



**Mathématiques
et sciences humaines**
Mathematics and social sciences

176 | 2006
176, Varia

Heroic periods

Périodes héroïques

Nicholas H. Bingham



Édition électronique

URL : <http://msh.revues.org/3641>
ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique
sociales de l'EHESS

Édition imprimée

Date de publication : 1 décembre 2006
Pagination : 31-42
ISSN : 0987-6936

Référence électronique

Nicholas H. Bingham, « Heroic periods », *Mathématiques et sciences humaines* [En ligne], 176 | Hiver 2006, mis en ligne le 28 juillet 2006, consulté le 05 décembre 2016. URL : <http://msh.revues.org/3641> ; DOI : 10.4000/msh.3641

Ce document est un fac-similé de l'édition imprimée.

© École des hautes études en sciences sociales

HEROIC PERIODS

Nicholas H. BINGHAM¹

RÉSUMÉ – Périodes héroïques

Cet article peut être vu comme un complément au livre de Stigler, The History of Statistics: The Measurement of Uncertainty before 1900, publié en 1986. Il est centré sur la période qui sépare les premiers travaux concernant la méthode des moindres carrés, écrits entre 1805 et 1815, et la découverte – puis le développement – des méthodes statistiques de régression et corrélation, entre 1885 et 1900. Des raisons possibles de cette « longue pause » entre deux « périodes héroïques » sont évoquées.

MOTS-CLÉS – Corrélation, Histoire des statistiques, Moindres carrés, Régression

SUMMARY – *This article complements Stigler's book of 1986, The History of Statistics: The Measurement of Uncertainty before 1900. We focus on the gap between the early work on the method of least squares, c. 1805-1815, and the discovery and development of regression and correlation, c. 1885-1900. Some possible reasons for this “long pause” between two “heroic periods” are suggested.*

KEY-WORDS – Correlation, History of statistics, Least squares, Regression

It is eighteen years now since the appearance of Stephen Stigler's book *The History of Statistics: The Measurement of Uncertainty before 1900*. Stigler's book ([Stigler, 1986] below, for brevity) breaks into three parts. Part One: The Development of Mathematical Statistics in Astronomy and Geodesy before 1827 – focuses on the introduction of the method of least squares in 1805 by Legendre (1752-1833) and Gauss and its treatment by Laplace – the Gauss-Laplace synthesis – leading up to his *Théorie Analytique des Probabilités* (TAP below) in its editions from 1812 on. It ends with the death of Laplace, and least squares and error analysis firmly established in astronomy and related sciences. Part Two: The Struggle to Extend a Calculus of Probabilities to the Social Sciences – traces events over the next half-century or so, with particular reference to the work of Quetelet, Poisson, Bienaymé and Lexis in the social sciences. Part Three: A Breakthrough in Studies of Heredity – focuses on the period 1885-1900 of the English School, of Galton, Pearson and Edgeworth, beginning with Sir Francis Galton's classic study of the heredity of human height. Galton's work gave us the terminology and concepts of regression and correlation, and – combined with the Gauss-Laplace synthesis of TAP, and developed further by the capable mathematical hands of Edgeworth and others – led to the familiar Linear Model of modern Statistics.

¹ Mathematics Department, Imperial College London, South Kensington, London SW7 2AZ, n.bingham@ic.ac.uk

On page 2 of his book, Stigler addresses the nub of the matter, in a characteristically pithy and amusing way.

For example, elementary statistics texts tell us that the method of least squares was first discovered about 1805. ... We also read that Sir Francis Galton discovered regression about 1885, in studies of heredity. Already we have a puzzle – a modern course in regression analysis is concerned almost entirely with the method of least squares and its variations. How could the core of such a course date from both 1805 and 1885? Is there more than one way a sum of squared deviations can be made small?

Our aim here is to return to Stigler's puzzle, and complement his book-length treatment from another point of view. In the compass of some ten pages, and from the point of view of the history of mathematics in general and of probability and statistics in particular, we discuss this same puzzle and advance some thoughts towards its possible resolution.

1. HEROIC PERIODS – BEFORE LEAST SQUARES

The first great puzzle that the history of probability and statistics presents us with is how mankind could have been exposed to chance events from time immemorial, and been in possession of powerful mathematical tools from the birth of the modern period onwards, with so little progress on the mathematics of chance for so long. The pre-history of our subject is discussed at monograph length in David's book *Games, Gods and Gambling* [David, 1962]. Certainly there are grounds for thinking that the pre-eminent role played in the thinking of our forefathers by matters of religion, fate and the like inhibited the application of rational, mathematical or scientific thinking to matters of chance.

The history proper of probability and statistics is usually taken as beginning with the Pascal-Fermat correspondence of 1654 – the 350th anniversary of which is aptly commemorated by this *Journée Bru*. After contributions by mathematicians of the calibre of Huygens, the first "immortal book" was the posthumous *Ars Conjectandi* (AC) of Jacob Bernoulli (1654-1705) in 1713 ("Bernoulli's theorem" – the weak law of large numbers for Bernoulli trials). The second "immortal book" was *The Doctrine of Chances* (DC) by Abraham de Moivre (1667-1754), in editions of 1718, 1738 and 1756. The normal distribution, and the Bernoulli case of the central limit theorem, date from de Moivre's note *Approximation...* of 1733, and appear in the 1738 edition of DC. Leading contributors from the rest of the eighteenth century were from Daniel Bernoulli (1700-1782), Thomas Bayes (1702-1761; *Essay* of 1763, posthumously) and Condorcet (1743-1794; *Essai* of 1785).

By this time, the first stirrings that led to least squares were being felt. One major precursor was the work of Roger Boscovich (1711-1787), who – motivated by problems of the figure of the earth – wrote on error analysis and the fitting of a linear relationship in 1760 (for background, see Stigler [1984]). Another was the work of Tobias Mayer on the libration of the moon [Stigler, 1986, Ch. 1]). Laplace also worked on error analysis in the late 1700s, motivated by problems in celestial mechanics. From this period date his use of the symmetric exponential distribution (density $\exp\{-|x|\}/2$) as an error curve, and the method of least absolute deviations (LAD). This method has enjoyed a recent revival in statistics. Laplace's magisterial *Traité de mécanique céleste* (MC) appeared (volumes 1-4) from 1799 to 1805.

2. THE FIRST HEROIC PERIOD: LEAST SQUARES

The first Heroic Period, however, concerns the introduction and absorption of the method of least squares. This first saw the light of day in 1805 in a supplement to Legendre's book *Nouvelle méthode pour la détermination des orbites des comètes*. This was followed in 1809 by Gauss in his *Theoria motus corporum coelestium in sectionibus conicis solem ambientem* (TM). Here Gauss made his famous remark:

On the other hand, our principle (principium nostrum), which we have made use of since the year 1795, has lately been published by Legendre....

Of the resulting priority dispute, we shall merely refer to Seal [1967], Plackett [1972] and Stigler [1977] for historical discussion, and point out that, in the academic and scholarly climate of those times, modern conventions regarding the supremacy of publication dates, and the need to acquaint one's contemporaries with the fruits of one's labours as they are harvested, had not yet become established.

Gauss' first work here, instead of centring on the minimization of a sum of squares (or, what is equivalent in a model with normal errors, maximization of the relevant likelihood function), assumes that the best estimate of the mean of n readings is their average (or sample mean), and uses Bayes' Theorem to *derive* the functional form of the normal law. Because of this, the normal law is often called the Gaussian law (in modern stochastic process theory, the corresponding term *Gaussian process* is always used). However, as Laplace had used the normal law earlier (as of course had de Moivre), some writers – particularly in the French school – call it the *Laplace* or *Laplace-Gauss* law.

Gauss returned to least squares in his long paper *Theoria combinationis observationum erroribus minimis obnoxiae* (TC), Part I (1812), II (1823), Supplement, 1826, in the Proceedings of the Royal Society of Göttingen. Here his earlier appeals to the "principle of the mean" and Bayes' Theorem are avoided. Instead, he interprets least squares in terms of minimizing a loss, using a quadratic loss function, an approach fully developed in modern Decision Theory. He restricted attention to linear estimators which are *consistent* (would be exact if there were no error – this is the same as unbiasedness in the normal case, but may not be so in non-normal situations). For background and discussion here, see Stigler [1975]; Sprott [1978]. Gauss also obtained the minimum-variance, or greatest-precision, property of least-squares estimators, now generally known as the Gauss-Markov theorem. (The name is a misnomer – (see [Seal, 1967] for the background on Neyman's misreading of Markov's book. But such mistakes are thematic here, as instances of Stigler's Law of Eponymy – that everything becomes known by the wrong name; see [Stigler, 1980]).

We turn now to the work of Laplace (Pierre Simon de Laplace, 1749-1827). Laplace's early work in probability and error analysis, culminating in his MC, has been touched on above. Here we come to his principal contribution, his *Théorie Analytique des Probabilités* (TAP) of 1812 (2nd ed., 1814; 3rd ed., 1820). From the last page of the Introduction (cvi pages):

On voit par cet Essai, que la théorie des probabilités n'est au fond, que le bon sens réduit au calcul... [Todhunter, 1865, p. 504].

The text is divided into:

Livre I: Du calcul des fonctions génératrices (p. 1-177),

Livre II: Théorie générale des probabilités (p. 179-461),

Additions (p. 462-484).

Augustus De Morgan comments on Laplace:

... his TAP is by very much the most difficult mathematical work we have ever met with [Todhunter, 1865, p. 539].

Ch. IV: De la probabilité des erreurs des resultants moyens d'un grand nombre d'observations, et des résultants moyens les plus avantageux (p. 304-348).

This chapter is the most important in Laplace's work, and perhaps the most difficult; it contains the remarkable theory which is called the method of least squares [Todhunter, 1865, p. 560].

Laplace has been called the French Newton. He was by far the greatest of the mathematicians considered so far, save only for Gauss, and his TAP was a milestone in probability theory of enormous importance. Some indication of this is given by the title of Todhunter's book of 1865 – A history of the mathematical theory of probability from the time of Pascal to that of Laplace – and the fact that Todhunter's last chapter (Ch. XX: Laplace) occupies pages 464-613 in his book of 624 pages. Among the many tributes to Laplace, we mention a description of TAP (quoted by Todhunter, p. 612):

to be one of the most splendid works of the greatest mathematicians of the past age.

And from Poisson's obituary of Laplace (*Comptes Rendus*, quoted by Todhunter, p. 613):

Sans doute, Laplace s'est montré un homme de génie dans la mécanique céleste...

Mais on peut dire que c'est encore plutôt dans le calcul des probabilités qu'il a été un grand géomètre...

Following Gauss' TM of 1809, there were now two ways in which the normal law entered the theory:

- a) via the Central Limit Theorem (in modern terminology), specifically its special case for Bernoulli trials, the de Moivre – Laplace theorem,
- b) through the link with the method of least squares.

Laplace recognised the importance of (a) and (b) together in a Supplement added to a paper of his in 1810, and incorporated this "Gauss-Laplace synthesis" into his TAP of 1812.

A detailed account of the Gauss-Laplace synthesis (Stigler's telling phrase) is contained in Chapter 4 of Stigler [1986].

3. THE MID-NINETEENTH CENTURY: HEROIC PERIOD OF MATHEMATICS

We turn now from our specific concern – the history of probability and statistics – to consider the history of mathematics in general. This is partly by way of setting context,

partly in order to contrast the very different situations in mathematics at large and in probability and statistics during this period.

The turn of the nineteenth century witnessed the emergence of Gauss (1777-1855), arguably the greatest mathematician who ever lived and certainly among the top three (with Newton and Archimedes). The first years of the century saw the epoch-making introduction of least squares, culminating in the Gauss-Laplace synthesis. The Napoleonic era saw, as well as extensive blood-letting and political upheaval, much constructive work in the sciences and the educational system. Thus Napoleon acted as a major patron of the sciences – honouring such figures as Laplace, Fourier and Monge, for example. He imposed the metric system on those parts of continental Europe he controlled. He founded the French *Grandes Ecoles*, including the *Ecole Polytechnique* at which Monge taught. With these new institutions, and the mathematical legacy of the great French mathematicians of the time – Lagrange, Laplace, Legendre, Fourier and their contemporaries – new mathematical textbooks appeared, and helped to confirm the place of France as the leading mathematical country of the time.

The stage was set for an heroic period in mathematics unequalled since the birth of the modern scientific age with Newton and Leibniz. It would take a book to do any sort of justice to this; we content ourselves here with the briefest of notes on eight major fields. Our aim is merely to substantiate that this period does indeed dominate the undergraduate curriculum, and so was indeed an heroic period for mathematics.

3.1. ANALYSIS

In analysis, Cauchy (1789-1857) transformed the field by his development of complex function theory, complex integration and the calculus of residues. Rigorous analysis gradually established itself, through the work of Bolzano (1781-1848), in his 1817 book *Rein analytische Beweis...*, and leading up to the work of Weierstrass (1815-1897), most influential during his own lifetime through his lectures at the University of Berlin.

3.2. ALGEBRA

The work of Abel (1802-1829) and Galois (1811-1832) gave mathematics the group concept. The work of Cayley (1821-1895) and Sylvester (1814-1897) in the 1850s introduced matrices, and led to linear algebra.

3.3. GEOMETRY

The work of Poncelet (1788-1867) systematized the projective geometry that was created earlier by Desargues. The non-Euclidean geometry of Lobachevskii (1793-1856) and Bolyai (1802-1860) in the 1830s changed the conceptual framework with which one thinks about space, and demonstrated a dramatic break with habits of thought established over two thousand years. No clearer illustration of the movement to the modern world could be given in mathematics.

3.4. NUMBER THEORY

The work of Riemann (1826-1866) of 1859 on analytic number theory applied the complex analysis of Cauchy to problems on the distribution of primes, and gave us the

Riemann Hypothesis still open today. The new algebra led, in the hands of Kronecker (1823-1891), Kummer (1810-1893) and Dedekind (1831-1916), to the study of ideals and algebraic number fields.

3.5. MECHANICS

The Newtonian mechanics of the *Principia* (1687) led on to the Lagrangian mechanics of the *Mécanique Analytique* (1788) and to Hamiltonian mechanics (1834). One spectacular triumph of Newtonian mechanics was its use in the discovery by Adams and Leverrier of the planet Neptune in 1845, from perturbations in the motion of Uranus.

3.6. ELECTROMAGNETISM

The modern age in electromagnetism might be dated from the essay of 1828 of George Green (1793-1841), *An essay on the application of mathematical analysis to the theories of electricity and magnetism* – itself heavily influenced by the work of the French school, of Laplace, Poisson and others. The discovery by Michael Faraday (1791-1867) of electromagnetic induction in 1831, and by James Clerk Maxwell (1831-1879) of the electromagnetic theory of light in 1864, were perhaps the two greatest single scientific advances of the nineteenth century.

3.7. STATISTICAL MECHANICS

The field of statistical mechanics rests principally on the work of three workers, Maxwell (above), Ludwig Boltzmann (1844-1906) and Josiah Willard Gibbs (1839-1903). This marriage of mechanics with statistical physics, together with the closely related field of thermodynamics, is still the basis of much science and was a precursor to the quantum theory of the next century.

3.8. PARTIAL DIFFERENTIAL EQUATIONS (PDE)

Much of classical physics can be expressed in the language of PDEs in general, and second-order linear PDEs with constant coefficients in particular. The classical trichotomy into elliptic (prototype: Laplace's equation), parabolic (prototype: the heat equation) and hyperbolic (prototype: the wave equation), and the development of the associated machinery of eigenfunction expansions, special-function theory and the like, dates from the mid-nineteenth century.

4. THE MID-NINETEENTH CENTURY: HEROIC PERIOD IN ENGINEERING, IN INDUSTRY, ETC.

The beginning of the nineteenth century saw the end of the old Europe of the Middle Ages, propelled into upheaval by political events – the French Revolution, the Napoleonic wars, etc. Even 1825, say, witnessed the old world recovering from this turmoil. But by fifty years later, we see instead an early form of the recognisably modern world of today. The Industrial Revolution had burgeoned. Railways had begun; so had deep coal mining (to provide fuel for the new steam age, as well as for domestic heating); so had steel smelting. The great works of Victorian engineering that still survive had begun: railway and road bridges of new types and unprecedented size,

steamships, tunnels through mountains and under rivers, public provision of such necessities of urban life as water and sewerage supply, the first skyscrapers, etc. All this was made possible by the application, and absorption into manufacturing, industrial and commercial life, of the first century and a half or so of the modern scientific and mathematical world. It is often said, and rightly so, that this great age of engineering was the age of Newtonian science applied.

During this time, we also see profound changes in institutions. The modern university emerged – new metropolises such as London acquiring their own universities, and the mediaeval seats of learning such as Oxford and Cambridge adapting to changed circumstances by radical revision of curricula and of access. The modern institutions of capitalism – the stock exchange, the limited liability company, etc. – emerged. The great industrial nations rushed to colonize the world (leading to the cataclysm of World War One in due course). These changes led to political repercussions – the emergence of the modern left-right divide in politics, etc.

By the last quarter of the century, we see a world as like that of a century later as of half a century earlier. We see a world in which universal suffrage is accomplished or coming, in which mass education is established, and in which the role of science and technology is recognised as decisive. We see a social contract in which governments recognise a duty to provide for their electorates, both for electoral survival and out of recognition of the need to nurture and develop the nation's human capital in the new and dynamic conditions. In short: we see a new world, in which everything has changed beyond recognition in half a century.

5. PROBABILITY AND STATISTICS: THE LONG PAUSE

Conspicuous by its failure to move with the hectic pace of change of these times is the field of probability and statistics. To return to Stigler's telling remark quoted earlier, we find this field in 1885 much as we left it in 1805.

Stigler devotes Part Two of his book – *The Struggle to Extend a Calculus of Probabilities to the Social Sciences* – to answering the question of why this period, so fertile elsewhere, should have been so comparatively barren here. He traces the work of Quetelet (1796-1874), his application of statistics to measurements of human heights and the like, and social questions concerning conviction rates etc., and his tendency to fit a normal distribution to any data set to hand (“Quetelismus”). He considers the work of Poisson (1781-1840), his social studies – *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, 1837 – and the origin of the Poisson distribution. The work of Bienaymé (1796-1878) is also briefly considered; (see [Heyde, Seneta, 1972]) for a book-length treatment.

Interesting as these contributions and their authors are, these developments hardly compare, either with those in mathematics during this period or with those in the broader spheres of engineering, industry and commerce or politics.

Meanwhile, the birth of the splendid Russian school of probability that was to prove so decisive in the next century had taken place. Probability was first studied in Russia at the Academy of Saint Petersburg (the capital in Tsarist days) by Euler and Daniel Bernoulli in the 18th century. The pre-eminence of the Russian school may be traced to the work of P. L. Chebyshev (or Tchebychev, to give just one other

transliteration) (1821-1894). Chebyshev gave what is now called Chebyshev's inequality in 1867 (*J. Math. Pures Appl.* 88, p. 177-184) – although this result was anticipated by Bienaymé in 1853. Later, he introduced the method of moments (1890-91). His most distinguished pupil was A. A. Markov (1856-1922), who succeeded him in 1883. Markov's 1912 book *Wahrscheinlichkeitsrechnung* gives us the concept of a Markov chain, and brings us to within sight of the modern age.

6. OUR SECOND HEROIC PERIOD – THE ENGLISH SCHOOL

Gauss had most of the essentials for the Linear Model, the core material for modern statistics. Key ideas that he lacked were

- a) regression and correlation,
- b) exact sampling distributions (chi-square, t , F etc.)

One of the key components – the bivariate normal distribution – was studied in a long paper by A. Bravais (1846): *Analyse mathématique sur les probabilités des erreurs de situation d'un point. Mém. Acad. Roy. Sci. Inst. France* 9, p. 255-332.

But this still did not lead to the required mathematical breakthrough. A modern treatment uses linear algebra and matrices. The mathematics for this did not then exist – as above, it was introduced in the next decade by Cayley and Sylvester; this may partly explain the delay.

Gauss, Laplace and Legendre were all motivated by the needs of *celestial mechanics*. The area which most stimulated further progress was *biology*. This had been revolutionized by Darwin's theory of natural selection (Charles Darwin (1809-1882); *The Origin of Species by Natural Selection*, 1859). The pioneer of the second heroic period in this account was the unlikely figure of Sir Francis Galton (1822-1911). Galton

... latterly devoted himself to heredity, founding and endowing the study of eugenics... (Chambers' Biographical Dictionary).

Galton exemplifies the supreme importance – in science, mathematics or any other field of enquiry – of *asking the right questions*. His first work relevant here was his book *Hereditary Genius: An Enquiry into its Laws and Consequences* of 1869. Motivated by an attempt to introduce quantitative methods into the study of heredity (this *after* Mendelian genetics was initiated – by Gregor Mendel (1822-1884) in 1865, but *before* it was rediscovered and widely noticed in 1900; see Mendel [1965], Galton undertook a statistical study of the heights of adult males, say (the “response variable”, y – actually, Galton used females also, but adjusted their heights to allow for gender) plotted against the average of the father's and mother's heights (mid-parental height – the “predictor variable”, x). His data set consisted of 928 adult offspring, born of 205 mid-parents. On plotting a histogram, Galton noticed empirically that the contours seemed elliptical in shape. Galton was himself a weak mathematician, but he obtained the help of a Cambridge mathematician, J. Hamilton Dickson, who was able to apply to Galton's study the relevant mathematics, that of the bivariate normal distribution (already studied by Bravais in 1846). Galton's study of 1885 was published in 1886. He concluded that, of every inch of mid-parental height above (or below) the mean, *on average* q inches were transmitted to the next generation, for some constant q between 0 and 1. Continuing in this way, after n generations, on average q^n inches are passed on. Since this tends to 0 as n increases, this led Galton to the – to him discouraging, in view

of his fascination with hereditary genius – conclusion that height (and by analogy, hereditary genius, intelligence etc.) *reverts towards the mean*. The coefficient ρ above – visibly a measure of linkage, or association, between x and y – Galton called the coefficient of co-relation. These Galtonian concepts of reversion and co-relation have been transmitted to us as *regression and correlation*. His work was published at book length in *Natural Inheritance* (1889). The histogram and elliptical plot from Galton's original data are reproduced in Stigler [1986, p. 286-287], and a full account of Galton's work, its motivation and historical context in Stigler [1986, Chapter 8 – The English Breakthrough: Galton].

Galton's work provided the stimulus for that of Karl Pearson (1857-1936). Pearson became Professor of Applied Mathematics at University College, London in 1884, and Galton Professor of Eugenics there in 1911. (One notes in passing the enthusiasm of progressives and scholars of these and later times for the term eugenics, long before it sank beneath the weight of opprobrium that attached to it following Nazi times.) Pearson wrote his influential *The Grammar of Science* – a textbook on statistics – in 1892 (I cannot resist noting in passing the parallel between Pearson's description of statistics in these terms and that of a genius of contemporary statistics, Persi Diaconis, who describes statistics as *the physics of numbers*). Pearson was much influenced by W. F. R. Weldon (1860-1906), Jodrell Professor of Zoology at University College, London (1890-1900) and founder-editor of the journal *Biometrika* in 1900. Pearson was noted for his introduction of the Pearson family of density curves of various types, which he used to describe characteristics of the shapes of shell-fish obtained by Weldon from the Bay of Naples. Pearson re-discovered and named the chi-square distribution (found earlier by Helmert in 1876), and used it as the basis for his *chi-square goodness-of-fit test* of 1900 – arguably, the beginning of modern statistics. His historical contributions include his *Life of Galton* (1914, 1930) and his *The History of Statistics in the 17th and 18th Centuries* (1978).

The next major figure of this period was F. Y. Edgeworth (1845-1926), who was born in Ireland, studied in Dublin and Oxford, and taught at King's College, London. Edgeworth worked on statistics from 1885 on, influenced in particular by Galton and Pearson. It is to Edgeworth, in two papers of 1892 and 1893, that the basic distribution theory of the Multivariate Normal distribution, the theoretical underpinning of modern Multivariate Analysis, is due. A full account of Edgeworth's work in statistics, and its importance, is given in Stigler [1986, Chapter 9]. Note that by this time, the apparatus of matrices and linear algebra, the natural language for describing multivariate distributions, was well developed and in use. Edgeworth's work is striking to the reader of today for its modernity.

This work was continued by G. U. Yule (1871-1951), a pupil of Pearson. Yule links the figures of this Heroic Period – the English breakthrough – with the major figures of British statistics in the 20th century. For instance, his *An Introduction to the Theory of Statistics* of 1911 became, from its eleventh edition of 1937 on, “Kendall and Yule”, co-authored with M. G. Kendall, a fore-runner of Kendall's *The Advanced Theory of Statistics* of 1943, which became “Kendall and Stuart” (and later, the “Kendall, Stuart and Ord” of today). A detailed appreciation of Yule's work is in Stigler [1986, Chapter 10: Pearson and Yule].

Pearson worked extensively on the distribution theory of the sample correlation coefficient r (the sample version of ρ). But the definitive step here was taken by Pearson's real successor, R. A. Fisher (1880-1962, later Sir Ronald Fisher), who found

the sampling distribution of r . Fisher's work is of enormous interest, partly because he was undoubtedly the greatest statistician who ever lived, partly because his work takes off so directly from that of the English breakthrough described above. But one must end somewhere. The second Heroic Period just described may be thought to end as well in 1900 as anywhere else – British statistics has continued to demonstrate great intellectual vitality ever since – and we will close our account, as Stigler did his, in 1900.

7. THE LONG PAUSE: WHY DID IT HAPPEN?

It is always difficult to “prove a negative”. The Long Pause *did* happen, and it stands in stark contrast with the Heroic Periods described above – in probability and statistics both before it and after, in mathematics during it, and in a number of other areas of human endeavour, including engineering, during it also. To go further is to speculate, which while interesting is also dangerous. We will content ourselves here with a few comments by way of summary.

First, the theme of Stigler's book of 1986 is to analyse the relationship of the Long Pause to its preceding and succeeding Heroic Periods (not by these names), with particular reference to the combination of observations and the use of probability models in inference, and the “horizontal” dimension – spread of ideas between disciplines – and the “vertical” – the development of probability from its humble beginnings to its modern form, and its symbiotic relationship to statistics and questions of inference. Stigler's thesis is given both succinctly in the Introduction to his book (p. 1-8) and in detail in Part Two: The Struggle to Extend a Calculus of Probabilities to the Social Sciences (p. 161-261).

Second, however great one's devotion to probability and statistics in general and to their remarkable history in particular, one must concede that their importance is outweighed by that of mathematics – not to mention science, engineering and human affairs more generally. It is at least arguable that one needs look no further for an explanation of the Long Pause than the remarkable – even explosive – developments in mathematics during this period. No doubt the diversion of the intellectual energies of our fore-fathers during this time into laying down the core material of the modern mathematics curriculum played at least some part in the absence of major developments in probability and statistics during the same period.

Thirdly, the Long Pause, and the circumstances of its ending, clearly demonstrates the vital importance in scientific enquiry of *asking the right questions*. The great men of the first Heroic Period were motivated by celestial mechanics – central to the Newtonian tradition. This tradition is clearly dominant in the story of the superb triumph of the detailed and recondite work concerning the Adams-Leverrier affair over Neptune in 1845, for example. Whether Gauss, Laplace and Legendre would have warmed to the study of the social and biological sciences by quantitative means we shall never know. What we do know is that Galton, an insignificant mathematician but an inquisitive scientist driven by an urge to understand heredity, was able to go beyond the achievements of these great mathematicians, and complete them by the concepts of regression and correlation, thereby giving us the foundations for the Linear Model ubiquitous in the statistics of today. Any serious study of the heredity of human height, for example, would rapidly and inevitably have led to Galtonian regression and correlation.

Fourthly, the crucial importance of the availability of the key mathematical tools – be they concepts, techniques or even merely notations – is illustrated here. Edgeworth’s Theorem of 1893, giving the functional form of the multivariate normal density in terms of the relevant parameters – the elements of the mean vector and the covariance matrix – needs for its very formulation the machinery of linear algebra – vectors, matrices and linear transformations – that was available to him but not to Gauss and his contemporaries, or to Bravais in 1846 (even though Gauss had another necessary ingredient, the determinant). One specific case may illustrate the importance of notation. Laplace used the symmetric exponential – density $\exp\{-|x|\}/2$ – as an error curve before his adoption of the normal. It is salutary to realize that this great man was hampered by lack of a proper notation for the absolute value sign [Stigler, 1986, p. 111], which was introduced only later, by Weierstrass in 1841 [Cajori, 1993, vol. 2, p. 123]. Finally, if any underlining of the need to *find the right notation* were needed, we refer the reader to Pólya’s magnificent book *How to Solve It* [Pólya, 1957].

7.1. POSTSCRIPT

We content ourselves here with two closing remarks.

1. The journal *Biometrika* mentioned above – founded by Weldon in 1900 – has always had an interest in the history of probability and statistics. This has found expression in a series of papers *Studies in the history of statistics and probability*, published in its pages at intervals since. Two volumes have been published in book form, *Volume I* (ed. E. S. Pearson & M. G. Kendall) and *Volume II* (ed. Sir Maurice Kendall & R. L. Plackett) (Charles Griffin, London, 1970, 1977 – “Studies I” and “Studies II” in the references). The series of papers continues. My recent contribution *Studies XLVI* – [Bingham, 2000] – takes the history of probability from 1900 to 1933, and makes my decision to stop here at 1900 the easier.
2. The subject-matter here centres on the role of the method of least squares. I first met this a final-year undergraduate (Oxford, 1965-6), learning statistics from Plackett’s *Regression Analysis* Plackett [1960], a valuable if somewhat terse exposition. The first sentence of the Preface reads:

The field of regression analysis is here supposed to consist of the algebraic theory and numerical methods associated with the principle of least squares, its application in the analysis of experimental data, and the construction of designed experiments.

I remember puzzling as a student as to why least squares should be called regression. When, long afterwards, I not only knew this, but found myself teaching this material, I determined that I would make quite sure that all my students at least knew this. My continuing interest in the history of our field stems from this. It is remarkable that the author of Plackett [1960] is – or developed into – Robin Plackett the historian of our subject.

It is a pleasure to close by saluting the work of Bernard Bru and his colleagues over twenty-five years of the Paris Seminar on the history of probability and statistics, and to express my pleasure in participating in this *Journée Bru*.

REFERENCES

- BINGHAM N. H., "Measure into probability: from Lebesgue to Kolmogorov. Studies in the history of statistics and probability", XLVI, *Biometrika* 87, 2000, p. 145-156.
- CAJORI F., *A history of mathematical notations*, Volumes 1, 2, Dover, 1993.
- DAVID F. N., *Games, Gods and Gambling*, London, Charles Griffin, 1962.
- HEYDE C. C., SENETA E., *I. J. Bienaymé: statistical theory anticipated*, New York, Springer, 1977.
- MENDEL G., *Experiments in Plant Hybridization*, Edinburgh, Oliver & Boyd, 1965.
- PLACKETT R. L., *Regression Analysis*, Oxford, Oxford University Press, 1960.
- PLACKETT R. L., "The discovery of the method of least squares", *Studies II*, 17, 1972.
- POLYA G., *How To Solve It*, New York, Doubleday, 1957, [Princeton University Press, 1945].
- SEAL H. L., "The historical development of the Gauss linear model", *Studies I*, 15, 1967.
- SPROTT D. A., "Gauss' contributions to statistics", *Historia Mathematica* 5, 1978, p. 183-203.
- STIGLER S. M., Napoleonic statistics: the work of Laplace, *Biometrika* 62, 1975, p. 503-517.
- STIGLER S. M., "An attack on Gauss, published by Legendre in 1820", *Historia Mathematica* 4, 1977, p. 31-35.
- STIGLER S. M., "Stigler's law of eponymy", *Transactions of the New York Academy of Sciences*, 2nd series, 39, 1980, p. 145-157.
- STIGLER S. M., "Boscovich, Simpson, and a 1760 manuscript note on fitting a linear relation", *Biometrika* 71, 1984, p. 615-620.
- STIGLER S. M., *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge (MA), Harvard University Press, 1986.
- TODHUNTER I. (1865): *A History of the Mathematical Theory of Probability, from the Time of Pascal to that of Laplace*, London, Macmillan, 1865, [New York, Chelsea, 1949].