TD2 de Statistique Inférentielle

Vraisemblance, Exhaustivité, Information de Fisher

1^{ere} Année Magistère 07-08

1 n-échantillon d'une loi de Poisson

On considère $X_1, X_2, \ldots, X_n$ un n-échantillon d'une loi de Poisson de paramètre θ inconnu, $\theta \in [0, \infty[$.

1) On cherche à estimer θ.
   a) L'estimateur $\overline{X} = \frac{X_1 + \ldots + X_n}{n}$ est-il sans biais ? Quel est son risque quadratique ?
   b) Montrer que, pour tout estimateur $h(X_1, \ldots, X_n)$, il existe un estimateur $\hat{h}(X_1 + \ldots + X_n)$ de θ de risque quadratique inférieur : $\overline{X}$ est une statistique exhaustive.
   c) Quel est l'estimateur du maximum de vraisemblance de θ ?
2) On cherche maintenant à estimer $e^{-l\theta}$, probablité pour que sur l expériences futures, on observe toujours 0.
   a) On propose l'estimateur, $e^{l\overline{X}}$? Est-il sans biais ? Quel est son risque quadratique ?
   b) Déterminer une fonction g de $\mathbb{N}$ dans $\mathbb{R}$ telle que $g(n\overline{X})$ soit un estimateur sans biais. Que se passe-t-il si on prend $n = 1, \quad l = 2$?

2 Loi de Poisson

Considérons le modèle dominé $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \{\mathcal{P}(\theta), \theta > 0\})^n$ où $\mathcal{P}(\theta)$ est une loi de Poisson de paramètre θ.

a) Montrer que $T(X) = \sum_{i=1}^n X_i$ est une statistique exhaustive sans utiliser le Théorème de Neyman-Fisher.
b) A l'aide du Théorème de Neyman-Fisher retrouver ce résultat.

3 Exhaustivité

Soit $(X_1, \cdots, X_n)$ un n-échantillon de loi générique celle de X. Le support de la loi de X est $v_1, \cdots, v_{k+1}$ avec les probabilités respectives $\theta_1, \cdots, \theta_{k+1}$, $(\sum_{i=1}^{k+1} \theta_i = 1)$. Notons l'ensemble des paramètres :

$$\Theta = \left\{ \theta = (\theta_1, \cdots, \theta_{k+1}), \theta_i > 0 \text{ et } \sum_{i=1}^{k+1} \theta_i = 1 \right\},$$

θ inconnu. Soit N_j le nombre des X_j qui sont égaux à v_j.

a) Notons $N = (N_1, \cdots, N_k)$. Montrer que N est exhaustive pour θ.
b) Quelle est la distribution de $(N_1, \cdots, N_{k+1})$?

4 Modèle exponentiel

a) Utiliser la propriété de reparamétrisation des familles exponentielles pour calculer dans le cas d'un n-échantillon $(X_1, \cdots, X_n)$ de loi gaussienne de paramètres μ et σ :

$$\mathbb{E}_\theta(\sum X_i), \mathbb{E}_\theta(\sum X_i^2), \mathrm{Var}_\theta(\sum X_i), \mathrm{Var}_\theta(\sum X_i^2).$$

b) Même question pour le modèle de Rayleigh : $f(x,\theta) = (x/\theta^2)\exp\left(-x^2/2\theta^2\right)$, pour $x > 0$ et $\theta > 0$.

5 Modèles exponentiels

Montrer que les distributions suivantes peuvent se mettre sous la forme d'un modèle exponentiel en θ, $\theta > 0$, dont on précisera à chaque fois le paramètre naturel et la satistique exhaustive.

- Loi de Poisson : $P(y = k) = e^{-\theta}\dfrac{\theta^k}{k!}$, $k \in \mathbb{N}$.
- Loi binomiale : $P(y = k) = \begin{pmatrix} n \\ k \end{pmatrix} \theta^k (1-\theta)^{n-k}$, $k = 1, \ldots, n$.
- Loi binomiale négative : $P(y = k) = \begin{pmatrix} n+k-1 \\ n-1 \end{pmatrix} \theta^n (1-\theta)^k$, $k \in \mathbb{N}$.
- Loi gamma : $f(x) = \dfrac{\theta^a}{\Gamma(a)} e^{-\theta x} x^{a-1}$, $x > 0$, $a > 0$.
- Loi de Weibull : $f(x) = \theta a x^{a-1} \exp\left(-\theta x^a\right)$, $x > 0$, $a > 0$.
- Loi de Pareto : $f(x) = \dfrac{\theta a^\theta}{x^{\theta+1}}$, $a > 0$, $x > a$.

6 Loi uniforme sur $[0, \theta]$

On considère une suite de variables aléatoires réelles $X_1, \ldots, X_n$ indépendantes, de même loi uniforme sur l'intervalle $[0, \theta]$, $\theta > 0$.

a) Donner une statistique exhaustive pour ce modèle.
b) Montrer que cette statistique est totale.
c) Trouver un estimateur sans biais de variance minimale de l'espérance $\theta/2$.

7 Estimation d'une probabilité de panne

Soit $X_1, \ldots, X_n$ une suite indépendante de temps de panne, supposés distribués selon une loi exponentielle de paramètre θ, $\theta > 0$.
Touver un estimateur sans biais de variance minimale de la probabilité de panne à l'instant t, $t > 0$ fixé.

8 Loi log-normale

Soit $X_1, \ldots, X_n$ une suite de variables aléatoires réelles positives, indépendantes de même loi, telle que $\log X_i$ soit distribué selon une loi normale de paramètre θ réel inconnu, et devariance 1.

a) Montrer que $\sum_{i=1}^n \log X_i$ est une statistique exhaustive totale et minimale.
b) Donner une estimateur sans biais de variance minimale de θ.
c) Donner l'information de Fisher du modèle en θ.
d) Donner un estimateur sans biais de e^θ.

9 Loi Hypergéométrique

Pour N et n fixes, on considère $\Theta = \{0, 1, \ldots, N\}$ et la famille $(P_\theta)_{\theta\in\Theta}$ des lois hypergéométriques sur $\Omega = \{0, 1, \ldots, n\}$ donnée par

$$P_\theta(k) = \frac{\binom{\theta}{k}\binom{N-\theta}{n-k}}{\binom{N}{n}}, \quad 0 \le k \le \inf(n, \theta).$$

On observe X de loi P_θ. Donner un estimateur uniformément de variance minimale parmi les estimateurs sans biais de θ. Calculer cette variance.

10 Efficacité

Soit $X_1, \ldots, X_n$ un n-échantillon de la loi F_θ. Discuter de l'efficacité de $\overline{X}$ comme estimateur de la moyenne dans les cas suivants : F_θ est la loi $\mathcal{B}(1, \theta)$, $\mathcal{N}(\theta, 1)$, $\mathcal{P}(\theta)$, $\mathcal{E}(\theta)$, Géométrique(θ) et enfin $\mathcal{U}([0, \theta])$.

11 Loi de Binomiale

Soit $X_1, \ldots, X_n$ un n-échantillon de la loi $\mathcal{B}(1, \theta)$, $0 < \theta < 1$. On pose $T = \overline{X}(1 - \overline{X})$ et $S = \frac{n}{n-1}\overline{X}(1 - \overline{X})$. On veut estimer $f(\theta) = \theta(1 - \theta)$.

a) Donner un estimateur uniformément de variance minimale parmi les estimateurs sans biais de $f(\theta)$.

b) Quelle est la borne de Cramér-Rao pour les estimateurs sans biais de $f(\theta)$?

c) On veut montrer que $\mathrm{Var}_\theta(S) = \frac{1}{n}\varphi(\theta) + O(n^{-2})$ et calculer $\varphi(\theta)$.

  α) Calculer $E_\theta f(\overline{X})$ et $E_\theta f(\overline{X})^2$ en utilisant la formule de Taylor appliquée à $f(\overline{X}) - f(\theta)$ et $f(\overline{X})^2 - f(\theta)^2$.

  β) Vérifier que $E_\theta(\overline{X} - \theta)^3$ et $E_\theta(\overline{X} - \theta)^4$ sont de l'ordre de $O(n^{-2})$.

  γ) Conclure.

d) Calculer $\mathrm{Var}_\theta(S)$.

12 Super efficacité

Soit $X_1, \ldots, X_n$ un n-échantillon de la loi $\mathcal{N}(\theta, 1)$ et on pose $\widehat{X} = \overline{X}\,\mathbb{1}_{|\overline{X}|>n^{-1/4}}$. Calculer le risque quadratique $R_n(\theta)$ de $\widehat{X}$ pour estimer θ ainsi que $\lim_{n\to\infty} nR_n(\theta)$ en $\theta = 0$. En déduire que pour $\theta = 0$ et pour n grand, $\widehat{X}$ est meilleur que $\overline{X}$, bien que $\overline{X}$ soit efficace. L'inégalité de Cramér-Rao est-elle contredite ?

13 Support d'une loi uniforme

Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes de même loi uniforme sur $[\theta - 1/2, \theta + 1/2]$. Quel est l'estimateur du maximum de vraisemblance de θ ?

14 Temps de panne censuré

On suppose que le temps de première panne de certains appareils suit une loi géométrique de paramètre θ (sur une échelle de temps *discrète*, le temps étant compté par exemple en jours, en mois, ...). On observe n objets et on note les instants de pannes inférieurs strictement à $r + 1$. Quel est l'estimateur du maximum de vraisemblance de θ ?

15 Loi de Pareto

On appelle la loi de Pareto unilatérale de paramètres α et r la loi de densité

$$f_{\alpha,r}(x) = \frac{\alpha r^{\alpha}}{x^{\alpha+1}} \mathbb{1}_{(x>r)} \quad \alpha > 0,\, r > 0.$$

a) Calculer sa moyenne (pour $\alpha > 1$) et sa variance (pour $\alpha > 2$).
b) On observe désormais un n-échantillon de cette loi. Donner une statistique exhaustive pour le paramètre (α, r).
c) En supposant α connu, déterminer l'estimateur du maximum de vraisemblance pour r et calculer sa loi. Cet estimateur est-il sans biais ?