Journal de la société française de statistique

Michel Armatte

Fréchet et la médiane : un moment dans une histoire de la robustesse

Journal de la société française de statistique, tome 147, nº 2 (2006), p. 23-37

<http://www.numdam.org/item?id=JSFS_2006__147_2_23_0>

FRÉCHET ET LA MÉDIANE : UN MOMENT DANS UNE HISTOIRE DE LA ROBUSTESSE

Michel ARMATTE *

RÉSUMÉ

L'article de Fréchet (1940) n'est pas à proprement parler discuté, mais il est plutôt replacé dans une perspective historique qui s'identifie aux différentes contestations du dogme de la théorie des erreurs, qui relie moyenne et loi des erreurs, et aux différentes tentatives de mettre en avant les propriétés intéressantes de la médiane simple ou pondérée, et plus généralement des quantiles et des statistiques d'ordre.

ABSTRACT

The Fréchet (1940) paper is not veritably discussed. It is replaced in an historical scope, and becomes a moment of the several disputations about "theory of errors" dogma which gives privilege to the average and the Laplace-Gauss law of errors. There are a lot of attempts to advance interesting properties of the simple or weighted median and more generally of the quantiles and the order statistics.

Le texte de Fréchet proposé à la discussion est un très bon point de départ pour discuter de la médiane, sa place parmi les différents résumés de valeur centrale, les propriétés de sa distribution d'échantillonnage, son rôle dans la théorie des erreurs, sa généralisation à des espaces métriques, et le jeu des interprétations qui lui ont été accolées. Cet article doit cependant être situé, d'une part dans l'histoire de la Statistique, d'autre part dans l'œuvre de Maurice Fréchet, et enfin dans sa postérité. Nous proposons ici de nous centrer sur les deux premiers points et d'esquisser un bref aperçu historique de la notion de médiane et, à travers elle, des statistiques de rang ainsi que d'une certaine idée de la robustesse.

1. Théorie des erreurs

Sans remonter aux premiers textes de Galilée sur le traitement des erreurs de mesure – on y trouve déjà l'idée d'une distribution paire, bornée et décroissante des erreurs – on doit partir de la théorie des erreurs pour bien identifier la place qu'y jouent les caractéristiques de valeur centrale. L'article de Jean III Bernoulli dans le Supplément à l'Encyclopédie Méthodique (1777) est une

* Université Paris-Dauphine et Centre A. Koyré
e-mail : michel.armatte@dauphine.fr

bonne entrée dans le problème : « *Quand on a fait plusieurs observations d'un même phénomène, et que les résultats ne sont pas tout à fait d'accord entre eux, on a coutume alors de prendre le milieu entre tous les résultats, parce que de cette manière les différentes erreurs se répartissent également dans toutes les observations, l'erreur qui peut se trouver dans le résultat moyen devient aussi moyenne entre toutes les erreurs* ». Remarquons dans ce texte la terminologie de milieu qui indique encore un choix possible entre différents *milieux* de l'erreur, et la fin de la phrase qui a déjà fait le choix de la moyenne. Une sorte de confirmation que la moyenne est la forme la plus habituelle de milieu à prendre. Remarquons aussi la confusion qui peut se lire entre les lignes : l'erreur sur une moyenne est bien une erreur moyenne, mais sa variabilité n'est pas la moyenne de la variabilité de l'erreur. C'est un des premiers théorèmes mathématiques dû à Simpson en 1755, qui, pour la première fois, introduit diverses lois des erreurs explicitement formulées pour montrer le gain de précision de la moyenne par rapport à celle d'une observation. L'auteur de l'article « Milieu » fait ensuite la revue des « *différents grands géomètres qui ont entrepris cette utile recherche* » et tenté de justifier ou d'améliorer le choix de la simple moyenne arithmétique : Boscovich, Lambert, Bernoulli (Daniel), Lagrange. Ces travaux, auxquels il faudrait rajouter le mémoire de Laplace de 1774, marquent le début d'un traitement mathématique de la question des erreurs, qui s'achèvera en 1820 quand Laplace nous offre dans sa deuxième édition de la *Théorie analytique* une magistrale synthèse de ses travaux, et de ceux de Gauss, synthèse selon laquelle la théorie des erreurs se conclut par un triple choix : celui de la moyenne comme milieu, de la seconde loi de Laplace (en e^{-x^2}) comme loi des erreurs, et de la méthode des moindres carrés comme principe d'ajustement. Ce triplet gagnant est consolidé par plusieurs théorèmes qui lient ces trois ingrédients et fournissent les bases de la statistique mathématique : Principe de la moyenne chez Encke et Gauss, théorème de la limite centrale de Laplace, théorème de Gauss-Markov sur l'optimalité des estimateurs des moindres carrés. Cette théorie va dominer le siècle suivant et opérer un grand nombre de révolutions et d'innovations dans les sciences physiques (astronomie, géodésie, balistique...) et sociales (physique sociale de Quetelet, biométrie de Galton, économétrie...). Mais qui dit dominer sous-entend que d'autres approches ont été dominées sans êtres réduites à l'inexistence. Principalement celles de la médiane.

Pour préciser ce dont on parle, et traduire les débats de l'époque sans trop les trahir, notons qu'un modèle d'observations, linéaire ou linéarisé par approximation, peut s'écrire

$O_i = a + \varepsilon_i$, dans le cas d'observations directes O_i de la grandeur a entachées d'erreurs ε_i aléatoires de loi L. Le modèle estimé sera $O_i = \hat{a} + e_i$ où $\hat{a}$ est un estimateur de a et e_i le résidu ou écart pour l'observation i. Mais ce modèle n'est qu'un cas particulier ($p = 0$) du modèle suivant :

$[M_p]$ $O_i = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + \ldots + a_p x_{i,p} + \varepsilon_i$. Dans ce cas dit « d'observations indirectes », les observations dépendent de variables observables $x_{i,1}, \ldots, x_{i,p}$, les paramètres inconnus sont $a_0, \ldots, a_p$. Le modèle estimé est : $O_i = \hat{a}_0 + \hat{a}_1 x_{i,1} + \hat{a}_2 x_{i,2} + \ldots + \hat{a}_p x_{i,p} + e_i$

Laplace a 24 ans et est déjà académicien quand il rédige son mémoire sur la probabilité des causes (1774) dans lequel il cherche à résoudre la question du « milieu à prendre entre plusieurs observations » dans le cas des observations directes (M_0) :

1) Pour une loi de densité f, il établit que, la médiane est le milieu qui minimise la somme des écarts absolus :

$$M_1 \text{ solution de } \int_0^{M_1} f(x)dx = 1/2 = M_2 \text{ solution de Min} \int_{-\infty}^{\infty} |x - M_2| f(x)dx.$$

2) il propose comme loi des erreurs *a priori*, découlant d'un principe assez fallacieux, ce que l'on appellera ensuite la première loi de Laplace : $\varphi(x) = (1/2)me^{-m|x|}$.

3) Par la méthode de la probabilité inverse – qu'on appelle ensuite la formule de Bayes – il en déduit une densité a posteriori pour a conditionnelle à 3 observations $O1$, $O2$, $O3$. Celle-ci n'a pas de forme analytique simple et il ne réussit pas à en déterminer la médiane. Notons qu'avec cette même densité, s'il avait choisi comme vraie valeur le mode et non la médiane de la loi *a posteriori*, la solution aurait été la médiane des 3 observations. C'est ce que Glaisher démontrera pour un nombre quelconque d'observations en 1872, donc un siècle plus tard !

4) Dans un second mémoire (1777), Laplace reprend la même méthode avec une densité logarithmique de la forme $\varphi(x) = (1/2a)\log(a/|x|)$ définie sur $[-a, a]$ avec une asymptote verticale en $x = 0$ sans aboutir davantage à un résultat exploitable. Mais la médiane comme milieu et l'écart absolu moyen à la médiane sont en place. Lagrange et Bernoulli (Daniel), dans deux mémoires contemporains, n'ont pas davantage réussi à résoudre cette question simple du « milieu le plus avantageux ».

La question se débloque à l'occasion des travaux sur la « figure de la terre » c'est-à-dire la controverse entre deux formes possibles du globe, une forme d'ellipsoïde aplatie aux pôles (c'est la théorie de Newton, et de Maupertuis) ou allongée aux pôles (c'est la théorie de Descartes et des Cassini, qui trustent sur trois générations la direction de l'Observatoire de Paris). Sous l'hypothèse d'un ellipsoïde, la longueur d'un degré de méridien en un point de latitude l est une fonction linéaire de $\sin^2 l$. Après les deux expéditions académiques de 1735 au Pérou et en Laponie, les mesures se multiplient, et l'on dispose vers 1770 d'un système de 13 relations de type M_1 à deux inconnues a_0 et a_1 ou, si l'on préfère, d'un nuage de 12 points qu'il faut ajuster par une seule droite, dont la pente va définitivement partager Maupertuis et les Cartésiens qui en sont aux mains !

C'est dans ce contexte que Joseph Roger Boscovich, jésuite d'origine dalmate né à Raguse (Dubrovnik), chargé de déterminer un degré de méridien de Rome, avait d'abord conclu d'après cinq observations que « la détermination de ces degrés ne peut pas être réconciliée avec l'ellipse de Newton ». Dans une nouvelle publication de 1770 (avec l'anglais Christopher Maire) il imagine la solution suivante : prenons pour droite estimée celle qui 1) assure une compensation des résidus e_i (toutes celles qui passent par le point moyen du nuage) 2) qui minimise la somme des valeurs absolues des mêmes résidus. Ce programme n'a pas de solution analytique, mais Boscovich remarque qu'en faisant tourner cette droite dans le sens de la montre depuis la verticale, on rencontre successivement les points (x_i, y_i) pour lesquels les rapports $(x_i - \overline{x})/(y_i - \overline{y})$ sont en ordre croissant. La somme des corrections va diminuer puis augmenter proportionnellement à $|x_i - \overline{x}|$. Pour avoir un minimum de cette somme, il suffit de s'arrêter dans cette rotation quand la somme des quantités $|x_i - \overline{x}|$, rangées dans l'ordre des pentes, atteint la moitié de la somme totale des mêmes quantités. Cette solution est une médiane pondérée, au sens où elle minimise $\sum |(a_i - a)(x_i - \overline{x})|$.

Laplace va reprendre cette méthode en 1786 et 1793 en la rebaptisant méthode de situation. Il y ajoute seulement une preuve que l'algorithme fournit bien la solution du programme, et, en 1799, une variante dans laquelle les observations sont pondérées. Mais aucun argument probabiliste ne vient étayer cette solution. En 1786 Laplace a proposé également une autre solution, celle du minimax qui consiste à prendre la droite qui minimise l'erreur maximale. C'est une première apparition d'une valeur extrême.

C'est à ce moment qu'intervient la solution des moindres carrés proposée par Legendre (1805), avec une antériorité revendiquée ensuite par Gauss. La seule justification donnée par Legendre est la simplicité (on aboutit à un jeu d'équations normales linéaires) et le fait que, pour les observations directes, la méthode redonne un milieu connu : la moyenne arithmétique; puisque $\mathrm{Min} \sum (x_i - a)^2$ a pour solution $a = \sum x_i/n$.

Vers 1810, rien n'est cependant résolu. On est en face d'un assez grand nombre de choix de milieux, de lois possibles des erreurs, et de critères d'ajustement. Ce sont les théorèmes successifs de Gauss et de Laplace qui vont construire une synthèse dans laquelle la moyenne arithmétique, la deuxième loi de Laplace en $exp(-x^2)$, et les moindres carrés constituent le triplet gagnant. Ces théorèmes ne prouvent pas que le bon milieu à prendre est la moyenne, ou que la bonne loi est la loi de Laplace-Gauss, mais ils associent ces résultats. Partant d'un principe, à savoir que le milieu des observations doit être la moyenne, et d'une méthode d'ajustement qui est le maximum de la probabilité *a posteriori* des paramètres conditionnellement aux observations, Gauss (1809) en déduit que la loi des observations est la deuxième loi de Laplace [1], dès lors dite de Laplace-Gauss. En 1810 Laplace justifie celle-ci par son fameux théorème de convergence – le théorème central limite – démontré par l'introduction de fonctions génératrices. Il l'étend aux observations indirectes en 1811, et la relie à la procédure des moindres carrés : si on a un grand nombre d'observations, les erreurs sur les paramètres (qui sont des combinaisons des ε_i) sont « normales » et le critère du minimum de « l'erreur moyenne à craindre » [2] – c'est-à-dire la formule de 1774 – conduit aux solutions des moindres carrés. Tout cela

1. En fait c'est la troisième si on compte celle de 1777, ce qu'on ne fait jamais.

2. Pour suivre dans le détail ces diverses démonstrations, voir – outre les œuvres originales – Andoyer (1908), Stigler (1986), Armatte (1995, 2004).

est repris dans le chapitre IV de la Théorie analytique et complété dans les Suppléments de la troisième édition. Gauss, dont Laplace a critiqué la justification de 1809 [3], a le dernier mot en 1821 en montrant que l'on peut se passer de la normalité et des conditions asymptotiques de Laplace, et justifier le procédé des moindres carrés (et la moyenne) par le choix d'un nouveau critère : la minimisation de l'erreur quadratique. Gauss a changé son fusil d'épaule : Mieux vaut «minimiser cette fonction de perte que de maximiser la probabilité d'une erreur nulle» dit-il en substance. Cette suite de théorèmes construit une théorie, mais une théorie qui ne justifie aucun choix pratique, puisqu'il faut chaque fois un ou deux principes (ou postulats) sur deux des éléments pour justifier la forme du troisième.

2. Médiane et robustesse

Au cours du XIXe siècle, la contestation se développe sur les trois éléments de la théorie des erreurs. Un astronome comme Encke (et plus tard Schiaparelli fit de même) tenta de justifier le choix de la moyenne par deux propositions axiomatiques : 1) la probabilité d'une erreur ne dépend que de son amplitude 2) le résultat le plus probable pour n observations est une combinaison des résultats probables donnés par différents sous-groupes qui ne fait pas intervenir les valeurs elles-mêmes. Soit encore que la densité est paire et que le résumé à choisir est exhaustif. Mais ni ce système axiomatique de Encke (1832) ni celui de Schiaparelli (1868) ne furent considéré comme des fondements corrects et suffisants du «principe de la moyenne». Ellis (1844) suggère que toute solution du programme $\mathrm{Min}\sum f(x_i - a) = 0$ est aussi intéressante que la moyenne obtenue pour le cas particulier où f est l'identité. Autant prendre directement comme axiome que la moyenne est le meilleur milieu. Aucune des justifications de la loi gaussienne des erreurs ne fut prise au sérieux complètement. Ni la théorie de Laplace malgré sa défense par Bienaymé contre Cauchy, ni celle des erreurs élémentaires mise en avant par Hagen ne purent venir à bout des expérimentateurs qui avaient parfois des raisons très sérieuses de ne pas accepter une seule loi des erreurs pour toutes les situations expérimentales [4]. Quant au principe des moindres carrés, face au poids déraisonnable qu'il donne aux grandes erreurs en valeur absolue, il fut bien souvent remplacé dans les applications pratiques par la méthode de

3. «Comme rien ne prouve que la règle de départ [la moyenne] donne le résultat le plus avantageux, la même incertitude existe par rapport à la seconde [les moindres carrés]» écrit-il.

4. Voir par exemple la lettre de Bravais à Quetelet reproduite dans Quetelet (1946) : Bravais y décrit différents cas de mesure où les erreurs ne sont pas gaussiennes.

Cauchy [5], encore plus simple, et surtout plus robuste ; et il finit par être retiré de l'enseignement des polytechniciens par la réforme Le Verrier de 1850.

N'y avait-il pas un autre milieu qui soit lié à une autre loi des erreurs par un autre principe d'optimisation ? C'est ce que tentèrent tardivement quelques géomètres en proposant la médiane des observations, laquelle ne jouissait ni des propriétés de linéarité, ni de celles d'exhaustivité, mais avait été utilisée depuis bien longtemps par les astronomes comme valeur intermédiaire entre plusieurs observations divergentes.

Dans le second Supplément (1818) de la *Théorie Analytique*, Laplace avait tenté de le faire en comparant deux solutions du modèle M_1 sans constante $y_i = ax_i + \varepsilon_i$. Rappelons :

– que la solution des moindres carrés est : $\hat{a}_L = \sum x_i y_i / \sum x_i^2$. Elle peut s'écrire $\sum (x_i^2 / \sum x_i^2) a_i$ et apparaît alors comme une moyenne des a_i pondérée par les x_i^2.

– que la solution de Mayer et Cauchy est dans ce cas $\hat{a}_{MC} = \sum y_i / \sum x_i = \overline{y}/\overline{x}$. Elle peut s'écrire $\sum (x_i / \sum x_i) a$, et apparaît alors comme une moyenne des a_i pondérée par x_i.

– et que la solution de Boscovich donne $\hat{a}_B = y_i / x_i$ où i représente le numéro de l'observation qui correspond à la médiane des x_i. Donc $\hat{a}_B$ est la médiane pondérée des $a_i = y_i / x_i$.

Dans le cas $x_i = 1$ (Modèle M_0), Laplace (1818) établit les distributions de probabilité asymptotiques séparées de deux « estimateurs » de a : celui des moindres valeurs absolues et celui des moindres carrés, pour des erreurs ε_i suivant une loi paire $\varphi(x)$ quelconque. La médiane pondérée de Boscovich est asymptotiquement distribuée selon une loi normale de variance $1/\{4\varphi^2(0)\sum e_i^2)\}$ où les e_i sont les résidus observés pour la solution médiane. Cette solution n'est préférable à celle des moindres carrés que sous la condition : $\varphi(0) > 1/2\sigma$, où σ est l'erreur quadratique moyenne, une condition que ne remplit pas la loi de Gauss ($\varphi(0) = 1/\sigma\sqrt{2}\pi$). Laplace, cherchant visiblement une combinaison des deux « estimateurs », établissait alors leur loi conjointe et l'exhaustivité de la solution des moindres carrés dans le cas d'erreurs gaussiennes, au sens suivant : l'estimateur des moindres carrés est meilleur que toute combinaison linéaire de cet estimateur avec celui de Boscovich. (Stigler 1973a) a vu dans ce travail une anticipation de la démonstration similaire que Fisher établit en 1920 lorsqu'il compare les deux mesures de dispersion que sont l'écart-type et l'écart absolu moyen. C'est encore ce même problème que reprendra Fréchet en 1940.

Cournot (1843) est à l'origine d'une réhabilitation de la médiane. Il substitue cette dénomination moderne à celle de « valeur probable » utilisée par les

5. Cauchy se propose de déterminer la meilleure approximation linéaire possible d'une quantité $y = f(t)$ par une série de variables x_1, x_2, x_3 qui résultent d'un développement en série de $f(t)$. On s'arrête dans l'approximation quand la fonction d'écart passe en dessous d'un seuil fixé. Cauchy ne choisit pas la somme des carrés comme fonction d'écart. Il propose un algorithme d'interpolation plutôt qu'une solution à un problème d'ajustement. Pour la première approximation de la forme M_1 sans constante, la solution de Cauchy est celle de Mayer.

astronomes allemands (Gauss et Bessel) ou encore à celle de « valeur centrale » de Gustave Fechner. Ce psycho-physicien a considéré, en 1878, la médiane comme un cas particulier des valeurs centrales définies par Min$\{\sum |y_i - m|^n\}$, et il a démontré, un peu plus généralement que Glaisher (1872), que si l'on cherche non pas le milieu d'erreur de la loi a posteriori des observations mais son maximum, alors une loi a priori conforme à la première loi de Laplace en $e^{-k|x|}$ conduit à prendre pour milieu la médiane des observations; et vice-versa le choix de la médiane comme la valeur la plus probable implique une loi des erreurs de ce type. Le critère de la probabilité maximale du système des observations conduit ainsi à deux associations concurrentes :

1. Moyenne – Loi de Gauss – Moindres carrés,
2. Médiane – première Loi de Laplace – Méthode de Situation.

Le principe de la médiane est donc tout aussi productif que le principe de la moyenne.

Y a-t-il d'autres associations possibles d'un milieu et d'une loi par ce même principe du maximum de la probabilité ? C'est une question que tente de résoudre Joseph Bertrand (1889, p. 181 et suivantes). Cherchant plus précisément la loi associée à un milieu quelconque $(x_1, x_2, ..., x_i, ..., x_n)$ entre n observations, il observe qu'une telle loi n'existe pas pour toute fonction sauf si celle-ci vérifie une condition sur la translation du genre $\varphi(\{x_i + a\}) = (\{x_i\}) + a$. La moyenne quadratique, la moyenne géométrique, par exemple, ne vérifient pas cette propriété et n'ont pas de loi associée. Avec la même problématique, on trouvera aussi un long appendice mathématique à un papier de Keynes (1909) sur les indices monétaires dans lequel il évalue, sur les traces d'Edgeworth, le transport analogique de la théorie des erreurs dans celle des indices.

Le poids trop important donné aux grandes erreurs par la méthode des moindres carrés pose des problèmes de robustesse, c'est-à-dire de sensibilité à des valeurs accidentelles. Il faut donc rediscuter de ce qui était au début de la théorie des erreurs, c'est-à-dire de l'élimination de mesures entachées d'erreurs systématiques ou « anormales ». Cela peut se faire par trois voies. L'élimination a posteriori des *outliers* sur la base d'une distribution normale est discutée par Benjamin Peirce (1852), Airy (1861), De Morgan (1861), Glaisher (1872, 1873). La seconde voie part du principe que les mesures sont de qualité hétérogène, et qu'il faudrait donner moins de poids aux observations les plus discordantes. Mais il faut aussi traduire cette hétérogénéité par une modification de la loi des erreurs. Dans cette veine il faut signaler d'abord les travaux de Simon Newcomb : il suggère d'ajuster un mélange de lois normales sur les résidus des moindres carrés, et prend pour estimateur la moyenne de la densité *a posteriori*. Stigler (1973b) a montré qu'il anticipe ainsi la théorie des M-estimateurs de Peter J. Huber. L'autre piste développée par Edgeworth consiste à dériver une loi des erreurs généralisée, à partir de la loi normale par un simple changement de variable. Le cas le plus intéressant est celui de la transformation logarithmique qui s'impose en psychologie (à la suite encore des travaux de Fechner sur la sensibilité stimulus-réponse) et en économie (à la suite des travaux d'Edgeworth et de Gibrat sur les distributions de prix et de revenus). L'origine de la loi log-normale apparaît dans les travaux de Fechner et de Galton en même temps que la discussion de la moyenne géométrique. Les deux pistes se rejoignent quand Edgeworth suggère en 1888 l'utilisation de la médiane dans le cas de distributions assimilables à des mélanges de lois normales.

Francis Galton a développé toute sa théorie de la régression et de la corrélation sur la base de données supposées suivre une loi normale. Cependant il est l'introducteur d'un usage systématique de la courbe des fréquences cumulées (« l'ogive ») et des quantiles. Sa loi de « régression vers la médiocrité » établie en 1885 s'énonce comme une linéarité des médianes de y conditionnelles à une plage de valeurs de x, et non comme le lieu des moyennes liées. Son découpage de la loi normale en fractiles lui permet toutes sortes d'interprétations directes de la loi en cloche en tant que dispositif d'interclassement des individus. C'est aussi dans une note de bas de page de *Natural inheritance* (1889) qu'apparaissent deux formules d'estimateurs de la moyenne et de l'écart-type d'une loi normale en fonction de deux quantiles d'ordre p et q, une intuition qui sera généralisée par Sheppard (1899). Une exploration systématique des propriétés des estimateurs linéaires fondés sur des statistiques d'ordre (L-estimateurs) se trouve dans un papier prémonitoire de Daniell (1920).

Ces différents travaux portent à la fois sur la recherche d'une loi des erreurs alternative, et sur d'autres caractéristiques de valeurs centrales. Ils forment le soubassement de ce que, depuis Box (1953), on a désigné par *robustesse*.

3. Maurice Fréchet et la médiane

Maurice Fréchet, normalien et agrégé de mathématiques, après différents postes en province (Besançon, Nantes) et une interruption due à la première guerre mondiale, rejoint la nouvelle université de Strasbourg où il enseignera les mathématiques puis la statistique de 1919 à 1927. Il y croise le sociologue Maurice Halbwachs et tous deux signent un petit ouvrage sur « le calcul des probabilités à la portée de tous » (Fréchet et Halbwachs 1924), qui recevra le prix Montyon de statistique en 1925. La comparaison des propriétés de la moyenne et de la médiane est déjà présente dans ce petit opuscule. Bien que les principales recherches de Fréchet avant 1940 se situent dans le domaine des mathématiques pures et des espaces abstraits, et lui valent sa réputation internationale, elles ont une résonance directe dans le domaine de la statistique mathématique. En effet, la notion de distance, généralisée aux espaces de fonctions ou de variables aléatoires, lui permet d'envisager une extension des notions de valeurs typiques, de dispersion, et de convergence stochastique. Par exemple il écrit (Fréchet 1947b) :

« Soient X un élément aléatoire, a un élément certain, appartenant tous deux à un espace distancié D. On appellera écart moyen d'ordre r de X la borne inférieure, finie ou non de $(E\ |X-a|^r)^{1/r}$ quand a varie. »

Ce qui permet de définir de nombreuses valeurs typiques dont la médiane et la moyenne respectivement pour $r = 1$ et 2. Fréchet généralise cette définition à des espaces fonctionnels et à des espaces vectoriels de Banach. Muni de cette idée de distance dans des espaces abstraits, Fréchet va revisiter systématiquement une bonne partie des concepts de la théorie des erreurs et de la statistique mathématique. Dans un autre article (Armatte 2002) nous avons relaté ses critiques du « soi-disant coefficient de corrélation » dans une campagne retentissante menée à l'IIS au moyen d'enquêtes par correspondance, de discussions en séances plénières, et de motions en bonne et due forme. Ces actions s'inscrivent directement dans une sorte de mission qu'il s'est donnée, et qu'il explicite dans ses nombreux discours. Prenant par exemple la Présidence de la Société Statistique de Paris en 1948, à la suite d'Alfred Sauvy, Fréchet explique dans son allocution de président entrant qu'il est, après Borel (1922) et Darmois (1938), le troisième mathématicien à présider la Société, et que sa mission est de conforter les bases de la méthode statistique, par l'usage des mathématiques, tout particulièrement les notions d'espace métrique qu'il a développées lui-même, dans un cadre international où « la vocation de la France est d'apporter ordre et lumière là où régnait confusion et obscurité ». A la suite de cette controverse sur le coefficient de corrélation, il publiera effectivement (Fréchet 1947a) de nouveaux indices de corrélation fondés sur cette notion de distance. La même notion lui a permis de revisiter la question de la théorie des erreurs d'observation dans plusieurs textes de la période 1922-1929. Il y propose une remise en question « de l'hypothèse de l'additivité partielle des erreurs » (1928) et de la loi normale qui en résulte. Il étudie la loi des extrêmes qui s'avère être du type des lois stables de Pareto-Lévy. Il propose de prendre pour résultat celui qui minimise la plus grande des erreurs partielles en valeur absolue, reprenant la vieille solution du minimax par Laplace. Fréchet a publié en Pologne (1927) des résultats sur la distribution de l'erreur maximale, qui ont suscité les commentaires peu enthousiastes de Paul Lévy [6]. Fréchet s'appuie sur les travaux de Fisher et sur ceux de Emil Julius Gumbel, dont il faut rappeler qu'il fut son protégé lors de son séjour en France de 1932 à 1940, à son arrivée d'abord (invité par Borel à l'IHP), dans ses multiples tentatives de naturalisation ensuite jusqu'en 1939, sous le gouvernement de Vichy enfin quand il est déchu comme juif de sa nationalité française (Hertz, 1997). Gumbel a publié des articles et plusieurs notes aux *Compte-Rendus à l'Académie des Sciences* sur la statistique des extrêmes entre 1932 et 1937, dans lesquelles lui aussi s'affranchit des hypothèses gaussiennes. Une de ces notes, présentée par Henri Eyraud, son directeur à *l'Institut des Sciences financières et actuariales* de Lyon, travaille la question de la médiane (Gumbel 1934), une autre note de 1937 traite « de la précision de la moyenne arithmétique et de la médiane ». Les travaux de Gumbel trouveront de nombreuses applications en démographie (durée extrême de vie), en hydrologie (pluies, crues et inondations), en physique (émissions radioactives).

Le papier que publie Fréchet dans le Journal de la Société de Statistique de Paris en 1940 « Sur une limitation très générale de la dispersion de la médiane », et qui est à l'origine du présent débat, s'inscrit en continuité dans

6. Ces résultats seront cités par Paul Lévy dans une conférence de 1930 à l'IHP. Mais une lettre de Paul Lévy à Fréchet d'octobre 1928 affirme que «l'article polonais est sans grand rapport avec la théorie des erreurs, en raison du caractère de la loi de composition adoptée qui ne me paraît guère avoir de chances de se rencontrer en pratique».

cette tentative de fonder une théorie alternative des erreurs. L'introduction revendique une nécessaire révision des standards de la méthode statistique au regard de la rigueur des concepts mathématiques, et lui permet pour un moment de s'identifier à Cauchy. Le propos plus modeste du texte est cependant de restreindre le monopole de la moyenne et de l'écart quadratique moyen, et de réhabiliter la médiane comme résumé de valeur centrale en montrant que sa dispersion peut être dans certains cas meilleure que celle de la moyenne. En 1940, Fréchet n'a pas encore pris connaissance de l'ensemble des travaux de Ronald Fisher – il aura un échange de correspondance assez difficile avec lui sur la question de l'estimation par la « probabilité fiduciaire » – et c'est par l'intermédiaire de Gumbel qu'il accède à sa formule de l'écart-type estimé de la médiane, que l'on pourra comparer avec celle de Laplace (1818). Dans le cas de lois paires dont le mode (qu'il appelle encore la dominante) s'identifie à la médiane, Fréchet note que la dispersion de la médiane reste bornée par un nombre supérieur à la dispersion de la moyenne si on utilise le critère de l'écart quadratique moyen ou écart-type. Mais cela ne veut pas dire qu'il n'y ait pas des lois pour lesquelles la dispersion de la médiane passe en dessous de celle de la moyenne. Le contre-exemple est tout trouvé avec la première loi de Laplace. Cependant si l'on change de critère de la mesure de dispersion et que l'on adopte l'écart absolu moyen à la médiane (écart probable), incontestablement plus facile à calculer pour les statisticiens rappelle-t-il, alors l'avantage de la médiane devient important car on ne peut borner l'écart absolu de la moyenne. Je ne sais pas quelle a été la réception de ce travail.

Fréchet va revenir sur la médiane quelques années plus tard, dans sa conférence au Palais de la Découverte intitulée « *Réhabilitation de la notion statistique de l'homme moyen* ». Il est vraisemblable que c'est avec Halbwachs qu'il eut les premières discussions sur la problématique de l'homme moyen chère à Quetelet, car Halbwachs avait fait sa thèse complémentaire sur ce sujet en 1912. Il y développait surtout une critique de la construction microprobabiliste de l'homme moyen comme résultante d'un grand nombre de petites variations indépendantes, et de l'assimilation consciente chez Quetelet d'une moyenne subjective (mesures d'objets différents) à une moyenne objective (mesure d'un même objet référent). En bon disciple de Durkheim, Halbwachs n'acceptait pas de réduire le type collectif à un type moyen, et il voyait d'autres mécanismes à l'œuvre dans la construction holiste de la société que la loi des grands nombres, par exemple la présence de normes sociales, ou bien encore l'adaptation au milieu dans un sens darwinien.

Fréchet s'attache davantage aux critiques logiques déjà faites par Cournot, Edgeworth, Bertrand et quelques autres sur l'impossible agrégation des moyennes partielles de chaque caractéristique en un homme moyen. Celui-ci n'est ni un modèle, ni même un représentant de son groupe. Il est tout simplement impossible, car il ne peut être moyen à la fois en taille et en poids ou volume (c'est la critique de Bertrand). L'homme moyen est tout simplement monstrueux donc inhumain. Fréchet va donc proposer deux autres choix de valeurs centrales pour lesquels l'agrégation produit un homme « possible ». La première consiste à remplacer l'homme moyen par l'homme médian ce qui amoindrit notablement ses difformités, puisque la sensibilité de la médiane aux valeurs extrêmes est moindre que celle de la moyenne. Mais il y a de grandes chances pour que l'homme médian reste irréaliste. La seconde solution consiste à définir l'homme typique comme un homme réel de la population qui se trouve « au milieu » de celle-ci du point de vue de l'ensemble de ses caractéristiques. Ce qui nous amène évidemment au choix d'une formule de la distance à minimiser entre cet homme et le groupe. Fréchet propose de prendre pour chaque couple de personnes un indice de dissemblance (une distance) en forme de médiane des écarts absolus sur chaque dimension, et d'agréger ces indices par une simple somme sur tous les couples possibles pour aboutir à une distance de chacun au groupe.

La détermination de la commune qui incarnerait le milieu de la France a fait l'objet de nombreuses controverses parce que justement le choix des sujets, des mesures, des pondérations à combiner dans une telle procédure est très ouvert. Bernard Monjardet (1991), dans un livre collectif qui s'attachait à cette question, a retracé la généalogie de cette problématique de la médiane métrique. La réponse esquissée par Fréchet dans sa conférence sur l'homme moyen, ainsi que plusieurs mémoires plus généraux sur les valeurs typiques en font partie (Fréchet 1948). La médiane statistique simple d'un échantillon est la solution de ce problème en dimension 1. En dimension 2, dans le plan, le problème a été posé pour 3 points par Fermat en 1629, et pour n points par l'économiste Alfred Weber (frère du sociologue Max Weber) à propos de la localisation des industries, en 1909. La question devint plus tard l'objet d'une controverse sur le « centre de la population américaine » que calculait l'office de recensement. Les statisticiens s'en sont emparés : Eells en 1930, Gini et Galvani en 1929, et Fréchet comme on l'a vu dans les années 1940. Une autre application des médianes métriques a été introduite par G.-Th. Guilbaud (1952) et Barbut (1980) dans les espaces métriques discrets, à propos de la théorie des votes chez Condorcet : si chaque votant formule son vote sur p candidats sous forme d'un ordre total, le classement par paires à la majorité, qui peut aboutir à des circuits du type $A > B > C > A$, peut s'analyser comme la recherche d'un tournoi médian dans l'espace métrique des tournois munis d'une distance fondée sur le nombre de désaccords, qui est du type de celle que Kendall a produite pour son coefficient de corrélation de rang.

L'objectif de ce papier était de retracer une histoire de la médiane jusqu'aux travaux de Fréchet et à « l'invention » de la robustesse par Box. C'était en 1953. C'est ce que l'on dit. En fait il s'agit du premier usage attesté du terme, et nous venons de voir que le concept a une histoire plus ancienne que le nom. Nous en resterons donc là sachant bien que cette histoire est loin d'être achevée.

Références

AIRY G.B. (1861). On the Algebraic and numerical Theory of Errors of Observations and the Combination of Observations, London Macmillan.

ANDOYER H. (1908). Théorie des Erreurs, *Encyclopédie des Sciences Mathématiques Pures et Appliquées*, Exposé d'après l'article allemand de J. BAUSCHINGER, édition française de J. Molk, Paris, Gauthier-Villars, p.161-195.

ARMATTE M. (1995). Histoire du Modèle linéaire. Formes et usages en Statistique et en Économétrie jusqu'en 1945, Thèse EHESS, sous la dir. de J. Mairesse.

ARMATTE M., (2002). Maurice Fréchet Statisticien, enquêteur et agitateur public, *Revue d'histoire des Mathématiques*, **7** (2001), p. 7-65.

ARMATTE M. (2004). La théorie des erreurs (1750-1820), enjeux, problématiques, résultats. *Histoires de Probabilités et de Statistiques*, E. Barbin et J.P. Lamarche (ed), IREM, Ellipses, p.141-160.

ARMATTE M. (2005). Lucien March : statistiques sans probabilité, *Journal électronique d'histoire des probabilités et de la statistique*, Vol.**1**/N°1, mars 2005. www.jehps.net.

BARBUT M. (1980). Médianes, Condorcet et Kendall, *Math.Sci.Hum.*, 69, p. 5-13.

BERNOULLI J. (III), (1789). Milieu, *Encyclopédie Méthodique*, tome II, p.404-409.

BERTRAND J., (1889). *Calcul des probabilités*; 1ère éd. 1889 Paris, Gauthier-Villars; 2ème éd. 1907, 3ème édition, New York, Chelsea, XLIX + 317 p.

BOSCOVICH R. J. et MAIRE C. (1770). Voyage Astronomique et Géographique dans l'Etat de l'Eglise entrepris par l'ordre du Pape Benoît XIV pour mesurer deux degrés de méridien et corriger la carte de l'Etat ecclésiastique, Paris, Tilliard; trad. de De Litteraria Expeditione per Pontificiam ditionem ad dimetienda duas Meridiani gradus, Rome, Palladis, 1755.

BOX G.E.P. (1953). Non Normality and Tests on Variances, *Biometrika*, 40, 318-35.

BRUNT D. (1931). *The combination of observations*, 2ème édition (1ère edition 1917), Cambridge Univ. Press.

CARVALLO E. (1912). *Le Calcul des Probabilités et ses applications*, Gauthier-Villars, Paris, 163 p.

COURNOT A. A. (1984). *Exposition de la théorie des chances et des probabilités*, Oeuvres Complètes, Ed Bru, Vrin, Paris, 1ère éd. 1843.

DANIELL P.J. (1920). Observations weighted according to order, *Amer. J. Math.*, 42, p. 222-36.

DE MORGAN A. (1861). On the Theory of Errors of Observations, Trans. Cambr. Philos. Soc., 10, 109-127.

EELLS W.C. (1930). A mistaken conception of the centre of a population, *J. Amer. Statist. Assoc.*, **25**, p.33-40.

ELLIS R. L. (1844). On the foundation of the Theories of Probabilities, *Trans. of Cambridge Phil. Soc.*, vol. VIII, part I, p. 204-219.

ENCKE J.F. (1832). Über die Methode der Kleinsten Quadrate, *Berliner Astronomisches Jahrbuch für 1834*, 249-312.

EISENHART C. (1977). Boscovich and the combination of observations, in Kendall et Plackett ed., Studies *in the history of Statistics and probability*, vol. 2, Griffin, London, p.88-100.

FAA DE BRUNO, (1869). *Traité élémentaire des erreurs*, Gauthier-Villars, Paris.

FELDMAN J. *et alii* (eds) (1991). *Moyenne, milieu, centre. Histoire et usages.* Paris, Ed. de l'EHESS.

FISHER R. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the lean square error, in Benett (ed), *Monthly notices of R.S.*, et *Collected Papers of R.A. Fisher*, N°12, Univ. of Adelaïde.

FRECHET M. (1922). *Sur la théorie des erreurs d'observation*, conférence non publiée de Berne.

FRECHET M. (1924). Sur la loi des erreurs d'observation, *Bulletin de la Société Mathématique de Moscou*, **33**, p. 5-8.

FRECHET M. (1925). Sur la loi des erreurs d'observation, *CRAS*, Paris, 1925, p. 204-206.

FRECHET M. (1927). Sur la loi de probabilité de l'écart maximum, *Annales de la Société Polonaise de Mathématiques*, **5**, p. 93-116.

FRECHET M. (1928). Sur l'hypothèse de l'additivité des erreurs partielles, *Bulletin des Sciences Mathématiques*, 2e série, **63**, p.203-216.

FRECHET M. (1940). Sur une limitation très générale de la dispersion de la médiane, *Journal de la Société Statistique de Paris*, p. 67-79.

FRECHET M. (1947a). Anciens et nouveaux indices de corrélation. Leur application au calcul des retards économiques, *Econometrica*, **15** (janvier 1947), p.1-30. Errata, octobre 1947, p. 374-375.

FRECHET M. (1947b). Les espaces abstraits et leur utilité en statistique théorique et même en statistique appliquée, *Journal de la Société Statistique de Paris*, p. 410-421.

FRECHET M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié, *Annales de l'I.H.P.*, **14**, p.215-310.

FRECHET M. (1949). Les valeurs typiques d'ordre nul ou infini et leur généralisation, in *Le calcul des Probabilités et ses applications*, Colloque de Lyon, Paris, ditions du CNRS, p. 47-51.

FRECHET M. (1950). Réhabilitation de la notion statistique de l'homme moyen, *conférence du Palais de la Découverte* du 15 octobre 1949, 24 p.; réédité in [Fréchet 1955, p. 317-341].

FRECHET M. (1955). *Les mathématiques et le concret*, Paris, PUF.

FRECHET M., HALBWACHS M. (1924). *Le Calcul des probabilités à la portée de tous*, Paris : Dunod, 297 p.

GALTON F. (1889). *Natural Inheritance*, London, Macmillan.

GALVANI, GINI C. (1929). Di talune estensioni dei concetti di media ai caratteri qualitative, *Metron*, **8**.

GINI C. (1914). L'huomo medio, Giornali degli economiste e revista de statistica, **48**, p. 1-24.

GLAISHER J. W. L. (1872). On the Law of Facility of Errors of Observations and on the Method of Least Squares, *Memoirs of the Royal Astronomical Society*, **39**, Part II, p. 75-124.

GLAISHER J.W.L. (1873). On the Rejection of Discordant Observations, *Monthly Notices of the Royal Astronomical Society*, 33, 391-402.

GOEDSEELS P. J. E. (1902). *Théorie des Erreurs d'Observation*, Paris, Gautier-Villars.

GUILBAUD G.-Th. (1952). Les théories de l'intérêt général et le problème logique de l'agrégation, *conomie Appliquée*, **5** (4), repris dans *Éléments de la théorie des jeux*, Paris, Dunod, 1968.

GUMBEL E. J. (1934). La distribution finale des valeurs voisines de la médiane, *Comptes Rendus Académie des Sciences*, **199** (22), p. 1174-1176.

GUMBEL E. J. (1935). Les valeurs extrêmes des distributions statistiques, *Annales de l'Institut Henri Poincaré*, **4** (2), p. 115-158.

GUMBEL E. J. (1937). La précision de la moyenne arithmétique et de la médiane, *Aktuarske Vedy*, **6** (4), Prague.

HALBWACHS M. (1912). La théorie de l'homme moyen. Essai sur Quetelet et la statistique morale, Alcan, Paris.

HERTZ S. (1997). *Emile Julius Gumbel (1891-1966) et la statistique des extrêmes*, Thèse de doctorat de l'Université de Lyon-1, dir. P. Crepel.

KEYNES J. M. (1909). The Method of Index Numbers with Special Reference to the Measurement of General Exchange Value, in *Collected Writings of John Maynard Keynes*, D. Moggridge (ed), Macmillan Cambridge University Press, 1983, Vol. XI, p.49-173.

LAPLACE P.S. (1774). Mémoire sur la probabilité des causes par les évènements, *Oeuvres Compl.*, 1891, **VIII**, 27-65 et 141-153.

LAPLACE P.S. (1812). Théorie analytique des probabilités, 1ère ed. 1912. In *Oeuvres Compl.*, 1886, **VII**; 2ème ed. 1814 : préface; 3ème éd.1820 : trois suppléments, Oeuvres Compl., 1904, XIII, 98-120, Oeuvres Compl.1886, VII, 531-580, Oeuvres Compl., 1886, VII, 580-616.

MONJARDET B. (1991). Éléments pour une histoire de la médiane métrique, in Feldman *et alii*, *Moyenne, milieu, centre. Histoire et Usages*, Paris, Ed. EHESS.

NEWCOMB S. (1886). A Generalized Theory of the Combination of Observations so as to obtain the Best Result, *American Journal of Mathematics*, 6, 343-66.

NEWCOMB S. (1912). Researches on the Motion of the Moon, II, *Astronomical Papers*, 9-240, U.S. Nautical Almanach Office.

PEIRCE B. (1852). Criterion for the rejection of doubtful observations, *Astronomical Journal* 2, 161-163; reprinted in Stigler 1980 American Contributions to Mathematical Statistics in the Nineteenth Century, Vol.2, N.Y., Arno Press.

QUETELET A. (1846). *Lettres à S.A.R. le Duc régnant de Saxe-Cobourg et Gotha sur la Théorie des Probabilités appliquées aux sciences morales et politiques*, Bruxelles, Hayez.

SEAL H.L. (1967). "The historical development of the Gauss linear model", repris in Kendall and Pearson, *Studies*, 1970.

SHEPPARD W.F. (1899), On the Statistical Rejection of Extreme Variations, Single or Correlated, *Proceedings of the London Mathematical Society*, 31, 70-99, London.

SHEYNIN O.B. (1973). Mathematical Treatment of Astronomical Observations, *Arch. Hist. Exact Sci.*, 11, 109-126.

SHEYNIN O.B. (1977). Laplace's Theory of Errors, *Arch. Hist. Exact Sci.*, **17**, 1-61.

STIGLER S. M. (1973a). Laplace, Fisher, and the discovery of the concept of sufficiency, *Biometrika*, **60**, 439 p.; in *Studies*, **2**, p. 271.

STIGLER S.M. (1973b). Simon Newcomb, Percy Daniell, and the history of robust estimation, 1885-1920, *JASA*, **68**, p. 872-879, reprinted in Kendall and Plackett, Studies..., 1977, p. 410-417.

STIGLER S. M. (1978). Francis Ysidro Edgeworth, statisticien, *Journal of the Royal Statistical Society*, **A**, **141**, Part 3, p. 287-322.

STIGLER S.M. (1986). *The history of statistics. The measurement of uncertainty before 1900*, Belknap Press of Harvard University Press, Cambridge and London.

ZIZEK F. (1913). *Statistical Averages. A methodological Study*, trad. américaine W.M.Persons, New York, Henry Holt and Cy, 392 p.