Hence

$$\frac{du}{dx} = f_1 (x, u, v, \ldots, w).$$

Similar results hold for the other functions. *The functions $u, v, \ldots, w$ are consequently the functions sought.*

CALCUL DES PROBABILITÉS. Par J. BERTRAND, de l'Académie Française, Secrétaire perpétuel de l'Académie des Sciences. Paris, Gauthier-Villars, 1889. 8vo., LVII + 332 pp.

THERE is possibly no branch of mathematics at once so interesting, so bewildering, and of so great practical importance as the theory of probability. Its history reveals both the wonders that can be accomplished and the bounds that cannot be transcended by mathematical science. It is the link between rigid deduction and the vast field of inductive science. A complete theory of probability would be a complete theory of the formation of belief. It is certainly a pity then, that, to quote M. Bertrand, "one cannot well understand the calculus of probabilities without having read La Place's work," and that "one cannot read La Place's work without having prepared himself for it by the most profound mathematical studies."

Though not otherwise is thorough knowledge to be gained, yet an exceedingly useful amount of knowledge is to be had without such effort. In fact, M. Bertrand's forty odd pages of preface on "The Laws of Chance" give an insight into the theory without the use of so much as a single algebraic symbol.

Listen to this *reductio ad absurdum* of Bernoulli's theory of moral expectation:

"'If I win,' says poor Peter, proposing a game of cards to Paul, 'you must pay three francs for my dinner.' 'Meal for meal,' replies Paul, 'you should pay twenty francs in case you lose, for that is the price of my supper.' 'If I lose twenty francs,' cries Peter, frightened out of his wits, 'I cannot dine to-morrow: without coming to that you might lose a thousand; put them up against my twenty. According to Daniel Bernoulli, you will still have the advantage.'"

Even more complete is the upsetting of Condorcet's calculation as to the probability of the sun's rising.

"Paul would wager that the sun rises to-morrow. The theory fixes the stakes. Paul shall receive a franc if the sun rises and will give a million if it fails to do so. Peter accepts the wager. Each morning he loses his franc and pays it. As the sun rises from morning to morning his chance daily diminishes. Paul conscientiously increases his stake; Peter as conscientiously pays his franc. The obligations remain equitable. The bettors travel through twenty countries from West to East. Peter always loses; he pursues his fortune however and takes Paul to the North; they cross the arctic circle; the sun stays a month below the horizon; Paul loses 30 millions and thinks the order of nature overturned."

Even La Place does not escape M. Bertrand's pleasant raillery, and M. Quetelet has his ideal average man dismissed as follows:

"In the body of the average man the Belgian author places an average soul. . . . The typical man would be without passions and without vices, neither foolish nor wise, neither ignorant nor learned, forever dozing: this is the mean between sleeping and waking; answering neither yes nor no; mediocre in everything. After having eaten for 38 years the average ration of a healthy soldier, he would die, not of old age, but of some average sickness that statistics would reveal to him."

But I must hasten on to the main body of the work: suffice it to say that in these few introductory pages is packed this variety of topics:

The Petersburg paradox, D'Alembert and Bernoulli's dispute as to the benefits of inoculation for small-pox, Bernoulli's theorem, the ruin of players, inverse probabilities; Poisson's law of large numbers, the application of the theory of probability to statistics, the theory of errors of observation, the probability of decisions.

Having seen so much done without any algebra at all, we are prepared to accept M. Bertrand's statement, as to the entire work, that "very few pages could embarrass a reader familiar with the elements of mathematics."

Scarce an integration sign, never a generating function, really it is charmingly simple and direct: and everywhere illuminated too by the common sense and mother wit that are so conspicuously displayed in the preface.

It is characteristic of the method of treatment that all lurking dangers, all insidious snares, are carefully pointed out by means of numerous concrete examples.

A good instance of this, though merely one of half a dozen bearing upon the same point, is the following problem to show the absurdity of trying to reckon probabilities when the favorable and possible cases are each infinite in number.

In a circle a chord is drawn at random. What is the probability that it shall be longer than the side of the inscribed equilateral triangle?

You might say: The probability is unchanged by fixing one end of the chord. The probability that it shall be long enough is then merely the probability that it does not lie without the angle made by the two chords of 120° meeting at the point. This probability is $\frac{1}{3}$.

Or, you might say: The direction matters not, provided the chord is not too far from the centre, viz., not more than half the radius of the circle. The probability is $\frac{1}{2}$.

Yet again: To choose a chord at random is to choose its middle point. In order that the chord shall be long enough its middle point must not be without a circle concentric with, and of half the radius of, the given circle. This gives the probability $\frac{1}{4}$.

Similarly, in a chapter on total and compound probabilities, are a number of problems showing how essential it is, in computing the probability of a compound event, that the probability to be given to the second composing event shall be that which it has when the first event is *known to have happened.*

Even Clerk Maxwell has violated this principle in obtaining the formula

$$\varphi(x) = Ge^{-k^2x^2}$$

in which x is the x-component of the velocity of a molecule of gas and $\varphi(x)$ its probability. He assumed that the x, y, and z components had independent probabilities. That they are not independent is plain enough from this: if x is the maximum velocity of a molecule, the movement is parallel to the x-axis, and y and z are nothing.

In the chapter on expectation the consideration of the Petersburg paradox gives an opportunity to again ridicule Daniel Bernoulli's moral expectation. I quote a few passages:

"You play, that is the hypothesis. Are you foolish or wise to do so? The question is not put."

"Peter, whose whole fortune is 100,000 francs, wishes a chance to gain 100 millions. 'Nothing is easier,' coolly replies the geometer whom he consults. 'If the game is equitable you will have 999 chances in a thousand of losing your 100,000 francs.'"

"The theory of moral expectation has become classic, never was the word more exactly used: it has been studied, it has been commented upon, it has been developed in works justly famous. Its success stops there, no one ever has made or ever can make any use of it."

Two chapters are devoted to James Bernoulli's famous theorem that no possible event can be so improbable but that there is as great a chance as you please of its happening some time or other, if only you wait long enough.

Three distinct demonstrations are given, of which the first is the most straightforward and complete. I give a sketch :

The probabilities of two contrary events being p and q, the most probable combination in μ trials is that in which the first event happens μp times and the second μq times. By Sterling's theorem its probability is

$$1 \,/\, \sqrt{2\pi\mu pq}\,;$$

and the probability that it happens $\mu p \pm h$ times is

$$e^{-h^2/2\mu pq} \,/\, \sqrt{2\pi\mu pq}.$$

Approximately true for small values of h, it does to take this formula true for large and even infinite values, because then both the true values and those given by the formula are so small as to be negligible ; e. g. :

Putting, $\mu = 1000$, $h = 100$, $p = q = \frac{1}{2}$, we get for the corresponding probability

$$e^{-20}\sqrt{2} \,/\, \sqrt{1000\,\pi} = 0.000\,000\,000\,520\,06.$$

By a simple artifice, we substitute for (1), giving the probability of an error h, the formula

$$e^{-z^2/2\mu pq}dz \,/\, \sqrt{2\pi\mu pq},$$

for the probability of an error between z and $z + dz$.

To test this formula, notice that the sum of all possible probabilities is certainty and that we should have and do have

$$\int_{-\infty}^{+\infty} e^{-z^2/2\mu pq}dz \,/\, \sqrt{2\pi\mu pq} = 1.$$

Other such tests can be given.

The probability then that an error shall be smaller than α is

$$2\int_0^{\alpha/2\mu pq} e^{-t^2}\,dt = \Theta\,(\alpha \,/\, 2\mu pq).$$

If α has a determinate value, however large, the probability that it shall not be surpassed approaches zero as μ is increased without limit, which proves Bernoulli's theorem.

The *relative* error grows smaller and smaller as the *absolute* error grows larger and larger.

To fix the meaning of this, consider how many times a coin would need to be tossed in order that the probability of obtaining heads at least a million more times than tails shall exceed 0.01. We get, putting μ for the required number,

$$0.99 = \Theta\left(1\,000\,000\ \sqrt{2}\,/\,\sqrt{\mu}\right) = \Theta\,(1.83)$$

and thence

$$\mu = 597\,211 \text{ millions.}$$

Not only when the probability of an event is known can the number of its happenings in a given number of trials be predicted almost with certainty, but when the probability is unknown, Bernoulli's theorem can be used inversely to find it. The ratio of the number of happenings of an event to the number of trials certainly approaches this unknown probability as the number of trials is increased.

Nevertheless, two conditions are necessary : the probability must not change during the trials, and it must have a determinate value.

"The King of Siam is forty years old, what is the probability that he shall live ten years? It is different for us than for those who have asked his doctor, different for the doctor than for those who have received his confidences; very different indeed for the conspirators who have taken steps to strangle him to-morrow."

In a word, Bernoulli's theorem applies to *objective*, not to *subjective*, probability.

An immediate consequence of the theorem is the inevitable ruin of any gambler who plays long enough at a fair game. But the number of games before ruin occurs may be enormous. Thus, in tossing coins at one franc a toss it requires 624 000 tossings to insure a probability of 0.9 that one player or the other shall lose 100 francs. Truly a courageous gambler would hardly be frightened at such a prospect.

If in this very game either party had started in with only a franc and was to receive a franc a game so long as he played without losing it, his mathematical expectation would be infinite.

Let us return to the inverse use of Bernoulli's theorem. Consider this problem :

An urn contains μ balls, some white, some black, in an unknown proportion. k drawings are made, the ball being replaced after each drawing. There are obtained m white and n black balls. What is the most probable composition of the urn?

Before the trial is made, all sorts of hypotheses are possible as to the composition of the urn. Suppose all compositions are equally probable. It follows that the probable ratio of white to black balls is $m : n$; and, that there should be a deviation ε from this ratio, the probability is

$$Ge^{-\epsilon^2 (m+n)/2pq},$$

where G is independent of ε.

The hypothesis that all compositions are *a priori* equally probable is rarely realized. Suppose that the balls were put into the urn by lot with a probability $\frac{1}{2}$ for each color.

We then get for the probable proportion of white balls

$$(\mu + 2m)/2\,(\mu + m + n),$$

which lies between $\frac{1}{2}$ and $m/(m + n)$.

If μ is very large this will approach $\frac{1}{2}$, no matter what the numbers m and n are; if, on the contrary, it is m and $m + n$ that are large, the fraction is very near to $m/(m + n)$, no matter what μ is.

The probability of causes is always thus affected by *a priori* probabilities.

To what is commonly known as the theory of least squares three chapters are devoted. In fact, a fourth chapter, on "Errors in the Position of a Point," is really an extension of Gauss's law of errors.

Very interesting is the criticism of Gauss's reasoning.

To begin with, can it be strictly maintained that the probability of an error Δ is $\varphi(\Delta)$?

"Does it not depend upon the quantity measured ?"

"If you take a weight, if you measure an angle, is there not a greater chance of a correct estimate if the weight is an exact number of milligrams, if the angle contains an exact number of seconds, than if it is necessary to add a fraction ? If this fraction, not given by the instrument, is exactly $\frac{1}{2}$, is there not a less chance of error in evaluating it than if it is 0.27 ?"

There is a case where the postulate is rigorously demonstratable, but the conclusion is nevertheless only approximate. Suppose, in fact, that the quantity to be measured is the proportion of white balls in an urn of unknown composition.

Of μ balls drawn, m are white.

The fraction m/μ is a measure of the ratio sought. The measure is the more precise as the number of balls drawn is greater. The operation repeated n times gives the n successive measures.

$$m_1/\mu,\ m_2/\mu,\ \ldots,\ m_n/\mu.$$

The most probable value of the ratio deduced from the drawings is

$$\Sigma\, m/n\mu = \Sigma\,(m/\mu)/n,$$

the arithmetical mean of n equally trustworthy measures.

Now, if in μ drawings from an urn we get m white balls, the probability that the ratio of white to the whole number of balls shall be m/μ is indeed *approximately*

$$\mu e^{-z^2\mu^2/2m(\mu-m)}/\sqrt{2\pi m(\mu-m)},$$

which is of the form

$$ke^{-k^2h^2}/\sqrt{\pi},$$

and precisely what Gauss's law would give; but if the law were rigorous the formula should be *exact.*

The hypothesis that the arithmetical mean of several quantities is the most probable value leads to inconsistencies. It requires, for example, that the most probable value of the square of the quantity shall be the arithmetical mean of the squares. Nor can the objection be avoided by making a distinction between measures directly observed and those resulting from calculation. A mechanic could easily attach to a balance a needle to indicate the square or the logarithm of a weight.

The same objection applies to expressing the probability of an error as a function of the error alone.

The problem is proposed, "If Gauss had adopted, instead of the mean, another mode of combination of the measures, what law of errors would he have deduced ?"

The problem is not solved, but it is shown that if

$$f(x_1, x_2, \ldots, x_n)$$

is the most probable value of a quantity of which $x_1, x_2, \ldots, x_n$ are measurements, then, in order that the probability of an error shall be a function of the error, f must be the arithmetical mean of the x's increased by some function of their differences. In other words, it must be such that if all the x's are increased each by α, it shall also be increased by α.

In spite, however, of all theoretical objections to Gauss's law, constantly accumulating experience completely justifies its adoption. As to the arithmetical mean, Ferrero has shown (see Charles S. Peirce's review in *Am. Journ. Math.* I., 59) that all functions of the measurements that it would not be absurd to take for the most probable value of the quantity measured, will, if the measurements are good, agree in their results; while, of course, if the measurements are bad, no treatment can be expected to give good results.

It is gratifying to find these careful definitions of precision and weight.

"The precision of one measure is said to be α times that of another measure when the probability that an error is contained between z and $z + dz$ for the one is the same as that it shall be contained between αz and $\alpha\ (z + dz)$ for the other."

"The weight of one observation is said to be β times that of another, when the consequences that can be deduced as to the value of a magnitude measured by an observation in the first system are equivalent to those that can be deduced from β observations in the second system, all giving the same result."

"If β is a fraction m/n, it is necessary that m concordant observations of the first system can be replaced by n concordant observations of the second."

"The system of observations that gives to the error z the probability

$$ke^{-k^2z^2}/\sqrt{\pi}$$

has k for its precision and k^2 for its weight, if we take for units of precision and weight those of one observation in the system for which the probability of an error z is proportional to e^{-z^2}."

M. Bertrand argues at some length for the rejection of doubtful observations. He gives no criterion, however, as Peirce has done, to determine when they shall be rejected. This is left to the judgment of the computer, with the caution that the number of retained observations must be large. The probable value of the square of an error smaller than λ, when those larger than λ have been rejected, is

$$\frac{1}{2nk^2} \cdot \frac{\Theta\,(k\lambda) - 2k\lambda e^{-k^2\lambda^2}/\sqrt{\pi}}{[\Theta\,(k\lambda)]^2}.$$

As for Gauss's attempt, in his last memoirs, to break away from all hypotheses of a law of errors in establishing the method of least squares, it is shown that neglecting the squares and powers of the errors is equivalent to assuming the exponential law.

The equating of the probable value of a function to its true value is not unobjectionable. The following example shows this.

Five angles l_1, l_2, l_3, l_4, l_5, have been measured. The geometrical conditions require

$$l_4 + l_1 - l_3 = 0,$$

$$l_5 + l_2 - l_3 = 0.$$

It is found, however, that

$$l_4 + l_1 - l_3 = h_1,$$

$$l_5 + l_2 - l_3 = h_2.$$

Designating the errors really committed by e_1, e_2, e_3, e_4, e_5, we have:

$$e_4 + e_1 - e_3 = h_1,$$

$$e_5 + e_2 - e_3 = h_2.$$

No matter what the multipliers λ_1, λ_2, λ_3 may be, the trinomial

$$\lambda_1 h_1^2 + \lambda_2 h_2^2 + \lambda_3 h_1 h_2$$

is known.

This trinomial is a homogeneous function of the second degree in the errors; and calling m^2 the probable value of the square of one of the errors, that of the product of two of them being nothing, we shall find for the probable value of the trinomial, calculated before the measures are taken,

$$m^2(3\lambda_1 + 3\lambda_2 + \lambda_3).$$

Equating this to the true value gives

$$m^2 = (\lambda_1 h_1^2 + \lambda_2 h_2^2 + \lambda_3 h_1 h_2)/(3\lambda_1 + 3\lambda_2 + \lambda_3),$$

an infinite number of different values for m^2.

This does not furnish, however, the most serious reason why the chances of error cannot be precisely evaluated.

"It is supposed, *a priori*, that all the measurements of a system are equally precise; it is impossible in the vast majority of cases to believe in such equality: it is from lack of knowing reasons for preference that the results are accepted as equivalent. But, known or unknown, these reasons, if they exist, must have an influence upon the error really committed and of which it is pretended to give the probability."

"After having, with immense labor, discussed the transit of Venus observations of 1761, Encke found for the parallax of the sun 8″.49 with a probable error 0″.06. He could therefore bet 300000 to one that the error would not reach 0″.42. Nevertheless, astronomers have just accepted the parallax 8″.91 corresponding exactly to the error 0″.42."

"We can simply affirm, and this is the important point, that if the sum of the squares of the corrections are small, there is great likelihood that the observations have been well made."

The extension of Gauss's law to errors in the situation of a point gives for the most probable position of a point the centre of gravity of the observed positions supposed equally weighted.

Restricting ourselves to a variation in two coordinates, "the probability that an error shall be comprised between u and $u+du$ for x and between v and $v+dv$ for y is

$$Ge^{-k^2u^2 - 2\lambda uv - k'^2v^2}\,dudv.$$

Points of equal probability are upon the same ellipse having for its equation

$$k^2u^2 + 2\lambda uv + k'^2v^2 = H,$$

u and v designating the differences between the coordinates of the point considered and the true position, the common centre of all the similar ellipses whose dimensions are proportional to $\sqrt{H}$.

A comparison is made between the theory and the results of firing 1000 shots at a target.

The law has been partially guessed by Galton in his *Discussion on the Data of Stature*, and more fully worked out by Mr. Hamilton Dickson. (See *Natural Inheritance*, p. 100 *et seq.*)

It seems a pity that in the chapter on the laws of statistics some slight reference at least should not have been made to Galton's investigations.

A point liable to be overlooked in applying the laws of probability to statistics is well stated.

"There are plenty of ways of consulting chance that will give the same mean without giving the same probabilities of error. Instead of drawing balls from an urn of a given composition, we could draw in order from many urns of various compositions. The average results would be the same as for drawings from an urn of average composition, the chances of error would not be."

"If, to take an extreme case, instead of drawing 10,000 times from an urn containing one white and one black ball, we draw alternately from two urns, one containing a white the other a black ball, we shall certainly get white 5000 times, the error will be nothing."

To represent tables of mortality, the substitution of several urns for a single one, seems, *a priori*, very plausible. Among individuals of the same age it is impossible not to find classes in which the chances of life are unequal."

The book ends very pleasantly with a chapter on the misapplications of the theory of probabilities to judicial decisions.

Apropos of this matter, does not the great probability that attaches to the results of concurrent independent judgments furnish the strongest possible argument for cultivating independence, for ridding ourselves of the systematic errors imposed by education and fashion, by party and sect?

I have very inadequately sketched this most admirable introduction to the science of probability: the life and vigor of the original cannot be reproduced in a brief review.

ELLERY W. DAVIS.

COLUMBIA, *August* 14, 1891.