Sur les résultats moyens d'observations de mme nature, et sur les résultats les plus probables

Mr. Augustin Cauchy[*]

We suppose m unknowns linked, by n linear and approximate equations, to n quantities furnished by some observations of like nature, and of which each involves a certain *error* ε. One will be able, from these equations multiplied by certain factors $\lambda_1, \lambda_2, \ldots, \lambda_n$, next added among them, to deduce a final equation proper to determine the first unknown x, and the value of x thus found will be that which one calls a *mean result*. If one knows the *law of facility* of error ε, and the limits between which this error is certainly contained, one will be able, from formulas established in the preceding Memoir, to deduce the probability P of the coincidence of the error ξ, which will involve the mean result, with a quantity numerically inferior to a certain limit v. This probability varies with the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$ that one is able to choose in a manner to obtain the greatest possible value of P; and to this greatest value of P will correspond *the most probable value* of x, which will depend generally on the limit v, and of the function $f(\varepsilon)$ proper to represent the law of facility of the error ε. Besides, ε coming to increase, the function $f(\varepsilon)$ is able to decrease rapidly enough in order that one is able, without sensible error, to neglect the values of this function corresponding to some values of ε situated beyond some limits between which the error ε is certainly contained. It is to this special case that the present Memoir is returned; and, by supposing the condition fulfilled which just was enunciated, I establish the very simple formulas which determine the most probable value of the unknown x.

According to these formulas, the most probable value of x will become independent of the value assigned to the limit v only for a special form of the function $f(\varepsilon)$, which contains two arbitrary constants c, N. From these two constants, the second N is the only one which serves, with the coefficients of the unknowns in the given equations, to determine the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$. If one supposes it reduced to the number 2, the most probable mean results will be precisely those which the method of least squares would provide. But it will be entirely otherwise from it if the number N [199] ceases to be equal to 2. We imagine, in order to fix the ideas, that the unknowns reduce themselves to a single x, and that the coefficients of this unknown, in the given equations, are unequal; then the most probable value of the unknown x will be furnished, if one supposes $N = 1$, by a single one of the given equations, namely, by that in which the

coefficient of x will offer the greatest numerical value, and, if one supposes N very great, by the final equation which one will obtain by adding to one another the given equations, prepared in a manner that in each of them the coefficient of x is positive.

§ I — *General considerations on the probability of the mean results, and on the most probable results.*

Let, as in the preceding Memoir:
$k_1, k_2, \ldots, k_n$ the quantities furnished by observation;
$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ the errors that they involve;
l any one of the whole numbers $1, 2, \ldots, n$;
ι, κ the inferior and superior limits between which the error ε_l is certainly contained;
$f(\varepsilon)\,d\varepsilon$ the probability of the coincidence of the error ε_l with a quantity contained between the infinitely close limits $\varepsilon, \varepsilon + d\varepsilon$.
We suppose further that, m unknowns $x, y, \ldots, v, w$ being linked to the quantities $k_1, k_2, \ldots, k_n$ by n approximate equations of the form

$$(1) \qquad a_l x + b_l y + \cdots + g_l v + h_l w = k_l,$$

one draws from these equations multiplied by certain factors, next added among them, the value of the unknown x; and we name $\lambda_1, \lambda_2, \ldots, \lambda_n$ these factors, chosen in a manner that one has

$$(2) \qquad S a_l \lambda_l = 1, \quad S b_l \lambda_l = 0, \ldots, \quad S h_l \lambda_l = 0,$$

the characteristic letter S indicating a sum of terms similar to one another. The value found of the unknown x, and the error ξ that will involve this error, will be

$$(3) \qquad x = S \lambda_l k_l, \quad \xi = S \lambda_l \varepsilon_l.$$

Let now P be the probability of the coincidence of the error ξ with a quantity contained between two given limits $\omega_{\prime}, \omega_{\prime\prime}$, and we put, for [200] brevity,

$$(4) \qquad \phi(\theta) = \int_{\iota}^{\kappa} e^{-\theta \varepsilon i} f(\varepsilon) d\varepsilon,$$

$$(5) \qquad \Phi(\theta) = \phi(\lambda_1 \theta)\phi(\lambda_2 \theta)\cdots\phi(\lambda_n \theta).$$

One will have

$$(6) \qquad \phi(0) = 1, \qquad\qquad (7) \qquad\qquad \Phi(0) = 1,$$

and formula (15) on page 159[1] will give

$$(8) \qquad P = \frac{1}{\pi} \int_{0}^{\infty} \frac{\sin \omega_{\prime\prime}\theta - \sin \omega_{\prime}\theta}{\theta} \Phi(\theta)\, d\theta.$$

---

[1] This is the formula $P = \frac{1}{2\pi} \int_{w_{\prime}}^{w_{\prime\prime}} \int_{-\infty}^{\infty} \mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n e^{\theta \tau i} d\theta d\tau$ given in "Mémoire sur les coefficients limitateurs ou restricteurs."

We consider especially the case where one could have

$$\iota = -\kappa, \quad \omega_I = -v, \quad \omega_{II} = v,$$

v being, as κ, a positive quantity, and moreover,

(9) $$f(-\varepsilon) = f(\varepsilon).$$

In the special case, formula (4) will give

(10) $$\phi(\theta) = 2 \int_0^\kappa f(\varepsilon) \cos \theta \varepsilon \, d\varepsilon,$$

and formula (17) on page 159^2 will give

(11) $$\begin{cases} P = \dfrac{2}{\pi} \int_0^v \int_0^\infty \Phi(\theta) \cos \theta \tau \, d\theta \, d\tau \\[2mm] \quad = \dfrac{2}{\pi} \int_0^\infty \dfrac{\sin \theta v}{\theta} \Phi(\theta) \, d\theta \\[2mm] \quad = \dfrac{2v}{\pi} \left[\int_0^\infty \Phi(\theta) \, d\theta - \dfrac{v^2}{2.3} \int_0^\infty \theta^2 \Phi(\theta) \, d\theta + \cdots \right] \end{cases}$$

(12) $$D_v P = \dfrac{2}{\pi} \int_0^\infty \Phi(\theta) \cos \theta v \, d\theta.$$

If one is not able to assign to ε_l any inferior or superior limit, one will have $\kappa = \infty$, and, hence, formula (10) will give

(13) $$\phi(\theta) = 2 \int_0^\infty f(\varepsilon) \cos \theta \varepsilon \, d\varepsilon.$$

Moreover, formula (13) comprehends, as a particular case, formula (10) [201] that one draws from it, by reducing $f(\varepsilon)$ to a discontinuous function which vanishes constantly, for some values of ε superior to κ.

If the function $f(\varepsilon)$, without being discontinuous, becomes very small, and decreases very rapidly for some values of ε superior to κ, so that one has sensibly

$$\int_\kappa^\infty f(\varepsilon) \, d\varepsilon = 0,$$

then, the numerical value of the integral $\int_\kappa^\infty f(\varepsilon) \cos \theta \varepsilon \, d\varepsilon$, always inferior to that of the integral $\int_\kappa^\infty f(\varepsilon) \, d\varepsilon$, since one has constantly $f(\varepsilon) > 0$, will be itself very small, and, hence, in the determination of the function $\phi(\theta)$, one will have, without sensible error, to substitute formula (13) in equation (10).

We observe further that equations (1), (2), (3) are not altered, when each of the quantities

$$a_l, b_l, \ldots, g_l, k_l, \lambda_l$$

---

^{2}Here we have the formula $P = \frac{1}{2\pi} \int_{-v}^v \int_{-\infty}^\infty \mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n e^{\theta \tau i} \, d\theta \, d\tau$ from the paper cited in the previous footnote.

come to change sign. There results from it that, in the case where condition (9) is verified, one is able to be limited to deduce from formulas (5) and (11), joined to formula (10) or (13), the values of P corresponding to some positive values of the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$.

We observe finally that from formula (13) one is able to deduce, not only the function $\phi(\theta)$, when one knows the function $f(\varepsilon)$, but reciprocally the function $f(\varepsilon)$, when one knows $\phi(\theta)$. In fact, we multiply the two members of this formula by $\cos\theta\tau$, next we integrate with respect to θ between the limits $\theta = 0$, $\theta = \infty$. Then, by replacing τ by ε, we will find

$$(14) \qquad f(\varepsilon) = \frac{1}{\pi} \int_0^\infty \phi(\theta)\cos\theta\varepsilon\,d\theta.$$

Similarly, one will draw from formula (11)

$$(15) \qquad \Phi(\tau) = \int_0^\infty \cos v\tau . D_v P \, dv.$$

The probability P, determined by formula (11), depends generally on the form assigned to the function $f(\varepsilon)$, and on the values attributed, not only to the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$, but also to the positive quantity v. By supposing invariable the form of the function $f(\varepsilon)$ and the value of v, one is able [202] to demand what values it is acceptable to attribute to the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$, in order that the value of P is the greatest possible, or, in other terms, in order that the first of the equations (3) furnishes the *most probable* value of x. Now, if one designates by aid of the characteristic letter δ some corresponding variations attributed to the quantities $\lambda_1, \lambda_2, \ldots, \lambda_n$, P, and if P is a continuous function of $\lambda_1, \lambda_2, \ldots, \lambda_n$, it will suffice ordinarily, in order to obtain the *maximum* of P, to subject $\lambda_1, \lambda_2, \ldots, \lambda_n$ to the condition

$$(16) \qquad \delta P = 0,$$

whatever be, besides, those of the variations $\delta\lambda_1, \delta\lambda_2, \ldots, \delta\lambda_n$ which will remain arbitraries, when one will have regard to the formulas (2) and, consequently, to the following:

$$(17) \qquad Sa_l\delta\lambda_l = 0, \quad Sb_l\delta\lambda_l = 0, \ldots, \quad Sh_l\delta\lambda_l = 0.$$

Therefore, in order to obtain the maximum of P, it will suffice ordinarily to express $\lambda_1, \lambda_2, \ldots, \lambda_n$ as function of m new factors $\alpha, \beta, \ldots, \eta$, by aid of equations of the form

$$(18) \qquad D_{\lambda_l}P = a_l\alpha + b_l\beta + \cdots h_l\eta,$$

then to determine these new factors, by aid of formulas (2) joined to formula (18).

§ II. — *On the conditions to fulfill in order that the most probable results become independent of the limits assigned to the errors which they involve.*

Let always ξ be the error which involves the value found of the unknown x, and P the probability of the coincidence of this error with a quantity contained between

4

the limits $-v, +v$. We admit besides, as we have done, for the errors that involve the quantities furnished by observation, one law of facility represented by a function $f(\varepsilon)$ which remains invariable when the error ε changes sign, and which decreases very rapidly for some values increasing from this same error. The probability P will be given by formula (11) from the first paragraph, and if, by supposing P a continuous function of the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$, one wishes to make so that P becomes a *maximum*, one must determine them by aid of the formula

$$(1) \qquad\qquad \delta P = 0,$$

in which δP represents the variation of P corresponding to the variations $\delta\lambda_1$, $\delta\lambda_2$, $\ldots$, $\delta\lambda_n$ of the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$. Besides, the values of $\lambda_1, \lambda_2, \ldots, \lambda_n$ [203] deduced from formula (1), combined with the formulas (17) of § I, and, hence, *the most probable* value x of the unknown x, will depend, in general, on values attributed not only to the coefficients

$$a_l, b_l, \ldots, \quad h_l, k_l,$$

which contain the given equations, but further to the positive quantity v; and if one wishes that x become independent of v, it will be necessary that, v coming to vary, the relation established by formula (1) among the quantities $\delta\lambda_1, \delta\lambda_2, \ldots, \delta\lambda_n$ remain invariable; in other terms, it will be necessary that one has

$$(2) \qquad\qquad D_v \delta P = 0.$$

But, in regard to formula (15) of § I, equation (2) will carry away the following:

$$(3) \qquad\qquad \delta\Phi(\tau) = 0,$$

whatever be, besides, the value attributed to τ. Therefore, if the most probable value x of the unknown x becomes independent of v, then in formula (1) one will be able to substitute equation (3) which, whatever be τ subsisting, will agree necessarily with the following:

$$(4) \qquad\qquad \delta\Phi(1) = 0.$$

On the other hand, if one puts, for brevity,

$$(5) \qquad\qquad \varpi(\theta) = \frac{D_\theta \phi(\theta)}{\phi(\theta)} = D_\theta \ln \phi(\theta),$$

one will have identically

$$(6) \qquad\qquad \delta\Phi(\tau) = \Phi(\tau) S\varpi(\lambda_l \tau)\delta\lambda_l;$$

and, hence, formulas (3), (4) will give

$$(7) \qquad\qquad S\varpi(\lambda_l \tau)\delta\lambda_l = 0,$$

$$(8) \qquad\qquad S\varpi(\lambda_l)\delta\lambda_l = 0.$$

5

Now, in order that the equations (7) and (8) agree between them, it will be necessary that one has

$$(9) \qquad \frac{\varpi(\lambda_1 \tau)}{\varpi(\lambda_1)} = \frac{\varpi(\lambda_2 \tau)}{\varpi(\lambda_2)} = \cdots = \frac{\varpi(\lambda_n \tau)}{\varpi(\lambda_n)}.$$

It is good to observe that equation (8), joined to formulas (17) of [204] § I, will determine the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$ as function of the coefficients $a_l, b_l, \ldots, h_l$. If these coefficients come to vary, $\lambda_1, \lambda_2, \ldots, \lambda_n$ will vary in consequence, and if, while they vary, the most probable value x of the unknown x remains independent of v, then, from formula (9) which will not cease to be verified, one will conclude that a ratio of the form

$$\frac{\varpi(\lambda \tau)}{\varpi(\lambda)},$$

offers a value independent of the value attributed to λ. One will have therefore then

$$(10) \qquad \frac{\varpi(\lambda \tau)}{\varpi(\lambda)} = \frac{\varpi(\tau)}{\varpi(1)};$$

and hence, by supposing τ positive,

$$(11) \qquad \varpi(\tau) = \tau^M \varpi(1),$$

M being a constant quantity. This put, formula (5) will give, for the positive values of θ,

$$(12) \qquad \phi(\theta) = e^{-c\theta^N},$$

the values of N, c being

$$N = M+1, \quad c = -\frac{\varpi(1)}{N}.$$

Besides, the value of $\phi(\theta)$ being determined by formula (12), formula (14) of § I will give

$$(13) \qquad f(\varepsilon) = \frac{1}{\pi} \int_0^\infty e^{-c\theta^N} \cos \theta \varepsilon \, d\theta.$$

Thus the law of facility of errors which involve the observations must be one of the laws which formula (13) supposes, if one wishes that the most probable value of each unknown becomes independent of the limits assigned to the error which involves this same value.

We suppose now the value of $f(\varepsilon)$ determined by formula (13); then, by attributing, as one is able to do it, some positive values to the factors $\lambda_1, \lambda_2, \ldots, \lambda_n$, one will draw from equation (12), joined to formulas (5) and (11) from § I,

$$(14) \qquad \Phi(\theta) = e^{-c\theta^N}$$

[205] and

$$(15) \quad \begin{cases} P = \dfrac{2}{\pi} \displaystyle\int_0^\infty e^{-s\theta^N} \dfrac{\sin \theta v}{\theta}\, d\theta \\[2mm] \quad = \dfrac{2v}{\pi} \dfrac{1}{N s^{\frac{1}{N}}} \left[\Gamma\!\left(\dfrac{1}{N}\right) - \dfrac{v^2}{2.3} \dfrac{1}{s^{\frac{2}{N}}} \Gamma\!\left(\dfrac{3}{N}\right) + \cdots \right], \end{cases}$$

the value of s being determined by the formulas

$$(16) \qquad\qquad s = c\Lambda, \quad \Lambda = \lambda_1^N + \lambda_2^N + \cdots + \lambda_n^N.$$

Then also, P will increase for the decreasing values of the two quantities s, Λ, and equation (8) will give

$$(17) \qquad\qquad S\lambda_l^{N-1}\delta\lambda_l = 0.$$

Therefore to formula (18) of § I, one will be able to substitute that here,

$$(18) \qquad\qquad \lambda_l^{N-1} = a_l\alpha + b_l\beta + \cdots + h_l\eta.$$

If one supposes the unknowns reduced to a single x, then, after having prepared the given equations in a manner that the coefficient a_l is always positive, one will draw from equation (18)

$$(19) \qquad \lambda_l^{N-1} = a_l\alpha, \qquad (20) \quad \lambda_l = a_l^{\frac{1}{N-1}} \alpha^{\frac{1}{N-1}},$$

and from equation (20) joined to the formula $Sa_l\lambda_l = 1$,

$$(21) \qquad\qquad \alpha^{\frac{1}{N-1}} = Sa_l^{\frac{N}{N-1}}.$$

Then also, in order to obtain the final equation which will furnish the most probable value of x, it will suffice to add to one another the given equations, after having respectively multiplied them by the diverse terms of the sequence

$$(22) \qquad\qquad a_1^{\frac{1}{N-1}}, \quad a_2^{\frac{1}{N-1}}, \ldots, \quad a_n^{\frac{1}{N-1}}.$$

If one reduces the exponent N to the number 2, equation (13) will give

$$(23) \qquad\qquad f(\varepsilon) = \left(\dfrac{k}{\pi}\right)^{\frac{1}{2}} e^{-h\varepsilon^2},$$

the value of k being

$$k = \dfrac{1}{2\sqrt{c}};$$

[206] then also formula (18), reduced to

$$(24) \qquad\qquad \lambda_1 = a_1\alpha + b_1\beta + \cdots + h_1\eta,$$

will lead precisely to the results that the method of least squares furnishes, that which agrees with the conclusions to which we are arrived in the preceding Memoir.

If one reduces the exponent N to unity, one will draw from equation (13)

$$(25) \qquad f(\varepsilon) = \frac{k}{\pi} \frac{1}{1+k^2 \varepsilon^2},$$

the value of k being $\frac{1}{c}$. Then also, by supposing the coefficients $a_1, a_2, \ldots, a_n$ unequal, and designating by a_1 the greatest of all, one will draw from formulas (20), (21) from very small values of the ratios $\frac{\lambda_2}{\lambda_1}, \frac{\lambda_3}{\lambda_1}, \ldots, \frac{\lambda_n}{\lambda_1}$. Therefore then, if the unknowns are reduced to a single x, the most probable value of x will be that which the single equation will furnish

$$a_1 x = k_1.$$

Finally, if the exponent N becomes very great, the diverse terms of the sequence (22) will be reduced sensibly to unity. Therefore then, if the unknowns are reduced to a single x, the most probable value of x will be drawn from the formula that one obtains, when one adds among them the given equations prepared in a manner that the coefficients of the unknown x in the first members are all positive.

"Mr. Augustin Cauchy presented also a Memoir which has for title: *Sur les résultats moyens d'un très-grand nombre d'observations*.

"**Mr. Bienaymé** remarked that the examination of a question as delicate as that which is treated by Mr. Cauchy in the Memoir of which he just gave a lecture could not be done with benefit in a verbal discussion. He believes, completely, that it would be possible to bring some arguments to the support of the opinion of Laplace, who thought that one had by right to apply the method of least squares, without knowing the law of probability provided that it was constant. Mr. Bienaymé will study therefore with care the new ideas issued on this subject, and he will communicate later that which could appear to him to justify the opinion of Laplace."