Problèmes inverses — Estimation par maximum de vraisemblance

Feuille d'exercices 1 — 14/04/2020

I. Modèles statistiques

Exercice 1

André et Berthe jouent à pile ou face, chacun avec sa propre pièce. Simultanément, ils lancent n fois leurs pièces de manière indépendante. A l'issue de chaque lancer, un joueur qui a obtenu pile est déclaré vainqueur si son adversaire a tiré face. Dans les autres cas, aucun vainqueur n'est déclaré. Proposer un modèle statistique pour cette expérience.

Exercice 2

Un circuit électrique est composé de deux types de diodes A et B montées en série. Les durées des vie des diodes sont indépendantes et de lois exponentielles de paramètres inconnus, éventuellement différents.

1. Quelle est la loi de la durée de vie du circuit ?

2. On observe la durée de vie de n circuits indépendants. Proposer un modèle statistique paramétrique et identifiable pour cette expérience.

3. On observe maintenant quel type de diode a défailli (A ou B). Proposer un modèle statistique identifiable pour cette expérience.

4. Proposer un expérience et un modèle statistiques tels que les deux paramètres des lois exponentielles de durées de vie des diodes soient identifiables.

Exercice 3

D'une génération à la suivante, le poids moyen des cochons d'un élevage augmente d'un facteur multiplicatif inconnu λ, mais cette tendance est perturbée de façon additive par une variable aléatoire gaussienne centrée, indépendante des générations précédentes et de variance fixe inconnue.

1. Exprimer le poids du cochon de la k-ième génération seulement en fonction de λ, du poids initial du cochon et des différentes variables aléatoires gaussiennes ajoutées à chaque génération.

2. Montrer que le poids du cochon de la k-ième génération est une variable aléatoire de loi gaussienne dont on déterminera les paramètres.

3. Quel est le modèle statistique associé à l'observation des poids des cochons des n premières générations ?

Exercice 4

On considère deux variables aléatoires réelles indépendantes X et R telles que R vaut 1 avec probabilité $1/2$ et -1 avec probabilité $1/2$, et X suit la loi $\mathcal{N}(0,1)$. Pour $\theta \in \mathbb{R}$, on note $Y_\theta = \theta R + X$.

1. Calculer la loi de Y_θ.

2. On observe n réalisations indépendantes de la loi de Y_θ, avec θ inconnu. Préciser le modèle statistique et donner la densité des lois intervenant dans ce modèle.

II. Estimation paramétrique

Exercice 5

Soit $(\mathbb{R}_+^n, \{\mathcal{Q}_\theta^{\otimes n}\}_{\theta>0})$ un modèle statistique tel que pour chaque $\theta > 0$, $\mathcal{Q}_\theta$ est la loi sur $\mathbb{R}_+$ de densité

$$f_\theta(x) = \frac{1}{2\sqrt{x\theta}} \mathbb{1}_{]0,\theta]}(x).$$

On note $(X_1, \ldots, X_n) \sim \mathcal{Q}_\theta^{\otimes n}$ et $\hat{\theta}_1 = \max_{1 \leqslant i \leqslant n} X_i$.

1. Calculer la fonction de répartition de $\hat{\theta}_1$ et en déduire sa densité.

2. Montrer que $\hat{\theta}_1$ est un estimateur de θ biaisé, mais asymptotiquement sans biais.

3. Calculer l'espérance de X_1 et en déduire à l'aide de la méthode des moments un autre estimateur $\hat{\theta}_2$ de θ, qui est sans biais.

4. Déterminer les risques quadratiques de $\hat{\theta}_1$ et $\hat{\theta}_2$. Lequel est préférable ? Sont-ils consistants ?

5. Trouver les vitesses et les lois limites des estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$.

Exercice 6

Soit le modèle statistique $(\mathbb{R}^n, \{\mathcal{Q}_\theta^{\otimes n}\}_{\theta>0})$ tel que, pour chaque $\theta > 0$, la loi $\mathcal{Q}_\theta$ admet un moment d'ordre 4 et la variance de $\mathcal{Q}_\theta$ vaut θ. On note $\hat{\sigma}_n^2$ la variance empirique, c'est-à-dire, si $(X_1, \ldots, X_n) \sim \mathcal{Q}_\theta^{\otimes n}$,

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

1. Montrer que $\hat{\sigma}_n^2$ est un estimateur consistant de θ.

2. Déterminer la vitesse et la loi limite de $\hat{\sigma}_n^2$.

Exercice 7

Soit $(\{0,1\}^n, \{\mathcal{B}(\theta)^{\otimes n}\}_{\theta \in [0,1]})$ le modèle du jeu de pile ou face introduit dans le cours. On note $\mathbb{P}_\theta = \mathcal{B}(\theta)^{\otimes n}$, $(X_1, \ldots, X_n) \sim \mathbb{P}_\theta$ et $(x_1, \ldots, x_n)$ les observations de cette expérience.

1. Proposer un estimateur $\hat{\theta}(X_1, \ldots, X_n)$ de θ par la méthode des moments.

2. Vérifier que $\theta \mapsto \mathbb{P}_\theta\{(X_1, \ldots, X_n) = (x_1, \ldots, x_n)\}$ atteint son maximum en $\hat{\theta}(x_1, \ldots, x_n)$.

Exercice 8

On considère le modèle de régression linéaire gaussien suivant $X = R\theta + \varepsilon$ où $X = (X_1, \dots, X_n)$ est un vecteur colonne d'observation dans $\mathbb{R}^n$ et $R = (R_1, \dots, R_k)$ est la matrice $n \times k$ des régresseurs (c'est-à-dire les variables explicatives des observations), où R_i est le vecteur colonne de $\mathbb{R}^n$ des valeurs du i-ième régresseur pour les n expériences réalisées. On suppose que $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 \mathrm{Id})$. Les paramètres inconnus de ce modèle sont le vecteur $\theta \in \mathbb{R}^k$ et le réel $\sigma > 0$.

1. Que signife l'hypothèse $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 \mathrm{Id})$? Expliquer comment retrouver, comme un cas particulier du précédent, le modèle apparemment plus général $Y = R\theta + \varepsilon$ où $\varepsilon \sim \mathcal{N}_n((m, \dots, m), \sigma^2 \mathrm{Id})$, où m est un paramètre inconnu supplémentaire.

2. Quel est le modèle statistique correspondant à cette expérience ? Préciser notamment quels sont les paramètres connus, les observations et les paramètres inconnus.

3. On suppose que la matrice R est de rang k (on peut toujours s'y ramener, quitte à réduire le nombre k de régresseurs). On note E le sous-espace vectoriel de $\mathbb{R}^n$ engendré par les vecteurs $\{R_1, \dots, R_k\}$ et Π_E la projection orthogonale sur E.

a) On définit l'estimateur $\hat{\theta} = (R^T R)^{-1} R^T X$ de θ. Vérifier que $\Pi_E X = R\hat{\theta}$ et montrer que $\hat{\theta}$ est un estimateur sans biais de θ.

b) Montrer que $X - \Pi_E X = \varepsilon - \Pi_E \varepsilon$ et en déduire que

$$\hat{\sigma}^2 = \frac{\|X - \Pi_E X\|^2}{n-k}$$

est un estimateur sans biais de σ^2.
(Indication : afin de calculer la loi de $\varepsilon - \Pi_E \varepsilon$, on utilisera que, pour toute base orthonormée $\{e_i\}_{1 \leqslant i \leqslant n}$ de $\mathbb{R}^n$, $G_1 e_1 + \ldots + G_n e_n$ a pour loi $\mathcal{N}_n(0, \mathrm{Id})$ si $G_1, \dots, G_n$ sont i.i.d. de loi $\mathcal{N}(0,1)$.)

c) Déterminer la densité des lois intervenant dans le modèle statistique de la question **2.** et vérifier que le maximum de la densité des observations, vue comme une fonction des paramètres inconnus (θ, σ^2), atteint son maximum en $(\hat{\theta}, \frac{n-k}{n}\hat{\sigma}^2)$.

III. Estimateur de maximum de vraisemblance (EMV)

Exercice 9

Soit $(\mathbb{R}_+^n, \{\mathcal{Q}_\theta^{\otimes n}\}_{\theta > 0})$ un modèle statistique tel que pour chaque $\theta > 0$, $\mathcal{Q}_\theta$ est la loi sur $\mathbb{R}_+$ de densité

$$\theta^2 x e^{-\theta x} \mathbb{1}_{\mathbb{R}_+}(x).$$

1. Calculer l'EMV de ce modèle.

2. Etudier le biais et le risque quadratique de l'EMV.

3. Déterminer la loi limite de l'EMV.

Exercice 10

Soit $(\mathbb{R}^n, \{\mathcal{Q}_\theta^{\otimes n}\}_{\theta\in\mathbb{R}})$ un modèle statistique tel que pour chaque $\theta \in \mathbb{R}$, $\mathcal{Q}_\theta$ est la loi de densité

$$e^{x-\theta}\mathbb{1}_{]-\infty,\theta]}(x).$$

1. Calculer l'EMV du modèle, noté $\hat{\theta}$ dans la suite.

2. Déterminer la loi de $n(\hat{\theta} - \theta)$.

3. En déduire le biais de $\hat{\theta}$.

4. Soit $\alpha \in]0,1[$. Construire un intervalle de confiance pour le paramètre θ au niveau de confiance $(1-\alpha)$.

Exercice 11

On considère une machine qui usine des pièces aéronautiques. Afin de satisfaire les spécifications du produit, une contrainte est imposée sur les tailles maximales et minimales des pièces. Le fabricant souhaite vérifier si sa machine est conforme aux spécifications et veut donc estimer les tailles maximales et minimales des pièces fabriquées par sa machine. Il suppose donc que les observations des tailles des n pièces fabriquées par sa machine suivent le modèle statistique $((\mathbb{R}_+^*)^n, \{\mathcal{U}([a,b])^{\otimes n}\}_{0<a<b})$ et il cherche à estimer les paramètres inconnus a et b.

1. Calculer l'EMV $(\hat{a}, \hat{b})$ de (a, b).

2. Prouver que $(\hat{a}, \hat{b})$ est consistant.

3. **a)** Déterminer la densité du couple $(\hat{a}, \hat{b})$.

b) En déduire la fonction de répartition de $\hat{b} - \hat{a}$.

c) Montrer que $\frac{\hat{b}-\hat{a}}{b-a}$ est une v.a. pivotale.

d) Soit $\alpha \in]0,1[$. En déduire un intervalle de confiance pour le paramètre $b-a$ au niveau de confiance $(1-\alpha)$.