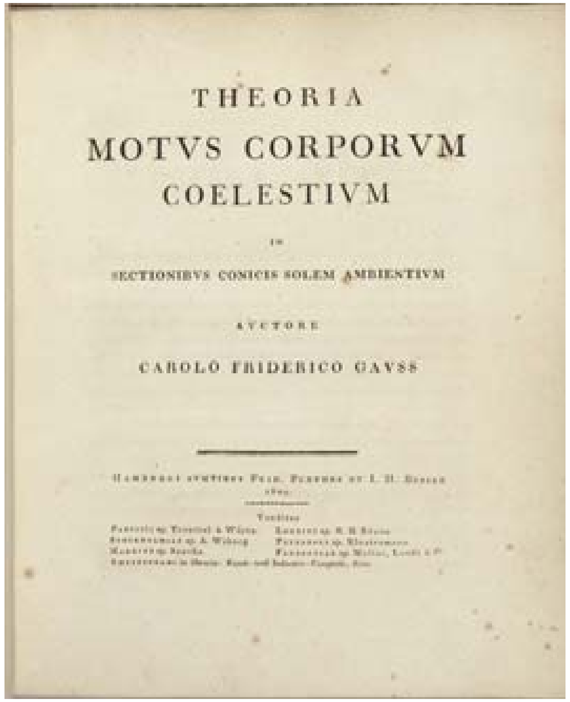
From 1809 and to 1925—threads in the theory of errors


THEORIA
MOTVS CORPORVM
COELESTIVM
IN
SECTIONIBVS CONICIS SOLEM AMBIENTIVM
AVCTORE
CAROLO FRIDERICO GAVSS


Statistical Methods for
Research Workers
BY
R. A. FISHER, M.A.
OLIVER AND BOYD
EDINBURGH: TWEEDDALE COURT
LONDON: 33 PATERNOSTER ROW, E.C.
1925

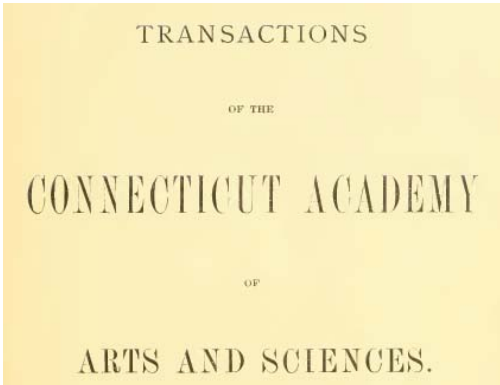
John Aldrich
University of Southampton

Seminar November 2010 Oxford

—Size

In 1877 Merriman found

- 408 items—earliest in 1749 (Euler)
- 153 in German, 110 in French, 90 in English, 16 in Latin…
- 13 “proofs of the method of least squares”



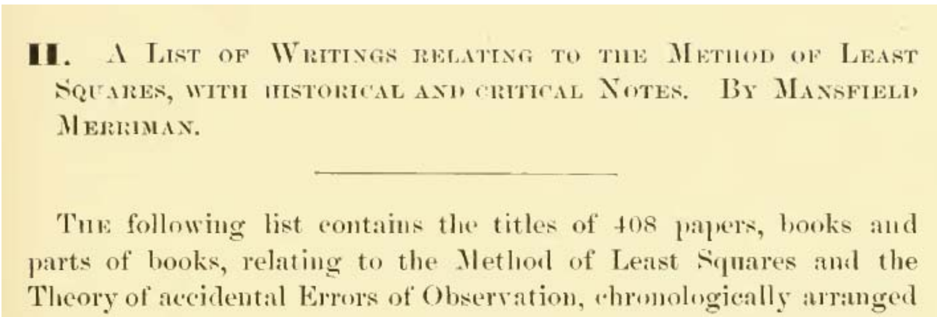
TRANSACTIONS

OF THE

CONNECTICUT ACADEMY

OF

ARTS AND SCIENCES.



II. A List of Writings relating to the Method of Least Squares, with historical and critical Notes. By Mansfield Merriman.

The following list contains the titles of 408 papers, books and parts of books, relating to the Method of Least Squares and the Theory of accidental Errors of Observation, chronologically arranged

I follow 2 threads—one with 3 main characters, the other with 2 mains

The talk draws on 2 papers:

- Lüroth, Dedekind and the Bayesian line in the theory of errors
- The theory of errors reborn as mathematical statistics—Fisher and Student 1908-25

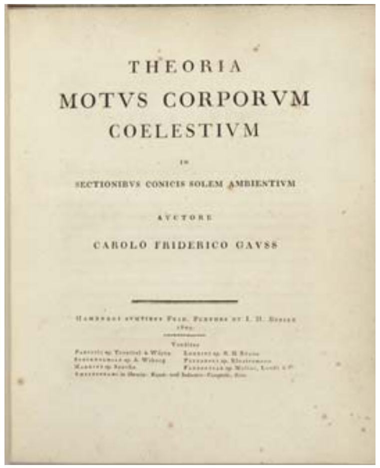
but treats only one issue

- inference to the mean of the normal distribution—ultimately the *t*-distribution

The threads

- Thread 1 begins in astronomy and is developed by mathematicians
- Thread 2 ends in one of a series of biological monographs and manuals
- mainly the work of a mathematician but it has come from laboratory work in a brewery.
- The threads are tenuously connected because thread 2 came from a corruption of thread 1…

Thread I : C. F. Gauss (1777-1855) mathematician, astronomer, geodesist


THEORIA
MOTVS CORPORVM
COELESTIVM
IN
SECTIONIBVS CONICIS SOLEM AMBIENTIVM
AVCTORE
CAROLO FRIDERICO GAVSS

- *Theoria Motus* (1809) introduces model for indirect observations

$$y \,|\, \beta, \sigma^2, X \sim N(X\beta, \sigma^2 I)$$

- Precision parametrised by

$$h = 1/\sigma\sqrt{2}$$

Uniform prior for β → max prob value given by least squares

Precision of βs given relative to h

But no inference on h.

W. F. Bessel (1784-1846) astronomer, geodesist

Ueber den Ort des Polarsterns, von Herrn *Friedr. Wilh. Bessel* Prof. der Astr. in Königsberg.

Unterm 29. Jun. 1815 eingesandt.

- Introduced the probable error
- mean ± probable error gives a 50% interval for a normally distributed quantity
- Bessel's estimate of p.e. based on mean error

$$0.8453 \frac{\sum |y_i - \bar{y}|}{m}$$

(direct observations)

Gauss 1816

Bestimmung der Genauigkeit der Beobachtungen.

....

Art. 176): oder die Wahrscheinlichkeit jedes Werthes von h ist der Grösse

$$h^m e^{-hh(\alpha\alpha+\beta\beta+\gamma\gamma+\text{ u.s.w.})}$$

proportional. Der *wahrscheinlichste* Werth von h ist folglich derjenige, für welchen diese Grösse ein Maximum wird, welchen man nach bekannten Regeln

$$=\sqrt{\frac{m}{2(\alpha\alpha+\beta\beta+\gamma\gamma+\text{ u.s.w.})}}$$

....

Es ist also eins gegen eins zu wetten, dass der wahre Werth von h

zwischen $H(1-\frac{\rho}{\sqrt{m}})$ und $H(1+\frac{\rho}{\sqrt{m}})$

- Estimates h.
- No mean or regression parameters
- Maximum posterior value of h assuming implicitly a uniform prior.
- 50% limits for h (normal approximation)

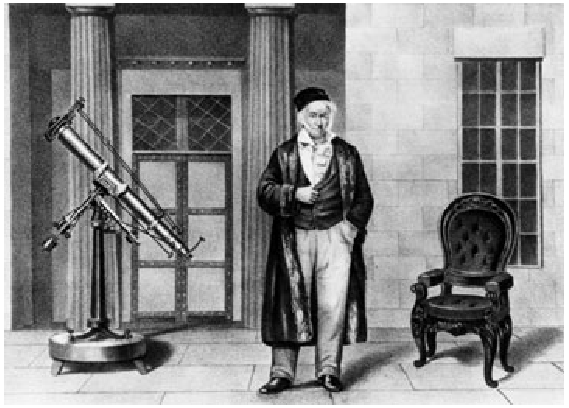
Then a change of direction

- Large sample properties of alternative estimators of h. Mean square estimator most efficient (Fisher-speak)

Second proof of least squares—Gauss 1822-3

- New frequentist approach to least squares based on "Gauss-Markov theorem."
- This was not widely taught in the nineteenth century.
- An innovation that *did* take root involved the estimation of precision: instead of sample size use the number of degrees of freedom (Fisher-speak).

Gauss taught least squares—both proofs—most years until his death in 1855



Students

—astronomers including

- Schumacher (1780-1850) who founded *Astronomische Nachrichten*
- Encke—see below

—mathematicians including

- Dedekind (1831-1917)—see below.

Richard Dedekind, student 1850

1860 DEDAKIND. 'Ueber die Bestimmung der Präcision einer Beobachtungsmethode nach der Methode der kleinsten Quadrate.' *Vierteljahrs. Gesell. Zürich*, Vol. V, pp. 76–83.

The most probable value of the measure of precision h is found to be $\sqrt{\frac{n-1}{2\sum x^2}}$ and not $\sqrt{\frac{n}{2\sum x^2}}$. See 1866 BÖRSCH.

quadraten ansieht. Der wahrscheinlichste Werth zu der Präcision ist daher wirklich

$$= \sqrt{\frac{m-n}{2\Omega_0}}.$$

…

$$\frac{c\,dx}{(X^2+\Omega_0)^{\frac{m-n+2}{2}}}$$

or $t(m-n+1)$

1860 paper

- has regression and precision parameters.
- maximises marginal posterior.
- focuses on precision.
- Confirms that least squares gives the most prob value of marginal posterior—for which he needs that marginal.

J. F. Encke (1791-1865) Gauss student 1811-13, astronomer, populariser and combiner

1832 ENCKE. 'Ueber die Methode der kleinsten Quadrate.' *Berlin. Astron. Jahrbuch* for 1834, pp. 249–312; for 1835, pp. 253–320; for 1836, pp. 253–308. —Republished in *Encke's astronomische Abhandlungen* (Berlin, 1866), Vol. I, Nos. xii, xiii, xiv.

These memoirs form a treatise on the Method of Least Squares, from which many text-books have been compiled.

Gauss and Bessel "combine the strictest theory with the happiest practical applications of theoretical truths."

- Gauss (1809): maximum probability estimate
- Gauss (1823): estimate of precision
- Bessel: probable error

Dedekind followed up: Jacob Lüroth (1844-1910) geometer with an interest in astronomy

Astronomische Nachrichten.

Expedition auf der Königlichen Sternwarte bei Kiel.

Herausgeber: Prof. Dr. C. A. F. Peters.

Bd. 87. Nr. 2078. 14.

Vergleichung von zwei Werthen des wahrscheinlichen Fehlers.

....

$$r_1 > R_1 > r_1 \sqrt{\frac{p}{p+1}}$$

ist; so dass die hier durchgeführte Methode der Berechnung des wahrscheinlichen Fehlers ihn stets kleiner liefert als die gewöhnliche, aber höchstens so, als ob eine Beobachtung mehr vorhanden wäre und die letztere Methode der Berechnung angewandt würde.

Carlsruhe, den 27. December 1875.

J. Lüroth.

- Compares probable error formula based on Dedekind distribution (R_1) with the usual (Encke) formula (r_1): p degrees of freedom
- No big deal

End of thread I

Merriman comments on Lüroth

> 1876 Lüroth. 'Vergleichung von zwei Werthen des wahrscheinlichen Fehlers.' *Astron. Nachr.*, Vol. LXXXVII, col. 209–220.
>
> The usual formula is compared with a new formula and shown to give larger values.

- Lüroth disappears until Pfanzagl and Sheynin (1996)--" A Forerunner of the *t*-Distribution"
- Similar analyses from Edgeworth (1883), Burnside (1923) and Jeffreys (1931)
- Perhaps elsewhere too?

A new thread … Student (Gosset) & Fisher

- Early C20 Britain a sleepy time in error theory
- A lively time in biometry
- Leading biometrician Karl Pearson was not much interested in error theory
- But he got entangled
- Student sought Pearson’s help on an error theory problem
- Later Fisher applied error theory ideas to Pearson’s projects

William Sealy Gosset chemist and brewer

THE APPLICATION OF THE "LAW OF ERROR" TO THE WORK OF THE BREWERY.

3rd November, 1904.

The following report has been made in response to an increasing necessity to set an exact value on the results of our experiments, many of which lead to conclusions which are probable but not certain. It is hoped that what follows may do something to help us in estimating the Degree of Probability of many of our results, and enable us to form a judgment of the number and nature of the fresh experiments necessary to establish or disprove various hypotheses which we are now entertaining.*

- E. S. Pearson: "All this is simply Airy or Merriman put by Gosset into the form most useful for his fellow brewers."
- A bit more ingenious but no real innovations

Airy or Merriman …

1861 AIRY. '*On the Algebraical and Numerical Theory of Errors of Observations and the Combination of Observations.*' Cambridge and London, 8vo, pp. xvi, 103. --Second ed., 1875.

Only LAPLACE's *Théorie analytique des Probabilités* was consulted in preparing this book, and as a consequence it is unreadable except by those already thoroughly acquainted with the subject.

A TEXT-BOOK

ON THE

METHOD OF LEAST SQUARES.

BY

MANSFIELD MERRIMAN,

MEMBER OF AMERICAN MATHEMATICAL SOCIETY.

- Merriman descended from Encke via Chauvenet: the Bayesian element attenuated.
- Airy not at all Bayesian—but not very informative about practice

1905 Gosset consults Karl Pearson, biometrician

VII. *Mathematical Contributions to the Theory of Evolution.*—IV. *On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation.*

By KARL PEARSON, *F.R.S., and* L. N. G. FILON, *B.A., University College, London.*

Pearson wasn't interested in error theory but had lots of ideas

- Bayesian arguments
- significance testing
- fitting by method of moments
- large sample theory
- correlation

1908 two half baked studies?

VOLUME VI　　MARCH, 1908　　No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

- Obtains sampling distribution
- Uses it as a posterior distribution

so that the odds are about 33 : 1 that small corn gives a better yield than large. These odds are those which would be laid, and laid rightly, by a man whose only knowledge of the matter was contained in the two experiments. Anyone con-

- hint here of analysis with uninformative prior?
- incomplete too?

PROBABLE ERROR OF A CORRELATION COEFFICIENT.

BY STUDENT.

AT the discussion of Mr R. H. Hooker's recent paper "The correlation of the weather and crops" (*Journ. Royal Stat. Soc.* 1907) Dr Shaw made an enquiry as to the significance of correlation coefficients derived from small numbers of cases.

- Obviously incomplete
- Fails to obtain sampling distribution for correlation—except for null case
- Bayesian scheme of combining sampling distribution with prior cannot be realised

The mean paper's distributions: *s* and *z*

I. The equation is determined of the curve which represents the frequency distribution of standard deviations of samples drawn from a normal population.

…

Proceeding in the spirit of Pearsonian curve fitting but calculating theoretical moments

Consequently a curve of Professor Pearson's type III. may be expected to fit the distribution of s^2.

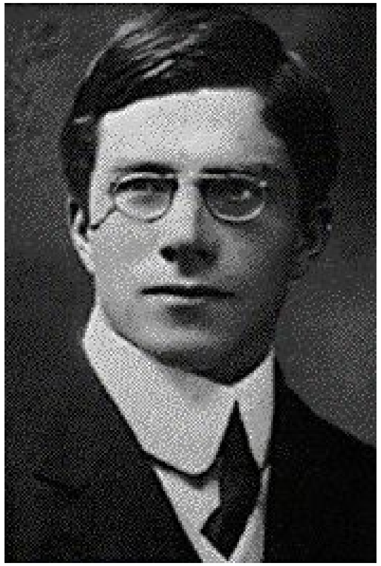
…

III. The equation is determined of the curve representing the frequency distribution of a quantity z, which is obtained by dividing the distance between the mean of a sample and the mean of the population by the standard deviation of the sample.

Following Pearson, the estimate *s* divides by the sample size.

Following nobody, the z ratio divides by s and *not* by $s/\sqrt{n}$.

1912 A half baked paper on estimation



ON AN ABSOLUTE CRITERION
FOR FITTING FREQUENCY CURVES.

By *R. A. Fisher*, Gonville and Caius College, Cambridge.

1. If we set ourselves the problem, in its essence one of frequent occurrence, of finding the arbitrary elements in a function of known form, which best suit a set of actual observations, we are met at the outset by an arbitrariness which appears to invalidate any results we may obtain. In

...

But we may solve the real problem directly.

Fisher proposed the de-Bayesed method of Merriman as a universal criterion—the future maximum likelihood.

Its virtue—absoluteness = invariance.

The Bayesian approach was rejected

Common ground with Student?

Evidently not—BUT

- the absolute criterion existed in 2 versions
- 2nd version: to estimate *h* use the density of μ^2 (sample variance).

160 *Mr. Fisher, On fitting frequency curves.*

If a frequency curve of unit area were drawn, showing the frequencies with which different values of μ^2 occur, for a given h, and if b were the ordinate corresponding to the observed μ^2, then we should expect the equation

$$\frac{\partial b}{\partial h} = 0$$

to give the most probable value of h. It is sufficient here,

Fisher had an estimation interest in the density of the sample variance—and derived it.

When he came into contact with Gosset he used the same method to work out the density of *z*.

1915: Fisher on correlation

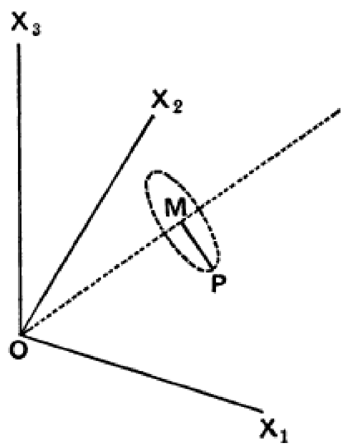
FREQUENCY DISTRIBUTION OF THE VALUES OF THE CORRELATION COEFFICIENT IN SAMPLES FROM AN INDEFINITELY LARGE POPULATION.

By R. A. FISHER.

....

The five quantities above defined have, in fact, an exceedingly beautiful interpretation in generalised space, which we may now examine.

....



- Fisher's derivation of Student's *z* was *not* published
- But Pearson published a later piece on the correlation coefficient
- Pearson thought this *is* important
- Gosset "there still remains the determination of the probability curve giving the probability of the real value"

1920 Fisher tells the astronomers

Monthly Notices of the Royal Astronomical Society

A Mathematical Examination of the Methods of determining the Accuracy of an Observation by the Mean Error, and by the Mean Square Error. By R. A. Fisher, M.A.

- Arthur Eddington claims that the mean error estimate is superior to the mean square estimate.
- Fisher says no, producing results first obtained by Gauss in 1816 and Helmert in 1876

But new idea of sufficiency

The whole of the information respecting σ, *which a sample provides, is summed up in the value of* σ_2.

1922 Regression moves

MISCELLANEA.

THE GOODNESS OF FIT OF REGRESSION FORMULÆ, AND THE DISTRIBUTION OF REGRESSION COEFFICIENTS.

By R. A. FISHER, M.A.

...

we are in a position to give an exact solution of the distribution of the regression coefficients. This problem has been outstanding for many years; but the need for its solution was recently brought home to the writer by correspondence with "Student," whose brilliant researches [7] in 1908 form the basis of the exact solution.

Tables of the Probability Integral of the above Type VII distribution have been prepared by "Student" [8], for values of $n - p$ from 0 to 30. These tables are in a suitable form for testing the significance of an observed regression coefficient. For larger samples the curve will be sufficiently normal for most purposes,

- Regression lived in the multinormal
- Gosset asked Fisher for the distribution of a regression coefficient
- looking for a follow-up to the correlation results of 1915
- He *got* a follow-up to Student on the mean
- As error theory natural –from direct to indirect observations

1922 z turns into t

…

…

The present tables, have, therefore, at Mr Fisher's suggestion been constructed with argument $t = z\sqrt{n}$, where n is now one less than the number in the sample.

Student 1925

- In 1922 Gosset decides to produce new tables of *z*
- Fisher has a new method of computing the values
- And there are two parallel projects
- By November 1922 they are using the *t* form
- New tables appear in 1925/6

Bayesian diversion 1923

Proceedings of the Cambridge Philosophical Society

- Burnside "On Errors of Observation"
- Fisher "Note on Dr. Burnside's Recent Paper on Errors of Observation"

Gosset to Fisher

- It is interesting to see how à priori probability has got him just off the line.
- … it is a good instance of the futility of á priori assumptions

Yet Gosset's position re Bayes remains unclear.

- In retirement the group theorist William Burnside takes up error theory
- Produces a Bayesian analysis of Student's problem
- Fisher writes indicating Student's priority
- Gosset underwhelmed by Burnside

A system of distributions 1924

ON A DISTRIBUTION YIELDING THE ERROR FUNCTIONS OF SEVERAL WELL KNOWN STATISTICS

By Mr. R. A. Fisher,
Statistical Department, Rothamsted Experimental Station, Harpenden, England.

...

The four chief cases of the z distribution have the following applications

I. Normal Curve	II. χ^2	III. Student's	IV. z
Many statistics from large samples	Goodness of fit of frequencies Index of dispersion for Poisson and Binomial samples Variance of normal samples	Mean Regression coefficient Comparison of means and regressions	Intraclass correlations Multiple correlation Comparison of variances Correlation ratio Goodness of fit of regressions

- A new *z*
- Convergence of different streams of research
- Some work coming out of Pearsonian statistics
- Some work coming out of a rejuvenated theory of errors

1925: *t* arrives

R. A. FISHER, M. A.
Rothamsted Experimental Station

Applications of "Student's„ Distribution.

1. - Introductory.

The Theory of Errors may be said to have taken its origin in the fact that the accuracy of the mean of a number of observations

- *Statistical Methods for Research Workers*
- "Applications of `Student's' distribution"

  and two articles about tables
- "Expansion of `Student's' integral in powers of n^{-1}“
- Student's "New tables for testing the significance of observations."

Beginnings and ends

Dedekind and Lüroth

- Contributions look like something Gauss could have written in 1816
- All the ideas were available then
- Dedekind—like Gauss focussed on precision
- Dedekind and Lüroth did not impose their constructions

Student and Fisher

- inherited a very thin version of the theory of errors
- had a productive interaction with biometry
- with Student a lot of confusion
- with Fisher a new subject