

Comment analyser la dynamique temporelle
d'associations statistiques dotées
d'une très forte inertie ?

Avancées récentes
en modélisation log-multiplicative

Louis-André Vallet (CNRS)

Observatoire Sociologique du Changement,
UMR 7049 CNRS & Sciences Po, Paris

L'objet de cette présentation est de *revisiter certaines avancées méthodologiques qui sont intervenues dans la modélisation statistique des tableaux de contingence* au cours des quinze dernières années.

Ces avancées ont trait à la façon d'analyser *comment l'association entre deux variables qualitatives diffère selon les modalités d'une troisième variable* (par exemple selon des groupes sociaux, selon des pays ou encore au cours du temps).

Et je voudrais aussi montrer que *ce progrès méthodologique a eu des conséquences importantes dans le domaine des études relatives à la stratification sociale*.

Plus précisément, en tirant parti de la puissance de ces instruments, *les sociologues sont aujourd'hui capables de mettre au jour des tendances temporelles* dans des associations statistiques caractérisées par une très forte inertie et qui, il y a une dizaine d'années, étaient généralement considérées comme extrêmement stables au cours du temps.

UN EXEMPLE

Supposons qu'à partir d'enquêtes nationales représentatives on observe *le tableau de contingence à deux dimensions qui croise* :

- *la catégorie socioprofessionnelle du père*
(variable 'Origine' O avec NO modalités)
- *avec le diplôme le plus élevé obtenu*
(variable 'Éducation' E avec NE modalités)

pour des ensembles d'individus qui appartiennent à NC cohortes de naissance quinquennales

(variable « couche » 'Cohorte' C avec NC modalités)

Les variables 'Origine' et 'Éducation' sont définies comme des variables qualitatives nominales, i.e. il n'y a pas d'ordre total de toutes les catégories socioprofessionnelles et de l'ensemble des diplômes.

Dans l'application que je présenterai sur des données françaises, *le tableau de contingence à trois dimensions $NO \times NE \times NC$ est de dimension $8 \times 7 \times 13$* avec un échantillon total de 240 000 individus issus de sept enquêtes nationales représentatives.

QUESTIONS SOCIOLOGIQUES ET ANALYSE « STANDARD »

Depuis la fin des années 1970 ou le début des années 1980, les sociologues de la stratification se sont typiquement posés les questions suivantes sur de telles données.

- *Existe-t-il, dans cette société, une association statistique entre la catégorie socioprofessionnelle du père et le diplôme le plus élevé du fils ou de la fille ?*
- *Si oui, cette association statistique est-elle restée constante dans la société au fil des cohortes de naissance ?*
- *Ou bien a-t-elle changé vers un affaiblissement ou un renforcement au fil des cohortes de naissance ?*

Pour répondre à ces questions, les sociologues de la stratification ont typiquement utilisé *l'analyse log-linéaire standard avec des modèles dits « de type hiérarchique »*.

QUELQUES NOTATIONS

Soit c l'une des cohortes de naissance (parmi NC), o et o' deux origines sociales quelconques (parmi NO), e et e' deux diplômes quelconques (parmi NE).

Avec ces notations, m_{oec} désigne l'effectif estimé dans la cellule (o, e, c) du tableau de contingence à trois dimensions.

Et od_c est le odds ratio qui mesure, dans la cohorte c , la force intrinsèque de l'association statistique entre ces origines sociales et ces diplômes :

$$od_c = \frac{m_{oec}/m_{oe'c}}{m_{o'ec}/m_{o'e'c}}$$

ou encore :

$$\text{Log}(od_c) = \text{Log}(m_{oec}) + \text{Log}(m_{o'e'c}) - \text{Log}(m_{oe'c}) - \text{Log}(m_{o'ec})$$

Les sociologues de la stratification ont, de manière habituelle, considéré *trois modèles log-linéaires hiérarchiques*.

Modèle d'absence d'association ou Modèle d'indépendance conditionnelle

$$\text{Log}(m_{oec}) = \lambda + \lambda_o^O + \lambda_e^E + \lambda_c^C + \lambda_{oc}^{OC} + \lambda_{ec}^{EC}$$

Estimé avec NC(NO-1)(NE-1) degrés de liberté, ce modèle implique :

$$\text{Log}(od_c) = 0 \quad \text{or} \quad od_c = 1$$

Il suppose donc que les variables 'Origine' et 'Éducation' sont statistiquement indépendantes dans chaque cohorte de naissance.

Par conséquent, ce modèle exprime l'hypothèse d'une complète égalité des chances sociales devant l'enseignement.

Et il nous fournit une référence pour apprécier le degré auquel des modèles plus réalistes s'ajustent plus correctement aux données.

Modèle d'association constante ou Modèle d'absence d'interaction des trois variables

$$\text{Log}(m_{oec}) = \lambda + \lambda_o^O + \lambda_e^E + \lambda_c^C + \lambda_{oc}^{OC} + \lambda_{ec}^{EC} + \lambda_{oe}^{OE}$$

Estimé avec (NO-1)(NE-1)(NC-1) degrés de liberté, ce modèle implique :

$$\text{Log}(od_c) = \lambda_{oe}^{OE} + \lambda_{o'e'}^{OE} - \lambda_{oe'}^{OE} - \lambda_{o'e}^{OE}$$

Il suppose donc que tous les odds ratios qui mesurent l'association statistique entre origine sociale et éducation obtenue sont constants au fil des cohortes de naissance.

Par conséquent, il exprime l'hypothèse d'une inégalité des chances sociales devant l'enseignement rigoureusement constante dans le temps.

Modèle saturé

$$\text{Log}(m_{oec}) = \lambda + \lambda_o^O + \lambda_e^E + \lambda_c^C + \lambda_{oc}^{OC} + \lambda_{ec}^{EC} + \lambda_{oe}^{OE} + \lambda_{oec}^{OEC}$$

Estimé avec 0 degré de liberté, le modèle saturé reproduit parfaitement les données observées et implique :

$$\text{Log}(od_c) = (\lambda_{oe}^{OE} + \lambda_{o'e'}^{OE} - \lambda_{oe'}^{OE} - \lambda_{o'e}^{OE}) + (\lambda_{oec}^{OEC} + \lambda_{o'e'c}^{OEC} - \lambda_{oe'c}^{OEC} - \lambda_{o'ec}^{OEC})$$

i.e. chaque odds ratio qui mesure l'association statistique entre deux origines sociales et deux diplômes varie librement d'une cohorte de naissance à une autre.

SUR LES DONNÉES FRANÇAISES...

Modèle	G ²	ddl	ID	rG ²	Bic
<i>Hommes et femmes (N=240 367)</i>					
1. Absence d'association	50 388.8	546	15.9	-	43 623.9
2. Association constante	4 107.8	504	4.1	91.8	-2 136.7
5. Modèle saturé	0.0	0	0	100.0	0.0

Le modèle d'absence d'association est extrêmement éloigné des données :

- le chi-deux du rapport de vraisemblance dépasse 50 000 ;
- selon l'indice de dissimilarité, le modèle classe de manière erronée près de 16 % de l'effectif total.

À l'inverse, le modèle d'association constante semble particulièrement apte à décrire la société française au fil de 13 cohortes quinquennales :

- il classe de manière erronée 4,1 % seulement de l'effectif total ;
- il élimine près de 92 % de la distance totale qui sépare les données du modèle d'absence d'association ;
- en raison de la valeur négative de la statistique Bic (*Bayesian Information Criterion*), le modèle d'association constante doit être préféré au modèle saturé.

QUEL EST LE PROBLÈME ?

Le modèle d'association constante apparaît relativement proche des données. *Cela suggère que l'association statistique entre origine sociale et éducation obtenue se caractérise par une forte inertie au fil des cohortes...*

...Et les sociologues ont donc été enclins à conclure en faveur d'une persistance remarquable de l'inégalité des chances devant l'enseignement (cf. Garnier & Raffalovich, 1984 ; Shavit & Blossfeld (dir.), 1993).

MAIS...

La différence entre le modèle d'association constante et le modèle saturé s'avère hautement significative.

Cependant, ce test a l'inconvénient de disséminer sa puissance statistique sur un ensemble de $(NO-1)(NE-1)(NC-1)$ degrés de liberté (504 ici !).

Enfin, il est inenvisageable d'examiner les 504 paramètres indépendants pour l'interaction entre les trois variables. En d'autres termes, on ne peut véritablement découvrir en quoi l'association entre origine sociale et éducation obtenue a varié au fil des cohortes de naissance !

AINSI, QUELLE LIMITE L'ANALYSE LOG-LINÉAIRE

STANDARD RENCONTRE-T-ELLE ?

Sauf si l'on décide de se focaliser sur un odds ratio particulier ou sur un petit nombre d'odds ratios, *l'analyse log-linéaire standard est incapable de modéliser de façon parcimonieuse la manière dont l'association statistique entre deux variables (ici 'Origine' & 'Éducation') varie selon les modalités d'une troisième variable (ici 'Cohorte')*.

MAIS...

En suivant une suggestion de Leo A. Goodman dans *International Statistical Review* (1986)...

UNE SOLUTION A ÉTÉ PROPOSÉE DANS LE CADRE

D'UNE MODÉLISATION LOG-MULTIPLICATIVE...

LE 'LOG-MULTIPLICATIVE LAYER EFFECT MODEL',
encore appelé LE MODÈLE UNIDIFF (pour 'uniform difference')

Il a été proposé en 1992, simultanément par Yu Xie (dans *American Sociological Review*) et par Robert Erikson et John Goldthorpe (dans leur livre *The Constant Flux*).

$$\text{Log}(m_{oec}) = \lambda + \lambda_o^O + \lambda_e^E + \lambda_c^C + \lambda_{oc}^{OC} + \lambda_{ec}^{EC} + \beta_c \psi_{oe}$$

Estimé avec (NO.NE-NO-NE)(NC-1) degrés de liberté (*i.e. seulement (NC-1) degrés de liberté de moins que le modèle d'association constante !*), ce modèle implique :

$$\text{Log}(od_c) = \beta_c (\psi_{oe} + \psi_{o'e'} - \psi_{oe'} - \psi_{o'e})$$

i.e. il décompose le logarithme de chaque odds ratio comme le produit d'une structure commune (l'expression formée des paramètres ψ_{oe}) et d'un paramètre spécifique à la cohorte (β_c).

Ainsi, *en supposant une structure stable pour l'association statistique entre origine sociale et éducation obtenue, ce modèle est apte à détecter des différences entre cohortes dans la force de l'association, i.e. dans le niveau général de l'inégalité des chances devant l'enseignement.*

Plus précisément, en supposant β_1 fixé à 1, estimer β_c comme étant inférieur à 1 (respectivement supérieur à 1) pour une cohorte ultérieure correspondra au fait que tous les *logged odds ratios* estimés se rapprochent de 0 (respectivement s'éloignent de 0), i.e. correspondra au fait que l'association est plus faible (respectivement plus forte) dans la cohorte considérée que dans la première cohorte.

Ce modèle suppose donc que tous les odds ratios varient dans la même direction d'une cohorte à la suivante et il exprime cette variation au moyen d'un paramètre unique.

En conséquence, *il est très puissant pour détecter une tendance dominante dans les données, mais il peut être aussi un peu fruste pour décrire précisément le changement qui est intervenu.*

SUR LES DONNÉES FRANÇAISES...

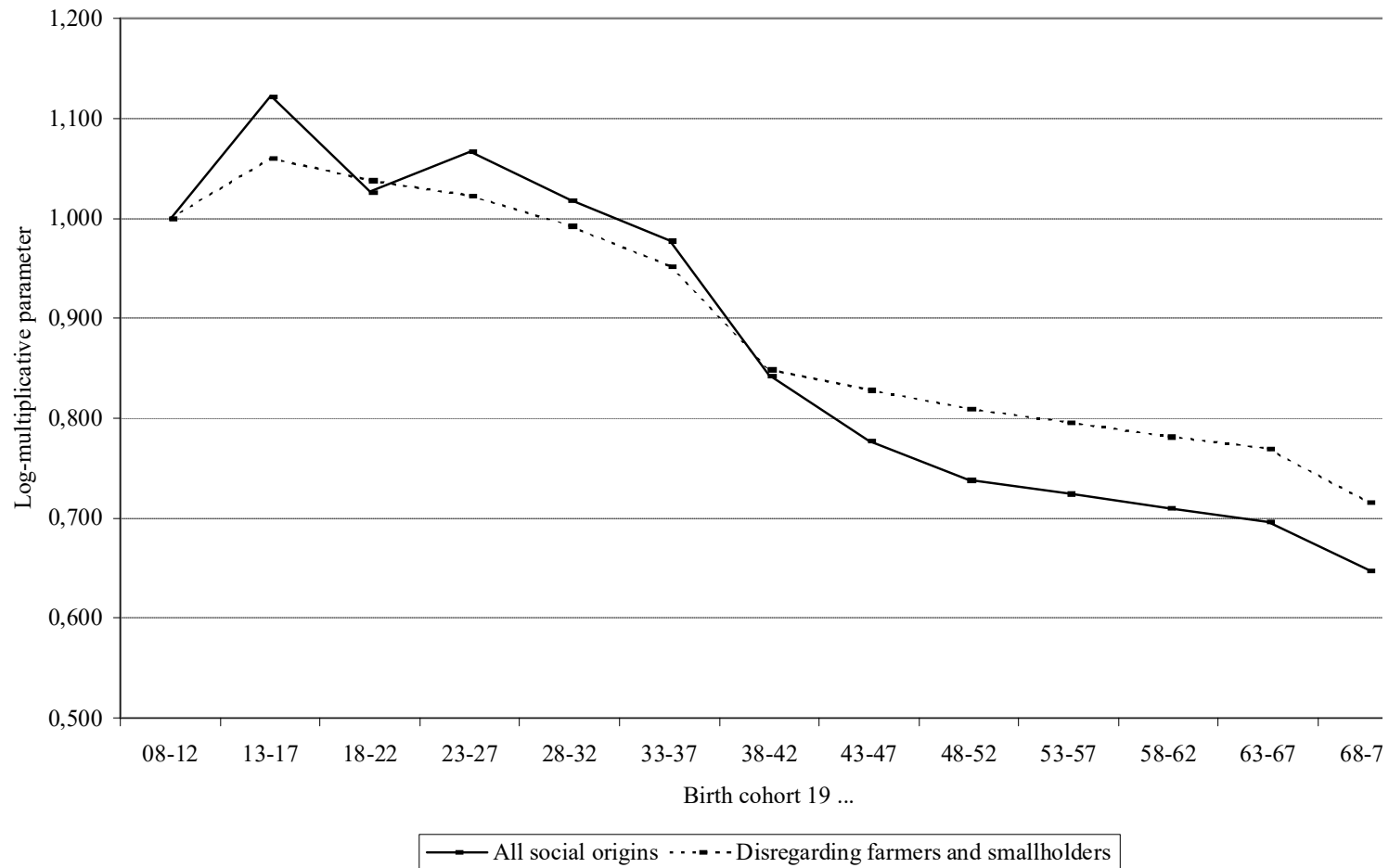
Modèle	G^2	ddl	ID	rG^2	Bic
<i>Hommes et femmes (N=240 367)</i>					
1. Absence d'association	50 388.8	546	15.9	-	43 623.9
2. Association constante	4 107.8	504	4.1	91.8	-2 136.7
3. Modèle Unidiff	3 442.1	492	3.9	93.2	-2 653.7
5. Modèle saturé	0.0	0	0	100.0	0.0

Le test de la différence entre le modèle Unidiff et le modèle d'association constante est hautement significatif: la différence pour le chi-deux du rapport de vraisemblance vaut 665,7 pour 12 degrés de liberté.

Et, d'après la valeur plus négative de la statistique Bic, le nouveau modèle doit aussi être préféré au modèle d'association constante.

On peut donc analyser la dynamique temporelle que les paramètres β_c révèlent...

CHANGE IN THE GENERAL STRENGTH OF ASSOCIATION BETWEEN SOCIAL ORIGIN AND
EDUCATIONAL DESTINATION – ‘MULTIPLICATIVE UNIFORM COHORT EFFECT’ ASSOCIATION MODEL
ESTIMATED OVER THIRTEEN FIVE-YEAR BIRTH COHORTS (N=240,367)



Dans la société française, la force générale de l'association statistique entre origine sociale et éducation obtenue a diminué de plus de 35 % (dans l'échelle des logarithmes des odds ratios) en soixante ans.

Presque monotone, l'affaiblissement du lien entre origine sociale et éducation obtenue a été spécialement prononcé entre les cohortes 1933-1937 et 1948-1952. Il s'est largement interrompu pour les trois cohortes quinquennales suivantes, mais a repris de nouveau dans la dernière (1968-1972).

La ligne en pointillés correspond à une seconde analyse qui ignore les enfants d'agriculteurs exploitants (en utilisant la technique dite « des zéros structurels »).

Elle montre que le changement intervenu dans les trajectoires scolaires des enfants d'agriculteurs a joué un rôle important pour renforcer la tendance générale, mais n'a pas constitué l'unique facteur à son origine.

Il est cependant possible d'être plus précis avec un second modèle log-multiplicatif un peu plus complexe...

LE 'LOG-MULTIPLICATIVE REGRESSION-TYPE MODEL'

C'est un modèle plus général proposé par Leo A. Goodman et Michael Hout dans *Sociological Methodology 1998 & 2001*.

$$\text{Log}(m_{oec}) = \lambda + \lambda_o^O + \lambda_e^E + \lambda_c^C + \lambda_{oc}^{OC} + \lambda_{ec}^{EC} + \lambda_{oe}^{OE} + \gamma_c \varphi_{oe}$$

Estimé avec (NO.NE-NO-NE)(NC-2) degrés de liberté, ce modèle implique :

$$\text{Log}(od_c) = (\lambda_{oe}^{OE} + \lambda_{o'e'}^{OE} - \lambda_{oe'}^{OE} - \lambda_{o'e}^{OE}) + \gamma_c (\varphi_{oe} + \varphi_{o'e'} - \varphi_{oe'} - \varphi_{o'e})$$

Les paramètres λ_{oe}^{OE} établissent *la structure de base de l'association statistique* entre origine sociale et éducation obtenue.

Les paramètres φ_{oe} représentent *la part de l'association statistique qui varie selon la cohorte de naissance*.

Enfin, la valeur du paramètre γ_c exprime *la force de l'ajustement de l'association pour la cohorte c*.

En conséquence, ce modèle est apte à détecter des différences entre cohortes à la fois dans la structure et dans la force de l'association.

Plus précisément, l'examen des paramètres φ_{oe} va mettre en lumière les combinaisons « origine sociale – diplôme » pour lesquelles la transformation au fil des cohortes a été la plus prononcée.

Et l'examen des paramètres γ_c montrera quelles cohortes de naissance ont été les plus affectées par la transformation de l'association statistique.

Les contraintes suivantes d'identification sont spécialement utiles :

$$\sum_o \lambda_{oe}^{OE} = \sum_e \lambda_{oe}^{OE} = \sum_o \varphi_{oe} = \sum_e \varphi_{oe} = 0 \quad , \quad \gamma_1 = 1 \quad \text{and} \quad \gamma_{NC} = 0$$

de sorte que les paramètres λ_{oe}^{OE} représenteront la structure de l'association dans la dernière cohorte (la plus jeune) et les paramètres φ_{oe} représenteront l'écart de la structure de l'association dans la première cohorte (la plus âgée) relativement à la dernière.

SUR LES DONNÉES FRANÇAISES...

Modèle	G^2	ddl	ID	rG^2	Bic
<i>Hommes et femmes (N=240 367)</i>					
1. Absence d'association	50 388.8	546	15.9	-	43 623.9
2. Association constante	4 107.8	504	4.1	91.8	-2 136.7
3. Modèle Unidiff	3 442.1	492	3.9	93.2	-2 653.7
4. Modèle de Goodman-Hout	799.2	451	1.7	98.4	-4 788.6
5. Modèle saturé	0.0	0	0	100.0	0.0

Le test de la différence entre le modèle de Goodman-Hout et le modèle Unidiff est hautement significatif : la différence pour le chi-deux du rapport de vraisemblance vaut 2 642,9 pour 41 degrés de liberté.

D'après la valeur plus négative de la statistique Bic, le nouveau modèle doit aussi être préféré au modèle Unidiff.

Enfin, la proximité du nouveau modèle aux données (ou au modèle saturé) est impressionnante : même si la taille d'échantillon dépasse 240 000, la statistique de chi-deux vaut moins de 800 pour 451 degrés de liberté et l'indice de dissimilarité vaut 1.7 % seulement.

STRUCTURE DE BASE DE L'ASSOCIATION (PARAMÈTRES λ_{oe}^{OE})
DANS LA COHORTE 1968-1972

λ_{oe}^{OE} parameters	No diploma	Primary education certificate	Lower secondary diploma	Lower vocational diploma	Upper secondary diploma	Lower tertiary degree	Upper tertiary degree
Farmers and smallholders	-0.190	0.492	-0.021	0.529	0.033	-0.132	-0.711
Artisans and shopkeepers	-0.036	0.042	0.032	0.125	-0.036	-0.054	-0.073
Higher-grade professionals and managers	-0.479	-1.064	-0.089	-0.640	0.284	0.650	1.339
Teachers and assimilated occupations	-0.583	-1.446	-0.201	-0.589	0.421	0.797	1.602
Lower-grade professionals and technicians	-0.271	-0.319	0.033	-0.203	0.123	0.294	0.342
Routine non manual workers	0.215	0.327	0.035	0.083	-0.072	-0.212	-0.375
Foremen and skilled manual workers	0.508	0.821	0.031	0.268	-0.281	-0.501	-0.847
Agricultural and unskilled manual workers	0.836	1.146	0.181	0.428	-0.472	-0.842	-1.277

Dans la cohorte 1968-1972, les diplômes les plus élevés (enseignement supérieur court ou long) sont fortement et positivement associés avec les origines sociales les plus favorisées (enseignants et assimilés, puis gros indépendants et cadres supérieurs).

Inversement, ils sont fortement et négativement associés avec les origines sociales les moins favorisées (ouvriers non qualifiés et agricoles, puis contremaîtres et ouvriers qualifiés, et agriculteurs exploitants pour les diplômes de l'enseignement supérieur long seulement).

Une structure opposée caractérise l'association statistique entre les mêmes groupes sociaux et les catégories « sans diplôme » et « certificat d'études primaires ».

Enfin, obtenir un diplôme de l'enseignement professionnel court est positivement associé avec une origine sociale parmi les agriculteurs exploitants et les ouvriers non qualifiés et agricoles, et négativement associé avec les deux origines sociales les plus favorisées.

ÉCART DE STRUCTURE DE L'ASSOCIATION (PARAMÈTRES φ_{oe})
 DANS LA COHORTE 1908-1912 (PAR RAPPORT À LA COHORTE 1968-1972)

φ_{oe} parameters	No diploma	Primary education certificate	Lower secondary diploma	Lower vocational diploma	Upper secondary diploma	Lower tertiary degree	Upper tertiary degree
Farmers and smallholders	2.325	0.883	-0.153	-0.828	-0.915	-1.006	-0.307
Artisans and shopkeepers	0.001	0.170	0.142	-0.032	-0.017	-0.235	-0.029
Higher-grade professionals and managers	-0.709	0.135	0.032	0.067	0.297	0.205	-0.027
Teachers and assimilated occupations	-1.333	0.227	0.544	-0.490	0.302	0.579	0.170
Lower-grade professionals and technicians	-0.847	-0.203	0.189	0.315	0.279	0.095	0.172
Routine non manual workers	-0.531	-0.399	0.107	0.282	0.008	0.279	0.256
Foremen and skilled manual workers	0.320	-0.400	-0.207	0.518	0.058	-0.021	-0.268
Agricultural and unskilled manual workers	0.774	-0.414	-0.653	0.169	-0.013	0.104	0.033

Prendre la valeur absolue des paramètres φ_{oe} et calculer leur moyenne par ligne et par colonne met en lumière les origines sociales dont la position relative a été la plus transformée et les diplômés qui ont contribué le plus au changement historique dans la structure de l'association.

En particulier, dans le passé et comparativement à la cohorte 1968-1972, la catégorie « sans diplôme » était plus fortement positivement associée avec les origines sociales les moins favorisées (tout spécialement agriculteurs exploitants) et plus fortement négativement associée avec les deux origines les plus favorisées comme avec le groupe des cadres moyens et celui des employés.

Étudier le profil des paramètres φ_{oe} par ligne révèle qu'en soixante ans et pour ce qui concerne leurs trajectoires scolaires, les fils et filles d'agriculteurs exploitants ont considérablement amélioré leur position relative.

Au contraire, la situation des enfants d'artisans et petits commerçants est demeurée beaucoup plus stable.

Dans les premières décennies du XX^e siècle, les hommes et femmes originaires de quatre groupes sociaux étaient caractérisés par des trajectoires scolaires encore plus favorables qu'elles ne le sont dans les cohortes les plus récentes : enseignants et assimilés, gros indépendants et cadres supérieurs, cadres moyens, employés.

Enfin, pour les enfants des fractions qualifiée et non qualifiée de la classe ouvrière, le changement historique des trajectoires scolaires apparaît plus complexe.

Les enfants de contremaîtres et ouvriers qualifiés ont davantage amélioré leur position relative que les enfants d'ouvriers non qualifiés et agricoles.

Dans la première cohorte relativement à la dernière, les premiers étaient plus positivement associés avec les catégories « sans diplôme » et « enseignement professionnel court » et plus négativement associés avec la catégorie « enseignement supérieur long ». Mais l'association statistique n'a guère varié entre le groupe des ouvriers non qualifiés et agricoles et les diplômes les plus élevés (baccalauréat ou équivalent, enseignement supérieur court ou long).

VALEUR DE L'AJUSTEMENT DE L'ASSOCIATION (PARAMÈTRES γ_c)

AU FIL DES COHORTES DE NAISSANCE

Birth cohort	1908-1912	1913-1917	1918-1922	1923-1927	1928-1932	1933-1937	1938-1942
γ_c parameters	1	0.891	0.758	0.867	0.772	0.673	0.519
Birth cohort	1943-1947	1948-1952	1953-1957	1958-1962	1963-1967	1968-1972	
γ_c parameters	0.357	0.241	0.200	0.175	0.101	0	

En dehors d'une exception entre les cohortes 1918-1922 et 1923-1927, la force de l'ajustement de l'association décline de façon monotone sur l'ensemble de la période.

De nouveau, le changement est spécialement marqué entre les cohortes 1933-1937 et 1948-1952.

SUR LE MODÈLE UNIDIFF DE NOUVEAU...

Dans l'ouvrage *Social Mobility in Europe* (Breen (dir.), 2004, Oxford University Press), le chapitre sur la France met en évidence un affaiblissement de l'association statistique entre classe d'origine (i.e. la classe du père, O) et classe de destination (i.e. la classe du répondant, D) pour les hommes âgés de 25 à 64 ans.

Le paramètre log-multiplicatif du modèle Unidiff (β_t), fixé à 1 en 1970, est estimé à :

0,953 en 1977

0,887 en 1985

et 0,873 en 1993.

Ainsi, la fluidité sociale entre générations a augmenté d'environ 13 % (dans l'échelle des logarithmes des odds ratios) en un quart de siècle.

Et, pour tenter d'expliquer un tel changement, on peut de nouveau utiliser des versions plus sophistiquées du modèle Unidiff...

INTRODUIRE DES TERMES LOG-MULTIPLICATIFS DANS UN MODÈLE TOPOLOGIQUE « MULTI-COUCHES » DU TABLEAU DE MOBILITÉ

Le modèle log-linéaire dit « topologique » du tableau de mobilité (Hauser, 1978) suppose que la structure et la force de l'association statistique entre classe d'origine et classe de destination peuvent être décrites par un nombre limité de paramètres d'interaction (ou « niveaux de densité »).

Erikson et Goldthorpe (1992) ont proposé une telle représentation théorique de la fluidité sociale dans les sociétés modernes. *Dans leur modèle « noyau », les « niveaux de densité » résultent de la combinaison de huit paramètres plus élémentaires.*

Ces paramètres capturent différents effets :

- *Hiérarchie sociale : HI1 (mobilité de courte distance) & HI2 (mobilité de longue distance) ;*
- *Hérédité sociale : IN1 (toutes les classes), IN2 (classes avec indépendants seulement) & IN3 (agriculteurs exploitants seulement) ;*
- *Secteur : SE (différencie entre le secteur primaire et les autres secteurs) ;*
- *Affinité : AF1 (négative) & AF2 (positive).*

REPRÉSENTATION THÉORIQUE DE LA FLUIDITÉ SOCIALE « NOYAU » DANS LA TABLE DE MOBILITÉ 7 x 7

D O	I+II	III	IVa+b	IVc	V+VI	VIIa	VIIb
I+II	IN1+IN2	HI1+AF2	HI1+AF2	HI1+SE	HI1	HI1+HI2	HI1+HI2+SE+AF1
III	HI1+AF2	IN1	-	SE	-	HI1	HI1+SE
IVa+b	HI1+AF2	-	IN1+IN2	SE+AF2	-	HI1	HI1+SE
IVc	HI1+HI2+SE	HI1+SE	HI1+SE+AF2	HI1+IN1+IN2+IN3	HI1+SE	SE+AF2	-
V+VI	HI1	-	-	SE	IN1	HI1+AF2	HI1+SE
VIIa	HI1+HI2	HI1	HI1	HI1+SE	HI1+AF2	IN1	SE
VIIb	HI1+HI2+SE+AF1	HI1+SE	HI1+SE	HI1	HI1+SE	SE+AF2	IN1

Les divisions hiérarchiques dans la structure sociale figurent horizontalement et les divisions sectorielles figurent verticalement.

Le tableau montre comment les paramètres élémentaires se combinent pour constituer les « niveaux de densité » caractéristiques de la table de mobilité 7 x 7.

Sur la base de cette représentation théorique, il est possible d'identifier les facteurs responsables de la variation de la fluidité sociale comme suit...

Model (seven-class schema)	G ²	df	DI	rG ²	Bic			
A. Constant social fluidity	300.1	108	2.4	98.6	-881.4			
Core Models								
B. Temporally invariant parameters	568.6	136	3.3	97.4	-919.1			
	<i>H11</i>	<i>H12</i>	<i>IN1</i>	<i>IN2</i>	<i>IN3</i>	<i>SE</i>	<i>AF1</i>	<i>AF2</i>
<i>France 1970, 1977, 1985, 1993</i>	-0.243	-0.570	0.399	0.824	1.140	-0.803	-0.817	0.451
<i>France 1970 (The Constant Flux, p. 147)</i>	-0.24	-0.47	0.41	0.92	1.00	-0.89	-0.75	0.47
C. Temporally changing parameters	451.0	112	2.9	97.9	-774.2			
	<i>H11</i>	<i>H12</i>	<i>IN1</i>	<i>IN2</i>	<i>IN3</i>	<i>SE</i>	<i>AF1</i>	<i>AF2</i>
<i>1970</i>	-0.318	-0.703	0.350	0.975	0.935	-0.931	-0.572	0.459
<i>1977</i>	-0.289	-0.574	0.393	0.801	1.424	-0.665	-1.156	0.470
<i>1985</i>	-0.171	-0.538	0.425	0.742	1.091	-0.777	-0.767	0.426
<i>1993</i>	-0.083	-0.441	0.481	0.809	1.264	-0.687	-0.665	0.407
D. Only H11 changes over time	515.5	133	3.1	97.6	-939.5			
E. Only H12 changes over time	544.7	133	3.3	97.5	-910.2			
F. Only IN1 changes over time	550.0	133	3.3	97.5	-904.9			
G. Only IN2 changes over time	532.4	133	3.2	97.5	-922.6			
H. Only IN3 changes over time	566.4	133	3.3	97.4	-888.5			
I. Only SE changes over time	555.2	133	3.3	97.4	-899.8			
J. Only AF1 changes over time	561.5	133	3.3	97.4	-893.4			
K. Only AF2 changes over time	567.8	133	3.4	97.4	-887.1			
L. H11 and H12 change over time	499.5	133	3.1	97.7	-955.4			
	<i>UNIDIFF for H11 and H12</i>	<i>1.000 (1970)</i>	<i>0.830 (1977)</i>	<i>0.619 (1985)</i>	<i>0.509 (1993)</i>			
M. Model L with a linear trend	500.7	135	3.1	97.7	-976.1			
	<i>Linear trend per year for H11 and H12</i>	<i>-0.0230</i>						
N. Model M + SE changes over time	489.7	132	3.0	97.7	-954.3			
	<i>Linear trend per year for H11 and H12</i>	<i>-0.0227</i>						
	<i>UNIDIFF for SE</i>	<i>1.000 (1970)</i>	<i>0.825 (1977)</i>	<i>0.839 (1985)</i>	<i>0.846 (1993)</i>			
O. Model N with equality constraint	489.8	134	3.0	97.7	-976.1			
	<i>Linear trend per year for H11 and H12</i>	<i>-0.0226</i>						
	<i>UNIDIFF (constrained) for SE</i>	<i>1.000 (1970)</i>	<i>0.834 (1977, 1985, 1993)</i>					

La version temporellement invariante du modèle noyau (modèle B) se compare très favorablement avec le modèle de fluidité sociale constante (modèle A) : sa statistique BIC est meilleure et elle classe de manière erronée 3,3 % seulement de l'effectif total. Les huit paramètres estimés à partir des quatre enquêtes sont aussi très proches de ceux obtenus par Erikson et Goldthorpe à partir de la seule enquête française de 1970.

Avec 24 paramètres supplémentaires, la version temporellement variable du modèle noyau (modèle C) fournit une statistique G^2 plus basse de 117,6 points. Dans le contexte du modèle noyau, cela représente toute la variation de la fluidité sociale intervenue en 23 ans, mais l'on comprendrait mieux cette variation si l'on était capable de la résumer en grande partie, en utilisant un petit nombre de paramètres, en plus des huit effets élémentaires.

Dans une première étape, j'estime une série de modèles qui incorporent un effet Unidiff au cours du temps pour *l'un seulement* des huit paramètres élémentaires. En utilisant la version temporellement invariante du modèle noyau comme référence, une forte amélioration de la statistique G^2 apparaît quand cet effet est appliqué à HI1 (modèle D), et secondairement à IN2 et HI2 (modèles G and E).

Dans le modèle L, j'impose donc un effet Unidiff sur les deux paramètres de hiérarchie *simultanément*, et cela procure à ce stade le modèle qui s'ajuste le mieux aux données. En outre, la variation monotone traduite par le paramètre Unidiff peut être résumée, sans perte d'information notable, par une tendance linéaire (modèle M).

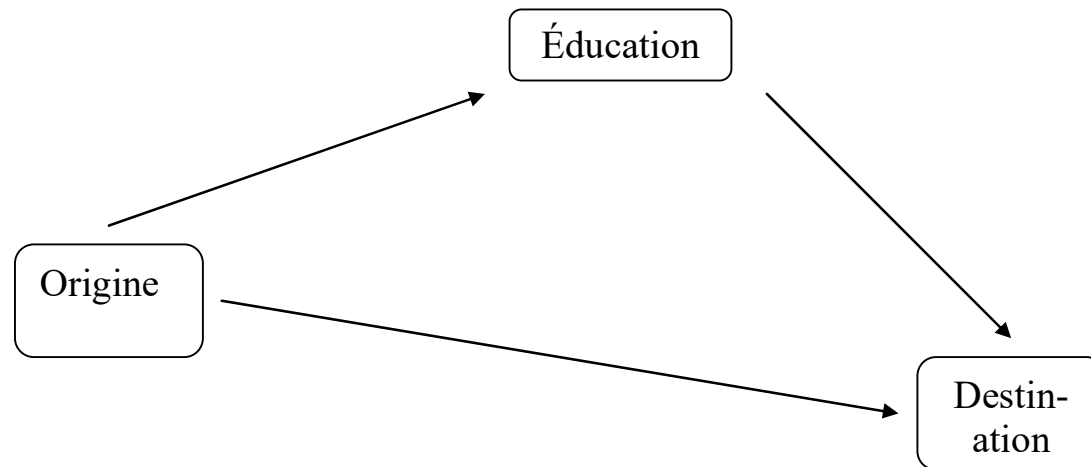
Enfin, ajouter un autre effet Unidiff au paramètre de secteur diminue significativement la statistique G^2 (modèle N) et cela peut être représenté par un effet de seuil qui oppose l'enquête de 1970 aux suivantes (modèle O).

Bien que j'ai examiné de nombreuses variantes supplémentaires du modèle noyau, je n'ai pu trouver de modèle plus puissant que le modèle O : avec seulement deux paramètres, il élimine 67,1 % de la distance entre les versions temporellement invariante et temporellement variable du modèle noyau.

Et ces paramètres nous fournissent l'interprétation suivante : l'accroissement de la fluidité sociale a résulté principalement d'un affaiblissement progressif de la dimension hiérarchique de la structure sociale ; il a aussi résulté d'une réduction de la distance entre les classes des agriculteurs exploitants et ouvriers agricoles et les autres classes.

ANALYSER LES TENDANCES TEMPORELLES DANS LES LIENS ENTRE ORIGINE SOCIALE, ÉDUCATION & POSITION SOCIALE

Si l'on introduit le niveau d'éducation comme variable intermédiaire, l'affaiblissement du lien entre classe d'origine et classe de destination peut résulter de quatre effets différents :



- un lien plus faible entre Origine et Éducation ;
- une association partielle plus faible entre Éducation et Destination ;
- une association partielle plus faible entre Origine et Destination ;
- un effet de composition (l'association partielle entre Origine et Destination est plus faible dans les groupes plus éduqués dont l'importance relative dans la population active a augmenté entre 1970 et 1993).

Et le test de ces quatre hypothèses peut être effectué ainsi...

Model	G ²	df	DI	rG ²	Bic
<i>Analysis of the Origin-Education-Time table</i>					
A. Conditional independence {OT}{ET}	12 720.2	192	16.5	-	10 619.8
B. Constant association {OT}{ET}{OE}	418.4	144	2.8	96.7	-1 156.9
C. UNIDIFF	335.9	141	2.5	97.4	-1 206.5
	<i>UNIDIFF parameters for {OE} 1.000(1970) 0.886(1977) 0.808(1985) 0.784(1993)</i>				
<i>Analysis of the Origin-Education-Destination-Time table</i>					
D. Conditional independence {OET}{DT}	41 547.0	1 482	32.6	-	25 334.8
E. Constant {OD} association {OET}{DT}{OD}	20 126.4	1 446	20.4	51.6	4 307.9
F. Constant {ED} association {OET}{DT}{ED}	15 920.4	1 434	17.1	61.7	233.3
G. Constant associations {OET}{DT}{OD}{ED}	2 449.1	1 398	5.8	94.1	-12 844.3
H. Only {OD} changes over time	2 428.1	1 395	5.7	94.2	-12 832.5
	<i>UNIDIFF parameters for {OD} 1.000 0.956 0.894 0.890</i>				
I. Only {ED} changes over time	2 342.4	1 395	5.5	94.4	-12 918.2
	<i>UNIDIFF parameters for {ED} 1.000 0.910 0.855 0.740</i>				
J. Both {OD} and {ED} change over time	2 325.5	1 392	5.4	94.4	-12 902.2
	<i>UNIDIFF parameters for {OD} 1.000 0.961 0.902 0.909</i>				
	<i>UNIDIFF parameters for {ED} 1.000 0.912 0.859 0.743</i>				
K. Model I + {OD} changes over education	2 238.7	1 387	5.4	94.6	-12 934.3
	<i>UNIDIFF (time) parameters for {ED} 1.000(1970) 0.906(1977) 0.851(1985) 0.740(1993)</i>				
	<i>UNIDIFF (education) parameters for {OD} 1a 1b 1c 2a 2b 2cgen 2cvoc 3a 3b</i>				
	<i>1.000 1.089 1.014 0.988 0.858 0.499 0.814 0.725 0.633</i>				

Premièrement, j'ai analysé la dynamique de l'association entre classe d'origine et éducation. Le modèle Unidiff est préférable au modèle d'association constante et ses paramètres révèlent une diminution de 21,6 % (dans l'échelle des logarithmes des odds ratios) pour la force du lien Origine – Éducation entre les populations enquêtées en 1970 et 1993 (modèle C).

Cette diminution s'affaiblit progressivement, conformément à ce qui a été montré plus haut (la plus grande partie du changement est intervenue entre les cohortes du milieu des années 1930 et celles du milieu des années 1950). *Première conclusion : une diminution de l'inégalité des chances scolaires est intervenue.*

Deuxièmement, j'ai analysé le tableau à trois dimensions Origine – Éducation – Destination dans une perspective temporelle.

La statistique rG^2 suggère clairement que l'association partielle Éducation – Destination est plus forte que l'association partielle Origine – Destination (par comparaison des modèles E et F).

Puis, partant du modèle qui incorpore ces deux associations (modèle G), j'introduis un effet Unidiff au cours du temps pour l'une, l'autre ou les deux associations (modèles H, I et J). Les statistiques G^2 et Bic sont clairement en faveur du modèle I qui révèle une diminution de 26,0 % (dans l'échelle des logarithmes des odds ratios) pour la force de l'association Éducation – Destination. *Seconde conclusion : un affaiblissement de l'avantage relatif procuré par le diplôme dans l'accès aux positions sociales est intervenu, mais l'effet direct de l'origine sur la position a peu varié en France sur la période 1970-1993.*

Enfin, un ajustement encore meilleur est obtenu en ajoutant au modèle I un effet Unidiff sur l'éducation pour l'association Origine – Destination (modèle K). Les paramètres estimés révèlent clairement que l'effet direct de l'origine sur la position occupée est généralement plus faible parmi les individus plus diplômés – spécialement à partir du milieu du secondaire (2b) jusqu'à l'enseignement supérieur court ou long (3a et 3b). *Troisième conclusion : un effet de composition est intervenu, accroissant progressivement le poids des catégories de diplômes pour lesquelles l'effet direct de l'origine sociale sur la position occupée est plus réduit.*

LA MODÉLISATION LOG-MULTIPLICATIVE

PEUT AISÉMENT ÊTRE IMPLÉMENTÉE

EN UTILISANT LE LOGICIEL

LEM

Développé par **Jeroen K. Vermunt (Université de Tilburg)**

<http://members.home.nl/jeroenvermunt/>