

Modèles logit

Fabian Bastin

Avril 2011

Motivation

Un des modèles les plus populaires.

Avantages : forme analytique.

Les termes d'erreur sont supposés i.i.d., et suivre une loi de Gumbel (appelée aussi valeur extrême de type I). Avec une moyenne μ et un facteur d'échelle λ , nous avons

$$f(x) = e^{-\frac{x-\mu}{\lambda}} e^{-e^{-\frac{x-\mu}{\lambda}}},$$
$$F(x) = e^{-e^{-\frac{x-\mu}{\lambda}}}.$$

En posant, le facteur d'échelle à 1.0, la variance de cette distribution vaut $\pi^2/6$. La moyenne n'a pas d'importance (pourquoi ?).

En supposant ceci, nous avons implicitement normalisé l'échelle des utilités.

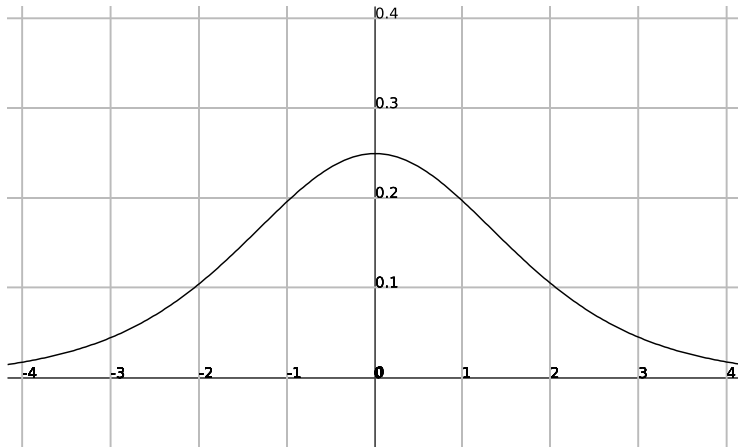
Pourquoi une Gumbel ?

Soit X et Y deux variables de Gumbel. Alors $Z = X - Y$ suit une distribution logistique de moyenne nulle, et de variance $\pi^2/3$. Sa fonction de répartition est

$$F_Z(x) = \frac{1}{1 + e^{-x}}.$$

Son graphe ressemble à une normale... or le praticien aime les normales pour représenter le bruit d'observation.

Distribution logistique



L'hypothèse d'indépendance n'est pas restrictive si on suppose que n'importe quel k , $V_{i,k}$ est spécifié suffisamment bien pour que $\epsilon_{i,k}$ ne fournisse pas d'information au chercheur sur l'erreur pour une autre alternative (bruit blanc).

Si ce n'est pas le cas, nous pouvons :

- 1 utiliser un modèle différent qui permet de la corrélation entre alternatives ;
- 2 respecifier la partie déterministe de l'utilité de manière à capturer les corrélations en dehors du terme d'erreur ;
- 3 utiliser les probabilités logit comme simples approximations.

Rappelons que

$$P_{i,k}[\beta] = P_{\epsilon}[\epsilon_{i,l} \leq \epsilon_{i,k} + V_{i,k}(\beta) - V_{i,l}(\beta), \forall l \in \mathcal{K}],$$

ce qui donne

$$\begin{aligned} P_{i,k}[\beta] &= \int_{-\infty}^{+\infty} P[\epsilon_{i,l} \leq \epsilon_{i,k} + V_{i,k}(\beta) - V_{i,l}(\beta), \forall l \in \mathcal{K} \mid \epsilon_{i,k}] dF_{\epsilon_{i,k}}, \\ &= \int_{-\infty}^{+\infty} \prod_{l \neq k} P[\epsilon_{i,l} \leq s + V_{i,k}(\beta) - V_{i,l}(\beta)] e^{-s} e^{-e^{-s}} ds \\ &= \int_{-\infty}^{+\infty} \prod_{l \neq k} e^{-e^{-(s+V_{i,k}-V_{i,l})}} e^{-s} e^{-e^{-s}} ds \\ &= \int_{-\infty}^{+\infty} e^{-e^{-s} - \sum_l e^{-(V_{i,k}-V_{i,l})}} e^{-s} ds \end{aligned}$$

Soit $t := e^{-s}$. L'intégrale devient

$$\begin{aligned} P_{i,k}[\beta] &= \int_0^{+\infty} e^{-t - \sum_l e^{-(V_{i,k} - V_{i,l})}} dt \\ &= \left[\frac{e^{-t - \sum_l e^{-(V_{i,k} - V_{i,l})}}}{-\sum_l e^{-(V_{i,k} - V_{i,l})}} \right]_0^{+\infty} \\ &= \frac{1}{\sum_l e^{-(V_{i,k} - V_{i,l})}} \\ &= \frac{e^{V_{i,k}}}{\sum_l e^{V_{i,l}}}. \end{aligned}$$

On remarque que plus l'utilité observée d'une alternative est grande, plus la probabilité de la choisir est grande.

Si la probabilité est très basse ou très élevée, l'effet d'un changement au niveau de V est faible.

Le point auquel l'augmentation dans l'utilité représentative a le plus grand effet est au voisinage d'une probabilité proche de $1/2$.

Au niveau opérationnel (par exemple en transport), cela nous indique qu'il est préférable de concentrer ses efforts sur des services de niveau moyen (pas efficace, mais pas terriblement pauvre non plus).

Normalisation d'échelle

Dans les modèles logit, les variances des termes d'erreur sont normalisées à $\pi^2/6$, en raison du choix d'un facteur d'échelle égal à 1. Si on choisit un autre facteur d'échelle λ , la variance vaudrait $\lambda^2\pi^2/6$. La probabilité de choix devient alors

$$\frac{e^{V_{i,k}/\lambda}}{\sum_l e^{V_{i,l}/\lambda}}.$$

Si on fait tendre λ vers ∞ ,

$$\lim_{\lambda \rightarrow \infty} P_{i,k} = \frac{1}{K}.$$

Le modèle n'explique plus rien (tout est dans le bruit).

A l'opposé,

$$\lim_{\lambda \rightarrow 0} P_{i,k} = \begin{cases} 1 & \text{si } V_{i,k} > V_{i,l} \forall l \neq k; \\ 0 & \text{si } V_{i,k} < \max_l \{V_{i,l}\}. \end{cases}$$

Exemple

Un ménage doit choisir entre deux systèmes de chauffage : gaz ou électricité.

Prenons les utilités

$$U^{(g)} = \beta_1 C_a^g + \beta_2 C_o^g + \epsilon^{(g)},$$

$$U^{(e)} = \beta_1 C_a^e + \beta_2 C_o^e + \epsilon^{(e)}$$

où g et e désignent gaz et électricité (respectivement), C_a et C_o sont les coûts d'achats et de fonctionnement (annuel).

Nous supposons de plus $\beta_1 < 0$, $\beta_2 < 0$.

Exemple (suite)

Si $\epsilon^{(g)}$ et $\epsilon^{(e)}$ sont i.i.d. Gumbel,

$$P_g = \frac{e^{\beta_1 C_a^g + \beta_2 C_o^g}}{e^{\beta_1 C_a^g + \beta_2 C_o^g} + e^{\beta_1 C_a^e + \beta_2 C_o^e}}.$$

β_2/β_1 représente la volonté de payer du couple (WTP : willingness to pay) pour une réduction du coût de fonctionnement.

Supposons par exemple $\beta_1 = -0.20$, $\beta_2 = -1.14$. Ce qui donne une WTP de 5.70, i.e. le ménage est prêt à payer 5.70 (euros, dollars, ...) pour un système dont le coût de fonctionnement annuel est d'une unité inférieure.

Exemple (suite)

Ajoutons une troisième option, par exemple le mazout :

$$U^{(m)} = \beta_1 C_a^m + \beta_2 C_o^m + \epsilon^{(m)}.$$

Alors,

$$P_g = \frac{e^{\beta_1 C_a^g + \beta_2 C_o^g}}{e^{\beta_1 C_a^g + \beta_2 C_o^g} + e^{\beta_1 C_a^e + \beta_2 C_o^e} + e^{\beta_1 C_a^m} + e^{\beta_2 C_o^m}}.$$

Forces et faiblesses du logit

- Le logit peut représenter les variations de goût (liées aux caractéristiques individuelles).
- Substitution proportionnelle à travers les alternatives.
- Ne peut capturer la dynamique de choix répétés.

Le logit peut capturer des variations de goût qui varient systématiquement par rapport aux variables observées, mais ne peut gérer des goûts aléatoires.

Exemple : choix de ménages parmi les marques et modèles de voitures à acheter.

Attributs :

- PA_k : prix d'achat du modèle/marque k ;
- El_k : espace intérieur.

Utilité :

$$U_{i,k} = \alpha_i PA_k + \beta_i El_k + \epsilon_{i,k},$$

où α_i et β_i sont des paramètres spécifiques au ménage i .

Exemple : suite

On suppose

$$\alpha_i = \frac{\theta}{W_i},$$
$$\beta_i = \rho M_i,$$

où M_i est le nombre de personnes dans le ménage. et W_i est le revenu du ménage.

Dès lors

$$U_{i,k} = \frac{\theta}{W_i} PA_k + \rho M_i EI_k + \epsilon_{i,k}.$$

Si les $\epsilon_{i,k}$ sont i.i.d. et suivent une Gumbel, nous obtenons un modèle logit classique, où les caractéristiques du ménage modulent les préférences.

Et on peut complexifier le modèle. . .

Variations de goût aléatoires

Supposons que les préférences des ménages sont influencées par d'autres aspects non directement observables.

Introduisons deux variables aléatoires μ_i et η_i comme suit :

$$\alpha_i = \frac{\theta}{W_i} + \eta_i,$$
$$\beta_i = \rho M_i + \mu_i,$$

ce qui conduit à

$$U_{i,k} = \frac{\theta}{W_i} PA_k + \eta_i PA_k + \rho M_i EI_k + \mu_i EI_k + \epsilon_{i,k}.$$

Variations de goût aléatoires (suite)

Puisque μ_i et η_i ne sont pas observés, nous devons incorporer $\eta_i PA_k$ et $\mu_i EI_k$ dans le terme d'erreur :

$$\bar{\epsilon}_{i,k} = \eta_i PA_k + \mu_i EI_k + \epsilon_{i,k}.$$

Les $\bar{\epsilon}_{i,k}$ ne sont plus indépendants, vu qu'ils partagent des termes communs entre alternatives, et ne sont pas identiquement distribués vu que PA_k et EI_k varient entre alternatives.

Une augmentation de la probabilité d'une alternative implique nécessairement une diminution dans la probabilité pour les autres alternatives.

Une propriété fondamentale des modèles logit est l'indépendance d'alternatives irrelevantes (I.I.A. : **independence from irrelevant alternatives**).

Considérons les alternatives k_1 et k_2 . Alors,

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

Le rapport des probabilités de deux alternatives ne dépend pas des autres alternatives.

Exemple des bus rouges/bleus (Ben Akiva et Lerman).

Décision par rapport à deux modes de transport : la voiture (c) ou le bus rouge (b_r).

On suppose que les probabilités de choix sont équiprobables :

$$P_c = P_{b_r} = \frac{1}{2}.$$

Ajout de l'alternative bus bleu (b_b). Alors

$$P_c = P_{b_r} = P_{b_b} = \frac{1}{3}.$$

Conséquences de l'I.I.A. (suite)

L'I.I.A. est vraie au niveau individuel, pas pour des parts de la population. Elle ne s'applique qu'à des segments homogènes de la population. Si nous sommes capables d'identifier des segments de marchés dans la population avec des préférences systématiques différences, l'IIA est moins problématique.

Avantages de l'I.I.A.

Si le nombre d'alternatives est très grands, le calcul des probabilités de choix peut être très coûteux.

Lors de l'estimation, on peut remplacer le dénominateur par la somme sur un sous-ensemble d'alternatives, tirées aléatoirement. Les paramètres estimés restent consistants par rapport au modèle logit estimés avec toutes les alternatives.

Lors de l'analyse des choix entre différentes alternatives, on peut aussi se limiter à des sous-ensembles d'alternatives.

- 1 Estimer le modèle.
- 2 Réestimer le modèle sur un sous-ensemble d'alternatives.
- 3 Si l'I.I.A. tient, les paramètres obtenus dans le sous-ensemble d'alternative ne variera pas significativement de ceux obtenus avec toutes les alternatives.

Mais si ces tests échouent, cela ne donne pas beaucoup d'information sur le modèle à utiliser à la place.

Données panel : un individu delivre plus d'une observation.
Sinon, on parle de données cross-sectional.

Si les facteurs non-observés qui affectent les preneurs de décision sont indépendants au cours des répétitions, alors le logit peut-être utilisé de la même manière. On peut modifier le modèle pour tenir compte de la dynamique entre les facteurs observés.

Si les observations sont dépendantes, le logit peut rester consistant. Toutefois, nous ne pouvons directement représenter la dynamique des facteurs non-observés.

Variables avec décalage

Idée : faire entrer comme variables explicatives des observations des étapes précédentes.

Exemple : inertie dans les choix :

$$V_{ikt} = \alpha y_{ik(t-1)} + \beta x_{ikt},$$

avec $\alpha > 0$ et

$$y_{ik(t-1)} = \begin{cases} 1 & i \text{ a choisi } k \text{ dans la période } t-1; \\ 0 & \text{sinon.} \end{cases}$$

Note : $y_{ik(t-1)}$ est indépendant de ϵ_{ikt} .

Surplus consommateur

But : mesurer les changements de surplus consommateur associé avec une politique.

Si nous ajoutons une alternative, ou modifions les attributs d'alternatives existantes, nous souhaitons mesurer les bénéfices et les comparer aux coûts.

Le surplus consommateur (SC) est l'utilité (en termes monétaires) que la personne i reçoit dans la situation de choix :

$$SC_i = \frac{1}{\alpha_i} \max_k \{U_{i,k}\},$$

où α_i est l'utilité marginale du revenu :

$$\frac{dU_i}{dY_i} = \alpha_i,$$

avec Y_i le revenu de la personne i .

Surplus consommateur (suite)

Dès lors,

$$\frac{1}{\alpha_j} = \frac{dY_j}{dU_k}.$$

Comme nous n'observons que les $V_{i,k}$, nous calculerons

$$E[SC_i] = \frac{1}{\alpha_j} E \left[\max_k \{ V_{i,k} + \epsilon_{i,k} \} \right],$$

Si les termes d'erreur sont i.i.d. Gumbel, et que l'utilité est linéaire par rapport au revenu (et donc α_j est une constante), on peut montrer que

$$E[SC_i] = \frac{1}{\alpha_j} \ln \left(\sum_{k=1}^K e^{V_{i,k}} \right) + C,$$

où C est une constante inconnue qui représente la valeur absolue de l'utilité ne pouvant être mesurée.

Surplus consommateur (suite)

Le terme

$$\sum_{k=1}^K e^{V_{i,k}}$$

qui apparaît également au dénominateur de la probabilité logit, est appelé terme logsum.

On peut également considérer toute la population en prenant

$$\frac{1}{I} \sum_{i=1}^I w_i E[CS_i].$$

Un changement sera évalué en calculant le surplus avant (0) et après (1) :

$$\Delta E[SC_i] = \frac{1}{\alpha_i} \left(\ln \left(\sum_{k=1}^{K_1} e^{V_{i,k}^1} \right) - \ln \left(\sum_{k=1}^{K_0} e^{V_{i,k}^0} \right) \right)$$

Habituellement, une variable de coût ou de prix entre dans l'utilité représentative, auquel cas (avec des utilités linéaires), α_j est simplement l'opposé de ce coefficient.