CONSIDÉRATIONS A l'appui de la décourverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés*

M. Bienaymé

Comptes rendus des Séances de l'Académie des Sciences, XXXVII, pp. 309–324
Liouville's Jour. Math. Pures Appl. (2) 12, (1867) pp. 158-176

Session of 20 August 1853

In the session of 8 August 1853, the Academy has extended some remarks, some reflections made by Mr. Cauchy and by Mr. Le Verrier on the method of least squares. These reflections have not been inserted into the *Comptes rendu*, to my great regret, because they had appeared of a true interest. There has been published only the Memoir of Mr. Cauchy, of which the lecture had furnished the occasion of this kind of discussion; and I have not at all seen reproduced an opinion which had seemed to me to direct it into his verbal observations. Mr. Cauchy had denied the exactness of the result so remarkable, discovered and demonstrated by Laplace, and which consists in this that the method of least squares is applied to the data of observations, whatever be the law of probability of errors. I have believed to be able to announce, then, that I could have some good reasons perhaps to present to the support of the opinion of Laplace. I am going to expose them the briefest possible; although I am by right to make remark however that the diverse Memoirs or pieces of Memoirs published by our scholarly colleagues, in Nos. 5, 6 and 7 of the *Compte rendu*, do not justify his assertion. Far from there, according to me, there are such parts of this analysis so fecund, and sometimes very ingenious, that, with very little change, would demonstrate plainly the discovery of Laplace. And if I limit myself to indicate it, it is that I do not wish even to have the air of making it a criticism of it.

Criticism is not at all my end. The critical works that I will have to present successively to the Academy will show that there is, in my eyes, between my hands, only an instrument of which I am obliged serve myself in order to arrive to the truth. The calculus of probabilities gives place to some singular illusions, to which the better minds

*Translated by Richard J. Pulskamp, Department of Mathematics & Computer Science, Xavier University, Cincinnati, OH. June 19, 2010 have not always been able to remove themselves. This calculation is the first not of Mathematics beyond the domain of absolute truth. They have made it only by groping, so to say. Laplace entitled one of his chapters: *Des illusions dans l'estimation des probabilités*. Also one finds more of an error buttressed by a recommendable name, more of a mistake escaped to some men who have authority. The science, although Pascal has created it two centuries ago, is not very far from its origin. The analytic development has progressed nearly alone. Now, in these commencements of a science, the one who seeks the true seems sometimes to use a quite rigorous criticism, while he strives alone to exit from obscurity, to make vanish the phantoms that one believes to perceive. It is seen to compel all at the same time to reduce to narrow limits the advantages that too much haste, too much imagination had made to exaggerate beyond measure; to put into evidence the inconveniences that the ardor of the first successes had concealed; sometimes also, to maintain the points justly to acquire the knowledge against some preventions, some inexact judgments which would make it retrograde. In a word, he is forced to bulldoze the land, and to defend it in order that his successors are able to advance with security. This is that which is able alone to make a criticism impartial. And next, it is necessary to say well, the same nature of the calculus of the probabilities, which treats some errors of every kind, some deviations of every genre, of the incompatibility of the observations resulting from the feebleness of human organs, of the discordance of given statistics, discordance which physical variations produce, very extended most often, and the negligence or the imperiousness with which these variations establish themselves; the nature of this calculation which treats of all these facts for which has been forged the word of hazard, so little intelligible; the same nature of this calculation is crucial.

I will have therefore need, for this character of many of my future communications, of a sympathetic disposition of the listener or of the reader. I solicit it in advance. It is, moreover, quite a long time that the greater part of my researches have been made. The demonstrations of certain errors go up again to more than thirty years. I dare to hope that one will wish well to understand, in the long silence where they have remained, the proof of a complete absence to desire to criticize. Besides, I will go as I have already done; I will reduce the criticism to the strict necessary, and I believe that Mr. Cauchy will see it in my actual communication.

And however the opinion of Laplace on the subject which occupies us would merit although all the possible resources were employed to sustain it. The question is not only of the method of least squares, of a combination of equations to resolve, that one would be able to replace, without great damage, by another combination. Perhaps, if it were a concern only of the errors of observations, I would have not taken the podium three weeks ago; although I ought have supposed that I was a small cause of the lessening that our scholarly colleague would wish to make subject to the method of least squares. But here is why I ought to speak. It is, that one knows it well, that if the discovery of Laplace is inexact, a considerable part of his great work, the better part, in that which it is the most applicable, the most practical, will be found reversed in a single coup. All the chapters which treat *of the research of the phenomena and of their causes; of probability of causes and of future events drawn from observed events; of the mean durations of life, of marriages and of any associations, of the benefits of the established bases on the probabilities of events, etc.*, will follow more or less completely the lot of the celebrated chapter IV *sur la probability des erreurs des résultats moyens*, where is found demonstrated the method that Legendre had founded on some considerations so different. Because all these chapters are supported on a common principle, on the reduction of any function to some terms common to all, and easy to calculate.

Also Laplace had sensed immediately the importance of his discovery. Hardly had he done it, that he brings it before this company, and that he announces that he is going to publish a Treatise on the probabilities. From 1770 to 1809, during nearly forty years, Laplace had given some numerous Memoirs on the probabilities: but, some interest that he had in these Memoirs, he had not wished to write them in general theory. As soon as he has recognized the property of the functions of probability, he sees clearly that it is a principle which rules nearly all the applications, and he composes his theory.

One has said often that this theory would render, alone, the name of Laplace immortal. But I am founded to believe that these words carried only on the beautiful developments of analysis which are due to him; and that the principle of probabilities, so general, so fecund, has struck the eyes only of a very small number of persons. If it was not thus, I am persuaded that instead of starting it, of revoking it in doubt, one would endeavor to consolidate it, and to show how Laplace has extended it; how one is able to abuse it by some ill conceived or incomplete applications, how finally the progress of analysis is able to serve to spread the light on the calculus of probabilities. But many have seen only one occasion of problems of analysis, and those who have sought the spirit of this analysis are exhausted by vain efforts in order to replace the calculus of Laplace by that which one calls *easy demonstrations*, *popular proofs*.

Until here, it does not seem possible to operate this replacement as soon as one wishes to know the magnitude of the probability of a given error. The analysis follows through Poisson, the analysis that Cauchy has just employed with a little more of that rigor which must rule Mathematics, giving, in fact of numerical calculations, of true practical applications, nothing more than the formulas so commodious that Laplace has known to draw from his.

But if one wishes to be limited to demonstrate the method of least squares by that which touches the combination of observations, without calculating the greatness of the error or only by proving that that method to operate in a manner to render it a minimum, transcendent analysis is no longer necessary, nor even great intensity of mind, one time that one has well envisioned the conditions which make a general principle of the discovery of Laplace.

Two things, indeed, have always surprised, and always will surprise, when, for the first time, one comes to consider the final result which he has delivered to the observers. It is, on one hand, that the integral $\frac{2}{\pi}\int_0^a dt\, e^{-t^2}$ reproduces itself without ceasing, as approximate expression of the probability; and on the other, that the deviations or the errors are proportionals to the limit of that integral, and to a function of the mean of the squares of the differences between each possible error and its mean. It is there all that which remains, in the approximation of Laplace, of the primitive character of the function of probability which ruled during the observations.

It seems, from the very first, that it had a mysterious link between the probabilities and that integral which is presented to Mr. Gauss in his first researches, as a consequence of the principle of the arithmetic mean that he adopted gratuitously, and which is again evident from the analysis of Laplace, based uniquely on this that the observations are in great number. Also, more as scholar has he thought that a better deeper knowledge of the theory of probabilities would bring understanding that the law of probability of the errors is represented by the function $\frac{1}{\sqrt{\pi}}e^{-t^2}$. But it is there one of these inexact views which leads to some erroneous consequences: and that has producted in great number, as much to the theoretical point of view as in the practical results. It will suffice to say here that there is not a necessary liaison between that exponential and the laws of probabilities; that it is only a means of very convenient approximation, but which would be able to be replaced by some other formulas; that this which restores it without ceasing, it is that it is very proper to represent a function in the environs of the maximum, but that it represents it only nearly so; and that even in the questions where it offers the most facility as approximation, it gives frequently some results of which the falsity is manifest as soon as one wishes to employ some reasonings a little complex, instead of holding it for that which it is really, that is to say for a pure arithmetic machine, good to the numerical calculations.

But it is not thus of the mean of the squares of the differences of the errors to their mean. It is not an arbitrary element of the approximation, nor, as Mr. Gauss believed it, an arbitrary measure of precision, to which one would be able to substitute each other mean of powers of even degree. All to the contrary, the mean of the squares contains the fundamental condition of the development of the probabilities in measure as the observations are multiplied; and if it is not permitted to say *a priori* that one could find some more advantageous function (other than the means of even degree), one is able to affirm that under the hypothesis of one such discovery the function would be able to be replaced by a use convenient to the mean of the squares.

The reason for it is quite simple. In order to explain it, I will take, as M. Cauchy has done it, the most ordinary case of the method of least squares.

When one wishes to resolve a system of equations of the first degree where some observed quantities ω_1, ω_2, ω_3, ... enter, which are affected of certain errors equal respectively to ϵ_1, ϵ_2, ϵ_3. ..., one multiplies each of these equations by one of the arbitrary factors h_1, h_2, h_3, ... subjected to make vanish all the unknowns, save a single one, that one obtains under the form

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

so that the error of x has for value

$$\xi = h_1\epsilon_1 + h_2\epsilon_2 + h_3\epsilon_3 + \cdots .$$

In general, the error ξ is susceptible to all possible values, because the signs of the coefficients and those of the errors are already determined, and that one is not able to dispose completely of the factors h, which must satisfy $i - 1$ equations, if there are i unknowns, in the end to eliminate $i - 1$ of them.

The question is therefore, in order to obtain the probability of ξ, to examine what is the law of probability of a sum of n products, if there were n equations, each product formed from an error multiplied by a given factor.

Let b_i be the probability of an error ϵ_i of a manner that the sum

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

as this must be; and we consider the polynomial

$$b_1 z^{\alpha_1} + b_2 z^{\alpha 2_1} + b_3 z^{\alpha_3} + \cdots = S.bz^{\alpha} = \phi(z),$$

in which an exponent α_i is a function determined from the error ϵ_i.

If one makes, likewise,

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

β, γ, etc., being some functions of the errors, it is evident that the product

$$S.bz^{\alpha} \times S.bz^{\beta} \times S.bz^{\gamma} \times \cdots = \phi(z).\psi(z).\chi(z).\cdots = P$$

will have in each of these terms, as exponent, one of the values of all the sums which one is able to form by adding n of the quantities α, β, γ, etc., taken at will, and for coefficient the probability of that value. Thus this product P, ordered with respect to the magnitude of the exponents of z, whatever they be, will present the law of probability of the sums of which one alone has been able to be encountered in the experiences or observations; likewise as $\phi(z) = S.bz^{\alpha}$ presents the law of probability of the quantities α. Designating by σ_i one of the sums

$$\alpha + \beta + \gamma + \cdots$$

and by B_i the corresponding probability, one will be able to write $P = S.bz^{\sigma}$, and that which was just said of $\phi(z)$ will be applied to the product P.

One knows that the logarithmic derivatives of $\phi(z)$ will be

$$\frac{\phi'(x)}{\phi(z)}, \quad \frac{\phi''(z)}{\phi(z)} - \left[\frac{\phi'(x)}{\phi(z)}\right]^2, \quad \frac{\phi'''(z)}{\phi(z)} - 3\frac{\phi''(z)\phi'(z)}{[\phi(z)]^2} + 2\left[\frac{\phi'(x)}{\phi(z)}\right]^3,$$
$$\frac{\phi'^{v}(z)}{\phi(z)} - 4\frac{\phi'''(z)\phi'(z)}{(\phi(z))^2} - 3\frac{[\phi''(z)]^2}{[\phi(z)]^2} + 12\frac{\phi''(z)[\phi'(z)]^3}{[\phi(z)]^3} - 6\frac{[\phi'(z)]^4}{[\phi(z)]^4}, \cdots$$

But one knows also that the logarithmic derivatives of a product

$$P = \phi(z).\psi(z).\chi(z).\cdots$$

are respectively the sums of the derivatives of like order of the logarithms of the factors

$$\frac{d^n \ln P}{dz^n} = \frac{d^n[\ln\phi(z)]}{dz^n} + \frac{d^n[\ln\psi(z)]}{dz^n} + \frac{d^n[\ln\chi(z)]}{dz^n} + \cdots$$

Whatever be that order, that derivative, if the product is reduced by one power, is therefore simply equal to n times the derivative of the unique factor of the power. That singular property is one of the foundations of the calculus of probabilities, as one knows by examining that which these derivatives become, while one makes the variable z vanish, which has been introduced only in order to carry the exponents.

One is able to remark, first, that the first derivative of $\phi(z)$ is

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

and that, consequently, for $z = 1$, that derivative is reduced to $S.b\alpha$, that is to say to the mean of the quantities α. One concludes from it that in the product P, the first logarithmic derivative is likewise the mean $S.B\sigma$; and as this first derivative is necessarily the sum of the first derivatives of each of the factors; it is evident from it immediately that the mean value of the sums σ of the quantities α, β, γ, etc., is the sum of the mean values of these quantities, that is to say that one has

$$S.B\sigma = S.B\alpha + S.B\beta + S.B\gamma + \cdots$$

If respectively $\alpha_i = \beta_i = \gamma_i = \cdots$, there comes

$$S.B\sigma = n.S.B\alpha.$$

If $\alpha_i = h_i\epsilon_i, \beta = h_i\epsilon_i, \gamma = h_i\epsilon_i$, etc., one obtains then

$$S.B\sigma = (h_1 + h_2 + h_3 + \cdots)\, S.B\epsilon.$$

These formulas were known. They express that which one is able to call the conservation of the arithmetic mean, in the succession of events composed of simple events subjected to a like law of probability. The mean of the sums of the events is the sum, or a multiple, of the mean of simple events. With some other conditions, there would be some other modes of combination of the means. The one there suffices for that of which there is question at this moment.

But the arithmetic mean is not at all the only one which is conserved thus.

One has clearly

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

hence, the second logarithmic derivative

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

and making $z = 1$,

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

Similarly,, the second derivative of the product

$$P = S.Bz^\sigma$$

will be, for $z = 1$,

$$S.B\sigma^2 - (S.B\sigma)^2 - S.B\sigma.$$

Whence one concludes immediately, by having regard to the value of $S.B\sigma$,

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

Putting

$$\mu_\alpha = Sb\alpha, \quad \mu_\beta = Sb\beta, \quad \mu_\gamma = Sb\gamma, \ldots, \quad \mu_\sigma = Sb\sigma,$$

one is able to remark that the mean of the squares, diminished from the square of the mean, is equal to the mean of the square of the differences of all the quantities to their mean; so that one finds

$$S.B(\sigma - \mu_\sigma)^2 = S.b(\alpha - \mu_\alpha)^2 + S.b(\beta - \mu_\beta)^2 + \cdots$$

For the case of equality of α, β, γ, etc., one comes therefore

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

and for the case of

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

as then

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

if $\mu = S.b.\epsilon$, one obtains

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

Consequently, one sees that the mean of the squares of the differences between the arithmetic mean and the diverse quantities of which it is composed is conserved in the same manner as the mean. In one sequence of events, it is a multiple of the mean of the squares of the deviations of the simple events.

According to these results, as Laplace has taught to be rid of the arithmetic mean, one is able to admit that the quantities α, β, γ, etc., are already diminished from their means in the polynomials $\phi(z)$, $\psi(z)$, $\chi(z)$, etc., and then $\phi'(z)$, $\psi'(z)$, $\chi'(z)$, etc., is reduced to zero for $z = 1$. The logarithmic derivatives will be reduced therefore to

$$\phi''(1), \quad \phi'''(1), \quad \phi'^v(1) - 3\left[\phi^v(1)\right]^2, \ldots$$

Nothing is easier than to conclude that

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

or that the mean of the cubes of the deviations of the composite events is again the sum of the means of the cubes of the deviations of the simple events. But since the 4th power it is no longer thus, and one finds, by some calculations analogous to the preceding,

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

There results from it that the mean of the 4th powers are not conserved. Indeed, in order to have this mean $S.B(\sigma - \mu_\sigma)^4$, it will be necessary to add to the two members of this relation $3\left[S.B(\sigma - \mu_\sigma)^2\right]^2$, which will be evidently of the order of n^2 for $\alpha = \beta = \gamma = \cdots$, and of the order of $(h_1^2 + h_2^2 + h_3^2 + \cdots)^2$, for the case of

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

Hence, this term will be of an order quite superior to the one of the terms of the second member of the relation, when n, or the number of observations, will be very great: these terms reach effectively only the order n, and offer different signs.

One will be able therefore to put

$$S.B(\sigma - \mu_\sigma)^4 = n^2C + nC_1;$$

and in the evaluation of this expression, it will be especially to the term of n^2 that it will be necessary to have regard, and for n very great, to that term nearly uniquely.

One recognizes thus that it is only

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

which conserves the order of the great number n.

Moreover, as

$$nC^2 = 3\left[S.b(\alpha - \mu_\alpha)^2 + S.b(\beta - \mu_\beta)^2 + \cdots\right]^2,$$

one will find

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

This which shows with evidence how the consideration of the 4th power would bring back to discuss simply the sum or the mean of the squares.

In the question which occupies us, the sum of the cubes, one sees it well, one could be of no utility, no more than any sum of odd powers, because these powers are affected of different signs in the terms of opposite sides from the mean; and, consequently, the sums and the means of odd powers are really only some differences which would not be able to serve to the reasonings which are going to follow. It suffices therefore to stop at the even powers.

But it is quite easy to be assured that each logarithmic derivative of even order $2i$ will lead, for the mean of the $2i$ powers, to some results analogous to the one that the 4th powers give; because it is clear that that derivative will contain $[\phi''(1)]^i$, and that, hence, $S.B(\sigma - \mu_\sigma)^{2i}$ will contain the power

$$\left[S.b(\alpha - \mu_\alpha)^2 + S.b(\beta - \mu_\beta)^2 + \cdots\right]^i,$$

which will be of order n^i, an order superior to the one of all the other terms. Thus this will be only $\sqrt[i]{S.B(\sigma - \mu_\sigma)^{2i}}$ which will conserve the order of n; and that uniquely by the great term containing the sum of the means of the squares. There will not be a mean to avoid the discussion of this sum.

These remarks put beyond doubt the error committed by Mr. Gauss, who, as I have said, had affirmed that the choice of the sum of the squares was arbitrary, and that one would be able at will to take for measure of the deviations of the observations a sum of any powers (see *Theoria combinationis observationum minimis erroribus obnoxiae*). This is said in passing. Here the same remarks establish that the sole mean of the squares themselves conserving in the necessary order, and representing themselves in all the sums of following powers, there will not be opening to new discussion.

Now, nothing will be easier than to recognize, by the expression

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

that it is necessary to render a minimum the sum of the squares of the factors h_i.

In the resolution of the given equations, these factors are generally of order of $\frac{1}{n}$; and, hence, the sum of their squares is of the same order.

Therefore the mean $S.B(\sigma - \mu_\sigma)^2$ will be so much more near to zero (or of the middle term of the polynomial which will have zero for exponent), as $S.h^2$ will be smaller or as n will be great. Therefore, this number increasing, all the probability in the product P will be concentrated in the terms near to the middle. It is necessary although it be so for it, since the number of terms of P is infinite, or, if it is finite, that it is very great of the order of n; and that the exponents σ becomes infinite, or at least of this same order very raised. If at a distance a little considerable from the middle term, of which the mean of the squares is approached without ceasing, it subsisted of the terms uniting any probability, it is clear that this bringing together would become impossible. For n infinitely great, the mean of the squares is infinitely small; now all these terms are positive: therefore all those in which enter some deviations which have some sensible value reunites only an infinitely small probability. There remains of probability only in the deviations nearest to zero.

It is necessary however well understood that a deviation of order of $\frac{1}{\sqrt{n}}$ will be yet very near to zero, since the question is of the mean of the squares which is of the order of $\frac{1}{n}$, and since thus the square of that deviation will be well of the same order of that mean.

Well before that n is infinite, one imagines without doubt actually in a clean manner that when this number will become very great, the probable deviations will be very small, according as the mean of their squares will become smaller. They will diminish with that mean, and reciprocally it will diminish with these deviations. In a way that by rendering that function a minimum, one will contain the error in the most narrow probable limits. But one knows that that condition of the minimum leads to the process that Legendre has called the *method of least squares*. It is therefore that method that it is necessary to follow when one treats equations more numerous than the unknowns sought.

If these considerations are not sufficient, one is able to calculate at least the form of the probability, and the march of the grandeur that it takes with the increasing number of observations.

We suppose, for example, that the error must be contained between the limits

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

After n observations, the mean of the squares being expressed by

$$S.h^2 \times S.b(\epsilon-\mu)^2 \quad \text{or even by} \quad \frac{k}{n}S.b(\epsilon-\mu)^2,$$

the number k being very small relatively to n, the limits in order to remain constants will take the form

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

But since the mean of the squares is $\frac{k}{n}S.b(\epsilon-\mu)^2$, it is clear that the terms of that mean placed beyond the limits above is able to furnish only a fraction f. The sum of these terms is therefore

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

On another side, p being the probability of all these terms beyond some limits, one will have evidently

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

θ being inferior to 1, since $t^2.2S.b(\epsilon-\mu)^2$ is the smallest of the deviations which enter into the terms in question.

Equaling these two values of S, one obtains

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

and the probability that the error falls between the assigned limits will be

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

Thus the probability increases constantly with the number of observations. For a like probability, the limits are tightened therefore, thus as it has been said.

The spirit of the analysis of Laplace is since then easily knowable; and if one wishes to calculate with precision the grandeur of the probability obtained, one is able to resume that analysis. One will be able, if one prefers it, to adapt also to this calculation that which Mr. Cauchy has given in the *Compte rendu* of the session of 16 August, p. 269[1] and later.

But these preceding considerations leave no doubt on the properties of the mean of the squares. As the reasoning has been general, it follows that the function of probability can be any.

[1]Translator's note: This refers to "Sur la probabilité des erreurs qui affectent des résultats moyens d'observations de même nature," item 527. On page 269 begins § II — On the probability of errors which affect the mean results.

I am in a hurry to arrive to the exceptions signaled by Mr. Cauchy, because they modify not at all the opinion of Laplace.

It seems, first of all, that he is not able to say *any function*, since it is necessary for the preceding discussion that the mean of the squares of the deviations, $S.b(\epsilon-\mu)^2$, be a finite quantity. But who does not see that Laplace has excluded each function of the errors which would not offer a finite value of the mean of the squares of their deviations? There was no need to say it, since that mean enters into all his calculations, and serves as measure to that which he has named the *weight* of the results. Without any doubt, he would habe been able to add expressly that exclusion of each function of probability incapable to give a finite mean of the squares of the deviations. It is probable that there is no point of it made, because he regarded as evident that in some observations, I will not say even accurate one, but only conducted with some knowledge of instruments, the errors are finite, and, by that same, all the functions of probabilities offer a finite mean for the squares, and much more for the superior powers of the deviations. These last means are useless, completely; they enter only in a remainder that Laplace neglects; and, in the passage cited by Mr. Cauchy, one will find, if one is satisfied with the motives of Laplace, another face of the same analysis which will be able to satisfy more those who will love an analytic evaluation. Unfortunately, it will not be more practical to deduce an exact numerical evaluation of the formulas of Mr. Cauchy.

Moreover, let one suppose one instant the errors without limits, and, hence, that the mean of the squares has no finite value. The same observations will warn the least attentive observer of them. Because it will be necessary that the great values of the errors had a notable probability; and since when they will present themselves, if not as often as the others, at least in great proportion. Thus, one will have some observations awfully discordant; and no doubt that they were rejected, and that the instruments or the process of observations were subjected to a very grounded correction.

Let one take for example the function of probability

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

indicated[2] by Mr. Cauchy, in the session of 8 August (p. 206 of the *Comptes rendus)*. One is going to understand that the instrument affected of a parallel law of probability would not be even put in sale by an ordinary artist. One could not know what name to give to the establishment that would have constructed it.

In order to simplify, I suppose the constant $k = 1$, and one has, for the probability of a numerical error inferior to c,

$$2\int_0^c d\epsilon\, f(\epsilon) = \frac{2}{\pi}\int_0^c \frac{de}{1+\pi^2} = \frac{2}{\pi}\arctan c.$$

[2]Translator's note: This refers to "Sur les résultats moyens d'observations de même nature, et sur les résultats les plus probables," item 526. See equation 25.

One draws thence the small Table that is here:

Probabilities.	Limit of error.	Probabilities.	Limit of error.
0.1	0.15838	0.6	1.37638
0.2	0.32492	0.7	1.96261
0.3	0.50953	0.8	3.07768
0.4	0.72654	0.9	6.31375
0.5	1	1	∞

One sees, some measure that one may wish to take for unity of error, that it will be all also easy to find an error six times greater, or six time smaller, than to encounter a neighbor of that unity. The great errors would be indicated therefore very quickly. A simple calculation shows that there are odds 2 against 1 to find one out of ten observations, and more than 7 against 1 if one executes only twenty of them.

The most ordinary attention will be soon to take care against a parallel instrument.

But there is more: it would be a quite singular instrument. There would be all so much security in a single observation as in the mean of 10, of 20, of 1000 or of any number of observations, or rather there would be the same danger in the mean of 1000 observations than with one alone. It is very easy, indeed, to be assured that the law of probability B of the mean μ of n observations is exactly the same,

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

whatever be n, small or great, as the law of a single error. One would possess then an instrument with which there would be only a single observation to execute.

One would avow that a similar example could not be raised against the discovery of Laplace. Certainly, the word *any function* is able to be understood only of functions capable to give some results either so much less exact, or so much more exact as the observer is giving the pain to multiply his difficult operations.

Thus Mr. Poisson, who has first remarked the function $\arctan \epsilon$, has not had less confidence in the method of least squares. He is content to say that without doubt one does not encounter that hypothesis in practice.

Beyond the exclusion of the functions of probability which have not finite mean of the squares of the errors, and which will be discovered by the observations, it is necessary again, in order that

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

becomes more and more small with $\frac{1}{n}$, or in measure as one make more observations; it is necessary visibly again that the factors $h_1,\ h_2,\ h_3,\ \ldots$, not form a series decreasing to infinity. It is again to Mr. Poisson who belongs the remark of this abstract case. I have published only after him an extract of my work on the effect of compound interest (*Procès-verbaux de la Société Philomathique*, 1839, and journal *l'Institut*, no. 286), where I have shown that the factors which multiply the deviations increase in geometric progression, in the practice of the financial establishments and of commerce, in a manner to require the multiplication of the affairs in a very short time. But one is able to put in question if a parallel sequence of factors will never be encountered among those which present themselves for resolving a system of equations, all capable of giving some values of the same unknowns. There will exist, indeed, errors only on the terms entirely known; and it would be necessary to admit that there are some equations in which these errors are insignificant, in order that there took place to multiply them by a sequence of factors which converge rapidly. However it is quite clear that, when even this case would happen to speak it, the discussion to which each observer must submit their data would warn immediately: likewise by the examination of the financial data, I have been warned from that action consumed of compound interest and of the expenses which influence in the manner of compound interest.

In this case therefore, it would not be necessary to say that the method has a defect: because it will not make a defect, then even, that to those who would apply it without knowing it well, and who would wish that it dispenses them in each examination. It is necessary to agree that it is there an impossible thing, and that the discussion of the observations must precede by quite far the application of the method of Legendre and Gauss, so much recommended by Laplace, and by Bessel, that consummate observer, to which person will be able to object the defect of practice, a defect which is made to sense more than one time, even according to the writing of Laplace and the writing of Poisson.

I believe to have furnished to the support of the discovery of Laplace of the reasons which are scarcely able to permit passage to the objections. It would have to give on the use of the method a great number of details in which it will be impossible to enter. I refer therefore in this regard to the original works of Gauss and of Laplace.