

GEV

Fabian Bastin

Avril 2011

La propriété IIA du logit est souvent indésirable.

Les modèles de valeur extrême généralisée visent à relâcher l'IIA en proposant d'autres schémas de substitution.

Comme le nom l'indique, il s'agit d'une généralisation du logit, qui peut être vu comme un cas particulier des modèles GEV.

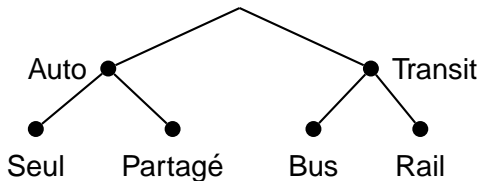
L'exemple le plus connu est le nested logit.

Avantage des GEV : on continue à avoir une forme analytique pour les probabilités de choix.

Nested logit

Le nested logit, que l'on pourrait traduire littéralement par logit niché, est approprié quand l'ensemble des alternatives peut être partitionné en sous-ensembles appelés nids (nests).

L'IIA tient à l'intérieur de chaque nid, mais pas pour des alternatives dans des nids différents.



Daly et Zachary (1978), McFadden (1978), Williams (1977) ont montré que le nested logit est consistant avec le principe de maximisation d'utilité.

Les K alternatives sont partitionnées en M nids, B_1, \dots, B_M , tels que

$$U_{m=1}^M \{j \in B_m\} = \mathcal{K}, \quad B_m \cap B_n = \emptyset \quad \forall m, n \quad m \neq n.$$

$\epsilon_i = \{\epsilon_{i,1}, \epsilon_{i,2}, \dots, \epsilon_{i,K}\}$ a comme fonction de répartition

$$e^{-\sum_{m=1}^M \left(\sum_{k \in B_m} e^{-\epsilon_{i,k}/\lambda_m} \right)^{\lambda_m}}$$

Si k_1 et k_2 sont dans le nest B_m , ϵ_{i,k_1} est corrélé avec ϵ_{i,k_2} .

Pour deux alternatives dans des nids différents,

$$\text{cov}(\epsilon_{i,k_1}, \epsilon_{i,k_2}) = 0, \quad k_1 \in B_m, \quad k_2 \in B_n \quad m \neq n.$$

λ_m est une mesure du degré d'indépendance dans l'utilité non-observée entre les alternatives du nid m .

Plus λ_m est grand, moins il y a de dépendance. Plus correctement, moins il y a de corrélation.

$1 - \lambda_m$ est une mesure de corrélation ; avec $\lambda_m = 1$, on retrouve le logit standard.

Probabilité pour l'individu i de choisir l'alternative k dans le nid B_m .

$$P_{i,k} = \frac{e^{V_{i,k}/\lambda_m} \left(\sum_{j \in B_m} e^{V_{i,j}/\lambda_m} \right)^{\lambda_m - 1}}{\sum_{\ell=1}^M \left(\sum_{j \in B_\ell} e^{V_{i,j}/\lambda_\ell} \right)^{\lambda_\ell}}$$

Vérifions que l'IIA tient à l'intérieur de chaque sous-ensemble d'alternatives, mais pas entre nids.

Nids différents : $k_1 \in B_{m_1}$, $k_2 \in B_{m_2}$, $m_1 \neq m_2$:

$$\frac{P_{i,k_1}}{P_{i,k_2}} = \frac{e^{V_{i,k_1}/\lambda_{m_1}} \left(\sum_{j \in B_{m_1}} e^{V_{i,j}/\lambda_{m_1}} \right)^{\lambda_{m_1} - 1}}{e^{V_{i,k_2}/\lambda_{m_2}} \left(\sum_{j \in B_{m_2}} e^{V_{i,j}/\lambda_{m_2}} \right)^{\lambda_{m_2} - 1}}$$

Même nid B_m :

$$\frac{P_{i,k_1}}{P_{i,k_2}} = \frac{e^{V_{i,k_1}/\lambda_m}}{e^{V_{i,k_2}/\lambda_m}}.$$

L'IIA devient IIN. Le rapport des probabilités ne dépend pas des attributs des alternatives dans les nids autres que ceux contenant k_1 et k_2 .

Ceci peut être utilisé comme un test d'IIA : on teste la contrainte $\lambda_m = 1$, et on utilise le test de rapport de vraisemblance.

Consistance avec la maximisation d'utilité

Si $0 < \lambda_m < 1$, $\forall m$, le modèle est consistant avec le principe de maximisation d'utilité.

Si un certain $\lambda_m > 1$, le modèle est consistant avec le comportement de maximisation d'utilité pour un certain domaine des variables explicatives, mais pas toutes les valeurs.

Si un certain $\lambda_m < 0$, le modèle est inconsistant avec le principe de maximisation d'utilité.

Dans la notation utilisée, chaque λ_m est un paramètre fixé, impliquant que les preneurs de décision ont les mêmes corrélations entre les facteurs non-observés.

Pour tenir compte de variations entre individus, on peut spécifier les λ_k comme des fonctions paramétriques des attributs démographiques, par exemple en prenant

$$\lambda_k = e^{\alpha Z_i}.$$

Décomposition en somme de deux logits

L'utilité peut être décomposée en deux parties :

- 1 w constante pour toutes les alternatives à l'intérieur d'un nid ;
- 2 y une partie qui varie entre les alternatives au sein d'un nid.

Ainsi,

$$U_{i,k} = w_{i,k} + y_{i,k} + \epsilon_{i,k}.$$

La probabilité de choix peut être vu comme conditionnelle au choix d'un nid :

$$P_{i,k} = P_{i,k | B_m} P_{i,B_m}.$$

Décomposition en somme de deux logits (suite)

$$P_{i,B_m} = \frac{e^{w_{i,m} + \lambda_m l_{i,m}}}{\sum_{\ell=1}^K e^{w_{i,\ell} + \lambda_\ell l_{i,\ell}}},$$
$$P_{i,k | B_m} = \frac{e^{y_{i,k} / \lambda_m}}{\sum_{j \in B_m} e^{y_{i,j} / \lambda_m}},$$
$$l_{i,m} = \log \sum_{j \in B_m} e^{y_{i,j} / \lambda_m}.$$

(Voir Train, Section 4.2.5 pour le développement)

On peut voir $l_{i,m}$ comme le logsum, et λ_m le coefficient du logsum.

Les portions non-observées de l'utilité de toutes les alternatives sont conjointement distribuées selon une valeur extrême généralisée.

Posons $Y_k = e^{V_k}$, et prenons

$$G = G(Y_1, \dots, Y_K).$$

Notons

$$G_i = \frac{\partial G}{\partial Y_i}.$$

Si G satisfait un certain nombre de conditions,

$$P_k = \frac{Y_k G_k}{G}$$

est une probabilité, consistante avec le principe de maximisation de l'utilité.

Posons $Y = (Y_1, \dots, Y_K)$.

- 1 $G \geq 0, \forall Y \geq 0$;
- 2 $G(Y)$ est homogène de degré $\kappa > 0$:

$$G(\rho Y) = \rho^\kappa G(Y).$$

Habituellement, on pose $\kappa = 1$;

- 3 $\lim_{Y \rightarrow \infty} G(Y) = \infty$;
- 4 pour n'importe quelle séquence distincte (k_1, \dots, k_n) , $\partial^n G / \partial Y_{k_1} \dots \partial Y_{k_n}$ est ≥ 0 si n est impair et ≤ 0 si n est pair.

Voir Section 4.4 de Train.