De Poisson à Lazarsfeld en passant chez Quetelet, Bertillon, Galton et Pearson

Jean-Jacques Droesbeke, Gilbert Saporta

Chapitre 1

DE POISSON À LAZARSFELD EN PASSANT CHEZ QUETELET, BERTILLON, GALTON ET PEARSON

Jean-Jacques Droesbeke et Gilbert Saporta

1.1 Introduction

L'émergence d'un concept ou d'une théorie est habituellement expliquée par de nombreux facteurs. Parmi ces derniers, on trouve les progrès des sciences et techniques à l'époque de leur développement, le contexte politique et social dans lequel ils ont évolué, les problèmes concrets qui les ont rendus nécessaires ou encore les qualités des hommes ou des femmes qui ont contribué à leur essor. Parmi ceux-ci, il y en a de célèbres et d'autres plus méconnus. Nous évoquerons ci-dessous quelques personnages qui ont participé à l'émergence des mélanges de distributions ou de la prise en compte de variables latentes. Nous rencontrerons à cette occasion Poisson, Quetelet, Galton, Pearson ou encore Spearman dont le parcours et les contributions au développement de la statistique sont relativement bien connus. Nous soulignerons quel est leur apport à la problématique développée dans cet ouvrage. Nous y trouverons aussi les noms de deux figures dont on a moins l'habitude de parler.

Le premier est médecin - on pourrait aussi le qualifier de démographe, d'anthropologue ou de statisticien, dans les acceptions qu'avaient ces qualificatifs au 19e siècle - et s'appelle Louis-Adolphe Bertillon ; il vécut de 1821 à 1883. Le second, un sociologue, Paul Felix Lazarsfeld est né en 1901, et décédé en 1976.

1.2 Les mélanges de distributions

1.2.1 La situation au début du 19e siècle

Une des caractéristiques importantes de l'évolution de la statistique au début du 19e siècle est certainement la volonté d'appliquer à l'étude des populations humaines un des outils majeurs de cette époque développé dans le cadre des mesures d'observation - et tout particulièrement de l'astronomie : la loi des erreurs qui portera plus tard le nom de loi normale. A cette époque, deux hommes mettent fin provisoirement à la recherche d'une loi des erreurs dont la crédibilité devient évidente aux yeux de leurs contemporains : Carl Friedrich Gauss (1777-1855) et Pierre Simon de Laplace (1749-1827). L'avantage de cette loi réside dans le fait qu'elle est non seulement associée au principe de prendre comme valeur centrale une moyenne, mais qu'elle possède en outre la qualité de reposer sur la méthode des moindres carrés comme choix optimal d'ajustement. Ajoutons encore à cela que Laplace justifie son importance par le recours à un théorème limite et on aura en main les principales raisons du succès que cette loi va connaître au cours des deux siècles qui vont suivre.

De nombreux scientifiques vont participer à ce mouvement, tout particulièrement en France. L'un d'entre eux peut à juste titre être associé à l'émergence des mélanges de lois ; il s'agit de Siméon Denis Poisson.

1.2.2 L'apport de Poisson

L'importance de Siméon Denis Poisson (1781-1840) dans le développement de la statistique et des probabilités au début du 19e siècle n'est plus à prouver. Il va bien sûr favoriser l'étude et la diffusion de la loi qui porte son nom, mais son oeuvre est loin de se réduire à cette unique contribution.

Né à Pithiviers le 21 juin 1781, Poisson fait rapidement état de son intérêt pour les mathématiques. Il entre à l'École polytechnique de Paris en 1798, ses qualités de mathématicien sont vite remarquées par ses professeurs[1], et notamment par Pierre Simon de Laplace et Joseph-Louis Lagrange, qui se liera d'amitié avec lui. Ses études se concrétisent par la rédaction de deux mémoires en 1800 dont l'un sera publié sur les recommandations des deux examinateurs, Sylvestre-François Lacroix et Adrien-Marie Legendre qui apporteront, eux aussi, une importante contribution à l'élaboration et à la diffusion des outils statistiques nouveaux évoqués plus haut.

Soutenu par Laplace, Poisson devient répétiteur à l'École polytechnique dès l'obtention de son diplôme avant d'être nommé professeur suppléant en 1802. Le départ de Joseph Fourier pour Grenoble lui permet de s'installer

[1] Le seul enseignement qui rebute Poisson est le cours de géométrie dispensé par Monge, en raison de son inhabilité, semble-t-il, à réaliser correctement des figures. Le sort les a peut-être réconciliés en les plaçant l'un à côté de l'autre leur nom sur le pourtour du premier étage de la Tour Eiffel parmi les 72 savants qui y sont mentionnés.

durablement dans le corps professoral de cette grande école, en 1806, ce qui constitue clairement un record dans la rapidité des promotions obtenues au sein de cette institution. Soutenu par Laplace, Poisson devient répétiteur à l'École polytechnique dès l'obtention de son diplôme avant d'être nommé professeur suppléant en 1802. Le départ de Fourier pour Grenoble en 1806 lui permet de s'installer durablement dans le corps professoral de cette grande école, ce qui constitue clairement un record dans la rapidité des promotions obtenues au sein de cette institution.

Ses nombreux travaux - on compte plus de 300 publications sous son nom - concernent surtout les mathématiques, la physique et l'astronomie. Deux articles importants paraissent en 1824 (*Sur la probabilité des résultats moyens des observations*) et 1829 (*Suite du Mémoire sur la probabilité...*). On y trouve une démonstration rigoureuse du théorème central limite concernant une somme de variables indépendantes, identiquement distribuées ou non. Dans son travail *sur les proportions de naissances des filles et des garçons* publié en 1830, Poisson établit que l'approximation normale d'une loi binomiale de paramètre p et d'exposant n n'est pas valable quand p tend vers 0 et que la moyenne np reste finie ; il établit la distribution limite qui sera appelée ultérieurement loi de Poisson. En 1835 et 1836, notre mathématicien s'attaque à l'étude de la loi des grands nombres pour des essais binomiaux dont les résultats seront repris et consolidés dans un mémoire publié en 1837 qui va faire parler de lui *Recherches sur la probabilité des jugements en matière criminelle et matière civile précédées des règles générales du calcul des probabilités.* La simple consultation de la table des matières de cet ouvrage fait apparaître cinq chapitres qui soulignent les centres d'intérêt de l'auteur :

Chapitre 1 : *Règles générales des probabilités*

Chapitre 2 : *Suites des règles générales ; probabilités des causes et des événements futurs, déduites de l'observation des événements passés*

Chapitre 3 : *Calcul des probabilités qui dépendent de très grands nombres* (Cas des chances constantes pendant les épreuves)

Chapitre 4 : *Suite du calcul des probabilités dépendantes de très grands nombres* (Cas des chances variables d'une manière quelconque, comprenant celui des chances constantes)

Chapitre 5 : *Applications des règles générales des probabilités aux décisions des jurys et aux jugements des tribunaux*

Vient d'abord un préambule de 29 pages dans lequel Poisson justifie d'emblée le titre de son ouvrage : « *Parmi les applications du calcul (des probabilités), une des plus importantes est celle qui se rapporte à la probabilité des jugements, ou, en général, des décisions rendues à la pluralité des voix* ». Poisson commence par rendre hommage à Condorcet dont *L'essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* souligne toute «l'utilité de ce genre de recherches ». Il s'attarde cependant peu sur ce personnage pour en venir immédiatement à sa référence essentielle en la matière : «*Dans son traité des probabilités, Laplace s'est aussi occupé du calcul des chances d'erreur à craindre dans le jugement rendu contre un accusé, à une majorité connue, par un tribunal ou un jury composé d'un nombre de personnes également connu. La solution qu'il a donnée de ce problème, l'un des plus délicats de la théorie des probabilités, est fondée sur le principe qui sert à déterminer les probabilités des causes diverses auxquelles on peut attribuer les faits observés ; principe que Bayes a présenté d'abord sous une forme un peu différente, et dont Laplace a fait ensuite le plus heureux usage*[2]*, dans ses mémoires et dans son traité, pour calculer la probabilité des événements futurs d'après l'observation des événements passés. (...) Laplace fait une hypothèse qui n'est point incontestable : il suppose que la probabilité qu'un juré ne se trompera pas est susceptible de tous les degrés également possibles, depuis la certitude, représentée par l'unité, jusqu'à l'indifférence, qui répond dans le calcul à la fraction* $\frac{1}{2}$*, et se rapporte à une égale chance d'erreur et de vérité. L'illustre géomètre fonde son hypothèse sur ce que l'opinion d'un juré a sans doute plus tendance vers la vérité que vers l'erreur ; ce qu'on doit admettre effectivement en général. Mais il existe une infinité de lois différentes de probabilité des erreurs qui satisfont à cette condition, sans qu'on soit obligé de supposer que la chance qu'un juré ne se trompera pas, ne puisse jamais descendre au-dessous de* $\frac{1}{2}$*, et qu'au-dessus de cette limite, toutes ses valeurs soient également possibles* ».

Poisson poursuit son exposé en regrettant de n'avoir pu présenter ses remarques lorsque Laplace était vivant : «*l'autorité de son nom m'en eût fait un devoir, que son amitié, dont je me glorifierai toujours, m'aurait rendu facile à remplir* ». Poisson souligne ensuite toute sa réticence à faire part des «*principales raisons qui m'ont déterminé à abandonner la dernière solution à laquelle Laplace s'était arrêté* », soulignant par là ses réticences à l'égard de son aîné, malgré le respect que l'on avait encore pour lui, dix ans après sa mort, et qui se prolongera dans la suite.

Après ces considérations liées à l'importance de Laplace dans le monde scientifique, Poisson en arrive à l'un des points centraux de son ouvrage : «*Les choses de toutes natures sont soumises à une loi universelle qu'on peut appeler* la loi des grands nombres. *Elle consiste en ce que, si l'on observe des nombres très considérables d'événements d'une même nature, dépendants de causes constantes et de causes qui varient irrégulièrement, tantôt dans un sens, tantôt dans l'autre, c'est-à-dire sans que leur variation soit progressive dans aucun sens déterminé, on trouvera, entre ces nombres, des rapports à très peu près constants* ».

Les quatre premiers chapitres de l'ouvrage de Poisson[3], occupent 288 pages

---

[2]Malgré cette présentation élogieuse, Laplace n'aurait probablement pas aimé cette phrase, lui qui a toujours nié avoir connu les résultats de l'homme d'église anglais avant de les trouver lui-même !

[3]La pagination mentionnée ici concerne la réédition de l'ouvrage réalisée par les Éditions

alors que le dernier n'en contient que 98, ce qui souligne clairement le souhait de Poisson de réaliser avant tout une étude théorique et de discuter les développements de son illustre prédécesseur. Notre intention n'est pas ici de commenter le contenu de ce mémoire mais plutôt de nous arrêter à quelques extraits situés principalement dans les chapitres 2 et 4 et destinés à expliciter la *loi des grands nombres*, dénomination qu'il a proposée, rappelons-le, dans le préambule.

Poisson construit un modèle dans lequel un événement - appelons-le E ou F comme lui - peut advenir au cours d'une épreuve que l'on peut répéter un grand nombre de fois : «*Je suppose que par la nature des événements (...), celui qui arrivera à chaque épreuve puisse être dû à l'une des causes C_1, C_2, C_3,..., C_y, dont y est le nombre, qui s'excluent mutuellement, et que je regarderai d'abord comme également possibles. Je désigne par c_i la chance que la cause quelconque C_i donnera à l'arrivée de l'événement E ; de manière qu'à une épreuve déterminée, à la première, par exemple, la chance de E soit c_1 quand ce sera la cause C_1 qui interviendra, c_2 quand ce sera C_2, etc. S'il n'y avait qu'une seule cause possible, la chance de E serait nécessairement la même à toutes les épreuves ; mais dans notre hypothèse, elle sera susceptible, à chaque épreuve, d'un nombre y de valeurs également probables, et variera en conséquence, d'une épreuve à l'autre* ».

Dans ses *Recherches*, Poisson reprend - comme nous l'avons déjà souligné - des résultats publiés antérieurement et redémontre le théorème central limite prouvant que la moyenne d'un grand nombre de «variables aléatoires indépendantes » admet asymptotiquement une loi normale. Ce qu'il est important de préciser, c'est que cette démonstration est réalisée pour des variables binomiales, multinomiales mais aussi continues.

En se plaçant dans le contexte présenté ci-dessus, et en utilisant des notations plus actuelles, Poisson établit qu'en notant $p_j(x)$ la probabilité d'avoir le résultat x quand il est causé par C_j , la probabilité d'avoir x est une combinaison linéaire convexe des $p_j(x)$ avec des poids w_j égaux aux probabilités d'avoir C_j ; il apparaît ainsi que la loi de x est un mélange des lois conditionnelles de x étant donné la cause où les poids de ces lois sont les w_j. En recourant deux fois de suite au théorème central limite, il prouve que la moyenne des résultats obtenus en répétant le processus un très grand nombre de fois admet asymptotiquement une loi normale. Poisson généralise ensuite son étude au cas où le nombre de causes peut être infini et où la loi du mélange peut être continue. Il faut reconnaître qu'à part l'ouvrage de Lacroix intitulé *Traité élémentaire du calcul des probabilités*, les différentes publications de l'époque n'auront pas la diffusion que pourrait justifier leur contenu, en dehors d'un cercle restreint de lecteurs avertis. Il n'est donc pas surprenant que l'étude des mélanges de lois ne se soit pas développée dans la foulée de l'ouvrage de Poisson, mais cela importe peu. Sa contribution n'en reste pas moins remarquable. En 1839, Irénée Jules Bienaymé étudie le cas particulier du mélange de lois binomiales.

Jacques Gabay en 2003.

Quatre ans plus tard, Antoine-Augustin Cournot présentera les résultats de ses prédécesseurs sous une forme plus accessible dans son *Exposition de la théorie des Chances et des probabilités* qu'il publie en 1843. A cette époque, la terminologie utilisée distingue clairement le mot *chance* qui découle d'une propriété objective d'un mécanisme aléatoire qui peut engendrer un événement du mot *probabilité* qui mesure plutôt les raisons de croire qu'un événement se produira ou non. Une caractéristique intéressante de Cournot est qu'il recourt parfois à l'usage de représentations graphiques pour soutenir son discours, contrairement à la plupart de ses prédécesseurs. Un autre personnage suivra souvent le même chemin : Adolphe Quetelet.

1.2.3 Quetelet et les mélanges de distributions

Le rôle d'Adolphe Quetelet (1796-1874) dans le développement de la statistique au début du 19e siècle est prépondérant et la «loi de possibilité » - c'est ainsi qu'il préfére appeler la loi qui deviendra «normale » plus tard - occupe dans son oeuvre une place essentielle. Le point de départ de Quetelet se situe bien sûr dans la continuité des travaux de Gauss et Laplace que nous avons évoqués ci-dessus.

Le mérite de Quetelet n'est pas d'avoir créé de nouveaux instruments ; d'autres l'ont fait avant lui comme on a déjà pu s'en rendre compte. Au cours de ses nombreux voyages en Europe, il va rencontrer les principaux artisans de la recherche en statistique et probabilités ou correspondre avec eux. Il s'intègre ainsi dans un réseau informel qui regroupe entre autres Laplace, Gauss, Fourier, Bravais, Poisson ou encore Malthus. Il va réussir là où les autres connaîtront un échec plus ou moins important : la vulgarisation de bon aloi. Son secret tient en grande partie aux qualités de sa personne et de sa méthode qui privilégient la synthèse sur l'analyse, le sens pédagogique à la rigueur de la formule (voir Armatte et Droesbeke [1997]).

Une des principales expressions de ces qualités se trouvent dans les *Lettres à S.A.R. Le Duc de Saxe-Cobourg et Gotha sur la théorie des probabilités appliquées aux sciences morales et politiques* qu'il publie en 1846 et qui seront traduites en anglais en 1849. L'originalité majeure de cet ouvrage est qu'il est composé de 46 lettres successives formant autant de chapitres. Cette publication est destinée à poursuivre par écrit les leçons commencées en 1836 et qui s'adressaient à deux neveux du Roi des Belges Léopold Ier, les princes Ernest et Albert de Saxe-Cobourg. Le premier devenu Duc régnant de Saxe-Cobourg en 1844 est le seul destinataire des *Lettres*, mais Quetelet entretiendra avec le second, qui épouse la Reine Victoria en 1840, des relations chaleureuses et durables. Le projet de Quetelet est de fournir à ses deux élèves une théorie de l'homme moyen utile dans leurs fonctions respectives et fondée sur le calcul des probabilités.

Les lettres XX et XXI nous intéressent tout particulièrement ici. Dans la première, Quetelet revient sur les deux notions de moyenne qu'il a introduites auparavant ; il distingue en effet une « moyenne proprement dite » qui s'applique aux mesures répétées d'une même grandeur d'une « moyenne arithmétique » qui concerne des mesures de grandeurs semblables : «*Dans le premier cas, la moyenne représente une chose existant réellement ; dans le second, elle donne, sous la forme d'un nombre abstrait, une idée générale de plusieurs choses essentiellement différentes, quoiqu'homogènes* ».

L'argument principal de la lettre XX peut être résumé en cinq points (voir Armatte et Droesbeke [1997]) :

[1] Mesurons mille fois le tour de poitrine d'une statue d'un gladiateur. Les mille mesures se répartiront autour de leur moyenne selon la loi de possibilité, « *en sorte que la maladresse, ou le hasard, si nous aimons mieux ce mot pour couvrir notre amour-propre, procède avec une régularité qu'on ne serait guère tenté de lui attribuer* », avec une dispersion que l'auteur chiffre par un écart probable de 1 mm.

[2] « *Si l'on avait à mesurer la poitrine d'une personne vivante au lieu de celle d'une statue, (...) la ligne qui les représenterait serait toujours la courbe de possibilité, mais dilatée dans le sens horizontal, proportionnellement à l'erreur probable* », du seul fait de la respiration.

[3] « *...supposons qu'on ait employé un millier de statuaires pour copier le gladiateur avec tout le soin imaginable (...) si leurs inexactitudes ne sont qu'accidentelles, les mille mesures, groupées par ordre de grandeur, présenteront encore une régularité remarquable et se succéderont dans l'ordre que leur assigne la loi de possibilité* ».

[4] « *On trouve, dans le troisième volume du journal médical d'Edimbourg, les résultats de 5738 mesures prises sur les poitrines de soldats de différents régimes écossais* ». Ces mesures sont réparties selon la loi de possibilité avec une erreur probable de 33 mm.

[5] « *Et si l'on nous donnait les deux séries de mesures sans les avoir désignées d'une manière particulière, nous serions très embarrassés de dire quelle série a été prise sur 5738 soldats différents et quelle série a été obtenue sur une seule et même personne, avec moins d'habitude et des moyens d'appréciation plus grossiers. (...) les choses se passent absolument comme si les poitrines qui ont été mesurées avaient été modelées sur un même type, sur un même individu, idéal si l'on veut, mais dont nous pouvons saisir les proportions par une expérience suffisamment prolongée* ».

Par cette approche, Quetelet introduit un être de référence équivalent à celui mesuré dans le premier cas : l'homme moyen, cet être idéal dont nous sommes en quelque sorte la copie.

La lettre XXI rebondit sur l'analyse faite dans la lettre précédente :

« *Depuis que j'ai écrit ma dernière lettre, j'ai songé qu'on pouvait faire une objection assez spécieuse à ce qu'elle contient. Cette objection n'aura certainement pas échappé à Votre Altesse, habituée comme elle l'est à voir des* *régiments et à tenir compte de la taille des soldats. On me demandera ce que deviendrait ma prétendue régularité dans la manière dont procèdent les mesures, si j'avais à opérer sur un régiment de cuirassiers, par exemple, ne contenant que des hommes très-grands et très-vigoureux, et sur un régiment de chasseurs, composé d'hommes beaucoup plus petits* ».

Quetelet ne renonce pas à son homme moyen « *dont tous les autres hommes s'écartent plus ou moins* » mais il précise sa pensée en prenant un autre exemple : « *Que demain on peuple une île déserte, en y plaçant 1000 hommes de la race la plus grande, des patagons par exemple, ayant tous 1m80 de hauteur, et 1000 lapons n'ayant que 1m40 de hauteur : la taille moyenne dans cette île sera de 1m60 et cependant pas un homme n'aura cette taille. En groupant les tailles par ordre de grandeur, nous ne pourrons former que deux groupes, et la loi de possibilité sera complètement en défaut, du moins en apparence. Mais on voit d'abord que le désaccord ne provient ici que de ce qu'on mêle des choses hétérogènes, des hommes de races différentes, et qui ont des lois différentes de développement* ».

Quetelet généralise ensuite son exemple en prenant des Patagons de tailles distinctes variant autour de 1m80 : « *quand on les groupera par ordre de grandeur, leur arrangement, nous le savons déjà, sera déterminé par la loi de possibilité. Si l'on en fait autant pour les lapons, il peut arriver qu'un certain nombre de ces derniers aient la taille des patagons les plus petits, et alors les deux lignes qui figurent leur arrangement empièteront l'une sur l'autre. (...) Si l'on avait à mesurer les tailles chez un peuple (mélangé), on pourrait ignorer qu'un pareil mélange a eu lieu, mais l'expérience le ferait connaître. La ligne qui représenterait les mesures aurait deux sommets, qui annonceraient deux races différentes ayant des moyennes inégales* ». Et d'ajouter pour convaincre son interlocuteur que « *la loi de possibilité a donc ce nouvel avantage qu'elle aide à résoudre un problème très-intéressant sous le rapport anthropologique* ».

Quetelet poursuit cette lettre dans le même esprit, ce qui nous permet de comprendre qu'il est persuadé qu'une loi normale peut être engendrée en mélangeant un grand nombre de lois normales. Il est assez étonnant qu'il n'ait pas explicité son raisonnement par le recours à des graphiques alors qu'il s'agit là d'un outil qu'il apprécie fortement. Un autre homme va le faire dans sa foulée : Louis-Adolphe Bertillon.

1.2.4 Louis-Adolphe Bertillon

Louis-Adolphe Bertillon naît à Paris le 2 avril 1821. Son père, Jean-Baptiste Bertillon, y tient un magasin, rue Montmorency, où le parisien peut découvrir la lumière du gaz qu'il avait appris à fabriquer à Londres, quand le gaz à éclairage était encore inconnu sur le continent. L'aisance que ce commerce apporte à la cellule familiale leur permet de quitter Paris vers 1828 pour s'installer dans une vaste propriété rurale près de Montargis. C'est en parcourant les bois de la propriété de ses parents que Louis-Adolphe se prend de passion pour l'étude des animaux qu'il est tout heureux, quand il en trouve un décédé, de le disséquer pour en comprendre l'usage et la disposition de ses organes. À la mort de sa mère emportée par le choléra en 1832, il est mis en pension à Paris où son père veut qu'il se forme pour se lancer à son tour dans le commerce. Ce projet marque le début de relations plus conflictuelles avec son père car il est plutôt décidé à se consacrer à sa passion en faisant des études de médecine, malgré l'opposition paternelle. La chimie l'intéresse tout particulièrement et, comme le dira plus tard le Docteur Chervin dans son éloge funèbre [1884] : « *afin de bien posséder cette science, il l'enseigna* ».

Pendant ses études de médecine, Bertillon suit aussi les cours du Collège de France où il admire Michelet dont il devient vite un familier, cimentant ainsi une amitié qui durera près de 30 ans. Devenu médecin, il pratique sa discipline avec sérieux mais, porté par des idées républicaines que l'on qualifierait de « gauche » à notre époque, il participe aussi activement à des débats et réunions consacrés à l'émancipation des ouvriers et qui se répandent peu après la révolution de 1848. C'est au cours de l'une d'entre elles qu'il rencontre celui qui va devenir son beau-père : Achille Guillard, ingénieur gazier de profession, mais qui se fait surtout remarquer par ses travaux en botanique et qui semble le premier à avoir utilisé le mot de *démographe* dans son ouvrage *Eléments de statistique humaine ou démographie comparée* (1855)..

Pendant les combats révolutionnaires, il met son activité au service des blessés en parcourant les barricades. Il s'investit aussi dans des projets de création de magasins où l'on propose à prix minimes des produits de première nécessité. Son implication dans des projets peu appréciés au niveau gouvernemental le conduit en prison où son codétenu n'est autre que son beau-père. Libéré par non-lieu dans la suite, Bertillon connaît à nouveau l'enfermement dans la nuit du 3 au 4 décembre 1851, dans la foulée du coup d'état de Louis-Napoléon Bonaparte ; il restera séquestré pendant plusieurs semaines avant de retrouver une vie plus sereine.

La thèse de doctorat en médecine qu'il soutient en 1852 l'oriente vers la démographie : *De quelques éléments de l'hygiène dans leur rapport avec la durée de vie*. Il effectue des études qui le persuadent que l'homme ne peut se développer harmonieusement et vivre plus longtemps que dans l'état de liberté.

Il consacre son énergie à l'étude de thèmes comme celui de la mortalité à chaque âge - en publiant notamment des *conclusions statistiques contre les détracteurs de la vaccine* - et s'intéresse particulièrement à la mortalité des nouveau-nés. Sa notoriété croissante va lui permettre de participer à la fondation de 2 sociétés savantes qui sont créées en ce milieu du 19e siècle : la *Société d'anthropologie* en 1859 et la *Société de statistique de Paris* en 1860. Il sera porté ultérieurement à la présidence de chacune de ces deux sociétés.

Il reprend des activités politiques qu'il avait mises en veilleuse quand, le 4 septembre 1870, la troisième République est proclamée, à la suite de la défaite de Napoléon III à Sedan. Il est alors nommé par le Gouvernement de la Défense nationale en tant que maire du 5ème arrondissement de Paris.

Á son retour à la vie privée, il retourne vers la publication scientifique. Citons en particulier la *Démographie figurée de la France* qu'il réalise en 1874 et plusieurs articles qu'il publie en 1876 dans le *Dictionnaire encyclopédique des sciences médicales* sur des sujets divers : mariage, France, mortalité, natalité, mort-né, etc. L'un d'entre eux est publié en 1876 et est intitulé « *Moyenne* » ; il va nous permettre de parler de Bertillon dans le contexte de cet ouvrage.

Il publie dans le dictionnaire encyclopédique d'autres contributions sur des sujets divers : mariage, France, mortalité, natalité, mort-né, etc. Il s'éteint en 1883 à la suite d'une longue maladie.

Cet article sur la moyenne est révélateur de l'esprit qui prévaut à l'époque parmi les statisticiens observateurs de la société. Grand admirateur de Quetelet, Bertillon reprend à son compte les notions introduites par ce dernier en leur donnant une nouvelle dénomination : moyenne prise comme valeur approchée d'une grandeur réelle, mais inconnue, ou *moyenne objective* ; *moyenne subjective* quand il s'agit de prendre une valeur imaginaire représentative d'un ensemble d'individus ou d'objets. Ce second cas est illustré par le calcul de la taille moyenne de 18 260 conscrits français que Bernard Garel évoquera dans le chapitre 3 de cet ouvrage. En examinant la taille de 9 002 conscrits mesurés entre 1851 et 1860 dans le département du Doubs, Bertillon constate que le graphique de la distribution de ces tailles (figure 1.1) fait apparaître deux modes distincts, ce qui ne correspond pas à ce qu'il attendait.

Son embarras s'amplifie quand il constate que ce phénomène persiste lorsqu'il scinde la période d'observation en deux : de 1851 à 1855 et de 1856 à 1860 ! Bertillon ne peut que se rendre à l'évidence : cette situation ne peut être accidentelle, elle est nécessairement due à une cause précise. S'ensuit alors une hypothèse selon laquelle ce phénomène est dû à la présence de deux ethnies distinctes dans le Doubs, l'une dont la taille est petite, l'autre avec une taille plus grande. Il semble que ce graphique soit la première représentation publiée qui ait engendrée l'hypothèse d'un mélange de populations. Le fin mot de cette histoire est dû à un autre personnage du nom de Livi qui, en 1896, s'apercevant que ce phénomène avait disparu, se rendra compte du fait que cette bimodalité résultait en réalité d'un changement d'unités opéré par Bertillon qui avait remplacé des mesures arrondies au centimètre le plus proche par des groupements en classes exprimées en pouces. Il n'empêche que l'idée d'être en présence d'un mélange de populations avait germé dans l'esprit de notre médecin-démographe-statisticien ! Bertillon ne connaîtra jamais cet épilogue car il s'éteindra en 1883, à la suite d'une longue maladie.

Notons encore pour terminer ce paragraphe que Louis-Adolphe aura deux fils qui s'illustreront eux aussi, mais de manière différente. Le premier, prénommé Jacques, suivra les traces de son père sur le chemin de la statistique. Comme lui, il dirigera le service de statistique de la ville de Paris - où il sortira plusieurs *Atlas de la statistique graphique de Paris* - et publiera en 1895 un *Cours*

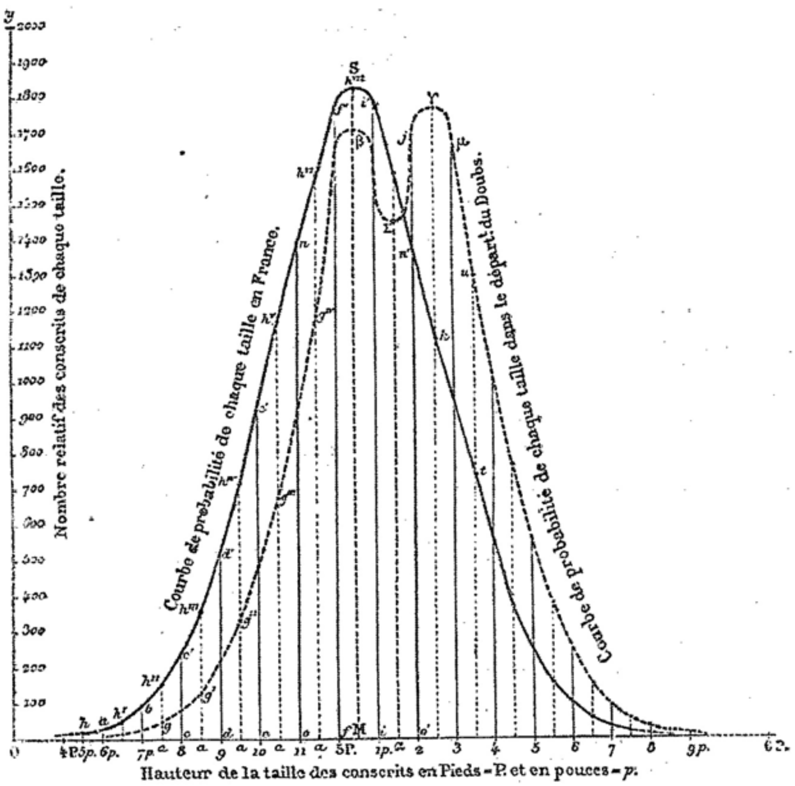

Figure 1.1 –

élémentaire de statistique administrative qui connaîtra un certain succès. Le second fils, Antoine, suit un parcours différent. Moins bon élève que son frère, il débute sa carrière en 1879 à la Préfecture de Police de Paris où il passe son temps à recopier sur des fiches les signalements des détenus qui y sont amenés. Désireux de construire une procédure destinée à retrouver fiablement des individus ayant déjà été fichés par la police, Antoine Bertillon propose un système d'identification des individus basé sur 14 mensurations distinctes du corps humain. Son système, connu sous le nom de « bertillonnage » va se répandre rapidement et sera utilisé dans de nombreux services d'identité judiciaire, en concurrence ou en complément selon le cas, à l'usage des empreintes digitales. Antoine a donc la possibilité de léguer son nom à la postérité comme l'un des inventeurs de la police scientifique. Mais un événement fâcheux va ternir sa réputation montante. Durant l'affaire Dreyfus, il va en effet construire une thèse dite de l'autoforgerie, qui va devenir une arme essentielle de l'accusation et des antidreyfusards.

Bien que n'étant pas expert en graphologie - il ne sera d'ailleurs jamais désigné officiellement comme expert juridique - il acceptera cependant, à la demande de l'accusation, de participer au débat qui a pour objet de savoir si l'écriture d'un bordereau annonçant l'envoi de documents secrets à l'ambassade d'Allemagne était identique à celle de lettres saisies au domicile de Dreyfus. Son incompétence en la matière ne l'empêchera cependant pas d'affirmer que les deux écritures sont identiques, ce qui contribue malheureusement à la condamnation de Dreyfus. Chacun sait que cette contre-vérité sera rejetée dans la suite et Dreyfus réhabilité. Antoine Bertillon a manqué là une bonne occasion de se taire !

L'histoire semble avoir retenu avant tout le criminologue et non le faux expert puisque le nom d'Antoine Bertillon a été donné à une rue du 15ème arrondissement de Paris.

Un visiteur célèbre voudra figurer dans le répertoire des photos d'identification de Bertillon : il s'agit de Sir Francis Galton dont nous n'évoquerons pas ici les découvertes liées à la corrélation ou la régression, ni ses études consacrées aux empreintes digitales, préférant nous concentrer sur sa participation à l'histoire des mélanges de distributions.

1.2.5 Galton et les mélanges

Francis Galton naît le 16 février 1822 à Sparkbrook, près de Birmingham, en Angleterre. Cousin de Charles Darwin, il effectue des études de médecine à Londres, puis à Cambridge, où il poursuit aussi une formation en mathématiques. Ses études ne sont pas particulièrement brillantes mais le décès de son père en 1844 lui assure une fortune considérable qui lui permettra d'éviter d'exercer une carrière médicale qui ne l'attirait pas réellement.

En revanche, il développe une passion pour les voyages qui vont le mener en Afrique, dans des régions peu explorées à l'époque. La qualité du récit de ses voyages et l' intérêt qu'il porte à la météorologie et la cartographie vont lui ouvrir les portes des cercles scientifiques fondés par les géographes.

Il devient rapidement obsédé par le besoin de mesurer, compter, construire des graphiques pour étudier les comportements humains et les caractères des individus. Limité dans sa maîtrise des mathématiques, Galton rencontre la loi qu'il désignera plus tard sous le qualificatif de *normale* dans un écrit d'un collègue géographe qui utilise la méthode proposée par Quetelet pour ajuster cette loi à des données. C'est ainsi que dans son premier grand ouvrage intitulé *Hereditary Genius* qu'il publie en 1869, Galton puise son inspiration statistique dans les *Lettres* de Quetelet dont nous avons parlé plus haut.

Galton se laisse aisément convaincre du fait que les caractéristiques des populations humaines suivent de façon souvent naturelle cette loi proposée par Gauss et Laplace et popularisée par Quetelet.

Sous l'influence des idées propagées par Darwin, Galton est persuadé que l'hérédité est le résultat de la combinaison d'un nombre fini de particules

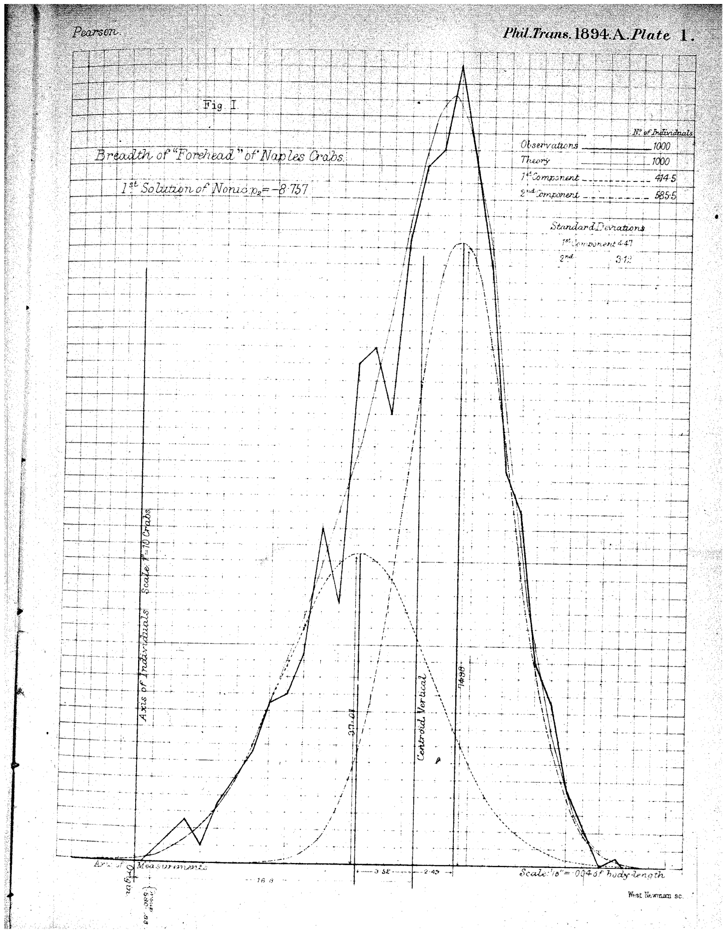

Figure 1.2 – Galton chez Bertillon

héréditaires dont les variations biologiques peuvent être décrites par une loi binomiale symétrique.

En 1875, il écrit : « *Pourquoi un mélange de séries radicalement différentes donnerait dans de nombreux cas des résultats apparemment identiques à ceux d'une simple série ?* ».

Il sous-tend sa réflexion par des approches de complexité croissante. Après avoir analysé le résultat du lancer de 130 pièces de monnaie de nature différente, il poursuit par des exemples plus élaborés dont celui-ci (troisième exemple). Formant la somme de trois lois binomiales symétriques $52B(x, 9) + 924B(x, 10) +$

$40B(x, 11)$ il s'aperçoit que le résultat est proche d'une loi valant $1024B(x, 10)$.

La construction du quinconce lui permet d'éviter de faire reposer ses déductions sur des développements mathématiques.

1.2.6 Les mélanges de Karl Pearson

Un dernier personnage va nous permettre de terminer notre voyage à travers le 19e siècle. Né en 1857, Carl Pearson va poursuivre des études brillantes à Cambridge qu'il termine en 1879 avant de séjourner un an en Allemagne où il prend l'habitude de remplacer la première lettre de son prénom, C, par la lettre K que chacun lui attribue encore aujourd'hui.

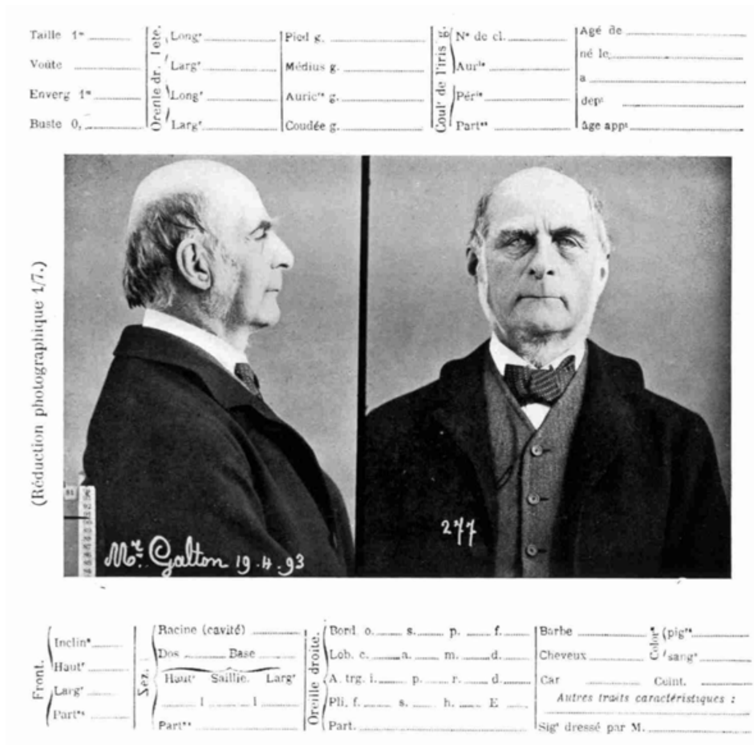

Figure 1.3 –

Á cette époque, les centres d'intérêt de Pearson sont multiples et engagés, allant de la physique aux mathématiques en passant par le théâtre, la poésie, l'histoire de l'art ou encore la littérature médiévale germanique. Il s'implique aussi pour soutenir des causes comme l'amélioration du statut de la femme dans la société ou la liberté de pensée. Son attrait pour la statistique ne se révélera en fait que vers 1890.

Pendant quelques années, Pearson occupe des postes d'enseignant de mathématiques avant de se fixer, en 1884, au *University College of London* dont il influencera le développement et les changements institutionnels. Dès ce moment, il se forge une stature d'enseignant respecté et écouté.

De 1891 à 1894, il occupe la chaire de géométrie au *Gresham College* dans la Cité de Londres. Il y dispensera 38 conférences dont les 8 premières constituent la trame de la *Grammar of Science* qu'il publie en 1892.

Pour éviter que ses étudiants ne soient rebutés par des développements mathématiques trop abstraits, Pearson recourt à la statistique pour les inciter à recourir à l'outil mathématique dans des disciplines comme les assurances ou le commerce. Il défend à cette occasion l'usage des graphiques et des représentations géométriques comme outil pédagogique. Il trouve aussi dans les études biologiques de son temps une source d'inspiration et de réflexion importante.

En 1892, Pearson fait la connaissance d'un zoologiste, Walter Frank Weldon, qui trouve auprès de lui l'aide dont il a besoin pour analyser les nombreuses données qu'il amasse ça et là. C'est ainsi que Weldon lui fournit de grandes quantités de mesures des caractéristiques physiques de crabes qui font apparaître des distributions de mesures dont la forme n'épouse pas réellement l'allure normale chère à Quetelet et Galton. C'est auprès de Pearson que Weldon va trouver le moyen d'expliquer les comportements bimodaux qu'il observe.

Pearson va trouver dans les mathématiques de la mécanique les outils et les dénominations - les *moments* en l'occurrence - qui vont l'aider à traiter les données de Weldon. La mise en évidence graphique de mélanges de distributions se retrouve dans un gros article intitulé *Contributions to the Mathematical Theory of Evolution* et publié par Karl Pearson en 1894. Nous reproduisons l'un de ces graphiques dans ce chapitre. Il concerne l'estimation des paramètres du mélange de deux lois normales que Bernard Garel évoquera dans le chapitre 2.

Pour Pearson, ajuster un mélange de deux lois normales équivaut à découper une courbe asymétrique en deux lois normales. Pour y arriver, il faut estimer cinq paramètres à partir des cinq premiers moments, ce qu'il propose de faire au moyen de la méthode destinée à devenir la *méthode des moments*. Cela le conduira à résoudre une équation polynomiale du neuvième degré (voir chapitre 2).

1.3 Les modèles à variables latentes

Expliquer des phénomènes observés par l'intervention de facteurs cachés est sans doute une pratique aussi vieille que l'humanité, mais l'étude des superstitions, croyances, magies etc. ne relève pas de cet ouvrage !

Á la suite de chercheurs en sciences sociales, en paraphrasant McCutcheon

[1987], on considère que de nombreux concepts ne peuvent être observés directement comme l'autoritarisme, les préjugés raciaux ou l'engagement religieux etc. Par exemple, alors que l'on ne peut observer directement l'engagement religieux, on peut raisonnablement penser qu'un haut niveau d'engagement conduit à une fréquentation élevée des lieux de culte, à prier plus souvent, à se conduire de façon plus conforme à l'orthodoxie... Si chacun de ces indicateurs est causé par une variable non observable ou latente, nous nous attendons à des covariations entre variables manifestes qui s'expliquent par la relation de chacune avec la (ou les) variables latentes.

Cela se traduit par l'hypothèse fondamentale d'indépendance conditionnelle : les variables manifestes sont indépendantes conditionnellement aux variables latentes.

1.3.1 Les différents modèles

L'origine de l'usage des modèles à variables latentes remonte aux travaux de Charles Spearman [1904] sur l'intelligence, qui ont conduit au développement de l'analyse factorielle. Spearman postulait que les corrélations entre résultats de tests cognitifs pouvaient s'expliquer par les variations d'un unique facteur commun noté g que l'on qualifia un peu rapidement d'« intelligence générale ». Cette hypothèse fut rapidement remplacée par celle d'un modèle plurifactoriel. L'analyse factorielle a donné lieu à une abondante littérature que nous ne pouvons résumer ici. Á ce premier modèle où variables manifestes (observées) et variables latentes (facteurs) sont de même nature quantitative, succédèrent les modèles de profils latents, de traits latents et de classes latentes, qui diffèrent selon la nature des variables. La table **??** (Bartholomew et Knott, [1999]) résume les différentes situations.

Table 1.1 – Les quatre modèles de structures latentes

	Variables latentes	
Variables manifestes	**qualitatives**	**quantitatives**
qualitatives	classes latentes	traits latents
quantitatives	profils latents	analyse factorielle

On a souvent considéré que les deux méthodes où les variables manifestes sont qualitatives se sont développées dans des domaines différents : « *Latent class analysis was developed mainly within the social and political sciences, whereas latent trait models have a clear psychometrical background.* » (Heinen, 1996, ix) ce qui est inexact puisque les traits latents et les classes latentes ont été introduits simultanément par P.F. Lazarsfeld sous le nom commun de *structures* *latentes*, comme on le verra plus tard. Notons que l'expression *traits latents* n'est plus guère utilisée et que les psychométriciens parlent plus volontiers d'*Item response Theory* .

Nous nous concentrerons ici sur le modèle de classes latentes. Notons que ce modèle est par nature unifactoriel : deux variables latentes qualitatives à m_1 et m_2 catégories sont équivalentes à une seule variable à $m_1 m_2$ catégories.

1.3.2 Le modèle des classes latentes

Ce modèle fut introduit en 1950 par Paul F. Lazarsfeld. On peut le considérer comme l'analogue de l'analyse factorielle lorsque variables manifestes et variables latentes sont toutes catégorielles.

Les travaux de Paul F. Lazarsfeld se situent dans le contexte de la deuxième guerre mondiale où de nombreuses études psycho-sociales furent menées par le département de la guerre américain sur le personnel militaire. Ces études furent publiées dans les quatre volumes de *The American Soldier* (Stouffer, [1949-50]). Le volume 4 contient les deux chapitres de Paul F. Lazarsfeld sur les structures latentes, précédés par huit chapitres de Louis Guttman. Lazarsfeld écrivit ensuite le chapitre *Latent Structure Analysis* de l'ouvrage monumental *Psychology : A Study of a Science*, [1959]. L'ensemble de ses travaux sur les structures latentes pendant vingt ans fut rassemblé dans le manuel *Latent Structure Analysis* paru en 1968 écrit en collaboration avec N.W. Henry.

Lazarsfeld [1950a] commence par décrire sous le nom de *trace lines* le modèle des traits latents (sans utiliser cette expression) où des réponses à des questions dichotomiques s'expliquent par un continuum. Il donne pour exemple la mesure de l'ethocentrisme x des soldats américains dans la deuxième guerre mondiale avec des questions comme « *I believe that our Europeans allies are much superior to us in strategy and fighting morale : YES NO* ». Les *trace lines* sont les courbes des probabilités conditionnelles de répondre OUI à la question n° i sachant x , $f_i(x)$, qu'il modélise par des polynômes. Lazarsfeld précise page 379 qu'il s'agit d'un modèle dont l'existence ne peut ni être prouvée, ni réfutée directement.

Il développe ensuite le modèle des classes latentes dans le cas de questions dichotomiques en commençant par un modèle à deux classes où dans chaque classe les variables observées sont mutuellement indépendantes. Il arrive alors à des systèmes d'équations (*accounting equations*) dont la résolution laborieuse à l'aide de déterminants occupe plusieurs dizaines de pages (Lazarsfeld, [1950b]). Il faudra attendre les travaux de Leo Goodman [1974] pour obtenir un algorithme d'estimation des paramètres par le maximum de vraisemblance en montrant que le modèle de classes latentes peut s'interpréter comme un modèle log-linéaire, et en considérant les classes latentes comme des données manquantes.

Dans le cas de p variables manifestes dichotomiques x_i le modèle s'écrit sous la forme suivante :

$$f(\mathbf{x}) = \sum_{j=1}^{k} \pi_j \prod_{i=1}^{p} p_{ij}^{x_i}(1-p_{ij})^{1-x_i} \tag{1.1}$$

où p_{ij} est la probabilité de répondre OUI à l'item i dans la classe j et π_j est la probabilité d'appartenance à la classe j .
On voit alors qu'il s'agit d'un modèle de mélange discret de p lois de Bernoulli. Le modèle de « profils latents », qui en est l'homologue pour des variables manifestes numériques, n'est qu'un cas particulier de mélange discret où les variables manifestes sont indépendantes à l'intérieur de chaque classe de la variable latente, ce qui revient à considérer que les corrélations entre variables manifestes n'était dues qu'à l'hétérogenéité des observations.

La méthode des classes latentes permet ensuite d'affecter chaque répondant à la classe la plus probable et peut donc être utilisée comme une méthode de classification non supervisée. On trouvera dans Aitkin *et al* [1987], Celeux [1988] ainsi que dans le chapitre 4 de Droesbeke *et al.* [2005] des comparaisons avec des méthodes issues de l'analyse des données comme la classification selon la distance du χ^2 .

1.3.3 A propos de Paul Lazarsfeld

L'inventeur de la méthode des classes latentes mérite d'être mieux connu des statisticiens. Né à Vienne en 1901, Paul Félix Lazarsfeld est selon le sociologue Raymond Boudon, « l'un des quatre ou cinq sociologues de sa génération qui auront le plus fortement marqué l'histoire de leur discipline » et le promoteur de la pensée mathématique en sociologie. Il a étudié le droit, l'économie et les mathématiques à l'université de Vienne où il a soutenu une thèse en mathématiques appliquées en 1924 sur des problèmes issus de la théorie de la gravitation d'Einstein : *Über die Berechnung der Perihelbewegung des Merkur aus der Einsteinischen Gravitationstheorie.* Il reconnaissait avoir été influencé par Ernst Mach, Henri Poincaré et Albert Einstein.

Le cercle de Vienne (R.Carnap, O.Neurath) développait à la même époque des travaux de philosophie des sciences, mais Lazarsfeld ne semble pas l'avoir fréquenté, ni non plus son contemporain Karl Popper, qu'il n'appréciait guère et jugeait trop puritain (Lautman *et al.* [1998]).

Il s'intéressa très vite aux enquêtes psycho-sociales et fut un pionnier des études media en réalisant en 1930-31 ce qui fut peut-être la première enquête scientifique auprès des auditeurs de la radio. Il resta toute sa vie un spécialiste des études de communication et de l'influence des media sur les choix politiques.

Son étude sur les chômeurs de Marienthal (l'édition française de 1981 est préfacée par Pierre Bourdieu) lui ouvrit les portes des États-Unis en 1933 grâce à une bourse Rockefeller, alors que militant socialiste actif il assistait à la montée du nationalisme. Á l'expiration de sa bourse, il retourne en Europe. En 1934, le Parti socialiste d'Autriche est déclaré illégal. La plupart des membres de sa famille sont arrêtés. Il perd son poste dans l'enseignement secondaire. Seule est maintenue la vague charge de cours à l'université de Vienne qui couvrait ses activités de recherche. Il décide d'employer les dollars qui lui restent à l'achat d'un billet de troisième classe pour les États-Unis. Il s'y établit et devint selon ses termes « un marxiste en congé ».

Il créa un institut de recherches à l'université de Newark (maintenant un campus de Rutgers University), conçu comme un pont entre les approches nord-américaines et européennes des études d'opinion. Son étude « *The Art of Asking Why* » [1935] eut une grande influence auprès des spécialistes. Á titre d'anecdote : pour faire croire que son institut avait un personnel plus nombreux, Lazarsfeld publiait également sous un pseudonyme. Il recherchait aussi des subsides d'entreprises et effectua par exemple des études sur la consommation pour l'industrie laitière et le magazine Cosmopolitan.

Avec un nouveau financement de la Fondation Rockefeller, il créa en 1937 un centre de recherches à l'université de Princeton consacré à l'étude des effets sociaux de la radio. Avec Frank Stanton, chef du département des études d'audience de CBS (F. Stanton devint président de CBS de 1946 à 1971), il mit au point en 1938 le premier appareil d'analyse de la satisfaction des auditeurs (avec des boutons satisfait, pas satisfait) le « Lazarsfeld-Stanton Program Analyzer » appelé familièrement *Little Annie*. L'*Office of Radio Research* fut transféré ensuite à l'université Columbia où Paul F. Lazarsfeld fut nommé professeur de sociologie et y resta jusqu'à sa retraite en 1969. On lira avec profit la communication de S. Cavin [2008] qui retrace l'histoire du *Princeton Radio Project*, son rôle pendant la deuxième guerre mondiale et pendant la guerre froide. En 1944 ce centre devint le Bureau of Applied Social Research. Ce bureau fut renommé *Center for the Social Sciences* en 1976, puis *Lazarsfeld Center for the Social Sciences* pour le vingtième anniversaire de son décès et devint une composante de l'*Institute for Social and Economic Research and Policy* (ISERP) fondé en 1999 (http ://www.iserp.columbia.edu/about/).

Paul F. Lazarsfeld étudia les déterminants des votes et élabora pour sa première grande étude électorale, lors des élections présidentielles de 1940, une méthode d'enquête (baptisée « panel ») consistant à mener, tout au long de la campagne électorale, des interviews répétées auprès d'un échantillon représentatif d'électeurs. Les résultats furent publiés dans *The People's Choice* [1944]. Á l'occasion des élections de 1948, Paul F. Lazarsfeld proposa d'utiliser des chaînes de Markov pour modéliser les changements d'opinion dans un panel. Curieusement, il ne résolut pas lui-même le problème qu'il confia à T.W. Anderson [1954], son voisin du département de statistique de Columbia, probablement pour l'attirer dans son orbite (Lautman, J. *et al* [1998] p.278).

Paul Lazarsfeld qui se décrivait comme un sociologue mathématicien partisan des recherches empiriques fut très fier d'être le premier *Quetelet Professor of Social Science* de l'Université Columbia en 1962. Paul F. Lazarsfeld a écrit avec David Landau l'article « Quetelet » de l'International Encyclopedia of the Social Sciences [1968], il considérait Quetelet plutôt qu'Auguste Comte, comme le véritable fondateur de la sociologie .

L'influence internationale de Lazarsfeld fut considérable comme en témoigne pour la critiquer d'un point de vue « gauchiste » l'article de Michael Pollak [1979] dans la revue Actes de la Recherche en Sciences Sociales, qui suscita une réaction de J. Dumazedier dans la même revue cinq années plus tard. Francophile convaincu, il fit de longs séjours à plusieurs reprises à Paris et fut nommé en 1972 professeur émérite de l'Université René Descartes, titre non réglementé à l'époque (communication personnelle de Raymond Boudon), à l'initiative de Jean Stoetzel (professeur de sociologie et fondateur de l'IFOP).

P.F.Lazarsfeld considérait qu'il n'avait eu que quatre idées originales dans sa vie dont l'analyse des panels et les structures latentes. Il décéda en 1976.

1.4 Bibliographie

Aitkin, M., Francis, B., et Raynal, N. [1987], Une étude comparative d'analyses des correspondances ou de classifications et des modèles de variables latentes ou de classes latentes. *Revue de Statistique Appliquée*, 35, 3, p.53-81.

Anderson, T.W. [1954], Probability models for analyzing time changes in attitudes. In : P.F. Lazarsfeld, Editor, *Mathematical Thinking in the Social Sciences*, Free Press, Glencoe, p.17-66.

Armatte, M., Droesbeke, J.J. [1997], Quetelet et les probabilités : le sens de la formule, in *Actualité et universalité de la pensée scientifique d'Adolphe Quetelet*, Bruxelles, 1996, Acad. Roy. Belg. Cl. Sci. Mém. Collect. 8(3) 13, p.107-135.

Bartholomew, D.J. et Knott, M. [1999], *Latent Variable Models and Factor Analysis*, Arnold.

Boudon, R. [non daté], *Paul Lazarsfeld*, Encyclopedia Universalis.

Cavin, S. [2008], Adorno, Lazarsfeld and The Princeton Radio Project, 1938-1941 *Paper presented at the annual meeting of the American Sociological Association Annual Meeting*, Sheraton Boston and the Boston Marriott Copley Place, Boston, MA, Jul 31, 2008.

Celeux, G. [1988], Classification et modèles. *Revue de Statistique Appliquée*, 36, 4, p.43-57.

Droesbeke, J.J., Lejeune, M. et Saporta G. (éditeurs) [2005], *Modèles sta-*

tistiques pour données qualitatives, Paris, Technip.

Goodman, L. A. [1974], Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models, *Biometrika*, 61, p.215-231.

Henry, N. [1999], Latent Structure Analysis at Fifty, *Joint Statistical Meetings*, Baltimore, http ://www.people.vcu.edu/ nhenry/LSA50.htm

Lautman, J. et Lécuyer, B.P. (éditeurs) [1998], *Paul Lazarsfeld (1901-1976), la sociologie de Vienne à New-York*, Paris, L'Harmattan.

Lazarsfeld, Paul F. [1959], Latent Structure Analysis, in *Psychology : A Study of a Science*, Vol. 3, S. Koch (ed.). New York : McGraw-Hill.

Lazarsfeld, P. F., et Henry, N. W. [1968], *Latent Structure Analysis*, Boston : Houghton Mifflin.

Lazarsfeld, Paul F. [1950a], The Logical and Mathematical Foundations of Latent Structure Analysis, dans Stouffer, S.A., et al. chapter 10, 362-412 *Measurement and Prediction, Volume IV of The American Soldier : Studies in Social Psychology in World War II.* Princeton University Press.

Lazarsfeld, Paul F. [1950b], Some Latent Structures, dans Stouffer, S.A., et al. chapter 11, p.413-472 *Measurement and Prediction, Volume IV of The American Soldier : Studies in Social Psychology in World War II.* Princeton University Press.

McCutcheon, A. L. [1987], *Latent Class Analysis*, Beverly Hills, Sage Publications.

Pollak, M. [1979], Paul F. Lazarsfeld, fondateur d'une multinationale scientifique, *Actes de la Recherche en Sciences Sociales*, 25, 1, p.45-59

Spearman, C. [1904], General Intelligence, Objectively Determined and Measured, *American Journal of Psychology*, 15, p.201-293.