

REVUE UNIVERSELLE
DES
MINES, DE LA MÉTALLURGIE

DES
TRAVAUX PUBLICS, DES SCIENCES ET DES ARTS
APPLIQUÉS A L'INDUSTRIE

LA STATISTIQUE MATHÉMATIQUE EN ANGLETERRE
LA LOI D'ERREUR de M. F. Y. EDGEWORTH

PAR

Edmond FRANCKEN,

Ingénieur principal aux Chemins de fer de l'Etat-Belge,

ET

Ernest MAHAIM,

Professeur à l'Université de Liège.

Depuis quinze à vingt ans, la statistique fait l'objet en Angleterre de travaux originaux, qui présentent un intérêt considérable au point de vue philosophique, aussi bien qu'au point de vue spécial de la statistique ou des mathématiques.

Chose singulière, les écrivains de langue française paraissent ne pas les connaître. Récemment encore, le regretté M. H. Laurent publiait un petit volume intitulé *Statistique mathématique* (1) plein d'aperçus intéressants d'ailleurs, mais qui ne contient pas la moindre allusion aux études des statisticiens anglais dont nous parlons.

Nous avons pensé que le moment était venu de rompre un injuste silence en faisant connaître au moins les

(1) Paris, O. Doin, 1908. 1 vol. in-16, iv-272 pages.

derniers résultats des recherches de l'un des premiers penseurs de la statistique anglaise, M. le professeur F. Y. Edgeworth.

Nous laissons à d'autres, plus compétents que nous, le rôle du critique. Nous nous contenterons d'exposer ce qui nous paraît essentiel.

Mais nous aurions fait tort à cet exposé même, si nous ne l'avions pas situé à sa place dans l'histoire de la statistique et dans celle du calcul des probabilités. C'est pourquoi nous commencerons par rappeler des faits et des notions dont il n'est pas inutile de montrer le développement.

I

L'histoire de la statistique présente des singularités, uniques peut-être dans l'histoire des sciences.

Conçue d'abord comme une science purement descriptive, destinée à renseigner les gouvernants et le public sur les ressources des Etats voisins, elle se borna pendant longtemps à exposer les « choses remarquables » des Etats. C'est ainsi que l'entendait Gottfried Achenwall, professeur à l'Université de Göttingen, quand il en inventa le nom, en 1749 (1). Il passait en revue l'Espagne, le Portugal, la France, la Grande-Bretagne, les Pays-Bas, la Russie, le Danemark et la Suède, et en faisait la description d'après un plan uniforme : les limites du territoire et leurs changements ; le sol, le climat, les fleuves, la nature du sol, sa répartition, le superflu et le manque de produits ; puis les habitants, leur nombre et leurs caractères propres ; le droit du souverain ; les Etats :

(1) Dans la préface à son ouvrage : *Abriss der neuesten Staatswissenschaft der heutigen vornehmsten europäischen Reiche*. La bibliothèque de l'Université de Liège en possède la quatrième édition : *Staatsverfassung der heutigen vornehmsten Europäischen Reiche im Grundrisse*. Göttingen, Vandenhoeck, 1762, 1 vol. n-8°, 552 pages.

noblesse et classes d'habitants; la constitution de la cour et du gouvernement; les lois et l'administration de l'église, de l'instruction et de la justice; l'industrie, le commerce intérieur et extérieur; les monnaies, les finances, les dettes, les forces de terre et de mer; enfin la politique, la vie politique et l'avenir du pays.

Tel est le cadre de toute étude statistique, notamment dans les Universités, jusque vers le premier tiers du dix-neuvième siècle.

On notera que cette « description » ne repose pas exclusivement, ni même principalement sur des nombres. D'une part, plus d'un objet n'en comportaient point (par exemple l'exposé de la constitution juridique du pays); d'autre part, on se contentait d'à peu près et de termes vagues pour les objets qui auraient pu être traduits en résultats numériques. Ainsi le voulait l'élégance de l'exposé et la tradition académique.

On remarquera encore que, pendant cette période, la statistique, fidèle à son étymologie (*statisticus*, de *status*, état; italien *statista*, homme d'état) est essentiellement et exclusivement une science politique, la science de l'Etat ou des Etats.

Or, dès la fin du dix-huitième siècle, le développement des sciences vint modifier profondément la situation. Une série de sciences nouvelles se fondent, qui s'installent sur le domaine de la statistique, s'y étendent, et finalement l'en expulsent.

C'est, avec Montesquieu, Kant (1796), Fichte (1796), la philosophie du droit qui s'empare de toute la partie juridique rationnelle; puis les auteurs de droit naturel, de droit public, de droit administratif se chargent d'exposer, d'expliquer les constitutions et les lois; la science politique ébauche les théories générales.

C'est, avec Adam Smith (1774), Jean Baptiste Say (1803) et Ricardo (1817), l'économie politique, qui réclame

bientôt, comme lui appartenant en propre, quantité de matières des plus importantes : tout ce qui touche au bien-être des Etats et de leurs habitants, y compris les finances publiques, la question monétaire, le commerce intérieur et extérieur.

C'est surtout la géographie, qui s'érige en rivale. En 1812 paraît la grande Géographie française de Malte-Brun, bientôt suivie, en Allemagne, de celle de Ritter (1817). Tout ce qui touche au sol, au territoire des Etats, et même les fameuses « choses remarquables » en général deviennent l'objet de la géographie. Ce qui, aujourd'hui encore, ressemble le plus à un ancien traité de statistique, c'est un traité de géographie à la manière d'Elisée Reclus.

De sorte qu'il vint un temps où la statistique n'eut plus d'objet propre : des spécialistes, à côté d'elle, lui avaient ravi successivement toutes les parties de son enseignement, les traitaient avec des méthodes appropriées, et en poursuivaient progressivement l'explication scientifique, tandis que la vieille statistique se contentait du procédé superficiel de la description. Il arriva alors ce qui devait arriver : elle fut abandonnée et elle disparut même des programmes des universités (1).

Chose étonnante, le nom en a survécu, pour s'appliquer à une science toute différente, qui d'une part, n'est plus une science politique, et d'autre part, repose essentiellement sur des données numériques.

(1) La statistique figure au programme de nos facultés de droit en Belgique dès la fondation des Universités de l'Etat (1816); maintenue par la loi organique du 27 septembre 1835, elle fut rayée par la loi du 15 juillet 1849, malgré les protestations de Heuschling, d'Omalius d'Halloy (Sénat, séance du 14 juillet 1849) et de l'Académie (séance publique du 16 décembre 1849. *Annuaire* de 1850, p. 116). C'est l'arrêté royal du 2 octobre 1893 qui l'a inscrite de nouveau au programme des cours : elle figure comme cours obligatoire à la licence en sciences administratives, et, depuis 1901, à la licence en sciences commerciales.

Deux ordres de circonstances ont contribué à opérer cette transformation, d'une façon insensible, mais très facile à discerner dans l'histoire : ce sont les progrès de la statistique officielle et certaines recherches scientifiques relatives aux lois des phénomènes de la population. Voici en peu de mots leur rôle respectif.

A coup sûr, l'administration, dans tous les temps, eut besoin de renseignements précis et périodiques, et l'on n'a pas de peine à trouver à Rome, en Grèce, chez les Hébreux, chez les Babyloniens et jusque chez les Chinois, des recensements, des relevés numériques de toute espèce. Mais, à partir de la formation des Etats modernes, ce besoin devient une nécessité. Plus le territoire s'étend, plus la machine administrative se complique, plus surtout les services publics, l'armée, les finances par exemple prennent de l'importance, et plus la connaissance précise de l'état de situation des peuples et des richesses est indispensable. Aussi, les monarques absolus qui s'entendaient aux affaires, comme Frédéric II, veulent avoir des rapports périodiques sur tous les objets qui peuvent les renseigner sur l'exécution des mesures administratives.

Avec la Révolution française, qui crée une nouvelle administration, un nouveau système de gouvernement, nous voyons croître encore les exigences. L'Assemblée Constituante demande à Lavoisier un « aperçu général sur la richesse territoriale et les revenus de la France » en 1790 (1). Necker organise un « Bureau de renseignements » qui est l'embryon des bureaux de statistique

(1) L'Assemblée nationale a fait imprimer en 1791 une partie de l'ouvrage, sous le titre : « Résultats extraits d'un ouvrage intitulé : *De la richesse territoriale du Royaume de France*, ouvrage dont la rédaction n'est point encore achevée ; remis au Comité de l'Imposition par M. Lavoisier... imprimé par ordre de l'Assemblée nationale. Paris. Imprimerie Nationale 1791. » 1 vol. in-8, 48 p.

officielle, et dès le commencement du dix-neuvième siècle, cette institution devient définitive. En 1804, Napoléon fait publier une Statistique générale de la France, d'après des mémoires statistiques dressés par les préfets. Successivement, tous les autres Etats imitent cet exemple : chacun a désormais son bureau de statistique, dont la mission d'information se concrétise en des relevés numériques. A l'envi, on s'efforce à perfectionner les méthodes et l'exposé. Les recensements généraux de la population, d'abord occasionnels, deviennent périodiques. Au début du dix-neuvième siècle, tous ces relevés ne sont pas encore exclusivement obtenus par des méthodes directes, touchant les faits eux-mêmes : on s'adresse aux maires des communes, aux personnes compétentes, aux experts, pour obtenir les nombres. Mais, de plus en plus, grandit le prestige du chiffre et le besoin d'exactitude. Si nous faisons abstraction du recensement de 1830, ordonné par le roi Guillaume des Pays-Bas sous l'inspiration de Quételet, c'est en Belgique, en 1846, que se place le premier recensement de la population obtenu d'après des questionnaires adressés aux chefs de ménages. Sur l'initiative de Quételet encore, les chefs de bureaux de statistique de tous les Etats prennent contact, se réunissent en Congrès en 1853, et voici qu'ils discutent classification, procédés, tableaux, moyennes, toute une technique formelle qui constitue une science et qui est indépendante des résultats objectifs. On appelle cela encore de la statistique (1), mais il est clair qu'on n'a plus affaire à l'antique discipline d'Achenwall.

(1) Quételet lui-même publie en 1828 des *Recherches statistiques sur le Royaume des Pays-Bas* qui ont encore le cadre des anciens traités : « Etendue du Royaume des Pays-Bas ; de la population ; des impôts et du commerce ; de la librairie et des journaux ; de l'instruction et des institutions de bienfaisance ; des crimes et des délits ; examen comparatif des différentes parties du royaume. »

Parallèlement à cet ordre de travaux, se développent des recherches scientifiques extrêmement intéressantes qui vont finir par s'y rattacher. Ce sont celles relatives aux lois de la population.

On trouve, paraît-il, des formes d'assurance sur la vie dès le moyen-âge, pour les gens qui font des voyages en mer et les pèlerins (1). Au milieu du dix-septième siècle, les Hollandais connaissaient déjà les principes de l'assurance : en 1671, Jean de Witt présentait aux États de Hollande et de West-Frise un calcul de la valeur des rentes sur la vie, qui, au dire de Bierens de Haan, est « le premier essai de table de mortalité et la première application des principes du calcul des probabilités à la notion de chance de vie ». Il semble d'ailleurs qu'elle n'était pas basée sur des relevés statistiques, mais sur des hypothèses (2). Il n'en est pas de même de la table de mortalité qui est due à Edmond Halley, le célèbre astronome (1656-1742), et qui date de 1693. Elle était construite d'après des relevés de décès pris dans les registres paroissiaux de la ville de Breslau et communiqués par un allemand, Gaspar Neumann, à la Royal Society de Londres (3).

Dès ce moment, l'attention est attirée sur les lois de la population. Un brave ecclésiastique allemand, Johann

(1) V. Meitzen : *Geschichte, Theorie und Technik der Statistik*, 2^e Aufl. Berlin. 1903. p. 17.

(2) M. le professeur Le Paige a bien voulu nous communiquer la réimpression faite en 1879 par la Société des sciences d'Amsterdam de cet ouvrage, qu'on considéra longtemps comme perdu. Son titre est : *Waerdye van Lyfrente naer proportie van Los renten*. In 's Graven-Hage, by Jacobus Scheltus... Anno 1671. 1 vol. in-folio (0,335 × 0,21). 24 pp.

(3) *An estimate of the Degrees of the mortality of mankind, drawn from curious Tables of the Births and Funerals of the City of Breslaw ; with an Attempt to ascertain the Price of Annuities upon Lives*. t. XVII (1693) des *Philosophical Transactions*. London.

Peter Süssmilch, y consacre en 1740, un livre (1), où il prouve « l'ordre divin qui préside aux vicissitudes de la race humaine dans la naissance, la mort et la reproduction ». Il voit dans la Genèse la révélation de l'ordonnance providentielle qui fait que sur 21 garçons il naît 20 filles, et que les individus des deux sexes sont en nombre égal à l'âge du mariage : ce qui est pour lui la preuve que la monogamie est commandée par Dieu. Il met en évidence certaines régularités dans les proportions de décès d'après l'âge. Bref, il pose le problème des lois qui ne va plus disparaître des préoccupations scientifiques et sur lequel le progrès du calcul des probabilités et la pratique des assurances vont jeter une lumière étonnante.

Assurances et probabilités ne se comprennent que si l'on a à sa disposition des données numériques tirées de dénombrements préalables. C'est là le lien qui va rattacher les mathématiques à la statistique, notamment à la statistique officielle. Celle-ci offre non seulement des relevés numériques, mais portent sur des *grands nombres*. Dès lors, la matière est fournie aux applications des formules (2).

Tandis que tous les grands mathématiciens depuis Pascal s'attachent à faire faire des progrès au calcul des probabilités, on voit les savants faire de plus en plus

(1) *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der Geburt, dem Tode, und der Fortpflanzung desselben erwiesen*. Berlin. 1740.

(2) M. Mansion, dans son beau discours comme directeur de la classe des sciences et président de l'Académie (16 décembre 1903) *Sur la portée objective du calcul des probabilités* a très bien saisi ce lien quand il a indiqué la fréquence et le nombre des dénombrements comme la cause et l'occasion de l'application du calcul des probabilités à la statistique. « Dans les pays civilisés, le dénombrement exact des naissances, des décès, des mariages, des délits, des crimes et d'une foule d'autres faits sociaux a permis de démêler certaines lois et d'en tirer des conséquences théoriques et pratiques ». *Bulletin de l'Académie (Classe des sciences)*, 1903, p. 1263).

emploi de relevés statistiques, en affiner et perfectionner à la fois la technique et la théorie.

Alors, apparaît Quételet, qui est le véritable père de la statistique moderne. Nous avons déjà signalé son influence sur la statistique officielle. Il n'aurait eu d'autre mérite que d'en avoir amélioré les procédés et d'avoir commencé par ses fameux Congrès officiels l'œuvre d'unification internationale, que son action eût déjà été d'importance.

Mais il a fait bien davantage. Le premier, il a eu une vue d'ensemble d'une sociologie, c'est-à-dire d'une science de la société, et il l'a appelée *Physique sociale* parce qu'il croyait que les régularités que l'on constate dans les phénomènes sociaux étaient l'expression de lois physiques, semblables à celle du monde inorganique.

Il généralise et systématise une conception de la nature et de la société humaine qu'on n'avait eue que partiellement avant lui, et cette conception repose sur le calcul des probabilités.

Pour Quételet, les individus sont pour ainsi dire, des approximations. Les différents caractères mesurables (et presque tous finiront par être mesurables) présentent chez chaque individu des écarts par rapport à une valeur moyenne, idéale, que la nature cherche de plus en plus à réaliser. Les tailles des conscrits, les circonférences de la poitrine des soldats écossais, par exemple, sont exactement la même chose que les mesures répétées qu'un observateur prendrait de la taille, de la circonférence thoracique d'une statue. De même que ces mesures répétées présenteraient des écarts, se groupant autour d'une valeur moyenne, d'autant plus nombreux qu'ils s'en rapprocheraient, d'autant plus rares qu'ils s'en éloigneraient, de même les tailles et les mesures individuelles se groupent autour d'une valeur moyenne en d'autant plus grand nombre qu'elles s'en rapprochent.

Les nains et les géants sont des « erreurs de la nature ». L'homme normal, est l'homme du type le plus fréquent : c'est celui qui se rapproche de l'idéal auquel tend la nature.

Sans doute cette conception méthaphorique (1) sur laquelle nous revenons plus loin, n'est pas propre à Quételet. Les lois des phénomènes sociaux avaient été, nous l'avons vu, pressenties avant lui, et avant lui aussi, bien d'autres avaient appliqué le calcul à la répartition de faits moraux ou sociaux autour d'une moyenne. Mais personne avant Quételet n'avait conçu la méthode comme un *système général* de nature à constituer une mécanique ou une physique sociale.

Nous n'avons pas à rechercher ici tous les défauts de sa conception de la société humaine, ni de sa Physique sociale (2). Ce qui nous intéresse, c'est de constater que Quételet a transformé la statistique elle-même.

Il est clair, en effet, que si cette méthode était légitime, elle pouvait s'étendre à d'autres objets qu'aux dénombrements relatifs à l'homme.

Dans son *Anthropométrie*, on aperçoit déjà une extension remarquable de la méthode. Il y groupe des études sur un grand nombre de caractères mesurables de l'Homme, de l'homme comme tel, et non comme être social. Il est évident que, déjà ici, la statistique ne peut être considérée comme une science politique.

En effet, à partir de Quételet, la destinée de la statis-

(1) Dans l'une de ses *Lettres sur la théorie des probabilités*, on voit très nettement la naissance de la métaphore : il y compare un grand nombre de statuaires copiant une statue à des observateurs mesurant la circonférence thoracique d'un grand nombre d'hommes. L'exemple en est très joliment amené. V. *Lettres*, p. 213.

(2) J. Bertrand, dans son *Calcul des probabilités*. Paris, 1889, pp. xli-lxiii, a accablé de « sarcasmes spirituels » comme dit M. Mansion (disc. cité p. 1238) sa conception de l'homme moyen.

tique est complètement modifiée. Elle n'est plus, à proprement parler, une science, mais un procédé scientifique qui consiste essentiellement à tirer des conclusions (des coefficients, des régularités, des lois) de dénombrements de faits concrets et nombreux.

Dès lors, loin d'être uniquement une science d'Etat, elle pénètre dans une foule de domaines scientifiques qui lui paraissaient étrangers : la voici en météorologie, la voici en médecine (1), la voici en physiologie, en zoologie, en botanique. Partout où l'on aura des relevés numériques de faits concrets qui pourront être assimilés à des erreurs d'observation, découlera une série de conclusions scientifiques. Partout, il y aura des règles à observer pour faire correctement un dénombrement, pour obtenir la signification d'une moyenne. Bref, une méthode générale est constituée, quel que soit l'objet concret auquel elle s'applique.

La théorie de l'évolution en science naturelle a été une des premières à profiter de ces progrès.

Les procédés que Quételet employait pour rechercher le *type* de l'homme, devaient fatalement convenir au *type* des différentes espèces. En outre, la notion d'hérédité se prêtait aussi à des recherches mathématiques, basées toujours sur des dénombrements.

Ce sont des études de ce genre qui ont attiré, notamment en Angleterre, l'attention d'un groupe important de savants. Francis Galton, l'ami de Herbert Spencer, paraît avoir été un des premiers à se livrer à ces problèmes. Celui qu'il appelle son « coadjuteur », le professeur Karl Pearson, y a voué une activité sans relâche. Depuis 1894, les *Philosophical Transactions* contiennent un nombre considérable de « contributions mathématiques » à l'étude de l'évolution.

(1) Gavarret, dès 1840, écrit des *Principes généraux de statistique médicale*.

Bien plus, une revue s'est fondée, qui se donne pour programme la mesure de la vie ou des phénomènes biologiques : la revue *Biometrika*, de Cambridge (1).

C'est là que s'est développée, par de nombreuses et journalières applications, une méthode de plus en plus variée, complexe et riche, dont la portée dépasse à coup sûr le champ où elle se cantonne, puisque les actuaires commencent à l'employer en matière d'assurances sur la vie (2).

Nous n'entreprendrons pas, dans ce travail, d'exposer l'ensemble de ces recherches, qui mériteraient une étude approfondie. Elles sont d'ailleurs trop intimement liées à l'évolution naturelle pour que l'on puisse aisément les exposer sans les y rattacher, ce qui nous entraînerait fort loin. Notre intention est de nous borner aujourd'hui à l'œuvre d'un seul représentant de l'Ecole anglaise moderne, M. F. Y. Edgeworth, professeur d'économie politique à l'Université d'Oxford.

M. Edgeworth est célèbre en économie politique par l'application qu'il y a faite des mathématiques, après Jevons, Cournot, Walras. Son livre « *Mathematical Psychics* » (1881) est classique sous ce rapport.

Son œuvre en théorie statistique n'a pas moins d'importance. Depuis 1883, en effet, il a publié dans divers recueils de nombreuses études qui paraissent avoir abouti récemment à une conclusion d'une portée tout à fait

(1) *Biometrika, a journal for the statistical study of biological problems, edited in consultation with FRANCIS GALTON, by W. F. R. WELDON, KARL PEARSON, and C. B. DAVENPORT.* Cambridge, The University Press. 8° (1^{er} numéro, octobre 1901).

(2) M. W. Palin Elderton, à la demande de l'Institut des actuaires anglais, a consacré tout un volume aux courbes de M. Pearson et à leur emploi dans les assurances : *Frequency Curves and Correlation.* London, S.-D. (1906) (Institute of Actuaries). 1 vol. 8°, XIII-172 pp.

générale. Sa théorie, en effet, concerne toutes les données statistiques qui satisferont à certaines conditions mathématiques définies. C'est dire qu'il ne s'agit de rien moins que d'une loi nouvelle, dont le domaine paraît être très étendu.

Mais pour en faire voir l'intérêt, il est nécessaire que nous remontions aux notions mathématiques dont elle est la conséquence.

II

On sait que c'est Gauss qui a donné l'expression généralement connue de la loi des erreurs accidentelles. Sous certaines réserves, elle peut être considérée comme une extension du théorème de Bernoulli.

Il importe de mettre brièvement cette filiation en évidence.

Considérons, pour fixer les idées, une urne qui renferme p_1 boules blanches et q_1 boules noires.

On en tire *au hasard* une boule : la probabilité qu'elle soit blanche est donnée par la fraction $p = \frac{p_1}{p_1 + q_1}$, qu'elle soit noire par la fraction $q = \frac{q_1}{p_1 + q_1}$.

La boule est remise dans l'urne et on répète le tirage : les probabilités de tirer une boule blanche ou bien une boule noire sont les mêmes que tout à l'heure.

Mais considérons l'ensemble des deux tirages et demandons-nous quelles sont *a priori* les probabilités de tirer successivement 2 boules blanches, 2 boules de couleurs différentes ou bien 2 boules noires. On trouve aisément la réponse. Ces probabilités sont :

$$p^2, \quad 2pq, \quad q^2.$$

Si on considérait une répétition de trois tirages, les

probabilités de tirer 3 boules blanches, 2 blanches et 1 noire, 1 noire et 2 blanches, ou bien 3 noires seraient :

$$p^3, \quad 3 p^2 q, \quad 3 p q^2, \quad q^3.$$

La généralisation de ces résultats s'établit aisément : dans le cas de m tirages répétés, les probabilités d'amener 0 blanche, 1 blanche et $m - 1$ noires, 2 blanches et $m - 2$ noires... i blanches et $m - i$ noires... seraient, en représentant, comme d'habitude, par $C_m^1 \dots$, $C_m^2 \dots$, $C_m^i \dots$, les combinaisons simples de m lettres 1 à 1, 2 à 2..., i à $i \dots$:

$$q^m, \quad C_m^1 q^{m-1} p, \quad C_m^2 q^{m-2} p^2, \quad C_m^i q^{m-i} p^i.$$

On reconnaît dans cette suite les termes consécutifs du développement du binôme newtonien $(q + p)^m$ et on pourra appeler « loi binomiale » la relation :

$$y_i = C_m^i q^{m-i} p^i = \frac{m!}{i! (m-i)!} q^{m-i} p^i \quad (1)$$

qui exprime la probabilité y_i d'amener en m épreuves, i boules blanches (de probabilité élémentaire p) et $m - i$ noires (de probabilité élémentaire q).

Le calcul montre que la valeur de y_i est d'autant plus grande que le rapport $\frac{i}{m-i}$ est plus voisin de $\frac{p}{q}$ ou de son égal $\frac{p_1}{q_1}$ ou encore que i est plus voisin de mp .

En particulier si $p_1 = q_1$ et si par suite $p = q$, y_i sera maximum pour $i = \frac{m}{2}$ si m est pair ou bien pour $i = \frac{m-1}{2}$ et pour $i = \frac{m+i}{2}$ si m est impair.

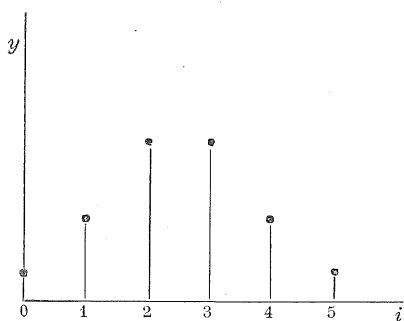
Par exemple si $p = q = \frac{1}{2}$ et si $m = 5$, on a :

$$y_0 = \frac{1}{32}, \quad y_1 = \frac{5}{32}, \quad y_2 = y_3 = \frac{10}{32}, \quad y_4 = \frac{5}{32}, \quad y_5 = \frac{1}{32}$$

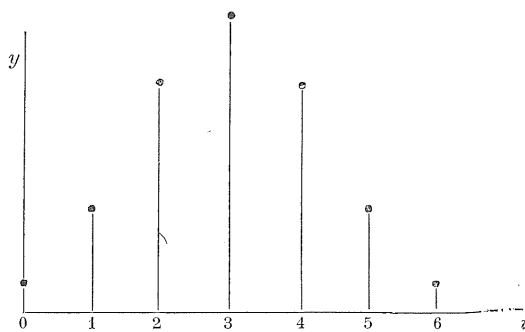
La sortie de 2 blanches et 3 noires et la sortie de 3 blanches et 2 noires sont également probable ; leur probabilité est $\frac{10}{32}$ et elle est supérieure à celle d'amener toute autre combinaison.

Pour $p = q$ et $m = 6$, on aurait :

$$y_0 = \frac{1}{64}, y_1 = \frac{6}{64}, y_2 = \frac{15}{64}, y_3 = \frac{20}{64}, y_4 = \frac{15}{64}, y_5 = \frac{6}{64}, y_6 = \frac{1}{64}$$



Diagr. 1.



Diagr. 2.

Ces résultats peuvent être traduits par les diagrammes (1) et (2), où on a porté en abscisses les valeurs discontinues de i et en ordonnées celles de y . Les sommets des ordonnées esquissent une courbe que nous allons bientôt voir se dessiner complètement.

Quand m est grand par rapport à l'unité, la valeur (1) de y se prête mal au calcul. Dans ce cas, l'emploi de la formule de Stirling qui donne comme valeur approchée du factoriel de n , quand n est suffisamment grand

$$e^{-n} n \sqrt{2 \pi n}$$

permet d'écrire une expression de y , qui en diffère fort peu et qui est d'un maniement aisé.

Appelons i_0 le nombre de boules blanches de la combinaison qui a le plus de chances de sortir : nous le savons, i_0 diffère peu de mp . Soit h la différence entre i_0 et le nombre i des boules blanches d'une autre combinaison quelconque. Le nombre h s'appelle l'écart de cette combinaison et il la définit. A l'écart $+5$ correspond la combinaison qui comprend $i_0 + 5$ boules blanches ; à l'écart -5 celle qui en comprend $i_0 - 5$ et nous pourrions parler indifféremment de la probabilité de la combinaison qui comprend $i + h$ boules blanches ou de la probabilité de l'écart h .

Ceci posé, l'analyse établit que cette dernière probabilité, sous réserve que h soit petit par rapport à $\sqrt{m}$, est à très peu de chose près égale à

$$\frac{1}{\sqrt{2\pi mpq}} e^{-\frac{h^2}{2mpq}}$$

Enfin, et cette dernière formule va résumer pour nous la portée du théorème de Bernoulli, la probabilité composée d'un écart compris entre des limites o et α comprenant les entiers $h_1, h_2, h_3, \dots, h_\alpha$, c'est-à-dire la somme des probabilités des écarts $[h_1, h_2, \dots$ et h_α ou encore la probabilité d'un écart ne dépassant pas α est très approximativement fournie par l'intégrale définie :

$$\frac{1}{\sqrt{2\pi mpq}} \int_0^\alpha e^{-\frac{z^2}{2mpq}} dz.$$

La démonstration de cette formule suppose, nous l'avons dit, que h soit petit par rapport à $\sqrt{m}$. Mais on peut écarter cette restriction en observant que les éléments de l'intégrale qui ont en exposant de grandes valeurs de z sont sensiblement nuls et qu'il en va de même des probabilités partielles correspondantes qui figurent dans la somme à calculer.

On a pu conserver avec quelque raison le nom de loi binomiale à la fonction

$$y = \frac{1}{\sqrt{2\pi mpq}} e^{-\frac{h^2}{2mpq}}$$

et à sa forme générale

$$y = A e^{-\frac{h^2}{c^2}}.$$

La courbe correspondante, en forme de cloche, est dite *courbe binomiale*, mais plus souvent *courbe normale* des écarts.

La probabilité d'un écart compris entre x et $x + dx$ est représentée par le trapèze élémentaire $AabB$.

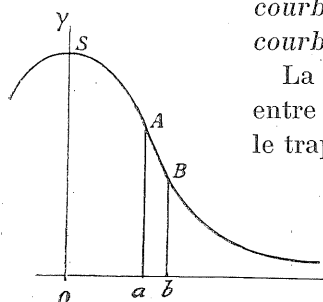


Fig. 1.

La surface totale comprise entre la courbe et l'axe des x représente la probabilité d'un écart compris entre

$-\infty$ et $+\infty$, c'est-à-dire la certitude. On doit donc avoir :

$$A \int_{-\infty}^{+\infty} e^{-\frac{h^2}{c^2}} dh = 1.$$

Or, on sait que
$$\int_{-\infty}^{+\infty} e^{-\frac{h^2}{c^2}} dh = c \sqrt{\pi}.$$

Il s'en suit qu'on doit prendre :

$$A = \frac{1}{c \sqrt{\pi}}.$$

Le comte de Buffon, déjà en 1777 (1), dans son *Arith-*

(1) V. *Essai d'Arithmétique morale*. Suppl. à l'*Histoire naturelle*, t. IV. Paris 1777.

métique morale et après lui un grand nombre de savants ont soumis le théorème de Bernoulli à des vérifications expérimentales : on a toujours constaté que les événements confirmaient les conclusions des calculs de la façon la plus satisfaisante.

Jusqu'ici, l'évènement, dont nous avons envisagé la probabilité, et les événements conjointement possibles étaient nettement définis, d'une manière très spéciale et particulièrement simple, l'un et les autres étaient des événements composés, par voie d'addition, d'un même nombre m , fixé une fois pour toutes, d'événements simples contradictoires, a (la sortie d'une boule blanche) ou b (la sortie d'une boule noire) et dont les probabilités p et q restaient constantes au cours des m répétitions.

Nous allons maintenant considérer un genre d'événements bien autrement compliqués, les erreurs accidentelles inhérentes à la mesure, à l'aide d'instruments physiques, d'une grandeur concrète. L'erreur en question est bien encore la résultante d'un groupe d'erreurs partielles, mais le nombre de celles-ci n'est plus connu, ni même constant, leurs probabilités sont variables et inconnues ; au surplus, les erreurs élémentaires peuvent se combiner différemment que par voie d'addition. Combien, dans ces nouvelles conditions, la détermination de la probabilité de l'erreur — qui correspond à l'écart h de tout à l'heure — constitue un problème mal déterminé et d'un abord malaisé pour l'analyse mathématique, il n'est guère besoin de le dire.

Gauss, dans ses *Disquisitiones Arithmeticae*, n'a pas hésité à l'aborder de front et, chose bien faite pour étonner au premier abord, il a retrouvé comme solution précisément la fonction de la loi de Bernoulli, $y = A e^{-\frac{h^2}{c^2}}$.

La vérification était facile. Pour juger la formule,

il suffisait de répéter un grand nombre de fois la mesure d'une grandeur connue, en prenant soin, bien entendu, d'écarter les erreurs systématiques.

Si la loi binomiale s'étend à ces événements, comme à ceux envisagés par Bernoulli, les mesures trouvées devront présenter avec la vraie grandeur en observation des écarts, disons des erreurs, qui se grouperont conformément à ses indications.

C'est, assure-t-on, ce que les physiciens et surtout les astronomes ont pu souvent constater.

En particulier, on cite les 470 observations sidérales de Bradley que Bessel a contrôlées. A titre d'exemple, voici quelques chiffres empruntés à cette expérience devenue célèbre. Il s'agit de l'ascension droite de l'étoile polaire déterminée de 1813 à 1815 à l'observatoire de Königsberg. Les chiffres observés sont mis en parallèle avec ceux que fournit la fonction

$$y = A e^{-\frac{h^2}{c^2}}$$

quand on y fait

$$c = \frac{1}{0.43539} \quad \text{et} \quad A = \frac{1}{c\sqrt{\pi}}.$$

Erreurs.	Nombre des erreurs.	
	Réel.	Calculé.
0''0 à 0''4	25	24,9
0''4 à 0''8	22	21,9
0''8 à 1''2	19	18,2
1''2 à 1''6	11	13,7
1''6 à 2''	9	9,5

Les concordances ne sont pas toutes aussi parfaites, mais toujours elles ont été jugées très suffisantes et les

observateurs n'ont pas cessé d'appliquer la loi binomiale, devenue *la loi des erreurs* de Gauss.

Cependant l'illustre géomètre, sans douter de la vérité de la loi qu'il avait découverte et qu'un succès constant confirmait chaque jour, se rendait parfaitement compte de la fragilité des raisonnements sur lesquels il l'avait fondée. La preuve en est dans les efforts qu'il dépensa pour asseoir sur une base indépendante et plus solide les conséquences tirées de sa loi.

M. J. Bertrand a mis en évidence les points faibles de la démonstration de Gauss (1).

Première objection : la loi de Gauss repose sur le postulat de la moyenne : on admet qu'en présence de m observations, toutes dignes de la même confiance, le parti le plus sage est de prendre leur valeur moyenne. Souvent ce parti sera le meilleur. Cependant, objecte M. Bertrand, si parmi les m observations la plupart sont étroitement groupées et si une ou deux — sans raison apparente, bien entendu — restent isolées, loin des autres, est-il bien sûr qu'il conviendra de prendre la moyenne de toutes ces données et qu'il ne vaudrait pas mieux abandonner les valeurs sporadiques pour ne tenir compte que du faisceau des autres ?

Ainsi le fondement de la démonstration, bien qu'acceptable d'une façon générale, est sujet à des exceptions qui ne pourront pas manquer de retentir dans la portée de la loi.

Deuxième objection : Gauss, sans en avertir, admet que la probabilité d'une erreur x est une fonction de x , indépendante de la grandeur z de l'objet à mesurer. Cette hypothèse est fort peu vraisemblable *a priori*.

Troisième objection, peut-être la plus grave : Gauss, au cours de sa démonstration, écrit que la valeur moyenne

(1) V. *Calcul des probabilités*. Paris. 1889.

de m observations d'une grandeur z est la valeur la plus probable de z . C'est trahir le postulat fondamental : d'après celui-ci, la moyenne est la valeur probable de z et non pas la valeur la plus probable. La distinction est d'importance : une urne renferme 14 pièces d'argent de 1/2 franc et 5 pièces d'or de 10 francs. Je puis prendre une pièce au hasard. Mon gain probable est $\frac{14 \times 1/2 + 5 \times 10}{19} = 3$ francs, tandis que mon gain le plus probable est de 1/2 franc seulement.

M. H. Poincaré, dans ses belles leçons professées à la Sorbonne en 1893 et 1894, a plaidé de son mieux, comme il le dit, en faveur de la loi de Gauss. Elle ne pouvait pas, à notre époque, rencontrer un plus sûr et plus brillant défenseur.

Pour ce qui est du postulatum de la moyenne, M. Poincaré concède que celle-ci n'est pas, en tous les cas, la valeur probable, mais il démontre que, si le nombre des mesures s'accroît de plus en plus, leurs moyennes successives sont de plus en plus probables.

M. Poincaré, en tenant compte des autres objections de M. Bertrand, établit par une analyse des plus pénétrante et qui ne semble plus laisser de prise à la critique, que l'erreur commise avec un instrument d'observation suivra la loi de Gauss si elle satisfait aux conditions suivantes :

1° Elle doit être la résultante d'un très grand nombre d'erreurs partielles et indépendantes ;

2° Chacune des erreurs partielles doit être très petite ;

3° Pour les erreurs partielles, la probabilité d'une erreur positive doit être la même que celle d'une erreur négative égale en valeur absolue.

Sous ces réserves, peu importe la ou les lois de fréquence suivies par les erreurs élémentaires.

En résumé, à condition de réserver le nom d'erreurs accidentelles à celles qui satisfont aux trois conditions que nous venons d'énoncer, la loi de Gauss peut être considérée comme rigoureusement démontrée (1).

III.

La loi de Gauss a été formulée en vue des erreurs accidentelles de l'observation d'une quantité, généralement concrète ou réelle.

Comment en est-on venu à l'appliquer à la statistique, c'est-à-dire à comparer *plusieurs* quantités réelles à des erreurs d'observation d'une seule quantité abstraite ou idéale ?

Il est tout d'abord bon de rappeler que déjà Jacques Bernoulli avait été attiré par la question de l'emploi du calcul des probabilités à d'autres objets que les jeux de hasard.

(1) Il est intéressant d'observer ici que Bravais, l'illustre cristallographe, avait eu l'intuition des raisons profondes qui confèrent à la loi binomiale son extraordinaire portée et que l'analyse de M. Poincaré a si bien dégagées.

Dans une correspondance avec Quételet, que celui-ci a reproduite en annexe à ses *Lettres sur la théorie des probabilités*, Bravais donne un exemple frappant de la tendance à s'équilibrer que présentent naturellement des chances inégales, à condition qu'elles soient amenées à se combiner un grand nombre de fois. Considérons un jeton. Si le jeton était parfaitement homogène et si le joueur qui le lance était exempt de tout mouvement intentionnel ou habituel, la probabilité de ramener la face, découverte avant le jeu, serait de $\frac{1}{2}$ tout comme celle d'amener la face opposée. Mais admettons qu'une cause perturbatrice quelconque donne $\frac{2}{3}$ de chance au retour de la face découverte contre $\frac{1}{3}$ au renversement de la pièce. Si en commençant le jeu, elle présente sa face, la probabilité d'amener face sera, au premier coup, de $\frac{2}{3}$, au second coup de $\frac{5}{9}$, au troisième coup de $\frac{14}{27}$, au quatrième de $\frac{41}{81}$, etc... Ainsi au premier coup, face avait deux fois plus de chances, que pile ; au quatrième déjà, son avantage n'est plus à peine que de $2\frac{1}{2}\%$; après quelques coups encore, il serait sensiblement réduit à rien.

Son *Ars conjectandi* (1) se divise en quatre parties : la première n'est qu'un commentaire du traité de Christ. Huyghens sur « les raisonnements dans le jeu de hasard » ; la seconde, la théorie des combinaisons et des permutations ; la troisième, des applications de celle-ci à toutes sortes de jeux de hasard. Mais la quatrième porte le titre suggestif : « de l'emploi et de l'application de la doctrine précédente dans les matières civiles, morales et économiques » (2).

Rien qu'aux préliminaires, où il pose une série de définitions importantes, on s'aperçoit que l'auteur commence une partie essentielle de son œuvre. C'est là qu'il fait, le premier, la distinction bien nette entre la détermination de la probabilité *a priori* et celle *a posteriori*, et se pose le problème « singulier » d'où il fait sortir la démonstration de son fameux théorème (3). Malheureusement, la mort a interrompu son œuvre, et l'on ne saura jamais comment il entendait faire ses applications aux « choses économiques », surtout à l'aide des moyens limités que lui fournissait la seule théorie des combinaisons.

(1) Nous devons à l'aimable obligeance de M. le professeur Le Paige la communication d'un exemplaire de l'édition princeps de ce livre célèbre, qu'on ne trouve ni à la Bibliothèque de l'Université de Liège, ni à la Bibliothèque Royale : *Jacobi Bernoulli Ars Conjectandi, opus posthumum*. Basileae impensis Thurnisiorum fratrum 1715. 1 vol. in-8°. (0,17 × 0,20) 306 et 35 pp.

(2) *Pars quarta, tradens usum et applicationem precedentis Doctrinae in civilibus, moralibus et Economicis* (pp. 210 à 239 de l'édition citée).

(3) Intitulé du chap. IV : *De duplici modo investigandi numeros casuum. Quid sentiendum, de illo qui instituitur per experimenta. Problema singulare eam in rem propositum*. Le problème est le suivant : Montrer qu'on peut faire autant d'épreuves que, étant donné autant de fois qu'on veut, il devienne plus vraisemblable que le nombre des observations favorables soit compris entre des limites (assignées).

On remarquera déjà la généralité du terme *observations*.

Il fallut cependant les progrès accomplis par les travaux de De Moivre (1), Stirling (2), Mac Laurin (3) et Euler (4) pour permettre à Laplace de donner au théorème de Bernoulli la forme analytique commode que nous avons rappelée plus haut (5).

Dès lors, la voie était ouverte pour assimiler les écarts d'une valeur quelconque qui se groupent autour d'une moyenne à des erreurs d'observation, surtout quand Poisson eut donné, sous le nom de *Loi des grands nombres*, la généralisation du théorème de Bernoulli au cas où les probabilités des événements changent d'une épreuve à l'autre (6).

C'est alors que la loi des erreurs de Gauss vint confirmer et préciser la règle depuis longtemps entrée dans la pratique que, pour avoir une mesure exacte, il fallait répéter les mesurages et en prendre la moyenne.

Or, dans les coefficients statistiques qui sont plus ou moins stables, tels que le rapport des naissances masculines aux naissances féminines, le rapport des décédés aux survivants d'un âge donné, etc., on avait une grandeur dont beaucoup d'autres s'écartaient plus ou moins symétriquement d'une moyenne, comme les erreurs d'observation. En outre, la fixité du coefficient se mon-

(1) *The doctrine of chances*. London 1718, et *Miscellanea Analytica de Seriebus et Quadraturis*, London 1730.

(2) *Methodus differentialis*... London. 1730.

(3) *Treatise of Fluxions*. Edinburg. 1742.

(4) *Institutiones calculi differentialis*. Petrop. 1755.

(5) V. la suite des raisonnements, des procédés, et la part de chacun de ces illustres géomètres dans E. Czuber : *Die Entwicklung der Wahrscheinlichkeitstheorie und ihrer Anwendungen*. Leipzig. 1897. Teubner. pp. 71-78.

(6) *Recherches sur la probabilité des jugements*. Paris 1837, chap. IV : Suite du calcul des probabilités de très grands nombres. (Cas des chances variables d'une manière quelconque, comprenant celui des chances constantes), pp. 246-317.

trait d'autant mieux que l'on avait affaire à de plus grands nombres d'observations. L'assimilation était aisée ; on prit souvent ces données comme bases d'exemples ou d'applications, et l'ont vint naturellement à les traiter de la même manière, sans se rendre compte peut-être de la différence entre les deux cas. Quand il s'agit de l'observation d'une mesure, par exemple, il s'agit de chercher la valeur vraie ou la valeur probable, d'une grandeur réelle ; dans le calcul d'un coefficient statistique stable, ce qui est vrai, ou réel, ce sont les données, c'est-à-dire les écarts, et la valeur cherchée est toute dans l'esprit.

De là, l'assimilation systématique de Quételet dont nous avons parlé plus haut. Il voit parfois dans la Nature une urne dont nous ne connaissons pas la composition, et dont les boules tirées sont les évènements ; parfois la Nature lui apparaît comme un tireur qui vise un but et dont les coups répétés sont les individus qui tendent à se rapprocher d'un type, qui est le but visé.

Il est peut-être chimérique de rechercher dans quelle mesure cette comparaison est légitime. Sa justification se trouve dans le grand nombre de cas où elle s'est montrée féconde pour aboutir à des résultats pratiques incontestables, tels que les assurances, ou les mesures anthropométriques.

N'en a-t-il pas été de même en physique ?

Maxwell admet que, dans un système gazeux isolé, des vitesses de toutes grandeurs et de tous sens, réparties conformément à la loi de Gauss, animent les molécules : cette hypothèse, impossible à contrôler directement, trouve d'étonnantes confirmations dans ses résultats : les lois de Mariotte, de Gay-Lussac et d'autres en découlent comme autant de corollaires.

Les mouvements Browniens, si mystérieux encore, mais déjà perceptibles à notre œil aidé du microscope, se plient également bien aux exigences de la loi de Gauss.

Néanmoins, à côté des phénomènes qui admettent volontiers ce mode de représentation, quantité d'autres y paraissent rebelles.

Ce sont ceux dont la courbe représentative ne présente pas un axe vertical de symétrie.

Deux voies s'ouvraient, dans ce cas, aux savants.

La première est sûre et aisée. Elle consiste à prendre les données empiriques comme abscisses et ordonnées d'une courbe spéciale à chaque cas et dont la théorie de l'interpolation donne la formule. La représentation est aussi parfaite qu'on le veut, mais les divers phénomènes représentés n'ont pas de lien général, et toute commune mesure de la dispersion de la fréquence devient impossible.

La seconde voie, moins simpliste, est plus scientifique. Elle consiste à considérer le cas de la courbe symétrique comme un cas particulier d'une courbe plus générale.

Si nous appelons :

$$y_0 = A e^{-mx^2}$$

ila loi normale, la loi générale sera une série de la forme

$$y = y_0 + y_1 + y_2 \dots$$

dont y_0 sera le premier terme et la première approximation.

C'est dans cette voie qu'a cherché M. le professeur Edgeworth (1).

Ses recherches remontent à plus de vingt-cinq ans.

(1) Parmi les devanciers qu'il aime à citer lui-même, il y a Bravais et Lexis. Celui-ci, professeur d'économie politique à l'Université de Göttingen, est l'auteur d'une théorie de la *dispersion* des erreurs, qui date de 1877, à l'aide de laquelle il a fourni notamment le moyen d'expliquer la stabilité de certains coefficients statistiques, même quand ils reposent sur des nombres d'observations relativement petits. (*Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft*. Freiburg, 1877, 1 vol. in-8° 95 pages.) Cette théorie l'a conduit en outre à imaginer, comme se trouvant

Quand on relit, en effet, des articles qu'il a publiés dans le *Philosophical Magazine* dès 1883, on voit qu'il a déjà alors cette conviction que la loi des erreurs s'applique à un plus grand nombre de phénomènes qu'on ne le croit d'habitude. Un travail qu'il présenta à la session jubilaire de la Royal Statistical Society en 1885 sur les *méthodes de la Statistique* est enrichi de multiples exemples qui n'ont pas d'autre but.

Puis, dans les nombreux articles qu'il a fait paraître dans divers recueils (1), la même pensée revient sous cette forme que la condition de symétrie de la répartition des écarts n'est nullement indispensable.

En 1902, il donna l'expression d'une seconde, et même d'une troisième approximation de la loi qui convenait à une courbe de fréquence disymétrique, et M. Bowley a indiqué toute la technique du travail numérique de l'application de ces formules.

Enfin, en 1905, il a exposé dans les *Cambridge Philosophical Transactions*, sous le titre de *The Law of Error* (la loi d'erreur), la théorie qui conduit à une expression tout à fait générale de la loi.

Nous ne pouvons songer à reproduire ici cette théorie. Qu'il nous suffise d'en montrer les grandes lignes.

L'auteur définit ainsi la loi d'erreur qu'il s'agit de chercher : « une expression approchée de la fréquence avec laquelle, à la longue, différentes valeurs sont prises par une grandeur qui dépend d'un nombre d'éléments variables ».

à la base des observations, des erreurs élémentaires dont la probabilité varie en série. Voir à ce sujet, Czuber, *op. cit.* p. 231 et L. von Bortkiewicz, dans l'*Encyclopédie des sciences mathématiques* de W. Meyer (Leipzig, Teubner, 1900-1904) : I. D. 4a. *Anwendungen der Wahrscheinlichkeitsrechnung auf Statistik*, pp. 829-835 (cette partie n'a pas encore paru dans la traduction française).

(1) Voir à la fin de notre mémoire une liste de ces articles.

La première approximation de cette loi a été donnée par Laplace (1); Poisson a recherché la seconde, à l'aide de la méthode des moindres carrés. M. Edgeworth en a indiqué sans la démontrer la troisième (2).

Il s'agit à présent de fournir la preuve que toutes ces expressions sont des cas particuliers d'une expression plus générale qui conduira à une quatrième, une cinquième, une *t^{me}* approximative.

L'auteur prend pour point de départ, non le cas le plus simple, mais le cas « typique », c'est-à-dire celui où toutes les conditions pour que la loi s'applique au mieux sont remplies. Ces conditions sont au nombre respectable de onze.

M. Edgeworth ne nous donne pas d'exemple concret, et ne nous indique pas de domaine de la vie réelle où se rencontre ce cas *typique*, et il est vraisemblable qu'on ne l'y trouvera jamais dans sa pureté.

Disons tout de suite d'ailleurs, que ce n'est qu'un procédé de raisonnement, et qu'une fois la démonstration faite pour ce cas compliqué, l'auteur va successivement biffer trois de ces conditions, et en relâcher cinq ou six autres, de façon à se rapprocher de plus en plus de cas se produisant dans l'expérience journalière.

Quoi qu'il en soit, après avoir posé les conditions du problème, il en entreprend la solution avec un luxe de preuves vraiment étourdissant.

Il nous offre, d'abord, une première méthode, qui lui est propre, bien qu'elle se rattache à celle de Laplace et de Poisson. Nous notons entre autres qu'elle repose

(1) *Théorie analytique des probabilités*, livre II, chap. IV : *De la probabilité des erreurs des résultats moyens d'un grand nombre d'observations et des résultats moyens les plus avantageux*, pp. 304 et suiv. de la 3^e édit. (1820).

(2) Pour la première fois au Supplément de l'*Encyclopædia Britannica*.

sur la théorie des fonctions génératrices de Laplace, qui fournit très heureusement le moyen d'exprimer en série les approximations cherchées.

La méthode de M. Edgeworth, que M. Bowley a appelée à juste titre un « triomphe d'analyse » (1), le conduit à la formule symbolique suivante :

$$y = e^{-\frac{x^2}{2k_0}} \left(\frac{d}{dx} \right)^3 + \frac{1}{4!} \left(\frac{d}{dx} \right)^4 + \dots + (-1)^t \frac{k_t}{(t+2)!} \left(\frac{d}{dx} \right)^{t+2} y_0$$

dans laquelle y_0 est mis pour $\frac{1}{\sqrt{2\pi k_0}}$, et $k_0, k_1, k_2, \dots, k_t$

sont des moyennes puissances des écarts des différentes valeurs par rapport à la moyenne.

Dans le développement en série de l'exponentielle, on a pour premier terme y_0 ou $\frac{1}{\sqrt{2\pi k_0}} e^{-\frac{x^2}{2k_0}}$ qui cons-

titue la première approximation ; pour second terme $-k_1 \frac{1}{3!} \cdot \frac{d^3 y_0}{dx^3}$ qui forme la seconde approximation,

puis $k_2 \frac{1}{4!} \cdot \frac{d^4 y_0}{dx^4} + \frac{k_1^2}{2!} \cdot \frac{1}{3! 3!} \cdot \frac{d^6 y_0}{dx^6}$ qui consti-

tue la troisième, et rien n'empêche de poursuivre jusqu'à la $t^{\text{ième}}$, à condition que t soit petit par rapport à certaines grandeurs indiquées.

Non content de sa démonstration originale, l'auteur entreprend la solution du problème par d'autres méthodes : celle due au professeur Morgan Crofton, qui a écrit l'article *Probabilités* dans l'*Encyclopédie britannique*, puis celle employée par Laplace lui-même, au livre II, ch. IV, art. 22 de sa *Théorie analytique des probabilités* et que M. Edgeworth s'attache à employer pour résoudre « un point difficile » signalé par Czuber (*Theorie der Beobachtungsfehler*, p. 271) ; enfin l'auteur

(1) Voir *Journal of Stat. Soc.*, t. LXVIII, 1905, p. 749.

donne encore une méthode « variante » qui emprunte ses procédés aux deux dernières.

M. Edgeworth nous offre alors des « vérifications » : pour montrer la validité des méthodes employées, il les applique à des cas plus simples qui peuvent être traités par des méthodes plus élémentaires.

C'est alors seulement qu'il montre dans quelle mesure sa formule reste utile pour traiter des cas moins complexes que le cas « typique » et qu'il supprime les conditions les plus rigoureuses posées au début.

On peut résumer ainsi avec M. Bowley (1) celles qui subsistent :

1° Il faut que la *grandeur complexe* dont il s'agit d'exprimer la loi de fréquence soit composée d'une somme (ou de toute autre fonction qu'on peut considérer approximativement comme une somme) de nombreux (*m*) *éléments*.

2° Ces éléments entrent, chacun, dans la formation de groupes, qui prennent eux-mêmes leurs valeurs *au hasard* selon une loi de fréquence.

3° Ces courbes de fréquence doivent avoir leurs moments d'ordres successifs inférieurs à

$$m^{-1/2} \quad , \quad m^{-1} \quad , \quad m^{-3/2} \dots\dots$$

Il faut entendre par là que les moyennes puissances successives des écarts (comptés à partir du centre de gravité comme origine) doivent former une série descendante.

4° En formant les groupes, les éléments doivent être indépendants les uns des autres ; mais cette indépendance n'a pas besoin d'être absolue. De même, en principe, chaque apport d'un groupe à une observation individuelle qui constitue la quantité complexe doit aussi être indépendant des autres ; mais cette indépendance n'a pas

(1) V. *Journal Stat. Soc.* t. LXVIII, 1905. p. 749.

besoin d'être absolue en ce sens qu'il suffit de disposer d'un plus grand nombre d'observations pour que la loi d'erreur soit satisfaite.

Supposons, pour fixer les idées, qu'il s'agisse de rechercher la loi de fréquence d'une quantité complexe, dont la moyenne des valeurs est A . Une observation individuelle sera représentée par son écart par rapport à cette moyenne :

$$A + x.$$

La valeur individuelle de l'observation p sera :

$$A + {}_p x,$$

que nous avons appelée plus haut un *groupe*.

C'est, en effet, un groupe, parce que A est fonction de beaucoup d'éléments, dont les valeurs moyennes peuvent être représentées par a .

$$A = a_1 + a_2 + \dots a_m.$$

Si nous appelons ξ les écarts par rapport à ces dernières moyennes, nous aurons pour la p^{me} observation :

$$A + {}_p x = (a_1 + {}_p \xi_1) + (a_2 + {}_p \xi_2) + \dots (a_m + {}_p \xi_m).$$

On demande que les valeurs prises par le même élément dans les différentes observations, par exemple ${}_1 \xi_m, {}_2 \xi_m, \dots {}_p \xi_m$, soient prises au hasard indépendamment l'une de l'autre.

On demande en outre que les valeurs prises par différents éléments à la même observation ${}_p \xi_1, {}_p \xi_2, {}_p \xi_m$, soient aussi prises au hasard indépendamment l'une de l'autre.

Cependant, il faut faire observer immédiatement que ce hasard ne doit pas être *parfait*, et est compatible avec une certaine corrélation entre les éléments et les groupes.

Si l'on veut traduire ces symboles dans un exemple plus concret — qui n'est pas d'ailleurs le meilleur — nous dirons que, s'il s'agit de rechercher la loi de

fréquence des salaires, la « quantité complexe » est le salaire, dont la moyenne est A . Le salaire d'un ouvrier individuel p sera $A + {}_p x$ et sera fonction (supposée additive) d'un grand nombre d'éléments : son intelligence, la durée de son travail, la localité où il travaille, la prospérité générale de l'industrie à laquelle il appartient, le nombre de ses concurrents, le nombre de patrons, etc. Pour former le salaire individuel, ces éléments prennent leurs valeurs au hasard, et on admettra que $a_1 + {}_p \xi_1$, la contribution de l'élément intelligence de l'ouvrier p ne dépende pas de celle d'un autre. On admettra, en outre, que chez le même ouvrier, chacun des éléments prenne des valeurs qui ne soient pas dépendantes les unes des autres.

Quant à la condition relative aux moments des courbes de fréquence, elle indique que les variations des différents éléments et des différents groupes soient sensiblement du même genre ; en ce sens que leurs dispersions, mesurées par leurs moments successifs, n'excèdent pas une certaine dispersion dépendant du nombre (m) d'éléments.

M. Edgeworth indique comme pouvant être soumises à sa loi généralisée d'erreur : le taux de la mortalité aux différents âges, les salaires, les caractères biologiques qui se mesurent, tels que la taille, etc. Il l'a appliquée également à des statistiques relatives aux hauteurs barométriques. Il ajoute que, dans son idée, « de même que la *ligne droite*, le *cercle*, l'*ellipse* et d'autres figures mathématiques simples sont continuellement, sous l'action de certaines forces, reproduites dans la nature et dans l'art, de même cette variation « sporadique » qui semble prévaloir à travers l'Univers, prend continuellement la forme *exponentielle* qu'exprime sa loi d'erreur ».

L'intérêt de ces recherches consiste donc, en résumé, en ceci, que nous possédons maintenant une définition

mathématique des conditions dans lesquelles on doit s'attendre à trouver en action la loi du hasard (1).

La difficulté consiste à reconnaître dans les données statistiques la présence de ces conditions. On nous offre une répartition des salaires d'un grand nombre d'ouvriers. Dépendent-ils d'un nombre suffisamment grand « d'éléments » ? Ceux-ci sont-ils suffisamment indépendants les uns des autres ?

C'est malheureusement ce qu'il est impossible de savoir *a priori*. Mais quand on essaie d'y appliquer la formule de la loi d'erreur, on est à même de dire si une cause prépondérante n'est pas venue déranger la répartition qui se réaliserait d'elle-même si toutes les conditions essentielles indiquées plus haut avaient été remplies.

(1) En employant ce mot, nous n'entendons pas lui donner une portée différente de celle d'une figure de langage habituelle. Nous ne voulons nullement « éblouir les profanes » comme l'indique M. Mansion, dans le discours cité (p. 1260), répétant J. Bertrand. La loi d'erreur de M. Edgeworth est parfaitement conciliable avec la conception subjectiviste du hasard. Si les salaires se groupent autour d'une valeur moyenne d'après une loi qui prend cette forme, ce n'est nullement parce que cette dernière serait l'expression d'une action mystérieuse du hasard.

D'ailleurs, comme parmi les éléments ou constituants de cette fréquence, il y a une infinité d'actions humaines qu'on peut considérer comme libres — « les seuls cas fortuits à parler rigoureusement » dit Bertrand — on conçoit que la dispersion des salaires suive la forme générale des phénomènes soumis au hasard. C'est pourquoi nous pouvons très bien souscrire aux paroles de M. Mansion à ce sujet : « Le calcul des probabilités a pour objet les événements que nous sommes fondés à considérer comme soumis à une loi partiellement inconnue, loi complexe d'ailleurs, résultante d'une loi principale, d'après laquelle des rapports numériques sont constants et de lois perturbatrices secondaires donnant naissance à de faibles variations dites fortuites par définition. » (*Op. cit.* p. 1260.)

IV

Pour terminer, nous allons appliquer la loi de M. Edgeworth à un exemple.

D'une façon générale, le problème à résoudre est le suivant : on donne les résultats d'un dénombrement, c'est-à-dire la valeur de la fréquence y en fonction des mesures x de l'une des caractéristiques du phénomène dénombré ; les valeurs de x varient d'une façon discontinue, par sauts brusques, mais égaux ; on demande de déterminer la fonction $y = f(x)$ qui relie les variations de y aux variations *continues* de x .

La théorie de M. Edgeworth fait connaître la forme de la fonction cherchée : il s'agit d'en déterminer les paramètres.

Prenons, comme exemple, les salaires des ouvriers (hommes) de plus de 16 ans en Belgique, d'après le *Recensement industriel* du 31 octobre 1896 (T. XVIII, p. 267).

Ce relevé peut être résumé dans le tableau suivant :

Catégories de salaires.	Nombre d'ouvriers.	Fréquence en millièmes.
Moins de fr. 1,50	22.276	48
De fr. 1,50 à 2,00	28.644	61
2,00 à 2,50	62.191	133
2,50 à 3,00	87.015	187
3,00 à 3,50	100.392	215
3,50 à 4,00	65.615	141
4,00 à 4,50	50.887	109
4,50 à 5,00	21.134	45
5,00 à 5,50	13.934	30
5,50 à 6,00	5.776	12
6,00 à 6,50	3.668	8
6,50 à 7,00	1.653	4
7,00 à 7,50	1.127	2
7,50 à 8,00	650	1,5
8,00 à 8,50	593	1
8,50 à 9,00	316	1
9,00 à 9,50	326	1
9,50 à 10,00	223	0,5
Totaux	463.185	1000

Ces résultats peuvent également être résumés graphiquement à l'aide d'un polygone de fréquence, où les surfaces de chaque échelon sont proportionnelles aux valeurs des fréquences y .

En vue de faciliter les calculs, nous avons transporté l'origine o au milieu de la base du premier échelon et pris comme unité d'abscisse la base d'un échelon.

On appelle i^{e} moment de l'ensemble de ces salaires, la somme des i^{e} puissances de chaque salaire multiplié par sa fréquence. Si on suppose les salaires de chaque échelon uniformément répartis, c'est-à-dire si on termine les échelons par une base supérieure rectiligne, on a, en vertu d'un théorème d'usage courant en mécanique :

$$\begin{aligned} M_0 &= \Sigma y = 1 \\ M_1 &= \Sigma x y \\ M_2 &= \Sigma x^2 y \\ M_3 &= \Sigma x^3 y, \text{ etc.} \end{aligned}$$

On observera qu'en appelant g le salaire moyen, on a :

$$g = \frac{y_1 + 2 y_2 + 3 y_3 + \dots}{y_1 + y_2 + y_3 + \dots} = \frac{\Sigma x y}{\Sigma y} = M_1.$$

Le point G d'abscisse g est dit le *centre de gravité* de la courbe. C'est à ce point pris pour origine qu'est rapportée la courbe donnée par M. Edgeworth.

Les moments rapportés à l'origine G ont pour expressions :

$$\begin{aligned} \mu_1 &= \Sigma (x - g) y = \Sigma x y - g \Sigma y \\ &= M_1 - g \\ &= 0 \\ \mu_2 &= \Sigma (x - g)^2 y = \Sigma x^2 y - 2g \Sigma x y + g^2 \Sigma y \\ &= M_2 - 2g M_1 + g^2 \\ &= M_2 - g^2 \\ \mu_3 &= \Sigma (x - g)^3 y = \Sigma x^3 y - 3g \Sigma x^2 y + 3g^2 \Sigma x y - g^3 \Sigma y \\ &= M_3 - 3g (\mu_1 + g^2) + 3g^3 - g^3 \\ &= M_3 - 3g \mu_2 - g^3 \\ \mu_4 &= M_4 - 4g \mu_3 + 6g^2 \mu_2 - g^4 \end{aligned}$$

Mais l'hypothèse d'une répartition uniforme des salaires entre tous les salariés d'un même échelon ne constitue évidemment qu'une approximation. On fait subir à celle-ci une correction, dite correction de Sheppard (1) du nom de son inventeur, en supposant que les divers échelons, tout en conservant leurs surfaces respectives $y_0, y_1, y_2, \dots$, sont terminés vers le haut par une courbe qui les raccorde, qui soit continue et qui présente avec l'axe des x à l'avant et à l'arrière, deux points de contact *complet*. Dans ces conditions, on établit aisément en appelant ν_i les moments corrigés, que

$$\begin{aligned}\nu_1 &= \mu_1 \\ \nu_2 &= \mu_2 - \frac{1}{12} \\ \nu_3 &= \mu_3 \\ \nu_4 &= \mu_4 - \frac{1}{2} \mu_2 + \frac{7}{240}, \text{ etc.}\end{aligned}$$

Une fois en possession des valeurs des moments 1^{er}, 2^e, 3^e, etc., de la distribution donnée, il va suffire de calculer ces mêmes moments en fonction des paramètres inconnus de la courbe représentative de M. Edgeworth. On prendra autant de moments qu'il y a de paramètres à calculer et on obtiendra ainsi le nombre d'équations nécessaires à leur détermination.

Si nous nous bornons à la troisième approximation, la loi générale de l'erreur peut s'écrire :

$$y = (A_0 + A_1x + A_2x^2 + A_3x^3 + A_4x^4 + A_5x^5 + A_6x^6) y_0 \quad (1)$$

en posant :

$$\begin{aligned}A_0 &= 1 + \frac{1}{2} \frac{k_2}{c^4} + \frac{5}{3} \frac{k_1^2}{c^6} \\ A_1 &= -2 \frac{k_1}{c^4}\end{aligned}$$

(1) Voir « On the calculation of the Average Square, Cube etc. of a large number of Magnitudes » (*Journal R. Stat. Society*, sept. 1897, pp. 698-703). — V. aussi *Biometrika*, vol. III, p. 308.

$$A_2 = -2 \frac{k_2}{c^6} + 10 \frac{k_1^2}{c^8}$$

$$A_3 = \frac{4}{3} \frac{k_1}{c^6}$$

$$A_4 = \frac{2}{3} \frac{k_2}{c^8} - \frac{20}{3} \frac{k_1^2}{c^{10}}$$

$$A_5 = 0$$

$$A_6 = \frac{8}{9} \frac{k_1^2}{c^{12}}$$

$$c = \sqrt{2 k_0}$$

$$y_0 = \frac{1}{c\sqrt{\pi}} e^{-\frac{x^2}{c^2}}$$

L'équation (1) renferme en fait, trois paramètres indépendants k_0 , k_1 et k_2 . Pour les déterminer, il nous suffira de calculer trois moments et de les égaier respectivement aux trois valeurs correspondantes de ν .

Appelons μ'_i le moment d'ordre i fourni par la fonction de M. Edgeworth.

On a :

$$\mu'_i = \int_{-\infty}^{+\infty} y x^i dx = \int_{-\infty}^{+\infty} (A_0 + A_1 x + \dots A_6 x^6) x^i y_0 dx.$$

Or, pour p impair, $\int_{-\infty}^{+\infty} x^p e^{-\frac{x^2}{c^2}} dx$ est nulle et pour p pair, sa valeur est :

$$1 \cdot 3 \cdot 5 \dots (p-1) \sqrt{\frac{\pi}{2^p}} \cdot c^{p+1}.$$

De là résulte immédiatement en faisant successivement $i = 0, 1, 2, 3, 4$, que

$$\mu'_0 = 1$$

$$\mu'_1 = 0$$

$$\mu'_2 = \frac{1}{2} c^2$$

$$\mu'_3 = k_1$$

$$\mu'_4 = 3 \mu'^2_2 + k_2$$

Par conséquent, en égalant les μ'_i aux ν_i pour $i = 0, 1, 2, 3$ et 4, on trouve :

$$k_0 = \frac{1}{2} c^2 \quad \text{et} \quad c = \sqrt{2\nu_2}$$

$$k_1 = \nu_3$$

$$k_2 = \nu_4 - 3\nu_2^2$$

Il ne reste plus qu'à calculer les $A_0, A_1, A_2, \dots, A_6$.

Au lieu d'écrire la fonction sous la forme :

$$y = (A_0 + A_1 x + \dots + A_6 x^6) y_0$$

on peut mettre en évidence les trois approximations successives et écrire :

$$y = (1 + \Delta_1 + \Delta_2) y_0$$

$$\text{où } y_0 = \frac{1}{c\sqrt{\pi}} e^{-\frac{x^2}{c^2}} \quad \text{et où } \Delta_1 = \frac{k_1}{c^3} \left(-2\xi + \frac{4}{3}\xi^3 \right)$$

$$\Delta_2 = \frac{k_2}{c^4} \left(\frac{1}{2} - 2\xi^2 + \frac{2}{3}\xi^4 \right) + \frac{k_1}{c^6} \left(-\frac{5}{3} + 10\xi^2 - \frac{20}{3}\xi^4 + \frac{8}{9}\xi^6 \right)$$

$$\text{et} \quad \xi = \frac{x}{c}.$$

Si l'on veut construire la courbe, on pourra passer aux logarithmes et développer en série le logarithme de $(1 + \Delta_1 + \Delta_2)$ en négligeant les termes de degré supérieur au premier en k_2 et au second en k_1 . Il vient :

$$\log. y = 0.4343 (m_0 + m_1 \xi + \dots + m_4 \xi^4)$$

en posant

$$m_0 = -\log \text{nép} (c\sqrt{\pi}) + \frac{1}{2} \frac{k_2}{c^4} - \frac{5}{3} \frac{k_1^2}{c^6}$$

$$m_1 = -2 \frac{k_1}{c^3}$$

$$m_2 = -1 - 2 \frac{k_2}{c^4} + 8 \frac{k_1^2}{c^6}$$

$$m_3 = \frac{4}{3} \frac{k_1}{c^3}$$

$$m_4 = \frac{2}{3} \frac{k_2}{c^4} - 4 \frac{k_1^2}{c^6}$$

Passons maintenant au calcul numérique.

$M_1 = \Sigma xy = 4,991$, qui est en même temps l'abscisse de l'ordonnée du centre de gravité g . Exprimée en francs, cette valeur = fr. 3,24, = la moyenne arithmétique des salaires.

$$M_2 \Sigma x^2 y \equiv = 29.891$$

$$M_3 \Sigma x^3 y \equiv = 209.093$$

$$M_4 \Sigma x^4 y \equiv = 1699.147$$

Rapportés à l'abscisse du centre de gravité comme origine, les moments deviennent :

$$\mu_1 = 0$$

$$\mu_2 = M_2 - g^2 = 4.981$$

$$\mu_3 = M_3 - 3 g \mu_2 - g^3 = 10.187$$

$$\mu_4 = M_4 - 4 g \mu_3 - 6 g^2 \mu_2 - g^4 = 130.806$$

Amendés enfin par les corrections de Sheppard :

$$\nu_1 = 0$$

$$\nu_2 = \mu_2 - \frac{1}{12} = 4.898$$

$$\nu_3 = \mu_3 = 10.187$$

$$\nu_4 = \mu_4 - \frac{1}{2} \mu_2 - \frac{7}{240} = 128.286$$

D'où l'on tire :

$$\text{le module, } c = \sqrt{2\nu_2} = 3.13$$

$$k_1 = \nu_3 = 10.187$$

$$k_2 = \nu_4 - 3\nu_2^2 = 56.314$$

$$\frac{k_1}{c^3} = 0.332$$

$$\frac{k_2}{c^4} = 0.5867$$

$$\frac{k_1^2}{c^6} = 0.1103$$

La formule

$$y = \frac{1}{c \sqrt{\pi}} \left[1 + \frac{k_1}{c^3} \left(-2\xi + \frac{4}{3}\xi^3 \right) + \frac{k_2}{c^4} \left(\frac{1}{2} - 2\xi^2 + \frac{2}{3}\xi^4 \right) + \frac{k_1^2}{c^6} \left(-\frac{5}{3} + 10\xi^2 - \frac{20}{3}\xi^4 + \frac{8}{9}\xi^6 \right) \right] e^{-\xi^2}$$

peut être mise, comme nous l'avons dit plus haut, sous la forme :

$$\log y = \log \frac{1}{\sqrt{\pi}} - \log c + \log e \left[-\xi^2 + \frac{k_1}{c^3} \left(-2\xi + \frac{4}{3}\xi^3 \right) + \frac{k_2}{c^4} \left(\frac{1}{2} - 2\xi^2 + \frac{2}{3}\xi^4 \right) + \frac{k_1^2}{c^6} \left(-\frac{5}{3} + 8\xi^2 - 4\xi^4 \right) \right]$$

$$\log c = 0.4955$$

$$\log \frac{1}{\sqrt{\pi}} = 1.7514$$

$$\log \frac{1}{\sqrt{\pi}} - \log c = 1.2559$$

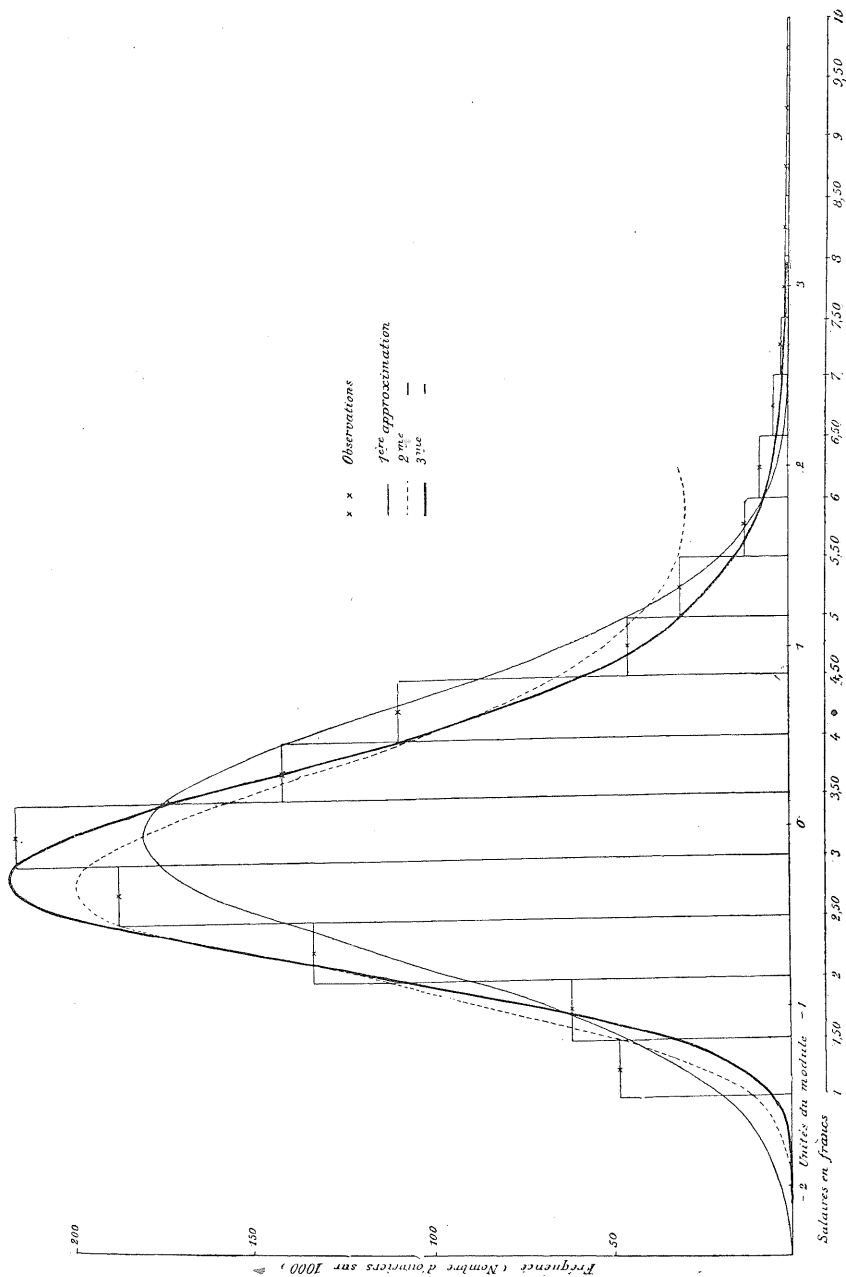
Posons $\varphi_1 = -2\xi + \frac{4}{3}\xi^3$

$$\psi_1 = \frac{1}{2} - 2\xi^2 + \frac{2}{3}\xi^4$$

$$\psi_2 = -\frac{5}{3} + 8\xi^2 - 4\xi^4$$

$$\log y = (1.2559 - 0.4343 \xi^2) + (0.144 \varphi_1) + (0.255 \psi_1 + 0.048 \psi_2)$$

On se rappellera que la première approximation y_0 est exprimée par la première parenthèse, la seconde y_1 par l'ensemble des deux premières, et la troisième y_2 par le tout. Le diagramme p. 259 donne le tracé du polygone de fréquence et des courbes de la première, de la deuxième et de la troisième approximation.



Pour mesurer la convenance de la courbe aux données, les statisticiens anglais évaluent en pour cent du nombre total des observations la somme des différences, prises en valeurs absolues, entre les observations et les valeurs calculées correspondantes.

Ce procédé nous donne, en nous arrêtant aux neuf premières données (1), les résultats suivants :

Observations	1 ^{re} approximation	Différences	2 ^e approximation	Différences	3 ^e approximation	Différences
48	30	18	26	22	15	33
61	66	5	83	22	67	6
133	116	17	145	12	145	12
187	152	35	193	6	212	25
215	180	35	179	36	200	15
141	160	19	132	9	141	0
109	114	5	85	24	83	26
45	66	21	53	8	43	2
30	30	0	35	5	22	8
969		145		144		127
Différences en % . . .		15		15		13

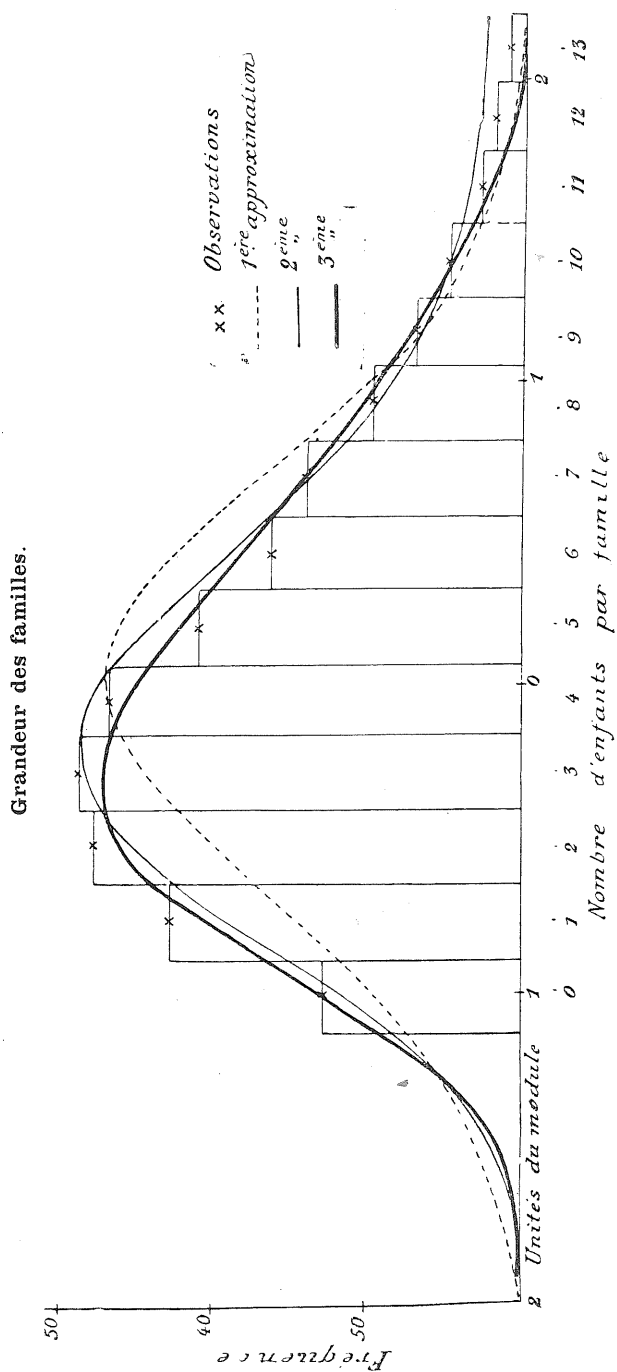
V

Les courbes de salaires ne sont pas naturellement celles qui donnent les meilleurs résultats.

Nous reproduisons ci-après deux exemples donnés par M. Edgeworth (2). Le premier est emprunté à une statistique de M. le professeur K. Pearson et exprime le nombre relatif de familles sur 1000, ayant un certain nombre d'enfants.

(1) La deuxième approximation ne fournit plus des valeurs comparables, quand x dépasse 2 unités du nodule.

(2) *Journal St. Soc.*, T. LXIX (1906), pp. 522 et suiv.

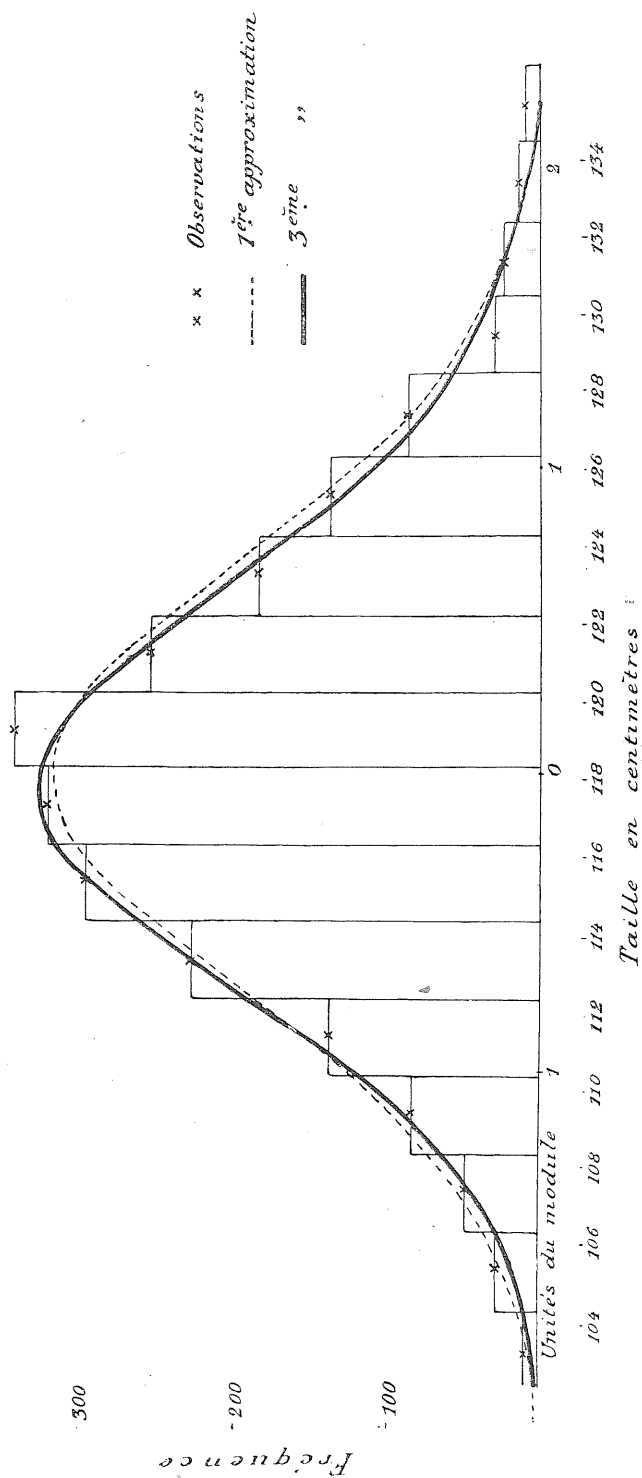


Nombre d'enfants par famille.	Nombre de familles sur 1000 (fréquence).	Première approximation et différences avec les observations.		Deuxième approximation.		Troisième approximation.	
0	64	48	16	58	6	64	0
1	116	73	43	92	24	102	14
2	140	101	39	127	13	130	10
3	145	127	18	142	3	135	10
4	134	134	0	139	5	130	4
5	106	131	25	120	14	111	5
6	82	113	31	95	13	92	10
7	72	87	15	70	2	73	1
8	49	60	11	48	1	53	4
9	37	37	0	34	3	36	1
10	25	20	5	23	2	20	5
11	13	10	3	17	4	10	3
12	10	4	6	14	4	4	6
13	5	2		?			
14	2	0		?			
15	0,4	0		?			
16	0,2	0		?			
17	0,4	0		?			
Différences en %		21		9		7	

Le second exemple représente la taille de 2192 petites filles mesurées dans les écoles de Saint-Louis (Amérique) en 1894.

	Première approximation.		Deuxième approximation.		Troisième approximation.	
Observations.	Avec les différences par rapport aux observations.					
1	0			—		—
2	4	2	2	0	2	0
8	9	1	8	0	8	0
27	24	3	22	5	20	7
42	39	3	44	2	42	0
84	88	4	94	10	84	0
137	152	15	158	21	148	11
222	222	0	230	8	226	4
297	260	28	284	13	290	7
321	295	26	312	9	321	0
342	316	26	300	42	311	31
243	267	24	260	17	258	15
183	205	22	200	17	127	14
138	127	11	128	10	124	14
79	68	11	79	0	78	1
28	39	11	44	16	40	12
21	22	1	21	0	21	0
10	6	4	8	2	8	2
5	2	3	4	1	5	0
	195		173		118	
Différence en %	9		8		2,4	

Taille des écolières à Saint-Louis (Amérique).



M. C. V. L. Charlier, le savant directeur de l'Observatoire de Lund (Suède) a publié, sur les courbes de M. Karl Pearson une étude qui éclaire également le travail numérique de celle de M. Edgeworth. V. Meddelanden från Lunds Astronomiska Observatorium. Série II, N° 24. *Researches into the theory of probability*. (Lund 1906. Håkan Ohlsson, 1 br. in-4° 51 pages avec des tables et de nombreuses figures.)

Appendice.

Voici une liste des articles de M. Edgeworth, que nous ne donnons pas comme complète :

The Law of Error. Philosophical Magazine (5^e série), t. XVI, 1883, pp. 300-309.

The Method of Least Squares. Ibid. 1883., pp. 360-375.

The Physical Basis of Probability. Ibid., 1883, pp. 433-435.

On the reduction of Observations. Ibid. t. XVII, 1884, pp. 135-141.

A priori Probabilities. Ibid. t. XVIII, 1884, pp. 204-210.

On Methods of Statistics. Jubilee volume of the R. Statistical Society. June 1885, pp. 181-217.

Methods of Ascertaining variations in the Rate of Births, Deaths, and Marriages. Journal of the Royal Statistical Society Déc. 1885, vol. XLVIII, p. 628.

The Law of Error and the Elimination of Chance. Phil. Mag. t. XXI. 1886. pp. 308-324.

On the Determination of the Modulus of Errors. Ibid. t. XXI. 1886. pp. 500-507.

Progressive Means. Journal St. Soc. t. XLIX, 1886, pp. 469-478.

The Mathematical Method of Statistics. Ibid. vol. XLIV, pp. 649-654.

Problems in Probabilities. Phil. Mag. t. XXII, 1886, pp. 371-384.

On discordant Observations. Ibid. t. XXIII, 1887, pp. 364-375.

The Choice of Means. Ibid. t. XXIV. 1887. pp. 268-271.

The empirical Proof of the Law of Error, Ibid. t. XXIV. 1887. pp. 330-342.

On a new Method of reducing Observations relating to several quantities. Ibid. t. XXV. 1888. pp. 184-191.

Problems in Probabilities. N° 2. Competitive Examinations. Phil. Mag. t. XXX. 1890. pp. 171-188.

The Law of Error and correlated Averages. Phil. Mag. t. XXXIV. 1892. pp. 429-438 et 518-526.

A new method of treating correlated Averages. Phil. Mag. t. XXXV. 1893. pp. 63-64.

Statistical Correlation between Social Phenomena. Journal Stat. Soc. 1893, t. LVI, pp. 670-675.

Exercises in the Calculations of Errors, Phil. Mag., t. XXXVI, 1903, pp. 98-111.

Note on the Calculation of Correlation between Organs. Phil. Mag., t. XXXVI, 1893, pp. 350-351.

Asymmetrical or Correlation of Social Phenomena. Journal Stat. Soc., t. LVII, 1894, pp. 563-568.

On some recent contributions to the Theory of Statistics. Journal Stat. Soc., t. LVIII, 1895, pp. 506-515.

The Asymmetrical Probability-Curve. Phil. Mag., t. XLI, 1896, pp. 90-99.

The Compound Law of Error. Phil. Mag., t. XLI, 1896, pp. 207-215.

Statistics of Unprogressive Communities. Journal Stat. Soc., t. LIX, 1896, pp. 358-386.

Supplementary Notes on Statistics. Journal Stat. Soc. ibid., pp. 529-539.

Miscellaneous applications of the Calculus of Probabilities. Journal Stat. Soc. 1897, t. LX, pp. 681-698, et 1898, t. LXI, pp. 119-131.

On the Representation of Statistics by Mathematical

Formulae. Journal Stat. Soc. 1898, t. LXI, pp. 670-700, 1899, t. LXII, pp. 125-140, 373-385, 534-555, et 1900, t. LXIII, pp. 72-81.

Methods of Representing Statistics of Wages and other Groups not fulfilling the normal Law of Error. Journal Stat. Soc. 1902, t. LXV, pp. 325-354.

The Law of Error. Cambridge Philosophical Transactions, t. XX, part. I (1905), pp. 36-141.

The Generalised Law of Error or Law of Great Numbers. Journal Stat. Soc. 1906, t. LXIX, pp. 497-530.
