1 Exercices

Une grande partie des modèles utilisés en pratique sont des modèles exponentiels (modèle gaussien, log-normal, exponentiel, gamma, Bernoulli, Poisson, etc). Nous allons étudier quelques propriétés de ces modèles. Les modèles exponentiels sont traités dans l'annexe E du polycopié.

Exercice 1 (Modèle exponentiel canonique). Soit $(\mathsf{X},\mathscr{X})$ un espace mesurable (nous prendrons $\mathsf{X}=\mathbb{R}^k$ ou $\mathsf{X}=\mathbb{N}^k$) et μ une mesure σ-finie sur $(\mathsf{X},\mathscr{X})$. Soit $T:\mathsf{X}\to\mathbb{R}$ et $h:\mathsf{X}\to\mathbb{R}^+$ deux fonctions mesurables. On appelle modèle exponentiel canonique associé au couple (T,h) une famille de lois ayant une densité par rapport à μ de la forme

$$x\mapsto q(\eta;x)=h(x)\exp\left(\eta T(x)-A(\eta)\right)\ ,\quad x\in\mathsf{X}\ , \tag{1}$$

où $A(\eta)$ est défini par :

$$A(\eta)=\log\int h(x)\exp\left(\eta T(x)\right)\mu(\mathrm{d}x)\ . \tag{2}$$

L'*espace des paramètres naturels* de la famille canonique associée à (T,h) est l'ensemble

$$\Xi=\{\eta\in\mathbb{R}:\ |A(\eta)|<\infty\}\ .$$

1. Montrer que la loi exponentielle de densité

$$x\mapsto p(\eta;x)=\eta\exp(-\eta x)\mathbb{1}_{\mathbb{R}^+}(x)$$

définit un modèle exponentiel canonique sur $\mathsf{X}=\mathbb{R}$. Préciser l'espace des paramètres naturels.

2. Montrer que la loi gaussienne $\mathrm{N}(\eta,1)$ définit un modèle exponentiel canonique. Précisez l'espace des paramètres naturels.

Nous allons tout d'abord établir certaines propriétés de ces modèles exponentiels canoniques

3. Montrer que l'espace des paramètres naturels est un sous-ensemble convexe de $\mathbb{R}$.
4. Montrer que la fonction A est convexe.

On suppose dans la suite que Ξ est un intervalle ouvert.

5. Soit g une fonction mesurable telle que pour tout $\eta\in\Xi$,

$$\int|g(x)|\exp\left(\eta T(x)\right)\mu(\mathrm{d}x)<\infty\ .$$

Pour $\eta\in\Xi$, on pose

$$G(\eta)=\int g(x)\exp\left(\eta T(x)\right)\mu(\mathrm{d}x)\ .$$

Montrer que G est infiniment différentiable sur Ξ et que, pour tout $k\in\mathbb{N}^*$ et $\eta\in\Xi$,

$$G^{(k)}(\eta)=\int g(x)T^k(x)\exp\left(\eta T(x)\right)\mu(\mathrm{d}x)\ .$$

Soit $(X_1,\dots,X_n)$ un n-échantillon du modèle exponentiel canonique associé à (T,h).

6. Déterminer l'estimateur des moments associé à la fonction T.
7. Déterminer l'estimateur du maximum de vraisemblance. Que remarque-t-on ?

Exercice 2 (Modèle exponentiel général). De façon générale, on appelle famille exponentielle toute famille de loi de densité

$$x \mapsto p(\theta;x) = h(x)\exp(\varphi(\theta)T(x) - B(\theta))$$

par rapport à une mesure σ-finie sur $\mathbb{R}$ ou $\mathbb{N}$.

On supposera que Θ est un intervalle ouvert de $\mathbb{R}$.

1. Montrer que la loi de Poisson définit un modèle exponentiel avec

$$\varphi(\theta) = \log(\theta), \quad B(\theta) = \theta, \quad T(x) = x, \quad h(x) = 1/x!.$$

   Préciser le modèle exponentiel canonique associé.

2. Montrer que la loi binomiale définit un modèle exponentiel avec

$$\varphi(\theta) = \log\left(\frac{\theta}{1-\theta}\right), \quad B(\theta) = -n\log(1-\theta), \quad T(x) = x, \quad h(x) = \binom{n}{x}.$$

   Préciser le modèle exponentiel canonique associé.

On suppose que la fonction φ définit un difféomorphisme de Θ sur Ξ l'espace des paramètres naturels associé. On dispose d'un n-échantillon $(X_1,\dots,X_n)$ du modèle

$$(\mathsf{X}, \mathscr{X}, \{p_\theta \cdot \mu \,:\, \theta \in \Theta\})$$

3. Déterminer l'estimateur du maximum de vraisemblance du paramètre θ [on pensera à utiliser la paramétrisation canonique].

Exercice 3 (Modèle auto-régressif). On considère l'observation $Z = (X_1,\dots,X_n)$, où les X_i sont issus du processus d'autorégression :

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

où $(\xi_1,\dots,\xi_n)$ sont des variables aléatoires i.i.d. de loi normale $\mathrm{N}(0,\sigma^2)$ et $\phi \in \mathbb{R}$.

1. Écrire le modèle statistique engendré par l'observation Z.
2. Déterminer l'estimateur du maximum de vraisemblance de $\theta = (\phi,\sigma^2)$.

Exercice 4. Soient $t_1,\dots,t_n$ des réels tels que $t_i \neq t_j$ pour deux indices $i \neq j$. On considère le modèle statistique

$$\left(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \left\{\bigotimes_{i=1}^n \mathrm{N}(\beta_1 + \beta_2 t_i, \sigma^2) \,:\, (\beta_1,\beta_2,\sigma) \in \mathbb{R}^2 \times \mathbb{R}_+^*\right\}\right) .$$

On note $\mathbf{1}$ et $\mathbf{t}$ les vecteurs de $\mathbb{R}^n$ définis par $\mathbf{1} = (1,\dots,1)^T$ et $\mathbf{t} = (t_1,\dots,t_n)^T$. Dans la suite, on pose

$$s = \frac{1}{n}\|\mathbf{t}\|^2 - (\bar{\mathbf{t}})^2 , \quad \text{avec} \quad \bar{\mathbf{t}} = n^{-1}\sum_{i=1}^n t_i .$$

1. Montrer que $s \neq 0$.
2. Déterminer les estimateurs du maximum de vraisemblance $(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2)$ de $(\beta_1, \beta_2, \sigma^2)$.
3. Déterminer la loi de $(\hat{\beta}_1, \hat{\beta}_2)$.
4. Montrer que $\hat{\beta}_1$ et $\hat{\beta}_2$ sont indépendants si et seulement si $\bar{\mathbf{t}} = 0$.
5. Déterminer la loi de $\hat{\sigma}^2$.
6. Montrer que $\hat{\sigma}^2$ et $(\hat{\beta}_1, \hat{\beta}_2)$ sont indépendants.

Exercice 5 (Durée de vie). On considère un système ne fonctionnant que si c'est le cas de deux machines de types différents. Les durées de vie X_0 et X_1 des deux machines suivent des lois exponentielles de paramètres λ_0 et λ_1. Les variables aléatoires X_0 et X_1 sont supposées indépendantes.

1. Montrer qu'une variable aléatoire X suit la loi exponentielle $\mathscr{E}(\lambda)$ si et seulement si
$$\forall x > 0 : \; \mathbb{P}(X > x) = \exp(-\lambda x).$$
2. Calculer la probabilité pour que le système ne tombe pas en panne avant la date t. En déduire la loi de la durée de vie Z du système. Le système tombe en panne, calculer la probabilité pour que la panne soit due à une défaillance de la machine 1.
3. Soit $I = 1$ si la panne du système est due à une défaillance de la machine 1, $I = 0$ sinon. Calculer $\mathbb{P}(Z > t; I = \delta)$ pour tout $t \geq 0$ et $\delta \in \{0, 1\}$. En déduire que Z et I sont indépendantes.
4. On dispose de n systèmes identiques et fonctionnant indépendamment les uns des autres dont on observe les durées de vie $Z_1, \dots, Z_n$.
   (a) Écrire le modèle statistique correspondant. Les paramètres λ_0 et λ_1 sont-ils identifiables ?
   (b) Supposons maintenant que l'on observe à la fois les durées de vie des systèmes $Z_1, \dots, Z_n$ et les causes de la défaillance correspondantes $I_1, \dots, I_n$, $I_i \in \{0, 1\}$. Écrire le modèle statistique dans ce cas. Les paramètres λ_0 et λ_1 sont-ils identifiables ?