JOURNAL

DE

MATHÉMATIQUES

PURES ET APPLIQUÉES

FONDÉ EN 1836 ET PUBLIÉ JUSQU'EN 1874

PAR JOSEPH LIOUVILLE

BIENAYMÉ

Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés

Journal de mathématiques pures et appliquées 2e série, tome 12 (1867), p. 158-176.

<http://www.numdam.org/item?id=JMPA_1867_2_12__158_0>

gallica NUMDAM

Article numérisé dans le cadre du programme Gallica de la Bibliothèque nationale de France
http://gallica.bnf.fr/

et catalogué par Mathdoc dans le cadre du pôle associé BnF/Mathdoc
http://www.numdam.org/journals/JMPA

CONSIDÉRATIONS

A l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés;

PAR M. BIENAYMÉ [*].

Extrait des *Comptes rendus des Séances de l'Académie des Sciences*, t. XXXVII, séance du 29 août 1853.

Dans la séance du 8 août 1853, l'Académie a entendu des remarques, des réflexions faites par M. Cauchy et par M. Le Verrier sur la méthode des moindres carrés. Ces réflexions n'ont pas été insérées dans le *Compte rendu*, à mon grand regret, car elles m'avaient paru d'un intérêt véritable. Il n'a été publié que le Mémoire de M. Cauchy, dont la lecture avait fourni l'occasion de cette espèce de discussion; et je n'ai point vu reproduite une opinion qui m'avait semblé le diriger dans ses observations verbales. M. Cauchy avait nié l'exactitude du résultat si remarquable, découvert et démontré par Laplace, et qui consiste en ce que la méthode des moindres carrés s'applique aux données des observations, quelle que soit la loi de probabilité des erreurs. J'ai cru pouvoir annoncer, alors, que j'aurais quelques bonnes raisons peut-être à présenter à l'appui de l'opinion de Laplace. Je viens les exposer le plus brièvement possible; quoique je sois en droit de faire remarquer cependant que les divers Mémoires ou morceaux de Mémoires publiés par notre savant confrère, dans les n[os] 5, 6 et 7 du *Compte rendu*, ne justifient nullement son assertion. Loin de là,

[*] Depuis longtemps nous nous proposions de reproduire dans le *Journal de Mathématiques* l'important travail de M. Bienaymé. Quatorze années se sont écoulées; nous croyons pourtant que nos lecteurs l'accueilleront encore aujourd'hui avec un vif intérêt. (J. L.)

selon moi, il y a telles parties de cette analyse si féconde, et parfois très-ingénieuse, qui, avec bien peu de changements, démontreraient pleinement la découverte de Laplace. Et si je me borne à l'indiquer, c'est que je ne veux pas même avoir l'air d'en faire la critique.

La critique n'est point mon but. Les travaux critiques que j'aurai à présenter successivement à l'Académie montreront que ce n'est, à mes yeux, entre mes mains, qu'un instrument dont je suis obligé de me servir pour parvenir à la vérité. Le calcul des probabilités donne lieu à de singulières illusions, auxquelles les meilleurs esprits n'ont pu toujours se soustraire. Ce calcul est le premier pas des Mathématiques hors du domaine de la vérité absolue. Elles ne l'ont fait qu'en tâtonnant, pour ainsi dire. Laplace intitule un de ses chapitres : *Des illusions dans l'estimation des probabilités.* Aussi trouve-t-on plus d'une erreur étayée d'un nom recommandable, plus d'une méprise échappée à des hommes qui font autorité. La science, quoique Pascal l'ait créée il y a deux siècles, n'est pas bien loin de son origine. Le développement analytique a progressé presque seul. Or, dans ces commencements d'une science, celui qui cherche le vrai semble parfois user d'une critique bien rigoureuse, alors qu'il s'efforce seulement de sortir de l'obscurité, de faire évanouir les fantômes qu'on y croit apercevoir. Il se voit contraint tout à la fois de réduire à d'étroites limites les avantages que trop de précipitation, trop d'imagination avaient fait exagérer outre mesure; de mettre en évidence les inconvénients que l'ardeur des premiers succès avait dissimulés; parfois aussi, de maintenir les points justement acquis par la science contre des préventions, des jugements inexacts qui la feraient rétrograder. En un mot, il est forcé de déblayer le terrain, et de le défendre pour que ses successeurs puissent y avancer avec sécurité. C'est ce que peut seule faire une critique impartiale. Et puis, il faut bien le dire, la nature même du calcul des probabilités, qui traite des erreurs de toute espèce, des écarts de tout genre, de l'incompatibilité des observations résultant de la faiblesse des organes humains, de la discordance des données statistiques, discordance que produisent les variations physiques, très-étendues le plus souvent, et la négligence ou l'impéritie avec lesquelles ces variations se constatent; la nature de ce calcul qui traite de tous ces faits pour lesquels a été forgé le mot de hasard, si peu intelligible; la nature même de ce calcul est critique.

J'aurai donc besoin, pour ce caractère de plusieurs de mes communications futures, d'une bienveillante disposition de l'auditeur ou du lecteur. Je la sollicite d'avance. Il y a, du reste, bien longtemps que la plupart de mes recherches ont été faites. Les démonstrations de certaines erreurs remontent à plus de trente ans. J'ose espérer qu'on voudra bien reconnaître, dans le long silence où elles sont demeurées, la preuve d'une absence complète du désir de critiquer. D'ailleurs, j'agirai comme je l'ai déjà fait; je réduirai la critique au strict nécessaire, et je crois que M. Cauchy le verra dans ma communication actuelle.

Et cependant l'opinion de Laplace sur le sujet qui nous occupe mériterait bien que toutes les ressources possibles fussent employées à la soutenir. Il ne s'agit pas ici seulement de la méthode des moindres carrés, d'une combinaison d'équations à résoudre, qu'on pourrait remplacer, sans grand dommage, par une autre combinaison. Peut-être, s'il ne s'était agi que des erreurs des observations, n'aurais-je pas pris la parole il y a trois semaines; bien que je dusse supposer que j'étais un peu cause de l'amoindrissement que notre savant confrère voulait faire subir à la méthode des moindres carrés. Mais voici pourquoi j'ai dû parler. C'est, qu'on le sache bien, que si la découverte de Laplace est inexacte, une partie considérable de son grand ouvrage, la partie la meilleure, en ce qu'elle est la plus applicable, la plus pratique, se trouvera renversée d'un seul coup. Tous les chapitres qui traitent *de la recherche des phénomènes et de leurs causes; de la probabilité des causes et des événements futurs tirée des événements observés; des durées moyennes de la vie, des mariages et des associations quelconques; des bénéfices des établissements basés sur la probabilité des événements, etc.*, suivront plus ou moins complétement le sort du célèbre chapitre IV *sur la probabilité des erreurs des résultats moyens*, où se trouve démontrée la méthode que Legendre avait fondée sur des considérations si différentes. Car tous ces chapitres sont appuyés sur un principe commun, sur la réduction d'une fonction quelconque à des termes communs à toutes, et faciles à calculer.

Aussi Laplace avait-il senti sur-le-champ l'importance de sa découverte. A peine l'eut-il faite, qu'il l'apporte devant cette compagnie, et qu'il annonce qu'il va publier un Traité des probabilités. De 1770 à 1809, pendant près de quarante ans, Laplace avait donné des Mémoires nombreux sur les probabilités: mais, quelque intérêt qu'il y eût dans ces Mémoires, il n'avait pas voulu les rédiger en théorie générale. Aussitôt qu'il a reconnu la propriété des fonctions de probabilités, il voit clairement que c'est un principe qui régit presque toutes les applications, et il compose sa théorie.

On a dit souvent que cette théorie rendrait, seule, immortel le nom de Laplace. Mais je suis fondé à croire que ces paroles ne portaient que sur les beaux développements d'analyse qui lui sont dus; et que le principe de probabilités, si général, si fécond, n'a frappé les yeux que d'un bien petit nombre de personnes. S'il n'en était ainsi, je suis persuadé qu'au lieu de l'ébranler, de le révoquer en doute, on s'efforcerait de le consolider, et de montrer comment Laplace l'a entendu; comment on peut en abuser par des applications mal conçues, ou incomplètes; comment enfin les progrès de l'analyse peuvent servir à répandre le jour sur le calcul des probabilités. Mais plusieurs n'y ont vu qu'une occasion de problèmes d'analyse, et ceux qui ont cherché l'esprit de cette analyse se sont épuisés en vains efforts pour remplacer les calculs de Laplace par ce qu'on appelle des *démonstrations faciles*, des *preuves populaires*.

Jusqu'ici, il ne semble pas possible d'opérer ce remplacement dès qu'on veut connaître la grandeur de la probabilité d'une erreur donnée. L'analyse suivie par Poisson, l'analyse que M. Cauchy vient d'employer avec un peu plus de cette rigueur qui doit régir les Mathématiques, ne donnent, en fait de calculs numériques, de vraies applications pratiques, rien de plus que les formules si commodes que Laplace a su tirer de la sienne.

Mais si l'on veut se borner à démontrer la méthode des moindres carrés en ce qui touche la combinaison des observations, sans calculer la grandeur de l'erreur et seulement en prouvant que cette méthode opère de manière à la rendre un minimum, il ne faut plus d'analyse transcendante, ni même de grande contention d'esprit, une fois qu'on a bien envisagé les conditions qui font un principe général de la découverte de Laplace.

Deux choses, en effet, ont toujours surpris, et surprendront toujours, quand, pour la première fois, on vient à considérer le résultat final qu'il a livré aux observateurs. C'est, d'une part, que l'intégrale $\frac{2}{\sqrt{\pi}}\int_0^a dte^{-t^2}$ se reproduit sans cesse, comme expression approchée de la probabilité; et de l'autre, que les écarts ou les erreurs sont proportionnels à la limite de cette intégrale, et à une fonction de la moyenne des carrés des différences entre chaque erreur possible et sa moyenne. C'est là tout ce qu'il reste, dans l'approximation de Laplace, du caractère primitif de la fonction de probabilité qui régnait pendant les observations.

Il semble, au premier abord, qu'il y ait un lien mystérieux entre les probabilités et cette intégrale qui s'est présentée à M. Gauss dans ses premières recherches, comme une conséquence du principe de la moyenne arithmétique qu'il adoptait gratuitement, et qui est ensuite ressortie de l'analyse de Laplace, basée uniquement sur ce que les observations sont en grand nombre. Aussi, plus d'un savant a-t-il pensé qu'une connaissance mieux approfondie de la théorie des probabilités amènerait à reconnaître que la loi de probabilité des erreurs est représentée par la fonction $\frac{1}{\sqrt{\pi}}e^{-t^2}$. Mais c'est là une de ces vues inexactes qui conduisent à des conséquences erronées : et celle-ci en a produit en grand nombre, tant au point de vue théorique que dans les résultats pratiques. Il suffira de dire ici qu'il n'y a pas de liaison nécessaire entre cette exponentielle et les lois de probabilités; qu'elle n'est qu'un moyen d'approximation très-commode, mais qui pourrait être remplacé par d'autres formules; que ce qui la ramène sans cesse, c'est qu'elle est très-propre à représenter une fonction aux environs du maximum, mais qu'elle ne la représente qu'à peu près; et que même dans les questions où elle offre le plus de facilité comme approximation, elle donne fréquemment des résultats dont la fausseté est manifeste dès qu'on veut l'employer à des raisonnements un peu complexes, au lieu de la tenir pour ce qu'elle est réellement, c'est-à-dire pour une pure machine arithmétique, bonne aux calculs numériques.

Mais il n'en est pas ainsi de la moyenne des carrés des différences des erreurs à leur moyenne. Ce n'est pas un élément arbitraire de l'approximation, ni, comme le croyait M. Gauss, une mesure arbitraire de la précision, à laquelle on pourrait substituer toute autre moyenne de puissances de degré pair. Tout au contraire, la moyenne des carrés renferme la condition fondamentale du développement des probabilités à mesure que les observations se multiplient; et s'il n'est pas permis de dire *à priori* qu'on ne saurait trouver quelque fonction plus avantageuse (autre que les moyennes de degré pair), on peut affirmer que dans l'hypothèse d'une telle découverte la fonction pourrait être remplacée par un emploi convenable de la moyenne des carrés.

La raison en est bien simple. Pour l'expliquer, je prendrai, comme l'a fait M. Cauchy, le cas le plus ordinaire de la méthode des moindres carrés.

Lorsqu'on veut résoudre un système d'équations du premier degré où entrent des quantités observées ω_1, ω_2, $\omega_3,\ldots$, qui sont affectées de certaines erreurs égales respectivement à ε_1, ε_2, $\varepsilon_3,\ldots$, on multiplie chacune de ces équations par un des facteurs arbitraires h_1, h_2, $h_3,\ldots$, assujettis à faire disparaître toutes les inconnues, sauf une seule, que l'on obtient sous la forme

$$x = h_1\omega_1 + h_2\omega_2 + h_3\omega_3 + \ldots,$$

de sorte que l'erreur de x a pour valeur

$$\xi = h_1\varepsilon_1 + h_2\varepsilon_2 + h_3\varepsilon_3 + \ldots.$$

En général, l'erreur ξ est susceptible de toutes les valeurs possibles, parce que les signes des coefficients et ceux des erreurs sont déjà déterminés, et qu'on ne peut disposer complétement des facteurs h, qui doivent satisfaire à $i-1$ équations, s'il y a i inconnues, afin d'en éliminer $i-1$.

Il s'agit donc, pour obtenir la probabilité de ξ, d'examiner quelle est la loi de probabilité d'une somme de n produits, s'il y avait n équations, chaque produit formé d'une erreur multipliée par un facteur donné.

Soit b_i la probabilité d'une erreur ε_i, de manière que la somme

$$b_1 + b_2 + b_3 + \ldots = 1,$$

comme cela doit être; et considérons le polynôme

$$b_1 z^{\alpha_1} + b_2 z^{\alpha_2} + b_2 z^{\alpha_3} + \ldots = \mathrm{S}.bz^{\alpha} = \varphi(z),$$

dans lequel un exposant α_i est une fonction déterminée de l'erreur ε_i.

Si l'on fait, de même,

$$\mathrm{S}.bz^{\beta} = \psi(z), \quad \mathrm{S}.bz^{\gamma} = \chi(z), \ldots,$$

β, γ, etc., étant des fonctions des erreurs, il est évident que le produit

$$\mathrm{S}.bz^{\alpha} \times \mathrm{S}.bz^{\beta} \times \mathrm{S}.bz^{\gamma} \times \ldots = \varphi(z).\psi(z).\chi(z)\ldots = \mathrm{P}$$

aura dans chacun de ses termes, comme exposant, l'une des valeurs de toutes les sommes que l'on peut former en ajoutant n des quantités α, β, γ, etc., prises à volonté, et pour coefficient la probabilité de cette valeur. Ainsi ce produit P, ordonné par rapport à la grandeur des exposants de z, quels qu'ils soient, présentera la loi de probabilité des sommes dont une seule a pu se rencontrer dans les expériences ou observations; de même que $\varphi(z) = \mathrm{S}.bz^{\alpha}$ présente la loi de probabilité des quantités α. Désignant par σ_i l'une des sommes

$$\alpha + \beta + \gamma + \ldots$$

et par B_i la probabilité correspondante, on pourra écrire $\mathrm{P} = \mathrm{S}.\mathrm{B}z^{\sigma}$, et ce qui va être dit de $\varphi(z)$ s'appliquera au produit P.

On sait que les dérivées logarithmiques de $\varphi(z)$ seront

$$\frac{\varphi'(z)}{\varphi(z)}, \quad \frac{\varphi''(z)}{\varphi(z)} - \left[\frac{\varphi'(z)}{\varphi(z)}\right]^2, \quad \frac{\varphi'''(z)}{\varphi(z)} - 3\,\frac{\varphi''(z)\,\varphi'(z)}{(\varphi z)^2} + 2\left[\frac{\varphi'(z)}{\varphi(z)}\right]^3,$$

$$\frac{\varphi^{\mathrm{iv}}(z)}{\varphi(z)} - 4\,\frac{\varphi'''(z)\,\varphi'(z)}{[\varphi(z)]^2} - 3\,\frac{[\varphi''(z)]^2}{[\varphi(z)]^2} + 12\,\frac{\varphi''(z)[\varphi'(z)]^2}{[\varphi(z)]^3} - 6\,\frac{[\varphi'(z)]^4}{[\varphi(z)]^4}, \ldots$$

Mais on sait aussi que les dérivées logarithmiques d'un produit

$$\mathrm{P} = \varphi(z)\,\psi(z)\,\chi(z)\ldots$$

sont respectivement les sommes des dérivées de même ordre des logarithmes des facteurs

$$\frac{d^n(\mathrm{l.\,P})}{dz^n} = \frac{d^n[\mathrm{l.}\,\varphi(z)]}{dz^n} + \frac{d^n[\mathrm{l.}\,\psi(z)]}{dz^n} + \frac{d^n[\mathrm{l.}\,\chi(z)]}{dz^n} + \dots$$

Quel que soit cet ordre, cette dérivée, si le produit se réduit à une puissance, est donc simplement égale à n fois la dérivée de l'unique facteur de la puissance. Cette propriété singulière est un des fondements du calcul des probabilités, comme on le reçonnaît en examinant ce que deviennent ces dérivées, alors qu'on y fait disparaître la variable z, qui n'a été introduite que pour porter les exposants.

On peut remarquer, d'abord, que la première dérivée de $\varphi(z)$ est

$$\frac{\varphi'(z)}{\varphi(z)} = \frac{\mathrm{S}.\,b\alpha z^{\alpha-1}}{\mathrm{S}.\,bz^{\alpha}},$$

et que, partant, pour $z = 1$, cette dérivée se réduit à $\mathrm{S}.b\alpha$, c'est-à dire à la moyenne des quantités α. On en conclut que dans le produit P, la première dérivée logarithmique est de même la moyenne $\mathrm{S.B}\sigma$; et comme cette première dérivée est nécessairement la somme des premières dérivées de chacun des facteurs, il en ressort immédiatement que la valeur moyenne des sommes σ des quantités α, β, γ, etc., est la somme des valeurs moyennes de ces quantités, c'est-à-dire qu'on a

$$\mathrm{S.B}\sigma = \mathrm{S}.b\alpha + \mathrm{S}.b\beta + \mathrm{S}.b\gamma + \dots$$

Si respectivement $\alpha_i = \beta_i = \gamma_i = \dots$, il vient

$$\mathrm{S.B}\sigma = n\mathrm{S}.b\alpha.$$

Si $\alpha_i = h_1\varepsilon_i$, $\beta_i = h_2\varepsilon_i$, $\gamma_i = h_3\varepsilon_i$, etc., on obtient alors

$$\mathrm{S.B}\sigma = (h_1 + h_2 + h_3 + \dots)\,\mathrm{S}.b\varepsilon.$$

Ces formules étaient connues. Elles expriment ce qu'on peut appeler la conservation de la moyenne arithmétique, dans les successions d'événements composés d'événements simples soumis à une même loi de probabilité. La moyenne des sommes d'événements est la somme, ou un multiple, des moyennes d'événements simples. Avec d'autres conditions, il y aurait d'autres modes de combinaison des moyennes. Celui-là suffit pour ce dont il s'agit en ce moment.

Mais la moyenne arithmétique n'est point la seule qui se conserve ainsi.

On a visiblement

$$\varphi''(z) = \mathrm{S}.b\alpha(\alpha - 1)z^{\alpha-2} = \mathrm{S}.b\alpha^2 z^{\alpha-2} - \mathrm{S}.b\alpha z^{\alpha-1};$$

par suite, la seconde dérivée logarithmique

$$\frac{\varphi''(z)}{\varphi(z)} - \left[\frac{\varphi'(z)}{\varphi(z)}\right]^2 = \frac{\mathrm{S}.b\alpha^2 z^{\alpha-2} - \mathrm{S}.b\alpha z^{\alpha-1}}{\mathrm{S}.bz^{\alpha}} - \left[\frac{\mathrm{S}.b\alpha z^{\alpha-1}}{\mathrm{S}.bz^{\alpha}}\right]^2;$$

et faisant $z = 1$,

$$\frac{\varphi''(1)}{\varphi(1)} - \left[\frac{\varphi'(1)}{\varphi(1)}\right]^2 = \mathrm{S}.b\alpha^2 - (\mathrm{S}.b\alpha)^2 - \mathrm{S}.b\alpha.$$

Semblablement, la seconde dérivée du produit

$$\mathrm{P} = \mathrm{S}.\mathrm{B}z^{\sigma}$$

sera, pour $z = 1$,

$$\mathrm{S}.\mathrm{B}\sigma^2 - (\mathrm{S}.\mathrm{B}\sigma)^2 - \mathrm{S}.\mathrm{B}\sigma.$$

D'où l'on conclut sur-le-champ, en ayant égard à la valeur de $\mathrm{S}.\mathrm{B}\sigma$,

$$\begin{aligned}\mathrm{S}.\mathrm{B}\sigma^2 - (\mathrm{S}.\mathrm{B}\sigma)^2 &= \mathrm{S}.b\alpha^2 - (\mathrm{S}.b\alpha)^2 \\ &+ \mathrm{S}.b\beta^2 - (\mathrm{S}.b\beta)^2 \\ &+ \mathrm{S}.b\gamma^2 - (\mathrm{S}.b\gamma)^2 + \ldots\end{aligned}$$

Posant

$$\mu_\alpha = \mathrm{S}\,b\alpha, \quad \mu_\beta = \mathrm{S}.b\beta, \quad \mu_\gamma = \mathrm{S}.b\gamma, \ldots, \quad \mu_\sigma = \mathrm{S}.\mathrm{B}\sigma,$$

on peut remarquer que la moyenne des carrés, diminuée du carré de la moyenne, est égale à la moyenne des carrés des différences de toutes les quantités à leur moyenne; de sorte qu'on trouve

$$\mathrm{S}.\mathrm{B}(\sigma - \mu_\sigma)^2 = \mathrm{S}.b(\alpha - \mu_\alpha)^2 + \mathrm{S}.b(\beta - \mu_\beta)^2 + \ldots$$

Pour le cas d'égalité des α, β, γ, etc., il vient donc

$$S.B(\sigma - \mu_\sigma)^2 = nS.b(\alpha - \mu_\alpha)^2;$$

et pour le cas de

$$\alpha_i = h_1 \varepsilon_i, \quad \beta_i = h_2 \varepsilon_i, \ldots,$$

comme alors

$$\mu^\alpha = h_1 \mu, \quad \mu_\beta = h_2 \mu, \ldots,$$

si $\mu = S.b\varepsilon$, on obtient

$$S.B(\sigma - \mu_\sigma)^2 = (h_1^2 + h_2^2 + h_3^2 + \ldots)\,S.b(\varepsilon - \mu)^2.$$

Par conséquent, on voit que la moyenne des carrés des différences entre la moyenne arithmétique et les diverses quantités dont elle se compose se conserve de la même manière que la moyenne. Dans une suite d'événements, elle est un multiple de la moyenne des carrés des écarts des événements simples.

D'après ces résultats, comme Laplace a enseigné à se débarrasser de la moyenne arithmétique, on peut admettre que les quantités α, β, γ, etc., soient déjà diminuées de leurs moyennes dans les polynômes $\varphi(z)$, $\psi(z)$, $\chi(z)$, etc., et alors $\varphi'(z)$, $\psi'(z)$, $\chi'(z)$, etc., se réduiront à zéro pour $z = 1$. Les dérivées logarithmiques se réduiront donc à

$$\varphi''(1), \quad \varphi'''(1), \quad \varphi^{\text{IV}}(1) - 3[\varphi^{\text{V}}(1)]^2, \ldots.$$

Rien n'est plus facile que d'en conclure que

$$\begin{aligned} S.B(\sigma - \mu_\sigma)^3 = {} & S.b(\alpha - \mu_\alpha)^3 \\ & + S.b(\beta - \mu_\beta)^3 \\ & + S.b(\gamma - \mu_\gamma)^3 + \ldots, \end{aligned}$$

ou que la moyenne des cubes des écarts des événements composés est encore la somme des moyennes des cubes des écarts des événements simples. Mais dès la puissance 4^e il n'en est plus ainsi, et l'on trouve, par des calculs analogues aux précédents,

$$\begin{aligned}&\mathrm{S.B}(\sigma-\mu_\sigma)^4-3\left[\mathrm{S.B}(\sigma-\mu_\sigma)^2\right]^2\\ &=\mathrm{S.B}(\alpha-\mu_\alpha)^4-3\left[\mathrm{S.B}(\alpha-\mu_\alpha)^2\right]^2\\ &+\mathrm{S.B}(\beta-\mu_\beta)^4-3\left[\mathrm{S.B}(\beta-\mu_\beta)^2\right]^2+\ldots\end{aligned}$$

Il en résulte que la moyenne des 4[es] puissances ne se conserve pas. En effet, pour avoir cette moyenne $\mathrm{S.B}(\sigma-\mu_\sigma)^4$, il faudra ajouter aux deux membres de cette relation $3\left[\mathrm{S.B}(\sigma-\mu_\sigma)^2\right]^2$, qui sera évidemment de l'ordre de n^2 pour $\alpha=\beta=\gamma=\ldots$, et de l'ordre de $(h_1^2+h_2^2+h_3^2+\ldots)^2$, pour le cas de

$$\alpha_i=h_1\varepsilon_i,\quad \beta_i=h_2\varepsilon_i,\ldots.$$

Partant, ce terme sera d'un ordre bien supérieur à celui des termes du second membre de la relation, quand n, ou le nombre des observations, sera très-grand : ces termes n'atteignent effectivement que l'ordre n, et offrent des signes différents.

On pourra donc poser

$$\mathrm{S.B}(\sigma-\mu_\sigma)^4=n^2\mathrm{C}+n\mathrm{C}_1;$$

et dans l'évaluation de cette expression, ce sera surtout au terme en n^2 qu'il faudra avoir égard, et pour n très-grand, à ce terme presque uniquement.

On reconnaît ainsi que c'est seulement

$$\sqrt{\mathrm{S.B}(\sigma-\mu_\sigma)^4}=n\sqrt{\mathrm{C}+\frac{1}{n}\mathrm{C}'}$$

qui conserve l'ordre du grand nombre n.

De plus, comme

$$n^2\mathrm{C}=3\left[\mathrm{S}.b(\alpha-\mu_\alpha)^2+\mathrm{S}.b(\beta-\mu_\beta)^2+\ldots\right]^2,$$

on trouvera

$$\sqrt{\mathrm{S.B}(\sigma-\mu_\sigma)^4}=\left[\mathrm{S}.b(\alpha-\mu_\alpha)^2+\mathrm{S.B}(\beta-\mu_\beta)^2+\ldots\right]\sqrt{3+\frac{3}{n}\cdot\frac{\mathrm{C}'}{\mathrm{C}}}.$$

Ce qui montre avec évidence comment la considération des 4^{es} puissances ramènerait à discuter simplement la somme ou la moyenne des carrés.

Dans la question qui nous occupe, la somme des cubes, on le voit bien, ne saurait être d'aucune utilité, non plus qu'aucune somme de puissances impaires, parce que ces puissances sont affectées de signes différents dans les termes de côtés opposés de la moyenne; et, par conséquent, les sommes et les moyennes de puissances impaires ne sont réellement que des différences qui ne pourraient servir aux raisonnements qui vont suivre. Il suffit donc de s'arrêter aux puissances paires.

Mais il est bien facile de s'assurer que toute dérivée logarithmique d'ordre pair $2i$ conduira, pour la moyenne des puissances $2i$, à des résultats analogues à celui que donnent les puissances 4^{es}; car il est manifeste que cette dérivée contiendra $[\varphi''(1)]^i$, et que, par suite, $\mathrm{S.B}(\sigma - \mu_\sigma)^{2i}$ contiendra la puissance

$$[\mathrm{S}.b(\alpha - \mu_\alpha)^2 + \mathrm{S}.b(6 - \mu_6)^2 + \ldots]^i,$$

qui sera de l'ordre n^i, ordre supérieur à celui de tous les autres termes. Ainsi ce sera seulement $\sqrt[i]{\mathrm{S.B}(\sigma - \mu_\sigma)^{2i}}$ qui conservera l'ordre de n; et cela uniquement par le grand terme contenant la somme des moyennes des carrés. Il n'y aura donc nul moyen d'éviter la discussion de cette somme.

Ces remarques mettent hors de doute l'erreur commise par M. Gauss, qui, comme je l'ai dit, avait affirmé que le choix de la somme des carrés était arbitraire, et qu'on pourrait à volonté prendre pour mesure des écarts des observations une somme de puissances quelconques (voir *Theoria combinationis observationum minimis erroribus obnoxiæ*). Cela soit dit en passant. Ici les mêmes remarques constatent que la seule moyenne des carrés se conservant dans l'ordre nécessaire, et se représentant dans toutes les sommes de puissances suivantes, il n'y aura pas ouverture à nouvelle discussion.

Maintenant, rien ne sera plus facile que de reconnaître, par l'expression

$$\mathrm{S.B}(\sigma - \mu_\sigma)^2 = (h_1^2 + h_2^2 + h_3^2 + \ldots)\,\mathrm{S.B}(\varepsilon - \mu)^2,$$

qu'il faut rendre un minimum la somme des carrés des facteurs h_i.

Dans la résolution des équations données, ces facteurs sont généralement de l'ordre de $\frac{1}{n}$; et, par suite, la somme de leurs carrés est du même ordre.

Donc la moyenne $S.B(\sigma - \mu_\sigma)^2$ sera d'autant plus rapprochée de zéro (ou du terme milieu du polynôme qui aura zéro pour exposant), que $S.h^2$ sera moindre ou que n sera plus grand. Donc, ce nombre croissant, toute la probabilité dans le produit P se concentrera dans les termes voisins du milieu. Il faut bien qu'il en soit ainsi, puisque le nombre des termes de P est infini, ou, s'il est fini, qu'il est très-grand de l'ordre de n; et que les exposants σ deviennent infinis, ou du moins de ce même ordre très-élevé. Si à une distance un peu considérable du terme milieu, dont la moyenne des carrés se rapproche sans cesse, il subsistait des termes réunissant quelque probabilité, il est manifeste que ce rapprochement deviendrait impossible. Pour n infiniment grand, la moyenne des carrés est infiniment petite; or tous ces termes sont positifs : donc tous ceux dans lesquels entrent des écarts qui ont quelque valeur sensible ne réunissent qu'une probabilité infiniment petite. Il ne reste de probabilité qu'aux écarts les plus voisins de zéro.

Il faut toutefois bien entendre qu'un écart de l'ordre de $\frac{1}{\sqrt{n}}$ sera encore très-voisin de zéro, puisqu'il s'agit de la moyenne des carrés qui est de l'ordre de $\frac{1}{n}$, et qu'ainsi le carré de cet écart sera bien de l'ordre même de cette moyenne.

Bien avant que n soit infini, on conçoit sans doute actuellement d'une manière nette que quand ce nombre deviendra très-grand, les écarts probables seront très-petits, selon que la moyenne de leurs carrés deviendra plus petite. Ils diminueront avec cette moyenne, et réciproquement elle diminuera avec ces écarts. De sorte qu'en rendant cette fonction un minimum, on renfermera l'erreur dans les limites probables les plus étroites. Mais on sait que cette condition du minimum conduit au procédé que Legendre a appelé *méthode des moindres carrés*. C'est donc cette méthode qu'il faut suivre quand on traite des équations plus nombreuses que les inconnues cherchées.

Si ces considérations ne suffisent pas, on peut calculer au moins la forme de la probabilité, et la marche de la grandeur qu'elle prend avec le nombre croissant des observations.

Supposons, par exemple, que l'erreur doive être comprise entre les limites

$$\pm t\sqrt{2 S.b(\varepsilon - \mu)^2}.$$

Après n observations, la moyenne des carrés s'exprimant par

$$S.h^2 \times S.b(\varepsilon - \mu)^2 \quad \text{ou bien par} \quad \frac{k}{n} S.b(\varepsilon - \mu)^2,$$

le nombre k étant très-petit relativement à n, les limites pour rester constantes prendront la forme

$$\pm \frac{t\sqrt{n}}{\sqrt{k}} \sqrt{2\frac{k}{n} S.b(\varepsilon - \mu)^2}.$$

Mais puisque la moyenne des carrés est $\frac{k}{n} S.b(\varepsilon - \mu)^2$, il est clair que les termes de cette moyenne placés au delà des limites ci-dessus ne peuvent en fournir qu'une fraction f. La somme de ces termes est donc

$$S = f \cdot \frac{k}{n} S.b(\varepsilon - \mu)^2.$$

D'un autre côté, p étant la probabilité de tous ces termes au delà des limites, on aura évidemment

$$S = p \times \frac{1}{\theta} \cdot \frac{t^2 n}{k} \cdot 2\frac{k}{n} S.b(\varepsilon - \mu)^2,$$

θ étant inférieur à 1, puisque $t^2 . 2S.b(\varepsilon - \mu)^2$ est le plus petit des écarts qui entrent dans les termes en question.

Égalant ces deux valeurs de S, on obtient

$$p = \frac{2t^2}{\theta f} \cdot \frac{k}{n},$$

et la probabilité que l'erreur tombe entre les limites assignées sera

$$1 - p = 1 - \frac{\theta f}{2t^2} \cdot \frac{k}{n}.$$

Ainsi la probabilité croît constamment avec le nombre des observations. Pour une même probabilité, les limites se resserrent donc, ainsi qu'il a été dit.

L'esprit de l'analyse de Laplace est dès lors facilement saisissable; et si l'on veut calculer avec précision la grandeur de la probabilité obtenue, on peut reprendre cette analyse. On pourra, si on le préfère, adapter aussi à ce calcul celle que M. Cauchy a donnée dans le *Compte rendu* de la séance du 16 août, p. 269 et antérieures.

Mais les considérations précédentes ne laissent aucun doute sur les propriétés de la moyenne des carrés. Comme le raisonnement a été général, il s'ensuit que la fonction de probabilité peut être quelconque.

J'ai hâte d'arriver aux exceptions signalées par M. Cauchy, car elles ne modifient nullement l'opinion de Laplace.

Il semble, au premier abord, qu'il n'ait pu dire *une fonction quelconque*, puisqu'il faut pour la discussion précédente que la moyenne des carrés des écarts, $S.B(\varepsilon - \mu)^2$, soit une quantité finie. Mais qui ne voit que Laplace a exclu toute fonction des erreurs qui n'offrirait pas une valeur finie de la moyenne des carrés de leurs écarts? Il n'avait pas besoin de le dire, puisque cette moyenne entre dans tous ses calculs, et sert de mesure à ce qu'il a nommé le *poids* des résultats. Sans nul doute, il aurait pu ajouter expressément cette exclusion de toute fonction de probabilité incapable de donner une moyenne finie des carrés des écarts. Il est probable qu'il ne l'a point fait, parce qu'il regardait comme évident que dans des observations, je ne dirai pas même précises, mais seulement conduites avec quelque connaissance des instruments, les erreurs sont finies, et, par cela même, toutes les fonctions de probabilités offrent une moyenne finie pour les carrés, et bien plus pour les puissances supérieures des écarts. Ces dernières moyennes sont inutiles, toutefois; elles n'entrent que dans un reste que Laplace néglige; et, dans le passage cité de M. Cauchy, on trouvera, si l'on n'est satisfait des motifs de Laplace, une autre face de la même analyse qui pourra contenter davantage ceux qui aimeront une évaluation analytique. Malheureusement, il ne sera pas plus praticable de déduire une évaluation numérique exacte des formules de M. Cauchy.

Au surplus, qu'on suppose un instant les erreurs sans limites, et, par suite, que la moyenne des carrés n'a point de valeur finie. Les observations mêmes en avertiront l'observateur le moins attentif. Car il faudra que les grandes valeurs des erreurs aient une probabilité notable; et dès lors elles se présenteront, sinon aussi souvent que les autres, du moins en proportion assez grande. Ainsi, l'on aura des observations effroyablement discordantes; et nul doute qu'elles ne soient rejetées, et que les instruments ou les procédés d'observations ne soient soumis à une correction très-fondée.

Qu'on prenne pour exemple la fonction de probabilité

$$f(\varepsilon) = \frac{k}{\pi} \cdot \frac{1}{1 + k^2 \varepsilon^2},$$

indiquée par M. Cauchy, dans la séance du 8 août (p. 206 des *Comptes rendus*). On va reconnaître que l'instrument affecté d'une pareille loi de probabilité ne serait pas même mis en vente par un artiste ordinaire. On ne saurait quel nom donner à l'établissement qui l'aurait construit.

Pour simplifier, je suppose la constante $k = 1$, et l'on a, pour la probabilité d'une erreur numériquement inférieure à c,

$$2\int_0^c d\varepsilon f(\varepsilon) = \frac{2}{\pi}\int_0^c \frac{d\varepsilon}{1+\varepsilon^2} = \frac{2}{\pi}\,\text{arc tang}\,c.$$

De là on tire la petite Table que voici :

Probabilités.	Limite d'erreur.
0,1	0,15838
0,2	0,32492
0,3	0,50953
0,4	0,72654
0,5	1

Probabilités.	Limite d'erreur.
0,6	1,37638
0,7	1,96261
0,8	3,07768
0,9	6,31375
1	∞

On voit, quelque mesure qu'on veuille prendre pour unité d'erreur, qu'il sera tout aussi facile de trouver une erreur six fois plus grande, ou six fois plus petite, que d'en rencontrer une voisine de cette unité. Les grandes erreurs se montreraient donc très-vite. Un calcul simple fait voir qu'il y a 2 contre 1 à parier d'en trouver une sur dix observations, et plus de 7 contre 1 si l'on en exécute seulement vingt.

L'attention la plus ordinaire sera bientôt mise en garde contre un pareil instrument.

Mais il y a plus : ce serait un instrument fort singulier. Il y aurait tout autant de sécurité dans une seule observation que dans la moyenne de 10, de 20, de 1000 ou d'un nombre quelconque d'observations, ou plutôt il y aurait le même danger dans la moyenne de 1000 observations qu'avec une seule. Il est très-facile, en effet, de s'assurer que la loi de probabilité B de la moyenne μ de n observations est exactement la même,

$$f(\mu) = \frac{1}{\pi}\cdot\frac{1}{1+\mu^2},$$

quel que soit n, petit ou grand, que la loi d'une seule erreur. On posséderait alors un instrument avec lequel il n'y aurait qu'une seule observation à exécuter.

On avouera qu'un pareil exemple ne saurait être élevé contre la découverte de Laplace. Certes, le mot *fonction quelconque* ne peut comprendre que des fonctions capables de donner des résultats tant soit peu exacts, et d'autant plus exacts que l'observateur se donne la peine de multiplier ses opérations difficiles.

Aussi M. Poisson, qui a le premier fait remarquer la fonction *arc tang* ε, n'en a-t-il pas eu moins de confiance dans la méthode des moindres carrés. Il se contente de dire qu'on ne rencontre sans doute pas cette hypothèse dans la pratique.

Outre l'exclusion des fonctions de probabilité qui n'ont pas de moyenne finie des carrés des erreurs, et qui seront décelées par les observations, il faut encore, pour que

$$S.B(\sigma - \mu_\sigma)^2 = (h_1^2 + h_2^2 + h_3^2 + \ldots) S.b(\varepsilon - \mu)^2$$

devienne de plus en plus petite avec $\frac{1}{n}$, ou à mesure qu'on fait plus d'observations; il faut visiblement encore que les facteurs h_1, h_2, $h_3, \ldots$, ne forment pas une série décroissante à l'infini. C'est encore à M. Poisson qu'appartient la remarque de ce cas abstrait. Je n'ai publié qu'après lui un extrait de mon travail sur *l'effet de l'intérêt composé* (*Procès-verbaux de la Société Philomathique*, 1839, et journal *l'Institut*, n° 286), où j'ai montré que les facteurs qui multiplient les écarts croissent en progression géométrique, dans la pratique des établissements financiers et du commerce, de manière à exiger la multiplication des affaires dans un temps très-court. Mais on peut mettre en question si une pareille série de facteurs se rencontrera jamais parmi ceux qui se présentent pour résoudre un système d'équations, toutes capables de donner des valeurs des mêmes inconnues. Il n'existera, en effet, d'erreurs que sur les termes tout connus; et il faudrait admettre qu'il y a des équations dans lesquelles ces erreurs sont insignifiantes, pour qu'il y eût lieu de les multiplier par une suite de facteurs qui convergent rapidement. Cependant il est bien clair que, quand même ce cas arriverait à l'improviste, la discussion à laquelle tout observateur doit soumettre ses données l'avertirait sur-le-champ : de même que par l'examen des données financières, j'ai été averti de cette action dévorante de l'intérêt composé et des dépenses qui influent à la manière de l'intérêt composé.

Dans ce cas donc, il ne faudrait pas dire que la méthode fait défaut : car elle ne fera défaut, alors même, qu'à ceux qui l'appliqueront sans la bien connaître, et qui voudront qu'elle les dispense de tout examen. Il faut convenir que c'est là une chose impossible, et que la discussion des observations doit précéder de bien loin l'application de la méthode de Legendre et de Gauss, tant recommandée par Laplace, et par

Bessel, cet observateur consommé, à qui personne ne pourra objecter le défaut de pratique, défaut qui se fait plus d'une fois sentir, même chez Laplace et chez Poisson.

Je crois avoir fourni à l'appui de la découverte de Laplace des raisons qui ne peuvent guère laisser passage aux objections. Il y aurait à donner sur l'emploi de la méthode un grand nombre de détails dans lesquels il serait impossible d'entrer. Je renvoie donc à cet égard aux ouvrages originaux de Gauss et de Laplace.