



Karl Pearson and Mathematical Statistics

Author(s): Burton H. Camp

Source: *Journal of the American Statistical Association*, Vol. 28, No. 184 (Dec., 1933), pp. 395-401

Published by: [American Statistical Association](#)

Stable URL: <http://www.jstor.org/stable/2278129>

Accessed: 02/10/2011 13:45

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*.

<http://www.jstor.org>

KARL PEARSON AND MATHEMATICAL STATISTICS

BY BURTON H. CAMP, *Wesleyan University*

The retirement of Karl Pearson as professor at the University of London and director of the Galton Laboratory marks the culmination of a most notable chapter in the development of statistics. From many parts of the world men and women have come to his laboratory to listen to his lectures and to conduct their own researches in his stimulating presence. His editorship of *Biometrika* has made for that journal its position of prime importance as a repository for contributions to theoretical statistics.

Before reviewing Pearson's mathematical work it is necessary to pay respect to his personal qualities as a teacher and a scholar. It would be impossible for one who has been in close touch with him not to feel compelled to do this, and in addition these qualities have an important bearing on a proper interpretation of his writings. First of all he is friendly. This is probably not appreciated to the degree to which it is true by those who have been only his readers, for there is much in what he has written that is caustic. His critics have been dealt with in severe and able language. Sometimes it has been obvious that this has been well deserved, when they saw only a little of what he meant and gave publicity to palpably incorrect interpretations or to naïve criticisms of his views. But sometimes it has not been deserved, or at all events not obviously deserved, and then of course it reflected adversely on its author, but it does not follow, as some may have supposed, that he is given to shallow judgment or that he is unkind. Rather, if I may apply an Americanism to so staunch a Briton, he is quick on the trigger. I once had a cowboy friend at Harvard who used to say that Cambridge was all right, but as for him, he preferred a country where there was "jest a little smell of gunpowder in the air, not enough to make it disagreeable, but jest enough to make everybody polite, one to the other." He would have loved the Galton Laboratory when Professor Pearson was about, and—this is the point of the story—the rest of us loved it too; for with the brilliant mind and its masterful repartee lives as warm and kind a heart as a teacher ever had. It cannot be said of him as of some that he is so engrossed in things scholarly as to leave out the human touch. Indeed, strange as it may seem, something which is almost the reverse of that is true: although one cannot be in his presence without recognizing that here is a distinguished person, one wants to be in his presence not because he is distinguished

but because he is lovable. Every year at the laboratory a reunion is held of such of his former associates and pupils as are near enough to come. What impresses the stranger most about these meetings is that these persons seem to have come to do honor not so much to the philosopher as to the friend.

At his laboratory there was truly an association of scholars. Although local students were there working for degrees, for the most part those who had come were working simply for the development of science. Professor Pearson was not only the acting head of his laboratory, but was vital in every one of its activities. Anthropologists, biologists, sociologists, psychologists, mathematicians and others were there together, each working on his own problem, and once, frequently twice, every day, Professor Pearson sat down with each individual and thought through his work with him. He was indeed so very helpful it was even embarrassing, for it is not always easy to show progress in research twice a day. Pearson is indefatigable. He arrived at the laboratory early in the morning before others were admitted and left long after others were excluded. He hurried through lunch and beat his staff back to the books. He did not attend the British Empire Exposition in 1924. It was only a ten minutes' ride from his office, but he said he did not have the time. He was even then, at age 67, working at home late at night. He was taking a month's so-called vacation in August, but carrying his work with him, and coming back to London once or twice a week.

He is painstaking in two important respects. First, his mathematics is essentially rigorous. I was somewhat surprised to find that this was so, for coming from a background of training in analysis and having read most of his papers, I had the feeling that his mathematics might be a bit on the hop, skip and jump order, but I found that although his writings did not always mention the fine points, still they were in his mind, and really had been taken care of. Secondly, his computation, though naturally accurate, was always thoroughly checked, and he has insisted on similar care among his associates. Much of Pearson's theoretical work will, of course, ultimately be rewritten, perhaps several times, but the voluminous tables which he and his staff have compiled will for the most part never be recomputed. It is a comfort to know that they are trustworthy. The problem of computing a truly reliable table is not the simple one which those who have not done it commonly suppose, and a prodigious amount of work, both of routine and of theoretical nature, has been done at the Galton Laboratory on tables. The following tables at least should be mentioned: *Tracts for Computers*, *Tables for Statisticians and Biometricians* (2 volumes),

Table of Twenty Place Logarithms, Tables of the Incomplete Gamma Function. In connection with the construction of the latter much theoretical work was done on the problem of interpolation (see also the following by Seimatsu Narumi, one of Pearson's pupils: "Some Formulae in the Theory of Interpolation of Many Independent Variables," *Tôhoku Mathematical Journal*, vol. xviii, pp. 309-321).

This account of Pearson's scientific activities will have to be restricted almost exclusively to the mathematical part, but, although probably his eminence is due primarily to his success as a mathematician, his contributions to other sciences have been very important indeed. It is difficult to do justice even to his mathematics without incursions into various other fields, as will be evident from some of the titles to be cited below. This is especially true of his papers in the *Draper's Company Memoirs*. The record of his work is scattered through many volumes. His writings in *Biometrika* alone total about 1,500 pages, not including papers under joint authorship and others obviously done under his immediate supervision. He has written no book on mathematical statistics. Many wish that he would do so, for his writings have a clearness of exposition hard to match and he has at his command a great wealth of illustrative material. Possibly now, after his retirement from the laboratory, this hope of his friends may be considered more favorably.

One of his most important early papers on statistics was "Skew Variation in Homogeneous Material," *Philosophical Transactions of the Royal Society, A*, vol. 186 (1895), pp. 343-415. This contains a complete exposition of his now well known frequency curves (the fundamental types). Other frequency curves have been suggested, such as the so-called Gram-Charlier series of Hermite's polynomials, which had been tabulated by Pearson in the guise of tetrachoric functions, and various generalizations of both types. For a time there was much discussion as to which sort of frequency curves was the most valuable. This was rather regrettable. Both the Pearson and Charlier types spring from natural assumptions and both are valuable aids in analysis. Although it is a striking fact that almost every natural frequency distribution can be fitted by one of Pearson's curves or by a few terms of Charlier's series, it does not follow that either of these systems comprises in some hidden sense a natural law, and prolonged argument as to which gives the better fit would not appear to be justified on that ground. Certain of Pearson's curves are, of course, coming into prominence now in another connection, namely as the theoretical forms which are satisfied by the sampling distribution of certain statistical parameters.

Pearson's discovery of the chi-square test of significance was published in the *Philosophical Magazine* in 1900 with tables, vol. 50, pp. 157-175. The theory as then announced was essentially sound and has been of great value. As pointed out by Fisher and others, that theory would better be modified if used otherwise than in the ideal case, that is, the case where the universe sampled is supposed known. This modification turns out to be quite simple fortunately, and, as clearly stated by Irwin in the *Journal of the Royal Statistical Society*, vol. 92 (1929), p. 264, it is not absolutely necessary. It is a matter of precisely what question in probability one wishes to solve. It should also be pointed out that, by using too fine a division, Pearson at first carried some of the implications of his theory to an unwarranted extreme.

The theory of sampling runs through many volumes of *Biometrika*. When this theory was developed the samples were supposed fairly large and for the most part the discussion had to do with the discovery of formulae for the standard deviations of various statistics, a very important matter which is basic to the whole theory of sampling. Pearson was not at that time interested in the modern question of small samples and again he sought usually a solution for the ideal case when the universe sampled was supposed known. Again it is true that the modern improvements are often made possible by shifting the questions in probability from the questions whose solution was sought by Pearson to similar but not exactly identical ones whose solution for small samples it is easier to obtain. These early papers of his on sampling are marked by a thoroughness and completeness that have not been fully appreciated. Together they form an admirable text on the foundations of the subject. Latterly he has contributed to the small sample theory. This he thinks of as valuable but not so valuable as it sometimes appears. It should not, he thinks, be swallowed whole:

Experimental work of a very useful kind has been started to discover how far the present mathematical theory of small samples can be extended to other than a single type of parent-population; but it is too early yet to be dogmatic as to the limits within which the application of such theory is valid. In particular I hold that the so-called "z" test as usually applied to small samples, especially when it is used to measure the probability or improbability of identity in the constants of small correlated samples, really requires further consideration. (1931.)

The idea involved in the coefficient of correlation was initially due to Galton, and it was originally called Galton's function, but Pearson's work on the development of this theory has been so important that the coefficient is now commonly known as his. The following papers should be mentioned here: "Regression, Heredity, and Panmixia," *Philosophical Transactions of the Royal Society, A*, vol. 187 (1896), pp.

253-318; "On the Influence of Natural Selection on the Variability and Correlation of Organs," same journal, A, vol. 200 (1903), pp. 1-66; "Novel Properties of Partial and Multiple Correlations," *Biometrika*, vol. 11 (1915-17), pp. 231-238. Pearson has investigated also other measures of interrelation such as the coefficient of contingency; e.g., "On the Theory of Contingency and Its Relation to Average and Normal Correlation," *Draper's Company Research Memoirs*, Biometric Series, vol. 1 (1904). These other coefficients are not so valuable as the coefficient of correlation, however, and the same is true of various coefficients advocated by others, and Pearson has been forced to spend a good deal of labor in proving this. His tetrachoric " r " is theoretically the best measure of interrelation in a fourfold table, being in fact the very " r " of that normal surface which precisely fits the table. For many years it suffered in popularity because of the difficulty in its computation. That difficulty is now completely removed with the publication in 1931 of his second set of Tables (cf. also *Biometrika*, vols. 11, 19, and 22). The problem of polychoric " r " is still in a less satisfactory state (cf. an article by K. and E. S. Pearson, *Biometrika*, vol. 14, pp. 127-157), and it is especially because of this fact that the coefficient of contingency is used, but the latter is an unsatisfactory substitute, partly because it does not depend on the order in which the columns (or rows) of the correlation table are arranged. In this connection it is pertinent to note that at an early date Pearson recognized the error in dealing with a merely ordered series as if it were measured, by the method of assigning to it arbitrary numbers, and emphasized as the only scientific basis of measurement the method of graduation by means of a normal curve. This method lies at the foundation of much of the technique of the psychologist and the educationalist and the use of the Kelly-Wood table and others.

Pearson has been much interested in the history of statistics and is an avid reader of the early masters of the theory of probability, Bernoulli, Laplace, and others. It was by a brilliant inference that he found a rare appendix to a volume of De Moivre which showed that De Moivre and not Gauss or Laplace was the real author of the normal law, in the sense that De Moivre first gave the relation between this exponential function and the point binomial of probability theory.

The above paragraphs have to do with Pearson's thoughts on some matters that are familiar to all of us. For the rest it is perhaps sufficient to pick out from a large number half a dozen subjects with brief references for each, merely to indicate the variety of his interest in mathematical statistics: Probability that two samples belong to the same population, *Biometrika*, vols. 8, 10, 24, 25; hypergeometric series,

simple and double, *Biometrika*, vol. 16 (*cf.* also Romonovsky, vol. 17); bivariate surfaces, *Biometrika*, vol. 17 (*cf.* also Rhodes, vol. 14, Narumi, vol. 15); properties of Student's z , *Biometrika*, vol. 23; ranked individuals, and ranked variations, vols. 23 and 24. His earlier work in the fields of engineering and of mathematical astronomy is also important, but would not particularly interest readers of this JOURNAL.

Pearson has given much of his energy to the study of eugenics and anthropology, and although these are not our primary interest, they are too interesting to omit altogether. To quote from the University College Magazine:

In the field of Eugenics, he has ever stressed the importance of the careful collection of information before any valid theories can be formed. *The Treasury of Human Inheritance* which has been published in a number of parts, represents the first and still the only attempt in England to provide material on an adequate scale for the study of human genetics. His contributions to the scientific study of physical anthropology have been perhaps as great as those of any other man. A recognition of their value was shown in 1932 when the Rudolf Virchow Medal was presented to him, the only anthropologist not a German to have received this honor. His contributions to medical knowledge were also recognized when he was made an Honorary Fellow of the Royal Society of Medicine, a very unusual honor for a layman, while he is the only man outside the insurance world to be a member of the Actuaries Club. The year 1930 saw the completion of the third and last volume of a great labor of love, *The Life and Letters of Francis Galton*. Those who glance at even a portion of it will begin to understand, not only what Galton was, but what Karl Pearson has been and is.

Pearson's scientific achievement is thus another excellent illustration of the old truth that progress in both mathematics and practical science is specially fostered when they are permitted to interact the one on the other. The modern mathematical theory of statistics apparently owes its existence to the need for solving practical problems in the theory of inheritance, and much of modern biometry would not exist if this study had not elicited the interest of a mathematician. At this moment a committee of the American Statistical Association is at work on the problem of how best to nurture in this country the development of mathematical statistics and how to supply mathematical tools to the so-called practical statistician. It would appear that the story of Pearson might give the best possible solution, namely the founding for scholars in this country of a laboratory similar to his, with a mathematician of his promise who will study all their problems with them. If the latter objective appears too difficult to realize, it affords for that very reason a striking commentary on what he has accomplished.

Professor Pearson retires after forty-two years of service at University College and twenty-four years as the head of the Galton Labora-

tory, this position having been transferred to him by Sir Francis Galton two years before his death in 1911. Pearson's position is now being shared by his son, Egon Pearson, who is head of the department of statistics at University College, and by R. A. Fisher, who is Galton Professor of Eugenics and in charge of the Galton Laboratory.