

## Legendre et la méthode des moindres carrés

La méthode du maximum de vraisemblance et celle des moindres carrés sont les outils de la théorie des erreurs ou de l'estimation, utilisés tous les jours dans toutes les sciences d'observation.

La théorie des erreurs a été développée pour résoudre trois problèmes : combiner les erreurs pour choisir une valeur "juste milieu", trouver la loi de densité de probabilité<sup>1</sup> des erreurs, choisir une démarche pour déterminer des quantités dans un système d'équations surdéterminé<sup>2</sup>.

L'estimation du maximum de vraisemblance est une méthode statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon de mesures donné.

Par ailleurs, la recherche de la valeur la plus probable d'une quantité observée par diverses mesures donne le résultat suivant : la valeur la plus probable est telle que la somme des carrés des différences entre les observations et cette valeur est minimum.

@@@@@@@

La première publication de la méthode des moindres carrés (destinée à déterminer des quantités dans un système d'équations surdéterminé) est due à Legendre qui l'a donnée en annexe d'un ouvrage intitulé *Nouvelles méthodes pour la détermination des orbites des comètes* (1805)<sup>3</sup>.

### Querelle entre Legendre et Gauss sur la paternité de la méthode

Gauss, dans son traité *Theoria motus corporum coelestium*<sup>4</sup> publié en 1809, indique qu'il utilisait cette technique depuis 1795. Il écrit :

---

1. Aujourd'hui, la loi de probabilité continue d'une variable aléatoire  $X$  est représentée par une densité de probabilité  $f$ , la fonction  $f$  étant positive, intégrable, d'intégrale égale à 1, telle que :  $P(X = x) = f(x) dx$  est la probabilité que la variable prenne ses valeurs dans un intervalle élémentaire  $(x, x + dx)$ . La primitive  $F$  de la densité de probabilité  $f$  est appelée fonction de répartition.

2. Dans ce cas, on dispose de nombreuses mesures (ou résultats d'observations) des valeurs exactes recherchées, donc de plus d'équations que d'inconnues : c'est en ce sens que le système est *surdéterminé*.

3. Le texte est daté du 15 ventôse an 13 (6 mars 1805). Il fut reproduit par Puissant dans son *Traité de Géodésie*, édité par Courcier toujours en 1805 (pages 137-141) puis republié dans la deuxième édition de l'ouvrage de Legendre en 1806. L'originale des *Nouvelles méthodes pour la détermination des orbites des comètes* de 1805 a pour éditeur Didot, et l'édition de 1806 a été éditée par Courcier.

4. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, Auctore Carolo Friderico Gauss, Hamburgi sumptibus Perthes et Besser, 1809, page 210 et seq.

*Au reste, ce principe, dont nous avons fait usage depuis 1795, a été donné dernièrement par Legendre dans ses Nouvelles méthodes pour la détermination des orbites des comètes, Paris, 1806 ; on trouvera dans cet ouvrage plusieurs conséquences que le désir d'abrégé nous a fait omettre.*



***Figure 1 : Portrait de Gauss (1777-1855) en 1840.***

Cette affirmation de Gauss irrita Legendre. Il écrit à Gauss, quelque temps après, en 1809 (Gauss, *Werke*, Band. 10/1, p. 380) :

*Je ne vous dissimulerai donc pas, Monsieur, que j'ai éprouvé quelque regret de voir qu'en citant mon mémoire ..., vous dites principum nostrum ... Il n'est aucune découverte qu'on ne puisse s'attribuer en disant qu'on avait trouvé la même chose quelques années auparavant ; mais si on n'en fournit pas la preuve en citant le lieu où on l'a publiée, cette assertion devient sans objet et n'est plus qu'une chose désobligeante pour le véritable auteur de la découverte. En Mathématiques il arrive très souvent qu'on trouve les mêmes choses qui ont été trouvées par d'autres et qui sont bien connues; c'est ce qui m'est arrivé nombre de fois, mais je n'en ai point fait mention et je n'ai jamais appelé principum nostrum un principe qu'un autre avait publié avant moi. Vous êtes assez riche de [votre] fonds, Monsieur, pour n'avoir rien à envier à personne; et [je suis] bien persuadé au reste que j'ai à me plaindre de l'expression seulement et nullement de l'intention ...*

**@@@@@@**

Dans une lettre à Laplace datée de 1812 (*Werke*, Band 10/1 pp. 373-374) Gauss déclare, pour sa part :

*J'ai fait usage de la méthode des moindres carrés depuis l'an 1795 (...) Cependant mes applications fréquentes de cette méthode ne datent que de l'année 1802, depuis ce temps j'en fait usage pour ainsi dire tous les jours dans mes calculs astronomiques sur les nouvelles planètes. (...). Au reste j'avais déjà communiqué cette même méthode, beaucoup avant la publication de l'ouvrage de M. Legendre, à plusieurs personnes, entre autres à Mr. Olbers en 1803. Ainsi, pouvais-je dans ma théorie parler de*

*la méthode des moindres carrés, dont j'avais fait depuis sept ans mille et mille applications (...). Pouvais-je parler de ce principe, -que j'avais annoncé à plusieurs de mes amis déjà en 1803 comme devant faire partie de l'ouvrage que je préparais,-comme d'une méthode empruntée de Mr. Legendre? Je n'avais pas l'idée, que Mr. Legendre pouvait attacher tant de prix à une idée aussi simple, qu'on doit plutôt s'étonner qu'on ne l'a pas eue depuis 100 ans, pour se fâcher que je raconte, que je m'en suis servi avant lui ? ... Mais j'ai cru que tous ceux qui me connaissent le croiraient même sur ma parole, ainsi que je l'aurais cru de tout mon coeur si Mr. Legendre avait avancé, qu'il avait possédé la méthode déjà avant 1795.*

La méthode des moindres carrés est qualifiée par Stigler dans son *History of Statistics* (Belknap Press, 1986) comme le thème dominant - le leitmotiv - des mathématiques statistiques au XIX<sup>e</sup> siècle. Stigler ajoute que la clarté de l'exposition de Legendre n'a pas été surpassée et que ce texte doit être compté dans l'histoire de la statistique comme l'une des introductions d'une nouvelle méthode les plus claires et les plus élégantes. Pour sa part, Joseph Bertrand, dans l'introduction de sa traduction de divers mémoires de Gauss intitulée *Méthode des moindres carrés*, (1855) affirme que Gauss considérait que la meilleure méthode pour combiner les observations était *un des problèmes les plus importants de la philosophie naturelle*.

### **LEGENDRE : UNE METHODE ALGEBRIQUE**

La méthode utilisée par Legendre pour exposer la méthode des moindres carrés n'est pas probabiliste mais purement algébrique : il traite le problème algébrique de la détermination des quantités dans un système d'équations surdéterminé.

On peut commenter la méthode de Legendre comme suit en reprenant les équations des erreurs écrites par Legendre au début de son mémoire :

$$E = a + bx + cy + fz + \dots$$

$$E' = a' + b'x + c'y + f'z + \dots$$

$$E'' = a'' + b''x + c''y + f''z + \dots$$

(etc.)

Où  $E, E', E''$ ...sont les erreurs à rendre minimales.

Legendre utilise une méthode empirique, simple à mettre en œuvre, qui consiste à rendre minimale la somme des carrés des erreurs et qu'il justifie par la phrase :

*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 (...) 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 somme des carrés des erreurs s'écrit :

$$S = E^2 + E'^2 + E''^2 + \dots = (a + bx + cy + fz + \dots)^2 + (a' + b'x + c'y + f'z + \dots)^2 + (a'' + b''x + c''y + f''z + \dots)^2 + \dots$$

et nous désirons que cette somme soit minimale par rapport aux inconnues  $x, y, z, \dots$

Les dérivées partielles  $\frac{dS}{dx}, \frac{dS}{dy}, \frac{dS}{dz}, \dots$  doivent donc être nulles.

Soit, pour la dérivée par rapport à  $x$

$$\frac{dS}{dx} = 2b(a + bx + cy + fz + \dots) + 2b'(a' + b'x + c'y + f'z + \dots) + 2b''(a'' + b''x + c''y + f''z + \dots) + \dots = 0$$

que l'on peut écrire, avec les notations de Legendre :

$$\int ab + x \int b^2 + y \int bc + z \int bf + \dots = 0, \text{ avec } \int ab = ab + a'b' + a''b'' + \dots$$

etc.

En prenant, de même, les dérivées partielles par rapport aux autres inconnues – qui sont en nombre  $n$  –, on obtient finalement  $n$  équations :

$$\int ab + x \int b^2 + y \int bc + z \int bf + \dots = 0$$

$$\int ac + x \int bc + y \int c^2 + z \int cf + \dots = 0$$

$$\int af + x \int bf + y \int fc + z \int f^2 + \dots = 0$$

.....

Le nombre d'équations étant égal aux nombres d'inconnues, il existe une solution et une seule, qui donne les valeurs des inconnues qui, selon l'approche empirique adoptée, sont les plus proches des valeurs réelles. Comme l'indique Legendre :

*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 moyenne arithmétique des mesures est un cas particulier lorsque l'on a une seule variable  $x$  avec un coefficient constant  $b = b' = b'' = \dots = 1$ . Legendre vérifie que, dans ce cas, la méthode des moindres carrés consiste à minimiser la somme

$$(x_1 - x)^2 + (x_2 - x)^2 + \dots + (x_n - x)^2,$$

soit en dérivant par rapport à  $x$

$$2(x_1 - x) + 2(x_2 - x) + \dots + 2(x_n - x) = 0$$

et donc  $x = \frac{x_1 + x_2 + \dots + x_n}{n}$ , c'est-à-dire la moyenne arithmétique des valeurs mesurées.

Legendre, ajoute la remarque suivante :

*Ces formules sont les mêmes par lesquelles on trouverait le centre de gravité commun de plusieurs masses égales, situés dans les points donnés (...) Si on divise la masse d'un corps en molécules égales et assez petites pour être considérées comme des points, la somme des carrés des distances des molécules au centre de gravité sera un minimum.*

La méthode, qui aboutit à ce résultat intuitif dans des cas simples, est néanmoins très facile à utiliser lorsque la situation est plus complexe. Par exemple, l'orbite elliptique d'une comète est déterminée par rapport au plan de l'écliptique par le calcul de six paramètres, inclinaison de l'orbite, demi grand-axe, excentricité, longitude du nœud ascendant, longitude du périhélie et date de passage de l'objet céleste au périhélie. En général, trois observations depuis la Terre donnant chacune six angles sont suffisantes pour résoudre le problème. La méthode de Gauss pour écrire les équations et les résoudre est alors la plus souvent employée.

### **L'APPROCHE PROBABILISTE DE LA METHODE DES MOINDRES CARRES CHEZ GAUSS ET LAPLACE**

Pour comprendre le problème sur le plan probabiliste nous allons dire quelques mots des méthodes de Gauss et de Laplace relatives au même problème.

On peut, en utilisant le concept de variable aléatoire, exposer comme suit le résultat de Gauss publié dans sa *Theoria motus corporum coelestium*<sup>5</sup> de 1809. Considérons que les erreurs sur une quantité  $X$  inconnue, pour laquelle on dispose de plusieurs mesures  $x_1, x_2, \dots, x_n$ , sont  $e_1 = x_1 - X, \dots, e_n = x_n - X$  et correspondent à des réalisations de  $n$  variables aléatoires indépendantes  $E_1, \dots, E_n$  de même loi continue de densité  $f$ , dépendant de la valeur inconnue de  $X$ .

Pour  $X$  donné, la probabilité d'effectuer des erreurs comprises entre  $e_1 + de_1, \dots, e_n + de_n$  est alors (compte tenu de l'indépendance)

$$f(e_1) \cdot f(e_2) \dots f(e_n) \cdot de_1 \dots de_n$$

On peut alors, selon la méthode de *Bayes ou de probabilité des causes*, les mesures  $x_1, \dots, x_n$  étant connues et considérées comme des conséquences de la distribution des erreurs, chercher quelle est la valeur de  $X$  (*la cause*) la plus vraisemblable.

Cette valeur de  $X$  rendra maximale la probabilité d'observation des mesures observées  $x_1, \dots, x_n$  et donc d'observation des erreurs  $e_1, \dots, e_n$ . Il s'agit de rechercher  $X$ , donc  $f$ , de sorte que  $f(x_1 - X) \dots f(x_n - X)$  soit maximum. Le produit  $f(x_1 - X) \dots f(x_n - X)$  est maximum lorsque la somme  $\log f(x_1 - X) + \dots + \log f(x_n - X)$  est maximale. En dérivant par rapport à  $X$  on obtient la condition

$$\frac{d \log f(x_1 - X)}{dX} + \dots + \frac{d \log f(x_n - X)}{dX} = 0 \quad (1)$$

On sait que la moyenne arithmétique  $\bar{X} = \frac{x_1 + \dots + x_n}{n}$  correspond à la valeur recherchée de  $X$ . Ceci correspond à un axiome dans la théorie de Gauss<sup>6</sup>. Cette moyenne vérifie l'équation en  $X$

$$\sum_{i=1}^n (x_i - X) = 0$$

Gauss induit que, pour vérifier l'équation (1), on a, pour  $i$  allant de 1 à  $n$ ,

$$\frac{d \log f(x_i - X)}{dX} = k(x_i - X).$$

5. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*, Auctore Carolo Friderico Gauss, Hamburgi sumptibus Perthes et Besser, 1809, page 210 et seq.

6. À ce propos, Gauss écrit : « On a coutume de regarder comme un axiome l'hypothèse que si une 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 ».

En prenant  $k(x_i - X)$ , Gauss prend ainsi une fonction linéaire en  $x_i$ , fonction *impaire* la plus simple (la fonction  $f$  d'erreur étant nécessairement paire – même probabilité d'erreur dans un sens ou dans l'autre –, la dérivée de son logarithme est nécessairement impaire). On a alors, en intégrant,

$$\log f(x_i - X) = -k \frac{(x_i - X)^2}{2} + cte, \text{ soit:}$$

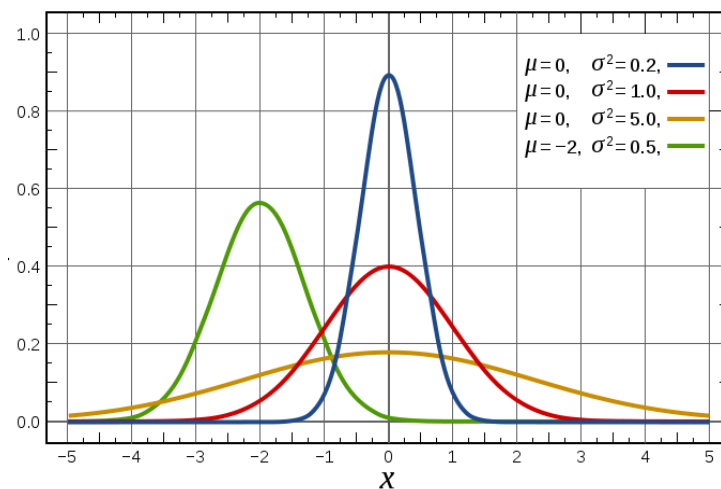
$$f(x_i - X) = Ce^{-k \frac{(x_i - X)^2}{2}}$$

c'est à dire la densité de la loi de probabilité dite normale.

On peut vérifier qu'avec cette loi, le produit  $f(x_1 - X) \dots f(x_n - X) = Ce^{-k \frac{\sum_{i=1}^n (x_i - X)^2}{2}}$  est maximum (ce qui est recherché d'après ce

qui précède) lorsque la somme des carrés des écarts des mesures  $\sum_{i=1}^n (x_i - X)^2$  est minimale.

La loi normale la plus générale (dite de Laplace-Gauss) fait intervenir une valeur moyenne  $\mu$  et un écart type  $\sigma$ .



**Figure 2 : Fonction densité de probabilité de la « loi normale » (loi de Gauss-Laplace) d'espérance  $\mu$  et de variance  $\sigma$ .**

En résumé, Gauss, dans son traitement de la théorie des erreurs de 1809, introduit en fait le principe du maximum de vraisemblance (puisqu'il cherche la valeur de la quantité inconnue  $X$  qui rendra maximale la probabilité d'observation des mesures observées  $x_1, \dots, x_n$  et donc des erreurs  $e_1, \dots, e_n$ ). Après ajout du postulat de la moyenne arithmétique, il obtient à la fois la loi de distribution normale et le principe des moindres carrés. La justification de ce principe des

moindres carrés est donc la maximalisation de la distribution *a posteriori* des erreurs.

@@@@@@

En ce qui concerne Laplace, il écrit de son côté, dans son mémoire de 1811<sup>7</sup> dans l'objectif de généraliser les résultats de Legendre et de Gauss :

*Lorsque l'on veut corriger par l'ensemble d'un grand nombre d'observations plusieurs éléments déjà connus à peu près, on s'y prend de la manière suivante. Chaque observation étant une fonction des éléments, on substitue dans cette fonction leurs valeurs approchées, augmentées respectivement de petites corrections qu'il s'agit de connaître. En développant ensuite la fonction en série, par rapport à ces corrections, et négligeant leurs carrés et leurs produits, on égale la série à l'observation, ce qui donne une première équation de condition entre les corrections des éléments. Une seconde observation fournit une équation de condition semblable, et ainsi du reste. Si les observations étaient rigoureuses, il suffirait d'en employer autant qu'il y a d'éléments ; mais, vu les erreurs dont elles sont susceptibles, on en considère un grand nombre, afin de compenser les unes par les autres ces erreurs, dans les valeurs des corrections que l'on déduit de leur ensemble. Mais de quelle manière, faut-il combiner entre elles les équations de condition pour avoir les corrections les plus précises ? C'est ici que l'analyse des probabilités peut être d'un grand secours. Toutes les manières de combiner ces équations se réduisent à les multiplier chacune par un facteur particulier et à faire une somme de tous ces produits : on forme ainsi une première équation finale entre les corrections des éléments. Un second système de facteurs donne une seconde équation finale, et ainsi de suite jusqu'à ce que l'on ait autant d'équations finales que d'éléments dont on déterminera les corrections en résolvant ces équations. Maintenant il est visible qu'il faut choisir les systèmes de facteurs, de sorte que l'erreur moyenne à craindre en plus ou en moins sur chaque élément soit un minimum. J'entends par erreur moyenne la somme des produits de chaque erreur par sa probabilité. En déterminant les facteurs par cette condition, l'analyse conduit à ce résultat remarquable, savoir que, si l'on prépare chaque équation de condition de manière que son second membre soit zéro, la somme des carrés des premiers membres est un minimum, en y faisant varier successivement chaque correction. Ainsi cette méthode, que MM. Legendre et Gauss ont proposée, et qui, jusqu'à présent, ne présentait que l'avantage de fournir, sans aucun tâtonnement, les équations finales nécessaires pour corriger les éléments, donne en même temps les corrections les plus précises.*

---

7. « Sur les intégrales définies et leur application aux probabilités et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations », *Mémoires de l'Institut National des sciences et des arts ; sciences mathématiques et physique*, 11, 1811, pp. 279-347.



En résumé, Gauss avait démontré en 1809 la loi normale en considérant comme un axiome l'hypothèse que si une quantité a été obtenue par plusieurs observations, faites avec le même soin dans des circonstances semblables, la moyenne arithmétique des valeurs observées sera la valeur la plus probable de cette quantité. Laplace, dans la suite de son mémoire de 1810, démontre que même en ignorant la distribution des erreurs sur chaque groupe de mesures, on peut, lorsque le nombre d'observations est très grand, remplacer les mesures par leur moyenne empirique et obtenir pour l'ensemble des observations une loi normale. Comme dans le mémoire de Gauss, Laplace obtient le meilleur résultat en minimisant la somme des carrés des erreurs de chaque résultat, multipliées respectivement par la plus grande ordonnée de la courbe de probabilité des erreurs (car la valeur de *l'inconnue*  $X$  doit rendre maximale la probabilité d'observation des mesures observées  $x_1, \dots, x_n$  et donc maximale la probabilité d'observation des erreurs  $e_1, \dots, e_n$ ). Il est donc évident que les résultats de Laplace sont obtenus, comme ceux de Gauss, à l'aide d'un raisonnement probabiliste.

### Une explication du raisonnement de Gauss par Joseph Bertrand



**Figure 3 :** *Le mathématicien Joseph Bertrand (1822-1900). Bertrand était polytechnicien (X1839), secrétaire perpétuel de l'Académie des sciences à partir de 1874, membre de l'Académie française en 1884.*

L'histoire de la méthode des moindres carrés est singulière. Legendre, qui, le premier, l'a proposée, en 1805, aux calculateurs, ne la rattachait à aucune idée théorique. Gauss, qui depuis longtemps en faisait usage, a publié, en 1809, une démonstration acceptée sans contestation.

(...)

Si des grandeurs  $l_1, l_2, \dots, l_n$  ont été observées, la probabilité d'une erreur comprise entre  $z$  et  $z + dz$ , pour l'une d'elles, étant  $\varphi(z)dz$ , la probabilité du concours des erreurs  $z_1, z_2, \dots, z_n$  est proportionnelle à

$$\varphi(z_1) \varphi(z_2) \dots \varphi(z_n) \quad (1)$$

et ce produit doit être rendu maximum.

La fonction  $\varphi(z)$  restant inconnue, le problème paraît insoluble. L'hypothèse la plus plausible est, d'après le résultat du calcul, celle qui rend minima la somme

$$z_1^2 + z_2^2 + \dots + z_n^2 \quad (2)$$

Si le produit (1) est maximum quand la somme (2) est minima, il doit rester constant en même temps que celle-ci et la fonction  $\varphi(z)$  doit satisfaire à une équation de la forme

$$\varphi(z_1) \varphi(z_2) \dots \varphi(z_n) = F(z_1^2 + z_2^2 + \dots + z_n^2)$$

On déduit de cette équation bien connue

$$\varphi(z) = G e^{-k^2 z^2}$$

C'est la loi proposée en 1809; elle seule peut justifier la règle qu'on en a déduite.

(...)

Toutes les équations, dans la théorie de Gauss, sont réduites au premier degré (...) on peut,  $z$  étant très petit, remplacer  $\varphi(z)$  par son développement

$$\varphi(0) + z\varphi'(0) + \frac{z^2}{2}\varphi''(0) + \frac{z^3}{6}\varphi'''(0) + \dots$$

Les erreurs constantes<sup>8</sup> étant écartées, c'est une des hypothèses de Gauss, on a  $\varphi(z) = \varphi(-z)$  et, par conséquent,  $\varphi'(0) = 0$  et  $\varphi'''(0) = 0$ .

On peut, puisqu'on est convenu de négliger  $z^2$  par rapport à  $z$ , négliger  $z^4$  par rapport à  $z^2$  et réduire  $\varphi(z)$  à la forme

$$\varphi(z) = \varphi(0) + \frac{1}{2}\varphi''(0)z^2 = a + bz^2$$

Mais on a, dans les mêmes conditions,

$$G e^{-k^2 z^2} = G - Gk^2 z^2$$

Si donc on pose  $G = a$ ,  $k = -b^2/a$ , on aura

$$\varphi(z) = G e^{-k^2 z^2}$$

La fonction  $\varphi(z)$ , quelle que soit sa forme véritable, est réductible, dans les conditions où l'on veut résoudre le problème, à celle que Gauss avait proposée d'abord.

(extrait de Joseph Bertrand, *Sur la méthode des moindres carrés*, C.R.A.S., Paris, Tome CVI, n°16, 1888, pp. 1115-1117).

8. Bertrand fait ici référence aux erreurs systématiques habituellement faites dans les mesures. La méthode des moindres carrés nécessite de les éliminer et de ne tenir compte que des erreurs aléatoires - d'où la symétrie de la courbe d'erreurs par rapport à l'axe des  $y$ .

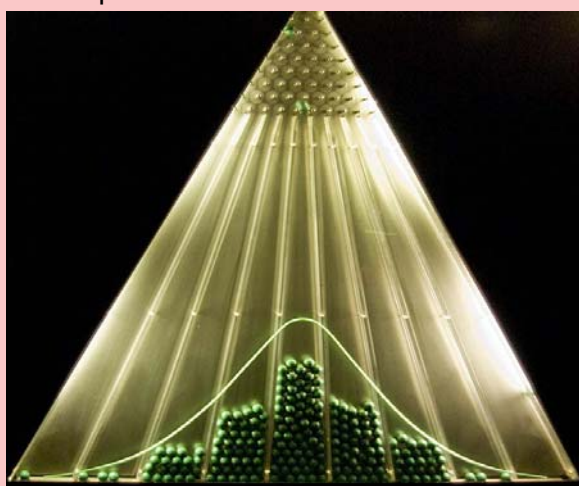
@@@@@@

L'établissement, en 1809 et 1810, de la *loi normale*, dont l'usage se révélera fondamental en statistique, s'est donc fait par deux voies : celle de Gauss relative à la théorie des erreurs aboutissant à la méthode des moindres carrés, et celle de Laplace qui a énoncé la première version du théorème de la limite centrale.

La synthèse entre ces deux approches impose, dès 1810, la loi normale comme loi quasi universelle. En effet, même si la distribution individuelle des erreurs ne suit pas une loi normale, celle des moyennes des erreurs suit sous certaines conditions (indépendance, lois identiques), une loi normale.

### Le théorème de la limite centrale

Le théorème de la limite centrale établit que la somme d'un grand nombre de réalisations de variables aléatoires indépendantes et ayant la même loi de distribution est approximativement distribuée suivant la loi normale si les variables aléatoires ont une variance finie. Abraham de Moivre avait établi le théorème en 1733 pour une loi binomiale et Laplace approxima en 1811 la loi binomiale par la loi normale. Le théorème de la limite centrale est appelé de nos jours le second théorème fondamental des probabilités, le premier étant la loi des grands nombres de Jacob Bernoulli qui indique que la moyenne des réalisations de variables aléatoires indépendantes ayant la même loi de distribution et une espérance finie converge vers la valeur de cette espérance mathématique.



**Figure 4 : Planche de Galton** (scientifique anglais, 1811-1922). Une bille lâchée en haut croise divers pitons sur son chemin et tombe dans un des compartiments du bas. Cette chute obéit à la loi de probabilités dite binomiale (voir les billes en bas), qui converge vers la loi de Gauss (courbe en cloche au-dessus).

Le nombre de travaux concernant ces théorèmes étant historiquement très élevé, Lucien Le Cam a commenté avec humour, en 1986, l'histoire du Théorème de la limite centrale [*The central limit theorem around 1935*. Statistical science, volume 1, 1986, pp.78-96.] (traduction):

*Au commencement, il y eut De Moivre, Laplace, et de nombreux Bernoullis et ils engendrèrent des théorèmes de la limite, et les hommes avisés dirent que cela était bien et ils donnèrent au théorème le nom de Gauss. Puis vinrent de nouvelles générations qui trouvèrent que le théorème avait une validité expérimentale mais manquait de rigueur. Alors vinrent Chebychev, Liapunov et Markov et ils prouvèrent le théorème et Polya voyant l'usage des moments déclara que le nom Théorème de la limite centrale devait être utilisé. Puis vint Lindeberg et il dit que cela était élémentaire car Taylor avait développé ce qui demandait à être développé, et il le dit deux fois (...)*

Compte tenu de l'utilisation permanente de la loi normale, on a, par ailleurs, défini une *fonction erreur* qui est l'intégrale numérique qui a été introduite par Gauss en 1816. On la trouve tabulée dans tous les manuels et elle est définie par

$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\zeta^2} d\zeta$  . La probabilité pour qu'une variable normale centrée réduite<sup>9</sup> prenne une valeur dans l'intervalle  $[-z, z]$  est  $erf\left(\frac{z}{\sqrt{2}}\right)$ .

### AUTRE PRESENTATION DE LA METHODE DES MOINDRES CARRÉS

Dans la théorie des moindres carrés, on cherche à estimer un vecteur  $\vec{\beta}$  [1 x r], ayant ses r composantes inconnues, à partir de N observations  $\vec{Y}$  reliées linéairement à  $\vec{\beta}$ , mais comportant des erreurs  $\vec{\varepsilon}$ .

On a  $\vec{Y} = M\vec{\beta} + \vec{\varepsilon}$  où

$M = \{x_{ij}\}$  est une matrice connue de N x r éléments avec  $N \geq r$ , la colonne de rang r étant pleine.

9. Une variable aléatoire centrée réduite a une espérance mathématique E(X) nulle et un écart type égal a 1. Par rapport à une variable aléatoire quelconque X, la variable centrée réduite est une variable translaturée et homothétique  $\frac{X-E(X)}{\sigma(X)}$ .

Le système des multiplicateurs  $k_{ij}$  utilisés par Legendre, Gauss et Laplace pour traiter la méthode des moindres carrés constituait en fait la recherche d'une matrice  $K$  de  $r \times n$  éléments  $K = \{k_{ij}\}$ , telle que l'on obtienne  $KM = I$ ,  $I$  étant la matrice unité.

L'estimation de la moyenne du vecteur  $\vec{\beta}$  noté  $\bar{\vec{\beta}}$ , consiste alors à choisir la matrice  $K$  selon une approche probabiliste ou une approche algébrique afin d'obtenir l'estimation

$$\bar{\vec{\beta}} = \vec{\beta} + K\vec{\varepsilon} = K\vec{Y}$$

L'approche de Legendre, purement algébrique, consiste à considérer la relation  $\vec{Y} = M\vec{\beta} + \vec{\varepsilon}$  comme un jeu surdéterminé d'équations linéaires ; la matrice  $K$  est alors établie en agissant sur les erreurs elles-mêmes. L'approche probabiliste consiste à établir  $K$  en imposant les distributions de probabilité sur les erreurs.

@@@@@@

Il est à noter qu'un très grand nombre de publications ont été faites, depuis celles de Legendre, Gauss et Laplace, sur ce problème des moindres carrés. Un travail remarquable d'André-Louis Cholesky (1910) montre que la décomposition de la matrice définie plus haut est possible en un produit d'une matrice triangulaire par sa matrice transposée, ramenant la résolution du système de départ à deux résolutions successives de systèmes plus aisés à résoudre<sup>10</sup>.

## CONCLUSION

Legendre a terminé son exposé de 1805 sur les moindres carrés par une application à la mesure des degrés du méridien à partir de données établies par Delambre et Méchain. La méthode des moindres carrés a été surtout utilisée au dix-neuvième siècle en astronomie et en géodésie, c'est pourquoi le titre de l'ouvrage de Legendre est relatif aux comètes et son exemple d'application à la géodésie.

---

<sup>10</sup> Roger Mansuy, analyse du manuscrit *Sur la résolution numérique des systèmes d'équations différentielles* écrit par Cholesky, [Bibnum](#) septembre 2008.

Cette méthode est aujourd'hui universellement utilisée de même que la loi normale dite de Laplace-Gauss et les quelques lignes qui précèdent ont tenté de rappeler les circonstances de leur établissement.



*(août 2010)*

## Annexe

212

### LIBR. II. SECT. III.

$$\frac{dv}{dr} \varphi' v + \frac{dv'}{dr} \varphi' v' + \frac{dv''}{dr} \varphi' v'' + \text{etc.} = 0$$

$$\frac{dv}{ds} \varphi' v + \frac{dv'}{ds} \varphi' v' + \frac{dv''}{ds} \varphi' v'' + \text{etc.} = 0$$

Hinc itaque per eliminationem problematis solutio plene determinata derivari poterit, quamprimum functionis  $\varphi'$  indoles innotuit. Quae quoniam a priori definiri nequit, rem ab altera parte aggredientes inquiremus, cuiam functioni, tacite quasi pro basi acceptae, proprie innixum sit principium triuivium, cuius praestantia generaliter agnoscitur. Axiomati scilicet loco haberi solet hypothesis, si quae quantitas per plures observationes immediatas, sub aequalibus circumstantiis aequalique cura institutas, determinata fuerit, medium arithmeticum inter omnes valores observatos exhibere valorem maxime probabilem, si non absoluto rigore, tamen proxime saltem, ita ut semper tutissimum sit illi inhaerere. Statuendo itaque  $V = V' = V'' \text{ etc.} = p$ , generaliter esse debebit  $\varphi'(M-p) + \varphi'(M'-p) + \varphi'(M''-p) + \text{etc.} = 0$ , si pro  $p$  substituitur valor  $\frac{1}{\mu} (M + M' + M'' + \text{etc.})$ , quemcunque integrum positivum exprimat  $\mu$ . Supponendo itaque  $M' = M'' = \text{etc.} = M - \mu N$ , erit generaliter, i. e. pro quovis valore integro positivo ipsius  $\mu$ ,  $\varphi'(\mu - 1)N = (1 - \mu)\varphi'(-N)$ , vade facile colligitur, generaliter esse debere  $\frac{\varphi' \Delta}{\Delta}$  quantitatem constantem, quam per  $k$  designabimus. Hinc fit  $\log \varphi \Delta = \frac{1}{2} k \Delta \Delta + \text{Const.}$ , siue designando basin logarithmorum hyperbolicorum per  $e$ , supponendoque  $\text{Const.} = \log x$ ,

$$\Delta \varphi = x e^{\frac{1}{2} k \Delta \Delta}$$

Porro facile perspicitur,  $k$  necessario negativam esse debere, quo  $\Omega$  revera fieri possit maximum, quamobrem statuemus  $\frac{1}{2} k = -hh$ ; et quum per theorema elegans primo ab ill. Laplace inuentum, integrale  $\int e^{-hh\Delta\Delta} d\Delta$ , a  $\Delta = -\infty$  vsque ad  $\Delta = +\infty$ , fiat  $= \frac{\sqrt{\pi}}{h}$ , (denotando per  $\pi$  semicircumferentiam circuli cuius radius 1), functio nostra fiet

$$\varphi \Delta = \frac{h}{\sqrt{\pi}} e^{-hh\Delta\Delta}$$

178.

Functio modo eruta omni quidem rigore errorum probabilitates exprimere certo non potest: quum enim errores possibiles semper limitibus certis coërceantur,

**La densité de la loi normale apparaît page 212 de l'ouvrage de Gauss,  
Theoria motus corporum, 1809.**