LE PREMIER PILLIER DE LA SAGESSE STATISTIQUE

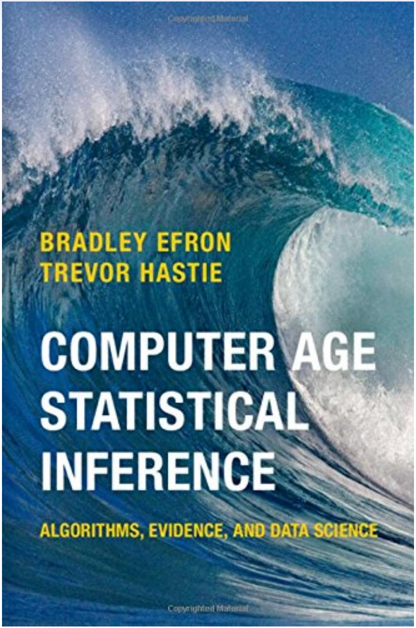
The first pillar of statistical wisdom

James A. Hanley

Department of Epidemiology, Biostatistics & Occupational Health

McGill University

~~2016.11.23~~ ~~2017.01.25~~ 2017.03.22
Département de mathématiques
Université du Québec à Montréal



Copyrighted Material
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE
Copyrighted Material

2016

OUTLINE

- An introduction to Stephen Stigler's book The Seven Pillars of Statistical Wisdom (and his 2 earlier books)

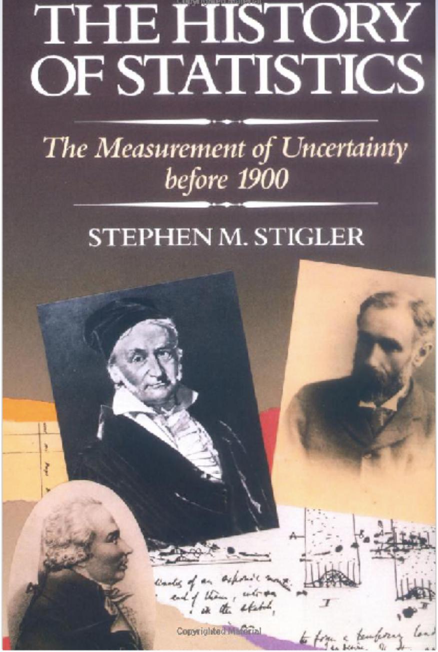
- The first of these pillars: 'Aggregation'

  - early instances of the sample mean in scientific work

  - multi-parameter situations [briefly]

  - some early error distributions

  - how their 'centres' were fitted



"An exceptionally searching, almost loving, study of the relevant inspirations and aberrations of its principal characters James Bernoulli, de Moivre, Bayes,

The book is a pleasure to read: the prose sparkles; the protagonists are vividly drawn; the illustrations are handsome and illuminating; the insights

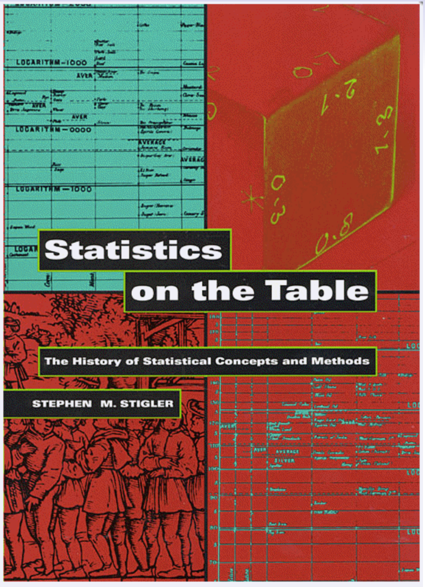
HANLEY.

The History of Statistics

With best wishes

Steve Stigler

6 Dec 2013


Statistics
on the Table
The History of Statistical Concepts and Methods
STEPHEN M. STIGLER

The Seven Pillars of Statistical Wisdom

STEPHEN M. STIGLER

Why do we continue to ask: What is Statistics?

- Not a single subject.
- Has changed dramatically, from a profession that
  - claimed such extreme objectivity that statisticians would only gather data – not analyze them
  - to a profession that seeks partnership with scientists in all stages of investigation, from planning to analysis.
- Different faces to different sciences: in some applications,
  - we accept the scientific model as derived from mathematical theory; in others
  - we construct a model that can then take on a status as firm as any Newtonian construction.
- In some, we are active planners and passive analysts; in others, just the reverse.

A unified discipline, even a science of our own?

- I will not try to tell you what Statistics is or is not.

- I will attempt to formulate seven principles, seven pillars that have supported our field in different ways in the past and promise to do so into the indefinite future.

- I will try to convince you that each of these was revolutionary when introduced, and remains a deep and important conceptual advance.

The 7 Pillars

‘Aggregation’ / ‘Combination of Observations’ / ‘Taking a mean’ (simplest e.g.)

- An old idea, revolutionary in an earlier day – and still so today, whenever it reaches into a new area of application.
- Given a no. of observations, you gain information by throwing information away!
- A simple arithmetic mean discards the individuality of the measures, subsuming them to one summary.
- It may come naturally now in repeated measurements of, say, a star position in astronomy. But in the seventeenth century it might have required ignoring the knowledge that the French observation was made by an observer prone to drink and the Russian observation was made by use of an old instrument, but the English observation was by a good friend who had never let you down.
- Details of individual observations ‘erased’ to reveal a better indication than any single observation could on its own.

Averages are many but they have a short history

- The earliest clearly documented use of an arithmetic mean was in 1635
- Other forms of statistical summary have a much longer history, back to Mesopotamia and nearly to the dawn of writing.
- Recent important instances of this first pillar are more complicated. The method of least squares and its cousins and descendants are all averages.
- 19th century: " combination of observations."

The taking of a mean of any sort is a rather radical step in an analysis

- statistician is discarding information in the data;
- the individuality of each observation is lost:
  - the order in which the measurements were taken
  - the differing circumstances in which they were made,
  - including the identity of the observer.

Examples

- 1860s: Pushback against Jevons' Commodities Index
- 1874: Determining dimensions of solar system using measurements during Transit of Venus
  Are those made with **different** equipment by observers of different skills at slightly different times at different places like enough to be meaningfully averaged?
- Are successive observations of a star position made by a single observer, acutely aware of every tremble and hiccup and distraction, sufficiently alike to be averaged?
- In ancient and even modern times, too much familiarity with the circumstances of each observation could undermine intentions to combine them.
- Strong temptation to select one observation thought to be the best, rather than to corrupt it by averaging with others of suspected lesser value.

‘Funes the Memorious’

(Jorge Luis Borges 1942)

Ireneo Funes found after an accident that he could remember absolutely everything. He could reconstruct every day in the smallest detail, and he could even later reconstruct the reconstruction, but he was incapable of understanding.

“To think is to forget details, generalize, make abstractions.

In the teeming world of Funes there were only details.”

Aggregation can yield great gains above the individual components.

Funes was big data without Statistics.

THE ARITHMETIC MEAN

1. When was it first used to summarize a data set?
2. When was this practice widely adopted?

1 : may be impossible to answer.

2: seems to be sometime in the 17th century, but being more precise about the date also seems intrinsically difficult.

To better understand the measurement and reporting issues involved, let us look at an interesting example, one that includes what may be the earliest published use of the phrase “arithmetical mean” in this context.

WIKIPEDIA
The Free Encyclopedia

Article | Talk | Read | Edit | View history | Search Wikipedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikipedia store

Interaction

- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact page

Tools

- What links here
- Related changes
- Upload file
- Special pages
- Permanent link
- Page information
- Wikidata item
- Cite this page

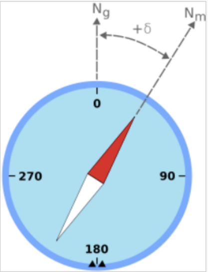
Print/export

Magnetic declination

From Wikipedia, the free encyclopedia

"Magnetic North" redirects here. For other uses, see Magnetic North (disambiguation).

Magnetic declination or **variation** is the angle on the horizontal plane between magnetic north (the direction the north end of a compass needle points, corresponding to the direction of the Earth's magnetic field lines) and true north (the direction along a meridian towards the geographic North Pole). This angle varies depending on position on the Earth's surface, and changes over time.



Example of magnetic declination showing a compass needle with a "positive" (or "easterly") variation from geographic north. N_g is geographic or true north, N_m is magnetic north, and δ is magnetic declination

Somewhat more formally, Bowditch defines variation as "the angle between the magnetic and geographic meridians at any place, expressed in degrees and minutes east or west to indicate the direction of magnetic north from true north. The angle between magnetic and grid meridians is called grid magnetic angle, grid variation, or grivation."[1]

By convention, declination is positive when magnetic north is east of true north, and negative when it is to the west. *Isogonic lines* are lines on the Earth's surface along which the declination has the same constant value, and lines along which the declination is zero are called *agonic lines*. The lowercase Greek letter δ (delta) is frequently used as the symbol for magnetic declination.

The term **magnetic deviation** is sometimes used loosely to mean the same as magnetic declination, but more correctly it refers to the error in a compass reading induced by nearby metallic objects, such as on board a ship or aircraft.

Magnetic declination should not be confused with **magnetic inclination**, also known as

▸ Link to Wikipedia

❡ In Limehouse the sixteenth of October. Anno 1580.

Fornoone.			Afternoone.				
Elevation of the Sunne.	Variation of the shadow from the North of the Needle to the Westwardes.		Elevation of the Sunne.	Variation of the shadow from the North of the Needle to the Eastwards.		Variation of the Needle from the Pole or Axis.	
Deg.	Degr.	Min.	Deg.	D.	M.	D.	M.
17	52	35	17	30	0	11	17 2/3
18	50	8	18	27	45	11	11 1/3
19	47	30	19	24	30	11	30
20	45	0	20	22	15	11	22 1/3
21	42	15	21	19	30	11	22 1/3
22	38	0	22	15	30	11	15
23	34	40	23	12	0	11	20
24	29	35	24	7	0	11	17
25	22	20	25	Fró N. to w. 0,8′		11	14

"I do finde the true variation of the Needle or Cumpas at Lymehouse to be about 11° 15', or 11° 20',

whiche is a poinct of the Cumpas just or a little more."

His 11° 15’ does not correspond to any modern summary measure

- It is smaller than the mean, median, midrange, and mode.
- It agrees with the value for 22° elevation, and could have been so chosen – but then why also give 11° 20’, the figure for 23° elevation?
- Or perhaps he rounded to agreement with “one point of the compass,” that is, the 11° 15’ distance between each of the 32 points of the compass?
- Regardless, it is clear Borough did not feel the necessity for a formal compromise.

Observations made at Diepford An. 1634 Iunij 12 before Noone

Alt: ☉ vera		Azim: Mag		Azim. ☉		variatio	
Gr.	Min.	Gr.	M.	Gr.	M.	Gr.	M.
44,	45.	106,	0	110	6	4.	6
46,	30,	109,	0	113	10	4,	10
48,	31,	113,	0	117	1	4.	1
50,	54,	118	0	122,	3	4.	3
54,	24,	127	0	130	55	3	55

After Noone the same day.

Alt. ☉ vera		Azi. Mag		Azim. ☉		Variation	
Gr.	Min.	Gr:	M.	G,	Mn.	Gr,	Min
44	37	114:	0	109.	53.	4:	7
40	48	108:	0	103,	50	4:	10
38	46	105.	0	100,	48	4.	12
36	43	102,	0	97.	56	4.	4
34	32	99,	0	95,	0	4:	0
32	10	96:	0	91.	55	4'	5

These Concordant Obſervations can not produce a variation greater then 4 gr. 12 min. nor leſſe then 3 gr. 55 min. the Arithmeticall meane limiting it to 4 gr. and about 4 minutes.

His "meane" is not the arithmetic mean of all 11; that would be 4° 5'

- Instead he gives the mean of the largest and smallest: what later statisticians would call a midrange
- As such it is not remarkable. While it is an arithmetic mean of two observations, there is scarcely any other way of effecting a compromise between two values.
- There were in fact several earlier astronomers who had done this or something similar when confronted with two values and in need of a single value - certainly Brahe and Johannes Kepler in the early 1600s, and possibly al-Biruni ca. 1000 CE.
- What was **new** with Gellibrand's work was the terminology – he **gives a name to the method used**. The name had been known to the ancients, but, as far as is now known, none of them had felt it useful or necessary to actually use the name in their written work.

Sign that statistical analysis of observations had entered into a new phase:

short note in the Transactions of the Royal Society in 1668

Observed June 13. 1668.							
Sun's Observ'd Altitude.		Magnetical Azimuth.		Suns true Azimuth.		Variat. Westerly.	
Gr.	M.	Gr.	M.	Gr.	M.	G.	M.
44	20	72	00	70	38	1	22
39	30	80	00	78	24	1	36
31	50	90	00	88	26	1	34
27	42	95	00	93	36	1	24
23	20	103	00	101	23	1	23

"In taking this *Table* [Captain Sturmy] notes the greatest distance or difference to be 14 minutes; and so taking the *mean* for the true Variation, he concludes it *then and there* to be just 1 deg. 27 min. *viz. June 13 1666*."

While the true mean is 1 deg. 27.8' and Captain Sturmy (or mathematician Staynred) rounded down

- It is in any event clear that the arithmetic mean had arrived by the last third of that [17th] century and been officially recognized as a method for combining observations.
- The date of birth may never be known, but the fact of birth seems undeniable.

Example: land surveying in the early 1500s

- The basic unit of land measure in those times was the rod, defined as 16 feet long.
- And in those days a foot meant a real foot, but whose foot?
- Surely not the king's foot, or each change of monarch would require a renegotiation of land contracts.

Simple and elegant solution reported by Köbel

- “Stand at the door of a church on a Sunday and bid 16 men to stop, tall ones and small ones, as they happen to pass out when the service is finished;

- then make them put their left feet one behind the other, and the length thus obtained shall be a right and lawful rood to measure and survey the land with,

- and the 16th part of it shall be the right and lawful foot.”

– It was truly a community rod!
– Functionally, it’s the arithmetic mean of the 16 individual feet,
– but nowhere was the mean mentioned. [cf. ’havaria’ in marine insurance]

COMBINATION OF OBSERVATIONS

(Multiparameter applications)

$\{\alpha,\ \beta,\ \theta\}$ – Mayer, 1750

Table 1.1. Mayer's twenty-seven equations of condition, derived fro
observations of the crater Manilius from 11 April 1748 through 4 M

Eq. no.	Equation
1	$\beta - 13°10' = +0.8836\alpha - 0.4682\alpha \sin\theta$
2	$\beta - 13°8' = +0.9996\alpha - 0.0282\alpha \sin\theta$
3	$\beta - 13°12' = +0.9899\alpha + 0.1421\alpha \sin\theta$
4	$\beta - 14°15' = +0.2221\alpha + 0.9750\alpha \sin\theta$
5	$\beta - 14°42' = +0.0006\alpha + 1.0000\alpha \sin\theta$
6	$\beta - 13°1' = +0.9308\alpha - 0.3654\alpha \sin\theta$
7	$\beta - 14°31' = +0.0602\alpha + 0.9982\alpha \sin\theta$
8	$\beta - 14°57' = -0.1570\alpha + 0.9876\alpha \sin\theta$
9	$\beta - 13°5' = +0.9097\alpha - 0.4152\alpha \sin\theta$
10	$\beta - 13°2' = +1.0000\alpha + 0.0055\alpha \sin\theta$
11	$\beta - 13°12' = +0.9689\alpha + 0.2476\alpha \sin\theta$
12	$\beta - 13°11' = +0.8878\alpha + 0.4602\alpha \sin\theta$
13	$\beta - 13°34' = +0.7549\alpha + 0.6558\alpha \sin\theta$
14	$\beta - 13°53' = +0.5755\alpha + 0.8178\alpha \sin\theta$
15	$\beta - 13°58' = +0.3608\alpha + 0.9326\alpha \sin\theta$
16	$\beta - 14°14' = +0.1302\alpha + 0.9915\alpha \sin\theta$
17	$\beta - 14°56' = -0.1068\alpha + 0.9943\alpha \sin\theta$
18	$\beta - 14°47' = -0.3363\alpha + 0.9418\alpha \sin\theta$
19	$\beta - 15°56' = -0.8560\alpha + 0.5170\alpha \sin\theta$
20	$\beta - 13°29' = +0.8002\alpha + 0.5997\alpha \sin\theta$
21	$\beta - 15°55' = -0.9952\alpha - 0.0982\alpha \sin\theta$
22	$\beta - 15°39' = -0.8409\alpha + 0.5412\alpha \sin\theta$
23	$\beta - 16°9' = -0.9429\alpha + 0.3330\alpha \sin\theta$
24	$\beta - 16°22' = -0.9768\alpha + 0.2141\alpha \sin\theta$
25	$\beta - 15°38' = -0.6262\alpha - 0.7797\alpha \sin\theta$
26	$\beta - 14°54' = -0.4091\alpha - 0.9125\alpha \sin\theta$
27	$\beta - 13°7' = +0.9284\alpha - 0.3716\alpha \sin\theta$

Source: Mayer (1750, p. 153).
Note: One misprinted sign in equation 7 has been corrected.

Table 1.1. Mayer's twenty-seven equations of condition, derived from observations of the crater Manilius from 11 April 1748 through 4 March 1749.

Eq. no.	Equation	Group
1	$\beta - 13°10' = +0.8836\alpha - 0.4682\alpha \sin\theta$	I
2	$\beta - 13°8' = +0.9996\alpha - 0.0282\alpha \sin\theta$	I
3	$\beta - 13°12' = +0.9899\alpha + 0.1421\alpha \sin\theta$	I
4	$\beta - 14°15' = +0.2221\alpha + 0.9750\alpha \sin\theta$	III
5	$\beta - 14°42' = +0.0006\alpha + 1.0000\alpha \sin\theta$	III
6	$\beta - 13°1' = +0.9308\alpha - 0.3654\alpha \sin\theta$	I
7	$\beta - 14°31' = +0.0602\alpha + 0.9982\alpha \sin\theta$	III
8	$\beta - 14°57' = -0.1570\alpha + 0.9876\alpha \sin\theta$	II
9	$\beta - 13°5' = +0.9097\alpha - 0.4152\alpha \sin\theta$	I
10	$\beta - 13°2' = +1.0000\alpha + 0.0055\alpha \sin\theta$	I
11	$\beta - 13°12' = +0.9689\alpha + 0.2476\alpha \sin\theta$	I
12	$\beta - 13°11' = +0.8878\alpha + 0.4602\alpha \sin\theta$	I
13	$\beta - 13°34' = +0.7549\alpha + 0.6558\alpha \sin\theta$	III
14	$\beta - 13°53' = +0.5755\alpha + 0.8178\alpha \sin\theta$	III
15	$\beta - 13°58' = +0.3608\alpha + 0.9326\alpha \sin\theta$	III
16	$\beta - 14°14' = +0.1302\alpha + 0.9915\alpha \sin\theta$	III
17	$\beta - 14°56' = -0.1068\alpha + 0.9943\alpha \sin\theta$	III
18	$\beta - 14°47' = -0.3363\alpha + 0.9418\alpha \sin\theta$	II
19	$\beta - 15°56' = -0.8560\alpha + 0.5170\alpha \sin\theta$	II
20	$\beta - 13°29' = +0.8002\alpha + 0.5997\alpha \sin\theta$	III
21	$\beta - 15°55' = -0.9952\alpha - 0.0982\alpha \sin\theta$	II
22	$\beta - 15°39' = -0.8409\alpha + 0.5412\alpha \sin\theta$	II
23	$\beta - 16°9' = -0.9429\alpha + 0.3330\alpha \sin\theta$	II
24	$\beta - 16°22' = -0.9768\alpha + 0.2141\alpha \sin\theta$	II
25	$\beta - 15°38' = -0.6262\alpha - 0.7797\alpha \sin\theta$	II
26	$\beta - 14°54' = -0.4091\alpha - 0.9125\alpha \sin\theta$	II
27	$\beta - 13°7' = +0.9284\alpha - 0.3716\alpha \sin\theta$	I

Source: Mayer (1750, p. 153).
Note: One misprinted sign in equation 7 has been corrected.

The Earth is not perfectly spherical

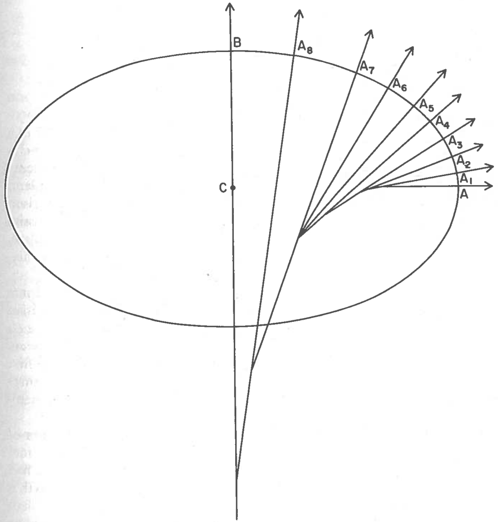


Figure 1.3. A side view of an exaggeratedly oblate earth, illustrating the lengthening of degrees of arc toward the pole. The meridian quadrant AB is broken into nine segments, each of 10° latitude. (Based upon Berry, 1898, p. 277.)

1755

Table 1.4. Boscovich's data on meridian arcs. "Y" "X"

Location	Latitude (θ)	Arc length (toises)	Boscovich's $\sin^2\theta \times 10^4$
(1) Quito	0°0′	56,751	0
(2) Cape of Good Hope	33°18′	57,037	2,987
(3) Rome	42°59′	56,979	4,648
(4) Paris	49°23′	57,074	5,762
(5) Lapland	66°19′	57,422	8,386

Source: Boscovich and Maire (1755, p. 500). Reprinted in Boscovich and Maire (1770, p. 482).

Note: Arc lengths are given as toises per degree measured, where 1 toise ≅ 6.39 feet. The value for $\sin^2\theta \times 10^4$ for the Cape of Good Hope is erroneous and is evidently based on 33°8′. The correct figure would be 3,014.

where he followed in a Newtonian tradition of giving geometric descriptions rather than analytic ones.[8] It will be easier, however, to relate Boscovich's different efforts to later work if we adopt an analytic formulation from the beginning. In analytic terms, Boscovich was faced with the equivalent of five observational equations,

$$a_i = z + y \sin^2 \theta_i ,$$

E[Y I X] = A + B X

where a_i and θ_i are the length of an arc (in toise per degree, 1 toise ≅ 6.39 feet) and the latitude of the midpoint of the arc, both at location i. The unknowns y and z are, respectively, the excess of a 1° arc at the pole over one at the equator and the length of a degree at the equator. **B** **A**

1793: 1 metre = 10,000,000th part of the meridian quadrant

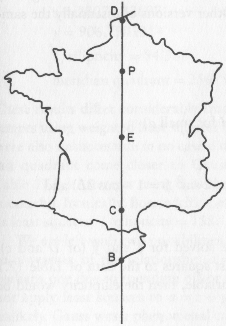


Figure 17.1
The French meridian arc, through Dunkirk (D), the Pantheon (P) in Paris, Evaux (E), Carcassone (C), and Barcelona (B).

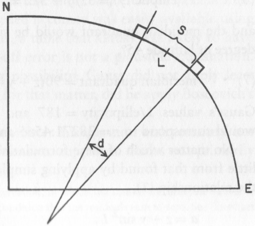


Figure 17.2
A meridian quadrant, from the Equator (E) to the North Pole (N), showing an arc segment of d degrees latitude and length S modules, centered at latitude L.

NOUVELLES MÉTHODES

POUR LA DÉTERMINATION

DES

ORBITES DES COMÈTES;

PAR A. M. LEGENDRE,

Membre de l'Institut et de la Légion d'honneur, de la Société royale de Londres, &c.

A PARIS,

Chez FIRMIN DIDOT, Libraire pour les Mathématiques, la Marine, l'Architecture, et les Éditions stéréotypes, rue de Thionville, n° 116.

AN XIII — 1805.

APPENDICE.

Sur la Méthode des moindres quarrés.

DANS la plupart des questions où il s'agit de tirer des mesures données par l'observation, les résultats les plus exacts qu'elles peuvent offrir, on est presque toujours conduit à un systême d'équations de la forme

$$E = a + bx + cy + fz + \&c.$$

dans lesquelles a, b, c, f, &c. sont des coëfficiens connus, qui varient d'une équation à l'autre, et x, y, z, &c. sont des inconnues qu'il faut déterminer par la condition que la valeur de E se réduise, pour chaque équation, à une quantité ou nulle ou très-petite.

Dans la plupart des questions où il s'agit de tirer des mesures données par l'observation, les résultats les plus exacts qu'elles peuvent offrir, on est presque toujours conduit à un système d'équations de la forme

$$E = a + bx + cy + fz + \&c.$$

dans lesquelles a, b, c, f, &c. sont des coëfficiens connus, qui varient d'une équation à l'autre, et x, y, z, &c. sont des inconnues qu'il faut déterminer par la condition que la valeur de E se réduise, pour chaque équation, à une quantité ou nulle ou très-petite.

Si l'on a autant d'équations que d'inconnues x, y, z, &c., il n'y a aucune difficulté pour la détermination de ces inconnues, et on peut rendre les erreurs E absolument nulles. Mais le plus souvent, le nombre des équations est supérieur à celui des inconnues, et il est impossible d'anéantir toutes les erreurs.

Dans cette circonstance, qui est celle de la plupart des problêmes physiques et astronomiques, où l'on cherche à déterminer quelques élémens importans, il entre nécessairement de l'arbitraire dans la distribution des erreurs, et on ne doit pas s'attendre que toutes les hypothèses conduiront exactement aux mêmes résultats; mais il faut sur-tout faire en sorte que les erreurs extrêmes, sans avoir égard à leurs signes, soient renfermées dans les limites les plus étroites qu'il est possible.

De tous les principes qu'on peut proposer pour cet objet, je pense qu'il n'en est pas de plus général, de plus exact, ni d'une application plus facile que celui dont nous avons fait usage dans les recherches précédentes, et qui consiste à rendre *minimum* la somme des quarrés des erreurs. Par ce moyen, il s'établit entre les erreurs une sorte d'équilibre qui empêchant les extrêmes de prévaloir, est très-propre à faire connoître l'état du système le plus proche de la vérité.

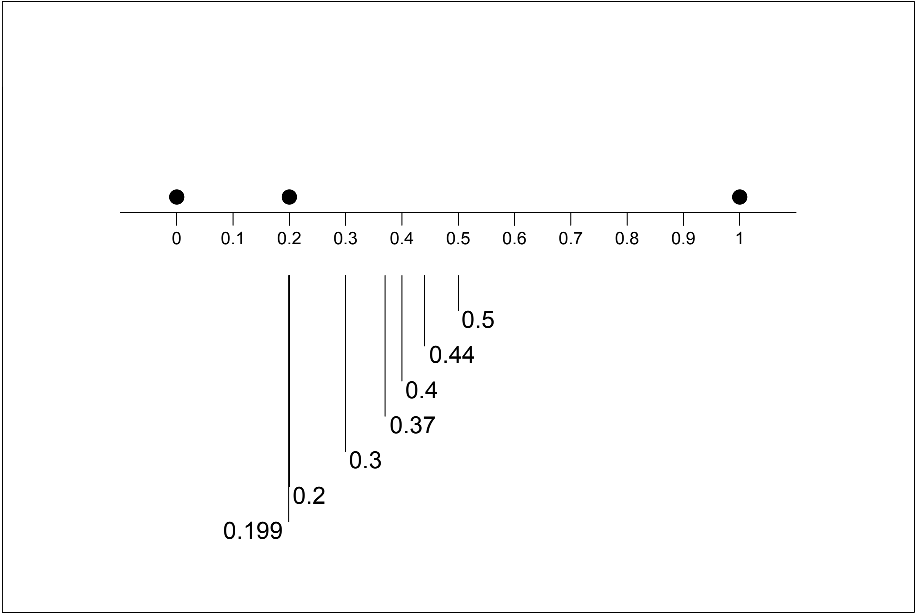
Lieu de l'observation.	*Sa latitude.*	*Arcs compris exprimés en modules.*	L′ — L	L′ + L
Dunkerque	51° 2′ 10″ 50			
		DP 62472.59	2° 11′ 20″ 75	99° 53′ 0″
Panthéon à Paris	48 50 49.75			
		PE 76145.74	2 40 7.25	95 1 32
Evaux............	46 10 42.50			
		EC 84424.55	2 57 48.10	89 23 37
Carcassonne.....	43 12 54.40			
		CM 52749.48	1 51 9.60	84 34 39
Montjouy	41 21 44 80			

EARLY ERROR DISTRIBUTIONS

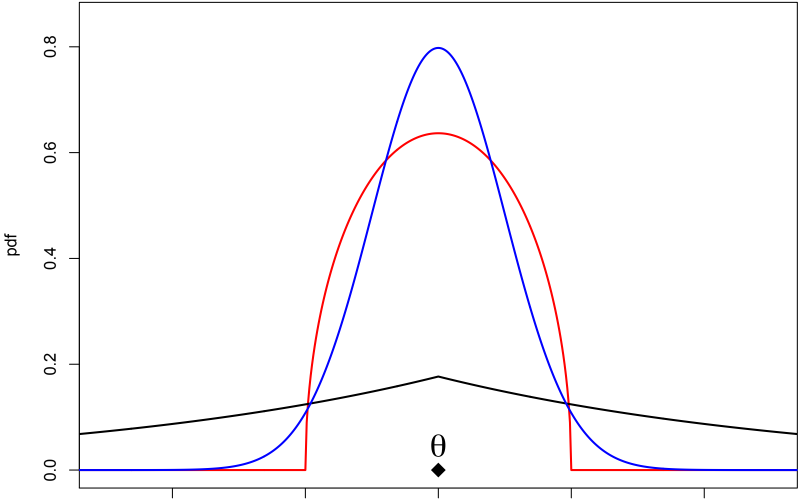
(and how their ‘centres’ were fitted)

3 discrepant observations

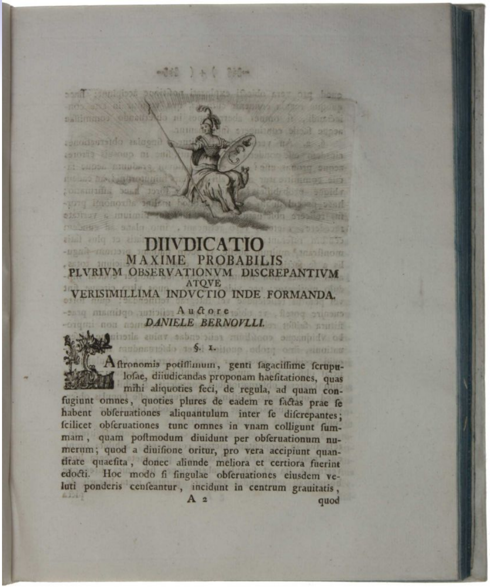
Various estimates of '**Centre**' of 3 discrepant observations



Possible Error Distributions



Daniel Bernoulli, 1778   Laplace, 1774   ???, 17??



DIIVDICATIO
MAXIME PROBABILIS
PLVRIVM OBSERVATIONVM DISCREPANTIVM
ATQVE
VERISIMILLIMA INDVCTIO INDE FORMANDA.

Auctore
DANIELE BERNOVLLI.

§. 1.

Aſtronomis potiſſimum, genti ſagaciſſime ſcrupuloſae, diiudicandas proponam haeſitationes, quas mihi aliquoties feci, de regula, ad quam confugiunt omnes, quoties plures de eadem re factas prae ſe habent obſeruationes aliquantulum inter ſe diſcrepantes; ſcilicet obſeruationes tunc omnes in vnam colligunt ſummam, quam poſtmodum diuidunt per obſeruationum numerum; quod a diuiſione oritur, pro vera accipiunt quantitate quaeſita, donec aliunde meliora et certiora fuerint edocti. Hoc modo ſi ſingulae obſeruationes eiusdem veluti ponderis cenſeantur, incidunt in centrum grauitatis,

A 2 quod

1778

DANIELE BERNOVLLI. “The most probable choice between several discrepant observations and the formation therefrom of the most likely induction”

After preliminaries re. choice of the radius of the controlling circle

ML Criterion

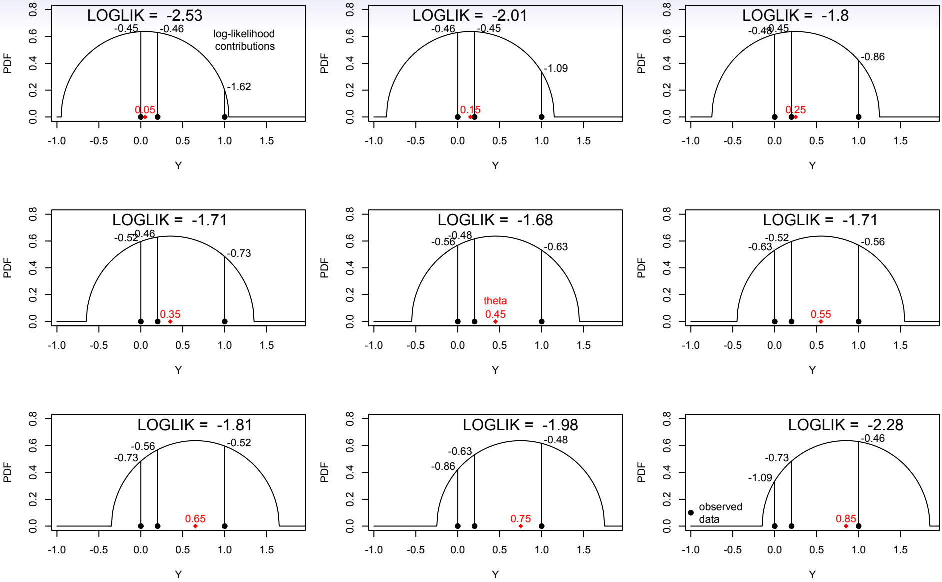
- it remains to determine the position of the controlling circle, since it is at the centre of this circle that the several observations should be deemed to be, as it were, concentrated.
- The aforesaid position is deduced from the fact that the **whole complex of observations would occur more easily, and therefore more probably, for this location than for any other position** of the circle.
- We shall have the true degree of probability for the whole complex of observations if we note the probability corresponding to the several observations that have been carried out and multiply all the probabilities by each other.
- Then the product of the multiplication is to be differentiated and the differential put = 0. In this way we shall obtain an equation whose root will give the distance of the centre from any given point.

The common rule gives $\hat{\theta} = 0.4$. Let us see the new one which to my mind is more probable, and let us put $r = 1$. The following purely numerical equation results

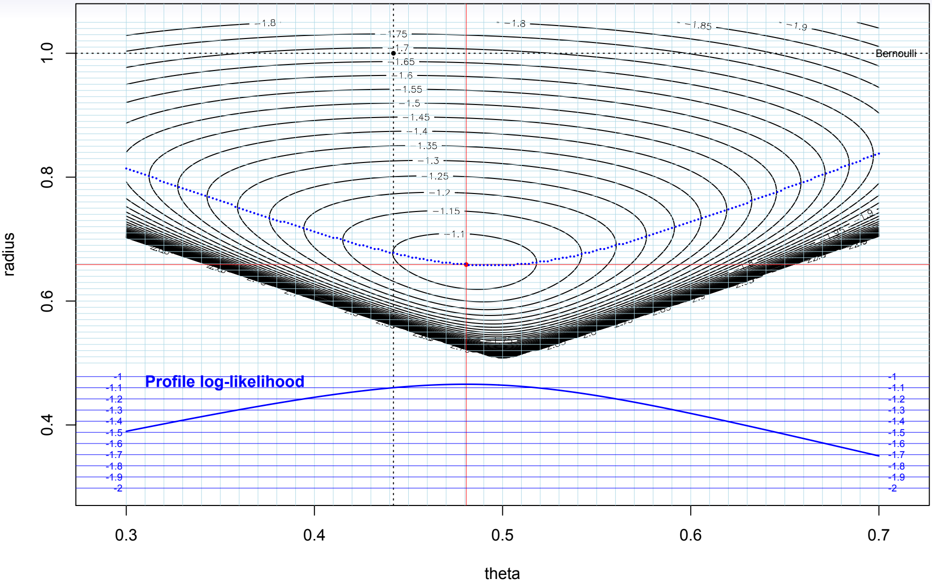
$$1.92 - 0.32\hat{\theta} - 12.96\hat{\theta}^2 + 4.64\hat{\theta}^3 + 12\hat{\theta}^4 - 6x\hat{\theta}^5 = 0,$$

the solution of which is approximately $\hat{\theta} = 0.44$, which exceeds the commonly accepted value by more than a tenth.

Bernoulli (<=1778), R(2016)



MLEs of centre (theta) and 'radius' of Bernoulli error model; data: y = {0, 0.2 and 1.0}



Daniel Bernoulli's 1769 manuscript, studied by Stigler

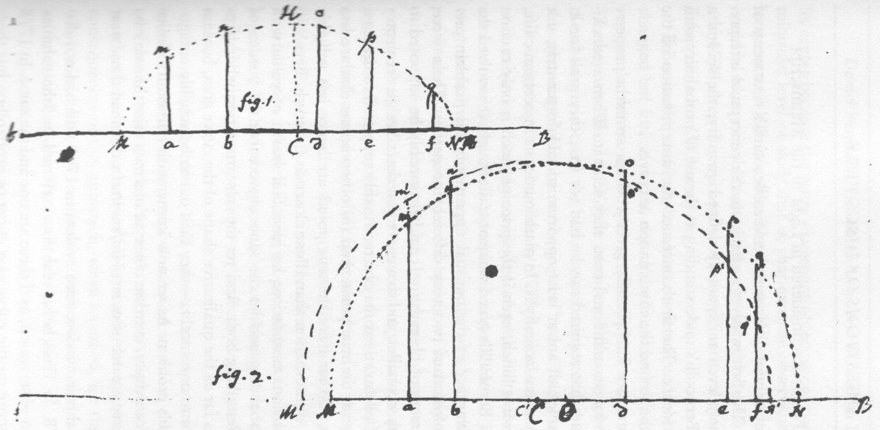
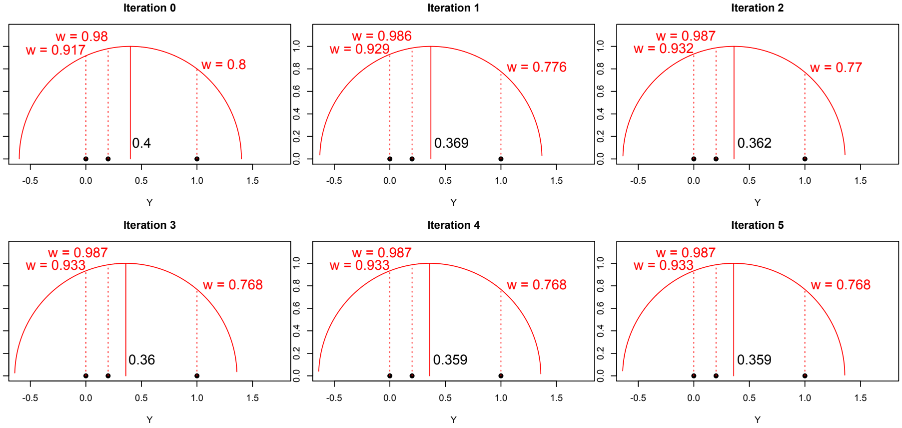
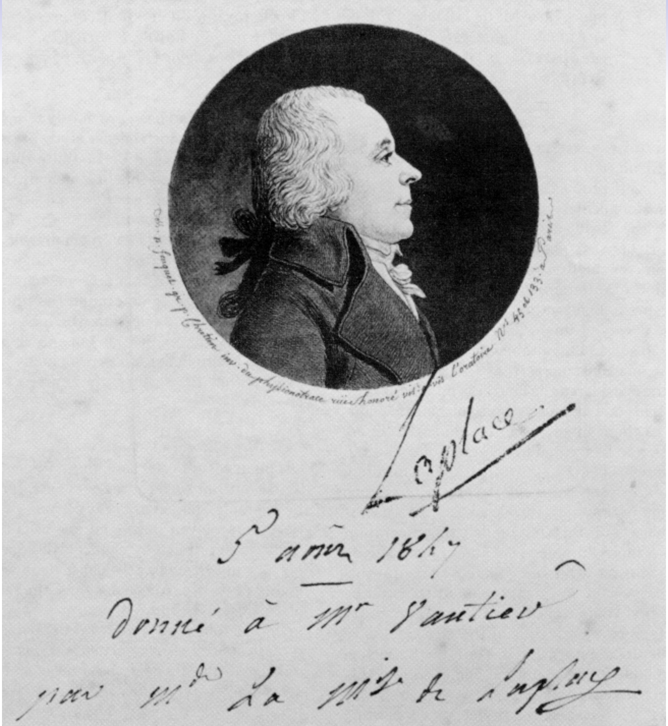


Figure 16.1 Daniel Bernoulli's manuscript drawing of his semicircular curve describing the frequency of errors, and showing the first two iterations of his procedure which used that curve to take a weighted average of the observations depending upon their distance from the previously determined average.

Bernoulli 1769: Robust (M-)Estimation of a Location Parameter, Huber 1964




REPUBLIQUE FRANÇAISE
30F
+9F
POSTES
LAPLACE
1749-1827
LEMAGNY
COTTET


Laplace
donné à Mr

MÉMOIRE

SUR

LA PROBABILITÉ DES CAUSES

PAR LES ÉVÉNEMENTS (¹).

Mémoires de l'Académie royale des Sciences de Paris (*Savants étrangers*), Tome VI, p. 621; 1774.

1774

I.

La théorie des hasards est une des parties les plus curieuses et les plus délicates de l'Analyse, par la finesse des combinaisons qu'elle exige et par la difficulté de les soumettre au calcul; celui qui paraît l'avoir traitée avec le plus de succès est M. Moivre, dans un excellent Ouvrage qui a pour titre : *Theory of chances;* nous devons à cet habile géomètre les premières recherches que l'on ait faites sur l'intégration des équations différentielles aux différences finies; la méthode qu'il a imaginée pour cet objet est fort ingénieuse et il l'a très heureusement appliquée à la solution de plusieurs problèmes sur les Probabilités; on doit convenir cependant que le point de vue sous lequel il a envisagé cette matière est indirect. Les équations aux différences finies sont susceptibles des mêmes considérations que celles aux différences infiniment petites, et doivent être traitées d'une manière analogue; la seule différence qui s'y rencontre est que, dans le cas des différences infiniment petites, on peut négliger certaines quantités qu'il n'est pas

(¹) Par M. de la Place, Professeur à l'École royale militaire.

Memoire on the Probability of the Causes of Events

Statistical Science
2013, Vol. 28, No. 3, 283–288
DOI: 10.1214/13-STS438

The True Title of Bayes's Essay

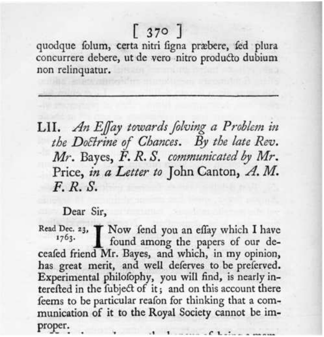
Stephen M. Stigler

Abstract. New evidence is presented that Richard Price gave Thomas Bayes's famous essay a very different title from the commonly reported one. It is argued that this implies Price almost surely and Bayes not improbably embarked upon this work seeking a defensive tool to combat David Hume on an issue in theology.

Key words and phrases: Thomas Bayes, Richard Price, Bayes's theorem, history.

Monday 23 December 2013 is the 250th anniversary of the date Richard Price presented Thomas Bayes's famous paper at a meeting of the Royal Society of London. The paper was published in 1764 as part of the 1763 volume of the *Philosophical Transactions* of the Royal Society, with the block of print shown in Figure 1 at its head. In December 1764 Richard Price read a follow-up paper with himself as author (Figure 2); it was published in 1765 as part of the volume for 1764. All modern readers have taken these article heads as the titles of the papers; the first as "An Essay toward solving a Problem in the Doctrine of Chances;" the second as "A Demonstration of the Second Rule in the Essay toward the Solution of a Problem in the Doctrine of Chances." But Richard Price (and perhaps Bayes as well) had very different titles in mind.

At that time, it was the occasional practice of the Royal Society to supply authors with offprints of published papers, generally before the appearance of the printed volume, based upon the same print block



[370]

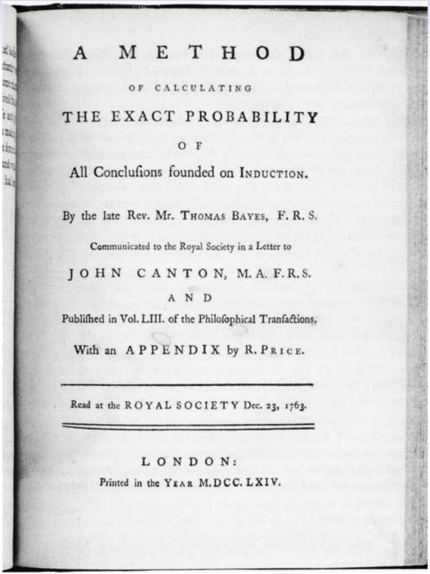
quodque ſolum, certa nitri ſigna præbere, ſed plura concurrere debere, ut de vero nitro producto dubium non relinquatur.

LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr*. Bayes, *F. R. S. communicated by Mr*. Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

FIG. 1. *The heading for Bayes (1764).*



A METHOD

OF CALCULATING

THE EXACT PROBABILITY

OF

All Conclufions founded on INDUCTION.

By the late Rev. Mr. THOMAS BAYES, F. R. S.

Communicated to the Royal Society in a Letter to

JOHN CANTON, M. A. F. R. S.

AND

Publifhed in Vol. LIII. of the Philofophical Tranfactions.

With an APPENDIX by R. PRICE.

Read at the ROYAL SOCIETY Dec. 23, 1763.

LONDON:

Printed in the YEAR M.DCC.LXIV.

FIG. 3. *The title page from the offprint of Bayes (1764). Source: Watson (2013).*

Memoire on the Probability of the Causes of Events

PRINCIPLE: If an event can be produced by a number n of different causes, the probabilities of these causes given the event are to each other as the probabilities of the event given the causes, and the probability of the existence of each of these is equal to the probability of the event given that cause, divided by the sum of all the probabilities of the event given each of these causes.

- Problem I: If an urn contains an infinity of black and white tickets in an unknown ratio, and we draw $p + q$ tickets from it, of which p are white and q are black, then we require the probability that when we draw a new ticket from the urn, it will be white.
- Problem II: Two players A and B, whose respective skills are unknown, play some game, for example piquet, where the first player to win a number n points receives a sum a deposited at the beginning of play. I suppose that the two players are forced to abandon play with player A lacking f points and player B lacking g points. In this situation, we ask how we should divide the sum a between the two players.
- Problem III: Determine the mean that one should take among 3 given observations of the same phenomenon.

Mémoire sur la probabilité des causes par des événements

PRINCIPE. — *Si un événement peut être produit par un nombre n de causes différentes, les probabilités de l'existence de ces causes prises de l'événement sont entre elles comme les probabilités de l'événement prises de ces causes, et la probabilité de l'existence de chacune d'elles est égale à la probabilité de l'événement prise de cette cause, divisée par la somme de toutes les probabilités de l'événement prises de chacune de ces causes.*

Problème I. — *Si une urne renferme une infinité de billets blancs et noirs dans un rapport inconnu, et que l'on en tire $p+q$ billets dont p soient blancs et q soient noirs; on demande la probabilité qu'en tirant un nouveau billet de cette urne il sera blanc.*

Problème II. — *Deux joueurs* A *et* B, *dont les adresses respectives sont inconnues, jouent à un jeu quelconque, par exemple au piquet, à cette condition que celui qui, le premier, aura gagné le nombre n de parties, obtiendra une somme a déposée au commencement du jeu; je suppose que les deux joueurs soient forcés d'abandonner le jeu, lorsqu'il manque f parties au joueur* A, *et h parties au joueur* B; *cela posé, on demande comment on doit partager la somme a entre les deux joueurs.*

PROBLÈME III. — *Déterminer le milieu que l'on doit prendre entre trois observations données d'un même phénomène.*

Fig. 1.

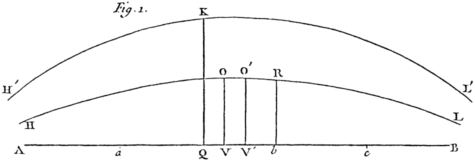


Fig. 2.

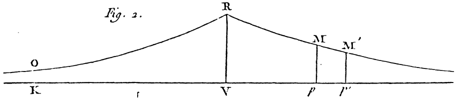
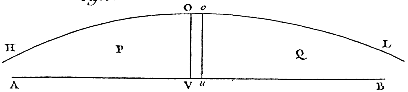


Fig. 3.



Ell. Haussard Sculp.

Suppose now (Figure 1) that the true instant of the phenomenon is at the point V, at the distance x from the point a. The probability that the three observations a, b, and c deviate by the distances Va, Vb, and Vc will be $\phi(x) \cdot \phi(p - x) \cdot \phi(p + q - x)$. If we suppose the true instant were at the point V' and that $aV' = x'$, then this probability would be $= \phi(x') \cdot \phi(p - x') \cdot \phi(p + q - x')$. It follows then from our fundamental principle of section II that the probabilities that the true instant of the phenomenon is at the points V or V', are to each other as $\phi(x) \cdot \phi(p - x) \cdot \phi(p + q - x) : \phi(x') \cdot \phi(p - x') \cdot \phi(p + q - x')$. Thus if we construct a curve HOL with the equation $y = \phi(x) \cdot \phi(p - x) \cdot \phi(p + q - x)$, the ordinates of this curve would represent the probabilities of the corresponding points on the abscissa.

Supposons maintenant (*fig.* 1) que le véritable instant du phénomène soit au point V, à la distance x du point a; la probabilité que les trois observations a, b et c s'écarteront aux distances Va, Vb et Vc sera

$$\varphi(x)\varphi(p-x)\varphi(p+q-x);$$

et, si nous supposons le véritable instant au point V', en sorte que $a\mathrm{V}'=x'$, cette probabilité sera

$$\varphi(x')\varphi(p-x')\varphi(p+q-x');$$

d'où il résulte, par notre principe fondamental de l'Article II, que les probabilités que le véritable instant du phénomène est aux points V ou V' sont entre elles comme

$$\varphi(x)\varphi(p-x)\varphi(p+q-x):\varphi(x')\varphi(p-x')\varphi(p+q-x').$$

Si donc on construit une courbe HOL, dont l'équation soit

$$y=\varphi(x)\varphi(p-x)\varphi(p+q-x),$$

les ordonnées de cette courbe pourront représenter les probabilités des points correspondants de l'abscisse. Cela posé :

In seeking the mean that we should choose among many observations, there are two objects we may have in mind.

The first is the instant such that it is equally probable that the true instant of the phenomenon falls before it or after it. We can call this instant the ~~*mean*~~ **Median** *of probability*.

The second is the instant that *minimizes* the sum of the errors to be feared multiplied by their probabilities. We can call this the *mean of error* or *astronomical mean*, since it is that which astronomers should give preference to.

To find the first mean, it is necessary to determine the ordinate *OV* which divides the area of the curve *HOL* in two equal parts, since then it is clearly as probable that the true instant of the phenomenon falls to the right as to the left of the point *V*.

To find the second mean, it is necessary to choose (Figure 3) a point *V* on the abscissa such that the sum of the ordinates of the curve *HOL*, multiplied by their distance from the point *V*, is a *minimum*. Now I claim that the second mean differs not at all from the first.

Par le milieu que l'on doit choisir entre plusieurs observations, on peut entendre deux choses qu'il importe également de considérer.

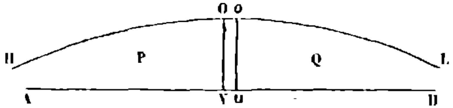
La première est l'instant tel qu'il soit également probable que le véritable instant du phénomène tombe avant ou après ; on pourrait appeler cet instant *milieu de probabilité*.

La seconde est l'instant tel qu'en le prenant pour milieu, la somme des erreurs à craindre, multipliées par leur probabilité, soit un *minimum* ; on pourrait l'appeler *milieu d'erreur* ou *milieu astronomique*, comme étant celui auquel les astronomes doivent s'arrêter de préférence.

Pour avoir le premier milieu, il faut déterminer l'ordonnée OV, qui divise l'aire de la courbe HOL en deux parties égales; car il y a visiblement alors autant de probabilité que le véritable instant du phénomène tombe à droite comme à gauche du point V.

Pour avoir le second milieu, il faut choisir (*fig.* 3) un point V sur

Fig. 3.



l'abscisse, tel que la somme des ordonnées de la courbe HOL, multipliées par leurs distances à ce point V, soit un minimum. Or je dis que ce second milieu ne diffère point du premier. Pour le faire voir, menons l'ordonnée *ou*, infiniment proche de OV. Soient

Laplace's (First) Error Distribution

differences, it follows that we must, subject to the rules of probabilities, suppose the ratio of two infinitely small consecutive differences to be equal to that of the corresponding ordinates. We thus will have

$$\frac{d\phi(x + dx)}{d\phi(x)} = \frac{\phi(x + dx)}{\phi(x)}.$$

Therefore

$$\frac{d\phi(x)}{dx} = -m\phi(x),$$

which gives $\phi(x) = Ce^{-mx}$. Thus, this is the value that we should choose for $\phi(x)$. The constant C should be determined from the supposition that the area of the curve ORM equals unity, which represents certainty, which gives $C = ½m$. Therefore $\phi(x) = (m/2)e^{-mx}$, e being the number whose hyperbolic logarithm is unity.

With *m* fixed, 'MEDIAN OF POSTERIOR' Estimator:

less than q. We suppose that p is greater than q in the following calculations; then to determine the distance x of the point a from the point V where we should fix the true instant of the phenomenon, we will have the following equation.

$$m^2 e^{-m(2p+q-x)} = m^2 e^{-m(p+q)}(1 + \tfrac{1}{3}e^{-mp} - \tfrac{1}{3}e^{-mq}),$$

from which we find

$$x = p + (1/m)\ln(1 + \tfrac{1}{3}e^{-mp} - \tfrac{1}{3}e^{-mq}).$$

$\hat{\theta} = 0.37$ if we fix $m = 1/\sqrt{8}$,

(so his error distribution has same variance as D.Bernoulli($r = 1$).

My textbook in 1966 – Cramér 1946, 10th printing

MATHEMATICAL METHODS OF STATISTICS

By

HARALD CRAMÉR

PROFESSOR IN THE UNIVERSITY OF STOCKHOLM

As shown in the preceding paragraph, the mean is characterized by a certain minimum property: the second moment becomes a minimum when taken about the mean. There is an analogous property of the median: *the first absolute moment* $E(|\xi - c|)$ *becomes a minimum when c is equal to the median.* This property holds even in the indeterminate case, and the moment has then the same value for c equal to any of the possible median values. Denoting the median (or, in the indeterminate case, any median value) by μ, we have in fact the relations

$$E(|\xi - c|) = \begin{cases} E(|\xi - \mu|) + 2\int_{\mu}^{c} (c - x)\, dF(x) & \text{for } c > \mu, \\ E(|\xi - \mu|) + 2\int_{c}^{\mu} (x - c)\, dF(x) & \text{»} \quad c < \mu. \end{cases}$$

The second terms on the right hand sides are evidently positive, except in the case when c is another median value (indeterminate case), when the corresponding term is zero.[1]) The proof of these relations will be left as an exercise for the reader.

Where to stand: 3 unequally spaced elevators

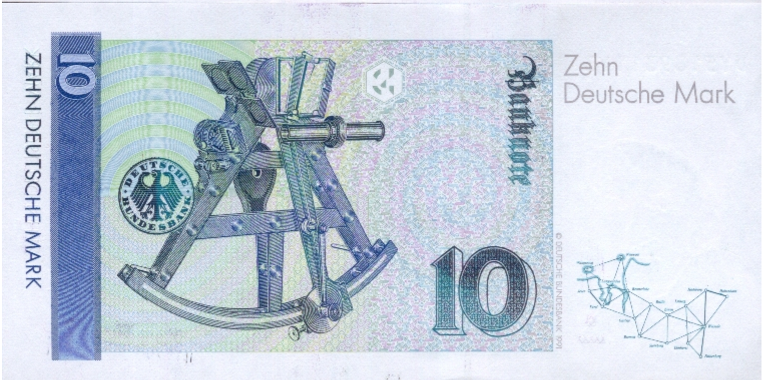
▸ http://www.medicine.mcgill.ca/epidemiology/hanley/elevator.html

Visualizing the median as the minimum deviation location.

Hanley JA, Joseph, L, Platt RW, Chung MK, Bélisle P

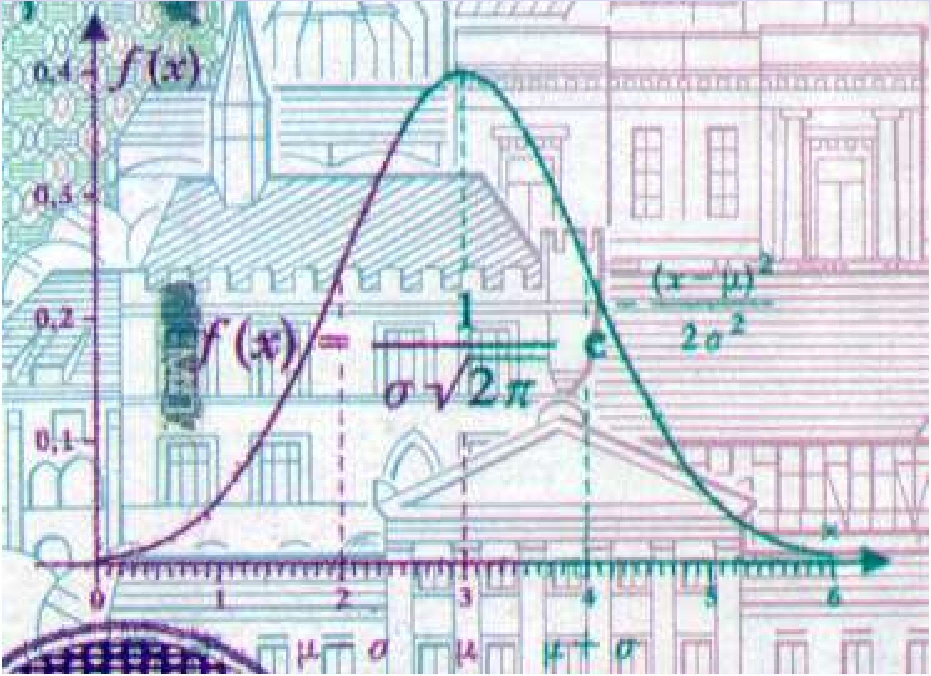
The American Statistician 55(2): 150-152, May 2001.


DN6833178D9
DEUTSCHE BUNDESBANK
Banknote
10
Deutsche Bundesbank
Frankfurt am Main
1.Oktober 1993
1777–1855 Carl Friedr. Gauß
ZEHN DEUTSCHE MARK
DN6833178D9


ZEHN DEUTSCHE MARK
10
DEUTSCHE BUNDESBANK
Banknote
Zehn
Deutsche Mark
10
© DEUTSCHE BUNDESBANK 1991


GL0011661A1
DEUTSCHE BUNDESBANK
Banknote
10
ZEHN DEUTSCHE MARK
1777–1855 Carl Friedr. Gauß
Deutsche Bundesbank
Frankfurt am Main
1.Oktober 1993
GL0011661A1

1777-1855



f(x)
0,4
0,3
0,2
0,1
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
0
1
2
3
4
5
6
μ − σ
μ
μ + σ

1809: "Theory of the Motion of Heavenly Bodies Moving about the Sun in Conic Sections"

The most probable value of a single unknown observed with equal care several times under the same circumstances is the arithmetic mean of the observations $y_1, y_2, \ldots$

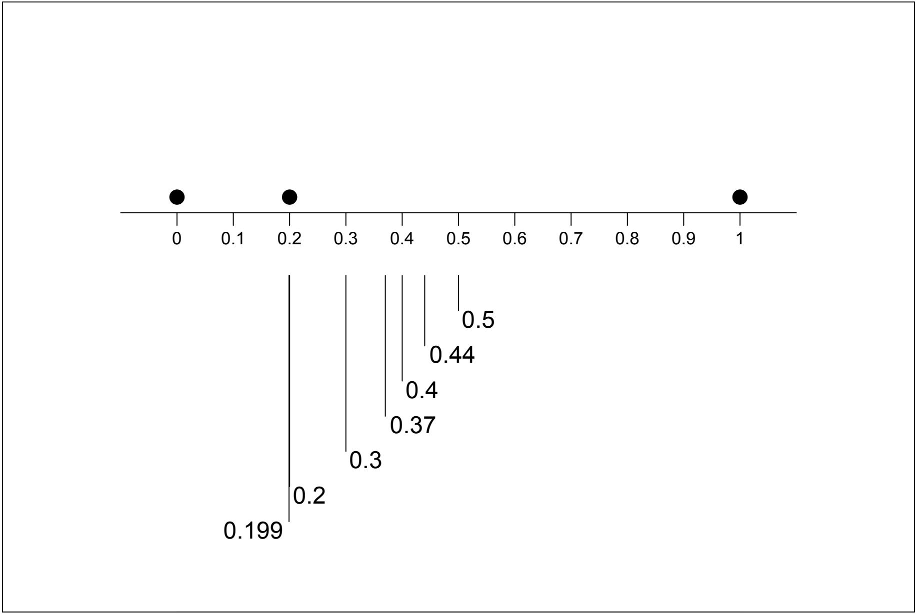
In this case $\bar{y}$ maximizes L only when

$$\phi(\epsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2 \epsilon^2}.$$

In the more general situation, this error distribution leads to the method of least squares as providing values that maximize L.

Stigler: The History of Statistics (1986)

Various estimates of '**Centre**' of 3 discrepant observations



Stigler didn't say why it took so long, ...

- To move up from probability theory and gambling to mathematical statistics, we had to wait for the infinitesimal calculus (Newton, Leibnitz, 2nd half of 1600s).
- The Enlightenment helped: "*Nullius in verba*", Latin for "on the word of no one" or "take nobody's word for it"; *sapere aude* Latin for "Dare to know".
- Laws derived from principles: e.g., ways to come up with error distributions.
- Surveying, astronomy, navigation, ...
- Estimands (parameters) before coming up with estimates
- It takes time to join dots. DeMoivre – Laplace – Gauss; Legendre – Galton.

TODAY, STATISTICAL HISTORY IS ONLY A CLICK AWAY

NOT TOO YOUNG/OLD TO
START/CONTINUE TO CONNECT THE DOTS

The 7 pillars rephrased: the usefulness of 7 basic statistical ideas

1. The value of data targeted reduction or compression of data
2. The diminishing value of an increased amount of data
3. How to put a probability measuring stick to what we do
4. How to use internal variation in the data to help in that
5. How asking questions from different perspectives can lead to revealingly different answers
6. The essential role of the planning of observations
7. How all these ideas can be used in exploring and comparing competing explanations in science

The revolutionary ideas pushed aside or overturned firmly held mathematical or scientific beliefs

- Discarding the individuality of data values
- Downweighting new and equally valuable data
- Overcoming objections to any use of probability to measure uncertainty outside of games of chance.
- How can the variability interior to our data measure the uncertainty about the world that produced it?
- Galton's multivariate analysis revealed to scientists that their reliance upon rules of proportionality dating from Euclid did not apply to a scientific world in which there was variation in the data – overthrowing 3000 years of mathematical tradition.
- Fisher's designs were in direct contradiction to what experimental scientists and logicians had believed for centuries; his methods for comparing models were absolutely new to experimental and required a change of generations for their acceptance.

- These ideas are not part of Mathematics, nor are they part of Computer Science.

- They are centrally of Statistics,

  `and I must now confess that while I began by explicitly denying that my goal was to explain what Statistics is, I may by the end of the book have accomplished that goal nonetheless.`

- All seven have ancient origins, and the modern discipline has constructed its many-faceted structure with great ingenuity and with a constant supply of exciting new ideas of splendid promise.

- But without taking away from that modern work, I hope to articulate a unity at the core of Statistics both across time and between areas of application.

1860s: Jevons versus critics of a Commodities Index that discarded information to increase information

absurd to average data on pig iron and pepper.

Individual commodities: investigators with detailed with historical knowledge were tempted to think they could “explain” every movement, every fluctuation, with some story of why that particular event had gone the way it did.

“Were a complete explanation of each fluctuation thus necessary, not only would all inquiry into this subject be hopeless, but the whole of the statistical and social sciences, so far as they depend upon numerical facts, would have to be abandoned.”

It was not that the stories told about the data were false; it was that they (and the individual peculiarities in the separate observations) had to be pushed into the background. If general tendencies were to be revealed, the observations must be taken as a set; they must be combined.

Combination of Observations - Multiparameter applications

- 1750 Mayer
- Boscovich 1755 (10 pairs of 2), 1757, 1760, 1770 (Least sum of absolute errors)
- Laplace 1783 ((Least maximum error – très pénible))
  1788 LaplaceSaturnData.pdf
  1789 (formalize Boscovich) 1799 ((Least sum of weighted absolute errors))
- ???? Legendre
- Gauss

1788 - Saturn Data - Laplace

Table 1.3. Laplace's Saturn data.

Eq. no.	Year (i)	$-a_i$	b_i	c_i	d_i	Laplace residual	Halley residual	L.S. residual
1	1591	1′11.9″	−158.0	0.22041	−0.97541	+1′33″	−0′54″	+1′36″
2	1598	3′32.7″	−151.78	0.99974	−0.02278	−0.07	+0.37	+0.05
3	1660	5′12.0″	−89.67	0.79735	0.60352	−1.36	+2.58	−1.21
4	1664	3′56.7″	−85.54	0.04241	0.99910	−0.35	+3.20	−0.29
5	1667	3′31.7″	−82.45	−0.57924	0.81516	−0.21	+3.50	−0.33
6	1672	3′32.8″	−77.28	−0.98890	−0.14858	−0.58	+3.25	−1.06
7	1679	3′9.9″	−70.01	0.12591	−0.99204	−0.14	−1.57	−0.08
8	1687	4′49.2″	−62.79	0.99476	0.10222	−1.09	−4.54	−0.52
9	1690	3′26.8″	−59.66	0.72246	0.69141	+0.25	−7.59	+0.29
10	1694	2′4.9″	−55.52	−0.07303	0.99733	+1.29	−9.00	+1.23
11	1697	2′37.4″	−52.43	−0.66945	0.74285	+0.25	−9.35	+0.22
12	1701	2′41.2″	−48.29	−0.99902	−0.04435	+0.01	−8.00	−0.07
13	1731	3′31.4″	−18.27	−0.98712	−0.15998	−0.47	−4.50	−0.53
14	1738	4′9.5″	−11.01	0.13759	−0.99049	−1.02	−7.49	−0.56
15	1746	4′58.3″	−3.75	0.99348	0.11401	−1.07	−4.21	−0.50
16	1749	4′3.8″	−0.65	0.71410	0.70004	−0.12	−8.38	+0.03
17	1753	1′58.2″	3.48	−0.08518	0.99637	+1.54	−13.39	+1.41
18	1756	1′35.2″	6.58	−0.67859	0.73452	+1.37	−17.27	+1.35
19	1760	3′14.0″	10.72	−0.99838	−0.05691	−0.23	−22.17	−0.29
20	1767	1′40.2″	17.98	0.03403	−0.99942	+1.29	−13.12	+1.34
21	1775	3′46.0″	25.23	0.99994	0.01065	+0.19	+2.12	+0.26
22	1778	4′32.9″	28.33	0.78255	0.62559	−0.34	+1.21	−0.19
23	1782	4′4.4″	32.46	0.01794	0.99984	−0.23	−5.18	−0.15
24	1785	4′17.6″	35.56	−0.59930	0.80053	−0.56	−12.07	−0.57

Source: Laplace (1788).
Note: Residuals are fitted values minus observed values.

Ceres to MH370

MH370

In 428 BCE, how to settle on a single figure?

Thucydides:

- Height of the enemy’s wall (in no. of bricks) were counted by many persons at once; and though some might miss the right calculation, most would hit upon it, particularly as they counted over and over again.
- The length required for the ladders was thus obtained.

The mode – the most frequently reported value.