Mémoire

sur

la probabilité des erreurs

d'aprés

la méthode des moindres carrés[*]

M. I.-J. Bienaymé

Inspector général des Finances

> Nos sequimur probabilia, nec ultra quam id, quod
> verisimile occurrerit, progredi possumus; et refellere
> sine pertinacia, et refelli sine iracundia, parati sumus.
> Tusculan. 2.   Book 2, §2.[1]

§I.

The method of least squares is so frequently employed today in the sciences of observation, that all that which is able to render the applications more sure becomes of great interest, as simple as it is besides. This consideration has made writing the following researches which have for object to modify the ordinary calculus of the probability of errors, not in the case where the observations are known only a great magnitude, but in the case much more multiplied where the observations give at the same time unknown magnitudes, linked by some equations with the observed magnitude. In other times, when the method was little used in France, and regarded, because of the long calculations that it requires, more as a scholarly curiosity than as a real instrument of the observer, the profound modification recognized in the probability was able to appear only of a secondary importance; but, at present, it seems truly useful to signal the faultiness of the ordinary calculus, for it touches in some works more numerous each day, and the observers, being able to sacrifice precious time to the verification of theories prone to raise difficulties, are obliged to accept the practical rules on the faith of their advancers, especially when those are men of great and just authority in science.

---

[*]Translated by Richard J. Pulskamp, Department of Mathematics & Computer Science, Xavier University, Cincinnati, OH. June 20, 2010

[1]We follow probabilities, nor are we able to go beyond that which may occur having the appearance of truth; and refute without persistance, and we are prepared to be refuted without anger.

There would be perhaps more of a defect to raise in the applications of the method of least squares: it will be however question here that the one who strikes most often the eyes, and of which here is the indication so simple, that in the first words everyone will encounter the existence of it, although the modifications that it requires are able to impose an analytic work rather complicated.

One knows that the method invented by Legendre, around fifty years ago, and published for the first time in his *Nouvelles Méthodes pour la détermination des orbites des comètes*, in-4°, 1805, is reduced to multiply each of the equations of the first degree formed between the magnitude observed and the unknowns, by the coefficient of each of these successive unknowns; to add the products given by coefficients of the same unknown, this which furnishes as many new equations as unknowns; finally to resolve these equations in the ordinary manner. The solutions thus obtained enjoy the property of containing only the least errors possible for a given probability. It is not an absolute minimum, as one has seemed to believe very often, but it is a minimum relative to the errors of observation and to the mode chosen for the combination of all the equations which they furnish. There would be able to be found other combinations more advantageous, and there would be to discuss them according to the cases.

The calculation of the magnitude of the existing errors and of the probability that they are able to have, is made according to the ordinary rules of the theory and entail only some analytic difficulties.

When there enters only one unknown in the equations, this calculation is exact, at least as to the theoretic foundations. But when there are many unknowns, the given rules in order to calculate the error and the probability of each of them furnish only the error and the probability that it would be able to have if it were alone, and some magnitudes that make the errors of the others.

Now, one of the first principles of the theory of probabilities, is that, when many events arrive simultaneously, the probability of the concurrence of these events is the product of the probabilities of each. So that the probability of the concurrence is inferior to the probability of each event taken apart: it is so much smaller, as there are more events.

Evidently, it is likewise of the errors of many unknowns; the probability that these errors remain all the time within certain limits is able to be only the product of the separated probabilities that each do not deviate from its proper limits, and, consequently, this probability of concurrence of the errors of limited magnitude must be notably inferior to the probability of the limits of each error considered in isolation, whatever the others are able to be.

It is therefore an imperfection that to assign as probability of error of an unknown making part of a system to determine, that which it would have if it were alone, instead of giving some rules in order to calculate the probability of the set of errors of the system which are not able, in reality, to be isolated from one another.

This defect becomes more evident still when one pays attention that the probability relative to a limit of error is so much greater, as the extent limited by this limit is greater: so that, if one wishes to conserve to the set of errors a little higher probability, it will be necessary to assign to each error some more extended limits than those which would require the same probability of the unknown were alone. Then therefore if one be stopped at these last limits, one has really only a very weak probability that they

2

are not passed. One will see, indeed, in that which follows, that it suffices with two unknowns in order that the extent of the errors be doubled for one same probability. The usual rules are born thence from the quite inexact ideas out of the knowledge of the unknown magnitudes that one is proposed to deduce from the observations, and they mislead at the same time on the value of these observations, in the same way as out of the number of good observations that it would be indispensable to make in order to arrive to an assigned result.

If it is permitted, on the subject of this number, of this multiplicity of the observations, quite extended from good observations, to remark that this is a domineering condition of human knowledge when the precision is not absolute. For men seek immediately so much of things, as the probability of concurrence of exact values of all these things is able to become very great only by the immense magnitude of the number of observations. It is, indeed, that magnitude and the continual repetition of the ordinary circumstances of life which sets beyond doubt the practices of common sense; and all the effectiveness of a sophism consists most often only to steal the view of the real multiplicity of the facts from a daily experience, by making believe in an artificial multiplicity of events possible without doubt, but which never arrive so to speak.

The defectiveness one time indicated, one sees that the concern is no longer to modify conveniently the demonstration of the methods, finally to calculate, no longer the probability of error of one unknown, but the probability of the set of errors. It was here to choose among the demonstrations.

To speak properly, Legendre has not given it at all. Only he endorses his process rather solidly on the advantages that it makes evident, and he insists principally on this that the arithmetic mean of the observations, that one takes ordinarily with confidence, is precisely only a particular case of the method of least squares.

It was Mr. Gauss who who first attached this method to the calculation of probabilities, some years after the publication of Legendre, and it was also in a work of Astronomy: *Theoria motus corporum cœlestium*, in-4°, 1809. The demonstration that Mr. Gauss gives reposes on the reciprocal of the remark by Legendre, that the arithmetic mean is deduced from his method, when this mean is able to take place. Reciprocally, if the arithmetic mean adopted generally is necessary, says Mr. Gauss, there results from it that a certain law of probability is necessary, and that the sole process to follow in order to combine the equations furnished by observation is the process which renders a minimum the sum of the squares of these equations. But one must agree that the hypothesis of the necessity to take the arithemtic mean of a mass of facts in order to obtain the most exact possible result, is totally gratuitous, and it could no longer could be admitted a priori as the same hypothesis of the necessity of the minimum of the squares. There is therefore thence only a proof restricted to the particular cases where the law of probability of errors, which always leads to the arithmetic mean, is encountered in the observations; and this is that which arrives very often by the nature of things. But, as one does not know most ordinarily, there exists no truly solid demonstration. Also, in a much more recent work, dedicated exclusively to the method of least squares, *Theoria combinationis observationum minimis erroribus obnoxiæ*, in-4°, 1823, Mr. Gauss has founded on some other considerations the use of this method. These are however only the considerations, and not the proofs; and one finds real demonstration only in the later works of Laplace.

It is in a Memoir published in 1811, that Laplace showed that, when the number of observations is large enough, the most restricted errors probably are given by the method of least squares. This Memoir is found among those of the Académie des Sciences for 1811, and it has been reproduced in the *Théorie analytique des Probabilités*, which appeared in 1812. The principles of the demonstration of Laplace are sheltered from all objection: the analytic means are able alone to raise some. They have each not at all the desirable rigor, and perhaps, in the applications, it would be acceptable to discuss well the particular cases which would be able to be presented. But, when one reflected that it permitted most often of some observations, added or subtracted, in order to render to the analytic expressions each their value, one recognized that the analysis of Laplace satisfied completely the general demonstrations. It is this analysis which will be employed in that which is going to follow, in order to supplement the omission of the probability of the concurrence of the diverse errors in the case of many unknowns, which presents itself nearly always. Only there will be made some simplifications, drawn especially from the beautiful work of Laplace and Mr. Gauss. If both have neglected this consideration of the concurrence of events, which increase so strongly the magnitude of the possible errors with a determined probability, they have at least furnished all the means to calculate it.

It would have been without doubt very useful to enter into more practical details and to give many examples of the calculations; but each application of this kind would require a rather long time.

One will see however how the probability of 1 million against 1 assigned by Laplace to the possible errors of the mass of Jupiter, of which he fixed the limit to $\frac{1}{100}$ of the value, must be reduced to 1160 against 1; even by supposing, as he has done, that the number of the equations employed, 129 only, permit applying the formula of approximation, and required not the addition of the terms habitually neglected. This diminution of the calculated probability suffices also, if one pays attention to the other defects which the 129 equations are able to contain, in order to make to cease the surprise that one had had to experience when it has seemed necessary to modify by $\frac{1}{50}$ the mass which seemed definitely calculated with a double precision. The new formulas put therefore here the calculation of the probabilities sheltered from the reproaches which have been able to be made to it a little hastily with this occasion. It is very probable that, in the other cases where this calculation appeared with defect, one would find equally the cause in some omission of the analysis or of the observation.

<center>§II.</center>

The application of the method of least squares supposes that a great number n of observations has given some results ω_1, ω_2, ..., ω_h, ..., ω_n which would have been able to be calculated in advance as linear function of many elements x_1, x_2, ..., x_i,..., x_m in number m, if these elements were known. Each observation furnishes then between the observed value and the corresponding calculated value, an equation such that

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

which is reported to the hth observation. The coefficients $a_{i,h}$ are known quantities, independent of the elements x_1, x_2, etc. The indices of which they are affected mark the

<center>4</center>

unknown and the equation to which the coefficient belongs. Thus, $a_{3,7}$ is the coefficient of the third unknown in the seventh equation.

The equations in number n that represent the expression (1) contain only m unknowns, it suffices that there is found m of them which do not return the ones into the others, in order that one is able to resolve them by the ordinary process. But as the observation does not give rigorously the value of the observed magnitude and as some error is always joined, it is clear that it will be more advantageous to make serve, by any combination, all the equations obtained in the determination of the unknowns, that there would not be able to be chosen m of them, perhaps without good motives. Good sense suffices, indeed, in order to presume that, in the reunion of so many values of which the errors are in all possible senses, some compensation will be made which will assure to the unknowns more exact values than the solutions of m isolated equations.

The most ordinary combinations of the equations of the first degree return to add them after having multiplied them respectively by some arbitrary factors, which serve to extricate the unknowns in being lent to diverse transformations. Calling $K_{i,1}$ the factor destined to the hth equation, one obtains a sum of equations such that

$$(2) \qquad x_1 S.K_{i,h}a_{i,h} + x_2 S.K_{i,h}a_{2,h} + \cdots + x_i S.K_{i,h}a_{i,h} + \cdots = S.K_{i,h}\omega_h,$$

and one will have immediately x_i if one subjects the n factors $K_{i,1}$, $K_{i,2},\ldots,K_{i,n}$ to the conditions after this, which render null the coefficients of all the unknowns, excepting the coefficient of x_i which they reduce to unity:

$$(3) \quad \begin{cases} S.a_{1,h}K_{i,h} = a_{1,1}K_{i,1} + a_{1,2}K_{i,2} + \cdots + a_{1,h}K_{i,h} + \cdots + a_{1,n}K_{i.n} = 0, \\ S.a_{2,h}K_{i,h} = a_{2,1}K_{2,1} + a_{2,2}K_{i,2} + \cdots + a_{2,h}K_{i,h} + \cdots + a_{2,n}K_{i.n} = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S.a_{i',h}K_{i,h} = a_{i',1}K_{i,1} + a_{i',2}K_{i,2} + \cdots + a_{i',h}K_{i,h} + \cdots + a_{i',n}K_{i.n} = 0, \\ S.a_{i,h}K_{i,h} = a_{i,1}K_{i,1} + a_{i,2}K_{i,2} + \cdots + a_{i,h}K_{i,h} + \cdots + a_{i,n}K_{i.n} = 1, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ S.a_{m,h}K_{i,h} = a_{m,1}K_{i,1} + a_{m,2}K_{i,2} + \cdots + a_{m,h}K_{i,h} + \cdots + a_{m,n}K_{i.n} = 0. \end{cases}$$

There remains then in equation (2) only

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

It is necessary to remark that the factors $K_{i,h}$ being in number n are not determined at all by the conditions (3), since these conditions are only in number m.

By taking another system of n factors, which one would designate likewise by $K_{i,h}$ one would obtain similarly the value of the first unknown x_i, provided that one submitted the n factors $K_{i,h}$ to m conditions similar to the conditions (3) which report to the unknown x_i.

The m unknowns will be therefore furnished by m systems of n factors, determined in part by m systems of conditions completely similar, and in number m for each unknown element. As the factors of a system do not enter into the others, the m^2 conditions will determine only an equal number of factors out of the mn employed, and $n-m$ will be independent in each system.

It is from these factors remained arbitrary that one will arrange in order to render the values

$$x_i = S.\omega_h K_{I,h}$$

the most exact possible; and it is this for why it is necessary to resort to the calculus of probabilities, under any form that one disguises it.

Effectively, the value of the coefficients $K_{i,h}$, whatever one wished to fix it, would modify by nothing the value found for x_i in function of these arbitraries, if the equations were rigorously exact. But, since it attaches always some error more or less great to the result of an observation, it would be necessary, in order to have some rigorous equations, to subtract from all the quantities ω_h the respective errors ε_h of which they are affected. One is not able, and the sum

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

remains affected subsequently with an error

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

The magnitude of this error will depend, one sees, on the choice of the coefficients K, and at the same time, on the law of probability of the possible errors, which will rule in the course of the observations.

There would be to make more than one remark on that which one must understand by this law of probability, and on the means to take the place of the ignorance where the observer is often found in this subject. But it is necessary to be limited here to the unique point in discussion. The law of errors ε_h will be supposed known.

Each of the values found for the unknowns $x_1, x_2, \ldots, x_m$ will be sullied by an error $r_1, r_2, \ldots, r_m$ respectively, which will be presented under the form which comes to be assigned to the error r_i of the element x_i:

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

The magnitude of each of these errors will result from that which it is being presented during the observations one system of errors $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_h, \ldots, \varepsilon_n$ rather than another, and one sees that all the errors r will be modified at the same time by the changes that the system of errors ε will be able to undergo. Each of the errors ε is unknown; but if the law is known, that is if one knows that to each error ε corresponds a certain probability, the probability of a given system $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ of errors in the set of observations will be also known and equal to the product of the unknown probabilities of each.

This product will be therefore the probability of the system (4) of errors $r_1, r_2, \ldots, r_m$ since these are all determined as soon as the errors ε_h of each observation receives given values.

6

Designating by $\phi\varepsilon$ the function which expresses the probability of the error ε, so that $\phi\varepsilon\,d\varepsilon$ is this infinitely small probability, one will have, between the limits in which the errors are able to be extended,

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

since it is the sum of the probabilities of all the possible cases. Moreover, if one calls μ_β the mean of the powers β of all the possible errors, one will have

$$\int \varepsilon^\beta\,d\varepsilon\,\phi\varepsilon = \mu_\varepsilon,$$

For example, the mean of all the possible errors will be expressed by

$$\mu_1 = \int \varepsilon\,d\varepsilon\,\phi\varepsilon;$$

the mean of their squares by

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

The probability of a system of values r_1, r_2, ..., r_m being the probability of the set of values ε_1, ε_2, ..., ε_n of which they are composed, will be expressed by the product

$$\phi\varepsilon_1\,d\varepsilon_1 \times \phi\varepsilon_2\,d\varepsilon_2 \times \phi\varepsilon_3\,d\varepsilon_3 \times \cdots \times \phi\varepsilon_n\,d\varepsilon_n.$$

There remains no more than to determine the systems of values of the quantities r_i for which this product is greatest, and beyond some limits of which it becomes as soon as not very probable that these quantities are able to be found. One will arrive evidently there if one employs the process of Laplace, which consists in multiplying it by a function of the errors r_1, r_2, ..., r_m and of their components ε_1, ε_2, ..., ε_n such that the subsequent integration, taken in all extent of the probability, leaves to subsist only the sum of the values corresponding to certain given magnitudes of r_1, r_2, ..., r_m. This process, of which Laplace has so much repeated the use without varying the form of it, is at base the same that Fourier, and more recently Mr. Lejeune-Dirichlet, have reproduced in another manner and with a so great success. In following Laplace, instead of forming first the necessary restrictive function, one will consider the product

$$P = {}_n\!\int d\varepsilon_1\,\phi\varepsilon_1 \times d\varepsilon_2\,\phi\varepsilon_2 \times \cdots \times d\varepsilon_h\,\phi\varepsilon_h \times \cdots \times d\varepsilon_n\,\phi\varepsilon_n$$

$$\times\, e^{\alpha_1\sqrt{-1}(\varepsilon_1 K_{1,1}+\varepsilon_2 K_{1,2}+\varepsilon_3 K_{1,3}+\cdots+\varepsilon_n K_{1,n})}$$

$$\times\, e^{\alpha_1\sqrt{-1}(\varepsilon_1 K_{2,1}+\varepsilon_2 K_{2,2}+\varepsilon_3 K_{2,3}+\cdots+\varepsilon_n K_{2,n})}$$

$$\cdots\cdots\cdots$$

$$\times\, e^{\alpha_1\sqrt{-1}(\varepsilon_1 K_{i,1}+\varepsilon_2 K_{i,2}+\varepsilon_3 K_{i,3}+\cdots+\varepsilon_n K_{i,n})}$$

$$\cdots\cdots\cdots$$

$$\times\, e^{\alpha_1\sqrt{-1}(\varepsilon_1 K_{m,1}+\varepsilon_2 K_{m,2}+\varepsilon_3 K_{m,3}+\cdots+\varepsilon_n K_{m,n})},$$

which contains by exposing the errors of m unknowns

$$r_i = \varepsilon_1 K_{i,1} + \varepsilon_1 K_{i,2} + \cdots + \varepsilon_1 K_{i,n} = S\varepsilon_h K_{i,h},$$

multiplied each by a special argument α_i, which is able to permit to distinguish it after the multiple integration, taken for all extent in which the errors ε are possible. If, after this integration relative to ε, one wishes to take account of the product P, it will be necessary to represent it as being no longer function but of r, and under the form

$$P = {}_m\!\int dr_1 dr_2 \ldots dr_m \Phi(r_1, r_2, \ldots, r_m) e^{r_1 \alpha_1 \sqrt{-1} + r_2 \alpha_2 \sqrt{-1} + \cdots + r_m \alpha_m \sqrt{-1}},$$

an expression in which the function Φ will be such, that it will reunite the probabilities relative to some equal values of the system of errors r. It is really this function, multiplied by the differential product $dr_1 dr_2 \ldots dr_m$, which is the probability of this system of errors, and, if one knew it, there would be no more but to integrate within the convenient limits in order to obtain the sought probability. In order to determine the function Φ, it suffices to multiply P successively by a series of m integrals of the form

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha_i e^{-r_1 \alpha_1 \sqrt{-1}}$$

for one knows that

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} d\alpha_i e^{-r_1 \alpha_1 \sqrt{-1}} \int du \Phi(u) e^{\alpha_i u \sqrt{-1}} = \Phi(r_i).$$

If therefore one repeats this nearly mechanical operation for all the arguments α_i and the errors r_i, in number m, one will obtain the law of probability of these errors. This will be the result Q that is here:

$$Q = \frac{1}{(2\pi)^m} \, {}_m\!\int_{-\infty}^{\infty} d\alpha_1 d\alpha_2 \ldots d\alpha_m e^{r_1 \alpha_1 \sqrt{-1} + r_2 \alpha_2 \sqrt{-1} + \cdots + r_m \alpha_m \sqrt{-1}} \times P.$$

Hence, $Q dr_1 dr_2 \ldots dr_m$ will be the probability of the system r_1, r_2, ..., r_m of the errors of the unknowns $x_1, x_2, \ldots, x_m$; and integrating between the limits which contain the magnitudes of these errors of which one will wish to know the common probability, one will obtain easily this probability, which will be

$$p = {}_m\!\int dr_1 dr_2 \ldots dr_m Q.$$

One is able to easily recognize, if one wishes, the composition of the restrictive function which has been introduced only successively, and it would be superfluous to be stopped. There remains to pursue the successive integrations.

To this effect, one modifies the product P by reuniting in the exponent all the terms which refer to one same observation, and separating the integrals relative to ε_1, ε_2, ...,

ε_n, one has

$$P = \int d\varepsilon_1 \phi \varepsilon_1 \, e^{\varepsilon_1(\alpha_1 K_{1,1} + \alpha_2 K_{2,1} + \alpha_3 K_{3,1} + \cdots + \alpha_m K_{m,1})\sqrt{-1}}$$

$$\times \int d\varepsilon_2 \phi \varepsilon_2 \, e^{\varepsilon_2(\alpha_1 K_{1,2} + \alpha_2 K_{2,2} + \alpha_3 K_{3,2} + \cdots + \alpha_m K_{m,2})\sqrt{-1}}$$

$$\int d\varepsilon_h \phi \varepsilon_h \, e^{\varepsilon_h(\alpha_1 K_{1,h} + \alpha_2 K_{2,h} + \alpha_3 K_{3,h} + \cdots + \alpha_m K_{m,h})\sqrt{-1}}$$

$$\times \cdots \cdots$$

$$\int d\varepsilon_n \phi \varepsilon_n \, e^{\varepsilon_n(\alpha_1 K_{1,n} + \alpha_2 K_{2,n} + \alpha_3 K_{3,n} + \cdots + \alpha_m K_{m,n})\sqrt{-1}}$$

Isolating one of these n integrals, one is able to write

$$\int d\varepsilon_h \phi \varepsilon_h \, e^{\varepsilon_h \sqrt{-1} \Sigma . K_{i,h}},$$

and likewise, finally for brevity, one is able to represent $\Sigma \alpha_i K_{i,h}$, a sum composed of m terms, by S_h, at least provisionally; one is able also to suppress the index h of ε, which has been employed only for clarity. One will have, by developing the exponential,

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

and, by integrating relatively to ε, which is explicit throughout, one will have

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

because of

$$\int d\varepsilon \phi \varepsilon = 1 \qquad \text{and} \qquad \int \varepsilon^\beta \, d\varepsilon \phi \varepsilon = \mu_\beta.$$

One will be able next to restore this series in exponential, this which gives

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

Nothing will be more easy, one sees it, than to suppose the law of probability $\phi \varepsilon$ variable from one observation to another, and to conserve this distinction in the rest of the calculus. But one will recognize soon that the method of least squares reposes on this hypothesis, that the means μ_1 and μ_2 of the possible errors ε and of their squares, not varying from one observation to another; so that the variation of the law of probability, reduces to the means of the superior powers, would have little interest in the actual question. It will be not therefore be taken account of it at all.

Under this hypothesis, the product P being formed of all similar integrals, will be presented under the form

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

9

the third and fourth terms of the exponent being likewise

$$- \frac{\mu_3 - 3\mu_2\mu_1 + 2\mu_1^4}{6}(S_1^3 + S_2^4 + \cdots + S_n^3)\sqrt{-1}$$

$$+ \frac{\mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4}{24}(S_1^4 + S_2^4 + \cdots + S_m^4).$$

If one examines at present each of the series which enter inter the different terms, one will see without difficulty that each variable α_i is multiplied in the one of them by all the factors $K_{i,h}$ corresponding to the same unknown x_i or to the same error r_i. They will be able therefore to be reunited under the symmetric forms, and, in order to be more clear, it will be lawful to attribute to the sums which comprehend n terms the sign S more specially, and to those which will contain only m, the sign Σ. One is able to observe that the first are those where the order of the observations in number n will bring forth so many different letters; and that the second are those which, depending on the order of the unknowns x_i in number m, could comprehend only this last number of letters.

One has thus, as previously,

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

where the sum Σ is relative to the index i, which itself varies only from 1 to m, the number of unknowns. There results from it, for the sum of the first powers of S_h,

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

that one would be able to write $S(\sigma \alpha_i K_{i,h})$, the finite sum S reporting back to the index h.

It is quite easy to see that

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

Likewise, because of

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

10

one has

$$S_1^2 + S_2^2 + \cdots + S_h^2 + \cdots + S_n^2$$
$$= \alpha_1^2(K_{1,l}^2 + K_{1,2}^2 + K_{1,3}^2 + \cdots + K_{1,h}^2 + \cdots + K_{1,n}^2)$$
$$= \alpha_2^2(K_{2,l}^2 + K_{2,2}^2 + K_{2,3}^2 + \cdots + K_{2,h}^2 + \cdots + K_{2,n}^2)$$
$$\cdots\cdots$$
$$= \alpha^2(K_{m,l}^2 + K_{m,2}^2 + K_{m,3}^2 + \cdots + K_{m,h}^2 + \cdots + K_{m,n}^2)$$
$$= 2\alpha_1\alpha_2(K_{1,l}K_{2,1} + K_{1,2}K_{2,2} + \cdots + K_{1,h}K_{2,h} + \cdots + K_{1,n}K2,n)$$
$$= 2\alpha_1\alpha_3(K_{1,l}K_{3,1} + K_{1,2}K_{3,2} + \cdots + K_{1,h}K_{3,h} + \cdots + K_{1,n}K3,n)$$
$$\cdots\cdots$$
$$= 2\alpha_i\alpha_{i'}(K_{i,1}K_{i',1} + K_{i,2}K_{i',2} + \cdots + K_{i,h}K_{i',h} + \cdots + K_{i,n}Ki',n)$$
$$\cdots\cdots$$
$$= 2\alpha_{m-1}\alpha_m(K_{m-1,1}K_{m,1} + \cdots + K_{m-1,h}K_{m,h} + \cdots + K_{m-1,n}Km,n),$$

or, for brevity,

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

Here the sums of the products are able to be as easily designated as the sums of the powers, since there is in each sum only a single variable. One would abbreviate yet further by writing

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

by having care, in order to form the second sum Σ, by taking the indices i and i' only under the condition $i < i'$, or else by erasing the coefficient 2 which precedes it.

Passing to S_h^3, it is clear that, following the analogous remarks

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

It is nearly useless to say that it is necessary to understand by the sums Σ relative to the indices i, when they are applied to the products, all the possible combinations, without double use.

Arriving finally to S_h^4, one will have absolutely in the same manner,

$$S_1^4 + S_2^4 + \cdots + S_h^4 + \cdots + S_n^4 = S(\Sigma \alpha_i K_{i,h})^4$$

$$= S \begin{pmatrix} \Sigma \alpha_i^4 K_{i,h}^4 + 4\Sigma \alpha_i^3 \alpha_{i'} K_{i,h}^3 K_{i',h} + 6\Sigma \alpha_i^2 K_{i,h} K_{i',h}^2 \\ + 12\Sigma \alpha_i^2 \alpha_{i'} \alpha_{i''} K_{i,h}^2 K_{i',h} K_{i'',h} \\ + 24\Sigma \alpha_i^2 \alpha_{i'} \alpha_{i''} \alpha_{i'''} K_{i,h}^2 K_{i',h} K_{i'',h} K_{i''',h} \end{pmatrix}$$

$$= \Sigma(\alpha_i^4 K_{i,h}^4 + 4\Sigma(\alpha_i^3 \alpha_{i'} SK_{i,h}^3 K_{i',h}) + 6\Sigma(\alpha_i^2 \alpha_{i'}^2 SK_{i,h}^2 K_{i',h}^2)$$
$$+ 12\Sigma(\alpha_i^2 \alpha_{i'} \alpha_{i''} SK_{i,h}^2 K_{i',h} K_{i'',h})$$
$$+ 24\Sigma(\alpha_i^2 \alpha_{i'} \alpha_{i''} \alpha_{i'''} K_{i,h}^2 K_{i',h} K_{i'',h} K_{i''',h}).$$

It was useful to show the forms of the third and fourth terms of the exponent, although one was able to understand for what very different motives they are neglected in the applications.

By designating by T_1, T_2, T_3, T_4 the terms of the exponent of e, which come to be developed, the product P is able to be written

$$P = e^{T_1 \sqrt{-1} - \frac{1}{2}T_2 - \frac{1}{6}T_3\sqrt{-1} + \frac{1}{24}T_4 + \cdots}$$

$$= e^{T_1\sqrt{-1} - \frac{1}{2}T_2} \left(1 - \frac{1}{6}T_3\sqrt{-1} + \frac{1}{24}T_4 + \cdots - \frac{1}{72}T_3^2 + \cdots \right),$$

by stopping at the terms of the fourth degree in α_i, as previously.

There will remain therefore no more to integrate than relative to these variables, in order to obtain the function Q, which is presented under the form

$$Q = \frac{1}{(2\pi)^m} m \int_{-\infty}^{\infty} d\alpha_1 d\alpha_2 \ldots d\alpha_m e^{-\Sigma r_1 \alpha_1 \sqrt{-1} - r_2 \alpha_2 \sqrt{-1} - \cdots - r_m \alpha_m \sqrt{-1}} \times P$$

$$= \frac{1}{(2\pi)^m} m \int_{-\infty}^{\infty} d\alpha_1 d\alpha_2 \ldots d\alpha_m e^{-\Sigma r_i \alpha_i \sqrt{-1} + \sqrt{-1}\mu_1 \Sigma(\alpha_i SK_{i,h}) - \frac{1}{2}T_2} \left(1 - \frac{\sqrt{-1}}{6}T_3 - \cdots \right).$$

It is clear that the first powers of α can be put in common factor, and that the exponent of e is equal to

$$-\alpha_i \sqrt{-1}(r_1 - \mu_1 SK_{1,h}) - \alpha_2 \sqrt{-1}(r_2 - \mu_1 SK_{2,h}) - \cdots$$
$$- \alpha_m \sqrt{-1}(r_m - \mu_1 SK_{m,h}) - \frac{1}{2}T_2.$$

Laplace and many other geometers have supposed the mean μ_1 reduced to zero. But this condition simplifies too little the calculations, and it requires in some sort to begin them when this mean is not able to be considered null. One alleges, indeed, that, in a system of of well directed observations, this mean μ_1, which is a true constant error, had to be recognized and subtracted from the observed values. Is it not possible, on the contrary, that one makes the observations only in order to determine the constant errors, and that one must conserve μ_1, which is then the sought thing. Here, in order

12

to conserve the facility to take this mean again, although the research of its value, according to the same observations, must not be a subject of examination, it will suffice to reduce to a single letter the terms of the form $r_i - \mu_1 SK_{i,h}$. One will introduce thus instead of the errors r of other quantities which will differ from them only by constant quantities. One is able to write, for example,

$$(5) \qquad\qquad r_i - \mu_1 SK_{i,h} + \rho_i \sqrt{2(\mu_2 - \mu_1^2)};$$

so that an error r_i will be calculated by the relation

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

and

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

In this manner, one sees that

$$\frac{1}{(2\pi)^m}\ {}_m\!\int_{-\infty}^{\infty} d\alpha_1 d\alpha_2 \ldots d\alpha_m\, e^{-\sqrt{-1}\sqrt{2(\mu_2-\mu_1^2)}\Sigma(\rho_i\alpha_i - \frac{1}{2}T_2}\left(1 - \frac{\sqrt{-1}}{6}T_3 - \cdots\right).$$

If one makes the α_i subject to an analogous change, that is that one puts

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

there will result from it

$$dr_i d\alpha_i = d\rho_i \sqrt{2(m_2 - \mu_1^2)} \times \frac{dz_i}{\sqrt{\frac{1}{2}(\mu_2 - \mu_1^2)}} = 2 d\rho_i dz_i.$$

Moreover, the factor $\frac{1}{2}(\mu_2 - \mu_1^2)$ will vanish from the term $\frac{1}{2}T_2$; so that the probability of the system of errors designated by ρ_i or by

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

which is equal to

$$p = {}_m\!\int dr_1 dr_2 \ldots dr_m Q,$$

will be calculated by the formula

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

by writing Z_2, Z_3, Z_4, for brevity, instead of the following terms that give the change of the α_i by $\frac{z_i}{\sqrt{\frac{1}{2}(\mu_2-\mu_1^2)}}$. One knows

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

13

and one will have more simply

$$A_2 = \Sigma(Z_i^2 SK_{i,h}^2) + 2\Sigma(z_i z_{i'} SK_{i,h} K_{i',h}).$$

Likewise,

(6)
$$
\begin{cases}
Z_3 = 2^{\frac{3}{2}} \dfrac{\mu_3 - 3\mu_3\mu_1 + \mu_1^3}{(\mu_2 - \mu_1^2)^{\frac{3}{2}}}
\begin{cases}
\Sigma(z_i^3 SK_{i,h}^3) + 3\Sigma(z_i^2 z_{i'} SK_{i,h}^2 K_{i',h}) \\
+6\Sigma(z_i z_{i'} z_{i''} SK_{i,h} K_{i',h} K_{i'',h})
\end{cases}, \\[2em]
Z_4 = 4\dfrac{\mu_1 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^3}{(\mu_2 - \mu_1)^2}
\begin{cases}
\Sigma(z_i^4 SK_{i,h}^4) + 4\Sigma(z_i^3 z_{i'} SK_{i,h}^3 K_{i',h}) \\
+6\Sigma(z_i^2 z_{i'}^2 SK_{i,h}^2 K_{i',h}^2) \\
+12\Sigma(z_i z_{i'} z_{i''} SK_{i,h}^2 K_{i',h} K_{i'',h}) \\
+24\Sigma(z_i z_{i'} z_{i''} z_{i'''} SK_{i,h} K_{i',h} K_{i'',h} K_{i''',h})
\end{cases}
\end{cases}.
$$

It is quite easy to note that the sums designated by S containing n terms will be of the form nM, M being a common value of these terms. So that it will suffice, in order to render very small the parts which contain them, to give to the variables z in the denominators in which are found n to a power superior to the first, and a common value M very nearly equivalent. One knows that it is there that which arrives in the analysis of Laplace.

In order to abridge yet the writing of the exponent of e and in order to develop it, one will be able to put indifferently

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

this which entails

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

In this manner, the exponent will be

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

If now one makes at the same time

(8)
$$
\begin{cases}
\beta_1 = z_1 h_{1,1} + z_2 h_{1,2} + z_3 h_{1,3} + \cdots + z_i h_{1,i} + \cdots + z_m h_{1,m} + t_1\sqrt{-1} \\
\beta_1 = \qquad\quad + z_2 h_{2,2} + z_3 h_{2,3} + \cdots + z_i h_{2,i} + \cdots + z_m h_{2,m} + t_2\sqrt{-1} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\beta_1 = \qquad\qquad\qquad\qquad\qquad\qquad + z_i h_{i,i} + \cdots + z_i h_{i,m} + t_i\sqrt{-1} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\beta_1 = \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad z_m h_{m,m} + t_m\sqrt{-1},
\end{cases}
$$

and that after having raised to the square each of the new varialbes β_i, one makes the sum of them, there will be found only four kinds of terms:

In z_i^2 of which the coefficient will be

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

and will contain i terms;

In $2z_i z_{i'}$ of which the coefficient will be likewise

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

containing i terms;

In $-2z_i\sqrt{-1}$ multiplied by the variable expression

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

containing i terms;

And, finally, the m squares of the t_i taken negatively.

If therefore one makes the coefficients of the z_i^2 equal to those that these variables have in the exponent of e, and that one acts likewise for the coefficients of the double products; this which will give

(9)
$$\left\{
\begin{aligned}
&b_{1,1}=h_{1,1}^2, \quad b_{1,2}=h_{1,1}h_{1,2}, \quad b_{1,3}=h_{1,1}h_{1,3}, \\
&\qquad\qquad b_{2,2}=h_{1,2}^2+h_{2,2}^2 \quad b_{2,3}=h_{1,2}h_{1,3}+h_{2,2}h_{2,3} \\
&\qquad\qquad\qquad\qquad\qquad\qquad b_{3,3}=h_{1,3}^2+h_{2,3}^2+h_{3,3}^2, \\
&b_{1,4}=h_{1,1}h_{1,4}, \\
&b_{2,4}=h_{1,2}h_{1,4}+h_{2,2}h_{2,4}, \\
&b_{3,4}=h_{1,3}h_{1,4}+h_{2,3}h_{2,4}+h_{3,3}h_{3,4}, \\
&b_{4,4}=h_{1,4}^2 \quad +h_{2,4}^2 \quad +h_{3,4}^2 \quad +h_{4,4}^2, \\
&b_{1,i}=h_{1,1}h_{1,i}, \\
&b_{2,i}=h_{1,2}h_{1,i}+h_{2,2}h_{2,i}, \\
&b_{3,i}=h_{1,3}h_{1,i}+h_{2,3}h_{2,i}+h_{3,3}h_{3,i}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&b_{i,i'}=h_{1,i}h_{1,i}+h_{2,i}h_{2,i'}+h_{3,i}h_{3,i'}+\cdots+h_{i',i'}h_{i',i}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&b_{i,i}=h_{1,i}^2 \quad +h_{2,i}^2 \quad +h_{3,i}^2 \quad +h_{4,i}^2+\cdots+h_{i,i},
\end{aligned}
\right.$$

relations of which the law is easy to know; next, that one subjects the new variables t_i, which alone multiply the first powers of the z_i, to the conditions

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

it is manifest that the sum of the squares of the variables β_i will be, save the sign, equal to the function of the z_i and of the ρ_i which formed the exponent (7), less the sum of

15

the squares of the variables t_i. Thus the exponent will be able to be replaced by the negative sum of the squares of the new variables β_i and t_i.

On the other hand, the relations (8) and (10) assure immediately

$$d\beta_1 d\beta_2 \ldots d\beta_m = dz_1 dz_2 \ldots dz_m h_{1,1} h_{2,2} \ldots h_{m,m},$$
$$d\rho_1 d\rho_2 \ldots d\rho_m = dt_1 dt_2 \ldots dt_m h_{1,1} h_{2,2} \ldots h_{m,m};$$

so that

$$d\rho_1 d\rho_2 d\rho_m \times dz_1 dz_2 \ldots dz_m = dt_1 dt_2 \ldots dt_m \times d\beta_1 d\beta_2 \ldots d\beta_m.$$

Hence, the variables are found completely separated in the multiple integral, the limits of the β_i remaining infinite, as those of the z_i or of the α_i from which they derive; and one obtains the probability

$$p = \frac{1}{\pi^m} \left\{ \begin{array}{c} {}_m \int dt_1 dt_2 \ldots dt_m \, e^{-t_1^2 - t_2^2 - \cdots - t_m^2} \\ \times \quad {}_m \int_{-\infty}^{\infty} d\beta_1 d\beta_2 \ldots d\beta_m \, e^{-\beta_1^2 - \beta_2^2 - \cdots - \beta_m^2} \left(1 - \frac{\sqrt{-1}}{6} Z_3 + \frac{1}{24} Z_4 - \frac{1}{72} Z_3^2 \right) \end{array} \right\}.$$

In this new expression, the multiple integral relative to β is obtained without other difficulty than to write, considering the known values

$$\int_{-\infty}^{\infty} d\beta \, e^{-\beta^2} = \sqrt{\pi}, \qquad \int_{-\infty}^{\infty} \beta^{2i+1} d\beta \, e^{-\beta^2} = 0,$$

$$\int_{-\infty}^{\infty} \beta^{2i} d\beta \, e^{-\beta^2} = \frac{1.3.5 \ldots 2i-1}{2.2.2 \ldots 2} \sqrt{\pi}.$$

One will have therefore

$$_m \int_{-\infty}^{\infty} d\beta_1 d\beta_2 \ldots d\beta_m \, e^{-\beta_1^2 - \beta_2^2 - \cdots - \beta_m^2} = (\sqrt{\pi})m.$$

But, when the function under the sign will be multiplied by the quantities represented in equation (6) by Z_3 and Z_4 or by Z_3^2, the result will be less simple. One knew immediately that it will contain only the parts of these expression multiplied by some even powers of β, the odd powers vanish. There remains that Z_3 will be changed into a function $B_3(\sqrt{\pi})^m$ of the t_i, containing only the first powers, second and third of these new variables; Z_4 will become $B_4(\sqrt{\pi})^m$ and will contain only the zero, first, second, third and fourth powers; finally Z_3^2 becoming $B_6(\sqrt{\pi})^m$ will offer the zero, first, second, third, fourth, fifth and sixth powers of t_i. These terms will not be developed, uniquely for brevity.

After all these remarks, there remains for the probability

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

One sole difficulty will seem perhaps to hinder this transformation (if one deviates, well understood, those which hold to the process of Laplace, of which there will not be question here). The difficulty of which there is concern, it is that which would carry on the determination of the arbitraries $h_{i,i'}$, which are deduced from the coefficients $b_{i,i'}$

16

only through some equations of the second degree (9), and would be able to become imaginaries. But it is thence that which could not arrive, seeing that the function of the second degree in z_1, z_2, ..., z_m is a sum of squares, as one is able to convince oneself; and this form permits to apply the following march in a Memoir presented to the Académie des Sciences in 1834 (Volume VI of the *Recueil des Savants étrangers*). The successive transformation of the variables would put into evidence the sums of squares which only enter into the radicals, and render them necessarily reals.

Restoring through all that which precedes to no longer contain but some integrals of the form

$$\int dt\, e^{-t^2} t^n,$$

nothing will be easier than to obtain the probability p when the limits of the variables t_i will be deduced, by means of relations (10), of those that one will wish to assign to the errors ρ_i. The magnitudes of these last are arbitrary; but one sees that the magnitude assigned to one of them will influence on the form of all the others, or all at least on the form of those which follow it in the order of the relations (10). If one would wish that the variables ρ_i be only proportionals to the terms $t_i h_{i,i}$, which enter only once into each respectively, one would arrive only by assigning to the variables t_i in the integral p of the limits which would depend on these variables; if also that it would be possible to evaluate p only by some very painful approximations. But there exists some combinations of errors of which the probability is able, to the contrary, to be expressed without too many difficulties.

These are those for which the exponent of e is able to take only the values inferior to a certain constant γ^2. One sees that it is necessary to integrate p under this hypothesis for all the values of t_i which satisfy the condition

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

One will arrive there through several well known methods. There could be found some interest, fifteen or twenty years ago, when from the first research on this subject, to develop these processes. It will suffice today to show that there is need only of very simple integrations.

By operating successively on the variables t_i, one would have to integrate each of them between the equal limits and of contrary signs

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

Now, between similar limits, it is clear that the integration relative to the function which has been designated by B_3 will give a null result. B_3 arises effectively from Z_3, where there enters only some products of odd degree of the z_i and where there remains only the terms affected of even powers of β_i, that is those which offer some odd powers of t_i. They will give place to some integrations of the form

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

Thus, all that which is relative to B_3 will vanish.

As for B_4 and B_6, it is found from the products of even powers of t_i, and the odd powers will be affected with $\sqrt{-1}$; those will make the imaginary terms vanish where they are encountered. But the integration will leave to subsist all the terms where there enters only some even powers. The result will not be null therefore. Only, all the terms of these functions acquire, by the substitution of the values of z_i as function of β_i and t_i of the divisors which contain the squares of the finite sums of the factors $K_{i,h}$, which are in number n in each sum. Thus, when the number n of observations will be very great, each of the terms will be of very small order $\frac{1}{n}$. It will be superfluous to be arrested on this point in the present question. One knows rather that it is precisely there the form that the analysis of Laplace brings forth. There would be only to show it by the same calculation, no other difficulty than the length of the writing of these rather complex expressions. However, it is necessary to observe that, when there are m elements or unknowns, and not one alone, the number of these terms of order $\frac{1}{n}$ depends on the number m. So that the set of these terms is only of the order of $\frac{m}{n}$. It is therefore indispensable, since one neglects constantly this part of the integral, to be assured that $\frac{n}{m}$ is a great number; great enough especially in order to counterbalance the influence of the powers of γ which enter into these terms to γ^6. There will be therefore in the applications a condition to not forget, that $\gamma^6 \frac{m}{n}$ remains of the order of the quantities that one will believe to be able to be permitted to neglect. If it is not thus, one would be able to be assured some exactitude only for some small values of γ; and it is this which has not always been made with the attention that this point merits.

After these considerations, there remains no more to be occupied but with the formula of approximation

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

Between the equal limits and of contrary signs, it is palpable that the negative values of the t_i give some results equal to those that produce the positive values. One is able therefore to double the result of each integration, or to multiply the integral above by 2^m, and make the calculation only from zero to the positive limits. This is that which is executed easily by transforming first one of the variables t_m for example, by means of the relation

$$t_1^2 + t_2^2 + \cdots + t_m^2 = u^2,$$
$$t_m = \sqrt{u^2 - t_1^2 - t_2^2 - \cdots - t_{m-1}^2},$$
$$t_m dt_m = u du,$$
$$dt_m = \frac{u du}{\sqrt{u^2 - t_1^2 - t_2^2 - \cdots - t_{m-1}^2}},$$

which carries for the limits of u the values zero to γ. The expression of p restores thus to

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

and under this form it is sufficient from the truly elementary single integral,

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

in order to reduce the multiple integral. The gamma functions have here for object only to abbreviate the writing. The formula which contains them is properly that which serves to reduce the terms contained in B_4 and B_6. It is necessary to make $\eta = 0$, in order to apply it to the approximate value of p.

It is evident, in this particular case,

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

Employing this expression $(m-1)$ times, by taking care to give successively to a the values $(u^2 - t_1^2 - t_2^2 - \cdots - t_i)$, which result from the disappearance of one of the variables t_i in each operations, one will obtain

$$_{m-1}\!\int \frac{dt_1 dt_2 \ldots dt_{m-1}}{\sqrt{u^2 - t_1^2 - t_2^2 - \cdots - t_{m-1}^2}} = u^{m-2}\left(\frac{\sqrt{\pi}}{2}\right)^{m-1} \frac{\Gamma\left(\frac{1}{2}\right)}{\Gamma(1)} \times \frac{\Gamma(1)}{\Gamma\left(\frac{3}{2}\right)} \times \frac{\Gamma\left(\frac{3}{2}\right)}{\Gamma(2)} \times \cdots \times \frac{\Gamma\left(\frac{m-1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}$$

$$= u^{m-1}\left(\frac{\sqrt{\pi}}{2}\right)^m \frac{2}{\Gamma\left(\frac{m}{2}\right)}$$

Substituting into p, there results from it

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

This quite simple result receives two very different forms according as the number m of the elements or of the unknowns is even or odd. It is most worthy to note that from the point of view of the numerical calculus, the probability is more simply expressed when there is presented an even number of unknowns, than even when there is concern of one alone.

One knows effectively that

$$\int_0^\gamma u^{m-1} du\, e^{-u^2} = -\gamma^{m-2}\frac{e^{-\gamma^2}}{2} - \frac{m-2}{2}\gamma^{m-4}\frac{e^{-\gamma^2}}{2} - \frac{m-2}{2}\cdot\frac{m-4}{2}\gamma^{m-6}\frac{e^{-\gamma^2}}{2} - \cdots$$

$$-\frac{m-2}{2}\cdot\frac{m-4}{2}\cdots\frac{m-2i}{2}\gamma^{m-2i-2}\frac{e^{-\gamma^2}}{2}$$

$$+\frac{m-2}{2}\cdot\frac{m-4}{2}\cdots\frac{m-2i}{2}\frac{m-2i-2}{2}\int_0^\gamma u^{m-2i-3} du\, e^{-u^2}.$$

So that, for $m = 2g$,

$$\int_0^\gamma u^{2g-1} du\, e^{-u^2} = \left\{ \begin{array}{l} -\gamma^{2g-2} - \frac{2g-2}{2}\gamma^{2g-4} - \frac{2g-2}{2}\cdot\frac{2g-4}{2}\cdot\gamma^{2g-6} - \cdots \\ \qquad - \frac{2g-2}{2}\cdot\frac{2g-4}{2}\cdots\frac{2}{2} \end{array} \right\} \frac{e^{-\gamma^2}}{2}$$

$$+ \frac{2g-2}{2}\cdot\frac{2g-4}{2}\cdots\frac{2}{2}\cdot\frac{1}{2},$$

that one is able to write

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

and, for $m = 2g - 1$,

$$\int_0^\gamma u^{2g-2} du\, e^{-u^2} = \left\{ \begin{array}{l} -\gamma^{2g-3} - \frac{2g-3}{2}\gamma^{2g-5} - \frac{2g-3}{2}\cdot\frac{2g-5}{2}\cdot\gamma^{2g-7} - \cdots \\ \qquad - \frac{2g-3}{2}\cdot\frac{2g-5}{2}\cdots\frac{3}{2}\gamma \end{array} \right\} \frac{e^{-\gamma^2}}{2}$$

$$+ \frac{2g-3}{2}\cdot\frac{2g-5}{2}\cdots\frac{3}{2}\cdot\frac{1}{2}\int_0^\gamma du\, e^{-u^2},$$

which will be written also, because of $\Gamma\left(\frac{2g-1}{2}\right) = \frac{2g-3}{2}\cdot\frac{2g-5}{2}\cdots\frac{3}{2}\cdot\frac{\sqrt{\pi}}{2}$,

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

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

One will have therefore finally

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

Such are the expressions approached from the probability that the errors ρ_i (10), expressed in functions of the variables t_i, are not able to be understood beyond the limits assigned by the condition

$$(12) \qquad\qquad t_1^2 + t_2^2 + \cdots + t_m^2 < \gamma^2.$$

If there were only three variables, one would express this condition by saying that the points that they determine, when one regards them as some rectangular coordinates, are not exited from the sphere of radius γ. That analogy leads therefore to say that the variables t are not able to exit from the extent of the analytic relationship

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

20

which permits none to exceed $\pm\gamma$.

One recognizes without doubt already, in the general forms of the probability (11), how much it will be different according to the number of the elements; one sees that it diminishes for one same value of γ, in measure as the number m of the elements increase, and it is that which must be, conformably to the principles of the calculus of the probabilities reported at the beginning of this work.

It was there the principal end that there was concern to attain; and, up to a certain degree, it was distant to be stopped there. For one recognized without difficulty that the coefficents $K_{i,h}$ are able to be those that give the method of least squares, since those will satisfy the little numerous conditions (3), which alone are imposed on the factors $K_{i,h}$, in order that there results values of the unknowns from it.

But, as the expressions of the errors of these elements are here different from those which are found widespread in the many works since the theories of Mr. Gauss and of Laplace, one would demand, without any doubt, if the results of the preceding analysis render or not necessary the factors $K_{i,h}$ particular to the method of least squares. It seems therefore indispensable to prove that, under the conditions posed, the value of the probability, or rather the expressions of the errors in which all the difficulties of the question are reunited, since the probability is only a pure constant that one is able to calculate in advance and of which one has some Tables, to which the concern is only to apply some errors of given magnitude; it seems indispensable to prove that the errors will be the smallest possible when one will employ in the elimination of the unknowns the factors $K_{i,h}$ assigned by the method of least squares. One will see then clearly that the omission committed on the value of the probability of the errors would alter only this probability, and that the modification which repairs it (formulas 11) carries no change in the mode of elimination prescribed by the method. This is that which is going to be done, before preceding to some numerical applications.

But since at present there is agreement to retell it, the formulas (11) and (10) repair the omission completely. One recognized, moreover, that they apply to all the systems of elimination by means of factors or of linear combinations of the equations given by observation. They permit thus to calculate the error and the probability in the number of cases where one does not wish to take all the pain that the method of least squares requires.

They contain effectively the calculation of the probability of a system of linear functions r_i (4) of errors ε_h submitted to a law of probability $\phi(\varepsilon)$, whatever be the origin of these functions, and whatever is able to be the determinations of the factors $K_{i,h}$. These formulas offer therefore the solution of a class of rather extended problems, in which the coexistence of many functions is necessary, and where one would know since then how to have in view only the probability of their set. It is besides rather worthy of remark that the expressions of the probability are the same as if the functions r_i or ρ_i were independent, and if there was concern only to make the product of their particular probabilities, next to determine from them the value under the condition (12); it is well known that, in this case, the relations (10) would not subsist. But it is necessary to be arrested in these indications, and to return to the method of least squares.

The concern is to research what are the values of the arbitrary factors $K_{i,h}$ which will give the narrowest limits to the errors ρ_i for a probability determined by the constant γ.

The question would be to know how to resolve without knowledge of the extent of the values only to be able to take an error ρ_i depending on the t_i, according as the formulas (10), when the t_i are subject to the condition (12),

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

Let one suppose first the variables t_i linked by the relation

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

and let one seek the greatest value that is able to receive

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

which contain only i of the m variables t. One will be able to consider these i variables, whatever be besides the $m - i$ remaining as linked by the equation

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

Hence, one will have, for the greatest value of ρ_i, the conditions

$$\frac{d\rho_i}{dt_{i'}} = h_{i',i} + h_{i,i}\frac{dt_i}{dt_{i'}} = 0,$$

$$t_{i'}dt_{i'} + t_i dt_i = 0,$$

$t_{i'}$ being any one of the i first variables, v being regarded as constant relative to these variables, and t_i being found function of the $i - 1$ which precede it. One deduces thence

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

It will be necessary therefore that the $t_{i'}$ are proportional to the $h_{i',i}$ corresponding. Putting

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

one concludes from it

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

Now, by virtue of the last of the relations (9), the sum which multiplies f^2 is precisely equal to the coefficient $b_{i,i}$; one has therefore

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

The greatest value of ρ_i will be

$$
\begin{aligned}
\rho_i &= f(h_{1,i}^2 + h_{2,i}^2 + \cdots + h_{i,i}^2) \\
&= b_{i,i}\sqrt{\frac{v}{b_{i,i}}} = \sqrt{vb_{i,i}}
\end{aligned}
$$

This value will be, indeed, the greatest absolutely, for one is able to be assured that the sign of the second total differential is negative or positive, according as one takes the factor f positive or negative, that is according as one takes ρ_i to its positive limit or to its negative limit. Instead of calculating the second differential, that which is not very complicated, one will abridge however, by substituting $fh_{i',i} + \delta_{i'}$ into ρ_i instead of $t_{i'}$, this which gives

$$
\begin{aligned}
\rho_i &= f(h_{1,i}^2 + h_{2,i}^2 + \cdots + h_{i,i}^2) + \delta_1 h_{1,i} + \delta_2 h_{2,i} + \cdots + \delta_i h_{i,i} \\
&= fb_{i,i} + \delta_1 h_{1,i} + \delta_2 h_{2,i} + \cdots + \delta_i h_{i,i}.
\end{aligned}
$$

Making the same substitution into v, there comes

$$
\begin{aligned}
v = {}& f^2(h_{1,i}^2 + h_{2,i}^2 + \cdots + h_{i,i}^2) \\
& + 2f(\delta_1 h_{1,i} + \delta_2 h_{2,i} + \cdots + \delta_i h_{i,i}) \\
& + \delta_1^2 + \delta_2^2 + \cdots + \delta_i^2,
\end{aligned}
$$

and, because of the value of f^2, there remains only

$$
\delta_1 h_{1,i} + \delta_2 h_{2,i} + \cdots + \delta_i h_{i,i} = -\frac{1}{2f}(\delta_1^2 + \delta_2^2 + \cdots + \delta_i^2).
$$

Thus,

$$
\rho_i = f\left[b_{i,i} - \frac{1}{2}(\delta_1^2 + \delta_2^2 + \cdots + \delta_i^2)\right].
$$

The value of

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

is therefore a maximum or a minimum, according as the sign of f, that is that it is also greatest absolutely.

It is not necessary to forget, however, that v has for superior limit u, this which supposes that all the t from t_{i+1} to t_m are nulls. According to that, the extreme limits of ρ_i are $\pm\sqrt{ub_{i,i}}$. Next, as u must be remaining inferior to γ^2, one sees that, definitely, for a probability determined by the constant γ, this constant fixes very simply the extent of the limits of the errors ρ_i under the form

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

Moreover, this form was quite easy to foresee according to the first of the relations (10) and the symmetry of all this calculus. There is, indeed, in the relations (10), a real symmetry that depends on this that one of the errors is able to occupy, at will, any rank in the transformations; so also, that it is permitted to affirm that that which is able to

23

be recognized out of the error expressed most simply, holds necessarily for all. But the asymmetrical appearance of the relations posed has made prefer a more striking proof.

If one recalls actually that $b_{i,i}$ is made only to replace $S.K_{i,h}^2$, one will have, for the limits of ρ_i,

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

There remains, since then, in order to restrict the most possible this extent of the errors ρ_i, only to find the means to render a minimum the sum of the squares $S.K_{i,h}^2$. Now one knows that it is there one of the properties of the factors by which one effects the elimination in the method of least squares.

One can scarcely, in order to demonstrate it briefly, take another way than the immediate comparison of the coefficients $K_{i,h}$ which serves to obtain the value

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

with the factors which would give the same unknown, according to the method of least squares.

Calling these factors $A_{i,h}$, it is clear that they will satisfy, as the $K_{i,h}$, the equations (3), in which it sufficed to substitute one letter for the other, and that one would find

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

But at the same time, if one had operated directly according to the prescriptions of the method by forming m equations, by aid of the successive multiplication of the n equations (1) by the coefficients of each unknown, these m equations would be

$$x_1 Sa_{1,h}a_{1,h} + x_2 Sa_{1,h}a_{2,h} + \cdots + x_i Sa_{1,h}a_{i,h} + \cdots = Sa_{1,h}\omega_h,$$
$$x_1 Sa_{2,h}a_{1,h} + x_2 Sa_{2,h}a_{2,h} + \cdots + x_i Sa_{2,h}a_{i,h} + \cdots = Sa_{2,h}\omega_h,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x_1 Sa_{i,h}a_{1,h} + x_2 Sa_{i,h}a_{2,h} + \cdots + x_i Sa_{i,h}a_{i,h} + \cdots = Sa_{i,h}\omega_h,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x_1 Sa_{m,h}a_{1,h} + x_2 Sa_{m,h}a_{2,h} + \cdots + x_i Sa_{m,h}a_{i,h} + \cdots = Sa_{m,h}\omega_h,$$

and one would eliminate by multplying by m factors B_1, B_2,..., B_i,..., B_m, which would be subject to as many relations, such that one had

$$x_i' = B_i Sa_{1,h}\omega_h + B_2 Sa_{2,h} + \cdots + B_m Sa_{m,h}\omega_h,$$

or the same value as by the coefficients $A_{i,h}$.

One will reunite in common factor all that which multiplies ω_1, ω_2, ..., ω_n, and there will result

$$x_i' = \omega_1 \Sigma B_{i'}a_{i',1} + \omega_2 \Sigma B_{i'}a_{i',2} \quad + \omega_3 \Sigma B_{i'}a_{i',3} + \cdots + \omega_h \Sigma B_{i'}a_{i',h} + \cdots$$
$$+ \omega_n \Sigma B_{i'}a_{i',n}.$$

By bringing together this form of the preceding

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

one recognized that the coefficients $A_{i,h}$, applied directly to the n given equations, are linked to the factors B_1, B_2, etc., by the relation

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

Let now one be reported to the relations (3) of the $K_{i,h}$ and let one subtract from them a system of similar relations in which one will have set $A_{i,h}$ in the place of $K_{i,h}$, thus as the remark has been made just now; the subtraction of two corresponding relations will give a system of m equations

$$Sa_{1,h}(K_{i,h} - A_{i,h}) = 0,$$
$$Sa_{2,h}(K_{i,h} - A_{i,h}) = 0,$$

$$Sa_{i',h}(K_{i,h} - A_{i,h}) = 0,$$
$$Sa_{i,h}(K_{i,h} - A_{i,h}) = 0,$$
$$Sa_{m,h}(K_{i,h} - A_{i,h}) = 0;$$

and it becomes manifest that by multiplying them by the factors $B_{i'}$ respectively, next adding them all, they will give

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

and, as $\Sigma B_{i'} a_{i',h}$ is nothing other than $A_{i,h}$ in the method of least squares, there will result from it

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

or else

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

a singular relation between the arbitrary factors $K_{i,h}$ and those that require the minimum of the squares. Thence nothing is more easy than to conclude the identity

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

next

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

and

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

an expression by which Mr. Gauss has demonstrated that, under the relations (3), the minimum of the sum of the squares of the factors $K_{i,h}$ holds when these factors are precisely the factors $A_{i,h}$ assigned by the method of Legendre.

This method reduces therefore the errors which will enter necessarily into the values of the x_i, to the most narrow possible limits, for a given probability. Reciprocally, if a system of limits determine a probability, as the quantity γ which serves to calculate it will be linked to all the limits by the relation

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

it is palpable that γ will be a maximum for the minimum of $S.K_{i,h}^2$, or for the result of the method of least squares; so that a system of limits being chosen, the probability that the errors will not exit from it will be the greatest possible, when one will have determined the unknowns by this method.

It has already been said effectively that, in the formulas (11), the value of p is increasing with γ. It is this which one recognizes immediately by the differentiation of these formulas which, both, when m is the number $2g$ or $2g - 1$ of the elements, have for derived the quantity

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

a very simple expression which would change in sign only with the constant γ, here regarded as variable.

One is able to be assured that, for $m = 1$, when there is only a single element or a single unknown x, all these formulas return completely into the known relations

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

limits of $\rho = \pm\gamma\sqrt{S.A_h^2}$,

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

One knew the simple processes given by Laplace and by Mr. Gauss in order to deduce the quantities μ_1 and μ_2 of the same observations. These processes, which depend on that which one called the *theory of probabilities a posteriori*, are not modified by the change which was just developed. It influences happily only on the magnitude of the probability, or rather on the extent of the most probable errors. There will therefore be no question of the calculation of μ_1 nor of μ_2. but it has seemed necessary, except to form some Tables of numerical values of the formulas (11), at least to present some interesting values.

Before proceding to this application, perhaps it is acceptable to make to observe that if the expression of the probability p given generally (page 16),

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

had been integrated relative to all the variables except one, the result would have been the probability of the limits assigned to the error ρ corresponding, however great that all the other errors be: in other terms, the probability of this error considered in isolation, and as if the others did not exist. One would arrive in this manner to the formulas relative to the case of one element only, such as they come to be written, that is in the ordinary manner of calculation of the probability. But, again one time, this mode gives it much too great, since it counts, in the calculation, all the probability of the combinations of errors, in which the other errors have some magnitudes which would no longer permit even to be trusted in the equations. One is going to recognize, moreover, the extreme difference of the two results.

The same considerations have prevented having regard to the mean values of the errors. One knows that one calls thus the sum of the products of the errors, taken all positively, by their respective probabilities. This arithmetic mean, in which each error enters proportionally to the number of chances which are able to bring it forth, are able to be an exact index of the importance of an error only when it is calculated in isolation. If, on the contrary, it was encountered that the greatest considerable magnitudes of the error of an unknown were precisely those which depend on systems of errors of which the probability is weak, one imagines that the mean would be able to give an erroneous idea of the most ordinary magnitude of the special error to which it is reported. In general, the use of the mean values is chosen delicately, unless they were not the special object of researches. One will sense immediately, for example, that even in the most simple case, the mean of the errors ρ_i, which is calculated by integrating

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

does not give an idea quite just, since the limits between which there are odds one against one that the value ρ_i is able to fall are determined by the equation

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

or else

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

where one draws

$$z = 0.476\,936.$$

$\frac{1}{\sqrt{\pi}}$ is, on the contrary, equal to 0.564 18.

The mean of the errors is therefore far from being encountered among the most probable errors. In truth, this mean is that of all the errors taken with the + sign, and the real mean is 0. There would be more than one observation to make on the usage of the means; but it is necessary here, in order to avoid too much length, not be be arrested there, no more than a great number of other useful points. It suffices to have shown that the means have not always the sense that the habits of the mind make attached, in the most ordinary circumstances; and that, consequently, they are not proper at all to the demonstration of the method of least squares. Also is it from this evaluation of the mean of the errors, taken with the positive sign, that Mr. Gauss forms a kind of objection to the analysis of Laplace (*Theoria Combinationis*, etc.). The fact is that the result would not be at all rigorous, if this mean were not found proportional to the limits of the errors. But it is likewise in the mean of the squares of the errors that Mr. Gauss believed to be able to adopt *a priori* as *criterium* of the precision. It is, on the contrary, the existence of the remarkable criterium that the analysis of Laplace demonstrates, likewise all that which precedes.

§IV.

In the applications, the value of the constant γ, the most difficult to find perhaps, if not the most useful, is that for which the probability is equal to $\frac{1}{2}$. There is, for this value, as many odds as the errors will fall within the limits which result from it, as there are they will exceed these limits.

Thus, first, in the case of a single unknown x, Bessel has given the value

$$\gamma = 0.47\,693\,64.$$

The last decimal appeared inexact, and it is necessary to take

$$\gamma = 0.47\,693\,627\,620.$$

But it is there a veritable luxury of decimals, in fact of probabilities; for it will be quite rare that the term neglected which depends on a large number n of observations permits to push the approximation beyond some first decimals.

One has therefore, for a single unknown, the probability

$$p_1 = \frac{2}{\sqrt{\pi}} \int_0^{0.4769\ldots} dt\, e^{-t^2} = \frac{1}{2},$$

that the error r of the value x is between the limits

$$r = \mu_1 S.A_h \pm 0.4769\ldots \sqrt{2(\mu_2 - \mu_1^2)S.A_h^2},$$

the factors A being those which indicate the method of least squares. As to the values μ_1 and μ_2, they must be sought, either beyond the observations, or by the observations themselves, and that, according to the known process for the probabilities *a posteriori*, thus as it has been said.

When nothing prevents raising the value of γ to 3 for example, the probability that the error is contained within the limits

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

is expressed very nearly by

$$\frac{2}{\sqrt{\pi}} \int_0^3 dt\, e^{-t^2} = 0.99\,997\,790.$$

It is this which one is able to see in the Tables which have been published for the values of the integral $\frac{2}{\sqrt{\pi}} \int_0^{\gamma} dt\, e^{-t^2}$ (notably in the *Exposition de la Théorie des Chances*, of Mr. Cournot.)

These values are precisely those which have been applied, by omission, to the problems which comprehend many unknowns. They are reported here only in order to facilitate the comparison with those which give formulas (11).

If there are two unknowns, it is necessary to employ the second of these formulas, and the probability is then very simply

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

It becomes equal to $\frac{1}{2}$, for $\frac{1}{2} = e^{-\gamma^2}$, or $\gamma = \sqrt{l2}$, the letter l indicates the Napierian logarithm. One will find without difficulty

$$\gamma = 0.83\,255\,461\ldots$$

Thus, since there are only two elements to deduce from observations, the limits comprehend an interval nearly double, and the errors are able to be much more greater, consequently.

One has the probability $\frac{1}{2}$ that the error of the element x_i is comprehended within the limits

$$r_i - \mu_1 S.A_{i,h} = \pm 0.83\,255\ldots \sqrt{2(\mu_2 - \mu_1^2)S.A_{i,h}^2},$$

or, if one wishes that the errors are able to vary, one will say that $\frac{1}{2}$ is the probability of the set of all the systems for which

$$r_1 - \mu_1 S.A_{1,h} = t_1 \sqrt{2(\mu_2 - \mu_1^2)S.A_{1,h}^2},$$
$$r_2 - \mu_1 S.A_{2,h} = t_1 \sqrt{2(\mu_2 - \mu_1^2)\frac{(S.A_{1,h}A_{2,h})^2}{S.A_{1,h}^2}}$$
$$t_2 \sqrt{2(\mu_2 - \mu_1^2)\left(S.A_{2,h}^2 - \frac{(S.A_{1,h}A_{2,h})^2}{S.A_{1,h}^2}\right)},$$

the variables t_1 and t_2 being subject to the condition

$$t_1^2 + t_2^2 < 0.69\,314\,718\ldots;$$

the quantities under the radicals being determined besides by the relations (9), where the factors A of the method of least squares replace the arbitraries K in the coefficients

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

When in this system one will wish to attain a probability $0.99\,997\,790\ldots$ equal to that which gives $\gamma = 3$ when there is only one unknown, it will be cecessary to resolve the equation

$$1 - e^{-\gamma^2} = 0.99\,997\,790\ldots;$$

which raises the value of γ to $3.27\,419\ldots$

Thus, for some very great probabilities, the limits differ less than for the weak probabilities, when one passes from the case of a single unknown to the case of two unknowns. It is this of which it is easy to render account, since it would be absolutely likewise if the two unknowns were independent of one another.

Here is now (finally for brevity) the table of values of γ which give the probability equal to $\frac{1}{2}$ in the formulas (11), from $m = 1$ to $m = 8$. It would not be difficult, but it

would be very long to make the same calculations for a greater number of elements:

$$
\begin{aligned}
m &= 1, & \gamma_1 &= 0.47\,693, \\
m &= 2, & \gamma_1 &= 0.83\,255, & &= 1.74\,56 \times \gamma_1, \\
m &= 3, & \gamma_1 &= 1.08\,76, & &= 2.28\,14 \times \gamma_1, \\
m &= 4, & \gamma_1 &= 1.29\,551, & &= 2.71\,64 \times \gamma_1, \\
m &= 5, & \gamma_1 &= 1.47\,50, & &= 3.09\,27 \times \gamma_1, \\
m &= 6, & \gamma_1 &= 1.63\,525, & &= 3.42\,87 \times \gamma_1, \\
m &= 7, & \gamma_1 &= 1.78\,12, & &= 3.73\,47 \times \gamma_1, \\
m &= 8, & \gamma_1 &= 1.91\,623, & &= 4.01\,78 \times \gamma_1.
\end{aligned}
$$

The concern is only to compare each of these numbers with the first, in order to recognize immediately the real extent of the limits of the probable error, already calculated in a system of observations. There exists, effectively, no change to make incur to the ordinary calculations, save to the one of the value γ. It is nearly useless to remember, moreover, that the probable error is thus named because it is precisely the value of the limit corresponding to the probabiltiy $\frac{1}{2}$.

Here is an example of this closeness. Bessel, in his Memoir on the Comet of Olbers (*Untersuchungen über die Bahn des Olberschen Kometen*), employed to the correction of the elements of the orbit three hundred forty-nine observations, of which he forms some equations in six unknowns. He concludes from it, for one of these unknowns, the correction of the eccentricity, a probable error of $0.00\,017\,127$ which he evaluates in time at around 101 days, out of the 74 years assigned to the revolution of the star.

In this calculation, Bessel has employed the factor

$$
\gamma_1 = 0.47\,693.
$$

Since there are six unknowns, it would be convenient to take, on the contrary,

$$
\gamma_6 = 1.63\,525 = 3.428\ldots \times \gamma_1,
$$

that is more than triple, and to elevate the probable error to $0.00\,058\,72$, this which entails without doubt a probable error near to one year out of the duration of the revolution.

One is able very well here to recognize the necessity to extend thus the limits of the probable error. One senses, indeed, that it is necessary, in order that the errors of the other elements are able to have the same probability, to take account of all the combinations in which each of their values are by right to enter.

Bessel remarks with all reason that it is much to regret that the circumstances of the geocentric movement had left one so great uncertainty on this element, while the others offer only some errors quite less comparatively. The comet of Olbers must return to perihelion only 9 February 1887, it will not be before thirty-five years that these calculations will be able to offer a positive interest. But one sees, in the Memoir of Bessel, that he assigned a probability of 66.65 against 1, to the limit of one year, while one is able to wager no more than 1 against 1 that the comet will not exceed this limit, in advance or in retard. In order to obtain 66.85 against 1, it would be necessary to resolve exactly the equation resulting from the second of the formulas (11), this which

would give to γ a value (2.812) more than quintuple of the value γ_1, and entailing a limit of possible errors, although little probable, around nineteen months.

It would be nearly useless to make exactly this last calculation, because the number of observations being only 349, and the number of the elements being elevated to 6, the divisor of the neglected terms attains only 58; so that these terms would be able to have a great influence on the first decimals of the probability. It suffices to have indicated the sense in which the omissions influence. As for the small values of γ, it is much more difficult that the neglected terms were an effect capable of altering sensibly the probability $\frac{1}{2}$, or each other as weak, and it is by this reason that the calculation has been made with exactitude.

Here is a second and last example of the application of the formulas (11).

In the first Supplement to the *Théorie analytique des Probabilités*, Laplace evaluated to $1\,000\,000$ against 1 the probability that the mass of Jupiter, which he corrected by the aid of 129 equations, in which he has made this mass equal to $\frac{1+z'}{1067.09}$, and which he reduced to $\frac{1}{1070.35}$, is not in error of $\frac{1}{100}$ of the first value. However, it is carried, actually, to $\frac{1}{1050}$ in the *Annuaire du Bureau des Longitudes*. The difference is around $\frac{1}{51.5}$; this which makes very nearly the double of the limits assigned with one so high probability.

This considerable difference is able to depend on the manner by which the 129 equations have been formed, and is able to be even one totally natural consequence of the defect of precision of the observations employed, or of the formulas of reduction. But, if one considers that the 129 equations contain 6 unknowns, and that, consequently, it would be necessary to attach to the limits of z' only the probability of the set of the limits assigned to all these unknowns, and to which they must satisfy at the same time, one finds a probability much inferior to $\frac{1\,000\,000}{1\,000\,001}$, or, more exactly, $1 - 0,00000\,1004$, which give the values reported by Laplace.

The value of the correction, such as Laplace gives it, is

$$z' = 0.00\,305 + \gamma\sqrt{\frac{1}{(345.885)^2}}.$$

The divisor under the radical is a little too great, because it would be necessary to take, for 129 observations and 6 unknowns, only 123 times (or $129 - 6$) a certain denominator. By taking account of this slight correction, depending on this that 6 is not to neglect next to 129, one has, for z', the limits

$$z' = 0.00\,305 \pm \gamma \times 0.00\,296.$$

The possible error of z' must be such, that the error of the mass is below $\frac{1}{100}$ of $\frac{1}{1067.09}$, there results from it

$$\frac{\gamma \times 0.00\,296}{1067.09} = \frac{1}{100} \cdot \frac{1}{1067.09},$$

or else

$$\gamma = 3.37\,745.$$

This value of γ would have yet given to Laplace the probability $0.999\,998\,21$, or very nearly odds of $560\,488$ against 1.

But, if one makes it enter into the second of the formulas (11), in putting $m = 6$, one finds

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

and, each calculation made,

$$p_6 = 0.9991389,$$

or very nearly $\frac{1160}{1161}$.

There would be therefore only odds of 1160 against 1, that the mass $\frac{1}{1070.35}$ were not in error by $\frac{1}{100}$.

This probability is able already to appear rather great; but it is not necessary to lose from view that the formulas (11), as those of Laplace, neglect some terms which vanish only when $\frac{m}{n}$ is rather small in order to permit employing some values of γ as great as 3.37745. These terms, without doubt, render here the formula erroneous, since $\frac{n}{m}$, far from being a great number, is only

$$\frac{129}{6} = 21.5.$$

The value

$$p = 0.9991389$$

resulting from the approximation, is therefore here very imperfect; and even then that one would be held to the formula of Laplace, the value that he has given of $\frac{1000000}{1000001}$ would not be able to be admitted because of the great uncertainty of the terms neglected by him.

It is necessary to recognize that 129 equations permit using this approximative analysis only for weak values of γ. But then, will one say with Mr. Gauss, the method of least squares is therefore no longer demonstrated for the most frequent cases in practice, for some numbers of observations very great relative to the pains that they give to the observers, but too small in order to assure a great probability; this method, which appeared precious, would be therefore most often only a convenient rule by its uniformity and by the advantages that Legendre had mentioned since the origin?

To these regrets too founded, it is necessary to respond affirmatively without the least hesitation. There is need, indeed, only of simple good sense, destitute from each calculation, in order to recognize that some observations affected with errors are able to give the values sought only when it is made from the compensations among the errors; and everyone senses that these compensations could take place only out of the large numbers. To employ any mean from a small number of facts, it is necessarily to incur the risk that the errors add. It is thus of them in all the circumstances where the proof is able to be repeated at will, and then the knowledge that one acquires must no longer be imprudently extended to beyond the limits of the greatest possible errors. Certainly, it would be a true aberration to claim that the arithmetic mean of 10 measures of which each is able to be affected by an error, by more or less, from 0 to 1 centimeter, will have a great advantage on each other combination of these 10 measures. Without doubt, it would be worth more to be able to demonstrate that it enjoys always, even for so small numbers, a superior probability; but, by the single nature of things, the

difference would be very small. Thus, is it superfluous to research that which is passed in small numbers? In this regard, one will be always and of total necessity constrained to return to the reflection of Gibbon: "The laws of probability, so true in general, so deceptive in particular." It would be necessary therefore, even when one would have proved that such or such process is the most probable, to serve only in taking account, no longer of the mean errors, but of the greatest possible errors; for the probability of the ones and the others will never differ enough in order that one is able to be permitted to neglect the ones rather than the others. Probability is founded only on the possibility of the things, and in a small number of observations, a possible event has chance to take its place as each other of the same possibility. It is only in the large numbers that certain classes of facts, certain combinations become impossible, or rather not very possible, consequently, improbable. And this urgent condition of the great number has nothing in particular to the method of least squares.

That it is necessary to continue to apply this method to small numbers of observations, it is this which is not doubtful at all. The analysis which demonstrates this remains true for small numbers likewise as that which proves the value of the mean results of each specie. Only, the neglected terms no longer allow to be evident an absolute maximum of probability; but, taking their influence, they come to alter a little the result of the approximation, and one sees that this result is not quite extended from the maximum. If it differs from it, it is only by small quantities of which the divisors increase proportionally to the number of observations; so although the method takes this property to give the maximum of advantage, since this number has some value, it is also before it is great enough in order to permit to apply, without scruple, the formulas of approximation to the calculation of the probability. It would have to add to this motives, drawn from the march of the calculations, all those that Legendre made first to be worth in favor of its process. But this is not the place at all; they are known and appreciated. There would be concern only to show the ways to fill that which had appeared to be a troublesome gap in a theory so useful. For nothing is more hurtful to the progress toward the truth than the erroneous confidence which is increased by possession of results of which science is yet remote.