Introduction à la Statistique Bayésienne

Bases théoriques et applications

Partie III

F. Mhamdi[1,2]

[1]Laboratoire des Signaux et Smart Systèmes (L3S-ENIT)

[2]Ecole Supérieure de la Statistique et de l'Analyse de l'Information (ESSAI)

ESSAI, 2024-2025

Table of Contents I

Contexte

$X_1, ..., X_n$ n-échantillon de X.
$L(X/\widetilde{\Theta} = \theta) \hookrightarrow \mathbb{P}_\theta = f(x/\theta)\mu$
$L((X_1, ..., X_n)/\widetilde{\Theta} = \theta) \hookrightarrow \mathbb{P}_\theta^n = \Pi_{k=1}^n f(x_k/\theta)\mu^{\otimes n}$
$\pi = \pi(\theta)\beth$: la loi à priori.
On a : $\Pi_{k=1}^n f(x_k/\theta)\pi(\theta) = \pi(\theta/x_1, ..., x_n) m_\pi(x_1, ..., x_n)$

Estimateur du maximum à postériori (MAP)

Contexte

$X_1, ..., X_n$ n-échantillon de X.
$L(X/\widetilde{\Theta} = \theta) \hookrightarrow \mathbb{P}_\theta = f(x/\theta)\mu$
$L((X_1, ..., X_n)/\widetilde{\Theta} = \theta) \hookrightarrow \mathbb{P}_\theta^n = \Pi_{k=1}^n f(x_k/\theta)\mu^{\otimes n}$
$\pi = \pi(\theta)\beth$: la loi à priori.
On a : $\Pi_{k=1}^n f(x_k/\theta)\pi(\theta) = \pi(\theta/x_1, ..., x_n)m_\pi(x_1, ..., x_n)$

Définition : Estimateur par maximum à postériori

L'estimateur du maximum à postériori (MAP) associé à π qu'on le note $\hat{\theta}_n^{MAP}(X_1, ..., X_n)$ est solution de :

$$\max_{\theta \in \Theta} \pi(\theta/X_1, ..., X_n)$$

ou bien

$$\max_{\theta \in \Theta} \Pi_{k=1}^n f(x_k/\theta)\pi(\theta)$$

Lois impropres

Dans les sections précédentes, l'espace du paramètre est "gouverné" par une loi de probabilité qu'on appelle loi à priori.
On peut étendre les notions ci-dessus à des mesures plus générales : les mesures σ-finies.
Soit $\daleth$ une mesure σ-finie à priori sur $\Theta(\subset E)$. Avec (E, d) un espace métrique (généralement $E = \mathbb{R}^m$).

Dans les sections précédentes, l'espace du paramètre est "gouverné" par une loi de probabilité qu'on appelle loi à priori.
On peut étendre les notions ci-dessus à des mesures plus générales : les mesures σ-finies.
Soit $\daleth$ une mesure σ-finie à priori sur $\Theta(\subset E)$. Avec (E, d) un espace métrique (généralement $E = \mathbb{R}^m$).

Définition : lois impropres

Une loi impropre est une mesure positive σ-finie, qui vérifie :
$\int_\Theta \pi(\theta) d\daleth(\theta) = +\infty$.
On suppose que $m_\pi(x) = \int_\Theta f(x/\theta)\pi(\theta)d\daleth(\theta) < +\infty$ μ.p.p.

Exemple :

$$X \hookrightarrow N(\mu, \sigma^2); \pi(\mu) = 1; \mu = \theta \in \Theta = \mathbb{R}.$$

Exemple :

$$X \hookrightarrow N(\mu, \sigma^2); \pi(\mu) = 1; \mu = \theta \in \Theta = \mathbb{R}.$$

$m_\pi(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{1}{2\sigma^2}(x-\mu)^2] d\mu < +\infty$

Donc la mesure de Lebesgue sur $\mathbb{R}$ est une loi impropre qu'on peut utiliser.

DÉFINITION : ESTIMATEUR BAYÉSIEN GÉNÉRALISÉ

On appelle estimateur bayésien généralisé associé à X (resp $(X_1, ..., X_n)$), la fonction de perte $L(.,.)$ et la mesure à priori $\daleth$ et qu'on le note $\hat{\theta}^{B.G}(X)$ (resp. $\hat{\theta}_n^{B.G}(X_1, ..., X_n)$) l'estimateur qui minimise le risque bayésien généralisé pour la mesure $\daleth$.

$$R_\pi^{B.G}(\hat{\theta}^{B.G}(X)) = \min_{\hat{\theta}(X)\in E} \int_\Theta R(\theta, \hat{\theta}) d\daleth(\theta)$$

resp.

$$R_\pi^{B.G}(\hat{\theta}_n^{B.G}(X_1, ..., X_n)) = \min_{\hat{\theta}_n(X_1,...,X_n)\in E} \int_\Theta R(\theta, \hat{\theta}_n(X_1, ..., X_n)) d\daleth(\theta)$$

Exercice 1

Soient :

- $X \hookrightarrow N(\mu, \sigma^2)$; σ^2 connue; $\eta = \frac{1}{\sigma^2}$
- La loi à priori pour $\mu \hookrightarrow N(\mu_0, \eta_0^{-1})$; avec $\eta_0 = \frac{1}{\sigma_0^2}$

1- Déterminer pour cet exemple la loi à postériori $\pi(./x)$

2- Comparer les espérances et les variances à postériori et à priori.

3- Déterminer $\delta^\pi(X)$ l'estimateur Bayésien avec la fonction perte quadratique.

4- Montrer que pour $(X_1, ..., X_n)$ un n-échantillon de X on obtient :

$\pi(\mu/x_1, ..., x_n) \propto exp[-\frac{1}{2}(n.\eta + \eta_0)(\mu - \frac{n.\eta\overline{X_n}+\eta_0\mu_0}{n.\eta+\eta_0})^2]$

5- En déduire la loi à postériori $\pi(./x)$.

6- Déterminer

$$\hat{\theta}_n^{MAP}(X_1, ..., X_n)$$

avec $\theta = \mu$.

7- Déterminer $\lim_{n\rightarrow+\infty} \hat{\theta}_n^{MAP}(X_1, ..., X_n)$

Exercice 2

Soient :

- $X \hookrightarrow Bn(n,p)$;
- π La loi à priori $\hookrightarrow \pi(u) = \frac{1}{\beta(\frac{1}{2},\frac{1}{2})} u^{-\frac{1}{2}}(1-u)^{-\frac{1}{2}}$

1- Montrer que $f(x,p) = \frac{C_n^x}{\beta(\frac{1}{2},\frac{1}{2})} p^{x-\frac{1}{2}}(1-p)^{n-x-\frac{1}{2}}$

2- Déduire que :

$$m_\pi(x) = \frac{\beta(x+\frac{1}{2}, n-x+\frac{1}{2})}{\beta(\frac{1}{2},\frac{1}{2})} C_n^x$$

3- En déduire la loi à postériori $\pi(./x)$

4- Déterminer l'estimateur Bayésien relatif à la fonction perte quadratique.

Exercice 3

Soient :

- $X \hookrightarrow N(\theta, \sigma^2)$; σ^2 connue.
- π La loi à priori sur θ de la forme $N(a, b^2)$

1- Déterminer $L(X/\widetilde{\Theta} = \theta)$

2- Déterminer $L(X, \widetilde{\Theta})$

3- Montrer que

$f(x, \theta) = \frac{1}{2\pi\sigma b} exp[-\frac{1}{2}\frac{\sigma^2+b^2}{\sigma^2 b^2}(\theta - \frac{\sigma^2 b^2}{\sigma^2+b^2}(\frac{a}{b^2} + \frac{x}{\sigma^2}))^2] exp[-\frac{1}{2}\frac{(x-a)^2}{\sigma^2+b^2}]$

4- Montrer que :

$$m_\pi(x) = \frac{1}{\sqrt{2\pi(\sigma^2+b^2)}} exp(-\frac{1}{2}\frac{(x-a)^2}{\sigma^2+b^2})$$

5- En déduire que $L(\widetilde{\Theta}/X = x) \hookrightarrow N(\frac{\sigma^2 b^2}{\sigma^2+b^2}(\frac{a}{b^2} + \frac{x}{\sigma^2}), \frac{\sigma^2 b^2}{\sigma^2+b^2})$

6- Déduire l'estimateur Bayésien relatif à la fonction perte quadratique.

Exercice 4

- $X = (X_1, ..., X_n)$ n observations indépendantes distribuées selon la loi $N(\theta, \sigma^2)$; σ^2 connue.
- π La loi à priori sur θ de la forme $\mathbb{N}(a, b^2)$

1- Déterminer $L((X_1, X_2, ..., X_n)/\widetilde{\Theta} = \theta)$
2- Déterminer $L(X_1, X_2, ..., X_n, \widetilde{\Theta})$
3- Montrer que
$f(x_1, ..., x_n, \theta) = \frac{1}{(2\pi)^{\frac{n+1}{2}}\sigma^n b} exp[-\frac{1}{2}(\frac{n}{\sigma^2} + \frac{1}{b^2})(\theta - \frac{1}{\frac{n}{\sigma^2}+\frac{1}{b^2}}(\frac{1}{\sigma^2}\sum_{k=1}^n x_k + \frac{a}{b^2}))^2] exp[-\frac{1}{2}(\frac{1}{\sigma^2}\sum_{k=1}^n x_k^2 + \frac{a^2}{b^2} - \frac{1}{\frac{n}{\sigma^2}+\frac{1}{b^2}}(\frac{1}{\sigma^2}\sum_{k=1}^n x_k + \frac{a}{b^2})^2)]$
4- En déduire la loi à postériori $L(\widetilde{\Theta}/(X_1, X_2, ..., X_n,) = (x_1, x_2, ..., x_n)) \hookrightarrow N(\frac{1}{\frac{n}{\sigma^2}+\frac{1}{b^2}}(\frac{1}{\sigma^2}\sum_{k=1}^n x_k + \frac{a}{b^2}), \frac{1}{\frac{n}{\sigma^2}+\frac{1}{b^2}})$
5- Montrer que $\delta^\pi(X_1, ..., X_n) = \frac{b^2}{b^2+\frac{\sigma^2}{n}}\overline{X_n} + \frac{\frac{\sigma^2}{n}a}{\frac{\sigma^2}{n}+b^2}$ est l'estimateur Bayésien relatif à la fonction perte quadratique.

Exercice 6

On considère maintenant la fonction perte L^1 avec $\Theta = \mathbb{R}$ c-à-d :
$L(\theta, \delta) = |\theta - \delta|$
Montrer que δ^π est la médiane de la loi à postériori.

Choix de la loi à priori

Le choix des lois a priori est une étape fondamentale en statistique bayésienne.
Les différents choix possibles peuvent être motivés par difféerents points de vue :

Le choix des lois a priori est une étape fondamentale en statistique bayésienne.
Les différents choix possibles peuvent être motivés par difféerents points de vue :

- Choix basé sur des expériences du passé ou sur une intuition du statisticien.

Choix de la loi à priori

Le choix des lois a priori est une étape fondamentale en statistique bayésienne.
Les différents choix possibles peuvent être motivés par difféerents points de vue :

- Choix basé sur des expériences du passé ou sur une intuition du statisticien.
- Choix basé sur la faisabilité des calculs.

Le choix des lois a priori est une étape fondamentale en statistique bayésienne.
Les différents choix possibles peuvent être motivés par difféerents points de vue :

- Choix basé sur des expériences du passé ou sur une intuition du statisticien.
- Choix basé sur la faisabilité des calculs.
- Choix basé sur la volonté de n'apporter aucune information nouvelle pouvant biaiser l'estimation.

Lois Subjectives

L'idée est d'utiliser les données antérieures. Dans un cas concret, il peut être judicieux de baser son raisonnement sur l'expertise de spécialistes. Par exemple, si on fait des biostatistiques, on s'appuiera sur l'expertise des médecins et des biologistes pour déterminer une loi a priori cohérente.

Définition : Conjuguées

Une famille $\mathcal{F}$ de distributions sur Θ est dite conjuguée pour la loi $f(x \mid \theta)$ si pour tout $\pi \in \mathcal{F}$; la distribution à posteriori $\pi(. \mid x)$ appartient également à $\mathcal{F}$.

Définition : Conjuguées

Une famille $\mathcal{F}$ de distributions sur Θ est dite conjuguée pour la loi $f(x \mid \theta)$ si pour tout $\pi \in \mathcal{F}$; la distribution à posteriori $\pi(. \mid x)$ appartient également à $\mathcal{F}$.

Avantages des familles conjugées

- Simplifier les calculs : Avant le développement des outils de calcul numérique, ces familles étaient pratiquement les seules qui permettaient de faire aboutir des calculs.

Définition : Conjuguées

Une famille $\mathcal{F}$ de distributions sur Θ est dite conjuguée pour la loi $f(x \mid \theta)$ si pour tout $\pi \in \mathcal{F}$; la distribution à posteriori $\pi(. \mid x)$ appartient également à $\mathcal{F}$.

Avantages des familles conjugées

- Simplifier les calculs : Avant le développement des outils de calcul numérique, ces familles étaient pratiquement les seules qui permettaient de faire aboutir des calculs.
- La mise à jour la loi se fait à travers les paramètres de la loi et donc l'interprétation est souvent bien plus facile.

$f(x \mid \theta)$	$\pi(\theta)$	$\pi(\theta \mid x)$
$\mathcal{N}\left(\theta, \sigma^2\right)$	$\mathcal{N}\left(\mu, \tau^2\right)$	$\mathcal{N}\left(\frac{\sigma^2\mu+\tau^2 x}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$
$\mathcal{P}(\theta)$	$Ga(\alpha, \beta)$	$Ga(\alpha + x, \beta + 1)$
$Ga(\nu, \theta)$	$Ga(\alpha, \beta)$	$Ga(\alpha + \nu, \beta + x)$
$B(n, \theta)$	$Be(\alpha, \beta)$	$Be(\alpha + x, \beta + n - x)$
$\mathcal{N}\left(\mu, \frac{1}{\theta}\right)$	$Ga(\alpha, \beta)$	$Ga\left(\alpha + \frac{1}{2}, \beta + \frac{(\mu-x)^2}{2}\right)$

Remarque

Une loi conjuguée peut être déterminée en considérant la forme de la vraisemblance $f(x \mid \theta)$ et en prenant une loi a priori de la même forme que cette dernière vue comme une fonction du paramètre.

REMARQUE

Une loi conjuguée peut être déterminée en considérant la forme de la vraisemblance $f(x \mid \theta)$ et en prenant une loi a priori de la même forme que cette dernière vue comme une fonction du paramètre.

EXERCICE

On considère une loi Pareto de paramètres (α, a), Supposons a connu :

$$f(x \mid \theta, a) = \frac{\theta a^{\theta}}{x^{\theta+1}} \chi_{[a,+\infty[}(x).$$

Question : Proposer une loi conjugée pour l'exemple précédent.

Remarque

Une loi conjuguée peut être déterminée en considérant la forme de la vraisemblance $f(x \mid \theta)$ et en prenant une loi a priori de la même forme que cette dernière vue comme une fonction du paramètre.

Exercice

On considère une loi Pareto de paramètres (α, a), Supposons a connu :

$$f(x \mid \theta, a) = \frac{\theta a^{\theta}}{x^{\theta+1}} \chi_{[a,+\infty[}(x).$$

Question : Proposer une loi conjugée pour l'exemple précédent.

Correction

$f(x \mid \theta) \propto \theta e^{\theta \log(a/x)} x^{-1} \chi_{[a,+\infty[}(x)$. On pourrait donc prendre une loi a priori de type Gamma.

Dans le cas où on dispose que de peu d'informations sur θ, on peut choisir des loi a priori dites peu ou non informatives. On souhaite que l'a priori intervienne de façon minimale dans la loi à postériori , i.e. que les donéees parlent d'elles même.

Rappel : Information de Fisher

On rappelle la définition de l'information de Fischer :

- **(a) Cas Unidimensionnel**

$$I(\theta) = \mathbb{E}\left[\left|\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right|^2\right]$$

qui, sous certaines conditions de régularité, peut se réécrire

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right]$$

Loi à priori de Jeffreys

Rappel : Information de Fisher

On rappelle la définition de l'information de Fischer :

- **(a) Cas Unidimensionnel**

$$I(\theta) = \mathbb{E}\left[\left|\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right|^2\right]$$

qui, sous certaines conditions de régularité, peut se réécrire

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X \mid \theta)\right]$$

- **(b) Cas multi-dimensionnel** Si $\theta \in \mathbb{R}^k$ alors $I(\theta)$ est une matrice dont les coefficients sont

$$I_{ij}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(X \mid \theta)\right].$$

Rappel : Information de Fisher

Soit $X_1, X_2, .., X_n$ n-échantillon de X d'un modèle régulier. L'information de Fisher associée à l'échantillon $(x_1, x_2, ..., x_n)$ est $I_n(\theta) = nI(\theta)$, avec $(I(\theta))_{i,j} = -E_\theta\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j} \text{ Log } f\,(x/\theta)\right)$

Définition : loi à priori de Jeffreys

La loi à priori de Jeffreys est donnée par :

- **(a) Cas Unidimensionnel ($\Theta \subset \mathbb{R}$) :**

$$\begin{aligned} \pi(\theta) &\propto \sqrt{I_n(\theta)} \\ &\propto \sqrt{I(\theta)} \end{aligned}$$

Définition : loi à priori de Jeffreys

La loi à priori de Jeffreys est donnée par :

- **(a) Cas Unidimensionnel ($\Theta \subset \mathbb{R}$) :**

$$\begin{aligned}\pi(\theta) &\propto \sqrt{I_n(\theta)}\\ &\propto \sqrt{I(\theta)}\end{aligned}$$

- **(b) Cas multi-dimensionnel ($\Theta \subset \mathbb{R}^m$) :**

$$\begin{aligned}\pi(\theta) &\propto \sqrt{\det(I_n(\theta))}\\ &\propto \sqrt{\det(I(\theta))}\end{aligned}$$

Remarques

Cette loi possède deux intérêts principaux :

- $I(\theta)$ est un indicateur de la quantité d'information apportée par le modèle $f(x \mid \theta)$. Donc il parait intuitivement justifié que les valeurs de θ pour lesquelles $I(\theta)$ est plus grande doivent être plus probables à priori.
- La loi de Jeffreys est invariante par reparamétrisation.

Remarques

Cette loi possède deux intérêts principaux :

- $I(\theta)$ est un indicateur de la quantité d'information apportée par le modèle $f(x \mid \theta)$. Donc il parait intuitivement justifié que les valeurs de θ pour lesquelles $I(\theta)$ est plus grande doivent être plus probables à priori.
- La loi de Jeffreys est invariante par reparamétrisation.

Exercice 1

Soit $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, avec σ^2 est connue et que l'on cherche à estimer $\theta = \mu$.

- Déterminer la loi de à priori de Jeffreys.
- Montrer que $\pi\left(\mu / x_1 \to x_n\right) \propto \exp\left(-\frac{n}{2\sigma^2}(\overline{x}_n - \mu)^2\right)$
- En déduire la loi à posteriori $\pi\left(\mu / x_1 \to x_n\right)$ associée à la loi à priori de Jeffreys.

Exercice 2

Soit $X_1, \ldots, X_n$ un n-échantillon d'une variable aléatoire réelle X de loi normale de moyeme 0 et de variance $\theta > 0$ incomnue.

- Calculer la loi à priori de Jeffreys associé que l'on notera π.
- Détorminer la loi à postérioni assciée à la loi à priori de Jeffreys.

Région α-crédible

Définition : Région α-crédible

Un sous ensemble C mesurable de Θ est dit α-crédible si :

$$\mathbb{P}^{\pi}(\theta \in C/X) = \int_C d\pi(\theta/X) \geq 1 - \alpha$$

Cas particulier

pour n échantillon de X :

$$\begin{aligned}\mathbb{P}^{\pi}(\theta \in C/X_1, X_2, ..., X_n) &= \int_C d\pi(\theta/X_1, X_2, ..., X_n) \\ &= \int_C \pi(\theta/x_1, x_2, ..., x_n) d\daleth(\theta) \geq 1 - \alpha\end{aligned}$$

Région α-crédible

Remarque

Il existe une infinité de région α crédible. Il est tout à fait logique de s'intéresser à la région qui a le volume minimal par rapport à la mesure $\gimel$-finie sur . C-à-d : on essaye de minimiser le volume de C définie par : $vol(C) = \gimel(C) = \int_C d\gimel(\theta)$

Notations :

Dans la suite, on pose :

Notations :

Dans la suite, on pose :

- $$\begin{aligned} C^{\pi}_{/x}(k) &= \{\theta \in \Theta / \pi(\theta/x) \geqslant k\} \\ &= (\pi(\cdot/x))^{-1}\left([k, +\infty[\right) \end{aligned}$$

Notations :

Dans la suite, on pose :

- $$C^{\pi}_{/x}(k) = \{\theta \in \Theta / \pi(\theta/x) \geqslant k\}$$
  $$= (\pi(\cdot/x))^{-1}([k, +\infty[)$$
- $$k_{\alpha}(x) = sup\{k/\mathbb{P}^{\pi}(C^{\pi}_{/x}(k)/X = x) \geq 1 - \alpha\}$$

Notations :

Dans la suite, on pose :

- $$\begin{aligned} C^{\pi}_{/x}(k) &= \{\theta \in \Theta / \pi(\theta/x) \geqslant k\} \\ &= (\pi(\cdot/x))^{-1}([k, +\infty[) \end{aligned}$$

- $$k_{\alpha}(x) = sup\{k/\mathbb{P}^{\pi}(C^{\pi}_{/x}(k)/X = x) \geq 1 - \alpha\}$$

Définition : Région H.P.D

$C^{\pi}_{\alpha/x}$ est une région à plus haute vraisemblance à postériori (H.P.D) ssi $C^{\pi}_{\alpha/x} = \{\theta \in \Theta / \pi(\theta/x) \geq k_{\alpha}(x)\}$

Région α-crédible

Théorème :

$C^{\pi}_{\alpha/x}$ est une région α-crédible et elle minimise le volume parmi les régions α-crédible.

Exercice :

On reprend l'exemple de l'exercice 4 : c-à-d :

- $X = (X_1, ..., X_n)$ n observations indépendantes distribuées selon la loi $N(\theta, \sigma^2)$; σ^2 connue.
- π La loi à priori sur θ de la forme $\mathbb{N}(a, b^2)$

On pose par la suite :

$$C^{\pi}_{/x}(k) = \{\theta \in \Theta / \pi(\theta/x) \geqslant k\} = [\alpha_k, \beta_k]$$

$$k_{\alpha}(x) = sup\{k/\mathbb{P}^{\pi}(C^{\pi}_{/x}(k)/X = x) \geq 1 - \alpha\} = [\alpha_{k_\alpha}, \beta_{k_\alpha}]$$

1- Montrer que $\beta_{k_\alpha} = h + \sqrt{z} q^{N(0,1)}_{1-\frac{\alpha}{2}}$, avec a et z sont deux quantités à déterminer.

2- Montrer que $\alpha_{k_\alpha} = h - \sqrt{z} q^{N(0,1)}_{1-\frac{\alpha}{2}}$

EXERCICE

Soit $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, avec σ^2 est connue et que l'on cherche à estimer $\theta = \mu$.

- Déterminer la région H.P.D de niveau $1-\alpha$ associée à la loi de à priori de Jeffreys.