

**LES MODELES
D'INTERACTION SPATIALE**

2^{ème} PARTIE

Christian CALZADA

DAEI / SES / DEE

*Club d'Echanges
sur le transport de marchandises*

Séance du 11 décembre 1998

LES MODELES D'INTERACTION SPATIALE

COMPLEMENTS

Ces vingt dernières années un certain nombre d'analogies ont été démontré entre les spécifications générales des modèles gravitaires et d'autres techniques : maximisation de l'entropie, programmation mathématique, problèmes de transport (approche combinée), modèles de choix discrets, théorie du chaos, etc..

L'objet de cette partie a trait à des compléments techniques sur les différents types de modèles d'interaction spatiale. Ceci sous la forme de fiches documentées sur chacun des thèmes suivants :

- MIS et calibrages
- MIS, statistique, occasions intervenantes et groupées
- MIS et formulation poissonnienne
- Une approche combinée distribution-affectation
- MIS et entropie
- MIS dynamiques
- MIS et recherche de partitionnement optimale
- MIS, réseaux de neurones et intelligence artificielle
- Bibliographie

FICHE N°1

CALIBRAGES

Introduction

Les procédures de calibrage d'un modèle remplissent traditionnellement deux grands objectifs :

1. estimer le degré d'adhésion du modèle conceptuel à la réalité (estimation),
2. estimer les paramètres du modèle de telle façon que les valeurs trouvées donnent la meilleure estimation (calibrage proprement dit).

Deux grands types de calibrage existent :

- la méthode du minimum quadratique : dans le cadre d'un processus itératif, le but est de réduire le carré de l'écart entre les flux observés et estimés,
- ceux basés sur la méthode du maximum de vraisemblance (EMV) : l'objectif vise à la recherche des valeurs des paramètres qui maximisent la probabilité d'une distribution réelle des données de flux.

Méthodes dites du minimum quadratique

- La méthode de linéarisation classique

cas d'un MIS sans contrainte :

Si l'on reprend les notations générales d'un MIS sans contrainte :

$$N_{ij} = k \cdot E_i^\mu \cdot A_j^\alpha \cdot c_{ij}^{-\beta}$$

pour passer à une forme linéaire, il suffit dès lors de passer aux logarithmes :

$$\ln N_{ij} = \ln k + \mu * E_i + \alpha * A_j - \beta \ln c_{ij}$$

dans le cadre d'un modèle sans contrainte paretien, de la même manière on obtiendra :

$$\ln N_{ij} = \ln k + \mu * E_i + \alpha * A_j - \beta c_{ij}$$

Si les hypothèses des Moindres Carrés Ordinaires (MCO) sont réunies, μ' , α' , β' , sont des estimateurs sans biais et robustes de μ , α , et β . Par contre $\exp(\ln k)'$ est un estimateur biaisé de k . k sera toujours sous-estimé en MCO, cette sous-estimation résulte de la relation :

$$\sum_i \sum_j N_{ij}' \leq \sum_i \sum_j N_{ij}$$

D'autre part des problèmes d'hétéroscédasticité peuvent subvenir suite à la transformation logarithmique, qui peuvent être contournés en recourant aux spécifications de type Moindres Carrés Pondérés ou Généralisés, avec un poids égal à $(\ln N_{ij})^{1/2}$.

cas d'un MIS à doubles contraintes :

La formulation générale (doubles contraintes) est la suivante :

$$N_{ij} = k \cdot E_i \cdot H_i \cdot A_j \cdot G_j \cdot F(c_{ij})$$

1. cas du modèle avec contrainte sur les origines :

on aura : $\beta_0 + \beta_i = \ln H_i + \ln E_i$

2. cas du modèle avec contrainte sur les destinations :

$$\text{on aura : } \beta_0 + \beta_j = \ln G_j + \ln A_j$$

3. cas du modèle avec doubles contraintes :

$$\beta_0 + \beta_i + \beta_j = \ln H_i + \ln E_i + \ln G_j + \ln A_j$$

Il existe au moins deux situations dans lesquelles le critère des MCO est plus maniable que celui de l'EMV :

- cas des MIS simples avec sans ou une seule contrainte,
- cas du modèle général de MIS sans contrainte lorsque l'estimation est menée simultanément sur l'ensemble des $2n+1$ paramètres $((E_i, A_j, \beta))$.

Avantages de la méthode :

♦ avantages de la méthode déterministe (et non probabiliste comme dans le cas de l'EMV) visant à minimiser les écarts entre matrice observée et matrice théorique. En particulier si l'on retient le critère de minimisation de la somme des carrés des résidus on peut dire que :

- ♦ ce critère est conforme à l'esprit du Problème de Maximisation de l'Entropie (PEM) : si l'hypothèse de normalité des résidus est respectée, alors les estimations par MCO et EMV sont identiques, de plus l'estimation par EMV satisfait au PEM.

- ♦ ce critère satisfait l'axiome de réflexivité : $\sum_{i=1}^n \sum_{j=1}^n (N_{ij} - \hat{N}_{ij}) = 0 \Leftrightarrow [N_{ij}] = [\hat{N}_{ij}]$

- ♦ ce critère offre des solutions analytiques simples
- ♦ ce critère conduit généralement à une solution unique

Autre approche :

Une étude de O' Kelly, Song and Shen (1995) propose une approche alternative à la confrontation des résultats estimés et observés. Dans le cadre d'un modèle sans contrainte et en recourant à la programmation linéaire, une estimation de l'impédance β et des facteurs d'attraction P_i et P_j est réalisée. Il s'agit par transformation logarithmique de revenir à un système linéaire, en ajoutant un terme d'erreur D_{ij} , écart entre le système idéal et la réalité.

$$T_{ij} = k(P_i P_j) / (c_{ij})^\beta \quad \forall i \neq j \quad k \text{ et } c_{ij} \text{ sont des constantes}$$

$$\ln P_i + \ln P_j - \beta (\ln c_{ij}) = \ln T_{ij} - \ln k$$

$$X_i + X_j - a_{ij} \beta = b_{ij} + D_{ij} \quad \forall i \neq j$$

avec :

$$\ln P_i = X_i \quad \ln P_j = X_j \quad a_{ij} = \ln c_{ij} \quad b_{ij} = \ln T_{ij} - \ln k$$

$$D_{ij} = D_{ij}^1 - D_{ij}^2 \quad |D_{ij}| = D_{ij}^1 + D_{ij}^2 \text{ pour tout } i \neq j$$

Pour l'estimation de X_i, X_j et β , deux techniques de calcul sont envisageables :

- Minimisation du maximum de l'erreur absolue W :

MinW

$$D_{ij}^1 - D_{ij}^2 = X_i + X_j - a_{ij} \beta - b_{ij}$$

$$W - D_{ij}^1 - D_{ij}^2 \geq 0$$

$$D_{ij}^1, D_{ij}^2, X_i, X_j, \beta, W \geq 0$$

- Minimisation de la somme des erreurs absolues :

$$\text{Min} \sum_i \sum_j (D_{ij}^1 + D_{ij}^2)$$

$$D_{ij}^1 - D_{ij}^2 = X_i + X_j - a_{ij}\beta - b_{ij}$$

$$D_{ij}^1, D_{ij}^2, X_i, X_j, \beta, W \geq 0$$

- Minimisation de la somme pondérée des erreurs absolues :

$$\text{Min} \sum_i \sum_j p_{ij} (D_{ij}^1 + D_{ij}^2)$$

$$D_{ij}^1 - D_{ij}^2 = X_i + X_j - a_{ij}\beta - b_{ij}$$

$$D_{ij}^1, D_{ij}^2, X_i, X_j, \beta, W \geq 0$$

avec p_{ij} poids.

Méthodes dites du maximum de vraisemblance

L'approche par maximisation de la vraisemblance est similaire de celle du principe de l'Entropie Maximum. On montre à partir des travaux de A.W. Evans (1971) et F.J. Cesario (1975) que l'estimateur $\hat{\beta}$ du maximum de vraisemblance du paramètre β d'un MIS sans contrainte satisfait au principe de l'Entropie Maximum.

Rappel :

Le Principe de l'Entropie Maximum (PEM) développé par Jaynes (1957) généralise le fameux "Principe de la Raison Suffisante" de Laplace.

Formulation :

On considère une variable aléatoire X où x_i , ($i = 1, \dots, n$.) représente les valeurs numériques de la variable aléatoire X et p_i , ($i = 1, \dots, n$.) les probabilités correspondantes d'apparition des valeurs x_i .

$$p_i \geq 0 ; \sum_{i=1}^n p_i = 1$$

Définition du problème :

Etablir la répartition p_i , sans connaître les fréquences relatives $f(x_i) = f_i$ de chaque événement i , en connaissant seulement quelques moments de la variable aléatoire X .

Solution : Recourir aux probabilités subjectives.

L'école des subjectivistes regarde les probabilités comme une expression de notre ignorance. La probabilité d'un événement constitue l'expression formelle de l'espérance que l'on a vis-à-vis de l'occurrence de cet événement (et non comme une propriété propre de cet événement basée sur l'expérience). Autrement dit :

- un test objectif énoncera : 'cette distribution représente-t-elle des fluctuations observables ?'. On recherchera alors une méthode d'échantillonnage qui évite au mieux les biais.
- un test subjectif : 'cette distribution représente-t-elle correctement l'état de notre connaissance ?'. On recherchera alors une distribution de probabilité la moins biaisée possible étant donné l'information disponible. Ce concept sert notamment quand il s'agit d'établir des fonctions de répartition pour des événements constitutifs de phénomènes aléatoires non directement observables, c'est le cas de la majorité des MIS.

On ne calculera donc pas p_i à partir des données d'expériences mais à partir d'un principe variationnel qui consiste à exprimer l'état de notre connaissance du phénomène en ces termes : 'Compte tenu de l'information disponible, le degré de notre incertitude est maximum'.

En utilisant le résultat de *Shannon* qui fait de l'entropie une mesure de l'incertitude, Jaynes formule mathématiquement que l'entropie $H(p_i)$ de la distribution p_i est maximum compte tenu de l'information disponible sous la forme de moments statistiques.

Dans le cas du moment de premier ordre, c'est à dire l'espérance mathématique $E(X)$ on aura :

$$\text{Max}H(p_i) = -K \sum_{i=1}^n p_i \text{Log} p_i$$

$$\text{avec : } \sum_{i=1}^n p_i = 1 \quad \text{et} \quad \sum_{i=1}^n p_i f_i = E(X)$$

On aboutit au résultat suivant :

$$p_i = \frac{1}{\phi(\beta)} \exp(-\beta f_i) \quad \text{où} \quad \beta \text{ est solution unique de : } \frac{\phi'(\beta)}{\phi(\beta)} = -E(X)$$

$$\text{avec } \phi(\beta) = \sum_{i=1}^n \exp(-\beta f_i)$$

Si l'on reprend la formulation suivante simplifiée d'un MIS écrite en termes de fréquences :

$$N_{ij} = E_i A_j \exp(-\beta c_{ij}) \quad (1)$$

$$\text{avec } \sum_{i=1}^n \sum_{j=1}^n N_{ij} = 1$$

On dispose d'une matrice échantillon $[N_{ij}^0]$ et l'on suppose que la fonction de probabilité jointe des fréquences $n_{ij}^0 = \frac{N_{ij}^0}{N^0}$ avec $N^0 = \sum_i \sum_j N_{ij}^0$ suit une loi multinomiale.

La fonction de vraisemblance s'écrira alors sous la forme :

$$L[N_{ij}^0] = K \prod_i \prod_j (n_{ij}^0)^{N_{ij}^0} \quad (2) \quad \text{où } K \text{ intervient comme facteur de proportionnalité.}$$

En substituant (1) dans (2) on obtient une fonction en $2n + 1$ paramètres. On maximise le logarithme de la fonction de vraisemblance :

$$\text{Log}L[N_{ij}^0] = \sum_i \sum_j N_{ij}^0 [\text{Log}E_i + \text{Log}A_j - \beta c_{ij}] + \text{Log}K$$

$$\text{sous la contrainte de cohérence : } \sum_i \sum_j E_i A_j \exp(-\beta c_{ij}) = 1$$

L'écriture du lagrangien donne :

$$L = \sum_i \sum_j N_{ij}^0 [\text{Log}E_i + \text{Log}A_j - \beta c_{ij}] - \lambda \left[\sum_i \sum_j E_i A_j \exp(-\beta c_{ij}) - 1 \right]$$

* L'annulation de la dérivée de L par rapport à E_i donne :

$$\sum_j E_i A_j \exp(-\beta c_{ij}) = \frac{1}{N^0} \sum_j N_{ij}^0, \quad (i = 1, \dots, n)$$

* L'annulation de la dérivée de L par rapport à A_j donne :

$$\sum_i E_i A_j \exp(-\beta c_{ij}) = \frac{1}{N^0} \sum_i N_{ij}^0, \quad (j = 1, \dots, n)$$

Les deux expressions plus haut expriment les contraintes sur les marges lignes et colonnes.

* L'annulation de la dérivée de L par rapport à β donne :

$$\sum_i \sum_j E_i A_j \exp(-\beta c_{ij}) = \frac{1}{N^0} \sum_i \sum_j N_{ij}^0 c_{ij} \text{ soit encore } \sum_i \sum_j n_{ij} c_{ij} = \sum_i \sum_j \frac{N_{ij}^0}{N^0} c_{ij} = \bar{C}^0$$

autrement dit : coût moyen théorique = coût moyen observé.

Comme une des conditions nécessaires qu'une distribution (N_{ij}) satisfasse le critère de l'entropie maximale et soit conforme à la formulation générale est qu'elle respecte le contrainte de coût moyen, alors on peut dire que les estimateurs du maximum de vraisemblance $\tilde{E}_i, \tilde{A}_j, \tilde{\beta}$ sont conformes au Principe de l'Entropie Maximum.

Algorithmes de maximisation de l'entropie

H. Persson propose en 1986, un algorithme de résolution du problème de maximisation de l'entropie.

Description de l'algorithme :

Soit m variable inconnues x_j ($j = 1, \dots, m$), l'algorithme consiste à maximiser :

$$F(x) = - \sum_j x_j \ln \frac{x_j}{x_j^0}$$

sous les contraintes :

$$\sum_j a_{kj} x_j \leq b_k \quad k \in I_1$$

$$\sum_j a_{kj} x_j = b_k \quad k \in I_2$$

$$x_j \geq 0$$

avec $\bar{x} = (x_1, \dots, x_m)$ et x_j^0, a_{kj}, b_k qui sont des constantes positives pour tout i, k et j .

Le lagrangien s'écrit :

$$L(\bar{x}, \bar{\lambda}) = - \sum_i x_i \ln(x_i / x_i^0) + \sum_k \lambda_k (b_k - \sum_i a_{ki} x_i)$$

avec $\bar{\lambda} = (\lambda_1, \dots, \lambda_l)$.

Les conditions nécessaires et suffisantes pour la maximisation sont données par :

$$\frac{\partial L}{\partial x_i} = - \ln \frac{x_i^*}{x_i^0} - 1 - \sum_k \lambda_k^* a_{ki} = 0,$$

$$\lambda_k^* \left(\sum_j a_{kj} x_j^* - b_k \right) = 0, \quad \lambda_k^* \geq 0, \quad k \in I_1,$$

$$\sum_j a_{kj} x_j^* = b_k, \quad k \in I_2.$$

La fonction duale Φ du problème d'optimisation est définie par :

$$\Phi(\bar{\lambda}) = \max L(\bar{x}, \bar{\lambda})$$

$$\begin{aligned} \text{Min } \Phi(\bar{\lambda}) \\ \lambda_k \geq 0 \quad , \quad k \in I_1 \end{aligned}$$

On aboutit à :

$$\Phi(\bar{\lambda}^*) = F(\bar{x}^*)$$

$$\Phi'_k{}^n = \partial\Phi(\bar{\lambda}^n) / \partial\lambda_k \quad \Phi''_{kk}{}^n = \partial^2\Phi(\bar{\lambda}^n) / \partial\lambda_k^2$$

$$\begin{aligned} \lambda_j^{n+1} &= \lambda_j^n \quad \text{per } j \neq k \quad , \\ \lambda_k^{n+1} &= \max \left\{ 0, \lambda_k^n - \gamma \frac{\Phi'_k{}^n}{\Phi''_{kk}{}^n} \right\} \quad , \quad k \in I_1 \\ \lambda_k^{n+1} &= \lambda_k^n - \gamma \frac{\Phi'_k{}^n}{\Phi''_{kk}{}^n} \quad , \quad k \in I_2 \end{aligned}$$

* Autres algorithmes de maximisation de la fonction d'entropie

Les algorithmes proposés par *Ericksson* (1980) et *Erlander* (1981) utilisent uniquement des contraintes d'égalité, contrairement à *Persson*, ils déterminent à chaque itération, les nouvelles valeurs pour chacune des variables duales.

Un autre célèbre algorithme est celui de *Bregman*, introduit par *Lamond* et *Stewart* en 1981, il consiste en la résolution d'un système d'équations non linéaires, très coûteuse en temps de calcul.

Autres méthodes de calibrage :

Nous n'aborderons pas ici les méthodes de calibrage ('*balancing methods*') développées dans le cadre du domaine des transports qui relèvent de problèmes de minimisation convexe :

- méthode de *Bregman*,
- algorithme de *Kruithof*,
- méthode de *Evans-Kirby*,
- tout ce qui a trait à la littérature sur les problèmes d'affectation.

Nous renvoyons le lecteur à la bibliographie de fin de document et aux annexes de ce document.

Bibliographie

- Baxter M.J. (1987), "Testing for misspecification in models of spatial flows", *Environment and Planning A*, **19**, 9, pp. 1153-1160.
- Baxter M.J. (1985), "Quasi-likelihood estimation and diagnostic statistics for spatial interaction models", *Environment and Planning A*, **17**, pp. 1627-1635.
- Davies R.B., Guy C.M. (1987), "The statistical modelling of flow data when the Poisson assumption is violated", *Geographical Analysis*, **19**, 4, pp. 300-314.
- Deming W.E., Stephan F.S. (1940), "On a least squares adjustment of sampled frequency tables when the expected marginal totals are known", *Annals of Mathematical Statistics*, **11**, pp. 427-446.
- Eriksson J. (1980), "A note on solution of large sparse maximum entropy problems with linear equality constraints", *Mathematical Programming*, **18**, pp. 146-154.
- Erlander S. (1981), "Entropy in linear programs", *Mathematical Programming*, **21**, pp. 137-151.
- Erlander S., Stewart F. (1985), The gravity model in transportation analysis: theory and extensions, chap. 7, pp. 159-172, Topics in Transportation, Utrecht, The Netherlands, Tokyo, Japan.
- Evans A.W. (1971), "The calibration of trip distribution models with exponential or similar cost functions", *Transportation Research*, **5**, pp. 15-38.
- Furness K.P. (1965), "Time function iteration, traffic engineering and control", **7**, pp. 458-460.
- Hyman G.M. (1969), "The calibration of trip distribution models", *Environment and Planning A*, **1**, pp. 105-112.
- Lamond B, Steward N.F. (1981), "Bregman's balancing method", *Transportation Research*, **15B**, pp. 239-248.
- O'Kelly M.E., Song W., Shen G. (1995), "New estimates of gravitational attraction by linear programming", *Geographical Analysis*, **27**, 4, pp. 271-285.
- McCullagh P., Nelder J.A. (1989), Generalized linear models, London: Chapman and Hall.
- Persson H. (1986), "Solving the entropy maximisation problem with equality and inequality constraints", *Environment and Planning A*, **18**, 12, pp. 1665-1676.
- Rao C.R. (1973), Linear statistical inference and its applications, New York, Wiley.
- Rogerson P.A. (1986), "Parameter estimation in the intervening opportunities model", *Geographical Analysis*, **18**, 4, pp. 357-360.
- Schneider M. (1959), "Gravity models and trip distribution theory", *paper and proceeding of the Regional Science Association*, **5**, pp. 51-56.
- Sen A., Smith T.E. (1995), Gravity models of spatial interaction behaviour, Advances in spatial and network economics, chap. 2, pp. 355-531.
- Sen A. (1986), "Maximum likelihood estimation of gravity model parameters", *Journal of Regional Science*, **26**, 3, pp. 461-474.
- Sen A., Matuszewski Z. (1991), "Properties of maximum likelihood estimates of gravity model parameters", *Journal of Regional Science*, **31**, 4, pp. 469-486.
- Smith T.E. (1984), "A foundational framework for poisson frequency analysis of weakly interacting behaviour", *Working papers in Regional Science and Transportation*, University of Pennsylvania, Philadelphia
- Smith T.E. (1987), "Poisson gravity models of spatial flows", *Journal of Regional Science*, **27**, 3, pp. 315-340.
- Yun S., Sen A. (1994), "Computation of maximum likelihood estimates of gravity model parameters", *Journal of Regional Science*, **34**, 2, pp. 199-216.
- Weber J.S. (1987), "Elasticities of constrained gravity models", *Journal of Regional Science*, **27**, 4, pp. 621-640.
- Weber J., Sen A. (1985), "Sensitivity of gravity model parameters", *Journal of Regional Science*, **25**, pp. 317-336.
- Wedderburg R.W.M. (1974), "Quasi-likelihood and the Gauss-Newton method", *Biometrika*, **61**, pp. 439-447.

FICHE N°2

MIS, STATISTIQUE, OCCASIONS INTERVENANTES ET GROUPEES

Rappel

Une formulation générale de la classe des modèles gravitaires peut s'écrire comme :

$$E_c(N_{ij}) = A_c(i)B_c(j)F_c(c_{ij}) \quad \text{avec } P = \{P_c : c \in C\}$$

où $A(\cdot)$ et $B(\cdot)$: qui sont des fonctions de poids ('weight functions') à l'origine et la destination (attributs locaux mais aussi une constante gravitationnelle démographique G) et $F(\cdot)$ une fonction de résistance ('distance deterrence function').

MIS et Statistique

L'on sait que les MIS (SIM's) ont un fondement macro-économique basé sur la maximisation de l'entropie, tandis que les Modèles de Choix Discrets (DCM's) ont eux un fondement micro-économique centré sur la maximisation d'utilités individuelles. On peut dès lors penser les MIS comme un cas particulier des DCM's.

En effet malgré les apparences les MIS se rapprochent plus des modèles à variables endogènes qualitatives que des modèles quantitatifs usuels. La nature quantitative de la variable $E_c(N_{ij})$ (flux migratoires, navettes domicile-travail, circulation de marchandises sur un réseau, ..) vient du fait qu'elle résulte de l'agrégation d'un ensemble de décisions individuelles. Le phénomène sous-jacent est donc fondamentalement de nature qualitative : choix de l'une des j destinations possible à partir de l'origine i , autrement dit probabilité que l'unité de base attirée par la destination j provienne de l'origine i .

Formellement :

- cas des MIS avec contrainte d'émission :

Exemple de flux (niveau moyen) :

Pour un lieu d'origine i des flux émis, la capacité d'émission du lieu i se répartit entre j destinations, avec une contrainte de répartition telle que :

$$\forall i, \sum_j N_{ij} \equiv N_i = E_i \quad \text{avec } E_i : \text{volume global émis par l'origine } i$$

On peut dès lors écrire :

$$p_{ij} = \frac{N_{ij}}{E_i} = k A_j F(c_{ij}) \quad \text{qui peut s'interpréter comme un distribution de probabilités sur l'ensemble}$$

des j destinations avec p_{ij} étant la probabilité pour que l'origine du flux en i soit localisé en j .

Si nous reprenons l'équation générale et en simplifiant l'écriture, l'on obtient :

$$E_c(N_{ij}) = A_c(i)B_c(j)F_c(c_{ij}) \quad \text{avec } P = \{P_c : c \in C\}$$

$$N_{ij} = k.E_i A_j F(c_{ij}) \quad \text{soit } N_{ij} = k.E_i p_{ij}$$

$$\text{avec } p_{ij} = G_j^{-1} A_j F(c_{ij}) \quad \text{où } G_j = \sum_j A_j F(c_{ij})$$

Remarque :

Dans le cas spécifique du modèle entropique [$F(c_{ij}, \beta) = \exp(-\beta c_{ij})$] avec forme log-linéaire [$\text{Log}(A_j) = N_j \delta$] pour les effets d'attraction, on retombe sous une forme de type logit indépendant :

$$p_{ij} = \frac{\exp(Y_j \delta - c_{ij} \beta)}{\sum_k \exp(Y_k \delta - c_{ik} \beta)} = \left[1 + \sum_{k \neq j} \exp[(Y_k - Y_j) \delta - (c_{ik} - c_{ij}) \beta] \right]^{-1}$$

• cas des MIS avec contrainte d'attraction :

Symétriquement, l'on a :

$$\forall j, \sum_i N_{ij} \equiv N_j = A_j \quad \text{avec } A_j : \text{volume global reçus par la destination } j$$

On peut dès lors écrire :

$$N_{ij} = k \cdot A_j q_{ij}$$

$$\text{avec } q_{ij} = H_i^{-1} E_i F(c_{ij}) \quad \text{où} \quad H_i = \sum_i E_i F(c_{ij})$$

• cas des MIS avec doubles contraintes :

On peut écrire :

$$N_{ij} = k E_i p_{ij} = k \cdot A_j q_{ij} = k (G_i H_i)^{-1} E_i A_j F(c_{ij})$$

avec :

$$p_{ij} = G_j^{-1} A_j F(c_{ij}) \quad \text{où} \quad G_j = \sum_j A_j F(c_{ij})$$

$$q_{ij} = H_i^{-1} E_i F(c_{ij}) \quad \text{où} \quad H_i = \sum_i E_i F(c_{ij})$$

Remarque :

Ayant établi la similitude avec les modèles de choix discrets (Discrete Choice Models, DCM's) et en particulier les spécifications des modèles polytomiques non ordonnés, les méthodes d'estimation sont celles du maximum de vraisemblance, ce qui reviendra (cas des contraintes d'émission) à rechercher les estimateurs β et δ qui maximisent la fonction :

$$l(\beta, \delta) = \sum_{i,j} N_j \ln p_{ij}(\beta, \sigma)$$

Cette équivalence mathématique entre modèles MIS et DCMs a été démontré par nombre d'auteurs :

- Anas (1975,1983) stipule que les modèles gravitaires ou entropiques ne sont pas moins des modèles de comportements que les modèles de choix discret et en particulier les MNL (*Multinomial Logit*).

⇒ Alex Anas, (1983), "Discrete Choice Theory, information theory and the multinomial logit and gravity models", *Transportation, Research B*, Vol. 17B, N°1, pp. 13-23.

- Leonardi (1985) quant à lui en partant de la théorie micro-économétrique, et en recourant à des techniques statistiques d'agrégation ('méthodes asymptotiques') obtient un modèle MNL.

⇒ Leonardi G. (1985), "Equivalenzia Asintotica tra la Teoria delle Utilita Casuali e la Massimizzazione dell'Entropia", *Territorio e Trasporti* (A. Reggiani, ed.), Franco Angeli, Milano, pp. 29-66.

Conclusions

- il apparaît une analogie entre la spécification du type MNL et les MIS. Pour les autres types de spécification des DCM's (Probit, GEVM, ...), à ce jour aucune analogie n'a pu être prouvée scientifiquement.
- les limites et axiomes inhérents aux DCM's (hypothèse IAA) s'appliquent aux modèles MIS à simple contrainte. Le modèle à doubles contraintes avec processus de choix séquentiel se rapproche de spécifications du type MNL hiérarchique.
- dans les modèles de choix discrets les fonctions d'utilité sont supposées linéaires pour les attributs des alternatives aussi bien que pour les individus, hypothèse non explicite dans la formulation générale des modèles MIS.

Modèles gravitaires d'occasions intervenantes (I)

Remarque préliminaire : le vocable 'occasion intervenante' est une traduction littérale de l'expression '*intervening opportunity*'. D'autres auteurs préfèrent parler d'occasion ou d'opportunité intermédiaire. Lorsqu'un individu se déplace d'une zone i vers une zone j , une occasion intervenante consiste en la possibilité pour cet individu de satisfaire son besoin avant qu'il n'atteigne la zone j .

La théorie sous-jacente à ce type de modèles a été établie par *Stouffer* (1940), elle formule l'hypothèse que la mobilité n'est pas nécessairement associée avec la distance : « le nombre de personnes franchissant une distance donnée pour relier une zone d'origine i à une zone de destination j est directement proportionnel au nombre d'opportunités ('*opportunities*') localisées en j et inversement proportionnel au nombre d'occasions intervenantes ('*intervening opportunities*') situées le long du trajet (i, j) ».

Le modèle de Schneider

Une adaptation au cadre du problème de la distribution, des hypothèses de *Stouffer* a été faite par *Schneider* (1960). Ce dernier s'est servi du cadre de la statistique mathématique et caractérise le comportement de déplacement d'un individu de la manière suivante : un individu se déplace d'une origine i à la plus proche occasion de destination, si cette destination est acceptée, le déplacement est terminé, sinon on il continue à la prochaine plus proche occasion de destination, et ainsi de suite. Il est à noter que l'ensemble des destinations est classé par rang décroissant de l'accessibilité à partir chaque zone d'émission.

Posons :

U_{ij}^k : probabilité que l'individu parti de la zone i pour se rendre dans la zone j ne soit pas satisfait des opportunités rencontrées dans la $k^{\text{ème}}$ zone suivant la zone de départ.

L : probabilité qu'une opportunité rencontrée satisfasse l'individu qui se déplace.

Soit D^1 le nombre d'occasions pouvant intervenir dans la 1^{ère} zone immédiatement contiguë à la zone de départ i .

La probabilité que l'individu ne satisfasse pas son besoin dans la zone 1 est complémentaire à la probabilité qu'il y trouve satisfaction :

$$U_{ij}^1 + LD^1 = 1$$

En termes de déplacements, la probabilité que l'individu continue son chemin au delà de la zone 1 est complémentaire à la probabilité que cette zone soit son point de destination finale.

La probabilité U_{ij}^2 que l'individu poursuive son chemin au delà de la zone 2 sera égale à la probabilité que cet individu ait dépassé la zone 1, multipliée par la probabilité qu'il ne trouve pas satisfaction dans la zone 2, soit :

$$U_{ij}^2 = U_{ij}^1(1 - LD^2)$$

et ainsi de suite :

$$U_{ij}^k = U_{ij}^{k-1}(1 - LD^k)$$

Schneider calcule alors le nombre A^k , somme des opportunités rencontrées sur le trajet allant de la zone i à la zone k soit : $A^k = \sum_{i=1}^k D^i$ comme $D^k = A^k - A^{k-1}$, on aura alors :

$$\frac{U_{ij}^k - U_{ij}^{k-1}}{U_{ij}^{k-1}} = -L(A^k - A^{k-1}) \quad \text{soit encore} \quad \frac{dU_{ij}}{U_{ij}} = -LdA$$

La résolution de cette équation différentielle conduit à :

$$\text{Log}U_{ij} = -LA + \text{constante} \quad \text{soit } U_{ij} = C_i \exp(-LA)$$

avec C_i : constante d'intégration.

En définissant N_{ij}^k , nombre de déplacements de la zone i à la $k^{\text{ième}}$ zone en direction de j et O_i comme le nombre total de déplacements quittant la zone i : $O_i = \sum_j N_{ij}^k$, on aura :

$$N_{ij}^k = O_i (U_{ij}^{k-1} - U_{ij}^k) \text{ soit encore :}$$

$$N_{ij}^k = C_i O_i [\exp(-LA^{k-1}) - \exp(-LA^k)]$$

N.B. : Il y a indépendance entre les deux événements suivants : « continuer au delà de la zone $k-1$ » et « ne pas être satisfait dans la zone k ».

La formulation générale du modèle de Schneider peut encore s'écrire :

$$N_{ij} = O_i [\exp(-\lambda U_j) - \exp(-\lambda U_{j+1})]$$

avec :

U_j : distribution cumulée des occasions intermédiaires entre i et j

λ : est la probabilité qu'une occasion de destination soit acceptée

Dans ce type de modélisation, la distribution spatiale des flux d'une origine donnée est gouvernée par les attractions exercées par les occasions existantes dans chaque destination.

L'on s'aperçoit dès lors que dans le cadre du modèle gravitaire classique, les effets des occasions des se déplacer ne sont pas prises en compte, a contrario le modèle d'occasions intermédiaires souffre de l'omission des effets non linéaires des accessibilités (distances) des destinations en concurrence.

Comment combiner dès lors ces deux approches ?

Baxter et Ewing (1979) ont d'abord intégré les occasions intermédiaires dans la spécification de la fonction de distance du modèle gravitaire conventionnel ce qui améliore la prédiction de 13%, ensuite *Wills* (1986) a développé un modèle combiné complexe.

Il faut attendre 1993 et la contribution originale de faite par *GonValves et Ulyssea-Neto* pour que la liaison entre les deux approches, intégration de l'effet de la distance et des occasions intermédiaires soit formalisée dans un même modèle.

Un certain nombre d'hypothèses sur le nombre de flux entre chaque paire d'OD dans un intervalle de temps donné, sont posées :

- ce nombre est proportionnel à la propension de la zone d'émission à générer des flux,
- ce nombre est proportionnel à l'attractivité de la zone à destination,
- ce nombre est inversement proportionnel à une mesure de la friction spatiale associée à la distance entre les zones interagissant,
- ce nombre est inversement proportionnel à une mesure de la friction spatiale associée au nombre d'occasions intermédiaires entre les zones interagissant.

L'approche développée est de type macro-économique, au sens où il existe un comportement homogène des individus qui n'interagissent pas entre eux.

Formulation :

$$P(i, j) = P(i) \cdot P(j/i)$$

Probabilité d'un flux d'une origine i à une destination j = probabilité de flux partant de i * probabilité conditionnelle à destination de j sachant qu'il est parti de i . Notons respectivement O_i le nombre de flux ayant pour origine i et D_j le nombre de flux attirés par j .

$$N = \sum_i O_i = \sum_j D_j$$

On a : $P(i) = \frac{O_i}{N}$ (hypothèse d'homogénéité des comportements)

Notons V_j le nombre d'occasions à destination de j et $L_{ij} = (l; 1 \leq l \leq J)$ où la zone l est une zone d'occasions entre i et j .

Le nombre d'occasions entre i et j est donné par : $W_{ij} = \sum_{l \in L_{ij}} V_l$

Inversement on sait par ailleurs que la probabilité $P(j/i)$ peut s'écrire en utilisant la formulation de Schneider comme :

$$P(j/i) = k_i \left\{ \exp(-\lambda W_{ij}) - \exp[-\lambda(W_{ij} + V_j)] \right\} = k_i \exp(-\lambda W_{ij}) [1 - \exp(-\lambda V_j)]$$

avec $\sum_j P(j/i) = 1$

k_i = constante de proportionnalité ('balancing factor') qui garantit que chaque flux ayant pour origine i aboutit à une destination j .

On aboutit par substitutions successives à l'équation :

$$P(i, j) = \frac{O_i k_i}{N} \exp(-\lambda W_{ij}) [1 - \exp(-\lambda V_j)]$$

La probabilité a priori de la distribution des flux (N_{ij}) est donnée par :

$$P(N_{ij}) = \prod_{i,j} \left[P(i, j)^{N_{ij}} \right] = \prod_{i,j} \left\{ \frac{O_i k_i}{N} \exp(-\lambda W_{ij}) [1 - \exp(-\lambda V_j)] \right\}^{T_{ij}}$$

avec $\sum_j k_i \exp(-\lambda W_{ij}) [1 - \exp(-\lambda V_j)] = 1 \quad \forall i = 1, 2, \dots, I$

Pour $M[N_{ij}]$, on peut écrire (cf. Wilson) : $W(N) = \frac{N!}{\prod_{i,j} N_{ij}!}$

On définit alors la probabilité $P(N)$ comme :

$$P(N) = \frac{N!}{\prod_{i,j} N_{ij}!} \cdot P(N_{ij})$$

autrement dit :

$$P(N) = \frac{N!}{\prod_{i,j} N_{ij}!} \cdot \prod_{i,j} \left\{ \frac{O_i \cdot k_i}{N} \cdot \exp(-\lambda \cdot W_{ij}) \cdot [1 - \exp(-\lambda V_j)] \right\}^{T_{ij}}$$

Le problème d'optimisation de *Wilson* revient alors à résoudre le système d'équations suivant :

$$\max_{N_{ij}} N = P(N)$$

sous les contraintes :

$$\sum_j N_{ij} = O_i, \quad \sum_i N_{ij} = D_j, \quad \sum_{i,j} N_{ij} c_{ij} = C.$$

Après maximisation du Lagrangien et plusieurs substitutions, on aboutit à :

$$N_{ij} = A_i \cdot B_j \cdot O_i \cdot D_j \cdot \exp(-\lambda \cdot W_{ij} - \beta \cdot c_{ij});$$

avec :

$$A_i = \left[\sum_j B_j \cdot D_j \cdot \exp(-\lambda \cdot W_{ij} - \beta \cdot c_{ij}) \right]^{-1}$$

$$B_j = \left[\sum_i A_i \cdot O_i \cdot \exp(-\lambda \cdot W_{ij} - \beta \cdot c_{ij}) \right]^{-1}$$

Le calibrage du modèle se fait par maximisation de la vraisemblance.

Exemple :

Un exemple pratique présenté par les auteurs concernait une matrice billetterie des déplacements par bus entre 44 municipalités du sud de l'état de *Santa Catarina* au Brésil (1985). Outre cette matrice on disposait d'une matrice-temps, des profils socio-économiques des municipalités (population, nombre d'inscriptions scolaires en secondaire, nombre de lits d'hôpitaux, nombre d'établissements commerciaux). Les coûts de trajets ont été approximés à partir des temps de trajets entre zones.

Le nombre d'occasions a été mesuré grâce un indice variant sur une échelle de 0 à 100, indice constitué comme une moyenne pondérée du nombre d'activités (shopping, professionnel, éducation, santé) associés aux motifs de déplacements.

En vue de définir les zones intermédiaires l entre deux zones i et j , a été énoncé :

- que le temps de trajet de i à l est plus court que de i à j ,
- que l est hiérarchiquement (cf. théorème des places centrales de *Christaller*) supérieur ou égal à i .

Les résultats ont abouti à des résultats très satisfaisants ($R^2=0,9378$; Erreur absolue moyenne normalisée = 59,7121).

Bibliographie

- Brown S. (1988), "The wheel of retail gravitation", *Environment and Planning A*, **24**, **10**, pp. 1409-1429.
- Fik T.J. (1988), "Hierarchical interaction : The modelling of a competition central place system", *Annals of Regional Science*, **22**, **2**, pp. 48-69.
- Goncalves M.B., Ulissea-Neto (1993), "The development of a new gravity-opportunity model for trip distribution", *Environment and Planning A*, **25**, **6**, pp. 817-826.
- Guy C.M. (1987), "Recent advances in spatial interaction modelling : An application to the forecasting of shopping travel", *Environment and Planning A*, **19**, **2**, pp. 173-186.
- Hua C. (1990), "A flexible and consistent system for modelling interregional trade flows", *Environment and Planning A*, **22**, **4**, pp. 439-457.
- Muranyi T.C., Miller J.D. (1966), *New applications of the opportunity model of traffic planning and design*, Chicago Area Transportation Study, 130 North Franklin Street, Chicago.
- Ruiter E.R. (1967), "Toward a better understanding of the intervening opportunities model", *Transportation Research*, **1**, pp. 47-57.
- Schneider M. (1959), "Gravity models and trip distribution theory", *Papers and proceedings of the Regional Science Association*, Vol. **5**, pp. 51-56.
- Stouffer S.A. (1940), "Intervening opportunities : a theory relating mobility and distance", *American Sociological Review*, **5**, pp. 845-867.
- Wills M.J. (1986), "A flexible gravity-opportunities model for trip distribution", *Transportation Research*, **20B**, pp. 89-111.

Modèles gravitaires d'occasions groupées (II)

Le modèle dit de *Fotheringham* développé ci-après emprunte la même philosophie que celui des occasions interposées mais en intégrant une contrainte supplémentaire liée à la prise en compte du phénomène spatial qui en fait une théorie à portée plus générale. Dans la théorie dite des occasions groupées ('*clustered opportunities*'), les occasions sont définies aussi par leur proximité spatiale ce qui peut les rendre plus attractives ('*effet d'agglomération*') ou moins attractives ('*effet de compétition*').

Il est clair que l'hypothèse qui veut, que chaque individu puisse avant de se déplacer, évaluer et donc avoir accès à la totalité de l'information sur les attributs de chaque destination, est irréaliste, dans la mesure où ces derniers ont une capacité de synthèse de l'information réduite (6 alternatives au maximum).

Une hypothèse plus acceptable revient à considérer que chaque individu plutôt que d'évaluer l'ensemble des alternatives disponibles, commence par évaluer des groupes d'alternatives pour enfin au sein d'un groupe classer les alternatives par ordre de préférence.

Ce processus de choix hiérarchique peut être formulé de la manière suivante :

$$N_{ij} = \frac{O_i V_{ij}}{\sum_j V_{ij}}$$

avec $V_{ij} = W_{1j}^{\alpha_{1i}} W_{2j}^{\alpha_{2i}} \dots W_{kj}^{\alpha_{ki}} * d_{ij}^{\beta_i}$

avec :

V_{ij} : somme de tous les attributs de la place j qui affectent le nombre de déplacements au départ de i

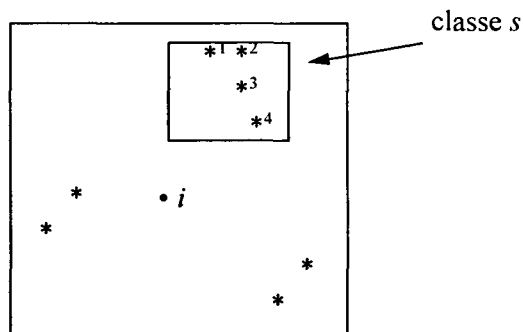
W_{1j} qui représente l'attribut N°1 de la destination j qui affecte son attraction totale.

α_{1i} : paramètre qui reflète la sensibilité de la décision de choix de destination à un changement de l'attribut N°1 de la destination pour une origine située en i .

k : nombre d'attributs

⇒ H_0 : pour des raisons de démonstration, on considérera par la suite que toutes les destinations potentielles sont également attractives pour une origine donnée i , autrement dit : $V_{ij} = V_i, \forall j$

⇒ H_1 : l'exemple suivant sera analysé : modélisation du volume de flux ayant pour origine i et à destination de 1, 2, 3 et 4, destinations qui appartiennent à la classe s .



Dans l'hypothèse d'évaluation de l'ensemble des destinations sans tenir compte de leur appartenance spatiale à une même classe s , l'on aura :

$$V_{is} = n_s V_i \quad (1)$$

V_{is} : attractivité de la classe s des flux ayant pour origine i

n_s : nombre d'alternatives dans la classe s

Autrement dit si le nombre d'alternatives dans la classe s augmente, l'attractivité de la classe augmentera d'un facteur constant V_i pour chaque alternative ajoutée : $\frac{\partial V_{is}}{\partial n_s} = V_i$.

Ce scénario n'est valide que dans le contexte d'alternatives peu nombreuses. Dans le cas contraire d'un grand nombre d'alternatives, le processus de choix sera hiérarchique et l'on aura :

$$V_{is} = (n_s V_i)^\alpha \quad \text{soit} \quad \frac{\partial V_{is}}{\partial n_s} = \alpha n_s^{\alpha-1} V_i^\alpha \quad \text{avec} \quad \alpha > 0$$

* différents cas se produisent alors :

$\alpha = 1$: on est ramené au cas de l'équation (1) et le modèle est correctement spécifié (hypothèse d'un modèle logit)

$\alpha > 1$: les grandes classes (grand nombre de destinations au sein de la classe) sont supposées plus attractives, et les petites classes moins attractives : exemple des centres commerciaux (effet d'agglomération)

$\alpha < 1$: les grandes classes sont supposées moins attractives, et les petites classes plus attractives : exemple des centres commerciaux (effet de compétition).

Reste un problème de fond, comment définir la nature des classes perçues par un individu ? C'est impossible sauf à considérer qu'en termes de perception de l'espace par l'individu, des destinations à proximité d'autres destinations pourront être incluses dans certaines classes. Il convient alors d'introduire cette variable de proximité (coefficient négatif : plus une destination est proche de toutes les autres et plus le flux migratoire sera faible, *ceteris paribus*) dans les équations.

Plusieurs formulations de la proximité sont disponibles, on utilise généralement un critère d'accessibilité potentielle :

$$A_j = \sum_{m \neq j} \frac{W_m}{d_{jm}}$$

A_j : potentiel d'accessibilité de la destination j par rapport aux m autres destinations potentielles

W_m : poids (population)

d_{jm} : distance entre j et m

Le modèle général s'écrit alors :

$$N_{ij} = \frac{O_i V_{ij} A_j^{\gamma_i}}{\sum_j V_{ij} A_j^{\gamma_i}}$$

avec

$$V_{ij} = W_{1j}^{\alpha_{1i}} W_{2j}^{\alpha_{2i}} \dots W_{kj}^{\alpha_{ki}} * d_{ij}^{\beta_i}$$

et γ_i : relation entre les flux de migration et la 'centralité' des destinations, de signe négatif si les migrants sélectionnent leurs destinations de manière hiérarchique et s'ils sous-estiment le nombre d'opportunités dans les grandes classes.

En termes d'utilités individuelles, on aura :

$$U(i, j) = \alpha_1 \ln(W_j) + \alpha_2 d_{ij} + \alpha_3 \ln(A_j) + \varepsilon_{ij}$$

Cette formulation permet de résoudre deux difficultés :

- l'une concerne l'hypothèse IAA : le ratio des probabilités de sélection de deux destinations est indépendant de toutes les autres destinations.

$$\frac{M_{i1}}{M_{i2}} = \frac{W_{1j}^{\alpha_{1i}} W_{2j}^{\alpha_{2j}} \dots W_{kj}^{\alpha_{ki}} d_{i1}}{W_{12}^{\alpha_{1i}} W_{22}^{\alpha_{2i}} \dots W_{k2}^{\alpha_{ki}} d_{i2}} : \text{ratio des probabilités d'un migrant choisissant deux destinations 1 et 2.}$$

Cela ne fait pas de différence si la destination 1 est entourée d'autres alternatives où si elle est isolée ! alors que dans la spécification introduisant une variable de proximité, on tient compte de cet aspect

$$\frac{M_{i1}}{M_{i2}} = \frac{W_{1j}^{\alpha_{1i}} W_{2j}^{\alpha_{2j}} \dots W_{kj}^{\alpha_{ki}} d_{i1} A_1^{\gamma_i}}{W_{12}^{\alpha_{1i}} W_{22}^{\alpha_{2i}} \dots W_{k2}^{\alpha_{ki}} d_{i2} A_2^{\gamma_i}}$$

- l'autre propriété indésirable du modèle général est qu'il est impossible pour la probabilité d'un migrant ayant sélectionné une alternative de croître avec l'ajout d'une nouvelle alternative dans le système. Ce n'est plus le cas dans la spécification intégrant la variable de proximité.

Le concept de modèles hiérarchiques n'est pas nouveau (cf. nested logit model (*Ben-Akiva et Lerman* (1985))), cependant les '*competing destinations models*' intègrent de façon explicite un modèle de choix spatial qui prend en compte le processus de choix hiérarchique des individus mais aussi le fait qu'ils sous-estiment généralement la taille de grandes classes d'alternatives. Un de leurs principaux avantages est de ne pas rendre nécessaire de spécifier pour chaque individu la forme de la représentation mentale hiérarchique de son espace.

Bibliographie

- Ewing G. (1986), "Spatial pattern in distance deterrence parameters and Fotheringham's theory of competing destinations", *Environment and planning A*, vol. **18**, pp. 547-551.
- Fotheringham (1983), "Modelling Hierarchical destination choice", *Environment and planning A*, vol. **18**, pp. 401-418.
- Fotheringham (1986), "Further discussion on the distance-deterrence parameters and the competing destination model", *Environment and planning A*, vol. **18**, pp. 553-556.
- Jayet H. (1990), "Spatial search process and spatial interaction : polarisation, concentration and spatial search equilibrium", *Environment and planning A*, vol. **22**, pp. 719-732.
- Jayet H. (1990), "Spatial search process and spatial interaction", *Environment and planning A*, vol. **22**, pp. 583-599.
- Kanaroglou Pavel, S. Ferguson Mark R. (1996), "Discrete spatial model for aggregate destination", *The Regional Science Research Corporation*, **36**, **2**, pp. 271-290.
- Lo L. (1992), "Destination interdependence and the competing destinations model", *Environment and planning A*, vol. **24**, pp. 1191-1204.
- Masahisa Fujita and Tony E. Smith (1990), "Additive-interaction models of spatial agglomeration", *Journal of regional science*, No 1, vol. **30**, pp. 51-74.
- Putman and Chung (1989), "Effetti della definizione di sistema spaziale sui modelli di interazione spaziale-1: Il problema di definizione del sistema", *Environment and planning A*, vol. **21**, pp. 27-46.

FICHE N°3

MIS ET FORMULATION POISSONNIENNE

Les données sur le comportement spatial des ménages sont typiquement des données de type non continu : le choix modal et de destination sont des variables qualitatives et la fréquence de déplacement est une variable discrète positive. Pour expliquer une telle variable discrète par d'autres variables exogènes qualitatives ou quantitatives, le modèle linéaire classique (Classical Linear Regression Model) se révèle inopérant : le nuage des observations n'a pas une forme adaptée à un ajustement linéaire, l'hypothèse de normalité ne peut être faite, puisque la variable endogène discrète prend un petit nombre de valeurs avec des probabilités strictement positives, les formules de prévision aboutissent alors à des valeurs impossibles à prendre.

Flowerdew et Aitkin dans leur célèbre article préconisent alors l'utilisation d'une méthode d'estimation basée sur le processus de Poisson.

Rappel :

Les modèles poissonniens appartiennent à la famille des modèles linéaires généralisés (GLIM).

Sachant x_i et que la loi de Y_i doit appartenir à la famille exponentielle, l'on a :

$$E(Y_i) = \mu_i = g^{-1}(x_i \beta) \quad \text{avec } g = \text{fonction de lien } (g^{-1} = F)$$

Deux choix sont alors à déterminer : le choix de la loi exacte de la famille exponentielle F et le lien g.

- si l'on retient la loi de Poisson (loi P à 1 paramètre λ) qui appartient à la famille exponentielle on a dans le cas discret :

$$P_\theta [Y_i = k] = e^{-\lambda} (\lambda^k / k!) = \exp[k * \log \lambda - \lambda - \log(k!)] = \exp[k * \theta - e^\theta - \log(k!)]$$

$$\text{avec : } E(y_i) = V(y_i) = b'(\theta) = b''(\theta) = e^\theta = \lambda$$

- si $g = \log$ alors $F = \exp$ on obtient un modèle log linéaire poissonnien.

On aboutit dans le cadre transport à la formulation suivante :

$$T_i \succ f(t_i) = \frac{e^{-\lambda_i} \lambda_i^{t_i}}{t_i!}; t_i = 0, 1, 2, \dots$$

$$\text{avec } E(T_i) = \lambda_i = \exp(X_i' \beta); i = 1, \dots, n$$

Si l'on revient maintenant au cadre des MIS, l'utilisation de la spécification dite Poissonnienne, revient à faire l'hypothèse que le phénomène observé apparaît comme la superposition d'événements rares, autrement dit dans le cas d'un modèle de flux de marchandises d'une région i vers une région j :

1. à chaque instant, le départ d'un camion de la région i vers la région j est parfaitement imprévisible, c'est à dire indépendant de ce qui a précédé,
2. les départs se font un à un,
3. dans un intervalle de temps fixé, les chances de départ sont contraintes.

Le choix de ce type de spécification est justifié lorsque la variable étudiée décrit le nombre de fois où un événement s'est produit pendant une certaine période et quand les hypothèses habituelles du processus de Poisson sont satisfaites :

Exemple : nombre d'arrivées journalières dans un aéroport en fonction de la date, nombre d'accidents de la route sur un tronçon en fonction des caractéristiques de celui-ci, nombre de trajets pour un motif particulier durant une période de temps déterminée, etc..

Quant aux hypothèses habituelles des formulations de Poisson :

- une des hypothèses fondamentales concerne l'indépendance entre le présent et le passé : les événements à prédire doivent s'être produits de manière indépendante dans le temps. C'est une hypothèse très restrictive, dans la mesure où l'occurrence d'un événement peut très bien conditionner la probabilité d'occurrence de l'événement suivant en l'accroissant ou en la réduisant.
- l'autre a trait à l'hypothèse d'égalité de la variance et de l'espérance. Si tel n'est pas le cas, les données seront dites sous- ('Underdispersion', Var. < Esp.) ou sur-dispersées ('Overdispersion', Var. > Esp.).

Pour pallier à cet effet on peut modéliser le paramètre λ comme une distribution Gamma de densité : $\Gamma(\cdot)$ est la fonction Gamma

$$T_i \succ g(t_i) = \frac{\Gamma(t_i + \gamma_i)}{\Gamma(t_i + 1)\Gamma(\gamma_i)} \left(\frac{\gamma_i}{\gamma_i + \mu_i}\right)^{\gamma_i} \left(\frac{\mu_i}{\gamma_i + \mu_i}\right)^{t_i} \quad t_i = 0, 1, \dots$$

Cameron et Trivedi (1986) ont ajouté :

$$\gamma_i = \exp(X_i' \beta) \quad \gamma = \frac{1}{\alpha} [\exp(X_i' \beta)]^k, \alpha > 0$$

$$E(T_i) = \exp(X_i' \beta) \quad \text{Var}(T_i) = E(T_i) + \alpha [E(T_i)]^{2-k}$$

On s'aperçoit alors qu'espérance et variance ne sont plus contraintes à être égales, les paramètres α , k , déterminant la forme de la relation entre espérance et variance.

Si l'on pose $\alpha = 0$ (H_0) et ceci $\forall k$: on est ramenés au cas de l'hypothèse d'absence de sous- ou sur-dispersion et donc au modèle poissonnien.

Rickard (1988) compare le recours à la distribution de Poisson avec celle de la loi binomiale négative dans le cadre des trajets à longue distance en train. Elle trouve que la distribution binomiale négative est plus appropriée, la distribution totale étant la somme des distributions de poisson, de segments la composant.

Il reste que même si ces hypothèses ne sont pas vérifiées, le modèle poissonnien peut néanmoins être valable en première approximation, si l'observation a lieu sur un intervalle de temps limité.

Suivant le type de fonction d'interaction spatiale, et la présence ou non de contraintes, divers résultats sont disponibles :

A. cas d'un MIS sans contrainte :

Si l'on reprend les notations générales d'un MIS sans contrainte :

$$N_{ij} = k \cdot E_i A_j F(c_{ij})$$

- dans l'approche dite Pareto où $F(c_{ij}) = \exp[-\theta \cdot c_{ij}]$, l'on peut écrire que le modèle poissonnien

$$\text{sous la forme : } N_{ij} = \exp(\beta_0 + \beta_1 \ln E_i + \beta_2 \ln A_j + \theta \ln c_{ij}) + \varepsilon_{ij}$$

$$\text{avec déviance} = 2 * \left[\sum_i \sum_j N_{ij}^* \cdot \text{Log}(N_{ij} / N_{ij}^*) \right]$$

on considérera alors que $\beta_1 = \beta_2 = 1$ (propriété dite de 'coût total contraint')

Dans ce cadre on aura la propriété suivante : $\sum_{ij} d_{ij} = \sum_{ij} \hat{d}_{ij}$ ou encore $\bar{d}_{ij} = \bar{\hat{d}}_{ij}$

- dans le cadre d'une fonction puissance de la fonction d'interaction spatiale

$$F(c_{ij}) = (c_{ij})^{-\theta} \quad (\theta > 0)$$

l'on aura la propriété suivante :

$$\left[\sum_i \sum_j N_{ij} \cdot \ln(c_{ij}) \right] = \left[\sum_i \sum_j \hat{N}_{ij} \cdot \ln(c_{ij}) \right]$$

- pour une fonction dite 'combinée' : $F(c_{ij}) = \exp(-\theta \ln(c_{ij}))$, l'on aura :

$$\sum_{ij} \ln(\hat{c}_{ij}) = \sum_{ij} \ln(c_{ij})$$

B. cas d'un MIS avec contraintes :

La formulation générale (doubles contraintes) est la suivante :

$$N_{ij} = k \cdot E_i H_i A_j G_j F(c_{ij})$$

1. cas du modèle avec contrainte sur les origines :

$$\text{on aura : } \beta_0 + \beta_i = \ln H_i + \ln E_i$$

2. cas du modèle avec contrainte sur les destinations :

$$\text{on aura : } \beta_0 + \beta_j = \ln G_j + \ln A_j$$

3. cas du modèle avec doubles contraintes :

$$\beta_0 + \beta_i + \beta_j = \ln H_i + \ln E_i + \ln G_j + \ln A_j$$

Remarques générales

sur les avantages et inconvénients d'une formulation poissonnienne :

1. Les avantages de la formulation de Poisson

- les avantages sont tout d'abord ceux inhérents aux *GLIM* (estimation par maximum de vraisemblance, ...), en particulier dans le cadre de la prévision du nombre de déplacements pour tel motif, on aboutit à des valeurs positives et à des probabilités.
- dans le cadre des MIS, la somme des flux observés est égale à la somme des flux estimés (*propriété de conservation des flux*).
- il est possible afin d'améliorer le modèle d'introduire des contraintes sur les origines et/ou sur les destinations et d'en faire un modèle doublement contraint. Notons que si de telles variables (*categorical variables*) sont introduites dans le modèle les facteurs d'origine et/ ou destination ne peuvent pas l'être en même temps.
- prendre en compte les interactions entre variables (notamment celle entre distance physique et distance d'appartenance).

2. Les limites

- hypothèse de départ : la variable étudiée décrit le nombre de fois où un événement s'est produit pendant une certaine période de temps et l'espérance du processus = sa variance. Les flux se font de manière indépendantes et avec la même probabilité, hypothèse qui est celle des MIS classiques. Hypothèse forte d'absence de coopération / jeux entre individus.
- risque d'« *overdispersion* », cas où la déviance est supérieure à la valeur critique du Chi2 (cas de tableaux grands)
 - omission de variables exogènes

- un processus de Poisson non adéquat : cas par exemple de flux migratoires qui ne sont plus indépendants et de probabilité inégale. Ainsi les déplacements de personnes au sein d'un même ménage ne sont pas indépendants.
- Risque d'« Underdispersion » la déviance est inférieure à la valeur critique du Chi2.
 - * matrices OD clairsemées (beaucoup de valeurs nulles / faibles valeurs).
 - * multicollinéarité entre variables exogènes: analyser les interactions entre exogènes.
- Présence d'autocorrélation spatiale.
- Modèles de taille de ménage

Bibliographie

- Flowerdew R., Aitkin M. (1982), "A method of fitting the gravity model based on the Poisson distribution", *Journal of Regional Science*, **22(2)**, pp. 191-202.
- Gouriéroux C. (1989), *Econométrie des variables qualitatives*, chapitre X, pp. 297-305, *Collection Economie et Statistiques Avancées*, 2^{ème} édition, Economica.
- Barmby T., Doornik J. (1989), "Modelling trip frequency as a poisson variable", *Journal of Transport Economics and Policy*, septembre.
- Cameron A.C., Trivedi P.K. (1986), "Econometric models based on count data : comparisons and applications of some estimators and tests", *Applied Econometrics*, (1), pp. 29-53.
- Rickard J.M. (1988), "Factors influencing long distance rail passengers trip rates in Great Britain", *Journal of Transport Economics and Policy*, **22**, pp. 209-233.

FICHE N°4

UNE APPROCHE COMBINÉE DISTRIBUTION-AFFECTATION

Plutôt que d'analyser des matrices OD de flux (trafics) de manière générale, il apparaît plus réaliste dans le domaine des transports de tenir compte que ces flux (trafics) sont affectés sur le réseau.

En faisant l'hypothèse que l'ensemble des états d'un système ont la même probabilité d'apparaître (hypothèse de *Boltzmann-Wilson*), le nombre d'états associés à une distribution particulière $N(T)$ peut s'écrire :

$$N(T) = \frac{\sum_i \sum_j N_{ij}!}{\prod_{i,j} N_{ij}!} = \frac{N!}{\prod_{i,j} N_{ij}!}$$

avec T : nombre total de flux dans le système

L'état le plus stable de la matrice OD est l'ensemble qui maximise le nombre d'états et peut être formulé de la façon suivante :

$$\text{Max} N(T) = \frac{N!}{\prod_{i,j} N_{ij}!}$$

sous les contraintes :

$$\sum_j N_{ij} = N_i, \forall i$$

$$\sum_i N_{ij} = N_j, \forall j$$

En prenant le logarithme de la fonction ci-dessus, l'on obtient :

$$\text{Max} \ln N(T) = \ln \frac{N!}{\prod_{i,j} N_{ij}!} = \ln N! - \sum_{i,j} \ln N_{ij}!$$

L'approximation de *Stirling* nous donne : $\ln(N_{ij}!) = N_{ij} \ln(N_{ij}) - N_{ij} + o(N_{ij})$

$$\text{où } \lim_{N_{ij} \rightarrow \infty} \frac{1}{N_{ij}} o(N_{ij}) = 0$$

En négligeant le terme $o(N_{ij})$, l'équation peut s'écrire sous la forme d'un modèle des distribution entropique classique :

$$\text{Max} - \sum_{i,j} (N_{ij} \ln N_{ij} - N_{ij})$$

sous les contraintes :

$$\sum_j N_{ij} = N_i, \forall i$$

$$\sum_i N_{ij} = N_j, \forall j$$

Comme la maximisation d'une fonction d'objectif équivaut à la minimisation de la forme négative de la fonction d'objectif et si nous couplons ce modèle de distribution avec le problème de transport à la *Hitchcock's*, un modèle combiné de distribution-affectation peut être écrit sous la forme :

$$\text{Min} \sum_{ij} c_{ij} N_{ij} + \frac{1}{\beta} \sum_{ij} (N_{ij} \ln N_{ij} - N_{ij}) \quad (1)$$

sous les contraintes :

$$\sum_j N_{ij} = N_i, \forall i \quad (2)$$

$$\sum_i N_{ij} = N_j, \forall j \quad (3)$$

$$N_{ij} \geq 0, \forall i, j \quad (4)$$

Suivant *Wilson*, la résolution des équations ci-dessus nous donne :

$$N_{ij} = A_i B_j N_i N_j e^{-\beta c_{ij}}, \forall i, j \quad (5)$$

$$A_i = \frac{1}{\sum_j B_j N_j e^{-\beta c_{ij}}}, \forall i \quad (6)$$

$$B_j = \frac{1}{\sum_i A_i N_i e^{-\beta c_{ij}}}, \forall j \quad (7)$$

Le paramètre β , coefficient de friction spatiale est un indicateur de dispersion spatiale des trafics hors de la solution de *Hitchcock*. Quand β est petit, la dispersion est grande, ce qui signifie un grand nombre d'interactions entre les unités spatiales. Quand β est grand, la dispersion est faible, ce qui signifie les flux sont concentrés sur un petit nombre de paires OD. Quand β tends vers l'infini, les équations (1), (2), (3), (4) reviennent à la résolution du *problème de Hitchcock* :

$$\sum_{ij} c_{ij} N_{ij} + \lim_{\beta \rightarrow \infty} \frac{1}{\beta} \sum_{ij} N_{ij} \ln N_{ij} = \sum_{ij} c_{ij} N_{ij}$$

La valeur de β peut être estimée par des méthodes de calibrage, par exemple en utilisant les équations (5).

La variation des N_{ij} peut être définie comme un écart à la moyenne, et donc la somme des carrés de ces écarts peut encore s'écrire :

$$\sum_k (N_k - \bar{N})^2 = \sum_k (\hat{N}_k - \bar{N})^2 + \sum_k e_k^2$$

Les équations simultanées (5), (6), (7), du modèle doublement contraint peuvent être résolues par une procédure du type D.S.F. (*Deming-Stephan-Furness*). La procédure a été introduite par *Kruithof* (1937), incorporée ensuite dans la littérature statistique par *Deming* et *Stephan* (1940), et dans la littérature transport par *Furness* (1965).

Cette procédure est équivalente à la méthode RAS (*Bachrach*, 1970) de la littérature économique ; encore appelée *Iterative Proportional Fitting Procedure* (IPF). *Goodman* (1968) a amélioré l'algorithme.

Première itération :

Pour satisfaire à la contrainte de la conservation des flux de l'équation (2), la somme lignes des flux doit être satisfaite, ce qui revient à :

$$\hat{N}_i = N_i \sum_j N_j e^{-\beta c_{ij}} \quad (8)$$

$$\hat{A}_i = \frac{N_i}{\hat{N}_i} = \frac{1}{\sum_j N_j e^{-\beta c_{ij}}}$$

ce qui revient à mettre $B_j=0$ et résoudre A_i .

L'équation (8) nous garantit que la contrainte lignes de l'équation (2) est satisfaite, mais ne nous dit rien sur la contrainte colonnes. On procède donc de manière identique pour la contrainte (3) :

$${}^1 N_j = N_j \sum_i N_i e^{-\beta c_{ij}} {}^1 A_i \quad (9)$$

$${}^1 B_j = \frac{N_j}{{}^1 N_j} = \frac{1}{\sum_i N_i e^{-\beta c_{ij}} {}^1 A_i}$$

Deuxième itération :

- conservation des sommes-lignes :

$${}^2 N_i = N_i \sum_j N_j e^{-\beta c_{ij}} {}^1 A_i {}^1 B_j \quad (8)$$

$${}^2 A_i = \frac{N_i}{{}^2 N_i} = \frac{1}{\sum_j N_j e^{-\beta c_{ij}} {}^1 A_i {}^1 B_j}$$

- conservation des sommes-colonnes :

$${}^2 N_j = N_j \sum_i N_i e^{-\beta c_{ij}} {}^1 A_i {}^1 B_j {}^2 A_i$$

$${}^2 B_j = \frac{N_j}{{}^2 N_j} = \frac{1}{\sum_i N_i e^{-\beta c_{ij}} {}^1 A_i {}^1 B_j {}^2 A_i}$$

et ainsi de suite

Le processus se continue jusqu'à ce que la différence entre les flux observés et estimés N_i et N_j sont inférieurs à un seuil ϵ prédéterminé.

Le modèle combiné distribution-affectation peut donc s'écrire :

$$N_{ij} = A_i B_j N_i N_j e^{-\beta c_{ij}}, \forall i, j$$

sous les contraintes :

$$A_i = {}^1 A_i {}^2 A_i {}^3 A_i \dots$$

$$B_j = {}^1 B_j {}^2 B_j {}^3 B_j \dots$$

Bibliographie

- Erlander S., Stewart F. (1985), The gravity model in transportation analysis : theory and extensions, chap. 8, pp. 175-199, Topics in Transportation, Utrecht, The Netherlands, Tokyo, Japan.
- Goodman, L.A. (1968), "The analysis of cross-classified data : independence, quasi-independence and interaction in contingency tables with or without missing cells", *Journal of the American Statistical Association*, **63**, pp. 1091-1131.
- Furness, K.P.(1965), "Time function iteration", *Traffic Engineering and Control*, **7**, pp. 458-460.
- Deming, W.E., Stephan, F.F. (1940), "On a least squares adjustment of sampled frequency tables when the expected marginal tables are known", *Annals of Mathematical Statistics*, **11**, pp. 427-444.

FICHE N°5

MIS ET ENTROPIE

Le Principe de l'Entropie Maximum (PEM) _

Le Principe de l'Entropie Maximum (PEM) développé par *Jaynes* (1957) généralise le fameux "Principe de la Raison Suffisante" de *Laplace*.

Formulation :

On considère une variable aléatoire X où x_i , ($i = 1, \dots, n$) représente les valeurs numériques de la variable aléatoire X et p_i , ($i = 1, \dots, n$) les probabilités correspondantes d'apparition des valeurs x_i .

$$p_i \geq 0 ; \sum_{i=1}^n p_i = 1$$

Définition du problème :

Etablir la répartition p_i , sans connaître les fréquences relatives $f(x_i) = f_i$ de chaque événement i , en connaissant seulement quelques moments de la variable aléatoire X .

Solution : recourir aux probabilités subjectives.

L'école des subjectivistes regarde les probabilités comme une expression de notre ignorance. La probabilité d'un événement constitue l'expression formelle de l'espérance que l'on a vis-à-vis de l'occurrence de cet événement (et non comme une propriété propre de cet événement basée sur l'expérience). Autrement dit :

- un test objectif énoncera : 'cette distribution représente-t-elle des fluctuations observables ?'. On recherchera alors une méthode d'échantillonnage qui évite au mieux les biais.
- un test subjectif : 'cette distribution représente-t-elle correctement l'état de notre connaissance ?'.

On recherchera alors une distribution de probabilité la moins biaisée possible étant donné l'information disponible. Ce concept sert notamment quand il s'agit d'établir des fonctions de répartition pour des événements constitutifs de phénomènes aléatoires non directement observables, c'est le cas de la majorité des MIS.

On ne calculera donc pas p_i à partir des données d'expériences mais à partir d'un principe variationnel qui consiste à exprimer l'état de notre connaissance du phénomène en ces termes : 'Compte tenu de l'information disponible, le degré de notre incertitude est maximum'. En utilisant le résultat de *Shannon* qui fait de l'entropie une mesure de l'incertitude, *Jaynes* formule mathématiquement que l'entropie $H(p_i)$ de la distribution p_i est maximum compte tenu de l'information disponible sous la forme de moments statistiques.

Dans le cas du moment de premier ordre, c'est à dire l'espérance mathématique $E(X)$ on aura :

$$\text{Max}H(p_i) = -K \sum_{i=1}^n p_i \text{Log}p_i$$

$$\text{avec : } \sum_{i=1}^n p_i = 1 \quad \text{et} \quad \sum_{i=1}^n p_i f_i = E(X)$$

On aboutit au résultat suivant :

$$p_i = \frac{1}{\phi(\beta)} \exp(-\beta f_i) \quad \text{où} \quad \beta \text{ est solution unique de : } \frac{\phi'(\beta)}{\phi(\beta)} = -E(X)$$

$$\text{avec } \phi(\beta) = \sum_{i=1}^n \exp(-\beta f_i)$$

La généralisation du PEM et son application aux MIS

Généraliser le PEM consiste à maximiser une fonction d'entropie sous un ensemble de contraintes quelconques, contraintes qui peuvent différer des formulations mathématiques des moments statistiques. C'est la démarche suivie par *Wilson* (1968,1969).

Rappel de la première formulation de Wilson :

Soit N_{ij} le nombre de déplacements d'une zone d'origine i vers une zone de destination j .

H₀: On suppose que le nombre de déplacements E_i émanant d'une zone i et le nombre de déplacements A_j arrivant dans une zone j sont connus, on aura alors :

$$\sum_{j=1}^n N_{ij} = E_i, \quad (i = 1, n) \quad (1) \quad \text{et} \quad \sum_{i=1}^n N_{ij} = A_j, \quad (j = 1, n) \quad (2)$$

H₁: On suppose également que le coût total C engendré par l'ensemble de tous les déplacements est une grandeur fixe, soit :

$$\sum_{i=1}^n \sum_{j=1}^n N_{ij} c_{ij} = C \quad (3) \quad \text{avec } c_{ij} : \text{coût moyen pour se déplacer de la zone } i \text{ vers la zone } j$$

Wilson énonce alors que 'La probabilité qu'une distribution (N_{ij}) survienne sur le réseau est proportionnelle au nombre d'arrangements que l'on peut effectuer entre les déplacements et qui donnent naissance à une telle distribution'. Dans cette première formulation, *Wilson* n'utilise que la seule sémantique de l'analyse combinatoire.

Si on note $T = \sum_i E_i = \sum_j A_j$ le nombre total de déplacements, alors le nombre d'arrangements

$$\text{distincts qui donne naissance à la distribution } (N_{ij}) \text{ s'écrit : } W(N_{ij}) = \frac{N!}{\prod_i \prod_j N_{ij}!}$$

Wilson maximise cette quantité appelée 'permutabilité' sous les contraintes (1), (2), (3) et en passant par la maximisation du Lagrangien, on obtient :

$$N_{ij} = E_i H_i A_j G_j \exp(-\beta c_{ij})$$

$$H_i = \left[\sum_j A_j G_j \exp(-\beta c_{ij}) \right]^{-1}$$

$$G_j = \left[\sum_i E_i H_i \exp(-\beta c_{ij}) \right]^{-1}$$

Cette formulation devait être reprise dans le cadre de la statistique mécanique par J.A. *Tomlin* et S.G. *Tomlin* (1968), dans celui de la théorie de l'information (*Wilson*, 1969).

Reformulations

On peut reformuler le modèle élémentaire de Wilson tel que décrit plus haut à partir du PEM : 'Maximiser une permutableté revient à maximiser une fonction d'entropie telle que définie par C.E. Shannon'.

On obtient alors :

$$\text{Max}H(n_{ij}) = -\sum_i \sum_j n_{ij} \text{Log}n_{ij}$$

$$\text{avec : } \sum_i n_{ij} = d_j \quad \sum_j n_{ij} = o_i \quad \sum_i \sum_j c_{ij} n_{ij} = \bar{C}$$

$$\text{et } d_j = \frac{D_j}{N} \quad o_i = \frac{O_i}{N} \quad \bar{C} = \frac{C}{N}$$

Les trois ensembles de contraintes représentent toute l'information géographico-économique disponible sur le réseau. A l'exception de cette information, l'incertitude sur l'état du système de distribution de déplacements est maximum.

La fonction d'objectif et l'ensemble des contraintes constituent ce que l'on appelle un programme d'interaction spatiale (PIS). La solution du PIS conduit généralement à une famille de distributions théoriques.

A partir de ceci on peut aboutir à des formalisations de MIS plus ou moins complexes suivant la quantité d'information disponible :

- cas d'une interaction non contrainte :
 - le PIS est réduit à sa fonction objective :

$$\text{Max}H(n_{ij}) = -\sum_i \sum_j n_{ij} \text{Log}n_{ij}$$

avec :

$$\sum_i \sum_j n_{ij} = 1 \quad , \quad n_{ij} \geq 0 \quad \text{avec} \quad n_{ij} = \frac{N_{ij}}{\sum_i \sum_j N_{ij}} \quad (i, j = 1, \dots, n)$$

un tel PIS correspond à la solution du Principe de la Raison Insuffisante de Laplace où la distribution des déplacements est uniforme, d'où : $N_{ij} = \frac{1}{n^2}$.

- introduction de la seule contrainte de coût :

Soit (c_{ij}) la matrice des coûts unitaires de déplacements et C budget constant des déplacements,

$$C = \sum_{i=1}^n \sum_{j=1}^n c_{ij} N_{ij} .$$

Le PIS s'écrit alors :

$$\text{Max}H(n_{ij}) = -\sum_i \sum_j n_{ij} \text{Log}n_{ij}$$

avec :

$$\sum_i \sum_j n_{ij} = 1 \quad , \quad n_{ij} \geq 0 \quad (i, j = 1, \dots, n)$$

$$\sum_{i=1}^n \sum_{j=1}^n c_{ij} n_{ij} = \bar{C}$$

après résolution du *lagrangien* on aboutit à la solution :

$$n_{ij} = k * \exp(-\beta c_{ij}) \quad \text{avec} \quad k = \left[\sum_i \sum_j \exp(-\beta c_{ij}) \right]^{-1}$$

Ainsi en l'absence de contraintes additionnelles de déplacements, les individus ont une propension à se déplacer qui varie en raison inverse à l'exponentielle du coût de déplacement.

La connaissance du nombre total de déplacements N permet de déduire, à partir de la distribution des déplacements (N_{ij}) :

$$N_{ij} = n_i n_j N, \quad N_{ij} = \frac{N}{n^2},$$

$$N_{ij} = k * \exp(-\beta_{cij}) \quad \text{avec } k = N * \left[\sum_i \sum_j \exp(-\beta_{cij}) \right]^{-1}$$

- interaction mono-contraite :

On peut prendre comme type de contraintes :

- contrainte globale : $\sum_{i=1}^n \sum_{j=1}^n N_{ij} = N$

La connaissance des n valeurs de O_i implique la connaissance du nombre total de déplacements, et

l'on aboutit à : $\sum_{i=1}^n O_i = \sum_{j=1}^n D_j = N,$

$$N_{ij} = k * \exp(-\beta_{cij}) \quad \text{avec } k = N * \left[\sum_i \sum_j \exp(-\beta_{cij}) \right]^{-1}$$

- contrainte sur les origines : $\sum_{j=1}^n N_{ij} = O_i$

$$N_{ij} = A_i * O_i * \exp(-\beta_{cij}) \quad \text{avec } A_i = \left[\sum_{j=1}^n \exp(-\beta_{cij}) \right]^{-1}, \quad \beta = \lambda N, \text{ multiplicateur de Lagrange}$$

associé à la contrainte de coût total et A_i facteur d'équilibre qui assure le respect de la contrainte d'émission.

- contrainte sur les destinations : $\sum_{i=1}^n N_{ij} = D_j$

$$N_{ij} = B_j * D_j * \exp(-\beta_{cij}) \quad \text{avec } B_j = \left[\sum_{i=1}^n \exp(-\beta_{cij}) \right]^{-1}, \quad \beta = \lambda N, \text{ multiplicateur de Lagrange}$$

associé à la contrainte de coût total et B_j facteur d'équilibre qui assure le respect de la contrainte d'attraction (ex. de modèles complexes de localisation résidentielle)

- PIS multi-contraintes : il s'agit ici du modèle élémentaire de *Wilson* à deux contraintes, ou de modèles plus complexes du type :
 - prise en compte des points de transit (exportations/importations/capacités portuaires),
 - prise en compte de contraintes technico-géographiques : analyse de matrices input-output,
 - etc. .

Bibliographie

- Wilson A.G. (1967), "A statistical theory of spatial distribution models", *Transportation Research*, Vol. 1, pp. 253-269.
- Wilson A.G. (1969), "The use of entropy maximising models", *Journal of transport economics and policy*, vol. 3, pp. 108-126.
- Wilson A.G. (1970), "Advances and problems in distribution modelling", *Transportation Research*, vol. 4, pp. 1-13.
- Wilson A.G. (1970), "Inter-regional commodity flows: entropy maximising methods", *Geographical Analysis*, vol. 2, pp. 255-282.
- Wilson A.G. (1973), "Further developments of entropy maximising transport models", *Transportation Planning and Technology*, vol. 1, pp. 183-193.
- Wilson S.R. (1976), "Statistical notes on the evaluation of calibrated gravity models", *Transportation Research*, Vol. 10, pp. 343-345.
- Jaynes E.T. (1957), "Information theory and statistical mechanics", *Physical Review*, vol. 106, pp. 620-630.
- Jaynes E.T. (1957), "Information theory and statistical mechanics", *Physical Review*, vol. 108, pp. 171-180.

FICHE N°6

LES MIS DYNAMIQUES

Il serait présomptueux de vouloir ici résumer l'ensemble des travaux menés sur le sujet, cependant un certain nombre d'éléments sont connus sur les propriétés dynamiques des MIS :

* l'analogie formelle existant entre les MIS et les modèles de choix discret (DCM's) peut être étendue au contexte dynamique. Un MIS peut être vu comme la solution optimale d'un problème d'interaction spatiale dynamique (MDIS).

* un MIS peuvent aussi expliquer les interactions entre activités dans un processus dynamique stochastique, exemple des MIS à simple contrainte qui présentent une structure stable même en présence de perturbations mineures (cf. théorie des catastrophes).

Analogie entre les MIS et la fonction logistique

Si nous considérons dans un contexte dynamique l'équivalence entre MIS et modèles de choix discrets, nous pouvons écrire la probabilité de choix de l'alternative j au temps t :

$$P_j = \frac{\exp(u_j)}{\sum \exp_i(u_i)}$$

Ainsi le taux de changement de P_j au cours du temps peut être exprimé sous la forme d'une structure de type *Volterra-Lotka* : $\dot{P}_j = \dot{u}_j P_j (1 - P_j) - P_j \sum_{j \neq i} \dot{u}_i P_i$

On s'aperçoit que le premier terme de cette équation est équivalente à la formulation d'un modèle logistique de croissance (cas d'un choix binaire). Cette formulation est bien connue par sa capacité à générer bifurcations et comportements chaotiques pour certaine valeurs des paramètres de croissance.

Chaos et bifurcation (*Bifurcation Theory* (BT))

La description de systèmes dynamiques se caractérise par un équilibre multiple qui peut basculer d'un point d'équilibre à un autre et produire des discontinuités. Ces discontinuités sont des propriétés du système plutôt que le résultat de chocs externes.

Lorsque plusieurs équilibres existent dans un système, la trajectoire de ce dernier dépendra de la configuration initiale et de la valeur des paramètres. Il suffit de fait d'un léger changement (perturbation) dû à une transformation de l'environnement dans lequel fonctionne le système, qui entraîne une petite variation de ces conditions initiales, ou un léger changement de la valeur d'un paramètre, ou l'effet d'une simple perturbation aléatoire, pour que l'on traverse la 'séparatrice' (ligne séparant deux bassins d'attraction de points d'équilibre). Les bifurcations sont d'autant plus nombreuses qu'elles concernent une échelle de niveau inférieur :

- nombreuses et souvent irréversibles au niveau de l'individu (naissance, déménagement, ...)
- à l'échelle d'une zone, la bifurcation exige du temps et se produit plus rarement (croissance au déclin d'une ville, choix de lieu de délocalisation d'usine.

I. Prigogine et I. Stengers (1979) soulignent que « la taille critique (ex. d'une ville) est d'autant plus élevée, et la fluctuation déstabilisante est donc d'autant plus rare, que la diffusion qui couple les régions du système - et en particulier la région fluctuante et son environnement - est élevée. En d'autres termes, plus rapide est la communication dans le système, plus grande est la proportion des fluctuations insignifiantes, incapables de transformer l'état du système : plus stable est cet état ».

Le système tend à amortir la plupart des fluctuations et il le fait d'autant mieux que l'information circule bien entre les composantes du système.

Plusieurs études analytiques de modèles dynamiques ont montré que les conditions susceptibles de mener vers une évolution de type chaotique sont très restrictives. La grande rareté des bifurcations a ainsi été démontrée pour les systèmes migratoires, à l'échelle inter-régionale ou interurbaine (Haag, 1989).

Formalisation de la dynamique d'un système de zones spatiales : la méthode des équations maîtresses

Il s'agit ici de décrire la dynamique d'un ensemble de zones spatiales en observant l'évolution de leur population.

Chaque zone est caractérisée par la taille de sa population.

A un instant t donné, on peut caractériser l'état du système spatial par un vecteur n tel que :

$n = (n_1, n_2, \dots, n_L)$ décrivant la répartition de la population totale N entre les L zones formant le système, n_i représente la population de la zone i .

L'analyse dynamique de ce système se traduit par l'étude de l'évolution du vecteur n dans le temps.

Deux possibilités se présentent :

- étudier l'évolution $\frac{dn_i}{dt}$ de la population de chacune des zones du système

étudier l'évolution de la probabilité $P(n, t)$ pour qu'à un instant t donné, on soit en présence d'une configuration $n(t)$ particulière : $n(t) = (n_1(t), n_2(t), \dots, n_L(t))$ avec $\sum_i n_i(t) = N(t), i = 1, \dots, L$.

A chaque configuration $n(t)$ on associe une probabilité d'apparition $P(n, t)$ avec $\sum_{\{n\}} P(n, t) = 1$

où $\{n\}$ représente l'ensemble de toutes les configurations possibles.

Ex. :

- La configuration $(0, 0, \dots, 0, N, 0, \dots, 0)$ décrit la configuration totale de toute la population sur un seul espace.
- La configuration $(N/L, N/L, \dots, N/L)$ décrit un système homogène avec équi-répartition.

L'équation maîtresse (Master equation) décrit la dynamique $\frac{dP(n, t)}{dt}$ de la probabilité $P(n, t)$

Le processus migratoire résulte de la conjugaison de deux processus simultanés :

1. les échanges avec l'extérieur du système : un solde global positif se traduit par une augmentation de la population urbaine totale N du système.
2. les échanges entre les zones du système qui influent sur la seule redistribution de la population entre les zones mais non sur la population totale N du système.

Si l'on considère un intervalle de temps dt suffisamment court pour qu'un seul individu migre d'une zone à l'autre entre t et $t + dt$ et en supposant que cette migration s'opère de la zone j vers la zone i , notons W_{ij} le flux migratoire de j à i , on aura :

$$n(t + dt) = n'_{ij} = \{n_1, n_2, \dots, n_i + 1, \dots, n_j - 1, \dots, n_L\}$$

Le processus de passage de la configuration $n(t)$ à la configuration $n(t + dt)$ peut dès lors être analysée à deux niveaux :

* au niveau microscopique des individus résidant dans les zones considérées :

On note $P(i, t + dt / j, t)$ la probabilité conditionnelle pour qu'un individu se trouve dans la zone i au temps $t + dt$ sachant qu'il résidait dans la zone j au temps t soit encore :

$$P(i, t + dt / j, t) = dt p_{ij}(t) + o(dt^2) \quad [\text{développement de Taylor}]$$

avec $p_{ij}(t) = \left. \frac{dP(i, t' / j, t)}{dt'} \right|_{t'=t}$: dérivée de la probabilité au point t , autrement dit le changement instantané de la probabilité de migrer de j à i durant une période de temps dt très courte.

Cette quantité est appelée 'taux de transition individuel'.

* au niveau macroscopique du système :

A la suite du déplacement d'un individu de la zone j à i , le système est passé d'une configuration n à une configuration n' .

La probabilité de ce changement de configuration s'écrit :

$P(n', t + dt / n, t)$ = probabilité conditionnelle de trouver la configuration n' au temps $t + dt$, sachant que l'on observait la configuration n au temps t , soit encore :

$$P(n', t + dt / n, t) = dt w_{ij}(n, t) + o(dt^2)$$

avec $w_{ij}(t) = \left. \frac{dP(n', t' / n, t)}{dt'} \right|_{t'=t}$: changement instantané de la probabilité de passer d'une configuration n à une configuration voisine n' .

Cette quantité est appelée 'taux de transition configurationnel'.

Les taux mentionnés ci-dessus sont reliés par la relation :

$$w_{ij}(n, t) = n_j p_{ij}(t)$$

Ainsi pour passer d'un macro-état à un autre immédiatement voisin, il suffit de connaître les caractéristiques des micro-états correspondants, c'est à dire les taux de transition individuels.

On aura :

$$\sum_i \sum_j w_{ij} = \sum_i \sum_j n_j p_{ij} \quad \text{où } i, j = 1, \dots, L \quad \text{avec } i \neq j$$

$$\frac{dP(n, t)}{dt} = \sum_i \sum_j w_{ij}(n'_i) P(n'_i, t) - \sum_i \sum_j w_{ij}(n) P(n, t) = \text{flux entrants} \quad - \text{flux sortants}$$

sous les hypothèses suivantes :

- les probabilités individuelles de migration ne sont complètement indépendantes (ex. des migrations d'accompagnement' des membres du ménage)
- les comportements migratoires varient en fonction des variables socio-démographiques des individus (notamment âge et milieu social)

Comme il s'avère difficile dans la pratique de disposer de plusieurs états du système (recensements) aussi travaille-t-on généralement sur le processus déterministe, par passage aux valeurs moyennes, associé au processus stochastique engendré par les équations maîtresses.

Si l'on fait l'hypothèse que la distribution des $P(n, t)$ est unimodale, alors on peut supposer que la configuration moyenne \bar{n} qui a la plus grande probabilité de réalisation coïncide avec la configuration observée.

Notons $\bar{n}(t) = (\bar{n}_1(t), \dots, \bar{n}_L(t))$ cette configuration moyenne.

La population moyenne $\bar{n}_k(t)$ d'une zone k est donnée par : $\bar{n}_k(t) = \sum_C n_k P(n,t)$ avec C ensemble des configurations possibles. C'est aussi l'espérance mathématique de la distribution de probabilité $P(n,t)$.

On démontre que :

$$d\bar{n}_k(t)/dt = \sum_C n_k dP(n,t)/dt = \sum_i \bar{n}_i p_{ki} - \sum_j \bar{n}_k p_{jk} \quad \text{pour } k = 1, \dots, L$$

La dynamique du système est donc décrite par un système de L équations différentielles, fonctions de la taille des zones n_i et des taux de transition individuels p_{ik} de migrer d'une zone à une autre. La variation de la population d'une zone k du système résulte de l'addition de l'ensemble des flux ayant k pour destination ($\sum_i \bar{n}_i p_{ki}$) auquel on soustrait tous les flux ayant k pour origine ($\sum_j \bar{n}_k p_{jk}$).

Reste à définir les taux de transitions individuels $p_{ij}(n)$

On notera : $p_{ij}(t) = v_{ij}(t) \exp(A_i(t) - A_j(t))$ soit $L * \frac{(L-1)}{2} * T$ paramètres à estimer

avec :

- une mesure symétrique de la proximité, au sens large, entre deux zones : $v_{ij} = v_{ji}$. Si l'on note : L le nombre de zones du système, T le nombre d'intervalles de temps considérés, on peut écrire :

$$v_{ij}(t) = v_0(t) f_{ij} \quad \text{soit } L * \frac{(L-1)}{2} + T \text{ paramètres à estimer avec :}$$

* $v_0(t)$: moyenne des indices $v_{ij}(t)$ pour la date t (mesure du niveau général de mobilité à cette date, $1 * T$ paramètres à estimer)

* $f_{ij} = f_{ji}$: facteur qui rend compte des effets symétriques globaux entre deux couples de zones ($L * (L-1)/2$ paramètres à estimer)

- une composante spécifique :

* A_i : attractivité spécifique de l'agglomération i (L paramètres à estimer)

* $\exp(A_i - A_j)$: gain que représente pour un individu le fait de résider dans la zone i plutôt que dans la zone j .

Cette formulation est à rapprocher de la théorie de l'utilité, les modèles de choix discret s'obtenant comme solution stationnaire du modèle dynamique.

L'équation générale du modèle s'écrit donc comme :

$$d\bar{n}_k/dt = \sum_i \bar{n}_i v_0(t) f_{ki} \exp(A_k - A_i) - \sum_j \bar{n}_k v_0(t) f_{jk} \exp(A_j - A_k)$$

Les méthodes d'estimation sont fondées sur des algorithmes classiques de type log-linéaire ou non linéaire.

Exemple d'application : évolution dynamique du système urbain français (L. Sanders).

Bibliographie

- Bahrenberg, Fisher, Nijkamp (1984), Recent developments in spatial analysis : methodology, measurement, models.
- Ben-Akiva, Lerman (1985), Discrete choice analyses : Theory and application to travel Demand.
- Birkin M., Clarke M. (1984), "Comprehensive Dynamic Urban Models : Integrating Macro and Micro Approaches", *Evolving Geographical Structures* (D.A. Griffith, A.C. Lea, eds.), pp. 164-191.
- Clarke M., Wilson A.G. (1986), "A framework for dynamic comprehensive urban models : the integration of accounting and microsimulation approaches", *Sistemi Urbani*, n° 2/3, pp. 145-177.
- Crouchley R. (1986), "An examination of the equivalence of three alternative mechanisms for establishing the equilibrium solutions of the production-constrained spatial interaction model", *Environment and Planning A*, 19, pp. 861-874.
- Dendrinou D.S., Mullally H. (1988), "Volterra-Lotka Ecological Dynamics, Gravitational Interaction and Turbulent Transportation : an integration", paper presented at the 35th North American Meeting of the Regional Science Association, Toronto, Canada.
- Fisher, Nijkamp (1987), From static towards dynamic discrete choice modelling : a state of the art review.
- Haag G., Weidlich W. (1988), Interregional migration, Dynamic Theory and comparative analysis, Springer Verlag.
- Haag G. (1989), Dynamic decision Theory : applications to urban and regional topics, Dordrecht, Kluwer Acad. Pub..
- Haag G., Gruetzmann K. (1993), "A new Theory of nested decision processes with memory", *Paper in Regional Science*, 72, 3, pp. 313-335.
- Harris, Wilson (1978), "Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models", *Environment and Planning A*, 10, pp. 371-388.
- Kaashoek, Vorst (1984), "The cusp catastrophe in the urban retail model", *Environment and Planning A*, 16, pp. 851-862.
- Nijkamp P., Poot J. (1986), "Dynamics of generalised spatial interaction models", *Regional sciences and Urban Economics*, 17, pp. 367-390.
- Nijkamp P., Reggiani A. (1988), "Dynamic spatial interaction models : new directions", *Environment and Planning*, 20, pp. 1449-1460.
- Nijkamp P., Reggiani A. (1988), "Analysis of dynamic spatial interaction models by means of optimal control", *Geographical Analysis*, 20, 1, pp. 21-29.
- Nijkamp, Reggiani (1988), Analysis of dynamic spatial interaction models by means of optimal control.
- Nijkamp, Reggiani (January 1991), "Chaos Theory and spatial dynamics", *Journal of transport economics and policy*, pp. 81-96.
- Nijkamp, Reggiani (1992), Interaction, Evolution and Chaos in Space, Springer Verlag.
- Oppenheim (1991), "Retail activity allocation modelling with endogenous retail prices and shopping travel costs", *Environment and Planning A*, 23, pp. 731-744.
- Sanders L. (1992), Systèmes de villes et synergie, collection villes, Anthropos-Economica.
- Slater P.B. (1991), "Equilibrium and nonequilibrium statistical thermodynamical approaches to modelling spatial interaction dynamics", *Environment and Planning A*, 24, pp. 441-446.
- Wilson (1981), Catastrophe Theory and bifurcation : applications to urban and regional systems (Croom Helm, Beckenham, Kent).

FICHE N°7

MIS & RECHERCHE DE PARTITIONNEMENT

INTRODUCTION

- Rappels sur la problématique des MIS :

Classiquement il s'agira de fixer une règle de mobilité f permettant de reconstituer au mieux une matrice de flux estimés $[F^*]$ à partir de E , R et D .

E et R : vecteurs des sommes de départ et d'arrivée.

D : matrice de coût généralisé moyen des déplacements.

$$[F^*] = f(E, R, [D])$$

“ L'hypothèse centrale de ces modèles (MIS) est que la distance, c'est à dire la proximité spatiale mesurée en kilomètres, temps ou coûts a une influence déterminante sur la probabilité de relation entre les habitants ou groupes sociaux répartis sur un territoire ”.

La modélisation de l'effet d'appartenance doit permettre de mettre en évidence “des formes d'intégration territoriale spécifique, autrement dit des portions de territoires dans lesquelles les habitants adoptent un comportement migratoire significativement différent de celui des habitants d'autres portions du territoire ” [Claude Grasland, 1996].

- Questions posées :

- L'introduction d'une ou plusieurs partitions territoriales a priori dans le modèle entraîne-t-elle un gain d'ajustement ?

- recherchons la partition optimale a posteriori qui garantisse le meilleur ajustement, ce qui nous conduit à résoudre des problèmes de partition d'ensembles connexes. La recherche de la configuration optimale parmi l'ensemble de toutes les configurations possibles peut vite s'avérer lourde et très coûteuse en temps de calcul. Il a donc fallu recourir aux techniques dites de *parallélisation*, dont schématiquement on peut décrire les différentes phases :

- * phase n°1 de parallélisation : répartition des calculs sur un nombre de processeurs p ,

- * phase n°2 : application de méthodes de répartition de charges afin d'éviter les déséquilibres entre les processeurs,

- * phase n°3 : Classification des différentes partitions par le biais de la fonction de coût associée au modèle.

La partition optimale, est celle dont la valeur de la fonction de coût est minimale (e).

$$\text{avec } e = \chi^2_{F,F^*} / \chi^2_{F,F^{\text{id}}}$$

CLASSIFICATION EXHAUSTIVE SUR L'ENSEMBLE DES PARTITIONS

Une étude exhaustive vise à rechercher la configuration optimale parmi l'ensemble de toutes les configurations possibles. Cette démarche s'avère très vite lourde et coûteuse en temps de calcul, en effet selon la configuration considérée le nombre d'expérimentations peut varier exponentiellement. Ainsi un graphe totalement connecté de taille 10 permet de construire $2^{10-1} = 512$ partitions un-deux, $3^{10-1} = 19\ 683$ partitions en trois, $4^{10-1} = 262\ 144$ partitions en quatre, Dans le même ordre d'idée, il est intéressant de noter que les trois formes de découpage de la France, ci-après retenues ont permis

de générer respectivement: 1 085 partitions France entière, 1 116 partitions France sans Paris, 30 103 partitions France avec Paris neutre.

Une fois les résultats obtenus, il devient possible de classer les différentes partitions par la biais de la fonction de coût associée au modèle. Ceci permet de définir la partition optimale, c'est à dire la partition dont la valeur de la fonction de coût est minimale.

Le principe de cette méthode de classification oblige à prendre en compte l'ensemble des cas possibles sans distinction. La complexité du traitement d'une structure géographique sera donc fonction de la taille de son « partitionnement ».

La complexité algorithmique de la fonction coût est en $O(n^2)$ (calcul d'un χ^2), donc celle de la fonction d'optimisation sera du même ordre. Afin de mieux appréhender l'impact de la complexité algorithmique sur les limites de la plate-forme expérimentale, celle-ci sera traduite en un temps de calcul moyen (les temps peuvent varier d'un facteur 2 selon la partition traitée) pour un modèle donné.

Le tableau permet de matérialiser le problème du temps de calcul nécessaire pour mener à bien cette expérimentation. Il apparaît clairement qu'un nombre excessif de partitions entraîne des temps de calcul ne permettant pas une utilisation aisée de la plate-forme. Afin de remédier à cela, une première « phase de parallélisation » a été effectuée, elle correspond à une répartition des calculs sur un nombre de processeurs p . Les partitions sont réparties sur chacun des processeurs uniformément. Ces derniers réalisent les opérations nécessaires pour chaque partition et enregistrent les résultats dans un fichier personnel. C'est lors du traitement de ces résultats que les fichiers sont fusionnés et triés. Chaque processeur traite donc un problème de taille inférieure. En première approximation, les temps de calculs sont en $T_p = O(TI/p)$, pour tout p inférieur ou égal au nombre total de partitions. L'avantage d'une telle technique est qu'elle ne présuppose aucune caractéristique particulière pour le matériel employé, ce qui lui confère une 'portabilité' maximale.

Dans un deuxième temps, il est nécessaire d'appliquer les méthodes de répartition de charges afin d'éviter les déséquilibres entre les processeurs. En effet les calculs nécessaires à l'optimisation de la fonction coût varient selon la partition qui lui est associée. Des situations où certains processeurs ont fini de traiter les partitions qui lui sont attribué, alors que d'autres sont en cours de calcul peuvent survenir. Une redistribution des partitions restantes permettrait d'augmenter l'efficacité de la parallélisation.

Une troisième voie de développement porterait sur la parallélisation de la fonction de coût. Celle-ci se base sur un calcul du χ^2 dont la complexité algorithmique est en $O(n^2)$ sur machine séquentielle et qui passe en $O(\log n)$ sur machine parallèle. Bien que les temps de calcul puissent être diminués par ces différentes options de parallélisme, certaines situations obligent à sélectionner les partitions à traiter lorsque celles-ci sont trop nombreuses.

ALGORITHMES DE PARTITIONNEMENT D'UN ESPACE GEOGRAPHIQUE

On s'intéresse ici aux effets de la partition géographique sur les flux migratoires. On cherche à mettre en évidence le rôle joué par le découpage territorial sur les flux migratoires, ce qui permet de développer une analyse tout autant sur la notion d'appartenance que sur celle de distances inter-unités. Pour répondre à ces attentes, le LMC (IMAG) s'est penché sur les problèmes de partition d'un ensemble connexe. L'objectif est de développer un algorithme qui génère automatiquement tous les découpages, en k parties connexes, de la structure géographique étudiée.

Analyse du problème

Le problème de la connexité d'un ensemble à n éléments a été largement traité dans la littérature algorithmique \site{bk:aho,bk:graham,bk:knuth}. Par contre il est beaucoup plus rare de trouver des ouvrages ou articles traitant de la génération exhaustive des partitions d'un ensemble.

Nous avons choisi de représenter la structure géographique étudiée par son graphe de contiguïté (Figure 1). De cette façon, nous pouvons appliquer toutes les théories et les algorithmes portant sur les graphes planaires \site{bk:graham,bk:sakarovitch}. En nous basant sur un algorithme récursif de parcours en profondeur du graphe, nous avons développé notre propre générateur automatique de partitions à partir de:

- X un ensemble de sommets modélisant les unités territoriales.
- E un ensemble d'arêtes modélisant la contiguïté entre unités territoriales.

Ce qui permet de définir:

qu'une partition du territoire, représentée par les ensembles de sommets $P_1 \dots P_k$, est cohérente (du point de vue des géographes), si et seulement si les sous-graphes définis par ces ensembles sont connexes.

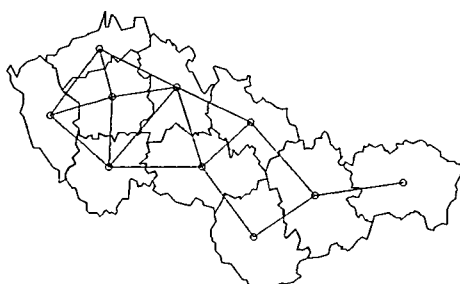


Figure 1: Graphe de contiguïté représentant la structure géographique étudiée

Cette définition permet de ne pas avoir à tester la totalité des partitions possibles (soit k^{n-1} pour une partition en k parties d'un graphe de dimension n) mais de ne retenir que celles dites "cohérentes".

Présentation de l'algorithme de génération des partitions

Principe de fonctionnement de l'algorithme

L'algorithme est basé sur le principe de parcours en profondeur récursif d'un graphe (graphe de contiguïté associé à l'ensemble X (exemple Figure 1), de type "Depth-First Search" \site{bk:aho}).

Définitions:

- Comme l'algorithme est basé sur un parcours en profondeur d'un graphe G , il faudra comprendre par descente dans le graphe l'examen d'une branche de l'arbre de récursions.
- L'ensemble des voisins du nœud x_i , V_{x_i} représente l'ensemble des nœuds contigus à x_i .
- L'ensemble des nœuds visités: $Visit$, représente l'ensemble des nœuds rencontrés lors du parcours du graphe.

Soit un graphe G et un nœud $\{y\}$ ($\{y\} \in G$). $\{y\}$ déconnecte le sous-graphe g ($g \subset G$) si il n'existe pas de sous graphe connexe g' ($g' \subset g$) tel que:

$$g' = g \setminus \{y\}$$

Le parcours du graphe initial G va permettre de générer des sous-graphes définissant les parties de la structure géographique étudiée. Ces sous-graphes sont représentés par des ensembles de nœuds rencontrés lors du parcours du graphe (X_k). Prenons l'exemple de la construction des partitions en deux parties connexes. L'algorithme va visiter tous les nœuds du graphe à partir d'un nœud

d'initialisation $\{x_0\}$. Le passage du nœud $\{x_i\}$ au nœud $\{x_j\}$ est défini grâce aux ensembles V_{x_i} et Visité. Le premier ensemble permet de déterminer les nœuds candidats, le second garantit de ne pas visiter deux fois le même nœud. Les nœuds rencontrés seront placés dans X_1 lors de la descente dans le graphe, puis dans X_2 lors de la remontée. Pour qu'une partition soit générée, il faut que l'ensemble V_{x_i} soit vide, c'est-à-dire qu'il n'existe plus de nœud voisin non visité. Dans ce cas, X_1 définira la première partie, et son complémentaire la deuxième. La partition sera définitivement validée si les deux sous-graphes, représentés par X_1 et X_2 , sont connexes (Ex: Figure 2).

Le parcours du graphe initial nous assure de générer la totalité des partitions possibles. De plus, puisqu'un nœud ne peut faire parti des deux ensembles en même temps, et qu'il ne peut être visité qu'une seule fois par descente dans le graphe, nous sommes certains de ne pas construire deux fois la même partition.

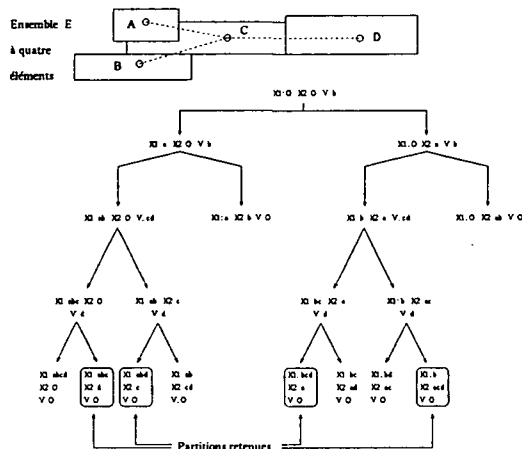


Figure 2 : Arbre d'exécution du deuxième algorithme sur un ensemble E à quatre éléments

Pour un graphe fortement connexe, la complexité de cette algorithme est exponentielle (fonction de la dimension de la structure géographique étudiée). Néanmoins, puisque la construction des partitions repose sur le parcours en profondeur du graphe de contiguïté, nous pouvons supposer que le nombre de partitions construites par l'algorithme sera fonction de la dimension du graphe, et du degré moyen de ces nœuds.

Construction récursive optimale

Le principe de l'algorithme est globalement équivalent au précédent. Si l'on reprend l'exemple précédent, la différence s'établit sur le mode de construction des deux parties.

L'initialisation s'effectue toujours grâce à un nœud de $\{x_0\}$ pris au hasard. Mais, lorsque l'on considère un nœud $\{x_j\}$ appartenant à V_{x_i} , on cherche à savoir s'il déconnecte ou non la deuxième partie X_2 (la première partie X_1 est constituée de nœuds voisins les uns des autres, elle est donc connexe par construction). Si x_j ne déconnecte pas X_2 , on cherche à l'englober dans un sous-graphe connexe de E. Si un tel sous-graphe existe, X_1 sera redéfini par son complémentaire. Une fois ces étapes réalisées, la descente dans le graphe se poursuit avec un nœud de V_{x_j} . Si par contre x_j déconnecte X_2 , la descente dans le graphe sera arrêtée et $\{x_j\}$ sera placé dans X_2 . L'algorithme reprendra la construction des partitions avec un autre élément de V_{x_i} .

De cette manière, dès qu'un nœud déconnecte X_2 , tout le parcours du graphe qui s'y rapporte est supprimé (Figure 3). De même, lorsque l'on a identifié le sous-graphe connexe incluant X_2 , nous

pouvons ré-initialiser X_1 par son complémentaire: $\text{Compl}X_2$, puisque les nœuds de $\text{Compl}X_2$ ne pourront pas faire partie de X_2 (Figure 4).

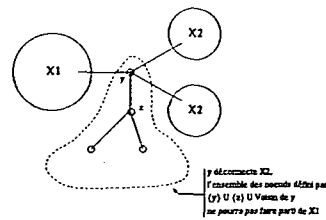


Figure 3 : Branche de parcours du graphe coupée par l'algorithme.

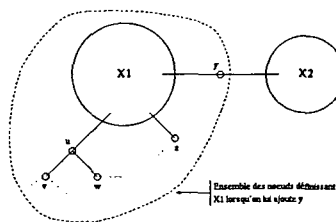


Figure 4 : Construction de X_1 .

L'exemple représenté par la Figure 5 illustre les deux situations types rencontrées lors de l'application de cet algorithme. E est équivalent à $X_1 \cup X_2 \cup \{y'\} \cup \{y''\}$. L'algorithme va traiter successivement $\{y'\}$ puis $\{y''\}$. Le premier nœud ne déconnecte pas X_2 dans la mesure où il existe un sous-graphe connexe de E , $X_2 \cup \{y''\}$, qui inclut X_2 . $\{y'\}$ sera donc ajouté à X_1 afin de former X_1' . Par contre le deuxième nœud déconnectera X_2 puisqu'il n'existe pas de sous-graphe connexe de E qui inclut X_1' ne sera donc pas construit par l'algorithme. $\{y''\}$ sera ajouté à X_2 , et une nouvelle descente dans le graphe sera effectuée à partir de X_1' et $X_2 \cup \{y''\}$.

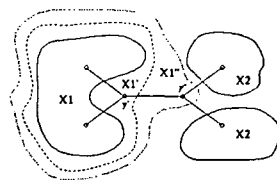


Figure 5 : Représentation du déroulement du troisième algorithme.

L'optimisation effectuée permet la suppression des branches de la récursion n'aboutissant pas à deux sous-graphes connexes. L'algorithme génère uniquement des partitions cohérentes au sens des géographes. Il possède donc une complexité équivalente au nombre de partitions générées.

Généralisation: partition en k parties connexes

La généralisation à la construction de k ($k > 2$) parties connexes s'effectue en subdivisant les sous-parties déjà formées. Ainsi, la génération des partitions en trois parties connexes s'effectue en appliquant l'algorithme une première fois sur l'ensemble E , puis en le réitérant une deuxième fois sur l'une des parties. Pour les partitions quatre parties, l'algorithme sera ré-appliqué sur les deux parties précédemment calculées, etc....

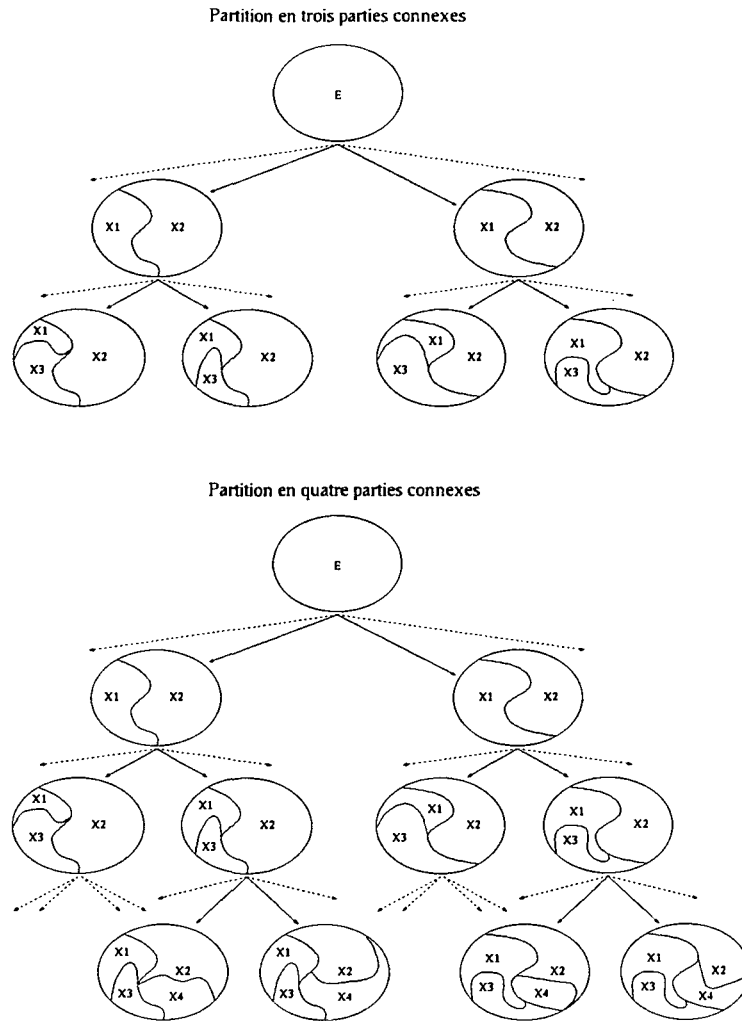


Figure 6 : Construction de k ($k > 2$) parties connexes.

Le tableau ci-dessous représente le nombre de partitions en 2, 3 ou 4 parties générées par le troisième algorithme. Les résultats fournis sont tirés de l'étude de l'ex-Tchécoslovaquie subdivisé en 10 régions ou 109 districts.

Nombre de partitions construites

10 unités	2	3	4
109 unités	48	557	760

EVALUATION DES PARAMETRES DU MODELE

L'évaluation des paramètres est effectuée à l'aide d'une fonction d'optimisation. Celle qui est actuellement disponible sur la plate-forme est nommée '*Hooke*'. Elle a été implantée par Mark G. Johnson (déc.1994), d'après la méthode de 'recherche directe' développée par Kooke et Jeeves. Le principe général de l'algorithme mis en place repose sur la 'recherche directe' d'une solution à des problèmes numériques ou statistiques. La méthode consiste à examiner un ensemble de solutions triviales afin de déterminer une stratégie de construction permettant de former un nouvel ensemble de solutions, inclus dans le premier, de meilleure qualité. Si nous prenons l'exemple d'une fonctionnelle

f à minimiser, ceci revient à évaluer sa valeur en un point initial x_0 et à trouver un nouveau point x_1 tel que:

$$x_1 = x_0 \pm \delta_0 * f(x_1) < f(x_0)$$

La valeur de δ_0 (pas de recherche de la solution) est fixée à l'initialisation. δ_0 est multiplié par un facteur $\rho < 1$ si $f(x_0 \pm \delta_0) \leq f(x_0)$. De cette façon, x_i converge vers \underline{x} qui minimise la fonctionnelle f . La condition d'arrêt de l'optimisation est réalisée lorsque $f(x_{i+1}) - f(x_i) < \nu$ ou lorsque le nombre d'itérations maximum est atteint.

Cette fonction d'optimisation a été choisie car elle ne nécessite pas de conditions particulières sur la linéarité ou la dérivabilité de la fonctionnelle considérée. Néanmoins, l'usage a montré que son comportement est très sensible aux conditions initiales (valeur de x_0 , δ_0 et ρ) et à la forme de la fonctionnelle étudiée (si l'on se trouve au milieu d'un plateau, de diamètre supérieur à $2 * \delta_0$, la fonction Hooke ne convergera pas). De plus l'optima trouvé peut être local ou global sans qu'on puisse le savoir.

Le principal défaut de cette fonction est son caractère itératif qui la rend très coûteuse en temps de calcul. Dans le cas qui nous concerne (utilisation du χ^2 comme fonction de coût), le problème de la dérivabilité ne se pose pas, néanmoins comme la formulation du modèle de Wilson est intrinsèquement de nature itérative et donc rend l'expression de la dérivée difficilement exploitable pour la détermination des paramètres optimaux.

Temps moyen de calcul par type de modèle utilisé en secondes et heures/minutes/secondes

<i>Nombre de partitions en 2 parties connexes</i>				
	<i>Référence</i>	<i>avec PARIS</i>	<i>sans PARIS</i>	<i>PARIS neutre</i>
	1	2 085	1 116	3 103
BARRIERE	2	4 170 01:09:30	2 232 00:37:12	6 206 01:43:26
PARETO+BARRIERE	8	16 680 04:38:00	8 928 02:28:48	2 484 06:53:44

EXEMPLE D'APPLICATION : Modélisation de l'effet de barrière

Cette étude visait un double objectif:

* *l'estimation d'un modèle à doubles contraintes* pour les déplacements inter-régionaux de voyageurs et de véhicules à longue distance par mode et tous motifs, à partir de différentes sources et pour différentes spécifications de la fonction de résistance,

* *la généralisation de ces modèles dans un deuxième temps*, en incluant de possibles effets d'appartenance (barrières ou anti-barrières géographiques). L'idée est de tester l'hypothèse d'un espace français des transports spatialement homogène et isotrope.

Ce dernier paramètre sera utilisé sous une forme multiplicative dans le modèle à doubles contraintes, on recherchera la partition en 2 optimale pour le critère du χ^2 pour l'ensemble connexe considéré, en neutralisant ou non l'effet de Paris; une plate-forme informatique de recherche de partitionnement étant 'implémentée' à cet effet.

Données utilisées

Les matrices de flux (voyageurs et trafics) utilisées sont issues de plusieurs sources:

* **Source n° 1: Modèle de prévision MATISSE**

Matrices multimodales de trafics en voyageurs pour le trafic français domestique en 1992, les zones origine-destination correspondant aux départements de la métropole (sauf les Hauts-de-Seine et le Val de Marne agrégés à Paris). Tous les couples de zones dont les centres sont distants de plus de 50 km à vol d'oiseau sont traités hors trajets internes à un département.

Les flux ne sont pas différenciés par sens et ont été agrégés au niveau régions administratives.

* **Source n° 2: Enquête Transports et Communications (ETC) 1993-94 (INSEE)**

Cette enquête transports ménages s'est déroulée sur la période allant du 03/05/93 au 30/04/94. Les données ici utilisées proviennent du bloc « déplacements à longue distance (+ 80 km à vol d'oiseau du lieu de domicile ». L'interview de la personne B pour ses voyages et déplacements pondérés porte sur les trois derniers mois. Les « moyens de transports principaux » considérés sont: la VP (3 ou 4 roues), le train.

* **Source N° 3: Matrices du SETRA**

Matrices de flux de véhicules VP intérieurs interrégionaux pour « une année 1993 » en TMJA, les deux sens confondus. Cette matrice provient de la construction d'une matrice interdépartementale, réalisée à partir des enquêtes de circulation menées dans les CETE au cours des 5 dernières années. L'exploitation des enquêtes a permis de renseigner 50% de la matrice, le reste a été modélisé à partir des relations connues à l'aide du logiciel TRIPS (MVA).

L'ensemble des matrices régionales (9) couvrent les deux sens de trajets confondus, elles ont en effet été 'symétrisées' ($f_{ij} + f_{ji}$) et le mode aérien a été mis de côté.

On dispose donc de 9 types de matrices:

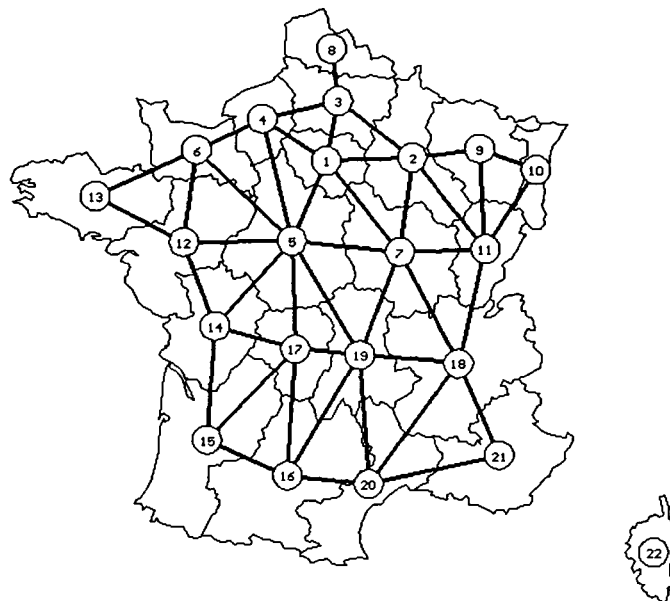
* 3 matrices issues de MATISSE: VP, Train et Total des flux,

* 3 matrices issues de l'ETC 93-94: VP, Train et Total des flux,

* 3 matrices routières en nombre de véhicules: VP-SETRA 1^{ère} version, VP-SETRA 2^{ème} version, VP-MATISSE.

Zonages

- un zonage spécifique a été défini à partir du fichier des zones d'emplois (INSEE),
- Topologie du graphe de contiguïté: un graphe de contiguïté [voir graphe ci-dessous] a été élaboré dont une région La Corse a été volontairement exclue. D'autre part l'option de non-rattachement des régions N°8 (Nord) et N°2 (Champagne) a été retenue même s'il s'en faut de quelques communes et compte tenu de l'allongement de la région Champagne-Ardenne.



Evaluation de l'influence de Paris

L'influence de l'Ile de France a été décomposée en trois parties :

- * cette région jouant un rôle identique à n'importe quelle autre région,
- * une région neutralisée (PARIS NEUTRE), au sens où les flux peuvent transiter par la région mais sans s'arrêter,
- * une région déconnectée des autres unités géographiques: aucun flux ne peut s'arrêter ni transiter par cette région.

Distances employées

La distance la plus simple à employer est la distance euclidienne ($DEucl$). Elle représente la distance « à vol d'oiseau » entre deux unités géographiques.

$$DEucl = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

(x_i, y_i) : centre de gravité de l'unité i

(x_j, y_j) : centre de gravité de l'unité j

Les coordonnées citées ci-dessus représentent le centre de gravité des unités i et j , mais elles peuvent tout aussi bien identifier la ville la plus importante, le chef lieu administratif, etc..

Bien que la distance euclidienne puisse paraître tout à fait adaptée à notre cas, il est possible d'employer n'importe quel type de distance, d'autant plus que les matrices de flux utilisées concernent les déplacements en moyenne à plus de 80 km à vol d'oiseau du lieu de domicile. A cette occasion trois nouvelles distances ont été construites. Elles permettent de prendre en compte la

population incluse dans les zones d'emplois INSEE. Cette subdivision des unités géographiques en zones d'emplois permet de pondérer l'action des flux en fonction de la population.

1. D_0 : se base sur la définition de la distance euclidienne pondérée par la proportion de la population des zones d'emplois.

Soit les unités géographiques i et j et les zones d'emplois k ($k \in i$) et l ($l \in j$). La distance $D_{0i,j}$ est définie par:

$$D_{0i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]}$$

avec

$pop_i[k]$: population de la zone d'emploi k de l'unité géographique i

2. D_{50} : se base sur la définition de la distance euclidienne pondérée par la proportion de la population des zones d'emploi séparées d'au moins 50 km.

Soit les unités géographiques i et j et les zones d'emplois k ($k \in i$) et l ($l \in j$). La distance $D_{50i,j}$ est définie par:

$$D_{50i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]} * \delta [DEucl_{k,l} \geq 50]$$

avec

$pop_i[k]$: population de la zone d'emploi k de l'unité géographique i

$\delta [DEucl_{k,l} \geq 50]$: vaut 1 si $DEucl_{k,l} \geq 50$, 0 sinon.

3. D_{80} : se base sur la définition de la distance euclidienne pondérée par la proportion de la population des zones d'emploi séparées d'au moins 80 km.

Soit les unités géographiques i et j et les zones d'emplois k ($k \in i$) et l ($l \in j$). La distance $D_{80i,j}$ est définie par:

$$D_{80i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]} * \delta [DEucl_{k,l} \geq 80]$$

avec

$pop_i[k]$: population de la zone d'emploi k de l'unité géographique i

$\delta [DEucl_{k,l} \geq 80]$: vaut 1 si $DEucl_{k,l} \geq 80$, 0 sinon.

Sélection de la matrice de distance

L'étude exhaustive croisant matrices de flux (9) * fonctions de distance (4) * distances interrégionales (5) correspondrait à 180 configurations possibles !. Il s'est vite avéré nécessaire d'effectuer un choix sur l'une de ces contraintes (flux, fonctions de distance, distances). A cet effet une étude complète des distances disponibles a été réalisée grâce au modèle à doubles contraintes associé à la fonction de la distance de Pareto [voir tableau ci-dessous].

L'analyse des erreurs présentée dans les tableaux ci-joints nous orientent vers le **choix de la distance D_0** comme distance la plus appropriée.

SELECTION DE LA MATRICE DE DISTANCE

Modèle de Pareto

	DEucl	D0	D50	D80	D50+Prop	D80+Prop	MIN	MEDIANE
	Err	Err	Err	Err	Err	Err		
M31	0,0163	0,0080	0,0080	0,0587	0,0081	0,0724	0,0080	0,0122
M32	0,0197	0,0191	0,0194	0,0363	0,0189	0,0410	0,0189	0,0195
M34	0,0299	0,0279	0,0286	0,0709	0,0278	0,0843	0,0278	0,0292
MS15	0,0482	0,0507	0,0522	0,1675	0,0506	0,1978	0,0482	0,0515
MS16	0,0265	0,0247	0,0248	0,0301	0,0246	0,0316	0,0246	0,0256
MS19	0,0467	0,0499	0,0511	0,1303	0,0497	0,1517	0,0467	0,0505
VLS	0,0250	0,0105	0,0109	0,0981	0,0107	0,1232	0,0105	0,0180
VLS2	0,0163	0,0080	0,0080	0,0587	0,0081	0,0724	0,0080	0,0122
VLS3	0,0312	0,0250	0,0251	0,0770	0,0250	0,0598	0,0250	0,0282

Spécifications

Afin d'aboutir au plus degré de généralité côté spécification des équations testées, il a été convenu de partir d'un modèle de Wilson à doubles contraintes sans conditionner l'ajustement dans la première phase par la disponibilité de variables économiques et démographiques pour qualifier les facteurs d'émission et d'attraction, ce qui est souvent le cas, pour dans des phases ultérieures permettre, si nécessaire, des procédures de réajustement aux marges plus larges.

La forme générale des modèles utilisés ci-après est celle des Modèles d'Interactions Spatiales dits à doubles contraintes. Cette dernière permet de combiner le pouvoir de rétention et d'attraction des unités géographiques, dans un objectif plus de prévision que de caractérisation.

On a:

$$F_{ij}^* = a_i O_i b_j D_j h(c_{ij})$$

avec:

$$a_i = \frac{1}{\sum_{i,j} b_j D_j h(c_{ij})} : \text{contrainte de la conservation des départs}$$

$$b_j = \frac{1}{\sum_{i,j} a_i O_i h(c_{ij})} : \text{contrainte de la conservation des arrivées}$$

O_i : somme des arrivées en i

D_j : somme des départs de j

$h(c_{ij})$: fonction d'interaction spatiale (résistance)

☞ Pour ce qui suit quatre types de spécification dont une dite de 'référence' ont été retenus:

- * modèle dit de « *référence* » (I): $F_{ij}^* = a_i O_i b_j D_j h(c_{ij})$ avec $h = Id$
- * modèle dit de *Pareto* (II): $F_{ij}^* = a_i O_i b_j D_j h(c_{ij})$ avec $h(c_{ij}) = cij^{-a}$
- * modèle dit de *Pareto+barrière* (III): $F_{ij}^* = a_i O_i b_j D_j h(c_{ij}) \gamma^{A_{ij}}$ avec $h(c_{ij}) = cij^{-a}$
- * modèle dit de *barrière seule* (IV): $F_{ij}^* = a_i O_i b_j D_j \gamma^{A_{ij}}$

$\gamma^{Aij} / Aij = 0$ (flux inter-régions), $Aij = 1$ (flux intra-régions)

Les erreurs de chaque type de modèle seront exprimées par rapport au modèle dit de référence (mod. I)

$$\text{Err}(F, F^*) = \sum_{ij} (Fij - Fij^*)^2 / Fij^* \quad (\text{critère du type CHI2: } \chi^2_{F, F^*})$$

- ◆ cinq types de distances sont prises en compte (segmentation par nombre de paires se situant à plus de) : distance euclidienne, distances à vol d'oiseau entre zones d'emploi: seuil=0, =50 km, =80. Ces distances ont été établies à partir du fonds de carte des zones d'emploi 1994 (INSEE).
- ◆ à titre de test deux nouvelles variables avaient été définies (prop_{ab} , \underline{d}_{ab}) :

* $Sij = 0$ ou $= 1$ (si zones d'emploi sont entre elles à un seuil de distance dij)

$$* \text{prop}_{ab} = [\sum_{iea} \sum_{jeb} Pi Pj Sij] / [\sum_{iea} \sum_{jeb} Pi Pj]$$

(proportion de paires d'OD qui vérifient la condition Sij)

$$* \underline{d}_{ab} = [\sum_{iea} \sum_{jeb} Pi Pj Sij dij] / [\sum_{iea} \sum_{jeb} Pi Pj Sij]$$

Les tests d'introduction des dites variables n'ont pas amélioré statistiquement les résultats d'estimation et donc celles-ci ont été de facto écartées.

Mesure de la qualité des modèles

La qualité du modèle est déterminée par l'erreur entre la matrice de flux observés et la matrice des flux estimés: $\text{Err}(F, F^*)$. Cette erreur permet d'apprécier l'adéquation du modèle utilisé par rapport au critères étudiés. Toute norme matricielle est utilisable pour mesurer cette erreur. Trois types de mesure sont communément employées:

1. la mesure en *valeur absolue*:

$$M1 = \text{Err}(F, F^*) = \sum_{ij} |Fij - Fij^*|$$

2. la mesure du *carré de l'erreur*:

$$M2 = \text{Err}(F, F^*) = \sum_{ij} (Fij - Fij^*)^2$$

3. la mesure du χ^2_{F, F^*} :

$$M3 = \text{Err}(F, F^*) = \sum_{ij} \frac{(Fij - Fij^*)^2}{Fij^*}$$

Cette dernière mesure M3 est plus coûteuse en temps de calcul que la seconde, elle-même plus coûteuse que la première ($M3 > M2 > M1$). Il faut en outre choisir une mesure permettant une interprétation claire des résultats. Les deux premières mesures (M1 et M2) seront utilisées en présence de données homogènes, alors que la troisième (M3) le sera pour des données hétérogènes. En effet, lorsqu'il s'agit de données hétérogènes, les deux premières mesures négligeront l'action des petites valeurs au profit des grosses, alors que la mesure du χ^2 (M3) réalisera une normalisation des données avant traitement. Cette normalisation s'avérera utile dans le cas d'unités géographiques très éloignées dont les valeurs de flux seront souvent faibles. Les résultats ainsi obtenus serviront à définir une fonction de coût dont la minimisation permettra l'ajustement des paramètres du modèle.

Plate-forme expérimentale utilisée

La plate-forme expérimentale se base sur une première version de la mesure de qualité des modèles réalisée par les géographes de l'équipe PARIS. Dans cette première version la formulation de l'erreur (Err) était:

$$MA(\chi^2) = \sum_{ij} \frac{(F_{ij} - MA(F))^2}{MA(F)} \quad \chi^2_{F,F^*} = \sum_{ij} \frac{(F_{ij} - F_{ij}^*)^2}{F_{ij}^*}$$
$$Err = \chi^2_{F,F^*} / MA(\chi^2)$$

avec :

$$MA(F): \text{Moyenne arithmétique des valeurs de la matrice des flux observés } F = \frac{1}{n^*(n-1)} \sum_{ij} F_{ij}$$

n: dimension de la matrice carrée.

Nous voyons ici que le χ^2 calculé est normalisé par un ' χ^2 moyen', Ce dernier nous permet d'apprécier la dispersion des flux observés par rapport à la moyenne.

Dans un même ordre d'idée, il nous a paru intéressant de calculer une erreur propre à la matrice de distance employée. Cette approche nous permet d'apprécier l'apport de la 'complexification' de la fonction de distance introduite. Pour se faire, nous calculons une *erreur de référence* pour chaque matrice en utilisant le modèle à doubles contraintes avec la fonction d'identité: χ^2_{F,F^*id} . Cette valeur nous servira à normaliser les erreurs calculées avec d'autres types de fonction de la distance.

La formulation de l'erreur du modèle (e) devient:

$$\chi^2_{F,F^*id} = \sum_i \frac{\sum_j (F_{ij} - F^* Id_{ij})^2}{F^* Id_{ij}} \quad e = \chi^2_{F,F^*} / \chi^2_{F,F^*id}$$

La définition de cette erreur nous permet de construire une fonction coût représentative de chaque option de modélisation.

Principaux résultats

Les principaux résultats présentés ne concernent ici que les meilleurs découpages de l'espace géographique, partitions de rang 1.

Ces résultats seront commentés en regard de plusieurs éléments:

- ♦ comparaisons pour un même mode entre matrices d'origine statistique différente,
- ♦ cartes modales des charges de réseau,
- ♦ 'cartes de potentiels' de population et de densité de population en France métropolitaine en 1990.

* une première remarque générale: le modèle simple de Pareto explique en moyenne plus de 80% des flux, avec une différence notable pour la matrice fer issue de l'ETC (56% de la variance expliquée).

* l'introduction de l'effet de barrière dans le modèle de Pareto apparaît significative pour les matrices fer, d'autant plus si l'on regarde les spécifications anti-barrière.

* si l'on cumule l'ensemble des partitions de toutes les matrices par type de modèle et si l'on regarde à la fois les cartes d'affectation des trafics et les cartes dites de potentiel de population établies par l'équipe PARIS, des analogies apparaissent immédiatement.

* la mesure des effets frontières à partir de la partition n°1 laisse apparaître une faible dispersion intermodale des valeurs [voir tableaux ci-après].

Part expliquée par le modèle
(en référence au modèle dit de référence)

Types de modèles	REFERENCE	PARETO	IS SEULE		PARETO+BARRIERE		GAINS III/II	
	I	II	IV		III		en points	
			B	AB	B	AB	B	AB
	ERR	ε	ε	ε	ε	ε		
ROUTE [MATISSE]	0,355	93,8%	41,9%	3,6%	95,2%	95,2%	1,4%	1,4%
FER [MATISSE]	0,108	82,3%	-90,8%	-216,8%	86,1%	87,0%	3,8%	4,7%
TOTAL [MATISSE]	0,288	90,3%	40,3%	3,5%	93,1%	92,4%	2,7%	2,0%
ROUTE [ETC]	0,569	91,1%	39,5%	2,3%	93,5%	93,1%	2,4%	2,1%
FER [ETC]	0,055	55,5%	38,5%	9,5%	60,2%	63,8%	4,7%	8,3%
TOTAL [ETC]	0,422	88,2%	39,3%	2,2%	90,3%	90,8%	2,1%	2,6%
VLS	0,444	97,7%	41,9%	2,8%	98,4%	98,2%	0,8%	0,5%
VLS2	0,228	96,5%	45,5%	5,5%	96,9%	96,9%	0,4%	0,4%
VLS3	0,194	87,1%	43,8%	7,7%	88,7%	88,1%	1,5%	1,0%

Mesure de l'effet frontière

$\gamma > 1$

		-	PARIS NEUTRE	SANS PARIS
MATISSE ROUTE	M31	1,4	1,7	1,7
ETC ROUTE	MS15	1,6	1,7	1,7
MATISSE FER	M32	1,5	1,7	1,7
ETC FER	MS16	1,5	1,9	2,0
MATISSE TOTAL	M34	1,5	1,7	1,7
ETC TOTAL	MS19	1,4	1,8	1,8
SETRA ¹	VLS	1,8	1,8	1,8
SETRA ²	VLS2	1,4	1,4	1,4
MATISSE	VLS3	1,9	1,5	1,7

Spécification: PARETO+BARRIERE

$\gamma > 1$

		-	PARIS NEUTRE	SANS PARIS
MATISSE ROUTE	M31	3,3	4,2	4,2
ETC ROUTE	MS15	3,5	4,4	4,4
MATISSE FER	M32	2,5	3,7	3,7
ETC FER	MS16	2,1	3,5	3,5
MATISSE TOTAL	M34	3,0	4,0	4,0
ETC TOTAL	MS19	3,3	4,8	4,8
SETRA ¹	VLS	4,3	5,0	5,0
SETRA ²	VLS2	3,6	4,3	4,3
MATISSE	VLS3	4,3	4,1	4,1

Spécification: BARRIERE

Bibliographie

- Aha, Hopcraft, Ullman (1974), The design and analysis of computer algorithms.
- Graham R.L., Grottschel M., Lovasz L. (1995), Handbook of combinatorics, North-Holland.
- Hooke R., Jeeves T.A. (1961), "Direct Search' solution of numerical and statistical problems", *Association for Computing Machine*.
- Sakarowitch M. (1984), Graphe et Programmation Linéaire, Hermann.
- Birkin M., Clarke M., George F. (1995), "The use of parallel computers to solve non-linear spatial optimisation problems : an application to network planning", *Environment and Planning A*, 27, pp. 1049-1068.
- Harris B. (1985), "Some notes on parallel computing with special reference to transportation and land use modelling", *Environment and Planning A*, 17, pp. 1275-1278.
- Openshaw S., Summer R. (1995), Parallel spatial interaction modelling on the KSR1-64 supercomputer, Working paper 95/15, School of Geography, University of Leeds.
- Turton I, Openshaw S. (199 ?), Modelling and optimising flows using parallel spatial interaction models.
- Vincent J.M., Ottogalli F.-G., Grasland C., Calzada C. (1997), "Déplacements de voyageurs interrégionaux à longue distance, MIS et recherche de partitionnement de l'espace", document interne SES, synthèse des résultats et rapport général.

FICHE N°8

MIS, RESEAUX DE NEURONES & INTELLIGENCE ARTIFICIELLE

Un formalisme issu de la modélisation de systèmes neurophysiologiques

Le terme Réseau de Neurones Formels (RNF) désigne un ensemble de modèles dont l'utilisation s'est largement répandue dans les quinze dernières années en touchant tous les domaines concernés par la modélisation et le traitement des données. Les premiers développements datent des débuts de l'ère informatique (années 40) et sont menés principalement par des chercheurs issus de disciplines comme la neurobiologie ou la psychologie. Après une période de relative marginalité dans les années 60, la discipline connexionniste a connu une véritable renaissance dans les années 80 sous l'impulsion cette fois de physiciens, de mathématiciens et d'informaticiens.

Parmi les utilisations types [R1] des RNF, citons les problèmes de Reconnaissance De Forme (classification, segmentation prédictive, détection), de prévision de trajectoire de systèmes dynamiques, de diagnostic, d'identification et de commandes non linéaires.

Connexionnisme et Economie des Transports

A la question: comment appréhender ce que la théorie économique propose comme des objets non représentables sous forme de fonctions, la logique connexionniste fournit un début de réponse.

Ces méthodes dont le domaine par excellence est celui de l'estimation de variétés non linéaires, possèdent des propriétés d'apprentissage, d'adaptation et de robustesse (résistance au bruit) qui sont chères aux macroéconomistes.

Les problèmes rémanents en analyse des données comme le traitement des valeurs manquantes, la prise en compte de variables qui interagissent fortement entre elles, la prédiction simultanée de plusieurs choix individuels de comportements ont trouvé des solutions intéressantes grâce aux RNF.

L'aspect neuromimétique de ces modèles

Les modèles de RNF reposent sur un principe hérité directement des réseaux de neurones biologiques, puisqu'ils sont constitués d'un réseau de processeurs élémentaires à fonction de transfert non linéaire, appelés neurones formels, fortement connectés les uns aux autres et qui communiquent en transmettant des niveaux d'activation via des connexions dans le réseau. L'aspect neuro-mimétique de ces modèles ne se limite pas à l'architecture et au mode de propagation de l'information, car ces modèles disposent aussi de propriétés d'adaptivité grâce auxquelles certains paramètres de l'architecture du réseau ne sont pas fixés a priori par le concepteur mais résultent d'un processus d'apprentissage à partir des données du problème particulier à traiter.

Différentes classes de RNF

De nombreux modèles de RNF sont étudiés et utilisés pour des applications, parmi lesquels on distingue deux catégories principales: les réseaux à couches et les réseaux classifieurs. Ces deux grandes familles de réseaux se différencient à la fois par leur fonction (association de formes d'entrée - sortie pour les réseaux à couches, classification de formes d'entrée pour le second) et par leur type d'apprentissage (apprentissage supervisé pour les réseaux à couches, c'est à dire utilisant l'information de sortie désirée au niveau de chaque forme d'entrée à traiter; apprentissage non supervisé pour les réseaux classifieurs). Citons encore deux autres catégories de modèles classiques mais moins utilisées en pratique : les réseaux à états d'équilibres ou les réseaux probabilistes.

Principes fondamentaux

Les caractéristiques communes aux modèles connexionnistes peuvent se résumer de la façon suivante:

- non linéarité,
- parallélisme (répartition de la représentation de l'information et du traitement de l'information),
- adaptativité.

Apprendre à faire apprendre

La mise en œuvre d'une application utilisant un modèle neuronal s'effectue suivant quatre phases successives :

- ◆ dans un premier temps on fait le choix du modèle à utiliser (type d'architecture, de processus de propagation de l'activation et de méthode d'apprentissage). Ce choix s'accomplit en fonction de la nature du problème à traiter (association entrées - sorties ou classification, données de nature temporelle ou non ... etc.) et des connaissances préalables dont le concepteur dispose sur le phénomène à modéliser. Ces connaissances sont aussi nécessaires à la détermination précise de l'architecture du réseau retenu (nombre de couches de neurones, nombre de neurones dans chaque couche, connectivité).
- ◆ vient ensuite la phase d'apprentissage qui consiste à ajuster itérativement certains paramètres du réseau (typiquement le poids des connexions) en fonction d'exemples fournis.
- ◆ la phase de test permet alors de qualifier le résultat de cet apprentissage en confrontant le réseau ainsi adapté à de nouveaux exemples et en testant ses capacités de généralisation.
- ◆ ce n'est que lorsque cette phase de test est jugée satisfaisante que le réseau pourra véritablement être utilisé.

On voit bien que la mise en œuvre d'une application à base de réseaux de neurones nécessite de porter un soin très attentif à la constitution de la base d'exemples. La capacité de généralisation du réseau, qui détermine son utilité pratique in fine, ne sera satisfaisante que si la base d'apprentissage est soigneusement déterminée, de taille suffisante et bien représentative du domaine sur lequel on veut utiliser le réseau.

L'utilisation de méthodes neuronales ne signifie donc pas que l'on s'affranchit des difficultés classiques de statistique inférentielle (« représentativité » de l'échantillon test), ni qu'on puisse s'abstenir d'un certain niveau d'expertise sur le domaine d'application (choix, codage des informations pertinentes). Remarquons également que si la mise au point et l'adaptation du réseau nécessitent en général un temps de réalisation non négligeable - ainsi qu'un certain savoir faire - la complexité du réseau n'est plus un obstacle en phase d'utilisation puisqu'une simple propagation d'information dans le réseau, opération quasi instantanée, fournit la réponse du système.

La démarche neuronale est sujette comme toute démarche algorithmique d'optimisation aux problèmes classiques de convergence: phase d'apprentissage souvent longue (choix du bon algorithme) - existence d'optima locaux, choix délicat des paramètres de gain.

La facilité de mise en œuvre de ces méthodes dans tout système de production au prix d'une maintenance minimale en a fait un outil pérenne, d'autant plus que leur caractère évolutif, autrement dit la possibilité de relancer l'apprentissage sur de nouveaux cas sans avoir à tout recommencer a joué en leur faveur.

De l'utilité des Réseaux de Neurones Formels

Les techniques neuronales sont des méthodes statistiques non linéaires. Les contraintes d'architecture et de traitement de l'information qu'on définit dans les réseaux confèrent à ces modèles un caractère particulier et original. On dispose de quelques résultats théoriques qui permettent de mieux situer ces

méthodes, mais il est clair qu'il manque à ce jour une théorie mathématique unifiée. Cependant l'introduction de neurones non plus seulement à valeurs réelles ou binaires mais à valeurs complexes ou vectorielles ainsi que la définition des réseaux logiques ont montré qu'une grande généralité tant sur le plan de la forme des données utilisées que sur celui de l'architecture s'avérait possible.

L'objection la plus couramment faite est celle du manque de transparence de l'outil qui en ferait une boîte noire de plus. Indiquons que des travaux portant sur l'extraction de règles ont montré que l'on peut traduire la connaissance issue de l'apprentissage sous une forme accessible bien que réductrice.

Insistons sur le fait que ces méthodes connexionnistes ne sont pas exclusives des méthodes classiques, on peut donc les concevoir comme un « sur-modèle » du modèle linéaire, dans le sens où on peut espérer faire mieux et en tout cas pas plus mal qu'avec un modèle classique linéaire, voire créer des modèles composites associant des modèles neuronaux séparés et des modèles conventionnels statistiques.

Modélisation neuronale des flux de télécommunications et confrontation des résultats avec les modèles gravitaires classiques

Dans la littérature concernant la modélisation des flux de télécommunications, l'approche gravitaire conventionnelle revient très souvent (*Pacey*, 1983; *Rietveld et Janssen*, 1990; *Rossera*, 1990; *Fisher et al.*, 1993; *Guildmann*, 1992) sous la forme suivante :

T_{rs}^{conv} : intensité du trafic téléphonique entre la région r et la région s , inter-régionale si $r \neq s$, intra-régionale si $r = s$.

$$T_{rs}^{conv} = k A_r^{\alpha_1} B_s^{\alpha_2} \exp(-\beta D_{rs}) \quad (r, s = 1, \dots, n)$$

avec :

- A_r = intensité des appels téléphoniques générée par la région d'origine r
- B_s = potentiel d'attractivité des appels de la région de destination s
- D_{rs} = facteur de séparation (mesuré en termes de distance) associé à chaque couple (r, s) représentant l'effet inhibiteur de la séparation géographique
- k = paramètre d'échelle
- $\alpha_1, \alpha_2, \beta$ = paramètres
- n = numéro de la région

L'exemple fait partie des réseaux de neurones à boucle de rétroaction à deux couches, dont la structure hiérarchique est constituée d'une couche (input) I_1 , d'une couche cachée I_2 , et d'une couche d'output I_3 .

$$\begin{aligned} X_{2,i_2} &= \sum_{i_1=1}^{I_1} W_{1,i_1,i_2} Y_{1,i_1} + W_{1,I_1+1,i_2} & (i_2 = 1, \dots, I_2) \\ Y_{2,i_2} &= f(X_{2,i_2}) & (i_2 = 1, \dots, I_2) \\ X_{3,i_3} &= \sum_{i_2=1}^{I_2} W_{2,i_2,i_3} Y_{2,i_2} + W_{2,I_2+1,i_3} & (i_3 = 1, \dots, I_3) \\ Y_{3,i_3} &= f(X_{3,i_3}) & (i_3 = 1, \dots, I_3) \end{aligned}$$

avec

- I_j = nombre de unités de process du $j^{\text{ème}}$ ensemble ($j = 1, 2, 3$)
- i_j = indices associés avec le $j^{\text{ème}}$ ensemble d'unités ($j = 1, 2, 3$)
- Y_{j,i_j} = *output* des unités i_j du $j^{\text{ème}}$ ensemble ($j = 2, 3$)
- X_{j,i_j} = *input* de l'unité de process i_j appartenant au $j^{\text{ème}}$ ensemble ($j = 2, 3$)
- $W_{k,i_k,i_{k+1}}$ = poids du $k^{\text{ème}}$ ensemble entre les $k^{\text{ème}}$ et $(k+1)^{\text{èmes}}$ ensembles d'unités avec :
 - a) $k = 1$: poids de connexion entre la couche cachée et la couche de sortie
 - b) $k = 2$: poids de connexion entre l'ensemble des inputs et la couche cachée
- W_{1,I_1+1,i_2} = unités de biais des unités cachées

- $W_{2,12+1,i3}$ = unités de biais des unités d'output
- $f(X_k, i_k)$ = fonction de transfert des unités cachées i_2 et d'output i_3

$f(X_k, i_k) = \frac{1}{1 + \exp(-aX_k, i_k)}$, $k = 2, 3$: fonction de transfert de forme sigmoïdale, avec :

$$df / dX_k, i_k = af(1 - X_k, i_k) > 0.$$

a) introduction d'une fonction additive d'erreurs (minimisation d'une erreur quadratique moyenne) :
S ensemble d'exemples

$$E(S) = \sum_{I \in S} e(y_I, x_I, \omega) = \frac{1}{2} \sum_{I \in S} [y_I - \hat{y}_I]^2$$

la fonction d'erreur e est une mesure de la distance entre la sortie désirée y et l'actuelle sortie \hat{y} .
La minimisation de E est appelé processus d'apprentissage.

b) rétro-propagation du gradient de l'erreur quadratique $E(S)$

$$W^{new} = W^{old} + \lambda * \Delta W$$

$$\Delta W_{k,i_k,i_{k+1}}(t+1) = -\eta \frac{\delta E(S)}{\delta W_{k,i_k,i_{k+1}}} + \gamma \Delta W_{k,i_k,i_{k+1}}(t), 0 \leq \gamma < 1.$$

c) le problème de surdétermination du modèle : la validation croisée.

Construction d'un indicateur de performance dit *Average Relative Variances* (ARV) :

$$ARV(S) = \frac{\sum_{I \in S} (y_I - \hat{y}_I)^2}{\sum_{i \in S} (y_i - \bar{y})^2} = \frac{1}{\hat{\sigma}^2} * \frac{1}{N_S} * \sum_{I \in S} (y_I - \hat{y}_I)^2$$

d) mesure comparée des estimations entre modèle gravitaire conventionnel T_{rs}^{conv} et modèle neuronal

T_{rs}^{neur} et valeurs observées T_{rs}^{obs} :

- ♦ comparaison entre résidus absolus $(T_{rs}^{obs} - T_{rs}^{conv})$ et $(T_{rs}^{obs} - T_{rs}^{neur})$
- ♦ comparaison entre résidus relatifs $\frac{(T_{rs}^{obs} - T_{rs}^{conv})}{T_{rs}^{obs}}$ et $\frac{(T_{rs}^{obs} - T_{rs}^{neur})}{T_{rs}^{obs}}$

Exemple : prévision des flux de trafics téléphoniques entre 32 districts australiens en 1991 [M. Fischer et S. Gopal].

- * incapacité des deux types de modélisation (conventionnelle et neuronale) à prédire les grands flux et sur-prédiction des flux faibles, et même configuration spatiale des résidus,
- * une prédiction légèrement meilleure par le modèle neuronal (EQM).

Note : nous ne discuterons pas ici du cas des algorithmes génétiques et renvoyons le lecteur aux articles indexées dans la bibliographie.

Bibliographie

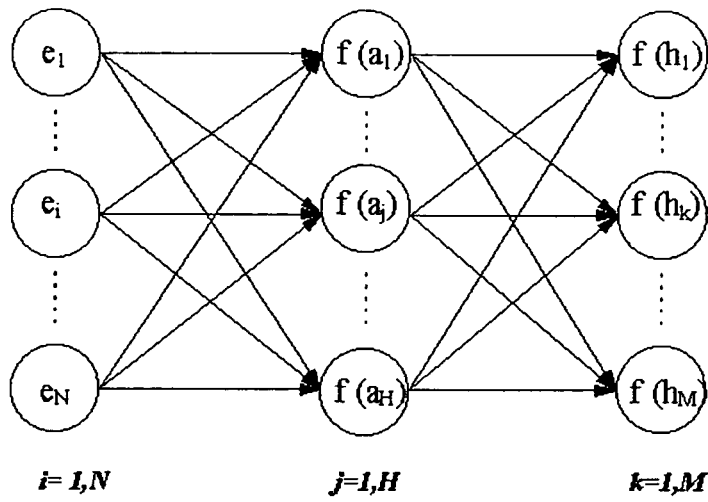
Réseaux de neurones

- Fischer M.M. (1998), "Computational neural networks. A new paradigm for spatial analysis", *Environment and Planning A*.
- Fischer M.M., Gopal S. (1996), "Learning in Single Hidden-Layer Feedforward Network Models: Backpropagation in a Spatial Interaction Modelling Context", *Geographical Analysis*, Vol.28, No.1, pp. 38-55.
- Fischer M.M., Gopal S. (1994), "Artificial Neural Networks: a New Approach to Modelling Interregional Telecommunication Flows", *Journal of Regional Sciences*, Vol. 34, No. 4, pp. 503-527.
- Fischer M. M., Essetzbichler J., Gassler H., Trichtl G. (1993), "Interregional and International Telephone Communication : Aggregate Traffic Models and Empirical Evidence for Austria", *Sistemi Urbani*, 15.
- Guldmann J. M. (1992), "Modelling Residential and Business Telecommunication Flows : A regional Point-to-point Approach", *Geographical Analysis*, Vol. 24, pp. 121-141.
- Holland J. H. (1975), *Adaptation in Natural and Artificial Systems*, Ann Arbor : The university of Michigan Press.
- Openshaw S. (1988), "Building an Automated Modelling System to Explore a Universe of Spatial Interaction Models", *Geographical Analysis*, Vol. 20, No. 1, pp. 31-46.
- Pacey P.L. (1983), "Long Distance Demand : A Point-to-point Model", *Southern Economic Journal*, Vol. 49, pp. 1094-1107.
- Rietveld P., Jansen L. (1990), "Telephone Calls and Communication Barriers : The Case of The Netherlands", *The Annals of Regional Science*, Vol. 24, pp. 307-318.
- Rossera F. (1990), "Discontinuities and Barriers in Communications : The Case of Swiss Communities of Different Language", *The Annals of Regional Science*, Vol. 24, pp. 319-336.
- Schwefel H. P. (1995), *Evolution and Optimization Seeking*, New York: Wiley.
- Wilson A. G. (1971), "A Family of Spatial Interaction Models and Associated Development", *Environment and Planning A*. Vol. 3, pp. 1-32.

Algorithmes génétiques

- Caudele TP, Dolan CP (1989), "Parametric connectivity : Training of constrained networks using genetic algorithms", in Schaffer JD (ed.) *Proceedings of the Third International Conference on Genetic Algorithms and their Applications*, pp.370-374. Morgan Kaufmann, San Mateo, CA.
- Diplock G., Openshaw S. (1996), "Using Simple Genetic Algorithms to Calibrate Spatial Interaction Models", *Geographical Analysis*, Vol. 28, No. 3, pp. 262-279.
- Fischer M.M., Leung Y. (1998), "A genetic-algorithms based evolutionary computational neural network for modelling spatial interaction data", *Annals of Regional Science*, Vol. 32, pp. 437-458.
- Goldberg D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, Pa.: Addison-Wesley.
- Koza JR (1993), *Genetic Programming*, the MIT Press, Cambridge, MA, London, UK.
- Michalewicz Z (1992), *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin Heidelberg, New York.
- Openshaw S. (1998), "Neural network, genetic and fuzzy logic models of spatial interaction", *Environment and Planning A*.
- Openshaw S., Oppenshaw C. (1997), *Artificial intelligence in geography*, Wiley, Chichester New York.
- Yao X (1993), "A review of evolutionary artificial neural networks", *International Journal of Intelligent Systems*, 8(4), pp. 539-567.

Couche d'entrée I Couche cachée J Couche de sortie K



BIBLIOGRAPHIE

OUVRAGES GENERAUX SUR LES MIS

- Courgeau D. (1980), Analyse quantitative des migrations humaines, Masson, Paris, 225 p..
- Erlander S., Stewart N. F. (1990), The Gravity Model in Transportation Analysis : Theory and Extensions, Topics in Transportation, VSP, Utrecht, The Netherlands, 226 p..
- Fotheringham A.S., O'Kelly M.E. (1989), Spatial Interaction Models : Formulations and applications, Kluwer Academic Publishers, London, 221 p..
- Fustier B. (1977), "Les interactions spatiales", dans Ponsard C., Analyse Economique Spatiale, Chap. 5, pp. 193-230.
- Pini G. (1992), "L'interaction spatiale", dans Bailly A., Ferras R., Pumain D., Encyclopédie de Géographie, Chap. 29, pp. 557-576.
- Sanders L. (1992), Système de villes et synergétique, Anthropos-Economica, Coll. Villes, Paris, 274 p..
- Stillwell J., Congdon P. (eds) (1992), Migration Models : macro and micro approaches, Belhaven Press, London and New-York, 330 p..
- Nijkamp P., Reggiani A. (1992), Interaction, Evolution and Chaos in Space, Springer-Verlag, 278 p..
- Vermot-Desroches B. (Mai 1986), Interdépendances spatiales et théorie moderne de l'information, Collection de l'I.M.E., n°30, LATEC, Université de Bourgogne, 617 p..
- Sen A., Smith T. E. (1995), Gravity Models of Spatial Interaction Behaviour, Advances in Spatial and Network Economics, Springer-Verlag, 572 p.

**ARTICLES PARUS
DANS LES 'REVUES TRANSPORT'**

1. Théorie Coût-Efficacité & Modèles d'Interaction Spatiale
2. Maximisation de l'entropie & Modèles d'Interaction Spatiale
3. Surplus/Maximisation de l'utilité et Modèles d'Interaction Spatiale
4. Modèle entropique, Modèle gravitaire and modèle Logit
5. Calibrage des Modèles d'Interaction Spatiale (Entropie / Logit)

6. Modèles entropiques hiérarchisés (Nested Entropy Models)
7. Modèles entropiques dynamiques

8. Algorithmes de maximisation des modèles entropiques
9. Algorithmes de type 'Bregman's Balancing' et 'Proximal Point'
10. Effets de barrière

1. Théorie Coût-Efficacité & Modèles d'Interaction Spatiale

- S. Erlander (1988), "On the Relationship between the Discrete and Continuous Models for Combined Distribution and Assignment", *Transportation Research*, Vol. **22B(5)**, pp. 371-382.
- E. Smith (1988), "A Cost-Efficiency Theory of Dispersed Network Equilibria", *Environment and Planning A*, Vol. **20**, pp. 231-266.
- E. Smith (1987), "Poisson Gravity Models of Spatial Flows", *Journal of Regional Science*, Vol. **27(3)**, pp. 315-340.
- E. Smith (1983), "A Cost-Efficiency Approach to the Analysis of Congested Spatial-Interaction Behaviour", *Environment and Planning A*, Vol. **15**, pp. 435-464.
- E. Smith (1978), "A Cost-Efficiency Principle of Spatial Interaction Behaviour", *Regional Science and Urban Economics*, Vol. **8**, pp. 313-337.

2. Maximisation de l'entropie & Modèles d'Interaction Spatiale

- Bruzelius (1981), "Accessibility, Entropy and the Distribution and Assignment of Traffic: A Comment", *Transportation Research*, Vol. **15B**, pp. 165-172.
- Decay and A. Norcliffe (1977), "A Flexible Doubly-constrained Trip Distribution Models", *Transportation Research*, Vol. **11B**, pp. 203-204.
- Guiasu and B. V-Desroches (1983), "Some Remarks on Wilson's Gravity Model and a Slight Generalisation of it", *Transportation Research*, Vol. **17B(4)**, pp. 333-337.
- Harris (1986), "A Continuous Dual Method for the Hitchcock Problem", *Transportation Research B*, Vol. **20B(1)**, pp. 71-75.
- O. Jornsten, T. Larsson, J. T. Lundgren and A. Migdalas (1990), "An Entropy Model with Variable Target", *Environment and Planning A*, Vol. **22**, pp. 493-506.
- O. Jornsten and J. T. Lundgren (1989), "An Entropy-Based Modal Split Model", *Transportation Research*, Vol. **23B(5)**, pp.345-359R. Roy and P. F. Lesse (1985), "Entropy Model with Entropy Constraints on Aggregated Events", *Environment and Planning A*, Vol. **17**, pp. 1669-1674.

- Sasaki (1972), "Estimation of Person Trip Patterns through Markov Chains", *Proceedings of the 5th ISTTT*, pp. 119-130.
- Sasaki (1967), "Probability Methods to Estimate Trip Distributions", *Annual Report of Roads 1967*, pp. 121-134.
- B. Slater (1989), "Transport Coefficients for the Equilibration of Spatial Interaction Processes", *Transportation Research*, Vol. **23B(1)**, pp. 49-52.
- K. Walsh and R. W. Gibberd (1983), "A Probability Analysis of some Spatial Interaction Models", *Transportation Research*, Vol. **17B(3)**, pp. 193-200.
- G. Wilson (1975), "Some New Forms of Spatial Interaction Model: A Review", *Transportation Research*, Vol. **9B**, pp. 167-179.
- Wilson (1970), "The Use of the Concept of Entropy in System Modelling", *Operations Research Quarterly*, Vol. **21**, pp. 247-265.
- Wilson (1969), "The Use of Entropy Maximising Models in the Theory of Trip Distribution, Mode Split and Route Split", *Journal of Transport Economics and Policy*, **3**, pp. 108-126.
- Wilson (1967), "A Statistical Theory of Spatial Distribution Models", *Transportation Research*, Vol. **1**, pp. 253-269.

3. Surplus/Maximisation de l'utilité et Modèles d'Interaction Spatiale

- Cochrane (1975), "A Possible Economic Basis for the Gravity Models", *Journal of Transport Economic and Policy*, Vol. **9**, pp. 34-49.
- Crouchley (1987), "An Examination of the Equivalence of Three Alternative Mechanisms for Establishing the Equilibrium - Solutions of the Production-Constrained Spatial Interaction Model", *Environment and Planning A*, Vol. **19**, pp. 861-874.
- Golob and M.J. Beckmann (1971), "A Utility Model for Travel Forecasting", *Transportation Science*, Vol. **5**, pp. 79-90.
- Harris and A. G. Wilson (1978), "Equilibrium Values and Dynamics of Attractiveness Terms in Production-Constrained Spatial-Interaction Models", *Environment and Planning A*, Vol. **10**, pp. 371-388.
- Leonardi (1978), "Optimum Facility Location by Accessibility Maximising", *Environment and Planning A*, Vol. **10**, pp. 1287-1305.
- Neuberger (1971), "User Benefits in Evaluation of Transport-land Use Plans", *Journal of Transport Economic and Policy*, Vol. **5**, pp. 52-75.
- Nijkamp (1975), "Reflection on Gravity and Entropy Models", *Regional Science and Urban Economics*, Vol. **5**, pp. 203-255.
- C. W. L. Williams and K. S. Kim (1990), "Location-Spatial Interaction Models: 2. Competition between Independent Firms", *Environment and Planning A*, Vol. **22**, pp. 1155-1168.
- C. W. L. Williams, K. S. Kim and D. Martin (1990), "Location-Spatial Interaction Models: 1. Benefit-Maximising Configurations of Services", *Environment and Planning A*, Vol. **22**, pp. 1079-1089.
- C. W. L. Williams (1977), "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit", *Environment and Planning A*, Vol. **9**, pp. 285-344.
- W.L. Williams (1976), "Travel Demand Models, Duality Relations and User Benefit Analysis," *Journal of Regional Science*, Vol. **16**, pp. 147-166.

4. Modèle entropique, modèle gravitaire and modèle Logit

- Anas (1983), "Discrete Choice Theory, Information Theory and the Multinomial Logit and Gravity Models", *Transportation Research*, Vol. **17B(1)**, pp. 13-23.
- Bruzelis (1981), "Accessibility, Entropy and the Distribution and Assignment of Traffic : A Comment", *Transportation Research*, Vol. **15B**, pp. 165-172.
- Erlander (1988), "On the Relationship between the Discrete and Continuous Models for Combined Distribution and Assignment", *Transportation Research*, Vol. **22B(5)**, pp. 371-382.

- Erlander (1982), "Accessibility, Entropy and the Distribution and Assignment of Traffic Revisited", *Transportation Research*, Vol. **16B**(6), pp. 471-472.
- Fisk (1985), "'Entropy and Information Theory': Critique, Comments, and Reply", *Environment and Planning A*, Vol. **17**, pp. 679-710.
- Florian (1979), "Utility, Entropy and a 'Paradox' of Traffic Flow", ?, pp. 1-10.
- M. Macgill (1977), "Theoretical Properties of Biproportional Matrix Adjustments", *Environment and Planning A*, Vol. **9**, pp. 687-701.
- Miyagi (1985), "On the Formulation of a Stochastic User Equilibrium Model Consistent with Random Utility Theory", Proceedings of the 4th WCTR, pp. 1619-1635.
- Morisugi and L. D. Hanh (1994), "Logit Model and Gravity Model in the Context of Consumer Behaviour Theory", *Journal of Infrastructure Planning and Management*, No. **488/E-23**, pp. 111-119.
- R. Roy (1981), "A Note on Consistency of Choice Units in Entropy Models", *Transportation Research*, Vol. **15B**, pp. 159-164.
- R. Roy and P. F. Lesse (1981), "On Appropriate Microstate Descriptions in Entropy Modelling", *Transportation Research*, Vol. **15B**, pp. 85-96.

5. Calibrage des Modèles d'Interaction Spatiale (Entropie / Logit)

- Anas (1988), "Statistical Properties of Mathematical Programming Models of Stochastic Network Equilibrium", *Journal of Regional Science*, Vol. **28**(4), pp.511-530
- Andersson (1981), "On the Convergence of Iterative Methods for the Distribution Balancing Problem", *Transportation Research*, Vol. **15B**, pp. 173-201.
- Boyce E., Chon K.S., Lee Y.J., Lin K.T. and LeBlanc L.J (1983), "Implementation and Computational Issues for Combined Models of Location, Destination, Mode, and Route Choice", *Environment and Planning A*, Vol. **15**, pp. 1219-1230.
- Cesario J. (1975), "Least-Squares Estimation of Trip Distribution Parameters", *Transportation Research*, Vol. **9B**, pp. 13-18.
- Erlander (1978), "A Note on Balancing Procedures for Trip Distribution Estimation", *Transportation Research*, Vol. **12B**, pp. 335-336.
- Erlander and Stewart N.F. (1978), "Interactivity, Accessibility and Cost in Trip Distribution", *Transportation Research*, Vol. **12B**, pp.291-293
- Fisk and G. R. Brown (1975), "The Role of Model Parameters in Trip Distribution Models", *Transportation Research*, Vol. **9B**, pp. 143-148.
- Gray H. and Sen A.K. (1983), "Estimating Gravity Model Parameters: A Simplified Approach Based on the Odds Ratio", *Transportation Research*, Vol. **17B**(2), pp. 117-131.
- Jerrerson J. and Scott C.H. (1979), "The Analysis of Entropy Models with Equality and Inequality Constraints", *Transportation Research*, Vol. **13B**, pp. 123-132.
- Malone M., E and F. DiCesare (1978), "Calibration of Gravity Models", *Transportation Research*, Vol. **12B**, pp. 181-184.
- Mohr, W. H. Crown and K. R. Polenske (1987), "A Linear Programming Approach to Solving Infeasible RAS Problems", *Journal of Regional Science*, Vol. **27**(4), pp.587-603.
- Richardson (1985), "Input-Output and Economic Base Multipliers: Looking Backward and Forward", *Journal of Regional Science*, Vol. **25**(4), pp. 608-661.
- Williams (1976), "A Comparison of some Calibration Techniques for Doubly Constrained Models with an Exponential Cost Function", *Transportation Research*, Vol. **10B**, pp. 91-104.
- Wills J. (1982), "On the Equivalence of Common Rearrangements for Sum Constrained Travel Estimation Problems", *Transportation Research*, Vol. **16B**, pp. 339-359.

6. Nested Entropy Models

- Boyce E., Chon K.S., Lee Y.J., Lin K.T. and LeBlanc L.J . (1983), "Implementation and Computational Issues for Combined Models of Location, Destination, Mode, and Route Choice", *Environment and Planning A* , Vol. **15**, pp. 1219-1230.

- Ben-Akiva (1974), "Structure of Passenger Travel Demand Models", *Transportation Research Record*, **526**, pp. 26-42.
- Brice (1989), "Derivation of Nested Transport Models within a Mathematical Programming Framework", *Transportation Research-B*, Vol. **23B(1)**, pp. 19-28.
- Daly (1987), "Estimating 'Tree' Logit Models", *Transportation Research-B*, Vol. **21B(4)**, pp. 251-267.
- Daly (1982), "Estimating Choice Models Containing Attraction Variables", *Transportation Research-B*, Vol. **16B(1)**, pp.5-15
- Fisk S. and Boyce D.E. (1984), "A Modified Composite Cost Measure for Probabilistic Choice Modelling", *Environment and Planning A*, Vol. **16**, pp. 241-248.
- Daly and S. Zachary (1978), "Improved Multiple Choice Models", In *Determinants of Travel Choice*(Ed. D.A. Hensher and M.Q. Dalvi), Saxon House.
- MacFadden (1981), "Econometric Models of Probabilistic Choice", In *Structural Analysis of Discrete Data with Economic Application* (Ed. C.F. Manski and D. MacFadden), MIT Press.
- Manheim (1973), "Practical Implications of Some Fundamental Properties of Travel Demand Models", *Highway Research Record*, **422**, pp. 21-38.
- Ortuzar (1983), "Nested Logit Models for Mixed-mode Travel in Urban Corridors", *Transportation Research*, Vol. **17A(4)**, pp. 283-299.
- Sheffi (1979), "Estimating Choice Probabilities among Nested Alternatives", *Transportation Research-B*, Vol. **13B**, pp. 189-205.
- Sobel (1980), "Travel Demand Forecasting by Using the Nested Multinomial Logit Models", *Transportation Research Record*, **775**, pp. 48-55.
- Williams C.W.L. and Ortuzar J.D. (1982), "Behavioural Theories of Dispersion and the Mis-specification of Travel Demand Models", *Transportation Research-B*, Vol. **16B(3)**, pp. 167-219.
- Williams C.W.L. (1977), "On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit", *Environment and Planning A*, Vol. **9**, pp. 285-344.

7. Dynamic Entropy Models

- Fischer and Nijkamp (1987), "From Static towards Dynamic Discrete Choice Modelling. A State of the Art Review", *Regional Science and Urban Economics*, Vol. **17**, pp. 3-27.
- Harris and Wilson A.G. (1978), "Equilibrium Values and Dynamics of Attractiveness Terms in Production-Constrained Spatial-Interaction Models", *Environment and Planning A*, Vol. **10**, pp. 371-388.
- Leonardi and Casti J. (1986), "Agglomerative Tendencies in the Distribution of Populations", *Regional Science and Urban Economics*, Vol. **16**, pp. 43-56.
- Nijkamp and A. Reggiani (1988), "Dynamic Spatial Interaction Models: New Directions", *Environment and Planning A*, Vol. **20**, pp. 1449-1460.
- Nijkamp (1987), "Discrete Spatial Choice Analysis. Editorial Introduction", *Regional Science and Urban Economics*, Vol. **17**, pp. 1-2.
- Nijkamp and J. Poot (1987), "Dynamics of Generalised Spatial Interaction Models", *Regional Science and Urban Economics*, Vol. **17**, pp. 367-390.
- Oppenheim (1989), "Equilibrium Commercial Activity and Travel Distributions: Incorporating Endogenous Prices and Travel Costs into the Harris-Wilson Framework, with some Numerical Experiments", *Transportation Research-B*, Vol. **23B(3)**, pp. 225-242.

8. Algorithmes de maximisation des modèles entropiques

- Censor, De Pierro, Elfving, Herman and Iusem (1990), "On Iterative Methods for Linearly Constrained Entropy Maximisation", *Numerical Analysis Mathematical Modelling*, Vol. **24**, pp. 145-163.
- Censor and Segman J. (1987), "On Block-Iterative Entropy Maximisation", *Journal of Information & Optimisation Sciences*, Vol. **8(3)**, pp. 275-291.

- Denbo S. (1976), "A Set of Geometric Programming Test Problems and their Solutions", *Mathematical Programming*, Vol. 10, pp. 192-213
- Evans W. (1970), "Some Properties of Trip Distribution Methods", *Transportation Research*, Vol. 4, pp. 19-36.
- Denbo S. (1979), "Second Order Algorithms for the Polynomial Geometric Programming Dual, Part E : Analysis", *Mathematical Programming*, Vol. 17, pp. 156-175.
- Erlander (1978), "A Note on Balancing Procedures for Trip Distribution Estimation", *Transportation Research*, Vol. 12, pp. 335-336.
- Evans P. and Kirby H.R. (1974), "A Three-Dimensional Furness Procedure for Calibrating Gravity Models", *Transportation Research*, Vol. 8, pp. 105-122.
- Evans (1976), "Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment", *Transportation Research*, Vol. 10B(1), pp. 37-57.
- Fang C. and Rajasekera J.R. (1989), "Quadratically Constrained Minimum Cross-Entropy Analysis", *Mathematical Programming*, Vol. 44, pp. 85-96.
- Lamond and Stewart N.F. (1981), "Bregman's Balancing Method", *Transportation Research-B*, Vol. 15B(4), pp. 239-248.
- Powell J.D. (1988), "An Algorithm for Maximising Entropy Subject to Simple Bounds", *Mathematical Programming*, Vol. 42, pp. 171-180
- Robillard and Stewart N.F. (1974), "Iterative numerical Methods for Trip Distribution Problems", *Transportation Research*, Vol. 8, pp. 575-582.

9. Algorithmes de type 'Bregman's Balancing' et 'Proximal Point'

- Bregman M. (1967), "Proof of the Convergence of Sheleikhovskii's Method for a Problem with Transportation Constraints", *Zh. vychisl. Mat. mat. Fiz.*, Vol. 7(1), pp. 147-156
- Bregman M. (1967), "The Relaxation Method of Finding the Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming", *Zh. vychisl. Mat. mat. Fiz.*, Vol. 7(3), pp. 620-631.
- Borwein M. and Lewis A.S. (1991), "Duality Relationships for Entropy-Like Minimisation Problems", *SIAM Journal of Control and Optimisation*, Vol. 29(2), pp. 325-338.
- Censor and Zenios S.A. (1992), "Proximal Minimisation Algorithm with D-Functions", *Journal of Optimisation Theory and Applications*, Vol. 73(3), pp. 451-464.
- Censor (1981), "Row-Action Methods for Huge and Sparse Systems and their Applications", *SIAM Review*, Vol. 23(4), pp. 444-466.
- Chen and Teboulle M. (1993), "Convergence Analysis of a Proximal-Like Minimisation Algorithm Using Bregman Functions", *SIAM Journal of Optimisation*, Vol. 3(3), pp. 538-543.
- De Pierro and Iusem (1986), "A Relaxed Version of Bregman's Method for Convex Programming", *Journal of Optimisation Theory and Applications*, Vol. 5(3), pp. 421-440.
- Eggermont P.B. (1990), "Multiplicative Iterative Algorithms for Convex Programming", *Linear Algebra and its Applications*, Vol. 130, pp. 25-42.
- Gamboa and Gassiat E. (1994), "The Maximum Entropy Method on the Mean : Applications to Linear Programming and Superresolution", *Mathematical Programming*, Vol. 66, pp. 103-122.
- Hettich (1986), "An Implementation of a Discretization Method for Semi-Infinite Programming", *Mathematical Programming*, Vol. 34, pp. 354-361.
- Iusem N., Svaiter B.F. and Teboulle M. (1994), "Entropy-Like Proximal Methods in Convex Programming", *Mathematics of Operation Research*, Vol. 19(4), pp. 790-814.
- Iusem N. (1991), "On Dual Convergence and the Rate of Primal Convergence of Bregman's Convex Programming Method", *SIAM Journal of Optimisation*, Vol. 1(3), pp. 401-423.
- Lamond and N. F. Stewart (1981), "Bregman's Balancing Method", *Transportation Research*, Vol. 15B(4), pp. 239-248.
- Lemaire (1989), "The Proximal Algorithm", *International Series of Numerical Mathematics*, Vol. 87, pp. 73-87.
- Teboulle (1992), "Entropic Proximal Mappings with Applications to Non-linear Programming", *Mathematics of Operations Research*, Vol. 17(3), pp. 670-690.

10. Effets de barrière

- Bopda A., Grasland C (1994), "Migrations, régionalisations et régionalismes au Cameroun ", *Espace, Population, Sociétés*, 4, pp. 109-129.
- Bröcker J., Rohweder H.C. (1990), "Barriers to international trade : methods of measurement and empirical evidence ", *Annals of Regional Science*, 24, pp. 289-305.
- Cattan N., Grasland C. (1996), "Air Traffic Fields of Western European Cities", dans Pumain D., Saint-Julien T., (eds), *Urban Networks in Europe*, Jon Libey Eutotext / INED , pp. 115-128.
- Dorigo G., Tobler W. (1983), "Push-Pull Migration Laws ", *Annals of the Association of American Geographers*, 73(1), pp. 1-17.
- Grasland C. (1994), "Limites politiques et barrières migratoires : l'exemple de l'ex-Tchécoslovaquie (1965-1989)", dans Galluser W. (eds), *Political Boundaries and Coexistence*, IGU Symposium, Basle, May 1994, pp. 45-440.
- Hägerstrand T. (1953), "Innovation diffusion as a spatial process", Translation by A. Pred, University of Chicago Press.
- Klaassen H., Wagenaar S., Van de Weg A. (1972), "Measuring psychological distances between the Flemings and the Walloons", *Papers of the Regional Science Association*, 29, pp. 45-62.
- Mackay J.R. (1958), "The interactance hypothesis and boundaries in Canada : a preliminary study ", *Canadian Geographer*, 11, pp. 1-8.
- Ravenstein E.G. (1885), "The laws of migration", *Journal of the Royal Statistic Society*, 48, pp. 167-177.
- Senior M. (1979), "From gravity modelling to entropy maximising : a pedagogic guide ", *Progress in Human Geography*.
- Stewart J.Q. (19XX ?), "Empirical mathematical rules concerning the distribution and equilibrium of population ", *Geographical Review*, 48, pp. 461-485.
- Stouffer S.A. (1940), "Intervening opportunities : a theory relating mobility and distance", *American Sociological Review*, V, pp. 845-867.

Extrait de :

- Erlander S., Stewart F. (1985), The gravity model in transportation analysis : theory and extensions, chap. 7, pp. 159-172, Topics in Transportation, Utrecht, The Netherlands, Tokyo, Japan.

Chapter 7 Balancing methods

In the historical background given in Section 2.1, reference was made to the 'commonly used balancing method' for finding the parameters r_i and s_j of the doubly constrained gravity model. In this chapter, a fairly complete treatment of the method will be given, insofar as applications related to the gravity model are concerned. Balancing methods have other applications outside of transportation planning, but these will not be discussed in detail here. Similarly, other classes of numerical methods which may profitably be applied in the context of the gravity model will not be discussed in full detail, and methods for solving the traffic assignment problem will not be discussed at all. While no treatment of the gravity model in transportation analysis would be complete without a careful presentation of the closely related balancing method, a wider treatment of the relative merits of this and other methods, for solving this and other problems, properly belongs in books on Numerical Analysis or Operations Research.

The plan for this chapter is similar to that for the book as a whole. Just as in Chapter 1 we introduced a certain convex minimization problem more or less without motivation, expecting the reader to take it on faith that many useful results would later be derived from it, so here we begin by introducing a certain general method, deriving special cases of the method in subsequent sections of the chapter.

7.1. Bregman's method

Consider problem (1.1) in the special case $f(y) \equiv 0$:

$$\text{Min}_{x \in \mathcal{F}} \sum_{k \in \mathcal{K}} x_k \{ \log(x_k/x_k^0) - 1 \}, \quad (7.1)$$

where \mathcal{F} is a certain compact polyhedron with

$$x \geq 0. \quad (7.2)$$

It is now convenient to distinguish two possibilities for the definition of \mathcal{F} (supposed non-empty):

$$\mathbf{Ax} = \mathbf{b}, \quad (7.3a)$$

$$\mathbf{Ax} \geq \mathbf{b}, \quad (7.3b)$$

where \mathbf{A} is an $m \times n$ matrix with no zero rows, and \mathbf{b} is an m -vector.

We shall sometimes want to refer to the problem defined by (7.3a), and at other times to the problem defined by (7.3b).

To motivate the method that we are about to present, we may write down the dual objective for the problem defined by (7.3a):

$$\mathcal{Q}(\lambda) = - \sum_{k \in \mathcal{L}} x_k^0 \exp \left(\sum_{i=1}^m \lambda_i a_{ik} \right) + \sum_{i=1}^m \lambda_i b_i,$$

obtained by substituting $x_k = x_k^0 \exp(\lambda^T \mathbf{a}_{\cdot k})$, $k \in \mathcal{L}$, into the expression for the minimum of the Lagrangian (the sign of λ has been changed relative to Chapter 1). Consider now a method that sequentially sets each component of the gradient of \mathcal{Q}

$$\frac{\partial}{\partial \lambda_i} \mathcal{Q}(\lambda) = - \sum_{k \in \mathcal{L}} a_{ik} x_k^0 \exp \left(\sum_{i=1}^m \lambda_i a_{ik} \right) + b_i,$$

to zero by modifying one component of the Kuhn-Tucker vector λ at a time, i.e. we look for $\lambda_i^{(p+1)}$ and $\mathbf{x}^{(p+1)}$ such that

$$\sum_{k \in \mathcal{L}} a_{ik} x_k^{(p+1)} - b_i = 0$$

and

$$x_k^{(p+1)} / x_k^{(p)} = \exp((\lambda_i^{(p+1)} - \lambda_i^{(p)}) a_{ik}), \quad k \in \mathcal{L}.$$

Bregman's method in the case of (7.3a) is defined by exactly these two equations, except that $k \in \mathcal{L}$ is replaced by $k \in \mathcal{D}$: the method can be proved to converge even in the case when certain x_k are forced to zero by (7.2) and (7.3a); that is, $x_k^{(p)} > 0$ for all finite p , but if $k \in \mathcal{D} \sim \mathcal{L}$, $\lim_{p \rightarrow \infty} x_k^{(p)} = 0$.

Bregman's method is more general than this, however, and in particular, a different specialization of the method yields an algorithm for (7.1), (7.2), and (7.3b), as described below.

We note in passing that Bregman's method is applicable also to the minimization of a wider class of objective functions subject to (7.2) and (7.3a) or (7.3b) than that defined by (7.1), and converges for a wider class of cyclic strategies for the balancing sequence (choice of constraint to be balanced at each step) than the particular strategy defined in the following algorithms.

The initial approximation $\mathbf{x}^{(0)}$ is taken to be the vector \mathbf{x}^0 .

Algorithm for (7.1), (7.2) and (7.3a):

Initialization

$$x_k^{(0)} \leftarrow x_k^0 \quad k \in \mathcal{D}$$

$$\lambda_i^{(0)} \leftarrow 0 \quad i = 1, \dots, m$$

$$p \leftarrow 0$$

$$i \leftarrow 1$$

General step (Balance constraint i)

Find the unique solution $\mathbf{x}^{(p+1)}$ and δ of

$$\log(x_k^{(p+1)} / x_k^{(p)}) + \delta a_{ik} = 0, \quad k \in \mathcal{D}, \quad (7.4)$$

and

$$\sum_{k \in \mathcal{D}} a_{ik} x_k^{(p+1)} = b_i. \quad (7.5)$$

Then

$$\lambda_i^{(p+1)} \leftarrow \lambda_i^{(p)} - \delta$$

$$i \leftarrow (i \text{ modulo } m) + 1.$$

$$p \leftarrow p + 1.$$

It can be shown that the above algorithm converges provided only that (7.2) and the *single* equation (7.5) do not alone force some $x_k^{(p+1)}$ to be zero.

The following algorithm is applicable in the case of the inequality constraints (7.3b), and convergence is guaranteed under the same weak hypothesis (mentioned in the previous paragraph).

Algorithm for (7.1), (7.2), and (7.3b):

Initialization

$$x_k^{(0)} \leftarrow x_k^0 \quad k \in \mathcal{Q}$$

$$\lambda_i^{(0)} \leftarrow 0 \quad i = 1, \dots, m$$

$$p \leftarrow 0$$

$$i \leftarrow 1$$

General step (Balance constraint i)

Compute

$$r_i = \sum_{k \in \mathcal{Q}} a_{ik} x_k^{(p)} - b_i.$$

If ($r_i = 0$) or ($r_i > 0$ and $\lambda_i^{(p)} = 0$) then

$$x_k^{(p+1)} \leftarrow x_k^{(p)} \quad k \in \mathcal{Q}, \quad \lambda_i^{(p+1)} \leftarrow \lambda_i^{(p)} \quad \{\text{goto End of step}\};$$

otherwise find the unique solution y and δ of

$$\log(y_k/x_k^{(p)}) + \delta a_{ik} = 0, \quad k \in \mathcal{Q}$$

$$\sum_{k \in \mathcal{Q}} a_{ik} y_k = b_i,$$

$$\text{if } (r_i < 0) \text{ or } (r_i > 0 \text{ and } \delta \leq \lambda_i^{(p)})$$

$$\text{then } \lambda_i^{(p+1)} \leftarrow \lambda_i^{(p)} - \delta$$

$$x_k^{(p+1)} \leftarrow y_k \quad k \in \mathcal{Q} \quad \{\text{goto End of step}\}$$

otherwise compute $x^{(p+1)}$ from

$$\log(x_k^{(p+1)}/x_k^{(p)}) + \lambda_i^{(p)} a_{ik} = 0, \quad k \in \mathcal{Q}$$

$$\lambda_i^{(p+1)} \leftarrow 0$$

{End of step}

$$i \leftarrow (i \text{ modulo } m) + 1$$

$$p \leftarrow p + 1.$$

7.2. The Kruithof algorithm for the ordinary gravity model

If the constraint matrix A defines the usual transportation constraints, then (7.4) becomes

$$\log(t_{ij}^{(p+1)}/t_{ij}^{(p)}) + \delta = 0, \quad (i, j) \in \mathcal{Q}, \quad (7.6)$$

that is, $t_{ij}^{(p+1)} = t_{ij}^{(p)} \cdot \exp(-\delta)$.

If i in the Bregman (equality constraints) algorithm corresponds to a row constraint, then (7.5) becomes

$$\sum_{j: (i,j) \in \mathcal{Q}} t_{ij}^{(p+1)} = a_i, \quad i \in \mathcal{I}, \quad (7.7a)$$

where i in (7.7a) has the meaning of (3.1) and (3.2); similarly, if i in the Bregman algorithm corresponds to a column constraint, then (7.5) becomes

$$\sum_{i: (i,j) \in \mathcal{Q}} t_{ij}^{(p+1)} = b_j, \quad j \in \mathcal{J}. \quad (7.7b)$$

Solving for $\exp(-\delta)$ in (7.6) and (7.7a), we obtain for row i :

$$t_{ij}^{(p+1)} = t_{ij}^{(p)} \left\{ a_i / \sum_{j: (i,j) \in \mathcal{Q}} t_{ij}^{(p)} \right\}, \quad (i, j) \in \mathcal{Q} \quad (7.8a)$$

and in (7.6) and (7.7b), we obtain for column j :

$$t_{ij}^{(p+1)} = t_{ij}^{(p)} \cdot \left\{ b_j / \sum_{(i,j) \in \mathcal{Q}} t_{ij}^{(p)} \right\}, \quad (i,j) \in \mathcal{Q}. \quad (7.8b)$$

The quantities in braces are the usual balancing factors, and (cycling through the rows and columns) we have obtained the classical Kruithof (Furness, Sheleikovskii) method for the ordinary gravity model.

We observe that it is not necessary to work with the trip matrix elements $t_{ij}^{(p)}$; we can, rather, leave the original $t_{ij}^{(0)} \equiv t_{ij}^0$ elements unmodified, and work directly with the multipliers r_i and s_j .

Proposition 1. Define $r_i^{(0)} = 1, i \in \mathcal{I}, s_j^{(0)} = 1, j \in \mathcal{J}$, and

$$r_i^{(p+1)} = r_i^{(p)} \left\{ a_i / \sum_{(i,j) \in \mathcal{Q}} t_{ij}^{(p)} \right\}, \quad i \in \mathcal{I}, \quad (7.9a)$$

$$s_j^{(p+1)} = s_j^{(p)} \left\{ b_j / \sum_{(i,j) \in \mathcal{Q}} t_{ij}^{(p)} \right\}, \quad j \in \mathcal{J}. \quad (7.9b)$$

Then

$$t_{ij}^{(p+1)} = r_i^{(p+1)} t_{ij}^0, \quad (7.10a)$$

and

$$t_{ij}^{(p+1)} = s_j^{(p+1)} t_{ij}^0, \quad (7.10b)$$

where it is understood that (7.8a), (7.9a), and (7.10a) are used together, and (7.8b), (7.9b), and (7.10b) are used together.

Proof. The proof follows by an easy induction on p . \square

Equation (7.9) summarizes the Kruithof algorithm as it should normally be applied in the simplest case, when the t_{ij}^0 (or $\exp(-\gamma^0 c_{ij}^0)$) are given. If there are other constraints (such as a cost constraint) to be satisfied, it may be added as a constraint to be balanced in the Bregman algorithm, although

this involves solution of the (non-linear) equations (7.4) and (7.5); alternatively, another numerical method may be applied to find the root of the function $C(t) - C = 0$, using the Kruithof algorithm to evaluate the function $C(t)$ subject to the constraints on the marginal totals. In the former case many cyclic strategies are possible, as mentioned above.

7.3. A method for relaxed gravity models

At the beginning of Section 6.8 we alluded to the possibility of replacing the equality constraints (3.1) with inequality (interval) constraints as in Section 5.1. In this case, the Bregman algorithm for (7.1), (7.2), and (7.3b) may be applied, and the method reduces to a balancing procedure that is scarcely more complicated than the ordinary Kruithof procedure.

Proposition 2. Let

$$t \in \mathcal{F} \Leftrightarrow \underline{a}_i \leq \sum_j t_{ij} \leq \bar{a}_i,$$

$$\underline{b}_j \leq \sum_i t_{ij} \leq \bar{b}_j,$$

$$t_{ij} \geq 0,$$

$$t_{ij} = 0, \quad (i,j) \notin \mathcal{Q}.$$

Then the Bregman algorithm applied to (7.1), (7.2), and (7.3b) reduces to (supposing that we balance first on the upper bounds \bar{a}_i , then \underline{a}_i , and similarly for the $\bar{b}_j, \underline{b}_j$):

Initialize:

$$r_i^L \leftarrow r_i^G \leftarrow s_j^L \leftarrow s_j^G \leftarrow 1, \quad i \in \mathcal{I}, \quad j \in \mathcal{J}.$$

Balance row i :

$$\left\{ \begin{array}{l} f = \bar{a}_i / \sum_j t_{ij}^{(p)} \\ t_{ij}^{(p+1)} \leftarrow t_{ij}^{(p)} \cdot \min \{ 1/r_i^L, f \}, \quad (i,j) \in \mathcal{Q} \\ r_i^L \leftarrow \min (1, r_i^L f) \end{array} \right.$$

$$\left\{ \begin{aligned} f &= \underline{a}_i / \sum_j t_{ij}^{(\rho+1)} \\ t_{ij}^{(\rho+2)} &\leftarrow t_{ij}^{(\rho+1)} \cdot \max \{1/r_i^G f\}, \quad (i, j) \in \mathcal{Q} \\ r_i^G &\leftarrow \max (1, r_i^G f) \end{aligned} \right.$$

Balance column j :

$$\left\{ \begin{aligned} f &= \bar{b}_j / \sum_i t_{ij}^{(\rho)} \\ t_{ij}^{(\rho+1)} &\leftarrow t_{ij}^{(\rho)} \min \{1/s_j^L, f\}, \quad (i, j) \in \mathcal{Q} \\ s_j^L &\leftarrow \min (1, s_j^L f) \end{aligned} \right.$$

$$\left\{ \begin{aligned} f &= \underline{b}_j / \sum_i t_{ij}^{(\rho+1)} \\ t_{ij}^{(\rho+2)} &\leftarrow t_{ij}^{(\rho+1)} \cdot \max \{1/s_j^G, f\}, \quad (i, j) \in \mathcal{Q} \\ s_j^G &\leftarrow \max (1, s_j^G f) \end{aligned} \right.$$

Proof. Consider the case of balancing row i for a constraint of the form $\sum_j t_{ij} \leq \bar{a}_i$. When the residual $\bar{a}_i - \sum_j t_{ij}^{(\rho)}$ in the Bregman algorithm (note that we must change the sign since (7.3b) involves \geq constraints) is zero, then $f = 1$, and when it is greater (less) than zero, then f is greater (less) than 1.

The algorithm given here involves only two cases:

- (1) if $r_i^L \cdot f \leq 1$, it modifies the $t_{ij}^{(\rho)}$ values, multiplying them by f .
- (2) if $r_i^L \cdot f > 1$, it modifies them by multiplying them by $1/r_i^L$.

Note that r_i^L always remains less than or equal to 1.

The Bregman algorithm begins by checking if $f = 1$ (this implies that we are in case 1, since $r_i^L \leq 1$), or if $f > 1$ and $r_i^L = 1$ (this implies case 2), and if so, the $t_{ij}^{(\rho)}$ and r_i^L are left unchanged, as is true for the algorithm given here for the supposed values of f and r_i^L .

For the constraint $\bar{a}_i - \sum_j t_{ij} \geq 0$ we have the correspondences

$\lambda_i^{(\rho)} \leftrightarrow -\log(r_i^L)$ and $\delta \leftrightarrow \log f$ relating the notation of the Bregman algorithm and that given here. The Bregman algorithm proceeds next to calculate $f = \exp(\delta)$ explicitly, and if $f < 1$ (case 1) or $f > 1$ and $f \leq 1/r_i^L$ (also case 1) then it balances using the factor f , as does the algorithm given here. Otherwise (case 2) it balances using the factor $\exp(\lambda_i^{(\rho)}) = 1/r_i^L$, which again corresponds to the algorithm given here.

The proof for the other types of constraint is similar. \square

Again, it may or may not be desired to include the other linear constraints such as $C(t) = C$ amongst the equations to be 'balanced' by the Bregman method.

Similar algorithms can be obtained when only some of the marginal constraints are replaced by interval constraints, or when certain of the constraints are dropped completely (which may be appropriate when certain of the zones are external to the area being studied, so that we do not have detailed information about the number of arrivals or departures).

7.4. The Evans-Kirby method

In Section 5.3 the Evans-Kirby model was discussed. In that model, the zonal pairs (i, j) were grouped into K classes

$$(i, j) \text{ belongs to class } k \Leftrightarrow \underline{c}_k \leq c_{ij} \leq \bar{c}_k, \quad k = 1, \dots, K.$$

The general form of a solution to the Evans-Kirby model is

$$t \in \mathcal{F} \text{ and } t_{ij} = \begin{cases} r_i s_j v_k, & r_i > 0, s_j > 0, (i, j) \in \mathcal{L} \\ 0, & (i, j) \notin \mathcal{L} \end{cases}$$

where k is determined by the condition that $\underline{c}_k \leq c_{ij} \leq \bar{c}_k$.

We may rewrite t_{ij} as $\sum_{k=1}^K t_{ijk}$, where $t_{ijk} = r_i s_j v_k \delta_{ijk}$, $\delta_{ijk} = 1$ for one value of k for each $(i, j) \in \mathcal{L}$, and the marginal constraints become

$$\sum_j \sum_k t_{ijk} = a_i, \quad i \in \mathcal{I},$$

and

$$\sum_i \sum_k t_{ijk} = b_j, \quad j \in \mathcal{J}.$$

If we suppose (for example) that the total number of trips in each cost class k is also given, so that

$$\sum_i \sum_j t_{ij} = d_k, \quad k = 1, \dots, K.$$

then the Bregman method can be applied to produce a three-dimensional balancing procedure exactly analogous to the ordinary two-dimensional Kruithof procedure. This is the Evans-Kirby method.

7.5. Relationship with Newton's method

Returning to the ordinary gravity model, we have seen that the Kruithof algorithm produces (in the limit) values r_i and s_j such that

$$\sum_{(i,j) \in \mathcal{I}} r_i t_{ij}^0 s_j = a_i, \quad i \in \mathcal{I} = \{1, \dots, I\},$$

$$\sum_{(i,j) \in \mathcal{J}} r_i t_{ij}^0 s_j = b_j, \quad j \in \mathcal{J} = \{1, \dots, J\}.$$

Since finding such r_i and s_j is just the problem of finding the root of a set of non-linear equations in r and s , it is natural to think of applying standard numerical methods such as Newton's method to the problem. As we mentioned at the beginning of the chapter, we shall not discuss this possibility in detail. However, it will be shown here that one possible implementation of a simplified form of Newton's method reduces to the ordinary Kruithof method.

To solve the nonlinear system $\mathbf{P}(r, s) = \mathbf{0}$ given above using Newton's method one iterates using:

$$\mathbf{P}'(r^{(v)}, s^{(v)}) \left\{ \begin{pmatrix} r^{(v+1)} \\ s^{(v+1)} \end{pmatrix} - \begin{pmatrix} r^{(v)} \\ s^{(v)} \end{pmatrix} \right\} = -\mathbf{P}(r^{(v)}, s^{(v)}),$$

where the quantity in large braces is the solution to the system of linear equations at step $v + 1$. The matrix \mathbf{P}' is given by

$\sum_j t_{ij} s_j$...	0	$r_i t_{i1}$...	$r_i t_{iJ}$
.
.
0	...	$\sum_j t_{ij} s_j$	$r_i t_{i1}$...	$r_i t_{iJ}$
$t_{i1} s_1$...	$t_{i1} s_1$	$\sum_i r_i t_{i1}$...	0
.
.
$t_{iJ} s_J$...	$t_{iJ} s_J$	0	...	$\sum_i r_i t_{iJ}$

Writing

$$x_i^{(v+1)} = (r_i^{(v+1)} - r_i^{(v)}) / r_i^{(v)}, \quad i = 1, \dots, I,$$

$$y_j^{(v+1)} = (s_j^{(v+1)} - s_j^{(v)}) / s_j^{(v)}, \quad j = 1, \dots, J,$$

we obtain

$$x_i^{(v+1)} \sum_j t_{ij}^{(v)} + \sum_j y_j^{(v+1)} t_{ij}^{(v)} = -\left(\sum_j t_{ij}^{(v)} - a_i \right),$$

$$\sum_i x_i^{(v+1)} t_{ij}^{(v)} + y_j^{(v+1)} \sum_i t_{ij}^{(v)} = -\left(\sum_i t_{ij}^{(v)} - b_j \right).$$

Now, if we simplify Newton's method by iterating on rows and columns separately (which corresponds to letting $y_j^{(v+1)} = 0$ and $x_i^{(v+1)} = 0$, respectively, in the equations above), we obtain solutions

$$x_i^{(v+1)} = a_i / \sum_j t_{ij}^{(v)} - 1, \quad \text{iterating on rows,}$$

$$Y_j^{(v+1)} = b_j / \sum_i t_{ij}^{(v+1/2)} - 1, \quad \text{iterating on columns,}$$

where

$$t_{ij}^{(v+1/2)} = [1 + X_i^{(v+1/2)}] t_{ij}^{(v)}.$$

Let

$$t_{ij}^{(v+1)} = [1 + Y_j^{(v+1)}] t_{ij}^{(v+1/2)}.$$

Let $r_i^{(v+1/2)} = r_i^{(v)} [1 + X_i^{(v+1/2)}]$, so that $t_{ij}^{(v+1/2)} = \{r_i^{(v+1/2)} / r_i^{(v)}\} t_{ij}^{(v)}$, and $s_j^{(v+1)} = s_j^{(v)} [1 + Y_j^{(v+1)}]$, so that $t_{ij}^{(v+1)} = \{s_j^{(v+1)} / s_j^{(v)}\} t_{ij}^{(v+1/2)}$. Then $t_{ij}^{(v+1)} = t_{ij}^{(v)} a_i / \sum_j t_{ij}^{(v)}$, and $t_{ij}^{(v+1)} = t_{ij}^{(v+1/2)} b_j / t_{ij}^{(v+1/2)}$, which is the Kruithof balancing procedure. Hence, the balancing procedure is just a Newton procedure where the updating is restricted to considering alternately only rows or columns at one time.

Before leaving Newton's method we alert the reader to the fact that the method can be (and has been) applied to variables other than the r_i and s_j . For example, Newton's method can be applied to find the roots α and β of the gradient of the dual objective function:

$$\mathcal{D}(\alpha, \beta) = - \sum_{ij} t_{ij}^0 \exp(-\alpha_i - \beta_j) - \sum_i \alpha_i a_i - \sum_j \beta_j b_j.$$

If $\nabla_{\alpha, \beta} \mathcal{D}(\alpha, \beta) = 0$, then

$$\sum_i t_{ij}^0 \exp(-\alpha_i - \beta_j) - a_i = 0, \quad i = 1, \dots, I,$$

$$\sum_j t_{ij}^0 \exp(-\alpha_i - \beta_j) - b_j = 0, \quad j = 1, \dots, J.$$

This leads to quite a different method.

7.6. Exercises

- (1) Prove Proposition 1.
- (2) Prove Proposition 2 in the case $\sum_i t_{ij} \geq a_i$.

- (3) *Evans-Kirby model.* Write down the three-dimensional balancing equations, in the case of the Evans-Kirby model, using a given matrix δ_{ijk} , in terms of successive approximations $r_i^{(p+1)}$, $s_j^{(p+1)}$ and $t_k^{(p+1)}$ to the balancing factors, analogous to those used in Proposition 1.
- (4) Show that the Bregman algorithm for inequality constraints reduces to the Bregman algorithm for equality constraints if all of the constraints in (7.3b) occur in pairs

$$\begin{cases} \mathbf{a}_i \cdot \mathbf{x} \geq b_i, \\ -\mathbf{a}_i \cdot \mathbf{x} \geq -b_i. \end{cases}$$

- (5) Derive Bregman's method for the minimization problem defined by the objective function

$$\sum_k (x_k - x_k^0)^2$$

in place of $\sum_k x_k \{\log(x_k/x_k^0) - 1\}$.

- (6) *Preconditioning of matrices.* To illustrate that the use of balancing methods is not confined to problems involving desired marginal totals, consider the problem of preconditioning a square matrix prior to computing its eigenvalues and eigenvectors. Let the elements x_k correspond to t_{ij}^2 , where $\{t_{ij}\}$ is a square $n \times n$ matrix, and let the matrix \mathbf{A} define the constraints

$$\sum_m t_{im}^2 = \sum_m t_{mi}^2, \quad i = 1, \dots, m.$$

Show that applying Bregman's method to this problem involves multiplying the k -th row, and dividing the k -th column, of $\{t_{ij}^2\}$ by

$$d_k = \left(\sum_{i \neq k}^n t_{ik}^2 \right)^{1/2} \left(\sum_{i \neq k}^n t_{ki}^2 \right)^{-1/2},$$

thus yielding a diagonally similar matrix. (It can be shown that the conditions for convergence are that the matrix $\{t_{ij}\}$ be irreducible by permutations.)

7.7. Notes

The material in this chapter is based largely on Lamond and Stewart (1981).

Other papers dealing with the application of Newton's method to the numerical solution of the gravity model include Robillard and Stewart (1974), Erlander (1977, 1978) and Eriksson (1980, 1981). Jörnsten (1981) has applied Benders' decomposition (Benders, 1962) to models related to those discussed in the next chapter. The relationship between different versions of Newton's method applied to the gravity model is discussed by Andersson (1981).

Another application of Bregman's balancing method within the field of transportation planning is the multiproportional balancing method proposed by Murchland (1977, 1978), which has been applied in particular to the problem of estimating the origin-destination matrix given link flows (Willumsen, 1978; van Zuylen, 1978).

Outside of the field of transportation planning, balancing methods are used for preconditioning of matrices prior to the solution of other problems, such as the numerical calculation of its eigenvalues and eigenvectors (Osborne, 1960; Parlett and Reinsch, 1969; Grad, 1971), and the solution of a corresponding linear programming problem (Tomlin, 1975; Larsson, 1985). Also, in the field of algebraic picture reconstruction, balancing methods are widely used (Gordon *et al.*, 1970; Elfving, 1980). Most (but not all) of the methods turn out to be special cases of Bregman's method, as is well known to workers in the field of picture reconstruction (Elfving, 1980; Censor and Lent, 1981).

7.1. Bregman's method (Bregman, 1967a,b) may be viewed (in the case of equality constraints) as a balancing method applied to the general matrix constraints $Ax = b$ in (7.1), (7.2) and (7.3a), as a coordinate-wise descent method applied to the dual objective function, or as the Gauss-Seidel method applied to the solution of the equations defined by the Kuhn-Tucker conditions for (7.1), (7.2), and (7.3a). In the case of inequality constraints, the method uses non-orthogonal projections onto hyperplanes to find a solution satisfying the Kuhn-Tucker conditions. Bregman proved convergence of his method under quite general conditions, and explicitly discussed 'controls' other than the simple cyclic procedure $i \rightarrow (i \text{ modulo } m) + 1$ discussed in the text. A more modern and more easily accessible reference, including a convergence proof for control strategies more general than those considered by Bregman, is Censor and Lent (1981).

7.2. As mentioned in Section 2.1, the ordinary balancing method was used by Kruithof (1937) in his work on the prediction of telephone traffic. We attribute the method to him, although Bregman (1967a,b) attributes its invention to Sheleikhovskii without giving a reference. The method is often called the Furness method (Furness, 1962) in transportation planning.

The first proof of the convergence of the ordinary Kruithof method appears to have been given by Gorman (1963). Many independent proofs have been given since.

The ordinary Kruithof method has also been used in regional analysis (Wilson, 1974), in economics (Bacharach, 1970), and in statistics, in the estimation of transition probabilities of a doubly stochastic Markov chain (Sinkhorn, 1964).

7.3. The application of Bregman's method to relaxed gravity models (Dacey and Norcliffe, 1977; Jefferson and Scott, 1979; Jörnsten, 1979; Lebeuf *et al.*, 1981) is very close to the applications in algebraic picture reconstruction (Elfving, 1980; Censor and Lent, 1981, and references cited therein). Censor and Lent (1981) show that by using a not strictly cyclic control strategy, it is possible to combine the multipliers r_i^L and r_i^G (and s_j^L and s_j^G) in the application of Bregman's method to relaxed distribution problems. The application of Bregman's method to partially constrained gravity models (MacGill, 1979), applicable in the case of zones external to the area being studied, is similar to the relaxed distribution case.

7.4. An application of the Kruithof method in matrices in three or more dimensions (as in the case of the Evans-Kirby model; Evans and Kirby, 1974) also occurs in statistics, in the adjustment of contingency tables (Ireland and Kullback, 1968, and references cited therein).