

JOURNAL  
DE  
MATHÉMATIQUES

PURES ET APPLIQUÉES

FONDÉ EN 1836 ET PUBLIÉ JUSQU'EN 1874

PAR JOSEPH LIOUVILLE

---

I.-J. BIENAYMÉ

**Sur la probabilité des erreurs d'après la méthode des moindres carrés**

*Journal de mathématiques pures et appliquées 1<sup>re</sup> série*, tome 17 (1852), p. 33-78.

[http://www.numdam.org/item?id=JMPA\\_1852\\_1\\_17\\_33\\_0](http://www.numdam.org/item?id=JMPA_1852_1_17_33_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>

---

SUR

# LA PROBABILITÉ DES ERREURS

D'APRÈS

## LA MÉTHODE DES MOINDRES CARRÉS;

**PAR M. I.-J. BIENAYMÉ,**

Inspecteur général des Finances.

Nos sequimur probabilia, nec ultrà quam id, quod  
verisimile occurrerit, progredi possumus; et refellere  
sine pertinacia, et refelli sine iracundia, parati sumus.  
TUSCULAN., 2.

---

§ 1<sup>er</sup>.

La méthode des moindres carrés est si fréquemment employée aujourd'hui dans les sciences d'observation, que tout ce qui peut en rendre les applications plus sûres devient d'un grand intérêt, quelque simple que ce soit d'ailleurs. Cette considération a fait rédiger les recherches suivantes qui ont pour objet de modifier le calcul ordinaire de la probabilité des erreurs, non dans le cas où les observations ne font connaître qu'une seule grandeur, mais dans les cas beaucoup plus multipliés où les observations donnent à la fois plusieurs grandeurs inconnues, liées par des équations avec la grandeur observée. Dans d'autres temps, lorsque la méthode était peu usitée en France, et regardée, à cause des longs calculs qu'elle exige, plutôt comme une curiosité savante que comme un instrument réel de l'observateur, la modification profonde reconnue dans la probabilité avait pu ne paraître que d'une importance secondaire; mais, à présent, il semble vraiment utile de signaler la défectuosité du calcul ordinaire, car elle touche à des travaux plus nombreux chaque jour, et les observateurs, ne pouvant sacrifier un temps précieux à la vérification de théories difficiles, sont obligés d'en accepter les règles

pratiques sur la foi de leurs devanciers, surtout quand ceux-ci sont des hommes de grande et juste autorité dans la science.

Il y aurait peut-être plus d'un défaut à relever dans les applications de la méthode des moindres carrés : il ne sera cependant question ici que de celui qui frappe le plus souvent les yeux, et dont voici l'indication si simple, qu'aux premiers mots tout le monde en reconnaîtra l'existence, bien que les modifications qu'il nécessite puissent imposer un travail analytique assez compliqué.

On sait que la méthode inventée par Legendre, il y a environ cinquante ans, et publiée pour la première fois dans ses *Nouvelles Méthodes pour la détermination des orbites des comètes*, in-4°, 1805, se réduit à multiplier chacune des équations du premier degré formées entre la grandeur observée et les inconnues, par le coefficient de chacune de ces inconnues successivement ; à ajouter les produits donnés par les coefficients de la même inconnue, ce qui fournit autant de nouvelles équations que d'inconnues ; enfin à résoudre ces équations de la manière ordinaire. Les solutions ainsi obtenues jouissent de la propriété de ne renfermer que les moindres erreurs possibles pour une probabilité donnée. Ce n'est pas un minimum absolu, comme on a semblé le croire assez souvent, mais c'est un minimum relatif aux erreurs des observations et au mode choisi pour la combinaison de toutes les équations qu'elles fournissent. Il pourrait se trouver d'autres combinaisons plus avantageuses, et il y aurait à les discuter suivant les cas.

Le calcul de la grandeur des erreurs subsistantes et de la probabilité qu'elles peuvent avoir, se fait suivant les règles ordinaires de la théorie et n'entraîne que des difficultés analytiques.

Quand il n'entre qu'une inconnue dans les équations, ce calcul est exact, du moins quant aux fondements théoriques. Mais quand il y a plusieurs inconnues, les règles données pour calculer l'erreur et la probabilité de chacune d'elles ne fournissent que l'erreur et la probabilité qu'elle pourrait avoir si elle était seule, et quelque grandes que fussent les erreurs des autres.

Or, un des premiers principes de la théorie des probabilités, c'est que, quand plusieurs événements arrivent simultanément, la probabilité du concours de ces événements est le produit des probabilités

de chacun. De sorte que la probabilité de ce concours est inférieure à la probabilité de chaque événement pris à part : elle est d'autant plus petite, qu'il y a plus d'événements.

Évidemment, il en est de même des erreurs de plusieurs inconnues; la probabilité que ces erreurs restent toutes à la fois dans certaines limites ne peut être que le produit des probabilités séparées que chacune ne s'écarte pas de ses limites propres, et, par conséquent, cette probabilité du concours des erreurs de grandeur limitée doit être notablement inférieure à la probabilité des limites de chaque erreur considérée isolément, quelles que puissent être les autres.

C'est donc une défectuosité que d'assigner comme probabilité de l'erreur d'une inconnue faisant partie d'un système à déterminer, celle qu'elle aurait si elle était seule, au lieu de donner des règles pour calculer la probabilité de l'ensemble des erreurs du système qui ne peuvent, en réalité, être isolées les unes des autres.

Ce défaut devient plus évident encore lorsqu'on fait attention que la probabilité relative à une limite d'erreur est d'autant plus grande, que l'étendue bornée par cette limite est plus grande : de sorte que, si l'on veut conserver à l'ensemble des erreurs une probabilité un peu élevée, il faudra assigner à chaque erreur des limites plus écartées que celles qu'exigerait la même probabilité si l'inconnue était seule. Lors donc qu'on s'arrête à ces dernières limites, on n'a réellement qu'une très-faible probabilité qu'elles ne sont pas dépassées. On verra, en effet, dans ce qui suit, qu'il suffit de deux inconnues pour que l'étendue des erreurs soit presque doublée pour une même probabilité. Les règles usuelles font naître par là des idées fort inexactes sur la connaissance des grandeurs inconnues qu'on s'est proposé de déduire des observations, et elles égarent même parfois sur la valeur de ces observations, ainsi que sur le nombre de bonnes observations qu'il serait indispensable de faire pour parvenir à un résultat assigné.

Qu'il soit permis, au sujet de ce nombre, de cette multiplicité des observations, bien entendu de bonnes observations, de faire remarquer que c'est là une condition impérieuse des connaissances humaines quand la précision n'est pas absolue. Car les hommes cherchent à la fois tant de choses, que la probabilité du concours des

valeurs exactes de toutes ces choses ne peut devenir très-grande que par la grandeur immense du nombre des observations. C'est, en effet, cette grandeur et la répétition continuelle des circonstances ordinaires de la vie qui met hors de doute les pratiques du sens commun; et toute l'efficacité d'un sophisme ne consiste le plus souvent qu'à dérober la vue de la multiplicité réelle des faits d'une expérience journalière, en faisant croire à une multiplicité factice d'événements possibles sans doute, mais qui n'arrivent pour ainsi dire jamais.

La défectuosité une fois indiquée, on voit qu'il ne s'agit plus que de modifier convenablement la démonstration de la méthode, afin de calculer, non plus la probabilité de l'erreur d'une inconnue, mais la probabilité de l'ensemble des erreurs. Il y avait ici à choisir parmi les démonstrations.

A proprement parler, Legendre n'en a point donné. Seulement il appuie son procédé assez solidement sur les avantages qu'il fait ressortir, et il insiste principalement sur ce que la moyenne arithmétique des observations, que l'on prend d'ordinaire avec confiance, n'est précisément qu'un cas particulier de la méthode des moindres carrés.

Ce fut M. Gauss qui rattacha le premier cette méthode au calcul des probabilités, quelques années après la publication de Legendre, et ce fut aussi dans un ouvrage d'Astronomie : *Theoria motus corporum cœlestium*, in-4°, 1809. La démonstration que M. Gauss y donne repose sur la réciproque de la remarque de Legendre, que la moyenne arithmétique se déduit de sa méthode, quand cette moyenne peut avoir lieu. Réciproquement, si la moyenne arithmétique adoptée généralement est nécessaire, dit M. Gauss, il en résulte qu'une certaine loi de probabilité est nécessaire, et que le seul procédé à suivre pour combiner les équations fournies par l'observation est le procédé qui rend un minimum la somme des carrés de ces équations. Mais on doit convenir que l'hypothèse de la nécessité de prendre la moyenne arithmétique d'une masse de faits pour obtenir le résultat le plus exact possible, est tout à fait gratuite, et ne saurait pas plus être admise à priori que l'hypothèse même de la nécessité du minimum des carrés. Il n'y a donc là qu'une preuve restreinte aux cas particuliers où la loi de probabilité des erreurs, qui conduit toujours à la

moyenne arithmétique, se rencontre dans les observations; et c'est ce qui arrive très-souvent par la nature des choses. Mais, comme on l'ignore le plus ordinairement, il n'existe pas là de démonstration vraiment solide. Aussi, dans un ouvrage bien plus récent, consacré exclusivement à la méthode des moindres carrés, *Theoria combinationis observationum minimis erroribus obnoxia*, in-4°, 1823, M. Gauss a-t-il fondé sur d'autres considérations l'emploi de cette méthode. Ce ne sont toutefois que des considérations, et non des preuves; et l'on ne trouve de démonstration réelle que dans les travaux antérieurs de Laplace.

C'est dans un Mémoire publié en 1811, que Laplace fit voir que, quand le nombre des observations est assez grand, les erreurs les plus restreintes probablement sont données par la méthode des moindres carrés. Ce Mémoire se trouve parmi ceux de l'Académie des Sciences pour 1811, et il a été reproduit dans la *Théorie analytique des Probabilités*, qui parut en 1812. Les principes de la démonstration de Laplace sont à l'abri de toute objection; les moyens analytiques peuvent seuls en soulever quelques-unes. Ils n'ont point toute la rigueur désirable, et peut-être, dans les applications, conviendrait-il de bien discuter les cas particuliers qui pourraient se présenter. Mais, lorsqu'on réfléchit qu'il suffirait le plus souvent de quelques observations, ajoutées ou retranchées, pour rendre aux expressions analytiques toute leur valeur, on reconnaît que l'analyse de Laplace satisfait complètement aux démonstrations générales. C'est cette analyse qui sera employée dans ce qui va suivre, pour suppléer l'omission de la probabilité du concours des diverses erreurs dans le cas de plusieurs inconnues, qui se présente presque toujours. Seulement il y sera fait quelques simplifications, tirées surtout des beaux travaux de Laplace et de M. Gauss. Si l'un et l'autre ont négligé cette considération du concours des événements, qui accroît si fortement la grandeur des erreurs possibles avec une probabilité déterminée, ils ont du moins fourni tous les moyens de la calculer.

Il aurait été sans doute très-utile d'entrer dans plus de détails pratiques et de donner plusieurs exemples de calculs; mais chaque application de ce genre exigerait un temps assez long.

On verra cependant comment la probabilité de 1 million contre 1

assignée par Laplace aux erreurs possibles de la masse de Jupiter, dont il fixait la limite à  $\frac{1}{100}$  de la valeur, doit être réduite à 1160 contre 1; même en supposant, comme il l'a fait, que le nombre des équations employées, 129 seulement, permît d'appliquer la formule d'approximation, et n'exigeât pas l'addition des termes négligés habituellement. Cette diminution de la probabilité calculée suffit bien, si l'on fait attention aux autres défauts que peuvent renfermer les 129 équations, pour faire cesser la surprise qu'on a dû éprouver quand il a paru nécessaire de modifier de  $\frac{1}{50}$  la masse qui semblait définitivement calculée avec une précision double. Les formules nouvelles mettent donc ici le calcul des probabilités à l'abri des reproches qui ont pu lui être faits un peu précipitamment à cette occasion. Il est très-probable que, dans les autres cas où ce calcul paraît en défaut, on en trouverait également la cause dans quelque omission de l'analyse ou de l'observation.

## § II.

L'application de la méthode des moindres carrés suppose qu'un grand nombre  $n$  d'observations a donné des résultats  $\omega_1, \omega_2, \dots, \omega_h, \dots, \omega_n$  qui auraient pu être calculés d'avance en fonction linéaire de plusieurs éléments  $x_1, x_2, \dots, x_i, \dots, x_m$  en nombre  $m$ , si ces éléments étaient connus. Chaque observation fournit dès lors entre la valeur observée et la valeur calculée correspondante, une équation telle que

$$(1) \quad a_{1,h} x_1 + a_{2,h} x_2 + \dots + a_{i,h} x_i + \dots + a_{m,h} x_m = \omega_h,$$

qui se rapporte à la  $h^{\text{ième}}$  observation. Les coefficients  $a_{i,h}$  sont des quantités connues, indépendantes des éléments  $x_1, x_2$ , etc. Les indices dont ils sont affectés marquent l'inconnue et l'équation auxquelles le coefficient appartient. Ainsi,  $a_{3,7}$  est le coefficient de la troisième inconnue dans la septième équation.

Les équations en nombre  $n$  que représente l'expression (1) ne renfermant que  $m$  inconnues, il suffit qu'il s'en trouve  $m$  qui ne rentrent pas les unes dans les autres, pour qu'on puisse les résoudre par les





sont nullement déterminés par les conditions (3), puisque ces conditions ne sont qu'au nombre de  $m$ .

En prenant un autre système de  $n$  facteurs, qu'on désignerait de même par  $K_{i,h}$ , on obtiendrait semblablement la valeur de la première inconnue  $x_i$ , pourvu qu'on astreignît les  $n$  facteurs  $K_{i,h}$  à  $m$  conditions semblables aux conditions (3) qui se rapportent à l'inconnue  $x_i$ .

Les  $m$  inconnues seront donc fournies par  $m$  systèmes de  $n$  facteurs, déterminés en partie par  $m$  systèmes de conditions tout à fait semblables, et en nombre  $m$  pour chaque élément inconnu. Comme les facteurs d'un système n'entrent pas dans les autres, les  $m^2$  conditions ne détermineront qu'un nombre égal de facteurs sur les  $mn$  employés, et  $n - m$  seront indépendants dans chaque système.

C'est de ces facteurs restés arbitraires qu'on disposera pour rendre les valeurs

$$x_i = S \cdot \omega_h K_{i,h}$$

le plus exactes possible; et c'est ce pour quoi il faut nécessairement recourir au calcul des probabilités, sous quelque forme qu'on le déguise.

Effectivement, la valeur des coefficients  $K_{i,h}$ , quelle qu'on voulût la fixer, ne modifierait en rien la valeur trouvée pour  $x_i$  en fonction de ces arbitraires, si les équations étaient rigoureusement exactes. Mais, puisqu'il s'attache toujours quelque erreur plus ou moins grande au résultat d'une observation, il faudrait, pour avoir des équations rigoureuses, retrancher de toutes les quantités  $\omega_h$  les erreurs respectives  $\varepsilon_h$  dont elles sont affectées. On ne le peut pas, et la somme

$$x_i = S \cdot \omega_h K_{i,h}$$

reste affectée par suite d'une erreur

$$r_i = S \cdot \varepsilon_h K_{i,h}$$

La grandeur de cette erreur dépendra, on le voit, du choix des coefficients  $K$ , et en même temps, de la loi de probabilité des erreurs possibles, qui aura régné dans le cours des observations.

Il y aurait à faire plus d'une remarque sur ce qu'on doit entendre

par cette loi de probabilité, et sur les moyens de suppléer à l'ignorance où l'observateur se trouve souvent à ce sujet. Mais il faut se borner ici au point unique en discussion. La loi des erreurs  $\varepsilon_h$  sera supposée connue.

Chacune des valeurs trouvées pour les inconnues  $x_1, x_2, \dots, x_m$  sera entachée d'une erreur  $r_1, r_2, \dots, r_m$  respectivement, qui se présentera sous la forme qui vient d'être assignée à l'erreur  $r_i$  de l'élément  $x_i$  :

$$(4) \quad \begin{cases} r_1 = S. \varepsilon_h K_{1, h}, \\ r_2 = S. \varepsilon_h K_{2, h}, \\ r_3 = S. \varepsilon_h K_{3, h}, \\ \dots \dots \dots \\ r_m = S. \varepsilon_h K_{m, h}. \end{cases}$$

La grandeur de chacune de ces erreurs résultera de ce qu'il s'est présenté pendant les observations un système d'erreurs  $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots, \varepsilon_h, \dots, \varepsilon_n$  plutôt qu'un autre, et l'on voit que toutes les erreurs  $r$  seront modifiées à la fois par les changements que pourra subir le système des erreurs  $\varepsilon$ . Chacune des erreurs  $\varepsilon$  est inconnue; mais si la loi en est connue, c'est-à-dire si l'on sait qu'à une erreur  $\varepsilon$  répond une certaine probabilité, la probabilité d'un système donné  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  d'erreurs dans l'ensemble des observations sera aussi connue et égale au produit des probabilités de chacune.

Ce produit sera donc la probabilité du système (4) des erreurs  $r_1, r_2, \dots, r_m$ , puisque celles-ci sont déterminées toutes dès que les erreurs  $\varepsilon_h$  de chaque observation reçoivent des valeurs données.

Désignant par  $\varphi_\varepsilon$  la fonction qui exprime la probabilité de l'erreur  $\varepsilon$ , de sorte que  $\varphi_\varepsilon d\varepsilon$  soit cette probabilité infiniment petite, on aura, entre les limites dans lesquelles les erreurs peuvent s'étendre,

$$\int \varphi_\varepsilon d\varepsilon = 1,$$

puisque c'est la somme des probabilités de tous les cas possibles. De plus, si l'on appelle  $\mu_\varepsilon$  la moyenne des puissances  $\varepsilon^6$  de toutes les erreurs possibles, on aura

$$\int \varepsilon^6 d\varepsilon \varphi_\varepsilon = \mu_\varepsilon.$$

Par exemple, la moyenne de toutes les erreurs possibles s'exprimera par

$$\mu_1 = \int \varepsilon d\varepsilon \varphi \varepsilon;$$

la moyenne de leurs carrés par

$$\mu_2 = \int \varepsilon^2 d\varepsilon \varphi \varepsilon, \text{ etc.}$$

La probabilité d'un système de valeurs  $r_1, r_2, \dots, r_m$  étant la probabilité de l'ensemble des valeurs  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  dont elles se composent, s'exprimera par le produit

$$\varphi_{\varepsilon_1} d\varepsilon_1 \times \varphi_{\varepsilon_2} d\varepsilon_2 \times \varphi_{\varepsilon_3} d\varepsilon_3 \times \dots \times \varphi_{\varepsilon_n} d\varepsilon_n.$$

Il ne reste plus qu'à déterminer les systèmes de valeurs des quantités  $r_i$  pour lesquels ce produit est le plus grand, et en dehors des limites desquels il devient dès lors très-peu probable que ces quantités puissent se trouver. L'on y parviendra évidemment si l'on emploie le procédé de Laplace, qui consiste à le multiplier par une fonction des erreurs  $r_1, r_2, \dots, r_m$  et de leurs composantes  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ , telle que l'intégration subséquente, prise dans toute l'étendue de la probabilité, ne laisse subsister que la somme des valeurs correspondantes à certaines grandeurs données des  $r_1, r_2, \dots, r_m$ . Ce procédé, dont Laplace a tant répété l'emploi sans en varier la forme, est au fond le même que Fourier, et plus récemment M. Lejeune-Dirichlet, ont reproduit d'une autre manière et avec un si grand succès. En suivant Laplace, au lieu de former d'abord la fonction limitative nécessaire, on considérera le produit

$$\begin{aligned} P = & \int d\varepsilon_1 \varphi_{\varepsilon_1} \times d\varepsilon_2 \varphi_{\varepsilon_2} \times \dots \times d\varepsilon_h \varphi_{\varepsilon_h} \times \dots \times d\varepsilon_n \varphi_{\varepsilon_n} \\ & \times e^{\alpha_1 \sqrt{-1} (\varepsilon_1 K_{1,1} + \varepsilon_2 K_{1,2} + \varepsilon_3 K_{1,3} + \dots + \varepsilon_n K_{1,n})} \\ & \times e^{\alpha_2 \sqrt{-1} (\varepsilon_1 K_{2,1} + \varepsilon_2 K_{2,2} + \varepsilon_3 K_{2,3} + \dots + \varepsilon_n K_{2,n})} \\ & \dots \dots \dots \\ & \times e^{\alpha_i \sqrt{-1} (\varepsilon_1 K_{i,1} + \varepsilon_2 K_{i,2} + \dots + \varepsilon_n K_{i,n})} \\ & \dots \dots \dots \\ & \times e^{\alpha_m \sqrt{-1} (\varepsilon_1 K_{m,1} + \varepsilon_2 K_{m,2} + \dots + \varepsilon_n K_{m,n})}, \end{aligned}$$

qui contient en exposant les erreurs des  $m$  inconnues

$$r_i = \varepsilon_1 K_{i,1} + \varepsilon_2 K_{i,2} + \dots + \varepsilon_n K_{i,n} = \sum \varepsilon_h K_{i,h},$$

multipliées chacune par un argument spécial  $\alpha_i$ , qui puisse permettre de la distinguer après l'intégration multiple, prise pour toute l'étendue dans laquelle les erreurs  $\varepsilon$  sont possibles. Si, après cette intégration relative aux  $\varepsilon$ , on veut se rendre compte du produit P, il faudra se le représenter comme n'étant plus fonction que des  $r$ , et sous la forme

$$P = \int_m dr_1 dr_2 \dots dr_m \Phi(r_1, r_2, \dots, r_m) e^{r_1 \alpha_1 \sqrt{-1} + r_2 \alpha_2 \sqrt{-1} + \dots + r_m \alpha_m \sqrt{-1}},$$

expression dans laquelle la fonction  $\Phi$  sera telle, qu'elle réunira les probabilités relatives à des valeurs égales du système des erreurs  $r$ . C'est réellement cette fonction, multipliée par le produit différentiel  $dr_1 dr_2 \dots dr_m$ , qui est la probabilité de ce système d'erreurs, et, si on la connaissait, il n'y aurait plus qu'à intégrer dans des limites convenables pour obtenir la probabilité cherchée. Pour déterminer la fonction  $\Phi$ , il suffit de multiplier P successivement par une série de  $m$  intégrales de la forme

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha_i e^{-r_i \alpha_i \sqrt{-1}},$$

car on sait que

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha_i e^{-r_i \alpha_i \sqrt{-1}} \int du \Phi(u) e^{z_i u \sqrt{-1}} = \Phi(r_i).$$

Si donc on répète cette opération presque mécanique pour tous les arguments  $\alpha_i$  et les erreurs  $r_i$ , en nombre  $m$ , on obtiendra la loi de probabilité de ces erreurs. Ce sera le résultat Q que voici :

$$Q = \frac{1}{(2\pi)^m} \int_m d\alpha_1 d\alpha_2 \dots d\alpha_m e^{-r_1 \alpha_1 \sqrt{-1} - r_2 \alpha_2 \sqrt{-1} - \dots - r_m \alpha_m \sqrt{-1}} \times P.$$

Partant,  $Q dr_1 dr_2 \dots dr_m$  sera la probabilité du système  $r_1, r_2, \dots, r_m$  des erreurs des inconnues  $x_1, x_2, \dots, x_m$ ; et intégrant entre les limites qui renferment les grandeurs de ces erreurs dont on voudra

connaître la probabilité commune, on obtiendra facilement cette probabilité, qui sera

$$p = \int_m dr_1 dr_2 \dots dr_m Q.$$

On peut aisément reconnaître, si l'on veut, la composition de la fonction limitative qui n'a été introduite que successivement, et il serait superflu de s'y arrêter. Il reste à poursuivre ces intégrations successives.

A cet effet, on modifie le produit P en réunissant dans l'exposant tous les termes qui se rapportent à une même observation, et séparant les intégrales relatives aux  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ , on a

$$\begin{aligned}
 P = & \int d\varepsilon_1 \varphi_{\varepsilon_1} e^{\varepsilon_1 (\alpha_1 K_{1,1} + \alpha_2 K_{2,1} + \alpha_3 K_{3,1} + \dots + \alpha_m K_{m,1})} \sqrt{-1} \\
 & \times \int d\varepsilon_2 \varphi_{\varepsilon_2} e^{\varepsilon_2 (\alpha_1 K_{1,2} + \alpha_2 K_{2,2} + \alpha_3 K_{3,2} + \dots + \alpha_m K_{m,2})} \sqrt{-1} \\
 & \dots \dots \dots \\
 & \times \int d\varepsilon_h \varphi_{\varepsilon_h} e^{\varepsilon_h (\alpha_1 K_{1,h} + \alpha_2 K_{2,h} + \alpha_3 K_{3,h} + \dots + \alpha_m K_{m,h})} \sqrt{-1} \\
 & \dots \dots \dots \\
 & \times \int d\varepsilon_n \varphi_{\varepsilon_n} e^{\varepsilon_n (\alpha_1 K_{1,n} + \alpha_2 K_{2,n} + \alpha_3 K_{3,n} + \dots + \alpha_m K_{m,n})} \sqrt{-1}.
 \end{aligned}$$

Isolant une de ces  $n$  intégrales, on peut l'écrire

$$\int d\varepsilon_h \varphi_{\varepsilon_h} e^{\varepsilon_h \sqrt{-1} \sum_i \alpha_i K_{i,h}},$$

et même, afin d'abrégier, on peut représenter  $\sum \alpha_i K_{i,h}$ , somme composée de  $m$  termes, par  $S_h$ , au moins provisoirement; on peut aussi supprimer l'indice  $h$  de  $\varepsilon$ , qui n'a été employé que pour la clarté. On aura, en développant l'exponentielle,

$$\int d\varepsilon \varphi_{\varepsilon} e^{\varepsilon S_h \sqrt{-1}} = \int d\varepsilon \varphi_{\varepsilon} \left( 1 + \varepsilon S_h \sqrt{-1} - \frac{\varepsilon^2 S_h^2}{2} - \frac{\varepsilon^3 S_h^3}{6} \sqrt{-1} + \frac{\varepsilon^4 S_h^4}{24} + \dots \right),$$

et, en intégrant relativement à  $\varepsilon$ , qui est partout explicite, on aura

$$1 + \mu_1 S_h \sqrt{-1} - \frac{\mu_2 S_h^2}{2} - \frac{\mu_3 S_h^3}{6} \sqrt{-1} + \frac{\mu_4 S_h^4}{24} + \dots,$$

à cause de

$$\int d\varepsilon \varphi \varepsilon = 1 \quad \text{et} \quad \int \varepsilon^6 d\varepsilon \varphi \varepsilon = \mu_6.$$

On pourra ensuite rétablir cette suite en exponentielle, ce qui donne

$$e^{\mu_1 S_h \sqrt{-1} - \frac{S_h^2}{2} (\mu_2 - \mu_1^2) - \frac{S_h^3}{6} (\mu_3 - 3\mu_2 \mu_1 + 2\mu_1^3) \sqrt{-1} + \frac{S_h^4}{24} (\mu_4 - 4\mu_3 \mu_1 - 3\mu_2^2 + 12\mu_2 \mu_1^2 - 6\mu_1^4) + \dots}$$

Rien ne serait plus facile, on le voit, que de supposer la loi de probabilité  $\varphi \varepsilon$  variable d'une observation à l'autre, et de conserver cette distinction dans le reste des calculs. Mais on reconnaîtra bientôt que la méthode des moindres carrés repose sur cette hypothèse, que les moyennes  $\mu_1$  et  $\mu_2$  des erreurs  $\varepsilon$  possibles et de leurs carrés, ne varient pas d'une observation à l'autre; de sorte que la variation de la loi de probabilité, réduite aux moyennes des puissances supérieures, aurait peu d'intérêt dans la question actuelle. Il n'en sera donc point tenu compte.

Dans cette hypothèse, le produit P étant formé d'intégrales toutes semblables, se présentera sous la forme

$$P = e^{\mu_1 (S_1 + S_2 + \dots + S_n) \sqrt{-1} - \frac{\mu_2 - \mu_1^2}{2} (S_1^2 + S_2^2 + \dots + S_n^2) - \dots},$$

les troisième et quatrième termes de l'exposant étant de même

$$- \frac{\mu_3 - 3\mu_2 \mu_1 + 2\mu_1^3}{6} (S_1^3 + S_2^3 + \dots + S_n^3) \sqrt{-1} \\ + \frac{\mu_4 - 4\mu_3 \mu_1 - 3\mu_2^2 + 12\mu_2 \mu_1^2 - 6\mu_1^4}{24} (S_1^4 + S_2^4 + \dots + S_n^4).$$

Si l'on examine à présent chacune des suites qui entrent dans les différents termes, on verra sans peine que chaque variable  $\alpha_i$  est multipliée dans l'une d'elles par tous les facteurs  $K_{i,h}$  correspondant à la même inconnue  $x_i$  ou à la même erreur  $r_i$ . Ils pourront donc être réunis sous des formes symétriques, et, pour plus de clarté, il sera loisible d'attribuer aux sommes qui comprennent  $n$  termes le signe S plus spécialement, et à celles qui n'en renfermeront que  $m$ , le signe  $\Sigma$ .

On peut observer que les premières sont celles où l'ordre des observations en nombre  $n$  amène autant de lettres différentes; et que les secondes sont celles qui, dépendant de l'ordre des inconnues  $x_i$  en nombre  $m$ , ne sauraient comprendre que ce dernier nombre de lettres.

On a ainsi, comme précédemment,

$$S_h = \alpha_1 K_{1,h} + \alpha_2 K_{2,h} + \dots + \alpha_i K_{i,h} + \dots \\ + \alpha_m K_{m,h} = \Sigma \alpha_i K_{i,h},$$

où la somme  $\Sigma$  est relative à l'indice  $i$ , qui lui-même ne varie que de 1 à  $m$ , nombre des inconnues. Il en résulte, pour la somme des premières puissances des  $S_h$ ,

$$S_1 + S_2 + \dots + S_h + \dots + S_n = \Sigma \alpha_i K_{i,1} + \Sigma \alpha_i K_{i,2} + \dots \\ + \Sigma \alpha_i K_{i,h} + \dots + \Sigma \alpha_i K_{i,n},$$

qu'on pourrait écrire  $S(\Sigma \alpha_i K_{i,h})$ , la somme finie  $S$  se rapportant à l'indice  $h$ .

Il est bien aisé de voir que

$$S(\Sigma \alpha_i K_{i,h}) = \Sigma \alpha_i (K_{i,1} + K_{i,2} + \dots + K_{i,h} + \dots + K_{i,n}) \\ = \alpha_1 S K_{1,h} + \alpha_2 S K_{2,h} + \dots + \alpha_m S K_{m,h} \\ = \Sigma (\alpha_i S K_{i,h}).$$

De même, à cause de

$$S_h^2 = (\alpha_1 K_{1,h} + \alpha_2 K_{2,h} + \dots + \alpha_m K_{m,h})^2 = (\Sigma \alpha_i K_{i,h})^2 \\ = \alpha_1^2 K_{1,h}^2 + \alpha_2^2 K_{2,h}^2 + \dots + \alpha_m^2 K_{m,h}^2 \\ + 2 \alpha_1 K_{1,h} (\alpha_2 K_{2,h} + \alpha_3 K_{3,h} + \dots + \alpha_m K_{m,h}) \\ + 2 \alpha_2 K_{2,h} (\alpha_3 K_{3,h} + \dots + \alpha_m K_{m,h}) \\ \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \\ + 2 \alpha_{m-1} K_{m-1,h} \times \alpha_m K_{m,h},$$

on a

$$\begin{aligned}
 & S_1^2 + S_2^2 + \dots + S_h^2 + \dots + S_n^2 \\
 &= \alpha_1^2 (K_{1,1}^2 + K_{1,2}^2 + K_{1,3}^2 + \dots + K_{1,h}^2 + \dots + K_{1,n}^2) \\
 &+ \alpha_2^2 (K_{2,1}^2 + K_{2,2}^2 + K_{2,3}^2 + \dots + K_{2,h}^2 + \dots + K_{2,n}^2) \\
 &\dots \dots \dots \\
 &+ \alpha^2 (K_{m,1}^2 + K_{m,2}^2 + \dots + K_{m,h}^2 + \dots + K_{m,n}^2) \\
 &+ 2\alpha_1 \alpha_2 (K_{1,1} K_{2,1} + K_{1,2} K_{2,2} + \dots + K_{1,h} K_{2,h} + \dots + K_{1,n} K_{2,n}) \\
 &+ 2\alpha_1 \alpha_3 (K_{1,1} K_{3,1} + K_{1,2} K_{3,2} + \dots + K_{1,h} K_{3,h} + \dots + K_{1,n} K_{3,n}) \\
 &\dots \dots \dots \\
 &+ 2\alpha_i \alpha_{i'} (K_{i,1} K_{i',1} + K_{i,2} K_{i',2} + \dots + K_{i,h} K_{i',h} + \dots + K_{i,n} K_{i',n}) \\
 &\dots \dots \dots \\
 &+ 2\alpha_{m-1} \alpha_m (K_{m-1,1} K_{m,1} + \dots + K_{m-1,h} K_{m,h} + \dots + K_{m-1,n} K_{m,n}),
 \end{aligned}$$

ou, en abrégant,

$$\begin{aligned}
 &= \alpha_1^2 SK_{1,h}^2 + \alpha_2^2 SK_{2,h}^2 + \alpha_3^2 SK_{3,h}^2 + \dots + \alpha_m^2 SK_{m,h}^2 \\
 &+ 2\alpha_1 (\alpha_2 SK_{1,h} K_{2,h} + \alpha_3 SK_{1,h} K_{3,h} + \dots + \alpha_m SK_{1,h} K_{m,h}) \\
 &+ 2\alpha_2 (\alpha_3 SK_{2,h} K_{3,h} + \dots + \alpha_m SK_{2,h} K_{m,h}) \\
 &\dots \dots \dots \\
 &+ 2\alpha_{m-1} \alpha_m SK_{m-1,h} K_{m,h}.
 \end{aligned}$$

Ici les sommes des produits ont pu être aussi facilement désignées que les sommes de puissances, puisqu'il n'y a dans chaque somme qu'une seule variable. On abrégait encore davantage en écrivant

$$\Sigma(\alpha_i^2 SK_{i,h}^2) + 2\Sigma(\alpha_i \alpha_{i'} SK_{i,h} K_{i',h}),$$

en ayant soin, pour former la seconde somme  $\Sigma$ , de ne prendre les indices  $i$  et  $i'$  que sous la condition  $i < i'$ , ou bien d'effacer le coefficient 2 qui la précède.

Passant aux  $S_h^3$ , il est visible que, suivant des remarques analogues,

$$\begin{aligned}
 & S_1^3 + S_2^3 + \dots + S_h^3 + \dots + S_n^3 \\
 &= \alpha_1^3 SK_{1,h}^3 + \alpha_2^3 SK_{2,h}^3 + \dots + \alpha_m^3 SK_{m,h}^3 \\
 &+ 3\alpha_1^2 (\alpha_2 SK_{1,h}^2 K_{2,h} + \alpha_3 SK_{1,h}^2 K_{3,h} + \dots + \alpha_m SK_{1,h}^2 K_{m,h}) \\
 &+ 3\alpha_2^2 (\alpha_1 SK_{2,h}^2 K_{1,h} + \alpha_3 SK_{2,h}^2 K_{3,h} + \dots + \alpha_m SK_{2,h}^2 K_{m,h}) \\
 &\dots \dots \dots \\
 &+ 6\alpha_1 \alpha_2 \alpha_3 SK_{1,h} K_{2,h} K_{3,h} + \dots + 6\alpha_i \alpha_{i'} \alpha_{i''} SK_{i,h} K_{i',h} K_{i'',h} + \dots \\
 &\quad \quad \quad + 6\alpha_{m-2} \alpha_{m-1} \alpha_m SK_{m-2,h} K_{m-1,h} K_{m,h} \\
 &= \Sigma(\alpha_i^3 SK_{i,h}^3) + 3\Sigma(\alpha_i^2 \alpha_{i'} SK_{i,h}^2 K_{i',h}) + 6\Sigma(\alpha_i \alpha_{i'} \alpha_{i''} SK_{i,h} K_{i',h} K_{i'',h}).
 \end{aligned}$$



Il est presque inutile de dire qu'il faut entendre par les sommes  $\Sigma$  relatives aux indices  $i$ , lorsqu'elles s'appliquent aux produits, toutes les combinaisons possibles, sans double emploi.

Arrivant enfin aux  $S_h^4$ , on aura absolument de la même manière,

$$\begin{aligned} & S_1^4 + S_2^4 + \dots + S_h^4 + \dots + S_m^4 = S(\Sigma \alpha_i K_{i,h})^4 \\ = & S \left( \begin{aligned} & \Sigma \alpha_i^4 K_{i,h}^4 + 4 \Sigma \alpha_i^3 \alpha_{i'} K_{i,h}^3 K_{i',h} + 6 \Sigma \alpha_i^2 \alpha_{i'}^2 K_{i,h}^2 K_{i',h}^2 \\ & + 12 \Sigma \alpha_i^2 \alpha_{i'} \alpha_{i''} K_{i,h}^2 K_{i',h} K_{i'',h} \\ & + 24 \Sigma \alpha_i \alpha_{i'} \alpha_{i''} \alpha_{i'''} K_{i,h} K_{i',h} K_{i'',h} K_{i''',h} \end{aligned} \right) \\ = & \Sigma (\alpha_i^4 SK_{i,h}^4) + 4 \Sigma (\alpha_i^3 \alpha_{i'} SK_{i,h}^3 K_{i',h}) + 6 \Sigma (\alpha_i^2 \alpha_{i'}^2 SK_{i,h}^2 K_{i',h}^2) \\ & + 12 \Sigma (\alpha_i^2 \alpha_{i'} \alpha_{i''} SK_{i,h}^2 K_{i',h} K_{i'',h}) \\ & + 24 \Sigma (\alpha_i \alpha_{i'} \alpha_{i''} \alpha_{i'''} SK_{i,h} K_{i',h} K_{i'',h} K_{i''',h}). \end{aligned}$$

Il était utile de montrer les formes des troisième et quatrième termes de l'exposant, afin qu'on pût reconnaître pour quels motifs très-différents ils sont négligés dans les applications.

En désignant par  $T_1, T_2, T_3, T_4$  les termes de l'exposant de  $e$ , qui viennent d'être développés, le produit  $P$  peut s'écrire

$$\begin{aligned} P &= e^{T_1 \sqrt{-1} - \frac{1}{2} T_2 - \frac{1}{6} T_3 \sqrt{-1} + \frac{1}{24} T_4 + \dots} \\ &= e^{T_1 \sqrt{-1} - \frac{1}{2} T_2} \left( 1 - \frac{1}{6} T_3 \sqrt{-1} + \frac{1}{24} T_4 + \dots - \frac{1}{72} T_3^2 + \dots \right), \end{aligned}$$

en s'arrêtant aux termes du quatrième degré en  $\alpha_i$ , comme précédemment.

Il ne restera donc plus à intégrer que relativement à ces variables, pour obtenir la fonction  $Q$ , qui se présente sous la forme

$$\begin{aligned} Q &= \frac{1}{(2\pi)^m} \int_{-\infty}^{\infty} d\alpha_1 d\alpha_2 \dots d\alpha_m e^{-r_1 \alpha_1 \sqrt{-1} - r_2 \alpha_2 \sqrt{-1} - \dots - r_m \alpha_m \sqrt{-1}} \times P \\ &= \frac{1}{(2\pi)^m} \int_{-\infty}^{\infty} d\alpha_1 d\alpha_2 \dots d\alpha_m e^{-\Sigma r_i \alpha_i \sqrt{-1} + \sqrt{-1} \mu_1 \Sigma (\alpha_i SK_{i,h}) - \frac{1}{2} T_2 \left( 1 - \frac{\sqrt{-1}}{6} T_3 - \dots \right)}. \end{aligned}$$

Il est visible que les premières puissances des  $\alpha$  peuvent se mettre

en facteur commun, et que l'exposant de  $e$  est égal à

$$- \alpha_1 \sqrt{-1} (r_1 - \mu_1 \text{SK}_{1,h}) - \alpha_2 \sqrt{-1} (r_2 - \mu_1 \text{SK}_{2,h}) - \dots \\ - \alpha_m \sqrt{-1} (r_m - \mu_1 \text{SK}_{m,h}) - \frac{1}{2} T_2.$$

Laplace et plusieurs autres géomètres ont supposé la moyenne  $\mu_1$  réduite à zéro. Mais cette condition simplifie très-peu les calculs, et elle oblige en quelque sorte à les recommencer quand cette moyenne ne peut être considérée comme nulle. On allègue, à la vérité, que, dans un système d'observations bien dirigées, cette moyenne  $\mu_1$ , qui est une véritable erreur constante, a dû être reconnue et retranchée des valeurs observées. N'est-il pas possible, au contraire, qu'on ne fasse les observations que pour déterminer les erreurs constantes, et qu'on doive conserver  $\mu_1$ , qui est alors la chose cherchée. Ici, pour conserver la facilité de reprendre cette moyenne, bien que la recherche de sa valeur, d'après les observations mêmes, ne doive pas faire un sujet d'examen, il suffira de réduire à une seule lettre les termes de la forme  $r_i - \mu_1 \text{SK}_{i,h}$ . On introduira ainsi au lieu des erreurs  $r$  d'autres quantités qui n'en différeront que de quantités constantes. On peut écrire, par exemple,

$$(5) \quad r_i - \mu_1 \text{SK}_{i,h} = \rho_i \sqrt{2(\mu_2 - \mu_1^2)};$$

de sorte qu'une erreur  $r_i$  se calculera par la relation

$$r_i = \mu_1 \text{SK}_{i,h} + \rho_i \sqrt{2(\mu_2 - \mu_1^2)},$$

et

$$dr_i = d\rho_i \sqrt{2(\mu_2 - \mu_1^2)}.$$

De cette manière, on voit que

$$Q = \frac{1}{(2\pi)^m} \int_{-\infty}^{\infty} d\alpha_1 d\alpha_2 \dots d\alpha_m e^{-\sqrt{-1} \sqrt{2(\mu_2 - \mu_1^2)} \sum \rho_i \alpha_i - \frac{1}{2} T_2} \left( 1 - \frac{\sqrt{-1}}{6} T_3 + \dots \right).$$

Si l'on fait subir aux  $\alpha_i$  un changement analogue, c'est-à-dire qu'on pose

$$\alpha_i = z_i \frac{1}{\sqrt{\frac{1}{2}(\mu_2 - \mu_1^2)}}, \quad d\alpha_i = \frac{dz_i}{\sqrt{\frac{1}{2}(\mu_2 - \mu_1^2)}}$$

il en résultera

$$dr_i d\alpha_i = d\rho_i \sqrt{2(\mu_2 - \mu_1^2)} \times \frac{dz_i}{\sqrt{\frac{1}{2}(\mu_2 - \mu_1^2)}} = 2 d\rho_i dz_i.$$

De plus, le facteur  $\frac{1}{2}(\mu_2 - \mu_1^2)$  disparaîtra du terme  $\frac{1}{2} T_2$ ; de sorte que la probabilité du système d'erreurs désignées par  $\rho_i$  ou par

$$r_i = \mu_1 SK_{i,h} + \rho_i \sqrt{2(\mu_2 - \mu_1^2)},$$

qui est égale à

$$p = \int_m dr_1 dr_2 \dots dr_m Q,$$

se calculera par la formule

$$p = \frac{1}{\pi^m} \int d\rho_1 d\rho_2 \dots d\rho_m \int_{-\infty}^{\infty} dz_1 dz_2 \dots dz_m e^{-2\sqrt{-1} \sum \rho_i z_i - Z_2} \left( 1 - \frac{\sqrt{-1}}{6} Z_1 + \frac{1}{24} Z_1 - \frac{1}{72} Z_1^2 \right),$$

en écrivant  $Z_2, Z_3, Z_4$ , pour abrégé, au lieu des termes suivants que

donne le changement des  $\alpha_i$  en  $\frac{z_i}{\sqrt{\frac{1}{2}(\mu_2 - \mu_1^2)}}$ . On avait

$$\frac{1}{2} T_2 = \frac{\mu_2 - \mu_1^2}{2} [\Sigma(\alpha_i^2 SK_{i,h}^2) + 2 \Sigma(\alpha_i \alpha_{i'} SK_{i,h} K_{i',h})];$$

et l'on aura plus simplement

$$Z_2 = \Sigma(z_i^2 SK_{i,h}^2) + 2 \Sigma(z_i z_{i'} SK_{i,h} K_{i',h}).$$

De même,

$$(6) \left\{ \begin{aligned} Z_3 &= 2^{\frac{3}{2}} \frac{\mu_2 - 3\mu_1\mu_1 + \mu_1^3}{(\mu_2 - \mu_1^2)^{\frac{3}{2}}} \left\{ \Sigma(z_i^3 SK_{i,h}^3) + 3 \Sigma(z_i^2 z_{i'} SK_{i,h}^2 K_{i',h}) \right\} \\ &\quad + 6 \Sigma(z_i z_{i'} z_{i''} SK_{i,h} K_{i',h} K_{i'',h}) \\ Z_4 &= 4 \frac{\mu_2 - 4\mu_1\mu_1 - 3\mu_1^2 + 12\mu_1\mu_1^2 - 6\mu_1^3}{(\mu_2 - \mu_1^2)^2} \left\{ \Sigma(z_i^4 SK_{i,h}^4) + 4 \Sigma(z_i^3 z_{i'} SK_{i,h}^3 K_{i',h}) \right\} \\ &\quad + 6 \Sigma(z_i^2 z_{i'}^2 SK_{i,h}^2 K_{i',h}^2) \\ &\quad + 12 \Sigma(z_i z_{i'} z_{i''} SK_{i,h} K_{i',h} K_{i'',h}) \\ &\quad + 24 \Sigma(z_i z_{i'} z_{i''} z_{i'''} SK_{i,h} K_{i',h} K_{i'',h} K_{i''',h}) \end{aligned} \right.$$

Il est bien aisé de remarquer que les sommes désignées par S contenant  $n$  termes seront de la forme  $nM$ ,  $M$  étant une valeur commune

de ces termes. De sorte qu'il suffira, pour rendre très-petites les parties qui les contiennent, de donner aux variables  $z$  des dénominateurs dans lesquels se trouvent  $n$  à une puissance supérieure à la première, et une valeur commune  $M$  à peu près équivalente. On sait que c'est là ce qui arrive dans l'analyse de Laplace.

Pour abrégier encore l'écriture de l'exposant de  $e$  et pour le développer, on pourra mettre indifféremment

$$b_{i,i'} = SK_{i,h} K_{i',h} = b_{i',i},$$

ce qui entraîne

$$b_{i,i} = SK_{i,h}^2.$$

De cette manière, l'exposant sera

$$(7) \quad \left\{ \begin{array}{l} - 2\rho_1 z_1 \sqrt{-1} - 2\rho_2 z_2 \sqrt{-1} - \dots - 2\rho_i z_i \sqrt{-1} - \dots - 2\rho_m z_m \sqrt{-1} \\ - z_1^2 b_{1,1} - 2z_1(z_2 b_{1,2} + z_3 b_{1,3} + \dots + z_i b_{1,i} + \dots + z_m b_{1,m}) \\ \quad - z_2^2 b_{2,2} - 2z_2(z_3 b_{2,3} + \dots + z_i b_{2,i} + \dots + z_m b_{2,m}) \\ \quad \dots \\ \quad \quad - z_i^2 b_{i,i} - 2z_i(z_{i+1} b_{i,i+1} + \dots + z_m b_{i,m}) \\ \quad \quad \quad \dots \\ \quad \quad \quad \quad - z_m^2 b_{m,m}. \end{array} \right.$$

Si maintenant on fait à la fois

$$(8) \quad \left\{ \begin{array}{l} \xi_1 = z_1 h_{1,1} + z_2 h_{1,2} + z_3 h_{1,3} + \dots + z_i h_{1,i} + \dots + z_m h_{1,m} + t_1 \sqrt{-1}, \\ \xi_2 = \quad \quad + z_2 h_{2,2} + z_3 h_{2,3} + \dots + z_i h_{2,i} + \dots + z_m h_{2,m} + t_2 \sqrt{-1}, \\ \quad \dots \\ \xi_i = \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad + z_i h_{i,i} + \dots + z_m h_{i,m} + t_i \sqrt{-1}, \\ \quad \dots \\ \xi_m = \quad + z_m h_{m,m} + t_m \sqrt{-1}, \end{array} \right.$$

et qu'après avoir élevé au carré chacune des nouvelles variables  $\xi_i$ , on en fasse la somme, il ne s'y trouvera que quatre espèces de termes :

En  $z_i^2$  dont le coefficient sera

$$h_{1,i}^2 + h_{2,i}^2 + \dots + h_{i-1,i}^2 + h_{i,i}^2,$$

et contiendra  $i$  termes;

En  $2 z_i z_{i'}$  dont le coefficient sera de même

$$h_{1,i} h_{1,i'} + h_{2,i} h_{2,i'} + \dots + h_{i-1,i} h_{i-1,i'} + h_{i,i} h_{i,i'}$$

contenant  $i$  termes;

En  $-2 z_i \sqrt{-1}$  multiplié par l'expression variable

$$t_1 h_{1,i} + t_2 h_{2,i} + \dots + t_i h_{i,i}$$

contenant encore  $i$  termes;

Et, enfin, les  $m$  carrés des  $t_i$  pris négativement.

Si donc on fait les coefficients des  $z_i^2$  égaux à ceux que ces variables ont dans l'exposant de  $e$ , et qu'on agisse de même pour les coefficients des doubles produits; ce qui donnera

$$(9) \left\{ \begin{array}{l} b_{1,1} = h_{1,1}^2, \quad b_{1,2} = h_{1,1} h_{1,2}, \quad b_{1,3} = h_{1,1} h_{1,3}, \\ \quad \quad \quad b_{2,2} = h_{1,2}^2 + h_{2,2}^2, \quad b_{2,3} = h_{1,2} h_{1,3} + h_{2,2} h_{2,3}, \\ \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad b_{3,3} = h_{1,3}^2 + h_{2,3}^2 + h_{3,3}^2, \\ \\ b_{1,4} = h_{1,1} h_{1,4}, \\ b_{2,4} = h_{1,2} h_{1,4} + h_{2,2} h_{2,4}, \\ b_{3,4} = h_{1,3} h_{1,4} + h_{2,3} h_{2,4} + h_{3,3} h_{3,4}, \\ b_{4,4} = h_{1,4}^2 + h_{2,4}^2 + h_{3,4}^2 + h_{4,4}^2, \\ \\ b_{1,i} = h_{1,1} h_{1,i}, \\ b_{2,i} = h_{1,2} h_{1,i} + h_{2,2} h_{2,i}, \\ b_{3,i} = h_{1,3} h_{1,i} + h_{2,3} h_{2,i} + h_{3,3} h_{3,i}, \\ \dots \dots \dots \\ b_{i,i'} = h_{1,i} h_{1,i'} + h_{2,i} h_{2,i'} + h_{3,i} h_{3,i'} + \dots + h_{i',i'} h_{i',i}, \\ \dots \dots \dots \\ b_{i,i} = h_{1,i}^2 + h_{2,i}^2 + h_{3,i}^2 + \dots + h_{i,i}^2 + \dots + h_{i,i} \end{array} \right.$$

relations dont la loi est facile à saisir; puis, qu'on assujettisse les nouvelles variables  $t_i$ , qui seules multiplient les premières puissances

des  $z_i$ , aux conditions

$$(10) \left\{ \begin{array}{l} \rho_1 = t_1 h_{1,1}, \\ \rho_2 = t_1 h_{1,2} + t_2 h_{2,2}, \\ \rho_3 = t_1 h_{1,3} + t_2 h_{2,3} + t_3 h_{3,3}, \\ \dots \\ \rho_i = t_1 h_{1,i} + t_2 h_{2,i} + t_3 h_{3,i} + \dots + t_i h_{i,i}, \\ \dots \\ \rho_m = t_1 h_{1,m} + t_2 h_{2,m} + t_3 h_{3,m} + \dots + t_m h_{m,m}; \end{array} \right.$$

il est manifeste que la somme des carrés des variables  $\xi_i$  sera, sauf le signe, égale à la fonction des  $z_i$  et des  $\rho_i$  qui formait l'exposant (7), moins la somme des carrés des variables  $t_i$ . Ainsi l'exposant pourra être remplacé par la somme négative des carrés des variables nouvelles  $\xi_i$  et  $t_i$ .

D'un autre côté, les relations (8) et (10) assurent immédiatement

$$\begin{aligned} d\xi_1 d\xi_2 \dots d\xi_m &= dz_1 dz_2 \dots dz_m h_{1,1} h_{2,2} \dots h_{m,m}, \\ d\rho_1 d\rho_2 \dots d\rho_m &= dt_1 dt_2 \dots dt_m h_{1,1} h_{2,2} \dots h_{m,m}; \end{aligned}$$

de sorte que

$$d\rho_1 d\rho_2 d\rho_m \times dz_1 dz_2 \dots dz_m = dt_1 dt_2 \dots dt_m \times d\xi_1 d\xi_2 \dots d\xi_m.$$

Partant, les variables se trouvent complètement séparées dans l'intégrale multiple, les limites des  $\xi_i$  restant infinies, comme celles des  $z_i$  ou des  $\alpha_i$  dont elles dérivent; et l'on obtient la probabilité

$$P = \frac{1}{\pi^m} \left\{ \int dt_1 dt_2 \dots dt_m e^{-t_1^2 - t_2^2 - \dots - t_m^2} \times \int_{-\infty}^{\infty} d\xi_1 d\xi_2 \dots d\xi_m e^{-\xi_1^2 - \xi_2^2 - \dots - \xi_m^2} \left( 1 - \frac{\sqrt{-1}}{6} Z_3 + \frac{1}{24} Z_4 - \frac{1}{72} Z_3^2 \right) \right\}.$$

Dans cette nouvelle expression, l'intégrale multiple relative aux  $\xi$  s'obtient sans autre peine que d'écrire, attendu les valeurs connues

$$\begin{aligned} \int_{-\infty}^{\infty} d\xi e^{-\xi^2} &= \sqrt{\pi}, \quad \int_{-\infty}^{\infty} \xi^{2i+1} d\xi e^{-\xi^2} = 0, \\ \int_{-\infty}^{\infty} \xi^{2i} d\xi e^{-\xi^2} &= \frac{1.3.5 \dots 2i-1}{2.2.2 \dots 2} \sqrt{\pi}. \end{aligned}$$

On aura donc

$$\int_{-\infty}^{\infty} d\xi_1 d\xi_2 \dots d\xi_m e^{-\xi_1^2 - \xi_2^2 - \dots - \xi_m^2} = (\sqrt{\pi})^m.$$

Mais, lorsque la fonction sous le signe sera multipliée par les quantités représentées dans l'équation (6) par  $Z_3$  et  $Z_4$  ou par  $Z_3^2$ , le résultat sera moins simple. On reconnaît sur-le-champ qu'il ne contiendra que les parties de ces expressions multipliées par des puissances paires de  $\xi$ , les puissances impaires disparaissant. Il en ressort que  $Z_3$  se changera en une fonction  $B_3 (\sqrt{\pi})^m$  des  $t_i$ , ne renfermant que les puissances première, deuxième et troisième de ces nouvelles variables;  $Z_4$  deviendra  $B_4 (\sqrt{\pi})^m$  et ne contiendra que les puissances zéro, première, deuxième, troisième et quatrième; enfin  $Z_3^2$  devenant  $B_6 (\sqrt{\pi})^m$  offrira les puissances zéro, première, deuxième, troisième, quatrième, cinquième et sixième des  $t_i$ . Ces termes ne seront pas développés, uniquement pour abrégé.

D'après toutes ces remarques, il reste pour la probabilité

$$p = \frac{1}{(\sqrt{\pi})^m} \int dt_1 dt_2 \dots dt_m e^{-t_1^2 - t_2^2 - \dots - t_m^2} \left( 1 - \frac{1}{6} B_3 + \frac{1}{24} B_4 - \frac{1}{72} B_6 \right).$$

Une seule difficulté semblerait peut-être entraver cette transformation (si l'on écarte, bien entendu, celles qui tiennent au procédé de Laplace, dont il ne sera pas question ici). La difficulté dont il s'agit, c'est celle qui porterait sur la détermination des arbitraires  $h_{i,y}$ , lesquelles ne sont déduites des coefficients  $b_{i,y}$  que par des équations du deuxième degré (9), et pourraient devenir imaginaires. Mais c'est là ce qui ne saurait arriver, attendu que la fonction du deuxième degré en  $z_1, z_2, \dots, z_m$  est une somme de carrés, comme on peut s'en convaincre; et cette forme permet d'y appliquer la marche suivie dans un Mémoire présenté à l'Académie des Sciences en 1834 (tome VI du *Recueil des Savants étrangers*). La transformation successive des variables mettrait en évidence les sommes de carrés qui seules entrent dans les radicaux, et les rendent nécessairement réels.

Ramenée par tout ce qui précède à ne plus contenir que des inté-

grales de la forme

$$\int dt e^{-t^n},$$

rien ne sera plus facile que d'obtenir la probabilité  $p$  lorsque les limites des variables  $t_i$  seront déduites, au moyen des relations (10), de celles qu'on voudra assigner aux erreurs  $\rho_i$ . Les grandeurs de ces dernières sont arbitraires; mais on voit que la grandeur assignée à l'une d'elles influera sur la forme de toutes les autres, ou tout au moins sur la forme de celles qui la suivent dans l'ordre des relations (10). Si l'on voulait que les variables  $\rho_i$  fussent seulement proportionnelles aux termes  $t_i h_{i,i}$ , qui n'entrent qu'une fois dans chacune respectivement, on n'y parviendrait qu'en assignant aux variables  $t_i$  dans l'intégrale  $p$  des limites qui dépendraient de ces variables; si bien qu'il ne serait possible d'évaluer  $p$  que par des approximations très-pénibles. Mais il existe des combinaisons d'erreurs dont la probabilité peut, au contraire, s'exprimer sans trop de difficultés.

Ce sont celles pour lesquelles l'exposant de  $e$  ne peut prendre que les valeurs inférieures à une certaine constante  $\gamma^2$ . On voit qu'il faut intégrer  $p$  dans cette hypothèse pour toutes les valeurs des  $t_i$  qui satisfont à la condition

$$t_1^2 + t_2^2 + \dots + t_i^2 + \dots + t_m^2 < \gamma^2.$$

On y parviendra par plusieurs méthodes bien connues. Il pouvait se trouver quelque intérêt, il y a quinze ou vingt ans, lors de la première recherche sur ce sujet, à développer ces procédés. Il suffira aujourd'hui de montrer qu'il n'est besoin que d'intégrations très-simples.

En opérant successivement sur les variables  $t_i$ , on aurait à intégrer chacune d'elles entre les limites égales et de signes contraires

$$t_i = \pm \sqrt{\gamma^2 - t_1^2 - t_2^2 - \dots - t_{i-1}^2}.$$

Or, entre de semblables limites, il est visible que l'intégration relative à la fonction qui a été désignée par  $B_3$  donnera un résultat nul.  $B_3$  provient effectivement de  $Z_3$ , où il n'entrait que des produits de degré impair des  $z_i$ , et où il n'est resté que les termes affectés de puis-



sances paires des  $\xi_i$ , c'est-à-dire ceux qui offrent des puissances impaires des  $t_i$ . Ils donneront lieu à des intégrations de la forme

$$\int_{-c}^c dt e^{-t^2} t^{2n+1} = 0.$$

Ainsi, tout ce qui est relatif à  $B_3$  disparaîtra.

Quant à  $B_4$  et  $B_6$ , il s'y trouve des produits de puissances paires des  $t_i$ , et les puissances impaires seront affectées de  $\sqrt{-1}$ ; celles-ci feront disparaître les termes imaginaires où elles se rencontrent. Mais l'intégration laissera subsister tous les termes où il n'entre que des puissances paires. Le résultat ne sera donc pas nul. Seulement, tous les termes de ces fonctions acquièrent, par la substitution des valeurs de  $z_i$  en fonction des  $\xi_i$  et  $t_i$ , des diviseurs qui contiennent les carrés des sommes finies des facteurs  $K_{i,h}$ , lesquels sont au nombre de  $n$  dans chaque somme. Ainsi, quand le nombre  $n$  des observations sera très-grand, chacun des termes sera de l'ordre très-petit  $\frac{1}{n}$ . Il serait superflu de s'arrêter sur ce point dans la question présente. On sait assez que c'est là précisément la forme qu'amène l'analyse de Laplace. Il n'y aurait à le montrer par le calcul même, d'autre difficulté que la longueur de l'écriture de ces expressions assez complexes. Toutefois, il faut faire observer que, quand il y a  $m$  éléments ou inconnues, et non une seule, le nombre de ces termes de l'ordre de  $\frac{1}{n}$  dépend du nombre  $m$ . De sorte que l'ensemble de ces termes n'est que de l'ordre de  $\frac{m}{n}$ . Il est donc indispensable, puisque l'on néglige constamment cette partie de l'intégrale, de s'assurer que  $\frac{n}{m}$  est un grand nombre; assez grand surtout pour contre-balancer l'influence des puissances de  $\gamma$  qui entrent dans ces termes jusqu'à  $\gamma^6$ . Ce sera donc dans les applications une condition à ne pas oublier, que  $\gamma^6 \frac{m}{n}$  reste de l'ordre des quantités qu'on croira pouvoir se permettre de négliger. S'il n'en est pas ainsi, on ne pourrait s'assurer quelque exactitude que pour de petites valeurs de  $\gamma$ ; et c'est ce qui n'a pas toujours été fait avec l'attention que ce point mérite.

D'après ces considérations, il ne reste plus à s'occuper que de la formule d'approximation

$$p = \frac{1}{(\sqrt{\pi})^m} \int dt_1 dt_2 \dots dt_m e^{-t_1^2 - t_2^2 - \dots - t_m^2}.$$

Entre les limites égales et de signes contraires, il est palpable que les valeurs négatives des  $t_i$  donnent des résultats égaux à ceux que produisent les valeurs positives. On peut donc doubler le résultat de chaque intégration, ou multiplier l'intégrale ci-dessus par  $2^m$ , et ne faire le calcul que de zéro aux limites positives. C'est ce qui s'exécute facilement en transformant d'abord l'une des variables  $t_m$  par exemple, au moyen de la relation

$$\begin{aligned} t_1^2 + t_2^2 + \dots + t_m^2 &= u^2, \\ t_m &= \sqrt{u^2 - t_1^2 - t_2^2 - \dots - t_{m-1}^2}, \\ t_m dt_m &= u du, \\ dt_m &= \frac{u du}{\sqrt{u^2 - t_1^2 - t_2^2 - \dots - t_{m-1}^2}}, \end{aligned}$$

qui entraîne pour les limites de  $u$  les valeurs zéro et  $\gamma$ . L'expression de  $p$  se ramène ainsi à

$$p = \left(\frac{2}{\sqrt{\pi}}\right)^m \int_0^\gamma u du e^{-u^2} \int_{m-1} \frac{dt_1 dt_2 \dots dt_{m-1}}{\sqrt{u^2 - t_1^2 - t_2^2 - \dots - t_{m-1}^2}},$$

et sous cette forme il suffit de la seule intégrale vraiment élémentaire,

$$\int_0^{\sqrt{a}} dt t^\eta (a - t^2)^{\frac{\delta}{2}} = a^{\eta + \frac{\delta+1}{2}} \frac{\Gamma\left(\eta + \frac{1}{2}\right) \Gamma\left(\frac{\delta}{2} + 1\right)}{2\Gamma\left(\eta + 1 + \frac{\delta+1}{2}\right)},$$

pour réduire l'intégrale multiple. Les fonctions gamma n'ont ici pour objet que d'abrégier l'écriture. La formule qui les contient est proprement celle qui sert à réduire les termes renfermés dans  $B_4$  et  $B_6$ . Il faut y faire  $\eta = 0$ , pour l'appliquer à la valeur approximative de  $p$ .

Elle devient, dans ce cas particulier,

$$\int_0^{\sqrt{a}} dt (a - t^2)^{\frac{\delta}{2}} = \frac{\sqrt{\pi}}{2} a^{\frac{\delta+1}{2}} \frac{\Gamma\left(\frac{\delta}{2} + 1\right)}{\Gamma\left(\frac{\delta+1}{2} + 1\right)}.$$

Employant  $(m - 1)$  fois cette expression, en ayant soin de donner successivement à  $a$  les valeurs  $(u^2 - t_1^2 - t_2^2 - \dots - t_i)$ , qui résultent de la disparition d'une des variables  $t_i$  à chaque opération, on obtiendra

$$\begin{aligned} \int \frac{dt_1 dt_2 \dots dt_{m-1}}{\sqrt{u^2 - t_1^2 - t_2^2 - \dots - t_{m-1}^2}} &= u^{m-2} \left(\frac{\sqrt{\pi}}{2}\right)^{m-1} \frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma(1)} \times \frac{\Gamma(1)}{\Gamma\left(\frac{3}{2}\right)} \times \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(2)} \times \dots \times \frac{\Gamma\left(\frac{m-1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \\ &= u^{m-1} \left(\frac{\sqrt{\pi}}{2}\right)^m \frac{2}{\Gamma\left(\frac{m}{2}\right)}. \end{aligned}$$

Substituant dans  $p$ , il en résulte

$$p = \frac{2}{\Gamma\left(\frac{m}{2}\right)} \int_0^\gamma u^{m-1} du e^{-u^2}.$$

Ce résidu fort simple reçoit deux formes très-différentes selon que le nombre  $m$  des éléments ou des inconnues est pair ou impair. Il est très-digne de remarque qu'au point de vue du calcul numérique, la probabilité soit plus simplement exprimée quand il se présente un nombre pair d'inconnues, que lors même qu'il s'agit d'une seule.

On sait effectivement que

$$\begin{aligned} \int_0^\gamma u^{m-1} du e^{-u^2} &= -\gamma^{m-2} \frac{e^{-\gamma^2}}{2} - \frac{m-2}{2} \gamma^{m-4} \frac{e^{-\gamma^2}}{2} - \frac{m-2}{2} \cdot \frac{m-4}{2} \gamma^{m-6} \frac{e^{-\gamma^2}}{2} - \dots \\ &\quad - \frac{m-2}{2} \cdot \frac{m-4}{2} \dots \frac{m-2i}{2} \gamma^{m-2i-2} \frac{e^{-\gamma^2}}{2} \\ &\quad + \frac{m-2}{2} \cdot \frac{m-4}{2} \dots \frac{m-2i}{2} \frac{m-2i-2}{2} \int_0^\gamma u^{m-2i-3} du e^{-u^2}. \end{aligned}$$

De sorte que, pour  $m = 2g$ ,

$$\int_0^\gamma u^{2g-1} du e^{-u^2} = \left\{ \begin{aligned} & -\gamma^{2g-2} - \frac{2g-2}{2} \gamma^{2g-4} - \frac{2g-2}{2} \cdot \frac{2g-4}{2} \cdot \gamma^{2g-6} - \dots \\ & - \frac{2g-2}{2} \cdot \frac{2g-4}{2} \dots \frac{2}{2} \\ & + \frac{2g-2}{2} \cdot \frac{2g-4}{2} \dots \frac{2}{2} \cdot \frac{1}{2}, \end{aligned} \right\} \frac{e^{-\gamma^2}}{2}$$

qu'on peut écrire

$$\int_0^\gamma u^{2g-1} du e^{-u^2} = \frac{1}{2} \Gamma\left(\frac{2g}{2}\right) - \frac{1}{2} \Gamma\left(\frac{2g}{2}\right) e^{-\gamma^2} \left[ \frac{\gamma^{2g-2}}{\Gamma\left(\frac{2g}{2}\right)} + \frac{\gamma^{2g-4}}{\Gamma\left(\frac{2g-2}{2}\right)} + \dots + \frac{\gamma^2}{\Gamma\left(\frac{4}{2}\right)} + 1 \right];$$

et, pour  $m = 2g - 1$ ,

$$\int_0^\gamma u^{2g-2} du e^{-u^2} = \left\{ \begin{aligned} & -\gamma^{2g-3} - \frac{2g-3}{2} \gamma^{2g-5} - \frac{2g-3}{2} \cdot \frac{2g-5}{2} \gamma^{2g-7} - \dots \\ & - \frac{2g-3}{2} \cdot \frac{2g-5}{2} \dots \frac{3}{2} \gamma \\ & + \frac{2g-3}{2} \cdot \frac{2g-5}{2} \dots \frac{3}{2} \cdot \frac{1}{2} \int_0^\gamma du e^{-u^2}, \end{aligned} \right\} \frac{e^{-\gamma^2}}{2}$$

qui s'écrira aussi, à cause de  $\Gamma\left(\frac{2g-1}{2}\right) = \frac{2g-3}{2} \cdot \frac{2g-5}{2} \dots \frac{3}{2} \cdot \frac{\sqrt{\pi}}{2}$ ,

$$\int_0^\gamma u^{2g-2} du e^{-u^2} = \Gamma\left(\frac{2g-1}{2}\right) \times \frac{1}{\sqrt{\pi}} \int_0^\gamma du e^{-u^2} - \frac{1}{2} \Gamma\left(\frac{2g-1}{2}\right) e^{\gamma^2} \left[ \frac{\gamma^{2g-3}}{\Gamma\left(\frac{2g-1}{2}\right)} + \frac{\gamma^{2g-5}}{\Gamma\left(\frac{2g-3}{2}\right)} + \dots + \frac{\gamma}{\Gamma\left(\frac{3}{2}\right)} \right].$$

On aura donc enfin

$$(11) \left\{ \begin{aligned} p_{2g-1} &= \frac{2}{\sqrt{\pi}} \int_0^\gamma du e^{-u^2} - e^{-\gamma^2} \left[ \frac{\gamma^{2g-3}}{\Gamma\left(\frac{2g-1}{2}\right)} + \frac{\gamma^{2g-5}}{\Gamma\left(\frac{2g-3}{2}\right)} + \dots + \frac{\gamma^3}{\Gamma\left(\frac{5}{2}\right)} + \frac{\gamma}{\Gamma\left(\frac{3}{2}\right)} \right], \\ p_{2g} &= 1 - e^{-\gamma^2} \left[ \frac{\gamma^{2g-2}}{\Gamma\left(\frac{2g}{2}\right)} + \frac{\gamma^{2g-4}}{\Gamma\left(\frac{2g-2}{2}\right)} + \dots + \frac{\gamma^2}{\Gamma\left(\frac{4}{2}\right)} + \frac{1}{\Gamma\left(\frac{2}{2}\right)} \right]. \end{aligned} \right.$$

Telles sont les expressions approchées de la probabilité que les

erreurs  $\rho_i$  (10), exprimées en fonctions des variables  $t_i$ , ne peuvent pas s'étendre au delà des limites assignées par la condition

$$(12) \quad t_1^2 + t_2^2 + \dots + t_m^2 < \gamma^2.$$

S'il n'y avait que trois variables, on pourrait exprimer cette condition en disant que les points qu'elles déterminent, quand on les regarde comme des coordonnées rectangulaires, ne sortent pas de la sphère de rayon  $\gamma$ . L'analogie conduit donc à dire que les variables  $t$  ne peuvent sortir de l'étendue de la relation analytique

$$t_1^2 + t_2^2 + \dots + t_m^2 = \gamma^2,$$

qui ne permet à aucune d'excéder  $\pm \gamma$ .

On reconnaît sans doute déjà, dans les formes générales de la probabilité (11), combien elle sera différente selon le nombre des éléments; on voit qu'elle diminue pour une même valeur  $\gamma$ , à mesure que le nombre  $m$  des éléments s'accroît, et c'est ce qui doit être, conformément aux principes du calcul des probabilités rappelés au commencement de ce travail.

C'était là le but principal qu'il s'agissait d'atteindre; et, jusqu'à certain degré, il serait loisible de s'y arrêter. Car on reconnaît sans peine que les coefficients  $K_{i,h}$  peuvent être ceux que donne la méthode des moindres carrés, puisque ceux-ci satisfont aux conditions peu nombreuses (3), qui seules sont imposées à ces facteurs  $K_{i,h}$ , pour qu'il en résulte des valeurs des inconnues.

Mais, comme les expressions des erreurs de ces éléments sont ici différentes de celles qui se trouvent répandues dans plusieurs ouvrages depuis les théories de M. Gauss et de Laplace, on se demanderait, sans nul doute, si les résultats de l'analyse précédente rendent ou non nécessaires les facteurs  $K_{i,h}$  particuliers à la méthode des moindres carrés. Il semble donc indispensable de prouver que, sous les conditions posées, la valeur de la probabilité, ou plutôt les expressions des erreurs dans lesquelles toutes les difficultés de la question sont réunies, puisque la probabilité n'est qu'une pure constante qu'on peut calculer d'avance et dont on a des Tables, auxquelles il ne s'agit que d'appliquer des erreurs de grandeur donnée; il

semble indispensable de prouver que les erreurs ne seront les plus petites possible que quand on emploiera à l'élimination des inconnues les facteurs  $K_{i,h}$  assignés par la méthode des moindres carrés. On verra alors clairement que l'omission commise sur la valeur de la probabilité des erreurs n'altérerait que cette probabilité, et que la modification qui la répare (formules 11) n'apporte aucun changement au mode d'élimination prescrit par la méthode. C'est ce qui va être fait, avant de procéder à quelques applications numériques.

Mais dès à présent il convient de le redire, les formules (11) et (10) réparent l'omission complètement. On reconnaît, de plus, qu'elles s'appliquent à tous les systèmes d'élimination au moyen de facteurs ou de combinaisons linéaires des équations données par l'observation. Elles permettent ainsi de calculer l'erreur et la probabilité dans nombre de cas où l'on ne veut pas prendre toutes les peines qu'exige la méthode des moindres carrés.

Elles renferment effectivement le calcul de la probabilité d'un système de fonctions linéaires  $r_i$  (4) d'erreurs  $\varepsilon_h$  soumises à une loi de probabilité  $\varphi(\varepsilon)$ , quelle que soit l'origine de ces fonctions, et quelles que puissent être les déterminations des facteurs  $K_{i,h}$ . Ces formules offrent donc la solution d'une classe de problèmes assez étendue, dans lesquels la coexistence de plusieurs fonctions est nécessaire, et où l'on ne saurait dès lors avoir en vue que la probabilité de leur ensemble. Il est d'ailleurs assez digne de remarque que les expressions de la probabilité soient les mêmes que si les fonctions  $r_i$  ou  $\rho_i$  étaient indépendantes, et qu'il ne se fût agi que de faire le produit de leurs probabilités particulières, puis d'en déterminer la valeur sous la condition (12); il est bien entendu que, dans ce cas, les relations (10) ne subsisteraient pas. Mais il faut s'arrêter dans ces indications, et revenir à la méthode des moindres carrés.

### § III.

Il s'agit de rechercher quelles sont les valeurs des facteurs arbitraires  $K_{i,h}$  qui donneront les plus étroites limites aux erreurs  $\rho_i$  pour une probabilité déterminée par la constante  $\gamma$ .

La question ne saurait se résoudre sans la connaissance de l'étendue

des valeurs que peut prendre une erreur  $\rho_i$  dépendant des  $t_i$ , selon les formules (10), quand les  $t_i$  sont soumises à la condition (12),

$$t_1^2 + t_2^2 + \dots + t_m^2 < \gamma^2.$$

Qu'on suppose d'abord les variables  $t_i$  liées par la relation

$$t_1^2 + t_2^2 + \dots + t_m^2 = u,$$

et qu'on cherche la plus grande valeur que puisse recevoir

$$\rho_i = t_1 h_{1,i} + t_2 h_{2,i} + \dots + t_i h_{i,i}$$

qui ne contient que  $i$  des  $m$  variables  $t$ . On pourra considérer ces  $i$  variables, quelles que soient d'ailleurs les  $m - i$  restantes comme liées par l'équation

$$t_1^2 + t_2^2 + \dots + t_i^2 = u - t_{i+1}^2 - t_{i+2}^2 - \dots - t_m^2 = v.$$

Partant, on aura, pour la plus grande valeur de  $\rho_i$ , les conditions

$$\begin{aligned} \frac{d\rho_i}{dt_{i'}} &= h_{i',i} + h_{i,i} \frac{dt_i}{dt_{i'}} = 0, \\ t_{i'} dt_{i'} + t_i dt_i &= 0, \end{aligned}$$

$t_{i'}$  étant l'une quelconque des  $i$  premières variables,  $v$  étant regardée comme constante relativement à ces variables, et  $t_i$  se trouvant fonction des  $i - 1$  qui la précèdent. On déduit de là

$$h_{i',i} - h_{i,i} \frac{t_{i'}}{t_i} = 0 \quad \text{ou} \quad \frac{t_{i'}}{t_i} = \frac{h_{i',i}}{h_{i,i}}.$$

Il faudra donc que les  $t_{i'}$  soient proportionnels aux  $h_{i',i}$  correspondants. Posant

$$t_{i'} = f h_{i',i},$$

on en conclut

$$f^2 (h_{1,i}^2 + h_{2,i}^2 + h_{3,i}^2 + \dots + h_{i,i}^2) = v.$$

Or, en vertu de la dernière des relations (9), la somme qui multiplie  $f^2$  est précisément égale au coefficient  $b_{i,i}$ ; on a donc

$$f^2 b_{i,i} = v \quad \text{et} \quad f = \sqrt{\frac{v}{b_{i,i}}}.$$

La valeur la plus grande de  $\rho_i$  sera donc

$$\begin{aligned}\rho_i &= f(h_{1,i}^2 + h_{2,i}^2 + \dots + h_{i,i}^2) \\ &= b_{i,i} \sqrt{\frac{\nu}{b_{i,i}}} = \sqrt{\nu b_{i,i}}.\end{aligned}$$

Cette valeur sera, en effet, la plus grande absolument, car on peut s'assurer que le signe de la différentielle seconde totale est négatif ou positif, selon qu'on prend le facteur  $f$  positif ou négatif, c'est-à-dire selon qu'on prend  $\rho_i$  à sa limite positive ou à sa limite négative. Au lieu de calculer la différentielle seconde, ce qui n'est pas très-compliqué, on abrégera cependant, en substituant  $fh_{i',i} + \partial_i$  dans  $\rho_i$  au lieu de  $t_{i'}$ , ce qui donne

$$\begin{aligned}\rho_i &= f(h_{1,i}^2 + h_{2,i}^2 + \dots + h_{i,i}^2) + \partial_1 h_{1,i} + \partial_2 h_{2,i} + \dots + \partial_i h_{i,i} \\ &= fb_{i,i} + \partial_1 h_{1,i} + \partial_2 h_{2,i} + \dots + \partial_i h_{i,i}.\end{aligned}$$

Faisant la même substitution dans  $\nu$ , il vient

$$\begin{aligned}\nu &= f^2(h_{1,i}^2 + h_{2,i}^2 + \dots + h_{i,i}^2) \\ &\quad + 2f(\partial_1 h_{1,i} + \partial_2 h_{2,i} + \dots + \partial_i h_{i,i}) \\ &\quad + \partial_1^2 + \partial_2^2 + \dots + \partial_i^2,\end{aligned}$$

et, par suite de la valeur de  $f^2$ , il ne reste que

$$\partial_1 h_{1,i} + \partial_2 h_{2,i} + \dots + \partial_i h_{i,i} = -\frac{1}{2f}(\partial_1^2 + \partial_2^2 + \dots + \partial_i^2).$$

Ainsi

$$\rho_i = f \left[ b_{i,i} - \frac{1}{2}(\partial_1^2 + \partial_2^2 + \dots + \partial_i^2) \right].$$

La valeur de

$$\rho_i = fb_{i,i}$$

est donc un maximum ou un minimum, suivant le signe de  $f$ , c'est-à-dire qu'elle est bien la plus grande absolument.

Il ne faut pas oublier, maintenant, que  $\nu$  a pour limite supérieure  $u$ , ce qui suppose que tous les  $t$  de  $t_{i+1}$  à  $t_m$  sont nuls. D'après cela, les limites extrêmes de  $\rho_i$  sont  $\pm \sqrt{ub_{i,i}}$ . Puis, comme  $u$  doit rester inférieur à  $\gamma^2$ , on voit que, définitivement, pour une probabi-



lité déterminée par la constante  $\gamma$ , cette constante fixe très-simplement l'étendue des limites des erreurs  $\rho_i$  sous la forme

$$\rho_i = \pm \gamma \sqrt{b_{i,i}}.$$

Au surplus, cette forme était bien facile à prévoir d'après la première des relations (10) et la symétrie de tout ce calcul. Il y a, en effet, dans les relations (10), une symétrie réelle qui dépend de ce qu'une des erreurs peut occuper, à volonté, un rang quelconque dans les transformations; si bien, qu'il est permis d'affirmer que ce qui peut se reconnaître sur l'erreur exprimée le plus simplement, a lieu nécessairement pour toutes. Mais l'apparence asymétrique des relations posées a fait préférer une preuve plus saisissable.

Si l'on se rappelle actuellement que  $b_{i,i}$  n'a fait que remplacer  $S.K_{i,h}^2$ , on aura, pour les limites de  $\rho_i$ ,

$$\rho_i = \pm \gamma \sqrt{S.K_{i,h}^2}.$$

Il ne reste, dès lors, pour restreindre le plus possible cette étendue des erreurs  $\rho_i$ , qu'à trouver les moyens de rendre un minimum la somme de carrés  $S.K_{i,h}^2$ . Or on sait que c'est là une des propriétés des facteurs par lesquels on effectue l'élimination dans la méthode des moindres carrés.

On ne peut guère, pour le démontrer brièvement, prendre d'autre voie que la comparaison immédiate des coefficients  $K_{i,h}$  qui servent à obtenir la valeur

$$x_i = S \omega_h K_{i,h},$$

avec les facteurs qui donneraient la même inconnue, d'après la méthode des moindres carrés.

Appelant ces facteurs  $A_{i,h}$ , il est clair qu'ils satisferont, comme les  $K_{i,h}$ , aux équations (3), dans lesquelles il suffirait de substituer une lettre à l'autre, et qu'on trouverait

$$x'_i = S \omega_h A_{i,h}.$$

Mais en même temps, si l'on avait opéré directement selon les prescriptions de la méthode en formant  $m$  équations, à l'aide de la multi-

plication successive des  $n$  équations (1) par les coefficients de chaque inconnue, ces  $m$  équations seraient

$$\begin{aligned} x_1 S a_{1,h} a_{1,h} + x_2 S a_{1,h} a_{2,h} + \dots + x_i S a_{1,h} a_{i,h} + \dots &= S a_{1,h} \omega_h, \\ x_1 S a_{2,h} a_{1,h} + x_2 S a_{2,h} a_{2,h} + \dots + x_i S a_{2,h} a_{i,h} + \dots &= S a_{2,h} \omega_h, \\ \dots & \\ x_1 S a_{i,h} a_{1,h} + x_2 S a_{i,h} a_{2,h} + \dots + x_i S a_{i,h} a_{i,h} + \dots &= S a_{i,h} \omega_h, \\ \dots & \\ x_1 S a_{m,h} a_{1,h} + x_2 S a_{m,h} a_{2,h} + \dots + x_i S a_{m,h} a_{i,h} + \dots &= S a_{m,h} \omega_h, \end{aligned}$$

et l'on éliminerait en multipliant par  $m$  facteurs  $B_1, B_2, \dots, B_i, \dots, B_m$ , qui seraient assujettis à autant de relations, telles que l'on eût

$$x'_i = B_1 S a_{1,h} \omega_h + B_2 S a_{2,h} \omega_h + \dots + B_m S a_{m,h} \omega_h,$$

ou la même valeur que par les coefficients  $A_{i,h}$ .

On pourra réunir en facteur commun tout ce qui multiplie  $\omega_1, \omega_2, \dots, \omega_n$ , et il en résultera

$$\begin{aligned} x'_i = \omega_1 \Sigma B_{i'} a_{i',1} + \omega_2 \Sigma B_{i'} a_{i',2} + \omega_3 \Sigma B_{i'} a_{i',3} + \dots + \omega_h \Sigma B_{i'} a_{i',h} + \dots \\ + \omega_n \Sigma B_{i'} a_{i',n}. \end{aligned}$$

En rapprochant cette forme de la précédente

$$x'_i = S \omega_h A_{i,h},$$

on reconnaît que les coefficients  $A_{i,h}$ , appliqués directement aux  $n$  équations données, sont liés aux facteurs  $B_1, B_2$ , etc., par la relation

$$A_{i,h} = \Sigma B_{i'} a_{i',h}.$$

Que maintenant on se reporte aux relations (3) des  $K_{i,h}$  et qu'on en retranche un système de relations semblables dans lesquelles on aura mis  $A_{i,h}$  à la place de  $K_{i,h}$ , ainsi que la remarque en a été faite tout à l'heure; la soustraction de deux relations correspondantes

donnera un système de  $m$  équations

$$\begin{aligned} Sa_{1,h}(K_{i,h} - A_{i,h}) &= 0, \\ Sa_{2,h}(K_{i,h} - A_{i,h}) &= 0, \\ \dots \dots \dots \dots \dots \dots \dots \\ Sa_{i',h}(K_{i,h} - A_{i,h}) &= 0, \\ \dots \dots \dots \dots \dots \dots \dots \\ Sa_{i,h}(K_{i,h} - A_{i,h}) &= 0, \\ \dots \dots \dots \dots \dots \dots \dots \\ Sa_{m,h}(K_{i,h} - A_{i,h}) &= 0; \end{aligned}$$

et il devient manifeste qu'en les multipliant par les facteurs  $B_{i'}$  respectivement, puis les ajoutant toutes, elles donneront

$$S[(\Sigma B_{i'} a_{i',h})(K_{i,h} - A_{i,h})] = 0;$$

et, comme  $\Sigma B_{i'} a_{i',h}$  n'est autre chose que  $A_{i,h}$  dans la méthode des moindres carrés, il en résultera

$$S[A_{i,h}(K_{i,h} - A_{i,h})] = 0,$$

ou bien

$$SA_{i,h} K_{i,h} = SA_{i,h}^2,$$

relation singulière entre les facteurs arbitraires  $K_{i,h}$  et ceux que nécessite le minimum des carrés. De là rien n'est plus facile que de conclure l'identité

$$SK_{i,h}^2 - 2SA_{i,h} K_{i,h} = SK_{i,h}^2 - 2SA_{i,h}^2,$$

puis

$$SK_{i,h}^2 - SA_{i,h}^2 = SK_{i,h}^2 - 2SK_{i,h} A_{i,h} + SA_{i,h}^2,$$

et

$$SK_{i,h}^2 = SA_{i,h}^2 + S(K_{i,h} - A_{i,h})^2,$$

expression par laquelle M. Gauss a démontré que, sous les relations (3), le minimum de la somme des carrés des facteurs  $K_{i,h}$  a lieu quand ces facteurs sont précisément les facteurs  $A_{i,h}$  assignés par la méthode de Legendre.

Cette méthode réduit donc les erreurs qui entreront nécessairement

dans les valeurs des  $x_i$ , aux plus étroites limites possibles, pour une probabilité donnée. Réciproquement, si un système de limites détermine une probabilité, comme la quantité  $\gamma$  qui sert à la calculer sera liée à toutes les limites par la relation

$$\rho_i = \gamma \sqrt{S \cdot K_{i,h}^2},$$

il est palpable que  $\gamma$  sera un maximum pour le minimum de  $S \cdot K_{i,h}^2$ , ou pour le résultat de la méthode des moindres carrés; de sorte qu'un système de limites étant choisi, la probabilité que les erreurs n'en sortiront pas sera la plus grande possible, quand on aura déterminé les inconnues par cette méthode.

Il a déjà été dit effectivement que, dans les formules (11), la valeur de  $p$  est croissante avec  $\gamma$ . C'est ce qu'on reconnaîtrait immédiatement par la différentiation de ces formules qui, toutes deux, lorsque  $m$  est le nombre  $2g$  ou  $2g - 1$  des éléments, ont pour dérivées la quantité

$$\frac{dp_m}{d\gamma} = 2e^{-\gamma^2} \frac{\gamma^{m-1}}{\Gamma\left(\frac{m}{2}\right)},$$

expression très-simple qui ne changerait de signe qu'avec la constante  $\gamma$ , ici regardée comme variable.

On peut s'assurer que, pour  $m = 1$ , quand il n'y a qu'un seul élément ou une seule inconnue  $x$ , toutes ces formules rentrent complètement dans les relations connues

$$r = \mu_1 S \cdot A_h + \rho \sqrt{2(\mu_2 - \mu_1^2)},$$

limites de  $\rho = \pm \gamma \sqrt{S \cdot A_h^2}$ ,

$$p_1 = \frac{2}{\sqrt{\pi}} \int_0^\gamma e^{-t^2} dt.$$

On connaît les procédés simples donnés par Laplace et par M. Gauss pour déduire les quantités  $\mu_1$  et  $\mu_2$  des observations mêmes. Ces procédés, qui dépendent de ce qu'on a appelé la *théorie des probabilités à posteriori*, ne sont pas modifiés par le changement qui vient d'être développé. Il n'influe heureusement que sur la grandeur de la proba-

bilité, ou plutôt sur l'étendue des erreurs les plus probables. Il ne sera donc pas question ici du calcul de  $\mu_1$  ni de  $\mu_2$ . Mais il a paru nécessaire, sinon de former des Tables des valeurs numériques des formules (11), du moins d'en présenter quelques valeurs intéressantes.

Avant de procéder à cette application, peut-être convient-il de faire observer que si l'expression de la probabilité  $p$  donnée généralement (page 54),

$$p = \frac{1}{(\sqrt{\pi})^m} \int dt_1 dt_2 \dots dt_m e^{-t_1^2 - t_2^2 - \dots - t_m^2} \left( 1 - \frac{\sqrt{n}}{6} B_3 + \frac{1}{24} B_4 - \frac{1}{72} B_6 \right),$$

avait été intégrée relativement à toutes les variables excepté une, le résultat aurait été la probabilité des limites assignées à l'erreur  $\rho$  correspondante, quelque grandes que fussent toutes les autres erreurs : en d'autres termes, la probabilité de cette erreur considérée isolément, et comme si les autres n'existaient pas. On arriverait de cette manière aux formules relatives au cas d'un seul élément, telles qu'elles viennent d'être écrites, c'est-à-dire au mode ordinaire de calcul de la probabilité. Mais, encore une fois, ce mode la donne beaucoup trop grande, puisqu'il compte, dans le calcul, toute la probabilité des combinaisons d'erreurs, dans lesquelles les autres erreurs ont des grandeurs qui ne permettraient même plus de se fier aux équations. On va reconnaître, au surplus, l'extrême différence des deux résultats.

Les mêmes considérations ont empêché d'avoir égard aux valeurs moyennes des erreurs. On sait qu'on appelle ainsi la somme des produits des erreurs, prises toutes positivement, par leurs probabilités respectives. Cette moyenne arithmétique, dans laquelle chaque erreur entre proportionnellement au nombre des chances qui peuvent l'amener, ne peut être un indice exact de l'importance d'une erreur que quand elle est calculée isolément. Si, au contraire, il se rencontrait que les grandeurs les plus considérables de l'erreur d'une inconnue fussent précisément celles qui dépendent de systèmes d'erreurs dont la probabilité est faible, on conçoit que la moyenne pourrait donner une idée erronée de la grandeur la plus ordinaire de l'erreur spéciale à laquelle elle se rapporte. En général, l'emploi des valeurs moyennes

est chose délicate, à moins qu'elles ne fassent l'objet spécial des recherches. On sentira sur-le-champ, par exemple, que même dans le cas le plus simple, la moyenne des erreurs  $\rho_i$ , qui se calcule en intégrant

$$\frac{2}{\sqrt{\pi}} \int_0^{\infty} dt e^{-t^2} t \sqrt{S \cdot K_{i,h}^2} = \frac{1}{\sqrt{\pi}} \sqrt{S \cdot K_{i,h}^2},$$

ne donne pas une idée bien juste, puisque les limites entre lesquelles il y a un contre un à parier que peut tomber la valeur  $\rho_i$  sont déterminées par l'équation

$$\int_0^z dt e^{-t^2} = \int_z^{\infty} dt e^{-t^2} = \frac{1}{2} \sqrt{\pi} - \int_0^z dt e^{-t^2},$$

ou bien

$$\int_0^z dt e^{-t^2} = \frac{1}{2} \sqrt{\pi};$$

d'où l'on tire

$$z = 0,476936.$$

$\frac{1}{\sqrt{\pi}}$  est, au contraire, égal à 0,56418.

La moyenne des erreurs est donc loin de se rencontrer parmi les erreurs les plus probables. A la vérité, cette moyenne est celle de toutes les erreurs prises avec le signe +, et la moyenne réelle est 0. Il y aurait plus d'une observation à faire sur l'usage des moyennes; mais il faut ici, pour éviter trop de longueurs, ne pas s'y arrêter, non plus qu'à un grand nombre d'autres points utiles. Il suffit d'avoir montré que les moyennes n'ont pas toujours le sens que les habitudes de l'esprit y font attacher, dans les circonstances les plus ordinaires; et que, en conséquence, elles ne sont point propres à la démonstration de la méthode des moindres carrés. Aussi est-ce de cette évaluation de la moyenne des erreurs, prises avec le signe positif, que M. Gauss forme une espèce d'objection à l'analyse de Laplace (*Theoria Combinationis*, etc.). Le fait est que le résultat ne serait point rigoureux, si cette moyenne ne s'était trouvée proportionnelle aux limites des erreurs. Mais il en est de même de la moyenne des carrés des erreurs que M. Gauss croit pouvoir adopter à priori comme *criterium* de la précision. C'est, au contraire, l'existence de ce criterium

remarquable que démontre l'analyse de Laplace, de même que tout ce qui précède.

#### § IV.

Dans les applications, la valeur de la constante  $\gamma$ , la plus difficile à trouver peut-être, sinon la plus utile, est celle pour laquelle la probabilité est égale à  $\frac{1}{2}$ . Il y a, pour cette valeur, autant à parier que les erreurs tomberont dans les limites qui en résultent, qu'il y a à parier qu'elles excéderont ces limites.

Ainsi, d'abord, dans le cas d'une seule inconnue  $x$ , Bessel a donné la valeur

$$\gamma = 0,4769364.$$

La dernière décimale paraît inexacte, et il faut prendre

$$\gamma = 0,47693627620.$$

Mais c'est là un véritable luxe de décimales, en fait de probabilités; car il sera bien rare que le terme négligé qui dépend du grand nombre  $n$  des observations permette de pousser l'approximation au delà des premières décimales.

On a donc, pour une seule inconnue, la probabilité

$$p_1 = \frac{2}{\sqrt{\pi}} \int_0^{0,4769\dots} dt e^{-t^2} = \frac{1}{2},$$

que l'erreur  $r$  de la valeur de  $x$  est entre les limites

$$r = \mu_1 S. A_h \pm 0,4769\dots \sqrt{2(\mu_2 - \mu_1^2)} S. A_h^2,$$

les facteurs  $A$  étant ceux qu'indique la méthode des moindres carrés. Quant aux valeurs  $\mu_1$  et  $\mu_2$ , elles doivent être cherchées, soit en dehors des observations, soit par les observations mêmes, et cela, d'après les procédés connus pour les probabilités à *posteriori*, ainsi qu'il a été dit.

Lorsque rien n'empêche d'élever la valeur de  $\gamma$  jusqu'à 3 par exemple, la probabilité que l'erreur est renfermée dans les limites

$$r - \mu_1 S. A_h = \pm 3 \sqrt{2(\mu_2 - \mu_1^2)} S. A_h^2$$

est exprimée à peu près par

$$\frac{2}{\sqrt{\pi}} \int_0^3 dt e^{-t^2} = 0,99997790.$$

C'est ce que l'on peut voir dans les Tables qui ont été publiées pour les valeurs de l'intégrale  $\frac{2}{\sqrt{\pi}} \int_0^\gamma dt e^{-t^2}$  (notamment dans l'*Exposition de la Théorie des Chances*, de M. Cournot.)

Ces valeurs sont précisément celles qui ont été appliquées, par omission, aux problèmes qui comprennent plusieurs inconnues. Elles ne sont rapportées ici que pour en faciliter la comparaison avec celles que donnent les formules (11).

S'il y a deux inconnues, il faut employer la deuxième de ces formules, et la probabilité est alors très-simplement

$$p_2 = 1 - e^{-\gamma^2}.$$

Elle devient égale à  $\frac{1}{2}$ , pour  $\frac{1}{2} = e^{-\gamma^2}$ , ou  $\gamma = \sqrt{12}$ , la lettre 1 indiquant les logarithmes népériens. On trouvera sans peine

$$\gamma = 0,83255461\dots$$

Ainsi, dès qu'il y a seulement deux éléments à déduire des observations, les limites comprennent un intervalle presque double, et les erreurs peuvent être bien plus grandes, par conséquent.

On a la probabilité  $\frac{1}{2}$  que l'erreur de l'élément  $x_i$  est comprise dans les limites

$$r_i - \mu_i S.A_{i,h} = \pm 0,83255\dots \sqrt{2(\mu_2 - \mu_1^2) S.A_{i,h}^2},$$

ou, si l'on veut que les erreurs puissent varier, on dira que  $\frac{1}{2}$  est la probabilité de l'ensemble de tous les systèmes pour lesquels

$$r_1 - \mu_1 S.A_{1,h} = t_1 \sqrt{2(\mu_2 - \mu_1^2) S.A_{1,h}^2},$$

$$r_2 - \mu_1 S.A_{2,h} = t_1 \sqrt{2(\mu_2 - \mu_1^2) \frac{(S.A_{1,h} A_{2,h})^2}{S.A_{1,h}^2}}$$

$$+ t_2 \sqrt{2(\mu_2 - \mu_1^2) \left( S.A_{2,h}^2 - \frac{(S.A_{1,h} A_{2,h})^2}{S.A_{1,h}^2} \right)},$$



les variables  $t_1$  et  $t_2$  étant assujetties à la condition

$$t_1^2 + t_2^2 < 0,69314718\dots;$$

les quantités sous les radicaux étant déterminées d'ailleurs par les relations (9), où les facteurs A de la méthode des moindres carrés remplacent les arbitraires K dans les coefficients

$$b_{i,i'} = S.K_{i,h} K_{i',h}.$$

Lorsque dans ce système on voudra atteindre une probabilité 0,99997790... égale à celle que donne  $\gamma = 3$  quand il n'y a qu'une inconnue, il faudra résoudre l'équation

$$1 - e^{-\gamma^2} = 0,99997790\dots;$$

qui élève la valeur de  $\gamma$  à 3,27419...

Ainsi, pour des probabilités très-grandes, les limites différent moins que pour de faibles probabilités, quand on passe du cas d'une seule inconnue au cas de deux inconnues. C'est ce dont il est facile de se rendre compte, puisqu'il en serait absolument de même si les deux inconnues étaient indépendantes l'une de l'autre.

Voici maintenant (afin d'abrégier) le tableau des valeurs de  $\gamma$  qui donnent la probabilité égale à  $\frac{1}{2}$  dans les formules (11), depuis  $m = 1$  jusqu'à  $m = 8$ . Il ne serait pas difficile, mais il serait très-long de faire les mêmes calculs pour un plus grand nombre d'éléments :

$m = 1,$	$\gamma_1 = 0,47693,$	
$m = 2,$	$\gamma_2 = 0,83255,$	$= 1,7456 \times \gamma_1,$
$m = 3,$	$\gamma_3 = 1,0876,$	$= 2,2814 \times \gamma_1,$
$m = 4,$	$\gamma_4 = 1,29551,$	$= 2,7164 \times \gamma_1,$
$m = 5,$	$\gamma_5 = 1,4750,$	$= 3,0927 \times \gamma_1,$
$m = 6,$	$\gamma_6 = 1,63525,$	$= 3,4287 \times \gamma_1,$
$m = 7,$	$\gamma_7 = 1,7812,$	$= 3,7347 \times \gamma_1,$
$m = 8,$	$\gamma_8 = 1,91623,$	$= 4,0178 \times \gamma_1.$

Il ne s'agit que de comparer chacun de ces nombres avec le pre-

mier, pour reconnaître sur-le-champ l'étendue réelle des limites de l'erreur probable, déjà calculées dans un système d'observations. Il n'existe, effectivement, aucun changement à faire subir aux calculs ordinaires, sauf celui de la valeur  $\gamma$ . Il est presque inutile de rappeler, de plus, que l'erreur probable est ainsi nommée parce que c'est précisément la valeur de la limite correspondant à la probabilité  $\frac{1}{2}$ .

Voici un exemple de ce rapprochement. Bessel, dans son Mémoire sur la Comète d'Olbers (*Untersuchungen über die Bahn des Olberschen Kometen*), emploie à la correction des éléments de l'orbite trois cent quarante-neuf observations, dont il forme des équations à six inconnues. Il en conclut, pour l'une de ces inconnues, la correction de l'excentricité, une erreur probable de 0,00 017 127 qu'il évalue en temps à environ 101 jours, sur les 74 ans assignés à la révolution de l'astre.

Dans ce calcul, Bessel a employé le facteur

$$\gamma_1 = 0,47 693.$$

Puisqu'il y a six inconnues, il convenait de prendre, au contraire,

$$\gamma_6 = 1,63 525 = 3,428... \times \gamma_1,$$

c'est-à-dire plus du triple, et d'élever l'erreur probable à 0,00 058 72, ce qui entraîne sans doute une erreur probable de près d'une année sur la durée de la révolution.

On peut très-bien ici reconnaître la nécessité d'étendre ainsi les limites de l'erreur probable. On sent, en effet, qu'il faut, pour que les erreurs des autres éléments puissent avoir la même probabilité, tenir compte de toutes les combinaisons dans lesquelles chacune de leurs valeurs sont en droit d'entrer.

Bessel fait observer avec toute raison qu'il est fort à regretter que les circonstances du mouvement géocentrique aient laissé une si grande incertitude sur cet élément, tandis que les autres n'offrent que des erreurs bien moindres comparativement. La comète d'Olbers ne devant revenir au périhélie que le 9 février 1887, ce ne sera pas avant trente-cinq ans que ces calculs pourront offrir un intérêt positif.

Mais on voit, dans le Mémoire de Bessel, qu'il assignait une probabilité de 66,85 contre 1, à la limite d'un an, tandis qu'on ne peut parier beaucoup plus de 1 contre 1 que la comète n'excédera pas cette limite, en avance ou en retard. Pour obtenir 66,85 contre 1, il faudrait résoudre exactement l'équation résultant de la deuxième des formules (11), ce qui donnerait à  $\gamma$  une valeur (2,812) plus que quintuple de la valeur  $\gamma_1$ , et entraînerait une limite d'erreurs possibles, bien que peu probables, d'environ dix-neuf mois.

Il serait presque inutile de faire exactement ce dernier calcul, car le nombre des observations n'étant que de 349, et le nombre des éléments s'élevant à 6, le diviseur des termes négligés n'atteint que 58; de sorte que ces termes pourraient avoir une grande influence sur les premières décimales de la probabilité. Il suffit d'avoir indiqué le sens dans lequel influaient les omissions. Quant aux petites valeurs de  $\gamma$ , il est bien plus difficile que les termes négligés aient un effet capable d'altérer sensiblement la probabilité  $\frac{1}{2}$ , ou toute autre aussi faible, et c'est par cette raison que le calcul a été fait avec exactitude.

Voici un second et dernier exemple de l'application des formules (11).

Dans le premier Supplément à la *Théorie analytique des Probabilités*, Laplace évalue à 1 000 000 contre 1 la probabilité que la masse de Jupiter, qu'il corrige à l'aide de 129 équations, dans lesquelles il a fait cette masse égale à  $\frac{1+z'}{1067,09}$ , et qu'il réduit à  $\frac{1}{1070,35}$ , n'est pas en erreur de  $\frac{1}{100}$  de la première valeur. Cependant elle est portée, actuellement, à  $\frac{1}{1050}$  dans l'*Annuaire du Bureau des Longitudes*. La différence est d'environ  $\frac{1}{51,5}$ ; ce qui fait à peu près le double des limites assignées avec une si haute probabilité.

Cette différence considérable peut dépendre de la manière dont les 129 équations ont été formées, et peut être même une conséquence toute naturelle du défaut de précision des observations employées, ou des formules de réduction. Mais, si l'on considère que les 129 équations

tions renferment 6 inconnues, et que, par suite, il ne fallait attacher aux limites de  $z'$  que la probabilité de l'ensemble des limites assignées à toutes ces inconnues, et auxquelles elles doivent satisfaire à la fois, on trouve une probabilité bien inférieure à  $\frac{1\ 000\ 000}{1\ 000\ 001}$ , ou, plus exactement,  $1 - 0,00\ 000\ 100\ 4$ , que donnent les valeurs rapportées par Laplace.

La valeur de la correction, telle que la donne Laplace, est

$$z' = 0,00\ 305 + \gamma \sqrt{\frac{1}{(345,885)^2}}$$

Le diviseur sous le radical est un peu trop grand, parce qu'il ne fallait prendre, pour 129 observations et 6 inconnues, que 123 fois (ou  $129 - 6$ ) un certain dénominateur. En tenant compte de cette légère rectification, dépendant de ce que 6 n'est pas à négliger auprès de 129, on a, pour  $z'$ , les limites

$$z' = 0,00\ 305 \pm \gamma \times 0,00\ 296.$$

L'erreur possible de  $z'$  devant être telle, que l'erreur de la masse soit au-dessous de  $\frac{1}{100}$  de  $\frac{1}{1067,09}$ , il en résulte

$$\frac{\gamma \times 0,00\ 296}{1067,09} = \frac{1}{100} \cdot \frac{1}{1067,09},$$

ou bien

$$\gamma = 3,37\ 745.$$

Cette valeur de  $\gamma$  aurait encore donné à Laplace la probabilité  $0,999\ 998\ 21$ , ou à peu près 560 488 à parier contre 1.

Mais, si on la fait entrer dans la seconde des formules (11), en posant  $m = 6$ , on trouve

$$p_6 = 1 - e^{-\gamma^2} \left( \frac{1}{2} \gamma^4 + \gamma^2 + 1 \right);$$

et, tout calcul fait,

$$p_6 = 0,999\ 13\ 89,$$

ou à peu près  $\frac{1160}{1161}$ .

Il n'y avait donc que 1160 à parier contre 1, que la masse  $\frac{1}{1070,35}$  n'était pas en erreur de  $\frac{1}{100}$ .

Cette probabilité peut déjà paraître assez grande; mais il ne faut pas perdre de vue que les formules (11), comme celle de Laplace, négligent des termes qui ne disparaissent que quand  $\frac{m}{n}$  est assez petit pour permettre d'employer des valeurs de  $\gamma$  aussi grandes que 3,37745. Ces termes, sans nul doute, rendent ici la formule erronée, puisque  $\frac{n}{m}$ , loin d'être un grand nombre, n'est que

$$\frac{129}{6} = 21,5.$$

La valeur

$$p = 0,9991389$$

résultant de l'approximation, est donc ici très-imparfaite; et lors même qu'on s'en tiendrait à la formule de Laplace, la valeur qu'il a donnée de  $\frac{1000000}{1000001}$  ne pourrait être admise à cause de la grandeur certaine des termes négligés par lui.

Il faut reconnaître que 129 équations ne permettent d'employer cette analyse approximative que pour de faibles valeurs de  $\gamma$ . Mais alors, dira-t-on avec M. Gauss, la méthode des moindres carrés n'est donc plus démontrée pour les cas les plus fréquents dans la pratique, pour des nombres d'observations très-grands relativement aux peines qu'ils donnent aux observateurs, mais trop petits pour assurer une grande probabilité; cette méthode, qui paraît précieuse, ne serait donc le plus souvent qu'une règle commode par son uniformité et par les avantages que Legendre avait signalés dès l'origine?

A ces regrets trop fondés, il faut répondre affirmativement sans la moindre hésitation. Il n'est besoin, en effet, que du simple bon sens, dépourvu de tout calcul, pour reconnaître que des observations affectées d'erreurs ne peuvent donner les valeurs cherchées que quand il se fait des compensations entre les erreurs; et tout le monde sent que ces compensations ne sauraient avoir lieu que sur de grands nombres.

Employer une moyenne quelconque d'un petit nombre de faits, c'est nécessairement courir le risque que les erreurs s'ajoutent. Il en est ainsi dans toutes les circonstances où l'épreuve ne peut être répétée à volonté, et alors la connaissance qu'on acquiert ne doit plus être imprudemment étendue au delà des limites des plus grandes erreurs possibles. Certes, ce serait une véritable aberration de prétendre que la moyenne arithmétique de 10 mesures dont chacune peut être affectée d'une erreur, en plus ou en moins, de 0 à 1 centimètre, aura un grand avantage sur toute autre combinaison de ces 10 mesures. Sans nul doute, il vaudrait mieux pouvoir démontrer qu'elle jouit toujours, même pour de si petits nombres, d'une probabilité supérieure; mais, par la nature seule des choses, la différence serait très-petite. Aussi, est-il superflu de rechercher ce qui se passe dans de petits nombres. A cet égard, on sera toujours et de toute nécessité contraint de revenir à la réflexion de Gibbon : « Les lois de la probabilité, si vraies en général, si trompeuses en particulier. » Il faudra donc, même quand on aurait prouvé que tel ou tel procédé est le plus probable, ne s'en servir qu'en tenant compte, non plus des erreurs moyennes, mais des erreurs les plus grandes possibles; car la probabilité des unes et des autres ne différera jamais assez pour qu'on puisse se permettre de négliger les unes plutôt que les autres. La probabilité n'est fondée que sur la possibilité des choses, et dans un petit nombre d'observations, un événement possible a chance de prendre sa place comme tout autre de même possibilité. Ce n'est que dans les grands nombres que certaines classes de faits, certaines combinaisons deviennent impossibles, ou plutôt très-peu possibles, par conséquent improbables. Et cette condition impérieuse du grand nombre n'a rien de particulier à la méthode des moindres carrés.

Qu'il faille continuer d'appliquer cette méthode à de petits nombres d'observations, c'est ce qui n'est point douteux. L'analyse qui la démontre reste vraie pour de petits nombres de même que celle qui prouve la valeur des résultats moyens de toute espèce. Seulement, les termes négligés ne laissent plus ressortir un maximum absolu de probabilité; mais, reprenant leur influence, ils viennent altérer un peu le résultat de l'approximation, et l'on voit que ce résultat n'est pas fort éloigné du maximum. S'il en diffère, ce n'est que de petites

quantités dont les diviseurs croissent proportionnellement au nombre des observations; si bien que la méthode reprend cette propriété de donner le maximum d'avantage, dès que ce nombre a quelque valeur, et cela bien avant qu'il soit assez grand pour permettre d'appliquer, sans scrupule, les formules d'approximation au calcul de la probabilité. Il y aurait à ajouter à ces motifs, tirés de la marche des calculs, tous ceux que Legendre fit d'abord valoir en faveur de son procédé. Mais ce n'en est point ici le lieu; ils sont connus et appréciés. Il ne s'agissait que de montrer les moyens de combler ce qui avait paru être une lacune fâcheuse dans une théorie si utile. Car rien n'est plus nuisible au progrès vers la vérité que la confiance erronée qui se croit en possession de résultats dont la science est encore éloignée.

---