



Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 1, n°1; Mars 2005

Lucien March (1859-1933) Une statistique mathématique sans probabilité?

Michel ARMATTE¹

Résumé

La carrière de Lucien March, ingénieur statisticien d'Etat entre 1900 et 1930, est éclatée entre les tâches administratives dues à sa position clé auprès de la *Statistique générale de la France*, les innovations mécanographiques, l'application des outils de la statistique de Karl Pearson aux questions démographiques et économiques, et la diffusion en France des thèses phénoénalistes et eugénistes qui les accompagnaient. Elle offre un point d'entrée intéressant pour comprendre les enjeux de la statistique mathématique du début XX^e, en particulier ses rapports difficiles avec le calcul des probabilités.

Abstract

Lucien March had a very large spectrum career of engineer and State-statistician between 1900 and 1930. It covers administration of the *Statistique Générale de la France*, use of machines in census, applications of Karl Pearson's statistical tools to demography and economics, and diffusion of Pearson's epistemology and eugenics. A study of this career is a good entrance into the mathematical statistics of the first nineteen's, especially its hard relation to the calculus of probabilities.

Il y a deux avantages à prendre la carrière de Lucien March comme fil conducteur d'une exploration de la statistique des premières décennies du XX^e siècle.

En premier lieu, le cas de Lucien March illustre parfaitement l'étonnante variété des activités et centres d'intérêt que semble réunir la profession de statisticien à cette époque. Si le personnage n'est pas inconnu, il n'est pas non plus le prototype du grand savant. Il est par contre de l'ingénieur statisticien d'État, placé au carrefour de recherches, de débats, et d'actions concernant tour à tour la mécanisation du recensement, les questions démographiques du moment entre natalisme et eugénisme, la philosophie des sciences, la méthodologie de la statistique économique (professions, chômage, salaires, prix), et le rôle des mathématiques dans cette méthodologie. Engagé à la fois dans les tâches routinières de la statistique administrative et dans des innovations fondamentales pour ces divers champs, Lucien March fait preuve d'un éclectisme qui n'est cependant pas isolé. Entre ingénierie de la production de chiffres, science administrative des "choses remarquables d'un Etat", science de la population, science de la société, discipline ancillaire de l'économie politique, et branche des mathématiques relative aux raisonnements inductifs, la Statistique n'arrête pas de chercher sa voie et de redéfinir son essence profonde, mais aussi son rôle social.

En second lieu, le cas de Lucien March illustre la position caractéristique d'une catégorie importante de statisticiens concernant l'articulation de la méthodologie statistique avec le calcul des probabilités. Dans la tradition laplacienne du début XIX^e, l'analyse statistique n'est rien d'autre que l'application du calcul des probabilités aux questions de la philosophie naturelle (par exemple la mécanique céleste) ou aux "sciences morales". Si Laplace entendait pas là principalement les questions de témoignages, de jugement des tribunaux, de décision des assemblées, et de durées de vie, ou encore de bénéfices des établissements, Poisson, Quetelet, L.A. Bertillon, et quelques autres, en avaient fait l'outil principal, via le concept d'un homme moyen, d'une analyse des sociétés en terme de causes constantes, qui permettait de fonder les sciences sociales de la seconde moitié du XIX^e. Mais à la suite des critiques de Bertrand, de l'école allemande de Lexis et Bortkiewicz, et de l'école

¹ Centre A. Koyré et Université Paris Dauphine, michel.armatte@dauphine.fr

biométrique anglaise, cette doxa identifiée par les statisticiens au modèle de l'urne invariante qui fonde le modèle binomial et la loi des erreurs, se trouve sérieusement mise à mal dans la communauté des statisticiens qu'ils soient démographes ou économistes. Dès lors deux tentations sont possibles : jeter l'enfant avec l'eau du bain en refusant tout modèle probabiliste comme définitivement trop éloigné des réalités observées, ou bien adopter de nouveaux modèles probabilistes plus généraux que le modèle normal. Le tiraillement entre ces trois attitudes vis-à-vis du calcul des probabilités – tradition laplacienne, rénovation et dépassement, ou rejet systématique – est particulièrement perceptible au tournant du siècle dans le domaine des statistiques économiques, jusque dans les années 1930, dont on sait qu'elles marquent un retour en grâce des probabilités à la fois chez les mathématiciens et logiciens grâce aux approches formelles, parfois axiomatiques, de Von Mises, Borel, Fréchet et Kolmogorov, et chez les économètres de la Cowles Commission.

1. La carrière d'un statisticien au tournant du siècle

Qui est donc Lucien March pour cumuler ainsi toutes les facettes d'une discipline qui n'est pas encore académique, et pour incarner le refus de toute une génération de voir cette discipline ravalée au rang d'application d'une branche des mathématiques, peu recommandable au demeurant? L'essentiel de ce que l'on sait sur lui, se trouve déjà exposé dans la thèse de Marietti sur la SGF, dans la notice nécrologique de Huber qui lui avait succédé, dans notre thèse [Armatte, 1995], dans les articles de Desrosières [1998] et Jovanovic et Le Gall [2001], et ne progressera guère en l'absence d'archives de la Statistique Générale de la France qui semblent bien avoir disparu à jamais. Nous nous contentons donc d'une présentation synthétique de sa biographie.

Né le 6 décembre 1859, Lucien March (prononcer Marc, comme le fait Bunle dans les entretiens accordés à A. Desrosières) est entré à l'Ecole polytechnique en 1878 et sorti dans le corps de l'artillerie de marine. Il a d'abord été ingénieur dans l'industrie mécanique avant d'être appelé en septembre 1892 à l'Office du Travail créé un an plus tôt, et dont le sous-directeur était le polytechnicien de sa promotion Arthur Fontaine (1860-1931). March est chargé du recensement professionnel couplé au recensement général de 1896. C'est à cette occasion qu'il rencontre à l'exposition universelle de 1889 l'ingénieur américain Hollerith (1860-1929) faisant démonstration de sa machine à cartons perforés initialement destinées au dépouillement du *Census* américain de 1890. Après une étude de son fonctionnement in situ lors du congrès IIS de Vienne en 1892, il fait adopter ces machines (perforatrice, trieuse, tabulatrice), et du personnel recruté en grand nombre (une centaine) par une société de droit français créée par Hollerith, pour le dépouillement mécanique du recensement des entreprises de 1896, celui de la population continuant à passer par les communes.

La Statistique Générale de la France, fondée par un arrêté du 3 avril 1840 sur la base d'un Bureau de statistique du Ministère du commerce mis en place par Thiers en 1833, a été dirigée successivement par Moreau de Jonnés, Legoyt (1852), Challot (1871), Toussain Loua (1875), Victor Turquan (1887), et Blancheville (1896). Elle fut ensuite placée sous l'autorité du directeur de l'Office du Travail, puis intégrée à celui-ci en août 1899 dans une Direction du Travail dirigée par A. Fontaine au sein du Ministère du Commerce, et installée dès 1901 dans les nouveaux locaux de l'avenue Rapp². Deux statisticiens adjoints, Huber et Risser, sont recrutés en 1901, et March devient le chef des services techniques de l'Office et de la SGF. Pour le dépouillement du recensement de 1901, il écarte les machines Hollerith, pour une question de coût, et aussi d'inadaptation à la décentralisation, mais récupère pour la SGF le

² Voir [Marietti 1947] et [Huber 1937]

personnel mis à sa disposition quatre ans plus tôt par Hollerith³. Il met au point une nouvelle machine, le classi-compteur dont le principe est plus simple et assez différent : pas de stockage de l'information dans les cartes perforées, ce qui est une régression vu d'aujourd'hui, mais une économie de temps à l'époque, et saisie des données des bulletins sur un clavier dont chacune des 60 touches est affectable à une modalité de réponse, et fait avancer d'une unité le compteur associé. Les compteurs se transforment en imprimante sur une feuille qui prend la trace de leurs valeurs. Il obtient également que les opérations ne passent plus par les communes et les préfectures mais soient entièrement centralisées à la SGF.

En 1906 et 1907, les services techniques de Lucien March sont pérennisés par un rattachement direct à la Direction du Travail, elle-même logée au nouveau Ministère du Travail. La SGF est réorganisée dans ce cadre avec un effectif statutaire de "2 statisticiens, 3 statisticiens adjoints⁴, 1 chef de travaux, 4 secrétaires comptables, 9 contrôleurs, 6 vérificatrices, 6 garçons, 1 concierge"⁵ et Lucien March en devient le chef (1907) puis le directeur (1910-1920) alors qu'elle prend son autonomie par rapport à la Direction du Travail.

Cette période correspond à un nouvel engagement de March dans des travaux de statistique économique, via le *Comité d'études relatives à la prévision des chômages industriels*, créé par un arrêté de juin 1911, auprès de la SGF, présidé successivement par Levasseur et Cauwès, puis Albert Thomas. Il fait suite à une première commission⁶ établie en 1905 et chargée de préparer un programme d'enquêtes annexées au recensement de la population, puis à une Commission des crises établie en 1908 et présidée par A. Picard⁷. Ce *Comité* avait demandé qu'un service d'observation des prix soit annexé à la SGF. Inscrit au projet de budget pour 1913, celui-ci ne reçut son statut et ses financements qu'après guerre, en octobre 1917. La première guerre mondiale entraîne en effet une rupture des crédits de la SGF, et une mobilisation des intellectuels sur de nouvelles tâches de réorganisation de la production, en particulier des mathématiciens Paul Painlevé (Ministre de la Guerre) et Émile Borel (secrétaire général de la Présidence du Conseil), de l'historien Simiand et du sociologue Halbwachs membres du cabinet du ministre de l'Armement Albert Thomas. Le service des prix, chargé de suivre les indices économiques et d'informer le gouvernement de ce qu'ils annonçaient, a joué un rôle important après guerre, dans la définition du programme de reprise de l'activité économique. A lui seul il comportait davantage de statisticiens que la SGF dix ans plus tôt. En 1919, à la veille du départ en retraite de March⁸, la SGF comprenait statutairement 43 personnes (dont 4 statisticiens et 6 statisticiens adjoints) et le Service d'observation des prix en comprenait 14 (dont 2 statisticiens et 3 statisticiens adjoints)⁹.

³ Voir [Peaucelle 2004] qui s'intéresse principalement aux coûts comparés de ces innovations

⁴ C'est à ce moment que furent recrutés les statisticiens adjoints Lenoir, Dugé de Bernonville, Bunle et Ville-Chabrol.

⁵ [Marietti 1947] p. 35.

⁶ Cette première commission présidée par Foville, l'ancien ministre Pierre Baudin, regroupe, outre 3 sénateurs, 2 députés, et 4 autres politiques, les statisticiens-économistes J. Bertillon, F. Faure, A. Fontaine, de Foville, Y. Guyot, C. Laurent, E. Levasseur, L. March.

⁷ [SGF 1913]

⁸ Lucien March quitte la Direction en 1920 pour être remplacé par Michel Huber. La SGF sera directement rattachée à la Présidence du Conseil par un décret de 1930. En 1942 elle est absorbée par le SNS fondé par Carmille. En 1946 le SNS devient l'INSEE.

⁹ Vingt ans plus tard Carmille crée un Service de Démographie d'une toute autre ampleur (925 emplois) et doté de plusieurs ateliers de mécanographie, qui absorbera la petite SGF en 1941, et consacra la victoire provisoire

Cette carrière d'une vingtaine d'années à la tête de la SGF fit de Lucien March le principal animateur des débats sur la statistique économique dans différents lieux savants. C'est d'abord la *Société de Statistique de Paris* (SSP) : il y entre en 1897 et la préside en 1907. C'est ensuite l'Université : il est le principal fondateur en 1922, avec Borel, Huber, et Darmais de *l'Institut de Statistique de l'Université de Paris*, placé sous la tutelle des quatre Facultés, et qui forme chaque année moins d'une dizaine d'étudiants, ce qui est peu, mais c'est la seule formation à la statistique mathématique, à côté des enseignements de Levasseur au Collège de France, de Liesse au Conservatoire des Arts et Métiers, puis de Divisia au CAM, à l'École polytechnique et aux Ponts et Chaussée, lesquels constituent plutôt une introduction au traitement de données économiques et sociales¹⁰. Il est surtout le représentant de la statistique officielle française à *l'Institut International de Statistique* (IIS) dont il fut membre en 1901. Principal organisateur de la session de Paris en 1909, une rencontre qui devait jouer un grand rôle dans la diffusion des méthodes nouvelles de l'école biométrique londonienne (Galton et Pearson), il est avec ses partenaires anglais, le responsable d'un double transport de la corrélation et de la régression vers le continent et vers le milieu des économistes¹¹. Ce rôle de passeur est accentué par le fait que March soit le premier français à bien connaître les travaux de Karl Pearson dont il adopte (et adapte, parfois de manière critique) les trois principaux ingrédients : les outils d'analyse statistique, exposés par Pearson dans les *Contributions mathématiques à la théorie de l'évolution* entre 1895 et 1906, – principalement la méthode des moments, le système de courbes de distribution, et la corrélation -, mais aussi comme nous le verrons, le programme eugénique, et la philosophie des sciences¹².

2. March et la méthodologie statistique

2.1. Distributions

Les premières études de March, en 1898, se proposent de chercher des fonctions mathématiques à ajuster sur des distributions empiriques de salaires qu'il souhaite comparer, et qui proviennent de diverses enquêtes¹³. Pourquoi les ajuster? Pour pouvoir les résumer par quelques valeurs caractéristiques, et la plus importante selon lui est le mode (qu'il appelle valeur normale) et que l'on obtient comme maximum de la densité. Lucien March choisit d'ajuster une forme sophistiquée de la courbe de Pareto. On en rappelle rapidement les caractéristiques.

La loi des revenus de Pareto apparaît pour la première fois dans le *Giornale degli Economisti* en 1895, puis dans un recueil publié par l'Université de Lausanne l'année suivante, et enfin dans le premier chapitre du livre III du *Cours d'Economie Politique*. Exploitant des données anglaises de Giffen de 1843 et 1879-80, ainsi que des données italiennes (Benini), prussiennes (Soetbeer, 1876-86), saxonnes (Ammon), bâloises, et parisiennes, Pareto propose de représenter dans un plan cartésien les logarithmes des revenus x en fonction des logarithmes du nombre $N(x)$ des revenus supérieurs à x . Or dit-il 1°) *Nous sommes de suite frappés du fait que les points ainsi déterminés ont une tendance très marquée à se disposer en ligne droite (...)* 2°) *Nous observons que les lignes droites ainsi tracées font avec l'axe des*

des "renseignements individuels continus" sur les "dénombrements anonymes et globaux", comme le dira Alfred Sauvy.

¹⁰ Sur l'ISUP, Voir [Pressat 1987]

¹¹ Voir les articles de Yule, Bowley et Edgeworth dans le BIIS de 1909.

¹² Voir [Porter 2004] ou notre note de lecture sur cet ouvrage dans Mathématiques et Sciences Humaines, 2005.

¹³ Une enquête de la chambre de commerce de Paris de 1860, une enquête de l'Office du travail de 1891-93, une enquête allemande sur les mineurs de Dortmund, et une enquête américaine du Département fédéral du Travail.

abscisses des angles qui ne sont pas très différents l'un de l'autre.(...) Nous nous trouvons ici en présence d'une loi naturelle, qui nous révèle une tendance des revenus à se grouper d'une certaine façon, au moins dans les sociétés et pour les époques considérées. (...) Ces résultats sont très remarquables. Il est absolument impossible d'admettre qu'ils sont dus au hasard. Remarquant que la pente de la droite ajustée peut être prise pour mesure de l'inégalité de la distribution, Pareto fait de son ajustement l'outil permettant de trancher le débat entre libéraux et socialistes sur la question de la paupérisation croissante des masses laborieuses, en faveur des premiers évidemment. Un de ces émules, Otto Ammon, se saisira de la loi de Pareto pour en établir des versions plus sophistiquées, et pour en faire le chaînon manquant permettant de fonder l'inégale distribution des revenus dans les sociétés industrielles sur la répartition naturelle normale des capacités humaines. Cette naturalisation des inégalités sociales est à l'oeuvre chez plusieurs économistes, comme par exemple André Liesse qui consacre son ouvrage sur *le Travail* (1901) à une refondation de l'économie sur un tel principe.

Chez Pareto, l'ajustement est totalement empirique et ne dérive d'aucun principe générateur de type probabiliste. Dans l'addition au cours de 1897, convaincu par les thèses de Lexis, il affirmera que s'il faut chercher un modèle, celui-ci doit être à plusieurs urnes, mais le modèle de combinaison de lois binomiales dont il accouche ne le satisfait pas, parce que pour lui, c'est la formule empirique qui précède la formule théorique, comme Kepler a précédé Newton, et qui doit concentrer d'abord notre attention. D'autre part Pareto rejette l'usage de la méthode des moindres carrés pour ajuster la droite sur le nuage de points. Dans un papier de 1909, il en donnera quatre justifications. 1°) Contrairement au cas de la théorie des erreurs de mesure, on n'observe pas n fois le même objet donc les écarts au modèle ne sont pas des erreurs 2°) la forme de la fonction f n'est pas connue a priori; 3°) les écarts (qui ne sont pas des erreurs) ne suivent pas une loi normale 4°) Dans le cas d'une variable transformée par le logarithme la méthode des moindres carrés minimise des erreurs relatives et pas des erreurs absolues sur la variable observable, donc ces dernières peuvent être très importantes. Dans un cas comme celui de la courbe des revenus que Pareto qualifie de problème d'interpolation – terme permettant de comparer plusieurs méthodes y compris celle des moindres carrés sans référence aux distributions de probabilités - il en conclut qu'il est préférable d'utiliser la méthode de Cauchy.

Dans son étude de 1898, Lucien March s'inscrit dans cette tradition. Il choisit comme formule de départ une formule empruntée dit-il à Otto Ammon, que l'on appelle parfois la 3^{ème} loi de Pareto, de densité telle que $\log y = A + B \cdot \log x + Cx$. Elle lui est également inspirée par le système de courbes de Pearson (type III). March réalise 3 ajustements de cette formule, par la méthode des moindres carrés, par la méthode des moments de Pearson, par la méthode de Cauchy, et conclut aisément que les deux dernières méthodes l'emportent. Ces conclusions insistent fortement sur l'intérêt de l'approche de Pearson, et confirment totalement ce que dit Pareto de la supériorité de la méthode de Cauchy sur celle des moindres carrés. Il est important de noter que l'approche de l'ajustement d'une distribution par ces trois auteurs s'oppose assez radicalement à celle que développent au même moment d'autres statisticiens. Dans une série d'articles (1883, 1892, 1896) de la même période, Edgeworth met en place une modélisation des distributions asymétriques par une loi des erreurs généralisée qui est un développement en série de la loi multinomiale, dont la loi normale est une première approximation. Il remarque que cette loi des erreurs généralisée peut se déduire par un simple changement de variable – une *translation* dit-il – de la densité normale, et il cherche à généraliser cette méthode de translation pour rendre compte de données comme celles des salaires :

"Je propose de traiter les groupes asymétriques d'observations comme s'ils étaient généralisés de la manière qui vient d'être décrite, comme si chaque membre du groupe observé

était une certaine fonction du membre d'un groupe normal; et de représenter le groupe par une formule qui, selon cette hypothèse, constitue la seconde approximation de la vraie distribution. Ce n'est pas que cette hypothèse soit universellement bonne, ni que la formule soit toujours la plus appropriée, mais elle est parfois la formule, et en générale une bonne formule."

"(...) La thèse principale de ce papier est que la préférence doit être donnée à des formules ayant une certaine affinité avec la loi normale, vues comme si elles étaient produites par quelques modification des conditions de convergence de la loi normale : ces conditions sont au nombre de deux : la coopération d'un nombre infini d'agents indépendants, et la petitesse des effets de chaque agent sur le résultat total. (...) Dans la recherche d'une base a priori pour une loi générale de fréquences, je me sépare des distingués statisticiens qui regardent la précision de l'ajustement comme le seul test d'une formule représentative. Ainsi le professeur Pareto, inventeur d'une belle et utile représentation de la fréquence de revenus de différentes grandeurs, semble opposé à l'idée même d'une courbe de probabilité généralisée.¹⁴"

Edgeworth et Bowley publient en 1902 une étude dans laquelle cette loi des erreurs généralisée est ajustée à des données de salaire. L'équation de translation n'est pas ajustée par les moindres carrés, mais par une méthode de médiane généralisée. Bowley, tout autant l'élève de Pearson que d'Edgeworth, réussit d'ailleurs l'exploit de ne pas évoquer les moindres carrés dans son traité, pas plus dans son long chapitre sur les méthodes d'interpolation que dans celui qui traite de l'ajustement, ou encore de la corrélation. La méthode n'apparaîtra que dans un dernier appendice de la quatrième édition.

2.2. Corrélations.

Un second exemple de recherches statistiques que March va développer est son travail sur "la comparaison de courbes statistiques" [March 1905]. Il y construit une lignée de coefficients de concordance entre deux séries temporelles sur la base de la notion de concomitance que l'on trouve dans la *Logique* de Mill, et d'un simple comptage des concordances et discordances emprunté à Fechner. Le coefficient de corrélation empirique du "product moment" de Pearson¹⁵ est situé dans cette lignée plus générale des coefficients de concordance et du problème nouveau qui se pose aux économistes de ce qu'il appelle la covariation de deux séries temporelles. La reconstruction de la corrélation en covariation est bien un emprunt à Pearson, mais un emprunt réapproprié qui s'accompagne 1°) d'un rejet de tout lien avec la notion de distribution normale – que le coefficient r de Pearson soit d'abord et avant tout un estimateur du paramètre ρ de la forme quadratique de la densité de la loi normale à deux dimensions n'intéresse pas March – et 2°) d'une affirmation de la spécificité des corrélations calculées en économie, à savoir la dimension temporelle des séries, laquelle constitue en quelque sorte leur élément commun, la source essentielle de leur liaison. Cette dimension temporelle explique d'ailleurs la réticence des économistes à passer du graphique temporel où sont superposées les évolutions de deux séries x et y au graphique cartésien dans lequel le nuage de points (x_i, y_i) révèle la liaison des variables sous condition d'une élimination de la variable temps. Elle explique aussi que de nombreux économistes – et March en est - pensent que le coefficient de corrélation ne se suffit pas à lui seul, et qu'il ne

¹⁴ [F.Y. Edgeworth 1898 et 1899].

¹⁵ Le coefficient de corrélation entre x et y est égal au rapport de la covariance entre x et y au produit des écarts-types de x et y .

prend sens qu'associé au graphique temporel. March traite dans le *JSSP* (1904-5) des représentations graphiques et propose à l'IIS une normalisation de celles ci qui permette des usages comparatifs. La même dimension temporelle des données économiques suggère aussi une série de traitements préalables de ces données avant de calculer une corrélation : on peut par exemple décomposer les séries en tendance et en cycles, ce que proposera Hooker en 1901, et la corrélation des séries en tendance ou en écarts à cette tendance ont en général des valeurs et des interprétations différentes. Il en est de même si l'on considère les variables en niveau ou en différences (ou accroissements), voire en taux de croissance (accroissements relatifs) et March [1928], très soucieux de pointer les risques d'une simple transposition des outils biométriques à l'économie, s'attachera à distinguer ce qu'il appelle *coefficient de covariation différentielle* et *coefficient de covariation tendancielle*. La notion de corrélation que March importe de l'école anglaise est donc totalement reconstruite dans un double souci de la décrocher totalement de tout contexte probabiliste et de coller au plus près des données observées et des significations spécifiques qu'elle prend dans le champ économique.

2.3. Indices et baromètres.

Dans le cadre du service d'observation des prix créé en 1917, March est chargé de la confection des *Indices du mouvement général des affaires*, édités d'abord en recueil par la *revue politique et parlementaire*, et quelques années plus tard placés sous le contrôle scientifique et financier de l'ISUP. Les *Indices* constituent une sorte de Baromètre économique à la française, pendant de ceux de Harvard (Persons), de Louvain (Dupriez), de Londres (Beveridge), de Berlin (Wagemann) et de Moscou (Kondratieff). Le recueil de ces *Indices*, publiés jusqu'en 1939, renferme chaque mois dix planches grand format. La première décrit graphiquement les évolutions de trois indicateurs (type A,B,C de Harvard¹⁶) pour la France le Royaume Uni et les Etats-Unis. Les autres planches, à l'exception des deux dernières consacrées au commerce et à la finance anglaise, représentent une cinquantaine d'indicateurs français (finance, commerce extérieur et intérieur). Un texte de deux pages accompagne ces planches, constitué d'un commentaire (variable) de la conjoncture et d'une note explicative sur les conditions de productions de ces indices. Techniquement l'élaboration des indices est assez simple : les indices sont exprimés en "valeur relative par rapport à une moyenne mensuelle calculée durant la période 1904-1913", ce qui n'élimine ni la tendance ni la saisonnalité; d'autre part, ils ne résultent pas de regroupements mécaniquement opérés sur la base des corrélations observées; ce ne sont le plus souvent que simples ratios, moyennes ou indices budgétaires.

March ne se contente pas de publier ces indices, il en discute les fondements théoriques (dans la théorie quantitative de la monnaie par exemple) et les principes de construction [March 1922, 1923, 1924, et 1927]. Sa conception des baromètres est intéressante à noter pour notre propos : il rejette à la fois l'indice unique qui résumerait l'ensemble du mouvement des affaires (comme Juglar le fera par exemple avec les encaisses de la banque de France), et l'indice synthétique élaboré comme moyenne pondérée de plusieurs séries. Il préfère juxtaposer les indices plutôt que les combiner. C'est que la moyenne n'a de sens pour lui que si elle prend bien en compte une totalité. Il s'oppose à l'interprétation probabiliste des indices, en terme directement dérivés de la théorie des causes constantes de Laplace, Poisson et Quetelet : puisqu'un indice des prix est une moyenne (pondérée) de prix relatifs, ces moyennes opèrent une compensation des effets des causes accidentelles et spécifiques à chaque bien, et ne retiennent que les effets des causes constantes

¹⁶ Les courbes A, B, C représentent respectivement le marché financier, le marché monétaire et le marché des produits industriels.

qui sont à l'œuvre, c'est-à-dire la variation de valeur de la monnaie. Cette théorie, défendue, sous certaines conditions, par Edgeworth dans 3 memoranda de 1887-89, l'amenait à préciser ce que devaient être les pondérations pour un indice monétaire. Mais elle est battue en brèche par plusieurs auteurs dont Keynes [1909] qui souligne que les conditions d'un transfert de la loi des erreurs ne sont pas réunies : les prix des biens ne sont pas indépendants, pas de même importance, et de plus la probabilité n'est pas pour lui une limite de fréquences : "*nous devons rejeter les canons de pondération du Professeur Edgeworth parce qu'ils sont fondés sur une analogie partiellement fautive avec la théorie des erreurs. (...) La simplicité apparente de la méthode est illusoire, et occulte la quantité importante d'information empirique que nous devons posséder [sur la loi de variation de la valeur d'échange] pour pouvoir l'appliquer avec succès. Il n'y a pas de méthode probabiliste a priori valable en toute circonstance pour la construction des nombres indices.*" Après 1914, l'inflation prit en Europe une importance très inquiétante. La question des pondérations fut à nouveau débattue et Irving Fisher entreprit une étude systématique de toutes les formules d'indices dans deux ouvrages [Fisher 1911 et 1922] d'où il ressortait finalement que la question des pondérations n'était pas aussi déterminante que celle de la précision des relevés de prix. L'approche de l'indice monétaire par L. March est plus proche de celle de Keynes, et plus tard de Divisia, et très critique par rapport à la justification probabiliste : parce que la moyenne peut cacher "des accumulations dissimulées", parce que les prix ne sont pas indépendants et obéissent à des lois de substitution, parce que l'indice ne prend pas en compte toutes les transactions, parce que sa forme varie suivant le temps et le lieu¹⁷.

3. La philosophie des sciences de Pearson et March.

Ce souci de développer une statistique qui soit au plus près des phénomènes vient d'une philosophie que March partage avec Pearson. C'est celle de la *Grammaire de la Science*, un ouvrage de Karl Pearson dont il traduit et préface la 3^{ème} édition en 1912. Cette préface ne laisse aucun doute sur son adhésion aux thèses phénoménaliste de Pearson et à la primauté de la notion de contingence : la science traite principalement de phénomènes qu'elle décrit, classe, et résume mais elle ne saurait se prononcer sur la chose en soi :

"L'homme de science ne met donc en œuvre pour ainsi dire que les numéros d'un catalogue d'impressions sensibles différenciées, rapports concrets ou idées abstraites. Entre ces rapports et ces idées il établit des relations nouvelles qui s'imposent à l'esprit avec d'autant plus de force qu'elles sont mieux en harmonie avec sa propre constitution. La certitude ou l'évidence scientifique résulte de l'accord de la pensée avec elle-même, soit dans l'exercice de sa faculté d'identifier, soit dans l'exercice de son pouvoir d'abstraction et de généralisation. Il n'est de certitude scientifique que dans le monde des concepts; non seulement le noumène, mais encore la perception sont en dehors du certain.

L'homme de science n'affirme ni ne nie la réalité du monde extérieur; il n'affirme ni ne conteste non plus la légitimité d'idées dont l'origine serait hors du fond commun des impressions sensibles immédiates ou accumulées. Il déclare simplement que, pour lui, l'envers des phénomènes et l'intuition spontanée, - même si celle-ci, comme on l'a supposé, est le fruit

¹⁷ Pour être complet, il faudrait aussi signaler les recherches empiriques sur les liens prix-salaires entreprises par le Directeur de la SGF [March/SGF 1910 et March 1922].

des acquisitions de l'espèce, - sont inaccessibles à l'intelligence, sont irréductibles aux rapports rationnels que peut former la pensée exercée"¹⁸

L'agnosticisme que March partage avec Pearson quant à la nature même des choses se conjugue avec un rejet de la nécessité et de la causalité, notions limitées qui n'ont plus qu'une place secondaire dans le discours scientifique, et que doit remplacer la notion bien plus générale de contingence : "*Puisqu'il n'est point de nécessité hors de la sphère des conceptions, les lois naturelles n'expriment que des contingences (...) Entre les changements de la cause et de l'effet il y a simplement accord statistique plus ou moins étroit : le professeur Pearson a consacré d'importants travaux aux moyens de mesurer les degrés de cet accord.*" En d'autres termes le coefficient de corrélation et la notion de régression sont les instruments de mesure de cette contingence des phénomènes à quoi se ramène toute science. Seule cette philosophie pouvait donner une telle importance à la statistique dans le concert des sciences. L'irruption de la contingence comme fondement des sciences d'observation est le point essentiel qui permet, comme March l'a bien compris, un renforcement mutuel du point de vue épistémologique et de la pratique scientifique : il faut bien mesurer ce que cette pensée philosophique doit à la statistique (les notions de classes d'événements, de similitude, de contingence) mais aussi la force que cette philosophie peut conférer à la pratique de la statistique, et la manière dont elle peut orienter son développement. Mais la contrepartie est sévère : il faut abandonner toute prétention à l'analyse causale. March y est prêt, et, en accord avec Pearson, il critique ou ignore ceux qui, comme ses contemporains Moore (1914) ou même Lenoir (1913), interprètent les corrélations en terme de causalité, et posent les premières pierres d'une économétrie structurelle faite de relations stables entre grandeurs macro économiques.

L'eugénisme est aussi un point de vue, et un programme d'action, qui donne sens à la statistique démographique, en même temps que celle-ci permet d'argumenter des discours eugénistes. March en épouse les thèses, au moins à partir de 1912, date du congrès international de Londres auquel il participe, avec Huber, Barriol et Faure, mais plus vraisemblablement avant cette date à l'occasion de ses premières recherches démographiques et de ses lectures des travaux de Pearson. En tout cas sa conviction est suffisante pour qu'il prenne l'initiative dès l'année suivante de la fondation d'une Société française d'eugénique dont il fut le trésorier archiviste. Anne Carol¹⁹ a montré cependant que l'eugénisme français est davantage une création des médecins hygiénistes, des spécialistes de la puériculture ou de "l'hominiculture" (titre d'un ouvrage de Jules Amar repris par le doyen Landouzy), que des statisticiens. La liste des 104 membres de la société publiée dans le premier numéro de la revue *Eugénique* comprend 65% de médecins. La société ne réussira pas à s'adosser à un laboratoire de recherche comme l'a fait Galton, et dès le milieu des années 1920 ses effectifs baissent, l'obligeant à se fondre dans la société d'anthropologie. March restera le plus actif des membres, avec 12 articles sur une trentaine publiés dans *Eugénique*, et c'est encore lui qui mènera la délégation française (avec Vacher de Lapouge et Cuenot) au second congrès international d'eugénique à New York (1921). Entre Bertillon père, auteur d'une *Démographie figurée de la France* et de quelques fameux articles du *Dictionnaire* de Dechambre ("moyenne", "taille", "mariage"...), et Bertillon fils, auteur d'un *Traité de démographie* incorporé à son *Cours élémentaire de Statistique* (1895) et fondateur d'une *Alliance nationale pour l'accroissement de la population française*, puis Alfred Sauvy et Alexis Carrel à la fin des années 30, March est certainement un maillon important d'une école démographique

¹⁸ [Pearson 1912], note du traducteur (L. March).

¹⁹ [Carol 1995]

française. Mais, comme l'a montré Desrosières, son attirance pour l'eugénisme est modérée par une vision nataliste partagée par la plupart des démographes français, et qui se démarque du modèle anglais : les français sont plus inquiétés par une baisse générale de la natalité que par les dégâts de la fécondité différentielle.

4. Lucien March et le calcul des probabilités

L'approche de la méthode statistique que développe Lucien March peut à maints égards être caractérisée de mathématique. Par opposition avec les analyses que l'on trouve dans le JSSP avant 1910, qui s'appuient principalement sur tableaux, moyennes et graphiques, le changement est quantitatif – la théorie statistique ne représente pas 10% des articles du JSSP dans la première décennie du XX^e - et qualitatif : de nouveaux outils mathématiques sont mis en œuvre comme les ajustements, la corrélation et la régression. Cependant cette statistique mathématique n'est pas celle que vont développer Darmon, Borel, Lévy et Fréchet deux décennies plus tard : non seulement, comme nous l'avons vu, elle privilégie les observations et ne trouve aucun principe pour se construire en théorie, mais elle ne s'appuie absolument pas sur le calcul des probabilités. Les premiers ouvrages du calcul des probabilités qui succèdent au traité de Bertrand [1889] sont ceux de Poincaré [1896], Borel [1909], de Laurent [1908], de Carvallo [1912], de Bachelier [1912] et de Keynes [1921]. Ils ne sont pas cités par Lucien March. Et ce n'est pas un simple oubli. Le rejet de la probabilité chez March est totalement explicite dans ses textes de méthodologie.

Dans son discours d'investiture comme président de la SSP, March [1907] affirme :
"[Aujourd'hui] on substitue au moins implicitement la notion de tendance à celle de loi, celle de fréquence à celle de probabilité. L'écart moyen le plus souvent utilisé est calculé sur les observations elles-mêmes et non d'après un rapport hypothétique de probabilité, (...) ou bien on introduit de nouveaux concepts comme celui de stabilité" [s'en suit une référence précise mais non explicite aux travaux de Dormoy et Lexis] (...) La forme classique de la courbe en cloche ne convient pas pour les phénomènes qui ne se développent pas symétriquement autour d'une valeur moyenne. "

Il publie en 1908 dans le JSSP un court article sur *la terminologie en statistique*, d'après une intervention au IV^e congrès des sciences mathématiques, dans lequel sa position vis-à-vis de la probabilité se fait très claire :

"Une différence essentielle sépare la fréquence observée en statistique et la probabilité mathématique. En statistique nous ignorons les circonstances initiales des faits observés; notre connaissance peut tout au plus s'étendre à quelques parties de l'enchaînement intermédiaire entre les conditions originelles et le résultat (...) Dans le schéma des probabilités les combinaisons et répétitions qui contiennent en puissance le résultat final sont complètement connues; seul le jeu du déclenchement qui fait apparaître certaines combinaisons demeure imperceptible en raison de l'exiguïté de son action. L'accord d'une distribution de fréquences, observée en statistique, avec une distribution de probabilités n'implique donc qu'une analogie apparente entre l'enchaînement des faits statistiques et la formation des probabilités."

(...) Aussi serait-il opportun, en statistique, de renoncer à l'emploi du mot probabilité pour exprimer l'attente que fait naître la constatation d'une fréquence; car si dans la théorie des probabilités, la convention sur laquelle repose cette attente inspire une parfaite confiance, en statistique le degré de confiance que mérite cette attente est souvent modifié par l'étude des liaisons des faits, par les enseignements des sciences sociales.

D'ailleurs, les applications du calcul seraient à peu près aussi commodes si l'on opérait sur des fréquences au lieu d'opérer sur des probabilités; les énoncés des propositions

devraient être transformés, mais ils seraient plus rigoureux et éviteraient bien des critiques adressées aux anciens énoncés.

Par exemple lorsque Laplace contestait le caractère accidentel des écarts du coefficient de natalité masculine observé à Paris, à Londres et à Naples, en assimilant la détermination du sexe à un tirage au sort, il s'exposait aux critiques que Bertrand a adressées à son énoncé.

(...) Dans les applications à la statistique il ne paraît pas suffisant de substituer à la notion de probabilité a posteriori celle de probabilité statistique que suggère Blaschke. Pour que les personnes insuffisamment averties ne se méprennent pas sur la portée de ces applications, il semble préférable d'écarter complètement de l'analyse des résultats statistiques le terme probabilité et le mot probable entendu dans le sens mathématique."

En 1907, la session de l'IIS à Paris est l'occasion d'une confrontation intéressante entre les positions de Lucien March, Wilhelm Lexis et de Francis Edgeworth sur cette question, ce dont témoigne le compte rendu du BIIS paru en 1909. Lexis affirme (p.55) que *"Le plus utile est de ramener les phénomènes à la loi de Gauss ou à des lois dérivées et non de chercher à construire des formules empiriques beaucoup moins intéressantes. Le fondement le plus solide doit toujours rester la loi de Gauss qui est une loi de la nature"*. Edgeworth renchérit et introduit son rapport en affirmant que *"la loi de Gauss est toujours la loi classique idéale de la statistique, même si elle n'est pas toujours exactement vérifiée dans la nature"*(p.92) . Un peu plus loin, il a cette belle phrase souvent citée : *"les probabilités ne se rattachent pas seulement à la statistique comme l'arbre à la vigne dont il supporte les branches; il y a aussi un entrelacement de racines, une liaison des premiers principes, qu'il est difficile de disséquer"*. Dans une autre déclaration comme président de séance (p. 132) il dit :

"La loi de Gauss est d'une importance capitale parce qu'elle est la loi des choses soumises à un mélange complexe de causes indépendantes. C'est ainsi qu'elle régit le monde des atomes. Là on ne peut la modifier, pas plus qu'on ne peut, dans l'expression de la loi de gravitation, remplacer le carré par le cube parce que le carré est la réalité. La loi des erreurs est aussi générale que la loi de la gravitation. (...) Si la loi des erreurs n'est pas toute la science statistique, c'en est une partie, une partie essentielle."

Le papier qu'il présente reprend ensuite systématiquement les apports du calcul des probabilités qui passent par la loi normale, la loi des erreurs généralisées dont nous avons déjà parlé, la loi log-normale de Galton-MacAllister, la théorie probabilité de l'indice monétaire, et les différentes méthodes d'ajustement, pour conclure *"c'est se méprendre que de supposer que les derniers développements des probabilités ont discrédité la loi des erreurs"*.

Dans la même session, March introduit son rapport en affirmant :

"Deux genres de théories s'opposent actuellement, l'une rattachant toutes les distributions à la loi des erreurs de Gauss, l'autre, avec Pearson, cherchant des lois où interviennent des probabilités variables. Il ne me semble pas qu'on doive condamner aucune de ces tendances".

Il développe ensuite de nombreux exemples d'ajustements qui prouvent que *"la loi de Gauss semble au premier abord peu en harmonie avec les faits naturels"*. Il rejette clairement les calculs d'erreur probable sur des résultats, et l'emploi des mots "probabilité" et "corrélation", mais pas l'usage des procédés mathématiques. De tels procédés opèrent, dit-il, *"comme un crible logique approprié à l'analyse et la synthèse des rapports numériques qui constituent le principal objet de la statistique"* et dont il donne 3 exemples : les indices, les enquêtes partielles par la méthode représentative, et la comparabilité des graphiques. Il obtiendra qu'un *Comité des méthodes de comparaison statistique* soit institué à l'IIS pour

traiter de ces sujets. Le rejet des probabilités n'est pas pour lui synonyme d'un rejet des mathématiques. Il a d'ailleurs proposé plusieurs fois un nouveau terme pour cette mathématique de la statistique, celui de *pléthométrie* qui figure pour la première fois dans un texte publié en 1910 sous le titre "*du mode d'exposer les principaux éléments de la théorie statistique*". "*Le principe de compensation, sur lequel est basée toute comparaison de faits collectifs, [et qui est] la généralisation de la formule du marché équitable que les hommes ont sans doute adoptée depuis qu'ils vivent en société*", lui semble une base suffisante pour fonder la statistique; en commençant par la moyenne, qu'il définit comme la grandeur qui, répétée n fois, fournit le total des quantités. Par marché équitable, il entend la condition que le total des quantités livrées soient égal au total des quantités demandées, quelles que soient les valeurs des livraisons partielles. Drôle de renversement qui consiste à fonder la statistique sur un principe économique, qui reprend donc l'origine de l'espérance mathématique dans les contrats aléatoires, en niant que cela se rapporte à la notion de probabilité. Supposant ensuite que chaque valeur observée est une valeur possible extraite d'une "distribution primaire" – nous dirions "la réalisation d'une variable aléatoire" – il retrouve toute la théorie de l'échantillonnage et de la décomposition de variance (ou "fluctuation") sous la forme d'un "principe de conservation" des moyennes et fluctuations divisées par n dans la transformation des distributions primaires en distribution secondaire, y compris les propriétés asymptotiques, la distribution normale des erreurs considérées comme somme d'erreurs élémentaires uniformément réparties, la typologie des courbes de Pearson, les droites de régression et la covariation ²⁰. Le défi est donc relevé : on peut exposer toute la statistique mathématique en un cours de 40 pages, sans un mot sur la probabilité.

Le principe exposé ici sera repris dans la seconde partie de son ouvrage testament, construit sur la base de ses cours à l'ISUP²¹, et publié en 1930 : *les principes de la méthode statistique*. Un ouvrage de 800 pages, d'une richesse incomparable puisqu'il aborde en plus des sujets classiques de l'analyse statistique, les méthodes de la statistique administrative (première partie), et des applications de la statistique assez originales à la conjoncture et aux affaires (marchés, stocks, contrôle de fabrication, coûts, bilans). Mais ce livre arrive trop tard aussi bien dans la vie de March que dans celle de la discipline. La statistique des années 1930 aura renoué avec d'autres sources, celles du calcul des probabilités renouvelé par l'école française, mais aussi celles de la statistique géométrique de Fisher.

Si l'on doutait encore de la pensée de March sur la probabilité, cet extrait du compte-rendu qu'il donne du traité de Bachelier en 1912 enfonce le clou :

"Pour ce qui intéresse la statistique, on peut poser qu'en principe la théorie de cette science n'a point nécessairement à invoquer le calcul des probabilités. Celui-ci est fondé sur des conditions, soit connues a priori, soit impliquant une suffisante continuité difficilement assimilable aux conditions fondamentales des faits d'observation dont traite la statistique (...) Mais le calcul des probabilités et les applications aux jeux de hasard constituent par la difficulté même des raisonnements et par l'attention qu'ils exigent une excellente gymnastique de l'esprit"

²⁰ Cela revient dit-il à considérer une armée de s régiments (groupes de distributions primaires) dont chacun comprend n compagnies, et à prendre un homme par compagnie pour faire l'échantillon qu'il appelle distribution secondaire.

²¹ Il est intéressant de noter ici qu'après 1933, le cours de "méthode statistique (éléments)" de March est repris par Huber puis Dugé de Bernonville, tandis que celui de "statistique mathématique" de Darmon (et assuré deux années par Borel) change radicalement de présentation en donnant toute sa place au calcul des probabilités (voir le tableau comparatif des contenus de ce cours donné par [Pressat 1987] p.25.

On retrouve dans cette dernière phrase un dernier argument qui a définitivement plombé le calcul des probabilités comme un ensemble de "jeux et casse-tête" sans aucune autre utilité sociale que de développer une gymnastique de l'esprit. De 1900 à aujourd'hui, la liste des cours et traités qui ont privilégié cet aspect de "problem solving" sans offrir aucune réflexion articulée sur la probabilité a été assez longue²². D'où d'ailleurs l'intérêt du petit livre de vulgarisation de Fréchet et Halbwachs en 1924 qui ouvrait une toute autre perspective.

5. La crise générale du modèle probabiliste standard

Le positionnement de March sur la méthodologie statistique et ses liens au calcul des probabilités ne sont pas atypiques. Il est tout à fait représentatif d'une situation de transition qui a duré environ 25 ans, entre 1895 et 1920, et dont l'enjeu est tout à fait clair si l'on se réfère à plusieurs arènes dans lesquelles l'avenir de la statistique est débattu, ou encore aux différentes sources qui permettent à l'historien d'aujourd'hui de s'en faire une idée. Il y a d'abord bien évidemment le corpus des traités de la discipline. Une étude statistique des contenus de ces traités²³ sur une période d'un siècle et demi révèle une évolution significative d'une science ayant un objet propre – successivement l'État, le territoire, la population, le système économique et social – à une science qui se définit par sa méthode - l'induction numérique – et peut s'appliquer à n'importe quel champ de connaissance, qu'il relève des sciences naturelles ou des sciences humaines. Cette évolution de la discipline se fait d'ailleurs en toute conscience et les statisticiens comme Meitzen [1890] ou Julin [1921, p. 43] font référence à la statistique comme méthode qui pourrait succéder à la statistique comme science. Poursuivie par des moyens statistiques après codage des contenus, l'étude du même corpus révèle un tournant plus précis qui s'opère dans les années 1900-1930 avec l'introduction dans ces traités d'une série de résultats mathématiques qui en deviennent petit à petit l'ossature. Les ouvrages de Maurice Block (1878), J. Bertillon (1895), Meitzen (1891) Von Mayr (1897), Faure (1906) sont représentatifs d'une statistique qui est d'abord une discipline administrative de production de l'information, et de traitement sommaire de celle-ci, avec des outils qui ne dépassent pas le tableau, le graphique et la moyenne. Tout au plus y trouve-t-on, avec ou sans critique, les traces d'une tentation d'inférence vers des régularités ou des lois au sens de Quetelet. Ce qui n'enlève rien à leur intérêt comme exposé des principes d'une science administrative et logique de construction des faits sociaux. Mais les ouvrages de Czuber (1921), Niceforo (1925), Fisher (1925) Jordan (1927), Darmois (1928), Rietz (1924 et 1927), Risser et Traynard (dans le traité de Borel, 1933) sont d'une toute autre facture : ce sont de véritables traités de statistique mathématique dont les objets principaux sont les théorèmes de Bernoulli et de Laplace, les méthodes d'ajustement de distributions, la modélisation des séries chronologiques, la théorie des moments, celle de la corrélation, et celle des sondages; toutes ces méthodes sont dédiées à des inférences inductives sur des distributions, des liaisons et des évolutions, et sont largement fondée sur le calcul des probabilités. Cependant le même corpus fait apparaître une catégorie de traités intermédiaire à la fois par ses dates de parution (entre 1910 et 1925) et par ses thématiques (ils sont à peu près tous centrés sur la méthodologie du traitement de données économiques) : Bowley (1901 et 5 éditions ultérieures), Yule (1910 et 14 éditions successives), Liesse (1905), King (1912), Secrist (1917), Jerome (1924), Julin (1921), Crum et Patton (1925), Young (1925), March

²² Comparer le traité de l'actuaire Laurent [1908] qui est une collection décousue de petits problèmes et celui de Carvallo [1912] bien structuré autour des théorèmes fondamentaux, de Bernoulli et Laplace est à ce titre instructif. Carvallo souligne d'ailleurs dans son introduction que son livre est né d'un besoin exprimé par March pour le recrutement de statisticiens.

²³ Voir Armatte (1991, 1995, 2001)

(1930), Galvani (1934). L'approche de la majorité de ces ouvrages, américains pour la plupart, et liés à l'introduction de la statistique juste avant que des chaires lui soient dédiées, est une approche empirique du traitement des données économiques, où la collecte, les graphiques, les résumés statistiques, les *index-numbers*, et les calculs de corrélation et régression dominant, sans que soient développés les bases probabilistes. On trouverait une évolution analogue dans les journaux des sociétés de statistique parisienne (la SSP) anglaise (RSS) et américaines (ASA).

Un troisième corpus peut être évoqué, c'est celui des deux éditions allemande et française de *l'Encyclopédie des Sciences mathématiques pures et appliquées*²⁴. Emmanuel Czuber, auteur de L'article "Wahrscheinlichkeitsrechnung" (calcul des probabilités) dans *l'Encyklopädie* de Klein considère que la théorie de Laplace "ne présente plus qu'un intérêt historique" et de nombreux indices montrent que le travail critique de Bertrand pour déconsidérer une grande partie de l'héritage laplacien a porté ses fruits : *"il n'est guère possible de donner une définition satisfaisante de la probabilité"* dit son traducteur français Le Roux. Suit une présentation très éclatée de petits problèmes de jeux qui témoigne de la crise du calcul des probabilités au tournant du siècle, avant les travaux de Borel, Bachelier, Fréchet, P. Lévy, Kolmogorov. L'article "Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik" (Application du calcul des probabilités à la Statistique) de la même encyclopédie dont l'auteur allemand est Ladislaus von Bortkiewicz est encore plus révélateur des effets de cette crise pour la statistique. Son papier est totalement éclaté entre d'une part une tentative de fondement de la statistique et d'autre part une longue discussion d'ajustements empiriques de données démographiques dans lequel l'apport de la probabilité est totalement inexistant. L'auteur français, l'actuaire Oltramare, a dû faire précéder la traduction d'une mise en garde sur les difficultés à se saisir des modèles probabilistes pour traiter des données statistiques :

"Vu la nature un peu spéciale du sujet, il est nécessaire d'expliquer à quels points de vue on peut envisager l'application aux données statistiques des règles du calcul des probabilités. Les points de vue concernant l'objet même de la statistique diffèrent en effet beaucoup les uns des autres. Le domaine à traiter est très étendu, il embrasse natalité, mortalité, morbidité, criminalité etc.; les résultats sont représentés le plus souvent par des rapports qui peuvent parfois être regardés comme nous fournissant des valeurs particulières de fonctions plus ou moins déterminées. Il s'agit surtout pour le statisticien, de chiffrer dans la mesure du possible la valeur ou le degré de précision des nombres obtenus, et le cas échéant, de déterminer le plus exactement possible la nature, la forme et les coefficients des fonctions dont il possède un certain nombre de valeurs particulières. Le calcul des probabilités nous donne pour cela des règles simples à suivre, mais dans le cas bien entendu où ses principes seraient réellement applicables. La première chose à faire est de s'en assurer.

(...)

En général toutefois la concordance n'existera pas (...) les écarts qui existent entre ces moyennes et les valeurs particulières qui ont servi à les former, bien que pouvant présenter le caractère d'erreurs accidentelles, dépassent souvent notablement les limites prévues par la théorie. Devons-nous dans ce cas repousser le calcul des probabilités, nous priver de ses avantages ou attendre pour l'appliquer que les causes réelles de l'événement soient mieux connues? Quelques statisticiens ne l'ont pas admis; par des artifices de calcul, la considération des probabilités variables, ils sont parvenus dans bien des cas à étendre le champ d'application de ses principes, et (...) à ramener les résidus à un caractère normal."

²⁴ Voir [Armatte, 2005].

Ces lignes reflètent un conflit latent sur l'héritage de Laplace. Ceux qui s'accrochent au dogme laplacien de la loi normale ne sont plus très nombreux. Et pourtant la théorie des erreurs de Gauss et Laplace qui a lié la méthode des moindres carrés à la loi normale aurait dû lui assurer une certaine pérennité. Ce fut le cas en astronomie et en géodésie, davantage d'ailleurs en Allemagne et en Belgique qu'en France, comme l'a montré la thèse de M.F. Jozeau [1997]. Ce fut le cas de manière inespérée dans les sciences sociales avec le succès de la physique sociale de Quetelet formée autour du concept d'homme moyen : son existence dépend d'une homogénéité attestée par la loi normale. Mais en démographie comme en économie, les économistes ont d'abord boudé la statistique jusque dans les années 1870²⁵, et quand ils commencent alors à l'adopter, dans le contexte des crises économiques de la fin de siècle, c'est pour chercher dans les chiffres accumulés les symptômes du cycle des affaires, pour en capter les composantes. Les tests de Dormoy et Lexis, comparant la dispersion empirique des fréquences de tel ou tel phénomène sur une ou deux décennies avec la dispersion que fournirait le modèle binomial ont conclu que les régularités de Quetelet – celles qu'il attribuait par exemple à la propension au crime – étaient presque toujours illusoire, à l'exception du sex-ratio à la naissance, et que décidément le modèle probabiliste de l'urne de composition fixe était de peu d'utilité pour rendre compte des séries observées²⁶. Constatant par exemple un rapport élevé entre ces deux dispersions pour la mortalité sur 10 ans, Bertrand conclut : *"L'assimilation à une urne de composition invariable n'est donc pas acceptable. La vicissitude des événements règle sans cesse la composition de l'urne. Tantôt c'est le choléra qui passe et y verse des boules noires. Ce sont des eaux plus pures et plus fraîches qui apportent les boules blanches."*

Si les erreurs astronomiques pouvaient être supposées normales c'est qu'elles avaient été réduites aux seules erreurs accidentelles, par rapport à un modèle d'orbites dérivé des formules de la mécanique de Newton. En économie point de tel modèle théorique. Point de séparation des causes accidentelles d'avec les causes variables qui s'enchevêtrent. C'est cette interdépendance possible des causes, soit qu'elles se renforcent soit qu'elles s'annulent qui conduit à des dispersions inférieures ou supérieures, et suggèrent des modèles que Lexis baptise "hyponormal" et "hypernormal". Une école alternative se dessine qui prend ses racines dans les travaux de Poisson, Cournot et Bienaymé sur les "modèles d'urne variable" et dont Lexis est le principal porte parole à la fin du XIX^e siècle. Il s'agit de sauver le calcul des probabilités en lui fournissant de nouvelles bases qui lui permette de produire à nouveau des modèles plausibles pour les données démographiques et sociales. Bortkiewicz (1868-1931), élève de Lexis à Göttingen, est sur la période qui nous intéresse le principal auteur engagé dans ce programme. Inventeur en 1898 de la "loi des petits nombres" - il aurait été préférable qu'il la baptise loi des événements rares - qui généralisait la loi limite de Poisson, il a fourni un premier modèle d'une loi asymétrique, de variance légèrement supérieure, vers laquelle converge la loi binomiale quand la probabilité de l'événement considéré dans l'épreuve répétée est faible. Cependant la carrière de Bortkiewicz est à peine commencée quand il rédige son article de l'*Encyklopädie* et c'est pour cela sans doute que l'article balance entre une reconstruction probabiliste nouvelle de la statistique mathématique (première partie) et une déclinaison très classique des ajustements empiriques des données démographiques, avec lesquels s'est longtemps confondue la statistique. Bortkiewicz s'est illustré ensuite par la poursuite des recherches de Lexis sur la stabilité et l'homogénéité des séries²⁷, mais aussi par

²⁵ Voir [Ménard 1977]

²⁶ Armatte (1989)

²⁷ Sa dernière publication, [Bortkiewicz 1931] est particulièrement riche.

plusieurs controverses importantes avec Gini, sur le sujet de la mesure des inégalités, et avec Karl Pearson²⁸, auquel il reprochait principalement la construction de formules d'ajustement sans aucun fondement théorique. On rappelle que le principe de ces formules était de dériver tout un "système de courbes" servant de modèles de distribution, dont les fameuses lois Beta, sur la seule base d'une équation aux différences ajustée à des lois binomiale et hypergéométrique. On ne peut douter que Bortkiewicz ait été finalement avec Lexis le meilleur représentant de ce courant d'un renouvellement du calcul des probabilités qui devait à la fois batailler sur un front contre les héritiers directs de Laplace encore attachés au seul modèle normal, comme Edgeworth et sur un autre front contre ceux qui pensaient pouvoir se passer des probabilités, comme Pareto, Pearson, et plus fortement encore, Lucien March, Warren Persons à Harvard, Knapp de Strasbourg, et plus tard encore François Divisia à Paris.

Conclusion

Deux éléments ressortent avec force de ce rapide balayage des activités de L. March. Les travaux statistiques de Lucien March donnent une première impression d'éclectisme qui caractérise bien la situation du statisticien du début XXe, capable de s'investir à la fois dans les dispositifs d'enquête et de recensement, dans le traitement mécanique des informations (mécanographie), dans la méthodologie d'analyse mathématique, dans les enjeux de l'évolution démographique et dans la fabrication d'outils de prévision économique, dans les débats de philosophie des sciences. En second lieu, l'impression d'éclectisme fait place à un sentiment de forte intégration de ces différents engagements. Ils font système. Le rôle attribué aux graphiques et aux indices non agrégés, le souci d'un ajustement le plus fin possible de modèles mathématiques aux distributions pour en faire des résumés sténographiques comme aurait dit Pearson s'expliquent assez bien par un souci constant de coller aux réalités des phénomènes. Le déni de réalisme attribué au calcul des probabilités et à ses modèles s'accorde avec un agnosticisme sur les essences et une application de la règle d'Occam chère au Pearson de la *Grammaire*. Ce rejet de l'alea ne doit pas être pour rien dans son désintérêt relatif pour les sondages, le seul des 3 thèmes du comité de méthodologie qu'il a animé à l'IIS sur lequel il n'ait pas travaillé, et qui n'occupe qu'une toute petite place dans son traité. Dans ses *Principes*, l'analyse statistique est d'abord présentée comme une comptabilité. Son engagement fort dans le recensement et son traitement mécanographique en ferait-il davantage un homme des registres qu'un homme des enquêtes, comme le sera également Carmille? Finalement la méthodologie statistique est prise dans une sorte de toile d'implications sociales et cognitives assez fortes faite de choix philosophiques (la contingence) et politiques (l'eugénisme, le natalisme, le taylorisme), accrochés à un système technique qui privilégie l'investigation systématique par recensement, le traitement mécanographique, la tabulation et la représentation graphique, aux dépens d'un recours à des modèles plus essentiels, que ce soit ceux du calcul des probabilités ou ceux de l'économie mathématique qui ne l'ont guère attiré.

Inventeur d'une terminologie qui ne lui survivra pas – pléthométrie, covariation, valeur normale - March est aussi bien le disciple contemporain de Karl Pearson, que de Cheysson, Le Chatelier, Colson, Rueff, Divisia et plus largement la famille des ingénieurs économistes français attachés aux valeurs traditionnelles de la famille, de l'Etat-Providence, et d'une science positive soucieuse de s'en tenir aux phénomènes, à leur objectivation sous forme de lois empiriques, et à l'exploitation de leur "valeur pratique" comme le disait Borel.

²⁸ Voir par exemple [Bortkiewicz 1915]

Abréviations :

JSSP : *Journal de la Société Statistique de Paris*

BIIS : *Bulletin de l'Institut International de Statistique*

JRSS : *Journal of the Royal Statistical Society*

Bibliographie

AMMON O., 1896, *Gesellschaftsordnung*, 2è ed., Iena.

ARMATTE M., 1989, "Modèles statistiques de l'homogénéité et de la stabilité d'une population au XIXème siècle", in *Les ménages; mélanges en l'honneur de J.Desabie*, Paris, INSEE

ARMATTE M., 1991, "Une discipline dans tous ses états: la Statistique à travers ses traités (1800-1914), *Revue de synthèse*, IVème série, N°2, Avril-juin 1991

ARMATTE M., 1992, "Conjonctions, conjoncture et conjecture. Les baromètres économiques", *Histoire et Mesure*, VII, 1-2, p. 99-149.

ARMATTE M., 1994, "François Divisia (1889-1964) ", in *Les professeurs du Conservatoire National des Arts et Métiers. Dictionnaire biographique 1794-1955*, C. Fontanon et A. Grelon (dir), Paris, INRP/CNAM, tome 1, p.424-440.

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., 2001, "Developments in Statistical reasoning and their links with mathematics" in *Changing Images in Mathematics, From the French Revolution to the new Millenium*, U. Bottazzini et A. Dahan Dalmedico (ed) , London, Harwood Acad. Publish, pp. 137-162.

ARMATTE M., 2005, "Les Probabilités et les Statistiques dans l'*Encyclopédie des Sciences Mathématiques Pures et Appliquées : promesses et déceptions.*", à paraître dans un ouvrage consacré à l'*Encyclopédie*, sous la direction d'Hélène Gispert et Catherine Goldstein.

BAUSCHINGER J. et ANDOYER H., 1908, "Théorie des Erreurs", *Encyclopédie des Sciences Mathématiques Pures et Appliquées*, édition française de J.Molk, Paris, Gauthier-Villars

BERTILLON J., 1895, *Cours élémentaire de Statistique Administrative*, Paris, Société des Editions Scientifiques, 593 p.

BERTRAND J., 1971, *Calcul des probabilités*, 3ème édition, New York, Chelsea, XLIX + 317 p.; 1ère ed. 1889; 2ème ed. 1907.

BORTKIEWICZ L. (von), 1909, "Statistique". *Encyclopedie des Sciences Mathématiques Pures et Appliquées*, Tome I, 4ème Vol., 3ème fasc., 453-480, Jules Molk (ed), Gauthier-Villars, Paris; exposé par F. Oltramare, d'après "Anwendungen der Wahrscheinlichkeitsrechnung auf statistik", *Encyklopädie der mathematischen Wissenschaften*, Leipzig, Teubner, 1901.

BORTKIEWICZ L. (von), 1915, "Realismus und Formalismus in der mathematischen Statistik, *Allgemeines statistisches Archiv*, 9, p. 222-256.

BORTKIEWICZ L. (von) , 1931, The relation between stability and homogeneity, *Ann.Math.Stat.*, 2, 1-22.

BOWLEY A. L., 1901, *Elements of Statistics*, King and Son, London. 2ème ed 1902, 335 p.; 4ème ed. 1920, 454 p.; Traduction française sur la 5ème édition par L. Suret et G. Lutfalla, 1929.

BUNLE H., 1911, "Relations entre les variations des indices économiques et le mouvement ds mariages", *JSSP*, p. 80-91.

CAROL A., 1995, *Histoire de l'eugénisme en France*.Paris, Le Seuil.

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

DARMOIS G., 1928, *Statistique Mathématique*, Librairie Douin, Paris.

DESROSIERES A., 1998, Lucien March (1859-1933) : A pioneer of Quantitative Economics, in Warren J. Samuels (ed), *European Economists of the early 20th Century*, Vol 1, Northampton, Edward Elgar.

- DIVISIA F., 1932, "Economie et Statistique. A propos d'un livre récent", *Revue d'Economie Politique*, Sept-Oct. 1932, p. 1457-1495. Extraits dans *Bulletin X-Crise*, 14-15, 1933-34.
- EDGEWORTH F. Y., 1925 (1887-1889), "Measurement of Change in Value of Money" et "Tests of accurate measurement", à partir de trois mémoires présentés à la BAAS en 1887-89, *Papers relating to political economy*, London, Macmillan, Vol. I, p. 195-335.
- EDGEWORTH F. Y., 1896, "On the asymmetrical probability-curve", *Phil Mag.*
- EDGEWORTH F. Y., 1898-1900, "On the Representation of Statistics by Mathematical Formulae", *Journal of the Royal Statistical Society*, 61, p.671-700; 62, p. 125-140, 373-385, 534-555; 63 (Suppl.), p. 72-81.
- EDGEWORTH F. Y., 1909, "Sur l'application du calcul des probabilités à la statistique", *BIIS*, Tome XVIII, p. 220-253.
- JOVANOVIC F. et LE GALL P. , 2001, March to numbers. The statistical style of Lucien March, *History of Political Economy*, Vol. 33, Supplément, p. 86-110.
- JOZEAU M.F., 1997, *Géodésie au XIXe siècle : de l'hégémonie française à l'hégémonie allemande. Regards belges*, Thèse de Doctorat, Université Paris VII, Edition IREM.
- JULIN A., 1921, *Principes de Statistique théorique et appliquée*, Préface L.March, tome1: Statistique théorique, Marcel Rivière, Paris, 712 p.
- KEYNES J. M., 1909, "The Method of Index Numbers with Special Reference to the Measurement of General Exchange Value", *Collected Writings of John Maynard Keynes*, D. Moggridge (ed), Macmillan Cambridge University Press, 1983, Vol. XI, p.49-173.
- LAPLACE P. S., 1986 (1825), *Essai philosophique sur les probabilités*, postface B.Bru, d'après la 5ème édition (1825), Paris, C. Bourgois.
- LENOIR M., 1913, *Etudes sur la Formation et le Mouvement des prix*, Thèse de Doctorat, Paris, Giard et Brière.
- LEXIS W., 1875 , *Einleitung in die Theorie de Bevölkerungs-statistik*, Strasbourg, Trübner.
- LEXIS W., 1877, *Zur Theorie der MassenerScheinungen in der menschlichen Gesellschaft*, Freiburg, F.Wagner.
- LEXIS W., 1879, "Ueber die Theorie der Stabilität statistischer Reihen", *Jahrbuch für Nationalöfonomie und Statistik*, Vol. 32, p. 60-98.
- MARCH L., 1898, "Salaires, quelques exemples de leur distribution", *JSSP*, p. 193-206, et 241-248.
- MARCH L., 1904-1905, "Les représentations graphiques et la statistique comparative", *JSSP*, 1904, p. 398, 407; 1905, p. 47-48.
- MARCH L., 1905, "Comparaison numérique de courbes statistiques", *JSSP*, p.255-277.
- MARCH L., 1907, "L'organisation de la statistique générale de la France", discours en prenant la présidence de la SSP, *JSSP*, p.35-43
- MARCH L., 1908, "Remarques sur la terminologie en statistique", extrait d'une communication au congrès de mathématiques de Rome, *JSSP*, p. 290-296.
- MARCH L., 1909, "De l'application de procédés mathématiques à la comparaison des statistiques", *BIIS* , Tome XVIII, 1^{ère} Livraison, p. 254-264.
- MARCH L., 1910a, "Essai sur un mode d'exposer les principaux éléments de la théorie statistique", *JSSP*, p. 447-486.
- MARCH L., 1910b, "Influence des variations de prix sur le mouvement des dépenses ménagères à Paris", *JSSP*, p. 45, 124, 136.
- MARCH L., 1911a, "Les moyens de rendre comparables les courbes statistiques", *BIIS*, Tome XIX, 1^{ère} Livraison, p.50.
- MARCH L., 1911b, "Statistique", in *De la méthode dans les sciences*, Paris Alcan.
- MARCH L., 1912a, Revue de "Calcul des probabilités" de Louis Bachelier, Paris Gauthiers-Villars, *JSSP*, p. 258.
- MARCH L., 1912b, "La théorie des salaires. A propos de l'ouvrage du Professeur Ludwell Moore : *Laws of Wages*", *JSSP*, p. 366-383.

- MARCH L., 1920, "La méthode statistique", *Métron*, Vol. I, N°1, p. 22-
- MARCH L., 1921a, "Les modes de mesure du mouvement général des prix", *Metron*, Vol. I, N°4, p. 73.
- MARCH L., 1921b, "La méthode statistique en économie politique", *Revue de métaphysique et de morale*, 28, p.137-173.
- MARCH L., 1922, *Mouvement des prix et des salaires pendant la guerre*, Paris, PUF.
- MARCH L., 1923a, "Les indices du mouvement général des affaires", *JSSP*, p. 166, p.221
- MARCH L., 1923b, "Rapport sur les indices de la statistique économique", *BIIS*, 2^{ème} Livraison, p.3.
- MARCH L., 1924, "Les indices économiques", *Metron*, Vol. III, N°3, p. 334-
- MARCH L., 1927, "F. Edgeworth et les indices des mouvements des prix", *JSSP*, p. 261-
- MARCH L., 1928, "Différences et corrélation en statistique", *JSSP*, p. 38-58.
- MARCH L., 1930, *Les principes de la méthode statistique*, Paris, Alcan, 796 p.
- MARIETTI P. G., 1947, *La Statistique Générale de la France*, Thèse de Doctorat en Droit, Univ.de Paris.
- MEITZEN A., 1890 (1886), *History, Theory, and Technique of Statistics*, American Academy of political and social science, Philadelphia; Trad.angl.et introd.R.P.Falkner d'après *Gechichte, Theorie und Technik der Statistik*, Berlin.
- MENARD C., 1977, "Trois formes de résistance aux statistiques: Say, Cournot, Walras", in *Pour une histoire de la Statistique*, tome 1, Paris, INSEE.
- PARETO V., 1896, *La Courbe de la Répartition de la Richesse*, Lausanne, Faculté de Droit / Librairie Viret-Genton.
- PARETO V., 1897 (1964) , *Cours d'Economie Politique*, Lausanne, F.Rouge et Paris, F.Pichon; *Oeuvres Complètes*, Dir. Giovanni Busino, Tome 1, Genève, Droz.
- PEARSON K., 1895, "Mathematical Contributions to the Theory of Evolution. II. Skew Variation in Homogeneous Material", *Phil. Trans. of the R.S.*, 186A, p. 343-414.
- PEARSON K., 1912 (1892), *La Grammaire de la Science*, traduction et préface de L.March sur la 3ème édition (1911), Paris, Alcan.
- PEAUCELLE J.L., 2004, "A la fin du XIXème siècle, l'adoption de la mécanographie est-elle rationnelle?", *Gérer et comprendre*, N°77, Septembre 2004, p.60-74.
- POINCARÉ H., 1912, *Calcul des Probabilités*, Leçons professées à la faculté des sciences de Paris (2ème semestre 1893-94), 2^{ème} édition (1^{ère} édition 1896), Paris, G.Carré, 274 p.
- PORTER T., 2004, *Karl Pearson. The scientific life in a statistical age*, Princeton UNiv. Press.
- PRESSAT R., 1987, "Quelques aspects de l'histoire de l'enseignement de la statistique", *JSSP*, 128-1, p.15-29.
- RIETZ H. L. (ed), 1924, *Handbook of mathematical statistics*, N.Y., Houghton Miffling Cy.
- RISSER R. et TRAYNARD C. E., 1933, *Les principes de la statistique mathématique*, Fascicule IV du tome I du *Traité du calcul des probabilités et de ses applications* par Emile Borel, Paris Gauthier-Villars.
- S.G.F., 1910, *Salaires et coûts de l'existence à différentes époques jusqu'en 1910*, note préliminaire de L.March, Paris, Imp.Nat.
- S.G.F., 1913, *Historique et travaux de la fin du XVIIe siècle au début du XXe, avec 103 tableaux graphiques relatifs aux travaux les plus récents* (Exposition universelle de Gand), Paris, Imprimerie nationale.
- YULE G. U., 1909, "Les applications de la méthode de corrélation aux statistiques sociales et économiques", C.R. XIIème session I.I.S., Paris, *B.I.I.S.*, tome XVIII, p. 265-277.
- ZIZEK F., 1913, *Statistical Averages. A methodological Study*, trad. américaine W.M.Persons, New York, Henry Holt and Cy, 392 p.