

JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MAURICE FRÉCHET

Sur les expressions analytiques de la mortalité valables pour la vie entière

Journal de la société statistique de Paris, tome 88 (1947), p. 261-285

http://www.numdam.org/item?id=JSFS_1947__88_261_0

© Société de statistique de Paris, 1947, 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

SUR LES EXPRESSIONS ANALYTIQUES DE LA MORTALITÉ

VALABLES POUR LA VIE ENTIÈRE

Remarque générale.

On sait combien il est utile de connaître une formule mathématique pour représenter un phénomène. Une courbe est plus commode pour en embrasser qualitativement l'allure. Une table numérique est plus utile pour traiter numériquement des cas particuliers. La courbe et la table sont accessibles à un plus vaste public. Mais la formule est d'un format plus réduit, elle suggère l'analogie avec des phénomènes de nature concrète très différente. Et, surtout, elle est la seule à pouvoir s'insérer dans une suite de déductions permettant d'arriver sans nouvelles observations ou expériences à de nouvelles lois.

Or, ce précieux avantage est en grande partie perdu quand on a affaire à une formule qui ne représente un phénomène que dans une section de son développement.

Cas de la mortalité.

Tel est malheureusement le cas pour la plupart des formules proposées pour

représenter l'évolution de la mortalité avec l'âge et en particulier pour la plus célèbre d'entre elles, la loi de Gompertz-Makeham

$$l_x = k s^x g^{c^x}$$

dans les notations classiques. D'après elle, le taux annuel de mortalité est :

$$q_x = \frac{l_x - l_{x+1}}{l_x} = 1 - s g^{c(x+1)-c^x}.$$

Il varie donc toujours dans le même sens quand l'âge x croît tandis que, comme on sait, ce taux annuel commence par décroître. En fait, cette formule ne commence à donner de résultats satisfaisants qu'à partir de 20 ans (1).

Pour établir la table AF des assurés français, on a employé la formule de Gompertz-Makeham de 25 ans à l'âge extrême de la table, soit 99 ans. Mais, pour l'intervalle de 1 an à 25 ans, on a représenté le nombre l_x de survivants par un polynôme du 6^e degré en x , assujetti bien entendu à donner pour 25 ans le même nombre que la formule de Gompertz-Makeham employée (2).

Ce sont là bien de complications et on ne peut songer à utiliser cette représentation analytique globalement, il faut se résigner à ne l'appliquer que par morceaux.

Formules vie entière.

Aussi, depuis assez longtemps, a-t-on songé à obtenir des formules valables pour la vie entière. Dans plusieurs publications et dans son important ouvrage : *Assurances sur la vie. Calcul des primes*, Gauthier-Villars, 1924, p. 295-300, Galbrun indique les calculs à faire pour y arriver en développant la fonction $1 - p(x, t)$ en série de Gram-Charlier (3). Mais il n'a pas fait ces calculs et on ne sait si la méthode donnerait une bonne approximation avec un nombre assez petit de termes. Aussi devons-nous nous rabattre sur les seules expressions parvenues à notre connaissance et qui sont d'une toute autre forme. Ce sont celles qui sont mentionnées dans un article de J. F. Steffensen : *Infantile mortality from an actuarial point of view*, Skandinavisk Aktuarietidskrift, 1930, p. 272-286. Cet auteur rappelle d'abord qu'Oppermann avait proposé la formule :

$$\text{colog } l_x = 0,0353923 \sqrt{x} - 0,0110583 x + 0,00159406 \sqrt{x^3}$$

pour représenter l_x dans les premiers âges de la vie. Les constantes sont celles qui correspondent à des données $l_0, l_1, l_4, l_9, l_{16}$ et l_{25} déduites de « Peerage Experience, Female ». Et l'approximation obtenue est assez bonne (de 0 à 24 ans). Oppermann a obtenu la formule :

$$\text{colog } l_x = 0,0429080 \sqrt{x} - 0,1163985 x + 0,001339 \sqrt{x^3}$$

(1) On est généralement satisfait de l'application de la loi de Makeham aux âges élevés. Cela tient peut-être à ce que pour ces âges les nombres de décès enregistrés sont faibles, la portion correspondante du graphique de q_x offre de grandes oscillations et à peu près n'importe quelle fonction rapidement croissante donnerait une approximation non pas aussi bonne, mais pas beaucoup plus mauvaise que celle de Makeham. En fait, Gumbel, qui a étudié particulièrement la mortalité des âges élevés, considère que la loi de Makeham ne leur convient pas.

(2) Encore a-t-on eu recours à un autre et troisième procédé pour les valeurs de l_x de mois en mois de 0 à 1 an.

(3) Pour notre problème, il faudrait prendre $t = 0$.

pour « l'Institution Danoise d'État pour l'assurance sur la vie ». L'approximation de cette seconde formule est assez bonne jusqu'à 6 ans, un peu moins de 6 à 13 ans.

Oppermann proposait pour aller au delà de l'enfance la formule générale :

$$\frac{l_x}{l_0} = (1 + ax + bx^2 + \dots) e^{px - qx^2}.$$

Mais s'il a proposé cette formule pour l'enfance et l'âge adulte, il n'est pas certain qu'il l'eût recommandé pour toute la vie; car il écrit dans un rapport qu'une formule valable pour la vie entière « serait si utile qu'elle serait une compensation suffisante à n'importe quelle perte de temps et d'effort consacrée à l'obtenir ».

Thiele réussit une première tentative de réalisation du vœu d'Oppermann en proposant les deux formules ci-dessous :

$$\mu_x = \frac{a_1}{\sqrt{x}} + a_2 + a_3 \sqrt{x} + a_4 x + a_5 \sqrt{x^3} + a_6 x^2 + a_7 \frac{e^{a_8 \sqrt{x}}}{\sqrt{x}}$$

pour la mortalité masculine,

$$\mu_x = \frac{b_1}{\sqrt{x}} + b_2 + b_3 \frac{e^{b_4 \sqrt{x}}}{\sqrt{x}} \cos(b_5 + b_6 \sqrt{x}) + b_7 \frac{e^{b_8 \sqrt{x}}}{\sqrt{x}}$$

pour la mortalité féminine

et Thiele s'est effectivement servi de ces formules pour ajuster la mortalité scandinave.

Nous venons de résumer l'intéressant historique de Steffensen dans le mémoire mentionné ci-dessus. Steffensen observe que si la formule de Thiele donne une approximation assez bonne pour toute la vie pour être considérée comme la première des formules applicables à la vie entière, toutefois elle contient beaucoup de constantes (huit) et paraît bien empirique. Il lui paraît plus naturel de choisir une formule qui ne s'écarte sensiblement de la formule de Gompertz-Makeham que pour les âges faibles et qui, pour ceux-ci, s'inspire de l'idée d'Oppermann d'après laquelle il conviendrait que le taux instantané de mortalité, μ_x tende vers l'infini avec $\frac{1}{x}$. Il propose donc d'ajouter à l'expression de Gompertz-Makeham un terme correctif correspondant à une valeur de μ_x de la forme

$$\frac{x}{2M} \cdot \frac{10^{-x\sqrt{x-\lambda}}}{\sqrt{x}}$$

(où M est la constante de Napier, $M = \log_{10} e$).

Steffensen fait voir qu'on peut choisir les deux constantes x et λ de sorte que la formule donne une bonne approximation pour les premiers âges et observe qu'alors le terme correctif devient rapidement négligeable, étant inférieur à $\frac{1}{10^5}$ pour $x \geq 27,3$. Il obtient ainsi la formule :

$$\mu_x = \frac{10^{1,00031 - 0,71282 \sqrt{x}}}{\sqrt{x}} + 0,002080 + 100,039668x + 4,00722$$

Il donne enfin une table des valeurs de q_x calculées et observées, correspondant

à la table D^M (5) de 0 à 29 ans (pour les autres âges, la formule donnant pratiquement les mêmes résultats que celle de Gompertz-Makeham). Nous donnons ci-après un extrait de la table de Steffensen.

TABEAU I.

AGES	0	1	2	3	10	15	20	29
q_x calculé.	0,09562	0,01367	0,00705	0,00474	0,00250	0,00254	0,00275	0,00358
q_x observé.	0,09364	0,01158	0,00469	0,00307	0,00127	0,00160	0,00274	0,00358

Steffensen reconnaît qu'un statisticien considérerait comme insuffisante l'approximation obtenue pour les premiers âges. Toutefois, sa formule reproduit bien, par exemple, le phénomène de la décroissance initiale de μ_x avec un minimum à l'adolescence (minimum calculé à 12 ans; observé à 10 ans). Le résultat obtenu lui paraît donc suffisant au point de vue de l'actuaire qui s'intéresse plus au futur qu'au passé (dont une description sommaire lui suffit).

Nouvelle formule.

Avant de connaître ces premières tentatives, nous avons pensé, comme Oppermann, qu'il serait très précieux d'avoir une formule valable pour toute la vie. Nous avons été guidé dans la recherche d'une telle formule par des considérations de nature très diverses; les unes théoriques, les autres pratiques. Il nous paraît inutile de reproduire celles-ci, l'intérêt de la formule obtenue étant surtout de montrer que des formules très différentes peuvent rapprocher du but et que des *essais successifs ultérieurs*, menés avec persévérance ou une tentative plus heureuse — l'intuition et la chance aidant — permettront certainement d'arriver à une formule simple donnant une bonne approximation pour toute la vie. Il est probable qu'on aurait intérêt à partir d'hypothèses basées sur les raisons physiologiques de la mortalité.

La formule que nous présenterons contient un polynôme du 6^e degré. On sera donc tenté de considérer le degré de ce polynôme comme trop élevé, et il est, en effet, souhaitable de diminuer le nombre (sept) des constantes utilisées dans la formule. Toutefois, si l'on se souvient que pour la table AF le polynôme du même 6^e degré employé n'était utilisé que de 1 à 25 ans alors que le nôtre (liant d'ailleurs une variable et une fonction différente de celles de la table AF) s'applique à la *vie entière*, on reconnaîtra le progrès obtenu. Notre formule, comme celles de Thiele et de Steffensen, reproduit le phénomène du minimum à l'adolescence de la mortalité. Nous estimons que les efforts ultérieurs devront tendre à reproduire, en outre, un *second phénomène* moins marqué et moins connu, celui de l'existence de deux points d'inflexion du graphique de q_x au voisinage de 30 ans, phénomène qui s'accroît pour certaines populations, comme la japonaise, jusqu'à se traduire par un maximum local de q_x aux environs de 32 ans (1).

(1) Le graphique I montre aussi l'existence dans la table RAV d'un maximum local du taux de mortalité entre 20 et 30 ans vers 25 ans.

Forme de la formule adoptée.

En appelant q_x le taux annuel de mortalité

$$\left(q_x = \frac{l_x - l_{x+1}}{l_x} \right)$$

nous avons cherché à représenter non pas q_x mais $\log q_x$ par un polynôme et en prenant comme variable non pas x mais d'abord $\log x$. Nous suivions, sans le savoir, la même idée qu'Oppermann et Steffensen en supposant q_x infini pour $x = 0$. D'ailleurs, réflexion faite, il nous avait paru qu'il serait plus naturel de faire remonter l'origine à l'époque de la conception, et nous avons pensé à prendre comme variable, non pas $\log x$ mais $\log (x + 0,75)$. Finalement, pour des raisons de calcul, nous avons choisi comme variable $\log (x + a)$, a étant une constante positive convenablement choisie, en procédant d'ailleurs par simple tâtonnement pour ce choix qui n'a qu'un effet secondaire.

Mode de calcul.

Nous avons déterminé les constantes de cette formule en calculant les coefficients du polynôme par la méthode des moindres carrés. Nous avons d'abord, pour simplifier, adopté, pour la première table ajustée par nous, des poids égaux pour les différentes observations. Puis, dans les derniers calculs, relatifs à la seconde table mentionnée plus loin, nous avons introduit une série de poids dont les valeurs étaient les nombres des « exposés au risque » à chaque âge, dans l'espérance d'améliorer l'approximation, ce qui a eu lieu sans toutefois donner la bosse vers 30 ans mentionnée plus haut. Quoi qu'il en soit, nous donnerons plus loin la première table, puis la formule d'approximation et la représentation graphique des résultats concernant la première et la seconde tables.

TABLES CONSIDÉRÉES

I. — *Table d'assurés suisses.*

Nous avons fait un premier essai en utilisant une table de mortalité que nous venions de recevoir. C'est une table figurant aux pages 74 à 77 du volume *Soixante-quinze ans d'activité, 1857-1932*, publié en 1933 par la Société Suisse d'Assurances générales sur la vie humaine, Zurich. A ces pages sont donnés les relevés de taux bruts de mortalité correspondant à deux catégories différentes d'assurés. Nous avons choisi celui qui correspondait au plus grand nombre d'assurés, c'est-à-dire à « l'Assurance populaire ». Ces relevés ont servi à ladite Société pour établir la table de taux ajustés dite :

Table pour les assurances en cas de décès de l'Assurance populaire	RAV	$\frac{1894/1930}{1921/1931}$
---	-----	-------------------------------

et qui figure aux pages 81 et 82 du même volume. On trouvera ces nombres dans les tables plus loin. Ces tables donnent les taux jusqu'à 89 ans; mais les exposés au risque étant déjà réduits à 10 à 83 ans, nous n'avons utilisé que les taux de 0 à 82 ans.

La Société a opéré l'ajustement par tranches d'âges de la façon suivante :
taux ajustés pris égaux aux taux bruts de 0 à 3 ans; de 3 à 12 ans, ajustement

par une parabole du 3^e degré; de 12 à 40 ans, ajustement par la méthode de King; à partir de 40 ans, ajustement par la formule de Makeham. La Société donne une idée de l'approximation obtenue dans les lignes suivantes : « Assurance populaire » (âge d'entrée de 0 à 80 ans) :

Nombre des décès présumés d'après la table ajustée.	4.932,9
Nombre des décès constatés	4.913
Écart. . .	19,9

L'écart représente ainsi 4 ‰ environ des décès constatés. »

Pour un premier essai, nous nous sommes contentés d'abord d'employer un polynôme du 3^e degré pour représenter à lui seul la loi de mortalité pour la vie entière (plus exactement de 0 à 82 ans). Le polynôme étant :

$$\eta = \alpha + \beta \xi + \gamma \xi^2 + \delta \xi^3$$

η représentait $\log q_x$ où q_x est le nombre de décès constaté sur 1.000 assurés d'âge x , et ξ représentait $\log (x + a)$. On a trouvé par la méthode des moindres carrés :

$$\alpha = 0,163; \quad \beta = -0,522; \quad \gamma = 2,045; \quad \delta = 1,866$$

et on a dressé un tableau des valeurs observées et calculées de q_x pour tous les âges entiers. Nous en extrayons ici ce qui correspond à l'extrait donné plus haut de la table de Steffensen,

TABLEAU II.

AGES	0	1	2	3	10	15	20	29
q_x calculé	0,00857	0,00871	0,00494	0,00376	0,00149	0,00136	0,00154	0,00247
q_x observé	0,01382	0,00600	0,00390	0,00302	0,00103	0,00150	0,00205	0,00235

en rappelant que notre formule ne contient que 4 coefficients contre 5 dans celle de Steffensen (que, d'ailleurs, nous ne connaissions pas à l'époque [1941] où ont été faits nos calculs). Il est vrai que la formule de Steffensen donne sans doute des résultats bien meilleurs que la nôtre pour les âges plus élevés.

Encouragé par ces résultats, nous avons fait appliquer la même méthode pour obtenir un polynôme du 6^e degré, dans la relation :

$$(1) \quad \eta' = \log q'_x + 3 = \alpha + \beta \xi + \gamma \xi^2 + \delta \xi^3 + \zeta \xi^4 + \psi \xi^5 + \lambda \xi^6$$

où $\xi = \log (x + 2) - 1,096$.

Le nombre 1,096 n'ayant aucune signification particulière, avait été introduit seulement pour faciliter les calculs numériques. On a trouvé :

$$\alpha = 0,15915; \quad \beta = 0,25607; \quad \gamma = 1,93027;$$

$$\delta = -1,93398; \quad \zeta = -0,90240; \quad \psi = 3,92202; \quad \lambda = 2,30667.$$

On trouvera plus loin le tableau V indiquant pour chaque âge x , la valeur observée du taux de mortalité q_x , la valeur correspondante η de $\log q_x$, la valeur calculée η' de η , d'après (1) et la valeur calculée correspondante q'_x de q_x .

On a fait un triple contrôle. D'une part, on a calculé les erreurs $\eta' - \eta$, qu'on trouvera dans le tableau V. On en a déduit leur valeur médiane qui est précisé-

ment égale à 0, puis pour les écarts en valeur absolue $|\eta' - \eta|$, leur médiane E, leur moyenne quadratique

$$\sigma = \sqrt{\frac{\sum (\eta' - \eta)^2}{83}}$$

et leur moyenne

$$\mu = \frac{\sum |\eta' - \eta|}{83}$$

L'intérêt que présente l'étude des erreurs $\delta_x = \eta' - \eta$, c'est qu'en posant :

$$q'_x - q_x = \Delta_x,$$

on a :

$$\delta_x = \log q'_x - \log (q'_x - \Delta_x)$$

$$\delta_x = -\log \left(1 - \frac{\Delta_x}{q'_x}\right) = -M \mathcal{L} \left(1 - \frac{\Delta_x}{q'_x}\right)$$

Or, quand $\frac{\Delta_x}{q'_x}$ est petit, ce rapport, qui est l'erreur relative sur q'_x est la partie principale de $-\mathcal{L} \left(1 - \frac{\Delta_x}{q'_x}\right)$. On a donc approximativement :

$$\delta_x \simeq M \frac{\Delta_x}{q'_x}$$

de sorte que les erreurs absolues δ_x sur η' sont à peu près proportionnelles aux erreurs relatives sur le taux de mortalité q'_x . On a trouvé :

$$E = 0,0505; \quad \mu = 0,065; \quad \sigma = 0,0824.$$

Les erreurs relatives sur le taux de mortalité sont à peu près $\frac{1}{M}$ fois plus grandes ($M = 0,434294$), donc à peu près deux fois plus grandes que les δ , en tous cas de l'ordre de grandeur moyen de 1 %; c'est-à-dire une erreur relative qui, sans être excellente, n'est pas excessive. Comparons maintenant les trois ajustements de $\eta = \log q_x$.

TABLEAU III.

ÉCARTS	AJUSTEMENT par la table RAV	NOTRE AJUSTEMENT PAR L'INTERMÉDIAIRE d'un polynôme	
		du 6 ^e degré	du 3 ^e degré
médian : E	0,084	0,0505	0,097
moyen : μ	0,0465	0,0650	0,109
quadr. moyen : σ	0,064	0,0824	0,131

On voit, comme on pouvait s'y attendre, qu'on a augmenté les erreurs relatives en remplaçant les quatre formules, relatives à quatre tranches d'âge, utilisées pour l'ajustement RAV par la formule, unique de l'un ou l'autre de nos deux ajustements. On augmente encore les erreurs relatives en utilisant pour notre ajustement un polynôme de degré plus bas (du 3^e au lieu du 6^e). Mais on observera que ces augmentations successives ne sont pas extrêmement fortes, chaque colonne n'étant environ que moitié plus forte que la précédente.

Un autre contrôle est basé sur la remarque suivante. Si les valeurs ajustées étaient exactes, les quantités δ pourraient être considérées comme 83 erreurs

d'observation. Il serait naturel de supposer qu'elles sont distribuées suivant la loi de Laplace (1). Or, pour celle-ci, on connaît les valeurs de $\frac{E}{\sigma}$ et $\frac{\mu}{\sigma}$. On a alors calculé ces valeurs pour les deux ajustements obtenus avec des polynômes du 3^e et du 6^e degré; on les a aussi calculées pour les δ correspondant à la table ajustée, RAV (basée sur les mêmes données brutes mais l'ajustement étant fait par 4 formules différentes relatives à 4 tranches d'âges différentes).

On a ainsi obtenu le tableau suivant :

TABLEAU IV.

	LOI de Laplace	AJUSTEMENT par la table RAV	NOTRE AJUSTEMENT PAR L'INTERMÉDIAIRE d'un polynôme	
			du 6 ^e degré	du 3 ^e degré
$\frac{E}{\sigma}$	0,674	0,532	0,624	0,741 0,833
$\frac{\mu}{\sigma}$	0,798	0,725	0,789	

(1) Loi appelée souvent, et d'ailleurs à tort, loi normale ou loi de Gauss. D'autres lois, très importantes, loi de Pareto..., ne peuvent être considérées comme anormales. Et Gauss avait trois ans quand Laplace a formulé la loi des erreurs d'observation (résultat essentiellement distinct de la formule d'approximation due à de Moivre).

TABLEAU V

Table de mortalité relative à l'assurance populaire basée sur les données brutes
trouvées dans « soixante-quinze ans d'activité »
publié par la Société Suisse d'Assurances générales et ajusté par la formule
 $\log. Q_x - 3 = \eta' = \text{polynôme du 6}^{\text{e}} \text{ degré en } \zeta = \log(x + 2) - 1,096.$

x (âge)	0	1	2	3	4	5	6	7	8
η observé	1,1405	0,7782	0,5911	0,4800	0,3874	0,2895	0,2718	0,2279	0,0969
$\eta' = \eta$ calculé	1,1237	0,8396	+0,6012	0,4307	0,3153	0,2401	0,1933	0,1666	0,1540
$\eta' - \eta$	-0,0169	+0,0614	+0,0101	-0,049	-0,0520	-0,0295	-0,0786	-0,0613	+0,0571
Q_x observé taux pour 1.000	13,82	6,00	3,90	3,02	2,33	1,86	1,87	1,69	1,25
Q'_x calculé	13,80	6,91	3,99	2,70	2,07	1,74	1,56	1,47	1,43
$Q'_x - Q_x$	-0,02	+0,91	+0,09	-0,32	-0,26	-0,12	-0,31	-0,22	+0,18
ρx	-0,88	1,15	0,16	-0,78	-0,76	-0,40	-1,11	-0,85	0,72

x (âge)	9	10	11	12	13	14	15	16	17
η observé	0,1206	0,0128	0,0828	0,0253	0,0864	0,0899	0,1761	0,2648	0,2648
$\eta' = \eta$ calculé	0,1512	0,1554	0,1644	0,1766	0,1910	0,2069	0,2237	0,2409	0,2585
$\eta' - \eta$	+0,0307	+0,1426	+0,0816	+0,1513	+0,1047	+0,1170	-0,0476	-0,0567	-0,0063
Q_x observé taux pour 1.000	1,32	1,03	1,21	1,0	1,22	1,23	1,50	1,99	1,84
Q'_x calculé	1,42	1,43	1,46	1,50	1,55	1,61	1,67	1,74	1,81
$Q'_x - Q_x$	+0,10	+0,40	+0,25	+0,44	+0,33	+0,38	+0,27	-0,25	-0,03
ρx	0,40	1,67	1,04	1,83	1,875	1,68	1,08	-1,00	-0,12

x (âge)	18	19	20	21	22	23	24	25	26
η observé	0,4099	0,4843	0,4698	0,4698	0,448	0,4756	0,4639	0,4564	0,4400
$\eta' = \eta$ calculé	0,2761	0,2937	0,3113	0,3287	0,3461	0,3633	0,3805	0,3977	0,4149
$\eta' - \eta$	-0,1338	-0,1906	-0,1685	-0,1411	-0,0588	-0,1153	-0,0834	-0,0587	-0,0260
Q_x observé taux pour 1.000	2,57	3,05	2,95	2,95	2,54	3,01	2,91	2,86	2,76
Q'_x calculé	1,89	1,97	2,05	2,13	2,22	2,31	2,40	2,50	2,60
$Q'_x - Q_x$	-0,68	-1,08	-0,94	-0,82	-0,32	-0,70	-0,51	-0,36	-0,16
ρx	-2,615	-4,15	-3,76	-3,15	-1,23	-2,59	-1,82	-1,29	-0,53

x (âge)	27	28	29	30	31	32	33	34	35
η observé	0,5328	0,4829	0,3711	0,4594	0,4216	0,3483	0,4409	0,5276	0,3927
$\eta' = \eta$ calculé	0,4322	0,4496	0,4672	0,4851	0,5032	0,5217	0,5406	0,5599	0,5793
$\eta' - \eta$	-0,1006	-0,0333	+0,0962	+0,0257	+0,0816	+0,1734	+0,0997	+0,0322	+0,1866
Q_x observé taux pour 1.000	3,41	3,04	2,35	2,88	2,64	2,23	2,76	3,37	2,47
Q'_x calculé	2,71	2,82	2,93	3,06	3,19	3,32	3,47	3,63	3,80
$Q'_x - Q_x$	-0,70	-0,22	+0,58	+0,18	+0,55	+1,09	+0,71	+0,26	+1,33
ρx	-2,33	-0,71	1,81	0,55	1,62	3,03	1,92	0,67	3,24

x (âge)	36	37	38	39	40	41	42	43	44
η observé	0,5539	0,5740	0,6031	0,6684	0,6580	0,6590	0,7523	0,7566	0,7723
$\eta' = \eta$ calculé	0,5999	0,6207	0,6422	0,6642	0,6869	0,7103	0,7344	0,7593	0,7849
$\eta' - \eta$	+0,0460	+0,0467	+0,0390	-0,0042	+0,0289	+0,0514	-0,0184	+0,0028	+0,0126
Q_x observé taux pour 1.000	3,58	3,75	4,01	4,66	4,55	4,56	5,66	5,71	5,92
Q'_x calculé	3,98	4,18	4,39	4,62	4,86	5,13	5,43	5,74	6,09
$Q'_x - Q_x$	+0,40	+0,43	+0,38	-0,04	+0,31	+0,57	-0,23	+0,03	+0,17
ρx	0,95	0,98	0,81	0,08	0,61	1,06	0,41	-0,05	0,27

x (âge)	45	46	47	48	49	50	51	52	53
η observé	0,7701	0,8287	0,8882	0,9031	0,9096	0,8893	1,0370	1,0752	1,1119
$\eta' = \eta$ calculé	0,8113	0,8396	0,8667	0,8955	0,9253	0,9560	0,9876	1,0200	1,0584
$\eta' - \eta$	0,0412	0,0099	-0,0215	-0,0076	+0,0158	+0,0667	-0,049	-0,0552	-0,0566
Q_x observé taux pour 1.000	5,89	6,74	7,73	8,00	8,12	7,75	10,89	11,89	12,94
Q'_x calculé	6,48	6,90	7,36	7,86	8,42	9,04	9,72	10,47	11,31
$Q'_x - Q_x$	+0,59	-0,16	+0,37	-0,14	+0,30	+1,29	-1,17	-1,42	-1,63
ρx	0,89	0,23	-0,50	-0,18	0,36	1,45	-1,19	-1,34	-1,41

x (âge)	54	55	56	57	58	59	60	61	62
η observé	1,0504	1,077	1,2531	1,2548	1,3532	1,3329	1,3090	1,4430	1,4333
$\eta' = \eta$ calculé	1,0877	1,1229	1,1592	1,1964	1,2345	1,2736	1,3136	1,3547	1,3968
$\eta' - \eta$	+0,0373	+0,0459	-0,0939	-0,0584	-0,1186	-0,0893	+0,0047	-0,0883	-0,0365
Q_x observé taux pour 1.000	11,23	11,94	17,91	17,98	22,55	23,06	20,37	27,73	27,15
Q'_x calculé	12,24	13,27	14,43	15,72	17,16	18,78	20,59	22,63	24,93
$Q'_x - Q_x$	+1,01	+1,33	-3,48	-2,26	-5,39	-4,28	+0,22	-5,10	-2,19
ρx	0,80	0,97	-2,19	-1,27	-3,13	-1,91	-0,0	-1,57	-0,58

x (âge)	63	64	65	66	67	68	69	70	71
η observé	1,4042	1,5270	1,4476	1,4349	1,5885	1,8280	1,5805	1,7614	1,8497
$\eta' = \eta$ calculé	1,4397	1,4837	1,5288	1,5748	1,6218	1,6697	1,7188	1,7687	1,8197
$\eta' - \eta$	+0,0356	-0,0433	+0,081	+0,1399	+0,0333	-0,1582	+0,1383	+0,0073	-0,0301
Q_x observé taux pour 1.000	25,36	33,65	28,03	27,22	38,77	67,29	38,06	57,73	70,75
Q'_x calculé	27,52	30,46	33,79	37,57	41,86	46,75	52,33	58,71	66,02
$Q'_x - Q_x$	+2,16	-3,19	+5,76	+10,35	+3,09	-20,54	14,27	+0,98	+4,73
ρx	0,50	0,65	1,02	1,28	0,82	-1,88	1,09	0,06	-0,28

x (âge)	72	73	74	75	76	77	78	79	80
η observé	1,8656	2,0238	1,9695	1,9830	2,0809	2,3500	2,1667	2,2730	2,1761
$\eta' = \eta$ calculé	1,8716	1,9245	+1,9785	2,0334	2,0894	2,1463	2,2042	2,2632	2,3229
$\eta' - \eta$	+0,0060	-0,0993	+0,0090	+0,0505	+0,0084	-0,2037	+0,0375	-0,0099	+0,1469
Q_x observé taux pour 1.000	73,39	105,63	93,22	96,15	120,48	223,68	146,79	187,5	150,0
Q'_x calculé	74,40	84,05	95,17	108,0	122,8	140,1	160,0	183,3	210,3
$Q'_x - Q_x$	+1,01	-21,58	+1,95	+11,8	+2,32	+83,8	+13,2	+4,2	+60,3
ρx	0,05	-0,93	0,07	0,39	0,06	1,97	0,27	-0,08	0,94

On voit que, pour notre ajustement, par l'intermédiaire d'un polynôme, ces résultats sont, comme on pouvait encore s'y attendre, plus près pour le 6^e degré que pour le 3^e degré, de ceux qui correspondent à la loi de Laplace. Ici, l'on comparait les valeurs de δ_x pour les différentes valeurs de x .

Un autre contrôle a été basé sur l'idée que la valeur observée q_x pouvait être considérée pour une valeur déterminée de x comme la fréquence, — dans un nombre d'épreuves égal au nombre e_x d'« exposés au risque », — d'un événement de probabilité q'_x .

L'écart quadratique moyen de cette fréquence est :

$$\sigma_x = \sqrt{\frac{q'_x(1 - q'_x)}{e_x}}$$

où le nombre e_x est donné dans la même table que celle des taux bruts q_x .

On forme le quotient :

$$\rho_x = \frac{q'_x - q_x}{\sigma_x} = \frac{\Delta_x}{\sigma_x}$$

et on n'acceptera l'ajustement que si la plupart des valeurs de ρ_x sont comprises entre moins 3 et plus 3. En fait, c'est ce qui a lieu, pour la plupart des valeurs de x , pour notre ajustement par l'intermédiaire d'un polynôme du 6^e degré, ajustement pour lequel la plus grande valeur de $|\rho_x|$ est 4,15 et où il n'y a que 10 valeurs de $|\rho_x|$ sur 83, qui soient supérieures à 2, dont 6 seulement sont supérieures à 3.

On trouvera ces différents résultats dans la table numérique V qui suit. On embrassera d'un coup d'œil deux des ajustements mentionnés sur les graphiques rejetés à la fin de cet article.

(Ce contrôle donne un résultat meilleur pour l'ajustement RAV, pour lequel la plus grande valeur de $|\rho_x|$ (pour $x \leq 82$) est 2,1. Mais il faut noter que c'est ce qui doit arriver quand on adopte des formules diverses pour diverses tranches d'âges, à moins que les formules employées ne soient trop mauvaises. En poussant la méthode à l'extrême, si chaque tranche était d'un seul chiffre, on devrait prendre $q'_x = q_x$ et on aurait ρ_x constamment nul, ce qui serait parfait en soi, sauf qu'il n'y aurait plus du tout d'ajustement.)

REMARQUE. — La discussion précédente nous paraît suffisante. Cependant, la nécessité de dissiper un malentendu qui s'est produit nous amène à la détailler un peu plus. Nous nous excusons d'avance des longueurs et des répétitions auxquelles nous avons été ainsi conduit.

Il est important, pour comparer les ajustements représentés par la figure de ne pas perdre de vue les deux buts différents auxquels visaient leurs auteurs.

On aurait, en tout cas, tout à fait tort de se placer uniquement au point de vue de l'approximation qu'ils réalisent. Faisons-le cependant un instant. Nous avons dit plus haut, en divers passages que la table RAV réalise une meilleure approximation que notre table (tableau V). Bien que nous l'ayons montré et dit nettement, on nous en a communiqué deux autres preuves que nous allons reproduire.

Si les nombres de notre table étaient exactement les probabilités de morta-

lité, parmi les nombres $\varphi_x = \frac{q'_x - q_x}{\sigma_x}$ cités plus haut, 4 seulement sur 83 . devraient être > 2 dont 0,22 seulement supérieurs à 3 alors qu'il y en a respectivement 10 et 6.

D'autre part, en appliquant le critère classique de :

$$\chi^2 = \sum \frac{e_x (q_x - q'_x)^2}{q'_x}$$

où e_x est le nombre des exposés au risque, on en a déduit que la probabilité pour une plus grande divergence entre la théorie et l'observation, c'est-à-dire entre les valeurs calculées de χ^2 selon notre ajustement et leur espérance mathématique serait très faible, de l'ordre de 10^{-6} , pendant que, pour l'ajustement RAV, cette probabilité est de 90 %.

Cela suffirait-il pour justifier le choix de l'ajustement RAV ? Pas encore.

La Société suisse avait à réaliser une *autre* condition : obtenir une courbe de mortalité suffisamment régulière pour qu'on puisse la considérer comme suffisamment indépendante des variations accidentelles qui se présentent dans les taux bruts. Si l'on renonçait à cette condition, si l'on se contentait de la précédente, relative à l'approximation la meilleure possible, alors l'ajustement RAV serait à la fois plus compliqué et moins bon que celui qui consiste tout simplement à prendre les taux calculés égaux aux taux bruts. *Toutes les raisons données plus haut* pour préférer la table RAV à notre ajustement, *feraient préférer l'absence d'ajustement, non seulement à la table RAV, mais à tout autre ajustement.* Grâce à l'absence d'ajustement, toutes les quantités $q'_x - q_x$ seraient, en effet, nulles, les quantités φ_x seraient nulles, χ^2 serait nul et la probabilité de 90 % serait remplacée par 100 %!...

Mais les actuaires de la Société suisse n'ont pas perdu de vue cette seconde condition. Quand on adopte un plus grand nombre de formules dans le même — plus grand — nombre d'intervalles, la condition de bonne approximation peut être *mieux* réalisée, mais la condition de régularité est *moins bien* réalisée. Il y a un *juste milieu* à tenir. La Société suisse a su trouver ce juste milieu, car tout en obtenant la bonne approximation qui se manifeste par la faible valeur de χ^2 , la courbe qu'elle a obtenue est d'une bonne régularité. On pourrait être tenté de préférer la nôtre au point de vue de la régularité, puisque la nôtre ne présente pas la bosse qui se manifeste dans la courbe RAV vers 30 ans. Et ce serait là une remarque juste si cette bosse était imputable aux erreurs accidentelles. Mais nous avons fait observer plus haut que cette bosse est une caractéristique générale des tables de mortalité vers les 30 ans. C'est donc, au contraire, une trop grande régularité de notre ajustement qui a fait disparaître cette particularité non accidentelle. Et nous avons exprimé plus haut le souhait qu'un perfectionnement de notre résultat vienne faire réapparaître cette bosse caractéristique.

En résumé, si l'on se place au point de vue qui a guidé la Société suisse et la plupart des auteurs de tables de mortalité, c'est-à-dire si l'on se propose d'obtenir une courbe ajustée assez régulière et qui, pourtant, ne s'écarte pas trop de la courbe des taux bruts, la table RAV est une excellente solution de ce problème. Elle est, en particulier, une meilleure solution du problème

qu'elle était destinée à résoudre que la solution qui serait présentée par notre courbe *pour un autre problème* que celui qui en était la raison d'être.

Par contre, notre courbe est une meilleure solution du problème que nous nous sommes proposé, que la courbe RAV dressée *dans un autre but*. Dans notre problème, il s'agissait de réaliser le mieux possible les deux conditions précédentes au moyen d'une seule formule, mais de le faire avec une formule qui *ne soit pas trop compliquée*, qui ne comporte pas un nombre excessif de constantes.

Or, observons que l'ajustement au moyen de formules relatives à des intervalles distincts *peut toujours s'exprimer* par une formule valable *vie entière* et exactement *équivalente*. Soit, en effet, un ajustement tel que q'_x soit représenté par l'expression $f_1(x)$ dans un ensemble E_1 de valeur de x , par $f_2(x)$ dans un ensemble E_2 de valeurs de x , etc... Soit alors $\varphi_k(x)$ une fonction égale à 1 sur E_k et à 0 en dehors. L'ajustement considéré pourra être représenté par la relation :

$$q'_x = \rho(x)$$

valable pour toute valeur de $x \geq 0$, en posant :

$$\rho(x) = f_1(x) \varphi_1(x) + f_2(x) \varphi_2(x) + \dots (1)$$

Si la formule $q'_x = f_k(x)$ dans E_k a fait intervenir c_k constantes, nous voyons que cet ajustement s'exprime par une formule valable pour la vie entière, mais, en général, plus compliquée que chacune des expressions f_k et comportant un nombre de constantes égal à $c_1 + c_2 + \dots$

Par exemple, la table RAV correspond, d'une part, aux formules suivantes :

$$\begin{aligned} q'_x &= \text{const.} = \alpha_1 = f_1(x) \text{ dans l'ensemble } E_1 \text{ constitué par } x = 0 \\ q'_x &= \text{const.} = \alpha_2 = f_2(x) \text{ dans l'ensemble } E_2 \text{ constitué par } x = 1 \\ q'_x &= \text{const.} = \alpha_3 = f_3(x) \text{ dans l'ensemble } E_3 \text{ constitué par } x = 2 \\ q'_x &= \alpha_4 + \alpha_5 x + \alpha_6 x^2 + \alpha_7 x^3 = f_4(x) \text{ dans l'ensemble } E_4 \text{ constitué par } 3 \leq x \leq 12 \\ \log(1 - q'_x) &= \alpha_8 + \alpha_9 (\alpha_{10})^x = \log[1 - f_5(x)] \text{ dans l'ensemble } E_5 \text{ constitué par } 40 \leq x \end{aligned}$$

qui comporte 10 constantes, réduites à 8 si l'on impose la continuité aux points $x = 12$, $x = 40$; d'autre part, à un ajustement, dont il est dit seulement qu'il est fait suivant la méthode de King, dans l'intervalle $12 \leq x \leq 40$.

Notre ajustement comporte 7 constantes, coefficients du polynôme du 6^e degré. Si, au lieu de poser $a = 0,75$, on laisse a arbitraire (d'ailleurs dans des limites étroites), on aurait même 8 constantes. Le nombre 1,096 qui figure dans l'expression de ξ n'a rien d'une constante arbitraire, il n'a été introduit que pour la commodité du calcul. On aurait eu une formule en ξ' différente mais absolument équivalente, si, au lieu de poser $\xi = \log(x + a) + 1,096$, on avait pris comme variable $\xi' = \log(x + a)$.

La table RAV comporte donc, par rapport à la nôtre un nombre supplémentaire de constantes égal à celui qui correspond à l'emploi de la méthode de King pour l'intervalle 12-40 ans, nombre qui n'est pas indiqué. Mais sur-

(1) Bien que les fonctions $\varphi_1(x)$, $\varphi_2(x)$..., soient discontinues, comme $y = f(x)$ est l'équation de la courbe représentative de l'ajustement par intervalles distincts, $f(x)$ sera continue, si, comme on le fait toujours, on a choisi f_1, f_2 ..., de sorte que cette courbe soit continue.

tout, si, pour faire la comparaison de deux formules valables vie entière, on représente analytiquement la table RAV au moyen d'une expression de la forme (2), on voit que notre formule est de loin la plus simple.

Mais ces deux remarques sont faites pour le principe; elles ont aussi peu de portée que celles qui avaient pour but d'établir que la table RAV donnait une approximation meilleure que la nôtre. Les buts visés par les auteurs des deux tables ne sont pas les mêmes et on ne saurait pas plus faire grief aux actuaires suisses de conduire à une formule (valable vie entière) trop compliquée, puisqu'ils ne cherchaient pas à aboutir à cette formule, que de nous objecter une approximation moins bonne puisqu'en nous astreignant à obtenir une formule valable vie entière, nous nous imposons une condition dont on sait d'avance qu'elle restreint l'approximation.

Laissant donc de côté la comparaison de notre table avec la table RAV (qui ne se justifie ni si l'on a en vue notre but, ni si l'on a en vue le but des actuaires de la Société suisse), il reste à juger en soi notre table. Ceux qui ont confiance dans des conclusions absolues tirées du χ^2 diront que notre table ne résout pas du tout le problème posé, ceci en ce qui concerne l'approximation désirée. La suite montrera combien cette conclusion serait précipitée et prouvera *une fois de plus* combien il serait dangereux de se laisser guider uniquement par le test χ^2 .

Bien entendu, les tables dressées avec plusieurs formules pour plusieurs intervalles ne résolvent pas non plus *notre* problème mais pour une autre raison. Le test du χ^2 fournissait un critère global. Si l'on veut serrer la question, le critère non global reposant sur les valeurs de φ_x , montre que notre ajustement, qui satisfait à ce critère pour la plupart des âges, n'y satisfait pas pour une dizaine d'entre eux. Si l'on inspecte la table à ce point de vue, on s'aperçoit que les valeurs de φ_x qui ne sont pas satisfaisantes se trouvent presque toutes dans l'intervalle d'âge de 18 ans à 35 ans, c'est-à-dire précisément dans cet intervalle où, comme nous l'avons déjà observé, notre ajustement ne reproduit pas la bosse qui caractérise cet intervalle. Nous vérifions donc ce qui a été dit plus haut, à savoir que l'approximation de notre ajustement est en gros satisfaisante, mais qu'un ajustement par une seule formule serait meilleur au point de vue de l'approximation s'il reproduisait cette bosse.

Pour montrer que notre ajustement est en gros satisfaisant, nous pouvons employer le critère invoqué par la Société suisse pour justifier son ajustement. Nous pouvons compter ainsi le nombre total de décès que notre ajustement aurait fait prévoir dans l'ensemble des exposés au risque de la Société.

On peut le calculer pour la période 0-83 ans; on trouve 4.924 décès au lieu de 4.841 observés, soit un écart de 16 ‰. Pour le comparer au résultat de la Société suisse, on doit prendre l'intervalle 0-80 ans pour lequel notre ajustement prévoit 4.913 décès et la Société suisse 4.000 au lieu de 4.830 décès observés, soit une erreur encore égale à 16 ‰ pour notre ajustement au lieu de 4 ‰ pour la Société suisse. Nous vérifions encore une fois qu'en ce qui concerne l'approximation numérique, notre ajustement est moins bon que celui de la table RAV lequel est à son tour comme précédemment moins bon que l'absence d'ajustement; double constatation à prévoir et qui n'est ni dans un cas, ni dans l'autre, suffisante pour en conclure la supériorité correspondante.

GRAPHIQUE I

*Reproduction du graphique publié par la Société Suisse d'Assurances générales
sur la vie humaine.*

TAUX ANNUELS DE MORTALITÉ

Observations 1921-1930, sans distinction de sexe ni d'année d'assurance.

Assurance populaire.

Taux de
mortalité

q_x
0,06

0,05

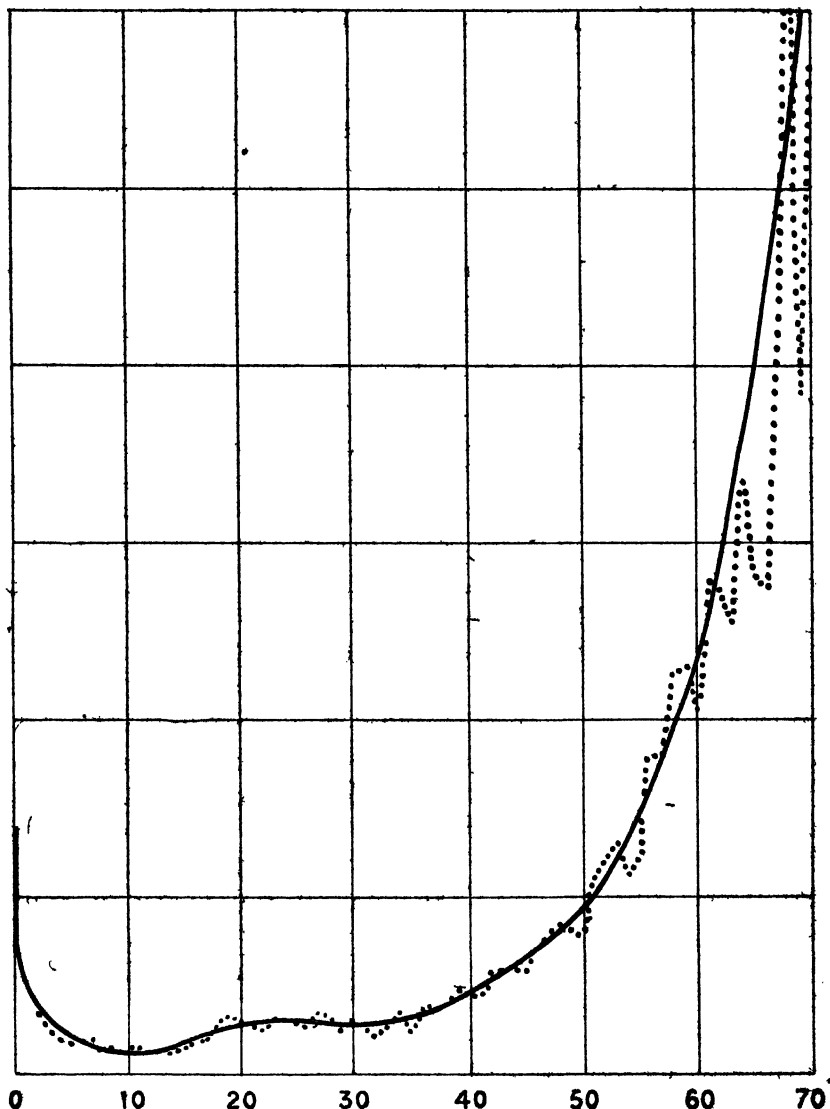
0,04

0,03

0,02

0,01

0



Age x

..... taux bruts.

——— taux ajustés par la table R.A.V.

Ce qui nous paraît plus important, c'est de constater que l'erreur globale donnée ainsi par notre ajustement est de moins de 2 %, c'est-à-dire d'un ordre de grandeur qui pourra généralement être admis dans la plupart des recherches démographiques tout en étant peut être insuffisant pour l'établissement des tarifs actuariels. Nous pensons que *pour le but* que nous avons en vue, le grossissement de l'erreur de 4 ‰ à 16 ‰ est largement compensé par la simplification obtenue en passant de la formule vie entière (2) à la formule vie entière (1). Et d'ailleurs la voie est libre pour une amélioration. Nous pensons que notre ajustement est inutilement précis pour les grands âges où les données sont trop imprécises et qu'en introduisant des poids égaux aux exposés au risque, on perdrait sans inconvénient une partie de cette précision inutile pour gagner de la précision du côté important de la bosse à faire apparaître et signalée plus haut plusieurs fois. C'est là un problème qui pourra tenter l'habile service de l'Actuariat de la Société suisse, lequel d'ailleurs pourrait chercher une autre amélioration en inventant une formule plus appropriée. Ayant ici montré que le difficile problème posé n'est pas sans espoir, que notre courbe reproduit assez bien, quoique encore insuffisamment, l'allure générale de la mortalité (de même que les précédentes de Thiele et de Steffensen), nous serions heureux si le présent travail pouvait inciter à entreprendre d'obtenir mieux encore.

II. — Table de la population féminine française.

Nous avons aussi utilisé les valeurs observées pour le taux annuel de mortalité q_x de la population féminine française en 1920-1921 telles qu'elles sont données dans le numéro de juillet 1928 du *Bulletin de la Statistique générale de la France*.

En posant $\eta' = \log q'_x + \text{constante}$, $\xi = \log(x + 1,25) - 1$ et représentant comme ci-dessus η' par un polynôme du 6^e degré en ξ , nous avons trouvé pour ces coefficients :

$$\begin{aligned} \alpha &= 2,497625; & \beta &= 0,608971; & \gamma &= 0,415098; & \delta &= -3,635284; \\ \varphi &= 3,360895; & \psi &= 3,921062; & \lambda &= -1,976273. \end{aligned}$$

Ils ont été obtenus par la méthode des moindres carrés, mais cette fois en introduisant pour chaque âge un poids égal au nombre d'exposés au risque.

En introduisant aussi ces mêmes poids pour le calcul des écarts E , σ des η définis comme ci-dessus, nous avons trouvé $E = 0,029123$, $\sigma = 0,04278$, valeurs moitié plus petites que pour la population d'assurés (1), qui nous ont donné le rapport :

$$\frac{E}{\sigma} = 0,681,$$

voisin de la valeur correspondante, 0,674, pour la loi de Laplace.

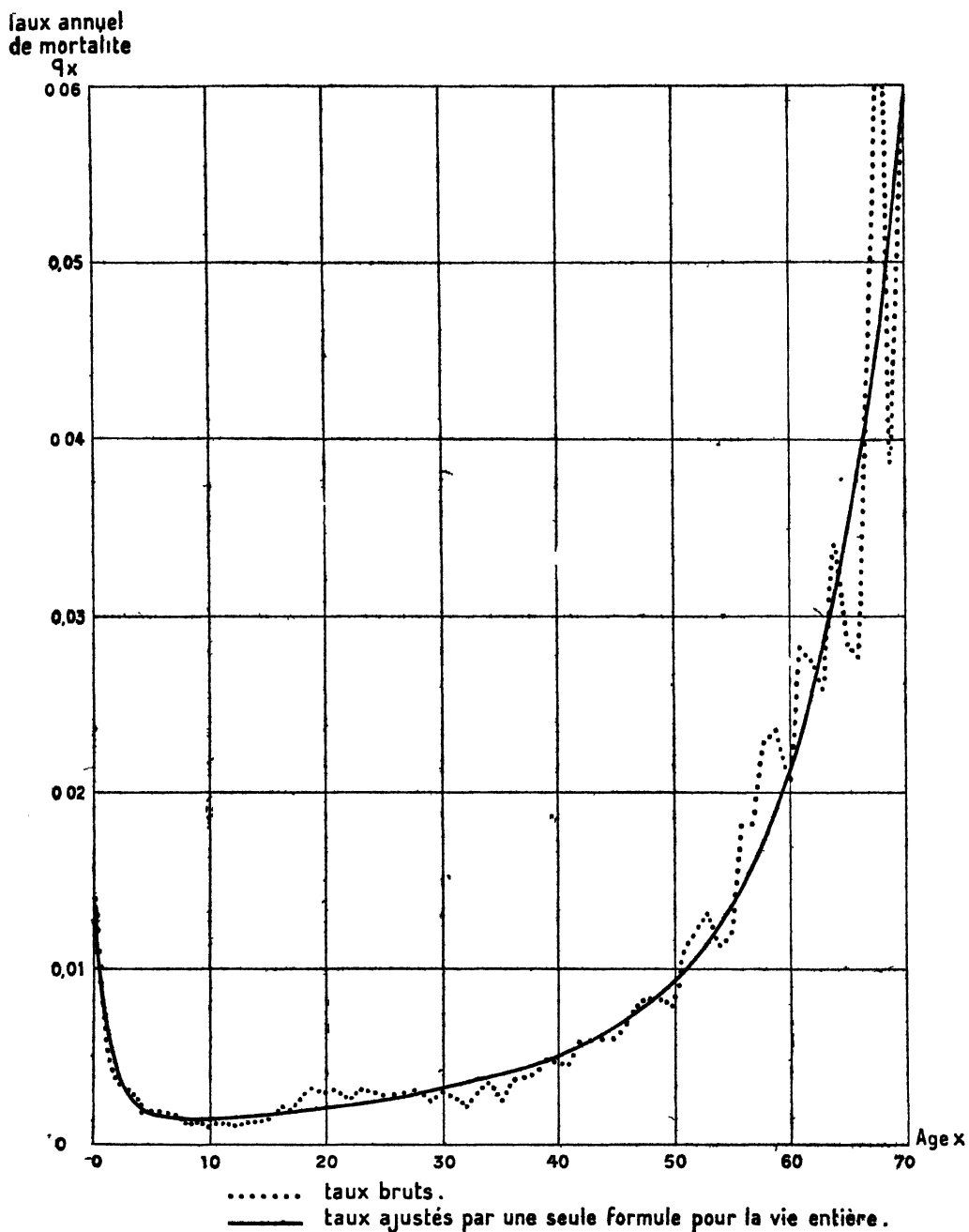
Les courbes représentatives correspondantes des taux observés et calculés sont données par le graphique III.

(1) Cette amélioration de l'approximation est peut-être due à l'emploi des poids, mais surtout, sans doute, à ce que la seconde population est beaucoup plus nombreuse et que par suite les taux observés varient plus régulièrement.

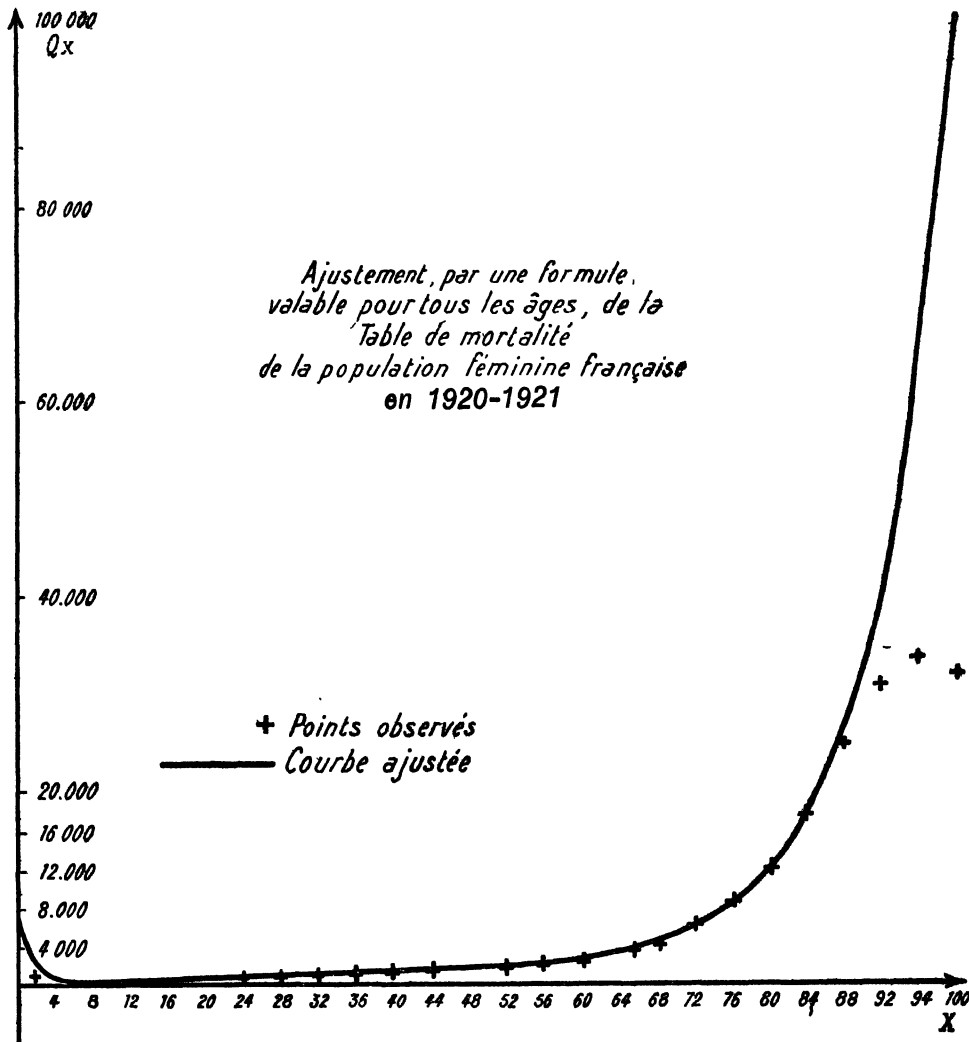
GRAPHIQUE II

*Courbe de mortalité des assurés (assurance populaire)
de la Société Suisse d'Assurances générales.*

Ajustement par une seule formule pour la vie entière.



GRAPHIQUE III



DISCUSSION

M. RISSER. — Le brillant exposé de M. Fréchet m'a vivement intéressé; aussi, suis-je convaincu que les recherches de notre éminent collègue dans ce domaine attireront certainement l'attention des statisticiens et des actuaires et suggéreront des travaux nouveaux. Qu'il me soit permis d'ajouter quelques remarques à la suite de sa très intéressante communication.

Faisant état des renseignements récapitulatifs de la brochure intitulée « Mouvement de la population. Les causes de décès », établie par les soins de la Statistique générale de la France, M. Duc, dans son étude sur une loi de mortalité d'origine statistique (voir *Bulletin de l'Institut des Actuaires français*, n° 162, septembre 1935), a dressé un tableau donnant, pour un groupe de 100.000 vivants à la naissance, la répartition des décès d'après leur cause, dans l'hypothèse où la population resterait constante; l'examen de ce tableau lui a suggéré une loi de mortalité.

Représentant graphiquement la répartition des décès d'après leur cause, telle qu'elle résultait du tableau déduit de la Statistique générale de la France, en portant en abscisses les âges et en ordonnées les nombres de décès, M. Duc a ainsi obtenu un certain nombre de courbes dont les plus importantes rappelaient par leur forme les courbes de Gauss. Il a été ainsi conduit à chercher si le nombre de décès à chaque âge dans un groupe complet de population pouvait être représenté par une somme d'équations de la forme de celle de Gauss, mais plus généralement par :

$$d(x) = \sum a_i e^{-b_i(x-x_i)^2},$$

dans laquelle $d(x)$ représente le nombre total des décès à chaque âge x , et les quantités a_i , b_i , x_i seraient des coefficients à déterminer.

A cet effet, l'auteur a utilisé la table CR brute, et constaté qu'une telle représentation pouvait être considérée comme satisfaisante; toutefois, il lui fallait examiner si la représentation simple reconnue comme suffisamment exacte pour les âges avancés de la vie pouvait être généralisée et étendue à la durée de la vie humaine.

C'est ainsi que M. Duc a été amené à porter à 5 le nombre total des groupes homogènes dans la population étudiée depuis la naissance, et à représenter le nombre des décès $d(x)$ par l'équation :

$$d(x) = a_1 e^{-b_1 x^2} + a_2 e^{-b_2(x-25)^2} + a_3 e^{-b_3(x-45)^2} + a_4 e^{-b_4(x-60)^2} + a_5 e^{-b_5(x-75)^2},$$

dans laquelle les coefficients a_i et b_i devaient être déterminés, et les constantes (25, 45, 60, 75) qui étaient évaluées d'une manière approximative, étaient susceptibles de modifications.

En définitive, cette fonction qui nécessitait l'introduction de 10 paramètres, avait l'avantage de s'appliquer à la durée entière de la vie humaine; elle permettait, de plus, de calculer l'annuité viagère à l'âge x et les primes pures qui s'y rattachent ainsi qu'aux tarifications des risques aggravés.

Dans une étude publiée dans le *Bulletin des Actuaire français (B. I. A. F.)* en 1909, nous avons étudié les formules représentatives du nombre des vivants aux différents âges.

Pour la période (0-12 mois), nous avons préconisé la formule :

$$v_n = 1000 - A \sqrt[n]{n} + n A',$$

(1.000 étant la population à l'âge 0 et n le nombre de mois).

Pour la période (1 an, 15 ans exclus), un polynôme du 4^e ou du 5^e degré, suivant en cela une méthode analogue à celle adoptée par les Compagnies françaises d'assurances sur la vie pour la table AF.

De 15 à 70 ans, c'est-à-dire pour la période *active* de la vie, nous avons pris pour v_x une fonction du type :

$$v_x = e^{-\beta'x - \beta''} - C_1 \int e^{(Bx + Cx^2 + \dots + Px^p)} dx,$$

déduite de l'expression du taux instantané de mortalité

$$\mu_x = \beta' + e^{Ax + Bx^2 + Cx^3 + \dots + Px^p},$$

basée sur l'extension de l'hypothèse de Gompertz.

On aboutissait ainsi, par une voie spéciale à un des groupes de formules de survivants trouvés par Quiquet dans son beau mémoire sur les lois de survie.

Alors que la méthode des moindres carrés a été appliquée à l'ajustement des taux de mortalité de 0 à 25 ans au moyen d'un polynôme du 6^e degré par le groupe des actuaires du Comité des Compagnies d'assurances en 1895, celle basée sur l'emploi des polynômes de Tchebycheff a été utilisée pour l'ajustement des mêmes taux de mortalité sous l'instigation de M. Fréchet par deux de ses élèves, et cette fois-ci en n'introduisant qu'un polynôme du 3^e degré.

Toujours en recourant à l'emploi des polynômes de Tchebycheff, notre collègue a procédé à une représentation du logarithme de l_x (assurés français), pour les têtes de 0 à 25 ans, au moyen d'un polynôme du 6^e degré.

En ce qui concerne les formules représentatives du nombre des vivants de 0 à ω (limite de la vie humaine), qui ont suscité de nombreuses études de statisticiens et d'actuaires, il nous faut rappeler spécialement, d'une part, celle de Thiele et, d'autre part, celle d'Oltamare.

La formule de Thiele se rattache à un développement limité du type $b_i \cdot a_i \sqrt{x}$:

$$\log V_x = c_v + c_1 \sqrt{x} + c_2 (\sqrt{x})^2 + \dots + c_{s-1} (\sqrt{x})^{s-1} + b_1 a_1 \sqrt{x} + b_2 a_2 \sqrt{x} + b_3 a_3 \sqrt{x}$$

Nous avons été amené, dans deux communications en 1927, à suggérer certaines remarques sur l'emploi de la méthode de Thiele, puis à analyser tout particulièrement un mémoire d'Oltamare concernant une formule relative au logarithme de la probabilité de vie :

$$\log p_x = \alpha_0 + \alpha_1 c^x + \alpha_2 c^{2x} + \beta_1 c^{-x} + \beta_2 c^{-2x};$$

Nous avons, dans une note parue aux comptes rendus de l'Académie des Sciences en 1927, montré que l'utilisation des formules :

$$(1) \log p_x = Q_n(x) + \alpha_1 c^x + \beta_1 c^{-x},$$

$$(2) \log p_x = Q_n(x) + \alpha_1 c^x + \alpha_2 c^{2x} + \beta_1 c^{-x} + \beta_2 c^{-2x},$$

[$Q_n(x)$ étant un polynôme], devait donner des résultats satisfaisants pour toute la durée (0, ω).

Si l'on prenait pour $Q_n(x)$, un polynôme du 3^e degré, l'utilisation de la formule (1) nécessite le calcul de 7 coefficients. Or, si l'on décompose la durée de la vie humaine en quatre périodes : (0-1 an), (1-25 ans exclus), (25, ω) et que l'on préconise respectivement pour ces périodes :

$$(0-12 \text{ mois}) \dots v_n = 1000 - A \sqrt[3]{n} - n A',$$

1 an-25 ans exclus, un polynôme du 3^e degré,

(25, ω), la formule $K S^x g^{cx}$, on devrait procéder au calcul de 10 coefficients.

Si l'on examine la forme des courbes de mortalité générale, et si l'on veut interpréter les deux points d'inflexion qui apparaissent sur la plupart d'entre elles, il serait peut-être nécessaire de substituer un polynôme de degré supérieur au 3^e. Il s'ensuit que, pour la représentation de v_x , l'on devrait introduire au minimum 10 constantes, alors que celle basée sur l'emploi de $\log p_x$ n'en ferait intervenir que 7 au minimum et 9 au maximum. Seule l'introduction du processus de Gompertz-Makeham a donné naissance à un type de lois de survie, lois qui ont été généralisées d'une manière profonde et heureuse par Quiquet;

peut-être un mode de raisonnement un peu différent de celui de Gompertz nous conduira-t-il à une représentation relativement simple et suffisamment approchée du nombre des vivants à tous les âges.

M. HUBER voudrait présenter deux observations au sujet de l'intéressante communication de M. Fréchet. Celui-ci a fait observer que certaines courbes de mortalité, notamment celles de la population du Japon, présentent deux points d'inflexion vers 30-32 ans. D'autres courbes relatives à la population masculine de certains pays offrent un mouvement beaucoup plus accentué : un maximum vers 23-24 ans, suivi d'un minimum avant la trentaine et la reprise de l'accroissement de la mortalité. C'est le cas pour la France (table 1898-1903, etc., jusqu'à 1920-1923, le mouvement s'atténue pour les tables postérieures), pour d'autres pays, notamment pour la Norvège où le phénomène est encore plus marqué. Cette anomalie ne se produit pas pour le sexe féminin.

Deuxième observation : M. Risser a signalé un essai de représentation de la loi de mortalité par la somme de deux ou trois courbes laplaciennes. Je voudrais rappeler que Karl Pearson a représenté la mortalité par âge depuis la conception, c'est-à-dire y compris la mortalité prénatale, par la somme de 5 courbes à un seul sommet, laplaciennes ou presque (*The chances of death*, vol. I, p. 26, Londres, Edward Arnold, 1897). Ainsi la courbe des décès masculins en Angleterre (1871-1880) est décomposée en 5 courbes : 1^o vieillards, courbe déviée, mode 72 ans, moyenne 67 ; 2^o adultes, courbe de Laplace, mode et moyenne 42 ans ; 3^o jeunes hommes, courbe de Laplace, mode et moyenne 23 ans ; 4^o enfants, courbe déviée, mode 3 ans, moyenne 6 ; 5^o nouveau-nés et mort-nés, courbe dont le maximum se place au début du dernier mois de la grossesse.

Il est intéressant de rappeler aussi la représentation pittoresque que K. Pearson a faite du phénomène, dans une gravure intitulée : *The bridge of life*.

Sur ce pont s'avancent, à la suite, un enfant, un jeune homme, un adulte, un vieillard. A l'entrée, un squelette écrasant un nouveau-né sous un crâne décharné, symbolise l'effet des tares héréditaires sur la mortalité prénatale et celle des premiers jours de la vie. La mortalité infantile (jusqu'à 1 an) est figurée par un autre squelette qui tire sur l'enfant avec une mitrailleuse. Un autre tire sur les adolescents avec un arc et des flèches, tir peu efficace qui correspond au minimum de la mortalité. Le tir d'un tromblon devient un peu plus meurtrier sur les adultes ; enfin, les hommes d'âge mûr essuient le feu d'un fusil moderne et la coupure brusque du pont figure la chute finale, inévitable.

M. DELAPORTE. — Toutes les formules d'ajustement du taux de mortalité selon l'âge indiquent que ce taux continue à croître jusqu'à la fin de la vie. Au contraire, les tables de mortalité expérimentales indiquent toujours une forte atténuation de la croissance du taux de mortalité au-dessus de 80 ans. Certaines tables, même, montrent une diminution de la mortalité brute à partir de cet âge. Nous oserions presque dire qu'on peut distinguer une table de mortalité ajustée sur une fonction mathématique d'une table de mortalité statistique en examinant si le taux de mortalité continue à croître régulièrement au delà de 80 ans. D'après environ 300 tables de mortalité que nous avons étudiées, il semble que les tables brutes n'indiquent pas d'augmentation de la mortalité après 80 ou 85 ans, tout au plus une légère croissance très infé-

rieure à celle des âges précédents. Une expression mathématique devrait donc tenir compte de ce ralentissement dans les âges supérieurs de la vie.

Les difficultés que M. Fréchet a rencontrées pour l'ajustement du maximum de mortalité suisse entre 20 et 30 ans nous semblent provenir de la nature même de ce maximum.

Dans les pays scandinaves, spécialement la Suède et la Norvège, il est une vaste boursoufflure de la surface des taux de mortalité en fonction de l'âge et de l'époque. Ce maximum du taux avec l'âge s'est développé vers la fin du XIX^e siècle, et a atteint son maximum avec l'époque vers 1900; depuis ce moment, il est en décroissance et a presque disparu maintenant. Si l'on examine les causes de décès dans ces pays, on constate que la boursoufflure des taux de mortalité correspond à l'évolution de la mortalité par tuberculose.

Dans d'autres pays d'Europe, tels que la France, les Pays-Bas et l'Italie, on observe un maximum des taux de mortalité masculins seulement, qui se situe de 21 à 23 ans, correspondant exactement aux âges du service militaire. Il semble que les conditions sanitaires que l'on rencontre à ce moment rendent mortelles des infections nouvelles ou acquises précédemment.

De toute manière, le maximum de mortalité que l'on observe entre 20 et 30 ans provient d'un phénomène, la tuberculose, qui s'ajoute au vieillissement normal des individus pour produire la mort.

Il semble donc que l'on doive utiliser une fonction mathématique supplémentaire pour représenter cette mortalité. Ceci nous amène ainsi à envisager que la mortalité de l'ensemble de la vie doit mathématiquement être formée par la somme de plusieurs expressions correspondant à chacune des phases physiologiques de la vie :

1^o La mortalité embryonnaire ou fœtale;

2^o La mortalité infantile, due à la transformation violente du mode de vie de l'enfant à la naissance, et à ses premières contaminations microbiennes.

Le taux de mortalité est d'autant plus faible qu'on s'éloigne davantage de la naissance;

3^o Une mortalité infectieuse ayant son maximum entre 20 et 30 ans;

4^o Le vieillissement qui fait croître le taux de mortalité avec l'âge selon la loi de Makeham;

5^o Un ralentissement de la mortalité par vieillissement au delà de 80 ou 85 ans.

M. HUBER, à propos de la mortalité aux âges élevés, exprime l'avis qu'il est difficile de porter un jugement sur son allure, au moins pour la population totale d'un pays. Les nombres de vivants et de décès à ces âges sont faibles et peu précis, surtout dans les pays où, il y a quatre-vingts ans, l'état civil n'existait pas ou seulement dans une forme rudimentaire, comportant beaucoup d'incertitude, ce qui explique, d'autre part, le nombre anormal de prétendus centenaires enregistrés au recensement dans certains de ces pays.

Note de MM. SCHWARTZ et CUZIN. — Nous avons été frappés de l'emploi fait par M. Fréchet des *logarithmes* dans sa courbe de mortalité et l'interprétation biologique donnée par M. Bourgeois nous semble judicieuse. Nous pensons qu'il serait intéressant de rapprocher cette observation des résultats auxquels conduit l'étude statistique de plusieurs paramètres biologiques.

Ainsi, la répartition des tailles dans une population de conscrits est sensiblement gaussienne; c'est un fait si bien connu qu'il finit par sembler intuitif. Or, si on étudie les poids de ces mêmes conscrits, on constate que leur courbe de répartition s'écarte d'une courbe de Gauss par une dissymétrie très nette (le mode est déporté à gauche de la moyenne). Cette différence entre le comportement du poids et celui de la longueur peut étonner, ces deux paramètres ayant chacun autant de raisons *a priori* d'être, ou de ne pas être, distribué en loi de Gauss; mais elle s'explique tout à fait d'après la loi d'allométrie de Teissier et Huxley : entre deux paramètres x et y , à peu près quelconques (de dimensions, chimiques, etc...), d'un individu existe la relation $y = kx^\alpha$, k et α restant constants pendant la période de croissance pour les individus de même type. Par suite, si un paramètre était distribué gaussiennement, tous ceux qui sont liés à lui par une relation d'allométrie où α diffère de 1, ne sauraient être distribués gaussiennement. Il est donc impossible avec une seule forme de loi (gaussienne ou non) de rendre compte de la distribution de l'ensemble de ces paramètres.

Au contraire, si l'on choisit comme paramètres les *logarithmes* des mesures, la loi d'allométrie indique clairement que ceux-ci ne diffèrent les uns des autres que par un facteur multiplicatif et une constante additive; ils ont alors tous la même loi de distribution. Cette seule considération devrait conduire à utiliser toujours les logarithmes, sous la seule condition que les paramètres envisagés vérifient la loi d'allométrie, ce qui est le cas général.

Or, une deuxième considération vient s'ajouter à celle-ci pour inciter à l'emploi des logarithmes : c'est que cette distribution unique semble bien être une distribution de Gauss; nous l'avons vérifié à plusieurs reprises dans un domaine assez large (feuilles de tabac, population de conscrits, etc...).

Les interprétations existantes de la loi d'allométrie, d'autre part, viennent confirmer de façon théorique ces vérifications. En effet, pour un paramètre x , la loi de croissance serait :

$$\frac{dx}{x} \text{ (ou } d\mathcal{L}x) = kda$$

a désignant l'« alimentation » dans son sens le plus large. Or, on conçoit beaucoup mieux la distribution gaussienne pour ce facteur alimentation que pour la longueur ou le poids d'un individu, d'abord parce qu'il est de caractère additif, ensuite parce qu'il est causal dans le morphogénèse, tandis que la longueur ou le poids sont des effets.

En conclusion, nous pensons qu'on peut, en règle presque générale, utiliser à des fins statistiques le *logarithme* des paramètres biologiques. L'intérêt est, d'une part, que la majorité des paramètres intéressant un individu vérifient la même loi de distribution, ce qui permet de ne faire qu'une seule étude d'ajustement, d'autre part, que cette distribution semble être, en général, comme des motifs théoriques l'expliquent assez bien, une distribution de Gauss. Ceci s'appliquerait aussi bien à des caractères dimensionnels ou chimiques des individus, qu'à des poids de récolte en vue de comparer des engrais, etc...

Dr MALTHETE. — Nous avons admiré le travail de M. Fréchet et nous pen-

sons comme lui que la double inflexion de la courbe de mortalité générale peut s'expliquer en partie par la surmortalité tuberculeuse aux âges correspondants.

Nous voudrions demander à M. Fréchet s'il ne croit pas que, dans la construction mathématique d'une courbe de mortalité française, on devrait essayer de tenir compte du facteur aisance ou pauvreté? Nous avouons notre complète ignorance en matière d'assurance sur la vie, mais nous voudrions savoir s'il existe des tables de survie tenant compte de la classe sociale des assurés. Les primes demandées sont-elles semblables, au même âge, pour un agriculteur et pour un médecin en bonne santé?

Nous nous souvenons d'un travail de Hersch sur « l'inégalité devant la mort d'après les statistiques de la ville de Paris » et nous croyons qu'il y aurait peut-être intérêt à construire plusieurs courbes de mortalité se rapportant à des classes sociales plus homogènes. Peut-être s'apercevrait-on que les courbes obtenues pourraient être construites plus facilement.

M. FRÉCHET. — Je remercie vivement les membres de la Société qui ont bien voulu accroître l'intérêt que présente le sujet de mon exposé par leurs importantes interventions.

MM. Risser et Huber ont apporté une contribution essentielle à mon exposé historique.

Retenons d'abord de l'exposé de M. Risser que, d'après lui, ses formules (1) et (2) devraient donner des résultats satisfaisants pour la vie entière. Et souhaitons que lui-même ou un jeune chercheur en fasse l'essai numérique sur une population déterminée.

Il est, d'autre part, intéressant de noter que le travail de M. Duc, signalé par M. Risser, paraît avoir été inspiré par les mêmes idées que le mémoire antérieur de K. Pearson cité par M. Huber. D'une part, comme il semble que M. Duc se soit contenté d'indiquer la forme générale de sa loi de mortalité, le travail de Karl Pearson a l'avantage de s'être traduit par une formule complète, c'est-à-dire à coefficients numériques dûment calculés et relatifs à une population déterminée (en l'espèce la population française) [1]. Mais, d'autre part, K. Pearson ne donne pas un fondement scientifique précis du choix de sa formule; tandis que M. Duc, en dressant les courbes de mortalité relatives à diverses causes de décès et montrant qu'elles se rapprochaient de la forme de Laplace, a justifié le principe de leur méthode. Ici se pose une question: si l'on veut, comme le recommande aussi M. Delaporte, représenter une fonction de la mortalité à l'âge x par une somme de termes dont les maxima ont lieu vers les époques assez similaires choisies respectivement par K. Pearson, M. Duc et M. Delaporte, quelle est la fonction à choisir? Sera-ce le taux de mortalité annuel q_x , le taux de survie annuel, p_x , la probabilité pour un nouveau-né de mourir, d_x , ou de survivre, l_x , à l'âge x ? En tout cas, K. Pearson, puis M. Duc, ont fait leur choix: pour eux c'est d_x .

Encore du point de vue historique, je sollicite un complément d'information.

(1) K. Pearson ne donne que peu de détails sur la réalisation numérique de son ajustement. Grâce à l'obligeance de son fils, le Professeur Egon Pearson, nous pouvons ajouter qu'on trouvera des détails complémentaires dans l'exemple IV figurant à la fin du mémoire intitulé *Contributions to the Mathematical Theory of evolution, II: Skew Variation in homogeneous Material*, publié par Karl Pearson dans les *Philosophical Transactions of the Royal Society, Séries A*, vol. 186 (1895), pp. 343-414.

L'un des auditeurs a cru se souvenir, sans en être certain, que dans un ouvrage en français de M. Lotka, l'auteur citerait une formule de mortalité valable vie entière et due à un auteur hollandais. Je souhaite qu'un des lecteurs de ce journal confirme et complète cette information et plus généralement que me soit adressée, par l'intermédiaire de notre dévoué secrétaire, M. Depoid, toute information relative à d'autres formules vie entière, parmi celles qui ont été éprouvées numériquement.

J'avais constaté et signalé l'existence d'une bosse vers les 30 ans sur un certain nombre de représentations graphiques du taux annuel de mortalité. Mais n'étant pas démographe de profession, mes observations ne pouvaient porter sur un aussi grand nombre de cas que celles de MM. Risser, Huber et Delaporte. Il m'est donc très précieux de pouvoir affirmer avec plus de force encore, grâce à leur témoignage, que cette bosse n'est pas due à un zigzag accidentel mais correspond à une réalité physiologique (qui, d'ailleurs, comme le suggèrent MM. Huber et Delaporte aurait tendance à s'effacer avec le temps).

En ce qui concerne les âges élevés, les interventions de MM. Delaporte et Huber m'inciteraient à penser que la question doit rester posée. D'une part, M. Delaporte constate que les tables brutes de mortalité présentent un taux de mortalité qui, contrairement à toutes les formules proposées, croît beaucoup moins rapidement après 80 et 85 ans que dans les âges élevés antérieurs ou même diminue. Mais, d'autre part, M. Huber nous rappelle que les données brutes après 80 ou 85 ans sont relatives à de petits nombres et surtout que leur exactitude est très suspecte.

Existe-t-il, demande le Dr Malthête, des tables de survie tenant compte du facteur aisance ou pauvreté? On peut, en réponse partielle, dire qu'en tout cas, les tables de survie des C^{ies} d'assurances sur la vie concernent des personnes plus aisées que la moyenne de celles qui constituent la population française pour laquelle il existe aussi une table de survie.

L'intervention de MM. Schwartz et Cusin est très suggestive. Il est clair qu'entre des variables statistiques associées comme la taille T et le poids P d'un individu doit exister une relation approximativement de la forme $\frac{T}{P^\alpha} = \text{const.}$, α étant ici égal à 3. D'où l'idée de généraliser, de prévoir la loi d'allométrie et ensuite de la vérifier dans un grand nombre de cas comme l'ont fait Teisner et Huxley.

Quand la loi s'applique, le choix du logarithme de la variable paraît s'imposer. Il faut être reconnaissant à MM. Schwartz et Cusin de nous signaler l'intérêt de ce choix et de nous informer des résultats de leurs observations.

Toutefois, dans le cas actuel, et bien que leur intervention ait pour but d'approuver le choix que j'avais fait du log. de l'âge comme variable, je me demande si la loi d'allométrie doit être invoquée ici. On ne voit pas trop quelles grandeurs physiologiques seraient proportionnelles au carré ou au cube de l'âge et même dans ce cas comment elles auraient à être associées aux tables de mortalité?

Pour conclure, je signale aux chercheurs les questions suivantes dont l'opportunité paraît se dégager de la discussion :

1° Réunir tous les renseignements bibliographiques concernant les formules

de mortalité valables vie entière autres que celles qui ont été signalées plus haut et choisies parmi celles qui ont été éprouvées numériquement, chacune au moins sur une table brute.

2° Comparer les valeurs de ces formules en tenant compte *simultanément* du nombre n de constantes qu'elles emploient et de l'approximation « qu'elles fournissent (1);

3° Chercher à faire mieux;

4° Faire une étude spéciale de la mortalité *brute* après 80 ans en la basant sur l'étude de collectivités où l'on ait des données certaines sur l'âge des individus. Par exemple, en considérant des personnages d'une situation sociale telle que les dates de naissance et de mort aient été publiées.
