Examen 2009-2010

Centrale - Statistiques avancées (première partie) - deuxième session

6 mai 2010

A- Régression : une nouvelle observation diminue l'incertitude

On considère un modèle de régression linéaire simple :

$$Y_i = a + bx_i + \epsilon_i, \quad i \in \{1, \dots, n\},$$

où les variables ϵ_i sont indépendantes et de même loi $\mathcal{N}(0, \sigma^2)$. On note

$$\bar{x}_n = \frac{x_1 + \cdots + x_n}{n} \quad \text{et} \quad S_n^2 = x_1^2 + \cdots + x_n^2.$$

1. Rappeler les estimateurs du maximum de vraisemblance pour $\hat{a}_n$ et $\hat{b}_n$ (on ne demande pas de justifier les formules).
2. Quelle est la loi de $\hat{b}_n$?
3. Quel est le risque quadratique de $\hat{a}_n$ pour l'estimation de a ?

On ajoute une nouvelle observation (x_{n+1}, y_{n+1}). On définit donc

$$\bar{x}_{n+1} = \frac{x_1 + \cdots + x_n + x_{n+1}}{n+1}, \quad S_{n+1}^2 = x_1^2 + \cdots + x_n^2 + x_{n+1}^2$$

et on note $\hat{a}_{n+1}$ et $\hat{b}_{n+1}$ les estimateurs du maximum de vraisemblance respectifs de a et b avec l'échantillon $(x_1, Y_1), \dots, (x_{n+1}, Y_{n+1})$.

4. Montrer que $\bar{x}_{n+1} = (n\bar{x}_n + x_{n+1})/(n+1)$.
5. Montrer que le risque quadratique de $\hat{b}_{n+1}$ est toujours plus faible que celui de $\hat{b}_n$.
6. Cela est vrai-il aussi pour l'estimation de l'ordonnée à l'origine : le risque quadratique de $\hat{a}_{n+1}$ est-il toujours plus faible que celui de $\hat{a}_n$?

B- Points de vue fréquentiste et bayésien

Pierre choisit un nombre entier strictement positif θ. Ensuite, il tire une suite de nombres aléatoires $X_1, \dots X_n$ indépendants de loi uniforme sur l'ensemble $\{1, \dots, \theta\}$, qu'il transmet à Jean. Jean cherche à estimer θ à partir des nombres fournis par Pierre.

7. Pour Jean, quel est l'estimateur du maximum de vraisemblance $\hat{\theta}_{ML}$?
8. Montrer que $\mathbb{P}\left(\hat{\theta}_{ML} \neq \theta\right) \leq \exp(-n/\theta)$.
9. Montrer que $\hat{\theta}_{ML}$ est consistant.

Jean apprend que Pierre a choisi θ au hasard : pour tout entier naturel non nul k la probabilité qu'il avait de choisir $\theta = k$ était de

$$\mathbb{P}(\theta = k) = \frac{c_a}{k^a}, \quad \text{avec} \quad c_a = \left(\sum_{k=1}^{\infty} \frac{1}{k^a}\right)^{-1}.$$

10. Déterminer la loi a posteriori, c'est-à-dire

$$\Pi_n(\{k\}) = \mathbb{P}(\theta = k | X_1, \ldots, X_n) = \sum_{x_1, \ldots, x_n \in \mathbb{N}^*} \mathbb{P}(\theta = k | X_1 = x_1, \ldots, X_n = x_n) \mathbb{1}_{\{X_1 = x_1, \ldots, X_n = x_n\}}$$

pour tout $k \geq 1$.

11. Que proposez-vous comme estimateur bayésien $\hat{\theta}_B$?
12. Montrer que $\Pi_n(\{\theta\}) \to 1$ en probabilité quand n tend vers l'infini.