En pratique, il est d'usage d'utiliser la méthode de Newton-Raphson pour maximiser numériquement $\ell(\underset{\sim}{\beta})$.

Pour ce faire, on pose $\hat{\underset{\sim}{\beta}}^{(k)}$, le vecteur contenant les valeurs éstimés pour $\underset{\sim}{\beta}$ après la k^e itération de l'algorithme.

En supposant que l'algorithme ait convergé après l'itération $(i+1)$, alors on a que:

$$\frac{\partial}{\partial \underset{\sim}{\beta}} \ell\left(\hat{\underset{\sim}{\beta}}^{(i+1)}\right) = 0$$

ou encore que:

△

$$S\left(\hat{\underset{\sim}{\beta}}^{(i+1)}\right) = 0$$

; où $S(\cdot)$ est appelé "vecteur score" et est donné par:

$$S(\underset{\sim}{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_0} \ell(\underset{\sim}{\beta}) \\ \frac{\partial}{\partial \beta_1} \ell(\underset{\sim}{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_p} \ell(\underset{\sim}{\beta}) \end{bmatrix}_{(p+1)\times 1}$$

Pour obtenir un algorithme permettant d'obtenir un estimateur récursivement (à partir de $\hat{\underset{\sim}{\beta}}^{(i)}$), il faut développer $S(\hat{\underset{\sim}{\beta}}^{(i+1)})$ autour de $\hat{\underset{\sim}{\beta}}^{(i)}$ à l'aide d'une série de Taylor:

$$(*) \quad S(\hat{\underset{\sim}{\beta}}^{(i+1)}) = S(\hat{\underset{\sim}{\beta}}^{(i)}) + \left[\frac{\partial}{\partial \underset{\sim}{\beta}}\left(S(\hat{\underset{\sim}{\beta}}^{(i)})\right)\right]\left(\hat{\underset{\sim}{\beta}}^{(i+1)} - \hat{\underset{\sim}{\beta}}^{(i)}\right)$$

$$(*) + (\Delta) \Rightarrow S(\hat{\underset{\sim}{\beta}}^{(i)}) + \frac{\partial}{\partial \beta} S(\hat{\underset{\sim}{\beta}}^{(i)}) \left(\hat{\underset{\sim}{\beta}}^{(i+1)} - \hat{\underset{\sim}{\beta}}^{(i)}\right) = 0$$

ou encore...

$$S(\hat{\underset{\sim}{\beta}}^{(i)}) + \left(-I(\hat{\underset{\sim}{\beta}}^{(i)})\right)\left(\hat{\underset{\sim}{\beta}}^{(i+1)} - \hat{\underset{\sim}{\beta}}^{(i)}\right) = 0$$

; où

$I(\underset{\sim}{\beta})$ est appelé matrice d'information de Fisher et est donnée par:

$$I(\underset{\sim}{\beta}) = \begin{bmatrix} \frac{\partial^2}{\partial \beta_0^2} \ell(\underset{\sim}{\beta}) & \frac{\partial^2}{\partial \beta_0 \partial \beta_1} \ell(\underset{\sim}{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_0 \partial \beta_p} \ell(\underset{\sim}{\beta}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial \beta_0 \partial \beta_p} \ell(\underset{\sim}{\beta}) & \frac{\partial^2}{\partial \beta_1 \partial \beta_p} \ell(\underset{\sim}{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_p^2} \ell(\underset{\sim}{\beta}) \end{bmatrix}_{(p+1)\times(p+1)}$$

Ainsi:

$$S(\hat{\beta}^{(i)}) = I(\hat{\beta}^{(i)})\left(\hat{\beta}^{(i+1)} - \hat{\beta}^{(i)}\right)$$

$\Rightarrow$

$$I^{-1}(\hat{\beta}^{(i)})\,S(\hat{\beta}^{(i)}) = \hat{\beta}^{(i+1)} - \hat{\beta}^{(i)}$$

$\Rightarrow$

$$\boxed{\underbrace{\hat{\beta}^{(i+1)}}_{\text{vecteur de param. mis à jour}} = \underbrace{\hat{\beta}^{(i)}}_{\text{Ancien vecteur de param.}} + \underbrace{I^{-1}(\hat{\beta}^{(i)})\,S(\hat{\beta}^{(i)})}_{\text{Produit matriciel entre l'inverse de la matrice de Fisher et le vecteur score}}}$$

4.4.3) Validation "global" du modèle avec la "DÉVIANCE" :

En réécrivant la fonction de vraisemblance $\ell(\beta)$ comme suit :

$$\ell(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \ln\left(f_Y(y_i) \right)$$

$$= \sum_{i=1}^{n} \ln\left(f_Y(y_i\,;\mu_i) \right) \quad ; \text{ où } \mu_i = g^{-1}(X_i\underset{\sim}{\beta})$$

$$= \ell(\underset{\sim}{y}, \underset{\sim}{\mu})$$

... on peut définir la statistique de déviance $D(\underset{\sim}{\beta})$ comme :

$$D(\underset{\sim}{\beta}) = 2 \times \left(\underbrace{\ell(\underset{\sim}{y}, \underset{\sim}{y})}_{\substack{\text{"Logvrais"}\\ \text{sous un} \\ \text{modèle} \\ \text{parfait}}} - \underbrace{\ell(\underset{\sim}{y}, \underset{\sim}{\mu})}_{\substack{\text{"Logvrais"} \\ \text{du modèle} \\ \text{obtenu}}} \right)$$

On remarque que dans le cas de la famille exponentielle, on aura :

$$\boxed{D(\underset{\sim}{\beta}) = 2 \times \sum_{i=1}^{n} \left\{ y_i \times \left(\theta(y_i) - \theta(\mu_i) \right) - b\left(\theta(y_i)\right) + b\left(\theta(\mu_i)\right) \right\}}$$

Ainsi:

(1) Loi Normale.

$$D(\beta) = 2\times \sum_{i=1}^{n} \left\{ y_i\left(y_i - \mu_i\right) - \frac{y_i^2}{2} + \frac{\mu_i^2}{2} \right\}$$

$$= \sum_{i=1}^{n} \left\{ 2y_i^2 - 2\mu_i y_i - y_i^2 + \mu_i^2 \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i^2 - 2\mu_i y_i + \mu_i^2 \right\}$$

$$= \sum_{i=1}^{n} \left(y_i - \mu_i\right)^2$$

$$= \sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

$$= SSE \;!$$

Conclusion:

Dans le cas de la loi normale, la déviance est égale à SSE

De fait, $D(\beta)$ constitue une généralisation de SSE qui sera valide avec toutes les sous-cas de la famille exponentielle!

(2) Loi de Poisson:

$$D(\beta) = 2 \times \sum_{i=1}^{n} \left\{ y_i \times \ln\left(\frac{y_i}{\mu_i}\right) - (y_i - \mu_i) \right\}$$

$\neq$ SSE!

(3) Binomiale (m_i, μ_i):

$$D(\beta) = 2 \times \sum_{i=1}^{n} \left\{ y_i \times \ln\left(\frac{y_i}{\mu_i}\right) + (m_i - y_i) \times \ln\left(\frac{m_i - y_i}{m_i - \mu_i}\right) \right\}$$

$\neq$ SSE!

(4) Gamma:

$$D(\beta) = 2 \times \sum_{i=1}^{n} \left\{ -\ln\left(\frac{y_i}{\mu_i}\right) + \frac{y_i - \mu_i}{\mu_i} \right\}$$

$\neq$ SSE!

(5) Inverse Gaussienne:

$$D(\beta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{\mu_i^2 / y_i}$$

$\neq$ SSE

Conclusion: On retombe sur SSE uniquement par la loi normale. Dans les autres, cas, on a une mesure plus générale!

4.4.4) Validation "local" du modèle avec des tests d'hypothèses et intervalles de confiances :

I) Tests d'hypothèse très général :

On introduit ici une version généralisée des tests de Fisher partiels introduits au chapitre 3.

Hypothèses à considérer : (très similaire à celle de 3.2.3 !)

H_0 : Un modèle réduit (noté M_0) qui est un sous-modèle de M_1 (modèle plus complet !) est statistiquement acceptable

H_1 : On doit utiliser le modèle plus complet (noté M_1).

Pour tester H_0 contre H_1, on utilise la statistique suivante :

$$\boxed{\begin{aligned}\chi^2_{OBS} &= D(\underset{\sim}{\beta}_{H_0}) - D(\underset{\sim}{\beta}_{H_1}) \\ &= 2\times\left(\ell(\underset{\sim}{\beta}_{H_1}) - \ell(\underset{\sim}{\beta}_{H_0})\right)\end{aligned}}$$

... et on rejette H_0 au niveau de confiance $100\times(1-\alpha)\%$ si

$$\boxed{\chi^2_{OBS} \geq \chi^2_{\alpha}\left(\begin{array}{c}\#\text{paramètres} \\ \text{dans } M_1\end{array} - \begin{array}{c}\#\text{paramètres} \\ \text{dans } M_0\end{array}\right)}$$

II) Intervalles de confiances:

Selon la théorie du maximum de vraisemblance, on a que la loi asymphthotique (lorsque $n \to \infty$) de $\hat{\beta}$ est une loi normale multi-dimensionnelle:

$$\underset{(p+1)\times 1}{\hat{\beta}} \xrightarrow[n \to \infty]{} N_{p+1}\left(\beta \, ; \, Var(\hat{\beta})\right)$$

; avec

$$Var(\hat{\beta}) = I^{-1}(\hat{\beta})$$: Borne inférieure de Cramer-Rao !

Ainsi, un intervalle de confiance au niveau $100 \times (1-\alpha)\%$ pour β_i est donné par:

$$\boxed{\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}} \times \sqrt{\left[Var(\hat{\beta})\right]_{i+1,\, i+1}}}$$

4.5) Modèle de régression normale: (un retour!)

Sachant maintenant les résultats plus généraux des GLM, il est intéressant de retrouver les concepts de la régression multiple!

Dans ce cas, on a que:

$$(Y_i \mid \underset{\sim}{X_i}) \sim N(\mu_i, \sigma^2);$$

avec

$$\mu_i = \underset{\sim}{X_i}\beta \Rightarrow g(x) = x \quad \left(\text{fonction de lien identité!}\right)$$

Et donc:

$$\ell(\underset{\sim}{\beta}) = \sum_{i=1}^{n} \ln\left(f_{Y_i}(y_i)\right)$$

$$= \underbrace{-\tfrac{1}{2}}_{\text{Signe négatif!}} \sum_{i=1}^{n} \left\{ \underbrace{\ln(2\pi\sigma^2)}_{\text{cte p/r } \beta} + \frac{1}{\sigma^2}(y_i - X_i\beta)^2 \right\}$$

... **Remarque**: Ici, maximiser $\ell(\underset{\sim}{\beta})$ revient à minimiser $\sum_{i=1}^{n}(y_i - X_i\beta)^2 = SSE$!

Par conséquent:

$$S(\beta) = \frac{\partial}{\partial \beta} \ell(\beta) = ?$$

On cherche le j^e élément de ce vecteur $S(\beta)$:

$$[S(\beta)]_j = \frac{\partial}{\partial \beta_i} \ell(\beta) = + \frac{2}{2\sigma^2} \sum_{i=1}^{n} (y_i - X_i \beta) X_{i,j}$$

$$= + \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} X_{ij} Y_i - \left(\sum_{i=1} X_i \times X_{ij} \right) \beta \right)$$

$$= \frac{1}{\sigma^2} \left(X_j' \cancel{X} - X_j' X \beta \right)$$

Ainsi, on déduit que:

$$S(\beta) = \frac{1}{\sigma^2} \left(X'Y - X'X\beta \right)$$

↑↑↑

Ici
19-11-2012