John Venn as statistician

A.W.F.Edwards

At the Open University History of Statistics meeting, 15th December 2004

John Venn was a logician, and latterly a historian, and laid claim to no other academic fields. We cannot speak of Venn the statistician, but only of Venn *as* statistician, for his logical work, more especially in his book *The Logic of Chance*, lead him only to dabble in statistical matters. Maynard Keynes is reported to have said that *The Logic of Chance* was 'strikingly original and considerably influenced the development of the theory of statistics', but that must be a reference not so much to the theory as to the *logic* of statistics and the influence Venn had on the formalisation and acceptance of the frequentist viewpoint. But, as we shall see, Venn did inspire Karl Pearson to write his first statistical paper, in 1893. [11,12.]

Of course, Venn is well-known for his logical diagram. [13.] I have told its history in my recent book *Cogwheels of the Mind – The Story of Venn Diagrams* in which you may also find some more information about Venn himself. [14.] In truth, there is not very much to tell since he led a quiet life as a Fellow of Caius College, Cambridge, for sixty-six years, dying in 1923. [15.] In 1862 he became College Lecturer in Moral Science, and was one of the group instrumental in getting the University's Moral Science Tripos started. He published *The Logic of Chance* in 1866, *Symbolic Logic* in 1881 – containing the Venn diagram – and *The Principles of Empirical or Inductive Logic* in 1889.

Naturally, when I became interested in Venn I went to the college library to see if we had any of his papers and manuscripts, so I have long known of the existence of a set of his lecture notes labelled 'Theory of Statistics', and ever since I have harboured a minor ambition to look at them carefully and see what they contained. My interest was heightened ten years ago when Eileen Magnello pointed out to me that in 1894 Pearson had written 'There has been up to the present time – with the honourable exception of courses by Dr Venn in Cambridge – no teaching of statistical theory in England'. But I had become hooked on Venn diagrams, and continued to neglect the study of the lecture notes. So I am grateful for the opportunity to talk today because it has forced me to look at them at last. And very uninteresting they are, except as a window on the pre-Pearsonian statistical world.

It must have been as he prepared the third edition of *The Logic of Chance* that Venn became interested in what he later called 'the theory of statistics'. The preface is dated December 1887, and the edition was published the following year. [16.] We should scan the book to see what Venn has been reading. Already in the first edition he owns the influence of Todhunter's *History* (published the previous year), of De Morgan's *Formal Logic* (1847) and Boole's *The Laws of Thought* (1854). Mill's *System of Logic* (1856) and Whewell's *Inductive Sciences* (1860) are mentioned, as well as the help of Henry Sidgwick – later to beat Venn to the Knightbridge Professorship of Philosophy. By the second edition of 1876 he is expressing his indebtedness to Mr C.J.Monro, 'late fellow of Trinity' (whom Jevons also thanks in the second edition of his *Principles of Science* of 1877) and to J.W.L.Glaisher for information about least squares. By the third edition Venn has encountered the work Edgeworth 'to ... whom I am also personally much obliged for many discussions, oral and written, and for his kindness in looking through the proof-sheets', and Mr Galton ‘to whom every branch of the theory of statistics owes so much’. He conducted extensive correspondence with both men.

From the third edition it is clear that Venn had also read Ellis and Quetelet, Jevons and Whitworth, Hume and Donkin, Poisson and Lexis, Cournot and Bertillon, Airy and Merriman, and many others, as one would expect of a diligent Victorian logician with a taste for history and book-collecting. He is familiar with the famous early tests of significance by Arbuthnott and Mitchell and, most remarkably, with Düsing’s paper on the sex-ratio of 1884. So it is more interesting to remark on what he had not read. No Pascal or Huygens or even De Moivre. No Bayes. No Gauss and only the *Essay philosophique* of Laplace. The Bernoullis barely get a mention (and the index does not distinguish between James and Daniel). Here is a Cambridge mathematician writing about induction whose imagination has not been fired by the chief works on the statistical approach to the problem.

There are hardly any numerical examples in *The Logic of Chance* but we may note the appendix to Chapter 10, which is an elaboration of his comments on the sex-ratio in families earlier in the chapter. Venn analyses the distribution of the sexes in families drawn from the Herald’s Visitations. ‘They are not sufficiently extensive yet for publication’, he writes, and limits himself to fitting binomial distributions for the cases of 4, 5 and 6 children in a family. But he is able to note that for each family-size the central numbers are deficient and the extreme values over-represented, by comparison with binomial expectation. He suggests a mixture distribution ‘but fuller statistics are needed’. Was this the first one? Perhaps here is the origin of Fisher’s similar example in Chapter 3 of *Statistical Methods for Research Workers*.

Another numerical example, in Chapter 18, refers very briefly to the asymmetrical distribution of barometric pressure taken at the same hour each day, but here Venn refers us to his letter to *Nature* on 1st September 1887 for an actual figure. The Caius archives contain the draft. [17.] In the letter Venn applauds Galton and Edgeworth for demonstrating how untenable was Quetelet’s insistence on only one law of error, and he offers this example of barometric pressures recorded by Mr W.E.Pain of Cambridge [17X.] over a period of thirteen years. You may wonder how I have his photograph. Here is the answer: [17Y.] his practice still exists. A second figure depicting maximum and minimum temperatures is also given, and its draft is also in Caius. Though tolerably symmetrical, both curves are notably platycurtic.

It was this letter to *Nature* depicting a ‘skew’ curve that inspired Pearson’s first statistical publication, in *Nature* for 26th October 1893. ‘I have recently obtained’, he writes, ‘a generalised form of the probability curve which fits with a great degree of accuracy such curves [as Dr Venn’s], and propose to discuss it at length shortly’. Pearson then goes on to suggest using a general binomial pending his publication, estimating p, n and c (a scale parameter) by the method of moments. In her D.Phil. dissertation Eileen Magnello tells the story from the Pearsonian viewpoint, and I defer to her for a judgment as to the extent to which Venn’s paper got Pearson going.

Finally we must not forget Venn’s diagram of a random walk in two dimensions at the end of Chapter 5, which, though not strictly statistics, seems to be the first drawing of a random walk, accompanied by remarks that show that Venn realised that in the normal limit the diagram was *fractal*. (I have printed this extract in my book of readings in statistics with Herb David.)

By comparison with these efforts Venn’s lectures on ‘Theory of Statistics’ are rather disappointing, but then they were not directed at mathematicians but undergraduates in the Moral Science Tripos. The *Cambridge University Reporter*

Lecture List for the Michaelmas Term 1890 announced that Dr Venn would start his lectures on 14th October. [18.] These are the ones that Pearson regarded as the first course in statistical theory in England. They are shown as taking place in Caius, but I do not know where. My figure shows the list of topics covered. [19.] Lots about life tables, something about the binomial and the 'exponential law of error', that is, the normal distribution; Galton's ogive, or cumulative distribution representation; some animadversions on averages and smoothing. In fact little of interest except as indicating the state of statistics in England before Pearson. I have had some of the notes scanned and can show them on request.

Some of the other figures in the same folder as the lecture notes interested me as well as the barometric readings. One [20.] shows Venn forming the sampling distribution of the mean of two observations from a peculiar distribution he has invented. The same example occurs in *The Logic of Chance.* [21.] Its purpose is to warn against the assumption that taking an average is always the best policy. Venn would have enjoyed the Cauchy distribution! Another figure shows the addition of two similar normal distributions such that the result is just unimodal. [22.]

A note very much to my own taste appears on the back of page 27 of the lectures. [23.] Venn has discovered for himself Euler's identity which arises from expanding the identity $(x+y)^r(x+y)^s = (x+y)^{r+s}$. Venn has $r = s$ and writes out the case for index 4. Donald Knuth once remarked of Pascal's triangle 'There are so many relations present that when somone finds a new identity, there aren't many people who get excited about it any more, except the discoverer!' but I don't suppose Venn was very excited. (For more on Pascal's arithmetical triangle, take a look at my book of that name. [24.])

If we shift the rows we have a bivariate binomial distribution:

1	4	6	4	1	
4	16	24	16	4	
6	24	36	24	6	
4	16	24	16	4	
1	4	6	4	1	
16	64	96	64	16	= 16 times 1 4 6 4 1.

As Venn has noted, the diagonal-wise sum 1 8 28 56 70 56 28 8 1 is the binomial $(1+1)^8$. Now, we know that the bivariate binomial must turn into a zero-correlation bivariate normal distribution in the limit, with the same variance of the marginal distribution in any direction. It is amusing to see if this holds for the diagonally-generated marginal binomial. The ordinary marginals have variance $n/4$, where n is the index. The diagonal marginal has a cell interval only $1/\sqrt{2}$ of the ordinary marginal, so its variance will be $2n/4$ x $(1/\sqrt{2})^2$ or $n/4$. Thus the equality of the variances works even for small n.

But to return to Dr Venn. On the 26th May the following year, 1891, he gave a paper to the Royal Statistical Society 'On the nature and uses of averages'. It is wonderfully worthy and wholly uninteresting. Both Galton and Edgeworth attended and were suitably polite in their comments. [25.]

Venn's statistical contributions were now nearly over, but Pearson, calling himself 'a perfect stranger', remembered his work when, on 28th February 1893, he asked if Venn would be prepared to give one of four Gresham lectures on his behalf. Pearson was unable to give his as planned 'owing to the need for complete rest' and wondered whether Venn might be able to talk on 'probability curves, means, doublehumped curves, etc. in which you are specially interested'. Pearson's letters are in the Caius archives. Venn agreed to lecture on 18th April, and Pearson's letter of thanks mentions his R.S.S. lecture. Venn sent his syllabus and Pearson was delighted, adding that the other three lecturers would be Sir Robert Ball, Weldon, and Whitworth. The lecture must have been a success, for it was in the following year that Pearson applauded Venn for his Cambridge course.

That seems to be all. Venn returned to Gonville and Caius College and set about writing the college's biographical history for which, among historians, he is justly famous, the first volume of which appeared in 1897.

We shall leave him, at the age of 86, taking tea at home with R.A.Fisher, aged 30. For in the Adelaide Fisher archive I found a letter from John Venn to Fisher misfiled under John Archibald Venn, our Venn's son. He wrote in reply to a letter of Fisher's (of which there is no copy, alas) on 16th December 1920,

> Dear Mr Fisher,
> Your letter came to me like a whiff from the past. I would gladly offer advice or help, if I could, - but you know, as well as I do, that, on a subject like the Theory of Statistics, a man who is not familiar with the last thing out is almost completely helpless. Years ago I suppose I knew as much on the subject as any one going, but since I ceased to lecture I have ceased to study, & have got drifted very far away from my original work. All the same I should be very glad to have a talk with you, and, - on your second topic, - may be able to speak to the purpose.
> Are you staying on in Cambridge for some days? If so, could you come in for a cup of tea at 4.30?
>
> Yours very truly
> J.Venn

But we shall never know if Fisher came to tea.