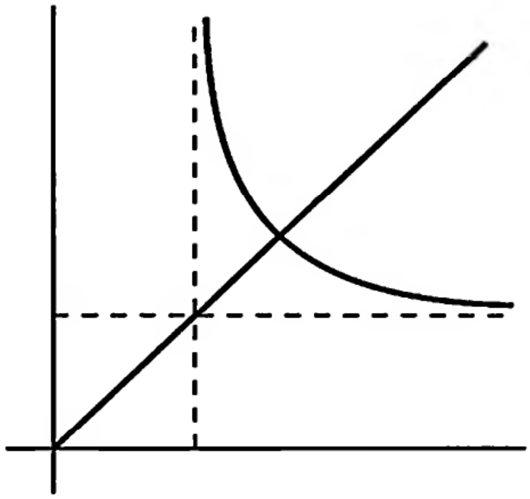

FIGURE 7

38. Show that

$$\int_0^1 \frac{dx}{x^x} = \sum_1^\infty \frac{1}{n^n}.$$

2. The Law of Errors

> Everybody believes in it, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact.
>
> —G. Lippmann (French physicist, 1845–1921)

We begin with a problem encountered by every experimental scientist.

No physical quantity can ever be measured with perfect accuracy—all observations are subject to error. And, if several measurements of the same magnitude are made, they will invariably differ from each other. The imperfections of our senses or of our best instruments imply an uncertainty that is said to be accidental, or random, and which tends to occur however great the skill and conscientiousness of the observer. A physicist, for example, who makes several determinations of the speed of light observes that they do not agree exactly. He rejects the notion that the speed is changing and attributes the variations in his results to unavoidable errors of observation. We do not know the origin of random errors—"we attribute them to chance because their causes are too complicated and too numerous." [14] Like the random noise that pervades an electronic communication channel, random errors are inevitable.

The problem was taken up by the great German mathematician Carl Friedrich Gauss. Gauss devoted much of his life to astronomy, and in his famous treatise of 1809, *Theoria Motus Corporum Coelestium in Sectionibus Conics Solum Ambientium* (Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections)—a masterful investigation of the mathematics of planetary orbits—he put forth a theory of errors of observation that survives to this day.

In that work, Gauss formulates a principle upon which his entire theory is to be built.

Principle of the Arithmetic Mean. *When any number of equally good observations have given*

$$x_1, x_2, \ldots, x_n$$

as the values of a certain magnitude, the most probable value is their arithmetic mean.

What could be a simpler or more natural way of arriving at an equitable balance, an equilibrium, among a given set of observations? In fact, when the x's are points in the plane or in space, their arithmetic mean is precisely the balancing point, or centroid, of the system formed by placing equal masses at each point. But Gauss's approach to the theory of errors is to be based, not on a mechanical equilibrium, but on probability. Without the quantification of uncertainty, an answer to the question "How good is best?" was not possible.[15] Starting from the principle of the arithmetic mean, Gauss derived his famous law of errors, an elegant law that governs the probability that a single measurement x will lie between two given limits.

If μ is the true value of the magnitude being observed, then the observational error is the deviation

$$\text{error} = \mu - x.$$

The problem is to determine the form of the *error function,* a positive function $\Phi(x)$ having the property that, for any given measurement,

[14] Henri Poincaré, *Science and Method* (Dover, 1952), p. 75. For an elementary discussion of random error in the context of precision weighing done at the National Bureau of Standards in Washington, see D. Freedman, R. Pisani, and R. Purves, *Statistics* [160], chapters 6, 24.

[15] S. M. Stigler, *The History of Statistics* [403], p. 140.

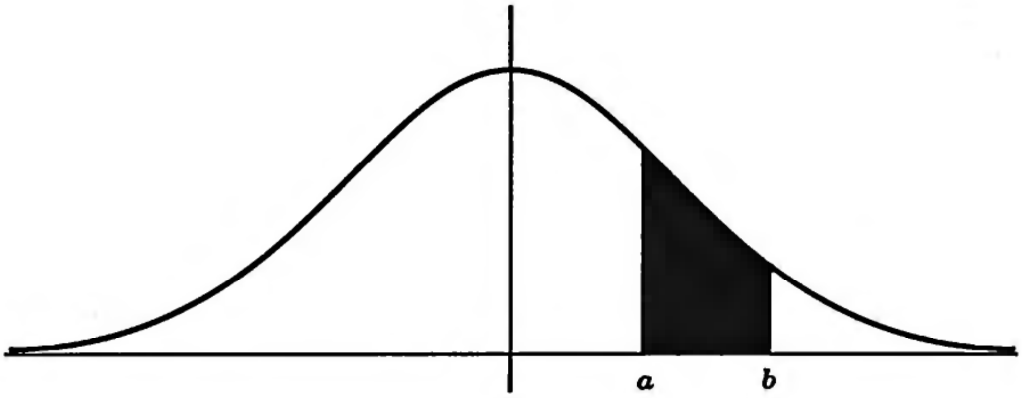

FIGURE 8

$$\text{Probability}\ [a < \text{ error } < b] = \int_a^b \Phi(x)\, dx.$$

Properly speaking, $\Phi(x)$ ought to be a discontinuous function, taking on only a finite number of distinct values. For any measuring device, no matter how sensitive it may be, cannot be accurate beyond a certain limit. But there is a more serious objection.

> As our mental eye penetrates into smaller and smaller distances and shorter and shorter times, we find nature behaving so entirely differently from what we observe in visible and palpable bodies of our surrounding that *no* model shaped after our large-scale experiences can ever be 'true'.... The idea of a *continuous range,* so familiar to mathematicians in our days, is something quite exorbitant, an enormous extrapolation of what is really accessible to us. The idea that you should *really* indicate the exact values of any physical quantity—temperature, density, potential, field strength, or whatever it might be—for *all* the points of a continuous range, say between zero and 1, is a bold extrapolation. We *never* do anything else than determine the quantity approximately for a very limited number of points and then 'draw a smooth curve through them'. This serves us well for many practical purposes, but from the epistemological point of view, from the point of view of the theory of knowledge, it is totally different from a supposed exact continual description.[16]

[16] E. Schrödinger, *Science and Humanism* [371], pp. 25ff.

To simplify the analysis, it is convenient to disregard these practical difficulties and to consider an ideal case in which measurements may range over all real values. The discrete error function can then be replaced by a continuous one.[17]

There are three simple properties that $\Phi(x)$ should satisfy:

i. Since errors that are of the same magnitude but of opposite sign are equally likely,

$$\Phi(x) = \Phi(-x).$$

ii. Since small errors are more likely than larger ones and extremely large errors are negligible, $\Phi(x)$ should decrease rapidly to zero as x approaches infinity.

iii. Since it is certain that each error will take on some real value,

$$\int_{-\infty}^{\infty} \Phi(x)\, dx = 1.$$

These conditions certainly do not determine $\Phi(x)$ uniquely, and many mathematicians before Gauss—most notably, Euler, Laplace, and Legendre—had long sought the elusive error function. It was Gauss who finally provided the solution.[18]

Theorem (The Normal Law of Errors). *The only error function that makes the arithmetic mean the most probable value is*

$$\Phi(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}.$$

The constant h serves as a measure of the accuracy of the observations.

Proof. While the proof is demanding, each step along the way is elementary, and the reader who perseveres will have mastered the ingredients of a most important mathematical model. It is convenient to present the proof in three steps.

---

[17] The idea of using the continuous case to model the discrete case was once a daring novel idea, known as *Boscovich's Hypothesis* (R. J. Boscovich, *Treatise on Natural Philosophy,* Venice, 1758).

[18] The basic result was obtained independently and almost simultaneously by the American mathematician Robert Adrian (1775–1843) [289], p. 149.

Step 1. What exactly is meant by "most probable value"? To motivate the answer, let us assume for a moment that we are dealing with the discrete case in which only a finite number of different measurements are possible. The error function is then discontinuous and $\Phi(x)$ represents the probability that a single measurement differs from the true value μ by the amount x.

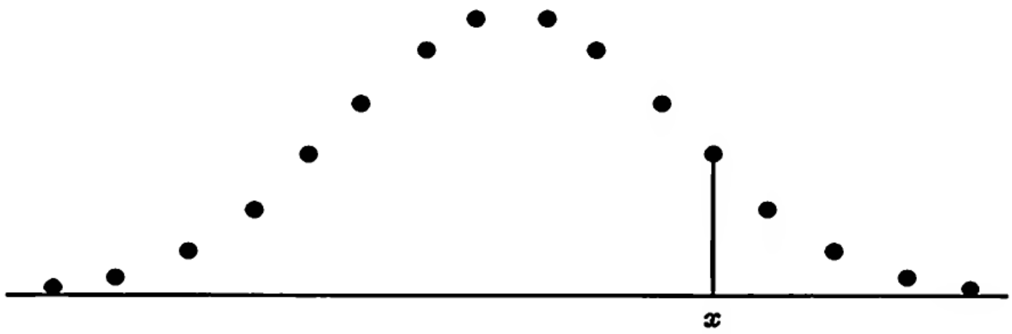

FIGURE 9

Now, it is entirely reasonable to suppose that repeated observations are *independent.* (Recall that two events are said to be independent if the probability of their joint occurrence is the product of the probabilities of their individual occurrences. It is natural to regard events which seem unrelated as being independent of each other. The mathematical rule of multiplication of probabilities is a way of formalizing this vague and intuitive notion.)

In view of the preceding comments, if n equally good observations have produced the values

$$x_1, x_2, \ldots, x_n$$

(with corresponding errors $\mu - x_1, \mu - x_2, \ldots, \mu - x_n$), then the product

$$\Phi(\mu - x_1)\Phi(\mu - x_2)\cdots\Phi(\mu - x_n)$$

represents the likelihood that all these values were observed.

But μ is unknown—it can never be known—so Gauss asserts that its most probable value is that for which the likelihood of the observations is a maximum. [19] By adopting the principle of the arithmetic mean, we are stipulating

[19] For a more complete explanation of why this choice of μ is the "most probable value" see Whittaker and Robinson, *The Calculus of Observations* [448], §112. In §110, it is shown how the principle of the arithmetic mean can be deduced from other axioms of a more elementary nature.

that this maximum must occur when

$$\mu = (x_1 + \cdots + x_n)/n.$$

To pass to the continuous case involves no essential change in the argument and we are thereby led to the same maximum problem (see, for example, [448], §112).

Step 2. Question: What form must the error function take if, given any three observations a, b, and c, the product

$$\Phi(x-a)\Phi(x-b)\Phi(x-c) \tag{1}$$

assumes its largest value when

$$x = \frac{a+b+c}{3}? \tag{2}$$

Let us suppose, for simplicity, that Φ is continuously differentiable. Setting the logarithmic derivative of (1) equal to zero to obtain the maximum, we find

$$\frac{\Phi'(x-a)}{\Phi(x-a)} + \frac{\Phi'(x-b)}{\Phi(x-b)} + \frac{\Phi'(x-c)}{\Phi(x-c)} = 0.$$

Let

$$F(x) = \frac{\Phi'(x)}{\Phi(x)},$$

which is then defined and continuous for all real values of x, and observe that (2) holds if and only if $(x-a)+(x-b)+(x-c)=0$. Since a, b, and c are arbitrary, it follows that $F(x)+F(y)+F(z)=0$ whenever $x+y+z=0$. Equivalently,

$$F(x+y) = F(x) + F(y) \tag{3}$$

for all x and y.

This functional equation is quite reminiscent of another that we encountered in connection with the self-similarity of the logarithmic spiral (see chapter 3, §2). Arguing as before, we find that the only continuous solutions of (3) are the scalar multiples of the identity function. Thus

$$\frac{\Phi'(x)}{\Phi(x)} = Ax$$

and

$$\Phi(x) = Be^{Ax^2/2}.$$

But the relation

$$\int_{-\infty}^{\infty} \Phi(x)\,dx = 1$$

shows that A must be negative, say $-2h^2$, so that

$$\frac{1}{B} = \int_{-\infty}^{\infty} e^{-h^2x^2}\,dx = \frac{\sqrt{\pi}}{h}.^{20}$$

Thus the only error function that makes the arithmetic mean the most probable value is

$$\Phi(x) = \frac{h}{\sqrt{\pi}}e^{-h^2x^2}. \tag{4}$$

Step 3. It remains only to verify the correctness of the solution, namely, to show that if Φ is given by (4) then the principle of the arithmetic mean is valid, not just for three observations, but for any finite number as well. If the values $x_1, x_2, \ldots, x_n$ have been observed, then the function to be maximized is

$$\Phi(x - x_1)\Phi(x - x_2)\cdots\Phi(x - x_n).$$

By combining the exponentials, we see that this entails *minimizing the sum of the squares of the errors*[21]

$$(x - x_1)^2 + (x - x_2)^2 + \cdots + (x - x_n)^2.$$

But the minimum is obtained when

$$(x - x_1) + (x - x_2) + \cdots + (x - x_n) = 0,$$

---

[20] The second equality follows from the beautiful and well-known result

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

For a proof see chapter 5, §2.

[21] This is an example of the famous *method of least squares,* introduced by Legendre in 1805 in his *Nouvelles méthodes pour la détermination des orbites des comètes.* Commenting on the method, Stigler writes:

> The method of least squares was the dominant theme—the leitmotif—of nineteenth-century mathematical statistics. In several respects it was to statistics what the calculus had been to mathematics a century earlier. "Proofs" of the method gave direction to the development of statistical theory, handbooks explaining its use guided the application of the higher methods, and disputes on the priority of its discovery signaled the intellectual community's recognition of the method's value. ([403], p. 11)

that is, when x is the arithmetic mean of the observations. This completes the proof.

The curve

$$y = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}$$

is known generally as the **normal curve**, or the **Gaussian curve**, although it was introduced in probability theory nearly a century earlier by De Moivre (1718). It is perhaps the most ubiquitous of curves, appearing throughout the physical, biological, and social sciences. Contained within it are two of the most famous constants in mathematics, π and e, which seem to appear mysteriously in so many settings involving randomness and chance.

For each choice of the constant h, the graph is a symmetric, bell-shaped curve, as pictured below. Gauss called h the *modulus of precision*, and it reflects the accuracy of the observations. For large values of h, most of the area under the curve is concentrated near the y-axis, indicating that in this case most of the observations fall very close to the true value. For small values of h, the curve flattens out and the bulk of the observations are diffuse.

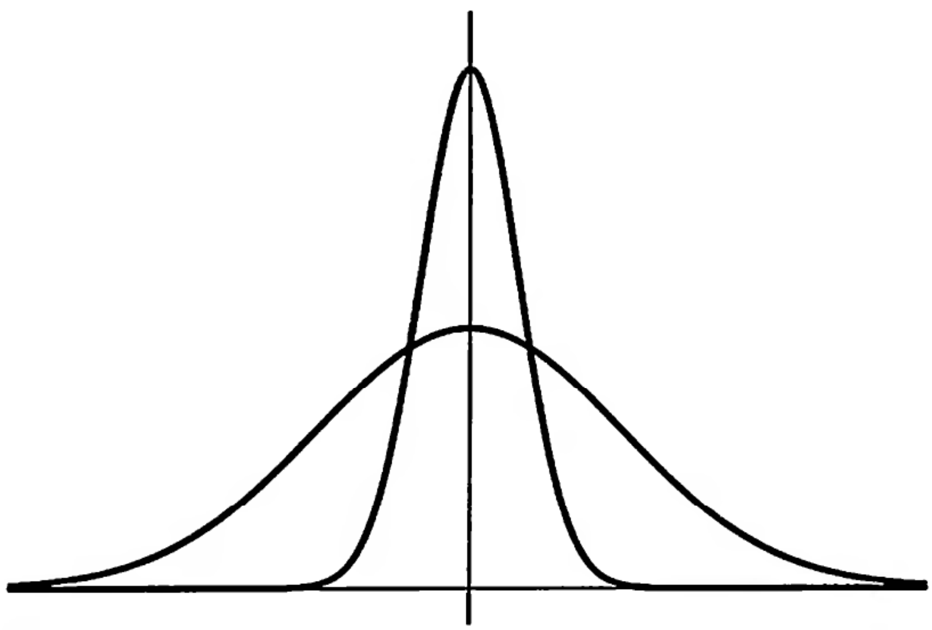

FIGURE 10

Gauss's derivation of the normal law rests on the principle of the arithmetic mean, yet, almost immediately, Laplace saw that a better rationale could be provided. If each deviation from the average measure were itself the cumulative effect of a very large number of independent causes, each one of them having only a very slight influence on the whole, then the normal law would follow.

Here, certainly, is one of the finest examples of order out of chaos in all of probability theory. A single component of an individual measurement can be totally unpredictable, and yet the cumulative effect of many such components will be governed by a deterministic law.

> We know not to what are due the accidental errors, and precisely because we do not know, we are aware they obey the law of Gauss. Such is the paradox. The explanation is nearly the same as in the preceding cases. We need know only one thing: that the errors are very numerous, that they are very slight, that each may be as well negative as positive. What is the curve of probability of each of them? We do not know; we only suppose it is symmetric. We prove then that the resultant error will follow Gauss's law, and this resulting law is independent of the particular laws which we do not know. Here again the simplicity of the result is born of the very complexity of the data.[22]

The central limit theorem, as Laplace's result came to be known, placed Gauss's principle of the arithmetic mean on a firm logical foundation. "Only after this work of Laplace did the widespread applications of probability theory become feasible as a scientifically justified method."[23]

While a precise formulation and proof of the central limit theorem lies well beyond our scope, it is instructive to illustrate it by mentioning a famous piece of apparatus that was devised by the English statistician and natural scientist Sir Francis Galton (1822–1911). Galton called it the *quincunx*,[24] and it illustrates the principle of the law of errors from the joint effect of a large number of small and independent deviations.

The quincunx had a glass face and a funnel at the top. Small ball bearings poured through the funnel cascaded through a triangular array of pins and ul-

---

[22] Henri Poincaré, op. cit. This translation comes from *The World of Mathematics* [316], volume 2, p. 1389.

[23] L. E. Maistrov, *Probability Theory: A Historical Sketch* [289], p. 148. With Laplace's theorem as a starting point, Adolphe Quetelet (1796–1874), the Belgian astronomer and statistician, succeeded in showing that the normal curve could be fitted to a large variety of empirical data taken from all corners of science. Heights and weights of individuals, sizes of skulls, and today even IQ scores appear to be normally distributed. Nature, it seems, can be counted on to obey the laws of probability.

[24] The word is used in *Natural Inheritance*, p. 64, to describe the arrangement of the pins. The description given here is adapted from Stigler [403], p. 276.

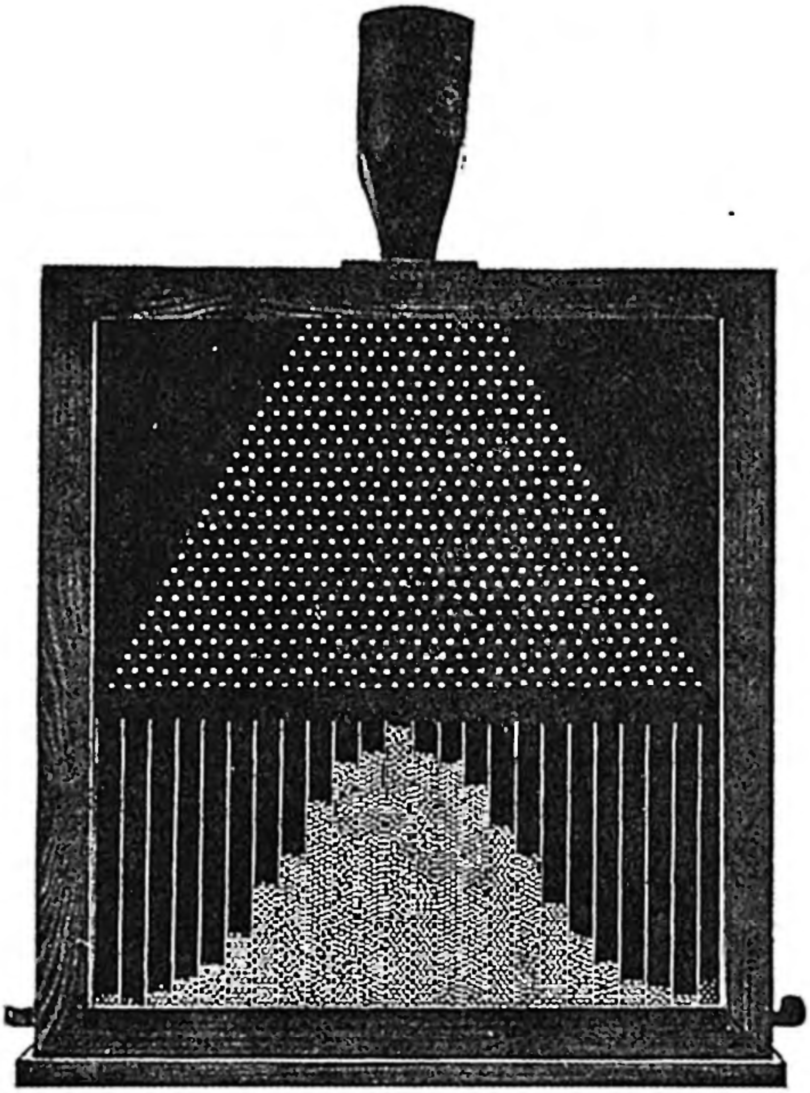

FIGURE 11
Galton's apparatus
From G. Weber, *Belysningsteknik*, 2nd edition, p. 132.

timately collected in compartments at the bottom. The construction was such that each ball bearing would strike one pin in every row and, at least in principle, fall either left or right with equal probabilities. What should we observe? Since each encounter with a pin subjects a ball bearing to a small displacement, equally likely to be left or right, the total displacement—as measured from the compartment directly below the funnel—was the sum of as many independent displacements as there were rows of pins (twenty-six in Figure 11). The resulting outline after many ball bearings were dropped should resemble a normal curve.

By examining the configuration of the pins and asking the obvious combinatorial question—In how many ways can a ball bearing reach a given pin?—the reader will discover almost at once the reappearance of Pascal's celebrated arithmetical triangle (problem 8). Here begins another fascinating excursion into the unexpected connection between the binomial theorem and the normal law (see, for example, [156], volume I, chapter VII).

PROBLEMS

1. In memoirs of 1777 and 1781, Laplace gave an extremely complex argument to show why the function

$$y = \frac{1}{2a} \log\left(\frac{a}{|x|}\right), \qquad |x| \le a$$

should be taken as an error distribution. Here, a represents the upper limit of the possible errors. Does Laplace's function satisfy all three properties required of an error function?

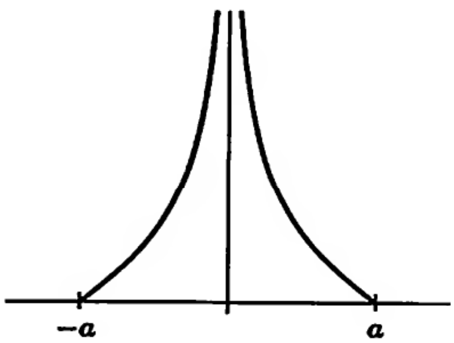

FIGURE 12

2. Given a triangle ABC, we form a new triangle $A'B'C'$ by connecting the midpoints of the sides (see Figure 13). When this process is repeated indefinitely the resulting sequence of similar triangles converges to a common intersection point P. Identify P in terms of A, B, and C.