



École nationale de la statistique
et de l'analyse de l'information

Année scolaire 2016 - 2017

PROJET MICROÉCONOMÉTRIE DE 2^e ANNÉE

Les mesures de la ségrégation sociale

Réalisé par :

Romain Cometx

Jérôme Hananel

Régis Relland

Table des matières

Introduction	2
1 Les mesures usuelles de la ségrégation scolaire	3
2 Polarisation spatiale des collègues	7
3 Mesurer la mobilité entre environnements sociaux	9
Conclusion	12

Introduction

Afin de juger de l'efficacité du système éducatif français, de nombreuses études sociologiques se sont penchées sur les facteurs qui favorisent la réussite ou l'échec des élèves. Il apparaît que l'origine sociale des jeunes joue un rôle important dans le choix de leur parcours ainsi que dans leur réussite. Pour limiter l'influence du milieu social des élèves et ne pas accentuer les inégalités, les autorités publiques cherchent à favoriser la mixité au sein des établissements scolaires. L'analyse et le pilotage du système éducatif nécessitent donc de mesurer l'état de la mixité au sein des collèges et des lycées.

Mais alors, comment mesurer la mixité sociale ? Pour répondre à cette demande, plusieurs indices ont été proposés. Tous ces indices mesurent une absence de mixité. On parle donc d'indices de ségrégation. Quel que soit l'indice, un postulat affirme que les élèves d'une unité (classe, collège, lycée) peuvent être séparés en différents groupes. Chaque groupe correspond à une origine sociale différente. De plus, pour ne pas fausser l'interprétation qui découle de leurs mesures, chaque indicateur doit en principe respecter quatre propriétés :

- si la répartition des élèves entre les différents groupes est la même entre une unité (par exemple un collège) et chaque sous-unité afférente (par exemple les classes de ce collège), l'indice de ségrégation doit rester stable en utilisant la nouvelle subdivision.
- il n'est pas soumis à la taille de la population si la répartition au sein des groupes ne varie pas. On parle **d'invariance d'échelle**.
- il doit être **invariable à la composition**. Cela signifie que si la proportion d'élèves issus d'un groupe augmente dans la population globale et que cette augmentation se répartit équitablement dans chaque unité alors l'indice de ségrégation ne varie pas.
- il doit être **décomposable**. Il est ainsi possible de se focaliser sur certains groupes sociaux par un découpage plus fin de ceux-ci. Il est aussi possible de décomposer l'indice suivant un regroupement des unités (par exemple les collèges du privé ou du public).

Dans un premier temps et après avoir rapidement présenté les principaux indices qui sont utilisés pour mesurer les ségrégations entre établissements, nous nous focaliserons sur l'indice d'entropie normalisé. A travers un exemple, nous verrons qu'il permet d'étudier la ségrégation sous différents angles.

Dans un second temps, nous nous attarderons sur l'indice de Moran qui propose une analyse spatiale. Il permet ainsi de mettre en évidence des corrélations entre des mesures géographiquement voisines. Un exemple concret permettra de rendre compte de l'existence de ce type de corrélations.

Enfin, dans une dernière partie, nous présenterons les limites de ces indices et nous proposerons une approche dynamique. La mixité sociale ne se mesure plus comme une photographie à l'instant t mais comme une évolution sur une période.

Chapitre 1

Les mesures usuelles de la ségrégation scolaire

Avant d'expliciter les formules qui conduisent au calcul des différents indices, il convient de définir un cadre formel. La segmentation utilisée concerne l'origine sociale des élèves des collèges. Le nombre de classes choisi pour l'illustration dépend de l'indice : deux classes (a, a^c) pour les indices de dissimilarité et d'exposition normalisée, quatre classes (a, b, c, d) pour l'indice d'entropie. Les classes sont ordonnées. Ainsi, un élève dans le groupe a est considéré comme très favorisé par son origine sociale alors qu'un élève du groupe d est supposé défavorisé. On note q_i la proportion d'élèves du groupe i dans la population de référence (par exemple les collégiens de l'académie).

Les indices de dissimilarité et d'exposition normalisée : des indices simples mais ne vérifiant pas toutes les propriétés souhaitées

L'indice de dissimilarité

L'indice de dissimilarité proposé par DUNCAN et DUNCAN en 1955 s'écrit, dans sa version simpliste, de la façon suivante :

$$D = \frac{1}{2} \sum_{k=1}^K |q_a^k - q_{a^c}^k|$$

L'interprétation qui en découle est donc aisée. Cet indice compare pour chaque unité k (par exemple un collège) la proportion d'élèves du groupe a à la proportion d'élèves des autres groupes (regroupés en une seule classe). La ségrégation est minimale lorsque les parts d'élèves sont les mêmes pour chaque unité et est maximale lorsque chaque unité est constituée d'élèves d'un seul groupe. Cet indice n'est pas décomposable et ne respecte donc pas la quatrième propriété.

Dans sa version multi-groupes, cet indice reste facilement interprétable et se réécrit comme suit :

$$D = \frac{1}{2I_{Simpson}} \sum_{k=1}^K \pi_k [|q_a^k - q_a| + |q_b^k - q_b| + |q_c^k - q_c| + |q_d^k - q_d|]$$

avec :

- π_k la part d'élèves de l'académie scolarisée dans le collège k , q_a la part d'élèves du groupe a dans l'académie et q_a^k cette même part dans le collège k .
- $I_{Simpson} = q_a(1 - q_a) + q_b(1 - q_b) + q_c(1 - q_c) + q_d(1 - q_d)$, l'interaction de Simpson. Elle peut donc s'interpréter comme la probabilité que deux individus tirés au hasard ne soient pas issus du même milieu social.

Lors de sa généralisation à plusieurs groupes, cet indice perd sa propriété d'invariance à la composition. Deux propriétés sur quatre ne sont alors pas respectées.

L'indice d'exposition normalisé

L'indice d'exposition normalisé mesure quant à lui la probabilité qu'un groupe a soit en contact avec d'autres groupes sociaux. L'intérêt d'un tel indicateur est de pouvoir mesurer la probabilité qu'un élève issu d'un milieu défavorisé soit «tiré vers le haut» par un élève issu d'un milieu favorisé. On part ici du principe que l'interaction entre les groupes est bénéfique. Il est défini par la formule suivante :

$$S = \frac{1}{q_a(1-q_a)} \sum_{k=1}^K \pi_k (q_a^k - q_a)^2$$

Facile à interpréter, cet indicateur ne respecte pas la propriété d'invariance à la composition.

L'indice d'entropie normalisé

L'indice d'entropie présente de bonnes propriétés et permet d'aller plus loin dans l'analyse. Alors qu'avec les autres indicateurs, la ségrégation est présentée comme une opposition binaire entre deux groupes, l'indice d'entropie propose une analyse plus fine de la diversité sociale (multi-groupes). De plus, il a l'avantage d'être décomposable suivant les particularités des établissements scolaires étudiés. Le champ de l'étude est alors affiné.

On définit d'abord l'entropie au sein de l'académie par :

$$h(P) = q_a * \ln\left(\frac{1}{q_a}\right) + q_b * \ln\left(\frac{1}{q_b}\right) + q_c * \ln\left(\frac{1}{q_c}\right) + q_d * \ln\left(\frac{1}{q_d}\right)$$

Quelques remarques :

- on mesure l'entropie correspondant à la distribution des élèves au sein de la population de référence. À ce stade, l'entropie n'est donc pas décomposée.
- plus un groupe est surreprésenté, plus l'entropie diminue vers zéro. Dans le cas extrême où un unique groupe est présent, l'entropie vaut zéro. On peut donc assimiler l'entropie à un score de mixité.
- l'entropie atteint son maximum lorsque les élèves sont équitablement répartis entre les quatre groupes.

À partir de l'entropie, on définit l'indice d'entropie normalisé :

$$H = \sum_{k=1}^K \pi_k \frac{h(P) - h(P_k)}{h(P)}$$

Cet indice compare donc pour chaque collège, sa diversité sociale à celle de la population globale (l'académie). La différence de diversité est pondérée par la taille du collège. La valeur finale est normalisée par l'entropie de l'académie.

L'indice d'entropie normalisé est ainsi compris entre 0 et 1. Il s'approche de 0 lorsqu'il ne détecte aucune ségrégation et tend vers 1 lorsque les groupes ne sont pas mélangés entre les collèges. Cet indice possède un double avantage pour l'analyse :

- étant décomposable, il permet de déterminer des ségrégations propres à un groupe (par exemple les élèves issus d'un milieu défavorisé).
- il est également décomposable selon des regroupements effectués pour les collèges. On peut alors mesurer la ségrégation pour des établissements du privé ou du public (d'autres types de regroupement sont possibles).

Une application de l'indice d'entropie normalisé : la mesure de la ségrégation dans les collèges français

Dans cet exemple, les données sont issues des bases *Scolarité*. On définit le milieu social de chaque élève à partir de la catégorie sociale du père (si l'information est manquante, c'est celle de la mère qui est conservée). Il existe quatre catégories : «défavorisée», «moyenne», «favorisée» et «très favorisée».

Des entropies parfois proches sur les territoires mais qui masquent de fortes disparités

L'application a ici pour but de déterminer des différences de composition sociale entre les collèges. Des départements et académies peuvent présenter un profil comparable (indice de ségrégation proche) mais des disparités très importantes entre les collèges. Ces disparités sont mesurables grâce à la propriété de décomposition de l'indice d'entropie.

Comme le montre le tableau 1.1, la répartition des collégiens au sein des quatre groupes est quasiment la même dans les départements des Vosges et de la Somme. L'entropie $h(P)$ qui mesure la diversité sociale du territoire est donc très proche pour ces deux départements (1,247 et 1,228). Or, les indices d'entropie normalisés H sont très différents puisqu'ils sont égaux à 0,038 pour les Vosges et à plus du double (0,087) pour la Somme. Les collèges de la Somme sont davantage concernés par la ségrégation. La mixité est donc moins importante au sein des établissements de la Somme que dans ceux des Vosges.

Une analyse similaire peut-être conduite sur les académies de Grenoble et de Lyon. Les entropies sont relativement proches (1,346 et 1,334) alors que les indices d'entropie normalisés diffèrent assez largement (0,054 pour l'académie de Grenoble et presque deux fois plus pour celle de Lyon). Ainsi, malgré des profils ressemblants entre ces deux académies, la ségrégation entre les collèges est bien plus forte dans l'académie de Lyon.

		Part de collégiens de catégorie... (en %)					Entropie $h(P)$	Indice d'entropie H (ensemble des collèges)
		Nombre de collèges	Très favorisée	Favorisée	Moyenne	Défavorisée		
Départements	Vosges	48	14,8	13,1	24,6	47,6	1,247	0,038
	Somme	67	15,6	11,0	24,7	48,7	1,228	0,087
Académies	Grenoble	323	24,7	14,5	27,1	33,6	1,346	0,054
	Lyon	302	27,0	13,6	23,8	35,6	1,334	0,092
France hors Mayotte		6 890	22,7	12,5	26,9	37,8	1,318	0,097

Champ : collèges de France hors Mayotte, secteurs public et privé sous contrat.

Source : MENESR-DEPP, système d'information Scolarité.

TABLE 1.1 – Exemples de départements ou d'académies au profil social "moyen" similaire mais présentant une ségrégation plus ou moins forte de leurs collèges, rentrée 2015

Une ségrégation qui croît avec le degré d'urbanisation

Il est également possible d'utiliser l'indice d'entropie pour comparer des territoires qui présentent des profils moyens différents. Il suffit pour cela de comparer les indices d'entropie. Les résultats laissent apparaître une ségrégation plus forte dans les départements où le degré d'urbanisation est élevé (voir figure 1.1). Ainsi, les départements d'Ile-de-France, les Bouches-du-Rhône ou encore les Alpes-Maritimes sont caractérisés par une forte ségrégation entre les collèges. À l'inverse, les collèges se ressemblent le plus dans la Lozère, la Creuse ou les Hautes-Alpes.

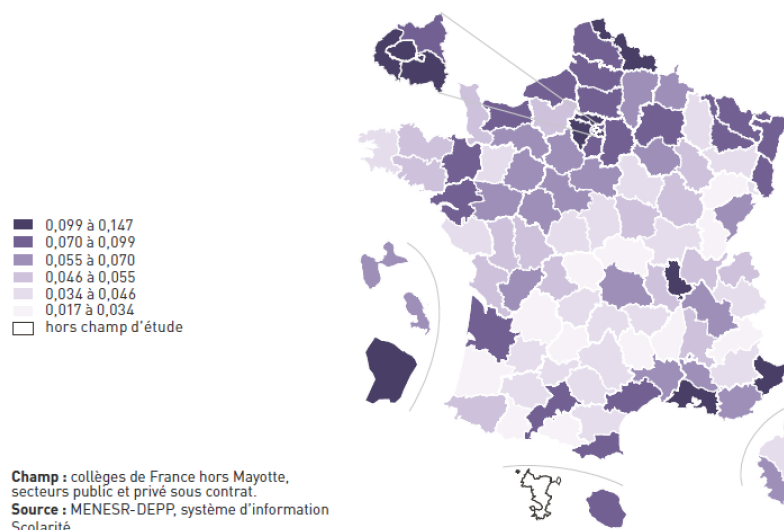


Figure 1.1 – Indice d'entropie à la rentrée 2015

La ségrégation suivant les secteurs Public/Privé

En utilisant la propriété de décomposition de l'indice d'entropie, nous pouvons analyser l'évolution de la ségrégation dans les collèges du secteur public par rapport à celle du secteur privé. Ainsi, il apparaît que depuis 2003 elle stagne sur l'ensemble des collèges alors qu'elle diminue dans le public.

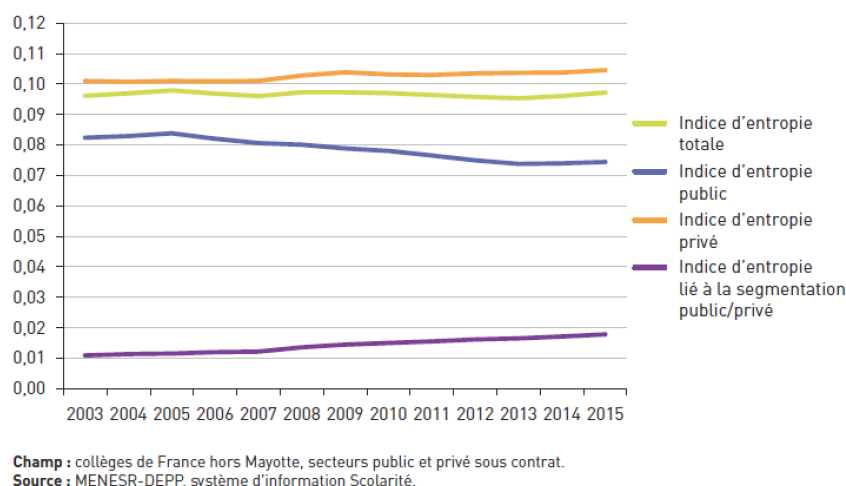


Figure 1.2 – Evolution de l'indice d'entropie suivant les secteurs Public/Privé

En outre, l'indice d'entropie mesurant les compositions sociales entre privé et public a augmenté sur la période. Autrement dit, alors que la ségrégation a diminué dans le public, les différences avec le privé se sont accrues. Cela peut s'expliquer par le fait que les établissements privés ont accueilli davantage d'élèves issus de milieux favorisés par rapport au Public.

Chapitre 2

Polarisation spatiale des collèges

La ségrégation scolaire est un phénomène qui est lié à la composition des élèves d'un établissement. Or, cette composition dépend de l'emplacement géographique, de l'environnement social à proximité, mais également des autres établissements alentour¹. L'affectation d'un élève à un collège est régie par la carte scolaire. Chaque collège public se situe dans un secteur, à l'intérieur duquel les élèves sont rattachés à cet établissement. Certains critères permettent toutefois de déroger à cette règle (handicap, bourses, élèves à la frontière de deux secteurs, etc ...). En revanche, les collèges privés ne sont pas soumis à cette carte scolaire, ce qui peut générer des phénomènes d'évitement : certains parents ne souhaitent pas envoyer leurs enfants dans le collège de secteur et préfèrent choisir un collège privé.

L'indice local de Moran

Un indicateur permet de mesurer cette ségrégation géographique : l'indice local de Moran. Il s'agit d'une mesure de l'autocorrélation spatiale dans la composition des collèges. On peut par exemple mesurer la sur ou sous-représentation d'élèves défavorisés dans un collège, et comparer avec les autres collèges alentour.

Si l'on s'intéresse aux élèves défavorisés d , l'indice de Moran I_d^i pour le collège i est défini par la formule suivante :

$$I_d^i = \frac{(q_d^i - q_d)(\sum_{k=1}^K w_{ik} q_d^k - q_d)}{\sum_{k=1}^K (q_d^k - q_d)^2}$$

avec :

q_d : la proportion d'élèves défavorisés sur l'ensemble de la zone d'étude ;

q_d^k : la proportion d'élèves défavorisés dans le collège k ;

w_{ik} : la matrice de pondération en fonction de l'éloignement entre les collèges i et k ².

Le terme $\sum_{k=1}^K w_{ik} q_d^k$ correspond à la part d'élèves de milieu défavorisé en moyenne dans les collèges alentour. Ainsi, si $(q_d^i - q_d)$ et $(\sum_{k=1}^K w_{ik} q_d^k - q_d)$ sont de même signe, cela signifie que le collège i et les collèges à proximité ont le même profil en ce qui concerne l'accueil des élèves défavorisés. Un collège "défavorisé" est, par exemple, entouré de collèges "défavorisés". À l'inverse, s'ils sont de signe opposés, I_d^i est négatif et le collège i diffère des collèges alentour. Un collège "favorisé" est alors entouré de collèges "défavorisés"³.

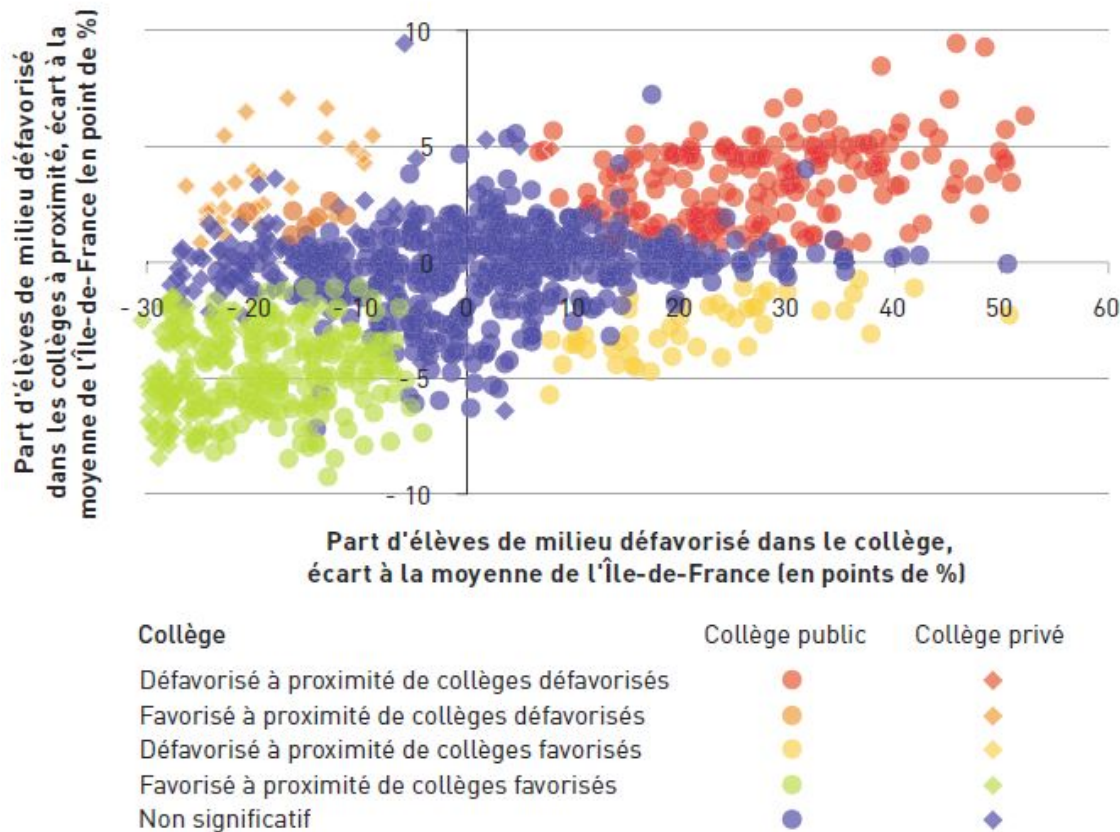
1. www.reformeducollege.fr/mieux-comprendre/comment-les-eleves-seront-ils-affectes-dans-un-college-la-carte-scolaire

2. Dans l'étude de Pauline Givord, c'est l'inverse de la distance à vol d'oiseau qui est retenue.

3. Le choix de la zone d'étude est crucial, il paraît donc délicat de se limiter à une académie donnée du fait de l'existence de frontières administratives : il serait donc plus cohérent d'observer les données sur la France entière. Malheureusement, les calculs sont complexes, c'est pourquoi les auteurs se sont limités à des zones plus petites.

L'exemple de l'Île-de-France

L'étude prend comme exemple l'Île-de-France, qui est symptomatique du phénomène de ségrégation scolaire. Les données concernent donc l'ensemble des collèges des académies de Paris, Créteil et Versailles. La figure suivante représente les collèges selon l'écart de la proportion d'élèves de milieu défavorisé dans le collège considéré par rapport à la moyenne (en abscisse) et selon l'écart entre la proportion d'élèves défavorisés dans les collèges alentour et la moyenne (en ordonnée).



Note : en abscisse : $(q_d^i - q_d)$, en ordonnée : $(\sum_{k=1}^K w_{ik} q_d^k) - q_d$

Source : MENESR-DEPP, système d'information Scolarité

Figure 2.1 – Part d'élèves de milieux défavorisés en Île-de-France, croisée avec celle de leur voisinage

Les collèges qui comptent les plus grands écarts positifs à la moyenne (à droite du graphique), et donc une part d'élèves défavorisés très supérieure à la moyenne, sont tous des collèges publics. Inversement, les collèges avec une part d'élèves défavorisés très inférieure à la moyenne (à gauche du graphique) sont en grande partie des collèges privés.

Le cadran nord-ouest (orange) est composé d'établissements "favorisés" entourés d'établissements défavorisés : 31 collèges sur 38 sont des établissements privés. À l'opposé, dans le cadran sud-est (jaune), les établissements "défavorisés" entourés de collèges favorisés sont tous des collèges publics. À ce titre, certains établissements privés peuvent être vus comme un moyen d'éviter des collèges publics réputés difficiles : les familles aisées à proximité de quartier sensible peuvent avoir tendance à inscrire leurs enfants dans des collèges privés plutôt que dans le collège public d'affectation.

Chapitre 3

Mesurer la mobilité entre environnements sociaux

Les différents indices de ségrégation ou plus généralement d'inégalités ont un défaut commun, il sont mesurés à un instant t figé. Ainsi, même si on peut toujours mesurer l'évolution de ces indices dans le temps, le calcul de ceux-ci ne prend pas en compte la mobilité des individus.

Ce chapitre, inspiré du document de travail *Measuring Social Environment Mobility* de l'Insee, sera alors consacré à la mesure de la mobilité dans les environnements sociaux, tout en présentant les limites des autres indicateurs usuels de ségrégation.

Des indicateurs usuels de ségrégation limités

Pour mesurer les évolutions salariales

Pour illustrer les limites des indicateurs usuels à la mesure de la mobilité, nous pouvons prendre un exemple simple. Cet exemple consiste en la mesure d'inégalités salariales entre deux individus A et B à deux instants donnés. Nous allons alors fixer les revenus au premier instant $t = 1$ et considérer deux scénarios différents à l'instant $t = 2$.

Individu A	1000 euros
Individu B	2000 euros

TABLE 3.1 – Revenu des deux individus à l'instant $t = 1$

Individu A	1000 euros
Individu B	2000 euros

TABLE 3.2 – Revenu des deux individus à l'instant $t = 2$: Scénario 1

Individu A	2000 euros
Individu B	1000 euros

TABLE 3.3 – Revenu des deux individus à l'instant $t = 2$: Scénario 2

Ainsi, au vu de ces deux distributions de revenus, il paraît évident que les différents indices d'inégalité des revenus seront identiques quelque soit le scénario. La mesure de l'évolution de cet indice ne sera donc pas suffisante pour mesurer la mobilité et l'inversion des distributions des revenus de ces individus.

Pour mesurer la mixité scolaire

Un autre exemple simple illustrant la difficulté d'expliquer les indicateurs de mobilité par les différents indicateurs de ségrégation est celui des inégalités scolaires. Considérons ici une population de $N = 8$ élèves, appartenant chacun à une certaine catégorie sociale (ici, on en considère seulement deux par souci de simplification, celles-ci sont représentées par des cases noircies ou non sur la figure

suivante). Ces 8 individus sont répartis la première année ($t = 1$) dans deux écoles (X et Y), et l'année suivante ($t = 2$), on observe la composition des nouvelles écoles (X' et Y').

Le point soulevé ici est le même que pour les inégalités salariales. En effet, les indicateurs de ségrégation ne pourront pas mesurer la mobilité, étant donné que les répartitions des différentes classes sociales sont identiques dans les deux cas.

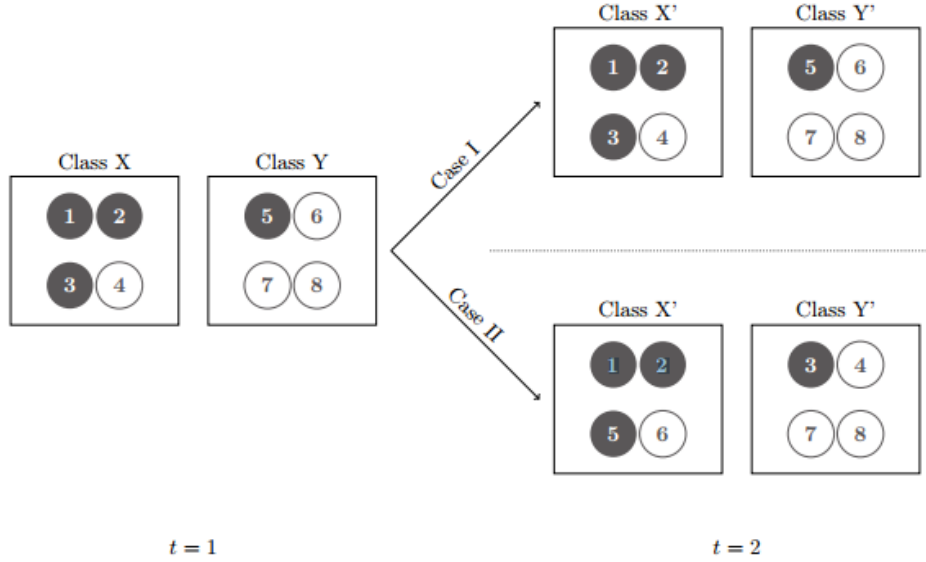


Figure 3.1 – Répartition de 8 élèves de 2 milieux sociaux différents entre 2 écoles différentes, entre deux instants

Ces exemples, aussi simples soient-ils, expliquent de manière efficace que la mesure de l'évolutions d'indicateurs figés dans le temps est insuffisante pour donner un aperçu de la mobilité, ou plus généralement sur les différents phénomènes qui ont pu se dérouler entre deux instants.

La mobilité peut également être mesurée

Certains indicateurs, moins utilisés dans la littérature courante, permettent néanmoins de mesurer la mobilité, sous couvert de certaines hypothèses. Prenons ici l'exemple basique d'une société à N catégories sociales distinctes $1, 2, \dots, N$, et supposons connue la distribution de probabilités $P = (p_{ij})$ correspondant aux différentes probabilités de passage d'un groupe i à un groupe j . Un indicateur de mobilité est alors :

$$M_T = \min\left(\frac{N - \text{trace}(P)}{N-1}; 1\right)$$

Ainsi, cet indicateur prend des valeurs comprises entre 0 et 1, 0 correspondant à une mobilité nulle, et 1 à une mobilité totale. De même, on peut construire des indicateurs permettant de mesurer la mobilité des revenus.

Application de cet indicateur : Mesurer la mobilité entre environnements scolaires

Pour appliquer cet indicateur, reprenons l'exemple précédent des 8 élèves, appartenant à 2 classes sociales distinctes, réparties dans 2 écoles différentes (voir figure 3.1). Pour cela, nous pouvons évaluer de manière empirique les matrices de transition. L'indicateur présenté révèle ici une forte mobilité dans le cas 2, contrairement à une mobilité nulle dans le cas 1.

Cas 1 : $M_T = 0$

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Cas 2 : $M_T = 1$

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

Conclusion

Pour favoriser la réussite scolaire de tous les jeunes, les autorités publiques cherchent à ne pas reproduire les inégalités sociales au sein des établissements scolaires. Pour y parvenir, les établissements doivent favoriser la mixité sociale. Différents indices permettent de mesurer la ségrégation entre les collèges, les lycées, les classes, etc. L'indice d'entropie normalisé est couramment utilisé car il a de bonnes propriétés et offre des résultats interprétables sous différents angles. Ainsi, il apparaît qu'en France, la ségrégation entre les collèges diffère selon les territoires. Les territoires urbanisés peinent d'ailleurs à favoriser la mixité sociale dans leurs établissements. Quant aux différences de composition des collèges et lycées entre le secteur public et le secteur privé, elles sont bien réelles et ont même tendance à se creuser.

Or, la composition des établissements scolaires est liée à l'emplacement géographique et à l'environnement social à proximité. L'indice de Moran offre une mesure de cette autocorrélation spatiale. L'analyse qui en découle permet par exemple de mettre en évidence, en Île-De-France, une part importante d'élèves défavorisés (par rapport à la moyenne) dans le secteur public et inversement dans le secteur privé.

Des études récentes proposent d'aller plus loin et préconisent de prendre en compte la mobilité sociale des jeunes. L'analyse n'est plus statique et il s'agit de raisonner dans un cadre dynamique. En effet, les indicateurs classiques donnent une vision limitée de la ségrégation, et permettent seulement de comparer l'état du monde à deux instants donnés. Pour étudier la mobilité, il est donc nécessaire d'instaurer de nouveaux indicateurs qui offrent une mesure plus proche de la réalité.