

**Direction des Études et Synthèses Économiques**

**G 2010 / 11**

**L'apport des méthodes à noyaux  
pour mesurer la concentration géographique  
Application à la concentration des immigrés  
en France de 1968 à 1999**

**Roland RATHELOT et Patrick SILLARD**

**Document de travail**



**Institut National de la Statistique et des Études Économiques**

# INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

*Série des documents de travail  
de la Direction des Études et Synthèses Économiques*

**G 2010 / 11**

**L'apport des méthodes à noyaux  
pour mesurer la concentration géographique  
Application à la concentration des immigrés  
en France de 1968 à 1999**

**Roland RATHELOT\* et Patrick SILLARD\*\***

JUILLET 2010

Nous remercions Didier BLANCHET, Mirna SAFI et Chloé TAVAN pour leurs précieux conseils et leurs relectures attentives, Jean-Michel FLOCH pour de fructueuses discussions ainsi que Didier LEFEBVRE pour son précieux travail sur les bases de données géographiques. Les documents relatifs à la base Saphir nous ont été fournis et expliqués par Bernard AUBRY. Ce travail a été commencé alors que les auteurs étaient chargés d'études à l'Insee. Toute erreur est de la seule responsabilité des auteurs. Les opinions exprimées ici sont celles des auteurs et n'engagent aucunement le CREST, le SG-CIV ou l'Insee.

---

\* Crest - Faisait partie du Département des Études Économiques d'Ensemble au début de cette étude

\*\* Secrétariat Général - Comité interministériel des villes - Faisait partie du Département des Études Économiques d'Ensemble au début de cette étude.

## L'apport des méthodes à noyaux pour mesurer la concentration géographique

Application à la concentration des immigrés en France de 1968 à 1999

### Résumé

La majorité des études s'intéressant à la localisation géographique d'un phénomène ou d'une population procèdent, préalablement à la représentation cartographique, à l'agrégation des données selon des unités géographiques administratives, qui n'ont pas de rapport avec la nature du phénomène étudié. Lorsque l'on étudie une population rare, et *a fortiori* lorsque l'on dispose de données non-exhaustives, les résultats dépendent largement des choix opérés au moment de l'agrégation. Dans cette étude, nous utilisons les estimateurs non paramétriques à noyaux pour estimer des rapports de densités de population. L'utilisation de ces méthodes présente, pour un coût technique très faible, l'intérêt d'agréger les données de manière à obtenir un arbitrage optimal entre précision et biais. Nous appliquons ces méthodes au répertoire historique des recensements Saphir (Insee) pour étudier l'évolution de la distribution géographique des immigrés sur le territoire français entre 1968 et 1999. Un deuxième apport de ce travail est de proposer un indice de concentration résidentiel adapté aux rapports de densités obtenus dans une première étape.

**Mots-clés** : immigration, localisation géographique, concentration résidentielle, statistique spatiale, estimation non-paramétrique, indices de ségrégation

---

## What kernel methods bring to the analysis of spatial concentration of migrants in France: 1968-1999

### Abstract

Most studies about the geographical location of a phenomenon or a population first aggregate data according to administrative boundaries which are generally not related to the issue. When the population under study is rare, and when the sample is not exhaustive, these aggregation choices often lead the results. In this study, we use non-parametric kernel techniques to estimate ratios of population densities. Using this kind of method is not technically costly and allows one to obtain the optimal trade-off between variance and bias. We apply this technique to the distribution of immigrants over the French territory, using data from population censuses between 1968 to 1999. Another contribution of this work is to propose a concentration index based on the density ratios obtained in the first step.

**Keywords**: spatial concentration, geographical location, segregation indices, non-parametric estimation, immigration

**Classification JEL** : J15, 018, R00

# 1 Introduction

L'analyse spatiale d'un phénomène nécessite souvent, en premier lieu, d'être capable d'en mesurer la densité, c'est-à-dire la probabilité que ce phénomène se produise en chaque point de l'espace. Lorsque l'on étudie une population, la densité permet de connaître la répartition précise de celle-ci sur un territoire. Comme dans les autres domaines de la statistique, la nature des données disponibles et le fait que le nombre d'observations est nécessairement fini (même lorsque l'on dispose d'une base exhaustive) limite la précision de l'estimation de la densité. La première étape de l'estimation consiste à agréger les données spatiales sur une unité géographique que le statisticien peut choisir. Ce choix est cependant guidé par l'arbitrage entre deux problèmes concurrents. Si l'unité choisie est trop petite, la densité ne dépendra, en chaque point, que d'un nombre très faible d'observations. Elle sera peu biaisée mais sera en contrepartie très imprécise. Si l'unité choisie est trop grande, l'estimateur de la densité utilisera beaucoup d'observations, dont certaines situées loin du point d'estimation. La variance de la densité sera donc assez faible, mais potentiellement biaisée. Cet arbitrage biais-variance est courant en statistique et a donné naissance à des méthodes pour lesquelles le choix de l'unité d'agrégation vise à minimiser un critère intégrant à la fois le biais et la variance, en fonction des données du problème considéré.

L'un des objectifs de ce papier est de souligner, dans le contexte de l'étude de populations rares, l'importance de l'échelle d'analyse et de montrer que les méthodes non paramétriques d'estimation de densités permettent d'utiliser l'information localisée à l'échelle la plus fine possible et de l'agréger avec parcimonie, pour un coût technique assez faible. Ce faisant, nous proposons une description spatiale et longitudinale des phénomènes de polarisation à l'œuvre dans la localisation des populations immigrées<sup>1</sup> en France, de 1968 à 1999.

La question de la répartition spatiale des immigrés en France est un sujet régulièrement étudié en sciences sociales. L'utilisation de ces méthodes montre que la plupart des caractéristiques géographiques de la localisation des populations étrangères sont infra-régionales, voire infra-départementales. Ainsi, l'analyse est susceptible d'être plus finement menée lorsque l'on utilise directement des données communales, convenablement agrégées grâce à des estimateurs à noyaux, plutôt que des données agrégées suivant les limites administratives (département, région).

Par ailleurs, une autre contribution de ce travail est de proposer, à partir des valeurs estimées des densités de population, une mesure de la concentration géographique des différentes populations. En particulier, le fait d'être capable de mesurer la concentration géographique est

---

<sup>1</sup>On rappelle que les termes d'*immigrés* et d'*étrangers* recouvrent deux réalités différentes : un immigré est né non-français hors de France et peut avoir acquis la nationalité française par la suite ; un étranger est de nationalité étrangère. Les deux individus vivent en France au moment du recensement. On rappelle également qu'un enfant né en France de parents étrangers vivant en France est considéré comme étranger jusqu'à 18 ans. À 18 ans, le jeune devient français automatiquement, sauf s'il exprime le souhait de conserver sa nationalité de naissance. Dans le recensement de 1999, 1 étranger sur 6 est né en France.

essentiel à la caractérisation et la compréhension des phénomènes de ségrégation spatiale. Pour les décideurs politiques, de telles mesures sont utiles pour guider la mise en place des politiques en faveur du développement local. La politique de développement des grands ensembles urbains faisant suite à la crise du logement des trente glorieuses a perduré jusque dans les années 1970 et a contribué à l'émergence de quartiers concentrant chômage et insécurité. En l'absence de politique correctrice, on constate une tendance forte à la polarisation des populations défavorisées, en particulier dans ces quartiers. Cette polarisation s'opère à une échelle fine et son appréciation requiert d'utiliser l'information au niveau le plus détaillé possible. Ainsi, nous proposons, à partir des mesures de densités de populations obtenues, une mesure de concentration géographique relative, comparable entre différentes populations, ou, pour une même population, au cours du temps.

Le calcul d'un indice de concentration, basé sur l'indice d'entropie, par groupe d'origine et par année de recensement, montre que la concentration géographique des groupes nationaux décroît régulièrement depuis 1968, à l'exception des origines d'immigration récente, comme les immigrés d'Asie et d'Océanie. Certains groupes sont fortement concentrés sur l'ensemble de la période 1968-1999. Il s'agit essentiellement des immigrés originaires d'Afrique subsaharienne, de Tunisie et d'Asie.

## 1.1 L'étude de la localisation des immigrés en France

La question de la localisation de la population immigrée en France a fait l'objet d'études démographiques et sociologiques essentiellement à partir de la fin des années 1980.<sup>2</sup> L'étude du géographe George (1986) utilise les données des recensements de 1954, 1975 et 1982 pour représenter, au niveau départemental, les effectifs et la répartition des étrangers en France. Desplanques et Tabard (1991) mobilisent les mêmes sources (recensements de 1975 et 1982) pour étudier l'évolution de la localisation des immigrés entre les deux années. L'analyse est réalisée pour l'essentiel à une échelle départementale. Les auteurs tentent de dégager les grandes lignes de la répartition nationale des immigrés selon la nationalité. Ils pointent aussi que l'échelle d'analyse pertinente est vraisemblablement plus fine et citent le quartier. Harnois (1992) utilise le recensement de 1990 pour mettre en évidence quelques nouvelles caractéristiques de la distribution des immigrés sur le territoire national, sans opérer de distinction selon la nationalité. Là encore, l'auteur privilégie, pour sa cartographie, une échelle départementale de représentation, en descendant au niveau communal sur les Yvelines et Paris. Desplanques (1996) discute du choix de l'échelle d'analyse de la localisation des immigrés en concluant que l'échelon départemental est le plus pertinent. Chenu (1996) se concentre sur les villes où la population immigrée est plutôt plus représentée qu'ailleurs. Toutefois, aucune analyse géographique n'est proposée. Simon (1996) se concentre sur l'habitat dans lequel s'installent les immigrés. Enfin, Aubry (1996) est le premier papier qui s'appuie sur la base de données Saphir utilisée ici. Aubry traite des

---

<sup>2</sup>On signalera la thèse de Mauco (1932), sans doute la première à aborder de manière systématique le thème de la localisation des étrangers sur le territoire français, mais dont les conclusions en termes de politiques publiques sont critiquables.

migrations internes à la France de la population immigrée et n'exploite pas réellement la dimension spatiale de l'information. Plus récemment, Jayet et Ukrayinchuk (2007) proposent une analyse économique de la répartition géographique des immigrés, en tentant d'en distinguer les déterminants. Ils utilisent également les données des recensements de 1968, 1982 et 1990, en utilisant l'échelle communale ou une agrégation à la zone d'emploi.

## 1.2 L'immigration vue par les statistiques du recensement

La France a une longue histoire d'immigration<sup>3</sup>. À l'aube du vingtième siècle, les immigrés représentent environ 3 % de la population totale vivant en France. À la fin de la Première Guerre Mondiale, le pays souffre d'un manque de main-d'œuvre et signe des conventions avec d'autres pays européens afin de faciliter les migrations. En 1931, les immigrés représentent 6,6 % de la population. Au cours des années trente, les conséquences de la Grande Dépression poussent le gouvernement à prendre des mesures pour endiguer les flux d'immigration. Mais à la fin des années trente, ils reprennent, en majorité depuis l'Espagne. Pourtant, en 1946, les immigrés représentent toujours moins de 5 % de la population. Après la Seconde Guerre Mondiale, les autorités encouragent l'immigration qu'elles souhaitent tout de même contrôler en fonction des besoins. Les années 1956-1973 sont une période où l'immigration suit un rythme relativement soutenu mais le choc pétrolier de 1974 conduit à nouveau le gouvernement à la freiner.

Tout au long de cette étude, nous utilisons la base de données Saphir de l'Insee. Cette base est une extraction de l'exploitation lourde des recensements généraux de la population pour les années 1962, 1968, 1975, 1982, 1990 et 1999. Toutes les variables des recensements ne sont pas retenues dans Saphir, mais celles qui le sont ont vu leur nomenclature faire l'objet d'une harmonisation. Nous utilisons dans cette étude toutes les années sauf 1962. Le taux de sondage est de 1/4 pour 1968, 1982, 1990 et 1999. Il est de 1/5 pour 1975.

Nous sélectionnons, parmi les variables contenues dans la base, la nationalité de naissance, le pays de naissance, ainsi que la commune de résidence de l'individu.<sup>4</sup> Nous distinguons ainsi d'une part les Français de naissance et, d'autre part, les immigrés, par nationalité ou groupe de nationalités.<sup>5</sup>

L'évolution du nombre de Français et immigrés vue par Saphir est donnée dans la figure 1, ainsi qu'un détail par nationalité d'origine pour ces derniers. Le graphique de gauche montre que le nombre d'immigrés croît sur toute la période. De 6,8 % de la population en 1962, la proportion

---

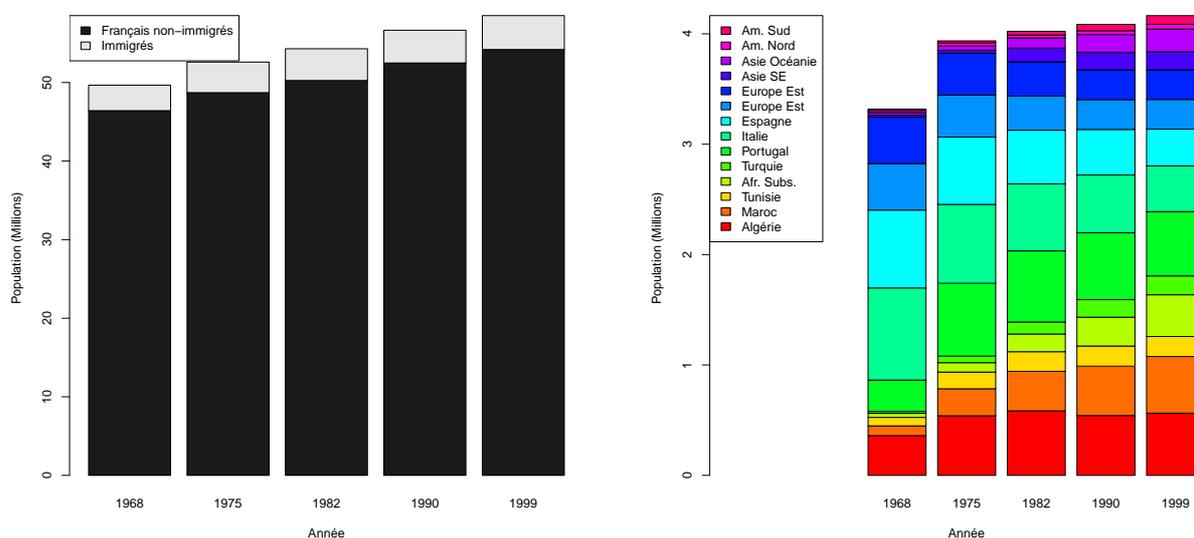
<sup>3</sup>Cf Noiriel (1988), Dewitte (2003) ou Insee (2005) pour davantage d'informations sur l'histoire de l'immigration en France.

<sup>4</sup>Les 15 nationalités les plus fréquentes sont renseignées *in extenso*. Les autres sont regroupées. Ces dernières ne sont pas utilisées dans l'étude, sauf lorsqu'on se réfère à la population immigrée, dans son intégralité.

<sup>5</sup>Les nationalités les plus représentées – Algérie, Espagne, Italie, Maroc, Portugal, Tunisie, Turquie – ont été conservées, les autres ont été regroupées – Afrique subsaharienne, Amérique du Nord, Amérique du Sud, Asie du Sud-Est, Reste de l'Asie et Océanie, Europe de l'Est et Europe de l'Ouest.

d'immigrés atteint 7,4 % en 1975 et stagne ensuite à ce niveau jusqu'à 1999. Le graphique de droite présente l'évolution de la composition de la population immigrée par nationalité d'origine. Parmi les nationalités d'origine les plus représentées, la population d'origine algérienne atteint 580 000 personnes en 1982 et se maintient à ce niveau par la suite. La population d'origine marocaine atteint 510 000 individus en 1999, son point maximum sur la période. Les immigrés d'origine portugaise, qui étaient très peu présents en 1962, arrivent massivement à partir des années soixante, pour atteindre un pic de 660 000 personnes en 1975. Le nombre d'immigrés d'origine italienne, arrivés en France principalement dans la première partie de vingtième siècle, décroît sur l'ensemble de la période étudiée, de 830 000 en 1968 à 410 000 en 1999.

FIG. 1 – Évolution de la population française et immigrée en France (gauche) ; Évolution de la composition immigrée par pays ou continent d'origine (droite)



Source : Saphir, Insee.

### 1.3 Le choix de l'échelle d'analyse

Le choix de l'échelle d'analyse n'est séparable ni des quantités étudiées ni de l'utilisation que l'on souhaite faire des résultats. Ainsi, lorsque l'on souhaite savoir si une statistique de ratio de la population étrangère sur la population totale doit être calculée à l'échelle d'un quartier, d'une commune, d'un département ou même d'un pays, il convient de se demander à quelle question cette statistique doit répondre. Selon qu'il s'agisse d'expliquer, par exemple, la localisation sur le territoire ou les phénomènes de ségrégation urbaine, l'échelle pertinente peut ne pas être la même. Dans le premier cas, qui est aussi celui de cette étude, il s'agira d'apprécier des phénomènes à l'échelle nationale : l'échelle pertinente sera sans doute supra-communale. Dans le second cas, il sera nécessaire d'observer la localisation des individus au sein des agglomérations,

à l'échelle du bâtiment ou au moins du quartier. De manière générale, il est toujours nécessaire de concilier une échelle fine et une précision correcte des estimateurs.

Ainsi que Desplanques (1996) l'a noté, le choix de l'échelle est délicat pour la question qui nous intéresse ici. Les données communales telles qu'elles sont disponibles dans les recensements paraissent en effet difficiles à exploiter et ce, pour au moins deux raisons. Premièrement, les communes sont extrêmement différentes les unes des autres, en surface comme en population. Par conséquent, les résultats obtenus sur des communes de type rural et urbain ne seront pas comparables. Deuxièmement, l'échantillonnage (au quart ou au cinquième selon les recensements) induit l'existence d'une incertitude d'autant plus importante que les effectifs sont réduits. Les communes peu peuplées ne sont donc pas isolables, car l'incertitude est dans ce cas trop importante.

La littérature traitant de cette question s'est donc souvent limitée à l'échelle départementale. Nous considérons qu'il s'agit d'un choix contestable et qu'il est possible de mieux saisir l'information disponible en agrégeant de manière différente les données à l'échelle la plus fine, soit, ici, l'échelle communale. Le département présente selon nous trois inconvénients majeurs :

- Un manque de *comparabilité*. Si les communes sont difficiles à comparer les unes aux autres, c'est également le cas des départements. On peut penser qu'en termes de population comme d'étendue géographique le département des Hauts-de-Seine est difficilement comparable à l'Aveyron.
- Une grande *taille*. Pour beaucoup de questions spatiales liées aux phénomènes d'immigration, l'échelle pertinente est plus proche de celle de la commune. Agréger les données au niveau du département conduit à modifier artificiellement le contraste de l'information et, le plus souvent, à le gommer.
- Un découpage *arbitraire*. Les contours des départements n'ont pas été tracés pour traiter des questions relatives à l'immigration. Même si l'on admettait que le département offrait une échelle adéquate, pour quelle raison devrait-on agréger l'information selon ces contours et pas selon d'autres de mêmes dimensions ?

Les deux derniers éléments sont liés au fait que les limites administratives sont, pour étudier la plupart des questions, arbitraires et non adaptées au cas considéré. Ces problèmes de statistique spatiale sont connues sous le nom de *Modifiable Areal Unit Problem* (MAUP)<sup>6</sup>.

Dans la section suivante, nous détaillons la méthode proposée pour estimer la répartition des populations, ainsi que la proportion de chacune d'elles dans la population totale. Un indice de concentration spatiale, basé sur l'indice d'entropie, est également introduit. Dans une troisième

---

<sup>6</sup>Le MAUP est la conséquence de l'utilisation d'un zonage pour présenter des statistiques de comptage ou de moyennes de phénomènes irrégulièrement répartis et éventuellement corrélés dans l'espace. À l'extrême, l'agrégation spatiale fait davantage ressortir la corrélation spatiale d'un phénomène que la réalité de celui-ci, cf. Openshaw et Taylor (1979).

partie, nous appliquons ces méthodes aux cinq derniers recensements, de 1968 à 1999, et estimons des proportions locales pour une série de groupes nationaux significatifs. Nous calculons l'indice de concentration pour ces cinq années et pour les groupes nationaux retenus, ce qui nous permet de formuler un diagnostic sur l'évolution de la concentration géographique des immigrés en France sur une période de trente ans.

## 2 Présentation de la méthode

### 2.1 L'objectif

Dans cette étude, nous souhaitons ré-examiner la répartition spatiale des immigrés sur le territoire français, de 1968 à 1999. D'un point de vue plus général, nous souhaitons procéder à l'analyse de la concentration d'une population particulière (ici celle des immigrés, voire des immigrés d'une origine donnée). Cette concentration se manifeste soit de manière absolue (la population étudiée se concentre en certains lieux de l'espace géographique et non de manière uniforme), soit de manière relative à une population de référence (ou par rapport à d'autres populations d'intérêt). La concentration absolue permet d'identifier des structures générales de la répartition de la population d'intérêt mais ces structures reproduisent, pour l'essentiel, la distribution géographique de la population totale. Par exemple, l'essentiel des habitants sont, en France, localisés dans les communes urbaines.<sup>7</sup> Au premier abord, les populations étudiées ici reproduisent logiquement ce schéma, s'installant auprès des pôles urbains au contact de la demande de travail et de l'offre de logement. Par conséquent, l'analyse de la concentration s'opère plutôt à grands pôles urbains donnés, débouchant ainsi sur l'idée de la concentration relative : la population générale constitue la référence par rapport à laquelle la répartition de la population d'intérêt est analysée. Plus précisément, nous souhaitons présenter une méthodologie permettant de répondre à trois familles de questions, relatives à la localisation et la concentration d'une population d'intérêt.

1. Dans quels lieux la population d'intérêt se concentre-t-elle ?
2. Comment la concentration relative à la population d'intérêt évolue-t-elle avec le temps ?
3. La population d'intérêt est-elle plus ou moins concentrée que la population de référence ?

Ces dimensions nécessitent la mise en place d'un cadre probabiliste qui est précisé à la sous-section 2.2. Puis, la sous-section 2.3 explicite le calcul et l'estimation d'un rapport entre population d'intérêt et population de référence en vue d'une représentation cartographique des lieux de concentration accrue. Enfin, la sous-partie 2.4 définit un indice de concentration dérivé de l'entropie de Shannon, explique les propriétés qui en motivent l'utilisation et précise son calcul en fonction des rapports de densités calculés dans la sous-section précédente.

---

<sup>7</sup>les 3000 communes de plus de 10 000 habitants (soit moins de 10 % des communes françaises) regroupent plus de 60 % de la population française (chiffres de 2005).

## 2.2 Densités de population

### 2.2.1 Cadre probabiliste

La position d'un individu appartenant à une population d'intérêt  $\mathcal{X}$  dans l'espace géographique (ici le territoire métropolitain de la France, noté  $\mathfrak{F}$ ) est supposée être une variable aléatoire bi-dimensionnelle notée  $\mathbf{X}^{\mathcal{X}}$ . On peut penser, par exemple, à sa longitude et sa latitude dans un référentiel donné. Chaque point  $\mathbf{x}$  de  $\mathfrak{F}$  est une valeur possible de la variable  $\mathbf{X}^{\mathcal{X}}$ . La répartition de la population d'intérêt  $\mathcal{X}$  peut donc être comprise comme la densité de la variable aléatoire  $\mathbf{X}^{\mathcal{X}}$  sur  $\mathfrak{F}$ . On notera  $f^{\mathcal{X}}(\mathbf{x})$  la densité de la variable  $\mathbf{X}^{\mathcal{X}}$  sous-jacente à la répartition spatiale de la population  $\mathcal{X}$ . Il est possible de construire un estimateur de la densité de  $\mathbf{X}^{\mathcal{X}}$  étant donné un échantillon  $(\mathbf{X}_1^{\mathcal{X}}, \dots, \mathbf{X}_{N^{\mathcal{X}}}^{\mathcal{X}})$  de variables caractérisant les localisations de  $N^{\mathcal{X}}$  individus de la même population  $\mathcal{X}$  (en d'autres termes, tous les  $\mathbf{X}_i^{\mathcal{X}}$  ont pour loi  $f^{\mathcal{X}}$  pour  $i \in \{1, \dots, N^{\mathcal{X}}\}$ ). Nous estimons cette densité en utilisant un estimateur à noyaux (Silverman 1986, Wand et Jones 1995) :

$$\hat{f}_{\mathbf{H}}^{\mathcal{X}}(\mathbf{x}) = \frac{1}{N^{\mathcal{X}}} \sum_{i=1}^{N^{\mathcal{X}}} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i^{\mathcal{X}}) \quad (1)$$

où  $K_{\mathbf{H}}(\cdot)$  est un noyau, par exemple gaussien, dépendant d'un paramètre matriciel de largeur de bande,  $\mathbf{H}$  :

$$K_{\mathbf{H}}(\mathbf{r}) = \frac{1}{2\pi\sqrt{|\mathbf{H}|}} \exp\left[-\frac{\mathbf{r}'\mathbf{H}^{-1}\mathbf{r}}{2}\right]$$

où  $|\mathbf{H}|$  désigne le déterminant de la matrice carrée symétrique définie positive  $\mathbf{H}$ .

On note que  $\int_{\mathbb{R}^2} K_{\mathbf{H}}(\mathbf{r}) d\mathbf{r} = 1$ , de sorte que  $\hat{f}_{\mathbf{H}}^{\mathcal{X}}$  est une densité de probabilité ( $\int \hat{f}_{\mathbf{H}}^{\mathcal{X}} = 1$ ).

### 2.2.2 Utiliser des données communales

Les données que nous utilisons ne fournissent qu'une information imparfaite de localisation. En effet, nous n'avons pas directement accès au vecteur  $\mathbf{X}_i$ , réalisation pour l'individu  $i$  de la variable  $\mathbf{X}$ , mais plutôt à une variable indicatrice précisant, pour chaque individu, sa commune de résidence. On notera  $I(j)$  l'ensemble des individus qui résident dans la commune  $j$ . L'ensemble des  $N$  communes françaises forme une partition du territoire métropolitain  $\mathfrak{F}$ . On note  $\mathbf{y}_j$  le centre de gravité géographique de la commune  $j$ .<sup>8</sup> En somme, pour l'individu  $i$ , on dispose d'une information de localisation approchée de son lieu de résidence par  $\mathbf{y}_j$  lorsque  $i \in I(j)$ . Dans la majeure partie des cas, l'approximation précédente entraîne un biais limité dans l'estimation de

---

<sup>8</sup>Avec les notations précédentes,  $\mathbf{y}_j = \frac{\int_{\mathfrak{C}_j} \mathbf{r} d\mathbf{r}}{\int_{\mathfrak{C}_j} d\mathbf{r}}$  où  $\mathfrak{C}_j$  désigne le territoire géographique de la commune  $j$ .

la densité (1). En effet,

$$\hat{f}_{\mathbf{H}}^{\mathcal{X}}(\mathbf{x}) = \frac{1}{N} \sum_{j=1}^N \frac{N_j^{\mathcal{X}} \times N}{N^{\mathcal{X}}} \left[ \frac{1}{N_j} \sum_{i \in I(j)} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) \right]$$

où  $N_j^{\mathcal{X}}$  est le nombre d'individus de la population  $\mathcal{X}$  résidant dans la commune  $j$ ,  $N_j$  le nombre d'individus dans la commune  $j$  et  $I(j)$  l'ensemble des individus de cette commune. Dans des conditions stylisées où le nombre d'individus par commune est identique<sup>9</sup> et où la distribution des individus au sein d'une commune est uniforme, alors, par la loi des grands nombres,

$$\left[ \frac{1}{N_j^{\mathcal{X}}} \sum_{i \in I(j)} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i) \right] \xrightarrow{p.s.} K_{\mathbf{H}}(\mathbf{x} - \mathbf{y}_j)$$

Naturellement, la vitesse de convergence et le biais induit par l'approximation précédente sont liés à l'écart des conditions réelles aux conditions stylisées, ainsi qu'au couple  $(\mathbf{H}, \mathbf{x})$  considéré. De même, la forme de  $K_{\mathbf{H}}$  a une incidence sur le biais. Dans tous les cas, l'imprécision de la localisation n'a véritablement d'influence que sur la forme localisée de  $\hat{f}_{\mathbf{H}}^{\mathcal{X}}$ . En d'autres termes, la connaissance de la forme générale de  $f^{\mathcal{X}}$  ne devait pas être perturbée par l'approximation ; seule la connaissance des détails locaux devrait en pâtir. En particulier, il est clair que les détails de la densité à une échelle infra-communale<sup>10</sup> ne seront pas accessibles compte tenu de l'approximation de la localisation retenue.

En assignant, par convention, l'ensemble des individus d'une commune à son centre de gravité, la densité (1) se simplifie en :

$$\hat{f}_{\mathbf{H}}^{\mathcal{X}}(\mathbf{x}) = \sum_{j=1}^N \tilde{f}_j^{\mathcal{X}} K_{\mathbf{H}}(\mathbf{x} - \mathbf{y}_j) \quad (2)$$

où  $\tilde{f}_j^{\mathcal{X}} = \frac{N_j^{\mathcal{X}}}{N}$  est la fraction de la population  $\mathcal{X}$  localisée dans la commune  $j$  ( $N_j^{\mathcal{X}}$  est le nombre d'individus de la population  $\mathcal{X}$  localisés dans la commune  $j$  et  $N$  le nombre total de communes en France) et  $\mathbf{y}_j$  est le centre de gravité de la commune  $j$ .

L'estimateur  $\hat{f}_{\mathbf{H}}^{\mathcal{X}}(\mathbf{x})$  est *a priori* défini pour tout  $\mathbf{x} \in \mathfrak{F}$ . En pratique, on définit une grille sur laquelle on estime ces densités. Les densités sont normalisées pour sommer à l'unité sur la grille.

---

<sup>9</sup>Dans ce cas, le terme  $\frac{N_j^{\mathcal{X}} \times N}{N^{\mathcal{X}}}$  est égal à 1.

<sup>10</sup>Le diamètre moyen d'une commune est de 4 km, donc il est illusoire d'étudier des détails de  $\hat{f}^{\mathcal{X}}$  d'amplitude inférieure à quelques kilomètres.

La détermination du paramètre matriciel de largeur de bande,  $\mathbf{H}$ , est présentée dans le paragraphe suivant.

### 2.2.3 La détermination du paramètre de largeur de bande

Considérons  $f^{\mathcal{X}}$  la densité d'une population  $\mathcal{X}$  sur un espace géographique  $\mathfrak{F}$ . On estime  $f^{\mathcal{X}}$  par un estimateur à noyaux  $\hat{f}_{\mathbf{H}}^{\mathcal{X}}$  conformément à la relation (2). Cet estimateur à noyau fait intervenir un paramètre matriciel  $\mathbf{H}$  qu'il est possible de choisir de façon à minimiser l'erreur moyenne intégrée carrée, définie par :

$$MISE(\mathbf{H}) = \mathbb{E} \int_{\mathfrak{F}} \left( f^{\mathcal{X}}(\mathbf{x}) - \hat{f}_{\mathbf{H}}^{\mathcal{X}}(\mathbf{x}) \right)^2 d\mathbf{x}$$

Cette erreur se décompose en une somme de deux termes : l'intégrale d'un terme de biais

$$B(\mathbf{x}; \mathbf{H}) = \left\{ \mathbb{E}(f^{\mathcal{X}}(\mathbf{x}) - \hat{f}_{\mathbf{H}}^{\mathcal{X}}(\mathbf{x})) \right\}^2$$

et, l'intégrale d'un terme de variance

$$\mathbb{E} \left[ \left( f^{\mathcal{X}}(\mathbf{x}) - \hat{f}_{\mathbf{H}}^{\mathcal{X}}(\mathbf{x}) - B(\mathbf{x}; \mathbf{H}) \right)^2 \right].$$

Le biais est croissant avec  $|\mathbf{H}|$  et la variance est décroissante avec  $|\mathbf{H}|$ , de sorte que l'erreur moyenne intégrée carrée présente un minimum.

Divers algorithmes existent pour calculer une valeur approchée de ce minimum. Nous choisissons la méthode par *plug-in* développée par Duong et Hazelton (2003).<sup>11</sup>

## 2.3 Représenter la concentration relative spatialisée

On se place dans la situation où l'on souhaite comparer la concentration d'une population relativement à une autre. On considère une population  $\mathcal{T}$  de référence par rapport à laquelle on veut étudier le caractère concentré ou non de la population  $\mathcal{X}$ . Dans notre cas,  $\mathcal{T}$  sera la population totale et  $\mathcal{X}$  un sous-ensemble de cette population, Français de naissance ou immigrés par nationalité d'origine. Sur un plan graphique, on est intéressé de préciser les lieux où la densité  $f^{\mathcal{X}}$  s'écarte significativement de la densité  $f^{\mathcal{T}}$ . Une solution couramment utilisée (voir par exemple Kelsall et Diggle (1998)) consiste à étudier la fonction

$$\rho^{\mathcal{X}/\mathcal{T}}(\mathbf{x}) = \frac{f^{\mathcal{X}}(\mathbf{x})}{f^{\mathcal{T}}(\mathbf{x})}$$

---

<sup>11</sup>Cette méthode repose sur un algorithme itératif. Une estimation pilote donne une première mesure de la densité et de ses propriétés locales (dérivées...). Ces quantités sont introduites dans les expressions du critère à minimiser, aboutissant à une première estimation de la matrice optimale. En utilisant celle-ci, une nouvelle estimation de la densité est réalisée, etc.

Hormis aux points où la population de référence est absente, le ratio précédent permet de préciser les lieux où la population  $\mathcal{X}$  est sur-représentée (dans ce cas  $\rho^{\mathcal{X}/\mathcal{T}} > 1$ ) ou sous-représentée (dans ce cas  $\rho^{\mathcal{X}/\mathcal{T}} < 1$ ) par rapport à la population de référence  $\mathcal{T}$ . L'estimation du ratio se construit naturellement à partir de l'estimation des densités selon la formule :

$$\hat{\rho}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}/\mathcal{T}}(\mathbf{x}) = \frac{\hat{f}_{\mathbf{H}^{\mathcal{X}}}^{\mathcal{X}}(\mathbf{x})}{\hat{f}_{\mathbf{H}^{\mathcal{T}}}^{\mathcal{T}}(\mathbf{x})} \quad (3)$$

Les paramètres matriciels de largeur de bande  $\mathbf{H}^{\mathcal{X}}$  et  $\mathbf{H}^{\mathcal{T}}$  sont optimaux, c'est-à-dire que chacun minimise, pour la densité concernée, l'erreur moyenne intégrée.<sup>12</sup>

La probabilité  $p^{\mathcal{X}/\mathcal{T}}(\mathbf{x})$  de rencontrer au point  $\mathbf{x}$  un individu appartenant à la population  $\mathcal{X}$  conditionnellement au fait de rencontrer quelqu'un, peut être facilement obtenue à partir de  $\rho^{\mathcal{X}/\mathcal{T}}(\mathbf{x})$ .

$$p^{\mathcal{X}/\mathcal{T}}(\mathbf{x}) = \frac{N^{\mathcal{X}}}{N^{\mathcal{T}}} \rho^{\mathcal{X}/\mathcal{T}}(\mathbf{x})$$

Cette probabilité est également la proportion locale du groupe  $\mathcal{X}$  dans la population totale. Le rapport  $\rho$  et la probabilité  $p$  sont égaux à une constante près : la proportion de  $\mathcal{X}$  dans  $\mathcal{T}$  sur l'ensemble du territoire. La probabilité  $p^{\mathcal{X}/\mathcal{T}}(\mathbf{x})$  s'estime naturellement par :

$$\hat{p}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}/\mathcal{T}}(\mathbf{x}) = \frac{N^{\mathcal{X}}}{N^{\mathcal{T}}} \hat{\rho}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}/\mathcal{T}}(\mathbf{x})$$

## 2.4 Mesurer globalement la concentration relative

Les quantités calculées dans les sections précédentes permettent d'obtenir des mesures locales de concentration. Dans cette section, on souhaite obtenir une mesure synthétique pour l'ensemble du territoire de la régularité (ou de l'irrégularité) de la distribution géographique d'une population.

### 2.4.1 Les mesures existantes de concentration et de ségrégation

On trouve dans la littérature sur la mesure de la ségrégation des outils utiles à la construction d'un tel indicateur. Le travail de Massey et Denton (1988) propose une typologie pertinente des indices de ségrégation résidentielle. Se plaçant dans le cas d'une population composée de deux groupes et disposées sur des unités géographiques, ils examinent chacun des indices de ségrégation et les classent selon leur signification. Ainsi, ils distinguent cinq dimensions de ségrégation

---

<sup>12</sup>Ceci ne signifie pas que l'erreur moyenne intégrée du ratio de densités soit minimale pour le couple de paramètres matriciels obtenus. Il pourrait donc s'avérer intéressant de rechercher le couple de paramètres  $(\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}})$  qui minimiserait l'erreur moyenne intégrée du ratio de densités. Kelsall et Diggle (1998) ont développé une telle approche pour des densités unidimensionnelles. À notre connaissance, aucun développement de ce type n'a été fait, à ce jour, pour le cas bi-dimensionnel.

– inégalité, contact, concentration, centralité et regroupement – correspondant à autant d’acceptions de ce terme.

- L’*inégalité* caractérise la distribution spatiale des populations entre les différentes unités géographiques. Les indices de dissimilarité, dont l’interprétation est aisée (il s’agit de la proportion d’individus de la minorité qui doivent changer d’unité spatiale pour parvenir à l’équidistribution) et qui sont fréquemment utilisés, appartiennent à cette famille.
- Le *contact* correspond aux possibilités qu’un individu de la population minoritaire a d’entretenir des relations avec le groupe majoritaire. Un groupe minoritaire ségrégué sera un groupe isolé au sens du contact : les membres de ce groupe n’ont pas la possibilité d’interagir avec les membres du groupe majoritaire. Pour le calcul d’indices de contact, la taille du groupe est primordiale : à distribution donnée, un groupe minoritaire sera d’autant plus fréquemment en contact avec les individus du groupe majoritaire qu’il compte peu de membres. Les indices d’interaction et d’isolement appartiennent à cette catégorie.
- La *concentration* lie la taille d’un groupe à l’aire qu’il occupe.
- La *centralité* est pertinente lorsque les unités spatiales sont réparties en centres et en périphéries. Dans ce cas, un groupe est d’autant plus ségrégué qu’une proportion importante réside dans des unités situées loin des centres.
- Enfin, le *regroupement* est la seule dimension à tenir compte explicitement de la géographie des unités spatiales. Ainsi, pour un niveau d’inégalité donné, un groupe sera plus ségrégué au sens du regroupement si les unités spatiales dans lesquelles il est le plus représenté sont proches. Les indices de proximité sont les plus utilisés pour mesurer la dimension de regroupement de la ségrégation.

Dans le cadre de Massey et Denton (1988), les observations concernent uniquement des décomptes de population par unité spatiale. Pour les trois dernières dimensions, les données doivent être enrichies par des informations supplémentaires sur la géographie des unités spatiales. L’aire est nécessaire pour les indices de concentration, la distance au centre pour la centralité et les distances réciproques (ou au moins la contiguïté) des unités pour le regroupement.

Pour une même dimension, sociologues et géographes ont proposé un nombre croissant d’indices, débattant de leurs mérites et défauts respectifs. Une approche axiomatique a également vu le jour. James et Taeuber (1985) a dégagé un certain nombre de propriétés qu’un indice de ségrégation devait satisfaire. D’autres travaux comme ceux d’Hutchens (2001) ou Reardon et Firebaugh (2002a) discutent et justifient ce type d’approche. Il en ressort que les indices les plus fréquemment utilisés, comme l’indice de dissimilarité et l’indice de Gini violent des propriétés importantes. En revanche, les indices issus des théories probabilistes de l’information paraissent mieux se comporter au regard de ces critères.

L’approche de Massey et Denton (1988) comporte un certain nombre de limites qui sont le pendant de sa relative simplicité. Tout d’abord, seuls deux groupes sont envisagés. Reardon

et Firebaugh (2002a) relâchent cette hypothèse en construisant des indices similaires et permettant d’inclure un nombre arbitraire de groupes. Ensuite, et c’est un problème plus grave, ils s’intéressent à des unités spatiales définies et disjointes, mais ne proposent pas de solutions pour des données continues. Une série de travaux, Wong (1993), Grannis (2002), Reardon et Firebaugh (2002b), Wong (2005) discutent le caractère spatial des indices de ségrégation et la pertinence de l’approche classique considérant les unités indépendamment les unes des autres. C’est la contribution de Reardon et O’Sullivan (2004) qui va le plus loin sur cette question, soulignant que les indices de regroupement et d’inégalité décrivent la même réalité lorsque la proximité spatiale des individus est explicitement prise en compte.

Au même titre que le calcul de distributions ou de proportions, le calcul d’indices de ségrégation pose le problème de l’unité spatiale considérée. La plupart des indices répertoriés dans Massey et Denton (1988) et généralisés par les travaux suivants reposent sur des comptages au niveau d’une unité spatiale. Le choix de l’unité ne doit pas nécessairement être celui de l’unité la plus fine. Ainsi, le concept d’exposition fait implicitement l’hypothèse que la plupart des rapports sociaux ont lieu à l’intérieur de l’unité spatiale. Dans le cas de l’Île-de-France et en utilisant les données du recensement de 1999, Gobillon et Selod (2007) s’interrogent sur le lien entre ségrégation résidentielle (en terme de catégories sociales et d’origine nationale) et performance sur le marché du travail (garder son emploi, retrouver un emploi). Ils utilisent la commune comme unité géographique pour calculer leurs indices de ségrégation.

Dans ce travail, nous tentons de contourner le problème d’échantillonnage de nos données à l’échelle communale pour calculer des indices de ségrégation aussi fins que les données le permettent et tenant compte de la distance réciproque des unités géographiques. Ce faisant, nous donnons un autre cadre à l’indice proposé par Reardon et O’Sullivan (2004).

#### **2.4.2 L’indice d’entropie comme mesure de concentration**

De nombreux indices de concentration ont été développés selon le problème traité. Il est utile de reconsidérer cette question en repartant du cadre probabiliste introduit à la section 2.2. D’une manière générale, plus la population  $\mathcal{X}$  étudiée plus haut est concentrée en certains lieux, plus le résultat de l’expérience aléatoire consistant à localiser un individu de la population  $\mathcal{X}$  conformément à la répartition nationale de cette population sera prévisible. À l’extrême, si tous les individus de  $\mathcal{X}$  sont localisés en un point (connu puisque la distribution spatiale est connue), le résultat de l’expérience aléatoire est certain : tout individu de  $\mathcal{X}$  est localisé en ce point. Avec les notations utilisés plus haut, cette expérience aléatoire est caractérisée par la variable  $\mathbf{X}^{\mathcal{X}}$ . Ainsi, la forte concentration de la population  $\mathcal{X}$  entraîne la présence de pics accentués pour la densité  $f^{\mathcal{X}}$  et une incertitude limitée quant à la localisation dans l’espace géographique complet : si la population est concentrée, alors les individus de la population se localisent prioritairement dans les lieux où d’autres individus se trouvent déjà. À l’inverse, la répartition uniforme de la population dans l’espace géographique représente la situation la plus défavorable pour prédire le

résultat de l'expérience aléatoire associée à la variable  $\mathbf{X}^{\mathcal{X}}$ . Ainsi, concentration géographique, densité d'une variable aléatoire de localisation géographique et incertitude associée à l'expérience aléatoire modélisée par la variable précédente sont intimement liées. Ce lien a été mis en évidence par les travaux de Shannon (1948). Shannon (voir annexe "L'entropie de Shannon") propose de mesurer le taux d'information associé à une réalisation de l'expérience aléatoire  $\mathbf{X}^{\mathcal{X}}$  par l'entropie de la variable définie par :

$$\mathbb{H}(\mathbf{X}^{\mathcal{X}}) = -\mathbb{E}(\ln(f^{\mathcal{X}}(\mathbf{X}^{\mathcal{X}}))) = -\int_{\mathfrak{F}} f^{\mathcal{X}}(\mathbf{x}) \ln [f^{\mathcal{X}}(\mathbf{x})] d\mathbf{x} \quad (4)$$

Par construction, l'entropie croît avec le niveau d'aléa associé à l'expérience aléatoire  $\mathbf{X}^{\mathcal{X}}$ , ce niveau étant maximal lorsque la variable suit une loi uniforme sur  $\mathfrak{F}$ . En effet, plus des événements élémentaires associés à l'expérience sont équiprobables, plus le résultat de l'expérience est difficile à prévoir, donc plus le niveau d'aléa est élevé. À l'inverse, plus la population  $\mathcal{X}$  est concentrée (*i.e.* certains événements élémentaires sont plus probables que d'autres), plus l'entropie est faible.

### 2.4.3 Construction d'un indice de concentration relative

Comme précédemment, on souhaite étudier la concentration de la population  $\mathcal{X}$ , sous-population de la population totale  $\mathcal{T}$ . Cependant, on cherche maintenant une mesure globale du caractère plus ou moins concentré de  $\mathcal{X}$  par rapport à  $\mathcal{T}$ . Comme  $\mathcal{X}$  est une sous-population de  $\mathcal{T}$ , les expériences aléatoires  $\mathbf{X}^{\mathcal{X}}$  et  $\mathbf{X}^{\mathcal{T}}$ , sous-jacentes au positionnement des individus dans l'espace, ne sont pas indépendantes. Il convient donc de faire intervenir la population complémentaire à  $\mathcal{X}$  dans  $\mathcal{T}$ , notée  $\mathcal{E}$  pour travailler sur des variables indépendantes statistiquement. Si on considère un individu de la population  $\mathcal{T}$  indépendamment de sa localisation dans l'espace, il appartient soit au groupe  $\mathcal{X}$ , soit au groupe  $\mathcal{E}$ . En somme, l'expérience aléatoire  $Y^{\mathcal{X}}$  à laquelle on s'intéresse se décompose en deux niveaux de choix :

1. **premier niveau de choix** : On tire un individu en population générale. Il est de type  $\mathcal{X}$  avec une probabilité  $\pi$ , ou de type  $\mathcal{E}$  avec une probabilité  $1 - \pi$ .  $\pi$  est donc la proportion de la population  $\mathcal{X}$  dans la population générale  $\mathcal{T}$ . Ce premier niveau de choix correspond à l'entropie d'une variable de Bernouilli notée  $W^{\mathcal{X}}$  :  $\mathbb{H}(W^{\mathcal{X}}) = -\pi \ln(\pi) - (1 - \pi) \ln(1 - \pi)$ .
2. **deuxième niveau de choix** : Une fois le choix de type réalisé, la localisation spatiale est le résultat d'une expérience aléatoire modélisée par la variable  $\mathbf{X}^{\mathcal{X}}$  si l'individu est de type  $\mathcal{X}$ , ou  $\mathbf{X}^{\mathcal{E}}$  si l'individu est de type  $\mathcal{E}$ .

En application des principes de calcul de l'entropie (voir annexe "L'entropie de Shannon"), l'entropie de l'expérience aléatoire  $Y^{\mathcal{X}}$  est :

$$\mathbb{H}(Y^{\mathcal{X}}) = \mathbb{H}(W^{\mathcal{X}}) + \pi \times \mathbb{H}(\mathbf{X}^{\mathcal{X}}) + (1 - \pi) \times \mathbb{H}(\mathbf{X}^{\mathcal{E}}) \quad (5)$$

L'entropie de cette expérience aléatoire est comparable d'une sous-population de  $\mathcal{T}$  à l'autre. Une transformation permet de normaliser la quantité précédente. En effet, on montre que (voir

annexe “Propriétés de l’indice (7)” :

$$-\mathbb{H}(W^{\mathcal{X}}) \leq \mathbb{H}(\mathbf{X}^{\mathcal{T}}) - \mathbb{H}(Y^{\mathcal{X}}) \leq 0 \quad (6)$$

Ce qui nous donne accès à un indice de concentration de la population  $\mathcal{X}$ , étant donnée la répartition spatiale de la population générale :<sup>13</sup>

$$I^{\mathcal{X}} = \frac{1}{\mathbb{H}(W^{\mathcal{X}})} [\mathbb{H}(\mathbf{X}^{\mathcal{T}}) - \pi \times \mathbb{H}(\mathbf{X}^{\mathcal{X}}) - (1 - \pi) \times \mathbb{H}(\mathbf{X}^{\mathcal{E}})] \quad (7)$$

L’une des propriétés de l’entropie de Shannon est qu’elle n’est pas affectée lorsque l’on intervertit les deux étapes de l’expérience aléatoire à laquelle elle se rapporte. Ainsi, si la première étape correspond à la localisation spatiale d’un individu quelconque et la seconde à la détermination de son type, conditionnellement à sa localisation, on obtient une nouvelle écriture de l’indice, conforme à celle développée par Reardon et O’Sullivan (2004, Eq. 7).<sup>14</sup>

$$I^{\mathcal{X}} = 1 - \frac{1}{\mathbb{H}(W^{\mathcal{X}})} \int f^{\mathcal{T}} \left( -p^{\mathcal{X}/\mathcal{T}} \ln(p^{\mathcal{X}/\mathcal{T}}) - (1 - p^{\mathcal{X}/\mathcal{T}}) \ln(1 - p^{\mathcal{X}/\mathcal{T}}) \right) \quad (8)$$

Compte tenu de la relation (5) et de l’inégalité (6),  $0 \leq I^{\mathcal{X}} \leq 1$ .

### Comment s’interprète la valeur de $I^{\mathcal{X}}$ ?

- $I^{\mathcal{X}} = 1$  lorsque, pour tout  $\mathbf{x} \in \mathfrak{F}$ , soit  $f^{\mathcal{X}}(\mathbf{x}) = 0$ , soit  $f^{\mathcal{E}}(\mathbf{x}) = 0$ . En d’autres termes, en tout point de l’espace géographique, la population totale se compose d’individus d’un seul type  $\mathcal{X}$  ou  $\mathcal{E}$ . La concentration est alors maximale puisque tous les individus d’une sous-population sont regroupés ensemble.
- À l’inverse,  $I^{\mathcal{X}} = 0$  lorsque pour tout  $\mathbf{x} \in \mathfrak{F}$ ,  $f^{\mathcal{X}}(\mathbf{x}) = f^{\mathcal{E}}(\mathbf{x}) = f^{\mathcal{T}}(\mathbf{x})$ . Les populations  $\mathcal{X}$  et  $\mathcal{E}$  sont, en tout point de l’espace géographique, réparties dans les mêmes proportions qu’en population totale non spatialisée. Dans ce cas, la concentration est minimale.

Avec l’indice (7), on peut comparer le niveau de concentration d’une sous-population  $\mathcal{X}_1$  avec celui d’une sous-population  $\mathcal{X}_2$  en calculant  $I^{\mathcal{X}_1}$  et  $I^{\mathcal{X}_2}$ . Si  $I^{\mathcal{X}_1} > I^{\mathcal{X}_2}$ , alors  $\mathcal{X}_1$  est plus concentrée que  $\mathcal{X}_2$ . On peut aussi suivre l’évolution de la concentration d’une sous-population  $\mathcal{X}$  dans le temps. Dans ce cas, on compare les indices  $I^{\mathcal{X}}(t = 1)$  et  $I^{\mathcal{X}}(t = 2)$  : si  $I^{\mathcal{X}}(t = 2) > I^{\mathcal{X}}(t = 1)$ , alors la concentration de la population  $\mathcal{X}$  s’est accentuée. Naturellement, plusieurs paramètres peuvent changer :

- Toutes choses égales par ailleurs, la population  $\mathcal{X}$  peut s’être regroupée dans l’espace géo-

---

<sup>13</sup>L’expression (7) mérite que l’on s’y attarde quelque peu. L’indice entropique  $\mathbb{H}(\mathbf{X}^{\mathcal{A}})$  croît à mesure que la loi de  $\mathbf{X}^{\mathcal{A}}$  se rapproche de la loi uniforme ( $\mathcal{A} \in \{\mathcal{X}, \mathcal{E}, \mathcal{T}\}$ ). Plus la densité de la variable  $\mathbf{X}^{\mathcal{A}}$  présente de modes accentués, plus  $\mathbb{H}(\mathbf{X}^{\mathcal{A}})$  est petite. Si la population  $\mathcal{X}$  est concentrée comme la population  $\mathcal{T}$ , alors  $\mathbb{H}(X^{\mathcal{X}})$  et  $\mathbb{H}(X^{\mathcal{T}})$  sont comparables (et par extension  $\mathbb{H}(X^{\mathcal{E}})$  et  $\mathbb{H}(X^{\mathcal{T}})$ ) et leur différence est proche de 0. Si, au contraire,  $\mathcal{X}$  est très concentrée par rapport à  $\mathcal{T}$ , alors  $\mathbb{H}(X^{\mathcal{T}}) \gg \mathbb{H}(X^{\mathcal{X}})$  (et par extension  $\mathbb{H}(X^{\mathcal{T}}) \gg \mathbb{H}(X^{\mathcal{E}})$ ) et dans ce cas,  $I^{\mathcal{X}}$  est proche de 1. Ainsi,  $I^{\mathcal{X}}$  apparaît comme une mesure normalisée de la concentration d’une sous-population  $\mathcal{X}$  relativement à une population totale ( $\mathcal{T}$ ) qui englobe la sous-population précédente.

<sup>14</sup>La démonstration formelle de l’équivalence des deux formes se trouve en annexe.

graphique, son poids dans la population totale demeurant inchangé. Dans ce cas,  $\mathbb{H}(\mathbf{X}^{\mathcal{X}}(t = 2)) < \mathbb{H}(\mathbf{X}^{\mathcal{X}}(t = 1))$  et l'indice de concentration  $I^{\mathcal{X}}$  augmente.

- Toutes choses égales par ailleurs, la population  $\mathcal{X}$  peut avoir crû sans que sa répartition spatiale n'ait été modifiée. On montre que pour des petites valeurs de  $\pi$ ,  $I^{\mathcal{X}}$  est une fonction croissante de  $\pi$  (à  $f^{\mathcal{X}}$  et  $f^{\mathcal{E}}$  données). En d'autres termes, si la population de type  $\mathcal{X}$  pèse peu dans la population totale  $\mathcal{T}$ , un accroissement de son poids relatif sans modification de répartition des sous-populations composant la population totale entraîne une augmentation de l'indice de concentration. Ce résultat se justifie dans la mesure où si la population de type  $\mathcal{X}$  est distribuée spatialement de manière différente de la population alternative  $\mathcal{E}$ , alors l'accroissement de son poids relatif contribue à accentuer les contrastes.

### Extension de l'indice de concentration à plusieurs sous-populations

Si l'on souhaite comparer la situation globale (et non population par population) entre différentes dates, par exemple, il est nécessaire de construire un indice tenant compte de toutes les populations existant sur un territoire. Pour ce faire, il est possible de généraliser l'indice de concentration (7) en travaillant sur un ensemble de sous-populations  $(\mathcal{X}_1, \dots, \mathcal{X}_m)$  d'une population  $\mathcal{T}$ . Le complément de  $(\mathcal{X}_1, \dots, \mathcal{X}_m)$  à  $\mathcal{T}$  est la population  $\mathcal{E}$ . L'expérience aléatoire  $W^{\mathcal{X}}$  décrite plus haut se généralise en une expérience aléatoire  $W^{\mathcal{X}_1, \dots, \mathcal{X}_m}$  dans laquelle l'individu est choisi parmi  $m + 1$  possibilités (les  $m$  sous-populations précédentes et la population complémentaire  $\mathcal{E}$ ) selon une loi de probabilité telle que :

$$\begin{cases} \mathbb{P}(W^{\mathcal{X}_1, \dots, \mathcal{X}_m} = i) &= \pi_i \\ \mathbb{P}(W^{\mathcal{X}_1, \dots, \mathcal{X}_m} = m + 1) &= 1 - \sum_{i=1}^m \pi_i \end{cases}$$

Avec ces notations, on peut définir un indice multi-populations par :

$$I^{\mathcal{X}_1, \dots, \mathcal{X}_m} = \frac{1}{\mathbb{H}(W^{\mathcal{X}_1, \dots, \mathcal{X}_m})} \left[ \mathbb{H}(\mathbf{X}^{\mathcal{T}}) - \left( \sum_{i=1}^m \pi_i \mathbb{H}(\mathbf{X}^{\mathcal{X}_i}) \right) - \left( 1 - \sum_{i=1}^m \pi_i \right) \mathbb{H}(\mathbf{X}^{\mathcal{E}}) \right] \quad (9)$$

où les variables aléatoires  $\mathbf{X}^{\mathcal{X}_i}$  modélisent les expériences aléatoires caractérisant le positionnement géographique d'un individu de la population  $\mathcal{X}_i$ , et  $\mathbb{H}(W^{\mathcal{X}_1, \dots, \mathcal{X}_m}) = - \sum_{i=1}^m \pi_i \ln(\pi_i) -$

$(1 - \sum_{i=1}^m \pi_i) \ln(1 - \sum_{i=1}^m \pi_i)$  est l'entropie de l'expérience aléatoire correspondant au premier niveau de choix (variable  $W^{\mathcal{X}_1, \dots, \mathcal{X}_m}$ ).

L'indice  $I^{\mathcal{X}_1, \dots, \mathcal{X}_m}$  présente les mêmes caractéristiques que  $I^{\mathcal{X}}$  : il est compris dans l'intervalle  $[0, 1]$  ; plus l'indice est proche de 1, plus les sous-populations d'intérêt sont concentrées (donc ségréguées dans une problématique de ségrégation) ; plus l'indice est proche de 0, moins les sous-populations sont concentrées par rapport à la population totale.

De même que précédemment, il peut également s'écrire :

$$I^{\mathcal{X}_1, \dots, \mathcal{X}_m} = 1 - \frac{\int f^{\mathcal{T}} \left[ -\sum_{i=1}^m p^{\mathcal{X}_i/\mathcal{T}} \ln(p^{\mathcal{X}_i/\mathcal{T}}) - (1 - \sum_{i=1}^m p^{\mathcal{X}_i/\mathcal{T}}) \ln(1 - \sum_{i=1}^m p^{\mathcal{X}_i/\mathcal{T}}) \right]}{\mathbb{H}(W^{\mathcal{X}_1, \dots, \mathcal{X}_m})}$$

**Estimer les indices de concentration** Estimer l'indice  $I^{\mathcal{X}}$  est aisé si l'on dispose déjà d'estimateurs pour les densités de population et pour le ratio. L'indice de concentration s'estime empiriquement par :

$$\hat{I}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}} = 1 - \frac{1}{-\hat{\pi} \ln(\hat{\pi}) - (1 - \hat{\pi}) \ln(1 - \hat{\pi})} \sum \hat{f}_{\mathbf{H}^{\mathcal{T}}}^{\mathcal{T}} \left[ -\hat{p}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}/\mathcal{T}} \ln(\hat{p}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}/\mathcal{T}}) - (1 - \hat{p}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}/\mathcal{T}}) \ln(1 - \hat{p}_{\mathbf{H}^{\mathcal{X}}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}/\mathcal{T}}) \right]$$

avec  $\hat{\pi} = N^{\mathcal{X}}/N^{\mathcal{T}}$ , la proportion de la population appartenant au groupe  $\mathcal{X}$ .

La démarche est identique pour l'indice multigroupe qui s'estime par :

$$\hat{I}_{\mathbf{H}^{\mathcal{X}_1, \dots, \mathcal{X}_m}, \mathbf{H}^{\mathcal{X}_m, \mathcal{H}^{\mathcal{T}}}}^{\mathcal{X}_1, \dots, \mathcal{X}_m} = 1 - \frac{\sum \hat{f}_{\mathbf{H}^{\mathcal{T}}}^{\mathcal{T}} \left[ -\sum_{i=1}^m \hat{p}_{\mathbf{H}^{\mathcal{X}_i}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}_i/\mathcal{T}} \ln(\hat{p}_{\mathbf{H}^{\mathcal{X}_i}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}_i/\mathcal{T}}) - (1 - \sum_{i=1}^m \hat{p}_{\mathbf{H}^{\mathcal{X}_i}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}_i/\mathcal{T}}) \ln(1 - \sum_{i=1}^m \hat{p}_{\mathbf{H}^{\mathcal{X}_i}, \mathbf{H}^{\mathcal{T}}}^{\mathcal{X}_i/\mathcal{T}}) \right]}{-\sum_{i=1}^m \hat{\pi}_i \ln(\hat{\pi}_i) - (1 - \sum_{i=1}^m \hat{\pi}_i) \ln(1 - \sum_{i=1}^m \hat{\pi}_i)}$$

avec, pour tout  $i$ ,  $\hat{\pi}_i = N^{\mathcal{X}_i}/N^{\mathcal{T}}$ , la proportion de la population appartenant au groupe  $\mathcal{X}_i$ .

### 3 Résultats : l'évolution de la localisation des immigrés en France

Dans cette section, nous appliquons les méthodes détaillées dans la section précédente pour analyser la localisation et la concentration géographique des principaux groupes d'origines, pour les cinq années de recensement : 1968, 1975, 1982, 1990 et 1999. L'ensemble des traitements de la base de données Saphir, des bases de données géographiques, le calcul des densités de population et des indices de concentration ainsi que le tracé des cartes ont été réalisés grâce au logiciel R et à ses paquetages mapprotools, spatstat et ks<sup>15</sup>

#### 3.1 L'évolution des proportions d'immigrés dans la population

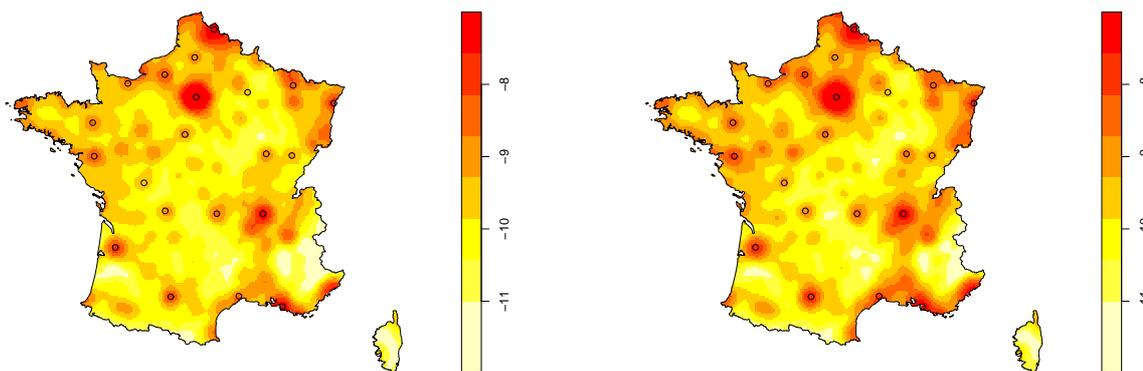
**Distribution géographique de la population totale sur le territoire** Nous présentons dans un premier temps, dans la figure 2, la distribution géographique de la population totale en

---

<sup>15</sup>cf. R Development Core Team (2007), Lewin-Koh et Bivand (2008), Baddeley et Turner (2005) et Duong (2008).

1968 et 1999.<sup>16</sup> À première vue, les cartes se ressemblent. Un examen attentif montre la nette diminution du poids des zones rurales au profit des agglomérations et des zones littorales, à l'exception des Alpes.

FIG. 2 – Log-densité de la population totale en 1968 (gauche) ; en 1999 (droite))

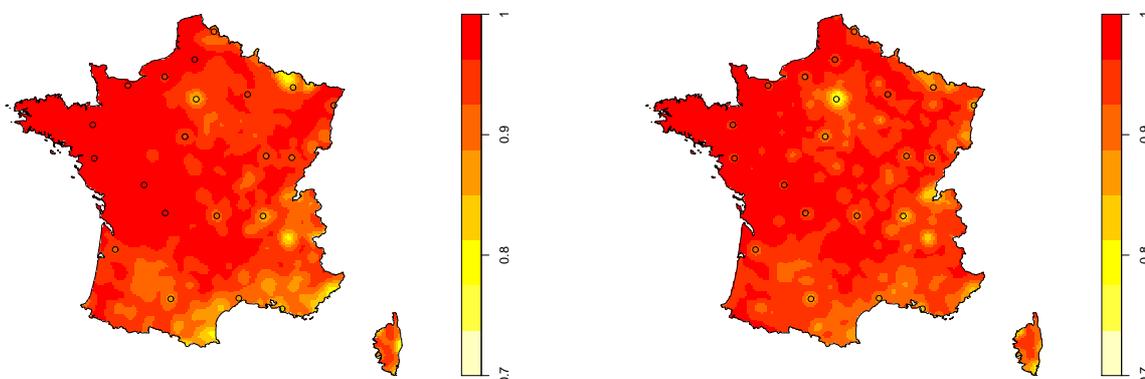


Source : Saphir (Insee).

**La distribution relative des Français de naissance** La proportion locale de Français de naissance dans la population totale est présentée dans la figure 3 pour les années de recensement 1968 et 1999. Bien que les différences entre les deux figures soient multiples, elles partagent des caractéristiques communes. D'abord, le ratio oscille partout entre 70 % et 100 %. Ensuite, les valeurs les plus importantes sont concentrées dans le nord-ouest et le centre du territoire ; dans ces zones, la probabilité qu'un habitant soit français de naissance vaut presque 1. Parmi les zones où les Français de naissance sont relativement moins nombreux, on retrouve : l'arc Méditerranéen, l'axe Bordeaux-Toulouse et les agglomérations de Paris, Lyon, Grenoble, Genève et Metz. En 1968, on remarque que les valeurs les plus basses de ce rapport sont localisées le long de la côte Méditerranéenne, de la frontière du Luxembourg et de l'Italie. En 1999, les régions où l'on rencontre le moins de Français de naissance sont l'Île-de-France, la frontière luxembourgeoise et le Genevois. Ponctuellement, des zones d'accueil se sont développées comme le long de la vallée du Rhône.

<sup>16</sup>Les densités sont présentées ici en logarithme, pour permettre la lisibilité malgré l'écart important entre zones urbaines et rurales. Les cercles évidés représentés sur toutes les cartes correspondent aux chefs-lieux de régions administratives.

FIG. 3 – Proportion locale de Français de naissance dans la population totale en 1968 (gauche) ; en 1999 (droite)



Source : Saphir (Insee).

Malgré cette stabilité apparente de la distribution géographique des Français de naissance sur trente ans, et donc de celle des immigrés, il est probable que ces tendances générales cachent des différences entre les groupes d'origines. Comme mentionné dans les précédentes études sur le sujet (cf. *supra*), ces différences sont susceptibles de provenir de plusieurs facteurs : la proximité du pays d'origine, la cohorte d'immigration, les causes d'immigration et, finalement, les caractéristiques socio-économiques des immigrés. Nous illustrons ceci dans la suite en concentrant l'étude sur six groupes d'immigrés, parmi les plus importants en effectifs : Algériens, Portugais, Marocains, Turcs, Espagnols, Italiens.

**La distribution relative des immigrés originaires d'Algérie ou du Portugal** Les immigrations algérienne et portugaise présentent quelques similitudes. Immigration de travail à l'origine, ces deux populations sont déjà massivement présentes en 1968. Cependant, elles connaissent un nouvel essor au cours de la période d'étude. La figure 4 présente la proportion locale des populations immigrées originaire d'Algérie et du Portugal dans la population totale, pour les années 1968, 1982 et 1999.

En dépit des similitudes, il existe de nombreuses différences quant à la localisation géographique des deux populations. Tout d'abord, sans doute pour des raisons de proximité géographique avec le Portugal, les immigrés d'origine portugaise vivant en France habitent plus souvent dans l'ouest du pays, en particulier sur la côte entre Bordeaux et les Pyrénées. Les immigrés algériens, quant à eux, sont majoritairement concentrés dans le sud-est, en Île de France

et sur la frontière belge. Au total, sur l'ensemble de la période, les immigrés portugais apparaissent plus régulièrement répartis que les algériens.

Ces derniers restent majoritairement concentrés autour de quelques grandes agglomérations : Paris, Lyon, Grenoble, Marseille, mais aussi à Alès ou le long des frontières allemande, luxembourgeoise et belge. À noter que si leur effectif total varie au cours des trente ans (en augmentation en 1968 et 1982, stable au-delà), leur répartition géographique change peu sur la période.

Les immigrés portugais, quant à eux, occupent quelques centres bien précis en 1968, comme la grande banlieue sud, est et ouest de Paris, l'agglomération de Clermont-Ferrand. Entre 1968 et 1982, leur présence augmente parallèlement à une diversification des espaces occupés et se localise en de multiples poches de concentration réparties partout en France, à l'exception de la Bretagne, de la Normandie et de l'arc méditerranéen. De ce point de vue, la répartition est plutôt plus équilibrée que celle d'autres groupes nationaux. Le recul constaté en 1999 du nombre d'immigrés d'origine portugaise laisse intact cet équilibre de leur distribution spatiale.

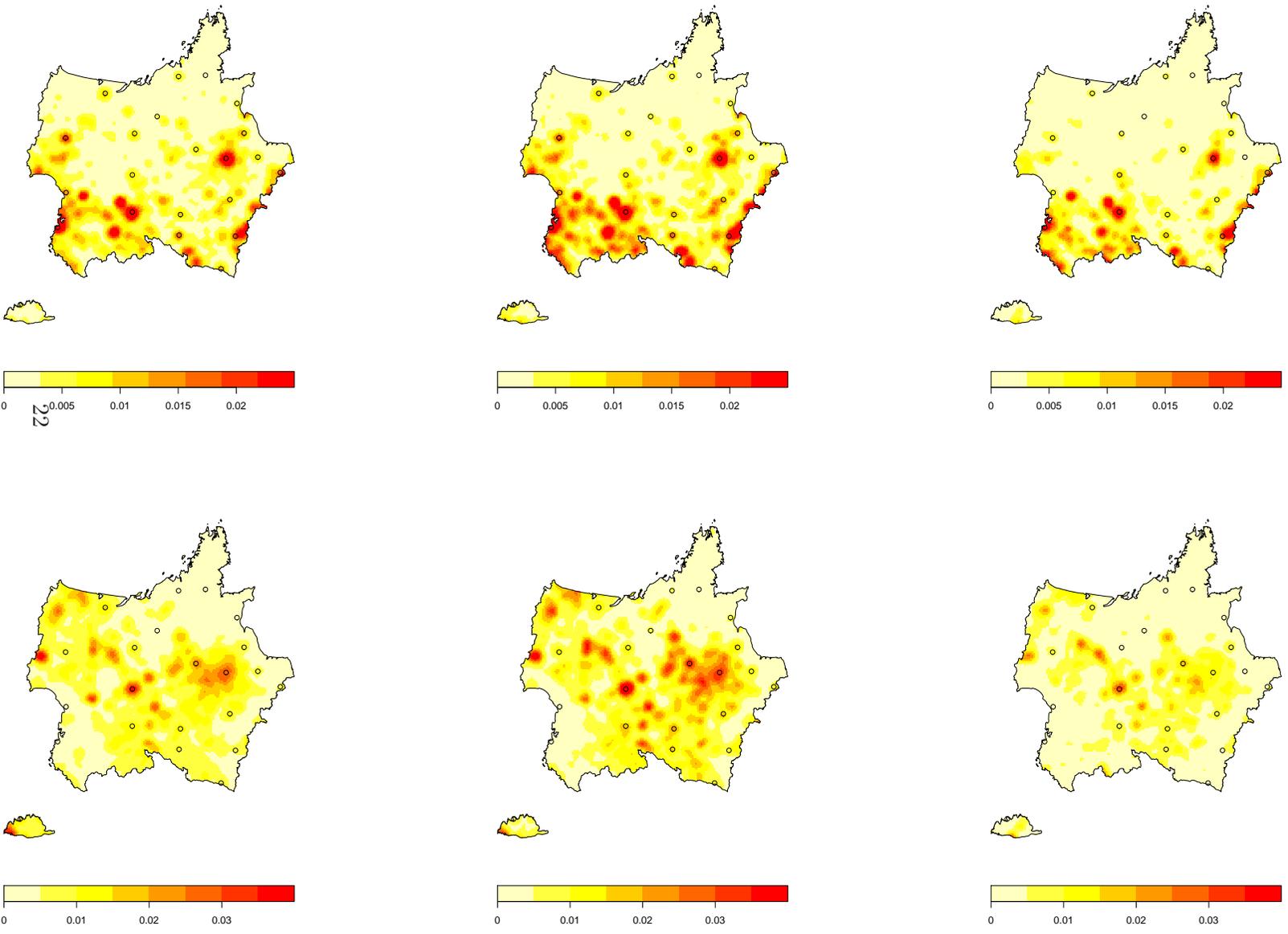
**La distribution relative des immigrés originaires du Maroc et de Turquie** Au contraire des immigrés algériens et portugais, les Marocains et les Turcs font partie de vagues migratoires plus récentes, prenant une réelle ampleur dans les années 1970 (début des années 1970 pour les Marocains, fin pour les Turcs). Leur nombre croît significativement entre 1982 et 1999. La figure 5 présente la proportion locale des immigrés marocains et turcs dans la population totale, pour les années 1968, 1982 et 1999.

En 1968, les deux populations sont quasiment absentes du territoire, à l'exception notable de la Corse, dans laquelle un nombre relativement important d'immigrés marocains s'est déjà installé. Cette particularité demeurera au cours des trente ans.

Entre 1968 et 1982, un nombre important de Marocains et de Turcs ont immigré. Leur localisation est toutefois très différente. Les immigrés turcs sont ainsi essentiellement présents le long des frontières allemande et suisse, ainsi que dans les régions grenobloise, parisienne et orléanaise. Les Marocains, outre la Corse, s'installent dans des zones assez lointaines les unes des autres : basse vallée du Rhône, vallée de la Garonne, Île-de-France (essentiellement en banlieue nord-ouest de Paris) et, globalement, autour de toutes les agglomérations de la moitié est du pays.

Entre 1982 et 1999, les effectifs des deux populations continuent d'augmenter, et les deux groupes semblent accentuer leur concentration, sur l'axe Montpellier-Montélimar et en Corse pour les Marocains, dans la région de Genève et le long de la frontière allemande pour les Turcs.

FIG. 4 – Proportion d'immigrés originaires d'Algérie (gauche) et du Portugal (droite) dans la population totale en 1968 (haut) ; en 1982 (milieu) et 1999 (bas)



En outre, la population immigrée turque, en 1982 et 1999, a tendance à être relativement plus présente dans les périphéries que dans les centres urbains (Paris, Lyon, Grenoble...).

**La distribution relative des immigrés d'origine espagnole et italienne** Les Italiens et les Espagnols proviennent d'une immigration plus ancienne. Tout au long de la période, leur nombre décroît de manière importante. Ceci s'accompagne d'une baisse de la concentration spatiale visible à la figure 6, qui présente la proportion locale d'immigrés espagnols et italiens sur le territoire français, en 1968, 1982 et 1999.

1968 est l'année où les deux groupes nationaux sont, des trois années représentées, les plus présents. Un élément particulièrement frappant pour ces deux populations est l'importance de la proximité du pays d'origine dans le choix de localisation. Ainsi les Espagnols sont-ils particulièrement concentrés dans une zone délimitée au nord par un axe Bordeaux-Montpellier, ainsi que dans une zone d'environ 100 à 200 km de large autour de l'arc méditerranéen. De manière moins évidente, ils se sont également installés dans la périphérie de Clermont-Ferrand et d'Orléans. Les Italiens se comportent de manière similaire, en s'installant en Corse et sur les flancs est des régions PACA et Rhône-Alpes. Ils sont également présents dans la région de Metz et dans une zone assez large autour de la vallée de la Garonne. Enfin, les deux groupes sont présents dans la périphérie de Paris et de Lyon.

En 1982 puis en 1999, la population de ces deux groupes chute. Les zones où ils demeurent restent les mêmes que précédemment, avec des concentrations plus manifestes à l'extrémité sud du Roussillon pour les Espagnols et dans la région de Nice et de Metz pour les Italiens.

**Comparaison avec des méthodes couramment utilisées** Agréger l'information par département n'est rien d'autre qu'un cas particulier de la méthode que nous utilisons. Au lieu de pondérer l'information en fonction de la distance au point d'intérêt, cela revient à assigner un poids unitaire (la matrice  $\mathbf{H}$  de l'équation 1) à toutes les observations d'un département et un poids nul à celles situées en dehors du département. On comprend qu'il s'agit d'une méthode d'autant plus arbitraire et sans fondement que les limites du département sont sans rapport avec les mécanismes de localisation des étrangers.

Nous comparons ici les résultats obtenus en utilisant notre méthode à ceux que l'on obtient par agrégation au département. Nous prenons l'exemple de la proportion des immigrés marocains dans la population totale en 1999. Tout d'abord, nous examinons les distributions par commune et par département. La première constatation est que les immigrés d'origine marocaine sont présents dans environ 12% des 36500 communes du territoire métropolitain. Si l'on restreint aux 4491 communes comptant au moins un immigré marocain en 1999, les quantiles

FIG. 5 – Proportion de d'immigrés originaires du Maroc (gauche) et de Turquie (droite) dans la population totale en 1968 (haut) ; en 1982 (milieu) et 1999 (bas)

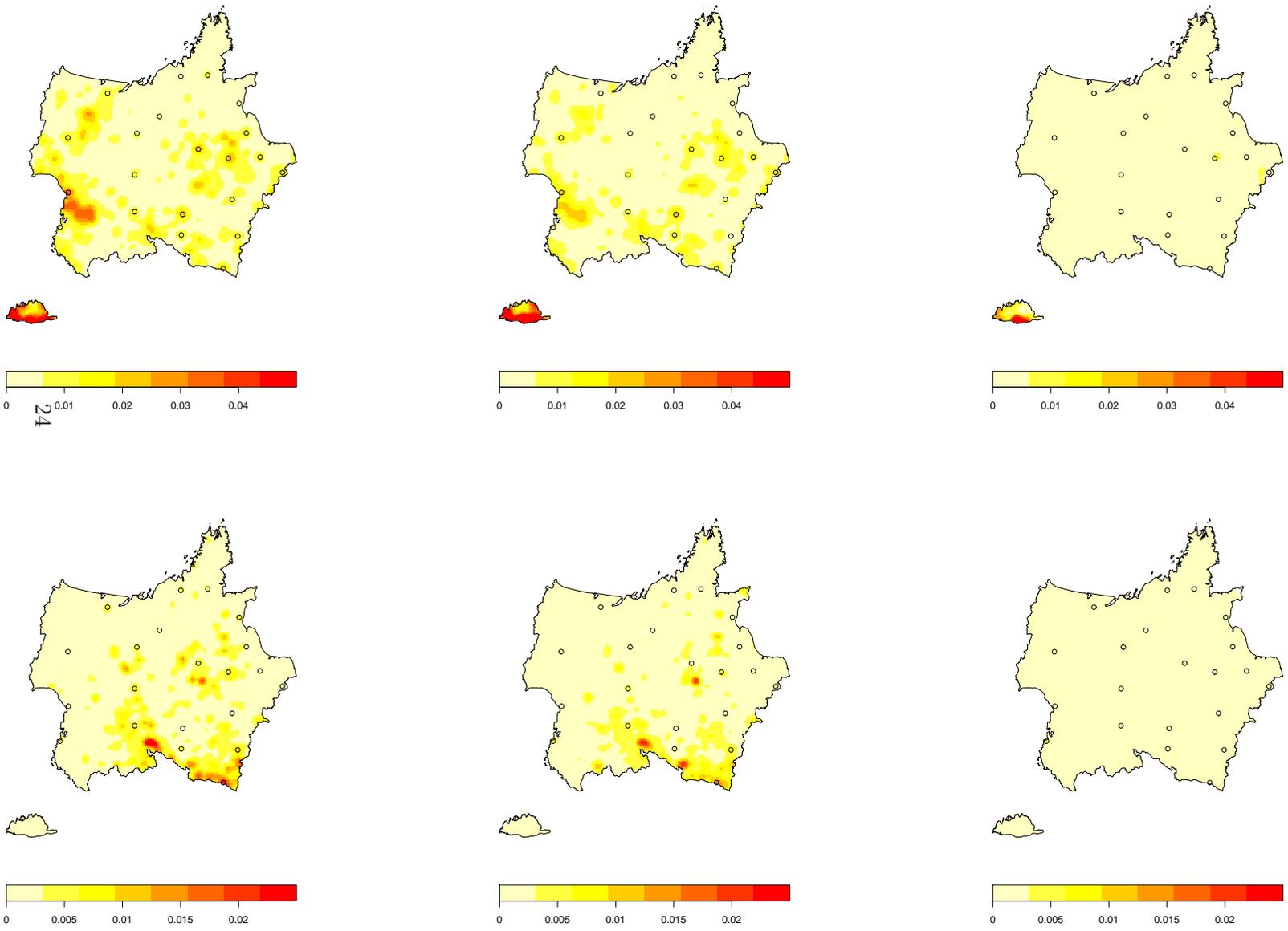
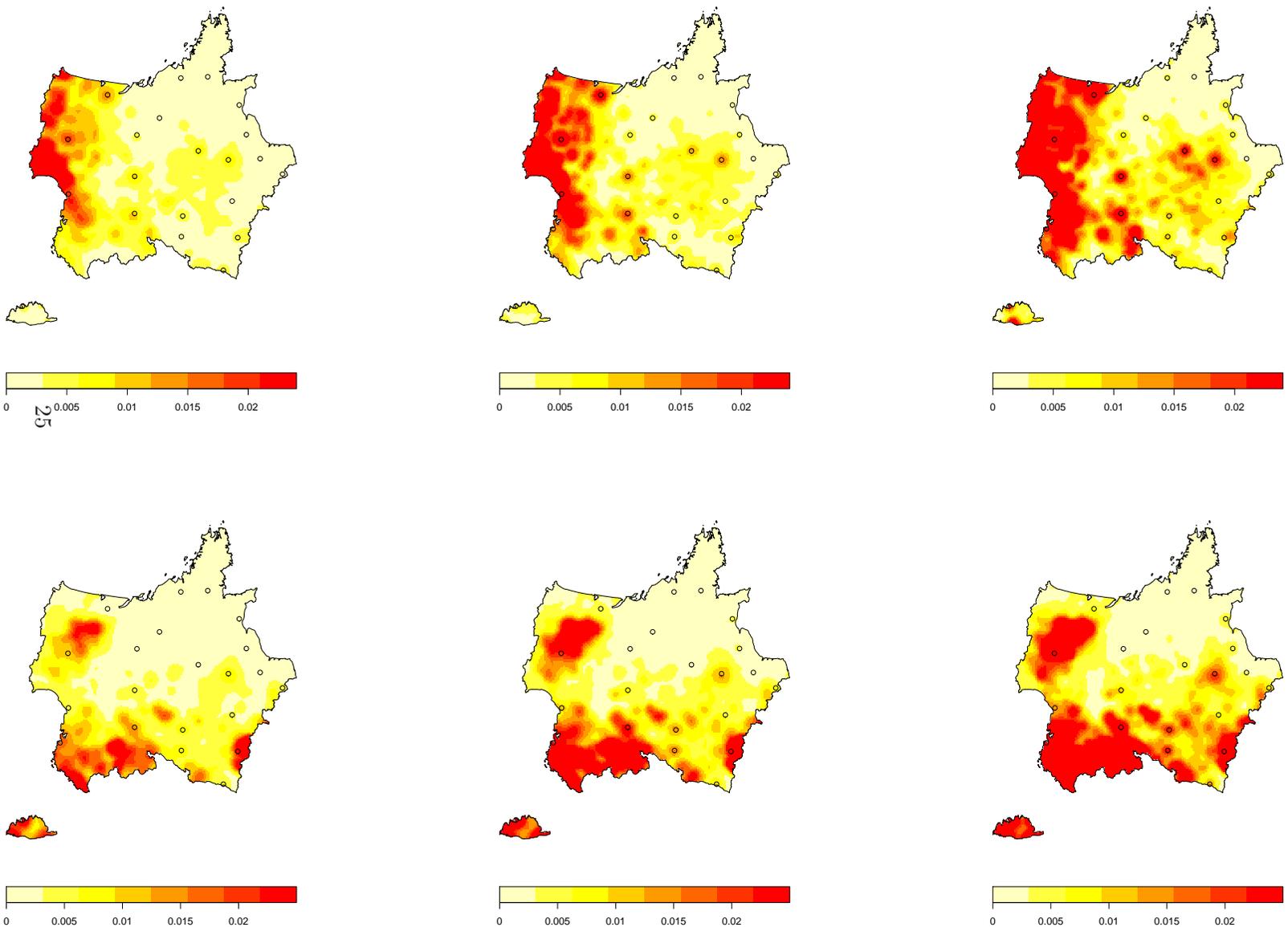


FIG. 6 – Proportion d'immigrés originaires d'Espagne (gauche) et d'Italie (droite) dans la population totale en 1968 (haut) ; en 1982 (milieu) et 1999 (bas)



de la distribution du ratio sont donnés dans le tableau 1.

TAB. 1 – Rapport entre population marocaine et la population française (de naissance) en 1999 : quartiles de distribution

Échelle	Min	$Q_{25\%}$	Médiane	$Q_{75\%}$	Max	N
Commune	0.0000	0.0000	0.0000	0.0000	1.216	36541
Commune <sup>1</sup>	0.0001	0.0036	0.0089	0.0214	1.216	4491
Département	0.0004	0.0037	0.0069	0.0116	0.0614	96

Source : Saphir (Insee).

(<sup>1</sup>) La deuxième ligne correspond aux communes comptant au moins un Marocain dans le recensement de 1999.

Dans la figure 7, on compare les distributions géographiques obtenues à partir des méthodes à noyaux et celle qui consiste à calculer une moyenne par département. Dans la méthode que nous préconisons, nous calculons chacune des densités à partir des données communales, avant d’en faire le rapport, point par point. Dans la méthode par département, nous agrégeons les effectifs de chaque population par département puis nous effectuons le rapport. Les deux cartes en haut de la figure 7 montrent bien l’apport de la méthode à noyaux pour visualiser les grandes tendances d’une répartition géographique, tout en permettant d’identifier les lieux de concentration à des échelles fines : l’agrégation départementale occulte en partie des concentrations identifiables sur la cartographie issue de la méthode des noyaux.

Assez logiquement, la méthode départementale a aussi tendance à atténuer les contrastes. Un premier exemple concerne la Corse : les immigrés originaires du Maroc y sont manifestement concentrés dans les communes du littoral de l’île, ce dont l’approche départementale ne peut rendre compte. Un autre exemple est celui de l’Hérault et du Gard, deux départements dans lesquels la proportion d’immigrés originaires du Maroc semble particulièrement importante. En réalité, la méthode par noyaux fait apparaître qu’il s’agit essentiellement d’un phénomène lié aux agglomérations de Montpellier et Nîmes. Ces exemples illustrent de manière assez claire la non pertinence du zonage géographique départemental pour l’étude du phénomène considéré.

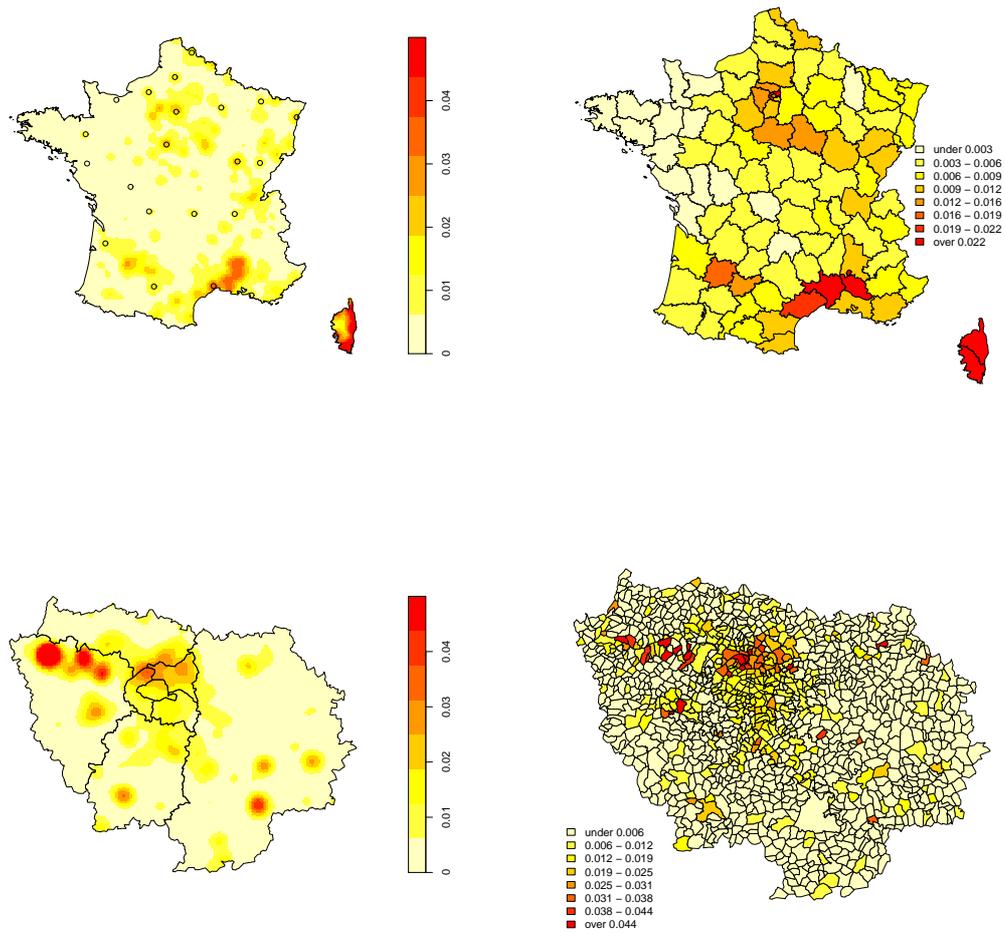
Sur un plan technique, la comparaison des tracés départementaux et national par noyaux mettent en évidence deux points.

- Au niveau communal, la valeur zéro pour le ratio est sur-représentée. Il s’agit d’un artefact statistique dû à l’échantillonnage et au fait que la population d’intérêt est une population rare. Pour remédier à ce problème, l’approche par noyau tire parti de l’information disponible dans les communes avoisinantes ; l’approche par département fait la moyenne au niveau de chaque département et globalise l’information.
- La figure 7 confirme l’existence d’une forte hétérogénéité infra-départementale. Ainsi, de

fortes valeurs du ratio au niveau départemental peuvent cacher deux phénomènes démographiquement très différents : des valeurs locales moyennes mais fréquentes ou un fort pic localisé. L'approche par noyau, contrairement aux moyennes départementales permet de distinguer ces deux hypothèses.

La figure 7 illustre également l'apport de la méthode par noyaux par rapport à une cartographie représentant directement les valeurs communales : la représentation lissée permet de visualiser les dominantes tout en évitant l'écueil d'une information fortement bruitée, semblable à une mosaïque sans structure, qui s'avère finalement peu lisible et biaisée. Par exemple, à l'extrême ouest de l'Île de France, la représentation par la méthode des noyaux met bien en évidence le pôle de Mantes-la-Joie et ses environs qui ressort de manière beaucoup moins nette de la représentation directe des informations communales.

FIG. 7 – Proportion locale de Marocains dans la population totale en 1999. Haut : méthode par noyau (gauche), agrégation par département (droite); Bas : méthode par noyau (gauche), commune (droite) en Île de France



Source : Saphir (Insee).

Note : l'échelle de couleurs des deux cartes du haut est différente; elle est identique dans les deux cartes du bas.

### 3.2 Mesurer les asymétries dans la répartition spatiale des immigrés sur le territoire

Le tableau 2 présente la valeur de l'indice de concentration détaillé à la sous-section 2.4, par pays d'origine pour l'année 1999, selon que l'on utilise les données issues du calcul par une méthode à noyaux ou celles issues d'une agrégation départementale. La théorie prévoit que le calcul à partir de données départementales est susceptible de biaiser l'indice à la baisse. Ainsi, si la population est très concentrée au sein des départements, arrondir au niveau des départements va sous-estimer la dissymétrie de la distribution par rapport à une population mieux répartie au sein des départements.

Dans le tableau 2, on constate que, pour toutes les nationalités, l'indice sur données continues est systématiquement plus élevé que celui obtenues à partir des données départementales.<sup>17</sup> Ainsi, comme prévu, l'indice départemental sous-estime la concentration des populations étrangères étudiées. La sous-estimation est d'ailleurs supérieure aux variations de concentration entre deux recensements (voir tableau 3), ce qui témoigne de la faiblesse de l'analyse départementale.<sup>18</sup> L'écart entre les deux indices a cependant une variance assez forte, notamment pour les populations les plus réduites. Dans certains cas, même le rang n'est pas respecté ; c'est le cas de l'Asie du Sud-Est, par exemple.

Les résultats du tableau 2 permettent de faire émerger quatre groupes de pays, selon le niveau de leur indice de concentration en 1999. Il est intéressant de constater que ces groupes transgressent largement les rapprochements que l'on opère habituellement entre les groupes d'immigrés. Ainsi, les plus concentrés sont les immigrés originaires d'Afrique subsaharienne, d'Asie-Océanie et de Tunisie. Très concentrés également, mais dans une moindre mesure, viennent les immigrés d'Asie du Sud-Est, d'Espagne, de Turquie, d'Amérique du Sud, d'Italie et d'Algérie. Les immigrés originaires d'Amérique du Nord et du Portugal sont moins concentrés que ces derniers. Les groupes ayant les indices de concentration les plus bas viennent d'Europe de l'Est, d'Europe de l'Ouest et du Maroc.

Le tableau 3 montre l'évolution de l'indice continu pour chaque nationalité entre les recensements de 1968 et 1999, ainsi que celle de l'indice multigroupe. Si l'on se réfère à l'indice multigroupe, la concentration géographique semble diminuer entre 1968 et 1975, puis se stabilise au-delà.<sup>19</sup> Le constat d'une diminution de la concentration pourrait rejoindre l'idée théorique selon laquelle les immigrés font des choix de localisation de moins en moins stéréotypés. En fait,

---

<sup>17</sup>Rappelons que plus l'indice de concentration est élevé, plus la population est concentrée. L'indice est compris dans l'intervalle  $[0, 1]$ .

<sup>18</sup>Le calcul des indices au niveau communal est impossible, étant donné que la proportion d'immigrés dans ces communes est fréquemment égal à zéro. L'indice que nous utilisons n'est pas défini dans ce cas.

<sup>19</sup>Voir également Safi (2007) pour un panorama de l'évolution de la concentration résidentielle des immigrés en France

la spécialisation des lieux par nationalité tend donc d'abord à se réduire, puis à se stabiliser.

Ce constat doit cependant être nuancé selon la nationalité. Pour la majorité des groupes étudiés, Espagne, Italie, Europe de l'Est, Europe de l'Ouest, Amérique du Nord, on constate une décroissance assez régulière, tout au long de la période de l'indice de concentration. Pour d'autres, Algérie, Maroc, Tunisie, Turquie, la baisse a lieu principalement en 1968 et 1975. Au cours de vingt-cinq ans qui suivent, la concentration semble plutôt stagner. Pour les immigrés d'origine portugaise, ou d'Afrique subsaharienne, l'indice n'évolue pas sensiblement tout au long de la période. La concentration des immigrés d'Asie du Sud-Est présente un profil atypique : une augmentation importante jusqu'en 1990 et une baisse significative de 1990 à 1999.

Ces résultats ne doivent pas être interprétés comme des résultats sur la ségrégation *stricto sensu*. La ségrégation résidentielle est un phénomène se déroulant à une échelle beaucoup plus fine, celle d'une commune ou même d'un quartier. Les données que nous exploitons ici permettent de s'intéresser à la répartition spatiale à l'échelle du territoire national, d'une région, d'un département ou d'un groupe de communes, mais pas d'étudier directement la ségrégation. La distinction est importante : il est très possible qu'alors qu'une population est mieux répartie dans l'ensemble sur le territoire, elle soit localement plus ségrégée dans certains pôles urbains ou même certains quartiers. Des données infra-communales seraient nécessaires pour pousser l'analyse à ce niveau.

TAB. 2 – Indice entropique par nationalité selon l'unité élémentaire, en 1999

	Unité élémentaire	
	noyau	département
Afrique Subsaharienne	0.127	0.093
Algérie	0.074	0.061
Amérique du Nord	0.067	0.062
Amérique du Sud	0.077	0.061
Asie du Sud Est	0.086	0.058
Espagne	0.079	0.076
Europe de l'Est	0.049	0.040
Europe de l'Ouest	0.036	0.034
France	0.077	0.059
Italie	0.076	0.069
Maroc	0.049	0.040
Asie, Océanie	0.143	0.103
Portugal	0.060	0.053
Tunisie	0.113	0.092
Turquie	0.078	0.063

Source : Saphir (Insee).

Note : Indice entropique à deux groupes, entre le groupe étudié et le reste de la population recensée en 1999.

TAB. 3 – Indice entropique par nationalité et par année de recensement (noyau)

	Année de recensement				
	1968	1975	1982	1990	1999
France	0.079	0.071	0.075	0.076	0.077
Afrique Subsaharienne	0.136	0.136	0.147	0.130	0.127
Algérie	0.099	0.075	0.077	0.076	0.074
Tunisie	0.159	0.113	0.118	0.118	0.113
Maroc	0.091	0.055	0.045	0.048	0.049
Turquie	0.115	0.060	0.075	0.073	0.078
Espagne	0.106	0.102	0.100	0.088	0.079
Italie	0.106	0.099	0.092	0.085	0.076
Portugal	0.059	0.065	0.066	0.062	0.060
Europe de l'Est	0.077	0.069	0.069	0.062	0.049
Europe de l'Ouest	0.055	0.047	0.041	0.037	0.036
Amérique du Nord	0.129	0.112	0.093	0.084	0.067
Amérique du Sud	0.110	0.122	0.127	0.105	0.077
Asie du Sud Est	0.075	0.093	0.096	0.101	0.086
Asie, Océanie	0.118	0.115	0.139	0.133	0.143
Multigroupe	0.095	0.084	0.085	0.082	0.079

Source : Saphir (Insee).

Note : Indice entropique à deux groupes et multigroupe (dernière ligne).

## Références

- AUBRY, B. (1996) : “La mobilité interne des étrangers et des immigrés,” *Espace, Population et Sociétés*, 2–3, 299–303.
- BADDELEY, A., ET R. TURNER (2005) : “Spatstat : an R package for analyzing spatial point patterns,” *Journal of Statistical Software*, 12(6), 1–42.
- CHENU, A. (1996) : “Les étrangers dans les agglomérations françaises,” dans *Données urbaines 1*, ed. D. Pumain, et F. Godard, pp. 225–234. INSEE.
- DESPLANQUES, G. (1996) : “La répartition des personnes d’origine étrangère en France,” *Espace, Population et Sociétés*, 2–3, 287–297.
- DESPLANQUES, G., ET N. TABARD (1991) : “La localisation de la population étrangère,” *Economie et Statistique*, 242, 51–62.
- DEWITTE, P. (2003) : *Deux siècles d’immigration en France*. La Documentation Française.
- DUONG, T. (2008) : *ks : Kernel smoothing*, R package version 1.5.6.
- DUONG, T., ET M. L. HAZELTON (2003) : “Plug-in bandwidth matrices for bivariate kernel density estimation,” *Nonparametric statistics*, 15(1), 17–30.
- GEORGE, P. (1986) : “Les étrangers en France : étude géographique,” *Annales de géographie*, 95(529), 273–300.
- GOBILLON, L., ET H. SELOD (2007) : “The Effect of Segregation and Spatial Mismatch on Unemployment : Evidence from France,” CEPR Working Paper 6198.
- GRANNIS, R. (2002) : “Discussion : Segregation Indices and their Functional Inputs,” *Sociological Methodology*, 32.
- HARNOIS, J. (1992) : “La localisation des étrangers en France,” *INSEE Première*, 177.
- HUTCHENS, R. (2001) : “Numerical Measures of Segregation : Desirable Properties and their Implications,” *Mathematical Social Sciences*, 42, 13–29.
- INSEE (ed.) (2005) : *Les immigrés en France*, Collection Références. Insee, 2005 edn.
- JAMES, D. R., ET K. E. TAEUBER (1985) : “Measures of Segregation,” *Sociological Methodology*, 14, 1–32.
- JAYET, H., ET N. UKRAYINCHUK (2007) : “La localisation des immigrants en France : Une première approche,” *Revue d’Économie Régionale et Urbaine*, 2007(4), 625–649.
- KELSALL, J. E., ET P. J. DIGGLE (1998) : “Spatial Variation in Risk : a Non-Parametric Binary Regression Approach,” *Applied Statistics*, 47, 559–573.
- KENDALL, M. G. (1973) : “Entropy, Probability and Information,” *International Statistical Review*, 47(1), 59–68.
- KULLBACK, S. (1959) : *Information Theory and Statistics*. Ed. Dover.
- LEWIN-KOH, N. J., ET R. BIVAND (2008) : *maptools : Tools for reading and handling spatial objects*, R package version 0.7-15.

- MASSEY, D. S., ET N. A. DENTON (1988) : “The Dimensions of Residential Segregation,” *Social Forces*, 67(2).
- MAUCO, G. (1932) : *Les étrangers en France : Leur rôle dans la vie économique*. A. Colin, Paris.
- NOIRIEL, G. (1988) : *Le Creuset français : Histoire de l’immigration XIXe-XXe siècle*. Le Seuil, Paris.
- OPENSHAW, S., ET P. TAYLOR (1979) : “A million or So Correlated Coefficients : Three Experiments on the Modifiable Areal Unit Problem,” dans *Statistical Applications in the Spatial Sciences*, ed. N. Wrigley, et B. R.J. Pion (London).
- R DEVELOPMENT CORE TEAM (2007) : *R : A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- REARDON, S. F., ET G. FIREBAUGH (2002a) : “Measures of Multigroup Segregation,” *Sociological Methodology*, 32, 33–67.
- (2002b) : “Response : Segregation and Social Distance : A Generalized Approach to Segregation Measurement,” *Sociological Methodology*, 32.
- REARDON, S. F., ET D. O’SULLIVAN (2004) : “Measures of Spatial Segregation,” *Sociological Methodology*, 34, 121–162.
- SAFI, M. (2007) : “Le devenir des immigrés en France. Barrières et inégalités,” Thèse de doctorat EHESS.
- SHANNON, C. E. (1948) : “A Mathematical Theory of Communication,” *The Bell System Technical Journal*, 27, 379–423, 623–656.
- SILVERMAN, B. W. (1986) : *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
- SIMON, P. (1996) : “Espace de vie des immigrés en France : une typologie,” *Espace, Population et Sociétés*, 2–3, 305–314.
- WAND, M. P., ET M. C. JONES (1995) : *Kernel Smoothing*. Chapman & Hall.
- WONG, D. W. (1993) : “Spatial Indices of Segregation,” *Urban Studies*, 30, 559–572.
- (2005) : “Formulating a General Spatial Segregation Measure,” *The Professional Geographer*, 57, 285–294.

## Annexe : Propriétés de l'entropie de Shannon

Le concept est construit à partir d'une variable aléatoire discrète mais se généralise sans difficulté à une variable continue conformément à la formulation (4). On considère une expérience aléatoire modélisée par une variable  $\mathbf{Z}$  prenant  $K$  valeurs différentes  $\{1, \dots, K\}$  de sorte que  $\mathbb{P}(\mathbf{Z} = k) = p_k$  avec  $\sum_{k=1}^K p_k = 1$ . Shannon définit trois propriétés que devrait vérifier une mesure  $\mathbb{H}(\mathbf{Z})$  caractérisant le taux d'information moyen apporté par une observation, les probabilités  $p_k$  étant connues :

- i.  $\mathbb{H}(\mathbf{Z})$  est une fonction continue des  $p_k$ .
- ii. Si tous les résultats possibles de l'expérience sont équiprobables (i.e.  $p_k = \frac{1}{K}$ ), alors  $\mathbb{H}(\mathbf{Z})$  est une fonction croissante de  $K$ .
- iii. Si le résultat de l'expérience est issu d'une série de choix imbriqués, alors  $\mathbb{H}(\mathbf{Z})$  ( $\mathbf{Z}$  se rapporte au résultat final) est une somme pondérée des quantités  $\mathbb{H}$  calculées à chaque niveau de choix, la pondération des niveaux de choix s'effectuant selon leur probabilité de survenue.

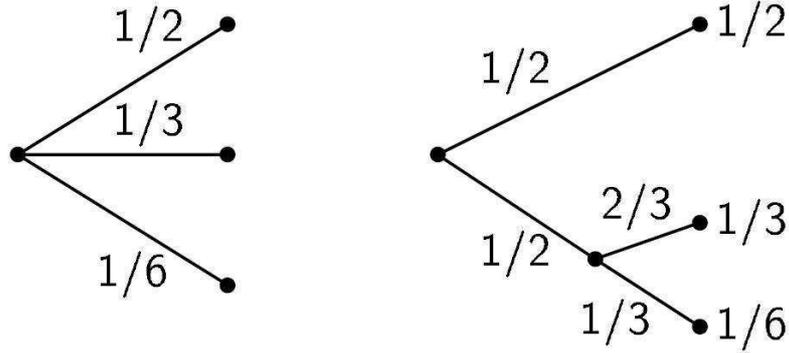
Shannon (1948) montre que la seule mesure qui vérifie ces trois propriétés est l'entropie définie par (à une constante près) :

$$\mathbb{H}(\mathbf{Z}) = - \sum p_k \ln(p_k) \tag{10}$$

On note que  $\mathbb{H}(\mathbf{Z}) = -\mathbb{E}(\ln(\mathbf{Z}))$ , ce qui permet de généraliser la définition au cas d'une variable continue.

Revenons sur les axiomes initiaux : la première propriété (i) est une propriété technique qui permet d'asseoir la mesure sur des critères qui vont rendre les différents états de probabilité continûment comparables. La deuxième propriété (ii) est reliée au fait que, toutes choses égales par ailleurs, plus le nombre d'états possibles est grand, plus la prédiction est incertaine. La dernière propriété (iii) est une propriété d'additivité de l'entropie par rapport à une séquence de choix successifs aboutissant au résultat.

Afin de mieux comprendre cette dernière propriété, prenons l'exemple proposé par Shannon de deux expériences conduisant à trois résultats possibles de même nature selon deux procédés de tirages différents :



La première expérience consiste à tirer parmi trois résultats possibles ; la seconde consiste à choisir, dans un premier temps, entre le premier résultat et le groupe des deux autres puis, si le second a été retenu, à choisir entre les deux derniers résultats. Dans les deux expériences, chacun des trois résultats a la même probabilité d'être retenue. L'entropie de la première est  $\mathbb{H}(\frac{1}{2}, \frac{1}{3}, \frac{1}{6})$ . L'entropie de la deuxième expérience se décompose en :

- l'entropie au premier choix :  $\mathbb{H}(\frac{1}{2}, \frac{1}{2})$  ;
- l'entropie au deuxième choix si le second groupe de résultats est retenu :  $\mathbb{H}(\frac{2}{3}, \frac{1}{3})$ .

La propriété d'additivité impose qu'il existe  $\alpha$  et  $\beta$  tels que

$$\mathbb{H}(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = \alpha \mathbb{H}(\frac{1}{2}, \frac{1}{2}) + \beta \mathbb{H}(\frac{2}{3}, \frac{1}{3})$$

Le premier choix auquel correspond l'entropie  $\mathbb{H}(\frac{1}{2}, \frac{1}{2})$  se produit dans tous les cas, donc  $\alpha = 1$ . Le second choix auquel correspond l'entropie  $\mathbb{H}(\frac{2}{3}, \frac{1}{3})$  se produit dans la moitié des cas, donc  $\beta = \frac{1}{2}$ .

Shannon considère cette propriété comme "naturelle" pour une fonction d'information (elle est aussi vérifiée par les informations de Fisher et Kullback), même si d'autres statisticiens ne la perçoivent pas comme telle (voir à ce propos la discussion de Kendall (1973)). Toujours est-il qu'il est assez remarquable que ces trois propriétés, somme toute relativement simples, conduisent à une unique notion mathématique : l'information au sens de Shannon ou entropie.

*L'entropie mesure le niveau d'aléa caractéristique de l'expérience aléatoire considérée. C'est un taux moyen, c'est-à-dire qu'il caractérise le niveau d'aléa associé à une seule réalisation de l'expérience, indépendamment du résultat obtenu. Si certains résultats possibles sont très probables et d'autres moins, le niveau d'aléa est plus faible que si tous les événements possibles sont équiprobables. Du point de vue de la densité de probabilité, l'existence d'événements plus probables que d'autres se traduit par des modes accentués. L'entropie ainsi définie, s'apparente donc à une mesure de non-concentration.*

**Quelques exemples d'entropies** Nous considérons trois expériences aléatoires : le cas d'une variable binaire, le cas d'une variable gaussienne et le cas d'une variable uniforme par morceaux.

- $\mathbf{Z} \in \{0, 1\}$  avec  $\mathbb{P}(\mathbf{Z} = 1) = p$ . Alors

$$\mathbb{H}(\mathbf{Z}) = -p \ln p - (1 - p) \ln(1 - p)$$

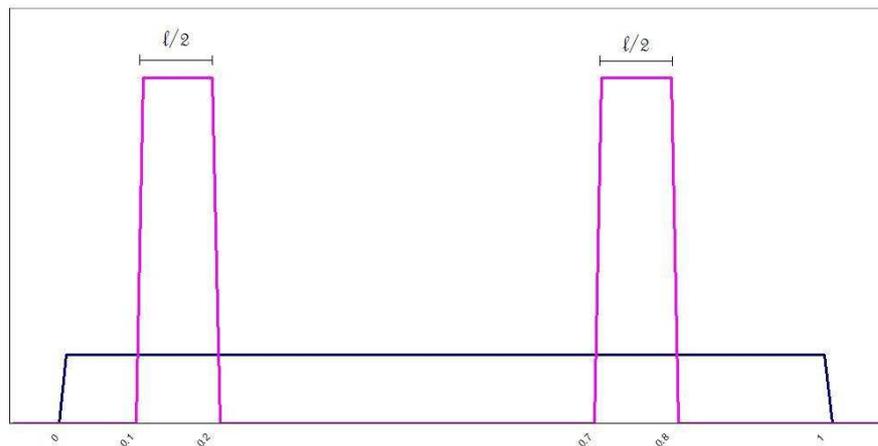
En application de la relation (10), on montre que  $\mathbb{H}(\mathbf{Z})$  est maximale (i.e. l'aléa associé à l'expérience  $\mathbf{Z}$  est maximal) lorsque  $p = \frac{1}{2}$ , ce qui correspond à la situation où les deux événements élémentaires possibles sont équiprobables.

- $\mathbf{Z}$  est une variable aléatoire gaussienne de moyenne quelconque et de variance  $\sigma^2$ . En application de la relation (4), on montre que

$$\mathbb{H}(\mathbf{Z}) = \ln(\sigma\sqrt{2\pi}) + \frac{1}{2}$$

Plus  $\sigma$  est petit, plus le mode principal de la densité est accentué et plus  $\mathbb{H}(\mathbf{Z})$  est faible.

- Considérons, d'une part une variable  $\mathbf{Z}^{(1)}$  suivant une loi uniforme sur  $[0, 1]$  et d'autre part, une variable  $\mathbf{Z}^{(2)}$  uniforme sur des sous-intervalles de  $[0, 1]$  de longueur totale  $\ell < 1$  (voir figure ci-dessous).



Alors, on montre que  $\mathbb{H}(\mathbf{Z}^{(1)}) = 0$  et  $\mathbb{H}(\mathbf{Z}^{(2)}) = \ln(\ell)$ . Ainsi, plus la variable comporte de mode(s) accentué(s) (i.e. plus  $\ell$  est petite), plus son entropie est faible.

## Annexe : Propriétés de l'indice (7)

### Démonstration de l'inégalité 6

Montrons tout d'abord que  $I^{\mathcal{X}}$  défini par (7) vérifie l'inégalité  $I^{\mathcal{X}} \leq 1$  et que  $I^{\mathcal{X}} = 1$  si et seulement si

$$\forall \mathbf{x} \in \mathfrak{F}, \min \{f^{\mathcal{X}}(\mathbf{x}), f^{\mathcal{E}}(\mathbf{x})\} = 0$$

Par construction, pour tout  $\mathbf{x} \in \mathfrak{F}$ ,  $\pi f^{\mathcal{X}}(\mathbf{x}) + (1 - \pi)f^{\mathcal{E}}(\mathbf{x}) = f^{\mathcal{T}}(\mathbf{x})$ . On considère la fonction

$$\left\| \begin{array}{l} \varphi : \mathbb{R}_*^+ \rightarrow \mathbb{R} \\ x \mapsto \varphi(x) = -x \ln(x) \end{array} \right.$$

On note que  $\varphi$  ainsi définie est prolongeable par continuité en 0 :  $\varphi(0) = 0$ .  $\varphi$  est une fonction convexe. Il en découle que

$$\varphi(\pi f^{\mathcal{X}}(\mathbf{x})) + \varphi((1 - \pi)f^{\mathcal{E}}(\mathbf{x})) \geq \varphi(f^{\mathcal{T}}(\mathbf{x}))$$

L'égalité précédente est vérifiée si et seulement si  $\pi f^{\mathcal{X}}(\mathbf{x}) = 0$  ou bien  $(1 - \pi)f^{\mathcal{E}}(\mathbf{x}) = 0$ .

On en déduit que, pour tout  $\mathbf{x} \in \mathfrak{F}$ ,

$$\pi f^{\mathcal{X}}(\mathbf{x}) \ln[\pi f^{\mathcal{X}}(\mathbf{x})] + (1 - \pi)f^{\mathcal{E}}(\mathbf{x}) \ln[(1 - \pi)f^{\mathcal{E}}(\mathbf{x})] \leq f^{\mathcal{T}}(\mathbf{x}) \ln[f^{\mathcal{T}}(\mathbf{x})]$$

On intègre l'inégalité précédente sur  $\mathfrak{F}$ , et on obtient :

$$\pi \ln(\pi) + (1 - \pi) \ln(1 - \pi) + \pi \int f^{\mathcal{X}} \ln f^{\mathcal{X}} + (1 - \pi) \int f^{\mathcal{E}} \ln f^{\mathcal{E}} \leq \int f^{\mathcal{T}} \ln f^{\mathcal{T}}$$

soit, le côté droit de l'inégalité (6) :

$$\mathbb{H}(\mathbf{X}^{\mathcal{T}}) - \mathbb{H}(Y^{\mathcal{X}}) \leq 0$$

L'égalité ne survient que lorsque

$$\forall \mathbf{x} \in \mathfrak{F}, \min\{f^{\mathcal{X}}(\mathbf{x}), f^{\mathcal{E}}(\mathbf{x})\} = 0$$

En d'autres termes, l'égalité est atteinte uniquement lorsque les individus de la population  $\mathcal{X}$  se mettent là où ceux de la population  $\mathcal{E}$  ne se mettent pas, et *vice-versa*.

Montrons à présent l'autre côté de l'inégalité (6). Par le théorème 3.1 (p. 14) de (Kullback 1959), on a, quelles que soient les densités  $f^{\mathcal{X}}$  et  $f^{\mathcal{T}}$  :

$$\int f^{\mathcal{X}} \ln \left( \frac{f^{\mathcal{X}}}{f^{\mathcal{T}}} \right) \geq 0$$

l'égalité étant assurée si et seulement si  $f^{\mathcal{X}} = f^{\mathcal{T}}$ . Cette propriété est aussi vérifiée en échangeant  $\mathcal{X}$  et  $\mathcal{E}$ . Il en découle que  $\int f^{\mathcal{X}} \ln f^{\mathcal{X}} \geq \int f^{\mathcal{X}} \ln f^{\mathcal{T}}$  et  $\int f^{\mathcal{E}} \ln f^{\mathcal{E}} \geq \int f^{\mathcal{E}} \ln f^{\mathcal{T}}$ . En combinant ces deux inégalités, on a :

$$\pi \int f^{\mathcal{X}} \ln f^{\mathcal{X}} + (1 - \pi) \int f^{\mathcal{E}} \ln f^{\mathcal{E}} \geq \int f^{\mathcal{T}} \ln f^{\mathcal{T}}$$

d'où le côté gauche de l'inégalité :

$$\mathbb{H}(\mathbf{X}^{\mathcal{T}}) - \mathbb{H}(\mathbf{Y}^{\mathcal{X}}) \geq -\mathbb{H}(\mathbf{W}^{\mathcal{X}})$$

L'égalité ne survient que lorsque  $f^{\mathcal{X}} = f^{\mathcal{E}} = f^{\mathcal{T}}$ .

### Démonstration de l'équivalence des écritures (7) et (8) de l'indice de concentration

L'indice sous sa forme (8) s'écrit :

$$I_{RO}^{\mathcal{X}} = 1 - \frac{1}{\mathbb{H}(W^{\mathcal{X}})} \int f^{\mathcal{T}} \left( -p^{\mathcal{X}/\mathcal{T}} \ln(p^{\mathcal{X}/\mathcal{T}}) - (1 - p^{\mathcal{X}/\mathcal{T}}) \ln(1 - p^{\mathcal{X}/\mathcal{T}}) \right)$$

Comme  $p^{\mathcal{X}/\mathcal{T}} = \frac{\pi f^{\mathcal{X}}}{f^{\mathcal{T}}}$  et  $1 - p^{\mathcal{X}/\mathcal{T}} = \frac{(1-\pi)f^{\mathcal{E}}}{f^{\mathcal{T}}}$ ,

$$I_{RO}^{\mathcal{X}} = 1 - \frac{1}{\mathbb{H}(W^{\mathcal{X}})} \int -\pi f^{\mathcal{X}} \ln \pi - \pi f^{\mathcal{X}} \ln f^{\mathcal{X}} + \pi f^{\mathcal{X}} \ln f^{\mathcal{T}} - (1-\pi) f^{\mathcal{E}} \ln(1-\pi) - (1-\pi) f^{\mathcal{E}} \ln f^{\mathcal{E}} + (1-\pi) f^{\mathcal{E}} \ln f^{\mathcal{T}}$$

En utilisant le fait que les densités s'intègrent à l'unité,

$$\int -\pi f^{\mathcal{X}} \ln \pi - (1 - \pi) f^{\mathcal{E}} \ln(1 - \pi) = \mathbb{H}(\mathbf{W}^{\mathcal{X}})$$

Par ailleurs,

$$\pi f^{\mathcal{X}} \ln f^{\mathcal{T}} + (1 - \pi) f^{\mathcal{E}} \ln f^{\mathcal{T}} = \mathbb{H}(\mathbf{X}^{\mathcal{T}})$$

De par ces deux dernières égalités, l'indice s'écrit :

$$I_{RO}^{\mathcal{X}} = 1 - \frac{1}{\mathbb{H}(W^{\mathcal{X}})} \int \pi \mathbb{H}(\mathbf{X}^{\mathcal{X}}) + (1 - \pi) \mathbb{H}(\mathbf{X}^{\mathcal{E}}) - \mathbb{H}(\mathbf{X}^{\mathcal{T}}) + \mathbb{H}(\mathbf{W}^{\mathcal{X}})$$

D'où  $I_{RO}^{\mathcal{X}} = I^{\mathcal{X}}$ .

G 9001	J. FAYOLLE et M. FLEURBAEY Accumulation, profitabilité et endettement des entreprises		Macro-economic import functions with imperfect competition - An application to the E.C. Trade		françaises : une évaluation empirique des théories de la structure optimale du capital	G 9412	J. BOURDIEU - B. CŒURÉ - B. COLIN-SEDILLOT Investissement, incertitude et irréversibilité Quelques développements récents de la théorie de l'investissement
G 9002	H. ROUSSE Détection et effets de la multicollinéarité dans les modèles linéaires ordinaires - Un prolongement de la réflexion de BELSLEY, KUH et WELSCH	G 9203	I. STAPIC Les échanges internationaux de services de la France dans le cadre des négociations multilatérales du GATT Juin 1992 (1ère version) Novembre 1992 (version finale)	G 9312	L. BLOCH - B. CŒURÉ Q de Tobin marginal et transmission des chocs financiers	G 9413	B. DORMONT - M. PAUCHET L'évaluation de l'élasticité emploi-salaire dépend-elle des structures de qualification ?
G 9003	P. RALLE et J. TOUJAS-BERNATE Indexation des salaires : la rupture de 1983	G 9204	P. SEVESTRE L'économétrie sur données individuelles-temporelles. Une note introductive	G 9313	Equipes Amadeus (INSEE), Banque de France, Métric (DP) Présentation des propriétés des principaux modèles macroéconomiques du Service Public	G 9414	I. KABLA Le Choix de breveter une invention
G 9004	D. GUELLEC et P. RALLE Compétitivité, croissance et innovation de produit	G 9205	H. ERKEL-ROUSSE Le commerce extérieur et l'environnement international dans le modèle AMADEUS (réestimation 1992)	G 9314	B. CREPON - E. DUGUET Research & Development, competition and innovation	G 9501	J. BOURDIEU - B. CŒURÉ - B. SEDILLOT Irreversible Investment and Uncertainty : When is there a Value of Waiting ?
G 9005	P. RALLE et J. TOUJAS-BERNATE Les conséquences de la désindexation. Analyse dans une maquette prix-salaires	G 9206	N. GREENAN et D. GUELLEC Coordination within the firm and endogenous growth	G 9315	B. DORMONT Quelle est l'influence du coût du travail sur l'emploi ?	G 9502	L. BLOCH - B. CŒURÉ Imperfections du marché du crédit, investissement des entreprises et cycle économique
G 9101	Equipe AMADEUS Le modèle AMADEUS - Première partie - Présentation générale	G 9207	A. MAGNIER et J. TOUJAS-BERNATE Technology and trade : empirical evidences for the major five industrialized countries	G 9316	D. BLANCHET - C. BROUSSE Deux études sur l'âge de la retraite	G 9503	D. GOUX - E. MAURIN Les transformations de la demande de travail par qualification en France Une étude sur la période 1970-1993
G 9102	J.L. BRILLET Le modèle AMADEUS - Deuxième partie - Propriétés variantielles	G 9208	B. CREPON, E. DUGUET, D. ENCAOUA et P. MOHNEN Cooperative, non cooperative R & D and optimal patent life	G 9317	D. BLANCHET Répartition du travail dans une population hétérogène : deux notes	G 9504	N. GREENAN Technologie, changement organisationnel, qualifications et emploi : une étude empirique sur l'industrie manufacturière
G 9103	D. GUELLEC et P. RALLE Endogenous growth and product innovation	G 9209	B. CREPON et E. DUGUET Research and development, competition and innovation : an application of pseudo maximum likelihood methods to Poisson models with heterogeneity	G 9318	D. EYSSARTIER - N. PONTY AMADEUS - an annual macro-economic model for the medium and long term	G 9505	D. GOUX - E. MAURIN Persistence des hiérarchies sectorielles de salaires: un réexamen sur données françaises
G 9104	H. ROUSSE Le modèle AMADEUS - Troisième partie - Le commerce extérieur et l'environnement international	G 9301	J. TOUJAS-BERNATE Commerce international et concurrence imparfaite : développements récents et implications pour la politique commerciale	G 9319	G. CETTE - Ph. CUNÉO - D. EYSSARTIER - J. GAUTIÉ Les effets sur l'emploi d'un abaissement du coût du travail des jeunes	G 9505	D. GOUX - E. MAURIN Bis Persistence of inter-industry wages differentials: a reexamination on matched worker-firm panel data
G 9105	H. ROUSSE Effets de demande et d'offre dans les résultats du commerce extérieur manufacturé de la France au cours des deux dernières décennies	G 9302	Ch. CASES Durées de chômage et comportements d'offre de travail : une revue de la littérature	G 9401	D. BLANCHET Les structures par âge importent-elles ?	G 9506	S. JACOBZONE Les liens entre RMI et chômage, une mise en perspective <i>NON PARU - article sorti dans Economie et Prévision n°122 (1996) - pages 95 à 113</i>
G 9106	B. CREPON Innovation, taille et concentration : causalités et dynamiques	G 9303	H. ERKEL-ROUSSE Union économique et monétaire : le débat économique	G 9402	J. GAUTIÉ Le chômage des jeunes en France : problème de formation ou phénomène de file d'attente ? Quelques éléments du débat	G 9507	G. CETTE - S. MAHFOUZ Le partage primaire du revenu Constat descriptif sur longue période
G 9107	B. AMABLE et D. GUELLEC Un panorama des théories de la croissance endogène	G 9304	N. GREENAN - D. GUELLEC / G. BROUSSAUDIER - L. MIOTTI Innovation organisationnelle, dynamisme technologique et performances des entreprises	G 9403	P. QUIRION Les déchets en France : éléments statistiques et économiques	G 9601	Banque de France - CEPREMAP - Direction de la Prévision - Erasme - INSEE - OFCE Structures et propriétés de cinq modèles macro-économiques français
G 9108	M. GLAUDE et M. MOUTARDIER Une évaluation du coût direct de l'enfant de 1979 à 1989	G 9305	P. JAILLARD Le traité de Maastricht : présentation juridique et historique	G 9404	D. LADIRAY - M. GRUN-REHOMME Lissage par moyennes mobiles - Le problème des extrémités de série	G 9602	Rapport d'activité de la DESE de l'année 1995
G 9109	P. RALLE et alii France - Allemagne : performances économiques comparées	G 9306	J.L. BRILLET Micro-DMS : présentation et propriétés	G 9405	V. MAILLARD Théorie et pratique de la correction des effets de jours ouvrables	G 9603	J. BOURDIEU - A. DRAZNIKES L'octroi de crédit aux PME : une analyse à partir d'informations bancaires
G 9110	J.L. BRILLET Micro-DMS <b>NON PARU</b>	G 9307	J.L. BRILLET Micro-DMS - variantes : les tableaux	G 9406	F. ROSENWALD La décision d'investir	G 9604	A. TOPIOL-BENSAÏD Les implantations japonaises en France
G 9111	A. MAGNIER Effets accélérateur et multiplicateur en France depuis 1970 : quelques résultats empiriques	G 9308	S. JACOBZONE Les grands réseaux publics français dans une perspective européenne	G 9407	S. JACOBZONE Les apports de l'économie industrielle pour définir la stratégie économique de l'hôpital public	G 9605	P. GENIER - S. JACOBZONE Comportements de prévention, consommation d'alcool et tabagie : peut-on parler d'une gestion globale du capital santé ? <i>Une modélisation microéconométrique empirique</i>
G 9112	B. CREPON et G. DUREAU Investissement en recherche-développement : analyse de causalités dans un modèle d'accélérateur généralisé	G 9309	L. BLOCH - B. CŒURE Profitabilité de l'investissement productif et transmission des chocs financiers	G 9408	L. BLOCH, J. BOURDIEU, B. COLIN-SEDILLOT, G. LONGUEVILLE Du défaut de paiement au dépôt de bilan : les banquiers face aux PME en difficulté	G 9606	C. DOZ - F. LENGART Factor analysis and unobserved component models: an application to the study of French business surveys
G 9113	J.L. BRILLET, H. ERKEL-ROUSSE, J. TOUJAS-BERNATE "France-Allemagne Couplées" - Deux économies vues par une maquette macro-économétrique	G 9310	J. BOURDIEU - B. COLIN-SEDILLOT Les théories sur la structure optimal du capital : quelques points de repère	G 9409	D. EYSSARTIER, P. MAIRE Impacts macro-économiques de mesures d'aide au logement - quelques éléments d'évaluation	G 9607	N. GREENAN - D. GUELLEC La théorie coopérative de la firme
G 9201	W.J. ADAMS, B. CREPON, D. ENCAOUA Choix technologiques et stratégies de dissuasion d'entrée	G 9311	J. BOURDIEU - B. COLIN-SEDILLOT Les décisions de financement des entreprises	G 9410	F. ROSENWALD Suivi conjoncturel de l'investissement		
G 9202	J. OLIVEIRA-MARTINS, J. TOUJAS-BERNATE			G 9411	C. DEFEUILLEY - Ph. QUIRION Les déchets d'emballages ménagers : une analyse économique des politiques française et allemande		

G 9608	N. GREENAN - D. GUELLEC Technological innovation and employment reallocation	G 9714	F. LEQUILLER Does the French Consumer Price Index Overstate Inflation?	G 9808	A. MOUROUGANE Can a Conservative Governor Conduct an Accommodative Monetary Policy ?	G 9913	Division « Redistribution et Politiques Sociales » Le modèle de microsimulation dynamique DESTINIE
G 9609	Ph. COUR - F. RUPPRECHT L'intégration asymétrique au sein du continent américain : un essai de modélisation	G 9715	X. BONNET Peut-on mettre en évidence les rigidités à la baisse des salaires nominaux ? Une étude sur quelques grands pays de l'OCDE	G 9809	X. BONNET - E. DUBOIS - L. FAUVET Asymétrie des inflations relatives et menus costs : tests sur l'inflation française	G 9914	E. DUGUET Macro-commandes SAS pour l'économétrie des panels et des variables qualitatives
G 9610	S. DUCHENE - G. FORGEOT - A. JACQUOT Analyse des évolutions récentes de la productivité apparente du travail	G 9716	N. IUNG - F. RUPPRECHT Productivité de la recherche et rendements d'échelle dans le secteur pharmaceutique français	G 9810	E. DUGUET - N. IUNG Sales and Advertising with Spillovers at the firm level: Estimation of a Dynamic Structural Model on Panel Data	G 9915	R. DUHAUTOIS Evolution des flux d'emplois en France entre 1990 et 1996 : une étude empirique à partir du fichier des bénéficiaires réels normaux (BRN)
G 9611	X. BONNET - S. MAHFOUZ The influence of different specifications of wages-prices spirals on the measure of the NAIRU : the case of France	G 9717	E. DUGUET - I. KABLA Appropriation strategy and the motivations to use the patent system in France - An econometric analysis at the firm level	G 9811	J.P. BERTHIER Congestion urbaine : un modèle de trafic de pointe à courbe débit-vitesse et demande élastique	G 9916	J.Y. FOURNIER Extraction du cycle des affaires : la méthode de Baxter et King
G 9612	PH. COUR - E. DUBOIS, S. MAHFOUZ, J. PISANI-FERRY The cost of fiscal retrenchment revisited: how strong is the evidence ?	G 9718	L.P. PELÉ - P. RALLE Âge de la retraite : les aspects incitatifs du régime général	G 9812	C. PRIGENT La part des salaires dans la valeur ajoutée : une approche macroéconomique	G 9917	B. CRÉPON - R. DESPLATZ - J. MAIRESSE Estimating price cost margins, scale economies and workers' bargaining power at the firm level
G 9613	A. JACQUOT Les flexions des taux d'activité sont-elles seulement conjoncturelles ?	G 9719	ZHANG Yingxiang - SONG Xueqing Lexique macroéconomique français-chinois, chinois-français	G 9813	A.Th. AERTS L'évolution de la part des salaires dans la valeur ajoutée en France reflète-t-elle les évolutions individuelles sur la période 1979-1994 ?	G 9918	Ch. GIANELLA - Ph. LAGARDE Productivity of hours in the aggregate production function: an evaluation on a panel of French firms from the manufacturing sector
G 9614	ZHANG Yingxiang - SONG Xueqing Lexique macroéconomique Français-Chinois	G 9720	M. HOUEBINE - J.L. SCHNEIDER Mesurer l'influence de la fiscalité sur la localisation des entreprises	G 9814	B. SALANIÉ Guide pratique des séries non-stationnaires	G 9919	S. AUDRIC - P. GIVORD - C. PROST Evolution de l'emploi et des coûts par qualification entre 1982 et 1996
G 9701	J.L. SCHNEIDER La taxe professionnelle : éléments de cadrage économique	G 9721	A. MOUROUGANE Crédibilité, indépendance et politique monétaire Une revue de la littérature	G 9901	S. DUCHÈNE - A. JACQUOT Une croissance plus riche en emplois depuis le début de la décennie ? Une analyse en comparaison internationale	G 2000/01	R. MAHIEU Les déterminants des dépenses de santé : une approche macroéconomique
G 9702	J.L. SCHNEIDER Transition et stabilité politique d'un système redistributif	G 9722	P. AUGERAUD - L. BRIOT Les données comptables d'entreprises Le système intermédiaire d'entreprises Passage des données individuelles aux données sectorielles	G 9902	Ch. COLIN Modélisation des carrières dans Destinie	G 2000/02	C. ALLARD-PRIGENT - H. GUILMEAU - A. QUINET The real exchange rate as the relative price of nontradables in terms of tradables: theoretical investigation and empirical study on French data
G 9703	D. GOUX - E. MAURIN Train or Pay: Does it Reduce Inequalities to Encourage Firms to Train their Workers?	G 9723	P. AUGERAUD - J.E. CHAPRON Using Business Accounts for Compiling National Accounts: the French Experience	G 9903	Ch. COLIN Evolution de la dispersion des salaires : un essai de prospective par microsimulation	G 2000/03	J.-Y. FOURNIER L'approximation du filtre passe-bande proposée par Christiano et Fitzgerald
G 9704	P. GENIER Deux contributions sur dépendance et équité	G 9724	P. AUGERAUD Les comptes d'entreprise par activités - Le passage aux comptes - De la comptabilité d'entreprise à la comptabilité nationale - A <i>paraître</i>	G 9904	B. CREPON - N. IUNG Innovation, emploi et performances	G 2000/04	Bilan des activités de la DESE - 1999
G 9705	E. DUGUET - N. IUNG R & D Investment, Patent Life and Patent Value An Econometric Analysis at the Firm Level	G 9801	H. MICHAUDON - C. PRIGENT Présentation du modèle AMADEUS	G 9905	B. CREPON - Ch. GIANELLA Wages inequalities in France 1969-1992 An application of quantile regression techniques	G 2000/05	B. CREPON - F. ROSENWALD Investissement et contraintes de financement : le poids du cycle Une estimation sur données françaises
G 9706	M. HOUEBINE - A. TOPIOL-BENSAÏD Les entreprises internationales en France : une analyse à partir de données individuelles	G 9802	J. ACCARDO Une étude de comptabilité générationnelle pour la France en 1996	G 9906	C. BONNET - R. MAHIEU Microsimulation techniques applied to inter-generational transfers - Pensions in a dynamic framework: the case of France	G 2000/06	A. FLIPO Les comportements matrimoniaux de fait
G 9707	M. HOUEBINE Polarisation des activités et spécialisation des départements en France	G 9803	X. BONNET - S. DUCHÈNE Apports et limites de la modélisation « Real Business Cycles »	G 9907	F. ROSENWALD L'impact des contraintes financières dans la décision d'investissement	G 2000/07	R. MAHIEU - B. SÉDILLOT Microsimulations of the retirement decision: a supply side approach
G 9708	E. DUGUET - N. GREENAN Le biais technologique : une analyse sur données individuelles	G 9804	C. BARLET - C. DUGUET - D. ENCAOUA - J. PRADEL The Commercial Success of Innovations An econometric analysis at the firm level in French manufacturing	G 9908	Bilan des activités de la DESE - 1998	G 2000/08	C. AUDENIS - C. PROST Déficit conjoncturel : une prise en compte des conjonctures passées
G 9709	J.L. BRILLET Analyzing a small French ECM Model	G 9805	P. CAHUC - Ch. GIANELLA - D. GOUX - A. ZILBERBERG Equalizing Wage Differences and Bargaining Power - Evidence from a Panel of French Firms	G 9909	J.P. ZOYEM Contrat d'insertion et sortie du RMI Evaluation des effets d'une politique sociale	G 2000/09	R. MAHIEU - B. SÉDILLOT Equivalent patrimonial de la rente et souscription de retraite complémentaire
G 9710	J.L. BRILLET Formalizing the transition process : scenarios for capital accumulation	G 9806	J. ACCARDO - M. JLASSI La productivité globale des facteurs entre 1975 et 1996	G 9910	Ch. COLIN - FI. LEGROS - R. MAHIEU Bilans contributifs comparés des régimes de retraite du secteur privé et de la fonction publique	G 2000/10	R. DUHAUTOIS Ralentissement de l'investissement : petites ou grandes entreprises ? industrie ou tertiaire ?
G 9711	G. FORGEOT - J. GAUTIÉ Insertion professionnelle des jeunes et processus de déclassement	G 9807	Bilan des activités de la Direction des Etudes et Synthèses Economiques - 1997	G 9911	G. LAROQUE - B. SALANIÉ Une décomposition du non-emploi en France	G 2000/11	G. LAROQUE - B. SALANIÉ Temps partiel féminin et incitations financières à l'emploi
G 9712	E. DUBOIS High Real Interest Rates: the Consequence of a Saving Investment Disequilibrium or of an insufficient Credibility of Monetary Authorities?			G 9912	B. SALANIÉ Une maquette analytique de long terme du marché du travail	G2000/12	Ch. GIANELLA Local unemployment and wages
G 9713	Bilan des activités de la Direction des Etudes et Synthèses Economiques - 1996			G 9912 Bis	Ch. GIANELLA Une estimation de l'élasticité de l'emploi peu qualifié à son coût	G2000/13	B. CREPON - Th. HECKEL - Informatisation en France : une évaluation à partir de données individuelles

G2001/01	- Computerization in France: an evaluation based on individual company data F. LEQUILLER - La nouvelle économie et la mesure de la croissance du PIB - The new economy and the measurement of GDP growth	G2002/01	F. MAGNIEN - J.-L. TAVERNIER - D. THESMAR Les statistiques internationales de PIB par habitant en standard de pouvoir d'achat : une analyse des résultats	G2002/16	F. MAUREL - S. GREGOIR Les indices de compétitivité des pays : interprétation et limites	G2004/06	M. DUÉE L'impact du chômage des parents sur le devenir scolaire des enfants
G2001/02	S. AUDRIC La reprise de la croissance de l'emploi profite-t-elle aussi aux non-diplômés ?	G2002/02	Bilan des activités de la DESE - 2001	G2003/01	N. RIEDINGER - E. HAUVY Le coût de dépollution atmosphérique pour les entreprises françaises : Une estimation à partir de données individuelles	G2004/07	P. AUBERT - E. CAROLI - M. ROGER New Technologies, Workplace Organisation and the Age Structure of the Workforce: Firm-Level Evidence
G2001/03	I. BRAUN-LEMAIRE Evolution et répartition du surplus de productivité	G2002/03	B. SÉDILLOT - E. WALRAET La cessation d'activité au sein des couples : y a-t-il interdépendance des choix ?	G2003/02	P. BISCOURP et F. KRAMARZ Création d'emplois, destruction d'emplois et internationalisation des entreprises industrielles françaises : une analyse sur la période 1986-1992	G2004/08	E. DUGUET - C. LELARGE Les brevets accroissent-ils les incitations privées à innover ? Un examen microéconométrique
G2001/04	A. BEAUDU - Th. HECKEL Le canal du crédit fonctionne-t-il en Europe ? Une étude de l'hétérogénéité des comportements d'investissement à partir de données de bilan agrégées	G2002/04	G. BRILHAULT - Rétropolation des séries de FBCF et calcul du capital fixe en SEC-95 dans les comptes nationaux français - Retropolation of the investment series (GFCF) and estimation of fixed capital stocks on the ESA-95 basis for the French balance sheets	G2003/03	Bilan des activités de la DESE - 2002	G2004/09	S. RASPILLER - P. SILLARD Affiliating versus Subcontracting: the Case of Multinationals
G2001/05	C. AUDENIS - P. BISCOURP - N. FOURCADE - O. LOISEL Testing the augmented Solow growth model : An empirical reassessment using panel data	G2002/05	P. BISCOURP - B. CRÉPON - T. HECKEL - N. RIEDINGER How do firms respond to cheaper computers? Microeconomic evidence for France based on a production function approach	G2003/04	P.-O. BEFFY - J. DERUYON - N. FOURCADE - S. GREGOIR - N. LAÏB - B. MONFORT Évolutions démographiques et croissance : une projection macro-économique à l'horizon 2020	G2004/10	J. BOISSINOT - C. L'ANGEVIN - B. MONFORT Public Debt Sustainability: Some Results on the French Case
G2001/06	R. MAHIEU - B. SÉDILLOT Départ à la retraite, irréversibilité et incertitude	G2002/06	C. AUDENIS - J. DERUYON - N. FOURCADE L'impact des nouvelles technologies de l'information et de la communication sur l'économie française - un bouclage macro-économique	G2003/05	P. AUBERT La situation des salariés de plus de cinquante ans dans le secteur privé	G2004/11	S. ANANIAN - P. AUBERT Travailleurs âgés, nouvelles technologies et changements organisationnels : un réexamen à partir de l'enquête « REPONSE »
G2001/07	Bilan des activités de la DESE - 2000	G2002/07	J. BARDAJI - B. SÉDILLOT - E. WALRAET Évaluation de trois réformes du Régime Général d'assurance vieillesse à l'aide du modèle de microsimulation DESTINIE	G2003/06	P. AUBERT - B. CRÉPON Age, salaire et productivité La productivité des salariés décline-t-elle en fin de carrière ?	G2004/12	X. BONNET - H. PONCET Structures de revenus et propensions différentes à consommer - Vers une équation de consommation des ménages plus robuste en prévision pour la France
G2001/08	J. Ph. GAUDEMET Les dispositifs d'acquisition à titre facultatif d'annuités viagères de retraite	G2002/08	J.-P. BERTHIER Réflexions sur les différentes notions de volume dans les comptes nationaux : comptes aux prix d'une année fixe ou aux prix de l'année précédente, séries chaînées	G2003/07	H. BARON - P.O. BEFFY - N. FOURCADE - R. MAHIEU Le ralentissement de la productivité du travail au cours des années 1990	G2004/13	C. PICART Évaluer la rentabilité des sociétés non financières
G2001/09	B. CRÉPON - Ch. GIANELLA Fiscalité, coût d'usage du capital et demande de facteurs : une analyse sur données individuelles	G2002/09	F. HILD Les soldes d'opinion résumant-ils au mieux les réponses des entreprises aux enquêtes de conjoncture ?	G2003/08	P.-O. BEFFY - B. MONFORT Patrimoine des ménages, dynamique d'allocation et comportement de consommation	G2004/14	J. BARDAJI - B. SÉDILLOT - E. WALRAET Les retraites du secteur public : projections à l'horizon 2040 à l'aide du modèle de microsimulation DESTINIE
G2001/10	B. CRÉPON - R. DESPLATZ Evaluation des effets des dispositifs d'allègements de charges sociales sur les bas salaires	G2002/10	I. ROBERT-BOBÉE Les comportements démographiques dans le modèle de microsimulation Destinie - Une comparaison des estimations issues des enquêtes Jeunes et Carrières 1997 et Histoire Familiale 1999	G2003/09	P. BISCOURP - N. FOURCADE Peut-on mettre en évidence l'existence de rigidités à la baisse des salaires à partir de données individuelles ? Le cas de la France à la fin des années 90	G2005/01	S. BUFFETEAU - P. GODEFROY Conditions de départ en retraite selon l'âge de fin d'études : analyse prospective pour les générations 1945 à 1974
G2001/11	J.-Y. FOURNIER Comparaison des salaires des secteurs public et privé	G2002/11	J.-P. ZOYEM La dynamique des bas revenus : une analyse des entrées-sorties de pauvreté	G2003/10	M. LECLAIR - P. PETIT Présence syndicale dans les firmes : quel impact sur les inégalités salariales entre les hommes et les femmes ?	G2005/02	C. AFSA - S. BUFFETEAU L'évolution de l'activité féminine en France : une approche par pseudo-panel
G2001/12	J.-P. BERTHIER - C. JAULENT R. CONVENEVOLE - S. PISANI Une méthodologie de comparaison entre consommations intermédiaires de source fiscale et de comptabilité nationale	G2002/12	F. HILD Prévisions d'inflation pour la France	G2003/11	P.-O. BEFFY - X. BONNET - M. DARRACQ-PARIES - B. MONFORT MZE: a small macro-model for the euro area	G2005/03	P. AUBERT - P. SILLARD Délocalisations et réductions d'effectifs dans l'industrie française
G2001/13	P. BISCOURP - Ch. GIANELLA Substitution and complementarity between capital, skilled and less skilled workers: an analysis at the firm level in the French manufacturing industry	G2002/13	M. LECLAIR Réduction du temps de travail et tensions sur les facteurs de production	G2004/01	P. AUBERT - M. LECLAIR La compétitivité exprimée dans les enquêtes trimestrielles sur la situation et les perspectives dans l'industrie	G2005/04	M. LECLAIR - S. ROUX Mesure et utilisation des emplois instables dans les entreprises
G2001/14	I. ROBERT-BOBÉE Modelling demographic behaviours in the French microsimulation model Destinie: An analysis of future change in completed fertility	G2002/14	E. WALRAET - A. VINCENT - Analyse de la redistribution intragénérationnelle dans le système de retraite des salariés du privé - Une approche par microsimulation - Intragenerational distributional analysis in the french private sector pension scheme - A microsimulation approach	G2004/02	M. DUÉE - C. REBILLARD La dépendance des personnes âgées : une projection à long terme	G2005/05	C. L'ANGEVIN - S. SERRAVALLE Performances à l'exportation de la France et de l'Allemagne - Une analyse par secteur et destination géographique
G2001/15	J.-P. ZOYEM Diagnostic sur la pauvreté et calendrier de revenus : le cas du "Panel européen des ménages"	G2002/15	P. CHONE - D. LE BLANC - I. ROBERT-BOBÉE Offre de travail féminine et garde des jeunes enfants	G2004/03	S. RASPILLER - N. RIEDINGER Régulation environnementale et choix de localisation des groupes français	G2005/06	Bilan des activités de la Direction des Études et Synthèses Économiques - 2004
G2001/16	J.-Y. FOURNIER - P. GIVORD La réduction des taux d'activité aux âges extrêmes, une spécificité française ?			G2004/04	A. NABOULET - S. RASPILLER Les déterminants de la décision d'investir : une approche par les perceptions subjectives des firmes	G2005/07	S. RASPILLER La concurrence fiscale : principaux enseignements de l'analyse économique
G2001/17	C. AUDENIS - P. BISCOURP - N. RIEDINGER Existe-t-il une asymétrie dans la transmission du prix du brut aux prix des carburants ?			G2004/05	N. RAGACHE La déclaration des enfants par les couples non mariés est-elle fiscalement optimale ?	G2005/08	C. L'ANGEVIN - N. LAÏB Éducation et croissance en France et dans un panel de 21 pays de l'OCDE

G2005/10	P.-O. BEFFY - C. L'ANGEVIN Chômage et boucle prix-salaires : apport d'un modèle « qualifiés/peu qualifiés »
G2005/11	B. HEITZ A two-states Markov-switching model of inflation in France and the USA: credible target VS inflation spiral
G2005/12	O. BIAU - H. ERKEL-ROUSSE - N. FERRARI Réponses individuelles aux enquêtes de conjoncture et prévision macroéconomiques : Exemple de la prévision de la production manufacturière
G2005/13	P. AUBERT - D. BLANCHET - D. BLAU The labour market after age 50: some elements of a Franco-American comparison
G2005/14	D. BLANCHET - T. DEBRAND - P. DOURGNON - P. POLLET L'enquête SHARE : présentation et premiers résultats de l'édition française
G2005/15	M. DUÉE La modélisation des comportements démogra- phiques dans le modèle de microsimulation DESTINIE
G2005/16	H. RAOUI - S. ROUX Étude de simulation sur la participation versée aux salariés par les entreprises
G2006/01	C. BONNET - S. BUFFETEAU - P. GODEFROY Disparités de retraite de droit direct entre hommes et femmes : quelles évolutions ?
G2006/02	C. PICART Les gazelles en France
G2006/03	P. AUBERT - B. CRÉPON - P. ZAMORA Le rendement apparent de la formation continue dans les entreprises : effets sur la productivité et les salaires
G2006/04	J.-F. OUVRARD - R. RATHELOT Demographic change and unemployment: what do macroeconomic models predict?
G2006/05	D. BLANCHET - J.-F. OUVRARD Indicateurs d'engagements implicites des systèmes de retraite : chiffrages, propriétés analytiques et réactions à des chocs démographiques types
G2006/06	G. BIAU - O. BIAU - L. ROUVIERE Nonparametric Forecasting of the Manufacturing Output Growth with Firm-level Survey Data
G2006/07	C. AFSA - P. GIVORD Le rôle des conditions de travail dans les absences pour maladie
G2006/08	P. SILLARD - C. L'ANGEVIN - S. SERRAVALLE Performances comparées à l'exportation de la France et de ses principaux partenaires Une analyse structurelle sur 12 ans
G2006/09	X. BOUTIN - S. QUANTIN Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises : 1984- 2002
G2006/10	C. AFSA L'estimation d'un coût implicite de la pénibilité du travail chez les travailleurs âgés
G2006/11	C. LELARGE Les entreprises (industrielles) françaises sont- elles à la frontière technologique ?
G2006/12	O. BIAU - N. FERRARI Théorie de l'opinion Faut-il pondérer les réponses individuelles ?
G2006/13	A. KOUBI - S. ROUX Une réinterprétation de la relation entre productivité et inégalités salariales dans les entreprises
G2006/14	R. RATHELOT - P. SILLARD The impact of local taxes on plants location decision
G2006/15	L. GONZALEZ - C. PICART Diversification, recentrage et poids des activités de support dans les groupes (1993-2000)
G2007/01	D. SRAER Allègements de cotisations patronales et dynamique salariale
G2007/02	V. ALBOUY - L. LEQUIEN Les rendements non monétaires de l'éducation : le cas de la santé
G2007/03	D. BLANCHET - T. DEBRAND Aspiration à la retraite, santé et satisfaction au travail : une comparaison européenne
G2007/04	M. BARLET - L. CRUSSON Quel impact des variations du prix du pétrole sur la croissance française ?
G2007/05	C. PICART Flux d'emploi et de main-d'œuvre en France : un réexamen
G2007/06	V. ALBOUY - C. TAVAN Massification et démocratisation de l'enseignement supérieur en France
G2007/07	T. LE BARBANCHON The Changing response to oil price shocks in France : a DSGE type approach
G2007/08	T. CHANEY - D. SRAER - D. THESMAR Collateral Value and Corporate Investment Evidence from the French Real Estate Market
G2007/09	J. BOISSINOT Consumption over the Life Cycle: Facts for France
G2007/10	C. AFSA Interpréter les variables de satisfaction : l'exemple de la durée du travail
G2007/11	R. RATHELOT - P. SILLARD Zones Franches Urbaines : quels effets sur l'emploi salarié et les créations d'établissements ?
G2007/12	V. ALBOUY - B. CRÉPON Aléa moral en santé : une évaluation dans le cadre du modèle causal de Rubin
G2008/01	C. PICART Les PME françaises : rentables mais peu dynamiques
G2008/02	P. BISCOURP - X. BOUTIN - T. VERGÉ The Effects of Retail Regulations on Prices Evidence from the Loi Galland
G2008/03	Y. BARBESOL - A. BRIANT Economies d'agglomération et productivité des

G2009/09	G. LALANNE - E. POULIQUEN - O. SIMON Prix du pétrole et croissance potentielle à long terme
G2009/10	D. BLANCHET - J. LE CACHEUX - V. MARCUS Adjusted net savings and other approaches to sustainability: some theoretical background
G2009/11	V. BELLAMY - G. CONSALES - M. FESSEAU - S. LE LAIDIER - É. RAYNAUD Une décomposition du compte des ménages de la comptabilité nationale par catégorie de ménage en 2003
G2009/12	J. BARDAJI - F. TALLET Detecting Economic Regimes in France : a Qualitative Markov-Switching Indicator Using Mixed Frequency Data
G2009/13	R. AEBERHARDT - D. FOUGÈRE - R. RATHELOT Discrimination à l'embauche : comment exploiter les procédures de <i>testing</i> ?
G2009/14	Y. BARBESOL - P. GIVORD - S. QUANTIN Partage de la valeur ajoutée, approche par données microéconomiques
G2009/15	I. BUONO - G. LALANNE The Effect of the Uruguay round on the Intensive and Extensive Margins of Trade
G2010/01	C. MINODIER Avantages comparés des séries des premières valeurs publiées et des séries des valeurs révisées - Un exercice de prévision en temps réel de la croissance trimestrielle du PIB en France
G2010/02	V. ALBOUY - L. DAVEZIES - T. DEBRAND Health Expenditure Models: a Comparison of Five Specifications using Panel Data
G2010/03	C. KLEIN - O. SIMON Le modèle MÉSANGE réestimé en base 2000 Tome 1 – Version avec volumes à prix constants
G2010/04	M.-É. CLERC - É. COUDIN L'IPC, miroir de l'évolution du coût de la vie en France ? Ce qu'apporte l'analyse des courbes d'Engel
G2010/05	N. CECI-RENAUD - P.-A. CHEVALIER Les seuils de 10, 20 et 50 salariés : impact sur la taille des entreprises françaises
G2010/06	R. AEBERHARDT - J. POUGET National Origin Differences in Wages and Hierarchical Positions - Evidence on French Full- Time Male Workers from a matched Employer- Employee Dataset
G2010/07	S. BLASCO - P. GIVORD Les trajectoires professionnelles en début de vie active : quel impact des contrats temporaires ?
G2010/08	P. GIVORD Méthodes économétriques pour l'évaluation de politiques publiques
G2010/09	P.-Y. CABANNES - V. LAPÈGUE - E. POULIQUEN - M. BEFFY - M. GAINI Quelle croissance de moyen terme après la crise ?
G2010/10	I. BUONO - G. LALANNE La réaction des entreprises françaises à la baisse des tarifs douaniers étrangers
G2008/04	D. BLANCHET - F. LE GALLO Les projections démographiques : principaux mécanismes et retour sur l'expérience française
G2008/05	D. BLANCHET - F. TOUTLEMONDE Évolutions démographiques et déformation du cycle de vie active : quelles relations ?
G2008/06	M. BARLET - D. BLANCHET - L. CRUSSON Internationalisation et flux d'emplois : que dit une approche comptable ?
G2008/07	C. LELARGE - D. SRAER - D. THESMAR Entrepreneurship and Credit Constraints - Evidence from a French Loan Guarantee Program
G2008/08	X. BOUTIN - L. JANIN Are Prices Really Affected by Mergers?
G2008/09	M. BARLET - A. BRIANT - L. CRUSSON Concentration géographique dans l'industrie manufacturière et dans les services en France : une approche par un indicateur en continu
G2008/10	M. BEFFY - É. COUDIN - R. RATHELOT Who is confronted to insecure labor market histories? Some evidence based on the French labor market transition
G2008/11	M. ROGER - E. WALRAET Social Security and Well-Being of the Elderly: the Case of France
G2008/12	C. AFSA Analyser les composantes du bien-être et de son évolution Une approche empirique sur données individuelles
G2008/13	M. BARLET - D. BLANCHET - T. LE BARBANCHON Microsimuler le marché du travail : un prototype
G2009/01	P.-A. PIONNIER Le partage de la valeur ajoutée en France, 1949-2007
G2009/02	Laurent CLAVEL - Christelle MINODIER A Monthly Indicator of the French Business Climate
G2009/03	H. ERKEL-ROUSSE - C. MINODIER Do Business Tendency Surveys in Industry and Services Help in Forecasting GDP Growth? A Real-Time Analysis on French Data
G2009/04	P. GIVORD - L. WILNER Les contrats temporaires : trappe ou marchepied vers l'emploi stable ?
G2009/05	G. LALANNE - P.-A. PIONNIER - O. SIMON Le partage des fruits de la croissance de 1950 à 2008 : une approche par les comptes de surplus
G2009/06	L. DAVEZIES - X. D'HAULTFOEUILLE Faut-il pondérer ?... Ou l'éternelle question de l'économètre confronté à des données d'enquête
G2009/07	S. QUANTIN - S. RASPILLER - S. SERRAVALLE Commerce intragroupe, fiscalité et prix de transferts : une analyse sur données françaises
G2009/08	M. CLERC - V. MARCUS Élasticités-prix des consommations énergétiques des ménages

G2010/11 R. RATHELOT - P. SILLARD  
L'apport des méthodes à noyaux pour mesurer la  
concentration géographique - Application à la  
concentration des immigrés en France de 1968 à  
1999