MOURRE

La formule de Pareto comme instrument statistique

Journal de la société statistique de Paris, tome 82 (1941), p. 85-108

<http://www.numdam.org/item?id=JSFS_1941__82__85_0>

II

LA FORMULE DE PARETO

COMME INSTRUMENT STATISTIQUE [1]

Vilfredo Pareto a donné sur la distribution des revenus une formule célèbre :

$$N = \frac{A}{x^\alpha},$$

où A et α sont des paramètres, constants pour une année et un pays déterminés.

On a beaucoup reproché à cette formule de ne donner que des résultats approximatifs.

Il faut d'abord marquer à ce sujet les points de vue différents qui séparent les statisticiens des économistes. Le statisticien cherche à s'appuyer sur des formules aussi précises que possible; il est satisfait quand une formule lui permet de retrouver d'une manière presque exacte les données de la statistique. Le point de vue du statisticien, qui ne considère que les limites de sa science, est certainement juste. Mais l'économiste vise souvent un autre but. Il se préoccupe moins de la perfection plus ou moins grande de la formule que de la lumière qu'elle peut projeter sur certains phénomènes qui lui étaient inconnus. Même, si la formule lui donne simplement le sens des variations qu'il n'aurait pu connaître sans elle, il estimera qu'il aura fait un pas en avant.

La formule de Pareto est pour les économistes un instrument de recherche singulièrement puissant. Elle projette sur la structure sociale un jour nouveau. Dans deux communications antérieures à la *Société de Statistique*, l'une en 1922, l'autre en 1929, j'ai montré que le revenu moyen croissait ou décroissait en même temps dans toutes les classes sociales et que, par conséquent, les riches et les pauvres étaient liés dans la prospérité et la misère. J'ai démontré également que l'inégalité sociale s'accroissait, à mesure que la richesse générale augmentait.

Je ne reviendrai pas aujourd'hui sur ces importantes conclusions, auxquelles peut conduire la formule de Pareto. Je vais simplement vous montrer, en appliquant la formule de Pareto à l'étude des revenus français en 1931 et à celle des revenus prussiens en 1913, que, bien qu'approximative, l'équation de Pareto, en raison de sa simplicité, peut rendre au statisticien les plus grands services.

Je parle ainsi de la formule simple de Pareto $N = \frac{A}{x^\alpha}$, car il existe deux autres formules de Pareto d'une précision plus grande :

$$N = A\,(x + a)^{-\alpha}\, e^{-\beta x}$$
$$N = A\, x^{-\alpha}\, e^{-\beta x}$$

(1) Communication faite le 20 novembre 1940.

A mon avis, elles n'ont pas grand intérêt, en raison de leur complication et il est préférable de leur substituer, si on cherche, en renonçant à la simplicité, des résultats plus exacts, d'autres formules qui sont meilleures.

Je citerai d'abord la belle formule de notre collègue, M. Gibrat :

$$y = \frac{1}{\sqrt{\pi}} e^{-z^2}$$

$$z = a \log (x - x_0) + b.$$

y est le rapport du nombre de personnes ayant un revenu déterminé au nombre total des possesseurs de revenus; x_0 représente le revenu minimum.

Cette formule donne des résultats beaucoup plus exacts que celle de Pareto; elle s'applique très bien aux cas où l'équation de Pareto ne donne que des résultats insuffisamment approchés, en particulier aux successions. En outre, son champ d'application est très large et s'étend à des répartitions variées en dehors du domaine de l'économie politique. Elle a donné des résultats remarquables dans certains problèmes de sciences physiques et naturelles pouvant être considérés comme liés à des phénomènes statistiques.

Malheureusement, pour ce qui concerne la répartition des fortunes, la formule de Gibrat a deux paramètres, dont un seul a possède une signification économique.

a joue dans cette formule le même rôle que α dans l'équation de Pareto. Quand a décroît, l'inégalité, telle que je l'ai définie, augmente et le revenu moyen s'accroît. Mais il reste un autre paramètre, b, auquel on ne trouve aucune signification économique et qui influe sur les variations de a. Il en résulte que toute déduction économique devient de ce fait imprécise.

Certes, la formule de Pareto a aussi deux paramètres, α et A. Mais ces deux paramètres présentent une signification économique.

α exprime à la fois la richesse et l'inégalité des fortunes. A varie proportionnellement à la population. Il est en outre fonction de α, c'est-à-dire que le nombre de personnes ayant un revenu déterminé varie entre un revenu minimum et un revenu maximum arbitrairement choisis en même temps que la richesse générale.

De plus, A s'élimine la plupart du temps dans les calculs. Il n'intervient ni dans la formule du revenu moyen, ni dans celle du rapport d'inégalité, tel que je l'ai défini, c'est à-dire du rapport de la somme des fortunes (ou des revenus) dépassant la fortune moyenne (ou le revenu moyen) à la somme des fortunes (ou des revenus) inférieurs à la fortune moyenne (ou au revenu moyen).

Une formule plus simple que celle de M. Gibrat et qui a probablement des avantages et des applications du même ordre a été proposée par notre Secrétaire général, M. Barriol.

Le nombre de personnes ayant un revenu supérieur à x est donné par l'équation :

$$\mathrm{N} = \frac{\mathrm{A}}{(x + a)^{\alpha}} - k.$$

Cette formule est beaucoup plus exacte que celle de Pareto et elle a le précieux avantage d'être intégrable.

En outre, elle ne conduit pas aux résultats absurdes que donne la formule de Pareto, quand on l'emploie jusqu'à la limite inférieure ou supérieure.

Alors que dans la formule de Pareto, il y a toujours une fraction d'individu ayant un revenu supérieur à n'importe quel chiffre, on peut écrire au moyen de la formule de Barriol que le nombre de personnes ayant un revenu supérieur au revenu maximum est nul. On a en effet :

$$N = \frac{A}{(M + a)^\alpha} - k = 0.$$

Dans la formule de Pareto pour $x = 0$, on a N infini.

Dans la formule de Barriol, on a, pour $x = 0$, :

$$N = \frac{A}{a^\alpha} - k.$$

Ce sont là des propriétés intéressantes.

En outre, les paramètres a et k ont peut-être une signification économique. En général, dans les pays où le taux de l'impôt est très lourd et très progressif, les détenteurs de grosses fortunes fraudent beaucoup. C'est peut-être une des causes qui font que la formule de Pareto est, en général, défectueuse du côté des gros revenus. Les statistiques fiscales indiquent donc, la plupart du temps, un nombre de contribuables inférieur à celui donné par la formule. Le paramètre k, qui influe beaucoup sur le chiffre des possesseurs de gros revenus et très peu sur celui des possesseurs de petits revenus, permet donc d'ajuster les résultats de la statistique avec ceux de la formule.

D'autre part, le paramètre a permet d'opérer un redressement de la courbe du côté des petites fortunes. En effet, a influe beaucoup sur le nombre des possesseurs de petits revenus et bien moins sur celui des possesseurs de gros revenus.

Or, en général, la statistique donne, de même que pour les gros revenus, des chiffres trop faibles pour les détenteurs de petits revenus. En effet, si les possesseurs de gros revenus pratiquent souvent une large fraude, beaucoup de détenteurs de petits revenus ne font aucune déclaration.

L'intéressante formule de M. Barriol mérite donc d'être étudiée dans ses applications statistiques. Je n'ai pas encore eu le temps d'entreprendre ces recherches.

Enfin, il existe une autre formule, donnée par notre collègue, M. Fréchet.

Cette formule, qui se déduit de la formule simple de Pareto, est :

$$\frac{R}{N x} = \text{constante} = \frac{\alpha}{\alpha - 1}$$

R, dans cette formule, représente la somme des revenus supérieurs à x, N le nombre des personnes ayant un revenu supérieur à x. $\frac{R}{N}$ représente la moyenne des revenus supérieurs à x et $\frac{R}{N x}$ égale le rapport de la moyenne des revenus supérieurs à x à ce revenu x.

Cette formule permet de vérifier d'une manière simple l'exactitude de la loi de Pareto. En outre, M. Fréchet signale, d'après M. Gibrat (1), que je cite n'ayant pas en mains le texte même de M. Fréchet, que cette équation se prête à l'étude de répartitions analogues à celles des revenus, telles que « celles de la population des villes suivant leur importance ». D'intéressantes recherches peuvent être entreprises dans cette direction.

Remarquons du reste que, si on prend un revenu maximum infini, la formule de Fréchet, de même que celle de Pareto, n'est plus applicable, quand α a une valeur voisine de 1.

Dans l'équation de Pareto, si le revenu maximum est infini, le revenu moyen est égal à $\left(\frac{\alpha}{\alpha - 1}\right) m$, m étant le revenu minimum. Ce revenu tend vers l'infini, quantd α tend vers 1.

Au contraire, si on étudie la formule de Pareto dans un intervalle déterminé, c'est-à-dire avec un revenu maximum fini, ainsi qu'il en est toujours dans la réalité, le revenu moyen est donné par la formule :

$$r = r_2 \frac{\alpha}{\alpha - 1} \cdot \frac{\rho^{\alpha - 1} - 1}{\rho^{\alpha} - 1}$$

r_2 étant le revenu maximum, ρ étant le rapport $\frac{r_2}{r_1}$ du revenu maximum au revenu minimum.

Or la quantité

$$\frac{\alpha}{\alpha - 1} \cdot \frac{\rho^{\alpha - 1} - 1}{\rho^{\alpha} - 1}$$

peut s'écrire ;

$$\frac{\alpha}{\rho^{\alpha} - 1} \cdot \frac{\rho^{\alpha - 1} - 1}{\alpha - 1}.$$

Quand α tend vers 1, le premier facteur de cette équation tend vers $\frac{1}{\rho - 1}$. On démontre que le second facteur tend vers le logarithme népérien de ρ.

Il ne faut pas croire du reste que les valeurs de α, égales à 1 ou voisines de 1, ne s'appliquent qu'à des cas théoriques, qu'on ne rencontre jamais dans la pratique. Pareto travaillant sur des documents du XVe siècle concernant la ville de Bâle, a trouvé, pour une époque où la richesse générale était vraisemblablement bien moindre qu'aujourd'hui, des valeurs de α légèrement supérieures à 1, ce qui, semblait il, décélait une fortune moyenne considérable. Il en parut surpris et douta de l'exactitude des statistiques utilisées pour ses recherches. Le fait n'a cependant rien d'extraordinaire. Pareto n'avait pas remarqué que les valeurs de α concernant les fortunes étaient toujours beaucoup plus faibles que celles s'appliquant aux revenus. Moi même, j'ai trouvé pour les sucessions françaises de 1929 une valeur de α à peine supérieure à 1. Les valeurs obtenues par Pareto pour Bâle sont plus élevées et correspondent à une richesse moindre. Les villes, du reste, au moyen âge, étaient habitées par une bourgeoisie commerçante assez riche et par des artisans aisés; les campagnes surtout étaient pauvres.

(1) GIBRAT, *Les Inégalités économiques*, pp. 44 et 45 (Librairie du Recueil, Sirey).

Remarquons encore que la formule de Pareto s'applique beaucoup mieux aux revenus qu'aux successions. Mais l'importance économique des revenus est plus grande que celle des capitaux. Les revenus, en effet, embrassent toute la richesse d'un pays; ils comprennent non seulement le produit de la fortune acquise, mais aussi celui beaucoup plus considérable du capital humain, c'est-à-dire les traitements, salaires et rémunérations de services divers. La formule de Pareto s'applique mieux au tout qu'à la partie.

Il est bien évident du reste que l'équation de Pareto, tout au moins quand on veut faire d'elle un instrument de mesure, doit être employée dans les genres de répartition où elle se montre à peu près exacte, plutôt que dans ceux où son approximation est très défectueuse.

En résumé, l'équation de Pareto présente sur les formules diverses qu'on a cherché à lui substituer l'avantage d'une plus grande simplicité. Elle est donc beaucoup plus maniable, mais elle a l'inconvénient d'une moins grande exactitude.

***

Détermination de α :

— Un point très important, qui ne paraît pas avoir suffisamment attiré l'attention des chercheurs, est la méthode de détermination de l'exposant α. Sans doute, si la formule était parfaite, toutes les méthodes devraient conduire au même résultat. Mais il n'en est pas ainsi et les valeurs trouvées pour α diffèrent d'une manière très large selon le mode de calcul adopté.

Une méthode, qui paraît très séduisante au premier abord, est celle qui consiste à cacluler α en partant du revenu moyen, pris dans un intervalle fini.

On se sert de la formule :

$$r = r_2 \frac{\alpha}{\alpha - 1} \cdot \frac{\rho^{\alpha - 1} - 1}{\rho^{\alpha} - 1}.$$

Les documents statistiques nous fournissent le revenu moyen r, le revenu minimum r_1, le revenu maximum r_2 et, par conséquent, le rapport ρ du revenu maximum au revenu minimum.

Quant au paramètre A, il sera calculé au moyen de la formule du revenu total dont le montant T nous est également fourni par la statistique.

Cette formule est :

$$T = \frac{A\,\alpha}{\alpha - 1} \cdot \left[\frac{1}{r_1^{\alpha - 1}} - \frac{1}{r_2^{\alpha - 1}}\right].$$

La méthode est donc rapide et simple; elle nous permet, en donnant lieu à une sorte de pondération des erreurs, d'obtenir une courbe de Pareto qui donne bien l'image de la répartition d'ensemble des revenus. Elle a toutefois l'inconvénient de faire négliger la dernière tranche, dont on ignore le revenu maximum que les statistiques fiscales n'indiquent pas. Mais cette dernière tranche n'exerce sur le revenu moyen qu'une influence insignifiante. Ainsi pour les revenus français de 1931, elle n'est que de 1 milliard environ, ce qui est peu par rapport aux totaux de 60 milliards environ avant déduction des charges et de 50 milliards environ après déduction des charges.

La méthode d'évaluation de α par le revenu moyen paraît donc tentante, mais cependant elle peut conduire à de graves erreurs, car la manière dont sont établis les documents officiels peut faire varier le revenu moyen dans une large mesure.

Prenons pour exemple les revenus français en 1931, que je viens de citer. L'année 1931 a été choisie au hasard. L'Administration des Finances nous donne une première évaluation des revenus faite avant l'application des déductions pour situation et charges de famille, ce qui donne un total de revenus de 59 milliards 823 millions environ. Après déduction pour situation et charges de famille, le nombre des contribuables étant resté le même, le revenu total s'abaisse à 49 milliards 910 millions environ, ce qui montre un écart de 10 milliards environ entre les deux évaluations. Le revenu moyen s'abaisse de 28.276 francs à 23.511 francs. Quant à l'exposant α calculé pour le revenu moyen, il s'élève de 1,45 à 1,687, dernière tranche négligée. On se trouve donc, en apparence, en présence de deux groupements très différents en richesse, alors qu'en réalité il s'agit toujours du même groupement.

Remarquons toutefois que la différence entre l'évaluation des revenus n'a nullement modifié l'ordre dans lequel sont rangés leurs détenteurs. Nous sommes donc amenés à nous servir pour le calcul de α, non pas du revenu moyen, mais de l'ordre de répartition des contribuables.

Choisissons une des méthodes qui nous permet, d'après cet ordre, de calculer α, par exemple celle des moindres carrés. On trouve $\alpha = 1,791$, ce qui correspond, en négligeant la dernière tranche, à 22.055 francs de revenu moyen au lieu de 23.511 francs, chiffre fourni par la statistique, dernière tranche négligée, déduction faite des charges de famille.

La méthode de calcul de α, d'après l'ordre dans lequel sont rangés les contribuables, nous permet de retrouver, à peu de chose près, le revenu moyen calculé après déduction des charges de famille.

Or quels sont les revenus réels? Sont-ce ceux qui sont établis avant la déduction des charges, ou ceux établis après la déduction des charges?

Ce sont évidemment les seconds. Bien entendu, il ne peut s'agir ici que de revenus nets. Or que faut-il entendre par un revenu net de 20.000 francs par exemple? C'est un revenu constitué par une somme de 20.000 francs, qui permet à son détenteur de se procurer d'abord ce qui est indispensable pour vivre, c'est-à-dire nourriture, logement, habillement, chauffage, etc..., et ensuite de faire quelques dépenses de luxe, s'il lui reste un excédent.

Mais il est bien évident que, si ce détenteur d'un revenu de 20.000 francs a à sa charge sept enfants en bas âge, ce n'est plus 20.000 francs qu'il aura à sa disposition, mais une somme beaucoup plus faible et, avec le pouvoir d'achat actuel de la monnaie, il sera réduit à la misère.

Les revenus doivent donc être considérés après déduction des charges de famille. Celles-ci pèsent surtout sur les petits revenus et n'exercent sur les gros revenus qu'une très faible influence. L'entretien d'enfants en bas âge ne constitue qu'une charge insignifiante pour les détenteurs d'un million de revenus.

Aussi, l'Administration fiscale ne fait-elle, avec sagesse, subir aux gros revenus qu'une déduction insignifiante du fait des charges de famille. Elle abaisse en effet la première tranche de la somme des revenus, celle qui va de 10.000 à 20.000 francs, de 25 milliards environ à 19 milliards environ, tandis que la tranche au-dessus de un million, de l'ordre de grandeur de un milliard, ne subit qu'une déduction de 4 millions environ.

L'ordre dans lequel sont rangés les possesseurs de revenus nous oriente donc vers la répartition la plus réelle et il est probable que cette constatation ne s'applique pas seulement aux revenus français, mais qu'elle est d'ordre général.

C'est donc cette répartition des contribuables le long de l'échelle des revenus qui doit nous servir à déterminer α.

Mais là encore nous avons le choix entre plusieurs méthodes. Il y a d'abord la méthode graphique.

On tire de l'équation $N = \frac{A}{x^\alpha}$

$$\log N = \log A - \alpha \log x.$$

Si les points représentatifs correspondant aux différentes valeurs envisagées pour log N sont en ligne droite, c'est que la formule de Pareto s'applique bien. On trouve ainsi pour les revenus de 1931 : $\alpha = 1{,}81$.

Cette méthode a l'avantage de la rapidité et de la simplicité, mais elle ne peut être employée dans les cas où la formule de Pareto s'applique mal et où les différents points ne forment pas une ligne tout à fait droite. On sera alors obligé de faire passer entre les points trouvés pour les différentes valeurs de N une ligne droite à laquelle on peut donner plusieurs positions différentes, ce qui entraînera beaucoup d'arbitraire et rendra très peu sûres les comparaisons d'une année à l'autre.

Il est donc, dans ce cas, nécessaire de substituer à la méthode graphique une méthode rigide qui fixe la valeur de α au moyen d'un procédé invariable, d'où l'arbitraire sera exclu, par exemple la méthode des moindres carrés, qui a pour but de réduire les erreurs au minimum.

Mais la méthode des moindres carrés est très longue et je propose de lui substituer une méthode beaucoup plus rapide qui conduit à des résultats à peu près identiques.

Ce qui empêche la formule de Pareto d'être exacte, c'est qu'on ne trouve pas dans toutes les parties de la courbe tracée au moyen des données statistiques une valeur identique pour α. En réalité, la courbe obtenue n'est pas une courbe parfaite de Pareto, mais peut être décomposée en une suite de courbes qui sont, sinon des courbes parfaites de Pareto, du moins des courbes qui s'en éloignent peu. Il n'y a donc qu'à prendre les différentes valeurs trouvées pour α dans chaque partie de la courbe et à en faire la moyenne arithmétique.

Soient les revenus x_1, x_2, etc... N_1 le nombre de personnes ayant plus que x_1 de revenus, N_2 le nombre de personnes ayant plus que x_2, etc... On tire de l'équation de Pareto l'équation :

$$\alpha = \frac{\log N_1 - \log N_2}{\log x_2 - \log x_1}$$

$$\alpha = \frac{\log N_2 - \log N_3}{\log x_3 - \log x_2}, \text{ etc...}$$

Nous obtiendrons ainsi pour les revenus français de 1931 les valeurs de α suivantes :

			α
De 10.000	à	20.000	1,675
20.000	à	30.000	1,759
30.000	à	40.000	1,751
40.000	à	50.000	1,731
50.000	à	100.000	1,657
100.000	à	200.000	1,686
200.000	à	500.000	1,971
500.000	à	1.000.000	2,100

L'α de la dernière tranche, au-dessus de 1 million, n'a pu être calculé, les statistiques fiscales ne nous indiquant pas le revenu maximum.

En prenant la moyenne arithmétique de ces différentes valeurs de α, on obtient 1,791, c'est à dire exactement le même chiffre que celui fourni par la méthode des moindres carrés.

Mais ce résultat, bien qu'il soit fourni simultanément par les deux méthodes, peut être amélioré. Il est en effet peu logique d'attribuer la même importance, comme on le fait quand on se borne à prendre la moyenne arithmétique, à chaque valeur trouvée pour α. Il est évident, par exemple, qu'une valeur de α calculée entre les points de la droite logarithmique très éloignés les uns des autres et s'appliquant par conséquent à un vaste ensemble de l'échelle des revenus, a une signification beaucoup plus importante que celle qu'on peut attribuer à une valeur calculée entre des points très rapprochés.

Il faut donc, avant d'établir une moyenne, commencer par pondérer chaque valeur de α. Quelle méthode faut-il suivre?

α étant calculé de la valeur de revenus x_1 à la valeur x_2, de la valeur x_2 à la valeur x_3, etc..., il nous a paru indiqué de pondérer α d'après la somme des revenus comprise entre x_1 et x_2, entre x_2 et x_3, etc... Nous tenons compte, en procédant ainsi, non seulement du nombre de contribuables, mais aussi de ce qu'ils possèdent. Si les détenteurs de revenus sont très peu nombreux pour les valeurs élevées de x, ils sont fort riches, de sorte que les dernières tranches représentant les sommes de revenus les plus élevées, ne sont pas dénuées de toute importance

Nous ferons remarquer toutefois que nous avons calculé les valeurs de α, non d'après la courbe de la somme des revenus, mais d'après celle des personnes ayant un revenu supérieur à x.

Il pourrait paraître plus logique de se servir de la même formule, tant pour le calcul des différentes valeurs de α que pour leur pondération. Mais les deux manières de procéder reviennent au même.

En effet, la somme des fortunes pour une valeur de x, ou plus exactement pour une valeur comprise entre x et x_1 est donnée par la formule :

$$y = \frac{A\,\alpha}{x^\alpha}.$$

De cette équation, nous tirons :

$$\alpha = \frac{\log y_1 - \log y_2}{\log x_2 - \log x_1} \qquad (1)$$

Or il ne nous est pas possible de déterminer α au moyen de cette dernière équation, car les données statistiques ne nous fournissent pas le moyen de calculer directement y_1 et y_2. L'Administration fiscale ne nous donne que la somme des revenus compris entre x_1 et x_2 et que le nombre des contribuables possédant cette somme de revenus. Nous sommes donc obligés de déterminer α au moyen de l'équation :

$$\alpha = \frac{\log N_1 - \log N_2}{\log x_2 - \log x_1} \qquad (2)$$

Or les deux valeurs de α sont évidemment les mêmes dans l'équation (1) et dans l'équation (2). Nous connaissons donc la valeur de α dans l'équation (1) et nous pondérons cette valeur au moyen de celle de la somme des revenus compris entre x_1 et x_2.

Nous obtenons les pondérations suivantes pour les différentes valeurs de α,

Revenus en francs	Sommes des revenus	Poids	Valeurs de α
—	—	—	—
De 10.000 à 20.000.	19.603.469.300	18,26	1,675
20.000 à 30.000.	8.023.099.200	7,29	1,759
30.000 à 40.000.	4.364.424.600	4,07	1,751
40.000 à 50.000.	2.732.695.500	2,55	1,731
50.000 à 100.000.	6.061.436.800	5,65	1,657
100.000 à 200.000.	3.881.558.600	3,62	1,686
200.000 à 500.000.	3.156.259.300	2,93	1,791
500.000 à 1.000.000.	1.073.375.000	1	2,100

Nous multiplions chaque valeur de α par son poids. Nous faisons la somme de ces produits que nous divisons par la somme des poids, et nous trouvons $\alpha = 1{,}725$, au lieu de $\alpha = 1.791$, chiffre obtenu sans pondération.

Le revenu moyen entre 10.000 francs, revenu minimum, et un million pris comme revenu maximum, calculé au moyen de cette valeur de α est 22.957 francs, au lieu de 23.511 francs, chiffre fourni par les données statistiques. L'écart est donc inférieur à 2 1/2 %, ce qui est une approximation très satisfaisante.

Pour obtenir le revenu moyen, il n'est pas nécessaire de calculer la valeur du paramètre A, qui ne figure pas dans la formule du revenu moyen. Mais il est utile de la connaître pour savoir dans quelle mesure la courbe obtenue avec une valeur de α, telle que nous l'avons déterminée, se rapproche de la courbe établie au moyen des données de la statistique.

On peut calculer A pour chaque tranche, en se servant de la valeur moyenne pondérée de $\alpha = 1{,}725$, au moyen des équations :

$$\log A_1 = \log N_1 + 1{,}725 \log x_1$$

$$\log A_2 = \log N_2 + 1{,}725 \log x_2,$$

etc..., etc...

On prend la moyenne de ces différentes valeurs de A. Mais, même en pondérant les valeurs ainsi trouvées, on arriverait à un résultat moins bon que celui procuré par une méthode plus simple. Elle consiste à choisir comme valeur de A celle qui est calculée pour la première tranche.

On a :

$$A = \log N_1 + 1{,}725 \log x..$$

L'erreur est nulle pour N_1, qui est le plus importante des différentes valeurs de N, puisque N_1 est le nombre total des contribuables.

Les résultats obtenus figurent dans le tableau ci-dessous :

$$\alpha = 1{,}725.$$

A ayant plus que	francs	A calculé par une moyenne non pondérée	Chiffres de la statistique	A calculé seulement dans la 1re tranche	Pourcentage des erreurs
N_1	10.000	2.052.000	2.080.164	2.080.164	0
N_2	20.000	620.700	651.669	629.200	3,4
N_3	30.000	308.300	319.150	312.600	2,1
N_4	40.000	187.800	192.838	190.300	1,3
N_5	50.000	127.800	131.110	129.500	1,2
N_6	100.000	38.650	41.519	39.180	5,7
N_7	200.000	11.690	12.897	11.850	8,9
N_8	500.000	2.407	2.119	2.440	15,3
N_9	1.000.000	728	494	738	49,3
					9,69 erreur moyenne

Mais de même que nous avons pondéré les différentes valeurs de α, il est logique de pondérer les erreurs. Les personnes ayant plus de un million de revenus sont si peu nombreuses que l'erreur les concernant ne peut avoir la même importance, à pourcentage égal, que celles concernant des contribuables beaucoup plus nombreux.

Le procédé de pondération le meilleur consiste à tenir compte à la fois du nombre de personnes et de leurs revenus, et, puisqu'il s'agit de personnes ayant plus que x_1, x_2, etc..., de pondérer d'après les sommes des revenus au-dessus de x_1, x_2, etc...

Nous obtenons les résultats suivants :

Revenus au-dessus de	francs	Poids	Pourcentage des erreurs avant pondération
x_1	49.910.486.900	49,2	0
x_2	30.307.017.600	29,8	3,4
x_3	22.283.918.400	21,9	2,1
x_4	17.649.493.800	17,4	1,3
x_5	15.186.798.300	14,9	1,2
x_6	9.125.361.500	8,9	5,7
x_7	3.243.802.900	3,2	8,9
x_8	2.087.545.600	2,1	15,3
x_9	1.014.168.600	1	49,3

En opérant le produit des erreurs par les poids et en divisant la somme de ces produits par celle des poids, on obtient une erreur moyenne de 2,93 %, au lieu de 9,69 %, résultat qui est tout à fait satisfaisant.

Nous sommes donc arrivés, au moyen de la formule de Pareto, à donner une représentation à peu près exacte des données de la statistique. Mais les résultats trouvés auraient été certainement meilleurs encore, si les chiffres mis à notre disposition par l'Administration fiscale n'avaient pas été altérés par la fraude.

On doit remarquer que la formule de Pareto s'applique mal aux deux extrémités de la courbe, de sorte qu'on n'obtient pas, en partant d'elle, une ligne logarithmique absolument droite. L'exposant α en effet, tout au moins pour les revenus français, au lieu d'être constant pour toute l'étendue de la courbe, a, du côté des petits revenus, une valeur trop faible, et du côté des gros revenus une valeur trop forte.

Or un très grand nombre de contribuables ayant de petits revenus ne fait aucune déclaration. D'autre part, beaucoup de détenteurs de gros revenus fraudent, parce que l'exagération des droits atteignant les grosses fortunes rend cette fraude très tentante. Si les déclarations de revenus étaient sincères, la courbe de Pareto serait relevée au voisinage de son point de départ, parce qu'un plus grand nombre de contribuables y figurerait, puis s'abaisserait plus rapidement. Donc l'exposant α croîtrait. A l'autre extrémité au contraire, la courbe s'abaisserait moins rapidement et α décroîtrait. Le paramètre α tendrait donc à avoir aux deux extrémités la même valeur qu'au centre de la courbe.

Quant à la valeur moyenne de α, elle devrait croître légèrement, car, par suite de la forte pondération dont est affectée la somme des petits revenus, l'accroissement de α, vers le point de départ de la courbe, exerce sur sa valeur moyenne une influence plus forte que sa diminution à l'autre extrémité de la courbe.

* * *

La formule de Pareto nous donne non seulement le moyen d'évaluer les revenus des contribuables d'un pays, mais elle nous fournit encore un nouvel avantage très précieux, celui de nous permettre des comparaisons d'un pays à un autre, alors que les statistiques s'appliquent à des périodes différentes, où le pouvoir d'achat n'est pas le même et alors que leurs chiffres sont compris entre des revenus maximum et minimum qui ne sont pas identiques.

Comparons ainsi les revenus français de 1931 que nous venons d'étudier, et les revenus prussiens de 1913, tels qu'ils nous sont donnés par le *Statistisches Jahrbuch für den Preussischen Staat*, année 1913.

La statistique prussienne nous fournit simplement le nombre des contribuables par tranches de fortune; ainsi de 900 marks à 3.000 marks, on compte 6.489.373 contribuables, etc... Cette statistique nous indique aussi le taux de l'impôt, c'est-à dire le *Steuersatz*, mais elle ne nous donne aucun renseignement sur la somme des revenus des contribuables. J'ai su, depuis que mon travail était terminé, que d'autres publications nous fournissaient des indications plus complètes, mais le temps m'a manqué pour les utlliser.

Le revenu minimum de la statistique prussienne est 900 marks et le revenu maximum 100.000 marks. Connaissant l'ordre dans lequel sont rangés les contribuables, nous pouvons, en nous servant de la méthode précédemment exposée, calculer la valeur de α.

Nous obtenons les résultats suivants :

Marks	α
De 900 à 3.000	1,81
3.000 à 6.500	1,68
6,500 à 9.500	1,42
9.500 à 30.500	1,36
30.500 à 100.000	1,46

On voit que la formule de Pareto s'applique moins bien aux revenus prussiens de 1913 qu'aux revenus français de 1931. Si nous prenons la moyenne arithmétique des différentes valeurs calculées pour α, nous trouvons : $\alpha = 1{,}54$.

La méthode graphique, bien que d'autant plus incertaine que les points trouvés, en passant des nombres aux logarithmes, ne sont pas en ligne droite, conduit cependant à placer la valeur de α entre 1,60 et 1,70. La valeur trouvée $\alpha = 1{,}54$ paraît donc très mauvaise et donne de moins bons résultats que la méthode graphique.

Cet exemple nous montre qu'il est d'autant plus indispensable de pondérer les résultats trouvés que la formule s'applique moins bien.

En étudiant les revenus français de 1931, nous avions pondéré au moyen de la somme des revenus. Mais la statistique prussienne ne nous la donne pas. Par contre, elle nous renseigne d'une manière très approximative sur le taux de l'impôt dans chaque tranche.

Nous voyons : *Steuersatz*, 6 marks—52 marks, pour la première tranche. Cela veut dire que les contribuables paient de 6 marks pour 900 marks, à 52 marks pour 3.000 marks et le total de l'impôt dans cette tranche a produit 107.535.261 marks.

Or ce taux de 6 marks pour 900 marks, revenu minimum de la tranche, représente 0,67 %. 52 marks pour 3.000 marks, revenu maximum de la tranche, représentent 1,70 %.

La moyenne de ces deux taux est 1,18 %

Il s'agit maintenant de savoir combien, dans ce produit de l'impôt s'élevant à 107.535.261 marks, il y a de fois 1,18. A chaque 1,18 correspondent 100 marks. En multipliant par 100 le quotient trouvé, on obtient 9.113.157.700 marks, c'est-à dire le total de la somme des revenus des contribuables ayant de 900 à 3.000 marks de revenu. Nous procédons de la même manière pour les tranches suivantes.

Nous arrivons donc à reconstituer la somme des revenus pour chaque tranche. Sans doute, ces résultats ne sont qu'approximatifs et, de ce fait, notre pondération sera moins bonne que si nous avions des chiffres plus exacts. Elle sera cependant suffisante pour nous permettre d'approcher d'assez près de la meilleure valeur de α.

Nous trouvons ainsi : $\alpha = 1{,}697$.

Si nous calculons le revenu moyen en divisant la somme des revenus, telle que nous l'avons approximativement déterminée, par le nombre des contribuables, nous trouvons qu'il s'élève à 2.101 marks. Si nous calculons de nouveau le revenu moyen en nous servant de la formule :

$$r = r_2 \frac{\alpha}{\alpha - 1} \cdot \frac{\rho^{\alpha-1} - 1}{\rho^{\alpha} - 1}$$

dans laquelle $\alpha = 1{,}697$, on trouve qu'il s'élève à 2.037 marks. L'écart entre le chiffre fourni par la statistique est donc de 3 % environ.

D'autre part, α s'élève pour les revenus prussiens à 1,697 contre 1,725 pour les revenus français.

Or l'exposant α nous donne le revenu moyen indépendamment du pouvoir d'achat de la monnaie. Il suffit pour cela dans la formule du revenu moyen de ramener le revenu maximum r_2 à une somme ayant le même pouvoir d'achat dans les deux cas comparés et de conserver constant le rapport ρ du revenu maximum au revenu minimum.

α ayant une valeur un peu plus faible pour les revenus prussiens en 1913 que pour les revenus français en 1931, nous sommes donc amenés à conclure que le revenu moyen prussien en 1913 était légèrement supérieur au revenu français en 1931.

Mais ce n'est pas là le résultat apparent fourni par les comparaisons des chiffres des deux statistiques fiscales, qui font ressortir une richesse beaucoup plus grande pour la France que pour la Prusse.

Nous avons trouvé en effet pour la France en 1931 : 23.511 francs, comme revenu moyen, ce qui, malgré l'affaiblissement du pouvoir d'achat du franc depuis 1913, est sensiblement plus que 2.101 marks.

Mais cette divergence ne correspond pas à la réalité. La statistique prussienne en effet s'étend aux contribuables les plus pauvres et supprime les contribuables les plus riches. Le fait est évident, à première vue, puisque la Prusse de 1913 et la France de 1931 avaient à peu près la même population. Or la statistique prussienne porte sur 7.313.382 contribuables, tandis que la statistique française porte sur 2.080.164 contribuables seulement. Quant aux limites du revenu minimum, il est en Prusse de 900 marks or et en France de 10.000 frans-papier. Le revenu maximum est : 100.000 marks-or en Prusse et un million de francs-papier en France.

La formule de Pareto nous a donc permis de comparer le revenu moyen pour les deux pays, alors que le simple examen des statistiques fiscales ne nous fournissait aucun renseignement à cet égard.

Mais cette comparaison peut être grandement améliorée. Le fait d'avoir considéré les revenus les plus petits a fait croître la valeur moyenne pondérée de α dans la statistique prussienne, puisque dans les trois dernières tranches $\alpha = 1,42$, $\alpha = 1,36$, $\alpha = 1,46$, tandis que dans la première tranche, il égale 1,81. Cette première tranche a été affectée d'un très gros coefficient. Si on l'avait réduite en supprimant un nombre important des contribuables la composant, ce coefficient aurait été réduit et on aurait obtenu une valeur moyenne de α sensiblement plus faible.

Nous allons donc supprimer la somme de revenus portant sur les contribuables ayant un revenu en marks-or inférieur à 10.000 francs-papier de 1931.

J'ai pris comme coefficient la hausse des prix de la vie 609, qui est celui des prix de détail à Paris en 1931, d'après la statistique générale de la France, la base étant 100 en 1914. Si l'année 1913 avait été choisie comme base, la différence entre les résultats obtenus aurait été insignifiante.

Sur cette base, 900 marks or de 1913, dont le pouvoir d'achat aurait varié dans la même proportion que celui du franc de 1913 à 1931, correspondent à 5.481 marks.

10.000 francs de 1931 correspondront à 8.000 marks transformés en pouvoir d'achat de 1931.

Nous précisons, pour qu'il n'y ait aucune équivoque, que la question n'est pas de savoir dans quelle proportion le Reichsmark de 1931 s'est déprécié par rapport au mark impérial de 1913.

Il s'agit de faire varier dans la même proportion le pouvoir d'achat des monnaies comparées. S'il nous plait, pour un motif imposé ou arbitraire, de multiplier par 10 une unité monétaire, l'unité monétaire du pays pris en comparaison doit être, elle aussi, multipliée par 10. Nous ne pouvons mesurer les dimensions relatives de deux objets en regardant l'un avec une loupe et l'autre à l'œil nu. Ils doivent être tous les deux observés avec la même loupe.

Le problème étant ainsi posé, il s'agit de retrancher les contribuables ayant moins de 8.000 marks, c'est-à dire en comptant dans l'unité adoptée de 1931, ayant des revenus compris entre 5.481 marks et 8.000 marks, ou en rétablissant le décompte en marks or, les contribuables ayant de 900 à 1.313 marks.

Pour cela, nous emploierons la représentation logarithmique de la courbe de Pareto relative à la somme des revenus qui est une droite. Nous pourrons ainsi déterminer par une interpolation la somme des revenus des contribuables en question.

En procédant ainsi, nous réduirons le poids de la première tranche de 12,75 à 7,69 et la valeur trouvée pour α sera 1,660 au lieu de 1,697 avant cette correction.

Comparons les différents revenus moyens trouvés pour la Prusse et pour la France, et, pour cela, servons-nous de la formule :

$$\frac{\alpha}{\alpha - 1} \cdot \frac{\rho^{\alpha-1} - 1}{\rho^{\alpha} - 1}.$$

On trouve :

Prusse 1913 (Pareto)	0,023 959
France 1931 (chiffres de la statistique)	0,023 511
France 1931 (Pareto)	0,022 957

c'est-à-dire qu'en même unité de mesure, en l'espèce, le franc de 1931, ces revenus moyens auraient été 23.959 francs, 23.511 francs et 22.957 francs.

Remarquons que, si j'avais multiplié les marks de 1913 par 5,98 (coefficient de hausse des prix de détail pour la France entière en 1931), au lieu de 6,09 (prix de détail pour Paris), ce qui aurait du reste mieux valu, j'aurais obtenu un revenu moyen plus élevé pour la Prusse, mais qui n'aurait différé du chiffre trouvé que d'une manière tout à fait insignificante.

On arrive donc à cette conclusion que le revenu moyen prussien en 1913 et le revenu moyen français en 1931 sont à peu près du même ordre.

Ce résultat est tout à fait vraisemblable. Comparons le avec d'autres chiffres indiqués par nos collègues, MM. Sauvy et Rivet dans leur belle étude « Fortune et Revenu national », parue dans la *Revue d'Économie politique* (janvier-février 1939).

Les revenus totaux distribués en France par tête d'habitant, en 1913, s'élevaient à 900 francs. D'autre part, MM. Sauvy et Rivet indiquent que l'Institut de conjoncture de Berlin estime à 750 marks les revenus distribués en Allemagne en 1913 par tête d'habitant, soit 937 fr. 50. Donc, en 1913, en France, 900 francs, et en Allemagne 937 fr. 50.

En 1936, d'après MM. Sauvy et Rivet, le revenu français distribué par tête s'élevait à 4.730 francs, ce qui, en appliquant le coefficient de hausse des prix de détail à Paris, (base 1914) c'est-à-dire 4,80, donne une hausse réelle d'un peu moins de 10 %. D'autre part, le revenu prussien se serait accru, d'après l'Institut de conjoncture de Berlin, de 7 % environ, de 1913-1914 à 1937.

Tous les chiffres obtenus concourent donc à prouver que les revenus français et les revenus allemands, par tête d'habitant, ne diffèrent que très peu.

Sans doute, les chiffres de MM. Sauvy et Rivet, ceux de l'Institut de conjoncture de Berlin et les chiffres que j'ai trouvés ne sont pas exactement comparables.

D'abord, mon étude a porté non sur l'Allemagne et la France, mais sur la Prusse et la France. Il est possible que la Prusse soit un peu plus riche que l'ensemble de l'Allemagne, parce que, dans son territoire se trouve la capitale de l'Empire, ensuite parce qu'elle détient la plus grande partie de l'industrie charbonnière.

De plus, je n'ai envisagé que le revenu moyen des contribuables, tandis que les derniers chiffres que j'ai cités s'appliquent à la totalité des revenus, divisés par le chiffre de la population. Ils sont donc forcément beaucoup plus faibles.

Enfin, de même que l'emploi de la formule de Pareto ne fournit que des résultats approchés, les méthodes qui procèdent au dénombrement direct des revebus comportent aussi forcément, quelle que soit la science avec laquelle elles sont maniées, une certaine part d'approximation. MM. Sauvy et Rivet font du reste remarquer « que des réserves doivent être faites sur le montant du revenu national allemand, en raison de l'incertitude subsistant sur les méthodes de calcul et notamment sur la possibilité des doubles emplois ».

Il résulte de cette étude que la formule de Pareto, bien qu'approximative, peut conduire, si on sait la manier, avec une technique prudente, à des résultats précis et qu'elle nous permet de comparer très rapidement la fortune de deux pays à des époques différentes.

Elle sera encore d'un emploi plus sûr, si on étudie l'évolution de la fortune dans un même pays. En effet, par suite de la stabilité des formes sociales, la distribution des revenus, même si elle s'éloigne de celle donnée par la formule de Pareto, se traduira toujours par une courbe de la même forme. Les divergences entre elles et la courbe de la réalité se trouveront donc annulées dans les comparaisons.

Pourquoi du reste être si sévère pour la formule de Pareto? Pour de nombreux phénomènes, les comparaisons statistiques ne se font-elles pas au moyen de données approximatives? L'estimation des variations des récoltes s'appuie-t-elle sur des statistiques exactes? Les nombres indices des prix, quelle que soit la méthode avec laquelle ils sont calculés, ne peuvent donner des résultats absolument précis. En effet, la notion du prix général, c'est-à-dire d'un prix s'appliquant à l'ensemble des marchandises, ne peut être définie que par une convention arbitraire, qui laisse la place libre à l'approximation.

Les autres sciences, par exemple la physiologie, n'emploient-elles pas des instruments de mesure qui manquent de précision et donnent cependant des indications d'une utilité certaine. Dans un passé encore récent, les appareils construits pour prendre la tension sanguine donnaient des chiffres qui avaient un écart de 10 à 20 % avec la réalité. Les renseignements fournis par eux aux praticiens étaient cependant très précieux.

La formule de Pareto peut conduire aux résultats les plus variés et les plus importants et j'estime que celui de pouvoir mesurer la richesse d'un pays au moyen de l'ordre dans lequel sont rangés les contribuables constitue à lui seul une grande découverte.

Baron MOURRE.

DISCUSSION

M. le PRÉSIDENT remercie M. le baron MOURRE et donne la parole à M. M. FRÉCHET, qui s'exprime ainsi :

M. FRÉCHET. — Je tiens d'abord à féliciter M. le baron MOURRE de son très intéressant exposé. Celui-ci a suscité dans mon esprit quelques réflexions que je voudrais vous soumettre.

I. — Que signifie la loi de PARETO ; que si l'on désigne par N le nombre des individus de revenus supérieurs à x, et si l'on reporte sur un papier quadrillé les points dont les coordonnées sont les logarithmes de N et de x, ces points qui, *a priori*, pourraient se distribuer très irrégulièrement (sauf à préserver la décroissance, qui a lieu, par définition de N, ou de log N, quand log x croît) se trouvent alignés, ou à peu près alignés.

C'est vraiment là un fait remarquable et qui peut illustrer la réponse qu'on peut opposer à ceux, de plus en plus rares, qui dénient droit de cité à la mathématique dans le domaine de l'économie politique. Ils répètent qu'un instrument aussi précis, aussi rigide que la mathématique ne peut avoir son usage dans une science où interviennent des causes si nombreuses, si complexes, si subtiles (comme la psychologie humaine). Il est certain qu'il faut employer cet instrument avec précaution. Mais comment ne pas être frappé de la réponse à ces objections fournie par la loi de PARETO. Qu'y a-t-il de plus complexe, de plus subtil que les raisons pour lesquelles s'établit la répartition des revenus, déterminée par la situation sociale, les aptitudes intellectuelles, le caractère des individus, et aussi les circonstances extérieures. Chaque point, reporté comme précédemment indiqué, est la résultante d'un monde de causes diverses. Et pourtant, c'est là un fait : les points obtenus sont à peu près alignés, avec tout juste un petit gauchissement aux extrémités.

Comment exprimer en langage non mathématique cette constatation surprenante, à moins qu'on ne redise, d'une manière moins précise et en phrases beaucoup plus longues, que ces points sont approximativement alignés.

II. — Je ferai maintenant une remarque au sujet des comparaisons du revenu moyen dans plusieurs pays. Ces revenus moyens sont calculés à partir de *statistiques fiscales*. Or celles ci ne concernent ni la population entière, ni la population active. Elles se limitent à ceux qui ont (ou déclarent) un revenu supérieur à un certain revenu de base. Celui ci varie d'un pays à l'autre. Il serait donc plus indiqué de parler, dans ces cas, non pas de revenu moyen tout court, mais de la moyenne des revenus supérieurs à telle ou telle somme donnée.

III. — Au sujet de la méthode de calcul de α dans la formule de PARETO, l'orateur fait observer avec juste raison que la méthode des moindres carrés conduit à des calculs pénibles. Il propose une autre méthode de calcul. Celle-ci peut s'exprimer géométriquement ainsi : on prend pour α (qui est le coefficient angulaire de la droite d'ajustement mentionnée plus haut) la moyenne arithmétique des coefficients angulaires des côtés du polygone ayant pour sommets les points observés. Cette méthode a donné de bons résultats au conférencier. Toutefois, elle conduit en général à ce résultat paradoxal que plus d'observations on fera, moins bien α risque d'être déterminé; en effet, de petites irrégularités dans les positions des sommets se traduiront, quand les points sont rapprochés, par de grandes irrégularités dans les valeurs des coefficients angulaires des côtés.

Si donc, on veut éviter cet inconvénient, sans recourir à de longs calculs, il nous paraîtrait préférable d'employer la méthode très élémentaire « des moindres écarts » qu'on trouvera exposée ailleurs (1).

IV. — Je m'associe entièrement à la dictinction faite par le conférencier entre les exigences du statisticien et celles de l'économiste, au regard de l'approximation des formules employées.

Pour l'économiste, si une formule plus simple donne cependant une bonne idée de l'ordre de grandeur et du sens de variation de la fonction elle sera bien souvent préférable.

Toutefois, il n'est pas inutile d'indiquer des raisons, valables même pour l'économiste, pour chercher des formules moins simples, mais plus approchées que celle de PARETO.

Tout d'abord, il serait très souvent utile, pour les économistes, d'être en possession de tableaux statistiques ne comportant pas seulement les deux colonnes donnant les valeurs correspondantes de N et de x, mais trois colonnes, la troisième donnant les valeurs de R, total des revenus supérieurs à x. Quelques administrations établissent cette troisième colonne. Or, si une formule représentait bien N en fonction de x, on pourrait en déduire R par intégration :

$$R_x = \int_{\infty}^{X} x \, d \, N \tag{1}$$

On pourrait donc, par ce moyen, *compléter sans nouveau recensement statistique* les autres tableaux parus sous cette troisième colonne. Et ceci est un exemple de plus, en passant, des services que peut rendre la mathématique à l'économie politique. Mais si l'expression de N en fonction de x n'est pas bien approchée, celle de R le sera souvent encore moins. C'est ce qui a lieu avec la formule de PARETO :

$$N = \frac{A}{x^\alpha} \tag{2}$$

elle permet bien d'arriver à une expression simple de R :

$$R = \frac{A\,\alpha}{\alpha - 1} \, \frac{1}{x^{\alpha - 1}} \tag{3}$$

(1) *Représentation des lois empiriques par des formules approchées*, à l'usage des chimistes, des physiciens, des ingénieurs et des statisticiens, par FRÉCHET et ROMANN, p. 49, chez Eyrolles, 1930.

Mais l'erreur relative faite sur le calcul de α, au moyen du tableau des N et des x se traduira par une plus grande erreur relative sur $\alpha - 1$ (car α étant généralement compris entre 1 et 2, $\alpha - 1$ sera compris entre 0 et 1) et, par suite, aussi sur R. On obtiendra encore une idée de l'ordre de grandeur de R, sans en connaître aucune valeur statistique directe, ce qui est en soi très remarquable, mais avec des erreurs relatives souvent excessives. Là encore se révèle l'utilité de chercher une expression plus précise que la loi de PARETO.

V. — La formule de M. BARRIOL est une de celles qui permettent une plus grande précision. On peut la juger à deux points de vue. Au point de vue d'une bonne représentation, d'un tableau numérique, elle réalise une supériorité certaine sur la formule de PARETO — et se conforme mieux à la réalité — en introduisant un revenu maximum. (Toutefois, on peut observer que la loi de PARETO fournit aussi un revenu maximum, à savoir la valeur de x pour laquelle $N = 1$, puisque N ne peut être inférieur à 1, sans être égal à zéro.) Mais si l'on considère la loi de PARETO comme une loi de probabilité plutôt que comme une loi de fréquence, il y aurait, au contraire, un inconvénient de principe à supposer qu'il existe une valeur que le revenu ne peut dépasser dans la population fictive illimitée à laquelle se réfère la définition de la probabilité.

Soit $f = \frac{N}{N_0}$ la fréquence des N revenus supérieurs à x parmi les N_0 revenus déclarés (supérieurs à x_0); si la loi de PARETO est une loi de fréquence, on aura :

$$f = \frac{A\,x^{-\alpha}}{A\,x_0^{-\alpha}} \qquad \text{d'où} \qquad f = \left(\frac{x_0}{x}\right)^{\alpha}$$

Mais si c'est une loi de probabilité, la probabilité P que N surpasse x sera obtenue en remplaçant f par p :

$$P = \left(\frac{x_0}{x}\right)^{\alpha} \tag{4}$$

En prenant x_1 assez grand, la probabilité que le revenu soit supérieur à x_1 pourra être rendue assez petite (par exemple, p inférieur à un milliardième), pour que cette circonstance devienne pratiquement impossible. Par suite, un maximum *pratique* du revenu sera assuré, ce qui réalise le desiderata de M. BARRIOL, sans qu'il soit nécessaire pour cela de supposer p nul au delà d'un certain revenu dont on ne voit pas comment ne pas le fixer arbitrairement.

On a là une circonstance analogue à celle qui se présente en théorie de la mortalité, où l'on a longtemps supposé un âge limite (vers les cent ans), tandis qu'actuellement, on préfère de plus en plus supposer seulement que la probabilité de survie à l'âge x tend vers zéro quand x croît indéfiniment.

VI. — Indiquons encore une raison que pourrait avoir un économiste, pour accepter des formules moins simples que celle de PARETO.

C'est d'abord que la formule de PARETO ne rend pas compte, ne traduit absolument pas, un phénomène constaté précisément par les économistes. Au lieu de représenter graphiquement la fonction N de x, on a eu besoin de représenter le taux d'accroissement de N par unité de revenu $\left(\text{c'est-à-dire } -\frac{dN}{dx}\right)$ et on a constaté que ce taux présente un maximum pour un revenu assez modeste, mais supérieur à celui qui assurerait strictement l'existence. Il commence donc par croître. C'est une remarque empirique qu'on aurait pu prévoir théoriquement, car il est certain qu'au dessous d'un certain revenu, un individu ne peut plus vivre, que par suite, N restant constant pour x assez petit, son taux d'accroissement sera d'abord nul. Il devra donc ensuite commencer par augmenter.

Or, d'après la formule de PARETO, on a :

$$-\frac{dN}{dx} = \frac{A\alpha}{x^{\alpha-1}}$$

de sortè que le taux d'accroissement décroîtrait constamment.

Il est clair que — question d'exactitude rigoureuse mise à part, — il importe au point de vue de l'économiste que la formule adoptée reflète cette propriété importante :

VII. — La formule

$$N = \frac{N_\omega}{2}\left(1 - \Theta\left[a \log (x - x_0) + b\right]\right) \tag{5}$$

avec :

$$a > 0,$$

et

$$\Theta(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$

que notre collègue GIBRAT a eu le mérite, en 1931, de justifier au moyen de nombreux exemples (et qui avait été proposée, depuis longtemps d'ailleurs, indépendamment, par Mc Alister en 1879, Edgeworth en 1898 et Kapteyn en 1903) est encore une de celles qui permettent de mieux approcher les relevés statistiques et qui échappent à l'inconvénient ci dessus.

Le conférencier a noté qu'il serait intéressant de donner une signification concrète au paramètre b de cette formule (a jouant, comme il l'a signalé et comme on le verra plus loin, un rôle analogue à celui de α dans la formule de PARETO). Cette interprétation est en fait connue; je vais la rappeler.

On voit d'abord que pour x tendant vers x_0 par valeurs positives $a \log (x - x_0) + b \to -\infty$, donc N tend vers son maximum N_ω. Ainsi x_0 est le revenu minimum, soit qu'il représente le revenu au-dessous duquel il n'y a plus de déclaration, N_ω ne désignant alors que le nombre de déclarations, soit que N_ω représente l'ensemble de la population et x_0 le revenu minimum au dessous duquel l'existence n'est plus possible.

Quant à la signification de b, il vaut mieux poser $b = -a\,c$ et donner la signification de c. En effet, quand $a \log (x - x_0) + b$ est nul, soit pour $x = p$, on a $N = \frac{N_\omega}{2}$ de sorte que p est le revenu probable ou tout au moins le revenu déclaré probable. On a alors :

$$a \log (p - x_0) - ac = o$$

d'où :

$$p - x_0 = e^c$$

Les significations des paramètres x_0 et c sont donc :

(6) $x_0 =$ revenu minimum,

(7) $c =$ log [(revenu déclaré probable) — (revenu minimum)].

Calcul explicite de R. — Pour avoir la signification de a, il est utile de calculer explicitement R. Je vais compléter ici un article précédent (1) où, ayant eu besoin d'exprimer R sous forme analytique simple, j'avais cru nécessaire de substituer à l'expression de Mc Alister une autre plus facilement intégrable. En fait, on peut aussi calculer R, et même très simplement, à partir de l'expression de Mc Alister.

Appliquons la formule (1); on peut écrire :

$$R_X = \int_{+\infty}^{X} (x - x_0)\, d\,N + x_0 N_X.$$

D'où, en posant :

$$z = a\left[(\log (x - x_0) - c\right],$$

ou

$$x - x_0 = e^{\frac{z}{a} + c}$$

$$R_X - x_0 N_X = \int_{+\infty}^{Z} e^{\frac{z}{a} + c}\, d\,N,$$

avec, d'après (5)

$$N = \frac{N_\omega}{2}\left[1 - \Theta(z)\right] \qquad d\,N = -\frac{N_\omega}{2}\frac{2}{\sqrt{\pi}} e^{-z^2} dz,$$

d'où :

$$R_X - x_0 N_X = \frac{N_\omega}{\sqrt{\pi}} \int_{Z}^{+\infty} e^{-z^2 + \frac{z}{a} + c}\, d\,z.$$

$$\frac{R_X - x_0 N_X}{N_\omega} = \frac{1}{\sqrt{\pi}} \int^{+\infty} e^{-\left(z - \frac{1}{2a}\right)^2 + c + \frac{1}{4a^2}}\, d\,z$$

$$= \frac{1}{\sqrt{\pi}} e^{c + \frac{1}{4a^2}} \int_{Z - \frac{1}{2a}}^{+\infty} e^{-t^2} dt = \frac{e^{c + \frac{1}{4a^2}}}{2}\left[1 - \Theta\left(Z - \frac{1}{2a}\right)\right]$$

Ainsi :

$$R - x_0 N = \frac{e^{c + \frac{1}{4a^2}}}{2} N_\omega \left[1 - \Theta\left(z - \frac{1}{2a}\right)\right]$$

D'où enfin :

$$R = \frac{N_\omega}{2}\left\{x_0\left[1 - \Theta(z)\right] + e^{c + \frac{1}{4a^2}}\left[1 - \Theta\left(z - \frac{1}{2a}\right)\right]\right\} \tag{8}$$

avec

$$z = a\left[\log (x - x_0) - c\right]$$

(1) *Sur les formules de répartition des revenus* (*Revue de l'Institut International de Statistique*, 1939).

On peut aussi écrire R plus explicitement :

$$(9) \qquad R = \frac{N_\omega}{2} \left\{ x_0 \left[1 - \Theta\left(a \left[\log(x - x_0) - c\right]\right)\right] + e^{c + \frac{1}{4a^2}} \left[1 - \Theta\left(a \left[\log(x - x_0) - c - \frac{1}{2a^2} \right] \right) \right] \right\}$$

On en tire le revenu moyen :

$$r = \frac{R_\omega}{N_\omega} = \frac{1}{2} \left\{ 2\, x_0 + 2\, e^{c + \frac{1}{4a^2}} \right\} \quad , \qquad r = x_0 + e^{c + \frac{1}{4a^2}}$$

Comme on a vu que $p - x_0 = e^c$, on en conclut :

$$\frac{r - x_0}{p - x_0} = e^{\frac{1}{4a^2}}$$

d'où :

$$(10) \qquad a = \frac{1}{2 \sqrt{\log \frac{r - x_0}{p - x^0}}}$$

$$(11) \qquad a = \frac{1}{2 \sqrt{\log \frac{\text{revenu moyen} - \text{revenu minimum}}{\text{revenu probable} - \text{revenu minimum}}}}$$

Voici donc obtenues par les formules (6), (7) et (11) les significations économiques des paramètres x_0, c et a figurant dans la formule de Mac Alister.

Remarque. — Il est intéressant de voir ce que deviennent les seconds membres de ces formules (qui ont un sens indépendant de la formule adoptée pour la représentation de N en fonction de x) quand on part de la loi de Pareto :

$$N = \frac{A}{x^\alpha}.$$

Appelons x_0 le revenu au-dessous duquel il n'y a pas de déclaration, — ou, suivant le cas, le revenu au-dessous duquel on ne peut vivre.

Le revenu moyen sera :

$$r = \frac{R_\omega}{N_\omega} = \frac{A \frac{\alpha}{\alpha - 1} \frac{1}{x_0^{\alpha - 1}}}{\frac{A}{x_0^\alpha}} = \frac{\alpha\, x_0}{\alpha - 1}.$$

Le revenu probable p sera tel que :

$$\frac{A}{p^\alpha} = \frac{N_\omega}{2} = \frac{A}{2\, x_0^\alpha}$$

d'où :

$$\frac{p}{x_0} = 2^{\frac{1}{\alpha}}$$

Dès lors, les valeurs des seconds membres de (6), (7), (11) deviennent :

(6) $$\text{revenu minimum} = x_0,$$

(7 *bis*) $$\log(p - x_0) = \log x_0 + \log\left(2^{\frac{1}{\alpha}} - 1\right)$$

(11 *bis*) $$\frac{1}{2\sqrt{\log\left(\frac{r-x_0}{p-x_0}\right)}} = \frac{1}{2\sqrt{\log\left[\frac{1}{(\alpha-1)\left(2^{\frac{1}{\alpha}}-1\right)}\right]}}$$

On vérifie facilement que a comme α sont des fonctions décroissantes du rapport $\frac{r - x_0}{p - x_0}$.

Nous avons déjà écrit (1), M. Halbwachs et moi : « On peut considérer que l'inégalité de la répartition des (revenus) est d'autant plus prononcée que le montant probable (p) est plus petit par rapport au montant unique qu'on constaterait si les (revenus) étaient également distribués, c'est à dire au montant moyen (r). On peut donc considérer le rapport $\frac{r}{p}$ comme mesurant ou mieux repérant l'inégalité de répartition des (revenus) ou tout au moins des revenus déclarés. » Il est clair que des raisons analogues pourraient conduire à faire considérer $\frac{r - x_0}{p - x_0}$ comme un tel indice. Il en résulte que a et α ont une signification commune, celle d'être d'autant plus petits que les revenus sont plus inégalement distribués.

M. Amy est très frappé par l'identité de la formule de Pareto avec celles utilisées par de nombreux auteurs dans l'étude de phénomènes n'ayant certainement aucune autre analogie. En voici quelques exemples :

La loi d'allométrie, qui représente la croissance des animaux ou des végétaux en fonction du temps. La formule d'adsorption de Freundlich qui relie la quantité de matières colorantes ou d'autres substances colloïdales fixées par un corps poreux à la concentration de cette substance en solution.

La loi de réciprocité qui relie l'intensité lumineuse et la durée de son action permettant d'obtenir un égal noircissement d'une émulsion photographique.

La loi reliant la concentration d'un ion en solution avec la quantité de cet ion fixé par une zéolithe, une permutite, un colloïde humique ou même un organisme vivant.

Il existe un grand nombre de phénomènes physiques où l'exposant prend une forme simple. Est-il besoin de rappeler la formule du pendule ou celle de la chute des corps avec les exposants 2 ou 1/2 ?

La loi d'Eötövs qui relie la tension superficielle, le poids moléculaire, la densité du liquide et de la vapeur et la température avec les puissances 2/3, etc..., etc...

Disons encore que pendant longtemps on a représenté l'équilibre de l'oxy-

---

(1) *Le calcul des probabilités à la portée de tous.* (Voir Exercice III, p. 145), chez Dunod, 296 pages.

gène et de l'oxyde de carbone avec l'hémoglobine dans le sang *in vivo et in vitro* par une équation identique. Or, on sait actuellement que c'est la loi d'action de masse qui régit ces équilibres. D'autre part, les équilibres des échangeurs d'air sont tout aussi bien (ou aussi mal) représentés par des formules exponentielles, la loi de réciprocité photographique ne s'applique plus dans les régions de sous et sur exposition et s'inverse complètement dans la région de solarisation.

On peut constater des faits du même genre avec la formule d'allométrie ou la loi de FREUNDLICH.

En fait, il n'y a que lorsque l'exposant est un entier ou une fraction très simple, la racine carrée ou la racine cubique par exemple, que la loi semble vraiment rigoureuse.

D'accord avec la plupart des auteurs qui ont discuté cette question en physique ou en physico chimie, il semble donc que les formules du type PARETO à coefficients complexes n'ont aucune signification théorique et sont un simple artifice mathématique commode. Il n'y a donc pas lieu d'attacher une signification directe et précise à la valeur du coefficient. Certes, dans la formule de PARETO le coefficient α varie dans le même sens que l'inégalité de fortunes, mais il me paraît dangereux de dire qu'il mesure cette inégalité.

L'expérience a de même montré que les modifications de la formule n'amélioraient les résultats que grâce à la présence de constantes supplémentaires. Des différentes modifications possibles, il est donc logique de choisir celles qui se prêtent aux calculs les plus aisés toujours sans chercher à attacher un sens théorique aux nouvelles constantes.

Les critiques que je crois devoir présenter ne me font pas sous-estimer l'intérêt de ces formules. Je citerais comme particulièrement démonstratif à cet égard que la représentation graphique de la loi d'allométrie a mis en évidence l'existence de deux évolutions nécessaires chez chaque individu, la discontinuité dans la croissance correspondant au début du développement sexuel.

Le travail si intéressant présenté ce soir par M. le baron MOURRE illustre mieux que tout autre le résultat que l'on peut attendre de l'emploi de formules empiriques.

M. BARRIOL dit qu'il a été en effet amené à étudier la formule de PARETO et à chercher diverses applications; il croit que l'engouement des statisticiens et des économistes mathématiciens pour cette formule résulte de la facilité de l'intégration de la fonction exponentielle; dans ses précédentes communications à la Société, M. le baron MOURRE s'est très heureusement et très habilement servi de cette propriété pour démontrer diverses propriétés qu'il a sommairement rappelées dans sa conférence.

La formule proposée à laquelle M. MOURRE fait allusion a quelques avantages sur la formule de PARETO grâce à deux coefficients simples qui ne compliquent pas son intégration et permettent d'étendre très largement le champ d'application de la formule.

En ce qui concerne ce que vient de dire si excellemment le professeur FRÉCHET, M. BARRIOL pense que l'ajustement des points pouvant être alignés d'une expression dépendant d'une fonction logarithmique se fait très facilement avec l'aide d'une droite tracée sur un transparent; mais il y a beaucoup de cas où la fonction logarithmique ne peut pas intervenir simplement. Il s'est souvent très bien trouvé de faire l'ajustement en se donnant *a priori* une valeur de l'erreur, à ne pas dépasser et à calculer les écarts avec la formule $\sqrt{\frac{2\,p\,(1-p)}{n}}$ qui tient compte du nombre des épreuves, p étant la probabilité et n le nombre d'épreuves.

Les calculs sont très rapides, l'ajustement se fait d'une façon très suffisante et le dessinateur est parfaitement guidé par le tracé de la courbe ajustée.

Les comparaisons faites par M. Mourre sont intéressantes et les valeurs de α pour les deux pays considérés sont assez voisines, et cependant les éléments statistiques sont très différents, tant au point de vue de la répartition de classes de revenus, que de la contexture même des éléments imposés. Il croit cependant qu'il serait préférable de diriger les recherches en considérant seulement un pays, et c'est dans ce sens qu'il a souvent encouragé des candidats au diplôme de statisticien de l'Institut de Statistique de l'Université de Paris à préparer une thèse.

M. le Président fait observer que les courbes représentatives des revenus d'une part et des fortunes d'autre part sont essentiellement différentes, bien qu'on ait souvent voulu les rapprocher, mais il lui apparaît que ces efforts ont été vains.

M. le baron Mourre répond qu'en effet la répartition des capitaux est très différente de celle des revenus. Pareto ne paraît pas l'avoir remarqué et c'est peut-être parce qu'il obéissait à l'idée préconçue que α ne pouvait varier qu'entre des limites assez étroites. Mais si l'on admet qu'on se trouve en présence de répartitions très différentes, quand il s'agit des fortunes et des revenus, il n'y a pas lieu de s'étonner que α ait une valeur beaucoup plus faible pour les successions que pour les revenus.