Modèles Statistiques Linéaires, Maîtrise MASS, année 2000/2001.

Correction Examen Décembre 2000

Exercice 1

1) On a $h(p) = p_1p_2$ et $h(\hat{p}) = \hat{p}_1\hat{p}_2 = \frac{y_1y_2}{n^2}$ On sait que

$$\sqrt{n}\begin{pmatrix}\hat{p}_1 - p_1\\ \hat{p}_2 - p_2\end{pmatrix} \to \mathcal{N}(0, h'(p)\Gamma(p)^t h'(p)) \text{ avec } \Gamma(p) = \begin{pmatrix}p_1(1-p_1) & -p_1p_2\\ -p_1p_2 & p_2(1-p_2)\end{pmatrix}$$

et $h'(p)\Gamma(p)^t h'(p) = p_1p_2(p_1 + p_2 - 4p_1p_2)$

2) L'intervalle de confiance est donc

$$\left[\frac{y_1y_2}{n^2} \pm l_\alpha\sqrt{\frac{\hat{p}_1\hat{p}_2(\hat{p}_1+\hat{p}_2-4\hat{p}_1\hat{p}_2)}{n}}\right]$$

avec $\hat{p}_1 = \frac{y_1}{n}$ et $\hat{p}_2 = \frac{y_2}{n}$.
l_α le quantile d'ordre $1-\alpha$ d'une $\mathcal{N}(0,1)$.

3) et 4) voir td 7
régle de décision: $T > \chi^2_{1-\alpha} \iff$ rejet de H_0,
avec $T = 2\sum_{i=1}^I \sum_{j=1}^3 y_{ij}\log\frac{y_{ij}n_+}{y_{+j}n_i}$ où $n_+ = \sum_{i=1}^I n_i$ et $y_{+j} = \sum_{i=1}^I y_{ij}$ et où $\chi^2_{1-\alpha}$ est le $(1-\alpha)$ quantile de la distribution de χ^2 de degré de liberté $2I-2$.

Exercice 2

1) On a
$Y_{1i} = a + u_{1i}$
$Y_{2i} = bY_{1i} + u_{2i} = ba + bu_{1i} + u_{2i}$.
$u \sim \mathcal{N}\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{pmatrix}\right]$
Donc $\begin{pmatrix}Y_{i1}\\Y_{i2}\end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix}a\\ba\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & b\sigma_1^2\\ b^2\sigma_1^2 & b\sigma_1^2+\sigma_2^2\end{pmatrix}\right]$

$\Theta = \mathbb{R}^2 \times (\mathbb{R}_*^+)^2$, $\mathbb{R}_*^+ = \{x : x > 0\}$.

2) La loi de u_i a pour densité

$$\frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{u_{1i}^2}{2\sigma_1^2} - \frac{u_{2i}^2}{2\sigma_2^2}\right)$$

La transformation: $\begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix} \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix}$ est linéaire de jacobien 1, la densité de Y_i est donc:

$$f(y_i;\theta) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{(y_{1i}-a)^2}{2\sigma_1^2} - \frac{(y_{2i}-by_{1i})^2}{2\sigma_2^2}\right].$$

La log-vraisemblance est

$$L_n(\theta) = -n\log 2\pi - \frac{n}{2}\log\sigma_1^2 - \frac{n}{2}\log\sigma_2^2 - \frac{1}{2\sigma_1^2}\sum_{i=1}^n (y_{1i}-a)^2 - \frac{1}{2\sigma_2^2}\sum_{i=1}^n (y_{2i}-by_{1i})^2$$

Donc on trouve les EMV $\hat{a}_n, \hat{b}_n, \hat{\sigma^2_{1n}}, \hat{\sigma^2_{2n}}$ (voir enoncé Examen).

3) La log densité est:
$\log f(y_i,\theta) = -\log 2\pi - \frac{1}{2}\log\sigma_1^2 - \frac{1}{2}\log\sigma_2^2 - \frac{(y_{1i}-a)^2}{2\sigma_1^2} - \frac{(y_{2i}-by_{1i})^2}{2\sigma_2^2}$.
Les dérivées secondes croisées en (a,b), (a,σ_2^2), (b,σ_1^2), (σ_1^2,σ_2^2) sont nulles. Les dérivées en (a,σ_1^2) et (b,σ_2^2) sont:

$\frac{\mathcal{D}^2 \log f}{\mathcal{D}a\mathcal{D}\sigma_1^2} = -\frac{y_{1i}-a}{\sigma_1^4} = -\frac{u_{1i}}{\sigma_1^4}$.

$\frac{\mathcal{D}^2 \log f}{\mathcal{D}b\mathcal{D}\sigma_2^2} = -\frac{y_{1i}(y_{2i}-by_{1i})}{\sigma_2^4} = -\frac{(a+u_{1i})u_{2i}}{\sigma_2^4}$

d'espérance nulle. La matrice d'information pour une observation est donc diagonale. Les termes diagonaux sont
$-E\frac{\mathcal{D}^2 \log f}{\mathcal{D}a^2} = \frac{1}{\sigma_1^2}$
$-E\frac{\mathcal{D}^2 \log f}{\mathcal{D}b^2} = \frac{1}{\sigma_2^2}EY_{1i}^2 = \frac{a^2+\sigma_1^2}{\sigma_2^2}$
$-E\frac{\mathcal{D}^2 \log f}{\mathcal{D}\sigma_1^4} = -\frac{1}{2\sigma_1^4} + \frac{1}{\sigma_1^6}E(Y_{1i}-a)^2 = \frac{1}{2\sigma_1^4}$
$-E\frac{\mathcal{D}^2 \log f}{\mathcal{D}\sigma_2^4} = -\frac{1}{2\sigma_2^6} + \frac{1}{\sigma_2^6}E(Y_{2i}-bY_{1i})^2 = \frac{1}{2\sigma_2^4}$

Donc la matrice de variance-covariance asymptotique de $\sqrt{n}(\hat{\theta}_n - \theta)$ est la matrice diagonale dont les termes diagonaux sont σ_1^2, $\frac{\sigma_2^2}{a+\sigma_1^2}$, $2\sigma_1^4$, $2\sigma_2^4$.

4) La statistique de Wald pour $H_0 : b = 1$ est:

$$\begin{aligned} \xi_n^W &= n(\hat{b}_n - 1)^2 \frac{\hat{a}_n^2 + \hat{\sigma^2_{1n}}}{\hat{\sigma^2_{2n}}} \\ &= \frac{n(\hat{b}_n-1)^2 \sum_{i=1}^n Y_{1i}^2}{\sum_{i=1}^n (Y_{2i} - \hat{b}_n Y_{1i})^2} \end{aligned}$$

La régle de décision est: rejet de $H_0 \iff \xi_n^W \geq \chi^2_{1-\alpha}(1)$.

5) On a $\hat{b}_n = b + \frac{\sum_{i=1}^n Y_{1i} u_{2i}}{\sum_{i=1}^n Y_{1i}^2}$.

Sous H_0,

$$\hat{b}_n = 1 + \frac{\sum_{i=1}^n Y_{1i} u_{2i}}{\sum_{i=1}^n Y_{1i}^2}.$$

Conditionnellement à $Y^1 = (Y_{11}, .., Y_{1n})$, la loi de $\hat{b}_n$ est:

$$\mathcal{N}\left(1, \frac{\sigma_2^2}{\sum_{i=1}^n Y_{1i}^2}\right)$$

puisque les (u_{2i}) et les (Y_{1i}) sont indépendants. Toujours conditionnellement à Y^1, $\frac{1}{\sigma_2^2} \sum_{i=1}^n (Y_{2i} - \hat{b}_n Y_{1i})^2$ suit $\chi^2(n-1)$ et est indépendant de $\hat{b}_n$ (résultat du modèle linéaire gaussien), donc:

$$\begin{aligned}
\frac{n-1}{n} \xi_n^W &= \left[\frac{(\hat{b}_n - 1) \left(\sum_{i=1}^n Y_i^2\right)^{1/2}}{\sigma_2} . \frac{1}{\left(\frac{1}{(n-1)\sigma_2^2} \sum_{i=1}^n (Y_{2i} - \hat{b}_n Y_{1i})^2\right)^{1/2}} \right]^2 \\
&= \left(\frac{U}{\sqrt{\frac{V}{n-1}}} \right)^2 = \frac{U}{\frac{V}{n-1}}
\end{aligned}$$

$U \sim \mathcal{N}(0,1)$, $V \sim \chi^2(n-1)$ indépendamment. Donc conditionnellement ξ_n^W suit la loi fixe $F(1, n-1)$ (ou la loi du carré d'une variable de Student $t(n-1)$) et donc cette loi est aussi la loi non conditionnelle de ξ_n^W.

Test exact de région critique:
$\frac{n-1}{n} \xi_n \geq F_{1-\alpha}(1, n-1)$

ou

$|\hat{b}_n - 1| \left[\frac{(n-1)(\sum_{i=1}^n Y_{1i}^2)}{\sum_{i=1}^n (Y_{2i} - \hat{b}_n Y_{1i})^2} \right]^{1/2} \geq t_{1-\alpha/2}(n-1)$.