MATHÉMATIQUES ET SCIENCES HUMAINES

ÉRIC TÉROUANNE

Comparaison de tendance centrale par l'analyse de transferts

Mathématiques et sciences humaines, tome 134 (1996), p. 63-76

<http://www.numdam.org/item?id=MSH_1996__134__63_0>

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1996, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (http://msh.revues.org/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Math. Inf. Sci. hum. (34e année, n°134, *1996*, pp.63-76)

COMPARAISON DE TENDANCE CENTRALE PAR L'ANALYSE DE TRANSFERTS

Éric TÉROUANNE[1]

RÉSUMÉ — *La différence de tendance centrale entre deux distributions sur un ensemble fini est représentée par une série de transferts entre les modalités. Un modèle unique est proposé qui permet d'analyser ces différences pour des variables nominales, ordinales ou métriques aussi bien que pour les variables numériques. En particulier on définit un indice de différence entre les distributions qui se ramène à l'indice de distorsion de Gini dans le cas d'une variable nominale et à la différence entre les moyennes dans le cas d'une variable numérique.*

SUMMARY — Comparison of central tendencies by means of transfer analysis.
The differences in the central tendency of two distributions on a finite set is represented by a series of transfers between modalities. A common model is presented which allows one to analyze these transfers for nominal, ordinal or metric variables, as well as for numerical ones. In particular, we define an index of difference between the distributions which boils down to Gini's distortion index in the case of a nominal variable and to the difference between the means in the case of a numerical variable.

1. INTRODUCTION

On distingue classiquement, dans la comparaison de deux distributions, ce qui concerne la tendance centrale et ce qui concerne la concentration ou dispersion. Faute d'un terme mieux adapté, et qui serait peut-être à trouver, nous utiliserons les termes de *tendance centrale* pour désigner tout ce qui ne concerne pas la dispersion-concentration. Ce faisant, on vise à dépasser la simple comparaison des résumés (mode, médiane ou moyenne) et à analyser plus finement les différences entre les deux distributions. Sans même introduire l'analyse des transferts, il y a bien des façons de le faire. Comparons par exemple deux distributions p et q d'une variable nominale à trois modalités (figure 1, à gauche).

Modalités	p	q
a	0,5	0,2
b	0,4	0,5
c	0,1	0,3

Ordres de centralité	p	q
1°	a	b
2°	b	c
3°	c	a

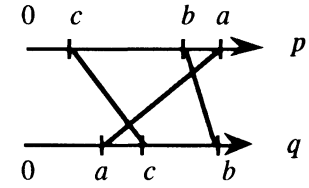

Figure 1. Divers niveaux de comparaison de deux distributions en termes de tendance centrale.

[1] Département de mathématiques et informatique appliquées, Université Paul Valéry, B.P. 5043, 34032 Montpellier Cedex. e-mail EricTerouanne@bred.univ-montp3.fr

On peut distinguer au moins trois niveaux de comparaison en termes de tendance centrale, tous trois fondés sur le même principe : dans le cas nominal, une modalité est d'autant plus centrale qu'elle est plus fréquente. Ces trois niveaux sont emboîtés : le dernier contient toute l'information du second qui lui-même contient celle du premier. On peut développer des outils de comparaison analogues dans le cas de variables non nominales.

i - La comparaison peut se réduire à celle de la modalité la plus centrale de chaque distribution, dans le cas nominal son mode. Dans notre exemple, on dira simplement que p et q n'ont pas le même mode (a pour p et b pour q).

ii - On peut, au lieu de ne s'intéresser qu'à la plus centrale, considérer le classement de toutes les modalités, de la plus centrale à la moins centrale, dans chacune des deux distributions (figure 1, au centre). On dira alors que la modalité la plus centrale pour p est la plus excentrée pour q, alors que la plus centrale pour q est en seconde position pour p. Ou encore que b et c gagnent chacune une place quand on passe de p à q. On peut même quantifier les désaccords entre les deux classements (ici deux désaccords sur un maximum de trois).

iii - On peut encore nuancer cette comparaison en situant les modalités sur l'échelle des fréquences (figure 1, à droite). On voit alors que le changement de classement de b quand on passe de p à q est dû à une plus faible différence que le changement de classement de c. Ou encore, que le changement de mode est dû à une forte différence en ce qui concerne a et à une bien plus faible en ce qui concerne b. Là aussi, divers indices permettent de quantifier ce type de différence.

L'analyse de transferts que nous développons ici est un nouvel outil de comparaison fine de deux distributions en termes de tendance centrale. Dans le cas nominal, la différence entre cet outil et le troisième niveau de comparaison décrit ci-dessus relève de la différence entre comparaison additive et multiplicative : au lieu de prendre en compte les différences entre les fréquences d'une même modalité dans les deux distributions, il prend en compte leurs rapports, autrement dit la densité d'une distribution par rapport à l'autre. C'est pourquoi au lieu de représenter ces distributions sur deux échelles parallèles (figure 1, à droite), nous les représenterons sur deux échelles perpendiculaires (figure 2).

Modalités	p	q	q/p
a	0,5	0,2	0,4
b	0,4	0,5	1,25
c	0,1	0,3	3

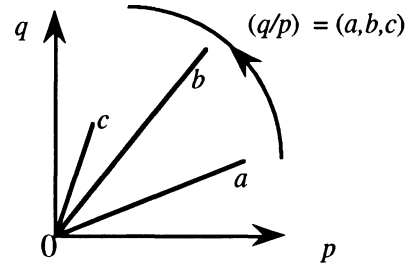

Figure 2. densité d'une distribution q par rapport à une distribution p sur $E = \{a,b,c\}$, et l'ordre (q/p) associé.

Cette façon de comparer deux distributions p et q d'une variable nominale nous a amené à introduire la notion de *distorsion* [11]. Cette grandeur est la somme des valeurs absolues des quantités $t_{x,y}(p,q) = p(x)q(y)-p(y)q(x)$. On part ici de deux remarques :

i - On peut interpréter les grandeurs $t_{x,y}(p,q)$ comme une famille de transferts entre modalités, qui change p en q (§.2).

ii - La distorsion, bien que parfois appelée indice de concentration de p par rapport à q, mesure en fait la différence entre p et q en termes de tendance centrale et non en termes de dispersion ou concentration (§.3).

L'étude de tels transferts est ensuite étendue du cas nominal au cas ordinal (§.4) puis au cas métrique (§.5), et plus généralement au cas de variables combinant ces deux structures, dont les variables numériques sont un cas particulier (§.6). Il suffit pour cela de doter les transferts d'une orientation ou d'une longueur.

On peut ainsi, dans un formalisme commun à une grande variété de structures, quantifier la différence de tendance centrale entre deux distributions et la décomposer en une somme de contributions de chacune des modalités ou de chacune des paires de modalités. L'indice global de transfert ainsi défini coïncide dans le cas nominal avec la généralisation de l'indice de Gini et dans le cas numérique avec la différence entre les moyennes. Dans le cas d'une structure purement métrique ou purement ordinale, il fournit des indices de différence qui manquaient à la panoplie des outils classiques.

Pour finir, on met en œuvre ces divers outils sur un même exemple simple mais réel (§.7), afin d'illustrer la variété et la finesse des analyses possibles.

2. TRANSFERTS CANONIQUES ET DISTORSION

Il y a toujours de nombreuses façons de passer d'une distribution p sur un ensemble fini E à une autre distribution q, au moyen d'une série de transferts entre modalités. Parmi toutes ces façons nous en distinguerons une que nous appellerons *transferts canoniques* changeant p en q. Le transfert canonique de la modalité x vers la modalité y pour changer p en q est le déterminant :

$$t_{x,y}(p,q)=\begin{vmatrix} p(x) & p(y) \\ q(x) & q(y) \end{vmatrix} = p(x)q(y) - p(y)q(x)$$

Si $t_{x,y}(p,q)$ est positif, il y a un transfert positif de x vers y. S'il est négatif, le transfert positif a lieu de y vers x. On a bien sûr :

$$t_{x,y}(p,q)= - t_{y,x}(p,q) = t_{y,x}(q,p)= - t_{x,y}(q,p)$$

On vérifie immédiatement que cette famille de transferts transforme bien p en q. On a en effet pour tout x dans E :

$$\sum_y t_{x,y}(p,q) = \sum_y p(x)\, q(y) - \sum_y p(y)\, q(x) = p(x) - q(x)$$

Dans le cas d'une variable dichotomique, avec $E = \{x,y\}$, le transfert $t_{x,y}(p,q)$ se réduit à $p(x)-q(x) = q(y)-p(y)$.

On appelle *ordre de distorsion* de q par rapport à p le préordre, noté (q/p), induit sur E par la densité de q par rapport à p (figure 2). Le couple (x,y) appartient à l'ordre (q/p), ce que l'on note $(x,y)(q/p)$, si :

$$p(x) = p(y) = 0 \quad \text{ou} \quad \frac{q(x)}{p(x)} \leq \frac{q(y)}{p(y)}$$

Ce qui revient à dire que, dans le passage de p à q, x est relativement moins favorisé (ou plus défavorisé) que y. Cela s'exprime encore :

$$(x,y)(q/p) \Leftrightarrow t_{x,y}(p,q) \geq 0$$

Autrement dit, tous les transferts canoniques positifs se font d'une modalité inférieure vers une modalité supérieure pour l'ordre de distorsion (q/p).

La *distorsion* entre p et q, définie comme la superficie de la *surface de distorsion* (figure 3) entre p et q, est la somme des transferts canoniques positifs [11]. Pour des raisons d'homogénéité avec les autres structures envisagées dans cet article, nous appellerons également cette grandeur le *transfert nominal* et la noterons :

$$t(\{p,q\}) = \sum_{(x,y)(q/p)} t_{x,y}(p,q)$$

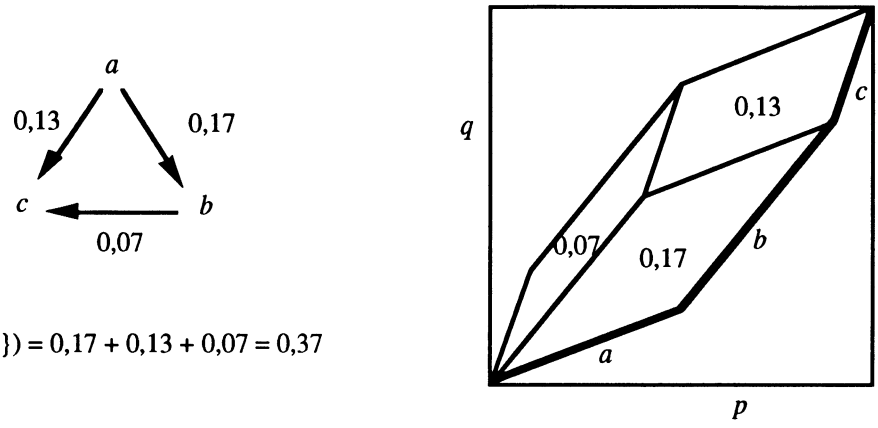

Figure 3. Transferts canoniques entre modalités et transfert nominal entre les distributions de la figure1.

Cet indice est toujours compris entre 0 et 1. Il est nul si et seulement si $p = q$. Il prend la valeur 1 si et seulement si les supports de p et q sont disjoints. C'est un écart sur l'espace des distributions, mais pas une distance car il peut violer l'inégalité triangulaire[2]. En posant :

$$P_{(q/p)}(x) = \sum_{(x,y)(q/p)} p(y) \quad \text{et} \quad Q_{(q/p)}(x) = \sum_{(x,y)(q/p)} q(y)$$

on a aussi :

$$t(\{p,q\}) = \sum_{x} p(x)Q_{(q/p)}(x) \; - q(x)P_{(q/p)}(x)$$

Le transfert nominal se décompose en somme des contributions de chacune des modalités ([11]). La contribution de la modalité x est la quantité :

$$t_x(\{p,q\}) = \frac{1}{2}\sum_{y} |t_{x,y}(p,q)| = \frac{1}{2}\sum_{y} (q/p)(x,y)t_{x,y}(p,q)$$

où $(q/p)(x,y)$ vaut 1 si $(x,y)(q/p)$ et -1 si $(y,x)(q/p)$.

---

[2] Par exemple entre les distributions (1/3,1/3,1/3), (1/2,1/2,0) et (0,1/2,1/2).

3. TRANSFERTS, CONCENTRATION ET TENDANCE CENTRALE

L'idée de comparer deux distributions statistiques en analysant les transferts qu'il faut effectuer entre les modalités pour passer de la première distribution à la seconde remonte au moins aux travaux de Dalton [3], formalisant une idée de Pigou [9] pour comparer des distributions de fortune ou de revenus. Les différences entre distributions étudiées dans ce cadre sont des différences de concentration : une distribution est plus concentrée qu'une autre si l'on peut passer de la première à la seconde au moyen de transferts qui tous prennent à un plus riche pour donner à un plus pauvre.

Au demeurant, peu importe de savoir si c'est la même personne ou non qui demeure la plus riche : cette notion de concentration est insensible aux permutations sur les modalités. Dans ces études, on oublie en quelque sorte les modalités, c'est à dire les personnes entre lesquelles sont distribués les biens, pour ne s'intéresser qu'aux fréquences, c'est à dire les fortunes. L'ordre partiel ainsi défini sur l'espace des distributions est décrit par la courbe de Lorenz [5]. Il a été abondamment étudié dans le cadre d'une théorie des inégalités que Marshall et Olkin présentent de façon détaillée [6]. Pour des développements et une bibliographie plus récents, le lecteur peut se reporter à [7] ou [8].

Le contour inférieur de notre surface de distorsion (trait gras dans la figure 3) se confond avec la courbe de Lorenz de la distribution q quand p est la distribution uniforme. C'est ce qui lui vaut habituellement le nom de *courbe de concentration* ([1],[2]) ou *courbe de concentration relative* ([12]) de q par rapport à p. Chez ces auteurs, ce que nous appelons transfert nominal, et qui généralise l'indice de Gini [4], est appelé *indice de concentration de q par rapport à p*. Regazzini [10] donne également une interprétation de cette grandeur en termes de transferts (mais seulement entre modalités successives dans l'ordre de distorsion) et la rebaptise *mean equalizing transfer changing p into q*.

On a justifié ([11]) le remplacement du concept de "concentration de q par rapport à p" par celui de "distorsion (ou tansfert nominal) entre p et q" par le fait que les mesures p et q y jouent des rôles symétriques. Il convient d'ajouter que le terme de concentration prête ici à confusion, car il s'agit bien de différences de tendance centrale. Comparons par exemple la distribution p de la figure 1 avec les cinq autres distributions de la même variable nominale obtenues à partir de p par simples permutations des fréquences (figure 4).

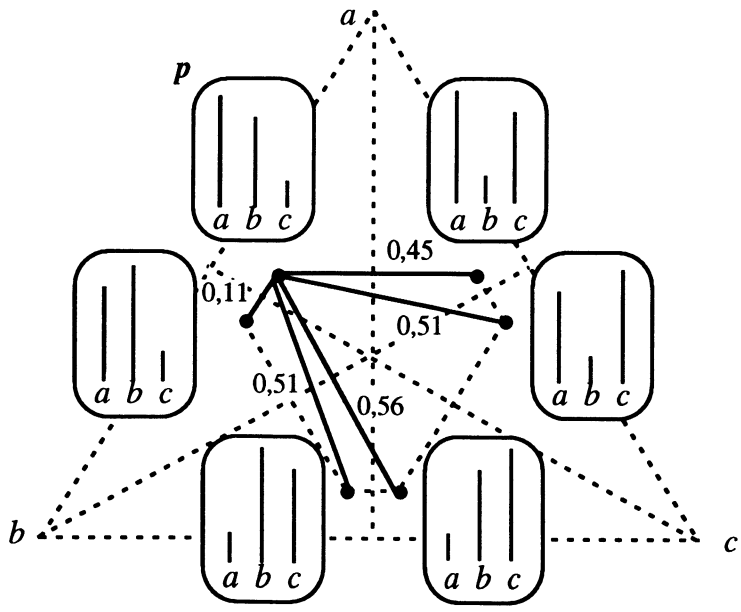

Figure 4. Transferts nominaux entre une distribution p et les cinq distributions de même concentration qu'elle.

Ces six distributions ont exactement la même concentration. Par contre elles diffèrent en termes de tendance centrale, la plus différente de p étant celle qui inverse l'ordre des fréquences : son mode est la modalité la plus rare de p et vice-versa. C'est bien cette différence de tendance centrale que mesure le transfert nominal.

4. TRANSFERT ORIENTÉ

On considère maintenant le cas où E est doté a priori d'un ordre total ω. Par convention, nous dirons que x est *à gauche* de y, ou y *à droite* de x, et nous écrirons $(x,y)\omega$, si le couple (x,y) appartient à ω. Si les ordres ω et (q/p) coïncident sur la paire $\{x,y\}$, c'est à dire si x est à gauche de y et que $t_{x,y}(p,q)$ est positif, ou si x est à droite de y et que $t_{x,y}(p,q)$ est négatif, le transfert canonique entre x et y se fait de gauche à droite. Il se fait de droite à gauche sinon. Examinons notre exemple des figures 1-2 avec pour ordre a priori $c < a < b$. Le transfert positif entre a et b se fait alors de gauche à droite tandis que les deux autres (de a vers c et de b vers c) se font de droite à gauche (figure 5, à gauche).

On appellera *transfert orienté de p à q pour l'ordre* ω, et on notera $t_\omega(p,q)$, la somme de tous les transferts canoniques de gauche à droite :

$$t_\omega(p,q) = \sum_{(x,y)\omega} t_{x,y}(p,q) = - t_\omega(q,p) = t_{\omega*}(q,p) = - t_{\omega*}(p,q)$$

où ω^* désigne l'ordre dual de ω. Cet indice varie entre $t(\{p,q\})$, quand ω et (q/p) coïncident, et $-t(\{p,q\})$, quand ω est le dual (p/q) de (q/p). En posant :

$$P_\omega(x) = \sum_{(x,y)\omega} p(y) \quad et \quad Q_\omega(x) = \sum_{(x,y)\omega} q(y)$$

On a :

$$t_\omega(p,q) = \sum_x p(x)Q_\omega(x) - q(x)P_\omega(x)$$

Si $t_\omega(p,q)$ est positif, on dira qu'il y a un *transfert vers la droite* positif quand on passe de p à q. S'il est négatif, $-t_\omega(p,q)$ représente un *transfert vers la gauche* positif.

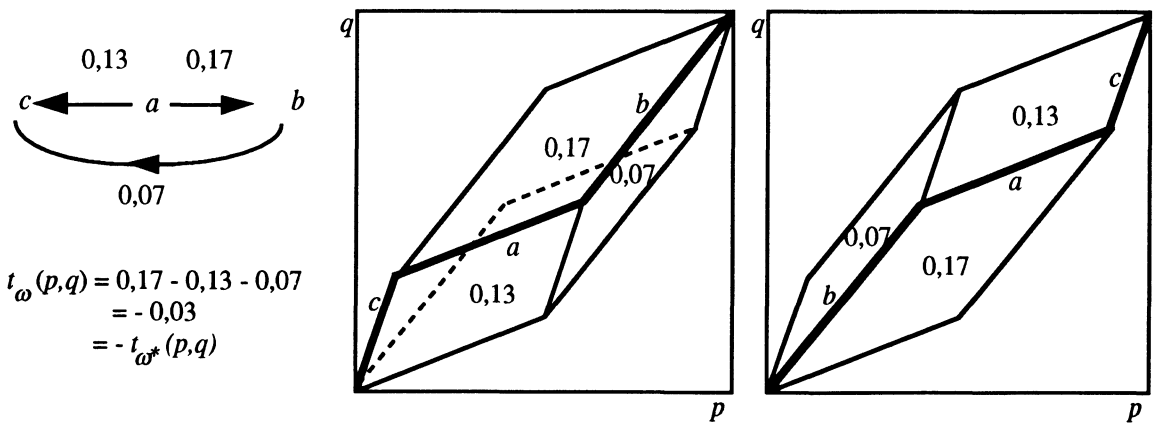

Figure 5. Transfert orienté de p à q pour l'ordre $\omega = (c,a,b)$ (à gauche) et pour l'ordre dual ω^* (à droite).

Graphiquement (figure 5, au centre), on peut représenter l'ordre a priori ω par une ligne brisée (trait gras) où les segments représentant les modalités se suivent dans l'ordre ω. Si l'on reconstitue le puzzle de la surface de distorsion en respectant cette ligne, celle-ci sépare celle-là en deux parties. La surface du dessus est constituée des parallélogrammes associés aux paires de modalités sur lesquelles ω et (q/p) coïncident, et celle du dessous de ceux pour lesquels ω et (q/p) s'opposent. Le transfert orienté de p à q est donc la somme algébrique de ces surfaces, comptées positivement au dessus et négativement au dessous de la ligne représentant ω. Avec les mêmes conventions, la figure 3 et la figure 5 (à droite) représentent les transferts orientés de p à q pour les ordres (q/p) et ω^* respectivement.

On peut montrer également que le transfert orienté de p à q pour ω est la mesure algébrique (avec la même convention de signe que ci-dessus) de la superficie comprise entre la ligne qui représente ω et sa symétrique par rapport au centre du carré (en pointillés sur la figure 5, au centre), qui représente l'ordre dual ω^*.

La contribution de la modalité x au transfert orienté entre p et q est la quantité :

$$t_{\omega,x}(\{p,q\}) = \frac{1}{2}\sum_{y} \omega(x,y)\ t_{x,y}(p,q)$$

où $\omega(x,y)$ vaut 1 si $(x,y)\omega$, -1 si $(y,x)\omega$.

Le transfert orienté mesure le bilan global, soit vers la droite soit vers la gauche, des transferts entre modalités quand on passe de p à q. Une autre grandeur intéressante dans la comparaison de deux distributions d'une variable ordinale est l'*homogénéité* de ces transferts. Quelle que soit la valeur absolue du transfert orienté, il est intéressant de savoir s'il résulte de l'addition de transferts entre modalités qui vont tous dans le même sens (*transferts homogènes*), ou bien si une grande partie d'entre eux, de signes opposés, se compensent.

On peut évaluer l'*homogénéité* des transferts en comparant la valeur absolue du transfert orienté au transfert nominal. On notera $h_\omega(p,q)$ cet *indice d'homogénéité* :

$$h_\omega(p,q) = \frac{|t_\omega(p,q)|}{t(\{p,q\})} = \frac{\left|\sum_{(x,y)\omega} t_{x,y}(p,q)\right|}{\sum_{(x,y)\omega} |t_{x,y}(p,q)|} = h_\omega(q,p) = h_{\omega*}(p,q) = h_{\omega*}(q,p)$$

L'homogénéité est comprise entre 0 et 1. Elle est de 100% pour l'ordre (q/p) ou son dual. Dans le cas de l'ordre ω de notre exemple (figure 5), elle vaut seulement 0,03 / 0,37, soit environ 8% : les transferts vers la gauche et vers la droite se compensent presque, avec un léger avantage aux premiers. Dans le cas de l'ordre $a < c < b$, on obtient un transfert orienté positif (0,13 + 0,17 - 0,07 = 0,23) et un indice d'homogénéité de 0,23 / 0,37 = 62%, indiquant que les transferts vers la droite l'emportent assez largement sur les transferts vers la gauche.

5. TRANSFERT MÉTRIQUE

Considérons maintenant le cas où E est muni a priori d'une structure métrique. Le transfert canonique entre deux modalités x et y se fait alors sur une certaine distance $d(x,y)$. On appellera *transfert métrique de x vers y* le produit $d(x,y)t_{x,y}(p,q)$. La somme des transferts métriques positifs, notée $t_d(\{p,q\})$, constitue le *transfert métrique entre p et q* :

$$t_d(\{p,q\}) = \sum_{(x,y)(q/p)} d(x,y)\ t_{x,y}(p,q)$$

Si l'on interprète $t(\{p,q\})$ comme la proportion d'"individus" qui se sont déplacés lors des transferts canoniques changeant p en q, la proportion d'individus qui n'ont pas bougé est $1\text{-}t(\{p,q\})$ et $t_d(\{p,q\})$ est la distance moyenne parcourue par tous les individus. Le transfert nominal $t(\{p,q\})$ est simplement le cas particulier où la distance d est uniformément égale à 1 hors de la diagonale.

6. TRANSFERT MÉTRIQUE ORIENTÉ

Si E est muni à la fois d'une métrique d et d'un ordre total ω, on définit le *transfert métrique orienté de p à q* par :

$$t_{\omega,d}(p,q) = \sum_{(x,y)\omega} d(x,y)t_{x,y}(p,q)$$

L'indice d'homogénéité est alors défini comme :

$$h_{\omega,d}(p,q) = \frac{|t_{\omega,d}(p,q)|}{t_d(\{p,q\})}$$

Le transfert orienté $t_\omega(\{p,q\})$ est le cas particulier où la distance d est uniformément égale à 1 hors de la diagonale.

Un autre cas particulier est celui où la métrique peut être définie au moyen d'une fonction croissante f de E dans R telle que, pour tout couple (x,y) :

$$d(x,y) = |f(x)\text{-}f(y)|$$

On a alors :

$$t_{\omega,d}(p,q) = \sum_{(x,y)\omega} (f(y)\text{-}f(x))(p(x)q(y) - p(y)q(x))$$

Or cette quantité est aussi égale à :

$$\sum_{(y,x)\omega} (f(y)\text{-}f(x))(p(x)q(y) - p(y)q(x))$$

On a donc :

$$\begin{aligned} t_{\omega,d}(p,q) &= \frac{1}{2}\sum_{x,y} (f(y)\text{-}f(x))(p(x)q(y) - p(y)q(x)) \\ &= \sum_{x,y} f(y)\,(p(x)q(y) - p(y)q(x)) \\ &= \sum_{y} f(y)q(y) - \sum_{y} f(y)p(y) \end{aligned}$$

Le transfert métrique orienté de p à q est donc la différence entre les moyennes, pour q et p respectivement, de la fonction f.

Un exemple d'une telle situation est une variable ordinale à laquelle on adjoint la métrique des rangs. Le transfert métrique orienté de p à q se confond alors avec la différence entre les

rangs moyens pour q et pour p. On a ainsi le choix, pour caractériser la différence de tendance centrale entre deux distributions d'une variable ordinale, entre le transfert métrique orienté calculé avec la métrique des rangs et le transfert orienté, c.à.d. le transfert métrique orienté calculé avec la métrique uniforme. Ce choix est exactement celui que l'on est amené à faire quand, pour caractériser la liaison entre deux variables ordinales, on choisit entre le Rho de Spearman (métrique des rangs) et le Tau de Kendall (métrique uniforme).

Un autre exemple est bien sûr celui d'une variable numérique : f est alors l'identité et le transfert métrique orienté $t_{\omega,d}(p,q)$ est simplement la différence $\bar{q} - \bar{p}$ entre les moyennes $\bar{q}$ et $\bar{p}$ des deux distributions. La contribution d'une modalité x à ce transfert s'écrit :

$$t_{\omega,d,x}(p,q) = \frac{1}{2}\begin{vmatrix} p(x) & x-\bar{p} \\ q(x) & x-\bar{q} \end{vmatrix}$$

7. UN EXEMPLE

Le tableau 1 donne la population de chacune des cinq villes les plus importantes de l'Hérault lors des recensements de 1881, 1968 et 1990 en nombre d'habitants (trois premières lignes) puis en pourcentages (trois lignes suivantes). Chacune des trois dernières lignes donne la densité de l'un des recensements par rapport à un plus ancien. Ainsi pour 1968/1881, la densité pour la ville v est le rapport $p_{68}(v)/p_{81}(v)$ où $p_a(v)$ est la proportion de la ville v dans le total des populations des cinq villes pour l'année a.

	Agde	Béziers	Lunel	Montpellier	Sète	Total
1881	8170	42915	6487	56005	35517	149094
1968	10348	82260	10922	167211	41044	311785
1990	17784	72362	18501	210866	41916	361429
1881	5,48%	28,78%	4,35%	37,56%	23,82%	100%
1968	3,32%	26,38%	3,50%	53,63%	13,16%	100%
1990	4,92%	20,02%	5,12%	58,34%	11,60%	100%
1968 / 1881	0,606	0,917	0,805	1,428	0,553	
1990 / 1968	1,483	0,759	1,461	1,088	0,881	
1990 / 1881	0,898	0,696	1,176	1,553	0,487	

Tableau 1. Populations de cinq villes de l'Hérault lors de trois recensements, distributions de fréquences correspondantes et densités de ces distributions les unes par rapport aux autres.

Nous allons comparer ces trois distributions en considérant d'abord que l'ensemble des villes est dépourvu de structure a priori, puis en le dotant d'une structure métrique, enfin en y adjoignant un ordre. Pour faire court, les villes pourront dans la suite être désignées par leurs initiales.

7.1. Transfert nominal

L'ordre de distorsion de 1968 par rapport à 1881 (7° ligne du tableau 1) est (S, A, L, B, M). Les plus grandes villes dominent (Montpellier) ou stagnent (Béziers), au détriment de la troisième (Sète) et, dans une moindre mesure, des deux plus petites (Lunel et Agde). Le tableau 2 indique les valeurs des transferts canoniques $t_{x,y}(p,q)$.

	Sète	Agde	Lunel	Béziers	Montpellier
Sète	0	0,0007	0,0026	0,0250	0,0783
Agde	- 0,0007	0	0,0005	0,0049	0,0169
Lunel	- 0,0026	- 0,0005	0	0,0014	0,0102
Béziers	- 0,0250	- 0,0049	- 0,0014	0	0,0553
Montpellier	- 0,0783	- 0,0169	- 0,0102	- 0,0553	0

Tableau 2. Transferts canoniques d'une ville (ligne) à une autre (colonne) lors du passage de la distribution de 1881 à celle de 1968.

Le transfert nominal est la somme des transferts positifs (triangle supérieur), soit 0,1957. On peut interprétéer ce nombre en disant que les transferts canoniques entre 1881 et 1968 ont concerné 19,57% de la masse totale. Il faut bien sûr prendre garde à ne pas interpréter ces transferts comme de véritables transferts de population au sens démographique du terme : non seulement les populations à ces deux dates sont quasi disjointes, mais en plus l'évolution entre deux recensements combine de tels transferts avec un solde naturel et des migrations depuis ou vers le reste du monde.

Pour interpréter ces transferts, il faut d'abord admettre qu'ils s'exercent au sein d'une population fictive de masse totale constante et égale à 1. Il faut ensuite admettre que ce ne sont pas les transferts les plus "économiques" : dans l'exemple ci-dessus, on peut passer de la distribution de 1881 à celle de 1968 en opérant uniquement des transferts depuis les quatre autres villes vers Montpellier, déplaçant alors seulement 16,07% de la masse totale (somme des différences positives entre les lignes 4 et 5 du tableau 1). Les transferts canoniques, eux, rendent compte, pour chaque paire de villes séparément, de la modification du rapport entre leurs masses : le transfert de x vers y est positif si le rapport entre les populations de x et y a diminué, négatif s'il a augmenté, nul s'il est resté le même.

On peut représenter ces transferts par un graphe (figure 6) dans lequel les villes sont placées de gauche à droite selon l'ordre de distorsion. Les transferts de 1881 à 1968 se font en grande majorité vers Montpellier (82% du total). Le transfert de Béziers vers Montpellier (28% du total) fait plus que compenser les gains de Béziers (16%). Sète assume à elle seule plus de la moitié des pertes.

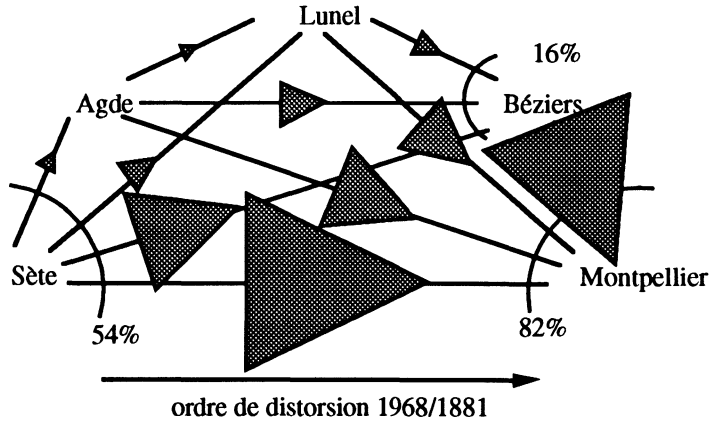

Figure 6. Graphe des transferts canoniques de 1881 à 1968.
Les pourcentages concernent des flèches issues de ou allant vers une même ville (arcs de cercle).

On applique le même traitement au passage de 1968 à 1990 (tableau 3 et figure 7). L'ordre de distorsion est alors (*B*, *S*, *M*, *L*, *A*). Le transfert nominal s'élève à 0,0976, soit la moitié du transfert nominal observé de 1881 à 1968. Les deux plus petites villes (Lunel et Agde) sont devenues bénéficiaires, avec des comportements très similaires : elles reçoivent chacune 16% des transferts sans pratiquement rien céder. Montpellier reste le plus gros gagnant, recevant 63% des transferts et en cédant par ailleurs 14%. Enfin c'est Béziers et non plus Sète qui est le plus gros perdant, avec 65% des transferts négatifs.

	Béziers	**Sète**	**Montpellier**	**Lunel**	**Agde**
Béziers	0	0,0042	0,0466	0,0065	0,0063
Sète	- 0,0042	0	0,0146	0,0027	0,0026
Montpellier	- 0,0466	- 0,0146	0	0,0070	0,0070
Lunel	- 0,0065	- 0,0027	- 0,0070	0	0,00005
Agde	- 0,0063	- 0,0026	- 0,0070	- 0,00005	0

Tableau 3. Transferts canoniques entre les distributions de 1968 et 1990.

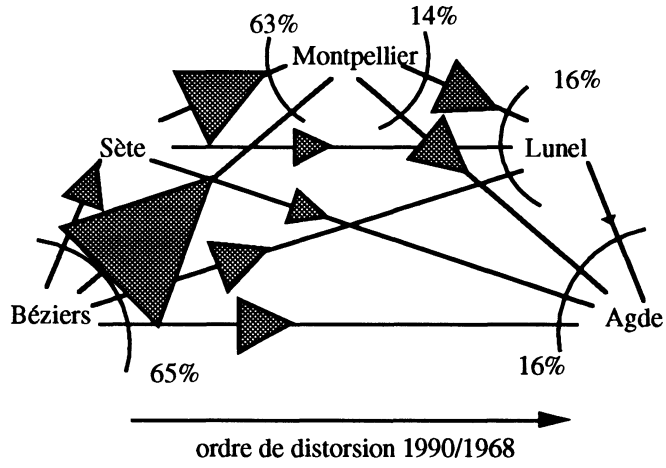

Figure 7.Graphe des transferts canoniques de 1968 à 1990.

L'ordre de distorsion de 1881 à 1990 (tableau 4) est encore différent. Le transfert nominal s'y élève à 0,2445, soit un peu moins que la somme de ses valeurs pour les deux périodes successives. Le graphe des transferts (figure 8) illustre l'écrasement par Montpellier des autres villes, à l'exception de Lunel dont le bilan est positif.

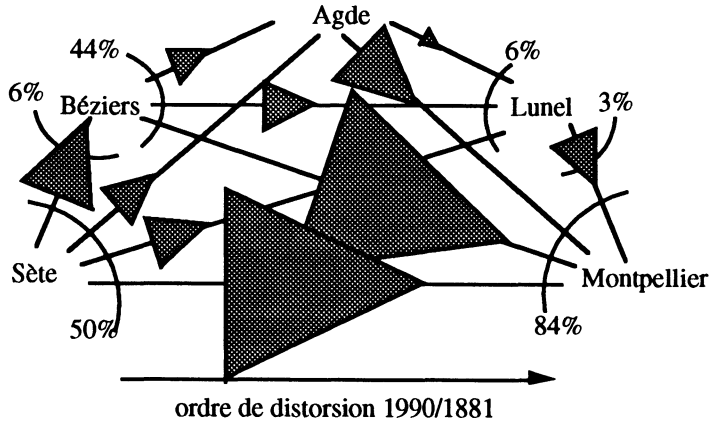

Figure 8. Graphe des transferts canoniques de 1881 à 1990.

	Sète	Béziers	Agde	Lunel	Montpellier
Sète	0	0,0143	0,0054	0,0071	0,0954
Béziers	- 0,0143	0	0,0032	0,0060	0,0927
Agde	- 0,0054	- 0,0032	0	0,0007	0,0135
Lunel	- 0,0071	- 0,0060	- 0,0007	0	0,0062
Montpellier	- 0,0954	- 0,0927	- 0,0135	- 0,0062	0

Tableau 4. Transferts canoniques entre les distributions de 1881 et 1990.

7.2. Transfert métrique

Introduisons maintenant une structure métrique (figure 9), celle des distances kilométriques entre les villes. On a considéré les distances par la route et on n'a pas tenu compte, pour simplifier, de la construction entre-temps d'une autoroute. Les distances non indiquées sont celles des plus courts chemins dans le graphe de la figure 9 (par exemple 96km entre Béziers et Lunel).

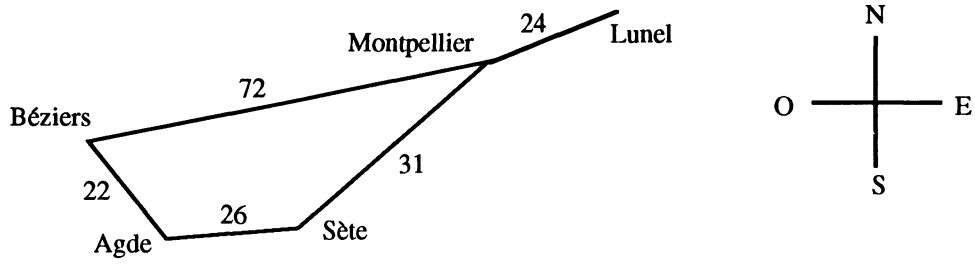

Figure 9. Disposition géographique des cinq villes
avec leurs distances routières (en km).

Les transferts métriques, calculés par exemple entre 1881 et 1990 (tableau 5), se font par définition dans le même sens que les transferts canoniques (tableau 4). Par contre ils s'expriment dans une unité, ici le kilomètre. Le transfert métrique entre ces deux distributions s'élève à 12,47km. Le même indice vaut respectivement 9,12km entre 1881 et 1968 et 5,56km entre 1968 et 1990.

	Sète	Béziers	Agde	Lunel	Montpellier
Sète	0	0,6870	0,1395	0,3931	2,9580
Béziers	- 0,6870	0	0,0702	0,5782	6,6762
Agde	- 0,1395	- 0,0702	0	0,0538	0,7688
Lunel	- 0,3931	- 0,5782	- 0,0538	0	0,1477
Montpellier	- 2,9580	- 6,6762	- 0,7688	- 0,1477	0

Tableau 5. Transferts métriques entre les distributions
de 1881 et 1990 (en km).

La différence entre l'analyse du transfert nominal et celle du transert métrique réside essentiellement dans les valeurs relatives des transferts. Ainsi les pertes de Sète, la ville la plus centrale géographiquement, chutent, passant de 50% du transfert nominal à 33% du transfert métrique. Par contre les soldes des villes les plus excentrées voient leurs valeurs relatives s'aggraver, passant respectivement de +3% du transfert nominal à +7% du transfert métrique pour Lunel et de -38% à -53% pour Béziers.

7.3. Transfert orienté

On peut également introduire des ordres totaux sur E. En effet la comparaison de ces trois recensements fait apparaître une dynamique, surtout sensible dans la deuxième période, qui favorise l'est du département au détriment de l'ouest. On peut représenter la disposition des villes d'ouest en est (figure 10) par un ordre total $\omega = (B, A, S, M, L)$. Le transfert orienté de 1968 à 1990 est la somme des transferts canoniques du tableau 3 qui vont d'une ville vers une autre située plus à l'est (autrement dit à l'orient !).

La plupart de ces transferts sont positifs, les seuls négatifs étant ceux de Montpellier, Sète et Lunel vers Agde (en gris sur la figure 10). Il en résulte un transfert orienté $t_\omega(1968,1990) = 0{,}0880$. Son signe positif confirme que les transferts vers l'est l'emportent sur ceux vers l'ouest. La valeur de l'indice d'homogénéité, $h_\omega(1968,1990) = 0{,}0880 / 0{,}0976$, soit 90%, confirme la force de cette tendance.

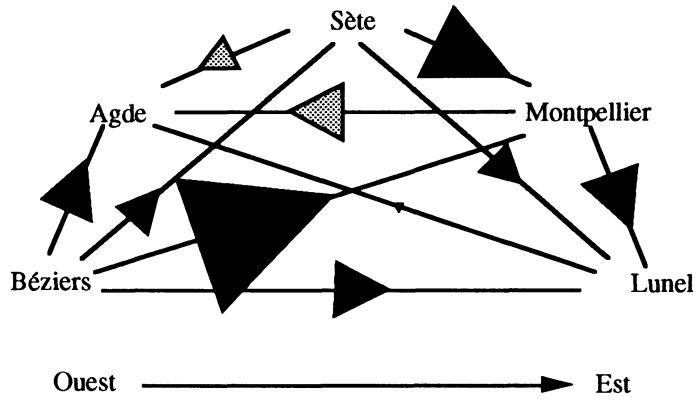

Figure 10. Transferts d'ouest en est positifs (noir) ou négatifs (gris) entre les distributions de 1968 et 1990.

Le transfert orienté est également positif de 1881 à 1968 : les seuls transferts négatifs (ou d'est en ouest) sont ceux de Sète vers Béziers et Agde et de cette dernière vers Béziers. On trouve $t_\omega(1881,1968) = 0{,}1651$ et $h_\omega(1881,1968) = 0{,}1651 / 0{,}1957 = 84\%$. Cette dernière valeur confirme que la dynamique d'ouest en est est moins forte dans cette première période que dans la seconde. On trouve enfin $t_\omega(1881,1990) = 0{,}2186$ et $h_\omega(1881,1990) = 89\%$.

On peut évidemment faire les mêmes calculs en prenant en compte la structure métrique. Les transfert métriques orientés correspondants sont 5,85 km de 1881 à 1968, 4,62 km de 1968 à 1990 et 10,52 km de 1881 à 1990. Les indices d'homogénéité sont alors respectivement de 83% entre 1968 et 1990, de 63% entre 1881 et 1968 et de 84% entre 1881 et 1990.

On peut étudier de la même façon une autre dynamique, surtout sensible dans la première période, qui favorise les grandes villes au détriment des petites. On représente alors la taille des villes par l'ordre total (A, L, S, B, M) (malgré que Lunel soit plus petit qu'Agde au premier recensement). On trouve là encore des transfert orientés positifs et pour indice d'homogénéité (dans le cas métrique) 97% de 1881 à 1968, 37% de 1968 à 1990 et 81% de 1881 à 1990.

7.4. Transfert numérique

On peut enfin utiliser une variable de type numérique pour analyse par exemple la dynamique d'ouest en est : il suffit pour cela de représenter chaque ville par sa longitude. Le transfert métrique orienté d'une distribution à l'autre sera alors la différence entre leurs longitudes moyennes. On obtiendrait un autre plongement de ce type, et des résultats

légèrement différents, en remplaçant la distance routière entre les villes par leur distance le long de la voie de chemin de fer qui les relie effectivement dans cet ordre.

Avec ce type de structure, et avec lui seulement, le transfert métrique orienté de 1881 à 1990 est la somme des transferts métriques orientés des deux périodes 1881-1968 et 1968-1990.

BIBLIOGRAPHIE

[1] BARBUT M., "Sur quelques propriétés élémentaires des fonctions de concentration de C. Gini", *Mathématiques, Informatique et Sciences humaines*, 88 (1984), 5-19.
[2] CIFARELLI D.M. et REGAZZINI E., "On a general definition of concentration function", *Sankhya*, 49 (B-3) (1987), 307-319.
[3] DALTON H., "The measurement of the inequality of incomes", *Econ. J.*, 30 (1920), 348-361.
[4] GINI C., "Sulla misura della concentrazione della variabilità dei caratteri", *Atti del Reale Istituto Veneto di Sci., Litt., Arti*, AA., 1913-1914, 73 (II) (1914), 1903-1942.
[5] LORENZ M.O., "Methods of measuring the concentration of wealth", *Publ. Amer. Statist. Assoc.*, 9 (1905), 209-219.
[6] MARSHALL A.W. et OLKIN J., *Inequalities : theory of majorization and its applications,* New York, Academic Press, 1979.
[7] MOSLER K., "Majorization in economic disparity measures", *Linear Algebra and its applications*, 199 (1994), 91-114.
[8] MOSLER K. et SCARSINI M., *Stochastic orders and decision under risk*, Hayward, California, IMS Lecture Notes - Monograph series (19), 1991.
[9] PIGOU A., *Wealth and Welfare*, New York, Macmillan, 1912.
[10] REGAZZINI E., "Concentration comparisons between probability measures", *Sankhya*, 54 (B) (1992), 129-149.
[11] TÉROUANNE E., "Distorsion entre deux distributions d'une variable nominale", *Mathématiques, Informatique et Sciences humaines*, 131 (1995), 29-38.
[12] YITZHAKI S. et OLKIN I., "Concentration indices and concentration curves", in *Stochastic ordrers and decision under risk*, K. Mosler et M. Scarsini Eds., IMS Lecture Notes - Monograph series (19) (1991), 380-392.