Brazilian Journal of Probability and Statistics
2021, Vol. 35, No. 4, 657–698
https://doi.org/10.1214/21-BJPS518

The foundations of statistical science: A history of textbook presentations

Alan Agresti

Department of Statistics, University of Florida, Gainesville, Florida 32611, USA, agresti@ufl.edu

Abstract. We discuss how the foundations of statistical science have been presented historically in textbooks, with focus on the first half of the twentieth century after the field had become better defined by advances due to Francis Galton, Karl Pearson, and R. A. Fisher. Our main emphasis is on books that presented the theory underlying the subject, often identified as *mathematical statistics*, with primary focus on books authored by G. Udny Yule, Maurice Kendall, Samuel Wilks, and Harald Cramér. We also discuss influential books on statistical methods by R. A. Fisher and George Snedecor that showed scientists how to implement the theory. We then survey textbooks published in the quarter century after World War 2, as Statistics gathered more visibility as an academic subject and Departments of Statistics were formed at many universities. We also summarize how Bayesian presentations of Statistics emerged. In each section, we describe how the books were evaluated in reviews shortly after their publications. We conclude by briefly discussing the recent past, the present, and the future of textbooks on the foundations of statistical science and include comments about this by several notable statisticians.

1 Introduction

In this article, we discuss how the foundations of statistical science have been presented in textbooks, with main focus on the first half of the twentieth century after the field had the great advances that define it today. In 1901 Arthur Bowley (1869–1957) published *Elements of Statistics*, the first of seven editions of a book for which its contents serves to illustrate how much would soon change in what was regarded as the essence of the field. By that time, Karl Pearson's interactions with the biologist W. F. R. Weldon and with Francis Galton had led to his conducting statistical work that helped to lay the foundations of modern statistical science (Magnello 2009), such as with his 1900 paper devising the chi-squared test of goodness-of-fit. In 1911 Pearson founded the first academic Statistics department, at University College, London. In the same year, George Udny Yule published *An Introduction to the Theory of Statistics*, which could be considered the first book that dealt with some aspects of statistical theory in a more modern, although relatively non-technical, sense.

Focusing mainly on the first half of the twentieth century, we discuss textbooks that followed the development of key modern concepts of statistical science, in particular, (1) Francis Galton's introduction in the late 1880s of concepts of correlation and regression and their subsequent connection with the least squares method that had been presented near the beginning of the century by Gauss, Legendre, and Laplace, (2) Karl Pearson's many innovations around the turn of the century, including the product-moment estimate of the correlation, families of skewed distributions and chi-squared testing of hypotheses about multinomial probabilities, principal components analysis, and the method of moments for estimating parameters, and (3) Ronald Aylmer (R. A.) Fisher's introduction in the early 1920s of fundamental concepts

Key words and phrases. History of statistics, mathematical statistics, statistical theory, G. Udny Yule, R. A. Fisher, Maurice Kendall, Samuel Wilks, Harald Cramér, George Snedecor.
Received September 2021; accepted September 2021.

such as maximum likelihood, sufficiency, and efficiency, and of widely-applicable methods such as the analysis of variance, and Fisher's promotion of Willam S. Gosset t distribution for applying significance tests about means. Although limiting our discussion to English-language textbooks in the twentieth century, we acknowledge that elements of modern statistics had been published as books well before then. Important examples are Pierre-Simon Laplace's 1812 book *Théorie Analytique des Probabilités*, which dealt with much more than probability theory, and Adolphe Quetelet's 1835 book *A Treatise on Man and the Development of His Faculties*, which introduced statistical methods to the social sciences by formulating the concept of the "average man" and fitting distributions to data.

In focusing on statistical foundations and thus on books about statistical *theory* or what later became identified as *mathematical statistics*, we purposely omit discussion of the quite large number of books that had the purpose of presenting statistical *methods* and their application in various fields. In this omission, however, we exempt two books because of the great impact they had both in presenting modern statistical methods for the first time as well as themselves introducing related methods, and as a consequence had immense popularity among scientists conducting data analysis during the middle half of the 20th century: the highly innovative and influential *Statistical Methods for Research Workers* by R. A. Fisher (1925) and the more reader-friendly *Statistical Methods* by George Snedecor (1937).

The many editions of Yule's *An Introduction to the Theory of Statistics*, continued by Maurice Kendall in 1937 with the 11th edition, seem to have been the primary textbook source on the foundations of Statistics for more than 30 years. Sections 2 and 3 of this article summarize key textbooks in the period before World War 2, with Section 2 discussing the Yule and Yule/Kendall contribution in Britain and Section 3 covering corresponding publications during this period in the United States. Section 4 then discusses the Fisher and Snedecor books on statistical methods that had such a great impact on scientists as well as statisticians during this period. The years during and immediately following World War 2 found publication of a very different type of book, the *mathematical statistics* form of text that has been written in various forms ever since then and that is the primary focus of this article. Section 5 discusses these highly influential books, authored by Samuel Wilks, Maurice Kendall, and Harald Cramér. Section 6 considers the explosion of mathematical statistics textbooks in North America and Europe that appeared in the quarter century following World War 2, during a period in which Statistics departments were formed in most large universities in North America. Section 7 considers the parallel development of textbooks that advocated the Bayesian approach to probability and statistical inference, beginning with the foundational book by Harold Jeffreys. In all these sections, we also summarize evaluations of the books, positive and negative, in book reviews shortly after their publications. The final section briefly summarizes developments in the past half century, as the field of Statistical Science has reached a mature stage and the related field of Data Science has gained more visibility, with new programs appearing every year now in universities around the world. That section considers the question of what aspects of statistical science should be presented as the foundations of the subject now and in the future. We conclude with comments about this by several notable statisticians.

2 Early twentieth-century statistical theory in Britain: Yule and Yule/Kendall textbooks

The preface of *Elements of Statistics* by Arthur Bowley (1901), based on lectures that he presented at the London School of Economics and Political Science in the five years following its foundation in 1895 (and which he continued for 40 years), states that "There seems to be no text-book in English dealing directly and completely with the common methods of statistics.

… In the excellent books in French, German, and Italian on this subject there is a general tendency to deal at length with the history of official statistics, the limits of the science, and particular applications of the theory of probability, to the exclusion of more general matter." Bowley's knowledge of Statistics seem to have been influenced more by his friendship with the Anglo-Irish economist/statistician Francis Ysidro Edgeworth than by Karl Pearson, who by then was in his mid-40s and establishing his authority over the subject in Britain. Yet Bowley's book keeps mathematics to a minimum and contains very little statistical theory. His stated goal was to put "in the simplest possible way those formulae and ideas which appear to be the most useful in the fields of economic and social investigation." The first 80% of the 330 pages focus on tabulations, averages, graphic methods, accuracy, index numbers, and interpolation. The short second part of the book focuses on basic laws of probability, with special attention to binomial probabilities and approximations for them using the normal distribution. It shows standard error formulas for means and differences of means, but otherwise does not consider statistical inference. The flavor of the presentation reflects Bowley's statement in the preface that the treatment in that second part "leaves very much in the background the Method of Least Squares; the phrase, useful in some connections, seems to make the application of the Law of Error to statistics unnecessarily complex." (He never defines "Law of Error" but seems to use it throughout the book to refer to the normal variability of sample means in summarizing many small, independent effects.) The final 10 pages of the book introduce the correlation, concluding by mentioning Galton, regression toward the mean, and regression equations and their importance in the study of evolution. The book had multiple editions, the final (7th) appearing in 1937.

George Udny Yule (1871–1951), who began working in University College, London, in 1893 as a demonstrator for Karl Pearson, was influenced by Pearson to make Statistics his profession. The extent of his accomplishments is perhaps not appreciated today as much as it should be. For instance, in an 1897 article on the theory of correlation, he extended Pearson's landmark 1896 article about the Pearson product moment correlation and multiple regression and correlation by applying least squares to calculate the partial regression coefficients and the multiple correlation and then introducing the *net* coefficient of correlation, what became known as the partial correlation. Yule showed that the conditional standard deviation around the fitted relationship would decrease with the addition of an explanatory variable unless its partial correlation were 0. Whereas Pearson had focused on finding appropriate frequency surfaces, Yule moved away from unneeded assumptions such as normality. In a 1907 paper about the theory of correlation and regression for multiple variables, he introduced what became standard notation for partial regression coefficients, showed that the estimated partial effect of x_j is identical to the regression of the residuals from regressing y on the explanatory variables other than x_j on the residuals from regressing x_j on those other explanatory variables, and he showed how the partial correlation is the ordinary correlation between residuals for two models. Several articles between 1900 and 1912 discussed association in contingency tables, including advocating the odds ratio and showing in 1903 the potential discrepancy between marginal and conditional associations, studied much later in 1951 by E. H. Simpson and now called *Simpson's paradox*. In this work, Yule was critical of Pearson's approach of assuming that continuous bivariate distributions underlie two-way contingency tables. In the 1920s he published influential papers that introduced a modern approach to time series analysis, including devising the correlogram and laying the foundation for autoregressive modeling. He often wrote about the dangers of spurious correlation, such as by pointing out how one can get nonsense correlations between two time series. He lectured about Statistics at University College from 1902 until 1912, with the 1911 publication of *An Introduction to the Theory of Statistics* summarizing his lectures. For more about Yule, see Stigler (1986, pp. 345–358) and Yates (1952).

Table 1 *Contents of Yule's* An Introduction to the Theory of Statistics (1911)

Chapter title	Pages range
1. Notation and terminology	7–16
2. Consistence	17–24
3. Association	25–41
4. Partial association	42–59
5. Manifold classification	60–74
6. The frequency-distribution	75–105
7. Averages	106–132
8. Measures of dispersion, etc.	133–156
9. Correlation	157–190
10. Correlation: Practical applications and methods	191–206
11. Miscellaneous theorems involving the use of the correlation-coefficient	207–224
12. Partial correlation	225–249
13. Simple sampling of attributes	250–271
14. Simple sampling continued: Effect of removing the limitation of simple sampling	272–286
15. The binomial distribution and the normal curve	287–312
16. Normal correlation	313–330
17. The simpler cases of sampling for variables: Percentiles and mean	331–351

The Table of Contents of the first edition of *An Introduction to the Theory of Statistics* is a good portrayal of the state of statistical science in Britain in 1911. Table 1 shows chapter titles, listing also the range of page numbers to indicate the amount of attention paid to each topic. Although the title alludes to *theory*, in the preface Yule states "The volume represents an attempt to work out a systematic introductory course on statistical methods … suited to those who possess only a limited knowledge of mathematics." The mathematical level is higher than in Bowley's book, but still low, mainly requiring algebra and analytic geometry with some subtle calculus in discussions of least squares. Unlike Bowley's book, Yule presented mathematical derivations of key results. Several chapters are quite different in content than in Bowley's book, for instance with substantial material included on association measures and on dealing with multiple variables and conditional as well as marginal associations. Specifically, topics in Yule's book but not in Bowley's include the Chapter 3 material on association in 2×2 contingency tables, the Chapter 4 material on generalizations to $2 \times 2 \times 2$ tables (basically explaining results from articles by Yule in 1900 and 1903) and the Chapter 5 extension to more general cases, the Chapter 12 derivation of partial regression coefficients for multiple regression models and the corresponding partial correlation measure, and the Chapter 16 discussion of correlation and regression in the context of the bivariate normal distribution and the extension to the multivariate normal.

Each chapter of Yule's textbook contains exercises, dealing both with theory and with analyses of real data. Here are typical ones of the first type, from Chapters 11 and 12:

- Find the correlation between $X_1 + X_2$ and $X_1 + X_3$; X_1, X_2, and X_3 being uncorrelated.
- If the relation $ax_1 + bx_2 + cx_3 = 0$ holds for all sets of values of x_1, x_2, and x_3, what must the partial correlation be?

Appendix II in the book lists 22 books published before Yule's, mainly focusing on probability, but including the books by Quetelet and Bowley. Reviews of Yule's book were generally complimentary, commending Yule on a useful introduction to the Galton/Pearson school of Statistics. For example, a 1911 review in *Journal of the Royal Statistical Society* (*JRSS*, by a reviewer identified as A. W. F.) states "The increasing tendency to rely on statistics in political and other discussions has made the work of systematising the somewhat scattered material of growing urgency. … In the work before us we have a more thorough and well-planned reduction to proper sequence of the different parts of the subject than has yet been produced in our own, or, so far as the writer is aware, in any language. … The doubts which have sometimes been expressed … whether there be a science of statistics, or only a method of statistical analysis, are now definitely resolved, for the work of systematization accomplished by Mr. Yule has left the scientific nature of statistical theory plain to the most hardened doubter." However, likely because Yule had proposed different approaches to modeling association with discrete data and not promoted the measures he had introduced, Karl Pearson was highly critical of Yule's book, stating in a 1913 article with D. Heron in *Biometrika*, "If Mr. Yule's views are accepted, irreparable damage will be done to the growth of modern statistical theory. … Unthinking praise has been bestowed on a text-book which at many points can only lead statistical students hopelessly astray." Not surprisingly, Yule left University College, moving in 1912 to Cambridge University for the remainder of his career.

A chance meeting with Maurice Kendall (1907–1983) at a time when Yule's health did not permit preparing new editions resulted in Kendall's continuing to revise the book, from the 11th edition in 1937 (which increased the 434-page 10th edition in length to 570 pages) until its final 700-page-long 14th edition in 1950 (Yule and Kendall 1950). Perhaps surprisingly, none of the editions solely authored by Yule or jointly with Kendall made more than a superficial mention of Fisher, focusing mainly on Fisher's correction of the degrees of freedom for Pearson's chi-squared test and the distribution of the correlation rather than Fisher's important contributions in the 1920s on statistical theory or on the design of experiments. For the editions with Kendall, the main additions were new chapters on moments and measures of skewness and kurtosis, the chi-squared distribution, small-sample methods, the analysis of variance, and time series. However, the book has no mention of maximum likelihood or the Neyman–Pearson theory of hypothesis testing. Kendall's contributions to Yule's book were only the beginning of an incredibly prodigious output of Statistics books, partly described in Sections 5 and 6 of this article.

To illustrate the influence of Yule's book, we quote from a review by Jerzy Neyman in *Nature* of the 1937 edition with Kendall: "Mr. Udny Yule's "Introduction" was the first book on statistics that came into my hands. This was about 1916. I liked it then and learned a lot from it. I like it even more now in its eleventh edition, when it appears in a revised and supplemented form. In fact, in my opinion, this is the best book on the theory of statistics that was ever written." Yet, Neyman also wrote, "We may regret only that it is an elementary introduction to the theory of statistics and not the theory itself. This still remains to be written." Neyman could not help but also use the review to apparently throw a barb at Fisher's popular book on statistical methods, suggesting that to appreciate how clearly Yule and Kendall defined terms, "the reader may wish to consult some of the popular books on the "Methods" where, instead of properly ordered definitions and deductions, he will frequently find a sort of mystic gospel full of mysterious terms and impressive formulae."

3 Early mathematical statistics textbooks in the U.S.

It may have taken some time for early editions of Yule's book to have much influence in North America. In 1918 Carl J. West, who was a mathematics professor at Ohio State from 1910 to 1918 before leaving academic life to join the U.S. Department of Agriculture, published *Introduction to Mathematical Statistics* (West 1918). Strongly influenced by Karl Pearson, he focused on descriptive statistics and graphics including methods of correlation, association in contingency tables, and Pearson curves, but assumed knowledge of only elementary mathematics. The book's exercises focus on calculations rather than theory. This was perhaps the first occurrence of the term "mathematical statistics" in a book title, although at least as far back as 1872 the term appears in an article. Wittstein (1872), based on a talk given at a conference in Hanover, Germany in 1865, argued for using that name for the new science that employs mathematics "to ascend from observations to the discovery of natural laws." (Stigler (1978) surveyed the history of mathematical statistics in the U.S. prior to 1885, and Chapter 8 of Stigler (1999) is a highly entertaining argument that mathematical statistics did not seriously begin until 1933.)

A stronger theoretical presentation of Statistics was made by Henry Lewis Rietz (1875–1943), who spent most of his career in the mathematics department at the University of Iowa. Rietz was a founding member of the Institute of Mathematical Statistics in 1935 and its first president. After noting in 1923 that he could find only seven institutions that offered courses in mathematical statistics (Rietz 1923), a year later he edited *Handbook of Mathematical Statistics*, with 12 chapters on topics such as random sampling, curve fitting by least squares, and partial and multiple correlation (Rietz 1924). In 1927 he published his own book, *Mathematical Statistics* (Rietz 1927). In its preface, Rietz mentions the goal of "shifting the emphasis and point of view in the study of Statistics in the direction of the consideration of the underlying theory involved in certain highly important methods of statistical analysis. ... Elementary methods of description and analysis of data by purely graphical methods and by the use of various kinds of averages and measures of dispersion are for the most part omitted owing to the fact that these methods are so available in recent elementary books." The book has somewhat higher mathematical level than Yule's, in many places assuming knowledge of differential and integral calculus.

Table 2 shows the table of contents for Rietz's textbook. Chapter 1 introduces the relative frequency definition of probability, mathematical expectation, and moments of random variables. Chapter 2 introduces the binomial distribution and its normal and Poisson approximations. Chapter 3 focuses on alternative distributions, with main focus on Pearson's families. Chapter 4 introduces the Pearson estimate of correlation, linear regression with least squares estimates and the bivariate normal distribution, and extensions such as the multiple correlation and partial correlation. Chapter 5 focuses on the standard error in estimation using random sampling, for a variety of measures, including the sample proportion, mean, variance, and median, and briefly mentions the 1908 Student (pseudonym for Gosset) paper for dealing with small samples. Chapter 6 presents ideas published in 1879 by the German statistician–economist Wilhelm Lexis about handling situations in which the probability of an event is constant from trial to trial within a set but varies from set to set, such as in time series or data stratified by region of observation. The highly-technical final chapter uses a Gram–Charlier series for improvements to the normal and Poisson approximations for the binomial distribution. Like Yule's many editions but unlike books published ten years later, this book does not discuss Fisher's landmark work five years earlier in introducing maximum likelihood and properties such as sufficiency, consistency, and efficiency. (His only mention of Fisher is of his 1915 paper about the distribution of the sample correlation coefficient.) In

Table 2 *Contents of Rietz's* Mathematical Statistics (1927)

Chapter title	Pages range
1. The nature of the problems and underlying concepts of mathematical statistics	1–21
2. Relative frequencies in simple sampling	22–45
3. Frequency functions of one variable	46–76
4. Correlation	77–113
5. On random sampling fluctuations	114–145
6. The Lexis theory	146–155
7. A development of the Gram-Charlier series	156–172

fact, the book does not discuss statistical inference, such as hypothesis testing. Unlike Yule's editions, Rietz's book did not contain exercises.

Rietz's book had only a brief review by E. B. Wilson in *Journal of the American Statistical Assoclation* (*JASA*) in 1927 but a very positive 1928 review by W. L. Crum in *Bulletin of the American Mathematical Society*, stating "The book does render available, as does no other volume written in English which is known to the reviewer, the essentials of an introductory survey of the underlying mathematical theory which must receive increasing attention from specialists in statistics, if the widespread use of statistical method is to be systematically helpful." A rather negative review in 1928 in *JRSS* (by a reviewer identified as E. C. R.) criticized the lack of real-data examples and interpretations and the choice of topics. For instance, "one feels that the illustrations of the various Pearson types are over-weighting this subject in a small book, the room could well be used for a talk on the inferences from a sample as to the aggregate. ... A good deal of the mathematical analysis dealing with multiple correlation might well give place to a talk on the meaning of the various measures of correlation (what is meant when we say that $r = .5$?)." In summary, "a book of this kind is useful to the mathematical statistician, even with its limitations, but one doubts its appeal to the thoughtful people who wish to extend their knowledge."

The year 1927 also saw the publication of *Introduction to The Mathematics of Statistics* by Robert Wilbur Burgess (1887–1969), a statistician at Western Electric Company and later the Director of the U.S. Census Bureau from 1953 to 1961. This book had lower mathematical level than Rietz's, not requiring calculus, and despite the title it may be charitable to call it a book on mathematical statistics. In 1928, Burgess's book had a mainly positive review in *JASA* by Henry Schultz, although critical of the brief treatment of statistical inference, lack of distinction between correlation and causation, and insufficient interpretation of coefficients in multiple regression. A review by Rietz in *The American Mathematical Monthly* was mixed, admiring the presentation of regression but criticizing the treatment of elementary descriptions such as ratios and percentages. Of higher mathematical level and apparently more influential was the 1928 publication *Probability and its Engineering Uses* by Thornton Fry (1892–1991), an industrial mathematician originally at Western Electric Company but at Bell Telephone Laboaratories for most of his career. Fry's book was based on a course he taught at Bell Labs and MIT on the theory of probability applied to electrical problems, in particular those that arise in the work of the telephone exchange, such as congestion. The title is somewhat misleading, because although the first six chapters introduce a mathematical approach to probability (including Bayes theorem), the last five chapters focus on applications to different subjects in Statistics, including a substantial treatment of curve fitting and chi-squared goodness of fit using the binomial, Poisson, Pearson curves, and Gram-Charlier series. Fry did not discuss correlation and regression. A *JRSS* review (by E. C. R.), although containing criticisms of some the applications of probability, states "we may venture to accord praise for a very fine piece of work" and points out the potential more general interest, "the problems met with in the consideration of an efficient telephone service might be of interest to medical statisticians, if they could find analogies between the crowding of subscribers on telephone exchanges and the crowding of patients into hospitals or of bacteria in the body." A second edition was published in 1965, for which Saul Blumenthal wrote a long and quite critical review in *Technometrics* in 1966.

In retrospect, Burgess's and Fry's books were not important ones, but in reading tables of contents of book reviews in major statistics journals before World War 2, one is struck by the relative banality of most titles. For instance, the June 1929 issue of *JASA* has book reviews of titles such as *Stabilization of Prices*, *The Construction of Index Numbers*, *The Prosperity of Australia*, and *Petroleum and Coal, the Keys to the Future*. The rare exception is the occasional Harold Hotelling review of new editions of Fisher's *Statistical Methods for Research Workers* and *The Design of Experiments*.

Table 3 *Contents of Fisher's* Statistical Methods for Research Workers (1925)

Chapter title	Pages range
1. Introduction	1–26
2. Diagrams	27–42
3. Distributions	43–76
4. Tests of goodness of fit, independence and homogeneity, with table of χ^2	77–100
5. Tests of significance of means, differences of means, and regression coefficients	101–137
6. The correlation coefficient	138–175
7. Intraclass correlations and the analysis of variance	176–210
8. Further applications of the analysis of variance	211–233

4 Fisher and Snedecor textbooks on statistical methods

Ronald Aylmer Fisher (1890–1962) started publishing statistical research papers in 1915, and his work beginning in 1919 at Rothamsted Experimental Station led naturally to his 1925 publication of *Statistical Methods for Research Workers*. This book contains material on applied statistics not found in any other books at that time. This was the first presentation in book form of maximum likelihood, t tests, the analysis of variance, and randomization and blocking in the design of experiments. Table 3 shows the table of contents for the first edition. Chapter 1, although called "Introductory," explains what is meant by consistent, efficient, and sufficient statistics and the method of maximum likelihood. Chapter 4 points out Karl Pearson's error in stating the degrees of freedom for the chi-squared test. Of particular note is Chapter 5, which applies Student's t distribution (introduced by W. S. Gosset in 1908) to significance testing of the mean with small samples. Tables at the end of the book include a table of quantiles of the t distribution for many quantiles (including 0.90, 0.95, 0.98, 0.99) and degrees of freedom values between 1 and 30 that made it much easier for research workers to start using the method with small samples. Fisher also thanked Gosset in the preface for reading the proofs and making suggestions. Chapter 7 seems to contain the first textbook example of the analysis of variance for an actual experiment, with Chapter 8 showing an extension to what is now called a split-plot design. The second edition, published three years later in 1928, added a chapter on "The principles of statistical estimation," and the same chapter titles remained for later editions. The 4th edition added the analysis of covariance as a topic. The 14th and final edition was published in 1970, with E. A. Cornish completing it based on notes made by Fisher before his death in 1962.

Karl Pearson's son Egon reviewed the first edition, giving a mixed review that reflected Fisher's neglecting most of Karl Pearson's ideas, although concluding "But if old methods are dismissed somewhat summarily there are several fresh results of considerable interest as well as new tables, and anyone interested in the theory of small samples can hardly pass over Mr. Fisher's contributions to the subject." Harold Hotelling wrote a very favorable review in *JASA*, finishing the review by stating "The author's work is of revolutionary importance and should be far better known in this country." He also reviewed the next six editions for the journal and wrote an article about Fisher's influence in an issue of the journal devoted to celebrating the 25th anniversary of the publication of the first edition (Hotelling 1951). The 1990 republication of this book together with Fisher's other two books to celebrate the centenary of his birth contains an informative foreward by Frank Yates, Fisher's successor at Rothamsted Experiment Station. Yates indicated that Fisher intended the book as a laboratory notebook rather than as a student textbook, and he felt it important to include real-data examples in each chapter as a template for use of the methods. Yates argued that because of Pearson's influence, other applied statistics books had placed too much attention on curve fitting and various forms of correlation and association and confused estimating the degree of association and testing the significance of its existence, and also failed to heed the needs of experimenters having only small samples.

Before Fisher's book, Stephen Stigler has noted to me (in email correspondence) that Karl Pearson's 1914 *Tables for Statisticians and Biometricians*, with later editions in 1924 and 1930 and a Part II in 1931, was perhaps more influential in instruction in applied statistics than any other book. Although not regarded as a textbook, its 143 pages of tables were preceded by a 71-page tutorial on how to use the tables. The 55 tables included standard normal quantiles, multiples of the standard deviation to find the probable error of the sample mean and standard deviation for sample sizes up to 1000, probable errors for the correlation, tail probabilities for Pearson's chi-squared statistic (though with incorrect degrees of freedom, see, e.g., 3 in testing for association in a 2×2 table), and related tables for fitting Pearson curves. A review by J. Arthur Harris in *Science* in 1914 praised this for the great labor needed to produce it by a single laboratory and for its originality of contents. In 1933, Fisher himself rather generously wrote in *Nature* that "statisticians have for long been familiar with Prof. K. Pearson's "Tables for Statisticians" as the standard exposition of the theoretical conceptions and practical procedures adopted by the Biometric Laboratory at University College. The handsome and expensive production of the tables, and the abundant introductory matter, were features which, from its first appearance, compensated for the partial or personal tone which characterised the treatment of controversial topics."

While Fisher's book influenced many scientists who analyzed data between its publication in 1925 and his death, it was not an easy read for most of them. The reviewer (Leon Isserlis) of the first edition for *JRSS* stated "Much is lacking if the book is to be regarded as an authoritative record of achievement in statistical method apart from Mr. Fisher's own contributions.... The book will undoubtedly prove of great value to research workers whose statistical series necessarily consist of small samples, but will prove a hard nut to crack for biologists who attempt to use it as a first introduction to statistical method." Gosset himself, in *Eugenics Review*, concludes a review by stating "Dr. Fisher's book will doubtless be found in the laboratories of those who realise the necessity for statistical treatment of experimental results, but it should not be expected that full, perhaps even in extreme cases any, use can be made of such a book without contact either personal or by correspondence with someone familiar with its subject matter." As a consequence, the book that probably had a greater impact on increasing the use of statistical methods during most of that period was George Snedecor's *Statistical Methods* (Snedecor 1937). Appointed to the mathematics department at Iowa State College in 1913, Snedecor (1881–1974) taught courses in Statistics there beginning in 1915, and in 1927 formed the Mathematics Statistical Service to provide consulting help. In 1933 the Statistical Laboratory was formed, with Snedecor as Director. In 1935 the Statistical Section of the Iowa Agricultural Experiment station was formed, with Snedecor as Section Head, and the first edition of his book added the subtitle *Applied to Experiments in Agriculture and Biology*. The topics covered in Snedecor's book were certainly greatly influenced by Fisher (who has 24 citations in its index, compared to 3 for Karl Pearson and none for Yule). Snedecor was one of first statisticians in the U.S. to appreciate the importance of Fisher, inviting him for extended visits to Iowa State in 1931 and 1936 for summer courses with scientists from around the country. In 1936, Iowa State gave Fisher his first honorary doctorate.

Table 4 shows the table of contents of the first edition of Snedecor's book. It presented the statistical methods in a more-detailed manner than Fisher's book that would be simpler for most scientists to follow. Each chapter contained numerous examples, many stated in the form of exercises for the reader to perform. Statistical inference appears right from the first chapter, with the third page distinguishing between sample and population, the following

Table 4 *Contents of Snedecor's* Statistical Methods (1937)

Chapter title	Pages range
1. Experiments on attributes	3–21
2. An experiment designed to compare measurements of individuals	22–42
3. Sampling from a normally distributed population	43–62
4. An experiment designed to compare two groups	63–74
5. Short cuts and approximations	75–88
6. Linear regression	89–114
7. Correlation	115–133
8. Large sample theory	134–153
9. Enumeration data with multiple degrees of freedom	154–170
10. Experiments involving more than two groups of measurement data. Analysis of variance	171–197
11. Analysis of variance with two criteria of classification	198–218
12. Two variates in two or more groups. Covariance	219–241
13. Multiple regression and covariance	242–263
14. Curvilinear regression	264–290
15. Individual degrees of freedom	291–316
16. Large samples of enumeration data. Binomial and Poisson distributions	317–334

page introducing hypotheses about parameters and then a chi-squared test of goodness of fit. Chapter 5 presented suggestions about computations, including methods to use with calculating machines. Snedecor was also influenced by Fisher's 1935 book on *Design of Experiments*, as that topic receives considerable attention throughout the book.

A review in *JASA* in 1938 by A. E. Treloar stated "With simplicity of verbal exposition as a keynote, the readers are addressed directly in the informal conversational style of a laboratory discussion between an understanding teacher and his responsive student" but criticizes the book for having insufficient depth about the theory: "There is great danger in teaching statistical methods without building at the same time a clear understanding of the reasons validating each procedure." Likewise in a 1939 review of the 2nd edition in *JASA*, W. Edward Deming complains that "nowhere in the book is there any attention given to the question, "When do these methods apply?" ... A large portion of the book is occupied with tests of hypotheses but without any discussion of how a test should be selected; the fundamental researches of Neyman and Pearson in this regard seem to be completely overlooked" yet he concludes by stating "It is always easy to find fault. The fact is that I have found the book to be of great assistance for reference, and I take pleasure in recommending it as a text, provided sufficient attention is given to guided supplemental reading to obtain perspective." Reviewing the 4th edition of Snedecor's book in 1946 in *JASA*, David J. Finney wrote "Since its first publication in 1937, this book has been one of the few to combine successfully a sound theoretical basis with an exposition sufficiently clear and detailed for those without statistical expertise."

The depth of the book improved when Snedecor added William Cochran as a co-author beginning with the 1956 5th edition; the final 8th edition appeared in 1989. Cochran had left Rothamsted in 1939 for Iowa State, where he stayed until taking positions after World War 2 at North Carolina, Johns Hopkins, and Harvard. Cochran's experience at Rothamsted and research in experimental designs and surveys made him an excellent complement to Snedecor. An obituary of Cochran in 1982 in *The Annals of Statistics* by Geoff S. Watson quotes Cochran's colleague Alexander Mood (who wrote his own book, discussed in Section 6) at Iowa State as saying "Almost from the day he arrived he was the pre-eminent statistical consultant in the U.S. He was marvelous at it and to my judgment in a class by himself. No one else had the breadth of experience with data from so many fields of statistical investigation; no one else had such universal knowledge of statistical techniques; probably no one else was such a comprehensive reader of statistical journals." With Cochran on board, many new topics were added to Snedecor's book, such as multiple comparisons, the Wilcoxon test, and expectations of mean squares in factorial ANOVA. Regarding the 5th edition, K. Alexander Brownlee stated in *JASA* that it "has been beloved by users of statistics in biology, and also other sciences, for talking to them in language that they can understand, without committing conceptual errors that bring down the wrath and scorn of the theoretical statisticians. … It is clear that *Statistical Methods*, which was already the best text of its kind, has undergone substantial improvement, and it can therefore be highly recommended to its potential readers." In an obituary of Snedecor, Oscar Kempthorne (1974) recalled that "the great majority of workers in noisy sciences loved the Snedecor "cookbook." … they could feel sympathetic to the examples, they could see how statistical methods advanced the quality of the picture, they could see how the same methods could be applied to their own problems, and finally they could *feel* that the process of following the Snedecor cookbook had improved their ability to interact constructively with their data. … it seems indubitable that Snedecor's presentation is the exemplar to be surpassed." Kempthorne pointed out that for some years Snedecor's book, especially in its later version with Cochran, was the most highly cited or one of the most highly cited scientific books and sold an impressive number of copies.

5 The World War 2 period: Wilks, Kendall, Cramér, and mathematical statistics

By the beginning of World War 2, the lack of an established book on statistical theory that presented the substantial Fisher and Neyman–Pearson advances paralleled the lack of recognition of Statistics as a field in British or American universities. In an article advocating serious expansion in the teaching of Statistics, John Wishart (1939) noted that University College was the only place in Britain where it was then possible to study for a degree in Statistics. After stating "We must all at some time or another have been indebted to Mr. Yule for his excellent *Introduction to the Theory of Statistics*, but the inference from calling the book an "Introduction" is that something more is eventually required. A step in the right direction has been taken by Mr. Yule, in association with Mr. Kendall, in revising his classic work." Wishart argued of the need of a new in-depth book on mathematical statistics, which he felt would be a major project and take a significant amount of time to write. He outlined the topics that he thought should be covered: "It should begin with a study of probability theory, a section which may well prove to be the most difficult of all to write. … Thereafter, there is a fairly clear road through distributional theory as an application of probability; the study of distribution functions and their parameters; the theory of estimation (an important section); theories of fitting; tests of goodness of fit; sampling distributions of statistics, leading to tests of significance, … working on by stages through problems in a single variable to those for many variables. … The whole would, of course, be treated mathematically, and since there is plenty of mathematics involved, on the one hand in pure development work in deriving statistics, particularly when it comes to regression and correlation problems in many variables, and analysis of variance problems, … it is clear that a fairly formidable mathematical methodological text would be the result, which would have the effect of putting statistics much more clearly on the map as an important and extensive branch of applied mathematics." In the discussion of Wishart's paper, however, Fisher warned against placing too much emphasis on mathematics at the expense of the original subject areas for the data, and Yule stated "The paper made me feel (rather depressingly) how different your statistical world is from mine. … I would question whether you do not too much ignore the non-mathematician." For economists and psychologists, Yule argued "A good part of the course should deal with the sources of his data, the special risks of misinterpretation, and the special methods used. … I should be inclined to hold that the non-mathematician might be the better teacher for non-mathematicians; he realized their difficulties, which the pure mathematician can hardly do.” Maurice Bartlett (1940) provided his view on the status of mathematical statistics at that time, arguing that mathematical statistics needed to put more emphasis on the theory of probability as its basis. In the discussion of Bartlett’s paper, however, Kendall argued that it was incorrect to identify statistical theory with mathematical statistics, as much of that theory was quite separate from mathematics or even probability.

In North America, Harold Hotelling (1940) likewise stated compelling arguments for why Statistics was well deserving of its own department in the university, pointing out that academia had not been well served by having Statistics taught in various departments by teachers inadequately prepared in statistical theory. His article states “A knowledge of theoretical statistics is not easy to obtain. There is no comprehensive treatise on the subject, starting with first principles, and proceeding by sound deductions and well-chosen definitions to the methods that need to be used in practice. … the seeker after truth regarding statistical theory must make his way through or around an enormous amount of trash and downright error.” That situation was soon to change, due to landmark textbooks by S. S. Wilks, Maurice Kendall, and Harald Cramér.

Samuel Stanley Wilks (1906–1964) received his PhD at the University of Iowa in 1931, with Henry Rietz supervising his doctorate. Following two years in England, where he worked with Karl Pearson and his son Egon and with John Wishart, Wilks received an appointment in mathematics at Princeton University, partly through the recommendation of Hotelling, whom he got to know on a postdoctoral visit to Columbia University. He was also a founding member of the Institute of Mathematical Statistics in 1935, and served as a highly effective editor of *Annals of Mathematical Statistics* from 1938 until 1949 as the journal’s identity changed dramatically to become one of the foremost Statistics journals. His landmark publication on the large-sample chi-squared distribution of the likelihood-ratio statistic appeared in that journal 1938. Wilks first taught Statistics at Princeton in the 1936–37 academic year, in a graduate course in the fall and an undergraduate course in the spring, with his 106 pages of notes made available by Princeton in 1937 in *The Theory of Statistical Inference*, which was very positively reviewed by Rietz in *JASA* in 1937. In 1943 his notes evolved into a lithoprinted publication, *Mathematical Statistics*, three times as long, later published in 1947 by Princeton University Press.

Table 5 shows the table of contents for Wilks’s 1943 textbook. Erich Lehmann (2008) regarded it as “the first modern graduate text in the field. By “modern,” I mean that it centered

Table 5 *Contents of Wilks’s* Mathematical Statistics (1943)

Chapter title	Pages range
1. Introduction	1–4
2. Distribution functions	5–46
3. Some special distributions	47–78
4. Sampling theory	79–97
5. Sampling from a normal population	98–121
6. On the theory of statistical estimation	122–146
7. Tests of statistical hypotheses	147–156
8. Normal regression theory	157–175
9. Applications of normal regression theory to analysis of variance problems	176–199
10. On combinatorial statistical theory	200–225
11. An introduction to multivariate statistical analysis	226–270

on the approach to statistical inference created by the Fisher, Neyman, and Pearson revolution." The outline is quite different from books in statistical theory before then, such as Rietz's, which ignored statistical inference, and not so different from how many instructors would present the topic today. Chapter 1 introduces probability as a long-run relative frequency under "randomness." Chapter 2, including topics such as the Stieltjes integral and moment generating functions, shows the higher mathematical level, compared to Rietz's book. Chapter 3 introduces the binomial, multinomial, Poisson, negative binomial, and normal distributions and Pearson system of distributions. Chapter 4 introduces moments, the Central Limit Theorem, and sampling theory for order statistics such as the median, minimum, and maximum response. Section 5 introduces sampling distributions that occur in sampling from a normal population—chi-squared, t, and F, the latter two under the names "Student t" and "Snedecor's F." The chapter includes the reproductive property of chi-squared, Cochran's theorem (from a 1934 article) about a condition for quadratic forms to be independent chi-squared random variables, and the independence of the sample mean and variance with normal sampling. Chapter 6 introduces confidence intervals and maximum likelihood estimation, with discussion of Fisher's principles of consistency, efficiency, and sufficiency, but without presenting Fisher's fiducial approach to interval estimation. Chapter 7 shows how statistical tests can be generated using confidence intervals, pointing out that "confidence intervals give us a far more complete judgement about the parameter than significance tests." This chapter also introduces likelihood-ratio tests, Wilks's result about its limiting chi-squared distribution, and the Neyman–Pearson theory and its concepts of a critical region and power. The application of regression to analysis of variance problems in Chapter 9 shows Fisher's influence, with discussion of topics such as randomized blocks, Latin squares, incomplete layouts, and the analysis of covariance. Topics in Chapter 10 include the theory of runs, matching theory, and Pearson and likelihood-ratio tests of independence in two-way contingency tables. The final chapter includes the Wishart distribution, Hotelling's multivariate version of the t test and its generalization for testing equality of multivariate means, multivariate regression and analysis of variance, principal component analysis, and canonical correlation theory. Overall, this book has much greater mathematical depth than the Kendall later editions of Yule's book, and Lehmann's praise seems highly appropriate. However, Wilks's book also had a very different style from Yule's, quite dry and relatively sparse in terms of motivation and failing to include real-data examples or exercises.

A review by Abraham Wald in *JASA* in 1943 stated "The study of Mathematical Statistics has been seriously hampered by the lack of good books and by the fact that many important developments of the theory can be found only in the original form and widely scattered in scientific periodicals. The author has performed a very valuable service to all interested in the field by writing the present excellent book. … Readers with some mathematical background will find the book an excellent introduction to the modern developments in the theory of mathematical statistics. Also teachers of mathematical statistics will find it very helpful in their classroom work. The book can be highly recommended to everyone interested in this field." Maurice Bartlett in *JRSS* in 1943 stated "This book will be useful to the professional statistician, also to mathematically-minded students provided that they are not solely reliant on it; it is not recommended to the amateur, nor to the student who is more interested in method than in theory." The book was also favorably reviewed by Jerzy Neyman in *Mathematical Reviews*, who pointed out that the book was unique in being the only one where a student could get information of the theories of testing hypotheses and on estimation. Lehmann (2008) noted "To pull these various contributions together and present them in a unified, coherent account, combined with the necessary mathematical and probabilistic preparation, was a major achievement. The book could well have served as the standard graduate introduction to the field, but it never got its due. This is partly the result of its having been published in wartime, when the profession had different concerns, and partly that Wilks considered it not sufficiently polished and he published it with a soft cover by a photographic process rather than regular print. Immediately after the war, competing books appeared, and its opportunity was gone."

Another book that appeared during the war with a second part shortly thereafter was Maurice Kendall's *The Advanced Theory of Statistics*. Unlike the book by Cramér discussed next, Kendall stated in the preface that it was "a book on statistics, not on statistical mathematics," with intention "to keep the mathematics to heel." Kendall stated "The need for a thorough exposition of the theory of statistics has been repeatedly emphasized in recent years. The object of this book is to develop a systematic treatment of that theory as it exists at the present time." The book was originally planned in 1938 as a joint project with M. S. Bartlett, J. O. Irwin, E. S. Pearson, and J. Wishart. After the outbreak of World War 2, however, Kendall had to proceed alone with the project. It began life as a two volume work (Volume 1, 1943; Volume 2, 1946) of nearly 1000 pages and grew steadily thereafter, as a sole-authored work until the late fifties, when Alan Stuart became involved and it was rewritten in three volumes, and then later continued and extended by other authors.

Table 6 shows the table of contents for volume 1. Kendall decided to discuss descriptive statistics and introduce standard probability distributions before discussing concepts of probability. Chapters 13–16 deal with association and correlation, with regression analysis left for volume 2. The statistician whose works is most commonly cited is R. A. Fisher, somewhat more than Karl Pearson, but primarily for articles dealing with distribution theory. A review in 1945 by Burton H. Camp in *JASA*, after mentioning topics still to come in volume 2, states "This volume includes pretty much everything else one can think of under the heading of mathematical statistics." The review is generally complimentary but with several criticisms, such as complaining about the multiple definitions of probability and the lack of discussion of when each is suitable.

Table 7 shows the table of contents for volume 2, which has five sections, with its greatest emphasis on results due to Fisher and to Neyman. The first section, comprising Chapters 17 to 20, deals with estimation, including a detailed discussion of properties of maximum likelihood such as consistency, sufficiency, and efficiency. The discussion of confidence intervals

Table 6 *Contents of Kendall's* The Advanced Theory of Statistics, *Volume* 1 (1943)

Chapter title	Pages range
1. Frequency distributions	1–28
2. Measures of location and dispersion	29–48
3. Moments and cumulants	49–89
4. Characteristic functions	90–115
5. Standard distributions (1)	116–136
6. Standard distributions (2)	137–163
7. Probability and likelihood	164–185
8. Random sampling	186–203
9. Standard errors	204–230
10. Exact sampling distributions	231–253
11. Approximations to sampling distributions	254–289
12. The χ^2 distribution	290–307
13. Association and contingency	308–323
14. Product moment correlation	324–367
15. Partial and multiple correlation	368–387
16. Rank correlation	388–437

Table 7 *Contents of Kendall's* The Advanced Theory of Statistics, *Volume* 2 (1946)

Chapter title	Pages range
17. Estimation: Likelihood	1–49
18. Estimation: Miscellaneous methods	50–61
19. Confidence intervals	62–84
20. Fiducial inference	85–95
21. Some common tests of significance	96–140
22. Regression	141–174
23. The analysis of variance (1)	175–217
24. The analysis of variance (2)	218–246
25. The design of sampling inquiries	247–268
26. General theory of significance tests (1)	269–306
27. General theory of significance tests (2)	307–327
28. Multivariate analysis	328–362
29. Time series (1)	363–395
30. Time series (2)	396–439

focuses on their long-run justification, the conservative performance with small samples from a discrete distribution, the use of information from the log-likelihood function for large samples, and generalizations for multiple parameters, illustrating with "studentisation" when the true standard deviation is unknown in sampling from a normal distribution. This section also includes a short chapter on Fisher's fiducial approach. Kendall notes that if the reader does not accept Fisher's argument for a new postulate of inference, he can "console himself with the thought that the results of the process are the same as those given by the theory of confidence intervals" and notes that Jeffreys's Bayesian approach gives the same results in many cases. He concludes by discussing the awkward Behrens–Fisher problem of comparing means from normal distributions having unequal variances, noting Fisher's criticism of the confidence interval method because of its being unable to handle that important case. The second section, comprising Chapters 21, 23, 24 and 26 to 28, covers the theory of statistical tests, including the analysis of variance and multivariate analysis. Its first chapter on significance tests presents detailed properties of the distribution of the t statistic, including cases of sampling from non-normal distributions, and of the F distribution (called "Fisher's z distribution"). The chapter presents a variety of tests, including goodness-of-fit tests, Fisher's combination of tests, tests of homogeneity, tests for correlations, and conditional tests (see, e.g., conditioning on a sufficient statistic to compare two Poisson means). The later two chapters on the general theory of significance tests present the Neyman–Pearson approach, with discussion of the concept of unbiased tests and likelihood-ratio tests, but mentioning Wilks's result about their asymptotic chi-squared distribution only for the case (in a 1935 paper by Wilks) of testing independence in a contingency table. The chapter on multivariate analysis presents the Wishart distribution, Hotelling's generalizations of the t test, Fisher's linear discriminant analysis, and canonical correlation analysis. The third section, consisting of Chapter 22, deals with regression analysis and completes the account of statistical relationship begun in Chapters 13 to 16 of Volume I. It includes discussion of the bivariate normal, fitting curvilinear relationships, orthogonal polynomials, and inference for regression coefficients. The fourth section, Chapter 25, gives an introductory account of theoretical considerations on sampling methods, including stratified sampling, and the design of experiments, including the influence of Fisher for concepts of orthogonality, randomisation, balance, and confounding and a presentation of the analysis of variance for Latin squares. Finally, the fifth section, comprising Chapters 29 and 30, deals with the analysis of time series, including moving averages and autoregressive series, trend-elimination, analyzing differences of successive values, tests for randomness, and analyzing serial correlation. Each volume has the style of numbered paragraphs within each each chapter, and each chapter concludes with notes and bibliographic references and numerous exercises. Most of the exercises are highly technical, often presenting recent research results. The bibliography itself in this massive two-volume work is 62 pages long.

In *Annals of Mathematical Statistics* in 1947, Maurice Bartlett was critical of the treatment of probability and of various aspects of the organization. For instance, with Kendall's inclusion of separate chapters for fiducial probability and the theory of confidence intervals, Bartlett questioned "whether strict impartiality is the best answer … with a field which is still a battleground." In the preface to Volume 2, Kendall had stated that in presenting the logic of statistical inference and discussing both confidence intervals and fiducial inference, "It is, I consider, unfair in a book of this kind not to present all sides of a case, particularly when there is so much disagreement among the authorities. Some day I hope to show that the disagreement is more apparent than real." Following two pages of critical comments and corrections and mention of results that could be added, Bartlett finished the review by stating "These criticisms … are not intended to detract seriously from what is a remarkable achievement. An excellent sense of proportion has been maintained throughout between mathematical theory and illustrative discussion and examples. This makes this treatise, if both the breadth and level of the subject matter are taken into account, at present unique. It will be an indispensable reference book to every teacher and advanced student of the theory of statistics." Browsing through these two volumes today, one cannot help but be greatly impressed by the breadth and depth of Kendall's treatment of the foundational results introduced in the previous quarter century, and done so during a highly stressful period when he was assistant general manager to the British Chamber of Shipping and had air-raid warden duties at night.

Other than Kendall's volumes and the 1947 version of Wilks's 1943 text on *Mathematical Statistics*, the landmark publication in statistical theory immediately following the war was Harald Cramér's 575-page-long 1946 text, *Mathematical Methods of Statistics*. Like Wilks's book, it was published by Princeton University Press. Harald Cramér (1893–1985) was Professor of Actuarial Mathematics and Mathematical Statistics and director of the Institute of Mathematical Statistics at the University of Stockholm and President of that university from 1950 to 1961, and the book was based on his Stockholm lectures since 1930. In the preface, Cramér states "During the last 25 years, statistical science has made great progress, thanks to the brilliant schools of British and American statisticians, among whom the name of Professor R. A. Fisher should be mentioned in the foremost place. During the same time, largely owing to the work of French and Russian mathematicians, the classical calculus of probability has developed into a purely mathematical theory satisfying modern standards with respect to rigour. The purpose of the present work is to join these two lines of development in an exposition of the mathematical theory of modern statistical methods, in so far as these are based on the concept of probability. A full understanding of the theory of these methods requires a fairly advanced knowledge of pure mathematics." Compared to existing books on statistical theory, the mathematical level is higher, with much more emphasis on probability. (Approaches for defining probability were not discussed by Wilks and were considered in only a few pages of Chapter 7 in Kendall's first volume.) Cramér's book has three parts: Mathematical Introduction, Random Variables and Probability Distributions (based on a separate tract published in 1937), and Statistical Inference. Table 8 outlines the contents.

The 12 chapters in the first part cover set theory, Lebesgue measure and integration, the Lebesgue–Stieltjes integral, Fourier integrals, and matrices and quadratic forms. The second part introduces probability as a long-run relative frequency, giving special credit to publications by Ludwig von Mises in 1931, 1936, and 1941 for building an axiomatic approach with

Table 8 *Contents of Cramér's* Mathematical Methods of Statistics (1946)

Chapter group	Pages range
Part 1: Mathematical introduction	
1–3. Sets of points	3–18
4–7. Theory of measure and integration in R_1	19–75
8–9. Theory of measure and integration in R_n	76–88
10–12. Various questions	89–136
Part 2: Random variables and probability distributions	
13–14. Foundations	137–165
15–20. Variables and distributions in R_1	166–259
21–24. Variables and distributions in R_n	260–322
Part 3: Statistical inference	
25–26. Generalities	323–340
27–29. Sampling distributions	341–415
30–31. Tests of significance, I	416–472
32–34. Theory of estimation	473–524
35–37. Tests of significance, II	525–556

this foundation but also mentioning subjective approaches of Jeffreys and Keynes and then also stating Kolmogorov's axiomatic approach in terms of random variables. He then covers the standard distributions, those that are based on the normal distribution (the chi-squared, t, F and beta), as well as others such as the Cauchy, Laplace, Pareto, and Pearson types. He presents convergence theorems, such as versions of the law of large numbers. He introduces the characteristic function and uses it to prove the Lindeberg/Lévy version of the Central Limit Theorem (from articles published in 1922 and 1935), and he discusses various generalizations such as what is now called the delta method. The third part introduces simple random sampling and sampling distributions and their moments and quantiles, with particular cases including the joint distribution of the sample mean and sample variance in samples from a normal distribution, the regression coefficient, and the correlation, partial correlation, and multiple correlation. The discussion of tests of significance first mainly focuses on Pearson's chi-squared tests of goodness of fit and related tests such as independence and homogeneity in contingency tables, and then in later chapters on more complex problems such as analysis of variance (including randomized blocks and Latin squares) and regression. The discussion emphasizes the Neyman–Pearson approach for tests and Fisher for the theory of estimation and in particular maximum likelihood estimation and its asymptotic properties and concepts of efficiency and sufficiency. Cramér also introduces confidence regions, focusing on Neyman's approach and a Bayesian approach rather than Fisher's fiducial method.

Each group of chapters finishes with references and a historical background. Two sets of chapters in Part 2 on probability distributions have exercises, but dealing with extensions that would be quite difficult for most students. Like the Kendall volumes, the book is incredibly up-to-date in terms of presenting the great Fisher and Neyman–Pearson theoretical advances on the previous quarter century, with careful explanation of hypothesis testing, point estimation, and confidence sets. A rather surprising omission was the lack of discussion of Wilks's general likelihood-ratio test result. Cramér does cite Wilks's 1937 notes as well as the 1940 12th edition of Yule and Kendall and Kendall's 1943 *Advanced Theory of Statistics*. Among other innovations, the book introduced what became known as the Cramér–Rao lower bound for the variance of an estimator and proved that the likelihood equation has a consistent solution.

The book received highly favorable reviews, particularly for its mathematical sophistication compared with other textbooks. Paul Hoel and Jacob Wolfowitz in *JASA* stated "This excellent book will be welcomed by mathematicians who wish to become acquainted with statistical methods and, even more so, by statisticians who desire a reference book for proofs of many of the fundamental theorems of statistical theory." William Feller, in *Annals of Mathematical Statistics*, stated that the book "fills an urgent need." Henry Scheffé, in *Bulletin of the American Mathematical Society*, stated "This book has long been needed, and its effect on the future development of mathematical statistics in both teaching and research will be sharp and lasting. The achievement of the author is to present the first treatise on statistics in which the mathematical developments are carried through according to standards of rigor comparable to those now customary in pure mathematics"; Harold Hotelling, in *Journal of Political Economy*, stated "This book meets an important and long-felt need for a really basic treatment of modern statistical theory." Lehmann (2008) recalled that the book "quickly established itself as the standard introduction to the theory of statistical inference developed by Fisher and Neyman–Pearson. Two of its outstanding features were mathematical rigor and readability."

Shortly after their appearance, the economist Paul Samuelson (1950) reviewed these and other recent textbooks in mathematical statistics. He reviewed Cramér especially positively, calling it "the long waited for definitive intermediate treatise. . . . Probably this is the book that will be most frequently cited in the technical literature and which the theoretical statistician will find himself taking off the shelf most often."

A much shorter book (153 pages) but one also that places strong emphasis on mathematics had been published in 1939 by Alexander C. Aitken (1895–1967), a New Zealand native who spent his academic career in Edinburgh, Scotland (Aitken 1939). The main purpose of *Statistical Mathematics* was to describe aspects of mathematics directly applicable to statistical theory and practice. Aitken had introduced the concept of generalized least squares in a 1935 paper, along with now standard vector/matrix notation for the linear regression model. This may have been the first book in which a systematic development of the theorems of mathematical statistics was made by means of moment generating functions. The main topics covered were an axiomatic approach to probability, probability distributions, least squares and curve fitting, and sampling distributions of statistical coefficients. The book continued through a 8th edition in 1957. Also noteworthy is a short (50-page) 1942 book, *On the Principles of Statistical Inference* by Abraham Wald, based on four lectures at Notre Dame University, that gave an overview of the Neyman–Pearson theory of testing hypotheses, Fisher's approach to estimation, and the theory of confidence intervals. A book by Wolfenden (1942) had "mathematical statistics" in its title but was geared toward an introductory course for actuarial students.

6 The quarter century after World War 2: Explosion in mathematical statistics

At the end of World War 2, as books on the theory of Statistics got some visibility among probabilists and statisticians and universities began to add courses on the topic, few such courses were in Statistics departments. Noether (1989) observed, "At many institutions, the center of statistics instruction had moved from departments of economics to departments of mathematics. As concerned statisticians began to realize, the move often had the unfortunate consequence that mathematicians entrusted with the teaching of introductory statistics courses preferred to concentrate on mathematical, rather than statistical ideas, resulting in the increasingly held belief that statistics was simply a branch of mathematics." Hotelling (1940) discussed how difficult being an effective teacher of Statistics is for a mathematician without training in Statistics and a sympathy and understanding for applications or for a subject-matter specialist such as an economist who uses Statistics but does not have theoretical training in it. In 1945 Jerzy Neyman (1894–1981), who had moved from University

College, London to the University of California at Berkeley in 1938, organized a symposium to contribute to the revival of scientific work in mathematical statistics following World War 2. The symposium included contributions from Hotelling, Wolfowitz, Doob, Feller, and Lehmann, with Hotelling's keynote address "The place of statistics in the university." The *Berkeley Symposium on Mathematical Statistics and Probability* (Neyman 1949) was the first of six, held roughly every five years. Neyman had founded a Statistical Laboratory at Berkeley, which led to the creation of a Department of Statistics in 1955. Soon thereafter, particularly with the explosion in growth of state universities in the 1960s with the baby boomer generation, Statistics departments in the U.S. became much more common (Agresti and Meng 2013).

As Statistics as a field received more attention in the form of university departments, so did the textbooks by Wilks, Kendall, and Cramér. Their importance over the years is reflected by citation counts in Google Scholar, for instance more than 16,000 for Cramér and more than 20,000 for the Kendall volumes. We next consider briefly some successors to the texts by Wilks, Kendall, and Cramér in the following quarter century, focusing on the texts that seem to have been most influential during a period of explosion of books about mathematical statistics. Some of these books had lower technical level, such as books by Hoel (1947), Weatherburn (1946), Wilks (1948), and Mood (1950). The Hoel and Mood books seem to have received the most attention, and we focus here on Alexander Mood's *Introduction to the Theory of Statistics*, which contained more substance on statistical inference than Hoel's book.

Alexander Mood (1913–2009) studied at Princeton under Sam Wilks (a fellow Texan native), receiving a PhD in 1940. After World War 2, he taught at Iowa State, recruited by William Cochran. Working with Cochran, he became interested in experimental design and agricultural statistics, initiated formation of the Statistics department, and wrote his textbook. In the book's preface, he mentions that the book developed from notes prepared in 1945 when no suitable texts were yet available. He states "Since then the situation has been relieved considerably, and had I known in advance what books were in the making it is likely that I should not have embarked on this volume. However, it seemed sufficiently different from other presentations to give prospective teachers and students a useful alternative choice." He acknowledges Wilks for having kindled his interest in Statistics and serving as his mentor and giving careful lectures that helped Mood himself to understand the subject. Although a book in theory, many examples and exercises were motivated by his working in the Statistics lab at Iowa State. By the time of its publication, he worked at the Rand Corporation. He then formed a company (General Analysis Corporation), served as President of the Institute of Mathematical Statistics in 1957, and eventually returned to academia at the University of California, Irvine. Table 9 shows the table of contents of Mood's 433 page textbook. The contents would not seem unusual in a book published today, and a first-year graduate student or advanced undergraduate student in Statistics would probably find this book much more readable than the earlier ones by Wilks, Kendall, and Cramér.

Mood begins by defining Statistics as "the technology of the scientific method." While calling the theory of statistics a branch of applied mathematics, he distinguishes this from the practice of statistics: "The use of statistical tools is not merely a matter of picking out the wrench that fits the bolt; it is more a matter of selecting the correct one of several wrenches whih appear to fit the bolt about equally well but none of which fit it exactly." Any statistical formula "is merely a tool, and moreover a tool derived from some simple mathematical model which cannot possibly represent the actual situation with any great precision. In using the tool one must make a whole series of judgments relative to the nature and magnitude of the various errors engendered by the discrepancies between the model and the actual experiment." He points out that therefore an applied statistician needs to be completely familiar with both

Table 9 *Contents of Mood's* Introduction to the Theory of Statistics (1950)

Chapter title	Pages range
1. Introduction	1–7
2. Probability and combinatorial methods	8–43
3. Discrete distributions	44–64
4. Distributions for continuous variates	65–90
5. Expected values and moments	91–106
6. Special continuous distributions	107–123
7. Sampling	124–146
8. Point estimation	147–164
9. The multivariate normal distribution	165–191
10. Sampling distributions	192–219
11. Interval estimation	220–244
12. Tests of hypotheses	245–288
13. Regression and linear hypotheses	289–315
14. Experimental designs and the analysis of variance	316–364
15. Sequential tests of hypotheses	365–384
16. Distribution-free methods	385–418

the theory and methodology of Statistics. In a later chapter, he gives a historical context to the development of methods of statistical inference, calling Karl Pearson "the first major contributor to the theory of statistics" and "the founder of the science of statistical inference," crediting Gosset for his contribution toward inference for small samples, credits Fisher as "the real giant in development of the theory of statistics," stating "this one man must be credited with at least half the essential and important developments as the theory now stands," and finally credits Neyman and E. S. Pearson for the general theory of testing hypotheses, including the crucial ingredient of the power of a test. The final two chapters contain material just recently developed—sequential tests and distribution-free methods. In the preface, Mood points out the importance of the book's 500 exercises for numerical examples, illustrating the theory, and introducing additional material. The exercises have quite a different style and would be much less difficult for most students than the exercises in the books by Kendall and Cramér.

Mood's book received several complimentary book reviews, including one by Jacob Wolfowitz in 1951 in *American Mathematical Monthly* that complained that although mathematical statistics had made great strides in the last twenty-five years, nearly all existing books in Statistics were merely "cookbooks." Although he found things to complain about in Mood's book (such as the meager treatment of a test of hypothesis, which seems surprising given the long chapter that included a variety of contexts including composite hypotheses and explanation of likelihood-ratio tests), he called it "an excellent introduction to statistics." When Mood did not have time to revise, McGraw Hill Publishing found a co-author (Franklin Graybill) for a 2nd edition in 1963. The 3rd edition added Duane Boes in 1974. The Hoel (1947) book was also successful, at a less technical level, continuing through a 5th edition published in 1984.

Beginning in 1958, Maurice Kendall took on Alan Stuart as a co-author for his book, expanding it from two volumes to three volumes. *Volume 1: Distribution Theory* in 1958 dealt with standard univariate distributions and distributions associated with the normal, the multivariate normal and quadratic forms, moments and characteristic functions, and sampling distributions and standard errors. *Volume 2: Inference and Relationship* in 1961 dealt with the theory of estimation and testing hypotheses, statistical relationship, distribution-free methods, and sequential analysis. *Volume 3: Design and Analysis, and Time Series* in 1961 dealt with the design and analysis of sample surveys and experiments, including analysis of variance and the theory of multivariate analysis and time series. Kendall was knighted in 1974 and passed away in 1983, and Stuart passed away in 1998. The most recent 6th edition of Volume 1 in 2010 (with title *Kendall's Advanced Theory of Statistics: Distribution Theory*) has J. Keith Ord as co-author and Volume 2 in 2015 (with title *Kendall's Advanced Theory of Statistics: Classical Inference and the Linear Model*) has Ord and Steven Arnold as co-authors. The 4th and final edition of Volume 3 appeared in 1982 with Ord as coauthor. Volume 2B on *Kendall's Advanced Theory of Statistics: Bayesian Inference* by Anthony O'Hagan was added in 1994, with 2nd edition in 2004 including Jonathan Forster as coauthor.

In 1959, a book was published that probably had more use than any mentioned yet on the teaching of mathematical statistics in the next two decades: *Introduction to Mathematical Statistics* by Robert Hogg (1924–2014) and Allen Craig (1905–1978). Craig, a Floridian, received the first masters degree in mathematics at the University of Florida in 1928 and then went to the University of Iowa at the same time as Sam Wilks, attracted by the presence of Henry Rietz. Hogg received his PhD at Iowa with Craig as advisor in 1950 and was founding chair of the Statistics Department in 1965 and served in that capacity for 19 years. The Hogg and Craig textbook differed from other books in thoroughly treating the change-of-variable method of deriving the distribution of a function of several random variables and also included considerable discussion of the order statistics, such as showing them to be sufficient statistics in the nonparametric case. In an interview (Randles 2007), Hogg stated that his name came before Craig's in the author list because Craig wanted to promote the younger Hogg and because Craig did not expect to work on later editions, although he did in fact contribute a bit to the next two. Hogg also mentioned that their book evolved out of a desire to do a monograph about sufficient statistics and dissatisfaction with the way existing books presented distribution theory. In fact, a book review by Benoit Mandelbrot in *Information & Computation* in 1960 praised the chapter on sufficient statistics and discussed its key role in the exponential family of distributions.

Hogg felt that the 1959 textbook served mainly as a preliminary edition, and the 1965 second edition expanded from 245 to 383 pages, with addition of material on decision theory, Bayesian inference, sequential analysis, and a chapter on the analysis of variance. Table 10 shows the table of contents of that edition. A 1966 review in *JASA* by Martin Fox states "The transition between editions is akin to that between lecture notes and a polished book. … Generally motivation is now much better. The probability chapters, inadequate in the first edition,

Table 10 *Contents of Hogg and Craig second edition of* Introduction to Mathematical Statistics (1965)

Chapter title	Pages range
1. Distributions of random variables	1–49
2. Conditional probability and stochastic independence	50–77
3. Some special distributions	78–106
4. Distributions of functions of random variables	107–150
5. Interval estimation	151–167
6. Order statistics	168–185
7. Limiting distributions	186–203
8. Sufficient statistics	204–236
9. Point estimation	237–253
10. Statistical hypotheses	254–284
11. Other statistical tests	285–307
12. The analysis of variance	308–342
13. Further normal distribution theory	343–365

are now quite complete for a book at this level." Perhaps surprising is regression having only 4 pages, with no mention of regression to the mean or Francis Galton or Karl Pearson and no mention of Fisher except for attributing the factorization criterion for a sufficient statistic to him and Neyman. The book included a very large number of exercises, nearly all dealing with theory rather than data analysis, and like the exercises in Mood's book, easier for the average student than exercises in the earlier books on statistical theory. This textbook is still used, now in its 8th edition published in 2018 with Joseph McKean as co-author. Shortly after 1959, other books were published with similar content as Hogg and Craig but slightly lower technical level, such as by Brunk (1960), Freund (1962), and Lindgren (1962).

Other books soon followed that also got attention from statisticians wanting to learn about statistical theory. In 1962, S. S. Wilks published a much expanded (644 page) version of his 1943 book that now included more than 400 exercises. However, a review by David Cox in 1962 in *Annals of Mathematical Statistics* commented that Wilks failed to examine the main concepts of the subject and connections between confidence intervals and tests. He stated "The book would, I think, have been much strengthened by the inclusion both of more extended motivation of some of the topics and also of critical discussion of such ideas as sufficiency, unbiasedness, and so forth." Cox also noted that the exponential family appears only briefly in exercises and that Bayes's theorem is not mentioned at all. He concludes that although the Kendall and Stuart volumes would probably be preferred by working statisticians looking for an introduction to a particular advanced topic, Wilks's book had a choice of topics more in line with conventional interests and was likely to be preferred as a textbook for advanced students. The same issue had a long review of Wilks's book by Wassily Hoeffding. (In an email, David Cox told me that the editor was unhappy with Cox's review.) However, Hoeffding had the same criticisms about the presentation of topics and pointed out many other insufficiencies as well as many errors. His review states "I think it would have been better if more attention had been given to those developments which have yielded fairly general and systematic methods for constructing statistical procedures with desirable properties and which bring out connections between seemingly unrelated topics. ... These remarks apply especially to the treatment of estimation and hypothesis testing." (He refers to the books by Lehmann (1959) and Fraser (1957) as more narrowly-focused ones that accomplished what Wilks failed to do in terms of showing how to construct good statistical procedures.) Wilks's student Alexander Mood reminisced (Mood 1990) "Sam was a true mathematician in that he always strove for elegance in proofs and was always most careful about details—a perfectionist. One unfortunate result was that his beautiful book (Wilks, 1962) on mathematical statistics was published about 20 years too late. A reasonably complete version was ready in 1942 and issued in lithographed form in 1943 for limited distribution. Sam kept tinkering with it year after year; then Harald Cramér published his book (Cramér, 1946) which covered much the same ground. Sam decided to make his much more comprehensive and tinkered with it another 15 years. Cramér's book enjoyed great prestige—something Sam's book would have had if he had been a little less concerned about impeccability."

Soon after, the 3rd edition of *Probability Theory and Mathematical Statistics* by the Polish probabilist and statistician Marek Fisz (1910–1963) was translated from Polish to English by Robert Bartoszynski and published shortly after Fisz's death in 1963. Table 11 shows the table of contents of this book, which devoted half of its 677 pages to probability. In the preface, Fisz indicates that besides providing a systematic introduction to modern probability theory and mathematical statistics and showing concrete examples of application of the theories, his goal was to provide extensive references, with indications of their contents. This was reflected by a 37 page bibliography, with many articles from the Russian and Polish literature, and by many of the exercises providing recent research results in complements. The mathematical level was quite a bit higher than the books by Mood and by Hogg and Craig, with

Table 11 *Contents of Fisz's* Probability Theory and Mathematical Statistics (1963)

much more emphasis on probability and stochastic processes. Fisz covered many classic results of K. Pearson, Fisher and Neyman–Pearson, yet some key results and concepts appeared only in complements, such as the likelihood-ratio test, or not at all, such as regression toward the mean and the influence of Francis Galton. A 1964 review by Seymour Geisser in *Technometrics* praised the probability half of the book but criticized the mathematical statistics part as having limited value "because of its preoccupation with order statistics, runs, goodness of fit tests and other non-parametric niceties so that analysis of variance, covariance, regression, correlation, and experimental design suffer by comparison." A review in *JASA* by I. Richard Savage thought the first part comparable to Cramér's treatment of probability, but with Cramér more self-contained in terms of needed mathematics and Fisz having a more extensive treatment of probability theory. Reviews noted that examples were somewhat artificial and exercises often contained important theoretical results but would be very difficult for the typical student to answer.

Another advanced book published soon after Wilks's and Fisz's was Calyampudi Radhakrishna (C. R.) Rao's (1965) *Linear Statistical Inference and Its Applications*, which evolved from Rao's 1952 book *Advanced Statistical Methods in Biometric Research*, one of the first books to discuss the application of multivariate statistical methodology for scientific research problems. Table 12 shows the table of contents. Rao was influenced by his PhD advisor Fisher as well as his friend P. C. Mahalanobis in stating in the preface that "Statistical inference techniques, if not applied to the real world, will lose their import and appear to be deductive exercises. Furthermore, it is my belief that in a statistical course emphasis should be given to both mathematical theory of statistics and to the application of the theory to practical problems. A detailed discussion on the application of a statistical technique facilitates better understanding of the theory behind the technique." The treatment of statistical inference in Chapters 4–7 had emphases on the Fisher and Neyman–Pearson schools as in the books twenty years earlier by Wilks, Kendall, and Cramér. A special feature of the final chapter on multivariate analysis was that instead of deriving features of the multivariate normal distribution through its probability density function, Rao used the simple characterization that every linear function of the variables has a univariate normal distribution. The mathematical level of Rao's textbook was considerably higher than those by Mood and by Hogg and Craig,

Table 12 *Contents of Rao's* Linear Statistical Inference and Its Applications (1965)

Chapter title	Pages range
1. Algebra of vectors and matrices	1–78
2. Probability theory, tools and techniques	79–154
3. Continuous probability models	155–219
4. The theory of least squares and analysis of variance	220–313
5. Criteria and methods of estimation	314–381
6. Large sample theory and methods	382–443
7. Theory of statistical inference	444–515
8. Multivariate analysis	516–604

beginning with a chapter on an advanced treatment of vector spaces and matrices and a second chapter on probability that included a measure-theoretic exposition. The textbook did not lend itself as well to being a course textbook as those books, because of the technical level, very terse style, and quite difficult exercises. An overall favorable 1966 review by William Cochran (in *J. Franklin Institute*) summarized "As a text, the compactness of the writing and the great variety of topics demands A grade students, despite the ample supply of exercises." A 1966 review in *Technometrics* by Norman L. Johnson criticized organizational aspects such as a poor index and lack of consistency in form of construction but concluded "The amount of detailed information succinctly presented is indeed the outstanding impression from reading this book. A close second is the sensible discussion of applications of theoretically developed methods." A 1966 review in *Biometrika* by John Aitchison had several criticisms, such as Rao's presentation of least squares material in an algebraic rather than geometric manner (using projections) and his attempting to cover too much, yet concluded "It is the duty of every self-respecting mathematical statistician to possess a well-thumbed copy of this book."

In the quarter century following World War 2, highly influential books were published about particular aspects of statistical theory and methods. Perhaps the most influential was Erich Lehmann's (1959) *Testing Statistical Hypotheses*, from which statisticians became more familiar with the Neyman–Pearson theory and its extensions and ramifications. Other influential books on various topics included Abraham Wald's (1947) *Sequential Analysis* and 1949 monograph *Statistical Decision Functions*, Frank Yates's (1949) *Sampling Methods for Censuses and Surveys*, the William Cochran and Gertrude Cox (1950) *Experimental Designs*, William Cochran's (1953) *Sampling Techniques*, Donald Fraser's (1957) *Nonparametric Methods in Statistics*, Theodore Anderson's (1958) *An Introduction to Multivariate Statistical Analysis*, Henry Scheffé's (1967) *The Analysis of Variance*, and Norman Draper and Harry Smith's (1967) *Applied Regression Analysis*. During this period, robust statistics and exploratory data analysis started to become popular, although John Tukey's seminal book *Exploratory Data Analysis* that generated a significant broadening of the field of Statistics did not appear until 1977. Two other influential books that appeared soon after this period were *Discrete Multivariate Analysis* by Yvonne Bishop, Stephen Fienberg, and Paul Holland (1975) and *Generalized Linear Models* by Peter McCullagh and John Nelder (1983). As follow-ups to the Fisher and Snedecor textbooks on statistical methods, during this period numerous books appeared that presented the methods with orientation toward particular areas of application. Popular examples were Hubert Blalock's (1960) *Social Statistics*, K. Alexander Brownlee's (1960) *Statistical Theory and Methodology in Science and Engineering*, C. R. Rao's (1952) *Advanced Statistical Methods in Biometric Research*, and 12 editions between 1937 and 1991 of Austin Bradford Hill's *Principles of Medical Statistics*.

Table 13 *Contents of Jeffreys's 2nd edition of* Theory of Probability (1948)

Chapter title	Pages range
1. Fundamental notions	1–46
2. Direct probabilities	47–98
3. Estimation problems	99–167
4. Approximate methods and simplifications	168–219
5. Significance tests: one new parameter	220–304
6. Significance tests: various complications	305–340
7. Frequency definitions and direct methods	341–371
8. General questions	372–395

7 Foundations of statistical science from a Bayesian perspective

Many statisticians regard the 1939 book by Cambridge professor Harold Jeffreys (1891–1989) on *Theory of Probability* as the foundational textbook on the Bayesian statistical approach. Jeffreys had already achieved scientific fame with his 1924 book *The Earth, Its Origin, History and Physical Constitution* that some regard as founding modern geophysics. He began the 1939 book's preface by stating "The chief object of this work is to provide a method of drawing inferences from observational data that will be self-consistent and can also be used in practice." He points out that the field had grown without much attention to logical foundations, and "my objection to current statistical theory is not so much to the way it is used as to the fact that it limits its scope at the outset in such a way that it cannot state the questions asked, or the answers to them, within the language that it provides for itself." Table 13 shows the table of contents of the 1948 second edition. Robert et al. (2009) give an informative chapter-by-chapter review of this book and its influence. They point out that the title is misleading, as the book does discuss estimation and significance tests and does not present mathematical bases of probability but rather the theory of *inverse probability* (the Jeffreys name for the Bayesian approach). The chapter on estimation introduces exponential families and the principle of Jeffreys noninformative priors. Jeffreys's book helped to initiate the objective Bayesian school of Statistics, stating the principle for deriving noninformative prior distributions from the sampled distribution, using Fisher information. The following chapter introduces Fisher's maximum likelihood estimate, but justifies it as being indistinguishable for large samples from an estimate using inverse probability with a uniform prior distribution. Not surprisingly, his treatment of significance tests makes much use of Bayes factors, which he had introduced in an article four years before the 1939 first edition. The book concludes with two chapters on the weaknesses of other theories and on some general scientific questions.

In a 1940 book review in *Nature*, Kendall discusses the controversial aspects of the personalitic view of probability but concludes that "Dr. Jeffreys' views deserve the most serious study. They may not be right; but they are more tenable than many statisticians have been willing to concede in the past, and it is hoped that this book will be widely read." In an obituary of Jeffreys in 1989 in *JRSS A*, Dennis Lindley wrote that Jeffreys, who actually achieved more fame as a geophysicist than as a statistician, was one of the finest writers of scientific English and "uses probability as the language of scientific method. … The demonstration of the central role that probability must play in all our affairs is one of the greatest achievements of the 20th century." Not all reviews were positive. In a 1941 review in *Biometrika*, S. S. Wilks stated that it is "doubtful that there will be many scholars thoroughly familiar with the system of statistical thought initiated by R. A. Fisher and extended by J. Neyman, E. S.

Pearson, A. Wald, and others who will abandon this system in favour of the one proposed by Jeffreys in which inverse probability plays the central role." In *JASA* in 1940, Jerzy Neyman wrote "We are told that the theories of probability developed by others are not satisfactory and that if modern statistical work does bring valuable results, it is only because the statisticians forget their own definitions and, without noticing it, follow the lines of thought of Dr. Jeffreys. ... The author's criticism of theories of probability, based on conceptions of measure and of frequency, is somewhat bitter and his description of these theories is hardly adequate." Neyman indicated puzzlement at Jeffreys's belief that in each case just one prior distribution is appropriate and in his choice of the improper prior proportional to $1/x$ when X must be positive. He concludes, "As this book is revolutionary, it is probable that readers will disagree with it, but all of them are likely to be very interested." R. A. Fisher respected Jeffreys but had already had arguments with him for several years before publication of his book (see Aldrich 2005), and Joan Fisher Box (1978, p. 441) quotes him as saying that Jeffreys "makes a logical mistake on the first page which invalidates all the 395 formulae in his book." Nonetheless, Jeffreys's book has since received about 15,000 citations according to Google Scholar. A later highly-influential book for the exposition of probability itself from a Bayesian perspective was Bruno De Finetti's 1970 *Teoria delle Probabilità* (translated into English in 1974 and 1975), based largely on a 1937 article of his. De Finetti's writings, themselves much influenced by Jeffreys, also emphasized the central role of exchangeability.

Textbooks presenting an overview of the theory of Statistics from a Bayesian perspective followed well after those with a frequentist perspective. Much influenced by Jeffreys, Irving John (I. J.) Good's (1950) *Probability and the Weighing of Evidence* was a short (119-page) book that moved in the direction of a Bayesian Statistics textbook. The final and longest chapter of Good's book discussed relevance to Statistics, for topics such as curve fitting, significance tests, and estimation, but with his own rather idiosyncratic ways of amending standard methods. University of Chicago professor Leonard Jimmy Savage (1917–1971), whose PhD students included the eminent Bayesian statisticians Morris DeGroot and Don Berry, published *The Foundations of Statistics* in 1954. Starting with six axioms and using rigorous proofs, Savage put forth a theory of subjective and personal probability and developed the theory of utility and optimal decision making. He warns the reader in the preface that "it cannot be a textbook, or manual of instruction, stating the accepted facts about its subject, for there scarcely are any." Table 14 shows the table of contents of the 2nd edition in 1972, nearly identical to the first but with additional references and footnotes and editorial comments in its preface. Its contents are dramatically different from contents of books discussed so far in this article. The first part, ending with Chapter 7, is an introduction to the personalistic (and thus Bayesian) tradition in probability and introducing the expected utility theory of von Neumann and Morgenstern. The second part provides a personalistic discussion of frequentist methods. In Chapter 8, he defines *statistics proper* as "the art of dealing with vagueness and with interpersonal difference in decision situations." Savage discusses conditions under which minimax rules are reasonable and how a group might use them to reach a compromise decision. The final three chapters on statistical inference, highly condensed into 43 pages, are more concerned with commenting on existing approaches rather than expounding positive approaches taking the Bayesian approach. The chapter on point estimation introduces maximum likelihood and its good properties and credits Fisher. The chapter on testing describes the Neyman–Pearson approach. In the preface of the 2nd edition, he states respect for that approach in terms of leading to admissibility as a criterion, though pointing out how that criterion again leads to personal probabilities as central. Savage also included a chapter on interval estimation but argues against its use, pointing out that users of the method "endlessly pass the buck, saying in effect, 'This assertion has arisen according to a system that will seldom lead you to make false assertions, if you adopt it. As for myself, I assert nothing but the properties of the system.'"

Table 14 *Contents of Savage's 2nd edition of* The Foundations of Statistics (1972)

Chapter title	Pages range
1. Introduction	1–5
2. Preliminary considerations in the face of uncertainty	6–26
3. Personal probability	27–55
4. Critical comments on personal probability	56–68
5. Utility	69–119
6. Observation	105–119
7. Partition problems	120–153
8. Statistics proper	154–157
9. Introduction to the minimax theory	158–171
10. A personalistic representation of the minimax theory	172–177
11. The parallelism between the minimax theory and the theory of two-person games	178–183
12. The mathematics of minimax problems	184–199
13. Objections to the minimax rules	200–207
14. The minimax theory applied to observations	208–219
15. Point estimation	220–245
16. Testing	246–256
17. Interval estimation and related topics	257–262

The table of contents of Savage's book doesn't convey the author's strong viewpoints. For example, in the preface of the second edition, Savage states that there is no justification for frequentist methods. He states "The theory of personalistic probability applied to statistics shows that many of the prominent frequentist devices can at best lead to accidental and approximate, not systematic and cogent, success. … Among the ill-founded frequentist devices are minimax rules, almost all tail-area tests, tolerance intervals, and, in a sort of class by itself, fiducial probability." The book does not mention Francis Galton, Karl Pearson, or G. Udny Yule, but does contain several citations to David Blackwell, Bruno De Finetti, R. A. Fisher, Maurice Kendall, Erich Lehmann, Oskar Morgenstern, Jerzy Neyman, John von Neumann, and Abraham Wald. The book received many reviews, mainly positive. A 1956 review in *JASA* by Francis Anscombe begins with "Despite the vastness of the literature on the theory of statistics and its applications, there is little to which a professional statistician can turn with hope of enlightenment on the fundamental ideas of his subject." He then puts Savage's book with Fisher's books and Jeffreys's *Theory of Probability* as meeting this need, but concludes that Savage exhibits excessive definiteness at the outset, and "Savage's theory of decisions is adequate for the discussion of many, but not all, problems in statistical methodology." In a 1955 review in *JRSS A*, however, I. J. Good states "The writing is always lucid, but the book as a whole is difficult to read, largely because of the conciseness of the style." In a later article, Savage (1961) stated "Good and I were both too deeply in the grip of frequentist tradition to do a thorough job." In fact, the chapters of Savage's book focusing on statistical inference dealt primarily with criticisms of the frequentist approach. Nonetheless, Savage's book has been highly influential, with more than 18,000 citations on Google Scholar.

Savage (1991) credited the 1959 textbook by Robert Schlaifer of the Harvard Business School as the first to be written "entirely and wholeheartedly from the Bayesian point of view". Schlaifer (1914–1994) was trained as a classical historian and Greek scholar, and when he self taught himself Statistics so that he could teach it in Harvard's business school, he soon decided that the prevailing Pearson/Fisher/Neyman synthesis was inadequate for business applications. Although this detailed 742-page book does not have the theoretical depth of other post World War 2 books described in this article, we include discussion of it because it was quite revolutionary in its Bayesian decision-making approach for presenting Statistics as a way of handling real-world problems. Table 15 shows its overall outline. In

Table 15 *Contents of Schlaifer's* Probability and Statistics for Business Decisions (1959)

Part title	Pages range
Introduction: The problem of decision under uncertainty	2–65
Part One: The use of probabilities based directly on experience	66–159
Part Two: Simple random processes and derived probabilities	160–329
Part Three: The use of information obtained by sampling	330–507
Part Four: The value of additional information	508–605
Part Five: Objective statistics: Tests of significance and confidence intervals	606–669

the preface, Schlaifer explained that he proposed analyzing business problems based on the theory of utility, the best decision being the one with its highest expected value when one assigned values to consequences and probabilities to events. The introductory section discussed the meaning of probability and defined expected value and utility. Part one of the text then introduced measures of location and showed how to assess probabilities by smoothing historical frequencies. Part two introduced standard discrete and continuous probability distributions with some normal approximations and ended with discussion of the Monte Carlo method as an alternative when mathematical analysis cannot be carried out analytically. Part three showed how to revise probabilities in the light of new information and gave results for sampling from normal populations. Part four considered decisions in the context of data obtained sequentially. The final part discussed the more classical approach to statistical inference using tests of hypotheses and confidence intervals as part of a sequential decision process, but evaluates consequent decisions in terms of expected loss.

In criticizing frequentist methods for statistical inference, Schlaifer states "Most *users* of tests of significance are intuitively interpreting statistical significance in a way which accords perfectly with Bayesian theory but not at all with classical theory by which they formally justify their procedures. A person responsible for a decision does not care about the *conditional* probabilities of making the wrong decision given each and every possible value of the basic random variables, but he may be very much interested in the *unconditional probability that the best terminal decision that he can make in the light of the evidence currently available will turn out to be wrong after the event*." He warns against making an error of the third kind by delivering a solution to the wrong problem, for instance by not taking into account utility and expected loss. He is critical of exact frequentist confidence intervals for a binomial proportion because of how the discreteness forces them to be conservative (e.g., the method for generating a 95% confidence interval from binomial tail probabilities merely ensuring that the a priori probability is *at least* 0.95 of capturing the parameter) unless one uses supplementary and irrelevant randomization. In any case, he argues that "*the long-run frequency with which a certain method of making statements would produce incorrect statements is of no real interest to anyone*" and that a statistician should be able to take account of other sources of information besides the particular sample of data. He then argues for interpreting confidence intervals for binomial parameters in terms of posterior probabilities that apply approximately with a uniform prior distribution for the parameter.

In a 1959 *JASA* review, Francis Anscombe begins "At the thought of yet another elementary book on statistics, this time for businessmen, over 700 pages long, suitable as a text for a one-year course to economics students—one's heart sinks. What a surprise when one summons up courage to open it! … This is an elementary exposition of the science of decision making under uncertainty, by one who has accepted L. J. Savage's *Foundations of Statistics* and related literature and is not hindered by a compulsion to reiterate what has been said in previous elementary texts. … Although the mathematical level is very low, it is a book for adults, and is not likely to prove useful for undergraduates. The integrity of the style is quite astonishing. The great length is due to careful and detailed explanations of calculations and to subtle discussions of the proper framing of problems." After describing how the subject matter differs greatly from what is usually called Business Statistics, including using a great variety of case studies to illustrate the role of utility in decision-making in production and marketing, Anscombe concludes, "Professor Schlaifer is to be congratulated on an extraordinary achievement."

Two years later, Schlaifer cowrote *Applied Statistical Decision Theory*, with Howard Raiffa of Harvard Business School. Their book provided a more detailed and systematic Bayesian treatment that also developed the idea of using conjugate prior and posterior distributions to simplify analyses computationally. In a 2008 interview by Stephen Fienberg in *Statistical Science*, Raiffa stated "The theme of our book was: it doesn't have to be too complicated; anything the classicists can do we can do also—only better. We discovered a simple algebraic way to go from priors to posteriors for sampling distributions that admitted fixed-dimensional sufficient statistics, like the exponential distributions." Table 16 shows its contents. A long preface provides justification and explanation for their approach. The authors stated "The logical and philosophical justification … has been fully developed by Savage in his *Foundations of Statistics*; the purpose of the present book is not to discuss these basic principles but to contribute to the body of analytical techniques and numerical results that are needed if practical decision problems are to be solved in accordance with them." After introducing the general decision process as a game against nature in which the optimum decision maximizes the expected value of a utility function, Bayes' theorem is introduced. A Bayesian sufficient statistic (equivalent to the traditional definition) is defined as the least you need from the data to get the same posterior distribution as you would get from all the data. Chapter 3 introduces conjugate prior and posterior distributions and summarizes weights in the posterior analysis that apply to the prior and the sample evidence. Later chapters introduce standard univariate probability distributions and the multivariate normal. They then apply the conjugate prior and posterior approach to estimating binomial and Poisson parameters and normal mean and variance parameters and regression parameters, with and without nuisance parameters being known. For stopping rules in collecting data, the authors rely essentially on the likelihood principle. The authors also consider optimal sample size, comparing two or multiple means, and selecting the best of several processes. "Terminal analyses" determine the expected utility of the optimal decision given each possible experimental outcome.

Table 16 *Contents of* Applied Statistical Decision Theory *by Raiffa and Schlaifer* (1961)

Chapter title	Pages range
1. The problem and the two basic modes of analysis	3–27
2. Sufficient statistics and noninformative stopping	28–42
3. Conjugate prior distributions	43–78
4. Additive utility, opportunity loss, and the value of information	79–92
5A. Linear terminal analysis	93–138
5B. Selection of the best of several processes	139–175
6. Problems in which the act and state spaces coincide	176–210
7. Univariate normalized mass and density functions	211–241
8. Multivariate normalized density functions	242–260
9. Bernoulli process	261–274
10. Poisson process	275–289
11. Independent normal process	290–309
12. Independent multinormal process	310–333
13. Normal regression process	334–353

A "preposterior" analysis considers the decision problem as it appears before collecting the data, when one can evaluate the expectation of the posterior expected utilities with respect to the prior distribution.

An overall positive 1961 review in *Technometrics* by Mervyn Stone suggests that whether methodologists believe that the book is a major contribution to analytical Bayesian techniques will depend on whether they agree with the authors' assertion that in common situations the family of conjugate distributions is sufficiently rich to adequately express prior information and beliefs. A 1962 review in *JASA* by Harry V. Roberts stated "I approached this book with a strong sympathy for the Bayesian point of view, so I was naturally predisposed to like the book. Yet my praise must exceed what I would accord on these grounds alone. By any standard this book is a remarkable achievement. It is clearly and logically presented. Much of it is novel. … Raiffa and Schlaifer have contributed a major work and a major stimulus to further work. I hope especially that mathematical statisticians will read it and be encouraged to develop the many fascinating lines of research that are suggested by it. I recommend it strongly to all statisticians seriously interested in applying statistics to real problems."

Later Statistics books having a Bayesian perspective include a two-volume work by Dennis Lindley (1965) and an introductory-level textbook by David Blackwell (1969). Books by I. J. Good (1965) and George E. P. Box and George Tiao (1973) gave overviews from applied perspectives, with Good focusing on estimating probabilities and Box and Tiao focusing on inference under the assumption of a normal distribution. Morris DeGroot (1975) wrote a mathematical statistics book with a Bayesian perspective that could better serve as a course textbook (later revised by Mark Schervish, with 4th edition in 2018). A later book by Bernardo and Smith (1994) gave a thorough review of the foundations, with emphasis on information-theoretic concepts and viewing statistical inference as a special case of decision theory. Since then, the textbook by Gelman et al. (1995, now in a 3rd edition) has been an especially popular one for presenting Bayesian inference.

8 The past, present, and future of statistical theory in the greater data science world

The main focus of this article has been on contributions in the first half of the twentieth century by Yule, Kendall, Wilks, Cramér, Fisher, and Snedecor that helped statisticians and scientists more fully to understand this rapidly-developing field. Fisher's contributions have been highly documented in multiple sources, such as by his daughter's biography (Box 1978), but some statisticians might not know much (if anything) about the others, all of whom made a multitude of important contributions to the field of Statistics besides their textbooks. For short but informative summaries of their contributions, see obituaries of Yule by Yates (1952), of Kendall by Stuart (1984), of Wilks by Mosteller (1964), of Cramér by Blom (1987), and of Snedecor by Kempthorne (1974). See also Chapters 21 and 56 of Lehmann (2008) for reminiscences of Wilks and Cramér.

In the past sixty years, numerous descendants of Yule's 1911 book have been published, with titles that include "mathematical statistics" or "theory of statistics" or even more all-encompassing titles, such as *All of Statistics* (Wasserman 2003). In the past 30 years, the textbook by Casella and Berger (1990, 2nd edition 2002) seems to have been the most popular one for first-year graduate courses in departments of Statistics and Biostatistics. Other textbooks that have received considerable use and positive recognition include Bickel and Doksum (1977, 2nd edition 2015), Cox and Hinkley (1974), Rohatgi (1976), and at a somewhat lower level, Rice (1988, 3rd edition 2013), and at a lower level yet, Mendenhall and Scheaffer (1973, 7th edition 2008 with Dennis Wackerly).

It is an ever-increasing challenge to write a book about the foundations of Statistics, because of the growth of the field in this era of data science, with continual introduction of new methods and new types of data and a broader scope that includes data preparation and visualization. Since the time period of primary focus in this article, Statistics as a field has grown and diversified so rapidly in response to widening needs across more and more contexts that it has become increasingly centered on computational methods and applications and less centered on narrow mathematical formulations and proofs. As the revolution instigated by Tukey (1977) in the data analysis expansion of the field of Statistics and by Breiman (2001) in formulating an algorithmic alternative to a modeling culture extends well beyond boundaries of mathematical statistics, statistical theory is likely to increasingly be viewed merely as part, and perhaps a very small part, of the broad vision of "greater data science" presented by Donoho (2017). In particular, in viewing Statistics in a general sense as the science of learning from data, it seems increasingly artificial for a book to place a boundary between "mathematical statistics" and methods of data analysis or to have purely a frequentist or Bayesian focus. So, considering this, for a book to be written now or in the future that includes discussion of the theoretical foundations of Statistics, what should be included and what is not fundamental enough for inclusion? The challenge is great when we keep in mind the limitations on how much can be reasonably devoured by a student in a one-year introductory course based on such a textbook.

In a book that I myself recently coauthored, focusing on the foundational concepts and methods with which we thought any undergraduate student who planned to be a data scientist should be familiar (Agresti and Kateri 2022), the titles of the first six of nine chapters were not much different from those in a mathematical statistics book such as Alexander Mood's seventy years ago. Our textbook focuses on descriptive statistics, probability distributions, sampling distributions, estimation and confidence intervals, significance testing, and linear models and least squares, before presenting material about some relatively newer areas such as generalized linear models, regularization methods, classification, and clustering. However, we thought it important also to rely strongly on simulations and on apps (such as Bernhard Klingenberg's outstanding apps at https://www.artofstat.com) and use examples employing `R` and `Python` to illustrate all concepts, thus avoiding any boundary between statistical theory and the rest of Statistics. In the future, books on theoretical statistics should and probably will also include more material on aspects of high-dimensional data and machine learning, such as now presented in specialized books such as Efron and Hastie (2016), Hastie, Tibshirani, and Friedman (2009), and Wainwright (2019).

To conclude this article, I asked Professor Xiao-Li Meng, with whom I had edited a book about the historical evolution of Statistics and Biostatistics departments in the U.S. (Agresti and Meng 2013), to contribute his thoughts about the presentation of the foundations of Statistics in the modern data science era. I also invited several reknowned statisticians, beginning with the first two winners of the International Prize in Statistics, Sir David Cox and Bradley Efron, to add brief comments with their thoughts about this or about the historical evolution of textbooks on the theory of Statistics. I am greatly appreciative of all the discussants for taking the time to add their thoughts, which follow:

Discussant 1: Xiao-Li Meng, Harvard University

Preeminent educator Alan Agresti's feat of historical introspection via reflection on representations of the foundations of statistical science in textbooks is likely to be as time-honored as it is timely. If we replace "statistical science" with "data science," "mathematics" with "statistics" or "mathematicians" with "statisticians", then various passages in Alan's article about the formation of statistics in the 20th century could be applied almost verbatim to a yet unwritten account of the present and prospective formation of data science in the 21st century.

About a century ago, statistics needed to attain the mathematical rigor necessary to establish itself as a respectable academic discipline, as evidenced by the "Explosion in Mathematical Statistics" highlighted in Section 6 of Alan's overview. A century later, an intense race is on to provide rigorous statistical foundations for many learning algorithms, which collectively form the most well-recognized—but the least well-understood—core of data science. This race is scientifically necessary both for the growth of the field of machine learning and for its general recognition as a coherent scientific discipline, as distinguished from an assemblage of trials and errors (and prayers). Rumor has it that while statisticians may not burnish their reputations as deep scholars by publishing in leading journals in machine learning, machine learners would command considerably more respect as rigorous researchers by publishing consistently in the *Annals of Statistics* and the like. Yet although this imbalance reflects a natural evolution of disciplines, it should also remind us to revisit our formative years to glean historical insights and lessons. Alan's article is timely precisely because it opens up a rich archive of such lessons with expertly arranged highlights. As such, it is amply suitable for a self-guided historical tour replete with reminiscences and reflections.

My own tour led me to take several long pauses or detours in the process of contemplating the contemporary implications of some historical lessons. For example, as Alan reported, the effort to rigorize statistics by introducing more mathematics within it had a profoundly negative impact on statistical education, especially at the introductory level. For decades, many introductory courses in statistics were taught by mathematicians who had no experience or even an interest in what makes statistics a separate discipline, that is, a discipline to principally explore and coherently govern the world of uncertainty. The nature of this world is inherently more nuanced and imprecise than can be tolerated or even comprehended by a precisely framed and deterministically oriented mindset. Drawing from this historical lesson, a prospective Alan Agresti of the future may well be wondering if the current attempt to rigorize learning algorithms via statistical interrogations could have a similar side effect—namely, the relegation of data science courses to statisticians who would teach data science just as statistics. (As an aside, I am acutely aware of the fact that there are still a good number of fellow statisticians who consider data science to be merely a "sexified" version of statistics, to use a term attributed to Nate Silver. While this is not the place to engage in this debate, I am preparing an editorial arguing that equating data science with statistics is tantamount to claiming that physical science is no more than physics.)

The concern that data science might end up being taught as "just statistics" is, then, legitimate. Having been a statistician for three decades, I am confident that with some preparation, I can teach almost any statistical topic reviewed in this article and, in doing so, provide decent insights and reasonable explanations (yes, even for fiducial inference). But no matter how much effort I might put into it, the same could not be said if I were to teach database management, distributed computing, human-computer interaction, etc. Anyone native to these fields would instantly recognize that, at best, I would be a foreigner struggling to make my utterances coherent. These are all essential fields in the endeavor to ensure that data science gets done (right), but most statisticians do not have experience or even an interest in such topics. Fortunately, unlike a century or even decades ago, where teaching was a mostly siloed endeavor with almost everyone left to sink or swim, team teaching is now rather common. I, for one, have been a part of the team of computer scientists and statisticians that offers Harvard's first introductory undergraduate course in data science. Nevertheless, the current development of data science education is far behind that of data science research, precisely because data science is not a single discipline. Hence, designing a coherent or even feasible curriculum is exceedingly challenging. Indeed, what would be a cogent curriculum for natural science, or social science, or humanities?

Table 17 *Contents of Wasserman's* All of Statistics (2004)

Chapter title	Pages range
1. Probability	3–18
2. Random variables	19–46
3. Expectation	47–62
4. Inequalities	63–70
5. Convergence of random variables	71–86
6. Models, statistical inference and learning	87–96
7. Estimating the CDF and statistical functionals	97–106
8. The bootstrap	107–118
9. Parametric inference	119–148
10. Hypothesis testing and p-values	149–174
11. Bayesian inference	175–192
12. Statistical decision theory	193–208
13. Linear and logistic regression	209–230
14. Multivariate models	231–238
15. Inference about independence	239–250
16. Causal inference	251–262
17. Directed graphs and conditional independence	263–280
18. Undirected graphs	281–290
19. Log-linear models	291–302
20. Nonparametric curve estimation	303–326
21. Smoothing using orthogonal functions	327–348
22. Classification	349–380
23. Probability redux: stochastic processes	381–402
24. Simulation methods	403–420

As another historical reminder, Arthur Bowley's *Elements of Statistics* (1901), which I had never heard of prior to reading Alan's article, earned a historical place not for being a founding document but rather a relic of late empire, as it were. One may wonder if any of current textbooks in statistics would receive a similar distinction from an Alan Agresti of the 22nd century. We certainly have come a long way as a field from Yule's 1911 book (Table 1 of the article), which typifies the formative era of systemized probabilistic treatments of our beloved subject. Comparing Kendall's *Advanced Theory of Statistics* (Table 6 and Table 7) with Larry Wasserman's *All of Statistics* from some 70 years later (Table 17) also provides us a general sense of both time invariant and time-varying topics and identifies issues that we should not overlook.

I choose Wasserman's *All of Statistics* as a comparand for Kendall's *Advanced Theory of Statistics* mostly because both strive for breadth. The former—which features as its title an inside academic joke, especially in view of its 2006 sequel *All of Non-parametric Statistics*—is, however, more ambitious than the latter, as it aims to cover both theory and methods. The topics covered in the first half of Wasserman's book—with the obvious exception of Bootstrap—can easily be found in Kendall's volumes. Other topics, in each textbook suggest a few clear historical indicators, as well as some less clear individual choices. For example, cumulants and fiducial inference rarely enter contemporary statistical textbooks, whereas causal inference and simulation methods are increasingly featured. However, the omission of data collection principles and methods (e.g., sampling, experimental designs) in *All of Statistics* would likely be cited by a future historian as an example of why statistics is currently often misperceived as only relevant for the data analysis step in the life cycle of data science. That is a misperception because both sampling and design were clearly in Kendall's volumes, and indeed Yule's 1911 textbook had a chapter on "Effect of removing the limitation of simple sampling." The issue of data collection design (or lack thereof) currently has increasingly vast implications as our society comes to rely more and more on "big data" that are not collected with much quality control as is offered by probabilistic sampling, simple or not. Therefore, the rich statistical insights and methods on data collections should be made known as widely as ever, especially in textbooks aimed for the broad data science community.

Historical introspections as such reminds us of the continuing effort that every generation must make to sustain what is time-honored, especially when our attention can be—and often is—easily consumed by the demands of what is timely. The healthy evolution of a discipline necessarily entails the conscious deepening of its foundations coupled with the expansion its horizons, especially when facing many new challenges. The field of statistics is certainly at such a historical juncture, as it has the opportunity and responsibility to be a leading (though by no means the sole) force in the rapidly evolving data science ecosystem. For that, we should all be grateful to Professor Alan Agresti for this extremely timely historical treasure.

Discussant 2: Sir David Cox, Nuffield College, Oxford University, UK

Congratulations on this massive effort. However, the trouble with restricting to textbooks is a distorted picture of the history, I think. In my view, subjective judgement of course, the three most important publications up to 1940 are Fisher's 1922 paper "On the mathematical foundations of theoretical statistics" (*Philosophical Transactions of the Royal Society A*), his 1925 paper "Theory of statistical estimation" (*Proceedings of the Cambridge Philosophical Society*), and his 1926 paper "The arrangement of field experiments" (*Journal of the Ministry of Agriculture of Great Britain*) which led to his 1935 book *The Design of Experiments*.

Cramér was indeed a towering figure in Sweden; the English translation of his general book was massively influential in the U.S. in particular. Sam Wilks' lecture notes from 1943 were influential but his much-criticized 1962 Wiley book made the mistake of trying to make the mathematics rigorous. Another major figure, more on the applied side, was Anders Hald in Copenhagen, who wrote a lot about history and was highly influential in various ways, such as setting up the Bernoulli Society; also, Frank Yates, whose books on field trials and sampling were very influential and he was a pioneer in statistical computing.

Kendall was a man of phenomenal energy. Kendall volume 1 was largely written in air-raid shelters. At the same time MGK had a demanding full-time job in shipping control. World War 2 had the effect of a massive increase in interest in academic statistics, in Cambridge, Oxford, Imperial in the UK and Harvard, Berkeley, Columbia, Stanford, etc. In my own case it is highly unlikely that I would have become a statistician without the War, and I am one of many.

Discussant 3: Bradley Efron, Stanford University, USA

Writing a good statistics textbook used to be a balancing exercise between mathematics and statistical inference, trying to not let mathematical accuracy drown out the inferential structure. These days there are three poles to balance: mathematics, statistics, and computation. Modern students, at least those at Stanford, love sitting in front of the computer more than paper and pencil work so tomorrow's successful textbook might have fewer theorems and proofs, and more algorithmic diagrams. Hastie, Tibshirani, and Friedman's *Statistical Learning* text is a highly successful example of this genre. In any case, mathematical rigor is an overated pedagogical tactic. This is especially true for elementary texts. Most often, mathematically sophisticated topics like Student's t preceed simpler computational methods such as Wilcoxon's test; permutations, bootstraps, cross-validations etc could be used to introduce basic statistical ideas (testing, accuracy, prediction), saving parametric theory for later.

That being said, I'm a little worried that the Big Data emphasis on nonparametric applications will push advanced parametric methods like classical multivariate analysis and exponential families into the realm of the past. The training of US naval officers requires time on three-masted sailing ships, the point being that a full understanding of something depends on knowledge of the basics. Cox and Hinkley's 1974 book *Theoretical Statistics* shows that it is possible to navigate deep statistical waters without getting stuck in the Sargasso Sea of asymptotics. A contemporary version that took on 21st Century topics would be most welcome. Meanwhile we should thank Professor Agresti for an ambitious and successful 150 year review to the textbook literature.

Discussant 4: Helen MacGillivray, Queensland University of Technology, Australia

As in other disciplines, the roles of mathematics in the statistical and data sciences are to help support, develop, justify, underpin and unify modelling, theories, principles, concepts, procedures, experimental work and usage. The books of Professor Agresti's beautifully constructed journey through the first 50–60 years of the twentieth century combined research and conceptual development with mathematical support for the 'professional' statistical work of that era. Because the statistical sciences should be viewed inclusively from the full processes of data investigations and analysis to stochastics, 'mathematical statistics' should perhaps now be considered as the mathematics support of everything involving data, variation and uncertainty. And we have increasingly seen also the blurring and dissolving of the boundaries between mathematical and computational support for the sciences of data, variation and uncertainty.

The books of the second half of the last century in Professor Agresti's discussion tend to also indicate the emerging variety supporting research, advanced, or 'introductory' learning, plus a flavour of the rapidly broadening sciences of statistics, stochastics and data. Both these aspects lead to his questions about content. Because both content potential and approach possibilities are now immense, it is imperative that books should be purposeful and manifest in their 'story'. I suggest the emphasis be that statistical (and data science) textbooks should be very clear on their purpose, audience, issues and assumptions articulated throughout, and that the chosen statistical 'journey' be consistent and coherent—the last does not preclude some modularisation but the development pathways need to be well-identified with consistent exposition level. Justifications should be lucid and relevant, whether they be mathematical, contextual, conceptual, data-driven, computational, visual, or any combination. And because developments in statistics and data science, no matter how theoretical or applied, have real problems at their root, connections and motivations, real contexts and data, visualisations, synthesis of findings and fusion of approach all contribute to a purposeful statistical narrative.

For advanced undergraduate and masters levels, the more unified the mathematics content the better, even if results are given with references to further exposition and proofs. Unification is particularly valuable for historical results—for example, likelihood results, the elegant matrix results unifying general (and generalised) linear model inference theory and the (different) matrix results doing the same for Markov processes. This also enables more succinct exposition and some important historical results to be retained in context with their underpinning role(s) explained. The range of statistical capabilities which have been given, or have arisen from, mathematical underpinning has increased so enormously alongside computational power that the above principles of textbook writing are critically necessary. Carefully selected proofs can provide valuable technical learning and elucidate the power of mathematical models at appropriate levels. However the statistical stories of all such results should also incorporate data, context and usage. Surely Covid-19 highlights that statistics must ensure ownership of probability and stochastic modelling, without any artificial separation from data and inference.

For foundational statistics at college and university, there have been some unfortunate repressive effects of minimising and removing mathematics without re-thinking structure, the statistical 'story', and the above principles of textbook writing. If an introductory textbook is intended to be data-driven, cultivating the full statistical investigation process, as increasingly recommended for both future statisticians and for those in other disciplines, then it should develop coherently for this purpose. Left-over 'ghosts' include arbitrary interruptions of old-style set-based probability, excessive focus on theoretical sampling distributions, and the suffocating effects of constraining foundational inference, no matter in what form, to just one or two variables. There is also no 'fixed' introductory inferential path: diverse statisticians have independently advocated various pathways including starting with categorical data; moving directly to many variables; working via bootstrapping; but always embedded within data investigations.

Whatever view of data science is taken, it provides both opportunities and challenges, including technology as support not domination. Just as the great pioneering minds of statistics used and grew mathematics to underpin statistical progress, so too can mathematics and technology together and in their own ways help to advance the statistical and data sciences in themselves and in their contributions across all disciplines.

Discussant 5: Susan Murphy, Harvard University

A first thought is that the future of textbooks on theory of Statistics is dim. Indeed, I keep few textbooks on my shelves and those kept function more as remembrances. However, this viewpoint is too narrow. I have virtual copies of many textbooks and I, as well as my students, refer to a number of these textbooks frequently. Two classical textbooks are van der Vaart's *Asymptotic Statistics* (1998) and van der Vaart & Wellner's *Weak Convergence and Empirical Processes* (1996). Even in these days of high dimensional data, classical theory for high dimensional models can provide a starting point! More recently statisticians and computer scientists have rediscovered the importance and challenges of sequential decision making. It is a joy to discover and read books that are likely to become classics such as book by Agarwal et al., *Reinforcement Learning: Theory and Applications*; this book mixes statistics (likelihood methods, confidence regions, probabilistic models, minimax theory), probability (high probability bounds) and optimization to provide/develop a modern theory for sequential decision making. Yes, the future of textbooks on theory in Statistics is bright, particularly in emerging areas such as sequential decision making and high dimensional data both of which require the development of statistical theory to make progress.

Discussant 6: Andrew Gelman, Columbia University, USA

It was fun to read Alan Agresti's review of a century of statistics textbooks. As we get older we all become historical experts of one sort or another, and often what's striking is not the changes in teaching and theory during fifty years or more, but the continuity. It has been thirty-five years since my first time teaching statistics, and it is my impression that during this time the teaching of advanced statistics has changed a lot–the central topics now are machine learning and Bayesian statistics, and even the core of classical theory has adapted to address problems of big data and big models–but introductory teaching has changed much less in its concepts. Computing has supplanted mathematics to some extent, but the basic sequence of displaying data, collecting data, basic probability, sampling distributions, unbiased estimation, hypothesis testing, and confidence intervals–that's pretty much unchanged, as can be seen in various well-regarded textbooks such as Moore, McCabe, and Craig (2021) or the U.S. high school Advanced Placement syllabus. I think we can and soon will do better, but this will require an integration of teaching with theory that cannot happen all at once. It will help for creators of new courses and writers of new textbooks to be aware of the problems of data science and to be willing to teach programming and numeracy instead of relying on algebra and calculus, but in addition I think it will be necessary to go beyond the existing frameworks of hypothesis testing, sampling theory, and static Bayesian inference and move to a more active role of statistical modeling to generalize from data to real-world conclusions. It is a saying beloved of teachers that the best way to learn a subject is to teach it. Continuing, the best way to teach a subject is to understand it, and applying the contrapositive of this principle suggests that challenges in teaching reflect gaps in our understanding. We are in an exciting era in which our understanding of principles and methods of statistics and data science is increasing rapidly–as demonstrated in many recent research articles and advanced textbooks–and I anticipate this will soon result in radically restructured and improved introductory teaching as well.

Discussant 7: Nicholas Jon Horton, Amherst College, USA

In 2003, esteemed theoreticians Nancy Reid, Bradley Efron, and Carl Morris participated in a panel at the Joint Statistical Meetings chaired by David Moore to discuss whether the "Math Stat" course was obsolete (see notes by Rossman and Chance, 2003). In his opening statement, Moore posited that "The math stat course has not changed in 40 years, whereas statistics has changed enormously, so how could the course not be obsolete"? This reality is even more true 18 years later.

Professor Agresti is to be commended for his important and insightful history of mathematical statistics textbooks. Whatever path we take next requires an understanding of our history. His review is a helpful step to help ground future developments.

A key shift in the discipline of statistics in recent decades has been the dramatic rise of computation. Modern statistics and data science methods leverage computation to an increasing extent (Statistics at a Crossroads, 2019; Hardin et al., 2021). I would argue that computation is now as important a pillar for statistics and data science as mathematics. This reality needs to be reflected throughout our curricula including our theory courses.

George Cobb (2011) outlined the importance of data and computing and highlighted approaches that he felt were necessary but not sufficient innovations (examples include Nolan and Speed (2000) and Horton (2013)). The mathematical statistics course has been gradually evolving to incorporate data (a long-standing joke is that traditional textbooks have always had lots of data: it's the little "x"'s!). Our theory courses need to embrace these shifts at the undergraduate and graduate levels to ensure that our students have the foundation to appropriately engage in statistics and data science practice.

I have been teaching a group and project-focused course using Rice's *Mathematical Statistics and Data Analysis* with the goal of infusing computation, fostering multiple aspects of statistical practice (e.g., collaboration and reproducibility), deepening student understanding, and helping them to build data acumen using parallel analytic and empirical problem solving (Horton, 2013). It was encouraging to see the outlines of the textbook that Professor Agresti and colleagues have proposed. I look forward to teaching from it!

Discussant 8: Maria Kateri, RWTH Aachen University, Germany

I enjoyed reading the enlightening overview by Alan Agresti on the roots of key concepts in statistics, commenting fundamental and highly influential literature. I thank him for inviting us to think on the need of reorganizing courses and textbooks about statistical theory—a fact that is also linked to concerns on the contemporary and future role of statistical science in the era of data science and big data. The core content of a textbook on the theory of Statistics remains the same (i.e., descriptive statistics, basics of probability theory and probability distributions, point and interval estimation, hypothesis testing and linear models). Depending on the respective level and the audience, further models and methods can be addressed with a clear focus on multivariate analysis. A necessary add-on should deal with fundamentals of simulation methods and computational statistics, and an introduction to, for example, bootstrap confidence intervals or Monte Carlo simulations. What needs to be updated is the way of motivating, presenting and justifying the topics, linking them to actual problems and applications. A problem-driven approach would be more attractive and effective. In this sense, I would also welcome a strengthened role of Bayesian procedures, presented in parallel with frequentist approaches. Furthermore, without abandoning mathematical proofs, some of them could be replaced by simulation-based verification and reasoning. To cover needs of machine learning applications, more emphasis should be given on topics like prediction and prediction intervals, model selection and verification (including, e.g., cross-validation). Such an orientation and broadening of Statistics textbooks would stimulate statistical thinking and, most importantly, would illustrate and emphasize the fundamental role of Statistics within Data Science.

Discussant 9: John Hinde, National University of Ireland Galway, Ireland

Alan Agresti has presented a highly illuminating and authoritative account of statistical inference texts that nicely charts the development of the subject in the 20th century. It is very interesting to see the evolution of the content and how, over the years, things come in and other things go out. Predominately, we see an increased mathematisation of the content consistent with the tools and skills that students required to understand, apply, and, subsequently, to advance the subject. The interesting question that faces us now is what should a text for the 21st century look like? What do students in the broader field of data science need to know to face the challenges of applying, understanding and extending appropriate inferential tools? For sure, any text needs to reflect the increased computerisation of the subject and the issues associated with large data—how do we do inference in these settings? But any text also needs to include the basic foundations, principles and techniques of small sample statistics—data summarisation, point and interval estimation, hypothesis testing, etc. However, to address the issues associated with the large and diverse data sources now routinely available I would like to see some focus on data-generating processes, for example to include streaming data, continuous space-time monitoring, etc; these are becoming commonplace in so many areas and are frequently inherently multivariate. I think that semi-parametric methods will also be increasing important, which in themselves will require exposure to additional mathematical and statistical ideas. Finally, I believe that texts should also include more on design, as the principles and methods that underpin the design of experiments are really at the foundations of our subject. This knowledge may be useful for handling the massive datasets that will be routine and the associated ideas will also be invaluable for the appropriate and efficient use of computer intensive methods for inference and exploration in the large data context.

Discussant 10: Gauss Cordeiro, Federal University of Pernambuco, Brazil

The twentieth century was a time of great progress in the field of mathematical statistics. I enjoyed reading about the first textbooks published in statistical methods and mathematical statistics by Yule, Rietz, Fisher, Snedecor, Wilks, Kendall (two volumes), Cramér, Mood, Hogg and Craig and Rao. The theory of Statistics from a Bayesian perspective was presented in the textbooks by Jeffreys, Savage and Schlaifer. The chapters of these textbooks are listed in tables along with historical facts and interesting comments, which greatly helps in understanding and absorbing the material. The paper is an excellent survey of the development of statistical textbooks during the first 60 years of the twenty century. When C. R. Rao received an honorary doctorate degree from University of Brasilia in the mid-1990s, he defined Statistics as a sum of science, technology, and art. Because of the increase of computer technology and knowledge, I believe that all statistical concepts in these old classical textbooks, will be applied in data science in several fields especially to improve quality of life, accelerate finding cures for specific diseases, and to help to solve many problems on health, social, economics and industry. My congratulations to Professor Agresti for this very valuable work!

Acknowledgments

I would like to thank Sir David Cox for numerous suggestions and historical insights. Thanks also to Stephen Stigler for telling me about Fry (1928) and Pearson (1914), to Malay Ghosh for reminding me about Fisz (1963), and to Phil Brown, David Hitchcock, Bernhard Klingenberg, Helen MacGillivray, Luigi Pace, Alessandra Salvan, and Gerhard Tutz for helpful comments. Xiao-Li Meng thanks Suzanne Smith for making his utterances coherent.

References

Agresti, A. and Kateri, M. (2022). *Foundations of Statistics for Data Scientists, with R and Python*. Boca Raton: CRC Press.

Agresti, A. and Meng, X.-L. (2013). *Strength in Numbers: The Rising of Academic Statistics Departments in the U.S.* New York: Springer. MR3077328 https://doi.org/10.1007/978-1-4614-3649-2

Aitken, A. C. (1939). *Statistical Mathematics*. Edinburgh: Oliver and Boyd. MR0001493

Aldrich, J. (2005). The statistical education of Harold Jeffreys. *International Statistical Review* **73**, 289–307.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley. MR0091588

Bartlett, M. S. (1940). The present position of mathematical statistics (with discussion). *Journal of the Royal Statistical Society* **103**, 1–29.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley. MR1274699 https://doi.org/10.1002/9780470316870

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day. MR0443141

Bickel, P. J. and Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics, Volume II*. London: Chapman and Hall / CRC. MR3287337

Bishop, Y. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis*. Cambridge: MIT Press. MR0381130

Blackwell, D. (1969). *Basic Statistics*. New York: McGraw-Hill. MR1535965 https://doi.org/10.2307/2316759

Blalock, H. M. (1960). *Social Statistics*. New York: McGraw-Hill.

Blom, G. (1987). Harald Cramér 1893–1985. *The Annals of Statistics* **15**, 1335–1350. MR0913560 https://doi.org/10.1214/aos/1176350596

Bowley, A. L. (1901). *Elements of Statistics*. London: P. S. King & Son.

Box, J. F. (1978). *R. A. Fisher: The Life of a Scientist*. New York: Wiley. MR0500579

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading: Addison-Wesley. MR0418321

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science* **16**, 199–231. MR1874152 https://doi.org/10.1214/ss/1009213726

Brownlee, K. A. (1960). *Statistical Theory and Methodology in Science and Engineering*. New York: Wiley. MR0119268

Brunk, H. D. (1960). *An Introduction to Mathematical Statistics*. Boston: Ginn and Company. MR0114257

Burgess, R. W. (1927). *The Mathematics of Statistics*. Boston: Houghton Mifflin.

Casella, G. and Berger, R. (1990). *Statistical Inference*. *Wadsworth & Brooks/Cole*. MR1051420

Cobb, G. W. (2011). Teaching statistics: Some important tensions. *Chilean Journal of Statistics* **2**, 31–62. http://soche.cl/chjs/volumes/02/01/Cobb(2011).pdf. MR2820017

Cochran, W. G. (1953). *Sampling Techniques*. New York: Wiley. MR0054199

Cochran, W. G. and Cox, G. (1950). *Experimental Designs*. New York: Wiley. MR0034558

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall. MR0370837

Cramér, H. (1937). *Random Variables and Probability Distributions*. *Cambridge Tract in Mathematics* **36**. Cambridge. MR0165599

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press. MR0016588

de Finetti, B. (1974, 1975). *Theory of Probability*. New York: Wiley. Two volumes translated by A. Machi and A. F. M. Smith. MR0440641

DeGroot, M. (1975). *Probability and Statistics*. Reading, MA: Addison Wesley. MR0373075

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics* **26**, 745–766. MR3765335 https://doi.org/10.1080/10618600.2017.1384734

Draper, N. R. and Smith, H. (1967). *Applied Regression Analysis*. New York: Wiley. MR0211537

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge: Cambridge University Press. MR3523956 https://doi.org/10.1017/CBO9781316576533

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd. (14th ed. 1970.) MR0346954

Fisher, R. A. (1990). *Statistical Methods, Experimental Design, and Scientific Inference*. London: Oxford University Press. MR0346955

Fisz, M. (1963). *Probability Theory and Mathematical Statistics*. New York: Wiley. MR0164358

Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics*. New York: Wiley. MR0083868

Freund, J. E. (1962). *Mathematical Statistics*. New York: Prentice-Hall. MR0166853

Fry, T. C. (1928). *Probability and Its Engineering Uses*. Princeton: D. Van Nostrand Company, Inc. MR0181027

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall. MR1385925

Good, I. J. (1950). *Probability and the Weighing of Evidence*. New York: Hafner. MR0041366

Good, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. Cambridge, MA: MIT Press. MR0185724

Hardin, J., Horton, N. J., Nolan, D. and Temple Lang, D. (2021). Computing in the statistics curricula: A 10-year retrospective. *Journal of Statistics and Data Science Education* **29**(sup1), S4–S6. https://doi.org/10.1080/10691898.2020.1862609

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer. MR2722294 https://doi.org/10.1007/978-0-387-84858-7

Hill, A. B. (1937). *Principles of Medical Statistics*. London: Lancet.

Hoel, P. G. (1947). *Introduction to Mathematical Statistics*. New York: Wiley.

Hogg, R. V. and Craig, A. T. (1959). *Introduction to Mathematical Statistics*. New York: Macmillan. MR0137186

Horton, N. J. (2013). I hear, I forget. I do, I understand: A modified Moore-method mathematical statistics course. *American Statistician* **67**, 219–228. MR3303815 https://doi.org/10.1080/00031305.2013.849207

Hotelling, H. (1940). The teaching of statistics. *The Annals of Mathematical Statistics* **11**, 457–470. MR0120695

Hotelling, H. (1951). The impact of R. A. Fisher on statistics. *Journal of the American Statistical Association* **46**, 35–46. MR0359184

Jeffreys, H. (1939). *Theory of Probability*. Oxford: Clarendon Press. MR0000924

Kempthorne, O. (1974). *International Statistical Review* **42**, 319–323. Snedecor, George W., 1881–1974. MR0386955

Kendall, M. G. (1943). *The Advanced Theory of Statistics, Vol. 1*. London: Griffin. MR0019869

Kendall, M. G. (1946). *The Advanced Theory of Statistics, Vol. 2*. London: Griffin. MR0019869

Kendall, M. G. and Stuart, A. (1958). *The Advanced Theory of Statistics. Volume 1: Distribution Theory*. London: Griffin. MR0019869

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. New York: Wiley. MR0107933

Lehmann, E. L. (2008). *Reminiscences of a Statistician*. Berlin: Springer. MR2367933 https://doi.org/10.1007/978-0-387-71597-1

Lindgren, B. W. (1962). *Statistical Theory*. New York: Macmillan. MR0143273

Lindley, D. V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint; Volume 1 Probability, Volume 2 Inference*. Cambridge: Cambridge University Press. MR0168083

Magnello, M. E. (2009). Karl Pearson and the establishment of mathematical statistics. *International Statistical Review* **77**, 3–29. MR0541179

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. London: Chapman & Hall. MR0727836 https://doi.org/10.1007/978-1-4899-3244-0

Mendenhall, W. and Scheaffer, R. L. (1973). *Mathematical Statistics with Applications*. N. Scituate: Duxbury Press. MR0353508

Mood, A. M. (1950). *An Introduction to the Theory of Statistics*. New York: McGraw-Hill. MR0033470

Mood, A. M. (1990). Miscellaneous reminiscences. *Statistical Science* **5**, 38–43. MR1054856

Mosteller, F. (1964). Samuel S. Wilks: Statesman of statistics. *American Statistician* **18**, 11–17. MR0191011

Neyman, J. (1949). *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press. MR1529649 https://doi.org/10.2307/2310188

Noether, G. E. (1989). The teaching of statistics in the USA: The early years. In *Sesquicentennial Invited Paper Sessions, Proceedings of the American Statistical Association*, 279–285. Alexandria, VA: American Statistical Association.

Nolan, D. and Speed, T. (2000). *Stat Labs: Mathematical Statistics Through Applications*. https://www.stat.berkeley.edu/users/statlabs.

Pearson, K. (1914). *Tables for Statisticians and Biometricians*. Cambridge: Cambridge University Press.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Harvard Business School Division of Research. (Later published by MIT Press and in 2000 in Wiley Classics Library.) MR1789469

Randles, R. H. (2007). A conversation with Robert V. Hogg. *Statistical Science* **22**, 137–152. MR2408664 https://doi.org/10.1214/088342306000000637

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: Wiley. MR0050824

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*. New York: Wiley. MR0221616

Rice, J. A. (1988). *Mathematical Statistics and Data Analysis*. Brooks/Cole Pub. Co.

Rietz, H. L., ed. (1923). On the subject matter of a course in mathematical statistics. *The American Mathematical Monthly* **30**, 155–166. MR1520205 https://doi.org/10.2307/2298448

Rietz, H. L., ed. (1924). *Handbook of Mathematical Statistics*. Boston: Houghton Mifflin.

Rietz, H. L. (1927). *Mathematical Statistics*. Mathematical Association of America.

Robert, C. P., Chopin, N. and Rousseau, J. (2009). Harold Jeffreys's theory of probability revisited. *Statistical Science* **24**, 141–172. MR2655841 https://doi.org/10.1214/09-STS284

Rohatgi, V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics*. New York: Wiley. MR0407916

Rossman, A. and Chance, B. (2003). Notes from JSM 2003 panel "Is the math stat course obsolete?". https://nhorton.people.amherst.edu/MathStatObsolete.pdf.

Samuelson, P. A. (1950). Recent texts in mathematical statistics. *Review of Economics and Statistics* **32**, 139–143.

Savage, L. J. (1954). *The Foundations of Statistics*. New York: Dover. MR0063582

Savage, L. J. (1961). The foundations of statistics reconsidered. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability; Volume 1: Contributions to the Theory of Statistics* (J. Neyman, ed.) 575–586. Berkeley: University of California. MR0133898

Scheffé, H. (1967). *The Analysis of Variance*. New York: Wiley. MR0116429

Schlaifer, R. (1959). *Probability and Statistics for Business Decisions*. New York: McGraw-Hill.

Snedecor, G. W. (1937). *Statistical Methods, Subtitle Applied to Experiments in Agriculture and Biology*. Ames, Iowa: Collegiate Press.

Statistics at a Crossroads (2019). https://www.nsf.gov/mps/dms/documents/Statistics_at_a_Crossroads_Workshop_Report_2019.pdf.

Stigler, S. M. (1978). Mathematical statistics in the early states. *The Annals of Statistics* **6**, 239–265. MR0483118

Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge: Harvard University Press. MR0852410

Stigler, S. M. (1999). *Statistics on the Table*. Cambridge: Harvard University Press. MR1712969

Stuart, A. (1984). Sir Maurice Kendall, 1907–1983. *Journal of the Royal Statistical Society Series A* **147**, 120–122. MR0748794

Tukey, J. (1977). *Exploratory Data Analysis*. Reading: Addison Wesley.

Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-asymptotic Viewpoint*. Cambridge: Cambridge University Press. MR3967104 https://doi.org/10.1017/9781108627771

Wald, A. (1942). *On the Principles of Statistical Inference*. *Notre Dame Mathematical Lectures* **1**. Notre Dame: University of Notre Dame Press. MR0006681

Wald, A. (1947). *Sequential Analysis*. New York: Wiley. MR0020764

Wald, A. (1949). *Statistical Decision Functions*. New York: Wiley. MR0036976

Wasserman, L. (2003). *All of Statistics: A Concise Course in Statistical Inference*. Berlin: Springer. MR2055670 https://doi.org/10.1007/978-0-387-21736-9

Weatherburn, C. E. (1946). *A First Course in Mathematical Statistics*. Cambridge: Cambridge University Press. MR0019262

West, C. J. (1918). *Introduction to Mathematical Statistics*. Columbus, Ohio: R. G. Adams and Company.

Wilks, S. S. (1937). *The Theory of Statistical Inference*. Princeton: Princeton University.

Wilks, S. S. (1947). *Mathematical Statistics*. Princeton: Princeton University Press. MR0008657

Wilks, S. S. (1948). *Elementary Statistical Analysis*. Princeton: Princeton University Press.

Wilks, S. S. (1962). *Mathematical Statistics*. New York: Wiley. MR0144404

Wishart, J. (1939). Some aspects of the teaching of statistics (with discussion). *Journal of the Royal Statistical Society* **102**, 532–564.

Wittstein, T. (1872). On mathematical statistics and its application to political economy and insurance. *Journal of the Institute of Actuaries and Assurance Magazine* **17**, 178–189. Translated by T. B. Sprague.

Wolfenden, H. H. (1942). *The Fundamental Principles of Mathematical Statistics*. New York: Macmillan. MR0006636

Yates, F. (1949). *Sampling Methods for Censuses and Surveys*. London: Charles Griffin and Co. MR1527705 https://doi.org/10.2307/2308128

Yates, F. (1952). George Udny Yule, 1871–1951. *Biographical Memoirs of Fellows of the Royal Society* **8**, 309–323. MR0051789

Yule, G. U. (1911). *An Introduction to the Theory of Statistics*. London: Charles Griffin and Co.

Yule, G. U. and Kendall, M. G. (1950). *An Introduction to the Theory of Statistics*, 14th ed. London: Charles Griffin and Co. MR0035938