CHAPITRE 1-4

MESURE DE L'INÉGALITÉ ET DE LA CONCENTRATION

Plan

CHAPITRE 1-4

MESURE DE L'INÉGALITÉ ET DE LA CONCENTRATION

Références : Arriaga, 1975, p. 65-71 ; Taylor, 1977, 179-185 ; Mills et Hamilton, 1989, p. 413-414 ; Kendall et Stuart (1991, p. 58) ; Jayet (1993, p. 18-29) ; Valeyre (1993) ; MacLachlan et Sawada (1997).

Nous nous attachons dans ce chapitre à l'examen des différentes valeurs d'une même variable dans un ensemble d'observations. Une mesure d'inégalité (on dit aussi « de disparité ») indique à quel degré les valeurs diffèrent les unes des autres. Prenons, par exemple, les revenus des habitants d'un pays ; une mesure d'inégalité du revenu sert à quantifier le degré d'inégalité de la distribution du revenu entre les habitants du pays, de façon à pouvoir le comparer à celui d'autres pays. Dans cet exemple, la variable examinée est le revenu et les observations correspondent aux habitants du pays.

Lorsque les observations correspondent à des catégories et que la variable examinée est le nombre d'individus (d'objets) d'une population donnée qui se trouve dans chaque catégorie, alors une mesure d'inégalité est aussi une mesure de concentration. Par exemple, si l'on considère la distribution de la population humaine entre les régions d'un pays, une mesure d'inégalité indique à quel point la population du pays est concentrée.

En sciences sociales, on s'est intéressé à la mesure d'inégalité dans plusieurs contextes différents : inégalité dans la distribution du revenu, concentration des parts de marché (mesure inverse du degré de concurrence), concentration spatiale des populations ou des activités économiques, etc.

La construction de mesures d'inégalité ou de concentration pose un problème analogue à celui de la multidimensionnalité dans la définition de nombres indices : il s'agit de résumer en un seul chiffre une caractéristique possédée par l'ensemble des valeurs que prend une variable. On peut donc s'attendre à ce qu'il n'y ait pas de solution unique.

En général, une mesure de l'inégalité compare la distribution observée avec une distribution de référence, qui représente l'égalité parfaite. Souvent, la distribution de référence reste implicite. Mais il est parfois nécessaire de l'expliciter. Par exemple, s'agissant de la répartition spatiale d'une population entre des régions, une concentration nulle correspond-elle à la situation où le nombre d'habitants est le même dans toutes les régions ? Ou correspond-elle plutôt à la situation où le nombre d'habitants est proportionnel à la superficie des régions ? Ou encore, à la superficie habitable ?

Quelles sont les propriétés désirables d'une mesure d'inégalité ? Valeyre (1993) propose les six propriétés suivantes :

1. Une mesure d'inégalité doit prendre des valeurs non négatives, puisqu'il s'agit d'une mesure de l'éloignement de la distribution observée par rapport à la distribution de référence.
2. Une mesure d'inégalité doit prendre la valeur zéro si, et seulement si, la distribution observée est identique à la distribution de référence.
3. Toutes les observations doivent être traitées de la même manière.
4. Une mesure d'inégalité doit être indépendante de la valeur moyenne de la variable examinée ; une mesure de concentration doit être indépendante de la taille de la population dont on étudie la distribution.
5. L'agrégation d'observations affichant le même degré de spécificité ne doit pas changer la valeur de la mesure [1].
6. Principe de transfert de Pigou-Dalton : une mesure d'inégalité doit diminuer si la distribution est modifiée d'une façon qui réduit incontestablement l'inégalité [2].

Ces principes permettent d'évaluer la validité des différentes mesures d'inégalité qui se proposent. Ainsi, l'écart-type ou la variance ne possèdent pas la propriété 4. Par contre, le coefficient de variation possède les six propriétés énoncées : correctement utilisé, il constitue donc une bonne mesure d'inégalité.

Voyons maintenant quelques autres exemples de mesures d'inégalité ou de concentration.

---

[1] La spécificité réfère au rapport entre une valeur observée de la variable étudiée et la valeur correspondante dans la distribution de référence. Par exemple, les quotients de localisation sont des indicateurs de spécificité. Une mesure de la concentration géographique de l'emploi d'une branche d'activité donnée ne devrait pas être affectée si l'on agrège deux régions dont les quotients de localisation sont égaux.

[2] Techniquement, cela se traduit par la condition suivante : si la valeur de la variable diminue pour une observation *i* et augmente d'un même montant pour une autre observation *j*, et si le degré de spécificité de l'observation *i* est supérieur à celui de l'observation *j*, alors la mesure d'inégalité doit diminuer.

1-4.1 Le coefficient de concentration de l'économie industrielle

Cette mesure est surtout utilisée en économie industrielle, mais elle a aussi été utilisée pour mesurer le degré de concentration de la distribution par taille des villes. C'est tout simplement la somme des parts des *n* plus grandes entités. Par exemple, Rosen et Resnick (1980) mesurent le degré de concentration d'une hiérarchie urbaine au moyen la fraction de la population urbaine totale qui se trouve dans les trois plus grandes villes. En économie industrielle, on mesure souvent la concentration de marché au moyen de la somme des parts des quatre plus grandes entreprises.

Cette mesure a l'avantage de ne pas être très exigeante en termes de données, mais il lui manque la plupart des propriétés désirables : elle ne possède que la première et la quatrième.

1-4.2 L'indice de concentration de Hirschman-Herfindahl

Cet indice est simplement par la somme des carrés des parts. Par exemple, pour mesurer le degré de concentration dans un système urbain qui comporte *n* villes, on peut calculer

$$H = \sum_{i=1}^{n} s_i^2$$

où s_i est la fraction de la population urbaine totale qui se trouve dans la ville *i*. L'indice *H* varie entre $1/n$ et 1 : il prend la valeur $1/n$ quand toutes les villes sont de taille égale, et la valeur 1 dans le cas où toute la population urbaine est concentrée dans une seule ville. On interprète parfois l'indice *H* en termes de « nombre équivalent », notamment en économie industrielle : dans un marché de, disons quarante entreprises, si l'indice *H* a une valeur de *x*, on dit que le degré de concentration « équivaut » à celle d'un marché de $1/x$ firmes ayant des parts de marché égales.

L'indice de Hirschman-Herfindahl ne possède pas les propriétés 2 et 5. En outre, il dépend du nombre d'observations *n*. Il est à noter enfin que l'indice *H* est très étroitement lié à la variance des parts : celle-ci est en effet égale à

$$\frac{1}{n}\sum_{i=1}^{n}\left(s_i - \frac{1}{n}\right)^2 = \frac{H}{n} - \frac{1}{n^2}$$

1-4.3 La courbe de Lorenz et l'indice de concentration de Gini

LA DIFFÉRENCE MOYENNE DE GINI

L'indice de concentration de Gini est ainsi nommé en l'honneur du statisticien italien Corrado Gini (1884-1965). Il mesure l'inégalité au moyen des différences entre toutes les paires d'observations (y_j, y_k). La somme pondérée des différences s'appelle la « différence moyenne de Gini » et elle se calcule, pour des données groupées, selon la formule suivante [3] :

$$\Delta = \frac{1}{N^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \left| y_j - y_k \right| f_j \, f_k$$

où

n est le nombre de valeurs distinctes observées

où f_j est la fréquence de la valeur y_j dans la distribution, de sorte que

$N = \sum_{j=1}^{n} f_j$ est le nombre d'observations

Par exemple, s'agissant de mesurer l'inégalité de la distribution du revenu au Québec, f_j serait le nombre de personnes qui ont un revenu de y_j ; N est le nombre de personnes dans la population.

Lorsque les observations sont groupées par classes, la valeur y_j est la valeur *moyenne* de la variable Y dans la classe j (et non pas le point milieu de l'intervalle de revenu de la classe j).

Écrivons

$v_j = \frac{f_j}{N}$, la fraction de la population appartenant à la classe j.

La valeur moyenne de la variable Y s'écrit alors

$$\mu = \frac{1}{N} \sum_{j=1}^{n} f_j \, y_j = \sum_{j=1}^{n} v_j y_j$$

---

3 Dans cette formule chaque observation est comparée à chacune des observations, y compris à elle-même ; c'est la différence moyenne *avec* répétition. Kendall et Stuart (1991, p. 58) donnent aussi la formule *sans* répétition. Lorsque N est grand, la différence est négligeable.

Soit

$$M = \sum_{j=1}^{n} f_j y_j$$, la somme des valeurs de la variable *Y*, et

$$w_j = \frac{f_j y_j}{\sum_{k=1}^{n} f_k y_k} = \frac{f_j y_j}{N\mu} = \frac{v_j y_j}{\mu}$$, la fraction de la somme allouée à la classe *j*.

Rangeons ensuite les observations, en vue de la construction d'une courbe de Lorenz (voir ci-après), par ordre croissant des rapports w_j / v_j . Complétons la notation en posant

$$Cw_j = \sum_{k=1}^{j} w_k$$

Cw_j est la fraction cumulée des classes de 1 à *j*.

En développant la formule de calcul de la différence moyenne de Gini, on obtient :

$$\Delta = 2\mu \left(1 - \sum_{j=1}^{n} v_j \; Cw_j \quad - \sum_{j=1}^{n} v_j \; Cw_{j-1} \right)$$

Cela est démontré à l'annexe 1-F.

CALCUL DE L'INDICE DE CONCENTRATION DE GINI

L'indice de concentration de Gini est simplement le rapport de la différence moyenne de Gini sur deux fois la moyenne :

$$G = \frac{\Delta}{2\mu} = 1 - \left(\sum_{j=1}^{n} v_j \; Cw_j \quad + \sum_{j=1}^{n} v_j \; Cw_{j-1} \right) = 1 - \sum_{j=1}^{n} v_j \left(Cw_j + Cw_{j-1} \right)$$

Arriaga (1975, p. 65-71), comme aussi plusieurs géographes, définit le coefficient Gini comme

$$G = \sum_{i=2}^{n} Cw_i \; Cv_{i-1} - \sum_{i=2}^{n} Cw_{i-1} \, Cv_i$$

où $Cv_j = \sum_{k=1}^{j} v_k$

Cette formule peut se déduire de la précédente.

$$G = 1 - \sum_{j=1}^{n} v_j \left(Cw_j + Cw_{j-1}\right)$$

$$G = 1 - \sum_{j=1}^{n} \left(Cv_j - Cv_{j-1}\right)\left(Cw_j + Cw_{j-1}\right)$$

$$G = 1 - \sum_{j=1}^{n} Cv_j\, Cw_j - \sum_{j=1}^{n} Cv_j\, Cw_{j-1} + \sum_{j=1}^{n} Cv_{j-1} Cw_j + \sum_{j=1}^{n} Cv_{j-1} Cw_{j-1}$$

où

$$\sum_{j=1}^{n} Cv_{j-1} Cw_{j-1} = \sum_{j=0}^{n-1} Cv_j\, Cw_j \text{ et } Cv_0 = Cw_0 = 0\text{, de sorte que}$$

$$- \sum_{j=1}^{n} Cv_j\, Cw_j + \sum_{j=1}^{n} Cv_{j-1} Cw_{j-1} = - \sum_{j=1}^{n} Cv_j\, Cw_j + \sum_{j=1}^{n-1} Cv_j\, Cw_j$$

$$= - Cv_n\, Cw_n = -1$$

Et

$$G = 1 - 1 - \sum_{j=1}^{n} Cv_j\, Cw_{j-1} + \sum_{j=1}^{n} Cv_{j-1} Cw_j$$

ce qui, puisque $Cv_0 = Cw_0 = 0$, équivaut à

$$G = - \sum_{j=2}^{n} Cv_j\, Cw_{j-1} + \sum_{j=2}^{n} Cv_{j-1} Cw_j$$

$$G = \sum_{j=2}^{n} Cw_j\, Cv_{j-1} - \sum_{j=2}^{n} Cw_{j-1} Cv_j$$

Lorsque les observations sont groupées, il suffit donc, pour calculer l'indice de concentration de Gini, de connaître la répartition entre les classes de la population (les v_j) et de la somme des valeurs de la variable Y (les w_j).

Le plus souvent, la population (ou les ménages) sont d'abord rangés en ordre croissant de revenu ; on définit ensuite des catégories de tailles égales : quartiles, quintiles, déciles, etc. On dira ainsi : « Les 20 % de la population avec les revenus les plus élevés (le quintile supérieur) accaparent *xx* % du revenu global, tandis que les 20 % avec les revenus les moins élevés (le quintile inférieur) n'en reçoivent que *zz* % ». De tels énoncés donnent aussi une mesure de la concentration, mais, contrairement au coefficient de Gini, ce sont des mesures partielles, qui ne tiennent compte que d'une partie de la distribution.

Voyons, par exemple, la répartition du revenu (*Y*) entre les familles au Canada en 1995 (la population considérée est donc celle des familles, et non des individus). Statistique Canada a récemment diffusé les données suivantes, tirées du Recensement de la population de 1996 [4].

4 *Le Quotidien*, 3 mars 1999. Il est à noter que les données du Recensement de 1996 sur les revenus annuels se rapportent à l'année précédente.

Tableau: Limites supérieures (en $ de 1995) des déciles du revenu familial et répartition du revenu global familial par décile, 1995

Décile	Limite supérieure	Part du revenu global (%)
Premier	15158	1,45
Deuxième	23184	3,55
Troisième	31097	4,96
Quatrième	38988	6,42
Cinquième	46951	7,86
Sixième	55355	9,37
Septième	64997	10,91
Huitième	77501	13,11
Neuvième	98253	15,85
Dixième		26,53

Les données de ce tableau peuvent être présentées autrement, comme ceci :

Classe de revenus ($ de 1995)	Fraction du nombre de familles (%)	Part du revenu global (%)
0-15158	10,00	1,45
15159-23184	10,00	3,55
23185-31097	10,00	4,96
31098-38988	10,00	6,42
38989-46951	10,00	7,86
46952-55355	10,00	9,37
55356-64997	10,00	10,91
64998-77501	10,00	13,11
77502-98253	10,00	15,85
98254 et plus	10,00	26,53

Dans le tableau qui précède, les w_j sont les parts du revenu global ; les v_j sont tous égaux à 10 %. À partir de ce tableau, on peut effectuer les calculs préliminaires comme dans le tableau suivant.

Classe de revenus ($ de 1995)	Fraction du nombre de familles (%)	Part du revenu global (%)			
	v_j	w_j	Cw_j	$v_j\,Cw_j$	$v_j\,Cw_{j-1}$
0-15158	10,00	1,45	0,0145	0,0015	0,0000
15159-23184	10,00	3,55	0,0500	0,0050	0,0015
23185-31097	10,00	4,96	0,0996	0,0100	0,0050
31098-38988	10,00	6,42	0,1638	0,0164	0,0100
38989-46951	10,00	7,86	0,2424	0,0242	0,0164
46952-55355	10,00	9,37	0,3361	0,0336	0,0242
55356-64997	10,00	10,91	0,4452	0,0445	0,0336
64998-77501	10,00	13,11	0,5763	0,0576	0,0445
77502-98253	10,00	15,85	0,7348	0,0735	0,0576
98254 et plus	10,00	26,53	1,0001	0,1000	0,0735
Total	100,00	100,00		0,3663	0,2663

L'indice de concentration de Gini du revenu familial au Canada par déciles, en 1995, est donc égal à :

$$G = 1 - (0{,}3663 + 0{,}2663) = 0{,}3675$$

Deux remarques s'imposent ici :

- Les données utilisées ici étaient d'emblée rangées par ordre croissant des rapports w_j / v_j .

  Ce n'est pas toujours le cas ! En général, avant de calculer l'indice de Gini, il faut préalablement ranger les données dans le bon ordre (voir l'exemple tiré de Taylor, 1977, ci-après).
- Avec des données groupées, l'indice de concentration de Gini dépend du groupement ou schème de classement utilisé. Si la population des familles avait été groupée par quintiles, ou par centiles, le résultat du calcul aurait été différent. Nous reviendrons sur ce point.

LA COURBE DE LORENZ

La courbe de Lorenz est un instrument de comparaison graphique entre deux distributions. Rappelons que

$$Cv_j = \sum_{k=1}^{j} v_k = \text{fraction cumulée de } X \text{ (par exemple, ci-haut, des familles)}$$

$$Cw_j = \sum_{k=1}^{j} w_k = \text{fraction cumulée de } Y \text{ (par exemple, ci-haut, des revenus)}$$

On a naturellement :

$Cv_n = Cw_n = 1$

Méthode de construction de la courbe de Lorenz (Voir l'exemple numérique ci-après, tiré de Taylor, 1977, p. 179) :

1. Calculer les rapports $\frac{w_i}{v_i}$ [5].
2. Réordonner les catégories en ordre croissant de $\frac{w_i}{v_i}$: $\frac{w_1}{v_1} < \frac{w_2}{v_2} < \cdots < \frac{w_n}{v_n}$
3. Calculer les fractions cumulatives Cv_i et Cw_i
4. La courbe de Lorenz est l'ensemble des points (Cv_i, Cw_i), où les Cv_i sont repérés sur l'axe horizontal.

La courbe de Lorenz a les propriétés suivantes :

1. $Cv_0 = Cw_0 = 0$ (par définition de Cv_i et de Cw_i) : la courbe part de l'origine ;
2. $Cv_n = Cw_n = 1$ (par définition de Cv_i et de Cw_i) : la courbe aboutit au point de coordonnées [1,1] (ou [100 %, 100 %]) ;
3. Lorsque les deux distributions sont identiques, on a, pour tout i,
   $Cv_i = Cw_i$
   c'est-à-dire que la courbe de Lorenz coïncide avec la diagonale.
4. $Cv_i \geq Cw_i$ pour i différent de 0 et de n (par construction, étant donné le réordonnancement des catégories) : la courbe se situe sous la diagonale ou coïncide avec elle ;
5. La pente de chaque segment de la courbe de Lorenz est égale à la valeur l'indicateur de spécificité associé à l'observation correspondante :
   $$\text{pente du segment } i = \frac{Cw_i - Cw_{i-1}}{Cv_i - Cv_{i-1}} = \frac{w_i}{v_i}$$

---

5 Ces rapports ne sont autre que les *spécificités* associées aux observations.

6. La courbe de Lorenz est concave vers le haut, c'est-à-dire que chaque segment a une pente plus abrupte que le précédent : cela découle de 5, puisque, par construction, $\frac{w_i}{v_i} < \frac{w_{i+1}}{v_{i+1}}$

CONSTRUCTION D'UNE COURBE DE LORENZ (EXEMPLE NUMÉRIQUE TIRÉ DE TAYLOR, 1977, P. 179)

Première étape : calcul des w_i / v_i

Zone	x_i Nombre de ménages de classe moyenne	v_i Distrib. de x	y_i Nombre de votes du parti Républi-cain	w_i Distrib. de y	w_i/v_i
A	30	0,25	30	0,30	1,20
B	20	0,17	15	0,15	0,90
C	10	0,08	8	0,08	0,96
D	10	0,08	5	0,05	0,60
E	20	0,17	19	0,19	1,14
F	30	0,25	23	0,23	0,92
Total	**120**	**1,00**	**100**	**1,00**	

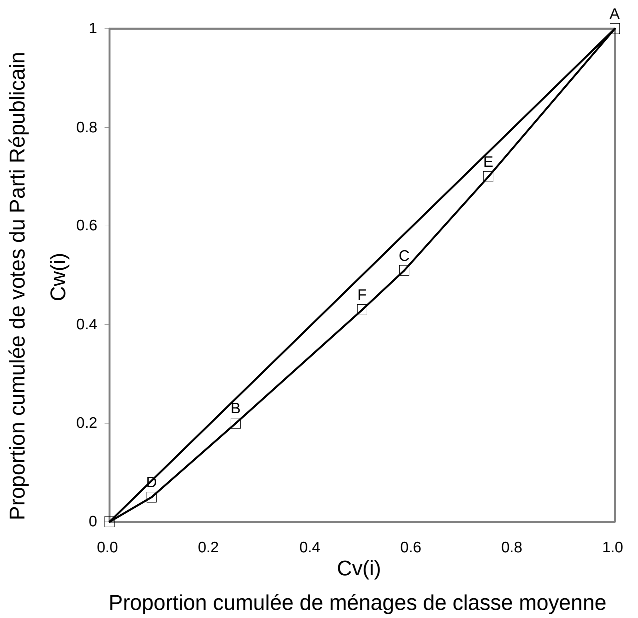
Deuxième étape : tri par ordre croissant des w_i/v_i

Troisième étape : calcul des Cv_i (abscisses) et des Cw_i (ordonnées)

Zone	x_i	v_i	y_i	w_i	w_i/v_i	Cv_i Abscisse	Cw_i Ordonnée	Écart $(Cv_i–Cw_i)$	Écart $\lvert v_i–w_i\rvert$
						0,00	0,00		
D	10	0,08	5	0,05	0,60	0,08	0,05	0,033	0,033
B	20	0,17	15	0,15	0,90	0,25	0,20	0,050	0,017
F	30	0,25	23	0,23	0,92	0,50	0,43	0,070	0,020
C	10	0,08	8	0,08	0,96	0,58	0,51	***0,073***	0,003
E	20	0,17	19	0,19	1,14	0,75	0,70	0,050	0,023
A	30	0,25	30	0,30	1,20	1,00	1,00	0,000	0,050
Total	**120**	**1,00**	**100**	**1,00**					**0,147**

Note : on peut constater que l'écart maximum entre la courbe de Lorenz et la diagonale est égal à $\frac{1}{2}\sum_i \lvert v_i - w_i \rvert$.

Courbe de Lorenz



Quatrième étape : calcul de l'indice de concentration de Gini

Zone	x_i	v_i	y_i	w_i	w_i/v_i	Cv_i	Cw_i	$v_i\,Cw_i$	$v_i\,Cw_{i-1}$
						Abscisse	Ordonnée		
						0,00	0,00		
D	10	0,08	5	0,05	0,60	0,08	0,05	0,004	0,000
B	20	0,17	15	0,15	0,90	0,25	0,20	0,033	0,008
F	30	0,25	23	0,23	0,92	0,50	0,43	0,108	0,050
C	10	0,08	8	0,08	0,96	0,58	0,51	0,043	0,036
E	20	0,17	19	0,19	1,14	0,75	0,70	0,117	0,085
A	30	0,25	30	0,30	1,20	1,00	1,00	0,250	0,175
Total	**120**	**1,00**	**100**	**1,00**				**0,554**	**0,354**

$G = 1 - (0{,}554 + 0{,}354) = 0{,}092$

CALCUL GÉOMÉTRIQUE DE L'INDICE DE GINI AU MOYEN DE LA COURBE DE LORENZ

Ce fut une réalisation remarquable de Corrado Gini que de démontrer, en 1914, que l'indice de concentration qui porte son nom est égal au rapport entre (1) la superficie comprise entre la diagonale et la courbe de Lorenz et (2) la superficie totale sous la diagonale :

$$G = \frac{\text{Superficie comprise entre la diagonale et la courbe de Lorenz}}{\text{Superficie totale sous la diagonale}}$$

La superficie totale du triangle sous la diagonale est donnée par

$$\frac{Cw_n \times Cv_n}{2} = \frac{1 \times 1}{2} = \frac{1}{2}$$

La superficie comprise entre la diagonale et la courbe de Lorenz est calculée comme la différence entre la superficie totale du triangle sous la diagonale (=$\frac{1}{2}$) et la superficie sous la courbe de Lorenz. La superficie sous la courbe de Lorenz (voir l'exemple numérique précédent et la figure ci-après) est la somme de *n* trapèzes dont chacun a une surface égale à

$$\frac{1}{2} v_i \left(Cw_i + Cw_{i-1} \right)$$

La superficie sous la courbe de Lorenz est donc la somme de ces *n* surfaces :

$$\frac{1}{2} \sum_{i=1}^{n} v_i \left(Cw_i + Cw_{i-1} \right)$$

Et le coefficient Gini est donné par

$$G = \frac{\left(\frac{1}{2}\right) - \left(\frac{1}{2} \sum_{i=1}^{n} v_i \left(Cw_i + Cw_{i-1} \right)\right)}{\left(\frac{1}{2}\right)} = 1 - \sum_{i=1}^{n} v_i \left(Cw_i + Cw_{i-1} \right) = \frac{\Delta}{2\mu}$$

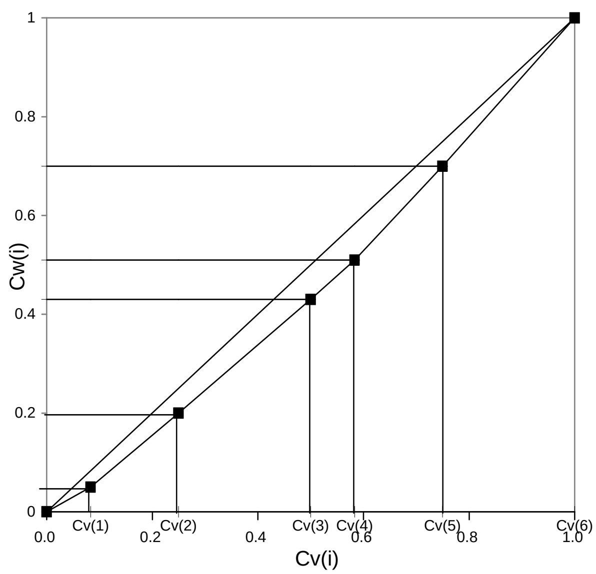
ce qui correspond bien à la formule énoncée précédemment.

Pour faciliter l'interprétation de la courbe de Lorenz et de l'indice de Gini qui lui est associé, il est utile de se rappeler que c'est la distribution *V* qui joue le rôle de distribution de référence (i.e. d'égalité parfaite ou de concentration nulle). Dans la courbe de Lorenz, les Cv_i sont repérés sur l'axe horizontal et les Cw_i, sur l'axe vertical.

Exemples :

- Si *V* est la répartition du territoire entre les zones et *W*, la répartition de la population, le coefficient Gini est une mesure de la concentration géographique de la population.
- Si *V* est une distribution de la population (ou des ménages) en *n* catégories et *W*, la distribution du revenu agrégé par catégorie, alors le coefficient Gini est une mesure de la concentration du revenu.

Calcul géométrique de l'indice de concentration de Gini



PROPRIÉTÉS DE L'INDICE DE CONCENTRATION DE GINI

L'indice de concentration de Gini possède les six propriétés que doit avoir une mesure d'inégalité, telles qu'énoncées au début de ce chapitre. Il a en outre les propriétés suivantes :

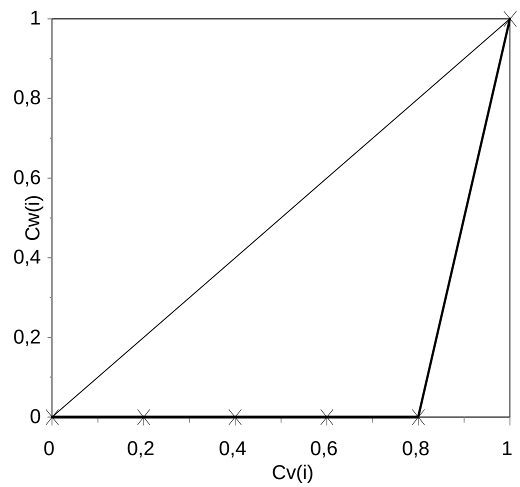
1. L'indice de Gini varie entre zéro et 1 [6]. La valeur minimum du coefficient Gini, zéro, est atteinte lorsque les deux distributions sont identiques. Sa valeur maximum théorique, 1, est atteinte lorsque la courbe de Lorenz suit la base et le côté droit de la « boîte » ; mais pour

[6] Ou entre 0 % et 100 % lorsqu'on l'exprime en pourcentage.

que ce maximum théorique soit atteint, il faut que le nombre de catégories tende vers l'infini, de sorte que v_n tende vers zéro [7].

2. On peut démontrer que l'indice de Gini est symétrique, c'est-à-dire que les rôles des deux distributions sont interchangeables ; en d'autres mots, si l'on intervertit les rôles, la valeur du coefficient Gini est inchangée.
3. Lorsque les données sont groupées, l'indice de Gini est sensible à la définition et au nombre des catégories utilisées (classes, zones).
4. Lorsqu'on l'utilise comme mesure de concentration spatiale, le Gini ne tient aucun compte de la proximité dans l'espace des différentes zones de forte densité (l'espace est traité comme un puzzle défait).

À propos de la valeur maximum que peut atteindre le coefficient de Gini lorsque le nombre de catégories n'est pas infini, précisons qu'elle est égale à $1 - v_n$. Cette propriété est illustrée dans la figure suivante.



Dans cet exemple, le lecteur peut facilement vérifier, en appliquant la méthode de calcul géométrique, que $v_n = 0{,}2$ et $G = (1 - 0{,}2) = 0{,}8$.

La troisième propriété mérite une attention plus poussée. Elle se manifeste notamment par ceci : l'agrégation de deux ou plusieurs catégories a *toujours* pour effet de réduire la valeur calculée du coefficient Gini (à moins que les deux catégories n'aient la même spécificité, auquel

7 Autrement, lorsque $v_n > 0$, la valeur maximum de G est égale à $1 - v_n$.

cas la propriété 5 des mesures d'inégalité se manifeste). Cela se vérifie aisément si l'on pense au calcul géométrique fait au moyen de la courbe de Lorenz : l'agrégation de deux catégories voisines réduit l'espace compris entre la courbe de Lorenz et la diagonale. Cela est également conforme à l'intuition que l'agrégation de catégories a pour effet de gommer une partie des différences.

Cette sensibilité du Gini à la définition des catégories peut sérieusement compromettre sa fiabilité comme mesure de la concentration, notamment lorsque les catégories sont de tailles inégales. Pour illustrer ce phénomène, imaginons que l'on veuille comparer la concentration de la population à deux moments dans le temps, sur un territoire découpé en trois zones de même superficie (disons égale à 1) :

	Superf.	Population		Densité	
		au temps 0	au temps *t*	au temps 0	au temps *t*
Zone 1	1	10	80	10	80
Zone 2	1	80	10	80	10
Zone 3	1	10	10	10	10

Il est évident dans cet exemple que la concentration est restée la même à l'échelle considérée ($G = 0{,}47$), même si le centre de gravité de la population s'est déplacé vers la Zone 1. Supposons maintenant que l'on ait agrégé les Zones 2 et 3 :

	Superf.	Population		Densité	
		au temps 0	au temps *t*	au temps 0	au temps *t*
Zone 1	1	10	80	10	80
Zones 2 et 3	2	90	20	45	10

Les données agrégées donnent l'illusion que la concentration a augmenté, puisqu'on a $G = 0{,}23$ au temps 0 et $G = 0{,}47$ au temps *t* (noter que l'indice de Gini calculé est plus faible avec les données agrégées au temps 0, mais qu'il est le même au temps *t*, puisque dans ce dernier cas, les zones agrégées sont de même densité, c'est-à-dire de même spécificité).

1-4.4 Pour conclure à propos de la mesure de l'inégalité...

Nous n'avons évoqué ici que quelques-unes des multiples mesures d'inégalités qui sont maintenant proposés. Parmi les mesures que nous avons laissées de côté, mentionnons celles qui sont des mesures d'entropie, comme la mesure de Shanon, ou la mesure du gain d'information de Kullback-Leibler (aussi associée au nom de Theil). Le lecteur intéressé pourra consulter le survol de Valeyre (1993).

Rappelons enfin que les mesures d'inégalité sont des mesures d'éloignement d'une distribution observée par rapport à une distribution de référence. Elles sont en cela étroitement apparentées aux mesures de dissimilarité, qui comparent deux distributions, dont les rôles sont toutefois symétriques (aucune des deux ne joue le rôle de référence).