EMPIRICAL LIKELIHOOD METHOD

by

Song Xi Chen

A thesis submitted for the degree of

Doctor of Philosophy

of

The Australian National University

Centre for Mathematics and Its Applications

November 1992

Declaration

I hereby declare that, apart from Chapter 1, this thesis describes my own original work, supervised by Professor P. G. Hall and published jointly with him or myself.

Song Xi Chen

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Peter Hall for his excellent supervision and insight. He introduced me to the subject treated in this thesis, and as a whole opened my eyes to modern techniques of statistics. His encouragements and constructive criticism were important stimuli for this thesis. I would like to thank Drs. Andrew Wood and Daryl Dayley of ANU, and Professor Art Owen of Stanford University for beneficial discussions and encouragements. A special thanks goes to my fellow student Robert Murison for his friendship and tennis games which I enjoyed so much during my two years stay in Canberra. I also would like to thank the people at the Centre for Mathematics and Its Applications and in both Statistics Departments of ANU for their support and assistance, which make my time at ANU the most enjoyable.

I am greatly indebted to my wife, Ling, for her continued help, encouragement and love over the last two years. It is never easy to be a Ph.D student's wife. I am also indebted to both families of us in China, for their endless love and moral support.

Finally, I would like to dedicate this thesis to my grandmother, Nai-Nai, for her 30 years constant love and bringing me up. I hope this could compensate some of my regrets of not being able to go back home to show my filial piety to her when she passed away from us in November 1991 in Beijing, at aged of 82.

Related Publications

The following papers have been submitted for publication from the work contained in this thesis:

CHEN, S.X. and HALL, P. (1991). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* to appear.

CHEN, S.X. (1991). On the accuracy of empirical likelihood confidence regions for linear regression model. *Ann. Inst. Statist. Math.* to appear.

CHEN, S.X. (1992a). Empirical likelihood confidence intervals for linear regression coefficients. Submitted to *Journal of Multivariate Analysis.*, May, 1992.

CHEN, S.X. (1992b). Comparing empirical likelihood and bootstrap hypothesis tests. Submitted to *Journal of Multivariate Analysis.*, August 1992.

Abstract

In this thesis we consider using empirical likelihood method for constructing nonparametric confidence regions (or intervals) and doing test in wide range of situations. Empirical likelihood, introduced by Owen (1988, 1990), is a nonparametric method of inference with sampling properties similar to those of the bootstrap. However, instead of assigning equal probabilities n^{-1} to all data values, empirical likelihood places arbitrary probabilities on the data points, say p_i on the i'th data value. The weights p_i are chosen by profiling a multinomial likelihood supported on the sample, and empirical likelihood confidence regions are constructed by contouring this multinomial likelihood. An attractive feature of empirical likelihood is that it produces confidence regions whose shapes and orientations are determined entirely by the data, and which have coverage accuracy at least comparable with those of bootstrap confidence regions. In Chapter 1 of this thesis we review the concepts of empirical likelihood and its developments. We also outline the notions of Edgeworth expansion which is an important tool of studying the coverage properties of empirical likelihood confidence regions.

However, so far all the work done on empirical likelihood is confined to the so called standard case where the parameter of interest is a smooth function of means and the sample is independent and identically distributed random vectors. The main motivation of this thesis is to establish the theory of empirical likelihood for other cases. In chapter 2 we consider constructing confidence intervals for population quantiles, which cannot be represented as a smooth function of means. We show that standard empirical likelihood confidence intervals for quantiles are identical to sign-test intervals. They have relatively large coverage error, of size $n^{-\frac{1}{2}}$, even though they are two-sided intervals. We show that smoothed empirical likelihood confidence intervals for quantiles have coverage error of order n^{-1}, and may be

Bartlett correctioned to produce intervals with an error order of only n^{-2}. Necessary and sufficient conditions on the smoothing parameter, in order for these sizes of error to be attained, are derived.

In Chapter 3 we consider the second non-standard case, which is to construct empirical likelihood confidence region for the regression coefficient vector β of a linear regression model $Y_i = x_i\beta + \epsilon_i$, $1 \leq i \leq n$. Due to the presence of the fixed design points, the observed random variables are independent but not identically distributed. So it is not the standard independent and identically distributed random sample case any more. Empirical likelihood methods were proposed by Owen (1991) for constructing confidence regions for β in the model (3.1.1). He derived a nonparametric version of Wilks' theorem, ensuring that empirical likelihood confidence regions for β have correct asymptotic coverages. We show that coverage errors of the empirical likelihood confidence regions for β are of order n^{-1}. Bartlett corrections may be employed to reduce the coverage errors to $O(n^{-2})$. For practical implementation of Bartlett correction, we also give an empirical Bartlett correction.

It is not enough to just construct confidence regions for β of a linear regression model. In practice, statisticians are often confronted with problems of constructing confidence intervals for a particular regression coefficient or for certain linear combinations of β. In Chapter 4 we address the above problem under the simple linear regression model: $y_i = a_o + b_o x_i + \epsilon_i, 1 \leq i \leq n$. Nonparametric versions of Wilks' theorem are proved for empirical likelihood of the slope parameter b_o and mean parameter $y_o = a_o + b_o\, x_o$ for any fised x_o, which enable us to construct empirical likelihood confidence intervals for these parameters. We also show that coverage errors of these confidence intervals are of order n^{-1} and can be reduced to order n^{-2} by Bartlett correction.

We see that almost all the work done on empirical likelihood concentrate on constructing confidence regions. After constructing an empirical likelihood confidence region, we can derive an empirical likelihood test about the parameter of interested by the duality between the confidence region and hypothesis test. However, so far little has been done on the aspects of power of empirical likelihood test. Surprisingly, no much has been done for that of a bootstrap test either! The contribution of Chapter 5 is developing high order expansions for the power function of the empirical likelihood and the bootstrap test for a mean against a series of local alternatives. A comparison between the empirical likelihood and the bootstrap tests for a mean parameter against a series of local alternative hypotheses is made. For univariate and bivariate cases, practical rules are proposed for choosing the more powerful test.

TABLES OF CONTENTS

CHAPTER THREE
ON THE COVERAGE OF EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR LINEAR REGRESSION MODEL

CHAPTER FOUR
EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR LINEAR REGRESSION COEFFICIENTS

CHAPTER FIVE
COMPARING EMPIRICAL LIKELIHOOD AND BOOTSTRAP HYPOTHESIS TESTS

獻給我奶奶

To My Grandmother

道可道，非常道。
名可名，非常名。
無名天地之始，
有名萬物之母。

The tao that can be said is not the everlasting Tao.

If a name can be named, it is not the everlasting Name.

That which has no name is the origin of heaven and earth;

That which has a name is the Mother of all things..

LAO TZU: "Tao Deh Ching"

CHAPTER 1

CONCEPTS OF EMPIRICAL LIKELIHOOD AND EDGEWORTH EXPANSIONS

1.1 Introduction

The coming of the computer age in the last a few decades has deeply changed the shapes and the ways of thinking of the centuries old displine called Statistics. The most notable event was the birth of the bootstrap method in 1979 by Efron (1979) and the following works done in 1980s, leading the bootstrap becoming a mature general statistical procedure with wide range of applications. Hall (1992) gave a full description of the development and theory of the bootstrap. An important feature of the bootstrap is the idea of "resampling". In the pre-bootstrap era, statisticians depended heavily on the Central Limit Theorem, which gives a normal approximation to a statistic of interest. However, this approximation only takes into account the first two moments without taking care of skewness and kurtosis of the statistic, causing that the accuracy of the approximation is only of first order. By generating a large number of resamples out of the original sample in a computer, the bootstrap implicitly corrects skewness and kurtosis during the resampling procedure. This leads to a more accurate approximation to the distribution of the statistic.

Empirical likelihood is another computer intensive method introduced by Owen (1988,1990). It constructs a likelihood function for a parameter of interest in a nonparametric setting, and uses the later to set up confidence regions for that parameter. Empirical likelihood has sampling properties similar to those of the bootstrap. However, instead of assigning equal probabilities n^{-1} to all data values, it places arbitrary probabilities on the data points, say p_i on the i'th data value. The weights,

p_i, are chosen by profiling a multinomial likelihood supported on the sample. Empirical likelihood confidence regions are constructed by contouring this multinomial likelihood. As mentioned by Owen (1988), the preliminary idea of empirical likelihood was used by Thomas and Grunkemeier (1975) to construct confidence intervals for survival probability. Those authors show that the confidence intervals have the desired property of respecting range, which is not generally held by normal approximation based methods. However, Owen was the first to systematically demonstrate that the idea has very wide range of applications.

It has been shown that nonparametric versions of Wilks' theorem and Bartlett correction hold true for empirical likelihood in a wide range of situations, akin to the usual parametric likelihood. However, compared with parametric likelihood, empirical likelihood is robust since it is constructed in a way which does not assume the form of the distribution. Compared with the bootstrap, empirical likelihood has several advantages. Hall and La Scala (1990) have identified the following attributes:

(1) Empirical likelihood enable the shape and orientation of a confidence region to be determined "automatically" by the sample, whereas construction of a multivariate bootstrap confidence region requires a decision on how the region should be shaped and oriented since the bootstrap itself cannot provide an answer to any of these. It can be very hard to decide whether to use an elliptical or rectangular confidence region.

(2) Empirical likelihood confidence regions are Bartlett correctionable, meaning that a simple adjustment for scale reduces the order of magnitude of coverage error from n^{-1} to n^{-2}, where n denotes sample size. (See DiCiccio, Hall and Romano 1991.) The bootstrap confidence region can attain the same order of coverage accuracy. However, the bootstrap achieves this at expensive of enormous computer hours.

(3) Empirical likelihood implicitly use the true scale parameter to construct

confidence regions, which consequently avoids the problem of estimating the scale. The bootstrap depends heavily on a stable estimate of the scale parameter, which can be very hard to be obtained especially for cases like correlation coefficient and ratio of means.

(4) Empirical likelihood confidence regions are range respecting, as was noticed by Thomas and Grunkemeier (1975). For example, the empirical likelihood confidence region for a correlation coefficient always lies within interval $(-1, 1)$. However, this property is not necessarily preserved by a bootstrap confidence region; consider for example confidence regions constructed using the percentile-t method.

The basic concepts of empirical likelihood are given in Section 1.2, together with some fundamental formulae and expansions used by empirical likelihood. In Section 3.1 we display some existing results of Edgeworth expansion, which will be the basic tool used in this thesis to study coverage accuracy and Bartlett correction of empirical likelihood confidence region. We provide an outlines of this thesis in Section 1.4

1.2 Concepts of Empirical likelihood

In this section we describe the basic concept of empirical likelihood. Suppose $X_1, \cdots, X_n$ are p-dimensional independent and identical distributed (i.i.d.) random vectors from unknown distribution F. Let $\theta = \theta(F)$ denote some characteristic of F, such as mean, variance etc, for which we want to construct a confidence region (or interval). Write $p_1, p_2, \cdots, p_n$ for nonnegative numbers adding to unity, and $\theta(p)$ for the value of θ when the distribution function F is replaced by

$$\hat{F}_p(x) = \sum_{i=1}^{n} p_i \, I(X_i \leq x),$$

where I is the indicator function. We can view $\hat{F}_p(x)$ as weighted empirical distribution function. For instance if θ denote the population mean, that is $\theta = \int x d\, F(x)$, then

$$\theta(p) = \int x d\, \hat{F}_p(x) = \sum_{i=1}^{n} p_i \, X_i.$$

The empirical likelihood for θ, evaluated at $\theta = \theta_1$, is defined to be

$$L(\theta_1) = \max_{\theta(p)=\theta_1, \sum p_i = 1} \prod_{i=1}^{n} p_i .$$

If we impose only one constraint, $\sum p_i = 1$, while maximizing $\prod_{i=1}^{n} p_i$, we get $p_i = n^{-1}$ for $i = 1, \cdots, n$, which gives us the bootstrap estimate $\hat{\theta} = \theta(\hat{F})$ for θ, where

$$\hat{F}(x) = \sum_{i=1}^{n} n^{-1} I(X_i \leq x)$$

is the empirical distribution function. Thus, we have

$$L(\hat{\theta}) = n^{-n}.$$

Now the empirical log-likelihood ratio , evaluated at $\theta = \theta_1$, is defined as

$$\begin{aligned} \ell(\theta_1) &= -2 \log\{L(\theta_1)/L(\hat{\theta})\} \\ &= -2 \min_{\theta(p)=\theta_1, \sum p_i = 1} \sum_{i=1}^{n} \log(n\, p_i). \end{aligned} \tag{1.2.1}$$

It is well-known that under certain regularity conditions, the usual parametric log-likelihood ratio has the following properties: (1) it converges in distribution to χ_p^2, the chi-square distribution with p degrees of freedom, as sample size n approaches to infinity; this is Wilks' theorem (Wilks 1938); (2) it is Bartlett correctable (Bartlett 1937, Lawley 1956). Wilks' theorem enable us to construct confidence regions by looking up the χ_p^2 tables, and Bartlett correction can be used to improve the coverage accuracy of the confidence region by simple adjustment to the mean of the log-likelihood ratio statistic.

Do these two properties hold true for the empirical likelihood ? Owen (1988, 1990) proved a nonparametric version of Wilks' theorem for empirical likelihood of mean. DiCiccio, Hall and Romano (1991) extended it to the case of a smooth function of means and proved the validity of Bartlett correction for this general case. Owen (1991) established a nonparametric version of Wilks' Theorem for the regression coefficient vector of a linear regression model. In the regression case, the random vectors involved are independent but not identically distributed. This is due to the presence of the fixed design points. We should mention here that all the results except the regression case listed above were proved under the following regularity conditions:

$$\begin{aligned}&\text{(i) } \Sigma = \mathrm{Cov}(X_1) \text{ is positive definite matrix; (ii) } E\|X_1\|^s < \infty;\\ &\text{(iii) for every positive b, the characteristic function } g \text{ of } X_1 \text{ satisfies}\\ &\text{Cramér's condition } \sup_{\|t\|>b} |g(t)| < 1,\end{aligned} \tag{1.2.2}$$

where $s = 5$ for the mean case and $s = 15$ for the smooth function of means case. The regularity conditions assumed for the regression case are described in Chapter 3.

In the rest of this section we give some basic formulae and algorithms for the case of $\theta = \mu = \int x\, dF(x)$, which has been used by previous authors and will be referenced repeatedly in this thesis for constructing empirical likelihood confidence regions in other situations.

According to (1.2.1), the empirical log-likelihood ratio for $\theta = \mu$, evaluated at $\mu = \mu_1$, is

$$\ell(\mu_1) = -2 \min_{\sum p_i X_i = \mu_1, \sum p_i = 1} \sum_{i=1}^{n} log(n\, p_i). \tag{1.2.3}$$

Using the Lagrange multiplier method to solve the above optimalization problem (1.2.3), it turns out that the optimal p_i's have the following form:

$$p_i = \frac{1}{n} \frac{1}{1 + t^T (X_i - \mu)}, \quad 1 \le i \le n, \tag{1.2.4}$$

where $t = (t_1, \cdots, t_p)^T$ satisfies

$$n^{-1} \sum \frac{X_i - \mu}{1 + t^T (X_i - \mu)} = 0. \tag{1.2.5}$$

Substituting (1.2.4) into (1.2.1) we obtain

$$\ell(\mu) = 2 \sum_{i=1}^{n} log\{1 + t^T (X_i - \mu)\}. \tag{1.2.6}$$

Let $\Sigma = \mathrm{Cov}(X_1)$, $z_i = \Sigma^{-1/2}(X_i - \mu)$ and z_i^j be the j'th component of p-dimensional vector z_i. Then using the new standardized variable z_i's, (1.2.5) and (1.2.6) become

$$\ell(\mu) = 2\sum_{i=1}^{n} log\{1 + \lambda^T z_i\}. \tag{1.2.7}$$

where $\lambda = (\lambda_1, \cdots, \lambda_p)^T = \Sigma^{\frac{1}{2}} t$ satisfies

$$n^{-1}\sum \frac{z_i}{1 + \lambda^T z_i} = 0. \tag{1.2.8}$$

Since analytic solution for λ in (1.2.8) is not attainable, we have to resort to expansion. Before doing that, let us first define

$$\begin{aligned} \alpha^{j_1 \cdots j_k} &= E\left(z_i^{j_1} \cdots z_i^{j_k}\right), \\ A^{j_1 \cdots j_k} &= n^{-1}\sum_{i=1}^{n} z_i^{j_1} \cdots z_i^{j_k} - \alpha^{j_1 \cdots j_k}. \end{aligned} \tag{1.2.9}$$

We see $\alpha^{j_1 \cdots j_k}$ is a k'th order multivariate moment of Z_i and $A^{j_1 \cdots j_k}$ is a k'th order central multivariate mean of Z_i's.

Owen (1990) set up an one-term Taylor expansion for $\ell(\mu)$:

$$\ell(\mu) = n A^j A^j + O_p(n^{-1/2}). \tag{1.2.10}$$

Throughout this thesis we use the summation convention that terms with repeated indices are to be summed over. From (1.2.10) we are able to prove the following nonparametric version of Wilks' theorem by assuming condition (i) of (1.2.2),

$$\ell(\mu) \xrightarrow{d} \chi_p^2, \qquad \text{as } n \to \infty, \tag{1.2.11}$$

since $\sqrt{n}\, A = (A^1, \cdots, A^p)$ converges to $N(0, I_p)$ in distribution by the Central Limit Theorem, where I_p is the p-dimensional identity matrix. An α-level confidence region for μ can be constructed in the following way. First find from the χ_p^2 tables the value c_α such that

$$P\left(\chi_p^2 < c_\alpha\right) = \alpha.$$

Then $I_\alpha = \{\beta \,|\, \ell(\beta) < c_\alpha\}$ is the α-level confidence region for μ, and (1.2.11) ensures that it has asymptotically correct coverage. To investigate the coverage accuracy of I_α, an Edgeworth expansion for the distribution function of $\ell(\mu)$ has to be developed. After Taylor expansion, as has been shown by DiCiccio, Hall and Romano (1988), λ has the following expansion:

$$\begin{aligned}\lambda^j &= A^j - A^{jk} A^k + \alpha^{jkl} A^k A^l + A^{jl} A^{kl} A^k + A^{jkl} A^k A^l \\ &\quad - \alpha^{klm} A^{jm} A^k A^l - 2\alpha^{jkm} A^{lm} A^k A^l + 2\alpha^{jkn} \alpha^{lmn} A^k A^l A^m \qquad (1.2.12)\\ &\quad - \alpha^{jklm} A^k A^l A^l + O_p(n^{-2}).\end{aligned}$$

Substituting (1.2.12) into (1.2.13), we obtain

$$\begin{aligned}n^{-1}\ell(\mu) &= A^j A^j - A^{jk} A^j A^k + \tfrac{2}{3}\alpha^{jkl} A^j A^k A^l + A^{jl} A^{kl} A^j A^K \\ &\quad + \tfrac{2}{3} A^{jkl} A^j A^k A^l - 2\alpha^{jkm} A^{lm} A^j A^k A^l + \alpha^{jkn}\alpha^{lmn} A^j A^k A^l A^m \\ &\quad - \tfrac{1}{2}\alpha^{jklm} A^j A^k A^l A^m + O_p(n^{-5/2}). \qquad (1.2.13)\end{aligned}$$

A signed root decomposition for $\ell(\mu)$ can be derived from (1.2.13), that is,

$$\ell(\beta) = \left(n^{1/2} R^T\right)\left(n^{1/2} R\right) + O_p(n^{-3/2}), \qquad (1.2.14)$$

where $R = R_1 + R_2 + R_3$ is a p-dimensional vector and $R_l = O_p\left(n^{-l/2}\right)$ for $l = 1, 2, 3$. Comparing terms in (1.2.13) with those in (1.2.14) yields,

$$R_1^j = A^j,$$

$$R_2^j = -\tfrac{1}{2} A^{jk} A^k + \tfrac{1}{3}\alpha^{jkm} A^k A^m \quad \text{and}$$

$$\begin{aligned}R_3^j &= \tfrac{3}{8} A^{jm} A^{km} A^k + \tfrac{1}{3} A^{jkm} A^k A^l - \tfrac{5}{12}\alpha^{jkm} A^{lm} A^k A^l \\ &\quad - \tfrac{5}{12}\alpha^{klm} A^{jm} A^k A^l + \tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} A^m A^k A^l - \tfrac{1}{4}\alpha^{jklm} A^m A^k A^l,\end{aligned}$$

where R_l^j is the j'th component of R_l. Notice that there exists a smooth function h_o such that $R = h_o(\overline{U_o})$, where $\overline{U_o} = (A^1, \cdots, A^p, A^{11}, \cdots A^{pp}, A^{111}, \cdots, A^{ppp})^T$ is a mean of i.i.d. random vectors. So, R is a smooth function of i.i.d. means. Thus, after calculating joint cumulants of R, and using the valid Edgeworth expansion developed by Bhattacharya and Ghosh (1978) for this case, it can be shown that under condition (1.2.2) for any $x > 0$,

$$P\{\ell(\mu) < x\} = P(\chi_p^2 < x) - \beta_0\, x\, g_p(x)\, n^{-1} + O(n^{-2}),$$

where g_p is the density of the χ_p^2 distribution and

$$\beta_0 = p^{-1}\left(\tfrac{1}{2}\alpha^{jjmm} - \tfrac{1}{3}\alpha^{jkm}\alpha^{jkm}\right). \tag{1.2.15}$$

This implies that

$$P(\mu \in I_\alpha) = \alpha - \beta_o\, c_\alpha\, g_p(c_\alpha)\, n^{-1} + O(n^{-2}),$$

which means that the coverage accuracy of the empirical likelihood confidence region I_α is of order n^{-1}. We know that in the parametric case, part of the coverage error of a confidence region constructed by the log-likelihood ratio method is due to the mean of the log-likelihood ratio not being equal to p, which is the mean of the χ_p^2 distribution. Bartlett correction can be used to improve the coverage accuracy by readjusting the mean of the log-likelihood ratio. For the case of smooth function of means, DiCiccio, Hall and Romano (1991) showed that the empirical likelihood confidence region is Bartlett correctable, which implies that a simple adjustment for the mean can reduce the coverage error from order n^{-1} to order n^{-2}. From expansion (1.2.13) the above authors showed that

$$E\{\ell(\beta)\} = p(1 + \beta_0 n^{-1}) + O(n^{-2}),$$

where β_0 is the Bartlett factor given by (1.2.15). It can be shown that

$$P\{\ell(\mu) < c_\alpha\left(1 + \zeta_0 n^{-1}\right)\} = \alpha + O(n^{-2}), \tag{1.2.16}$$

where ζ_0 is either β_0 or a root-n consistent estimate of β_0. From (1.2.16), we can correct the confidence region I_α by defining

$$I_\alpha^c = \{\mu \,|\, \ell(\mu) < c_\alpha\left(1 + \zeta_0 n^{-1}\right)\},$$

where (1.2.16) shows that the corrected confidence region I_α^c has coverage error of order n^{-2}.

1.3 Edgeworth Expansions

From Section 1.2 we see that Edgeworth expansion plays an important role for determining the coverage accuracy and availability of Bartlett correction for empirical likelihood confidence regions. So it is worthwhile to devote this section on it. In this section we display some existing results on Edgeworth expansions, for example as described in Bhattacharya and Rao (1976) and Bhattacharya and Ghosh (1978). These results will be used repeatedly in this thesis to derive asymptotic expansions of distributions of the empirical log-likelihood ratio statistics in various situations. In particular, we are interested in Edgeworth expansions for distributions of smooth function of a mean, where the mean could be an average of either i.i.d. random vectors or independent but not identically distributed random vectors. Before doing this we give some notation.

Let F be the distribution function of a random vector $X \in \mathbf{R}^k$ with characteristic function φ. If $\int \|x\|^s dF(x) < \infty$, we may have the following Taylor expansion

$$\log\{\varphi(t)\} = \sum_{|v|\leq s} \chi_v (it)^v / v! + o(\|t\|^s), \quad \text{as} \quad t \to 0, \tag{1.3.1}$$

where $t = (t_1, \cdots, t_k)$ and $v = (v_1, \cdots, v_k)$ is a nonnegative vector of integers with operations $|v| = \sum_{i=1}^k v_i$ and $(it)^v = (it_1)^{v_1} \cdots (it_k)^{v_k}$. The coefficient χ_v appearing in (1.3.1) is called the v'th cumulant of F. For a given set of χ_v, we define polynomials

$$\chi_l(z) = l! \sum_{|v|=l} \frac{\chi_v}{v!} z^v$$

for any positive integer l, where $z^v = z_1^{v_1} \cdots z_k^{v_k}$ for $z = (z_1, \cdots, z_k) \in \mathbf{R}^k$. Moreover, we define polynomials $\tilde{P}_s(z : \{\chi_v\})$ by the following formal equation in a real variable u,

$$1 + \sum_{s=1}^{\infty} \tilde{P}_s(z : \{\chi_v\}) u^s = 1 + \sum_{l=1}^{\infty} \frac{1}{l!} \{\sum_{s=1}^{\infty} \frac{\chi_{s+2}(z)}{(s+2)!} u^s\}^l.$$

Let $V = \mathrm{Cov}(X)$, $\phi_{0,V}$ and $\Phi_{0,V}$ be the normal density and distribution functions in $\mathbf{R}^k$ with zero mean and covariance matrix V respectively, and put

$$D^v \phi_{0,V} = \frac{\partial^{v_1}}{(\partial X_1)^{v_1}} \phi_{0,V}(x) \cdots \frac{\partial^{v_k}}{(\partial X_k)^{v_k}} \phi_{0,V}(x).$$

We define a function $P_r(-\phi_{0,V} : \{\chi_v\})$ by formally replacing $(it)^v$ in the polynomial $\tilde{P}_r(z : \{\chi_v\})$ with $(-1)^{|v|} D^v \phi_{0,V}$, that is

$$P_r(-\phi_{0,V} : \{\chi_v\}) = \tilde{P}_r(-D : \{\chi_v\})\phi_{0,V}.$$

Furthermore, let $P_r(-\Phi_{0,V} : \{\chi_v\})$ be the finite signed measure on $\mathbf{R}^k$ with density $Pr(-\phi_{0,V} : \{\chi_v\})$.

1.3.1 Edgeworth Expansions for i.i.d. Case

Suppose $X_1, \cdots, X_n$ are i.i.d random vectors drawn from distribution F with mean μ, covariance matrix V and characteristic function φ. Let

$$\overline{W} = n^{-1/2} \sum_{i=1}^{n} (X_i - \mu).$$

Then we have the following theorem due to Esseen (1945) and Bhattacharya (1968):

Theorem 1.3.1 *Assume that F has finite s'th absolute moment for some integer $s \geq 3$, and satisfies the Cramér's condition $sup_{\|t\|>b}|\varphi(t)| < 1$ for any positive b. Then,*

$$\sup_{B \in \mathcal{B}} |P(W \in B) - \sum_{r=0}^{s-2} n^{-r/2} P_r(-\Phi_{0,V} : \{\chi_v\})(B)| = o(n^{-(s-2)/2}), \qquad (1.3.2)$$

where $\mathcal{B}$ is any class of Borel sets satisfying

$$\sup_{B \in \mathcal{B}} \int_{(\partial B)^\epsilon} \phi_{0,V}(v) d v = O(\epsilon), \qquad \epsilon \downarrow 0, \qquad (1.3.3)$$

and ∂B and $(\partial B)^\epsilon$ are the boundary of B and ϵ-neighborhood of ∂B respectively.

Let $f_1, \cdots, f_m$ be real-valued Borel measurable functions on $\mathbf{R}^k$, h be a smooth real-valued function on $\mathbf{R}^m$, and $Q_i = (f_1(X_i), \cdots, f_m(X_i))$ for $1 \leq i \leq n$. Consider a statistic

$$T_n = n^{1/2}\{h(\overline{Q}) - h(\mu_q)\},$$

where $\overline{Q} = n^{-1} \sum_{i=1}^{n} Q_i$ and $\mu_q = E(Q_1)$. Clearly T_n is a smooth function of $\overline{Q}$. Put

$$d_{i_1,\cdots,i_k} = (D_{i_1} \cdots D_{i_k} h)(\mu_q), \quad 1 \leq i_1, \cdots, i_k \leq m,$$

as derivative of h, where D_{i_k} denotes differentiation with respect to the i_k'th coordinate. Furthermore, we define

$$\sigma_q^2 = (d_1, \cdots, d_k) V (d_1, \cdots, d_k)^T .$$

Consider the following Taylor expansion of T_n around μ_q,

$$T_n^{'} = n^{1/2} \sum_{k=1}^{s-1} \sum_{i_1, \cdots, i_k} (k!)^{-1} d_{i_1, \cdots, i_k} (\overline{Q}^{i_1} - \mu_q^{i_1}) \cdots (\overline{Q}^{i_k} - \mu_q^{i_k}),$$

where $\overline{Q}^{i_k}$ and $\mu_q^{i_k}$ are the i_k'th components of $\overline{Q}$ and μ_q respectively. Using the delta method we may expect that the Edgeworth expansion of the distribution of T_n and $T_n^{'}$ generally disagree only in terms of order $n^{-(s-2)/2}$ or smaller, i.e

$$P(T_n \leq x) = P(T_n^{'} \leq x) + o(n^{-(s-2)/2}),$$

since $T_n - T_n^{'} = o_p(n^{-(s-2)/2})$. Now, the cumulants of $T_n^{'}$ are much easier calculated than those of T_n, since $T_n^{'}$ is a multivariate polynomial in $\overline{Q} - \mu_q$. If Q_1 has sufficiently many moments then, as shown by James and Mayne (1962), the j'th cumulant $k_{j,n}$ of $T_n^{'}$ is given by

$$k_{j,n} = \tilde{k}_{j,n} + o(n^{-(s-2)/2}),$$

where $\tilde{k}_{j,n}$ is an "approximate cumulant" of $T_n^{'}$, having the form

$$\tilde{k}_{j,n} = \begin{cases} \sum_{i=1}^{s-2} n^{-i/2} b_{ji} & \text{if } j \neq 2; \\ \sigma_q^2 + \sum_{i=1}^{s-2} n^{-i/2} b_{2i} & \text{if } j = 2, \end{cases}$$

and $b_{j,i}$'s depend only on the moments of Z_1 and on derivatives of h at μ_q. The characteristic function of $T_n^{'}$ (or T_n) can be approximated by

$$\Upsilon_n(t) = \exp\{it\, \tilde{k}_{1,n} + \frac{(it)^2}{2!}(\tilde{k}_{2,n} - \sigma_q^2) + \sum_{j=3}^{s} \frac{(it)^j}{j!} \tilde{k}_{j,n}\} \exp(-\sigma^2 t^2/2). \qquad (1.3.4)$$

After expanding the first exponential factor in (1.3.4), we obtain

$$\Upsilon_n(t) = \exp(-\sigma^2 t^2/2)\{1 + \sum_{r=1}^{s-2} n^{-r/2} \pi_r(it)\} + o(n^{-(s-2)/2}),$$

where π_r's are polynomials whose coefficients do not depend on n. Then, the formal Edgeworth expansion $\Psi_{s,n}$ of the distribution of T_n is defined as

$$\Psi_{s,n}(u) = \int_{-\infty}^{u} \psi_{s,n}(v)dv,$$

where

$$\psi_{s,n} = \{1 + \sum_{r=1}^{s-2} n^{-r/2}\, \pi_r(-d/dv)\}\, \phi_{\sigma_q^2}(v).$$

Bhattacharya and Ghosh (1978) proved that $\Psi_{s,n}$ is a valid Edgeworth expansion of the distribution of T_n. Part of their results are stated in the following theorem:

Theorem 1.3.2 *Assume that (i) h has continous derivatives up to order $s \geq 3$ in a neighborhood of μ_q; (ii) $E|Q_1|^s$ is finite; (iii) Cramér's condition holds for Q_1, that is* $\limsup_{|t|\to\infty} |E\{\exp(i < t, Q_1 >)\}| < 1$, *where $< >$ denotes the Euclidean inner product on $\mathbf{R}^k$. Then,*

$$\sup_{B \in \mathcal{B}} |P(T_n \in B) - \int_B \psi_{s,n}(v)dv| = o(n^{-(s-2)/2})$$

uniformly holds over the class of $\mathcal{B}$ satisfying (1.3.3).

1.3.2 Edgeworth Expansion for a non-i.i.d case

In this subsection we display result for setting up Edgeworth expansion for a non-i.i.d case. The case we consider is that the sample $X_1, \cdots, X_n$ are independent but not necessarily identically distributed random vectors in $\mathbf{R}^k$. This is just the situation of a linear regression model, where the presence of the fixed design points makes the response random variables are independent but not identically distributed.

Let $X_1, \cdots, X_n$ be independent random vectors of $\mathbf{R}^k$, with mean zero and finite s'th absolute moments for some integer $s \geq 3$ for each X_i $1 \leq i \leq n$. Define $V_n = n^{-1} \sum_{i=1}^n \mathrm{Cov}(X_i)$, $v_{k,n}$ be the smallest eigenvalues of V_n, and $\overline{\chi}_{j,n}$ is the average j'th cumulant of $V_n^{-1/2} X_i$ for $1 \leq i \leq n$ where $V_n^{-1/2}$ is the inverse of the square root matrix of V_n. Furthermore, put $S_n = n^{-1/2} V_n^{-1/2} \sum_{i=1}^n X_i$. The following theorem, as an extension of Theorem 1.3.1, is an direct corollary of Theorem 20.6 of Bhattacharya and Rao (1976).

Theorem 1.3.3 *Assume that:*

(i) $v_{k,n}$ *is uniformly bounded away from zero; (ii) the average s-th absolute moments* $n^{-1}\sum_{i=1}^{n} E(\|X_i\|)^s$ *are bounded away from infinity for* $s \geq 3$*;*

(iii) for each positive ϵ, $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\int_{\|X_j\|>\epsilon n^{1/2}} \|X_i\|^s = 0$*; (iv) the* (1.3.5)

characteristic functions g_n *of* X_n *satisfies Cramér's condition*

$\limsup_{n\to\infty} \sup_{\|t\|>b} |g_n(t)| < 1,$ *for every positive b.*

Then

$$\sup_{B\in\mathcal{B}} |P(S_n \in B) - \sum_{r=0}^{s-2} n^{-r/2} P_r(-\Phi_{0,V} : \{\overline{\chi}_{j,n}\})(B)| = o(n^{-(s-2)/2}),$$

over the class of $\mathcal{B}$ *satisfying (1.3.3).*

1.3.3 Transformation of Edgeworth Expansion

We show in Theorem 1.3.2 that the Edgeworth expansion for the distribution of an i.i.d. mean can be transformed by a smooth function to yield another valid Edgeworth expansion. We also show that this expansion may be calculated from the cumulants obtained by using the delta method, i.e the cumulants formally calculated from a Taylor expansion omitting terms of higher order. Skovgaard (1981) generalized the above result of Bhattacharya and Ghosh (1978). He demonstrated, using the delta method, that *any* (not just for i.i.d. mean) valid Edgeworth expansion may be transformed by a *sequence* (not just a single smooth function) of sufficiently smooth functions to get another valid Edgeworth expansion.

Consider a statistic $U_n = (U_n^1, \cdots, U_n^k)$ with zero mean and unit variance on $\mathbf{R}^k$, constructed from a sample of size n. Suppose that for some $s \geq 3$, $E(\|U_n\|^s) < \infty$, and that there has been a valid Edgeworth expansion of the distribution of U_n of the form

$$P(U_n \in B) = \int_B \xi_n \, du + o(\beta_{s,n}),$$

uniformly in $B \in \mathcal{B}$ satisfying (1.3.3), where

$$\xi_n = \sum_{r=o}^{s-2} P_r(-\phi_{0,I_k} : \{\chi_{v,n}\})(u),$$

$$\beta_{s,n} = \left[\sup\{\, \|\chi_{v,n}\|^{1/(|v|-2)} \,|\, 3 \le |v| \le s \,\}\right]^{s-2} = o(1),$$

and $\{\chi_{v,n},\ 3 \le |v| \le s\}$ are the cumulants of U_n. Clearly we have $\chi_{v,n} = 0$ for $|v| = 1$, since $E(U_n) = 0$. When U_n is a normalized sum of independent and identical distributed random vectors, we have $\beta_{s,n} = O(n^{-(s-2)/2})$ as $\chi_{v,n} = O(n^{-(|v|-2)/2})$.

Let $\{h_n\}$ be a sequence of functions mapping $\mathbf{R}^k$ into $\mathbf{R}^m$ for $m \le k$. For each n, h_n is p-times differentiable at zero ($p \ge 2$) and satisfying $h_n(0) = 0$ and the Jacobian matrix of h_n at zero, say $Dh_n(0)$, is of rank m. Put

$$B_n = \{Dh_n(0)\}\{Dh_n(0)\}^T, \quad \text{and} \quad f_n = B_n^{-1} h_n,$$

so that $f_n(U_n)$ has asymptotic variance I_m. We shall show that under certain conditions on the smoothness of h_n, a valid Edgeworth expansion of the distribution of $f_n(U_n)$ may be established from the approximate cumulants of $f_n(U_n)$ obtained by using the delta method. Let

$$\tilde{d}_{i_1,\cdots,i_l} = (D_{i_1} \cdots D_{i_l} f_n)(0), \quad 1 \le i_1, \cdots, i_l \le k.$$

Taylor expanding $f_n(U_n)$ around zero, we have

$$Y_n = \sum_{l=1}^{p-1} \sum_{i_1,\cdots,i_l} (l!)^{-1} \tilde{d}_{i_1,\cdots,i_l} U_n^{i_1} \cdots U_n^{i_l}.$$

The moments of Y_n are much easier to calculate than those of $f_n(U_n)$, since Y_n is a polynomial in U_n. Let $\{\eta_{v,n}\}$, $1 \le |v| \le q$, be the first q'th order cumulants of Y_n, computed from the first q'th order moments of Y_n. Notice that the computations of the first q'th order moments of Y_n may involve moments of U_n of higher orders than s, which may not exist. To solve this problem, we define the formal cumulants of U_n, say $\{\chi^*_{v,n}\}$, such that

$$\chi^*_{v,n} = \begin{cases} \chi_{v,n}, & \text{if } |v| \leq s; \\ 0, & \text{if } |v| \geq s. \end{cases}$$

By the well-known formulae connecting moments and cumulants we are able to define the formal moments of U_n, which will be used to calculate the moments and the cumulants of Y_n.

Neglecting the terms at smaller order of $\beta_{s,n}$ in $\eta_{v,n}$ we obtain $\tilde{\eta}_{v,n}$, the approximate cumulants of Y_n, where

$$\eta_{v,n} = \tilde{\eta}_{v,n} + o(\beta_{s,n}).$$

Let ζ_n be the density of the finite signed measure with characteristic function

$$\hat{\zeta}_n = \exp(i < t, \tilde{\eta}_{1,n} > -\tfrac{1}{2}\tilde{\eta}_{2,n}\, \|t\|^2) \textstyle\sum_{r=0}^{q-2} \tilde{P}_r(it : \{\tilde{\eta}_{v,n}\})$$

where $\tilde{\eta}_{1,n}$ is an m-dimensional vector consisting of all $\tilde{\eta}_{v,n}$ with $|v| = 1$, $\tilde{\eta}_{2,n}$ is an $m \times m$ matrix with all $\tilde{\eta}_{v,n}$ $|v| = 2$ as its elements, and $<>$ denotes the Euclidean inner product of vectors.

Now the problem becomes how to choose q such that

$$\sup_{B \in \mathcal{B}} |P\{f_n(U_n) \in B\} - \int_B \zeta_n(u)\, du| = o(\beta_{s,n}), \tag{1.3.6}$$

where $\mathcal{B}$ is defined by (1.3.3). To this end we define, for $\alpha > 0$,

$$\rho(\alpha) = \{(2+\alpha)\, log(\beta_{s,n}^{-1})\}^{1/2} \quad \text{and} \quad H_n(\alpha) = \{t \in \mathbf{R}^k \mid \|t\| \leq \rho(\alpha)\},$$

and assume the following regularity condition:

(i) f_n is p times continously differentiable on $H_n(\alpha)$ and
$\sup\{\|D^p f_n(t)\| \,|\, t \in H_n(\alpha)\} \neq o(\beta_{s,n})$; (ii) with (1.3.7)
$\lambda_n = \sup\{(\|D^j f_n(0)\|/j!)^{1/(j-1)} \,|\, 2 \leq j \leq p-1\}$, we have $\lambda_n^{p-1} \neq o(\beta_{s,n})$.

Now we have the following theorem due to Skovgaard (1981):

Theorem 1.3.4: *Assume condition (1.3.7). Then if* $\lambda_n^{q-1} = o(\beta_{s,n})$ *and* $q \geq s$, *(1.3.6) is true uniformly over all* $B \in \mathcal{B}$.

1.4 Motivation and Summary of Thesis

Since Owen's pioneering papers in 1988 and 1990, empirical likelihood has been drawing increasing attention as a nonparametric method of constructing confidence regions and doing tests. However, almost all theoretical developments of empirical likelihood have focussed on the case where the parameter of interest is a smooth function of means and the sample is i.i.d. It is only in this case that coverage error has been shown to be of order n^{-1}, reducible to n^{-2} by Bartlett correction. Hall and La Scala (1990) gave a survey of developments in this setting. At the same time, the majority of published work concentrated on constructing confidence regions, with little attention being paid to aspects of hypothesis testing and to power properties of the empirical likelihood test.

The main contributions of this thesis are: (1) developing the high-order theory of empirical likelihood in new settings, which include the cases of quantiles and regression; (2) calculating the power of empirical likelihood tests.

The first non-standard case considered in this thesis is that of an empirical likelihood confidence interval for a population quantile. Owen (1988) has noted that, when applied to the problem of constructing confidence intervals for a population quantile (in particular, for the median), empirical likelihood reproduces precisely the so-called sign-test or binomial-method interval. This is reassuring, but it does show that in the context of quantile estimation, straight empirical likelihood has nothing to offer over existing techniques. One of the disadvantages of the sign test method is that it is usually unable to deliver confidence intervals with coverage accuracy better than $n^{-1/2}$, even for two-sided intervals. The reason of the poor performance is the discreetness of the binomial distribution, to which the empirical likelihood ratio converges. In Chapter 2 we show that by appropriately smoothing the empirical likelihood, coverage accuracy may be improved from order $n^{-1/2}$ to order n^{-1}. We demonstrate that this improvement is available for a wide range of choices of the smoothing parameter, so that it is not necessary to accurately determine an "optimal" value of the parameter. Furthermore, we show that smoothed empirical likelihood is Bartlett correctionable. That is, an empirical correction for scale can reduce the size of coverage error from order n^{-1} to order n^{-2}.

In Chapter 3 we consider constructing a confidence region for the regression coefficient vector, say β, of a linear regression model. Due to the presence of the fixed design points, the responses of the model are independent but not identically distributed random variables. Owen (1991) proposed using empirical likelihood to construct confidence region for β. He derived a nonparametric version of Wilks' theorem, ensuring that the empirical likelihood confidence regions have correct asymptotic coverage. However, questions regarding the coverage accuracy and Bartlett correctability of the confidence region remain to be addressed. We show in Chapter 3 that the coverage accuracy of an empirical likelihood confidence region for the regression coefficient vector is of order of n^{-1}, and that Bartlett correction can be implemented to improve the coverage accuracy from oder of n^{-1} to n^{-2}. We also give an empirical Bartlett factor for practically implementing the Bartlett correction.

However, it is not enough to just construct confidence regions for the regression coefficient vector β. In practice, statisticians are often confronted with problems of constructing confidence intervals for a particular regression coefficient or for certain linear combinations of β. In Chapter 4 we consider constructing empirical likelihood confidence intervals for the slope and means parameter of a simple linear regression model, by proving nonparametric versions of Wilks' Theorem for these parameters. We also show that the coverage accuracy of confidence intervals is of order n^{-1}, and that Bartlett correction can be used to further improve this accuracy.

After constructing an empirical likelihood confidence region, we can derive an empirical likelihood test for the parameter of interest by using the duality between the confidence region and hypothesis test. However, so far little has been done on aspects of the power of the empirical likelihood test. And, surprisingly, little has been done for the case of a bootstrap test. The contribution of Chapter 5 is to develop high-order expansions for the power function of empirical likelihood and bootstrap tests for a mean against a series of local alternatives. A comparison between empirical likelihood and bootstrap tests for a mean parameter, against a series of local alternative hypotheses is made. For univariate and bivariate cases, practical rules are proposed for choosing the more powerful test.

CHAPTER 2

EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR QUANTILES

2.1 Introduction

We noted in Chapter 1 that most work on empirical likelihood has concentrated on the case where the parameter of interest is a smooth function of means. In this chapter we consider constructing confidence intervals for population quantiles, which cannot be represented as a smooth function of means.

Owen (1988) has noted that, when applied to the problem of constructing confidence intervals for a population quantile (in particular, for the median), empirical likelihood reproduces precisely the so-called sign-test or binomial-method interval. This is reassuring, but it does show that in the context of quantile estimation, straight empirical likelihood has nothing to offer over existing techniques. One problem associated with the sign test method is that it is usually unable to creat confidence intervals with coverage accuracy better than order of $n^{-1/2}$ even for two-sided intervals. The reason for the poor coverage performance of the sign test intervals is due to the discreteness of the binomial distribution, which determines the true coverage probability.

Our aim in this paper is to show that coverage accuracy of an empirical likelihood confidence interval for quantiles may be improved from order $n^{-\frac{1}{2}}$ to order n^{-1}, by appropriately smoothing the empirical likelihood. We demonstrate that this improvement is available for a wide range of choices of the smoothing parameter, so that it is not necessary to accurately determine an "optimal" value of the parameter. Furthermore, we show that smoothed empirical likelihood is Bartlett correctable. That is, an empirical correction for scale can reduce the size of coverage error from order n^{-1} to order n^{-2}.

We also establish a very general version of Wilks' theorem in the context of empirical likelihood for quantiles. This result provides necessary and sufficient conditions on the range within which the smoothing parameter must lie if the asymptotic distribution of the empirical log likelihood ratio statistic is to be (central) chi-squared. Furthermore, we derive necessary and sufficient conditions on the smoothing parameter for the error in the chi-squared approximation to be $O(n^{-1})$, and also for the error after Bartlett correction to be $O(n^{-2})$. We suggest a particularly simple version of the Bartlett correction that produces confidence intervals with coverage error $o(n^{-1})$, although not quite $O(n^{-2})$.

Section 2.2 discusses unsmoothed empirical likelihood confidence intervals for quantiles. Section 2.3 describes smoothed empirical likelihood methods for quantiles, and proves a nonparametric version of Wilks' theorem. We also study in that section the coverage accuracy and Bartlett correctability of the confidence intervals . A simulation study is presented in Section 2.4. All proofs are deferred to Section 2.5.

2.2 Unsmoothed Empirical Likelihood Confidence Intervals for Quantiles

Let $X_1, \cdots, X_n$ be an i.i.d. sample from an unknown distribution F with $\theta_q = F^{-1}(q)$ as its unique q'th quantile. We wish to construct a confidence interval for θ_q. Let $p = (p_1, \cdots, p_n)$ with p_i's being nonnegative numbers adding to unity. We define the weighted empirical distribution function of F as

$$\hat{F}_p(x) = \sum_{i=1}^{n} p_i \, I(X_i \leq x),$$

where I is the indicator function. Then, empirical likelihood for θ_q, evaluated at θ, is defined to be

$$L(\theta) = \sup_{p:\hat{F}_p(\theta)=q \ \sum p_i=1} \prod_{i=1}^{n} p_i. \tag{2.2.1}$$

If we drop the constraint $\hat{F}_p(\theta) = q$ in (2.2.1), the profile likelihood is maximised by taking $p_i = n^{-1}$ for $1 \leq i \leq n$. For this choice of p we have $\theta(p) = \hat{\theta}$, the so-called bootstrap estimator of θ. This implies that

$$L(\hat{\theta}) = n^{-n}.$$

The empirical likelihood ratio is given by

$$R(\theta) = L(\theta)/L(\hat{\theta}) = \sup_{p:\hat{F}_p(\theta)=q} \prod_{i=1}^{n} (np_i). \tag{2.2.2}$$

Let $\theta(p)$ be the q'th quantile of the weighted empirical distribution function $\hat{F}_p(x)$. Then $\theta(p) = \inf\{x : \hat{F}_p(x) \geq q\}$. Let us re-index the sample such that $X_i = X_{(i)}$, denoting the i'th largest data value in the sample. Clearly the range of $\theta(p)$ is the set of ordered statistics $\{X_{(1)}, \cdots, X_{(n)}\}$. According to (2.2.2), we have for any $1 \leq i \leq n$,

$$R\{X_{(i)}\} = L\{X_{(i)}\}/L(\hat{\theta}) = \sup_{p:\theta(p)=X_{(i)},\, \sum p_i=1} \prod_{i=1}^{n} n\, p_i. \tag{2.2.3}$$

It is obvious that (2.2.3) can be reformulated as an optimization problem with the following form:

$$R\{X_{(i)}\} = \sup \prod_{i=1}^{n} (n\, p_i),$$

subject to

$$\begin{cases} \sum_{j=1}^{n} p_j = 1, \\ \sum_{j=1}^{i} p_j \geq q, \\ \sum_{j=1}^{i-1} p_j < q, \\ p_j \geq 0, \quad \text{for } 1 \leq j \leq n. \end{cases} \tag{2.2.4}$$

Since the objective function $\prod_{i=1}^{n}(n\, p_i)$ is a concave function of p, and the feasible set of p satisfying (2.2.4) is convex, then any local maximum is also a globe maximum. Using the Kuhn-Tucker theorem we may show that the optimal p has the following form:

$$p_j = \begin{cases} q/i, & 1 \leq j \leq i; \\ (1-q)/(n-i), & i+1 \leq j \leq n. \end{cases}$$

Thus we have

$$R\{X_{(i)}\} = n^n\ (q/i)^i\ \{(1-q)/(n-i)\}^{n-i}. \tag{2.2.5}$$

Some simple calculation reveals that $R\{X_{(i)}\}$ is an unimodal function satisfying

$$\begin{cases} R\{X_{(i)}\} \leq R(X_{(i+1)}) & \text{if } i \leq [n\, p]; \\ R\{X_{(i)}\} \geq R(X_{(i+1)}) & \text{if } i \geq [n\, p], \end{cases}$$

where $[n\,p]$ represents the largest integer not exceeding np. This enables us to define an empirical likelihood confidence interval for θ_q to be

$$I(c) = \{\ \theta :\ R(\theta) > c\ \} = [X_{(r_1)}, X_{(r_2)}],$$

where r_1, r_2 are respectively the smallest, largest integers such that

$$n^n\ (q/i)^i\ \{(1-q)/(n-i)\}^{n-i} \geq c\,.$$

According to David (1981, p.15), if F has a density, the exact coverage probability of the confidence interval $I(c)$ is given by

$$\begin{aligned} P\{\theta_q \in I(c)\} &= P\{X_{(r_1)} \leq \theta_q \leq X_{(r_2)}\} \\ &= \sum_{i=r_1}^{r_2-1} \binom{n}{i} q^i (1-q)^{n-i} \qquad (2.2.6) \\ &= P(r_1 \leq M \leq r_2 - 1), \end{aligned}$$

where M is a binomial $Bi(n,q)$ random variable. Formula (2.2.6) implies that the empirical likelihood confidence interval for a quantile is equivalent to that obtained by the so-called "sign test". This coverage probability cannot rendered closer than order $n^{-1/2}$ to any predetermined nominial coverage level, such as 0.95, no matter how the integers r_1, r_2 are selected. To appreciate this point, notice that due to the discreteness of the binomial distribution the coverage probability of $I(c)$ given by (2.2.6), can take only a finite number of values. This means that for any α between 0 and 1 it is very likely that you cannot have an exact α level confidence interval for θ_q. By the DeMoivre-Laplace theorem, we can approximate a binomial distribution by a normal distribution. In particular, using Kalinin's result (Johnson and Kotz, 1969, p.62f.), we have

$$P(r_1 \leq M \leq r_2 - 1) = \Phi(y_2) - \Phi(y_1) + \sum_{j=1}^{\infty} \{n\,q\,(1-q)\}^{-j/2}\,Q_j\,, \qquad (2.2.7)$$

where Φ is the standard normal distribution function, Q_j's are known function of ω, y_1 and y_2, where ω is the continuity correction which can be assigned arbitrarily (usually, we choose $\omega = 0.5$), and

$$y_1 = \frac{r_1 - (1-\omega) - nq}{\sqrt{n\,q\,(1-q)}}, \quad y_2 = \frac{r_1 + (1-\omega) - nq}{\sqrt{n\,q\,(1-q)}}.$$

Thus, for any $0 \leq \alpha \leq 1$, by appropriately choosing r_1 and r_2, we can let

$$\Phi(y_2) - \Phi(y_1) = \alpha.$$

Then put $c_\alpha = \max\{R_{(r_1)}, R_{(r_2)}\}$, from (2.2.6) amd (2.2.7) we have

$$P(\theta_q \in I_{c_\alpha}) = \alpha + \{q(1-q)\}^{-1/2} Q_1 n^{-1/2} + O(n^{-1}). \qquad (2.2.8)$$

This means that the empirical likelihood confidence interval for a quantile has coverage error no better than $O(n^{-1/2})$.

2.3 Smoothed Empirical Likelihood Confidence Intervals for Quantiles

We showed in the previous section that due to the discreteness of the binomial distribution, the coverage of the empirical likelihood confidence interval for a quantiles is in error by a term of size $n^{-1/2}$. To improve coverage accuracy we construct a smoothed empirical likelihood for a quantile in this section, by smoothing the weighted empirical distribution function $\hat{F}_p(x)$. We show that this smoothed empirical likelihood admits a nonparametric version of Wilks' theorem, which allows us to construct a confidence interval for θ_q by consulting the χ_1^2 tables. Furthermore, we show that by appropriately choosing the smoothing parameter, the coverage error of the smoothed empirical likelihood confidence interval is of order n^{-1} and can be further reduced to order of n^{-2} by employing Bartlett correction. These are significant improvements over the confidence interval obtained by the "sign test".

We divide this section into three parts. In subsection 2.3.1 we give some notation and lemmas, and introduce smoothed empirical likelihood. In subsection 2.3.2 we prove a nonparametric version of Wilks' theorem for smoothed empirical likelihood. In subsection 2.3.3 we establish an Edgeworth expansion for the distribution of the smoothed empirical likelihood, which enables us to derive the coverage accuracy and Bartlett correctability of smoothed empirical likelihood confidence intervals for θ_q.

2.3.1 Notation and Lemmas

In this subsection we give some notation and lemmas, and introduce smoothed empirical likelihood which will be used in the rest parts of this section. To define smoothed empirical likelihood we have to first give some notation and concepts of kernel smoothing.

Let K denote an r'th order kernel, of the type commonly used in nonparametric density estimation or regression (e.g. Silverman 1986, p.66ff; Härdle 1990, p.141f). That is, for some integer $r \geq 2$ and constant $\kappa \neq 0$, K is a function satisfying

$$\int u^j K(u)\, du = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{if } 1 \leq j \leq r-1 \\ \kappa & \text{if } j = r. \end{cases} \tag{2.3.1}$$

The case $r = 2$ is the most common, and there we take K to be a symmetric probability density. Larger values of r produce curve estimators with smaller variance. Define $G(x) = \int_{y<x} K(y)\, dy$. In this notation we put $G_h(x) = G(x/h)$. When $r = 2$ and K is a density, G and G_h are proper distribution functions. The h appearing in $G_h(x)$ is called the "bandwidth" or "smoothing parameter" and satisfies

$$h \to 0, \quad \text{as } n \to \infty. \tag{2.3.2}$$

Let f be the density function of F and $f^{(i)}$ the i'th derivative of f. We assume that

$$\begin{aligned} &f \text{ and } f^{(r-1)} \text{ exist in a neighbourhood of } \theta_q \text{ and are continuous} \\ &\text{at } \theta_q; \text{and } f(\theta_q) > 0. \end{aligned} \tag{2.3.3}$$

The moments of $G_h(\theta_q - X)$ are calculated in the following lemma:

Lemma 2.3.1. *Assume conditions (2.3.1) - (2.3.3), and that the kernel K is bounded and compactly supported. Then*

$$(i) \qquad E\{G_h(\theta_q - X)\} = q + \{(-h)^r/r!\} f^{(r-1)}(\theta_q)\,\kappa + o(h^r),$$

$$(ii) \qquad E\{G_h^{m+1}(\theta_q - X)\} = q - (m+1)\, h\, f(\theta_q)\, b_m + o(h),$$

where m is any positive integer and $b_m = \int_{-\infty}^{\infty} u\, G^m(u)\, K(u)\, du$.

Proof: We first prove (i). Using integration by parts,

$$E\{G_h(\theta_q - X)\} = \int_{-\infty}^{\infty} G_h(\theta_q - x)\, dF(x) = -\int_{-\infty}^{\infty} G(u)\, dF(\theta_q - h\, u)$$
$$= \int_{-\infty}^{\infty} F(\theta_q - h\, u)\, K(u)\, du.$$

By Taylor expansion of $F(\theta_q - h\, u)$ around θ_q, and noticing that K is an r'th order kernel,

$$E\{G_h(\theta_q - X)\} = F(\theta_q) + \{(-h)^r / r!\} f^{(r-1)}(\theta_q)\, \kappa \qquad (2.3.4)$$
$$+ (-h)^r / r! \int_{-\infty}^{\infty} u^r\, \{f^{(r-1)}(\theta_q - \omega\, h\, u) - f^{(r-1)}(\theta_q)\}\, K(u)\, du,$$

where $\omega = \omega(u) \in (0,1)$. Since K is bounded and compactly supported, and $f^{(r-1)}$ is continous at θ_q, it may be shown that

$$\lim_{h \to 0} \int_{-\infty}^{\infty} u^r\, \{f^{(r-1)}(\theta_q - \omega\, h\, u) - f^{(r-1)}(\theta_q)\}\, K(u)\, du = 0.$$

Substituting this into (2.3.4) and noting that $F(\theta_q) = q$ we have proved (i).

To prove (ii), we first notice that the conditions of K being bounded and compactly supported imply that b_m is finite for each positive integer m. Again using integration by parts,

$$E\{G_h^{m+1}(\theta_q - X)\} = \int_{-\infty}^{\infty} G_h^{m+1}(\theta_q - X)\, dF(x)$$
$$= -\int_{-\infty}^{\infty} G_h^{m+1}(u)\, dF(\theta_q - h\, u)$$
$$= -(m+1) \int_{-\infty}^{\infty} F(\theta_q - h\, u)\, G_h^m(u)\, K(u)\, du. \qquad (2.3.5)$$

Based on an one-term Taylor expansion of $F(\theta_q - h\, u)$ around θ_q and the continuity of f at θ_q, we can show from (2.3.5) that

$$E\{G_h^{m+1}(\theta_q - X)\} = q - (m+1)\, h\, f(\theta_q)\, b_m + o(h).$$

Thus, (ii) is proved. □

We define

$$w_i(\theta) = G_h(\theta - X_i) - q \quad \text{for } 1 \le i \le n,$$

and $\mu_k = E\{w_1(\theta_q)^k\}$ for $k = 1, 2 \cdots$. Using Lemma 2.3.1, we have

$$\begin{cases} \mu_1 = c_0\, h^r + o(h^r), \\ \mu_2 = q - q^2 + O(h), \\ \mu_i = q + \sum_{l=1}^{i-1} (-1)^l \binom{i}{l} q^{l+1} + (-1)^i q^i + O(h), \quad i \ge 3, \end{cases} \tag{2.3.6}$$

where $c_o = (-1)^r \,\kappa\, f^{(r-1)}(\theta_q)/r!$.

Now we may construct smoothed empirical likelihood for θ_q. We first smooth the weighted empirical distribution function $\hat{F}_p$ by defining

$$\hat{F}_{p,h}(\theta) = \sum_{i=1}^{n} p_i G_h(\theta - X_i).$$

We see that the smoothing is achieved by replacing the indicator function $I(X_i \le \theta)$ in $\hat{F}_p$ with $G_h(\theta - X_i)$. Replacing the constraint $\hat{F}_p(\theta) = q$ by its smoothed counterpart $\hat{F}_{p,h}(\theta) = q$ in (2.2.2), and taking the logarithm, we get the smoothed empirical log likelihood ratio for θ_q evaluated at $\theta_q = \theta$,

$$\ell_h(\theta) = \inf_{p:\hat{F}_{p,h}(\theta)=q;\sum p_i=1.} -2\sum_{i=1}^{n} \log(n\,p_i).$$

Using the Lagrange multiplier method, we may prove that the optimal point occurs with $p_i = n^{-1}\{1 + \lambda(\theta)\, w_i(\theta)\}^{-1}$, whence

$$\ell_h(\theta) = 2\sum_{i=1}^{n} \log\{1 + \lambda(\theta)\, w_i(\theta)\},$$

where $\lambda(\theta)$ is determined by

$$\sum_{i=1}^{n} w_i(\theta)\,\{1 + \lambda(\theta)\, w_i(\theta)\}^{-1} = 0\,. \tag{2.3.7}$$

The solution of equation (2.3.7), $\lambda(\theta)$, satisfies the following Lemma 2.3.2, whose proof is deferred to Section 2.5.

Lemma 2.3.2: *Assume that K satisfies (2.3.1), and is bounded and compactly supported. Then $\lambda(\theta_q) = O_p(n^{-1/2} + h^r)$, where $\lambda(\theta_q)$ is determined by (2.3.7) with $\theta = \theta_q$.*

FIGURE 2.1: Unsmoothed (step function) via smoothed empirical likelihood ratio functions for median based on sample $\mathcal{A}$, with various choices of bandwidth h: (1) $h = n^{-1/4}$, (2) $h = n^{-1/2}$, (3) $h = n^{-3/4}$ and (4) $h = n^{-1}$.

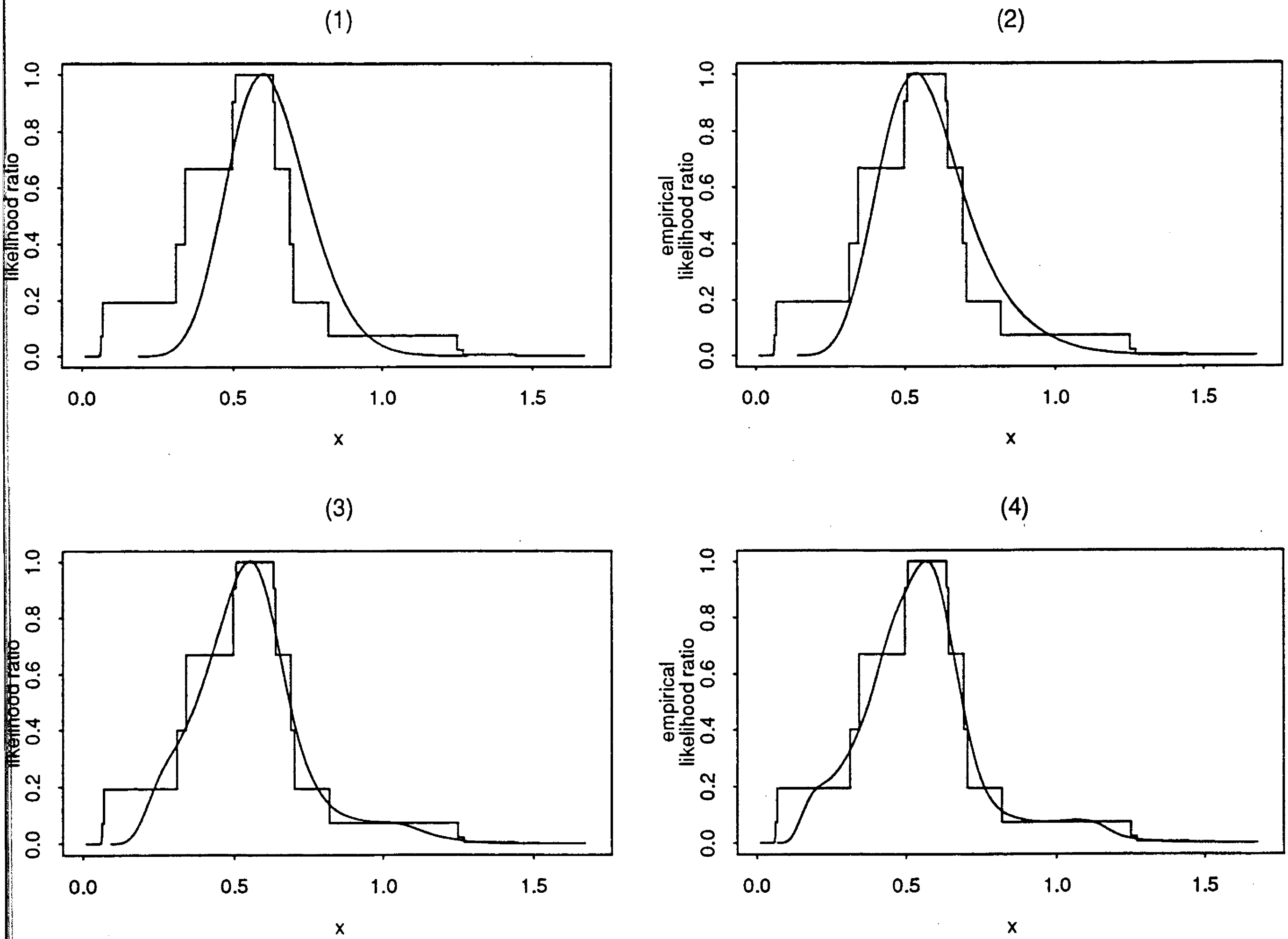


Figure 2.1 shows the unsmoothed empirical likelihood ratio function $R\{X_(i)\}$ graphed against the smoothed empirical likelihood ratio function $\ell_h(\theta_q)$ for a median ($q = 1/2$), based on a random sample

$$\mathcal{A} = \{0.011, 0.024, 0.055, 0.06, 0.068, 0.313, 0.341, 0.496, 0.506, 0.633,$$
$$0.639, 0.689, 0.70, 0.817, 1.251, 1.271, 1.445, 1.662, 1.678\}$$

generated from the χ_1^2 distribution.

2.3.2 Wilks' Theorem and Coverage Accuracy

As pointed out in Chapter 1, a fundamental result of empirical likelihood is that, like parametric likelihood, it admits a nonparametric version of Wilks' theorem. We have mentioned in Chapter 1 that the Wilks' theorem holds true for the case of smooth function of means, which enables us to construct an empirical likelihood confidence interval by looking up the chi-square tables. For our current problem of constructing confidence intervals for a quantile, we would like to first prove the Wilks' theorem for $\ell_h(\theta_q)$, which will give us a smoothed empirical likelihood confidence interval with correct asymptotic coverage. Then we would like to investigate coverage accuracy and Bartlett correctability of the confidence interval. In particular, we wish the order of magnitude of coverage error to be of smaller order than $n^{-1/2}$, which is the order of the coverage error of unsmoothed empirical likelihood confidence intervals (as shown in Section 2.2). The aim of this subsection is to address these problems by giving three theorems (Theorems 2.3.3 - 2.3.5). The proofs of these theorems are deferred to Section 2.5.

Our first result establishes necessary and sufficient conditions on the choice of bandwidth, h, such that $\ell_h(\theta_q)$ has an asymptotic χ_1^2 distribution.

Theorem 2.3.3: *Assume that*

K satisfies (2.3.1), and is bounded and compactly supported; that

f and $f^{(r-1)}$ exist in a neighbourhood of θ_q and are continuous (2.3.8)

at θ_q; that $f(\theta_q) > 0$; and that for some $t > 0$, $nh^t \to 0$ as $n \to \infty$.

Then $\ell_h(\theta_q)$ has an asymptotic χ_1^2 distribution if $nh^{2r} \to 0$, and this condition is also necessary if $f^{(r-1)}(\theta_q) \neq 0$.

Let us explain the implications of condition (2.3.8). The first part of (2.3.8) asks that K be a kernel of order r. The requirements that K be bounded and compactly supported implies that G_h is bounded, so as to get the result in Lemma 2.3.2 which is used to prove Theorem 2.3.2. However, we could obtain the result in Theorem 2.3.2 by imposing other similar conditions on the kernel. The second part asks that the distribution function F be sufficiently smooth in a neighbourhood of θ_q; the condition that r continuous derivatives of the target function (here, F) exist is the usual smoothness assumption imposed when working with an r'th order kernel. Requiring that $f(\theta_q) > 0$ ensures that the asymptotic variance of the sample quantile is of order n^{-1}. Without that assumption the order of magnitude of variance is strictly larger than n^{-1}, and the asymptotic theory is quite different. Finally, asking that $nh^t \to 0$ as $n \to \infty$ ensures that the bandwidth does not converge to zero too slowly. This is actually a very weak condition on h, since there is no restriction on t.

If K is a second-order kernel (i.e. $r = 2$) and $f'(\theta_q) \neq 0$ then $\ell_h(\theta_q)$ is asymptotically χ_1^2 if and only if $h = o(n^{-\frac{1}{4}})$. Such a bandwidth is of smaller order of magnitude than that which is usually appropriate for minimising error of a curve estimator; the latter h is of size $n^{-\frac{1}{5}}$, as shown for example by Silverman (1986, p.40ff). When $f^{(r-1)}(\theta_q) = 0$, it is possible for $\ell_h(\theta_q)$ to have an asymptotic χ_1^2 distribution yet nh^{2r} to be bounded away from zero.

If (2.3.8) is true and we choose the bandwidth h such that $n\,h^{2r} \to 0$, then by the theorem we can construct an α-level smoothed empirical likelihood confidence interval for θ_q as follows. First find from the χ_1^2 tables the value c_α such that

$$P(\chi_1^2 \leq c_\alpha) = \alpha.$$

Then, $I_{hc_\alpha} = \{\theta : \ell_h(\theta) \leq c_\alpha\}$ is a smoothed empirical likelihood confidence interval with nominal coverage level α. However, our objective of smoothing is not to get a result like this. Instead, we wish to find a suitable range of h such that the coverage error of I_{hc_α} is of smaller order than $n^{-1/2}$, which would show that I_{hc_α} has better coverage that the unsmoothed interval I_{c_α} given in Section 2.2.

Based on Theorem 2.3.3, we assume that

$$nh^{2r} \to 0 \quad \text{as } n \to \infty. \tag{2.3.9}$$

Clearly (2.3.9) implies (2.3.2). To establish an expansion of Edgeworth type for the distribution function of $\ell_h(\theta_q)$, we assume that

$$nh/\log n \to \infty, \quad \text{as } n \to \infty. \tag{2.3.10}$$

The coverage accuracy of I_{hc_α} is discussed in the following theorem:

Theorem 2.3.4: *Assume conditions (2.3.8) - (2.3.10). Then a sufficient condition for*

$$P(\theta_q \in I_{hc_\alpha}) = \alpha + O(n^{-1}) \tag{2.3.11}$$

as $n \to \infty$, is that nh^r is bounded. This condition is also necessary if $f^{(r-1)}(\theta_q) \neq 0$.

Theorem 2.3.4 implies that the smoothed empirical likelihood confidence interval I_{hc_α} has coverage error of order n^{-1} if the bandwidth h is properly chosen as recommended by the theorem. This is a significant improvement over the unsmoothed empirical likelihood confidence interval I_{c_α} given in Section 2.2. Notice that the boundness of nh^r is sufficient for condition (2.3.9) to be true. If the order of the kernel K is $r \geq 2$, we can choose $h = O(n^{-1/r})$. It is obvious that for such h, nh^r is bounded and $nh/\log n \to \infty$. Theorem 2.3.4 assures that this choice of h leads to coverage accuracy of order n^{-1}.

2.3.3. Bartlett Correction

From the proof of Theorem 2.3.4, which is deferred to Section 2.5, we see that no matter what the value of $c_\alpha > 0$, the right-hand side of (2.3.11) cannot be rendered equal to $\alpha + o(n^{-1})$ by appropriately choosing h. This means that smoothing cannot give us better coverage accuracy than $O(n^{-1})$. To further improve coverage accuracy we use Bartlett correction. It is well-known that part of the coverage error of I_{hc_α} is due to the fact that the mean of $\ell_h(\theta_q)$ is not equal to 1, which is the mean of the χ_1^2 distribution. Bartlett correction is a way to eliminate the approximating error by rescaling $\ell_h(\theta_q)$, so that it has correct means. We start with calculating the expectation of $\ell_h(\theta_q)$, which is given in the following lemma.

Lemma 2.3.5: *Assume conditions (2.3.8) and (2.3.9). Then,*

$$E\{\ell_h(\theta_q)\} = 1 + n^{-1}\beta + n\,\mu_1^2\,\mu_2^{-1} + o(nh^{2r}) + O(h^{3r} + n^{-1}h^r + n^{-2}),$$

where $\beta = \frac{1}{6}(3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2)$ *and* $\mu_j = E[G\{(\theta_q - X_i)/h\} - q]^j$.

We see from Lemma 2.3.5 that the difference between the expectations of $\ell_h(\theta_q)$ and its approximating chi-squared distribution is dominated by term $n^{-1}\beta + n\,\mu_1^2\,\mu_2^{-1}$. So if we choose bandwidth h such that $n\,h^{2r} = O(n^{-2})$ then we have

$$E\{\ell_h(\theta_q)\} - E(\chi_1^2) = n^{-1}\beta + O(n^{-2}).$$

We may reason that the expectation of $\ell_h(\theta_q)/(1 + n^{-1}\beta)$ differs from that of the χ_1^2 distribution only in terms of order n^{-2}, by using bandwidth h such that $n\,h^{2r} = O(n^{-2})$. However, β is usually unknown in practice and must be estimated. To this end, we define

$$\hat{\mu}_j = n^{-1}\sum_{i=1}^{n}[G\{(\hat{\theta}_q - X_i)/h\} - q]^j$$

and $\hat{\beta} = \frac{1}{6}(3\hat{\mu}_2^{-2}\hat{\mu}_4 - 2\hat{\mu}_2^{-3}\mu_3^2)$, where $\hat{\theta}_q$ is a root-n consistent estimate of θ_q. By the smoothness of G and Taylor expansion, we may show that $\hat{\beta} = \beta + O(n^{-1/2})$. Put $d(c_\alpha,\gamma) = c_\alpha(1 + n^{-1}\gamma)$ where γ is either β or $\hat{\beta}$. We prove in the following theorem that by appropriately choosing h, the Bartlett-corrected confidence interval $I_{h,d(c_\alpha,\gamma)} = \{\theta\,|\ell_h(\theta) \le d(c_\alpha,\gamma)\}$ has smaller coverage error than I_{hc_α}, no matter whether β or $\hat{\beta}$ is used.

Theorem 2.3.6. *Assume conditions* (2.3.8) *and* (2.3.10). *Then a sufficient condition for*

$$P(\theta_q \in I_{h,d(c_\alpha,\gamma)}) = \alpha + O(n^{-2}), \tag{2.3.12}$$

for either $\gamma = \beta$ or $\gamma = \hat{\beta}$, is that $n^3 h^{2r}$ be bounded. If $f^{(r-1)}(\theta_q) \neq 0$ then the boundedness of $n^3 h^{2r}$ is also necessary for (2.3.12).

From (2.3.6), we know that

$$\mu_2 = q(1-q) + O(h),$$
$$\mu_3 = q(1-q)(1-2q) + O(h),$$
$$\mu_4 = q(1-q)(1-3q+3q^2) + O(h).$$

Define $\beta_o = \frac{1}{6} q^{-1}(1-q)^{-1}(1-q+q^2)$. Then we have $\beta = \beta_0 + O(h)$. Since β_o is known, and if h is small enough, β_o will be a good approximation of β. For example, if h satisfies the requirement of Theorem 2.3.6 and K is a second-order kernel then $\beta = \beta_0 + O(n^{-3/4})$. Define the "partial" Bartlett-corrected confidence interval $I_{h,d(c_\alpha,\beta_o)} = \{\theta \,|\ell_h(\theta) \leq c_\alpha (1 + \beta_o n^{-1})\}$. It may be shown that the result in (2.3.12) can be changed to

$$P(\theta_q \in I_{h,d(c_\alpha,\beta_o)}) = \alpha + O(n^{-1}h). \tag{2.3.13}$$

Suppose we use a second order kernel and choose the bandwidth h of order $n^{-3/4}$, as suggested by Theorem 2.3.6. Then we obtain

$$P(\theta_q \in I_{h,d(c_\alpha,\beta_o)}) = \alpha + O(n^{-7/4}).$$

So the coverage error is just a factor $O(n^{1/4})$ larger than that of the full Bartlett correction confidence interval.

From a practical viewpoint, this simple "partial" Bartlett correction approach is particularly attractive. Although it does not enjoy quite the same asymptotic performance as the "full" correction discussed earlier, the simulation study in the next section shows that it performs commendably well in practice. This is presumably because the "full" correction needs to estimate μ_j for $j = 1,2,3$, which are relatively sensitive to bandwidth choice, and such estimators can be rather variable in small samples.

2.4 Simulation Study

In this section we present a simulation study designed to investigate the performance of smoothed empirical likelihood confidence intervals for quantiles, by using various bandwidth h when comparing with the unsmoothed confidence intervals. In particular, we wish to give examples of simple rules that are suggested by Theorems 2.3.4 and 2.3.6 for selecting bandwidth. We want to see if the empirical outcomes from our simulations are consistent with our theoretical findings.

Throughout this section we smooth using the so-called Bartlett or Epanechnikov kernel,

$$K(u) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}u^2\right) & \text{if } |u| \le \sqrt{5} \\ 0 & \text{otherwise} . \end{cases}$$

Since K is symmetric about the origin, it is a second-order kernel (i.e. $r = 2$). We concentrate on confidence intervals for quartiles and the median (i.e. $q = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$), and take the parent distribution F to be chi-squared with a variety of different degrees of freedom. We choose nominal coverages of $\alpha = 0.90$, and $\alpha = 0.95$, employ a variety of different formulae for h, and check on the performance of unsmoothed, smoothed and Bartlett correction confidence intervals. In the latter we take two different version of Bartlett correction confidence intervals. They are the "partially" corrected interval $I_{h,d(c_\alpha,\beta_o)}$ and "fullly" corrected interval $I_{h,d(c_\alpha,\gamma)}$ with $\gamma = \hat{\beta}$; we do not treat the interval $I_{h,d(c_\alpha,\gamma)}$ with $\gamma = \beta$ since β is usually unknown and therefore this method is not of practical interest. Formulae for I_{hc}, $I_{h,d(c_\alpha,\beta_o)}$, $I_{h,d(c_\alpha,\hat{\beta})}$, β_0 and $\hat{\beta}$ are given in Sections 2.3.2 and 2.3.3.

Recall from Theorems 2.3.4 and 2.3.6 that when $r = 2$, the bounds $h = O(n^{-\frac{1}{2}})$ and $h = O(n^{-\frac{3}{4}})$ define the largest h for which the uncorrected interval has coverage error $O(n^{-1})$ and the Bartlett-corrected interval with $\gamma = \hat{\beta}$ has coverage error $O(n^{-2})$, respectively.

Table 2.1 summarises results for the χ^2_m distributions with $m = 1,3,5$ and sample sizes $n = 10,15,20,30$. Figure 2.2 illustrates how coverage accuracy varies over different degrees of freedom and different sample sizes. Each point in the table and figure is based on 10,000 simulations. The chi-squared variables were produced by adding squares of independent normal variables given by the routine in Numerical Recipes (Press *et al.* 1989).

The following broad conclusions may be drawn from those results. First, smoothed empirical likelihood intervals have greater coverage accuracy than their unsmoothed counterparts, and further improvement is offered by Bartlett correction. Secondly, the "theoretical" Bartlett correction (based on the value β_0) performs similarly to the "empirical" Bartlett correction (using $\hat{\beta}$). Since β_0 is simpler than $\hat{\beta}$ to implement, it is to be recommended. Thirdly, choices of h in the range $n^{-\frac{1}{2}}$, $n^{-\frac{3}{4}}$ generally provide quite good coverage accuracy. However, when the underlying distribution is heavily skewed (e.g. χ_1^2), less smoothing than this is desirable.

TABLE 2.1: Estimated true coverages, from 10,000 simulations, of α-level smoothed empirical likelihood confidence intervals for the q'th quantile of the χ^2_m distribution. Rows headed "uncorr.", "β_0" and "$\hat{\beta}$" give the uncorrected interval and the Bartlett-corrected intervals computed with $\gamma = \beta_0$ and $\gamma = \hat{\beta}$, respectively.

$m = 1$, $n = 10$

h	q / α	0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.9239	0.9436	0.9345	0.9345	0.9240	0.9240
n^{-1}	uncorr.	0.8788	0.9278	0.8860	0.9430	0.9150	0.9594
	β_0	0.8916	0.9351	0.8935	0.9562	0.9491	0.9616
	$\hat{\beta}$	0.8920	0.9353	0.8950	0.9569	0.9501	0.9621
$n^{-\frac{3}{4}}$	uncorr.	0.8600	0.9266	0.8851	0.9460	0.9086	0.9465
	β_0	0.8745	0.9349	0.8934	0.9519	0.9296	0.9512
	$\hat{\beta}$	0.8757	0.9358	0.8945	0.9527	0.9307	0.9517
$n^{-\frac{1}{2}}$	uncorr.	0.8822	0.9377	0.8567	0.9184	0.8944	0.9268
	β_0	0.8943	0.9459	0.8686	0.9263	0.9068	0.9338
	$\hat{\beta}$	0.8939	0.9454	0.8720	0.9293	0.9088	0.9357
$n^{-\frac{1}{4}}$	uncorr.	0.8707	0.9306	0.7400	0.8295	0.8592	0.9064
	β_0	0.8827	0.9405	0.7518	0.8431	0.8695	0.9142
	$\hat{\beta}$	0.8814	0.9388	0.7550	0.8462	0.8717	0.9160

$m = 1,\ n = 15$

h	q / α	0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.9232	0.9787	0.9300	0.9693	0.9026	0.9694
n^{-1}	uncorr.	0.8821	0.9335	0.8844	0.9476	0.8807	0.9697
	β_0	0.8903	0.9394	0.8892	0.9509	0.8856	0.9701
	$\hat{\beta}$	0.8872	0.9407	0.8899	0.9511	0.8866	0.9710
$n^{-\frac{3}{4}}$	uncorr.	0.8562	0.9182	0.8890	0.9408	0.8815	0.9623
	β_0	0.8649	0.9250	0.8947	0.9443	0.8879	0.9645
	$\hat{\beta}$	0.8665	0.9261	0.8957	0.9451	0.8893	0.9651
$n^{-\frac{1}{2}}$	uncorr.	0.8710	0.9303	0.8701	0.9299	0.8895	0.9497
	β_0	0.8810	0.9357	0.8772	0.9339	0.8994	0.9544
	$\hat{\beta}$	0.8820	0.9365	0.8785	0.9349	0.9011	0.9552
$n^{-\frac{1}{4}}$	uncorr.	0.8748	0.9320	0.7190	0.8106	0.8740	0.9294
	β_0	0.8845	0.9384	0.7264	0.8199	0.8824	0.9321
	$\hat{\beta}$	0.8831	0.9378	0.7299	0.8227	0.8843	0.9339

$m = 1,\ n = 20$

h \ q / α		0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.8739	0.9348	0.9217	0.9734	0.9048	0.9718
n^{-1}	uncorr.	0.8848	0.9394	0.8921	0.9488	0.8821	0.9559
	β_0	0.8896	0.9427	0.8966	0.9500	0.8856	0.9566
	$\hat{\beta}$	0.8899	0.9432	0.8972	0.9508	0.8856	0.9566
$n^{-\frac{3}{4}}$	uncorr.	0.8575	0.9204	0.8925	0.9460	0.8833	0.9488
	β_0	0.8637	0.9265	0.8976	0.9493	0.8896	0.9510
	$\hat{\beta}$	0.8648	0.9288	0.8980	0.9502	0.8897	0.9509
$n^{-\frac{1}{2}}$	uncorr.	0.8596	0.9255	0.8849	0.9350	0.8915	0.9440
	β_0	0.8676	0.9312	0.8900	0.9376	0.9009	0.9482
	$\hat{\beta}$	0.8673	0.9317	0.8906	0.9382	0.9022	0.9483
$n^{-\frac{1}{4}}$	uncorr.	0.8894	0.9405	0.6903	0.7967	0.8851	0.9391
	β_0	0.8950	0.9449	0.6965	0.8030	0.8905	0.9427
	$\hat{\beta}$	0.8945	0.9443	0.6974	0.8047	0.8914	0.9432

$m = 1,\ n = 30$

h \ α \ q		0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.9388	0.9678	0.9292	0.9706	0.8806	0.9544
n^{-1}	uncorr.	0.8959	0.9456	0.8930	0.9525	0.9080	0.9417
	β_0	0.9005	0.9490	0.8953	0.9540	0.9093	0.9431
	$\hat{\beta}$	0.9007	0.9492	0.8953	0.9541	0.9094	0.9434
$n^{-\frac{3}{4}}$	uncorr.	0.8671	0.9285	0.8987	0.9472	0.9024	0.9418
	β_0	0.8711	0.9317	0.9010	0.9493	0.9050	0.9452
	$\hat{\beta}$	0.8718	0.9324	0.9012	0.9495	0.9056	0.9454
$n^{-\frac{1}{2}}$	uncorr.	0.8349	0.8972	0.8913	0.9427	0.8924	0.9462
	β_0	0.8391	0.9017	0.8940	0.9451	0.8961	0.9489
	$\hat{\beta}$	0.8397	0.9022	0.8948	0.9452	0.8966	0.9496
$n^{-\frac{1}{4}}$	uncorr.	0.8954	0.9453	0.6616	0.7647	0.8993	0.9431
	β_0	0.8993	0.9479	0.6665	0.7703	0.8938	0.9464
	$\hat{\beta}$	0.8989	0.9475	0.6675	0.7714	0.8942	0.9472

$m = 3,\ n = 10$

h	q / α	0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.9239	0.9436	0.9345	0.9345	0.9240	0.9240
n^{-1}	uncorr.	0.9139	0.9530	0.8885	0.9254	0.9187	0.9697
	β_0	0.9368	0.9570	0.8924	0.9678	0.9602	0.9712
	$\hat{\beta}$	0.9376	0.9577	0.8930	0.9680	0.9604	0.9716
$n^{-\frac{3}{4}}$	uncorr.	0.9040	0.9388	0.8871	0.9390	0.9160	0.9571
	β_0	0.9191	0.9436	0.8931	0.9529	0.9468	0.9598
	$\hat{\beta}$	0.9201	0.9442	0.8937	0.9532	0.9472	0.9599
$n^{-\frac{1}{2}}$	uncorr.	0.8854	0.9246	0.8852	0.9491	0.9057	0.9479
	β_0	0.8958	0.9309	0.8928	0.9573	0.9250	0.9519
	$\hat{\beta}$	0.8961	0.9317	0.8956	0.9583	0.9262	0.9529
$n^{-\frac{1}{4}}$	uncorr.	0.8736	0.9206	0.8830	0.9430	0.8965	0.9312
	β_0	0.8849	0.9284	0.8914	0.9477	0.9068	0.9380
	$\hat{\beta}$	0.8868	0.9317	0.8941	0.9499	0.9085	0.9394

$m = 3,\ n = 15$

h	q / α	0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.9232	0.9787	0.9300	0.9693	0.9026	0.9694
n^{-1}	uncorr.	0.8853	0.9672	0.8826	0.9582	0.8802	0.9769
	β_0	0.8899	0.9689	0.8850	0.9599	0.8829	0.9774
	$\hat{\beta}$	0.8912	0.9693	0.8856	0.9599	0.8835	0.9777
$n^{-\frac{3}{4}}$	uncorr.	0.8920	0.9542	0.8908	0.9517	0.8873	0.9707
	β_0	0.9002	0.9567	0.8954	0.9537	0.8926	0.9726
	$\hat{\beta}$	0.9025	0.9576	0.8960	0.9540	0.8938	0.9727
$n^{-\frac{1}{2}}$	uncorr.	0.8880	0.9442	0.8902	0.9443	0.8889	0.9651
	β_0	0.8956	0.9481	0.8958	0.9476	0.8945	0.9667
	$\hat{\beta}$	0.8926	0.9488	0.8964	0.9477	0.8959	0.9674
$n^{-\frac{1}{4}}$	uncorr.	0.8852	0.9331	0.8931	0.9472	0.8919	0.9511
	β_0	0.8925	0.9382	0.8985	0.9507	0.9001	0.9542
	$\hat{\beta}$	0.8951	0.9397	0.9008	0.9515	0.9027	0.9547

$m = 3,\ n = 20$

h \ q, α		0.25, 0.90	0.25, 0.95	0.50, 0.90	0.50, 0.95	0.75, 0.90	0.75, 0.95
0		0.874	0.935	0.922	0.973	0.905	0.972
n^{-1}	uncorr.	0.8844	0.9565	0.8866	0.9559	0.8848	0.9600
	β_0	0.8866	0.9586	0.8881	0.9569	0.8876	0.9610
	$\hat{\beta}$	0.8910	0.9593	0.8888	0.9708	0.8874	0.9610
$n^{-\frac{3}{4}}$	uncorr.	0.8825	0.9472	0.8903	0.9487	0.8831	0.9408
	β_0	0.8885	0.9498	0.8938	0.9507	0.8898	0.9422
	$\hat{\beta}$	0.8896	0.9506	0.8939	0.9508	0.8896	0.9427
$n^{-\frac{1}{2}}$	uncorr.	0.8947	0.9443	0.8985	0.9482	0.8894	0.9467
	β_0	0.9015	0.9484	0.9029	0.9498	0.8961	0.9489
	$\hat{\beta}$	0.9027	0.9501	0.9032	0.9500	0.8964	0.9491
$n^{-\frac{1}{4}}$	uncorr.	0.8901	0.9410	0.8931	0.9444	0.8968	0.9466
	β_0	0.8960	0.9453	0.8971	0.9476	0.9002	0.9507
	$\hat{\beta}$	0.8969	0.9463	0.8986	0.9487	0.9027	0.9509

$m = 3,\ n = 30$

h \ q, α		0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.9388	0.9678	0.9292	0.9706	0.8806	0.9544
n^{-1}	uncorr.	0.9027	0.9397	0.8980	0.9538	0.9076	0.9457
	β_0	0.9049	0.9409	0.8992	0.9549	0.9082	0.9464
	$\hat{\beta}$	0.9049	0.9410	0.8993	0.9549	0.9083	0.9465
$n^{-\frac{3}{4}}$	uncorr.	0.8970	0.9462	0.8973	0.9518	0.9066	0.9427
	β_0	0.9007	0.9487	0.8994	0.9532	0.9076	0.9441
	$\hat{\beta}$	0.9010	0.9490	0.8995	0.9533	0.9076	0.9443
$n^{-\frac{1}{2}}$	uncorr.	0.8949	0.9465	0.8956	0.9476	0.9039	0.9442
	β_0	0.8985	0.9494	0.8980	0.9492	0.9072	0.9465
	$\hat{\beta}$	0.8991	0.9501	0.8983	0.9495	0.9073	0.9467
$n^{-\frac{1}{4}}$	uncorr.	0.8950	0.9487	0.8976	0.9495	0.8980	0.9488
	β_0	0.8990	0.9506	0.8995	0.9512	0.9002	0.9518
	$\hat{\beta}$	0.8996	0.9512	0.9004	0.9520	0.9030	0.9526

$m = 5, n = 10$

h		q = 0.25, α = 0.90	0.25, 0.95	0.50, 0.90	0.50, 0.95	0.75, 0.90	0.75, 0.95
0		0.9239	0.9436	0.9345	0.9345	0.9240	0.9240
n^{-1}	uncorr.	0.9138	0.9651	0.8852	0.9219	0.9236	0.9702
	β_0	0.9287	0.9677	0.8884	0.9729	0.9651	0.9721
	$\hat{\beta}$	0.9291	0.9680	0.8889	0.9733	0.9655	0.9723
$n^{-\frac{3}{4}}$	uncorr.	0.9148	0.9519	0.8866	0.9332	0.9179	0.9648
	β_0	0.9255	0.9549	0.8918	0.9657	0.9553	0.9676
	$\hat{\beta}$	0.9266	0.9555	0.8924	0.9660	0.9555	0.9681
$n^{-\frac{1}{2}}$	uncorr.	0.8987	0.9285	0.8929	0.9478	0.9127	0.9558
	β_0	0.9108	0.9342	0.8990	0.9627	0.9347	0.9586
	$\hat{\beta}$	0.9118	0.9351	0.9010	0.9636	0.9363	0.9595
$n^{-\frac{1}{4}}$	uncorr.	0.8755	0.9233	0.8831	0.9480	0.9010	0.9382
	β_0	0.8860	0.9293	0.8924	0.9547	0.9161	0.9442
	$\hat{\beta}$	0.8876	0.9306	0.8951	0.9554	0.9180	0.9458

$m = 5,\ n = 15$

h	q / α	0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.9232	0.9787	0.9300	0.9693	0.9026	0.9694
n^{-1}	uncorr.	0.8761	0.9617	0.8874	0.9599	0.8790	0.9793
	β_0	0.8805	0.9632	0.8899	0.9606	0.8812	0.9801
	$\hat{\beta}$	0.8813	0.9633	0.8902	0.9606	0.8817	0.9804
$n^{-\frac{3}{4}}$	uncorr.	0.8872	0.9526	0.8909	0.9525	0.8844	0.9755
	β_0	0.8941	0.9549	0.8942	0.9543	0.8893	0.9766
	$\hat{\beta}$	0.8955	0.9556	0.8950	0.9544	0.8899	0.9766
$n^{-\frac{1}{2}}$	uncorr.	0.8873	0.9531	0.8899	0.9504	0.8890	0.9642
	β_0	0.8962	0.9568	0.8942	0.9531	0.8966	0.9660
	$\hat{\beta}$	0.8983	0.9577	0.8949	0.9534	0.8977	0.9664
$n^{-\frac{1}{4}}$	uncorr.	0.8878	0.9408	0.8961	0.9499	0.8870	0.9543
	β_0	0.8965	0.9458	0.9022	0.9540	0.8965	0.9580
	$\hat{\beta}$	0.8989	0.9473	0.9040	0.9549	0.8982	0.9585

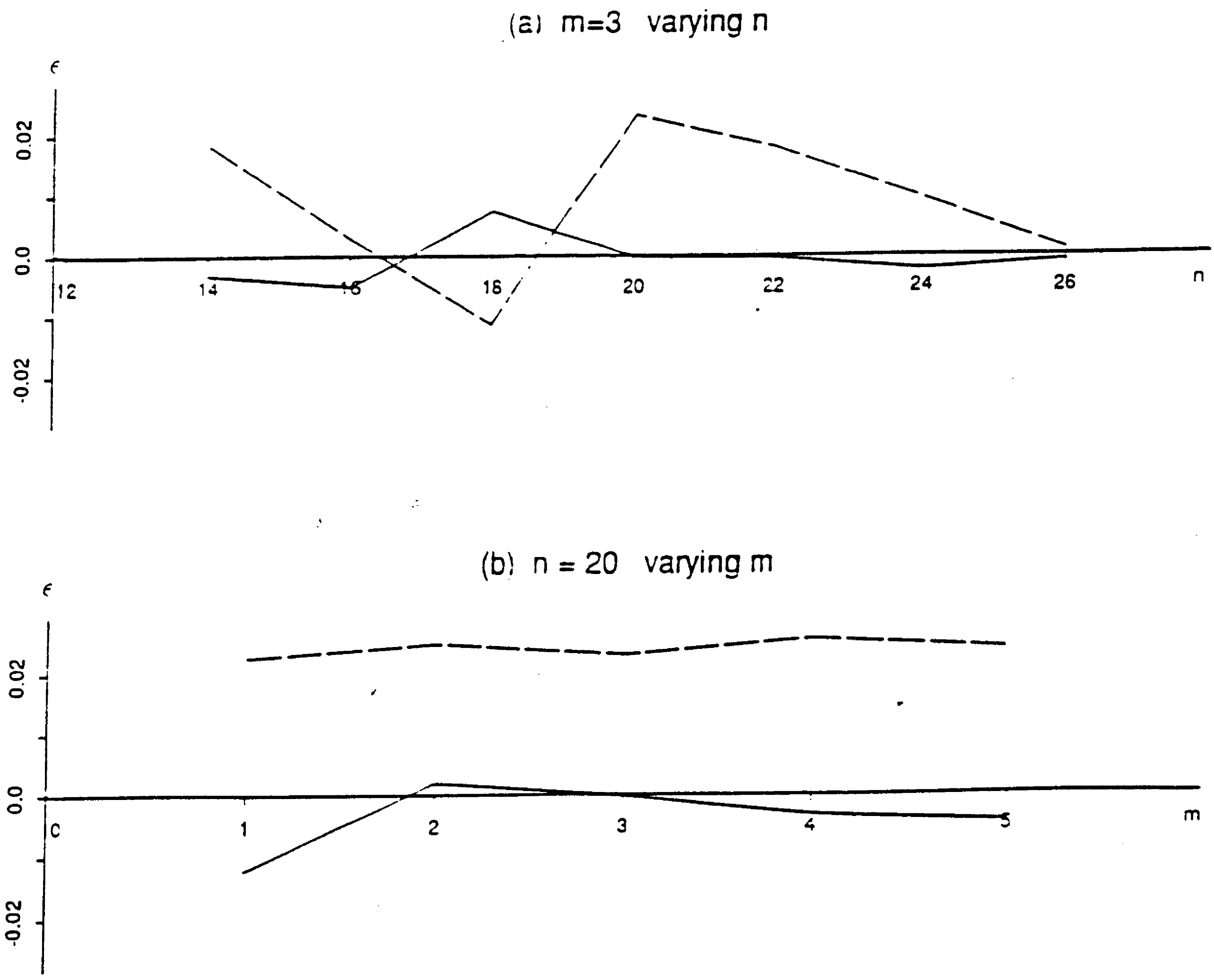
$m = 5,\ n = 20$

h \ q, α		0.25		0.50		0.75	
		0.90	0.95	0.90	0.95	0.90	0.95
0		0.8739	0.9348	0.9217	0.9734	0.9048	0.9718
	uncorr.	0.8848	0.9394	0.8921	0.9488	0.8821	0.9538
n^{-1}	β_0	0.8794	0.9567	0.8925	0.9565	0.8796	0.9559
	$\hat{\beta}$	0.8803	0.9571	0.8925	0.9565	0.8772	0.9606
	uncorr.	0.8858	0.9479	0.8889	0.9461	0.8797	0.9441
$n^{-\frac{3}{4}}$	β_0	0.8922	0.9488	0.8924	0.9471	0.8864	0.9480
	$\hat{\beta}$	0.8933	0.9493	0.8924	0.9473	0.8914	0.9496
	uncorr.	0.8947	0.9468	0.8948	0.9466	0.8827	0.9486
$n^{-\frac{1}{2}}$	β_0	0.9008	0.9493	0.8986	0.9478	0.8913	0.9506
	$\hat{\beta}$	0.9034	0.9504	0.8989	0.9478	0.8932	0.9508
	uncorr.	0.8837	0.9422	0.8925	0.9431	0.8916	0.9421
$n^{-\frac{1}{4}}$	β_0	0.8998	0.9457	0.8962	0.9447	0.8998	0.9474
	$\hat{\beta}$	0.9022	0.9473	0.8962	0.9446	0.9048	0.9485

$m = 5,\ n = 30$

h		q = 0.25, α = 0.90	0.25, 0.95	0.50, 0.90	0.50, 0.95	0.75, 0.90	0.75, 0.95
0		0.9388	0.9678	0.9292	0.9706	0.8806	0.9544
n^{-1}	uncorr.	0.9073	0.9407	0.9026	0.9544	0.9140	0.9422
	β_0	0.9089	0.9417	0.9034	0.9548	0.9147	0.9428
	$\hat{\beta}$	0.9089	0.9419	0.9034	0.9548	0.9147	0.9428
$n^{-\frac{3}{4}}$	uncorr.	0.9025	0.9484	0.9008	0.9556	0.9058	0.9428
	β_0	0.9050	0.9507	0.9031	0.9563	0.9080	0.9443
	$\hat{\beta}$	0.9051	0.9508	0.9032	0.9565	0.9083	0.9444
$n^{-\frac{1}{2}}$	uncorr.	0.8994	0.9453	0.9013	0.9480	0.9045	0.9472
	β_0	0.9033	0.9489	0.9039	0.9500	0.9075	0.9494
	$\hat{\beta}$	0.9035	0.9491	0.9040	0.9500	0.9075	0.9495
$n^{-\frac{1}{4}}$	uncorr.	0.8988	0.9478	0.9009	0.9494	0.8978	0.9489
	β_0	0.9025	0.9505	0.9041	0.9508	0.9019	0.9514
	$\hat{\beta}$	0.9032	0.9512	0.9050	0.9512	0.9025	0.9518

FIGURE 2.2: The graphs depict coverage error, given by $\epsilon =$ true coverage $- 0.95$, of smoothed (————) and unsmoothed (– – – – –) 95% confidence intervals for the median (i.e. $q = \frac{1}{2}$). In the case of the smoothed confidence interval, the bandwidth is $h = n^{-\frac{3}{4}}$ and Bartlett correction is employed with $\gamma = \beta_0$. Throughout, the underlying distribution is χ^2_m. Panel (a) illustrates the case where $m = 3$ is fixed and n varies; panel (b) illustrates $n = 20$ and varying m.



2.5 Proofs

In this section we give detailed proofs of Lemmas 2.3.2 and 2.3.5, and Theorems 2.3.3, 2.3.4 and 2.3.6.

2.5.1 Proof of Lemma 2.3.2

Lemma 2.3.2: *Assume that K satisfies (2.3.1), and is bounded and compactly supported. Then $\lambda(\theta_q) = O_p(n^{-1/2} + h^r)$, where $\lambda(\theta_q)$ is determined by (2.3.7) with $\theta = \theta_q$.*

Proof: Define $w_i = w_i(\theta_q) = G_h(\theta_q - X_i) - q$, and let $\lambda = \lambda(\theta_q)$ denote a solution of the equation

$$\sum_{i=1}^{n} w_i(1 + \lambda w_i)^{-1} = 0. \tag{2.5.1}$$

Since K is a bounded and compactly supported kernel satisfying (2.3.1), and

$$G(x) = \int_{y<x} K(y)\, dy,$$

then G is uniformly bounded on $\mathbf{R}^1$. Therefore there exists a positive number d_o such that

$$|w_i| \leq d_o, \quad \text{for } 1 \leq i \leq n. \tag{2.5.2}$$

From (2.5.1) and (2.5.2),

$$\begin{aligned}
0 = n^{-1}|\sum_{i=1}^{n} \frac{w_i}{1+\lambda w_i}| &= n^{-1}|\sum_{i=1}^{n}\left(w_i - \lambda\frac{w_i^2}{1+\lambda w_i}\right)| \\
&\geq n^{-1}|\lambda|\sum_{i=1}^{n}\frac{w_i^2}{|1+\lambda w_i|} - |\bar{w}_1| \\
&\geq \frac{|\lambda|}{1+d_o|\lambda|} n^{-1}\sum_{i=1}^{n} w_i^2 - |\bar{w}_1| \\
&= \frac{|\lambda|}{1+d_o|\lambda|}|\bar{w}_2| - |\bar{w}_1|,
\end{aligned}$$

where $\bar{w}_j = n^{-1}\sum_{i=1}^{n} w_i^j$ for $j = 1, 2, \cdots$. Therefore, $|\lambda|\,\bar{w}_2 \leq (1 + d_o|\lambda|)\,|\bar{w}_1|$, or equivalently,

$$|\lambda|\,(\bar{w}_2 - d_o|\bar{w}_1|) \leq |\bar{w}_1|. \tag{2.5.3}$$

Observe that $|\overline{w}_j|$ is average of i.i.d. random variables, so $\bar{w}_j - E(\bar{w}_j) = O_p(n^{-\frac{1}{2}})$. We know from (2.3.6) that $E(\bar{w}_1) = O(h^r)$ and $E(\bar{w}_2) = q(1-q) + o(1)$. Therefore, by (2.5.3),

$$|\lambda|\{q(1-q) + o_p(1)\} \leq O_p(n^{-\frac{1}{2}} + h^r),$$

which immediately gives us the result $\lambda = \lambda(\theta_q) = O_p(n^{-1/2} + h^r)$. Hence the lemma is proved. □

2.5.2 Proof of Theorem 2.3.3

Theorem 2.3.3: *Assume that*

> *K satisfies (2.3.1), and is bounded and compactly supported; that*
>
> *f and $f^{(r-1)}$ exist in a neighbourhood of θ_q and are continuous*
>
> *at θ_q; that $f(\theta_q) > 0$; and that for some $t > 0$, $nh^t \to 0$ as $n \to \infty$.*

Then $\ell_h(\theta_q)$ has an asymptotic χ_1^2 distribution if $nh^{2r} \to 0$, and this condition is also necessary if $f^{(r-1)}(\theta_q) \neq 0$.

Proof: We start with developing a Taylor expansion for λ which is the solution of (2.5.1). Using Lemma 2.3.2, for each $j \geq 1$ we have

$$\begin{aligned} 0 &= n^{-1} \sum_{i=1}^{n} w_i\{1 - \lambda w_i + (\lambda w_i)^2 - (\lambda w_i)^3 + \ldots\} \\ &= \bar{w}_1 - \lambda\, \bar{w}_2 + \cdots + (-\lambda)^j\, \bar{w}_{j+1} + (-\lambda)^{j+1}\, n^{-1} \sum_{i=1}^{n} \frac{w_i^{j+2}}{1 + \lambda w_i}. \end{aligned}$$

Inverting the above equation we have for $j = 1$,

$$\lambda = \bar{w}_2^{-1}\, \bar{w}_1 + T_1,$$

where according to Lemma 2.3.2

$$T_1 = \lambda^2\, \bar{w}_2^{-1}\, n^{-1} \sum_{i=1}^{n} \frac{w_i^3}{1 + \lambda w_i} = O_p\{(n^{-1/2} + h^r)^2\}. \tag{2.5.4}$$

For $j = 2$,

$$\begin{aligned} \lambda &= \bar{w}_2^{-1}\, \bar{w}_1 + \bar{w}_2^{-1}\, \bar{w}_3\, \lambda^2 - \bar{w}_2^{-1}\, \lambda^3\, n^{-1} \sum_{i=1}^{n} \frac{w_i^4}{1 + \lambda w_i} \\ &= \bar{w}_2^{-1}\, \bar{w}_1 + \bar{w}_2^{-3}\, \bar{w}_3\, \bar{w}_1^2 + T_2, \end{aligned}$$

where

$$T_2 = 2\,\bar{w}_2^{-2}\,\bar{w}_3\,\bar{w}_1\,T_1 + \bar{w}_2^{-1}\,\bar{w}_3\,T_1^2 - \bar{w}_2^{-1}\,\lambda^3\,n^{-1}\sum_{i=1}^{n}\frac{w_i^4}{1+\lambda w_i}.$$

Using Lemma 2.3.2 again and (2.5.4) we have

$$T_2 = O_p\{(n^{-1/2}+h^r)^3\}. \tag{2.5.5}$$

For $j = 3$,

$$\begin{aligned}\lambda &= \bar{w}_2^{-1}\,\bar{w}_1 + \bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + 2\,\bar{w}_2^{-3}\,\bar{w}_3^2\,\bar{w}_1\,\lambda^2 - \bar{w}_2^{-1}\,\bar{w}_4\,\lambda^3\\ &\quad - 2\,\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1\,\lambda^3\,n^{-1}\sum_{i=1}^{n}\frac{w_i^4}{1+\lambda w_i} + \bar{w}_2^{-1}\,\bar{w}_3\,T_1^2 + \bar{w}_2^{-1}\,\lambda^4\,n^{-1}\sum_{i=1}^{n}\frac{w_i^5}{1+\lambda w_i}\\ &= \bar{w}_2^{-1}\,\bar{w}_1 + \bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + 2\,\bar{w}_2^{-5}\,\bar{w}_3^2\,\bar{w}_1^3 - \bar{w}_2^{-4}\,\bar{w}_4\,\bar{w}_1^3 + T_3,\end{aligned}$$

where

$$\begin{aligned}T_3 &= 2\,\bar{w}_2^{-3}\,\bar{w}_3^2\,\bar{w}_1\,\{2\,\bar{w}_2^{-1}\,\bar{w}_1(\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + T_2) + (\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + T_2)^2\}\\ &\quad - \bar{w}_2^{-1}\,\bar{w}_4\,\{3\,\bar{w}_2^{-2}\,\bar{w}_1^2\,(\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + T_2) + 3\,\bar{w}_2^{-1}\,\bar{w}_1\,(\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + T_2)^2\\ &\quad + (\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + T_2)^3\} - 2\,\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1\,\lambda^3\,n^{-1}\sum_{i=1}^{n}\frac{w_i^4}{1+\lambda w_i}\\ &\quad + \bar{w}_2^{-1}\,\bar{w}_3\,T_1^2 + \bar{w}_2^{-1}\,\lambda^4\,n^{-1}\sum_{i=1}^{n}\frac{w_i^5}{1+\lambda w_i}.\end{aligned}$$

By Lemma 2.3.2, (2.5.4) and (2.5.5) we obtain that

$$T_3 = O_p\{(n^{-1/2}+h^r)^4\}.$$

In general we have

$$\begin{aligned}\lambda =& \bar{w}_2^{-1}\,\bar{w}_1 + \bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^2 + (2\bar{w}_2^{-5}\,\bar{w}_3^2 - \bar{w}_2^4\,\bar{w}_4)\,\bar{w}_1^3\\ &+ \sum_{k=4}^{j} R_{1k}\,\bar{w}_1^k + O_p\{(n^{-\frac{1}{2}}+h^r)^{j+1}\},\end{aligned} \tag{2.5.6}$$

where R_{1k} denotes $\bar{w}_2^{-(2k-1)}$ multiplied by a polynomial in $\bar{w}_2, \ldots, \bar{w}_{k+1}$, with constant coefficients. Expansion (2.5.6) is a little longer than are necessary for our present proof of Theorem 2.3.3. However, the additional details given here will be needed in the proof of Theorem 2.3.4.

Substituting the above Taylor expansion for $\lambda(\theta_q)$ into $\ell_h(\theta_q)$,

$$\begin{aligned}
\ell_h(\theta_q) &= 2\sum_{i=1}^{n} \log(1+\lambda w_i) \\
&= 2n \sum_{k=1}^{j+1} (-1)^{k+1}\, k^{-1}\, \lambda^k\, \bar{w}_k + O_p\{n(n^{-\frac{1}{2}}+h^r)^{j+2}\} \\
&= n\{\bar{w}_2^{-1}\,\bar{w}_1^2 + \tfrac{2}{3}\,\bar{w}_2^{-3}\,\bar{w}_3\,\bar{w}_1^3 + (\bar{w}_2^{-5}\,\bar{w}_3^2 - \tfrac{1}{2}\,\bar{w}_2^{-4}\,\bar{w}_4)\,\bar{w}_1^4 \\
&\quad + (8\bar{w}_2^{-6}\,\bar{w}_3\,\bar{w}_4 - 8\bar{w}_2^{-7}\,\bar{w}_3^3 - \tfrac{8}{5}\,\bar{w}_2^{-5}\,\bar{w}_5)\,\bar{w}_1^5\} \\
&\quad + n\sum_{k=5}^{j} R_{2k}\,\bar{w}_1^{k+1} + O_p\{n(n^{-\frac{1}{2}}+h^r)^{j+1}\}. \qquad (2.5.7)
\end{aligned}$$

The third identity follows on substituting (2.5.6) into the second identity, and noting that $\bar{w}_1 = \bar{w}_1 - E(\bar{w}_1) + E(\bar{w}_1) = O_p(n^{-\frac{1}{2}} + h^r)$.

Put $Z = n^{\frac{1}{2}}(\bar{w}_1 - \mu_1)\mu_2^{-\frac{1}{2}}$, where $\mu_j = E(\bar{w}_j)$. It is readily proved that under condition (2.3.8), Z is asymptotically normal $N(0,1)$. Furthermore, $\bar{w}_2 = \mu_2 + o_p(1) = q(1-q) + o_p(1)$, and $\bar{w}_j = O_p(1)$ for $j \geq 3$. Hence by (2.5.7) we have for any $j > 3$,

$$\ell_h(\theta_q) = n\mu_2^{-1}\,\bar{w}_1^2\{1+o_p(1)\} + O_p\{n(n^{-\frac{1}{2}}+h^r)^3\} + O_p\{n(n^{-\frac{1}{2}}+h^r)^j\}.$$

Since $nh^t \to 0$ for some $t > 0$, we have $n(n^{-\frac{1}{2}}+h^r)^j = o(1)$ for sufficient large $j > 3$. Also notice that

$$n(n^{-\frac{1}{2}}+h^r)^3 = O(n^{\frac{1}{2}}h^{2r} + nh^{3r}) + o(1) = o(n^{\frac{1}{2}}h^r + nh^{2r}) + o(1)$$

Hence,

$$\ell_h(\theta_q) = (n^{\frac{1}{2}}\,\mu_2^{-\frac{1}{2}}\,\mu_1 + Z)^2 + o_p(n^{\frac{1}{2}}h^r + nh^{2r}) + o_p(1).$$

Note that

$$(n^{\frac{1}{2}}\,\mu_2^{-\frac{1}{2}}\,\mu_1 + Z)^2 = Z^2 + O(n^{\frac{1}{2}}\mu_1 + n\mu_1^2) + o(n^{\frac{1}{2}}h^r + nh^{2r}) + o(1).$$

Thus, $\ell_h(\theta_q)$ has an asymptotic central chi-squared distribution with one degree of freedom if and only if $n^{\frac{1}{2}}\,\mu_1 \to 0$. By (2.3.6), $\mu_1 = c_o\,h^r + o(h^r)$ where $c_o = (-1)^r\,\kappa\,f^{(r-1)}(\theta_q)/r!$. Therefore, $n^{\frac{1}{2}}\,\mu_1 \to 0$ if $nh^{2r} \to 0$; and if $f^{(r-1)}(\theta_q) \neq 0$ then $n^{\frac{1}{2}}\,\mu_1 \to 0$ implies $nh^{2r} \to 0$. So the theorem proved. □

2.5.3 Proof of Theorem 2.3.4

Before proving Theorem 2.3.4, we first give the following lemma which establishes an analogue of Cramér's condition for the random vector $(G_h(X-\theta_q), G_h^2(X-\theta_q), \cdots, G_h^j(X-\theta_q))$, where j is the interger appeared in (2.5.6) and X is a random variable with distribution F.

Lemma 2.5.1: *Assume conditions (2.3.8), (2.3.9) and (2.3.10). Then for each $\epsilon > 0$ there exists a constant $C(\epsilon) > 0$ such that for all sufficiently small h,*

$$\sup_{t_1,\ldots,t_j:\Sigma|t_k|>\epsilon} \left| h\int_{-\infty}^{\infty} \exp\left\{ i\sum_{k=1}^{j} t_k\, G(u)^k \right\} f(\theta_q - hu)\, du \right| \leq 1 - C(\epsilon)\, h\,,$$

where $i = \sqrt{-1}$ and $G(x) = \int_{y<x} K(y)\, dy$.

Proof: Let u denote a random variable uniformly distributed on the unit interval $[0,1]$, and $U = (u, u^2, \cdots, u^j)$. Put $t = (t_1, \cdots, t_j)$ and define

$$I(t) = \int_0^1 \exp(i\sum_{k=1}^{j} u^k) du.$$

Clearly $I(t)$ is the characteristic function of U. Thus, by the Riemann-Lebesgue Lemma

$$\lim_{\|t\|\to\infty} I(t) = 0.$$

Since $\|t\| \to \infty$ if and only if $\sum_{k=1}^{j} |t_k| \to \infty$, we obtain that

$$\lim_{\sum_{k=1}^{j} |t_k|\to\infty} I(t) = 0. \tag{2.5.8}$$

From (2.3.8) we may assume that the compact support of kernel K is some interval $[a, b]$. Then using (2.5.8) and Lemma 4.2 of Hall (1991), we may show that

$$\lim_{\eta\to\infty} \lim_{h\to 0} \sup_{\sum |t_k|>\eta} \left| \int_a^b \exp\{i\sum_{k=1}^{j} t_k\, G_h^k(u)\, f(\theta_q - hu) du \right| = 0. \tag{2.5.9}$$

With the above preparation, we define

$$J(t) = \int_a^b \exp\{i\sum_{k=1}^{j} t_k\, G_h^k(u)\}\, f(\theta_q - hu)\, du$$

and

$$S(t) = h \int_{-\infty}^{\infty} \exp\{i \sum_{k=1}^{j} t_k G_h^k(u)\} f(\theta_q - hu)\, du$$

$$= h \int_{-\infty}^{a} f(\theta_q - hu)\, du + h \int_{b}^{\infty} \exp(i \sum_{k=1}^{j} t_k) f(\theta_q - hu)\, du + h\, J(t)$$

$$= 1 - F(\theta_q - ha) + \exp(i \sum_{k=1}^{j} t_k) F(\theta_q - hb) + h\, J(t). \tag{2.5.10}$$

By (2.5.9), there exists h_o and $\eta_o > 0$ such that for any $0 < h < h_o$ and $\sum |t_k| > \eta_o$,

$$|J(t)| \leq \tfrac{1}{3}(b - a) f(\theta_q). \tag{2.5.11}$$

Since condition (2.3.8) implies that f is continous at θ_q and $f(\theta_q) > 0$, we can choose h sufficiently small for

$$F(\theta_q - ha) - F(\theta_q - hb) \geq \tfrac{2}{3}(b - a)\, h\, f(\theta_q).$$

From (2.5.10),

$$\sup_{\sum |t_k| > \eta_o} |S(t)| \leq 1 - F(\theta_q - ha) + F(\theta_q - hb) + h\, |J(t)|$$

$$\leq 1 - \tfrac{2}{3}(b - a)\, h\, f(\theta_q) + \tfrac{1}{3}(b - a)\, h\, f(\theta_q)$$

$$\leq 1 - \tfrac{1}{3}(b - a)\, h\, f(\theta_q). \tag{2.5.12}$$

To prove the lemma it is sufficient to show that for each $\epsilon > 0$ there exists a constant B_o such that

$$\sup_{\epsilon < \sum |t_k| \leq \eta_o} |S(t)| \leq 1 - B_o\, h\, f(\theta_q). \tag{2.5.13}$$

Put $\xi(u,t) = \sum_{k=1}^{j} t_k G_h^k(u)$. By the continuity of f at θ_q, we can split $J(t)$ as follows:

$$J(t) = f(\theta_q) \int_{a}^{b} \exp\{i\xi(u,t)\}\, du + R(t),$$

where $\sup_{-\infty < t < \infty} R(t) \to 0$ as $h \to 0$. From (2.5.10),

$$|S(t)| \leq |1 - F(\theta_q - ha) + h\, f(\theta_q) \int_{a}^{b} \exp\{i\xi(u,t)\}\, du| + F(\theta_q - hb) + h\, |R(t)|. \tag{2.5.14}$$

It can be shown that for any real number v and w, by choosing h sufficiently small we have

$$\begin{aligned}|1 - F(\theta_q - ha) + h\,(v + iw)|^2 &= \{1 - F(\theta_q - ha) + h\,v\}^2 + (h\,w)^2 \\ &\leq \left\{1 - F(\theta_q - ha) + h\,v + \frac{h^2\,(v^2 + w^2)}{1 - F(\theta_q)}\right\}^2.\end{aligned}$$

Thus

$$|1 - F(\theta_q - ha) + h\,(v + iw)| \leq 1 - F(\theta_q - ha) + h\,v + \frac{h^2\,(v^2 + w^2)}{1 - F(\theta_q)}. \qquad (2.5.15)$$

Put

$$v = f(\theta_q) \int_a^b \cos\{\xi(u,t)\}\,du \quad \text{and} \quad w = f(\theta_q) \int_a^b \sin\{\xi(u,t)\}\,du.$$

Hence,

$$f(\theta_q) \int_a^b \exp\{i\xi(u,t)\}\,du = v + i\,w.$$

Notice that for the v and w defined above, we have

$$\begin{aligned}v^2 + w^2 &= f^2(\theta_q)\left\{\left(\int_a^b \cos\{\xi(u,t)\}\,du\right)^2 + \left(\int_a^b \sin\{\xi(u,t)\}\,du\right)^2\right\} \\ &\leq 2\,f^2(\theta_q)\,(b-a)^2.\end{aligned}$$

Using the above result, (2.5.14) and (2.5.15), we obtain

$$\begin{aligned}|S(t)| \leq\, & 1 - F(\theta_q - h\,a) + F(\theta_q - h\,b) + h\,f(\theta_q) \int_a^b \cos\{\xi(u,t)\}\,du \\ & + \frac{2\,h^2}{1 - F(\theta_q)}\,f^2(\theta_q)\,(b-a)^2 + h\,|R(t)|.\end{aligned}$$

By the continuity of f at θ_q,

$$F(\theta_q - h\,a) - F(\theta_q - h\,b) = h \int_a^b f(\theta_q - h\,u)\,du \;=\; h\,f(\theta_q)\,(b-a) + h\,o_h(1).$$

Now from (2.5.14),

$$\begin{aligned}|S(t)| \leq\, & 1 - h\,f(\theta_q) \int_a^b [1 - \cos\{\xi(u,t)\}]\,du \\ & + \frac{2\,h^2}{1 - F(\theta_q)}\,f^2(\theta_q)\,(b-a)^2 + h\,\{R(t) + o_h(1)\}.\end{aligned}$$

By choosing h sufficient small we have from the above eqation that

$$|S(t)| \leq 1 - \tfrac{1}{2} h\, f(\theta_q) \int_a^b [1 - \cos\{\xi(u,t)\}]\, du.$$

Since G is an increasing function in $[a,b]$, we have

$$B_o = \tfrac{1}{2} \inf_{\epsilon < \sum |t_k| \leq \eta_o} \int_a^b [1 - \cos\{\xi(u,t)\}]\, du > 0,$$

Thus we obtain (2.5.13). Therefore the lemma is readily proved by using (2.5.12) and (2.5.13). □

Now we are able to give the proof of Theorerm 2.3.4.

Theorem 2.3.4: *Assume conditions (2.3.8) - (2.3.10). Then a sufficient condition for*

$$P(\theta_q \in I_{hc_\alpha}) = \alpha + O(n^{-1})$$

as $n \to \infty$, *is that* nh^r *is bounded. This condition is also necessary if* $f^{(r-1)}(\theta_q) \neq 0$.

Proof: To prove this theorem, we have to develop an Edgeworth expansion of the distribution function of $\ell_h(\theta_q)$. Recall (2.5.7),

$$\begin{aligned} \ell_h(\theta_q) = n\big\{ & \bar{w}_2^{-1}\, \bar{w}_1^2 + \tfrac{2}{3}\, \bar{w}_2^{-3}\, \bar{w}_3\, \bar{w}_1^3 + \big(\bar{w}_2^{-5}\, \bar{w}_3^2 - \tfrac{1}{2}\, \bar{w}_2^{-4}\, \bar{w}_4\big)\, \bar{w}_1^4 \\ & + \big(8\bar{w}_2^{-6}\, \bar{w}_3\, \bar{w}_4 - 8\bar{w}_2^{-7}\, \bar{w}_3^3 - \tfrac{8}{5}\, \bar{w}_2^{-5}\, \bar{w}_5\big)\, \bar{w}_1^5 \big\} \\ & + n \sum_{k=5}^{j} R_{2k}\, \bar{w}_1^{k+1} + O_p\{n(n^{-\frac{1}{2}} + h^r)^{j+1}\}. \end{aligned}$$

Taking the signed square root of the right-hand side we may write

$$\ell_h(\theta_q) = (n^{\frac{1}{2}}\, S_j')^2\,,$$

where

$$\begin{aligned} S_j' &= \bar{w}_2^{-\frac{1}{2}} \Big\{ \bar{w}_1 + \tfrac{1}{3}\, \bar{w}_2^{-2}\, \bar{w}_3\, \bar{w}_1^2 + \big(\tfrac{4}{9}\, \bar{w}_2^{-4}\, \bar{w}_3^2 - \bar{w}_3^{-3}\, \bar{w}_4\big)\, \bar{w}_1^3 \\ & \quad + \big(\tfrac{112}{27}\, \bar{w}_2^{-6}\, \bar{w}_3^3 + \tfrac{97}{12}\, \bar{w}_2^{-5}\, \bar{w}_3\, \bar{w}_4 - \tfrac{4}{5}\, \bar{w}_2^{-4}\, \bar{w}_5\big)\, \bar{w}_1^4 \\ & \quad + \sum_{k=5}^{j} T_k\, \bar{w}_1^k \Big\} + U_{1j} \\ &= S_j + U_{1j} \end{aligned}$$

say, where T_k denotes $\bar{w}_2^{-2(k-1)}$ multiplied by a polynomial in $\bar{w}_2, \ldots, \bar{w}_k$ with constant coefficients, and $U_{1j} = O_p\{(n^{-\frac{1}{2}} + h^r)^{j+1}\}$. Noting that $nh^t \to 0$ for some $t > 0$, a little additional analysis shows that by choosing j sufficiently large we may ensure that for $\ell = 1$,

$$P(|U_{\ell j}| > n^{-\frac{5}{2}}) = O(n^{-2}). \tag{2.5.16}$$

Hence using the delta-method we have for $x > 0$,

$$\begin{aligned} P\{\ell_h(\theta_q) \le x^2) &= P(-x \le n^{\frac{1}{2}} S_j' \le x) \\ &\left\{\begin{matrix}\le\\ \ge\end{matrix}\right\} P(-x \mp n^{-2} \le n^{\frac{1}{2}} S_j \le x \pm n^{-2}) + O(n^{-2}), \end{aligned} \tag{2.5.17}$$

where the inequalities and plus/minus signs are to be taken respectively, in the indicated orders.

We see from (2.5.17) that the Edgeworth expansion for the distribution of $\ell_h(\theta_q)$ can be derived by an Edgeworth expansion for the distribution of $n^{\frac{1}{2}} S_j$. So our next step is to develop an Edgeworth expansion for the distribution of $n^{\frac{1}{2}} S_j$. Observe that S_j is a smooth function of $\bar{w}_1, \ldots, \bar{w}_j$. Denote that function by s_j. Put $\mu_k = E(\bar{w}_k)$, $\mu = (\mu_1, \ldots, \mu_j)$, $u = (u_1, \ldots, u_j)$, $V_k = \bar{w}_k - \mu_k$, $V = (V_1, \ldots, V_j)$,

$$d_{k_1 \ldots k_m} = \left(\prod_{\ell=1}^{m} \partial/\partial u_{k_\ell}\right) s_j(u_1, \ldots, u_j)\Big|_{u=\mu},$$

$$p(u) = s_j(\mu) + \sum_{m=1}^{6} (m!)^{-1} \sum_{k_1, \ldots, k_m \in \{1, \ldots, j\}} d_{k_1 \ldots k_m} u_{k_1} \ldots u_{k_m}.$$

Then p is a polynomial, and $p(V)$ represents a Taylor approximation to S_j with an error of order n^{-3}:

$$S_j = p(V) + U_{2j},$$

where $U_{2j} = O_p(n^{-3})$. A little additional analysis shows that (2.5.16) holds for $\ell = 2$, and so

$$P(n^{\frac{1}{2}} S_j \le x)\left\{\begin{matrix}\le\\ \ge\end{matrix}\right\} P\{n^{\frac{1}{2}} p(V) \le x \pm n^{-2}\} + O(n^{-2}). \tag{2.5.18}$$

By developing Taylor expansion formulae for the quantities $d_{k_1 \dots k_m}$, and for the cumulants of V, calculations deferred to Appendix 2 show that the cumulants $k_1, k_2, \dots$ of $n^{\frac{1}{2}} p(V)$ satisfy the following formulae:

$$\begin{aligned} k_1 &= n^{\frac{1}{2}} s_j(\mu) - \tfrac{1}{6} \mu_2^{-\frac{3}{2}} \mu_3 n^{-\frac{1}{2}} + O(n^{-\frac{1}{2}} h^r + n^{-\frac{3}{2}}), \\ k_2 &= \sigma^2 + (\tfrac{1}{2} \mu_2^{-2} \mu_4 - \tfrac{13}{36} \mu_2^{-3} \mu_3^2) n^{-1} + O(n^{-1} h^r + n^{-2}), \qquad (2.5.19) \\ k_3 &= O(n^{-\frac{1}{2}} h^r), \quad k_4 = O(n^{-1} h^r), \quad k_\ell = O(n^{-(\ell-2)/2}) \text{ for } \ell \geq 5, \end{aligned}$$

where

$$n^{\frac{1}{2}} s_j(\mu) = n^{\frac{1}{2}} \mu_2^{-\frac{1}{2}} \mu_1 + o(nh^{2r}).$$

and

$$\begin{aligned} \sigma^2 &= \sum_{k_1}^{j} \sum_{k_2}^{j} d_{k_1} d_{k_2} E\{(w_1^{k_1} - \mu_{k_1})(w_1^{k_2} - \mu_{k_2})\} \\ &= 1 + \tfrac{1}{3} \mu_2^{-2} \mu_1 \mu_3 + (\tfrac{7}{9} \mu_2^{-4} \mu_3^2 - \tfrac{1}{4} \mu_2^{-1} - \tfrac{7}{12} \mu_2^{-3} \mu_4) \mu_1^2 + O(h^{3r}). \end{aligned}$$

Let $cf(t)$ be the characteristic function of $n^{\frac{1}{2}} p(V)$. Then using (2.5.19)

$$\begin{aligned} cf(t) &= \exp(-\frac{t^2}{2}) \exp\{k_1 (it) + (k_2 - 1) \frac{(it)^2}{2!} + \sum_{j=3}^{\infty} k_j \frac{(it)^j}{j!}\} \\ &= \exp(-\frac{t^2}{2}) \exp\{k_1 (it) + (k_2 - 1) \frac{(it)^2}{2!} + k_3 \frac{(it)^3}{3!}\} + O(n^{-1} h^r + n^{-2}). \end{aligned}$$

This allows us to develop a formal Edgeworth expansion for the distribution of $p(V)$: assuming $nh^{2r} \to 0$,

$$\begin{aligned} P\{n^{\frac{1}{2}} p(V) \leq x\} &= \Phi(x) - \tfrac{1}{12} n^{-1} \{6\mu_2^{-1} (n\mu_1)^2 + 3\mu_2^{-2} \mu_4 - 2\mu_2^{-3} \mu_3^2\} x\phi(x) \\ &\quad + Q(x)\phi(x) + o(nh^{2r}) + O(n^{-2}), \qquad (2.5.20) \end{aligned}$$

where Φ, ϕ denote the standard normal distribution, density functions respectively, and $Q(x)$ is a even polynomial in x. Hence for $x > 0$,

$$\begin{aligned} P\{-x \leq n^{\frac{1}{2}} p(V) \leq x\} &= 2\Phi(x) - 1 - \tfrac{1}{6} n^{-1} \{6\mu_2^{-1} (n\mu_1)^2 + 3\mu_2^{-2} \mu_4 - 2\mu_2^{-3} \mu_3^2\} x\phi(x) \\ &\quad + o(nh^{2r}) + O(n^{-2}). \qquad (2.5.21) \end{aligned}$$

In view of (2.5.17), (2.5.18), (2.5.21) and the fact that $2\Phi(x)-1 = P(\chi_1^2 \le x^2)$, we obtain

$$P\{\ell_h(\theta_q) \le x^2\} = P(\chi_1^2 \le x^2) - \tfrac{1}{6} n^{-1}\{6\mu_2^{-1}(n\mu_1)^2 + 3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2\} x^2\phi(x^2)$$
$$+ o(nh^{2r}) + O(n^{-2}).$$

Recall that the α-level empirical likelihood confidence interval $I_{hc_\alpha} = \{\theta | \ell_h(\theta) \le c_\alpha\}$. And from (2.3.6), μ_j for $j = 2,3,4$ have the following forms:

$$\mu_1 = (-h)^r (r!)^{-1}\kappa f^{(r-1)}(\theta_q) + o(h^r), \quad \mu_2 = q(1-q) + o(1),$$
$$\mu_3 = q - 3q^2 + 2q^3 + o(1), \quad \mu_4 = q - 4q^2 + 6q^3 - 3q^4 + o(1).$$

Thus, using the above expansion for the distribution of $\ell_h(\theta_q)$, we have

$$\begin{aligned} P(\theta_q \in I_{hc_\alpha}) &= P\{\ell_h(\theta_q) \le c_\alpha\} \\ &= \alpha - n^{-1}\{(r!)^{-2}\kappa^2 f^{(r-1)}(\theta_q)^2 (nh^r)^2 q^{-1}(1-q)^{-1} \\ &+ \tfrac{1}{6} q^{-1}(1-q)^{-1}(1-q+q^2)\} c_\alpha\phi(c_\alpha) + o(n^{-1} + nh^{2r}) \end{aligned} \quad (2.5.22)$$

Now if nh^r is bounded we see clearly from (2.5.22) that

$$P(\theta_q \in I_{hc_\alpha}) = \alpha + O(n^{-1}). \quad (2.5.23)$$

By (2.5.22),

$$P(\theta_q \in I_{hc_\alpha}) = \alpha - \{(r!)^{-2}\kappa^2 f^{(r-1)}(\theta_q)^2 n\, h^{2r} q^{-1}(1-q)^{-1})\} c_\alpha\phi(c_\alpha)$$
$$+ o(n\, h^{2r}) + O(n^{-1}).$$

if (2.5.23) holds and $f^{(r-1)}(\theta_q) \neq 0$, then nh^{2r} must be bounded. In fact we can show that the error term in (2.5.23) cannot be at smaller order of n^{-1} if $nh^r \to C$, $0 \le C < \infty$. To appreciate this we note that from (2.5.22) that

$$P(\theta_q \in I_{hc_\alpha}) - \alpha = o(n^{-1})$$

if and only if

$$(r!)^{-2}\kappa^2 f^{(r-1)}(\theta_q)^2 C^2 + \tfrac{1}{6}(1 - q + q^2) = 0.$$

But, the left-hand side is strictly positive for all $0 < q < 1$.

It remains to check that the formal expansion (2.5.20) is valid. This may be done by developing an Edgeworth expansion of the multivariate distribution of $n^{\frac{1}{2}} V = n^{\frac{1}{2}}(V_1, \cdots, V_j)$ where $V_k = \bar{w}_k - \mu_k$ for $1 \le k \le j$, with the form

$$P(n^{\frac{1}{2}} V \in B) = \Phi_{0,\Sigma}(B) + \sum_{k=1}^{m} n^{-k/2} \int_B p_k(x)\, \phi_{0,\Sigma}(x)\, dx + O(n^{-(m+1)/2}) \quad (2.5.24)$$

uniformly in j-variate sets B from any class $\mathcal{B}$ satisfying

$$\sup_{B \in \mathcal{B}} \Phi_{0,\Sigma}\{(\partial B)^{\epsilon}\} = O(\epsilon)$$

as $\epsilon \downarrow 0$. In these formulae, $\Sigma = \mathrm{Cov}(V)$; $\Phi_{0,\Sigma}$ and $\phi_{0,\Sigma}$ denote the distribution and density functions of the $N(0, \Sigma)$ distributions; p_k is a polynomial of degree $k+2$ with uniformly bounded coefficients; $m \ge 1$ is any integer; and $(\partial B)^{\epsilon}$ is the set of all points distant at most ϵ from the boundary of B. Noting that V is a mean of a sum of independent and identically distributed random variables, this result may be proved using techniques from Bhattacharya and Rao (1976, p.192ff), based on an analogue of Cramér's condition for V established in Lemma 2.5.1. The methods used to get (2.5.24) are those given by Hall (1991). This proves the theorem. □

2.5.4 Proof of Lemma 2.3.5

Lemma 2.3.5: *Assume conditions (2.3.8) and (2.3.9). Then,*

$$E\{\ell_h(\theta_q)\} = 1 + n^{-1}\beta + n\,\mu_1^2\,\mu_2^{-1} + o(nh^{2r}) + O(h^{3r} + n^{-1}h^r + n^{-2}),$$

where $\beta = \frac{1}{6}(3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2)$ *and* $\mu_j = E[G\{(\theta_q - X_i)/h\} - q]^j$.

Proof: From (2.5.7) we know that

$$\begin{aligned}\ell_h(\theta_q) = n\{&\bar{w}_2^{-1}\bar{w}_1^2 + \tfrac{2}{3}\bar{w}_2^{-3}\bar{w}_3\bar{w}_1^3 + (\bar{w}_2^{-5}\bar{w}_3^2 - \tfrac{1}{2}\bar{w}_2^{-4}\bar{w}_4)\,\bar{w}_1^4 \\ &+ (8\bar{w}_2^{-6}\bar{w}_3\bar{w}_4 - 8\bar{w}_2^{-7}\bar{w}_3^3 - \tfrac{8}{5}\bar{w}_2^{-5}\bar{w}_5)\,\bar{w}_1^5\} \\ &+ n\sum_{k=5}^{j} R_{2k}\,\bar{w}_1^{k+1} + O_p\{n(n^{-\frac{1}{2}} + h^r)^{j+2}\}.\end{aligned}$$

To get $E\{\ell_h(\theta_q)\}$, we calculate the expectation of each term on the right hand side of the above equation. The following formulae will be used in the calculations:

$$E(\bar{w}_1 - \mu_1)^2 = n^{-1}(\mu_2 - \mu_1^2), \quad E(\bar{w}_2 - \mu_2)^2 = n^{-1}(\mu_4 - \mu_2^2),$$

$$E\{(\bar{w}_1-\mu_1)(\bar{w}_2-\mu_2)\} = n^{-1}(\mu_3-\mu_1\mu_2),$$

$$E\{(\bar{w}_1-\mu_1)(\bar{w}_3-\mu_3)\} = n^{-1}(\mu_4-\mu_1\mu_3)$$

$$E\{(\bar{w}_2-\mu_2)(\bar{w}_3-\mu_3)\} = n^{-1}(\mu_5-\mu_2\mu_3)$$

$$E\{(\bar{w}_1-\mu_1)(\bar{w}_4-\mu_4)\} = n^{-1}(\mu_5-\mu_1\mu_4)$$

$$E(\bar{w}_1-\mu_1)^3 = n^{-2}(\mu_3-3\mu_1\mu_2+3\mu_1^3)$$

$$E\{(\bar{w}_1-\mu_1)(\bar{w}_2-\mu_2)^2\} = n^{-2}(\mu_4-2\mu_1\mu_3+\mu_1\mu_2-\mu_2^2+2\mu_1^2\mu_2).$$

Now by Taylor expansion we have

$$\begin{aligned}\bar{w}_2^{-1}\bar{w}_1^2 &= \mu_2^{-1}\sum_{k=0}^{\infty}(-\mu_2)^{-k}(\bar{w}_2-\mu_2)^k(\bar{w}_1-\mu_1+\mu_1)^2\\ &= \mu_2^{-1}\{\mu_1^2\sum_{k=0}^{2}(-\mu_2)^{-k}(\bar{w}_2-\mu_2)^k+2\mu_1(\bar{w}_1-\mu_1)\sum_{k=0}^{1}(-\mu_2)^{-k}(\bar{w}_2-\mu_2)^k\\ &\quad+(\bar{w}_1-\mu_1)^2\}+\nu_1,\end{aligned}$$

where $E(\nu_1)=O(n^{-2})$. Taking expactation we obtain,

$$\begin{aligned}E(\bar{w}_2^{-1}\bar{w}_1^2) &= \mu_1^2\mu_2^{-1}+\mu_1^2\mu_2^{-3}E(\bar{w}_2-\mu_2)^2-2\mu_1\mu_2^{-2}E\{(\bar{w}_1-\mu_1)(\bar{w}_2-\mu_2)\}\\ &\quad+\mu_2^{-1}E(\bar{w}_1-\mu_1)^2\\ &= \mu_1^2\mu_2^{-1}+(1-2\mu_1\mu_2^{-2}\mu_3+\mu_1^2\mu_2^{-3}\mu_4)n^{-1}+O(n^{-2}).\end{aligned} \tag{2.5.25}$$

For the second term,

$$\begin{aligned}\bar{w}_2^{-3}\bar{w}_3\bar{w}_1^3 &= \mu_2^{-3}\{1-3\mu_2^{-1}(\bar{w}_2-\mu_2)+\cdots\}\{(\bar{w}_3-\mu_3)+\mu_3\}\\ &\quad\times\{(\bar{w}_1-\mu_1)^3+3\mu_1(\bar{w}_1-\mu_1)^2+3\mu_1^2(\bar{w}_1-\mu_1)+\mu_1^3\}\\ &= \mu_2^{-3}\{\mu_1^3\mu_3+3\mu_1^3\mu_3(\bar{w}_1-\mu_1)+3\mu_1\mu_3(\bar{w}_1-\mu_1)^2\\ &\quad+3\mu_1^2(\bar{w}_1-\mu_1)(\bar{w}_3-\mu_3)-3\mu_1^3\mu_2^{-1}(\bar{w}_2-\mu_2)(\bar{w}_3-\mu_3)\\ &\quad-9\mu_1^2\mu_2^{-1}\mu_3(\bar{w}_1-\mu_1)(\bar{w}_2-\mu_2)\}+\nu_2,\end{aligned}$$

where $E(\nu_2)=O(n^{-1}h^{3r}+n^{-2})$. Hence,

$$\begin{aligned}E(\bar{w}_2^{-3}\bar{w}_3\bar{w}_1^3) &= \mu_1^3\mu_2^{-3}\mu_3+n^{-1}(3\mu_1\mu_2^{-2}\mu_3+3\mu_1^2\mu_2^{-2}\mu_4-9\mu_1^2\mu_2^{-4}\mu_3^2\\ &\quad-3\mu_1^3\mu_2^{-4}\mu_5+6\mu_1^3\mu_2^{-3}\mu_3)+O(n^{-1}h^{3r}+n^{-2}).\end{aligned} \tag{2.5.26}$$

Similarly, we can show that

$$E(\bar{w}_2^{-5}\,\bar{w}_3^2\,\bar{w}_1^4) = \mu_1^4\,\mu_2^{-5}\,\mu_3^2 + 6\,n^{-1}\,\mu_1^2\,\mu_2^{-4}\,\mu_3^2 + O(n^{-1}\,h^{3r} + n^{-2}), \quad (2.5.27)$$

$$E(\bar{w}_2^{-4}\,\bar{w}_4\,\bar{w}_1^4) = \mu_1^4\,\mu_2^{-4}\,\mu_4 + 6\,n^{-1}\,\mu_1^2\,\mu_2^{-3}\,\mu_4 + O(n^{-1}\,h^{3r} + n^{-2}), \quad (2.5.28)$$

and the expectations of all other terms on the right hand side of (2.5.7) are at order of $O(n^{-1}\,h^{3r} + n^{-2})$.

Summarizing (2.5.25) - (2.5.33) we have

$$E\{\ell_h(\theta_q)\} = 1 + n^{-1}\,\beta + n\,\mu_1^2\,\mu_2^{-1} + o(nh^{2r}) + O(h^{3r} + n^{-1}\,h^r + n^{-2}),$$

where $\beta = \frac{1}{6}(3\mu_2^{-2}\,\mu_4 - 2\mu_2^{-3}\,\mu_3^2)$. So the lemma is proved. □

2.5.5 Proof of Theorem 2.3.6

Theorem 2.3.6. *Assume conditions* (2.3.8) *and* (2.3.10). *Then a sufficient condition for*

$$P(\theta_q \in I_{h,d(c_\alpha,\gamma)}) = \alpha + O(n^{-2}) \quad (2.5.29)$$

for either $\gamma = \beta$ or $\gamma = \hat{\beta}$, is that $n^3 h^{2r}$ be bounded. If $f^{(r-1)}(\theta_q) \neq 0$ then the boundedness of $n^3 h^{2r}$ is also necessary for (2.5.29).

Proof: We first prove in the case $\gamma = \beta$, where $\beta = \frac{1}{6}(3\mu_2^{-2}\,\mu_4 - 2\mu_2^{-3}\,\mu_3^2)$. Recall the Edgeworth expansion for the distribution of $\ell_h(\theta_q)$ developed in Section 2.5.3,

$$P\{\ell_h(\theta_q) \le x^2\} = P(\chi_1^2 \le x^2) - \tfrac{1}{6}\,n^{-1}\{6\mu_2^{-1}(n\mu_1)^2 + 3\mu_2^{-2}\,\mu_4 - 2\mu_2^{-3}\,\mu_3^2\}\,x^2\phi(x^2)$$
$$+ o(nh^{2r}) + O(n^{-2}).$$

Thus

$$P\{\ell_h(\theta_q) \le x^2(1 + \beta n^{-1})\} = P\{\chi_1^2 \le x^2(1 + \beta n^{-1})\}$$
$$- \tfrac{1}{6}\,n^{-1}\{6\mu_2^{-1}(n\mu_1)^2 + 3\mu_2^{-2}\,\mu_4 - 2\mu_2^{-3}\,\mu_3^2\}\,x^2\phi(x^2)$$
$$+ o(nh^{2r}) + O(n^{-2}). \quad (2.5.30)$$

Let g_1 denote the density function of the χ_1^2 distribution. Then

$$P\{\chi_1^2 \le x^2(1+\beta n^{-1})\} = P(\chi_1^2 \le x^2) + \beta\, x^2\, g_1(x^2)\, n^{-1} + O(n^{-2}),$$

and $x^2\, g_1(x^2) = x\,\phi(x)$. Substituting the above formulae into (2.5.30), and replacing x^2 by c_α satisfying $P(\chi_1^2 < c_\alpha) = \alpha$, we obtain

$$\begin{aligned} P\{\theta_q \in I_{h,d(c_\alpha,\beta)}\} &= P\{\ell_h(\theta_q) \le c_\alpha\,(1+\beta n^{-1})\} \\ &= \alpha - \tfrac{1}{6}\, n^{-1}\{6\mu_2^{-1}(n\mu_1)^2\}\, c_\alpha \phi(c_\alpha) + o(nh^{2r}) + O(n^{-2}) \\ &= \alpha - \tfrac{1}{6}\, n^{-1}\,\{(r!)^{-2}\,\kappa^2\, f^{(r-1)}(\theta_q)^2 n\, h^{2r}\, q^{-1}(1-q)^{-1}\}\, c_\alpha \phi(c_\alpha) \\ &\quad + o(nh^{2r}) + O(n^{-2}). \end{aligned} \tag{2.5.31}$$

Therefore if $n^3 h^{2r}$ is bounded then $n\,h^{2r} = O(n^{-2})$. From (2.5.31) we immediately see that (2.5.29) holds true. If $f^{(r-1)}(\theta_q) \ne 0$, then (2.5.29) implies that $n\,h^{2r} = O(n^{-2})$, which in turn means that $n^3 h^{2r}$ is bounded.

For the case of $\gamma = \hat\beta$, the proof of (2.5.31) cam be handed similarly. Since $\hat\beta$ is a root-n consistent estimate of β, we have

$$1 + \hat\beta\, n^{-1} = 1 + \beta\, n^{-1} + O_p(n^{-\frac{3}{2}}).$$

Using the delta-method we may show in a way similar to that which we used to derive (2.5.31), that

$$\begin{aligned} P\{\theta_q \in I_{h,d(c_\alpha,\hat\beta)}\} &= \alpha - n^{-1}\,\{(r!)^{-2}\,\kappa^2\, f^{(r-1)}(\theta_q)^2 n\, h^{2r}\, q^{-1}(1-q)^{-1}\}\, c_\alpha \phi(c_\alpha) \\ &\quad + o(nh^{2r}) + O(n^{-\frac{3}{2}}). \end{aligned} \tag{2.5.32}$$

However, by an argument based on the oddness and evenness of polynomials in the Edgeworth expansion for the distribution of $\ell_h(\theta_q)$, for example, by Barndorff-Nielson and Hall (1988), the $O(n^{-\frac{3}{2}})$ term in (2.5.32) is actually $O(n^{-2})$. To be more rigorous, we may apply Edgeworth expansion directly to $\ell_h(\theta_q) - x^2\,\hat\beta\, n^{-1}$, which can be fitted into a smooth function of means model, by using the same method as that for Edgeworth expansion for $\ell_h(\theta_q)$, although the analysis is far more tedious. Therefore,

$$P\{\theta_q \in I_{h,d(c_\alpha,\hat{\beta})}\} = \alpha - n^{-1}\{(r!)^{-2}\kappa^2 f^{(r-1)}(\theta_q)^2 n h^{2r} q^{-1}(1-q)^{-1}\} c_\alpha \phi(c_\alpha)$$
$$+ o(nh^{2r}) + O(n^{-2}).$$

Now, the rest of treatment for the case $\gamma = \hat{\beta}$ is exactly same with that for the case $\gamma = \beta$. □

Appendix 2 Calculation of Cumulants k_ℓ

In this appendix we calculate the cumulants $k_1, k_2, \ldots$ of $n^{\frac{1}{2}}P(V)$ which were used in the proof of Theorem 2.3.4. Let $k^{i_1 i_2 \cdots i_p}$ be the p'th order multivariate cumulants of $V = (V_1, \ldots, V_j)$. According to results given by James and Mayne (1962), the k_j's may be expressed as follows,

$$k_1 = n^{\frac{1}{2}}\{S_j(u) + \tfrac{1}{2} d_{ij}k^{ij} + \tfrac{1}{6} d_{ijk}k^{ijk} + \tfrac{1}{8} d_{ijk\ell}k^{ij}k^{k\ell}\} + O(n^{-\frac{5}{2}}) \quad (2.5.33)$$

$$k_2 = n\{d_i d_j k^{ij} + d_{ij}d_k k^{ijk} + (d_{ijk}d_\ell + \tfrac{1}{2} d_{ik}d_{j\ell})k^{ij}k^{k\ell}\} + O(n^{-2}) \quad (2.5.34)$$

$$\begin{aligned} k_3 = n^{\frac{3}{2}} \{ & d_i d_j d_k k^{ijk} + 3d_{ik}d_j d_\ell k^{ij}k^{k\ell} + \tfrac{3}{2} d_{ij}d_k d_\ell k^{ijk\ell} \\ & + (3d_{ij\ell}d_k d_m + \tfrac{3}{2} d_{i\ell m}d_j d_k + 3d_{ij}d_{k\ell}d_m + 3d_{i\ell}d_{jm}d_k)k^{ijk}k^{\ell m} \\ & + \tfrac{3}{2} d_{ijkm}d_\ell d_n k^{ij}k^{k\ell}k^{mn} + (3d_{ijk}d_{\ell m}d_n \\ & + 3d_{ikm}d_{j\ell}d_n + d_{ik}d_{jm}d_{in})k^{ij}k^{k\ell}k^{mn}\} + O(n^{-\frac{5}{2}}), \end{aligned} \quad (2.5.35)$$

$$\begin{aligned} k_4 = n^2\{ & d_i d_j d_k d_\ell k^{ijk\ell} + 12 d_{i\ell}d_j d_k d_m k^{ijk}k^{\ell m} \\ & + (4d_{ikm}d_j d_\ell d_n + 12 d_{ik}d_{jm}d_\ell d_n)k^{ij}k^{k\ell}k^{mn}\} + O(n^{-2}), \end{aligned} \quad (2.5.36)$$

$$k_\ell = O(n^{-(\ell-2)/2}), \qquad \ell \geq 5,$$

where

$$d_{j_1 \cdots j_m} = \Big(\prod_{\ell=1}^{m} \partial/\partial u_{j_\ell}\Big) S_j(u_1, \ldots, u_j)\big|_{u=\mu}.$$

It may be shown after some calculations that

$$d_1 = \mu_2^{-\frac{1}{2}} + \tfrac{2}{3}\mu_3\mu_1 + O(n^{-\frac{1}{2}}h^r + n^{-\frac{3}{2}}),$$

$$d_2 = -\tfrac{1}{2}\mu_3^{-\frac{3}{2}}\mu_1 + O(n^{-\frac{1}{2}}h^r + n^{-\frac{3}{2}}), \qquad d_\ell = O(h^{2r}), \qquad \ell \geq 3$$

$$d_{11} = \tfrac{2}{3}\mu_2^{-\frac{5}{2}}\mu_3 + O(h^r), \qquad d_{12} = -\tfrac{1}{2}\mu_2^{-\frac{3}{2}} + O(h^r),$$

$$d_{\ell m} = O(h^r), \qquad \text{for all other second derivatives},$$

$$d_{111} = -\tfrac{3}{2}\mu_2^{-\frac{7}{2}}\mu_4 + \tfrac{8}{3}\mu_2^{-\frac{9}{2}}\mu_3^2 + O(h^r), \qquad d_{112} = -\tfrac{5}{2}\mu_2^{-\frac{7}{2}}\mu_3 + O(h^r),$$

$$d_{113} = \tfrac{2}{3}\mu_2^{-\frac{5}{2}} + O(h^r), \qquad d_{122} = \tfrac{3}{4}\mu_2^{-\frac{5}{2}} + O(h^r),$$

$$d_{ijk} = O(h^r), \qquad \text{for all other third derivatives}.$$

Moreover, we have

$$\begin{aligned}
k^{11} &= n^{-1}(\mu_2 - \mu_1^2), \qquad k^{12} = n^{-1}(\mu_3 - \mu_1\mu_2), \\
k^{13} &= n^{-1}(\mu_4 - \mu_1\mu_3),, \qquad k^{22} = n^{-1}(\mu_4 - \mu_2^2), \\
k^{111} &= n^{-2}(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3), \qquad k^{112} = n^{-2}(\mu_4 - 2\mu_1\mu_3 + 2\mu_1\mu_2 - \mu_2^2), \\
k^{1111} &= n^{-3}(\mu_4 - 2\mu_2^2) + O(n^{-3}\mu_1), \qquad k^{1112} = n^{-3}(\mu_5 - 2\mu_2\mu_3) + O(n^{-3}\mu_1).
\end{aligned}$$

Substituting the above derivatives and the multivariate cumulants of V into (2.5.33)–(2.5.36), we are able to prove that

$$\begin{aligned}
k_1 &= n^{\frac{1}{2}} s_j(\mu) - \tfrac{1}{6}\mu_2^{-\frac{3}{2}}\mu_3 n^{-\frac{1}{2}} + O(n^{-\frac{1}{2}}h^r + n^{-\frac{3}{2}}), \\
k_2 &= \sigma^2 + (\tfrac{1}{2}\mu_2^{-2}\mu_4 - \tfrac{13}{36}\mu_2^{-3}\mu_3^2)n^{-1} + O(n^{-1}h^r + n^{-2}), \\
k_3 &= O(n^{-\frac{1}{2}}h^r), \quad k_4 = O(n^{-1}h^r), \quad k_\ell = O(n^{-(\ell-2)/2}) \text{ for } \ell \geq 5.
\end{aligned}$$

CHAPTER THREE

ON THE ACCURACY OF EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR LINEAR REGRESSION MODEL

3.1. Introduction

In Chapter 2 we considered constructing empirical likelihood confidence intervals for a population quantile. We also mentioned there that the quantile case is non-standard in the sense that the parameter of interest (i.e. quantile) cannot be represented as a smooth function of means. In this chapter we consider our second non-standard case, which is to construct empirical likelihood confidence region for the regression coefficient vector of a linear regression model. We shall see shortly that due to the presence of the fixed design points, the observed random variables are independent but not identically distributed. So it is not the standard i.i.d case any more.

Let us consider a linear regression model of the form

$$Y_i = x_i\beta + \epsilon_i, \qquad 1 \leq i \leq n, \tag{3.1.1}$$

where β is a $p \times 1$ vector of unknown parameters and x_i is a $1 \times p$ vector of the i'th fixed design point, for which scalar Y_i is the response. We allow the ϵ_i's to be heteroscedastic, that is, the ϵ_i's are independent random variables with mean zero and variances $\sigma^2(x_i)$. The data are observed in the form $\{(x_i, Y_i) \,|1 \leq i \leq n\}$.

A classical problem for linear regression model is that of how to construct confidence regions for β when the distribution functions of ϵ_i's are unknown. In these nonparametric settings the bootstrap has been used to construct confidence regions for β. But one drawback of the bootstrap is that it needs some subjective instructions on the shapes and orientations of confidence regions. Empirical likelihood methods, as an alternative to the bootstrap method for constructing confidence regions nonparametrically, were introduced by Owen (1988,1990). An important feature of empirical likelihood is that it uses only the data to determine the shape and orientation of a confidence region. Furthermore, in certain regular cases as pointed out in Chapter 1, empirical likelihood confidence regions are Bartlett correctable, meaning that simple empirical adjustments for scale can reduce coverage error from $O(n^{-1})$ to $O(n^{-2})$.

Empirical likelihood methods were proposed by Owen (1991) for constructing confidence regions for β in the model (3.1.1). He derived a nonparametric version of Wilks' theorem, ensuring that empirical likelihood confidence regions for β have correct asymptotic coverages. However, there are still two questions to be answered. They are, "How accurate are the empirical likelihood confidence regions ?" and "Are the empirical likelihood confidence regions Bartlett correctable ?"

This chapter aims to answer these two questions. We demonstrate in Section 3.2 that the coverage errors of empirical likelihood confidence regions for β are of order n^{-1}. In Section 3.3 we show that Bartlett correction may be used to reduce the order of magnitude of the coverage errors to n^{-2} . An empirical Bartlett correction is given, which allows one to practically implement the Bartlett correction. A simulation study is presented in Section 3.4. Detailed proofs and calculations of cumulants are given in Section 3.5 and Appendix 3, respectively.

We close this section with some notation. Let X be an $n \times p$ matrix with x_i as the i'th row, let β_{LS} denote the least squares estimator of β, $\beta_{LS} = (X^T X)^{-1} \sum x_i Y_i$, and put $\hat{\epsilon}_i = Y_i - x_i \beta_{LS}$.

3.2 Wilks' Theorem and Coverage Accuracy

As mentioned in Section 3.1, Owen (1991) proved a nonparametric version of Wilks' theorem for the empirical log-likelihood ratio of β, which enables us to construct confidence regions with correct asymptotic coverages. In this section we investigate the second order property of those confidence regions. We first give a Taylor expansion for empirical log-likelihood ratio, denoted by $\ell(\beta)$. Then we set up an Edgeworth expansion for the distribution function of $\ell(\beta)$, which allows us to evaluate coverage accuracy of empirical likelihood confidence regions.

For the linear regression model (3.1.1) we know that

$$E(Y_i | x_i) = x_i\beta, \qquad E(\epsilon_i) = 0, \qquad \mathrm{Var}(\epsilon_i) = \sigma^2(x_i).$$

Notice that we assume the variance of ϵ_i is related to the i'th fixed design point x_i, which implies heteroscedascity of the model. We define auxilliary variables $z_i = x_i^T(Y_i - x_i\beta)$, for $1 \le i \le n$, and

$$V_n = n^{-1}\sum_{i=1}^{n} \mathrm{Cov}(z_i) = n^{-1}\sum_{i=1}^{n} x_i^T x_i \sigma^2(x_i),$$

and let v_{1n} and v_{pn} denote the largest and smallest eigenvalues of V_n, respectively.

The problem of testing whether or not β is the true parameter is equivalent to testing whether $E\{z_i\} = 0$, for $1 \le i \le n$. Let $p_1, \cdots, p_n$ be nonnegative numbers summing to unity. Then the empirical log-likelihood ratio, evaluated at true parameter value β, is defined by

$$\ell(\beta) = -2 \min_{\sum p_i z_i = 0} \sum \log(np_i).$$

Using the Lagrange multiplier method, the optimal value for p_i may be shown to be given by

$$p_i = \frac{1}{n}\frac{1}{1+\lambda^T z_i} \qquad 1 \le i \le n.$$

This gives

$$\ell(\beta) = 2\sum \log\left(1 + \lambda^T z_i\right),$$

where λ is a $p \times 1$ vector satisfying

$$n^{-1}\sum \frac{z_i}{1+\lambda^T z_i} = 0.$$

In terms of studentized variables $w_i = V_n^{-\frac{1}{2}} z_i$, for $1 \le i \le n$, we have

$$\ell(\beta) = 2 \sum \log\left(1 + \lambda^T w_i\right), \tag{3.2.1}$$

where λ satisfies

$$n^{-1} \sum \frac{w_i}{1 + \lambda^T w_i} = 0. \tag{3.2.2}$$

Since analytic solution of equations (3.2.1) and (3.2.2) can rarely be achieved, we have to derive an asymptotic expansion for $\ell(\beta)$. To this end, we assume the following regularity condition.

There exist positive constants C_1 and C_2 such that uniformly in n,

$$C_1 < v_{pn} \le v_{1n} < C_2; \text{ and } n^{-2} \sum_{j=1}^{n} E\|z_j\|^4 \to 0, \text{ where } \|\ \| \text{ denotes} \tag{3.2.3}$$

the Euclidean norm.

Under condition (3.2.3), Owen (1991) showed that the λ appearing in (3.2.2) satisfies

$$\lambda = O_p(n^{-\frac{1}{2}}).$$

We define

$$\overline{\alpha}^{j_1 \cdots j_k} = n^{-1} \sum E\left(w_i^{j_1} \cdots w_i^{j_k}\right),$$
$$A^{j_1 \cdots j_k} = n^{-1} \sum \left(w_i^{j_1} \cdots w_i^{j_k} - \overline{\alpha}^{j_1 \cdots j_k}\right),$$

where w_i^j is the j'th component of w_i. In particular, $\overline{\alpha}^j = 0$, $\overline{\alpha}^{j\,k} = \delta^{j\,k}$, $\delta^{j\,k}$ is the Kronecker delta. Notice that the $\overline{\alpha}^{j_1 \cdots j_k}$ is a generalization of $\alpha^{j_1 \cdots j_k}$ defined in (1.2.9) for our current independent but not identically distributed case.

Notice that $\ell(\beta)$, given by (3.2.1) and (3.2.2), is similar to the empirical log-likelihood ratio for means in the independent and identically distributed case. The only difference is that $\{w_i\}_{i=1}^n$ are independent but not identically distributed random variables due to the presence of the fixed design points. However, by modifying the expansion (1.2.13) we may obtain the following expansion for $\ell(\beta)$,

$$\begin{aligned} n^{-1}\ell(\beta) = & A^j A^j - A^{j\,k} A^j A^k + \tfrac{2}{3}\overline{\alpha}^{j\,k\,l} A^j A^k A^l + A^{j\,l} A^{k\,l} A^j A^K \\ & + \tfrac{2}{3} A^{j\,k\,l} A^j A^k A^l - 2\overline{\alpha}^{j\,k\,m} A^{l\,m} A^j A^k A^l + \overline{\alpha}^{j\,k\,n} \overline{\alpha}^{l\,m\,n} A^j A^k A^l A^m \\ & - \tfrac{1}{2}\overline{\alpha}^{j\,k\,l\,m} A^j A^k A^l A^m + O_p(n^{-\frac{5}{2}}). \end{aligned} \tag{3.2.4}$$

Based on expansion (3.2.4), we have

$$\ell(\beta) = \left(n^{\frac{1}{2}} R^T\right) \left(n^{\frac{1}{2}} R\right) + O_p(n^{-\frac{3}{2}}), \tag{3.2.5}$$

where $R = R_1 + R_2 + R_3$ is a p-dimensional vector and $R_l = O_p\left(n^{-l/2}\right)$ for $l = 1, 2, 3$. Comparing terms in (3.2.4) with those in (3.2.5) yields,

$$\begin{aligned}
R_1^j &= A^j, \\
R_2^j &= -\tfrac{1}{2} A^{j\,k} A^k + \tfrac{1}{3} \overline{\alpha}^{j\,k\,m} A^k A^m \quad \text{and} \\
R_3^j &= \tfrac{3}{8} A^{j\,m} A^{k\,m} A^k + \tfrac{1}{3} A^{j\,k\,m} A^k A^l - \tfrac{5}{12} \overline{\alpha}^{j\,k\,m} A^{l\,m} A^k A^l \\
&\quad - \tfrac{5}{12} \overline{\alpha}^{k\,l\,m} A^{j\,m} A^k A^l + \tfrac{4}{9} \overline{\alpha}^{j\,k\,n} \overline{\alpha}^{l\,m\,n} A^m A^k A^l - \tfrac{1}{4} \overline{\alpha}^{j\,k\,l\,m} A^m A^k A^l,
\end{aligned} \tag{3.2.6}$$

where R_l^j is the j'th component of R_l. In particular,

$$R_1 = n^{-1} \sum w_i = n^{-1} \sum V_n^{-\frac{1}{2}} x_i^T \left(Y_i - x_i \beta\right).$$

The leading term in (3.2.5) is

$$\begin{aligned}
n R_1^T R_1 &= n^{-1} \{ \sum (Y_i - x_i\beta) x_i \} V_n^{-1} \{ \sum x_i^T (Y_i - x_i \beta) \} \\
&= (\beta_{LS} - \beta)^T (X^T X) \{ \sum x_i^T x_i \sigma^2 (x_i) \}^{-1} (X^T X) (\beta_{LS} - \beta).
\end{aligned}$$

It is well-known that, for the heteroscedastic linear regression model (3.1.1),

$$\operatorname{Var}(\beta_{LS}) = (X^T X)^{-1} \{ \sum x_i^T x_i \sigma^2 (x_i) \} (X^T X)^{-1}.$$

Thus,

$$\begin{aligned}
\ell(\beta) &= n^{-1} \{ \sum (Y_i - x_i \beta) x_i \} V_n^{-1} \{ \sum x_i^T (Y_i - x_i \beta) \} + O_p(n^{-\frac{1}{2}}) \\
&= (\beta_{LS} - \beta)^T \{ \operatorname{Var}(\beta_{LS}) \}^{-1} (\beta_{LS} - \beta) + O_p(n^{-\frac{1}{2}}).
\end{aligned} \tag{3.2.7}$$

Also,

$$\beta_{LS} - \beta = (X^T X)^{-1} \sum x_i^T \epsilon_i = (X^T X)^{-1} n \overline{Z},$$

where $Z_i = x_i^T (y_i - \beta x_i)$. By Cramér-Wold device, we have

$$\overline{Z} \xrightarrow{d} N\left(0, n^{-1} V_n\right).$$

Thus $\beta_{LS} - \beta$ converges to $N\{0, \mathrm{Var}(\beta_{LS})\}$ in distribution under condition (3.2.3). Therefore,

$$(\beta_{LS} - \beta)^T \mathrm{Var}^{-1}(\beta_{LS})(\beta_{LS} - \beta) \xrightarrow{d} \chi_p^2, \qquad \text{as } n \to \infty.$$

where $\to^d$ denotes converging in distribution. Hence we obtain

$$P\{\ell(\beta) < c\} = P\left(\chi_p^2 < c\right) + o(1) \qquad \text{as } n \to \infty, \tag{3.2.8}$$

which is a nonparametric version of Wilks' Theorem, first proved by Owen (1991).

From (3.2.7) we see that $\ell(\beta)$ implicitly uses the true variance of β_{LS} to construct confidence regions for β. This is an advantage of empirical likelihood over other resampling techniques, such as the jackknife and the bootstrap, which depend on explicit estimates of $\mathrm{Var}(\beta)$ and consequently pose problems resulting from the quality of these estimators. This point was noted by Wu (1986). Empirical likelihood can avoid this problem, reflecting the feature "let the data themselves decide". And also note that the first term on the right of (3.2.7) is different from that given by Owen (1991), who uses an estimate of $\mathrm{Var}(\beta_{LS})$. However, the difference has no first order effect.

Using (3.2.8), a confidence region for β with nominal coverage level α can be constructed as follows. First find from the χ_p^2 tables the value c_α such that

$$P\left(\chi_p^2 < c_\alpha\right) = \alpha.$$

Then $R_\alpha = \{\, \beta \,|\, \ell(\beta) < c_\alpha \,\}$ is the α-level confidence region for β, and (3.2.8) ensures that it has correct asymptotic coverage.

Before discussing the coverage accuracy of R_α, let us define $j_1 = (p^2 + p)/2$, $j_2 = j_1/2 + p(p+1)(2p+1)/12$, and

$$\overline{U} = (A^1, \cdots, A^p, A^{11}, \cdots, A^{pp}, A^{111}, \cdots, A^{ppp}),$$

being the $p + j_1 + j_2$-dimensional vector consisting of all distinct first three order multivariate central moments of $w_i = V_n^{-\frac{1}{2}} z_i$'s. Note that there are j_1 and j_2 distinct second and third order multivariate central moments in $\overline{U}$. Let $T_n = n\text{Cov}(\overline{U}) = n^{-1}\sum \text{Cov}(U_i)$ where

$$U_i = \left[x_i^T V_n^{-\frac{1}{2}}\epsilon_i, (x_i \otimes x_i)B_1^T\{\epsilon_i^2 - E(\epsilon_i^2)\}, (x_i \otimes x_i \otimes x_i)B_2^T\{\epsilon_i^3 - E(\epsilon_i^3)\}\right]^T.$$

Let S_i be the $(p-i)\times p$ matrix obtained by removing top i rows of $V_n^{-\frac{1}{2}}$, and $V_{nj}^{-\frac{1}{2}}$ be the j-th row of $V_n^{-\frac{1}{2}}$. Clearly $S_o = V_n^{-\frac{1}{2}}$ and $S_{p-1} = V_{np}^{-\frac{1}{2}}$. We define $j_1 \times p^2$ and $j_2 \times p^3$ matrices B_1 and B_2 as follows,

$$B_1 = \begin{pmatrix} V_{n1}^{-\frac{1}{2}} \otimes S_o \\ \vdots \\ V_{np}^{-\frac{1}{2}} \otimes S_{p-1} \end{pmatrix} \text{ and } B_2 = \begin{pmatrix} V_{n1}^{-\frac{1}{2}} \otimes V_{n1}^{-\frac{1}{2}} \otimes S_o \\ \vdots \\ V_{n1}^{-\frac{1}{2}} \otimes V_{np}^{-\frac{1}{2}} \otimes S_{p-1} \\ V_{n2}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \otimes S_2 \\ \vdots \\ V_{n2}^{-\frac{1}{2}} \otimes V_{np}^{-\frac{1}{2}} \otimes S_{p-1} \\ \vdots \\ V_{np}^{-\frac{1}{2}} \otimes V_{np}^{-\frac{1}{2}} \otimes S_{p-1} \end{pmatrix}$$

Then it may be shown that

$$T_n = \begin{pmatrix} I_p & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12}^T & \Gamma_{22} & \Gamma_{23} \\ \Gamma_{13}^T & \Gamma_{23}^T & \Gamma_{33} \end{pmatrix}$$

where

$$\Gamma_{12} = V_n^{-\frac{1}{2}}\{n^{-1}\sum x_i^T(x_i \otimes x_i)E(\epsilon_i^3)\}B_1^T,$$

$$\Gamma_{13} = V_n^{-\frac{1}{2}}\{n^{-1}\sum x_i^T(x_i \otimes x_i \otimes x_i)E(\epsilon_i^4)\}B_2^T,$$

$$\Gamma_{22} = B_1\left[n^{-1}\sum (x_i^T \otimes x_i^T)(x_i \otimes x_i)\{E(\epsilon_i^4) - E^2(\epsilon_i^2)\}\right]B_1^T,$$

$$\Gamma_{23} = B_1\left[n^{-1}\sum (x_i^T \otimes x_i^T)(x_i \otimes x_i \otimes x_i)\{E(\epsilon_i^5) - E(\epsilon_i^2)E(\epsilon_i^3)\}\right]B_2^T,$$

$$\Gamma_{33} = B_2\left[n^{-1}\sum (x_i^T \otimes x_i^T \otimes x_i^T)(x_i \otimes x_i \otimes x_i)\{E(\epsilon_i^6) - E^2(\epsilon_i^3)\}\right]B_2^T,$$

To derive an Edgeworth expansion for the distribution of $\ell(\beta)$, we have to use Theorem 1.3.3. Notice that the first condition of (1.3.5) demands the smallest eigenvalue of T_n is bounded away from zero.

We establish an Edgeworth expansion for the distribution of $\ell(\beta)$ in the following theorem.

Theorem 3.2.1: *Assume that*

(i) there exist positive constants C_1, C_2 such that uniformly in n, $C_1 \leq v_{pn} \leq v_{1n} \leq C_2$; (ii) the $\|x_i\|$'s for $1 \leq i \leq n$ are uniformly bounded; (iii) $\sup_n n^{-1} \sum_{j=1}^{n} E|\epsilon_j|^{15} < \infty$; (iv) for every positive τ, $\lim_{n \to \infty} n^{-1} \sum_{j=1}^{n} \int_{|\epsilon_j| > \tau n^{\frac{1}{2}}} |\epsilon_j|^{15} = 0$; (v) the characteristic function g_n of ϵ_n satisties Cramér's condition, i.e. for every positive b, $\overline{\lim}_{n \to \infty} \sup_{|t| > b} |g_n(t)| < 1$; (vi) the smallest eigenvalue of T_n is bounded away from zero. (3.2.9)

Then

$$P\{\ell(\beta) < c_\alpha\} = \alpha - a\, c_\alpha\, g_p(c_\alpha)\, n^{-1} + O(n^{-\frac{3}{2}}),$$

where g_p is the density of χ^2_p distribution, $P\left(\chi^2_p < c_\alpha\right) = \alpha$, and

$$a = p^{-1}\left(\tfrac{1}{2}\overline{\alpha}^{j\,j\,m\,m} - \tfrac{1}{3}\overline{\alpha}^{j\,k\,m}\overline{\alpha}^{j\,k\,m}\right). \tag{3.2.10}$$

Theorem 3.2.1 states that the coverage error of the empirical likelihood confidence region R_α is of order n^{-1}, that is

$$P\{\ell(\beta) < c_\alpha\} = \alpha + O(n^{-1}).$$

In (iii) of condition (3.2.9) we assume the average 15-th moment of ϵ_j's is uniformly bounded. This is to ensure, together with (ii) of condition (3.2.9), that the average fifth moment of U_j's is uniformly bounded in order to obtain a uniform error term at order of $n^{-\frac{3}{2}}$ in the Edgeworth expansion for the distribution of $\overline{U}$.

From (3.2.10) and the definitions of $\overline{\alpha}^{j\,j\,m\,m}$ and $\overline{\alpha}^{j\,k\,m}$, we have

$$a = p^{-1}\left[\tfrac{1}{2}n^{-1}\textstyle\sum_{i=1}^{n} E(\epsilon_i^4)\left(x_i V_n^{-1} x_i^T\right)^2 - \tfrac{1}{3}n^{-2}\sum_{i,l}\{E(\epsilon_i^3)E(\epsilon_l^3)(x_i V_n^{-1} x_l)^3\}\right].$$

This reveals that the coverage error depends on a combination of the following five factors: (1) the moments of ϵ_i's, (2) the nominal coverage level, (3) the configuration of the fixed design points, (4) the sample size n, and (5) dimension, p.

3.3. Bartlett Correction

In Section 3.2 we showed that the coverage errors of empirical likelihood confidence regions for β are of order n^{-1}. It is well-known that part of the coverage error is due to the fact that the mean of $\ell(\beta)$ does not agree with the mean of χ_p^2, that is $E\{\ell(\beta)\} \neq p$. The coverage accuracy of empirical likelihood confidence region can be improved by rescaling $\ell(\beta)$ to reduce this disagreement. We demonstrate in this section that the empirical likelihood confidence region for β is Bartlett correctable. Thus, a simple empirical correction for scale can reduce the size of coverage error from order n^{-1} to order n^{-2}. For practical implementation of Bartlett correction, we propose an empirical Bartlett correction.

From expansion (3.2.4) we may obtain an expansion for $E\{\ell(\beta)\}$ as follows,

$$E\{\ell(\beta)\} = p(1 + an^{-1}) + O(n^{-2}), \tag{3.3.1}$$

where a is given by (3.2.10). The Bartlett correctability of empirical likelihood confidence regions for β is discussed in the following theorem.

Theorem 3.3.1: *Assume condition (3.2.9). For any $c_\alpha > 0$,*

$$P\{\ell(\beta) < c_\alpha\left(1 + a\,n^{-1}\right)\} = \alpha + O(n^{-2}).$$

where $P\left(\chi_p^2 < c_\alpha\right) = \alpha$.

However, the Bartlett factor a is usually unknown in practice, because V_n and the moments of ϵ_i's are unknown. Suppose $\hat{a}$ is a root-n consistent estimate of a. We claims that by slightly modifying condition (3.2.9), Theorem 3.3.1 holds true when a is replaced by its root-n consistent estimate $\hat{a}$. To appreciate this, note that

$$1 + \hat{a}\,n^{-1} = 1 + a\,n^{-1} + O_p(n^{-\frac{3}{2}}).$$

Using the delta-method, we may show that

$$\begin{aligned} P\{\ell(\beta) < c_\alpha\left(1 + \hat{a}\,n^{-1}\right)\} &= P\{\ell(\beta) < c_\alpha\left(1 + a\,n^{-1}\right)\} + O(n^{-\frac{3}{2}}) \\ &= \alpha + O(n^{-\frac{3}{2}}). \end{aligned} \tag{3.3.2}$$

To be more rigorous, by slightly modifying condition (3.2.9) we may develop an Edgeworth expansion directly to $\ell(\beta) - c_\alpha \, \hat{a} \, n^{-1}$, under a smooth function of mean model although the analysis is far more tedious. By an argument based on the oddness and evenness of polynomials in the Edgeworth expansion, the $O(n^{-\frac{3}{2}})$ term in (3.3.2) is actually $O(n^{-2})$.

In the rest of this section we give a root-n consistent estimate of a. From (3.2.10) we know that the Bartlett correction is given by

$$a = p^{-1} \left(\tfrac{1}{2} \overline{\alpha}^{j\,j\,m\,m} - \tfrac{1}{3} \overline{\alpha}^{j\,k\,m} \overline{\alpha}^{j\,k\,m} \right),$$

where

$$\overline{\alpha}^{j\,k\,m} = n^{-1} \sum_{i=1}^{n} E\left(\epsilon_i^3\right) V_{n\,j}^{-\frac{1}{2}} x_i^T V_{n\,k}^{-\frac{1}{2}} x_i^T V_{n\,m}^{-\frac{1}{2}} x_i^T,$$

$$\overline{\alpha}^{j\,j\,m\,m} = n^{-1} \sum_{i=1}^{n} E\left(\epsilon_i^4\right) \left(x_i V_n^{-1} x_i^T\right)^2,$$

and $V_{n\,j}^{-\frac{1}{2}}$ is the j'th row of $V_n^{-\frac{1}{2}}$. We define

$$\hat{V}_n = n^{-1} \sum x_i^T x_i \hat{\epsilon}_i^2,$$

which is an estimator of covariance matrix V_n. Accordingly, we let $\hat{V}_n^{-1}$ be the inverse matrix of $\hat{V}_n$ and $\hat{V}_n^{-\frac{1}{2}}$ be the positive definite square root matrix of $\hat{V}_n^{-1}$. Now an estimate of a, $\hat{a}$ say, may be defined as follows:

$$\hat{a} = p^{-1} \left(\tfrac{1}{2} \hat{\overline{\alpha}}^{j\,j\,m\,m} - \tfrac{1}{3} \hat{\overline{\alpha}}^{j\,k\,m} \overline{\alpha}^{j\,k\,m} \right), \tag{3.3.3}$$

where

$$\hat{\overline{\alpha}}^{j\,k\,m} = n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^3 \hat{V}_{n\,j}^{-\frac{1}{2}} x_i^T \hat{V}_{n\,k}^{-\frac{1}{2}} x_i^T \hat{V}_{n\,m}^{-\frac{1}{2}} x_i^T$$

and

$$\hat{\overline{\alpha}}^{j\,j\,m\,m} = n^{-1} \sum_{i=1}^{n} \hat{\epsilon}_i^4 (x_i \hat{V}_n^{-1} x_i^T)^2.$$

We can see that $\hat{\overline{\alpha}}^{j\,k\,m}$ and $\hat{\overline{\alpha}}^{j\,j\,m\,m}$ are established by replacing ϵ_i, $V_{n\,j}^{-\frac{1}{2}}$ in $\overline{\alpha}^{j\,k\,m}$ and $\overline{\alpha}^{j\,j\,m\,m}$ with their corresponding estimates $\hat{\epsilon}_i$ and $\hat{V}_{n\,j}^{-\frac{1}{2}}$, where $\hat{V}_{n\,j}^{-\frac{1}{2}}$ denotes the j'th row of $\hat{V}_n^{-\frac{1}{2}}$.

We wish to prove that $\hat{a}$ is a root-n consistent estimate of a. To this end we assume that

$$\begin{gathered}\text{there exist positive constants } C_1,\ C_2 \text{ such that uniformly in n,}\\ C_1 \le v_{pn} \le v_{1n} \le C_2;\ \text{and there exist constants } q_1, q_2 > 0 \text{ such that}\\ q_1 \le \inf \|x_i\| \le \sup \|x_i\| \le q_2;\ \text{and } \sup_n n^{-1} \sum E(\epsilon_i^8) \le +\infty.\end{gathered} \tag{3.3.4}$$

Theorem 3.3.2: *Assume condition (3.3.4). Then,*

$$\hat{a} = a + O_p(n^{-\frac{1}{2}}).$$

The proof of Theorem 3.3.2 is deferred to Section 3.5.

After some simplification we may show that the empirical Bartlett factor $\hat{a}$ has the following explicit form:

$$\hat{a} = p^{-1}\left[\frac{1}{2}n^{-1}\sum_{i=1}^{n} \hat{\epsilon}_i^4 \left(x_i \hat{V}_n^{-1} x_i^T\right)^2 - \frac{1}{3}n^{-2}\sum_{i,l} \{\hat{\epsilon}_i^3 \hat{\epsilon}_l^3 (x_i \hat{V}_n^{-1} x_l)^3\}\right].$$

In some special cases, $\hat{a}$ has a simpler form, as we now indicate.

(1) If $\epsilon_1, \cdots, \epsilon_n$ are i.i.d, which implies that model (3.1.1) is a homoscedastic regression model, then

$$\hat{a} = p^{-1}n\left[\frac{1}{2}\hat{\mu}_{4\,\epsilon}\hat{\sigma}^{-4}\sum \{x_i\,(X^T X)^{-1}\,x_i^T\}^2 - \frac{1}{3}\hat{\mu}_{3\,\epsilon}^2\hat{\sigma}^{-6}\sum_{i\,l} \{x_i\,(X^T X)^{-1}\,x_l^T\}\right],$$

where $\hat{\mu}_{k\,\epsilon} = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^k$ for $k = 3, 4$, $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i^2$.

(2) If $\epsilon_1, \cdots, \epsilon_n$ are i.i.d and have a symmetric distribution then the model implies $E(\epsilon_i^3) = 0$, and we may take

$$\hat{a} = p^{-1}n\frac{1}{2}\hat{\mu}_{4\,\epsilon}\hat{\sigma}^{-4}\sum \{x_i\,(X^T X)^{-1}\,x_i^T\}^2.$$

Our simulation results in the next section show that $\hat{a}$ is a reliable estimator of a.

3.4. Simulation Study

In this section we use Monte Carlo simulation to examine the coverages of the empirical likelihood confidence regions proposed in previous sections. Under consideration is the following simple linear regression model:

$$Y_i = 1 + x_i^o + \epsilon_i, \quad i = 1, \cdots, n.$$

The data set x_i^0 for $1 \le i \le 150$ is displayed in Table 3.1. For sample size $n \le 150$, we use the first n x_i^o as the fixed design points. Four error patterns were considered. They are two homoscedastic error patterns $\epsilon_i = N(0,1)$ and $\epsilon_i = \mathcal{E}(1.00) - 1.00$, and two heteroscedastic error patterns $\epsilon_i = (1/2x_i^0)^{1/2}\ N(0,1)$ and $\epsilon_i = (1/2x_i^0)^{1/2}\ \{\mathcal{E}(1.00) - 1.00\}$, where $N(0,1)$ and $\mathcal{E}(1.00)$ are random variables with standard normal distribution and exponential distribution with unit mean, respectively. For each of these four error patterns we chose sample sizes $n = 30, 50, 100, 150$, and nominal coverage levels $\alpha = 0.90, 0.95$. The normal and exponential random variables were generated by the routines of Press et al. (1989).

We give in Table 3.2 the coverages of the uncorrected confidence regions and two corrected confidence regions based on 20,000 simulations. One of the corrected confidence regions uses the theoretical Bartlett correction a, another uses the empirical Bartlett correction $\hat{a}$. Since we know the error pattern, sample size and nominal coverage level α, we can calculate the theoretical coverages up to second order by using Edgeworth expansion in Theorem 3.2.1. Because the theoretical coverages can be computed without simulation, we call these "predicted coverages". We compare the "predicted coverages" with the uncorrected coverages in order to see if the theoretical results are consistent with the empirical outputs. Also, standard errors are given for each simulated coverage and these serve as one of the criteria for comparing accuracies among different kinds of simulated coverages.

The following conclusions may be drawn from the results shown in Table 3.2:

1) The simulated uncorrected coverages converge to the "predicted coverages"

as n increases. This empirically justifies the Edgeworth expansion developed in Theorem 3.2.1.

2) Standard errors and absolute coverage errors both show that the Bartlett corrected confidence regions have more accurate coverage than corresponding uncorrected ones.

3) The empirically corrected confidence regions perform similarly to their theoretically corrected counterparts, except for the cases of skewed error patterns with sample sizes n=30 and 50. It seems that we need a larger sample size to ensure $\hat{a}$ as a good estimator of a when the errors are skewed,

Comparing Table 3.2(a) with Table 3.2(b), we observe that skewness in the error patterns reduces the overall coverages. However this has little surprise for us since it has been foreseen by their corresponding "predicted coverages". In the examples considered we see some reduction in coverages caused by heterscedasticity when n is small. Nevertheless there is no clear evidence to say generally that heterscedasticity reduces coverage accuracy when sample size is large. Our theory shows that real coverage depends on the configuration of the fixed design points and the moments of the residuals when sample size, nominal coverage level and dimensionality are all fixed.

TABLE 3.1: The data set x_i^0 for $1 \le i \le 150$.

i	x_i^0	i	x_i^0	i	x_i^0	i	x_i^0	i	x_i^0
1	1.00	31	8.90	61	14.89	91	23.80	121	37.20
2	1.40	32	9.30	62	15.01	92	24.10	122	37.60
3	1.50	33	9.70	63	15.67	93	24.20	123	37.80
4	1.70	34	9.90	64	15.71	94	24.70	124	38.30
5	2.00	35	10.00	65	15.85	95	24.98	125	38.70
6	2.30	36	10.30	66	15.97	96	25.30	126	38.90
7	2.50	37	10.40	67	16.29	97	26.00	127	39.40
8	2.67	38	10.55	68	16.38	98	27.00	128	39.80
9	3.00	39	10.70	69	16.71	99	29.00	129	40.00
10	3.30	40	11.00	70	17.00	100	29.50	130	40.50
11	3.46	41	11.23	71	17.20	101	29.90	131	40.90
12	3.50	42	11.47	72	17.35	102	30.10	132	41.10
13	4.00	43	11.66	73	17.62	103	30.60	133	41.60
14	4.40	44	11.89	74	18.00	104	31.00	134	42.00
15	4.50	45	12.09	75	18.50	105	31.20	135	42.20
16	4.90	46	12.21	76	18.50	106	31.70	136	42.70
17	5.00	47	12.43	77	19.00	107	32.10	137	43.10
18	5.20	48	12.64	78	19.33	108	32.30	138	43.30
19	5.50	49	12.91	79	19.42	109	32.80	139	43.80
20	6.00	50	13.00	80	19.78	110	33.20	140	44.20
21	6.30	51	13.23	81	19.98	111	33.40	141	44.40
22	6.70	52	13.44	82	20.02	112	33.90	142	44.90
23	6.85	53	13.51	83	20.51	113	34.30	143	45.30
24	7.00	54	13.66	84	21.00	114	34.50	144	45.50
25	7.15	55	13.79	85	21.31	115	35.00	145	46.00
26	7.30	56	13.81	86	21.79	116	35.40	146	46.40
27	7.70	57	13.81	87	22.69	117	35.60	147	46.60
28	8.00	58	14.04	88	22.81	118	36.10	148	47.10
29	8.20	59	14.19	89	23.00	119	36.50	149	47.50
30	8.50	60	14.34	90	23.40	120	36.70	150	47.70

TABLE 3.2: Estimated true coverages, from 20,000 simulations, of α-level empirical likelihood confidence regions for β. Rows headed "predic.", "uncorr.", "a" and "$\hat{a}$" give the predicted, uncorrected and Bartlett-corrected coverages respectively. The figures in parentheses are 10^2 times the standard errors associated with the coverage probabilities.

(a) Normal error patterns

ϵ_i		$N(0,1)$		$(\frac{1}{2}x_i^0)^{\frac{1}{2}} N(0,1)$	
n	α	0.90	0.95	0.90	0.95
30	predic.	0.872	0.931	0.868	0.930
	uncorr.	0.839 (0.26)	0.904 (0.21)	0.833 (0.26)	0.897 (0.21)
	a	0.870 (0.24)	0.924 (0.19)	0.867 (0.24)	0.921 (0.19)
	$\hat{a}$	0.867 (0.24)	0.922 (0.19)	0.858 (0.25)	0.915 (0.20)
50	predic.	0.884	0.939	0.884	0.939
	uncorr.	0.872 (0.24)	0.928 (0.18)	0.869 (0.24)	0.927 (0.18)
	a	0.888 (0.22)	0.939 (0.17)	0.886 (0.22)	0.940 (0.17)
	$\hat{a}$	0.887 (0.22)	0.939 (0.17)	0.883 (0.23)	0.938 (0.17)
100	predic.	0.891	0.944	0.889	0.943
	uncorr.	0.890 (0.22)	0.942 (0.17)	0.888 (0.22)	0.941 (0.17)
	a	0.899 (0.21)	0.948 (0.16)	0.899 (0.21)	0.948 (0.16)
	$\hat{a}$	0.899 (0.21)	0.948 (0.16)	0.897 (0.21)	0.947 (0.16)
150	predic.	0.894	0.946	0.894	0.946
	uncorr.	0.894 (0.22)	0.946 (0.16)	0.893 (0.22)	0.948 (0.16)
	a	0.900 (0.21)	0.949 (0.15)	0.898 (0.21)	0.951 (0.15)
	$\hat{a}$	0.900 (0.21)	0.949 (0.15)	0.898 (0.21)	0.951 (0.15)

(b) Exponential error patterns

ϵ_i		$\mathcal{E}(1.00) - 1.00$		$(\frac{1}{2}x_i^0)^{\frac{1}{2}}\,\{\mathcal{E}(1.00) - 1.00\}$	
n	α	0.90	0.95	0.90	0.95
30	predic.	0.835	0.908	0.829	0.904
	uncorr.	0.800 (0.28)	0.864 (0.24)	0.788 (0.29)	0.854 (0.25)
	a	0.863 (0.24)	0.914 (0.20)	0.847 (0.25)	0.906 (0.21)
	$\hat{a}$	0.838 (0.26)	0.895 (0.22)	0.812 (0.28)	0.874 (0.23)
50	predic.	0.863	0.926	0.863	0.926
	uncorr.	0.837 (0.26)	0.900 (0.21)	0.836 (0.26)	0.898 (0.21)
	a	0.872 (0.24)	0.927 (0.18)	0.872 (0.24)	0.924 (0.18)
	$\hat{a}$	0.860 (0.25)	0.919 (0.19)	0.853 (0.25)	0.910 (0.20)
100	predic.	0.880	0.937	0.876	0.934
	uncorr.	0.871 (0.24)	0.926 (0.18)	0.869 (0.24)	0.924 (0.18)
	a	0.893 (0.22)	0.942 (0.17)	0.892 (0.22)	0.942 (0.17)
	$\hat{a}$	0.888 (0.22)	0.938 (0.17)	0.880 (0.22)	0.932 (0.17)
150	predic.	0.888	0.942	0.886	0.941
	uncorr.	0.884 (0.23)	0.939 (0.17)	0.884 (0.23)	0.934 (0.17)
	a	0.896 (0.22)	0.947 (0.16)	0.897 (0.22)	0.945 (0.16)
	$\hat{a}$	0.895 (0.22)	0.946 (0.16)	0.895 (0.22)	0.944 (0.16)

3.5 Proofs

In this section we display proofs Theorems 3.2.1, 3.3.1 and 3.3.2.

3.5.1 Proof of Theorem 3.2.1

Theorem 3.2.1: *Assume that*

(i) there exist positive constants C_1, C_2 such that uniformly in n,

$C_1 \le v_{pn} \le v_{1n} \le C_2$; *(ii) $\|x_i\|'s$ for $1 \le i \le n$ are uniformly bounded;*

(iii) $\sup_n n^{-1} \sum_{j=1}^{n} E|\epsilon_j|^{15} < \infty$; *(iv) for every positive τ,*

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} \int_{|\epsilon_j| > \tau n^{\frac{1}{2}}} |\epsilon_j|^{15} = 0; \quad \text{(v) the characteristic function} \tag{3.2.9}$$

g_n of ϵ_n satisties Cramér's condition, i.e. for every positive b,

$\overline{\lim_{n\to\infty}} \sup_{|t|>b} |g_n(t)| < 1.$

Then
$$P\{\ell(\beta) < c_\alpha\} = \alpha - a\, c_\alpha\, g_p(c_\alpha)\, n^{-1} + O(n^{-\frac{3}{2}}),$$

where g_p is the density of the χ_p^2 distribution, $P\left(\chi_p^2 < c_\alpha\right) = \alpha$, and

$$a = p^{-1}\left(\tfrac{1}{2}\overline{\alpha}^{j\,j\,m\,m} - \tfrac{1}{3}\overline{\alpha}^{j\,k\,m}\overline{\alpha}^{j\,k\,m}\right).$$

Proof: To prove the theorem we first derive an Edgeworth expansion for the distribution of $n^{\frac{1}{2}}R$. By the expansion $R = R_1 + R_2 + R_3$ and expressions for R_l, $l = 1, 2, 3$, calculations deferred to Appendix 3 show that the cumulants $k_1, k_2, \cdots$ of $n^{\frac{1}{2}}R$ have the following forms:

$$\begin{aligned} k_1 &= n^{-\frac{1}{2}}\mu + O\left(n^{-\frac{3}{2}}\right), \\ k_2 &= I_p + n^{-1}\Delta + O\left(n^{-2}\right), \\ k_j &= O\left(n^{-\frac{3}{2}}\right) \qquad j \ge 3, \end{aligned} \tag{3.5.1}$$

where I_p is the $p \times p$ identity matrix, $\mu = (\mu^1, \cdots, \mu^p)^T$, $\Delta = (\Delta_{ij})_{p\times p}$ and

$$\mu^j = -\tfrac{1}{6}\overline{\alpha}^{j\,k\,k}, \quad \Delta_{ij} = \tfrac{1}{2}\overline{\alpha}^{i\,j\,m\,m} - \tfrac{1}{3}\overline{\alpha}^{i\,k\,m}\overline{\alpha}^{j\,k\,m} - \tfrac{1}{36}\overline{\alpha}^{i\,j\,m}\overline{\alpha}^{m\,k\,k}.$$

Let $\mathcal{B}$ be a class of Borel sets satisfying

$$\sup_{B\in\mathcal{B}} \int_{(\partial B)^{\epsilon}} \phi(v)dv = O(\epsilon), \qquad \epsilon \downarrow 0, \tag{3.5.2}$$

where ∂B and $(\partial B)^{\epsilon}$ are the boundary of B and ϵ-neighborhood of ∂B respectively. A formal Edgeworth expansion for the distribution function of $n^{\frac{1}{2}}R$ is given as follows,

$$\sup_{B\in\mathcal{B}} |P(n^{\frac{1}{2}}R \in B) - \int_B \pi(v)\phi(v)dv| = O(n^{-\frac{3}{2}}), \tag{3.5.3}$$

where

$$\pi(v) = 1 + n^{-\frac{1}{2}}\mu^T v + \tfrac{1}{2}n^{-1}\{v^T\ (\mu\mu^T + \Delta)\ v - \mathrm{tr}\ (\mu\mu^T + \Delta)\},$$

$\phi(v)$ is the density function of the standard p-dimensional normal distribution, and tr is the trace operation for square matrices.

Accepting that the Edgeworth expansion (3.5.3) may be justified, we shall develop an Edgeworth expansion for the distribution of $\ell(\beta)$. Put

$$H = (h_{ij})_{p\times p} = \mu\mu^T + \Delta.$$

From (3.2.5) and by the symmetry of $\phi(v)$ we have

$$\begin{aligned} P\{\ell(\beta) < c_\alpha\} &= P\{(n^{\frac{1}{2}}R)^T(n^{\frac{1}{2}}R) < c_\alpha\} + O(n^{-\frac{3}{2}}) \\ &= \int_{\|v\|<c_\alpha^{\frac{1}{2}}} \pi(v)\phi(v)dv + O(n^{-\frac{3}{2}}) \\ &= P(\chi_p^2 < c_\alpha) + \tfrac{1}{2}n^{-1}\int_{\|v\|<c_\alpha^{\frac{1}{2}}} \{\textstyle\sum_{i=1}^p h_{ii}(v_i^2-1) + \sum_{i\neq j} h_{ij}v_iv_j\}\phi(v)dv \\ &\quad + O(n^{-\frac{3}{2}}) \\ &= \alpha - p^{-1}\sum_{i=1}^p h_{ii}\, c_\alpha\, g_p(c_\alpha)\, n^{-1} + O(n^{-\frac{3}{2}}). \end{aligned} \tag{3.5.4}$$

After some simple algebra we may show that

$$p^{-1}\sum_{i=1}^p h_{ii} = p^{-1}\left(\tfrac{1}{2}\overline{\alpha}^{jjmm} - \tfrac{1}{3}\overline{\alpha}^{jkm}\overline{\alpha}^{jkm}\right).$$

Thus from (3.5.4) we obtain

$$P\{\ell(\beta) < c_\alpha\} = \alpha - a\, c_\alpha\, g_p(c_\alpha)n^{-1} + O(n^{-\frac{3}{2}}).$$

It remains to check that the formal expansion (3.5.3) is valid. Since

$$\overline{U} = (A^1, \cdots, A^p, A^{11}, \cdots, A^{pp}, A^{111}, \cdots, A^{ppp})^T = n^{-1} \sum U_i,$$

where

$$U_i = \left[x_i^T V_n^{-\frac{1}{2}} \epsilon_i, (x_i \otimes x_i) B_1^T \{\epsilon_i^2 - E(\epsilon_i^2)\}, (x_i \otimes x_i \otimes x_i) B_2^T \{\epsilon_i^3 - E(\epsilon_i^3)\} \right]^T,$$

we see that $\overline{U}$ is the mean of independent but not identically distributed random vectors due to the presence of the fixed design points. However, from Theorem 1.3.3 an Edgeworth expansion for this case may be established. It may be shown that conditions in (3.2.9) implies the conditions of Theorem 1.3.3. In particular, (ii) and (iii) of condition (3.2.9) implies that $n^{-1} \sum_{j=1}^{n} E(\|U_j\|)^5$ is bounded away from infinity. Thus, we may establish the following Edgeworth expansion for the distribution of $\overline{U}$ under condition (3.2.9),

$$\sup_{B \in \mathcal{B}} |P(\overline{U} \in B) - \int_B \xi_{n5}(u) du| = O(n^{-\frac{3}{2}}), \tag{3.5.5}$$

for every class $\mathcal{B}$ of Borel sets satisfying (3.5.2). In (3.5.5),

$$\xi_{n5}(u) = \sum_{r=0}^{3} P_r(-\phi : \{\chi_{\nu n}\})(u),$$

$\{\chi_{\nu n}\}, 1 \le \nu \le 5$, are the first five cumulants of $\overline{U}$, $P_r(-\phi : \{\chi_{\nu n}\})(u)$ is the density of the finite signed measure with characteristic function $\tilde{P}_r(it : \{\chi_{\nu n}\}) \exp(-\frac{1}{2} t^T t)$, and $\tilde{P}_r$ is the Edgeworth-Cramér polynomial. From the expression for R we see that there exists a smooth function f_n such that $n^{\frac{1}{2}} R = f_n(\overline{U})$. Hence, using Theorem 1.3.4 we may show in our case that the Edgeworth expansion (3.5.5) may be transformed by sufficiently smooth function f_n, to yield a valid Edgeworth expansion (3.5.3) under condition (3.2.9). □

3.5.2 Proof of Theorem 3.3.1

Theorem 3.3.1.: *Assume condition (3.2.9). For any* $c_\alpha > 0$,

$$P\{\ell(\beta) < c_\alpha\left(1 + a\,n^{-1}\right)\} = \alpha + O(n^{-2}).$$

where $P\left(\chi_p^2 < c_\alpha\right) = \alpha$.

Proof: According to Theorem 3.2.1, under conditions in (3.2.9),

$$P\{\ell(\beta) < c_\alpha\left(1 + an^{-1}\right)\} = P\{\chi_p^2 < c_\alpha\left(1 + an^{-1}\right)\} - a\,c_\alpha\,g_p\{c_\alpha\left(1 + an^{-1}\right)\}\,n^{-1} + O(n^{-\frac{3}{2}}). \tag{3.5.6}$$

Note that $g_p(v)$ is the density of the χ_p^2 distribution,

$$P\{\chi_p^2 < c_\alpha\left(1 + an^{-1}\right)\} = P(\chi_p^2 < c_\alpha) + a\,c_\alpha\,g_p(c_\alpha)\,n^{-1} + O(n^{-2}), \tag{3.5.7}$$

and that

$$g_p\{c_\alpha\left(1 + an^{-1}\right)\} = g_p(c_\alpha) + O(n^{-1}). \tag{3.5.8}$$

Substituting (3.5.7) and (3.5.8) into (3.5.6) gives

$$P\{\ell(\beta) < c_\alpha\left(1 + an^{-1}\right)\} = P(\chi_p^2 < c_\alpha) + O(n^{-\frac{3}{2}}). \tag{3.5.9}$$

Moreover, by an argument based on the oddness and evenness of polynomials in the Edgeworth expansion (see for example Barndorff - Nielsen and Hall 1988), the $O(n^{-\frac{3}{2}})$ term in (3.5.9) is actually $O(n^{-2})$. □

3.5.3 Proof of Theorem 3.3.2

Theorem 3.3.2: *Assume condition (3.3.4). Then,*

$$\hat{a} = a + O_p(n^{-\frac{1}{2}}).$$

In views of (3.3.3), we see that Theorem 3.3.2 is an immediate result of the following Lemma 3.5.1

Lemma 3.5.1: *Assume condition (3.3.4). Then,*

$$\hat{\overline{\alpha}}^{j\,j\,m\,m} = \overline{\alpha}^{j\,j\,m\,m} + O_p(n^{-\frac{1}{2}}) \quad \text{and} \quad \hat{\overline{\alpha}}^{j\,k\,m} = \overline{\alpha}^{j\,k\,m} + O_p(n^{-\frac{1}{2}}).$$

To prove Lemma 3.5.1 we need the following Lemmas 3.5.2 and 3.5.3.

Lemma 3.5.2: *Assume (3.3.4). Then there exist positive constants q_3 and q_4 such that, uniformly for any i, j, $1 \le i, j \le n$,*

$$(i) \quad |V_{n\,j}^{-\frac{1}{2}} x_i^T| \le q_3.$$

$$(ii) \quad \|\,(n^{-1} X^T X)^{-1}\, x_i^T \,\| \le q_4.$$

Proof: (i) Since V_n is positive definite, it has the following orthogonal decomposition:

$$V_n = \sum_{l=1}^{p} v_{l\,n} \zeta_l \zeta_l^T,$$

where $v_{1\,n} \ge \cdots \ge v_{p\,n}$ and $\zeta_1, \cdots, \zeta_p$ are eigenvalues and corresponding orthogonal unity eigenvectors of V_n. This yields

$$V_n^{-\frac{1}{2}} = \sum v_{l\,n}^{-\frac{1}{2}} \zeta_l \zeta_l^T.$$

Hence,

$$V_{n\,j}^{-\frac{1}{2}} = \sum v_{l\,n}^{-\frac{1}{2}} \zeta_l^j \zeta_l^T.$$

Condition (3.3.4) ensures that there exist C_1,$C_2 > 0$ such that $C_1 < v_{p\,n} < v_{1\,n} < C_2$. uniformly in n. Thus, uniformly for any $1 \le i, j \le n$,

$$\begin{aligned} |V_{n\,j}^{-\frac{1}{2}T} x_i^T| &= |\sum v_{i\,n}^{-\frac{1}{2}} \zeta_l^j \zeta_l^T x_i^T| \\ &\le C_1^{-\frac{1}{2}} \|x_i\| \\ &\le C_1^{-\frac{1}{2}} q_2. \end{aligned}$$

The first part of the lemma is proved by choosing $q_3 = C_1^{-\frac{1}{2}} q_2$.

(ii) Since

$$n^{-1} X^T X = \frac{1}{n} \sum x_i^T x_i$$

then

$$\operatorname{tr}(n^{-1}X^TX) = \frac{1}{n}\sum x_i x_i^T.$$

By (3.3.4) we know that $q_1 \le |x_i x_i^T| \le q_2$ for $1 \le i \le n$. Thus,

$$q_1 < \operatorname{tr}(n^{-1}X^TX) < q_2.$$

This means that all eigenvalues of $n^{-1}X^TX$ are bounded between q_1 and q_2. Hence the eigenvalues of $(n^{-1}X^TX)^{-1}$ are bounded between q_2^{-1} and q_1^{-1}. Suppose $\alpha_1,\cdots,\alpha_p$ and $\xi_1,\cdots,\xi_p$ are the eigenvalues and corresponding unit orthogonal eigenvectors of $(n^{-1}X^TX)^{-1}$, that is

$$(n^{-1}X^TX)^{-1} = \sum \alpha_i \xi_i \xi_i^T$$

where $q_2^{-1} \le \alpha_i \le q_1^{-1}$ and $\|\xi_i\| = 1$ for $1 \le i \le n$. Thus,

$$\|(n^{-1}X^TX)^{-1}x_i^T\| = \|\sum \alpha_j \xi_i \xi_i^T x_i\| \le p\, q_1^{-1} q_2$$

The second part of the lemma is proved by allowing $q_4 = p\, q_1^{-1} q_2$.

Lemma 3.5.3: *Assume condition (3.3.4). Then*

$$\hat{V}_n = V_n + O_p(n^{-\frac{1}{2}}).$$

Proof: Put $S_n = n^{-1}\sum Z_i Z_i^T$. Our first step is to prove

$$S_n = V_n + O_p(n^{-\frac{1}{2}}). \tag{3.5.10}$$

Let $V_n^{(lk)}$ and $S_n^{(lk)}$ denote the l'th row, k'th column elements of $p \times p$ matrices V_n and S_n respectively. For any $1 \le l,\ k \le p$, using Chebyshev's inequality,

$$\begin{aligned}
P(n^{\frac{1}{2}}|S_n^{(lk)} - V_n^{(lk)}| > M_\epsilon) &= P[n^{\frac{1}{2}}|\sum\{Z_i^l Z_i^k - E(Z_i^l Z_i^k)\}| > M_\epsilon] \\
&\le M_\epsilon^{-2} n^{-1} E[\sum\{Z_i^l Z_i^k - E(Z_i^l Z_i^k)\}]^2 \\
&= M_\epsilon^{-2} n^{-1} \sum E\{Z_i^l Z_i^k - E(Z_i^l Z_i^k)\}^2 \\
&= M_\epsilon^{-2} n^{-1} \sum \{E\,(Z_i^l Z_i^k)^2 - E^2\,(Z_i^l Z_i^k)\} \\
&\le M_\epsilon^{-2} n^{-1} \sum E\|Z_i\|^4.
\end{aligned}$$

Notice that condition (3.3.4) implies that $n^{-1}\sum E\|Z_i\|^4$ is uniformly bounded from above. Thus for any $\epsilon > 0$, we may choose $M_\epsilon = (\epsilon^{-1}n^{-1}\sum E\|Z_i\|^4)^{\frac{1}{2}}$ such that

$$P(n^{\frac{1}{2}}|S_n^{(lk)} - V_n^{(lk)}| > M_\epsilon) < \epsilon.$$

So we have

$$S_n^{(lk)} = V_n^{(lk)} + O_p(n^{-\frac{1}{2}}).$$

Thus, (3.5.10) holds.

It can be shown that for each integer k there exist constants $D_{k\,0}$ such that

$$|\hat{\epsilon}_i^k - \epsilon_i^k| \le D_{k\,0}\,|x_i\,(\beta_{LS} - \beta)\,|\,\{|\epsilon_i|^{k-1} + |x_i\,(\beta_{LS} - \beta)\,|^{k-1}\}. \tag{3.5.11}$$

Using (3.5.11) and the boundness of the $\|x_i\|$'s,

$$\begin{aligned}
|\hat{V}_n^{(lk)} - S_n^{(lk)}| &= n^{-1}\sum x_i^l x_i^k\,(\hat{\epsilon}_i^2 - \epsilon_i^2) \\
&\le D_{2\,0}\|\beta_{LS} - \beta\|\,n^{-1}\sum |x_i^l x_i^k|\,\|x_i\|\,(|\epsilon_i| + \|x_i\|\,\|\beta_{LS} - \beta\|) \\
&\le D_1\|\beta_{LS} - \beta\|\,n^{-1}\sum |\epsilon_i| + D_2\|\beta_{LS} - \beta\|, \qquad (3.5.12)
\end{aligned}$$

where D_1 and D_2 are constants only related to the q_1, q_2 and $D_{2\,0}$. Condition (3.3.4) enables us to use Chebyshev's law of large number, which implies that

$$n^{-1}\sum |\epsilon_i| = O_p(1).$$

Since β_{LS} is a root-n consistent estimator of β, we readily obtain from (3.5.12) that

$$\hat{V}_n^{(lk)} = S_n^{(lk)} + O_p(n^{-\frac{1}{2}}).$$

This together with (3.5.10) enables us to show that

$$\hat{V}_n^{(lk)} = V_n^{(lk)} + O_p(n^{-\frac{1}{2}}).$$

Thus we have proved the lemma.

Now we are able to prove Lemma 3.5.1.

Proof of Lemma 3.5.1. We know from Lemma 3.5.2 that there exist positive constants q_3 and q_4 such that for any $1 \leq i, j \leq n$,

$$|V_{nj}^{-\frac{1}{2}} x_i^T| \leq q_3, \qquad \| (n^{-1} X^T X)^{-1} x_i^T \| \leq q_4. \tag{3.5.13}$$

Put

$$\overline{\alpha}_0^{j\,k\,m} = n^{-1} \sum_{i=1}^{n} \epsilon_i^3 V_{nj}^{-\frac{1}{2}} x_i^T V_{nk}^{-\frac{1}{2}} x_i^T V_{nm}^{-\frac{1}{2}} x_i^T \quad \text{and}$$

$$\overline{\alpha}_0^{j\,j\,m\,m} = n^{-1} \sum_{i=1}^{n} \epsilon_i^4 \left(x_i V_n^{-1} x_i^T\right)^2.$$

For any $M > 0$, using (3.5.13) and Chebyshev's inequality,

$$\begin{aligned}
&P\{n^{\frac{1}{2}}(\overline{\alpha}_0^{j\,j\,m\,m} - \overline{\alpha}^{j\,j\,m\,m}) > M\} \\
&= P\left[n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\epsilon_i^4 - E(\epsilon_i^4)\} V_{nj}^{-\frac{1}{2}} x_i^T V_{nk}^{-\frac{1}{2}} x_i^T V_{nm}^{-\frac{1}{2}} x_i^T > M\right] \\
&\leq M^{-2} n^{-1} E\left[\sum_{i=1}^{n} \{\epsilon_i^4 - E(\epsilon_i^4)\} V_{nj}^{-\frac{1}{2}} x_i^T V_{nk}^{-\frac{1}{2}} x_i^T V_{nm}^{-\frac{1}{2}} x_i^T\right]^2 \\
&\leq M^{-2} q_3^6 n^{-1} \sum E\{\epsilon_i^3 - E(\epsilon_i^3)\}^2.
\end{aligned}$$

From (3.3.4) we know that $n^{-1} \sum E\{\epsilon_i^3 - E(\epsilon_i^3)\}^2$ is uniformly bounded from above. Therefore for any $\epsilon > 0$, there exists a $M_\epsilon > 0$ such that for any $M > M_\epsilon$,

$$P\{n^{\frac{1}{2}}(\overline{\alpha}_0^{j\,j\,m\,m} - \overline{\alpha}^{j\,j\,m\,m}) > M\} < \epsilon$$

uniformly in n. Thus,

$$\overline{\alpha}_0^{j\,j\,m\,m} = \overline{\alpha}^{j\,j\,m\,m} + O_p(n^{-\frac{1}{2}}). \tag{3.5.14}$$

In a similar way we can prove that

$$\overline{\alpha}_0^{j\,k\,m} = \overline{\alpha}^{j\,k\,m} + O_p(n^{-\frac{1}{2}}). \tag{3.5.15}$$

To prove Lemma 3.5.1 it is sufficient to show that

$$\hat{\overline{\alpha}}^{j\,k\,m} = \overline{\alpha}_0^{j\,k\,m} + O_p(n^{-\frac{1}{2}}), \tag{3.5.16}$$

$$\hat{\overline{\alpha}}^{j\,j\,m\,m} = \overline{\alpha}_0^{j\,j\,m\,m} + O_p(n^{-\frac{1}{2}}). \tag{3.5.17}$$

We give only the proof of (3.5.17) here, since (3.5.16) may be handled similarly. By Lemma 3.5.3,

$$\hat{V}_n = V_n + O_p(n^{-\frac{1}{2}}),$$

which implies that

$$\hat{V}_n^{-1} = V_n^{-1} + O_p(n^{-\frac{1}{2}}) \quad \text{and} \quad \hat{V}_{nj}^{-1} = V_{nj}^{-1} + O_p(n^{-\frac{1}{2}})\ 1 \le j \le p. \tag{3.5.18}$$

By Taylor expansion, (3.5.18) and the Schwarz inequality, we have

$$(x_i \hat{V}_n^{-1} x_i^T)^2 - (x_i V_n^{-1} x_i^T)^2 = 2(x_i V_n^{-1} x_i^T)\{x_i(\hat{V}_n^{-1} - V_n^{-1})x_i^T\} + o_p(n^{-\frac{1}{2}}), \tag{3.5.19}$$

$$|x_i V_n^{-1} x_i^T| = |x_i V_n^{-\frac{1}{2}} V_n^{-\frac{1}{2}} x_i^T| \le \|x_i V_n^{-\frac{1}{2}}\|\,\|V_n^{-\frac{1}{2}} x_i^T\| \le p\, q_3^2. \tag{3.5.20}$$

Following (3.5.18) and (3.5.20), we may show that there exist positive constants C_3 and C_4 such that for $1 \le i \le n$,

$$|x_i \hat{V}_n^{-1} x_i^T| \le C_3 + |\Delta_1|, \tag{3.5.21}$$

$$|(x_i \hat{V}_n^{-1} x_i^T)^2 - (x_i V_n^{-1} x_i^T)^2| \le C_4\, \|\hat{V}_n^{-1} - V_n^{-1}\| + |\Delta_2|, \tag{3.5.22}$$

where $|\Delta_i| = O_p(n^{-\frac{1}{2}})$ for $i = 1, 2$ and $\|A\| = \max_{ij} |a_{ij}|$ for any matrix $A = (a_{ij})$.

From the fact that $\hat{\epsilon}_i = (\beta - \beta_{LS})x_i + \epsilon_i$, and using the Binomial Theorem, we may show that for each integer k there exists a constant D_k such that

$$|\hat{\epsilon}_i^k - \epsilon_i^k| \le D_k\, |x_i(\beta_{LS} - \beta)|\,\{|\epsilon_i|^{k-1} + |x_i(\beta_{LS} - \beta)|^{k-1}\}. \tag{3.5.23}$$

Now from (3.5.21) - (3.5.23),

$$\begin{aligned}
&|\hat{\alpha}^{jjmm} - \overline{\alpha}_0^{jjmm}| \\
&= n^{-1}\left|\sum \left[(\hat{\epsilon}_i^4 - \epsilon_i^4)\left(x_i \hat{V}_n^{-1} x_i^T\right)^2 + \epsilon_i^4\{(x_i \hat{V}_n^{-1} x_i^T)^2 - (x_i V_n^{-1} x_i^T)^2\}\right]\right| \\
&\le D_4 \|\beta_{LS} - \beta\|\, n^{-1} \sum \|x_i\|\,\{|\epsilon_i|^3 + \|x_i\|^3 \|\beta_{LS} - \beta\|^3\}\left(x_i \hat{V}_n^{-1} x_i^T\right)^2 \qquad (3.5.24) \\
&\quad + \left(C_4 \|\hat{V}_n^{-1} - V_n^{-1}\| + |\Delta_2|\right) n^{-1} \sum \epsilon_i^4 \\
&\le \left(q_2\, D_4\, \|\beta_{LS} - \beta\|\, n^{-1} \sum |\epsilon_i|^3 + q_2^4\, C_3\, D_4\, \|\beta_{LS} - \beta\|^4\right)(C_3 + |\Delta_1|) \\
&\quad + \left(C_4\, \|\hat{V}_n^{-1} - V_n^{-1}\| + |\Delta_2|\right) n^{-1} \sum \epsilon_i^4.
\end{aligned}$$

Since (3.3.4) implies that both $n^{-1}\sum|\epsilon_i|^3$ and $n^{-1}\sum\epsilon_i^4$ are uniformly bounded from above, (3.5.17) can be proved from (3.5.18), (3.5.24) and the fact that $\beta_{LS} = \beta + O_p(n^{-\frac{1}{2}})$. $\square$

Appendix 3: Calculations of the Cumulants of $n^{\frac{1}{2}} R$

In this appendix we detail the calculations of cumulants of $n^{\frac{1}{2}} R$ as shown in (3.5.1), which have been used to derive Theorem 3.2.1. Here we generalize the technique used by DiCiccio, Hall and Romano (1988) for *i.i.d.* case to independent but not identically distributed situation.

We first need some basic formulae for calculationg means of moments of independent but not identically distributed samples.

Let $X_1, \cdots, X_n$ be independent but not identically distributed random variables, and $h^1, h^2, \cdots$ be real-valued functions such that $E\{h^j(X_i)\} = 0$, $j = 1, 2, \cdots$, $i = 1, 2, \cdots, n$. Let $H_i^j = h^j(X_i)$ and $C^j = n^{-1} \sum h^j(X_i)$. Then

$$
\begin{aligned}
E(C^j C^k) &= n^{-2} \sum E(H_i^j H_i^k), \\
E(C^j C^k C^l) &= n^{-3} \sum E(H_i^j H_i^k H_i^l), \\
E(C^j C^k C^l C^m) &= n^{-4} \sum_{i_1 i_2} E(H_{i_1}^j H_{i_1}^k) E(H_{i_2}^l H_{i_2}^m)[3] + O(n^{-3}), \qquad (3.A.1) \\
E(C^j C^k C^l C^m C^n) &= n^{-5} \sum_{i_1 i_2} E(H_{i_1}^j H_{i_1}^k) E(H_{i_2}^l H_{i_2}^m H_{i_2}^n)[10] + O(n^{-6}), \\
E(C^j C^k C^l C^m C^n C^o) &= n^{-5} \sum_{i_1 i_2 i3} E(H_{i_1}^j H_{i_1}^k) E(H_{i_2}^l H_{i_2}^m) E(H_3^n H_3^o)[15] + O(n^{-6}),
\end{aligned}
$$

where

$$
\begin{aligned}
E(H_{i_1}^j H_{i_1}^k) E(H_{i_2}^l H_{i_2}^m)[3] = {} & E(H_{i_1}^j H_{i_1}^k) E(H_{i_2}^l H_{i_2}^m) + E(H_{i_1}^j H_{i_1}^l) E(H_{i_2}^k H_{i_2}^m) \\
& + E(H_{i_1}^j H_{i_1}^m) E(H_{i_2}^l H_{i_2}^k),
\end{aligned}
$$

and a similar rule applies for

$$
E(H_{i_1}^j H_{i_1}^k) E(H_{i_2}^l H_{i_2}^m H_{i_2}^n)[10] \text{ and } E(H_{i_1}^j H_{i_1}^k) E(H_{i_2}^l H_{i_2}^m) E(H_3^n H_3^o)[15].
$$

We shall use the above formulae very intensively to calculate the joint cumulants of $n^{\frac{1}{2}} R$, denoted by $k_j, j = 1, 2, 3, \ldots$.

According to (3.2.6),

$$
R = R_1 + R_2 + R_3
$$

where

$$R_1^j = A^j,$$

$$R_2^j = -\tfrac{1}{2}A^{j\,k}A^k + \tfrac{1}{3}\overline{\alpha}^{j\,k\,m}A^kA^m,$$

$$R_3^j = \tfrac{3}{8}A^{j\,m}A^{k\,m}A^k + \tfrac{1}{3}A^{j\,k\,m}A^kA^l - \tfrac{5}{12}\overline{\alpha}^{j\,k\,m}A^{l\,m}A^kA^l$$

$$-\tfrac{5}{12}\overline{\alpha}^{k\,l\,m}A^{j\,m}A^kA^l + \tfrac{4}{9}\overline{\alpha}^{j\,k\,n}\overline{\alpha}^{l\,m\,n}A^mA^kA^l - \tfrac{1}{4}\overline{\alpha}^{j\,k\,l\,m}A^mA^kA^l,$$

The joint first-order cumulants of R are given by

$$\mathrm{cum}(R^j) = E(R_1^j) + E(R_2^j) + E(R_3^j).$$

From the definition of R_1, R_2 and R_3 in (3.2.6), and using the formulae given in (3.A.1), we see that

$$E(R_1^j) = 0,$$

$$E(R_2^j) = -\tfrac{1}{2}E\left(A^{j\,k}A^k\right) + \tfrac{1}{3}\overline{\alpha}^{j\,k\,m}E\left(A^kA^m\right)$$

$$= -\tfrac{1}{2}n^{-1}\overline{\alpha}^{j\,k\,k} + \tfrac{1}{3}n^{-1}\overline{\alpha}^{j\,k\,m}\overline{\alpha}^{k\,m}$$

$$= -\tfrac{1}{6}n^{-1}\overline{\alpha}^{j\,k\,k},$$

$$E(R_3^j) = O(n^{-2}).$$

Thus,

$$\mathrm{cum}(R^j) = -\tfrac{1}{6}n^{-1}\overline{\alpha}^{j\,k\,k} + O(n^{-2}).$$

Consequently, by putting $\mu^j = -\frac{1}{6}n^{-1}\overline{\alpha}^{j\,k\,k}$ we have

$$k_1 = n^{\frac{1}{2}}\mathrm{cum}(R) = n^{-\frac{1}{2}}\mu + O(n^{-\frac{3}{2}}). \qquad (3.A.2)$$

According to definition, the joint second-order cumulants of R are defined to be

$$\mathrm{cum}(R^j, R^k) = E(R^jR^k) - E(R^j)E(R^k)$$

$$= E(R_1^jR_1^k) + E(R_1^jR_2^k)[2] + E(R_1^jR_3^k)[2] + E(R_2^jR_2^k)$$

$$+ E(R_2^jR_3^k)[2] + E(R_3^jR_3^k) - \tfrac{1}{36}\overline{\alpha}^{j\,m\,m}\overline{\alpha}^{j\,k\,k} + O(n^{-3}). \qquad (3.A.3)$$

It follows from the formulae in (3.A.1) that

$$E(R_1^jR_1^k) = E(A^jA^k) = n^{-1}\delta^{j\,k}, \qquad (3.A.4)$$

$$E(R_1^j R_2^k) = -\tfrac{1}{2}E\left(A^{k\,m} A^j A^m\right) + \tfrac{1}{3}\overline{\alpha}^{k\,m\,l} E\left(A^j A^m A^l\right)$$

$$= -\tfrac{1}{2}n^{-2}\left(\overline{\alpha}^{j\,k\,m\,m} - \delta^{j\,k}\right) + \tfrac{1}{3}n^{-2}\overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l}, \qquad (3.A.5)$$

$$E(R_1^j R_3^k) = \tfrac{3}{8}E\left(A^{k\,m} A^{l\,m} A^j A^l\right) + \tfrac{1}{3}E\left(A^{k\,m\,l} A^j A^m A^l\right) - \tfrac{10}{12}\overline{\alpha}^{k\,m\,l} E\left(A^{l\,n} A^j A^m A^n\right)$$

$$+ \left(\tfrac{4}{9}\overline{\alpha}^{k\,m\,o}\overline{\alpha}^{n\,l\,o} - \tfrac{1}{4}\overline{\alpha}^{k\,m\,l\,n}\right) E\left(A^j A^m A^l A^n\right). \qquad (3.A.6)$$

To compute $E(R_1^j R_3^k)$ we use the basic formulae in (3.A.1) again, obtaining that

$$E\left(A^{k\,m} A^{l\,m} A^j A^l\right) = n^{-2}\left(\overline{\alpha}^{j\,k\,m\,m} - \delta^{j\,k} + \overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l} + \overline{\alpha}^{j\,k\,m}\overline{\alpha}^{m\,l\,l}\right) + O(n^{-3}),$$

$$E\left(A^{k\,m\,l} A^j A^m A^l\right) = 3n^{-2}\overline{\alpha}^{j\,k\,m\,m} + O(n^{-3}),$$

$$E\left(A^{l\,n} A^j A^m A^n\right) = 2n^{-2}\left(\overline{\alpha}^{j\,m\,l} + \overline{\alpha}^{m\,n\,n}\delta^{j\,l}\right) + O(n^{-3}),$$

$$E\left(A^j A^m A^l A^n\right) = n^{-2}\left(\delta^{j\,m}\delta^{l\,n} + \delta^{j\,l}\delta^{m\,n} + \delta^{j\,n}\delta^{m\,l}\right) + O(n^{-3}).$$

Substituting the above equations back into the expression for $E(R_1^j R_3^k)$, we have

$$E(R_1^j R_3^k) = n^{-2}\left(\tfrac{5}{8}\overline{\alpha}^{j\,k\,m\,m} - \tfrac{3}{8}\delta^{j\,m}\delta^{k\,l}\delta^{m\,l} - \tfrac{29}{72}\overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l} - \tfrac{1}{72}\overline{\alpha}^{j\,k\,m}\overline{\alpha}^{m\,l\,l}\right)$$

$$+ O(n^{-3}). \qquad (3.A.7)$$

Similarly, since

$$E\left(A^{j\,m} A^{k\,l} A^m A^l\right) = n^{-2}\left(\overline{\alpha}^{j\,m\,m}\overline{\alpha}^{k\,l\,l} + \overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l} + \overline{\alpha}^{j\,k\,m\,m}\right) + O(n^{-3}),$$

$$E\left(A^{k\,n} A^n A^m A^l\right) = n^{-2}\left(\overline{\alpha}^{k\,n\,n}\delta^{m\,l} + \overline{\alpha}^{k\,n\,m}\delta^{l\,n} + \overline{\alpha}^{k\,n\,l}\delta^{m\,n}\right) + O(n^{-3}),$$

$$E\left(A^m A^l A^n A^o\right) = n^{-2}\left(\delta^{m\,l}\delta^{n\,o} + \delta^{m\,n}\delta^{l\,o} + \delta^{m\,o}\delta^{l\,n}\right) + O(n^{-3}),$$

then

$$E(R_2^j R_2^k) = \tfrac{1}{4}E\left(A^{j\,m} A^{k\,l} A^m A^l\right) - \tfrac{1}{6}\overline{\alpha}^{j\,m\,l} E\left(A^{k\,n} A^n A^m A^l\right)$$

$$- \tfrac{1}{6}\overline{\alpha}^{k\,m\,l} E\left(A^{j\,n} A^n A^m A^l\right) + \tfrac{1}{9}\overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,n\,o} E\left(A^m A^l A^n A^o\right)$$

$$= n^{-2}\left(\tfrac{1}{4}\overline{\alpha}^{j\,k\,m\,m} + \tfrac{1}{36}\overline{\alpha}^{j\,m\,m}\overline{\alpha}^{k\,l\,l} - \tfrac{7}{36}\overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l} - \tfrac{1}{4}\delta^{j\,m}\delta^{k\,l}\delta^{m\,l}\right)$$

$$+ O(n^{-3}). \qquad (3.A.8)$$

Substituting (3.A.4) - (3.A.8) into (3.A.3) we derive

$$\mathrm{cum}(R^j, R^k) = n^{-1}\delta^{j\,k} + n^{-2}\left(\tfrac{1}{2}\overline{\alpha}^{j\,k\,m\,m} - \tfrac{1}{3}\overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l} - \tfrac{1}{36}\overline{\alpha}^{j\,k\,m}\overline{\alpha}^{m\,l\,l} + \tfrac{1}{36}\overline{\alpha}^{j\,m\,m}\overline{\alpha}^{k\,l\,l}\right)$$

$$- \tfrac{1}{36}n^{-2}\overline{\alpha}^{j\,m\,m}\overline{\alpha}^{k\,l\,l} + O(n^{-3}).$$

$$= n^{-1}\delta^{j\,k} + n^{-2}\left(\tfrac{1}{2}\overline{\alpha}^{j\,k\,m\,m} - \tfrac{1}{3}\overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l} - \tfrac{1}{36}\overline{\alpha}^{j\,k\,m}\overline{\alpha}^{m\,l\,l}\right) + O(n^{-3}).$$

Put $\Delta = (\Delta_{j\,k})_{p \times p}$ where

$$\Delta_{j\,k} = \frac{1}{2}\overline{\alpha}^{j\,k\,m\,m} - \tfrac{1}{3}\overline{\alpha}^{j\,m\,l}\overline{\alpha}^{k\,m\,l} - \tfrac{1}{36}\overline{\alpha}^{j\,k\,m}\,\overline{\alpha}^{m\,l\,l},$$

and let I_p denote the $p \times p$ identity matrix. Thus we derive the second order cumulant of $n^{\frac{1}{2}}R$ to be

$$k_2 = I_p + n^{-1}\Delta + O(n^{-2}).$$

To compute the third-order joint cumulants of R we notice that

$$\begin{aligned} cum(R^j, R^k, R^h) &= E(R^j R^k R^h) - E(R^j)E(R^k R^h)[3] + 2E(R^j)E(R^k)E(R^h) \\ &= E(R_1^j R_1^k R_1^h) + E(R_2^j R_1^k R_1^h)[3] - E(R_2^j)E(R_1^k R_1^h)[3] \\ &\quad + O(n^{-3}). \end{aligned} \tag{3.A.9}$$

Again it follows from the formulae in (3.A.1) that

$$\begin{aligned} E(R_1^j R_1^k) &= n^{-1}\delta^{j\,k}, \\ E(R_1^j R_1^k R_1^h) &= n^{-2}\overline{\alpha}^{j\,k\,h}, \\ E(R_2^j) &= -\tfrac{1}{6}n^{-1}\overline{\alpha}^{j\,m\,m}, \\ E(R_2^j R_1^k R_1^h) &= n^{-2}\left(-\tfrac{1}{6}\overline{\alpha}^{j\,m\,m}\delta^{k\,h} - \tfrac{1}{3}\overline{\alpha}^{j\,k\,h}\right) + O(n^{-3}). \end{aligned}$$

Therefore,

$$E(R_2^j R_1^k R_1^h) = E(R_2^j)E(R_1^k R_1^h) - \tfrac{1}{3}E(R_1^j R_1^k R_1^h) + O(n^{-3}). \tag{3.A.10}$$

Hence from (3.A.9), we obtain

$$\mathrm{cum}(R^j, R^k, R^h) = O(n^{-3}).$$

Consequently we have

$$k_3 = O(n^{-\frac{3}{2}}).$$

At last we calculate the joint fourth-order cumulants of $n^{\frac{1}{2}}R$. By definition,

$$cum(R^j, R^k, R^h, R^m) = E(R^j\, R^k\, R^h\, R^m) - E(R^j\, R^k)\, E(R^h\, R^m)[3]$$

$$
\begin{aligned}
&- E(R^j)\,E(R^k\,R^h\,R^m)[4] + 2\,E(R^j)\,E(R^k)\,E(R^h\,R^m) \\
&- 6\,E(R^j)\,E(R^k)\,E(R^h)\,E(R^m) \\
&= E(R^j_1\,R^k_1\,R^h_1\,R^m_1) + E(R^j_2\,R^k_1\,R^h_1\,R^m_1)[4] + E(R^j_3\,R^k_1\,R^h_1\,R^m_1)[4] \\
&+ E(R^j_2\,R^k_2\,R^h_1\,R^m_1)[6] - E(R^j_1\,R^k_1)\,E(R^h_1\,R^m_1)[3] \\
&- E(R^j_2\,R^k_1)\,E(R^h_1\,R^m_1)[12] - E(R^j_3\,R^k_1)\,E(R^h_1\,R^m_1)[12] \\
&- E(R^j_2\,R^k_2)\,E(R^h_1\,R^m_1)[6] - E(R^j_2)\,E(R^k_1\,R^h_1\,R^m_1)[4] \\
&- E(R^j_2)\,E(R^k_2\,R^h_1\,R^m_1)[12] + 2\,E(R^j_2)\,E(R^k_2)\,E(R^h_1\,R^m_1)[6] \\
&+ O(n^{-4}). \qquad (3.A.11)
\end{aligned}
$$

Put

$$t_1 = \overline{\alpha}^{j\,k\,h\,m}, \quad t_2 = \delta^{j\,k}\,\delta^{h\,m},$$

$$t_3 = \overline{\alpha}^{j\,k\,h}\,\overline{\alpha}^{m\,n\,n} + \overline{\alpha}^{j\,k\,m}\,\overline{\alpha}^{h\,n\,n} + \overline{\alpha}^{j\,h\,m}\,\overline{\alpha}^{k\,n\,n} + \overline{\alpha}^{k\,h\,m}\,\overline{\alpha}^{u\,n\,n},$$

$$t_4 = \overline{\alpha}^{j\,k\,n}\,\overline{\alpha}^{h\,m\,n} + \overline{\alpha}^{j\,h\,n}\,\overline{\alpha}^{k\,m\,n} + \overline{\alpha}^{j\,m\,n}\,\overline{\alpha}^{k\,h\,n}.$$

It may be shown that

$$E(R^j_1\,R^k_1\,R^h_1\,R^m_1) - E(R^j_1\,R^k_1)\,E(R^h_1\,R^m_1)[3] = n^{-3}\,(t_1 - t_2),$$

$$E(R^j_2\,R^k_1\,R^h_1\,R^m_1)[4] - E(R^j_2\,R^k_1)\,E(R^h_1\,R^m_1)[12] = n^{-3}\,(-6\,t_1 + 2\,t_2 - \tfrac{1}{6}\,t_3 + \tfrac{2}{3}\,t_4),$$

$$E(R^j_2\,R^k_2\,R^h_1\,R^m_1)[6] - E(R^j_2\,R^k_2)\,E(R^h_1\,R^m_1)[6] = n^{-3}\,(3\,t_1 - t_2 + \tfrac{1}{6}\,t_3 - \tfrac{5}{9}\,t_4),$$

$$E(R^j_3\,R^k_1\,R^h_1\,R^m_1)[4] - E(R^j_3\,R^k_1)\,E(R^h_1\,R^m_1)[12] = n^{-3}\,(2\,t_1 - \tfrac{1}{9}\,t_4).$$

Also from (3.A.10),

$$E(R^j_2)\,\{E(R^k_1\,R^h_1\,R^m_1)[4] - E(R^k_2\,R^h_1\,R^m_1)[12] + 2\,E(R^k_2)\,E(R^h_1\,R^m_1)[6]\} = O(n^{-4}).$$

Hence, substituting the above formulae into (3.A.11), we get

$$\mathrm{cum}(R^j, R^k, R^h, R^m) = O(n^{-4}),$$

which means that

$$k_4 = O(n^{-\frac{3}{2}}).$$

Based on the general results given by James and Mayne (1962), we have for $j \geq 5$, that

$$k_j = O(n^{-(j-2)/2}).$$

CHAPTER 4
EMPIRICAL LIKELIHOOD CONFIDENCE INTERVALS FOR LINEAR REGRESSION COEFFICIENTS

4.1 Introduction

In Chapter 3 we considered constructing confidence region for the linear regression coefficient vector β. We showed that the coverage errors of the empirical likelihood confidence regions proposed by Owen (1991) are of order n^{-1} and that they can be reduced to order n^{-2} by employing Bartlett correction. However, it is not enough to just construct confidence regions for β. In practice, statisticians are often confronted with problems of constructing confidence intervals for a particular regression coefficient or for certain linear combinations of β.

In this chapter we address the above problem under the simple linear regression model. A simple linear regression model is

$$y_i = a_o + b_o x_i + \epsilon_i, \qquad 1 \leq i \leq n, \tag{4.1.1}$$

where all the variables appearing in (4.1.1) are scalars. Among them, x_i and y_i are the i'th fixed design point and response respectively, the ϵ_i 's are independent and identically distributed random errors with mean zero and variance σ^2, and a_o and b_o are the unknown intercept and slope parameters respectively.

This chapter has two aims. We first show how to construct empirical likelihood confidence intervals for the slope parameter b_o and means $y_o = a_o + b_o x_o$, for any fixed x_o under model (4.1.1). Obviously the later case includes the intercept parameter a_o when choosing $x_o = 0$. Then, we study the coverage accuracy and Bartlett correctability of empirical likelihood confidence intervals for these parameters.

We introduce some basic notation and formulae in Section 4.2. In Sections 4.3 and 4.4 we propose procedures for constructing confidence intervals for b_o and y_o, and show that both empirical likelihood confidence intervals have coverage errors of order n^{-1}. Furthermore, we demonstrate that Bartlett correction can be used to reduce the coverage error from order n^{-1} to order n^{-2}. A simulation study is presented in Section 4.5. All the proofs are deferred to Section 4.6. The calculations of cumulants are presented in Appendix 4.

4.2 Preliminaries

In this section we introduce some notation and basic formulae which will be used throughout this paper. We denote by $\hat{a}_o$ and $\hat{b}_o$ the least squares estimates of a_o and b_o respectively, and use μ_j for the j'th moment of ϵ_1 for $j = 1, 2$, and $\overline{x}$ and $\overline{y}$ for the means of x_i's and y_i's respectively. We define auxilliary variables

$$z_i(a,b) = (1,\, x_i)^T(y_i - a - bx_i) \quad 1 \le i \le n$$

where a and b are any candidate values for a_o and b_o respectively. Specifically we write z_i as $z_i(a_o, b_o)$. Furthermore, we put

$$\sigma_x^2 = n^{-1}\sum (x_i - \overline{x})^2, \quad m_j = n^{-1}\sum (x_i - \overline{x})^j, \quad j = 3,4$$

$$\hat{\sigma}^2 = n^{-1}\sum \hat{\epsilon}_i^2, \quad \hat{\mu}_j = n^{-1}\sum \hat{\epsilon}_i^j, \quad j = 3,4,$$

$$\overline{\epsilon} = \overline{y} - a_0 - b_o\overline{x}, \quad \text{where} \quad \hat{\epsilon}_i = y_i - \hat{a} - \hat{b}x_i.$$

Let

$$V_n = \sigma^2 \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & n^{-1}\sum x_i^2 \end{pmatrix}$$

be the average covariance matrix of auxilliary variables z_i's, let v_{1n} and v_{2n} be the largest and smallest eigenvalues of V_n respectively, and let

$$U_n = \begin{pmatrix} u_1^1 & u_1^2 \\ u_2^1 & u_2^2 \end{pmatrix} = V_n^{-\frac{1}{2}}$$

be the inverse of the square root matrix of V_n. Define

$$g_{j_1 j_2 \cdots j_k}(x_i) = \prod_{l=1}^{k} (u_{j_l}^1 + u_{j_l}^2 x_i),$$

$$\overline{\alpha}^{j_1 j_2 \cdots j_k} = n^{-1} \sum E\{g_{j_1 j_2 \cdots j_k}(x_i) \epsilon_i^k\},$$

$$A^{j_1 j_2 \cdots j_k}(a,b) = n^{-1} \sum g_{j_1 j_2 \cdots j_k}(x_i)(y_i - a - bx_i)^k - \overline{\alpha}^{j_1 j_2 \cdots j_k}.$$

For simplicity of notation, write

$$A_o^{j_1, j_2, \cdots, j_k} = A^{j_1, j_2 \cdots, j_k}(a_o, b_o), \quad \text{and} \quad A^{j_1, j_2, \cdots, j_k} = A^{j_1, j_2, \cdots, j_k}(a,b).$$

We assume the following regularity conditions.

There exist positive constants C_1 and C_2 such that uniformly in n,

$$C_1 < v_{pn} \le v_{1n} < C_2; \text{ and } n^{-2} \sum_{j=1}^{n} E\|z_j\|^4 \to 0, \tag{4.2.1}$$

where $\| \, \|$ is the Euclidean norm; and for candidate values a and b of a_o and b_o,

$$a = a_o + O_p(n^{-\frac{1}{2}}) \quad \text{and} \quad b = b_o + O_p(n^{-\frac{1}{2}}). \tag{4.2.2}$$

Let $\ell(a,b)$ be the empirical log-likelihood ratio evaluated at (a,b). Write $p_1, \cdots, p_n$ for nonnegative numbers adding to unity. Then, according to the definition of empirical likelihood,

$$\ell(a,b) = -2 \min_{\sum p_i z_i(a,b) = 0} \sum_{i=1}^{n} \log(np_i).$$

Using the Lagrange multiplier method gives us

$$\ell(a,b) = 2 \sum \log\{1 + \lambda (1, x_i)^T (y_i - a - bx_i)\},$$

and $\lambda = (\lambda_1, \lambda_2)$ satisfies

$$\sum \frac{(1, x_i)^T (y_i - a - bx_i)}{1 + \lambda (1, x_i)^T (y_i - a - bx_i)} = 0.$$

Since the analytic solutions for both λ and $\ell(a,b)$ are not obtainable, we have to resort to expansions. Using (3.2.4) of Chapter 3, under conditions (4.2.1) and

(4.2.2), we have the following Taylor expansion for $\ell(a,b)$:

$$\begin{aligned} n^{-1}\ell(a,\, b) = {} & A^j A^j - A^{j\,k} A^j A^k + \tfrac{2}{3}\overline{\alpha}^{j\,k\,l} A^j A^k A^l + A^{j\,l} A^{k\,l} A^j A^k \\ & + \tfrac{2}{3} A^{j\,k\,l} A^j A^k A^l - 2\,\overline{\alpha}^{j\,k\,m} A^{l\,m} A^j A^k A^l \\ & + (\overline{\alpha}^{j\,k\,n}\overline{\alpha}^{l\,m\,n} - \tfrac{1}{2}\overline{\alpha}^{j\,k\,l\,m}) A^j A^k A^l A^m + O_p(n^{-\frac{5}{2}}). \end{aligned} \qquad (4.2.3)$$

Here we use the summation convention according to which, if an index occurs more that once in an expression, summation over that index is understood.

4.3 Empirical Likelihood Confidence Interval for Slope Parameter

In this section we show how to construct empirical likelihood confidence intervals for the slope parameter b_o, and analyse the coverage properties of these confidence intervals. We first prove a nonparametric version of Wilks' theorem for the empirical log-likelihood ratio for b_o (Theorem 4.3.1). Then we develop an Edgeworth expansion of the distribution of the empirical log-likelihood ratio for b_o (Theorem 4.3.2), which is used to show that the coverage errors of the confidence intervals are of order n^{-1}. Furthermore we demonstrate that the empirical likelihood confidence intervals are Bartlett correctable (Theorem 4.3.3). This means that simple scale adjustments can reduce the coverage errors from $O(n^{-1})$ to $O(n^{-2})$.

The empirical log-likelihood ratio for b_o may be obtained by minimizing $\ell(a,b_o)$ respect to a, which is treated as a nuisance parameter in this section, since we are interested only in constructing confidence intervals for b_o. Let $\tilde{a}$ be the optimal a which minimizes $\ell(a,b_o)$. Then

$$\ell(b_o) = \ell(\tilde{a},b_o) = \min_a \ell(a,b_o),$$

where

$$\ell(a,b_o) = 2\sum \log\{1 + \lambda\,(1\,,x_i)^T\,(y_i - a - b_o x_i)\},$$

and $\lambda = (\lambda_1,\, \lambda_2)$ satisfies

$$\sum \frac{(1\,,x_i)^T\,(y_i - a - b_o x_i)}{1 + \lambda\,(1\,,x_i)^T\,(y_i - a - b_o x_i)} = 0.$$

4.3.1 Wilks' theorem

We first give an expansion for $\tilde{a}$. From (4.2.3) and using the notation given in Section 4.2, we know that

$$\begin{aligned} n^{-1}\ell(a,\, b_o) =& A^j(a,b_o)A^j(a,b_o) - A^{jk}(a,b_o)A^j(a,b_o)A^k(a,b_o) \\ &+ \{\tfrac{2}{3}\overline{\alpha}^{jkl}A^l(a,b_o) + A^{jl}(a,b_o)A^{kl}(a,b_o)\}\, A^j(a,b_o)A^k(a,b_o) \\ &+ \{\tfrac{2}{3}A^{jkl}(a,b_o) - 2\overline{\alpha}^{jkml}A^{lm}(a,b_o)\}\; A^j(a,b_o)A^k(a,b_o)A^l(a,b_o) \\ &+ (\overline{\alpha}^{jkn}\overline{\alpha}^{lmn} - \tfrac{1}{2}\overline{\alpha}^{jklm})A^j(a,b_o)A^k(a,b_o)A^l(a,b_o)A^m(a,b_o) \\ &+ O_p(n^{-\frac{5}{2}}). \end{aligned} \tag{4.3.1}$$

Consider an expansion of

$$\tilde{a} = \hat{a} + a_1 + a_2 + a_3,$$

where $a_j = O_p(n^{-j/2})$, $j = 1,2,3$. We will determine a_1, a_2, a_3 successively. Put

$$\gamma_j = n^{-1}\sum g_j(x_i), \quad \gamma_{jk} = n^{-1}\sum g_{jk}(x_i),$$
$$\gamma_{jkl} = n^{-1}\sum g_{jkl}(x_i),$$
$$\gamma_{jk,1}(a,b) = n^{-1}\sum g_{jk}(x_i)(y_i - a - bx_i),$$
$$\gamma_{jkl,2}(a,b) = n^{-1}\sum g_{jkl}(x_i)(y_i - a - bx_i)^2.$$

Under (4.2.2),

$$\begin{aligned} n^{-1}\ell(a,\, b_o) &= A^j(a,b_o)A^j(a,b_o) + O_p(n^{-\frac{3}{2}}), \\ &= \{A^j(\hat{a},b_o) - \gamma_j(a-\hat{a})\}\{A^j(\hat{a},b_o) - \gamma_j(a-\hat{a})\} + O_p(n^{-\frac{3}{2}}). \end{aligned}$$

Solving for a_1 requires minimizing $\gamma_j\gamma_j a_1^2 - 2A^j(\hat{a},b_o)\gamma_j a_1$. Hence we have

$$a_1 = A^j(\hat{a},b_o)\gamma_j/\gamma_j\gamma_j. \tag{4.3.2}$$

By the definition of $A^j(\hat{a},b_o)$ and γ_j, a_1 has another form,

$$a_1 = \overline{x}(\hat{b}_o - b_o). \tag{4.3.3}$$

To find a_2, note that $\gamma_{jk,1} = O_p(n^{-\frac{1}{2}})$ under (4.2.2),

$$n^{-1}\ell(a,\, b_o) = 2\,\gamma_j\,\{A^k(\hat{a},b_o) - \gamma_k a_1\}\,\{A^{jk}(\hat{a},b_o) - \overline{\alpha}^{jkl}(A^l(\hat{a},b_o) - \gamma_l a_1)\}a_2$$
$$+\,\gamma_j\gamma_j a_2^2 + R_1 + O_p(n^{-\frac{5}{2}}),$$

where R_1 denotes term of not involving a_2. By minimization,

$$a_2 = -(\gamma_i\gamma_i)^{-1}\gamma_j\{A^k(\hat{a},b_o) - \gamma_k a_1\}\{A^{jk}(\hat{a},b_o) - \overline{\alpha}^{jkl}(A^l(\hat{a},b_o) - \gamma_l a_1)\}.$$

After some algebra we may show that

$$n^{-1}\ell(a,\, b_o) = -2\,\{A^j(\hat{a},b_o)\gamma_j - \gamma_j\gamma_j a_1\}a_3 + R_2 + O_p(n^{-\frac{5}{2}}),$$

where R_2 denotes a term not involving a_3. Thus, (4.3.2) implies that a_3 will not appear in the $O_p(n^{-2})$ or larger terms in the expansion of $n^{-1}\ell(b_o)$, so we need not to calculate a_3 any more. In summary we have

$$\tilde{a} = \hat{a} + \overline{x}(\hat{b} - b_o)$$
$$-\,(\gamma_i\gamma_i)^{-1}\gamma_j\,\{A^k(\hat{a},b_o) - \gamma_k a_1\}[A^{jk}(\hat{a},b_o) - \overline{\alpha}^{jkl}\{A^l(\hat{a},b_o) - \gamma_l\, a_1)\}].$$

The above formula suggests using $\hat{a}+\overline{x}(\hat{b}-b_o)$ as an initial value for a in numerically searching for $\tilde{a}$. In the author's experience, this works well. Now, substituting $\tilde{a}$ into (4.3.1), the empirical likelihood ratio statistic at b_o is given by

$$\begin{aligned}
&n^{-1}\ell(b_o)\\
&= \{A^j(\hat{a},b_o) - \gamma_j a_1\}\{A^j(\hat{a},b_o) - \gamma_j a_1\} - \gamma_j\gamma_j a_2^2\\
&\quad - \{A^{jk}(\hat{a},b_o) - 2\gamma_{jk,1}(\hat{a},b_o)a_1 + \gamma_{jk}a_1^2\}\{A^j(\hat{a},b_o) - \gamma_j a_1\}\,\{A^k(\hat{a},b_o) - \gamma_k a_1\}\\
&\quad + \tfrac{2}{3}\,\overline{\alpha}^{jkl}\{A^j(\hat{a},b_o) - \gamma_j a_1\}\{A^k(\hat{a},b_o) - \gamma_k a_1\}\{A^l(\hat{a},b_o) - \gamma_l a_1\}\\
&\quad + A^{jl}(\hat{a},b_o)A^{kl}(\hat{a},b_o)\{A^j(\hat{a},b_o) - \gamma_j a_1\}\{A^k(\hat{a},b_o) - \gamma_k a_1\} \qquad (4.3.4)\\
&\quad + \tfrac{2}{3}\{A^{jkl}(\hat{a},b_o) - \gamma_{jkl,2}(\hat{a},b_o)a_1 - 3\,\overline{\alpha}^{jkm}A^{lm}(\hat{a},b_o)\}\,\{A^j(\hat{a},b_o) - \gamma_j a_1\}\\
&\quad \times\{A^k(\hat{a},b_o) - \gamma_k a_1\}\{A^l(\hat{a},b_o) - \gamma_l a_1\}\\
&\quad + (\overline{\alpha}^{jkn}\overline{\alpha}^{lmn} - \tfrac{1}{2}\,\overline{\alpha}^{jklm})\{A^j(\hat{a},b_o) - \gamma_j a_1\}\{A^k(\hat{a},b_o) - \gamma_k a_1\}\\
&\quad \times\{A^l(\hat{a},b_o) - \gamma_l a_1\}\{A^m(\hat{a},b_o) - \gamma_m a_1\} + O_p(n^{-\frac{5}{2}}).
\end{aligned}$$

For ease of analysis we next express $\ell(b_o)$ in terms of powers of $(\hat{b}-b_o)$. Define $\eta_j = \sigma_x^2 u_j^2$, where u_j^2 is the $(j, 2)$ element in the matrix U_n. Notice that

$$
\begin{aligned}
A^j(\hat{a}, b_o) - \gamma_j a_1 &= n^{-1} \sum g_j(x_i)(y_i - \hat{a} - a_1 - b_o x_i) \\
&= n^{-1}(u_j^1, u_j^2) \sum \begin{pmatrix} 1 \\ x_i \end{pmatrix} (y_i - \hat{a} - a_1 - b_o x_i) \\
&= (u_j^1, u_j^2) \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & n^{-1}\sum x_i^2 \end{pmatrix} \left\{ (\hat{a}, \hat{b})^T - (\hat{a} + a_1, b_o)^T \right\} \\
&= (u_j^1, u_j^2) \{-a_1 + \overline{x}(\hat{b} - b_o), -\overline{x} a_1 + n^{-1} \sum x_i^2 (\hat{b} - b_o)\} \\
&= (u_j^1, u_j^2)(0, \sigma_x^2)(\hat{b} - b_o) \\
&= \eta_j (\hat{b} - b_o).
\end{aligned}
$$

From the definition of U_n we know that $u_j^2 u_j^2 = 1/(\sigma_x^2 \sigma^2)$, so that

$$\{A^j(\hat{a}, b_o) - \gamma_j a_1\} \{A^j(\hat{a}, b_o) - \gamma_j a_1\} = (\hat{b} - b_o)^2 \sigma_x^2/\sigma^2.$$

Moreover,

$$
\begin{aligned}
\gamma_{jk,1}(\hat{a}, b_o) &= \gamma_{jk,1} - \gamma_{jk}(\hat{a} - a_0), \\
a_2 &= -(\gamma_l \gamma_l)^{-1} \gamma_j \eta_k \{A_o^{jk} - \overline{\alpha}^{jkl} \eta_l (\hat{b} - b_o)\}(\hat{b} - b_o), \\
A^{jk}(\hat{a}, b_o) &= A_o^{jk} - 2\gamma_{jk,1}(\hat{a} - a_0) + \gamma_{jk}(\hat{a} - a_0)^2, \\
A^{jk}(\hat{a}, b_o) - 2\gamma_{jk,1}(\hat{a}, b_o) a_1 + \gamma_{jk} a_1^2 &= A_o^{jk} - 2\gamma_{jk,1}\overline{\epsilon} + \gamma_{jk}\overline{\epsilon}^2,
\end{aligned}
$$

and

$$
\begin{aligned}
A^{jkl}(\hat{a}, b_o) - \gamma_{jkl,2}(\hat{a}, b_o) a_1 &= A_o^{jkl} - 3\gamma_{jkl,2}(a_1 + \hat{a} - a_0) + O_p(n^{-1}), \\
&= A_o^{jkl} - 3\gamma_{jkl,2}(\overline{y} - a_o - b_o \overline{x}).
\end{aligned}
$$

Substituting the above formulae into (4.3.4) it may be shown that

$$
\begin{aligned}
n^{-1}\ell(b_o) = {} & \frac{\sigma_x^2}{\sigma^2}(\hat{b} - b_o)^2 - \eta_j \eta_k A_o^{jk}(\hat{b} - b_o)^2 + \tfrac{2}{3}\overline{\alpha}^{jkl}\eta_j \eta_k \eta_l (\hat{b} - b_o)^3 \\
& + \eta_j \eta_k (\gamma_{jk,1}\overline{\epsilon} - \gamma_{jk}\overline{\epsilon}^2 + A_o^{jl} A_o^{kl})(\hat{b} - b_o)^2 \\
& - (\gamma_l \gamma_l)^{-1} \gamma_j \gamma_m \eta_k \eta_n \{A_o^{jk} A_o^{mn} - 2\overline{\alpha}^{jkl}\eta_l A_o^{mn}(\hat{b} - b_o) \\
& + \overline{\alpha}^{jkl}\overline{\alpha}^{mnp}\eta_l \eta_p (\hat{b} - b_o)^2\}(\hat{b} - b_o)^2 \\
& + \tfrac{2}{3}\eta_j \eta_k \eta_l (A_o^{jkl} - 3\gamma_{jkl,2}\overline{\epsilon} - 2\overline{\alpha}^{jklm} A_o^{lm})(\hat{b} - b_o)^3 \\
& + (\overline{\alpha}^{jkn}\overline{\alpha}^{lmn} - \tfrac{1}{2}\overline{\alpha}^{jklm})\eta_j \eta_k \eta_l \eta_m (\hat{b} - b_o)^4 + O_p(n^{-\frac{5}{2}}) \qquad (4.3.5).
\end{aligned}
$$

The following nonparametric version of Wilks' theorem is a direct consequence of expansion (4.3.5).

Theorem 4.3.1 (Wilks' theorem) *Assume conditions (4.2.1). Then,*

$$P\{\ell(b_o) < c\} = P(\chi_1^2 < c) + o(1), \quad \textit{as} \quad n \to \infty.$$

Proof: Since $\mathrm{Var}(\hat{b} - b_o) = n^{-1}\,\sigma^2/\sigma_x^2$, by the Central Limit Theorem, we know that $n^{\frac{1}{2}}(\hat{b} - b_o)\,\sigma_x/\sigma$ has asymptotically a standard normal distribution. Thus from (4.3.5),

$$\ell(b_o) = \frac{n\sigma_x^2}{\sigma^2}(\hat{b} - b_o)^2 + O_p(n^{-\frac{1}{2}}) = \chi_1^2 + o_p(1).$$

Hence the theorem is proved. □

From Theorem 4.3.1 an empirical likelihood confidence interval for b_o with nominal coverage level α can be constructed as follows. First find from χ_1^2 tables the value c_α such that $P(\chi_1^2 < c_\alpha) = \alpha$. Then $I_\alpha = \{b_o | \ell(b_o) < c_\alpha\}$ is the α-level confidence interval for b_o. Theorem 4.3.1 ensures that I_α has correct asymptotic coverage.

4.3.2 Coverage Accuracy and Bartlett Correction

In this subsection we investigate coverage accuracy of the empirical likelihood confidence interval I_α for the slope parameter b_o. To this end, we decompose $\ell(b_o)$ from (4.3.5) as follows:

$$\ell(b_o) = n\,R_b^2 + O_p(n^{-\frac{5}{2}}), \tag{4.3.6}$$

where $R_b = R_{b1} + R_{b2} + R_{b_3}$ and $R_{bj} = O_p(n^{-j/2})$ for $j = 1, 2, 3$.

Put

$$C_1 = -\tfrac{1}{2}\sigma^2\overline{\alpha}^{j\,k\,l}\,\overline{\alpha}^{m\,n\,p}\,\eta_k\eta_l\eta_n\eta_p\,(\gamma_j\gamma_m + \tfrac{1}{9}\,\sigma_x^{-2}\,\eta_j\eta_m)$$
$$+\,\eta_j\eta_k\eta_l\eta_m\,(\tfrac{1}{2}\,\overline{\alpha}^{j\,k\,n}\,\overline{\alpha}^{l\,m\,n} - \tfrac{1}{4}\,\overline{\alpha}^{j\,k\,l\,m}).$$

Comparing (4.3.5) with (4.3.6) yields

$$
\begin{aligned}
R_{b1} &= \frac{\sigma_x}{\sigma}(\hat{b} - b_o), \\
R_{b2} &= -\tfrac{1}{2}\sigma_x^{-1}\,\sigma\,\eta_j\eta_k\,A_o^{jk}(\hat{b} - b_o) + \tfrac{1}{3}\sigma_x^{-1}\,\sigma\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,(\hat{b} - b_o)^2, \\
R_{b3} &= \sigma_x^{-1}\,\sigma\,\eta_j\eta_k\,(A_o^{jk}\,\overline{\epsilon} - \tfrac{1}{2}\gamma_{jk}\,\overline{\epsilon}^2 + \tfrac{1}{2}A_o^{jl}A_o^{kl})(\hat{b} - b_o) + C_1\,\sigma_x^{-1}\,\sigma\,(\hat{b} - b_o)^3 \\
&\quad - \sigma_x^{-1}\,\sigma\,(\tfrac{1}{2}\sigma^2\,\gamma_k\gamma_n\eta_j\eta_m + \tfrac{1}{8}\sigma^2\,\sigma_x^{-2}\eta_j\eta_k\eta_m\eta_n)\,A_o^{jk}A_o^{mn}\,(\hat{b} - b_o) \qquad (4.3.7) \\
&\quad + \tfrac{1}{3}\sigma_x^{-1}\,\sigma\,\eta_j\eta_k\eta_l\,A_o^{jkl}\,(\hat{b} - b_o)^2 \\
&\quad + \sigma_x^{-1}\,\sigma\,\{\sigma^2\,\overline{\alpha}^{jkl}\,\eta_k\eta_m\eta_l\,(\gamma_j\gamma_n + \tfrac{1}{6}\sigma_x^{-2}\,\eta_j\eta_n) - \overline{\alpha}^{jkm}\,\eta_j\eta_k\eta_n)\}A_o^{mn}\,(\hat{b} - b_o)^2.
\end{aligned}
$$

Before we develop an Edgeworth expansion for the distribution of $\ell(b_o)$ we introduce some notation. From (4.3.7) we see that there exists a smooth function h such that $R_b = h(\overline{U})$, where

$$\overline{U} = (\hat{b} - b_o, \overline{\epsilon}, A_o^{11}, A_o^{12}, A_o^{22}, A_o^{111}, A_o^{112}, A_o^{122}, A_o^{222}).$$

Let

$$B_1 = \begin{pmatrix} V_{n1}^{-\frac{1}{2}} \otimes V_{n1}^{-\frac{1}{2}} \\ V_{n1}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \\ V_{n2}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \end{pmatrix}$$

and

$$B_2 = \begin{pmatrix} V_{n1}^{-\frac{1}{2}} \otimes V_{n1}^{-\frac{1}{2}} \otimes V_{n1}^{-\frac{1}{2}} \\ V_{n1}^{-\frac{1}{2}} \otimes V_{n1}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \\ V_{n1}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \\ V_{n2}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \otimes V_{n2}^{-\frac{1}{2}} \end{pmatrix}$$

be 3×4 and 4×8 matrices respectively, where $\otimes$ is the Kronecker product of matrices and $V_{nj}^{-\frac{1}{2}}$ is the j'th row of $V_n^{-\frac{1}{2}}$, $j = 1, 2$. From the definition of A_o^{jk} and A_o^{jkl}, $\overline{U}$ can be expressed as $\overline{U} = n^{-1}\sum U_i$ where U_i is a vector of 9 dimensions having the form

$$U_i = \left[\sigma_x^{-2}(x_i - \overline{x})\epsilon_i, \epsilon_i, \{(1\ x_i) \otimes (1\ x_i)\}\ B_1^T, \{(1\ x_i) \otimes (1\ x_i) \otimes (1\ x_i)\}\ B_2^T\right].$$

Put $T_n = n^{-1}\sum \mathrm{Cov}(U_i)$, being the average covariance matrix of U_i, let g_1 be the density function of the χ_1^2 distribution and

$$t_1 = \frac{\mu_4}{\sigma^4\sigma_x^4}\,m_4, \quad t_2 = \frac{\mu_3^2}{\sigma^6\sigma_x^6}\,m_3^2 \quad \text{and} \quad m_j = n^{-1}\sum (x_i - \overline{x})^j, \quad \text{for} \quad j = 3, 4.$$

Then, we have the following theorem whose proof is deferred to Section 4.6.

Theorem 4.3.2: *Assume that*

(i) there exist positive constants C_1, C_2 such that uniformly in n

$C_1 \le v_{2n} \le v_{1n} \le C_2$; *(ii) the $|x_i|$'s for $1 \le i \le n$ are uniformly bounded;*

(iii) $E|\epsilon_1|^{15} < \infty$; *(iv) for every positive τ,* $\lim_{n\to\infty} \int_{|\epsilon_1| > \tau n^{\frac{1}{2}}} \|\epsilon_1\|^{15} = 0$; (4.3.8)

(vi) the smallest eigenvalue of T_n is bounded away from zero; (v) the characteristic function h of ϵ_1 satisties Cramér condition:

$$\limsup_{|t|\to\infty} |h(t)| < 1.$$

Then,

$$P\{\ell(b_o) < c_\alpha\} = \alpha - (1 + \tfrac{1}{2} t_1 - \tfrac{1}{3} t_2) n^{-1} c_\alpha g_1(c_\alpha) + O(n^{-\frac{3}{2}}). \qquad (4.3.9)$$

Theorem 4.3.2 states that the empirical likelihood confidence interval I_α has coverage error of order n^{-1}. By looking at the coefficient of the n^{-1} term in the Edgeworth expansion of the distribution function of $\ell(b_o)$, we see that the coverage error is dominated by a combination of four factors: the moments of ϵ_i, the "moments" of the fixed design points, the nominal coverage level and the sample size n.

Based on the expression for R_{bj}, $j = 1, 2, 3$ in (4.3.7), we may show that

$$\begin{aligned} E\{\ell(b_o)\} =& n\,\{E(R_{b1})^2 + 2\,E(R_{b1}\,R_{b2}) + E(R_{b2})^2 + 2\,E(R_{b1}\,R_{b3})\} + O(n^{-2}) \\ =& 1 + (1 + \tfrac{1}{2} t_1 - \tfrac{1}{3} t_2) n^{-1} + O(n^{-2}). \end{aligned}$$

We see that the difference between the means of $\ell(b_o)$ and limiting χ_1^2 distribution is of order n^{-1}. Next we are going to show that Bartlett correction can reduce the coverage errors of empirical likelihood confidence intervals to order n^{-2}. Let

$$\rho_b = 1 + \tfrac{1}{2} t_1 - \tfrac{1}{3} t_2$$

be the Bartlett factor for $\ell(b_o)$. We have the following theorem about the Bartlett correctability of the confidence interval I_α:

Theorem 4.3.3: *Assume condition (4.3.8). Then,*

$$P\{\ell(b_o) < c_\alpha(1 + \rho_b n^{-1})\} = \alpha + O(n^{-2}).$$

Define the Bartlett corrected confidence interval

$$I_{\alpha\rho_b} = \{b_o \,|\, \ell(b_o) < c_\alpha(1 + \rho_b n^{-1})\}.$$

Theorem 4.3.3 maintains that $I_{\alpha\rho_b}$ has coverage error of order n^{-2}, which is one order of magnitude more accurate than I_α. However, ρ_b is usually unknown because of unknown σ^2, μ_3 and μ_4 in t_1 and t_2, where σ^2, μ_3 and μ_4 are the second, third and fourth moments of ϵ_1. To empirically employ Bartlett correction, we have to give a root-n consistent estimate of ρ_b. To this end, put

$$\hat{t}_1 = \frac{\hat{\mu}_4}{\hat{\sigma}^4 \sigma_x^4} m_4, \quad t_2 = \frac{\hat{\mu}_3^2}{\hat{\sigma}^6 \sigma_x^6} m_3^2,$$

where $\hat{\sigma}^2$, $\hat{\mu}_3$ and $\hat{\mu}_4$ are the moment estimators of σ^2, μ_3 and μ_4 respectively, and

$$m_j = n^{-1} \sum (x_i - \overline{x})^j, \quad \text{for} \quad j = 3, 4.$$

We define a root-n consistent estimator of ρ_b, denoted by $\hat{\rho}_b$, to be

$$\hat{\rho}_b = 1 + \tfrac{1}{2}\hat{t}_1 - \tfrac{1}{3}\hat{t}_2.$$

We may show that under moderate conditions, such as that the joint distribution of components of the $\ell(b_o)$ and $\hat{\rho}_b$ admits multivariate Edgeworth expansions, or the distribution of $\ell(b_o) - c_\alpha\, \hat{\rho}_b\, n^{-1}$ admits an Edgeworth expansion under a smooth-function-of-means model, the same order of accuracy holds true if we replace ρ_b by $\hat{\rho}_b$ in Theorem 4.3.3. This implies that the empirical Bartlett-corrected confidence interval

$$I_{\alpha\hat{\rho}_b} = \{b_o \,|\, \ell(b_o) < c_\alpha(1 + \hat{\rho}_b n^{-1})\}$$

also has coverage error of order n^{-2}. Our simulation results in Section 4.5 show that the coverage of $I_{\alpha\hat{\rho}_b}$ is very close to that of $I_{\alpha\rho_b}$, and that both are more accurate than I_α.

4.4 Empirical Likelihood Confidence Interval for Means

In this section we construct empirical likelihood confidence intervals for the mean value $y_o = E(y|x = x_o) = a_o + b_o x_o$, for any fixed x_o. Since $y_o = a_o$ when $x_o = 0$, we may confine our attention to constructing empirical likelihood confidence intervals for a general y_o. The empirical log-likelihood ratio for y_o, denoted by $\ell(y_o)$, may be obtained by minimizing $\ell(a, b)$ given in (4.2.3), under the constraint of $a + bx_o = y_o$; that is,

$$\ell(y_o) = \ell(\tilde{a}, \tilde{b}) = \min_{a+bx_o=y_o} \ell(a, b),$$

where

$$\ell(a, b) = 2 \sum \log\{1 + \lambda\,(1\,, x_i)^T\,(y_i - a - bx_i\},$$

and $\lambda = (\lambda_1,\ \lambda_2)$ satisfies

$$\sum \frac{(1\,, x_i)^T\,(y_i - a - bx_i)}{1 + \lambda\,(1\,, x_i)^T\,(y_i - a - bx_i)} = 0.$$

4.4.1 Wilks' theorem

To obtain the limiting distribution of $\ell(y_o)$, we have to find out $\tilde{a}$ and $\tilde{b}$. Suppose $\tilde{a}$ and $\tilde{b}$ have expansions

$$\tilde{a} = \hat{a} + a_1 + a_2 + a_3 \quad \text{and} \quad \tilde{b} = \hat{b} + b_1 + b_2 + b_3,$$

where $a_j, b_j = O_p(n^{-j/2})$, $j = 1, 2, 3$. Note that we use notation $\tilde{a}$ and a_j again here, but with different meanings from those in the Section 3. In the following, a_j, b_j, $j = 1, 2, 3$ are determined successively. Put

$$\beta_j = n^{-1} \sum g_j(x_i)\, x_i, \quad \beta_{jk} = n^{-1} \sum g_{jk}(x_i)\, x_i,$$

$$\beta_{jkl} = n^{-1} \sum g_{jkl}(x_i)\, x_i, \quad \beta_{jk2} = n^{-1} \sum g_{jk}(x_i)\, x_i^2,$$

$$\beta_{jk,1}(a, b) = n^{-1} \sum g_{jk}(x_i)\,(y_i - a - bx_i),$$

$$\beta_{jkl,2}(a, b) = n^{-1} \sum g_{jkl}(x_i)\,(y_i - a - bx_i)^2.$$

Under (4.2.2),

$$n^{-1}\ell(a,b) = A^j(a,b)\,A^j(a,b) + O_p(n^{-\frac{3}{2}}),$$

$$= \{A^j(\hat a,\hat b) - \gamma_j(a-\hat a) - \beta_j(b-b_o)\}\{A^j(\hat a,\hat b) - \gamma_j(a-\hat a) - \beta_j(b-b_o)\}$$

$$+ O_p(n^{-\frac{3}{2}}).$$

Solving for a_1 and b_1 requires minimizing

$$\gamma_j\gamma_j\, a_1^2 + \beta_j\beta_j\, b_1^2 - 2\,A^j(\hat a,\hat b)(\gamma_j\, a_1 + \beta_j\, b_1^2) + 2\,\gamma_j\beta_j\, a_1\, b_1,$$

subject to $\hat a + a_1 + (\hat b + b_1)\,x_o = y_o$. By the Langrage multiplier method, a_1 and b_1 satisfy

$$\begin{aligned}(\gamma_j\beta_j - \beta_j\beta_j\, x_o)\,a_1 + (\beta_j\beta_j - \gamma_j\beta_j x_o)\,b_1 &= A^j(\hat a,\hat b)(\beta_j - \gamma_j\, x_o),\\ a_1 + b_1\, x_o &= (b_o - \hat b)\,x_o + (a_o - \hat a).\end{aligned} \tag{4.4.1}$$

By the definitions of $g_j(x_i)$, γ_j and β_j, and from (4.4.1), we obtain

$$a_1 = -\frac{\sigma_x^2 + \overline{x}(\overline{x}-x_o)}{\sigma_x^2 + (\overline{x}-x_o)^2}W_o, \quad b_1 = \frac{(\overline{x}-x_o)}{\sigma_x^2 + (\overline{x}-x_o)^2}W_o, \tag{4.4.2}$$

where $W_o = \hat a + \hat b\, x_o - y_o = (\hat a - a_o) + (\hat b - b_o)\,x_o$. Clearly $W_o = O_p(n^{-\frac{1}{2}})$.

From the first equation of (4.4.1) we immediately have

$$\{A^j(\hat a,\hat b) - \gamma_j\, a_1 - \beta_j\, b_1\}(\beta_j - \gamma_j x_o) = 0. \tag{4.4.3}$$

To determine a_2 and b_2, we notice that

$$\begin{aligned}n^{-1}\ell(a,b) =&A^j A^j - A^{jk}(a,b_o)\,A^j A^k + \frac{2}{3}\overline{\alpha}^{jkl} A^j A^k A^l + O_p(n^{-2})\\ =&\gamma_j\gamma_j\, a_2^2 + 2\,\gamma_j\beta_j\, a_2\, b_2 + \beta_j\beta_j\, b_2^2 - 2\,\{A^j(\hat a,\hat b) - \gamma_j\, a_1 - \beta_j\, b_1\}(\gamma_j\, a_2 + \beta_j\, b_2)\\ &+ 2\,A^{jk}(\hat a,\hat b)\,\{A^j(\hat a,\hat b) - \gamma_j\, a_1 - \beta_j\, b_1\}(\gamma_k\, a_2 + \beta_k\, b_2)\\ &- 2\,\overline{\alpha}^{jkl}\,\{A^j(\hat a,\hat b) - \gamma_j\, a_1 - \beta_j\, b_1\}\,\{A^k(\hat a,\hat b) - \gamma_k\, a_1 - \beta_k\, b_1\}\times\\ &\times\{A^l(\hat a,\hat b) - \gamma_l\, a_1 - \beta_l\, b_1\} + R_3 + O_p(n^{-\frac{5}{2}}),\end{aligned}$$

where R_3 denotes a term not involving a_2 and b_2. Using (4.4.1), (4.4.3) and the Langrage multiplier method to minimize $\ell(a,b)$ under the constraint of $a_2 + b_2 x_o = 0$, we end up with $a_2 = -b_2\, x_o$ where

$$\begin{aligned}b_2 =&\{(\beta_m - \gamma_m\, x_o)(\beta_m - \gamma_m\, x_o)\}^{-1}\,\{A^j(\hat a,\hat b) - \gamma_j\, a_1 - \beta_j\, b_1\}\\ &\times\left[\overline{\alpha}^{jkl}\,\{A^k(\hat a,\hat b) - \gamma_k\, a_1 - \beta_k\, b_1\}(\beta_l - \gamma_l\, x_o) - A^{jk}(\hat a,\hat b)(\beta_k - \gamma_k\, x_o)\right].\end{aligned}$$

It can be shown after some algebra that

$$\ell(y_o) = \{A^j(\hat{a},\hat{b}) - \gamma_j\, a_1 - \beta_j\, b_1\}(\gamma_j\, a_3 + \beta_j\, b_3) + R_4 + O_p(n^{-\frac{5}{2}}),$$

where R_4 denotes a term not involving a_3 and b_3. Since $a_3 + b_3\, x_o = 0$, and in view of (4.4.3), we have

$$\ell(y_o) = R_4 + O_p(n^{-\frac{5}{2}}).$$

This means that a_3 and b_3 will not appear in the expansion of $\ell(y_o)$ up to the order of $O_p(n^{-2})$, so we need not concern with them any further.

Put $t^j = A^j(\hat{a},\hat{b}) - \gamma_j\, a_1 - \beta_j\, b_1$. Substituting $\tilde{a} = \hat{a} + a_1 + a_2 + a_3$ and $\tilde{b} = \hat{b} + b_1 + b_2 + b_3$ into the formula for $n^{-1}\,\ell(\tilde{a},\tilde{b})$, we obtain that

$$\begin{aligned} n^{-1}\ell(y_o) =& t^j\, t^j + \tfrac{2}{3}\,\overline{\alpha}^{j\,k\,l}\, t^j\, t^k\, t^l - (\beta_j - \gamma_j x_o)(\beta_j - \gamma_j x_o)\, b_2^2 \\ & - \{A^{jk}(\hat{a},\hat{b}) - 2\,\gamma_{jk,1}\, a1 - 2\,\beta_{jk,1}\, b_1 + \gamma_{jk}\, a_1^2 + 2\,\beta_{jk}\, a_1\, b_1 + \beta_{jk2}\, b_1^2\}\, t^j\, t^k \\ & + A^{jl}(\hat{a},\hat{b})\, A^{kl}(\hat{a},\hat{b})\, t^j\, t^k + \tfrac{2}{3}\,\{A^{jkl}(\hat{a},\hat{b}) - 3\,\gamma_{jkl,2}\, a_1 - 3\,\beta_{jkl,2}\, b_1\}\, t^j\, t^k\, t^l \\ & - 2\,\overline{\alpha}^{j\,k\,m}\, A^{lm}(\hat{a},\hat{b})\, t^j\, t^k\, t^l + (\overline{\alpha}^{j\,k\,n}\overline{\alpha}^{l\,m\,n} - \tfrac{1}{2}\,\overline{\alpha}^{j\,k\,l\,m})\, t^j\, t^k\, t^l\, t^m \\ & + O_p(n^{-\frac{5}{2}}). \end{aligned} \tag{4.4.4}$$

Define

$$\alpha^2(x_o) = \frac{\sigma_x^2}{\sigma_x^2 + (\overline{x} - x_o)^2} \quad \text{and} \quad \xi^j = u_{j1} + u_{j2} x_o \;\; j = 1,2.$$

Then we have

$$t^j = \alpha^2(x_o)\, W_o^2\, \xi^j \quad \text{and} \quad t^j\, t^j = \alpha^2(x_o)\,\sigma^{-2}\, W_o^2.$$

Since $\hat{a} + a_1 - a_o = -(\hat{b} + b_1 - b_o)\, x_o$ we have

$$\begin{aligned} & A^{jk}(\hat{a},\hat{b}) - 2\,\gamma_{jk,1}\, a1 - 2\,\beta_{jk,1}\, b_1 + \gamma_{jk}\, a_1^2 + 2\,\beta_{jk}\, a_1\, b_1 + \beta_{jk2}\, b_1^2 \\ =& A_o^{jk} - 2\,(\beta_{jk,1} - \gamma_{jk,1}\, x_o)(\hat{b} + b_1 - b_o) + (\beta_{jk2} - 2\,\beta_{jk}\, x_o + \gamma_{jk}\, x_o^2)(\hat{b} + b_1 - b_o)^2. \end{aligned}$$

From the expressions for b_2 and the fact that $(\beta_j - \gamma_j\, x_o)(\beta_j - \gamma_j\, x_o) = \sigma_x^2\, \sigma^{-2}\, \alpha^{-2}(x_o)$,

$$b_2 = \sigma_x^{-2}\, \sigma^2\, \alpha^4(x_o)\,\{\overline{\alpha}^{j\,k\,l}\, \alpha^2(x_o)\, W_o^2\, \xi^j\, \xi^k\, (\beta_l - \gamma_l\, x_o) - W_o\, A_o^{jk}\, \xi^j\, (\beta_k - \gamma_k\, x_o)\}.$$

Moreover,

$$A^{jkl}(\hat{a},\hat{b})-3\,\gamma_{jkl,2}\,a_1-3\,\beta_{jkl,2}\,b_1 = A_o^{jkl}-3\,\sigma^2\,(\beta_{jkl}-\gamma_{jkl}\,x_o)\,(\hat{b}+b_1-b_o)+O_p(n^{-1}).$$

Substituting the above formulae into (4.4.4), we obtain

$$\begin{aligned}
&n^{-1}\ell(y_o)\\
&=\alpha^2(x_o)\,\sigma^{-2}\,W_o^2-\alpha^4(x_o)\,\xi^j\,\xi^k\,A_o^{jk}\,W_o^2+\tfrac{2}{3}\,\alpha^6(x_o)\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^l\,W_o^3\\
&\quad-\sigma_x^{-2}\,\sigma^2\,\alpha^6(x_o)\,\{\overline{\alpha}^{jkl}\,\alpha^2(x_o)\,\xi^j\,\xi^k\,(\beta_l-\gamma_l\,x_o)\,W_o^2-\xi^j\,(\beta_k-\gamma_k\,x_o)\,A_o^{jk}\,W_o\}^2\\
&\quad+2\,\alpha^4(x_o)\,\xi^j\,\xi^k\,(\beta_{jk,1}-\gamma_{jk,1}\,x_o)\,(b_1+\hat{b}-b_o)\,W_o^2\\
&\quad-\alpha^4(x_o)\,\xi^j\,\xi^k\,\{(\beta_{jk2}-2\,\beta_{jk}\,x_o+\gamma_{jk}\,x_o^2)\,(b_1+\hat{b}-b_o)^2-A_o^{jl}\,A_o^{kl}\}\,W_o^2\\
&\quad+\alpha^6(x_o)\,\xi^j\,\xi^k\,\xi^l\,\{\tfrac{2}{3}\,A_o^{jkl}-2\,(\beta_{jkl}-\gamma_{jkl}\,\sigma^2\,x_o)\,(b_1+\hat{b}-b_o)-2\,\overline{\alpha}^{jkm}\,A_o^{lm}\}\,W_o^3\\
&\quad-2\,\alpha^6(x_o)\,\xi^j\,\xi^k\,\xi^l\,W_o^3+\alpha^8(x_o)\,(\overline{\alpha}^{jkn}\overline{\alpha}^{lmn}-\frac{1}{2}\overline{\alpha}^{jklm})\xi^j\,\xi^k\,\xi^l\,\xi^m\,W_o^4\\
&\quad+O_p(n^{-\frac{5}{2}}).
\end{aligned}\tag{4.4.5}$$

The following nonparametric version of Wilks' theorem is a direct conclusion of (4.4.5).

Theorem 4.4.1 (Wilks' theorem) *Assume conditions (4.2.1) and (4.2.2). Then,*

$$P\{\ell(y_o)<c\}=P(\chi_1^2<c)+o(1)\quad n\to\infty.$$

Proof: From (4.4.5) we know that

$$\ell(y_o)=\;n\,\alpha(x_o)^2\,\sigma^{-2}\,W_0^2+O_p(n^{-\frac{1}{2}})=\;n\,\sigma^{-2}\,\frac{\sigma_x^2}{\sigma_x^2+(\overline{x}-x_o)^2}\,W_0^2+O_p(n^{-\frac{1}{2}}).$$

Thus the theorem is proved by the fact that W_o is asymptotically Normal with mean zero and variance $n^{-1}\,\sigma^2\,\sigma_x^{-2}\,\{\sigma_x^2+(\overline{x}-x_o)^2)\}$. □

Now an empirical likelihood confidence interval for y_o with asymptotic coverage level α can be constructed as $J_\alpha = \{y_o|\ell(y_o) < c_\alpha\}$ where $P(\chi_1^2 < c_\alpha) = \alpha$. Theorem 4.4.1 assures that J_α has correct asympototic coverage.

4.4.2 Coverage Accuracy and Bartlett correction

In this subsection we shall investigate the second-order properties of J_α, that is the coverage accuracy and Bartlett correctability of J_α. We start with a signed root decomposition of $\ell(y_o)$, which can be obtained from (4.4.5) as follows:

$$\ell(y_o) = n\, R_{y_o}^2 + O_p(n^{-\frac{5}{2}}),$$

where $R_{y_o} = R_{y_o 1} + R_{y_o 2} + R_{y_o 3}$, and $R_{y_o j} = O_p(n^{-j/2})$ for $j = 1,2,3$. A little algebra shows that

$$\begin{aligned}
R_{y_o 1} &= \alpha(x_o)\,\sigma^{-1}\, W_o, \\
R_{y_o 2} &= \alpha^3(x_o)\,\sigma\,\xi^j\,\xi^k\,\{-\tfrac{1}{2} A_o^{jk}\, W_o + \tfrac{1}{3}\alpha^2(x_o)\,\overline{\alpha}^{j\,k\,l}\,\xi^l\, W_o^2\}, \\
R_{y_o 1}\, R_{y_o 3} &= \alpha^4(x_o)\,\xi^j\,\xi^k\,(\beta_{jk,1} - \gamma_{jk,1}\, x_o)(b_1 + \hat{b} - b_o)\, W_o^2 + C_2\, W_o^4 \\
&\quad - \tfrac{1}{2}\alpha^4(x_o)\,\xi^j\,\xi^k\,(\beta_{jk2} - 2\,\beta_{jk}\, x_o + \gamma_{jk}\, x_o^2)(b_1 + \hat{b} - b_o)^2\, W_o^2 \\
&\quad - \tfrac{1}{2}\alpha^6(x_o)\,\sigma^2\,\xi^j\,\xi^m\,\sigma_x^{-2}\,(\beta_k - \gamma_l\, x_o)(\beta_n - \gamma_n\, x_o)\, A_o^{jk}\, A_o^{mn}\, W_o^2 \\
&\quad - \tfrac{1}{8}\alpha^6(x_o)\,\sigma^2\,\xi^j\,\xi^m\,\xi^k\,\xi^n\, A_o^{jk}\, A_o^{mn}\, W_o^2 \\
&\quad + \alpha^8(x_o)\,\sigma^2 \Big\{ \sigma_x^{-2}\,\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^m\,(\beta_l - \gamma_l\, x_o)(\beta_n - \gamma_n\, x_o) \\
&\quad + \tfrac{1}{6}\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\,\xi^m\,\xi^n - \alpha^{-2}(x_o)\,\sigma^{-2}\,\overline{\alpha}^{j\,k\,m}\,\xi^j\,\xi^k\,\xi^n \Big\}\, A_o^{mn}\, W_o^3 \\
&\quad + \tfrac{1}{2}\alpha^4(x_o)\,\xi^j\,\xi^k\, A_o^{jl}\, A_o^{kl}\, W_o^2 + \tfrac{1}{3}\alpha^6(x_o)\,\xi^j\,\xi^k\,\xi^l\, A_o^{jkl}\, W_o^3 \\
&\quad - \alpha^6(x_o)\,\sigma^2\,\xi^j\,\xi^k\,\xi^l\,(\beta_{jkl,2} - \gamma_{jkl,2}\, x_o)(b_1 + \hat{b} - b_o)\, W_o^3,
\end{aligned}$$

where

$$\begin{aligned}
C_2 = &- \tfrac{1}{2}\alpha^{10}(x_o)\,\sigma^2\,\sigma_x^{-2}\,\{\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,(\beta_l - \gamma_l\, x_o)\}^2 - \tfrac{1}{18}\alpha^{10}(x_o)\,\sigma^2\,\{\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\}^2 \\
&+ \alpha^8(x_o)\,(\tfrac{1}{2}\overline{\alpha}^{j\,k\,n}\overline{\alpha}^{l\,m\,n} - \tfrac{1}{4}\overline{\alpha}^{j\,k\,l\,m})\,\xi^j\,\xi^k\,\xi^l\,\xi^m.
\end{aligned}$$

Put

$$s_1 = \alpha^4(x_o)\,\sigma^{-4}\,\mu_4\, q_1, \quad s_2 = \alpha^6(x_o)\,\sigma^{-6}\,\mu_3^2\, q_2^2 \quad \text{and} \quad s_3 = \alpha^4(x_o)\, q_3,$$

where

$$q_1 = 1 + 6\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2} - 4\,\frac{(\overline{x} - x_o)^3}{\sigma_x^6}\, m_3 + \frac{(\overline{x} - x_o)^4}{\sigma_x^8}\, m_4,$$

$$q_2 = 1 + 3\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2} - \frac{(\overline{x} - x_o)^3}{\sigma_x^6}\, m_3 \quad \text{and}$$

$$q_3 = 1 - 3\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2} + \frac{(\overline{x} - x_o)^2}{\sigma_x^6}\, m_4 + \frac{(\overline{x} - x_o)^4}{\sigma_x^4} + 2\left\{ \frac{(\overline{x} - x_o)^3}{\sigma_x^6} - \frac{(\overline{x} - x_o)}{\sigma_x^4} \right\} m_3.$$

The coverage accuracy of confidence interval J_α is discussed in the following theorem, whose proof is deferred to Section 4.6.

Theorem 4.4.2: *Assume condition (4.3.8). Then,*

$$P\{\ell(y_o) < c_\alpha\} = \alpha - (\tfrac{1}{2}\, s_1 - \tfrac{1}{3}\, s_2 + s_3)\, n^{-1}\, c_\alpha\, g_1(c_\alpha) + O(n^{-\frac{3}{2}}). \qquad (4.4.6)$$

Theorem 4.4.2 states that the coverage errors of empirical likelihood confidence intervals for $y_o = a + b\,x_o$ are of order of n^{-1}, provided that x_o is fixed and independent of sample size n. From the n^{-1} order term in (4.4.6) and the definitions of s_1, s_2 and s_3, we see that the coverage error is dominated by the combination of the following five factors: the moments of ϵ_i, the "moments" of the fixed design points, the nominal coverage level, the sample size n, and the size of $(\overline{x} - x_o)/\sigma_x$ —i.e. the standard distance between x_o and the centre, $\overline{x}$, of the design points.

In analogy with the Bartlett correction for the slope parameter b_o developed in Theorem 4.3.3, we can do the same thing here for y_o. Calculations reveal that

$$E\{\ell(y_o)\} = n\,\{E(R_{y_o})^2\} + O(n^{-2}) = 1 + (\tfrac{1}{2}\, s_1 - \tfrac{1}{3}\, s_2 + s_3)\, n^{-1} + O(n^{-2}).$$

Define

$$\rho_{y_o} = (\tfrac{1}{2}\, s_1 - \tfrac{1}{3}\, s_2 + s_3),$$

an ingredient of the Bartlett correction for $\ell(y_o)$. The Bartlett correction property for the empirical likelihood confidence interval for y_o is considered by the following theorem:

Theorem 4.4.3: *Assume conditions (4.3.8). For any $x > 0$ and fixed x_o,*

$$P\{\ell(y_o) < c_\alpha\,(1 + \rho_{y_o}\, n^{-1})\} = \alpha + O(n^{-2}).$$

We shall not give the proof of Theorem 4.4.3 since it is almost the same as that of Theorem 4.3.3. Theorem 4.4.3 states that a simple scale adjustment can improve the coverage accuracy of empirical likelihood confidence intervals for y_o from $O(n^{-1})$ to $O(n^{-2})$. Define the Bartlett corrected confidence interval

$$J_{\alpha\rho_{y_o}} = \{\, y_o \,|\, \ell(y_o) < c_\alpha\,(1 + \rho_{y_o}\, n^{-1})\}.$$

Theorem 4.4.3 ensures that

$$P(y_o \in J_{\alpha\rho_{y_o}}) = \alpha + O(n^{-2}).$$

However, ρ_{y_o} is usually unknown because σ^2, μ_3 and μ_4 are unknown. A root-n consistent estimate $\hat{\rho}_{y_o}$ of ρ_{y_o} can be obtained by replacing σ^2, μ_3 and μ_4 by $\hat{\sigma}^2$, $\hat{\mu}_3$ and $\hat{\mu}_4$ repectively in s_1 and s_2. Hence,

$$\hat{\rho}_{y_o} = (\tfrac{1}{2}\,\hat{s}_1 - \tfrac{1}{3}\,\hat{s}_2 + s_3)$$

where

$$\hat{s}_1 = \alpha^4(x_o)\,\hat{\sigma}^{-4}\,\hat{\mu}_4\,q_1 \quad \hat{s}_2 = \alpha^6(x_o)\,\hat{\sigma}^{-6}\,\hat{\mu}_3^2\,q_2^2.$$

Put

$$J_{\alpha\hat{\rho}_{y_o}} = \{\, y_o \,|\, \ell(y_o) < c_\alpha\,(1 + \hat{\rho}_{y_o}\, n^{-1})\}.$$

It may be shown that under moderate conditions, which ensure that $\ell(y_o) - c_\alpha\,\hat{\rho}_{y_o}\,n^{-1}$ admits an Edgeworth expansion under a smooth-function-of-means model, we may get same order of accuracy by replacing ρ_{y_o} with $\hat{\rho}_{y_o}$ in Theorem 4.4.3. Therefore, we have

$$P(y_o \in J_{\alpha\hat{\rho}_{y_o}}) = \alpha + O(n^{-2}).$$

Our simulation results in the next section confirm this.

4.5 Simulation Study

This section describes simulation experiments carried out to examine the coverage properties of the empirical likelihood confidence intervals for b_o and y_o proposed in the previous sections. The following simple linear regression model was treated:

$$y_i = 1 + x_i + \epsilon_i, \quad i = 1, \cdots, n.$$

The data set x_i is the one which displayed in Table 3.1 of Chapter 3. We chose sample sizes $n = 15, 30, 50$ and nominal coverage level $\alpha = 0.90, 0.95$. We assigned two error patterns for ϵ_i. One was $\epsilon_i = N(0,1)$, another was $\epsilon_i = E(1.00) - 1.00$, where $N(0,1)$ and $E(1.00)$ were random variables with the standard normal distribution and the exponential distribution with unit mean, respectively. The normal and exponential random variables were generated by the routines of Press et al.(1989).

For each combination of n, α and ϵ_i we display in Table 4.1 the coverages of the uncorrected confidence intervals and two Bartlett corrected confidence intervals based on 10,000 simulations. One of the corrected confidence intervals uses the theoretical Bartlett correction, another uses the empirical Bartlett correction. Standard errors are given for each of the simulated coverages. To empirically justify the expansions developed in Theorems 4.3.2 and 4.4.2, we also give theoretical coverages up to the second order in Edgeworth expansions for $\ell(b_o)$ and $\ell(y_o)$. Since the coverages can be obtained without simulation, they are called "predicted coverages".

The following broad conclusions may be drawn from the results summarized in Table 4.1. Firstly, the differences between the uncorrected coverages and their corresponding "predicted coverages" converge to zero as n increases. This gives empirical justification for Theorems 4.3.2 and 4.4.2. Secondly, substantial improvements on coverage accuracy have been made by implementing Bartlett corrections. This can be observed by looking at both the standard errors and absolute errors. Thirdly, the empirical Bartlett correction performs similarly to its theoretical Bartlett correction counterpart, except for the small sample skewed case.

TABLE 4.1: Estimated true coverages, from 10,000 simulations, of α-level empirical likelihood confidence regions for b_o and y_o's. Rows headed "predic.", "uncorr.", "b_o" or "y_o" and "$\hat{b}_o$" or "$\hat{y}_o$"give the predicted, uncorrected and Bartlett-corrected coverages respectively. The figures in parentheses are 10^2 times the standard errors associated with the simulated coverages.

(1) Coverages for slope parameter b_o

	ϵ_i	$N(0,1)$		E(1.00)-1.00	
n	α	0.90	0.95	0.90	0.95
15	predic.	0.840	0.909	0.750	0.849
	uncorr.	0.803 (0.40)	0.860 (0.35)	0.789 (0.41)	0.858 (0.35)
	ρ_{b_o}	0.859 (0.35)	0.911 (0.28)	0.904 (0.30)	0.950 (0.22)
	$\hat{\rho}_{b_o}$	0.853 (0.35)	0.906 (0.29)	0.856 (0.35)	0.916 (0.28)
30	predic.	0.878	0.935	0.845	0.913
	uncorr.	0.862 (0.35)	0.919 (0.27)	0.840 (0.37)	0.902 (0.30)
	ρ_{b_o}	0.884 (0.32)	0.935 (0.25)	0.880 (0.31)	0.931 (0.24)
	$\hat{\rho}_{b_o}$	0.883 (0.32)	0.934 (0.25)	0.871 (0.34)	0.928 (0.26)
50	predic.	0.888	0.939	0.870	0.930
	uncorr.	0.882 (0.32)	0.9386 (0.24)	0.860 (0.35)	0.926 (0.26)
	ρ_{b_o}	0.896 (0.31)	0.948 (0.22)	0.887 (0.32)	0.944 (0.23)
	$\hat{\rho}_{b_o}$	0.896 (0.31)	0.948 (0.22)	0.880 (0.33)	0.938 (0.24)

(2) Coverages for intercept parameter a_o

n	ϵ_i / α	$N(0,1)$ 0.90	0.95	E(1.00)-1.00 0.90	0.95
15	predic.	0.858	0.921	0.802	0.884
	uncorr.	0.822 (0.38)	0.884 (0.32)	0.805 (0.40)	0.868 (0.34)
	ρ_{y_o}	0.861 (0.35)	0.918 (0.27)	0.883 (0.32)	0.927 (0.26)
	$\hat{\rho}_{y_o}$	0.857 (0.35)	0.915 (0.28)	0.848 (0.36)	0.900 (0.30)
30	predic.	0.880	0.937	0.865	0.921
	uncorr.	0.864 (0.34)	0.922 (0.27)	0.840 (0.37)	0.901 (0.30)
	ρ_{y_o}	0.888 (0.32)	0.937 (0.24)	0.874 (0.33)	0.933 (0.25)
	$\hat{\rho}_{y_o}$	0.884 (0.32)	0.936 (0.24)	0.863 (0.34)	0.922 (0.27)
50	predic.	0.887	0.941	0.871	0.931
	uncorr.	0.883 (0.32)	0.933 (0.25)	0.860 (0.35)	0.920 (0.27)
	ρ_{y_o}	0.894 (0.31)	0.942 (0.23)	0.884 (0.32)	0.942 (0.23)
	$\hat{\rho}_{y_o}$	0.894 (0.31)	0.942 (0.23)	0.877 (0.33)	0.933 (0.25)

(3) Coverages for mean parameter y_o with $x_o = 5.00$

n	ϵ_i / α	$N(0,1)$ 0.90	$N(0,1)$ 0.95	E(1.0)-1.00 0.90	E(1.0)-1.00 0.95
15	predic.	0.865	0.926	0.840	0.909
	uncorr.	0.837 (0.37)	0.899 (0.30)	0.815 (0.39)	0.869 (0.34)
	ρ_{y_o}	0.871 (0.34)	0.924 (0.27)	0.868 (0.34)	0.908 (0.29)
	$\hat{\rho}_{y_o}$	0.867 (0.34)	0.922 (0.27)	0.846 (0.36)	0.893 (0.31)
30	predic.	0.885	0.940	0.875	0.934
	uncorr.	0.882 (0.32)	0.936 (0.25)	0.861 (0.35)	0.922 (0.27)
	ρ_{y_o}	0.897 (0.30)	0.946 (0.23)	0.884 (0.32)	0.938 (0.24)
	$\hat{\rho}_{y_o}$	0.897 (0.30)	0.946 (0.23)	0.876 (0.33)	0.932 (0.25)
50	predic.	0.889	0.943	0.879	0.936
	uncorr.	0.887 (0.32)	0.937 (0.24)	0.871 (0.33)	0.923 (0.27)
	ρ_{y_o}	0.898 (0.30)	0.945 (0.23)	0.891 (0.31)	0.939 (0.25)
	$\hat{\rho}_{y_o}$	0.897 (0.30)	0.944 (0.23)	0.884 (0.32)	0.933 (0.25)

(4) Coverages for mean parameter y_o with x_o = 10.00

ϵ_i		$N(0,1)$		E(0,1)-1.00	
n	α	0.90	0.95	0.90	0.95
15	predic.	0.832	0.904	0.744	0.844
	uncorr.	0.785 (0.41)	0.849 (0.36)	0.763 (0.39)	0.831 (0.37)
	ρ_{y_o}	0.855 (0.35)	0.905 (0.29)	0.884 (0.34)	0.928 (0.26)
	$\hat{\rho}_{y_o}$	0.847 (0.36)	0.899 (0.30)	0.833 (0.36)	0.889 (0.31)
30	predic.	0.876	0.934	0.846	0.914
	uncorr.	0.860 (0.35)	0.913 (0.28)	0.833 (0.37)	0.892 (0.27)
	ρ_{y_o}	0.885 (0.32)	0.931 (0.25)	0.874 (0.33)	0.933 (0.24)
	$\hat{\rho}_{y_o}$	0.882 (0.32)	0.930 (0.26)	0.863 (0.34)	0.919 (0.25)
50	predic.	0.889	0.943	0.877	0.935
	uncorr.	0.881 (0.32)	0.939 (0.24)	0.862 (0.35)	0.925 (0.26)
	ρ_{y_o}	0.893 (0.31)	0.948 (0.22)	0.884 (0.32)	0.942 (0.23)
	$\hat{\rho}_{y_o}$	0.892 (0.31)	0.947 (0.22)	0.875 (0.33)	0.936 (0.25)

4.6 Proofs

In this section we give proofs of Theorems 4.3.2, 4.3.3, 4.4.2 and 4.4.3.

4.6.1 Proof of Theorem 4.3.2

Theorem 4.3.2: *Assume that*

(i) there exist positive constants C_1, C_2 such that uniformly in n $C_1 \le v_{2n} \le v_{1n} \le C_2$; (ii) the $|x_i|$'s for $1 \le i \le n$ are uniformly bounded; (iii) $E|\epsilon_1|^{15} < \infty$; (iv) for every positive τ, $\lim_{n\to\infty} \int_{\|\epsilon_1\| > \tau n^{\frac{1}{2}}} |\epsilon_1|^{15} = 0$; (vi) the smallest eigenvalue of T_n is bounded away from zero; (v) the characteristic function h of ϵ_1 satisties Cramér condition:

$$\limsup_{|t|\to\infty} |h(t)| < 1.$$

Then,

$$P\{\ell(b_o) < c_\alpha\} = \alpha - (1 + \tfrac{1}{2}t_1 - \tfrac{1}{3}t_2)n^{-1}c_\alpha g_1(c_\alpha) + O(n^{-\frac{3}{2}}). \tag{4.6.1}$$

Proof: Let k_{bj} be the j'th cumulant of $n^{\frac{1}{2}}R_b$. Calculations deferred to Appendix 4.1 show that

$$\begin{aligned} k_{b1} &= -\tfrac{1}{6}t_2^{\frac{1}{2}}n^{-\frac{1}{2}} + O(n^{-\frac{3}{2}}), \\ k_{b2} &= 1 + (1 + \tfrac{1}{2}t_1 - \tfrac{13}{36}t_2)n^{-1} + O(n^{-2}), \\ k_{bj} &= O(n^{-\frac{3}{2}}), \quad j \ge 3. \end{aligned} \tag{4.6.2}$$

A formal Edgeworth expansion for the distribution function of R_b can be constructed as follows,

$$P(n^{\frac{1}{2}}R_b < x) = \int_{-\infty}^{x} \Psi(v)\phi(v)\,dv + O(n^{-\frac{3}{2}}), \tag{4.6.3}$$

where $\Psi(v) = 1 + \frac{1}{6}t_2^{\frac{1}{2}}vn^{-\frac{1}{2}} + \frac{1}{2}(1 + \frac{1}{2}t_1 - \frac{1}{3}t_2)(v^2 - 1)n^{-1}$. Accepting that expansion (4.6.3) may be justified, we establish an Edgeworth expansion for the distribution of $\ell(b_o)$, as follows:

$$\begin{aligned} P\{\ell(b_o) < c\} &= P(-c^{\frac{1}{2}} < n^{\frac{1}{2}} R_b < c^{\frac{1}{2}}) + O(n^{-\frac{3}{2}}) \\ &= \int_{-c^{\frac{1}{2}}}^{c^{\frac{1}{2}}} \Psi(v)\,\phi(v)\,dv + O(n^{-\frac{3}{2}}) \\ &= \alpha - (1 + \tfrac{1}{2} t_1 - \tfrac{1}{3} t_2)\, n^{-1} c g_1(c) + O(n^{-\frac{3}{2}}), \end{aligned}$$

where g_1 is the density function of the χ_1^2 distribution.

It remains to check that expansion (4.6.3) is valid. Remember that $R_b = h_o(\overline{U})$ where h_o is a sufficient smooth function and

$$\overline{U} = (\hat{b} - b_o, \overline{\epsilon}, A_o^{11}, A_o^{12}, A_o^{22}, A_o^{111}, A_o^{112}, A_o^{122}, A_o^{222}).$$

is the mean of independent but not identically distributed random variable U_i's. For this case, Theorem 1.3.3 ensures a valid Edgeworth expansion. It may be shown that condition (4.2.7) implies the conditions of Theorem 1.3.3. Thus, a valid Edgeworth expansion for $\overline{U}$ can be obtained. Consequently, the Edgeworth expansion of $\overline{U}$ may be transformed by a smooth function h_o to yield another valid Edgeworth expansion (4.3.6) for R_b by using Theorem 1.3.4. Therefore the theorem is proved. □

4.6.2 Proof of Theorem 4.3.3

Theorem 4.3.3: *Assume condition (4.3.8). Then,*

$$P\{\ell(b_o) < c_\alpha(1 + \rho_b n^{-1})\} = \alpha + O(n^{-2}).$$

Proof: The method of proof is similar to that of Theorem 3.3.1. Recalling (4.6.1), and noting that $\rho_b = 1 + \frac{1}{2} t_1 - \frac{1}{3} t_2$, we have

$$\begin{aligned} P\{\ell(b_o) < c_\alpha(1 + \rho_b n^{-1})\} = {} & P\{\chi_1^2 < c_\alpha(1 + \rho_b n^{-1})\} \\ & - \rho_b\, n^{-1}\, c_\alpha\,(1 + \rho_b n^{-1})\, g_1\{c_\alpha(1 + \rho_b n^{-1})\} \\ & + O(n^{-\frac{3}{2}}). \end{aligned} \tag{4.6.4}$$

Note too that

$$P\{\chi_1^2 < c_\alpha(1 + \rho_b n^{-1})\} = P(\chi_1^2 < c_\alpha) + \rho_b\, c_\alpha\, g_1(c_\alpha)\, n^{-1} + O(n^{-2}) \tag{4.6.5}$$

and

$$g_1\{c_\alpha(1+\rho_b n^{-1})\} = g_1(c_\alpha) + O(n^{-1}), \tag{4.6.6}$$

where g_1 is the density function of the χ_1^2 distribution.

Substituting (4.6.5) and (4.6.6) into (4.6.4) yields,

$$P\{\ell(b_o) < c_\alpha(1+\rho_b n^{-1})\} = \alpha + O(n^{-\frac{3}{2}}). \tag{4.6.7}$$

By the parity property of the polynomials in the coefficients of the above Edgeworth expansion, it can be shown that the $O(n^{-\frac{3}{2}})$ term in (4.6.7) is actually $O(n^{-2})$. Thus the theorem is proved. □

4.6.3 Proof of Theorem 4.4.2

Theorem 4.4.2: *Assume condition (4.3.8). Then,*

$$P\{\ell(y_o) < c_\alpha\} = \alpha - (\tfrac{1}{2}s_1 - \tfrac{1}{3}s_2 + s_3)n^{-1}\,c\,g_1(c) + O(n^{-\frac{3}{2}}).$$

Proof: Let $k_{y_o j}$, $j = 1, 2, \cdots$ denote the j'th cumulants of $n^{\frac{1}{2}} R_{y_o}$. Calculations deferred to Appendix 4.2 show that

$$\begin{aligned} k_{y_o 1} &= -\tfrac{1}{6}s_2^{\frac{1}{2}}n^{-\frac{1}{2}} + O(n^{-\frac{3}{2}}), \\ k_{y_o 2} &= 1 + (\tfrac{1}{2}s_1 - \tfrac{13}{36}s_2 + s_3)n^{-1} + O(n^{-2}), \\ k_{y_o j} &= O(n^{-\frac{3}{2}}), \quad j \geq 3. \end{aligned} \tag{4.6.8}$$

A formal Edgeworth expansion for the distribution of $n^{\frac{1}{2}} R_{y_o}$ can be set up from (4.4.4) as follows:

$$P(n^{\frac{1}{2}} R_b < x) = \int_{-\infty}^{x} \Pi(v)\,\phi(v)\,dv + O(n^{-\frac{3}{2}}), \tag{4.6.9}$$

where

$$\Pi(v) = 1 + \tfrac{1}{6}s_2^{\frac{1}{2}}\,v\,n^{-\frac{1}{2}} + \tfrac{1}{2}(\tfrac{1}{2}s_1 - \tfrac{1}{3}s_2 + s_3)(v^2-1)n^{-1}.$$

The validity of expansion (4.6.9) can be argued in the following way. Notice from expressions for $R_{y_o j}$ $j = 1, 2, 3$ that there is a smooth function Q_1 such that $R_{y_o} = Q_1(\overline{S})$, where

$$\overline{S} = (W_o, b_1 + \hat{b} - b_o, A_o^{11}, A_o^{12}, A_o^{22}, A_o^{111}, A_o^{112}, A_o^{122}, A_o^{222}).$$

Since

$$b_1 = (\overline{x} - x_o)\{\sigma_x^2 + (\overline{x} - x_o)^2\}^{-1} W_o \quad \text{and} \quad W_o = \overline{\epsilon} + (\overline{x} - x_o)(\hat{b} - b_o),$$

there exists another smooth function Q_2 such that $\overline{S} = Q_2(\overline{U})$, where

$$\overline{U} = (\hat{b} - b_o, \overline{\epsilon}, A_o^{11}, A_o^{12}, A_o^{22}, A_o^{111}, A_o^{112}, A_o^{122}, A_o^{222}).$$

Thus there exists a smooth function $Q = Q_1 Q_2$ such that $R_{y_o} = Q(\overline{U})$. It can be shown that condition (4.2.7) implies the condition of Theorem 1.3.3. Thus, using Theorems 1.3.3 and 1.3.4, we obtain the validaty of the Edgeworth expansion (4.6.9). From (4.6.9) and integrating , we immediately get the conclusion of Theorem 4.4.2.

□

Appendix 4 Calculations of Cumulants

In this appendix we display the calculation for the cumulants of $n^{\frac{1}{2}} R_b$ and $n^{\frac{1}{2}} R_{y_o}$, which were used in the proof of Theorems 4.3.2 and 4.4.2.

Appendix 4.1 Calculation of Cumulants of $n^{\frac{1}{2}} R_b$

In this section we calculate the cumulants of $n^{\frac{1}{2}} R_b$ which were shown in (4.6.2). Recall that

$$R_b = R_{b1} + R_{b2} + R_{b3},$$

where

$$\begin{aligned}
R_{b1} &= \frac{\sigma_x}{\sigma}(\hat{b} - b_o),\\
R_{b2} &= -\tfrac{1}{2}\sigma_x^{-1}\sigma\,\eta_j\eta_k\,A_o^{jk}(\hat{b} - b_o) + \tfrac{1}{3}\sigma_x^{-1}\sigma\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,(\hat{b} - b_o)^2,\\
R_{b3} &= \sigma_x^{-1}\sigma\,\eta_j\eta_k\,(A_o^{jk}\,\overline{\epsilon} - \tfrac{1}{2}\gamma_{jk}\,\overline{\epsilon}^2 + \tfrac{1}{2}A_o^{jl}A_o^{kl})(\hat{b} - b_o) + C_1\,\sigma_x^{-1}\sigma(\hat{b} - b_o)^3\\
&\quad - \sigma_x^{-1}\sigma\,(\tfrac{1}{2}\sigma^2\,\gamma_k\gamma_n\eta_j\eta_m + \tfrac{1}{8}\sigma^2\,\sigma_x^{-2}\eta_j\eta_k\eta_m\eta_n)\,A_o^{jk}A_o^{mn}\,(\hat{b} - b_o)\\
&\quad + \tfrac{1}{3}\sigma_x^{-1}\sigma\,\eta_j\eta_k\eta_l\,A_o^{jkl}\,(\hat{b} - b_o)^2\\
&\quad + \sigma_x^{-1}\sigma\,\{\sigma^2\,\overline{\alpha}^{jkl}\,\eta_k\eta_m\eta_l\,(\gamma_j\gamma_n + \tfrac{1}{6}\sigma_x^{-2}\,\eta_j\eta_n) - \overline{\alpha}^{jkm}\,\eta_j\eta_k\eta_n)\}\,A_o^{mn}\,(\hat{b} - b_o)^2
\end{aligned}$$

and

$$\begin{aligned}
C_1 &= -\tfrac{1}{2}\sigma^2\overline{\alpha}^{jkl}\,\overline{\alpha}^{mnp}\,\eta_k\eta_l\eta_n\eta_p\,(\gamma_j\gamma_m + \tfrac{1}{9}\sigma_x^{-2}\,\eta_j\eta_m)\\
&\quad + \eta_j\eta_k\eta_l\eta_m\,(\tfrac{1}{2}\overline{\alpha}^{jkn}\,\overline{\alpha}^{lmn} - \tfrac{1}{4}\overline{\alpha}^{jklm}).
\end{aligned}$$

From the definition of U_n we have the following basic formulae:

$$u_j^1 u_j^1 = \frac{n^{-1}\sum x_i^2}{\sigma^2\sigma_x^2}, \quad u_j^1 u_j^2 = -\frac{\overline{x}}{\sigma^2\sigma_x^2}, \quad \text{and} \quad u_j^1 u_j^1 = \frac{1}{\sigma^2\sigma_x^2}. \tag{4.A.1}$$

Put

$$t_1 = \frac{\mu_4}{\sigma^4\sigma_x^4}\,m_4 \quad \text{and} \quad t_2 = \frac{\mu_3^2}{\sigma^6\sigma_x^6}\,m_3^2.$$

Using (4.A.1) and the facts that $\eta_j = \sigma_x^2 u_j^2$ and $\overline{\alpha}^{jkl} = n^{-1}\sum g_{jkl}(x_i)\,\mu_3$, we have

$$\begin{aligned}
\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l &= \mu_3\,\sigma^{-6}\,m_3,\\
\eta_j\eta_k\,n^{-1}\sum_{i=1}^{n} g_{jk}(x_i)(x_i - \overline{x}) &= \sigma^{-4}\,m_3.
\end{aligned} \tag{4.A.2}$$

Since $E(R_{b1}) = 0$, we obtain

$$\begin{aligned} E(R_b) =& E(R_{b2}) + O(n^{-2}) \\ =& -\tfrac{1}{2}\sigma\,\sigma_x^{-1}\,\eta_j\eta_k\,E\{A_o^{jk}\,(\hat{b}-b_o)\} + \tfrac{1}{3}\sigma\,\sigma_x^{-1}\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,E(\hat{b}-b_o)^2 + O(n^{-2}) \\ =& -\tfrac{1}{2}\sigma\,\sigma_x^{-3}\,\eta_j\eta_k\,n^{-1}\textstyle\sum g_{jk}(x_i)(x_i-\overline{x})\,\mu_3 n^{-1} + \tfrac{1}{3}\sigma\,\sigma_x^{-3}\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,n^{-1} \\ & + O(n^{-2}) \\ =& -\tfrac{1}{6}\,t_2^{\frac{1}{2}}n^{-1} + O(n^{-2}). \end{aligned}$$

Thus we have

$$k_{b1} = -\tfrac{1}{6}\,t_2^{\frac{1}{2}}n^{-\frac{1}{2}} + O(n^{-\frac{3}{2}}). \tag{4.A.3}$$

To calculate k_{b2}, notice that

$$E(R_b^2) = E(R_{b1})^2 + 2\,E(R_{b1}\,R_{b2}) + E(R_{b2})^2 + 2\,E(R_{b1}\,R_{b3}) + O(n^{-3}).$$

Clearly $E(R_{b1})^2 = n^{-1}$. Since

$$\eta_j\eta_k\,n^{-1}\sum g_{jk}(x_i)(x_i-\overline{x})^2 = \sigma^{-4}\,m_4, \tag{4.A.4}$$

then (4.A.2) and (4.A.4) imply

$$\begin{aligned} E(R_{b1}R_{b2}) =& -\tfrac{1}{2}\,\eta_j\eta_k\,E\{A_o^{jk}\,(\hat{b}-b_o)^2\} + \tfrac{1}{3}\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,E(\hat{b}-b_o)^3 \\ =& -\tfrac{1}{2}\,(\mu_4-\sigma^4)\,\sigma_x^{-4}\,\eta_j\eta_k\,n^{-1}\textstyle\sum g_{jk}(x_i)(x_i-\overline{x})^2\,n^{-2} \\ & + \tfrac{1}{3}\,\mu_3\,\sigma_x^{-6}\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,m_3\,n^{-2} + O(n^{-3}) \\ =& \{-\tfrac{1}{2}\,(\mu_4-\sigma^4)\,\sigma_x^{-4}\,\sigma^{-4}\,m_4 + \tfrac{1}{3}\,t_2\}\,n^{-2} + O(n^{-3}). \end{aligned}$$

From (4.A.1) we know that

$$\eta_j\eta_k\,\eta_m\eta_n\,n^{-1}\sum g_{jkmn}(x_i) = \sigma^{-8}\,m_4. \tag{4.A.5}$$

Using (4.A.2) - (4.A.5) yields

$$
\begin{aligned}
E(R_{b2})^2 &= \tfrac{1}{4}\,\sigma^2\,\sigma_x^{-2}\,\eta_j\eta_k\,\eta_m\eta_n\,E\{A_o^{jk}\,A_o^{mn}(\hat b - b_o)^2\} \\
&\quad - \tfrac{1}{3}\,\sigma^2\,\sigma_x^{-2}\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\eta_m\eta_n\,E\{A_o^{mn}(\hat b - b_o)^3\} \\
&\quad + \tfrac{1}{9}\,\sigma^2\,\sigma_x^{-2}\,\overline{\alpha}^{jkl}\,\overline{\alpha}^{mnp}\,\eta_j\eta_k\eta_l\eta_m\eta_n\eta_p\,E(\hat b - b_o)^4 \\
&= \tfrac{1}{4}\,\sigma^4\,(\mu_4 - \sigma^4)\,\sigma_x^{-4}\,\eta_j\eta_k\,\eta_m\eta_n\,n^{-1}\textstyle\sum g_{jkmn}(x_i)\,n^{-2} \\
&\quad + \tfrac{1}{2}\,\sigma^2\,\mu_3^2\,\sigma_x^{-6}\,\{\eta_j\eta_k\,n^{-1}\textstyle\sum g_{jk}(x_i)(x_i - \overline{x})\}\,n^{-2} \\
&\quad + \tfrac{1}{3}\,\sigma^6\,\sigma_x^{-6}\,\left(\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\right)^2\,n^{-2} \\
&\quad - \sigma^4\,\mu_3\,\sigma_x^{-6}\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\eta_m\eta_n\,n^{-1}\textstyle\sum g_{mn}(x_i)(x_i - \overline{x})\,n^{-2} + O(n^{-3}) \\
&= \{\tfrac{1}{4}(\mu_4 - \sigma^4)\sigma_x^{-4}\,\sigma^{-4}\,m_4 - \tfrac{1}{6}\,t_2\}\,n^{-2} + O(n^{-3}).
\end{aligned}
$$

To calculate $E(R_{b1}R_{b3})$, note that

$$
\begin{aligned}
E(R_{b1}\,R_{b3}) &= \eta_j\eta_k\,E\{A_o^{jk}\,\overline{\epsilon}(\hat b - b_o)^2\} - \tfrac{1}{2}\gamma_{jk}\,\eta_j\eta_k\,E\{\overline{\epsilon}^2\,(\hat b - b_o)^2\} \\
&\quad - \left(\tfrac{1}{2}\sigma^2\,\gamma_k\gamma_n\eta_j\eta_m + \tfrac{1}{8}\sigma^2\,\sigma_x^{-2}\eta_j\eta_k\eta_m\eta_n\right)\,E\{A_o^{jk}A_o^{mn}\,(\hat b - b_o)^2\} \\
&\quad + \left(\sigma^2\,\overline{\alpha}^{jkl}\,\gamma_j\gamma_n\eta_k\eta_m\eta_l + \tfrac{1}{6}\sigma^2\,\sigma_x^{-2}\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\eta_m\eta_n - \overline{\alpha}^{jkm}\,\eta_j\eta_k\eta_n\right) \\
&\quad \times E\{A_o^{mn}\,(\hat b - b_o)^3\} + \tfrac{1}{2}\,\eta_j\eta_k\,E\{A_o^{jl}A_o^{kl}\,(\hat b - b_o)^2\} \\
&\quad + \tfrac{1}{3}\,\eta_j\eta_k\eta_l\,E\{A_o^{jkl}\,(\hat b - b_o)^2\} + C_1\,E\{(\hat b - b_o)^4\}. \qquad (4.A.6)
\end{aligned}
$$

We shall calculate the right-hand side of (4.A.6) term by term. Now, (4.A.1) implies that

$$\eta_j\eta_k\;\gamma_{jk} \;=\; \sigma_x^2\,\sigma^{-4}.$$

Neglecting terms of order of $O(n^{-3})$, we see that the sum of the first two terms on the right-hand side of (4.A.6) is

$$
\begin{aligned}
&\eta_j\eta_k\,E\{A_o^{jk}\,\overline{\epsilon}\,(\hat b - b_o)^2\} - \tfrac{1}{2}\,\gamma_{jk}\,\eta_j\eta_k\,E\{\overline{\epsilon}^2\,(\hat b - b_o)^2\} \\
&= \tfrac{1}{2}\,\eta_j\eta_k\;\gamma_{jk}\,\sigma_x^{-2}\,\sigma^4\,n^{-2} \;=\; \tfrac{1}{2}\,n^{-2}. \qquad (4.A.7)
\end{aligned}
$$

From (4.A.1) we know that

$$
\begin{aligned}
\gamma_k\gamma_n\,\eta_j\eta_m\,n^{-1}\textstyle\sum g_{jklm}(x_i) &= \sigma_x^2\,\sigma^{-8}, \\
\eta_j\eta_k\eta_m\eta_n\,n^{-1}\textstyle\sum g_{jklm}(x_i) &= m_4\,\sigma^{-8}, \\
\eta_j\gamma_k\,n^{-1}\textstyle\sum g_{jk}(x_i)(x_i - \overline{x}) &= \sigma_x^2\,\sigma^{-4}.
\end{aligned}
$$

Thus,

$$-\left(\tfrac{1}{2}\sigma^2\,\gamma_k\gamma_n\,\eta_j\,\eta_m+\tfrac{1}{8}\sigma^2\,\sigma_x^{-2}\eta_j\,\eta_k\,\eta_m\,\eta_n\right)\;E\{A_o^{jk}A_o^{mn}\,(\hat{b}-b_o)^2\}$$

$$=-\left(\tfrac{1}{2}\sigma^2\,\gamma_k\gamma_n\,\eta_j\,\eta_m+\tfrac{1}{8}\sigma^2\,\sigma_x^{-2}\eta_j\,\eta_k\,\eta_m\,\eta_n\right)\;\sigma_x^{-4}\{n^{-1}\sum g_{jklm}(x_i)\,\sigma_x^2\sigma^2(\mu_4-\sigma^4)$$

$$+\,2\,n^{-1}\sum g_{jk}(x_i)(x_i-\overline{x})\,n^{-1}\sum g_{mn}(x_i)(x_i-\overline{x})\,\mu_3^2\} \tag{4.A.8}$$

$$=-\tfrac{1}{2}(\mu_4-\sigma^4)\,\sigma^{-4}\,n^{-2}-\tfrac{1}{8}(\mu_4-\sigma^4)\,\sigma_x^{-4}\sigma^{-4}m_4\,n^{-2}-\mu_3^2\sigma^{-6}\,n^{-2}-\tfrac{1}{4}t_2\,n^{-2}.$$

To calculate the fourth term, observe that

$$C_1=-\tfrac{1}{2}\sigma^2\overline{\alpha}^{jkl}\,\overline{\alpha}^{mnp}\,\eta_k\,\eta_l\eta_n\,\eta_p\,(\gamma_j\gamma_m+\tfrac{1}{9}\,\sigma_x^{-2}\eta_j\,\eta_m)$$

$$+\,\eta_j\,\eta_k\,\eta_l\eta_m\,(\tfrac{1}{2}\,\overline{\alpha}^{jkn}\,\overline{\alpha}^{lmn}-\tfrac{1}{4}\,\overline{\alpha}^{jklm})$$

$$=\tfrac{4}{9}\,\mu_3^3\,\sigma_x^{-2}\,\sigma^{-10}\,m_3^2-\tfrac{1}{4}\,\mu_4\,\sigma^{-8}\,m_4.$$

Therefore,

$$C_1\,E(\hat{b}-b_o)^4=-\tfrac{3}{4}\,\mu_4\,\sigma^{-4}\,\sigma_x^{-4}\,n^{-1}\textstyle\sum(x_i-\overline{x})^4\,n^{-2}+\tfrac{4}{3}\,t_2\,n^{-2}. \tag{4.A.9}$$

Using (4.A.1) again we have

$$\overline{\alpha}^{jkl}\,\gamma_j\,\eta_k\,\eta_l=\mu_3\,\sigma_x^2\,\sigma^{-6}$$

and

$$\overline{\alpha}^{jkm}\,\eta_j\,\eta_k\,\eta_n\,n^{-1}\sum g_{mn}(x_i)(x_i-\overline{x})=\mu_3\,\sigma_x^{-2}\,\sigma^{-8}\,(\sigma_x^6+m_3^2).$$

So the fifth term is

$$\{\sigma^2\,\overline{\alpha}^{jkl}\,\eta_k\,\eta_m\,\eta_l\,(\gamma_j\,\gamma_n+\tfrac{1}{6}\,\sigma_x^{-2}\,\eta_j\,\eta_n)-\overline{\alpha}^{jkm}\,\eta_j\,\eta_k\,\eta_n)\}\,E\{A_o^{mn}\,(\hat{b}-b_o)^3\}$$

$$=3\,(\sigma^2\,\overline{\alpha}^{jkl}\,\gamma_j\,\gamma_n\,\eta_k\,\eta_m\,\eta_l+\tfrac{1}{6}\,\sigma^2\,\sigma_x^{-2}\overline{\alpha}^{jkl}\,\eta_j\,\eta_k\,\eta_l\eta_m\,\eta_n-\overline{\alpha}^{jkm}\,\eta_j\,\eta_k\,\eta_n)$$

$$\times\,n^{-1}\sum g_{mn}(x_i)(x_i-\overline{x})\,\mu_3\,\sigma_x^{-4}\,\sigma^2\,n^{-2}$$

$$=3\,\mu_3^2\,\sigma^{-6}\,n^{-2}+\tfrac{1}{2}\,t_2\,n^{-2}-3\,\mu_3^2\,\sigma^{-6}\,n^{-2}-3\,t_2\,n^{-2}=-\tfrac{5}{2}\,t_2\,n^{-2}. \tag{4.A.10}$$

Observe that (4.A.1) implies

$$\eta_j\,\eta_k\,n^{-1}\sum g_{jkll}=\sigma_x^{-2}\,\sigma^{-6}\,(\sigma_x^4+m_4),$$

$$\eta_j\,\eta_k\,n^{-1}\sum g_{jl}(x_i-\overline{x})\,n^{-1}\sum g_{kl}(x_i-\overline{x})=\sigma_x^{-2}\,\sigma^{-6}\,(\sigma_x^6+m_3^2),$$

so that we have the sixth term

$$
\begin{aligned}
&\tfrac{1}{2}\,\eta_j\eta_k\,E\{A_o^{jl}A_o^{kl}\,(\hat{b}-b_o)^2\}\\
&=\tfrac{1}{2}\,\eta_j\eta_k\,\sigma_x^{-4}\{n^{-1}\textstyle\sum g_{jkll}(x_i)\,\sigma_x^2\,\sigma^2\,(\mu_4-\sigma^4)\\
&\quad+2\,n^{-1}\textstyle\sum g_{jl}\,(x_i-\overline{x})\,n^{-1}\textstyle\sum g_{kl}\,(x_i-\overline{x})\,\mu_3^2 \qquad (4.A.11)\\
&=\tfrac{1}{2}\,(\mu_4-\sigma^4)\,\sigma^{-4}\,n^{-2}+\tfrac{1}{2}\,(\mu_4-\sigma^4)\,\sigma^{-4}\,\sigma_x^{-4}\,m_4\,n^{-2}+\mu_3^2\,\sigma^{-6}\,n^{-2}+t_2\,n^{-2}.
\end{aligned}
$$

Finally, since

$$\eta_j\eta_k\eta_l\,n^{-1}\sum g_{jkl}(x_i-\overline{x}) = m_4\,\sigma^{-6},$$

we have

$$
\begin{aligned}
\tfrac{1}{3}\,\eta_j\eta_k\eta_l\,E\{A_o^{jkl}\,(\hat{b}-b_o)^3\} &= \eta_j\eta_k\eta_l\,n^{-1}\textstyle\sum g_{jkl}(x_i-\overline{x})\,\sigma^{-4}\,\sigma^2\,\mu_4\\
&= \mu_4\,\sigma^{-4}\,\sigma_x^{-4}\,m_4\,n^{-2} = t_1\,n^{-2}. \qquad (4.A.12)
\end{aligned}
$$

Substituting (4.A.7)-(4.A.12) into (4.A.6), we obtain

$$
\begin{aligned}
E(R_{b1}\,R_{b3}) &= \tfrac{1}{2}\,n^{-2}+\tfrac{3}{8}\,(\mu_4-\sigma^4)\,\sigma^{-4}\,\sigma_x^{-4}\,m_4\,n^{-2}-\tfrac{5}{12}\,t_2\,n^{-2}\\
&\quad+\tfrac{1}{4}\,t_1\,n^{-1}\textstyle\sum(x_i-\overline{x})^4\,n^{-2}+O(n^{-3}).
\end{aligned}
$$

In summary,

$$
\begin{aligned}
E(R_b^2) &= E(R_{b1})^2+2\,E(R_{b1}\,R_{b2})+E(R_{b2})^2+2\,E(R_{b1}\,R_{b3})+O(n^{-3})\\
&= n^{-1}+((1+\tfrac{1}{2}\,t_1-\tfrac{1}{3}\,t_2)\,n^{-2}+O(n^{-3}).
\end{aligned}
$$

Since $k_{b2} = n\{E(R_b)^2 - E^2(R_b)\}$, we obtain

$$k_{b2} = 1+(1+\tfrac{1}{2}\,t_1-\tfrac{13}{36}\,t_2)\,n^{-1}+O(n^{-2}). \qquad (4.A.13)$$

Next we calculate k_{b3}. By definition,

$$
\begin{aligned}
k_{b3} &= n^{\frac{3}{2}}\{E(R_b^3)-3\,E(R_b)\,E(R_b^2)+2\,E^3(R_b)\}\\
&= n^{\frac{3}{2}}\{E(R_{b1}^3)+3\,E(R_{b1}^2\,R_{b2})-3\,E(R_{b2})\,E(R_{b1}^2)+O(n^{-\frac{3}{2}})\}.
\end{aligned}
$$

Since $E(\hat{b}-b_o)^3 = \mu_3\,\sigma_x^{-6}\,m_3\,n^{-2}+O(n^{-3})$, we immediately have

$$E(R_{b1}^3) = t_2^{\frac{1}{2}}\,n^{-2}+O(n^{-3}). \qquad (4.A.14)$$

Using (4.A.2),

$$E(R_{b1}^2\, R_{b2}) = -\tfrac{1}{2}\,\sigma_x\,\sigma^{-1}\,\eta_j\eta_k\,E\{A_o^{jk}\,(\hat{b}-b_o)^3\} + \tfrac{1}{3}\,\sigma^{-1}\,\sigma_x\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,E(\hat{b}-b_o)^4$$

$$= -\tfrac{3}{2}\,\sigma\,\mu_3\,\sigma_x^{-3}\,\eta_j\eta_k\,n^{-1}\sum g_{jk}(x_i)(x_i-\overline{x})\,n^{-2} + \tfrac{1}{3}\,\sigma^3\,\sigma_x^{-3}\,\overline{\alpha}^{jkl}\,\eta_j\eta_k\eta_l\,n^{-2}$$

$$+\,O(n^{-3})$$

$$= -\tfrac{3}{2}\,t_2^{\frac{1}{2}}\,n^{-2} + t_2^{\frac{1}{2}}\,n^{-2} + O(n^{-3})$$

$$= -\tfrac{1}{2}\,t_2^{\frac{1}{2}}\,n^{-2} + O(n^{-3}). \tag{4.A.15}$$

From the early calculations of $E(R_b)$ and $E(R_b^2)$,

$$E(R_{b2})\,E(R_{b1}^2) = -\tfrac{1}{6}\,t_2^{\frac{1}{2}}\,n^{-2} + O(n^{-3}). \tag{4.A.16}$$

Now (4.A.14) - (4.A.16) imply that

$$k_{b3} = O(n^{-\frac{3}{2}}). \tag{4.A.17}$$

From the definition of k_{b4} and the result that $k_{b3} = O(n^{-\frac{3}{2}})$,

$$k_{b4} = n^2\,\{E(R_b^4) - 3\,E^2(R_b^2) - 4\,E(R_b)\,E(R_b^3) + 12\,E^2(R_b)\,E(R_b^2) - 6\,E^4(R_b)\}$$

$$= n^2\,\{E(R_b^4) - 3\,E^2(R_b^2)\} - 4\,n^{\frac{1}{2}}\,E(R_b)\,k_{b3} + 2\,k_{b1}^4$$

$$= n^2\,\{E(R_b^4) - 3\,E^2(R_b^2)\} + O(n^{-2}).$$

$$= n^2\,\{E(R_{b1}^4) - 3\,E^2(R_{b1}^2) + 4\,E(R_{b1}^3\,R_{b2}) - 12\,E(R_{b1}^2)\,E(R_{b1}\,R_{b2})$$

$$+\,6\,E(R_{b1}^2\,R_{b2}^2) - 6\,E(R_{b1}^2)\,E(R_{b2}^2) + 4\,E(R_{b1}^3\,R_{b3})$$

$$-\,12\,E(R_{b1}^2)\,E(R_{b1}\,R_{b3})\} + O(n^{-2}). \tag{4.A.18}$$

From (4.A.2) - (4.A.5) we may show that

$$E(R_{b1}^4) - 3\,E^2(R_{b1}^2) = (\mu_4 - 3\,\sigma^4)\,\sigma^{-4}\,\sigma_x^{-4}\,m_4\,n^{-3},$$

$$4\,E(R_{b1}^3\,R_{b2}) - 12\,E(R_{b1}^2)\,E(R_{b1}\,R_{b2}) = \{-6\,(\mu_4 - \sigma^4)\,\sigma^{-4}\,\sigma^{-4}\,m_4 + \tfrac{4}{3}\,t_2\}n^{-3} + O(n^{-4}),$$

$$6\,E(R_{b1}^2\,R_{b2}^2) - 6\,E(R_{b1}^2)\,E(R_{b2}^2) = \{3\,(\mu_4 - \sigma^4)\,\sigma^{-4}\,\sigma^{-4}\,m_4 - t_2\}n^{-3} + O(n^{-4}).$$

Using the same argument which yields $E(R_{b1}\,R_{b3})$, we may show that

$$4\,E(R_{b1}^3\,R_{b3}) - 12\,E(R_{b1}^2)\,E(R_{b1}\,R_{b3}) = (2\,t_1 - \tfrac{1}{3}\,t_2)\,n^{-3} + O(n^{-4}).$$

Substituting the above four expressions into (4.A.18), we get

$$k_{b4} = O(n^{-2}). \tag{4.A.19}$$

From the results given by James and Mayne (1964),

$$k_{bj} = O(n^{-\frac{3}{2}}), \text{for any } j \geq 5. \tag{4.A.20}$$

In view of (4.A.3), (4.A.13), (4.A.17), (4.A.19) and (4.A.20), we readily derive (4.6.2).

Appendix 4.2. Calculations of Cumulants of $n^{\frac{1}{2}} R_{y_o}$

In this part of the appendix we give our derivation of (4.6.8), which was used to prove Theorem 4.4.2. It turns out that the calculations of the cumulants of $n^{\frac{1}{2}} R_{y_o}$ given in this section is very similar to that in Appendix 4.1. Put

$$q_1 = 1 + 6\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2} - 4\,\frac{(\overline{x} - x_o)^3}{\sigma_x^6}\,m_3 + \frac{(\overline{x} - x_o)^4}{\sigma_x^8}\,m_4,$$
$$q_2 = 1 + 3\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2} - \frac{(\overline{x} - x_o)^3}{\sigma_x^6}\,m_3$$

and

$$\begin{aligned} q_3 =& 1 - 3\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2} + \frac{(\overline{x} - x_o)^2}{\sigma_x^6}\,m_4 + \frac{(\overline{x} - x_o)^4}{\sigma_x^4} \\ &+ 2\,\{\frac{(\overline{x} - x_o)^3}{\sigma_x^6} - \frac{(\overline{x} - x_o)}{\sigma_x^4}\}m_3, \end{aligned}$$

and define

$$s_1 = \alpha^4(x_o)\,\sigma^{-4}\,\mu_4\,q_1, \quad s_2 = \alpha^6(x_o)\,\sigma^{-6}\,\mu_3^2\,q_2^2, \quad s_3 = \alpha^4(x_o)\,q_3.$$

Remember that $R_{y_o} = R_{y_o1} + R_{y_o2} + R_{y_o3}$ where

$$R_{y_o1} = \alpha(x_o)\sigma^{-1} W_o,$$

$$R_{y_o2} = \alpha^3(x_o)\sigma\, \xi^j \xi^k \{-\tfrac{1}{2} A_o^{jk} W_o + \tfrac{1}{3}\alpha^2(x_o)\overline{\alpha}^{jkl}\xi^l W_o^2\},$$

$$\begin{aligned}
R_{y_o1} R_{y_o3} = {} & \alpha^4(x_o)\xi^j \xi^k (\beta_{jk,1} - \gamma_{jk,1} x_o)(b_1 + \hat{b} - b_o) W_o^2 + C_2 W_o^4 \\
& - \tfrac{1}{2}\alpha^4(x_o)\xi^j \xi^k (\beta_{jk2} - 2\beta_{jk} x_o + \gamma_{jk} x_o^2)(b_1 + \hat{b} - b_o)^2 W_o^2 \\
& - \alpha^6(x_o)\sigma^2 \xi^j \xi^m \left\{\tfrac{1}{2}\sigma_x^{-2}(\beta_k - \gamma_l x_o)(\beta_n - \gamma_n x_o) + \tfrac{1}{8}\xi^k \xi^n\right\} A_o^{jk} A_o^{mn} W_o^2 \\
& + \alpha^8(x_o)\sigma^2 \left\{\sigma_x^{-2}\overline{\alpha}^{jkl}\xi^j \xi^k \xi^m (\beta_l - \gamma_l x_o)(\beta_n - \gamma_n x_o)\right. \\
& \left. + \tfrac{1}{6}\overline{\alpha}^{jkl}\xi^j \xi^k \xi^l \xi^m \xi^n - \alpha^{-2}(x_o)\sigma^{-2}\overline{\alpha}^{jkm}\xi^j \xi^k \xi^n\right\} A_o^{mn} W_o^3 \\
& + \tfrac{1}{2}\alpha^4(x_o)\xi^j \xi^k A_o^{jl} A_o^{kl} W_o^2 + \tfrac{1}{3}\alpha^6(x_o)\xi^j \xi^k \xi^l A_o^{jkl} W_o^3 \\
& - \alpha^6(x_o)\sigma^2 \xi^j \xi^k \xi^l (\beta_{jkl,2} - \gamma_{jkl,2} x_o)(b_1 + \hat{b} - b_o) W_o^3,
\end{aligned}$$

and

$$\begin{aligned}
C_2 = {} & -\tfrac{1}{2}\alpha^{10}(x_o)\sigma^2\sigma_x^{-2}\{\overline{\alpha}^{jkl}\xi^j \xi^k (\beta_l - \gamma_l x_o)\}^2 - \tfrac{1}{18}\alpha^{10}(x_o)\sigma^2\{\overline{\alpha}^{jkl}\xi^j \xi^k \xi^l\}^2 \\
& + \alpha^8(x_o)(\tfrac{1}{2}\overline{\alpha}^{jkn}\overline{\alpha}^{lmn} - \tfrac{1}{4}\overline{\alpha}^{jklm})\xi^j \xi^k \xi^l \xi^m.
\end{aligned}$$

From the early definition in Section 4.4, we know that

$$W_o = \hat{a} + \hat{b} x_o - a_o - b_o x_o = \overline{\epsilon} - (\overline{x} - x_o)(\hat{b} - b_o),$$

and $\xi^j = u_{j1} + u_{j2} x_o$ $j = 1,2$, which implies the following basic formulae:

$$\begin{aligned}
\xi^j \xi^k \gamma_{jk} &= \alpha^{-2}(x_o)\sigma^{-4}, \\
\overline{\alpha}^{jkl}\xi^j \xi^k \xi^l &= \mu_3 \sigma^{-6}\{1 + 3\frac{(\overline{x} - x_o)^2}{\sigma_x^2} - \frac{(\overline{x} - x_o)^3}{\sigma_x^6} m_3\}, \qquad (4.A.21) \\
\xi^j \xi^k n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x}) &= \sigma^{-4}\{\frac{(\overline{x} - x_o)^2}{\sigma_x^{-4}} m_3 - 2(\overline{x} - x_o)\}
\end{aligned}$$

We start with the calculation of k_{y_o1}. Notice that $E(R_{y_o}) = E(R_{y_o2}) + O(n^{-2})$, since $E(R_{y_o1}) = 0$. Using (4.A.21),

$$\begin{aligned}
E(R_{y_o2}) &= -\tfrac{1}{2}\alpha^3(x_o)\sigma\,\xi^j \xi^k E(A_o^{jk} W_o) + \tfrac{1}{3}\alpha^5(x_o)\sigma\,\overline{\alpha}^{jkl}\xi^j \xi^k \xi^l E(W_o^2), \\
&= -\tfrac{1}{2}\alpha^3(x_o)\sigma\,\xi^j \xi^k \{\gamma_{jk} - \frac{(\overline{x} - x_o)}{\sigma_x^2} n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x})\}\mu_3 n^{-1}
\end{aligned}$$

$$+\ \tfrac{1}{3}\,\alpha^5(x_o)\,\sigma\,\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\,\{1+\frac{(\overline{x}-x_o)^2}{\sigma_x^2}\}\,\sigma^2\,n^{-1}+O(n^{-2})$$

$$=-\ \tfrac{1}{6}\,\alpha^3(x_o)\,\sigma^{-3}\,\mu_3\,\{1+3\,\frac{(\overline{x}-x_o)^2}{\sigma_x^2}-\frac{(\overline{x}-x_o)^3}{\sigma_x^6}\,m_3\}\,n^{-1}+O(n^{-2})$$

$$=-\ \tfrac{1}{6}\,s_2^{\frac{1}{2}}\,n^{-1}+O(n^{-2}). \qquad (4.A.22)$$

Thus we have

$$k_{y_o1}=-\tfrac{1}{6}\,s_2^{\frac{1}{2}}\,n^{-\frac{1}{2}}+O(n^{-\frac{3}{2}}). \qquad (4.A.23)$$

For the convenience of computation, we calculate the third cumulant k_{y_o3} before k_{y_o2}. Observe that

$$k_{y_o3}=n^{\frac{3}{2}}\{E(R_{y_o}^3)-3\,E(R_{y_o})\,E(R_{y_o}^2)+2\,E^3(R_{y_o})\}$$

$$=n^{\frac{3}{2}}\{E(R_{y_o1}^3)+3\,E(R_{y_o1}^2\,R_{y_o2})-3\,E(R_{y_o2})\,E(R_{y_o1}^2)\}+O(n^{-\frac{3}{2}}).$$

We first have

$$E(R_{y_o1}^3)=\alpha^3(x_o)\,\sigma^{-3}\,\mu_3\,\{1+3\,\frac{(\overline{x}-x_o)^2}{\sigma_x^2}-\frac{(\overline{x}-x_o)^3}{\sigma_x^6}\,m_3\}\,n^{-2}+O(n^{-3})$$

$$=s_2^{\frac{1}{2}}\,n^{-2}+O(n^{-3}). \qquad (4.A.24)$$

Using the formulae in (4.A.21),

$$E(R_{y_o1}^2\,R_{y_o2})$$

$$=-\ \tfrac{1}{2}\,\alpha^5(x_o)\,\sigma^{-1}\,\xi^j\,\xi^k\,E(A_o^{jk}\,W_o^3)+\tfrac{1}{3}\,\alpha^7(x_o)\,\sigma^{-1}\,\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\,E(W_o^4)$$

$$=-\ \tfrac{3}{2}\,\alpha^3(x_o)\,\sigma\,\mu_3\,\xi^j\,\xi^k\,\{\gamma_{jk}-\frac{(\overline{x}-x_o)}{\sigma_x^2}\,n^{-1}\sum g_{jk}(x_i)(x_i-\overline{x})\}\,n^{-2}$$

$$+\ \alpha^3(x_o)\,\sigma^3\,\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\,n^{-2}+O(n^{-3})$$

$$=\alpha^3(x_o)\,\sigma^{-3}\,\mu_3\,\{-\tfrac{1}{2}-\tfrac{3}{2}\,\frac{(\overline{x}-x_o)^2}{\sigma_x^2}\,n^{-2}+\tfrac{1}{2}\,\frac{(\overline{x}-x_o)^3}{\sigma_x^6}\,m_3\}\,n^{-2}+O(n^{-3})$$

$$=-\ \tfrac{1}{2}\,s_2^{\frac{1}{2}}\,n^{-2}+O(n^{-3}). \qquad (4.A.25)$$

Moreover from (4.A.22), (4.A.24) , (4.A.25) and the fact that $E(R_{y_o1}^2)=n^{-1}$, we obtain

$$E(R_{y_o1}^3)+3\,E(R_{y_o1}^2\,R_{y_o2})=3\,E(R_{y_o2})\,E(R_{y_o1}^2).$$

Therefore,

$$k_{y_o3}=O(n^{-\frac{3}{2}}). \qquad (4.A.26)$$

Next we compute $k_{y_o 2}$. Note that $k_{y_o 2} = n\,\{E(R_b^2) - E^2(R_b)\}$ and

$$E(R_{y_o}^2) = E(R_{y_o 1})^2 + 2\,E(R_{y_o 1}\,R_{y_o 2}) + E(R_{y_o 2})^2 + 2\,E(R_{y_o 1}\,R_{y_o 3}) + O(n^{-3}). \tag{4.A.27}$$

We are going to compute each term appearing on the right-hand side of (4.A.27). Obviously $E(R_{y_o 1})^2 = n^{-1}$. Since

$$\begin{aligned} E(R_{y_o 1}\,R_{y_o 2}) &= -\tfrac{1}{2}\,\alpha^4(x_o)\,\xi^j\,\xi^k\,E(A_o^{jk}\,W_o^2) + \tfrac{1}{3}\,\alpha^6(x_o)\,\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\,E(W_o^3) \\ &= -\tfrac{1}{2}\,\alpha^4(x_o)\,(\mu_4 - \sigma^4)\,\xi^j\,\xi^k\,\{\gamma_{jk} - 2\,\frac{(\overline{x} - x_o)}{\sigma_x^2}\,n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x}) \\ &\quad + \frac{(\overline{x} - x_o)^2}{\sigma_x^4}\,n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x})^2\}\,n^{-2} \\ &\quad + \tfrac{1}{3}\,\alpha^6(x_o)\,\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\,\mu_3\,q_2\,n^{-2} + O(n^{-3}), \end{aligned}$$

then using (4.A.21) and the fact that

$$\xi^j\,\xi^k\,n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x})^2 = \sigma^{-4}\,\sigma_x^2\,\{1 - 2\,\frac{(\overline{x} - x_o)}{\sigma_x^4}\,m_3 + \frac{(\overline{x} - x_o)^2}{\sigma_x^6}\,m_4\},$$

we get

$$E(R_{y_o 1}\,R_{y_o 2}) = -\tfrac{1}{2}\,\frac{\alpha^4(x_o)\,(\mu_4 - \sigma^4)}{\sigma^4}\,q_1\,n^{-2} + \tfrac{1}{3}\,s_2\,n^{-2} + O(n^{-3}). \tag{4.A.28}$$

To calculate $E(R_{y_o 2})^2$, noticing that

$$\xi^j\,\xi^k\,\xi^m\,\xi^n\,\gamma_{jkmn} = \sigma^{-8}\,q_1,$$

and using (4.A.21) again, we obtain

$$\begin{aligned} E(R_{y_o 2})^2 &= \tfrac{1}{4}\,\alpha^6(x_o)\,\sigma^2\,\xi^j\,\xi^k\,\xi^m\,\xi^n\,E(A_o^{jk}\,A_o^{mn}\,W_o^2) \\ &\quad - \tfrac{1}{3}\,\alpha^8(x_o)\,\sigma^2\,\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l\,\xi^m\,\xi^n\,E(A_o^{mn}\,W_o^3) \\ &\quad + \tfrac{1}{9}\,\alpha^{10}(x_o)\,\sigma^2\,(\overline{\alpha}^{j\,k\,l}\,\xi^j\,\xi^k\,\xi^l)^2\,E(W_o^4) \\ &= \tfrac{1}{4}\,\alpha^4(x_o)\,\sigma^2\,(\mu_4 - \sigma^4)\,\xi^j\,\xi^k\,\xi^m\,\xi^n\,\gamma_{jkmn}\,n^{-2} \\ &\quad + \tfrac{1}{4}\,\alpha^6(x_o)\,\sigma^2\,\mu_3^2\,[2\,\xi^j\,\xi^k\,\xi^m\,\xi^n\,\gamma_{jk}\,\gamma_{mn} \\ &\quad - 4\,\sigma_x^{-2}\,(\overline{x} - x_o)\,\xi^j\,\xi^k\,\gamma_{jk}\xi^m\,\xi^n\,n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x}) \end{aligned}$$

$$+ 2\,\frac{(\overline{x} - x_o)^2}{\sigma_x^4}\,\{\xi^m\,\xi^n\,n^{-1}\sum g_{mn}(x_i)(x_i - \overline{x})\}^2]\,n^{-2}$$

$$- \alpha^6(x_o)\,\sigma^4\,\mu_3^2\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^l\,\xi^m\,\xi^n$$

$$\times \{\gamma_{mn} - \sigma_x^{-2}\,(\overline{x} - x_o)\,n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x})\}\,n^{-2}$$

$$+ \tfrac{1}{3}\,\alpha^6(x_o)\,\sigma^6\,(\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^l)^2\,n^{-2} + O(n^{-3})$$

$$= \alpha^4(x_o)\,\frac{(\mu_4 - \sigma^4)}{\sigma^4}\,q_1\,n^{-2} - \tfrac{1}{6}\,s_2\,n^{-2} + O(n^{-3}). \qquad (4.A.29)$$

Furthermore,

$$E(R_{y_o1}\,R_{y_o3}) = \alpha^4(x_o)\,\xi^j\,\xi^k\,E\{(\beta_{jk,1} - \gamma_{jk,1}\,x_o)\,(\hat{b} + b_1 - b_o)\,W_o^2\}$$

$$- \tfrac{1}{2}\,\alpha^4(x_o)\,\xi^j\,\xi^k\,(\beta_{jk2} - 2\,\beta_{jk}\,x_o + \gamma_{jk}\,x_o^2)\,E\{(\hat{b} + b_1 - b_o)^2\,W_o^2\}$$

$$- \alpha^6(x_o)\,\sigma^2\,\{\tfrac{1}{2}\,\sigma_x^{-2}\,\xi^j\,\xi^m\,(\beta_k - \gamma_l\,x_o)\,(\beta_n - \gamma_n\,x_o) + \tfrac{1}{8}\,\xi^j\,\xi^k\,\xi^m\,\xi^n\}$$

$$\times E(A_o^{jk}\,A_o^{mn}\,W_o^2) + C_2\,E(W_o^4)$$

$$+ \alpha^8(x_o)\,\sigma^2\,\{\sigma_x^{-2}\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^m\,(\beta_l - \gamma_l\,x_o)\,(\beta_n - \gamma_n\,x_o)$$

$$+ \tfrac{1}{6}\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^l\,\xi^m\,\xi^n - \alpha^{-2}(x_o)\,\sigma^{-2}\,\overline{\alpha}^{jkm}\,\xi^j\,\xi^k\,\xi^n\}\,E(A_o^{mn}\,W_o^3)$$

$$+ \tfrac{1}{2}\,\alpha^4(x_o)\,\xi^j\,\xi^k\,A_o^{jl}\,A_o^{kl}\,W_o^2 + \tfrac{1}{3}\,\alpha^6(x_o)\,\xi^j\,\xi^k\,\xi^l\,E(A_o^{jkl}\,W_o^3)$$

$$- \alpha^6(x_o)\,\sigma^2\,\xi^j\,\xi^k\,\xi^l\,(\beta_{jkl,2} - \gamma_{jkl,2}\,x_o)\,E\{(\hat{b} + b_1 - b_o)\,W_o^3\}. \qquad (4.A.30)$$

Each term on the right-hand side of (4.A.30) is computed below. Firstly, since

$$\xi^j\,\xi^k\,n^{-1}\sum g_{jk}(x_i)\,(x_i - \overline{x})(x_i - x_o) = \sigma^{-4}\,\sigma_x^2\,[1 - 2\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2} + \frac{(\overline{x} - x_o)^2}{\sigma_x^6}\,m_4$$

$$+ \{-2\,\frac{(\overline{x} - x_o)}{\sigma_x^4} + \frac{(\overline{x} - x_o)^3}{\sigma_x^6}\}\,m_3]$$

and

$$\xi^j\,\xi^k\,n^{-1}\sum g_{jk}(x_i)\,(x_i - x_o) = \sigma^{-4}\{-(\overline{x} - x_o) + \frac{(\overline{x} - x_o)^3}{\sigma_x^2} + \frac{(\overline{x} - x_o)^2}{\sigma_x^4}\,m_3\},$$

then we find the first term on the right-hand side of (4.A.30) to be

$$\alpha^4(x_o)\,\xi^j\,\xi^k\,E\{(\beta_{jk,1} - \gamma_{jk,1}\,x_o)\,(\hat{b} + b_1 - b_o)\,W_o^2\}$$

$$= \alpha^4(x_o)\,\sigma^2\,\sigma_x^{-2}\,\xi^j\,\xi^k\,\{n^{-1}\sum g_{jk}(x_i)\,(x_i - \overline{x})(x_i - x_o)$$

$$+ (\overline{x} - x_o)\,n^{-1}\sum g_{jk}(x_i)\,(x_i - x_o)\}n^{-2} + O(n^{-3})$$

$$= \alpha^4(x_o)\,q_3 n^{-2} + O(n^{-3}). \qquad (4.A.31)$$

Since

$$\xi^j\,\xi^k\,n^{-1}\sum g_{jk}(x_i)(x_i - x_o)^2 = \sigma^{-4}\,\sigma_x^2\,q_3,$$

then the second term has the form

$$
\begin{aligned}
&-\frac{1}{2}\,\alpha^4(x_o)\,\xi^j\,\xi^k\,(\beta_{jk2} - 2\,\beta_{jk}\,x_o + \gamma_{jk}\,x_o^2)\,E\{(\hat{b} + b_1 - b_o)^2\,W_o^2\}\\
&= -\tfrac{1}{2}\,\alpha^4(x_o)\,\sigma^4\,\sigma_x^{-2}\,\xi^j\,\xi^k\,n^{-1}\sum g_{jk}(x_i)(x_i - x_o)^2\,n^{-2} + O(n^{-3})\\
&= -\tfrac{1}{2}\,\alpha^4(x_o)\,q_3\,n^{-2} + O(n^{-3}).
\end{aligned}
\tag{4.A.32}
$$

To calculate the third term on the right-hand side of (4.A.30), note that

$$
\begin{aligned}
\xi^j\,(\beta_k - \gamma_l\,x_o)\,\gamma_{jk} &= 0,\\
\xi^j\,\xi^m\,(\beta_k - \gamma_l\,x_o)(\beta_n - \gamma_n\,x_o)\,\gamma_{jkmn} &= \sigma^{-8}\,\sigma_x^2\,q_3,\\
\xi^j\,\xi^k\,\xi^m\,\xi^n\,\gamma_{jkmn} &= \sigma^{-8}\,q_1,
\end{aligned}
$$

and

$$\xi^j\,(\beta_k - \gamma_l\,x_o)\,n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x}) = \sigma^{-4}\,\sigma^2\,\{1 - \frac{(\overline{x} - x_o)^2}{\sigma_x^2} - \frac{(\overline{x} - x_o)}{\sigma_x^4}\,m_3\}.$$

Using (4.A.21) again, we have

$$
\begin{aligned}
&\alpha^6(x_o)\,\sigma^2\,\{\tfrac{1}{2}\,\sigma_x^{-2}\,\xi^j\,\xi^m\,(\beta_k - \gamma_l\,x_o)(\beta_n - \gamma_n\,x_o) + \tfrac{1}{8}\,\xi^j\,\xi^k\,\xi^m\,\xi^n\}\,E(A_o^{jk}\,A_o^{mn}\,W_o^2)\\
&= \alpha^6(x_o)\,\sigma^2\left\{\tfrac{1}{2}\,\sigma_x^{-2}\,\xi^j\,\xi^m\,(\beta_k - \gamma_l\,x_o)(\beta_n - \gamma_n\,x_o) + \tfrac{1}{8}\,\xi^j\,\xi^k\,\xi^m\,\xi^n\right\}\\
&\times\left\{\sigma^2(\mu_4 - \sigma^4)\,\alpha^{-2}(x_o)\gamma_{jkmn} + 2\,\gamma_{jk}\,\gamma_{mn}\,\mu_3^2 - 4\,\frac{(\overline{x} - x_o)}{\sigma_x^2}\,\mu_3^2\,\gamma_{jk}\,n^{-1}\sum g_{mn}(x_i)(x_i - \overline{x})\right.\\
&\left.+ 2\,\frac{(\overline{x} - x_o)^2}{\sigma_x^4}\,\mu_3^2\,n^{-1}\sum g_{jk}(x_i)(x_i - \overline{x})\,n^{-1}\sum g_{mn}(x_i)(x_i - \overline{x})\right\}n^{-2} + O(n^{-3})\\
&= \alpha^4(x_o)\,(\mu_4 - \sigma^4)\,\sigma^{-4}(\tfrac{1}{2}\,q_3 + \tfrac{1}{8}\,q_1)\,n^{-2} + \tfrac{1}{4}\,s_2\,n^{-2} + \alpha^6(x_o)\,\mu_3^2\,\sigma^{-6}\,\frac{(\overline{x} - x_o)^2}{\sigma_x^2}\\
&\times\{1 - \frac{(\overline{x} - x_o)^2}{\sigma_x^2} - \frac{(\overline{x} - x_o)}{\sigma_x^2}\,n^{-1}\sum (x_i - \overline{x})^3\}\,n^{-2} + O(n^{-3}).
\end{aligned}
\tag{4.A.33}
$$

Moreover, because

$$\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,(\beta_l - \gamma_l\,x_o) = \mu_3\,\sigma^{-6}\,\{\frac{(\overline{x} - x_o)^3}{\sigma_x^2} - (\overline{x} - x_o) + \frac{(\overline{x} - x_o)^2}{\sigma_x^4}\,m_3\}$$

and

$$\overline{\alpha}^{jkm}\,\xi^j\,\xi^k\,\xi^m\,n^{-1}\sum g_{mn}(x_i)(x_i-\overline{x}) = \mu_3\,\sigma^{-8}\,(\overline{x}-x_o)\,\{-2-\alpha^2(x_o) + 3\,\frac{(\overline{x}-x_o)}{\sigma_x^4}\,m_3 - \frac{(\overline{x}-x_o)^2}{\sigma_x^8}\,m_3^2\},$$

the fourth term becomes

$$\begin{aligned}
&\alpha^8(x_o)\,\sigma^2\,\{\sigma_x^{-2}\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^m\,(\beta_l-\gamma_l\,x_o)(\beta_n-\gamma_n\,x_o)+\tfrac{1}{6}\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^l\,\xi^m\,\xi^n \\
&-\alpha^{-2}(x_o)\,\sigma^{-2}\,\overline{\alpha}^{jkm}\,\xi^j\,\xi^k\,\xi^n\}\,E(A_o^{mn}\,W_o^3) \\
&=3\,\alpha^6(x_o)\,\sigma^4\,\mu_3\{\sigma_x^{-2}\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^m\,(\beta_l-\gamma_l\,x_o)(\beta_n-\gamma_n\,x_o)+\tfrac{1}{6}\,\overline{\alpha}^{jkl}\,\xi^j\,\xi^k\,\xi^l\,\xi^m\,\xi^n \\
&-\alpha^{-2}(x_o)\,\sigma^{-2}\,\overline{\alpha}^{jkm}\,\xi^j\,\xi^k\,\xi^n\}\,\{\gamma_{mn}-\frac{(\overline{x}-x_o)}{\sigma_x^2}\,n^{-1}\sum g_{mn}(x_i)(x_i-\overline{x})\}n^{-2} \\
&+O(n^{-3}) \\
&=-\tfrac{5}{2}\,s_2\,n^{-2}+O(n^{-3}).
\end{aligned} \tag{4.A.34}$$

To get the next term, notice that

$$\begin{aligned}
\xi^j\,\xi^k\,\gamma_{jl}\,\gamma_{kl} &= \sigma^{-6}\,\alpha^{-2}(x_o), \\
\xi^j\,\xi^k\,\gamma_{jkll} &= \sigma^{-6}\,\{2+\frac{(\overline{x}-x_o)^2}{\sigma_x^2}-2\,\frac{(\overline{x}-x_o)}{\sigma_x^4}\,m_3+\frac{(\overline{x}-x_o)^2}{\sigma_x^6}\,m_4\}, \\
\xi^j\,\xi^k\,\gamma_{jl}\,n^{-1}\sum g_{kl}(x_i)(x_i-\overline{x}) &= \sigma^{-6}\,\{-2\,(\overline{x}-x_o)+\frac{(\overline{x}-x_o)}{\sigma_x^4}\,m_3\},
\end{aligned}$$

and

$$\xi^j\xi^k\,n^{-1}\sum g_{jl}(x_i)(x_i-\overline{x})\,n^{-1}\sum g_{kl}(x_i)(x_i-\overline{x}) = \sigma^{-6}\,[(\overline{x}-x_o)^2+\{\sigma_x-(\overline{x}-x_o)\,m_3\}^2].$$

Thus we have

$$\begin{aligned}
&\tfrac{1}{2}\,\alpha^4(x_o)\,\xi^j\,\xi^k\,A_o^{jl}\,A_o^{kl}\,W_o^2+\tfrac{1}{3}\,\alpha^6(x_o)\,\xi^j\,\xi^k\,\xi^l\,E(A_o^{jkl}\,W_o^3) \\
&=\tfrac{1}{2}\,\alpha^4(x_o)\,(\mu_4-\sigma^4)\,\sigma^2\,\xi^j\,\xi^k\,\gamma_{jkll}+\alpha^4(x_o)\,\mu_3^2\,\xi^j\,\xi^k\,\{\gamma_{jl}\,\gamma_{kl} \\
&-2\,\frac{(\overline{x}-x_o)}{\sigma_x^2}\,\gamma_{kl}\,n^{-1}\sum g_{jl}(x_i)(x_i-\overline{x})+\frac{(\overline{x}-x_o)^2}{\sigma_x^4}\,n^{-1}\sum g_{jl}(x_i)(x_i-\overline{x}) \\
&\times n^{-1}\sum g_{kl}(x_i)(x_i-\overline{x})\,n^{-2}+O(n^{-3}) \\
&=\tfrac{1}{2}\,\alpha^2(x_o)\,\frac{(\mu_4-\sigma^4)}{\sigma^6}\,\{2+\frac{(\overline{x}-x_o)^2}{\sigma_x^2}-2\,\frac{(\overline{x}-x_o)}{\sigma_x^4}\,m_3\,n^{-2}+s_2\,\}n^{-2}+O(n^{-3}) \\
&+\frac{\alpha^2(x_o)\,\mu_3^2}{\sigma^6}\,\frac{(\overline{x}-x_o)^2}{\sigma_x^2}\,\{1-\frac{(\overline{x}-x_o)^2}{\sigma_x^2}-\frac{(\overline{x}-x_o)}{\sigma_x^4}\,m_3\}^2\,n^{-2}+O(n^{-3}).
\end{aligned} \tag{4.A.35}$$

To calculate the seventh term in (4.A.30), note that

$$\tfrac{1}{3}\,\alpha^6(x_o)\,\xi^j\,\xi^k\,\xi^l\,E(A_o^{jkl}\,W_o^3)$$
$$=\alpha^4(x_o)\,\sigma^2\,\mu_4\,\xi^j\,\xi^k\,\xi^l\,\{\gamma_{jkl}-\frac{(\overline{x}-x_o)}{\sigma_x^2}\,n^{-1}\sum g_{jkl}(x_i)(x_i-\overline{x})\}\,n^{-2}+O(n^{-3}).$$

Since

$$\xi^j\,\xi^k\,\xi^l\,n^{-1}\sum g_{jkl}(x_i)(x_i-\overline{x})$$
$$=\sigma^{-6}\,\{-3\,(\overline{x}-x_o)+3\,\frac{(\overline{x}-x_o)^2}{\sigma_x^4}\,m_3-\frac{(\overline{x}-x_o)^3}{\sigma_x^6}\,n^{-1}\sum(x_i-\overline{x})^4\}$$

then using (4.A.21), we have

$$\tfrac{1}{3}\,\alpha^6(x_o)\,\xi^j\,\xi^k\,\xi^l\,E(A_o^{jkl}\,W_o^3)=s_1\,n^{-2}+O(n^{-3}). \qquad (4.A.36)$$

Because

$$E\{(\hat{b}+b_1-b_o)\,W_o^3\}=O(n^{-3}),$$

we immediately have

$$-\alpha^6(x_o)\,\sigma^2\,\xi^j\,\xi^k\,\xi^l\,(\beta_{jkl,2}-\gamma_{jkl,2}\,x_o)\,E\{(\hat{b}+b_1-b_o)\,W_o^3\}=O(n^{-3}). \qquad (4.A.37)$$

Finally it can be shown that

$$C_2=\alpha^4(x_o)\,\sigma^{-4}\,(\tfrac{4}{9}\,s_2-\tfrac{1}{4}\,s_1)\,n^{-2}+O(n^{-3}).$$

Thus,

$$C_2\,E(W_o^4)=(\tfrac{4}{3}\,s_2-\tfrac{3}{4}\,s_1)\,n^{-2}+O(n^{-3}). \qquad (4.A.38)$$

Substituting (4.A.31)-(4.A.38) into (4.A.30), we end up with

$$E(R_{y_o1}\,R_{y_o3})=\tfrac{1}{2}\,\alpha^4(x_o)\,q_3\,n^{-2}+\tfrac{3}{8}\,\frac{\alpha^4(x_o)\,(\mu_4-\sigma^4)}{\sigma^4}\,q_1\,n^{-2}-\tfrac{5}{12}\,s_2\,n^{-2}$$
$$+\tfrac{1}{4}\,\frac{\alpha^4(x_o)\,\mu_4}{\sigma^4}\,q_1\,n^{-2}+O(n^{-3}). \qquad (4.A.39)$$

Again substituting (4.A.28), (4.A.29) and (4.A.39) into (4.A.27), we obtain that

$$E(R_{y_o}^2)=E(R_{y_o1})^2+2\,E(R_{y_o1}\,R_{y_o2})+E(R_{y_o2})^2+2\,E(R_{y_o1}\,R_{y_o3})$$
$$=n^{-1}+(\tfrac{1}{2}\,s_1-\tfrac{1}{3}\,s_2+s_3)\,n^{-2}+O(n^{-3}).$$

Therefore,

$$k_{y_o 2} = 1 + (\tfrac{1}{2} s_1 - \tfrac{13}{36} s_2 + s_3) n^{-1} + O(n^{-2}). \tag{4.A.40}$$

Finally, we derive the fourth cumulant of $n^{\frac{1}{2}} R_{y_o}$. By definition,

$$\begin{aligned} k_{y_o 4} &= n^2 \{E(R_{y_o}^4) - 3 E^2(R_{y_o}^2) - 4 E(R_{y_o}) E(R_{y_o}^3) \\ &\quad + 12 E^2(R_{y_o}) E(R_{y_o}^2) - 6 E^4(R_{y_o})\} \\ &= n^{-2} \{E(R_{y_o}^4) - 3 E^2(R_{y_o}^2)\} - 4 n^{\frac{1}{2}} E(R_{y_o}) k_{y_o 3} + 2 k_{y_o 1}^4 \\ &= n^{-2} \{E(R_{y_o}^4) - 3 E^2(R_{y_o}^2)\} + O(n^{-2}). \\ &= n^{-2} \{E(R_{y_o 1}^4) - 3 E^2(R_{y_o 1}^2) + 4 E(R_{y_o 1}^3 R_{y_o 2}) - 12 E(R_{y_o 1}^2) E(R_{y_o 1} R_{y_o 2}) \\ &\quad + 6 E(R_{y_o 1}^2 R_{y_o 2}^2) - 6 E(R_{y_o 1}^2) E(R_{y_o 2}^2) + 4 E(R_{y_o 1}^3 R_{y_o 3}) \\ &\quad - 12 E(R_{y_o 1}^2) E(R_{y_o 1} R_{y_o 3}). \end{aligned}$$

Using these formulae in (4.A.21) and neglecting terms of order of n^{-4}, we may show that

$$E(R_{y_o 1}^4) - 3 E^2(R_{y_o 1}^2) = \alpha^4(x_o)(\mu_4 - 3\sigma^4)\sigma^{-4} q_1 n^{-3},$$

$$4 E(R_{y_o 1}^3 R_{y_o 2}) - 12 E(R_{y_o 1}^2) E(R_{y_o 1} R_{y_o 2}) = \{-6 \alpha^4(x_o)(\mu_4 - \sigma^4)\sigma^{-4} q_1 + \tfrac{4}{3} s_2\} n^{-3},$$

$$6 E(R_{y_o 1}^2 R_{y_o 2}^2) - 6 E(R_{y_o 1}^2) E(R_{y_o 2}^2) = \{3 \alpha^4(x_o)(\mu_4 - \sigma^4)\sigma^{-4} q_1 - s_2\} n^{-3}.$$

By the early formulae used to derive $E(R_{y_o 1} R_{y_o 3})$, we may show that

$$4 E(R_{y_o 1}^3 R_{y_o 3}) - 12 E(R_{y_o 1}^2) E(R_{y_o 1} R_{y_o 3}) = (2 s_1 - \tfrac{1}{3} s_2) n^{-3} + O(n^{-4}).$$

Thus we have

$$k_{y_o 4} = O(n^{-2}). \tag{4.A.41}$$

Based on the results given by James and Mayne (1964), we have

$$k_{y_o, j} = O(n^{-\frac{3}{2}}), \quad j \geq 5. \tag{4.A.42}$$

Hence, in summary of (4.A.23), (4.A.26), (4.A.40) and (4.A.41), we have proved (4.6.8).

CHAPTER 5
COMPARING EMPIRICAL LIKELIHOOD AND BOOTSTRAP HYPOTHESIS TESTS

5.1 Introduction

From the reviews on developments of empirical likelihood given in Chapter 1 and the work discussed in Chapter 2, 3 and 4, we see that almost all the research done on empirical likelihood concentrate on constructing confidence regions. After constructing an empirical likelihood confidence region, we can derive an empirical likelihood test about the parameter of interested by using the duality between confidence regions and hypothesis tests. However, so far little has been done on the aspect of power of empirical likelihood tests. Surprisingly, little has been done on that of a bootstrap test either! The contribution of this chapter is to develope high-order expansions for the power function of empirical likelihood and bootstrap tests for a mean against a series of local alternatives. A comparison between empirical likelihood and bootstrap tests for a mean parameter against a series of local alternative hypotheses is made. For univariate and bivariate cases, practical rules are proposed for choosing the more powerful test.

Let $X_1, \cdots, X_n$ be an independent and identically distributed (i.i.d.) random sample of p-dimensions from an unknown distribution with mean parameter μ and covariance matrix Σ. We consider using empirical likelihood and bootstrap methods to test the null hypothesis $H_o : \mu = \mu_o$ against a series of local alternatives H_n : $\mu = \mu_o + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau$, ,where both μ_o and τ are constant p dimensional vectors. The empirical likelihood and the bootstrap hypothesis tests for H_o can be formulated from the well-known duality between confidence regions and hypothesis tests.

Owen (1990) showed in the i.i.d. sample mean case that the power of an α-level empirical log-likelihood ratio test is asymptotically $P\{\chi_p^2(\|\tau\|^2) > \chi^2_{p,1-\alpha}\}$, where $\chi_p^2(\|\tau\|^2)$ is the noncentral chi-squared random variable with p degrees of freedom and noncentrality parameter $\|\tau\|^2$, and $\chi^2_{p,1-\alpha}$ is the $1-\alpha$ upper percentile of the central chi-squared distribution χ_p^2 . However, it is not difficult to show that the corresponding bootstrap test also achieves the same asymptotic power. Thus, to really compare powers of the empirical likelihood and the bootstrap tests, we have to develop higher-order expansions for the powers of these two tests, which will give us some insight into the problem.

In Section 5.2 we define the empirical likelihood and the bootstrap tests. After developing higher-order expansions for the power functions in Section 5.3, we propose in Section 5.4 two rules for practically choosing between the empirical likelihood and the bootstrap tests for univariate and bivariate cases. In the univariate case, the rule says that the empirical likelihood test is more powerful than the corresponding bootstrap test when $\tau\alpha_3 > 0$, and vice versa when $\tau\alpha_3 < 0$, where α_3 is the population skewness parameter. For higher dimensional cases, similar rules may be developed. In Section 5.5 we present simulation studies. We display our calculations of cumulants in Appendix 5.

5.2. Empirical Likelihood and Bootstrap Hypothesis Tests

Let $X_1, \cdots, X_n$ be a p dimension i.i.d. sample from unknown distribution F with mean μ and covariance matrix Σ. We want to test null hypothesis $H_o : \mu = \mu_o$ against a series of local alternatives $H_n : \mu = \mu_o + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau$, where μ_o and τ are p dimension constant vectors.

Put $Z_i = \Sigma^{-\frac{1}{2}}(X_i - \mu)$ and let Z_i^j be the j'th component of Z_i. We define

$$\alpha^{j_1 j_2 \cdots j_k} = E\left(Z_i^{j_1} \cdots Z_i^{j_k}\right),$$

$$A^{j_1 j_2 \cdots j_k} = n^{-1} \sum Z_i^{j_1} \cdots Z_i^{j_k} - \alpha^{j_1 j_2 \cdots j_k},$$

as the standardized multivariate moments of $X_1, \cdots, X_n$. Note that $\alpha^j = 0$ and $\alpha^{jk} = \delta^{jk}$ where δ^{jk} is the Kronecker delta. Throughout this paper we assume the following regularity condition:

(i) $\Sigma = Cov(X_1)$ is a positive definite matrix; (ii) $E\|X_1\|^{15} < \infty$;

(iii) the characteristic function g_1 of X_1 satisties Cramér's condition, (5.2.1)

for every positive b, $\sup_{\|t\|>b} |g_1(t)| < 1$.

5.2.1 Empirical Likelihood Tests

Write $p_1, p_2, \cdots, p_n$ for nonnegative numbers adding to unity. Then, the empirical log-likelihood ratio for μ is defined to be

$$\ell(\mu) = -2 \min_{\sum p_i X_i = \mu} \sum_{i=1}^{n} log\,(np_i)\,.$$

Based on the nonparametric version of Wilks' theorem given by Owen (1990), a $1-\alpha$ level confidence region for μ is defined as $I_{1-\alpha} = \{\mu | \ell(\mu) < c_\alpha\}$, where c_α is chosen from the χ_p^2 tables such that $P(\chi_p^2 > c_\alpha) = \alpha$. According to the duality between confidence regions and hypothesis tests, we define an α-level empirical likelihood test for the null hypothesis H_0 to be

$$\phi_e = \begin{cases} 1, & \text{if } \ell(\mu_0) > c_\alpha; \\ 0, & \text{otherwise.} \end{cases}$$

By Wilks' theorem the asymptotic significant level of ϕ_e is α. Let us define "type I accuracy" as the difference between the actual and nominal significant levels of a test. Using the results given by Hall and La Scala (1990), we may show that

$$P(\phi_e = 1 | H_0) = \alpha + O(n^{-1}),$$

which means that type I accuracy of empirical likelihood test ϕ_e is of order n^{-1}. Since empirical likelihood confidence regions are Bartlett correctable in this case, as shown by DiCiccio, Hall and Romano (1991), we may define the Bartlett-corrected empirical likelihood test to be

$$\phi_{ec} = \begin{cases} 1, & \text{if } \ell(\mu_0) > c_\alpha(1 + \hat{\beta}/n); \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\hat{\beta} = p^{-1}\left(\hat{\alpha}^{jjkk}/2 - \hat{\alpha}^{jkl}\hat{\alpha}^{jkl}/3\right)$$

is the empirical Bartlett factor, and $\hat{\alpha}^{jjkk}$ and $\hat{\alpha}^{jkl}$ are the usual moment estimates of α^{jjkk} and α^{jkl} respectively. Let

$$\beta = p^{-1}\left(\alpha^{jjkk}/2 - \alpha^{jkl}\alpha^{jkl}/3\right)$$

be the theoretical Bartlett factor. Clearly we have $\hat{\beta} = \beta + O_p(n^{-\frac{1}{2}})$. Since

$$P(\phi_{ec} = 1|H_0) = \alpha + O(n^{-2}),$$

type I accuracy of the corrected empirical likelihood test ϕ_{ec} is of order n^{-2}, which is the same order as that of bootstrap test as will be shown shortly.

5.2.2 Bootstrap Test

Let $\overline{x} = n^{-1}\sum X_i$ and $\hat{\Sigma} = n^{-1}\sum(X_i - \overline{x})(X_i - \overline{x})^T$ be the sample mean and sample covariance matrix respectively. To give an α-level bootstrap test of H_o, let $\overline{x}^*$ and $\hat{\Sigma}^*$ be the bootstrap version of $\overline{x}$ and $\hat{\Sigma}$ respectively, computed from a resample χ^* instead of the entire sample $\chi = \{X_1, \cdots, X_n\}$. Put

$$S(\tau) = n^{\frac{1}{2}}\hat{\Sigma}^{-\frac{1}{2}}(\overline{x} - \mu + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau).$$

We define a bootstrap test of H_o to be

$$\phi_b = \begin{cases} 1, & \text{if } S^T(\tau)\,S(\tau) > \hat{c}_\alpha; \\ 0, & \text{otherwise,} \end{cases}$$

where $\hat{c}_\alpha$ is determined by

$$P\{n\,(\overline{x}^* - \overline{x})^T\,\hat{\Sigma}^{*-1}(\overline{x}^* - \overline{x}) > \hat{c}_\alpha|\,\chi\} = \alpha$$

and can be empirically calculated by Monte Carlo simulations. It has been pointed out by Hall (1992) that

$$P(\phi_b = 1|H_0) = \alpha + O(n^{-2}),$$

which means type I accuracy of the bootstrap test ϕ_b is of order n^{-2}

Owen (1990) showed that for our current null and alternative hypothesis setting, the power of the uncorrected empirical likelihood test ϕ_e (also the corrected empirical likelihood test ϕ_{ec} as shown in Section 5.3) converges to $P\{\chi^2_p(\|\tau\|^2) > \chi^2_{p,1-\alpha}\}$, where $\chi^2_p(\|\tau\|^2)$ is the noncentral chi-squared random variable with noncentral term $\|\tau\|^2$. It is not difficult to show that the bootstrap test achieves the same asymptotic power as well. In order to compare the power performances of these tests we have to find higher-order expansions for the power functions of the empirical likelihood and bootstrap tests. To make the comparison fairly, we should only compare the corrected empirical likelihood test ϕ_{ec} with the bootstrap test ϕ_b, since both have the same type I accuracy of order n^{-2}. In theory we could adjust the test's level so that they are exactly equal. However, from a practical point of view a difference of order n^{-2} between the levels of the tests is fair enough to make our comparison. In the rest of this paper, when we say the empirical likelihood test we mean the Bartlett-corrected test ϕ_{ec}.

Before we finish this section, we should mention that the shape of the rejection region of the empirical likelihood test is determined automatically by the sample itself, whereas that of the bootstrap test is subjectively given by us as the complement of an elliptical region. This is an advantage of empirical likelihood over the bootstrap.

5.3 Power Expansions

In this section we calculate the powers of the empirical likelihood and bootstrap tests of null hypothesis $H_o : \mu = \mu_o$ against $H_n : \mu = \mu_o + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau$. Since analytic expressions for these power functions are difficult to obtain, we have to develop expansions for them.

Let $pow(\phi_{ec};\tau)$ and $pow(\phi_b;\tau)$ denote the powers of the α-level empirical likelihood tests ϕ_{ec} and the bootstrap test ϕ_b respectively, under the alternative hypothesis H_n. We shall calculate them one by one.

5.3.1 Power of ϕ_{ec}

According to the definition of power of a test, we have

$$\begin{aligned} pow(\phi_{ec};\tau) =& P(\phi_{ec} = 1|H_n) \\ =& P\{\ell(\mu_0) > \tilde{c}_\alpha | \mu = \mu_0 + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau\} \\ =& P\{\ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau) > \tilde{c}_\alpha\} \end{aligned}$$

where $\tilde{c}_\alpha = c_\alpha(1+\hat{\beta}/n)$. To calculate $pow(\phi_{ec};\tau)$ we first set up a Taylor expansion for $\ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau)$, from which an Edgeworth expansion of $pow(\phi_{ec};\tau)$ will be derived. By the definition of empirical likelihood,

$$\begin{aligned} \ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau) =& -2 \min_{\sum p_i X_i = \mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau} \sum \log(np_i) \\ =& -2 \min_{\sum p_i Z_i = -n^{-\frac{1}{2}}\tau} \sum \log(np_i) \end{aligned}$$

where $Z_i = \Sigma^{\frac{1}{2}}(X_i - \mu)$. Slightly modifying (3.7) of DiCiccio, Hall and Romano (1988), we have

$$\begin{aligned} & n^{-1}\ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau) \\ =& (A + n^{-\frac{1}{2}}\tau)^j(A + n^{-\frac{1}{2}}\tau)^j \\ & - \{A^{jk} + n^{-\frac{1}{2}}\tau^j A^k[2] + n^{-1}\tau^j\tau^k\}(A + n^{-\frac{1}{2}}\tau)^j(A + n^{-\frac{1}{2}}\tau)^k \\ & + \tfrac{2}{3}(\alpha^{jkl} + A^{jkl} + n^{-\frac{1}{2}}\tau^j\delta^{kl}[3] - 2\,\alpha^{jkm}A^{lm})(A + n^{-\frac{1}{2}}\tau)^j(A + n^{-\frac{1}{2}}\tau)^k(A + n^{-\frac{1}{2}}\tau)^l \\ & + (\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{2}\alpha^{jklm})(A + n^{-\frac{1}{2}}\tau)^j(A + n^{-\frac{1}{2}}\tau)^k(A + n^{-\frac{1}{2}}\tau)^l(A + n^{-\frac{1}{2}}\tau)^m \\ & + A^{jl}A^{kl}(A + n^{-\frac{1}{2}}\tau)^j(A + n^{-\frac{1}{2}}\tau)^k + O_p(n^{-\frac{5}{2}}), \end{aligned} \tag{5.3.1}$$

with $\tau^j A^k[2] = \tau^j A^k + \tau^k A^j$ and the same rule applies for $\tau^j\delta^{kl}[3]$. From (5.3.1) we can derive the following signed root decomposition for $\ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau)$:

$$\ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau) = n\{R_1(\tau) + R_2(\tau) + R_3(\tau)\}^T\{R_1(\tau) + R_2(\tau) + R_3(\tau)\} + O_p(n^{-\frac{5}{2}}),$$

where $R_l(\tau) = O_p(n^{-l/2})$ for $l = 1, 2, 3$, and

$$R_1^j(\tau) = (A + n^{-\frac{1}{2}}\tau)^j,$$

$$R_2^j(\tau) = -\tfrac{1}{2}A^{jk}(A+n^{-\frac{1}{2}}\tau)^k + \tfrac{1}{3}\alpha^{jkl}(A+n^{-\frac{1}{2}}\tau)^k(A+n^{-\frac{1}{2}}\tau)^l,$$

$$\begin{aligned} R_3^j(\tau) = & (-\tfrac{1}{2}\tau^j\tau^k n^{-1} + \tfrac{3}{8}A^{jl}A^{kl} - \tfrac{1}{2}\tau^j A^k[2]n^{-\frac{1}{2}})(A+n^{-\frac{1}{2}}\tau)^k \\ & + \{\tfrac{1}{3}(A^{jkl} + n^{-\frac{1}{2}}\tau^j\delta^{kl}[3]) - \tfrac{5}{6}\alpha^{jkm}A^{lm}\}(A+n^{-\frac{1}{2}}\tau)^j(A+n^{-\frac{1}{2}}\tau)^k \\ & + (\tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})(A+n^{-\frac{1}{2}}\tau)^k(A+n^{-\frac{1}{2}}\tau)^l(A+n^{-\frac{1}{2}}\tau)^m. \end{aligned} \tag{5.3.2}$$

Put

$$R(\tau) = R_1(\tau) + R_2(\tau) + R_3(\tau).$$

Let $k_l^{j_1,\cdots,j_l}$ denote the joint l'th order cumulant of $n^{\frac{1}{2}}R(\tau)$. Calculations deferred to Appendix 5 show that

$$\begin{aligned} k_1^j &= \tau^j + k_{11}^j n^{-\frac{1}{2}} + k_{12}^j n^{-1} + O(n^{-\frac{3}{2}}), \\ k_2^{jk} &= \delta^{jk} + k_{21}^{jk} n^{-\frac{1}{2}} + k_{22}^{jk} n^{-1} + O(n^{-\frac{3}{2}}), \\ k_3^{jkl} &= k_{32}^{jkl} n^{-1} + O(n^{-\frac{3}{2}}), \\ k_4^{jklm} &= O(n^{-\frac{3}{2}}), \\ k_l^{j_1,\cdots,j_l} &= O(n^{-\frac{3}{2}}), \quad \text{for } l \geq 5, \end{aligned} \tag{5.3.3}$$

where

$$\begin{aligned} k_{11}^j &= (\tfrac{1}{3}\alpha^{jlm}\tau^l\tau^m - \tfrac{1}{6}\alpha^{jkk}), \quad k_{21}^{jk} = \tfrac{1}{3}\alpha^{jkl}\tau^l, \\ k_{12}^j &= \tfrac{1}{2}\tau^j\tau^k\tau^k + \tfrac{7}{24}\tau^k\alpha^{jkll} + (\tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})\tau^k\tau^l\tau^m \\ & \quad + (-\tfrac{p}{6} - \tfrac{5}{24})\tau^j - \tfrac{7}{18}\alpha^{jkm}\alpha^{mll}\tau^k + \tfrac{1}{18}\alpha^{jkm}\alpha^{klm}\tau^l, \\ k_{22}^{jk} &= \tfrac{1}{2}\alpha^{jkmm} - \tfrac{1}{3}\alpha^{jml}\alpha^{kml} - \tfrac{1}{36}\alpha^{jkm}\alpha^{mll} + \tfrac{1}{3}\delta^{jk}\tau^l\tau^l + \tfrac{5}{12}\tau^j\tau^k \\ & \quad + (\tfrac{8}{9}\alpha^{jkm}\alpha^{lmn} - \tfrac{1}{9}\alpha^{jmn}\alpha^{klm} - \tfrac{7}{12}\alpha^{jkln})\tau^l\tau^n, \\ k_{32}^{jkl} &= -\tfrac{1}{2}\alpha^{jklm}\tau^m + \tfrac{5}{36}\tau^n\alpha^{jmn}\alpha^{klm}[3]. \end{aligned} \tag{5.3.4}$$

Note that the last result in (5.3.3) is obtained from the general results given by James and Mayne (1962).

Let ϕ be the density of $N(0, I_p)$, $H_l(v_{j_1}, \cdots, v_{j_l})$ be the l'th order multivariate Chebyshev-Hermite polynomials defined by Barndorff-Nielson and Cox (1989), and

$$\mathcal{D}_\tau(x) = \{v \mid \|v + \tau\| > x\}.$$

Then, we define

$$E_2(x,\tau) = \int_{\mathcal{D}_\tau(x)} \{k_{11}^j v_j + \tfrac{1}{2}k_{21}^{jk}(v_j v_k - \delta^{jk})\}\phi(v)dv \tag{5.3.5}$$

and

$$\begin{aligned} E_3(x,\tau) = &\int_{\mathcal{D}_\tau(x)} \{k_{12}^j v_j + \tfrac{1}{2}(k_{22}^{jk} + k_{11}^j k_{11}^k)(v_j v_k - \delta^{jk})\}\phi(v)dv \\ &+ \int_{\mathcal{D}_\tau(x)} (\tfrac{1}{6}k_{32}^{jkl} + \tfrac{1}{2}k_{11}^j k_{21}^{kl})H_3(v_j, v_k, v_l)\phi(v)dv, \\ &+ \int_{\mathcal{D}_\tau(x)} \tfrac{1}{8}\, k_{21}^{jk} k_{21}^{lm}\, H_4(v_j, v_k, v_l, v_m)\phi(v)dv. \end{aligned}$$

Put $\overline{U} = (A^1, \cdots, A^p, A^{11}, \cdots, A^{pp}, A^{111}, \cdots, A^{ppp})^T$. Note that only A^{jk} and A^{jkl} with $j \le k \le l$ appear in $\overline{U}$. With above preparations we are able to prove the following theorem which will lead to an Edgeworth expansion for $pow(\phi_{ec};\tau)$. The proof of Theorem 5.3.1 is deferred to Section 5.6.

Theorem 5.3.1 *Assume condition (5.2.1). Then, for any $x > 0$,*

$$P\{\ell(\mu - n^{-\frac{1}{2}}\Sigma^{1/2}\tau) > x\} = P\{\chi_p^2(\|\tau\|) > x\} + E_2(x,\tau)\,n^{-\frac{1}{2}} + E_3(x,\tau)\,n^{-1} + O(n^{-\frac{3}{2}}).$$

From Theorem 5.3.1 and using the delta method we have

$$\begin{aligned} pow(\phi_{ec};\tau) =& P\{\ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau) > c_\alpha(1 + \hat{\beta}/n)\} \\ =& P\{\ell(\mu - n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau) > c_\alpha(1 + \beta/n)\} + O(n^{-\frac{3}{2}}) \\ =& P\{\chi_p^2(\|\tau\|) > c_\alpha\} + E_2(c_\alpha,\tau)\,n^{-\frac{1}{2}} \\ &+ \{E_3(c_\alpha,\tau) - \beta\, g_{p\tau}(c_\alpha)\}\, n^{-1} + O(n^{-\frac{3}{2}}), \end{aligned} \tag{5.3.7}$$

where $g_{p\tau}$ is the density of the $\chi_p^2(\|\tau\|)$ distribution and

$$\beta = p^{-1}\left(\alpha^{jjkk}/2 - \alpha^{jkl}\alpha^{jkl}/3\right)$$

is the Bartlett factor.

Substituting k_{11}^j and k_{21}^j in (5.3.4) into the expression for E_2, we obtain

$$E_2(c_\alpha,\tau) = \int_{\mathcal{D}_\tau(c_\alpha)} (\tfrac{1}{3}\alpha^{jlm}\tau^l\tau^m - \tfrac{1}{6}\alpha^{jkk})\, v_j + \tfrac{1}{6}\,\alpha^{jkl}\tau^l(v_j v_k - \delta^{jk})\}\phi(v)dv. \tag{5.3.8}$$

Thus the second order term of the power of the empirical likelihood test depends on population skewness parameter α^{jlm}, on τ and on sample size n.

Remark: If $X_1, \cdots, X_n$ are independent but not identically distributed with the same mean parameter μ, we can modify the definitions of Σ and $\alpha^{j_1 j_2 \cdots j_k}$ by $\Sigma = n^{-1} \sum \mathrm{Cov}(X_i)$ and $\alpha^{j_1 j_2 \cdots j_k} = n^{-1} \sum E\left(Z_i^{j_1} \cdots Z_i^{j_k}\right)$, and replace condition (5.2.1) with the following condition:

(i) Let v_{pn} and v_{1n} be the smallest and largest eigenvalues of Σ.

There exist positive constants C_1, C_2 such that, uniformly in n,

$C_1 \le v_{pn} \le v_{1n} \le C_2$. (ii) $\sup\limits_n n^{-1} \sum\limits_{j=1}^{n} E\|X_j\|^{15} < \infty$. (iii) for every

positive τ, $\lim\limits_{n \to \infty} n^{-1} \sum\limits_{j=1}^{n} \int_{\|X_j\| > \tau n^{\frac{1}{2}}} \|X_j\|^{15} = 0$. (v) The characteristic

function g_j of X_j satisties the Cramér's condition, for every positive b,

$\limsup\limits_{j \to \infty} \sup\limits_{\|t\| > b} |g_j(t)| < 1.$

Then, we may have Theorem 5.3.1 and (5.3.7) for this non-i.i.d. case by developing Edgeworth expansions using Theorems 1.3.3 and 1.3.4. This means that we can calculate higher-order expansions of the power of empirical likelihood test for the regression coefficient vector β considered in Chapter 3, and for the slope parameter b_o and means y_o of a simple linear regression model considered in Chapter 4.

5.3.2 Power of ϕ_b

In this subsection we give an expansion for the power of the bootstrap test ϕ_b as we have done for the empirical likelihood test in Subsection 3.1. According to our definition,

$$\begin{aligned} pow(\phi_b; \tau) &= P(\phi_b = 1 | H_n) \\ &= P\{S^T(\tau) S(\tau) > \hat{c}_\alpha\}, \end{aligned} \tag{5.3.9}$$

where $\hat{c}_\alpha$ is determined by equation

$$P\{n\,(\overline{x}^* - \overline{x})^T \hat{\Sigma}^{*-1} (\overline{x}^* - \overline{x}) > \hat{c}_\alpha | \chi\} = \alpha$$

and $S(\tau) = n^{\frac{1}{2}}\hat{\Sigma}^{-\frac{1}{2}}(\overline{x} - \mu + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau)$. Since

$$\hat{\Sigma} = n^{-1}\sum_{i=1}^{n} X_i X_i^T - \overline{X}\,\overline{X}^T = \Sigma^{\frac{1}{2}}(I + \Delta_1 - \Delta_2)\Sigma^{\frac{1}{2}}$$

where $\Delta_1 = (\Delta^1_{jk})_{p\times p} = (A^{jk})_{p\times p}$ and $\Delta_2 = (\Delta^2_{jk})_{p\times p} = (A^j A^k)_{p\times p}$, we have

$$\hat{\Sigma}^{-1} = \Sigma^{-\frac{1}{2}}\{I + \Delta_1 - \Delta_2 - (\Delta_1 - \Delta_2)^2\}\Sigma^{-\frac{1}{2}} + \Delta_3,$$

where $\Delta_3 = (\Delta^3_{jk})_{p\times p}$ is a $p \times p$ matrix with each of its elements $\Delta^3_{jk} = O_p(n^{-\frac{3}{2}})$.

It can be show that

$$\{I + \Delta_1 - \Delta_2 - (\Delta_1 - \Delta_2)^2\}^{-\frac{1}{2}} = I - \tfrac{1}{2}\Delta_1 + \tfrac{1}{2}\Delta_2 + \tfrac{3}{8}\Delta_1^2 + O_p(n^{-\frac{3}{2}}).$$

Put

$$S_o(\tau) = (I - \tfrac{1}{2}\Delta_1 + \tfrac{1}{2}\Delta_2 + \tfrac{3}{8}\Delta_1^2)(n^{\frac{1}{2}}A + \tau). \tag{5.3.10}$$

Thus

$$\begin{aligned} S^T(\tau)S(\tau) &= n(\overline{X} - \mu + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau)^T\hat{\Sigma}^{-1}(\overline{X} - \mu + n^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}\tau) \\ &= (n^{\frac{1}{2}}A + \tau)^T\{I + \Delta_1 - \Delta_2 - (\Delta_1 - \Delta_2)^2\}(n^{\frac{1}{2}}A + \tau) + O_p(n^{-\frac{3}{2}}) \\ &= S_o^T(\tau)S_o(\tau) + O_p(n^{-\frac{3}{2}}) \end{aligned}$$

where $A = (A^1, \cdots, A^p)^T = \Sigma^{-\frac{1}{2}}(\overline{X} - \mu)$. Using the delta method,

$$\begin{aligned} pow(\phi_b;\tau) &= P\{S^T(\tau)S(\tau) > \hat{c}_\alpha\}, \\ &= P\{S_o^T(\tau)S_o(\tau) > \hat{c}_\alpha\} + O(n^{-\frac{3}{2}}). \end{aligned} \tag{5.3.11}$$

Let $\xi_l^{j_1,\cdots,j_l}$ denote the joint l'th order cumulant of $S_o(\tau)$. Calculations presented in Appendix 5 reveal that

$$\begin{aligned} \xi_1^j &= \tau^j + \xi_{11}^j n^{-\frac{1}{2}} + \xi_{12}^j n^{-1} + O(n^{-\frac{3}{2}}), \\ \xi_2^{jk} &= \delta^{jk} + \xi_{21}^{jk} n^{-\frac{1}{2}} + \xi_{22}^{jk} n^{-1} + O(n^{-\frac{3}{2}}), \\ \xi_3^{jkl} &= \xi_{31}^{jkl} n^{-\frac{1}{2}} + \xi_{32}^{jkl} n^{-1} + O(n^{-\frac{3}{2}}), \\ \xi_4^{jklm} &= \xi_4^{jklm} n^{-1} + O(n^{-\frac{3}{2}}), \\ \xi_l^{j_1,\cdots,j_l} &= O(n^{-\frac{3}{2}}), \quad \text{for} \quad l \geq 5. \end{aligned} \tag{5.3.12}$$

where

$$\xi_{11}^{j} = -\tfrac{1}{2}\alpha^{jkk},\ \xi_{12}^{j} = \tfrac{1}{2}\tau^{j} + \tfrac{3}{8}\tau^{l}(\alpha^{jkkl} - \delta^{jl}),$$
$$\xi_{21} = -\alpha^{jkm}\,\tau^{m},$$
$$\xi_{22}^{jk} = (p+2)\delta^{jk} + \alpha^{jlm}\,\alpha^{klm} + \tfrac{3}{4}\alpha^{jkm}\,\alpha^{llm} + \tfrac{1}{4}(\alpha^{jklm} - \delta^{jk})\tau^{l}, \tag{5.3.13}$$
$$\xi_{31}^{jkl} = -2\alpha^{jkl},\ \xi_{32}^{jkl} = \tau^{j}\delta^{kl}[3] + \tfrac{5}{4}\tau^{n}\,\alpha^{jkm}\,\alpha^{lmn}[3],$$
$$\xi_{4}^{jklm} = -2\,\alpha^{jklm} + 4\,\alpha^{jmn}\alpha^{kln}[3] + 4\,\delta^{jm}\,\delta^{kl}.$$

Let g_p be the density function of the χ_p^2 distribution, $\mathcal{D} = \{v \mid \|v\| \geq c_\alpha\}$ where $v = (v_1, \cdots, v_n)$, and $K_1 = c_\alpha^{-1} g_p^{-1}(c_\alpha)$. Moreover we define $\xi_{22}^{jk}(0)$ to be the value of ξ_{22}^{jk} when $\tau = 0$. It turns out that we have the following Cornish -Fisher expansion for $\hat{c}_\alpha$:

$$\hat{c}_\alpha = c_\alpha(1 + \beta_1 n^{-1}) + O_p(n^{-\frac{3}{2}}),$$

where

$$\begin{aligned}\beta_1 =K_1\Big\{ &\tfrac{1}{2}(\xi_{22}^{jj}(0) + \xi_{11}^{j}\xi_{11}^{j})\int_{\mathcal{D}} H_2(v_j)\phi(v)dv \\ &+ (\tfrac{1}{24}\xi_4^{jjjj} + \tfrac{1}{6}\xi_{11}^{j}\xi_{31}^{jjj})\int_{\mathcal{D}} H_4(v_j)\phi(v)dv \\ &+ (\tfrac{1}{8}\xi_4^{jjkk} + \tfrac{1}{2}\xi_{11}^{j}\xi_{31}^{jkk})\int_{\mathcal{D}j\neq k} H_2(v_j)H_2(v_k)\phi(v)dv \\ &+ \tfrac{1}{72}\xi_{31}^{jjj}\,\xi_{31}^{jjj}\int_{\mathcal{D}} H_6(v_j)\phi(v)dv \\ &+ (\tfrac{1}{12}\xi_{31}^{jjk}\,\xi_{31}^{kkk} + \tfrac{1}{8}\xi_{31}^{jkk}\,\xi_{31}^{jkk})\int_{\mathcal{D}j\neq k} H_2(v_j)H_4(v_k)\phi(v)dv\Big\}.\end{aligned}$$

To develop an Edgeworth expansion for the power of the bootstrap test, we define

$$F_2(x,\tau) = \int_{\mathcal{D}_\tau(x)} \{\xi_{11}^{j}\,v_j + \tfrac{1}{2}\xi_{21}^{jk}(v_j v_k - \delta^{jk}) + \tfrac{1}{6}\,\xi_{31}^{jkl} H_3(v_j, v_k, v_l)\}\phi(v)dv \tag{5.3.14}$$

and

$$\begin{aligned}F_3(x,\tau) = &\int_{\mathcal{D}_\tau(x)} \{\xi_{12}^{j}v_j + \tfrac{1}{2}(\xi_{22}^{jk} + \xi_{11}^{j}\xi_{11}^{k})(v_j v_k - \delta^{jk})\}\phi(v)dv \\ &+ \int_{\mathcal{D}_\tau(x)} (\tfrac{1}{6}\xi_{32}^{jkl} + \tfrac{1}{2}\xi_{11}^{j}k_{21}^{kl})H_3(v_j, v_k, v_l)\phi(v)dv\end{aligned}$$

$$+ \int_{\mathcal{D}_\tau(x)} (\tfrac{1}{24} k_4^{jklm} + \tfrac{1}{6} k_{21}^{jk} k_{21}^{lm} + \tfrac{1}{6} \xi_{11}^{j} \xi_{31}^{klm}) H_4(v_j, v_k, v_l, v_m) \phi(v) dv,$$

$$+ \int_{\mathcal{D}_\tau(x)} \tfrac{1}{12} \xi_{21}^{jk} \xi_{31}^{lmp} H_5^{jklmp}(v_j, v_k, v_l, v_m, v_p) \phi(v) dv$$

$$+ \int_{\mathcal{D}_\tau(x)} \tfrac{1}{72} \xi_{31}^{jkl} \xi_{31}^{mpq} H_6^{jklmp}(v_j, v_k, v_l, v_m, v_p, v_q) \phi(v) dv.$$

Now we are able to give an Edgeworth expansion for the distribution of $S_o^T(\tau)\, S_o(\tau)$, in the following theorem.

Theorem 5.3.2 *Assume condition (5.2.1). Then for any real x,*

$$P\{S^T(\tau)\, S(\tau) > x\} = P\{\chi_p^2(\|\tau\|) > x\} + F_2(x, \tau)\, n^{-\frac{1}{2}} + F_3(x, \tau)\, n^{-1} + O(n^{-\frac{3}{2}}).$$

We do not give the proof of Theorem 5.3.2 here, since it may be derived straight forwardly from (5.3.12), (5.3.13) and Theorem 1.3.1.

From Theorem 5.3.2 and using the delta method, we obtain the following expansion for the power of the bootstrap test ϕ_b:

$$pow(\phi_b; \tau) = P(S^T S > \hat{c}_\alpha) \tag{5.3.15}$$
$$= P\{\chi_p^2(\|\tau\|) > c_\alpha\} + F_2(c_\alpha; \tau)\, n^{-\frac{1}{2}} + \{F_3(c_\alpha, \tau) - \beta_1 g_{p\tau}(c_\alpha)\}\, n^{-1} + O(n^{-\frac{3}{2}}).$$

From (5.3.13) and (5.3.14)

$$F_2(c_\alpha; \tau) = -\int_{\mathcal{D}_\tau(c_\alpha)} \{\tfrac{1}{2}\alpha^{jkk}\, v_j + \tfrac{1}{2}\alpha^{jkl}\tau^l (v_j v_k - \delta^{jk}) + \tfrac{1}{3}\alpha^{jkl} H_3(v_j, v_k, v_l)\}\, \phi(v) dv. \tag{5.3.16}$$

From the above formula and (5.3.8) we see that the powers of the empirical likelihood and bootstrap tests have different second-order terms.

5.4. Power Comparisons

In this section we use the power expansions of the empirical likelihood and bootstrap tests developed in the previous section, to compare the powers of these two tests. Two rules are proposed for choosing practically the more powerful test. One is for the univariate case, another is for the bivariate case.

From (5.3.7) and (5.3.15) we know that both ϕ_{ec} and ϕ_b have the same first order term $P\{\chi_p^2(\|\tau\|) > c_\alpha\}$ in their power functions. Thus, a comparison should be made of higher-order terms. However, we shall only compare the second-order terms. The reason is that when the sample size n is large enough, the difference in the power of the two tests is dominated by the difference between the second order terms $E_2(c_\alpha;\tau)$ and $F_2(c_\alpha;\tau)$. From (5.3.8) and (5.3.16) we have

$$E_2(c_\alpha;\tau) = \int_{\mathcal{D}_\tau(c_\alpha)} \{(\tfrac{1}{3}\alpha^{jlm}\tau^l\tau^m - \tfrac{1}{6}\alpha^{jkk})v_j + \tfrac{1}{6}\alpha^{jkl}\tau^l(v_jv_k - \delta^{jk})\}\phi(v)dv,$$

$$F_2(c_\alpha;\tau) = -\int_{\mathcal{D}_\tau(c_\alpha)} \{\tfrac{1}{2}\alpha^{jkk}v_j + \tfrac{1}{2}\alpha^{jkl}\tau^l(v_jv_k - \delta^{jk}) + \tfrac{1}{3}\alpha^{jkl}H_3(v_j,v_k,v_l)\}\phi(v)dv.$$

5.4.1 The Univariate Case

For the univariate case (i.e. $p = 1$), put $w_1 = \sqrt{c_\alpha} - \tau$, $w_2 = \sqrt{c_\alpha} + \tau$ and $\alpha_3 = \alpha^{111}$. Then, (5.3.8) and (5.3.16) have the following forms:

$$\begin{aligned}
E_2(c_\alpha;\tau) &= \tfrac{1}{6}\alpha_3\left[(2\tau^2 - 1)\{\phi(w_1) - \phi(w_2)\} + \tau\{w_1\phi(w_1) + w_2\phi(w_2)\}\right],\\
F_2(c_\alpha;\tau) &= \tfrac{1}{6}\alpha_3\left[-\{\phi(w_1) - \phi(w_2)\} - 3\tau\{w_1\phi(w_1) + w_2\phi(w_2)\}\right.\\
&\quad \left. - 2\{w_1^2\phi(w_1) - w_2^2\phi(w_2)\}\right].
\end{aligned}$$

Thus we obtain,

$$\begin{aligned}
E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau) &= \tfrac{1}{6}\alpha_3\left[2\tau^2\{\phi(w_1) - \phi(w_2)\} + 4\tau\{w_1\phi(w_1) + w_2\phi(w_2)\}\right.\\
&\quad \left. + 2\{w_1^2\phi(w_1) - w_2^2\phi(w_2)\}\right]\\
&= \tfrac{1}{3}\alpha_3\{(\tau + w_1)^2\phi(w_1) - (\tau - w_2)^2\phi(w_2)\}\\
&= \tfrac{1}{3}\alpha_3 c_\alpha\{\phi(w_1) - \phi(w_2)\}.
\end{aligned}$$

Since $\phi(w_1) - \phi(w_2)$ is positive when $\tau > 0$, and negative when $\tau < 0$, we obtain

$$E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau) \begin{cases} \geq 0, & \text{if } \alpha_3\tau > 0;\\ < 0, & \text{otherwise.}\end{cases} \tag{5.4.1}$$

From (5.4.1) we see that in the univariate case the relative powerfulness of the two tests depends on whether the skewness parameters α and τ have identical sign. If α and τ have identical sign, the empirical likelihood test is more powerful; and vice visa if α and τ have different signs. Since α_3 is usually unknown, we may estimate it by its sample version $\hat{\alpha}_3$. Now we establish the following rule for the univariate case, which suggests when to use the empirical likelihood test and when to use the bootstrap test.

The Univariate Rule *When the sample size is reasonably large, we may choose the more powerful test between the empirical likelihood and bootstrap tests by the following rule:*

$$\begin{cases} \textit{use the empirical likelihood test,} & \textit{if } \hat{\alpha}_3\,\tau > 0; \\ \textit{use any of the two tests,} & \textit{if } \hat{\alpha}_3\,\tau = 0; \\ \textit{use the bootstrap test,} & \textit{if } \hat{\alpha}_3\,\tau < 0. \end{cases}$$

5.4.2 The Bivariate Case

For the bivariate case (i.e. $p = 2$) we write $\tau = (\tau_1, \tau_2)$ and define

$$I_1(\tau) = \int_{\mathcal{D}_\tau(x)} v_1\phi(v)dv,\ I_2(\tau) = \int_{\mathcal{D}_\tau(x)} v_2\phi(v)dv,\ I_{12}(\tau) = \int_{\mathcal{D}_\tau(x)} v_1v_2\phi(v)dv,$$

$$I_{11}(\tau) = \int_{\mathcal{D}_\tau(x)} (v_1^2 - 1)\phi(v)dv,\ I_{22}(\tau) = \int_{\mathcal{D}_\tau(x)} (v_2^2 - 1)\phi(v)dv,$$

$$I_{111}(\tau) = \int_{\mathcal{D}_\tau(x)} (v_1^3 - 3v_1)\phi(v)dv,\ I_{112}(\tau) = \int_{\mathcal{D}_\tau(x)} (v_1^2 - 1)v_2\phi(v)dv,$$

$$I_{122}(\tau) = \int_{\mathcal{D}_\tau(x)} v_1(v_2^2 - 1)\phi(v)dv,\ I_{222}(\tau) = \int_{\mathcal{D}_\tau(x)} (v_2^3 - 3v_2)\phi(v)dv.$$

We have from (5.3.8) and (5.3.16) that

$$\begin{aligned} E_2(c_\alpha;\tau) = \alpha^{111}&\left\{\tfrac{1}{6}(2\,\tau_1^2 - 1)J_1(\tau) + \tfrac{1}{6}\tau_1 J_{11}(\tau)\right\} + \alpha^{222}\left\{\tfrac{1}{6}(2\,\tau_2^2 - 1)J_2(\tau) + \tfrac{1}{6}\tau_2 J_{22}(\tau)\right. \\ &+ \alpha^{112}\left\{\tfrac{2}{3}\tau_1\tau_2 J_1(\tau) + (\tfrac{1}{3}\tau_1^2 - \tfrac{1}{6})J_2(\tau) + \tfrac{1}{6}\tau_2 J_{11}(\tau) + \tfrac{1}{3}\tau_1\, J_{12}(\tau)\right\} \\ &+ \alpha^{122}\left\{\tfrac{1}{6}(2\,\tau_2^2 - 1)J_1(\tau) + \tfrac{2}{3}\tau_1\tau_2 J_2(\tau) + \tfrac{1}{6}\tau_1 J_{22}(\tau) + \tfrac{1}{3}\tau_2\, J_{12}(\tau)\right\} \end{aligned}$$

and

$$\begin{aligned} F_2(c_\alpha;\tau) =&\,\alpha^{111}\left\{-\tfrac{1}{2}\,J_1(\tau) - \tfrac{1}{2}\,\tau_1\,J_{11}(\tau) - \tfrac{1}{3}\,J_{111}(\tau)\right\} \\ &+ \alpha^{222}\left\{-\tfrac{1}{2}\,J_2(\tau) - \tfrac{1}{2}\,\tau_2\,J_{22}(\tau) - \tfrac{1}{3}\,J_{222}(\tau)\right\} \\ &+ \alpha^{112}\left\{-\tfrac{1}{2}\,J_2(\tau) - \tfrac{1}{2}\,\tau_2\,J_{11}(\tau) - \tau_1\,J_{12}(\tau) - J_{112}(\tau)\right\} \\ &+ \alpha^{122}\left\{-\tfrac{1}{2}\,J_1(\tau) - \tfrac{1}{2}\,\tau_1\,J_{22}(\tau) - \tau_2\,J_{12}(\tau) - J_{122}(\tau)\right\} \end{aligned}$$

Therefore,

$$E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau) = \alpha^{111}\, J_{111}(\tau) + \alpha^{112}\, J_{112}(\tau) + \alpha^{122}\, J_{122}(\tau) + \alpha^{222}\, J_{222}(\tau), \tag{5.4.2}$$

where

$$\begin{aligned}
J_{111}(\tau) &= \tfrac{1}{3}(\tau_1^2+1)\, I_1(\tau) + \tfrac{2}{3}\tau_1\, I_{11}(\tau) + \tfrac{1}{3}\, I_{111}(\tau),\\
J_{222}(\tau) &= \tfrac{1}{3}(\tau_2^2+1)\, I_2(\tau) + \tfrac{2}{3}\tau_2\, I_{22}(\tau) + \tfrac{1}{3}\, I_{222}(\tau),\\
J_{112}(\tau) &= \tfrac{2}{3}\tau_1\tau_2\, I_1(\tau) + \tfrac{1}{3}(\tau_1^2+1)\, I_2(\tau) + \tfrac{2}{3}\tau_2\, I_{11}(\tau) + \tfrac{4}{3}\tau_1\, I_{12}(\tau) + I_{112}(\tau),\\
J_{122}(\tau) &= \tfrac{1}{3}(\tau_2^2+1)\, I_1(\tau) + \tfrac{2}{3}\tau_1\tau_2\, I_2 + \tfrac{2}{3}\tau_1\, I_{22}(\tau) + \tfrac{4}{3}\tau_2\, I_{12}(\tau) + I_{122}(\tau).
\end{aligned} \tag{5.4.3}$$

Notice from (5.4.3) that $J_{111}(\tau)$, $J_{112}(\tau)$, $J_{122}(\tau)$ and $J_{222}(\tau)$ only depend on τ and not otherwise on the underlying distributions. All of them can be calculated numerically for each given $\tau = (\tau_1, \tau_2)$.

To find out the sign of $E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau)$ given in (5.4.2), we also estimate α^{111}, α^{112}, α^{122} and α^{222} by their sample versions $\hat\alpha^{111}$, $\hat\alpha^{112}$, $\hat\alpha^{122}$ and $\hat\alpha^{222}$, respectively. Then we define an estimator of $E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau)$, which is

$$\widehat{\{E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau)\}} = \hat\alpha^{111}\, J_{111}(\tau) + \hat\alpha^{112}\, J_{112}(\tau) + \hat\alpha^{122}\, J_{122}(\tau) + \hat\alpha^{222}\, J_{222}(\tau).$$

Now we are able to give the following rule for choosing a test for the bivariate case.

The Bivariate Rule *When the sample size is reasonably large, we may choose the more powerful test between the empirical likelihood and bootstrap tests by the following rule:*

$$\begin{cases}
\textit{use the empirical likelihood test,} & \textit{if } \widehat{\{E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau)\}} > 0;\\
\textit{use any of the two tests,} & \textit{if } \widehat{\{E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau)\}} = 0;\\
\textit{use the bootstrap test,} & \textit{if } \widehat{\{E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau)\}} < 0.
\end{cases}$$

Remark 1 For the sake of conciseness, we shall not develop rules for cases where $p \geq 3$ in this thesis. However, one may develop some rules in the same way as for the case $p = 2$ by employing general formulae for $E_2(c_\alpha;\tau)$ and $F_2(c_\alpha;\tau)$ given in (5.3.8) and (5.3.16).

Remark 2 We must emphasise that the above rules are based on the large sample properties of the tests. However, they do give us some indication of what is really going on even when the sample size is not too large.

5.5 Simulation Study

In this section we run simulations to see if the theoretical rules developed in Section 5.4 are consistent with empirical outcomes. We considere two univariate cases and one bivariate case. The first univariate case is that where the samples are drawn from $N(0,1)$, the standard normal distribution; we want to test $H_o : \mu = 0$ against $H_n : \mu = n^{-1/2}\tau$. In the second univariate case we drew samples from $Exp(1.0)$, the exponential distribution with unit mean, and we tested the hypotheses $H_o : \mu = 1$ against $H_n : \mu = 1 + n^{-1/2}\tau$. In the bivariate case, we took random vectors $X_i = (X_i^1, X_i^2)$ for $i = 1, \cdots, n$,

$$\begin{cases} X_i^1 = y_i^o + y_i^1, \\ X_i^2 = y_i^o + y_i^2, \end{cases}$$

where y_i^o, y_i^1, y_i^2 were drawn independently from the exponential distribution $Exp(1.0)$. We chose sample size $n = 15$ and 30 for each of the univariate case, and $n = 30$ for the bivariate case. We fixed the level of the tests to be 0.90 in all the cases considered. The normal and exponential random variables were generated by the routines of Press el al. (1989).

The power curves of the empirical likelihood and bootstrap tests appearing in Figure 5.1 were obtained by running 5000 simulations at each of 19 values of τ, equally spaced within the interval (-4.5, 4.5). When calculating the power of the bootstrap test, we generated 499 resamples for each of the 5000 simulated samples. For the bivariate case, we calculated the powers of the two tests at 225 points of $\tau = (\tau_1, \tau_2)$ within the rectangular area $(-3.5, 3.5) \times (-3.5, 3.5)$, based on 5000 simulations and 999 resamples for each simulated sample. A contour plot of the difference between power functions of the empirical likelihood and the bootstrap tests is shown in Figure 5.2.

In the first univariate case we have $\alpha_3 = 0$ since the random variables were drawn from $N(0,1)$. According to the Univariate Rule, we can use any one of the two tests since the powers of the tests should be very same regardless of the value of τ. This is just what we see from parts (a) and (b) of Figure 5.1. The underlying reason for this similarity is that $\alpha_3 = 0$ makes both E_2 and F_2 vanish. Consequently, the difference between the powers of the empirical likelihood and the bootstrap tests is of order n^{-1}, rather than $n^{-1/2}$. For the second univariate case we know that $\alpha_3 = 2$. So the Univariate Rule predicts that the empirical likelihood test is more (less) powerful than the bootstrap test if $\tau < 0$ ($\tau > 0$). This is again just what parts (c) and (d) of Figure 5.1 try to tell us. Notice that when $\tau \in (0.5, 3.5)$, the empirical likelihood test is about 20 per cents more powerful than the bootstrap test. However, when sample size is $n = 15$, which is small, we observe in the normal case that the empirical likelihood test is marginally more powerful than the bootstrap test over all values of τ. At meanwhile, in the exponential case the empirical likelihood performs similarly with the bootstrap test in the range of $\tau < 0$, where the bootstrap tests should perform better. This may be due to the fact that the bootstrap test has to use to an explicit variance estimate, which can be very unreliable when the sample size is small, whereas the empirical likelihood test implicitly uses the true variance.

For the bivariate case, it can be shown that $\alpha^{111} = \alpha^{222} = 2(a+b)^3 + 2(a^3+b^3)$ and $\alpha^{112} = \alpha^{122} = 2ab(a+b) + 2(a+b)^3$, where $a = 0.5\,(1 + 1/\sqrt{3})$ and $b = 0.5\,(-1 + 1/\sqrt{3})$. After numerically calculating J_{111},J_{112}, J_{122} and J_{222}, it can be shown that

$$E_2(c_\alpha;\tau) - F_2(c_\alpha;\tau) \begin{cases} \geq 0 & \text{if } \tau_1 + \tau_2 \geq 0; \\ < 0 & \text{if } \tau_1 + \tau_2 < 0. \end{cases}$$

So the Bivariate Rule would suggest using the empirical likelihood test when $\tau_1 + \tau_2 > 0$, using the bootstrap test when $\tau_1 + \tau_2 < 0$, and using either of the two tests if $\tau_1 + \tau_2 = 0$. In Figure 5.2, we give a contour plot of $pow(\phi_{ec}) - pow(\phi_b)$, which is very consistent with the prediction made from the Bivariate Rule.

Figure 5.1: The graphs depict powers curves of empirical likelihood test (solid curves) and Bootstrap test (dashed curves) as functions of τ. In case (a) and (b) the samples were generated from $N(0,1)$, we tested $H_o : \mu = 0$; against $H_n : \mu = n^{1/2}\,\tau$. In case (c) and (d) the sample were generated from $Exp(1.00)$, we tested $H_o : \mu = 1$; against $H_n : \mu = 1 + n^{-1/2}\,\tau$. The level of the test was 0.90 and the sample size $n = 15$ in (a) (c) and $n = 30$ in (b) (d).

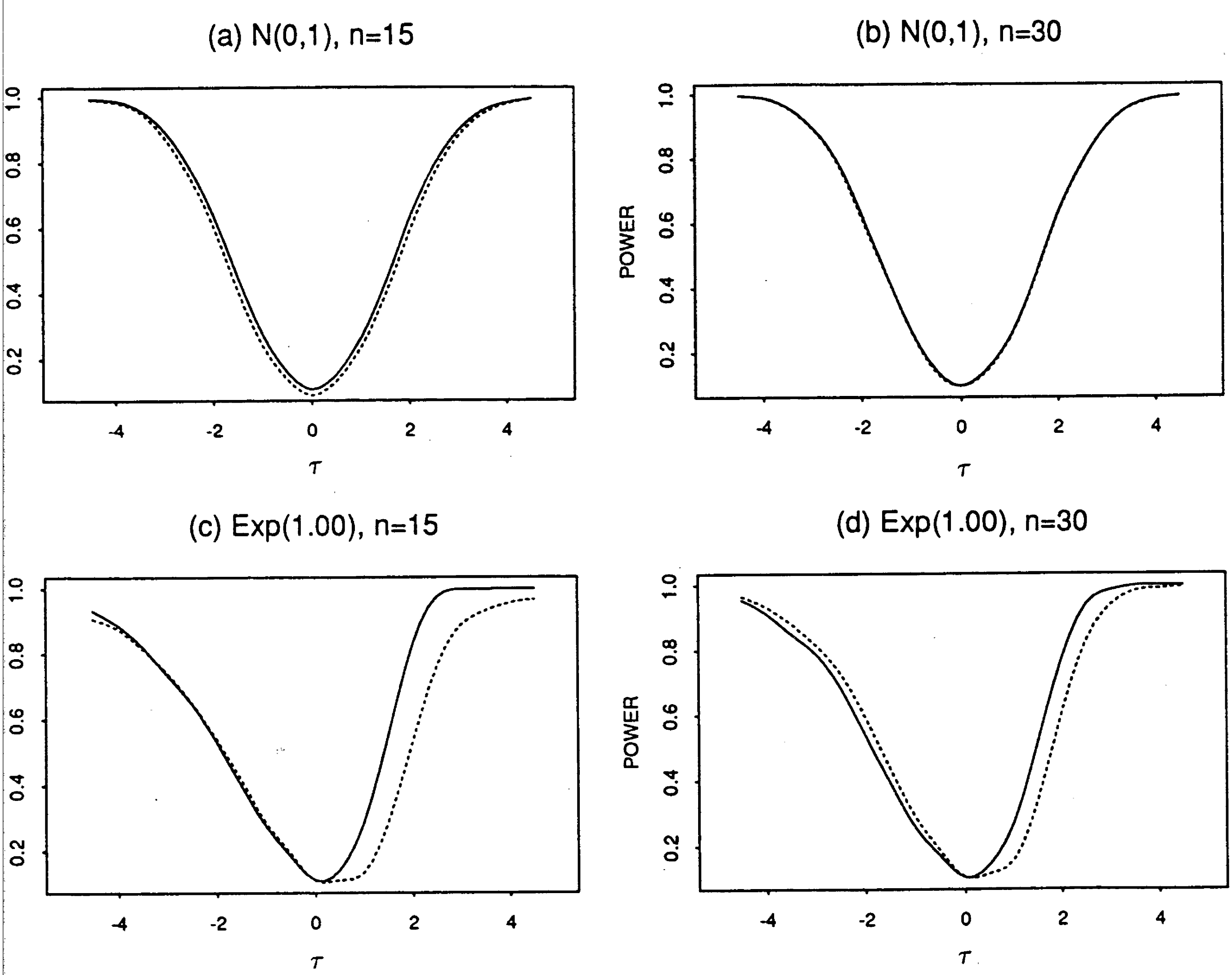
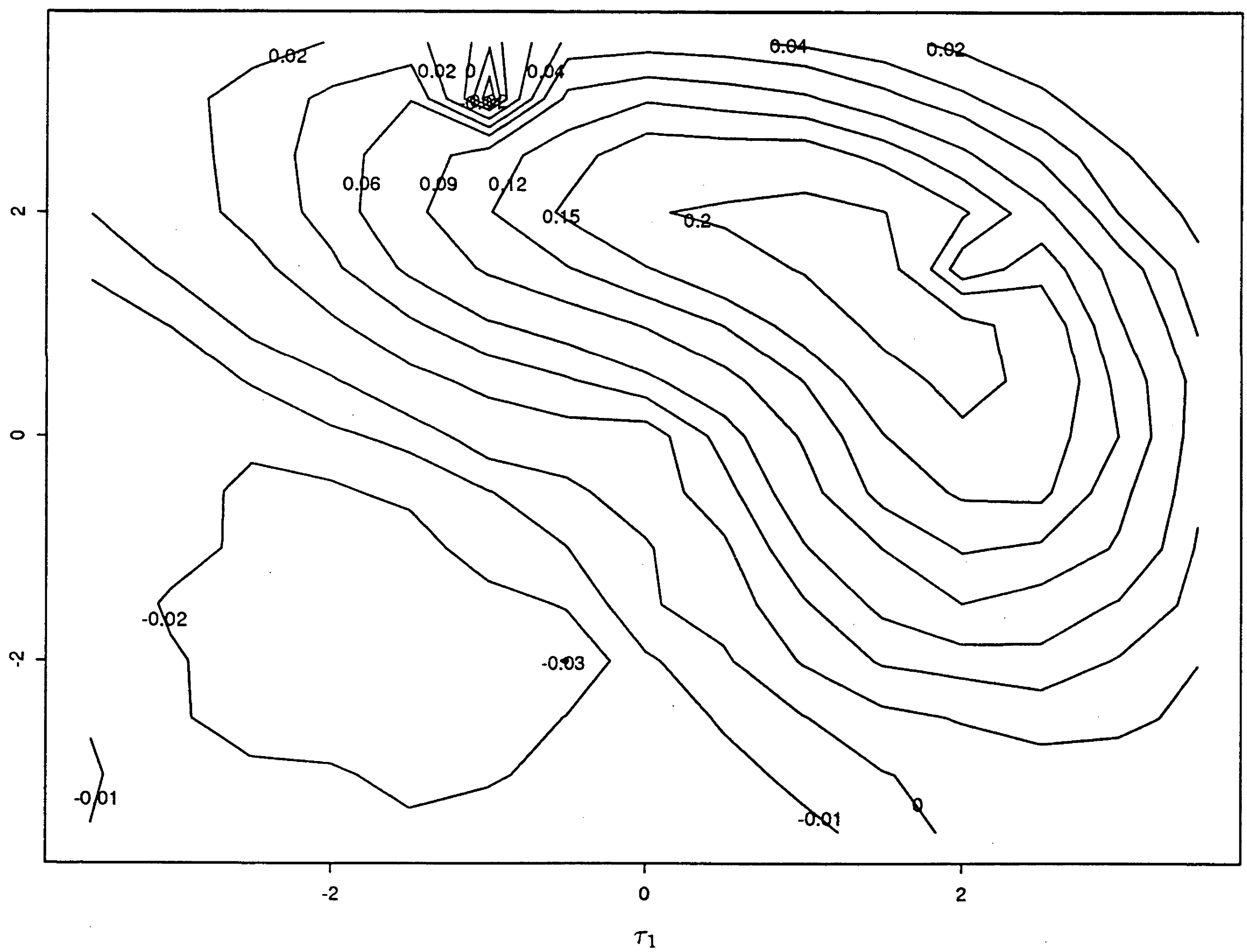


Figure 5.2: Contour plot of the difference between the powers of empirical likelihood and bootstrap tests from 5,000 simulation. The random samples were $(y_i^0 + y_i^1, y_i^0 + y_i^2)$ where y_i^l $1 \le l \le 3$ were drawn independently from $Exp(1.0)$. We tested $H_o : \mu = (2,2)^T$ against $H_n : \mu = (2,2)^T + n^{-1/2}\,\Sigma^{1/2}(\tau_1,\tau_2)^T$, where $\Sigma(1,1) = \Sigma(2,2) = 2$ and $\Sigma(1,2) = \Sigma(2,1) = 1$. The level of the test was 0.90 and the sample size $n = 30$.



5.6 Proof

In this section we give the proof of Theorem 5.3.1.

Theorem 5.3.1 *Assume condition (5.2.1). Then, for any $x > 0$,*

$$P\{\ell(\mu - n^{-\frac{1}{2}}\Sigma^{1/2}\tau) > x\} = P\{\chi_p^2(\|\tau\|) > x\} + E_2(x,\tau)n^{-\frac{1}{2}} + E_3(x,\tau)n^{-1} + O(n^{-\frac{3}{2}}).$$

Proof: Define

$$\begin{aligned}\Pi(v) = 1 &+ n^{-\frac{1}{2}}\{k_{11}^{j}\, v_j + \tfrac{1}{2}k_{21}^{jk}(v_j v_k - \delta^{jk})\} + n^{-1}\{k_{12}^{j}v_j + \\ &+ \tfrac{1}{2}(k_{22}^{jk} + k_{11}^{j}k_{11}^{k})(v_j v_k - \delta^{jk}) + (\tfrac{1}{6}k_{32}^{jkl} + \tfrac{1}{2}k_{11}^{j}k_{21}^{kl})H_3(v_j, v_k, v_l) \\ &+ \tfrac{1}{6}\, k_{21}^{jk}k_{21}^{lm}\, H_4(v_j, v_k, v_l, v_m)\},\end{aligned}$$

where the k's are given by (5.3.4). From (5.3.3) and (5.3.4), a formal Edgeworth expansion for the distribution function of $n^{\frac{1}{2}}R(\tau)$ can be constructed as follows,

$$P(n^{1/2}\, R(\tau) < x) = \int_{-\infty}^{x} \Pi(v)\,\phi(v)\,dv + O(n^{-\frac{3}{2}}). \qquad (5.3.6)$$

Accepting that expansion (5.3.6) may be justified, we establish an Edgeworth expansion for the distribution of $\ell(\mu - n^{-\frac{1}{2}}\Sigma^{1/2}\tau)$ as follows,

$$\begin{aligned}&P\{\ell(\mu - n^{-\frac{1}{2}}\Sigma^{1/2}\tau) > x\} \\ &= P(nR^T R > x) + O(n^{-\frac{3}{2}}) = \int_{\mathcal{D}_\tau(x)} \Pi(v)\phi(v)dv + O(n^{-\frac{3}{2}}) \\ &= P\{\chi_p^2(\|\tau\|) > x\} + E_2(x,\tau)n^{-\frac{1}{2}} + E_3(x,\tau)n^{-1} + O(n^{-\frac{3}{2}}),\end{aligned}$$

where $E_2(x,\tau)$ and $E_3(x,\tau)$ are given in (5.3.5).

It remains to check that expansion (5.3.6) is valid. Since

$$\overline{U} = (A^1, \cdots, A^p, A^{11}, \cdots, A^{pp}, A^{111}, \cdots, A^{ppp})^T,$$

We see that $\overline{U}$ is the mean of i.i.d. random vectors with mean 0, and there exists a smooth function h such that $R(\tau) = h(\overline{U})$. Thus, we can justify the expansion (5.3.6) by using Theorem 1.3.2 under condition (5.2.1). Therefore the theorem is proved. □

Appendix 5 Calculations of Cumulants

In this appendix we present our calculations of cumulants of $n^{\frac{1}{2}} R(\tau)$ and $n^{\frac{1}{2}} S(\tau)$ respectively.

Appendix 5.1 Cumulants of $n^{\frac{1}{2}} R(\tau)$

Let $k_l^{j_1,\cdots,j_l}$ denote the joint l'th order cumulant of $n^{\frac{1}{2}} R(\tau)$. In this part of the appendix, we present the calculations of $k_l^{j_1,\cdots,j_l}$'s. Recall that

$$R(\tau) = R_1(\tau) + R_2(\tau) + R_3(\tau),$$

where

$$\begin{aligned}
R_1^j(\tau) &= (A + n^{-\frac{1}{2}}\tau)^j, \\
R_2^j(\tau) &= -\tfrac{1}{2}A^{jk}(A + n^{-\frac{1}{2}}\tau)^k + \tfrac{1}{3}\alpha^{jkl}(A + n^{-\frac{1}{2}}\tau)^k(A + n^{-\frac{1}{2}}\tau)^l, \\
R_3^j(\tau) &= (-\tfrac{1}{2}\tau^j\tau^k n^{-1} + \tfrac{3}{8}A^{jl}A^{kl} - \tfrac{1}{2}\tau^j A^k[2]n^{-\frac{1}{2}})(A + n^{-\frac{1}{2}}\tau)^k \qquad (5.A.1) \\
&\quad + \{\tfrac{1}{3}(A^{jkl} + n^{-\frac{1}{2}}\tau^j\delta^{kl}[3]) - \tfrac{5}{6}\alpha^{jkm}A^{lm}\}(A + n^{-\frac{1}{2}}\tau)^j(A + n^{-\frac{1}{2}}\tau)^k \\
&\quad + (\tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})(A + n^{-\frac{1}{2}}\tau)^k(A + n^{-\frac{1}{2}}\tau)^l(A + n^{-\frac{1}{2}}\tau)^m.
\end{aligned}$$

Put

$$\begin{aligned}
\tilde{R}_1^j &= A^j, \\
\tilde{R}_2^j(\tau) &= -\tfrac{1}{2}A^{jk}(A + n^{-\frac{1}{2}}\tau)^k + \tfrac{1}{3}\alpha^{jkl}(A^kA^l + n^{-\frac{1}{2}}\tau^kA^l[2]), \qquad (5.A.2) \\
\tilde{R}_3^j(\tau) &= -\tfrac{1}{2}\tau^j\tau^kA^kn^{-1} + \tfrac{3}{8}A^{jl}A^{kl}(A + n^{-\frac{1}{2}}\tau)^k - \tfrac{1}{2}\tau^jA^k[2](A + n^{-\frac{1}{2}}\tau)^kn^{-\frac{1}{2}}) \\
&\quad + \tfrac{1}{3}(A^{jkl} + n^{-\frac{1}{2}}\tau^j\delta^{kl}[3])(A^kA^l + n^{-\frac{1}{2}}\tau^kA^l[2]n^{-\frac{1}{2}}) + \tfrac{1}{3}A^{jkl}\tau^k\tau^ln^{-1} \\
&\quad - \tfrac{5}{6}\alpha^{jkm}A^{lm}(A + n^{-\frac{1}{2}}\tau)^k(A + n^{-\frac{1}{2}}\tau)^l \\
&\quad + (\tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})(A + n^{-\frac{1}{2}}\tau)^k(A + n^{-\frac{1}{2}}\tau)^l(A + n^{-\frac{1}{2}}\tau)^m,
\end{aligned}$$

and

$$\Delta^j = \tau^jn^{-\frac{1}{2}} + \tfrac{1}{3}\alpha^{jkl}\tau^k\tau^ln^{-1} + \{-\tfrac{1}{2}\tau^j\tau^k\tau^k + \tfrac{1}{3}(\tau^j\delta^{kl})[3]\tau^k\tau^l\}n^{-\frac{3}{2}}.$$

It is clear that

$$R^j(\tau) = \Delta^j + \tilde{R}^j(\tau),$$

where

$$\tilde{R}^j(\tau) = \tilde{R}^j_1 + \tilde{R}^j_2(\tau) + \tilde{R}^j_3(\tau).$$

Let $\tilde{k}_l^{j_1,\cdots,j_l}$ denote the joint l'th order cumulants of $n^{\frac{1}{2}}\tilde{R}$. Since

$$k_l^{j_1,\cdots,j_l} = \begin{cases} n^{\frac{1}{2}}\Delta^j + \tilde{k}_l^{j_1,\cdots,j_l}, & \text{if } l = 1; \\ \tilde{k}_l^{j_1,\cdots,j_l}, & \text{if } l \geq 2, \end{cases} \tag{5.A.3}$$

and $\tilde{k}_l^{j_1,\cdots,j_l}$ are more easily calculated than $k_l^{j_1,\cdots,j_l}$, we shall directly calculate $\tilde{k}_l^{j_1,\cdots,j_l}$ and get $k_l^{j_1,\cdots,j_l}$ via (5.A.3). For the cases $l \geq 5$, using the results given by James and Mayne (1962), we have

$$k_l^{j_1,\cdots,j_l} = O(n^{-\frac{3}{2}}), \quad \text{for } l \geq 5. \tag{5.A.4}$$

So in the following we only give calculations of $\tilde{k}_l^{j_1,\cdots,j_l}$ for $l = 1, 2, 3$ and 4.

We start by computing the first order cumulants $\tilde{k}_1^j$ of $n^{\frac{1}{2}}\tilde{R}(\tau)$ for $j = 1, \cdots, p$. Since

$$E(A^{j_1\cdots j_m}) = 0,$$

$$E(A^{i_1\cdots i_{m_1}}\, A^{i_1\cdots j_{m_2}}\, A^{l_1\cdots l_{m_3}}) = O(n^{-2})$$

for any integers m, m_1, m_2 and m_3 larger than 1, we have from (5.A.2) that

$$\begin{aligned} E(\tilde{R}^j_1) &= 0, \\ E\{\tilde{R}^j_2(\tau)\} &= -\tfrac{1}{2}E(A^{jk}\,A^k + \tfrac{1}{3}\alpha^{jkl}E(A^k\,A^l) \\ &= -\tfrac{1}{2}\alpha^{jkk}\,n^{-1} + \tfrac{1}{3}\alpha^{jkl}\,\delta^{kl}\,n^{-1} \\ &= -\tfrac{1}{6}\alpha^{jkk}\,n^{-1} \end{aligned}$$

and

$$\begin{aligned} E\{\tilde{R}^j_3(\tau)\} = {} & \tfrac{3}{8}E(A^{jl}A^{kl})\tau^k\,n^{-\frac{1}{2}} - \tfrac{1}{2}E\{(\tau^j A^k)[2]\,A^k\}n^{-\frac{1}{2}} \\ & + \tfrac{1}{3}E\{A^{jkl}\,(\tau^k\,A^l)[2]\}\,n^{\frac{1}{2}} + \tfrac{1}{3}\tau^j\delta^{kl}[3]\,E(A^k\,A^l)\,n^{-\frac{1}{2}} \\ & - \tfrac{5}{6}\alpha^{jkm}\,E\{A^{lm}\,(A^k\,\tau^l)[2]\,n^{-\frac{1}{2}} \\ & + (\tfrac{4}{9}\alpha^{jkn}\,\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})\,E(A^k\,A^l\,\tau^m[3])\,n^{-\frac{1}{2}} + O(n^{-2}) \end{aligned}$$

$$= \tfrac{3}{8}(\alpha^{jkll} - \delta^{jk})\tau^k n^{-\frac{3}{2}} - \tfrac{1}{2}(\tau^j \delta^{kk} + \tau^k \delta^{jk}) n^{-\frac{3}{2}}$$

$$+ \tfrac{1}{3}(\alpha^{jkll}\tau^k + \alpha^{jkkl}\tau^l) n^{-\frac{3}{2}} + \tfrac{1}{3}(\tau^j \delta^{kl})[3]\,\delta^{kl} n^{-\frac{3}{2}}$$

$$- \tfrac{5}{6}\alpha^{jkm}(\alpha^{mll}\tau^k + \alpha^{klm}\tau^l) n^{-\frac{3}{2}}$$

$$+ (\tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})(\tau^k \delta^{lm}[3]) n^{-\frac{3}{2}}$$

$$+ O(n^{-2})$$

$$= \tfrac{7}{24}\alpha^{jkll}\tau^k + (-\tfrac{1}{6}p - \tfrac{5}{24})\tau^j n^{-\frac{3}{2}}$$

$$+ (\tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})\tau^k \tau^l \tau^m n^{-\frac{3}{2}}$$

$$+ (-\tfrac{7}{18}\alpha^{jkm}\alpha^{mll}\tau^k + \tfrac{1}{18}\alpha^{jkm}\alpha^{klm}\tau^l) n^{-\frac{3}{2}} + O(n^{-2}).$$

Thus,

$$k_1^j = \Delta^j n^{\frac{1}{2}} + \tilde{k}_1^j = \Delta^j n^{\frac{1}{2}} + E\{\tilde{R}^j(\tau)\} n^{\frac{1}{2}}$$

$$= \tfrac{1}{6}\alpha^{jkk} n^{-\frac{1}{2}} + \tfrac{7}{24}\alpha^{jkll}\tau^k n^{-1} + (-\tfrac{1}{6}p - \tfrac{5}{24})\tau^j n^{-1}$$

$$+ (\tfrac{4}{9}\alpha^{jkn}\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})\tau^k \tau^l \tau^m n^{-1}$$

$$+ (-\tfrac{7}{18}\alpha^{jkm}\alpha^{mll}\tau^k + \tfrac{1}{18}\alpha^{jkm}\alpha^{klm}\tau^l) n^{-1} + O(n^{-2}). \qquad (5.A.5)$$

Next we calculate $\tilde{k}_2^{jk}$ for any $1 \le j,k \le p$. Notice that

$$\tilde{k}_2^{jk} = n\,[E\{\tilde{R}_1^j \tilde{R}_1^k\} + E\{\tilde{R}_1^j \tilde{R}_2^k(\tau)\}[2] + E\{\tilde{R}_1^j \tilde{R}_3^k(\tau)\}[2]$$

$$+ E\{\tilde{R}_2^j(\tau)\tilde{R}_2^k(\tau)\} + E\{\tilde{R}_2^j(\tau)\tilde{R}_3^k(\tau)\}[2] + E\{\tilde{R}_3^j(\tau)\tilde{R}_3^k(\tau)\}$$

$$- E\{\tilde{R}_1^j\}\,E\{\tilde{R}_1^k\}]. \qquad (5.A.6)$$

In the following we compute each term on the right-hand side of (5.A.6), obtaining

$$E(\tilde{R}_1^j \tilde{R}_1^k) = E(A^j A^k) = \delta^{jk} n^{-1},$$

$$E\{\tilde{R}_1^j \tilde{R}_2^k(\tau)\} = -\tfrac{1}{2} E\{A^j A^{kl}(A + n^{-\frac{1}{2}}\tau)^l$$

$$+ \tfrac{1}{3}\alpha^{klm} E\{A^j A^l A^m + A^j(\tau^l A^m[2]) n^{-\frac{1}{2}}\}$$

$$= -\tfrac{1}{2}\{(\alpha^{jkll} - \delta^{jk}) n^{-2} + \alpha^{jkl}\tau^l n^{-\frac{3}{2}}\}$$

$$+ \tfrac{1}{3}\alpha^{klm}\{\alpha^{jlm} n^{-2} + (\tau^l \delta^{jm} + \tau^m \delta^{jl})\}$$

$$= \tfrac{1}{6}\alpha^{jkl}\tau^l n^{-\frac{3}{2}} + (-\tfrac{1}{2}\alpha^{jkll} + \tfrac{1}{2}\delta^{jk} + \tfrac{1}{3}\alpha^{jlm}\alpha^{klm}) n^{-2},$$

$$E\{\tilde{R}_1^j\,\tilde{R}_3^k(\tau)\} = -\tfrac{1}{2}\tau^k\,\tau^l\,E(A^j\,A^l)\,n^{-2} + \tfrac{3}{8}\,E\{A^j\,A^{kl}\,A^{ml}\,(A + n^{-\frac{1}{2}}\,\tau)^m\}$$

$$- \tfrac{1}{2}\,E\{A^j\,(\tau^k\,A^l[2])\,(A + n^{-\frac{1}{2}}\,\tau)^l\} + \tfrac{1}{3}\,n^{-1}\,\tau^l\,\tau^m\,E(A^j\,A^{klm})$$

$$+ \tfrac{1}{3}\,E[A^j\,\{A^{klm} + n^{-\frac{1}{2}}\,(\tau^k\,\delta^{lm}[3])\}\,\{A^l A^m + n^{-\frac{1}{2}}\,(\tau^m\,A^l[2])\}]$$

$$- \tfrac{5}{6}\alpha^{jkm}\,E\{A^j\,A^{mn}\,(A + n^{-\frac{1}{2}}\,\tau)^l\,(A + n^{-\frac{1}{2}}\,\tau)^n\}$$

$$+ (\tfrac{4}{9}\alpha^{kln}\,\alpha^{mpn} - \tfrac{1}{4}\alpha^{klmp})$$

$$\times\, E\{A^j\,(A + n^{-\frac{1}{2}}\,\tau)^l\,(A + n^{-\frac{1}{2}}\,\tau)^m\,(A + n^{-\frac{1}{2}}\,\tau)^p\} + O(n^{-3})$$

$$= (\tfrac{5}{8}\,\alpha^{jkmm} - \tfrac{3}{8}\,\delta^{jm}\,\delta^{kl}\,\delta^{ml} - \tfrac{29}{72}\,\alpha^{jml}\,\alpha^{kml} - \tfrac{1}{72}\,\alpha^{jkm}\,\alpha^{mll})\,n^{-2}$$

$$+ (\tfrac{1}{3}\,\tau^j\,\tau^k + \tfrac{1}{6}\,\delta^{jk}\,\tau^l\,\tau^l - \tfrac{5}{12}\,\alpha^{jklm}\,\tau^l\,\tau^m)\,n^{-2}$$

$$+ (\tfrac{1}{18}\,\alpha^{jmn}\,\alpha^{kln}\,\tau^l\,\tau^n + \tfrac{4}{9}\,\alpha^{jkn}\,\alpha^{lmn}\,\tau^l\,\tau^m)\,n^{-2} + O(n^{-\frac{5}{2}})$$

and

$$E\{\tilde{R}_2^j(\tau)\,\tilde{R}_2^k(\tau)\} = \tfrac{1}{4}\,E\{A^{jl}\,(A + n^{-\frac{1}{2}}\,\tau)^l\,A^{km}\,(A + n^{-\frac{1}{2}}\,\tau)^m\} - \tfrac{1}{3}\,\alpha^{kmn}$$

$$\times\, E\{A^{jl}\,(A + n^{-\frac{1}{2}}\,\tau)^l\,A^m\,A^n + n^{-\frac{1}{2}}\,(\tau^m\,A^n[2])\,A^{jl}\,(A + n^{-\frac{1}{2}}\,\tau)^l\}$$

$$+ \tfrac{1}{9}\,\alpha^{jlp}\,\alpha^{kmn}\,E[\{A^l\,A^p + n^{-\frac{1}{2}}\,(\tau^l\,A^p[2])\}\,\{A^m\,A^n + n^{-\frac{1}{2}}\,(\tau^m\,A^n[2])\}]$$

$$= (\tfrac{1}{4}\,\alpha^{jkmm} + \tfrac{1}{36}\,\alpha^{jmm}\,\alpha^{kll} - \tfrac{7}{36}\,\alpha^{jml}\,\alpha^{kml} - \tfrac{1}{4}\,\delta^{jm}\,\delta^{kl}\,\delta^{ml})\,n^{-2}$$

$$+ (\tfrac{1}{4}\,\alpha^{jklm}\,\tau^l\,\tau^m - \tfrac{1}{4}\,\tau^j\,\tau^k - \tfrac{2}{9}\,\alpha^{jln}\,\alpha^{kmn}\,\tau^l\,\tau^m)\,n^{-2} + O(n^{-\frac{5}{2}}).$$

Also it is easy to show that

$$E\{\tilde{R}_2^j(\tau)\tilde{R}_3^k(\tau)\} = O(n^{-\frac{5}{2}}) \quad \text{and} \quad E\{\tilde{R}_3^j(\tau)\tilde{R}_3^k(\tau)\} = O(n^{-\frac{5}{2}}).$$

Thus, substituting the above formulae into (5.A.6), and using the earlier result of

$$E(\tilde{R}^j) = -\tfrac{1}{6}\,\alpha^{jkk}\,n^{-1} + O(n^{-\frac{3}{2}})$$

and (5.A.3), we end up with

$$k_2^{jk} = \delta^{jk} + \tfrac{1}{3}\,\alpha^{jkl}\,\tau^l\,n^{-\frac{1}{2}} + (\tfrac{1}{3}\,\delta^{jk}\,\tau^l\,\tau^l + \tfrac{5}{12}\,\tau^j\,\tau^k)\,n^{-1}$$

$$+ (\tfrac{1}{2}\,\alpha^{jkmm} - \tfrac{1}{3}\,\alpha^{jml}\,\alpha^{kml} - \tfrac{1}{36}\,\alpha^{jkm}\,\alpha^{mll})\,n^{-1}$$

$$+ (-\tfrac{1}{9}\,\alpha^{jmn}\,\alpha^{klm}\,\tau^l\,\tau^n + \tfrac{8}{9}\,\alpha^{jkn}\,\alpha^{lmn}\,\tau^l\,\tau^m)\,n^{-1}$$

$$- \tfrac{7}{12}\,\alpha^{jklm}\,\tau^l\,\tau^m\,n^{-1} + O(n^{-\frac{3}{2}}). \tag{5.A.7}$$

By the definition,

$$
\begin{aligned}
k_3^{jkl} &= n^{\frac{3}{2}} \, [E\{\tilde{R}^j(\tau)\,\tilde{R}^k(\tau)\,\tilde{R}^l(\tau)\} - E\{\tilde{R}^j(\tau)\}\,E\{\tilde{R}^k(\tau)\,\tilde{R}^l(\tau)\}[3] \\
&\quad + 2\,E\{\tilde{R}^j(\tau)\}\,E\{\tilde{R}^k(\tau)\}\,E\{\tilde{R}^l(\tau)\}] \\
&= n^{\frac{3}{2}} \, [E(\tilde{R}_1^j\,\tilde{R}_1^k\,\tilde{R}_1^l) + E\{\tilde{R}_2^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\}[3] \\
&\quad + E\{\tilde{R}_3^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\}[3] + E\{\tilde{R}_1^j\,\tilde{R}_2^k(\tau)\,\tilde{R}_2^l(\tau)[3]\} \\
&\quad - E\{\tilde{R}_2^j(\tau)\}\,E\{\tilde{R}_1^k\,\tilde{R}_1^l\}[3] - E\{\tilde{R}_3^j(\tau)\}\,E\{\tilde{R}_1^k\,\tilde{R}_1^l\}[3] \\
&\quad - E\{\tilde{R}_2^j(\tau)\}\,E\{\tilde{R}_1^k\,\tilde{R}_2^l(\tau)\}[6]] + O(n^{-\frac{3}{2}}). \qquad (5.A.8)
\end{aligned}
$$

After some algebra we may show that

$$E(\tilde{R}_1^j\,\tilde{R}_1^k\,\tilde{R}_1^l) = E(A^j\,A^k\,A^l) = \alpha^{jkl}\,n^{-2}$$

and

$$
\begin{aligned}
E\{\tilde{R}_2^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\} &= -\tfrac{1}{2}\,E\{A^{jm}\,(A + n^{-\frac{1}{2}}\,\tau)^m\,A^k\,A^l\} \\
&\quad + \tfrac{1}{3}\,\alpha^{jmn}\,E\{A^m\,A^n\,A^k\,A^l + n^{-\frac{1}{2}}\,(\tau^m\,A^n[2])\,A^k\,A^l\} \\
&= (-\tfrac{1}{6}\,\alpha^{jmm}\,\delta^{kl} - \tfrac{1}{3}\,\alpha^{jkl})\,n^{-2} \\
&\quad + \{-\tfrac{1}{2}\,(\alpha^{jklm} - \delta^{jm}\,\delta^{kl})\,\tau^m + \tfrac{2}{3}\,\alpha^{jmn}\,\alpha^{kln}\,\tau^m\}\,n^{-\frac{5}{2}} \\
&\quad + O(n^{-3}).
\end{aligned}
$$

Notice that

$$E(A^{i_1\cdots i_{m_1}}\,A^{j_1\cdots j_{m_2}}\,A^{l_1\cdots l_{m_3}}\,A^{h_1\cdots h_{m_4}}\,A^{q_1\cdots q_{m_5}}) = O(n^{-3})$$

for any integers $m_l \geq 1$, $l = 1, \cdots, 5$. Hence,

$$
\begin{aligned}
& E\{\tilde{R}_3^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\} \\
&= \tfrac{3}{8}\,E(A^{jn}\,A^{mn}\,A^k\,A^l)\,\tau^m\,n^{-\frac{1}{2}} - \tfrac{1}{2}\,E\{(\tau^j\,A^m[2])\,A^k\,A^l\,A^m\}\,n^{-\frac{1}{2}} \\
&\quad + \tfrac{1}{3}\,(\tau^j\,\delta^{mn}[3])\,E(A^k\,A^l\,A^m\,A^n)\,n^{-\frac{1}{2}} + \tfrac{1}{3}\,E\{A^{jml}\,A^k\,A^l\,(\tau^m\,A^n[2])\}\,n^{-\frac{1}{2}} \\
&\quad - \tfrac{5}{6}\,\alpha^{jmn}\,E\{A^{mp}\,A^k\,A^l\,(\tau^p\,A^n[2])\} \\
&\quad + (\tfrac{4}{9}\alpha^{jkn}\,\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})\,E\{A^k\,A^l\,(A^m\,A^n\,\tau^p[3])\}\,n^{-\frac{1}{2}} \\
&\quad + (\tfrac{4}{9}\alpha^{jkn}\,\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})\,\delta^{kl}\,\tau^m\,\tau^n\,\tau^p\,n^{-\frac{5}{2}} + O(n^{-3})
\end{aligned}
$$

$$= \tfrac{7}{24}\,\alpha^{jmnn}\,\delta^{kl}\,\tau^m\,n^{-\frac{5}{2}} - (\tfrac{p}{6} + \tfrac{13}{24}\tau^j\,\delta^{kl}\,n^{-\frac{5}{2}} + \tfrac{1}{6}\,(\delta^{jk}\,\tau^l + \delta^{jl}\,\tau^k)\,n^{-\frac{5}{2}}$$

$$- \tfrac{1}{6}\,\alpha^{jklm}\,\tau^m\,n^{-\frac{5}{2}} + \tfrac{31}{71}\,(\alpha^{jkm}\,\alpha^{lmn} + \alpha^{jlm}\,\alpha^{kmn})\,\tau^n\,n^{-\frac{5}{2}}$$

$$- \tfrac{7}{18}\,\alpha^{jmn}\,\alpha^{ppn}\,\delta^{kl}\,n^{-\frac{5}{2}} - \tfrac{7}{9}\,\alpha^{jmn}\,\alpha^{klm}\,\tau^n\,n^{-\frac{5}{2}}$$

$$+ \tfrac{1}{18}\,\alpha^{jmn}\,\alpha^{pmn}\,\delta^{kl}\,\tau^p\,n^{-\frac{5}{2}} + O(n^{-3}).$$

Moreover,

$$E\{\tilde{R}_1^j\,\tilde{R}_2^k(\tau)\,\tilde{R}_2^l(\tau)\}$$

$$= \tfrac{1}{4}\,E\{A^j\,A^{kn}\,A^{lm}\,(A + n^{-\frac{1}{2}}\,\tau)^m\,(A + n^{-\frac{1}{2}}\,\tau)^n$$

$$- \tfrac{1}{6}\,\alpha^{lmn}\,E[A^j\,A^{kp}\,(A + n^{-\frac{1}{2}}\tau)^p\{A^m\,A^n + n^{-\frac{1}{2}}\,(\tau^m\,A^n[2])\}]$$

$$- \tfrac{1}{6}\,\alpha^{kmn}\,E[A^j\,A^{lp}\,(A + n^{-\frac{1}{2}}\tau)^p\{A^m\,A^n + n^{-\frac{1}{2}}\,(\tau^m\,A^n[2])\}]$$

$$+ \tfrac{1}{9}\,\alpha^{khp}\,\alpha^{lmn}\,E[A^j\{A^h\,A^p + n^{-\frac{1}{2}}\,(\tau^h\,A^p[2])\}\,\{A^m\,A^n + n^{-\frac{1}{2}}\,(\tau^m\,A^n[2])\}]$$

$$= \tfrac{1}{4}\,E\{A^j\,A^{kn}\,A^{lm}\,A^m)\,\tau^n\,n^{-\frac{1}{2}} + \tfrac{1}{4}\,E\{A^j\,A^{kn}\,A^{lm}\,A^n)\,\tau^m\,n^{-\frac{1}{2}}$$

$$- \tfrac{1}{6}\,[\alpha^{lmn}\,E\{A^j\,A^{kp}\,(\tau^p\,A^m\,A^n[3])\} - \alpha^{kmn}\,[E\{A^j\,A^{lp}\,(\tau^p\,A^m\,A^n[3])\}]\,n^{-\frac{1}{2}}$$

$$+ \tfrac{1}{9}\,\alpha^{khp}\,\alpha^{lmn}\,[E\{A^j\,A^m\,A^n\,(\tau^h\,A^p[2])\} + E\{A^j\,A^h\,A^p\,(\tau^m\,A^n[2])\}]\,n^{-\frac{1}{2}}$$

$$+ O(n^{-3})$$

$$= (\tfrac{1}{2}\,\alpha^{jklp} - \tfrac{1}{4}\,\tau^k\,\delta^{jl} - \tfrac{1}{4}\,\tau^l\,\delta^{jk} - \tfrac{1}{36}\,\alpha^{jkp}\,\alpha^{lmm}\,\tau^p - \tfrac{1}{36}\,\alpha^{jlp}\,\alpha^{kmm}\,\tau^p$$

$$- \tfrac{11}{36}\,\alpha^{jkm}\,\alpha^{lpm}\,\tau^p - \tfrac{11}{36}\,\alpha^{jlm}\,\alpha^{kpm}\,\tau^p)\,n^{-\frac{5}{2}} + O(n^{-3}).$$

Furthermore, from the earlier calculation we have

$$E\{\tilde{R}_2^j(\tau)\}\,E(\tilde{R}_1^k\,\tilde{R}_1^l) = \tfrac{1}{6}\,\alpha^{jmm}\,\delta^{kl}\,n^{-2},$$

$$E\{\tilde{R}_3^j(\tau)\}\,E(\tilde{R}_1^k\,\tilde{R}_1^l) = \tfrac{7}{24}\,\alpha^{jmnn}\,\delta^{kl}\,\tau^m\,n^{-\frac{5}{2}} + (-\tfrac{1}{6}\,p - \tfrac{5}{24})\,\tau^j\,\delta^{kl}\,n^{-\frac{5}{2}}$$

$$+ (-\tfrac{7}{18}\,\alpha^{jmn}\,\alpha^{mpp}\,\tau^n\,\tfrac{1}{18}\,\alpha^{jmn}\,\alpha^{mnp}\,\tau^p)\,\delta^{kl}\,n^{-\frac{5}{2}} + O(n^{-3})$$

and

$$E\{\tilde{R}_2^j(\tau)\}\,E\{\tilde{R}_1^k\,\tilde{R}_2^l(\tau)\} = -\tfrac{1}{18}\,\alpha^{jpp}\,\alpha^{klm}\,\tau^m\,n^{-\frac{5}{2}} + O(n^{-3}).$$

Substituting the above results into (5.A.8), we have

$$k_3^{jkl} = (-\tfrac{1}{2}\,\alpha^{jklm}\,\tau^m + \tfrac{5}{36}\,\alpha^{jkm}\,\alpha^{lmn}\,\tau^n$$

$$+ \tfrac{5}{36}\,\alpha^{jlm}\,\alpha^{kmn}\,\tau^n + \tfrac{5}{36}\,\alpha^{klm}\,\alpha^{jmn}\,\tau^n)\,n^{-1} + O(n^{-\frac{3}{2}}). \qquad (5.A.9)$$

To calculate k_4^{jklm}, we split $\tilde{R}_2^j(\tau)$ and $\tilde{R}_3^j(\tau)$ as follows:

$$\tilde{R}_2^j(\tau) = \tilde{R}_{21}^j + \tilde{R}_{22}^j(\tau) \quad \text{and} \quad \tilde{R}_3^j(\tau) = \tilde{R}_{31}^j + \tilde{R}_{32}^j(\tau), \tag{5.A.10}$$

where

$$\tilde{R}_{21}^j = \tilde{R}_2^j(0),\ \tilde{R}_{22}^j(\tau) = -\tfrac{1}{2} A^{jk}\,\tau^k\, n^{-\frac{1}{2}} + \tfrac{1}{3}\alpha^{jkl}\,(\tau^k A^l[2])\, n^{-\frac{1}{2}},\ \tilde{R}_{31}^j = \tilde{R}_3^j(0)$$

and

$$\begin{aligned}
\tilde{R}_{32}^j(\tau) = & -\{\tfrac{1}{2}\tau^j\,\tau^k A^k + \tfrac{1}{3}A^{jkl}\,\tau^k\,\tau^l - \tfrac{5}{6}\alpha^{jkm}\, A^{lm}\,\tau^k\,\tau^l\}\, n^{-1} \\
& + \tfrac{1}{3}(\tau^j\,\delta^{kl}[2])\,(\tau^k A^l[2])\, n^{-1} + \tfrac{1}{3}\{(\tau^j\,\delta^{kl}[3])\, A^k A^l + A^{jkl}\,(\tau^k A^l[2])\}\, n^{-\frac{1}{2}} \\
& + \{\tfrac{3}{8}A^{jk}\,A^{kl}\,\tau^k - \tfrac{1}{2}(\tau^j A^k[2])\,(A + n^{-\frac{1}{2}}\tau)^k + \tfrac{1}{3}A^{jkl}\,(\tau^k A^l[2])\}\, n^{-\frac{1}{2}} \\
& + (\tfrac{4}{9}\alpha^{jkn}\,\alpha^{lmn} - \tfrac{1}{4}\alpha^{jklm})\,(\tau^k A^l A^m[3]n^{-\frac{1}{2}} + \tau^k\,\tau^l A^m[3]\, n^{-1} + \tau^k\,\tau^l\,\tau^m\, n^{-\frac{3}{2}}) \\
& - \tfrac{5}{6}\alpha^{jkm}\, A^{lm}\,(\tau^k A^l[2])\, n^{-\frac{1}{2}}.
\end{aligned} \tag{5.A.11}$$

Noticing that

$$n^{\frac{1}{2}}\, E\{\tilde{R}^j(\tau)\}\, k_3^{klm} = O(n^{-\frac{3}{2}}),$$

we have

$$\begin{aligned}
k_4^{jklm} &= n^2\{E(\tilde{R}^j\,\tilde{R}^k\,\tilde{R}^l\,\tilde{R}^m) - E(\tilde{R}^j\,\tilde{R}^k)\,E(\tilde{R}^l\,\tilde{R}^m)[3] \\
&\quad - E(\tilde{R}^j)\,E(\tilde{R}^k\,\tilde{R}^l\,\tilde{R}^m)[4] + 2\,E(\tilde{R}^j)\,E(\tilde{R}^k)\,E(\tilde{R}^l\,\tilde{R}^m) \\
&\quad - 6\,E(\tilde{R}^j)\,E(\tilde{R}^k)\,E(\tilde{R}^l)\,E(\tilde{R}^m)\} \\
&= n^2[E(\tilde{R}_1^j\,\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m) + E\{\tilde{R}_2^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m\}[4] + E\{\tilde{R}_3^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m\}[4] \\
&\quad + E\{\tilde{R}_2^j(\tau)\,\tilde{R}_2^k(\tau)\,\tilde{R}_1^l\,\tilde{R}_1^m\}[6] - E(\tilde{R}_1^j\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[3] \\
&\quad - E\{\tilde{R}_2^j(\tau)\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m\}[12] - E\{\tilde{R}_3^j(\tau)\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m\}[12] \\
&\quad - E\{\tilde{R}_2^j(\tau)\,\tilde{R}_2^k(\tau)\}\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[6] - E\{\tilde{R}_1^j\,\tilde{R}_2^k(\tau)\}\,E\{\tilde{R}_1^l\,\tilde{R}_2^m(\tau)\}] \\
&\quad + O(n^{-\frac{3}{2}}).
\end{aligned} \tag{5.A.12}$$

It may be shown that

$$E(\tilde{R}_1^j\,\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m) - E(\tilde{R}_1^j\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[3] = (\alpha^{jklm} - \delta^{jk}\,\delta^{lm})\, n^{-3}, \tag{5.A.13}$$

$$E(\tilde{R}_{21}^j\,\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m) - E(\tilde{R}_{21}^j\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[3]$$
$$= (-6\,\alpha^{jklm} + 2\,\delta^{jk}\,\delta^{lm} - \tfrac{1}{6}\,\alpha^{jkl}\,\alpha^{mnn}[4] + \tfrac{2}{3}\,\alpha^{jkn}\,\alpha^{lmn}[3])\,n^{-3} + O(n^{-4}),$$

and using the third formula in (3.A.1) of Chapter 3,

$$\begin{aligned}
&E\{\tilde{R}_{22}^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m\} - E\{\tilde{R}_{22}^j(\tau)\,\tilde{R}_1^k\}\,E(\tilde{R}_1^l\,\tilde{R}_1^m\}[3]\\
&= -\tfrac{1}{2}\,\{E(A^{jn}\,A^k\,A^l\,A^m) - E(A^{jn}\,A^k)\,E(A^l\,A^m)[3]\}\,\tau^n\,n^{-\frac{1}{2}}\\
&\quad + \tfrac{1}{3}\,\alpha^{jnp}\,\{E(A^k\,A^l\,A^m\,A^n)\,\tau^p - E(A^k\,A^l)\,E(A^m\,A^n)\,\tau^p\\
&\quad + E(A^k\,A^l\,A^m\,A^p)\,\tau^n - E(A^k\,A^l)\,E(A^m\,A^p)\,\tau^n\}\,n^{-\frac{1}{2}}\\
&= O(n^{-\frac{7}{2}}).
\end{aligned}$$

Thus

$$\begin{aligned}
&E\{\tilde{R}_2^j(\tau)\,\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m\} - E\{\tilde{R}_2^j(\tau)\,\tilde{R}_1^k\}\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[3] \qquad (5.A.14)\\
&= (-6\,\alpha^{jklm} + 2\,\delta^{jk}\,\delta^{lm} - \tfrac{1}{6}\,\alpha^{jkl}\,\alpha^{mnn}[4] + \tfrac{2}{3}\,\alpha^{jkn}\,\alpha^{lmn}[3])\,n^{-3} + O(n^{-\frac{7}{2}}).
\end{aligned}$$

From (5.A.11), we can write

$$\tilde{R}_{32}^j(\tau) = H_1\,n^{-\frac{1}{2}} + H_2\,n^{-1} + H_3\,n^{-\frac{3}{2}},$$

where

$$H_1 = \sum a_{i_1\cdots i_{m_1},j_1\cdots j_{m_2}}(\tau)\,A^{i_1\cdots i_{m_1}}\,A^{j_1\cdots j_{m_2}},$$

$$H_2 = \sum b_{l_1\cdots l_{m_3}}(\tau)\,A^{l_1\cdots l_{m_1}} \quad \text{and} \quad H_3 = \sum c_{n_1\cdots n_{m_4}}(\tau),$$

and $a_{i_1\cdots i_{m_1},j_1\cdots j_{m_2}}(\tau)$, $b_{l_1\cdots l_{m_3}}(\tau)$ and $c_{n_1\cdots n_{m_4}}(\tau)$ are non-random terms only related to τ.

Using the formulae given in (3.A.1) and noting that $\tilde{R}_1^k = A^k$, we see that

$$E(H_1\,n^{-\frac{1}{2}}\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m) - E(H_1\,n^{-\frac{1}{2}}\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[3] = O(n^{-\frac{7}{2}}),$$
$$E(H_2\,n^{-1}\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m) - E(H_2\,n^{-1}\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[3] = O(n^{-4})$$

and

$$E(H_3\,n^{-\frac{3}{2}}\tilde{R}_1^k\,\tilde{R}_1^l\,\tilde{R}_1^m) - E(H_3\,n^{-\frac{3}{2}}\,\tilde{R}_1^k)\,E(\tilde{R}_1^l\,\tilde{R}_1^m)[3] = O(n^{-4}).$$

Therefore

$$E\{\tilde{R}^j_{32}(\tau)\,\tilde{R}^k_1\,\tilde{R}^l_1\,\tilde{R}^m_1\} - E\{\tilde{R}^j_{32}(\tau)\,\tilde{R}^k_1\}\,E(\tilde{R}^l_1\,\tilde{R}^m_1)[3] = O(n^{-\frac{7}{2}}). \qquad (5.A.15)$$

Thus

$$\begin{aligned} & E\{\tilde{R}^j_3(\tau)\,\tilde{R}^k_1\,\tilde{R}^l_1\,\tilde{R}^m_1\} - E\{\tilde{R}^j_3(\tau)\,\tilde{R}^k_1\}\,E(\tilde{R}^l_1\,\tilde{R}^m_1)[3] \\ &= E(\tilde{R}^j_{31}\,\tilde{R}^k_1\,\tilde{R}^l_1\,\tilde{R}^m_1) - E(\tilde{R}^j_{31}\,\tilde{R}^k_1)\,E(\tilde{R}^l_1\,\tilde{R}^m_1)[3] \\ &= \{2\,\alpha^{jklm} - \tfrac{1}{9}\,(\alpha^{jkn}\,\alpha^{lmn}[3])\,n^{-3} + O(n^{-4}). \qquad (5.A.16) \end{aligned}$$

Notice that

$$\begin{aligned} & E\{\tilde{R}^j_2(\tau)\,\tilde{R}^k_2(\tau)\,\tilde{R}^l_1\,\tilde{R}^m_1\} - E\{\tilde{R}^j_2(\tau)\,\tilde{R}^k_2(\tau)\}\,E(\tilde{R}^l_1\,\tilde{R}^m_1) \\ &= E(\tilde{R}^j_{21}\,\tilde{R}^k_{21}\,\tilde{R}^l_1\,\tilde{R}^m_1) - E(\tilde{R}^j_{21}\,\tilde{R}^k_{21})\,E(\tilde{R}^l_1\,\tilde{R}^m_1) \\ &\quad + E\{\tilde{R}^j_{21}\,\tilde{R}^k_{22}(\tau)[2]\,\tilde{R}^l_1\,\tilde{R}^m_1\} - E\{\tilde{R}^j_{21}\,\tilde{R}^k_{22}(\tau)\}[2]\,E(\tilde{R}^l_1\,\tilde{R}^m_1) \\ &\quad + E\{\tilde{R}^j_{22}(\tau)\,\tilde{R}^k_{22}(\tau)[2]\,\tilde{R}^l_1\,\tilde{R}^m_1\} - E\{\tilde{R}^j_{22}(\tau)\,\tilde{R}^k_{22}(\tau)\}[2]\,E(\tilde{R}^l_1\,\tilde{R}^m_1). \qquad (5.A.17) \end{aligned}$$

Using the arguments employed to derive (5.A.15), we may show that

$$E\{\tilde{R}^j_{21}\,\tilde{R}^k_{22}(\tau)[2]\,\tilde{R}^l_1\,\tilde{R}^m_1\} - E\{\tilde{R}^j_{21}\,\tilde{R}^k_{22}(\tau)\}[2]\,E(\tilde{R}^l_1\,\tilde{R}^m_1) = O(n^{-\frac{7}{2}}).$$

Note that

$$\begin{aligned} & E\{\tilde{R}^j_{22}(\tau)\,\tilde{R}^k_{22}(\tau)[2]\,\tilde{R}^l_1\,\tilde{R}^m_1\} - E\{\tilde{R}^j_{22}(\tau)\,\tilde{R}^k_{22}(\tau)\}[2]\,E(\tilde{R}^l_1\,\tilde{R}^m_1) \\ &= \tfrac{1}{4}\,\{E(A^{jn}\,A^{kp}\,A^l\,A^m) - E(A^{jn}\,A^{kp})\,E(A^l\,A^m)\}\,\tau^n\,\tau^p\,n^{-1} \\ &\quad - \tfrac{1}{6}\,\alpha^{kpq}\,[E\{(A^{jn}\,(\tau^p\,A^q\,[2])\,A^l\,A^m\} - E\{(A^{jn}\,(\tau^p\,A^q\,[2])\}\,E(A^l\,A^m)]\,\tau^n\,n^{-1} \\ &\quad - \tfrac{1}{6}\,\alpha^{jpq}\,[E\{(A^{kn}\,(\tau^p\,A^q\,[2])\,A^l\,A^m\} - E\{(A^{kn}\,(\tau^p\,A^q\,[2])\}\,E(A^l\,A^m)]\,\tau^n\,n^{-1} \\ &\quad + \tfrac{1}{9}\,\alpha^{jnp}\,\alpha^{kqs}\,[E\{(\tau^n\,A^p\,[2])\,(\tau^q\,A^s\,[2])\,A^l\,A^m\} \\ &\quad - E\{(\tau^n\,A^p\,[2])\,(\tau^q\,A^s\,[2])\}\,E(A^l\,A^m)]\,n^{-1} \\ &= \tfrac{1}{4}\,(\alpha^{jln}\,\alpha^{kmp} + \alpha^{jmn}\,\alpha^{klp})\,\tau^n\,\tau^p\,n^{-3} \\ &\quad - \tfrac{1}{6}\,\alpha^{kpq}\,\{(\alpha^{jln}\,\delta^{mq} + \alpha^{jmn}\,\delta^{lq})\,\tau^p + (\alpha^{jln}\,\delta^{mp} + \alpha^{jmn}\,\delta^{lp})\,\tau^q\}\,\tau^n\,n^{-3} \end{aligned}$$

$$- \tfrac{1}{6}\, \alpha^{jpq}\, \{(\alpha^{kln}\, \delta^{mq} + \alpha^{kmn}\, \delta^{lq})\, \tau^p + (\alpha^{kln}\, \delta^{mp} + \alpha^{kmn}\, \delta^{lp})\, \tau^q\}\, \tau^n\, n^{-3}$$

$$+ \tfrac{1}{9}\, \alpha^{jnp}\, \alpha^{kqs}\, \{(\tau^n\, \delta^{lp} + \tau^p\, \delta^{ln})(\tau^q\, \delta^{ms} + \tau^s\, \delta^{mq})$$

$$+ (\tau^n\, \delta^{mp} + \tau^p\, \delta^{mn})(\tau^q\, \delta^{ls} + \tau^s\, \delta^{lq})\} + O(n^{-\frac{7}{2}})$$

$$= \tfrac{1}{36}\, (\alpha^{jln}\, \alpha^{kmp} + \alpha^{jmn}\, \alpha^{klp})\, \tau^n\, \tau^p\, n^{-3} + O(n^{-\frac{7}{2}}).$$

Substituting the above results into (5.A.17), obtain that

$$\begin{aligned} & E\{\tilde{R}_2^j(\tau)\, \tilde{R}_2^k(\tau)\, \tilde{R}_1^l\, \tilde{R}_1^m\} - E\{\tilde{R}_2^j(\tau)\, \tilde{R}_2^k(\tau)\}\, E(\tilde{R}_1^l\, \tilde{R}_1^m) \\ &= E(\tilde{R}_{21}^j\, \tilde{R}_{21}^k\, \tilde{R}_1^l\, \tilde{R}_1^m) - E(\tilde{R}_{21}^j\, \tilde{R}_{21}^k)\, E(\tilde{R}_1^l\, \tilde{R}_1^m) \\ &\quad + \tfrac{1}{36}\, (\alpha^{jln}\, \alpha^{kmp} + \alpha^{jmn}\, \alpha^{klp})\, \tau^n\, \tau^p\, n^{-3} + O(n^{-\frac{7}{2}}) \\ &= (3\, \alpha^{jklm} - \delta^{jk}\, \delta^{lm} + \tfrac{1}{6}\, \alpha^{jkl}\, \alpha^{mnn}\, [3] - \tfrac{5}{9}\, \alpha^{jkn}\, \alpha^{lmn}\, [3])\, n^{-3} \\ &\quad + \tfrac{1}{36}\, (\alpha^{jln}\, \alpha^{kmp} + \alpha^{jmn}\, \alpha^{klp})\, \tau^n\, \tau^p\, n^{-3} + O(n^{-\frac{7}{2}}). \end{aligned} \tag{5.A.18}$$

Moreover

$$E\{\tilde{R}_1^j\, \tilde{R}_2^k(\tau)\}\, E\{\tilde{R}_1^l\, \tilde{R}_2^m(\tau)\} = \tfrac{1}{36}\, \alpha^{jkn}\, \alpha^{lmp}\, \tau^n\, \tau^p\, n^{-3} + O(n^{-\frac{7}{2}}). \tag{5.A.19}$$

Substituting (5.A.13), (5.A.14), (5.A.16), (5.A.18) and (5.A.19) into (5.A.12), we finally obtain that

$$k_4^{jklm} = O(n^{-\frac{3}{2}}). \tag{5.A.20}$$

Thus we see that the results in (5.3.3) follow from (5.A.4), (5.A.5), (5.A.7), (5.A.9) and (5.A.20).

Appendix 5.2 Cumulants of $n^{\frac{1}{2}}\, S(\tau)$

Let $\xi_l^{j_1 \cdots j_l}$ denote the joint l'th order cumulants of $n^{\frac{1}{2}}\, S_o(\tau)$. In this part of appendix 5 we give calculations of $\xi_l^{j_1 \cdots j_l}$'s. By (5.3.10),

$$S_o(\tau) = (I - \tfrac{1}{2}\, \Delta_1 + \tfrac{1}{2}\, \Delta_2 + \tfrac{3}{8}\, \Delta_1^2)\, (n^{\frac{1}{2}} A + \tau). \tag{5.A.21}$$

where

$$\Delta_1 = (A^{jk})_{p \times p} \quad \text{and} \quad \Delta_2 = (A^j\, A^k)_{p \times p},$$

and $A = (A^1, \cdots, A^p)$.

From (5.A.21), the j'th component of $S_o(\tau)$ has form

$$S_o^j(\tau) = n^{\frac{1}{2}} \{n^{-\frac{1}{2}} \tau^j + S_1^j + S_2^j(\tau) + S_3^j(\tau)\} \tag{5.A.22}$$

where

$$S_1^j = A^j, \quad S_2^j(\tau) = -\tfrac{1}{2} A^{jk} (A^k + n^{-\frac{1}{2}} \tau^k),$$
$$S_3^j(\tau) = \tfrac{1}{2} A^j A^k (A^k + n^{-\frac{1}{2}} \tau^k) + \tfrac{3}{8} A^{jk} A^{kl} (A^l + n^{-\frac{1}{2}} \tau^l). \tag{5.A.23}$$

To calculate ξ_1^j, we note that

$$E(S_1^j) = 0, \quad E\{S_2^j(\tau)\} = -\tfrac{1}{2} \alpha^{jkk} \text{ and} \tag{5.A.24}$$
$$E\{S_3^j(\tau)\} = \tfrac{1}{2} \delta^{jk} \tau^k n^{-\frac{3}{2}} + \tfrac{3}{8} (\alpha^{jkkl} - \tau^{jl}) \tau^l n^{-1} + O(n^{-\frac{3}{2}}).$$

Thus

$$\xi_1^j = n^{\frac{1}{2}} E\{S_o^j(\tau)\}$$
$$= \tau^j - \tfrac{1}{2} \alpha^{jkk} n^{-\frac{1}{2}} + (\tfrac{1}{8} \tau^j + \tfrac{3}{8} \alpha^{jkkl} \tau^l) n^{-1} + O(n^{-\frac{3}{2}}). \tag{5.A.25}$$

Notice that

$$\xi_2^{jk} = n\,[E(S_1^j S_1^k) + E\{S_1^j S_2^k(\tau)\}[2] + E\{S_1^j S_3^k(\tau)\}[2] + E\{S_2^j(\tau) S_2^k(\tau)\}$$
$$- E\{S_1^j + S_2^j(\tau) + S_3^j(\tau)\} E\{S_1^k + S_2^k(\tau) + S_3^k(\tau)\}]. \tag{5.A.26}$$

Observe that

$$E(S_1^j S_1^k) = \delta^{jk} n^{-1},$$
$$E\{S_1^j S_2^k(\tau)\} = -\tfrac{1}{2} E\{A^j A^{kl} (A^l + n^{-\frac{1}{2}} \tau^l)\}$$
$$= -\tfrac{1}{2} \{\alpha^{jkl} \tau^l n^{-\frac{3}{2}} + (\alpha^{jkll} - \delta^{jk}) n^{-2}\},$$
$$E\{S_1^j S_3^k(\tau)\} = \tfrac{1}{2} E\{A^j A^k A^l (A^l + n^{-\frac{1}{2}} \tau^l)\} + \tfrac{3}{8} E\{A^j A^{kl} A^{lm} (A^m + n^{-\frac{1}{2}} \tau^m)\}$$
$$= \{\tfrac{1}{2} (p+2) \delta^{jk}) + \tfrac{3}{8} (\alpha^{jkll} - \delta^{jk} + \alpha^{jkl} \alpha^{lmm} + \alpha^{jlm} \alpha^{klm})\} n^{-2}$$
$$+ O(n^{-\frac{5}{2}}),$$

and

$$E\{S_2^j(\tau)\,S_2^k(\tau)\} = \tfrac{1}{4}\,E\{A^{jl}\,(A^l + n^{-\frac{1}{2}}\,\tau^l)\,A^{km}\,(A^m + n^{-\frac{1}{2}}\,\tau^m)\}$$

$$= \tfrac{1}{4}\,(\alpha^{jkll} - \delta^{jk} + \alpha^{jll}\,\alpha^{kmm} + \alpha^{jlm}\,\alpha^{klm}$$

$$+ \alpha^{jklm}\,\tau^l\,\tau^m - \tau^j\,\tau^k)\,n^{-2} + O(n^{-\frac{5}{2}}).$$

Substituting the above formulae and (5.A.24) into (5.A.26), we obtain that

$$\xi_2^{jk} = \delta^{jk} - \alpha^{jkl}\,\tau^l\,n^{-\frac{1}{2}}$$

$$+ \{(p+2)\,\delta^{jk} + \tfrac{3}{4}\,\alpha^{jkl}\,\alpha^{lmm} + \alpha^{jlm}\,\alpha^{klm}\}\,n^{-1}$$

$$+ \tfrac{1}{4}\,(\alpha^{jklm}\,\tau^l\,\tau^m - \tau^j\,\tau^k)\}\,n^{-1} + O(n^{-\frac{3}{2}}). \qquad (5.A.27)$$

To calculate ξ_3^{jkl} we observe that

$$\xi_3^{jkl} = n^{\frac{3}{2}}\,[E(S_1^j\,S_1^k\,S_1^l) + E\{S_2^j(\tau)\,S_1^k\,S_1^l\}[3] + E\{S_3^j(\tau)\,S_1^k\,S_1^l\}[3]$$

$$+ E\{S_1^j\,S_2^k(\tau)\,S_2^l(\tau)\}[3] - E\{S_2^j(\tau)\}\,E(S_1^k\,S_1^l)[3] - E\{S_3^j(\tau)\}\,E(S_1^k\,S_1^l)[3]$$

$$- E\{S_2^j(\tau)\}\,E(S_2^k\,S_1^l)[6]] + O(n^{-\frac{3}{2}}). \qquad (5.A.28)$$

Using (3.A.1) we may show that

$$E(S_1^j\,S_1^k\,S_1^l) = \alpha^{jkl}\,n^{-2},$$

$$E\{S_2^j(\tau)\,S_1^k\,S_1^l\} = -\tfrac{1}{2}\,E\{A^{jm}\,A^k\,A^l\,(A^m + n^{-\frac{1}{2}}\,\tau^m)\}$$

$$= -\tfrac{1}{2}\,\{(\alpha^{jmm}\,\delta^{kl} + 2\,\alpha^{jkl})\,n^{-2} + (\alpha^{jklm}\,\tau^m - \tau^j\,\delta^{kl})\,n^{-\frac{5}{2}}\} + O(n^{-3}),$$

$$E\{S_3^j(\tau)\,S_1^k\,S_1^l\} = \tfrac{1}{2}\,E\{A^j\,A^k\,A^l\,A^m\,(A^m + n^{-\frac{1}{2}}\,\tau^m)\}$$

$$+ \tfrac{3}{8}\,E\{A^{jm}\,A^k\,A^l\,A^{mn}\,(A^n + n^{-\frac{1}{2}}\,\tau^n)\}$$

$$= \tfrac{1}{2}\,(\tau^j\,\delta^{kl}\,[3])\,n^{-\frac{5}{2}} + \tfrac{1}{8}\,(\alpha^{jmmn}\,\delta^{kl}\,\tau^n - \tau^j\,\delta^{kl})\,n^{-\frac{5}{2}}$$

$$+ \tfrac{1}{8}(\alpha^{jkm}\,\alpha^{lmn} + \alpha^{jlm}\,\alpha^{kmn})\,\tau^n\,n^{-\frac{5}{2}} + O(n^{-3})$$

and

$$E\{S_1^j\,S_2^k(\tau)\,S_2^l(\tau)\} = \tfrac{1}{4}\,E\{A^j\,A^{km}\,A^{ln}\,(A^m + n^{-\frac{1}{2}}\,\tau^m)(A^n + n^{-\frac{1}{2}}\,\tau^n)\}$$

$$= \tfrac{1}{4}\,\{E(A^j\,A^{km}\,A^{ln}\,A^m)\,\tau^n + E(A^j\,A^{km}\,A^{ln}\,A^n)\,\tau^m\}\,n^{-\frac{1}{2}}$$

$$+ O(n^{-3})$$

$$= \tfrac{1}{4}\,\{2\,\alpha^{jklm}\,\tau^m - \tau^k\,\delta^{jl} - \tau^l\,\delta^{jk} + \alpha^{jkm}\,\alpha^{lnn}\,\tau^m$$

$$+ \alpha^{jlm}\,\alpha^{knn}\,\tau^m + \alpha^{jkn}\,\alpha^{lmn}\,\tau^m + \alpha^{jln}\,\alpha^{kmn}\,\tau^m\}\,n^{-\frac{5}{2}}$$

$$+ O(n^{-3}).$$

Moreover, using the formulae given in the calculations of ξ_1^j and ξ_2^{jk} we know that

$$E\{S_2^j(\tau)\}\,E(S_1^k\,S_1^l) = -\tfrac{1}{2}\,\alpha^{jmm}\,\delta^{kl}\,n^{-2} + O(n^{-3}),$$

$$E\{S_3^j(\tau)\}\,E(S_1^k\,S_1^l) = (\tfrac{1}{8}\,\tau^j\,\delta^{kl} + \tfrac{3}{8}\,\alpha^{jmmn}\,\delta^{kl}\,\tau^n)\,n^{-\frac{5}{2}} + O(n^{-3})$$

and

$$E\{S_2^j(\tau)\}\,E\{S_2^k(\tau)\,S_1^l\,[2]\} = \tfrac{1}{2}\,\alpha^{jmm}\,\alpha^{kln}\,\tau^n\,n^{-\frac{5}{2}} + O(n^{-3}).$$

Substituting the above formulae into (5.A.28), we have

$$\xi_3^{jkl} = -2\,\alpha^{jkl}\,n^{-\frac{1}{2}} + (\tau^j\,\delta^{kl}\,[3] + \tfrac{5}{4}\,\alpha^{jkm}\,\alpha^{lmn}\,\tau^m\,[3])\,n^{-1} + O(n^{-\frac{3}{2}}). \qquad (5.A.29)$$

By our definition,

$$\begin{aligned}
\xi_4^{jklm} = {} & n^2\,[E(S_1^j\,S_1^k\,S_1^l\,S_1^m) + E\{S_2^j(\tau)\,S_1^k\,S_1^l\,S_1^m\}\,[4] + E\{S_3^j(\tau)\,S_1^k\,S_1^l\,S_1^m\}\,[4] \\
& + E\{S_2^j(\tau)\,S_2^k\,S_1^l\,S_1^m\}\,[6] - E(S_1^j\,S_1^k)\,E(S_1^l\,S_1^m)[3] - E\{S_2^j(\tau)\,S_1^k\}\,E(S_1^l\,S_1^m)[12] \\
& - E\{S_3^j(\tau)\,S_1^k\}\,E(S_1^l\,S_1^m)[12] - E\{S_2^j(\tau)\,S_2^k(\tau)\}\,E(S_1^l\,S_1^m)[6] \\
& - E\{S_2^j(\tau)\,S_1^l\}\,E\{S_2(\tau)^k\,S_1^m\}[12] - E\{S_2^j(\tau)\}\,E(S_1^k\,S_1^l\,S_1^m)\,[4] \\
& - E\{S_2^j(\tau)\}\,E(S_2^k(\tau)\,S_1^l\,S_1^m)\,[12] + 2\,E\{S_2^j(\tau)\}\,E\{S_2^k(\tau)\}\,E(S_1^l\,S_1^m)\,[6]] \\
& + O(n^{-\frac{3}{2}}). \qquad (5.A.30)
\end{aligned}$$

Using (3.A.1) agian, we have

$$E(S_1^j\,S_1^k\,S_1^l\,S_1^m) - E(S_1^j\,S_1^k)\,E(S_1^l\,S_1^m)[3] = (\alpha^{jklm} - \delta^{jk}\,\delta^{lm}\,[3])\,n^{-3},$$

$$\begin{aligned}
& E\{S_2^j(\tau)\,S_1^k\,S_1^l\,S_1^m\} - E\{S_2^j\,S_1^k\}\,E(S_1^l\,S_1^m)[3] \\
= & -\tfrac{1}{2}\,[E\{A^{jn}\,(A^n + n^{-\frac{1}{2}}\,\tau^n)\,A^k\,A^l\,A^m\} - E\{A^{jn}\,(A^n + n^{-\frac{1}{2}}\,\tau^n)\,A^k\}\,E(A^l\,A^m)\,[3]] \\
= & -\tfrac{1}{2}\,\{E(A^{jn}\,A^n\,A^k\,A^l\,A^m\} - E(A^{jn}\,A^n\,A^k)\,E(A^l\,A^m)\,[3]\} + O(n^{-\frac{7}{2}}) \\
= & -\tfrac{1}{2}\,(12\,\alpha^{jklm} - 4\,\delta^{jk}\,\delta^{lm}\,[3] + \alpha^{jnn}\,\alpha^{klm}\,[4] + 4\,\alpha^{jkn}\,\alpha^{lmn}\,[3])\,n^{-3} + O(n^{-\frac{7}{2}}),
\end{aligned}$$

$$E\{S_3^j(\tau)\, S_1^k\, S_1^l\, S_1^m\} - E\{S_3^j(\tau)\, S_1^k\}\, E(S_1^l\, S_1^m)[3]$$

$$= \tfrac{1}{2}\, [E\{A^j\, A^n\, (A^n + n^{-\frac{1}{2}}\, \tau^n)\, A^k\, A^l\, A^m\}$$

$$- E\{A^j\, A^n\, (A^n + n^{-\frac{1}{2}}\, \tau^n)\, A^k\}\, E(A^l\, A^m)\, [3]]$$

$$+ \tfrac{3}{8}\, [E\{A^{jp}\, A^{np}\, (A^n + n^{-\frac{1}{2}}\, \tau^n)\, A^k\, A^l\, A^m\}$$

$$- E\{A^{jp}\, A^{np}\, (A^n + n^{-\frac{1}{2}}\, \tau^n)\, A^k\}\, E(A^l\, A^m)\, [3]\}]$$

$$= \tfrac{1}{2}\, \{E(A^j\, A^n\, A^n\, A^k\, A^l\, A^m) - E(A^j\, A^n\, A^n\, A^k)\, E(A^l\, A^m)\, [3]\}$$

$$+ \tfrac{3}{8}\, \{E\{A^{jp}\, A^{np}\, A^n\, A^k)\, E(A^l\, A^m) - E(A^{jp}\, A^{np}\, A^n\, A^k\}\, E(A^l\, A^m)\, [3]\}\}$$

$$+ O(n^{-\frac{7}{2}})$$

$$= (4\, \delta^{jk}\, \delta^{lm}\, [3] + 3\, \alpha^{jkn}\, \alpha^{lmn}\, [3])\, n^{-3} + O(n^{-\frac{7}{2}})$$

and

$$E\{S_2^j(\tau)\, S_2^k\, S_1^l\, S_1^m\}\, [6] - E\{S_2^j(\tau)\, S_2^k(\tau)\}\, E(S_1^l\, S_1^m)[6]$$

$$- E\{S_2^j(\tau)\, S_1^l\}\, E\{S_2^k(\tau)\, S_2^m(\tau)\}[12]$$

$$= \tfrac{1}{4}\, [E\{A^{jn}\, A^{kp}\, (A^n + n^{-\frac{1}{2}}\, \tau^n)(A^p + n^{-\frac{1}{2}}\, \tau^p)\, A^l\, A^m\}$$

$$- E\{A^{jn}\, A^{kp}\, (A^n + n^{-\frac{1}{2}}\, \tau^n)(A^p + n^{-\frac{1}{2}}\, \tau^p)\}\, E(A^l\, A^m)]\, [6]$$

$$- \tfrac{1}{4}\, \alpha^{jkn}\, \alpha^{lmp}\, \tau^n\, \tau^p\, [12]\, n^{-3}$$

$$= (3\, \alpha^{jklm} - \delta^{jk}\, \delta^{lm}\, [3] + \tfrac{3}{2}\, \alpha^{jnn}\, \alpha^{klm}\, [4] + 3\, \alpha^{jkn}\, \alpha^{lmn}\, [3])\, n^{-3} + O(n^{-\frac{7}{2}}).$$

Furthermore, from the calculation of ξ_3^{jkl} displayed earlier we know that

$$E\{S_2^j(\tau)\}\, E(S_1^k\, S_1^l\, S_1^m)\, [4] + E\{S_2^j(\tau)\}\, E\{S_2^k(\tau)\, S_1^l\, S_1^m\}\, [12]$$

$$= -\tfrac{1}{2}\, \alpha^{jnn}\, \{\alpha^{klm} - \tfrac{1}{2}\, (\alpha^{kpp}\, \delta^{lm} + 2\, \alpha^{klm})\, [3]\}\, [4]\, n^{-3} + O(n^{-\frac{7}{2}})$$

$$= (\alpha^{jnn}\, \alpha^{klm}\, [4] + \tfrac{1}{2}\, \alpha^{jnn}\, \alpha^{kpp}\, \delta^{lm}\, [6])\, n^{-3} + O(n^{-\frac{7}{2}})$$

and

$$E\{S_2^j(\tau)\}\, E\{S_2^k(\tau)\}\, E(S_1^l\, S_1^m) = \tfrac{1}{4}\, \alpha^{jnn}\, \alpha^{kpp}\, \delta^{lm}\, n^{-3}.$$

Substituting the above results into (5.A.30), we obtain that

$$\xi_4^{jklm} = (-2\, \alpha^{jklm} + 4\, \delta^{jk}\, \delta^{lm}\, [4] + 4\, \alpha^{jkn}\, \alpha^{lmn}\, [3])\, n^{-1} + O(n^{-\frac{3}{2}}). \qquad (5.A.31)$$

Finally using the results given by James and Mayne (1962),

$$\xi_l^{j_1 \cdots j_l} = O(n^{-\frac{3}{2}}) \quad \text{for} \quad l \geq 5. \qquad (5.A.32)$$

In summary of (5.A.25), (5.A.27), (5.A.29), (5.A.31) and (5.A.32), we obtain (5.3.12) and (5.3.13).

REFERENCES

BARNDORFF-NIELSEN, O.E. and COX, D.R. (1989). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London and New York.

BARNDORFF-NIELSEN, O.E. and HALL, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistics. *Biometrika* **75**, 374-378.

BARTLETT, M.S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Soc. A* **160**, 268-82.

BHATTACHARYA, R.N. (1968). Berry-Esseen bounds for the multi-dimensional central limit theorem. *Bull. Am. Math. Soc.* **75**, 68-86.

BHATTACHARYA, R.N. and GHOSH, J.K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6**, 434 - 451.

BHATTACHARYA, R.N. and RAO, R.R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

DAVID, H.A. (1981). *Order Statistics.* Wiley, New York.

DICICCIO, T.J., HALL, P., and ROMANO, J.P (1988). Bartlett adjustment for empirical likelihood. Tech. Report No. 298. Depart. of Statistics, Stanford University.

DICICCIO, T.J., HALL, P., and ROMANO, J.P (1991). Bartlett adjustment for empirical likelihood. *Ann. Statist.* **19**, 1053-1061.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

ESSEEN, C.G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta. Math.* **77**, 1-125.

HALL, P. (1990). Pseudo-likelihood theory for empirical likelihood. *Ann. Statist.*

18, 121-140.

HALL, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *statistics* **22** 2, 215-232. Akademie Verlag.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag.

HALL, P. and LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *International Statist. Review.* **58**, 2, 109-127.

HÄRDLE, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

JAMES, G.S. and MAYNE, A.J. (1962). Cumulants of functions of random variables. *Sankhya, Ser. A* **24**, 47–54.

JOHNSON, N.I. and KOTZ, S. (1969). *Discrete Distributions.* Houghton Miffin Corp..

KOLACZYK, E.D. (1992). Empirical likelihood for generalized linear models. Tech. Report No. 389. Depart. of Statistics, Stanford University.

LAWLEY, D. N. (1956). A general method for approximating the distribution of the likelihood ratio criteria. *Biometrika* **43**, 295-303.

OWEN, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237-249.

OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90-120.

OWEN, A (1991) Empirical likelihood for linear model. *Ann. Statist.* **19** 1725-1747.

PRESS, W.H., FLANNERY, B.F., TEUKOLSKY, S.A. and VETTERLING, W.T. (1989). *Numerical Recipes: the Art of Scientific Computing.* Cambridge University Press, Cambridge

SKOVGAARD, IB,M. (1981). Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist.* **8**, 207-217.

THOMAS, D.R. and GRUNKEMEIER, G.L. (1975). Confidedence interval estimation of survival probabilities for censored data. *J. Am. Statist. Assoc.* **70**, 865-871.

WILKS, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* **9**, 60-62.

WU, C. F. J. (1986). Jacknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1261- 1350.