UNCERTAIN INFERENCE[1]

By Ronald Aylmer Fisher

Received July 27, 1936 Presented October 14. 1936

At a Tercentenary Celebration we shall do well to look both to the past and to the future. In undertaking to address a mathematical audience, at the present time, on the subject of Uncertain Inference my chief care will naturally be to set forth, at least in outline, those very recent advances which have resolved effectively and conclusively the doubts, confusions, and ambiguities which we can now see clouded the views, and arrested the progress, of those great predecessors to whom our subject owes its gradual development. But just as, behind the Harvard of to-day, the fully developed *alma mater* where future generations of Americans will train their minds, and form their characters, we perceive the struggling college of the seventeenth century, without which this other could not have been what it is; so we can only gain a just perspective of my present topic by recalling the steps, some hesitating, some even false, by which men have come gradually to understand how their reason may be applied to uncertainties, yet applied with logical rigour, and how, in particular, it may be applied to observational facts with all their limitations, their paucity in number and their imperfect precision, and yet draw from them precisely those inferences which the observations warrant.

The first great step was the development of the concept of mathematical probability. Much as this word has since been misapplied, to the writers of the seventeenth and eighteenth centuries its meaning was plain and unequivocal. For centuries, no doubt, *expectations* had been deemed capable of evaluation. Expectations under wills, and expectations from uncompleted trading ventures, had been bought and sold. In games of chance such expectations seemed capable of rigorous calculation. The structure of the game, and its condition when broken off, made it possible to assign to each player a calculable fraction of the amount at stake. This fraction, the ratio of the expectation to the prize which might be won, supplied the essentially new concept of probability. To Thomas Bayes, indeed, this was its definition.

The idea of probability seems to have been an essentially new one in mathematical thought. So far as we know it was unknown to the

[1] A paper delivered at the Tercentenary Conference of Arts and Sciences at Harvard University, September, 1936.

Greek and to the Islamic mathematicians. It was a concept *sui generis,* rather like the notion of temperature in Physics, and it was novel particularly in this, that it brought uncertain consequences within the domain of exact or rigorous thought. If the apparatus used in gambling were true or unbiassed, and were fairly used, the probabilities of the game could be calculated with exactitude. From this point in time there was no excuse for mathematicians to confuse rigour with certainty. In the discussions on probability the uncertainty remained an integral part of the situation, but the concept of probability allowed the nature and extent of this uncertainty to be specified with rigour.

The possibilities of this situation were, of course, only slowly appreciated. Not until our own days has it been realized that the fact that some uncertain inferences are rigorously expressible in terms of probability does not imply that the same concept is capable of providing an exact specification of the nature of uncertainty in all cases. We are now, indeed, familiar with logical situations of a different type which require to be specified in terms of *mathematical likelihood;* and there is, as yet, no assurance that even probability and likelihood together will suffice for the specification of every kind of logical uncertainty which may be profitably discussed.

For centuries, however, it was assumed that if uncertain inferences were to be made they must be made in terms of mathematical probability. It was, I believe, this assumption, more than any other factor, which has led to efforts to define probability in more general, and usually in psychological, terms, and has introduced infinite confusion into the use of this once well defined concept.

Thomas Bayes' paper of 1763 was the first attempt known to us to rationalize the process of inductive reasoning. From time immemorial, of course, men had reasoned inductively; sometimes, no doubt, well, and sometimes badly, but the uncertainty of all such inferences from the particular to the general had seemed to cast a logical doubt on the whole process. By the middle of the eighteenth century, however, experimental science had taken its first strides, and all the learned world was conscious of the effort to enlarge knowledge by experiment, or by carefully planned observation. To such an age the limitations of a purely deductive logic were intolerable. Yet it seemed that mathematicians were willing to admit the cogency only of purely deductive reasoning. From an exact hypothesis, well defined in every detail, they were prepared to reason with precision as to its various particular consequences. But, faced with a finite, though representative, sample of observations, they could make no rigorous statements about the population from which the sample had been drawn.

Bayes perceived the fundamental importance of this problem and framed an axiom, which, if its truth were granted, would suffice to bring this large class of inductive inferences within the domain of the theory of probability; so that, after a sample had been observed, statements about the population could be made, uncertain inferences, indeed, but having the well-defined type of uncertainty characteristic of statements of probability. Bayes' technique in this feat is ingenious. His predecessors had supplied adequate methods, given a well-defined population, for stating the probability that any particular type of sample might result. His problem was: given a particular kind of sample, to state with what probability a particular type of population might have given rise to it. He imagines, in effect, that the possible types of population have themselves been drawn, as samples, from a super-population, and his axiom defines this super-population with exactitude. His problem thus becomes a purely deductive one to which familiar methods were applicable.

There is one point for which Bayes is seldom given enough credit. He had doubts as to the necessary truth of his axiom. So serious were these doubts that he withheld his entire treatise from publication until they should be resolved; and it appears that they never were resolved, for his paper was published by his friends after his death.

That Bayes' axiom was designed to meet a real need is shown by the eagerness and rapidity with which his work became the common property of European mathematicians. Laplace, in particular, incorporated it into the foundations of his "Théorie Analytique des Probabilités," cruelly twisting the definition of probability itself in order to accommodate the doubtful axiom. It is certain that Laplace had no appreciation of Bayes' scientific caution. He says of Bayes, "Et il y est parvenu d'une manière fine et très ingenieuse, quoi qu'un peu embarrassée."

Substantial errors are so rare in the history of mathematics that mathematicians are remarkably unsuspicious of the work of their greater predecessors. The illustrious authority of Laplace thus explains in some sort why Bayes' doctrine in its new dress was embodied without a query into the mathematical teaching of full two generations. To practical thinkers it seemed to meet a practical need. To mathematicians it appeared robed in the authority and in the analytic elegance of Laplace's "Théorie." To De Morgan in 1838 it was still unquestioned gospel, and one of the great steps forward in the history of his subject.

The first serious criticism was developed by Boole in his "Laws of Thought" in 1854. In that extraordinary work Boole anticipated many subsequent attempts to develop a symbolical logic, with particular reference to problems in probability. He recognizes the contradictions and inherent arbitrariness of Bayes' axiom, as developed by Laplace, and quite properly treats it as an attempt to supply by hypothesis something which the data themselves lack. He writes: "These results only illustrate the fact, that when the defect of data is supplied by hypothesis, the solution will, in general, vary with the nature of the hypotheses assumed; so that the question still remains, only more definite in form, whether the principles of the theory of probabilities serve to guide us in the election of such hypotheses. I have already expressed my conviction that they do not." Boole gives fresh reasons and adds:—"Still, it is with diffidence that I express my dissent on these points from mathematicians generally, and more especially from one who, of English writers, has most fully entered into the spirit and the methods of Laplace; and I venture to hope that a question, second to none other in the theory of probabilities in importance, will receive the careful attention which it deserves."

Boole's criticism worked its effect only slowly. In the latter half of the nineteenth century the theory of inverse probability was rejected more decisively by Venn and by Chrystal, but so retentive is the tradition of mathematical teaching that I may myself say that I learned it at school as an integral part of the subject, and for some years saw no reason to question its validity. Mathematicians were averse from abandoning a theory, which often led to plausible conclusions, and, above all, which they had nothing to replace. Its acceptance as orthodox effectively concealed from the majority the fact that, not a mere restatement in more accurate terms, but a fundamentally new approach, was required. As late as 1908 we find Edgeworth, vague but definitely defensive: "I submit that very generally we are justified in assuming an equal distribution of *a priori* probabilities over that tract of the measurable with which we are here concerned."

Why should a mathematician defend a procedure for which he can say no more than that? And why, to take another example, should Karl Pearson, a few years later (1920) put forward what he, and I believe he alone, regarded as a *proof* of the disputed axiom. Such stubborn unwillingness to abandon a false position, to admit ignorance, and to start again, can only be due to mathematicians having so seldom experience of situations which call for an orderly retreat!

The need for an exact procedure of inductive inference was essentially a practical one, and the means for meeting it were being prepared by mathematicians having practical interests beyond those discussed by specialists in the academic theory of probability. Let us turn to Gauss and the foundations of the theory of estimation. As is well known, Gauss, at one time, developed his method of least squares by a formulation identical with that now used in the method of maximum likelihood, but which he justified as taking for the estimate the value of the unknown which had the highest probability. That would be, of course, the *mode* of its frequency distribution, if any such distribution could be assigned to it. Later, as he explained in a letter to Bessel, he let this argument fall into the background, through the conviction that maximizing the probability was less important than minimizing the injurious effects of the actual errors of estimation. To measure these injurious effects by the square of the error he regarded as arbitrary, though convenient.

Modern research has reconciled the two aims discussed by Gauss. If, for any frequency distribution of a variable x,

$$df = y(x)dx,$$

where the frequency density y depends on some unknown parameter θ, we calculate

$$\int \frac{1}{y}\left(\frac{dy}{d\theta}\right)^2 dx,$$

over all possible values of x; then this quantity is invariant for transformations of x, and measures the amount of information which a single observation x contains respecting θ. If x is itself an estimate of θ derived from a sample, the expression measures the intrinsic accuracy of an estimate having the sampling distribution given. For the particular and important case of the normal or Gaussian distribution, the intrinsic accuracy is the invariance, or the reciprocal of the mean square. Error curves of forms other than the Gaussian can then be compared in their precision. When this is done it appears that the estimate obtained by maximizing the likelihood is in general the one for which the intrinsic accuracy is greatest.

A knowledge of the likelihood function thus takes the place of knowledge of a probability distribution in that type of uncertain inference with which the theory of estimation is concerned. This logical situation is one of wide occurence in the discussion of scientific theories of all kinds. It presupposes a hypothesis containing one or more arbitrary parameters. The hypothesis is capable of specifying the probability or frequency of occurrence, of each of the observational facts which can be distinguished. The probabilities of the observable occurrences are then functions of the parameters, and functions of known mathematical form. Only the values of the parameters are unknown. The theory of estimation discusses the advantages of the different methods by which these values can be estimated from an observational record. Clearly, there can be no operation properly termed "estimation," until the parameter to be estimated has been well defined, and this requires that the mathematical form of the distribution shall be given. Nevertheless, we need not close our eyes to the possibility that an even wider type of inductive argument may some day be developed, which shall discuss methods of assigning from the data the functional form of the population. At present it is only important to make clear that no such theory has been established.

The direct assessment of the amount of information supplied by a body of data, the sample of observations, and by a parallel and independent process, of the amount of information extracted from the data, and contained in the estimate, brings to light the important fact that in some special, but specially important cases, these amounts are equal. The estimate exhausts the whole value of the data; once the estimate has been calculated, the remaining facts which the data provide are entirely irrelevant to the value of the unknown parameter. Their distributions are, in fact, independent of the value of this parameter, so that we have the enlightening situation, of which the arithmetic mean of a normal sample, or of a sample from a Poisson Series, are examples, in which, given the value of the first, or Sufficient estimate, the sampling distribution of any alternative estimate is independent of the quantity of which it is designed to indicate the value. All such alternative estimates are therefore worthless. The existence of Sufficient statistics, in the sense defined above, is not only of theoretical interest as a possibility, but of great practical importance, for the cases in which they exist cover many of the forms most used by statisticians in practice.

Theoretically, however, the existence of sufficient statistics is exceptional, dependent as it is from a special functional relationship. When no sufficient statistic exists then no single estimate can contain the whole of the information supplied by the sample. There appears to be an inevitable loss, and, in these cases, the method of maximum likelihood is only preeminent in making this loss as small as is possible. The next task of the theory is to trace the cause of this loss, and to discover in what way it may be made good.

Before turning to this fascinating enquiry, we must recall another development of modern mathematical statistics, in which again the practical requirements of research have moulded the mathematical structure. I refer to the establishment of exact tests of significance. These are now somewhat numerous, and of many kinds, designed to cover the various cases which commonly arise in practice. They are all of quite recent origin, and I may take as typical the test of significance of the mean of a normal sample. This was published in 1908, which year, you may notice, is the same from which I have quoted Edgeworth's defense of inverse probability. Its author was a young man, then unknown, who chose to publish under the now celebrated pseudonym of "Student."

The classical procedure, dating at least from the time of Gauss, for testing the significance of the difference between the observed mean of a normal sample, and zero, or any other value chosen for comparison, is to divide the difference by its standard error, as estimated from the sample. If $\bar{x}$ is the observed mean of n observations, and μ the true mean of the population from which the sample was drawn, then it has long been known that $\bar{x}$ is distributed in different samples in a normal distribution, with its centre at μ, and having a variance one nth of that of the population sampled. If, therefore, we knew the true standard deviation, σ, of this population, we should know that

$$\frac{(x - \mu)\sqrt{n}}{\sigma}$$

was distributed normally with unit variance, and so could assign with exactitude the probability with which any chosen value would be exceeded. In fact, the true value, σ, is not known, but we have in its place an entirely satisfactory estimate, s, defined by,

$$s^2 = \frac{1}{n-1} S(x - \bar{x})^2,$$

where S stands for summation over the sample. This estimate is, in fact, a sufficient one; but it is, none the less, a fact that the value of s arrived at will usually differ more or less from the true value, σ. Consequently, if we substitute s for σ, and calculate

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{s},$$

we are not justified in asserting that t will be distributed in the normal distribution. The originality of "Student's" approach lay in enquiring how in fact the ratio t is distributed, when calculated from samples of n observations. The exact solution is found to be given by the frequency element.

$$df = \frac{\frac{n-2}{2}!}{\frac{n-3}{2}!\sqrt{\pi(n-1)}} \cdot \frac{dt}{\left(1+\frac{t^2}{n-1}\right)^{n/2}},$$

a distribution very different in mathematical character from the Gaussian, though progressively approaching this form as n is indefinitely increased. The distribution is, however, exact, and capable of tabulation for each size of sample possible. It has, indeed, at various times been rather thoroughly tabulated. Consequently, in place of asserting that there is a probability of one chance in forty that

$$\frac{(x-\mu)\sqrt{n}}{\sigma} > 1.960,$$

an assertion which would only be directly useful if σ were known with exactitude, it is equally open to us, if, for example, our mean were based on fifteen observations, to assert that

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{s}$$

has a probability of one in forty of exceeding the value 2.145. This statement is directly useful, for s is not unknown, but is calculable with exactitude from the observations.

Armed with this new tool, it was natural for practical experimenters to take a further logical step of great theoretical importance, namely to use the ratio, e.g., 2.145, appropriate to the level of significance chosen, to multiply this by the standard error of the mean *as estimated*, to add or substract the product to or from the observed mean, and so to obtain working limits for the values of the unknown mean of the population.

In fact, since the distribution of t is known with exactitude, and since t is given by the formula

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{s},$$

which involves, apart from μ, directly calculable quantities only, namely $\bar{x}$ and s, both of which are sufficient statistics, we may infer, without any use of probabilities *a priori*, a frequency distribution for μ which shall correspond with the aggregate of all such statements as that made above, to the effect that the probability that μ is less than $\bar{x} - 2.145\ s/\sqrt{n}$ is exactly one in forty.

It is, at first sight, easy to confuse probability statements respecting unknown parameters, derived by arguments similar to the above, with statements of *inverse* probability. Indeed, attempts have been made to use these arguments, by identifying the results to which they lead with statements of inverse probability, as a means of ascertaining which particular hypothesis of probabilities *a priori* should be adopted in order to lead to equivalent conclusions. In reality the statements with which we are concerned differ materially in logical content from inverse probability statements, and it is to distinguish them from these that we speak of the distribution derived as a *fiducial* frequency distribution, and of the working limits, at any required level of significance, that may be derived from it as the *fiducial limits* at this level. This distinctive terminology is not intended to suggest that fiducial probability is not in the strictest sense a mathematical probability, like any other to which the term ought to be applied, but that it has been derived by a form of argument very different from that introduced by Bayes, and one which was unknown to all the early writers on the theory of probability.

It is a matter of some historical interest to examine why a mode of reasoning so essentially simple, and so cogent, as that outlined above, should have escaped the penetration of the early writers, who include some of the most illustrious of mathematicians. There are two circumstances which may help to make clear this difficulty. The distributions studied by the early writers were nearly all discontinuous distributions, distributions in particular, of which the variates are frequencies. When applied to these the fiducial type of argument does not lead us to an exact frequency distribution of the unknown parameters, but only to a series of inequalities which add little in intelligibility to the tests of significance from which they may be derived. The neglect of the frequency distributions of continuous variates, until they were forced on the notice of mathematicians by the requirements of the quantitative sciences, is, I believe, one potent reason why early writers on probability were not led to use arguments of the fiducial type. For such arguments to be fruitful, moreover, the distributions considered must be not only continuous, but mathematically exact. Exact solutions of all the more important and immediate problems were possible by analytic methods certainly within the capacity of the greater writers of the last 150 years. That their existence remained for so long unknown, can only, I believe, be explained by the absence of any steady conviction that inferences involving an element of uncertainty *deserve* anything better than rough and approximate discussion.

Two subsidiary circumstances, also, have in our own time greatly facilitated the new approach, and have, indeed, made its development inevitable. One is the convenient practice of tabulating the distributions required, at a series of definite levels of significance, i.e., of expressing the variate in terms of the probability, in place of regarding the probability as a function of the variate. The second circumstance is the abandonment of the inverse type of argument, since, so long as statements of inverse probability were held to be the aim, the possibility of making inferences of fiducial probability, which differ from the former in logical content, was very naturally overlooked.

There is one peculiarity of uncertain inference which often presents a difficulty to mathematicians trained only in the technique of rigorous deductive argument, namely, that our conclusions are arbitrary, and therefore invalid, unless all the data, exhaustively, are taken into account. In rigorous deductive reasoning we may make any selection from the data, and any certain conclusions which may be deduced from this selection will be valid, whatever additional data we may have at our disposal. The more philosophic writers on probability, however, such as Venn, have emphasized the fact that conclusions in this field are relative, not only to what is known, but also to what is undetermined. Venn, for example, contrasts the conclusions to be drawn from such items of information as that the death-rate of Englishmen is higher in Madeira than in England, and that the death rate of tuberculous patients is higher in England than in Madeira. The probable effect of a change of residence is different for the contrasted cases of a man chosen at random from the English population, as against one chosen at random from the tuberculous patients of that country. The additional datum that the individual chosen is tuberculous must not be ignored in drawing inferences from the remaining data.

This peculiarity appears to be characteristic of uncertain inference in general. It is certainly as important in inductive reasoning from observational data, as in the purely deductive inferences of the classical theory of probability. Every statistician is conscious that if he were to allow himself to make an arbitrary selection among the observational material available, then the most orthodox operations of his craft could be made to lead to almost any desired conclusion. The political principle that "Anything can be proved by statistics" thus enshrines a subtle truth, which requires to be the more carefully borne in mind, the more we rely on mathematical techniques developed with only *certain* inferences in view.

This consideration is vital to the fiducial type of argument, which purports to infer exact statements of the probabilities that unknown hypothetical quantities, or that future observations, shall lie within assigned limits, on the basis of a body of observational experience. No such process could be justified unless the relevant information latent in this experience were exhaustively mobilized and incorporated in our inference.

We may now appreciate the necessity of the condition I mentioned, in connection with "Student's" test of significance, for the mean of a normal sample; namely, that the quantities $\bar{x}$ and s, which, together with the unknown parameter μ, appear in the expression for t, should be Sufficient estimates of the mean and standard deviation of the population sampled. For this is a guarantee that they have, together, tapped all the information the sample has to give respecting the nature of the population. If alternative estimates had been used; if, for example, we had found the median in place of the arithmetic mean, $\bar{x}$, or, if we had used Peter's Formula, based on the mean deviation, in place of Bessel's formula, based on the mean square, we might have derived an entirely valid test of significance; that is to say we could have found a quantity t', with a distribution exactly known for samples of a given size, and expressible, like t, in terms of the unknown parameter, together with directly calculable quantities only. But, if we had gone further, and, substituting for t' in terms of μ, had derived a fiducial distribution of the unknown parameter, the distribution we should obtain would be based only on that part of the information available, which our special estimates of the mean and standard deviation had conserved. The distribution obtained would differ from that found by using the sufficient estimates, and the probability statements which it embodies would be discrepant. Without the requirement that the information available should be exhausted, a host of discrepant inferences would appear equally admissible, each dependent from the personal choice of the statistician, through his choice of the method of estimation to be employed.

When Sufficient estimation is possible, there is here no problem;

but the exhaustive treatment of the cases in which no Sufficient estimate exists is now seen to be an urgent requirement. This at present is in the interesting stage of being possible sometimes, though, so far as we know, not always. I have spoken of the Sufficient estimates as containing in themselves the whole of the information provided by the data. This is not strictly accurate. There is always one piece of additional, or ancillary, information which we require, in conjunction with even a Sufficient estimate, before this can be utilized. That piece of information is the size of the sample; or, in general, the extent of the observational record. We always need to know this in order to know how reliable our estimate is. Instead of taking the size of the sample for granted, and saying that the peculiarity of the cases where sufficient estimation is possible lies in the fact that the estimate then contains all the further information required, we might equally well have inverted our statement; and, taking the estimate of maximum likelihood for granted, have said that the peculiarity of these cases was that, in addition, nothing more than the size of the sample was needed for its complete interpretation. This reversed aspect of the problem is the more fruitful of the two, once we have satisfied ourselves that, when information is lost, this loss is minimized by using the estimate of maximum likelihood. The cases in which Sufficient estimation is impossible are those in which, in utilizing this estimate, other ancillary information is required from the sample beyond the mere number of observations which compose it. The function which this ancillary information is required to perform is to distinguish among samples of the same size those from which more or less accurate estimates can be made; or, in general, to distinguish among samples having different likelihood functions, even though they may be maximized at the same value. Ancillary information never modifies the value of our estimate; it determines its precision.

The procedure of this kind the most general possible would be, from a sample of n observations, to specify (a) the estimate, or set of estimates of the unknown parameters, having the greatest likelihood; and (b) a set of functionally independent ancillary statistics, sufficient in conjunction with (a) to allow the observations to be reconstructed in their entirety, and having the additional property that these ancillary quantities shall be all distributed in samples in distributions independent of the unknown parameters. It is easy to see that this can be done in certain simple cases. For example, if μ is the only unknown parameter in a frequency distribution specified by the differential element

$$df = \phi(x - \mu)dx,$$

then the differences between successive observations, when these are arranged in order of magnitude, supply $n - 1$ functionally independent quantities, calculable from the sample, the sampling distribution of each of which is evidently independent of μ. We may, therefore, regard such a set of differences as specifying the configuration of the sample, and, in interpreting our estimate, may take as its sampling distribution that appropriate to only those samples which have the actual configuration observed.

Here, then, is a second group of solutions, by which estimation may be made exhaustive, like the Sufficient statistics in depending from a special functional relationship, like them, also, in resolving a wide class of the problems arising in practice. And my final word on this topic is a query, the answer to which so far is unknown, and which is, therefore, at present a challenge to our mathematical intuition. May I put the problem in this form?:—

> The agricultural land of a pre-dynastic Egyptian village is of unequal fertility. Given the height to which the Nile will rise, the fertility of every portion of it is known with exactitude, but the height of the flood affects different parts of the territory unequally. It is required to divide the area, between the several households of the village, so that the yields of the lots assigned to each shall be in pre-determined proportion, whatever may be the height to which the river rises.

If this problem is capable of a general solution, then it is possible in general to recognize something corresponding with the configuration of the sample in the simple case discussed above, and one of the primary problems of uncertain inference will have reached its complete solution. If not, there must remain some further puzzles to unravel.

REFERENCES

BAYES, T.
1763. An Essay towards solving a problem in the Doctrine of chances. Phil. Trans. 53, p. 370.
BOOLE, G.
1854. Laws of thought, Cambridge. p. 375.
DE MORGAN, A.
1838. An Essay on Probabilities and on their application to life contingencies and Insurance Offices. Preface vi.
EDGEWORTH, F. L.
1908. On the probable errors of frequency constants J.R.S.S. 71, p. 387.
FISHER, R. A.
1925. Theory of Statistical Estimation. Proc. Camb. Phil. Soc. 22, pp. 700–725.
1933. Two new properties of mathematical likelihood. Proc. Roy. Soc. A, 144, pp. 285–307.
GAUSS, K. F.
1809. Theoria Motus Corporum Coelestium. Hamburg. p. 210.
1839. Briefwechsel zwischen Gauss und Bessel. Leipzig (1880).
LAPLACE, P. S.
1814. Théorie analytique des probabilités. Paris. Second Edition, p. ciii.
PEARSON, K.
1920. The fundamental problem of practical statistics. Biometrika xiii. pp. 1–16.
'STUDENT'
1908. The probable error of a mean. Biometrika. 6, pp. 1–25.
VENN, J. A.
1866. The Logic of Chance. Cambridge.