Université Paris Dauphine — Année 2001-2002
U.F.R. Mathématiques de la Décision
MD4 - Modèles Linéaires et généralisations

Correction Examen 17/12/2001

Exercice 1

1 Nous avons $X'X = \sum_{i=1}^{n} x_i^2$ et $X'Y = \sum_{i=1}^{n} x_i Y_i$.

Par conséquent, $\widehat{\alpha} = (X'X)^{-1} X'Y = \dfrac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$.

Par ailleurs, $\mathbb{V}(\widehat{\alpha}) = \dfrac{\sum_{i=1}^{n} x_i^2 \mathbb{V}(Y_i)}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \sigma^2 \dfrac{\sum_{i=1}^{n} x_i^2 k_i}{\left(\sum_{i=1}^{n} x_i^2\right)^2}$.

2 Nous avons $X'\Sigma^{-1}X = \dfrac{1}{\sigma^2}\sum_{i=1}^{n} \dfrac{x_i^2}{k_i}$ et $X'\Sigma^{-1}Y = \dfrac{1}{\sigma^2}\sum_{i=1}^{n} \dfrac{x_i Y_i}{k_i}$.

Par conséquent, $\widetilde{\alpha} = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}Y = \dfrac{\sum_{i=1}^{n} \frac{x_i Y_i}{k_i}}{\sum_{i=1}^{n} \frac{x_i^2}{k_i}}$.

Par ailleurs, $\mathbb{V}(\widetilde{\alpha}) = \mathbb{V}\left(\dfrac{\displaystyle\sum_{i=1}^{n}\frac{x_i Y_i}{k_i}}{\displaystyle\sum_{i=1}^{n}\frac{x_i^2}{k_i}}\right) = \dfrac{\displaystyle\sum_{i=1}^{n}\frac{x_i^2\mathbb{V}(Y_i)}{k_i^2}}{\left(\displaystyle\sum_{i=1}^{n}\frac{x_i^2}{k_i}\right)^2} = \dfrac{\sigma^2}{\displaystyle\sum_{i=1}^{n}\frac{x_i^2}{k_i}}.$

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

$$\left(\sum_{i=1}^{n} x_i^2\right)^2 = \left(\sum_{i=1}^{n}\sqrt{k_i}x_i\frac{x_i}{\sqrt{k_i}}\right)^2 \leq \sum_{i=1}^{n} k_i x_i^2 \times \sum_{i=1}^{n}\frac{x_i^2}{k_i}.$$

Par conséquent, $\dfrac{\displaystyle\sum_{i=1}^{n} x_i^2 k_i}{\left(\displaystyle\sum_{i=1}^{n} x_i^2\right)^2} \geq \dfrac{1}{\displaystyle\sum_{i=1}^{n}\frac{x_i^2}{k_i}}$

De ce fait, $\mathbb{V}(\widehat{\alpha}) \geq \mathbb{V}(\widetilde{\alpha})$.

4 Nous avons $\widetilde{\alpha} - \alpha \sim \mathrm{N}\left(0, \dfrac{\sigma^2}{\displaystyle\sum_{i=1}^{456}\frac{x_i^2}{k_i}}\right)$

et $R = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{456}(Y_i - \widetilde{\alpha}x_i)^2 \sim \chi^2(455)$.

Par ailleurs, les variables aléatoires réelles $\widetilde{\alpha}$ et R sont indépendantes.

Par conséquent, $\sqrt{455\displaystyle\sum_{i=1}^{456}\frac{x_i^2}{k_i}}\left(\dfrac{\widetilde{\alpha} - \alpha}{\sqrt{\displaystyle\sum_{i=1}^{456}(Y_i - \widetilde{\alpha}x_i)^2}}\right) \sim \mathrm{T}(455)$

À l'aide de la variable pivotale précédente, nous construisons un intervalle de confiance de niveau 95 % pour le paramètre α.

La densité de la loi de Student étant symétrique par rapport à l'axe des ordonnées, nous choisissons un intervalle de confiance à risques symétriques. Nous obtenons :

$$\begin{aligned} \mathrm{IC}_{95\%}(\alpha) &= \left[0.6 - 1.965\left(\frac{\sqrt{6}}{\sqrt{455\times 25}}\right), 0.6 + 1.965\left(\frac{\sqrt{6}}{\sqrt{455\times 25}}\right)\right] \\ &= [0.555, 0.645]. \end{aligned}$$

Exercice 2

1 Soit L la log-vraisemblance de $x_1, \ldots, x_n, y_1, \ldots, y_n$.

Nous avons :

$$L\left(x_1, \ldots, x_n, y_1, \ldots, y_n; \mu_X, \mu_Y, \sigma^2\right) =$$

$$-2n\log\left(\sqrt{2\pi}\right) - 2n\log\left(\sqrt{\sigma^2}\right) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}\left(x_i - \mu_X\right)^2 + \sum_{i=1}^{n}\left(y_i - \mu_Y\right)^2\right).$$

Par ailleurs, $\dfrac{\delta L}{\delta \mu_X} = \dfrac{n}{\sigma^2}\left(\overline{x} - \mu_X\right)$, $\dfrac{\delta L}{\delta \mu_Y} = \dfrac{n}{\sigma^2}\left(\overline{y} - \mu_Y\right)$

et $\dfrac{\delta L}{\delta \sigma^2} = -\dfrac{n}{\sigma^2} + \dfrac{1}{2\sigma^4}\left(\sum_{i=1}^{n}\left(x_i - \mu_X\right)^2 + \sum_{i=1}^{n}\left(y_i - \mu_Y\right)^2\right)$

Par conséquent, $\widehat{\mu_X} = \overline{X}$, $\widehat{\mu_Y} = \overline{Y}$

et $\widehat{\sigma^2} = \dfrac{1}{2n}\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right)$.

2 Nous avons : $\dfrac{\delta^2 L}{\left(\delta \mu_X\right)^2} = -\dfrac{n}{\sigma^2}$, $\dfrac{\delta^2 L}{\left(\delta \mu_Y\right)^2} = -\dfrac{n}{\sigma^2}$, $\dfrac{\delta^2 L}{\delta \mu_X \delta \mu_Y} = 0$,

$\dfrac{\delta^2 L}{\delta \mu_X \delta \sigma^2} = -\dfrac{n\left(\overline{x} - \mu_X\right)}{\sigma^4}$, $\dfrac{\delta^2 L}{\delta \mu_Y \delta \sigma^2} = -\dfrac{n\left(\overline{y} - \mu_Y\right)}{\sigma^4}$

et enfin $\dfrac{\delta^2 L}{\left(\delta \sigma^2\right)^2} = \dfrac{n}{\sigma^4} - \dfrac{\left(\sum_{i=1}^{n}\left(x_i - \mu_X\right)^2 + \sum_{i=1}^{n}\left(y_i - \mu_Y\right)^2\right)}{\sigma^6}$.

Par conséquent, $\mathrm{I}\left(\mu_X, \mu_Y, \sigma^2\right) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 & 0 \\ 0 & \frac{n}{\sigma^2} & 0 \\ 0 & 0 & \frac{n}{\sigma^4} \end{bmatrix}$.

3 Test de Wald $H_0 : \{\mu_X = \mu_Y\} : \{h(\mu_X, \mu_Y, \sigma^2) = \mu_X - \mu_Y = 0\}$.

Nous avons $\sqrt{n}\left[\left(\overline{X} - \overline{Y}\right) - (\mu_X - \mu_Y)\right] \longrightarrow \mathrm{N}(0, 2\sigma^2)$.

Par conséquent, sous H_0 :

$$W = \frac{n^2\left(\overline{X} - \overline{Y}\right)^2}{\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2\right)} \longrightarrow \chi^2(1).$$

Pour un risque de première espèce asymptotiquement égal à α, nous rejetons H_0 si $\dfrac{n^2\left(\overline{x} - \overline{y}\right)^2}{\left(\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 + \sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2\right)} > F^{-1}_{\chi^2(1)}(1-\alpha)$.

Test du rapport de vraisemblance
$H_0 : \{\mu_X = \mu_Y\} : \{h(\mu_X, \mu_Y, \sigma^2) = \mu_X - \mu_Y = 0\}$.

Soit L^0 la log-vraisemblance de $x_1, \ldots, x_n, y_1, \ldots, y_n$ sous l'hypothèse H_0. Soit $\widehat{\mu_X}^0$ et $\widehat{\sigma^2}^0$ les estimateurs du maximum de vraisemblance de μ_X et σ^2 sous l'hypothèse H_0.
Nous avons $\widehat{\mu_X}^0 = \dfrac{\overline{X} + \overline{Y}}{2}$
et $\widehat{\sigma^2}^0 = \dfrac{1}{2n}\left(\sum_{i=1}^{n}\left(X_i - \widehat{\mu_X}^0\right)^2 + \sum_{i=1}^{n}\left(Y_i - \widehat{\mu_X}^0\right)^2\right)$.
La statistique du rapport de vraisemblance est alors :

$$R = 2\left(L\left(X_1, \ldots, X_n, Y_1, \ldots, Y_n; \widehat{\mu_X}, \widehat{\mu_Y}, \widehat{\sigma^2}\right)\right.$$
$$\left. - L^0\left(X_1, \ldots, X_n, Y_1, \ldots, Y_n; \widehat{\mu_X}^0, \widehat{\sigma^2}^0\right)\right).$$

Sous l'hypothèse H_0, nous avons $R \longrightarrow \chi^2(1)$. Pour un risque de première espèce asymptotiquement égal à α, nous rejetons H_0 si $r > F^{-1}_{\chi^2(1)}(1-\alpha)$ où r est la réalisation de la variable aléatoire R.

4 Nous avons $\dfrac{400(3-4.2)^2}{80} = 7.2 > 3.841$.

Par conséquent, d'après le test de Wald nous rejetons l'hypothèse selon laquelle $\mu_X = \mu_Y$.

Exercice 3

1 Soit L la log-vraisemblance de $y_1, \ldots, y_n$. Nous avons :

$$L\left(y_1, \ldots, y_n; \alpha, \beta\right) = -\sum_{i=1}^{n} \exp\left(\alpha + \beta x_i\right) + \sum_{i=1}^{n} y_i \left(\alpha + \beta x_i\right) - \sum_{i=1}^{n} \log\left(y_i!\right).$$

Par conséquent, les équations de vraisemblance sont alors :

$$-\sum_{i=1}^{n} \exp\left(\alpha + \beta x_i\right) + \sum_{i=1}^{n} y_i = 0,$$

$$-\sum_{i=1}^{n} x_i \exp\left(\alpha + \beta x_i\right) + \sum_{i=1}^{n} y_i x_i = 0$$

Elles sont très difficiles à manipuler et nous ne pouvons pas mettre en évidence les expressions analytiques des estimateurs du maximum de vraisemblance.

2 Nous avons :

$$\frac{\delta^2 L}{(\delta\alpha)^2} = -\sum_{i=1}^{n} \exp\left(\alpha + \beta x_i\right),$$

$$\frac{\delta^2 L}{(\delta\beta)^2} = -\sum_{i=1}^{n} x_i^2 \exp\left(\alpha + \beta x_i\right),$$

$$\frac{\delta^2 L}{\delta\alpha\delta\beta} = -\sum_{i=1}^{n} x_i \exp\left(\alpha + \beta x_i\right).$$

Par conséquent,

$$\mathrm{I}(\alpha, \beta) = \begin{bmatrix} \sum_{i=1}^{n} \mu_i & \sum_{i=1}^{n} x_i \mu_i \\ \sum_{i=1}^{n} x_i \mu_i & \sum_{i=1}^{n} x_i^2 \mu_i \end{bmatrix}.$$

3 Nous avons

$$(\mathrm{I}(\alpha,\beta))^{-1} = \left(\frac{1}{\sum_{i=1}^{n} \mu_i \sum_{i=1}^{n} x_i^2 \mu_i - \left(\sum_{i=1}^{n} x_i \mu_i \right)^2} \right) \begin{bmatrix} \sum_{i=1}^{n} x_i^2 \mu_i & -\sum_{i=1}^{n} x_i \mu_i \\ -\sum_{i=1}^{n} x_i \mu_i & \sum_{i=1}^{n} \mu_i \end{bmatrix}$$

Par conséquent,

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \longrightarrow$$

$$\mathrm{N}\left(0, \frac{n \sum_{i=1}^{n} \exp(\beta x_i)}{\exp(\alpha) \left(\sum_{i=1}^{n} \exp(\beta x_i) \sum_{i=1}^{n} x_i^2 \exp(\beta x_i) - \left(\sum_{i=1}^{n} x_i \exp(\beta x_i) \right)^2 \right)} \right).$$

4 D'après la question précédente, sous l'hypothèse H_0, nous avons :

$$T = \frac{1}{n} \left(\widehat{\beta}\right)^2 \exp(\widehat{\alpha}) \left(n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right) \longrightarrow \chi^2(1).$$

Ainsi, pour un risque de première espèce asymptotiquement égal à α, nous rejetons H_0 si $t > F^{-1}_{\chi^2(1)}(1-\alpha)$ où t est la réalisation de la variable aléatoire T.