Estimation Bayésienne

1.Problèmatique

Soit $X = (X_1, ..., X_n)$ un échantillon (n variables aléatoires indépendantes et de même loi) dont la loi dépend d'un paramètre inconnu $\theta \in \mathbb{R}$. On cherche à estimer ce paramètre θ à partir des observations $x = (x_1, ..., x_n)$ et à partir d'une information a priori sur le paramètre θ. L'approche Bayésienne suppose que l'on connaît les quantités suivantes :

- la densité a priori du paramètre θ notée $p(\theta)$ qui résume l'information dont on dispose sur le paramètre θ,
- la densité de la variable aléatoire X_i conditionnelle au paramètre θ notée $p(x_i|\,\theta)$.

On en déduit à l'aide du théorème de Bayes la loi a posteriori du paramètre θ

$$\boxed{p(\theta|\, x) = \frac{p(x|\theta)p(\theta)}{p(x)}}$$

avec

$$\begin{aligned} p(x|\,\theta) &= \prod_{i=1}^{n} p(x_i|\,\theta) \\ p(x) &= \int p(x|\,\theta)\, p(\theta)\, d\theta \end{aligned}$$

Les estimateurs Bayésiens du paramètre θ sont construits à partir de la loi a posteriori $p(\theta|\, x)$ par minimisation d'une fonction de coût appropriée. Les estimateurs Bayésiens les plus utilisés dans les applications du traitement du signal sont **l'estimateur MMSE** et **l'estimateur MAP**.

2. Estimateur MMSE

L'estimateur MMSE (Minimum Mean Square Error) de θ noté $\widehat{\theta}_{\text{MMSE}}(x)$ est l'estimateur qui minimise le coût quadratique

$$c\left(\widehat{\theta}(x)\right) = E\left[\left(\widehat{\theta}(x) - \theta\right)^2\right]$$

Il est défini par :

$$\boxed{\widehat{\theta}_{\text{MMSE}}(x) = E\left[\theta | x\right] \qquad \forall x}$$

L'estimateur MMSE est donc la moyenne de la loi a posteriori du paramètre θ.

Preuve

$$\begin{aligned} E\left[\left(\widehat{\theta}(x) - \theta\right)^2\right] &= E\left[E\left[\left.\left(\widehat{\theta}(x) - \theta\right)^2\right| x\right]\right] \\ &= \int E\left[\left.\left(\widehat{\theta}(x) - \theta\right)^2\right| x\right] p(x)dx \end{aligned}$$

Puisque $p(x) \geq 0$, l'estimateur qui minimise $E\left[\left(\widehat{\theta}(x) - \theta\right)^2\right]$ est tel que $\widehat{\theta}(x)$ minimise $E\left[\left.\left(\widehat{\theta}(x) - \theta\right)^2\right| x\right]$ pour tout x.

$$\begin{aligned} E\left[\left.\left(\widehat{\theta}(x) - \theta\right)^2\right| x\right] &= E\left[\theta^2 | x\right] - 2\widehat{\theta}(x) E\left[\theta | x\right] + \widehat{\theta}^2(x) \\ &= \text{var}\left[\theta | x\right] + \left(\widehat{\theta}(x) - E\left[\theta | x\right]\right)^2 \\ &\geq \text{var}\left[\theta | x\right] \end{aligned}$$

et l'égalité est atteinte pour

$$\boxed{\widehat{\theta}_{\text{MMSE}}(x) = E\left[\theta | x\right]}$$

3. Estimateur MAP

L'estimateur MAP de θ est construit par minimisation de la fonction de coût suivante (appelée parfois fonction de coût 0-1)

$$c\left[\left(\widehat{\theta}(x)-\theta\right)\right]=\begin{cases}1 \text{ si } \left|\widehat{\theta}(x)-\theta\right|>\frac{\Delta}{2}\\ 0 \text{ sinon}\end{cases}$$

où Δ est un nombre réel positif arbitrairement petit (mais non nul). Après quelques calculs que l'on trouvera dans la plupart des ouvrages de statistique, on obtient :

$$\boxed{\widehat{\theta}_{\text{MAP}}(x)=\arg\max_{\theta} p\left[\theta \mid x\right]}$$

L'estimateur MAP est donc obtenu en cherchant le mode de la loi a posteriori du paramètre θ.

Remarque : en minimisant la fonction de coût

$$c\left[\widehat{\theta}(x)\right]=\left|\widehat{\theta}(x)-\theta\right|$$

on obtient la médiane de la loi a posteriori $p\left[\theta \mid x\right]$.

INP ENSEEIHT

MODAP : Estimation Bayésienne

Exercice 1 :

On considère la mesure d'une résistance prélevée dans un lot de valeur nominale de précision connue. Ce lot étant fabriqué à l'issue d'un processus complexe, une résistance prélevée au hasard dans celui-ci aura une valeur modélisée par une variable aléatoire normale $\mathcal{N}(m, \sigma^2)$, m représentant la valeur nominale et $\frac{2\sigma^2}{m}$ la précision relative du lot. Une résistance est prélevée du lot et on désire connaître sa valeur θ. La valeur de la résistance est mesurée à l'aide de n appareils indépendants. On dispose donc de n observations x_i que l'on modélise par $x_i = \theta + b_i$ où $b_i \sim \mathcal{N}(0, \sigma_b^2)$. Déterminer les estimateurs de la moyenne et du maximum a posteriori du paramètre θ construits à partir de $(x_1, ..., x_n)$. Commenter les résultats obtenus (examiner en particulier les cas limites $n \to \infty$ et $\sigma^2 \to 0$).

Exercice 2 :

Considérons un problème de trafic, par exemple l'arrivée d'appels téléphoniques sur un faisceau de lignes d'un central téléphonique. On peut admettre que pour un faisceau particulier, le nombre d'arrivée d'appels pendant l'unité de temps suit une loi de Poisson de paramètre inconnu $\theta > 0$. On sait alors que sous cette hypothèse, la durée T séparant deux arrivées successives d'appels téléphoniques sur ce faisceau (avec l'unité de temps précédente) suit une loi exponentielle de paramètre θ soit :

$$f_T(t) = \theta e^{-\theta t} 1_{\mathbb{R}^+}(t)$$

On désire estimer le paramètre θ inconnu à l'aide de l'observation de n durées $t_i, i = 1, ..., n$ séparant des arrivées successives d'appels téléphoniques sur ce faisceau.

1) dans un premier temps, on suppose qu'aucune information a priori n'est disponible sur θ (à part $\theta > 0$ bien sur). Déterminer à partir des observations t_i l'estimateur du maximum de vraisemblance de θ noté $\widehat{\theta}_{\text{MV}}$.

2) Il est connu que pour l'ensemble des faisceaux téléphoniques, le nombre moyen θ d'arrivées d'appels téléphoniques pendant l'unité de temps est distribuée suivant une loi exponentielle de paramètre λ connu :

$$g(\theta) = \lambda e^{-\lambda \theta} 1_{\mathbb{R}^+}(\theta)$$

La densité $g(\theta)$ représente pour cet exemple la loi a priori sur le paramètre θ. Déterminer à partir des observations t_i les estimateurs de la moyenne a posteriori et du maximum a posteriori notés respectivement $\widehat{\theta}_{\text{MMSE}}$ et $\widehat{\theta}_{\text{MAP}}$.

3) Montrer que les estimateurs $\widehat{\theta}_{\text{MV}}$, $\widehat{\theta}_{\text{MMSE}}$ et $\widehat{\theta}_{\text{MAP}}$ sont équivalents lorsque n est grand et interprèter ce résultat.

Rappel : on rappelle la densité de la loi Gamma $G(\alpha, \beta)$ (avec $\alpha > 0$ et $\beta > 0$) :

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{\mathbb{R}^+}(x)$$

qui admet pour moyenne $E[X] = \frac{\alpha}{\beta}$ et variance $Var[X] = \frac{\alpha}{\beta^2}$.