Université de Montpellier — Année 2015-2016
HMMA304 - Modèles linéaires généralisés

Examen de contrôle continu
sans document - 2 novembre 2015

Exercice 1 (4 pts)

Nous considérons le modèle de régression linéaire tel que pour tout $i \in \{1, \ldots, n\}$, nous avons

$$Y_i = \alpha x_i + E_i$$

où α est un paramètre inconnu et E_i est un variable aléatoire réelle telle que : $\mathbb{E}(E_i) = 0$, $\mathbb{V}(E_i) = \sigma^2 k_i$ où $k_i > 0$ pour tout $i \in \{1, \ldots, n\}$ et $cov(E_i, E_j) = 0$ lorsque $i \neq j$.

Posons $Y = (Y_1, \ldots, Y_n)$, $X = (x_1, \ldots, x_n)$, $E = (E_1, \ldots, E_n)$ et $\mathbb{V}(E) = \Sigma$.

1 (1 pt) Donner l'expression de l'estimateur des Moindres Carrés Ordinaires $\widehat{\alpha}$ de α. Calculer sa variance.

2 (1 pt) Calculer l'expression de l'estimateur des Moindres Carrés Généralisés $\widetilde{\alpha}$ de α. Calculer sa variance.

3 (2 pts) Comparer $\mathbb{V}(\widehat{\alpha})$ et $\mathbb{V}(\widetilde{\alpha})$.

Exercice 2 (6 pts)

On considére n variables aléatoires indépendantes de loi de Poisson de paramètre $\mu_i = \mathbb{E}(Y_i)$. Nous supposons que $\log(\mu_i) = \alpha + \beta x_i$ où α et β sont des paramètres inconnus.

1 (1 pt) Montrer qu'il s'agit d'un modèle linéaire généralisée.

2 (3 pts) Donner la log-vraisemblance et les équations de vraisemblance. Pouvons-nous calculer les expressions analytique des estimateurs du maximum de vraisemblance de α et β ?

3 (2 pts) Calculer la matrice d'information de Fisher apportée par $(Y_1, \ldots, Y_n)$ sur les paramètres α et β notée $I_n(\alpha, \beta)$.

Exercice 3 (6 pts)

On considère n variables aléatoires indépendantes $Y_1, \ldots, Y_n$ telles que $Y_i \sim \mathcal{B}(\pi_i)$. Pour tout $i = 1, \ldots, n$, on suppose que $F^{-1}_{\mathcal{E}(1)}(\pi_i) = \alpha + \beta x_i$ où $x_i \in \mathbb{R}$ est supposé connu, α et β sont des paramètres inconnus, et $F^{-1}_{\mathcal{E}(1)}()$ est l'inverse de la fonction de répartition d'un loi exponentielle de paramètre 1.

1 (1 pt) Montrer qu'il s'agit d'un modèle linéaire généralisée.

2 (3 pts) Donner la log-vraisemblance et les équations de vraisemblance. Pouvons-nous calculer les expressions analytique des estimateurs du maximum de vraisemblance de α et β ?

3 (2 pts) Calculer la matrice d'information de Fisher apportée par $(Y_1, \ldots, Y_n)$ sur les paramètres α et β notée $I_n(\alpha, \beta)$.

Exercice 4 (4 pts)

On considère n variables aléatoires indépendantes $Y_1, \ldots, Y_n$ telles que $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$. On suppose que $n = kT$ où k et T sont des entiers donnés strictement supérieurs à 1. Enfin, on suppose que l'application qui à i associe μ_i est périodique de période T ($\mu_{i+jT} = \mu_i$ pour tout $i = 1, \ldots, T$ et pour tout $j = 0, \ldots, k-1$). On pose $\theta_i = \mu_i$ pour tout $i = 1, \ldots, T$ et $\theta = (\theta_1, \ldots, \theta_T)$.

1 (1 pt) Écrire le modèle linéaire précédent, appelé modèle (1), sous forme matricielle $Y = X\theta + E$, il s'agit de préciser la matrice X.

2 (2 pts) Pour tout $i = 1, \ldots, T$, donner l'estimateur des moindres carrés $\widehat{\theta_i}$ de θ_i. On pose $\widehat{\theta} = \left(\widehat{\theta_1}, \ldots, \widehat{\theta_T}\right)$. Donner la distribution de $\widehat{\theta}$.

3 (1 pt) Proposer un estimateur $\widehat{\sigma^2}$ non biaisé de σ^2.

Correction examen de contrôle continu HMMA 304
02/11/2015

Exercice 1

1] $\hat{\alpha} = \arg\min\limits_{\alpha \in \mathbb{R}} \sum_{i=1}^{n} (y_i - \alpha x_i)^2$

$$\hat{\alpha} = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2$$

2] $\tilde{\alpha} = \arg\min\limits_{\alpha \in \mathbb{R}} \sum_{i=1}^{n} \frac{(y_i - \alpha x_i)^2}{k_i}$

$$\tilde{\alpha} = \sum_{i=1}^{n} \frac{x_i}{k_i} y_i \Big/ \sum_{i=1}^{n} \frac{x_i^2}{k_i}$$

3] $\mathbb{V}(\hat{\alpha}) = \sigma^2 \sum x_i^2 k_i \Big/ \left(\sum x_i^2\right)^2$

$$\mathbb{V}(\tilde{\alpha}) = \sigma^2 \Big/ \sum \frac{x_i^2}{k_i}$$

$$\mathbb{V}(\hat{\alpha}) \geq \mathbb{V}(\tilde{\alpha})$$

$$\iff \quad \sum x_i^2 k_i \Big/ \left(\sum x_i^2\right)^2 \geq \frac{1}{\sum \frac{x_i^2}{k_i}}$$

$$\iff \quad \left(\sum x_i^2\right)^2 \leq \sum x_i^2 k_i \sum \frac{x_i^2}{k_i}$$

D'après l'inégalité de Cauchy Schwarz, nous avons

$$\left(\sum x_i^2\right)^2 = \left(\sum \sqrt{k_i}\, x_i \frac{x_i}{\sqrt{k_i}} \right)$$

$$\leq \left(\sum k_i x_i^2 \right) \left(\sum \frac{x_i^2}{k_i} \right)$$

Exercice 2

1) Modèle de régression poissonienne

$$f(y_i; \alpha, \beta) = \frac{e^{-e^{\alpha+\beta x_i}}}{y_i!} e^{y_i(\alpha+\beta x_i)} \mathbb{1}_{\mathbb{N}}(y_i)$$

$$f(y; \theta, \varphi) = \exp\left\{ \frac{y\theta - b(\theta)}{\varphi} + c(y, \varphi) \right\}$$

par rapport à la mesure de référence $\nu(dy)$

Ici $c(\phi, y_i) = \dfrac{\mathbb{1}_{\mathbb{N}}(y_i)}{y_i!}$, $\theta_i = \alpha + \beta x_i$, $\phi = 1$

et $b(\theta_i) = e^{\theta_i}$.

Nous sommes bien dans le cas de la famille exponentielle avec un paramètre de nuisance.

Par ailleurs $\mathbb{E}(y_i) = e^{\alpha + \beta x_i}$

$$= \pi(\alpha + \beta x_i)$$

La fonction de lien réciproque de π est la fonction $\log(.)$.

2) $L(\alpha, \beta) = \log\left[\prod_{i=1}^{n} f(y_i; \alpha, \beta)\right]$

$y_i \in \mathbb{N}$ $\forall i = 1, \dots, n$

$$L(\alpha, \beta) = -\sum_{i=1}^{n} \log(y_i!) + \sum_{i=1}^{n} y_i(\alpha + \beta x_i)$$

$$- \sum_{i=1}^{n} e^{\alpha + \beta x_i}$$

Les équations de vraisemblance sont

$$\frac{dL}{d\alpha}(\alpha,\beta) = 0 \text{ et } \frac{dL}{d\beta}(\alpha,\beta) = 0.$$

Nous obtenons

$$\frac{dL}{d\alpha}(\alpha,\beta) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} e^{\alpha+\beta x_i} = 0$$

$$\frac{dL}{d\beta}(\alpha,\beta) = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i e^{\alpha+\beta x_i} = 0$$

Nous ne sommes pas en mesure de déterminer une solution analytique au système d'équation précédent. Comme $L(\ )$ est strictement concave (on le montre facilement, voir question suivante), on ne peut pas ainsi calculer l'expression analytique de l'estimateur du maximum de vraisemblance.

3

3] $\frac{d^2 L}{(d\alpha)^2}(\alpha,\beta) = -\sum_{i=1}^{n} e^{\alpha+\beta x_i}$

$$\frac{d^2 L}{(d\beta)^2}(\alpha,\beta) = -\sum_{i=1}^{n} x_i^2 e^{\alpha+\beta x_i}$$

$$\frac{d^2 L}{d\alpha\, d\beta}(\alpha,\beta) = -\sum_{i=1}^{n} x_i e^{\alpha+\beta x_i}$$

$$\underline{I}_n = \begin{bmatrix} \sum_{i=1}^{n} e^{\alpha+\beta x_i} & \sum_{i=1}^{n} x_i e^{\alpha+\beta x_i} \\ \sum_{i=1}^{n} x_i e^{\alpha+\beta x_i} & \sum_{i=1}^{n} x_i^2 e^{\alpha+\beta x_i} \end{bmatrix}$$

Exercice 3

1] $f(y_i;\alpha,\beta) = \left[F_{\Sigma(1)}(\alpha+\beta x_i)\right]^{y_i} \left[1 - F_{\Sigma(1)}(\alpha+\beta x_i)\right]^{1-y_i} \mathbb{1}_{\{0,1\}}(y_i)$

$$F_{\Sigma(1)}(u) = \left[1 - e^{-u}\right] \mathbb{1}\{u \geq 0\}$$

è finita ----

Esercizio 4

1] $Y = (Y_1, ----, Y_T, Y_{T+1}, ----, Y_{2T}, ----, Y_{kT})$

$$X = \begin{bmatrix} I_T \\ \hline I_T \\ \hline \vdots \\ \hline I_T \end{bmatrix}$$ matrice di dimensione $kT \times T$

2] $\widehat{\Theta} = (X^T X)^{-1} X^T Y$

$$\Rightarrow \widehat{\theta}_j = \frac{1}{k} \sum_{i=1}^{k} Y_{j+(i-1)T}$$

$\forall j = 1, ----, T$

3] $$\widehat{\sigma}^2 = \left(\frac{1}{kT - T}\right) \sum_{j=1}^{T} \sum_{i=1}^{k} \left[Y_{j+(i-1)T} - \widehat{\theta}_j\right]^2$$

$$\mathbb{E}(\widehat{\sigma}^2) = \sigma^2, \forall \sigma^2 > 0$$