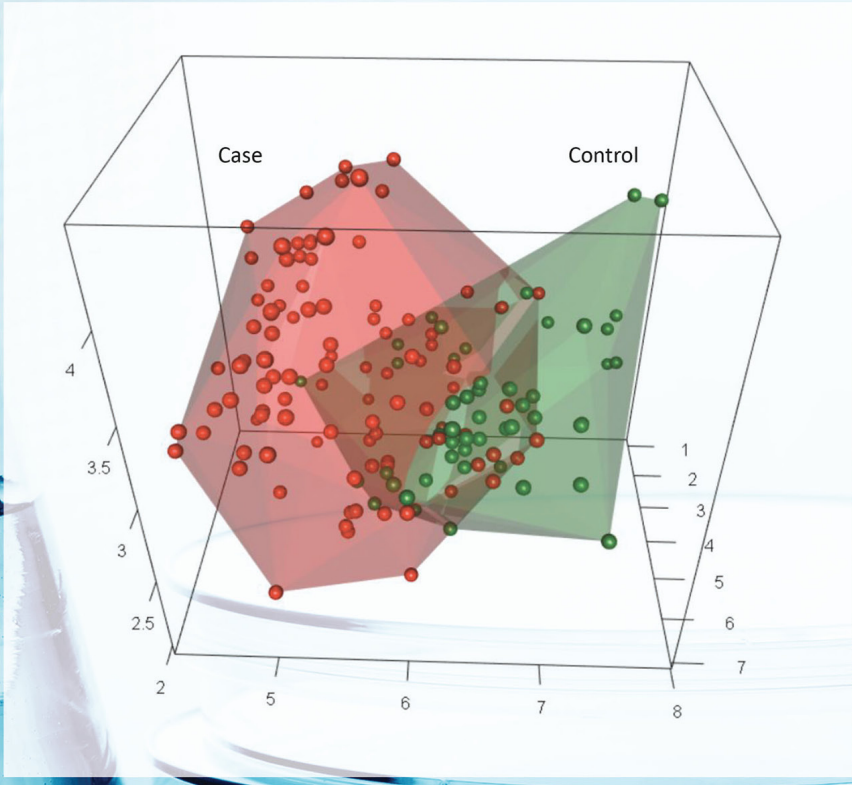
# EMPIRICAL LIKELIHOOD METHODS IN BIOMEDICINE AND HEALTH

Albert Vexler
Jihnhee Yu




CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

# Empirical Likelihood Methods in Biomedicine and Health

# Empirical Likelihood Methods in Biomedicine and Health

Albert Vexler
Jihnhee Yu


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



Printed on acid-free paper

International Standard Book Number-13: 978-1-4665-5503-7 (Hardback)



**Library of Congress Cataloging-in-Publication Data**

Names: Vexler, Albert, author. | Yu, Jihnhee, author.
Title: Empirical likelihood methods in biomedicine and health / by Albert Vexler, The State University of New York at Buffalo, USA and Jihnhee Yu, The State University of New York at Buffalo, USA.
Description: Boca Raton, Florida : CRC Press, [2018] | Includes bibliographical references and index.
Identifiers: LCCN 2017060382| ISBN 9781466555037 (hardback : alk. paper) | ISBN 9781351001526 (e-book)
Subjects: LCSH: Medical informatics--Research--Methodology. | Bioethics--Research--Methodology.
Classification: LCC R858 .V49 2018 | DDC 610.285072--dc23
LC record available at https://lccn.loc.gov/2017060382

**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**and the CRC Press Web site at**
**http://www.crcpress.com**

*To my parents, Octyabrina and Alexander, and my son, David*

**Albert Vexler**

*To Joonyeong*

**Jihnhee Yu**

# Contents

# *Preface*

Empirical likelihood (EL) is a nonparametric likelihood approach that has been used frequently in recent statistical tool developments. The method tends to be more robust than purely parametric approaches and demonstrates its applicability in many data analytical problems. As distributions of data in the real world are commonly unknown, data-driven approaches such as the EL method should be more competitive than purely parametric approaches, given the lack of the knowledge of true distributions. As the EL methods are often comparatively efficient when compared to other existing approaches (such as *t*-test-based schemes) even with normal underlying distribution cases, more active use of the methods seems warranted. However, the method may be unfamiliar to some statistical researchers and potential end-users for data analysis, and thus it is difficult to find applications of the method in publications related to practical areas such as medicine and epidemiological research. For a more active use of the method, researchers need to convince the utility, accuracy, and efficiency of the method. We hope that this book will be truly successful toward that endeavor.

This book can be used as a textbook for a one or two-semesters advanced graduate course. The material in the book can be appropriate for use both as a text and as a reference. We hope that the mathematical level and breadth of examples will recruit students and teachers not only from statistics and biostatistics, but from a broad range of fields. We hope this book to be a connecting dot that leads interested readers to some technical details of subject areas more easily. The authors of this book have been working on the topics of the EL approaches to tackle problems related to clinical and epidemiological data. We believe that the research areas of EL are rich in yet-to-be-found applications and theoretical developments on many statistical problems and often those findings could provide better solutions than existing approaches. However, the concept of the EL may be foreign to people who do not have exposures to the approach and that fact would make new researchers hesitate considering EL for tool developments. In this regard, through this book, readers may be familiar with our developmental scheme of EL approach. Especially, Chapters 3 through 8 contain subject areas that the authors heavily worked on, and their contents will provide analytical issues and motivational questions, theoretical developments, software implementation, brief simulation results, and data applications, which pretty much sum up our procedures to develop *new* EL methods. For the theoretical developments in those chapters, readers will find recurring formal patterns almost similar to a "ceremony." In that patterns of developments, we tried to provide enough details of theoretical statements that cater the need of prospective EL method developers.

Chapter 1 offers the overview of statistical hypothesis tests and rational of using the EL approach. This chapter addresses the benefit of using the likelihood approach in details including the principal idea of the Neyman–Pearson Lemma, likelihood ratio tests, and maximum likelihood. Then the EL is introduced as a data-driven likelihood function that is nonparametric and comparatively powerful. This chapter further discusses the EL's benefits such as constructing efficient statistical tests using Bayesian methods in a similar manner to the parametric likelihood and setting up the EL statistics as composite semi- or nonparametric likelihoods. Chapter 2 focuses on the performance of EL constructs relative to ordinary parametric likelihood ratio-based procedures in the context of clinical experiments. This chapter first offers an overview of the classical EL methods. It explains the similarity between EL functions and parametric likelihood functions, detailed expressions of the Lagrange multipliers used in EL statements up to the fourth derivatives, and asymptotic properties of the EL likelihood functions. The chapter also touches the topics of extra estimating equation information, density-based EL methods, building composite hypotheses tests, Bayesian approaches, Bartlett correction, interpretation of the EL as an empirical goodness-of-fit test, and some comparison with bootstrap methods. Chapter 3 discusses how to incorporate EL in the Bayesian framework by showing a novel approach for developing the nonparametric Bayesian posterior expectation, the nonparametric analog of James–Stein estimation, and the nonparametric Bayesian confidence interval estimation. The chapter explains posterior expectations of general functionals. Chapter 4 discusses a general scheme to extend the conventional EL inference, considering the probability weighted moments (PWMs). The main task consists of forming constraints relevant to PWMs and showing that the developed EL test follows the classical asymptotic theories. The statistical test and confidence interval estimation of the PWMs are derived based on the proposed asymptotic proposition. Chapter 5 discusses methods to combine likelihood functions in parametric or empirical form in the setting of two-group comparison. It demonstrates an inference on incomplete bivariate data using a method that combines the parametric model and ELs. This chapter starts with discussions of two-group comparison of means where the EL ratio (ELR) test statistic carries out the mean-specific comparisons unlike other available nonparametric tests. It discusses comparison of multivariate means as a simple extension of univariate two-group comparison. Then, the likelihood ratio test based on the combined likelihood for the incomplete and complete data is developed to compare two treatment groups. Chapter 6 discusses the quantile estimation using the EL in the settings of testing one group and two groups. The Bahadur representation of the maximum EL estimator (MELE) of the quantile function is presented. Testing methods consist of the conventional EL method and the plug-in method. Chapter 7 discusses the ELR test with the constraints in the form of U-statistics. It first provides a general explanation of U-statistics including the variance estimation. Then, it discusses the EL test statistic with

U-statistic type constraints. The chapter discusses EL approaches for univariate and multivariate one-group and two-group U-statistics and provides some suitable examples including a multivariate rank statistic and an application to crossover designs. Chapter 8 starts with the general introduction of the receiver operating characteristic (ROC) curves and then discusses the construction of the EL statistic for the nonparametric estimator of the whole or partial area under the ROC curve (AUC) that has a form of the U-statistic. It discusses the best combinations of multiple biomarkers using ROC curve analysis. An important task of constructing the EL statistic is to incorporate the correct variance estimate to the EL statistic as discussed in Chapter 7. The problem is that the typical variance formula for U-statistics is inaccurate to estimate the variability of the estimator as plug-in estimators of the quantiles used. In this context, the chapter provides details of a correct variance estimation strategy. Finally, as an introductory manner, Chapter 9 presents several interesting topics that are discussed in the EL literature. This overview will demonstrate that the EL approach has a flexibility to be applied to various topics of interest as far as users can formulate a statistical question in a form of the estimating equations. Discussions of regression methods include incorporating validation data with error-prone covariates, analyzing longitudinal data, handling incomplete membership information, and regression with surrogate covariates. Discussions of censored data analyses include testing hazard functions, quantile function estimation, testing mean survival times, analyzing mean quality-adjusted lifetime (QAL) with censored data, and regression approach with censored data. Discussions of missing data include imputation, methods incorporating missing probabilities, and handling missing covariates. The chapter concludes introducing a pseudo-EL approach in survey sampling. The chapter provides some details in terms of describing analytical issues, building the constraints and relevant inferential results.

When we refer the Appendix in each chapter, it indicates the Appendix at the end of that chapter. In this book, we provide R codes that are readily usable and probably just enough to carry out the task we explained. We note that the software code mentioned in this book certainly can be improved, optimized, and extended.

As the statistical methodology has been continuously developed to tackle various data analytical issues, it is hard to cover all new developments; nevertheless, we hope that this book is a helpful introduction to show versatility and applicability of the EL method.

# *Authors*

**Albert Vexler** obtained his PhD degree in Statistics and Probability Theory from the Hebrew University of Jerusalem, Israel, in 2003. His PhD advisor was Moshe Pollak, a fellow of the *American Statistical Association*, and Marcy Bogen Professor of Statistics at Hebrew University. Dr. Vexler was a postdoctoral research fellow in the Biometry and Mathematical Statistics Branch at the National Institute of Child Health and Human Development (National Institutes of Health, USA). Currently, Dr. Vexler is a tenured full professor at the State University of New York at Buffalo, Department of Biostatistics. Dr. Vexler has authored and coauthored various publications that contribute to both the theoretical and applied aspects of statistics in medical research. Many of his papers and statistical software developments have appeared in statistical/biostatistical journals, which have the top-rated impact factors and are historically recognized as the leading scientific journals, and include:
*Biometrics, Biometrika, Journal of Statistical Software, The American Statistician, The Annals of Applied Statistics, Statistical Methods in Medical Research, Biostatistics, Journal of Computational Biology, Statistics in Medicine, Statistics and Computing, Computational Statistics and Data Analysis, Scandinavian Journal of Statistics, Biometrical Journal, Statistics in Biopharmaceutical Research, Stochastic Processes and their Applications, Journal of Statistical Planning and Inference, Annals of the Institute of Statistical Mathematics, The Canadian Journal of Statistics, Metrika, Statistics, Journal of Applied Statistics, Journal of Nonparametric Statistics, Communications in Statistics, Sequential Analysis, The STATA journal, American Journal of Epidemiology, Epidemiology, Paediatric and Perinatal Epidemiology, Academic Radiology, The Journal of Clinical Endocrinology & Metabolism, Journal of Addiction Medicine, Reproductive Toxicology,* and *Human Reproduction*.

Dr. Vexler was awarded National Institutes of Health (NIH) grants to develop novel nonparametric data analysis and statistical methodology. Dr. Vexler's research interests are related to the following subjects: receiver operating characteristic curves analysis, measurement error, optimal designs, regression models, censored data, change point problems, sequential analysis, statistical epidemiology, Bayesian decision-making mechanisms, asymptotic methods of statistics, forecasting, sampling, optimal testing, nonparametric tests, empirical likelihoods, renewal theory, Tauberian theorems, time series, categorical analysis, multivariate analysis, multivariate testing of complex hypotheses, factor and principal component analysis, statistical biomarkers evaluations, and best combinations of biomarkers. Dr. Vexler is associate editor for *Biometrics* and the *Journal of Applied Statistics*. These journals belong to the first cohort of academic

literature related to the methodology of biostatistical and epidemiological research and clinical trials.

**Dr. Jihnhee Yu** received her PhD degree in Statistics from Texas A&M University, College Station, Texas, in 2003. Her PhD advisor was Thomas E. Wehrly, PhD. Currently, Dr. Yu is a tenured associate professor at the State University of New York at Buffalo, New York, Department of Biostatistics. Also, Dr. Yu is the director of the Population Health Observatory, School of Public Health, and Health Professions at the State University of New York at Buffalo. Before her university carrier, she was a senior biostatistics consultant at Roswell Park Cancer Institute, Buffalo, New York. Dr. Yu has been working on cancer study designs; prospective clinical trials; retrospective data analysis; secondary data analyses; epidemiological studies; relevant statistical methodology development, and data analysis in epidemiology, psychology, pediatrics, oral biology, neurosurgery, and health behavior, and other population health- and medical-related areas. Her theoretical research interests are nonparametric tests and clinical trial methodologies. In particular, she has worked on group sequential designs, unbiased estimators associated with group sequential designs, exact methods, and few empirical likelihood approaches on several topics. She authored and coauthored many statistical and collaborative papers that were published in top-ranking journals. Dr. Yu was awarded National Institutes of Health (NIH) grants to develop novel nonparametric data analysis and statistical methodology. Dr. Yu is an associate editor for *Journal of Applied Statistics.*

# 1

# *Preliminaries*

## 1.1 Overview: From Statistical Hypotheses to Types of Information for Constructing Statistical Tests

Most experiments in biomedicine and other health-related sciences involve mathematically formalized comparisons, employing appropriate and efficient statistical procedures in designing clinical studies and analyzing data. Decision making through formal rules based on mathematical strategies plays important roles in medical and epidemiological discovery, policy formulation, and clinical practice. In this context, the statistical discipline is commonly required to be applied to make conclusions about populations on the basis of samples from the populations.

The aim of methodologies in decision making is to maximize quantified gains and at the same time minimize losses to reach a conclusion. For example, statements of clinical experiments can target gains such as accuracy of diagnosis of medical conditions, faster healing, and greater patient satisfaction, while they minimize losses such as efforts, durations of screening for disease, and side effects and costs of the experiments.

There are generally many constraints and desirable characteristics for constructing a statistical test. An essential part of the test development is that statistical hypotheses should be clearly and formally set up with respect to objectives of clinical studies. Oftentimes, statistical hypotheses and clinical hypotheses are associated but stated in different forms and orders. In most applications, we are interested in testing characteristics or distributions of one or more populations. In such cases, the statistical hypotheses must be carefully formulated, and formally stated, depicting, e.g., the nature of associations in terms of quantified characteristics or distributions of populations. The term *Null Hypothesis*, symbolized $H_0$, commonly is used to point out our primary statistical hypothesis. For example, when one wants to test that a biomarker of oxidative stress has different circulating levels for patients with and without atherosclerosis (clinical hypothesis), the null hypothesis (statistical hypothesis) can be proposed corresponding to the *assumption* that levels of the biomarker in individuals with and without atherosclerosis

are distributed equally. Note that the clinical hypothesis points out that we want to show the discriminating power of the biomarker, whereas $H_0$ says there are no significant associations between the disease and biomarker's levels. The reason of such null hypothesis specification lies in the ability to formulate $H_0$ clearly and unambiguously as well as to measure and calculate expected errors in decision making. Probably, if the null hypothesis would be formed in a similar manner to the clinical hypothesis, we could not unambiguously determine which sort of links between the disease and biomarker's levels should be tested.

The null hypothesis is usually a statement to be statistically tested. In the context of statistical testing that provides a formal test procedure and compares mathematical strategies to make a decision, algorithms for monitoring statistical test characteristics associated with the probability to reject a correct hypothesis should be considered. While developing and applying test procedures, the practical statistician faces a task to control the probability of the event that a test outcome rejects $H_0$ when in fact $H_0$ is correct, called a *Type I error* (TIE) rate.

Obviously, in order to construct statistical tests, we must review the corresponding clinical study, formalizing objectives of the experiments and making assumptions in hypothesis testing. A violation of the assumptions can pose incorrect results of the test and a vital malfunction of the TIE rate control procedure. Moreover, should the user verify that the assumptions are satisfied, errors of the verifications itself can affect the TIE rate control.

Interests of clinical investigators give rise to a mathematically express procedure or statistical decision rules that are based on sample from populations. When constructing decision rules, two additional information resources can be incorporated. The first is a defined function that consists of the explicit, quantified gains and losses and their relative weights to reach a conclusion. Frequently, this function defines the expected loss corresponding to each possible decision. This type of information can incorporate a *loss function* into the statistical decision-making process. The second information source is a prior knowledge. Commonly, in order to derive prior information, researches should consider past experiences about similar situations. The Bayesian methodology (e.g., Berger, 2010) formally provides clear technique manuals on how to construct efficient statistical decision rules with prior information in various complex problems related to clinical experiments, employing prior information.

## 1.2 Parametric Approach

In constructing decision rules, a statistician may use a sort of technical statements relevant to observed data. Some information used for test construction can give rise to technical statements that oftentimes are called assumptions

regarding the distribution of data. The assumptions often define a fit of the data distribution to a functional form that is completely known or known up to parameters. A complete knowledge of the distribution of data can provide all the information that investigators need for efficient applications of statistical techniques. However, in many scenarios, the assumptions are reasonably guessed and very difficult to be proven or tested for their propriety. Widely used assumptions in biostatistics are that data derived via a clinical study follow one of the commonly used distribution functions such as the Normal, Lognormal, $t$, $\chi^2$, Gamma, F, Binominal, Uniform, Wishart, and Poisson. The distribution function of the data can be defined including parameters. For example, the normal distribution $N(\mu,\sigma^2)$ has the shape of the famous bell curve, where the parameters $\mu$ and $\sigma^2$ representing a mean and variance of a population define the distribution. Values of the parameters may be assumed to be unknown. Mostly in such cases, assumed functional forms of the data distributions are involved to make statistical decision rules via the use of statistics from the sample, which we call *Parametric Statistics*. If certain key assumptions are met, parametric methods can yield very simple, efficient, and powerful inferences.

## 1.3 Warning—Parametric Approach and Detour: Nonparametric Approach

The statistical literature widely addresses an issue that parametric methods are often sensitive to moderate violations of parametric assumptions and hence are nonrobust (e.g., Freedman, 2009). In order to reduce a risk to apply an incorrect parametric approach, the parametric assumptions can be tested. In this case, statisticians can try to verify the assumptions while making decisions with respect to main objectives of the clinical study. This leads to complicated topics dealt with in multiple testing. Also, it turns out that a computation of an expected risk that may lead to a wrong decision strongly depends on errors that can be made by failing to reject the parametric assumptions. The complexity of this problem can increase when researchers examine various functional forms to fit the data distribution in order to apply parametric methods. A substantial theoretical and experimental literature has discussions of the pitfalls of multiple testing that places blame squarely on the shoulders of the many clinical investigators who examine their data before deciding how to analyze it or neglecting to report the statistical tests that may not have supported their theses (e.g., Austin et al., 2006). In this context, one can present various cases, both hypothetical and actual, to get to the heart of issues arising especially in the health-related sciences. Note also that in many situations, due to the wide variety and complex nature of problematic real data, e.g. incomplete data subject

to instrumental limitations of studies (e.g., Vexler et al., 2008a,b), statistical parametric assumptions are hardly satisfied, and their relevant formal tests are complicated or oftentimes are not readily available. Unfortunately, even clinical investigators trained in statistical methods do not always verify the corresponding parametric assumptions and do not attend to probabilistic errors of the corresponding verification, when they use well-known basic parametric statistical methods, e.g., the *t*-test.

It is known that when the key assumptions are not met the parametric approach may be extremely biased and inefficient when compared to their robust nonparametric counterparts. Statistical inference under the nonparametric regime offers decision-making procedures, avoiding or minimizing the use of the assumptions regarding functional forms of the data distributions. In general, the choice between nonparametric and parametric approaches can boil down to expected efficiency versus robustness to assumptions. Thus, an important issue is to preserve efficiency of statistical techniques through the use of robust nonparametric likelihood methods, minimizing required assumptions about data distributions.

## 1.4 A Brief Ode to Likelihood

Testing statistical hypotheses based on the *t*-test or its modifications is one of the traditional instruments used in medical experiments and drug development. Despite the fact that these tests are straightforward with respect to their applications to clinical and medical settings, it should be noted that there has been a huge literature on the criticism of *t*-test type statistical tools. One major issue, which has been widely recognized, is with respect to the significant loss of efficiency of these procedures under different distributional assumptions. The legitimacy of *t*-test type procedures also comes into question in the context of inflated TIEs when data distributions differ from normal and the number of available observations is limited. The recent biostatistical literature has addressed the arguments well that values of biomarker measurements tend to follow skewed distributions, e.g. a lognormal distribution (Limpert et al., 2001), and hence the use of *t*-test type techniques in this setting is suboptimal and accompanied by difficulties to control the corresponding TIE rates.

Consider the following example based on data from a study evaluating biomarkers related to atherosclerotic coronary heart disease (Schisterman et al., 2001). A cross-sectional population-based sample of randomly selected

residents (age 35–79) of Erie and Niagara counties of the state of New York, United States, was used for the analysis. The New York State Department of Motor Vehicles drivers' license rolls were employed as the sampling frame for adults between the ages of 35 and 65, whereas the elderly sample (age 65–79) was randomly selected from the Health Care Financing Administration database. Participants provided a 12-hour fasting blood specimen for biochemical analysis at baseline, and a number of characteristics were evaluated from fresh blood samples. Figure 1.1 depicts a screenshot, demonstrating the example though the use of R, a powerful and flexible statistical software language (e.g., Crawley, 2012 for its introduction).

The samples $X$ and $Y$ present 50 measurements (mg/dL) of the biomarker high-density lipoprotein (HDL) cholesterol obtained from healthy patients. These measurements were divided into the two groups (i.e., $X$ and $Y$). Although one can reasonably expect the samples are from the same population, the $t$-test shows a significant difference of their distributions. Perhaps, the following issues may be taken into account to explain reasons of this incorrect output of the $t$-test. The histograms displayed in Figure 1.1 indicate that the distributions of the $X$ and $Y$ probably are skewed. In a non-asymptotic context, when the sample sizes are relatively small, one can show that the $t$-test-statistic is a product of likelihood ratio-type considerations based on normally distributed observations (e.g., Lehmann and Romano, 2005). That is, the $t$-test is a parametric test and the parametric assumption seems to be violated, in this example.

Thus, in many settings, it may be reasonable to propose an approach for developing statistical tests, attending data distributions, to provide procedures that are efficient as the $t$-test based on normally distributed observations. Toward this end the likelihood methodology can be employed.

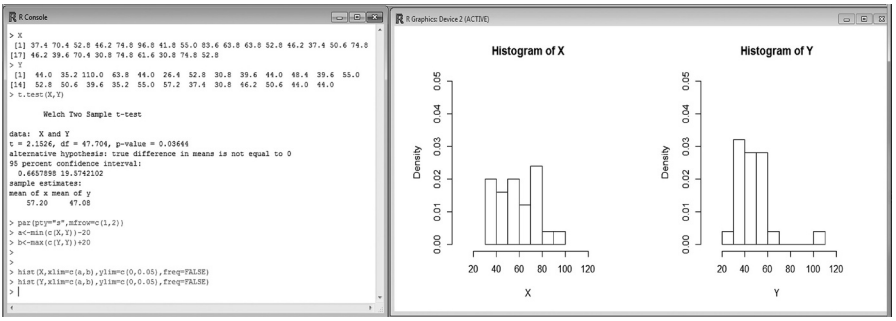



**FIGURE 1.1**
R data analysis output for measurements of HDL cholesterol levels in healthy individuals.

### 1.4.1 Likelihood Ratios and Optimality

Now we outline the likelihood principle. When the forms of data distributions are assumed to be known the likelihood principle is a central tenet for developing powerful statistical inference tools. The *likelihood method* or simply the *likelihood* is arguably the most important concept for inference in parametric modeling (Neyman and Pearson, 1992), and this fact equally applies when the underlying data are subjected to different problems and limitations related to medical and epidemiological studies, e.g. in the context of the analysis of survival data. Likelihood-based testing that we know was mainly found and formulated in a series of fundamental papers published in the period of 1928–1938 by Jerzy Neyman and Egon Pearson (Neyman and Pearson, 1928–1938). In 1928, the authors introduced the generalized likelihood ratio test and its association with chi-squared statistics. Five years later, the Neyman–Pearson Lemma (Neyman and Pearson, 1933) was introduced showing the optimality of the likelihood ratio test. These seminal works provided us with the familiar notions of simple and composite hypotheses and errors of the first and second kind, thus defining formal decision-making rules for testing. Without loss of generality, the principle idea of the proof of the Neyman–Pearson Lemma can be shown by using the trivial inequality

$$\big(A-B\big)\big(I\{A\geq B\}-\delta\big)\geq 0, \tag{1.1}$$

for any real numbers $A$, $B$, where $\delta\in[0,1]$ and $I\{\cdot\}$ denote the indicator function. For example, suppose we would like to classify independent identically distributed (i.i.d.) biomarker measurements $\{X_i, i=1,\ldots,n\}$ corresponding to hypotheses of the form $H_0$: $X_1$ is from a density function $f_0$ versus $H_1$: $X_1$ is from a density function $f_1$. In this context, to construct the likelihood ratio test statistic, we should consider the ratio between the joint density function of $\{X_1,\ldots,X_n\}$ obtained under $H_1$ and the joint density function of $\{X_1,\ldots,X_n\}$ obtained under $H_0$, and then define $\prod_{i=1}^{n} f_1(X_i)\big/\prod_{i=1}^{n} f_0(X_i)$ to be the likelihood ratio. In this case the likelihood ratio test is uniformly most powerful. This proposition directly follows from the expected value under $H_0$ of the inequality (1.1), where we define $A=\prod_{i=1}^{n} f_1(X_i)/f_0(X_i)$, $B$ to be a test-threshold (i.e., the likelihood ratio test rejects $H_0$ if and only if $A\geq B$), and $\delta$ is assumed to represent any decision rule based on $\{X_i, i=1,\ldots,n\}$. The Appendix contains details of the proof. This simple proof-technique was used to show optimal aspects of different statistical decision-making policies based on the likelihood ratio concept applied in clinical experiments (e.g., Vexler, Wu, and Yu, 2008; Vexler and Wu, 2009; Vexler and Gurevich, 2011).

### 1.4.2 The Likelihood Ratio Based on the Likelihood Ratio Test Statistic Is the Likelihood Ratio Test Statistic

The Neyman–Pearson concept to test, fixing the probability of a TIE, comes under some criticism by epidemiologists. One of the critical points is related to the Type II error, the incorrect decision by failing to reject the null hypothesis when the alternative hypothesis is true. For example, Freiman et al. (1978) pointed out results of 71 clinical trials that reported no *significant* differences between the compared treatments. The authors found that in the great majority of these trials the strong effects of new treatment are reasonable. On failing to reject the null hypothesis, the investigators in such trials inappropriately accepted the null hypothesis as correct, which probably resulted in the Type II error. In the context of likelihood ratio-based tests, we present the following result that demonstrates an association between the probabilities of the Type I and II errors.

Suppose we would like to test for $H_0$ versus $H_1$, employing the likelihood ratio $L = f_{H_1}(D) / f_{H_0}(D)$ based on data $D$, where $f_{H_i}$ defines a density function that corresponds to the data distribution under the hypothesis $H_i$. Say, for simplicity, we reject $H_0$ if $L > C$, where $C$ is a presumed threshold. In this case, one can then show that

$$f^L_{H_1}(u) = u\, f^L_{H_0}(u), \tag{1.2}$$

where $f^L_H(u)$ is the density function of the test statistic $L$ under the hypothesis $H$ and $u > 0$. Details of the proof of this fact are shown in the Appendix. Thus, we can obtain the probability of a Type II error in the form of

$$\begin{aligned}\Pr\{\text{the test does not reject } H_0 \mid H_1 \text{ is true}\} &= \Pr\{L \le C \mid H_1 \text{ is true}\} \\ &= \int_0^C f^L_{H_1}(u)\,du = \int_0^C u f^L_{H_0}(u)\,du.\end{aligned}$$

Now, if the density function $f^L_{H_0}(u)$ is assumed to be known to control the TIE rate, then the probability of the Type II error can be easily computed.

The likelihood ratio property $f^L_{H_1}(u) / f^L_{H_0}(u) = u$ can be applied to solve different issues related to performances of the likelihood ratio test. For example, in a term of the bias of the test, one can request to find a value of the threshold $C$ that maximizes

$$\Pr\{\text{the test rejects } H_0 \mid H_1 \text{ is true}\} - \Pr\{\text{the test rejects } H_0 \mid H_0 \text{ is true}\},$$

where the probability Pr{the test rejects $H_0 \mid H_1$ is true} depicts the power of the test. This equation can be expressed as

$$\Pr\{L > C \mid H_1 \text{ is true}\} - \Pr\{L > C \mid H_0 \text{ is true}\} = \left(1 - \int_0^C f_{H_1}^L(u)du\right) - \left(1 - \int_0^C f_{H_0}^L(u)du\right).$$

Let the derivative of this notation equal zero and solve the equation:

$$\frac{d}{dC}\left[\left(1 - \int_0^C f_{H_1}^L(u)du\right) - \left(1 - \int_0^C f_{H_0}^L(u)du\right)\right] = -f_{H_1}^L(C) + f_{H_0}^L(C) = 0.$$

By virtue of property (1.2), this implies $-Cf_{H_0}^L(C) + f_{H_0}^L(C) = 0$ and then $C = 1$ that provides the maximum discrimination between the power and the probability of a TIE of the likelihood ratio test.

*In words, an interesting fact is that the likelihood ratio $f_{H_1}^L / f_{H_0}^L$ based on the likelihood ratio $L = f_{H_1}/f_{H_0}$ becomes to be the likelihood ratio, that is, $f_{H_1}^L(L)/f_{H_0}^L(L) = L$. We leave interpretations of this statement, may be in terms of information, to the reader's imagination.*

## 1.5 Maximum Likelihood: Is It the Likelihood?

Various real-world data problems require considerations of statistical hypotheses with structures, which depend on unknown parameters. In this case, the maximum likelihood method proposes to approximate the most powerful likelihood ratio, employing a proportion of the maximum likelihoods, where the maximizations are over values of the unknown parameters belonging to distributions of observations under the corresponding hypotheses. We shall assume the existence of essential maximum likelihood estimators. The influential theorem of Wilks (1938) provides the basic rational as to why the maximum likelihood ratio approach has had tremendous success in statistical applications. Wilks showed that under regularity conditions, asymptotic null distributions of maximum likelihood ratio test statistics are independent of nuisance parameters. That is, the TIE rates of the maximum likelihood ratio tests can be controlled asymptotically and approximations of the corresponding p-values can also be computed.

Thus, if certain key assumptions are met one can show that parametric likelihood methods are very powerful and efficient statistical tools. We should emphasize that the discovery related to the likelihood ratio methodology in statistical developments may be comparable with the development of the

assembly line technique of mass production. The likelihood ratio principle gives clear instructions and technique manuals on how to construct efficient statistical decision rules in various complex problems related to clinical experiments. For example, Vexler et al. (2011c) developed a maximum likelihood ratio test for comparing populations based on incomplete longitudinal data subjected to instrumental limitations.

Although many statistical publications continue to contribute to the likelihood paradigm and are very important in the statistical discipline (an excellent account can be found in Lehmann and Romano, 2005), several significant questions arise naturally about the general applicability of the maximum likelihood approach. Conceptually, there is an issue specific to classifying maximum likelihoods in terms of likelihoods that are given by joint density (or probability) functions based on data. Integrated likelihood functions, with respect to arguments related to data points, are equal to one; however, accordingly integrated maximum likelihood functions often have values that are indefinite. Thus, although likelihoods present full information regarding the data, the maximum likelihoods might lose information conditional on the observed data. Consider the simple example: Suppose we observe $X_1$, that is assumed to be from a normal distribution $N(\mu,1)$ with mean parameter $\mu$. In this case the likelihood has the form $(2\pi)^{-0.5}\exp(-(X_1-\mu)^2/2)$ and correspondingly $\int(2\pi)^{-0.5}\exp(-(X_1-\mu)^2/2)\,dX_1=1$, whereas the maximum likelihood, i.e., the likelihood evaluated at estimated $\mu, \hat{\mu}=X_1$, is $(2\pi)^{-0.5}$, which clearly does not represent the data and is not a proper density. This demonstrates that as the Neyman–Pearson lemma is fundamentally found on the use of the density-based constitutions of likelihood ratios, maximum likelihood ratios cannot be optimal in general. That is, the likelihood ratio principle is in general not robust when the hypothesis tests have corresponding nuisance parameters to consider, e.g. testing a hypothesized mean given an unknown variance. An additional inherent difficulty of the likelihood ratio test occurs when a clinical experiment is associated with an infinite-dimensional problem with the number of unknown parameters being relatively large. In this case, Wilks theorem should be re-evaluated and nonparametric approaches can be considered in the contexts of reasonable alternatives to the parametric likelihood methodology (e.g., Fan et al., 2001).

The ideas of likelihood and maximum likelihood ratio testing may not be fiducial and applicable in general nonparametric function estimation/testing settings. It is also well known that when key assumptions are not met, parametric approaches may be suboptimal or biased as compared to their robust counterparts across the many features of statistical inferences. For example, in a biomedical application, Ghosh (1995) proved that the maximum likelihood estimators for the Rasch model are inconsistent as the number of nuisance parameters increases to infinity (Rasch models are often utilized in clinical trials that deal with psychological measurements, e.g., abilities, attitudes, personality traits). Due to the structure of likelihood functions based on products of densities or conditional density functions,

relatively nonsignificant errors of classifications of data distributions can lead to vital problems related to applications of likelihood ratio-type tests (e.g., Gurevich and Vexler, 2010). Moreover, one can note that given the wide variety and complex nature of biomedical data, e.g. incomplete data subject to instrumental limitations or complex correlation structures, parametric assumptions are rarely satisfied. The respective formal tests are complicated or oftentimes are not readily available.

## 1.6 Empirical Likelihood

The empirical likelihood (EL) approach is based on a data-driven likelihood function, thus is intrinsically nonparametric and comparatively powerful (Lazar and Mykland, 1998). An advantage of using the EL method is that it does not require the distribution assumption. The EL is able to incorporate known constraints on parameters in an inferential setting under both the null and alternative hypotheses. EL hypothesis tests maintain a prespecified TIE rate relatively well with various underlying distributions. In two group comparisons, they offer robust testing procedures under violations of the exchangeability assumptions (e.g., Yu et al., 2011). Historically the EL method was first introduced for the analysis of censored data (Thomas and Grunkemeier, 1975; Owen, 1991). Owen (1988) introduced the empirical likelihood ratio (ELR) approach to construct confidence intervals. Since being introduced into the statistical literature, the EL approach has demonstrated its practical applicability via extensions to a variety of statistical problems (e.g., Yang and Zhao, 2007; Vexler and Gurevich, 2010; Wang et al., 2010). The EL method incorporates information or assumptions regarding the parameters and translates those to the distribution-free likelihood estimation; thus the method can be used to combine additional information about parameters of interest (Qin and Lawless, 1994).

In comparison with classical testing methods based on normal approximations, the EL ratio test statistic does not rely on symmetric rejection regions, thus giving rise to more accurate tests (Hall and La Scala, 1990). Owen (1990) showed that the EL ratio provides confidence intervals less affected by the skewness of distribution comparing with methods based on the central limit theorem. DiCiccio et al. (1991) demonstrated that the EL method can achieve an excellent coverage rate for confidence intervals by applying some parametric techniques such as the Bartlett correction. The EL method for constructing confidence regions for parameters has comparable sampling properties of the bootstrap. Although the bootstrap uses resampling, the EL

method computes the profile likelihood of a general multinomial distribution based on data points. The property that EL produces regions that reflect emphasis in the observed dataset and involves no predetermined assumptions about the shape has considerable potential in the construction of confidence bands for curve estimators. Chen (1994) compared the powers of EL ratios and bootstrap tests for a mean parameter against a series of local alternative hypotheses (see Chapter 2 for details). It is shown that the EL ratio test can be more powerful than the bootstrap test depending on the population skewness parameter.

Versatility of the EL method is demonstrated in many different data analytical settings. Researchers have worked in the area specifically related to quantiles using the EL method; e.g., Chen and Hall (1993) used a smoothed EL approach to estimate confidence intervals for quantiles using kernel functions. They showed that the coverage accuracy may be improved from order $n^{-1/2}$ to order $n^{-1}$ by appropriately smoothing the EL method. The improvement is available for a wide range of choices of the smoothing parameter so that accurate choice of an *optimal* value of the parameter is not necessary. Chen and Chen (2000) investigated the properties of EL quantile estimation in large samples; Zhou and Jing (2003a) proposed an alternative smoothed EL approach where the EL ratio has an explicit form based on the concept of the M-estimators; and Lopez et al. (2009) investigated testing general parameters that are determined by the expectation of non-smooth functions. No distributional assumptions of EL allow the method to be used for analyzing data with complicated underlying distributions. The EL provides better performance with the confidence interval for the mean of a population with many zeros comparing with the method using parametric likelihood, whereas overall coverage properties are similar for both methods under various distribution assumptions (Chen et al., 2003; Kang et al., 2010). Qin and Leung (2005) used a semiparametric likelihood approach to estimate the distribution of the malaria parasite level, which is a mixture distribution where a component of the mixture distribution was again the mixture of discrete and continuous distributions. Qin (2000) showed an inference on incomplete bivariate data using a method that combines the parametric model and ELs. His work was extended to the group comparison using the EL by Yu et al. (2010). The EL method also incorporates auxiliary information of variables in a form of constraints, which can be obtained from reliable resources such as census reports (e.g., Qin and Lawless, 1994; Chen and Qin, 1993).

We conclude this section emphasizing the following important aspects: (1) The EL concept can provide efficiently nonparametric approximations to optimal (e.g., most powerful) parametric statistical schemes; (2) the EL methodology yields robust algorithms to solve a variety of complex problems related to clinical trials; (3) EL functions can be easily combined with parametric and

semiparametric likelihood functions to develop statistical procedures that demonstrate attractive properties in complicated model settings. These factors adduce evidence that EL techniques have great potentials to be adopted as primary statistical tools employed by clinical investigators.

## 1.7 Why Empirical Likelihood?

### 1.7.1 The Necessity and Danger of Testing Statistical Hypothesis

The ubiquitous use of statistical decision-making procedures in the current medical literature displays the vital role that statistical hypothesis testing plays in different branches of biomedical sciences. The benefits and fruits of statistical tests based on mathematical-probabilistic techniques, in epidemiology or other health-related disciplines, strongly depend on successful formal presentations of statements of problems and a description of nature. Oftentimes, certain assumptions about the observations used for the tests provide the probability statements that are required for the statistical tests. These assumptions do not come for free and ignoring their appropriateness can cause serious bias or inconsistency of statistical inferences, even when the test procedures thyself are carried out without mistakes. The sensitivity of the probabilistic properties of a test to the assumptions is referred to as the lack of robustness of the test (e.g., Wilcox, 1998).

Various statistical techniques require parametric assumptions that are to define forms of data distributions to be known up to parameters' values. For example, in the conventional *t*-test, the assumptions are that the observations of different individuals are realizations of independent, normally distributed, random variables, with the same expected value and variance for all individuals within the investigated group. Such assumptions are not automatically satisfied, and for some assumptions it may be doubted whether they are ever satisfied exactly. The null hypothesis $H_0$ and alternative hypothesis $H_1$ are statements which, strictly speaking, imply these assumptions, and which therefore are not each other's complement. There is a possibility that the assumptions are invalid, and neither $H_0$ nor $H_1$ is true. Thus, we can reject a statement related to clinical trials' interests just because the assumptions are not met. This issue is an impetus to departure from parametric families of data distributions, employing nonparametric test strategies.

One of the advantages of EL techniques lies in their generality and an assessment of their performance lies under conditions that are commonly unrestricted by parametric assumptions.

### 1.7.2 The Three Sources That Support the Empirical Likelihood Methodology for Applying in Practice

When in doubt about the best strategy to make statistical decision rules, the following arguments can be accepted in favor of EL methods:

1. The EL methodology employs the likelihood concept in a simple nonparametric fashion to approximate optimal parametric procedures. The benefit of using this approach is that the EL techniques are often robust and highly efficient. In this context, we also may apply EL functions to replace parametric likelihood functions in known and well-developed constructions. Consider the following example. The statistical literature widely suggests applying Bayesian methods for various tasks of clinical experiments, for example, when data are subjected to complex missing data problems, e.g. parts of data are not manifested as numerical scores (Daniels and Hogan, 2008). Commonly, to apply a Bayesian approach, one needs to assume functional forms corresponding to the distribution of the underlying data and parameters of interest. Lazar (2003) demonstrated potentials of constructing nonparametric Bayesian inference based on ELs that take the role of model-based likelihoods. This research demonstrated that the EL is a valid function for Bayesian inference. Vexler et al. (2013a) recommended applying EL functions to create Bayes Factor (BF)-type nonparametric procedures. The BF, a practical tool of applied biostatistics, has been dealt with extensively in the literature in the context of hypothesis testing (e.g., Carlin and Louis, 2000). The EL concept was shown to be very efficient when it is employed for modifying BF-type procedures to the nonparametric setting.
2. Similar to the parametric likelihood concept including Bayesian approaches, the EL methodology gives relatively simple systematic directions for constructing efficient statistical tests that can be applied in various complex clinical experiments.
3. Perhaps, the extreme generality of EL methods and their wide scope of usefulness partly follow on abilities to easily set up EL statistics as components of composite parametric/semi- and nonparametric likelihood-based systems, efficiently attending any observed data and relevant information. Parametric, semiparametric, and EL methods play roles complementary to one another, providing powerful statistical procedures for complicated practical problems.