Journal de la société statistique de Paris

MOURRE

Définitions de l'inégalité des revenus et utilisation de la formule de Pareto

Journal de la société statistique de Paris, tome 84 (1943), p. 228-240
<http://www.numdam.org/item?id=JSFS_1943__84__228_0>

III

DÉFINITIONS DE L'INÉGALITÉ DES REVENUS ET UTILISATION DE LA FORMULE DE PARETO [1]

Certains auteurs ont fait remarquer qu'une mesure de l'inégalité des revenus correctement établie devait être indépendante du pouvoir d'achat de la monnaie et du chiffre de la population.

La première de ces assertions est exacte, mais la seconde est erronée.

Le choix de l'unité monétaire ne peut en effet modifier la répartition des revenus. Par contre, les rapports qu'ont entre eux les différents revenus sont fonction du chiffre de la population. Un pays très peuplé doit en effet compter plus de grosses fortunes qu'un petit pays.

Cela résulte du principe très général d'après lequel les éléments revêtus d'un caractère exceptionnel sont d'autant moins rares que le nombre total des éléments mis en cause est plus grand. Considérons, par exemple, deux pays de population inégale, mais également évolués au point de vue d'un sport tel que le tennis. Chacun d'eux doit avoir des nombres

1) Communication présentée à la séance du 20 octobre 1943.

de mauvais joueurs, de joueurs moyens et de bons joueurs à peu près proportionnels au chiffre de leur population, mais le grand pays aura plus de chances que le petit pays de posséder le meilleur joueur international.

Pour ce qui concerne les revenus, la détention du maximum le plus élevé pour le grand pays acquiert un caractère de certitude, quand on fait usage de la formule de Pareto. Cette formule est :

$$N_x = \frac{A}{x^\alpha} \quad (1)$$

où N_x représente le nombre d'individus possédant un revenu supérieur à x, et où A et α sont les constantes pour une année et une répartition déterminées.

Nous nous servirons de la double représentation logarithmique, c'est-à-dire que nous porterons en abscisse log x et en ordonnée log N_x. Chaque répartition est représentée par un segment de droite. Des répartitions de même structure, c'est-à-dire ayant le même α, sont représentées par les segments de droites parallèles.

Nous désignerons par m le revenu minimum d'une répartition, c'est-à-dire le revenu immédiatement inférieur à celui qui est possédé par l'individu le plus pauvre soumis à la loi de Pareto.

Quant au revenu maximum, si nous supposons que la loi de Pareto s'applique aux revenus les plus élevés, le point où la droite rencontre l'axe des abscisses correspond à une valeur nulle du logarithme de N_x, pour laquelle $N_x = 1$. Cela nous permettra de fixer le revenu maximum.

Considérons deux répartitions de même structure, c'est-à-dire possédant le même α, mais de populations différentes.

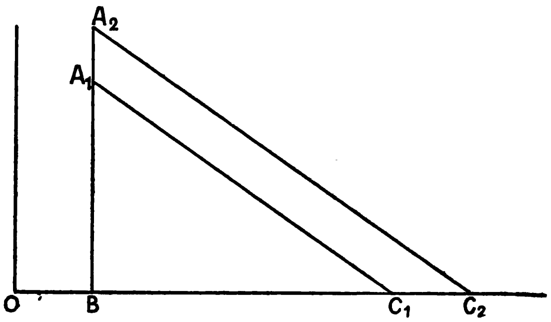

Fig. 1.

Leur représentation se fera suivant les deux segments A_1 C_1 et A_2 C_2. L'abscisse commune de A_1 et de A_2 est O B = log m ; leurs ordonnées représentent log N_1 et log N_2. C_1 et C_2 ont pour abscisses respectivement log M_1 et log M_2.

On voit que le plus grand pays a un revenu maximum plus élevé que le petit pays.

Certes, dans la pratique, le phénomène n'a pas le degré de certitude que lui confère la formule. La loi de Pareto est en défaut à l'extrémité riche de la répartition, où les grands nombres ne jouent plus. Il est possible que l'homme le plus riche de l'univers se trouve dans la République de Saint-Marin et non pas dans un des grands pays du monde. Mais il est peu vraisemblable qu'il en soit ainsi, car, si la loi de Pareto ne s'applique pas d'une manière rigoureuse dans la zone des gros revenus, elle subsiste néanmoins comme loi de tendance et le chiffre du revenu maximum fourni par elle, s'il n'est pas exact, indique du moins un ordre de grandeur vraisemblable.

Mais le fait qu'un pays, parce qu'il est plus peuplé, possède des détenteurs de revenus plus élevés qu'un autre, qui est moins peuplé, bien que toutes choses soient égales d'ailleurs, ne conduit-il pas à penser que la fortune est plus inégalement distribuée dans le premier que dans le second?

Nous sommes donc amenés à introduire dans les formules que nous emploierons pour définir l'inégalité le chiffre de la population.

Le revenu maximum ne nous est pas, en général, fourni par les statistiques fiscales. Choisissons pour le fixer la valeur de x pour laquelle log $N_x = 0$ et $N_x = 1$. Il y aurait donc un individu dont le revenu serait au moins égal à x. Bien entendu, ce choix du revenu maximum est conventionnel, mais, ainsi que le montre le calcul, il conduit à des résultats vraisemblables et pratiquement utilisables.

Nous appellerons ce revenu M_T, la lettre T indiquant que nous le considérons comme un revenu théorique. Son calcul est facile.

La formule de Pareto peut être mise sous la forme :

$$\frac{N_x}{N} = \left(\frac{m}{x}\right)^\alpha. \qquad (2)$$

Si, dans cette formule, on fait $N_x = 1$, on a :

$$\frac{1}{N} = \left(\frac{m}{M_T}\right)^\alpha, \text{ d'où } M_T = m . N^{\frac{1}{\alpha}}. \qquad (3)$$

Telle est la valeur du revenu maximum, en fonction de la population, que nous adopterons.

Le chiffre de la population influe non seulement sur l'inégalité, mais aussi sur le revenu moyen de l'ensemble de la répartition.

Si l'on désigne le revenu moyen par μ et si l'on pose :

$$\sigma = \frac{x_1}{x_2} \text{ et } x_1 < x_2,$$

on obtient :

$$\mu = \frac{\alpha}{\alpha - 1}\, x_1 . \frac{1 - \sigma^{\alpha - 1}}{1 - \sigma^\alpha}. \qquad (4)$$

Si l'on fait $x_1 = m$ et $x_2 = M_T$, on a :

$$\mu = \frac{\alpha}{\alpha - 1}\, m \left(\frac{1 - \dfrac{1}{N^{\frac{\alpha - 1}{\alpha}}}}{1 - \dfrac{1}{N}}\right). \qquad (5)$$

Nous pouvons calculer le revenu moyen d'une autre manière. La somme des revenus supérieurs à x_1 et au plus égaux à x_2 est donnée par la formule :

$$T_{x_1}^{x_2} = \frac{\alpha}{\alpha - 1}\, m^\alpha N \left(\frac{1}{x_1^{\alpha - 1}} - \frac{1}{x_2^{\alpha - 1}}\right) \qquad (6)$$

ce qui peut s'écrire :

$$T_{x_1}^{x_2} = \frac{\alpha}{\alpha - 1}\, m\, N \left[\left(\frac{m}{x_1}\right)^{\alpha - 1} - \left(\frac{m}{x_2}\right)^{\alpha - 1}\right]$$

Faisons maintenant $x_1 = m$ et $x_2 = M_T$.

On a :

$$T = \frac{\alpha}{\alpha - 1}\, N\, m \left[1 - \sigma^{\alpha - 1}\right]$$

Or $\mu = \dfrac{T}{N}$,

d'où :

$$\mu = \frac{\alpha}{\alpha - 1}\, m \left[1 - \sigma^{\alpha - 1}\right]$$

ce qui peut s'écrire :

$$\mu = \frac{\alpha}{\alpha - 1}\, m \left[1 - \frac{1}{N^{\frac{\alpha - 1}{\alpha}}}\right] \qquad (7)$$

La valeur de μ est différente de celle trouvée dans la formule (5), car le dénominateur de (5) dans le terme entre crochets se trouve supprimé.

Pourquoi n'avons-nous pas de dénominateur? Le nombre des contribuables dont le revenu est supérieur à x_1 et en plus égal à x_2 est $Nx_1 - Nx_2$; en particulier le nombre total des contribuables est :

$$N = N_m - N_M. \qquad \text{Mais } N_m = N\ \text{-}$$

Nous sommes donc conduits à négliger N_M, c'est-à-dire σ^α. Autrement dit nous ne pouvons nous servir de la loi de Pareto pour la partie riche, même en admettant qu'elle s'applique rigoureusement, qu'en négligeant σ^α, ce que nous ferons toujours.

Une objection vient immédiatement à l'esprit. Pourquoi faire usage d'une formule qui s'étend à toute la répartition, alors que la loi de Pareto ne s'applique plus du côté du revenu maximum? On fausse ainsi les résultats obtenus en y introduisant une source d'erreurs qui disparaîtrait, si on restreignait son étude à la portion de la répartition où la loi de Pareto s'adapte d'une façon très satisfaisante à la réalité.

Je vais montrer que la considération des revenus les plus élevés n'a, du fait que la loi de Pareto est une loi approchée, qu'une influence insignifiante sur le résultat du calcul du revenu moyen et que par contre il y a de grands avantages à se servir d'une formule embrassant l'ensemble de la répartition.

Utilisons pour cela la statistique des revenus français de 1931 (l'année 1931 a été prise au hasard).

La valeur paraissant le mieux convenir pour α étant 1,712 et le nombre total des contribuables fourni par la statistique étant 2.080.164, on obtient comme revenu maximum :

$$M_T = 48.950.000 \text{ francs.}$$

Si, en conservant la même valeur de α, nous abaissons le nombre de détenteurs de revenus de 2.080.164 à 1 million, nous aurons :

$$M_T = 31.920.000 \text{ francs.}$$

Le chiffre de la population exerce donc une influence considérable sur le revenu maximum théorique. Je vais montrer qu'il ne fait varier la valeur du revenu moyen que d'une quantité très petite.

Ainsi avec les données numériques précédemment indiquées, le chiffre calculé du revenu moyen est 23.988 francs contre 23.993, chiffre observé.

Si on abaisse le chiffre de la population à 1 million, α restant constant, le revenu moyen calculé est seulement abaissé de 23.988 francs à 23.968 francs, ce qui ne représente qu'une très faible diminution.

Cet exemple nous montre que ce qui se passe dans la région des très gros revenus n'exerce sur l'ensemble de la répartition qu'une influence bien faible.

Il est vrai qu'on aurait un écart plus grand, si on avait une valeur de α plus petite, car celle de 1,712 paraît plutôt élevée. Mais, même dans le cas de valeurs de α faibles, qui du reste ne doivent pas tomber dans la pratique courante au-dessous de 1,5, pour ce qui concerne les statistiques de revenus, on peut encore maintenir l'affirmation précédente.

Il faut du reste préciser la nature de l'influence du revenu maximum sur l'ensemble de la répartition. L'apport de très gros revenus, du seul fait d'une population plus élevée, accroît le revenu moyen de la population totale, mais ne change rien à la situation individuelle de l'ensemble des contribuables. Certes ceux-ci ressentiront plus profondément leur médiocrité en la comparant à la richesse des possesseurs de très grosses fortunes, mais leur revenu moyen sera exactement le même que si ceux-ci n'existaient pas.

Il n'y a pour s'en convaincre qu'à se reporter à la figure 1. Supposons que la répartition $A_1 C_1$ se transforme en la répartition $A_2 C_2$, le revenu moyen des individus situés dans la répartition $A_1 C_1$ restera exactement le même; il sera toujours exprimé par la formule (4), où figureront les mêmes nombres.

L'extrémité riche de la répartition n'exerçant qu'une faible influence sur la répartition d'ensemble, il n'y a donc que peu d'inconvénients à utiliser une formule qui embrasse la totalité des données de la statistique.

Mais on peut ne pas se tenir satisfait de ce raisonnement.

Deux cas, dira-t-on, sont à envisager. Ou α est petit et l'influence de l'extrémité riche de la répartition sur l'ensemble de celle-ci devient sensible; ou α est grand et elle est très faible.

Dans le premier cas, la formule (7) ne donne que des résultats trop peu approchés et il vaut mieux prendre comme revenu maximum une valeur plus petite que M_T.

Dans le second cas, pourquoi ne pas préférer à la formule proposée la formule du revenu moyen où M est infini, beaucoup plus simple et donnant lieu à des calculs très rapides? L'écart entre les résultats fournis par les deux formules est en effet très faible.

Je répondrai à la première objection que, comme je l'ai déjà signalé, les valeurs de α très faibles ne se rencontrent pas quand il s'agit des revenus. On ne les trouve que pour les répartitions successorales, auxquelles s'adapte mal la formule de Pareto. Quoi qu'il en soit, du reste, l'expression proposée aura toujours le mérite de donner sur la répartition totale des indications moins défectueuses que la formule du revenu moyen pris dans un intervalle infini.

Dans le deuxième cas, celui où α est grand, on peut certainement se servir avantageusement de la formule simple, où le revenu maximum est infini, mais il peut aussi être intéressant d'utiliser une expression qui donne des résultats légèrement meilleurs. Tout dépend du degré de précision qu'on veut atteindre.

Énumérons maintenant les avantages qu'il y a à employer la formule (7) :

1° L'extrémité riche de la répartition, bien qu'elle exerce peu d'influence sur l'ensemble des éléments considérés, constitue leur partie la plus dynamique. Or il est utile d'avoir sur elle quelques indications. Sans doute la loi de Pareto ne peut nous les fournir très précises, parce qu'elle s'applique mal dans la région des très gros revenus, mais elle nous donne au moins des ordres de grandeur qu'il peut être intéressant de connaître;

2° En étudiant une répartition dans l'intervalle total m, MT, on utilise une formule un peu plus simple que si on considérait un intervalle quelconque x_1 x_2;

3° En embrassant la totalité de la répartition, on se sert d'un procédé rigide qui exclut tout arbitraire. Si on s'arrête à une valeur quelconque x du revenu, parce qu'on estime que la loi de Pareto s'applique au delà de cette valeur avec une approximation insuffisante, quelle sera cette valeur de x? Supposons que la répartition s'étende de 10.000 francs, revenu minimum à 50 millions, revenu maximum, choisira-t-on comme limite supérieure de ses investigations 100.000 francs, 200.000 francs ou 1 million? Ce choix sera arbitraire et les résultats obtenus donneront souvent lieu à des confusions regrettables. J'estime, pour ma part, que c'est un très grand avantage de les éviter.

Munis d'une formule satisfaisante, nous sommes à même de nous en servir, s'il y a lieu, pour définir l'inégalité. N'ayant pas le temps de traiter la question à fond, je me bornerai à citer les deux définitions les plus connues de l'inégalité, sans même les critiquer. et j'en proposerai deux nouvelles.

L'inégalité des revenus a été définie, le rapport au revenu total de la somme des écarts en valeur absolue de tous les revenus individuels avec le revenu moyen.

Il existe aussi une autre définition très connue de l'inégalité, c'est celle de M. Corrado Gini.

M. Gini, au lieu de prendre l'écart de chaque revenu individuel avec le revenu moyen, considère la somme des écarts qu'ont entre eux les revenus individuels et il appelle *différence moyenne* la moyenne arithmétique de ces écarts pris en valeur absolue.

Ces deux définitions de l'inégalité sont excellentes. Aussi peut-on se demander quel besoin il y a d'en proposer de nouvelles.

Cependant, je crois qu'il peut être utile d'y ajouter d'autres définitions, non pas pour les supplanter, mais pour les compléter.

La question de l'inégalité divise les sociétés modernes; elle provoque entre les classes des luttes ardentes et même sanglantes. Si l'on arrive à réaliser l'égalité des fortunes, aux yeux des uns le monde connaîtra une ère de prospérité radieuse. Les autres, au contraire, croient que si la société humaine veut s'affranchir de la loi inéluctable de l'inégalité, à laquelle est soumise la nature entière, elle entrera dans la nuit. Même les esprits les moins réfléchis et les plus simples comprennent obscurément que la solution qu'on donnera au problème de l'inégalité sera grave pour l'avenir des hommes.

L'étude scientifique de l'inégalité ne peut donc pas rester le domaine exclusif des statisticiens et des mathématiciens. Les économistes, même dépourvus de culture mathématique, doivent pouvoir y apporter leur concours. Ils verront les phénomènes sous un angle différent de celui des mathématiciens et des statisticiens; les uns et les autres se complèteront mutuellement. Il faut aussi que l'homme de la rue puisse s'intéresser à la question, car la connaissance de la vérité est le meilleur moyen d'atténuer la lutte des classes.

La première tâche à remplir est d'établir si le système capitaliste a fait croître l'inégalité, comme l'affirment ses adversaires, ou l'a fait décroître, ainsi que Paul Leroy-Beaulieu a tenté de l'établir, ou l'a laissée stationnaire, ainsi que l'a soutenu M. Gibrat dans son ouvrage *Les inégalités économiques*.

Mais, pour résoudre cette question, une définition de l'inégalité est nécessaire. Nous la chercherons accessible à tous les esprits.

Or les définitions de l'inégalité, précédemment énumérées, sont trop abstraites pour être saisies par tous. Si l'on propose à l'homme de la rue, comme mesure de l'inégalité, le rapport en revenu total de la somme des écarts en valeur absolue de tous les revenus individuels avec le revenu moyen, il ne comprendra pas de quoi on lui parle; la différence moyenne, exprimée par le rapport de concentration, lui paraîtra une notion encore plus obscure. Il faut donc trouver autre chose.

Aussi, dès 1922, j'ai défini devant la Société de Statistique l'inégalité au moyen du rapport de la somme des revenus au-dessus du revenu moyen à la somme des revenus au-dessous du revenu moyen. J'opposais ainsi la masse des revenus des riches à la masse des revenus des pauvres.

Mais cette définition de l'inégalité a encore quelque chose d'abstrait. Il n'apparaît pas d'une manière évidente que l'inégalité augmente, parce que ce rapport s'accroît. On démontre, si la répartition obéit à la loi de Pareto, que, quand le revenu moyen augmente, le nombre de personnes possédant plus que le revenu moyen diminue. On pourrait être tenté d'intervertir l'interprétation du rapport et de dire que plus il décroît, plus l'inégalité s'accroît. Il est donc nécessaire de disposer d'une meilleure définition.

Les statisticiens anglo-saxons désignent par le terme de revenu équatorial le revenu qui sépare par moitiés égales le revenu total, la somme des revenus au-dessus de ce revenu étant égale à la somme des revenus au-dessous de lui.

J'ai donc pensé à considérer non plus le rapport des sommes des revenus, mais celui du nombre des individus possédant un revenu supérieur au revenu équatorial au nombre des individus possédant un revenu inférieur à ce revenu. Plus le nombre des premiers décroît par rapport à celui des seconds, plus, en d'autres termes, la moitié du revenu total entrée en des mains moins nombreuses, plus l'irrégularité augmente.

C'est la conception qui a servi de base au « slogan » des 200 familles.

Mais là encore le rapport d'inégalité n'est pas décelé d'une manière absolument claire. Il n'est pas évident qu'elle s'accroisse, parce que le rapport en question diminue. De plus, ce rapport d'inégalité présente l'inconvénient, s'il est pris dans un intervalle fini et si on veut l'exprimer au moyen de la formule de Pareto, d'être trop compliqué pour être pratiquement utilisable.

M. Fréchet, dans les observations qu'il a faites à la suite de ma communication à la Société de Statistique en novembre 1940, a signalé l'importance du revenu médian.

Cette remarque m'a suggéré l'idée qu'il y aurait avantage à substituer dans le rapport d'inégalité que je viens de proposer le revenu médian au revenu moyen.

Je rappelle que le revenu médian est celui possédé par l'individu qui, à une unité près, si le nombre total des individus est pair, se trouve au milieu de la répartition. Il y aura donc un nombre égal d'individus plus riches que lui et un nombre égal d'individus plus pauvres que lui.

Je prends comme rapport d'inégalité la somme des revenus au-dessus de ce revenu médian, possédée par la moitié la plus riche des individus, à la somme des revenus au-dessous de ce revenu médian possédée par la moitié la plus pauvre des individus. La société est donc séparée en deux classes, la première et la seconde. Plus les revenus de la première croîtront relativement à ceux de la seconde, plus l'inégalité augmentera.

Cette définition ne peut prêter à aucune erreur d'interprétation et elle est susceptible d'être comprise par tout le monde. Elle a donc la valeur sociale que nous recherchons.

Appelons R ce nouveau rapport d'inégalité. Pour une répartition quelconque, il s'exprime par la formule :

$$R = \frac{T_p}{T - T_p} \tag{8}$$

Si on opère dans la répartition de Pareto et si on fait le revenu maximum infini, on a :

$$R = \frac{1}{2^{\frac{\alpha-1}{\alpha}} - 1} \tag{9}$$

Si on considère un intervalle fini, on a :

$$R = \frac{\frac{\alpha}{\alpha - 1}\frac{m}{\mu}\left[\frac{1}{2^{\alpha-1}} - \sigma^{\alpha-1}\right]}{1 - \frac{\alpha}{\alpha-1}\frac{m}{\mu}\left[\left(\frac{1}{2}\right)^{\frac{\alpha-1}{\alpha}} - \sigma^{\alpha-1}\right]} \tag{10}$$

Si on veut opérer rapidement, on calculera l'inégalité en se servant de la formule de Pareto avec le revenu maximum infini (formule 9). Sous la condition que α soit suffisamment grand, les erreurs seront légères et, comme elles auront toujours lieu dans le même sens, elles seront dans bien des cas sans importance, par exemple, si l'on veut comparer l'inégalité pour différents pays et différentes époques de structure voisine.

Si on veut arriver à des résultats précis, on emploiera la formule complète. Toutefois, il sera encore plus précis, quoique un peu plus long, d'utiliser directement les données de la statistique, en faisant les interpolations nécessaires.

Je proposerai encore une autre définition de l'inégalité, non pas au titre de sa valeur sociale, mais parce qu'elle a le double avantage, si on considère une répartition soumise à la loi de Pareto, d'être très simple et de conduire à un rapport immédiatement calculable avec les données de la statistique.

Au lieu de partir du concept de l'*écart* pour mesurer l'inégalité de différentes grandeurs entre elles, nous nous appuierons sur le concept du *rapport*. Ces deux concepts su retrouvent du reste dans le langage courant. Si on distribue des parts de gâteau à des enfants, l'un d'eux, mécontent de son lot, pourra dire en parlant de son voisin : « Il en a en plus que moi. » C'est le concept de l'écart. Mais il se plaindra peut-être en ces termes : « Il en a eu deux fois plus que moi. » C'est le concept de la proportion.

Considérons une répartition quelconque de revenus et choisissons comme valeur repère le revenu minimum m.

Nous représenterons l'inégalité relative à un revenu x_1 individuel par $\frac{x_1}{m}$.

S'il y a n_{x_1} individus ayant le revenu x_1, l'inégalité pour le revenu x_1 sera $n_{x_1}\frac{x_1}{m}$.

Nous mesurerons l'inégalité de la répartition par la moyenne de toutes ces inégalités individuelles.

L'inégalité totale sera :

$$\frac{n_{x_1}\frac{x_1}{m} + n_{x_2}\frac{x_2}{m} \text{ et...} \qquad + n_M \frac{M}{m}}{N}$$

N étant le nombre total des individus,

$$N = n_{x_1} + n_{x_2} + \ldots\ldots n_M.$$

Mais ce rapport peut s'écrire :

$$\frac{n_{x_1}\, x_1 + n_{x_2}\, x_2 + \ldots\ldots n_M\, M}{m\, N}.$$

Mais $n_{x_1}\, x_1 + n_{x_2}\, x_2 + \ldots\ldots n_M\, M = T$ (revenu total).

Donc on a : $\frac{T}{m\,N}$.

Mais $\frac{T}{N}$ égalant μ, on obtient $\frac{\mu}{m}$.

L'inegalité est donc définie par le rapport du revenu moyen au revenu minimum.

Ce rapport $\frac{\mu}{m}$ a la simplicité d'un indice. On a en effet souvent proposé les indices de l'inégalité, qui la définissent d'une manière moins parfaite, mais plus simple que les rapports d'inégalité que nous avons énumérés. Mais $\frac{\mu}{m}$ est une définition précise de l'inégalité et est d'une telle simplicité que l'emploi des indices devient inutile.

Si l'on veut toutefois se servir d'un indice, celui qui nous paraît le plus recommandable est l'exposant α, ou mieux son inverse $\frac{1}{\alpha}$. Tous les rapports d'inégalité, dignes d'être retenus, varient en effet dans le même sens que $\frac{1}{\alpha}$. Il suffit du reste de jeter un coup d'œil sur la représentation doublement logarithmique de la fonction de Pareto, pour voir que plus $\frac{1}{\alpha}$ croît, moins la répartition est ramassée sur elle-même et devient, de ce fait, moins inégale. Aussi M. Gibrat a-t-il eu raison de prendre comme indice de l'inégalité $\frac{1}{c}$ dans la formule dont il a proposé l'emploi, a jouant dans cette formule le même rôle que α dans celle de Pareto.

Mais $\frac{\mu}{m}$ l'emporte aussi par son exactitude absolue, pourvu toutefois qu'on calcule l'exposant α par la formule du revenu moyen.

On procédera pour ce calcul de la manière suivante :

Si on fait M infini, on emploiera la formule :

$$\mu = \frac{\alpha}{\alpha - 1}\, m$$

et si on étudie la répartition de m à M_T, on se servira de la formule (7) :

$$\mu = \frac{\alpha}{\alpha - 1}\, m \left[1 - \frac{1}{N^{\frac{\alpha - 1}{\alpha}}}\right]$$

en utilisant les données de la statistique, m et μ pour la première formule, m, μ et N pour la seconde. Dans le cas où on fera usage de l'équation (7), on déterminera α par tâtonnements, en recherchant celle des valeurs de α qui satisfait à cette équation.

J'avais déjà, en 1940, dans la communication que j'ai faite ici, proposé cette méthode, mais je n'en avais pas retenu l'emploi, « parce que la manière dont sont établis les documents officiels peut faire varier le revenu moyen dans une large mesure ».

Cette objection n'a pas de sens. Nous sommes bien forcés d'utiliser les documents statistiques, tels qu'ils nous sont fournis. L'erreur qu'ils peuvent contenir sur le revenu moyen se répercutera sur tous les autres éléments de la répartition. On ne pourra jamais viser qu'à construire une courbe de Pareto se rapprochant le plus possible de celle de la statistique, que celle-ci soit exacte ou non.

$\frac{\mu}{m}$ est donc d'une exactitude absolue. Que nous calculions directement sa valeur d'après les données, $\frac{T}{m\,N}$, ou que nous la déterminions au moyen de la valeur de α, le résultat sera évidemment le même dans les deux cas, puisque nous avons déterminé α au moyen du calcul de $\frac{\mu}{m}$ tiré des données.

Or je soutiens que la méthode de détermination de α par la formule du revenu moyen est la meilleure.

On peut diviser les méthodes de détermination de α en deux groupes, celles où l'on ne pondère pas et celles où l'on pondère.

Les premières sont caractérisées par le fait qu'elles attachent une importance sinon égale, tout au moins très peu différente, aux différents points de la répartition.

Les méthodes de pondération procèdent autrement. Elles substituent à la représentation réelle un schéma de Pareto, où α est calculé d'une manière telle, qu'on attribue à chaque erreur, c'est-à-dire à chaque discordance avec la répartition réelle, seulement l'importance qu'elle mérite.

En d'autres termes, on s'efforce dans les méthodes de pondération d'obtenir les chiffres qui se rapprochent le plus possible de ceux des données essentielles et on se préoccupe peu des caractères secondaires de la répartition.

Toutes les méthodes, du reste, telles que la méthode graphique, la méthode des moindres carrés, la méthode des moindres écarts, la méthode de la moyenne arithmétique des différentes valeurs de α, se prêtent à la pondération.

La méthode graphique entraîne toujours une certaine pondération, tout à fait arbitraire, d'ailleurs. On aura, en effet en l'employant, une tendance à attribuer une moindre importance à l'extrémité riche de la répartition, où s'incurve la ligne droite logarithmique de l'équation de Pareto.

La méthode de détermination de α par la formule du revenu moyen est par excellence une méthode de pondération, car chaque revenu individuel figure dans la somme des revenus pour sa valeur exacte, donc pour son importance exacte, et chaque individu figure aussi dans la somme des individus pour son importance exacte, c'est-à-dire que chacun des individus étant compté pour 1, ils ne se différencient que par leur fortune.

Je dis que les résultats fournis par les méthodes de pondération sont meilleurs que ceux obtenus à l'aide des autres méthodes.

Pour le prouver, je ferai d'abord appel au bon sens. Dans le cas où il est impossible de déterminer exactement tous les caractères d'un phénomène, ne donnera-t-on pas une image plus fidèle de la réalité en s'attachant à ses caractères essentiels et en laissant dans l'ombre ses caractères secondaires?

Supposons par exemple qu'un peintre soit chargé de faire un portrait en pied et qu'on ne lui donne pour cela qu'un nombre d'heures tout à fait insuffisant. Que fera-t-il? Il consacrera la plus grande partie de son temps à dessiner et à peindre la tête et ne fera qu'esquisser le vêtement.

Voici un nouvel exemple où la pondération s'impose d'une manière encore plus frappante, parce qu'il serait absurde de ne pas tenir compte de l'importance des éléments considérés. Supposons que la France soit divisée en deux partis politiques, les jaunes et les verts et qu'on fasse une enquête sur l'opinion publique d'une ville de 1 million d'habitants. L'enquête révèle que cette ville doit compter 60 % de verts et 40 % de jaunes. Puis, poursuivant l'enquête, on constate dans un petit bourg de 1.000 habitants l'existence de 90 % de jaunes et de 10 % de verts. Si l'on prend la moyenne arithmétique du pourcentage de l'opinion dans les deux villes, sans tenir compte du nombre de leurs habitants, on concluera que les jaunes sont en grande majorité, ce qui est évidemment absurde.

Il faut donc donner à chaque recensement un coefficient proportionnel à la population sur lequel il porte, car l'élément essentiel que l'enquête doit mettre en lumière, c'est l'opinion de la population et non pas celle d'un groupement, abstraction faite de sa population.

Les méthodes de pondération doivent donner des résultats plus exacts et, par conséquent, des erreurs moins fortes que les méthodes sans pondération, pour les caractères essentiels d'une répartition de revenus, tels que le revenu total, le revenu moyen, l'inégalité et, par contre, des erreurs plus fortes pour les caractères secondaires tels que le nombre des contribuables possédant des revenus très élevés.

Et ceci nous conduit à la pondération des erreurs. Du fait que les divers chiffres fournis par les données n'ont pas la même importance, il en résulte que les erreurs commises à leur sujet n'ont pas non plus la même importance.

Reprenons l'exemple des deux partis politiques, les jaunes et les verts. Une erreur de 25 % sur le pourcentage des opinions d'une ville de 1 million d'habitants a une importance mille fois plus grande qu'une erreur sur l'opinion d'un village de 1.000 habitants.

Des arguments s'appuyant sur le bon sens pour montrer la supériorité d'une méthode sont déjà très puissants. Il serait cependant utile de leur donner une base plus précise.

Pour cela il faut partir d'une estimation satisfaisante de l'écart entre la répartition théorique et les données statistiques. Je propose la suivante :

Abandonnons la représentation doublement logarithmique et définissons ce que nous appelons courbe de la répartition observée.

Nous interpolerons dans chaque classe de revenus fournie par la statistique un arc de courbe de Pareto. C'est l'ensemble de ces arcs que nous désignerons par courbe de la répartition observée.

La courbe de la représentation théorique et celle de la représentation observée se coupent en un certain nombre de points. Tantôt l'une des deux courbes est au-dessous de l'autre et tantôt c'est le contraire. Nous ferons la somme des aires comprises entre les deux courbes et nous admettrons que plus elle est petite, plus l'erreur est faible.

Comment évaluer la surface d'erreur? Une intégration graphique, étant donné que la courbe de Pareto est très rapprochée de celle des données serait peu pratique. Ce quil y a de mieux est de calculer les différentes aires d'erreur. Pour cela il faut d'abord déterminer les points d'intersection des deux courbes.

Nous supposerons que la formule de Pareto s'applique d'une manière parfaite dans la répartition observée entre chaque valeur de x repère fournie par les données statistiques. α_i étant l'exposant pour l'interpolation entre x_i et x_{i+1}, nous le déterminerons au moyen de l'équation :

$$\alpha_i = \frac{\log Nx_i - \log Nx_{i+1}}{\log x_{i+1} - \log x_i}$$

Si nous prenons les valeurs de N_x des données statistiques, telles que $N x_i$ Nx_{i+1} et si nous comparons ces valeurs à celles que l'on obtient au moyen de la formule de Pareto, nous avons une erreur soit en plus, soit en moins. Si pour deux valeurs successives de x fournies par la statistique, telles que Nx_i, Nx_{i+1}, l'erreur est de signe contraire c'est que la courbe de Pareto coupe celle des données entre les points de coordonnées $N x_i$, x_i et Nx_{i+1}, x_{i+1}. Aux points d'intersection les valeurs de N_x sont égales.

Soit X_i la valeur de x pour laquelle les deux courbes se coupent entre x_i et x_{i+1} et α' l'exposant α dans le schéma de Pareto, où il conserve une valeur constante, nous rappelant que $A = N m^\alpha$, nous écrirons N_{X_i} de deux manières différentes :

$$N_{X_i} = \frac{N\ m^{\alpha'}}{X_i^{\alpha'}} = \frac{N\ m^{\alpha_i}}{X^{\alpha_i}}$$

et cela nous permet de calculer X_i qui est la seule inconnue dans cette relation. Nous procéderons de même pour les autres points d'intersection et nous obtenons X_i, X_j, X_k, etc..., rangés par valeurs croissantes.

Pour la répartition observée il faut avoir la somme des revenus correspondant à chaque arc d'interpolation. Cette somme nous est donnée directement par la statistique pour tout arc sur lequel ne se trouve pas de point X_i. Nous ferons le calcul seulement pour les autres points.

Par ailleurs, on calculera les diverses sommes de revenus dans le schéma de Pareto. On fera la différence de ces sommes dans les deux répartitions et on obtiendra ainsi les aires d'erreur.

Ces calculs ne sont pas compliqués, mais ils sont assez longs et le temps m'a manqué pour les entreprendre. En attendant je me suis contenté du mode d'estimation des erreurs généralement employé, mais en l'améliorant par la pondération. Je vais montrer que, grâce à la pondération, les résultats procurés par cette méthode, bien que très inférieurs à ceux qu'on obtiendrait par la méthode que je viens de proposer, sont cependant influencés par les surfaces d'erreur.

On a l'habitude de prendre les rapports entre les valeurs de N_x fournies par la statistique et les valeurs correspondantes obtenues au moyen du schéma de Pareto. Ces rapports représentent l'erreur pour chaque nombre. On prend la moyenne arithmétique de tous les rapports qui donne le pourcentage d'erreurs général.

Ainsi que je l'ai exposé dans ma communication de 1940, j'améliore cette méthode par une double pondération. Je pondère chaque valeur de N_x d'après la somme des revenus que possèdent les individus considérés; je pondère ensuite chaque erreur, toujours d'après la somme des revenus.

Les aires interviennent donc aussi dans cette méthode, mais d'une manière moins directe que dans le premier mode d'évaluation des erreurs. Là les erreurs étaient évaluées par des aires non pondérées; ici elles le sont par les segments d'ordonnées pondérées par les aires. Les deux méthodes s'apparentent donc entre elles, mais la seconde, recommandable par sa simplicité, demeure, même améliorée par la pondération, grossière vis-à-vis de la première.

Quel que soit le monde d'estimation des erreurs, le résultat obtenu dépend de la façon dont la statistique est établie, nous voulons dire du choix des valeurs des revenus limitant les classes des contribuables. Mais le fait que certaines des valeurs N_x se trouvant aux points d'intersection des deux courbes seront très éloignées ou très voisines des valeurs de N_x, valeurs repères données par la statistique, a peu d'influence dans la première méthode et en a beaucoup dans la seconde.

Supposons que, par hasard, avec la méthode usuelle, les valeurs N_x et N_{x_i} soient égales, l'erreur apparaîtra nulle, alors qu'elle ne l'est pas. Dans le cas inverse, l'erreur sera exagérément grossie.

Voici un tableau des résultats obtenus :

Revenus français de 1931.

$\alpha = 1{,}712$.

	Nombre des individus — Observés	Nombre des individus — Calculés (A a été calculé seulement dans la première tranche)	Pourcentage des erreurs avant pondération	Produit des erreurs par les poids
Plus de 10.000 francs	2.080.164	2.080.164	0	0
20.000 —	651.169	634.900	— 2,49	75,65
30.000 —	319.150	317.100	— 0,64	19,99
40.000 —	192.838	193.800	+ 0,50	9,10
50.000 —	131.110	132.300	+ 0,90	13,45
100.000 —	41.519	40.370	— 2,80	25,23
200.000 —	12.897	12.320	— 4,47	23,62
500.000 —	2.119	2.567	+21	44,13
1.000.000 —	494	783	+59	59,32
TOTAUX				264,49

Divisons la somme du produit des erreurs par les poids par la somme des poids, c'est-à-dire 264,49 par 152,63, on obtient comme erreur 1,733 %.

On voit que l'approximation est excellente. Elle se montre très supérieure à celle que j'avais obtenue avec la méthode exposée ici en 1940, qui consistait à prendre la moyenne arithmétique des différentes valeurs de α pondérées. La valeur ainsi obtenue pour α était 1,725 et l'erreur totale était de 2,93 %.

Avec les méthodes sans pondération, les valeurs de α auraient été encore plus élevées et l'approximation aurait été médiocre.

Faut-il s'étonner de pareils résultats? Non, car ils paraissent logiques. Je puis, en effet, raisonnant par syllogisme, dire :

Les méthodes de pondération donnent la meilleure approximation.

La meilleure pondération est donnée par le calcul de α par la formule du revenu moyen.

Donc le calcul de α par la méthode du revenu moyen donne la meilleure approximation.

En d'autres termes, parmi le nombre infini de courbes de Pareto qu'on peut tracer pour exprimer une distribution de revenus, celle où α est déterminé par la formule du revenu moyen est privilégiée par rapport à toutes les autres et donne en valeur absolue la somme minima de surfaces d'erreurs.

Mais un raisonnement syllogistique n'est pas rigoureux, car il fait état de vraisemblances et il est impossible de démontrer mathématiquement que la courbe de Pareto d'exposant α déterminé par la formule du revenu moyen donne en valeur absolue la surface minima d'erreurs. On ne peut que le constater empiriquement. Comme je l'ai dit, des calculs vont être entrepris pour vérifier cette hypothèse.

Quoi qu'il en soit du reste, la valeur de α déterminée par la formule du revenu moyen est celle qu'il faut choisir. S'il en existait une meilleure, comment pourrait-on la découvrir? On la reconnaîtrait, d'après notre point de vue, à ce signe qu'elle serait celle de la répartition de Pareto donnant la surface d'erreurs minima. On serait obligé de la calculer par tâtonnement. Ce procédé serait si long et procurerait de si minimes avantages qu'il doit être rejeté.

En outre, la fonction de Pareto, où α a été déterminé par la valeur du revenu moyen, a cette immense supériorité sur toutes les autres fonctions de Pareto, représentatives d'une distribution donnée, qu'elle offre une somme de surfaces d'erreurs nulle, si l'on tient compte du signe des erreurs. En effet, il faut bien qu'il en soit ainsi, pour que la somme totale des revenus dans cette répartition de Pareto soit égale à la somme totale des revenus de la répartition observée. De plus, cette fonction de Pareto est la seule à jouir d'une telle propriété, car à une valeur donnée de T et de N correspond une seule valeur de α.

En outre, un avantage très précieux de la méthode est qu'avec elle le calcul est extrêmement rapide. Si on emploie la formule $\mu = \frac{\alpha}{\alpha - 1} m$, M étant pris infini, il est presque instantané. Si l'on fait M fini, on est obligé d'utiliser la formule (7), mais le calcul, quoique un peu plus compliqué n'est pas long.

Un autre avantage de ce mode de détermination de α est que, par le seul fait que son emploi se recommande à tous, il doit permettre d'uniformiser les observations faites dans le cadre de la fonction de Pareto.

La méthode de détermination de α par le revenu moyen est donc la méthode par excellence; elle surpasse tellement toutes les autres, qu'il devient inutile de discuter la valeur respective de ces dernières. Elle était déjà du reste connue, car je la trouve signalée par Bortkiewicz qui, dans son Mémoire (1), la qualifie de « méthode interne », l'opposant aux

(1) *Institut international de Statistique.* Tokio, 1930.

autres méthodes qu'il qualifie de « méthodes externes ». Mais Bortkiewicz ne lui donne pas plus d'importance qu'aux autres modes de calcul de α et je crois que, jusqu'à présent, personne n'en avait démontré l'excellence, autant du moins que je puisse le savoir, car je n'ai pas encore terminé l'étude bibliographique de la fonction de Pareto.

α est, en général, déterminé par n'importe quelle méthode et parfois même très mal. Ainsi, si on se sert de la méthode indiquée par Vilfredo Pareto (1) et appliquée par M. Bresciani-Turroni, on trouve pour les revenus prussiens de 1913 : $\alpha = 1{,}51$, ce qui donne une erreur de 20 % environ sur le revenu moyen (2).

Une telle manière de calculer α est-elle admissible? Ce qui augmente encore la confusion, au sujet des résultats des différentes observations, c'est que ,la plupart du temps, on ne prend même pas la peine d'indiquer de quelle manière on a calculé α.

Une méthode de détermination de α, supérieure aux autres par sa rapidité et sa précision, si elle était adoptée par l'ensemble des statisticiens, permettrait d'obtenir de la formule de Pareto une approximation bien meilleure et d'étendre par conséquent son champ d'application.

Une telle méthode doit avoir son emploi non seulement dans la statistique des revenus, mais dans tous les cas où on se sert de la formule de Pareto. Or celle-ci est applicable, ainsi que l'a montré M. Amy dans les observations qu'il a faites à la suite de ma communication de 1940, à un grand nombre de phénomènes divers. Une meilleure détermination de α est donc susceptible de rendre des services non seulement aux économistes, mais à de nombreux chercheurs.

Du reste, la méthode qui consiste à traduire un phénomène observé par une expression analytique, où intervient la moyenne comme élément principal, est absolument générale. Elle peut s'appliquer à toutes les fonctions à un seul paramètre.

Même si la fonction a plusieurs paramètres, elle est encore utilisable; elle impose alors aux différents paramètres la condition de satisfaire à la formule de la moyenne et restreindra ainsi le nombre de valeurs qu'on peut leur donner.

Certes, il n'y a dans l'extension possible de la méthode de la moyenne qu'une hypothèse, mais si cette hypothèse est confirmée par les faits, elle peut conduire à de très intéressants résultats.

Je résume ce qui a été acquis au cours de cette étude :

1° Il a été démontré que le chiffre de la population exerce une influence sur l'inégalité des revenus et sur le revenu moyen;

2° Deux définitions nouvelles de l'inégalité ont été proposées, l'une remarquable par sa valeur sociale, l'autre par sa simplicité;

3° La supériorité d'une des méthodes de détermination de α a été démontrée.

MOURRE.

DISCUSSION

M. FRECHET. — Le nombre des auditeurs a déjà montré à M. le baron MOURRE combien ses communications sont appréciées. On y trouve toujours matière à réflexion.

Parmi les différents points qui ont attiré mon attention, j'indiquerai d'abord l'influence du chiffre de la population. Cette influence se fait sentir dans toutes les populations soumises à une loi de probabilité déterminée. C'est ainsi que si l'on admet la loi de Laplace pour les erreurs d'observation, l'intervalle de variation d'une erreur (qui est infini au sens des probabilités) est fini pour n erreurs et croît avec n. C'est ainsi que M. GUMBEL a pu étudier la variation de l'âge extrême avec le chiffre de la population.

Il est donc vrai ou faux que le revenu moyen, l'indice d'inégalité, etc... dépendent du chiffre de la population selon qu'on les définit par les fréquences ou par les probabilités.

Si l'on définit par les probabilités, la comparaison de deux pays de population inégale en sera facilitée en principe; elle pourra, au contraire, perdre un sens précis en les définissant par les fréquences. Autrement dit si on compare deux pays de même population et si dans l'un les gros revenus sont beaucoup plus élevés que dans l'autre, il sera naturel de conclure à une différence dans la répartition des revenus. Mais si ces gros revenus sont plus élevés dans le pays le plus peuplé, une telle conclusion serait non fondée et pourrait être même contraire aux faits.

M. MOURRE a cité quelques définitions de l'indice d'inégalité des revenus. Nous en indiquerons d'autres plus loin. La plus ancienne est celle de Pareto qui le mesure par le coefficient α de sa loi. C'est une très mauvaise définition, de même que celle qui le mesure par le coefficient a de Galton Mac Alister. Une bonne définition doit être indépendante de la formule mathématique choisie pour la représentation de la répartition des revenus. Une telle définition étant choisie, on peut alors, pour des motifs pratiques, chercher à prouver que

(1) Il metodo di figurare i fenomeni economici, *Giornale degli Economisti*, 1896, p. 84.
(2) Annual Survey of Statistical Data; Pareto's Law and the Index of Inequality of Incomes. *Econometrica*, April 1939.

pour telle ou telle de ces formules représentatives, l'inégalité ainsi définie statistiquement peut être répérée au moyen d'un des coefficients de ces formules.

Le premier, sans doute, à avoir introduit une définition statistique de l'inégalité est sans doute l'ingénieur Herzen qui, dans une communication intéressante à plusieurs titres et peu connue (1) a proposé en 1900 de prendre pour indice $\frac{r'' - r'}{r}$ où r est le revenu moyen, r'' la moyenne des revenus supérieurs à r, r' celle des revenus inférieurs à r.

Halbwachs et moi (2) nous avions proposé comme indice le rapport du revenu moyen, au revenu probable, et Gumbel (3) avait choisi $1 - 2\frac{\nu'}{\nu}$ où ν' est le nombre de ceux des ν membres de la collectivité qui ont un revenu supérieur au revenu « équatorial ».

Passons à une autre question. M. Mourre a très justement fait observer que l'inégalité des revenus a une très grande importance sociale, mais que pour pouvoir en discuter, il faudrait pouvoir avoir une bonne définition de l'indice d'inégalité afin d'en comparer les valeurs suivant l'espace et le temps.

En principe, celles des diverses définitions rappelées plus haut qui sont indépendantes de toute hypothèse sur une représentation mathématique de la répartition, paraissent satisfaisantes et devoir conduire à des résultats à peu près équivalents. Mais il existe *actuellement* une difficulté *infiniment plus grave*, qui s'oppose radicalement à l'emploi légitime d'un indice d'inégalité des revenus, pour en tirer des conclusions quant à la comparaison des revenus de deux populations.

C'est qu'actuellement nous ne possédons pas les éléments statistiques relatifs à l'*ensemble* de la population. Les premières comparaisons publiées se sont basées sur les déclarations des revenus. Or non seulement celles-ci sont, de notoriété publique, inférieures à la réalité mais elles ne concernent que les revenus moyens ou supérieurs. Pour entreprendre l'application d'ordre social de M. Mourre, il faudrait arriver à une évaluation de la répartition de tous les revenus d'un même pays.

M. Huber estime qu'avec des mesures appropriées, on pourrait arriver à préciser la distribution et la limite des revenus les plus élevés, mais que la même opération se heurterait, pour les revenus les plus faibles, à une impossibilité pratique.

M. Fréchet constate avec satisfaction que M. Huber est d'accord avec lui pour signaler le danger de l'utilisation d'un indice d'inégalité fondé sur des statistiques partielles comme si cet indice concernait l'ensemble de la population, qu'il est en outre d'accord avec M. Huber pour reconnaîre la très grande difficulté d'établir une statistique globale concernant à la fois salaires et revenus. Mais il croit fermement qu'une telle entreprise est possible. Elle n'aboutirait d'abord qu'à des résultats très grossièrement approchés, mais qu'on réussirait peu à peu à préciser. Reconnaître quels peuvent être les montants de revenus réels correspondant à des revenus déclarés paraît une entreprise aussi difficile et pourtant nous avons entendu citer une estimation d'une insuffisance de 30 à 40 % sur l'ensemble des déclarations. Même si une telle estimation n'est pas très approchée de la réalité, elle donne une idée de son ordre de grandeur. Sa valeur pourrait d'ailleurs être contrôlée au moyen de sondages.

M. Huber fait observer que ces sondages sont effectivement pratiqués dans les enquêtes sur les budgets familiaux. Or ces enquêtes ont toutes fait ressortir les énormes difficultés rencontrées pour faire préciser les recettes des ménages les plus pauvres. On peut en voir une preuve dans le fait qu'une proportion souvent élevée de ces budgets se présente en déficit. Or un budget familial ne peut pas normalement être en déficit, il faut qu'il soit bouclé d'une manière quelconque, par endettement, par des dons en nature ou en espèces ou par des revenus que l'on préfère ne pas indiquer. Aussi M. Huber a la conviction que toute tentative de préciser avec une approximation suffisante la distribution des revenus les plus faibles et leur limite inférieure rencontrerait des difficultés pratiques insurmontables.

M. Mourre pense, comme M. Fréchet, qu'on ne peut appliquer d'une manière certaine les conclusions tirées des variations de l'inégalité dans les classes moyennes et riches aux classes les plus pauvres qui échappent en général aux statistiques fiscales. Il fait cependant observer que les statistiques saxonnes et prussiennes renseignent sur les revenus des classes les plus pauvres. Les statistiques saxonnes indiquent le montant des revenus inférieurs à 300 marks, ainsi que le nombre de leurs détenteurs. Il en est de même des statistiques prussiennes pour les revenus inférieurs à 900 marks.

On peut donc, pour ces deux pays, calculer le rapport d'inégalité, non pas en se servant de la formule de Gini, ou le $\frac{\mu}{m}$, qui toutes deux, impliquent la connaissance du revenu m minimum, non indiqué dans les statistiques saxonnes et prussiennes, mais en employant soit le rapport :

$$\frac{\text{Somme des revenus supérieurs au revenu médian}}{\text{Somme des revenus inférieurs au revenu médian}}$$

(1) *La répartition des revenus*, Bulletin de la Société Vaudoise des Sciences Naturelles, vol. XXXV, p. 282-295.
(2) *Le calcul des probabilités à la portée de tous*, chez Dunod, 1924, p. 145, exercice III.
(3) *Ein Mass der Konzenration bei pekuniären Verteilungen*, Archiv für Sozialwiss. und Sozialpol. Bd 58, pp. 1927, p. 113-139.

soit le rapport :

$$\frac{\text{Somme des écarts de chaque revenu individuel avec le revenu moyen}}{\text{Revenu total.}}$$

M. MOURRE reconnaît le bien-fondé de l'observation de M. FRÉCHET d'après laquelle le choix arbitraire qu'on peut faire du revenu minimum modifie l'expression de l'inégalité. Mais cette critique n'est valable que si on emploie la formule complète. Dans cette formule en effet figure N, nombre total des contribuables, qui est fonction du revenu minimum.

Mais, si on prend le revenu maximum infini, ce qui n'entraîne qu'une erreur insignifiante pourvu que α soit suffisamment grand, on a $\frac{\mu}{m} = \frac{\alpha}{\alpha - 1}$.

L'expression, α étant fixé, quel que soit le revenu minimum choisi, est une constante.

M. FRÉCHET voudrait enfin dire un mot de l'évaluation de l'erreur commise en ajustant la répartition des revenus. La méthode des aires indiquée par M. MOURRE ne paraît pas d'une application très aisée. Le problème est de réduire l'*ensemble* des erreurs et le moyen le plus simple est de définir une sorte d'écart entre la courbe observée et la courbe calculée et de chercher à rendre cet écart minimum par un choix convenable des coefficients (A et α pour la loi de Pareto). On est amené à un pur problème mathématique classique : rendre minimum une fonction connue des coefficients. Mais c'est la définition de l'écart qui déborde la science mathématique et qui, sans être entièrement arbitraire, l'est au moins partiellement. Dans la méthode des aires, c'est la somme des valeurs absolue des aires qui sert à repérer l'écart mais il y a toute une variété de définitions également plausibles et qui conduiraient à des valeurs différentes des coefficients (généralement assez peu différentes pour α).