JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

EDMOND MALINVAUD

Relations entre la composition des familles et le taux de masculinité

Journal de la société statistique de Paris, tome 96 (1955), p. 49-64

<http://www.numdam.org/item?id=JSFS_1955__96__49_0>

IX

RELATIONS ENTRE LA COMPOSITION DES FAMILLES ET LE TAUX DE MASCULINITÉ

Chacun d'entre nous connaît sans doute des mères qui n'ont eu que des garçons ou que des filles malgré un nombre assez respectable de maternités. On entend alors souvent dire que ces femmes ne *pouvaient* avoir que des garçons ou que des filles. Les personnes habituées au raisonnement statistique formulent une question plus nuancée mais assez voisine : « Si j'ai déjà eu cinq filles et pas de garçons et que j'attende un nouveau bébé, ai-je plus de chances que ce soit une fille qu'un garçon? » Autrement dit, le sexe d'un enfant à naître dépend-il du sexe des enfants déjà nés des mêmes parents? Dans cet article, vont être présentés quelques résultats statistiques nouveaux qui permettent de répondre, dans une certaine mesure, aux questions précédentes. Une brève analyse interprétative en sera fournie de façon à dégager quelques conclusions intéressantes pour les recherches futures.

Phénomène essentiellement biologique, la détermination du sexe à la naissance recevra son explication dernière de la biologie. Pour le moment, les éléments fragmentaires dont dispose cette science, tendent à montrer que la formation du sexe semble influencée par de nombreux facteurs. Il est cependant encore impossible de donner du phénomène une analyse quantitative suffisamment précise pour répondre valablement aux questions soulevées ci-dessus. Dans ces conditions, l'observation statistique doit permettre de dégager les caractéristiques essentielles de la question. Judicieusement interprétée, elle doit apporter au biologiste des informations fondamentales pour sa recherche. Il est alors nécessaire que le statisticien synthétise les données empiriques et les présente sous la forme de quelques schémas mathématiques simples qui ne laisseraient subsister que les grandes lignes du phénomène. Tel fut d'ailleurs l'esprit d'une communication présentée en 1950 à la Société de Statistique de Paris par M. Schützenberger qui s'était attaché à étudier la succession des sexes dans les familles nombreuses (1).

Les données présentées ici résultent du dépouillement systématique des statistiques françaises d'état civil de 1946 à 1950. Chaque bulletin de naissance porte, en effet, mention du sexe du nouveau-né, et aussi du sexe de tous les enfants déjà nés de la même mère. Il a donc été possible d'étudier la liaison entre ces deux renseignements. La documentation obtenue porte sur près de quatre millions de naissances et semble bien être unique au monde, de sorte que son intérêt doit largement dépasser le cadre de notre pays. L'initiative de la collecte et du dépouillement de ces données revient à M. Gasc, administrateur à l'I. N. S. E. E., qui a bien voulu mettre à notre disposition ses résultats et permettre ainsi une étude qui n'aurait jamais vu le jour sans sa diligence (2).

Naissances suivant le sexe des enfants déjà nés

Les résultats obtenus pour la période 1946 à 1950 peuvent être présentés dans un tableau à double entrée, portant en colonnes le nombre de garçons déjà nés et en lignes le nombre de filles déjà nées. Ainsi, dans le tableau I, la case située en haut à gauche intéresse les premières naissances, tandis que la case située immédiatement à droite de la précédente correspond aux naissances provenant de mères qui ont déjà eu un seul garçon et pas de fille. Afin de supprimer toute difficulté d'interprétation, le dépouillement a été limité aux naissances légitimes et aux cas dans lesquels la mère n'avait contracté qu'un seul mariage et n'avait pas eu d'enfant avant ce mariage. Ainsi, chaque case du tableau I correspond, à proprement parler, à un type de famille. Il faut encore préciser que le nombre de garçons et de filles déjà nés pris en considération pour la détermination du type de famille comprend toutes les naissances antérieures d'enfants vivants ou mort-nés. Dans le cas de naissances multiples, de deux garçons par exemple, on a donné un ordre arbitraire à ces naissances et on a déterminé le type de famille avant la naissance du second garçon en tenant compte de son frère jumeau. Ainsi deux jumeaux se trouvent-ils enregistrés dans deux cases voisines du tableau I et non dans la même case.

Les tableaux I et II ci-dessous sont relatifs respectivement aux naissances vivantes et aux mort-nés. Dans chaque case, le nombre supérieur indique l'effectif en milliers des naissances observées pour le type de famille correspondant, tandis que le nombre inférieur mesure le taux de masculinité observé parmi ces naissances et exprimé en nombre moyen de garçons pour 100 naissances. Comme il apparaît clairement, l'effectif des naissances par type de famille décroît très rapidement lorsqu'on va de la gauche vers la droite, ou du haut en bas, dans les tableaux I et II. Ainsi les taux de masculinité calculés

---

(1) M. P. Schützenberger, « Résultats d'une enquête sur la distribution du sexe dans les familles nombreuses. » (*La Semaine des Hôpitaux de Paris*, 1949, n° 61) ; «Nouvelles recherches sur la distribution du sexe à la naissance. » (*Ibidem*, 1950, n° 86).

(2) Qu'il nous soit permis de remercier également nos camarades Croze, Desabie et Febvay, administrateurs à l'I. N. S. E. E., qui ont suivi ce travail au fur et à mesure de son avancement, ainsi que le Dr Sutter de l'I. N. E. D. qui a mis à notre disposition sa bibliographie sur le sujet.

seront-ils très précis pour les types de familles comportant peu d'enfants, mais assez incertains pour ceux qui comprennent plus de quatre garçons ou plus de quatre filles. Dans l'examen des taux de masculinité, il faudra donc retenir surtout la partie située à gauche et en haut du tableau I.

Quoi qu'il en soit, une tendance très nette se dégage : le taux de masculinité, pour les naissances vivantes comme pour les mort-nés, est d'autant plus élevé que la famille comprenait déjà à la naissance un plus grand nombre de garçons et un moins grand nombre de filles. Parmi les taux observés pour les naissances vivantes, le plus fort (54,5) correspond aux familles comprenant six garçons et une fille, le plus failbe (47,9) est obtenu pour les familles contenant six filles et deux garçons. Sur quarante-six taux calculés, treize sont inférieurs à 50, correspondant tous à des familles comprenant plus de filles que de garçons; douze sont supérieurs à 52, correspondant tous à des familles comprenant plus de garçons que de filles. Toutefois les écarts constatés restent relativement faibles puisque tous les taux sont pratiquement compris entre 48 et 55. La faible ampleur de ces variations exige que l'on examine avec soin les fluctuations aléatoires des taux observés ainsi que les particularités du dépouillement susceptibles d'avoir affecté les résultats obtenus.

TABLEAU I

Naissances vivantes et leurs taux de masculinité suivant le nombre des enfants déjà nés des mêmes parents (1)

(France 1946 à 1950)

NOMBRE DE FILLES déjà nées	NOMBRE DE GARÇONS DÉJA NÉS						
	0	1	2	3	4	5	6
0	1.499 51,44	552 51,87	163 52,7	45 52,5	13 54,4	4 53,5	2 52,1
1	506 50,88	290 51,16	114 51,4	43 52,4	16 52,5	7 53,3	3 54,5
2	142 50,2	107 50,5	60 50,7	30 51,3	15 51,9	7 52,4	3 51,8
3	37 49,3	38 49,3	29 50,3	19 51,5	11 50,7	6 51,0	3 52,7
4	10 48,3	14 49,5	13 49,7	11 50,2	8 50,7	5 51,6	3 52,1
5	3 48,7	6 49,0	6 49,7	6 48,7	5 51,2	3 51,4	
6	1 48,4	2 48,2	3 47,9	3 50,1	2 48,6		

(1) Il a été tenu compte uniquement des naissances légitimes. On s'est de plus limité aux cas dans lesquels la mère n'avait contracte qu'un seul mariage et n'avait pas eu d'enfants avant ce mariage. La composition de la famille a eté définie à partir de toutes les naissances antérieures, y compris les mort-nés.

Dans chaque cas, le nombre supérieur indique l'effectif des naissances en milliers, le nombre inférieur mesure le taux de masculinite exprımé en nombre de garçons pour cent naissances.

TABLEAU II

Mort-nés et leurs taux de masculinité
suivant le nombre des enfants déjà nés des mêmes parents (1)
(France 1946 à 1950)

NOMBRE DE FILLES déjà nées	NOMBRE DE GARÇONS DÉJA NÉS		
	0	1	2
0	35 57,7	8 58,7	3 58,6
1	7 55.7	5 58,3	2 58,0
2	2 54,8	2 55,6	1 57,7

Précision des taux de masculinité obtenus

Pour mesurer les fluctuations aléatoires affectant les taux moyens observés, il faut estimer les écarts-types de ces taux. Deux méthodes ont été retenues ici. Leurs résultats figurent simultanément dans le tableau III ci-dessous. Avant d'examiner ce tableau, indiquons brièvement comment les écarts-types ont été calculés.

On peut d'abord utiliser les résultats de calculs portant séparément sur chacune des années 1946 à 1950 et, pour chaque type de famille, estimer la dispersion des taux de masculinité annuels d'après la dispersion des cinq valeurs obtenues pour ce taux. Si p_{it} est le taux moyen observé pour les familles de composition i pendant l'année t, la variance de p_{it} peut être estimée par :

$$\text{var}\ (p_{it}) = \frac{1}{4} \sum_{t=1946}^{1950} (p_{it} - \bar{p}_i)^2$$

avec

$$\bar{p}_i = \frac{1}{5} \sum_{t=1946}^{1950} p_{it}$$

puisque les nombres de naissances ont été assez voisins pour les différentes années.

L'écart-type du taux moyen p_i obtenu sur l'ensemble des cinq années sera alors voisin de :

$$\sigma\ (p_i) = \sqrt{\frac{1}{5}\ \text{var}\ (p_{it})}$$

C'est par ce procédé qu'a été estimé le double de l'écart-type σ (p_i) tel qu'il est indiqué par le nombre supérieur de la case correspondante du tableau tableau III.

(1) Cf. note 1 du tableau I.

TABLEAU III

Double de l'écart-type affectant
le taux de masculinité présenté dans le tableau I (1)

NOMBRE DE FILLES déjà nées	NOMBRE DE GARÇONS DÉJÀ NÉS						
	0	1	2	3	4	5	6
0	0,12 0,08	0,45 0,13	0,4 0,3	0,5 0,5	0,5 0,9	1,7 1,5	1,8 2,5
1	0,28 0,14	0,21 0,19	0,3 0,3	0,6 0,5	1,0 0,8	1,2 1,2	1,8 1,9
2	0,5 0,3	0,3 0,3	0,6 0,4	0,4 0,6	1,1 0,8	1,3 1,2	1,6 1,5
3	0,8 0,5	0,5 0,5	0,6 0,6	0,3 0,7	0,5 0,9	1,0 1,2	0,6 1,6
4	1,1 1,0	0,9 0,8	1,2 0,8	1,2 1,0	1,0 1,1	1,5 1,4	2,2 2,0
5	1,4 1,8	0,6 1,3	1,3 1,2	1,4 1,2	1,8 1,4	1,5 1,8	
6	2,0 3,1	1,9 2,2	1,4 2,0	1,3 1,8	2,4 2,1		

Cependant, cette estimation précédente de l'écart-type $\sigma\ (p_i)$ reste très imprécise en raison du petit nombre d'observations sur lesquelles elle repose. En fait, si σ_i est la valeur vraie mesurée par $\sigma\ (p_i)$, $4\ \frac{\sigma^2\ (p_i)}{\sigma_i^2}$ suit une loi de χ^2 à quatre degrés de liberté. On en déduit qu'il y a encore environ 2,5 chances sur cent pour que :

$$\sigma\ (p_i) \leqq 0{,}35\ \sigma_i$$

et un nombre de chances du même ordre pour que :

$$\sigma\ (p_i) \geqq 1{,}65\ \sigma_i$$

Un intervalle de confiance à taux de signification 5 % sera donc :

$$0{,}6\ \sigma\ (p_i) \leqq \sigma_i \leqq 2{,}8\ \sigma\ (p_i)$$

On voit que la précision obtenue par cette méthode reste faible.

Aussi, peut-on avoir recours à un autre procédé d'estimation, en supposant que toutes les familles de même composition i ont la même probabilité d'avoir un garçon lorsqu'elles attendent une nouvelle naissance. La loi qui détermine le nombre de garçons nés est alors binomiale et l'écart-type du taux moyen p_i pour l'ensemble des cinq années peut être estimé très valablement par :

$$\sigma_i = 100 \sqrt{\frac{p_i\ (1 - p_i)}{n_i}} \# \frac{50}{\sqrt{n_i}}$$

si n_i est le nombre de naissances observées pour les familles de type i. Le nombre

(1) Pour chaque type de famille, le nombre supérieur a ete obtenu par estimation à partir des résultats annuels; tandis que le nombre inferieur représente la valeur theorique vraie dans le cas de l'hypothèse binomiale. (Pour plus de précisions, se reporter au texte.)

qui figure dans la partie inférieure de chaque case du tableau III a été calculé comme $\frac{100}{\sqrt{n_i}}$, où n_i est le nombre de naissances indiqué dans la case correspondante du tableau I. Il fournit donc une estimation précise du double de l'écart-type sous réserve que l'hypothèse binomiale soit vérifiée.

Les deux estimations reprises dans le tableau III donnent des ordres de grandeur comparables. Sauf dans deux cas, les écarts-types calculés par la seconde méthode restent à l'intérieur de l'intervalle de confiance défini ci-dessus, ce qui est tout à fait normal. Devant cet état de choses, il n'y a aucune raison pour ne pas conserver la seconde estimation malgré l'hypothèse quelque peu restrictive sur laquelle elle repose. Elle fournit, en effet, une évaluation plus précise. C'est elle que l'on retiendra dans ce qui suit.

Effet des naissances multiples (1)

Le tableau I présente les variations observées dans le taux de masculinité en liaison avec la composition de la famille. Il montre que le taux de masculinité pour les naissances vivantes est d'autant plus élevé que le nombre de garçons déjà nés est plus important et le nombre de filles déjà nées plus faible. Ce résultat n'est-il pas dû au procédé particulier employé pour l'enregistrement des naissances multiples? Rappelons, en effet, que pour ces naissances un ordre arbitraire a été adopté, de sorte que deux jumeaux sont portés dans deux cases voisines du tableau et non dans la même case. Dans le type de famille comportant un garçon et pas de fille, on trouvera donc non seulement une partie des naissances de second rang, mais aussi certains enfants nés d'un premier accouchement, à savoir tous les enfants jumeaux auxquels on aura donné le second rang et dont le frère sera un garçon. La proportion des garçons dans ce second groupe sera voisine des deux tiers puisque c'est à peu près la proportion des jumeaux de même sexe. Ainsi, même si le taux de masculinité parmi les naissances de second rang proprement dites était le même que celui observé pour les naissances de premier rang (51,44), quel qu'ait été le sexe du premier enfant, le taux de masculinité observé dans notre tableau pour le type de famille comportant un garçon et pas de fille serait supérieur à 51,44. Cette particularité due au procédé d'enregistrement des naissances multiples, est étrangère au phénomène que nous entendons étudier. Aussi allons-nous chercher à l'éliminer.

Pour ce faire. le plus simple consiste à rechercher quels auraient été les taux de masculinité moyens pour chaque type de famille avec les principes d'enregistrement adoptés si le taux de masculinité des accouchements était toujours égal à 51,50, indépendamment de la composition antérieure de la famille. Le calcul, effectué en tenant compte des nombre de jumeaux observés pendant la période 1946 à 1950, conduit aux résultats repris dans le tableau IV (2). La distorsion des taux introduite par la convention adoptée présente un caractère assez important pour qu'il soit nécessaire de l'éliminer avant de procéder à des interprétations sur les résultats du tableau I.

(1) Ce paragraphe n'a pas été traité dans l'exposé oral.
(2) En fait, pendant les premières années, les procédés d'enregistrement prévus n'avaient pas été rigoureusement suivis ce qui a compliqué le calcul.

Les nombres obtenus après élimination apparaissent sur le tableau V. Pour procéder à cette élimination, on s'est d'ailleurs contenté de retrancher des taux observés, l'excès par rapport à 51,50 du taux correspondant calculé dans le tableau IV. Ainsi, pour la case correspondant à un garçon et pas de fille, le calcul est le suivant :

$$51{,}68 = 51{,}87 - (51{,}69 - 51{,}50)$$

Tableau IV

Taux de masculinité théoriques dans l'hypothèse où le sexe à l'accouchement ne dépendrait pas des naissances antérieures (1)

Nombre de filles déjà nées	Nombre de garçons déjà nés				
	0	1	2	3	4
0	51,50	51,69	51,7	51,8	51,9
1	51,26	51,47	51,6	51,7	51,8
2	51,2	51,4	51,5	51,6	51,7
3	51,2	51,3	51,4	51,5	51,6
4	51,1	51,2	51,3	51,4	51,5

Tableau V

Taux de masculinité corrigés en liaison avec la composition de la famille (2)

Nombre de filles déjà nées	Nombre de garçons déjà nés						
	0	1	2	3	4	5	6
0	51,44 (0,08)	51,68 (0,13)	52,3 (0,3)	52,3 (0,5)	54,0 (0,9)	53,0 (1,5)	51,6 (2,5)
1	51,12 (0,14)	51,19 (0,19)	51,3 (0,3)	52,2 (0,5)	52,2 (0,8)	52,9 (1,2)	54,1 (1,9)
2	50,5 (0 3)	50,6 (0,3)	50,7 (0,4)	51,2 (0,6)	51,7 (0,8)	52,1 (1,2)	51,4 (1,5)
3	49,6 (0,5)	49,5 (0,5)	50,4 (0,6)	51,5 (0,7)	50,6 (0,9)	50,8 (1,2)	52,4 (1,6)
4	48,7 (1,0)	49,8 (0,8)	49,9 (0,8)	50,3 (1,0)	50,7 (1,1)	51,5 (1,4)	51,9 (2,0)
5	49,2 (1,8)	49,4 (1,3)	50,0 (1,2)	48,9 (1,2)	51,3 (1,4)	51,4 (1,8)	
6	48,9 (3,1)	48,6 (2,2)	48,3 (2,0)	50,4 (1,8)	48,8 (2,1)		

Dans les paragraphes qui vont suivre, seuls seront retenus les taux corrigés du tableau V. Afin de faciliter leur utilisation, l'indication du taux a été com-

(1) Pour la signification exacte de ces taux, se reporter au texte.
(2) Pour le calcul des taux se reporter au texte. Le nombre porté entre parenthèses indique le double de l'écart-type qui affecte le taux de masculinité correspondant.

plétée par celle du double de son écart-type tel qu'il peut être évalué à partir de l'hypothèse binomiale. Le nombre correspondant figure entre parenthèses dans chaque case du tableau.

Formule simplifiée reliant le taux de masculinité à la composition de la famille

Afin d'utiliser les données rassemblées dans le tableau V, il est sans doute utile de chercher à les synthétiser par une formule analytique simple, de façon à pouvoir se passer du tableau détaillé dans certains raisonnements rapides. Il semble assez naturel de rechercher une formule linéaire du type :

$$(3) \qquad p_i = p_o + \alpha\, n_i + \beta\, m_i$$

où p_i, n_i et m_i désignent respectivement, pour le type de famille i, le taux de masculinité, le nombre de garçons déjà nés et le nombre de filles déjà nées. α et β sont des constantes inconnues à déterminer.

Avec un tel schéma, et connaissant les écarts-types des erreurs qui affectent les taux observés, la méthode linéaire d'estimation la plus précise peut être appliquée. A peu de choses près, elle consiste à rechercher le minimum de la somme pondérée des carrés des écarts, les coefficients de pondération étant inversement proportionnels aux écarts-types (1). Le calcul conduit alors à la loi :

$$(4) \qquad \begin{array}{l} p_i = 51{,}45 + 0{,}32\, n_i - 0{,}53\, m_i \\ \quad (\pm\, 0{,}04)\ (\pm\, 0{,}03)\ \ (\pm\, 0{,}03) \end{array}$$

(1) En effet, le taux de masculinité p_i observé pour le type de famille i est, à vrai dire, une variable aléatoire qui peut s'exprimer suivant la formule :

$$(1) \qquad p_i = p_o + \alpha\, n_i + \beta\, m_i + \varepsilon_i\, \sigma_i$$

où σ_i est l'écart-type défini précédemment et ε_i une variable aléatoire de moyenne nulle et de variance égale à l'unité.

Il en résulte que :

$$(2) \qquad \frac{p_i}{\sigma_i} = p_o \frac{1}{\sigma_i} + \alpha \frac{n_i}{\sigma_i} + \beta \frac{m_i}{\sigma_i} + \varepsilon_i$$

Les σ_i étant supposés connus, le problème revient à estimer les coefficients p_o, α et β de la loi ci-dessus. On sait que la méthode la plus satisfaisante pour effectuer cette estimation consiste à rechercher le minimum de la somme des carrés des écarts :

$$\Sigma \left(\frac{p_i}{\sigma_i} - p_o \frac{1}{\sigma_i} - \alpha \frac{n_i}{\sigma_i} - \beta \frac{m_i}{\sigma_i}\right)^2$$

Les calculs sont classiques. Ils conduisent à la loi :

$$\frac{p_i}{\sigma_i} = 51{,}45 \frac{1}{\sigma_i} + 0{,}32 \frac{n_i}{\sigma_i} - 0{,}53 \frac{m_i}{\sigma_i}$$

La somme des carrés des 46 résidus est égale à :

$$\Sigma\, \varepsilon_i^2 = 63$$

Comme la somme des carrés des $\frac{p_i}{\sigma_i}$ est égale à 496, la proportion de la variance expliquée par l'ajustement est égale à :

$$R^2 = \frac{496 - 63}{496} = 0{,}87$$

Si la loi (2) était rigoureusement vérifiée la somme des carrés des résidus suivrait une loi de χ^2 à 43 degrés de liberté, donc sensiblement normale de moyenne 43 et d'écart-type 9,3. La probabilité d'observer une valeur au moins egale à 63 n'est donc que de 1,6 %.

La matrice des covariances des paramètres estimés p_o, α et β est donnée par :

$$10^{-4} \begin{bmatrix} 13{,}9 & 3{,}9 & 3{,}6 \\ 3{,}9 & 10{,}0 & 4{,}4 \\ 3{,}6 & 4{,}4 & 10{,}3 \end{bmatrix}$$

pour les taux de masculinité corrigés de l'influence des naissances multiples. La proportion de la variance totale expliquée par cet ajustement est égale à $R^2 = 0{,}87$. Les écarts-types qui affectent les coefficients α et β estimés se trouvent égaux à 0,03. Compte tenu de cette imprécision, on pourra utiliser la loi plus simple :

$$p_i = 51{,}45 + 0{,}3\, n_i - 0{,}5\, m_i \tag{5}$$

On peut encore se demander si la loi obtenue explique bien l'ensemble des taux observés. Pour cela, on appliquera un test d'ajustement sur les résidus. Celui-ci fait apparaître des résidus un peu trop forts puisque la somme de leurs carrés n'aurait qu'une probabilité de 1,6 % d'être atteinte si le schéma linéaire était rigoureusement valable. L'écart semble d'ailleurs attribuable presque entièrement aux taux observés pour les deux types de famille comprenant déjà un seul enfant. Après élimination de ces deux taux, le test d'ajustement donnerait une probabilité de 10 % pour une valeur au moins aussi forte de la somme des carrés des résidus.

En conclusion, la loi linéaire ci-dessus explique bien les taux de masculinité corrigés de l'influence des naissances multiples, sauf pour les deux types de famille comprenant déjà un seul enfant pour lesquels elle conduit à un écart trop fort par rapport à 51,45.

Hypothèses explicatives du taux de masculinité

Les tableaux statistiques et l'ajustement obtenu permettent d'examiner quelques hypothèses simples émises sur la détermination des sexes dans les naissances. Rappelons d'abord que les taux p_i considérés mesurent la proportion de garçons dans l'ensemble des naissances enregistrées de la part des familles d'une composition donnée. Comme nous ignorons le sexe des enfants avant leur naissance, ce taux correspond bien à une caractéristique purement biologique. Il ne semble pas qu'il puisse être affecté au moins directement par aucun facteur social ou psychologique ni par la limitation des naissances. C'est donc sans doute un instrument assez adéquat pour l'étude biologique.

Peut-on d'abord admettre que tous les couples ont, à tout moment, la même probabilité d'engendrer un garçon s'ils attendent une naissance? Dans la terminologie précédente ceci reviendrait à admettre, quel que soit i :

$$p_i = p_o$$

Les valeurs trouvées précédemment pour les coefficients de n_i et m_i sont très significativement différentes de zéro, de sorte que cette hypothèse peut être rejetée sans le recours à des procédés statistiques très complexes qui seraient appliqués sur les résultats du tableau V.

Suivant une autre explication, qui a reçu quelque crédit, la probabilité d'engendrer un garçon dépendrait du sexe du dernier enfant né et serait, à part cela, la même pour tous les couples. Elle prendrait donc l'une ou l'autre des deux valeurs p_g et p_f suivant que le dernier enfant né aurait été un garçon ou une fille. Certaines conséquences en découleraient pour les taux p_i que nous avons étudiés. Par exemple, dans les familles qui ont eu précédemment au moins un garçon mais aucune fille, le taux p_i devrait être indépendant du nombre de garçons déjà nés. Ceci est en contradiction avec notre formule et avec les résultats du tableau V. Ainsi, le taux de masculinité pour les familles comprenant déjà deux garçons est significativement plus fort que pour celles possédant un seul garçon (1).

Ceci semble tout d'abord contredire des résultats établis par M. Schützenberger. Étudiant des familles nombreuses, cet auteur a montré que les séries d'enfants du même sexe dans des familles de composition donnée sont plus fréquentes qu'elles devraient l'être si les sexes des enfants successifs étaient indépendants. Il a alors été conduit à introduire l'hypothèse précédente avec pour p_g et p_f les valeurs 52,1 et 50,9. La contradiction s'explique sans doute par le fait que l'étude avait un domaine de validité différent de la nôtre. Alors que M. Schützenberger étudiait les familles de cinq enfants et plus, nos données statistiques ne sont bonnes que pour les familles peu nombreuses, pour lesquelles l'espacement des naissances est plus marqué. Elles ne renseignent pas sur l'ordre des enfants, et ne se prêtent donc pas à l'analyse du taux de masculinité suivant le sexe et la date de la dernière naissance. Ces facteurs doivent jouer un certain rôle. Mais ils n'expliquent pas, à eux seuls, toutes les variations du taux de masculinité.

La question que nous avions formulée au début consistait à se demander si la probabilité d'engendrer un garçon ne serait pas variable suivant les couples, certains étant plutôt prédestinés à donner naissance à des garçons, d'autres à engendrer des filles. C'est ce type de questions que nous allons envisager maintenant.

Taux de masculinité variable suivant les couples

Pour cette étude, nous formulerons les deux hypothèses suivantes :

Hypothèse 1. — Pour tout couple donné, la probabilité qu'une naissance attendue soit masculine a une valeur q, valable pour toute la vie de ce couple et indépendante du nombre et du sexe des enfants déjà nés.

Hypothèse 2. — La fécondité des couples est indépendante de la probabilité q définie ci-dessus, c'est-à-dire, qu'elle est en moyenne la même pour les couples dans lesquels q est élevé que pour ceux pour lesquels il est bas.

Il est alors naturel de rechercher la loi de distribution de q dans la population française, c'est-à-dire la fonction non décroissante F (q) qui définit la proportion des ménages pour lesquels la probabilité étudiée ne dépasse pas q. Une relation étroite existe évidemment entre F (q) et les taux de masculinité p_i observés

---

(1) On peut procéder à un test sur l'ensemble des familles telles que $n_i = o$, $m_i > o$. On aurait en effet :

$$p_i = p_q + \sigma_i \varepsilon_i$$

La même méthode que précédemment conduirait à $p_q = 51{,}82$. La somme des carrés des résidus ε_i* serait alors de 39, valeur qui a une probabilité extrêmement faible dans une loi χ^2 à 5 degrés de liberté. De même, dans les familles telles que $m_i = 0$, $n_i > o$, on aurait :

$$p_i = p_f + \sigma_i \varepsilon_i$$

L'ajustement donne $p_f = 50{,}89$. La somme des carrés des résidus ε_i* serait alors de 66, donc encore plus forte que ci-dessus.

Ces deux tests permettent donc dé rejeter l'hypothèse simple formulée dans le texte.

pour les différents types de familles. Si les hypothèses 1 et 2 sont satisfaites, elles se traduisent par (1) :

$$(6) \qquad p_i = \frac{\int_0^1 q^{n_i+1} (1 - q)^{m_i} \, dF(q)}{\int_0^1 q^{n_i} (1 - q)^{m_i} \, dF(q)}$$

Cette égalité établit des relations entre les taux de masculinité des divers types de familles et les moments de la loi de distribution F (q). Si, changeant un peu les notations, on désigne par p_{nm} le taux des familles comprenant n garçons et m filles, les égalités suivantes se vérifient facilement :

$$\begin{array}{lll} (7) & p_{00} = E(q) & \\ & p_{10}\, E(q) = E(q^2) & p_{01}\,[1 - E(q)] = E(q) - E(q^2) \\ & p_{20}\, E(q^2) = E(q^3) & p_{11}\,[E(q) - E(q^2)] = E(q^2) - E(q^3) \end{array}$$

etc...

Si l'examen est limité aux cas pour lesquels $n_i \leq 4$ et $m_i \leq 4$, il existera 25 relations semblables entre les 9 premiers moments de F (q). Elles ne pourront être satisfaites que si les taux de masculinité p_{nm} satisfont certaines conditions :

$$(1 - p_{00}) = p_{00}\,(1 - p_{10}) \qquad (1 - p_{10})\, p_{11} = p_{10}\,(1 - p_{20})$$

et plus généralement :

$$(8) \qquad (1 - p_{nm})\, p_{n,\, m+1} = p_{nm}\,(1 - p_{n+1,\, m})$$

Ces conditions auraient d'ailleurs pu être établies directement à partir des hypothèses 1 et 2. Elles doivent permettre de tester la validité de ces hypothèses.

Il est facile de voir qu'aucune loi linéaire du type (3) ne peut satisfaire exactement les relations (8). Cependant les écarts resteraient dans les limites des fluctuations aléatoires affectant notre étude si l'on avait :

$$p_0 \# 0{,}50 \quad \text{et} \quad -\frac{\beta}{\alpha} \# 1$$

Si la loi (5) ne respecte pas les conditions (8), c'est surtout parce qu'elle implique une décroissance pour p_{nn} quand n augmente. Nous reviendrons sur ce point un peu plus loin.

Si la loi (3) était remplacée par la formule voisine :

$$(9) \qquad p_{nm} = p_0 + \alpha\,(n - m)$$

les hypothèses initiales pourraient être maintenues, compte tenu des erreurs d'échantillonnage. La distribution F (q) aurait alors approximativement pour

---

(1) En effet, si f désigne la probabilité pour un couple d'avoir $n_i + m_i + 1$ enfants, le nombre de garçons nés de familles possédant déjà n_i garçons et m_i filles, dans un ensemble de 1.000 couples est égal à :

$$C^{n_i}_{n_i + m_i}\, f\, p_i \int_0^1 q^{n_i} (1 - q)^{m_i}\, dF(q); \text{ mais il est aussi égal à :}$$

$$C^{n_i}_{n_i + m_i}\, f \int_0^1 q^{n_i+1} (1 - q)^{m_i}\, dF(q).$$

moyenne p_0, pour variance $p_0\,\alpha$, pour moment centré du troisième ordre $2\,p_0\,\alpha^2$ et pour moment centré du quatrième ordre $3\,p_2^0\,\alpha^2$. (Tous ces résultats se déduisent directement des formules (7). F (q) serait voisine d'une distribution normale.

Si donc on néglige la décroissance du taux de masculinité avec le rang de naissance, les taux observés pourraient s'expliquer à l'aide des hypothèses 1 et 2 avec une loi de distribution normale de moyenne $p_0 = 0{,}515$ et d'écart-type : $\sigma = 0{,}045$. La plupart des couples auraient une probabilité q d'engendrer des garçons comprise entre 40 % et 60 %. Les écarts entre les différents couples seraient donc relativement faibles.

Bien entendu la loi F (q) n'est pas déterminée avec une bien grande précision par notre procédé d'investigation, de sorte que sa forme exacte reste encore très imprécise. Cependant, on peut exclure certaines lois de répartition. Par exemple, une distribution qui comporterait une proportion, même faible, de couples ne pouvant avoir que des garçons ou que des filles, conduirait pour p_{no} ou p_{om} à des valeurs présentant une évolution exponentielle en contradiction avec les résultats observés. Il nous faut d'ailleurs examiner d'un peu plus près la décroissance du taux de masculinité avec le rang de naissance.

Taux de masculinité et rang de naissance

Les données statistiques analysées ici permettent aussi d'établir le tableau VI qui fait apparaître une décroissance lente du taux de masculinité avec le rang de naissance. Ce phénomène est conforme aux résultats de la formule (5) puisque la répartition des familles selon leur composition est sensiblement symétrique. Il est confirmé par des statistiques sur les naissances aux États-Unis entre 1942 et 1950, soit environ 28 millions de naissances (1).

TABLEAU VI

Taux de masculinité suivant le rang de naissance

(Nombre de garçons pour 100 naissances légitimes vivantes)
(France entière - 1946 à 1950)

NOMBRE D'ENFANTS DÉJA NÉS	0	1	2	3	4	5	6	7	8	9	10
Taux de masculinité	51,5	51,4	51,3	51,1	50,9	51,0	51,1	50,9	50,4	51,4	50,9
Double de l'écart-type sur ce taux	0,1	0,1	0,1	0,2	0,2	0,3	0,4	0,5	0,6	0,7	0,9

Le phénomène constaté ici est en contradiction avec les hypothèses 1 et 2 introduites dans le paragraphe précédent. Il pourrait s'expliquer par une modification de ces hypothèses. Par exemple, la probabilité q d'engendrer des garçons pourrait diminuer dans chaque famille avec l'âge des parents ou le nombre d'enfants déjà nés. Peut-être aussi la fécondité des couples n'est-elle pas indépendante de q. Si elle est d'autant plus forte que q est plus faible, on

(1) MYERS Robert-J. : « The Effect of Age Mother and Birth Order on Sex Ratio at Birth - The Milbank Memorial Fund Quarterly », juillet 1954.

trouvera dans les familles qui attendent un cinquième enfant relativement plus de couples avec des probabilités q basses que dans les familles attendant une première naissance. Ceci expliquerait une baisse du taux de masculinité avec le rang de naissance.

Composition des familles et fécondité

Le dépouillement des naissances présenté ici permet aussi d'étudier la composition des familles qui ont déjà n enfants et attendent une $n + 1^{me}$ naissance. Le tableau VII fournit la répartition des familles de n enfants dans lesquelles il y a eu une nouvelle naissance (la $n + 1^{me}$) suivant le nombre de garçons dans ces familles. Ces résultats permettent de constater que la fécondité des familles est un peu plus forte que la moyenne pour celles qui ont déjà eu surtout des garçons ou surtout des filles. Ce fait est probablement dû à une limitation des naissances moins fréquente dans les familles qui n'ont pas eu de fille, ou pas de garçon que dans celles qui ont eu un nombre sensiblement égal d'enfants des deux sexes.

A titre d'exemple, considérons le cas des familles dans lesquelles s'est produite une quatrième naissance. Sur 1.000 familles, 149 avaient déjà trois garçons. Cette proportion est-elle supérieure à celle que l'on observerait si la fécondité des familles était indépendante de leur composition? Une première réponse à cette question pourrait être obtenue en comparant la proportion observée au résultat du calcul rapide suivant. Le taux de masculinité pour l'ensemble de la population (mort-nés compris) étant de 51,5 %, on devrait avoir $1.000 \times (0{,}515)^3$ familles de trois garçons parmi 1.000 familles attendant une quatrième naissance. Le nombre correspondant serait donc de 137, notablement inférieur aux 149 observés. Ce raisonnement repose sur l'hypothèse simplifiée suivante :

TABLEAU VII

Répartition des familles au moment d'une nouvelle naissance suivant leur composition

(en pour mille)

NOMBRE (1) DE GARÇONS déjà nés	NOMBRE (1) D'ENFANTS DANS LA FAMILLE AVANT LA NOUVELLE NAISSANCE							
	0	1	2	3	4	5	6	7
0	1.000	479	238	122	63	33	17	9
1		521	487	353	231	145	90	53
2			275	376	366	295	218	152
3				149	260	312	303	257
4					80	171	239	271
5						44	108	176
6							25	68
7							.	14
Ensemble	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

(1) Ces nombres correspondent à l'ensemble des naissances précédentes. Ils comprennent donc les enfants encore vivants, les enfants déjà décédés et les mort-nés.

TABLEAU VIII

Composition des familles dans lesquelles est survenue une quatrième naissance

(En pour mille)

COMPOSITION DE LA FAMILLE	RÉPARTITION observée	HYPOTHÈSE 3	TAUX de masculinité observés et hypothèse 4
3 garçons	149	137	142
2 garçons, 1 fille	376	386	381
1 garçon, 2 filles	353	363	360
3 filles	122	114	117
Ensemble	1.000	1.000	1.000

Hypothèse 3. — La fécondité des familles de *n* enfants est indépendante de la composition de la famille. Le taux de masculinité est égal à 51,5 % pour toutes les familles de la population étudiée.

Nous avons vu qu'il y avait justement de bonnes raisons de penser que le taux de masculinité variait suivant les familles. Cependant, si l'on admet pour les familles de composition donnée les taux de masculinité observés, on peut encore calculer la proportion de familles de trois garçons en retenant seulement l'hypothèse suivante :

Hypothèse 4. — La fécondité des familles de *n* enfants est indépendante de leur composition.

Les résultats du calcul effectué de cette façon figurent à la dernière colonne du tableau IX (1). Le nombre de 142 familles de 3 garçons, quoique plus fort que celui obtenu avec l'hypothèse 3, est notablement inférieur au nombre observé. Des résultats analogues peuvent être établis pour les familles de 2, 4, 5, 6 et 7 enfants. Pour être parfaitement précis, il faut d'ailleurs tenir compte des erreurs d'échantillonnage qui affectent à la fois la première et la dernière colonne du tableau IX. En fait, pour la comparaison de ces colonnes, l'écart-type à faire intervenir est toujours inférieur à 1. La différence est donc nettement significative.

CONCLUSION

Quelles lumières les développements précédents apportent-ils pour la compréhension des phénomènes liés à la détermination du sexe à la naissance? Ils semblent suggérer que la capacité à engendrer des garçons varie quelque peu suivant les couples, ou peut-être suivant les mères et qu'elle décroît avec l'âge des parents. Mais ces phénomènes restent d'une ampleur assez limitée. Ils paraissent d'ailleurs confirmer les idées actuelles des généticiens d'après lesquelles le taux de masculinité dépendrait des conditions dans lesquelles

(1) En fait, il a fallu corriger quelque peu les taux figurant dans le tableau I de façon à obtenir des taux valables pour l'ensemble « naissances vivantes et mort-nés » et non plus seulement pour les seules naissances vivantes. En revanche, il n'y a pas lieu de tenir compte des règles spéciales introduites pour l'enregistrement des naissances multiples puisqu'elles jouent de la même façon sur les taux et sur la répartition des familles d'après leur composition.

s'effectuent la conception et le développement du fœtus, plutôt que de caractères proprement génétiques.

DISCUSSION

M. Vincent. — J'ai suivi avec d'autant plus d'intérêt l'exposé de M. Malinvaud, que j'ai moi-même entrepris des recherches sur la succession des sexes, dans un ensemble d'environ 8.500 familles d'au moins 9 enfants, qui satisfont aux conditions posées par M. Malinvaud (enfants tous nés postérieurement à l'unique mariage de la mère), et d'où en outre les familles à jumeaux ont été éliminées. Je dispose de l'intervalle entre naissances successives, ce qui me permet d'examiner l'hypothèse de la « contagion » entre sexes des enfants successifs, en fonction de cet intervalle. J'ai aussi l'intention d'étudier la fréquence des séquences de plusieurs naissances du même sexe.

Malheureusement, ces recherches n'en sont qu'au stade préliminaire, de sorte que je ne puis vous communiquer que des impressions provisoires, et non de véritables résultats.

Ces impressions se résument essentiellement en ceci — que la communication de M. Malinvaud est venue confirmer à mes yeux — : les résultats que l'on obtient sont plutôt négatifs, en ce sens qu'ils conduisent généralement à éliminer des hypothèses plutôt qu'à en confirmer. C'est ainsi qu'il ne semble pas qu'il y ait « contagion » entre les sexes des enfants successifs. Et s'il existe des différences entre les répartitions observées et celles que donnerait l'hypothèse binomiale, elles sont extrêmement faibles et par conséquent très difficiles à mettre en évidence.

Certaines de ces différences peuvent d'ailleurs s'expliquer par des facteurs psychologiques. C'est ainsi que le désir d'avoir des enfants de sexes différents suffirait amplement à expliquer certaines fréquences anormalement élevées, observées pour des familles à enfants tous de même sexe.

Je crois donc que l'incidence de la fécondité différentielle suivant la composition de la famille a des chances d'être sensible, et qu'elle mérite une attention spéciale. On pourrait sans doute la mettre en évidence en comparant entre eux certains milieux socio-économiques, où l'on peut présumer que le désir d'avoir un héritier mâle, par exemple, est d'intensité très différente.

Le tableau de taux de masculinité obtenu par M. Malinvaud, en tout cas, s'affranchit des conséquences de cette fécondité différentielle, et il me paraît constituer le premier argument probant qui ait jamais été apporté, en faveur de l'hypothèse selon laquelle certains couples auraient une prédisposition à procréer des enfants d'un certain sexe. Remarquons toutefois que cette prédisposition est très faible.

« M. Féron fait remarquer que le problème traité par E. Malinvaud a déjà été étudié par Z. Sougarev dans sa thèse à l'Institut de Statistique de Paris : « Taux de masculinité dans les naissances », Paris, 1937.

Se basant sur les données statistiques de D. A. Geissler qui portent sur 4.800.000 naissances provenant d'environ 1 million de familles (Geissler, *Beitrage zur Frage des Geschlechtsverhältnisses des Geboren-Zeits K. Sachrichen Stat*

Bureau, 1889), Z.Sougarev obtient des résultats tout à fait en accord avec ceux de Malinvaud. Il essaye ensuite d'expliquer les résultats obtenus et examine successivement les hypothèses de Geissler, Lexis, Fisher, Halbwachs (*J. Soc. Stat.*, 1933) et finalement montre que l'on a une bonne explication des faits observés en combinant les deux hypothèses suivantes dues à G. Darmois :

1° *Première hypothèse.* — La probabilité d'une naissance masculine n'a pas une valeur constante pour toutes les familles; les familles se répartissant suivant leurs prédispositions à produire plus de mâles ou de femelles.

2° *Deuxième hypothèse.* — Chaque naissance d'un garçon (ou d'une fille) agit sur l'organisme de la femme de façon à modifier sa prédisposition à avoir encore un garçon et il propose un schéma contagieux qui rend compte de cette hypothèse.

Les hypothèses de G. Darmois seraient faciles à vérifier dans le cas où les données expérimentales se présentent sous la forme de celles considérées par P. Vincent pour le prix Cognacq. Mais grâce aux calculs de Z. Sougarev il serait possible de confronter partiellement ces hypothèses avec l'expérience même quand les tableaux expérimentaux ont la forme donnée par Malinvaud ».