
PRISE EN COMPTE DU CARACTÈRE ORDONNÉ DES CATÉGORIES SOCIALES DANS LA MESURE DE LA SÉGRÉGATION

Fabrice MURAT (*)

(*) DEPP, Sous-direction des évaluations et de la performance

fabrice.education@education.gouv.fr

Mots-clés : Inégalités, ségrégation, économétrie, éducation

Domaine concerné : Indicateurs, échelles, indices - Statistique descriptive, analyse factorielle

Résumé

La ségrégation est la propension qu'ont les situations locales à s'écarter de la situation moyenne. L'absence de ségrégation se traduit par une répartition aléatoire de la population dans l'espace étudié, aboutissant à des répartitions spatiales semblables pour les différents groupes sociaux (Floch, 2017). Ces dernières années, une attention particulière a été accordée à cette question, dans le domaine de l'éducation, notamment en ce qui concerne les écarts sociaux entre collèges (Givord et alii, 2016).

Le choix entre les différents indicateurs de ségrégation disponibles dépend en partie de la nature de la variable décrivant le milieu social. Longtemps, les travaux sur la ségrégation ont porté sur des distinctions dichotomiques (favorisés/défavorisés ; noirs/blancs) et l'indice de Duncan, et dans une moindre mesure l'indice d'exposition normalisé, ont été privilégiés. Ces indices se relient assez facilement à des indicateurs de dispersion (l'écart absolu moyen et variance).

Cependant, résumer le milieu social à une distinction dichotomique est réducteur et des variables plus fines sont maintenant utilisées (distinction en 4 groupes selon la profession des parents dans les données de la DEPP ; tranches de revenu dans les données Insee). Pour ce type de variable, c'est l'indice d'entropie qui semble le plus pertinent.

Pour une unité i avec une répartition en K groupes sociaux ($p_{1i} \dots p_{Ki}$) et un effectif n_i , l'entropie est

définie par $h_i = \sum_{k=1}^K p_{ik} \ln\left(\frac{1}{p_{ik}}\right)$ et mesure la diversité des milieux sociaux dans l'unité. l'indice d'entropie fait le rapport entre la diversité moyenne et la diversité sur l'ensemble de la population

$$H = 1 - \sum_{i=1}^I \left(\frac{n_i}{n}\right) \frac{h_i}{h}$$

, où n est le nombre total d'individus, I le nombre total d'unités et h l'entropie sur l'ensemble des individus. Cet indicateur, compris entre 0 et 1, a de bonnes propriétés statistiques, (notamment en termes de décomposabilité par sous-populations).

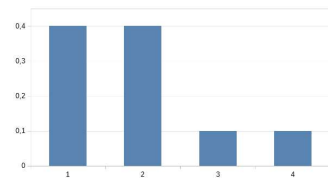
Cet indicateur ne tient pas compte du caractère ordonné des variables, s'il existe. Cela conduit à mettre sur le même plan des ségrégations d'une nature assez différente. L'exemple fictif ci-dessous l'illustre. Dans les deux populations, les individus se répartissent de la même façon dans la catégorie sociale en 4 groupes (25% dans chaque). L'entropie dans chaque établissement est identique (1,19 environ). Par conséquent, l'indice d'entropie est identique dans les deux populations.

Cette conclusion ne pose pas trop de problème si la catégorie sociale n'est pas ordonnée (même si qualitativement, les deux ségrégations ont une forme un peu différente). Même si la catégorie sociale

est ordonnée, on peut ne pas vouloir en tenir compte, en considérant que la ségrégation est un problème global et que dans une unité qui regroupe beaucoup d'individus très défavorisés, peu importe que les autres soient plus souvent un peu défavorisés ou très favorisés. Cependant, on peut aussi avoir une vision plus « unidimensionnelle » de la ségrégation et considérer que la population A est plus ségréguée, car elle oppose des unités très défavorisées à des unités très favorisées, diagnostic moins évident dans la population B.

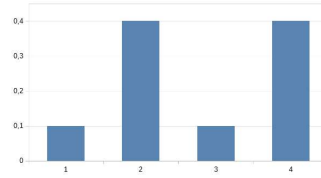
Population A

50 % d'unités

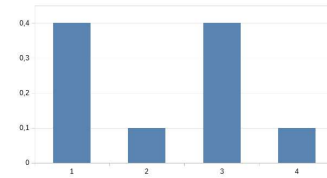


Population B

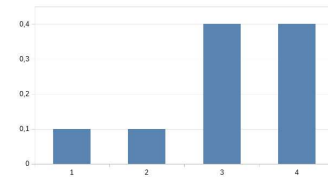
25 % d'unités



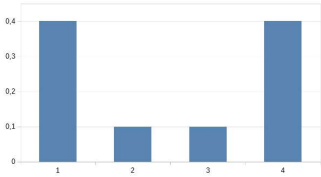
25 % d'unités



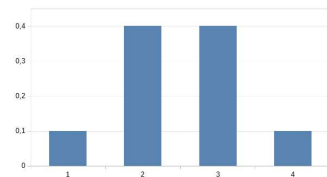
50 % d'unités



25 % d'unités



25 % d'unités



Pour tenir compte de cela, deux possibilités existent actuellement.

- Un indice d'entropie tenant compte du caractère ordonné des variables a été par exemple utilisé par Floch (2017) pour étudier la ségrégation des revenus entre territoires. Il se fonde sur les indices d'entropie dichotomiques calculés pour les individus sous un certain seuil, en augmentant régulièrement ce seuil (par exemple, en suivant les déciles de revenus). Plus que la synthèse qui est faite des indices partiels, avec une pondération adaptée, c'est le profil donné par l'ensemble de ces indices qui semble intéressant (signalant dans l'étude plus de ségrégation pour les hauts revenus que pour les bas revenus).
- Une autre approche consiste à se placer dans un cadre complètement quantitatif. Dans certains cas, c'est relativement simple : plutôt qu'étudier les tranches de revenus, il est possible d'utiliser directement le revenu comme variable quantitative à expliquer. L'indicateur le plus naturel est alors le R^2 , la part de variance du revenu entre unités divisée par la variance totale du revenu. Cependant, les catégorisations sociales utilisant la profession ne sont pas fondées sur une variable quantitative. Attribuer des valeurs arbitraires (1 aux plus défavorisés, 2 à la catégorie suivante...) est une solution simple, mais dont la robustesse est incertaine. Il est sans doute préférable d'utiliser une quantification fine des situations sociales fondée sur un modèle théorique explicite, comme l'Indice de Position Sociale développé à la DEPP (Rocher, 2016).

Nous allons dans ce travail proposer une autre approche, que l'on peut relier avec la vision économétrique de la deuxième possibilité indiquée au-dessus. Repartons d'une variable sociale dichotomique. L'indice d'exposition normalisé peut s'exprimer comme le rapport entre la variance entre unités de cette variable et la variance totale (ici simplement $p*(1-p)$). C'est aussi le R^2 de la régression **linéaire** de cette variable par les indicatrices d'appartenance aux unités. Cela établit un lien connu entre la mesure de la ségrégation et les indicateurs de corrélation (Reardon, Firebaugh, 2002). Une ségrégation importante implique que connaître l'unité d'appartenance d'un individu nous permet de prédire, avec une marge d'erreur assez raisonnable, sa catégorie sociale.

Dans ce cadre, il peut sembler problématique d'utiliser une spécification linéaire pour « expliquer » l'appartenance à une catégorie sociale. Nous montrons la pertinence théorique, dans certains cas, d'une spécification **logistique**, fondée sur l'hypothèse d'une variable latente continue décrivant le milieu social, dont la variable observée serait une réalisation dichotomique. Dans cette régression, les coefficients associés aux unités sont les logits des proportions d'individus favorisés, à une constante

près. On peut aussi relier certains indicateurs sur la qualité de l'ajustement, les pseudo-R² à certains indices de ségrégation : l'indice d'entropie est ainsi le pseudo-R² de MacFadden de ce modèle.

Ce résultat peut être prolongé à une catégorisation sociale en K groupes. L'indice d'entropie multigroupe est le pseudo-R² de MacFadden de la régression polytomique **non ordonnée** de la catégorie sociale par les indicatrices d'appartenance aux différentes unités. Chaque unité est caractérisée par $K-1$ coefficients (la somme des proportions par unité faisant 100 %, cela impose une contrainte identifiante) qui sont à une constante près, les logarithmes des rapports entre les proportions d'individus de type k et de ceux de type K pris en référence.

Le lien entre régressions logistiques et indices de ségrégation étant établi, nous cherchons alors à tenir compte du caractère ordonné de la catégorie sociale en utilisant une régression polytomique **ordonnée**. Nous donnons un cadre théorique à cette approche en reprenant le principe de variable sociale latente proposé pour le cas dichotomique. Le pseudo-R² de MacFadden de ce modèle paraît alors un indicateur pertinent de la ségrégation tenant compte du caractère ordonné de la catégorie sociale. La régression ordonnée étant un modèle plus simple que la régression non ordonnée (chaque unité est caractérisée par un seul coefficient et non $K-1$), l'indice ainsi obtenu est toujours plus petit que l'indice d'entropie habituel. C'est une propriété intéressante, car l'écart entre les deux valeurs permet de tester si la ségrégation est ordonnée (les deux valeurs sont proches) ou non (les valeurs sont assez éloignées : on ne peut déduire dans une unité simplement la répartition dans les $K-1$ premiers groupes en connaissant la proportion du type K ; la ségrégation est « multidimensionnelle »).

Une dernière perspective intéressante concerne l'utilisation de modèles multiniveaux à effets aléatoires. En effet, ce qui vient d'être présenté revient à faire un modèle multiniveaux à effets fixes, sans variable explicative. Passer à un modèle à effets aléatoires pour étudier la ségrégation a déjà été proposé quand la catégorisation sociale est une variable quantitative (Danhier & alii, 2017). Cela peut se faire pour une distinction dichotomique avec une spécification logistique. Pour une catégorisation en plus de 2 groupes, il apparaît nécessaire d'entrer dans le cadre ordonné que nous proposons ici. L'intérêt des modèles à effets aléatoires est notamment qu'ils isolent les fluctuations provoquées par la petite taille de certaines unités, qui est un problème bien connu dans l'étude de la ségrégation.

Une partie de ce travail est bien sûr consacrée à la mise en œuvre de ces méthodes d'abord sur des données simulées, afin de bien montrer dans quels cas, les indicateurs divergent. Les nouveaux indicateurs sont confrontés à l'indice d'entropie classique, non ordonné, mais aussi à celui utilisé par JM Floch dans son article de 2017, ainsi qu'à une mesure de la ségrégation pour une variable sociale quantitative. Le même travail est aussi mené sur des données réelles (notes et milieu social des élèves de 3^e). L'indice d'entropie ordonné apparaît alors assez proche de l'indice d'entropie habituel et les classements des départements selon ces deux indicateurs sont aussi cohérents. Il y a cependant un décalage, d'environ 20 % entre les valeurs, un peu plus grand dans certains départements : ces départements sont plutôt ruraux et cette ségrégation « non ordonnée » non négligeable pourrait correspondre à la situation des enfants d'agriculteurs, dont la répartition dans les établissements suit sans doute une logique plus géographique que sociale.

Bibliographie

- [1] Givord P., Guillermin M., Monso O., Murat F. (2016). « Comment mesurer la ségrégation dans le système éducatif ? », *Éducation & formations*, n° 91, pp 21-51, septembre 2016.
- [2] Danhier J., Martin E., Alarcon-Henriquez A., Kaelen R., Jacobs D. (2017). « Une ségrégation peut en cacher une autre. La répartition des élèves entre classes à prendre au sérieux », *Revue française de pédagogie*, n°199, pp 117-138.
- [3] Floch, J.-M. (2017). « Niveau de vie et ségrégation dans douze métropoles françaises » *Economie et Statistique / Economics and Statistics*, 497-498, pp 73-96.
- [4] Reardon S.F., Firebaugh G., 2002, « Measures of Multigroup Segregation », *Sociological methodology*, N°32, pp. 33-67.
- [5] Rocher T. (2016) « Construction d'un indice de position sociale », *Éducation & formations*, n° 90, pp 5-27.