Introduction à la statistique bayésienne

Anne Philippe

Université de Nantes, LMJL 2020

plan du cours

1. Modélisation bayésienne
2. Estimation et prévision bayésienne
3. Construition des lois a priori

Chapitre 1

Modélisation bayésienne

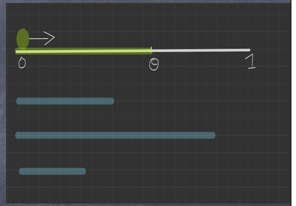
Définitions et exemples

Exemple et

rappels sur le calcul des lois conditionnelles

Exemple du billard

- On lance une bille qui s'arrête à un point $\theta \in [0,1]$ uniformément distribué.
- Question comment déterminer la valeur de θ sans effectuer de mesures ?
- On repète la même expérience N fois de façon indépendante et on note X le nombre de fois où elle s'arrête à gauche du point d'arrêt
- Comment estimer θ à partir de X ?



Approche fréquentiste

- X suit une loi binomiale $\mathscr{B}(N, \theta)$ où θ est un paramètre inconnu
- La vraisemblance s'écrit $V(\theta, X) = \binom{N}{X} \theta^X (1-\theta)^{N-X}$
- L'estimateur du MV est égal à $\hat{\theta}_N^{MV} = \frac{X}{N}$
- Cet estimateur n'utilise pas la 1er expérience aléatoire.

Alternative

- θ est une v.a. distribuée
- la loi de θ est la loi uniforme sur [0,1]
- la loi binomiale $\mathscr{B}(N, \theta)$ est la loi conditionnelle de X sachant θ
- Quelle est la loi de θ sachant X ?

Conditionnement

Soit (X,Y) un couple de variables aléatoires discrètes.

la loi marginale de Y s'écrit

$$P(Y = y) = \sum_x P(X = x, Y = y)$$

La loi conditionnelle de X sachant Y est donnée par la formule de Bayes :

$$P(X = x \,|\, Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Conditionnement

Soit (X,Y) un couple de variables aléatoires continues.
On note f la densité du couple (X,Y).

la loi marginale de Y admet une densité égale à

$$f_Y(y) = \int f(x, y)\, dx$$

La loi conditionnelle de X sachant Y admet une densité définie par

$$f_{X|Y}(x\,|\,y) = \frac{f(x, y)}{f_Y(y)}$$

Formule de BAYES

Soit (X,Y) un couple de variables aléatoires

La formule de Bayes est donnée par

$$P(X \in A\,, Y \in B) = \begin{cases} \displaystyle\int_B P(X \in A\,|\,Y = y) f_Y(y)\, dy & \text{si } Y \text{ est continue} \\ \displaystyle\sum_{y \in B} P(X \in A\,|\,Y = y)\, P(Y = y) & \text{si } Y \text{ est discrète} \end{cases}$$

C'est l'outil central pour calculer les lois conditionnelles.

Exemple du billard (cont.)

- On connait
  $\theta \sim U(0,1)$ et on note sa densité : $\pi(\vartheta) = I_{[0,1]}(\vartheta)$
  la loi conditionnelle $P(X = x\,|\,\theta = \vartheta) = \binom{N}{x} \vartheta^x (1-\vartheta)^{N-x}$

- Formule de BAYES :
  $P(X = x, \theta \in B) = \int_B P(X = x\,|\,\theta = \vartheta)\pi(\vartheta)\, d\vartheta = \int_B \binom{N}{x} \vartheta^x (1-\vartheta)^{N-x}\, I_{[0,1]}(\vartheta)\, d\vartheta$

  $P(X = x, \theta \in B) = P(\theta \in B\,|\,X = x) P(X = x)$

Exemple du billard (cont.)

- La loi marginale de X est
  $$P(X = x) = P(X = x, \theta \in [0,1]) = \binom{N}{x} \int_0^1 \vartheta^x (1-\vartheta)^{N-x}\, d\vartheta$$

- En appliquant la formule de Bayes, on obtient

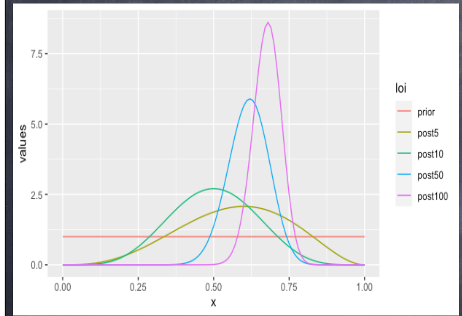
$$P(\theta \in B\,|\,X = x) = \frac{P(X = x\,, \theta \in B)}{P(X = x)} = \frac{\int_B \vartheta^x (1-\vartheta)^{N-x}\, d\vartheta}{\int_0^1 \vartheta^x (1-\vartheta)^{N-x}\, d\vartheta} = \int_B \pi(\vartheta\,|\,X = x) d\vartheta$$

$$\pi(\vartheta\,|\,X = x) = \frac{\vartheta^x (1-\vartheta)^{N-x}}{\int_0^1 \vartheta^x (1-\vartheta)^{N-x}\, d\vartheta}$$

c'est la loi beta de paramètres (x+1 , N-x+1)

expériences numériques

N	5	10	50	100
X	3	5	31	68
MV	0,6	0,5	0,62	0,68



Representation des lois conditionnelles de θ sachant les N observations

Loi		mean	quantile 2.5%	quantile 97.5 %
(prior) π		.50	.03	.98
$\pi(\,.\,\vert X)$ posteriori	N = 5	.57	.22	.88
	N = 10	.50	.24	.77
	N = 50	.62	.48	.74
	N=100	.68	.58	.76

Définition du modèle bayésien

loi a priori

Modèle paramétrique bayésien

- On considère $P_\theta^n(\cdot), \theta \in \Theta$ une famille de lois indexées par θ.
  On note $X = (X_1, \ldots, X_n)$ les observations.
- On suppose que θ est une VARIABLE ALÉATOIRE
- la loi $P_\theta^n(\cdot)$ est interprété comme la loi conditionnelle de $X = (X_1, \ldots, X_n)$ sachant le paramètre θ

$$P_\theta^n(\cdot) = P^n(\cdot \,|\, \theta)$$

Définition :

la loi du paramètre θ est appelée « loi a priori »

- cette loi est construite à partir des informations disponibles sur le paramètre θ avant de collecter des données
- L'information (dite a priori) provient
  - d'avis d'expert
  - de résultats d'experiences précédentes dont les résultats sont supposé similaires
  - des propriétés physiques
- Exemple du billard :
  On sait que θ est la position du point d'arrêt de la première boule. D'après l'expert ce point est uniformément distribué

Fréquentiste vs Bayésien

- $X_1, \ldots, X_n$ observations
- il existe $\theta_O \in \Theta$ inconnu
- $P^n_{\theta_0}(\cdot)$ est la loi des observations

- $X_1, \ldots, X_n$ observations
- θ variable aléatoire
- $P^n_\theta(\cdot) = P^{(n)}(\cdot \,|\, \theta)$ est la loi conditionnelle des observations sachant θ

Inférence

estimateur $\hat{\theta}(X_1, \ldots, X_n) \xrightarrow{P-ps-L^p} \theta_0$

intervalle de confiance

?

Inférence

loi a posteriori

Inférence

- L'objectif est de mettre à jour la loi a priori sur θ à partir des observations.
- En combinant la loi a priori et la loi des observations sachant le paramètre θ, on peut calculer
  - La loi jointe de $(X_1, \ldots, X_n, \theta)$
  - La loi marginale de $(X_1, \ldots, X_n)$ appelée loi prédictive a priori
  - La loi conditionnelle de θ sachant $(X_1, \ldots, X_n)$

Définition :
la loi conditionnelle du paramètre θ sachant les observations est appelée « loi a posteriori »

---

calcul de la loi a posteriori continue/continue

- π : densité de la loi a priori
- $f^{(n)}(x_1, \ldots, x_n \,|\, \theta)$: densité la loi de $(X_1, \ldots, X_n)$ sachant θ

- La loi jointe : $g_n(x_1, \ldots, x_n, \vartheta) = \pi(\vartheta) f^{(n)}(x_1, \ldots, x_n \,|\, \vartheta)$
- La loi marginale : $m_n(x_1, \ldots, x_n) = \int_\Theta \pi(\vartheta) f^{(n)}(x_1, \ldots, x_n \,|\, \vartheta)\, d\vartheta$
- La loi a posteriori : $\pi(\vartheta \,|\, x_1, \ldots, x_n) = \dfrac{\pi(\vartheta) f^{(n)}(x_1, \ldots, x_n \,|\, \vartheta)}{m_n(x_1, \ldots, x_n)}$

calcul de la loi a posteriori : discrète / discrète

- $\Theta = \{\theta_1, \ldots, \theta_p\}$ la loi a priori est définie par $\pi_i = P(\theta = \theta_i)$
- $P^{(n)}(X_1 = x_1, \ldots, X_n = x_n \mid \theta)$: loi de $(X_1, \ldots, X_n)$ sachant θ

- La loi marginale :

$$P^{(n)}(X_1 = x_1, \ldots, X_n = x_n) = \sum_1^p \pi_i P^{(n)}(X_1 = x_1, \ldots, X_n = x_n \mid \theta = \theta_i)$$

- La loi a posteriori :

$$P(\theta = \theta_i \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{\pi_i P^{(n)}(X_1 = x_1, \ldots, X_n = x_n \mid \theta = \theta_i)}{P^{(n)}(X_1 = x_1, \ldots, X_n = x_n)}$$

calcul de la loi a posteriori : continue/discrète

- π : densité de la loi a priori
- la loi de $(X_1, \ldots, X_n)$ sachant θ est discrète

- La loi marginale :

$$P_n(X_1 = x_1, \ldots, X_n = x_n) = \int_\Theta \pi(\vartheta) P^{(n)}(X_1 = x_1, \ldots, X_n = x_n \mid \vartheta)\, d\vartheta$$

- La loi a posteriori :

$$\pi(\vartheta \mid x_1, \ldots, x_n) = \frac{\pi(\vartheta) P^{(n)}(X_1 = x_1, \ldots, X_n = x_n \mid \vartheta)}{P_n(X_1 = x_1, \ldots, X_n = x_n)}$$

calcul de la loi a posteriori : discrète / continue

- $\Theta = \{\theta_1, \ldots, \theta_p\}$ la loi a priori est définie par $\pi_i = P(\theta = \theta_i)$
- $f^{(n)}(x_1, \ldots, x_n \mid \theta)$: densité la loi de $(X_1, \ldots, X_n)$ sachant θ

- La loi marginale : $m_n(x_1, \ldots, x_n) = \sum_1^p \pi_i f^{(n)}(x_1, \ldots, x_n \mid \theta_i)$

- La loi a posteriori : $P(\theta = \theta_i \mid X_1, \ldots, X_n) = \dfrac{\pi_i f^{(n)}(X_1, \ldots, X_n \mid \theta_i)}{m_n(X_1, \ldots, X_n)}$

Remarque

- A partir de la loi a priori et la conditionnelle de X sachant θ on connait la loi a posteriori à une constante multiplicative près.

- par exemple si θ est une va continue on a

$$\pi(\vartheta \mid x_1, \ldots, x_n) \propto \pi(\vartheta) f^{(n)}(x_1, \ldots, x_n \mid \vartheta)$$

ou

$$\pi(\vartheta \mid x_1, \ldots, x_n) \propto \pi(\vartheta) P^{(n)}(X_1 = x_1, \ldots, X_n = x_n \mid \vartheta)$$

- Comme on cherche une loi de probabilité, la loi a posteriori est bien définie à partir des expressions de droite

Chapitre 2

Estimation et prévision bayésienne

Estimation bayésienne = estimation probabiliste

- a partir du modèle bayésien, on obtient une loi de probabilité sur le paramètre : la loi a posteriori
- cette loi résume l'information provenant des données $X = (X_1, \ldots, X_n)$ et de l'information a priori

Loi a priori — Observations

↓

Loi a posteriori

Estimation par intervalle

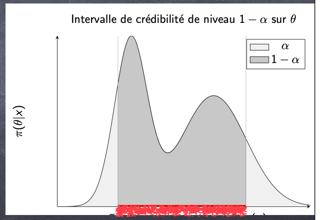
crédibilité

intervalle de crédibilité

- On fixe $\alpha \in]0,1/2[$.
  on cherche un intervalle $[l(X); u(X)] \subset \Theta$ tel que

$$P(\theta \in [l(X); u(X)] \mid X) = 1 - \alpha$$

- Interprétation : Ayant observé X, l'intervalle contient le paramètre θ avec une probabilité $1 - \alpha$



- exemple :
  $l(X) = q(\alpha/2, X)$ et $u(X) = q(1 - \alpha/2, X)$

où $q(\alpha, X)$ quantile d'ordre α de la loi a posteriori

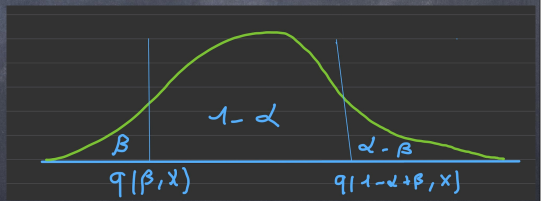
intervalles de crédibilité

- Il n'y a pas unicité des Iss'intervalle,
- tous les intervalles de la forme suivante sont des intervalles crédibilité

$$[q(\beta,X)\,;\,q(1-\alpha+\beta,X)] \quad \text{avec } 0 \le \beta \le \alpha$$

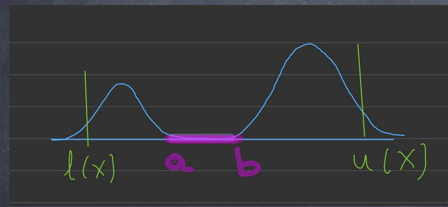
- On cherche le plus court intervalle :
  on cherche la valeur de β qui minimise la longueur

$$q(1-\alpha+\beta,X) - q(\beta,X)$$



défauts des IC

1. Soit $[l(X),u(X)]$ le plus court intervalle de crédibilité de niveau $1-\alpha$



$[a,b] \subset [l(X),u(X)]$ vérifie
$P(\theta \in [a,b] \,|\, X_1,\ldots,X_n) \approx 0$

$\Rightarrow [l(X),a] \cup [b,u(X)]$ est une région de niveau $1-\alpha$ plus courte que IC optimal

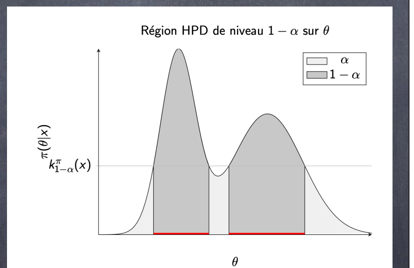
2. La généralisation en dimension supérieure est difficile

Région Highest Posterior Density (HPD)

- On cherche la plus petite région $H \subset \Theta$ telle que $P(\theta \in H \,|\, X_1,\ldots,X_n) = 1-\alpha$
- ces régions sont de la forme

$$H(K) = \{\theta \,:\, \pi(\theta \,|\, X_1,\ldots,X_n) > K\}$$

- On choisit la valeur $K := K_{1-\alpha}(X)$ telle que
  $P(\theta \in H(K_{1-\alpha}(X)) \,|\, X_1,\ldots,X_n) = 1-\alpha$



La région $\{\theta : \pi(\theta \,|\, X_1,\ldots,X_n) > K_{1-\alpha}(X)\}$ est appelé région HPD de niveau $1-\alpha$

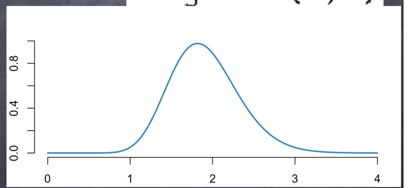
Propriétés

- la définition est indépendante de la dimension de Θ
- En dimension 1 si la distribution est unimodale alors la région HPD est un intervalle qui coincide avec le plus court intervalle de crédibilité
- la définition se généralise aux lois discrètes en prenant
  $H(K) = \{\vartheta \,:\, P(\theta = \vartheta \,|\, X_1,\ldots,X_n) > K\}$

Exemple

- On modélise $X_1, \ldots, X_n$ le nombre de pannes par une loi de poisson de paramètre $\theta > 0$
- La loi priori est la loi exponentielle de paramètre 1
- Calcul de la loi a posteriori
  - $P(X_1 = x_1, \ldots, X_n x_n \mid \vartheta) = e^{-n\vartheta} \vartheta^{\sum x_i} \frac{1}{x_1! \ldots x_n!}$
  - $\pi(\vartheta \mid) \propto e^{-\vartheta} I_{\vartheta > 0}$
  - $\pi(\vartheta \mid x_1, \ldots, x_n) \propto e^{-(n+1)\vartheta} \vartheta^{\sum x_i} I_{\vartheta > 0}$
- On observe $n = 10$ et $\sum X_i = 20$
- loi uniondale : IC et HPD coïncident

loi gamma (21,11)

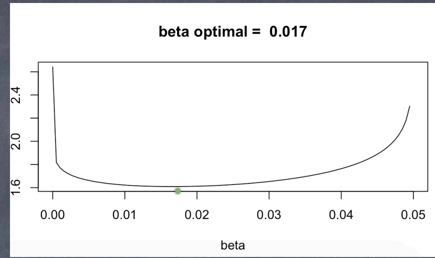


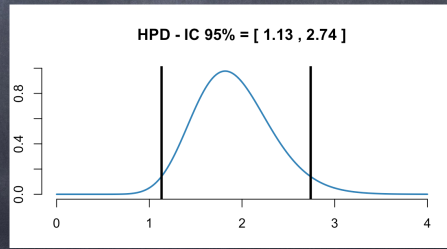
En utilisant R
qgamma quantile de la loi gamma
which.min recherche du min

(1) on calcule et represente la longueur en fonction de β

(2) on détermine la valeur de β qui minimise la longueur

(3) on en déduit le plus court intervalle de crédibilité qui coïncide avec la région HPD





Lien avec les intervalles de confiance.

- un intervalle de confiance de niveau $1-\alpha$ est une intervalle aléatoire [a(X) , b(X)] tel que, pour tout θ :

$$P_\theta(\,[a(X), b(X)] \ni \theta) = 1 - \alpha$$

$1-\alpha$ % des intervalles de confiance contiennent la vraie valeur du paramètre

- un intervalle de crédibilité de niveau $1-\alpha$ est une intervalle [l(X) u(X)] tel que

$$P(\theta \in [l(x), u(X)] \mid X) = 1 - \alpha$$

Ayant observé X, le paramètre appartient à l'intervalle avec une probabilité $1-\alpha$

Probabilité fréquentiste d'un intervalle de crédibilité

- Soit [l(X) u(X)] un intervalle de crédibilité de niveau $1-\alpha$
- Pour le modèle $\{P_\theta, \theta \in \Theta\}$, sa probabilité fréquentiste est égale à

$$P_\theta(\,[l(X), u(X)] \ni \theta) = \beta(\theta, n)$$

- ⚡⚡ En général $\beta(\theta, n) \neq 1 - \alpha$

Exemple : le modèle exponentiel

- conditionnellement à θ, $X_1,\dots,X_n$ iid suivant une loi exponentielle $\theta > 0$
- La loi priori est la gamma de paramètre (a,b)

1. La loi a posteriori est la loi gamma $(a+n, b+\sum X_i)$
2. Pour tout $\beta \in (0,\alpha)$: $\left[\dfrac{g_{n+a}(\beta)}{b+\sum X_i}; \dfrac{g_{n+a}(1-\alpha+\beta)}{b+\sum X_i}\right]$ est un intervalle de crédibilité de niveau $1-\alpha$

Notation : $g_a(\beta)$ est le quantile d'ordre β et G_a la fonction de répartition de la loi gamma (a, 1)

- le niveau fréquentiste est $\beta(n,\theta) = G_n(g_{n+a}(1-\alpha+\beta) - b\theta) - G_n(g_{n+a}(\beta) - b\theta)$
- A n fixé, quand a-> 0 et b-> 0 alors le niveau fréquentiste converge vers $1-\alpha$
- Approximation de la loi gamma
  - Quand $n \to \infty$, on a $g_n(\beta) \sim q_\beta\sqrt{n} + n$
  - Si $x_n \sim \sqrt{nx} + n$ avec x≠0 alors $G_n(x_n) \sim \Phi(x)$ quand $n \to \infty$
- A (a,b) fixés, le niveau fréquentiste converge vers $1-\alpha$ quand $n \to \infty$. L'intervalle de crédibilité de niveau $1-\alpha$ est un intervalle de confiance asymptotiquement de niveau $1-\alpha$

Notation : q_β est le quantile d'ordre β et Φ la fonction de répartition de la loi gaussienne N(0,1)

Crédibilité d'une hypothèse

- On considère le test statistique

$$H_0 : \theta \in \Theta_0 \text{ contre } H_1 : \theta \in \Theta_1$$

- On suppose que $P(\theta \in \Theta_i) \neq 0$ pour i=1,2
- Règle de décision Bayésienne :
  - On calcule les probabilités a posteriori des hypothèses :

    $P(\theta \in \Theta_i | X)$ pour i=1,2
  - On dit que H_0 est plus crédible que H_1 si

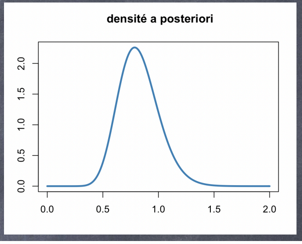
$$P(\theta \in \Theta_0 | X) \geq P(\theta \in \Theta_1 | X)$$

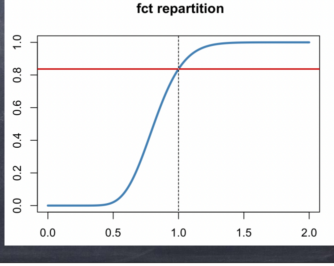
- La valeur de la probabilité a posteriori quantifie la crédibilité de l'hypothèse.

Exemple : modèle exponentiel (cont.)

- La loi a posteriori est la loi gamma $(n+a, b+\sum X_i)$
- On prend a=b= 1
- On veut tester $\theta \leq 1$ contre $\theta > 1$
- On évalue la pobabilitité a posteriori de l'hypothèse nulle, c'est à dire

$P(\theta \leq 1 | X) = G_{n+a}(b + \sum X_i)$ (= 83 %)





Prévision Bayésienne

Prévision en loi

Loi prédictive

- Objectif : on veut prévoir X_{n+1} à partir des observations passées $(X_1, \ldots, X_n)$ dans le modèle bayesien

Loi a priori

Loi conditionnelle des observations sachant le paramètre

$\Rightarrow$ Loi a posteriori $\Rightarrow$ Loi prédictive

Définition : la loi prédictive est la loi conditionnelle de X_{n+1} sachant $(X_1, \ldots, X_n)$.

Calcul de la loi prédictive

- La loi prédictive s'écrit :

$$p(x_{n+1} \mid x_1, \ldots, x_n) = \begin{cases} \int_\Theta p(x_{n+1} \mid \vartheta, x_1, \ldots, x_n)\, \pi(\vartheta \mid x_1, \ldots, x_n)\, d\theta & \text{continue} \\ \sum_{\vartheta \in \Theta} p(x_{n+1} \mid \vartheta, x_1, \ldots, x_n)\, P(\theta = \vartheta \mid x_1, \ldots, x_n) & \text{discrète} \end{cases}$$

- Où $p(x_{n+1} \mid \vartheta, x_1, \ldots, x_n)$ est la densité de la loi conditionnelle de X_{n+1} sachant θ et le passé $X_1, \ldots, X_n$:

$$p(x_{n+1} \mid \vartheta, x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_{n+1} \mid \vartheta)}{f(x_1, \ldots, x_n \mid \vartheta)}$$

La loi prédictive est un mélange de loi.

- A partir de la loi prédictive on construit

1. Un intervalle de prévision de niveau $1 - \alpha$, c'est un intervalle $[P_1 ; P_2]$ tel que

$$P(X_{n+1} \in [P_1 ; P_2] \mid X_1, \ldots, X_n) = 1 - \alpha$$

2. Un prédicteur ponctuel : Pour l'erreur L^2, la meilleure approximation de X_{n+1} par une fonction de $X_1, \ldots, X_n$ est l'espérance conditionnelle

$$\hat{X}_{n+1} = E(X_{n+1} \mid X_1, \ldots, X_n)$$

Exemple : observations iid suivant $f(\cdot\,|\,\theta)$

- L'indépendance conditionnellement à θ implique que $p(x_{n+1}\,|\,\vartheta, x_1,\ldots,x_n) = f(x_{n+1}\,|\,\vartheta)$
- D'ou la loi prédictive $p(x_{n+1}\,|\,x_1,\ldots,x_n) = \int_\Theta f(x_{n+1}\,|\,\vartheta)\,\pi(\vartheta\,|\,x_1,\ldots,x_n)\,d\vartheta$
- Le prédicateur ponctuel est $\hat{X}_{n+1} = \int_\Theta E(X_{n+1}\,|\,\vartheta)\pi(\vartheta\,|\,X_1,\ldots,X_n)\,d\vartheta$

Exemple : regression linéaire suivant $X = \theta t + \epsilon \quad \epsilon$ iid $N(0,\sigma^2)$

- L'indépendance conditionnellement à θ implique que $p(x_{n+1}\,|\,\vartheta, x_1,\ldots,x_n) = f(x_{n+1}\,|\,\vartheta)$ c'est la densité de la loi gaussienne de moyenne $\theta\, t_{n+1}$ et de variance σ^2
- D'ou la loi prédictive
- $p(x_{n+1}\,|\,x_1,\ldots,x_n) = \int_{\Theta\times R^+} \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2\sigma^2}(x_{n+1}-\theta t_{n+1})^2}\,\pi(\sigma^2,\vartheta\,|\,x_1,\ldots,x_n)\,d\vartheta d\sigma^2$
- Le prédicteur ponctuel est égal à $\hat{X}_{n+1} = E(\theta\,|\,X_1,\ldots,X_n)\,t_{n+1}$

Estimateurs de Bayes

Construction d'estimateurs à partir de la loi a posteriori

Estimateurs de Bayes

- Soit L une fonction de coût : elle permet évaluer la qualité d'un estimateur $\delta := \delta(X)$ du paramètre θ

  Exemple : $L_2(\delta,\theta) = (\delta-\theta)^2$ coût/erreur quadratique L^2
  $L_1(\delta,\theta) = |\delta-\theta|$ coût/erreur absolue L^1

  plus généralement c'est une fonction positive L: $\Theta\times\Theta \to R^+$ telle que $L(\delta;\theta) = 0 \Leftrightarrow \delta = \theta$
- Le risque bayesien d'un estimateur δ du paramètre θ est

$$r(\delta,\pi) = E(L(\delta(X),\theta)) = \int L(\delta(x),\vartheta) g_n(x,\vartheta)\,dx_1\ldots dx_n d\vartheta$$
$$= \int L(\delta(x),\vartheta)\pi(\vartheta\,|\,x)\,d\vartheta\, m_n(x_1,\ldots,x_n)\,dx_1\ldots dx_n$$

Définition : Un estimateur δ^π est un estimateur de Bayes sous coût L, s'il minimise le risque bayesien c'est à dire $r(\delta^\pi,\pi) \le r(\delta,\pi)$ pour estimateur δ

Construction des estimateurs de Bayes

- On définit $\rho(\delta) = \int_\Theta L(\delta, \vartheta)\pi(\vartheta \mid X_1, \dots X_n)\, d\vartheta$
- Théorème :

  $\delta^\pi(X) = \text{argmin}_\delta\, \rho(\delta)$ est un estimateur de Bayes
- Cas particulier
  - Coût quadratique : l'estimateur de Bayes est l'espérance de la loi a posteriori

    $\delta^\pi(X) = E(\theta \mid X_1, \dots X_n)$
  - Coût absolue : l'estimateur de Bayes est la médiane de la loi a posteriori

Propriétés asymptotiques

–

Contexte

- on suppose que le modèle $\{P_\theta^n = P^n(\cdot \mid \theta)\ \theta \in \Theta\}$ est régulier
- Sous ces hypothèse : l'estimateur du maximum de vraisemblance θ_n^{MV} converge presque surement et θ_n^{MV} est asymptotiquement efficace
- On suppose que les observations sont iid suivant f_{θ_0} où θ_0 appartient à l'intérieur de Θ
- Théorème 1
  si π_1, π_2 sont deux lois a priori telles que $\pi_i(\theta) > 0, \forall \theta \in \Theta$
  alors pour tout $A \subset \Theta$:

  $$\int_A |\pi_1(\vartheta \mid X) - \pi_2(\vartheta \mid X)|\, d\vartheta \xrightarrow[n\to\infty]{p.s.} 0$$

Théorème 2

Si π est une loi a priori telle que $\pi(\theta) > 0, \forall \theta \in \Theta$ et π est C^1 sur Θ alors

1) Pour tout intervalle ouvert U contenant θ_0
   on a $P(\theta \in U \mid X) \xrightarrow[n\to\infty]{ps} 1$

2) $E(\theta \mid X) \xrightarrow[n\to\infty]{P} \theta_0$

3) $Var(\theta \mid X) \xrightarrow[n\to\infty]{P} 0$

Théorème 3

Si π est une loi a priori telle que $\pi(\theta) > 0, \forall \theta \in \Theta$ et π est C^2 sur Θ
alors

1) On peut approcher la loi a posteriori par la loi gaussienne de moyenne $E(\theta|X)$ et de variance $Var(\theta|X)$

2) On peut approcher la loi a posteriori par la loi gaussienne de moyenne θ_n^{MV} et de variance $n^{-1}I^{-1}(\theta_n^{MV})$
où I est l'information de Fisher apportée par une observation

3) On a $\sqrt{n}(E(\theta|X) - \theta_n^{MV}) \xrightarrow[P]{n\to\infty} 0$

Conséquence du théorème 3

1) L'estimateur de Bayes sous coût quadratique c'est à dire $\delta^\pi(X) = E(\theta|X_1, \ldots X_n)$ est asymptotiquement efficace

On applique le théorème de Slutski
$\sqrt{n}(E(\theta|X) - \theta_0) = \sqrt{n}(E(\theta|X) - \theta_n^{MV}) + \sqrt{n}(\theta_n^{MV} - \theta_0)$
Le premier terme converge vers 0 en proba et le confond en loi vers $N(0, I^{-1}(\theta_0))$

4) Le niveau fréquentiste des intervalles de crédibilité de niveau $1-\alpha$ converge vers $1-\alpha$

$$\beta(\theta, n) \xrightarrow[n\to\infty]{} 1-\alpha$$

Chapitre 3

Loi a priori

Lois informatives

Loi discrète

- On suppose que le paramètre appartient à un ensemble fini $\Theta = \{\theta_1, \ldots, \theta_p\}$ avec les probabilités $\pi_1, \ldots, \pi_p$ c'est à dire $P(\theta = \theta_i) = \pi_i$
- Source d'information :
  - les résultats d'études précédentes supposées similaires
  - les avis de p experts et la proba représente la confiance accordée à chaque expert.

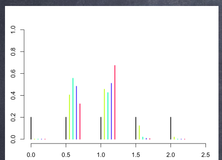
- La loi a posteriori est aussi une loi discrète à valeurs dans Θ
- Pour tout i=1…p, on a

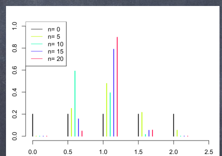
$$P(\theta = \theta_i | X) = \frac{f_n(X|\theta_i)\pi_i}{\sum_{j=1}^{p} \pi_j f_n(X|\theta_j)}$$

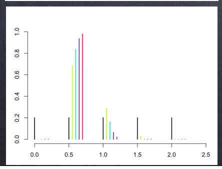
- Si $X \sim P_{\theta_0}$ avec $P(\theta = \theta_0) = 0$
  alors la loi a posteriori ne se concentre pas autour de la vraie valeur c'est à dire
  il existe un intervalle U tel que $\theta_0 \in U$ et
  $P(\theta \in U | X) \nrightarrow 1$

Illustration

- la loi a priori est uniforme sur {0, 1/2, 1 ,3/2, 2}
- les observations sont simulées suivant la loi exponentielle de paramètre 0.75







On représente la loi a posteriori en fonction de n pour 3 échantillons.

Evolution des probabilités a posteriori en fonction de n

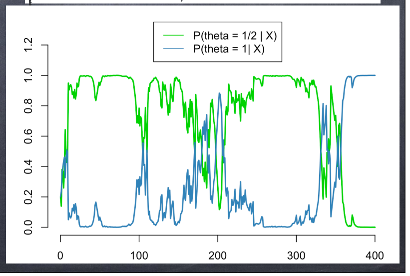
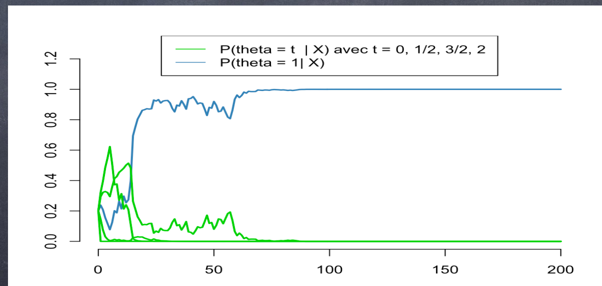


Illustration (cont.)

- les observations sont simulées suivant la loi exponentielle de paramètre 1
- On a $P(\theta = 1) = 1/5$

Evolution des probabilités a posteriori en fonction de n



- On observe que

  $P(\theta = 1 | X) \to 1$

Histogramme

- On relaxe le contrainte de finitude de $\Theta = \{\theta_1, \ldots, \theta_p\}$ en prenant comme support un intervalle
- On ordonne les valeurs θ_i, $i = 1,\ldots,p$
- on ajoute une borne inférieure : $\theta_0 < \theta_1 < \theta_2 < \cdots < \theta_p$
- On construit l'histogramme

$$\pi(\theta) = \sum_{i=1}^{p} \frac{\pi_i}{\theta_i - \theta_{i-1}} 1_{[\theta_{i-1},\theta_i[}(\theta)$$

  - Cette loi vérifie $P(\theta \in [\theta_{i-1}, \theta_i[) = \pi_i$
  - $\pi(\theta) > 0$ pour tout $\theta \in [\theta_0; \theta_p[$

- La loi a posteriori s'écrit

$$\pi(\theta \,|\, X_1, \ldots X_n) \propto \sum_{i=1}^{p} \frac{\pi_i}{\theta_i - \theta_{i-1}} f(X_1, \ldots X_n \,|\, \theta)\, 1_{[\theta_{i-1},\theta_i[}(\theta)$$

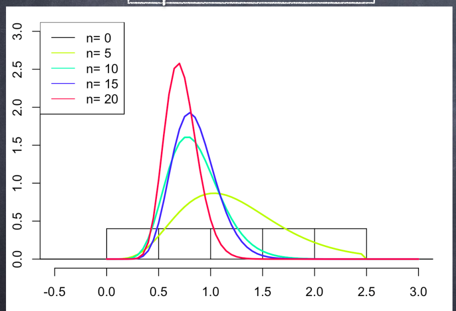
D'où

$$\pi(\theta \,|\, X_1, \ldots X_n) = \frac{\sum_{i=1}^{p} \frac{\pi_i}{\theta_i - \theta_{i-1}} f(X_1, \ldots X_n \,|\, \theta)\, 1_{[\theta_{i-1},\theta_i[}(\theta)}{\sum_{i=1}^{p} \frac{\pi_i}{\theta_i - \theta_{i-1}} \int_{\theta_{i-1}}^{\theta_i} f(X_1, \ldots X_n \,|\, \theta) d\theta}$$

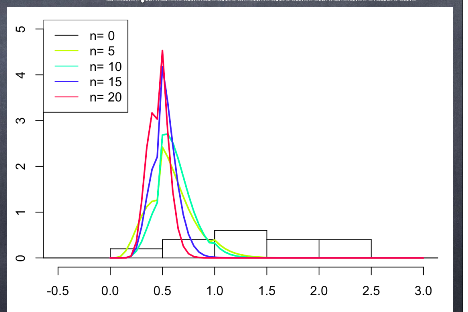
Illustration (cont.)

- les observations sont simulées suivant la loi exponentielle de paramètre .75.

A priori uniforme



A priori non uniforme



Famille de lois conjuguées

- On considère $\mathscr{P} = \{\pi_\lambda, \lambda \in \Lambda\}$ une famille paramétrique de lois sur Θ.
- Définition
  On dit que $\mathscr{P}$ est une famille conjuguée avec $\mathscr{F} = \{f_n(\cdot \,|\, \theta), \theta \in \Theta\}$ si la loi a priori appartient à $\mathscr{P}$ alors la loi a posteriori appartient aussi à $\mathscr{P}$.

  Autrement dit
  $\forall \lambda \in \Lambda$ si $\theta \sim \pi_\lambda$ alors $\exists \lambda(X) \in \Lambda$ tel que $\pi(\theta \,|\, X) = \pi_{\lambda(X)}(\theta)$

Exemple de Famille de lois conjuguées

- $\mathcal{F}$ est la famille des lois exponentielles
$f(x_1,\ldots,x_n\,|\,\theta) = \theta^n e^{-\theta n \bar{X}_n}$
  - La famille des lois Gamma est une famille de lois conjuguées :
  $\pi(\theta) = b^a \frac{1}{\Gamma(a)} e^{-b\theta}\theta^{a-1} 1_{\theta>0}$ a> 0 et b> 0.
  On a $\lambda = (a,b) \longrightarrow \lambda(X) = (n+a, b+n\bar{X}_n)$
- $\mathcal{F}$ est la famille des lois uniformes sur $[0,\theta]$
$f(x_1,\ldots,x_n\,|\,\theta) = \frac{1}{\theta^n} 1_{M_n \leq \theta}$ avec $M_n = \max(X,\ldots,X_n)$
  - La famille des lois de pareto est une famille de lois conjuguées :
  $\pi(\theta) = a\frac{b^a}{\theta^{a+1}} 1_{\theta>b}$ avec a> et b>0.
  On a $\lambda = (a,b) \longrightarrow \lambda(X) = (n+a, \max(b, M_n))$

Choix de l'hyperparamètre λ

- on fixe la valeur de λ à partir de l'information disponible a priori
- Exemple 1. Information a priori : θ est autour de 1
  - On choisit λ tel que $E(\theta) = \int \vartheta \pi_\lambda(\vartheta) d\vartheta = 1.$ En fonction de la dimension de Λ on pourra aussi ajouter une contrainte sur la variance de θ qui traduit la confiance accordée à l'information
- Exemple 2. Information a priori : $\theta \in A$ (avec une forte probabilité)
  - On fixe λ tel que $P(\theta \in A) = \int_A \pi_\lambda(\vartheta) d\vartheta = 95\,\%$ ou 80%, 99% ... en fonction de la confiance accordée à l'information

Mélange d'experts a priori

- Proposition : Soit $\mathcal{P} = \{\pi_\lambda, \lambda \in \Lambda\}$ une famille de lois conjuguées, la famille des mélanges de lois de $\mathcal{P}$ forme aussi une famille de lois conjuguées

  Rappel : un mélange s'écrit $\sum_{j=1}^{k} p_j \pi_{\lambda_j}(\theta)$ avec $(\lambda_1,\ldots,\lambda_k) \in \Lambda^k$,
  $(p_1,\ldots,p_k) \in [0,1]^k$ et $\sum_{j=1}^{k} p_j = 1$
- Application : on peut prendre en compte différentes sources d'information et accorder des poids différents en fonction de la fiabilité des sources

Illustration : modèle exponentiel

Expert 1 : $\theta = 1 \pm .5$

$\frac{a}{b} = 1 \qquad \frac{a}{b^2} = .5^2$

A priori $\Gamma(4,4)$

A posteriori $\Gamma(n+4\,, n\bar{X}_n + 4)$

Mélange

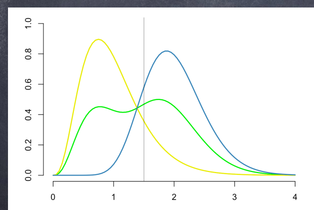
1/2 -1/2

Expert 2 : $\theta = 2 \pm .5$

$\frac{a}{b} = 2 \qquad \frac{a}{b^2} = .5^2$

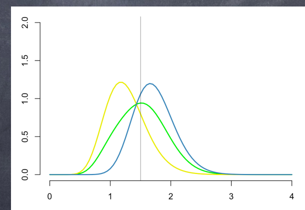
A priori $\Gamma(16,8)$

A posteriori $\Gamma(n+16\,, n\bar{X}_n + 8)$



A priori

n= 10



A posteriori

Lois non informatives

Loi a priori impropre

- On considère $\pi : \Theta \mapsto R^+$ telle que $\begin{cases} \sum_{\vartheta \in \Theta} \pi(\vartheta) = \infty & \text{discrète} \\ \int_\Theta \pi(\vartheta)\, d\vartheta = \infty & \text{continue} \end{cases}$
- π ne définit pas une loi de probabilité sur Θ
- Définition
  on dit que $\pi : \Theta \mapsto R^+$ est une loi impropre pour le modèle $\mathcal{F} = \{f_n(\cdot\,|\,\theta), \theta \in \Theta\}$ si
  $$\int_\Theta \pi(\theta) f_n(x_1, \ldots x_n\,|\,\theta)\, d\theta < \infty \text{ presque surement}.$$
- Si π est une loi impropre alors la loi a posteriori est bien définie par
  $$\pi(\theta\,|\,X) = \frac{\pi(\theta) f_n(X\,|\,\theta)}{\int_\Theta \pi(\vartheta) f_n(X\,|\,\vartheta)\, d\vartheta}$$

Si π est une loi impropre alors pour tout C>0, $\iota(\theta) = C\pi(\theta)$ est aussi une loi impropre.
A partir de ces deux lois impropres, on obtient la même loi a posteriori

Exemple : modèle exponentiel

- On considère $\pi(\theta) = 1_{R^+}(\theta)$
- on a
  $$\int_{R^+} \pi(\vartheta)\, d\vartheta = \infty$$
  $$\int_\Theta \pi(\vartheta) f_n(x_1, \ldots x_n\,|\,\vartheta)\, d\vartheta = \int_{R^+} \vartheta^n e^{-\vartheta n \bar{X}_n}\, d\vartheta < \infty \Leftrightarrow (n > 1 \text{ et } \bar{X}_n > 0).$$
- La fonction π définit donc une loi impropre si et seulement n>1
- Pour n> 1, la loi a posteriori est la loi gamma $\Gamma(n+1, n\bar{X}_n)$

Loi a priori de Laplace

- Si Θ est un ensemble fini ou de mesure de Lebesgue finie ($\int_\Theta d\vartheta < \infty$) alors la loi a priori de Laplace est la loi uniforme sur Θ
- Si $\begin{cases} \Theta \text{ infini dénombrable} \\ \sum_{\vartheta \in \Theta} f_n(X\,|\,\vartheta) < \infty \end{cases}$ ou $\begin{cases} \int_\Theta d\vartheta = \infty \\ \int_\Theta f_n(X\,|\,\vartheta)\, d\vartheta < \infty \end{cases}$

  alors la loi a priori de Laplace est une loi impropre définie par
  $$\pi(\theta) \propto 1_\Theta(\theta).$$
- Proposition : Si la loi a priori de Laplace existe alors la loi a posteriori vérifie $\pi(\theta\,|\,X) \propto f_n(X\,|\,\theta)$

Loi Non informative de Jeffreys

- On suppose que le modèle $\mathcal{F} = \{f_n(\cdot\,|\,\theta)\,,\, \theta \in \Theta\}$ est régulier
- Soit I_n l'information de Fisher et $|I_n|$ son déterminant
- Si $\int_\Theta \sqrt{|I_n(\vartheta)|}\, d\vartheta < \infty$

  ou

  si $\int_\Theta \sqrt{|I_n(\vartheta)|})\, d\vartheta = \infty$ et $\int_\Theta \sqrt{|I_n(\vartheta)|} f_n(X\,|\,\vartheta)\, d\vartheta < \infty$

  alors la loi de Jeffreys est définie par $\pi(\vartheta) \propto \sqrt{|I_n(\vartheta)|}$

Loi Non informative de Jeffreys

- La loi de Jeffreys favorise les régions où l'information de Fisher prend des grandes valeurs c'est à dire les régions où les données apportent plus d'information sur le paramètre
- La loi de Jeffreys est invariante par reparamétrisation

Exemple : modèle exponentiel

- On considère n variables aléatoires iid suivant la loi exponentielle
- L'information de Fisher est donnée par $\dfrac{n}{\theta^2}$
- $\int_O^\infty \dfrac{1}{\vartheta}\, d\vartheta = \infty$ et $\int_O^\infty \vartheta^{n-1} e^{-n\theta \bar{X}_n}\, d\vartheta < \infty$ ps car n>0 et $\bar{X}_n > 0$ ps
- La loi de Jeffreys est une loi impropre définie $\pi(\theta) \propto \dfrac{1}{\theta} 1_{R^+}(\theta)$
- La loi a posteriori est la loi gamma $\Gamma(n, n\bar{X}_n)$