SCIENCE

VOL. LXIII MARCH 19, 1926 No. 1629

CONTENTS

SCIENCE: A Weekly Journal devoted to the Advancement of Science, edited by J. McKeen Cattell and published every Friday by

THE SCIENCE PRESS

Lancaster, Pa. Garrison, N. Y.

New York City: Grand Central Terminal.

Annual Subscription, $6.00. Single Copies, 15 Cts.

SCIENCE is the official organ of the American Association for the Advancement of Science. Information regarding membership in the Association may be secured from the office of the permanent secretary, in the Smithsonian Institution Building, Washington, D. C.

Entered as second-class matter July 18, 1923, at the Post Office at Lancaster, Pa., under the Act of March 8, 1879.

STATISTICAL INFERENCE[1]

(1) IT was two years ago almost to a day that I was last here at the Johns Hopkins University for a brief visit. It was a pleasure to come then and now to be here once more, only I miss that kind soul, Dr. Halsted, and so must you, very much.

We are both engaged in a new undertaking—the teaching in a large way of the sciences basal to the public health. In a preliminary, a small, but I think very useful way we at Harvard, in cooperation with the Massachusetts Institute of Technology, started systematic instruction, I think, before you did, but to you undoubtedly will be given the credit, and quite properly, of making the first large start under your perennial leader, Dr. Welch. We shall keep step with you as well as we can.

And may I say that it appears to me to be of very good omen that it is Hopkins and Harvard who are working together along this new line. These two institutions have been leaders in American university education, leaders in their insistence on the university point of view, on scientific investigation as the normal university life and as the necessary precursor and accompaniment of sound applications and effective teaching of knowledge, whether old or new. There is in the pursuit of the public health unlimited opportunity for the satisfaction of a natural emotional desire to aid our fellow man, but the aid will be most effective, most permanently effective only if it is forwarded on patient scientific study. The traditions of our respective institutions augur well for the proper grounding of our work.

(2) When I looked about to choose a subject for discussion with you to-day it seemed to me that we might do well to consider a while together the question of statistical inference. Advisedly I say consider the question of inference, for I doubt whether the matter is yet sufficiently settled so that it has passed beyond the stage of query. Moreover, it seemed as though I should be likely to find here an audience not unwonted to reflect on this problem. Forty years ago you had Charles S. Peirce, a leader in the study of logic and one who did not shun the difficulties of probability and of statistical inference. You also drew on Simon Newcomb, who pondered much on the intricacies of the reduction of observations and was so astute in his own reductions that despite the advances of physics and astronomy some of his determinations have not been bettered to this

[1] DeLamar Lecture, School of Hygiene and Public Health, Johns Hopkins University, February 25, 1924.

day. And you have now the most active and scientific department of biometry and vital statistics in America and perhaps in the world, for I am not sure that we have to make any exception of the laboratory of Karl Pearson, the founder of this science.

Indeed, although I am thus assured of an especially informed audience, and because I am, I feel somewhat as Walter Nernst did on beginning his lectures on thermodynamics at Yale some fifteen years ago when after a reference to the work of Willard Gibbs he remarked that he was very keen to come there to expound this subject, meaning by keen, I suppose, *kühn*.

(3) Let us start with ordinary syllogistic deduction and with a simple example. The syllogism consists of three statements, two of which are premises and the third conclusion.

First premise:	All men are mortal.
Second premise:	Socrates is a man.
Conclusion:	Socrates is mortal.

The conclusion follows ineluctably from the premises. It follows whether the premises be right or wrong. Even if all men were not mortal and Socrates were not a man, the conclusion would follow. Thus

First premise:	All men are oysters.
Second premise:	Socrates is a man.
Conclusion:	Socrates is an oyster.

Or again:

First premise:	All men are mortal.
Second premise:	Cerberus is a man.
Conclusion:	Cerberus is mortal.

It is the very essence of syllogistic deduction that it is sure. One may say that we have here a form or formula of deduction. The method of thought can not lead one into error. Any sources of error must be found elsewhere than in the deduction; they must lie in the premises.

Now it should be clear that it is advantageous to separate our liabilities to error. Perhaps we can never be entirely certain of both our premises and so we may not be sure of the truth of our conclusion. Yet we are sure of its truth if the premises are true. We may introduce explicitly into the syllogistic formula our possible doubts as to the premises by making them hypothetical: *If* all men are mortal, and *if* Socrates is a man, then Socrates must be mortal.

(4) Pure or abstract mathematics shares with logic this distinction of being unerring in its deductions.[2] I do not mean that a mathematician may not blunder —we are all frail creatures. But the mathematician is concerned with forms of drawing conclusions and with the precision of those forms, not with the existential content of the premises or the conclusions. There was a long discussion in geometry as to the truth of Euclid's parallel postulate. The termination came when the non-Euclidean geometries were advanced about a century ago, geometries in which the parallel postulate was replaced by a variant and which were none the less logically consistent within themselves. There is no mathematical sense to the question: Is the parallel postulate true? Whatever sense there is is physical, *i.e.*: Do we have the most satisfactory geometry for the codification of natural laws when we use the Euclidean system?

The advantage of mathematics to the person who would apply it is precisely in the certainty of its analysis; it does not guarantee conclusions but serves to separate the processes of drawing conclusions from the difficulties inherent in the acceptance or rejection for practical purposes of the premises. The whole business of pure mathematics is to be self-consistent; truth and falseness do not enter into its field. Herein, too, lies the disadvantage of mathematics in the sciences. For in order to apply any formula or conclusion of mathematics it is necessary first to convince oneself that the formula or conclusion is in truth applicable to the case in hand. Logically we should have to verify the truth for the case in hand of every premise in the whole argument by which the formula had been derived—we could trust the professional reputation of the mathematician as to the consistency of his deduction, but we have ourselves the responsibility of verifying the various premises.

(5) Practically such verification is impossible for two reasons. First, the chain of argument is often so long and often so unintelligible as to wear us out; but there is another and more serious difficulty. That which I have stated as the manner of mathematics is really but an ideal manner, a party etiquette, not an everyday habit. Few mathematicians actually state all their premises in any field of work, and in many fields of mathematical prowess those premises are not even precisely known by anybody. They are to-day known for arithmetic of several kinds, algebra of several kinds, geometry of several kinds. I doubt if they are known for mechanics or any branch of physics. How forsooth is one who would apply a formula to verify premises which no one has yet known how to state?

What we actually do is to use the formula and trust to its being right for our case; but the responsibility is ours. It is we who assert that for our case the formula should give the right result. Unfortunately there are many persons who do not seem to realize their responsibility in this matter. They seem for some reason to believe that a mathematical formula is eternally true. Their attitude is Shamanistic. They go through with magic propitiatory rites,

[2] See, for example, B. Russell, "The Principles of Mathematics," 1903, Chap. I.

idolatrous of mathematics, ignorant of what it can and can not do for them. And I am not quite sure that the high priests of this pure and undefiled science do not somewhat aid and abet the idolatry.

(6) Many of us do not stop to think what very simple mathematical formulas may break down in practice. Consider, for example, $100 = 2 \times 50$. Is this true? It is of course neither true nor false in general. How about a particular case? If I have fifty apples in each of two baskets, I have one hundred apples altogether. The result is true. The enumeration of apples is something to which the theory of cardinal number is applicable. Consider another case. On one day I notice that the temperature is 100° F. in the shade, and on another day I see that it is 50°. Is the temperature twice as much the first day as it is the second day. It is not. The formula is false this time. Or again I mark one student 100 and another 50. Does the first know precisely twice as much as the second? The measurement of temperature and the estimation of knowledge do not lend themselves to the laws of cardinal number. It is no fault of arithmetic that persons who multiply temperatures or marks may obtain bizarre results.[3]

A long acquaintance both with mathematics and with a reasonably wide variety of applications thereof has made me somewhat skeptical of formulas, and I have perhaps unwisely gone so far as to state in print that I do not believe formulas. It is unwise merely because it is shocking to a prevailing Shamanism. When in Rome one really should do as the Romans do, and if one happens to reside among the Ostiaks or Samoyeds of Siberia it is perhaps still wiser to conform. With my audience to-day I am however quite safe; for no university has stood more resolutely than the Johns Hopkins for straight, independent thinking on the facts as they are. I shall therefore pursue my course a bit farther.

(7) Nearly a year ago I was invited to speak at Yale on the subject of "The Statistical Significance of Experimental Data." My communication was written somewhat in the lighter vein and it was printed in SCIENCE with disastrous consequences.[4] If Dr. Cattell had been the good friend I have liked to think him, he would have returned the manuscript with kind words of caution born of his profound knowledge of psychology and proper to a dignitary about to be elevated to the pontificate of organized science in America. He might have pointed out that SCIENCE was not a funny paper and have respectfully urged upon me as a more appropriate medium of publication either *Life* or *Punch*. To be sure, that might have given me the opportunity to retort that I did not consider *Life's* labored anti-vaccination or anti-vivisection material half so humorous as the anti-Bryan round-robin on evolution recently printed in SCIENCE. Dr. Cattell is too wily to lay himself open to any such thrust, and yet too kindly to suggest, as have some good friends, that my address was fit only for a broad Latin comic *a l'assiette au beurre!*

What were some of the terrible things I said? First and last my main contention was that it took real brains to be a good statistician. This has offended both parties to the context. Let me recant. We are in a mechanistic, nay a mechanical age, and America is the country par excellence of mechanical devices; it is not the illumined and generalizing minds of Farr and Galton that guide us on our way, but the cheerful chatter of the Hollerith tabulator and the Millionair calculator. Then, too, it appears that I slipped in the unfortunate remark that the mature "hunch" of a genius is better than many a scientific demonstration. I must apologize for a low form of hero-worship; it should not be tolerated in a democracy; besides, we are in an industrial era and it has been maintained statistically, has it not, that the best workmen are the morons? Again I recant. But this is aside from my subject to-day.

(8) There was, however, one statement, perhaps the most distressing, which bears on our discussion. I quoted the formula or rule that "the probable error in the mean or average of a number of observations is equal to the probable error of the individual observations divided by the square root of their number." This is the sacred cow of quite a hierarchy of the statistical fraternity. Pray, note that I do not call it the sacred "bull." Experience is a dear teacher, and even I am learning discretion. Later I remarked, just offhand and a bit flippantly, that I did not believe the rule. To be sure, I had shown that it did not work out right in the case in hand. But none the less every brother of the Shamanistic order for the Preservation of the Sanctity of Ancient Icons rose to protect dear bossy from my rough dehorning hands. And just here is where I do not yet recant. I do not believe the formula or rule for the computation of the error in a mean or average from the mean error of the observations. More than that, I do not believe any formula, not even $100 = 2 \times 50$, and particularly when I have already verified for a case in hand that it does not work.

What was my process of showing that the formula was inapplicable? I should have liked to analyze the proof of the rule into all its various steps so that I might examine the many premises to ascertain which ones were violated. This can not be done. Probability and statistics are not yet on a sound logical basis like algebra and geometry. We do not know

[3] We need a finite ordinal arithmetic for our marking system.

[4] SCIENCE, August 10, 1923, Vol. 58, No. 1493, pp. 93–100.

precisely what our premises really are. The best set of postulates or axioms we have is due to Keynes and may be found in Part II of his "Treatise on Probability." We owe him a great deal for this analysis; but this sort of rational founding of a branch of science is very difficult and is the work not of one man; it requires the accumulation of efforts of many students spread over much time.

As is natural in cases where a complete postulational basis has not yet been established and accepted, the proof of the theorem or formula is itself not very clear or convincing, it is an argument rather than a proof. In such cases it is necessary to turn to the end result and see whether in a particular case it seems to work. This is what I did and found that, assuming the formula to be true and applicable, the probability of getting the experimental results that were found was itself very much less than the probability of the conclusion it was desired to establish from the results. Such is the method of *reductio ad absurdum.* We ordinarily accept a statistical thesis as proved if the chances in its favor appear to be better than say twenty to one; but if by using entirely similar calculations on the same data we can show that our data themselves have not one chance in a thousand of existence, we can make no inference other than that the formulas are inapplicable to the material in hand.

(9) There is much of mystery in the situation. We have admirable proofs that no matter what the law of distribution of our observations, the mean or average values of reasonable numbers of the observations are distributed upon the Gaussian or normal curve or error. Edgeworth, who is as clear thinking an investigator as we have in statistics, has worked out simple cases theoretically and has verified them experimentally and has shown how very few need be the number of observations which we average before the mean values do actually distribute themselves very satisfactorily on the standard curve with predetermined parameter.[5] I have worked out such cases myself. There is no doubt, I think, but theoretically and in a great many practical cases the rule that the probable error in the mean is equal to the probable error of the observations divided by the square root of their number is clearly indicated as applicable just as is the rule $100 = 2 \times 50$. Yet I believe there are exceptions, probably broad categories of exceptions, and that the last word has by no means been said on the subject. Probably there are in the theoretical proofs subtle premises, perhaps only tacitly assumed, which if they could be dragged clearly into the light would give us a better understanding of the matter.

In my article in Science I was dealing with short runs of data such as often occur in experimental work; there were only seven to twelve observations in each series. Inasmuch as many of the proofs of the rule for the error in the mean depend on the assumption that the observations are numerous it might be thought that runs so short as seven to twelve would be adequate reason for not expecting the formula to apply, for excusing us from believing in it in such cases. But Edgeworth's proofs and experiments and some similar investigations of my own seem to indicate a very rapid convergence to regularity in the distribution of the means so that ten observations should be enough to average. There is some mystery lurking here.

(10) Suppose we turn our attention from the short runs of experimental work to the long runs we use in vital statistics. Should we believe the rule for very long runs, for large collections of data? Let us perform an experiment, but only in imagination. Suppose I draw upon the blackboard a line about eight inches long, and let us set about determining the length of that line by the simple process of each one estimating the length. If we could find one hundred compliant individuals and get an independent estimate from each, we could then add all the one hundred values together and divide by one hundred to find the mean estimated length of the line. We may well assume that on the average the length can be estimated to about an inch. Of course some of you would do better than that. But allowing an error in the individual estimate of one inch we should by the formula have for the error in the mean of one hundred estimates only one inch divided by the square root of one hundred, which is ten. The mean should therefore be accurate to the tenth of an inch. We can all believe that this is reasonable.

Imagine, however, that we desire a greater accuracy. We have then merely to go out into the highways and byways and hedges to round up let us say ten thousand interested individuals who could give us estimates. The mean of the ten thousand estimates should be accurate to the hundredth of an inch. And so if we get a million estimates we shall have a mean good to a thousandth of an inch. The process could be continued further to insure a greater accuracy—

[5] Edgeworth, Introductory description (especially pp. vi, vii) to the "Representation of Statistics by Mathematical Formulae," 1900, printed for private circulation and being in the main matter reprinted from *J. Roy. Statist. Soc.*, March, 1900 (among the Miscellanea), Vol. 61, 1898, pp. 670 *ff.*, Vol. 62, p. 125 *ff.*, p. 373 *ff.*, p. 534 *ff.* Reference is made here to Burton, *Phil. Mag.*, Vol. 28, 1889, p. 483, and Edgeworth, *Phil. Mag.*, Vol. 16, 1883, p. 301, *J. Roy. Statist. Soc.*, Vol. 51, 1888, p. 116. One may also consult Bowley, "Elements of Statistics," p. 289, 421.

at least, it could if we still believe the formula. But you may recall that on principle we should not believe formulas; that truth and falseness come in only when the formulas are applied, and that the responsibility lies with those who use the formulas. Is it inherently reasonable that by getting one million fair estimates of the length of the line we should have in their average a value for the length good to one thousandth of an inch? My own faith in the formula does not extend so far.

It would of course be a simple matter to determine by precise physical measurements the length of a line to much greater accuracy than this. Fine mechanical processes require gauging to half a thousandth of an inch or better. But a chalk line on a blackboard is a very indefinite physical object. Did you ever look at one under a reasonable magnification and see that it has neither beginning nor end; that its appearance of continuity and definiteness are due precisely to the fact that we do not examine it under power? How are you by any mathematical formula to determine the length of a line more accurately than the line exists?[6]

(11) Let me leave it as a query. I came not to resolve the difficulties that have their lairs in various corners of this vast labyrinth of probability and statistics, but merely to discuss the matter, to ask questions.

You may think I have hardly a fair experiment when I ask you to estimate the length of the line. Why not measure it, why use such a crude illustration? But note that I was willing to assume that individually you could estimate to an error of around 10 per cent.—one inch in eight. This is not bad accuracy with which to start. I fear my colleague, Dr. Richard Cabot, might not be so generous to the clinician in diagnosis.

Physicists, who are used to precise quantitative measurements under excellent control, are not too sure that the error in a mean can be had by the simple rule. Simon Newcomb, discussing impersonally his determination of the velocity of light, said: "So far as could be determined from the discordance of the separate measures the mean error of Newcomb's result would be less than $\pm$ 10 km. But making allowance for the various sources of systematic error the actual probable error was estimated at $\pm$ 30 km." He so far lacked confidence in the rule that he allowed a factor of safety of three in stating his precision. Perhaps Newcomb was conservative. From a recent study on which I have collaborated with Dr. W. J. Luyten, of our observatory, it appears that the good-to-best modern photographic stellar parallaxes need in their computed probable errors[7] a factor of safety of only one and one fourth to one and one half.

(12) This sort of knowledge is of course familiar to you. Probably Newcomb himself set it forth here some twoscore years ago. I was taught it in college more than twenty-five years since by Robert W. Willson, and a good chance indeed I had to learn the difference between probable errors computed by rote and rule, from the discordance of the observations, and the presumptive actual probable error. We students had an old meridian circle. It had been a good one in its time but had been neglected and maltreated and then reconditioned. You could find a new source of error in the instrument almost every day to increase your probable error until you thought you were done —and then a hard freeze would come along and heave one of the piers on which it rested. Yet somehow I doubt if with all its age and vagaries this former instrument of precision was less reliable than a modern death certificate. Why is it that we find need in precise physical work for a factor of safety in our probable errors to convert them from hypothetical pure mathematics to presumptive physical facts, while in economic and vital statistics we appear to lay no stress on them? Is it because the worse our data, the sounder our conclusions? It may be.

The question antedates my college days; antedates Newcomb and Peirce. Two centuries and a score of years ago, before Lexis, Quetelet, Gauss or Laplace, it was discussed in general terms between Jacques Bernoulli and Leibnitz. The former was inclined apparently to trust his figures further than the latter. Leibnitz wrote that the estimation of probabilities was very useful but that in affairs of state and many others it was not so much refinement of calculation that counted as an accurate consideration of all the circumstances; that there seemed to him to be an inherent difficulty in the cleancut determination of empirical probabilities because nature, though she had her habits, due to recurrence of causes, did not follow them except in a general way; that new diseases arose and whatever observations you should make on deaths would not thereby constrain nature.

If I should tremble to have the clever eye of Bernoulli fall upon my poor contribution in SCIENCE, I should nevertheless not hesitate with a profound bow of respect to lay a copy on the desk of Leibnitz.

And after all perhaps there is less difference in points of view than the heat of argument, the flush of individual exertion and the glowing dialectic of ill-

[6] It may be of comfort to some to say that we are not determining the length of the line but merely an estimate of its length and that the estimate may well be more precise than the length estimated.

[7] Newcomb, Encyclop. Brit., Vol. 11, p. 625, 11th ed. Wilson and Luyten, Proc. Nat. Acad. Sci., Vol. 10, April, 1924.

defined technical terms might lead us to imagine. Leibnitz and Bernoulli, Pearson and Keynes might practically agree in their practical inferences from the same data. There may be here not so much a difference of science and statistics as of temperament, of endocrine function.

(13) From classical logic we learn that the premises "All men are mortal" and "Socrates is a man" lead ineluctably to the conclusion "Socrates is mortal." We know also that the premises "Some men are liars" and "Munchausen is a man" lead to no conclusion at all. It is here that Charles Peirce[8] comes to his introduction of *probable deduction.* From the premises "Ninety-nine Cretans in a hundred are liars" and "Epimenides is a Cretan" he draws the probable deduction "There are ninety-nine chances in a hundred that Epimenides is a liar."

He does not fail to note that this is a very different sort of conclusion. You do not assert that Epimenides is a liar because he is a Cretan and 99 per cent. of Cretans are liars; but merely that the probability that Epimenides is a liar is ninety-nine in a hundred. Where in deductive logic the conclusion makes a definite attribution of a predicate to the subject, in probable deduction the conclusion makes a statement about probability. It is not true that Epimenides is 99 per cent. a liar. He may be a liar or he may not be; what we assert is that he belongs to a class of individuals ninety-nine out of each one hundred of whom are liars and we abbreviate this to the assertion that the chances he is himself a liar are ninety-nine in a hundred. Peirce emphasizes the fact that we must obtain our premises in good faith. If we happen to know Epimenides personally that may greatly alter our conclusion—even to the affirmation that he is no liar at all.

(14) He next goes on to *statistical deduction.* The premises here are "The proportion p of X's are Y's" and "S is a numerous set taken at random from among the X's," with the conclusion that "probably and approximately the proportion p of the S's are Y's." The inference is not sure but probable, not exact but approximate. The statement of the second premise brings in the term random—a set taken at random from a specified class. He remarks that there is no way of insuring randomness, *i.e.,* fairness or lack of bias in drawing the set of S's from the X's, except by faith in the honesty and open-mindedness of those that make the selection. How would you select one thousand dwellings at random within the city limits of Baltimore? Probably you would not trust your judgment but would resort to a lottery, inscribing on similar slips of paper the street and number of each dwelling, mixing the slips thoroughly and having a blindfolded person select one thousand.

One thing, however, is sure about statistical deduction, namely, that if you persist in repeating your drawings of samples you will ultimately vindicate the conclusion. Indeed the very notion of a random selection is that if the selection is indefinitely repeated you will select each element of the original set the same proportionate number of times. Thus if there are one hundred thousand X's and you select as the set S a particular one thousand not once but very many times at random, you will get any special X in the set S about once in a hundred times. With any specified set which you are sampling this indefinite continuation of the process is conceivable and the result is sure. You verify[9] in the long run that for a random set S the proportion p of elements that are Y is the same as in the class X from which S is selected. And this is precisely what we mean by the statement that in a particular drawing of one set S the proportion is probably approximately equal to p.

(15) When we pass from the larger set to the smaller we deduce and in the long run verify. Suppose we try to pass the other way, ascending from the smaller sample by induction to the larger universe from which it is drawn. The syllogism of *statistical induction* would read as follows: A large random set S is selected from the X's—which is a much larger and in general unknown class. Of the S's the proportion p are Y's. Hence probably and approximately the same proportion p of all the X's are Y's. Clearly it is this process of induction which is used so much in practice. One looks over a considerable number of apples in a barrel, finds a small proportion of poor ones, and buys the barrel with the confidence that the proportion of poor apples is similarly small in the whole lot. Of course in the good old times when the New England farmer "deaconed" his apples you shouldn't sample merely the top layers—and you didn't. Again the fairness of the sample is taken for granted.

In the case of deduction you know the proportion in the whole set and by repeated sampling verify it in the samples. For induction you know the proportion in the sample, and, as Peirce points out, you do not verify it but *modify* it by repeated sampling. This is truly a very great difference. And particularly so inasmuch as the general universe is unknown

[8] C. S. Peirce, "Theory of Probable Inference" in "Studies in Logic by Members of the Johns Hopkins University," Little, Brown and Co., Boston, 1883, pp. 126–203, especially pp. 127, 134, 137, 152, 154, 175. Also the "Probability of Induction," *Pop. Sci. Mon.,* April, 1878, reprinted in "Chance, Love and Logic," Harcourt, Brace and Co., 1923, especially p. 100.

[9] If the verification does not come, within reasonable limits, in due time you have to conclude that your sample is not truly random.

and may be variable. In practice we take a sample S of what we believe to be some large class of X's and find a proportion p of Y's. We take another sample S′ and find the proportion p′ of Y's—and so on to as many samples as we choose to make. When we deal with deduction we can continue the process indefinitely in imagination.

Can we so continue it in the case of induction? Clearly not; for the class of X's which we are sampling is changing like everything else in nature, and we have no means of determining offhand[10] how much of the variation in the successive proportions p, p′, p″ . . . found in the different samples represents fluctuations in the universe itself or fluctuations in or departures from real randomness of our sampling. We can not resort to our lottery, inscribing each element of our unknown universe X on slips of paper, and drawing samples at random. Induction is a much more dangerous process than deduction and one can not be so sure of it. There are limitations, as Leibnitz pointed out to Bernoulli in regard to the perfection, or, as Peirce says, modification of our estimates from sampling as a method of induction.

This may be illustrated by Peirce's own example. He remarks in the returns of the census of 1870 there were among native white infants under one year of age 478,774 males to 463,320 females, a proportion of 0.5082 of males; but among colored children there were 75,985 males to 76,639 females, a proportion of 0.4977 males. He infers that generally there is a larger proportion of male infants among whites than among negroes. That is, we propose to regard the returns of the 1870 census as representing a fair sample as between whites and negroes with respect to the ratio of the sexes of children.

Now it is of the essence of a sample to be a small fraction of the whole universe, and as we have enumerated all the white and colored children in 1870 in these United States we must for our induction consider a universe of all white and all negro infants at all times or in all places or both—or at the very least at many times and in many places. This is the kind of situation that meets us constantly in vital statistics. It is very different from drawing as a sample a cupful of beans out of a bag well shaken up; it is more like picking out as a sample one bag of beans from among a hypothetical class of bags of beans. Still, although recognizing the difference, I do not object to the inference that there are in general relatively more males than females born to whites than to negroes—provided that all the pertinent circumstances have been considered, and I believe that Peirce sufficiently emphasized the necessity of considering them as a matter of common honesty and sound logic.

(16) Later, however, he goes on to what I fear Leibnitz would consider refinement of calculation and states the conclusion that the discrepancy between the proportions 0.5082 for whites and 0.4977 for blacks is so large that the result would happen only once out of ten billion censuses in the long run. Now is there any significance at all to such a statement, and, if so, what is the significance? Clearly it can have nothing to do with vital statistics—with natural phenomena. The best estimate we have of the lapse of time since the earth's crust solidified is about one billion years. I have not seen any estimate of the future duration of the earth as a habitation for either whites or blacks that is longer than that. The human race has probably not existed a million years and may not exist for many millions. Even with the maximum increase of interest in vital statistics which we can stimulate and the maximum multiplication of governmental activities which we can imagine and deplore, ten billion censuses seems a large order. This is, however, but a drop in the bucket, for we are talking of results which would happen only once in ten billion censuses in the long run. Is not the calculation refined to the extreme? Would not a superficial examination of a few of the circumstances surrounding this problem have led us to the conclusion that the calculation could but waste our time and maybe mislead us into thinking that it meant something very different from what it possibly can mean.

Where do we stand logically when we compare Peirce's fundamental logical justification for induction, namely, the self-correcting or modifying property of indefinite repetition of sampling on the one hand, with such a calculation as he makes here on the other hand? Do we not need first to conclude that he is using sampling in two different senses? In modifying an empirically estimated probability by further sampling we are taking new censuses—new in time, new in place, different in numbers of population, etc. But we are taking real censuses and from them we shall get a variety of determinations of the empirical probabilities which we may interpret as best we can. In the second case we are not taking ten billion censuses nor in fact any except the one we have taken; we are doing a piece of pure mathematics which may or may not correspond even with gross approximation to what would happen if repeated censuses were taken.

Indeed the second proposition is not a piece of induction at all but of deduction. The statement is something like this. We imagine two hypothetical universes; one of many millions of white infants of which half are males, another of millions of black infants of which also half are males. By an impar-

[10] There are refined methods of making some such determination, but they are often neglected and of course can not be applied at all when we base our induction on a single sample, as is often done.

tial lottery we draw about one million whites from one universe and about 150,000 blacks from the other, and we call these drawings a census (which is perhaps too graphic an expression). Then we should have but one chance in ten billion of finding so great a difference between the proportions of males and females in the white and negro parts of this artificial census as was actually found in the real census of 1870. We could have as well called the white infants kittens and the black infants mice. We are talking merely about a certain lottery or pair of lotteries, not about blacks or whites or infants or kittens. The theorem is true in the imagined lottery game; it has no truth or falseness relative to actual happenings in actual censuses, and can have none until we have shown, as we can not, that, to the high degree of approximation needed, the actual censuses do justify a calculation based on this lottery.

(17) It is fortunate that there is no physical truth implied by these meticulously accurate calculations. You have heard of the proposition in the theory of probability known as the Gambler's Ruin. Betting ten billion to one would soon ruin one even with the best of luck if he were betting real goods in a real world of actual events; and it would ruin scientific reputations if people didn't realize that we were just playing the high priest in a Shamanistic rite, which is a much safer and perhaps an admirable rôle.

It appears, however, that there are some who can not cheer the game along. Such was the miserable "knocker" of a Leibnitz. Such is that wretched "kill-joy" Keynes. He considers it absurd that Pearson after a lengthy analysis should conclude as follows: If a sample of one hundred of a population shows 10 per cent. affected with a certain disease, then in a second sample of one hundred it is even betting that the percentage affected will be between 7.83 and 13.71. Keynes appears to think that Pearson is talking nosometry or vital statistics. Not at all, Mr. Keynes. The proposition is purely mathematical and identically the same as if it had read: If you come across one hundred dairy cows and 10 per cent. of them being milked, then when again you find one hundred dairy cows, it is even money that not fewer than 7.83 of them nor more than 13.71 will be milking. Or again: If you buy one hundred hellgramites and use them up catching ten bass, then you can bet even that in the long run you will come home half your bass fishing days with not less than 7.83 bass nor more than 13.71, provided you use just one hundred hellgramites each day. This form of statement might appeal more than the others to Izaak Walton; but there is no trusting him, because like Mr. Keynes he may be a realist and believe that fishing is fishing and not higher mathematics.

And then there is Whitehead, mathematician, philosopher, physicist and logician par excellence, who in a paper entitled "Uniformity and Contingency" delivered as his presidential address before the Aristotelian Society[11] says: The latest and sublest analysis of the difficulties which cluster around the notion of Induction is to be found in Part III of J. M. Keynes's "Treatise on Probability." What can we say to him? The joint author with Russell of the "Principia Mathematica" can not be suspected for a moment of not appreciating the significance of abstract mathematics. May it be that he believes induction belongs to that real nature which has her habits but only in a general way and that refined calculations mislead? Perhaps he will sometime come and tell us himself.

(18) Fortunately the solution, if there be a solution, to these difficult questions which lie at the basis of statistical theory is not of pressing importance in most practical applications of statistics. There is nothing practical about odds of one in a billion. Had I an adequate grounding in the biological and medical sciences, I might be less concerned with such doubts; but coming as I do to vital statistics through the more exact sciences, I believe that for me to ponder in this way is a very useful help against overstepping or overstumbling the bounds of what is reasonable in drawing conclusions. Did not Pettenkofer demonstrate a relation between the level of the ground water and typhoid in Munich,[12] and were not terrific odds calculated in support of his theory? Was it useful to build the chances quite so towering high that the fall thereof should the more resound?

I seem not to have reached any very definite conclusions; I came but to discuss. And as I remarked at the outset that coming here with any reflections on statistics was but carrying coals to Newcastle, I may only hope at the last that my cargo may not be wholly confiscate as slate.

EDWIN BIDWELL WILSON

SCHOOL OF PUBLIC HEALTH,
HARVARD UNIVERSITY

[11] *Proc. Aristot. Soc.*, Vol. 23, p. 18.

[12] I do not wish to be misunderstood. There was undoubtedly at Munich at the time a high correlation between the height of the ground water, or variations in the height, with the rise and fall of typhoid; so much is description, not inference. The correlation coefficient many times exceeded its probable error and such an excess in a game of chance would be excessively rare; so much is pure mathematics, not statistical inference. The inference was drawn that there was a direct causative relation, that typhoid entered the body not through the alimentary but through the respiratory tract, that the relation was general both in respect to time and to place and even with respect to disease. This led to a return of *mal-aria* or *miasm* as fundamental in epidemiology, and the point of view prevailed in some quarters for many years and was applied to malaria itself to combat the rising mosquito theory (London, *Epidemiological Society*, Vol. 17, pp. 76–77).