Quelques regards sur l'histoire pour éclairer l'enseignement des probabilités et statistiques

Anne Boyé
IREM des Pays de la Loire

« Qu'est-ce qu'une possibilité abstraite, une chance, une probabilité ?... De quoi les nombres donnent-ils la valeur dans ce domaine ? Que représentent des valeurs pour l'exactitude ? Peut-on évaluer l'incertitude, faire une théorie scientifique des erreurs ? Deux siècles d'interrogations dont la richesse et la difficulté, de Pascal à Huygens (du droit au gain à la chance de gagner), de Huygens à Laplace (de la valeur de la chance à l'évaluation de l'incertitude), auraient un pouvoir pédagogique malheureusement encore très peu utilisé aujourd'hui. »

Pierre Raymond, *De la combinatoire aux probabilités, 1975.*

Les origines de ce travail

Comme de nombreux enseignants de mathématiques, j'éprouvais un sentiment mitigé vis-à-vis de l'enseignement des statistiques, et une certaine difficulté à me situer entre mathématiques « sciences exactes », et mathématiques de l'approximation. Les probabilités et les statistiques remettent en cause la vision cartésienne et binaire du monde, que l'on s'est efforcé d'introduire dans d'autres domaines des mathématiques.

L'enseignement des probabilités développé depuis ces dernières années au lycée et maintenant au collège, s'appuie fortement sur les statistiques, non seulement descriptives, mais aussi et surtout inférentielles. Comment passe-t-on des statistiques au modèle mathématique ? Comment éclairer les liens entre statistiques et probabilité ?

Nous sommes ici dans le domaine, incertain et séduisant à la fois, des « mathématiques du réel », qui mêlent hasard, sondages, fluctuation d'échantillonnages, loi des grands nombres, ... Nos élèves sont familiers de ces mots, mais au milieu de confusions multiples de langage, d'informations pseudo-scientifiques, donnant un tour plus ou moins magique et irrationnel au hasard, et se heurtant à la difficulté largement partagée de penser le collectif en lieu et place de l'individu.

Peut-on se contenter par exemple de la définition qu'Armand Julin[1] donnait des statistiques en 1921 :

> Méthode qui, par le relevé en masse et l'expression numérique de ses résultats, arrive à la description des phénomènes collectifs et permet de reconnaître ce qu'ils représentent de permanent et de régulier dans leur variété, comme de variable dans leur apparente uniformité.

[1] Julin, Armand, Principes de statistique théorique et appliquée, tome 1, Paris, 1921.

Peut-on *a fortiori* se retrouver dans les définitions ou non définitions d'une probabilité, la définition classique de la probabilité numérique, celle que nous donne Lacroix[2], par exemple, dans son *Traité élémentaire de calcul des probabilités.*

> La probabilité mathématique se forme, comme on le voit, en divisant le nombre des chances favorables à l'événement par le nombre total des chances ; mais il faut bien faire attention que toutes les chances comparées soient également possibles.

dont Poincaré[3] souligne les incertitudes en 1902 :

> La probabilité a-t-elle été définie ? Et, si elle peut l'être, comment ose-t-on en raisonner ? La définition, dira-t-on, est bien simple :
>
> La probabilité d'un événement est le rapport du nombre de cas favorables à cet événement au nombre total de cas possibles, ... pourvu que ces cas soient également probables. Nous voilà donc réduits à définir le probable par le probable. Comment saurons-nous que deux cas possibles sont également probables ? Sera-ce par une convention ?
>
> La conclusion qui semble résulter de tout cela, c'est que le calcul des probabilités est une science vaine, qu'il faut se défier de cet instinct obscur que nous nommions bon sens et auquel nous demandions de légitimer nos conventions.
>
> Mais, cette conclusion, nous ne pouvons plus y souscrire ; cet instinct obscur, nous ne pouvons nous en passer ; sans lui la science serait impossible ; sans lui nous ne pourrions ni découvrir une loi, ni l'appliquer.

Nous proposons ici à la réflexion quelques textes fondateurs, qui peuvent donner quelques réponses, qui peuvent permettre d'appréhender toute la richesse de cet enseignement, et de constater comme souvent qu'en ce domaine aussi l'histoire donne du sens, et permet de comprendre certains obstacles. Nous constaterons aussi que ce détour par l'histoire permet de réinterroger certaines routines, comme le calcul de la moyenne, aussi bien en probabilité qu'en statistiques.

Nous distinguerons pour cela trois grands moments : la loi des grands nombres autour de Jaques Bernoulli et de ses successeurs, la méthode des moindres carrés et la loi des erreurs, autour de Gauss et Laplace, enfin de Quetelet aux années 1930, les statistiques sociales et l'émergence des statistiques mathématiques.

I- Première époque : autour de l'Ars conjectandi et de la loi des grands nombres : De l'Art de conjecturer

L'Ars conjectandi, est publié en 1713, huit ans après la mort de Jacques Bernoulli (1654-1705).

Jacques Bernoulli y fait le lien entre la probabilité *a priori* d'un événement que l'on peut calculer avant la réalisation d'une expérience aléatoire (par exemple, la probabilité d'obtenir un 6 en lançant un dé bien équilibré est 1/6), et la probabilité *a posteriori*, que l'on peut

---

[2] Lacroix, Sylvestre François, *Traité élémentaire de calcul des probabilités*, Bachelier, Paris, 1816.

[3] Poincaré, Henri, *La science et l'hypothèse*, 1902, Flammarion, Paris, 1968.

déduire de la fréquence d'apparition de cet événement après qu'on ait répété un grand nombre de fois l'expérience aléatoire.

Outre l'aspect mathématique, son objectif est beaucoup plus large : il s'agit de montrer « *comment les bases de l'Art de conjecturer peuvent s'appliquer à la vie civile, morale et politique* ».

C'est la quatrième partie (inachevée), qui traite de ce dernier aspect, et qui contient, entre autres, la modélisation du réel par une urne contenant des pierres blanches et des pierres noires ainsi que l'énoncé de la loi des grands nombres.

Les trois premières parties reprennent les travaux et résultats antérieurs :

- première partie : le traité de « l'illustre » Huygens, *De ratiociniis in ludo aleae* ;
- deuxième partie : la théorie des permutations et des combinaisons ;
- troisième partie : applications aux divers tirages au sort et jeux de hasard ;
- quatrième partie : *L'Art de conjecturer traitant de l'usage et de la doctrine précédente aux affaires civiles, morales et économiques.*

Les extraits que nous proposons sont issus de la traduction publiée par Norbert Meusnier[4] à l'IREM de Rouen.

Premier extrait où l'on s'interroge sur ce qu'est une probabilité

> [...] Ce qui par révélation, raison, sensation, expérience, témoignage de nos propres yeux ou autrement est tellement évident que non ne pouvons aucunement douter de son existence présente ou future jouit d'une certitude totale et absolue. Tout le reste acquiert dans nos esprits une mesure moins parfaite, plus ou moins grande, selon que les probabilités sont plus ou moins nombreuses qui nous persuadent qu'une chose est, sera ou a été.
>
> La *probabilité* est en effet un degré de la certitude et en diffère comme la partie diffère du tout. Évidemment si la certitude intégrale et absolue, que nous désignons par la lettre *a* ou par l'unité 1, est constituée de - supposons par exemple - cinq probabilités ou parties dont trois militent pour qu'un événement existe ou se produise, les autres s'y opposant : nous dirons que cet événement a $\frac{3}{5}a$, ou $\frac{3}{5}$ de certitude.
>
> Sera donc dit *plus probable* qu'autre chose ce qui a une plus grande part de certitude, bien qu'en fait, d'après le langage courant, soit dit *probable,* ce dont la probabilité dépasse notablement la moitié de la certitude. Je dis : *notablement* ; car ce qui équivaut à la moitié de la certitude environ est dit *douteux* ou ambigu. Ainsi ce qui a $\frac{1}{5}$ de certitude est plus probable que ce qui n'en a qu' $\frac{1}{10}$; même si ni l'un ni l'autre n'est en fait probable.
>
> Ce à quoi est attribuée une part de certitude, si peu même que cela soit, cela est *possible* ; lorsque cette part est nulle ou infiniment petite, cela est *impossible.* Ainsi est possible ce qui a $\frac{1}{20}$ ou $\frac{1}{30}$ de certitude.
>
> Ce qui est certain et indubitable, nous disons que nous le *savons* ou que nous le *comprenons* ; tout le reste nous le *conjecturons* ou le *présumons.*

[4] Bernoulli, Jacques, *Ars conjectandi*, édition bilingue, trad. Norbert Meusnier, IREM de Rouen, 1987.

> *Conjecturer* quelque chose, c'est mesurer sa probabilité : ainsi *l'art de conjecturer* ou la *stochastique* se définit pour nous comme l'art de mesurer aussi exactement que possible les probabilités des choses. Le but est que, dans nos jugements et dans nos actions, nous puissions toujours choisir ou suivre le parti que nous aurons découvert comme meilleur, préférable, plus sûr ou mieux réfléchi. C'est en cela seulement que réside toute la sagesse du Philosophe et toute la sagacité du Politique.
>
> Les probabilités sont estimées d'après le *nombre* et aussi le *poids des arguments* qui de quelque manière prouvent ou révèlent que quelque chose est, sera ou a été. En outre, par le *poids,* j'entends la force de ce qui prouve.

Est-ce toujours aussi simple ?

> [...] On en est ainsi venu à ce point que pour former selon les règles des conjectures sur n'importe quelle chose, il est seulement requis, d'une part, que les nombres de cas soient soigneusement déterminés et, d'autre part, que soit défini combien les uns peuvent arriver plus facilement que les autres. Mais c'est ici enfin que surgit une difficulté, nous semble-t-il : cela peut se voir à peine dans quelques très rares cas et ne se produit presque pas en dehors des jeux de hasard que leurs premiers inventeurs ont pris soin d'organiser en vue de se ménager l'équité, de telle sorte que fussent assurés et connus les nombres de cas qui doivent entraîner le gain ou la perte, et de telle sorte que tous ces cas puissent arriver avec une égale facilité. En effet, lorsqu' il s'agit de tous les autres résultats, dépendant pour la plupart soit de l'œuvre de nature soit de l'arbitre des hommes, cela n'a pas du tout lieu.
>
> Ainsi, par exemple, les nombres de cas sont connus lorsqu'il s'agit des dés, car pour chacun des dés les cas sont manifestement aussi nombreux que les bases, et ils sont tous également enclins à échoir ; car, à cause de la similitude des bases et du poids uniforme des dés, il n'y a point de raison pour qu'une des bases soit plus encline à échoir que l'autre, comme cela arriverait si les bases étaient de formes dissemblables, ou si le dé était constitué d'un côté d'une matière plus lourde que de l'autre. [..] Mais qui donc parmi les mortels définira par exemple le nombre de maladies, qui sont autant de cas, qui ont le pouvoir d'envahir les innombrables parties du corps humain à l'âge qu'on voudra et qui ont le pouvoir de nous apporter la mort ? Qui définira combien est plus facile à celle-ci qu'à celle-là, la peste ou l'hydropisie, l'hydropisie ou la fièvre, d'anéantir un homme, en sorte qu'à partir de là puisse être formée une conjecture sur un état futur de vie ou de mort ?

Deuxième extrait, où s'impose une solution : *a priori*, *a posteriori* ; le modèle de l'urne

> [...] Mais à la vérité ici s'offre à nous un autre chemin pour obtenir ce que nous cherchons. Ce qu'il n'est pas donné d'obtenir *a priori* l'est du moins *a posteriori*, c'est-à-dire qu'il sera possible de l'extraire en observant l'issue de nombreux exemples semblables ; ou on doit présumer que, par la suite, chaque fait peut arriver et ne pas arriver dans le même nombre de cas qu'il avait été constaté auparavant, dans un état de choses semblables, qu'il arrivait ou n'arrivait pas.

[...] Enfin il ne peut échapper à personne que, pour juger par ce moyen de quelque événement, il ne suffirait pas d'avoir fait choix d'une ou deux expériences, mais qu'il serait requis une grande quantité d'expériences.[...]

Troisième extrait : le modèle de l'urne et la loi des grands nombres

[...] Il reste assurément à chercher si, en augmentant ainsi le nombre des observations, nous augmentons continuellement. la probabilité d'atteindre le rapport réel entre les nombres de ceux qui font qu'un événement peut arriver et le nombre de ceux qui font qu'il ne peut arriver, de sorte que cette probabilité dépasse enfin un degré quelconque donné de certitude ; ou si le Problème, pour ainsi dire, a son Asymptote, c'est-à-dire s'il existe un degré de certitude qu'il n'est jamais possible de dépasser, de quelque manière qu'on multiplie les observations, d'avoir découvert le vrai rapport des cas ; que, par exemple, nous ne pouvons jamais obtenir de certitude au-delà de la moitié, ou de $\frac{2}{3}$ ou de $\frac{3}{4}$. Un exemple rendra clair ce que je voudrais dire.

Je suppose que, dans une urne, à ton insu soient placées trois mille pierres blanches et deux mille pierres noires ; je suppose, que pour connaître leurs nombres par expérience, tu tires une pierre après l'autre (en replaçant cependant chaque fois la pierre que tu as tirée avant de choisir la suivante pour que le nombre des pierres ne diminue pas dans l'urne) ; tu observes combien de fois sort une pierre blanche et combien de fois une noire. On demande si tu peux le faire tant de fois qu'il devienne dix fois, cent fois, mille fois, etc.., plus probable (c'est-à-dire qu'il devienne moralement certain) que le nombre de fois où tu choisis une pierre blanche et le nombre de fois où tu choisis une pierre noire soient dans ce même rapport sesquialtère où se complaisent à être entre eux les nombres de pierres ou de cas, plutôt que dans tout autre rapport différent de celui-ci. Car si cela ne se produisait pas, j'avoue que c'en serait fait de notre effort pour rechercher expérimentalement le nombre de cas. Mais si nous l'obtenons et si nous acquérons enfin par ce moyen la certitude morale (et je montrerai dans le chapitre suivant que cela aussi se produit réellement), nous aurons trouvé *a posteriori* les nombres de cas presque comme s'ils nous étaient connus *a priori*.

Mais pour que cela ne soit pas compris autrement qu'il ne convient, il faut bien noter ce qui suit ; je voudrais que le rapport entre les nombres de cas que nous entreprenons de déterminer expérimentalement ne fût pas pris de façon nette et sans partage (car ainsi, c'est tout le contraire qui arriverait et il deviendrait d'autant moins probable de découvrir le vrai rapport qu'on ferait de plus nombreuses observations), mais je voudrais que le rapport fût admis avec une certaine latitude, c'est-à-dire compris entre une paire de limites, pouvant être prises aussi rapprochées qu'on voudra.

> Assurément, si dans l'exemple des pierres proposé plus haut nous prenons les deux rapports $\frac{301}{200}$ et $\frac{299}{200}$, ou $\frac{3001}{2000}$ et $\frac{2999}{2000}$, etc... dont le sesquialtère est très près et du plus grand et du plus petit, on montrera que l'on peut arriver à ce que le rapport trouvé grâce à des expériences recommencées de nombreuses fois tombe entre ces limites du rapport sesquialtère plus probablement, de toute probabilité donnée, qu'en dehors.

Exposition et démonstration de ce résultat

> Soit donc le nombre de cas fertiles au nombre de cas stériles précisément ou approximativement dans le rapport $\frac{r}{s}$ et qu'il soit en conséquence, au nombre de tous dans le rapport $\frac{r}{r+s}$ ou $\frac{r}{t}$, rapport qu'encadrent les limites $\frac{r+1}{t}$ et $\frac{r-1}{t}$. Il faut montrer que l'on peut concevoir des expériences en un nombre tel qu'il soit plus vraisemblable d'autant de fois que l'on veut (soit c) que le nombre des observations tombe à l'intérieur de ces limites plutôt qu'en dehors, c'est-à-dire que le nombre des observations fertiles soit au nombre de toutes les observations dans un rapport ni plus grand que $\frac{r+1}{t}$, ni plus petit que $\frac{r-1}{t}$.

Ceci constitue en quelque sorte ce qui deviendra la loi « faible » des grands nombres :

Il est très peu probable que, si l'on fait un nombre suffisant d'expériences, la fréquence d'apparition d'un événement s'écarte notablement de sa probabilité.

Théoriquement, il faudrait faire un nombre infini d'expériences pour connaître la probabilité « vraie » ; Bernoulli démontre qu'un nombre fini d'expériences peut suffire pour connaître une «bonne» valeur approchée de la probabilité. Bernoulli ne s'est pas contenté d'une vague formulation sur la convergence des fréquences, il a en quelque sorte inventé l'idée d'intervalle de confiance.

Siméon-Denis Poisson (1781-1840) utilise pour la première fois en 1837 l'expression « loi des grands nombres » :

> Les choses de toutes natures sont soumises à une loi universelle qu'on peut appeler la loi des grands nombres.

Loi ? Que l'on démontre comme un Théorème ?

D'une certaine manière, pour faire des probabilités, (et surtout les appliquer), tout se passe comme si on devait admettre que la nature suit (au moins localement) des lois, qui permettent l'élaboration d'une théorie, à l'intérieur de laquelle on démontre ensuite cette loi, dans une posture proche de celle du physicien.

Dans les faits, ces questions ne se sont posées que lorsqu'on a voulu appliquer les probabilités à d'autres domaines que les jeux de hasard.

Elles recoupent aussi des questions métaphysiques : si la nature suit la loi des grands nombres, c'est qu'il y a « quelqu'un » qui a fabriqué le dé...

Quoi qu'il en soit, l'importance de son résultat n'échappe pas à Jacques Bernoulli :

> J'estime cette invention bien davantage que si j'avais livré la quadrature même du cercle, car si celle-ci était effectivement trouvée, son utilité serait peu considérable.

Abraham de Moivre (1667-1754) poursuivra le travail de Bernoulli[5] avec des méthodes différentes. Il investit dans le calcul des probabilités les idées du calcul infinitésimal et donne une expression plus complète du théorème de J. Bernoulli sur les épreuves répétées, avec une estimation de la loi de l'écart entre probabilité *a priori* et probabilité *a posteriori*.

Il établit la formule dite de Stirling (et Moivre) pour approcher $n!$ pour n très grand.

$$1 \times 2 \times 3 \times \ldots \times n \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Ceci lui permet d'explorer les limites de ce que nous nommons maintenant la loi binomiale en établissant ce qui deviendra un jour la loi normale.

De là, il s'interroge naturellement sur la nature du hasard.

> De ce qui a été dit, il suit que le Hasard perturbe très peu les événements qui dans leur institution naturelle sont amenés à se produire ou non, suivant une certaine loi déterminée.
>
> Bien que le Hasard produise des irrégularités, cependant avec l'avancement du temps, ces irrégularités n'auront aucune proportion avec le retour de l'ordre qui résulte naturellement du DESSEIN ORIGINEL.

Pierre-Simon Laplace (1749-1827)[6] reprend la démonstration de « la loi des grands nombres » et les travaux de De Moivre. Nous y reviendrons dans notre deuxième époque. Il développe à cette occasion sa profession de foi déterministe.

> Nous devons envisager l'état présent de l'univers comme l'effet de son état antérieur, et comme la cause de celui qui va suivre. Une intelligence qui pour un instant donné connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir comme le passé serait présent à ses yeux.

La loi des grands nombres, permettant de définir une probabilité *a posteriori*, semble paradoxalement faire disparaître l'objet des probabilités, le hasard.

II- Deuxième époque : Loi des erreurs, méthode des moindres carrés

De la détermination de la « vraie » valeur

À l'origine on trouve … l'astronomie et la géodésie.

Vers le milieu du XVIII[e] siècle, la modélisation des erreurs de mesure va devenir centrale pour l'existence même d'une science. Les instruments de mesure se sont améliorés.

---

[5] De Moivre, Abraham, *Doctrine of chances*, 1718-1738-1756.

[6] Laplace, Pierre Simon (de), *Théorie analytique des probabilités*, Courcier, Paris, 1812.

En astronomie, par exemple, les mesures se ramènent essentiellement à la mesure d'angles. Jusqu'à la révolution, l'instrument principal est le quart de cercle ou le secteur et il est possible de mesurer des angles à une demi-minute de degré près.

Cette précision ne suffit pas à réduire les erreurs. Il y a l'imprécision irréductible des résultats de l'observation. Or, la précision est une condition de la prévision. Par exemple, Lalande, Clairaut et Reine Lepaute, après un an de calculs, se trompent d'un mois sur le passage au périhélie de la comète de Halley en 1759.

Au delà des enjeux de pure connaissance scientifique, se profilent des enjeux de société de premier plan : l'enjeu de la puissance maritime, donc commerciale et militaire, liée entre autres au problème de la mesure de la longitude.

C'est aussi une occasion de prendre conscience de la difficulté de mesurer, de se rendre à l'évidence qu'il est impossible, expérimentalement, d'obtenir un résultat dont on puisse affirmer que c'est la « vérité ».

Se pose alors la question : puisqu'on ne peut avoir la « vraie » valeur, comment l'approcher le mieux possible, et comment avoir une idée de l'erreur commise ?

Le travail des mathématiciens va être de trouver les moyens de combiner de nombreuses mesures d'un même phénomène, qui réduisent le plus possible l'erreur finale.

Plusieurs idées vont se faire jour. Lagrange[7] par exemple va suivre la voie probabiliste, dans son *Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations, dans lequel on examine les avantages de cette méthode par le calcul des probabilités, et où l'on résout différents problèmes relatifs à cette matière.*

> Quand on fait plusieurs observations d'un même phénomène dont les résultats ne sont pas tout à fait d'accord, on est sûr que ces observations sont toutes, ou au moins en partie, peu exactes, de quelque source que l'erreur puisse provenir ; alors, on a coutume de prendre le milieu entre tous ces résultats, parce que de cette manière les différentes erreurs se répartissant également dans toutes les observations, l'erreur qui peut se trouver dans le résultat moyen devient aussi moyenne entre toutes les erreurs. Or, quoique tout le monde reconnaisse l'utilité de cette pratique pour diminuer, autant qu'il est possible, l'incertitude qui naît de l'imperfection des instruments et des erreurs inévitables des observations, j'ai cru cependant qu'il serait bon d'examiner et d'apprécier par le calcul les avantages qu'on peut espérer de retirer d'une semblable méthode ; c'est l'objet que je me suis proposé dans ce mémoire.

Voici les grandes lignes de la méthode de Lagrange. Il détermine *a priori* ce qu'il appelle la loi de *facilité* des erreurs, nom qui sera repris par Legendre et Gauss :

Il suit la « règle ordinaire du calcul des probabilités », rapport du nombre de cas favorables aux nombres de cas possibles, en évaluant des probabilités *a priori.*

Il évalue alors la probabilité que l'erreur commise en prenant le résultat « moyen » soit comprise entre des limites données.

Ayant ainsi construit un « modèle », il montre comment déterminer *a posteriori* la loi de probabilité des erreurs.

Une partie de la question tourne autour de « prendre le milieu ». Voici par exemple l'article « milieu » du Supplément à l'Encyclopédie de Diderot et d'Alembert, par Jean III Bernoulli, 1770.

---

[7] Lagrange, Joseph-Louis, 1774, in Miscellanea Taurinensia, Turin.

> Il n'est pas douteux que cette pratique (prendre le milieu entre tous les résultats) ne soit très utile pour déterminer l'incertitude qui naît de l'imperfection des instruments et des erreurs inévitables des observations ; mais il est aisé de s'apercevoir qu'elle ne la diminue pas autant qu'on le désirerait, et qu'elle est susceptible à plus d'un égard d'être perfectionnée, parce qu'en prenant simplement le milieu arithmétique, on ne tient pas compte du plus ou moins de probabilité de l'exactitude des observations, des différents degrés d'habileté des observateurs, etc. Différents grands géomètres ont entrepris cette recherche utile, ils l'ont considérée sous différents points de vue et l'ont traitée plus ou moins en détail ; il est fort à souhaiter que les astronomes, les physiciens et généralement tous les observateurs profitent des résultats de ces recherches dans la discussion de leurs observations.

Milieu = moyenne, moyenne pondérée, médiane, … Le mot est imprécis, et la pratique ne va pas de soi.

Lors de l'expédition à l'équateur, les mesures sur le terrain, faites par La Condamine et Bouguer étaient peu concordantes. La Condamine proposa alors à Bouguer de faire la moyenne des mesures (faites avec le plus grand soin) ; Bouguer refuse, estimant que ce serait une tricherie.

Parfois aussi on estime que l'un des participants est « meilleur » que les autres, donc il est normal de prendre sa mesure pour la meilleure.

La multiplication des observations est-elle nécessaire ? Les instruments étant plus perfectionnés, une seule bonne mesure suffit peut-être.

Un milieu ayant été choisi, la deuxième étape est l'estimation de l'erreur, par rapport à la « vraie » valeur, toujours inconnue. On cherche alors la loi « de facilité des erreurs ».

Une autre question surgit : la précision du milieu, dépendra-t-elle alors de la loi de facilité des erreurs ?

Pour établir un bon modèle de la loi des erreurs, il faut tenir compte des contraintes « naturelles » qui ont déjà été signalées par Galilée :

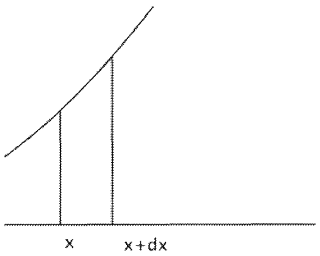
- les observations sont distribuées de façon symétrique autour de la « vraie » valeur ; c'est-à-dire que les erreurs sont distribuées de façon symétrique autour de zéro. (*a priori* on fait autant d'erreurs dans un sens que dans l'autre) .
- les petites erreurs arrivent plus fréquemment que les grandes erreurs (parce que l'observateur est consciencieux, donc il ne devrait pas faire beaucoup de grosses erreurs).

Gauss exprimera ces contraintes dans sa « *Méthode des moindres carrés* » en 1821.

> Les erreurs qui, dans des observations de même espèce, proviennent d'une cause simple et déterminée se trouvent renfermées entre certaines limites (…)
>
> Désignons par $F(x)$ la facilité relative d'une erreur x : on doit entendre par là, à cause de la continuité des erreurs, que $F(x)\mathrm{d}x$ est la probabilité que l'erreur soit comprise entre les limites x et $x + \mathrm{d}x$. Il n'est pas possible, en général, d'assigner la forme de la fonction F, et l'on peut même affirmer que cette fonction ne sera jamais connue dans la pratique. On peut néanmoins établir plusieurs caractères généraux qu'elle doit nécessairement présenter : (…) Elle s'annule pour toutes les valeurs de x non comprises entre les valeurs extrêmes. Pour toute valeur comprise entre ces limites, la fonction est positive. (…)

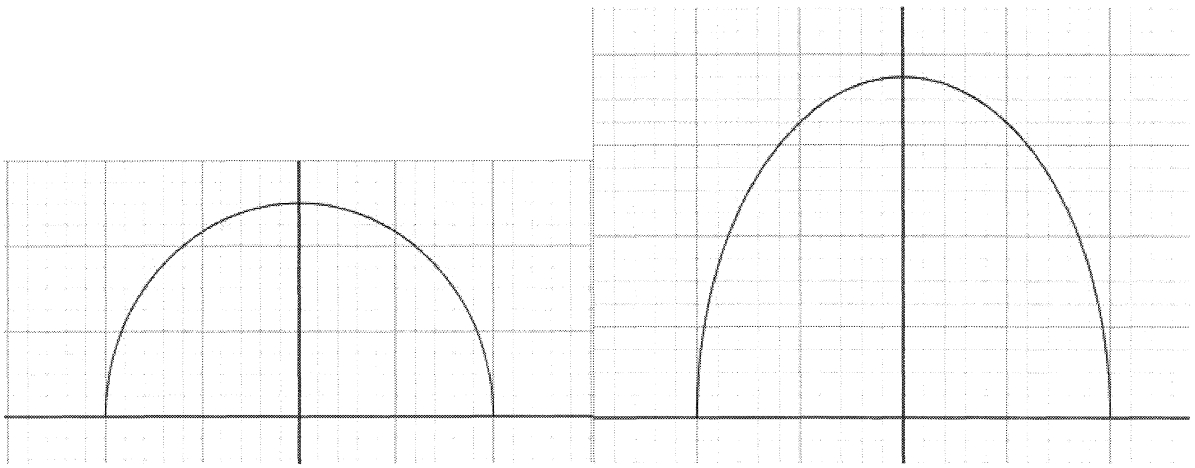
> Dans la plupart des cas, les erreurs égales et de signes contraires seront également probables. (…) Enfin, comme les petites erreurs sont plus facilement commises que les grandes, $F(x)$ sera en général maximum pour $x = 0$ et diminuera sans cesse lorsque x croîtra.

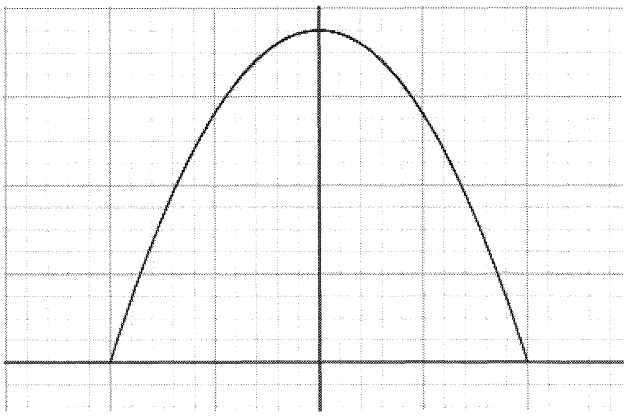
Pour une meilleure compréhension, nous ajoutons ce petit graphique, où la courbe serait celle de la fonction F, et la probabilité serait l'aire sous la courbe entre x et $x + dx$.



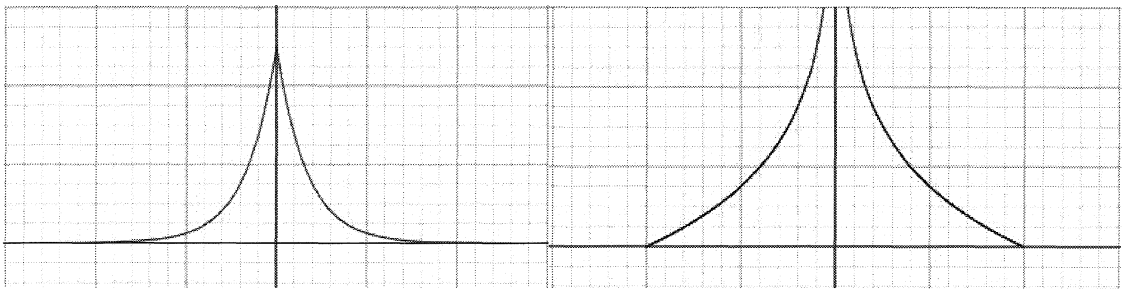
Il s'agit donc de trouver un modèle pour la « courbe d'erreur », qui obéit à ces contraintes. Plusieurs sont proposés dont voici quelques exemples.

Daniel Bernoulli en a proposé trois : en arc de cercle, d'ellipse ou de parabole.





Laplace propose deux modèles de courbes d'erreurs : une en $e^{-m|x|}$ sur $]-\infty,+\infty[$ (c'est la première loi des erreurs de Laplace) et l'autre en $\log(a/|x|)$ sur $]-a,+a[$.



Distribution en $e^{-m|x|}$ Distribution en $\log(a/|x|)$

Lagrange envisage une distribution cosinusoïdale.

Toutes ces courbes ont les caractéristiques demandées par Gauss.

Le problème du choix du « milieu » est lié au choix d'un modèle de distribution des erreurs et à l'estimation de la « vraie » valeur. Le meilleur milieu sera celui qui minimisera les erreurs.

La première idée, la plus naturelle, est de minimiser la somme des écarts à la vraie valeur. On démontre alors que c'est la médiane qui minimise cette somme.

Dans le modèle de Laplace, en exponentielle, le milieu est la médiane[8].

La méthode des moindres carrés ; la loi de Laplace-Gauss

Gauss choisit de minimiser la somme des carrés des écarts à la vraie valeur. Ce qui revient à privilégier la moyenne. Idée étrange *a priori*, mais la moyenne était finalement utilisée de façon empirique depuis « toujours ». Du moins, c'est une sorte de postulat.

> Il est devenu habituel de considérer comme axiome l'hypothèse que, si une quantité a été déterminée par plusieurs observations directes, faites dans les mêmes circonstances et avec le même soin, la moyenne arithmétique des valeurs observées donne la valeur la plus probable, si ce n'est exactement, du moins approximativement, si bien qu'il est toujours plus sûr d'y recourir.

Ainsi, le choix de la minimisation de la somme des carrés des écarts va s'imposer. Citons Legendre (*Nouvelles méthodes pour la détermination des orbites des comètes - 1805*).

> De tous les principes que l'on peut proposer pour cet objet, je pense qu'il n'en est pas de plus général, de plus exact, ni d'une application plus facile que celui dont nous avons fait usage dans les recherches précédentes et qui consiste à rendre minimum la somme des carrés des erreurs. Par ce moyen, il s'établit entre les erreurs une sorte d'équilibre qui, empêchant les extrêmes de prévaloir, est très propre à faire connaître l'état du système le plus proche de la vérité.
>
> La règle par laquelle on prend le milieu entre les résultats de différentes observations n'est qu'une conséquence très simple de notre méthode générale, que nous appellerons méthode des moindres carrés.

On ne sait lequel de Legendre ou Gauss a eu l'idée en premier. Legendre a la priorité de la publication, en 1805.

---

[8] Le terme « médiane » ne sera introduit qu'en 1843, par Cournot, dans « *Exposition de la théorie des chances et des probabilités* ». En 1757, Boscovitch est un des premiers à souligner l'intérêt de cette valeur centrale.

Gauss a affirmé qu'il la connaissait et l'utilisait longtemps avant (publiée en 1809, mais utilisée prétend-il, dès 1795, lors d'observations astronomiques quand il était encore étudiant[9]).

Le modèle qui privilégie la moyenne et donc la méthode des moindres carrés, va s'imposer comme étant **La** loi de distribution des erreurs, non sans mal.

Voici schématiquement le passage de la méthode des moindres carrés à la loi gaussienne.

Soit q la quantité inconnue sur laquelle on possède plusieurs mesures.

On constate d'abord que l'estimation de q rendant minimale la somme des carrés des erreurs est la moyenne des mesures par annulation de la dérivée pour $x = (x_1+x_2+\ldots+x_n)/n$

$$[(x-x_1)^2 + (x-x_2)^2 +\ldots+ (x-x_n)^2]'=2(x-x_1) + (x-x_2) +\ldots+ (x-x_n)$$

Puis, si l'on considère que les erreurs sont aléatoires, alors, Gauss montre que la loi validant la moyenne et la méthode des moindres carrés comme meilleure approximation, est la loi « gaussienne ». Gauss exprime sa « loi des erreurs » par

$$\Phi(x) = \frac{1}{h\sqrt{\pi}} e^{-x^2/h^2},$$

où h est un coefficient proportionnel à un nombre m qu'il appelle « erreur moyenne à craindre », qui sera nommé en 1893 « standard deviation » par Karl Pearson. Il le notera σ et c'est notre écart-type ; $\Phi(x)$ est ce que nous nommons la densité de probabilité.

Cette équation est apparemment très compliquée, mais sa forme très connue : il s'agit de la « courbe en cloche ».

La difficulté, c'est que l'on tourne un peu en rond : la méthode des moindres carrés et la « loi de Gauss » se confortent l'une l'autre, mais est-ce la meilleure adéquation à la réalité ?

Si l'on suppose que les mesures sont distribuées suivant une « courbe de Gauss », la valeur la plus vraisemblable de la constante cherchée est celle que donne la méthode des moindres carrés, c'est-à-dire la moyenne arithmétique ; et c'est la moyenne arithmétique qui se trouve également avoir une loi de « facilité » du type gaussien.

C'est la critique de Laplace[10] :

« *Mais, comme rien ne prouve que la première de ces règles* (la règle du milieu arithmétique) *donne le résultat le plus avantageux, la même incertitude existe par rapport à la seconde* (la règle des moindres carrés) ».

Laplace reprend le problème et essaie de donner une justification qui ne s'appuie plus sur le principe de moyenne, mais uniquement sur le calcul de probabilité. Il ne se situe plus dans la problématique de la recherche de la « vraie valeur » et du calcul d'erreurs, mais dans le contexte de la loi binomiale de Bernoulli, prolongeant les travaux de De Moivre.

Il montre que la somme, ou la moyenne arithmétique d'un grand nombre de variables indépendantes, dont les courbes de facilité sont quelconques, suit asymptotiquement une loi gaussienne. C'est le théorème de Laplace. Et la loi « de Gauss » devient la loi de Laplace-Gauss.

En fait, bien sûr, on ne connaît toujours pas la « vraie valeur ». On minimise le mieux possible les erreurs, mais on aurait pu les minimiser autrement.

Dans le texte suivant, Gauss lui-même met très bien en évidence l'arbitraire de ses choix.

---

[9] Chabert, Jean-Luc, « Gauss et la méthode des moindres carrés », *Revue d'histoire des sciences*, 1989.

[10] Laplace, Pierre-Simon (de), *Théorie analytique des probabilités*, 1812.

La question qui nous occupe a, dans sa nature même, quelque chose de vague et ne peut être bien précisée que par un principe jusqu'à un certain point arbitraire. La détermination d'une grandeur par l'observation peut se comparer, avec quelque justesse, à un jeu dans lequel il y aurait une perte à craindre et aucun gain à espérer : chaque erreur commise étant assimilée à une perte que l'on fait, la crainte relative à un pareil jeu doit s'exprimer par la perte probable, c'est-à-dire par la somme des produits des diverses pertes possibles par leurs probabilités respectives. Mais quelle perte doit-on assimiler à une erreur déterminée ?

C'est ce qui n'est pas clair en soi ; cette détermination dépend en partie de notre volonté. Il est évident, d'abord, que la perte ne doit pas être regardée comme proportionnelle à l'erreur commise ; car, dans cette hypothèse, une erreur positive représentant une perte, l'erreur négative devrait être regardée comme un gain : la grandeur de la perte doit, au contraire, s'évaluer par une fonction de l'erreur dont la valeur soit toujours positive.

Parmi le nombre infini de fonctions qui remplissent cette condition, il semble naturel de choisir la plus simple, qui est, sans contredit, le carré de l'erreur, et, de cette manière, nous sommes conduits au principe proposé plus haut.

Laplace a considéré la question d'une manière analogue, mais en adoptant, pour mesure de la perte, l'erreur elle-même prise positivement. Cette hypothèse, si nous ne nous faisons pas d'illusion, n'est pas moins arbitraire que la nôtre : faut-il, en effet, regarder une erreur double comme plus ou moins regrettable qu'une erreur simple répétée deux fois, et faut-il, par suite, lui assigner une importance double ou plus que double ? C'est une question qui n'est pas claire, et sur laquelle les arguments mathématiques n'ont aucune prise ; chacun doit la résoudre à son gré. On ne peut nier pourtant que l'hypothèse de Laplace ne s'écarte de la loi de continuité et ne soit, par conséquent, moins propre à une étude analytique ; la nôtre, au contraire, se recommande par la généralité et la simplicité de ses conséquences.

Gauss a choisi comme erreur moyenne à craindre (= écart type de la loi des erreurs, supposée centrée, dont le carré m^2 est la variance) :

$$m = \sqrt{\int_{-\infty}^{+\infty} x^2 \varphi(x) \mathrm{d}x},$$

là où Laplace choisissait :

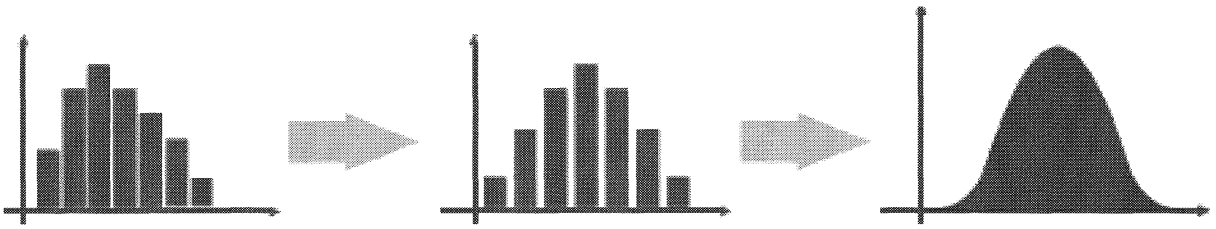
$$\int_{-\infty}^{+\infty} |\Delta| \varphi(\Delta) \mathrm{d}\Delta.$$

Cette théorie va résister aux mises à l'épreuve successives. De nombreuses améliorations seront données à la théorie des erreurs tout au long du XIX^{ème} siècle et début du XX^{ème} et la méthode des moindres carrés se répand, avec la loi de Laplace-Gauss. Une des grandes justifications est celle de Laplace, cette loi apparaissant comme limite de la loi binomiale sous certaines conditions.

De la loi binomiale à la loi normale :

Loi Binomiale
Bi (n ; p)

Loi normale (de Laplace Gauss)
$N(\mu ; \sigma^2)$



p tend vers 0,5

n supérieur à 25

Elle donne pourtant un poids important aux grandes erreurs. Elle pose des problèmes de robustesse, de sensibilité aux valeurs accidentelles. Elle n'est pas adaptée à toutes les situations.

C'est pourquoi la médiane est réhabilitée par Cournot (1801-1877) et Galton (1822-1911), et même, la moyenne géométrique est réintroduite (en effet, la comparaison de deux valeurs n'est pas seulement basée sur leur différence, mais aussi sur leur rapport), ou même la moyenne harmonique (exemple de la vitesse moyenne entre l'aller et le retour).

En fait, la loi de Laplace-Gauss est petit à petit considérée, dans les manuels, comme la loi des erreurs par excellence, bien que peu de vérifications empiriques aient pu être faites.

Nous sommes dans un monde statistique, régi par des lois arbitraires, faute de mieux. Et cela ne fonctionne pas trop mal. Il faut cependant conserver son sens critique.

Voici une boutade rapportée par Poincaré :

> Tout le monde y croit cependant, me disait un jour M. Lippman, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental.

Finalement, l'ensemble moindres carrés, moyenne, loi gaussienne, consacre l'usage de la moyenne (et donc de l'écart type, ou erreur moyenne à craindre), les uns justifiant les autres.

La moyenne a pris cependant un rôle prépondérant dans la recherche des « vraies » valeurs des phénomènes, à tel point que certains ont eu l'idée d'appliquer ces résultats à d'autres domaines.

III- Troisième période : les moyennes « subjectives » et les statistiques sociales

Les premiers contacts de nos élèves avec LA moyenne sont souvent la moyenne des notes, la moyenne des tailles des élèves, …

Est-ce le même problème que précédemment ? Y a-t-il une « vraie » note ? Une « vraie » taille ? N'y aurait-il pas une notion de normalité qui se dégagerait de cette question ?

Les méthodes de calcul de ces moyennes sont exactement les mêmes que pour déterminer les « vraies valeurs ».

Ces moyennes d'un autre type, qui donnent des renseignements d'un autre ordre, ont-elles quelques relations avec les précédentes ?

L'un de ceux qui vont contribuer à introduire les méthodes statistiques dans les questions de sociologie est le mathématicien et astronome belge, Adolphe Quételet (1796-1874).

> L'homme que je considère ici est, dans la société, l'analogue du centre de gravité dans les corps ; il est la moyenne autour de laquelle oscillent les éléments sociaux : ce sera si l'on veut un être fictif pour qui toutes les choses se passeront conformément aux résultats moyens obtenus pour la société. Si l'on cherche à établir, en quelque sorte, les bases d'une physique sociale, c'est lui qu'on doit considérer, sans s'arrêter aux cas particuliers ni aux anomalies.

Voici le raisonnement suivi pour expliquer l'analogie entre la moyenne d'une série de *n* mesures d'un même objet (la moyenne objective), et la moyenne d'une série de *n* objets différents (moyenne subjective).

Si on fait mille mesures du tour de poitrine d'une statue, et que l'on représente graphiquement la série de mesures, on obtient une courbe de Gauss.

Si, au lieu de mesurer mille fois la statue, on demande à mille sculpteurs de la copier, les mesures obtenues sur le tour de poitrine des mille statues suivront aussi une loi de Gauss.

Maintenant, si, au lieu de mesurer mille statues, on mesure le tour de poitrine de mille individus, où est la différence ? On obtiendra aussi, nécessairement, une courbe de Gauss.

Avant d'aller plus avant, nous devons signaler que la courbe de Gauss ou « courbe en cloche », n'apparaissait pas dans les premiers écrits. Ce n'est que peu à peu qu'elle a commencé de remplir les ouvrages, au point de devenir emblématique, entre autre avec Quételet.

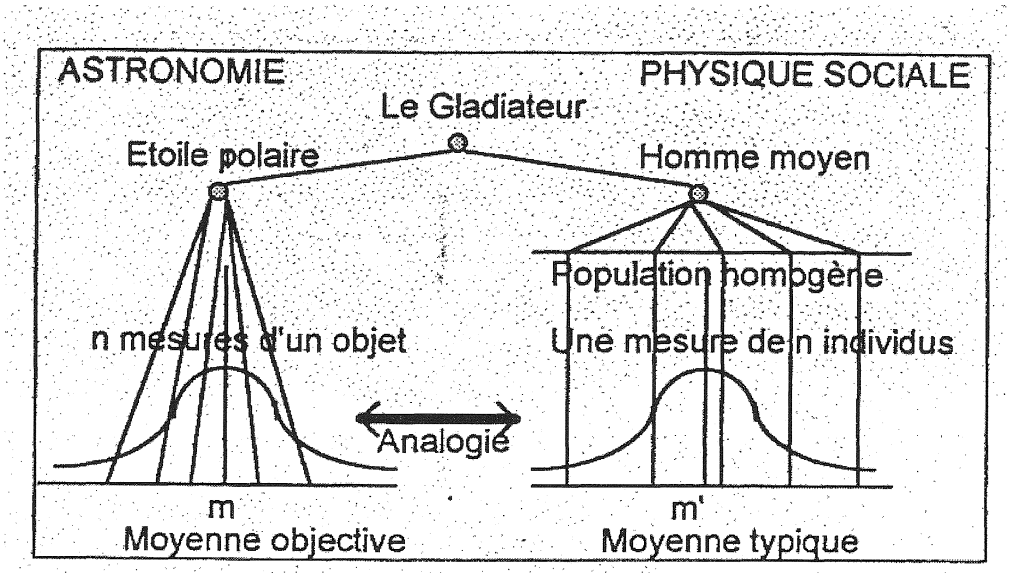
> Je demande maintenant s'il serait exagéré de parier un contre un qu'une personne peu exercée à prendre des mesures sur le corps humain va se tromper de 33 millimètres environ, en mesurant une poitrine de plus d'un mètre de circonférence ? Eh bien, en admettant cette erreur probable, 5738 mesures prises sur une même personne ne se regrouperaient pas avec plus de régularité, quant à l'ordre de grandeur, que les 5738 mesures prises sur les soldats écossais. Et si l'on connaît 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 plus grossiers.
>
> L'exemple que je viens de citer mérite, je crois, toute notre attention : il nous montre que 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. Si telle n'était pas la loi de la nature, les mesures ne se regrouperaient pas malgré leurs défectuosités, avec l'étonnante symétrie que leur assigne la loi de possibilité[11].

Quételet avait observé que les distributions « en chapeau de gendarme » apparaissaient non seulement en astronomie, mais aussi dans d'autres mesures, comme celle des tailles des conscrits. Au bout de quelque temps, il en vint à imaginer un « homme moyen », qui serait le plus équilibré possible, mais sans qualité distinctive, donc représentatif de n'importe qui.

---

[11] Quételet, Adolphe, *Lettres à son Altesse Royale le Duc de Saxe Cobourg et Gotha sur la théorie des probabilités*, Bruxelles, 1846.

Les humains réels se distribueraient autour de cet homme moyen abstrait, comme les erreurs dans la mesure d'une grandeur physique.



Le transport analogique de la moyenne

La démarche statistique fait émerger une tendance globale des aléas individuels.

Ce sont les Bertillon (père et fils), qui, à la fin du XIX^{ème} siècle, distingueront la « moyenne objective », celle de la théorie des erreurs, correspondant à un objet réel, de la « moyenne subjective », celle de l'homme moyen, qui fonde un objet fictif nouveau, sous la condition de « normalité ». Du coup, la loi de Gauss deviendra le plus souvent la loi « normale ».

La statistique moderne, celle que l'on utilise par exemple dans les sciences sociales, naîtra en Angleterre. Il s'agit, dépassant Quételet, de le transformer complètement, en particulier sur la question de l'homme moyen. Ce sont les travaux de Francis Galton (1822-1911), cousin de Darwin, qui sont à l'origine de cette statistique. La loi normale va maintenant être plutôt utilisée comme un outil pour étalonner les aptitudes. Et, bien sûr, il s'agit d'améliorer l'espèce en favorisant ceux qui sont au-dessus de la moyenne. Il est facile de voir les glissements possibles vers des théories eugénistes.

Galton examine la dispersion des résultats en développant les notions de médiane et de quartiles.

Karl Pearson (1857-1936) poursuivra les travaux de Galton, en particulier l'étude de la corrélation et invente le critère du χ^2 permettant de mesurer la qualité d'ajustement d'une distribution théorique à une distribution observée.

Il fait pour cela des « expériences », de lancers de dés, de pièces de monnaie, … lui-même ou ses élèves, ou ses proches.

Il faudrait, pour être complet, citer les travaux de Ronald Fisher (1890-1962) et William Gosset dit Student (1876-1937), fondateurs de la biométrie, et les premiers à faire une différence entre le domaine des probabilités et celui des statistiques maintenant mathématiques.

Conclusion

Pour conclure sur cette histoire, nous pourrions dire que les statistiques actuelles sont évidemment héritières de ces divers courants. Mais, du moins si on se laisse porter par les

apparences, la statistique est plutôt devenue une méthodologie, un ensemble de techniques, que le spécialiste sur son terrain appliquera, en interprétant les résultats dans le champ qui lui est propre. Il semble pourtant essentiel d'examiner sans cesse à quoi servent les outils statistiques, pourquoi et dans quel contexte philosophique ou social ils ont été pensés.

Du côté des probabilités, Giorgio Israel[12] souligne les dangers d'une axiomatique pure qui pourrait les couper des relations au réel.

> Une fois de plus, c'est la notion générale de hasard qui est en jeu. Une définition de ce concept en termes empiriques soulève une foule de problèmes presque inextricables. On peut certes recourir à une définition éliminant tous les problèmes, une définition abstraite et axiomatique qui vide les concepts de leur contenu et les rend inoffensifs. L'axiomatique est tout à fait capable de réaliser ce genre de miracle. C'est en effet un point de vue axiomatique qui a été introduit dans le calcul des probabilités au début des années 1930 par les mathématiciens soviétiques Alexandre Iakovlévitch Khintchin (1894-1959) et Andréï Nikolaïévitcth Kolmogorov (1903-1987). Il a connu depuis un immense développement. Comme dans d'autres domaines, il s'est imposé comme moyen de sortir des difficultés logiques posées par les approches classiques. (...) Encore une fois c'est dans le cadre de la mathématisation des sciences morales, sociales et économiques, que se pose la question du rôle des probabilités et de la dialectique entre vision objectiviste et vision subjectiviste du hasard. (...) Dans cette perspective, les approches objectivistes, comme l'axiomatique, conduisent tout simplement à nier ou à sous estimer la valeur probabiliste des sciences non physiques.

Du côté des statistiques, Florence Nightingale (1820-1910) plaidera largement en faveur de leur enseignement, étant donné leur utilité dans de nombreux domaines.

> Je n'ai pas besoin de vous dire, à vous qui êtes l'inventeur, le maître de toute la science statistique et de nous tous, l'extrême importance pour le législateur anglais, l'homme d'état, le « Member of Parliament », de se familiariser de bonne heure, étant encore à l'université, avec vos recherches, avec vos découvertes, puisque toute la science, je dirai plutôt la pratique de l'administration, de la politique, de la législation, de l'économie sociale, doit reposer sur les bases que vous avez jetées. [...]
>
> Il me tarde de la voir prendre (la physique sociale) sa place comme un sujet reconnu parmi les études et les examens de nos grandes universités[13].

Au final, les calculs mathématiques, en tant que tels et pris isolément, ne sont pas contestables. L'usage qu'on en fait demande un regard critique. Il faut reconnaître que cela ne semble pas si mal marcher. Une fois de plus, les mathématiques serviraient à décrire et comprendre le réel. Il faut prendre garde, cependant, que ce regard sur le réel ne soit pas biaisé. Un regard sur la construction des concepts peut contribuer à nous préserver de cet écueil.

---

[12] Israel, Giorgio, *La mathématisation du réel*, Seuil, Paris, 1996.

[13] Lettre de Florence Nightingale à Quételet, 8 novembre 1872.

Éléments de bibliographie

Bernoulli, Jacques, *Ars conjectandi*, traduction et commentaires de Norbert Meusnier et Bernard Lalande, IREM de Rouen, 1987.

Boursin, Jean-Louis, *Les structures du hasard*, Points Seuil, 1986.

Boyé, Anne, Comairas, Marie-Céline, Moyenne, médiane, écart type, quelques regards sur l'histoire pour éclairer l'enseignement des statistiques au lycée, in *Repères IREM*, 48, 2002.

Boyé, Anne, Comairas, Marie-Céline, *De l'urne de Bernoulli à l'homme moyen*, IREM des Pays de la Loire, 2003.

Boyé, Anne, *Des difficultés d'enseigner le hasard et les probabilités*, IREM des Pays de la Loire, 1996.

Chabert, Jean-Luc, « Gauss et la méthode des moindres carrés », in *Revue d'Histoire des sciences*, vol. 42, n° 1, Paris, 1989.

Courtebras, Bernard, *A l'école des probabilités*, Presses universitaires de Franche-Comté, 2006.

Dahan Dalmedico, Amy, « Le déterminisme de Pierre-Simon Laplace et le déterminisme aujourd'hui », in *Chaos et déterminisme*, Seuil, Paris, 1992.

Gauss, Carl Friedrich, *Méthode des moindres carrés*, traduit par J. Bertrand, Mallet-Bachelier, Paris, 1855, rééd. IREM Paris VII, 1996.

Israel, Giorgio, *La mathématisation du réel*, Seuil, Paris, 1992.

Lacroix, Sylvestre-François, *Traité élémentaire de calcul des probabilités*, 2e édition, Paris, 1822, rééd. IREM Paris VII, 1994.

Lagrange, Joseph-Louis, « Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations », in *Miscellanea Taurinensia*, 1774.

Laplace, Pierre-Simon de, *Théorie analytique des probabilités*, Paris, Courcier, 1812.

Poincaré, Henri, *La science et l'hypothèse*, 1902, Flammarion, Paris, 1968.

Quételet, Adolphe, *Lettres à son Altesse Royale le Duc de Saxe Cobourg et Gotha sur la théorie des probabilités*, Bruxelles, 1846.

Quételet, Adolphe, *Sur l'homme et le développement de ses facultés ou Essai de physique sociale*, Bachelier, Paris, 1835.

Quinio Benamo, Martine, *Probabilités et statistiques aujourd'hui*, L'Harmattan, Paris, 2009.

Raymond, Pierre, *De la combinatoire aux probabilités*, Maspero, Paris, 1975.

Stigler, Stephan M., *The history of statistics, The measurement of uncertainty before 1900*, 9e édition, Belknap Harvard, Cambridge, USA, 2003.