta_n)\pi(\theta_1, \ldots, \theta_n|\alpha)\pi(\alpha)$$

et

$$\pi(\theta_1, \ldots, \theta_n|x_1, \ldots, x_n) \propto \int f(x_1, \ldots, x_n|\theta_1, \ldots, \theta_n)\pi(\theta_1, \ldots, \theta_n|\alpha)\pi(\alpha)\, \mathrm{d}\alpha$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Echangeabilité

Si aucune information n'est disponible pour distinguer les θ_j les uns des autres.

- aucun ordre sur les $\theta_1, \dots, \theta_n$
- aucun regroupement

  On doit supposer que les paramètres sont interchangeables

Cette propriété correspond à la propriété d'échangeabilité pour une loi de probabilité :

la loi a priori sur $\theta_1, \dots, \theta_n$ est invariante par permutation des indices (1, …, n).

Forme des lois échangeables

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

La forme la plus simple d'une distribution échangeable est de supposer les variables $\theta_1, \dots, \theta_n$ indépendantes et identiquement distribuées suivant une loi paramétrée par α

$$p(\theta_1, \dots, \theta_n|\alpha) = \prod_{i=1}^{n} p(\theta_i|\alpha)$$

En général, α est inconnu, et il devient un paramètre (hyper paramètre) du modèle :

$$p(\theta_1, \dots, \theta_n, \alpha) = \prod_{i=1}^{n} p(\theta_i|\alpha)\pi(\alpha)$$

et

$$p(\theta_1, \dots, \theta_n) = \int_A \prod_{i=1}^{n} p(\theta_i|\alpha)\pi(\alpha)\, \mathrm{d}\alpha$$

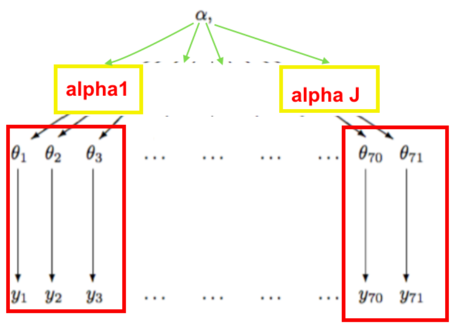
- Conditionnellement à α, les variables $\theta_1, \dots, \theta_n$ sont indépendantes
- Mais les variables $\theta_1, \dots, \theta_n$ ne sont pas indépendantes

Echangeabilité et informations supplémentaires

Multilevel model

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

- Si les observations peuvent être regroupées, on construit un modèle hiérarchique où chaque groupe a son propre sous-modèle.
  On choisit des lois échangeables pour chacun des groupes



Echangeabilité et covariables

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

La façon habituelle de modéliser l'échangeabilité avec les covariables $z_1, \dots, z_n$ est de supposer l'indépendance conditionnelle.
La loi a priori s'écrit

$$p(\theta_1, \dots, \theta_n, \alpha|z_1, \dots, z_n) = \prod_{i=1}^{n} p(\theta_i|\alpha, z_i)p(\alpha|z_1, \dots, z_n)$$

et

$$p(\theta_1, \dots, \theta_n|z_1, \dots, z_n) = \int \prod_{i=1}^{n} p(\theta_i|\alpha, z_i)p(\alpha|z_1, \dots, z_n)\, \mathrm{d}\alpha$$

Prévision dans un modèle hierarchique

- On observe $x_1, \ldots, x_n$ avec $x_i \sim f(\cdot|\theta_i)$
- Les paramètres du modèle sont $(\theta_1, \ldots, \theta_n, \alpha)$
- On veut prévoir x_{n+1}

$$p(x_{n+1}|x_1, \ldots, x_n) = \int f(x_{n+1}|\theta_{n+1}) p(\theta_{n+1}|\alpha) \pi(\alpha|x_1, \ldots, x_n) \, \mathrm{d}\alpha \, \mathrm{d}\theta_{n+1}$$

A partir d'un échantillon $(\alpha_1, \ldots, \alpha_M)$ simulé suivant la loi a posteriori de α,

1. On simule $\theta_{n+1}(i)$ suivant $p(\theta_{n+1}|\alpha_i)$ pour tout $i = 1, \ldots, M$
2. On simule $x_{n+1}(i)$ suivant $f(x_{n+1}|\theta_{n+1}(i))$ pour tout $i = 1, \ldots, M$

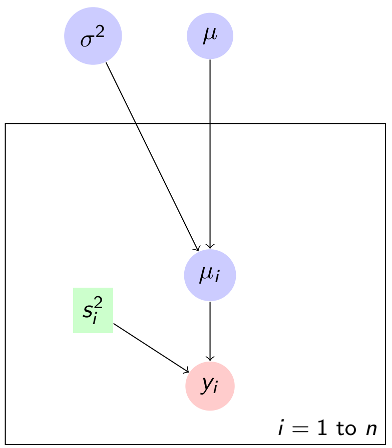
Modèle hiérarchique Normal-Normal

On suppose que

1. les observations sont indépendantes gaussiennes $y_i \sim N(\mu_i, s_i^2)$
$$y_i = \mu_i + s_i \varepsilon_i$$
   - les s_i sont connues
   - les ε_i sont iid $N(0, 1)$.
2. Hypothèses sur les μ_i :
$$\mu_i = \mu + \sigma \tilde{\varepsilon}_i$$
   - σ mesure la dispersion des μ_i autour de μ
   - les $\tilde{\varepsilon}_i$ sont iid $N(0, 1)$.
   - Loi échangeable sur $\mu_1, .., \mu_n$ car conditionnellement à (μ, σ^2), $\mu_1, .., \mu_n$ sont iid suivant la loi $N(\mu, \sigma^2)$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Loi a priori sur les hyper paramètres



On choisit

$$p(\sigma^2, \mu) = p(\sigma^2) p(\mu)$$

avec

$$\sigma^{-2} \sim \text{loi gamma}(\nu, \lambda)$$

$$\mu \sim N(m_\mu, V_\mu)$$

Remarque

Attention si on prend une loi impropre pour σ^2 (par exemple $\frac{1}{\sigma^2}$) la loi a posteriori n'est pas toujours définie.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Alternative pour la loi sur la variance

On fixe $s_0 > 0$.
La loi de σ^2 est la loi de shrinkage uniforme de paramètre s_0 si

$$s_0^2/(\sigma^2 + s_0^2) \sim \mathcal{U}(0, 1)$$

La densité est de la forme

$$p(\sigma^2) = \frac{s_0^2}{(s_0^2 + \sigma^2)^2}.$$

Choix de s_0
Si les s_i^2 sont connus on prend

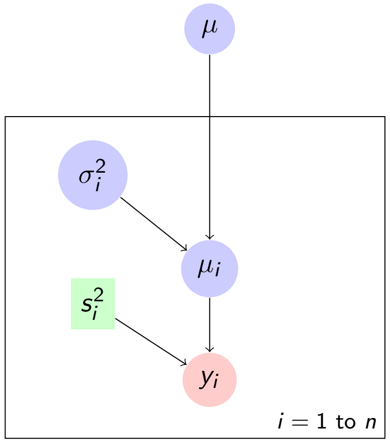
$$\frac{1}{s_0^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{s_i^2}.$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Extension : vers un modèle robuste

Pour obtenir un modèle robuste, on ajoute un effet individuel sur la variance

$$\mu_i = \mu + \sigma_i \tilde{\varepsilon}_i$$



Loi a priori sur les σ_i. Ils sont iid suivant la loi de Shrinkage uniforme.

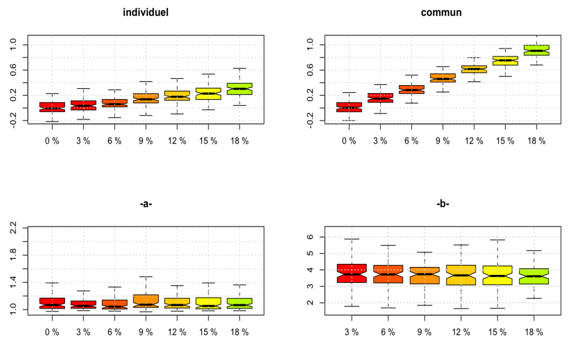
Illustration :

1. Pour illustrer la robustesse on simule un échantillon de taille 100 et on remplace une proportion q des valeurs par des outliers.
2. On répète 500 fois l'expérience
3. On représente l'évolution des estimateurs de Bayes de μ, σ_i en fonction de q.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Résultats numériques

Comparaison des modèles avec ou sans effet individuel sur la variance



Comparaison des valeurs de σ_i entre une observation "standard" et un outlier

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Description du problème

Soit $\mathcal{M}_k$ $k = 1, \dots K$ une collection de modèles.

- A partir des observations $x = (x_1, ..., x_n)$: on veut choisir le meilleur modèle
- Pour $k = 1, .., K$, le modèle $\mathcal{M}_k$ est défini par
  - $x \sim f_k(x|\theta_k) = L(\theta_k)$
  - $\theta_k \in \Theta_k$

Approche classique

- Pour chaque modèle, on estime le paramètre $\theta_k \in \Theta_k$ à partir des observations $x \rightsquigarrow \hat{\theta}_k$ l'estimateur du Maximum de Vraisemblance
- On utilise un critère de sélection de modèle :
  1. $AIC = -2\ln(L(\hat{\theta}_k)) + 2k$
  2. $AICc = AIC + \frac{2k(k+1)}{n-k-1}$
  3. $BIC = -2\ln(L(\hat{\theta}_k)) + \ln(n)k$
- On sélectionne le modèle que minimise le critère : k^*

Décision

- On estime θ par $\hat{\theta}_{k^*}$
- On prévoit en utilisant le modèle $\mathcal{M}_{k^*}$

Version Bayésienne

- Soit $\mathcal{M}_k$ $k = 1, \dots K$ une collection de modèles.
- pour $i = 1, ..., K$ on note
  - $x \sim f_i(x|\theta_i)$
  - $\theta_i \in \Theta_i$
  - $\theta_i \sim \pi_i(\theta_i)$
- On construit un méta modèle :
  l'indice du modèle devient aussi un paramètre du modèle
- soit $P(M_k)$ les probabilités a priori des K modèles.
  $\sum P(M_k) = 1$

Règle de décision

- On calcule les probabilités a posteriori des K modèles

$$\pi(M_i|x) = \frac{P(M_i)\int f_i(x|\theta_i)\pi_i(\theta_i)\,\mathrm{d}\theta_i}{\sum_{j=1}^{k} P(M_j)\int f_j(x|\theta_j)\pi_j(\theta_j)\,\mathrm{d}\theta_j}$$

- La règle de décision :

  On sélectionne la valeur k^* qui maximise $P(M_k|x)$

  ou

  On construit un modèle moyenné avec des poids (BMA)

Bayesian Model Averaging

Cox : "All models are wrong, some are useful"

Soit M_k $k = 1, \dots K$ une collection de modèles.
Pour chaque modèle de la collection on note

- $x \sim f_k(x|\theta_k)$
- $\theta_k \in \Theta_k$
- $\theta_k \sim \pi_k(\theta_k)$

Idée

- On estime et prévoit à partir d'un modèle moyenné
- Tous les modèles n'ont pas la même contribution
- Les poids des modèles individuels sont les probabilités a posteriori des modèles, c'est à dire

$$\pi(M_k|x) = \frac{P(M_k) \int f_k(x|\theta_k)\pi_k(\theta_k) \, \mathrm{d}\theta_k}{\sum_{i=1}^K P(M_i) \int f_i(x|\theta_i)\pi_i(\theta_i) \, \mathrm{d}\theta_i}$$

La prévision par mélange

Ayant observé $x_1, \dots, x_n$,

- la densité prédictive de x_{n+1} est

$$f(y|x_1, \dots, x_n) = \sum_k \pi(M_k|x_1, \dots, x_n) f_k(y|x_1, \dots, x_n)$$

où $f_k(y|x_1, \dots, x_n)$ est la densité de la loi prédictive dans le modèle M_k

$$f_k(y|x_1, \dots, x_n) = \int f_k(y|x_1, \dots, x_n, \theta_k)\pi_k(\theta_k|x_1, \dots, x_n) \, \mathrm{d}\theta_k$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Intervalle de prévision et prévision ponctuelle

- Les intervalles de prévision se calculent sur la loi prédictive
$f(y|x_1, \dots, x_n) = \sum_k \pi(M_k|x_1, \dots, x_n) f_k(y|x_1, \dots, x_n)$
- Le prédicteur optimal est

$$\hat{x}_{n+1} = \sum_k \pi(M_k|x_1, \dots, x_n)\hat{x}_{n+1}(k)$$

où $\hat{x}_{n+1}(k)$ est le predicteur ponctuel dans le modèle M_k

$$\hat{x}_{n+1}(k) = \int y f_k(y|x_1, \dots, x_n) \, \mathrm{d}y$$

Remarque

- *Tous les modèles contribuent au calcul de la prévision*
- *Dans l'approche classique, on calcule la prévision dans le modèle sélectionné.*

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Estimation et BMA

Soit M_k $k = 1, \dots K$ une collection de modèles.
Pour chaque modèle de la collection on note

- $x \sim f_k(x|\theta_k)$
- $\theta_k \in \Theta_k$
- $\theta_k \sim \pi_k(\theta_k)$

On veut estimer un paramètre d'intérêt $\tilde{\theta}$ (commun à tous les modèles)

$$\tilde{\theta} \in \tilde{\Theta} \subset \Theta_i \text{ pour tout } \ i$$

Le modèle bayesien moyenné est défini par la loi a posteriori

$$\pi^{BMA}(\tilde{\theta}|x) = \sum_{k=1}^K \pi(\tilde{\theta}|M_k, x)\pi(M_k|x)$$

- $\pi_k(\tilde{\theta}|M_k, x)$ est la loi a posteriori du paramètre $\tilde{\theta}$ pour le modèle M_k
- $\pi(M_k|x)$ est la probabilité a posteriori du modèle k

C'est la moyenne pondérée des distributions a posteriori de $\tilde{\theta}$ pour chacun des modèles

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Lien entre l'estimation et la prévision

On suppose que $\theta = \tilde{\theta}$.
La prévision par mélange coincide avec la loi prédictive construite à partir de la loi a posteriori $\pi^{BMA}(\theta|x)$

$$f(y|x_1,\dots,x_n) = \int f(y|x_1,\dots,x_n,\theta)\pi^{BMA}(\theta|x_1,\dots,x_n)\,\mathrm{d}\theta$$

Statistique Bayésienne
Anne Philippe

Regression

Le modèle s'écrit

$$y_i = x_i^\top \beta + \varepsilon_i.$$

On suppose qu'il y a p variables explicatives donc 2^p modèles

- La loi a priori de Zellner's g-prior est la loi multivariée gaussienne

$$\beta|\psi \sim \mathrm{N}[\beta_0, g\psi^{-1}(X^\top X)^{-1}].$$

  où $\sigma^2 = \psi^{-1}$ est la variance de ε_i
- g est un scalaire à fixer. Le choix classique 'unit information prior' (UIP), $g = N$ le nombre d'observations.
- $\pi(\sigma) = \frac{1}{\sigma}$

Statistique Bayésienne
Anne Philippe

Regression (cont)

- On choisit la loi uniforme sur les modèles.
- La loi sur le nombre de variables incluses n'est pas uniforme.
- Loi a posteriori BMA pour les coefficients de la régression

$$\pi(\beta|x,y) = \sum_{k=1}^{2^p} \pi(\beta|M_k,x,y)\pi(M_k|x,y)$$

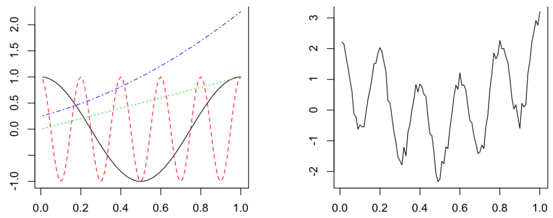
Remarque

En grande dimension on ne peut pas estimer tous les modèles. On conserve que les modèles de plus forte probabilité a posteriori.

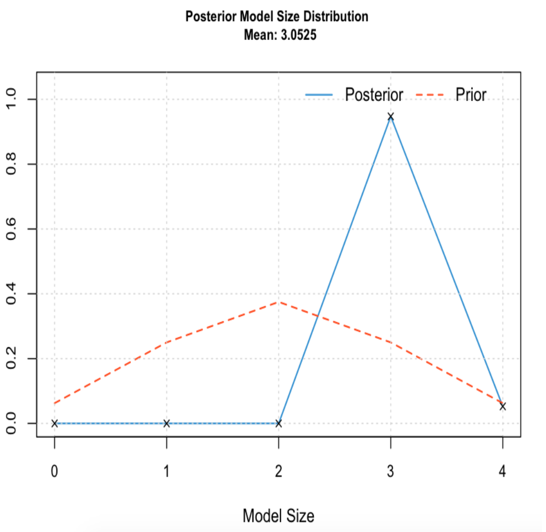
Statistique Bayésienne
Anne Philippe

Application sur données simulés (R : BMS)

- 4 variables explicatives.
- Données simulées avec 3 variables explicatives : $y = 1 * x1 + 1.5 * x2 + .5 * x3 + \varepsilon$
- la variable 4 n'apparait pas dans le modèle



loi a posteriori sur le nombre de variables



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

loi a posteriori sur les modèles

	0d	0e	0f	0c	07	05	06	04	03	08	09
x1	1.0	1.00	1.00	1	0	0	0	0	0	1	1
x2	1.0	1.00	1.00	1	1	1	1	1	0	0	0
x3	0.0	1.00	1.00	0	1	0	1	0	1	0	0
x4	1.0	0.00	1.00	0	1	1	0	0	1	0	1
PMP	0.5	0.44	0.05	0	0	0	0	0	0	0	0

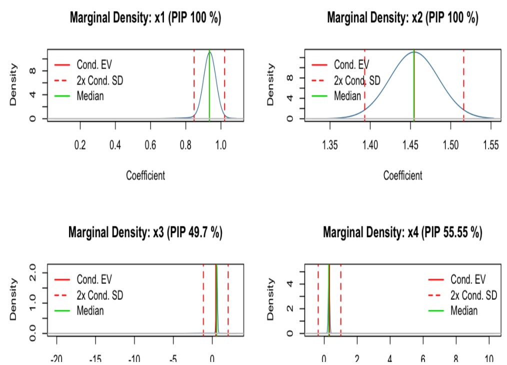
PMP : probabilité a posteriori du modèle

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Estimation des paramètres

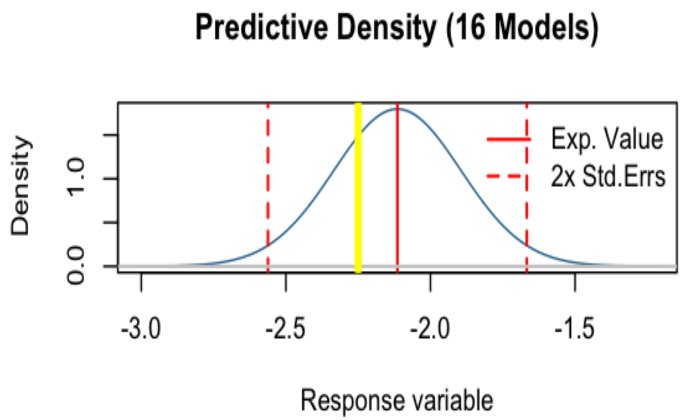
	PIP	Post Mean	Post SD
x2	1.0000000	1.4546208	0.03075993
x1	1.0000000	0.9341147	0.04337772
x4	0.5555022	0.1787574	0.30165430
x3	0.4970332	0.2239837	0.60424542

PIP : probabilité a posteriori que la variable soit intégrée dans le modèle



Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Prévision



Le trait jaune représente la réalisation

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Facteur de Bayes

Définition

Le facteur de Bayes est défini par

$$B_{0/1} = \frac{P(M_0|x)}{P(M_1|x)} \Big/ \left(\frac{P(M_0)}{P(M_1)}\right)$$

Le FB élimine bien l'influence des poids a priori des deux modèles et se comporte comme un rapport de vraisemblance

$$B_{0/1} = \frac{\int f_0(x|\theta_0)\pi_0(\theta_0)\ \mathrm{d}\theta_0}{\int f_1(x|\theta_1)\pi_1(\theta_1)\ \mathrm{d}\theta_1}$$

Interprétation du FB

Une valeur de $B_{0/1} > 1$ signifie que M_0 est plus vraisemblable M_1

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

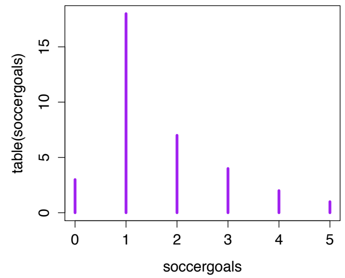
Table donnée par Kass et Raftery

$\log_{10} B_{0/1}$	confiance en faveur de M_0
0 - 0.5	faible
0.5 - 1	substantielle
$1 - 2$	forte
> 2	décisive.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

choix de modèles entre plusieurs lois a priori

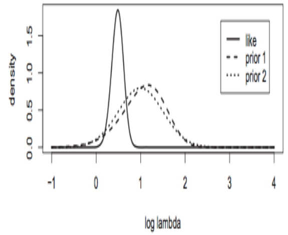
- Données le nombre de buts marqués par match pendant une saison (pour une équipe)
- On dispose d'un échantillon de taille 35.
- On modélise ces données par une loi de Poisson de paramètre λ
- On compare quatre modèles qui correspondent à quatre choix de lois a priori

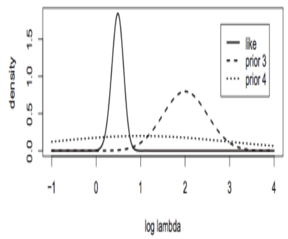


Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Lois a priori

1. λ suit une loi Gamma de paramètres $(4.57, 1.43)$. $E(\lambda) = 3$ et $P(\lambda < 2.1) = P(\lambda > 4.04) = .25$.
2. $log(\lambda)$ suit une loi $\mathcal{N}(1, 1/4)$ et $P(\lambda < 1.94) = P(\lambda > 3.81) = .25$.
3. $log(\lambda)$ suit une loi $\mathcal{N}(2, 1/4)$ et $P(\lambda < 5.27) = P(\lambda > 10.35) = .25$.
4. $log(\lambda)$ suit une loi $\mathcal{N}(1, 4)$ et $P(\lambda < 1.92) = P(\lambda > 28.5) = .25$.





calcul des facteur de Bayes

Les FB s'écrivent

$$B_{i/j} = \frac{\int f(x|\theta)\pi_i(\theta)\, d\theta}{\int f(x|\theta)\pi_j(\theta)\, d\theta} = \frac{m_i(x)}{m_j(x)}$$

où m_i est la loi marginale de x.
Résultats numériques

modèles	MAP	SD a post	$\log(m(x))$
1	0.5248047	0.1274414	-1.502977
2	0.5207825	0.1260712	-1.255171
3	0.5825195	0.1224723	-5.076316
4	0.4899414	0.1320165	-2.137216

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Facteur de Bayes

modèles i/j	2/1	2/3	2/4
$(B_{i/j})$	1.28	45.7	2.42
$log(B_{i/j})$	.24	3.82	0.88
en faveur de	2	2	2
confiance	faible	décisive	substantielle

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Tests

On veut tester si $\theta \in \Theta_0$ contre $\theta \in \Theta_1$.
Le FB s'écrit

$$B_{0/1} = \frac{P(\theta \in \Theta_0|x)}{P(\theta \in \Theta_1|x)} \Big/ \left(\frac{P(\theta \in \Theta_0)}{P(\theta \in \Theta_1)}\right) = \frac{\int_{\Theta_0} f(x|\theta)\pi_0(\theta)\, d\theta}{\int_{\Theta_1} f(x|\theta)\pi_1(\theta)\, d\theta}$$

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Hypothèses ■ classiques ■

- Pour le test $\{\theta_0\}$ contre $\{\theta_1\}$
  - π_i sont les masses de Dirac en θ_i
  - le BF s'écrit
$$B_{0/1} = \frac{f(x|\theta_0)}{f(x|\theta_1)}$$
- Pour le test $\{\theta_0\}$ contre $\{\theta \neq \theta_0\}$
  - π_0 sont les masses de Dirac en θ_0
  - π_1 admet pour densité g sur $\mathbb{R}$
  - la loi a priori est donnée par
$$\pi(\,\mathrm{d}\theta) = p\delta_{\theta_0}(\,\mathrm{d}\theta) + (1-p)g(\theta)\,\mathrm{d}\theta$$
  - le BF s'écrit
$$B_{0/1} = \frac{f(x|\theta_0)}{\int_{\mathbb{R}} f(x|\theta)g(\theta)\,\mathrm{d}\theta}$$

Comparaison avec Neyman Pearson

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

On teste pour un échantillon gaussien : $\mu = 0$ contre $\mu \neq 0$
On choisit comme loi a priori

$$\pi(\,\mathrm{d}\theta) = p\delta_0(\,\mathrm{d}\theta) + (1-p)g(\theta)\,\mathrm{d}\theta$$

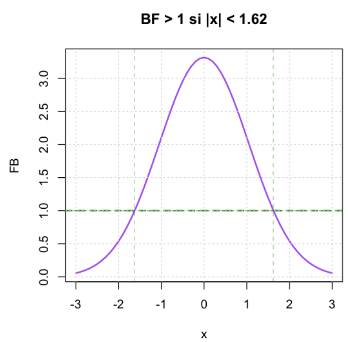
où g est la densité de la loi gaussienne de paramètre $(0, 10)$
Le facteur de Bayes est égal à

$$B_{0/1} = \sqrt{11}\exp(-10x^2/22)$$

Interprétation

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

- Le FB atteint son maximum en 0
- Le FB est supérieur à 1 (décision favorable à l'hypothèse nulle) si $|x| < 1.62$
- On retrouve la région du test de Neyman Pearson au seuil 10 %.



Remarque sur les lois a priori impropres

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Test sur la moyenne d'un échantillon gaussien

On teste pour un échantillon gaussien $\mathcal{N}(\theta, 1)$: $\theta = 0$ contre $\theta \neq 0$
Si on prend la loi de Jeffreys $g(\theta) = C\,\mathbb{I}_{\mathbb{R}}$, La constante C est arbitraire

$$\pi(\,\mathrm{d}\theta) = p\delta_0(\,\mathrm{d}\theta) + (1-p)C\,\mathrm{d}\theta$$

le BF s'écrit

$$B_{0/1} = \frac{f(x|0)}{C\int_{\mathbb{R}} f(x|\theta)\,\mathrm{d}\theta} = \frac{1}{C\sqrt{2\pi}}e^{-x^2/2}$$

le BF depend de C qui est arbitraire !

Conclusion : lorsque l 'on calcule un FB, la loi a priori doit être une probabilité.

Modèlisation par mélanges

Motivations

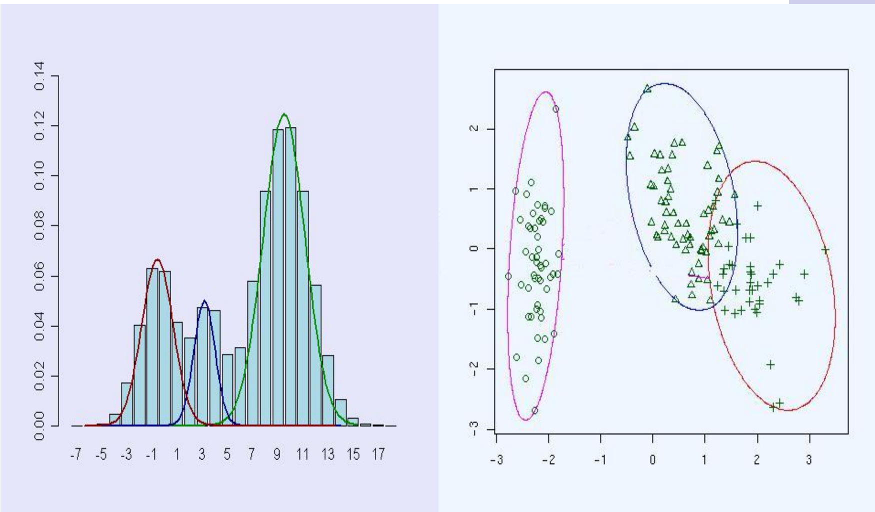
1. Phénomènes complexes // Structures multimodales
2. Populations hétérogènes et classes homogènes
3. Discrimination/Classification

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]

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

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

variable latente

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

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

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

Choix de la loi a priori

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Paramètres :

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

On décompose la loi a priori de la forme suivante

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

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

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

Choix de la loi a priori [suite]

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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

$$(p_1, \dots, p_k) \sim \mathrm{Dirichlet}(\alpha_1, \dots, \alpha_k)$$

de densité

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

Identifiabilité

Pour que le modèle soit identifiable on impose une contrainte sur les paramètres.
On peut prendre par exemple $\theta_1 < \dots < \theta_p$

Classification

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

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

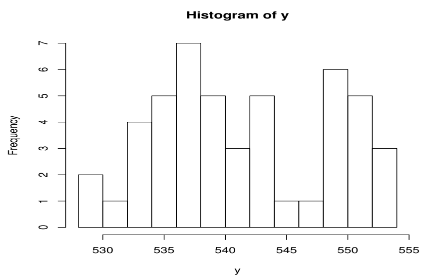
On décide que l'observation x_i est issue de $f_{J(i)}$ où

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

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



Bowmaker et al (1985) analyse data on the peak sensitivity wavelengths for individual microspectrophotometric records on a small set of monkey's eyes. (48 measurements).
Le modèle considéré est

$$p\mathcal{N}(\lambda_1, \sigma^2) + (1-p)\mathcal{N}(\lambda_2, \sigma^2)$$

- On introduit des variables latentes

$$z = \begin{cases} 1 & \text{si } x \sim \mathcal{N}(\lambda_1, \sigma^2) \\ 2 & \text{si } x \sim \mathcal{N}(\lambda_2, \sigma^2) \end{cases}$$

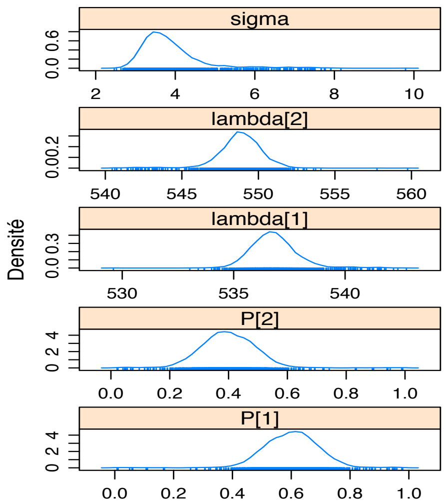
- Le choix de la loi a priori sur p est la loi uniforme.
- Le choix des lois a priori sur λ_i et σ sont les lois conjuguées.
  [loi gaussienne sur λ_i et loi inverse gamma sur σ_i^2]
- Pour les z_i on prend

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

.

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

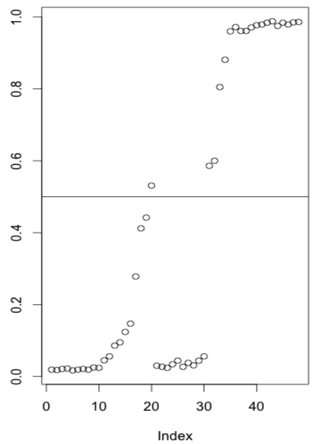
densityplot : lois a posteriori marginales

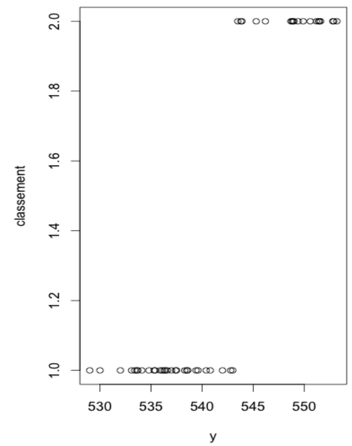


Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Estimation des variables cachées / manquantes

- Le graphique de gauche représente les probabilités $P(z_i = 1|x)$ en fonction de i (les données sont triées par ordre croissant)
- Le graphique de droite représente les estimations des z_i

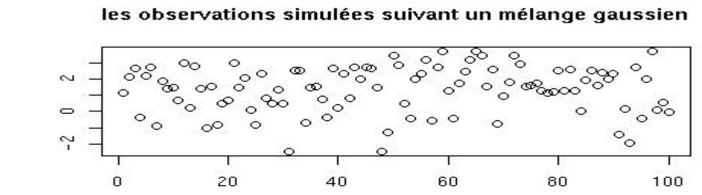


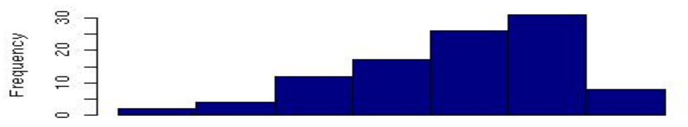


Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hierarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

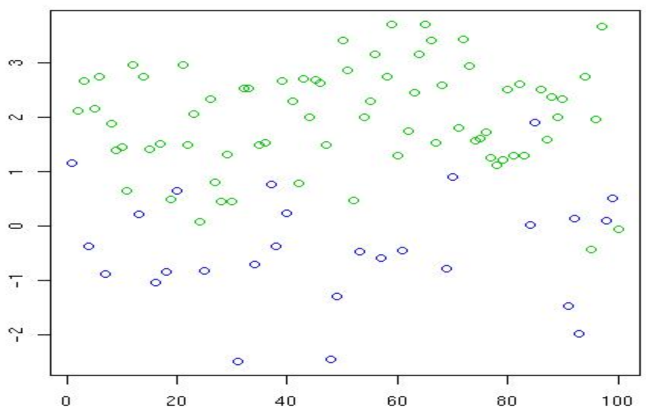
Données simulées

- On évalue la qualité de la classification sur des données simulées.
- On simule un échantillon 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





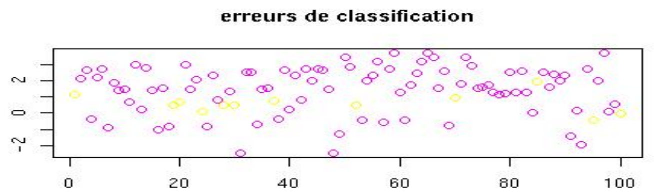
Les données manquantes ....

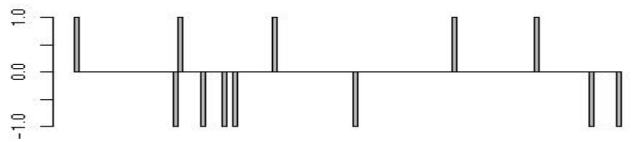


Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Qualité de la classification

Les points jaunes représentent les variables mal-classées





Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Estimation du nombre de composantes

C'est un problème de sélection de modèles

- On dispose d'une famille de modèles $\{M_k; i \in K\}$ où M_k est le mélange de lois à k composantes.
- Pour chaque modèle, on dispose d'une structure paramétrique $\theta^{(k)} \in \Theta_k$ qui regroupe les paramètres des k composantes du mélange.
- On suppose que le nombre de composante k est inconnu.
  Il est inclus dans l'ensemble des paramètres à estimer

Statistique Bayésienne
Anne Philippe
Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Description du modèle bayésien

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

Estimateurs

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

$$p(x_{n+1}|x_1, ..., x_n) = \sum_{k_0=1}^{K} P(k = k_0|x) \int f_{k_0}(x|\theta_{k_0})\pi_{k_0}(\theta_{k_0}|x)\, d\theta_{k_0}$$

Méthode de Monte Carlo

Statistique Bayésienne
Anne Philippe

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

Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Sortie de ces algorithmes

Statistique Bayésienne
Anne Philippe

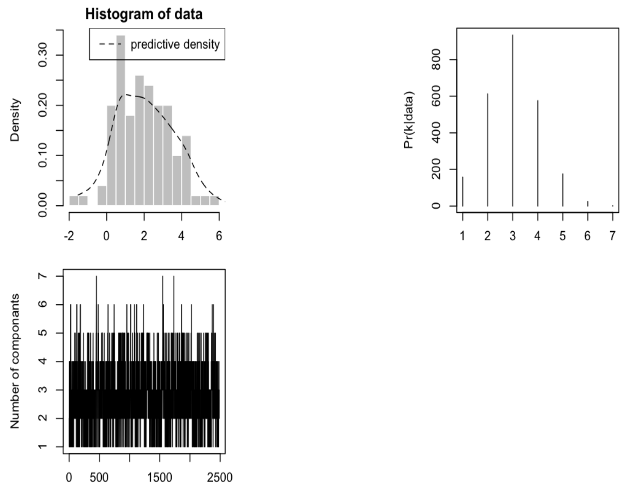
On simule un processus (k_t, θ_t) pour $t = 1, ...M$

1. (k_t) est un échantillon suivant une approximation de la loi a posteriori du nombre de composantes
2. L'approximation de $P(k = k_0|\mathbf{x})$ par la méthode de Monte Carlo s'écrit $\frac{1}{M}\sum_{t=1}^{M} \mathbb{I}_{k_t=k_0}$
3. la dimension du vecteur θ_t varie avec t. $\theta_t \in \Theta_{k_t}$ l'ensemble des paramètres du mélange à k_t composantes.
4. $\{\ \theta_t\,|k_t = k, t = 1, ..., M\}$ est un échantillon suivant une approximation de la loi a posteriori de $\theta^{(k)}$
5. L'approximation de $\mathbb{E}(\theta^{(k)}|\mathbf{x}, k)$ par la méthode de Monte Carlo s'écrit $\frac{1}{M}\sum_{t=1}^{M} \theta_t\,\mathbb{I}_{k_t=k}$

Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters

Exemple suite : estimation du nombre de clusters

Statistique Bayésienne
Anne Philippe



Résultats obtenus avec la librarie R `bmixture`

Modèle Bayésien
Inférence
Estimateurs de Bayes
Régions de crédibilité
Prévision des futures observations
Lois a priori
Approche subjective
Modèle hiérarchique
Approche non informative
Modèles Hiérarchiques
Paramètres multi-variés et données historiques.
Effet individuel
Choix de modèles et BMA
Sélection de modèle
Bayesian Model Averaging
Facteur de Bayes
FB et choix de la loi a priori
FB et Test
Classification bayésienne
Modèle de mélange
Nombre de clusters