Sur la plus grande erreur à craindre dans un résultat moyen, et sur le système de facteurs qui rend cette plus grande erreur un minimum

Mr. Augustin Cauchy*

Comptes Rendus Hebd. Séances Acad. Sci. 37, 328-334.
OC I, 12 (529), 114-124.
Read 22 August 1853. Published 29 August 1853.

§ I. — *On the greatest error to fear in a mean result.*

Let, as in the preceding Memoir:
$k_1, k_2, \ldots, k_n$ be the quantities furnished by observation;
$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ the errors that they involve;
l any one of the whole numbers $1, 2, \ldots, n$;

We suppose besides that, the positive or negative errors being equally probable, one names
$-\kappa, \kappa$ the limits between which the error ε_l is certainly contained.

Finally, we suppose that, m unknowns $x, y, \ldots, w$ being linked to the quantities k_1, k_2, $\ldots$, k_n by n linear and approximate equations of the form

$$a_l x + b_l y + \cdots + g_l v + h_l w = k_l,$$

one deduces from these equations multiplied by certain factors λ_1, λ_2, $\ldots$, λ_n, next added to one another, the final equation which furnishes immediately the value of the unknown n. This final equation will be

$$x = \lambda_1 k_1 + \lambda_2 k_2 + \cdots + \lambda_n k_n, \tag{1}$$

the factors λ_1, λ_2, $\ldots$, λ_n being chosen in a manner to confirm the conditions

$$Sa_l\lambda_l = 1, \quad Sb_l\lambda_l = 0, \ldots, \quad Sh_l\lambda_l = 0, \tag{2}$$

and the error ξ, which will affect the value of x, will be

$$\xi = \lambda_1\varepsilon_1 + \lambda_2\varepsilon_2 + \cdots + \lambda_n\varepsilon_n. \tag{3}$$

*Translated by Richard J. Pulskamp, Department of Mathematics & Computer Science, Xavier University, Cincinnati, OH. July 7, 2010

[327] We imagine at present that one names δ the greatest error to fear, for a given system of the factors, on the value of the unknown x. δ will be, by virtue of formula (3), the product of the limit κ by the sum of the numerical values of the factors λ_1, λ_2, …, λ_n; and one will have, consequently

$$\delta = \kappa\Lambda, \tag{4}$$

the value of Λ being

$$\Lambda = \sqrt{\lambda_1^2} + \sqrt{\lambda_2^2} + \cdots + \sqrt{\lambda_n^2} = S\sqrt{\lambda_l^2}.$$

It is good to observe that, the n factors λ_1, λ_2, …, λ_n being linked to one another by formulas (2), one will be able generally to express m of among them, for example the factors

$$\lambda_1, \quad \lambda_2, \ \ldots, \quad \lambda_m,$$

as function of $n-m$ remaining factors

$$\lambda_{m+1}, \quad \lambda_{m+2}, \ \ldots, \quad \lambda_n.$$

One must only except the particular case where one could have

$$S(\pm a_1 b_2 \ldots g_{m-1} h_m) = 0. \tag{6}$$

Moreover, by leaving aside this exceptional case, one will be able to eliminate from the sum Λ the factors λ_1, λ_2, …, λ_m. In order to arrive there, it will suffice to subtract from formula (5) the equation which one obtains by adding to one another equations (2), respectively multiplied by some arbitrary coefficients $\alpha, \beta, \ldots, \eta$, next to choose these coefficients in a manner to make vanish in the value of Λ the terms proportional to the factors λ_1, λ_2, …, λ_m. One will find thus, in first place,

$$\Lambda = \alpha + \alpha_1\lambda_1 + \alpha_2\lambda_2 + \cdots + \alpha_n\lambda_n, \tag{7}$$

the value of α_l being

$$\alpha_l = \frac{\lambda_l}{\sqrt{\lambda_l^2}} - a_l\alpha - b_l\beta - \cdots - h_l\eta, \tag{8}$$

next follows

$$\Lambda = \alpha + \alpha_{m+1}\lambda_{m+1} + \alpha_{m+2}\lambda_{m+2} + \cdots + \alpha_n\lambda_n, \tag{9}$$

the coefficients $\alpha, \beta, \ldots, \eta$ being determined by the system of formulas

$$\alpha_1 = 0, \quad \alpha_2 = 0, \ldots, \quad \alpha_m = 0. \tag{10}$$

[328] Besides, save the exceptional case where the condition (6) would be verified, formulas (10) will furnish always some finite and determined values of $\alpha, \beta, \ldots, \eta$. These

values always will depend on signs attributed to the factors λ_1, λ_2, ..., λ_m provided that the ratio

$$\frac{\lambda_l}{\sqrt{\lambda_l^2}}$$

will be reduced either to $+1$, or to -1, according as the factor λ_l will be positive or negative.

We remark again that, among the factors λ_1, λ_2, ..., λ_n, the number of those which will be reduced to zero will be able generally to be superior to $n-m$. Because, $n-m$ factors being supposed null, the m remaining factors will be found, for the ordinary, completely determined by formulas (2) which will furnish for these m factors values generally distinct from zero.

§ II. — *On the system of factors for which the greatest error to fear in the value of an unknown becomes the least possible.*

One is able to demand what is the system of factors for which the error δ, that is to say the greatest error to fear in the value of the unknown x, becomes the least possible.

When the given linear equations contain a single unknown x, the question is resolved immediately by aid of formulas (5), (6) from page 270.[1] By virtue of these formulas, in order that the greatest error to fear on the value of x is the least possible, it will be necessary, as one has said, that the signs of the factors λ_1, λ_2, ..., λ_n be precisely the signs of the coefficients $a_1, a_2, \ldots, a_n$; and one names δ the greatest error to fear, the error δ, by becoming the least possible, will be reduced, excepting sign, to the least of the ratios $\frac{\kappa}{a_1}, \frac{\kappa}{a_2}, \ldots, \frac{\kappa}{a_n}$. Consequently, the least value of δ will be

$$\delta = \frac{\kappa}{\sqrt{a_1^2}}, \tag{1}$$

if a_1 is the one of the coefficients $a_1, a_2, \ldots, a_n$ which offers the greatest numerical value; and, besides, in order to obtain this value of δ, one must suppose

$$\lambda_1 = \frac{1}{a_1}, \quad \lambda_2 = 0, \ldots, \quad \lambda_n = 0. \tag{2}$$

These conclusions would cease to be legitimate, if the proposed equations [329] would contain many unknowns. But whatever be the number m of these unknowns, one is able, by aid of the principles established in § I, to determine the least value of δ, and the system of corresponding factors. In fact, in this system, saver the exceptional cases where the coefficients $a_l, b_l, \ldots, h_l$ could satisfy certain conditions, m factors at least

$$\lambda_1, \lambda_2, \ldots, \lambda_m$$

[1] *Translator's note*: This refers to "Sur la probabilité des erreurs qui affectent des résultats moyens d'observations de même nature," item 527 of the Oeuvres Complète. Equation (5) is $\xi = \frac{\lambda_1\varepsilon_1+\lambda_2\varepsilon_2+\cdots+\lambda_n\varepsilon_n}{a_1\lambda_1+a_2\lambda_2+\cdots+a_n\lambda_n}$ and equation (6) is $\xi = \frac{n\lambda\kappa}{A}$.

will acquire some values distinct from zero, and, in order to eliminate these same factors of the sum Λ, it will suffice to subject the coefficients $\alpha, \beta, \ldots, \eta$ to the conditions (10) of § I, that is to say to the formulas

$$(3) \qquad \alpha_1 = 0, \quad \alpha_2 = 0, \ldots \quad \alpha_m = 0,$$

next to replace formula (5) by formula (9). Then also, save the exceptional cases, the quantities

$$\alpha_{m+1}, \quad \alpha_{m+2}, \ldots \quad \alpha_n$$

will be generally distinct from zero, and, hence, it will be necessary that the factors λ_{m+1}, λ_{m+2}, ..., λ_n be reduced all to zero. Because if one has not

$$(4) \qquad \lambda_{m+1} = 0, \quad = 0, \ \ldots, \quad \lambda_n = 0,$$

if, for example, λ_n differed from zero, it would suffice to attribute to λ_n an infinitely small increase, but affected by a sign contrary to the sign of α_n, in order to make the quantity Λ decrease and hence the error δ. Therefore then the error δ will not be, as one supposes it, the least possible. Besides, when formulas (4) will be verified, the values of λ_1, λ_2, ..., λ_m will be immediately furnished by the equations

$$(5) \qquad Sa_l\lambda_l = 1, \quad Sb_l\lambda_l = 0, \ldots, \quad Sh_l\lambda_l = 0,$$

and formulas (5), (9), (4) of § I will give

$$(6) \qquad \Lambda = \sqrt{\lambda_1^2} + \sqrt{\lambda_2^2} + \cdots + \sqrt{\lambda_m^2} = \alpha,$$

$$(7) \qquad \delta = \kappa\alpha.$$

Therefore the least value of δ will be generally of the form $\kappa\alpha$, α being a positive quantity, determined by the system of m equations analogous to formulas (3), that is to say by m equations of the form

$$(8) \qquad a_l\alpha + b_l\beta + \cdots + h_l\eta = \frac{\lambda_l}{\sqrt{\lambda_l^2}},$$

and generally also the factors λ_1, λ_2, ..., λ_m, proper to furnish this [330] smallest value, will vanish, except the factors corresponding to the values of l written as subscripts of the letters a, b, ..., h, λ, in the equations of which one will draw the values of α.

We add that, among the values of α determined as we have just said, one must choose the least of all. By substituting that into equation (7), one will obtain precisely the sought value of δ.

Applied to the case where the given linear equations contain a single unknown x, the method that we just exposed reproduces formulas (1) and (2).

When the given linear equations contain two unknowns x, y, one has, by virtue of formulas (4) and (5),

$$a_1\lambda_1 + b_2\lambda_2 = 1, \quad b_1\lambda_1 + b_2\lambda_2 = 0, \quad \lambda_3 = 0, \quad \lambda_4 = 0, \ldots, \lambda_n = 0,$$

and from these equations joined to formulas (6) and (7) one draws

$$\lambda_1 = \frac{\frac{1}{b_1}}{\frac{a_1}{b_1} - \frac{a_2}{b_2}}, \quad \lambda_2 = \frac{\frac{1}{b_2}}{\frac{a_2}{b_2} - \frac{a_1}{b_1}} \quad \lambda_3 = 0, \dots, \lambda_n = 0, \tag{9}$$

$$\delta = \kappa \frac{\frac{1}{\sqrt{b_1^2}} + \frac{1}{\sqrt{b_2^2}}}{\sqrt{\left(\frac{a_1}{b_1} - \frac{a_2}{b_2}\right)^2}}. \tag{10}$$

Therefore then, in order to find the least value of δ, it suffices to write, one below the other, the two sequences

$$\begin{cases} \frac{1}{b_1}, & \frac{1}{b_2}, \dots, & \frac{1}{b_n}, \\ \frac{a_1}{b_1}, & \frac{a_2}{b_2}, \dots, & \frac{a_n}{b_n}, \end{cases} \tag{11}$$

next to multiply by κ the least of the ratios that one obtains when one divides the sum of the numerical values of two terms of the first sequence by the difference between the numerical values of the corresponding terms of the second sequence. If the least of these ratios is formed with the first terms of the two sequences, the least value of δ will be furnished, with the corresponding values of the factors λ_1, λ_2, ..., λ_n, by formulas (9) and (10).

It is good to observe that one draws from formulas (9)

$$a_1\lambda_1 = \frac{1}{1-\rho}, \quad a_2\lambda_2 = \frac{1}{1-\rho^{-1}}, \tag{12}$$

[331] the value of ρ being

$$\rho = \frac{a_2 b_1}{a_1 b_2}.$$

Hence, the products $a_1\lambda_1, a_2\lambda_2$ will be both positive, if the ratio ρ is negative. But, if this ratio is positive, then unity being contained between the limits ρ and ρ^{-1}, the products $a_1\lambda_1, a_2\lambda_2$ will be, one positive, the other negative.

One should evidently, in formulas (9), (10) etc., exchange among them the letters *a* and *b*, if the question was to make in a way that the greatest error to fear, no longer on the value of *x*, but on the value of *y*, became the least possible.

We have, in that which precedes, set aside the exceptional cases where the coefficients $a_l, b_l, \dots, h_l$ verify certain conditions, for example, the condition (6) of § I. In order to resolve the problem in these exceptional cases, it will suffice ordinarily to substitute in the coefficients $a_l, b_l, \dots, h_l$ some other coefficients which differ from it infinitely little and cease to fulfill the conditions of which there is question. Moreover, it will be generally easy to see how the formulas established above must be modified, in the exceptional cases.

We consider, in order to fix the ideas, the case where, the unknowns being reduced to a single *x*, many of the coefficients $a_1, a_2, \dots, a_n$, for example of the *l* coefficients

$a_1, a_2, \ldots, a_n$, offer some numerical equal values, but superior to those of all the others. Then the least value of δ will be always determined by formula (1). But the values corresponding to λ_1, λ_2, ..., λ_n will not be necessarily those which furnish equations (2), and will be able to be again all those that one deduces from the formula

$$\alpha_1\lambda_1 + \alpha_2\lambda_2 + \cdots + \alpha_l\lambda_l = 1, \tag{13}$$

by attributing to the products $\alpha_1\lambda_1$, $\alpha_2\lambda_2$, ..., $\alpha_l\lambda_l$ some positive values, or, in other terms, by attributing respectively to the factors λ_1, λ_2, ..., λ_l the signs of the coefficients $a_1, a_2, \ldots, a_l$, consequently, all those which verify the condition

$$\sqrt{\lambda_1^2} + \sqrt{\lambda_2^2} + \cdots + \sqrt{\lambda_l^2} = \frac{1}{\sqrt{a_1^2}}. \tag{14}$$

§ III. — *Conclusions.*

Let, as in § I, ξ be the error of the unknown x, and δ the greatest value that this error is able to acquire for a given system of factors. Let, besides, $-v$, v be the inferior and superior limits between which [332] one wishes to contain the error ξ, and P the probability of coincidence of this error with a quantity contained between the limits $-v$, v. If, by attributing to the factors λ_1, λ_2, ..., λ_n of such values, that the greatest error κ becomes the least possible, one puts precisely $v = \delta$, the probability P will be changed into certitude, and will acquire thus the greatest value possible. Hence, if one attributes to v a value which is inferior to the value of δ, determined in § II, but which differs from it very little, the system of factors which will furnish the greatest value of P will differ very little from the system which corresponds to that value of δ.

Thus, for example, if, by supposing the unknowns reduced to a single x, and by designating by a_1 the one of the coefficients $a_1, a_2, \ldots, a_n$ which offer the greatest numerical value, one attributes to v a value inferior to the ratio $\frac{\kappa}{\sqrt{a_1^2}}$; but very little different from this ratio, the system of factors which will furnish the greatest value of P will differ very little from the system determined by formulas

$$\lambda_1 = \frac{1}{a_1}, \quad \lambda_2 = 0, \quad \lambda_3 = 0, \ldots \quad \lambda_n = 0. \tag{1}$$

Besides, this last system will be, in general, very different from the one that the method of least squares would furnish. Therefore, for some values of v sufficiently great, the method of least squares will be far from furnishing the value x of x, corresponding to the greatest value of P. This conclusion, which subsists, whatever be the limit κ and the number n of the given equations, extends evidently to the case where, these equations containing many unknowns, one replaces the system of factors which determine formulas (1) by the one which renders then the of δ the least possible value. In consequence, one is able to enunciate generally the following proposition:

If one names v the limit below which one wishes to lower the error ξ of the unknown x, and P the probability of the coincidence of this error with a quantity contained between the limits $-v, +v$, the system of factors corresponding to the greatest value of P will be ordinarily, for the values of v sufficiently great, very different from the one *which the method of least squares would give, whatever be besides the number n of the quantities furnished by observation, and whatever be the limit* κ *assigned to the errors which involve these same quantities.*

It would seem at first that, in the case where the number n becomes very great, one would be able to draw some conclusions different or even opposed from a formula established in § I of the preceding Memoir. It [333] seems, in fact, that, for some great values of n, the products $a_1\lambda_1$, $a_2\lambda_2$, $\dots a_n\lambda_n$ subjected to verify the condition

$$\alpha_1\lambda_1 + \alpha_2\lambda_2 + \cdots + \alpha_n\lambda_n = 1, \tag{2}$$

and, hence, the factors λ_1, λ_2, ..., λ_n must generally be reduced to some very small quantities of order $\frac{1}{n}$; and if this reduction takes place, if besides, by attributing to the number Θ a very great value of an order superior to the one of $\sqrt{n}$, but inferior to the order of n, one neglects the integral (23) from page 268[2] vis-à-vis of the integral (22), then the value of P would appear it must be obviously that which gives formula (33) of page 269,[3] that is to say that of which the maximum is furnished by the method of least squares. But formula (33), established as we have just said, reposes evidently on the hypothesis which is not able to be realized.

In first place, of that which one attributes to the number n a very great value, there does not result from it necessarily that the factors λ_1, λ_2, ..., λ_n are all very small. The contrary will arrive if one attributes to the greater part of among them some null values, as in § II of the present Memoir. There is more; among the factors λ_1, λ_2, ..., λ_n many will be able to conserve finite values in the same case where one will suppose these factors determined by the method of least squares.

In fact, we consider especially the case where, the unknowns being reduced to a single x, the coefficients $a_1, a_2, \dots, a_n$ of this unknown, in the given linear equations, form a geometric progression of which the first term is a, and the ratio r. Then one will have

$$a_l = ar^l; \tag{3}$$

and by supposing the factors λ_1, λ_2, ..., λ_n respectively proportional to the coefficients $a_1, a_2, \dots, a_n$, consistently to the rule furnished by the method of least squares, one will have again, in regard to equation (2),

$$\frac{\lambda_1}{1} = \frac{\lambda_2}{r} = \cdots = \frac{\lambda_n}{r^{n-1}} = \frac{1}{a}\frac{1-r}{1-r^n}. \tag{4}$$

Now, if, the value of a not being very great, one attributes to r a value contained between the values $0, 1$, but obviously distinct from these limits, the terms of the sequence

$$\lambda_1, \quad \lambda_2, \; \dots, \quad \lambda_n$$

[334] determined by formula (4), will not be all very small, for some great values of n. The first terms, for example, will conserve some finite values, by reducing quite closely

[2] *Ibid.* Memoir 527. Equation (23) on page 268 is $\int_\Theta^\infty \Phi(\theta)\cos\theta v\, d\theta$.

[3] *Ibid.* Equation (33) on page 269 is $P = \frac{2}{\sqrt{\pi}}\int_0^{\frac{v}{2\sqrt{s}}} e^{-\theta^2} d\theta$.

to the corresponding terms of the geometric progression

$$\frac{1}{2}, \quad \frac{1}{4}, \quad \frac{1}{8}, \ldots,$$

if, n being a very great number, one supposes $a = 1$, $n = \frac{1}{2}$.

In second place, the integral (23) of § I of the preceding Memoir is not able always to be neglected vis-a-vis the integral (22); and, moreover, in order that formula (9) of the same paragraph is able to be reduced to formula (33), it is necessary that the value of v does not overtake a certain limit. This is that which proves already the analysis exposed above, and that which shows also the formulas relative to the special case where the number n acquires a very great value, as we will explicate in a forthcoming article.

Mr. A. Cauchy presented further to the Academy a *Mémoire sur les résultats moyens d'un très-grand nombre d'observations.*