

LETTERS TO THE EDITOR.

[The Editor does not hold himself responsible for opinions expressed by his correspondents. Neither can he undertake to return, or to correspond with the writers of, rejected manuscripts intended for this or any other part of NATURE. No notice is taken of anonymous communications.]

A New Step in Statistical Science.

CONTRIBUTIONS of an abstract kind to statistical science are so little read by the bulk of statisticians, that the scope of the remarkable memoir by Prof. Karl Pearson may be unappreciated by them unless attention is pointedly directed to it.¹

Statisticians are conversant with the use of curves to epitomise masses of data. The forms of the majority of these curves are skewed or humped, and have hitherto been nondescript, except as Prof. K. Pearson's previous memoir showed, some of them may be dissected into two or more normal curves, having different constants. It is only a few curves that are symmetrical and conform closely to the normal law of facility of error. These few have been much studied, the numerous and valuable properties of the normal curve being extremely helpful in dealing with them. When the conformity between the observations and the normal law ceases to be close, the latter must be applied warily. When the discrepancy is serious, it becomes unsafe to trust the theory at all. Not a few statisticians have chafed under these limitations, and felt that the normal law is too limited in its grasp to satisfy their needs. Now at length, it turns out, thanks to Prof. K. Pearson, that the normal law admits of being regarded as nothing more than a very special case of a highly generalised theory, by whose aid curves may be drawn that shall fit every one of the observed curves he has tried as yet. The shapes of these are curiously varied. Their list includes the dimensions of shrimps and prawns, of American recruits and of school-girls, of Bavarian skulls, of frequency of enteric fever, and of divorces, of variation in house value, in butter-cups and in clover, in pauper percentages and in a mortality curve. Only those who have studied the delicate and oddly-shaped drawings of these observed curves on a large scale, upon which the equally delicate *fitted* curves have been superimposed, can adequately appreciate the wonderful closeness between the pairs of outlines. There can be no doubt that the descriptive efficiency of Prof. Karl Pearson's method is of the highest order.

The question of the utility of the method to ordinary statisticians has now to be considered. First, as to its descriptive powers. We are already able to describe the whole of any normal series by means of only two numbers; say, by the average of all the members of the series, and by the mean of the several departures of its individual members from that average. Henceforth, by the use of more constants, we shall be able to do the same to any series. A few figures will always serve as the equivalents of a vast amount of tabular matter.

The second and higher use is to afford a clue to the cause or causes of variety. It has long been a dream to me to select a peculiar and often recurring form of curve, and to study with all possible care the conditions under which it has been produced, so that whenever a new curve of that same form was met with, there should exist some guidance towards discovering the cause of its production. I have made not a few attempts from time to time, but was discouraged by the then impossibility of sufficiently defining the curves that were dealt with. That difficulty seems now removed. To explain myself further, let us suppose that a man finds the mark of a more or less incomplete circle on the ground, and wishes to discover how it was made. Various possibilities exist, which might have been recorded to help him: (1) The mark may have been made by a basin, &c., turned in a lathe, or, what come to the same thing in principle, by something revolving round a fixed centre. (2) It may have been stamped by a hoop, formed by bending an elastic rod until its ends met, the circular form depending on the equal distribution of stress. (3) It may have been the mark of a withy that had bound a faggot, which, after setting into shape, had broken away, the circular form arising from the compression of the sticks of the faggot within the smallest possible girth. (4) It may have been

the mark of a warped sheet of bark, hide, &c., the circular form of which depended on the unequal contraction of its outer and inner layers. (5) It could have been made by a projecting nail, near the angle formed by two straight rods securely joined at one end; the apparatus being caused to straddle and press upon two pegs in the ground, and moved in the only way possible under those conditions. Here the fragment of a circle would be due to the constancy of the angle subtended by the same chord. A catalogue of these and other possibilities, which are numerous and include circular forms of animal and vegetable growth, would certainly enlarge the speculations of the observer as to the cause of the circular mark. So it would be with the curves of which I spoke. Each form of curve would be a serious study in itself; still the results would gradually accumulate, and it is reasonable to look forward to a time when a set of such curves, each defined by Prof. K. Pearson's method, shall have been studied.

I venture upon one criticism as to the completeness of his generalisation. Variation of the normal kind is supposed to be due to the combined effect of (1) an infinite number of causes, that are (2) equally likely to err in excess or deficiency, and (3) that are independent of one another. These three restrictions are removed in the generalised curve, which bears a certain relation to the binomial point curve formed by the expansion of $(p + q)^n$ when (1) n need not be infinite, (2) p need not be equal to q , while (3) the binomial form which implies independence of the contributory causes, is modified. Now though the condition (3) is removed, it does not, as yet seem to me that the supposition which replaces it, and which is based on such considerations as the effect of withdrawing r cards from a pack of ns cards containing s suits, is analogous to what commonly occurs in *rerum naturæ*. Namely, the intermingling of contributory causes of various degrees of efficiency, some of which are very few in number and have large effects. Thus the number of persons who walk day by day down St. James' Street, is occasionally vastly augmented by some national spectacle, and it is considerably and irregularly affected by changes in the weather. It does not wholly depend on a multitude of equipotent causes. So again, the time of ripening of the fruit on a tree generally, is much affected by the aspects of the particular branches on which it grows. I therefore conclude that the effects of an aggregate of large and small causes, ought to be distinctly included in a thoroughly generalised formula of variation.

FRANCIS GALTON.

The Kinetic Theory of Gases.

I THINK sufficient stress has not been laid on the distinction between the *purely mathematical* proof of the fundamental determinantal relation connecting the differentials of the co-ordinates and momenta of dynamical systems and the *purely statistical* applications of that relation which form the subject of the Kinetic Theory.

The well-known determinantal equation is perfectly general and applicable to any dynamical system whatever. It merely asserts that if the initial coordinates and momenta receive any small independent variations whatever, the resulting variations in the final coordinates and momenta after any fixed interval of time are so related to them that the differential product of the former variations is equal to that of the latter, conformably to the well-known laws of Jacobians, by which the variables are changed in a multiple differential or integral.

In general the variations in question are *purely hypothetical*, just as, in the principle of Least Action, the actual motion is compared with varied motions which have no real existence. If, for example, we consider the system to be the Earth, the variations could only be made real by making the Earth move differently to what it actually moves, and doing this in every possible way.

But when it is required to assign any physical meaning to the determinantal relation, these hypothetically varied motions must be represented by actual motions, and this can only be done by taking an indefinitely large number of independent systems, all similar to the one we started with, and setting them moving in all possible ways. The determinantal relation then tells us that if the coordinates and momenta of these systems are initially distributed according to what in my Report I have called

¹ "Mathematical Contributions to the Theory of Evolution" (Part ii.), by Karl Pearson. Read at the Royal Society, January 24, 1895.