Biometrika (1965), **52**, 1 *and* 2, *p.* 3
Printed in Great Britain

Studies in the history of probability and statistics. XIV Some incidents in the early history of biometry and statistics, 1890–94*

By E. S. PEARSON
University College London

1. Introduction

Perhaps the two great formative periods in the history of mathematical statistics were the years 1890–1905 and 1915–30. In both, the remarkable leap forward was made in answer to a need for new theory and techniques to help in solving very real problems in the biological field. In the earlier years the original questions posed concerned the interpretation of data bearing on theories of heredity and evolution; in the later period the first call was to sharpen and develop the tools used in agricultural experimentation.

There is considerable fascination in trying to find out how things looked at the time to the men concerned in such pioneer movements, from what background they started and what was the combination of circumstances which lead to the particular lines of advance which they followed. Below I shall try to describe some of the history of the first few years of the 1890–1905 period. A good deal of this has already been put on record, for example, by K. Pearson (1906, 1930) and I shall draw freely on this material, but the availability of certain letters between Francis Galton (1822–1911), F. Y. Edgeworth (1845–1926), Karl Pearson (1857–1936) and W. F. R. Weldon (1860–1906)† has made it possible to add some illuminating personal touches to what is already on record.

The final event which brought about the association of Pearson and Weldon, leading 10 years later to the founding of *Biometrika*, was Weldon's election in 1890 to the Jodrell Chair of Zoology at University College, London where Pearson had been Professor of Applied Mathematics since 1884. But to understand the basis of the co-operation between these two men we must look still further back. The threads were gathered from many sources.

In the second half of the 1880's Pearson was by profession an applied mathematician, a good deal of whose teaching was to students of engineering. Between 1884 and 1893 he had first prepared for the press W. K. Clifford's unfinished manuscript of *The Common Sense of the Exact Sciences* and had then undertaken the far more arduous task of completing Isaac Todhunter's *A History of the Theory of Elasticity*; the second volume of this, containing some 1300 pages were almost entirely Pearson's contribution.

But throughout the '80's his research energies were also occupied in a quite different direction, in the study of mediaeval and Renaissance German literature and folk lore. This work is recorded in the series of essays, most of them first given as lectures, which were later published in *The Ethic of Freethought* (1888) and *The Chances of Death* (1897). The

* This article is enlarged from a talk originally given to students at University College in 1960 and later, in 1961, at Princeton University from where it was issued as No. 45 in the *Statistical Techniques Research Group Reports* (1962).

† Many of the letters are in possession of University College London with whose permission they are quoted.

link between these investigations and *The Grammar of Science* of 1892 lay as he wrote himself in a 'fundamental note of the author's thought, namely: the endeavour to see all phenomena, physical and social, as a connected growth, and describe them as such in the briefest formula possible'. The manner in which his applied mathematics could best be harnessed in this endeavour was at first less obvious; but several events helped to shape the course of his activities.

Early in 1889, shortly after its appearance, he had read Galton's *Natural Inheritance*. From the paper which he presented soon afterwards to a small discussion club* it would seem that he was then not altogether ready to accept Galton's approach and was perhaps rather critical of the popular way in which Galton expounded his subject. But it is clear from what he wrote at a much later date (see E. S. Pearson, 1938, pp. 18–19) that from this time onwards he began to be aware of a door which might be opened into new and exciting fields.

Another lead opened up at this juncture. Pearson had attended Todhunter's classes when at Cambridge between 1876 and 1879, and his later connexion with *A History of the Theory of Elasticity* may have increased his interest in Todhunter's earlier publication, *A History of the Mathematical Theory of Probability*. At any rate when in 1890 he applied for the vacant Lectureship in Geometry† at Gresham College in the City of London, he included among the subjects which he offered to present, 'graphical statistics' and 'the theory of probability', as subjects likely 'to supply a want felt by clerks and others engaged during the day in the City'.

The first eight of the Gresham Lectures, given in March and April 1891, fell under the heading: 'The Scope and Concepts of Modern Science'; they contained the material later developed and enlarged in *The Grammar of Science* (1892). The first edition of this book had chapters on 'Cause and Effect—Probability' and on 'Life', but the treatment follows Pearson's philosophic approach to scientific concepts built up during the 1880's, i.e. belongs to what may be termed the pre-Galton–Weldon period of his development.

In the same way the second series of Gresham Lectures under the general heading 'The Geometry of Statistics and the Laws of Chance' began with twelve lectures (November 1891, January and May 1892) which appear from the syllabuses (E. S. Pearson, 1938, pp. 142–53) to have been concerned with a somewhat formal account of methods of presentation of descriptive statistics.

It is only later, from November 1892 through 1893 and 1894, that we begin to see the subject taking a new life: the introduction of experiment; the comparison of theory with observation, whether the latter resulted from coin tossing or measurements taken on organs of certain groups of animals; use of the mean, the standard deviation and the coefficient of correlation; frequency curves, symmetrical, skew and double humped; problems of evolution; of differential death rates and selection; illustrations made on data of Galton and Weldon; the study of racial differences through measurement of human skulls.

* This small forward-looking club had been formed in 1885 by a group of men and women who were convinced that some of the social problems of their day could only be furthered by a deeper understanding, based partly on historical research, of the relationship between the sexes. It was for the light which it might throw on the laws of heredity that Pearson had picked on Galton's *Natural Inheritance* for review. Gaussian distributions and the calculus of probabilities were not of primary interest to the Club members.

† The duty of the Lecturer seems to have been to give a dozen end-of-the-day lectures a year to an extra mural audience.

It seems to me that reading through these old syllabuses of 1891–94 we can get in summary form the picture of how under the stimulus of contact with Weldon at University College—and here I would place Weldon's influence before that of Galton—the applied mathematician in Pearson had at last discovered what he was needing, a field in which his special powers could be brought into action in solving some of the problems of life—*metron* applied to *bios*.

Weldon approached the unborn subject of biometry from a quite different angle. He had taken the Cambridge Natural Science Tripos in 1881, with zoology as his principle interest. After a period occupied in research and demonstrating he had been appointed in 1884 to a University Lectureship in Invertebrate Morphology. Having much of the outlook of a field naturalist, he spent many of his vacations when at Cambridge and later in collecting and 'dredging', sometimes abroad and often, after its completion, at the Marine Biological Laboratory in Plymouth.

He had started as most of the younger men of his day with an immense enthusiasm for the Darwinian theory of natural selection and sought the opportunity to contribute to the proofs of what could be described as only a working hypothesis. He had turned first to the current morphological and embryological methods of attack, but by the late 1880's he was beginning to doubt whether much progress could be made this way. So it was that his thoughts began to turn to the study of variation and correlation in organic characters. It was at this juncture that Galton's *Natural Inheritance* brought him sudden illumination. Here, he felt, were the statistical methods of measurement, description and analysis which might help to establish evidence supporting the Darwinian theory.

His paper of 1890, showing that the distributions of four different measurements (expressed as ratios to total length) made on several different local races of the shrimp (*Crangon vulgaris*) closely followed the normal or Gaussian law, was almost certainly the first paper in which statistical methods were applied to biological types other than man. In the statistical treatment of the data he had received help from Galton as a referee, but the credit for making the vast number of measurements and for seeing the bearing of such results on the problems of evolution was his own. A second paper in the series (1892) gave the coefficients or correlation between organs in the same individual and compared these for the four local races.

These two early papers were but first steps, showing that the simple models of statistical theory then current, the univariate and bivariate normal distributions, were relevant to zoological data. The further programme of research which was taking shape in his mind was to be set out in the third paper (1893, p. 329), where he wrote:

> It cannot be too strongly urged that the problem of animal evolution is essentially a statistical problem: that before we can properly estimate the changes going on in a race or species we must know accurately (*a*) the percentage of animals which exhibit a given amount of abnormality with regard to a particular character; (*b*) the degree of abnormality of other organs which accompanies a given abnormality in one; (*c*) the difference between the death rate per cent. in animals of different degrees of abnormality with respect to any organ; (*d*) the abnormality of offspring in terms of the abnormality of parents and vice versa. These are all questions of arithmetic; and when we know the numerical answers to these questions for a number of species we shall know the deviation and the rate of change in these species at the present day—a knowledge which is the only legitimate basis for speculations as to their past history, and future fate.

To handle these questions with any degree of assurance required in fact something more than arithmetic—the development of a more advanced theory of mathematical statistics.

Weldon's own mathematical knowledge at this period was limited and he realized that much more would be needed if the new tools were to deal adequately with the problems he began to see ahead. He set about removing this disadvantage in two ways; by himself beginning a study of the great French writers on the calculus of probability; and by seeking the co-operation of a mathematician in this project of demonstrating Darwinian evolution by statistical inquiry. After failing to get help from Cambridge he turned very naturally to his colleague Karl Pearson at University College.

In this sketch of antecedents I have not attempted to give any account of Francis Galton. His position as friend and counseller of the two younger men cannot be questioned; it was he who had taken the initial step in developing the ideas of correlation and regression which were to hasten the introduction of quantitative analysis into fields of biological, medical and sociological research. But in 1890 Galton was 68 years old, a man of established reputation whose long history of achievement has been told elsewhere. It was Weldon and Pearson who were to bring a fresh impetus into the field and it is their approach to our period of history which is therefore of special interest.

In the pages which follow I shall try to bring out some of the human side of the venture: the conflict of view-points of the biologist and the mathematician; the time so often taken, when on the fringe of the unknown, to see that next step forward which now seems so obvious to us, years afterwards, the enthusiasm which could lead the protagonists, though attached to the same College, to speed a second letter by the midnight post with some modified calculations after a first, sent on the same evening. I shall deal almost entirely here with problems coming under the head of variation, hoping at a later date to introduce some of the discussions on correlation.

2. The place of the normal curve

Whether we turn to *Natural Inheritance*, to the correspondence between Galton, Weldon and Pearson or to the Notes of Pearson's statistics lectures taken down in 1894–96 by Udny Yule (1871–1951), we see the central, predominating place which the Error Curve and the Binomial held in statistical thinking in 1890. As a result some effort was needed to break free from certain traditions. Starting from the work of De Moivre in 1733,* what we now call the Normal Curve was first derived as a mathematical approximation to the point binomial. Later on, when the Normal Curve was found to give an admirable fit to numerous recorded distributions of errors of observation, it had become customary to explain this good graduation in terms of a theory of the super-position of a series of elementary errors; text books on the Theory of Errors contained, for example, so called 'proofs' of the normal law.†

Galton had given to this idea a visual significance, by making a small mechanical model, often termed his Quincunx, and fully described with a diagram on pp. 63–5 of *Natural Inheritance*. In this model, lead shot from a funnel are dropped onto a succession of rows of equally spaced pins and collected below in a number of vertical compartments. In the

* Published in the *Supplementum* to his 1730 *Miscellanea Analytica*.

† As a reaction to this view among astronomers I remember how Sir Arthur Eddington in his Cambridge lectures about 1920 on the Combination of Observations used to quote the remark that 'to say that errors must obey the normal law means taking away the right of the free-born Englishman to make any error he darn well pleases!'

happy phrasing which he used when appealing to his reader's imagination, Galton wrote

> 'The shot passes through the funnel and issuing from its narrow end scampers deviously down through the pins in a curious and interesting way; each of them darting to the right or left, as the case may be, every time it strikes a pin. The pins are disposed in a quincunx fashion, so that every descending shot strikes against a pin in each successive row....The outline of the columns of shot that accumulate in successive compartments approximates to the Curve of Frequency, and is closely of the same shape however often the experiment is repeated....
>
> The principle on which the action of the apparatus depends is, that a number of small and independent accidents befall each shot in its career. In rare cases, a long run of luck continues to favour the course of a particular shot towards either outside place, but in the large majority of instances the number of accidents that cause Deviation to the right, balance in a greater or less degree those that cause Deviation to the left....'

Possibly for use in his Gresham College Lectures on the Theory of Probability given in 1893, Pearson constructed a modification of Galton's model, in which the shot fell onto a succession of rows of small triangles projecting from the back-board, which could be progressively stepped sideways from row to row. As a result the chance of a shot bouncing to the right (p) was not equal to that of it bouncing to the left (q). The distribution collected in the compartments should then not be symmetrical, but correspond roughly to the terms of the binomial $(q+p)^n$ with $p \neq q$.

The emphasis which Galton placed on the Normal curve can be best illustrated by quoting some further passages from his *Natural Inheritance*.

> I need hardly remind the reader that the Law of Error upon which these Normal Values are based, was excogitated for the use of astronomers and others who are concerned with extreme accuracy of measurement, and without the slightest idea until the time of Quetelet that they might be applicable to human measures. But Errors, Differences, Deviations, Divergencies, Dispersions, and individual Variations, all spring from the same kind of causes. Objects that bear the same name, or can be described by the same phrase, are thereby acknowledged to have common points of resemblance, and to rank as members of the same species, class, or whatever else we may please to call the group. On the other hand, every object has Differences peculiar to itself, by which it is distinguished from others.
>
> This general statement is applicable to thousands of instances. The Law of Error finds a footing wherever the individual peculiarities are wholly due to the combined influence of a multitude of 'accidents', in the sense in which that word has already been defined. All persons conversant with statistics are aware that this supposition brings Variability within the grasp of the laws of Chance, with the result that the relative frequency of Deviations of different amounts admits of being calculated, when those amounts are measured in terms of any self-contained unit of variability, such as our Q (pp. 54–5).

Again, in a section headed *The Charms of Statistics*, we find

> It is difficult to understand why statisticians commonly limit their inquiries to Averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once. An Average is but a solitary fact, whereas if a single other fact be added to it, an entire Normal Scheme, which nearly corresponds to the observed one, starts potentially into existence.
>
> Some people hate the very name of statistics, but I find them full of beauty and interest. Whenever they are not brutalized, but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary. They are the only tools by which an opening can be cut through the formidable thicket of difficulties that bars the path of those who pursue the Science of man (pp. 62–3).

Finally, under *Order in Apparent Chaos*, he writes

> I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error'. The law would have been personified by the

Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshalled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along. The tops of the marshalled row form a flowing curve of invariable proportions; and each element, as it is sorted into place, finds, as it were, a pre-ordained niche, accurately adapted to fit it. If the measurement at any two specified Grades in the row are known, those that will be found at every other Grade, except towards the extreme ends, can be predicted in the way already explained, and with much precision (p. 66).

This enthusiastic placing of the Normal distribution in the forefront of the study of what he terms the Science of man was not made without considerable background investigation. At the International Exhibition held in London in 1884, Galton had had an Anthropometric Laboratory and he includes in *Natural Inheritance* a summary of some of the data which he had collected, by taking measurements on visitors to the Exhibition. Thus he measured nine physical characters in men and women, the frequencies for the 18 distributions varying between 212 and 1013. For each distribution he had found the standardized deviates (measured from the median in terms of the probable error, not the standard deviation) to the eleven percentage points shown below. He found the mean of these standardized deviates for the 18 distributions and compared them with the Normal curve values as follows:

Cumulative % from lower tail		...	5	10	20	30	40	50	60	70	80	90	95
Mean of 18	Observed		−2·44	−1·87	−1·24	−0·77	−0·40	0	0·38	0·75	1·21	1·92	2·47
deviates	Normal		−2·44	−1·90	−1·25	−0·78	−0·38	0	0·38	0·78	1·25	1·90	2·44

Over 70 years ago these results must have appeared as remarkable pointers in a hitherto unexplored field. It is hardly surprising that Weldon, when collecting the measurements of physical characters in animal populations should start with the assumption that these would be normally distributed within a homogeneous race.

3. The double humped curve

In this early work there appeared to be one noticeable exception to the fit of Normal curves to Weldon's distribution, that of the relative frontal breadth of the Naples crabs. He thought that this might arise from the mixture of two local races, providing perhaps evidence of the beginning of some process of selection at work. With much arithmetical labour, using trial and error, he graduated the distribution with the sum of two Normal curves. On the 27 November 1892, he wrote to both Galton* and Pearson telling of his achievement; to the latter he says:

Out of the mouths of babes and sucklings hath He perfected praise! In the last few evenings I have wrestled with a double humped curve, and have overthrown it. Enclosed is the diagram... [He adds numerical results]. If you scoff at this, I shall never forgive you.

It was this problem of dissection of a frequency distribution into two Normal components which led to Pearson's first statistical memoir, presented to the Royal Society in the autumn of 1893 and published in the *Philosophical Transactions* in the following year. In the most general case, the method involved the determination of the roots of a nonic whose

* At the end of his letter to Galton he remarked: 'Therefore, either Naples is the meeting point of two distinct races of crabs, or a 'sport' is in process of establishment. You have so often spoken of this kind of curve as certain to occur, that I am glad to send you the first case which I have found.'

parameters were derived from the first five moments of the observed composite distribution. Note that the introduction of the method of moments into the fitting of frequency curves is here described as giving a utilitarian answer to a practical problem; the possibility of other better solutions is admitted. Pearson recognized that some objective method of measuring goodness of fit had yet to be found; a possible method suggested was the comparison of the next, here the sixth, moments of the observed and fitted distributions. In fact, of course, the standard error of a sixth moment was so large that little could be achieved from this method of attack, but the derivation of the sampling errors of high moments lay in the future.

The question 'does a Normal curve fit this distribution and what does this mean if it does not?' was clearly prominent in their discussions. There were three obvious alternatives:

(*a*) The discrepancy between theory and observation is no more than might be expected to arise in random sampling.

(*b*) The data are heterogeneous, composed of two or more Normal distributions.

(*c*) The data are homogeneous, but there is real asymmetry in the distribution of the variable measured.

The conclusion (*c*) may have been hard to accept, such was the prestige surrounding the Normal law. But if it was accepted, it was still possible to retain the concept of a finite number of under-lying contributory causes, by deriving a continuous asymmetrical frequency curve from an asymmetrical binomial, on the lines of Pearson's Gresham College shot-model referred to above.

Weldon explored the position in an empirical way, both by tossing dice and by calculating the terms of a number of binomial distributions, $N(q+p)^n$, with $p \neq q$. The following extract from a letter to Pearson refers to this latter type of investigation:

23 April 1893

I have had a shock!

It seemed to me that the apparent symmetry of variation in animals showed that every 'accident' occurred about as often as any other, and that there was not, in any animal I had seen a 'tendency', as biologists have it 'to vary in one direction rather than in the opposite'.

But certain words of yours remained in my mind; and on Friday night I expanded $(0{\cdot}6+0{\cdot}4)^{20}$ and $(0{\cdot}7+0{\cdot}3)^{20}$ with most alarming results. I enclose a diagram of the appreciable terms in $(0{\cdot}7+0{\cdot}3)^{20}$. I should certainly not appreciate so slight a degree of asymmetry in an experimental curve zig-zag-ing about the diagram. But I should say that the observations varied symmetrically about the 7th term, $(p^{14}q^6)$.

So that no result of a kind which I had fondly hoped for can be drawn from these curves! I hoped that if an organ varied in a particular direction—that is if p became greater than q—the asymmetry of the curve would give some sort of measure of the difference between the two; and so a sort of kinetic of variation might be built up.

But if p may be more than twice as great as q with the abominable result which I enclose, that little hope goes to pieces.

It may seem curious to us today that so much weight was given to the idea of an underlying model in which contributory causes or factors led to discrete distributions which could be approximated by a continuous curve. As far as I know neither Weldon or Pearson made any serious attempt to identify the parameters of the discrete distributions with any biological phenomena. That all the component 'accidents' of Weldon's model would have a common p seems most unlikely.

As is well known Pearson (1895) obtained the fundamental differential equation of his generalized frequency curves as the limit of the slope/ordinate ratio of a hypergeometric series, and this was the approach which he still followed in his lecture presentation of the 1920's. In theoretical development, the 200-year-old tradition of deriving the Normal curve as an approximation to the binomial had a lasting hold on the imagination, though it is doubtful whether in practice, even in early days, he was much concerned about the physical meaning of his hypergeometric parameters.

The immediate stimulus for the development of theory leading to a system of skew frequency curves seems, however, to have come from Edgeworth, not from Galton or Weldon. Writing to his friend W. H. Macaulay of Kings College, Cambridge on 18 August 1895, Pearson remarks:

> There is a long tale as to the skew curves. Edgeworth came to me with some skew price curves nearly 18 months or two years ago [letters from Edgeworth suggest it was in the autumn of 1893] and asked me if I could discover any means of dealing with skewness. I had come to skewness also in my Gresham lectures. I went to him in about a fortnight and said I think I have got a solution out, here is the equation, and told him my chief (assumed) discoveries. I further said I don't intend to publish till I have illustrated every point from practical statistics....

In this connexion we find him writing to Galton on 19 November 1893:

> If you will suggest any type of statistics which you think ought better than another to give Macalister's curve,*...my Demonstrator Mr Yule and I will endeavour to fit them as an example of a type or class of asymmetrics.

The first Brunsviga calculator was not purchased until 1894, so that moment-calculation and curve fitting required in illustrating theory from 'practical statistics' proved rather laborious procedures involving a number of numerical slips which had later to be corrected. In writing to Yule in November 1894, Pearson remarked: 'I want to purchase a Brunsviga calculating machine before anything else, and am making inquiries about it. I think it would make moment-calculating fairly easy.' And later: 'Henrici† speaks well of the Brunsviga, but doubts whether it will not wear out with a few years' use'. But Henrici, of course, was wrong! To the end of his life, Pearson used as a spare machine at home a Brunsviga which must have been of the beginning-of-the-century vintage, while Maurice Kendall has told me that he has and still uses the Brunsviga which Yule purchased for work on his first (1910) edition of *An Introduction to the Theory of Statistics*.

4. Wanted: A test for goodness-of-fit

It seems appropriate to quote at the head of this section some remarks which Pearson wrote long afterwards about Galton in the third volume of his *Life* (1930, **3**A, p. 6).

> Again, if the reader anticipates that Galton was a faultless genius, who solved his problems straight away without slip or doubtful procedure, he is bound to be disappointed. Some creative minds may have done that, or appear to have done it, because, the building erected, they left no signs of the scaffolding; but the majority of able men stumble and grope in the twilight like their lesser brethren, only they have the persistency and insight which carries them on to the dawn.

Weldon's extensive dice tossing had another object in view. No valid criterion existed at this date by which to judge whether the differences between a series of group frequencies

* This was the log-normal curve which did not of course belong to his system, although agreeing very closely in shape with a corresponding four-moment Type VI curve.

† O. Henrici had been Pearson's predecessor in the Applied Mathematics chair at University College; he had moved from there to the City and Guilds of London Institute.

and a theoretical law, taken as a whole, were or were not more than might be attributed to the chance fluctuations of random sampling. Weldon, with his flair for empiricism, therefore decided to explore the kind of random fluctuations which one might expect to get in sampling, by comparing his tossing results with theoretical binomial expectations.

It was at this juncture that an incident occurred leading to the series of letters quoted below; unfortunately scarcely any of the letters written *to* Weldon have been preserved.

1. *Weldon to Galton*, from 30a Wimpole Street, W., 4. ii. 94.

Dear Mr Galton,

Will you be kind enough to give me your opinion on the following point?

I have collected 26,306 tosses of groups of 12 dice, for use at the Royal Institution. In each group the event recorded is the number of dice with 5 or 6 points, so that the chance of success with each die is 1/3. I enclose the result, which seems to me good.

A certain set of 7000 tosses, forming part of this result, was made for me by a clerk in the office of University College, whose accuracy in work of another kind I have had occasion to test by asking him to copy 24,000 numbers of 3 figures each, with excellent results.

A day or two ago, Pearson wanted some records of the kind in a hurry, in order to illustrate a lecture,* and I gave him the record of the clerk's 7000 tosses, together with some others.

I gave him the 7000 separately from the rest, and on examination he rejects them, because he thinks the deviation from the theoretically most probable result is so great as to make the record intrinsically incredible.

You will see how serious a matter this is. On the one hand I feel that I have no right to reject an experimental result for this kind of reason; and on the other, the result itself does not seem to me incredible.

I am anxious, however, not to rely upon my own judgement in so difficult a matter—I have therefore resolved to consult as many people as possible. Last night I saw Greenhill, whose experience in target practice at Woolwich makes him know this kind of thing statistically as well as mathematically—he is of opinion that the record is perfectly credible, and that I have no shadow of reason to disregard it.

Today I am sending it to you and to Edgeworth.†

Forgive me for troubling you, when I know how busy you are; but my need is very great.

Yours very truly,

W. F. R. WELDON

The 'most probable' result in each of the enclosed tables is obtained from the expansion of

$$\left(\tfrac{1}{3}+\tfrac{2}{3}\right)^{12}$$

and not from any form of approximation.

We can get a clue as to why Pearson considered 'the record as intrinsically incredible' from Yule's notes, taken down at Pearson's lectures in the autumn of 1894. In column 2 of the table below are presumably the result of the clerk's dice throwing, 7006 (not 7000) tosses of 12 dice. Column 3 shows the binomial expected frequencies as given by Yule, from the expansion of $(\frac{2}{3}+\frac{1}{3})^{12}$. Note that these figures add to 7010·2 not 7006; Yule does not however quote totals. He remarks:

The fit is good except at the ordinates four and five; 4 is 98 too low, 5 is 69 too high. This is odd: can the experimenter have inadvertently booked his results in the wrong column? Let us find out first what the chance against the above combined occurrence is.

* This lecture is almost certainly one of the four lectures on 'The Geometry of Chance' given that winter by Pearson at Gresham College.

† For the correspondence with Edgeworth which resulted, see pp. 13–15 below.

No. of 5's or 6's	Observed frequency	Expected frequency	$O-E$	$\frac{(O-E)^2}{E}$
0	45	54·0	−9·0	1·50
1	327	323·7	3·3	0·03
2	886	890·2	−4·2	0·02
3	1475	1483·7	−8·7	0·05
4	1571	1669·2	−98·2	5·78
5	1404	1335·3	68·7	3·53
6	787	788·9	−1·9	0·00
7	367	333·8	33·2	3·30
8	112	104·3	7·7	0·57
9	29	23·2	5·8	1·45
10	2	3·5		
11	1	0·4	−0·9	0·21
12	0	0·0		
Totals	7006	7010·2	−4·2	16·44 = χ^2

The argument used in the Notes, which was presumably the lecturer's, was based on the following analysis in which I have made the totals 7000 to fit in with Yule's arithmetic.

Group	Observed		Expected	
$x = 4$	1571		1669	
$x = 5$	1404	5429	1335	5331
Remainder	4025		3996	
Total	7000		7000	

Directing attention to the most exceptional group, that containing four '5's or 6's', Yule gives the standard error of the observed frequency as

$$\left(7000.\frac{1669}{7000}.\frac{5331}{7000}\right) = 35{\cdot}65$$

which gives $(1571-1669)/35{\cdot}65 = -2{\cdot}75$ as the ratio of the deviation from expectation to its standard error. Using the normal approximation to the sum of binomial terms, the probability of an absolute deviation (positive or negative) from expectation as large or larger than that observed is 0·0060 or $\frac{1}{167}$.*

The argument of the Notes now runs as follows:

> But now what is the chance of this being combined with the deviation of 69 in the next ordinate? We must be careful how we proceed here. We cannot simply work out, *as above*, the chance of the deviation 69 and multiply by the chance above, for the two are *not* independent. If the 98 has already been lost (or gained) it must be made up *somehow* by the other ordinates. We must remove the 1571 whose position has already been allotted and deal only with the rest.

He now finds a standard error for the 'five' group of

$$\left(5429.\frac{1335}{5331}.\frac{3996}{5331}\right) = 31{\cdot}92,$$

an expectation in the group of $5429 \times 1335/5331 = 1359$ and a ratio of deviate to standard error of $(1404-1359)/31{\cdot}92 = 1{\cdot}39$. The probability of an absolute deviation exceeding

* Quoting what he says is a table given in Pearson's 'Gresham College Lecture Notes', Yule gives the probability as $\frac{1}{136}$.

1·39 is, from the normal approximation, $0{\cdot}16\bar{5}$. It is now argued that the probability of a pair of deviations, having *opposite signs*, as large or larger than those observed is

$$\tfrac{1}{2} \times 0{\cdot}0060 \times 0{\cdot}165 = 0{\cdot}00049$$

or approximately $\frac{1}{2000}$.* Yule adds that 'if we went through the whole lot of all other deviations, the total chance for them might be somewhat smaller'. 'Consequently it appears not unreasonable' his Notes add, 'to conclude that the experimenter has made some slip or other in entering results in the wrong column'.

To these quotations from Yule's Notes must be added some correspondence with F. Y. Edgeworth which occurred in February 1894 when the argument was still at its height.

2. *Edgeworth to Weldon*, from All Souls College, Oxford, 7. ii. 94.

Dear Weldon,

The tests which I have applied to the cases with *four* and *five* dice do not yield a result which excites much suspicion. I shall be curious to know your final decision.

Yours very truly,

F. Y. EDGEWORTH

From some loose notes with the letters it appears that Edgeworth's calculations had been rather perfunctory. He had: (*a*) correctly found the separate moduli (standard error $\times\sqrt{2}$) for the 'four' group and the 'five' group as 50·4 and 46·4 (S.E.'s of 35·6 and 32·8); (*b*) divided these into the observed deviations from expectation of 98·2 and 68·7; (*c*) commented that in the first case: 'the ratio C is not quite 2, corresponding to odds against of rather more than 200 to 1, which can't be thought prohibitive I think'; (*d*) in the second case, remarked that the ratio $C = 1{\cdot}4$ corresponded to a very 'moderate improbability' and added: 'I write without a Table by me. But I know that the odds are nothing out of the way, say 50 or 100 to 1'.

We do not know what Pearson had first written to Edgeworth but he may well have proposed a procedure on the lines set out in Yule's Notes. There are, however, three further letters which have survived and are of interest in showing how at the beginnings men 'stumble in the twilight'.

3. *Edgeworth to Pearson*, from All Souls College, Oxford, 9. ii. 94.

My dear Pearson,

Your method would be all right as long as you are given only *two* results of the kind operated on. But it is not open to you I think to apply twelve (independent) tests to a composite event such as that considered; to select *two* which accuse, as the French say, a cause other than accident; and multiply the (im)probability of each to find the (im)probability of the system. To take a simple case, suppose each result presented one of two alternatives, either (*a*) ordinary, (*b*) improbable in the degree 1:100. Suppose that having *fifty* returns such as those before us (or any similar data) you look through them and find two events the probability of each of which is only 1/100, I don't think it is open to you to say that the probability of the system is 1/10,000. You should consider the growing likelihood, as you increase the number of your trials, that such extraordinary results will be presented. You surely would not make the same assertion if there had been *fifty thousand* data.

So it seems to me at present; but I know how kaleidoscopic these problems are.

In haste

Yours very truly,

F. Y. EDGEWORTH

* Through what was possibly a numerical slip afterwards corrected in red ink, the original calculations in the notes gave 1/14,416. Whether it was this probability, rather than the 1/2000, which induced Pearson to tell Weldon that the record was 'intrinsically incredible', one cannot now say!

4. *Pearson to Edgeworth*, from 7 Well Road, Hampstead, N.W., 10. ii. 94

My dear Edgeworth,

Probabilities are very slippery things and I may very well be wrong, but I do not clearly follow
A| your reasoning or illustrations. You say take 50 returns of 1:100 degree of probability. If two
occurred should I calculate the chance of the system as 1/10,000? Certainly *not*. Following the
B| method I applied to the dice, I should ask what is the chance of one 1:100 event occurring *in 50*
throws. This is $\frac{1}{2}$, and after its having occurred what is the chance of another like event in the
remaining 49 occurrences. It is 49/100. The combined chance is $49/100 \times \frac{1}{2} = 0{\cdot}245$, strikingly close
to the $0{\cdot}25 = \frac{1}{4}$, which I assume you to mean to be the probability of the event you suggest.

Now look at the dice problem in the same way. Disregard all but 5 and 6 occurring 4 times in
the 12 dice. I calculated the chances that in 7000 throws there should be a defect of 98 or more.
Chance = 1/270. Now make another experiment with the same 12 dice, take 7000 throws and
inquire how often there will be 5 and 6 occurring 5 times a certain number of times in excess. Suppose
C| this came out 1/40. Surely the combined chance against the two experiments would be $1/270 \times 1/40$?
They are quite independent.

Now what I contended was this, that admitting a defect of 98, this defect ought to be distributed
theoretically along the whole line of groups and that having done this, the distribution of the
C| remaining number of throws among the 5, 6, 7, 0, 1, 2, etc., was a practically new and independent
experiment.

Your method of looking at the matter leads me into difficulties in the following way. Suppose
we are quite certain that a population follows a normal frequency curve. We have, we will say,
discovered this by measurements on several 100,000's. Now we take a sample 10,000, and draw its
frequency curve, with a result when compared with the normal curve like this:

[Here Pearson sketched roughly a normal curve and a frequency polygon with a single 'hump'
rising well above the curve at X.]

Here the hump at X is counterbalanced by proportionate diminution of all the other ordinates.
The chance of this hump we will take to be 1/270. Now suppose instead of this, which fits the curve
very well except at A, we had a result like that over page, with a marked deficit at Y.

[Here is another diagram in which the frequency polygon has a marked dip at Y as well as the
hump at X, the defect at Y being rather less than excess at X.*]

D| Are you prepared to say that both these systems are equally probable and both improbabilities
are to be measured, hump X being more improbable than dip Y by the improbability of X? It
seems to me that the appearance of another anomaly like Y which almost counterbalances X must
much increase the improbability of this second system as compared with the first. You say, No!
E| Chance of $X = 1/270$, chance of $Y = 1/40$, and chance of whole system is the greater of these $= 1/270$,
F| and is not touched by whether Y exists or not. This does not seem to me at all satisfactory. I quite
agree X and Y are not independent. Well, make them so by cutting X off and distributing it in
C| proper proportions round the curve. Y will take some of it, but *not* all. Having done this the reduced
Y is an independent event, is an independent discrepancy in the normal frequency curve.

You say but there are other defects besides Y. Certainly, but when we have filled up Y again
they are of such minor importance that they are hardly worth considering—they fall, continuing
the process, so low down in the scale of fractions of the corresponding S.D.S. Here seems to me our
G| difference. You appear to calculate the improbability of a given distribution by its chief irregularity.
I assert that weight must be given to other irregularities. Now if dice gave a symmetrical curve we
should compare the theoretical and experimental Standard deviations and thus get a test of the
H| system as a *whole*. Surely the experimental S. deviation pays attention to Y as well as to X. It
diverges more from the theoretical, because X is not proportionately distributed but is collected
largely at Y. Now a similar result, it seems to me, must hold for these skew systems. Accordingly
I contend that your 1/270—mind you in itself a somewhat improbable result—does not represent
I| the badness of the experiment, that irregularity in other columns is also to be taken into account.
J| I admit that the proper way of doing this is quite open to question, but I think whatever way it is
done the great dip at Y will immensely increase the odds. Please note that a *second* 7000 experiments
by the same clerk, calculated from the two worst columns only give a chance of 1/250 instead of
1/3240 showing a marked improvement. While 19,000 experiments give a chance of 1/160 (1st 7000
K| *excluded*) by same method.

Yours very sincerely,

KARL PEARSON

* I have changed Pearson's original letters, A and B, to X and Y in order to avoid confusion with the letters A–K which Edgeworth added to Pearson's letter for reference purposes when he returned it with his own answer.

5. *Edgeworth to Pearson*, from 5 Mt. Vernon, Hampstead, 12. ii. 1894.

Dear Pearson,

Excuse my returning your letter for convenience of reference.*

A. I mean *fifty* data such as your numbers of *fours, fives*, etc.; or rather the fifty *independent* observations such as those which you derive these from.

For C, (p. 4 and after), I quite agree as to the 'independence' of which you speak.

B. The chance of an event 'of 1:100 degree of probability' (p. 1) occurring (at least) once in fifty (independent) trials is $1-(99/100)^{50}$. The chance of its occurring as much as twice is

$$1-\left\{\left(\frac{99}{100}\right)^{50}+50\times\left(\frac{99}{100}\right)^{49}\times\frac{1}{100}\right\}.$$

D. Certainly not. I am not prepared to take *no* account of B (Y in your figure). I only say it is very *difficult* to take account of it where the case is not so simple as my 1/100.

The significance of X and Y is modified according to the number of 'trials' or independent events there may be (I have tried to indicate these by so many intervals).

F. I here meant that—exactly. 'Not much suspicion' are I think my words to Weldon.

G. I never emphasized the *chief*.

H. Surely. But what I complain [of] is that you don't take the system as a whole but content yourself with a method proper to a *single datum or pair*. Having n observations you look out for one or two improbable results; and of course you will find them if n is large. There is our difference. You seem to think the size of n makes no difference. I say then as before let n be 50,000. It is chock sure that you will have two of the events (considered as independent) occurring, although the probability of each is 1/100. See your A. The improbability is *not* 1/10,000 but

$$1-\left\{\left(\frac{99}{100}\right)^{50,000}+50,000\left(\frac{99}{100}\right)^{49,999}\times\frac{1}{100}\right\}.$$

There is the issue; I subscribe to J.

K. Well then it is no wonder that we should have some deviations just to keep up the average.

If you are at University College tomorrow afternoon I may see you. I shall be there about 6.15.

Yours very truly,

F. Y. Edgeworth

I write in great haste and may well have made slips.

There are two further letters from Edgeworth to Pearson written later that February which show that the methods of dealing with multiple discrepancies were still being discussed. But the topic has shifted from the immediate problem of Weldon's 7000 dice throws, and I have not space to quote them here.

The interested reader must choose his own method of tackling this problem. Edgeworth was clearly criticizing Pearson for having picked out the largest and second largest of a set of differences from expectation without making proper allowance for the fact that they were the largest in a group of 13. But even if this were recognized, no theory existed for dealing with the largest of a set of mutually but unequally correlated differences.

Pearson's (1900) χ^2-test for goodness-of-fit, derived some five years later after the theory of multiple correlation had been developed, leads to the calculations shown in columns 4 and 5 of my table. The probability of obtaining a χ^2 based on 10 degrees of freedom, exceeding 16·44, if the discrepancies in the table taken as a whole were due to chance, is 0·088. Certainly this does not seem exceptional, but of course the χ^2-test does not take account of certain kinds of coincidence. Perhaps the position of the two large discrepancies next to each other in the table might justify some suspicion.

Had there been prior grounds for thinking that Mr Hull, the clerk, might have mixed up

* The letters A–K refer to the corresponding passages which Edgeworth had marked on Pearsons' letter of 10 February.

entries in these two particular columns, of 'fours' and 'fives', a sensitive test would consist in applying χ^2 to the following table.*

	Observed	Expected	Difference
$x = 4$	1571	1669	-98
$x = 5$	1404	1335	$+69$
Remainder	4031	4002	$+29$
Total	7006	7006	

Here, $\chi^2 = 9{\cdot}53$ and $P\ (\chi^2 \geqslant 9{\cdot}53 | \nu = 2) = 0{\cdot}009$, so that the result appears more exceptional, though hardly so even here on the basis of critical odds used by Edgeworth. However, we have been given no prior reasons for supposing that Mr Hull would confuse these two columns, and the selection of this most unfavourable three-category table would seem to be without justification. Theory, in fact, would be hard put to it to disprove Weldon, Greenhill and George Darwin's instinctive reaction to the figures; but the discussion must have emphasized the need for more thought and more mathematical research.

Yule's lecture notes mention an empirical measure of goodness-of-fit which was in use at the time. This was the ratio of (*a*) the area between the theoretical curve and the observed frequency polygon (not the histogram), taken everywhere as positive,† to (*b*) the total area under the curve. This may be set down roughly as

$$R = \Sigma |O - T| / \Sigma T$$

where O is the observed and T the theoretical frequency in a group. Yule quotes numerical values of this ratio, expressed as a percentage, for the cases of normal curves fitted to eleven different frequency distributions. The coefficients range between 5·85 and 13·5%, with a mean value of 8·0%. He remarks that when skew or compound curves were fitted to the same data, much better fits resulted, figures of '4% or so' being obtained.

During 1894 Weldon was much occupied with proposals for work which he planned to submit to a small, newly formed Royal Society 'Committee for conducting statistical inquiries into the measureable characteristics of plants and animals', of which he was secretary and Galton chairman. But among his letters to Galton discussing experiments to be sanctioned by the committee, there are many references to Pearson's asymmetrical frequency curves. It seems that neither Galton nor Weldon felt at home in the mathematics of Pearson's second Royal Society memoir (published in 1895); they were perhaps looking for a physical explanation of the fundamental differential equation, which they could not discover. Pearson also, with the enthusiasm of the creator of so elastic a system, was perhaps excusably trying out his curves on any set of non-normal data which came his way, without considering very deeply the biological meaning of the asymmetry. Thus, as Weldon pointed out, a skew distribution of measurements on the breadth of foreheads of Crabs or the stature of St Louis schoolgirls, might only reflect the fact that growth was going on within the age limits covered by the data.

The final letter which I shall quote, of a year later, bears on this theme and again well illustrates the slow process of bringing together the approaches of the biologist and the mathematician.

* I have amended Yule's reduced table given on p. 12 above so that the totals now agree with the original data quoted.

† The use of the Drawing Office planimeters made this an easy quantity to measure.

6. *Weldon to Galton*, from 30*a* Wimpole Street, W., 6. iii. 1895.

Dear Mr Galton,

Let me congratulate you heartily upon your recovery; I shall look forward with great pleasure to seeing you next week.

Pearson *does* admit that he omits to consider the moving mean in his theory of skew curves—or he did so nearly a fortnight ago, when I charged him with it—we had a delightful afternoon, abusing each other in a friendly way about this point for some hours; he promised more consideration of it; but since then he has been in bed with influenza—I hope he will be well enough to take a short holiday in a day or two, but he cannot work for some time.

Ten of our men at University College, and in many classes half the students, are in bed: so that I, who never get anything worse than a bad cold, feel like the Wandering Jew in time of plague.

About the mathematicians. I feel the force of what you say, naturally. But I am horribly afraid of pure mathematicians with no experimental training.

Consider Pearson. He speaks of the curve of frontal breadths, tabulated in the report, as being a disgraceful approximation to a normal curve. I point out to him that I know of a few great causes (breakage and regeneration) which will account for these few abnormal observations: I suggest that these observations, because of the existence of exceptional causes, are not of the same value as the great mass of the others, and may therefore justly be disregarded. He takes the view that the curve of frequency representing the observations, must be treated as a purely geometrical figure, all the properties of which are of equal importance; so that if the two 'tails' of the curve, involving only a dozen observations, give a peculiarity to its properties, this peculiarity is of as much importance as any other property of the figure.

For this reason he has fitted a 'skew' curve to my 'frontal breadths'. This skew curve fits the dozen observations at the two ends better than a normal curve; it fits the rest of the curve, including more than 90 % of the observations, *worse*. This sort of thing is always being done by Pearson, and by any 'pure' mathematician.

Greenhill, to whom I took my troubles, laughs at the whole thing. You know that his chief business is to teach the properties of probability surfaces to artillery officers in connection with target practice; and he has a good deal of experience of curves made with your quincuncial screen of pins.

Greenhill is quite ready to admit the necessity of ignoring the few aberrant observations. George Darwin says that Pearson pays too much attention to the higher moments (which of course depend chiefly on the character of extreme observations.)

Now these are the two men working at the applications of Probability who know, not only mathematics, but the degree of approximation to be expected from an experiment. This sort of instinct as to what may be expected of an experiment and what may not is a quality very rare among young mathematicians, so far as I know them.

But enough of them—I shall look forward to your proposals next week.

Yours very truly,

W. F. R. Weldon

The Herring, which makes a skew curve are very heterogeneous. The mean value of the length from snout to anus, on 717 males, was very widely different from that given by 990 males—the extra 270 being obtained by opening another of the cases of herrings. I have not the figures at hand, because I sent them to Pearson, as basis for his curve; but he says that '*the material is homogeneous, with skew variation about one mean*'. I don't believe it!

5. Concluding remarks

These letters are in several ways revealing; it would have been easy for the 'young mathematician' and the younger zoologist (Weldon was three years Pearson's junior) to drift apart, but the compelling urgency of the field for exploration which lay ahead, a field in which they were in so many ways fitted to co-operate, held them together. No doubt Galton played an important part in bridging the gap between them, so that six years later with friendship firmly established they were planning the first issue of *Biometrika*.

In the context of the 1894 discussions, they were still 'stumbling in the twilight'. There were many defects in Weldon's crustacean data: breakage and regeneration in individuals (as he himself pointed out); lack of homogeneity; the unknown effect of age on relative growth of parts and other disturbing factors. These particular series of observations were indeed unlikely to lead to any clear evidence, based on the fit of frequency curves, bearing on the process of natural selection. But the arguments which arose undoubtedly helped to bring out the need for more thorough investigation, both experimental and theoretical.

Pearson, too, showed an unsureness in the handling of the theory of probability. I suppose that he was never really interested in this calculus for its own sake, as a pure mathematician might have been; he needed its help as a tool in the solution of problems which *did* hold his interest. It was perhaps typical of the early British approach to mathematical statistics that he could write 'probabilities are very slippery things' in answer to Edgeworth's 'I know how kaleidoscopic these problems are'!

But reflexions of this kind cannot conceal the fact that out of these arguments developed the ever expanding structure of the theory of mathematical statistics which we know today. That seems ample justification for putting these incidents on record.

REFERENCES

GALTON, FRANCIS (1889). *Natural Inheritance.* London: Macmillan and Co.

PEARSON, E. S. (1938). *Karl Pearson. An Appreciation of Some Aspects of His Life and Work.* Cambridge University Press.*

PEARSON, KARL (1888). *The Ethic of Freethought, a Selection of Essays and Lectures.* London: T. Fisher Unwin. Republished (1901) by A. and C. Black.

PEARSON, KARL (1892). *The Grammar of Science.* London: Walter Scott. (1900, 1911), with additions, republished by A. and C. Black. (1937), first edition, republished in the Everyman Library (no. 939) by J. M. Dent and Sons Ltd.

PEARSON, KARL (1894). Contributions to the mathematical theory of evolution. *Phil. Trans.* A, **185**, 71–110.

PEARSON, KARL (1895). Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Phil. Trans.* A, **186**, 343–414.

PEARSON, KARL (1897). *The Chances of Death and other Studies in Evolution.* London: Edward Arnold.

PEARSON, KARL (1906). Walter Frank Raphael Weldon. 1860–1906. A Memoir. *Biometrika*, **5**, 1–52.

PEARSON, KARL (1930). *The Life, Letters and Labours of Francis Galton.* **3***A*. Cambridge University Press.

WELDON, W. F. R. (1890). The variations occurring in certain Decapod Crustacea. I. *Crangon vulgaris. Proc. Roy. Soc.* **47**, 445–53.

WELDON, W. F. R. (1892). Certain correlated variations in *Crangon vulgaris. Proc. Roy. Soc.* **51**, 2–21.

WELDON, W. F. R. (1893). On certain correlated variations in *Carcinus moenas. Proc. Roy. Soc.* **54**, 318–29.

* This book, now out of print, put together two articles previously published in *Biometrika* (1936), **28**, 193–257 and (1937), **29**, 161–248. Unfortunately two Appendices, one giving the Gresham College Syllabuses and the other G. U. Yule's summary of his 1894–96 lectures notes, are not included in the *Biometrika* articles.