MAP433 Statistique
PC3: maximum de vraisemblance

12 septembre 2014

1. Durées de connection

On peut modéliser la durée d'une connection sur le site `www.Cpascher.com` par une loi gamma$(2, 1/\theta)$ de densité

$$\theta^{-2} x e^{-x/\theta} 1_{[0,+\infty[}(x).$$

Pour fixer vos tarifs publicitaires, vous voulez estimer le paramètre θ à partir d'un échantillon $X_1, \dots, X_n$ de n durées de connexion. On vous donne $\mathbb{E}_\theta(X_i) = 2\theta$ et $\mathrm{var}_\theta(X_i) = 2\theta^2$.

1. Calculez l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ.
2. Que vaut $\mathbb{E}(\hat{\theta}_n)$? Quelle est la variance de $\hat{\theta}_n$?

2. Modèle exponentiel

On rappelle que le modèle exponentiel non-courbé est une famille de lois $\{\mathbb{P}_\theta,\ \theta \in \Theta\}$ ayant une densité

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

par rapport à une mesure μ que l'on supposera être la mesure de Lebesgue sur $\mathbb{R}$ ou la mesure de comptage sur $\mathbb{N}$. On supposera que Θ est un intervalle ouvert de $\mathbb{R}$, et $c(\cdot) \in C^2$, $c(\theta) > 0$ pour tout $\theta \in \Theta$ et on rappelle que $\varphi(\theta) := \mathbb{E}_\theta\left(f(X)\right) = -\frac{d}{d\theta}\log(c(\theta))$.

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

3. Modèle d'autorégression

On considère les observations $X_1, \dots, X_n$, où les X_i sont issus du *modèle d'autorégression* :

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

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).

4. 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 $\sqrt{n}(\hat{\theta}_n-\theta)$.