Vol. LXXVI.] [Part II.

JOURNAL

OF THE ROYAL STATISTICAL SOCIETY.

JANUARY, 1913.

On the USE *of the* THEORY *of* PROBABILITIES *in* STATISTICS RELATING *to* SOCIETY.

The PRESIDENTIAL ADDRESS *of* PROFESSOR F. Y. EDGEWORTH, F.B.A. DELIVERED *before the* ROYAL STATISTICAL SOCIETY, DECEMBER 17 1912.

MOST of those of whom I am the unworthy successor have signalised their entrance on the office of President by applying statistics to some problem of conspicuous practical interest. It was thus that Giffen, in 1883, exhibited the progress of the working classes during the preceding half century. So Caird discoursed on the condition of agriculture, Goschen on the distribution of incomes, Fowler on municipal finance. And, not to multiply distinguished instances, my immediate predecessor brought to bear on the problems of pauperism the statistics which he had assisted in compiling as a member of the Royal Commission on the Poor Law. All these, with a practical sagacity which I cannot hope to imitate, have fulfilled the purpose of our Society as laid down by its founders: the ascertaining and bringing together of those "facts which are "calculated to illustrate the conditions and prospects of Society." They also, I think, have served that purpose who, with Newmarch, have been content to prospect and report upon methods of statistical science. These are the only precedents which I can attempt to follow. I follow at a great distance. For whereas Newmarch, in his inaugural address on the progress and present condition of statistical enquiry (1869), ranges over *eighteen* "fields of statistical research "which in this country must require early attention," I confine myself to this eighteenth and last division: "the investigation of "the mathematics and logic of statistical evidence; that is to say "the true construction and use of averages, the deduction of "probabilities." The purport of the enquiry which Newmarch thus prescribes is more fully brought out by Dr. Guy, when referring in his inaugural address, a few years later, to "the principles of "the numerical method of the logic of large numbers," he complains of "the hitherto imperfect treatment of the principle involved in "the known reproduction year by year of like figures." No serious attempt, he says, has been made to place this matter before us in its true light and in all its fulness. The reproach made by Dr. Guy has at length in our day been removed by some of the more recent contributions to the *Journal*, in particular those of Mr. Bowley and Mr. Yule. But there is perhaps still room for some additional observations *in pari materiâ.*

Before entering on this abstract enquiry there are some concrete statistics relating to the history of our Society during the past session which it devolves upon me as President to communicate. The death rate among our Honorary Fellows has, I regret to report, been considerable. The loss of Levasseur alone would be great and irreparable. Since June of last year the Society has also lost Dr. Emil Blenck and Dr. Enrico Raseri. Among other losses by death is included that of two past Presidents of the Society, the Earl of Onslow, President in 1905–06, and Sir Francis S. Powell, President in 1904–05. By the death of Mr. A. H. Bailey the Council have lost one of their oldest and most respected colleagues.

Turning from losses to gains, it is satisfactory to be able to record a steady increase in the number of Fellows since 1909. The number of Fellows now on the Society's books is greater than in any year since 1906; but I think I should emphasize the opinion already expressed by the Council that in view of the widely extended interest in statistics the membership of the Society is still much below the number that might reasonably be expected. The extension of the Society's usefulness can, and should, be fostered by the general body of Fellows; and it rests with them, as well as with the Council, to ensure that the advantages which the Society now offers to those interested in statistical work are more widely known and appreciated.

It is pleasant to report that there is no falling off in the quality of our work. The Session has been a most successful one, not only as regards the character of the papers read, but also as regards the interest and value of our discussions.

In connection with the work of the Session we may look back with particular satisfaction upon the work done by the Special Committee on Infantile Mortality which came under discussion at the last Ordinary Meeting. Acknowledgment has already been made of the indebtedness of the Council to Dr. Dudfield and his colleagues, but I am glad to avail myself of this opportunity of expressing my personal appreciation of a statistical task so ably and conscientiously performed.

I cannot conclude this brief retrospect without voicing what is, I believe, the general feeling that our Society has been singularly fortunate during the past two Sessions in having possessed in Lord George Hamilton a President whose association with our activities has been a constant example and stimulus. We recall with satisfaction and gratitude the valuable services which he rendered to the Society not only at its Ordinary Meetings, but also on the Council and Committees.

I now return to the consideration of the subject which has been announced.

1. The use of Probabilities in Statistics relating to human affairs may properly be introduced by the use of the theory in physics. For as there is not one sort of arithmetic for social and another for physical phenomena, so the principle of Probabilities is essentially the same in these two regions. That principle may best be discerned and recommended by considering its manifestations in the more abstract and better accredited department of science. It was not without reason that Quetelet writing to Farr as to the conduct of the fourth International Congress "insists strongly that in this "country the scientific element of statistics should be developed in "all the sections so as to maintain its early alliance with the strict "sciences through the Calculus of Probabilities." The connexion between Probabilities and the strict sciences has become much closer since the days of Farr and Quetelet. Throughout the extensive part of Nature in which matter assumes the form of a gas, physical laws are presented as the average result effected by many miniature masses rushing hither and thither at random. Thus the pressure of a compressed gas, air for instance, against the interior surface of the vessel by which it is confined is accounted for by the force of the blows delivered by the molecules of the gas as they dash against the resisting barrier. The mixture of two gases which occurs when a partition between them is removed is likewise explained as a consequence of random movements and collisions. In the authoritative words of Maxwell "the constancy and uniformity of "the properties of the gaseous medium is the direct result of the "inconceivable irregularity of the motion of agitation of its "molecules." Those of the ancient philosophers who sought the origin of things in an atomic chaos were not so mistaken as they

appeared to the defenders of common sense. A large part is really played in nature by the random rush of clashing atoms—

"Innumerabilibus plagis vexata per aevum."

Nor are we concerned only with *atoms* as understood by the modern chemist, the comparatively large bodies which by their combination form the still larger *molecules* of air and water and less familiar substances. Also the thousands of much smaller particles or *corpuscles* into which each of those atoms is presumably removable are amenable to statistical treatment. The random flight of so-called "alpha" particles from "radio-active" substances presents some analogy to the departures from this life occurring in a population subject to a constant mortality.

In short, Statistics reigns and revels in the very heart of physics. "Probability-uniformities" are placed by Dr. Venn, in his masterly treatise on the modern material logic, as only one among six kinds of uniformity. But the most recent science suggests that this one is destined to swallow up many of the others: that sequences which now pose as laws of nature, co-existences which now seem ultimate facts, may one day prove to be the average outcome of movements in the invisible world of atoms and corpuscles.

Nor is it to be supposed, because the number of the constituents which go to an average in this kind of statistics is often exceeding great, that, therefore, the "average regularity" differs not only in degree, but in kind, from that which the statistical practitioner experiences. It is true that the multitudinous thud of trillions of molecules against the sides of an enclosing vessel presents itself to the senses as a single simple fact—the phenomenon of pressure. But theory suggests that if we could have a microscope sufficiently powerful to observe the motions in more detail the irregularity characteristic of statistics would appear. Experiment confirms this suggestion. M. Perrin has arranged granules (of "mastic" or other suitable substance) such and so small that floating in a liquid they are not uniformly affected by the impulse of impinging molecules. Struck in different places with different degrees of force they are driven hither and thither with varying velocities. While small enough to be thus affected, the granules are large enough to be seen through a microscope thus behaving like gigantic molecules. M. Perrin happily illustrates the action of invisible molecules on visible granules by the tossing of a distant ship, due to ripples not visible at a distance.

Moreover, in the interior world of corpuscles it has proved possible to observe and register the number of "alpha" particles which are discharged from a radio-active substance per minute or smaller unit of time The records made by Dr. Geiger show a variation in the number of exits from moment to moment comparable with the variation from year to year in the number of deaths occurring in a uniform population, such as males of the same age in the same occupation. These physical statistics have quite a human character.

Thus the main characteristic of statistical probability, collective constancy combined with individual irregularity—observations hovering about a mean towards which they converge—is conspicuously fulfilled in this domain of physics. Here, too, we may encounter the difficulties which beset the relation of an average to its constituents. There is first a question akin to the controversies about Free-will, which have exercised even writers on statistics. How can we reconcile the treatment of the individual constituents as fortuitous while it is believed that every particular event obeys the law of causation; as in the case of molecular motion, every single movement is determined by the strict rules of mathematical physics? Again, there is the antinomy which Buckle emphasized between the apparent freedom of the individual and the collective constancy of statistics. The regularity of the total is not simply explicable by the parts obeying rule. The truths of pure statistics have not the character of the simple propositions adduced in ordinary logic. A distinguished writer on the logic of Probabilities, Von Kries, seems even to deny that the truths which form our science are obtainable empirically. I will not follow him in attempting to describe the indescribable. I rather acquiesce in the dictum of a great mathematical writer on our subject, Poincaré: "There is here something mysterious inaccessible to the "mathematician."

2. The prestige of physical science attaches not only to the fundamental principle of Probability, but also to the *law* or higher theory which is built thereon. The law which I proceed to consider is comparable in respect of quantitative precision with the laws of physics, for instance that of gravitation. As that law informs us that the distance through which a body falls (in vacuo, starting from rest) increases proportionally to the *square* of the time it takes to fall, or, in other words, that the time increases proportionally to the *square root* of the distance, so our law, the law of normal frequency as it may be called, informs us that the precision of an average increases proportionately to the *square root* of the number of (independent) observations averaged. But there is a marked difference, though perhaps at bottom only one of degree, between the physical law and the law of chance. The latter retains the character of Probability, and is true only on an average. If we exhibit, say, by the position of a "hand" on a dial, like that employed in weighing machines, the relation between the (square

root of the) distance and the time for several careful experiments on falling bodies, we may expect that the index will remain constantly at the same point of the dial, or at most will show only slight tremors due to imperfections of observation. But the tremors which are incidental in ordinary physics are essential in Probabilities. We can only hope for an index hovering about a point.

The allusion to errors in physical observations is more than incidental to our present purpose. Not without reason has the law which I am proceeding to consider been designated the *law of error.* For errors of observation present an important, though only a particular, example of the law with which we are now concerned. And, indeed, the theory of errors of observation is one of the principal lessons which the statistician may obtain from the practice of physicists. But as I have dwelt at length on that lesson in former numbers of our *Journal,* I now confine myself to the illustrations which are afforded by molecular physics.

It is a circumstance of momentous interest to the mathematical statistician that the law which constitutes his main implement for dealing with statistics of the visible world is accredited by its complete fulfilment in the world beyond the senses revealed by the new physics. Consider, for example, the millions of trillions of air-molecules which within this hall are rushing hither and thither in every direction and at various rates of velocity. As the railway expert may classify the trains according to their speed—express or ordinary or slow—and state the numbers of each class on any line, so by the normal theory of frequency the molecules are divided into different classes according to their velocities and the proportionate numbers of each class assigned. The commonest or most typical class is that which is characterised by a velocity of about a quarter of a mile per second. The frequency or probability of a molecule being at rest is negligible. Only one air-molecule in two millions, if I rightly calculate, has a velocity of a mile per second. And so on for different rates of velocity.

Lucretius, if he had been imbued with the new atomic philosophy, would have rejoiced to see an example so splendid of order emerging from promiscuous collisions in a molecular chaos. But he would have found the mysteries of modern science even more difficult than the recondite doctrines of Greek philosophers—"Graiorum "obscura reperta"—to elucidate in Latin verse. Even into English prose it is not easy to translate the mathematical theory from its mother-tongue of symbol. Its general features need only be recalled here. It will be remembered that the normal law of frequency in its simplest form is represented by a symmetrical curve shaped much like a bow when strung. If along the string placed horizontally there are measured from the central point (numbers of) intervals or degrees corresponding to different extents of deviation from an average, the corresponding distance at each degree of the string from the bow (the ordinate, or small strip of area, corresponding to each distance on the horizontal axis) represents the frequency with which each particular degree of deviation is apt to occur. These frequencies diminish rapidly as you move from the maximum at the centre, and die away ultimately into nothing. The metaphor of mortality is indeed appropriate in that a group of population at the central age of life, say the number of persons each aged 50 years, would in successive years diminish at an increasing rate of mortality per cent. But the analogy is imperfect in this respect that the increase in the rate of human mortality is more than proportionate to the increase of age above the central epoch. Whereas the characteristic of our curve is that the diminution (per cent.) of the frequency (represented by the ordinate of the curve) is exactly proportional to the distance from the centre.

To exhibit the fulfilment of this law in a molecular chaos let it be allowed in framing a model to make some alteration of the dimensions. First let us leave out altogether the *third* dimension and consider only movements of molecules on a plane. Also let us increase their dimensions some hundred million-fold so that the magnified bodies may be about the size of a billiard ball. Let them, also, for convenience of enunciation, be all of the same shape; all equal and perfectly elastic balls. We are then to conceive trillions of such balls rushing hither and thither with repeated collisions over a perfectly smooth billiard-table of immense extent with perfectly elastic cushions. Under such conditions the velocities of the billiard balls will be distributed according to the normal law of frequency.

The fulfilment of the law may be discerned most clearly by considering the velocity of a ball moving in any direction, say south-west to north-east, as resolvable into two component velocities, one in the direction south to north, the other in the direction west to east. Thus, if the sides of our (rectangular) billiard-table lie east to west and south to north (parallel to lines of latitude and longitude) the velocity of any ball moving in any direction, that is the distance that it would move if unimpeded in a second (or other unit of time), may be supposed to be *projected* on the eastern side of the table. The projections might be conceived as shadows thrown by the horizontal western sun on a properly arranged screen along the eastern side of the table. It is these shadows which must directly and obviously fulfil our law. For consider the set of balls starting at the same instant from a set of points along a horizontal line (of any length anywhere in the table). And let us suppose—a very violent supposition, certainly—that these balls can move for a second without clashing against each other, or against some other balls. Or, at least, let it be possible to represent on the screen a corresponding set of shadows moving, say, northward, from a certain starting point. The velocities of these shadows will be distributed according to the normal law. If each individual stops dead at the point which it has reached at the end of a second from the time of starting, the heap thus presented will be shaped like our normal bow. Corresponding to the highest point is the case of most frequent occurrence, namely, that of no motion, stopping still at the starting-point. Similar statements are true of the velocities resolved in a direction perpendicular to that which has been considered, say on a screen placed along the northern side of the table.

This description may seem inconsistent with the statement above made that the commonest or typical velocity of molecules (of air, in a room of ordinary temperature) was about a quarter of a mile per second. But the inconsistency is only seeming; the reference being there to actual velocities in any direction, here to velocities resolved in a particular direction. If a fountain or garden-hose playing in every direction flings water to different distances (in the same time we should have to add to make the illustration perfect) over a smooth lawn, there is no inconsistency in saying that the portion of the lawn nearest the jet receives the greatest number of drops per square inch; but that of the rings which are formed by describing circles at equal distances from each other with the origin as centre, the one which receives the most drops is at a certain distance from the centre. The square inches contained in the rings nearer the centre receive less, for much the same reason that black sheep eat less than white ones—because there are fewer of them. That proposition so interpreted would remain true up to a point even if black sheep had in fact larger appetites.

Why, it may now be asked, is this beautiful theory fulfilled by clashing molecules? The statistician naturally seeks an explanation in the causes with which he is familiar. These causes have often been described in our *Journal*, and need only be mentioned summarily now. The condition that a set of magnitudes should obey the normal law is, in brief, that each should be a simple combination of numerous independently fluctuating elements. By a simple combination is understood especially a sum or a simple average (arithmetical mean), or a weighted average, or a "weighted "sum," as we may call the *numerator* of a weighted average. For example, the sum of 25 digits taken at random and each multiplied by ·2 is a weighted sum, which, as shown in the Appendix, fairly well fulfils the normal law.

The propriety of the particular weighting just instanced is that it keeps the dispersion or "spread" of the compound magnitudes just the same as what it is when the number of the components is different and the weight correspondingly altered. Thus the sums of 16 digits each divided by 4, as shown in the Appendix, present the same dispersion as the sums of 25 digits each divided by 5. Generally, if it is required to put together two statistical quantities each having the same dispersion in such wise that the dispersion of the compound may be the same as that of each of the components, it is proper to form a weighted sum of the two components with factors, such that the sum of their squares is equal to unity. Any number of such factors can be obtained from well-known (trigonometrical) tables; for instance, ·5 and ·86602 . .

This principle may be employed to illustrate the distribution of velocities in a molecular chaos by a fanciful distribution of property as follows:—Imagine a community numbering a million, each starting in life with a portion of gold and likewise one of silver, assigned according to some random principle of distribution. This initial distribution is then transformed by a series of—business or gambling—transactions of a peculiar type. Suppose citizen A, having initially a of gold and α of silver deals with citizen B, who has b of gold and β of silver. The result of the transaction will be that A will have a portion of gold which is a "weighted sum" of the four quantities, a, α, b, β, the weights being constructed, in a manner more particularly described in the Appendix, by the use of a pair of factors, such as the ·5 and ·866 . . above instanced. Similar transactions are effected by other pairs of citizens. Similarly A, after dealing with B, deals with some other citizen, say M; whether M retains his initial portions, or has already had dealings, say, with N. So brisk are the dealings that soon none will retain their original portions. After thousands of transactions each one's portion of gold and likewise of silver will be a weighted sum of some thousand fortuitously distributed elements, the weighting being such as to preserve the dispersion constant.

The interpretation of this parable is not far to seek. Each pair of letters, one Roman and one Greek, stands for the two velocities—in the directions south to north and west to east—into which the velocity of a particular molecule, as represented by one of our billiard balls, may be resolved. The velocities of a second ball, with which this first collides, are similarly represented by b and β. The result of the collision will depend not only on these velocities but also on the manner of their encounter, whether they meet full tilt or only graze each other. The velocities being assigned, this datum will be obtained when the direction of the line joining the centres is known. Suppose, for instance, that line to make an angle of 30° with the horizontal line from east to west; then the factors just now mentioned, namely, ·5 and ·866 . . would be appropriate. There will, of course, occur every variety of (possible) factors in the course of indefinitely numerous collisions.

But it should be remarked that these factors do not constitute so many independent elements; since, given the initial velocities, all the subsequent factors (or inclinations of lines joining centres at moments of contact) are theoretically given implicitly. If we start with a trillion molecules, we have *only* two (or in three dimensions, three) trillion *independent* initial velocities. The velocities after a quadrillion collisions would not *perfectly* fulfil the conditions of the normal law. There would be a certain *interdependence* in the contributory causes. There would, strictly speaking, be fulfilled not so much the normal law of frequency, as the more comprehensive "law of great numbers" which has been described in our *Journal*. But, I think, it would be a long time before the difference would be noticed. Nor would the fact, undoubtedly inconsistent with the *perfect* normal law, that no one molecule ever takes on more than the entire energy of the system (!) force itself practically into notice. With these slight reservations, it is deducible from the theory of Probabilities that the velocities in a molecular chaos will be distributed according to the normal law of frequency.

The simplicity of this deduction contrasts somewhat suspiciously with the stupendous demonstrations offered in the treatises on the Kinetic Theory of Gases. Nor would I claim for Probabilities more than the position of a buttress to a construction which must be rested mainly on a mechanical foundation. Indeed, when we consider the colossal substructures employed by the mathematical physicists, the thought occurs that the buttress itself may partly rest on that foundation; that the principles of Probabilities owe something to the principles of mechanics which rule the movement of molecules, which underlies the phenomena of chance.

But whatever the corner-stone and ultimate foundation, there can be no doubt about the stability and splendour of the edifice. There can be no doubt that the leading law of probabilities derives added validity and majesty from its connection with molecular physics. With something of the confidence inspired by physical science we turn to the applications of the law in statistics relating to society.

3. In the first example which I adduce the normal law is generated by exactly the same process as that which I have

employed to illustrate a physical thereom, namély the random selection of digits from a table of logarithms. I refer to the specimen of the method of sampling given by Mr. Bowley in his Presidential Address to the British Association in 1906. By the use of random digits Mr. Bowley selects a sample of 400 from a set of statements as to the percentage yield of different investments numbering nearly 4,000. From the sample he obtains with adequate accuracy the average yield of the investments, and the proportionate numbers or frequency of different classes of investment divided according to the amount of yield. Mr. Bowley allows me to cite (in my Appendix) another instructive experiment in sampling which he has more recently performed. Mr. Bowley rightly anticipates much saving of trouble from the use of such methods: "There is no need to make a house to house visitation to "learn the conditions of a district; it is sufficient to enumerate the "houses, to choose a certain proportion at random and investigate "carefully the state of their inhabitants."

It must ever be remembered that the conclusions obtained by the method of sampling at its best are only probable; and that the improbable will sometimes occur. Mr. Bowley had an experience of this sort. For whereas from his sample of 400 it was to be expected that the percentage of investments having a yield between 8*l.* and 10*l.* 19*s.* 9*d.* was 7·25; in fact (the whole set of nearly 4,000 having been examined) the percentage number of this class proved to be only 3·8. The improbability of such a divergence between theory and fact is calculated by Mr. Bowley to be about the same as that of drawing two named cards from a complete pack: that is, very considerable. It should be observed, however, that the probability thus calculated relates to the occurrence considered as a single event. Whereas, when there are several trials, as in the case before us there were several percentages distinguished, the probability of failure in some one trial becomes serious. If you go on exposing yourself to risk, you must expect to get hit. This is the rationale of the "paradox" adduced by De Morgan in his *Budget:* that in the expansion of the constant π the number of times that the digit *seven* occurs differs from the number of times which is *a priori* probable (the total number of digits in the expansion under observation divided by 10) to an extent which cannot be accounted for by mere chance. As Dr. Venn has observed, the probability of the occurrence is not simply the probability of drawing from an urn in which balls of ten different colours are mixed up in equal proportions a sample in which the proportion of balls of a *particular assigned* colour differs to a specified extent from the expected proportion, one in ten; but the probability that *some one* (at least) of the colours should present that discrepancy. When, as in the case now before us, the numbers of balls of each colour—the numbers of investments in each of the classes into which the sample is divided—are given, it is proper to apply Professor Pearson's beautiful criterion in order to determine the (im)probability of the composite event. Applying the criterion to Mr. Bowley's sample of 400 I find that nothing very unexpected has occurred. The *a priori* odds against the occurrence are about one to twenty; the probability is rather greater than that of drawing a black king (spades or clubs) from a pack of cards. The moral seems to be that we may obtain from samples a general outline of the facts—often sufficient for the initiation of a project like that of Insurance—rather than the features in detail.

A model of sampling less perfect in theory but closer to practice is afforded by some Norwegian statistics which have been marshalled by Dr. Kiär. A sample numbering 11,427 of statistics relating to the income and property of male persons at different ages was taken from the results of the Norwegian census, by three successive operations of more or less perfectly random sifting. First, certain localities were fixed on, apparently rather according to some official classification than by a genuinely random process. Then, as representative of persons at different periods of life, those whose age fell just between the round numbers 15, 20, 30, 35 . . ., that is, persons of the ages 17, 22, 27, 32 . . ., were selected. The initial letter of the name formed a third basis of selection.

There is generally a weak point in methods of sampling other than the most abstract; but, prior to experience, it is difficult to say where the point will be. In the example before us it might have been expected that the first step would prove treacherous. The third also, sampling by the initial letter of names, would justly be suspected in some countries—in the lands of the "M'cs" or the "O's." In fact, however, it seems that doubt attaches chiefly to the *second* operation, for the reason, if I understand rightly, that, as people advance in age, they are more and more apt to return for their age a round number, such as 45 or 50, rather than the intermediate 47. Accordingly the numbers of persons at intermediate ages, which are taken as typical in the sample, are apt to be unduly thinned at the later as compared with the earlier ages.

In spite of these imperfections, a very good result seems to have been obtained, as I infer, by comparing many of the percentages deduced from the sample with the actual figures of the census. Applying the Pearsonian criterion to several tables, I find the results to be so good that only about once in a thousand times could one expect a better result from ideally perfect sampling. To the statistical practitioner the following tests will perhaps be more satisfactory. Out of *fifteen* figures obtained as the percentages of the population in towns, married or unmarried, at different ages, I find that in the case of *eleven* the difference between the true and the calculated figure is less than 5 per cent. of the true figure; there is *one* error of just 5 per cent.; *two* errors between 10 and 20 per cent., and one just above 20 per cent. The distribution of the rural population, married and unmarried, at different age periods, is given with nearly equal accuracy by the sample. Out of *sixteen* figures (omitting one for which the figure given by the census was 0, and accordingly the ·1 given in the sample might be construed as showing an *infinite* relative error!) I find that *eight* show an error less than 5 per cent. of the true figure; there are *six* errors between 10 and 20 per cent., and one above 20 per cent. It must be remembered that these errors are per cent. of quantities which are themselves small percentages of the total populations. Other tables afford similar comparisons.

The worth of the sample is confirmed by its agreeing, when comparable, with a second sample taken, as I understand, with a more directly practical purpose connected with Insurance. The need of such a method for the purpose in hand is strongly suggested by the number of questions, some fifty, as to occupation of father, expenditure on food, fuel, clothing, &c., number of days lost by sickness and so on. It would be expensive and probably nugatory to ask all these questions in a general census.

The method of sampling under the designation of "The "Representative system" appears to be a permanent institution in Norway. Dr. Kiär, the distinguished Director of the Statistical Bureau at Christiania, writes: "This method has according to our "experience given very good results, and we have thereby acquired "valuable statistical information which it would probably have "been impossible to obtain in any other way." An important safety appliance is obtained by checking, or more exactly "con-"trolling," the results of the sample by the complete statistics with respect to some of the heads which admit of this test.

Altogether, these statistics appear amply to support the case for sampling, as ably maintained by Dr. Kiär at the meeting of the Fifth International Statistical Institute at Berne. His opponents could only object that the practice was very dangerous—especially, added Dr. Mayr, in view of the proclivities of the mathematicians who prefer calculating to observing. The danger attending mathematical machinery must frankly be admitted. But there should be balanced against it not only the saving of expense, but also the greater accuracy which may be attainable when elaborate questions are put to a select few rather than to the general population. The limited number, as Professor Schmoller, quoting Hesiod, told the Institute, may be "better than the whole."

The danger of dealing with samples according to mathematical rules is particularly great when the samples are not, as in the preceding examples, selected by a (more or less) random process from a given set of statistics; but the samples are the given statistics considered as specimens of an indefinitely larger set to which the statistics belong, a logical class rather than a particular multitude. For instance, suppose the subject in hand to be the proportion between married to unmarried men at a certain period of age. It is one thing to take the figure obtained by a carefully instituted sampling as representative of the true proportion given by a census (actual or potential); it is another matter to take the figures given by a census as representative of the relation at different times and places. Not only is there now the usual hazard attending the inductive leap from the known to the unknown; but also there is not the same security that the conditions of good sampling have been complied with. Even in the best constructed urns, as Mr. Yule has reminded us, the balls may not behave in perfectly random fashion; those of a particular colour may be more polished in such wise as to evade the hand of the operator. Very frequently the balls which represent concrete phenomena are apt to be stuck together, so that the number of independent causes, the true n of the normal formula, is not what it appears to be, not simply the number of balls. Neglect of this consideration has stultified many elaborate calculations of probability. Volumes have been written to recommend the use of Poisson's formulæ in medical statistics. But the cases observed in hospitals cannot, in general, be treated like so many balls drawn independently at random from an urn with a fixed (or even with a wavering) proportion of balls of different kinds. There are large common causes affecting considerable numbers of patients; for instance, the weather at different periods, or the circumstance, which may not appear in the statistics, that the character of an epidemic, whether mild or severe, is apt to affect large batches of patients identically.

I take these objections from Von Kries, one of the writers on Probabilities who may be compared with the historical school of economists as critics of classical authorities. They have no doubt performed a useful work; and a pleasant one, so far as accompanied with a sense of superiority over intellectual progenitors. But the revision of inspired originals may easily be carried too far. "To err "with Plato" is sometimes preferable to the common sense of commentators. In the matter before us the condemnation of the classical authorities requires to be softened by three extenuating circumstances.

First, the conditions requisite for the application of Laplace's, or Poisson's more general, formula (for testing the significance of differences in proportions) are more generally fulfilled than some of the critics have supposed. The fulfilment of the conditions is by no means confined to the one instance commonly admitted, the proportion of the sexes at birth. In several important classes of statistics relating to mortality the conditions seem to be adequately fulfilled. It is hardly too much to say that in the majority of statistics pertaining to social phenomena the concept of pure sortition is appropriate, provided that the number of observations with which we are dealing is not very large (as to the significance of which condition see Professor Bortkevitch's admirable article on the Applications of Probabilities to Statistics, in the *Encyclopädie der Mathematischen Wissenschaften*). Dr. J. H. Peek, of Overeen, in Holland, deserves especial mention as having empirically established the wide applicability of the *urnen-schema*, the analogy between the fluctuation of concrete statistics and that of balls drawn at random from an ideal urn. Dr. Peek has observed this character of pure sortition in (a ten-year series of) death rates of the general population at particular ages; and for other even less narrowly defined categories. Suppose that instead of considering the numbers of deaths varying from year to year the writer had considered the varying amounts of money paid on death (a fixed sum being payable at each death); evidently a comparison of calculated and observed deviations of payments from their value could equally have been employed to verify the validity of the *urnen-schema*. Such, as I understand, is the significance of certain statistics presented in Dr. Peek's paper—"On the "Application of Probabilities"—read at the Congress of Mathematicians at Cambridge last August. The correspondence between the results of his "first method" (inference from the *urnen-schema*) and the second more empirical determination of the fluctuation in question is very convincing.

Judging from my own experience in this matter, partly recorded in the *Journal* for 1885—experience much less extensive than Dr. Peek's—I should say that he had been rather fortunate in his examples. Dealing with English death-rates (possibly less homogeneous than Dutch ?), I only succeeded in verifying the character of pure sortition—the *urnen-schema*—for classes very much narrowed, deaths of persons at the same age in the same occupation.

Secondly, the empirical fact which has just been noticed is corroborated by theory; the important theory due to Professor Bortkevitch that statistics relating to rare events (of which the probability is a very small fraction) are apt to fluctuate in almost perfect accordance with the scheme of an ideal urn. The theory of the matter has lately been discussed afresh by Dr. Mortara, Professor of Statistics at Messina, and he has added some new and striking empirical verifications of the theory. Some of his specimens could not have been obtained in this country; such as marriages between aunts and nephews. The event is, indeed, rare in Italy, but not unknown nor unrecorded—about two such marriages each year in Sicily, two in three years in Lombardy. The statistics fulfil with remarkable precision Professor Bortkevitch's law of small numbers.

Thirdly, even though the *urnen-schema* is not fulfilled in fact, nor expected by theory, this imperfection is not fatal to the use of the normal law as an aid to induction. Provided that the degree of imperfection—the "coefficient of divergence" (from the normal type), in the phrase of a leading writer of the subject—is ascertainable and constant, the normal law is still available, as shown in the Jubilee volume of our Society. The reasoning by which significant differences in statistics are distinguished from fortuitous fluctuations is still substantially the same.

Yet the importance generally attached to the presence or absence of the conditions proper to the urn scheme seems not to be altogether without foundation. Let me illustrate the point by a trivial example. In the volume referred to statistics as to the number of wasps going into or coming out of a nest were adduced; and a coefficient was empirically obtained for testing what differences in the rate of movement were not fortuitous, but significant of changed conditions. It was subsequently ascertained that this coefficient is exactly that which is presented on the supposition that the individuals going in and out are random selections from the large total at work. In fact the wasps dart forth with much the same regular irregularity as the "alpha" particles to which I have referred as specimens of physical statistics. Well, the foreknowledge of this fact would surely be an asset of some importance in fine reasoning as to the interpretation of the statistics. The datum gives an advantage to those statistics over the only other statistics *in pari materiâ* with which I am acquainted. I refer to the observations made by the distinguished entomologists, Mr. and Mrs. Peckham, on a large wasp's nest in Wisconsin; a record of exits and entrances kept continuously from early morning to noon one day in August, 1886. The character of normal dispersion is conspicuously absent from this record. The British communities which I have observed contrast favourably in respect of regular movement with the more turbulent republic of the West.

I have now reached a point which marks the transition from discrete attributes to continuous quantity. The difference is very fine and of little importance so far as the working of the normal law of frequency is concerned. There is no essential difference, for example, between the use of the law in reasoning such as I have suggested about the *ratios* of the numbers of wasps issuing per minute to a certain total and the *absolute length* in inches of cuckoo's eggs, which forms the object of an elegant enquiry instituted by Mr. Latter. He proposes the questions whether the eggs of cuckoos deposited in the nest of any one species "stand out as a set apart "from cuckoo's eggs deposited elsewhere," and if so, whether they deviate in such a direction as to approximate in size to the egg of the foster parent. By the use of the normal law he ascertains that "hedge-sparrow cuckoos"—that is those who lay their eggs in the nests of hedge-sparrows—and certain other similarly defined classes of cuckoos, "do present differences marking them out as "distinct sets," and that some of them at least do differ from the main body "in the sense of the particular species of foster parent."

I have now entered on a large topic, the analogues in social statistics of the theory of errors in physics. But, as already intimated, I do not propose now to retouch that subject. The omission involves examples of sums (as well as averages) such as that which Mr. Bowley has dealt with in forming an estimate of income other than wages below the limit of exemption. Merely recognising the great lacuna which is thus made in this summary sketch, I go on to a distinct topic.

4. In the preceding examples the normal law of frequency has been manufactured by averaging statistics. In the cases now to be considered the averages, or rather "weighted sums," fulfilling the law are furnished, ready made, by the nature of things. We have already contemplated and explained a normality of this character in molecular statistics. It may be conjectured that the frequent appearance of the normal law in biological statistics admits of a similar explanation. Hereditary attributes are distributed among the members by continual crossing in the long course of generations, somewhat as resultant velocities are distributed among molecules in the course of repeated collisions.

Quetelet rather than Laplace is the path-breaker here. To Quetelet belongs the honour of having recognised the general prevalence of the normal law among the members of natural groups; and having assigned the presumable cause of that prevalence, the co-operation of numerous independent agencies. He, indeed, too much ignored the asymmetry—inconsistent with perfect normality —which is commonly present in some degree in actual groupings. But the omission is the more excusable in that the causation assigned by Quetelet leads straight to a correction of the normal law which often suffices to make the fit satisfactory (that second approximation due to Poisson, familiar to the students of Mr. Bowley's *Elements*). Such was the position when Professor Pearson's second Contribution to the Mathematical Theory of Evolution made its epochal appearance. Though the need of the new methods was not at once universally recognised, experience and reflection have shown that Professor Pearson was rightly inspired in making a new departure. No mere development of (the hypothesis underlying) the normal law (on the lines of Poisson) is adequate to represent the variety of concrete groupings.

This series of attempts to represent natural groups by mathematical forms may be illustrated by the history of astronomy—to compare certain things with uncertain. There is first the *circle* prescribed on the authority of Aristotle as the pattern of celestial movements. The pre-eminence attached to the circle was no doubt somewhat superstitious, though, as Whewell points out, the properties of the circle are really very important, "circular "functions" being still required to represent complicated astronomical movements. The circle having failed to represent the phenomena adequately, "epicycles" are superadded with some success, corresponding to the (Poissonian) correction of the normal law. This correction proving inadequate, Professor Pearson, like another Kepler, proposes a better fitting curve, one different from, but having some relation to the normal curve, as the ellipse may be regarded as a defective circle. But, note that it is with Kepler's ellipse, not Newton's, which we have to do. *This* ellipse is not rested on a physical cause. It is at least as open, as the other statistical schemes, to the Miltonian description of the astronomical tentatives.

How, will they wield
The mighty frame! How build, unbild, contrive
To save appearances! How gird the sphere
With centric and eccentric scribbled o'er
Cycle in epicycle . . . !

The *statistical* "cycle" and "epicycle" have indeed an advantage over other constructions, in that their forms are deducible from a known cause. Unfortunately the cause is not always present! But the "ellipse" only claims to "save appearances." It is not rested on a foundation of physical cause; in this respect recalling Bacon's complaint against the Copernican theory.

The following hypothesis, however, may be lent to the Pearsonian method. As a perfectly chaotic mixture of agencies results in the normal law of frequency, and a less perfect chaos, in the generalised form which has been described in our *Journal* as the "law of great "numbers"; so in the gradual inroad of law upon chaos, in the scale of fortuitousness, there must be a stage characterised by forms bearing still some resemblance to the normal type, but with features considerably defaced—a deformation much greater than that of the aforesaid "generalised" law. For example, the three stages, in reverse order, might presumably be presented by a mixture of gases working down from a violent initial disturbance to the normal state. The Pearsonian construction seems well calculated to express this sort of deformation.

But on this hypothesis it occurs that some other deformation of the ideal "circle"—some *oval* other than the "ellipse"—may be equally qualified to save appearances. A variant of this character is suggested in a Paper contributed to the Statistical Section of the Congress of Mathematicians at Cambridge this year. I cannot pretend to be an impartial judge in this matter. I will only observe that the issue is not so much a question of right, as of expediency, especially in the way of saving trouble. The new construction comes not to destroy, but to fulfil Professor Pearson's theories. For example, it confirms the rule which he gives as to the relation which commonly prevails (in skew groups) between the different kinds of average—the mode, the median, and the mean.

5. Passing from *curves* to variation in two dimensions, normal *surfaces*—of which a perfect specimen has already been presented in relation to molecular physics—I remark that we have here a new departure, a higher stage, in so far as a fresh instance is here presented of an exact quantitative *law* governing the phenomena of chance. I refer, of course, to the *ratio* or *right line* which forms the measure of so-called regression or correlation. To Galton belongs the honour of having first discerned the statistical significance of regression. A debt is also due to Professor Pearson for having pointed out the best method of determining the coefficient which measures correlation between two organs or attributes. It deserves attention that he obtained this result by a stroke of Inverse Probability, a method of reasoning which narrow-minded critics of the classical authors are wont to decry.

I need not dwell on the properties of normal correlation, as they have been fully elucidated by Mr. Yule in our *Journal*. I will only offer a few remarks on some declensions from the pure normal type which often occur in practice.

First suppose that—other things being in order—the statistics do not perfectly fulfil the normal law, but, on the contrary, are somewhat "skew." I hope that practical statisticians will give a trial to the methods of representing such groups which have been proposed in the Paper already referred to with reference to skew *curves*.

Next suppose that the material is indeed normal, or would be if it were perfectly measured, but that perfect measurements are not to hand. The milder species of this defect is when a complete set of measurements is not given or is, to save trouble, ignored; as has been done with remarkable success by Dr. Macdonell with reference to anthropometry. The extrication of the required coefficient from data thus imperfect has been accomplished ingeniously and adequately by Dr. Sheppard, more generally—with signal mathematical skill—by Professor Pearson. Of more philosophical interest is the application of these methods to cases in which the defect of measurement is due to the imperfection of our faculties, as in the case of colours, which we can arrange in a scale but cannot measure numerically. Always presuming that the normal law is in a sense fulfilled, it appears possible to determine an exact quantitative correlation (between classes of persons) in respect of such unquantified attributes, as eyecolour or good temper. The practice of the classical writers on Probabilities who did not hesitate to make "moral," in the sense of psychical, advantage a subject of calculation, seems to be countenanced by this modern art of measurement.

Now let both the defects which have been noticed be present together. Then we fall back upon a case, already considered, in which the proportion of a species defined by a certain attribute is different for different classes—as when the proportion of deaths is different for patients (otherwise similar) differently treated. How far now, in such circumstances, is it advisable to employ the term coefficient of correlation? How much will coefficients of association conduce to the practical object of curing patients? But I am aware that these are just now burning questions, and I bear in mind what Bagehot says about a controversial topic: "If you say "anything about the Act of 1844, it is little matter what else you "say, for few will attend to it" . . . "a single sentence respecting "it is far more interesting to very many than a whole book on any "other part of the subject." So I pass on to a final topic.

6. I cannot conclude without adverting to one of the principal social uses of Probabilities, the application of that calculus to the business of Insurance. Contemplating the construction of Life-Tables every Statistician must join with Dr. Farr in his tribute to "the illustrious Halley who by his table showed that the generations "of men, like the heavenly bodies, have prescribed orbits which

"analysis can trace." But, as Dr. Farr's encomium suggests, it must be remarked that the subject is only in part within our province. For, as well observed by one of the most lucid writers on the subject, "the problem of a mortality table [*Absterbe-ordnung*] "seems to resolve itself partly into a physical [naturwissent-"schaftliches], and partly into a Probability problem." In fact we are here on the confines of law and chance.

> "Here Nature first begins
> "Her farthest verge, and Chaos to retire
> "As from his outmost works a broken foe."

So far as actuarial mathematics deal with laws of nature in the ordinary sense of the term excluding chance, they are not here relevant. I must not therefore dwell on the niceties of interpolation, in particular the method which seems to be much in vogue with the actuaries known as "osculatory interpolation." The "fascinating "problem," as it is called by a distinguished expert, cannot here be entertained. I advert only to the problems which involve an element of Probabilities. They are mainly two, I think: the estimate of the *risk* run by an Insurance Company, and the representations of a mortality table by a law of frequency of the kind proper to Probabilities.

As to the first topic, I venture to express the opinion that for the general mathematical reader, if the expression may be allowed, the best, or at least a very good, introduction to the subject is given by two of the older writers in the Theory of Probabilities, Laplace and De Morgan. It may be remarked that Laplace's calculation of risk does not involve the assumption that deaths at different ages vary as so many balls of different colours taken from an ideal urn at random. There is only postulated the statistical constancy (not the normal fluctuation) of the proportions. De Morgan, indeed, postulates more when he deals with the probability that the proportions are liable to a certain error—a calculation proper to a special Life-Table.

The problem of "graduating" a life-table has long fascinated mathematicians. The records of experience have been "scribbled "o'er," to repeat our astronomical illustration, with a variety of formulæ adapted "to save appearances." Especial prestige attaches to Professor Pearson's scheme for representing the data by the repeated use of his flexible formula. For the data on the confines of law and chance appear to be exactly of the kind to which his formula is specially applicable.

A similar claim may be made on behalf of a certain method practised by Mr. Hardy, which consists in so adapting or "trans-"lating" a normal curve so as to fit data which are far from normal.

The general idea—independently struck out by the author—is the same as that of the method which I have proposed; while the ingenuity and success of the application are all his own.

But I dare say that neither of the methods distinguished will be much used by the practical actuary. He will attend not to the element of chance in the phenomenon, but to the element of law—the law of Gompertz and Makeham—in favour of which there is not only fact and reason (the rationale assigned by those authors) but also convenience, the law embodying the one formula which lends itself to the ready calculation of annuities for joint lives. In fact, to parody a daring French epigram, if the law did not exist it would be almost necessary to invent one. It is allowable, at least, according to distinguished writers on the calculation of Life-Tables, to "stretch a point" in favour of the Gompertz-Makeham law.

In leaving this subject I cannot say with Laplace: "I will not "dwell further on matters connected with insurance, because they "do not present any difficulties." On the contrary, difficulties which the practical details of the subject present to a layman compel me to be content with these meagre generalities.

It remains only to gather up the lessons which our rapid survey may convey.

Our first exemplification of Probabilities suggests a comparison between the uses of the higher mathematics in Political Economy and in Statistics. The connection of Probabilities with molecular physics has for the statistician the theoretical interest which the pre-eminence of the principle of *maximum* in mathematical physics generally may have for the economist. The statistician finds in the world of atoms an ideal model of that law in which he should exercise himself day and night; the economist is conducted by the theory of maxima to the contemplation of interdependent variables and all that is implied in the conception of a "margin." But whereas these general ideas form the principal or, according to some authorities, the only contribution of the higher mathematics to political economy, the Calculus of Probabilities affords more tangible service by directing the operations of sampling. I refer especially to the artificial kinds of sampling which have been instanced. For as to the constructive samples presented by data, such as the experience of hospitals, there is reason to fear that the Calculus of Probabilities—even if supplemented by the refinements of "Association" or "Correlation"—may not prove so powerful an aid to the ordinary methods of Induction. The *independence* of the observations postulated by the calculus is too often wanting. There is, indeed, as I have noticed, a remedy for this defect. But its application, requiring a prolonged series of observations, seems hardly available in statistics relating to society; however appropriate elsewhere, in metereology for instance, or with respect to the *unprogressive* communities of the animal world. Human society will not stand still long enough for "coefficients of divergence" from the ideal kind of frequency to be calculated. But the attempt may be not wasted so far as it cultivates a power of dealing with figures differing in degree of accuracy and what may be called a sense of probable error, which great statisticians like Giffen seem to possess instinctively.

This remark is equally applicable to the large branch of the subject which I have been compelled to omit, the analogues of the theory of errors in physics, such as that which the construction of index-numbers may present.

The character of progress in human institutions which has been mentioned is also unfavourable to the employment of analytical curves and surfaces to represent groups of statistics. But those methods seem particularly well adapted to investigate the slower progress of biological evolution. If the term "relating to society" in our title were interpreted so widely as to include biology, more than a passing tribute of admiration would be due to the band of mathematical statisticians who under the auspices of Professor Pearson have improved the weapons of Probabilities and brought new regions under her sway. Of statistics in the narrower sense, pertaining to states and statesmen, those parts, I think, are most amenable to the new methods which are most nearly related to our physical nature—in particular vital statistics.

On the whole, I accept the practical answer which our Society gives to the question: what is the use of mathematical methods for our purposes? The general tenor of our statistical reasoning remains what it was in the days of Porter and Newmarch. But we permit an occasional outbreak of mathematics. As one of my predecessors put it: there should be once in a session a paper which no one—or hardly any one—can understand. To have had once in an age a presidential address of this character will, I hope, not seem excessive.

APPENDIX.

Appended are some notes and references relating to the above sections in the order of succession.

1. (p. 167.) Reference is made to Dr. Farr's Inaugural Addresses to the Statistical Society, 1871 and 1872. Dr. Farr observes in this connection: "This Society owes its origin to two of the foremost "mathematicians and physicists of the age on the special ground

"that they saw in statistical phenomena a wide field, beyond "expression interesting to men, under the domain of law." Compare the resolution carried by Quetelet at the sixth meeting of the International Statistical Congress (quoted by Samuel Brown, *Journal of the Statistical Society*, vol. xxxi, p. 22).

(p. 168.) Dr. Venn's classification of "Uniformities" is given in his *Empirical Logic*, p. 94. His view of statistical "series" in his *Logic of Chance* should be compared.

(p. 168.) M. Jean Perrin's experiments are described in the *Annales de Chimie et de Physique*, 8me Series, September, 1909; translated by F. Soddy, as *Brownian Movement and Molecular Reality*, 1910. (Compare M. Perrin's lectures to the Royal Institute, February 24, reported in the *Chemical News* for October and November, 1912.) The Brownian movements, named from their first observer, are not to be confounded with the dancing of dust motes in the sunlight which Lucretius has described in terms singularly appropriate to the true molecular action:—

Multa videbis enim plagis ibi percita cœcis
Commutare viam retroque repulsa reverti
Nunc huc nunc illuc in cunctas denique partes.

. . . .

Sic a principiis ascendit motus et exit
Paulatim nostros ad sensus.

(*De Rerum Naturâ, lib.* II, v. 129 *et sqq.*)

(p. 168.) Dr. Geiger's experiments are described in the *Philosophical Magazine* for 1910, vol. 20, p. 700.

(p. 169.) Among statisticians who have approached the problem of philosophical determinism or theological predestination may be mentioned the Prince Consort in his Presidential Address to the International Statistical Congress of 1860 (*Journal of the Statistical Society*, vol. xxiii).

The title of Von Kries' logical disquisition is *Die Principien der Wahrscheinlichkeits-Rechnung*, 1886. See his remarks on "das "eigentlich Unempirische Princip," p. 170, and on "Spielräume" *passim.*

2. (p. 170.) The cognate papers in the *Journal* to which reference is here made are chiefly "Methods of Statistics," *Jubilee Volume*, 1885; "The Generalised Law of error," 1906; "The "probable errors of frequency-constants," 1908. Compare the contribution to (the *Bulletin de*) *l'Institut International de Statistique*, 1909.

(p. 170.) The illustration of molecular movements in two dimensions by a crowd of billiard balls is suggested by Mr. Jeans in his *Dynamical Theory of Gases*, p. 5. The same work may be referred to as illustrating the stupendous character of the received demonstrations; the author starts his principal proof by supposing six times as many dimensions (in hyper-space) as there are molecules under consideration—presumably some trillions!

(p. 172.) The genesis of the law of error is thus illustrated by linear functions of digits taken at random. Below are forty-eight statistics, each of which is obtained as follows:—Twenty-five digits are taken at random from a table of logarithms, each forming the seventh place of decimals in successive logarithms, beginning with the logarithm of 101 (and ending with the logarithm of 1,300). From the sum of the twenty-five digits which go to each of the forty-eight statistics there is subtracted 112·5 = 25 times 4·5; 4·5 being the mean value of an indefinitely great number of digits taken at random. This difference divided by 5 (the square root of 25) forms one of the group of statistics given below; a group which ought approximately to conform to the normal error-curve of which the central point is zero and the parameter or modulus the square root of twice the mean square of deviation of random digits from 4·5; that mean square being 8·25.

Below zero, without their (negative) sign, in a descending order of absolute magnitude (not the order of occurrence):—

6·9, 6·5, 5·5, 3·5, 2·7, 2·7, 2·5, 2·5, 2·5, 2·3, 2·1, 1·9, 1·9, 1·7, 1·7, 1·5, 1·3, 1·3, 1·3, 1·3, 1·1, ·9, ·3.

Above zero, without their (positive) sign, rearranged in an ascending order of absolute magnitude:—

·1, ·1, ·1, ·3, ·5, ·5, ·7, ·7, ·7, ·9, 1·1, 1·5, 1·5, 1·7, 1·9, 1·9, 2·3, 2·7, 2·9, 3·5, 4·9, 5·1, 5·3, 5·5, 6·5.

It may be noticed that the *Median* (the point which has as many observations above it as below it) is not, as it should be theoretically, zero, but + ·15. But this deviation from the theoretical value is well within the probable error which characterises the probability that if the forty-eight given statistics were really samples of a perfectly normal indefinitely large group, the observed irregularity might occur, *i.e.*, about ·33.

Likewise the *Quartiles* (the two points below which and above which respectively occur a quarter of the total number of observations) are − 1·9 and + 1·6 (half-way between + 1·5 and + 1·7). Whereas the theoretical deviation of each from zero is ·4769 Modulus = 1·94 nearly. There is thus (in one case) a difference between the observed and theoretical value of ·34, which is nothing extraordinary considering that the probable error to which the determination is liable is about ·37.

So, there ought to be theoretically four-fifths of the total number, say thirty-eight or thirty-nine observations, between the limits + 3·68 and − 3·68 (the Deciles = ·906 Modulus). And in fact forty of the observations lie between those limits.

The mean (deviation from zero) of the forty-eight observations taken without their signs in absolute magnitude is 2·266; while theoretically it should be Modulus/ $\sqrt{\pi}$, *i.e.*, 2·29. The difference between the actual and the theoretical values is less than ·03, the probable error, about ·05.

Aggregates of *sixteen* digits present similar characters, but less perfectly owing to the smaller number of constituents which go to each observation. Below are forty-eight statistics, each formed from the last sixteen digits in one of the batches above described. From the sum of each set of sixteen there was subtracted 72 (16 times 4·5), and the difference was divided by 4 (the square root of 16).

Below zero in the order of magnitude:—

5·5, 5·5, 4·75, 3·25, 3·25, 3·25, 3, 3, 2·5, 2·5, 2·5, 2·5, 2·5, 2, 1·75, 1·75, 1·75, 1·5, 1·5, 1·25, 1·25, 1·25, 1, ·75, ·75, ·5.

At zero, one observation.

Above zero:—

·25, ·5, 1, 1·25, 1·25, 1·5, 2, 2·25, 2·25, 2·25, 3, 3·25, 3·25, 4, 5, 5·25, 5·75, 5·75, 6·5.

The mean deviation of the forty-eight observations in absolute magnitude is 2·5, showing a difference from the theoretical mean, 2·29, not enormously greater than the probable error, viz. (in the sense above explained), about ·05.

(p. 173.) In the variant proof which is offered for Maxwell's theory of the distribution of velocities in a molecular chaos the factors a and α are, of course, to be replaced by sin θ and cos θ. By substituting for these cos α, cos β, cos γ, the direction co-sines of any line in space, the proof may readily be extended to three dimensions.

From the similarity of the expression for the mean velocities, say, u and v in two dimensions, it is perhaps evident that the mean squares of the velocity are equal. Or this may be considered as given by the facts of pressure; as in the paper on the subject in the *Philosophical Magazine* for January, 1913. Or it may be deduced from a smaller assumption, namely, that the line joining the centres of a colliding pair of (spherical equal perfectly elastic) molecules is as likely to be in one direction as another. Let θ be the angle made with the axis of x on any occasion. Then if u_s, v_s, and u_t, v_t, are the velocities, of the two molecules before collision resolved in the direction of the axis OX, we have for U_s, the velocity of the first in the direction of the line joining the centres, $u_s \cos\theta + v_s \sin\theta$, and for V_s the velocity in the direction perpendicular thereto, $-u_s \sin\theta + v_s \cos\theta$; with corresponding expressions for the molecule t. The velocities U_s and U_t are exchanged in consequence of the collision, and accordingly we have for u_s', the velocity of the s molecule after collision

$$u_s' = U_t \cos\theta - V_t \sin\theta\,;$$

$$u_s' = \cos\theta\,(u_t \cos\theta + v_t \sin\theta) + \sin\theta\,(v_s \cos\theta - u_s \sin\theta).$$

Now suppose every variety of u and v to occur, θ retaining the particular value assigned to it; then, as by hypothesis θ is independent of the velocities, we obtain an expression for the mean value of u_s' for the particular θ in terms of the mean velocities (in the directions of the axes). Those mean velocities are, indeed, zero (the centre of gravity of the system being at rest). But not so the mean square of velocities, which, in consequence of the circumstance (deducible from the theorem in the text) that the u's and v's are distributed normally, and the further assumption that they are practically, though (for a reason suggested in the text) not *perfectly* uncorrelated, becomes *clear of products*. If then we put $[u_s^2]$, $[u_t^2]$, &c., for mean square of velocity, and denote by $(u_s'^2)$ mean square of velocity in the direction OX of molecules colliding with line joining their centres at the particular angle θ, we have

$$(u_s'^2) = [u_t^2]\cos^4\theta + [v_t^2]\sin^2\theta\cos^2\theta + [v_s^2]\sin^2\theta\cos^2\theta + [u_s^2]\sin^2\theta.$$

For example, if $\theta = \frac{\pi}{2}$, $(u_s'^2)$ reduces to $[u_s^2]$; as it evidently ought to do, since the movement in the direction OX is unaffected by such a collision. Or if $\theta = o$, (u_s^2) reduces to $[u_t^2]$ as it ought to do, since the spheres in every such collision exchange velocities in the direction OX, Now (θ being independent of the u's and v's) it is proper to integrate the right-hand side of the above equation in order to obtain an expression for $[u_s^2]$, the mean of (u_s^2) for all possible values of θ, that is the mean of u_s^2 for all possible values of u. We have thus, integrating between limits O and $\frac{\pi}{2}$,

$$\frac{\pi}{2}\,[u_s']^2 = [u_t]^2\,\tfrac{3}{16}\,\pi + [v_t^2]\,\tfrac{1}{16}\,\pi + [v_s^2]\,\tfrac{1}{16}\,\pi + [u_s^2]\,\tfrac{3}{16}\,\pi\,;$$

$$\tfrac{1}{2}\,[u_s^2] = \tfrac{1}{2}\,[u_t^2].\text{—Q.E.D.}$$

By parity of reasoning it may be shown that if there are two sets of molecules, each like that above considered but with different masses, the mean square of velocity for each set will be in inverse proportion to the mass of the molecules of that set. There will be an equal partition of mean energy between the two sets.

The Calculus of Probabilities may be employed more generally to verify Maxwell's theory of the Equipartition of energy.

3. (p. 175.) Mr. Bowley's Presidential Address to Section F of the British Association is printed in the *Journal of the Statistical Society* for 1906. Here is the additional sample of the practice of Sampling which Mr. Bowley has supplied. It was required for a certain purpose

to ascertain the distribution of the population of England and Wales in rural districts, the unit being the civil parish, and the objective the density. For this purpose the civil parishes in rural districts were numbered consecutively from 1 to 12,830. Next 250 numbers were selected from logarithm tables in such a way that every number from 1 to 12,830 had an equal chance of inclusion. The parishes corresponding to the numbers were then selected and tabulated. In order to test the sufficiency of the sample, advantage was taken of a tabulation according to number of persons in a parish given in the Census, Cd. 6,258, p. 428. Here all parishes are included, but those in urban districts were subtracted. The result of the sample of 250, tabulated in accordance with the census classification and compared therewith, is as follows:—

Proportion of persons in parishes per 1,000.

	—100.	100—	200—	300—	400—	500—	1,000—
Sample........	140	208	168	108	80	164	132
Fact............	152	192	147	108	80	173	146

(p. 175.) De Morgan's paradox is discussed by Dr. Venn at p. 247 of his *Logic of Chance*, third edition. As to the principle, compare Cournot's *Exposition des Chances* (§ 114); with whom I may say, "Je ne dissimule pas ce qu'il y a de délicat dans toute "cette discussion." By a summary application of the principle, Dr. Venn, dealing with 708 digits, finds that the odds are reduced from the *primâ facie* 44 to 1 down to 4 to 1. By the Pearsonian criterion (dealing with the 608 digits analysed by De Morgan, *loc. cit.*, p. 291) I find that the odds against the observed event occurring by chance are only 3 to 1.

(p. 176.) Dr. Kiär's exemplifications of the method of sampling are given in the *Allgemeines Statistisches Archiv*, vol. v (1879), p. 1, *et sqq.*

(p. 176.) The caution required in applying the analogy of balls drawn at random from a bag to concrete statistics is well inculcated by Mr. Yule in his *Theory of Statistics*, ch. xiv. It should be observed that the "sampling" here explicitly treated corresponds to his "sampling of attributes"; while his "sampling for variables" is nearly coincident with the analogues of the Theory of Errors, which are here omitted as having been treated in former papers.

(p. 178.) The remarks in the text on the use of Probabilities in Medical Statistics are fully sustained by Gavarret's *Statistique Médicale*, Liebermeister, and other nineteenth-century writers *in pari materiâ.*

3. (p. 179.) Dr. Peek's principal observations on the coincidence

between the combinational and physical dispersion (to use Lewis' distinction) are given in *Zeitung für Versicherungs Recht u. Wissenschaft*, 1899.

(p. 179.) The reference to the *Journal of the Statistical Society* is not—or not principally—to the paper in the *Jubilee Volume*, but to that entitled *Methods of Ascertaining Variations in the rates of Births, Deaths, &c.*, in the December number of that year.

(p. 180.) Dr. Mortara's contributions to the "law of small "numbers" are given in the *Annali de Statistica*, series v. vol. 4, 1912.

(p. 180.) Mr. and Mrs. Peckham's observations are recorded in their article entitled, "Some Senses of Wasps," published by the History Society of Wisconsin, April, 1887.

(p. 180.) The "divergence" from ideal sortition ascertained by Dormoy and Lexis is, of course, not inconsistent with the fulfilment of the normal law by the *average*, with which alone we are concerned in this section (provided a sufficient number of *independent* elements go to the average).

(p. 180.) Contributions to entomological statistics will be found in the *Journal of the Statistical Society*, 1885, *Jubilee Volume*, p. 209; 1896, p. 358, *Statistics of Unprogressive Communities*, and p. 529, *Further Notes*; Biometrika, vol. v, *Statistics of Wasps and Bees;* where the normal character of the dispersion is pointed out.

(p. 181.) For Mr. Latter's application of the theory of observations to cuckoo's eggs, see *Biometrika*, vol. i, and vol. iv, part iii.

4. (p. 182.) "Ejus sunt viri qui quidvis in naturâ fingere, modo "calculi bene cedant, nihil putet" is Bacon's striking description of the Copernican theory (*Descriptio Globi Intellectualis*, ch. 6); not so absurd a view, in the absence of the physical basis supplied by later astronomical science.

5. (p. 185.) The authorities on actuarial mathematics to whom I specially refer are :—

Laplace, *Théorie Analytique des Probabilités*, Book II, chap. ix, Art. 40, and context.

De Morgan, *Theory of Probabilities*, part of the *Encyclopædia Metropolitana* (sect. 146 and context and earlier relevant sections).

E. Blaschke, *Vorlesungen über Mathematische Statistik*, 1906.

G. F. Hardy, *Theory of the Construction of Tables of Mortality*, 1909.

The "method of employing the 'normal' frequency-curve to "represent the series" given by Mr. Hardy at his p. 91 *et seq.* is of a piece with the method described in this *Journal* as "translation"; at least when we "treat z as a parabolic function of t" (Hardy, *loc. cit.*)—a treatment which is, I think, recommended by simplicity and a certain affinity to Taylorian expansion.