GALTON'S STATISTICAL METHODS.

Natural Inheritance. By Francis Galton, F.R.S. London and New York. Macmillan & Co. 1889. Pp. ix, 259.

This work is of double interest. Its primary purpose is biological, being to subject the question of heredity to accurate quantitative and mechanical treatment. As such it is doubtless the ablest work on the subject extant. But in the course of his investigation Galton has collected a large mass of statistical information, and, what is more important, has developed some new and interesting statistical methods. The key-note to the statistical side of the work is contained in Galton's statement that statisticians are apt to be content with averages, while an average is only an isolated fact. What is wanted is a method of *calculating distribution*, and a graphic scheme for reading the distribution. For example, compared with the knowledge of the average income of an English family, a knowledge of how the total income of England was distributed would be much more important. This would tell us the proportion which had incomes of every grade from the lowest to the highest, and would enable us to rank any given family at its place in the scale.

Particularly, in dealing with problems of heredity, is a scheme of distribution necessary. The knowledge of the average stature of a kinsfolk conveys little; the knowledge of how this faculty is distributed among the members of the kindred would be valuable. Galton's first work was to invent a scheme of distribution. The data to be dealt with, for example, are the strength of pull of 519 males as registered on a Salter's machine. The following are the figures: —

STRENGTH OF PULL.	PERCENTAGE.
Under 50 pounds, ...	2
" 60 "	10
" 70 "	37
" 80 "	70
" 90 "	91
" 100 "	95
Over 100 "	100

This might be illustrated by a diagram: —

The percentages of strength are marked off on a base line, the number of pounds on the right-hand perpendicular line. Then from each per cent, 2, 10, 37, etc., is erected a perpendicular to a height equal to the corresponding number of pounds,— *i. e.*, from 37 per cent would be erected a perpendicular to a height equal to 80 pounds, since 37 per cent pulled less than 80 pounds. While a line connecting these various perpendiculars will be broken, it is evident that, if the data were numerous enough, and the strengths more closely taken, say to every pound, we would get, approximately, a curved line. The figure, bounded by a curve of this kind is a scheme of distribution. (The perpendiculars, since they serve only for scaffolding, would not appear in an ordinary scheme.) By taking the measured strength of any individual on the side scale, say 74 pounds, carrying over a horizontal line until it meets the curve, and then dropping a perpendicular to meet the base line, the proportionate rank of the individual may be read off,— in this case 50 degrees. In other words, since 50 per cent exceed and 50 per cent fall short of his strength, he occupies a medium position, and his strength is mediocre. This position Galton always designates by M, and this M is always one of the chief constants in Galton's scheme. He notes that the M has three properties. The chance is an equal one that any previously unknown rank falls short of or exceeds M. The most probable value of any previously unknown measure is M; and, if the curve of the scheme is bilaterally symmetrical as respects M, M is identical with the ordinary average or arithmetical mean. It is evident that we have the start for an application of the theory of probability. Now, if the deviation of any grade from M is considered, that is. the error as respects the mean, we find that every measure in a scheme may be expressed by $M+(\pm D)$, the $+$ or — signifying up or down from M.

Galton's other constant he designates by Q. It is obtained as follows. Take the perpendicular at 75°, and that at 25°. Subtract the latter from the former and divide by 2, and we get a measure of the general *slope* of the curve of distribution, just as M measures the average height of the curved boundary. What it *really* gives is the deviation from the average, both in excess and in deficiency, of one half the number taken. For example, the Q of the scheme of distribution of stature is 1.7 inches. This means that one half the population differs less than 1.7 inches one way or the other from the average of the whole population.

Now, although this Q stands on its own independent basis, and can be derived from any scheme by dividing the differences of the ranks of 25° and 75° by 2, if the curve is symmetrical, it will be identical with what the mathematicians call the Probable Error. It thus becomes possible to apply the whole calculus of probability to any data capable of being expressed in a normal case of the scheme.

As matter of fact, Galton found a remarkable parallelism between results obtained by observation and those theoretically deduced from the mathematical calculations. He took, for example, eighteen schemes of distribution, including stature, weight, breathing capacity, strength of blow, keenness of sight, for both sexes, calculated the Q in each, or the measure of deviation that half the deviations exceeded, and half fell short of, and then calculated a mean Q, as it were, for the whole eighteen. The result differs but little from that theoretically obtained by the law of frequency of error. It is evident that Galton might express his scheme in the well-known curve of error, but his curve of distribution contains all that the curve of error contains, and much besides.

For the particular results obtained, I must refer to the book itself. One is so remarkable that it may be specified. If we call the M of the stature of the whole population P, and the mean stature of the parents $P \pm D$, the stature of the offspring will be, on the average, $P \pm \frac{1}{3}D$. In other words, upon the average, children of parents who are exceptional, or who deviate from the mean, will themselves deviate from the mean only one third of their parents' deviation. Considering the character of his results, it is not wonderful that Galton says: "I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'law of the frequency of error.' The law would have been personified by the Greeks and deified if they had known of it. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand, and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along."

It is to be hoped that statisticians working in other fields, as the industrial and monetary, will acquaint themselves with Galton's development of new methods, and see how far they can be applied in their own fields. It is, of course, clear that any data dealing with the proportions of distribution of anything whatever can be diagrammatically expressed in Galton's scheme; but if it is to be anything more than a picture for the eye, it must be possible to establish an M and a Q from which the entire scheme of deviations and their relations may be, in turn, deduced, at least approximately. In other words, such a scheme, if its curve were wholly irregular, would not be likely to yield any results. Furthermore, no curve is likely to be regular unless it expresses traits which are the result of accidents, that is, of circumstances which *do* bring about certain results, though they were not intended for that purpose. For example, if we had (what we are not likely to have) accurate data regarding the accumulation of wealth in families from parents to children, there is no great reason for expecting that in two generations we would get results akin to those of Galton regarding natural heredity. The tendency of wealth to breed wealth, as illustrated by any interest table, and the tendency of extreme poverty to induce conditions which plunge children still deeper into poverty, would probably prevent the operation of the law of regression toward mediocrity. It is not likely that children of the poor would be better off, and children of the wealthier poorer in anything like the ratio of $\frac{2}{3}$. But if we took generations enough, the operation of "accidents," such as imprudent, extravagant, and dissipated habits among the children of the rich, the growth of new industrial conditions which lessen the value of old forms of wealth, the emergence of money-makers among the poor, the development of social relations which would increase the ambition and chances of the poor, etc., we might find a similar law. That is, these accidents, or circumstances which, although in themselves irrelevant to the distribution of wealth, yet in the long run, largely determine it, would pelt down as it were the swells in the curve, and bulge out its depressions into something like a normal curve of distribution. Whether or not there is any truth in our example, it will serve to illustrate the nature of the data to which Galton's methods may be applied.

JOHN DEWEY.

UNIVERSITY OF MICHIGAN.