Mathematics Monograph Series 17

Empirical Likelihood in Nonparametric and Semiparametric Models

Liugen Xue and Lixing Zhu

（非参数和半参数模型中的经验似然）

SCIENCE PRESS
Beijing

(O-3923.0101)

Empirical Likelihood in Nonparametric and Semiparametric Models

This book is composed of ten chapters. The first chapter contains the preliminary knowledge about empirical likelihood and other relevant nonparametric methods. Chapters 2 and 3 analyze the section-data using the single-index model and the partially linear single-index model. Chapters 4 through 6 investigate the longitudinal data using the partially linear model, the varying coefficient model and a nonparametric regression model. Chapter 7 discusses nonlinear errors-in-covariables models with validation data. Chapters 8 through 10 investigate missing data under the framework of the linear model, a nonparametric regression model and the partially linear model. Every chapter, except for Chapter 1, of this book is self-contained so that the reader could focus on any chapter without much effect on the understanding of the others, and hence can read any chapters according to reader's own interest. The emphasis of this book is on methodologies rather than on theory, with a particular focus on applications of the empirical likelihood techniques to various semiparametric regression models. Key technical arguments are presented in the "proofs sections" at the end of each chapter. This gives interested researchers an idea of how the theoretical results are obtained. Also from the style of material organization, this book is more likely a lecture note, rather than a textbook. Most materials come from authors' research articles.

This book intends to provide a useful reference for researchers and to serve as a lecture note to postgraduate students. It is especially for the people working in the nonparametric and semiparametric statistics areas or applying the empirical likelihood method to other areas.

ISBN 978-7-03-027834-0

9 787030 278340 >

定 价：68.00 元

Mathematics Monograph Series 17

Liugen Xue and Lixing Zhu

Empirical Likelihood in Nonparametric and Semiparametric Models

(非参数和半参数模型中的经验似然)

SCIENCE PRESS
Beijing

Responsible Editor: Yuzhuo Chen

Published by Science Press
16 Donghuangchenggen North Street
Beijing 100717, P. R. China

Printed in Beijing

ISBN 978-7-03-027834-0

Preface

Recent years, the empirical likelihood method has received great attention when we deal with statistical inference for nonparametric and semiparametric regression models. These models include fully nonparametric regression, single-index, partially linear single-index, varying coefficient models, so on so forth. However, how to efficiently apply the empirical likelihood to these models is of particular interest and challenging. This is because for such models, classical empirical likelihood is not asymptotically distribution-free any more. The main reason that causes this difficulty is that in such models, there are two unknowns: the parameters of interest and some nonparametric link functions or additive functions, of which we need to regard them as infinite-dimensional nuisance parameters. Clearly, when we consider constructing confidence regions for the parameters of interest in these models, plug-in estimators are needed to replace the unknown nonparametric link functions. This is a commonly used method in the literatures, but it causes why the classic empirical likelihood does not have tractable limiting distribution. We in recent years have been studying this problem and proposed several bias correction methods to make the empirical likelihood more useful for these models.

Owen (2001) is the only comprehensive book in the empirical likelihood. As the pioneer in this area, Owen did the fundamental work and collected many important works in his book. However, for confidence region construction and hypothesis testing, Owen's book does not contain the materials about nonparametric and semiparametric regression models, especially bias correction approaches. Our book will present these different methods and the applications. Specifically, we will describe and illustrate the empirical likelihood method with "bias-correction" for constructing empirical likelihood ratios.

This book is composed of ten chapters. The first chapter will contain some preliminary knowledge. Chapters 2 and 3 will analyze the cross-section data using the single-index model and the partially linear single-index model. Chapters 4 through 6 will investigate the longitudinal data using the partially linear model, the varying coefficient model and a nonparametric regression model. Chapter 7 will discuss nonlinear errors-in-covariables models with validation data. Chapters 8 through 10 will investigate missing data using the linear model, a nonparametric regression model and the partially linear model. Each chapter, except for Chapter 1, of this book will be self-contained so that the reader could focus on any chapter without much effect on the understanding of the others, and hence can read any

chapters of interest according to reader's own interest. The emphasis of this book is on methodologies rather than on theory, with a particular focus on applications of the empirical likelihood techniques to various semiparametric regression models. Key technical arguments are presented in the "proofs sections" at the end of each chapter. This gives interested researchers an idea of how the theoretical results are obtained. Hopefully, the reader would find the technical arguments useful to his/her future research. Also from the style of material organization, this book is more likely a lecture note, rather than a textbook. Most materials come from our research articles.

This book intends to provide a useful reference for researchers and to serve as a lecture note to postgraduate students. It is especially for the people working in the nonparametric and semiparametric statistics area or applying the empirical likelihood method to other areas. As the empirical likelihood has been one of the most important tools in statistical analysis, the body of such people is fairly large. The people who work in this area should be interested in seeing the new approaches so as to apply it. We hope that the reader will be stimulated to choose or develop his/her own methodology in the universe of nonparametric and semiparametric statistics.

We owe much to our friends and colleagues. Chapter 7 is based on the joint work with W. Stute who had great contribution when we prepared that paper.

Xue's research was supported by the National Natural Science Foundation of China (10871013), the Beijing Natural Science Foundation (1102008) and the Ph.D. Program Foundation of Ministry of Education of China (20070005003) and PHR (IHLB). Zhu's research was supported by two grants from the Research Grants Council of Hong Kong, and partly supported by Yunnan University of Finance and Economics for his frequent visits to do joint research.

We express our deep gratitude to Ms. Yuzuo Chen for her constant help during the writing of this book.

Liugen Xue and Lixing Zhu
January, 2010

Contents

Chapter 1

Preliminary knowledge

1.1 Empirical likelihood (EL)

1.1.1 Definition of EL

The method of empirical likelihood developed by Owen (1998, 1990) provided a means of determining nonparametric confidence regions for statistical functionals. The method is likelihood based, but does not require the assumption of a parametric family for the data. Let $X_1 \cdots, X_n$ be independent random vectors in $\mathbf{R}^p$, for $p \geqslant 1$, with a common distribution function F_0, and let F_n be the empirical distribution function, which assigns probability mass n^{-1} to each of the observed data points. Then F_n maximises the nonparametric likelihood function

$$L(F) = \prod_{i=1}^{n} F\{X_i\},$$

where F is any probability measure on $\mathbf{R}^p$ and $F\{X_i\}$ is the probability of getting the value X_i in a sample from F. Following Owen (1988, 1990), the empirical likelihood ratio function is defined as

$$R(F) = \frac{L(F)}{L(F_n)}.$$

When there are no ties among the X_i, the empirical likelihood ratio function takes the form

$$R(F) = \prod_{i=1}^{n} np_i, \quad p_i = F\{X_i\}.$$

Owen (1988) showed that this formula is still appropriate even when there are ties in data, with the natural modification $\sum_{j:X_j=X_i} p_j = F\{X_i\}$. Taking the supremum of $R(F)$ subject to the constraint $T(F) = t$, forces $p_i = p_j$ whenever $X_i = X_j$. In other words, ties among data do not affect this natural re-expression of the likelihood ratio.

Suppose that interest centres on $T(F_0)$, where $T(\cdot)$ is a statistical functional. The nonparametric maximum likelihood estimate of $T(F_0)$ is $T(F_n)$. Owen (1988, 1990) showed that, under some reasonable conditions, sets of the form

$$\{T(F)|R(F) \geqslant r\}$$

may be used as confidence region for $T(F_0)$. In order to keep such regions well behaved, the class of distributions over which T is evaluated is restricted to those whose support is the observed sample, denoted by $F \ll F_n$.

Owen (1988, 1990) showed that under quite general conditions $-2\log\{R(F)\}$ converges to χ_q^2 in distribution, where χ_q^2 is the chi-square distribution with q degrees of freedom, and q is the number of parameters being estimated. Based on this asymptotic property, appropriate cut-off levels can be determined for empirical likelihood confidence regions of a specified coverage.

We proceed by analogy with parametric likelihood. Suppose that we are interested in a parameter $\theta = T(F)$ for some function T of distributions. The F is a member of a set $\mathcal{F}$ of distributions. In some cases we may take $\mathcal{F}$ to be the set of all distributions on $\mathbf{R}$. More often, we use a smaller set of distributions. Define the profile likelihood ratio function:

$$\mathcal{R}(\theta) = \sup\{R(F)|T(F) = \theta, F \in \mathcal{F}\}.$$

Empirical likelihood hypothesis tests reject $T(F_0) = \theta$, when $\mathcal{R}(\theta_0) < r_0$ for some threshold value r_0. Empirical likelihood confidence regions are of the form

$$\{\theta|\mathcal{R}(\theta) \geqslant r_0\}.$$

In many settings, the threshold r_0 may be chosen using an empirical likelihood theorem (ELT), a nonparametric analogue of Wilks' theorem.

1.1.2 EL for mean

Suppose that F_0 has mean $\mu_0 = (\mu_{01}, \cdots, \mu_{0p}) \in \mathbf{R}^p$ and variance V_0 of full rank. In order to form an empirical likelihood confidence region for μ_0, we define the profile empirical likelihood ratio function

$$\mathcal{R}(\mu) = \max\left\{\prod_{i=1}^{n} np_i \middle| p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i X_i = \mu\right\}.$$

Owen (1990) showed that $-2\log\mathcal{R}(\mu) \to \chi_p^2$ in distribution as $n \to \infty$, which is analogous to the parametric case shown by Wilks (1938). Therefore, to construct an approximate $(1-\alpha)$-level confidence region for μ_0, one computes the set

$$C_{\mu_0} = \{\mu \in \mathbf{R}^p | -2\log\mathcal{R}(\mu) \leqslant \chi_p^2(1-\alpha)\},$$

where $\chi_p^2(1-\alpha)$ is defined such that $P\{\chi_p^2 \leqslant \chi_p^2(1-\alpha)\} = 1-\alpha$.

A discussion of the computation of $\mathcal{R}(\mu)$ can be found in Owen (1990, Section 3). The problem of maximizing $\prod_{i=1}^{n} np_i$, subject to the constraints $p_i \geqslant 0$, $\sum_{i=1}^{n} p_i = 1$,

and $\sum_{i=1}^{n} p_i X_i = \mu$, is shown, using a Lagrange multiplier argument, to be equivalent to minimizing the expression $-\sum \log(1 + \lambda^{\mathrm{T}}(X_i - \mu))$ over $\lambda \equiv \lambda(\mu) \in \mathbf{R}^p$, when μ is in the convex hull of the data, i.e. $ch(\{X_i, \cdots, X_n\})$. This alternative version of the problem is the convex dual of the original. Instead of attempting to solve a constrained maximization problem, the researcher is faced with the much easier task of finding the unconstrained minimum of a convex function, a problem for which many algorithms exist.

1.1.3 Estimating equations

Estimating equations provide an extremely flexible way to describe parameters and the corresponding statistics. For a random variable $X \in \mathbf{R}^d$, a parameter $\theta \in \mathbf{R}^p$, and a vector-valued function $m(X, \theta) \in \mathbf{R}^s$ suppose that

$$E[m(X, \theta)] = 0. \tag{1.1.1}$$

The usual setting has $p = s$ and then under conditions on $m(X, \theta)$ and possibly on F, there is a unique solution θ. In this just determined case, the true value θ_0 may be estimated by solving

$$\frac{1}{n} \sum_{i=1}^{n} m(X, \hat{\theta}) = 0 \tag{1.1.2}$$

for $\hat{\theta}$. To write a vector mean by equation (1.1.1), we take $m(X, \theta) = X - \theta$, and then equation (1.1.2) gives $\hat{\theta} = \bar{X}$. For $P(X \in A)$ take $m(X, \theta) = I\{X \in A\} - \theta$. For a continuously distributed scalar X and $\theta \in R$, the function $m(X, \theta) = I\{X \leqslant \theta\} - \alpha$ defines θ as the α quantile of X. Owen (2001, Section 3.6) described tail probabilities and quantiles in more detail.

Equation (1.1.2) is known as an estimating equation, and $m(X, \theta)$ is called estimating function. Most maximum likelihood estimators are defined through estimating equations.

The underdetermined case $s < p$ can also be useful. Then (1.1.1) and (1.1.2) might each have an $s - p$ dimensional solution set of θ values. Some functions of θ may be precisely determined from the data, while the others will not.

In econometrics, considerable interest attaches to the overdetermined case with $s > p$. In problems with $s > p$ the fact that (1.1.1) holds is a special feature of F and constitutes important side information. Even when (1.1.1) holds for the true F_0, it will not ordinarily hold for the nonparametric maximum likelihood estimate $\hat{F}$, in which case (1.1.2) has no solution. The generalized method of moments looks for a value $\hat{\theta}$ that comes close to solving (1.1.1). An empirical likelihood approach to this problem was described in Owen (2001, Section 3.10).

The empirical likelihood and estimating equations are well suited to each other. The empirical likelihood ratio function for θ is defined by

$$\mathcal{R}(\mu) = \max\left\{\prod_{i=1}^{n} np_i \Big| p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i m(X_i, \theta) = 0\right\}.$$

Owen (2001) showed that $-2\log\mathcal{R}(\mu) \to \chi_p^2$ in distribution as $n \to \infty$, where θ_0 satisfies $E[m(X, \theta)] = 0$.

1.1.4 Advantages of EL

For parametric models, the empirical likelihood has been proved to be a very powerful tool. It has many advantages over normal approximation based method and the bootstrap method for constructing confidence intervals. First, the empirical likelihood-based confidence region does not need to impose prior constraints on the region shape, and the region is range preserving and transformation respecting (see Hall and La Scala, 1990). In addition, the empirical likelihood does not require the construction of a pivotal quantity. Second, as DiCiccio, Hall and Romano (1991) proved, the empirical likelihood is Bartlett correctable, and thus has an advantage over the bootstrap method. Third, the empirical likelihood does not involve a plug-in estimation for the limiting variance, which is used to make tests scale invariant. This is of particular importance especially for model checking. As is known, the limiting variance of the residuals is model dependent, and gets larger under alternatives than it does under null. Any plug-in estimator is then also model dependent, and in most of cases, becomes larger under the alternatives. This often deteriorates the power performance. Reader can refer to Stute, Thies and Zhu (1998) and Stute and Zhu (2005).

1.1.5 Related literature

The first use of an empirical likelihood ratio function to set confidence intervals appears to be Thomas and Grunkemeier (1975). Their application was to survival probabilities estimated by the Kaplan-Meier cure. Thomas and Grunkemeier provide a heuristic argument to show that empirical likelihood ratio intervals for a survival probability based on the χ_1^2 distribution have asymptotically correct coverage levels outside $[0, 1]$. This is especially appealing for survival probabilities near 0 or 1. Cox and Oakes (1984, Section 4.3) independently obtained the same intervals.

The empirical likelihood has parallels in the bootstrap literature. The Bayesian bootstrap of Rubin (1981) generated reweighted empirical distributions $\sum_{i=1}^{n} g_i \delta_{x_i}$, where the g_i's are positive random variables with unit sum. In the simplest case

they follow a unit Dirichlet distribution and may be sampled by taking the n gaps formed by 0, 1 and $n-1$ independent $U[0,1]$ random variables.

Owen (1991) and Chen (1993, 1994) applied the empirical likelihood to linear regression models, and proved that the empirical log-likelihood ratio is asymptotically chi-squared. This leads a direct use of limit distribution to construct confidence regions/intervals of regression parameters with asymptotically correct coverage probabilities. Kolaczyk (1994) made further extensions to generalized linear models. Qin and Lawless (1994) developed an empirical likelihood methodology based on general estimating equations. Wang and Jing (1999) and Shi and Lau (2000) considered a partial linear model with fixed design: see also Diciccio et al. (1991), Chen and Qin (1993), Qin (1993), Qin and Lawless (1994), Kitamura (1997) and Zhang (1997). Xue and Zhu (2006) and Zhu and Xue (2006) investigated the empirical likelihood confidence regions of the parameters in a single-index model and a partially linear single-index model. Kernel methods for the empirical likelihood have been looked at earlier, see, for example, Hall and Owen (1993), Chen and Qin (2000), Li and Van Keilegom (2002) and Owen (2001, Chapter 5). Xue and Zhu (2007a) investigated the local empirical likelihood-based inference for a varying coefficient model with longitudinal data. Xue and Zhu (2007b) investigated the issues of estimation and confidence region construction for a partially linear model with longitudinal data. Stute, Xue and Zhu (2007) studied inference in parametric-nonparametric errors-in-covariables regression models using an empirical likelihood approach based on validation data. Using local polynomial fitting, Xue (2010) studied construction of pointwise confidence intervals and simultaneous confidence bands for the nonparametric regression functions and their derivatives under clustered data. Under missing data, Xue (2009a, b, c) and Xue (2010) studied the nonparametric regression model, the linear model and the partially linear model. The other related works are: Diciccio et al. (1991), Chen and Hall (1993), Li (1995), Chen and Sitter (1999), Wang and Rao (2001, 2002a, b). Peng (2004), Wang, Linton and Härdle (2004), Xue and Zhu (2005), and Qin and Zhang (2007), among others. Owen (2001) is a fairly comprehensive reference book. Existing methods provide a valuable approach for confidence interval construction and tests in a nonparametric context.

1.2 Bootstrap method

Efron (1979) introduced a very general resampling procedure, called Bootstrap, for estimating distributions of statistics based on independent observations. The procedure is more widely applicable and has more sound of theoretical basis than the popular Quenoille-Tukey jackknife. Efron investigated a number of statistical

problems and demonstrated the feasibility of the bootstrap method. In past three decades, the bootstrap quantiles are frequently used for the purpose of constructing bootstrap-based confidence intervals. Like any other estimation procedures, the accuracy of the quantile estimators produced by the bootstrap method needs to be assessed.

A formal description of the bootstrap goes as follows. Let $X_1, \cdots, X_n$ be a random sample of size n from a population with unknown distribution F, and let $T_n = T_n(X_1, \cdots, X_n)$ be a statistic of interest. Let F_n be the empirical distribution function of X and let $X_1^*, \cdots, X_n^*$ be a random sample drawn from F_n. $\{X_1, \cdots, X_n\}$ is called a bootstrap sample. The bootstrap method estimates the distribution of T_n through the conditional distribution of $T_n^* = T_n(X_1^*, \cdots, X_n^*)$, given $X_1, \cdots X_n$. This conditional distribution is called the bootstrap distribution of T_n, and T_n^* is called the bootstrap statistic of T_n. In particular, the bootstrap estimates the variance of T_n by the conditional variance of T_n^*. Assume that $H_n(x) = P(R_n \leqslant x)$, where $R_n = R_n(T_n, F)$ is a real-valued functional of F. Then, a bootstrap estimator of $\hat{H}_n(x) = P^*(R_n^* \leqslant x)$, where $R_n^* = R_n(T_n^*, F)$ and P^* is the bootstrap conditional probability given $X_1, \cdots, X_n$. Since the bootstrap samples are generated from F_n, this method is called the nonparametric bootstrap. Note that $\hat{H}_n(x)$ will depend on F_n and hence itself is a random variable. To be specific, $\hat{H}_n(x)$ will change as the data $\{x_1, \cdots, x_n\}$ change. Recall that a bootstrap analysis is run to assess the accuracy of some primary statistical results. This produces bootstrap statistics, like standard errors or confidence intervals, which are assessments of error for the primary results.

Now we rewrite the above (generic) nonparametric bootstrap procedure into the following steps as follows. Refer to Efron and Tibshirani (1993) for detailed discussions.

Step 1. Construct an empirical probability distribution, F_n, from the sample by placing a probability of $1/n$ at each point, $X_1, \cdots, X_n$ of the sample. This is the empirical distribution function of the sample, which is the nonparametric maximum likelihood estimate of the population distribution, F.

Step 2. From the empirical distribution function, F_n, draw a random sample of size n with replacement. This is a *resample.*

Step 3. Calculate the statistic of interest, T_n, for this resample, yielding T_n^*.

Step 4. Repeat steps 2 and 3 B times, where B is a large number, in order to create B resamples. The practical size of B depends on the tests to be run on the data. Typically, B is at least equal to 1000 when an estimate of confidence interval around T_n is required.

Step 5. Construct the relative frequency histogram from the B number of T_n^*'s