

# A History of Mathematical Statistics from 1750 to 1930

ANDERS HALD

Formerly Professor of Statistics  
University of Copenhagen  
Copenhagen, Denmark



A Wiley-Interscience Publication

JOHN WILEY & SONS, INC.

New York · Chichester · Weinheim · Brisbane · Singapore · Toronto

# Contents

<b>Preface</b>	<b>xv</b>
<b>1. Plan of the Book</b>	<b>1</b>
1.1. Outline of the Contents	1
1.2. Terminology and Notation	7
1.3. Biographies	8
<b>PART I DIRECT PROBABILITY, 1750–1805</b>	
<b>2. Some Results and Tools in Probability Theory by Bernoulli, de Moivre, and Laplace</b>	<b>11</b>
2.1. The Discrete Equiprobability Model	11
2.2. The Theorems of James and Nicholas Bernoulli, 1713	13
2.3. The Normal Distribution as Approximation to the Binomial. De Moivre's Theorem, 1733, and Its Modifications by Lagrange, 1776, and Laplace, 1812	17
2.4. Laplace's Analytical Probability Theory	25
<b>3. The Distribution of the Arithmetic Mean, 1756–1781</b>	<b>33</b>
3.1. The Measurement Error Model	33
3.2. The Distribution of the Sum of the Number of Points by $n$ Throws of a Die by Montmort and de Moivre	34
3.3. The Mean of Triangularly Distributed Errors. Simpson, 1756–1757	35
3.4. The Mean of Multinomially and Continuously Distributed Errors, and the Asymptotic Normality of the Multinomial. Lagrange, 1776	40
3.5. The Mean of Continuous Rectangularly Distributed Observations. Laplace, 1776	50
3.6. Laplace's Convolution Formula for the Distribution of a Sum, 1781	55

<b>4. Chance or Design. Tests of Significance</b>	<b>65</b>
4.1. Moral Impossibility and Statistical Significance	65
4.2. Daniel Bernoulli's Test for the Random Distribution of the Inclinations of the Planetary Orbits, 1735	68
4.3. John Michell's Test for the Random Distribution of the Positions of the Fixed Stars, 1767	70
4.4. Laplace's Test of Significance for the Mean Inclination, 1776 and 1812	74
<b>5. Theory of Errors and Methods of Estimation</b>	<b>79</b>
5.1. Theory of Errors and the Method of Maximum Likelihood by Lambert, 1760 and 1765	79
5.2. Theory of Errors and the Method of Maximum Likelihood by Daniel Bernoulli, 1778	83
5.3. Methods of Estimation by Laplace before 1805	87
<b>6. Fitting of Equations to Data, 1750–1805</b>	<b>91</b>
6.1. The Multiparameter Measurement Error Model	91
6.2. The Method of Averages by Tobias Mayer, 1750	94
6.3. The Method of Least Absolute Deviations by Boscovich, 1757 and 1760	97
6.4. Numerical and Graphical Curve Fitting by Lambert, 1765 and 1772	103
6.5. Laplace's Generalization of Mayer's Method, 1787	107
6.6. Minimizing the Largest Absolute Residual. Laplace, 1786, 1793, and 1799	108
6.7. Laplace's Modification of Boscovich's Method, 1799	112
6.8. Laplace's Determination of the Standard Meter, 1799	116
6.9. Legendre's Method of Least Squares, 1805	118
 <b>PART II INVERSE PROBABILITY BY BAYES AND LAPLACE, WITH COMMENTS ON LATER DEVELOPMENTS</b>	
<b>7. Induction and Probability: The Philosophical Background</b>	<b>125</b>
7.1. Newton's Inductive-Deductive Method	125
7.2. Hume's Ideas on Induction and Probability, 1739	126
7.3. Hartley on Direct and Inverse Probability, 1749	129
<b>8. Bayes, Price, and the Essay, 1764–1765</b>	<b>133</b>
8.1. Lives of Bayes and Price	133

8.2. Bayes's Probability Theory	136
8.3. The Posterior Distribution of the Probability of Success	138
8.4. Bayes's Scholium and His Conclusion	142
8.5. Price's Commentary	145
8.6. Evaluations of the Beta Probability Integral by Bayes and Price	147
<b>9. Equiprobability, Equipossibility, and Inverse Probability</b>	<b>155</b>
9.1. Bernoulli's Concepts of Probability, 1713	155
9.2. Laplace's Definitions of Equiprobability and Equipossibility, 1774 and 1776	157
9.3. Laplace's Principle of Inverse Probability, 1774	159
9.4. Laplace's Proofs of Bayes's Theorem, 1781 and 1786	164
<b>10. Laplace's Applications of the Principle of Inverse Probability in 1774</b>	<b>167</b>
10.1. Introduction	167
10.2. Testing a Simple Hypothesis against a Simple Alternative	167
10.3. Estimation and Prediction from a Binomial Sample	169
10.4. A Principle of Estimation and Its Application to Estimate the Location Parameter in the Measurement Error Model	171
10.5. Laplace's Two Error Distributions	176
10.6. The Posterior Median Equals the Arithmetic Mean for a Uniform Error Distribution, 1781	180
10.7. The Posterior Median for Multinomially Distributed Errors and the Rule of Succession, 1781	181
<b>11. Laplace's General Theory of Inverse Probability</b>	<b>185</b>
11.1. The Memoirs from 1781 and 1786	185
11.2. The Discrete Version of Laplace's Theory	185
11.3. The Continuous Version of Laplace's Theory	188
<b>12. The Equiprobability Model and the Inverse Probability Model for Games of Chance</b>	<b>191</b>
12.1. Theoretical and Empirical Analyses of Games of Chance	191
12.2. The Binomial Case Illustrated by Coin Tossings	192
12.3. A Solution of the Problem of Points for Unknown Probability of Success	196
12.4. The Multinomial Case Illustrated by Dice Throwing	197

12.5. Poisson's Analysis of Buffon's Coin-Tossing Data	198
12.6. Pearson and Fisher's Analyses of Weldon's Dice-Throwing Data	200
12.7. Some Modern Uses of the Equiprobability Model	201
<b>13. Laplace's Methods of Asymptotic Expansion, 1781 and 1785</b>	<b>203</b>
13.1. Motivation and Some General Remarks	203
13.2. Laplace's Expansions of the Normal Probability Integral	206
13.3. The Tail Probability Expansion	210
13.4. The Expansion about the Mode	212
13.5. Two Related Expansions from the 1960s	216
13.6. Expansions of Multiple Integrals	218
13.7. Asymptotic Expansion of the Tail Probability of a Discrete Distribution	220
13.8. Laplace Transforms	222
<b>14. Laplace's Analysis of Binomially Distributed Observations</b>	<b>229</b>
14.1. Notation	229
14.2. Background for the Problem and the Data	230
14.3. A Test for the Hypothesis $\theta \leq r$ Against $\theta > r$ Based on the Tail Probability Expansion, 1781	232
14.4. A Test for the Hypothesis $\theta \leq r$ Against $\theta > r$ Based on the Normal Probability Expansion, 1786	234
14.5. Tests for the Hypothesis $\theta_2 \leq \theta_1$ Against $\theta_2 > \theta_1$ , 1781, 1786, and 1812	235
14.6. Looking for Assignable Causes	240
14.7. The Posterior Distribution of $\theta$ Based on Compound Events, 1812	242
14.8. Commentaries	245
<b>15. Laplace's Theory of Statistical Prediction</b>	<b>249</b>
15.1. The Prediction Formula	249
15.2. Predicting the Outcome of a Second Binomial Sample from the Outcome of the First	249
15.3. Laplace's Rule of Succession	256
15.4. Theory of Prediction for a Finite Population. Prevost and Lhuillier, 1799	262
15.5. Laplace's Asymptotic Theory of Statistical Prediction, 1786	264
15.6. Notes on the History of the Indifference Principle and the Rule of Succession from Laplace to Jeffreys (1948)	268

<b>16. Laplace's Sample Survey of the Population of France and the Distribution of the Ratio Estimator</b>	<b>283</b>
16.1. The Ratio Estimator	283
16.2. Distribution of the Ratio Estimator, 1786	284
16.3. Sample Survey of the French Population in 1802	286
16.4. From Laplace to Bowley (1926), Pearson (1928), and Neyman (1934)	289
<b>PART III THE NORMAL DISTRIBUTION, THE METHOD OF LEAST SQUARES, AND THE CENTRAL LIMIT THEOREM. GAUSS AND LAPLACE, 1809–1828</b>	
<b>17. Early History of the Central Limit Theorem, 1810–1853</b>	<b>303</b>
17.1. The Characteristic Function and the Inversion Formula for a Discrete Distribution by Laplace, 1785	303
17.2. Laplace's Central Limit Theorem, 1810 and 1812	307
17.3. Poisson's Proofs, 1824, 1829, and 1837	317
17.4. Bessel's Proof, 1838	327
17.5. Cauchy's Proofs, 1853	329
17.6. Ellis's Proof, 1844	333
17.7. Notes on Later Developments	335
17.8. Laplace's Diffusion Model, 1811	337
17.9. Gram-Charlier and Edgeworth Expansions	344
<b>18. Derivations of the Normal Distribution as a Law of Error</b>	<b>351</b>
18.1. Gauss's Derivation of the Normal Distribution and the Method of Least Squares, 1809	351
18.2. Laplace's Large-Sample Justification of the Method of Least Squares and His Criticism of Gauss, 1810	357
18.3. Bessel's Comparison of Empirical Error Distributions with the Normal Distribution, 1818	360
18.4. The Hypothesis of Elementary Errors by Hagen, 1837, and Bessel, 1838	365
18.5. Derivations by Adrain, 1808, Herschel, 1850, and Maxwell, 1860	368
18.6. Generalizations of Gauss's Proof: The Exponential Family of Distributions	373
18.7. Notes and References	380

<b>19. Gauss's Linear Normal Model and the Method of Least Squares, 1809 and 1811</b>	<b>381</b>
19.1. The Linear Normal Model	381
19.2. Gauss's Method of Solving the Normal Equations	383
19.3. The Posterior Distribution of the Parameters	386
19.4. Gauss's Remarks on Other Methods of Estimation	393
19.5. The Priority Dispute between Legendre and Gauss	394
<b>20. Laplace's Large-Sample Theory of Linear Estimation, 1811–1827</b>	<b>397</b>
20.1. Main Ideas in Laplace's Theory of Linear Estimation, 1811–1812	397
20.2. Notation	398
20.3. The Best Linear Asymptotically Normal Estimate for One Parameter, 1811	399
20.4. Asymptotic Normality of Sums of Powers of the Absolute Errors, 1812	401
20.5. The Multivariate Normal as the Limiting Distribution of Linear Forms of Errors, 1811	402
20.6. The Best Linear Asymptotically Normal Estimates for Two Parameters, 1811	405
20.7. Laplace's Orthogonalization of the Equations of Condition and the Asymptotic Distribution of the Best Linear Estimates in the Multiparameter Model, 1816	410
20.8. The Posterior Distribution of the Mean and the Squared Precision for Normally Distributed Observations, 1818 and 1820	418
20.9. Application in Geodesy and the Propagation of Error, 1818 and 1820	424
20.10. Linear Estimation with Several Independent Sources of Error, 1820	430
20.11. Tides of the Sea and the Atmosphere, 1797–1827	431
20.12. Asymptotic Efficiency of Some Methods of Estimation, 1818	444
20.13. Asymptotic Equivalence of Statistical Inference by Direct and Inverse Probability	452
<b>21. Gauss's Theory of Linear Unbiased Minimum Variance Estimation, 1823–1828</b>	<b>455</b>
21.1. Asymptotic Relative Efficiency of Some Estimates of the Standard Deviation in the Normal Distribution, 1816	455

21.2. Expectation, Variance, and Covariance of Functions of Random Variables, 1823	459
21.3. Gauss's Lower Bound for the Concentration of the Probability Mass in a Unimodal Distribution, 1823	462
21.4. Gauss's Theory of Linear Minimum Variance Estimation, 1821 and 1823	465
21.5. The Theorem on the Linear Unbiased Minimum Variance Estimate, 1823	467
21.6. The Best Estimate of a Linear Function of the Parameters, 1823	476
21.7. The Unbiased Estimate of $\sigma^2$ and Its Variance, 1823	477
21.8. Recursive Updating of the Estimates by an Additional Observation, 1823	480
21.9. Estimation under Linear Constraints, 1828	484
21.10. A Review	488

#### **PART IV SELECTED TOPICS IN ESTIMATION THEORY, 1830–1930**

<b>22. On Error and Estimation Theory, 1830–1890</b>	<b>493</b>
22.1. Bibliographies on the Method of Least Squares	493
22.2. State of Estimation Theory around 1830	494
22.3. Discussions on the Method of Least Squares and Some Alternatives	496
<b>23. Bienaymé's Proof of the Multivariate Central Limit Theorem and His Defense of Laplace's Theory of Linear Estimation, 1852 and 1853</b>	<b>501</b>
23.1. The Multivariate Central Limit Theorem, 1852	501
23.2. Bravais's Confidence Ellipsoids, 1846	504
23.3. Bienaymé's Confidence Ellipsoids and the $\chi^2$ Distribution, 1852	506
23.4. Bienaymé's Criticism of Gauss, 1853	509
23.5. The Bienaymé Inequality, 1853	510
<b>24. Cauchy's Method for Determining the Number of Terms To Be Included in the Linear Model and for Estimating the Parameters, 1835–1853</b>	<b>511</b>
24.1. The Problem	511
24.2. Solving the Problem by Means of the Instrumental Variable $\pm 1$ , 1835	512



24.3.	Cauchy's Two-Factor Multiplicative Model, 1835	516
24.4.	The Cauchy-Bienaymé Dispute on the Validity of the Method of Least Squares, 1853	520
<b>25.</b>	<b>Orthogonalization and Polynomial Regression</b>	<b>523</b>
25.1.	Orthogonal Polynomials Derived by Laplacean Orthogonalization	523
25.2.	Chebyshev's Orthogonal Polynomials, Least Squares, and Continued Fractions, 1855 and 1859	525
25.3.	Chebyshev's Orthogonal Polynomials for Equidistant Arguments, 1864 and 1875	535
25.4.	Gram's Derivation of Orthogonal Functions by the Method of Least Squares, 1879, 1883, and 1915	540
25.5.	Thiele's Free Functions and His Orthogonalization of the Linear Model, 1889, 1897, and 1903	550
25.6.	Schmidt's Orthogonalization Process, 1907 and 1908	556
25.7.	Notes on the Literature after 1920 on Least Squares Approximation by Orthogonal Polynomials with Equidistant Arguments	557
<b>26.</b>	<b>Statistical Laws in the Social and Biological Sciences. Poisson, Quetelet, and Galton, 1830–1890</b>	<b>567</b>
26.1.	Probability Theory in the Social Sciences by Condorcet and Laplace	567
26.2.	Poisson, Bienaymé, and Cournot on the Law of Large Numbers and Its Applications, 1830–1843	571
26.3.	Quetelet on the Average Man, 1835, and on the Variation around the Average, 1846	586
26.4.	Galton on Heredity, Regression, and Correlation, 1869–1890	599
26.5.	Notes on the Early History of Regression and Correlation, 1889–1907	616
<b>27.</b>	<b>Sampling Distributions under Normality</b>	<b>633</b>
27.1.	The Helmert Distribution, 1876, and Its Generalization to the Linear Model by Fisher, 1922	633
27.2.	The Distribution of the Mean Deviation by Helmert, 1876, and by Fisher, 1920	641
27.3.	Thiele's Method of Estimation and the Canonical Form of the Linear Normal Model, 1889 and 1903	645
27.4.	Karl Pearson's Chi-Squared Test of Goodness of Fit, 1900, and Fisher's Amendment, 1924	648

27.5. “Student’s” $t$ Distribution by Gosset, 1908	664
27.6. Studentization, the $F$ Distribution, and the Analysis of Variance by Fisher, 1922–1925	669
27.7. The Distribution of the Correlation Coefficient, 1915, the Partial Correlation Coefficient, 1924, the Multiple Correlation Coefficient, 1928, and the Noncentral $\chi^2$ and $F$ Distributions, 1928, by Fisher	675
<b>28. Fisher’s Theory of Estimation, 1912–1935, and His Immediate Precursors</b>	<b>693</b>
28.1. Notation	693
28.2. On the Probable Errors of Frequency Constants by Pearson and Filon, 1898	695
28.3. On the Probable Errors of Frequency Constants by Edgeworth, 1908 and 1909	697
28.4. On an Absolute Criterion for Fitting Frequency Curves by Fisher, 1912	707
28.5. The Parametric Statistical Model, Sufficiency, and the Method of Maximum Likelihood. Fisher, 1922	713
28.6. Efficiency and Loss of Information. Fisher, 1925	720
28.7. Sufficiency, the Factorization Criterion, and the Exponential Family. Fisher, 1934	727
28.8. Loss of Information by Using the Maximum Likelihood Estimate and Recovery of Information by Means of Ancillary Statistics. Fisher, 1925	729
28.9. Examples of Ancillarity and Conditional Inference. Fisher, 1934 and 1935	732
28.10. The Discussion of Fisher’s 1935 Paper	733
28.11. A Note on Fisher and His Books on Statistics	734
<b>References</b>	<b>741</b>
<b>Index</b>	<b>777</b>