The Enigma of Karl Pearson
and Bayesian Inference

John Aldrich
University of Southampton

Karl Pearson sesquicentenary conference, the Royal Statistical Society Friday 23 March 2007
An enigmatic position in the history of the theory of probability is occupied by Karl Pearson.*

Harold Jeffreys  *Theory of Probability* (1939)

* the theory of probability = Bayesian theory
Pearson was a Bayesian in philosophy of science but *not* in *statistics* where he used the method of moments in estimation and *P* values in testing.

HJ felt the enigma because

- He first knew Pearson as a philosopher of science
- Pearson should have written the *Theory of Probability* – he should have been me!
Historians of economics have an expression …

- In the 19th century German scholars wondered how Adam Smith could have written two very different books on ethics and economics.

- So das Adam Smith problem.

So with reference to the integrity of KP’s work

**das Karl Pearson Problem**
An easy solution to the problem?

- The philosophy that impressed Jeffreys is in the *Grammar of Science*.

- The relevant parts were written in 1891-2 *before* KP did any statistical work.

**KP changed his mind or he never had a mind to do Bayesian statistics**
but ... but ... but

- KP believed his philosophical tenets were relevant to his statistical practice and never changed his mind about this or about the tenets.

- KP did Bayesian statistical work at all stages of his career (HJ seems to have been unaware of this).

- KP even held that inverse probability is the basis of modern statistical theory.

The problem is deeper than HJ realised.
My goals

- to describe Pearson’s Bayesian philosophy of science
- to describe his Bayesian statistics
- to speculate on the link between this Bayesianism and the frequentism in most of his statistical writing
Most of his statistical writing

- KP wrote 400 papers over 40 years (1894-1936)

- 400 observations on his doing

- the doing changed

- no *Theory of Probability* — no summa
Identifying the big picture …

No treatise or textbook by KP, only

- *Tables for Statisticians and Biometricians* 1914 & 1931
- Lectures as reported by Yule 1894-6 Gosset 1907 and Isserlis 1913
- KP’s own scrappy lecture notes
Yule later reflected on his 1894-6 notes

- A straightforward, organized, logically developed course could hardly then exist when the very elements of the subject were being developed...

but to modern eyes

- there was never a “straightforward, organized, logically developed course”
Textbooks by the followers?

- W. P. Elderton *Frequency-Curves and Correlation* (1906)
  faithful but limited

  more an introduction to Yulean statistics
Retrospectives on KP’s statistics

- Egon Pearson’s biography (1936/8)
- Churchill Eisenhart’s *DSB* article (1974)

neither very detailed

Egon notices the Bayesian project(s)—*he was part of them* from 1920

Eisenhart misses them—more or less
Other impressions from the literature

- Bayes gone by 1900 or never there
  
  Stigler and Porter—general histories

- Bayes (or Laplace) always there
  
  Dale *History of Inverse Probability from Thomas Bayes to Karl Pearson*

  Hald *A History of Mathematical Statistics from 1750 to 1930*
The career

- 1870s undergraduate—Bayesian training
- 1880s otherwise occupied—elasticity & other matters
- 1890s establishing the permanent way
- 1900s the steady state—Bayes and non-Bayes coexist
- 1910s accommodating small sample theory
- 1920s Bayesian break-out
- 1930s the end
When Karl was Carl

The undergraduate Carl Pearson produced a nice set of notes on the Theory of Probability from lectures by Henry Martyn Taylor.

Our statistical inference appears under the headings

- Probability of Future Events based on Experience
- Errors of Observation & Method of Least Squares
“Of the probability of future events deduced from experience”

This section covers

- *our* Bayes’ theorem
- the rule of succession

the probability of a future success given *m* successes in *n* Bernoulli trials with a uniform prior for the probability of a success closer to *their* Bayes’ theorem
Theory of Errors and Least squares

- To find the most probable value of the true measurement is a “problem in inverse chances”

- Normal errors and uniform prior for the regression coefficients

- Ultimately Gauss’s method (1809).
Over the next decade

Pearson

- does all sorts of things

- establishes himself as an applied mathematician specialising in elasticity.

- but no probability
1890s: laying the permanent way—the lines on which thoughts travel

- probability in the *Grammar of Science* 1892
- statistical techniques needed for biometry in publications from 1894 onwards
The probability line: probability and projectability

- That a certain sequence has occurred and recurred in the past is a matter of experience to which we give expression in the concept causation.

- That it will continue to recur in the future is a matter of belief to which we give expression in the concept probability.
Subjective or objective probability?

KP after a "middle road" between

- De Morgan, Mill and Jevons who are "pushing the possibilities of the theory of probability in too wide and unguarded a manner"

and

- Boole and Venn who are "taking a severely critical and in some respects too narrow view of them."
We believe in recurrence because

- we have seen it in the past
- we can apply the rule of succession
- this is based on a uniform prior or equal distribution of ignorance but that is not unreasonable
A uniform knowledge prior: against Venn & Boole and after Edgeworth

- We take our stand upon the fact that probability-constants occurring in nature present every variety of fractional value.

- and that natural constants in general are found to show no preference for one number rather than another.
this uniform knowledge prior is unquestioned until 1917/1920

- It is in all KP’s lecture notes and in all his publications on prediction

- In 1917/20 KP challenges it in 2 opposite ways
  - Specific experience may contradict it
  - It is an unnecessary assumption
The statistical lines

- Method of moments 1894 non-Bayes
- Pearson curves 1895 non-Bayes
- Correlation & regression 1896 Bayes ?
- Probable errors 1898 non-Bayes
- $\chi^2$ and P-values 1900 non-Bayes
Estimation problems in biometry

- A mixture of normals by the method of moments (1894)
- Pearson curves by the method of moments (1895)
- Normal regression and correlation by the Gaussian method (1896)
Finding the “best value of correlation” for a bivariate normal population using the Gaussian method is straightforward.

Estimating a mixture of normals using the Gaussian method is not.
What comes naturally II

To the elastician

- Think of the distribution of the observations as a material lamina with mechanical properties—moments

- Imagine the visible lamina is formed from two invisible component laminae

- The moments of the visible lamina are determined by the moments of the invisibles and vice versa
KP’s comments on methods

- The method adopted [method of moments] has only been chosen after many trials and errors

- Other methods produce exponential equations the solution of which seems more beyond the wit of man than that of a numerical equation even of the ninth order
Horses for courses?

- KP never applied or discussed the application of the Gaussian method to estimating the Pearson curves.

- The method of moments took on a life of its own in curve fitting and perhaps KP judged the Gaussian method unsuited or impractical.
1900s: an equilibrium?

- Bayes for prediction
- non-Bayes for everything else
Equilibrium I: repeated sampling

- If a number of random samples be taken any statistical constant will vary from sample to sample, and its variation is distributed according to some law round the actual value of the constant for the total population.

- This variation would be very properly measured by the standard deviation of the constant for an indefinitely great series of random samples.

  probable errors papers in 1903, -13, -20
One and all, we act on the principle that the statistical ratio determined from our past experience will hold, at any rate approximately, for the near future.

I start as most mathematical writers have done with “the equal distribution of ignorance,” or Bayes’ Theorem.

The frequency of future samples is given by a certain hypergeometrical series, which is not at all closely approximated by a Gaussian curve ....
The fundamental problem in physics is the permanence of repeated sequences, without change of conditions.

The fundamental problem in statistics is the permanence of statistical ratios, without change of conditions.
Projectability not just for philosophers

- The medical statistician Major Greenwood did the calculations Pearson planned.

- The tables appear in *Tables for Statisticians and Biometricians* 1914.

Examples include:

In a batch of 79 recruits from a certain regiment four were found to be syphilitic. What number of syphilitics may be anticipated in a further batch of 40 recruits?
The 1917 development—out of small sample theory

- Pearson 1896/8 had found the large sample distribution of the correlation coefficient—though if you blink it’s a Bayes posterior.

- Student (1908) investigated the form of the small sample distribution.

- Soper (1913) tried to improve on Student (“I am indebted to Professor Karl Pearson for drafting the lines of this investigation and for critical supervision.”)
R. A. Fisher 1915

- In 1915 Pearson published Fisher’s paper on the extract distribution of the correlation coefficient

- KP was most struck by Fisher’s adjustment of $r$ to obtain the “most probable value” of $\rho$

$$\rho = r \left(1 - \frac{1 - r^2}{2n}\right)$$
KP on Fisher’s “use” of a uniform prior

- Statistical workers cannot be too often reminded that there is no validity in a mathematical theory pure and simple.

- Bayes’ Theorem must be based on experience, the experience where we are *à priori* in ignorance all values are equally likely to occur.

- It has unfortunately been made into a fetish by certain purely mathematical writers on the theory of probability who have not adequately appreciated the limits of Edgeworth’s justification of the theorem by appeal to general experience.
The tables permitted the construction of most probable values based on *informative* priors.

“The arithmetic involved has been of the most strenuous kind and has needed months of hard work on the part of the computers involved.”

The tables were re-issued in *Tables II* 1931.
Correlation in Small Samples. $\rho = 0.6$. Frequency curves for samples of sizes two to twenty-five, showing the changes in type from a skew "cocked hat" to J- and U-forms. Model A.
The 1920 development
“The Fundamental Problem of Practical Statistics”

An "event" has occurred \( p \) times out of \( p + q = n \) trials, where we have no a priori knowledge of the frequency of the event in the total number of occurrences. What is the probability of its occurring \( r \) times in a further \( r + s = m \) trials?

Pearson argued that the fundamental result of Laplace in no way depends upon the equal distribution of ignorance.
A lot was at stake

inverse probabilities are

- not only the basis of modern statistical theory

- but also the arithmological justification for the most ordinary actions and beliefs of the practical man.
Pearson’s argument taken in the obvious way—that the prior makes no difference—had to be wrong.

- Edgeworth and Burnside told him so.

- Pearson wrote two pieces to clarify but withdrew under a smokescreen.

  "Dr. Burnside, I venture to think, does not realise either the method in which I approach Bayes' Theorem, or the method in which Bayes approached it himself."
The 1920 development is a big psychological puzzle

- It went against decades of thinking
- It went against KP’s recent (1917) attack on Fisher

Egon who was there reflected

Perhaps it was due to a temporary lack of clearness in thought, a fault to which, I suppose, all of us succumb at times! (1938)
A population numbers $N$ individuals, of whom $p$ are marked by a special characteristic and $q$ not so.

a sample $n$ is taken and in this sample $r$ are found marked and $s$ not so.

What is the distribution of $p$?
The paper has an appendix: a note on the theory of inverse probabilities

- This revisits Venn, Edgeworth and the Cooperative Study

- It appears to endorse the practice of maximising the posterior probability obtained from a uniform prior without endorsing any of the justifications for it!
Summary and re-statement of the problem

- Bayes’ theorem always in the background

- The possibility of statistics depended on the stability of statistical ratios

- Which depended on a Bayesian argument

- The argument would come to the foreground when Pearson periodically invented a better way of calculating the integrals involved
There was glue for sticking frequentist procedures to Bayesian foundations

- Student 1908 looked for the sampling distribution for $r$ to combine it with the prior for $\rho$.

- Yule 1911 identified the posterior standard error with the sample standard error in large samples

**but** Pearson did not apply it
Pearson’s Bayesian statistics

- prediction using the rule of succession
- normal correlation
- sample to population inference

Quantitatively negligible—2 tables in 100 plus

*Tables of the Incomplete Beta Function*
Why was not there more Bayesian statistics?

- lack of a spirit of system
- excess of a spirit of problem solving
- difficulties in solving problems using the Gaussian method
- success with other methods
Enigma or case of unreasonable expectations?

End
Thanks to

Anthony Edwards
Ted Porter
Did anyone care about das Karl Pearson problem?

Probably only

- Harold Jeffreys
- Egon Pearson as KP’s student, collaborator & biographer
PS II

Did any of the work have lasting influence?

- The large sample Bayes work of 1896-8 reborn as Fisher’s large sample maximum likelihood distribution theory

- ESP writes of the 1920 error “that it acted as a stimulant on various members of his class; we discussed and argued and experimented with it, and from the new ideas which were set in train we were the gainers.”