

DANIEL SCHWARTZ

A. BOYCE

Le moment de la naissance et le moment de la mort. Analyse biométrique de la durée de gestation et de la durée de vie humaine, le point de vue du biologiste et le point de vue du statisticien

Journal de la société statistique de Paris, tome 116 (1975), p. 2-23

http://www.numdam.org/item?id=JSFS_1975__116__2_0

© Société de statistique de Paris, 1975, 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/>

COMMUNICATIONS

LE MOMENT DE LA NAISSANCE ET LE MOMENT DE LA MORT**Analyse biométrique de la durée de gestation et de la durée de vie humaine
le point de vue du biologiste et le point de vue du statisticien**

(Communication faite le 17 octobre 1974 devant la Société de statistique de Paris)

The variability of the human life span within a population can either be expressed in terms of the distribution of this duration (the statistician's approach) or in terms of functions which are derived from this distribution such as the force of mortality or the expectation of life (the biologist's approach). This second approach is by far the richer in its interpretation. The authors propose to apply this method to the study of other durations, and in particular to that of gestation. Then, returning to the statistical point of view, they propose a distribution model for the duration of life which leads them finally to a comparative assessment of the two points of view in the study of the durations

Die Unterschiede in der Lebensdauer einer gegebenen Bevölkerung können ausgedrückt werden unter den Gesichtspunkten der Statistik und der Biologie. Der zweite Gesichtspunkt gestattet eine grossere Anzahl von Erklärungen. Die Verfasser schlagen deshalb vor ihn für andere Zeitdauern zu verwenden, besonders für die Dauer der Schwangerschaft. Sie kommen dann auf den statistischen Gesichtspunkt zurück um ihn auf die Lebensdauer anzuwenden und schlagen dafür ein Verteilungsgesetz vor. Dieses Verteilungsgesetz führt sie zum Schluss zu einer vergleichenden Bilanz der zwei Gesichtspunkte für das Studium der Dauer eines Phänomens.

La variación de las duraciones de la vida humana en una población puede expresarse de dos maneras : estadística o biológicamente. Este segundo punto de vista permite interpretaciones mucho más ricas. Los autores proponen, pues, utilizarlo, para otras duraciones, especialmente la duración de la gestación. En cuanto al punto de vista estadístico para las duraciones de la vida, proponen una ley de distribución, lo que les conduce finalmente a un balance comparado de los dos puntos de vista en el estudio de la duración de un fenómeno.

Le but de ce travail, au titre quelque peu long et composite, est de présenter sur deux exemples — la durée de gestation et la durée de vie — deux manières d'interpréter la distribution de certaines variables, en fait essentiellement des durées, et de poser à cette occasion quelques questions sur le sens qu'on peut accorder à une distribution.

Nous définirons ces deux types d'interprétation à propos de la durée de vie humaine, après un bref rappel des données et indices classiques.

La mortalité est étudiée à partir de tables, dont il existe différents types; le plus courant est la *table du moment* qu'on peut ainsi définir très schématiquement : on constitue une génération fictive, comprenant au départ un nombre donné de sujets, disons 10 000, nés au *moment* considéré, on leur fait parcourir les âges successifs en appliquant à chaque âge les quotients annuels de décès observés dans la population étudiée, au moment considéré, ceci jusqu'à extinction : c'est ce que représentent, par exemple, les deux premières colonnes de la table 1 pour le sexe féminin, en France, en 1969, d'après Vallin [1].

TABLE 1

Fonctions déduites de la table de mortalité du moment

(France, 1969, sexe féminin)

*Calculs d'après les tables de Vallin [1] pour des intervalles $\Delta x = 5$ ans
(sauf 85 — 89)*

Début de l'intervalle x en années révolues	S	$-\Delta S = [\theta] \Delta x$	$[\mu] = -\frac{\Delta S}{S \Delta x}$	$[E]$	$x + [E]$
0	10 000	200		74,99	74,99
5	9 800	17	0,00035	71,47	76,47
10	9 783	14	0,00029	66,59	76,59
15	9 769	27	0,00055	61,68	76,68
20	9 742	32	0,00066	56,84	76,84
25	9 710	36	0,00074	52,02	77,02
30	9 674	44	0,00091	47,20	77,20
35	9 630	71	0,00147	42,40	77,40
40	9 559	105	0,00220	37,70	77,70
45	9 454	168	0,00351	33,09	78,09
50	9 288	230	0,00495	28,64	78,64
55	9 058	328	0,00724	24,30	79,30
60	8 780	491	0,01125	20,12	80,12
65	8 239	757	0,01838	16,17	81,17
70	7 482	1 172	0,03133	12,55	82,55
75	6 310	1 629	0,05163	9,42	84,42
80	4 681	1 878	0,08024	6,83	86,83
85	2 803	1 358	0,12112	4,73	89,73
89	1 445	1 445		3,30	92,30

La distribution des durées de vie dans la génération peut donner lieu à l'élaboration de divers indices :

1° La *fonction de survie* $S(x)$ est le nombre de survivants à l'âge x , soit multipliée par 10 000, la probabilité de survivre jusqu'à cet âge.

2° La *fonction de distribution* $\theta(x)$ est, multipliée par 10 000, la densité de probabilité de décès à l'âge x , ou encore l'ordonnée de la courbe de distribution; elle est égale au nombre de décès ($-\Delta S$) dans l'intervalle dx au voisinage de x , divisée par dx , soit :

$$\theta = -\frac{dS}{dx} \quad (1)$$

θ est donc la dérivée, changée de signe, de S .

3° Le *taux instantané de mortalité* $\mu(x)$ est la densité de probabilité de décès à l'âge x pour un survivant à cet âge; c'est donc le quotient de la densité ($\theta/10\ 000$) par la probabilité de survivre à l'âge x ($S/10\ 000$) :

$$\mu = \frac{\theta}{S} = -\frac{1}{S} \frac{dS}{dx} \quad (2)$$

μ est la dérivée logarithmique, changée de signe, de S .

4° L'*espérance de vie* $E(x)$ est le temps moyen qui reste à vivre aux survivants à l'âge x

$$E(x) = \frac{\int_0^{\infty} x\theta(t)(t-x) dt}{S(x)} \quad (3)$$

Souvent aussi on considère un 5^e indice, la *vie moyenne* des survivants à l'âge (x), soit $V(x) = x + E(x)$.

Ces indices sont généralement remplacés par des valeurs approchées, correspondant à des intervalles Δx finis (1 an, 5 ans...); à la table 1 figurent ainsi pour θ , μ et E les valeurs approchées $[\theta]$, $[\mu]$ et $[E]$, calculées à partir de $\Delta x = 5$ ans :

- $[\theta] = -\frac{\Delta S}{\Delta x}$; en fait on donne $[\theta] \Delta x = -\Delta S$, nombre de décès dans les 5 ans, et ordonnée de l'histogramme correspondant à cet intervalle de classe,

- $[\mu] = \frac{[\theta]}{S} = -\frac{1}{S} \frac{\Delta S}{\Delta x}$, qui est pour un sujet d'âge x , une approximation de la probabilité de mourir dans l'année (approximation parce que calculée sur 5 années); c'est le 1/5 du classique quotient quinquennal de mortalité ${}_5q_x$.

- $[E]$ est la moyenne arithmétique des durées de vie, calculées avec groupement par classes de 5 ans (pour la classe débutant à 89 ans, le centre de classe a été estimé d'après la table complète de Vallin).

Les 4 fonctions ont l'aspect bien connu indiqué figure 1, dont seule nous intéressera pour la suite la partie correspondant à l'état adulte (disons après 20 ans).

Les 4 fonctions S , θ , μ et E présentent entre elles des relations permettant de déduire l'une en fonction de l'autre et, calculées sur les mêmes données, elles traduisent toutes le même phénomène; mais elles le traduisent chacune à sa manière, et on peut centrer préférentiellement l'intérêt sur l'une ou l'autre. C'est ici que, parmi les choix adoptés, deux attitudes émergent d'emblée, que Teissier [2] appelait le point de vue du « statisticien » et celui du « biologiste », en les définissant ainsi : le statisticien centre son intérêt sur θ , constatant simplement que la longévité est, comme tout caractère biologique, entachée de variabilité : autour d'une longévité moyenne, dont la valeur n'est pas à expliquer, les événements fortuits entraînent des fluctuations dont la fonction θ donne l'entière description. Le biologiste, refusant de se borner à ce constat de variabilité, veut voir dans une des autres fonctions, essentiellement μ , un indice dont la variation avec l'âge reflète, au niveau de l'individu, la marque du vieillissement, et pourrait permettre d'en chercher une explication.

Les qualifications de « statistique » et « biologique » appellent certes des réserves, que nous développerons plus loin. Mais nous les conserverons provisoirement pour définir deux démarches dont l'aspect, à première vue du moins, paraît radicalement différent.

Le plan de notre exposé sera le suivant : 1° nous montrerons d'abord, dans le domaine classique de la mortalité, la richesse d'interprétation qu'a connue l'attitude du « biologiste »;

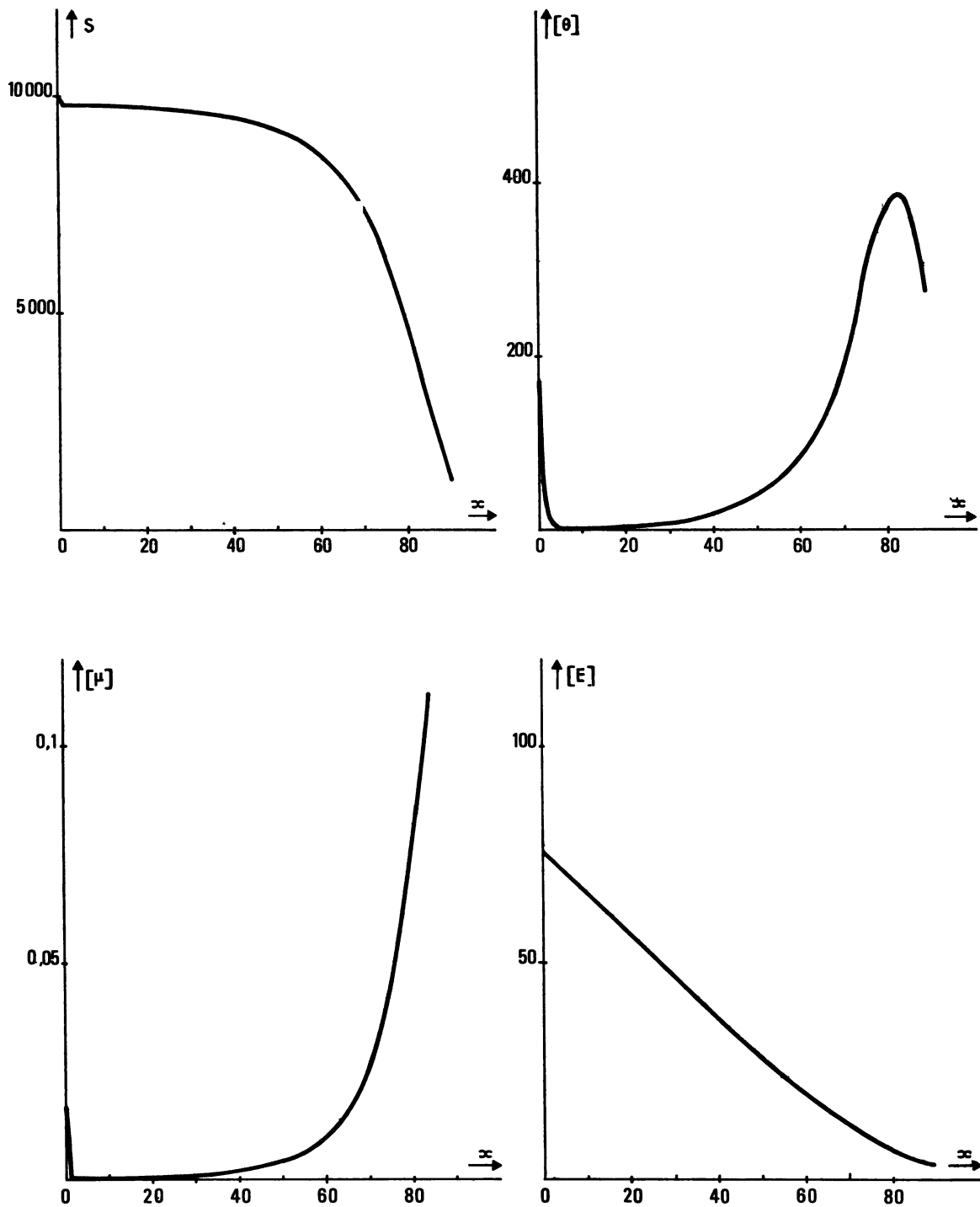


FIG. 1. — Représentation des fonctions S , $[\theta]$, $[\mu]$ et $[E]$ pour le sexe féminin, France 1969, à partir de S , $[\theta]$, $[\mu]$ et $[E]$ pour des intervalles annuels donnés par les tables de mortalité de Vallin [1], en fonction de l'âge x en années.

2° nous tenterons ensuite d'étendre cette attitude à un domaine beaucoup moins classique, celui des durées de gestation ; 3° revenant à la mortalité, nous développerons le point de vue du « statisticien », et 4° nous procéderons finalement à une confrontation des deux attitudes.

*
* *

Le succès de l'attitude « biologique » en mortalité se comprend aisément.

L'indice μ a d'abord un sens concret : rappelons que μdx est la probabilité, pour un sujet d'âge x , de mourir dans l'instant dx ; cette probabilité, comme le montre la figure 1, croît constamment et de plus en plus fort avec l'âge, traduisant ainsi un déclin de la vitalité; aussi la fonction μ a-t-elle été appelée depuis longtemps « force de mortalité ».

Il apparaît ensuite à première vue que μ , en échelle logarithmique, est une fonction sensiblement linéaire de l'âge.

Enfin, ce modèle simple se traduit, sur le plan biologique, d'une manière également simple, car si :

$$\text{Log } \mu = qx + C^{\text{te}} \quad (4)$$

ou
$$\mu = e^{qx + C^{\text{te}}} = be^{qx} \quad (5)$$

on a
$$\frac{d\mu}{\mu dx} = q \quad (6)$$

c'est-à-dire qu'à chaque instant dx , la force de mortalité augmente d'une quantité relative constante, ou encore que la résistance de l'organisme s'affaiblit d'une fraction constante. C'est bien sûr un des modèles les plus simples du vieillissement; il avait été formulé par Gompertz dès 1825 [3].

Au niveau de la fonction de survie, puisque $d \frac{(\text{Log } S)}{dx} = -\mu$, on déduit de (5) :

$$\text{Log } S = -\frac{b}{q} e^{qx} + K \quad (7)$$

formule écrite classiquement par les actuaires, en posant $q = \text{Log } c$, $b = -\text{Log } g \text{ Log } c$, et $K = \text{Log } k$

$$S = kg e^{cx} \quad (8)$$

Cette « loi de Gompertz » donne un ajustement convenable — dans nos contrées et à notre époque — pour la gomme approximative 55-80 ans.

Cependant cette loi, supposée traduire le vieillissement, ne concerne que les décès en relation avec l'âge; Makeham a proposé dès 1860 [4] de faire intervenir en outre une probabilité de mort dont la densité, soit a , serait en première approximation indépendante de l'âge, pour des causes de décès telles que certaines maladies infectieuses, relativement fréquentes à cette époque, et éventuellement certains types d'accidents. Le taux instantané de mortalité s'écrit alors (en négligeant la probabilité de mort par causes simultanées) :

$$\mu = be^{qx} + a \quad (9)$$

d'où résulte, en posant $a = -\text{Log } s$, la formule également classique de Gompertz-Makeham :

$$S = ks^x g e^{cx} \quad (10)$$

qui donne un ajustement convenable pour la quasi-totalité de la vie adulte — disons de 20 à 80 ans : c'est ce qu'on peut voir, figure 2, pour les données de la table 1, où le coefficient a a été choisi au mieux par une méthode graphique ($a = 4 \cdot 10^{-4}$).

La formule de Gompertz-Makeham a fait l'objet de multiples utilisations et interprétations.

La première utilisation est bien sûr la prévision. Grâce à la qualité de son ajustement, la formule de Gompertz-Makeham a rendu à cet égard aux actuaires les plus grands services.

Le pas suivant peut être qualifié d'utilisation interprétative. Lorsqu'on veut étudier la variation de la mortalité en fonction du pays, de l'époque, ou de telle condition particulière, d'environnement, l'évidente utilité d'un modèle est de résumer les données de mortalité par un petit nombre de coefficients, ce qui entraîne à la fois une simplification et, éventuellement, un début d'interprétation. La loi de Gompertz-Makeham ne fait intervenir que 3 coefficients (le coefficient k précisant seulement l'effectif initial), que nous retiendrons sous la forme utilisée en (9), soit a , b et q .

Sans doute convient-il de traiter à part le terme de Makeham. Disons que le coefficient a , qui mesure le risque de morts « accidentelles » a grandement décliné au cours du temps : par exemple, en France et pour le sexe féminin, sa valeur a passé de 70 à 10 pour 10 000 par an, du début au milieu de notre siècle [5]; pour l'année 1969, d'après les données de Vallin évoquées ci-dessus, nous trouvons 4 pour 10 000. Ainsi a-t-on été amené à supposer que a tend vers 0, tendance qui s'accorderait bien avec la disparition graduelle des décès par maladies infectieuses.

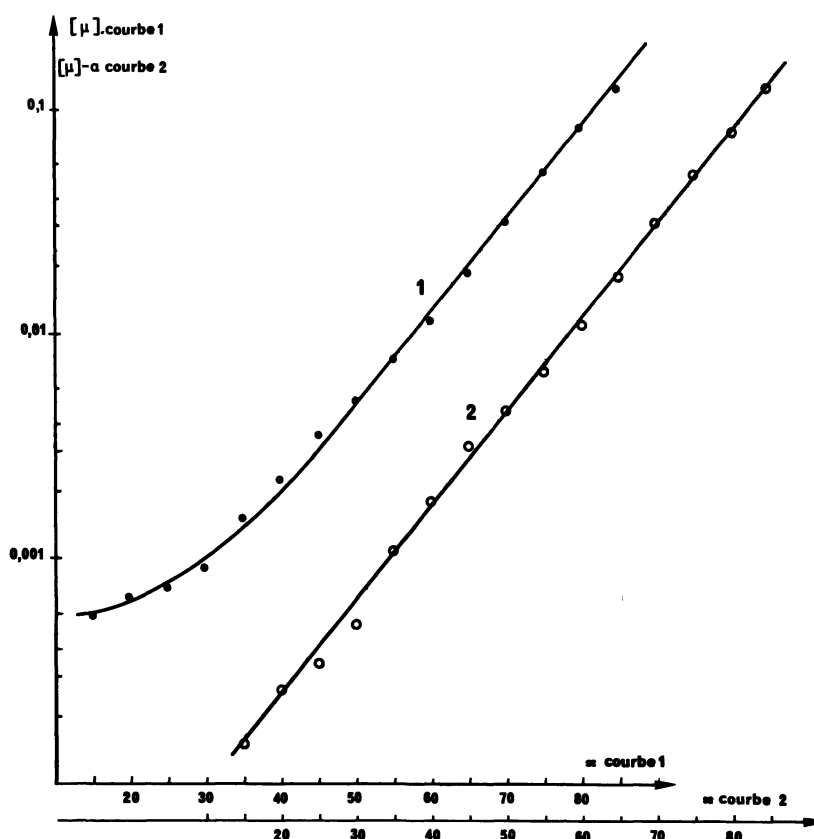


FIG. 2. — Variation de $[\mu]$ (courbe 1) et $[\mu]-a$ (courbe 2) en échelle logarithmique, en fonction de l'âge x en années, pour les données de la table 1 ($a = 4 \cdot 10^{-4}$).

Quant aux 2 coefficients de la loi de Gompertz, le « taux d'usure » q et b définissant un « état initial », leur variation dans le temps revêt un aspect qui au premier abord peut surprendre; alors qu'on aimerait gagner « sur les deux tableaux », c'est-à-dire diminuer à la fois q et b , les faits paraissent montrer qu'une diminution de b s'accompagne généralement d'une augmentation de q . C'est ce que constatent par exemple Redington [6] en comparant la mortalité en Grande-Bretagne au début et au milieu de ce siècle, et plus nettement encore Strehler et Mildvan [7] sur l'ensemble des contrées présentant des données valables : la corrélation entre b et q y apparaît fortement négative. En d'autres termes, si on revient à la droite exprimant $\text{Log } \mu$ en fonction de l'âge, disons pour simplifier la « droite de Gompertz », on doit conclure que plus son ordonnée à l'origine baisse, plus elle devient pentue.

Cette sorte de compensation n'est pas seulement « décevante », elle comporte aussi, au premier abord, une conséquence étrange : de deux « droites de Gompertz », celle qui démarre le plus bas est plus avantageuse que l'autre avant leur point d'intersection, mais plus désavantageuse après : si on gagne pour les jeunes, on perd pour les vieux, et inversement, ce qui rejoint à certains égards le paradoxe énoncé par Gombel [8]; cet auteur soutenait en effet, que des tables de vie défavorables ont un « dernier âge » plus élevé que les favorables et justifiait ainsi les âges élevés dans l'histoire des populations primitives.

Ces résultats doivent être rapprochés de considérations touchant, d'une part l'inégalité des gains possibles selon l'âge, d'autre part le concept d'une « vie limite ».

Les progrès de l'hygiène ont permis, dans nos contrées, une diminution de mortalité à chaque âge, mais qui affecte inégalement les classes d'âges : on le voit très clairement sur la figure 3, reproduite du travail de Sutter et Tabah [9], par la baisse graduelle des courbes de mortalité depuis le xviii^e siècle jusqu'à nos jours; dans la zone adulte qui nous intéresse, on aperçoit bien que le progrès, très considérable disons à 20 ans, diminue ensuite fortement avec l'âge, à tel point que le gain obtenu pour les âges extrêmes est quasi nul.

Le concept d'une durée de vie limite a été basé sur les considérations suivantes : dans la mesure où on remplace μ par son approximation établie sur 1 an $[\mu] = -\Delta S/S$ (qui n'est autre que le quotient de mortalité annuel ${}_1q_x$), l'âge ξ pour lequel $[\mu]$ atteint 1 est tel que la génération s'éteint dans l'année qui suit, de sorte que l'âge $(\xi + 1)$ n'est jamais atteint, et peut être considéré comme l'âge limite. Il suffit donc de porter les quotients annuels en échelle logarithmique en fonction de l'âge, de tracer au mieux la droite entre les points obtenus, et l'intersection avec l'ordonnée 1 donne ξ . Vincent [10] utilisant des données recueillies selon une méthode ingénieuse, dite des « générations éteintes », a pu, sur un effectif important de vieillards, d'âge au décès parfaitement connu, déterminer ainsi un âge limite, d'environ 110 ans, dont la caractéristique est d'être pratiquement le même pour les deux sexes, dans les différents pays d'Europe et les différentes époques de son étude. Ainsi était suggérée la notion d'une « limite théorique de la vie humaine ».

Si cette limite ξ est un véritable invariant, il faut en conclure que les « droites de Gompertz » convergent toutes au point d'abscisse ξ , d'ordonnée ${}_1q_x$ ou $[\mu] = 1$. Les progrès de l'hygiène, dans un même pays par exemple, se traduiraient par une droite, passant toujours par ce point, mais de plus en plus pentue. Ainsi s'expliqueraient, à la fois la baisse du gain possible avec l'âge, et l'évolution en sens inverse des deux coefficients de la loi de Gompertz.

Des recherches ont été menées avec une autre approche, basée sur la notion, introduite par Bourgeois-Pichat [11], de mortalités *exogène* et *endogène*, la première étant due aux facteurs du milieu et accessible à un effort de prévention, la seconde étant due à l'usure de l'organisme et difficilement modifiable dans l'état actuel des connaissances médicales. Pour évaluer la mortalité endogène, Bourgeois-Pichat compare différents pays et différentes

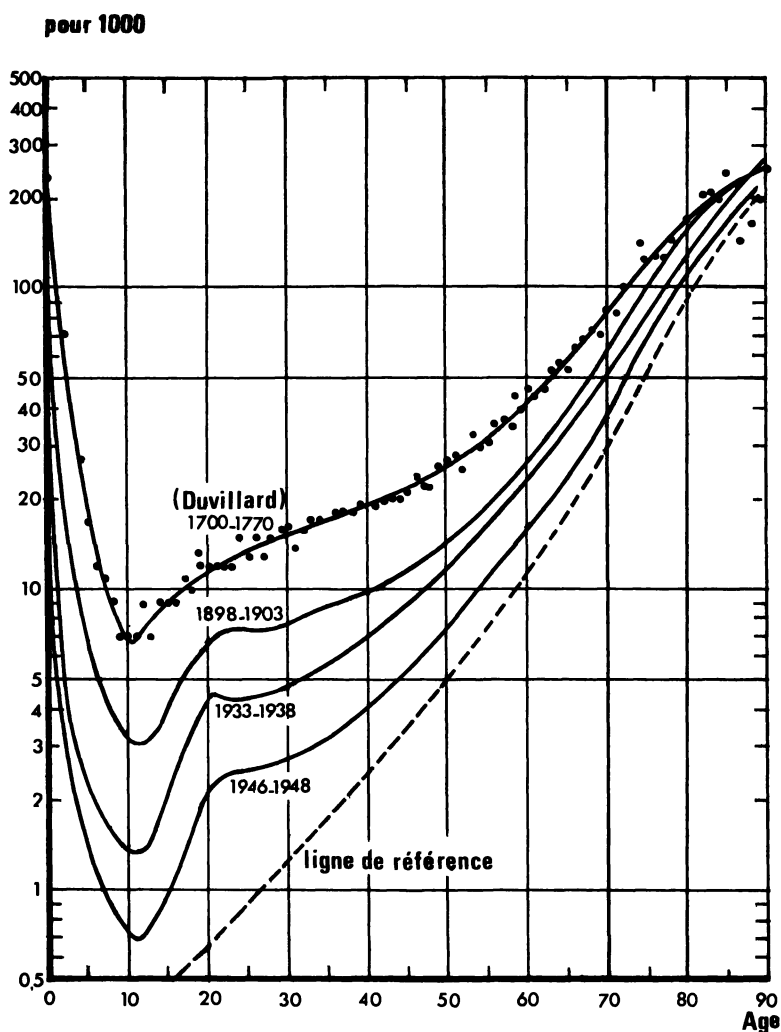


FIG. 3. — Variation du quotient de mortalité, en échelle logarithmique, en fonction de l'âge à diverses époques en France : 1700-1770 (table de Duvillard), 1898-1903, 1933-1938 et 1946-1948 (ensemble de la population). D'après Sutter et Tabah [9].

époques, et adopte la situation donnant la mortalité la plus basse (Norvège, époque récente) et en retire la mortalité pour certaines maladies infectieuses ou de l'appareil respiratoire. Il obtient ainsi une « mortalité biologique limite » pour laquelle les valeurs de $\text{Log} [\mu]$, de 30 à 85 ans donnent des points dessinant une droite, dont le prolongement après 85 ans se superpose assez bien avec le tracé de Vincent pour les grands vieillards.

En fait une partie de ces interprétations appelle des réserves, notamment le concept d'une durée de vie limite, ne fût-ce que parce qu'il repose sur l'approximation de μ par un quotient de mortalité calculé pour un délai fini, approximation très mauvaise aux âges extrêmes; $[\mu]$ ou ${}_tq_x$ est bien une probabilité, ne pouvant dépasser 1, mais μ est une densité de probabilité, qui peut dépasser 1, et qui tend vers l'infini dans la loi de Gompertz; cette loi n'admet pas de vie limite « théorique », mais seulement telle limite pratique qu'on définit

en fonction d'une probabilité minime choisie [12, 13]. Par ailleurs les faits même démentent le caractère invariant de la « vie limite », puisqu'un travail récent [14], conduit avec la méthode de Vincent montre qu'elle a augmenté de plus de 7 ans depuis le travail initial, correspondant à la période d'avant-guerre, et diffère maintenant de 2 ans selon les sexes.

Ce qui, en dépit de ces réserves, paraît bien émerger de cet ensemble de travaux, est la corrélation négative entre b et q , d'où résulte que le gain possible est d'autant plus faible que l'âge est plus élevé, la mortalité à chaque âge ayant peut-être — dans les conditions actuelles de vie — une limite inférieure déjà pratiquement atteinte chez les grands vieillards. Mais dire cela c'est admettre que la notion de vie limite, conceptuellement fautive, pourrait bien être « à peu près vraie »...

D'autres résultats concernent la variation des coefficients b et q selon l'environnement, en particulier l'exposition à des radiations. L'exposition continue d'animaux à des radiations de haute énergie tend à accroître q , ceci proportionnellement à la dose, tandis que l'exposition à une dose unique de radiations ionisantes n'affecte guère q , mais accroît b proportionnellement à la dose [15-17].

C'est cet ensemble de faits dont il faudrait tenir compte quand on arrive à la dernière étape concernant la loi de Gompertz, celle qui vise à « expliquer » pourquoi elle rend compte de la mortalité en fonction de l'âge.

Les « explications » sont multiples. Certaines, très simples, reprennent l'idée de Gompertz d'une vitalité décroissant d'une proportion constante par unité de temps, mais tâchent de donner au mot vitalité un contenu concret. Ainsi a-t-on donné une liste de fonctions physiologiques de l'homme qui varient avec l'âge selon une loi de Gompertz [7]. Il est facile, bien sûr, de supposer que l'altération de telle ou telle de ces fonctions, ou de leur ensemble, traduit le vieillissement et, à partir d'un seuil donné, la mort. Des hypothèses plus précises ont été formulées à propos de cellules vitales qui ne se reproduisent pas, celles du cerveau par exemple, et dont une proportion, croissante avec l'âge, pourrait présenter une altération au niveau des membranes [18]. Mais on a proposé aussi des modèles beaucoup plus élaborés, à mi-chemin entre les mathématiques et la biologie : Strehler et Mildvan [7] arrivent à la loi de Gompertz en supposant que le maintien en vie nécessite d'incessants efforts de restauration, dont l'ampleur est distribuée énergétiquement entre les molécules selon une loi de Maxwell-Boltzmann; Sacher et Trucco [19] représentent l'état de l'individu, décrit par n caractéristiques, par un point dans l'espace, supposent qu'il est soumis à un mouvement Brownien et qu'il y a mort s'il rencontre une certaine surface; Brillinger [20] démontre très élégamment que si un ensemble est formé d'éléments soumis à des distributions de décès identiques et indépendantes, et que la mort survienne au premier élément atteint, la mortalité de l'ensemble est régie par une famille de courbes dont la plus simple est celle de Gompertz.

Le fait que tel ou tel de ces modèles conduise à une loi de Gompertz ne suffit pas pour autant, bien sûr, à en prouver le bien-fondé. Un pas de plus peut être franchi si la théorie permet de prévoir les variations de b et q observées, telles que leur corrélation négative ou la relation avec une exposition aux radiations. C'est le cas, dans une certaine mesure, pour la théorie de Strehler et Mildvan [7]. Il reste que cet ensemble de théories prouve surtout que la loi de Gompertz peut résulter facilement d'une très grande variété de mécanismes.

L'approche « biologique », telle que nous l'avons définie au départ n'est d'ailleurs pas liée obligatoirement aux lois de Gompertz ou de Gompertz-Makeham, qui en ont été le point de départ. De nombreuses lois, autres que celles-ci, ont été proposées, notamment par Perks [21], visant à un meilleur ajustement. Elles n'apportent cependant qu'un progrès

modeste, au prix d'une complication qui fait perdre de vue le sens « biologique » de la loi. Mais leur postulat de base reste toujours l'existence, pour l'individu, d'une « force de mortalité ».

On peut encore ajouter que le concept de « force de mortalité » a été un outil de travail commode, en dehors de ces problèmes portant sur la mortalité générale, dans des domaines où on étudie une cause de décès donnée, suffisamment importante pour qu'on puisse négliger les autres pendant une période déterminée : c'est le cas, par exemple, dans des études d'accidents, et de mortalité de cancéreux, où on a admis comme modèle celui d'une force de mortalité constante.

Il convient enfin d'indiquer qu'à côté de la fonction μ , une autre fonction de type « biologique » a été un outil de travail très utile : c'est l'espérance de vie; soit du point de vue théorique — Teissier [2] a proposé un modèle — soit surtout du point de vue pratique : l'utilisation de l'espérance de vie à divers âges, selon que l'on désire inclure ou au contraire exclure la mortalité périnatale, infantile, juvénile, etc.. a connu les succès qui sont bien connus de tous.

*
* *

Revenons maintenant à l'autre extrémité de la vie. Chercher à « expliquer » la naissance, comme on l'a fait pour la mort, c'est tenter de savoir pourquoi se termine la grossesse; le biométricien se trouve, ici encore, confronté à l'étude d'une durée, la durée de gestation. On peut alors se demander ce que représentera, dans ce domaine, l'alternative entre les deux attitudes, décrite pour la durée de vie humaine.

La distribution des durées de gestation est bien connue; avec son allure de courbe en cloche dissymétrique, plus largement étendue vers la gauche, elle ressemble beaucoup à la distribution des durées de vie humaine. Il ne semble pas qu'on ait proposé, pour en rendre compte, de lois à la fois simples et réalisant un ajustement correct. C'est dire que l'attitude du « statisticien » n'a guère réussi. Mais il ne paraît pas qu'on ait, en contrepartie, tenté ici la démarche du « biologiste » avec ses indices familiers au démographe et les riches interprétations qu'ils ont suggérées pour la mort. Pourquoi ne pas le faire? Non pour le simple plaisir de rapprocher la naissance d'une mort, d'énoncer que la naissance est une fin de vie, la vie intra-utérine; mais parce qu'au-delà de ces mots il paraît logique d'appliquer les mêmes techniques à l'étude d'une même variable, la durée d'un phénomène.

Les données de départ sont des distributions de durée de gestation, cette durée étant comptée depuis les dernières règles jusqu'à l'accouchement. Nous ne considérons que les naissances, à l'exclusion des avortements, c'est-à-dire selon la définition usuelle, les durées de gestation d'au moins 28 semaines révolues, ce qui nous situe, comme pour la durée de vie humaine, dans la période adulte de la vie (fœtale). Pour établir, à partir d'une distribution de durée de gestation, l'équivalent d'une table de mortalité, il suffit de constituer une cohorte, en fixant arbitrairement, disons à 10 000, l'effectif au départ (28^e semaine révolue, jour 196) et en calculant, à partir de la distribution, les effectifs progressivement décroissants des femmes encore enceintes (c'est-à-dire n'ayant pas encore accouché) jusqu'à extinction de la cohorte.

Il faut insister d'emblée sur la piètre qualité de ces données de départ, comparée à celle des durées de vie humaine; en premier lieu la variable intéressante sur le plan biologique est, bien sûr, la durée comptée depuis la conception, tandis qu'on connaît seulement celle qui part des dernières règles, la différence entre les deux varie assez notablement et peut atteindre des valeurs élevées; en outre la date des dernières règles est sujette, dans l'inter-

rogatoire, à des erreurs multiples pouvant atteindre un mois. C'est dire d'emblée qu'on ne peut espérer obtenir des ajustements très corrects.

La table 2 donne les résultats correspondant à une des plus importantes séries dont nous disposons : celle de Bjerkedal et al. portant sur 125 485 naissances [22]. Nous présentons les données sous forme d'une « table de natalité », semblable à des tables de mortalité, avec mention des fonctions analogues S , θ , μ et E , ainsi que la « vie moyenne » $x + E$, qui n'est autre que la durée moyenne de gestation pour les femmes n'ayant pas encore accouché au temps x . Ces fonctions représentées graphiquement figure 4, ressemblent fortement à celles données figure 1 pour la mortalité, ce qui était prévisible vu la ressemblance des courbes de distribution (il y a toutefois, en fin de courbe, des différences sur lesquelles nous reviendrons). Il convient maintenant de tenter une interprétation des indices, à la lumière de celles proposées pour la mortalité.

TABLE 2

Table de « Natalité » et fonctions correspondantes pour la série de Bjerkedal et al. [22] portant sur 125 485 naissances non gemellaires

Calcul pour des intervalles $\Delta x = 7$ jours (sauf pour le premier)

Début de l'intervalle x		S	$-\Delta S = [\theta]\Delta x$	$[\mu] = -\frac{\Delta S}{S\Delta x}$	$[E]$	$x + [E]$
en semaines révolues	en jours					
28-31	196	10 000	55	0,0002	85,70	281,70
32	224	9 945	31	0,0004	58,10	282,10
33	231	9 914	42	0,0006	51,37	282,37
34	238	9 872	65	0,0009	44,47	282,47
35	245	9 807	119	0,0017	37,75	282,75
36	252	9 688	191	0,0028	31,18	282,18
37	259	9 497	361	0,0054	24,74	282,74
38	266	9 136	654	0,0134	18,60	284,60
39	273	8 282	1 969	0,0340	13,21	286,21
40	280	6 313	2 828	0,0640	9,39	289,39
41	287	3 486	2 072	0,0849	7,58	294,58
42	294	1 413	950	0,0960	7,30	301,30
43	301	463	275	0,0849	9,13	310,13
44	308	188	101	0,0767	11,10	319,10
45	315	87	87		13,50	328,50

1) Nous examinerons d'abord l'espérance E . L'examen, table 2, de la colonne $(x + E)$ montre que la durée de gestation, évaluée à la 28^e semaine — ou, ce qui revient au même, en début de grossesse si on ne tient pas compte des avortements — doit être augmentée de 2 jours pour les femmes encore enceintes au jour 259, c'est-à-dire ayant passé le cap des prématurés, de 7,7 jours au jours 280, qui est le centre de l'époque la plus probable, de 12,9 jours au jour 287. Ce jour-là, les femmes qui n'ont pas encore accouché, relativement nombreuses (35 % de l'effectif initial) ont largement dépassé les valeurs moyenne et modale et sont persuadées — leur accoucheur l'est aussi — que leur accouchement est plus qu'imminent; certes le jour le plus probable est ce jour même, mais sa probabilité n'est pas forte, et la durée qui reste à couvrir est bel et bien (voir colonne E) de 7,6 jours. C'est ici qu'on mesure la différence d'état d'esprit avec le domaine de la mortalité : une femme française en 1969, âgée de 85 ans, a aussi dépassé la moyenne et le mode; or on retient, non que l'année

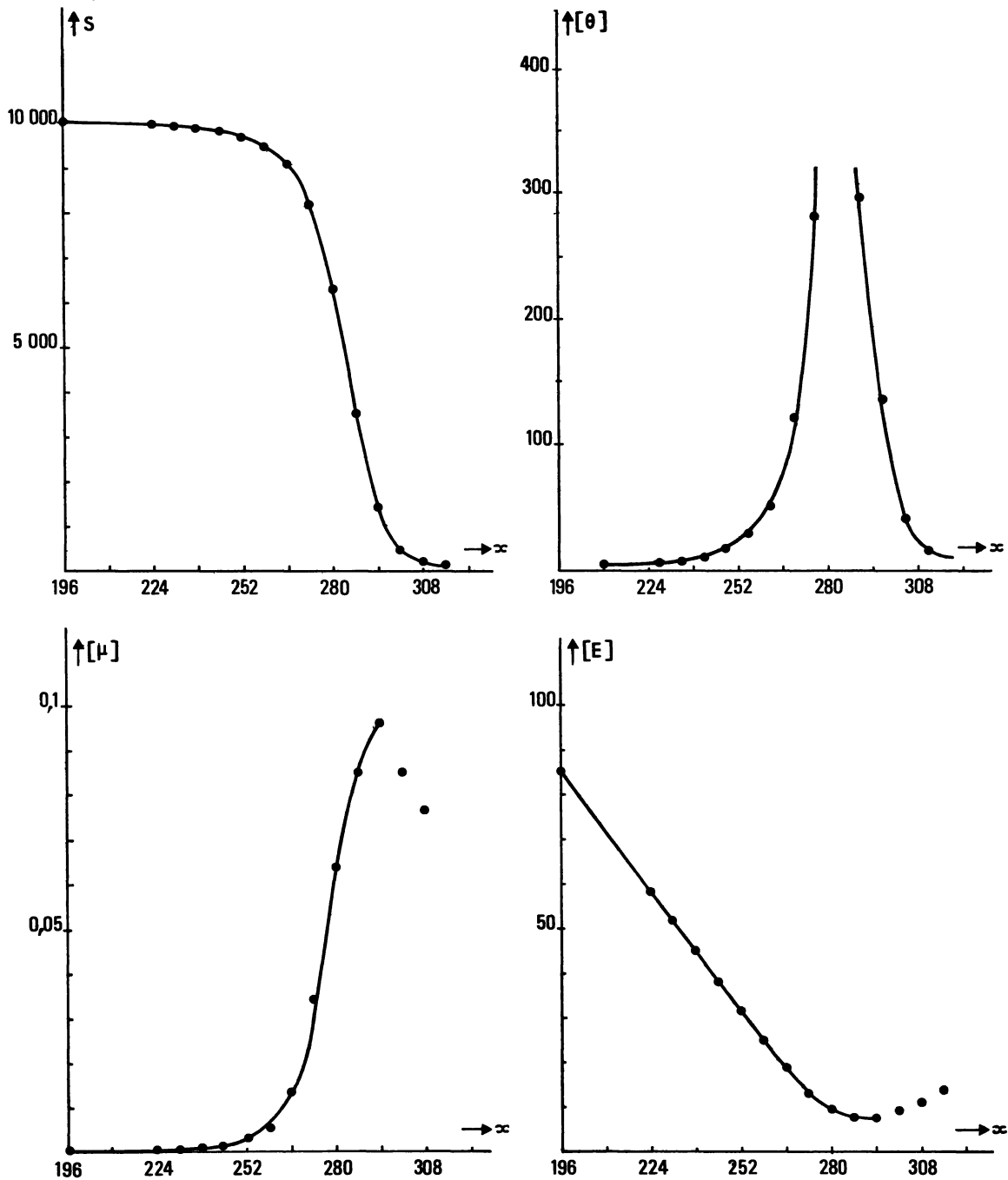


FIG. 4. — Représentation des fonctions S , $[\theta]$, $[\mu]$ et $[E]$ dans la série de 125 485 naissances de Bjerkedal et al. [22] à partir de S , $[\theta]$, $[\mu]$ et $[E]$ pour des intervalles de 7 jours, d'après la table 2, en fonction de la durée de gestation x en jours.

de décès la plus probable est celle qui vient, mais que son espérance de vie est encore de 4,7 ans. Sans doute la femme est-elle plus pressée d'accoucher que de mourir, mais on ne voit pas que ce soit là motif suffisant pour délaissier la moyenne au profit du mode...

La considération de l'« espérance de vie *in utero* » ou de la durée moyenne de gestation ($x + E$) mérite des commentaires :

a) La croissance progressive de la durée de gestation en fonction du temps x traduit le fait mathématique évident qu'on ampute de plus en plus la courbe de distribution de durées courtes. Ce fait, bien connu pour la vie humaine, gagnerait à l'être pour la durée de gestation. On éviterait ainsi des erreurs d'interprétation ou des biais. On entend dire, par exemple, que les femmes qui n'accouchent pas d'un prématuré ont une durée de gestation « plus grande que les autres » : plus grande que la moyenne générale, qui comprend les prématurés, certes, rien de plus. De même le « retard » des femmes ayant dépassé le mode peut être considéré, mais en se souvenant quand même que la moyenne et le mode sont établis sur une distribution qui tient compte de ces retards. Ainsi que le retard de femmes venues pour un examen biologique donné qui peut s'expliquer par l'exclusion des femmes ayant accouché avant cet examen, etc.

b) Le recul progressif de la date moyenne d'accouchement à mesure que la grossesse se poursuit pourrait intervenir dans la prédiction de cette date, soit pour la patiente, soit statistiquement pour une planification des lits dans les maternités.

c) Enfin l'espérance E à un moment donné pourrait, comme l'espérance de vie, être un moyen commode pour comparer des séries dans l'espace et dans le temps. Ainsi pourrait-on utiliser, selon le cas, l'espérance à 28 semaines, ou bien à 37 après le cap des prématurés ou plus tard encore...

d) Sans doute la série présentée comporte-t-elle une fréquence élevée de grossesses très tardives, dont beaucoup proviennent sûrement d'erreurs. Mais il en est de même dans les autres séries dont nous disposons, et c'est sur de telles séries que reposent les évaluations classiques de la durée de gestation.

De toute manière nous donnons, table 3, la « table de natalité » après élimination des accouchements postérieurs à la 42^e semaine révolue. On vérifie bien sûr que la variation de ($x + E$) tout au long de la grossesse conserve le même aspect.

TABLE 3

Table de « Natalité » et fonctions correspondantes pour la série de Bjerkedal et al. [22] après suppression des cas postérieurs à 42 semaines révolues

Début de l'intervalle x		S	$-\Delta S = (\theta) \Delta x$	$[\mu] = -\frac{\Delta S}{S \Delta x}$	$[E]$	$x + [E]$
en semaines révolues	en jours					
28-31	196	10 000	57	0,0002	84,32	280,32
32	224	9 943	33	0,0005	56,72	280,72
33	231	9 910	44	0,0006	49,90	280,90
34	238	9 866	68	0,0010	43,11	281,11
35	245	9 798	125	0,0018	36,39	281,39
36	252	9 673	200	0,0030	29,82	281,82
37	259	9 473	379	0,0057	23,39	282,39
38	266	9 094	895	0,0141	17,24	283,24
39	273	8 199	2 065	0,0360	11,79	284,79
40	280	6 134	2 965	0,0891	7,75	287,75
41	287	3 169	2 173	0,0980	5,20	292,20
42	294	996	996	—	3,00	297,00

2) Centrons maintenant notre attention sur μ , comme on l'a fait pour la mortalité; fonction que nous appellerons, par pure analogie de mots, « force de natalité »; et que nous porterons, ici encore, en échelle logarithmique. La courbe est tracée figure 4. On aperçoit d'abord une notable différence avec la mortalité en fin de course, comme on l'a déjà souligné pour l'espérance E : la courbe devient stationnaire, et même décroît. Cette singularité résulte certainement, pour une grande part, d'erreurs de déclaration dans la date des dernières règles ou d'ovulations tardives. Y a-t-il, en plus des raisons précédentes, un certain abaissement de la « force de natalité » chez les enfants naissant tard? De toute manière, quelle que soit l'interprétation, le seul fait que la courbe de μ — comme celle de E — mettent clairement en évidence une singularité qui n'apparaît pas dans la courbe de distribution θ , montre l'intérêt que présente l'examen de ces courbes.

Toutefois, pour la recherche d'un modèle, il est plus prudent de ne pas tenir compte des points correspondant aux cas tardifs.

On obtenait, en mortalité (fig. 2) un bon ajustement linéaire pour $\text{Log } \mu$ de 55 à 80 ans, et sur une gamme plus large pour $\text{Log } (\mu - a)$, après choix de la meilleure valeur pour a . On perçoit de même un certain alignement, figure 5, pour $\text{Log } \mu$ de la 37^e à la 39^e semaine, et pour $\text{Log } (\mu - a)$ de la 33^e à la 40^e, après choix de la meilleure valeur $a = 3 \cdot 10^{-3}$ (1). Ces alignements sont moins nets que pour la mortalité, et portent sur un nombre de points plus faible — mais avec des intervalles de temps aussi grands, et des données dont nous avons souligné la piètre qualité, pouvait-on attendre mieux? Aussi nous bornerons-nous à dire qu'un ajustement au modèle de Gompertz-Makeham « n'est pas impossible », nous gardant bien de dire qu'il soit, non seulement prouvé, mais même fortement suggéré par la figure.

Rien n'empêche d'ailleurs de supposer le modèle inadéquat, et d'envisager une autre loi pour $\text{Log } \mu$, par exemple une loi sigmoïde, comme c'est le cas pour beaucoup de caractéristiques du fœtus, à commencer par sa taille. Un aspect intéressant, dans cette hypothèse, serait le moment du point d'inflexion qui paraît être, figure 5, au voisinage de la 36^e semaine révolue, c'est-à-dire près du seuil de prématurité.

Il ne paraît pas possible d'aller plus loin. Nous nous contenterons de suggérer une discussion sur le concept de « force de natalité », et plus généralement sur l'emploi en natalité de techniques ayant conduit, en mortalité, à un riche faisceau d'hypothèses. Y a-t-il aussi, pour la natalité, des causes « exogènes » et « endogènes » les premières pouvant être peu à peu éliminées tandis qu'il subsisterait, même pour les accouchements très précoces, une proportion de cas d'origine endogène, représentant un « taux de natalité limite » qu'on ne saurait réduire dans l'état actuel des connaissances? La natalité « endogène » obéit-elle à un modèle « invariant » dans l'espèce humaine, ou dans la négative de quels facteurs et comment en dépend-elle?

* * *

Après avoir montré l'intérêt de l'attitude du « biologiste » en mortalité, et l'avoir proposée pour la natalité, nous allons maintenant, revenant à la mortalité, y développer le point de vue du « statisticien ».

Un simple coup d'œil, figure 1, sur l'allure de la courbe $\theta(x)$, en cloche dissymétrique s'étendant plus longuement vers la gauche (mode supérieur à la moyenne) écarte l'idée

1. Ou $4 \cdot 10^{-4}$ quand l'unité est le jour (fig. 5).

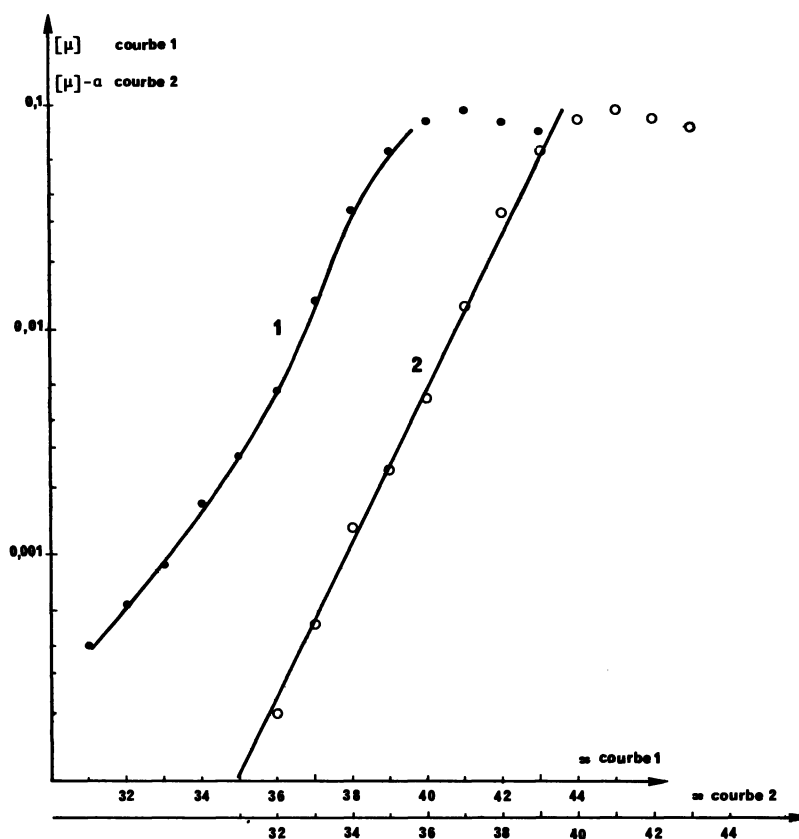


FIG. 5. — Variation de $[\mu]$ (courbe 1) et $[\mu]-a$ (courbe 2), en échelle logarithmique, en fonction de la durée de gestation x en jours pour les données de la table 2 ($a = 4 \cdot 10^{-4}$).

des lois les plus classiques (normale, log normale). Des lois plus complexes doivent alors être envisagées; d'illustres statisticiens s'y sont essayé, ne fût-ce que Karl Pearson, et de nombreux modèles ont été proposés; lois de Pearson de différents types, mélanges de plusieurs lois [23-27], dont le but a été le plus souvent d'obtenir un ajustement correct. Dans certains cas, mais peu fréquents, la loi est suggérée par des hypothèses biologiques, le plus souvent d'ordre génétique, la distribution des longévités reproduisant, d'une manière ou de l'autre, une variabilité génétique. La théorie la plus élaborée est sans doute celle de Szilard [28] pour qui la mort résulte de l'altération d'une proportion déterminée de chromosomes; ceux-ci pouvant être altérés, soit en fonction de l'âge avec une probabilité constante par unité de temps, soit au départ dans le patrimoine génétique; il y aurait alors un âge maximum, pour l'individu sans altération au départ, et des raccourcissements proportionnels au nombre initial de chromosomes lésés, qui suivrait une loi de Poisson. Szilard a ajouté en plus un certain nombre de raffinements permettant un meilleur ajustement aux données observées.

Il ne semble pas qu'on ait proposé jusqu'ici de loi fournissant un ajustement correct et qui soit en même temps assez simple pour étayer une approche d'« explication », double qualité qui a fait le succès de la loi de Gompertz.

Nous suggérons, dans cette optique, la prise en considération d'un modèle relativement simple; la distribution des décès a manifestement l'aspect de la courbe symétrique d'une

loi log normale. On peut donc se demander quelle qualité d'ajustement apporterait une distribution de type $\log(l - x)$; l serait alors un âge limite, et c'est l'écart à l'âge limite qui serait distribué selon une loi log normale.

Pour juger de l'adéquation de ce modèle, sur les données de la table 1, par exemple, on peut d'abord chercher une confirmation rudimentaire comme suit : laissant de côté les décès de l'enfance et nous bornant à la période adulte, on fera démarrer la cohorte, disons à 20 ans, ce qui revient à ramener le nombre de survivants à cet âge à 10 000, d'où la table de vie :

âge x	20	25	30	35	...
S	10 000	9 967	9 930	9 885	...

Pour une valeur donnée de l , mettons $l = 100$ ans, on porte sur un graphique, en regard des abscisses $(l - x)$ de chaque âge, en échelle logarithmique, le pourcentage de survivants S ci-dessus, en échelle probit. Si $\text{Log}(l - x)$ est distribué normalement, les points doivent être sensiblement alignés. Les résultats sont donnés, figure 6, pour différentes valeurs de l . On observe que le tracé, légèrement convexe pour $l \leq 100$ et concave pour $l \geq 110$, est très proche d'une droite pour $l = 100$ et 110, et qu'il doit y avoir une valeur optimum, entre 100 et 110, donnant un ajustement linéaire très correct.

Ce résultat justifie de rechercher une confirmation plus poussée par une méthode plus rigoureuse, qu'on peut ainsi conduire : la table de vie ci-dessus donne une distribution

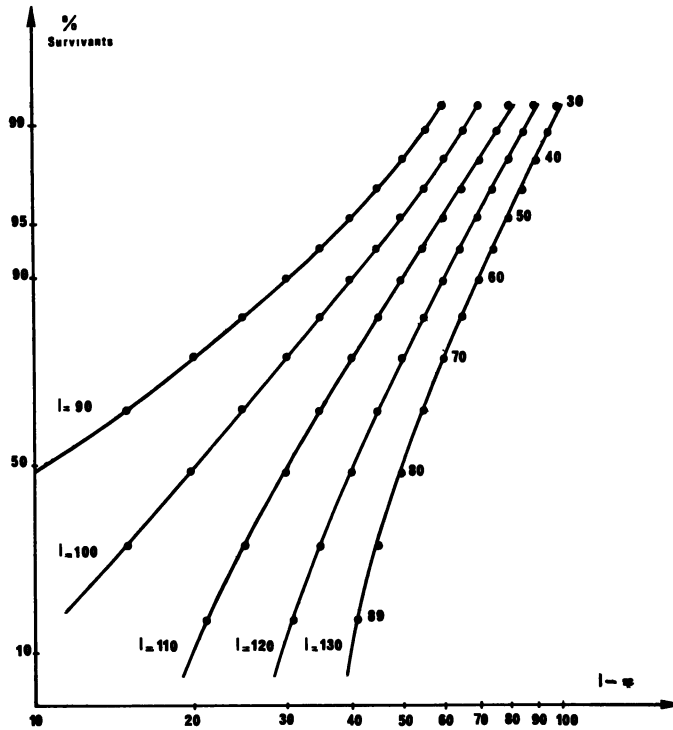


FIG. 6. — Pourcentage de survivants (en échelle probit) en fonction de l'écart $(l - x)$ à l'âge limite (en échelle logarithmique) pour différentes valeurs de l . Au-dessus des points figure l'âge x . La génération étudiée est celle de la table 1.

observée des décès, et il est possible de trouver, par une méthode d'optimisation, la loi log normale de $(1 - x)$ s'ajustant au mieux à la distribution observée, compte tenu de la « troncature » au-dessous de 20 ans (effectif inconnu) et de la « censure » au-dessus de 89 (les sujets sont répartis en une classe unique).

Cette démarche est possible sur ordinateur avec un programme convenable. Nous la menons actuellement (1), et bien que le travail ne soit pas terminé, il semble que la qualité de l'ajustement soit du même ordre, pour la gamme d'âges considérée, qu'avec la formule de Gompertz-Makeham. Le modèle proposé est donc relativement satisfaisant.

On notera que le modèle n'a pas distingué les mortalités « accidentelle » et « de vieillissement ». Mais la mortalité « accidentelle » de Makeham est pratiquement négligeable dans la table de vie considérée (on a vu que $a = 4 \cdot 10^{-4}$); le programme permet d'ailleurs de prendre en compte une mortalité de taux constant avec l'âge, a , et de l'estimer au mieux; dans le cas présent on trouve une valeur très voisine de celle ci-dessus ($a = 5 \cdot 50^{-4}$).

Il convient maintenant de voir, comme on l'a fait pour la loi de Gompertz-Makeham, quel parti peut être tiré du modèle pour l'utilisation et l'interprétation.

L'utilisation pour la prédiction ne pose pas de problème, puisque l'ajustement est satisfaisant.

Le pas suivant, que nous avons qualifié d'utilisation interprétative, est l'étude des paramètres du modèle en fonction du pays, de l'époque, ou des conditions d'environnement. Elle mérite d'être faite; les paramètres en cause : vie limite, moyenne et variance de la loi log normale, éventuellement taux d'accidents, ont un sens au moins aussi clair que ceux de la loi de Gompertz-Makeham.

La dernière étape, enfin, est celle de l'interprétation : le modèle permet-il, dans une certaine mesure, de suggérer des hypothèses sur les mécanismes du vieillissement et de la mort? Pour une telle démarche, le « statisticien », avec sa distribution θ des longévités est moins bien placé que le « biologiste », dont la fonction μ chiffre l'évolution d'un processus au niveau de l'individu. La démarche, cependant, mérite d'être tentée, et nous avons vu Szilard [28] déduire la distribution des décès de théories basées sur la proportion de chromosomes altérés à chaque âge.

La première conclusion à tirer du modèle est celle d'une durée de vie limite l . La notion d'une limite *stricte* est certes contestable, ne serait-ce que parce qu'il existerait alors, comme disait Feller [11], des individus atteignant l'âge $(l - \varepsilon)$ qui auraient la certitude de mourir dans le délai ε . Dans le modèle de Gompertz au contraire, la durée de vie limite, quoique parfois invoquée, n'existe pas. C'est là une différence, sur le plan conceptuel, fondamentale. Mais les développements relatifs à la loi de Gompertz nous ont conduits à suggérer que la notion de vie limite pourrait être « à peu près vraie ». Et dans les modèles « statistiques », certains, comme celui de Szilard, font état d'une vie limite, en ajoutant quelques raffinements qui lui retirent son côté de « limite stricte ». Nous admettrons donc, puisque notre modèle rend assez bien compte de la réalité, que « tout se passe comme si », en première approximation il y avait une vie limite l , et une distribution log normale de $(1 - x)$.

Nous nous bornerons alors, pour la suite de l'interprétation, à un abord simple — simpliste peut-être. Puisque $\log(1 - x)$ est distribué normalement, nous nous demanderons s'il est possible d'accorder, au fait de la distribution normale, une « signification ». C'est là un si vieux débat que nous osons à peine l'aborder.

Pour certains, l'adéquation à une loi normale, ou log normale, ne saurait avoir aucune valeur explicative, car cette adéquation, de moins en moins bonne à mesure que la taille de l'échantillon augmente, ne peut qu'être mauvaise à la limite; c'est qu'en fait une variable, telle que le poids de naissance par exemple, n'est sûrement pas distribuée normalement, car il y aurait alors une probabilité non nulle de valeurs infiniment grandes, ou négatives.

Pour d'autres, une distribution « pratiquement normale » reflète le théorème central limite et montre que la variable résulte d'effets nombreux, indépendants, additifs; et une distribution log normale la même situation pour des effets multiplicatifs, ou additifs en valeur relative. C'est là un début « d'explication » dans le déterminisme de la variable. Nous voudrions ajouter un autre argument. L'expérience statistique a largement montré que, si une variable est distribuée normalement, un facteur de variation modifie la moyenne mais pas la variance, et en translatant ainsi la distribution, indique que l'effet produit est le même quelle que soit la valeur de la variable; et ceci n'aura pas lieu pour une fonction de la variable non distribuée normalement, telle que son carré, ou son logarithme... Au contraire si x n'est pas distribuée normalement, mais que $f(x)$ le soit, par exemple $\text{Log } x$, c'est pour $f(x)$ qu'aura lieu la propriété précédente. Et ne devons-nous pas penser que, dans ce cas, la variable « qui compte » est $f(x)$?

Dans cette optique, pour la durée de vie humaine, une distribution quasi normale de $\text{Log}(1-x)$ suggère que les sujets reçoivent au départ de la vie un capital 1, auxquels les aléas opéreraient un retrait distribué de manière log normale; ce qui serait « expliqué » serait, non de la durée de vie, mais le retrait par rapport au capital initial.

L'idée d'un tel « capital » a déjà été suggérée, notamment Béard [29, 30] qui suppose les individus dotés au départ d'un certain nombre d'« unités de résistance ». Elle rejoint aussi la théorie de Szilard [28] où la vie a une durée maximale, diminuée d'un retrait reflétant les altérations du patrimoine génétique.

Si nous admettons que la variable « expliquée » est $\text{Log}(1-x)$ nous aurons tendance à penser que « ce qui compte » dans le vieillissement c'est, non l'addition de 1 an par exemple, mais la diminution relative du temps de vie restant; pour $l = 110$ par exemple, passer de 50 à 80 ans ou de 100 à 105 seraient des vieillissements « équivalents » ($110-x$) étant chaque fois divisé par 2. Une année d'âge « compterait » beaucoup plus (du point de vue du vieillissement) pour un sujet âgé que pour un sujet jeune.

Sans doute ce langage est-il un peu osé, c'est qu'en fait le « statisticien », à partir de la dispersion des longévités, raisonne ici sur un individu suivi au cours du temps, il parle en « biologiste ». Mais laissons-le un instant s'égarer sur cette voie... Rappelons la notion de « temps physiologique » évoquée par Lecomte de Nouy à propos de la cicatrisation des blessures, et qui serait le logarithme du temps chronologique. C'est encore le logarithme de ce temps qui devrait « compter » pour des sujets atteints de maladies graves, comme le cancer, puisque les survies ont une distribution log normale [31].

Il est frappant de constater ce contraste entre ces situations et celle du vieillissement; dans les premières situations, ce qui « compte » est $\log x$, dans la seconde c'est $\log(1-x)$; dans le premier cas, l'unité de temps a de moins en moins d'importance à mesure que le temps s'écoule, dans le deuxième cas elle en a de plus en plus; les deux images sont symétriques, et dans le vieillissement la vie fait figure d'un compte à rebours. Et devant cette image d'un vieillissement régi, dès son début, par sa fin, on ne peut s'empêcher de penser aux « droites de Gompertz » dont la pente au départ obéit à la nécessité de passer, à peu de choses près, par un point final...

Après cette description de l'attitude du « statisticien » et de celle du « biologiste », et du parti qui a pu être tiré de chacune, le moment est venu de procéder à leur confrontation sur un plan général.

Historiquement, la distribution de θ — la *courbe de mortalité* — figure dans les travaux du XVIII^e siècle, tandis que la notion de *quotient de mortalité* n'apparaît qu'au siècle suivant. Puis les différents auteurs utilisent, soit la courbe, soit le quotient, le plus généralement sans se préoccuper le moins du monde de l'autre indice; quelques-uns cependant font remarquer cette dualité [19, 26, 27, 32] pour prôner, soit l'une des attitudes, soit leur emploi simultané.

Nous voudrions, à propos de cette dualité, faire quelques commentaires.

Le premier concerne la portée du choix entre les deux attitudes : est-ce simple querelle de modes, ou différence profonde entre deux approches visant à décrire, l'une le vieillissement d'un individu, l'autre la variabilité des longévités dans une population?

On serait tenté d'abord de pencher pour la querelle mineure. Ne fût-ce que parce que les deux indices μ et θ peuvent se déduire l'un de l'autre : comment alors leurs apports seraient-ils différents? Et puis cette fonction, qui vise à traduire le vieillissement individuel, qu'est-elle sinon une probabilité, exprimant que certains sujets vont mourir et d'autres non, c'est-à-dire la variabilité dans la population?

Qu'il en soit ainsi au départ, certes. Mais l'utilisateur du quotient de mortalité introduit au moins deux notions nouvelles. Il postule d'abord que la probabilité de mort, indice collectif de la population, peut exprimer le degré de vieillissement d'un individu moyen. Il postule ensuite que l'individu moyen, aux divers âges, est le même individu. Alors naît une approche nouvelle, et ce n'est plus querelle mineure, mais bien différence profonde d'attitudes. Sans doute les qualificatifs de « statistiques » et « biologique » sont-ils trop absolus, puisque la distribution de θ a parfois été le point de départ, ou d'arrivée, de considérations biologiques; nous-mêmes y avons quelque peu cédé. Ces termes conservent cependant, une grande part de vérité. D'autres dichotomies ont été proposées, comme celle entre « prédestination » pour l'usage de θ , et « processus » pour celui de μ . L'essentiel en fin de compte est qu'un des points de vue concerne le groupe, et l'autre individu.

Cependant le double postulat, qui permet au « biologiste » de remonter de la connaissance transversale d'un groupe à la description longitudinale d'un phénomène individuel, demande réflexion.

Une probabilité peut-elle d'abord être tenue pour une caractéristique individuelle, ayant un sens concret? C'est là bien sûr un vieux débat, que nous nous bornerons à illustrer sur un exemple. La probabilité, pour un individu, de faire un infarctus dans l'année, est un indice collectif; mais l'augmentation de cette probabilité, à mesure que le sujet avance en âge, peut raisonnablement traduire l'augmentation d'un indice individuel de vulnérabilité. La croissance du risque de mort avec l'âge s'apparente bien à cette dernière situation.

En ce qui concerne le deuxième postulat, il est clair que si la population est hétérogène, formée d'individus à durée de vie systématiquement différentes, la variation de μ avec l'âge pose un redoutable problème; car l'individu moyen considéré aux divers âges résulte alors de la sélection qui élimine progressivement les individus à vie courte au profit de ceux à vie longue. De sorte que la courbe du biologiste concerne un individu constamment différent, ce qui rend son interprétation bien délicate. Sans doute le statisticien sera-t-il, lui aussi, gêné, mais il sera moins désarmé.

Notre deuxième commentaire vise à élargir le débat. Ce contraste entre les deux attitudes, qu'on observe dans l'étude de la mortalité, pourquoi ne l'observe-t-on pas, en général, dans d'autres domaines? Serait-ce que la durée de vie est une variable particulière?

Et si oui, ne partage-t-elle pas cette particularité avec d'autres variables, pour lesquelles on pourrait alors définir les deux attitudes?

Nous ouvrirons ce débat par le constat d'un apparent paradoxe. Disposant au départ de la distribution θ d'une variable x , la durée de vie, nous nous sommes d'emblée trouvés, dans l'attitude du biologiste, amenés à étudier la variation de la force de mortalité en fonction de l'âge. Comment, partis de données concernant une seule variable, avons-nous pu déduire une relation entre deux variables? C'est qu'en fait la durée de vie d'un sujet, $x = 80$ ans par exemple, nous indique que ce sujet, suivi au cours du temps, a été vivant pendant 80 ans et mort ensuite, nous disposons donc pour lui, non d'une information unique, mais d'informations à tous les instants, c'est-à-dire d'une relation exprimant son état en fonction du temps. Disons que la durée de vie n'est pas une variable comme les autres, son contenu informatif est beaucoup plus riche, les variables usuelles apportent une information de type ponctuel, celle-ci se situe au niveau d'une courbe. Que nous ayons alors une relation entre deux variables pour l'individu moyen de la population n'est pas pour nous étonner.

C'est pour des variables de ce type qu'on peut, à côté de l'attitude statistique, forger des indices concernant une évolution individuelle. C'est le cas des durées d'une manière générale, c'est bien pourquoi il nous a paru logique d'étudier ainsi la durée de gestation. Mais d'autres variables que les durées présentent un contenu informatif de même ordre, c'est le cas par exemple d'une dose létale, plus généralement du stimulus produisant un effet donné, variables dépassant largement le domaine de la biologie.

En définitive, il nous apparaît que, pour des variables ayant un contenu informatif déterminé, deux voies d'approche sont possibles, ayant chacune leurs conditions d'application; lorsque celles-ci sont vérifiées, on ne peut que gagner à les utiliser, l'une comme l'autre, en raison de l'éclairage différent qu'elles apportent.

Travail de l'Unité de recherches statistiques de l'I. N. S. E. R. M.

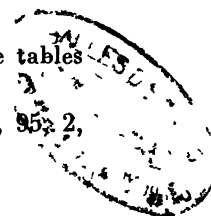
Daniel SCHWARTZ

Directeur de l'Unité de recherches, président de la Société de statistique de Paris

A. BOYCE

BIBLIOGRAPHIE

- [1] VALLIN J. — La mortalité par générations, en France, depuis 1899. Travaux et documents de l'INED. Cahier n° 63, P. U. F., 1973.
- [2] TEISSIER G. — Recherches sur le vieillissement et sur les lois de mortalité. *Ann. de Physiologie*, X, 2, 1934.
- [3] GOMPERTZ — On the Nature of the Function Expressive of the Law of Human Mortality. *Phil. Trans. Roy. Soc.*, 1825, 522.
- [4] MAKEHAM W.-M. — On the Law of Mortality and the Construction of Annuity Tables. *J. Inst. Actuaries*, 1860, 8, 301-310.
- [5] MERLIN P. — Recherches sur la mortalité française depuis 1900 et établissement de tables de générations jusqu'à la fin du siècle. Thèse I. S. U. P., Paris, oct. 1974.
- [6] REDINGTON F.-M. — An exploration into patterns of mortality. *J. Inst. Actuaries*, 95, 2, 401, 1969, 243-317.



- [7] STREHLER B.-L. et MILDVAN A.-S. — General theory of Mortality and Aging. A stochastic model relates observations on aging, physiologic decline, mortality and radiation. *Science*, 132, july 1960.
- [8] GOMBEL E.-J. — La durée extrême de la vie humaine. Herman, Paris, 1937.
- [9] SUTTER J. et TABAH L. — La mortalité, phénomène biométrique. *Population*, janv.-mars 1952.
- [10] VINCENT P. — La mortalité des vieillards. *Population*, 2, avril-juin 1951.
- [11] BOURGEOIS-PICHAT M.-J. — Essai sur la mortalité « biologique » de l'homme. *Population*, 3, juil.-sept. 1952.
- [12] DE FINETTI B. — A proposito di una ricerca sull'andamento della mortalita nelle et a senili. *Rivista italiana di economia demografia e statistica*, vol. 7, 2-3-4, déc. 1953.
- [13] LEBRAS H., BOYCE A. et SCHWARTZ S. — Travail en cours.
- [14] DEPOID D. — La mortalité des grands vieillards. *Population*, 4-5, juil.-oct. 1973.
- [15] HENSHAW P., STAPLETON G. et RILEY E.-R. — *Radiology*, 49, 349, 1947.
- [16] FURTH J., UPTON A.-C., CHRISTENBERRY K.-W., BENEDICT W.-H. et MOSHMAN J. — *Ibid.* 63, 562, 1954.
- [17] SACHER G. — *J. Natl. Cancer Inst.*, 15, 1125, 1955.
- [18] STREHLER B.-L. — Mortality Trends and Projections. *Transactions of the Society of Actuaries*, 19, 1967.
- [19] SACHER C.-A. et TRUCCO E. — Stochastic Theory of Mortality. *Ann. New York Acad. Sc.*, 96, March 1962.
- [20] BRILLINGER D.-R. — A justification of some common laws of mortality. *Transactions of the Society of Actuaries*, 13, 1961.
- [21] PERKS W. — On some experiments in the graduation of mortality statistics. *J. Inst. Actuaries*, 73, 1, 305, 1932, 12-57.
- [22] BJERKEDAL T., BAKKETEIG L. et LEHMANN E.-H. — Percentiles of birth weights of single, live births at different gestation periods. *Acta Paediat. Scand.*, 62, 1973, 449-457.
- [23] PEARSON K. — *Early Statistical Papers*. Cambridge University Press, 1948, 105-107.
- [24] BEARD R.-E. — Some notes on graduation. *J. Inst. Actuaries*, 77, 1951, 382-395.
- [25] BEARD R.-E. — Some further experiments in the use of the incomplete gamma function for the calculation of actuarial functions. *J. Inst. Actuaries*, 78, 1952, 341-343.
- [26] PHILLIPS W. — A basic curve of deaths. *J. Inst. Actuaries*, 80, 1954, 289-305.
- [27] LEXIS W. — Cité par M.-E. Ogborn, On the nature of the function expressive of the law of human mortality. *J. Inst. Actuaries*, 79, 1953, 174.
- [28] SZILARD L. — On the nature of the aging process. *Proceedings National Academy of Sciences*, 1959.
- [29] BEARD R.-E. — Dans la discussion de C.-D. Rich, A general theory of mortality. *J. Inst. Actuaries*, 70, 1939, 373.
- [30] BEARD R.-E. — Some experiments in the use of the incomplete gamma function for the approximate calculation of actuarial functions. *Proceedings of the Centenary Assembly of the Inst. Actuaries*, 2, University Press, Cambridge 1950, 89-107.
- [31] BOAG J.-W. — Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J. Royal Stat. Soc.*, B, 11, 1, 1949.
- [32] BARNETT H.-A.-R. — The components of mortality. *J. Inst. Actuaries*, 81, 1955, 105-131.

INTERVENTIONS

M. G. CALOT. — Je voudrais insister sur le fait que la courbe des quotients de mortalité par âge est une courbe de *fréquences moyennes*. En effet tout indique qu'à un âge donné la probabilité de décéder dans les douze mois à venir n'est pas la même pour tous les individus constituant une population. Par ailleurs, il n'est pas sûr qu'il y ait indépendance entre les risques de décès de deux individus donnés. Par conséquent un schéma binomial, ou même un schéma d'urnes de Poisson, ne conviennent peut-être pas pour décrire le mécanisme de la mort considérée comme un événement aléatoire.

A l'occasion de travaux que j'ai effectués à l'I. N. S. E. E. en 1965 (cf. *Études et Conjoncture*, n° 11, novembre 1965) sur la mortalité masculine par catégorie socio-professionnelle, j'ai pu vérifier qu'un « modèle » du genre ci-après s'ajustait assez convenablement aux données observées :

Repérons par l'indice j les catégories socio-professionnelles, par i les individus et par x l'âge. Soit $q_{ij}(x)$ le quotient de mortalité de l'individu i appartenant à la catégorie socio-professionnelle j à l'âge x . Tout se passe comme si :

— le quotient $q_{ij}(x)$ était tiré au hasard avec remise dans une population de quotients de mortalité, dont la moyenne $Q_j(x)$ varie selon la loi de Gompertz :

$$Q_j(x) = \exp(\alpha_j x + \beta_j),$$

et dont la variance est quelconque;

— il y a indépendance entre $q_{ij}(x_1)$ et $q_{ij}(x_2)$ pour $x_1 \neq x_2$;

— il y a indépendance entre les risques de décès entre deux individus quelconques.

Si on observe $n_j(x)$ individus d'âge x appartenant à la catégorie socio-professionnelle j , le nombre de décès $d_j(x)$ suit la loi binomiale :

$$d_j(x) = \mathcal{B}[n_j(x), \exp(\alpha_j x + \beta_j)]$$

et ces lois binomiales sont indépendantes entre elles lorsque x varie. Les paramètres α_j et β_j peuvent être estimés par la méthode du maximum de vraisemblance et les écarts entre effectifs théoriques et effectifs observés peuvent faire l'objet de tests d'adéquation.

Ce modèle convenait de façon satisfaisante pour 14 des 16 catégories socio-professionnelles étudiées (hommes de 35 à 70 ans).

M. MALIGNAC. — La formule de Gompertz-Makeham a été établie à l'époque où on ne connaissait pas les accidents d'automobile et leurs conséquences importantes sur la mortalité, conséquences variables avec l'âge. Ne conviendrait-il pas d'introduire ce facteur dans la formule de Makeham?

Réponse : Il est possible, en effet, que les accidents soient inclus dans la formule de Makeham.