

MARC BARBUT

Homme moyen ou homme extrême ? De Vilfredo Pareto (1896) à Paul Lévy (1936) en passant par Maurice Fréchet et quelques autres

Journal de la société française de statistique, tome 144, n° 1-2 (2003), p. 113-133

http://www.numdam.org/item?id=JSFS_2003__144_1-2_113_0

© Société française de statistique, 2003, tous droits réservés.

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

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

HOMME MOYEN OU HOMME EXTREME ? de Vilfredo Pareto (1896) à Paul Lévy (1936) en passant par Maurice Fréchet et quelques autres

Marc BARBUT *

RÉSUMÉ

On présente sommairement l'histoire de la découverte empirique, puis de la théorisation probabiliste, d'une famille de lois dans lesquelles l'accent est mis sur les « valeurs extrêmes » plus que sur les « valeurs moyennes » : les distributions statistiques de V. Pareto et les lois stables de P. Lévy. Diverses applications de ces lois dans les sciences sociales et celles de la Nature sont évoquées**.

ABSTRACT

The history of a family of statistical distributions and associated probability laws is briefly introduced, from their empirical discovery by V. Pareto to their theorization by P. Lévy : the non-normal stable distributions. In these laws, emphasis is put on « extreme values » rather than on « mean values ». Applications in natural sciences and in social sciences are considered**.

1. Introduction

Parmi les variables utilisées dans les applications de la statistique, il y en a qui sont par nature toujours positives : taille ou poids des conscrits, populations des agglomérations urbaines, revenus des contribuables, longueur des fleuves, surface des lacs, etc. Certaines relèvent des sciences de l'Homme, d'autres de celles de la Nature.

Du point de vue statistique, une classification pertinente est celle-ci : certaines de ces variables sont concentrées autour d'une valeur centrale (moyenne, mode ou médiane), et les grandes valeurs sont quasiment impossibles (négligeables) ; pour d'autres, par contre, les grandes (et même les très grandes) valeurs ne sont pas rares et ont des fréquences de l'ordre de quelques pourcents, allant parfois jusqu'à 10 %.

Les premières sont réputées « normales ». Mais qu'en est-il des autres ?

* École des Hautes Études en Sciences Sociales – Centre d'Analyse et de Mathématiques Sociales – 54 Boulevard Raspail, F-72570 Paris Cedex 06.

** Pour l'essentiel, cet article reprend un texte inclus dans « Pareto aujourd'hui », publié par Alban Bouvier, éditeur (Paris, P.U.F., 1999) avec son aimable autorisation.

Dès 1895-1896, Vilfredo Pareto découvre empiriquement toute une famille de distributions « autres », bien des fois redécouvertes depuis, et dans des domaines les plus divers, ceux des sciences sociales notamment, mais aussi dans les sciences naturelles. C'est une théorie élaborée en calcul des probabilités, du début des « années 20 » au milieu des « années 30 », par Paul Lévy qui donne la principale explication de cette universalité : la théorie des *lois stables* et de leurs domaines d'attraction.

2. Découverte d'un nouveau type de distribution statistique

Si, à propos du principal apport de Vilfredo Pareto à la Statistique (apport capital, on le verra, et toujours très actuel), je propose que l'on parle de son *homme extrême*, c'est bien sûr par opposition à *l'homme moyen* d'Adolphe Quetelet.

2.1. L'homme moyen : critique et réhabilitation

L'« homme moyen » d'une population c'est, selon A. Quetelet, celui dont la taille est égale à la moyenne des tailles des individus constituant cette population, le tour de poitrine est la moyenne des tours de poitrine, le poids la moyenne des poids, etc.

On sait la critique fondamentale que fit A. Cournot : l'homme moyen ainsi obtenu a toute chance de ne guère ressembler à un homme « en chair et en os », car chez un homme véritable ces diverses mesures ne sont pas indépendantes les unes des autres, et il n'y a aucune raison pour que les relations qu'elles ont entre elles soient conservées lorsque l'on prend la moyenne de chacune. Et de citer l'exemple d'une « population » de triangles rectangles : le triangle moyen calculé à la manière de Quetelet n'est pas rectangle en général.

Joseph Bertrand reprit cette critique en 1889, dans la préface à son traité de *Calcul des probabilités*; il fit notamment observer que le poids d'un individu étant *grosso modo* proportionnel à son volume, il varie comme le cube de la taille : or la moyenne des cubes d'une grandeur n'est évidemment pas le cube de la moyenne.

Bien plus tard, Maurice Fréchet, éminent mathématicien et statisticien avisé, entreprit de réhabiliter la notion introduite par Quetelet (voir Fréchet, 1955). Il fit notamment cette remarque très judicieuse : si chacune des grandeurs en jeu est peu dispersée autour de sa moyenne, comme c'est le cas pour la taille et le poids des conscrits, alors l'individu moyen « à la mode Quetelet » diffère peu de ce que pourrait être un individu réel. En reprenant l'exemple de Cournot, si chacun des côtés des triangles rectangles est très concentré autour de sa moyenne, le « triangle moyen » sera approximativement rectangle : cela se calcule aisément.

Or, la taille et le poids des conscrits sont à peu près distribués « normalement », c'est-à-dire selon la loi de Laplace et Gauss (la fameuse « courbe en cloche »), autour de leurs moyennes respectives.

Cette distribution a pour densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x-m)^2}{2\sigma^2} \right]. \quad (1)$$

Dans cette expression, m est la moyenne théorique et σ l'écart type; $f(x)$ est ainsi une fonction exponentielle du *carré de l'écart* de la variable x à sa moyenne; il en résulte une décroissance extrêmement rapide de $f(x)$ lorsque cet écart s'accroît : la courbe représentative, le « chapeau de gendarme », vient quasiment s'écraser sur l'asymptote que constitue l'axe des x . La « loi normale » convient pour des variables concentrées symétriquement autour de leur moyenne, et dont les valeurs possibles sont toutes *du même ordre de grandeur*.

Dans le cas des conscrits français de la classe 1922, 83 % des tailles sont, selon M. Fréchet (1955), comprises entre 1,57 et 1,815 m; à partir de ces données, on peut calculer que, dans l'hypothèse d'une distribution « normale » des tailles, la fréquence des statures supérieures à 2 m est de l'ordre de 10^{-4} . Quant à une taille dépassant le double de la moyenne (soit 3,40 m environ), c'est un événement quasiment impossible; et de fait, nul d'entre nous ne s'attend à rencontrer un géant d'une stature aussi monstrueuse au coin de la rue ni même dans une baraque de foire.

Ainsi, oui, Maurice Fréchet a raison : l'homme moyen a bien un sens. Mais à condition de ne considérer que celles des variables qui, comme les mesures anthropométriques, sont approximativement distribuées selon la « loi des erreurs » de Laplace et Gauss (ou éventuellement selon une autre loi statistique, mais fortement concentrée) et pour lesquelles par conséquent *les grandes valeurs ont un poids négligeable*.

Est-ce toujours le cas pour les « mesures de l'homme » ?

2.2. Des grandes valeurs qui ne sont pas rares

C'est ici qu'intervient Vilfredo Pareto. Aux alentours de 1895, installé depuis deux ans dans la chaire d'Économie politique de l'université de Lausanne, il étudie empiriquement les distributions de revenus et de richesses. La question qu'il se pose, comme d'autres économistes à la même époque, est la suivante : les revenus (ou les patrimoines) sont-ils distribués « au hasard », ou bien y a-t-il une loi « mécanique » qui expliquerait leur répartition observée ?

Or, pour Pareto comme pour la quasi-totalité de ses contemporains, une distribution « au hasard » ne peut être que celle des tirages aléatoires dans une urne de Bernoulli, et partant celle qui en dérive directement : la distribution de Laplace et Gauss, dite « loi des erreurs » en cette fin du XIX^e siècle. Si donc on observe que les revenus se répartissent différemment, on aura la preuve « *qu'il y a une cause qui produit [leur] tendance à se disposer suivant une certaine courbe* » (Pareto, 1965, p. 5 et 1896, 1964, p.312, § 960).

Tel est le raisonnement de V. Pareto. Qu'il permette de conclure à un déterminisme dans la répartition des revenus, c'est une autre question, sur laquelle on reviendra. Pareto lui-même ne le croyait guère, comme on le verra.

Pour l'instant, considérons le tout premier exemple publié par Pareto (1965, p. 2), celui de la répartition des revenus fiscaux en Angleterre pour l'année 1843. Selon ces données, le revenu moyen est de 450 livres environ ; or parmi les 106 637 contribuables, 7 923 ont plus du double (soit 900 livres) de ce revenu moyen : 7,4 % du total. Cette fréquence est loin d'être négligeable. Et il y a encore 2,6 % de contribuables ayant un revenu supérieur au quadruple de la moyenne. Nous sommes en présence d'une variable pour laquelle *les grandes valeurs ne sont pas rares*, et d'un tout autre ordre de phénomènes que ceux dont rend bien compte la « loi normale » : le revenu ne se distribue certainement pas comme la taille ou le poids des conscrits.

Cette considération sur la fréquence des grandes valeurs¹, est-ce elle qui a guidé Pareto dans sa découverte ? Je n'en sais rien ; à ma connaissance, il n'a jamais écrit ce qui l'avait mis sur la voie.

2.3. Découverte de la loi

Il est par contre tout à fait explicite sur le procédé qu'il a employé. Après quelques tâtonnements, sans doute, il a une première idée, très novatrice à son époque : c'est de considérer non l'histogramme des fréquences de revenus découpés en « tranches », comme l'auraient fait tous les statisticiens contemporains, mais la distribution *cumulée*.

À chaque niveau x_i de revenu observé, il associe donc le nombre $N(x_i)$ des titulaires de revenus supérieurs ou égaux à x_i .

Puis il eut une seconde idée, qui était bien de celles que pouvait avoir un ingénieur : on sait que Pareto exerça cette profession pendant plus de vingt ans en Toscane, dans une compagnie de chemins de fers, puis dans l'industrie du matériel ferroviaire. Cette idée d'ingénieur, ce fut de figurer les données observées en coordonnées logarithmiques (sur du « papier bilogarithmique »).

Suit le constat : les points de coordonnées $(\text{Log } x_i, \text{Log } N(x_i))$ sont à peu près alignés, sur une droite de pente négative, et ceci quel que soit le pays et la période considérés. Pareto a ainsi examiné quelques dizaines de *petites séries* – dix à vingt observations – de revenus (voir Pareto, 1896, 1964 et 1965). Ce qui conduit à une relation :

$$\text{Log } N(x) = -\alpha \text{Log } x + \text{Log } A \quad (\alpha > 0). \quad (2)$$

Soit encore :

$$N(x) = \frac{A}{x^\alpha} = N_0 \left(\frac{x_0}{x} \right)^\alpha \quad x \geq x_0 \quad (3)$$

où N_0 est le nombre total des revenus observés, et x_0 un seuil minimal pour les revenus observables.

1. Évaluer la fréquence de valeurs supérieures à quelques multiples de la moyenne pour étudier le comportement asymptotique de la distribution n'a évidemment de sens que pour des variables qui ne prennent que des valeurs positives ; c'est le cas de toutes celles que l'on envisagera ici.

Il généralisa immédiatement cette expression en ² :

$$N(x) = \frac{A}{(x+c)^\alpha} = N_0 \left(\frac{x_0+c}{x+c} \right)^\alpha, x \geq x_0 > -c \quad (4)$$

Il fallait faire cette généralisation, car une grandeur mesurable n'est d'ordinaire définie qu'à un changement d'unité près (c'est le rôle du paramètre x_0 dans l'expression (3)), mais également à un changement d'origine près, et c'est le rôle du paramètre c de l'expression (4) ³.

Quant à l'exposant α , qui est le paramètre intrinsèque de la distribution, ses valeurs estimées se situent toujours entre 1 et 2, pour les séries de revenus ou de patrimoines traitées par Pareto (de nos jours, ceci n'est plus vrai pour les revenus dans les pays occidentaux; le paramètre α y est largement supérieur à 2).

Enfin, comme les statistiques disponibles ne portent que sur les *revenus fiscaux*, et non sur la totalité des revenus, l'expression (4) et son cas particulier (3) ne sont à prendre en compte qu'à partir d'un certain seuil x_0 . Celui-ci correspond en général à la plus petite observation, mais peut éventuellement être estimé à partie des données.

Il est clair qu'une courbe d'équation (4) n'est pas du tout du même type que le « chapeau de gendarme » d'équation (1), car sa densité est, pour x grand, en $1/x^{\alpha+1}$; la « fonction puissance » croit beaucoup moins vite que l'exponentielle, et *a fortiori* que l'exponentielle du carré de x . Les *comportements asymptotiques sont très nettement différents*.

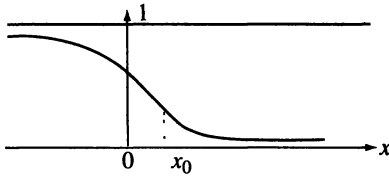
Et ceci, Pareto l'a bien vu. Lisons en effet sa conclusion (1896 (1964), p. 315, § 962) : « À première vue, la courbe de répartition des revenus ressemble à la courbe des probabilités, bien connue sous le nom de "courbe des erreurs" ... »

Cette ressemblance apparente, elle vient de ce que, au-delà d'un seuil x_0 (supérieur à la moyenne dans le cas de la distribution « normale »), l'une et l'autre courbe s'incurvent en se rapprochant indéfiniment de l'asymptote horizontale (voir figures 1 et 2).

2. Pareto proposa également une troisième formule qui ne fait qu'ajouter un paramètre supplémentaire, et qu'il s'empressa d'abandonner après ne l'avoir appliquée qu'une seule fois (voir Pareto, 1965, p. 6), à la statistique des revenus du grand-duché d'Oldenburg en 1890.

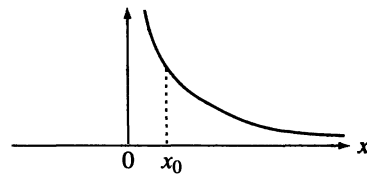
Pareto a aussi tenté, sans y avoir beaucoup réfléchi apparemment, de fournir une démonstration mathématique (voir 1896 (1964), note bas des pages 315 à 317) de ce que la distribution trouvée empiriquement, ne pouvait s'obtenir (par un « passage à la limite » convenable) à partir de la loi binomiale; malencontreusement, son calcul a probablement été bâclé; il fait quelques erreurs et aboutit à une densité en $(K/x)^{x+1/2}$ qui n'a *pas* grand-chose à voir, elle non plus, avec celle de la loi de Laplace-Gauss, pourtant établie (par un passage à la limite correctement effectué) dès 1733 par Abraham de Moivre. Comment n'a-t-il pas vu la contradiction entre son résultat « mathématique » erroné et la comparaison, justifiée celle-là, qu'il fait dans les mêmes pages entre sa courbe et la « courbe des erreurs » ? Déjà une énigme; nous en trouverons d'autres.

3. Une famille de distributions définies au choix près de l'origine et de l'unité de la variable constitue ce qu'on appelle un type de distribution. Ce terme sera fréquemment utilisé dans la suite.



La « normale » (cumulée)

FIG. 1.



La « paretienne » cumulée

FIG. 2.

Mais poursuivons la lecture :

« Il n'en est rien. Le profil qui résulterait de la loi des probabilités est *beaucoup plus creusé*⁴ que ne l'est celui de la courbe de la figure [*i.e.* la courbe représentative de la distribution de Pareto (3)]. En d'autres termes, la courbe des probabilités se rapproche des axes beaucoup plus que la courbe de la figure. »

Peut-être, comme on l'a dit plus haut, n'est-ce pas le comportement asymptotique qui a guidé Pareto dans sa découverte; ce texte nous prouve par contre que c'est bien là le critère qui lui a permis, après coup, de faire la différence.

Si d'ailleurs nous calculons maintenant, au moyen de l'estimation des paramètres de (4) fournie par Pareto pour les revenus de 1843 en Angleterre ($\alpha = 1,5$ et $c = 0$), la fréquence théorique des revenus supérieurs au double du revenu moyen, nous trouvons 6,8 %, valeur proche de celle de la fréquence empirique constatée plus haut, soit 7,4 %. De même, la fréquence théorique des revenus supérieurs au quadruple du revenu moyen est de 2,4 %, à comparer à la valeur empirique de 2,6 %.

2.4. L'homme moyen : critique sans réhabilitation

Ainsi, l'ajustement théorique de forme (3) ou (4) proposé par Pareto rend parfaitement compte de ce qui nous a servi ici de point de départ : il y a des variables pour lesquelles les grandes valeurs ne sont pas rares. Et traiter ces variables comme si elles étaient gaussiennes, c'est-à-dire en privilégiant la moyenne, ne peut conduire qu'à des absurdités.

L'une des pires, à mon sens, est celle qui est liée, vers la même époque, à la construction du QI.

À supposer que l'intelligence soit mesurable, ce que l'on peut tenir pour douteux, il est clair que nous sommes là dans l'ordre de phénomènes où « l'homme moyen » n'a guère de sens : tout enseignant sait bien que les très bonnes performances des élèves (comme aussi les très mauvaises) si elles ne sont pas très fréquentes, sont loin d'être quasi impossibles.

Or, comment a-t-on étalonné le QI? On a fait en sorte que les scores se répartissent, autour d'une moyenne fixée à 100, selon la distribution de

4. L'auteur de cet article tient à souligner cette opinion de Pareto.

Laplace-Gauss ! Dans ce choix, parfaitement arbitraire, l'influence, directe ou indirecte, mais combien néfaste en l'occurrence, des idées de Quetelet est évidente. Et ceci, à l'époque où non seulement les travaux de Pareto, mais aussi, dans un tout autre contexte (celui, notamment, des statistiques relatives au suicide) ceux du statisticien allemand von Bortkiewicz montraient que le monopole de la « loi normale » était injustifié.

C'est en 1898, en effet, que Ladislaus von Bortkiewicz publie aux Éditions Teubner de Leipzig *Das Gesetz der kleinen Zahlen (La loi des petits nombres)*. Il s'agit en fait de la distribution dite de Poisson, dont celui-ci (Siméon-Denis Poisson) avait fait la théorie plus d'un demi-siècle avant, mais qui ne deviendra populaire chez les statisticiens – pour la modélisation de questions de téléphonie et de files d'attente, notamment – qu'à partir des « années 30 ». Maurice Fréchet d'ailleurs, celui-là même qui a réhabilité l'homme moyen (mais on a vu à quelles conditions précises, pour quelles variables anthropométriques) fit remarquer à partir d'études empiriques, que parmi les nombreuses variables qui suivent approximativement la loi de Pareto, il y avait... les notes d'examen ! (voir Fréchet, 1955, page 105, dans le chapitre IV consacré aux *mathématiques dans les sciences humaines*, pages 89-117).

Quoi qu'en aient pensé A. Binet, T. Simon et leurs successeurs, l'intelligence ne relève pas de l'homme moyen, mais bien plutôt de *l'homme extrême*. Quand les éducateurs et pédagogues de toute obédience finiront-ils par comprendre cela et à en tirer les conséquences, notamment pour la docimologie et les instructions officielles sur la notation des examens ? Ce qui était excusable au moment où les outils d'analyse statistique dont on parle ici commençaient à être élaborés, est-ce encore admissible un siècle plus tard ?

Notons enfin que, dans son article de réhabilitation de l'« homme moyen », Fréchet (1955) propose de le remplacer par un « homme typique », plus général et de définition moins contraignante. Par exemple, l'« homme médian », qui ne repose que sur les propriétés ordinales, serait sûrement mieux adapté pour les questions de docimologie.

3. La comparaison des inégalités

La question la plus brûlante, socialement et politiquement, soulevée par la répartition des revenus, des richesses et des patrimoines, c'est celle des inégalités. Pareto, évidemment, s'y est beaucoup intéressé, et a cherché à voir comment sa distribution pouvait en rendre compte. Et, disons le d'emblée, il s'est, sur ce point, totalement fourvoyé.

3.1. Où Pareto se trompe

On remarquera d'abord qu'une distribution théorique de type (3) ou (4) *n'a pas de variance* lorsque l'exposant α est inférieur ou égal à 2, ce qui était le cas pour les valeurs estimées empiriquement par Pareto⁵. Plus précisément,

5. D'une façon générale, le moment d'ordre k n'existe que pour $k < \alpha$ (inégalité stricte).

la variance théorique serait infinie : effet du poids des grandes valeurs, une fois de plus.

Par contre, lorsque α dépasse 2, la variance a une valeur finie bien déterminée qui se calcule facilement et qui est une fonction décroissante de α . Celle-ci tend vers zéro lorsque α augmente indéfiniment.

Cette simple remarque suffit pour voir que la dispersion varie en sens inverse de α : plus α est grand, plus la distribution est concentrée. Et si α augmente indéfiniment, une distribution de type parétien (4) a pour limite celle (de type « Dirac ») pour laquelle tous les revenus seraient égaux au seuil x_0 : c'est l'équipartition, la distribution la plus égalitaire possible.

On peut donc inférer que l'inégalité (une définition précise et opératoire pour la Statistique de cette notion ne viendra que quelque vingt-cinq ans plus tard) varie en sens inverse de α .

Eh bien, cela, non seulement Pareto ne l'a pas vu, mais il semble même avoir toujours été persuadé du contraire. Pour lui, plus α est grand, plus il y a d'inégalité ; il le croit à tel point que, lorsque Georges Sorel, dès juillet 1897, lui fait remarquer son erreur (voir Sorel, 1897) avec l'argument de l'augmentation de la concentration justement, il admet seulement, dans sa réponse (1965, p. 43 et sq.) que le concept d'inégalité est ambigu. Il y a là pour moi l'un des mystères (il y en aura d'autres) de Pareto : sa bonne pratique des mathématiques – en témoignent des écrits tels que Pareto (1911), *Quelques exemples d'application des méthodes d'interpolation à la statistique*, dans Pareto (1989), pages 76- 88, ou encore *Sur les fonctions génératrices d'Abel*, dans Pareto (1989), pages 31-64 – aurait dû lui éviter de se tromper.

La question sera progressivement élucidée dans les décennies suivantes.

En 1905, l'économiste américain M.O. Lorenz introduit les *courbes de concentration*, qui permettent une comparaison intuitive de l'inégalité des distributions. L'ordre de comparaison ainsi obtenu est d'ailleurs *partiel*. Dans le cas des distributions de Pareto, l'équation de ces courbes, facile à obtenir, montre à l'évidence que l'inégalité varie en sens inverse de α , et qu'elle est d'autant plus grande que α est proche de 1⁶.

Quelques années plus tard, C. Gini définit au moyen des courbes de concentration son indice d'inégalité, qui vaut, pour les distributions parétiennes du premier type (3) :

$$G = \frac{\beta - 1}{\beta + 1} \quad \left(\beta = \frac{\alpha}{\alpha - 1} \right) \quad (5)$$

c'est bien une fonction croissante de β , et donc décroissante de α . Gini le fait remarquer, avec éclat sans doute, selon son habitude.

Pour autant que je sache, Pareto n'est jamais revenu sur son erreur. Pire, lorsqu'il publie en 1906 le *Manuel*, de même que dans l'édition française de 1909 – voir Pareto (1966), chapitre 7 –, il reprend (p. 389-393) l'argumentation

6. Lorsque $0 < \alpha \leq 1$, les distributions parétiennes non seulement n'ont pas de variance, mais elles n'ont même plus de moyenne et la courbe de concentration n'existe plus. C'est un cas – rarement rencontré dans les sciences sociales, et jamais dans les travaux de Pareto – que l'on ne traitera pas ici.

de 1896-1897, et invoque les « démonstrations mathématiques » *du Cours* (1964).

À sa décharge, on rappellera que c'est seulement au tout début des années 1920 que l'économiste britannique A. C. Pigou énonce le principe (nous dirions aujourd'hui l'axiome) qui fonde la notion *statistique* d'inégalité : transférer un franc d'un revenu à un revenu inférieur diminue l'inégalité de la répartition. C'est d'ailleurs ce principe simple (dit *principe des transferts* de Pigou et Dalton) qui légitime, avec quelques autres axiomes assez naturels, l'usage des courbes de concentration de Lorenz et Gini pour la comparaison des inégalités.

3.2. Moyennes conditionnelles et inégalités

C'est de Maurice Fréchet que viendra, en 1925, la remarque qui tranchera définitivement la question. Ici, on ne peut plus rien reprocher à Pareto : il est mort depuis deux ans.

Ce que remarque Fréchet, c'est ceci : pour chaque niveau de revenu x , considérons la *valeur moyenne* $m(x)$ *des revenus supérieurs* à x ; $m(x)$ se calcule, *mutatis mutandis*, comme l'espérance de vie à l'âge x des démographes et des actuaires ; on peut l'appeler *moyenne conditionnelle en x* . Il est clair que plus $m(x)$ est grand par rapport à x , plus l'inégalité peut paraître forte. Pour une distribution de Pareto de type (3), cette moyenne « théorique » vaut (il faut $\alpha > 1$) :

$$m(x) = \beta x \quad \text{où } \beta = \frac{\alpha}{\alpha - 1}. \quad (6)$$

La relation entre α et β est d'ailleurs symétrique et peut s'écrire :

$$\frac{1}{\alpha} + \frac{1}{\beta} = 1$$

L'équation (6) signifie que, si β est par exemple égal à 3 (c'est-à-dire $\alpha = 1,5$), ce revenu moyen est le triple de x , et ceci quel que soit le niveau x ; si β ne vaut que 2, il n'est plus que le double : l'inégalité (et le sentiment qu' en peut avoir le titulaire du revenu x) a diminué. Ainsi, l'inégalité « au sens de Fréchet » varie bien dans le même sens que β et donc en sens inverse de α . En particulier, si β tend vers 1 (c'est-à-dire si α tend vers l'infini), $m(x)$ diffère peu de x ; et pour $\beta = 1$ (c'est-à-dire $\alpha = \infty$), $m(x)$ est confondu avec x : c'est l'équirépartition.

La remarque de M. Fréchet s'étend, par un calcul « taupinal » très élémentaire, aux distributions de Pareto du type général (4). La moyenne conditionnelle est alors :

$$m(x) = \beta x + h = \beta x + c(\beta - 1). \quad (7)$$

Pour les distributions parétiennes, nous avons ainsi une représentation *fonctionnelle* de l'inégalité, analogue (mais non équivalente) à la représentation par les courbes de Lorenz. Ici, les « courbes » sont des droites, et la fonction est la plus simple qui soit : la fonction *linéaire* (ou affine).

Graphiquement, pour deux distributions à comparer, tant que la droite d'équation (7) de la première est au-dessus de celle de la seconde, c'est que la première est la plus inégalitaire; lorsque les deux droites se coupent, les moyennes conditionnelles ne permettent pas la comparaison. Comme dans le cas des courbes de Lorenz, l'ordre de comparaison est partiel.

On remarquera en outre que, pour les *grandes valeurs* (toujours elles) de x , la constante h de (7) est négligeable; l'ordre de grandeur de β suffit pour indiquer le degré d'inégalité. Pour l'ensemble de la distribution, par contre, il faut tenir compte des deux paramètres, et l'interprétation est plus délicate (voir Barbut, 1989 et 1998). À titre indicatif, on notera que le coefficient de Gini peut s'écrire, dans les mêmes notations que supra :

$$G = \frac{\beta - 1}{\beta + 1} \left(1 + \frac{c}{m_0} \right) \quad (8)$$

où $m_0 = m(x_0)$ est la moyenne de l'ensemble de la distribution.

En ce qui concerne les revenus fiscaux dans les pays européens à la fin du XIX^e siècle, les données publiées par V. Pareto (1896 et 1965) donnent des valeurs de α aux alentours de 1,5, ce qui correspond à $\beta = 3$: il y avait une forte inégalité.

Cent ans plus tard (et même, en fait, dès le milieu du XX^e siècle) ces valeurs sont inversées; c'est β qui est aux alentours de 1,5, parfois même encore plus proche de 1, et α qui vaut 3 ou plus : l'inégalité de la distribution des *revenus fiscaux* a très fortement diminué. Il faut y voir, sans nul doute, en partie l'effet des transferts opérés par la redistribution (impôts et cotisations sociales); le *principe des transferts* de Pigou et Dalton, évoqué ci-dessus, a bien eu sa conséquence logique.

Les valeurs élevées, toujours supérieures à 2, de l'exposant α pour les statistiques actuelles montrent que l'on est maintenant dans le cas de distributions ayant une variance, et devenues plus « normales ». En fait, c'est la distribution dite Log-Normale de R. Gibrat (dans laquelle le logarithme du revenu suit une loi de Laplace-Gauss) qui s'ajuste le mieux aux données (voir Armatte, 1995 et Gibrat, 1931). Cette remarque vaut pour les distributions de revenus, mais pas pour celles d'autres variables, dont on parlera plus loin, qui continuent à être parétiennes (α y est inférieur à 2).

3.3. Ajustements parétiens

La linéarité des moyennes conditionnelles dans le cas des distributions parétiennes, il est tout à fait surprenant – voici le deuxième « mystère Pareto » – que Pareto lui-même ne l'ait pas vue. Et pourtant, la formule (7) figure en toutes lettres dans son *Cours* – 1896 (1964), p. 313, formule (4 bis) et p. 321, formule (3) –, et le calcul qui conduit à l'expression simplifiée (6) *supra* se trouve déjà dans l'une de ses toutes premières publications sur ces questions (1965, p. 9-10). Seulement, voilà : Pareto avait essentiellement en vue, en faisant ce calcul, la détermination de l'effet de certains types de prélèvements (impôt proportionnel ou impôt progressif, par exemple) sur la distribution

des revenus; aussi n'a-t-il calculé les expressions (6) et (7) que pour une seule valeur de x , le seuil minimal x_0 , des revenus. Et il n'a pas pensé à faire varier ce seuil, ce qui l'aurait nécessairement conduit aux moyennes conditionnelles. Cet oubli est très regrettable pour deux raisons.

La première est que cela lui aurait évité de « vasouiller » comme on l'a vu le faire quant au sens de variation de l'inégalité en fonction de la variation des paramètres.

La seconde, et la plus importante à mes yeux, est que cela lui aurait fourni une méthode simple et efficace d'*estimation des paramètres* de la distribution; nous touchons là au troisième « mystère Pareto ».

La seule façon d'estimer les paramètres de la distribution qu'il indique dans ses publications est celle qui utilise la transformation en logarithmes; à savoir, dans le cas particulier des distributions de type (3), la relation (2) :

$$\text{Log } N(x) = -\alpha \text{Log } x + \text{Log } A.$$

On fait un ajustement linéaire sur le nuage de points ($\text{Log } x_i, \text{Log } N(x_i)$); α est estimé par la pente de la droite d'ajustement, $\text{Log } A$ – et donc A – par son ordonnée à l'origine.

Fort bien, mais cela a *deux inconvénients majeurs*.

Le premier est que faire un ajustement sur les données elles-mêmes (toutes les méthodes consistent à minimiser une fonction croissante des *écarts* entre observations et valeurs *théoriques* correspondantes), ou faire cette opération sur les *écarts des logarithmes* :

« *Ce n'est pas du tout la même chose.* » Ici, c'est Pareto lui-même qui parle (*Quelques exemples d'application des méthodes d'interpolation à la statistique*, p. 80 et *Tables pour faciliter l'application de la méthode des moindres carrés*, p. 92, dans *Statistique et économie mathématique*, 1989). Et il souligne bien qu'ajuster sur les logarithmes revient à minimiser une fonction des *écarts relatifs*, et donc, notamment, à fortement sous-évaluer le poids des grandes valeurs. Or, c'est précisément de bien prendre en compte celles-ci que les distributions parétiennes tirent leur principale efficacité.

Cette contradiction a toujours tourmenté Pareto, les textes que l'on vient d'évoquer en font foi. Mais qui, de nos jours, les a lus ?

Quant au *second inconvénient*, c'est que, dans le cas général (4), le « passage aux logarithmes » donne :

$$\text{Log } N(x) = -\alpha \text{Log } (x + c) + \text{Log } A. \quad (9)$$

Comme c , qui est l'un des paramètres à estimer, est inconnu, on ne peut déterminer les $(x_i + c)$ entrant sous le logarithme du second membre de (9); il devient donc impossible de procéder ainsi à l'ajustement linéaire.

Pareto fournit souvent des estimations de ce paramètre c pour les statistiques de revenus ou de patrimoines publiées dans (Pareto, 1896) ou dans (Pareto, 1965). Comment y parvenait-il? Mystère. Probablement par tâtonnements.

Il est également possible (c'est une hypothèse qu'il faudrait vérifier) qu'il n'utilisait pour réaliser l'ajustement que les points du graphique bilogarithmique correspondant aux x_i assez grands pour que la différence entre x_i et $(x_i + c)$ soit négligeable. Il n'a en tout cas jamais indiqué de méthode.

Or l'expression (7) de la moyenne conditionnelle donne une solution évidente à ce problème d'estimation. Il suffit, à chaque niveau x_i observé de revenu, d'associer la moyenne empirique mi des revenus supérieurs ou égaux à x_i (moyenne que les statistiques publiées par l'administration compétente permettent en général de calculer), et de procéder à l'ajustement linéaire :

$$m_i = \beta x_i + h + \varepsilon_i \quad (\text{où } \varepsilon_i \text{ est le « résidu »}). \quad (10)$$

De l'estimation de la pente β et de la constante h , on déduit immédiatement celle de c (puisque l'on a $h = c(\beta - 1)$), et celle de $\alpha = \frac{\beta}{\beta - 1}$.

Les ajustements réalisés par cette méthode sont toujours très bons (lorsque la distribution empirique est approximativement parétienne, bien sûr), et usuellement meilleurs que ceux obtenus par « passage aux logarithmes » (voir Barbut, 1998).

Dans les cas, peu fréquents, où l'exposant α de Pareto est inférieur à 1 ($0 < \alpha \leq 1$), ni la moyenne de la distribution, ni les moyennes conditionnelles n'ont plus de sens; de même, d'ailleurs, que les courbes de concentration de Lorenz et Gini, dont le calcul théorique suppose l'existence de ces moyennes.

La méthode d'estimation par ajustement linéaire préconisée ici reste néanmoins applicable, à condition de remplacer la moyenne conditionnelle par la *médiane* conditionnelle $\mu(x)$: le revenu médian des revenus supérieurs à x . Pour les distributions parétiennes, $\mu(x)$ est en effet fonction linéaire de x et l'on a, toujours dans les mêmes notations :

$$\mu(x) = 2^{1/\alpha} x + c(2^{1/\alpha} - 1). \quad (11)$$

4. Une loi universelle

4.1. Postérité et universalité

Dans les décennies qui suivirent la découverte par Pareto de sa loi de distribution des revenus, celle-ci fut retrouvée – dans l'ignorance, le plus souvent, des travaux de Pareto – dans d'autres domaines, bien éloignés de l'Économie.

Ce fut d'abord J.-B. Estoup, auteur d'un manuel de sténographie (1916) qui observe (dès 1907, peut-être) que, dans des textes assez longs écrits en français, la fréquence d'apparition de chaque mot de la langue est grossièrement proportionnelle à l'inverse du rang, lorsqu'on range les mots du texte par fréquence décroissante (voir Petruszewycz, 1973).

C'est la loi appelée «rang-fréquence» par les lexicologues ; elle peut s'écrire :

$$f_r = \frac{K}{r} \quad (12)$$

où f_r est la fréquence du mot de rang r .

La fréquence d'un mot peut être assimilée à la «richesse» de celui-ci dans le texte, et son rang est le nombre des mots qui sont plus «riches» que lui. En posant $f_r = x$ et $r = N(x)$, la formule (12) se réécrit donc :

$$r = \frac{K}{f_r} \text{ c'est-à-dire } N(x) = \frac{K}{x}. \quad (13)$$

C'est une loi de Pareto d'exposant α égal à 1.

La formulation de la loi «rang-fréquence» fut améliorée ultérieurement par G. K. Zipf (aussi est-elle souvent appelée chez les linguistes «loi de Zipf») qui retrouve cette loi dès 1929 et la formule – sous la forme (12) – dans son célèbre ouvrage publié en 1949. C'est surtout Benoît Mandelbrot (1955 et 1966) qui, dans les années 1950, montre le lien avec la loi de Pareto, et remplace (12) par l'expression plus générale et mieux ajustée aux données :

$$f_r = \frac{K}{(r + \rho)^B} \quad B > 0, \rho \text{ est une constante.} \quad (14)$$

Plus conforme à l'esprit parétien est la relation :

$$r = N(x) = \frac{K}{(f_r + c)^\alpha} = \frac{K}{(x + c)^\alpha}, \quad \alpha = \frac{1}{B} > 0. \quad (15)$$

Ces expressions de la répartition statistique des mots de textes assez longs semblent vérifiées dans la plupart des langues – sur toute cette question, on peut lire une synthèse dans Mandelbrot (1955) et dans Petruszewycz (1973).

En 1913, c'est le géographe allemand F. Auerbach qui, rangeant les villes de son pays par taille (population) décroissante, découvre la loi «rang-taille» ; il l'exprime sous la forme (12) ou (13), où f_r , doit être interprété comme la taille de la ville de rang r .

Actuellement, tous les spécialistes de Géographie urbaine ou de Démographie utilisent cette loi, sous les diverses formes (12), (13) ou (15) voir par exemple Pumain (1982) ou Barbut (1998).

On peut encore citer, entre autres exemples de répartition parétienne, celle des propriétés agricoles en fonction de leur surface, en France notamment, ou les sociétés cotées en bourse selon leur capitalisation.

De tous ces exemples, on pourrait être tenté de conclure qu'il s'agit d'un type de distribution spécifique des phénomènes sociaux ou humains. Pareto semble l'avoir pensé : «*La tendance des revenus à se grouper suivant une certaine loi pourrait bien dépendre, en partie, de la nature même des hommes*» (1965, p. 7).

Il n'en est rien : Maurice Fréchet (1941) a montré que la longueur des fleuves, par exemple, la surface des lacs ou celle des îles ont également une distribution approximativement et asymptotiquement (voir définition précise plus loin) parétienne.

Actuellement, ce type de distribution a de multiples applications; outre les domaines déjà mentionnés, citons les Mathématiques financières ou, en Sciences de la matière, la « physique du tas de sable » et la granulométrie.

4.2. Paul Lévy, les lois limites et les lois stables

Nous sommes donc en présence d'une loi universelle, aussi universelle au moins que ne l'est la loi de Laplace et Gauss. Cette universalité doit avoir une explication.

La plus forte, et la plus simple à la fois, des explications possibles, c'est dans des travaux s'échelonnant de 1922 à 1936 qu'elle est donnée par le très grand mathématicien probabiliste que fut Paul Lévy.

D'où le privilège accordé à la « loi normale » vient-il ? Essentiellement de deux de ses propriétés, d'ailleurs liées entre elles et qui ont une longue histoire; elle commence en 1689 avec la « loi des grands nombres » de J. Bernoulli, et s'achève aux alentours de 1930 avec P. Lévy, A. Khintchine, H. Cramer et quelques autres.

La première de ces propriétés, c'est le *théorème central limite* qui, dans une formulation simplifiée due à P. S. Laplace en 1812, dit ceci : d'une variable aléatoire numérique X ayant une distribution théorique quelconque, de moyenne m et de variance finie σ^2 , tirons au hasard, les tirages étant indépendants les uns des autres, n réalisations X_i (un échantillon aléatoire de taille n); alors la variable :

$$Y_n = \frac{\sum_{i=1}^n (X_i - m)}{\sigma\sqrt{n}} \quad (16)$$

obéit approximativement à la loi de Laplace-Gauss centrée réduite (c'est-à-dire de moyenne nulle et de variance égale à 1). L'approximation est d'autant meilleure que n est grand; en pratique, elle est excellente pour des échantillons de quelques dizaines d'unités.

Ainsi, la loi de Laplace-Gauss centrée réduite est la « loi limite », qui a pu être tabulée une fois pour toutes, des « moyennes » de variables aléatoires indépendantes calculées selon l'expression (16) : on remarquera que le passage des X_i à Y_n nécessite un changement d'origine sur chaque X_i , et un *changement d'unité* proportionnel non à la taille n de l'échantillon, mais à sa racine carrée.

La seconde propriété des distributions de Laplace-Gauss qui en explique l'universalité, c'est que toute somme de variables aléatoires indépendantes (*v.a.i.*) gaussiennes est encore une variable gaussienne : les lois de Laplace-Gauss sont de type *stable par rapport à l'addition des v.a.i.*

Il est clair que la loi limite d'une variable Y_n telle que celle de la formule (16) doit posséder cette seconde propriété : si nous « moyennons » selon le procédé (16) des *v.a.i.* appartenant à la loi limite, nous devons retomber sur cette même loi, par définition même du mot limite.

Mais l'importance des lois stables par rapport à l'addition tient aussi à la façon dont nous « modélisons » les phénomènes observables. Si une variable Z est pensée comme fonction d'autres variables U_1, U_2, \dots, U_k indépendantes entre elles mais relevant d'un même type de distribution symétrique, le premier « modèle » qui vient à l'esprit, le plus simple des modèles possibles, c'est le *modèle linéaire* dans lequel la variable « expliquée » Z est en première approximation combinaison linéaire positive des variables « explicatives » U_1, U_2, U_k . Si celles-ci suivent des lois de type stable par rapport à l'addition, il en sera de même de Z : variable « expliquée » et variables « explicatives » pourront ainsi être considérées comme relevant d'une même catégorie de phénomènes.

La question qui se pose alors, et que Paul Lévy achève de résoudre en 1936, est la suivante : quelles sont *toutes* les distributions stables par rapport à l'addition, et pouvant donc être approximativement celles de *v.a.i.* convenablement « moyennées » ?

Et voici la réponse : outre la loi de Laplace et Gauss, il y a une famille de lois qui toutes sont *asymptotiquement parétiennes* (quand x augmente indéfiniment, la fonction de répartition tend vers zéro comme $1/x^\alpha$ pour un exposant α tel que $0 < \alpha < 2$).

Elles peuvent s'obtenir comme lois limites (voir Lévy, 1937, p. 203) de *v.a.i.* parétiennes X_i d'exposant α « moyennées » selon le procédé :

$$Y_n = \frac{\sum_{i=1}^n (X_i - m)}{n^{1/\alpha}} \quad \text{pour } 1 < \alpha < 2 \quad (17)$$

$$Y_n = \frac{\sum_{i=1}^n X_i}{n^{1/\alpha}} \quad \text{pour } 0 < \alpha \leq 1. \quad (18)$$

La différence entre (17) et (18) provient de ce que, si $\alpha \leq 1$, la distribution de Pareto parente n'a, comme on l'a vu, même plus de moyenne m .

On remarquera que l'expression (16) du cas Gaussien correspond à $\alpha = 2$ dans (17) ; et que le dénominateur $n^{1/\alpha}$ de (17) et (18) tend vers l'infini plus rapidement que \sqrt{n} lorsque n augmente indéfiniment ; pour $\alpha = 1$, il vaut n ; pour $\alpha = 1/2$ ce diviseur est n^2 .

La famille des distributions de type stable par rapport à l'addition se trouve ainsi entièrement définie. Chacune est déterminée, aux changements d'origine et d'unité près, par la valeur du paramètre α ($0 < \alpha \leq 2$) :

– pour $\alpha = 2$, c'est la distribution « normale » de Laplace-Gauss ;

– pour $\alpha = 1$, c'est la distribution dite de Cauchy, dont l'expression mathématique importe peu ici⁷ ;

– pour chaque autre valeur de α , il n'y a pas de « formule » donnant l'expression de la fonction de répartition $F(x)$; celle-ci est cependant bien définie par sa transformée de Fourier :

$$\varphi(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x). \quad (19)$$

Pour les lois stables de P. Levy, $\varphi(t)$ a l'expression explicite :

$$\varphi(t) = \exp. \left[-a + ib \frac{|t|}{t} \right] |t|^\alpha \quad (20)$$

avec : $0 < \alpha \leq 2$, $a > 0$, $\left| b \cos \frac{\pi}{2} \alpha \right| \leq a \sin \frac{\pi}{2} \alpha$.

Le paramètre a prend en compte les changements d'unité, et b une éventuelle dissymétrie. Retenons finalement que :

– la loi stable de paramètre α peut être engendrée à partir de la loi de Pareto d'exposant α , en « moyennant » les variables paréliennes selon les expressions (17) ou (18), et en passant à la limite ;

– cette loi stable est elle-même *asymptotiquement parélienne* : pour les grandes valeurs de la variable, elle se comporte comme une distribution de Pareto en $1/x^\alpha$.

Pour ces raisons, si l'on tient à leur donner un patronyme, il convient, comme l'a proposé Mandelbrot, de les appeler *distributions de Pareto-Lévy* : Pareto en a découvert empiriquement le premier exemple statistique, Lévy en a trouvé les principales propriétés mathématiques.

Pour un exposé plus détaillé des principales étapes de l'histoire, qui s'étend sur deux siècles et demi, évoquée dans ce paragraphe, on peut consulter Barbut (1991).

4.3. Où l'on retrouve l'homme extrême

Ces propriétés signifient que les distributions *approximativement* et *asymptotiquement* paréliennes sont au coeur du Calcul des probabilités et de la Statistique probabiliste. Mais une propriété supplémentaire, démontrée par Maurice Fréchet (encore lui !) illustre bien le lien entre distribution de Pareto, phénomènes aléatoires, et grandes valeurs de la variable : d'une variable aléatoire X , tirons un échantillon (tirages indépendants) de taille n ; et considérons non plus une « moyenne » des réalisations X_i , mais la plus grande

7. Notons cependant que sa densité dessine une très belle « courbe en cloche », bien symétrique, mais pour laquelle les grandes valeurs ne sont pas négligeables : elle n'a ni variance, ni moyenne, et est asymptotiquement en $1/x^2$. Notons surtout que, par cet exemple entre autres, A. L. Cauchy, vers 1853, voulait déjà montrer qu'il y avait d'autres « courbes des erreurs » possibles que celle de Laplace et Gauss. Laplace lui-même avait d'ailleurs utilisé, à la fin du XVIII^e siècle, une première loi, dont la densité est une exponentielle simple de la valeur absolue de l'écart, et non de son carré.

d'entre elles, leur maximum :

$$Z_n = \max_i X_i \quad (21)$$

Au milieu des « années 20 », Fréchet recherche quelles peuvent être les lois limites de Z_n , et par la suite quelles sont les distributions *stables pour le maximum*, c'est-à-dire telles que Z_n ait une distribution de même type que celle des X_i . Et il montre que, sous certaines conditions, il s'agit encore, approximativement et pour les grandes valeurs de la variable, des distributions de Pareto⁸.

Stabilité pour le maximum : l'homme de Pareto est bien l'homme des valeurs extrêmes.

4.4. Distributions paréliennes et causalité

En présence d'une variable dont la distribution empirique est à peu près parélienne, certains ont cru pouvoir affirmer que cette variable n'était pas aléatoire et que le phénomène en jeu était régi par une loi de nature déterministe ; c'était encore écrit naguère dans quelques manuels d'Économie. Or ce que disent les théorèmes de P. Lévy et M. Fréchet, c'est exactement le contraire ; l'explication la plus simple et la plus naturelle du comportement parélien d'une variable observée, c'est qu'il s'agit d'un phénomène purement aléatoire, dû au seul hasard ; autrement dit, c'est que nous sommes dans l'ignorance des causes qui auraient pu le produire. Telle est la conclusion qui s'impose.

De cette méprise sur la nature éventuellement aléatoire des variables paréliennes, Pareto lui-même est largement responsable. On y reviendra.

Pour l'instant, ajoutons que si l'explication par le hasard (c'est-à-dire l'absence d'explication) est la plus immédiate, elle n'est pas la seule.

Il existe en effet des « modèles », certains très simples, purement algorithmiques, pour engendrer des distributions de Pareto. En voici un exemple classique.

Pour partager un gâteau entre individus d'une population nombreuse, procédons par étapes de la façon suivante : on commence par attribuer à une proportion p , fixée une fois pour toutes, des individus une proportion q (également fixée, $q < p$) du gâteau ; puis, deuxième étape, à la proportion p de ceux qui n'ont pas encore été servis, on affecte la proportion q de ce qui reste du gâteau ; et ainsi de suite... Un calcul élémentaire montre que la répartition finale du gâteau entre les individus est parélienne.

Il existe aussi des modèles « markoviens », dans lesquels le processus de répartition est aléatoire, mais avec des probabilités qui, à chaque étape, dépendent de ce qui s'est passé aux étapes précédentes. On pourra voir, à ce sujet, les écrits de Fréchet (*Le calcul des probabilités dans les sciences sociales*,

8. Dans d'autres conditions que celles formulées par Fréchet, on peut être conduit, pour la loi limite du maximum, à l'une ou l'autre de deux lois de type différent, mais apparentées à celle de Pareto (voir Barbut, 1998).

dans *Les Mathématiques et le concret*, 1955, p. 250-273, ainsi que l'article de 1945), de Mandelbrot (1960) et de Petruszewycz (1963).

Tout ceci signifie qu'il y a une grande diversité d'explications mathématiques possibles, allant du déterminisme pur à l'aléatoire pur, lorsque l'on observe empiriquement une distribution de Pareto. Et ce n'est pas au mathématicien ni au statisticien qu'il revient d'en proposer une; seul le spécialiste (qu'il soit sociologue, démographe ou économiste) du phénomène étudié est compétent pour le faire; en collaboration avec le statisticien, le cas échéant. C'est ainsi que devrait fonctionner une interdisciplinarité bien comprise.

4.5. Et Pareto lui-même, qu'en pensait-il ?

Mathématicien par sa formation, statisticien, mais aussi et surtout économiste et sociologue, il réalisait dans sa seule personne l'interdisciplinarité souhaitable.

À ce point, il faut, je crois, distinguer deux Pareto.

Il y a d'une part l'économiste ardemment libéral, qui veut prouver – ce fut, on l'a vu, son point de départ – que la distribution des revenus obéissant à une loi quasiment mécanique, les mesures de redistribution prônées par les socialistes seront sans effet sur leur répartition. D'où l'interprétation, erronée à mon sens – mais à laquelle cette thèse invite –, du fameux titre du § 962 du *Cours* : « *La répartition des revenus n'est pas l'effet du hasard* ».

Que Pareto, par cette phrase (qui revient à plusieurs reprises dans ses écrits et qui est la seule que trop de gens ont retenue), ait notamment voulu plaider qu'il y avait une causalité déterministe à la répartition des revenus, c'est particulièrement net dans l'étonnant parallèle (voir Pareto, 1965, p.23, dans sa réponse à des critiques de l'économiste britannique F.Y. Edgeworth) qu'il établit, sous la forme d'un tableau synoptique à deux colonnes, entre les lois du mouvement des planètes et sa loi de distribution des richesses : Pareto soi-même y est successivement le Kepler qui découvre empiriquement la loi, le Newton qui en fait la théorie et la mathématise et l'intemporel épistémologue qui en souligne, à juste titre, le caractère approximatif; comme c'est le cas de toute expression mathématique d'une loi de la nature, d'ailleurs. Argumentation que je crois largement polémique et qui, si j'ose dire, « pousse la métaphore » un peu loin.

Mais à côté du militant de l'économie libérale, il y a aussi le Pareto statisticien, et d'esprit « scientifique ». Et celui-ci, qu'écrivait-il ?

« La courbe que nous avons obtenue pour la répartition des revenus n'est probablement qu'un cas particulier de la courbe bien connue qui indique la fréquence des écarts. On donne à cette courbe le nom de courbe des erreurs. On la retrouve dans les circonstances les plus variées : dans le tir à la cible, dans les mesures d'une certaine quantité, dans l'observation des statures des conscrits, etc. Quand les écarts sont également possibles en plus ou en moins, la courbe prend la forme d'un chapeau de gendarme. Pour les revenus, les écarts ne sont plus également faciles dans les deux sens, et c'est probablement à cette circonstance qu'est due la forme singulière de la courbe de répartition

des revenus» (art. dans *Le monde économique* du 25 juillet 1896, reproduit dans l'ouvrage publié en 1965, p. 16-18).

C'est clair, et la même idée revient d'autres fois sous sa plume : pour le Pareto qui écrit ceci, sa loi, loin d'être déterministe, n'est qu'un avatar de la « loi des erreurs » de Laplace-Gauss. La distribution des revenus est bien assimilable à celle d'une variable aléatoire. Et ce qui *n'est pas l'effet du hasard*, c'est la *forme de la distribution*, et non les valeurs prises par la variable répartie.

Voici du reste, à ce sujet, son dernier mot. Dans une addition écrite en annexe du *Cours* ([1896 (1964)], p. 416 et sq., § 962), mais qui n'a guère été lue, Pareto essaye de construire un « modèle mathématique » engendrant sa distribution. Il commence ainsi :

« Nous avons vu que la répartition des revenus ne dépendait pas *seulement* du hasard ; demandons nous, maintenant, quelle loi d'*hétérogénéité* des individus suppose la *forme* que nous avons *découverte* pour la courbe des revenus. »

« Nous avons une urne contenant des boules blanches et des boules noires. On extrait μ boules, et l'on répète un grand nombre de fois... »

Mais cette urne de Bernoulli ne saurait convenir, compte tenu de la répartition observée. Donc : « ... il faut exclure l'hypothèse d'une seule urne, et admettre qu'il y en a plusieurs. »

Suit la formulation, discursive et mathématique, du modèle. Celui-ci est probabiliste : on tire non dans une seule urne, mais dans une population d'urnes dont Pareto fait varier l'effectif en fonction de la composition (la proportion p des boules blanches, celles qui font gagner 1 franc chaque fois qu'on en tire une) de façon à obtenir, pour les revenus ainsi tirés au hasard, la répartition voulue. En fait, il suppose p inversement proportionnel à une puissance positive de l'effectif g de l'urne : plus celle-ci a une population nombreuse et plus faible est la probabilité de tirer une « bonne boule ». Dans son esprit et dans son texte, chaque urne correspond à une strate de sa « pyramide sociale » ; c'est sa façon de prendre en compte l'« hétérogénéité des individus »⁹.

La formulation est maladroite sans doute, encore trop rudimentaire peut-être, et entachée d'une pétition de principe que l'on peut estimer assez naïve.

Un demi-siècle environ plus tard, M. Fréchet (toujours lui!) proposera des modèles stochastiques plus convaincants et correctement traités ; la composition de l'urne s'y modifie à chaque tirage en fonction des résultats des tirages antérieurs (voir *Le calcul des probabilités dans les sciences sociales*, dans *Les Mathématiques et le concret*, p. 250- 273, ou encore l'article de 1945).

Il n'en reste pas moins que le raisonnement et les calculs de Pareto sont probabilistes, et c'est là l'essentiel quant à ce que fut le fond de sa pensée.

9. Sur l'importance qu'eut la loi de distribution des revenus dans l'élaboration de la sociologie parétienne, lire par exemple Bernard Valade [1990].

Ainsi commence en 1896, dans l'histoire des idées de la Statistique et du Calcul des probabilités une grande aventure qui traversera le siècle en passant par les modèles de M. Fréchet et par les fractals de B. Mandelbrot, et aura culminé avec P. Lévy et les lois stables. Pareto lui-même n'en connaîtra presque rien : il ne semble plus guère s'intéresser à ces questions dans la dernière partie de sa vie ; et surtout, il est mort (en 1923) quand les événements décisifs se produisent.

Si l'on s'en tient à ce qu'il a vraiment écrit, il est probable qu'il en a pressenti quelque chose. Pour aller plus loin qu'il ne l'a fait, il aurait fallu qu'il en ait le goût ; mais aussi qu'il dispose des bons outils – l'inversion de la transformée de Fourier ne sera définitivement mise au point que vers 1925 par P. Lévy – et surtout d'un talent (voire d'un génie) mathématique comparable à celui des continuateurs dont on vient de citer les noms.

Pareto s'est voulu le successeur de A. Cournot et de L. Walras dans la fondation d'une économie mathématique. Sa théorie de l'équilibre (le maximum d'ophélimité¹⁰, appelé maintenant *optimum de Pareto*) n'a d'ailleurs guère vieilli. Dans ce qui a été étudié ici, s'il a montré quelques faiblesses de technicité et laissé passer une ou deux bévues, il a surtout fait preuve de son inventivité créatrice, et c'est ce qui compte pour le jugement que l'on peut porter *a posteriori* : il est sûr qu'il fut, pour l'évolution de la Statistique, à la fois un novateur et un précurseur d'une importance capitale.

RÉFÉRENCES

- ARMATTE M., (1995), Robert Gibrat et la loi de l'effet proportionnel, *Mathématiques, informatique et sciences humaines*, **129**, 5-35.
- BARBUT M. (1989), Distributions de type parétien et représentation des inégalités, *Mathématiques, informatique et sciences humaines*, **106**, 53-69.
- BARBUT M. (1991), Note sur les moyennes de variables aléatoires dans *Moyenne, milieu, centre*, ouvrage collectif. Paris, EHESS, 31-43.
- BARBUT M. (1998), Une famille de distributions : des parétiennes aux contreparétiennes. Applications à l'étude de la concentration urbaine et de son évolution, *Mathématiques, informatique et sciences humaines*, **141**, 43-72.
- BERTRAND J. (1899), *Calcul des probabilités*. Paris, Gauthier-Villars.
- BORTKEWITSCH L. VON (1898), *Das Gesetz der kleinen Zahlen*. Leipzig, Teubner.
- ESTOUP J.-B. (1916), *Gammes sténographiques*, 4^e édition. Paris, Institut sténographique.
- FRECHET M. (1925), *Une nouvelle représentation analytique de la répartition des revenus*, XVI^e session de l'Institut international de statistique, Rome.
- FRECHET M. (1941), Sur la loi de répartition de certaines grandeurs géographiques, *Journal de la Société statistique de Paris*, **81**, 115-123.
- FRECHET M. (1945), Nouveaux essais d'explication de la répartition des revenus, *Revue de l'Institut International de Statistique*, **13**, 1/4, 16-32.
- FRECHET M. (1955), *Les Mathématiques et le concret*. Paris, Presses Universitaires de France.

10. Mot forgé par V. Pareto, du grec ophélimos qui signifie utile

HOMME MOYEN OU HOMME EXTREME?

- GIBRAT R. (1931), *Les inégalités économiques*. Paris, Sirey.
- JOHNSON N.L. et KOTZ, S. (1970), *Distributions in Statistics : Continuous Univariate Distributions* – 1, 2. New York, Wiley.
- LEVY P. (1937), *Théorie de l'addition des variables aléatoires*. Paris, Gauthier-Villars.
- MANDELBROT B. (1955), Linguistique statistique macroscopique, dans *Logique, langage et théorie de l'information*, J. Piaget (éd.). Paris, PUF.
- MANDELBROT B. (1960), The Pareto-Levy law and the distribution of income, *International Economic Review*, vol. 4, janvier, 79-106.
- MANDELBROT B. (1960), On the theory of word frequencies and on related markovian models of discourse, in R. Jacobson (éd.), *Structure of language and its mathematical aspects*, 120-219, Providence, RI, Am. Math. Soc.
- MANDELBROT B. (1966), Les constantes chiffrées du discours, dans *Le Langage* (sous la dir. d'A. Martinet). Paris, Gallimard, Encyclopédie de la Pléiade.
- PARETO V. (1896), *Cours d'économie politique*, volume 2. Lausanne, F. Rouge. Nouv. éd. par G. Busino. Genève, Droz, 1964 t. II, liv. III, chap. I.
- PARETO V. (1909), *Manuel d'économie politique*. Paris, Girard. Nouv. éd. par G. Busino. Genève, Droz, 1966, chap. VII.
- PARETO V. (1911), Économie mathématique, in *Encyclopédie des sciences mathématiques*. Paris, Gauthier-Villars, 319-368.
- PARETO V. (1965), *Écrits sur la courbe de répartition de la richesse*, réunis et présentés par G. Busino. Genève, Droz.
- PARETO V. (1989), *Statistique et économie mathématique*, textes réunis par G. Busino. Genève, Droz.
- PUMAIN D. (1982), *La dynamique des villes*. Paris, Economica.
- PETRUSZEWCZ M. (1963), Loi de Pareto et processus markoviens, *Mathématiques et sciences humaines*, 3, 21-29.
- PETRUSZEWCZ M. (1973), Histoire de la loi d'Estoup-Zipf : documents, *Mathématiques et sciences humaines*, 44, 41-56.
- SOREL G. (1897), La loi des revenus, *Le devenir social*, 7, 577-607.
- VALADE B. (1990), *Pareto, la naissance d'une autre sociologie*. Paris, Presses Universitaires de France.
- ZIPF G. K. (1949), *Human Behavior and the principle of least effort*. New York, Hafner.