



HAL
open science

A propos des problèmes causés par les indices de mesure d'inégalité de Gini et de Kakwani

Louis de Mesnard

► **To cite this version:**

Louis de Mesnard. A propos des problèmes causés par les indices de mesure d'inégalité de Gini et de Kakwani. [Rapport de recherche] Laboratoire d'analyse et de techniques économiques(LATEC). 1997, 16 p., Figure, ref. bib.: 26 ref. hal-01527267

HAL Id: hal-01527267

<https://hal.science/hal-01527267>

Submitted on 24 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LATEC

LABORATOIRE D'ANALYSE ET DE TECHNIQUES ÉCONOMIQUES

U.R.A. 342 C.N.R.S.

DOCUMENT de TRAVAIL



UNIVERSITE DE BOURGOGNE

FACULTE DE SCIENCE ECONOMIQUE ET DE GESTION

4, boulevard Gabriel - 21000 DIJON - Tél. 80 39 54 30 - Fax 80 39 56 48

ISSN : 0292-2002

n° 9713

**A propos des problèmes causés par les indices
de mesure d'inégalité de Gini et de Kakwani**

Louis de MESNARD*

juin 1997

Professeur - LATEC (UMR 5601 CNRS)

A propos des problèmes causés par les indices de mesure d'inégalité de Gini et de Kakwani

Louis de Mesnard

LATEC (UMR CNRS 5601), Université de Bourgogne

Résumé

On fait le point sur les faiblesses des indices de Gini et de Kakwani lorsque la dominance entre distributions n'est pas garantie : à un même Gini peuvent correspondre plusieurs distributions différentes et à une même valeur du Kakwani peuvent correspondre une infinité de couples d'indices de Gini. On propose alors d'approximer la courbe de Lorenz typiquement en deux ou trois parties linéaires, puis de décrire la position de son coude ou de ses deux coudes par deux ou quatre nombres qui suffisent caractériser complètement sa forme et la catégorie à laquelle elle appartient. Une procédure de choix entre distributions est proposée, basée sur une application du maximin : entre deux distributions à un coude, celle dont la courbe de concentration est la plus tordue à droite est déclarée la moins inégalitaire.

I. INTRODUCTION

A. L'indice de Gini

L'inégalité est un concept central de l'économie sociale qui ne peut être apprécié que par une mesure pertinente. Beaucoup d'auteurs utilisent l'indice de Gini comme une mesure d'inégalité¹. L'objectif de ce papier est d'exposer les problèmes causés par les indices d'inégalité (l'indice de Gini et ses dérivés comme l'indice de Kakwani) et de déterminer les manières de les résoudre quand la dominance stochastique entre distributions n'est pas garantie. Il y a beaucoup de formules équivalentes pour calculer l'indice de Gini². Pour une variable discrète Berrebi et Silber (1987) exposent une formule très simple où les x_i sont triés en ordre décroissant : $I_G = 1 - \frac{1}{N^2 \bar{x}} \sum_{i=1}^N (2i-1)x_i$. Cette formule est facile à comprendre. Considérons la figure 1 dans laquelle les N individus i sont représentés sur l'axe des x , la variable cumulée est sur l'axe des y (x peut être le revenu gagné par chaque individu), triée par ordre décroissant et considérons la courbe de Lorenz obtenue en liant chaque individu au revenu cumulé de tous les individus placés avant lui plus son propre revenu. L'aire du rectangle est $N^2 \bar{x}$. Le côté horizontal du rectangle est N , le côté vertical est $\sum_{i=1}^N x_i = N \bar{x}$. L'aire A sous la courbe est la somme des aires des trapèzes :

¹Par exemple Lecaillon (1970, pp. 26-36) ou Lachaud (1992).

²Voir Theil (1967), Atkinson (1970), Gaswirth (1971), Sen (1973), Fei, Ranis et Kuo (1978), Dorfman (1979). Nous laissons à part le cas des sous - groupes dans la population ; voir Satchell (1987), Lambert et Aronson (1993).

$$A = \frac{x_N}{2} + \frac{(x_N + x_{N-1}) + x_N}{2} + \frac{(x_N + x_{N-1} + x_{N-2}) + (x_N + x_{N-1})}{2} + \dots$$

$$\dots + \frac{(x_N + x_{N-1} + \dots + x_1) + (x_N + \dots + x_2)}{2}$$

et l'indice de Gini est :

$$I_G = 1 - 2 \frac{A}{N^2 \bar{x}}$$

Une façon plus compliquée de calculer l'indice de Gini pour une distribution discrète est la suivante. Pour N individus, les revenus étant triés par l'ordre croissant, la courbe normalisée de concentration $L(i)$ relie la fréquence cumulée $\sum_{j=1}^i n_j / \sum_{j=1}^N n_j$ en abscisse à la masse relative cumulée (Combrouze, 1993, pp. 785-791). La figure est contenue dans un carré de côté égal à 1. L'indice de Gini est calculé comme $1 - 2A$, où A est la somme des superficies des trapèzes entre la courbe G et l'axe des x , c.a.d. :

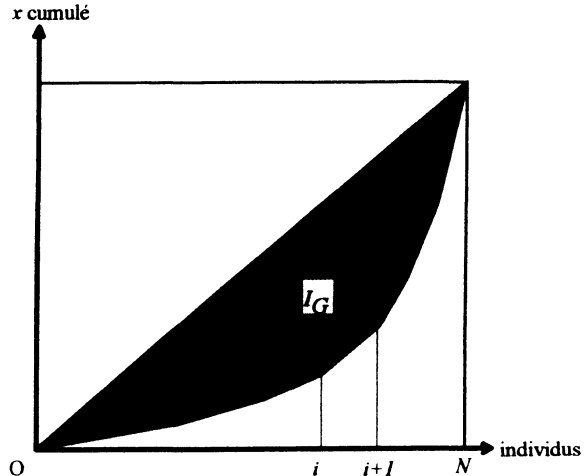


Figure 1

$$I_G = 1 - \left(\frac{n_1}{\sum_{j=1}^N n_j} \right) \left(\frac{n_1 x_1}{\sum_{j=1}^N n_j x_j} \right) + \sum_{i=1}^N \left(\frac{n_{i+1}}{\sum_{j=1}^N n_j} \right) \left(\frac{n_{i+1} x_{i+1} + 2 \sum_{j=1}^i n_j x_j}{\sum_{j=1}^N n_j x_j} \right)$$

Pour une variable continue x , l'indice de Gini noté $I_G(x)$, décrit l'inégalité dans la distribution $f(x)$ de x , quand x positive est ordonnée. La courbe de concentration normalisée $L(x)$ relie la fonction de distribution cumulée $F(x) = \int_0^x f(t) dt$ en abscisse (où t est la variable d'intégration), c.a.d. le pourcentage des individus gagnant moins que x , au moment de premier ordre de la distribution $F_1(x) = \int_0^x t f(t) dt / E(x)$ en ordonnée où $E(x) = \int_0^{+\infty} t f(t) dt$ est la valeur espérée de x c.a.d. la masse relative cumulée ou encore la part des revenus totaux gagnée par ces mêmes individus (Kendall et Stuart, 1969, p. 48) (Saporta, 1990, pp. 122 -124)³. Le rapport entre $F(x)$ et $F_1(x)$ est la courbe de concentration aussi appelée courbe de Lorenz⁴. La figure est contenue dans un carré de côté égal à 1. Le double de l'aire comprise entre la première diagonale et la fonction de concentration est l'indice de Gini $I_G(x) = 1 - 2 \int_0^1 L(x)$: il varie entre 0 et 1. Si cette aire est nulle, alors $I_G(x) = 0$ et la distribution cumulative des revenus est égalitaire : chacun individu a le même revenu. Si l'aire est égale à la moitié de l'aire du carré, alors $I_G(x) = 1$ et la distribution cumulative est la plus inégalitaire qu'il est

³Gaswirth (1971, 1972) a proposé certaines autres formules en termes de la fonction de distribution cumulée inverse pour le cas continu.

⁴Comme le note Kakwani (1977, pp. 719-720), Lorenz a inventé la courbe de concentration dans le cas où la variable est le revenu.

possible : un individu a tout le revenu, les autres ont zéro ⁵. En fait courbes de concentration et courbes de Lorenz ne sont pas exactement la même chose. Soit une fonction $g(x)$ non négative continue, avec une dérivée première, et soit le moment modifié de premier ordre $F_1[g(x)] = \frac{1}{E[g(x)]} \int_0^x g(x) f(x) dx$.

Le rapport entre $F(x)$ et $F_1(x)$ est la courbe de concentration $C[g(x)]$. Courbe de Lorenz et courbe de concentration coïncident si $g(x) = x$ ou si $g(x)$ est strictement monotone c.a.d. si $g'(x) > 0$ (Kakwani, 1977). L'indice de concentration est $I_C[g(x)] = 1 - 2 \int_0^1 L(x)$. Pour l'indice de concentration et l'indice de Gini, c'est la même chose : $I_G(x) = I_C(g(x))$ si $g'(x) > 0$.

Dans beaucoup de papiers, on considère seulement de petits transferts de revenu qui ne changent pas l'ordre des individus ⁶. Le point de vue est axiomatique. L'indice de Gini et les mesures similaires d'inégalité respectent cette condition qui nous garantit que la mesure est consistante. Le point de vue est "micro-économique". Cependant, ici nous acceptons de grands changements dans les revenus même s'ils changent l'ordre des individus dans la distribution, parce que notre objectif est d'analyser la pertinence de l'indice de Gini quand il est employé pour comparer deux distributions, par exemple pour deux pays ou le même pays à deux dates différentes, alors que la dominance stochastique n'est pas garantie. La question est : si deux distributions ont le même indice de Gini, sont-elles également égalitaires ? Le point de vue est macroéconomique. Rappelons qu'il est prouvé qu'il y a dualité entre la courbe de Lorenz $L(x)$ et la distribution cumulée $F(\theta x)$ où θ est un paramètre constant arbitraire (Fellman 1976) ⁷. Ici les changements que nous considérerons pourront violer cette hypothèse parce que nous ne pensons pas en termes de changement mais en termes de comparaison de distributions : quand on compare deux distributions, les changements peuvent être de toutes sortes. Deuxièmement, la question n'est pas de déterminer si deux distributions différentes peuvent avoir la même courbe de Lorenz, mais si deux distributions peuvent avoir la même valeur de l'indice de Gini. La réponse à cette question est oui comme on va le voir dans ce qui suit.

Les faiblesses de l'indice de Gini tiennent essentiellement dans le fait que plusieurs distributions de revenu, néanmoins bien différentes, peuvent sembler présenter un même niveau d'inégalité mesuré par l'indice de Gini. C'est le problème principal avec l'indice de Gini : il ne discrimine pas entre certaines distributions de la variable x même s'ils sont évidemment différents. Par exemple dans la figure suivante, les deux distributions continues sont évidemment différentes même si elles correspondent au même indice de Gini parce qu'elles sont symétriques par rapport à la deuxième diagonale. Est-ce la distribution $f_1(x)$ (courbe I dans la figure) ou la distribution $f_2(x)$ (courbe II) qui est la plus égalitaire ? Dans ce cas, l'indice de Gini ne nous aide pas à décider. A titre d'exemple, considérons deux distributions discrètes de revenu, $\{2, 2, 2, 4\}$ et $\{1, 3, 3, 3\}$. Pour les deux distributions, l'indice de Gini est $I_G = 0,15$, la somme des revenus et la moyenne sont les mêmes.

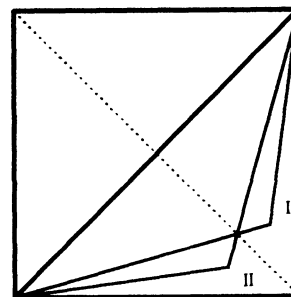


Figure 2

Cependant, quelle est la distribution la plus égalitaire, la première de type I ou la seconde de type II ?

⁵Il est possible de construire un indice de Gini en prenant l'aire sous la courbe de distribution cumulée : l'indice varie entre 1 et zéro, mais l'interprétation est similaire. On définit aussi l'indice de Gini relatif pour l'étude des systèmes de taxation, par exemple (x devient le revenu après l'impôt), en remplaçant le nombre d'individus en abscisse par la part du revenu avant l'impôt. Dans la suite, nous travaillons seulement sur les courbes de Lorenz ordinaires. On définit encore l'indice de Gini généralisé (Salles 1995), une mesure basée sur les rangs.

⁶Voir par exemple Ben Porath et Gilboa (1994), Barrett et Salles (1995).

⁷C'est pourquoi l'indice de Gini est insensible à un changement homothétique dans les revenus.

Comme dit plus haut, l'axiomatique des mesures d'inégalité insiste sur les petits changements dans les distributions obtenus par de petits transferts de revenus : la mesure doit être sensible à de petits transferts de revenu qui conservent le rang des revenus. Le problème est que pour le même pays, de grands changements dans le revenu et dans la distribution de la population peuvent être cachés par l'indice de Gini : il peut sembler stable en ce qui concerne l'inégalité de revenu, alors que la structure de ce revenu a changé énormément. Dans l'exemple ci-dessus, pendant le passage de la première distribution à la seconde, une partie de la classe moyenne/basse s'appauvrit, le reste s'enrichit et passe dans la classe moyenne / haute, tandis que les la classe aisée chute dans la classe moyenne / haute.

La différence entre deux distributions, pourtant symétriques par rapport à la deuxième diagonale peut être grande avec le même indice de Gini.

Par exemple, les deux distributions du tableau 1 ont la même valeur de l'indice de Gini ($I_G = 0,3$) et sont symétriques même si elles sont très différentes. En partant de la première distribution pour aller à la seconde, un transfert de 15 individus de la

| Revenus par tête | Nombre d'individus | Revenus par tête | Nombre d'individus |
|------------------|--------------------|------------------|--------------------|
| 1 | 40 | 1 | 25 |
| 4 | 10 | 4 | 25 |

Tableau 1

la classe la plus pauvre à la classe la plus riche n'affecte pas l'inégalité : ceci laisse perplexe !

La figure 3 indique comment varie l'indice de Gini quand on transfère les individus d'une catégorie à l'autre sans faire varier les revenus, en partant de 0% de pauvres pour aller à 100% (et de 100% de gens riches pour aller à 0%) : on peut voir que I_G augmente puis diminue, avec un maximum situé à 2/3 de pauvres avec 1/3 de la richesse et 1/3 de gens riches avec 2/3 de la richesse. L'indice de Gini se comporte ainsi parce que, après un moment, il y a si peu des gens riches que l'inégalité baisse même si le revenu moyen baisse aussi ⁸. Dans la deuxième figure suivante, la courbe en trait plein montre le parcours du coude avec le transfert d'individus ci-dessus et la ligne pointillée montre le parcours du coude qui est nécessaire pour obtenir un I_G constant : c'est une droite mais les revenus doivent aussi varier pour maintenir I_G fixe : seulement deux distributions ont le même Gini quand les revenus sont fixes pour deux catégories de revenu seulement ⁹.

Ce phénomène de symétrie est indirectement évoqué par Thon (1982, p. 142) dans son axiome "SC" pour *Strong Comparability* :

L'axiome SC introduit une propriété commode de comparabilité mais il est assez critiquable en lui-même; il implique par exemple que si $n=2$ (c.a.d. 50% de la population) reçoit tout le revenu, le niveau d'inégalité devrait être le même comme quand, par exemple, $n=100$ et quelqu'un monopolise le revenu entier (c.a.d. 1% de la population). Cela viole le bon sens ...

⁸Rappelons qu'on peut multiplier la moyenne par n'importe quel nombre positif arbitraire sans changer la valeur de l'indice de Gini ; donc on peut maintenir le revenu moyen à un niveau fixe.

⁹Rappelons qu'un transfert de revenus des plus riches aux plus pauvres diminue toujours l'indice de Gini. Par exemple, si le revenu de toutes les personnes riches passe de 4 à 3 et si le revenu de toutes les personnes pauvres passe de 1 à 2. Notons aussi que, comme on a le droit de changer homothétiquement le budget ou la population, les mêmes courbes sont obtenues quand le budget est constant ou quand la population est constante.

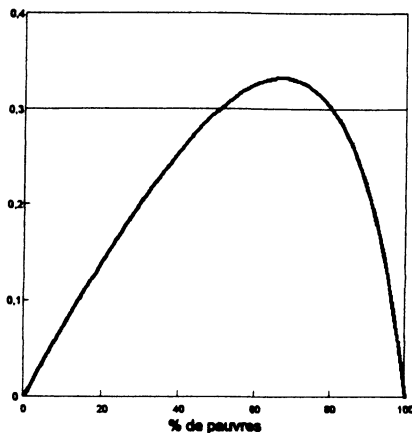


Figure 3. Indice de Gini et transfert de population

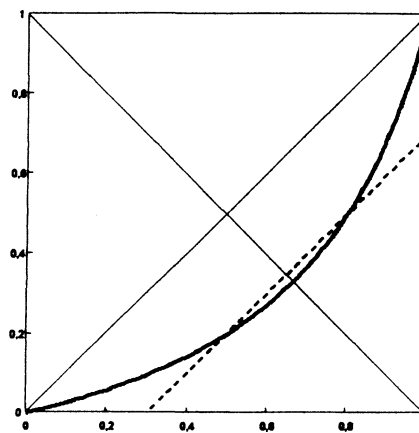


Figure 4. Position du coude et transfert de population

Je suis complètement d'accord avec cette citation, mais Thon ne poursuit pas plus avant dans cette voie. En fait à une même valeur de l'indice de Gini peut correspondre une infinité de distributions, pas seulement obtenues par une symétrie effectuée par rapport à la seconde diagonale. Par exemple, les quatre distributions du tableau 2 ont le même indice de Gini ($I_G = 0,75576$)¹⁰ :

| Revenus | Distribution 1 Nombre d'individus | Distribution 2 Nombre d'individus | Distribution 3 Nombre d'individus | Distribution 4 Nombre d'individus |
|---------|---|---|---|---|
| 1 | 100 000 | 100 000 | 10 000 000 | 10 000 |
| 10 | 10 000 | 100 000 | 1 000 000 | 100 000 |
| 100 | 1 000 | 10 000 | 353 516 | 100 000 |
| 1 000 | 100 | 100 000 | 100 | 100 000 |
| 10 000 | 10 | 100 000 | 100 | 63 945 |

Tableau 2

Un autre exemple (tableau 3). En science de gestion (gestion de stock, direction du personnel, gestion de clientèle, etc.), la « loi » des "80 - 20" est connue. Elle est appelée comme cela parce que typiquement 80% de la population représente 20% du chiffre d'affaires, et 20% de la population représente 80% du chiffre d'affaires¹¹ (voir figure 4). Supposons maintenant que dans un cas 80% des personnes gagnent 20% de revenus totaux et 20% des autres personnes gagnent 80% des revenus comme dans la figure ci-dessus. Dans un autre cas, 85% des personnes gagnent 25% des revenus et les 15% autres gagnent les 75% des revenus restant (quand un des plus pauvres gagne 1, un des plus riches gagne 17). Dans un dernier cas, 75% des personnes gagnent 15% des revenus et 25% gagnent 85% (quand un des plus pauvres gagne 1, un des plus riches gagne 17). Quel est le cas le plus égalitaire, sachant que dans ces trois cas, la valeur de l'indice de Gini est 0,6 (et que les cas 2 et 3 sont symétriques par rapport à la deuxième diagonale)¹² ?

¹⁰A nouveau, il faut rappeler que l'indice de Gini est insensible à un changement homothétique dans les revenus. Ainsi pour la deuxième distribution l'indice ne change pas si le nombre d'individus est uniformément égal à 1 au lieu de 10 000. Pour la même raison, le nombre total d'individus n'importe pas : on peut toujours normaliser pour avoir le même nombre d'individus dans chaque distribution.

¹¹Cette figure est obtenue pour la distribution suivante : chacun de 80% d'individus a 1 comme revenu et chacun de 20% autres a 16 comme revenu ; ici le coude est exactement sur la seconde diagonale, avec un indice de Gini égal à 0.6.

¹²Ces problèmes sont évidemment accrus pour les distributions bidimensionnelles (la distribution conjointe du revenu et du patrimoine, par exemple) ou multidimensionnelles, l'indice de Gini tra-

| Revenus en % | Distribution 1 % des indiv. Loi des « 80 - 20 » | Revenus en % | Distribution 2 % des indiv. | Revenus en % | Distribution 3 % des indiv. |
|--------------|---|--------------|-----------------------------|--------------|-----------------------------|
| 20 | 80 | 25 | 85 | 15 | 75 |
| 80 | 20 | 75 | 15 | 85 | 25 |

Tableau 3

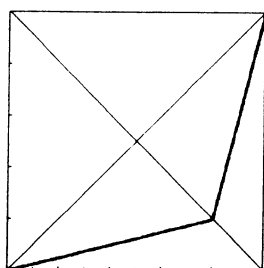


Figure 5. Loi des "80 - 20"

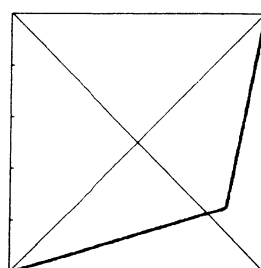


Figure 6. Cas 2

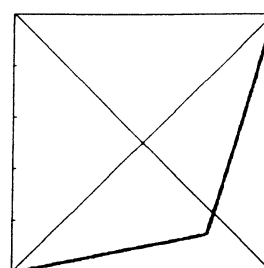


Figure 7. Cas 3

Aussi, un indice plus discriminant pourrait être préférable ¹³ et nous avons besoin d'un système pour voir si le coude de la distribution est essentiellement à droite ou à gauche de la deuxième diagonale, et de combien il en est. Seulement après cela, il sera possible de discuter pour savoir quelle distribution est la plus inégalitaire ¹⁴.

B. L'indice de Kakwani

L'idée originale de Kakwani (1977), et aussi de Fellman (1971), était la suivante. Il voulait évaluer les effets d'un système d'impôt, en comparant la courbe de Lorenz avant et après impôt. Il montre que si l'impôt est strictement proportionnel au revenu, alors les deux courbes sont confondues ; si le taux moyen d'impôt augmente avec le revenu, alors la courbe de Lorenz après impôt est sous la courbe correspondante avant impôt ¹⁵. Pour la comparaison de deux systèmes d'imposition, si pour le même impôt moyen l'élasticité d'imposition d'un système est uniformément plus forte que pour un autre système d'imposition, alors le revenu du premier système est plus égalitaire. La distance entre deux courbes de concentration dépend de la progressivité d'imposition et aussi du taux moyen d'imposition mais, comme indiqué par Kakwani (1977, p. 723), puisqu'il est possible d'augmenter le taux moyen d'imposition sans accroître la progressivité, la distance entre les deux courbes n'est pas une bonne mesure de progressivité. D'où l'idée est de calculer l'aire entre les deux courbes : l'indice de Kakwani est comparé deux distributions pour ce qui est de leur niveau d'inégalité. Soit deux distributions $f_1(x_1)$ et $f_2(x_2)$

ditionnel traitant seulement des distributions monodimensionnelles (comme la distribution des revenus dans un pays). Il faut noter que les difficultés ci-dessus sont complètement similaires à ce qu'on rencontre quand on veut décrire la forme d'une série statistique par un indicateur comme le moment (moyenne, variance, etc.), la médiane, la mode, etc.

¹³On ne doit pas confondre. Ce phénomène, des distributions multiples pouvant correspondre au même indice de Gini, ne peut pas être éliminé même avec la courbe de Lorenz généralisée. Il est prouvé qu'il y a une dualité entre la distribution et la courbe de Lorenz généralisée qui en découle (Thistle 1989) (voit aussi Iritani et Kuga (1983)).

¹⁴ Une autre voie pourrait consister à calculer le ratio de l'aire à droite de la deuxième diagonale sur l'aire à la gauche de cette diagonale, afin de donner une idée du degré de déformation mais alors, la même difficulté qu'avec l'indice de Gini surviendrait, c'est à dire qu'à une même valeur du ratio pourra correspondre une infinité de distributions.

¹⁵Ce résultat est aussi dans Fellman (1971).

et l'indice de Gini correspondant $I_G(x_1)$ et $I_G(x_2)$. L'indice de Kakwani $I_K(x_1, x_2)$ est simplement la différence entre les deux indices de Gini : $I_K(x_1, x_2) = I_G(x_1) - I_G(x_2)$ et si cette grandeur est positive alors $f_1(x_1)$ est plus inégalitaire que $f_2(x_2)$ ¹⁶.

L'indice de Kakwani est un complément de l'indice de Gini parce qu'il sert comparer des situations d'inégalité. Comme il provient de l'indice de Gini, l'indice de Kakwani rencontre les mêmes problèmes. Cependant, un problème supplémentaire survient : une infinité de couples d'indices de Gini correspond au même indice de Kakwani (chacun des indices de Gini correspondant à une infinité de distributions).

Par exemple, un indice de Kakwani nul peut correspondre à deux distributions qui sont égalitaires ou à deux distributions que ne sont pas complètement égalitaires ; parallèlement, un indice de Kakwani positif peut correspondre, quand on compare deux distributions, à une distribution égalitaire et à une distribution un peu inégalitaire ou à une distribution un peu inégalitaire et une distribution surtout inégalitaire. C'est ennuyeux. Ainsi, même si on admet l'idée que les indices de Gini décrivent correctement l'inégalité, les indices de Kakwani échouent à décrire correctement les différences d'inégalité. Par exemple, une valeur donnée de cet indice, qui indiquerait une stabilité dans la situation relative de deux pays du point de vue de l'inégalité, peut cacher une évolution importante. Un pays peut passer d'un indice de Gini nul à une valeur de 0,3, tandis que l'autre pays passe de 0,3 à 0,6 : le Kakwani est inchangé, même si on perçoit bien que les deux situations ne soient pas comparables, le premier pays passant même de l'absence d'inégalité à un niveau d'inégalité (sous réserve que l'indice de Gini décrive convenablement l'inégalité) égal au niveau d'inégalité initiale du deuxième pays.

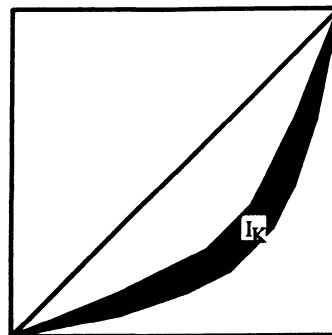


Figure 8. Indice de Kakwani

II. RECONNAISSANCE DE FORME ET MESURE DE L'INEGALITE

Les difficultés évoquées à propos des indices de Gini et de Kakwani correspondent pour l'essentiel à un problème bien connu, celui des courbes de Lorenz qui se coupent ¹⁷ : quand deux courbes de Lorenz se coupent, il n'est pas possible de dire qu'une distribution domine une autre et c'est sur cette question que tourne ce papier.

A. Reconnaissance de forme

Détecter la position de la médiane ¹⁸ n'est pas suffisant en soi-même pour caractériser la forme de la courbe de concentration parce que beaucoup de formes peuvent correspondre à une même médiane. Une porte de sortie peut être la définition de formes de courbes de concentration typiques, correspondant à des situations reconnues d'inégalité. Des formes typiques sont donc proposées. On obtient ainsi une typologie de courbes de concentration, sans jugements de valeur.

Chaque distribution peut être vue comme ayant trois parties, une avec une pente quasi-horizontale (la dérivée première est faible, A dans le figure 9), une avec une pente moyenne (la dérivée est moyenne, B) que nous pouvons qualifier de "marais", par référence à la situation politique pendant la Révolution Française, une avec une pente quasi-verticale (la dérivée première est forte, C). Une distribution conti-

¹⁶Parallèlement à l'indice de Gini, on définit l'indice de Suits à partir de l'indice de Gini relatif, au lieu de l'indice Kakwani.

¹⁷Voir Yitzhaki et Olkin (1986), Muliere et Scarsini (1989), Thistle (1989).

¹⁸Le nombre sur l'axe de fréquences cumulées tel que $I_G = 0.5$.

nue peut en général être décomposée de cette manière ¹⁹. Souvent, la partie B disparaît et la figure a seulement un coude comme dans la loi des « 80 - 20 » (figure 10) ²⁰.

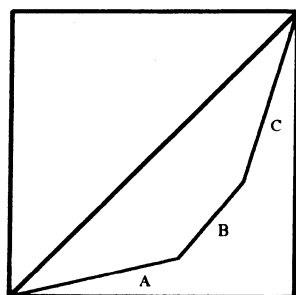


Figure 9. Décomposition en deux coudes

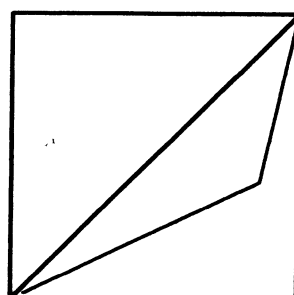


Figure 10. Décomposition en un seul coude

Il y a des courbes continues dans lesquelles la zone B est placée symétriquement par rapport à la deuxième diagonale ou des courbes discrètes dans lesquelles le coude est placé sur la deuxième diagonale (comme pour la figure de la loi des « 80 - 20 »). Il est faux de considérer ce type de distribution comme neutre.

Il y a des courbes tordues à gauche, c.a.d. avec une zone B située à gauche de la deuxième diagonale (inégalité en défaveur des plus pauvres, avec une classe haute - moyenne plus nombreuse, comme dans la deuxième distribution donnée dans l'exemple ci-dessus). La courbe peut être tordue à gauche même s'il n'y a aucun coude. Par exemple, une distribution uniforme (même nombre d'individus pour chaque revenu c.a.d. $n_j = n$ pour tout j , avec la même distance entre chaque revenu c.a.d. $x_j = a + b j$ pour tout j) correspond à une courbe de Lorenz tordue à gauche. La fréquence cumulée et la masse relative cumulée deviennent alors

$$\sum_{j=1}^i 1 / \sum_{j=1}^N 1 = i/N$$

$$\text{et } \sum_{j=1}^i x_j / \sum_{j=1}^N x_j = a i + b \frac{i(i+1)}{2} / a N + b \frac{N(N+1)}{2}$$

Ici le point de tangence est toujours à l'abscisse 0.5.

¹⁹Cette décomposition est très connue dans la pratique managériale : c'est la loi dite A-B-C, généralisation de la loi des « 80-20 ». Par l'exemple pour la gestion des stocks, où les pièces sont gérées suivant une décomposition en trois : une part importante des pièces du stock a un turn-over bas, une autre part a un turn-over plus important, et pour une petite part a un turn-over très fort.

²⁰ Parfois pour les distributions continues, et mêmes pour les distributions discrètes, les coudes ne sont pas clairs. Comme remplacement du coude, nous verrons qu'on peut prendre le point où la courbe est tangente à la parallèle à la première diagonale pour une distribution continue ou le segment de courbe de pente sensiblement égale à 1 pour une distribution discrète.

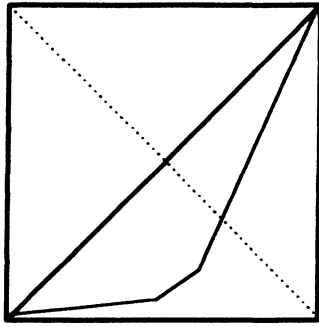


Figure 11. Courbe tordue à gauche

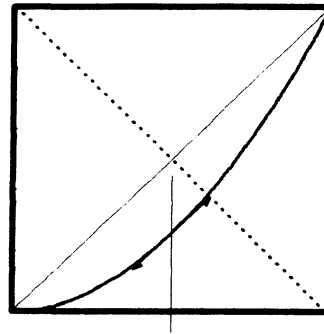


Figure 12. Courbe tordue à gauche, sans coude

La courbe peut être tordue à droite, c.a.d. avec une zone B ou un coude situés à droite de la deuxième diagonale (inégalité dans la faveur du plus riche, avec plus de pauvres, comme pour la première distribution donnée dans l'exemple), et toutes combinaisons admissibles des cas précédents, comme une grande zone B à la gauche de la deuxième diagonale.

Parfois, la zone B peut être à cheval sur la deuxième diagonale, plus à la gauche ou à droite. La largeur de la zone B est aussi importante. Quand la largeur est grande, il y a une importante classe moyenne.

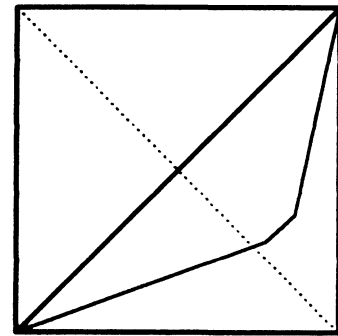


Figure 13. Courbe tordue à droite

B. Caractérisation des courbes

La position du point de tangence, du coude ou de la zone B, à droite ou à gauche de la seconde diagonale, et sa distance à celle-ci, prend plus d'importance que la valeur de l'indice de Gini. La difficulté consiste à détecter la configuration pour classer la distribution étudiée dans la bonne classe. Premièrement, rappelons qu'un point dans un espace euclidien à deux dimensions nécessite seulement deux nombres réels pour être défini complètement. Il y a une infinité de façons pour choisir ces deux nombres. Ici nous proposerons certaines, mais on peut en choisir une autre. Cependant, ce nouvel indice restera basé sur la courbe de Lorenz normalisée. Premièrement, on mesure l'angle(s) entre le coude(s) et la deuxième diagonale. Deuxième on mesure la "distance" entre le centre de la figure et le coude(s).

Ainsi, pour une distribution discrète à double coude, par définition, l'angle α caractérise le coude droit et l'angle β le coude gauche. Les deux angles sont positifs si les angles tournent dans le sens inverse des aiguilles d'une montre. Dans la figure, α est positif et β est négatif. Les angles α et β sont suffisants pour caractériser la catégorie de la courbe. Pour une distribution discrète à un seul coude ou pour une distribution continue avec une zone B étroite, le principe est le même mais les nombres d et d' sont égaux, comme les angles α ou β .

Si $\alpha = \beta > 0$ alors il y a seulement un coude à droite et les pauvres sont plus nombreux : plus l'angle est grand, plus il y a de personnes dans la classe pauvre. Si $\alpha = \beta < 0$ alors il y a seulement un coude à gauche et la classe haute - moyenne est plus nombreuse : plus l'angle

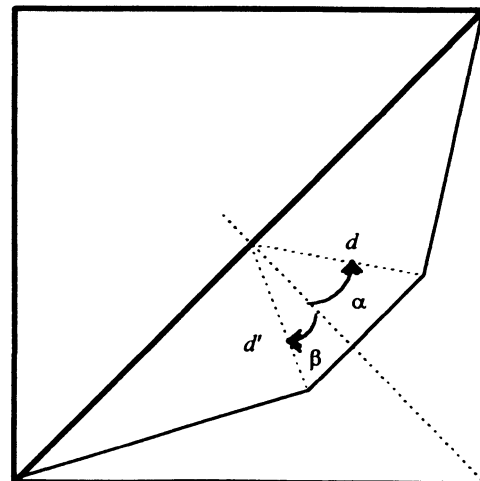


Figure 14

est grand, plus il y a de personnes riches. Si $\alpha \neq \beta$ il y a une zone B ; le nombre $|\alpha| + |\beta|$ indique la largeur de cette zone B : plus il est grand, plus la zone B est grande. Si $\alpha > 0$ et $\beta < 0$ la zone B est à cheval sur la deuxième diagonale. Si les deux angles sont positifs la zone B est entièrement à droite et inversement. Si α est grand et positif et $|\beta|$ petit ou si β est grand et positif et $|\alpha|$ est petit, alors la zone B démarre près de la deuxième diagonale. Si $\alpha = -\beta$ alors la zone B est symétrique par rapport à la deuxième diagonale. Les distances d et d' finissent de caractériser la position des coudes ; ils définissent aussi la pente et la position de la zone B.

Une courbe type A-B-C (linéaire "par morceaux") ou une courbe approximée par une courbe de type A-B-C, est complètement et suffisamment décrite par les quatre nombres α, β, d, d' et l'indice de Gini devient inutile. Pour de telles courbes, concernant les indices de Gini, tous les problèmes décrits ci-dessus sont éliminés. Ces quatre nombres prolongent l'analyse simple basée sur le calcul de la médiale.

En fait, il pourrait être plus facile de calculer sur coordonnées cartésiennes au lieu de coordonnées polaires. Notons (X_L, Y_L) et (X_R, Y_R) les coordonnées cartésiennes des coudes gauche et droit. On a $I_G = X_R - Y_R + X_L Y_R - X_R Y_L$. Dans le cas de deux coudes, comme la deuxième diagonale est donnée par $Y = 1 - X$, si $Y_L < 1 - X_L$ le premier coude est à gauche de la deuxième diagonale. Dans le cas d'un seul coude, $X_L = X_R = X$, $Y_L = Y_R = Y$ et $I_G = X - Y$ alors, comme la deuxième diagonale est donnée par $Y = 1 - X$, la courbe est tordue à gauche si $Y < 1 - X$ et l'indice est égal à une constante I_G^* si $Y = X - I_G^*$.

Le problème est de savoir comment détecter les coudes. Premièrement, rappelons que la courbe est convexe si les différences secondes de la courbe pour une courbe discrète (la dérivée seconde pour une courbe continue) sont toujours positives ou nulles. Aussi, le problème est de détecter le point où les différences secondes, ou la dérivée seconde, s'annulent ou sont proches de zéro (en ce point la courbe est une ligne droite ou une quasi ligne droite et où elles sont significativement positives (et il y a clairement un coude).

Quand il n'y a aucun coude évident, c.a.d. quand les différences secondes pour une courbe discrète (ou la dérivée seconde pour une courbe continue) changent trop lentement pour que le coude soit détectable, une bonne approche consiste à remplacer le coude par un autre point significatif et à le considérer comme s'il était un coude ; on peut prendre le point où la courbe est tangente à la parallèle à la première diagonale (pente égale à 1) pour une courbe continue ou le segment de courbe de pente sensiblement égale à 1 pour une distribution discrète²¹.

C. Pour revenir à l'évaluation de l'inégalité

Maintenant nous avons un petit ensemble de nombres, deux ou quatre, pour caractériser la forme de courbes de concentration. Cependant, il reste les questions suivantes. Quel est la distribution la moins égalitaire ? Quelle distribution préférer ?

1. Un coude

Dans le premier cas du premier exemple dans l'introduction, $\{2, 2, 2, 4\}$ et $\{1, 3, 3, 3\}$, on a un riche face à trois personnes de la classe basse - moyenne et dans le second, on a un pauvre face à trois personnes de la classe haute - moyenne mais le choix n'est pas clair. Pour la plupart des gens, la deuxième distribution semble moins égalitaire que la première parce que le ratio du plus haut revenu au plus bas

²¹ Une autre approche basée sur un autre couple de nombres, les traditionnels indice de Gini et la médiale, n'est pas convenable parce que la médiale est dépendante de l'indice de Gini. Donc en regardant seulement la médiale d'une courbe, on ne peut pas dire à quelle catégorie elle appartient.

est 3 quand il est seulement 2 dans la première, et en même temps, les pauvres sont moins pauvres mais ils sont beaucoup plus nombreux dans le premier cas.

L'exemple du tableau 4 est proche du dernier exemple de l'introduction (tableau 3, indice de Gini égal à 0.6). Du cas tordu à droite au cas tordu à gauche, la proportion du revenu le plus fort sur le plus bas reste fixe (154,8462 : 1) aussi cette proportion ne joue aucun rôle ; il y a des personnes plus riches dans le cas tordu à gauche mais en même temps il y a moins de pauvres !

C'est pourquoi l'indice de Gini échoue à discriminer entre les cas tordus à droite et à gauche. Du cas neutre au cas tordu à droite ou à gauche, les personnes les plus riches ont un revenu plus fort (154,8462 au lieu de 16) si chaque pauvre gagne toujours la même chose soit 1, mais du cas neutre au cas tordu à droite, les plus riches sont moins nombreux, alors qu'ils sont plus nombreux du cas neutre au cas tordu à gauche et l'indice de Gini échoue aussi à discriminer. Pour un budget constant donné (celui du cas tordu à gauche, c.a.d. 61 pour 100 individus), les pauvres doivent avoir chacun 24,035 et les personnes riches doivent avoir chacun 3721,677 dans le cas tordu à droite (soit $24,035 \times 99 + 3721,677 \times 1 \cong 6100$), mais seulement 1 et 154,846 respectivement dans le cas tordu à gauche (soit $1 \times 61 + 154,846 \times 39 \cong 6100$) ; aussi les pauvres sont moins pauvres dans le cas tordu à droite que dans le cas tordu à gauche.

| Neutre | | Tordu à droite | | Tordu à gauche | |
|--------------|-------------------------|----------------|-------------------------|----------------|-------------------------|
| Revenus en % | Distrib. 1 % d'individ. | Revenus en % | Distrib. 4 % d'individ. | Revenus en % | Distrib. 5 % d'individ. |
| 20 | 80 | 39 | 99 | 1 | 61 |
| 80 | 20 | 61 | 1 | 99 | 39 |

Tableau 4.

De plus, en examinant les cas limites on voit que la population entière est dans une seule catégorie dans le cas tordu à droite (les plus riches ont un revenu qui tend vers l'infini mais leur proportion tend vers zéro) alors qu'elle est séparée en deux dans le cas tordu à gauche, avec une partie très pauvre (les plus pauvres dont la proportion tend vers 100% ont un revenu qui tend vers zéro). En un certain sens, le cas tordu à droite est plus égalitaire.

| Tordu à droite | | Tordu à gauche | |
|----------------|------------------------------|----------------|------------------------------|
| Revenus en % | Distribution 0 % d'individus | Revenus en % | Distribution 6 % d'individus |
| 40 | 100 | 0 | 60 |
| 60 | 0 | 100 | 40 |

Tableau 5. Cas limites

Donner le plus possible aux plus pauvres correspond au critère du maximin rawlsien, critère que je propose donc d'adopter ici. Mais le maximin originel ne peut pas directement être pris en compte parce qu'il ne peut pas discriminer entre les cas dotés du même I_G : on a vu que lorsqu'il y a seulement deux classes de revenu, si le revenu des plus pauvres est augmenté et le revenu des plus riches est diminué, l'indice de Gini varie et l'inégalité aussi. Il faut donc adapter le maximin qui doit devenir :

Entre deux distributions à un coude avec la même valeur de l'indice de Gini, nous préférons comme moins inégalitaire la distribution qui est la plus tordue à droite.

L'indicateur peut être indifféremment l'angle α ou l'abscisse X du coude qui doivent être le plus fort possible. Donc parmi deux distributions avec seulement deux classes de revenu nous préférons la distribution dans laquelle les pauvres ont la plus grande partie du revenu, par exemple $\{2, 2, 2, 4\}$ contre $\{1, 3, 3, 3\}$, et aussi le cas tordu à droite du tableau 4 ; en regardant les deux distributions du deuxième exemple d'introduction (tableau 1), nous préférons la première distribution (80% de pauvres mais avec 50% du revenu au lieu de 50% de pauvres mais avec 20% du revenu).

Ceci fonctionne aussi pour deux distributions avec le même indice de Gini mais qui sont toutes les deux tordues à droite (les coudes sont alignés le long d'une ligne à 45°). Nous préférons ainsi le cas limite tordu à droite du tableau 5 au cas tordu à droite du tableau 4. Dans la figure 16, nous préférons la distribution *I* parce que $X_1 > X_2$. C'est la même chose pour deux distributions tordues à gauche : *I* est préféré parce que les pauvres sont moins pauvres (figure 17).

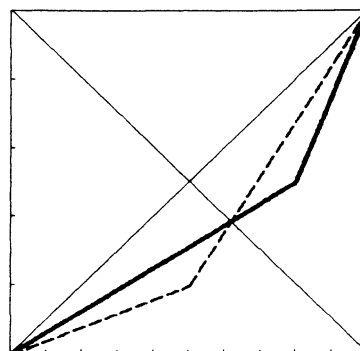


Figure 15. La distribution préférable, tordue à droite

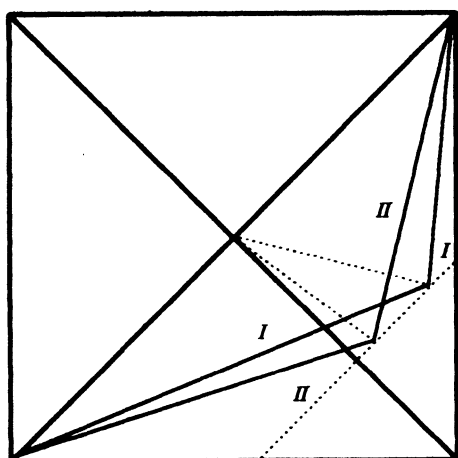


Figure 16. Même Gini, tordue à droite

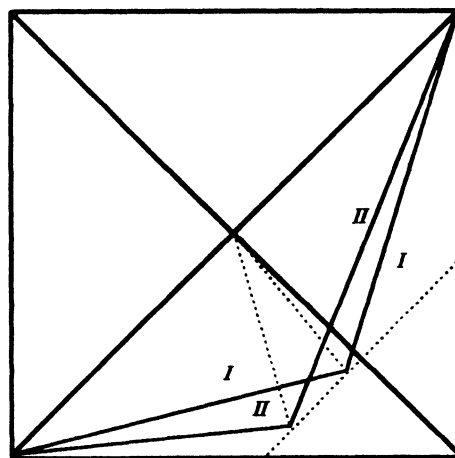


Figure 17. Même Gini, tordue à gauche

Qu'arrive-t-il pour deux distributions avec le même angle α , mais avec une valeur différente de d ? La réponse est claire : plus d est petit, plus l'indice de Gini est faible. Aussi on mesure l'indice de Gini de la distribution puis on considère l'angle ; mais on ne veut pas bâtir un nouvel indice avec un nombre unique pour ne pas rester piégé dans la trappe dans laquelle tombent les indices synthétiques comme l'indice de Gini.

2. Deux coudes

Pour les distributions avec deux coudes, c'est plus compliqué parce que nous avons deux degrés de liberté en plus et, pour un indice de Gini donné, les coudes ne sont pas nécessairement alignés sur une droite à 45° . Mais l'idée générale est la même. D'abord, nous identifions le type de la courbe (est-ce que la zone B existe, est-elle grande, est-elle à droite ou à gauche de la seconde diagonale, etc. ?). Deuxièmement, nous préférons les angles positifs, au moins α mais aussi β si possible mais nous devons adapter les critères. Quand on compare deux courbes dotées du même indice de Gini, le problème apparaît quand la hiérarchie des angles α et β est inversée. Alors, les distances d et d' jouent un rôle²². Dans la suite α_I et β_I sont les angles de la première courbe et α_{II} et β_{II} sont les angles de la

²²Même si, étant donné α et β , les distances d et d' sont liées à la valeur de l'indice de Gini.

seconde courbe. Supposons que $\alpha_I < \alpha_{II}$, $\beta_I > \beta_{II}$ et $d_I > d_{II}$, $d'_I < d'_{II}$ pour le même indice de Gini. Il n'y a aucune façon systématique de répondre concernant l'inégalité. On note (X_L^I, Y_L^I) et (X_R^I, Y_R^I) les coordonnées des coudes gauche et droit de la première courbe et (X_L^{II}, Y_L^{II}) et (X_R^{II}, Y_R^{II}) les coordonnées des coudes gauche et droit de la deuxième courbe. Ce type de courbe décrit une population avec trois classes, la plus pauvre est à gauche sur la courbe de Lorenz. Le coude gauche décrit l'inégalité entre la classe pauvre et la classe moyenne et le coude droit décrit l'inégalité entre l'agrégat des pauvres et des classes moyennes d'une part et la classe riche d'autre part. Si on pense que le premier phénomène est le plus important, on considère le coude gauche. Dans la figure 18, au coude gauche, le critère fonctionne et les pauvres sont plus nombreux et plus riches dans la distribution *I* que dans la distribution *II*. En marquant les coudes principaux par les lettres *a* et *b* (la courbe *I* a un coude principal en *a* et la courbe *II* en *b*), si le critère du coude principal est appliqué, c.a.d. si les courbes sont approximées comme ayant un seul coude, alors la courbe *I* est tordue à droite et pas la courbe *II*.

Les coudes peuvent aussi être alignés le long d'une ligne avec une pente dépassant 90° comme pour le coude gauche dans la figure 19 ; il semble y avoir un problème de choix mais en fait non, car alors le critère n'est plus nécessaire : si $X_L^I > X_L^{II}$, on est certain que la courbe *II* est moins inégalitaire que la courbe *I* au coude gauche parce que les pauvres sont moins nombreux mais plus riches dans *II* que dans *I*.

Mais si on pense que le deuxième phénomène - l'inégalité entre la classe pauvre et moyenne d'une part et la classe riche d'autre part - est la plus importante, on doit considérer le coude droit. Il n'y a aucune symétrie entre les coudes droit et gauche : les classes pauvre et moyenne seulement sont comparées au coude gauche tandis que toutes les classes, sauf la classe riche, sont comparées à la classe riche au coude droit. Dans les deux figures 18 et 19, le choix entre les deux courbes ne requiert pas l'aide du critère ci-dessus au coude droit : *II* est choisi au lieu de *I* à ce coude parce que les deux classes les plus basses sont plus nombreuses mais moins riches dans *I* que dans *II* au coude droit, même si courbe *I* est toujours choisie suivant le critère du coude principal.

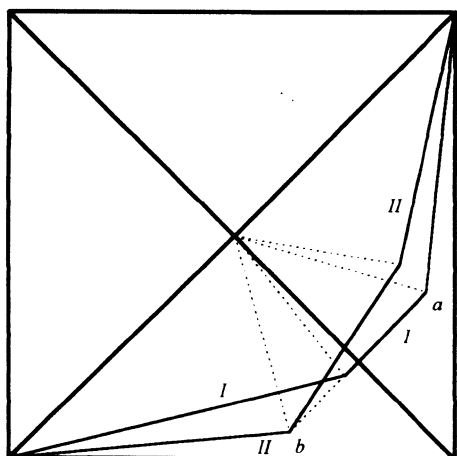


Figure 18. Coude à gauche avec choix axiomatique

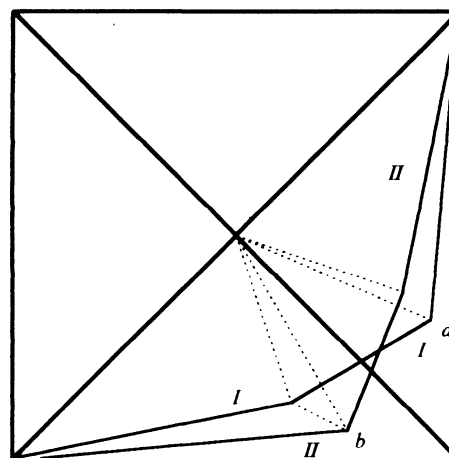


Figure 19. Coude à gauche avec choix évident

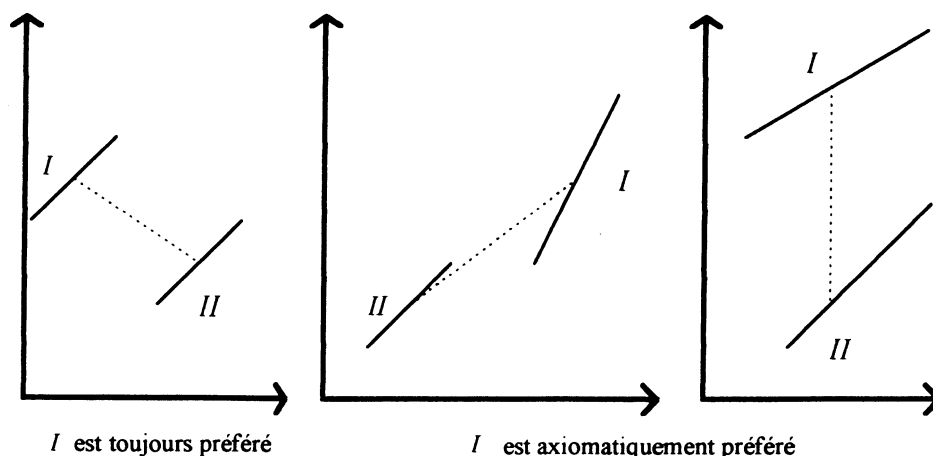


Figure 20. Résumé : choix à l'endroit d'un coude

Cela pourrait être plus compliqué quand les courbes se croisent deux fois (figure 21) : la courbe II est choisie avec la règle de coude principal (coude *b*), mais la distribution I est choisie au coude droit tandis que l'inégalité entre la classe pauvre et la classe moyenne est grande dans la distribution I (au coude gauche). Le choix entre I et II dépend de ce qu'on considère comme important. Il faut noter que ce problème survient généralement avec les courbes qui se croisent deux fois, dans lesquelles la hiérarchie des angles est respectée : $\alpha_{II} > \alpha_I$ et $\beta_{II} > \beta_I$.

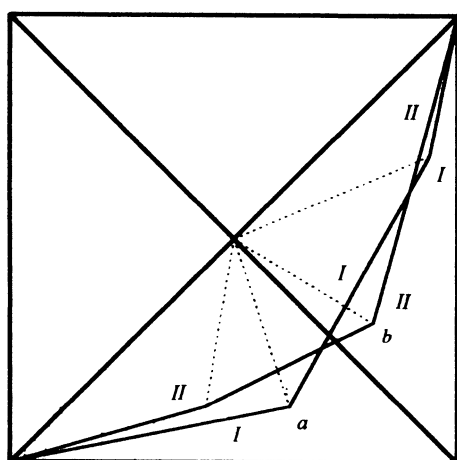


Figure 21. Deux croisements

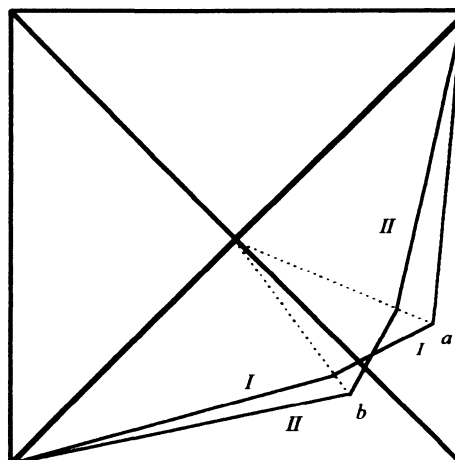


Figure 22. Cas particulier : inversion des *d*

Comme cas particulier, supposons que les deux distributions soient telles que $\alpha_I = \alpha_{II}$, $\beta_I = \beta_{II}$ et $d_I < d_{II}$, $d'_I < d'_{II}$. Les courbes ne croisent pas et la courbe I a le plus bas Gini. Maintenant supposons que les angles soient comme ci-dessus mais que les distances soient inversées, $d_I > d_{II}$ et $d'_I < d'_{II}$ (figure 22) et que la valeur de l'indice de Gini soit la même pour les deux courbes. Là encore, approximé par une courbe à un seul coude, la courbe I est la plus tordue à droite et nous la préférons comme moins inégalitaire.

Comme autre cas particulier, la part peut être la même pour I et II, c.a.d. $X_{II}^R = X_I^R$ (figure 23). Alors le critère du maximin dit que nous devons choisir II au coude droit.

III. CONCLUSION

Nous avons exposé les faiblesses des indices de Gini et de Kakwani quand ils sont utilisés comme outils pour évaluer et comparer des distributions du point de vue de l'inégalité. Essentiellement, ces indices ne sont pas discriminants : beaucoup de distributions avec un profil différent peuvent correspondre à la même valeur de l'indice de Gini et une infinité de couples de distributions peuvent correspondre à un même indice de Kakwani.

Donc nous proposons d'approximer la distribution par une distribution à deux ou trois morceaux linéaires : la nouvelle distribution est entièrement décrite par la position de son ou de ses deux coudes, permettant l'évaluation complète de l'inégalité et la comparaison de distributions.

Dans le cas le plus simple où il y a seulement un coude, une application du maximin de Rawls conduit à préférer la distribution la plus tordue à droite comme moins inégalitaire : les pauvres sont moins pauvres. Pour des distributions à deux coudes, le problème est plus compliqué, même si le même principe est appliqué : on considère soit le coude gauche, soit le coude droit, suivant que l'on pense que le phénomène le plus important est la relation entre la classe pauvre et la classe moyenne, ou entre la classe riche et le reste de la population.

Avec cette approche on quitte partiellement les indices numériques pour aller vers la reconnaissance de formes et l'analyse qualitative de l'inégalité, certes moins satisfaisants pour le bon sens, mais moins sujets à critique au plan scientifique parce que permettant d'éviter une quantification trompeuse.

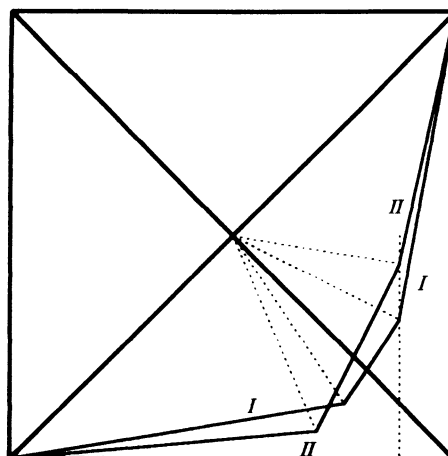


Figure 23. Même part de la population au coude droit

IV. REFERENCES BIBLIOGRAPHIQUES

- ATKINSON Anthony B. ((1970) "On the measurement of inequality", *Journal of Economic Theory*, 2, pp. 244-263.
- BARRETT C.R. and Maurice SALLES (1995) "On the generalization of the Gini coefficient", *Mathematical Social Sciences*, 30, pp. 235-244.
- BEN PORATH Elchanan and Itzhak GILBOA (1994) "Linear measures, the Gini index, and the income-equality trade-off", *Journal of Economic Theory*, 64, pp. 443-467.
- BERREBI Z.M. and Jacques SILBER (1987) "Dispersion, asymmetry and the Gini index of inequality", *International Economic Review*, 28(2), pp. 331-338.
- COMBROUZE A. (1993) *Probabilités et statistiques*, PUF, Paris.
- DORFMAN Robert (1979) "A formula for the Gini coefficient", *The Review of Economics and Statistics*, 61(1), pp. 149-149.
- FEI John. C. H., Gustav RANIS and Shirley W. Y KUO (1978) "Growth and the family distribution of income by factors components", *Quarterly Journal of Economics*, 92, pp. 17-53.
- FELLMAN J. (1976) "The effect of transformations on Lorenz curves", *Econometrica*, July, 823-24.
- GASWIRTH Joseph L. (1971) "A general definition of the Lorenz curve", *Econometrica*, 39, pp. 1037-1039.
- GASWIRTH Joseph L. (1972) "The estimation of the Lorenz curve and the Gini index", *The Review of Economics and Statistics*, 54, pp. 306-316.
- IRITANI Jun and Kiyoshi KUGA (1983) "The duality between the Lorenz curve and the income distribution functions", *Economic Studies Quarterly*, April, pp. 9-21.
- KAKWANI Nanak (1977) "Applications of Lorenz curves in economic analysis", *Econometrica*, 45, 3, pp. 719-727.

- KENDALL, M.G. and A. STUART (1969) *The advanced theory of statistics*, Vol. 1, Charles Griffin & Company Limited, London.
- LACHAUD Claire (1992) *Équité dans le financement et la prestation de soins de santé en France : une approche par les courbes de concentration*, Thèse, Université de Lyon I.
- LAMBERT Peter J. and J. Richard ARONSON (1993) "Inequality decomposition analysis and the Gini coefficient revisited", *The Economic Journal*, 103, pp. 1221-1227.
- LECAILLON Jacques (1970) *L'inégalité des revenus*, Cujas, Paris.
- MOYES Patrick (1989) "Some classes of functions that preserve the inequality and welfare orderings of income distributions", *Journal of Economic Theory*, 49, pp. 347-359.
- MULIERE Pietro and Marco SASSINI (1989) "A note on stochastic dominance and inequality measures", *Journal of Economic Theory*, 49, pp. 314-323.
- SAPORTA Gilbert (1990) *Probabilités, analyse de données et statistique*, Editions Technip, Paris.
- SATCHELL S. E. (1987) "source and subgroup decomposition inequalities for the Lorenz curve", *International Economic Review*, Vol. 28, n° 2, pp. 323-329.
- SEN Amartya (1973) *On economic inequality*, Oxford University Press, Dehli.
- _____ (1974) "Informational bases of alternative welfare approaches", *Journal of Public Economy*, 3, pp. 387-403.
- THEIL Henri (1967) *Economics and Information Theory*, Rand McNally, Chicago.
- THISTLE Paul. D. (1989) "Ranking distributions with generalized Lorenz Curves", *Southern Economic Journal*, 56(1), pp. 1-12.
- THON Dominique (1982) "An axiomatization of the Gini coefficient", *Mathematical Social Science*, 2, pp. 131-143.
- YITZHAKI S. and I. OLKIN (1986) "Concentration curves", *Department of Statistics*, Stanford University.