

THE GAUSSIAN LAW OF ERROR FOR ANY NUMBER OF VARIABLES*

BY

J. L. COOLIDGE

The exponential law for the distribution of accidental errors of observation, discovered by Gauss, has been a mathematical classic for over a century. Many have been the attempts to prove it, all based, necessarily, on more or less arbitrary assumptions. Perhaps the most searching examination of it was given by Poincaré in his *Calcul des probabilités*; his final opinion seems to be contained in the following phrase: †

“J’ai plaidé de mon mieux jusqu’ici en faveur de la loi de Gauss dont nous allons maintenant tirer les conséquences. Peut-être pourtant la cause n’était-elle pas parfaitement bonne.

“Elle ne s’obtient pas par des déductions rigoureuses, plus d’une démonstration qu’on a voulu en donner est grossière, entre autres celle qui s’appuie sur l’affirmation que la probabilité des écarts est proportionnelle aux écarts. Tout le monde y croit cependant, me disait un jour M. Lippmann, car les expérimentateurs s’imaginent que c’est un théorème de mathématiques, et les mathématiciens, que c’est un fait expérimental.”

The law has been extended to include the distribution of errors depending upon two variables, and in this form it has a certain importance in the theory of ballistics, and in that of statistical correlation; even the case of three variables has been slightly treated. The general case of n variables has never been taken up except in two recent articles by von Mises.‡ The treatment here is based on a very general form of analysis showing how an arbitrary distribution function will lead asymptotically to an exponential form. The analysis is very careful, the point of view extremely abstract, with little relation to practical applications. Moreover, the author gives no indication how the constants should be calculated in any particular case. It is the object of the present paper to deduce the Gaussian law for n variables by a method based upon the classical one for a single variable, but with somewhat broader and more explicit assumptions. In the second part we shall make

* Presented to the Society, December 27, 1922.

† Poincaré, *Calcul des Probabilités*, Paris, 1896, pp. 196 and 149.

‡ *Fundamentalsätze der Wahrscheinlichkeitsrechnung*, *Mathematische Zeitschrift*, vol. 4 (1919), and *Grundlagen der Wahrscheinlichkeitsrechnung*, *ibid.*, vol. 5 (1920). See also Dodd, *Functions of measurements*, *Skandinavisk Aktuarietidskrift*, 1922.

the additional assumptions necessary to determine the coefficients in any particular case, and show how these latter may then be calculated in a simple manner.

1. THE DEDUCTION OF THE LAW

Suppose that we are concerned with measuring groups of m quantities. We shall, for simplicity, assume that all groups are equally trustworthy, although the extension to the case of differently weighted groups is not difficult. We shall make certain assumptions about the distribution of errors, meaning, thereby, *accidental errors*, for we assume that constant errors have been removed.

ASSUMPTION 1: *The a priori probability that a group of quantities to be measured should take values in the infinitesimal region*

$$X \pm \frac{1}{2}dX, \quad Y \pm \frac{1}{2}dY, \quad Z \pm \frac{1}{2}dZ, \quad \dots,$$

where the points X, Y, Z, \dots lie in a certain continuous m -dimensional manifold S , will differ by an infinitesimal of higher order from the expression

$$f(X, Y, Z, \dots) dX dY dZ \dots,$$

where the function f is continuous with continuous first derivatives in S .

ASSUMPTION 2. *The probability that a group of quantities whose true values are X, Y, Z, \dots in S should be observed to have values, after the removal of constant errors, which lie in the infinitesimal region*

$$x \pm \frac{1}{2}dx, \quad y \pm \frac{1}{2}dy, \quad z \pm \frac{1}{2}dz, \quad \dots,$$

where (x, y, z, \dots) is a point of S , will differ by an infinitesimal of higher order from

$$\Phi(X, Y, Z, \dots, x, y, z, \dots) dx dy dz \dots,$$

where the function Φ is continuous with continuous first and second partial derivatives, and has a value independent of the choice of origin.

The last part of the assumption is plausible in practice, because if we are, for instance, measuring a length on a scale, the accidental errors will arise from various physical causes independent of the position of the 0. Moreover, it has a momentous consequence, for

$$\begin{aligned} \Phi(X, Y, Z, \dots, x, y, z, \dots) \\ = \Phi|0, 0, 0, \dots, x - X, y - Y, z - Z, \dots| \end{aligned}$$

Writing in the explicit values of the errors we have

$$\begin{aligned} \xi = x - X, \quad \eta = y - Y, \quad \zeta = z - Z, \quad \dots \\ \Phi \equiv \Phi(\xi, \eta, \zeta, \dots). \end{aligned}$$

that is to say, the probability for a system of errors is a function of those errors, and not of the true values and observed values considered as independent variables, a point which has been a stumbling block to some writers.

ASSUMPTION 3. *The mean value for the error on an individual variable is 0.*

This again is plausible, for a contrary assumption would show a tendency to favor positive or negative errors, and such a tendency we should naturally class with the constant errors, not with the accidental ones. As a further matter of notation let us write the averages

$$(1) \quad \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}, \quad \bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}, \\ \bar{z} = \frac{z_1 + z_2 + \cdots + z_n}{n}, \quad \dots$$

The actual errors are

$$(2) \quad \xi_i = x_i - X, \quad \eta_i = y_i - Y, \quad \zeta_i = z_i - Z, \quad \dots$$

The residual errors are

$$(3) \quad \delta_i = x_i - \bar{x}, \quad \epsilon_i = y_i - \bar{y}, \quad \dots, \\ x_i - \bar{x} = x_i - X - (\bar{x} - X), \\ \delta_i = \xi_i - \frac{\sum_j \xi_j}{n}.$$

The individual groups of observations are independent of one another; hence, by Assumption 3,

$$\text{Mean value } \xi_i \xi_j = 0, \\ \text{Mean value } \delta_i^2 = \frac{n-1}{n} \text{Mean value } \xi_i^2, \\ (4) \quad \text{Mean value } \xi_i^2 = \text{Mean value } \frac{\sum_j \delta_j^2}{n-1}, \\ \text{Mean value } \xi_i \eta_i = \text{Mean value } \frac{\sum_j \delta_j \epsilon_j}{n-1}.$$

ASSUMPTION 4. *If the infinitesimal increments dx , dy , dz , \dots be sufficiently small, the probability that the true values lie in the region*

$$\bar{x} \pm \frac{1}{2}dx, \quad \bar{y} \pm \frac{1}{2}dy, \quad \bar{z} \pm \frac{1}{2}dz, \quad \dots$$

is greater than that they lie in any other region of like structure about any other point.

We have now made a sufficient number of assumptions to enable us to deduce the analytic form for our functions. We do this, following the original method of Gauss, by calculating the probability that a given set of observations should have resulted from observing a group of quantities of assumed true value. The probability that the measurements $x_1, y_1, z_1, \dots, x_2, y_2, z_2, \dots, x_n, y_n, z_n, \dots$ were made on quantities whose true values are $X, Y, Z,$

... is by Bayes' theorem

$$\frac{f(X, Y, Z, \dots) \Phi_1 \Phi_2 \dots \Phi_n dXdYdZ \dots}{\int \dots \int f(X, Y, Z, \dots) \Phi_1 \Phi_2 \dots \Phi_n dXdYdZ \dots},$$

$$\Phi_i = \Phi(\xi_i, \eta_i, \zeta_i, \dots).$$

The integration in the denominator is supposed to be extended throughout the whole region S . This expression will be a maximum with the logarithm of its numerator. Equating to 0 the partial derivatives to X, Y, Z, \dots , we get

$$-\frac{\partial \log f}{\partial X} + \frac{\partial \log \Phi_1}{\partial \xi_1} + \frac{\partial \log \Phi_2}{\partial \xi_2} + \dots + \frac{\partial \log \Phi_n}{\partial \xi_n} = 0,$$

$$-\frac{\partial \log f}{\partial Y} + \frac{\partial \log \Phi_1}{\partial \eta_1} + \frac{\partial \log \Phi_2}{\partial \eta_2} + \dots + \frac{\partial \log \Phi_n}{\partial \eta_n} = 0.$$

One set of solutions will arise in case all of the observations have been correct, i.e.,

$$-\frac{\partial \log f}{\partial X} + n \frac{\partial \log \Phi(\xi, \eta, \dots)}{\partial \xi} = 0, \quad \xi = \eta = \dots = 0,$$

$$-\frac{\partial \log f}{\partial Y} + n \frac{\partial \log \Phi(\xi, \eta, \dots)}{\partial \eta} = 0, \quad \xi = \eta = \dots = 0.$$

Now, by definition, f is independent of n , hence

$$\frac{\partial \log f}{\partial X} = \frac{\partial \log f}{\partial Y} = \dots = 0, \quad f = \text{const.}$$

Let us underline the fact that we are considering probabilities for observations which do not go outside of the region S . We could not have f a constant throughout all space without a contradiction, and it is also evident that f must be rigorously 0 throughout most of space. The partial differential equations now take the simpler form

$$(5) \quad \begin{aligned} \frac{\partial \log \Phi_1}{\partial \xi_1} + \frac{\partial \log \Phi_2}{\partial \xi_2} + \dots + \frac{\partial \log \Phi_n}{\partial \xi_n} &= 0, \\ \frac{\partial \log \Phi_1}{\partial \eta_1} + \frac{\partial \log \Phi_2}{\partial \eta_2} + \dots + \frac{\partial \log \Phi_n}{\partial \eta_n} &= 0, \\ \dots &\dots \end{aligned}$$

These equations hold whenever

$$\begin{aligned} X = \bar{x}, \quad Y = \bar{y}, \quad Z = \bar{z}, \quad \dots, \\ \xi_1 + \xi_2 + \dots + \xi_n = 0, \\ \eta_1 + \eta_2 + \dots + \eta_n = 0, \\ \dots \end{aligned}$$

arose from taking all of these variables equal to zero. Furthermore, since this is a maximum, we know that the form is definite, i.e.,

The homogeneous quadratic form ψ^2 is positive and definite, with a non-vanishing discriminant.

2. DETERMINATION OF THE CONSTANTS*

It should be emphasized that everything which we have done so far is under the assumption that we are dealing with observations in the region S . We have found the probability that an observation in the region S should lie in a certain infinitesimal sub-region. In practice this is of no interest whatever until we have some idea of what the region S may be. It certainly could not be the whole of space, as the assumption that f is everywhere constant leads to a contradiction. On further consideration we notice two things. First of all, it seems quite plausible that f might be constant throughout a certain region, and equal to 0 almost everywhere else. Second, the expression (6) is excessively small, except in a very strictly confined space, and this rapid diminution of (6) would produce a result close to that of the vanishing of f . In other words, the error in calculating the constants will be very small if we assume that the formula (6) is universally valid. On the strength of this we make

ASSUMPTION 5. *For the purpose of calculating constants, formula (6) may be assumed true throughout all space.*

We note, secondly, that the only method for calculating our constants is to assume that certain observed values may be identified with their mean values as calculated by formula. The right sides of the last two equations (4) are proportional to the mean values of the averages of certain observed quantities, and we know by Tchebycheff's theorem† that it is highly likely that the value of an average shall be close to its mean value. This leads to

ASSUMPTION 6. *When the number of groups is large, the mean values of ξ^2 , η^2 , ξ , η , \dots may be equated to the observed values*

$$\frac{\sum_j \delta_j^2}{n-1}, \quad \frac{\sum_j \epsilon_j^2}{n-1}, \quad \frac{\sum_j \delta_j \epsilon_j}{n-1}, \quad \dots$$

These quantities give, of course, the probable errors of individual measure-

* The mathematical manipulation that follows depends on obvious applications of the theory of determinants. The methods and final formula are very close to Greiner, *Zeitschrift für Mathematik und Physik*, vol. 57, pp. 226 ff., and Pearson, *Philosophical Transactions of the Royal Society*, vol. 187, pp. 299 ff. Pearson assigns the credit to Edgeworth, *Philosophical Magazine*, ser. 5, vol. 34, p. 201. I must confess to finding Edgeworth so obscure that I do not know whether his result is like mine or not. Moreover, none of these writers seem to me to set forth the underlying assumptions with desirable clearness.

† Tchebycheff, *Oewres*, Petrograd, 1899, vol. 1, p. 687.

ments and their correlation coefficients two by two. It is interesting that these should be the only independent constants.

In order to clarify the manipulation, we shall at this point take the perilous step of changing our notation. Logically, the notation we are now going to adopt might well have been used from the start, but the resulting summation formulas, by their very compactness, are rather obscure, and it is easier to see what is really going on, by using the more diffuse symbolism with many continuation signs which we have employed so far. For the actual errors committed, we shall write

$$(7) \quad \xi = x_1, \quad \eta = x_2, \quad \zeta = x_3, \quad \dots,$$

there being m in all. The n sets of m residuals shall be written

$$\delta_{11}, \delta_{12}, \dots, \delta_{1m}; \quad \delta_{21}, \delta_{22}, \dots, \delta_{2m}; \quad \dots; \quad \delta_{n1}, \delta_{n2}, \dots, \delta_{nm}.$$

Our fundamental formula (6) may now be written

$$(8) \quad \Phi \equiv R e^{-\sum a_{ij} x_i x_j}, \quad a_{ij} = a_{ji}.$$

The assumptions 5 and 6 may be expressed by the equation

$$(9) \quad p_{ij} \equiv \frac{\sum_k \delta_{ki} \delta_{kj}}{n-1} = R \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i x_j e^{-\sum a_{ij} x_i x_j} dx_1 dx_2 \dots dx_m.$$

Since the discriminant of our quadratic form is not zero, we may find a linear transformation

$$(10) \quad x_i = \sum_k c_{ik} x'_k, \quad |c_{ij}| \neq 0,$$

such that

$$(11) \quad \begin{aligned} \sum_{i,j} a_{ij} x_i x_j &\equiv \sum_{i,j,k,l} c_{ik} c_{jl} a_{ij} x'_k x'_l \\ &= \sum_r b_r x_r'^2, \\ \sum_{i,j} c_{ik} c_{jl} a_{ij} &= 0, \quad k \neq l, \\ \sum_{i,j} c_{ir} c_{jr} a_{ij} &= b_r. \end{aligned}$$

Since the discriminant is an invariant of weight 2,

$$b_1 b_2 \dots b_m = |c_{ij}|^2 |a_{ij}|.$$

The inverse of the substitution, contragredient to (10), is

$$\begin{aligned} w'_k &= \sum_i c_{ik} w_i, \\ w_k'^2 &= \sum_{i,j} c_{ik} c_{jk} w_i w_j. \end{aligned}$$

In the projective space of $n-1$ dimensions where a point has the homogeneous coördinates x_1, x_2, \dots, x_n the hyperquadric

$$\sum_i a_{ij} x_i x_j = 0$$

has the tangential equation

$$\sum_{i,j} A_{ij} w_i w_j = 0, \quad A_{ij} = \frac{\partial |a_{ij}|}{\partial a_{ij}}.$$

In terms of the new variables we have

$$\begin{aligned} \sum_r b_r x_r'^2 &= 0, & \sum_r \frac{w_r'^2}{b_r} &= 0, \\ (12) \quad |b_1 b_2 \cdots b_m| \sum_r \frac{w_r'^2}{b_r} &= |c_{ij}|^2 \sum_{i,j} A_{ij} w_i w_j, \\ |a_{ij}| \sum_r \frac{c_{ir} c_{jr}}{b_r} &= A_{ij}. \end{aligned}$$

We may express (9) in terms of the new variables. The only point to remember is that the jacobian of the transformation is $|c_{ij}|$;

$$p_{ij} = R |c_{ij}| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{k,l} c_{ik} c_{jl} x'_k x'_l e^{-\sum_r b_r x_r'^2} dx'_1 dx'_2 \cdots dx'_m.$$

This simplifies greatly because

$$\int_{-\infty}^{\infty} x'_k e^{-b_k x_k'^2} dx'_k = 0.$$

Hence

$$p_{ij} = R |c_{ij}| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_k c_{ik} c_{jk} x_k'^2 e^{-\sum_r b_r x_r'^2} dx'_1 dx'_2 \cdots dx'_m.$$

We have the well known integrals

$$\int_{-\infty}^{\infty} e^{-b_k x_k'^2} dx'_k = \frac{\sqrt{\pi}}{\sqrt{b_k}}; \quad \int_{-\infty}^{\infty} x_k'^2 e^{-b_k x_k'^2} dx'_k = \frac{\sqrt{\pi}}{2b_k^{3/2}}.$$

Hence

$$p_{ij} = R |c_{ij}| \frac{\pi^{m/2}}{2\sqrt{b_1 b_2 \cdots b_m}} \sum_k \frac{c_{ik} c_{jk}}{b_k},$$

or, by (12),

$$p_{ij} = \frac{R}{2} \frac{|c_{ij}|}{|a_{ij}|} \frac{\pi^{m/2}}{\sqrt{b_1 b_2 \cdots b_m}} A_{ij} = \frac{R \pi^{m/2}}{2} \frac{A_{ij}}{|a_{ij}|^{3/2}}.$$

Furthermore, since the probability of some group of errors is 1,

$$\begin{aligned} 1 &= R \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum a_{ij} x_i x_j} dx_1 \cdots dx_n, \\ 1 &= \frac{R \pi^{m/2} |c_{ij}|}{\sqrt{b_1 b_2 \cdots b_m}} = \frac{R \pi^{m/2}}{\sqrt{|a_{ij}|}}. \end{aligned}$$

Dividing out R , we have

$$p_{ij} = \frac{A_{ij}}{2|a_{ij}|}.$$

In these equations the quantities p_{ij} are known; we wish to find the quantities a_{ij} . We first introduce one more symbol:

$$P_{ij} = \frac{\partial |p_{ij}|}{\partial p_{ij}}.$$

Since the process of interchanging each element of a non-vanishing determinant with its cofactor is an involutory one, except for multiplication by a power of the determinant, we must have

$$a_{ij} = MP_{ij}.$$

We calculate M by a little jugglery:

$$\begin{aligned} |a_{ij}| &= M^m |p_{ij}|^{m-1}, \\ |p_{ij}| &= \frac{|A_{ij}|}{2^m |a_{ij}|^m} = \frac{1}{2^m |a_{ij}|}, \\ M &= \frac{1}{2|p_{ij}|}. \end{aligned}$$

Let us exhibit our results in the form of a table:

Assumed errors: x_1, x_2, \dots, x_n .

Gaussian law of error for n variables: $\Phi = Re^{-\sum a_{ij}x_j}$.

Observed residuals: $\delta_{11}, \delta_{12}, \dots, \delta_{1m}; \delta_{21}, \delta_{22}, \dots, \delta_{2m}; \dots; \delta_{n1}, \delta_{n2}, \dots, \delta_{nm}$.

$$\begin{aligned} p_{ij} &= \sum_k \frac{\delta_{ki} \delta_{kj}}{n-1}; & P_{ij} &= \frac{\partial |p_{ij}|}{\partial p_{ij}}; \\ a_{ij} &= \frac{P_{ij}}{2|p_{ij}|}; & R &= \frac{1}{\sqrt{(2\pi)^m |p_{ij}|}}. \end{aligned}$$

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.