Université de Montpellier
M2 SSD - HAX918X

Année 2024-2025

Examen partiel - 2 octobre 2024
Durée 1h - Documents interdits

Exercise 1 (8 pts) Les aéroports se doivent de respecter certaines normes concernant les bruits émis par des avions au décollage et à l'atterrissage. Ainsi pour les zones habitées proches d'un aéroport, la limite tolérée se situe à environ 80 décibels. Les habitants d'un des villages proches d'un aéroport assurent que le bruit atteint la valeur limite de 84 décibels en moyenne pour un certain type d'avions. L'aéroport affirme qu'il n'est que de 70 décibels. Des experts sont convoqués pour trancher entre les deux parties en présence. Ils admettent que l'intensité du bruit causé par un avion de ce type suit une loi gaussienne de moyenne θ et de variance 49. Ils enregistrent l'intensité du bruit provoqué par le passage de ces avions sur un échantillon de taille $n = 100$. Ils trouvent que $\bar{x}_{100} = \frac{1}{100}\sum_{i=1}^{100} x_i = 77$.

L'espace paramétrique est réduit à deux valeurs $\Theta \in \{70, 84\}$. Pour une loi a priori uniforme sur Θ et la fonction de perte ci-dessous (fonction de perte qui reflette le fait que le décideur souhaite avoir des certitudes quant à la mise en place couteuse d'un dispositif anti-bruit)

$$L(\theta, d) = \begin{cases} 0 & \text{si} \quad \theta = d \\ 1 & \text{si} \quad \theta = 84 \quad \text{et} \quad d = 70 \\ 3 & \text{si} \quad \theta = 70 \quad \text{et} \quad d = 84 \end{cases}$$

donner la réponse bayésienne.

Correction L'espace des paramètres est restreint à deux valeurs : $\Theta = \{70, 84\}$. En supposant une loi a priori uniforme sur Θ, et en considérant la fonction de perte $L(\theta, d)$, qui reflète le fait que le décideur souhaite être certain avant d'implanter un dispositif anti-bruit coûteux, la méthode consiste à choisir d qui minimise l'espérance a posteriori de la perte, c'est-à-dire :

$$d_{\text{bayes}} = \arg \min_{d \in \{70,84\}} \mathbb{E}[L(\theta, d) \mid \bar{x}_{100} = 77].$$

Pour cela, il est nécessaire de calculer les probabilités a posteriori des deux valeurs possibles de θ à partir de l'observation $\bar{x}_{100} = 77$. Comme $\bar{x}_{100} \sim \mathcal{N}(\theta, 49/100)$, les vraisemblances sont respectivement :

$$f(\bar{x}_{100} \mid \theta = 70) = \frac{1}{\sqrt{2\pi \times \frac{49}{100}}} \exp\left(-\frac{(\bar{x}_{100} - 70)^2}{2 \times \frac{49}{100}}\right),$$

et

$$f(\bar{x}_{100} \mid \theta = 84) = \frac{1}{\sqrt{2\pi \times \frac{49}{100}}} \exp\left(-\frac{(\bar{x}_{100} - 84)^2}{2 \times \frac{49}{100}}\right).$$

Nous remarquons que pour $\bar{x}_{100} = 77$, ces deux termes sont égaux, ceci est très intuitif dans la mesure où la valeur observée 77 est à égale distance des deux moyennes possibles, 70 et 84. Comme la probabilité a priori est uniforme, nous en déduisons que

$$\mathbb{P}^{\pi}(\theta = 77 \mid \bar{x}_{100} = 77) = \mathbb{P}^{\pi}(\theta = 84 \mid \bar{x}_{100} = 84) = 1/2.$$

Aussi, si $d = 70$ alors la moyenne a postériori de la fonction de perte est égale à 1/2 et elle égale à 3/2 si $d = 84$. La décision bayésienne est donc $\hat{\theta} = 70$.

Exercice 2 (2 pts) Le nombre x d'arrivés dans une file d'attente suit une loi binomiale négative de paramètre (k, θ) pour $x \in \mathbb{N}$
$$f(x|k,\theta) = C_x^{k+x-1}\theta^k(1-\theta)^x$$
avec $\theta \in]0,1[$. Nous supposons k fixé et souhaitons estimer θ. Nous nous plaçons dans le paradigme bayésien. Quelle est la famille de loi conjuguée pour le paramètre θ ?

Correction La famille de lois a priori conjugués pour θ est la famille des lois Beta
$$\theta \sim \text{Beta}(\alpha, \beta)\,.$$

Exercice 3 (4 pts) La durée de vie x d'un composant électrique suit une loi log-normale de paramètre $(\theta, 1)$ pour $x \in \mathbb{R}_+^*$
$$f(x|\theta,1) = \frac{1}{x\sqrt{2\pi}}\exp\left(-\frac{1}{2}[\log(x)-\theta]^2\right)$$
avec $\theta \in \mathbb{R}$.
Nous disposons d'un n-échantillon $x_1, \ldots, x_n$ de durées de vie. Nous nous plaçons dans le paradigme bayésien et supposons que θ est distribué suivant une loi gaussienne centrée réduite. Calculer l'estimateur bayésien associé à la fonction de perte quadratique.

Correction La fonction de vraisemblance pour une observation unique x_i est donnée par la distribution log-normale
$$f(x_i|\theta) = \frac{1}{x_i\sqrt{2\pi}}\exp\left(-\frac{1}{2}[\log(x_i)-\theta]^2\right)\,.$$
Pour le n-échantillon $x_1, x_2, \ldots, x_n$, la fonction de vraisemblance est le produit des vraisemblances individuelles
$$L(\theta|x_1,\ldots,x_n) = \prod_{i=1}^{n}\frac{1}{x_i\sqrt{2\pi}}\exp\left(-\frac{1}{2}[\log(x_i)-\theta]^2\right)$$
$$L(\theta|x_1,\ldots,x_n) \propto \exp\left(-\frac{1}{2}\left[n\theta^2 - 2\theta\sum_{i=1}^{n}\log(x_i)\right]\right)\,.$$
Nous supposons une loi priori gaussienne pour θ
$$\theta \sim \mathcal{N}(0,1)\,.$$
Nous avons
$$\pi(\theta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\theta^2}{2}\right)\,,$$
en utilisant le théorème de Bayes, la distribution a postériori est telle que
$$\pi(\theta|x_1,\ldots,x_n) \propto L(\theta|x_1,\ldots,x_n)\pi(\theta)\,,$$
$$\pi(\theta|x_1,\ldots,x_n) \propto \exp\left(-\frac{1}{2}\left[(n+1)\theta^2 - 2\theta\sum_{i=1}^{n}\log(x_i)\right]\right)\,.$$
Il s'agit d'une forme quadratique en θ et la distribution a posteriori est donc gaussienne
$$\theta|x_1,\ldots,x_n \sim \mathcal{N}\left(\frac{\sum_{i=1}^{n}\log(x_i)}{n+1}, \frac{1}{n+1}\right)\,.$$
L'estimateur bayésien de θ pour la fonction de perte quadratique est donc
$$\hat{\theta} = \frac{\sum_{i=1}^{n}\log(x_i)}{n+1}\,.$$

Exercice 4 (6 pts) Le temps d'attente t entre deux appels téléphoniques dans un centre d'appel suit une loi exponentielle de paramètre $\lambda > 0$. Nous disposons d'un n-échantillon $t_1, \ldots, t_n$ de temps d'attente. Nous nous plaçons dans le paradigme bayésien.

1) (3 pts) Donner la loi a priori de Jeffreys pour le paramètre λ. Est-elle propre ?

Correction Pour la distribution exponentielle, l'information de Fisher est

$$I(\lambda) = \frac{1}{\lambda^2}\,.$$

La loi a priori de Jeffreys est proportionnel à la racine carrée de l'information de Fisher

$$\pi(\lambda) \propto \left(\frac{1}{\lambda}\right) \mathbf{1}_{\lambda>0}\,.$$

Il s'agit d'une loi a priori impropre car $\int_0^{+\infty} \left(\frac{1}{\lambda}\right) = \infty$.

2) (3 pts) Pour la loi a priori de Jeffreys, calculer l'estimateur bayésien de λ associé à la fonction de perte quadratique.

Correction La fonction de vraisemblance pour la distribution exponentielle est

$$f(t|\lambda) = \lambda \exp(-\lambda t)\,.$$

Pour un échantillon $t_1, t_2, \ldots, t_n$, la vraisemblance est la suivante

$$L(\lambda|t_1, \ldots, t_n) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} t_i\right)\,.$$

En multipliant par la loi a priori de Jeffreys, nous obtenons

$$\pi(\lambda|t_1, \ldots, t_n) \propto \lambda^{n-1} \exp\left(-\lambda \sum_{i=1}^{n} t_i\right) \mathbf{1}_{\lambda>0}\,.$$

Il s'agit d'une distribution Gamma

$$\lambda|t_1, \ldots, t_n \sim \text{Gamma}\left(n, \sum_{i=1}^{n} t_i\right)\,.$$

L'estimateur de λ associé à la fonction de perte quadratique est

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i}\,.$$