Institut National d'Etudes Démographiques

Nicolas BROUARD

L'EXTINCTION DES NOMS DE FAMILLE EN FRANCE: UNE APPROCHE

Dossiers et recherches
n° 27

Novembre 1989

INED, 27, rue du Commandeur, 75675 Paris Cedex 14
Tél. 43.20.13.45

Ce numéro des "Dossiers et Recherche" de l'INED correspond à la publication d'un rapport effectué en 1981 pour Monsieur J.F.X Deniau, Député. Un projet de loi assouplissant la transmission et la francisation des noms avait été déposé à cette époque.

Un nouveau projet de loi allant dans ce sens a été très récemment déposé à l'Assemblée Nationale et le présent rapport a été demandé à plusieurs reprises à l'INED. Une véritable publication aurait exigé une mise à jour en tenant compte des travaux effectués depuis dans ce domaine ; ce numéro des "Dossiers et Recherche" permet d'en assurer une diffusion immédiate sous une forme sensiblement améliorée.

Dessin : Nicole Berthoux
Typographie : Nicole Froment

L'extinction des noms de famille en France : une approche

Nicolas Brouard
Chargé de recherche à
l'Institut National d'Etudes Démographiques

Avril 1981

L'extinction des noms de famille est un sujet qui a été souvent abordé par les statisticiens et a donné naissance à toute une partie de la théorie statistique à savoir les processus dits de branchements. On attribue au Révérend H. W. Watson et F. Galton la première solution à ce problème. Ce dernier exprimait en 1874 les préoccupations de l'époque :

> The decay of the families of men who occupied conspicious positions in past times has been a subject of frequent remark, and has given rise to various conjectures. It is not only the families of men of genious or those of the aristocracy who tend to perish, but it is those of all with whom history deals, in any way, even of such men as the burgesses of towns, concerning whom Mr. Doubleday has inquired and written. The instances are very numerous in which surnames that were once common have since become scarce or have wholly disappeared. The tendency is universal, and, in explanation of it, the conclusion has been hastily drawn that a rise in physical comfort and intellectual capacity is necessarily accompanied by diminution in "fertility" – using that phrase in its widest sense and reckoning abstinence from marriage as sterility. If that conclusion be true, our population is chiefly maintained through the "proletariat", and thus a large element of degradation is inseparably connected with those other elements which tend to ameliorate the race. On the other hand Mr. Alphonse De Candolle has directed attention to the fact that, by the ordinary law of chances, a large proportion of families are continually dying out, and it evidently follows that,

> until we know what the proportion is, we cannot estimate whether any observed diminution of surnames among the families whose history we can trace, is or is not a sign of their diminished "fertility".

En fait il semble que ce problème ait été soulevé avant Galton et Watson par Bienaymé qui écrivait en 1845 : *on s'est beaucoup occupé de la multiplication possible des hommes ; et récemment diverses observations très curieuses ont été publiées sur la fatalité qui s'attacherait aux corps de noblesse, de bourgeoisie, aux familles des hommes illustres, etc.; fatalité qui, dit-on, ferait disparaître inévitablement ce qu'on a nommé des familles fermées.*

Depuis, de nombreux démographes et statisticiens ont abordé cette question, et nous pourrions citer A. J. Lotka (1920), T. E. Harris (1963), D. G. Kendall (1966), N. Keyfitz (1968). Qu'ils aient approfondi la théorie des processus de branchement (Harris), ou uniquement appliqué les résultats simples du processus (A. Lotka pour les Etats Unis en 1920, et N. Keyfitz pour le Japon et Israël en 1960), ils se sont attachés à la seule résolution du problème que Galton avait soumis dans un journal pour mathématiciens, le "Educational Times", et non, nous le verrons, au problème global du nombre d'extinction des noms d'une population. Le Révérend H. W. Watson avait ainsi répondu à la petite annonce suivante :

> Problem 4001 : A large nation, of whom we will only concern ourselves with the adult males, N in number, and who each bear separate surnames, colonise a district. Their law of population is such that, in each generation, a per cent of the adult males have no male children who reach adult life ; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five.
>
> Find (1) what proportion of the surnames will have become extinct after r generations; and (2) how many instances there will be of the same surname being held by m persons.

Nous utiliserons dans cette note les résultats de ces auteurs. En les appliquant à des données françaises actuelles, nous pourrons ainsi donner les probabilités d'extinction d'un nom de famille au bout d'une, deux ou plusieurs générations sachant que ce nom est initialement porté par un nombre donné d'hommes, avec bien entendu les réserves nécessaires qu'imposeront nos hypothèses.

Mais nous tenterons aussi d'aborder la question beaucoup plus délicate et peu étudiée à notre connaissance qui concerne le nombre de noms perdus chaque année ou en une génération moyenne dans un pays comme la France. La résolution du problème de Galton-Watson fournit un élément de réponse à notre seconde question si nous connaissons le nombre et la distribution des noms de famille suivant le nombre de porteurs. Mais c'est cette distribution statistique qu'il est difficile d'obtenir. En effet aucun sondage ne permet de répondre pleinement à la question, et seul un dépouillement exhaustif permettrait d'y répondre. En espérant qu'une telle information soit un jour disponible, il n'est guère possible pour le moment que d'étudier les distributions obtenues à partir de quelques sources (sécurité sociale militaire, état civil). On examinera alors les quelques méthodes statistiques qui permettent d'estimer la distribution totale. Le problème est très délicat, voir insoluble car si les fréquences des noms fréquents comme le nom Martin, 0,2 % de la population, sont aisément estimables et transposables à l'ensemble de la population, il n'en est rien des noms rares qui bien évidemment sont les plus menacés d'extinction : un nom rare au niveau national a une probabilité très faible d'apparaître dans un sondage.

Cette note comprendra donc deux parties, l'une concerne l'estimation et la validation du processus de Galton-Watson, l'autre concerne les statistiques réelles sur les noms de famille en France.

1 Pourquoi les noms de famille s'éteignent-ils ?

Le patrimoine national que constitue l'ensemble des noms de famille portés par les français s'est formé au Moyen Age et s'est pratiquement fixé sous François I puis sous Napoléon. Comme les créations et francisations de noms sont relativement peu fréquentes, notre patrimoine ne peut donc que s'appauvrir. Essayons de décrire les mécanismes de cet appauvrissement afin de construire un modèle qui puisse le quantifier.

Comme l'origine du terme l'indique, un patronyme s'hérite de père en fils. Ainsi, si un homme n'a pas de fils, son nom s'éteint à moins qu'il ne soit pas l'unique porteur du nom. Si au contraire il a plusieurs descendants de sexe masculin, la perte du nom est repoussée d'une ou plusieurs générations, et cela d'autant plus facilement que les descendants sont nombreux. Mais si en moyenne les descendants d'un homme sont nombreux et si la mortalité ne sévit pas trop pour permettre à ces descendants d'être eux-mêmes en âge de procréer, on peut comprendre que la population peut croître, et rendre l'extinction d'un nom moins probable. Pourtant même dans une population croissante, le seul fait que la taille des familles puisse varier rend la probabilité d'extinction parfois non négligeable.

Voyons comment nous pouvons formaliser un tel processus, le simplifier, le paramètrer et l'estimer.

2 La modélisation du processus et son estimation

Pour modéliser le processus, on peut considérer des probabilités de mariage suivant l'âge, tenir compte des probabilités de survie jusqu'à cet âge et utiliser un modèle de constitution de famille qui rende bien compte après que la période féconde soit terminée, des distributions de taille des familles réellement observées dans les enquêtes statistiques sur les familles.

Mais un tel modèle utilise généralement la simulation, et il n'est peut-être pas utile d'en utiliser un ici car on peut simplifier facilement notre processus. Il suffit par exemple de distinguer les hommes ou femmes qui se marieront un jour (90 %), des autres. Pour les premiers on considère que leur descendance sera distribuée suivant celle observée dans les enquêtes sur les familles, et qu'elle sera constituée au bout d'un temps moyen, celui de l'âge moyen à la paternité soit 28 ans environ. On ne connaît pas la distribution de la descendance des hommes ou femmes qui ne se sont pas mariés, mais il y a environ 8 % de naissances illégitimes chaque année. Ce chiffre n'est pas négligeable, mais ces naissances illégitimes proviennent en grande partie de jeunes femmes qui se marieront sans doute. Il est alors impossible de connaître la distribution de la descendance des hommes ou femmes qui ont terminé leur vie féconde, mais qui n'étant pas mariés, ne rentrent pas dans le champ des enquêtes sur les familles. Rappelons qu'un recensement ne fournit aucun renseignement utile puisqu'on ne peut connaître l'existence que des seuls enfants âgés de moins de seize ans. Les enquêtes de l'INED ne sont généralement pas assez importantes pour qu'on puisse obtenir de tels résultats avec une bonne précision.

Comme on manque d'information sur les non mariés, on s'est donc attaché à la seule distribution des nombres d'enfants par famille complète fournie par l'enquête sur les familles de 1962 et reproduite dans le tableau 1

Tableau 1: Nombre d'enfants par famille complète. Enquête sur les familles INSEE 1962.

Nombre d'enfants	0	1	2	3	4	5	6 et +	TOTAL
Distribution	165	249	238	145	82	47	74	1000

On sous-estime donc la probabilité d'avoir aucun enfant.

De même, la probabilité de survie jusqu'à l'âge de 28 ans avoisinant 98 %, on a négligé la mortalité. Par contre il a fallu transformer ce tableau en celui de la distribution du nombre de garçons par famille sachant qu'il naît en moyenne 105 garçons pour 100 filles. Les probabilités E_s se transforment en G_s par la formule suivante :

$$G_S = \sum_{i=1}^{\infty} E_i\binom{i}{s} g^s(1-g)^{i-s} \quad \text{où } g{=}105/205. \tag{1}$$

et conduisent au tableau 2

Tableau 2: NOMBRE DE GARÇONC PAR FAMILLE COMPLÈTE.

Nombre de garçons	0	1	2	3	4	5	6et+	TOTAL
Distribution	3751	3328	1760	764	298	84	11	10 000

Ces probabilités sont les données de base de notre modèle. Le nombre moyen de descendants d'un homme est donc $m = \sum_s sG_s = 1.08$ en une génération moyenne; ceci correspond à un taux de croissance annuel de 3 pour mille : $(1,03)^{28} \simeq 1,08$. D'ailleurs peu nous importe le taux de croissance, seule une répartition convenable des tailles des familles nous intéresse. En effet nous allons raisonner à un horizon de quelques générations seulement, et les propriétés asymptotiques fondamentales dans la théorie des processus de Galton-Watson, où on y distingue les trois cas de population croissante, stationnaire et décroissante et qui correspondent à des études théoriques très différentes, ne nous concernent pas car elles ne jouent pas un très grand rôle à court terme.

3 Probabilités d'extinction d'un nom après une génération suivant le nombre de porteurs initial du nom

La distribution $G_0, G_1, \ldots, G_6$ donne les probabilités qu'un monohomonyme disparaisse, reste monohomonyme, ou soit porté par 2, 3 ou plus descendants masculins à la génération suivante. Ainsi si la probabilité d'extinction d'un monohomonyme est $G_0 = 37.5$ %, celle d'un nom porté par deux individus est $G_0^2 = 14$ % et celle d'un nom porté par n individus G_0^n.

On a représenté sur la figure 1 les distributions complètes suivant que

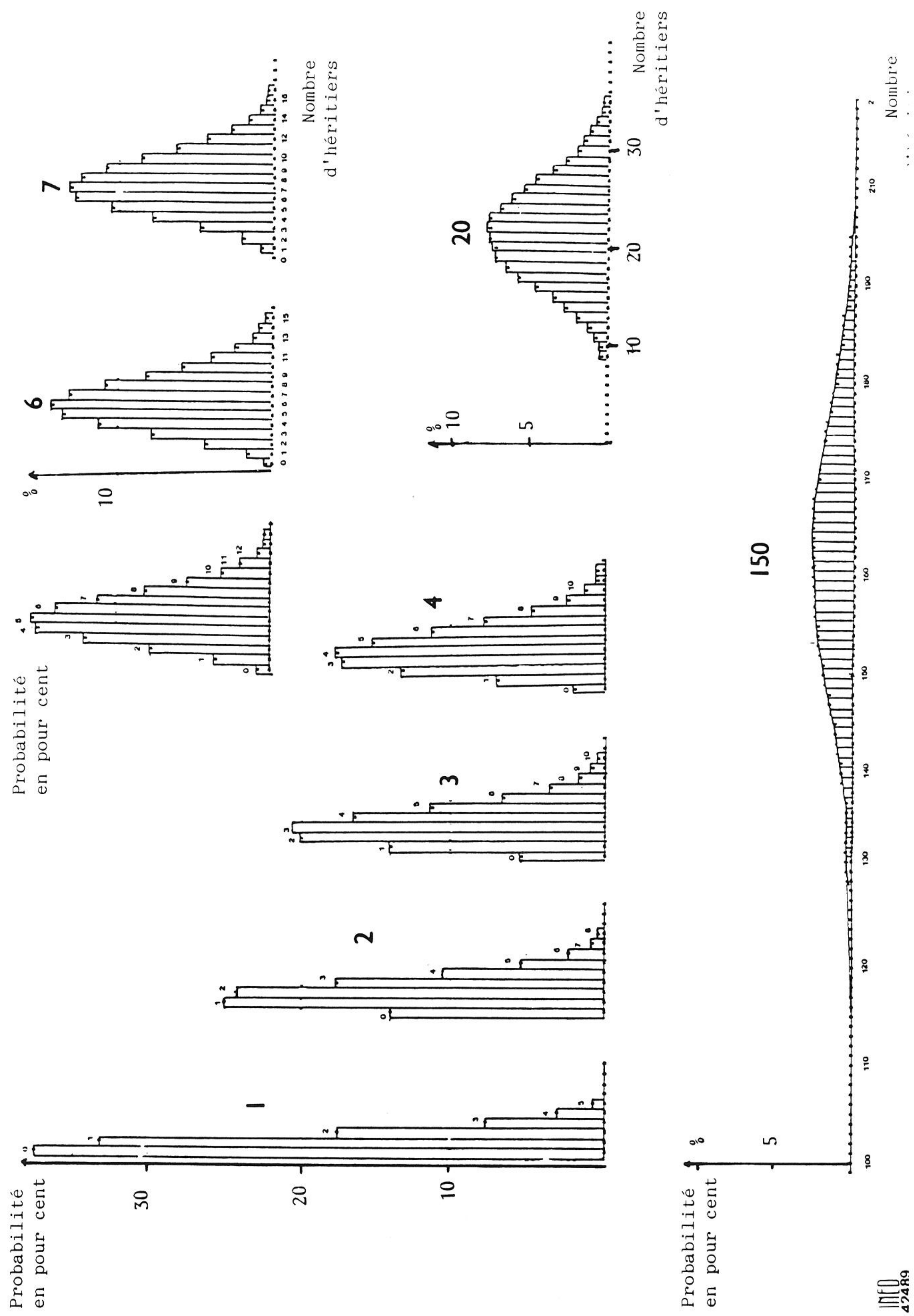


Figure 1 : Distribution de probabilité du nombre d'héritiers d'un même nom après une génération suivant que ce nom est porté

le nom était porté par 1, 2, 3, 4, 5, 6, 7, 20 et 150 individus. Pour un nom porté par 7 individus, la probabilité d'extinction n'est plus que de 1 pour 1000. Mais nous comprenons que si le nombre de noms portés par 7 hommes est supérieur à mille, un de ces noms sera au moins perdu.

Au delà d'une dizaine de porteurs ($n > 10$), la distribution est très bien représentée par une loi normale (loi des grands nombres) de moyenne nm, et de variance $n\sigma^2$ où $m = 1,08$ et $\sigma^2 = 1,28$. Ceci signifie qu'un nom porté par 150 individus sera porté après une génération par un nombre de descendants qui aura 95 % de chance de se trouver entre $162 - 2\sqrt{150 \times 1.28} = 134$ et $162 + 2\sqrt{150 \times 1.28} = 190$. Regardons maintenant ce qu'il en advient au bout de plusieurs générations.

4 Distribution de probabilité des descendants d'un monohomonyme après plusieurs générations

Comme le nombre de descendants d'un monohomonyme ne s'étale après une génération qu'entre 0 et 6, il suffit de pondérer les 6 premières distributions de la figure 2, et de les sommer pour obtenir la distribution des descendants après deux générations. Ainsi on peut voir sur la figure 2 que la probabilité d'extinction d'un monohomonyme est de 53 % au bout de deux générations,, puis de 61 % au bout de trois, 66 % au bout de quatre, etc. Celle-ci tend vers une probabilité limite de 86 % solution de l'équation :

$$s = G_0 + G_1 s + G_2 s^2 + \cdots + G_6 s^6 . \quad (2)$$

La moyenne de cette suite de distributions suit une loi exponentielle en m^n dite de Malthus[1].

Ainsi une partie de la masse de la distribution va se fixer en zéro tandis que l'autre partie 1 - 0,86 = 0,14 va s'étaler à l'infini comme on peut le voir pour la vingtième génération.

Regardons maintenant ce qu'il advient de la descendance d'un nom porté par plusieurs individus.

[1]Sa variance vaut $\frac{\sigma^2 m^n(m^n-1)}{m^2-m}$

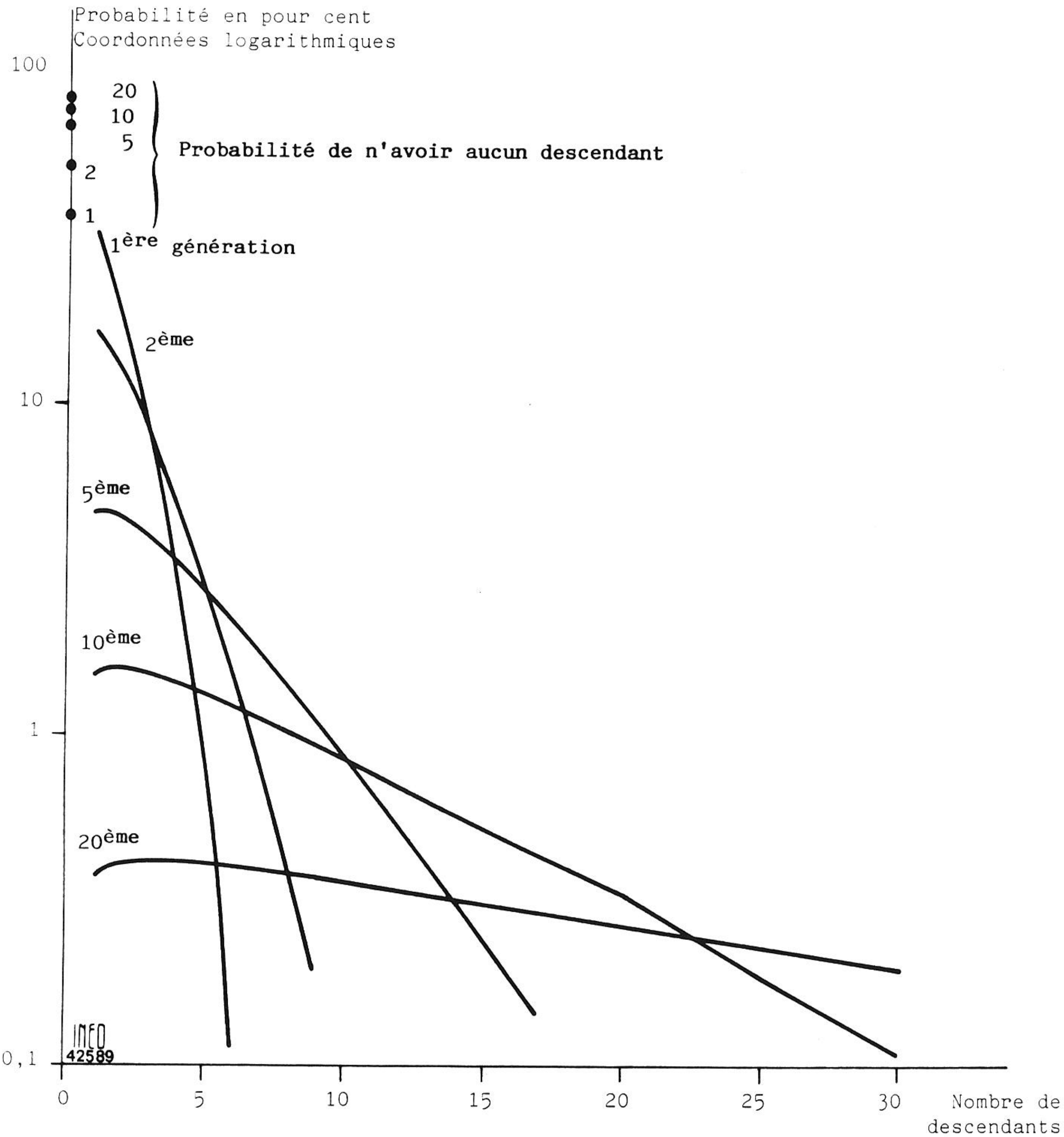


Figure 2 : Distribution de probabilité du nombre de descendants mâles d'un homme après 1, 2, 5, 10 et 20 générations.

5 Descendance après plusieurs générations d'un nom porté par un nombre quelconque d'individus

On a représenté sur la figure 3 la distribution de probabilité des descendants mâles de 5 hommes portant le même nom après 1, 2, 3, 5 et 10 générations. A la seconde génération la probabilité d'extinction est de 4.1 %, puis elle passe à 8.7 % à la troisième et 13 % à la quatrième. Au bout de dix générations elle atteint 32 % mais ce point n'a pas été représenté sur la figure 3 pour ne pas le surcharger.

On donne de même sur la figure 4 la distribution des descendances pour vingt individus portant le même nom. En cinq générations la probabilité d'extinction est de 1 pour mille. Celle-ci passe à 1 % en 10 générations, puis à 2.4 % en 15 et 3.6 en vingt générations.

De même nous avons donné les distributions successives pour un nom porté par 150 individus[2]. On peut ainsi voir sur la figure 5 l'effet d'une population légèrement croissante, puisque l'ensemble de la courbe se déplace vers la droite au fur et à mesure des générations, mais son écart-type croît lui aussi. La probabilité d'extinction toujours négligeable va croître de 10^{-31} à 10^{-21} en ces trois générations.

Comme nous avons étudié les mécanismes d'extinction des noms et pu les quantifier à l'horizon de quelques générations dans un cas simple et avec les paramètres de la démographie actuelle, nous pouvons nous intéresser maintenant à des statistiques sur l'ensemble des noms portés par les Français. Nous nous intéresserons dans cette seconde partie à un échantillon particulièrement important que nous commencerons par décrire. Cet échantillon unique et le plus important à notre connaissance, permettra d'étudier deux sous-problèmes du problème général de l'extinction des noms, problème général que nous pourrons mieux cerner sans pouvoir le résoudre.

[2]Pour un nombre élevé d'individus et de générations le calcul informatique devient compliqué. Pour un nombre intermédiaire, on ne doit mettre en mémoire qu'une sous-matrice de la matrice de passage de la chaîne de Markov associée et utiliser évidemment l'approximation par la loi normale. Les calculs exécutés en double précision exigent un temps d'ordinateur non négligeable.

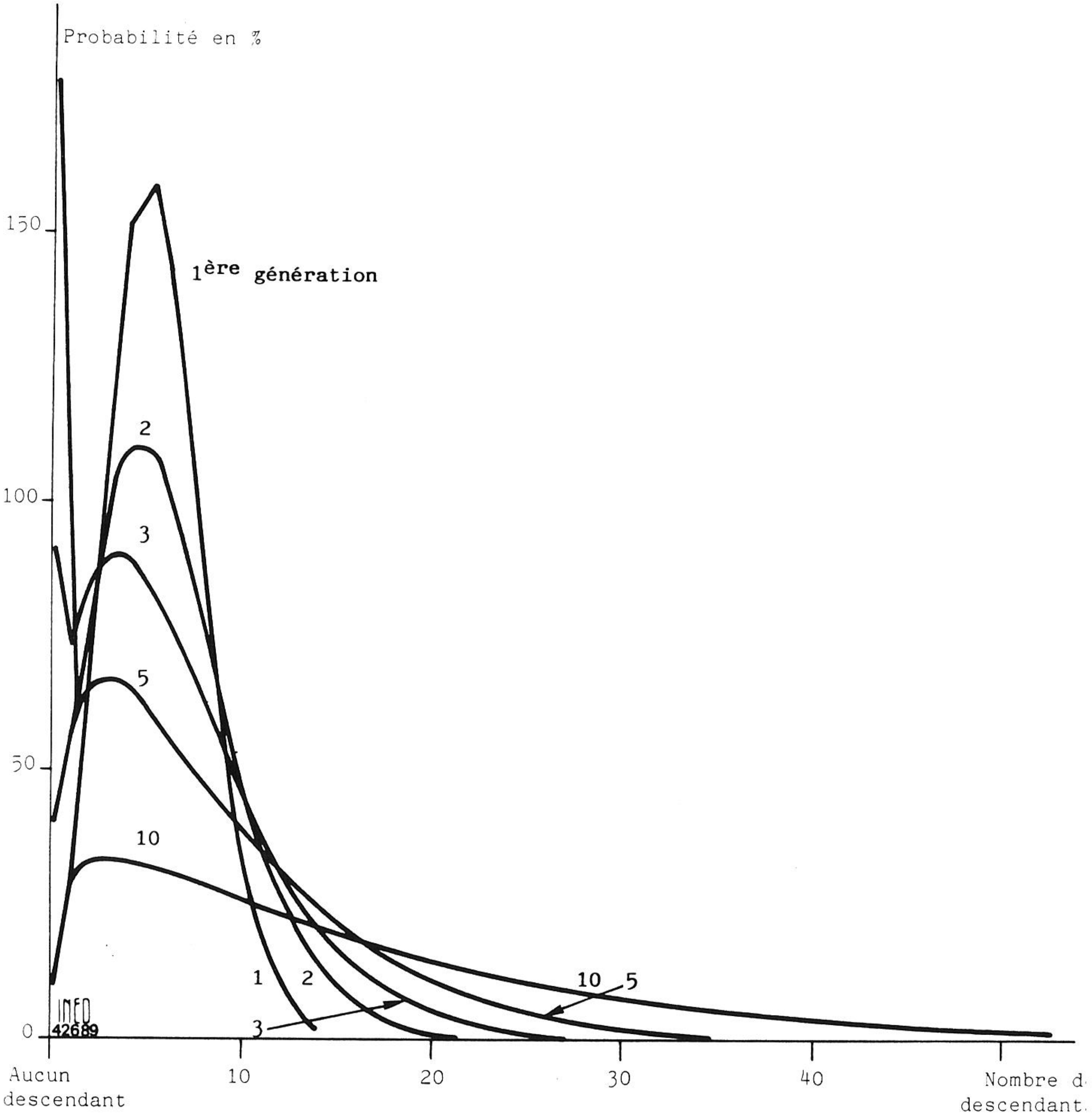
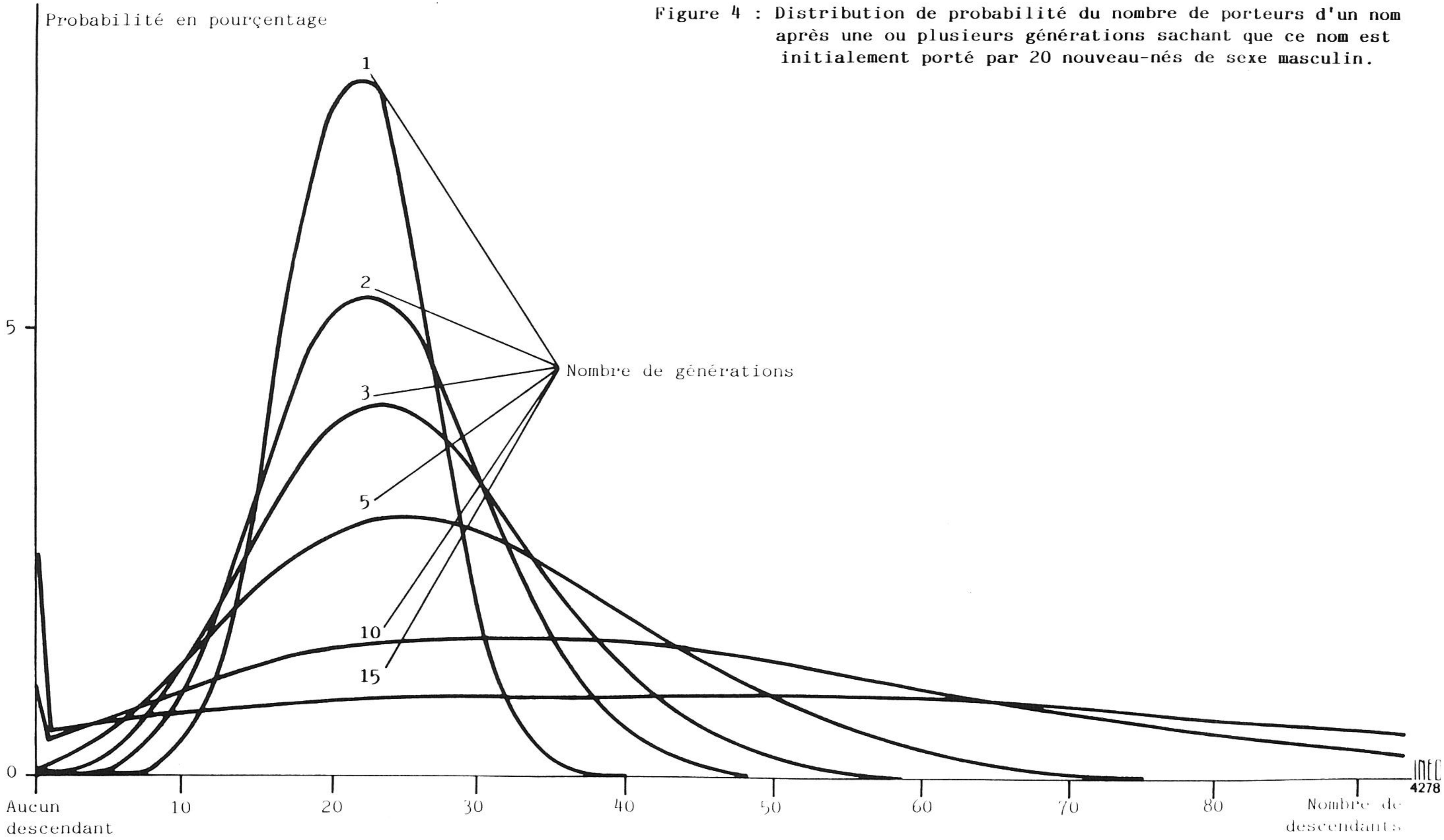


Figure 3 : Distribution de probabilité du nombre d'hommes descendants portant un même nom après 1, 2, 3, 5, ou 10 générations, sachant qu'initialement ce nom était porté par 5 hommes.

Figure 4 : Distribution de probabilité du nombre de porteurs d'un nom après une ou plusieurs générations sachant que ce nom est initialement porté par 20 nouveau-nés de sexe masculin.



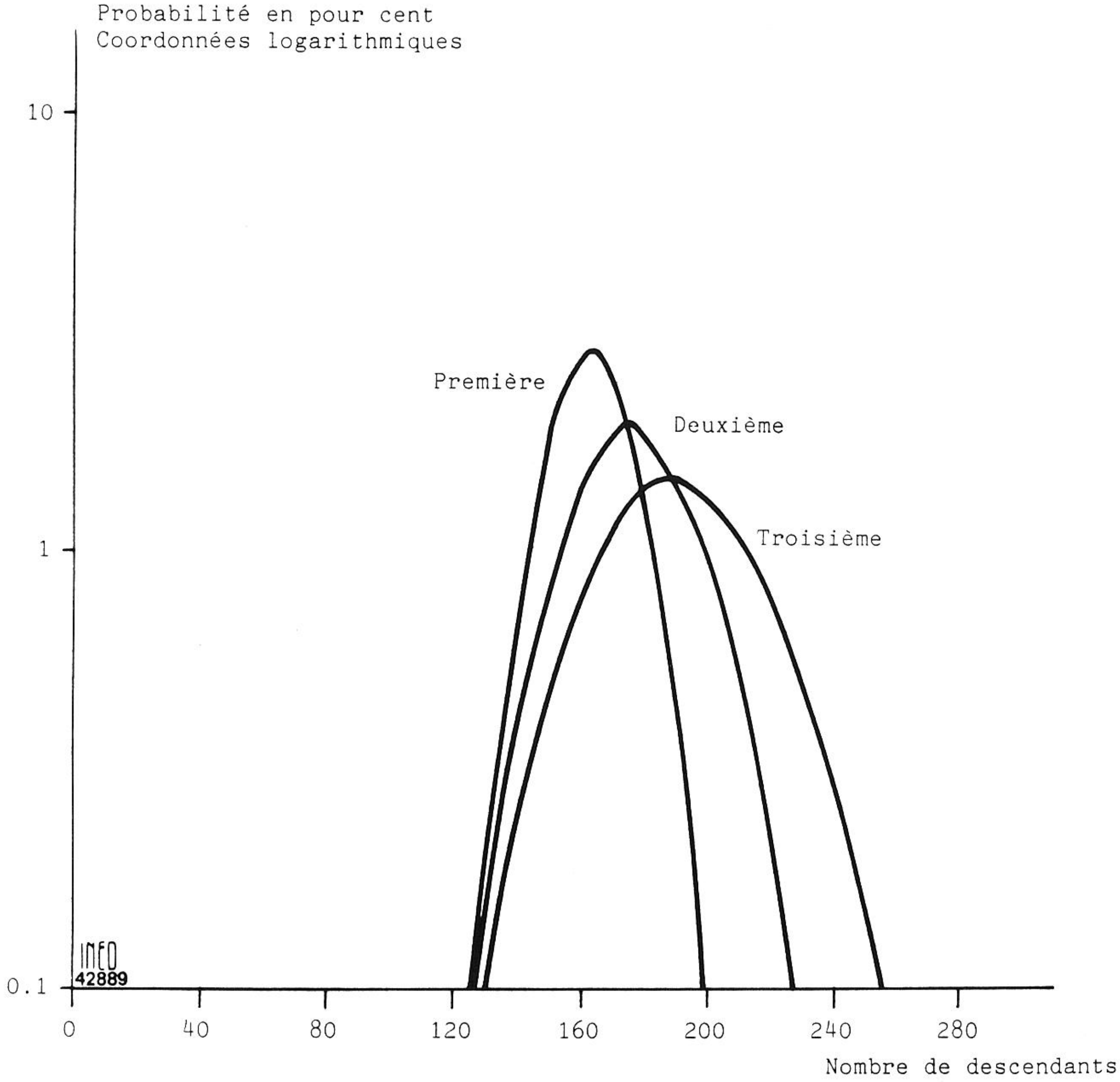


Figure 5 : Distribution de probabilité du nombre de descendants (de sexe masculin) de 150 hommes après 1, 2, puis 3 générations.

6 Essai d'évaluation du nombre des noms de famille et de son évolution

6.1 Description d'un échantillon important

Une source importante de renseignements d'ordre statistique sur la fréquence des noms nous est fournit par un échantillon de l'ensemble des naissances de l'année 1950 dans 56 départements français, soit 413 000 naissances. Cet échantillon devait permettre aux statisticiens de l'INSEE de mesurer la longueur moyenne des noms en vue d'un codage informatique. Un article de la revue Economie et Statistique[3] illustre ces renseignements et nous donne comme sous-produit la fréquence des noms les plus fréquents. Un dépouillement plus approfondi du listing de l'ensemble de ces noms, effectué par F.X. Deniau[4] fournit la distribution du nombre n_i des noms en fonction du nombre i des porteurs. Les chiffres figurent en annexe, mais on a représenté cette distribution en coordonnées doublement logarithmiques sur le graphique 6. Nous avons ainsi 68 365 monohomonymes sur un total de 116 400 noms. Nous remarquons surtout un alignement remarquable des différents noms montrant ainsi une dépendance qu'on retrouve dans beaucoup de domaines des sciences humaines comme la linguistique et la géographie sans n'avoir jamais bien pu l'expliquer. Ce sont des lois qui portent le nom de Pareto, Zipf, Mandelbrot, etc.

Remarquons pour se donner une idée du caractère néanmoins particulier de cette représentation, que les points 68 365 et 116 400, c'est-à-dire le nombre de monohomonymes et le nombre total de noms sont relativement proches sur le dessin. Un ajustement, par les moindres carrés par exemple, donnerait ainsi une fourchette très importante sur le nombre total de noms portés.

Pour compléter la description, nous avons tracé la courbe (graphique 7) croisant les deux fonctions cumulatives du nombre d'enfants en abscisse et du nombre de noms en ordonnée. Ainsi 50 % des enfants ont un nom porté à moins de 7 exemplaires, ce qui représente 97 % des noms. Pour une description plus détaillée de cette distribution, on peut se reporter à la liste des 2500 noms les plus fréquents du mémoire de F.X. Deniau déjà cité, mais pour le moment voyons si cet échantillon peut nous être utile pour le problème global que nous nous posons. Nous ne saurons pas traiter le

[3] "Monsieur Dupont s'appelle Martin et son prénom est Jean"

[4] "Les noms de famille en France, essai de dénombrement" F.X. Deniau. Mémoire de Démographie, ENSAE Juin 1977.

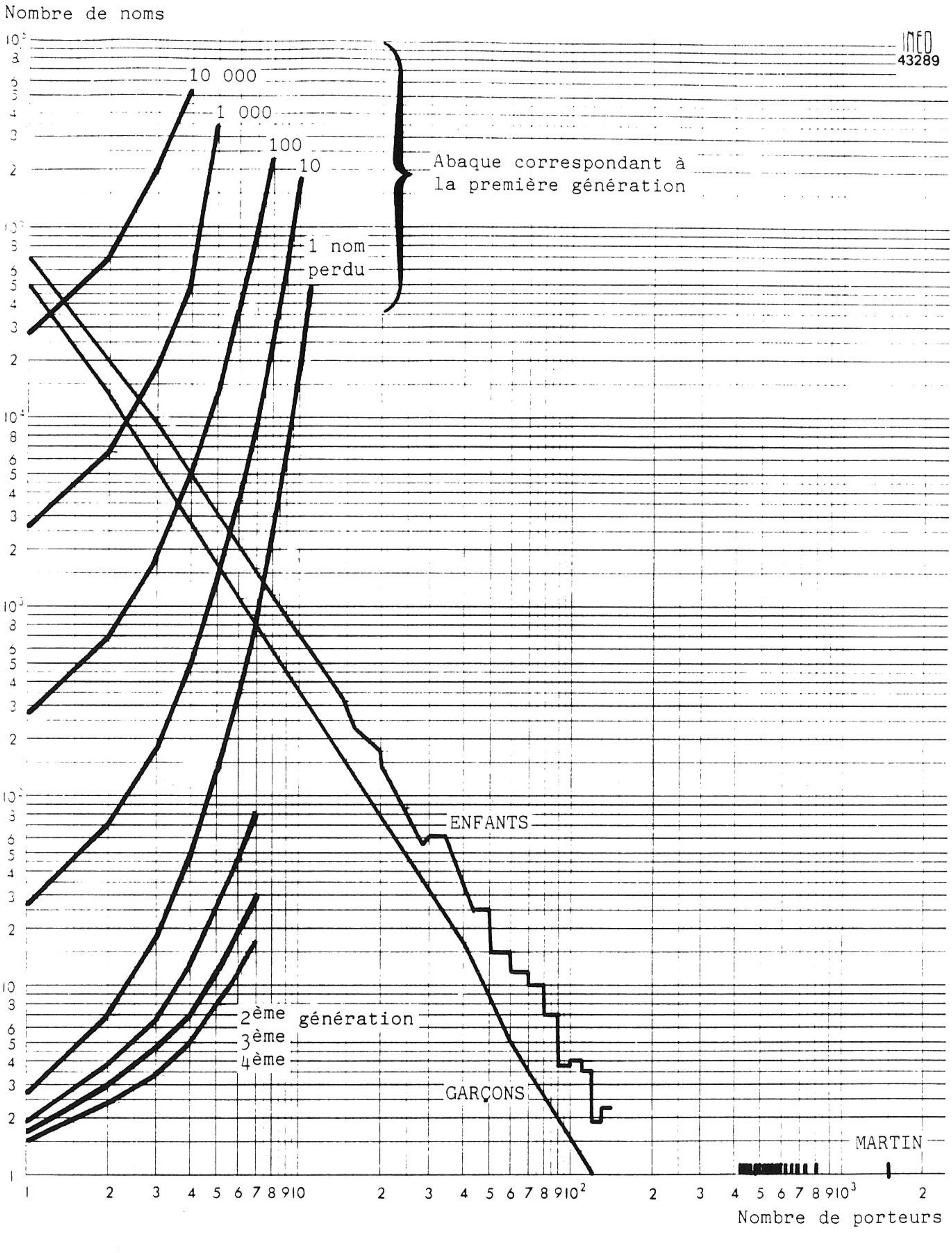


Figure 6 : Distribution du nombre des noms en fonction du nombre de porteurs. Echantillon des 417 000 enfants nés en 1950 dans 56 départements.
- Echantillon des seuls garçons correspondants.
- Abaque donnant le nombre de noms perdus.

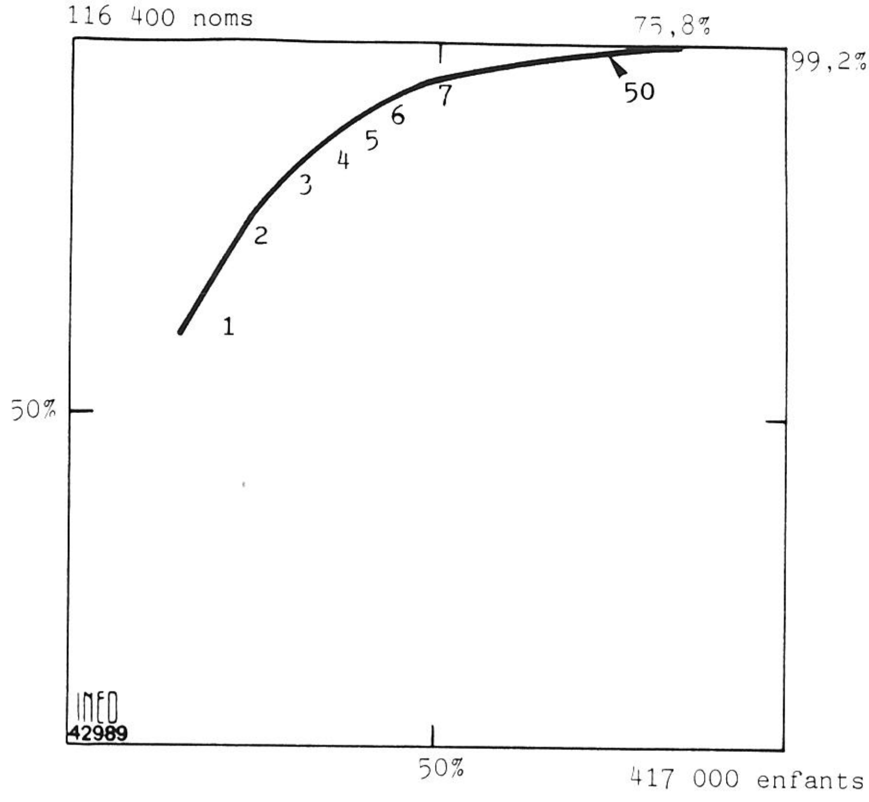


Figure 7 : Fonction cummulative du nombre des noms croisée avec la distribution du nombre des enfants.

problème global, mais nous pourrons exposer et répondre aux deux sous-problèmes suivants :

1. Le premier est la suite logique de la première partie sur le processus de Galton-Watson. Supposons que les 413 000$\frac{105}{205}$ = 211 530 garçons migrent seuls sans parents dans une région inhabitée et qu'ils vivent là avec les conditions de reproduction observées actuellement en France, alors nous pouvons étudier le nombre de noms perdus au cours des générations et les variations de la distribution des fréquences.

2. Le second problème consiste à supposer que l'échantillon des 413 000 noms est un sondage (environ au 1/100) de la population des 41,944 millions de Français en 1950, et d'élaborer un modèle théorique permettant de relier la distribution observée sur l'échantillon à la distribution réelle sur l'ensemble de la France.

7 Un échantillon important mais qui n'apporte qu'une information maigre sur l'ensemble de la population.

Notre échantillon est constitué des enfants nés en 1950 dans 56 départements. Un échantillon sur les 90 départements aurait été bien plus intéressant pour éliminer un peu la composante régionale de la répartition des noms, mais il n'existe pas pour le moment. De toute façon on comprend aisément pourquoi cette distribution n'a pas un énorme intérêt pour le problème général complexe que nous nous posons. En effet, les 68 365 monohomonymes décomptés sont au moins des bihomonymes dans la population totale, car chaque nouveau-né a obligatoirement un père vivant. Si nous ne tenons pas compte des naissances gémellaires, nous obtenons facilement un échantillon de taille double pour le même nombre de noms. On pourrait même calculer la probabilité pour qu'un nouveau-né ait un grand-père paternel vivant,voire un bisaïeul. Notre échantillon d'individus pourrait ainsi peut-être tripler, mais nous n'irions pas plus loin car il nous faudrait un modèle de famille pour connaître les frères et cousins collatéraux. Or ces modèles sont très complexes comme on peut l'imaginer, et n'existent pas. Des essais de microsimulations sur des petites populations, coûteux en temps de calculs et en hypothèses et paramètres, ont vu le jour mais n'ont pas abouti dans un but opérationnel.

8 Evolution du patrimoine des noms de la génération des garçons nés en 1950

Nous nous sommes donc limités au premier problème exposé plus haut. Pour cela nous avons dû raisonner sur la population masculine et donc dû calculer la distribution des noms en fonction du nombre de porteurs, en ne tenant compte que des garçons.

8.1 Passage de la distribution des enfants à celle des garçons

Les 68 365 noms d'enfants monohomonymes se répartiront en moyenne en 68 365$\frac{105}{205}$ noms monohomonymes portés par des garçons et 33 340 noms monohomonymes portés par des filles. Pour un nom bihomonyme, les probabilités de trouver 2 garçons, c'est à dire un nom de garçon bihomonyme est g^2 ou $g = \frac{105}{205}$; celle de trouver 1 garçon 1 fille soit un nom de garçon monohomonyme est $2g(1-g)$; celle de n'avoir aucun nom bihomonyme d'enfants porté par des garçons est $(1-g)^2$. Ainsi sur 19 044 noms d'enfants bihomonymes, le nombre de nom de chaque sorte est donné par la loi multinomiale. Mais on peut remplacer ces nombres aléatoires par leurs moyennes car les erreurs relatives sont faibles.

Par contre si 20 noms sont portés chacun par 50 enfants, la répartition des 20 noms entre les 51 différents cas possibles (aucun garçon, 1 garçon, ldots, 50 garçons) est plus aléatoire[5].

Sans connaissance de la distribution empirique exacte du nombre de noms suivant le nombre de porteurs mâles, on peut estimer cette distribution par l'espérance g^p du nombre aléatoire de noms pour un nombre de porteur p fixé, par la formule $g_p = \sum_{l=p}^{L} C_1^p g^p (1-g)^{1-p} n_1$, où n_1 est la distribution correspondant aux enfants en général.

On a représenté cette nouvelle distribution sur le même graphique 6, elle correspond à un effectif moitié, soit 417 000$\frac{105}{205}$ = 213 00 garçons, mais à un patrimoine de 77 400 noms sur 116 400 originaux.

8.2 Nombre de noms perdus et évolution de la distribution suivant les générations

Supposons que ces 213 000 garçons nés en 1950 dans 56 départements de France constituent une population à part ; ceux-ci représentent alors un patrimoine de 77 400 noms répartis suivant le nombre de porteurs comme l'indique la figure 6. Il nous est maintenant facile compte tenu des résultats de la première partie d'étudier son évolution au cours des générations.

Les calculs montrent que nous perdrions ainsi 20 700 noms en l'espace d'une génération (28 ans), 30 800 noms seraient perdus en l'espace de deux générations, 37 000 soit près de la moitié du patrimoine en l'espace de trois

---

5

- La probabilité d'avoir p garçons dans un lot de n enfants est $C_n^p g^p (1-g)^{n-p} = X_n^p$:
- si q noms sont portés chacun par n enfants, la probabilité d'une répartition de ces q noms entre les $n+1$ configurations possibles, 0 garçon, 1 garçon,..., n garçons est donnée par la loi multinomiale :

$$\Pr(Y_o = a_o, \ldots, Y_n = a_n) = \frac{q!}{a_0 \ldots a_n!}(X_n)^{a_0} \ldots (X_n^q)^{a_n},$$

avec $a_0 + \ldots + a_n = q$.

(85 ans).

On représenté sur le graphique 8 l'évolution du nombre des noms au cours des 10 générations suivantes. L'évolution de la population elle-même que nous avions choisi croissante est aussi représentée à une échelle quatre fois moindre. Après une génération environ, un nom serait en moyenne porté par 4 hommes contre 1 sur 2,75 au départ.

Comme nous pouvons le voir aussi sur le tableau 3, la perte des noms

Tableau 3: NOMBRE DE NOMS PERDUS AU COURS DES GÉNÉRATIONS.

Génération	Nombre de noms	Noms perdus entre 2 génératons	Population
0	77 400	20 729	213 500
1	56 673	10 137	230 580
2	46 536	6 200	249 026
3	40 336	4 231	268 948
4	36 105	3 086	290 464
5	33 019	2 354	313 701
6	30 665	1 855	338 797
7	28 810	1 498	365 173
8	27 312	1 233	395 173
9	26 079	1 029	426 787

est très rapide au cours des premières générations puis s'amoindrit lors des générations suivantes.

En effet, si nous regardons sur la figure 9 comment la distribution évolue, nous nous apercevons que la droite que nous avions observée au départ n'existe plus dès la première génération.

- Le nombre des noms rares (noms portés par moins de 10 personnes) a fortement diminué. Cette diminution des noms rares s'effectue, non pas comme on le dit par abus de langage au profit des noms plus fréquents dont ils sont totalement indépendants, mais une partie des noms rares va disparaître tandis que l'autre partie va tendre à être portée par plus d'individus et faire partie des noms plus fréquents.

- La population des Martin doit en moyenne croître au taux de croissance mais son évolution est aléatoire aussi.

- Sur l'ensemble des différentes populations que constituent les différents noms de fréquence moyenne, certaines peuvent croître très vite par le

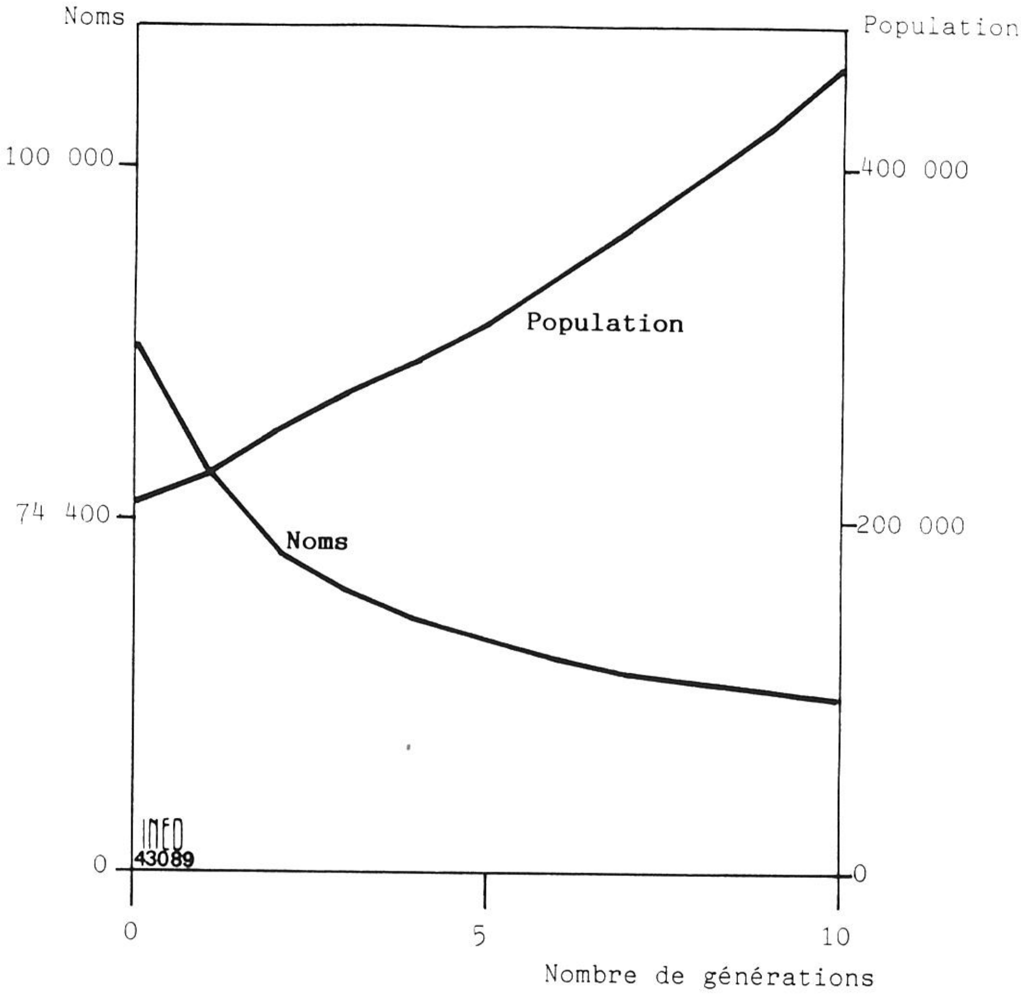


Figure 8 : Evolution du nombre des 77 400 noms de 213 000 garçons nés en 1950 dans 56 départements et évolution de leur population

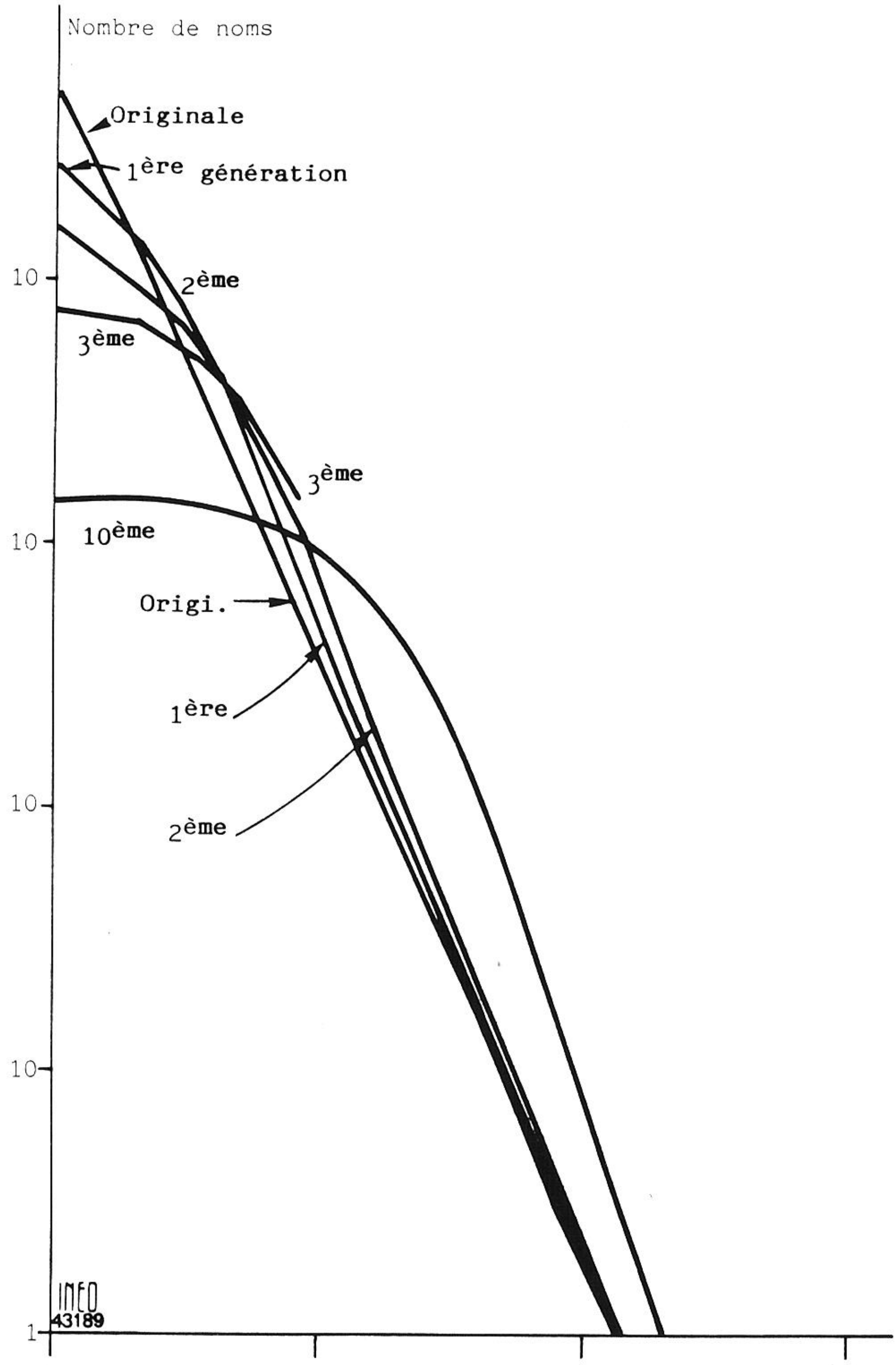


Figure 9 : Evolution de la distribution du nombre des noms en fonction du nombre des porteurs après 1, 2, 3, et 10 générations. Calculs établis à partir des 213 000 garçons nés en 1950.

seul fait du hasard, et d'autres décroître, alors qu'en moyenne la population totale constituée par l'ensemble de ces populations va croître au taux de croissance [6].

8.3 Des abaques de noms perdus

Nous avons aussi représenté sur le graphique des abaques permettant de donner rapidement un ordre de grandeur du nombre de noms perdus compte tenu de la distribution du nombre des noms en fonction du nombre des porteurs.

L'abaque "1 nom perdu" coupe ainsi la distribution relative aux garçons, au voisinage du 7ième porteur : ainsi sur les 800 noms portés chacun par 7 garçons, 1 nom sera perdu. Au delà de 7 porteurs la perte est donc encore plus négligeable.

Plus de 10 000 noms vont être perdus parmi les monohomonymes (18 600), entre 10 000 et 1000 pour les bihomonymes, 1000 et 100 pour les trihomonymes, entre 100 et 10 pour les noms portés par 4 personnes, et près de 10 seront perdus sur les 1700 noms portés chacun par 5 garçons.

Ces abaques sont très générales, elles permettent de calculer le nombre de noms perdus pour une courbe quelconque, en particulier pour la courbe correspondant à la France entière si nous la connaissions. Elle permet surtout de comprendre que pour évaluer le nombre de noms perdus à la centaine près, il suffit de connaître la distribution des noms rares jusqu'au 6 ou 7^{e} porteur, et à la dizaine près au maximum au 10^{e} porteur.

On a aussi représenté les abaques "1 nom perdu " qui correspondent à un horizon de deux, trois et quatre générations. Ainsi pour connaître le nombre de noms perdus à la centaine près au bout de quatre générations ¡footnote Il suffit de translater l'abaque "1 nom perdu à la 4^{e} génération" de deux unités correspondant à une multiplication par 100 en coordonnées logarithmiques., il faut connaître la distribution jusqu'au septième porteur environ ce qui représente, dans notre échantillon 50 % de la population et pèrs de 90 % de l'ensemble des noms.

[6] Plus exactement cette population somme, plus importante aura des fluctuations relatives plus faibles

8.4 Pourquoi cette loi de Pareto observée n'est-elle pas stable ?

Il est en effet tout à fait étrange que la distribution empirique observée s'ajuste si bien sur une loi de Pareto, alors que le processus d'évolution décrit par le modèle détruit cette stabilité[7].

On peut invoquer différentes raisons à ce fait :

- tout d'abord, l'apport en une génération de noms nouveaux, donc rares, peut contrebalancer exactement les pertes. Cette loi de Pareto serait alors justifiée ; et on peut alors penser à différents modes de création : noms étrangers, francisation de nom et déformation par l'écriture des noms existants.
- en second lieu, la courbe inconnue qui correspond à l'ensemble de la population française est peut-être elle-même incurvée vers le bas pour les noms rares, et montre ainsi déjà le processus d'appauvrissement.

Il nous est impossible de répondre au premier point. Dans le cas du second point, la perte des noms serait moins importante que dans une vraie droite de Pareto.

[7] Remarquons néanmoins que les points qui correspondent aux noms très rares (1 à 2 porteurs) se trouvent un peu au-dessous de la droite théorique.

Dossiers et recherches*

N° 1.- Georges TAPINOS, *Les méthodes d'analyse en démographie économique*, 1976, 288 p.

N° 2.- Claude LEVY, *Aspects socio-politiques et démographiques de la planification familiale en France, en Hongrie et en Roumanie*, 1977, 248 p.

N° 3.- Paul PAILLAT, *Le vécu du vieillissement en 1979*, 1981, 114 p.

N° 4.- Graziella CASELLI, Jacques VALLIN, J. VAUPEL et A. YASHIN, *L'évolution de la structure par âge de la mortalité en Italie et en France depuis 1900*, 1986, 28 p.

N° 5.- Jacques VALLIN et France MESLE, *Les causes de décès en France de 1925 à 1978*, 1986, 36 p.

N° 6.- Philippe FARGUES, *Un apport potentiel des formations sanitaires pour mesurer la mortalité dans l'enfance en Afrique*, 1986, 34 p.

N° 7.- Jacques VALLIN, France MESLE et Alfred NIZARD, *Reclassement des rubriques de la 8ème révision de la Classification internationale des maladies selon l'étiologie et l'anatomie*, 1986, 56 p.

N° 8.- Didier BLANCHET, *Equilibre malthusien et liaison entre croissances économique et démographique dans les pays en développement : un modèle*, 1986, 20 p.

N° 9.- Didier BLANCHET, *Deux études sur les relations entre démographie et systèmes de retraite*, 1986, 26 p.

N° 10.- Philippe FARGUES, *La migration obéit-elle à la conjoncture pétrolière dans le Golfe ? L'exemple du Koweït*, 1987, 30 p.

N° 11.- Gilles PISON, *Les jumeaux en Afrique au Sud du Sahara : fréquence, statut social et mortalité*, 1987, 48 p.

N° 12.- Philippe FARGUES, *Les saisons et la mortalité urbaine en Afrique. Les décès à Bamako de 1974 à 1985*, 1987, 38 p.

N° 13.- Kuakuvi GBENYON et Thérèse LOCOH, *Différences de mortalité selon le sexe, dans l'enfance en Afrique au Sud du Sahara*, 1987, 30 p.

N° 14.- Jacques VALLIN, *Théorie(s) de la baisse de la mortalité et situation africaine*, 1987, 44 p.

---

* Ces documents sont disponibles chez l'auteur.

N° 15.- Peter AABY, *Le surpeuplement, un facteur déterminant de la mortalité par rougeole en Afrique*, 1987, 52 p.

N° 16.- Gérard CALOT et Graziella CASELLI, *La mortalité en Chine d'après le recensement de 1982.*
I .- *Analyse selon le sexe et l'âge au niveau national et provincial*, 1988, 72 p.
II.- *Tables de mortalité par province*, 1988, 112 p.

N° 17.- Jacques VALLIN, *Evolution sociale et baisse de la mortalité : conquête ou reconquête d'un avantage féminin ?* 1988, 36 p.

N° 18.- Jacques VALLIN, *La mortalité en Europe de 1720 à 1914 : tendances à long terme et changements de structure par âge et par sexe*, 1988, 40 p.

N° 19.- Henri LERIDON, *Analyse des biographies matrimoniales dans l'enquête sur les situations familiales*, 1988, 64 p.

N° 20.- France MESLE, *Morbidité et causes de décès chez les personnes âgées*, 1988, 18 p.

N° 21.- Noël BONNEUIL et Philippe FARGUES, *Prévoir les "caprices" de la mortalité. Chronique des causes de décès à Bamako de 1964 à 1985*, 1989, 44 p.

N° 22.- Benoit RIANDEY, *Un échantillon probabiliste de A à Z : l'exemple de l'enquête Peuplement et dépeuplement de Paris. INED (1986)*, 1989.

N° 23.- Georges TAPINOS, Didier BLANCHET et Olivia EKERT-JAFFE, *Population et demande : changements démographiques, demande et structure de consommation*, 1989, 46 p.

N° 24.- Graziella CASELLI et Jacques VALLIN, *Mortalité et vieillissement de la population*, 1989, 30 p.

N° 25.- Alain BLUM et Philippe FARGUES, *Estimation de la mortalité maternelle dans les pays à données incomplètes. Une application à Bamako (1974-1985) et à d'autres pays en développement*, 1989, 36 p.

N° 26.- Gilles PISON, Monique LEFEBVRE, Catherine ENEL et Jean-François TRAPE *L'influence des changements sanitaires sur l'évolution de la mortalité. Le cas de Mlomp (Sénégal) depuis 50 ans*, 1989, 36 p.

N° 27.- Nicolas BROUARD, *L'extinction des noms de famille en France: une approche*, 1989, 22 p.