Modélisation de la demande de transport
4A
Théorie des choix discrets

Fabien Leurent
ENPC / LVMT

Plan

Définitions

- Choix discret
  - Répartition entre des options en nombre fini
  - Extensions : options en nb infini (entier ou intervalle)
- Interprétation microéconomique
  - La répartition découle d'un choix fait par un acteur économique : le « décideur »
  - Préférences, rationalité, information : rationalité individuelle si le décideur choisit ce qu'il préfère, d'après l'information dont il dispose
- Vocabulaire
  - Univers de choix : un ensemble d'options ; chaque option est décrite par des caractères
  - Problème de choix : le décideur et son univers particulier (*attention aux captivités* !)

Historique et domaines d'application

- Origine vers 1920, en biométrie : modèle probit pour interpréter des observations 0-1
- vers 1950 : modèle logit agrégé (Berkson, Theil), prix-temps (Quandt)
- vers 1960 : modèle prix-temps-revenu (Marche) ; position microéco de la décision discrète (Block et Marchak), modèle d'adresse (Lancaster)
- vers 1970 : interprétation microéconomique du logit (McFadden), application à la demande de transport
- Depuis 1980 : généralisation en marketing

Approche empirique d'une répartition

- On observe un certain phénomène sur un ensemble Ω de situations d'occurrence ω
- Dans chaque situation ω, il y a un résultat : une certaine modalité *m*
  - soit $m^*(\omega)$ la fonction qui associe la modalité à la situation
  - Ou description par variables indicatrices, y_m dans {0, 1} telles que valeur 1 si $m = m^*(\omega)$ ou 0 sinon
- Exemples
  - Vote pour un candidat *m* des électeurs ω
  - Choix du mode *m* par des demandeurs
- Deux niveaux d'analyse
  - Niveau macroscopique, agrégé : *m*
  - Niveau microscopique, désagrégé : ω

Modèle d'analyse

- Simplifier
  - Car on ne peut (ou ne veut) pas observer le détail des ω
  - On s'intéresse à la répartition des modalités
- Chaque modalité a une proportion

$$\pi_m = \frac{\sum_{\omega \in \Omega} y_m(\omega) f(\omega)}{\sum_{\omega \in \Omega} f(\omega)}$$

  avec $f(\omega)$ le poids particulier de l'option ω
- Les proportions varient en fonction de facteurs *X* ?
  - Si $m^*(\omega, X)$ dépend de *X*, ou si $f(\omega, X)$
  - Alors on exprime $\pi_m(X)$
- Exemple
  - Si la participation à une élection dépend de la météo

Fonctions de répartition

- Fonction de répartition
  - modèle $\pi_m(X)$ des proportions pour les modalités *m* de *M*, en fonction de facteurs exogènes *X*
  - C'est un modèle probabiliste « macro »
- Formulation $\pi_m(X, \Theta)$ avec un paramètre Θ
  - L'essentiel est la formule de π_m

Formulations typiques

- Tout-ou-rien : pour une option *m*, $\pi_m = 1$, et 0 pour les autres
- Forme quotient : fonctions non-négatives $J_m(X, \Theta)$ telles que

$$\pi_m(X) = \frac{J_m(\Theta, X)}{J_M(\Theta, X)} \quad \text{avec} \quad J_M(\Theta, X) = \sum_{m \in M} J_m(\Theta, X)$$

- Forme distribution : liée à une distribution de probabilité « micro » au niveau des ω, telle que

$$\pi_m(\Theta, X) = \frac{\int \tilde{\pi}_m(\underline{\Theta}, X, \omega) f(\omega)}{\int f(\omega)}$$

  ait une forme analytique simple liée à la répartition au niveau micro et à la fonction de représentation *f* (le paramètre Θ inclut $\underline{\Theta}$ et aussi des paramètres de *f*)

Position microéconomique

- Situation de choix
  - Un décideur, avec ses caractères
  - Des options, avec leurs caractères : univers de choix
- Préférences individuelles
  - Le décideur sait comparer des options pour en choisir une
- Rationalité microéconomique
  - Le décideur est rationnel compte tenu de ses préférences et de son information : il optimise son choix
  - Information totale ou partielle, sûre ou incertaine
- Aspects individuels
  - Un individu-décideur qui décide lui-même
  - Et selon ses propres préférences

Le coût généralisé

- Hypothèses
  - Décideur *u*, caractères Y_u (observés) et Z_u (estimés)
  - Options *m*, attributs X_m
- Le coût généralisé est une fonction
  $G_m(u)$ = F(Y_u, Z_u, X_m)
- Principe de choix : le décideur choisit l'option *m* dont le coût $G_m(u)$ est minimum

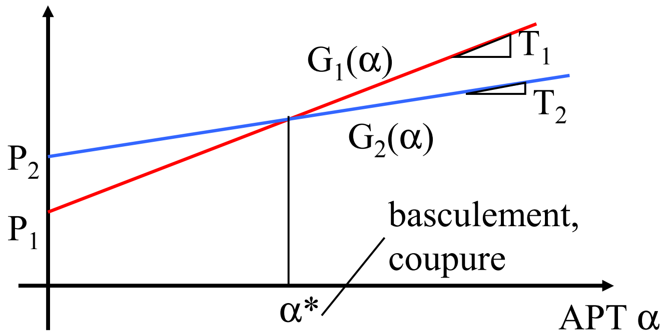
Distribution des décideurs

- Hypothèse : quand, pour un univers de choix, il y a une population de décideurs, avec des préférences / comportements variés
- Raisons des variations
  - Selon certains caractères : conditions individuelles d'accès ou d'usage des modes, niveau d'information
    => *Traitement par une segmentation discrète*
  - Selon les préférences : arbitrages entre les caractères des options
    => *Traitement par une distribution statistique des préférences*
- Modèle prix-temps
  - distribution de l'arbitrage prix-temps

Modèle bicritère prix-temps

- Hypothèses relatives aux services
  - L'option *m* est caractérisée en prix et temps, $X_m = (P_m, T_m)$
- Hypothèses relatives à la demande
  - Chaque demandeur a un arbitrage prix-temps α (en €/h ou autre unité)
  - Distribution des arbitrages α dans la population des décideurs, avec une fonction de répartition cumulée A(α)
  - Un demandeur d'arbitrage α perçoit pour l'option *m* un coût généralisé de $G_m(\alpha) = P_m + \alpha T_m$
- Principe du modèle
  - Chaque demandeur est affecté à une option de coût minimal pour lui
  - Les choix individuels induisent l'usage de chaque option

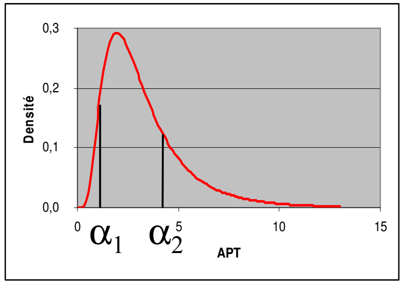
Comparaison des options

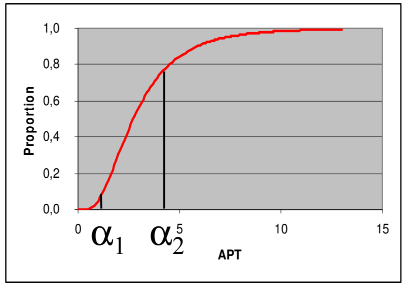


- Si $\alpha \leq \alpha^*$: option 1 meilleure pour α
- Si $\alpha > \alpha^*$: option 2 meilleure pour α

Les proportions d'usage

- Chaque option reçoit les clients qui la préfèrent
- Option *m* optimale sur $]x, y]$ =>
  Proportion d'usage $\pi_m = A(y) - A(x)$





Applications du modèle prix-temps

- Concurrences modales
  - Train classique et TGV
  - TGV et avion
  - Transport interurbain de fret
- Niveau de confort
  - Train 1ère et 2ème classe, si arbitrage selon niveau de confort
- Choix d'itinéraire
  - Réseau urbain ou interurbain, en présence de péages
- Liaison entre revenu et APT
  - Modèle prix-temps-revenu (Roger Marche, Claude Abraham, Jean-Didier Blanchet)

Exercice

- Hypothèses
  - Mode 1 : prix 10 €, temps 2 h
  - Mode 2 : prix 20 €, temps 1 h 30
  - Distribution des APT : uniforme sur [5, 25] €/h
- Questions
  - Arbitrage de coupure ?
  - Parts de marché ?
  - Quelles conséquences, si on fait varier P_2 de 0 à 50 € ?
- Solution
  - APT* = (20-10)/(2-1.5) = 20 €/h
  - Part de marché : le mode 2 est meilleur pour les APT > APT*
    donc $\pi_2 = 1 - A(20) = ¼$

Utilité aléatoire

- Représentation
  - Aléas ω : circonstances particulières que l'analyste traite de manière statistique
  - Pour l'option *m* et le décideur *u* : utilité aléatoire $U_m(u,\omega)$
  - Pour *u*, l'option *m* est meilleure dans une proportion des cas

$$\pi_m(u) = \Pr\{\omega : U_m(u,\omega) \geq U_r(u,\omega)\ \forall r\}$$

- Principe de choix
  - Dans chaque cas ω, le client choisit l'option d'utilité maximale => Rationalité économique
  - Donc *u* choisit l'option *m* avec la proportion $\pi_m(u)$

Choix discret de la zone de destination

- Depuis une origine *o*, une destination *d* présente une utilité (aléatoire) $U_{od} = S_d - G_{od} + \varepsilon_{od}$
  - S_d utilité brute
  - G_{od} coût généralisé
  - ε_{od} perturbation *aléatoire*
- Part modale

$$\Pr\{d \mid o\} = \Pr\{U_{od} \geq U_{od'} \mid \forall d'\}$$

- L'utilité moyenne de *d* sachant *o* est $V_{d|o} = S_d - G_{od}$

Choix entre 2 activités

ALEA / ACTIVITE	PLUIE	BEAU TEMPS
Promenade en forêt	1	7
Séance de cinéma	5	5

- Dans chaque case = utilité de l'activité pour le client, selon l'état météo

Analyse du choix

- Résultats
  - Cas de pluie : le décideur choisit le cinéma car 5 > 1
  - Cas de beau temps : le décideur choisit la forêt car 7 > 5
  - Répartition probabiliste due aux circonstances. Si 80% de beau temps, alors 80% forêt et 20% cinéma. L'utilité espérée moyenne est 80%.7 + 20%.5 = 6.6
- Formulation avec les notations génériques
  - Forêt : utilité moyenne = 6, perturbation ε = +1 si beau temps ou -5 si pluie
  - Cinéma : utilité moyenne = 5, perturbation ε = 0
- *Sur le traitement probabiliste*
  - *On oublie le détail des causes, afin de simplifier la description*
  - *On conserve et on privilégie, les effets avec leur probabilité d'occurrence*
- *Dans l'exemple : choix fait à 80% et 20%*

Principes fondamentaux

- La fonction d'utilité représente l'option
  - $U_m(u,\omega)$ pour le décideur *u*, aussi pour l'option *m* relativement aux autres
  - Dépendances statistiques entre les options ?
- Proportion de choix

  $\pi_m(u) = \Pr\{\omega : U_m(u,\omega) \geq U_r(u,\omega)\ \forall r\}$
- Fonction de satisfaction pour l'utilité moyenne
  - Le choix donne une utilité maximale au décideur *u*

    $U_M(u,\omega) = \max\{U_m(u,\omega)\ :\ m \in M\}$
  - Fonction de satisfaction $V_M(u) = \mathrm{E}_\omega[U_M(u,\omega)]$
- Élasticité(s)
  - Si les U_m dépendent de *Z* alors

    $e_{V_M:Z} = \dfrac{\partial \ln V_M}{\partial \ln Z} = \sum_{m \in M} \pi_m . \dfrac{\partial \ln V_m}{\partial \ln Z}$

Le cas indépendant

- Hypothèse
  - Les $U_m(u,\omega)$, pour les options *m* de *M*, sont indépendantes
- Conséquences
  - Pour décrire statistiquement la distribution conjointe des U_m , il suffit de connaître les fonctions de répartition F_m
  - Proportion de choix d'une option

$$\begin{aligned}\pi_m(u) &= \Pr\{U_m > U_r\,,\ \forall r \neq m\}\\ &= \int \Pr\{x > U_r\,,\ \forall r \neq m \mid U_m = x\}.\Pr\{U_m = x\}\\ &= \int \Pr\{x > U_r, \forall r \neq m\}\,\mathrm{d}F_m(x) \text{ en cas d'indépendance}\\ &= \int [\textstyle\prod_{r \neq m} \Pr\{x > U_r\}]\,\mathrm{d}F_m(x) \text{ en cas d'indépendance}\\ &= \int [\textstyle\prod_{r \neq m} 1 - F_r(x)]\,\mathrm{d}F_m(x)\end{aligned}$$

- Applications
  - Modèle logit
  - Cf information prévisionnelle en affectation TC

Fonction d'utilité additive

- Formule additive $U_m(u,\omega) = V_m(u) + \varepsilon_{um}(\omega)$
  - $V_m(u)$ partie déterministe
  - $\varepsilon_{um}(\omega)$ perturbation aléatoire
- Sur la partie déterministe
  - Typiquement, $V_m(u) = -G_m(u)$
  - En première approche, $V_m(u) = \beta_m^0 - \sum_k \beta_m^k . X_m^k$
  - Constante modale β_m^0
- Sur la perturbation aléatoire
  - A priori on centre chaque ε i.e. E[$\varepsilon_m(u,\omega)$] = 0
  - Dépendances statistiques ?
    On fait souvent l'hypothèse d'indépendance, qui simplifie grandement mais qui est très forte

Modèle binaire

- Proportions dans le cas binaire

$$\begin{aligned}p_1 &= \Pr\{U_1 \geq U_2\}\\ &= \Pr\{-G_1 + \varepsilon_1 \geq -G_2 + \varepsilon_2\}\\ &= \Pr\{\varepsilon_2 - \varepsilon_1 \leq G_2 - G_1\}\\ &= F_{\Delta\varepsilon}(G_2 - G_1)\end{aligned}$$

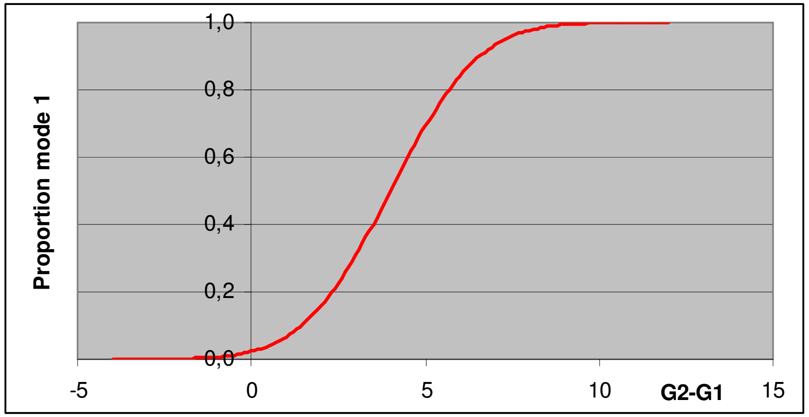
  avec $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$ et sa fonction de répartition $F_{\Delta\varepsilon}$
- C'est un modèle formel de répartition
  - Avec une bonne interprétation économique : si ΔG croît alors la proportion du mode moins cher augmente
  - Facile à traiter

Le modèle probit additif

- Hypothèses
  - Utilité additive $U = V + \varepsilon$
  - La perturbation $[\varepsilon_{um}]_{m \in M}$ est gaussienne multivariée, avec une moyenne μ et une matrice de covariance $\mathbf{C} = [v_{mr}]_{m,r \in M}$ avec $v_{mm} = \sigma_m^2$
- Modèle probit binaire
  - Alors $\Delta\varepsilon \approx \mathrm{N}(\mu_2 - \mu_1, \sigma_{\Delta\varepsilon}^2)$ avec $\sigma_{\Delta\varepsilon}^2 = \sigma_2^2 + \sigma_1^2 - 2v_{12}$
  - Cas indépendant $v_{12} = 0$
  - Formule de répartition $\mathrm{F}_{\Delta\varepsilon}(x) = \Phi\left(\frac{x - (\mu_2 - \mu_1)}{\sigma_{\Delta\varepsilon}}\right)$

La fonction Φ, fonction de répartition d'une variable gaussienne centrée et normée, est bien connue des statisticiens. Elle est disponible dans les tableurs et logiciels scientifiques
Sous Excel, fonction LOI.NORMALE

Fonction de réponse du probit



Modèle logit multinomial

- Hypothèses
  - Utilité additive $U_m(u,\omega) = V_m(u) + \varepsilon_{um}(\omega)$
  - Les perturbations aléatoires ε_{um} sont indépendantes et identiquement distribuées (i.i.d.) : loi de Gumbel de paramètre θ
  - Fonction de répartition dite « double exponentielle »,

    $$F_{G\theta}(x) = \exp[-e^{\gamma} e^{-\theta x}]$$

    avec γ # 0.5772… la constante d'Euler
  - Variance $\sigma^2 = \pi^2 / (6\theta^2)$
- Intérêt
  - Cette famille de variables aléatoires est **stable** pour l'opération « prendre le maximum »

Logit multinomial : propriétés

- Proportions des options

  $$\pi_m = \frac{\exp[\theta V_m]}{\exp[\theta V_M]} \text{ avec } \exp[\theta V_M] = \sum_{m \in M} \exp[\theta V_m]$$
- L'utilité maximale $U_M(u,\omega) = \max\{U_m(u,\omega) : m \in M\}$
  - c'est une VA de Gumbel de variance σ² (donc paramètre θ) et de moyenne la fonction de satisfaction $V_M(u)$
- Fonction de satisfaction : formule du log-somme

  $$V_M(u) = \frac{1}{\theta} \ln \sum_{m \in M} \exp[\theta V_m(u)]$$
- Cas binaire
  - $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$ a pour fonction de répartition

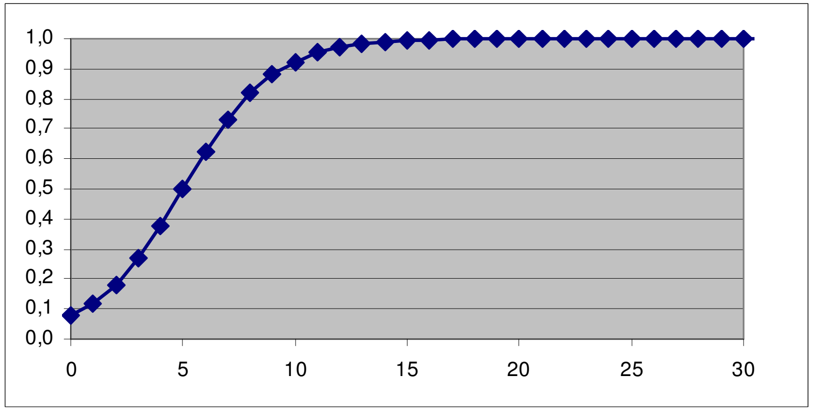
    $$\begin{aligned} \Pr\{\Delta\varepsilon = \varepsilon_2 - \varepsilon_1 \le x\} &= \Pr\{\varepsilon_2 - \varepsilon_1 \le V_1 - V_2\} = \Pr\{V_2 + \varepsilon_2 \le V_1 + \varepsilon_1\} \\ &= \pi_1(x) \text{ pour } x = V_1 - V_2 \\ &= 1/[1 + \exp(-\theta x)] \end{aligned}$$

    $\Delta\varepsilon$ est une variable aléatoire logistique

Exercice

- Hypothèses
  - Mode 1, prix 10 €, temps 2 h
  - Mode 2, prix P_2 =20 €, temps 1 h 30
  - APT = 10 €/h
- Questions
  - Part modale du mode 1 si θ = 1 /€ ?
  - Part modale du mode 1 si θ = 2 /€ ?
  - Pour θ = .5 /€, tracer π_1 en fonction de P_2 de 0 à 50 €

Réponse du logit



Conclusion provisoire

- Domaine très large d'application
  - En cas d'indépendance, équivaut pratiquement au probit
  - Calculs simples
  - Estimation simple
- Extension
  - Captivités
  - Non-linéarités
  - Distribution des coefficients
  - Structure de choix

Surplus de la demande

- Fonction de satisfaction

$$U_M(\omega) = \max_m U_m(\omega)$$
$$V_M = \mathrm{E}_\omega[U_M]$$

Formule du log-somme dans le cas logit

- La fonction de satisfaction mesure le surplus du consommateur, à une constante près
- Applications : modification d'une option, et aussi introduction d'une nouvelle option

Tunnel Prado-Carénage : contexte

- Centre-ville de Marseille
  - Tunnel de 2.5 km de long, ouvert en septembre 1993
  - Gain de temps médian : 11 minutes
  - 1.5 € par passage, en 1995
- Une expérience innovante
  - Première expérience française d'autoroute urbaine à péage, réservée aux véhicules légers
  - Objectif = estimer la distribution des APT
- Dispositif d'étude
  - Enquêtes de Préférences Déclarées
  - Préférences Révélées : enquêtes aux relations O-D sur les itinéraires concurrents, d'une part le tunnel, d'autre part en surface
  - Relations O-D regroupées par tranche de ΔDistance et de ΔTemps entre tunnel et surface

Modèle prix-temps stochastique

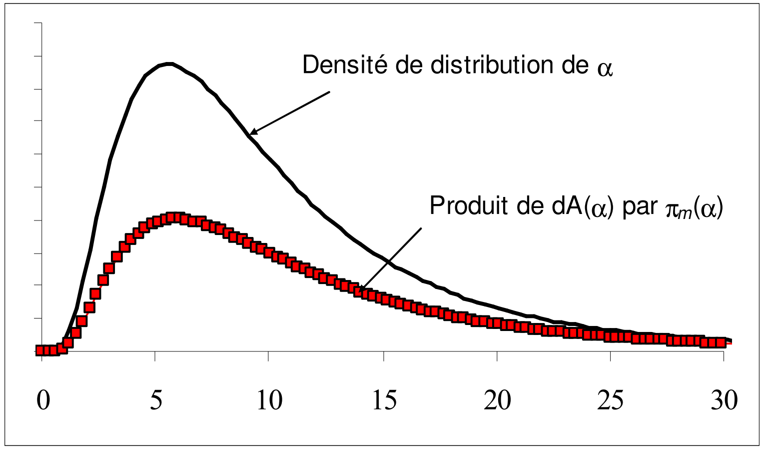
- Univers de choix : 2 options, Tunnel ou Surface
- Fonction d'utilité
  - P prix du péage, si en tunnel
  - Temps x APT
  - Perturbation aléatoire (nulle, ou logit, ou probit)
  - Distribution des APT (constante, ou log-normale, ou log-logistique, ou uniforme)
- Proportion d'usage du tunnel
  - A(α) fonction de répartition des APT

$$\pi_T = \int_\alpha \pi_T(\alpha)\, dA(\alpha)$$

  - Si perturbations logit,

$$\pi_T(\alpha) = \Pr\{\omega : U_T(\alpha,\omega) > U_S(\alpha,\omega)\} = \frac{1}{1+\exp[\theta(P+\alpha(T_T - T_S))]}$$

Résultats



Modèles à coefficients distribués

- Croisement entre les perturbations aléatoires et la distribution des décideurs

$$\Pr\{U_m \ge U_\ell\} = \int_\alpha \Pr\{U_{\alpha m} \ge U_{\alpha \ell} \mid \alpha\}\, d\Pr\{\alpha\}$$

- Mixed logit : perturbation logit et paramètres distribués dans le coût généralisé
- Cas d'une distribution d'APT uniforme sur [*a*, *b*]

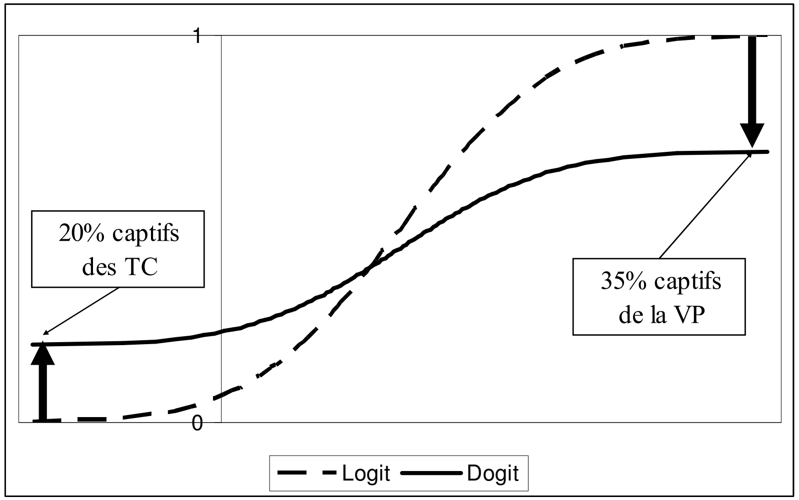
$$p_1 = 1 - \frac{\pi}{\sqrt{3}} \frac{\sigma_\Delta}{(b-a)\Delta T} \ln \frac{1+\exp\left(-\frac{\pi}{\sqrt{3}} \frac{\Delta P + a\Delta T - \Delta\mu}{\sigma_\Delta}\right)}{1+\exp\left(-\frac{\pi}{\sqrt{3}} \frac{\Delta P + b\Delta T - \Delta\mu}{\sigma_\Delta}\right)}$$

Captivités

- Définitions : exclusion / captivité
  - Exclusion : quand une option est indisponible pour un décideur
  - Captivité : vis-à-vis des options non-exclues
- Captivités en transport urbain de voyageurs
  - TC si distances longues et non motorisé
  - VP si distances longues et TC impraticables
- Modèle dogit
  - Proportion η_m imposée pour les captifs de l'option *m*
  - Le reliquat $1-\overline{\eta}$ avec $\overline{\eta}=\sum_m \eta_m$ est en situation de choix
  - Proportion d'usage

$$\pi_m = \eta_m + (1-\overline{\eta})\frac{\exp(\theta V_m)}{\exp(\theta V_M)}$$

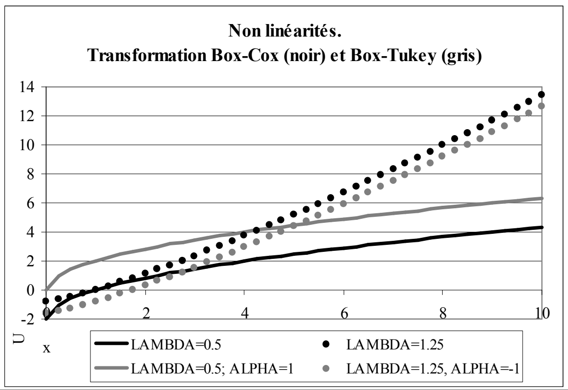
Fonction de réponse d'un dogit



Non-linéarités « élastiques »

- Moduler des élasticités dans la fonction d'utilité
  - modèles de Box-Cox et Box-Tukey

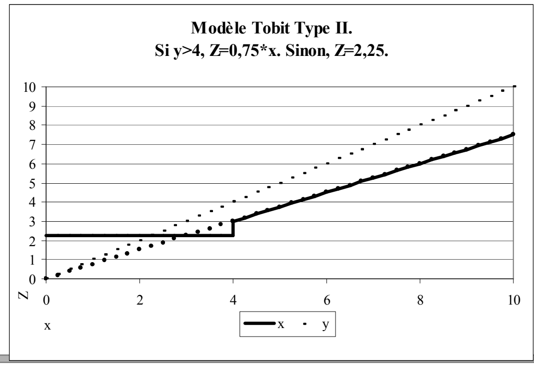
$V_m = -\sum_k \theta_k B_k[X_{mk}]$ avec $B_k[x]=\dfrac{(x-\alpha_k)^{\lambda_k}-1}{\lambda_k}$



Non-linéarités « abruptes »

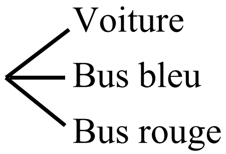
- Modèles Tobit : conditionner la fonction d'utilité par une variable

Type I $z(x)=\begin{cases} f(x) & \text{si } x>k \\ k & \text{sinon}\end{cases}$ Type II $z(x|y)=\begin{cases} f(x) & \text{si } y>k \\ c & \text{sinon}\end{cases}$



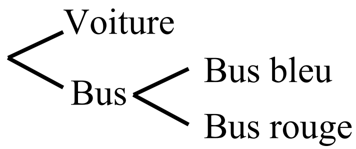
Structure de choix

- Motivation : relation entre les options



- Repeindre certains bus ne crée pas un nouveau mode !
- Remarque : pour une différence plus significative, comparer Bus Propre contre Bus Sale

Solution : nested logit



- Un modèle de choix par étage
  - Choix du type de bus, conditionnel au choix du *bus composite*
  - Mode *bus composite* : les caractères des deux lignes sont pris en compte au travers de l'utilité composite
  - Relation entre les étages : $\theta_1 \leq \theta_2$ car variance plus grande au premier embranchement
- Nested logit = logit emboîté, ou logit à étages

Conclusion

- Représentation du comportement individuel
- Cadre flexible et puissant ; influence des caractères du décideur et des options
- Modèles simples
  - Prix-temps, logit, probit
  - Simplicité du traitement formel
  - Risque de caractère réducteur
- Modèles sophistiqués
  - par adaptation, métissage, ou par structuration de modèles simples
  - Mixed logit
  - Modèle dogit
  - Nested logit
- Applications très larges