Modélisation et simulation du comportement — Modèles de choix discret

Michel Bierlaire

michel.bierlaire@epfl.ch

Laboratoire Transport et Mobilité — EPFL

Plan du cours

- Introduction et théorie
- De l'observation à la prédiction
- Laboratoires

Introduction

Motivation:

- Comprendre
- Prédire
- Influencer

Domaines :

- Marketing
- Transport
- Politique
- Management
- Nouvelles technologies

Introduction

Domaine :
→ Marketing
→ Transport
→ Politique
→ Management
→ Nouvelles technologies

Type de comportement :
→ Choix d'une marque
→ Choix d'un mode de transport
→ Choix d'un président
→ Choix d'une politique de gestion
→ Choix d'investissements

Prix Nobel 2000 à D. McFadden "for his development of theory and methods for analyzing discrete choice"

Introduction

Dimensions comportementales en logistique

“La qualité du service logistique fait partie des critères que retiennent les acheteurs de la grande distribution pour sélectionner leurs fournisseurs. Mais peut-on dire que le consommateur est, lui aussi, sensible à la logistique ? Quels sont les cas dans lesquels cette sensibilité est la plus forte ?”

Lichtle M.-Ch., Manzano M. et Plichon V.(2000) *La sensibilité du consommateur à la logistique : mise en évidence des variables déterminantes*
Rencontres Internationales de la Recherche en Logistique

Introduction

Dimensions comportementales en logistique

"Les recherches qui ont été menées sur le comportement visionnaire des dirigeants de petites firmes ont donné lieu à une abondante littérature scientifique. Elles ont essayé de décrire le comportement anticipatoire et mobilisateur d'entrepreneurs soucieux de générer une dynamique autour d'un projet collectif et informel : la vision."

Smida A. et Condor R. (2002) *Interactions entre vision, intention et anticipation chez les dirigeants des petites entreprises*

Gestion 26(4), HEC Montréal

Introduction

Dimensions comportementales en logistique

"Le but de l'étude était de tester l'intérêt des entreprises pour une implantation à Dourges en fonction d'une part du prix de location..."

Geerts J.-F. (1997) *Etude de pré-commercialisation de la plate-forme logistique d'intérêt européen de Dourges* STRATEC SA, Bruxelles.

Introduction

Dimensions comportementales en logistique

- Comportement du client
- Comportement des dirigeants
- Comportement des entreprises
- Comportement des fournisseurs
- Comportement des collaborateurs
- etc.

Introduction

Dimensions comportementales en transport

- Choix d’une origine [localisation]
- Choix d’une destination
- Choix de mode(s) de transport
- Choix du temps de départ
- Choix de l’itinéraire
- etc...

Exemple

Vous avez le choix entre trois ordinateurs :

- Ordinateur A : 1500 €
- Ordinateur B : 2000 €
- Ordinateur C : 2500 €

Lequel choisiriez-vous ?

Exemple

Commentaires :

- Attributs des options
- Comparaison numérique
- Rationalité

Exemple

Vous avez le choix entre trois ordinateurs :

- Ordinateur A : 1500 €, 75 Mhz
- Ordinateur B : 2000 €, 150 Mhz
- Ordinateur C : 2500 €, 200 Mhz

Lequel choisiriez-vous ?

Exemple

Commentaires :

- Plusieurs attributs des options
- Comment comparer ?
- Idée : convertir en unités communes
- Exemple de règle : “retirer 10 points par €, et ajouter 20 points par Mhz”

Exemple

- Ordinateur A : (-10/€) 1500 €
  + (20/Mhz) 75 Mhz
  = -13500
- Ordinateur B : (-10/€) 2000 €
  + (20/Mhz) 150 Mhz
  = -17000
- Ordinateur C : (-10/€) 2500 €
  + (20/Mhz) 200 Mhz
  = -21000

Exemple

- Ordinateur A : $(-3/€)\ 1500\ €$
  $+\ (20/\text{Mhz})\ 75\ \text{Mhz}$
  $=\ -1000$
- Ordinateur B : $(-3/€)\ 2000\ €$
  $+\ (20/\text{Mhz})\ 150\ \text{Mhz}$
  $=\ -1000$
- Ordinateur C : $(-3/€)\ 2500\ €$
  $+\ (20/\text{Mhz})\ 200\ \text{Mhz}$
  $=\ -3500$

Exemple

Commentaires :

- Comment connaître les facteurs de conversion ?
- Comment régler les égalités ?

Exemple

- Ordinateur PC : (-3/€) 1500 €
  + (20/Mhz) 75 Mhz
  = -1000
- Ordinateur Mac : (-3/€) 2000 €
  + (20/Mhz) 150 Mhz
  = -1000
- Ordinateur IBM : (-3/€) 2500 €
  + (20/Mhz) 200 Mhz
  = -3500

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Exemple

Commentaires :

- Attributs non observables [ex: mauvaise expérience passée]
- Aspects culturels, subjectifs [ex: “Je boycotte Bill Gates”]
- Variations de personne à personne
  - Secrétaire retraitée
  - Jeune étudiant en informatique

Modèles de choix discret

Défis de la modélisation

- Appréhender les aspects rationnels du comportement
- Identifier les attributs pertinents des options
- Estimer leur influence respective
- Tenir compte des caractéristiques du preneur de décision
- Incorporer l'irrationnel et les incertitudes

Modèles de choix discret

Hypothèses :

- Options
- Attributs
- Preneur de décision
- Règles de décision

Modèles de choix discret

Hypothèses : Options :
Ensemble de choix

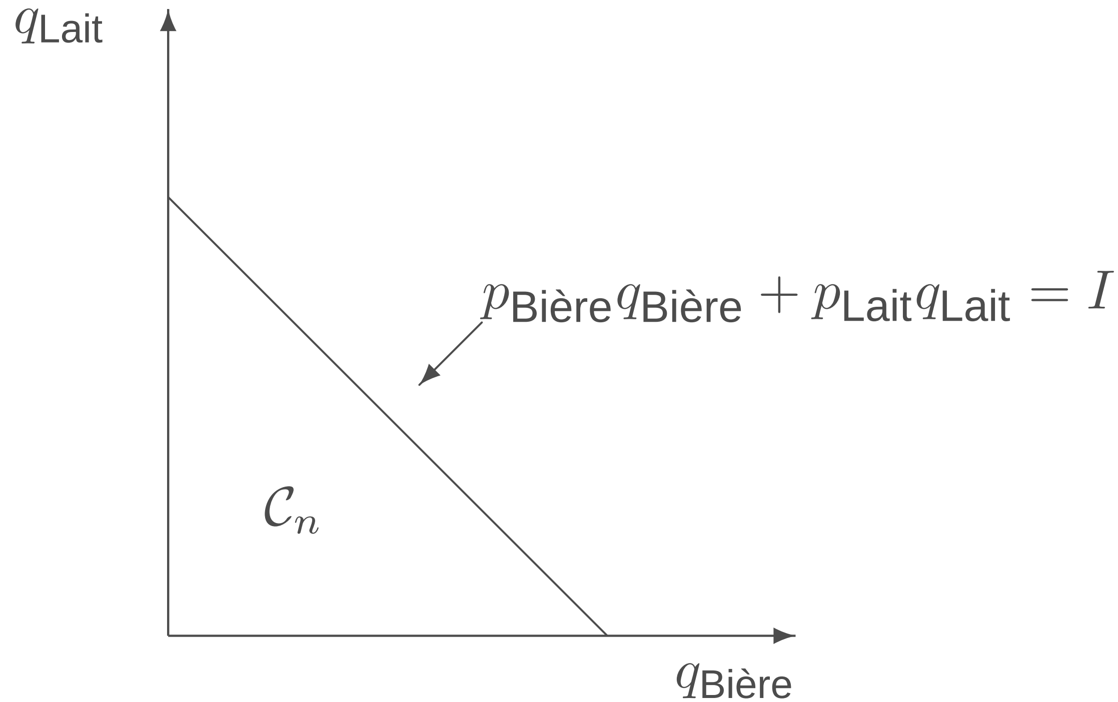
- peut être différent pour chaque personne
- certaines options peuvent être non disponibles [*Mon revendeur ne vend pas IBM*]
- le preneur de décision peut ignorer des options disponibles

- Ensemble de choix universel : $\mathcal{U}$
- Ensemble de choix individuel : $\mathcal{C}_n$

Modèles de choix discret

Hypothèses : Options :

Ensemble de choix continus :



Ensemble de choix discret :

$$\mathcal{C}_n = \{ \text{ PC, Mac, IBM } \}$$

Modèles de choix discret

Hypothèses : Attributs :

- prix
- vitesse
- capacité du disque
- design
- type de souris
- etc...

- génériques et spécifiques
- quantitatif et qualitatifs
- perception

Modèles de choix discret

Hypothèses : Attributs quantitatifs :
Exemples:

- Prix (en €)
- Vitesse (en tours/min.)
- Capacité du disque (en GO)

Simples à modéliser

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèles de choix discret

Hypothèses : Attributs qualitatifs:

Exemples:

- Design
- Type de souris

Autres exemples (transport) :

- Confort
- Fiabilité
- etc...

Comment modéliser de tels attributs ?

Modèles de choix discret

Hypothèses : Attributs qualitatifs :

1. Identification de tous les niveaux possibles
   - Design moderne
   - Design luxe
   - Design classique
   - Design bas de gamme
2. Choix d'un niveau de base : bas de gamme
3. Définition d'attributs numériques
4. Adoption d'une convention de codage

Modèles de choix discret

Hypothèses : Attributs qualitatifs :

Introduction d'un attribut 0/1 pour chaque niveau *sauf* le niveau de base

- z_a pour *moderne*
- z_b pour *luxe*
- z_c pour *classique*

Modèles de choix discret

Hypothèses : Attributs qualitatifs : Codage

	z_a	z_b	z_c
moderne	1	0	0
luxe	0	1	0
classique	0	0	1
bas de gamme	0	0	0

Si un attribut qualitatif possède n niveaux, on introduit $n-1$ variables 0/1 dans le modèle

Modèles de choix discret

Hypothèses : Attributs :

Modélisation des ℓ caractéristiques de l'option i pour l'individu n:

$$z^{in} = \begin{pmatrix} z_1^{in} \\ \vdots \\ z_\ell^{in} \end{pmatrix} \in \mathbb{R}^\ell$$

Modèles de choix discret

Hypothèses : Preneur de décision :

Exemples :

- Revenu
- Age
- Sexe
- etc.

Quantitatifs et qualitatifs

Modèles de choix discret

Hypothèses : Preneur de décision :

Modélisation des p caractéristiques socio-économiques de l'individu n:

$$S^n = \begin{pmatrix} S_1^n \\ \vdots \\ S_p^n \end{pmatrix} \in \mathbb{R}^p$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèles de choix discret

Hypothèses : Règle de décision :
Concept d'utilité, associée par l'individu n à l'alternative i

$$U_{in} = U_{in}(z^{in}, S^n)$$

Option choisie : celle associée à l'utilité maximale

Théorie économique néo-classique

Modèles de choix discret

Hypothèses : Règle de décision :

Le preneur de décision

- possède une parfaite capacité de discrimination
- est tout à fait rationnel
- est toujours cohérent

L'analyste

- a une parfaite connaissance des attributs et caractéristiques
- a une parfaite connaissance de la fonction d'utilité
- ne commet aucune erreur de mesure

Modèles de choix discret

Hypothèses : Règle de décision :

Idée : utilité modélisée par une variable aléatoire

$$U_{in} = V_{in} + \varepsilon_{in}$$

où

$$V_{in} : \mathbb{R}^{\ell+p} \longrightarrow \mathbb{R} : (z^{in}, S^n) \rightsquigarrow V_{in}(z^{in}, S^n)$$

et

ε_{in} est une variable aléatoire

Modèles de choix discret

Hypothèses : Règle de décision :

Dans ce cas, le modèle devient probabiliste

$$
\begin{aligned}
P(i|\mathcal{C}_n) &= P(U_{in} = \max_{j \in \mathcal{C}_n} U_{jn}) \\
&= P(U_{in} \geq U_{jn} \; \forall j \in \mathcal{C}_n)
\end{aligned}
$$

Modèles de choix discret

Hypothèses : Règle de décision :

Pour obtenir un modèle opérationnel, il faut

- préciser V_{in}
- préciser ε_{in}
- calculer la probabilité

Fonction d'utilité : le terme déterministe

Pour simplifier les notations, on utilise $x^{in} \in \mathbb{R}^{\ell+p}$ et $m = \ell + p$:

$$x^{in} = \begin{pmatrix} z^{in} \\ S^n \end{pmatrix} \in \mathbb{R}^m$$

et donc

$$V_{in} = V_{in}(x^{in})$$

Modèles de choix discret

Fonction d'utilité : le terme déterministe
Hypothèse la plus simple :

$$V_{in} = \beta_1 x_1^{in} + \ldots + \beta_m x_m^{in} = \sum_{j=1}^{m} \beta_j x_j^{in}$$

Question : est-ce réaliste ?

Modèles de choix discret

Fonction d'utilité : le terme déterministe

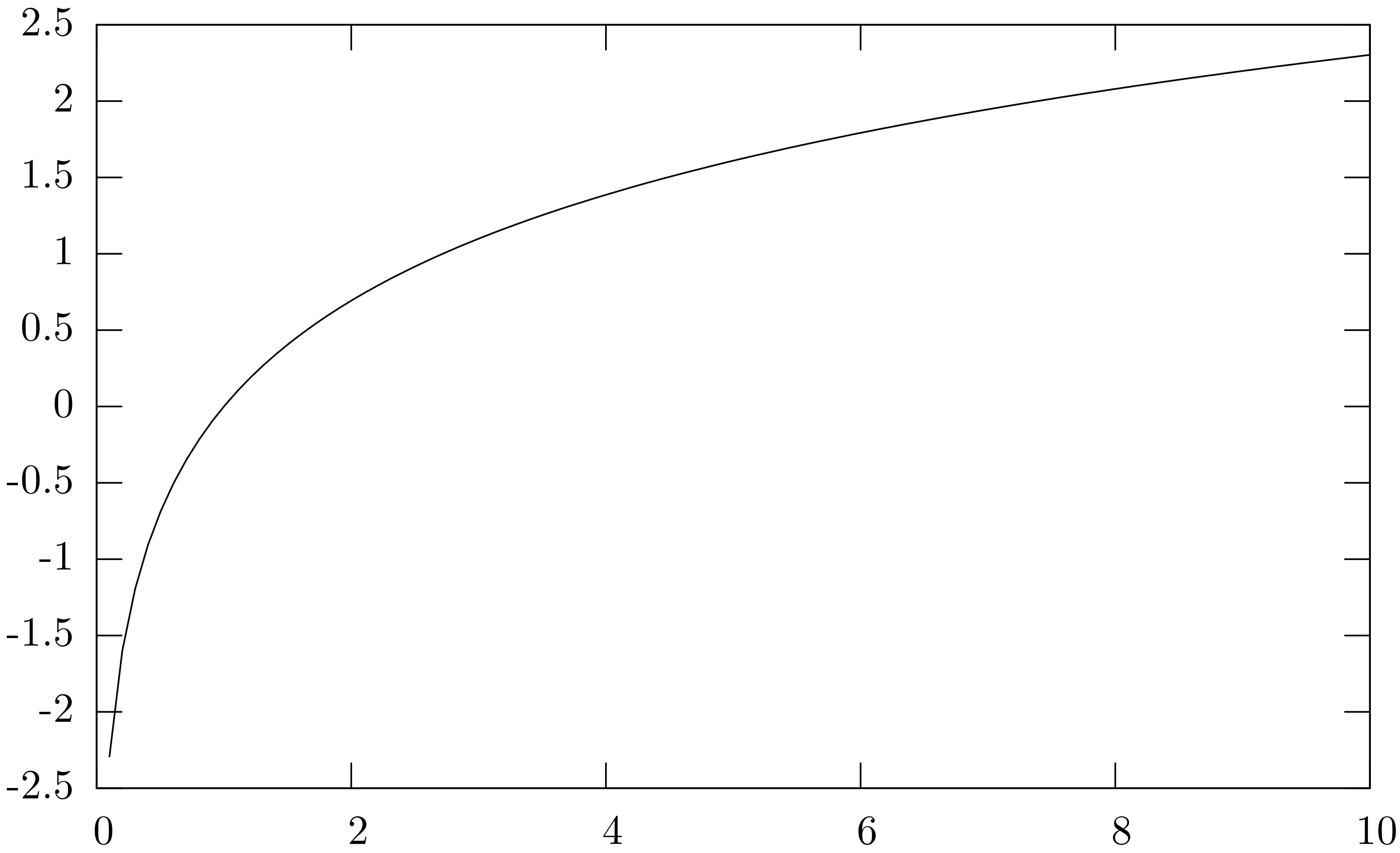
Exemple :

- M. A compare un ordinateur à 1000 € avec un à 2000 €.
- Mme B compare un ordinateur à 10000 € avec un à 11000 €.

Hypothèse de linéarité peu crédible.

Modèles de choix discret

Fonction d’utilité : le terme déterministe



Modèles de choix discret

Fonction d'utilité : le terme déterministe

Solution: considérer le logarithme du prix

- M. A associe le nombre 3 au premier et le nombre 3.301 au second
- Mme B associe le nombre 4 au premier et le nombre 4.041 au second

Même si $x_j^{in} = log(\text{prix})$, on peut supposer

$$V_{in} = \beta_1 x_1^{in} + \ldots + \beta_m x_m^{in} = \sum_{j=1}^{m} \beta_j x_j^{in}$$

Modèles de choix discret

Fonction d'utilité : le terme déterministe

Exemple :

$$
\begin{aligned}
V_{1A} &= \beta_1 \log(\mathsf{prix}_1) + \beta_2 \mathsf{âge}_A + \beta_3 \mathsf{femme}_A \\
V_{2A} &= \beta_1 \log(\mathsf{prix}_2) + \beta_2 \mathsf{âge}_A + \beta_3 \mathsf{femme}_A \\
V_{1B} &= \beta_1 \log(\mathsf{prix}_1) + \beta_2 \mathsf{âge}_B + \beta_3 \mathsf{femme}_B \\
V_{2B} &= \beta_1 \log(\mathsf{prix}_2) + \beta_2 \mathsf{âge}_B + \beta_3 \mathsf{femme}_B
\end{aligned}
$$

Modèles de choix discret

Fonction d'utilité : le terme déterministe

Au lieu de penser *prix*, on peut penser *poste budgétaire*

$$
\begin{aligned}
V_{1A} &= \beta_1 \log\left(\frac{\text{prix}_1}{\text{revenu}_A}\right) + \beta_2 \text{âge}_A + \beta_3 \text{femme}_A \\
V_{2A} &= \beta_1 \log\left(\frac{\text{prix}_2}{\text{revenu}_A}\right) + \beta_2 \text{âge}_A + \beta_3 \text{femme}_A \\
V_{1B} &= \beta_1 \log\left(\frac{\text{prix}_1}{\text{revenu}_B}\right) + \beta_2 \text{âge}_B + \beta_3 \text{femme}_B \\
V_{2B} &= \beta_1 \log\left(\frac{\text{prix}_2}{\text{revenu}_B}\right) + \beta_2 \text{âge}_B + \beta_3 \text{femme}_B
\end{aligned}
$$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire

Trois questions :

1. Quelle est sa moyenne ?
2. Quelle est sa variance ?
3. Quelle est sa distribution ?

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la moyenne

$$
\begin{aligned}
P(1|\{1,2\}) &= P(U_{1n} \geq U_{2n}) \\
&= P(V_{1n} + \varepsilon_{1n} \geq V_{2n} + \varepsilon_{2n})
\end{aligned}
$$

Supposons que $E[\varepsilon_{1n}] = \beta_0^{1n}$ et $E[\varepsilon_{2n}] = \beta_0^{2n}$.
Posons

$$
\begin{aligned}
\varepsilon'_{1n} &= \varepsilon_{1n} - \beta_0^{1n} \\
\varepsilon'_{2n} &= \varepsilon_{2n} - \beta_0^{2n}
\end{aligned}
$$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la moyenne

Ainsi

$$E[\varepsilon'_{1n}] = E[\varepsilon_{1n}] - \beta_0^{1n} = 0$$

$$E[\varepsilon'_{2n}] = E[\varepsilon_{2n}] - \beta_0^{2n} = 0$$

Si on écrit

$$\begin{aligned} U_{in} &= V_{in} && + \quad \beta_0^{in} + \varepsilon'_{in} \\ &= \textstyle\sum_{j=1}^{m} \beta_j x_j^{in} && + \quad \beta_0^{in} + \varepsilon'_{in} \end{aligned}$$

on a que la moyenne de l'erreur est nulle.

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la moyenne

Le paramètre β_0^{in} est appelé

constante spécifique à l'alternative i

Elle appréhende la partie subjective, non mesurée de l'utilité d'une alternative

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la moyenne

- Ordinateur PC : 1500 €, 100 Mhz
- Ordinateur Mac : 1500 €, 100 Mhz

$$
\begin{aligned}
U_{\text{PC}} &= \beta_1\, 1500 + \beta_2\, 100 + \beta_0^{\text{PC}} + \varepsilon'_{\text{PC}} \\
U_{\text{Mac}} &= \beta_1\, 1500 + \beta_2\, 100 + \beta_0^{\text{Mac}} + \varepsilon'_{\text{Mac}}
\end{aligned}
$$

$$
\begin{aligned}
P(\text{PC}|\{\text{PC},\text{Mac}\}) &= P(U_{\text{PC}} \geq U_{\text{Mac}}) \\
&= P(\beta_0^{\text{PC}} + \varepsilon'_{\text{PC}} \geq \beta_0^{\text{Mac}} + \varepsilon'_{\text{Mac}}) \\
&= P(\beta_0^{\text{PC}} - \beta_0^{\text{Mac}} \geq \varepsilon'_{\text{Mac}} - \varepsilon'_{\text{PC}})
\end{aligned}
$$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la moyenne

Si la partie déterministe de la fonction d'utilité contient une constante spécifique à l'alternative (CSA), on peut supposer que la moyenne de ε_{in} est nulle sans perte de généralité

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la moyenne

$$
\begin{aligned}
U_{1n} &= V_{1n} + \beta_0^{1n} + \varepsilon'_{1n} \\
U_{2n} &= V_{2n} + \beta_0^{2n} + \varepsilon'_{2n}
\end{aligned}
$$

Si on soustrait β_0^{2n} des deux utilités, on ne change rien

$$
P(U_{1n} \geq U_{2n})
$$

et donc il est équivalent d'avoir

$$
\begin{aligned}
U_{1n} &= V_{1n} + \bar{\beta}_0^{1n} & +\varepsilon'_{1n} \\
U_{2n} &= V_{2n} & +\varepsilon'_{2n}
\end{aligned}
$$

avec $\bar{\beta}_0^{1n} = \beta_0^{1n} - \beta_0^{2n}$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la moyenne

Il est impossible d'identifier toutes les CSA car seules leurs différences sont pertinentes. Dès lors, on impose une valeur arbitraire à l'une d'entre elles (typiquement 0)

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèles de choix discret

Notes : Les attributs du preneur de décision ne peuvent apparaître dans toutes les fonctions d'utilité

$$\left.\begin{array}{rcl} V_1 & = & \beta_1 x_{11} + \beta_2 \text{age} \\ V_2 & = & \beta_1 x_{21} + \beta_2 \text{age} \\ V_3 & = & \beta_1 x_{31} + \beta_2 \text{age} \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{rcl} V_1' & = & \beta_1 x_{11} \\ V_2' & = & \beta_1 x_{21} \\ V_3' & = & \beta_1 x_{31} \end{array}\right.$$

En général : attributs spécifiques aux alternatives

$$\begin{array}{rclr} V_1 & = & \beta_1 x_{11} + \beta_2 \text{age} & \\ V_2 & = & \beta_1 x_{21} + \beta_3 \text{age} & +\beta_4 \text{male} \\ V_3 & = & \beta_1 x_{31} & +\beta_5 \text{male} \end{array}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la variance

Nous avons

$$P(U_{1n} \geq U_{2n}) = P(\alpha U_{1n} \geq \alpha U_{2n}) \quad \forall \alpha > 0$$

On peut donc multiplier par n'importe quel α strictement positif sans rien changer au modèle.

De plus

$$\begin{aligned} \text{Var}(\alpha U_{1n}) &= \alpha^2 \text{Var}(U_{1n}) \\ \text{Var}(\alpha U_{2n}) &= \alpha^2 \text{Var}(U_{2n}) \end{aligned}$$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la variance

Si les variances sont identiques pour toutes les alternatives, leur valeur n'a aucune importance, et peut donc être arbitrairement choisie.

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

Hypothèse 1 : ε_{in} est la somme de beaucoup de variables aléatoires, représentant l'humeur de n, les erreurs de mesure, etc.

Théorème central-limite : la somme de nombreuses variables aléatoires indépendantes suit approximativement une loi normale

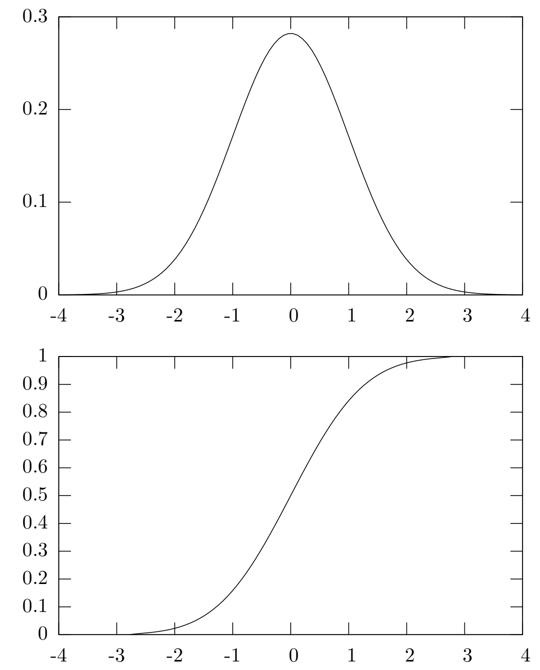
$$\varepsilon_{in} \sim N(0, 1)$$

Modèles de choix discret

Fonction d’utilité : le terme aléatoire : la distribution

Loi normale :

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}t^2}$$



Rappel: si $\varepsilon \sim N(0,1)$, alors

$$P(c \geq \varepsilon) = F(c) = \int_{-\infty}^{c} f(t)dt$$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

$$
\begin{aligned}
P(1|\{1,2\}) &= P(U_{1n} \geq U_{2n}) \\
&= P(V_{1n} + \varepsilon_{1n} \geq V_{2n} + \varepsilon_{2n}) \\
&= P(V_{1n} - V_{2n} \geq \varepsilon_{2n} - \varepsilon_{1n}) \\
&= P(V_{1n} - V_{2n} \geq \varepsilon)
\end{aligned}
$$

avec

$$
\begin{aligned}
\varepsilon_{1n} &\sim N(0,1) \\
\varepsilon_{2n} &\sim N(0,1) \\
\varepsilon &\sim N(0,2)
\end{aligned}
$$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

$$\begin{aligned} P(1|\{1,2\}) &= P(V_{1n} - V_{2n} \geq \varepsilon) \\ &= F(V_{1n} - V_{2n}) \end{aligned}$$

$$\boxed{P(1|\{1,2\}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{V_{1n}-V_{2n}} e^{\frac{1}{2}t^2} dt}$$

Intégrale non analytique !

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

S'il y a plus de deux options...

$$P(i|\mathcal{C}_n) = P(U_j - U_i \leq 0 \; \forall j \in \mathcal{C}_n)$$

$$\Delta_i U = [U_1 - Ui, \ldots, U_{i-1} - U_i, U_{i+1} - U_i, \ldots, U_j - U_i]^T$$

$$\Delta_i U \rightsquigarrow N(\Delta_i V, \Delta_i \Sigma \Delta_i^T)$$

$$f_i(x) = 2\pi^{-\frac{J-1}{2}} \left|\Delta_i \Sigma \Delta_i^T\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\Delta_i V)^T (\Delta_i \Sigma \Delta_i^T)^{-1}(x-\Delta_i V)}$$

$$P(i|\mathcal{C}_n) = \int_0^\infty \cdots \int_0^\infty f_i(x) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_J$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

Hypothèse 2 : ε_{in} est le maximum de beaucoup de variables aléatoires, représentant l'humeur de n, les erreurs de mesure, etc.

Théorème de Gumbel : le maximum de nombreuses variables aléatoires indépendantes suit approximativement une loi de Gumbel

$$\varepsilon_{in} \sim \text{Gumbel}(0, \mu)$$

Modèles de choix discret

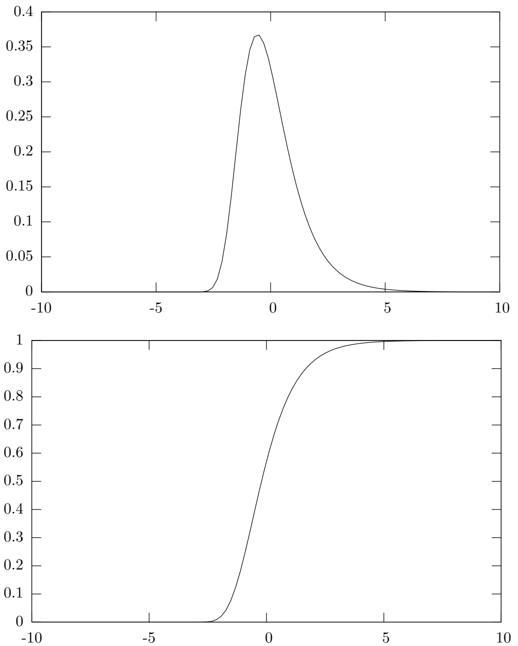
Fonction d'utilité : le terme aléatoire : la distribution

Loi de Gumbel(η, μ), avec $\mu > 0$:

$$f(t) = \mu e^{-\mu(t-\eta)} e^{-e^{-\mu(t-\eta)}}$$

Si $\varepsilon \sim$ Gumbel(η, μ), alors

$$\begin{aligned} P(c \geq \varepsilon) = F(c) &= \int_{-\infty}^{c} f(t)dt \\ &= e^{-e^{-\mu(t-\eta)}} \end{aligned}$$



Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

Si

$$\varepsilon \sim \text{Gumbel}(\eta, \mu)$$

alors

$$E[\varepsilon] = \eta + \frac{\gamma}{\mu} \quad \text{et} \quad \text{Var}[\varepsilon] = \frac{\pi^2}{6\mu^2}$$

avec

$$\gamma = \lim_{k\to\infty} \sum_{i=1}^{k} \frac{1}{i} - \ln k \approx 0.5772 \quad \text{constante d'Euler}$$

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

$$\begin{aligned} P(1|\{1,2\}) &= P(U_{1n} \geq U_{2n}) \\ &= P(V_{1n} + \varepsilon_{1n} \geq V_{2n} + \varepsilon_{2n}) \\ &= P(V_{1n} - V_{2n} \geq \varepsilon_{2n} - \varepsilon_{1n}) \\ &= P(V_{1n} - V_{2n} \geq \varepsilon) \end{aligned}$$

avec

$$\begin{aligned} \varepsilon_{1n} &\sim \text{Gumbel}(0,1) \\ \varepsilon_{2n} &\sim \text{Gumbel}(0,1) \\ \varepsilon &\sim \text{Logistique}(0,1) \end{aligned}$$

Modèle Logit

Modèles de choix discret

Fonction d'utilité : le terme aléatoire : la distribution

Pour la loi logistique (0,1), on a

$$P(c \geq \varepsilon) = F(c) = \frac{1}{1+e^{-c}}$$

$$P(1|\{1,2\}) = P(V_{1n} - V_{2n} \geq \varepsilon) = \frac{1}{1+e^{V_{2n}-V_{1n}}}$$

$$\boxed{P(1|\{1,2\}) = \frac{e^{V_{1n}}}{e^{V_{1n}} + e^{V_{2n}}}}$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Modèles de choix discret

Fonction d’utilité : le terme aléatoire : la distribution

S’il y a plus de deux options...

$$\boxed{P(i|\mathcal{C}_n) = \frac{e^{V_{in}}}{\sum_{j \in \mathcal{C}_n} e^{V_{jn}}}}$$

Modèle logit multinomial

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Exemple

Vous avez le choix entre trois ordinateurs :

- Ordinateur PC : (-10/€) 1500 €

  + (20/Mhz) 75 Mhz

  = -13500

- Ordinateur Mac : (-10/€) 2000 €

  + (20/Mhz) 150 Mhz

  = -17000

- Ordinateur IBM : (-10/€) 2500 €

  + (20/Mhz) 200 Mhz

  = -21000

Exemple

$$
\begin{aligned}
V_{\text{PC}} &= -13500 \\
V_{\text{Mac}} &= -17000 \\
V_{\text{IBM}} &= -21000
\end{aligned}
$$

$$
\begin{aligned}
P(\text{PC}) &= \frac{e^{-13500}}{e^{-13500}+e^{-17000}+e^{-21000}} = 45.95\% \\
P(\text{Mac}) &= \frac{e^{-17000}}{e^{-13500}+e^{-17000}+e^{-21000}} = 32.37\% \\
P(\text{IBM}) &= \frac{e^{-21000}}{e^{-13500}+e^{-17000}+e^{-21000}} = 21.70\%
\end{aligned}
$$

Plan du cours

- Introduction et théorie
- De l'observation à la prédiction
- Laboratoires

Estimation des modèles

$$\begin{array}{rcccccc} U_{\text{PC}} & = & \beta_1 & x_1^{\text{PC}} & + & \beta_2 & x_2^{\text{PC}} \\ & = & (-10/€) & 1500€ & + & (20/\text{Mhz}) & 75\text{Mhz} \end{array}$$

Comment trouver les valeurs de β ?

Estimation des modèles

Collecte de données

M. Hewlett :

	€	Mhz
PC	1200	50
Mac	1700	100
IBM	2500	500

Et il a choisi ... un Mac

M. Packard :

	€	Mhz
PC	1200	50
Mac	1700	100
IBM	2500	500

Et il a choisi ... un PC

M. Bill :

	€	Mhz
PC	1300	60
Mac	1800	70
IBM	2600	300

Et il a choisi ... un IBM

Estimation des modèles

Prédiction du modèle

Pour M. Hewlett

$$P_{\text{Hewlett}}(\text{Mac}|\{\text{PC}, \text{Mac}, \text{IBM}\}) = \frac{e^{V_{\text{Mac}}}}{e^{V_{\text{PC}}} + e^{V_{\text{Mac}}} + e^{V_{\text{IBM}}}}$$

avec

$$\begin{aligned} V_{\text{PC}} &= 1200\beta_1 + 50\beta_2 \\ V_{\text{Mac}} &= 1700\beta_1 + 100\beta_2 \\ V_{\text{IBM}} &= 2500\beta_1 + 500\beta_2 \end{aligned}$$

$$P_{\text{Hewlett}}(\text{Mac}|\{\text{PC}, \text{Mac}, \text{IBM}\}) = f_{\text{Hewlett}}(\beta_1, \beta_2)$$

Estimation des modèles

Prédiction du modèle
Pour tous

$$P_{\text{Hewlett}}(\text{Mac})P_{\text{Packard}}(\text{PC})P_{\text{Bill}}(\text{IBM})$$

c'est-à-dire

$$f_{\text{Hewlett}}(\beta_1, \beta_2) f_{\text{Packard}}(\beta_1, \beta_2) f_{\text{Bill}}(\beta_1, \beta_2) = \mathcal{L}(\beta_1, \beta_2)$$

Vraisemblance de l'échantillon

Estimation des modèles

Prédiction du modèle

On cherche les paramètres (β_1^*, β_2^*) tels que

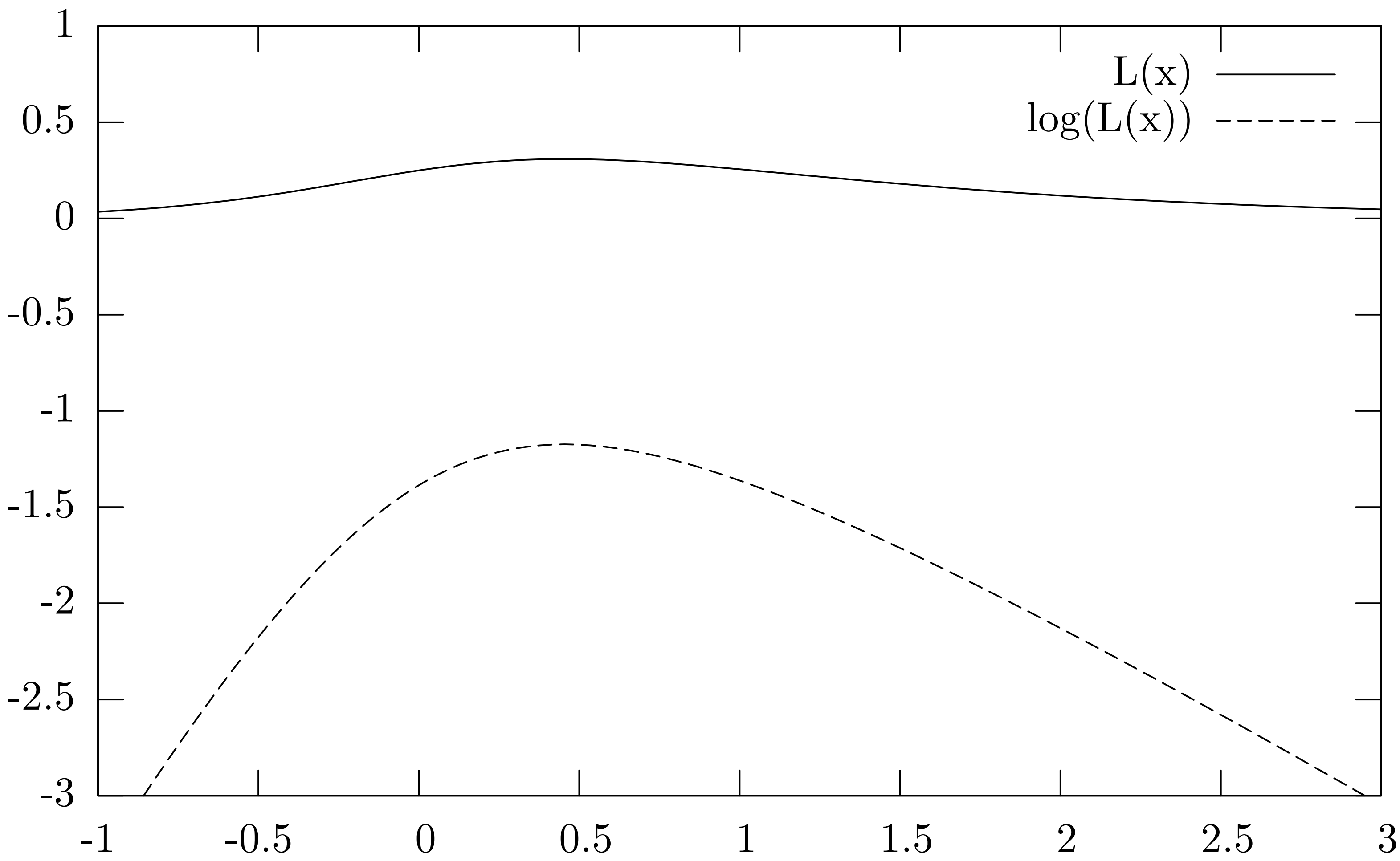
$$(\beta_1^*, \beta_2^*) = \text{argmax}_{\beta_1, \beta_2} \mathcal{L}(\beta_1, \beta_2)$$

En général

$$(\beta_1^*, \beta_2^*) = \text{argmax}_{\beta_1, \beta_2} \log \mathcal{L}(\beta_1, \beta_2)$$

$\boxed{\text{Estimation par maximum de vraisemblance}}$

Estimation des modèles



Estimation des modèles

Si

$$\mathcal{L}(\beta_1,\beta_2) = f_{\text{Hewlett}}(\beta_1,\beta_2) f_{\text{Packard}}(\beta_1,\beta_2) f_{\text{Bill}}(\beta_1,\beta_2)$$

alors

$$\begin{array}{rcl} \log \mathcal{L}(\beta_1,\beta_2) & = & \log f_{\text{Hewlett}}(\beta_1,\beta_2) \\ & + & \log f_{\text{Packard}}(\beta_1,\beta_2) \\ & + & \log f_{\text{Bill}}(\beta_1,\beta_2) \end{array}$$

BIOGEME

Objectifs

- Estimation des modèles de choix discrets
- Utilisation de divers algorithmes d'optimisation
- Logiciel "open source"
- Auteur : Michel Bierlaire, EPFL
- Versions : Linux, Windows, Mac

`http://biogeme.epfl.ch`

BIOGEME

BIerlaire's Optimization toolbox for GEV Models Estimation

Développement :

- Version 0.0: juillet 2001
- ...
- Version 1.5: juin 2007

BIOGEME

BIOGEME lit :

- un fichier de paramètres
- un fichier décrivant le modèle
- un fichier contenant les données

BIOGEME génère :

- un fichier contenant les résultats
- un fichier décrivant le modèle final
- un fichier contenant des statistiques sur les données

BIOGEME

Fichier de paramètres :

default.par

```
[GEV]
//gevAlgo = "CFSQP"
//gevAlgo = "SOLVOPT"
gevAlgo = "DONLP2"
gevTtestThreshold = 1.96
```

BIOGEME

Exemple : *Ben-Akiva & Lerman (1985) Discrete Choice Analysis: Theory and Applications to Travel Demand, MIT Press (p.88)*
Choix entre Auto et Bus

$$\begin{aligned} V_A &= \beta_1 + \beta_2 T_A \\ V_B &= \beta_2 T_B \end{aligned}$$

où T_A est le temps de trajet en auto (min), et T_B le temps en bus (min)

BIOGEME

Données :

#	Temps auto	Temps Bus	Choix	#	Temps auto	Temps Bus	Choix
1	52.9	4.4	B	11	99.1	8.4	B
2	4.1	28.5	B	12	18.5	84.0	V
3	4.1	86.9	V	13	82.0	38.0	V
4	56.2	31.6	B	14	8.6	1.6	B
5	51.8	20.2	B	15	22.5	74.1	V
6	0.2	91.2	V	16	51.4	83.8	V
7	27.6	79.7	V	17	81.0	19.2	B
8	89.9	2.2	B	18	51.0	85.0	V
9	41.5	24.5	B	19	62.2	90.1	V
10	95.0	43.5	B	20	95.1	22.2	B
				21	41.6	91.5	V

BIOGEME

Fichier de données :

akiva.dat

```
TempsVoiture    TempsBus    Choix
    52.9           4.4          2
     4.1          28.5          2
     4.1          86.9          1
    56.2          31.6          2
    51.8          20.2          2
     0.2          91.2          1
    27.6          79.7          1
    89.9           2.2          2
    41.5          24.5          2
    95.0          43.5          2
    99.1           8.4          2
...
```

BIOGEME

Description du modèle :

akiva.mod

```
// Example Ben-Akiva & Lerman (1985) p. 88
// Michel Bierlaire, EPFL

[Choice]
Choice

[Beta]
// Name Value  LowerBound UpperBound  status (0=variable, 1=fixed)
BETA1      0.0     -100.0     100.0         0
BETA2      0.0     -100.0     100.0         0

[Utilities]
// Id Name  Avail  linear-in-parameter expression (beta1*x1 + beta2*x2 +
1  Auto avail BETA1 * one + BETA2 * AutoTime
2 Bus avail BETA2 * BusTime
```

BIOGEME

Description du modèle (suite) :

akiva.mod

```
[Expressions]
// Define here arithmetic expressions for name that are not directly
// available from the data
one = 1
avail = 1

[Model]
$MNL
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BIOGEME

```
c:\biogeme\akiva> biogeme akiva akiva.dat
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
BIOGEME Version 1.5 [Sat Jun 30 12:17:17 GMT 2007]
Michel Bierlaire, EPFL
-- Compiled by Michel on MINGW32_NT-5.1
See http://biogeme.epfl.ch
                    !! CFSQP is NOT available !!
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
"In every non-trivial program there is at least one bug."

...
```

BIOGEME

Résultats :

`akiva.html`

Prédiction

Donc

$$\begin{array}{rcrcl} V_A & = & -0.2375 & - & 0.0531\,T_A \\ V_B & = & & - & 0.0531\,T_B \end{array}$$

Supposons $T_A = T_B = 50.0$
Alors $V_A = -2.8925$ et $V_B = -2.655$

$$\begin{array}{rcccl} P(A) & = & \frac{e^{V_A}}{e^{V_A}+e^{V_B}} & = & 44.1\% \\ P(B) & = & \frac{e^{V_B}}{e^{V_V}+e^{V_B}} & = & 55.9\% \end{array}$$

Validité du modèle

- Est-ce qu’on peut faire confiance à ces valeurs ?
- Si j’avais utilisé d’autres observations, les valeurs auraient été différentes. Un peu ? Beaucoup ?
- Quelle est la qualité du modèle ?

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Validité du modèle

Exemple :

Intégrons un paramètre non pertinent dans le modèle : la couleur des yeux

Validité du modèle

Fichier de données :

akiva2.dat

EyesColor	AutoTime	BusTime	Choice
1	52.9	4.4	2
0	4.1	28.5	2
1	4.1	86.9	1
0	56.2	31.6	2
1	51.8	20.2	2
1	0.2	91.2	1
1	27.6	79.7	1

...

Validité du modèle

Description du modèle :

akiva2.mod

```
[Beta]
// Name Value  LowerBound UpperBound  status (0=variable, 1=fixed)
BETA1      0.0     -100.0     100.0         0
BETA2      0.0     -100.0     100.0         0
BETA3      0.0     -100.0     100.0         0
[Utilities]
// Id Name  Avail  linear-in-parameter expression
1       Auto avail   BETA1 * one + BETA2 * AutoTime + BETA3 * EyesColor
2       Bus  avail                 BETA2 * TempsBus
```

Validité du modèle

Résultats :

akiva2.rep

```
Utility parameters
******************
Name  Value   Std err t-test p-val
----  -----   ------- ------ -----
BETA1  0.118  1.21      0.10  0.92  *
BETA2 -0.0537 0.0206  -2.61  0.01
BETA3 -0.588  1.56    -0.38  0.71  *
```

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Validité du modèle

Supposons $T_A = T_B = 50.0$

Couleur des yeux	V_A	V_B	P_A	P_B
0	-2.57	-2.68	53.0%	47.0%
1	-3.15	-2.68	38.5%	61.5%

$\boxed{\text{Cela n'a aucun sens !}}$

Validité du modèle

Résultats :

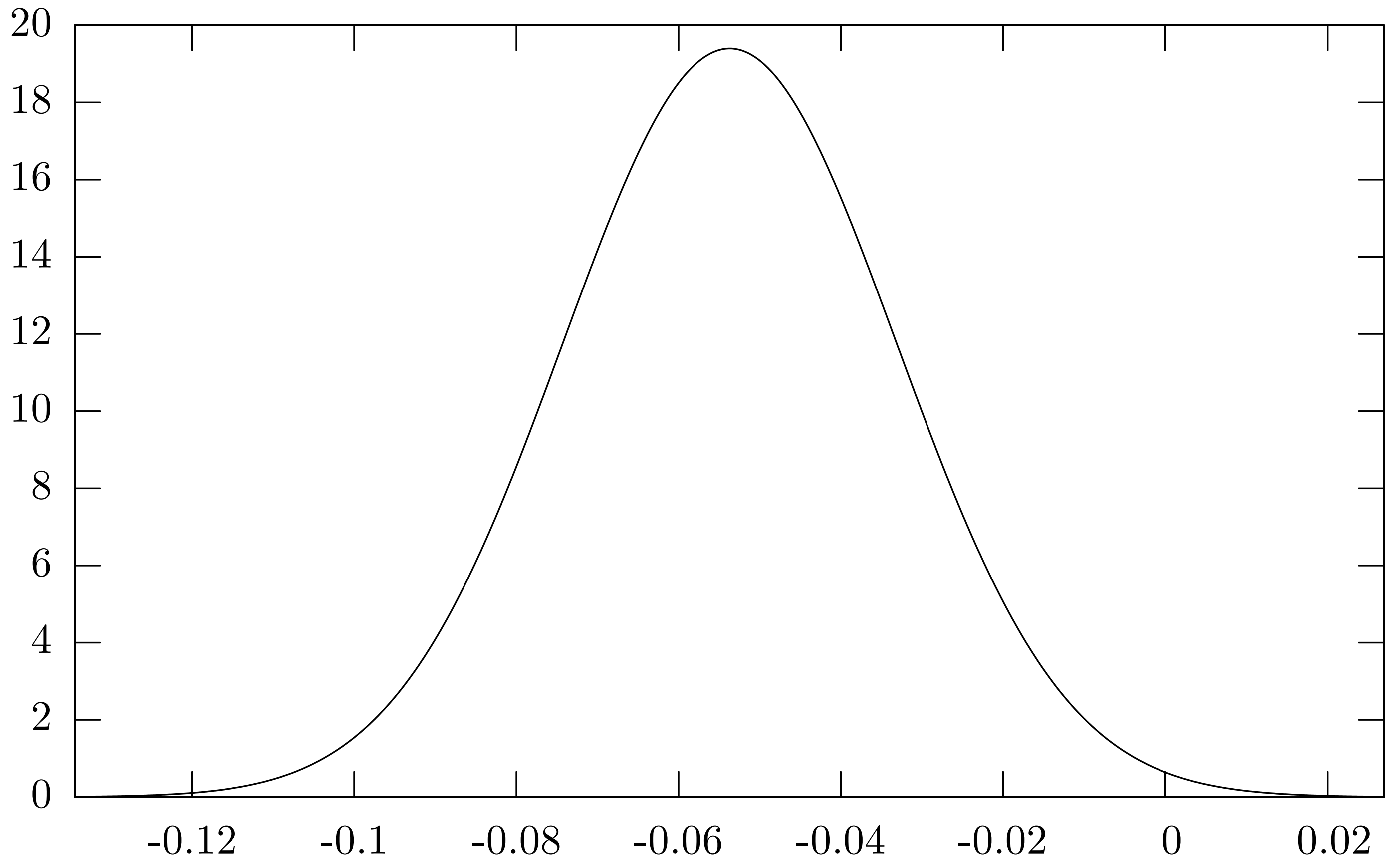
- BIOGEME ne fournit pas la valeur réelle du paramètre
- Il fournit une estimation de cette valeur
- La valeur réelle est une variable aléatoire
- Ici, on peut supposer qu'elle suit une loi normale

$$\beta \sim N(\hat{\beta}, \sigma)$$

$$\boxed{\text{BIOGEME calcule } \hat{\beta} \text{ et } \sigma}$$

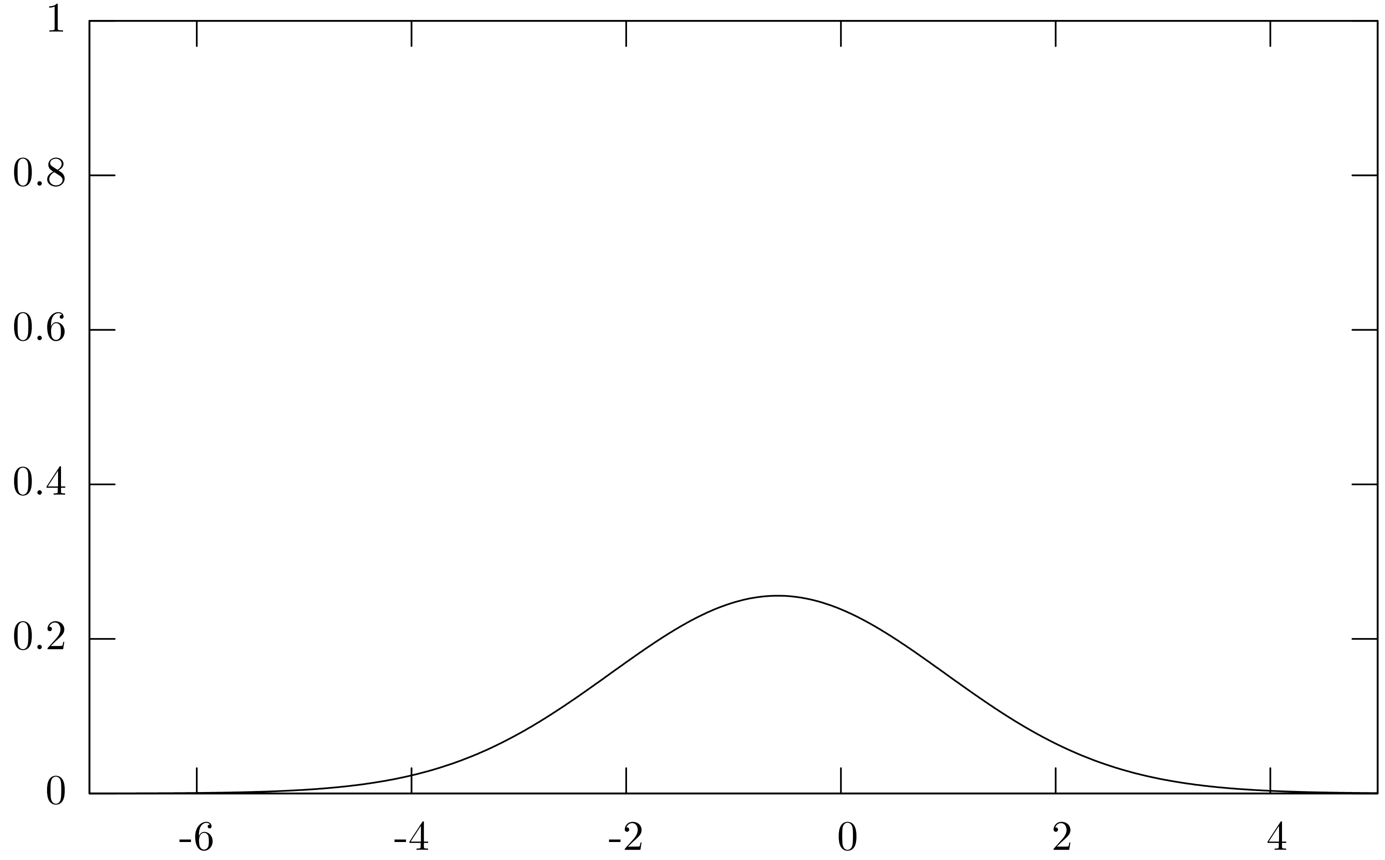
Validité du modèle

Name	Value	Std err	t-test	p-val
BETA2	-0.0537	0.0206	-2.61	0.01

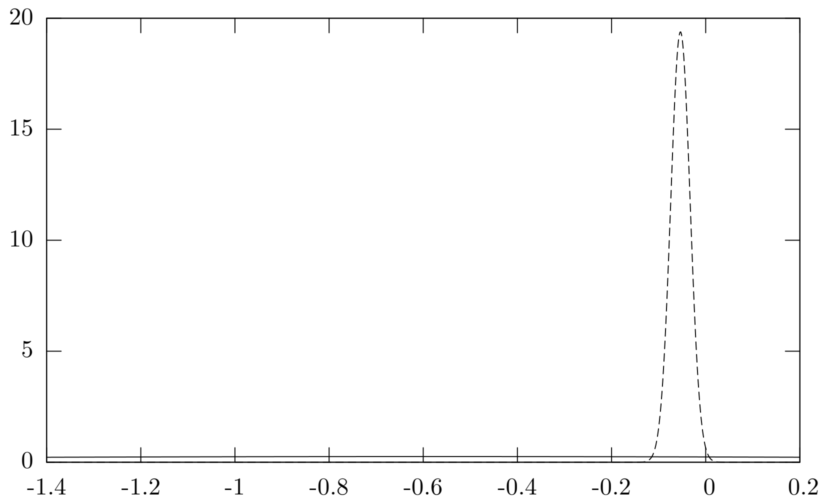


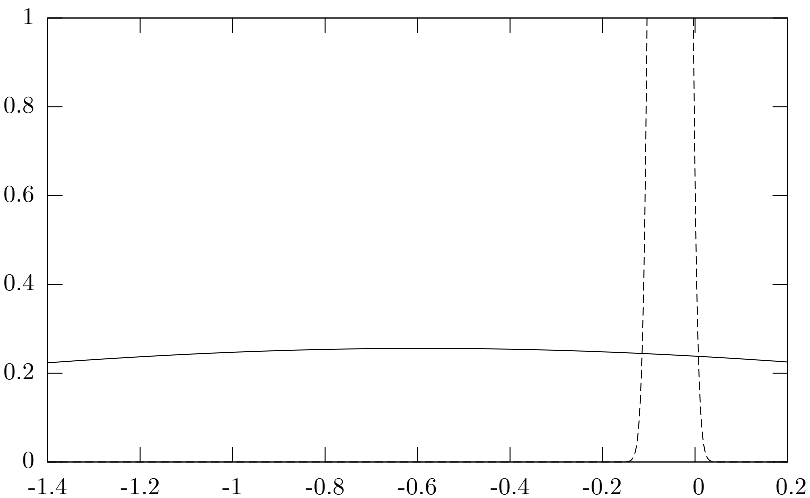
Validité du modèle

```
Name  Value   Std err t-test p-val
----  -----   ------- ------ -----
BETA3 -0.588  1.56    -0.38  0.71  *
```



Validité du modèle





Validité du modèle

Quelle confiance mettre dans la valeur $\hat{\beta}$?

Si $\beta \sim N(\hat{\beta}, \sigma)$, alors il y a 95% de chance que la valeur réelle se trouve dans l'intervalle

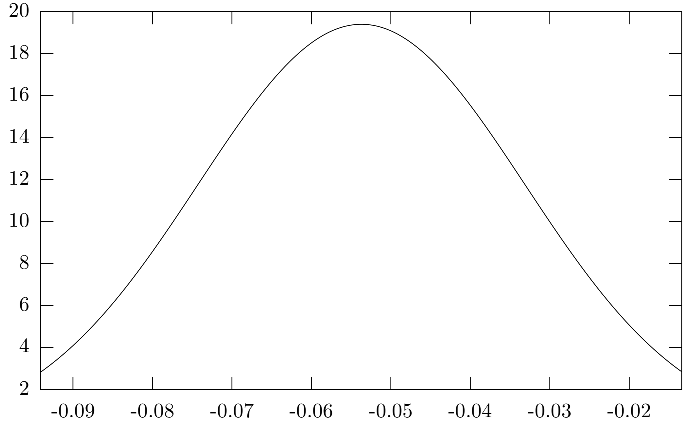
$$[\hat{\beta} - 1.96\ \sigma, \hat{\beta} + 1.96\ \sigma]$$

Intervalle de confiance

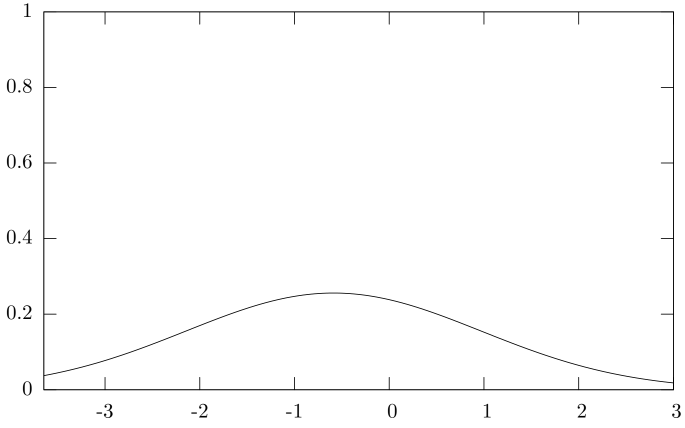
TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Validité du modèle



Temps de trajet



Couleur des yeux

Validité du modèle

Question : est-ce que la vraie valeur peut être nulle ?

Hypothèse : $\beta = 0$

On peut rejeter cette hypothèse si

$$\left| \frac{\hat{\beta}}{\sigma} \right| \geq 1.96$$

Name	Value	Std err	t-test	p-val	
BETA1	0.118	1.21	0.10	0.92	*
BETA2	-0.0537	0.0206	-2.61	0.01	
BETA3	-0.588	1.56	-0.38	0.71	*

Validité du modèle

Description du modèle :

akiva3.mod

```
[Beta]
// Name Value  LowerBound UpperBound  status (0=variable, 1=fixed)
BETA2      0.0     -100.0     100.0         0
[Utilities]
// Id Name  Avail  linear-in-parameter expression
1 Auto avail   BETA2 * AutoTime
2 Bus  avail   BETA2 * BusTime
```

Validité du modèle

Résultats :

akiva3.rep

```
Utility parameters
******************
Name  Value   Std err t-test p-val
----  -----   ------- ------ -----
BETA2 -0.0525 0.0203  -2.59  0.01
```

Validité du modèle

Récapitulons :

Modèle	$\mathcal{L}(\hat{\beta})$	# param.
`trivial`	-14.5561	0
`akiva`	-6.16604	2
`akiva2`	-6.09313	3
`akiva3`	-6.21701	1

A priori `akiva2` est meilleur. Pourtant, il comprend des paramètres non pertinents.

Validité du modèle

Soient deux modèles, A et B, comprenant resp. p et q paramètres, $p < q$.

Hypothèse : modèle A et modèle B sont équivalents.

On peut rejeter cette hypothèse si

$$-2(\mathcal{L}_A - \mathcal{L}_B) \geq F_{\chi^2}$$

où F_{χ^2} est le seuil à 95% de la loi χ^2 à $q - p$ degrés de libertés.

χ^2

$q-p$	90%	95%	$q-p$	90%	95%
1	2.706	3.841	16	23.542	26.296
2	4.605	5.991	17	24.769	27.587
3	6.251	7.815	18	25.989	28.869
4	7.779	9.488	19	27.204	30.144
5	9.236	11.070	20	28.412	31.410
6	10.645	12.592	21	29.615	32.671
7	12.017	14.067	22	30.813	33.924
8	13.362	15.507	23	32.007	35.172
9	14.684	16.919	24	33.196	36.415
10	15.987	18.307	25	34.182	37.652
11	17.275	19.675	26	35.563	38.885
12	18.549	21.026	27	36.741	40.113
13	19.812	22.362	28	37.916	41.337
14	21.064	23.685	29	39.087	42.557
15	22.307	24.996	30	40.256	43.773

Validité du modèle

A	B	$\mathcal{L}_A$	$\mathcal{L}_B$	$-2(\mathcal{L}_A - \mathcal{L}_B)$	$q-p$	F_{χ^2}
akiva3	akiva	-6.21701	-6.16604	0.10194	1	3.841
akiva	akiva2	-6.16604	-6.09313	0.14582	1	3.841
akiva3	akiva2	-6.21701	-6.09313	0.24776	2	5.991
trivial	akiva3	-14.5561	-6.21701	16.6782	1	3.841

Lorsque les modèles sont équivalents, il faut préférer le plus simple, celui comportant le moins de paramètres

Résumé

- Exemples
- Hypothèses et modèles
- Estimation
- Tests