De Cassini à Gauss : du calcul d'erreurs aux probabilités.

Xavier LEFORT, Anne BOYÉ

L'application des probabilités à la mesure des phénomènes physiques demeure tardive, plus dépendante bien sûr des progrès de la physique que du développement de la théorie mathématique. Si, traditionnellement, il est possible de voir l'origine du calcul des probabilités dans les travaux de Pascal et Fermat, son utilisation en terme de calcul d'erreurs ou de recherche de valeur vraie ne s'entrevoit qu'à peine dans les écrits de Jacques Bernouilli, pour ne prendre de l'importance qu'à la fin du XVIII° siècle et au début du XIX° siècle, avec Laplace et Gauss. Mais il s'agit alors plus d'un élément moteur de la recherche, que d'une simple application, puisque ce sont des considérations sur des problèmes de mesures qui génèrent, en particulier chez Gauss, l'établissement de la "loi du hasard", dite aussi "loi normale", ou "loi de Gauss", voire, "loi de Gauss-Laplace".

Nous vous proposons, à partir de quelques textes, de suivre l'émergence des relations entre l'étude d'un phénomène et sa mesure d'une part, et l'utilisation de la théorie mathématique d'autre part.

L'atelier consiste alors en une présentation des auteurs, suivie de la lecture commentée et discutée, voire exploration, de leurs écrits.

Jacques Bernouilli[1]

Né à Bâle en 1654, Jacques Bernouilli est un mathématicien incontournable dans la genèse du calcul des probabilités. Si son propos ne concerne pas directement la recherche d'une valeur exacte lors d'expériences répétées, et subséquemment le calcul d'erreurs, son ouvrage *Ars Conjectandi*, sera durant longtemps la référence de ceux qui se pencheront sur ces problèmes.

L'Ars conjectandi est un ouvrage posthume, paru en 1713, alors que l'auteur décède en 1705. Quelques difficultés familiales ont retardé sa parution. Il semble que, dès 1689, Jacques Bernouilli se soit intéressé à la question, mais ait différé la parution de son travail, souhaitant sans doute à la fois l'améliorer et l'illustrer d'exemples concrets.

Un des passages les plus représentatifs pour notre propos reste la proposition principale de la IV° partie:

Jacques BERNOUILLI "Ars Conjectandi" 1713

Propos. Princip. Il s'ensuit enfin la proposition elle-même, pour laquelle tout cela a été formulé, mais dont la démonstration se fait maintenant par la seule application des lemmes préparatoires à l'objet présent. Pour éviter la fatigue d'une circonlocution, j'appellerai "*féconds*" ou "*fertiles*" les cas dans lesquels un événement peut se produire, et "*stériles*" ceux dans lesquels le même événement ne peut se produire : de même, j'appellerai expériences "*fécondes*" ou "*fertiles*" celles pour lesquelles on constate qu'un des cas fertiles peut survenir, et "inféconde" ou "*stériles*" celles pour lesquelles on observe qu'un des cas stériles se produit. 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 $\frac{r+l}{t}$ et $\frac{r-l}{t}$. 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 $\frac{r+l}{t}$, ni plus petit que $\frac{r-l}{t}$.

1 Voir au sujet de J.Bernouilli le travail de N.Meusnier, IREM de ROUEN, 1987

Cesar-François Cassini de Thury

En 1668, l'Académie Royale des Sciences avait fait venir d'Italie Jean-Dominique Cassini pour diriger l'Observatoire Royal, puis pour travailler à l'établissement de la première carte de France à l'échelle[2]. L'ouvrage était de longue haleine : son fils Jacques Cassini, puis son petit fils, Cesar-François Cassini de Thury le poursuivirent, et ce n'est qu'après la révolution française qu'il fut complètement achevé.

Le travail de cartographie nécessitait un nombre considérable de mesures sur le terrain, en particulier pour déterminer le plus précisément possible la position géographique des lieux remarquables, à commencer par l'observatoire de Paris. On remarquera dans le texte suivant issu de *La description géométrique de la France* parue en 1783, d'une part le soin apporté aux mesures de "hauteur du pôle" (latitude), d'autre part la manière "subjective"de déterminer la valeur la plus vraisemblable.

Cesar-François CASSINI DETHURY
"Description géométrique de la France" 1783

Sur la latitude & la longitude de l'Observatoire Royal.

Il était nécessaire, pour faire usage des distances à la méridienne & à la perpendiculaire de l'Observatoire, pour déterminer la latitude & la longitude des principales Villes du Royaume, de connaître exactement la latitude du lieu où l'on devait rapporter la position de tous les autres compris dans l'étendue du Royaume : cet élément le plus important de l'Astronomie, a été l'objet des recherches des plus célèbres Astronomes. Je rapporterai ici les résultats de leurs observations & j'exposerai les raisons qui m'ont déterminé à adopter celle que j'ai employée dans le calcul.

J.D. Cassini avait déterminé en 1672, la hauteur du pole apparente de 48° 51' 0", c'est la plus petite ; car il a remarqué des variation très-singulières, qu'on pourrait attribuer à des irrégularités dans les réfractions, causées par les feux & la fumée d'une grande Ville, au Midi de laquelle l'Observatoire et placé. M. de la Hire en était si persuadé, qu'il n'a jamais fait usage des observations de l'Étoile polaire, pour en déduire la latitude de l'observatoire, & qu'il a préféré les observations faites au Midi, où il ne pouvait soupçonner aucunes irrégularités dans les réfractions.

M. Lemonnier, en 1737, a trouvé la hauteur du pole de	48° 51' 4"
M. Maraldy, en 1733,	48 51 1
M. Legentil, en 1764,	48 51 2
Je l'ai trouvée en 1744,	48 51 1

2 Voir aussi: Boyé-Lefort, 1990, *Mesurer aussi bien la terre que le ciel* , IREM des Pays de Loire,

Les observations faite à la porte Montmartre, par M. Picard, réduites à l'Observatoire Royal, donnent la hauteur du pôle de 48° 51' 0".

Celles faites à S. Jacques de la Boucherie, [3] donnent 48° 51' 2".

Les observations du Chevalier de Louville, faites à l'hôtel Taranne 48° 50' 58".

Par les observations faites à Orléans, il a trouvé quelques secondes de plus.

Voilà donc sept déterminations, dont la plus éloignée ne diffère de la mienne que de trois secondes ; & comme il n'est guère possible de porter la précision plus loin, j'ai regardé le résultat de mes observations, qui tient un milieu entre tous les autres, comme le plus approchant du vrai, & je m'y suis arrêté, pour fixer la latitude apparente de l'Observatoire de 48° 51' 1"

Et en supposant la réfraction de 49", la hauteur vraie sera de 48° 50' 12".

Lorsque j'ai vérifié en 1755, les divisions du quart de cercle que j'avais employé, j'avais que sur l'arc de 90", il y avait à peine dix à douze secondes d'erreur, & qu'elle n'était point proportionnelle. J'en ai dressé une Table, & ayant déterminé avec le plus grand foin la hauteur solsticiale du bord supérieur du Soleil, ayant égard à la correction indiquée dans la Table, j'avais trouvé l'obliquité moyenne en 1755, 23° 28' 13".

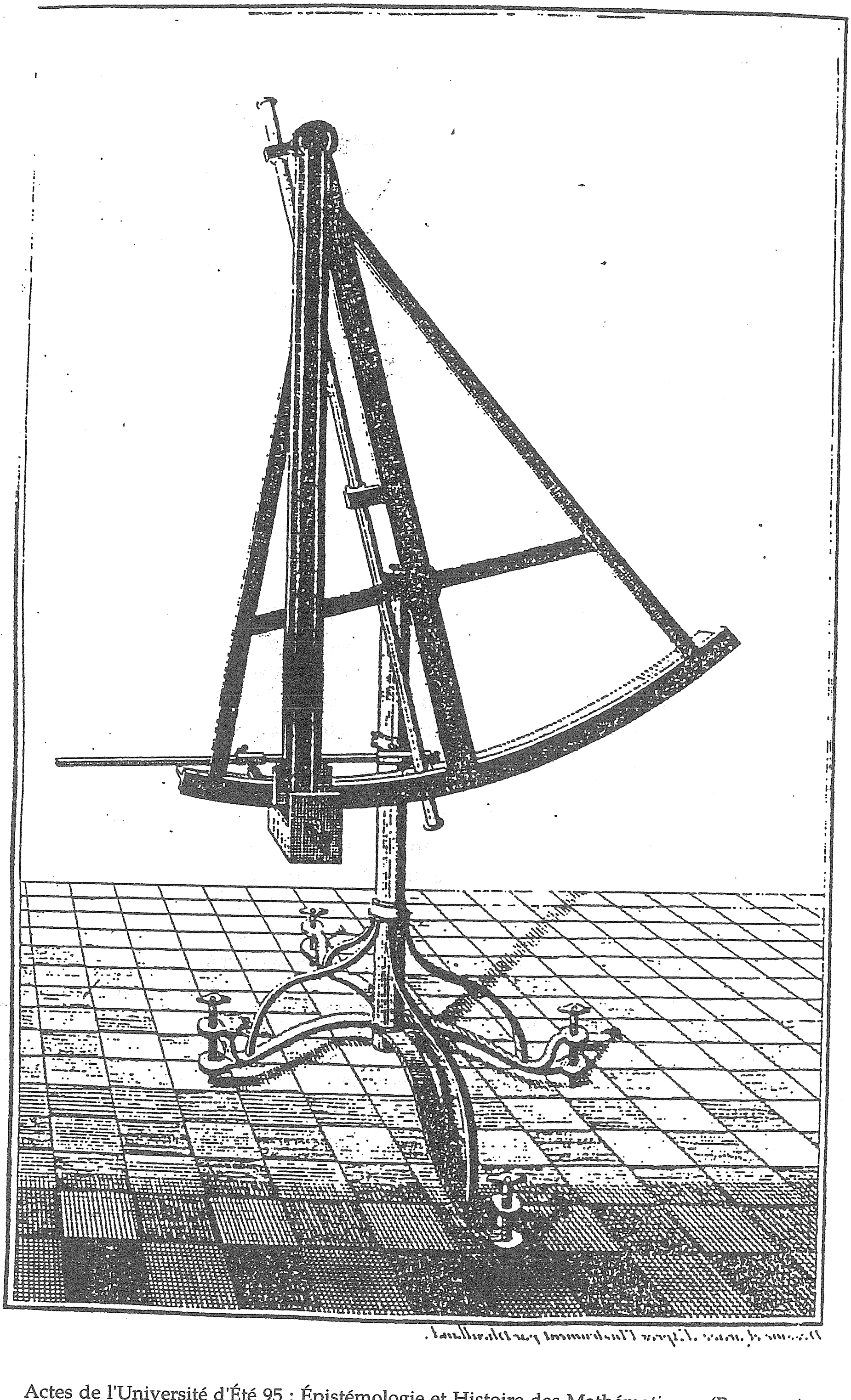
M. le Monnier ayant bien voulu me communiquer son observation, dont le résultat donnait la hauteur solsticiale, apparente de 64° 54' 9".

Tandis que l'observation de M. de la Caille donnait 64° 54' 7".

Et la mienne 64° 54' 9".

Le résultat de ces trois observations, qui ne différait que de deux secondes, était une preuve de la certitude de celles que j'avais employée pour la connaissance d'un second élément qui entre dans tous les calculs astronomiques. Il n'est pas, selon moi, de moyen plus certain pour vérifier les instruments, que de les comparer entr'eux & avec ceux qui sortent des mains des plus habiles Artistes, & je ne regarderai jamais comme une erreur certaine, une quantité qui n'excédera pas dix secondes, & qui peut être variable, parce que les instruments de telle matière qu'ils soient composés, éprouvent des altérations dont on reconnaît les effets, sans pouvoir en tenir compte.

3 Tables de la Hire.



Jean-Henri Lambert

Connu pour d'autres travaux, Lambert a publié en 1760 les résultats de ses recherches en optique: *Photometria sive de mensura*[4]. Confronté à des expériences répétées, recherchant la loi du phénomène observé, il se livre là à certaines réflexions sur la recherche des valeurs les plus probables, se référant aux travaux de Jacques Bernouilli.

Ce travail a été réalisé dans les années 1750, alors que Lambert était précepteur des enfants du Comte de Salis à Coire (Chur). Né à Mulhouse en 1728, notre auteur, autodidacte, avait trouvé cet emploi qui lui permit ensuite de voyager d'université en université avant d'être titularisé à l'Académie de Berlin en 1765.

La *Photométrie* de Lambert est remarquable en ce que cet ouvrage prend le parti d'utiliser l'expérience pour étayer les hypothèses. Tout est discuté, détaillé, commenté, quelques fois à l'excès. Ce souci entraîne l'auteur à représenter graphiquement la répartition des erreurs, ce qui est une réelle innovation.

Jean Henri LAMBERT "Photometria sive de mensura ..." 1760

295 - Puisque dans chaque expérience de photométrie, comme dans des expériences quelconques, les erreurs ne sont pas également fréquentes, on donnera une autre méthode pour déterminer la valeur moyenne à partir d'un nombre fini de celles-ci, et de manière qu'il soit très probable qu'elle s'écarte le moins de la valeur moyenne vraie. En effet, puisqu'il n'est pas possible d'obtenir une certitude absolue, on doit conduire toute chose pour obtenir la probabilité la plus grande. Il est donc utile d'exposer plus généralement cette méthode.

296 - Soit AC la valeur vraie à déterminer au moyen d'expériences, soit CB et CD les erreurs maximales de part et d'autre. Que les ordonnées QM,PN, RL, SK de la courbe BMLD représente les retours des erreurs CQ, CP, CR, CS, à savoir leurs fréquences. Nous appellerons celles-ci fréquences absolues ou vraies pour les distinguer des fréquences observées. Il est clair que celles-ci coïncident avec celles-là si l'expérience est infiniment répétée. Mais puisque cela n'arrive jamais, il faut découvrir ce qui demeure pour un nombre fini d'expériences.

297 - D'abord il est assurément facile de voir que si les fréquences observées pour chaque erreur sont proportionnelles aux fréquences vraies telles que PN, QM, SK, etc... la valeur vraie AC s'en dégage de toute façon dans ce cas. En effet, quand les fréquences observées divisées par les vrais ont même quotients, la courbe obtenue au-dessus de l'axe ACD est en-dessous de la courbe BMLD. Comme ceci arrive très rarement, il n'est pas possible de se servir de ce moyen.

4 La traduction de *Photometria sive de mensura* est en cours de parution, traduit et commenté par Mme Boyé et Ms Boyé, Coutil, Saillard.

298 - Par ailleurs, il est non moins évident qu'à cause des distances données des points P, Q, R, S les uns par rapport aux autres, si l'on donne la distance CQ d'un seul des points Q à partir du point C, c'est-à-dire l'erreur vraie, on donne aussi les erreurs vraies de tous les autres points CP, CR, CS. Ainsi le problème proposé se ramène-t-il à l'unique erreur vraie recherchée.

299 - Supposons que les quantités AP, AQ, AR, AS ont été observées n, m, l, k fois, on recherche tous les cas possibles pour lesquels ceci peut se produire. Pour trouver cela selon la théorie des combinaisons et des permutations :

1) Si le nombre des observations faites est N, et que chacune d'entre-elles se produit une fois, le nombre des cas possibles est le même que le nombre de permutations, c'est-à-dire $1 \times 2 \times 3 \times 4 \times 5 \times \ldots \times N$.

2) Si une observation arrive plusieurs fois, par exemple p fois, le nombre de cas possibles sera déterminé et sera $\frac{1x2x3x4x5x...xN}{1x2x3x4x5x...xp}$

Cette conclusion s'obtient lorsque toutes les observations sont également probables.

3) Si une ou plusieurs observations sont plus probables que d'autres et que cependant chacune arrive seulement une fois, il faudra multiplier le nombre de cas $1 \times 2 \times 3 \times \ldots \times N$ possibles par la fréquence vraie.

4) Enfin si les possibilités sont diverses et si certaines observations se produisent plusieurs fois, le nombre trouvé par la proposition 2) doit être multiplié par les fréquences vraies élevées à la puissance qui correspond à la fréquence observée.

300 - Ceci s'applique facilement à la question posée :

- le nombre des observations est $m + n + l + k$
- les fréquences observées sont m, n, l, k donc si l'on appelle N le nombre de cas possibles, les fréquences observées sont PN, QM, RL, SK, nous aurons

$$N = \frac{1x2x3x...x\ (m+n+l+k)\ x\ (PN^n)\ x\ (QM^m)\ x\ (RL^l)\ x\ (SK^k)}{(1x2x3x...m)\ x\ (1x2x3x...n)\ x\ (1x2x3x...l)\ x\ (1x2x3x...k)}$$

Comme les fréquences observées n, m, l, k, demeurent constantes, le facteur qui vient des permutations reste constant et nous aurons

$$N \simeq PN^n\, QM^m\, RL^l\, SK^k$$

Ainsi le nombre de cas possibles sera comme le produit des fréquences vraies élevées à leur puissances qui sont égales aux fréquences observées.

301 - Si l'on donnait l'erreur vraie CQ, on donnerait du même coup CP, CR, CS et aussi les fréquences vraies PN, QM, RL, SK puisque ici la courbe BMD est supposée donnée. De là serait donné aussi le nombre des cas possibles. Mais puisque nous sommes priés de nous contenter ici d'une probabilité, il faut s'engager dans une autre voie.

302 - Considérons l'erreur CQ comme variable ; si elle s'accroit ou diminue, les erreurs CP, CR, CS seront aussi augmentées ou diminuées puisque la distance respective des points P, Q, R, S est constante. C'est pourquoi seront changés aussi

les ordonnées, i.e. les fréquences vraies PN, QM, RL, SK et ainsi le nombre de cas possibles.

303 - Puisque en l'espèce le cas le plus probable est celui qui est le plus fréquent, il est logique que $PN^n \times QM^m \times RL^l \times SK^k$ soit égal au nombre le plus grand. Cette quantité est fonction du nombre de cas possibles et le cas le plus probable sera celui pour lequel la quantité sera la plus grande de toutes.

304 - Nous aurons donc
$n \,\mathrm{Log}\, PN + m \,\mathrm{Log}\, QM + l \,\mathrm{Log}\, RL + k \,\mathrm{Log}\, sk$ maximum
et en différenciant

$$n\frac{dPN}{PN} + m\frac{dQM}{QM} + l\frac{dRL}{RL} + k\frac{dSK}{SK} = 0$$

Considérons les droites tangentes aux points N, M, L et K et leurs sous-tangentes la probabilité sera la plus grande lorsque nous aurons

$$\frac{n}{\nu} + \frac{m}{\mu} + \frac{l}{\lambda} + \frac{k}{\Delta} = 0$$

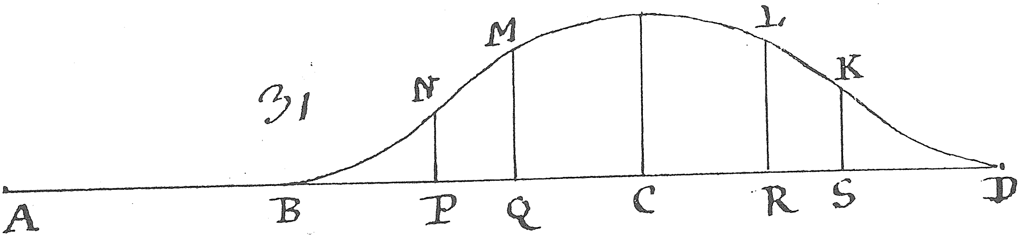
dans cette formule les sous-tangentes sont positives ou négatives de même que les erreurs auxquelles elles correspondent sont négatives ou positives.

305 - Si la courbe BMLD est identique de part et d'autre du centre C, la valeur moyenne AC trouvée par cette méthode coïncide assez souvent avec la valeur de la moyenne arithmétique, dont nous avons parlé dans les pages précédentes. Ainsi, par exemple, si l'on fait les deux observations AQ et AR seulement, nous aurons $m = l = 1$ et ainsi

$$\frac{1}{\lambda} + \frac{1}{\mu} = 0 \text{ ou } \mu = -\lambda$$

ce qui ne peut se produire que si le point C est à égal distance de P et Q. Dans ce cas AC est la moyenne arithmétique des deux observations.

306 - De façon similaire, ce résultat sera évident lorsque reprenant les quatre observations les distances PQ et RS sont égales. Le point C sera alors à égale distance des points Q, R ou R, S. Le même résultat sera obtenu avec trois observations dont les points extrêmes sont à égale distance du point central. Le point central coïncide alors avec le point C. D'ailleurs, il est évident que cette méthode est plus élégante que pratique puisqu'elle conduit à des calculs plus longs. A partir d'ici nous ne nous attarderons pas à un examen plus approfondi d'autant plus que la quantité trouvée par cette méthode ne diffère pratiquement pas de la moyenne arithmétique.



Adrien Marie Legendre

Adrien Marie Legendre est né en 1752 à Toulouse. En 1787 il est chargé d'opérations de géodésie, ce que la Convention en 1793 lui proposera à nouveau. Ces travaux peut-être, mais aussi ceux qu'il effectua en astronomie l'ont conduit à exposer le principe des "moindres carrés"[5]. Legendre traversera la révolution sans encombre, et travaillera sur la théorie des nombres et sur la géométrie, avant de mourir en 1833.

La méthode des moindres carrés figure en appendice des *Nouvelles méthodes pour la détermination des orbites des comètes,* paru en 1805.

Adrien Marie LEGENDRE "Nouvelles Méthodes ..."

Dans la plupart des questions où il s'agit de tirer des mesures données par l'observation les résultats les plus exacts qu'elles peuvent offrir; on est presque toujours conduit à un système de la forme :

$$E = a + b\,x + c\,y + f\,z + \text{etc.}$$

dans lesquelles a, b, c, f, etc. sont des coefficients connus qui varient d'une équation à l'autre, et x, y, z, etc. sont des inconnues qu'il faut déterminer par la condition que la valeur de E se réduise, pour chaque équation, à une quantité nulle ou très petite. Si l'on a autant d'équations que d'inconnues x, y, z, etc., il n'y a aucune difficulté pour la détermination de ces inconnues, et on peut rendre les erreurs E absolument nulles. Mais le plus souvent le nombre des équations est supérieur à celui des inconnues, et il est impossible d'anéantir toutes les erreurs.

Dans cette circonstance, qui est celle de la plupart des problèmes physiques et astronomiques, où l'on cherche à déterminer quelques éléments importants, il entre nécessairement de l'arbitraire dans la distribution des erreurs, et on ne doit pas s'attendre que toutes les hypothèses conduiront exactement aux mêmes résultats : mais il faut surtout faire en sorte que les erreurs extrêmes, sans avoir égard à leurs signes, soient renfermées dans les limites les plus étroites qu'il est possible.

De tous les principes qu'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.

---

5 Voir aussi: Chabert et Alii, 1994, *Histoire d'algorithmes,* Paris, Belin, p.324 et suivantes.

La somme des carrés des erreurs $E^2 + E'^2 + E''^2$ etc. étant

$$+ (a + bx + cy + fz + \text{etc.})^2$$
$$+ (a' + b'x + c'y + f'z + \text{etc.})^2$$
$$+ (a'' + b''x + c''y + f''z + \text{etc.})^2$$
$$+ \text{etc.} ;$$

si l'on cherche son *minimum* en faisant varier x seule, on aura l'équation

$$0 = \int ab + x \int b^2 + y \int bc + z \int bf + \text{etc} .$$

dans laquelle par $\int ab$ on entend la somme des produits semblables $ab + a'b' + a''b'' + \text{etc.}$, par $\int b^2$ la somme des carrés des coefficients de x, savoir, $b^2 + b'^2 + b''^2 + \text{etc.}$ ainsi de suite.

Le *minimum* par rapport à y donnera semblablement

$$0 = \int ac + x \int bc + y \int c^2 + z \int fc + \text{etc} .$$

et le *minimum* par rapport à z,

$$0 = \int af + x \int bf + y \int cf + z \int f^2 + \text{etc} .$$

où l'on voit que les mêmes coefficients $\int bc$, $\int bf$, etc. sont communs à deux équations, ce qui contribue à faciliter le calcul.

En général, pour former l'équation du minimum par rapport à l'une des inconnues, il faut multiplier tous les termes de chaque équation proposée par le coefficient de l'inconnue dans cette équation, pris avec son signe, et faire une somme de tous ces produits.
On obtiendra de cette manière autant d'équations du minimum qu'il y a d'inconnues, et il faudra résoudre ces équations par les méthodes ordinaires.

[...]

La règle par laquelle on prend le milieu entre les résultats de diverses observations (pour un seul élément), n'est qu'une conséquence très simple de notre méthode générale, que nous appellerons *méthode des moindres carrés*. En effet, si l'expérience a donné diverses valeurs a', a", a"', etc. pour une certaine quantité x, la somme des carrés des erreurs sera $(a' - x)^2 + (a'' - x)^2 + (a''' - x)^2 + \text{etc.}$; et en égalant cette somme à un *minimum*, on a

$$0 = (a' - x) + (a'' - x) + (a''' - x) + \text{etc.}$$

d'où résulte $x = \dfrac{a + a' + a'' + \ldots + \text{etc.}}{n}$, *n* étant le nombre des observations.

Carl Friedrich Gauss

Il y eut querelle d'antériorité entre Legendre et Gauss au sujet de la méthode des moindres carrés. Si Legendre l'avait exposée dans ses *Nouvelles méthodes...* en 1805, Gauss l'aurait conçue avant, pour ne la publier qu'en 1809 dans sa *Theoria motis corporum coelestium.*

Gauss est né en 1777 à Göttingen et sa précocité lui permit de poursuivre ses études aux frais d'une personnalité de sa région. Professeur d'astronomie à Göttingen en 1807, il fit paraître la *Theoria motis...* en 1809, où il expose non seulement la méthode des moindres carrés, mais il explicite surtout la loi de probabilité des erreurs.

Plus tard Gauss s'occupera de géodésie et de topographie dans la province du Hanovre, ce qui lui donnera l'occasion de mettre cette loi en application.

Gauss décédera en 1855.

Karl Friedrich GAUSS "Theoria motis corporum coelest ium" 1809

EXPOSITION DE LA MÉTHODE DES MOINDRES CARRÉS

(Extrait du Theoria Motus Corporums coelestium.)

1

..... Abordons maintenant une recherche beaucoup plus générale et des plus fécondes dans toute application du calcul aux phénomènes naturels. Soient V, V', V", etc., μ fonctions des v inconnues p, q, r, s, etc., et supposons que des observations directes aient donné, pour ces fonctions, les valeurs

$$V = M, \quad V' = M', \quad V'' = M'', \dots$$

En général, le calcul de ces inconnues constituera un problème indéterminé, déterminé ou plus que déterminé, suivant que l'on aura

$$\mu < v, \ \mu = v, \ \mu > v$$ [6]

Nous ne nous occuperons ici que du dernier cas, dans lequel évidemment il ne serait possible d'obtenir une représentation exacte de toutes les observations, que si ces observations n'étaient affectées d'aucune erreur. Mais comme cela n'a jamais lieu dans la nature, on devra regarder comme possible tout système de valeurs des inconnues p, q, r, s, etc., desquelles résultent, pour les fonctions V – M, V' –M', V" – M", des valeurs qui ne surpassent pas les limites des erreurs que l'on peut commettre dans les observations, mais on ne doit pas regarder tous ces systèmes possibles comme jouissant du même degré de probabilité.

Supposons d'abord, dans toutes les observations, un état de choses tel, qu'il n'y ait pas lieu de regarder l'une d'elles comme également probables. La probabilité qu'une erreur Δ soit commise dans l'une des observations sera une fonction de Δ, que

6 Si, dans ce troisième cas, μ + 1 - v des fonctions V, V', V", etc., pouvaient étre regardées comme des fonctions de toutes les autres, le problème deviendrait plus que déterminé relativement à ces fonctions, mais indéterminé relativement à p, q, r, s, etc. On ne pourrait pas en déduire les valeurs de ces dernières, même si les valeurs des fonctions V, V', V", etc., étaient d'une exactitude absolue : mais nous excluons ce cas particulier de nos recherches.

nous nommerons $\varphi(\Delta)$. Quoique cette fonction ne puisse être assignée d'une manière précise, on peut du moins affirmer qu'elle doit devenir maximum pour $\Delta = 0$, avoir dans la plupart des cas la même valeur pour des valeurs de Δ égales et de signes contraires, et, enfin, s'évanouir quand on donne à Δ une valeur égale ou supérieure à l'erreur maximum ; $\varphi(\Delta)$ doit donc, à proprement parler, être rapportée à la classe des fonctions discontinues, et, si nous nous permettons, pour la facilité du calcul, d'y substituer une fonction analytique, il faudra que cette dernière soit choisie de telle sorte qu'elle tende rapidement vers 0 à partir de deux valeurs de Δ, l'une supérieure, l'autre inférieure à 0, et qu'en dehors de ces deux limites on puisse la considérer comme nulle. Or la probabilité que l'erreur soit comprise entre Δ et une quantité $\Delta + d\Delta$ qui en diffère infiniment peu, sera exprimée par $\varphi(\Delta)\,.\,d\Delta$, et, par suite, la probabilité que l'erreur est comprise entre D et D', par

$$\int_{D}^{D'} \varphi(\Delta)\, d\Delta.$$

Cette intégrale, prise depuis la plus grande valeur négative de Δ jusqu'à sa plus grande valeur positive, ou plus généralement depuis $\Delta = -\infty$ jusqu'à $\Delta = \infty$, devra nécessairement être égale à 1. On aura donc

$$\int_{-\infty}^{\infty} \varphi(\Delta)\, d\Delta = I.$$

Supposons donc qu'on ait un système déterminé de valeurs des quantités p, q, r, s, etc. : la probabilité que l'observation donnera pour V la valeur M, sera exprimée par f (M – V), après qu'on aura substitué dans V les valeurs de p, q, r, s, etc. ; de même $\varphi(M' - V')$, $\varphi(M'' - V'')$, etc., exprimeront les probabilités pour que les observations donnent aux fonctions V', V", etc., les valeurs M', M", etc. C'est pourquoi, tant qu'on pourra considérer toutes les observations comme des événements indépendants les uns des autres, le produit

$$\varphi(M - V)\, \varphi(M' - V')\, \varphi(M'' - V'') \ldots = \Omega$$

exprimera la probabilité que toutes ces valeurs résulteront en même temps des observations.

2

De même qu'en se donnant des valeurs quelconques des inconnues, il en résulte, avant toute observation, une probabilité déterminée pour un système de valeurs des fonctions V, V', V", etc., de même, après que l'observation aura donné pour ces fonctions des valeurs déterminées, il en résultera pour chaque système de valeurs des inconnues qui en découleront, une probabilité déterminée : car il est clair qu'on devra considérer comme les plus probables les systèmes qui donnent à l'événement observé la plus grande probabilité. L'appréciation de cette probabilité peut s'obtenir par le théorème suivant :

Si, en adoptant une certaine hypothèse II, la probabilité d'un événement déterminé E est h, mais qu'en adoptant une autre hypothèse H', exclusive de la première et ayant à priori la même probabilité, la probabilité du même événement soit h' : je dis que lorsque l'événement E aura eu lieu la probabilité que H soit la vraie hypothèse sera à la probabilité que H' soit la vraie hypothèse comme h est à h'.

Pour le démontrer et afin de distinguer toutes les circonstances d'où peut dépendre, soit que l'hypothèse H ou H', ou toute autre ait lieu, l'arrivée d'un événement E ou d'un autre événement, formons un système des cas différents qui peuvent se présenter et que nous regarderons comme également probables à priori (c'est-à-dire tant qu'il y a doute si c'est l'événement E ou un autre qui aura lieu). Ces cas peuvent être ainsi distribués :

nombre de cas	HYPOTHÈSE propre à ces cas	ÉVÉNEMENT qui doit en résulter.
m	H	E
n	H	Différent de E
m'	H'	E
n'	H'	Différent de E
m"	Différente de II et de II'	E
n"	Différente de H et de H'	Différent de E

On aura d'après cela

$$h = \frac{m}{m+n}, \; h' = \frac{m'}{m'+n'}$$

Or avant l'arrivée de l'événement, la probabilité de l'hypothèse H était

$$\frac{m+n}{m+n+m'+n'+m''+n''}$$

Après l'événement qui exclut n + n' + n" cas, parmi ceux qui sont possibles, cette probabilité sera

$$\frac{m}{m+m'+m''}$$

De même les probabilités de l'hypothèse H' avant et après l'événement sont respectivement

$$\frac{m'+n'}{m+n+m'\,n'+m''+n''} \text{ et}$$

$$\frac{m'}{m+m'+m''}$$

mais, comme on a supposé que les hypothèses H et H' avaient avant l'événement la même probabilité, on aura

$$m + n = m' + n'$$

d'où résulte immédiatement la vérité du théorème.

Si maintenant on suppose qu'on n'a, pour déterminer les inconnues, que les observations

$$V = M,\; V' = M',\; V'' = M'', \ldots ,$$

et que tous les systèmes de valeurs des inconnues étaient également probables avant ces observations, il est visible que la probabilité d'un certain système, après ces observations, sera proportionnelle à Ω. C'est-à-dire que $\lambda\Omega dpdqdr$ ... exprimera la probabilité que les valeurs des inconnues soient respectivement comprises dans les limites infiniment voisines p et p + dp, q et q + dq, r et r + dr, ..., λ représentant une quantité indépendante de p, q, r, s, etc. ; et l'on aura évidemment

$$\frac{I}{\lambda} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \ldots \Omega \, dp \, dq \, dr \ldots$$

De là résulte naturellement que le système le plus probable des valeurs de p, q, r, etc., correspondra au maximum de Ω, et se tirera des v équations

$$\frac{d\Omega}{dp} = 0, \; \frac{d\Omega}{dq} = 0, \; \frac{d\Omega}{dr} = 0$$

si l'on pose

$$V - M = v,\; V' - M' = v',\; V'' - M'' = v'', \ldots \text{ et}$$

$$\frac{d.\varphi(\Delta)}{\varphi(\Delta).d\Delta} = \varphi'(\Delta),$$

ces équations prendront la forme suivante :

$$\frac{dv}{dp}\varphi'(v) + \frac{dv'}{dp}\varphi'(v') + \frac{dv''}{dp}\varphi'(v'') + \ldots = 0$$

$$\frac{dv}{dq}\varphi'(v) + \frac{dv'}{dq}\varphi'(v') + \frac{dv''}{dq}\varphi'(v'') + \ldots = 0$$

De là résulte qu'on pourra obtenir par l'élimination une solution pleinement déterminée du problème, dès que la nature de la fonction φ' sera connue. Mais comme cette fonction ne peut être définie à priori, abordons la question à un autre point de vue et cherchons une fonction acceptée tacitement comme base, en vertu d'un principe simple et généralement admis. Or on a coutume de regarder comme un axiome l'hypothèse que si une grande quantité a été obtenue par plusieurs observations immédiates, faites avec le même soin dans des circonstances semblables, la moyenne arithmétique des valeurs observées sera la valeur la plus probable de cette quantité, sinon en toute rigueur, du moins avec une grande approximation, de telle sorte que le plus sûr soit toujours de s'y arrêter. Si donc l'on pose

$$V = V' = V'' \ldots = p,$$

et

$$p = \frac{M + M' + M'' + \ldots}{\mu},$$

on devra avoir en général

$$\varphi'(M - p) + \varphi'(M' - p) + \varphi'(M'' - p) + \ldots = 0$$

pour toute valeur entière et positive de μ. Faisant ensuite

$$M' = M'' \ldots = M - \mu N,$$

on aura généralement

$$\varphi'\,[(\mu - 1)\,N] = (1 - \mu)\,\varphi'\,(-N),$$

d'où l'on tire facilement que $\frac{\varphi'(\Delta)}{\Delta}$ doit être en général une constante k. On aura donc

$$\log \varphi(\Delta) = \frac{1}{2}\,k\Delta^2 + \text{const.} = \frac{1}{2}\,\lambda\,\Delta^2 + \log x,$$

d'où

$$\varphi(\Delta) = xc^{\frac{1}{2}\lambda\Delta^2}$$

Or on voit facilement que la constante k doit être négative, pour que Ω puisse devenir maximum : posons donc

$$\frac{1}{2}\,k = -\,h^2$$

et comme, d'après un élégant théorème de Laplace, on a

$$\int_{-\infty}^{+\infty} e^{-h^2\Delta^2}\,d\Delta = \frac{\sqrt{\pi}}{h}$$

notre fonction deviendra

$$\varphi(\Delta) = \frac{h}{\sqrt{\pi}}\,e^{-h^2\Delta^2}.$$

Pierre Simon Laplace

Né en Normandie en 1749, Laplace est introduit à l'Académie Royale des Sciences dès 1773, malgré sa jeunesse. Il participa aux travaux de la Commission des Poids et Mesures. Il sera honoré par le régime impérial, puis, rallié à la royauté après 1815, Laplace décédera en 1827.

En ce qui concerne notre propos, Laplace publie en 1814 "la Théorie analytique des probabilités". Sa lecture n'est pas facile, l'auteur passant souvent très vite sur des notions ou résultats pas nécessairement évidents. Dans le texte choisi, par exemple, il appelle implicitement le calcul

$$\int_{\infty}^{+\infty} e^{-x^2}\,dx = \sqrt{\pi}\,,$$

théorème que, comme il est possible de le lire dans le texte précédent, Gauss attribue également à Laplace.

Pierre Simon LAPLACE "Théorie analytique des probabilités" 1814

23. Nous venons de rechercher le résultat moyen que des observations nombreuses et non faites encore, doivent indiquer avec le plus d'avantage, et la loi de probabilité des erreurs de ce résultat. Considérons présentement le résultat moyen des observations déjà faites, et dont on connaît les écarts respectifs. Pour cela, concevons un nombre s d'observations du même genre, c'est-à-dire, telles que la loi des erreurs soit la même pour toutes. Nommons A le résultat de la première ; A + q, celui de la seconde : $A + q^{(1)}$, celui de la troisième, et ainsi de suite ; q, $q^{(1)}$, $q^{(2)}$, etc. étant des quantités positives et croissantes, ce que l'on peut toujours obtenir par une disposition convenable des observations. Désignons encore par φ(z), la probabilité de l'erreur z pour chaque observation, et supposons que A + x soit le vrai résultat. L'erreur de la première observation est alors – x ; q – x, $q^{(1)} - x$, etc. sont les erreurs de la seconde, de la troisième, etc. La probabilité de l'existence simultanée de toutes ces erreurs, est le produit de leurs probabilités respectives ; elle est donc

$$\varphi(-x)\,.\,\varphi(q-x)\,.\,\varphi(q^{(1)}-x)\,.\ \text{etc.}$$

Maintenant, x étant susceptible d'une infinité de valeurs ; en les considérant comme autant de causes de l'événement observé, la probabilité de chacune d'elles sera, par le n° 1,

$$\frac{dx.\varphi(-x).\ \varphi(q-x).\ \varphi(q^{(1)}-x).\ \text{etc.}}{\int dx.\varphi(-x).\varphi(q-x).\ \varphi(q^{(1)}-x).\text{etc.}}\,,$$

l'intégrale du dénominateur étant prise pour toutes les valeurs dont x est susceptible. Nommons $\frac{1}{H}$ ce dénominateur. Cela est posé, imaginons une courbe dont x soit l'abscisse, et dont l'ordonnée y soit

$$H \,.\, \varphi(-x) \,.\, \varphi(q-x) \,.\, \varphi(q^{(1)}-x) \,.\, \text{etc.} ;$$

cette courbe sera celle des probabilités des valeurs de x. La valeur qu'il faut choisir pour résultat moyen, est celle qui rend l'erreur moyenne à craindre, un *minimum*. Toute erreur, soit positive, soit négative, devant être considérée comme un désavantage ou une perte réelle au jeu ; on a le désavantage moyen, en prenant la somme des produits de chaque désavantage, par sa probabilité ; la valeur moyenne de l'erreur à craindre, est donc la somme des produits de chaque erreur, abstraction faite du signe, par sa probabilité. Déterminons l'abscisse qu'il faut choisir pour que cette somme soit un *minimum*. Pour cela, donnons aux abscisses, pour origine, la première extrémité de la courbe précédente, et nommons x' et y' les coordonnées de la courbe, à partir de cette origine. Soit l la valeur qu'il faut choisir. Il est clair que si le vrai résultat était x', l'erreur du résultat l serait, abstraction faite du signe, l – x', tant que x' serait moindre que l ; or y' est la probabilité que x' est le résultat vrai ; la somme des erreurs à craindre, abstraction faite du signe, multipliées par leur probabilité, est donc pour toutes les valeurs de x', moindres que l, $\int (l - x') \,.\, y'dx'$, l'intégrale étant prise depuis x' = 0 jusqu'à x' = l.

On verra de la même manière, que pour les valeurs de x' supérieures à l, la somme des erreurs à craindre, multipliées par leur probabilité, est $\int (x' - l) .\, y'dx'$, l'intégrale étant prise x' = l jusqu'à l'abscisse x' correspondante à la dernière extrémité de la courbe ; la somme entière des erreurs à craindre, abstraction faite du signe, et multipliées par leurs probabilités respectives, est donc

$$\int (l - x') .\, y'dx' + \int (x' - l) \,.\, y'\, dx'.$$

La différentielle de cette fonction, prise par rapport à l, est

$$dl.\int y'dx' - dl.\int y'dx' ;$$

car on a la différentielle de $\int (l - x') \,.\, y'dx'$, en différentiant d'abord la valeur de l sous le signe f, et en ajoutant à cette différentielle, l'accroissement qui résulte de la variation de la limite de l'intégrale, limite qui se change en l + dl. Cet accroissement est égal à l'élément $(l - x').y'dx'$, à la limite où x' = l ; il est donc nul, et $dl.\int y'dx'$ est la différentielle de l'intégrale $\int (l - x').y'\, dx'$. On verra de la même manière, que $- dl.\int y'dx'$ est la différentielle de l'intégrale $\int (x' - l).\, y'\, dx'$. La somme de ces différentielles est nulle relativement à l'abscisse l, pour laquelle l'erreur moyenne à craindre est un *minimum* ; on a donc, relativement à cette abscisse,

$$\int y'dx' = \int y'dx'$$

la première intégrale étant prise depuis x' = 0 jusqu'à x' = l, et la seconde étant prise depuis x' = l jusqu'à la valeur extrême de x'.
Il suit de là que l'abscisse qui rend l'erreur moyenne à craindre, un *minimum*, est celle dont l'ordonnée divise l'aire de la courbe en deux parties égales. Ce point jouit encore de la propriété d'être celui en deçà duquel il est aussi probable que le vrai résultat tombe, qu'au-delà ; et par cette raison, il peut encore être nommé *milieu de probabilité*. Des géomètres célèbres ont pris pour le milieu qu'il faut choisir, celui qui rend le résultat observé, le plus probable, et par conséquent l'abscisse qui répond à la plus grande ordonnée de la courbe ; mais le milieu que nous adoptons, est évidemment indiqué par la théorie des probabilités.
Si l'on met $\varphi(x)$ sous la forme d'exponentielle, et qu'on le désigne par $c^{-\psi(x^2)}$, afin qu'il puisse également convenir aux erreurs positives et négatives ; on aura

$$y = H.c^{-\psi(x^2) - \psi(x-q)^2 - \psi(x-q^{(1)})^2 - \text{etc.}} \quad (1)$$

Si l'on fait $x = a + z$, et que l'on développe l'exposant de c par rapport aux puissances de z, y prendra cette forme,

$$y = H.c^{-M - 2Nz - Pz^2 - Qz^3 - \text{etc.}}$$

expression dans laquelle on a

$$M = \psi(a^2) + \psi(a-q)^2 + \psi(x-q^{(1)})^2 + \text{etc.},$$
$$N = a\,\psi'(a^2) + (a-q)\,.\,\psi'(a-q)^2 + (a-q^{(1)})\,.\,\psi'(a-q^{(1)})^2 + \text{etc.},$$
$$P = \psi'(a^2) + \psi'(a-q)^2 + \psi'(a-q^{(1)})^2 + \text{etc. } 2a^2\,.\,\psi(a^2)$$
$$+2(a-q)^2.\,\psi''(a-q)^2 + 2(a-q^{(1)})^2\,.\,\psi''(a-q^{(1)})^2 + \text{etc.},$$

etc.,

$\psi'(t)$ étant le coefficient de dt dans la différentielle de $\psi(t)$, $\psi''(t)$ étant le coefficient de dt dans la différentielle de $\psi'(t)$, et ainsi de suite.
Supposons le nombre s des observations, très-grand, et déterminons a par l'équation N = 0 que donne la condition du *maximum* de y ; alors on a

$$y = H.c^{-M - Pz^2 - Qz^3 - \text{etc.}}.$$

M, P, Q, etc. sont de l'ordre s ; or, si z est très petit de l'ordre $\frac{1}{\sqrt{s}}$,Qz^3 devient de l'ordre $\frac{1}{\sqrt{s}}$, et l'exponentielle $c^{-Qz^3 - \text{etc.}}$ peut se réduire à l'unité. Ainsi dans l'intervalle depuis z = 0 jusqu'à $z = \frac{r}{\sqrt{s}}$,on peut supposer

$$y = H.\ c^{-M - Pz^2}.$$

Au-delà, et lorsque z est d el'ordre $s^{-\frac{m}{2}}$ m étant plus petit que l'unité, Pz^2 devient de l'ordre s^{1-m} ; par conséquent c^{-Pz^2} devient ainsi que y, insensible ; ensorte que l'on peut, dans toute l'étendue de la courbe, supposer

$$y = H.\ c^{-M - Pz^2}.$$

La valeur de a donnée par l'équation N = 0, ou

$$0 = a.\psi'(a^2) + (a - q).\ \psi'(a - q)^2 + (a - q^{(1)}).\psi'(a - q^{(1)})^2 + \text{etc.},$$

est alors l'abscisse x correspondante à l'ordonnée qui divise l'aire de la courbe en parties égales. La condition que l'aire entière de la courbe doit représenter la certitude ou l'unité, donne

$$\frac{1}{H} = \int dz.c^{-M-Pz^2}$$

l'intégrale étant prise depuis $z = -\infty$ jusqu'à $z = \infty$, ce qui donne

$$H = \frac{cM\ .\ \sqrt{P}}{\sqrt{\pi}}.$$

L'erreur moyenne à craindre en plus ou en moins, en prenant a pur résultat moyen des observations, est ? ∫ zydz, l'intégrale étant prise depuis z nul jusqu'à z infini, ce qui donne pour cette erreur

$$\approx \frac{1}{2\ .\ \sqrt{\pi.\ P}}.$$

Mais l'ignorance entière où l'on est de la loi $c^{-\psi(x^2)}$ des erreurs de chaque observation, ne permet pas de former l'équation

$$0 = a.\psi'(a^2) + (a - q).\ \psi'(a - q)^2 + \text{etc.},$$

Ainsi la connaissance des valeurs de q, q(1), etc., ne donnant *à posteriori*, aucune lumière sur le résultat moyen *a* des observations ; il faut s'en tenir au résultat le plus avantageux déterminé *à priori*, et que l'on a vu être celui que fournit la méthode des moindres carrés des erreurs.

Cherchons la fonction $\psi(x^2)$ qui donne constamment la règle des milieux arithmétiques, admise par les observateurs. Pour cela, concevons que sur les *s* observations, les *i* premières coïncident, ainsi que les *s – i* dernières. L'équation N = 0 devient alors

$$0 = i\ .\ a.\psi'(a^2) + (s - i)\ .\ (a - q)\ .\ \psi'(a - q)^2$$

La règle des milieux arithmétiques donne

$$a = \frac{s - i}{s}.\ q\ ;$$

l'équation précédente devient ainsi

$$\psi'\ [(\frac{s\text{-}i}{s})^2.\ q^2] = \psi'(\frac{i^2}{s^2}.\ q^2).$$

Cette équation devant avoir lieu quels que soient $\frac{i}{s}$ et q, il est nécessaire que $\psi(t)$ soit indépendant de t, ce qui donne

$$\psi'(t) = k,$$

k étant une constante. En intégrant, on a

$$\psi(t) = kt - L,$$

L étant une constante arbitraire ; partant,

$$c^{-\psi(x^2)} = c^{L-kx^2}$$

Telle est donc la fonction qui peut seule, donner généralement la règle des milieux arithmétiques. La constante L doit être déterminée de manière que l'intégrale $\int dx.c^{L-kx^2}$, prise depuis $x = -\infty$ jusqu'à $x = \infty$, soit égale à l'unité ; car il est certain que l'erreur x d'une observation doit tomber dans ces limites ; on a donc

$$c^{L} = \sqrt{\frac{k}{x}}\ ;$$

par conséquent la probabilité de l'erreur x est $\sqrt{\frac{k}{x}}.c^{-kx2}$

En guise de conclusion

Le traité de Laplace sera longtemps la référence en théorie des probabilités. Sans doute, à la fin du XIX° siècle, Tchebichev, puis Markov lui donneront toute la rigueur voulue, avant que d'autres travaux n'en élargissent considérablement la conception. Mais, en ce qui concerne la traitement statistique des erreurs, les résultats de Laplace, de Gauss, voire de Legendre, sont d'utilisation quotidienne. De multiples ouvrages d'enseignement, tant en physique qu'en technologie, présentent en préliminaire la "courbe de distribution normale", ses propriétés et son utilisation. Elle est en effet mise en oeuvre en permanence et doit faire partie du bagage théorique de tout "technicien supérieur".