

MODÈLE « LOGIT » (J2)

(16 / 06 / 2020, © Monfort, Dicostat2005, 2005-2020)

Un **modèle logit**, ou **modèle Logit**, est un modèle dont la **variable endogène** est une **variable qualitative** particulière (**variable dichotomique**) (cf **variable binaire**) et dont les **variables exogènes** sont, en général, des **variables quantitatives** (ie chacune à valeurs dans $\mathbf{R}$).

(i) Soit $(\Omega, \mathcal{F}, \mathcal{P})$ un **modèle statistique** et $(\xi, \eta) : \Omega \mapsto \mathbf{R}^K \times N_1$ un **couple aléatoire** constitué d'une liste de variables exogènes (réelles) ξ_k (avec $k \in N_K^*$) et d'une variable endogène qui est une **variable indicatrice** (avec $N_1 = \{0, 1\}$). On note F la **fr** de la **loi logistique** (centrée réduite) $\mathcal{L}(0, 1)$:

$$(1) \quad F(z) = (1 + e^{-z})^{-1}, \quad \forall z \in \mathbf{R}.$$

On suppose que la loi conditionnelle P^η de η relativement à ξ est de la forme :

$$(2) \quad P^\eta(y) = [F(\xi' b)]^y \cdot [1 - F(\xi' b)]^{1-y}, \quad \forall y \in N_1,$$

dans laquelle $b \in \mathbf{R}^K$ est un **paramètre d'intérêt**.

On appelle alors **modèle « Logit »** le modèle statistique défini par donnée de N **copies** iid y_n de η (cf **suite iid**) et d'une (N, K) -**matrice d'observation** X , dont la ligne X_n représente, $\forall n \in N_N^*$, la N -ième observation de ξ correspondant à la valeur y_n de l'endogène.

(ii) La **vraisemblance** de ce modèle s'écrit :

$$(3) \quad L(y, b) = \prod_{n=1}^N P^\eta(y_n) = \prod_{n=1}^N \{F(X_n b)^{y(n)} \cdot [1 - F(X_n b)]^{1-y(n)}\},$$

où les $y(n)$ désignent les y_n .

Elle permet l'**estimation** de b par la **méthode du maximum de vraisemblance**, l'**équation de vraisemblance** associée s'écrivant :

$$(4) \quad X' y = X' \Psi(b),$$

avec (vecteur en ligne) :

$$(5) \quad \Psi(b) = (F(X_1 b) \parallel F(X_N b))',$$

où $\parallel$ désigne des sauts de lignes.

(iii) Sous des conditions générales, on montre que :

(a) **convergence presque sûre** : l'équation (4) admet (presque sûrement) une solution $\hat{b}_N$ qui converge presque sûrement vers b lorsque $N \rightarrow +\infty$;

(b) **normalité asymptotique** : la solution précédente est asymptotiquement gaussienne, ie :

$$(6) \quad \{\sum_{n=1}^N X_n' X_n f(X_n \hat{b}_N)\}^{1/2} (\hat{b}_N - b) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_K(0, I_K)$$

(loi normale multidimensionnelle centrée), où $f = dF / dz$ désigne la **densité de probabilité** associée à F .