MATHÉMATIQUES ET SCIENCES HUMAINES

F. BOURGUIGNON

Note sur les propriétés de décomposabilité des mesures d'inégalité

Mathématiques et sciences humaines, tome 93 (1986), p. 41-52

<http://www.numdam.org/item?id=MSH_1986__93__41_0>

Math. Sci. hum.(*23*^{e}année, n°93, *1986*, p.41-52)

NOTE SUR LES PROPRIETES DE DECOMPOSABILITE DES MESURES D'INEGALITE[☆]

F. BOURGUIGNON

Cette note présente les principaux résultats existants concernant la décomposition des mesures d'inégalité et en donne quelques exemples d'application dans le cas du coefficient de Theil.

AXIOMES ELEMENTAIRES CONCERNANT LA MESURE DE L'INEGALITE

Soit $Y = (y_1, y_2, \ldots y_n)$ une "distribution" de revenus dans une population de n individus. On notera $I(Y)$ ou $I_n(y_1, y_2, \ldots y_n)$ une mesure de l'inégalité de cette distribution, c'est-à-dire une application $\mathbb{R}^n \rightarrow R$.

On impose généralement aux mesures $I_n(\ )$ de satisfaire un certain nombre d'axiomes élémentaires.

a/ Anonymat ou symétrie

$I_n(y_1, y_2, \ldots y_n)$ est invariante à toute permutation de ses arguments. Par exemple, $I_n(y_1, y_2, y_3, \ldots y_n) = I_n(y_2, y_1, y_3, \ldots y_n)$. Implicitement, ceci suppose que les individus ont les mêmes "besoins" ou tirent la même "utilité" d'un revenu identique.

b/ Invariance par rapport à l'échelle des revenus

L'inégalité n'est pas modifiée si tous les revenus individuels sont multipliés par le même nombre. En d'autres termes, $I_n(\ )$ est supposé homogène de degré zéro :

☆ Cette note résume les principaux résultats de certains travaux. L'auteur dirige le Centre d'Economie Quantitative et Comparative de l'E.H.E.S.S., Paris.

$$I_n(\lambda y_1, \lambda y_2, \dots \lambda y_n) = I_n(y_1, y_2, \dots y_n) \; \forall \lambda \geq 0$$

Cet axiome est discutable si l'on considère que l'aversion d'une société pour l'inégalité peut être fonction de son niveau moyen de revenu ou de bien-être (voir S.-C. Kolm).

c/ Invariance par rapport à l'échelle de la population

L'inégalité associée à une distribution ne se modifie pas si chaque individu est "dupliqué" un nombre donné de fois. En d'autres termes, l'inégalité en France est la même que dans un pays imaginaire qui serait la réplique à l'échelle 2, 3, ..., n, de la population française. Comme pour l'axiome précédent, ceci revient à supposer que la mesure de l'inégalité est indépendante de la taille de la population.

d/ Transferts égalisateurs

Tout transfert de revenus d'un individu i à un individu j moins riche diminue (ou n'accroît pas) la mesure de l'inégalité.

$$I_n(y_1, \dots y_i - h, \dots y_j + h, \dots y_n) \leqslant I_n(y_1, \dots y_i, \dots y_j, \dots y_n) \; \forall \; y_i > y_j$$

PROPRIETES DE DECOMPOSABILITE

Supposons la population partitionnée en m groupes k (=1 ... m) d'effectifs n_k, et dénotons y_{ki} le revenu du i-ème individu (i=1 ... n_k) du groupe k. On dira qu'une mesure $I_n(\;)$ est "agrégative" si elle peut s'exprimer sous la forme :

$$I_n(y_{11}, y_{12}, \dots, y_{mn_m}) =$$

$$F^m \{I_{n_1}(y_{11}, y_{12}, \dots, y_{1n_1}), \dots I_{n_m}(y_{m1}, \dots, y_{mn_m}) \; ;$$

$$\bar{y}_1, \dots, \bar{y}_m \; ; n_1, \dots, n_m \}$$

avec $\bar{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{ki}$, pour toute partition (m ; $n_1, \dots, n_m$) de la population.

En d'autres termes, l'inégalité totale (I_n) s'exprime comme une fonction des inégalités <u>internes</u> aux groupes k(I_{n1}, ..., I_{nm}) et des caractéristiques <u>agrégées</u> (revenus moyens, effectifs) de ces groupes. On n'a pas besoin de connaître les distributions propres à chaque groupe pour calculer l'inégalité totale, mais seulement l'<u>inégalité</u> pour chaque groupe.

Si la fonction F^m est <u>additivement séparable</u> par rapport aux inégalités intra-groupes I_{n_1}, ..., I_{n_m}, on dira que la mesure $I_n(\ )$ est <u>décomposable</u>. En d'autres termes, l'inégalité totale est la somme des inégalités intra-groupes, pondérées par des fonctions ne dépendant que des revenus moyens et des effectifs, et d'un terme dépendant également de ces seules caractéristiques agrégées. Plus précisément, on montre que, dans ce cas, $F^m(\ )$ prend la forme :

$$I_n(\ ) = \sum_{k=1}^{m} G(\bar{y}_k,\ n_k\ ;\bar{y},\ n) I_{n_k}(\ ) + I_n(\bar{y}_1,\ldots,\bar{y}_1,\ \bar{y}_2,\ \ldots,\bar{y}_m,\ldots,\bar{y}_m)$$

Le dernier terme correspondant à l'inégalité de la distribution hypothétique où tous les individus d'un groupe donné auraient le même revenu. Les mesures décomposables différeront essentiellement par la fonction G( ) qui pondère les inégalités intra-groupes.

LA FAMILLE DES MESURES DECOMPOSABLES

Il a été montré que les seules mesures décomposables satisfaisant les axiomes indiqués plus haut et étant continument différentiables sont les suivantes, (F. Bourguignon (1979), A. Shorrocks (1980)) :

$$I_n(\ ) = \frac{1/n}{(c-1)} \sum_i \left[\left(\frac{y_i}{\bar{y}}\right)^c - 1\right] \quad c \neq 0,\ 1$$

$$I_n(\ ) = \frac{1}{n} \sum_i \log \frac{\bar{y}}{y_i} \quad \text{(déviation logarithmique moyenne)}$$

$$I_n(\ ) = \frac{1}{n} \sum_i \frac{y_i}{\bar{y}} \log \frac{y_i}{\bar{y}} \quad \text{(coefficient de Theil)}$$

les deux dernières mesures étant des cas particuliers de la première avec c = 0, ou c = 1 .

Dans le cas général, les coefficients de pondération des inégalités intra-groupe sont les fonctions :

$$G(\bar{y}_k,\ n_k\ ;\ \bar{y},\ n) = \frac{n_k}{n} \left(\frac{\bar{y}_k}{\bar{y}}\right)^c$$

Pour la déviation logarithmique moyenne (DLM, c = 0), on a donc la formule de décomposition :

$$DLM = \sum_k \frac{n_k}{n} DLM_k + \frac{1}{n} \sum_k n_k \log \frac{\bar{y}}{\bar{y}_k}$$

et pour le coefficient de Theil (T)

$$T = \sum_k \frac{n_k y_k}{n\bar{y}} T_k + \frac{1}{n} \sum_k \frac{n_k \bar{y}_k}{\bar{y}} \log \frac{\bar{y}_k}{\bar{y}}$$

On peut noter que dans le cas général (c $\neq$ 0, 1), la somme des coefficients de pondération n'est pas nécessairement égale à l'unité.

EXEMPLE : "UNE ANALYSE DE DECOMPOSITION DE L'INEGALITE DES REVENUS INDIVIDUELS EN FRANCE"*

Décomposition de l'indicateur de Theil et mesure de la contribution d'une variable à l'inégalité totale des revenus

L'indicateur de Theil se définit comme :

$$(1) \qquad T = \frac{1}{N} \sum_i \frac{y_i}{\bar{y}} \operatorname{Log} \frac{y_i}{\bar{y}}$$

où N est la taille de la population et $\bar{y}$ son revenu moyen. Supposons maintenant que la population est divisée en G groupes g (= 1, 2, ..., G), d'effectifs N_g et de revenu moyen $\bar{y}_g$. *La propriété de décomposabilité de l'indicateur de Theil* est que (1) est équivalent à :

$$(2) \qquad T = \sum_g \frac{N_g}{N} \frac{\bar{y}_g}{\bar{y}} T_g + \sum_g \frac{N_g}{N} \frac{\bar{y}_g}{\bar{y}} \operatorname{Log} \frac{\bar{y}_g}{\bar{y}} = T_I + T_E$$

où T_g est l'indicateur de Theil défini selon (1) pour le *seul groupe* d'individus g. Sous cette forme, on voit donc que l'inégalité totale (T) s'exprime comme la somme de deux termes :

- une moyenne pondérée de l'inégalité *interne* à chaque groupe $g(T_I)$;
- l'inégalité existant *entre* les groupes g, en supposant que les individus ont tous le même revenu, $\bar{y}_g$, à l'intérieur de chaque groupe $g(T_E)$.

Cette propriété est clairement analogue à celle, bien connue, de la variance selon laquelle "la variance totale est égale à la moyenne des variances plus la variance des moyennes". Il faut seulement noter que la "moyenne" des inégalités est dans (2) une moyenne pondérée non par la part de chaque groupe dans la population totale, comme c'est le cas pour la variance, mais par la part de chaque groupe dans le *revenu* total ($N_g \bar{y}_g / N \bar{y}$).

Examinons à présent comment la propriété de décomposabilité peut être utilisée pour mesurer la contribution d'une variable donnée à l'inégalité totale des revenus. Soit par exemple la variable âge, considérée de façon *discrète*, ce qui revient à décrire l'âge par un ensemble de variables indicatrices correspondant chacune à une tranche d'âge particulière. Considérée ainsi, la variable âge permet également d'effectuer une partition de la population et de lui appliquer la propriété de décomposabilité de l'indicateur

---

*Extrait de Revue économique 36, n° 4, juillet 1985, pp. 741-777.

de Theil. Les "groupes" g apparaissant dans (2) sont donc maintenant des groupes d'individus appartenant à une même tranche d'âge.

Comment se traduirait, dans le cadre de l'expression (2), l'affirmation "l'inégalité totale est exclusivement due à la variable âge" ? Evidemment, par le fait que les individus appartenant au même groupe d'âge ont des revenus identiques et que la première composante de (2) est nulle puisque chaque terme T_g est nul. La mesure traditionnellement utilisée de la *contribution d'une variable à l'inégalité totale* (C) *est donc le rapport de la seconde composante de* (2), *inégalité "entre" les groupes définis par la variable considérée, à l'inégalité totale* :

$$(3) \qquad C = [\sum_g \frac{N_g}{N} \frac{\bar{y}_g}{\bar{y}} \text{Log} \frac{\bar{y}_g}{\bar{y}}] \; / \; T = T_E \; / \; T.$$

Cette mesure n'est cependant pas sans ambiguïté. Dans l'exemple précédent, en effet, la contribution de la variable âge aurait pu également être définie comme la diminution de l'inégalité obtenue en *égalisant les revenus moyens des divers groupes d'âge*, sans modifier l'inégalité interne à chaque groupe. A partir de (2), on voit que cette seconde mesure, C', s'écrirait :

$$C' = [\sum_g \frac{N_g}{N} \frac{\bar{y}_g}{\bar{y}} T_g + \sum_g \frac{N_g}{N} \frac{\bar{y}_g}{\bar{y}} \text{Log} \frac{\bar{y}_g}{\bar{y}} - \sum_g \frac{N_g}{N} T_g] \; / \; T,$$

soit :

$$(4) \qquad C' = [\sum_g \frac{N_g}{N} (\frac{\bar{y}_g}{\bar{y}} - 1) T_g + \sum_g \frac{N_g}{N} \frac{\bar{y}_g}{\bar{y}} \text{Log} \frac{\bar{y}_g}{\bar{y}}] \; / \; T.$$

Par rapport à la définition précédente (3), on constate que la contribution C' tient compte non seulement du fait que l'égalisation des revenus moyens des groupes g annule l'inégalité intergroupe (T_E), mais aussi de ses effets sur la pondération de chaque groupe dans le calcul de l'inégalité interne moyenne (T_I). Initialement, cette pondération correspondait à la part du revenu d'un groupe dans le revenu total ; après égalisation, elle était égale au poids de chaque groupe dans la population totale.

On constate aisément que les contributions C et C' ne coïncident que dans le cas très particulier où l'inégalité interne est la même dans tous les groupes d'individus. Si l'on veut s'en tenir au coefficient de Theil pour mesurer l'inégalité, un choix est donc nécessaire. Dans ce qui suit, nous avons retenu la définition traditionnelle, C.

Mesure de la contribution de plusieurs variables à l'inégalité (mesurées par l'indicateur de Theil) et de leurs interactions

Au lieu de partitionner la population par rapport à une seule variable, l'âge, on peut utiliser simultanément plusieurs variables, l'âge et la CSP par exemple. Un groupe g est donc identifié par la tranche d'âge et la CSP auxquelles appartiennent tous ses membres. Dans la formule (2), le rapport T_E/T indique alors la contribution simultanée des variables âge et CSP à l'inégalité totale.

La comparaison de cette contribution simultanée et des contributions isolées de chaque de ces deux variables est intéressante. Dénotons C_x la contribution isolée [au sens de la définition (3)] de la variable X à l'inégalité totale, et T_E (X) la seconde composante de (2) lorsque la partition G de la population est effectuée par rapport à la variable X. On a donc, pour l'âge, considéré isolément :

$$C_A = T_E\ (A)\ /\ T$$

pour la CSP, considérée isolément :

$$C_{CSP} = T_E\ (CSP)\ /\ T$$

et pour les deux variables prises simultanément :

$$C_{A,\ CSP} = T_E\ (A,\ CSP)\ /\ T.$$

Le point intéressant est qu'il n'y a aucune raison a priori pour que

$$C_{A,\ CSP} = C_A + C_{CSP}.$$

Autrement dit, l'effet joint de l'âge et de la CSP sur l'inégalité n'est pas forcément la somme de leurs effets isolés. La différence, si elle existe, doit naturellement être attribuée à l'interaction existant entre ces deux variables dans la formation des inégalités, $I_{A,\ CSP}$. On a donc :

$$C_{A,\ CSP} = C_A + C_{CSP} + I_{A,\ CSP}$$

où, comme dans l'analyse de variance, $I_{A,\ CSP}$ dépend de la corrélation liant âge et CSP et de leur effet simultané sur les revenus individuels.

La contribution jointe $C_{A,\ CSP}$ peut également être obtenue par une procédure de décomposition *conditionnelle* de l'indicateur de Theil. Supposons la formule (2) appliquée aux groupes d'âge, et appelons T_a l'inégalité interne au groupe d'âge a. Cette inégalité peut à son tour être décomposée selon la formule (2) par rapport à la CSP et l'on peut en déduire la contribution de la CSP à l'inégalité du groupe d'âge a : $C_{CSP/a}$. La contribution jointe $C_{A,\ CSP}$ s'écrit alors :

$$C_{A,\ CSP} = \sum_a C_{CSP/a} \cdot \frac{T_a}{T} + C_A = C_{CSP/A} + C_A$$

et l'interaction $I_{A,\ CSP}$ définie précédemment s'écrit :

$$I_{A,\ CSP} = C_{CSP/A} - C_{CSP}$$

autrement dit, l'écart entre la contribution conditionnée par l'âge de la variable CSP et sa contribution isolée.

Ces concepts et méthodes se généralisent aisément au cas d'un nombre quelconque de variables. Nous utiliserons dans ce qui suit les trois variables : âge, CSP et sexe.

La structure de l'inégalité française en 1975

L'inégalité entre l'ensemble des actifs

Celle-ci se décompose en deux parts, presque égales, l'une expliquée par les trois variables disponibles (sexe, âge et CSP), l'autre résiduelle, qui correspond aux inégalités internes à chaque groupe, ou cellule, défini conjointement par ces trois variables.

Les contributions du sexe, de l'âge et de la CSP

Les contributions simples de l'âge et du sexe à l'inégalité totale s'avèrent assez faibles : 6,9 et 5,8 % respectivement. Ce résultat pour l'âge peut être comparé à celui de M. Sollogoub qui concerne les revenus en 1970 des ménages classés selon l'âge du chef de ménage. Celui-ci indique le Gini total, .432, et le Gini correspondant à une inégalité intra-cohorte nulle, .134. Si l'on estime la contribution de l'âge au Theil global à partir des

Tableau 1. Contributions simples et croisées des variables C.S.P., âge et sexe à l'inégalité totale.

		C.S.P.	Age	Sexe
Contributions simples [a]		44.2	6.9	5.8
Interaction simple avec : [b]				
	- C.S.P.	-	-1.9	.7
	- Age	-1.9	-	.4
	- Sexe	.7	.4	-
Contribution double avec : [b]				
	- C.S.P.	-	49.2	50.7
	- Age	49.2	-	13.1
	- Sexe	50.7	13.1	-
Interaction double avec :				
	- C.S.P. × Age	-	-	.4
	- Age × Sexe	-1.9	-	-
	- Sexe × C.S.P.	-	-2.2	-
Contribution triple avec :				
	- C.S.P. × Age	-	-	55.4
	- Age × Sexe	55.4	-	-
	- Sexe × C.S.P.	-	55.4	-
Inégalité expliquée par les trois variables :	55.4			
Inégalité résiduelle ("cellulaire") :	44.6			

a) Avec les notations de la section 1, ces pourcentages correspondent à l'inégalité entre groupes (définis par la C.S.P., l'âge ou le sexe), T_E, rapportée à l'inégalité totale, T.

b) Il s'agit de la composante d'interaction I (cf. section 1). Ainsi on passe des contributions simples C_{CSP} : 44,2 %, $C_{\hat{a}ge}$ = 6,9 % à la contribution double âge-C.S.P. par addition du terme interactif -1,9 % : $C_{\hat{a}ge,CSP}$ = 44,2 + 6,9 - 1,9 = 49,2.

données utilisées par M. Sollogoub, on obtient 9,7 %, soit des Theils respectivement de .342 et .0333.

La contribution simple de la variable CSP est beaucoup plus élevée puisqu'elle atteint près de la moitié du Theil global. Le fait qu'une telle part de l'inégalité résulte des différences de revenus moyens entre CSP retient l'attention à plusieurs titres. Elle justifie d'abord les études et les débats, nombreux en France, sur l'évolution à "court" terme (moins de 10 ans) de ces écarts de revenu. D'autre part, ces écarts représentent le facteur d'inégalité lié le plus directement aux variables économiques, en l'occurrence les dotations des agents en capital humain, en capital matériel, ..., les conditions d'équilibre sur les marchés du travail, du capital... Il faut cependant rappeler que ces écarts sont surestimés puisqu'il s'agit des revenus déclarés. En effet, comme on le sait, les revenus réels de certains indépendants, notamment ceux des exploitants agricoles dont les revenus déclarés sont les plus faibles de tous (20 % de moins que celui d'un manoeuvre), sont nettement supérieurs. Mais il nous était impossible de procéder à un redressement significatif de ces revenus. Les seules informations disponibles sont de caractère macro-économique. A ce niveau, le biais n'a qu'une incidence mineure sur la valeur du Theil total, parce qu'en relevant le revenu moyen des exploitants agricoles, on diminue la contribution de la variable CSP à l'inégalité, mais on augmente la pondération du Theil interne à cette CSP (qui est le Theil le plus élevé de tous les Theils intra CSP). Il reste vrai que ce biais conduit à surestimer la contribution de la variable CSP.

Au total, la contribution conjointe des trois variables atteint 55,4%; ce chiffre est très légèrement inférieur à la somme des trois contributions simples, ce qui indique une assez faible interaction entre les trois variables considérées. L'inégalité résiduelle (44,6 %) correspond, par ailleurs, à la somme des contributions à l'inégalité totale de la dispersion des revenus observée au sein de chaque "cellule" définie par rapport aux trois variables (par exemple, les femmes, employées et âgées de 40 à 44 ans).

La contribution de la CSP à l'inégalité dans chaque groupe d'âge, et réciproquement celle de l'âge à l'inégalité à l'intérieur de chaque CSP, varient comme en témoigne le tableau 2 (colonnes 4 et 8). Au sein des cohortes, on constate une incidence croissante de la CSP avec l'âge. Elle est faible pour les moins de 30 ans (.9/5, soit 18 %, et 2./6,7, soit 30 %), passe à

Tableau 2 . Contributions marginales de la C.S.P. et de l'âge à l'inégalité totale (Hommes + Femmes)

C.S.P.	Inégalité interne aux C.S.P.	Contribution à l'inégalité totale (%)	Dont : due à l'âge (%)	Tranches d'âge	Inégalité interne aux tranches d'âge	Contribution à l'inégalité totale (%)	Dont : due à C.S.P.
Exploitants agricoles	.722	5.7	.2	< 25 ans	.162	5.0	.9
Professions libérales	.226	2.3	.1	25 à 29 ans	.168	6.7	2.0
Ouvriers qualifiés	.096	3.6	.3	30 à 34 ans	.218	7.7	3.1
Ouvriers spécialisés	.113	3.5	.3	35 à 39 ans	.310	10.9	4.4
Manoeuvres	.132	1.2	.1	40 à 44 ans	.329	13.3	7.2
Industriels et gros commerçants	.350	4.0	.2	45 à 49 ans	.372	15.4	8.0
Artisans et petits commerçants	.407	11.0	.4	50 à 54 ans	.424	16.5	8.1
Ingénieurs et cadres supérieurs	.143	5.5	.7	55 à 60 ans	.432	6.9	3.5
Instituteurs et enseignants	.171	4.0	1.2	> 60 ans	.565	10.7	5.3
Employés	.145	7.3	.6				
Salariés agricoles et marins-pêcheurs	.259	.8	.0				
Contremaîtres, techniciens, cadres moyens	.145	6.9	.9				
Total	.337	55.8	5.0			93.1	41.3
Contribution isolée de la C.S.P.		44.2	1.9[a)]				
Contribution isolée de l'âge						6.9	2.9[a)]
Contribution jointe des 2 variables	49.2						

a) terme d'interaction C.S.P.-âge.

40 % pour les 30 à 40 ans et dépasse 50 % pour les 40-50 ans. Ce phénomène résulte de la forme des profils de revenu : ils sont d'autant plus pentus, d'habitude, que le revenu moyen de la CSP est élevé ; il en résulte une contribution croissante avec l'âge des différences de revenu par CSP à l'inégalité totale dans chaque cohorte. La contribution de l'âge à l'inégalité intra-CSP varie également : elle est très faible pour les indépendants et les salariés qualifiés (13 %), pour les contremaîtres, techniciens, cadres moyens et supérieurs, le maximum étant atteint parmi les instituteurs et professeurs (29 %), ce qui s'explique aisément par les caractéristiques des profils de traitement pour chaque catégorie d'enseignants.

BIBLIOGRAPHIE

BOURGUIGNON, F., "Decomposable income inequality measures", Econometrica, 47 , 4 (1979), 901-920.

BOURGUIGNON, F. et MORRISSON, C., "Une analyse de décomposition de l'inégalité des revenus individuels en France", Revue Economique, 36, 4 (1985), 741-777.

KOLM, S.-C., "Unequal inequalities, I and II", Journal of Economic Theory, 12-13 (1976), 416-442, 82-111.

SHORROCKS, A., "The class of additively decomposable inequality measurement", Econometrica, 48, 3 (1980), 613-625.

SOLLOGOUB, M., "La comparaison de l'inégalité dans la répartition personnelle des revenus : note sur l'étude de l'OCDE et sur les cas français et américains (1962-1970)", Revue d'économie politique, 90, 3 (1980),