

# CONTENTS

<b>PREFACE</b>	<b>xvi</b>
<b>ACKNOWLEDGMENTS</b>	<b>xix</b>
<b>INTRODUCTION: LANDMARKS IN PRE-LAPLACEAN STATISTICS</b>	<b>xx</b>
<b>PART ONE: LAPLACE</b>	<b>1</b>
<b>1 The Laplacean Revolution</b>	<b>3</b>
1.1 Pierre-Simon de Laplace (1749–1827),	3
1.2 Laplace’s Work in Probability and Statistics,	7
1.2.1 “Mémoire sur les suites récurro-récurrentes” (1774):	
Definition of Probability,	7
1.2.2 “Mémoire sur la probabilité des causes par les	
événements” (1774),	9
1.2.2.1 Bayes’ Theorem,	9
1.2.2.2 Rule of Succession,	13
1.2.2.3 Proof of Inverse Bernoulli Law. Method of Asymptotic	
Approximation. Central Limit Theorem for Posterior	
Distribution. Indirect Evaluation of $\int_0^\infty e^{-t^2} dt$ ,	14
1.2.2.4 Problem of Points,	18
1.2.2.5 First Law of Error,	19
1.2.2.6 Principle of Insufficient Reason (Indifference),	24
1.2.2.7 Conclusion,	25

- 1.2.3 “Recherches sur l’intégration des équations différentielles aux différences finis” (1776), 25
  - 1.2.3.1 Integration of Difference Equations. Problem of Points, 25
  - 1.2.3.2 Moral Expectation. On d’Alembert, 26
- 1.2.4 “Mémoire sur l’inclinaison moyenne des orbites” (1776): Distribution of Finite Sums, Test of Significance, 28
- 1.2.5 “Recherches sur le milieu qu’il faut choisir entre les resultants de plusieurs observations” (1777): Derivation of Double Logarithmic Law of Error, 35
- 1.2.6 “Mémoire sur les probabilités” (1781), 42
  - 1.2.6.1 Introduction, 42
  - 1.2.6.2 Double Logarithmic Law of Error, 44
  - 1.2.6.3 Definition of Conditional Probability. Proof of Bayes’ Theorem, 46
  - 1.2.6.4 Proof of Inverse Bernoulli Law Refined, 50
  - 1.2.6.5 Method of Asymptotic Approximation Refined, 53
  - 1.2.6.6 Stirling’s Formula, 58
  - 1.2.6.7 Direct Evaluation of  $\int_0^{\infty} e^{-t^2} dt$ , 59
  - 1.2.6.8 Theory of Errors, 60
- 1.2.7 “Mémoire sur les suites” (1782), 62
  - 1.2.7.1 De Moivre and Generating Functions, 62
  - 1.2.7.2 Lagrange’s Calculus of Operations as an Impetus for Laplace’s Generating Functions, 65
- 1.2.8 “Mémoire sur les approximations des formules qui sont fonctions de très grands nombres” (1785), 70
  - 1.2.8.1 Method of Asymptotic Approximation Revisited, 70
  - 1.2.8.2 Stirling’s Formula Revisited, 73
  - 1.2.8.3 Genesis of Characteristic Functions, 74
- 1.2.9 “Mémoire sur les approximations des formules qui sont fonctions de très grands nombres (suite)” (1786): Philosophy of Probability and Universal Determinism, Recognition of Need for Normal Probability Tables, 78
- 1.2.10 “Sur les naissances” (1786): Solution of the Problem of Births by Using Inverse Probability, 79
- 1.2.11 “Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités” (1810): Second Phase of Laplace’s Statistical Career, Laplace’s First Proof of the Central Limit Theorem, 83
- 1.2.12 “Supplément au Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités” (1810): Justification of Least Squares Based on Inverse Probability, The Gauss–Laplace Synthesis, 90
- 1.2.13 “Mémoire sur les intégrales définies et leur applications aux probabilités, et spécialement à la recherche du milieu qu’il faut choisir entre les résultats des observations” (1811): Laplace’s Justification of Least Squares Based on Direct Probability, 90

- 1.2.14 *Théorie Analytique des Probabilités (1812)*: The de Moivre–Laplace Theorem, 90
- 1.2.15 Laplace’s Probability Books, 92
  - 1.2.15.1 *Théorie Analytique des Probabilités (1812)*, 92
  - 1.2.15.2 *Essai Philosophique sur les Probabilités (1814)*, 95
- 1.3 The Principle of Indifference, 98
  - 1.3.1 Introduction, 98
  - 1.3.2 Bayes’ Postulate, 99
  - 1.3.3 Laplace’s Rule of Succession. Hume’s Problem of Induction, 102
  - 1.3.4 Bertrand’s and Other Paradoxes, 106
  - 1.3.5 Invariance, 108
- 1.4 Fourier Transforms, Characteristic Functions, and Central Limit Theorems, 113
  - 1.4.1 The Fourier Transform: From Taylor to Fourier, 114
  - 1.4.2 Laplace’s Fourier Transforms of 1809, 120
  - 1.4.3 Laplace’s Use of the Fourier Transform to Solve a Differential Equation (1810), 122
  - 1.4.4 Lagrange’s 1776 Paper: A Precursor to the Characteristic Function, 123
  - 1.4.5 The Concept of Characteristic Function Introduced: Laplace in 1785, 127
  - 1.4.6 Laplace’s Use of the Characteristic Function in his First Proof of the Central Limit Theorem (1810), 128
  - 1.4.7 Characteristic Function of the Cauchy Distribution: Laplace in 1811, 128
  - 1.4.8 Characteristic Function of the Cauchy Distribution: Poisson in 1811, 131
  - 1.4.9 Poisson’s Use of the Characteristic Function in his First Proof of the Central Limit Theorem (1824), 134
  - 1.4.10 Poisson’s Identification of the Cauchy Distribution (1824), 138
  - 1.4.11 First Modern Rigorous Proof of the Central Limit Theorem: Lyapunov in 1901, 139
  - 1.4.12 Further Extensions: Lindeberg (1922), Lévy (1925), and Feller (1935), 148
- 1.5 Least Squares and the Normal Distribution, 149
  - 1.5.1 First Publication of the Method of Least Squares: Legendre in 1805, 149
  - 1.5.2 Adrain’s Research Concerning the Probabilities of Errors (1808): Two Proofs of the Normal Law, 152
  - 1.5.3 Gauss’ First Justification of the Principle of Least Squares (1809), 159
    - 1.5.3.1 Gauss’ Life, 159
    - 1.5.3.2 Derivation of the Normal Law. Postulate of the Arithmetic Mean, 159
    - 1.5.3.3 Priority Dispute with Legendre, 163

- 1.5.4 Laplace in 1810: Justification of Least Squares Based on Inverse Probability, the Gauss–Laplace Synthesis, 166
- 1.5.5 Laplace’s Justification of Least Squares Based on Direct Probability (1811), 169
- 1.5.6 Gauss’ Second Justification of the Principle of Least Squares in 1823: The Gauss–Markov Theorem, 177
- 1.5.7 Hagen’s Hypothesis of Elementary Errors (1837), 182

## **PART TWO: FROM GALTON TO FISHER** **185**

### **2 Galton, Regression, and Correlation** **187**

- 2.1 Francis Galton (1822–1911), 187
- 2.2 Genesis of Regression and Correlation, 190
  - 2.2.1 Galton’s 1877 Paper, “Typical Laws of Heredity”: Reversion, 190
  - 2.2.2 Galton’s Quincunx (1873), 195
  - 2.2.3 Galton’s 1885 Presidential Lecture and Subsequent Related Papers: Regression, Discovery of the Bivariate Normal Surface, 197
  - 2.2.4 First Appearance of Correlation (1888), 206
  - \*2.2.5 Some Results on Regression Based on the Bivariate Normal Distribution: Regression to the Mean Mathematically Explained, 209
    - 2.2.5.1 Basic Results Based on the Bivariate Normal Distribution, 209
    - 2.2.5.2 Regression to the Mean Mathematically Explained, 211
- 2.3 Further Developments after Galton, 211
  - 2.3.1 Weldon (1890; 1892; 1893), 211
  - 2.3.2 Edgeworth in 1892: First Systematic Study of the Multivariate Normal Distribution, 213
  - 2.3.3 Origin of Pearson’s  $r$  (Pearson et al., 1896), 220
  - 2.3.4 Standard Error of  $r$  (Pearson et al., 1896; Pearson and Filon, 1898; Student, 1908; Soper, 1913), 224
  - 2.3.5 Development of Multiple Regression, Galton’s Law of Ancestral Heredity, First Explicit Derivation of the Multivariate Normal Distribution (Pearson et al., 1896), 230
    - 2.3.5.1 Development of Multiple Regression. Galton’s Law of Ancestral Heredity, 230
    - 2.3.5.2 First Explicit Derivation of the Multivariate Normal Distribution, 233
  - 2.3.6 Marriage of Regression with Least Squares (Yule, 1897), 237
  - 2.3.7 Correlation Coefficient for a  $2 \times 2$  Table (Yule, 1900). Feud Between Pearson and Yule, 244
  - 2.3.8 Intraclass Correlation (Pearson, 1901; Harris, 1913; Fisher, 1921; 1925), 253
  - 2.3.9 First Derivation of the Exact Distribution of  $r$  (Fisher, 1915), 258
  - 2.3.10 Controversy between Pearson and Fisher on the Latter’s Alleged Use of Inverse Probability (Soper et al., 1917; Fisher, 1921), 264

- 2.3.11 The Logarithmic (or Z-) Transformation (Fisher, 1915; 1921), 267
- \*2.3.12 Derivation of the Logarithmic Transformation, 270
- 2.4 Work on Correlation and the Bivariate (and Multivariate) Normal Distribution Before Galton, 270
  - 2.4.1 Lagrange's Derivation of the Multivariate Normal Distribution from the Multinomial Distribution (1776), 271
  - 2.4.2 Adrain's Use of the Multivariate Normal Distribution (1808), 275
  - 2.4.3 Gauss' Use of the Multivariate Normal Distribution in the *Theoria Motus* (1809), 275
  - 2.4.4 Laplace's Derivation of the Joint Distribution of Linear Combinations of Two Errors (1811), 276
  - 2.4.5 Plana on the Joint Distribution of Two Linear Combinations of Random Variables (1813), 276
  - 2.4.6 Bravais' Determination of Errors in Coordinates (1846), 281
  - 2.4.7 Bullet Shots on a Target: Bertrand's Derivation of the Bivariate Normal Distribution (1888), 288

### 3 Karl Pearson's Chi-Squared Goodness-of-Fit Test

293

- 3.1 Karl Pearson (1857–1936), 293
- 3.2 Origin of Pearson's Chi-Squared, 297
  - 3.2.1 Pearson's Work on Goodness of Fit Before 1900, 297
  - 3.2.2 Pearson's 1900 Paper, 299
- 3.3 Pearson's Error and Clash with Fisher, 306
  - 3.3.1 Error by Pearson on the Chi-Squared When Parameters Are Estimated (1900), 306
  - 3.3.2 Greenwood and Yule's Observation (1915), 308
  - 3.3.3 Fisher's 1922 Proof of the Chi-Squared Distribution: Origin of Degrees of Freedom, 311
  - \*3.3.4 Further Details on Degrees of Freedom, 313
  - 3.3.5 Reaction to Fisher's 1922 Paper: Yule (1922), Bowley and Connor (1923), Brownlee (1924), and Pearson (1922), 314
  - 3.3.6 Fisher's 1924 Argument: "*Coup de Grâce*" in 1926, 315
    - 3.3.6.1 The 1924 Argument, 315
    - 3.3.6.2 '*Coup de Grâce*' in 1926, 317
- 3.4 The Chi-Squared Distribution Before Pearson, 318
  - 3.4.1 Bienaymé's Derivation of Simultaneous Confidence Regions (1852), 318
  - 3.4.2 Abbe on the Distribution of Errors in a Series of Observations (1863), 331
  - 3.4.3 Helmert on the Distribution of the Sum of Squares of Residuals (1876): The Helmert Transformations, 336
  - \*3.4.4 Derivation of the Transformations Used by Helmert, 344

### 4 Student's $t$

348

- 4.1 William Sealy Gosset (1876–1937), 348
- 4.2 Origin of Student's Test: The 1908 Paper, 351

- 4.3 Further Developments, 358
  - 4.3.1 Fisher's Geometrical Derivation of 1923, 358
  - 4.3.2 From Student's  $z$  to Student's  $t$ , 360
- 4.4 Student Anticipated, 363
  - 4.4.1 Helmert on the Independence of the Sample Mean and Sample Variance in a Normal Distribution (1876), 363
  - 4.4.2 Lüroth and the First Derivation of the  $t$ -Distribution (1876), 363
  - 4.4.3 Edgeworth's Derivation of the  $t$ -Distribution Based on Inverse Probability (1883), 369

## 5 The Fisherian Legacy

371

- 5.1 Ronald Aylmer Fisher (1890–1962), 371
- 5.2 Fisher and the Foundation of Estimation Theory, 374
  - 5.2.1 Fisher's 1922 Paper: Consistency, Efficiency, and Sufficiency, 374
    - 5.2.1.1 Introduction, 374
    - 5.2.1.2 The Criterion of Consistency, 375
    - 5.2.1.3 The Criterion of Efficiency, 377
    - 5.2.1.4 The Criterion of Sufficiency, 377
  - 5.2.2 Genesis of Sufficiency in 1920, 378
  - 5.2.3 First Appearance of "Maximum Likelihood" in the 1922 Paper, 385
  - 5.2.4 The Method of Moments and its Criticism by Fisher (Pearson, 1894; Fisher, 1912; 1922), 390
  - 5.2.5 Further Refinement of the 1922 Paper in 1925: Efficiency and Information, 396
  - 5.2.6 First Appearance of "Ancillary" Statistics in the 1925 Paper: Relevant Subsets, Conditional Inference, and the Likelihood Principle, 403
    - 5.2.6.1 First Appearance of "Ancillary" Statistics, 403
    - 5.2.6.2 Relevant Subsets. Conditional Inference, 412
    - 5.2.6.3 Likelihood Inference, 417
  - 5.2.7 Further Extensions: Inconsistency of MLEs (Neyman and Scott, 1948), Inadmissibility of MLEs (Stein, 1956), Nonuniqueness of MLEs (Moore, 1971), 419
  - 5.2.8 Further Extensions: Nonuniqueness of Ancillaries and of Relevant Subsets (Basu, 1964), 421
- 5.3 Fisher and Significance Testing, 423
  - 5.3.1 Significance Testing for the Correlation Coefficient (Student, 1908; Soper, 1913; Fisher, 1915; 1921), 423
  - 5.3.2 Significance Testing for a Regression Coefficient (Fisher, 1922), 424
  - 5.3.3 Significance Testing Using the Two-Sample  $t$ -test Assuming a Common Population Variance (Fisher, 1922), 427
  - 5.3.4 Significance Testing for Two Population Variances (Fisher, 1924), 428
  - 5.3.5 *Statistical Methods for Research Workers* (Fisher, 1925), 429
- 5.4 ANOVA and the Design of Experiments, 431
  - 5.4.1 Birth and Development of ANOVA (Fisher and Mackenzie, 1923; Fisher, 1925), 431

- 5.4.2 Randomization, Replication, and Blocking (Fisher, 1925; 1926), Latin Square (Fisher, 1925), Analysis of Covariance (Fisher, 1932), 441
  - 5.4.2.1 Randomization, 441
  - 5.4.2.2 Replication, 442
  - 5.4.2.3 Blocking, 442
  - 5.4.2.4 Latin Square, 444
  - 5.4.2.5 Analysis of Covariance, 445
- 5.4.3 Controversy with Student on Randomization (1936–1937), 448
- 5.4.4 *Design of Experiments* (Fisher, 1935), 456
- 5.5 Fisher and Probability, 458
  - 5.5.1 Formation of Probability Ideas: Likelihood, Hypothetical Infinite Populations, Rejection of Inverse Probability, 458
  - 5.5.2 Fiducial Probability and the Behrens-Fisher Problem, 462
    - 5.5.2.1 The Fiducial Argument (1930), 462
    - 5.5.2.2 Neyman's Confidence Intervals (1934), 467
    - 5.5.2.3 The Behrens-Fisher Problem (1935), 470
    - 5.5.2.4 Controversy with Bartlett (1936–1939), 473
    - 5.5.2.5 Welch's Approximations (1938, 1947), 476
    - 5.5.2.6 Criticism of Welch's Solution (1956), 483
  - 5.5.3 Clash with Jeffreys on the Nature of Probability (1932–1934), 487
- 5.6 Fisher Versus Neyman–Pearson: Clash of the Titans, 502
  - 5.6.1 The Neyman-Pearson Collaboration, 502
    - 5.6.1.1 The Creation of a New Paradigm for Hypothesis Testing in 1926, 502
    - 5.6.1.2 The 'Big Paper' of 1933, 514
  - 5.6.2 Warm Relationships in 1926–1934, 520
  - 5.6.3 1935: The Latin Square and the Start of an Ongoing Dispute, 522
  - 5.6.4 Fisher's Criticisms (1955, 1956, 1960), 528
    - 5.6.4.1 Introduction, 528
    - 5.6.4.2 Repeated Sampling, 528
    - 5.6.4.3 Type II Errors, 532
    - 5.6.4.4 Inductive Behavior, 534
    - 5.6.4.5 Conclusion, 536
- 5.7 Maximum Likelihood before Fisher, 536
  - 5.7.1 Lambert and the Multinomial Distribution (1760), 536
  - 5.7.2 Lagrange on the Average of Several Measurements (1776), 541
  - 5.7.3 Daniel Bernoulli on the Choice of the Average Among Several Observations (1778), 544
  - 5.7.4 Adrain's Two Derivations of the Normal Law (1808), 550
  - 5.7.5 Edgeworth and the Genuine Inverse Method (1908, 1909), 550
- 5.8 Significance Testing before Fisher, 555
  - 5.8.1 Arbuthnot on Divine Providence: The First Published Test of a Statistical Hypothesis (1710), 555
  - 5.8.2 's Gravesande on the Arbuthnot Problem (1712), 562

- 5.8.3 Nicholas Bernoulli on the Arbuthnot Problem: Disagreement with ‘s Gravesande and Improvement of James Bernoulli’s Theorem (1712), 565
- 5.8.4 Daniel Bernoulli on the Inclination of the Planes of the Planetary Orbits (1735). Criticism by d’Alembert (1767), 571
- 5.8.5 Michell on the Random Distribution of Stars (1767): Clash Between Herschel and Forbes (1849), 578
  - 5.8.5.1 Michell on the Random Distribution of Stars (1767), 578
  - 5.8.5.2 Clash Between Herschel and Forbes (1849), 582
- 5.8.6 Laplace on the Mean Inclination of the Orbit of Comets (1776), 588
- 5.8.7 Edgeworth’s “Methods of Statistics” (1884), 588
- 5.8.8 Karl Pearson’s Chi-squared Goodness-of-Fit Test (1900), 590
- 5.8.9 Student’s Small-Sample Statistics (1908), 590

## **PART THREE: FROM DARMOIS TO ROBBINS 591**

### **6 Beyond Fisher and Neyman–Pearson 593**

- 6.1 Extensions to the Theory of Estimation, 593
  - 6.1.1 Distributions Admitting a Sufficient Statistic, 594
    - 6.1.1.1 Fisher (1934), 594
    - 6.1.1.2 Darmois (1935), 595
    - 6.1.1.3 Koopman (1936), 597
    - 6.1.1.4 Pitman (1936), 599
  - 6.1.2 The Cramér–Rao Inequality, 602
    - 6.1.2.1 Introduction, 602
    - 6.1.2.2 Aitken & Silverstone (1942), 603
    - 6.1.2.3 Fréchet (1943), 607
    - 6.1.2.4 Rao (1945), 611
    - 6.1.2.5 Cramér (1946), 614
  - 6.1.3 The Rao–Blackwell Theorem, 618
    - 6.1.3.1 Rao (1945), 618
    - 6.1.3.2 Blackwell (1947), 620
  - 6.1.4 The Lehmann–Scheffé Theorem, 624
    - 6.1.4.1 Introduction, 624
    - 6.1.4.2 The Lehmann–Scheffé Theorem. Completeness (1950), 626
    - 6.1.4.3 Minimal Sufficiency and Bounded Complete Sufficiency (1950), 629
  - 6.1.5 The Ancillarity–Completeness–Sufficiency Connection: Basu’s Theorem (1955), 630
  - 6.1.6 Further Extensions: Sharpening of the CR Inequality (Bhattacharyya, 1946), Variance Inequality without Regularity Assumptions (Chapman and Robbins, 1951), 632
- 6.2 Estimation and Hypothesis Testing Under a Single Framework: Wald’s Statistical Decision Theory (1950), 634
  - 6.2.1 Wald’s Life, 634

- 6.2.2 Statistical Decision Theory: Nonrandomized and Randomized Decision Functions, Risk Functions, Admissibility, Bayes, and Minimax Decision Functions, 636
- 6.2.3 Hypothesis Testing as a Statistical Decision Problem, 641
- 6.2.4 Estimation as a Statistical Decision Problem, 642
- 6.2.5 Statistical Decision as a Two-Person Zero-Sum Game, 643
- 6.3 The Bayesian Revival, 645
  - 6.3.1 Ramsey (1926): Degree of Belief, Ethically Neutral Propositions, Ramsey's Representation Theorem, Calibrating the Utility Scale, Measuring Degree of Belief, and The Dutch Book, 646
  - 6.3.2 De Finetti (1937): The Subjective Theory, Exchangeability, De Finetti's Representation Theorem, Solution to the Problem of Induction, and Prevision, 656
  - 6.3.3 Savage (1954): The Seven Postulates, Qualitative Probability, Quantitative Personal Probability, Savage's Representation Theorem, and Expected Utility, 667
  - 6.3.4 A Breakthrough in "Bayesian" Methods: Robbins' Empirical Bayes (1956), 674

**REFERENCES****681****INDEX****714**