Université Paris Dauphine Année 2005-2006
UFR Mathématiques de la Décision
M1 MIDO - Modèles linéaires et ses généralisations

Correction du partiel

Exercice 1 Partie A.
1) Sous H_0, $\mu = \pm 1$ donc on sait que $\hat{\mu}_0 = \pm 1$. On écrit la log-vraisemblance :

$$\begin{aligned} L_n(\mu, \sigma^2) &= -\sum_{i=1}^{n} \frac{(y_i - \mu)^2}{2\sigma^2} - \frac{n}{2} \log(2\pi\sigma^2) \\ &= -n \frac{S_n^2 + (\bar{Y}_n - \mu)^2}{2\sigma^2} - \frac{n}{2} \log(2\pi\sigma^2) \end{aligned}$$

On veut maximiser en (μ, σ^2) sur l'espace $\{\mu^2 = 1, \sigma^2 > 0\}$. ON maximise à μ fixé :

$$\begin{aligned} \frac{\partial L_n(\mu, \sigma^2)}{\partial \sigma^2} &= n \frac{S_n^2 + (\bar{Y}_n - \mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} \\ &= 0 \end{aligned}$$

si et seulement si :

$$\hat{\sigma}^2(\mu) = S_n^2 + (\bar{Y}_n - \mu)$$

Il faut maintenant choisir μ qui maximise

$$L_n(\mu, \hat{\sigma}^2(\mu)) = -\frac{n}{2} - \frac{n}{2} \log(2\pi[S_n^2 + (\bar{Y}_n - \mu)^2])$$

Cette fonction est décriossante en $(\bar{Y}_n - \mu)^2$ et $\mu = \pm 1$ donc $\hat{\mu}_0 = 1$ si $(\bar{Y}_n - 1)^2 < (\bar{Y}_n + 1)^2$, c'est à dire ssi $\bar{Y}_n > 0$.
b)

$$\xi_n^R = 2(l_n(\hat{\sigma}_0^2, \hat{\mu}_0) - l_n(\hat{\sigma}^2, \hat{\mu}))$$

Il est maintenant bien connu que $\hat{\mu} = \bar{Y}_n$ et $\hat{\sigma} = S_n$, on remarque que $S_n^2 + (\bar{Y}_n - \hat{\mu})^2 = \hat{\sigma}^2$, ainsi

$$\xi_n^R = n\left[\log(S_n^2 + (\bar{Y}_n - \hat{\mu}_0)^2) - \log S_n^2\right]$$

Le modèle gaussien est un modèle régulier (car exponentiel, sur un ouvert le support étant indépendant des paramètres) le test sécrit sous la forme :

$H_0 : g(\mu, \sigma^2) = 0$ avec $g(\mu, \sigma^2) = \mu^2 - 1$, g est de classe C^1 $\nabla g = (2\mu, 0)$ est de rang 1 sous H_0, donc on sait que sous H_0 en loi,

$$\lim_{n\to\infty} \xi_n^R = \chi^2(1)$$

ainsi le test du rapport de vraisemblance de niveau asymptotique α s'écrit par sa région de rejet définie par :

$$\{\xi_n^R \geq F^{-1}_{\chi^2(1)}(1-\alpha)\}.$$

2) Test UPP pour $H_0 : \mu < 0$ contre $H_1 : \mu \geq 0$
- On montre que le modèle est à rapport de vraisemblance monotone. Soit $\mu_1 < \mu_2$

$$\begin{aligned} \frac{f(x|\mu_1)}{f(x|\mu_2)} &= \exp\{-(x-\mu_1)^2/2 + (x-\mu_2)^2/2\} \\ &= \exp\{x(\mu_1 - \mu_2) - \mu_1^2/2 + \mu_2^2/2\} \end{aligned}$$

c'est une fonction décroissante en x. Pour trouver le sens des inégalités, il faut prendre $\mu_1 < 0$ et $\mu_2 > 0$, on sait alors qu'on accepte H_0 si le rapport ci-dessus est grand, ce qui revient à x petit, on obtient ainsi la région de rejet définie par

$$\{\sum_i Y_i > k_\alpha\},$$

où k_α est défini par

$$P_{\mu=0}[\sum_i Y_i > k_\alpha] = \alpha.$$

Sous H_0

$$\sum_i Y_i \sim \mathcal{N}(0, n\sigma^2)$$

, donc

$$\sum_i Y_i \geq k_\alpha \quad \Leftrightarrow \quad \frac{\sqrt{n}\bar{Y}_n}{\sigma} \geq \frac{\sqrt{n}k_\alpha}{\sigma} = \Phi^{-1}(1-\alpha),$$

où $\Phi(x)$ est la fonction de répartition d'une loi normale centrée réduite. Ainsi $k_\alpha = \sigma\Phi^{-1}(1-\alpha)/\sqrt{n}$.

Partie B

1) Vraisemblance :

$$V_n(a,b) = \exp\{-\sum_{i=1}^n (Y_i - ax_i - b)^2/(2\sigma^2)\} - \frac{n}{2}\log(2\pi\sigma^2)$$

Le maximum de vraisemblance annule le score :

$$\frac{\partial L_n(a,b)}{\partial a} = \frac{\sum_i x_i(Y_i - ax_i - b)}{\sigma^2} = 0$$

et

$$\frac{\partial L_n(a,b)}{\partial b} = \frac{\sum_i (Y_i - ax_i - b)}{\sigma^2} = 0$$

La deuxième équation est équivalent à

$$\bar{Y}_n - a\bar{x}_n = b$$

ainsi en recentrant x_i dans l'équation 1 on obtient :

$$\begin{aligned} \sum_i (x_i - \bar{x}_n)(Y_i - ax_i - b) + \bar{x}_n \sum_i (Y_i - ax_i - b) &= \sum_i (x_i - \bar{x}_n)(Y_i - \bar{Y}_n - a(x_i - \bar{x}_n)) \\ &= 0 \end{aligned}$$

ssi

$$\hat{a} = \frac{\sum_i (Y_i - \bar{Y}_n)(x_i - \bar{x}_n)}{\sum_i (x_i - \bar{x}_n)^2} \quad \hat{b} = \bar{Y}_n - \hat{a}\bar{x}_n$$

2)

$$\begin{aligned} E[\hat{a}] &= E\left[\frac{\sum_i (Y_i - \bar{Y}_n)(x_i - \bar{x}_n)}{\sum_i (x_i - \bar{x}_n)^2}\right] \\ &= \frac{\sum_i [E(Y_i - \bar{Y}_n)](x_i - \bar{x}_n)}{\sum_i (x_i - \bar{x}_n)^2} \\ &= \frac{\sum_i (ax_i + b - a\bar{x}_n - b)(x_i - \bar{x}_n)}{\sum_i (x_i - \bar{x}_n)^2} \\ &= \frac{a\sum_i (x_i - \bar{x}_n)^2}{\sum_i (x_i - \bar{x}_n)^2} \\ &= a \end{aligned}$$

car les x_i sont des coonstantes et $E(y_i) = ax_i + b$, $E(\bar{Y}_n) = n^{-1}(\sum_i E(y_i)) = a\bar{x}_n + b$. Comme

$$\sum_i (x_i - \bar{x}_n) = 0,$$

$$\sum_i (Y_i - \bar{Y}_n)(x_i - \bar{x}_n) = \sum_i Y_i(x_i - \bar{x}_n)$$

les Y_i sont des variables indépendantes donc la variance de la somme est la somme des variances et

$$\begin{aligned} var(\hat{a}) &= \frac{\sum_{i=1}^n var(Y_i)(x_i - \bar{x}_n)^2}{\left(\sum_i (x_i - \bar{x}_n)^2\right)^2} \\ &= \frac{\sigma^2}{\sum_i (x_i - \bar{x}_n)^2} \end{aligned}$$

$\hat{a}$ est une combinaison linéaire de loi normales indépendantes, donc $\hat{a}$ est une variable gaussienne d'espérance a et de variance $\sigma^2/(\sum_i (x_i - \bar{x}_n)^2$.

3) D'après la question précédente

$$\frac{\hat{a}\left(\sum_i (x_i - \bar{x}_n)^2\right)^{1/2}}{\sigma} \sim \mathcal{N}(0,1)$$

donc ξ_n est distribué comme le carré d'une normale centrée réduite, cette statistique suit donc un $\chi^2(1)$.
$\hat{a}$ est un estimateur de a, s'il est grand cela veut dire que a est différend de zéro, la région de rejet du test est donc définie par :

$$\{\xi_n \geq F^{-1}_{\chi^2(1)}(1-\alpha)\}$$

heuristiquement $\hat{a}$ va converger vers a, $\sum_i (x_i - \bar{x}_n)^2$ ressemble à n fois un estimateur de variance donc si $a \neq 0$ $a \sum_i (x_i - \bar{x}_n) \simeq ns^2 a \to \infty$
4) Application numérique

$$\hat{a} = \frac{65^2}{20} = 211,5$$

si on prend $\alpha = 0.05$ par exemple $P[\chi^2(1) > q] = P[-\sqrt{q} < \mathcal{N}(0,1) < \sqrt{q}] = 0.05$ ssi $P[\mathcal{N}(0,1) < q] = 0.025$, donc $\sqrt{q} \leq 2$ et $q \leq 4 \ll 211,5$ on refuse H_0.

Exercice 2
1) $S_n = \sum_i X_i$, les $X_i \sim \mathcal{B}(p)$ et sont indépendants donc $S_n \sim \mathcal{B}(n,p)$.
2) test UPP : hypothèse ponctuelles, mais variables discrètes, le test UPP est donné par

$$\phi(X_1, ..., X_n) = \begin{cases} 1 & \text{si} \quad V_n(0.03)/V_n(0.1) < k_\alpha \\ \gamma & \text{si} \quad V_n(0.03)/V_n(0.1) = k_\alpha \\ 0 & \text{si} \quad V_n(0.03)/V_n(0.1) > k_\alpha \end{cases}$$

où k_α est donné par :

$$P_0[V_n(0.03)/V_n(0.1) < k_\alpha] + \gamma P_0[V_n(0.03)/V_n(0.1) k_\alpha] = \alpha$$

il faut donc étudier la loi du rapport de vraisemblances

$$V_n(p) = \prod_i p_i^X (1-p)^{1-X_i}$$

donc

$$\frac{V_n(0.03)}{V_n(0.1)} = \left(\frac{0.03*0.9}{0.97*0.1}\right)^{S_n} \left(\frac{0.97}{0.9}\right)^n$$

ainsi

$$\frac{V_n(0.03)}{V_n(0.1)} > k_\alpha, \quad \Leftrightarrow S_n \log\left(\frac{0.03*0.9}{0.97*0.1}\right) \geq k'_\alpha$$

comme $0.03*0.9 < 0.97*0.1$ l'inégalité précédente est équivalent à $S_n > c_\alpha$; où c_α est défini par

$$P_\alpha[S_n > c_\alpha] + \gamma P_\alpha[S_n = c_\alpha]\alpha$$

avec $S_n \sim \mathcal{B}(n, 0.03)$.
3) On remarque que le modèle binomial est à rapport de vraisemblance monotone ainsi le test précédent est aussi le test UPP pour le problème

$$H_0 p = 0.03, \quad H_1 p > 0.03$$

4) Score

$$L_n(p) = S_n \log p + (n - S_n)\log(1-p)$$

donc

$$S_n(p) = n^{-1/2}\left(S_n(\frac{1}{p} + \frac{1}{(1-p)}) + \frac{n}{1-p}\right)$$

le maximum de vraisemblance contraint vaut $\hat{p}^0 = 0.03$, Pour le calcul de l'info de Fisher , il faut considérer $n = 1$, l'info de Fisher est égale à la variance du score

$$I(p) = V[X]\left[\frac{1}{p} + \frac{1}{(1-p)}\right]^2 = \frac{p(1-p)}{p^2(1-p)^2}$$

On obtient finalement

$$\xi_n^S = n^{-1}\left(\frac{S_n}{p_0(1-p_0)} + \frac{n}{(1-p)}\right)^2 p_0(1-p_0) = \frac{(S_n - np_0)^2}{np_0(1-p_0)}$$

il suffit ensuite de remplacer p_0 par 0.03.