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 $Var[X] = \frac{\alpha}{\beta^2}$.

Exercice 1

vraisemblance $x_i \sim N(\theta, \sigma_b^2)$

$$p(x_1, \dots x_n \mid \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_b^2}} \exp - \frac{(x_i-\theta)^2}{2\sigma_b^2}$$

$$= \boxed{\frac{1}{(2\pi\sigma_b^2)^{n/2}} \exp\left[-\frac{1}{2\sigma_b^2}\sum_{i=1}^{n}(x_i-\theta)^2\right]}$$

Loi a priori $\theta \sim N(m, \sigma^2)$

$$\boxed{p(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\theta-m)^2}{2\sigma^2}\right]}$$

Loi a posteriori

$$p(\theta \mid x_1 \dots x_n) \propto \exp\left[-\frac{1}{2\sigma_b^2}\sum_{i=1}^{n}(x_i-\theta)^2 - \frac{1}{2\sigma^2}(\theta-m)^2\right]$$

c'est une loi normale de densité

$$\frac{1}{\sqrt{2\pi v^2}} \exp\left[-\frac{(\theta-\mu)^2}{2v^2}\right]$$

Pour trouver v et μ, on pose $q(\theta) = \frac{(\theta-\mu)^2}{2v^2}$

$$q'(\theta) = \frac{(\theta-\mu)}{v^2} \qquad q''(\theta) = \frac{1}{v^2}$$

Mais $q(\theta) = \frac{1}{2\sigma_b^2}\sum_{i=1}^{n}(x_i-\theta)^2 + \frac{1}{2\sigma^2}(\theta-m)^2$

donc $q'(\theta) = \frac{-2}{2\sigma_b^2}\sum(x_i-\theta) + \frac{2}{2\sigma^2}(\theta-m)$

$$q''(\theta) = \frac{n}{\sigma_b^2} + \frac{1}{\sigma^2} = \frac{1}{v^2} \Rightarrow \boxed{v^2 = \frac{\sigma^2\sigma_b^2}{n\sigma^2+\sigma_b^2}}$$

②

$$g'(\theta) = -\frac{1}{\sigma_b^2}\left(\sum x_i - n\theta\right) + \frac{1}{\sigma^2}(\theta - m) = \frac{\theta - \mu}{v^2}$$

coefficient de θ : $\frac{n}{\sigma_b^2} + \frac{1}{\sigma^2} = \frac{n\sigma^2 + \sigma_b^2}{\sigma^2 \sigma_b^2} = \frac{1}{v^2}$ (on le savait !!)

coefficient constant

$$-\frac{\mu}{v^2} = \frac{-\sum x_i}{\sigma_b^2} - \frac{m}{\sigma^2}$$

donc $\mu = \frac{\sigma^2 \sigma_b^2}{n\sigma^2 + \sigma_b^2}\left(\frac{n}{\sigma_b^2}\frac{1}{n}\sum x_i + \frac{m}{\sigma^2}\right)$

$$\boxed{\hat{\mu} = \frac{n\sigma^2}{n\sigma^2 + \sigma_b^2}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) + \frac{\sigma_b^2}{n\sigma^2 + \sigma_b^2}\, m}$$

Interprétations

$n \to +\infty$ (beaucoup de données) — $\hat{\mu} \approx \frac{1}{n}\sum_{i=1}^{n} X_i$ on fait confiance aux données

$n \to 0$ (peu de données) — $\tilde{\mu} \approx m$

$\sigma_b^2 \to 0$ (précision des mesures très bonne) — $\tilde{\mu} \approx \frac{1}{n}\sum_{i=1}^{n} X_i$

$\sigma_b^2 \to +\infty$ (~~mau~~ mauvaise précision des mesures) — $\tilde{\mu} \sim m$

$\sigma^2 \to 0$ (précision de l'info a priori) — $\tilde{\mu} \approx m$

$\sigma^2 \to +\infty$ — $\tilde{\mu} \approx \frac{1}{n}\sum_{i=1}^{n} X_i$