Université Paris Dauphine Année 2003-2004
U.F.R. Mathématiques de la Décision
MD4 - Modèles Linéaires et Généralisations

Examen Terminal

Durée deux heures, Sans Documents, Deux exercices indépendants

Exercice 1 (13 pts)

On considère n variables aléatoires indépendantes $Y_1, \ldots, Y_n$ telles que $Y_i \sim \mathcal{N}\left(\mu_i, \sigma^2\right)$. 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 (1.5 pt) 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.

4 (1.5 pt) Donner un intervalle de confiance de niveau 0.95 pour le paramètre θ_i.

5 (5 pts) On suppose dans cette question que μ_i est de la forme

$$\mu_i = \alpha_0 + \sum_{j=1}^{l} \left(\alpha_j \cos\left(\frac{2\pi ji}{T}\right) + \beta_j \sin\left(\frac{2\pi ji}{T}\right) \right)$$

où, $j = 1, \ldots, l$, α_0, α_j et β_j ($j = 1, \ldots, l$) sont des paramètres inconnus et l est donné tel que $2l + 1 < T$.

5.1 On note $\tau = (\alpha_0, \alpha_1, \beta_1, \ldots, \alpha_l, \beta_l)$. Écrire le modèle précedent, appelé modèle (2), sous forme matricielle $Y = U\tau + \widetilde{E}$, il s'agit de préciser la matrice U.

5.2 Donner les estimateurs des moindres carrés $\widehat{\alpha_0}$, $\widehat{\alpha_j}$ et $\widehat{\beta_j}$ ($j = 1, \ldots, l$) de α_0, α_j et β_j. On note $\widehat{\tau} = \left(\widehat{\alpha_0}, \widehat{\alpha_1}, \widehat{\beta_1} \ldots, \widehat{\alpha_l}, \widehat{\beta_l}\right)$. Donner la distribution de $\widehat{\tau}$.
On rappelle que pour tout $j \in \mathbb{N}^*$ et pour tout $k \in \mathbb{N}^*$:

$$\sum_{i=1}^{n} \sin\left(\frac{2\pi ji}{T}\right) = \sum_{i=1}^{n} \cos\left(\frac{2\pi ji}{T}\right) = \sum_{i=1}^{n} \cos\left(\frac{2\pi ji}{T}\right) \sin\left(\frac{2\pi ki}{T}\right) = 0$$

$$\sum_{i=1}^{n} \cos\left(\frac{2\pi ji}{T}\right) \cos\left(\frac{2\pi ki}{T}\right) = \sum_{i=1}^{n} \sin\left(\frac{2\pi ji}{T}\right) \sin\left(\frac{2\pi ki}{T}\right) = \begin{cases} 0 & \text{si} \quad j \neq k \\ n/2 & \text{sinon} \end{cases} .$$

5.3 Montrer que $Im(U)$ est inclus dans $Im(X)$ et ainsi que le modèle (2) et un sous-modèle du modèle (1) par une hypothèse linéaire.
On rappelle que tout $j \in \mathbb{N}^*$ et pour tout $k \in \mathbb{N}^*$:

$$\cos\left(\frac{2\pi j}{T}\right) = \cos\left(\frac{2\pi j(T+k)}{T}\right)$$
$$\sin\left(\frac{2\pi j}{T}\right) = \sin\left(\frac{2\pi j(T+k)}{T}\right).$$

5.4 Expliciter le test de Fisher de l'hypothèse H_0 selon laquelle le modèle (2) est vrai contre l'hypothèse H_1 selon laquelle le modèle (1) est vrai.

6 (3 pts) On désire prédire Y_h pour $h > n$. Pour cela donner un estimateur $\widehat{\mu_h}$ de μ_h en extrapolant seulement l'hypothèse de périodicité du modèle (1). Donner la loi de probabilité de $Y_h - \widehat{\mu_h}$ et en déduire un intervalle de prévision de Y_h, i.e. un intervalle aléatoire dépendant seulement des Y_i $(i = 1 \ldots, n)$ et contenant Y_h avec une probabilité donnée et égale à $1 - \gamma$.

Exercice 2 (7 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 et α, β sont des paramètres inconnus.

1 (2 pts) Donner la log-vraisemblance, les équations de vraisemblances et la matrice d'information de Fisher apportée par $(Y_1, \ldots, Y_n)$ sur les paramètres α et β notée $I_n(\alpha, \beta)$.

2 (1 pt) Nous considérons le test de $H_0 : \beta = 0$ contre $H_1 : \beta \neq 0$. Donner la région critique du test de Wald.

3 (1 pt) Pour $n = 100$, nous avons obtenu les résultats suivants :

```
Analysis of Parameter Estimates

Parameter    DF  Estimate   Standard
                            Error
Interecept    1      1.57       1.23
X             1      0.9        2.45
Scale         1      1.00       0.00
```

Identifier les estimations du maximum de vraisemblance des paramètres α et β et conclure pour le test de la question 2).

4 (3 pts) On suppose maintenant que $Y_i \sim \mathcal{B}(\pi)$. On se place dans le paradigme bayésien et on suppose que π suit une loi Béta de paramètres (a, b). Donner la loi a posteriori de π. Expliquer clairement comment tester $H_0 : \pi > 0.5$ contre $H_1 : \pi \leq 0.5$.
On rappele que si U suit une loi Béta de paramètres (a, b), $f_U(u) \propto u^{a-1}(1-u)^{b-1}$.