

Modélisation de la demande de transport

5

La distribution spatiale du trafic

Fabien Leurent
ENPC / LVMT

Modélisation de la demande de transport

Distribution spatiale

Problèmes de distribution

- Distribution des établissements, des activités
 - Localisation d'un établissement, d'une activité
 - Répartition d'établissements (ou d'activités) d'une certaine nature ou de plusieurs natures complémentaires, sur un territoire
- Distribution spatiale du trafic
 - Flux par relation O-D = un transfert, un échange
 - Distribution d'un flux global, entre des relations O-D
 - À origine fixée, choix de destination
- Distribution du fret
 - Transport de sources à puits
 - Problème de tournée

Modélisation de la demande de transport

Distribution spatiale

Objectif (et plan)

- Modèles empiriques de distribution : « mesurer » les flux O-D
 - Traitement par O-D
 - Méthodes d'observation
 - Inférence d'une matrice O-D par maximum d'entropie
 - Estimation statistique
- Modèles économiques : « expliquer » les flux O-D
 - Modèle gravitaire
 - Choix discret
 - Modèle des opportunités
 - Choix d'activité, et accessibilité
 - Équilibre spatial
 - Modèles input-output

Modélisation de la demande de transport

Distribution spatiale

Relation O-D

- Chaque Dt a un lieu initial : son origine ; et un lieu final : sa destination
- Relation O-D = association d'une zone d'origine et d'une zone de destination
- Matrice O-D = ensemble de relations O-D
 - Pour zonage origine et zonage destination à préciser
 - Motif, horaire, mode...
 - Variables : flux, coûts, temps, paramètres de fonctions...

Modélisation de la demande de transport

Distribution spatiale

Matrices O-D

- Exemples
 - Flux : tableau import-export, en valeur, en tonnage, par type de produit
 - Distancier
- Représentations
 - Tableau
 - Lignes de désir : carte des zones, avec par couple de centroïdes un segment d'épaisseur proportionnelle au flux O-D

Exemple de tableau O-D

O \ D	1	2	3	4	5	6	7	8	9	10	11	12
1	1	3	460	91	65	960	21	440	96	340	527	16
2	130	1	78	29	44	7	108	69	12	89	11	16
3	3750	49	1	1666	396	616	0	479	9857	513	799	69
4	22	227	1636	1	809	836	0	1731	299	2952	193	225
5	122	161	359	249	1	62	104	7	350	0	15	0
6	463	18	4316	2124	38	1	917	7276	73	1272	69	26
7	1762	30	1923	746	136	889	1	10	2764	175	341	4
8	404	31	643	62	192	10695	124	1	3	304	1375	3
9	29	1	7568	812	1755	1	819	1771	1	1190	1096	354
10	270	46	0	0	0	77	1070	506	0	1	417	0
11	201	49	779	700	2	6	15	588	2488	213	1	87
12	0	0	0	254	0	0	0	271	0	0	162	1

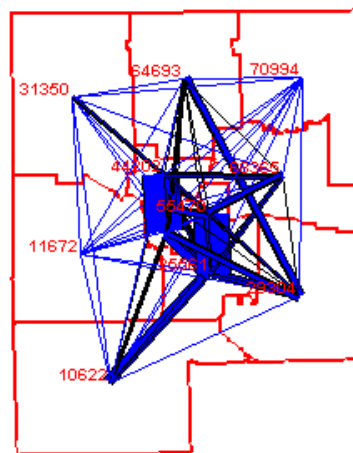
Modélisation de la demande de transport

Distribution spatiale

Modélisation de la demande de transport

Distribution spatiale

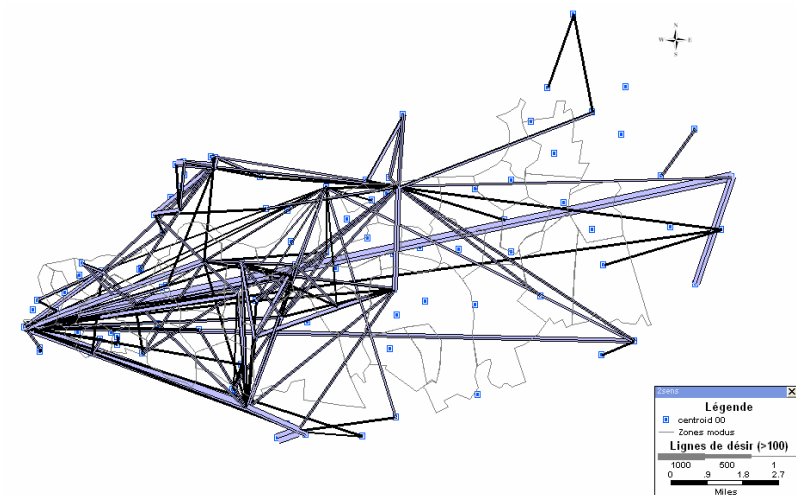
Lignes de désir



Modélisation de la demande de transport

Distribution spatiale

Lignes de désir des déplacements routiers à Marne la Vallée



Modélisation de la demande de transport

Distribution spatiale

Analyse par O-D

- O-D importantes en transport de "mobiles"
 - Comme en télécommunication = transport de message
 - Différent de l'électricité, de l'hydraulique
- Dimension des problèmes de choix
 - Nombre de relations O-D, fois le nombre de classes d'utilisateurs
- Conséquences informatiques
 - Volume de calcul multiplié par le nb d'O-D
 - Algorithmes pour traiter génériquement, collectivement, toutes les destinations depuis une origine

Observer les flux O-D

- Sur le réseau
 - Comptages sur les axes
 - Enquêtes aux O-D
- Auprès des ménages
 - Enquêtes-ménages
- Cas des entreprises
 - Douanes ; base de données SITRAM
- Nouvelles technologies : téléphonie mobile

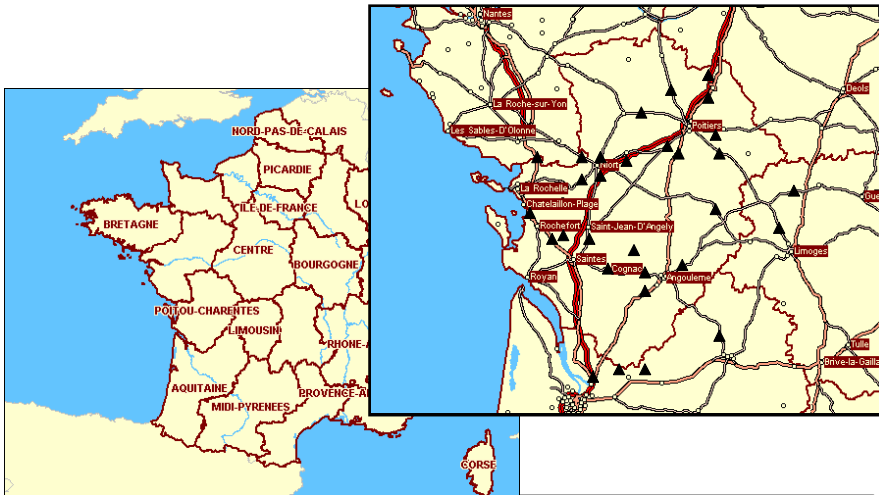
Comptage sur un axe

- Circonstances
 - Période : quelques heures ou jours, ou année
 - Capteur sur l'axe, ou billetterie
- Information recueillie
 - Nb de personnes, ou véhicules, ou paires d'essieux
 - Éventuellement, nb par classe
 - Répartition temporelle
 - Éventuellement, vitesse

Enquête aux O-D

- Circonstances
 - Période, heure / jour
 - Lieu : mode routier : enquête en bord de route
 - Lieu : mode TC : enquête à quai ou embarquée
- Interview
 - Usager : sexe, âge, profession, résidence
 - Déplacement : instant, origine, destination, motif O-D
 - Longueur et durée selon le mode : si TC en interurbain, questions sur les habitudes et les préférences

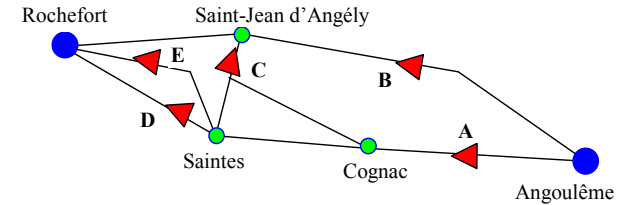
Enquêtes en région Poitou-Charentes



Modélisation de la demande de transport

Distribution spatiale

O-D Angoulême-Rochefort



Axe	Nb jours comptage	$E[X_a]$	$S^*[X_a]$	Nb interviews	Dont sur l'O-D i
A	15	10,029	3,824	1,332	11
B	5	3,739	1,260	676	13
C	10	7,107	2,720	1,243	0
D	20	9,735	3,645	1,388	14
E	12	5,736	3,400	1,554	14

Modélisation de la demande de transport

Distribution spatiale

Interprétation statistique

- Questions
 - Plusieurs estimateurs ! Lequel choisir ?
 - Fiabilité de l'estimateur ?
- Traitement
 - Chaque enquête est une source d'information imprécise: le résultat est une variable aléatoire, avec une estimation de la moyenne et de la variance
 - Mélanger les moyennes et les imprécisions, en parallèle et en série
 - Attention aux corrélations

Modélisation de la demande de transport

Distribution spatiale

Méthode 1/5 : analyse d'une source

- Cas type : observation d'une variable aléatoire X
- *Donner une image : l'estimateur*
 - Tirer un échantillon : valeurs $(X_k)_{k=1..K}$
 - Estimer la moyenne : moyenne d'échantillon

$$\hat{X} = \frac{1}{K} \sum_{k=1}^K X_k$$
 - Estimer la variance : variance d'échantillon

$$S^{*2} = \frac{1}{K-1} \sum_{k=1}^K (X_k - \hat{X})^2$$
- *L'estimateur est-il fidèle ? Quel biais ?*
 - Fidélité si, en moyenne sur tous les échantillons possibles, l'estimateur est égal à la grandeur observée (i.e. biais nul)

Modélisation de la demande de transport

Distribution spatiale

Méthode, 2/5 : imprécision de l'estimateur

- *L'imprécision de l'estimateur*
 - elle se mesure par la variance de l'estimateur (sur tous les échantillons possibles), notée SE^2
 - Intervalle de confiance $\hat{X} \pm 2SE_{\hat{X}}$ à 95%
- *Imprécision d'une moyenne d'échantillon*
 - Pour l'estimateur d'une moyenne d'échantillon

$$SE_{\hat{X}}^2 = V[\hat{X}] = V[X] / K$$
 - Estimé avec la variance d'échantillon

$$SE_{\hat{X}}^2 \approx S^{*2} / K$$

Méthode, 3/5 : estimateur-produit

- *Le produit est une combinaison en série*
 - En cas d'indépendance de X et Y
 - Moyenne $E[XY] = E[X]E[Y]$
 - Variance $V[XY] = V[X]V[Y] + E[X]^2 V[Y] + E[Y]^2 V[X]$
- *Application à l'estimateur « Flux d'axe et O-D »*
 - Variable X estime le flux X_a de l'axe a
 - Variable Y estime la proportion $p_{i/a}$ de l'O-D i
 - XY estime $Q_{ia} = X_a p_{i/a}$ le flux de $i \cap a$

Application, 1/3

	Moy(X)	SE(X)	Moy(p)	SE(p _{ia})	Moy(Q _{ia})	SE(Q _{ia})
A	10 029	987,4	0,83%	0,25%	82,8	26,3
B	3 739	563,5	1,92%	0,53%	71,9	22,7
C	7 107	860,1	0,00%	0,00%	0,0	0,0
D	9 735	815,0	1,01%	0,27%	98,2	27,5
E	5 706	981,5	0,90%	0,24%	51,4	16,5

Une proportion est la moyenne d'une variable 0-1 :
valeur 1 si Oui, 0 si Non

Méthode, 4/5 : estimateur-somme

- *La somme est une combinaison en parallèle*
 - Moyenne $E[\lambda X + \mu Y] = \lambda E[X] + \mu E[Y]$
 - Variance $\text{var}[\lambda X + \mu Y] = \lambda^2 \text{var}[X] + \mu^2 \text{var}[Y] + 2\lambda\mu \text{cov}[X, Y]$
 - En cas d'indépendance de X et Y , covariance nulle
- *Application à l'estimateur « Flux O-D sur faisceau »*
 - La moyenne de la somme, est la somme des moyennes
 - Indépendance donc la variance de la somme, est la somme des variances

Application, 2/3

Combinaison	Moy(Flux)	SE(Flux)
C+D+E	149,6	32,0
CDE + B	221,5	39,3
A + B	154,7	34,7

- Somme intermédiaire : postes C, D, E
- 1er flux total de l'O-D : postes C,D,E,B
- 2ème flux total de l'O-D : postes A, B

Méthode, 5/5 : Mélange (bayésien) d'indicateurs

- Le mélange de deux estimateurs
 - Pour utiliser des observations redondantes
 - Ici, cas de deux estimateurs en dimension 1 et indépendants

- Imprécision

$$1/SE_{X \otimes Y}^2 = 1/SE_X^2 + 1/SE_Y^2$$

- Moyenne

$$E[X \otimes Y] = SE_{X \otimes Y}^2 \left(\frac{E[X]}{SE_X^2} + \frac{E[Y]}{SE_Y^2} \right)$$

Application, 3/3

Combinaison	Moy(Flux)	SE(Flux)
CDE avec A	109,7	20,3
B + Mélange	181,6	30,5

- Mélange de A et de C+D+E
- Somme de B avec le mélange
 - L'imprécision, 30.5^2 est inférieure à celles de chaque estimateur : 34.7^2 pour A+B et 39^2 pour C+D+E+B

Coûts des enquêtes O-D

- Postes de coûts
 - Conception du questionnaire (fixe)
 - Administration (à rapporter à la durée d'enquête = débit x taux de sondage x durée d'interview)
- Coûts d'exploitation
 - Saisie informatique x nb interview x durée saisie
 - Calcul des flux O-D : quelle automatisation ?
- Optimiser les coûts
 - Optimiser le schéma d'enquête
 - Utilité de l'information antérieure

Enquêtes dérivées

- Enquête cordon
 - Pour un générateur spécifique, ou autour d'une agglo
 - Enquête papillon
 - Sur certains axes, des enquêteurs collent (en entrée) ou relèvent (en sortie) des papiers colorés suivant l'axe d'entrée
 - Connaissance grossière du transit et des échanges
- Relevé de plaques minéralogiques
 - Pour cordon, ou pour mesurer des temps de parcours
- Téléphonie mobile
 - Articles de recherche, étude DRE Limousin

Sur les enquêtes-ménages

- Marges des zones
 - Émissions, réceptions
- O-D par grandes zones
 - Taille de zone pour assurer 30 Dt minimum par relation
- Répartition par tranches
 - Tranches de distance, de temps, de coût généralisé
 - Application pour estimer des paramètres
 - À utiliser avec un modèle d'affectation des Dt sur réseau

Cas du fret

- Enquêtes aux douanes : Cf enquête-cordon
 - Flux par nature de chargement ; nationalité du véhicule
- Système d'Information TRANsport de Marchandises
 - Rail, voie navigable : exhaustif
 - Route : 1 poids lourd sur 10, une semaine par an : tous mouvements et chargements
 - D'où O-D entre départements, en tonnage, par nature de chargement, par mouvement et par type de véhicule
- Enquêtes auprès des chargeurs (dont ECHO)
 - Cf enquêtes ménages, mais sur les envois d'entreprises

Conclusions sur l'observation

- Traitement par relation origine-destination
- Plusieurs types d'observations
 - Pour chaque type : son résultat, sa procédure de recueil, son protocole de traitement, son coût
- Exploitation statistique
 - il faut quelques notions statistiques de base
- Liaison à la modélisation
 - Mesure de la matrice O-D : modèle empirique pour généraliser à partir d'observations partielles
 - Ou modèle économique de distribution

Inférence d'une matrice O-D

- Introduction
- Méthode des facteurs de croissance
 - Algorithme bi-proportionnel *de balancement*
- Le principe du maximum d'entropie
 - Entropie d'une matrice O-D
 - Analyse des conditions d'optimalité
- Algorithme multi-proportionnel
 - Tirer parti de diverses contraintes d'information
- *L'entropie généralisée (numérique)*

Estimation versus Inférence

- Définition de l'estimation
 - Fixer les valeurs de paramètres dans un problème statistique bien posé : cadre probabiliste clair, davantage d'observations que de paramètres, et *identifiabilité*
- Définition de l'inférence
 - Choisir une valeur vraisemblable compte tenu des informations disponibles
 - Parmi plusieurs valeurs vraisemblables possibles : un principe de sélection, le maximum d'entropie

Méthode des facteurs de croissance

- Résultat
 - Matrice O-D $[q_i]_{i \in I}$ compatible avec des marges d'émission et de réception fixées
- Inputs
 - Matrice O-D préalable, ancienne : $[g_i]$ En l'absence de références, prendre des flux =1 ou modélisés $D_i(G_i)$
 - Marges émission $\sum_{i \in o} q_i = E_o$
 - Et/ou marges réception $\sum_{i \in d} q_i = R_d$

Traitement par normalisation

- Marges en une seule dimension : normalisation
$$q_i \leftarrow \frac{E_o}{\sum_{j \in I} g_j} \cdot g_i$$
 Contraintes d'émission aux origines : par origine o , on traite chaque relation $i = (o, d)$ en actualisant par un facteur E_{new}/E_{old}
- Marges en 2 dimensions : algorithme bi-proportionnel
 - Par itération k , traitement avec 2 étapes $[2k+1]$ et $[2k+2]$
 - **Début** : poser $k \leftarrow 0$ et $q_i^{[0]} \leftarrow g_i$ pour toute i
 - Pour chaque origine o , pour tout $i = (o, d)$, poser $q_i^{[2k+1]} \leftarrow \frac{E_o}{\sum_{i \in o} q_i^{[2k]}} \cdot q_i^{[2k]}$
 - Pour chaque destination d , pour tout $i = (o, d)$, poser $q_i^{[2k+2]} \leftarrow \frac{R_d}{\sum_{i \in d} q_i^{[2k+1]}} \cdot q_i^{[2k+1]}$
 - Itérer jusqu'à **convergence**

Exemple

- 3 zones de demande
- Problème avec marges en émission et réception

$$\begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix} \begin{bmatrix} 460 \\ 384 \\ 311 \end{bmatrix}$$
$$[368 \quad 533 \quad 254]$$

- Connaissance antérieure $\begin{bmatrix} 107 & 160 & 100 \\ 160 & 210 & 107 \\ 88 & 123 & 100 \end{bmatrix}$

Convergence

- Résultat après 3 itérations : précision de 5 chiffres significatifs

$$\begin{bmatrix} 140.77 & 217.13 & 102.10 \\ 133.66 & 180.96 & 69.37 \\ 93.57 & 134.91 & 82.52 \end{bmatrix} \begin{bmatrix} 460 \\ 384 \\ 311 \end{bmatrix}$$
$$[368 \quad 533 \quad 254]$$

Questions

- Justification technique
 - Ça marche bien !
 - Ça marche vite !
 - Et c'est simple !
 - Justification scientifique ?
 - Économique ?
 - Statistique ?
- ⇒ Un **principe d'inférence** : le maximum d'entropie
- ⇒ *Principe dû à Claude Shannon (1951)*
 - ⇒ *Matrices O-D : Murchland (1966), Wilson (1967)*

Modèles entropiques

- Entropie simple
- Entropie croisée
- Observations pour une matrice O-D
- Algorithme multi-proportionnel
- Entropie généralisée (numérique)

L'entropie simple

- Principe
 - Entropie = mesure du **désordre au niveau microscopique**
 - Pour un système composé d'éléments microscopiques, un état macroscopique donné peut correspondre à divers états microscopiques
 - L'entropie simple mesure le nb d'états micro compatibles avec un certain état macro, à un log et une constante près

Evaluer l'entropie

- Répartition de N boules dans I urnes
 - Etat macroscopique = nb de boules n_i dans urne i
 - Nombre de configurations microscopiques compatibles = nb de tirages sans remise, de n_1 boules dans N , puis de n_2 boules dans $N - n_1$, ... enfin de n_I dans $N - \sum_{i=1}^{I-1} n_i$
 - Il y en a
$$\binom{N}{n_1} \binom{N-n_1}{n_2} \dots \binom{N-\sum_{i=1}^{I-1} n_i}{n_I} = \frac{N!}{\prod_{i=1}^I n_i!}$$
- Formule de Stirling $\ln(x!) \approx x \cdot [\ln(x) - 1]$
$$\ln \frac{N!}{\prod_{i=1}^I n_i!} = \ln(N!) - \sum_{i=1}^I \ln(n_i!) \approx N \cdot [\ln(N) - 1] - \sum_{i=1}^I n_i \cdot [\ln(n_i) - 1]$$
$$= N \ln(N) - \sum_{i=1}^I n_i \cdot \ln(n_i) = -N \cdot \sum_{i=1}^I \frac{n_i}{N} \cdot \ln\left(\frac{n_i}{N}\right)$$

Principe du maximum d'entropie

- Principe du maximum d'entropie
 - Choisir la répartition macro, compatible avec les contraintes macro, qui maximise le **désordre au niveau microscopique** (en nombre de configurations micro admissibles)
- Formulation de base
 - I cases i , proportions p_i à déterminer
 - Contraintes d'information, linéaires $V_a = \sum_i \gamma_{ai} p_i$
 - Fonction d'entropie $H(p) = -\sum_{i \in I} p_i \ln(p_i)$
 - Résoudre $\max_p H(p)$ sous $p \geq 0$, $\sum p_i = 1$, $V = \Gamma p$
avec $p = [p_i]$, $u = [1_i]$, $\Gamma = [\gamma_{ai}]$, $V = [v_a]$

L'entropie croisée

- Mesure le supplément d'ordre dans une répartition $[p_i]_{i \in I}$ par rapport à une répartition de référence $[\rho_i]_{i \in I}$
- Fonction d'entropie croisée
$$H^\times(p) = \sum_{i \in I} p_i \ln(p_i / \rho_i)$$
- Principe du minimum d'entropie croisée
 - minimiser le supplément d'ordre, sous les contraintes macro
 - Problème $\min_p H^\times(p)$ sous les contraintes précédentes
- Si $\rho_i = \frac{1}{I}$, alors entropie simple \Leftrightarrow croisée

Entropie d'une matrice O-D

- Fonction d'entropie croisée
 - Proportions $p_i = \frac{q_i}{N_q}$ et $\rho_i = \frac{g_i}{N_g}$
 - Fonction $H^\times(p) = \sum_{i \in I} \frac{q_i}{N_q} [\ln(\frac{q_i}{g_i}) + \ln \frac{N_g}{N_q}]$
simplifiée en $J(q) = \sum_{i \in I} q_i \ln(q_i / g_i)$
- Matrice de référence
 - Ancienne matrice de proportions ρ_i
 - Ou modèle, avec $\rho_i = Z(i) / \sum_{i \in I} Z(i)$
- Contraintes macroscopiques
 - Marges en émission, en réception des zones
 - Comptages locaux $V_a = \sum_i \gamma_{ai} \cdot q_i$

Analyse mathématique

- Objectif = exprimer des conditions d'optimalité pour soutenir la discussion économique
- Fonction objectif $J(q) = \sum_{i \in I} q_i \ln(q_i / g_i)$
- Contraintes

$$\left\{ \begin{array}{l} q_i \geq 0 \\ \sum_{i \in o} q_i = E_o \leftarrow \alpha_o \\ \sum_{i \in d} q_i = R_d \leftarrow \beta_d \\ \sum_i \gamma_{ai} q_i = V_a \leftarrow \lambda_a \end{array} \right\}$$
- Lagrangien $\mathcal{L}[q, \alpha, \beta, \lambda] \equiv J(q) - \alpha_o \cdot (E - \Delta_o q) - \beta_d \cdot (R - \Delta_d q) - \lambda_a \cdot (V - \Gamma q)$

Conditions d'optimalité

- Toute solution du problème d'optimisation précédent est caractérisée par les conditions suivantes (de Kuhn-Tucker) sur les variables primales et duales
- Conditions d'optimalité sur les variables primales

$$\begin{cases} \partial L / \partial q_i \geq 0 \\ q_i \cdot \partial L / \partial q_i = 0 \end{cases} \text{ donc } \ln\left(\frac{q_i}{g_i}\right) + 1 + \alpha_o + \beta_d + \sum_a \lambda_a \gamma_{ai} = 0$$
- Conditions d'optimalité sur les variables duales

$$\partial L / \partial \alpha_o = 0 \text{ donc } \sum_{i \in o} q_i = E_o$$

$$\partial L / \partial \beta_d = 0 \text{ donc } \sum_{i \in d} q_i = R_d$$

$$\partial L / \partial \lambda_a = 0 \text{ donc } \sum_i \gamma_{ai} q_i = V_a$$

Formules de la solution

$$q_i = g_i \exp[-1 - \alpha_o - \beta_d - \sum_a \lambda_a \gamma_{ai}]$$

$$E_o = \exp[-\alpha_o - 1] \cdot \sum_{i \in o} g_i \exp[-\beta_d - \sum_a \lambda_a \gamma_{ai}]$$

$$R_d = \exp[-\beta_d - 1] \cdot \sum_{i \in d} g_i \exp[-\alpha_o - \sum_a \lambda_a \gamma_{ai}]$$

$$V_a = \sum_i \gamma_{ai} g_i \exp[-1 - \alpha_o - \beta_d - \sum_a \lambda_a \gamma_{ai}]$$

- Remarquer le g_i en facteur
- Justifie l'algorithme multi-proportionnel

Algorithme multi-proportionnel

- Début $k \leftarrow 0, q_i^{[0]} \leftarrow \rho_i$
- Traiter les contraintes d'indice a ou o
 - $\forall a$ résoudre en λ_a : $\sum_i \gamma_{ai} q_i \exp[-\gamma_{ai} \lambda_a] = V_a$
 - et $\forall i$ actualiser $q_i \leftarrow q_i \exp[-\lambda_a \gamma_{ai}]$
- Destinations
 - $\forall d$ calculer $f_d \leftarrow R_d / \sum_{i \in d} q_i^{[2k+1]}$
 - et $\forall i \in d$ actualiser $q_i^{[2k+2]} \leftarrow q_i^{[2k+1]} f_d$
- Convergence ? Si pas atteinte, $k \leftarrow k + 1$

Suite de l'exemple

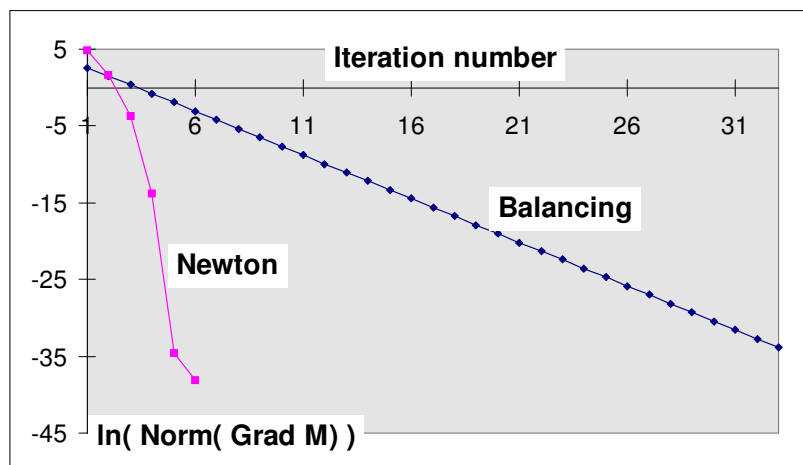
- On ajoute 2 classes de distances pour les O-D
 - Somme des cases en grisé = 214
 - Somme des autres cases = 941
- Résultats

143.00	216.26	100.74
132.01	180.99	71.00
92.98	135.75	82.26
368	533	254

460
384
311
- au lieu de

140.77	217.13	102.10
133.66	180.96	69.37
93.57	134.91	82.52

Convergence de 2 algorithmes



Entropie généralisée

- Pour un paramètre θ_ℓ
 - Points de support $z_{\ell i}$
 - Poids respectifs $p_{\ell i}$, avec connaissance a priori $g_{\ell i}$
 - Combinaison convexe $\theta_\ell = \sum_i p_{\ell i} z_{\ell i}$
- Entropie « numérique »

$$H^\times(p) = \sum_\ell \sum_i p_{\ell i} \ln(p_{\ell i} / g_{\ell i})$$
- Erreurs d'observation

$$\begin{cases} V_a = [\sum_\ell \gamma_{a\ell} \theta_\ell] + \varepsilon_a \\ \varepsilon_a = \sum_j p_{aj} z_{aj} \end{cases}$$
- L'entropie numérique a des **propriétés statistiques**

Conclusion sur l'inférence

- Des algorithmes efficaces
 - Avec des contraintes variées
- Le maximum d'entropie
 - Un principe puissant
 - Relativement intuitif
 - Propriétés statistiques de l'entropie généralisée, dans les cas avec plus d'observations que de variables
- La démarche d'inférence reste nécessaire quand il y a plus de variables que d'observations

État des logiciels

- Inférence par maximum d'entropie
 - Tous les grands logiciels offrent cette fonction, avec les marges d'émission et réception, et les comptages
 - Mais pas tous les logiciels pour les tranches de coût
- Estimation statistique
 - Emme/2 et TRIPS/CUBE sont bons (pas parfaits pour la spécification des erreurs)

Estimation statistique

- Approche intuitive : ajustement d'un modèle gravitaire
- Position statistique
 - Modèle économique : une équation structurale
 - Observations : des comptages et autres « moments statistiques »
 - Qualification de chaque observation : liaison au modèle économique, et position probabiliste
 - Choisir une méthode d'estimation : et donc une fonction-objectif (écart quadratique, ou vraisemblance)
 - Optimiser la fonction pour obtenir l'estimation (valeur particulière de l'estimateur pour les observations)

Ajuster un modèle gravitaire

- Cas de deux paramètres α, β
- Modèle économique
$$q_{od} = E_o R_d \exp[-\alpha - \beta G_{od}]$$
- Observation : un total pondéré $\sum_{od} X_{od} \cdot q_{od} = X_{tot}$
- Fonction d'écart entre le simulé et l'observé
$$(X_{tot} - \sum_{od} X_{od} q_{od})^2$$
- Retenir la valeur de (α, β) qui minimise l'écart
- Calcul par tabulation, ou par tâtonnement

Équation structurale

- Fonction de demande
 - I relations origine-destination $i = (o, d)$
 - Sur chaque relation i , en fonction de caractères X_i

$$q_i = D_i(\Theta, X_i)$$

- Hypothèses
 - On se donne D_i
 - On connaît les X_i
 - On recherche Θ
 - Erreur du modèle de demande : η_i telle que

$$q_i = D_i(\Theta, X_i) + \eta_i$$

Observations

- Forme de l'observation d'indice a :

$$V_a = \sum_{i \in I} \gamma_{ai} \cdot q_i + \varepsilon_a$$

- Instances
 - Comptage local du trafic sur un arc a : γ_{ai} est la probabilité d'emprunter l'arc a pour la relation O-D i
 - Marges de matrice : dans le cas d'une origine i.e. $a = o$, $\gamma_{ai} = 1$ si i a pour origine o , ou 0 sinon
 - Moments par classe d'O-D, ex. les temps passés pour les O-D d'une certaine tranche de distance : γ_{ai} sert à filtrer les O-D concernées (nul sinon) pour lesquelles il est égal au temps par unité de flux
 - Influence du dispositif d'observation : V_a et ε_a dépendent de la méthode d'observation : durée du recueil, taux d'échantillonnage...

Analyse économétrique

- Qualification d'une observation

$$V_a = \Delta_a \hat{V}_a(\Theta) + \tilde{\varepsilon}_a \quad \text{avec} \quad \begin{cases} \hat{V}_a(\Theta) = \sum_{i \in I} \gamma_{ai} D_i(\Theta, X_i) \\ \tilde{\varepsilon}_a = \varepsilon_a + \Delta_a \sum_{i \in I} \gamma_{ai} \eta_i \\ \Delta_a \text{ durée de l'observation } a \end{cases}$$

- Distribution de l'erreur $\tilde{\varepsilon}_a$: hypothèses usuelles
 - Processus de Poisson (hypothèse classique pour un comptage) : moyenne égale à la variance, donc $\tilde{\varepsilon}_a$ est « Poisson décalée », centrée de variance V_a
 - Ou loi gamma pour moduler la dispersion relative
 - Dans tous les cas, on note $\text{var}[\tilde{\varepsilon}_a] = \sigma_a^2$
- Dépendances entre les erreurs ?
 - Pour simplifier, hypothèse d'indépendance...

Moindres carrés

- Fonction d'écart

$$J(\Theta) = \sum_{a \in \tilde{A}} \left[\frac{\Delta_a \hat{V}_a(\Theta) - V_a}{\sigma_a} \right]^2$$

avec \tilde{A} : ensemble des observations

- A minimiser en fonction de Θ
- Quand les D_i sont linéaires en fonction de Θ , on dispose de formules analytiques pour la solution
 - Pour des erreurs $\tilde{\varepsilon}_a$ de moyenne $\bar{\varepsilon}_a$ et matrice de variance-covariance C , la formule est

$$\hat{\Theta} = [Z^T C^{-1} Z]^{-1} Z^T C^{-1} (V - \bar{\varepsilon})$$

en notant $Z = \Gamma \cdot [X_i^T]_{i \in I}$ les coefficients de Θ dans $\Delta_a \hat{V}_a(\Theta)$

Maximum de vraisemblance

- Fonction de vraisemblance

$$L(\Theta) = \prod_{a \in \tilde{A}} L_a(\Theta) \text{ en cas d'indépendance}$$

avec \tilde{A} : ensemble des observations

- A maximiser en fonction de Θ (ou maximiser le ln)

- Modèle poissonnien
$$L_a(\Theta) = \frac{(\Delta_a \hat{V}_a)^{V_a}}{V_a!} \exp(-\Delta_a \hat{V}_a)$$

$\Lambda_a(\Theta) \equiv \ln L_a(\Theta) \approx V_a [\ln(\frac{\Delta_a \hat{V}_a}{V_a}) - 1] - \Delta_a \hat{V}_a$ grâce à la formule de Stirling

- Modèle gamma
$$f(x|\alpha, \beta) = \beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha)$$
 - Fonction de densité à 2 paramètres α et β ,
$$\Lambda_a(\Theta) = (\alpha - 1) \ln \hat{V}_a(\Theta) - \beta \hat{V}_a(\Theta) + C^{te}$$
 pour contrôler la dispersion relative : prendre $\alpha = \sigma_a^2 / V_a^2$ et $\beta = \alpha / V_a$

Conclusion sur l'estimation

- Fondement statistique
- Traitement efficace
- Conditions nécessaires
 - Modèle de demande
 - Disposer d'observations
 - Analyse économétrique des observations
 - Logiciel spécialisé ! (sauf si modèle de demande directe)
- Autres fondements statistiques
 - Approches bayésiennes : pour mélanger une connaissance préalable et des observations nouvelles

Modèles économiques

- Rappels
- Modèle gravitaire
- Modèle des opportunités
- Modèle de choix discret
- Modèle d'accessibilité aux activités
- Equilibre spatial et modèles de commerce
- Modèles input-output de production et d'échange
- Conclusion

Rappels

- Analyse économique de la demande de Dt
 - Évaluation économique : Jules Dupuit, 1854
 - Connaître la fonction de demande
- Les deux effets
 - En volume : la consommation de déplacement
 - En structure : la répartition par relation O-D
- Interprétation micro-économique
 - Depuis une origine : choisir une destination
 - Choix d'activité ou choix de zone ?

Modèle gravitaire

- Outputs = volumes O-D
- Inputs = attributs des zones d'origine (Y_o^k) et de destination (Z_d^ℓ), et des services de transport (X_i^n)

- Formule d'origine
$$q_i = K \frac{P_o A_d}{d_i^2}$$

- Formule plus générale

$$q_i = K \exp(\sum_k \Psi_k Y_o^k) \exp(\sum_\ell \zeta_\ell Z_d^\ell) \cdot f(X_i^n)$$

- Souvent, f est une fonction du coût généralisé G_i , de forme $f(G_i) = \exp(-\lambda \cdot G_i) \cdot G_i^{-\mu}$

Analyse critique

- Formule intuitive, grâce à l'analogie physique
 - Plus il y a de population ou d'emplois dans au moins une des deux zones, plus il y a de déplacements entre elles
 - Plus le transport est coûteux, en argent ou en temps, moins il y a de déplacements
- Contraintes d'observation pour la répartition :
 - marges des zones
 - regroupement par tranche de distance / coût

Modèle des opportunités

- Depuis une origine o
 - découverte progressive des zones d de destination
 - avec dans chaque zone un nombre A_d d'opportunités, i.e. activités potentielles
- Hypothèse de choix
 - Un demandeur qui rencontre une opportunité, l'accepte avec une probabilité L
- Part de marché d'une zone

$$\Pr\{d|o\} = (1-L)^{\sum_{n<d} A_n} [1 - (1-L)^{A_d}]$$

Car choisir d implique de ne pas avoir choisi une zone $n < d$, et sachant cela, d'avoir choisi d , événement dont la probabilité complémentaire est $(1-L)^{A_d}$

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

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

- *Choix discret conjoint de l'origine et la destination*
 - Utilité $U_{od} = \beta_o^+ + \beta_d^- + \sum_k \beta_k X_{od}^k + \varepsilon_{od}$
 - Équivalent au modèle entropique doublement contraint, et donne lieu au modèle gravitaire (Anas, 1984)

Le modèle logit

- Si les ε_{od} sont distribuées Gumbel i.i.d. de moyenne nulle et même variance $\sigma^2 = \pi^2 / (6\theta^2)$ alors

$$\Pr\{d|o\} = \frac{\exp[\theta(S_d - G_{od})]}{\sum_d \exp[\theta(S_d - G_{od})]}$$

- Utilité conjointe, composite : formule du *log-somme*

$$V_o = \frac{1}{\theta} \ln \sum_d \exp[\theta(S_d - G_{od})]$$

- Fonction de demande : $E_o = D_o[\bar{U}_o]$ pour les émissions de la zone o

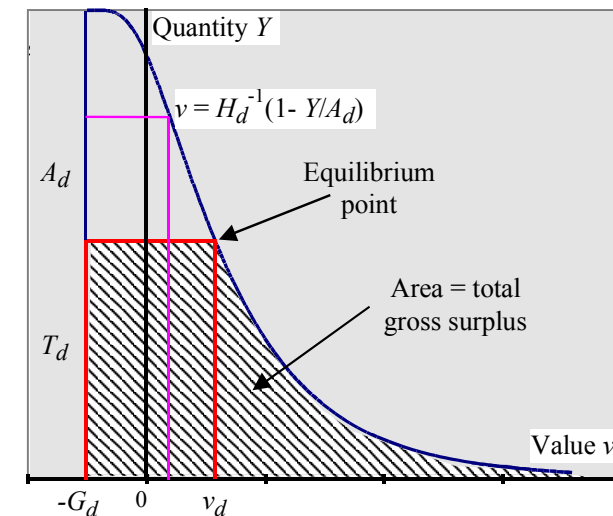
Discussion

- Formules analytiques simples
- Expression de l'utilité brute ?
 - Niveau activité ? Ex. salaire journalier
 - Niveau zone : $S_d = \ln A_d$ indépendant du découpage
- Perturbations
 - Effet d'abondance : plus les options sont nombreuses, plus l'utilité composite est forte ! Pas absurde, mais application délicate
- Congestion des activités : la même pour tous ?

Modèle AVA

- AVA = Accessibility to Vacant Activities
- L'offre d'activités
 - A_d activités dans la zone d
 - H_d fonction de répartition des valeurs brutes
- La demande d'activités
 - T_d clients
 - Chaque client choisit l'activité vacante de valeur brute maximale

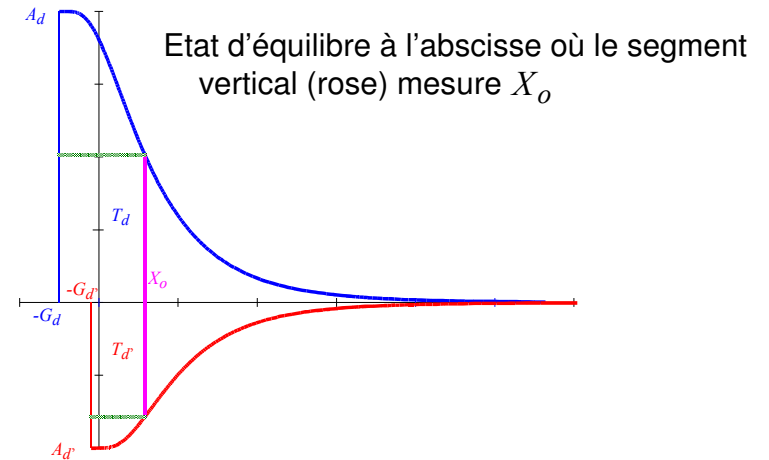
Équilibre du marché



Choix de destination

- L'offre d'activités dans le cas binaire
 - Zone d avec A_d activités et la fonction de répartition H_d
 - Zone d' avec $A_{d'}$ activités et la fonction de répartition $H_{d'}$
- Demande d'activité depuis la zone o
 - X_o clients = $T_d + T_{d'}$
 - Chaque client choisit l'activité vacante dont la valeur nette, égale à la valeur brute moins le coût de Dt (G_d ou $G_{d'}$), est maximale

Equilibre spatial entre destinations



Conséquences

- Formule du surplus pour une zone d'origine
 - Par destination
 - Surplus total depuis l'origine
$$J_{od}(T_d) \equiv \int_0^{T_d} (H_d^{-1}(1 - x/A_d) - G_{od}) dx$$
- Le comportement des consommateurs d'activités, revient à maximiser le surplus net total de l'origine

$$J(T_d; T_{d'}) = J_{od}(T_d) + J_{od'}(T_{d'})$$

$$\max_{T_d, T_{d'} \geq 0} J(T_d; T_{d'})$$
- Définition économique de l'accessibilité à une zone = utilité brute de la meilleure activité vacante

Cas exponentiel et modèle gravitaire

- Si les utilités des activités suivent une distribution exponentielle de paramètre λ_d et décalée de m_d
- Alors $H_d^{-1}(y) = m_d - \ln(1 - y) / \lambda_d$
- Surplus O-D

$$J_{od}(T_d) = T_d \left[m_d + \frac{1 - \ln(T_d / A_d)}{\lambda_d} \right]$$
- Si les distributions sont les mêmes dans les deux zones de destination, et hors saturation

$$\frac{T_d}{T_{d'}} \approx \frac{A_d \exp[-\lambda G_{od}]}{A_{d'} \exp[-\lambda G_{od}]}$$

Modèles de production et d'échange

- Définition générale
- Description
- Tableau des flux
- Fonction de production
- Consommation et commerce
- Formulation mathématique
- Logique d'usage
- Discussion

Définitions

- Modèle Input-Output
 - Relations macroéconomiques entre des secteurs de production : coefficients techniques pour les consommations intermédiaires
 - Relation avec les consommations finales des secteurs
- Modèle de commerce
 - Échanges entre zones : import, export
 - Coefficients de commerce
 - Coûts de transport
- Modèle d'I/O spatial
 - Production par secteur et par zone
 - Échanges entre les secteurs : modèle d'input-output
 - Échanges entre les zones et les secteurs, avec modèle de commerce

Description

- Types de biens/secteurs m, n
- Systèmes de zones
 - Index r, s
 - Extérieur noté e
- Dans la zone r
 - Consommation finale Y_r^m
 - Production locale X_r^m
 - Consommation intermédiaire K_r^{nm}
 - Marché local $X + \text{Import} = K + Y + \text{Export}$
 - Commerce : import I_{sr}^m ou I_{er}^m , export E_{rs}^m ou E_{re}^m

Tableau des flux

Zone r	Bien m	Bien n	Conso finale	Export interne	Export externe
Bien m	K_r^{mm}	K_r^{mn}	Y_r^m	E_{rs}^m	E_{re}^m
Bien n	K_r^{nm}	K_r^{nn}	Y_r^n	E_{rs}^n	E_{re}^n
			Valeur ajoutée		
	X_r^m	X_r^n	Production, en valeur		
	I_{sr}^m	I_{sr}^n	Import interne		
	I_{er}^m	I_{er}^n	Import externe		

Fonction de production

- Localement dans la zone r

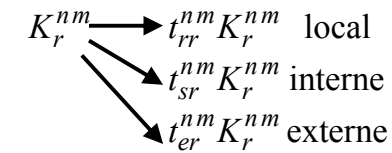
$$X_r^m \Rightarrow a_r^{nm} X_r^m = K_r^{nm}$$

$$X_r^m \Rightarrow a_r^{mm} X_r^m = K_r^{mm}$$

- a_r^{nm} est le "coefficient technique" d'implication du bien n dans la production du bien m ; à décrire localement

Consommation et commerce

- Principe



- Même décomposition pour les consommations finales Y_r^m que pour les consommations K
- Simplification pour déterminer les coefficients : hypothèse de Chenery-Moses $t_{sr}^{mn} = t_{sr}^m$

Coefficients de commerce

- Observation

$$t_{er}^m = I_{er}^m / (Y_r^m + \sum_n K_r^{mn})$$

$$t_{rr}^m = 1 - (I_{er}^m + \sum_s I_{sr}^m) / (Y_r^m + \sum_n K_r^{mn})$$

Les autres par enquête statistique

- Modélisation, par zone r d'import

$$t_{sr}^m = \frac{\exp[-(p_s^m + g_{sr}) / \theta_{rm}]}{\sum_s \exp[-(p_s^m + g_{sr}) / \theta_{rm}]}$$

Formulation mathématique

- Marché r , bien m
 - Import depuis s et e
 - Export vers s et e (contrepartie des imports)
 - Acquisition de la conso intermédiaire
 - Acquisition de la conso finale
 - Bilan offre-demande local (d'où X)

$$X = TK + TY + Y_e$$

- Fonction de production $K = AX$
- Au final $X = (I - TA)^{-1} (TY + Y_e)$

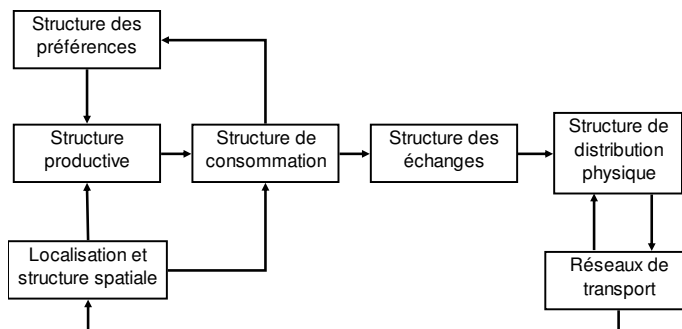
Logique d'usage

- 0- Estimer T et A
- 1- Hypothèses sur Y et Y_e , pour simuler X et en déduire K , les imports, les exports
- 2- Relier T aux prix sur les marchés locaux et aux coûts de transport
- 3- Relier les coûts de transport aux volumes transportés
- 4- Relier les prix des marchés, aux coûts et volumes de la production locale
- 5- Relier A aux coûts des inputs

Discussion

- Modèle agrégé
- Modèle en flux : place des stocks ? Aspects dynamiques ?
- Coûts indépendants du groupage ?
- Flux monétaires : qu'en est-il des flux physiques

Synthèse graphique



Conclusions

- La distribution spatiale, étape cruciale
- Utiliser une démarche économétrique, conciliant
 - Spécification économique, et interprétation
 - Estimation statistique : à maîtriser !
- Modèles avancés
 - Localisation des ménages en fonction des emplois et des services, et du transport
 - Localisation des entreprises en fonction du travail, des marchés (local ou accessible par le transport)
 - Modèles spatialisés d'input / output
- *Etat des connaissances*
 - *Encore imparfait... formules de surplus, peu convaincantes*
 - *Mode d'emploi d'un modèle tel que le gravitaire, généralement assez critiquable*