

Identification de modèles

Fabian Bastin

Avril 2011

Fonction d'utilité

Rappelons l'hypothèse

$$U_k(\beta) = V_k(\beta) + \epsilon_k.$$

La plupart des modèles supposent également que U_k (donc V_k) est linéaire par rapport à β :

$$V_k(\beta) = \beta^T x_k,$$

où x_k est le vecteur d'attributs associés à l'alternative k .

Il est possible de fixer une composante de x_k à 1, et en notant

$$\beta = (\beta_0, \beta_1, \dots, \beta_n),$$

nous obtenons

$$V_k(\beta) = \beta_0 + \sum_{i=1}^n \beta_i x_k^{(i)},$$

où $x_k^{(i)}$ est la i^e composante de x_k .

Constante spécifique d'alternative

Rappelons que seule la différence entre les utilités importe.

Par conséquent, β_0 doit être décomposé entre les alternatives :

$$\beta_0 = \left(\beta_0^{(1)}, \beta_0^{(2)}, \dots, \beta_0^{(K)} \right),$$

et

$$V_k(\beta) = \beta_0^{(k)} + \sum_{i=1}^n \beta_i x_k^{(i)}.$$

$\beta_0^{(k)}$ est appelé constante spécifique d'alternative, et capture l'effet moyen sur l'utilité de tous les facteurs non inclus dans le modèle (vu que $E[\epsilon_k] = 0$).

Une des constantes est posée à 0 (pourquoi ?).

Variables socio-démographiques

Les valeurs des attributs viennent souvent des alternatives. Considérons par exemple le choix entre le bus et la voiture. Nous pouvons alors considérer par exemple le temps et le coût comme attributs des utilités.

Il est toutefois raisonnable de croire que les caractéristiques propres à l'individu peuvent aussi entrer dans les fonctions d'utilité.

Ces caractéristiques peuvent entrer dans le modèle si elles sont spécifiées de manière à créer des différences dans les utilites entre alternatives.

Exemple

A partir de maintenant, nous ajouterons également un indice i pour désigner l'individu en considération :

- les individus font habituellement face à des valeurs d'attributs différents pour une même alternative ;
- ceci permet de représenter les variables socio-économiques.

Exemple : voiture (c) vs bus (b).

$$U_{i,c} = \beta_T T_{i,c} + \beta_M M_{i,c} + \epsilon_{i,c},$$

$$U_{i,b} = \beta_{0,b} + \beta_T T_{i,b} + \beta_M M_{i,b} + \theta_{i,b} w_i + \epsilon_{i,b},$$

où nous prenons T pour time, M pour money, et w pour wealth, représentant le revenu individuel.

Exemple (suite)

$\theta_b^0 = \theta_b - \theta_c$ donne l'effet différentiel du revenu sur l'utilité du bus comparativement à la voiture.

Nous pouvons aussi diviser le coût par le revenu, afin de représenter une sensibilité différente au coût :

$$U_{i,c} = \beta_T T_{i,c} + \beta_M \frac{M_{i,c}}{w_i} + \epsilon_{i,c},$$

$$U_{i,b} = \beta_{0,b} + \beta_T T_{i,b} + \beta_M \frac{M_{i,b}}{w_i} + \theta_{i,b} w_i + \epsilon_{i,b},$$

Quand les variables socio-démographiques interagissent directement avec les attributs des alternatives, il n'y a pas besoin de normaliser.

L'échelle (ordre de grandeur) des utilités est sans importance.
Les modèles

$$U_{i,k} = V_{i,k} + \epsilon_{i,k}, \quad k = 1, \dots, K,$$

$$U_{i,k} = \lambda V_{i,k} + \lambda \epsilon_{i,k}, \quad k = 1, \dots, K,$$

sont équivalents, pour tout $\lambda > 0$.

Ceci a toutefois un impact sur le terme d'erreur, vu que

$$\text{Var}[\lambda \epsilon_{i,k}] = \lambda^2 \text{Var}[\epsilon_{i,k}].$$

Nous devons donc veiller à normaliser les termes d'erreur.

Interprétation : permet de prendre en compte les variations d'utilités entre individus présentant une même composante déterministe.

La densité $f(\epsilon)$ est la distribution de la portion non-observée de l'utilité au sein de la population qui partage la même portion observée de l'utilité.

Pour le chercheur, ce terme couvre

- 1 la probabilité subjective que l'utilité non-observée prendra des valeurs données, i.e. la portion de l'utilité due aux individualités ;
- 2 les facteurs étranges qui conduisent la personne à choisir l'alternative i étant donné les facteurs observés.

Plusieurs hypothèses peuvent être faites sur les termes d'erreur.

Homoscédasticité : même variance, i.e.

$$\text{Var}[\epsilon_{i,k}] = \sigma^2,$$

pour tout i, k .

On normalisera souvent la variance à 1, ce qui donne

$$U'_{ik} = \frac{V_{ik}}{\sigma} + \epsilon'_{ik}.$$

Rapports entre coefficients

Le même principe lorsque nous estimons un modèle sur différents jeux de données.

Considérons les coefficients fictifs ci-dessous

Ville	Bruxelles	Paris
Coût	-0.55	- 0.81
Temps	-1.78	-2.69
Rapport	0.309	0.301

Les rapports sont similaires, mais les coefficients sont 50% plus grands pour Paris que pour Bruxelles.

Cela indique plus de variance pour Bruxelles que pour Paris.

Hétéroscédasticité

La variance peut être différente pour différents segments de la population.

On normalise la variance pour un segment, et on estime la variance pour chaque autre segment, relativement au segment de référence.

Supposons par exemple que nous voulons utiliser les données de Paris et Bruxelles conjointement :

$$U_{i,k} = \alpha T_{i,k} + \beta M_{i,k} + \epsilon_{i,k}^{\{B,P\}}.$$

Le rapport des variances est

$$\kappa = \frac{\text{Var} \left[\epsilon_{i,k}^P \right]}{\text{Var} \left[\epsilon_{i,k}^B \right]}$$

Hétéroscédasticité (suite)

Nous pouvons diviser l'utilité pour Paris par $\sqrt{\kappa}$ pour normaliser à la variance de $\epsilon_{i,k}^B$, vu que

$$\text{Var} \left[\frac{\epsilon_{i,k}^P}{\sqrt{\kappa}} \right] = \frac{1}{\kappa} \text{Var} [\epsilon_{i,k}^P] = \text{Var} [\epsilon_{i,k}^B].$$

Ceci donne

$$\text{Bruxelles : } U_{i,k} = \alpha T_{i,k} + \beta M_{i,k} + \epsilon_{i,k},$$

$$\text{Paris : } U_{i,k} = \alpha \frac{T_{i,k}}{\sqrt{\kappa}} + \beta \frac{M_{i,k}}{\sqrt{\kappa}} + \epsilon_{i,k}.$$

κ doit être estimé en même temps que α et β . Ici, $\kappa \approx 1.2$.

La situation se complexifie si les erreurs entre alternatives sont dépendantes, en particulier s'il existe des corrélations entre les composantes du vecteur

$$\epsilon = (\epsilon_{i,1}, \dots, \epsilon_{i,K}).$$

Nous noterons la matrice de variance-covariance par Σ .

Ce modèle est équivalent à celui dans lequel toutes les utilités sont différenciées de, disons, la première alternative :

$$\tilde{U}_{i,k} = U_{i,k} - U_{i,1},$$

et

$$\tilde{\epsilon}_i = (\epsilon_{i,2} - \epsilon_{i,1}, \epsilon_{i,3} - \epsilon_{i,1}, \dots, \epsilon_{i,K} - \epsilon_{i,1}).$$

Erreurs corrélées : normalisation

Notons $\sigma_{k,l}$ la corrélation entre les termes d'erreur associés aux alternatives k et l .

La matrice de variance-covariance du modèle devient

$$\begin{pmatrix} \sigma_{11}^2 + \sigma_{22}^2 - 2\sigma_{12} & \sigma_{11}^2 + \sigma_{23} - \sigma_{12} - \sigma_{13} & \dots & \sigma_{11}^2 + \sigma_{2K} - \sigma_{12} - \sigma_{1K} \\ \sigma_{11}^2 + \sigma_{23} - \sigma_{12} - \sigma_{13} & \sigma_{11}^2 + \sigma_{33}^2 - 2\sigma_{13} & \dots & \sigma_{11}^2 + \sigma_{3K} - \sigma_{13} - \sigma_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{11}^2 + \sigma_{2K} - \sigma_{12} - \sigma_{1K} & \sigma_{11}^2 + \sigma_{KK}^2 - 2\sigma_{1K} & \dots & \sigma_{11}^2 + \sigma_{KK}^2 - 2\sigma_{1K} \end{pmatrix}$$

Il est courant de normaliser à 1 une des variances des différences des termes d'erreur, par exemple en divisant Σ par

$$\sqrt{\sigma_{11}^2 + \sigma_{22}^2 - 2\sigma_{12}}.$$

Notons aussi la diminution du nombre de paramètres à estimer dans Σ , passant de $K(K+1)/2$ à $K(K-1)/2 - 1$.

Il faut rester prudent si nous souhaitons tirer des statistiques agrégées à partir des modèles de choix discrets, car un modèle de choix discret n'est pas linéaire par rapport aux utilités.

Par conséquent, par exemple, l'espérance de la probabilité de choix (sur la population) ne peut être calculée en prenant l'espérance des utilités.

L'approche la plus directe est alors, pour une population d'individus $i = 1, \dots, I$, est de prendre la moyenne empirique :

$$\hat{P}_k = \frac{1}{I} \sum_{i=1}^I P_{ik}.$$

Dans certains cas, on affectera des poids w_i aux individus, par exemple pour tenir compte des proportions dans la population :

$$\hat{P}_k = \frac{1}{I} \sum_{i=1}^I w_i P_{ik},$$

avec

$$\sum_{i=1}^I w_i = 1.$$

Choisir des w_i différents revient à stratifier l'échantillon. Les poids peuvent par exemple représenter la probabilité d'appartenance à un groupe donné, et la probabilité de choix, la probabilité sous la condition d'appartenir à ce groupe. Souvent, cela permet de réduire la variance de l'estimateur.

Soit $\hat{l}_k = I \cdot \hat{P}_k$. Le rapport

$$\frac{\hat{l}_k}{I}$$

est un estimateur de la part de marché de l'alternative k .

Des utilités non-linéaires sont parfois nécessaires dans certains contextes, mais sont plus difficiles à estimer.

La fonction de log-vraisemblance n'est plus forcément concave, et quelle est la bonne forme d'utilité à choisir ?

Mauvaise idée parfois rencontrée : attaquer le problème d'estimation au moyen de métaheuristiques : lent et imprécis.

Utilités non-linéaires : exemple

Choix du mode de transport (voiture ou bus) et du nombre d'heures de travail.

Travail : obtention de biens (B) ; salaire horaire : w .
Loisir(L).

Fonction d'utilité retenue :

$$U(\beta) = (1 - \beta) \ln G + \beta \ln L.$$

Choix du nombre d'heures à travailler par jour pour maximiser l'utilité si il y consommation totale de la richesse produite pendant les heures de loisir restantes ?

Exemple (suite)

Chaque mode de transport à un coût et un temps d'utilisation, qui doivent être pris en compte dans la maximisation d'utilité.

Train et McFadden (1978) ont alors obtenu la fonction d'utilité

$$U_k(\beta) = -\alpha \left(\frac{c_k}{w^\beta} + w^{1-\beta} t_k \right)$$

où k désigne la voiture ou le bus.