JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

ALFRED J. LOTKA

Applications de l'analyse au phénomène démographique

Journal de la société statistique de Paris, tome 74 (1933), p. 336-343
<http://www.numdam.org/item?id=JSFS_1933__74__336_0>

© Société de statistique de Paris, 1933, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

III

VARIÉTÉ

Applications de l'analyse au phénomène démographique.

A plusieurs reprises les lecteurs du « Journal de la Société de Statistique » ont été avisés indirectement de certaines méthodes d'analyse apportées par l'auteur au problème du mouvement de la population. A ce propos, M. Husson s'exprime dans ces ces mots : « Lotka montre que la composition par l'âge d'une population fermée, à tout instant, est la solution d'une certaine équation intégrale. Mais cette équation intégrale n'a pu être résulue jusqu'à présent qu'en supposant les fonctions de fécondité et de mortalité indépendantes du temps. » Il serait donc bien à propos d'indiquer comment la méthode des équations intégrales peut être étendue à certains cas d'un intérêt particulier, dans lesquels la fécondité n'est plus constante, mais au contraire varie avec le temps. Une certaine extension de l'analyse porte alors encore au cas où la mortalité aussi est variable.

Relation entre le nombre de la population, les naissances annuelles et la table de survie. — Rappelons tout d'abord brièvement le développement de l'équation fondamentale. Soit N (t) le nombre de la population à l'instant t, et B (t) le nombre de naissances annuelles à cette même époque. Soit encore $p(a)$ la probabilité au moment de sa naissance qu'un individu quelconque atteigne l'âge a. Cette probabilité, dans notre première discussion, nous la supposerons indépendante du temps t.

Évidemment, le nombre N (t) d'individus en vie à l'instant t se compose de la somme de tous ceux qui, ayant été nés à un instant t-a, ont survécu jusqu'à l'âge a.

Nous posons donc :

$$N(t) = \int_0^\infty B(t-a)\, p(a)\, da \tag{1}$$

Application à une population croissant d'après la loi dite « logistique ». — Un cas d'intérêt particulier est celui d'une population croissant d'après la loi de Verhulst-Pearl, dite « logistique », qui, jusqu'à l'année 1930, représente bien fidèlement la courbe d'accroissement de la population des États-Unis. Cette loi s'exprime par la formule bien simple :

$$N(t) = \frac{N_\infty}{1 + e^{-rt}} \tag{2}$$

les quantités N_∞ et r étant deux constantes caractéristiques de la population. Une troisième constante encore est cachée dans la formule, car celle-ci n'a cette forme toute simple qu'à la condition que le temps t soit mesure d'une origine particulière, qui, elle aussi, représente une constante caractéristique de la population.

Il conviendra, pour le but que nous nous posons, de développer le membre droit de (2) en série ordonnée suivant les puissances croissantes de e^{rt}. On aura alors, en vue de (1).

$$\int_0^\infty B(t-a)\, p(a)\, da = N_\infty \left(e^{rt} - e^{2rt} + e^{3rt} - \ldots\right) \tag{3}$$

On s'assure sans difficulté que la substitution

$$B(t) = U_1 e^{r} - U_2 e^{2rt} + \ldots \tag{4}$$

satisfait à l'équation (3), pourvu que

$$U_k = \frac{N_\infty}{\int_0^\infty e^{-kra}\, p(a)\, da} \tag{5}$$

De cette façon, nous avons obtenu une solution de l'équation (2), car dans (5), les quantités N_∞ et r, ainsi que la fonction $p(a)$ sont des données caractéristiques de la population, de sorte que chacune des constantes U_k se détermine par une intégration numérique qui n'offre, en principe, aucune difficulté.

Cependant, la solution, dans cette forme, est peu pratique, comme son application numérique demanderait la sommation d'une série jusqu'à des termes d'ordre élevé. Une transformation nous permettra de remédier à cet inconvénient, en même temps qu'elle nous donnera une idée plus claire de la nature de la fonction $B(t)$ définie par l'équation (3).

Pour obtenir ce résultat, nous développons les constantes U_k en séries ordonnées suivant les puissances croissantes de kr, c'est-à-dire, nous posons :

$$U_k = \frac{N_\infty}{\int_0^\infty e^{-kra}\, p(a)\, da} = \left\{c_0 + c_1 kr + c_2 k^2 r^2 + \ldots\right\} N_\infty \tag{6}$$

Introduisant ces expressions dans (4), nous obtenons, après avoir ordonné les termes suivant les puissances croissantes de r,

$$B(t) = N_\infty e^{rt} \left\{\frac{c_0}{1+e^{rt}} + \frac{c_1 r}{(1+e^{rt})^2} + \frac{c_2 r^2 (1-e^{rt})}{(1+e^{rt})^3} + \ldots\right. \tag{7}$$

On reconnaît que les termes successifs du membre droit sont, au facteur $c_j N_\infty$ près, les dérivées d'ordre ascendant, de la fonction

$$\varphi(t) = \frac{1}{1+e^{-rt}} = 1/2 + 1/2 \tanh \frac{rt}{2} \tag{8}$$

de façon que la formule (7) se présente sous la forme excessivement simple :

$$B(t) = N_\infty \sum_{j=0}^{j=\infty} c_j \varphi^j(t) \tag{9}$$

le symbole $\varphi^j(t)$ signifiant la dérivée d'ordre j de la fonction φ. Les coefficients c sont définis par l'équation (6). Ils sont en première ligne fonctions des moments L_n de la courbe de survie $p(a)$, c'est-à-dire de

$$L_n = \int_0^\infty a^n \, p(a) \, da \tag{10}$$

Ils se réduisent, cependant, à des expressions particulièrement simples, lorsqu'on remplace le variable t dans le membre droit de (7) par :

$$t' = t + \frac{L_1}{L_0} = t + \lambda_1 \tag{11}$$

Dans ce cas, les coefficients c_j prennent la forme

$$\left.\begin{aligned}
c_0 &= \frac{1}{L_0} \\
c_1 &= 0 \\
c_2 &= -\frac{\lambda_2}{2!\,L_0} \\
c_3 &= \frac{\lambda_3}{3!\,L_0} \\
c_4 &= -\frac{1}{4!\,L_0}\left\{\lambda_4 - 3\lambda_2^2\right\} \\
c_5 &= \frac{1}{5!\,L_0}\left\{\lambda_5 - 10\lambda_2\lambda_3\right\}
\end{aligned}\right\} \tag{12}$$

les quantités λ étant certaines fonctions bien connues, dites semi-invariants de Thiele, des moments L_n, de la courbe de survie $p(a)$.

Ces substitutions faites dans (9), ayant égard à (8), (11) et (12), nous arrivons enfin à la formule la plus caractéristique exprimant les naissances annuelles $B(t)$ en fonction du temps.

$$\left.\begin{aligned}
B(t) = \frac{N_\infty}{L_0}\Bigg\{&\frac{1}{2}\left(1 + \tanh\frac{rt'}{2}\right) + \frac{\lambda_2 r^2}{2!\,4}\left(\tanh\frac{rt'}{2} - \tanh^3\frac{rt'}{2}\right) \\
&- \frac{\lambda_3 r^3}{3!\,8}\left(1 - 4\tanh^2\frac{rt'}{2} + 3\tanh^4\frac{rt'}{2}\right) \\
&- \frac{(\lambda_4 - 3\lambda_2^2)r^4}{4!\,4}\left(2\tanh\frac{rt'}{2} - 5\tanh^3\frac{rt'}{2} + 3\tanh^5\frac{rt'}{2}\right) \\
&+ \frac{(\lambda_5 - 10\lambda_2\lambda_3)\,r^5}{5!\,8}\left(2 - 17\tanh^2\frac{rt'}{2} + 30\tanh^4\frac{rt'}{2} - 15\tanh^6\frac{rt'}{2}\right) - \ldots\Bigg\}
\end{aligned}\right\} \tag{13}$$

La fonction $\varphi(t)$ ne se distingue de la tangente hyperbolique que par la constante 1/2 et le facteur 1/2, de sorte que toutes les valeurs de cette fonction ainsi que de ses dérivées nécessaires pour le calcul s'obtiennent à l'aide de tables bien connues.

Mais l'avantage principal de la formule (13), c'est que le premier terme seul de la série, dans les cas qui se rencontrent en pratique, dominera complètement le caractère de la fonction $B(t)$. Les termes de second ordre et au delà ne fonctionnent que comme termes de correction presque négligeables; la figure 1, qui représente les résultats de calculs appliqués aux données numériques de la population des États-Unis, fait bien voir ceci. On y observe que la première composante, c'est-à-dire le premier terme de la série, représente, à une minuscule correction près, l'ensemble de la fonction $B(t)$. En effet, pour faire bien ressortir les composantes d'ordre 2, 3, 4, 5, il a fallu les représenter dans le second cadre de la figure 1 sur une échelle vingt-cinq fois plus grande que celle du premier cadre.

Taux de natalité. — Ayant obtenu une formule pour les naissances annuelles, on déduit immédiatement le taux de natalité en fonction du temps :

$$b(t) = \frac{B(t)}{N(t)} = \frac{1}{L_0}\left\{\frac{\varphi(t')}{\varphi(t)} - \frac{\lambda_2}{2!}\frac{\varphi''(t')}{\varphi(t)} + \ldots\ldots\right\}$$
$$= \frac{1}{L_0}\left\{1 + e^{r\lambda_1}\varphi(-t') - \frac{\lambda_2}{2!}\frac{\varphi''(t')}{\varphi(t)} + \ldots\ldots\right\} \quad (14)$$

C'est-à-dire, si l'on néglige les termes en φ'' et en dérivées d'ordre supérieur, le taux de natalité se présente sous la forme d'une constante à laquelle s'ajoute une courbe logistique inverse ayant pour centre le même instant $t = -\lambda_1$ qui est aussi le centre de la courbe des naissances annuelles. Il faut, cependant, signaler ici une différence importante : Aussi longtemps que nous nous occupons du nombre total des naissances annuelles, la première composante seule nous donne une idée assez juste de cette quantité, même pour les grandes valeurs négatives de t. De plus, le centre de la courbe se trouve effectivement, à très peu près, à l'instant $t' = 0$, c'est-à-dire à $t = -\lambda_1$. Ni l'une ni l'autre de ces remarques ne s'applique à la courbe du taux de natalité. On obtiendrait des valeurs très inexactes pour cette quantité en ne retenant que la première composante de l'expression (14) pour $b(t)$; et le centre (approximatif, car il n'y a pas symétrie exacte) de la courbe, est déplacé bien sensiblement de l'instant $t' = 0$. La raison de cette situation réside dans le fait que le nombre B (t) de naissances annuelles varie de zéro à plusieurs millions, tandis que les variations du taux $b(t)$ de natalité se tiennent entre des limites toutes étroites, disons entre 0,05 et 0,015. Une erreur de 10 ou même 20 %, qui aurait peu d'importance dans la partie primitive de la courbe de naissances (qui correspond à quelques milliers, tandis que sa partie ultérieure s'élève à 2 ou 3 millions), serait fatale lorsqu'il s'agit d'un taux de natalité. Il faut donc, en calculant le taux de natalité, retenir les termes d'ordre supérieur, ou, ce qui est plus simple, lorsqu'on aura obtenu les valeurs de B (t), en déduire directement celles de $b(t)$ en divisant par N (t).

Mortalité variable. — L'analyse exacte tenant compte des variations, non seulement de la fécondité (1), mais encore de la mortalité, présente des difficultés. Une approximation, par contre, se présente immédiatement, sans même changer la forme extérieure de nos formules. Pourvu que la fonction $p(a)$, que nous devons maintenant écrire $p(a, t)$, ne change que lentement avec le temps, il suffit que nous entendions désormais par les coefficients c_j, ainsi par les semi-invariants λ, des fonctions du temps, ce qui vient tout naturellement en conséquence de l'altération séculaire de la courbe de survie.

C'est dans ce sens que la formule (13) a été appliquée aux données de la statistique américaine, en négligeant provisoirement la partie jouée par l'immigration, correction étant apportée dans la suite pour ce facteur. C'est-à-dire, les valeurs de B (t) et ensuite de $b(t)$ ont été calculées comme si la croissance de la population s'était effectuée entièrement par l'excès des naissances par-dessus les morts. Des valeurs de $b(t)$ obtenues de cette façon, on a ensuite déduit une correction approximative correspondant à l'immigration nette effectivement enregistrée. C'est ainsi qu'ont été obtenues les valeurs de $b(t)$ du tableau I, où se trouvent aussi les taux de natalité effectivement observés. On voit qu'il y a correspondance bien satisfaisante, vu que nous avons quitté le terrain de l'analyse exacte lorsque nous admettions une table de survie variable avec le temps. De plus, notre correction pour l'immigration ne ferait aucune prétention à l'exactitude.

Répartition par l'âge. — Étant donné les naissances annuelles en fonction du temps, ainsi que la table de survie, il est évident que la répartition par l'âge est complètement déterminée (dans une population fermée). Le calcul a été appliqué premièrement sur base d'une table de survie fixe (États-Unis, 1919-1920, sexe féminin, race blanche) aux mêmes données américaines dont nous nous sommes déjà servis comme

(1) La croissance, selon la loi logistique, sous le régime d'une table de survie constante telle que nous l'avons supposée dans ce qui précède, demande nécessairement une fécondité decroissante.

exemple. Les résultats, en représentation graphique, forment une surface courbe caractéristique illustrée dans la figure 2.

Le calcul de la répartition par l'âge tenant compte de l'altération séculaire de la table de survie est une affaire bien laborieuse. Il n'a donc été fait que pour un seul cas. Commençant par la table anglaise de 1838-1854 (en défaut d'une table américaine applicable à cette époque), et continuant en six étapes à l'aide d'une série de tables aboutant à la table américaine de 1919-1920, le calcul mène à la courbe de répartition reproduite dans la figure 3. Dans ce diagramme, la courbe de répartition est superposée sur l'histogramme (« courbe escalier ») de la répartition effectivement enregistrée parmi la population indigène, de race blanche, de sexe féminin, en 1920.

En vue des approximations qui entrent dans nos formules lorsqu'il s'agit d'une mortalité variable, et en vue, de plus, des approximations sous-entendues dans l'application de ces formules à une population non fermée, (l'influence de l'immigration n'est que partiellement eliminée en se limitant à la population indigène), on ne s'attendrait qu'à une correspondance qualitative entre les chiffres calculés et l'observation. En effet, la courbe résultant des calculs suit l'histogramme des observations avec une fidélité surprenante.

*
* *

Si dans l'exposé qui a été donné ci-dessus de l'application d'une équation intégrale au phénomène démographique le cas d'une population croissant d'après la loi logistique a servi de paradigme, c'est que ce cas présente un intérêt particulier. La même méthode est pourtant applicable encore dans tous les cas où la croissance admet une représentation par une serie exponentielle quelconque. Au point de vue théorique donc, la méthode a un champ d'application assez étendu.

Il reste à dire que ce n'était pas l'objectif de l'auteur de faire une théorie de prévision. On cherchera la valeur de la méthode illustrée plutôt dans le coup d'œil qu'elle nous permet de jeter sur certaines caractéristiques fondamentales d'une population, et, surtout, sur les relations entre ces caractéristiques.

New-York.

Alfred J. Lotka.

Tableau I

Taux de natalité dans une population logistique. — Comparaison des résultats du calcul avec les taux effectivement enregistrés. — États-Unis. — Calcul basé sur une série de tables de survie (mortalité variable).

Année	Taux de natalité — Calculé — Brut	Taux de natalité — Calculé — Corrigé pour l'immigration	Taux de natalité — Enregistré
—	—	—	—
1915 (1) .	28	25	25
1916 . . .	28	25	25
1917 . . .	27	24	24
1918 . . .	27	24	24
1919 . . .	26	23	22
1920 . . .	26	23	23
1921 . . .	26	23	24
1922 . . .	25	22	22
1923 . . .	25	22	22
1924 . . .	24	21	22
1925 . . .	24	21	21
1926 . . .	23	20	20
1927 . . .	23	20	20
1928 . . .	22	19	19
1929 . . .	22	19	19

(1) Date à laquelle commencent les registres officiels des taux de natalité dans les « Birth Registration States ».

Tableau II

Valeurs des caractéristiques sur lesquelles sont basés les calculs.

	TABLE DE SURVIE					
	Angleterre et Pays de Galles, sexe féminin			États-Unis (1), race blanche, sexe féminin		
	1838-1854	1871-1880	1881-1890	1901	1910	1919-1920
Moments de la courbe de survie.						
L_0	41 8752	44.6244	47.1866	51.1019	53,6412	57,5173
L_1	1352,762	1466,602	1567,981	1740,587	1845,173	2014,934
L_2	63298,28	69005 92	73947,0	84027,48	89253,3	99179,1
L_3	3482718	3801783	4071170	4731810	5019800	5671231
L_4	210666300	230039100	245594100	291690700	308588100	354072100
L_5	13580900000	14807750000	15761150000	19112870000	20149620000	23455470000
Semi-invariants de la courbe de survie.						
λ_1.	32,30460	32,86547	33,22934	34,06113	34,39842	35,03147
λ_2.	468,0053	466,2323	462,9282	484,1531	480,6429	497,1154
λ_3.	4100,083	3727,031	3438,151	3606,870	3278,992	3364,727
λ_4.	— 175584,4	— 175350,8	— 181354,4	— 202749,2	— 203815,0	— 223390,4
λ_5.	— 8391050	— 8002230	— 6917734	— 7689962	— 6858849	— 7424076

Taux primitif d'accroissement r = 031425. L'origine de t est située à la date 1914,1. La population ultime N_∞ = 197.493.000.

(1) 1901 et 1910 les 10 États compris dans le territoire primitif d'enregistration; 1919-1920, territoire comprenant 27 États, d'après Foudray.

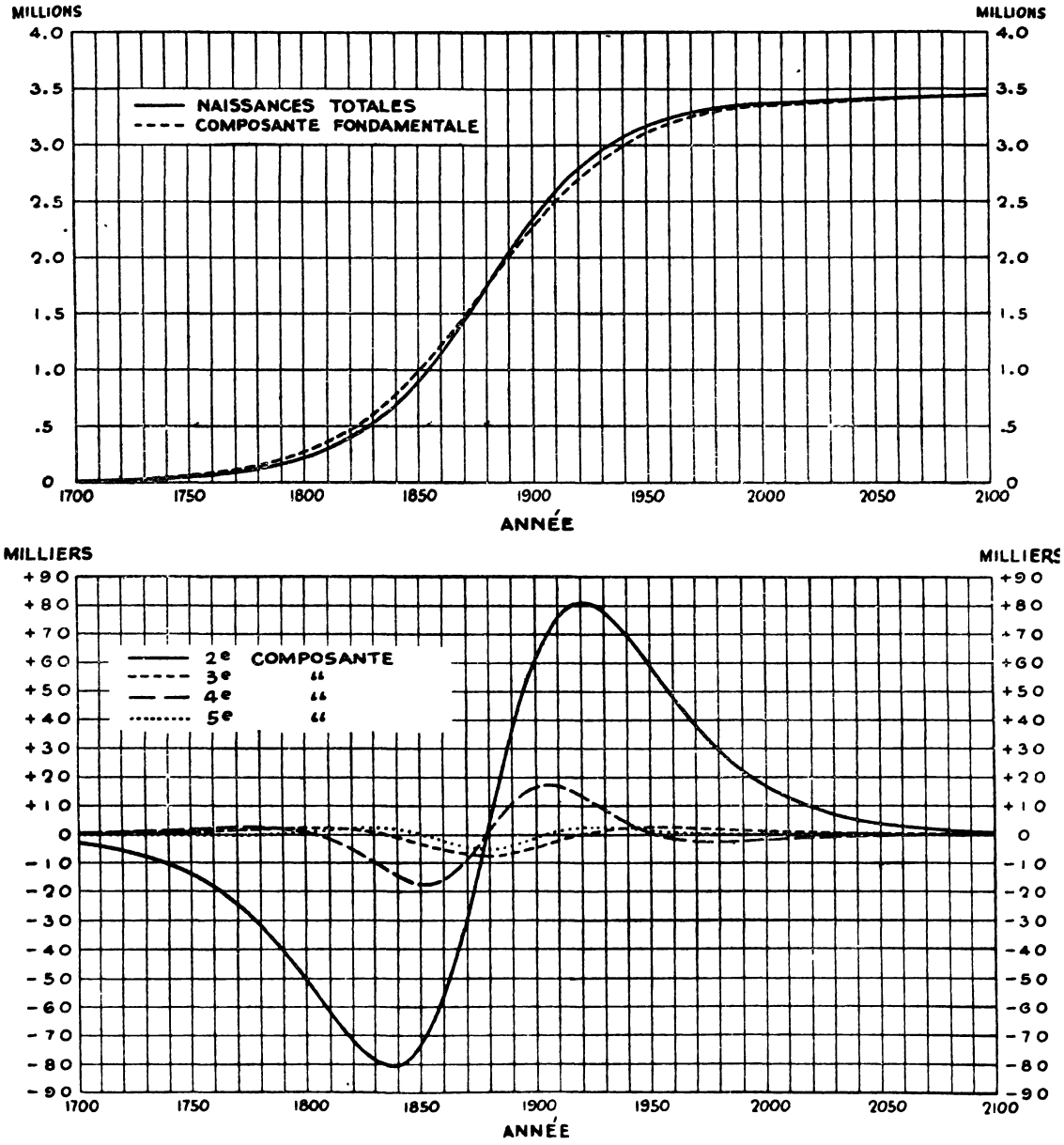

FIG. 1. — Naissances annuelles dans une population logistique. — Courbe calculée d'après les caractéristiques de la population des États-Unis et la table de survie de 1919-1920. (Mortalité invariable.)

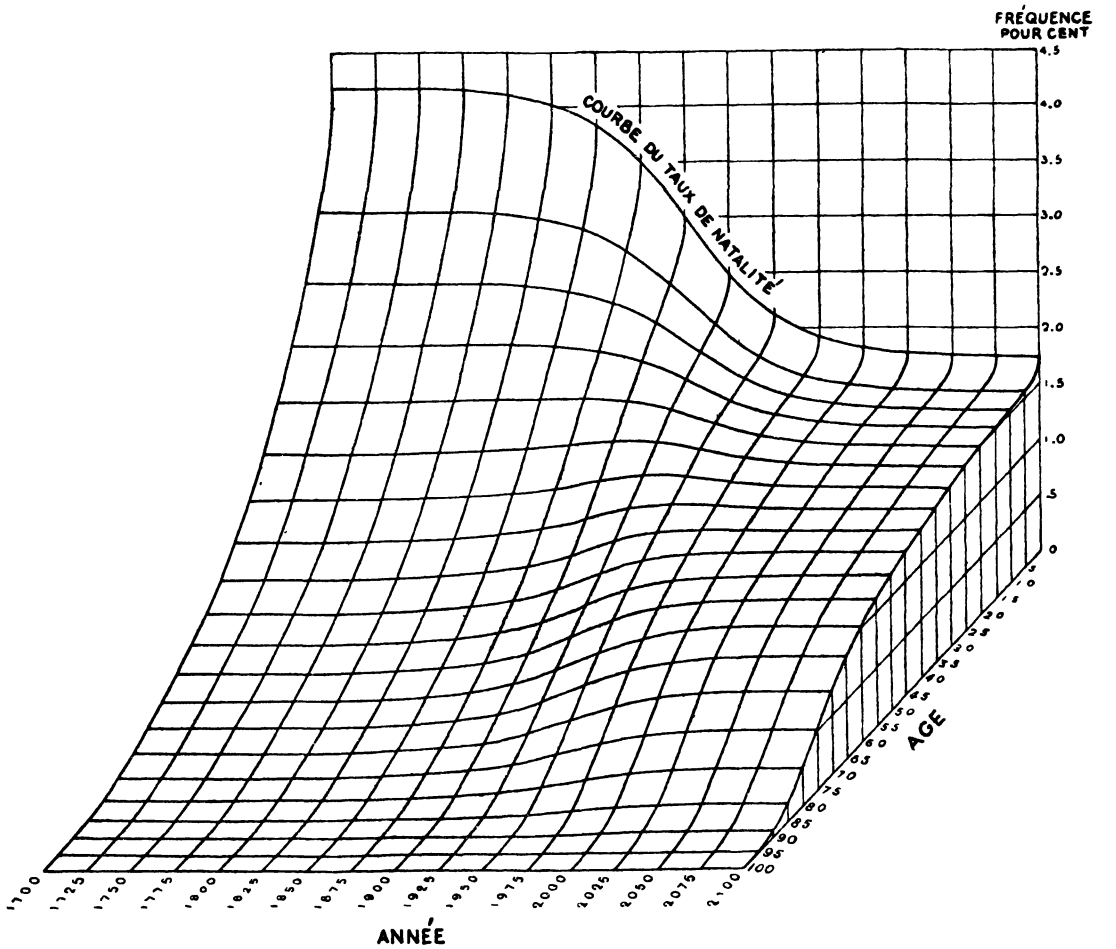

Fig. 2. — Répartition par l'âge dans une population logistique. — Modèle calculé d'après les caractéristiques de la population des États-Unis et la table de survie de 1919-1920. (Mortalité invariable.)

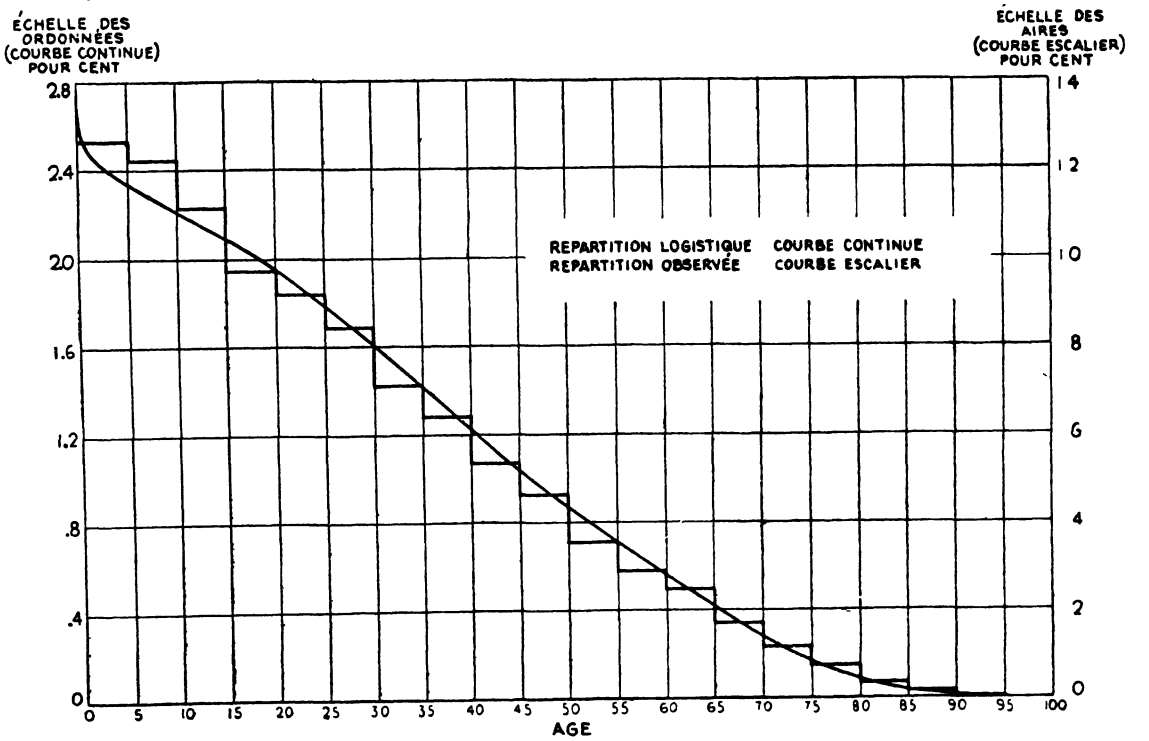

Fig. 3. — Répartition par l'âge dans une population logistique. — Calcul basé sur une série de tables de survie (mortalité variable avec le temps). — Résultats du calcul (courbe continue) comparés avec la répartition effectivement enregistrée aux États-Unis (1920), population indigène, race blanche, sexe féminin (courbe escalier).