Statistiques Mathématiques, Feuille 1

Modèles statistiques

Les modèles statistiques sont faciles à poser lorsqu'on dispose d'observations *iid*. Les difficultés apparaissent lorsque la donnée $\boldsymbol{x}$ possède une structure plus complexe et qu'il faut « modéliser » la loi $\mathbb{P}_\theta$ de $\boldsymbol{X}$.

1) Considérer le cas d'un modèle statistique où les observations X_i, $i = 1,\ldots, n$, sont indépendantes, chacune de loi $N(\alpha + \beta z_i, \sigma^2)$ où les z_i sont des constantes données. Dans ce contexte, les paramètres α, β et σ sont inconnus.

a) Déterminer l'espace paramétrique Θ et la vraisemblance $L(\boldsymbol{x}, \theta)$.

b) Supposer que les données X_i, $i = 1,\ldots, n$, sont indépendantes mais proviennent maintenant de la loi $N(\alpha_i + \beta z_i, \sigma^2)$, $i = 1,\ldots, n$ où le vecteur <u>inconnu</u> $\boldsymbol{\alpha} = (\alpha_1\ldots, \alpha_n)^{\mathrm{T}} \in S$, un sous-espace vectoriel de dimension $s < n$ de $\mathbb{R}^n$. On modélise cette situation en choisissant une base « intéressante » $(b_1,\ldots, b_s)$ de S et en écrivant

$$\boldsymbol{\alpha} = \sum_{j=1}^{s} a_j b_j\ .$$

Déterminer alors l'espace paramétrique Θ et la vraisemblance $L(\boldsymbol{x}, \theta)$.

c) Application de *a*) : On s'intéresse au rendement X_i de plants de tomate (en kg de tomates produites). Le i-ème plant reçoit une dose de z_i grammes d'engrais ($i = 1,\ldots, n$) et, pour chaque dose, on suppose que la relation entre z et X est constante et, en moyenne, linéaire. Si par ailleurs, les X_i sont indépendants, de loi normale avec variance σ^2, déterminer la vraisemblance du modèle statistique de cette expérience.

d) Application de *b*) : Considérer J ($< n$) groupes de n_j plants de tomate chacun ($n_1 + \ldots + n_J = n$). La variété des plants de tomate est la même à l'intérieur de chacun des groupes mais varie d'un groupe à l'autre en donnant des rendements différents. Le i-ème plant du j-ème groupe reçoit la doze z_{ij} grammes d'engrais et on mesure son rendement X_{ij}. On suppose que tous les plants sont indépendants et que, par ailleurs, les X_{ij} sont de loi normale de variance σ^2. Si la relation entre z et X est en moyenne linéaire de paramètres constants à l'intérieur de chaque groupe et avec une pente identique d'un groupe à l'autre, déterminer la vraisemblance du modèle statistique de cette expérience.

2) Une action est cotée en bourse au jours J_0. A la fin de cette journée, elle vaut X_0 € que l'on suppose de loi $N(\mu, 1)$. On suppose qu'à partir de ce moment, sa valeur à la fin de la journée J_i est donnée par $X_i = \mu + \rho(X_{i-1} - \mu) + \varepsilon_i$, où les ε_i sont *iid* de loi $N(0, 1)$ et où $|\rho| < 1$. Déterminer la vraisemblance de la donnée $\boldsymbol{x} = (x_0, x_1,\ldots, x_n)$.

Information de Fisher

3) Déterminer l'information de Fisher pour les modèles suivants :

a) $X_1, X_2, \ldots, X_n$ *iid* $\sim E(\theta)$.
b) $X_1, X_2, \ldots, X_n$ *iid* de densité $f(x;\, \theta) = \theta x^{\theta - 1}$ où $0 < x < 1$ et $\theta > 0$.
c) $X_1, X_2, \ldots, X_n$ *iid* $\sim \Gamma(\alpha, \beta)$.
d) $X_1, X_2, \ldots, X_n$ *iid* $\sim \mathrm{C}(\mu, \sigma)$.
e) $X_1, X_2, \ldots, X_n$ *iid* $\sim \mathrm{B}(\alpha, \beta)$.
f) $X_1, X_2, \ldots, X_n$ *iid* de densité $f(x;, \mu, \lambda) = \dfrac{\sqrt{\lambda}}{\sqrt{2\pi x^3}} \exp\left\{-\dfrac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}$ si $x > 0$ avec $\mu, \lambda > 0$. On donne que $\mathbb{E}(X) = \mu$ et $\mathbb{V}(X) = \mu^3/\lambda$.
g) Pour le modèle de l'exercice 1 *a*).

4) L'information de Fisher existe-elle pour la loi $U[0, \theta]$?

5) Considérer le modèle d'échantillonnage de n observation $X_1, X_2, \ldots, X_n$ *iid* $\sim Ber(\theta)$ et d'information de Fisher $\mathcal{I}_{\mathcal{M}_n}(\theta)$. Supposons que l'on reparamétrise la densité en posant $\lambda = g(\theta) = \log(\theta/(1 - \theta))$.
a) Écrire la vraisemblance des données en fonction du « nouveau » paramètre λ.
b) Déterminer l'information de Fisher $\mathcal{J}_{\mathcal{M}_n}(\theta)$ pour λ.
c) Montrer que $\mathcal{J}_{\mathcal{M}_n}(\theta) = \left[h'(\lambda)\right]^2 \mathcal{I}_{\mathcal{M}_n}(h(\lambda))$ où $h(\bullet) = g^{-1}(\bullet)$.

6) Généraliser le résultat 5-*c*) à des lois et des reparamétrisations quelconques. A titre de cas particulier, considérer le cas d'observations $X_1, X_2, \ldots, X_n$ *iid* $\sim N(\theta_1, \theta_2)$ et d'information de Fisher $\mathcal{I}_{\mathcal{M}_n}(\theta)$. On reparamétrise la vraisemblance des observations en posant

$$\lambda_1 = \frac{\theta_1}{\theta_2} \text{ et } \lambda_2 = -\frac{1}{2\theta_2}.$$

Déterminer $\mathcal{J}_{\mathcal{M}_n}(\theta)$