MAP433 Statistique
PC3: maximum de vraisemblance

11 septembre 2015

Exercice 1 (Durées de connection)**.** On peut modéliser la durée d'une connection sur le site `www.Cpascher.com` par une loi gamma(a, b) ($a > 0$, $b > 0$) de densité

$$p_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \mathrm{e}^{-bx} \mathbb{1}_{[0,+\infty)}(x).$$

On note $\theta = (a, b)$. Pour fixer vos tarifs publicitaires, vous voulez estimer le paramètre θ à partir d'un échantillon $X_1, \ldots, X_n$ de n durées de connexion.

1. Proposez un estimateur par la méthode des moments.
2. Ecrire les équations de vraisemblance. Déterminer pour une valeur de a fixée, le maximum $\hat{b}_n(a)$ de la fonction de vraisemblance.
3. Montrer que l'estimateur du maximum de vraisemblance est $\hat{\theta}_n = (\hat{a}_n, \hat{b}(\hat{a}_n)$ où $\hat{a}_n$ est le maximum de la fonction $a \mapsto L_n(a)$
$$L_n(a) = na \ln(a) - na \ln(\bar{X}_n) - n \ln(\Gamma(a)) + n(a-1)\overline{\ln(X)}_n - na$$
où $\overline{\ln(X)}_n = n^{-1} \sum_{i=1}^n \ln(X_i)$.
4. Proposez une méthode numérique pour déterminer l'estimateur du maximum de vraisemblance.
5. Question bonus : vous trouverez sur moodle un fichier de données (format texte). Implémentez la méthode pour calculer l'estimateur du maximum de vraisemblance. Vous proposerez aussi une méthode pour calculer des régions de confiance par bootstrap (voir la méthode du bootstrap sur moodle).

Exercice 2 (Modèle exponentiel)**.** Considérons une famille de fonctions de répartition $\{F_\theta,\ \theta \in \Theta\}$ ayant une densité par rapport à la mesure de Lebesgue sur $\mathbb{R}$ de la forme

$$p_\theta(x) = c(\theta) \exp\left(\theta f(x) + h(x)\right).$$

On suppose que Θ est un intervalle ouvert de $\mathbb{R}$, et $c(\cdot) \in C^2$, $c(\theta) > 0$ pour tout $\theta \in \Theta$. On note $\varphi(\theta) := \mathbb{E}_\theta\left(f(X)\right) = -\frac{d}{d\theta}\log(c(\theta))$.

Soit $X_1, \ldots, X_n$ un échantillon i.i.d. de densité p_θ, avec θ inconnu. Calculez l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ (s'il existe).

Exercice 3 (Modèle d'autorégression)**.** On considère les observations $X_1, \ldots, X_n$, où les X_i sont issus du *modèle d'autorégression* :

$$X_i = \theta X_{i-1} + \xi_i, \quad i = 1, \ldots, n, \qquad X_0 = 0,$$

avec ξ_i i.i.d. de loi normale $\mathcal{N}(0, \sigma^2)$ et $\theta \in \mathbb{R}$. Calculez l'estimateur du maximum de vraisemblance $(\hat{\theta}_n, \hat{\sigma}_n^2)$ de (θ, σ^2).

Exercice 4. Soient $\{(Y_i, Z_i)\}_{i=1}^n$ n vecteurs aléatoires i.i.d. ; On suppose que Y_1 et Z_1 sont indépendants et distribués suivant des lois exponentielles de paramètres $\lambda > 0$ and $\mu > 0$.

1. On observe $\{(Y_i, Z_i)\}_{i=1}^n$. Donnez le modèle statistique et déterminez l'estimateur du MV de λ et μ.
2. Déterminer la distribution limite de cet estimateur.
3. On observe $\{(X_i, \Delta_i)\}_{i=1}^n$ où $X_i = \min(Y_i, Z_i)$ et $\Delta_i = 1$ si $Y_i = X_i$ $\Delta_i = 0$ autrement. Donnez le modèle statistique et l'estimateur de vraisemblance de λ et μ dans ce modèle.
4. Déterminer la distribution limite de cet estimateur.

Exercice 5 (Répartition de génotypes dans une population)**.** Quand les fréquences de gènes sont en équilibre, les génotypes AA, Aa et aa se manifestent dans une population avec probabilités $(1-\theta)^2$, $2\theta(1-\theta)$ et θ^2 respectivement, où θ est un paramètre inconnu. Plato *et al.* (1964) ont publié les données suivantes sur le type de haptoglobine dans un échantillon de 190 personnes :

Type de haptoglobine :

$Hp - AA$	$Hp - Aa$	$Hp - aa$
10	68	112

1. Comment interpréter le paramètre θ ? Proposez un modèle statistique pour ce problème.
2. Calculez l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ.
3. Donnez la loi asymptotique de $\sqrt{n}(\hat{\theta}_n - \theta)$.