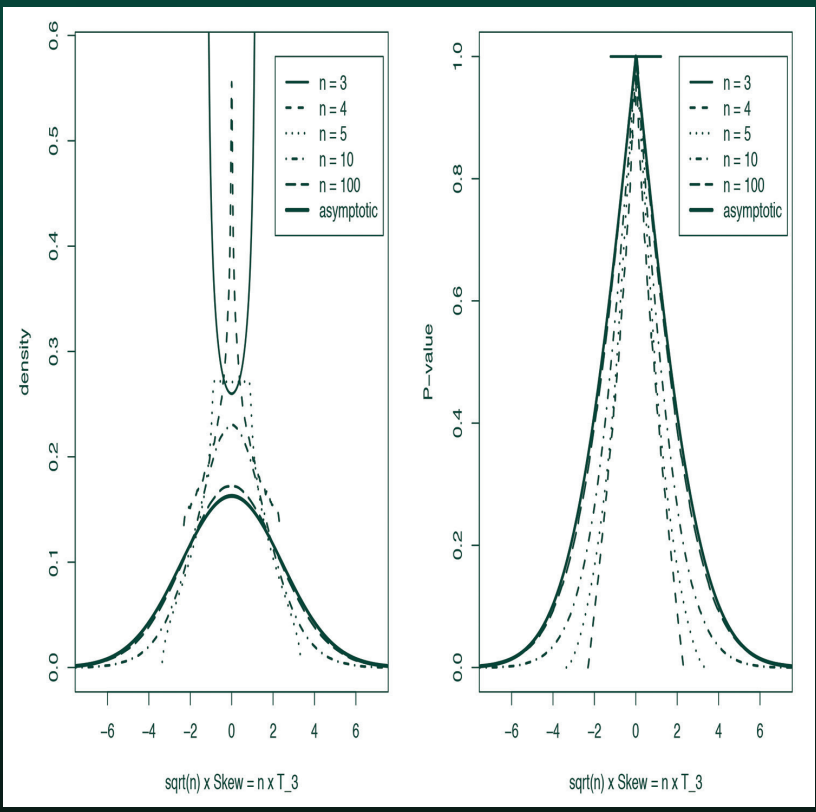
Monographs on Statistics and Applied Probability 144

Measuring Statistical Evidence Using Relative Belief



Michael Evans

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Measuring Statistical Evidence Using Relative Belief

MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

F. Bunea, V. Isham, N. Keiding, T. Louis, R. L. Smith, and H. Tong

1. Stochastic Population Models in Ecology and Epidemiology *M.S. Barlett* (1960)
2. Queues *D.R. Cox and W.L. Smith* (1961)
3. Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)
4. The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)
5. Population Genetics *W.J. Ewens* (1969)
6. Probability, Statistics and Time *M.S. Barlett* (1975)
7. Statistical Inference *S.D. Silvey* (1975)
8. The Analysis of Contingency Tables *B.S. Everitt* (1977)
9. Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)
10. Stochastic Abundance Models *S. Engen* (1978)
11. Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)
12. Point Processes *D.R. Cox and V. Isham* (1980)
13. Identification of Outliers *D.M. Hawkins* (1980)
14. Optimal Design *S.D. Silvey* (1980)
15. Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)
16. Classification *A.D. Gordon* (1981)
17. Distribution-Free Statistical Methods, 2nd edition *J.S. Maritz* (1995)
18. Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)
19. Applications of Queueing Theory, 2nd edition *G.F. Newell* (1982)
20. Risk Theory, 3rd edition *R.E. Beard, T. Pentikäinen and E. Pesonen* (1984)
21. Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)
22. An Introduction to Latent Variable Models *B.S. Everitt* (1984)
23. Bandit Problems *D.A. Berry and B. Fristedt* (1985)
24. Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)
25. The Statistical Analysis of Composition Data *J. Aitchison* (1986)
26. Density Estimation for Statistics and Data Analysis *B.W. Silverman* (1986)
27. Regression Analysis with Applications *G.B. Wetherill* (1986)
28. Sequential Methods in Statistics, 3rd edition *G.B. Wetherill and K.D. Glazebrook* (1986)
29. Tensor Methods in Statistics *P. McCullagh* (1987)
30. Transformation and Weighting in Regression *R.J. Carroll and D. Ruppert* (1988)
31. Asymptotic Techniques for Use in Statistics *O.E. Bandorff-Nielsen and D.R. Cox* (1989)
32. Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)
33. Analysis of Infectious Disease Data *N.G. Becker* (1989)
34. Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)
35. Empirical Bayes Methods, 2nd edition *J.S. Maritz and T. Lwin* (1989)
36. Symmetric Multivariate and Related Distributions *K.T. Fang, S. Kotz and K.W. Ng* (1990)
37. Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)
38. Cyclic and Computer Generated Designs, 2nd edition *J.A. John and E.R. Williams* (1995)
39. Analog Estimation Methods in Econometrics *C.F. Manski* (1988)
40. Subset Selection in Regression *A.J. Miller* (1990)
41. Analysis of Repeated Measures *M.J. Crowder and D.J. Hand* (1990)
42. Statistical Reasoning with Imprecise Probabilities *P. Walley* (1991)
43. Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)
44. Inspection Errors for Attributes in Quality Control *N.L. Johnson, S. Kotz and X. Wu* (1991)
45. The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)
46. The Analysis of Quantal Response Data *B.J.T. Morgan* (1992)
47. Longitudinal Data with Serial Correlation—A State-Space Approach *R.H. Jones* (1993)

48. Differential Geometry and Statistics *M.K. Murray and J.W. Rice* (1993)
49. Markov Models and Optimization *M.H.A. Davis* (1993)
50. Networks and Chaos—Statistical and Probabilistic Aspects
O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (1993)
51. Number-Theoretic Methods in Statistics *K.-T. Fang and Y. Wang* (1994)
52. Inference and Asymptotics *O.E. Barndorff-Nielsen and D.R. Cox* (1994)
53. Practical Risk Theory for Actuaries *C.D. Daykin, T. Pentikäinen and M. Pesonen* (1994)
54. Biplots *J.C. Gower and D.J. Hand* (1996)
55. Predictive Inference—An Introduction *S. Geisser* (1993)
56. Model-Free Curve Estimation *M.E. Tarter and M.D. Lock* (1993)
57. An Introduction to the Bootstrap *B. Efron and R.J. Tibshirani* (1993)
58. Nonparametric Regression and Generalized Linear Models *P.J. Green and B.W. Silverman* (1994)
59. Multidimensional Scaling *T.F. Cox and M.A.A. Cox* (1994)
60. Kernel Smoothing *M.P. Wand and M.C. Jones* (1995)
61. Statistics for Long Memory Processes *J. Beran* (1995)
62. Nonlinear Models for Repeated Measurement Data *M. Davidian and D.M. Giltinan* (1995)
63. Measurement Error in Nonlinear Models *R.J. Carroll, D. Rupert and L.A. Stefanski* (1995)
64. Analyzing and Modeling Rank Data *J.J. Marden* (1995)
65. Time Series Models—In Econometrics, Finance and Other Fields
D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (1996)
66. Local Polynomial Modeling and its Applications *J. Fan and I. Gijbels* (1996)
67. Multivariate Dependencies—Models, Analysis and Interpretation *D.R. Cox and N. Wermuth* (1996)
68. Statistical Inference—Based on the Likelihood *A. Azzalini* (1996)
69. Bayes and Empirical Bayes Methods for Data Analysis *B.P. Carlin and T.A Louis* (1996)
70. Hidden Markov and Other Models for Discrete-Valued Time Series *I.L. MacDonald and W. Zucchini* (1997)
71. Statistical Evidence—A Likelihood Paradigm *R. Royall* (1997)
72. Analysis of Incomplete Multivariate Data *J.L. Schafer* (1997)
73. Multivariate Models and Dependence Concepts *H. Joe* (1997)
74. Theory of Sample Surveys *M.E. Thompson* (1997)
75. Retrial Queues *G. Falin and J.G.C. Templeton* (1997)
76. Theory of Dispersion Models *B. Jørgensen* (1997)
77. Mixed Poisson Processes *J. Grandell* (1997)
78. Variance Components Estimation—Mixed Models, Methodologies and Applications *P.S.R.S. Rao* (1997)
79. Bayesian Methods for Finite Population Sampling *G. Meeden and M. Ghosh* (1997)
80. Stochastic Geometry—Likelihood and computation
O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (1998)
81. Computer-Assisted Analysis of Mixtures and Applications—Meta-Analysis, Disease Mapping and Others
D. Böhning (1999)
82. Classification, 2nd edition *A.D. Gordon* (1999)
83. Semimartingales and their Statistical Inference *B.L.S. Prakasa Rao* (1999)
84. Statistical Aspects of BSE and vCJD—Models for Epidemics *C.A. Donnelly and N.M. Ferguson* (1999)
85. Set-Indexed Martingales *G. Ivanoff and E. Merzbach* (2000)
86. The Theory of the Design of Experiments *D.R. Cox and N. Reid* (2000)
87. Complex Stochastic Systems *O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg* (2001)
88. Multidimensional Scaling, 2nd edition *T.F. Cox and M.A.A. Cox* (2001)
89. Algebraic Statistics—Computational Commutative Algebra in Statistics
G. Pistone, E. Riccomagno and H.P. Wynn (2001)
90. Analysis of Time Series Structure—SSA and Related Techniques
N. Golyandina, V. Nekrutkin and A.A. Zhigljavsky (2001)
91. Subjective Probability Models for Lifetimes *Fabio Spizzichino* (2001)
92. Empirical Likelihood *Art B. Owen* (2001)
93. Statistics in the 21st Century *Adrian E. Raftery, Martin A. Tanner, and Martin T. Wells* (2001)
94. Accelerated Life Models: Modeling and Statistical Analysis
Vilijandas Bagdonavicius and Mikhail Nikulin (2001)

95. Subset Selection in Regression, Second Edition *Alan Miller* (2002)
96. Topics in Modelling of Clustered Data *Marc Aerts, Helena Geys, Geert Molenberghs, and Louise M. Ryan* (2002)
97. Components of Variance *D.R. Cox and P.J. Solomon* (2002)
98. Design and Analysis of Cross-Over Trials, 2nd Edition *Byron Jones and Michael G. Kenward* (2003)
99. Extreme Values in Finance, Telecommunications, and the Environment *Bärbel Finkenstädt and Holger Rootzén* (2003)
100. Statistical Inference and Simulation for Spatial Point Processes *Jesper Møller and Rasmus Plenge Waagepetersen* (2004)
101. Hierarchical Modeling and Analysis for Spatial Data *Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand* (2004)
102. Diagnostic Checks in Time Series *Wai Keung Li* (2004)
103. Stereology for Statisticians *Adrian Baddeley and Eva B. Vedel Jensen* (2004)
104. Gaussian Markov Random Fields: Theory and Applications *Håvard Rue and Leonhard Held* (2005)
105. Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition *Raymond J. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu* (2006)
106. Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood *Youngjo Lee, John A. Nelder, and Yudi Pawitan* (2006)
107. Statistical Methods for Spatio-Temporal Systems *Bärbel Finkenstädt, Leonhard Held, and Valerie Isham* (2007)
108. Nonlinear Time Series: Semiparametric and Nonparametric Methods *Jiti Gao* (2007)
109. Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis *Michael J. Daniels and Joseph W. Hogan* (2008)
110. Hidden Markov Models for Time Series: An Introduction Using R *Walter Zucchini and Iain L. MacDonald* (2009)
111. ROC Curves for Continuous Data *Wojtek J. Krzanowski and David J. Hand* (2009)
112. Antedependence Models for Longitudinal Data *Dale L. Zimmerman and Vicente A. Núñez-Antón* (2009)
113. Mixed Effects Models for Complex Data *Lang Wu* (2010)
114. Intoduction to Time Series Modeling *Genshiro Kitagawa* (2010)
115. Expansions and Asymptotics for Statistics *Christopher G. Small* (2010)
116. Statistical Inference: An Integrated Bayesian/Likelihood Approach *Murray Aitkin* (2010)
117. Circular and Linear Regression: Fitting Circles and Lines by Least Squares *Nikolai Chernov* (2010)
118. Simultaneous Inference in Regression *Wei Liu* (2010)
119. Robust Nonparametric Statistical Methods, Second Edition *Thomas P. Hettmansperger and Joseph W. McKean* (2011)
120. Statistical Inference: The Minimum Distance Approach *Ayanendranath Basu, Hiroyuki Shioya, and Chanseok Park* (2011)
121. Smoothing Splines: Methods and Applications *Yuedong Wang* (2011)
122. Extreme Value Methods with Applications to Finance *Serguei Y. Novak* (2012)
123. Dynamic Prediction in Clinical Survival Analysis *Hans C. van Houwelingen and Hein Putter* (2012)
124. Statistical Methods for Stochastic Differential Equations *Mathieu Kessler, Alexander Lindner, and Michael Sørensen* (2012)
125. Maximum Likelihood Estimation for Sample Surveys *R. L. Chambers, D. G. Steel, Suojin Wang, and A. H. Welsh* (2012)
126. Mean Field Simulation for Monte Carlo Integration *Pierre Del Moral* (2013)
127. Analysis of Variance for Functional Data *Jin-Ting Zhang* (2013)
128. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition *Peter J. Diggle* (2013)
129. Constrained Principal Component Analysis and Related Techniques *Yoshio Takane* (2014)
130. Randomised Response-Adaptive Designs in Clinical Trials *Anthony C. Atkinson and Atanu Biswas* (2014)
131. Theory of Factorial Design: Single- and Multi-Stratum Experiments *Ching-Shui Cheng* (2014)
132. Quasi-Least Squares Regression *Justine Shults and Joseph M. Hilbe* (2014)
133. Data Analysis and Approximate Models: Model Choice, Location-Scale, Analysis of Variance, Nonparametric Regression and Image Analysis *Laurie Davies* (2014)
134. Dependence Modeling with Copulas *Harry Joe* (2014)
135. Hierarchical Modeling and Analysis for Spatial Data, Second Edition *Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand* (2014)

136. Sequential Analysis: Hypothesis Testing and Changepoint Detection *Alexander Tartakovsky, Igor Nikiforov, and Michèle Basseville* (2015)
137. Robust Cluster Analysis and Variable Selection *Gunter Ritter* (2015)
138. Design and Analysis of Cross-Over Trials, Third Edition *Byron Jones and Michael G. Kenward* (2015)
139. Introduction to High-Dimensional Statistics *Christophe Giraud* (2015)
140. Pareto Distributions: Second Edition *Barry C. Arnold* (2015)
141. Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data *Paul Gustafson* (2015)
142. Models for Dependent Time Series *Granville Tunnicliffe Wilson, Marco Reale, John Haywood* (2015)
143. Statistical Learning with Sparsity: The Lasso and Generalizations *Trevor Hastie, Robert Tibshirani, and Martin Wainwright* (2015)
144. Measuring Statistical Evidence Using Relative Belief *Michael Evans* (2015)

Monographs on Statistics and Applied Probability 144

Measuring Statistical Evidence Using Relative Belief

Michael Evans
University of Toronto
Canada

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20150417

International Standard Book Number-13: 978-1-4822-4280-5 (eBook - PDF)

To my wife Rosemary and daughter Heather.

Contents

Preface

The concept of statistical evidence is somewhat elusive. Virtually all approaches to statistical inference refer to the statistical evidence, or the evidence, for something being true or false. But, to our knowledge, no existing theory defines explicitly what this evidence is or at least prescribes how it is to be measured. It is argued here that not to define how to measure evidence is a significant failure for any proposed theory of statistical inference. After all, the purpose of any statistical analysis is to summarize what the statistical evidence is saying about questions of interest. It seems paradoxical that we should hope to do this without being explicit about how to measure statistical evidence.

It is the purpose of this text to provide an overview of recent work on developing a theory of statistical inference that is based on measuring statistical evidence. Of course, one might ask why there is any need to do this beyond perhaps the satisfaction of having a theory that is honest about such a basic concept.

There is a range of approaches to statistical theory from Bayesian theories at one end of the spectrum, to pure likelihood theory and various frequency-based approaches at the other. Many statisticians feel comfortable fitting themselves somewhere along this scale and ignore the failure to adequately deal with statistical evidence. Others even see virtue in adopting different approaches for different problems, as in wearing a Bayesian hat today and a frequentist hat tomorrow. To an extent, these attitudes are based on issues of practicality as, in the end, a practicing statistician has to get on with the business of doing statistical analyses. While this is understandable, this ignores answering the basic question of statistics: what is a correct statistical analysis? The failure to answer this question is a profound and unacceptable gap in the subject of statistics. It almost certainly undermines confidence in the subject and, to an extent, promotes an "almost anything goes" attitude.

Part of the purpose of this book is to show that being explicit about how to measure statistical evidence allows us to answer the basic question of when a statistical analysis is correct. In fact, such a definition prescribes how the inference step should proceed. As one might expect, however, there is more to the story than simply providing a definition. The approach advocated needs to be judged in its entirety. The theory must provide a logical, coherent framework for conducting statistical analyses that can be implemented in problems of practical importance. Furthermore, the theory must be seen to possess properties that add conviction concerning its suitability.

There are some basic issues that underlie many of the controversies and disagreements in statistics. These include things like the meaning of probability, the role of subjectivity, the meaning of objectivity, the role of infinity and continuity and the

relevance of the concept of utility. A fairly strong position is taken in this text on all of these. For example, it is argued that statistics is essentially subjective simply because statisticians always make choices in carrying out a statistical analysis. However, objectivity plays a key role through the data, in assessing the relevance of these choices. In essence, subjectivity can never be avoided but its effects can be assessed and to a certain extent controlled. Undoubtedly, the author's position on subjectivity and objectivity is in disagreement with those who advocate pure subjectivity and with those who believe that there is an objective theory of statistics. It is our contention that inferences derived via relative belief, together with checking the model and prior against the data, place statistics on a much firmer logical foundation with greater relevance for scientific applications.

One caveat needs to be stated for what is being proposed. The developments in this text represent an attempt to establish a gold standard for how a statistical analysis should proceed. A gold standard is something to strive to attain in an application, but, for various reasons, we may fall short. The result of such a failure does not entirely invalidate the analysis but it does suggest that the results have to be suitably qualified. Statisticians are already familiar with making such compromises. Consider the first, and perhaps most important, part of a statistical investigation, namely, the collection of the data. The gold standard here is random sampling from populations and the controlled allocation of values of predictor variables, but this is often not realized. Yet statistical analyses are conducted and useful information is acquired in spite of the deficiencies. Any conclusions drawn, however, must be suitably qualified when there is a failure to attain the highest standard in data collection. The difficulties entailed in guaranteeing that all the necessary ingredients hold for an application of a theory do not justify an attitude that the existence of such a theory is irrelevant.

Chapter 1 discusses some basic features of our overall vision, such as the roles of subjectivity, objectivity, infinity and utility in statistical analyses. In developing a theory of statistical inference, it is necessary to carefully delineate the problems to which the theory is to be applied. As such, the domain of application of the theory is provided, namely, what constitutes a statistical problem and what are the ingredients that a statistician needs to specify to conduct a statistical analysis. In this chapter a simple example is presented of a statistical analysis that satisfies our criteria. Chapter 2 considers the meaning of probability and the various positions taken on probability. This topic lies at the heart of many disagreements in statistics and the author contends that there are a number of reasonable ways to think about probability. Indeed, there is no claim concerning the correctness of one approach to probability over others. The theory of statistical inference presented here is basically independent of these interpretations, although, however probabilities are assigned, they are considered to be measuring belief. Chapter 3 begins the discussion of the heart of the matter, namely, attempts to deal with the concept of statistical evidence. This chapter demonstrates that, while many theories of inference make mention of statistical evidence, they don't adequately define what it is or, more important, how it is to be measured. Furthermore, this failure leads to anomalies for these theories. In Chapter 4 a method is provided for measuring statistical evidence and, based on this method, a theory of inference is developed. This theory is based on the assumption that the

ingredients chosen for a statistical analysis are *correct*. Chapter 5 discusses how a statistician is to go about choosing the ingredients for a statistical problem and how these choices are to be checked for their *correctness* in an application. Of course, the meaning of correct in this context requires considerable discussion. It is a key point in our development that the theory and application of inference are logically separate from the checking phase, and these shouldn't be confounded, as the problems are quite different. For practical applications Chapter 5 should precede Chapter 4 but the focus in this text is on measuring statistical evidence. Chapter 6 summarizes the text and points to further possible developments.

Certainly what is being advocated here is not completely divorced from what has been discussed by others. In fact, the author would describe the essence of the relative belief approach to statistical inference as simply being careful about the definition and usage of the Bayes factor. Furthermore, there are many similarities with pure likelihood theory, and relative belief could be described as filling in, from the author's point of view, the logical gaps in that theory. Also, the frequentist approach plays a key role, not in determining inferences, but in checking the suitability of the ingredients as well as providing optimality properties of the inferences.

The author is solely responsible for all errors and omissions. Thanks are owed to many. In particular, Luai Al-Labadi, Zeynep Baskurt, Shelly Cao, Zvi Gilula, Irwin Guttman, Gun Ho Jang, Shaocheng Liu, Hadas Moshonov, Saman Muthukumarana, Mohammed Shakhatreh, Tim Swartz and Tianli Zou were all co-authors on publications connected with the contents of this text and made key contributions. Irwin Guttman introduced me to Bayesian inference and this has had a major influence. Gun Ho Jang contributed numerous ingenious solutions to technical problems. Many readers of the manuscript provided valuable input, including students Stephen Marsh and Yang Guan Jian Guo. Keith O'Rourke made many useful suggestions. The reviewers also provided valuable feedback and the author would especially like to thank Jay Kadane and David Nott. My interest in problems concerned with the foundations of statistical inference was stimulated by Professor D. A. S. Fraser and I am grateful for that and also for instilling in me the belief that these problems are resolvable.

Chapter 1

Statistical Problems

1.1 Introduction

This book is about measuring statistical evidence. More precisely, a definition of statistical evidence is proposed based on the ingredients of a statistical problem as specified by the statistician. A direct consequence of this definition is a theory of statistical inference that has some unique and appealing features.

It may come as a surprise to the lay reader that exactly how one is to measure statistical evidence is not well-resolved in the scientific literature. Most reasonably numerate individuals have encountered the notions of p-values, standard errors, etc., and understand that these concepts are central to how to reason in statistical problems and that they have something to do with characterizing statistical evidence. Yet it is a fact that experienced, professional statisticians can disagree quite dramatically about the right way to reason in statistical contexts.

On examining the various approaches to inference, it will be seen that they commonly fail to precisely define what statistical evidence is. This can be regarded as a significant omission. In fact, it is our view that any valid theory of statistical inference must specify exactly what is meant by statistical evidence. A definition of statistical evidence should serve as a core of the theory of statistical inference and basically dictate how statistical problems are to be solved, namely, the statistical evidence should tell us what the solution to a problem is.

Our proposal for a definition of statistical evidence is provided in Chapter 4. Any such definition is based upon the ingredients of a statistical problem as specified by a statistician. So the current chapter is concerned with discussing exactly what is meant by a statistical problem and what ingredients need to be specified by the statistician. This leads us to exclude certain problems and ingredients that others might prefer to be included. Our defence for this is that our approach covers the vast majority of practically meaningful statistical problems and that by being exclusionary, a lot of unnecessary complexity and ambiguity is eliminated. Above all, our goal is a logical and complete theory of statistical inference that has practical relevance rather than some kind of mathematical generality. In fact, our view is that attempts at mathematical generality often mislead as to what is appropriate statistical reasoning.

Probability is a key concept in any theory of statistical inference. This is of such importance that the entirety of Chapter 2 is devoted to this topic.

1.2 Statistical Problems

The first question to be answered is: what is a statistical problem? The following example characterizes what could be called the archetypal statistical problem. The discussion in this text is restricted to the consideration of such problems and close relatives. We argue that such restrictions are necessary and moreover apply in the vast majority of applications.

Example 1.2.1 *The Archetypal Statistical Problem.*
Suppose there is a *population* Ω with $\#(\Omega) < \infty$, where $\#(A)$ denotes the cardinality of the set A. So Ω is just a finite set of objects. Further suppose that there is a *measurement* $X : \Omega \to \mathscr{X}$. A measurement X is a function defined on Ω taking values in the set $\mathscr{X}$. So $X(\omega) \in \mathscr{X}$ is the measurement of object $\omega \in \Omega$. For example, Ω could be the set of all students enrolled at a particular school and $X(\omega)$ the height in centimeters of student ω. So, in this case, $\mathscr{X}$ is a subset of R^1. As another example, Ω could be the set of all students enrolled at a particular school and $X(\omega) = (X_1(\omega), X_2(\omega))$, where $X_1(\omega)$ is the height in centimeters of student ω and $X_2(\omega)$ is the gender of student ω, and so, in this case, $\mathscr{X}$ can be taken to be a subset of $R^1 \times \{M, F\}$.

When considering a variable like height, it is common to treat this as possibly taking on a continuous range of values and to allow the set of possible values to be unbounded. While this may seem innocuous, as argued in Section 1.4, we need to be careful when using infinities in discussing statistical problems. As such, because of the finiteness of Ω, the finite accuracy with which our height measurements are made, and the fact that the measurements will occur within known bounds, $\mathscr{X}$ can be taken to be the set of all the possible values of $X(\omega)$ and this is a finite set. Infinite sets are introduced in Section 1.4 when considering approximations to statistical problems.

The fundamental object of interest in a statistical problem is then the *relative frequency distribution* of X over Ω. For a subset $A \subset \mathscr{X}$ the relative frequency of A is given by

$$r_X(A) = \frac{\#(\{\omega : X(\omega) \in A\})}{\#(\Omega)}.$$

So $r_X(A)$ is just the proportion of elements in Ω whose X measurement is in A. Clearly, knowing the relative frequency distribution is equivalent to knowing the *relative frequency function*

$$f_X(x) = \frac{\#(\{\omega : X(\omega) = x\})}{\#(\Omega)},$$

for $x \in \mathscr{X}$, as one can be obtained from the other. Notice that the frequency distribution is defined no matter what the set $\mathscr{X}$ is.

If we can conduct a *census*, where we obtain $X(\omega)$ for each $\omega \in \Omega$, then f_X is known exactly and there is nothing left to do from a statistical perspective. Of course, statistics exists as a subject precisely because it is, at the very least, generally uneconomical to conduct a census, and typically it is impossible to do so. Sometimes statistical problems are expressed as wanting to know about some aspect of f_X rather

than f_X itself. For example, if X is real-valued one may be interested to know the *mean*

$$\mu_X = \sum_{x \in \mathscr{X}} x f_X(x)$$

and *variance*

$$\sigma_X^2 = \sum_{x \in \mathscr{X}} (x - \mu_X)^2 f_X(x)$$

of the relative frequency distribution or perhaps some other characteristic of f_X. Certainly there is nothing wrong with focusing on some characteristics of f_X, but knowing f_X represents full statistical information. As such, we will express our discussion in terms of knowing the true f_X.

The fundamental question of statistics is then, based on partial information about the true f_X, how do we make inferences about the true f_X? Of course, it has to be made clear what is meant by partial information, and subsequent sections will do this, but as part of any statistical problem there is the *observed data* $x_1 = X(\omega_1), \ldots, x_n = X(\omega_n)$ where $\{\omega_1, \ldots, \omega_n\} \subset \Omega$ is selected, in some fashion, from the population. ■

Note that in Example 1.2.1 there are no infinities and everything is defined simply in terms of counting. Also there is no mention of probability. There is, however, a major *uncertainty* in that f_X is unknown without conducting a census. This is the fundamental uncertainty lying at the heart of all of statistical problems.

Undoubtedly Example 1.2.1 seems very restrictive but there are a number of ways in which it can be generalized without violating its basic characteristic of everything being finite and obtainable via counting. For example, we can consider several finite populations $\Omega_1, \ldots, \Omega_m$ with respective measurements $X_1, \ldots, X_m$ and relative frequency functions $f_{X_1}, \ldots, f_{X_m}$ and then discuss making comparisons among them. Also, the relation of so-called *measurement error* problems to Example 1.2.1 and the use of infinities and continuity in statistical modeling are examined in Section 1.4.

Perhaps most important, problems where the interest is with relationships among variables arise as generalizations of Example 1.2.1. The concept of relationship between variables is based on the concept of a *conditional relative frequency distribution*. Suppose there are two measurements X and Y defined on a population Ω with $(X(\omega), Y(\omega)) \in \mathscr{X}$. The *conditional relative frequency function* of Y given $X = x$ is then defined for $(x, y) \in \mathscr{X}$ by

$$f_{Y|X}(y|x) = \frac{\#(\{\omega : X(\omega) = x, Y(\omega) = y\})}{\#(\{\omega : X(\omega) = x\})} = \frac{f_{(X,Y)}(x,y)}{f_X(x)}$$

whenever $\#(\{\omega : X(\omega) = x\}) \neq 0$ or, equivalently, $f_X(x) \neq 0$. Notice that $f_{Y|X}$ is again obtained by simple counting and that it can be obtained from the *joint relative frequency function* $f_{(X,Y)}$ and the *marginal relative frequency function* f_X. The following makes use of conditional relative frequency and represents the most important application of statistics.

Example 1.2.2 *Relationships among Variables.*
Suppose there is a measurement $(X,Y):\Omega\to\mathscr{X}$ and our interest is in whether or not there is a *relationship* between the variables X and Y. There is a basic definition of what it means for variables to be related.

Definition 1.2.1 *Variables X and Y, defined on population Ω, are related variables if for some x_1,x_2 such that $f_X(x_1)\neq 0$ and $f_X(x_2)\neq 0$, then $f_{Y|X}(\cdot\,|x_1)\neq f_{Y|X}(\cdot\,|x_2)$.*

So two variables are related whenever changing the conditioning variable can result in a change in the conditional distribution of the other variable. Note that it is clearly the case that there is no relationship when all the conditional distributions are the same; in fact, this is equivalent to the *statistical independence* of the variables. In general, the *form of the relationship* is given by how $f_{Y|X}(\cdot\,|x)$ changes as x changes. From a practical viewpoint, formally at least, most variables are related by this definition, as it seems unlikely that these conditional distributions will always be the same. But the relationship between X and Y can be very weak and the changes deemed to be irrelevant for the application at hand.

It is common in statistical applications for various assumptions to be made about the form of $f_{Y|X}(\cdot\,|x)$. For example, a *regression assumption* says that, for a real-valued Y, at most the *conditional means*

$$\mu_Y(x)=\sum_{y\in\{Y(\omega):X(\omega)=x\}} y f_{Y|X}(y\,|x))$$

are changing as we change x. Often a *linear regression assumption* is also made where it is assumed that μ_Y is in some finite linear span of functions of x, namely, $\mu_Y\in L\{v_1,\ldots,v_k\}$ where the v_i are real-valued functions defined on $\{X(\omega):\omega\in\Omega\}$. Of course, these are assumptions and the methods of Chapter 5 are needed to see if these makes sense in an application. Actually, the regression assumptions make the most sense when Y is allowed to take on a continuous range of values, as discussed in Section 1.4. ■

Although we will often express concepts in terms of the archetypal statistical problem, it will be seen that these apply much more generally. Section 1.3 is particularly relevant in this regard.

1.3 Statistical Models

The fundamental problem in statistics arises because a census cannot be conducted and so relative frequency distributions such as f_X in Example 1.2.1 cannot be known exactly. Note that, because Ω is finite and because each component of X is bounded and measured to finite accuracy, there are only finitely many possibilities for f_X. For example, if $\#(\Omega)=10^4, X(\omega)$ is height recorded in centimeters and all heights are in $(0,300]$, then f_X is in a finite set of cardinality considerably less than $10^4\times 300$. The important point here is that the set of possibilities for f_X is finite.

In many statistical problems, the statistician is willing to *assume* that f_X is in a restricted set of possible relative frequency functions. We index these possible functions by a variable θ, called the *model parameter,* taking values in a set Θ,

called the *model parameter space*, to obtain the *statistical model* $\{f_\theta : \theta \in \Theta\}$. So $f_X \in \{f_\theta : \theta \in \Theta\}$ and note that, at this point, Θ is finite. Accordingly, instead of the true relative frequency function, we can speak equivalently about the *true value* of the model parameter since by assumption there is a unique value $\theta_{true} \in \Theta$ such that $f_{\theta_{true}} = f_X$.

More generally, we can define a *parameter* ψ *of the model* as $\psi = \Psi(\theta)$ where $\Psi : \Theta \overset{\text{onto}}{\to} \Psi$, and for convenience we use the Ψ symbol for both the function and its range. So there is a true value for ψ given by $\psi_{true} = \Psi(\theta_{true})$. In general, we want to make inferences about a *parameter of interest* ψ (which could be θ if we take Ψ to be the identity) and refer to all aspects of θ that distinguish values in $\Psi^{-1}\{\psi\}$ as *nuisance parameters*. We provide a simple example.

Example 1.3.1

Consider a population Ω of eligible voters where $\#(\Omega) = 20,000$ and $X(\omega) = 1$, if ω will vote in the next election, and $X(\omega) = 0$, otherwise. So there are in total exactly 20,001 different possibilities for f_X. Suppose further, however, that it is known that at least 5,000 and no more than 15,000 voters will indeed vote, where this information is based on historical records of elections. So, noting that $1/20,000 = 5 \times 10^{-5}$, the statistical model $\{f_\theta : \theta \in \Theta\}$ is given by $\theta \in \Theta = \{0.25000, 0.25005, 0.25010, \ldots, 0.75000\}$ where $f_\theta(1) = \theta$ and $f_\theta(0) = 1 - \theta$.

Rather than the model parameter θ, one might be interested in the odds parameter $\psi = \Psi(\theta) = \theta/(1-\theta)$ where the range is $\{0.25/(1-0.25), 0.25005/(1-0.25005), \ldots, 0.75/(1-0.75)\}$. In this case Ψ is 1 to 1, so there are no nuisance parameters. ■

Many attempts at developing theories of inference run into problems when considering inferences for an arbitrary $\psi = \Psi(\theta)$. This is referred to as the *nuisance parameter problem*.

An important point to note about a statistical model is that, unless it includes all the possible distributions, $\{f_\theta : \theta \in \Theta\}$ is an assumption and as such could be incorrect because $f_X \notin \{f_\theta : \theta \in \Theta\}$. As you might expect, when statistical analysis is based on incorrect assumptions, then we have to question the validity of the analysis. If so, why not always take $\{f_\theta : \theta \in \Theta\}$ to be the set of all possible distributions? Certainly in Example 1.3.1 it seems simple to avoid any assumptions.

There are several reasons why model assumptions are made. First and foremost is that there may be definite information about the form of f_X and this information leads to improved inferences when true. Second, and perhaps most common, is that for very complicated situations, model assumptions are made to simplify the analysis and it is felt that the error introduced by these simplifications will not have a material effect on the inferences drawn. For example, in Example 1.2.2 it seems very unlikely that the regression assumption ever holds exactly. But perhaps the deviation from this assumption is so small that it is immaterial, while the added simplicity of only looking at the conditional mean to examine the relationship is of great benefit.

There is the possibility, however, that the model $\{f_\theta : \theta \in \Theta\}$ could be grossly in error. So it is necessary to consider how to assess and deal with this as part of a

statistical analysis. This is discussed in part in Section 1.5 and is more thoroughly treated in Chapter 5.

1.4 Infinity and Continuity in Statistics

So far all the sets introduced have been finite. It is our belief that this finiteness holds in any practical application of statistics. Infinite sets can be used as part of a simplifying approximation but with the awareness that this can bring with it problems of interpretation that can lead us astray when we are not careful. Consider the following example.

Example 1.4.1 *Likelihood Functions.*
A probability density f_θ gives rise to a probability measure P_θ on $\mathscr{X}$ via integration of f_θ over relevant sets. Suppose it is assumed that data $x \in \mathscr{X}$ has been generated from one of the probability distributions in the model $\{f_\theta : \theta \in \Theta\}$. Now we wish to make an inference about the true $\theta \in \Theta$. In such a situation, methods based upon the likelihood function are commonly recommended.

Definition 1.4.1 *For observed data x and model $\{f_\theta : \theta \in \Theta\}$, the likelihood function is defined to be the function $L(\cdot\,|\,x) : \Theta \to [0,\infty)$ given by $L(\theta\,|\,x) = kf_\theta(x)$ for any fixed $k > 0$.*

In reality the likelihood function is an equivalence class of functions, as the constant k is arbitrary and can be chosen for convenience. This indeterminacy causes no problems because likelihood inferences only depend on the ratios of likelihood values.

The motivation behind considering the likelihood function lies in saying that θ_1 is at least as preferable (as a guess or inference about the true value of θ) as θ_2 whenever the *likelihood ratio* $L(\theta_1\,|\,x)/L(\theta_2\,|\,x) \geq 1$. This imposes a complete preference ordering on Θ. This *likelihood preference ordering* is natural when each distribution is discrete because

$$\frac{L(\theta_1\,|\,x)}{L(\theta_2\,|\,x)} = \frac{kf_{\theta_1}(x)}{kf_{\theta_2}(x)} = \frac{P_{\theta_1}(\{x\})}{P_{\theta_2}(\{x\})}$$

is the ratio of the probability of observing x when θ_1 is true to the probability of observing x when θ_2 is true. Given that we have observed x, it is natural to prefer those θ values which give a higher probability to the observed data.

For the situation where continuous probability distributions are employed, let us suppose, for the moment, that $\mathscr{X}$ is Euclidean, $N_\varepsilon(x)$ is an open ball about x of radius ε and f_θ is continuous and positive at x for each θ. Letting $\mathrm{Vol}(A)$ denote the Euclidean volume of $A \subset \mathscr{X}$, we have that

$$P_\theta(N_\varepsilon(x)) = \int_{N_\varepsilon(x)} f_\theta(z)\,dz \sim f_\theta(x)\mathrm{Vol}(N_\varepsilon(x))$$

as $\varepsilon \to 0$, since $P_\theta(N_\varepsilon(x))/f_\theta(x)\mathrm{Vol}(N_\varepsilon(x)) \to 1$ as $\varepsilon \to 0$. So for small ε

$$\frac{L(\theta_1\,|\,x)}{L(\theta_2\,|\,x)} = \frac{kf_{\theta_1}(x)}{kf_{\theta_2}(x)} \approx \frac{P_{\theta_1}(N_\varepsilon(x))}{P_{\theta_2}(N_\varepsilon(x))}$$

and again the likelihood ratio can be seen as comparing the probabilities of observing x. As such, the likelihood preference ordering makes sense in the continuous context too.

But notice a key assumption in this argument, namely, that f_θ is continuous at x for each θ. If g_θ is another integrable function that differs from f_θ at most on a set of volume measure 0, then $P_\theta(A) = \int_A f_\theta(z)\,dz = \int_A g_\theta(z)\,dz$ for every Borel set A. So g_θ could just as easily serve as a density for P_θ. As is well known, f_θ can be modified at countably many points to obtain a valid density g_θ and it is not necessary to use a density that is continuous at each x, at least for the computation of probabilities.

Now this anomaly may be considered a minor irritation, but consider a practical context. In such a situation the value of X is measured to a finite accuracy and as such every coordinate in x is a rational number. The set of $x \in \mathscr{X}$ with rational coordinates is necessarily countable and so a density g_θ for P_θ can be chosen such that $g_\theta(z) = 0$ (or some other constant) for every z with rational coordinates. Therefore, if we use such a g_θ in the definition of the likelihood function, for any actually observed data x the likelihood is identically 0 and the likelihood preference ordering doesn't distinguish among the θ. Clearly this is absurd, unless you don't believe in the relevance of the likelihood preference ordering to inference. ■

One way out of the dilemma posed with continuous models in Example 1.4.1 is to simply demand that the densities in the definition of the likelihood be continuous at each $x \in \mathscr{X}$. But what aspect of an application implies such a restriction? For us this restriction is imposed by the fact that all sets in a statistical application are finite and, when an infinite set is used, this is as an approximation to a finite object. If this approximation aspect is ignored, then absurdities can arise as in the discussion in Example 1.4.1. If $f_\theta(x)$ can be arbitrarily defined on a set of measure 0, as is certainly mathematically acceptable when considering densities just as mathematical objects, then the notion of an approximation is lost.

Various treatments of statistical theory treat infinite sets as basic ingredients that represent reality. For the developments here, however, while we want to make use of the simplicities available with infinite sets, conditions must be placed on such objects to ensure that they behave appropriately as approximations to entities that are in fact finite. As such, it is required that a density f_θ be defined as a limit. For example, in Example 1.4.1, for each $x \in \mathscr{X}$, it is required that

$$f_\theta(x) = \lim_{\varepsilon \to 0} \frac{P_\theta(N_\varepsilon(x))}{\mathrm{Vol}(N_\varepsilon(x))} \tag{1.1}$$

as this ensures that $P_\theta(N_\varepsilon(x)) \approx f_\theta(x)\mathrm{Vol}(N_\varepsilon(x))$ for small ε. The definition of densities is discussed more generally in Appendix A but it is noted that, if a version of f_θ exists that is continuous at x, then it is given by (1.1). This leads to the usual representative densities and in general the density calculated by differentiating a distribution function will satisfy (1.1). Any time density is used it is assumed that it is a density with respect to the volume measure on the respective space (counting measure is volume measure on discrete sets) and that the density arises as a limit as in (1.1); see Appendix A.