8-2016

Maximum empirical likelihood estimation in U-statistics based general estimating equations

Lingnan Li
Purdue University

Graduate School Form
30 Updated 12/26/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Lingnan Li

Entitled
Maximum Empirical Likelihood Estimation in U-statistics Based General Estimating Equations

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Hanxiang Peng
Chair

Benzion Boukai

Guang Cheng

Wei Zheng

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Hanxiang Peng

Approved by: Evgeny Mukhin 7/23/2016

Head of the Departmental Graduate Program Date

MAXIMUM EMPIRICAL LIKELIHOOD ESTIMATION IN U-STATISTICS

BASED GENERAL ESTIMATING EQUATIONS

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Lingnan Li

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

August 2016

Purdue University

Indianapolis, Indiana

ACKNOWLEDGMENTS

I want to acknowledge the effort from Professor Hanxiang Peng, who graciously became to be my thesis advisor four years ago and inspired me to start the research in the area of empirical likelihood.

I would like to thank Professor Fei Tan, who helped me on the projects; to thank Professor Benzion Boukai, who admitted me to the statistics program long time ago and supported me through the time; to thank Professor Jyoti Sarkar, Professor Fang Li, and Professor Wei Zheng; and to thank Professor Guang Cheng for serving on my dissertation committee.

I would also like to acknowledge the effort from Professor Zhongmin Shen, who helped me become part of the Ph.D. program and one of Professor Peng's students.

Finally, I would like to thank the Department of Mathematical Sciences of IUPUI, without whom this thesis would not have been possible at all.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

SYMBOLS

$X, Y, Z, \ldots$	column random variables (vectors)
$x, y, z, \ldots$	realizations of $X, Y, Z, \ldots$
$a, b, c, \ldots$	constants
$\alpha, \beta, \theta, \ldots$	column parameter vectors
$N(\theta_0)$	a neighborhood of θ_0
$\mathscr{N}(\mu, \Sigma)$	normal distribution with μ and variance (matrix) Σ
$\mathbb{R}$	the real space
$\mathbb{R}^m$	the m dimensional real (column) vector space
$\mathscr{R}_n, \hat{\mathscr{R}}_n$	empirical likelihood (ratio)
$\pi_1, \ldots, \pi_n$	probability weights
P, E, Var, Cov	probability, expected value, variance, covariance
$f, g, h, \ldots$	column vector functions
$\tilde{f}$	the centered vector function of f i.e. $\tilde{f} = f - E(f(Z))$
$U_{nm}(h)$	a U-statistic of order m with kernel h
θ_0	the true value of parameter
Θ	the parameter space
$\|a\|$	the euclidean norm of a vector a
$a^\top$	the transpose of a vector a
$A^\top$	the transpose of a matrix A
A^{-1}	the inverse of a matrix A
$A^{\otimes 2}$	$AA^\top$
$\dot{f}(z;\theta)$	$\frac{\partial}{\partial\theta} f(z;\theta) \in \mathbb{R}^k$ (column vector) for $\theta \in \mathbb{R}^k, f \in \mathbb{R}$
$\frac{\partial}{\partial\theta^\top} f(z;\theta)$	a row vector of partial derivatives for $\theta \in \mathbb{R}^k, f \in \mathbb{R}$
$\dot{f}(z;\theta)$	$\frac{\partial}{\partial\theta^\top} f(z;\theta)$ the $m \times k$ matrix for $\theta \in \mathbb{R}^k, f \in \mathbb{R}^m$
$\dot{f}(z;\theta)^\top$	$\frac{\partial}{\partial\theta} f(z;\theta)^\top$ the $k \times m$ matrix for $\theta \in \mathbb{R}^k, f \in \mathbb{R}^m$

$\ddot{f}(z;\theta)$	$\frac{\partial^2}{\partial\theta\partial\theta^\top} f(z;\theta) = \frac{\partial}{\partial\theta^\top}\frac{\partial}{\partial\theta} f(z;\theta)$ the $k \times k$ matrix for $\theta \in \mathbb{R}^k, f \in \mathbb{R}$
$\to_p$	convergence in probability
$\Rightarrow$	convergence in distribution

ABBREVIATIONS

a.s.	almost sure convergence
DF	Distribution Function
CDF	Cumulative Distribution Function
EDF	Empirical Distribution Function
EL	Empirical Likelihood
GEE	General Estimating Equation
i.i.d.	Independent and Identically Distributed
JEL	Jackknife Empirical Likelihood
MELE	Maximum Empirical Likelihood Estimate
MVUE	Minimum Variance Unbiased Estimate
r.v.	Random Variable
UGEE	U-statistics Based General Estimating Equation

ABSTRACT

Li, Lingnan Ph.D., Purdue University, August 2016. Maximum Empirical Likelihood Estimation in U-statistics Based General Estimating Equations. Major Professor: Hanxiang Peng.

In the first part of this thesis, we study maximum empirical likelihood estimates (MELE's) in U-statistics based general estimating equations (UGEE's). Our technical maneuver is the jackknife empirical likelihood (JEL) approach. We give the local uniform asymptotic normality condition for the log-JEL for UGEE's. We derive the estimating equations for finding MELE's and provide their asymptotic normality. We obtain easy MELE's which have less computational burden than the usual MELE's and can be easily implemented using existing software. We investigate the use of side information of the data to improve efficiency. We exhibit that the MELE's are fully efficient, and the asymptotic variance of a MELE will not increase as the number of UGEE's increases. We give several important examples and demonstrate that efficient estimates of moment based distribution characteristics in the presence of side information can be obtained using JEL for U-statistics.

In the second part, we propose several JEL goodness-of-fit tests for spherical symmetry, rotational symmetry, antipodal symmetry, coordinatewise symmetry and exchangeability. We employ the jackknife empirical likelihood for vector U-statistics to incorporate side information. We use estimated constraint functions and allow the number of constraints and the dimension to grow with the sample size so that these tests can be used to test hypotheses for high dimensional symmetries. We demonstrate that these tests are distribution free and asymptotically chisquare distributed. We conduct extensive simulations to evaluate the performance of these tests.

1. INTRODUCTION

Empirical likelihood (EL) is a data-driven likelihood approach with nonparametric nature which is effective and requires few assumptions about the distribution of the data. Owen (1988, 1990, 1991) showed that the empirical likelihood ratio statistics have the limiting chi-square distribution under mild conditions. He also demonstrated that tests and confidence intervals can be constructed. The empirical likelihood theory has been successfully extended to various areas of statistics with tremendous accomplishments. These include Bartlett correction (DiCiccio, *et al.*, 1991), generalized linear models (Kolaczyk, 1994), heteroscedastic partially linear models (Lu, 2009), partially linear models (Shi and Lau, 2000; Wang and Jing, 2003), parametric and semiparametric models in multiresponse regression (Chen and Van Keilegom, 2009), right censored data (Li and Wang, 2003), U-statistics with side information (Yuan, *et al.*, 2012), and stratified samples with nonresponse (Fang, *et al.*, 2009). Qin and Lawless (1994) linked empirical likelihood with finitely many estimating equations and investigated maximum empirical likelihood estimators. Chen, *et al.* (2009) obtained asymptotic normality for the number of constraints growing to infinity. Hjort, *et al.* (2009) and Peng and Schick (2013a, 2013b) generalized the empirical likelihood approach to allow for the number of constraints to grow with the sample size and for the constraints to use estimated criteria functions. Algorithms, calibration and higher-order precision of the approach can be found in Hall and La Scala (1990), Emerson and Owen (2009) and Liu and Chen (2010) among others.

In Owen's homepage (http://statweb.stanford.edu/∼owen/empirical/) software can be found. Here are two algorithms from this site: *scel.R* (R function to compute empirical likelihood using a self-concordant convex criterion) and *el.R* (Mai Zhou's R code for empirical likelihood, with an emphasis on survival analysis).

U-statistics is a class of statistics which is especially useful in estimation. Many popular statistics such as high order moments information can be expressed by U-statistics, see e.g. Serfling (1980), Kowalski and Tu (2008) and Lee (1990). Yuan, *et al.* (2012) explored maximum empirical likelihood estimates (MELE's) in U-statistics with side information. However, usual EL method runs into serious computational difficulties when it's applied to U-statistics. U-statistics are not independent but correlated so that they do not satisfy the independence or at least asymptotic independence which is assumed by the definition of empirical likelihood. Moreover, unlike the usual empirical likelihood, the nonlinearality of EL weights π_j's in the constraints equations results in that there are no explicit solutions for the EL weights. Jing, *et al.* (2009) identified the asymptotic independence of the jackknife pseudo values of a U-statistic and introduced their jackknife empirical likelihood (JEL) for U-statistics, and showed its effectiveness in handling one- or two-sample U-statistics. Some nice properties of the jackknife pseudo values of a U-statistic were exploited to establish the Wilks theorems for their cases. For example, the average of the jackknife pseudo values is equal to the U-statistic, and the sample variance of them is an asymptotically unbiased estimator of the asymptotic variance of the U-statistic.

Motivated by applications to goodness of fit U-statistic testing, Peng and Tan (2016) gave two approaches to justify the JEL for vector U-statistics and proved the Wilks theorems. They extended empirical likelihood for general estimating equations (GEE's) to U-statistics based general estimating equations (UGEE's). The results were extended to allow for the use of estimated constraints and for the number of constraints to grow with the sample size. They exhibited that the JEL can be used to construct EL tests for moment based distribution characteristics (e.g. skewness, coefficient of variation) with less computational burden and more flexibility than the usual EL. This can be done in the U-statistic representation approach and the vector U-statistic approach which were illustrated with several examples including JEL tests for Pearson's correlation, Goodman-Kruskal's Gamma, overdisperson, U-quantiles, variance components, and simplicial depth. They showed that tests are asymptotically distribution free. They ran simulations to exhibit power improvement of the tests with incorporation of side information.

Soon it was realized that it can also be used to construct point estimators. Qin and Lawless (1994) linked empirical likelihood with GEE's and investigated maximum empirical likelihood estimators (MELE's). They established consistency and asymptotic normality of MELE's under the usual regularity conditions, and demonstrated that the variance of a MELE will not increase when the number of estimating equations is increased. Furthermore, they showed that MELE's are fully semiparametrically efficient in the sense of least dispersed regular estimators (Bickel, *et al.* (1993), Van der Vaart (2000)). Peng and Schick (2013) explored MELE's in the case of constraint functions that may be discontinuous and/or depend on additional parameters. The latter is the case in applications to semiparametric models where the constraint functions may depend on the nuisance parameter. Zhang (1995, 1997) used the method of MELE's to construct improved estimates in M-estimation and quantile processes with the availability of auxiliary (side) information. He established consistency and asymptotic normality, and proved that the asymptotic variances of the resulting estimators are smaller than those of the usual sample M-estimators and sample quantiles. It was utilized by Hellerstein and Imbens (1999) for the least squares estimators in a linear regression model and the application to a real data set was presented. These authors dealt with finitely many of constraints. Peng and Schick (2013) has employed on-step estimator to construct MELE's. Peng (2015) developed a class of easy MELE's which is computationally more efficient. Recently, Tang and Leng (2012) used this idea to construct more efficient estimators of parameters in quantile regression.

In this thesis, we study MELE's in UGEE's and their asymptotic behaviors. Our technical maneuver is the jackknife empirical likelihood approach. It is well known that the asymptotic behaviors of the U-statistic $U_n(h)$ is determined by h_1 (see Chapter 2). Here we shall apply the theory of Qin and Lawless (1994) on h_1 to derive the asymptotic behaviors of the MELE's of the JEL for UGEE's. These results for the UGEE's are parallel to those of Qin and Lawless (1994). We obtain the uniform local asymptotic normality for the logarithm of the JEL ratio in Chapter 3. We derive the estimating equations for the MELE's in UGEE's in Chapter 4. Here we also give a class of easy MELE's and establish their asymptotic distribution. In Chapter 5, we provide a number of examples. Here we demonstrate that efficient estimates of moment based distribution characteristics in the presence of side information can be obtained using JEL for U-statistics. In Chapter 6, we propose several JEL tests for various multivariate and high dimensional symmetries. Some of the technical details are provided in Chapter 7.

2. JACKKNIFE EMPIRICAL LIKELIHOOD FOR VECTOR U-STATISTICS

In this chapter, we recall some facts about one-sample multivariate U-statistics and introduce the JEL approach.

2.1 Vector U-statistics

Let $(\mathscr{Z}, \mathscr{S})$ be a measurable space and P be a probability measure on this space. Let $Z_1, \ldots, Z_n$ be independent copies of a $\mathscr{Z}$-valued random variable Z with cumulative distribution function F under P. Let $h : \mathbb{R}^m \mapsto \mathbb{R}^d$ be a known function that is permutation symmetric in its m arguments. $\theta \in \Theta$ is a parameter we are interested in. A multivariate or vector U-statistic with kernel h of order m is defined as

$$U_n =: U_{nm}(h) =: \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \ldots < i_m \leq n} h(Z_{i_1}, \ldots, Z_{i_m}; \theta), \quad n \geq 2.$$

Throughout we assume h is F^m-square integrable, that is, $h \in L_2(F^m)$, where $L_2(F^m) = \left\{f : \int \|f\|^2 \, dF^m < \infty\right\}$, where $\|v\|$ denotes the euclidean norm of vecotr v. We assume throughout that

$$E(h(;\theta)) =: E(h(Z_1, \ldots, Z_m; \theta)) = 0. \tag{2.1.1}$$

This of course implies $E(U_n) = 0$. Also, we shall abbreviate $P_n f = n^{-1} \sum_{j=1}^{n} f(Z_j)$ and $Pf = E(f(Z))$. Let $h_m = h$ and

$$h_c(z_1, \ldots, z_c; \theta) = E(h(z_1, \ldots, z_c, Z_{c+1}, \ldots, Z_m; \theta)), \quad c = 1, \ldots, m-1.$$

Then h_c is a version of the conditional expectation, that is,

$$h_c(z_1, \ldots, z_c; \theta) = E(h(Z_1, \ldots, Z_m; \theta) | Z_1 = z_1, \ldots, Z_c = z_c).$$

Let δ_z be the point mass at $z \in \mathcal{Z}$. We now define

$$h_c^*(z_1,\ldots,z_c) = (\delta_{z_1} - P)\ldots(\delta_{z_c} - P)P^{m-c}h, \quad c = 0,1,\ldots,m.$$

Throughout we let

$$\tilde{f} = f - Pf$$

be the centered version of a function f for which Pf is well defined. Clearly $h_1^* = \tilde{h}_1$.

2.2 JEL for vector U-statistics

Let $U_{n-1}^{(-j)}$ denote the U-statistic based on the $n-1$ observations $Z_1, \ldots, Z_{j-1}$, $Z_{j+1}, \ldots, Z_n$. The jackknife pseudo values of the U-statistic are defined as

$$V_{nj} = nU_n - (n-1)U_{n-1}^{(-j)}, \quad j = 1,\ldots,n.$$

Let $R_{nj} = \tilde{V}_{nj} - m\tilde{h}_1(Z_j;\theta)$. It has been shown in (4.6) of Peng and Tan (2016) that each component of R_{nj} is of $O_p(n^{-1/2})$, hence

$$\tilde{V}_{nj}(\theta) = m\tilde{h}_1(Z_j,\theta) + O_p(n^{-1/2}), \quad j = 1,\ldots,n.$$

As argued in Peng and Tan (2016), this shows that each jackknife value $\tilde{V}_{nj}$ depends asymptotically on Z_j, so that $\tilde{V}_{nj}, j = 1,\ldots,n$ are asymptotically *independent*. As a result, if π_j is a probability mass placed at Z_j, then approximately the same probability mass π_j is placed at the jackknife value $\tilde{V}_{nj}$ for $j = 1,\ldots,n$; because of the the asymptotic independence of the jackknife values, the joint likelihood is approximately the product of these π_j's. Consequently, it is justified to introduce the JEL for the vector U-statistic $U_n(h)$ as follows:

$$\hat{\mathscr{R}}_n(\theta) = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j\tilde{V}_{nj}(\theta) = 0\Big\}, \quad \theta \in \Theta, \qquad (2.2.1)$$

where $\mathscr{P}_n$ denotes the closed probability simplex in dimension n, i.e.

$$\mathscr{P}_n = \big\{\boldsymbol{\pi} = (\pi_1,\ldots,\pi_n)^\top \in [0,1]^n : \pi_1 + \cdots + \pi_n = 1\big\}.$$

Using Lagrange multipliers, Owen (1988) derived

$$\pi_j = \frac{1}{n}\frac{1}{1+\xi^\top \tilde{V}_{nj}(\theta)}, \quad j = 1, \dots, n, \tag{2.2.2}$$

where ξ satisfies the equation

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\tilde{V}_{nj}(\theta)}{1+\xi^\top \tilde{V}_{nj}(\theta)} = 0. \tag{2.2.3}$$

As (2.1.1) holds, $\tilde{V}_{nj} = V_{nj}$. For notational brevity, we sometimes write $\tilde{V}_{nj} = V_{nj}$. Peng and Tan (2016) showed that the JEL for vector U-statistics are asymptotically chi-square distributed under the same usual assumption as for the asymptotic normality of vector U-statistics, that is, if $Var(h_1(Z))$ is nonsingular then $-2\log \hat{\mathscr{R}}_n(\theta_0)$ is asymptotic chi-square distributed with d degrees of freedom, i.e.

$$-2\log \hat{\mathscr{R}}_n(\theta_0) \implies \chi^2_d.$$

3. MAXIMUM EMPIRICAL LIKELIHOOD ESTIMATION

In this chapter, we give the local uniform asymptotic normality condition for the log-JEL for UGEE's.

3.1 Asymptotic behaviors of the local logarithm of the JEL ratio

Let $(\mathscr{Z}, \mathscr{S})$ be a measurable space, $\mathscr{Q}$ be a family of probability measures on $\mathscr{S}$, and θ be a parameter of interest which is from an open subset Θ of $\mathbb{R}^k$. Let $Z_1, ..., Z_n$ be independent and identically distributed (i.i.d.) $\mathscr{Z}$-valued random variables with an unknown distribution Q belonging to the model $\mathscr{Q}$. Recall that a kernel function $h : \mathbb{R}^m \mapsto \mathbb{R}^d$ is permutation symmetric about its m arguments and satisfies (2.1.1). We are interested in inference about the characteristic θ and work with the jackknife empirical likelihood (JEL) ratio $\hat{\mathscr{R}}_n(\theta)$ in (2.2.1).

Qin and Lawless (1994) studied the maximum empirical likelihood estimator (MELE),

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \hat{\mathscr{R}}_n(\theta). \tag{3.1.1}$$

Recall that the local JEL ratio is defined as

$$\hat{\mathscr{L}}_n(t) = \log \frac{\hat{\mathscr{R}}_n(\theta_0 + n^{-1/2}t)}{\hat{\mathscr{R}}_n(\theta_0)}, \quad t \in \mathbb{R}^k, \quad \theta_0 + n^{-1/2}t \in \Theta.$$

For a function f on $\mathscr{Z} \times \Theta$, let $\dot{f}, \ddot{f}$ denote the first and second partial derivative of f with respect to parameter θ, that is,

$$\dot{f}(z;\theta) = \frac{\partial}{\partial \theta} f(z;\theta), \quad \ddot{f}(z;\theta) = \frac{\partial^2}{\partial \theta \partial \theta^\top} f(z;\theta), \quad z \in \mathscr{Z}, \quad \theta \in \Theta.$$

Recall $\tilde{h} = h - E(h)$. Let us introduce the following assumptions.

(A1) There exist a neighborhood $N(\theta_0)$ of θ_0 and a square-integrable function G on $\mathscr{Z}^m$ such that $h(z_1, \ldots, z_m; \theta)$ is twice continuously differentiable with respect to θ for every $(z_1, \ldots, z_m) \in \mathscr{Z}^m$ with the first partial derivative $\dot{h}$ of full rank and second partial derivative $\ddot{h}$ satisfying

$$\|\dot{h}(z_1, \ldots, z_m; \theta)\| + \|\ddot{h}(z_1, \ldots, z_m; \theta)\| \leq G(z_1, \ldots, z_m), \quad \theta \in N(\theta_0).$$

(A2) $\sup_{\theta \in N(\theta_0)} E\|h(Z_1, \ldots, Z_m; \theta)\|^2 < \infty$.

(A3) $W = m^2 E\big(\tilde{h}_1(Z; \theta_0)^{\otimes 2}\big)$ is positive definite.

Remark 3.1.1 (A1) implies that $\|h(z_1, \ldots, z_m; \theta)\|$ is also bounded by a square-integrable function $G^*(z_1, \ldots, z_m)$ on $\mathscr{Z}^m \times N(\theta_0)$ provided that $N(\theta_0)$ is bounded (by K). In fact, it follows from the mean value theorem that for $\theta \in N(\theta_0)$, there exist θ^* lying between θ_0 and θ such that

$$\begin{aligned}\|h(z_1, \ldots, z_m; \theta)\| &\leq \|h(z_1, \ldots, z_m; \theta_0)\| + \|\dot{h}(z_1, \ldots, z_m; \theta^*)\|\|\theta - \theta_0\| \\ &\leq \|h(z_1, \ldots, z_m; \theta_0)\| + G(z_1, \ldots, z_m)\|\theta - \theta_0\| \\ &\leq \|h(z_1, \ldots, z_m; \theta_0)\| + 2KG(z_1, \ldots, z_m) \\ &:= G^*(z_1, \ldots, z_m).\end{aligned}$$

Let $u_n = u_n(\theta_0)$ where

$$u_n(\theta) = n^{-1/2} \sum_{j=1}^{n} m\tilde{h}_1(Z_j; \theta). \tag{3.1.2}$$

We have the following uniform local asymptotic normality.

Theorem 3.1.1 *Assume that (A1)-(A3) are satisfied. Then it holds the expansions*

$$\sup_{\|t\| \leq C} \| - 2\log \hat{\mathscr{R}}_n(\theta_0 + n^{-1/2}t) - (u_n - At)^\top W^{-1}(u_n - At)\| = o_P(1) \tag{3.1.3}$$

and

$$\sup_{\|t\| \leq C} \|\hat{\mathscr{L}}_n(t) - t^\top A^\top W^{-1} u_n + 1/2 t^\top A^\top W^{-1} At\| = o_P(1) \tag{3.1.4}$$

for every finite constant C, where $A = -E[m\dot{\tilde{h}}_1(Z; \theta_0)]$.

Note that A has full rank by (A1) and plays the same role as the quantity $-E[\dot{u}(Z;\theta)]$ does in Qin and Lawless (1994). Note also that the quadratic function q defined by

$$q(t) = t^\top A^\top W^{-1} u_n - 1/2 t^\top A^\top W^{-1} A t, \quad t \in \mathbb{R}^k, \tag{3.1.5}$$

is uniquely maximized by $\hat{t} = (A^\top W^{-1} A)^{-1} A^\top W^{-1} u_n$. This shows that $\theta \mapsto \hat{\mathscr{R}}_n(\theta)$ has a local maximizer $\hat{\theta}$ such that

$$n^{1/2}(\hat{\theta} - \theta_0) - (A^\top W^{-1} A)^{-1} A^\top W^{-1} u_n = o_P(1).$$

Therefore,

$$\hat{\theta} = \theta_0 + (A^\top W^{-1} A)^{-1} \frac{m}{n} \sum_{j=1}^n A^\top W^{-1} \tilde{h}_1(Z_j;\theta_0) + o_P(n^{-1/2}). \tag{3.1.6}$$

This of course implies the asymptotic normality of $\hat{\theta}$, i.e.

$$\sqrt{n}(\hat{\theta} - \theta_0) \implies \mathscr{N}(0, (A^\top W^{-1} A)^{-1}). \tag{3.1.7}$$

Substituting (2.2.2) in $\hat{\mathscr{R}}_n(\theta)$, we get

$$-\log \hat{\mathscr{R}}_n(\theta) = \sum_{j=1}^n \log(1 + \xi^\top V_{nj}(\theta)), \quad \theta \in \Theta, \tag{3.1.8}$$

where ξ satisfies (2.2.3). It is not difficult to see that under (A1) – (A3) the random function $\theta \mapsto \hat{\mathscr{R}}_n(\theta)$ is continuously differentiable. Consequently, $\hat{\theta}$ and $\hat{\xi} = \xi(\hat{\theta})$ must satisfy

$$B_{1n}(\theta,\xi) = 0, \quad B_{2n}(\theta,\xi) = 0, \tag{3.1.9}$$

where

$$B_{1n}(\theta,\xi) = \frac{1}{n} \sum_{j=1}^n \frac{V_{nj}(\theta)}{1 + \xi^\top V_{nj}(\theta)}, \tag{3.1.10}$$

and

$$B_{2n}(\theta,\xi) = \frac{1}{n} \sum_{j=1}^n \frac{1}{1 + \xi^\top V_{nj}(\theta)} \frac{\partial V_{nj}(\theta)^\top}{\partial \theta} \xi. \tag{3.1.11}$$

Note that (3.1.9) are the estimating equations for the MELE $\hat{\theta}$. Summarizing the above discussion, we have the following result.

Theorem 3.1.2 *Suppose (A1) – (A3) hold. Then there is a maximizer $\hat{\theta}$ for the JEL $\theta \mapsto \hat{\mathscr{R}}_n(\theta)$ such that $\hat{\theta}$ solves (3.1.9) and satisfies the stochastic expansion (3.1.6) hence (3.1.7).*

3.2 Some lemmas

To prove Theorem 3.1.1, we need the following lemmas with the proof delayed to Chapter 7.

Lemma 3.2.1 *Assume (A1)-(A2) are met. Then it holds for every finite C,*

$$\hat{D}_n(C) = \sup_{\|t\| \le C} \frac{1}{n} \sum_{j=1}^{n} \|V_{nj}(\theta_0 + n^{-1/2}t) - m\tilde{h}_1(Z_j;\theta_0)\|^2 = O_P(1/n). \tag{3.2.1}$$

Lemma 3.2.2 *Under assumptions (A1)-(A2), it holds the expansion,*

$$\sup_{\|t\| \le C} \|n^{-1/2} \sum_{j=1}^{n} (V_{nj}(\theta_0 + n^{-1/2}t) - m\tilde{h}_1(Z_j;\theta_0)) + At\| = o_P(1) \tag{3.2.2}$$

for every finite constant C, with $A = -E[m\dot{\tilde{h}}_1(Z;\theta_0)]$.

To complete the proof of Theorem 3.1.1, we need a general result from Peng and Schick (2013). Let $\mathbb{T}_{n1}(t), ..., \mathbb{T}_{nn}(t)$ be d-dimensional random vectors indexed by $t \in \mathbb{R}^k$, where $k \le d$. We are interested in the asymptotic behavior of the empirical likelihood process

$$\mathscr{R}_n(t) = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^{n} \pi_j \mathbb{T}_{nj}(t) = 0 \Big\}, \quad \|t\| \le C,$$

where C is a positive constant. We shall use the following result which is a special case of Lemma 5.2 of Peng and Schick (2015).

Lemma 3.2.3 *Let $x_1, ..., x_n$ be d-dimensional vectors. Set*

$$x^* = \max_{1 \le j \le n} \|x_j\|, \quad \bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j, \quad S = \frac{1}{n} \sum_{j=1}^{n} x_j x_j^\top,$$

and let λ denote the smallest and Λ the largest eigen value of the matrix S. Then the inequality $\lambda > 5\|\bar{x}\|x^*$ implies

$$\| -2\log \mathscr{R} - n\bar{x}^\top S^{-1}\bar{x}\| \le (\Lambda + \frac{\Lambda^3}{4\Lambda^2})\frac{2n\|\bar{x}\|^3 x^*}{(\lambda - \|\bar{x}\|x^*)^3}, \tag{3.2.3}$$

where

$$\mathscr{R} = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j x_j = 0\Big\}.$$

Motivated by this we introduce the quantities

$$\mathbb{T}_n^*(t) = \max_{1\le j\le n} \|\mathbb{T}_{nj}(t)\|, \quad \bar{\mathbb{T}}_n(t) = \frac{1}{n}\sum_{j=1}^n \mathbb{T}_{nj}(t), \quad \mathbb{S}_n(t) = \frac{1}{n}\sum_{j=1}^n \mathbb{T}_{nj}(t)\mathbb{T}_{nj}^\top(t).$$

We impose the following conditions.

(B1) $\sup_{\|t\|\le C} \mathbb{T}_n^*(t) = o_P(n^{1/2})$.

(B2) There is a positive definite $d\times d$ matrix S such that

$$\sup_{\|t\|\le C} \|\mathbb{S}_n(t) - S\| = o_P(1).$$

(B3) There exist k-dimensional random vectors u_n and and $d \times k$ matrix A of full rank k such that $u_n = O_P(1)$ and

$$\sup_{\|t\|\le C} \|\sqrt{n}\bar{\mathbb{T}}_n(t) - u_n + At\| = o_P(1).$$

We have the following result with the proof delayed to Chapter 7.

Lemma 3.2.4 *Suppose (B1)-(B3) hold. Then*

$$\sup_{\|t\|\le C} \| -2\log \mathscr{R}_n(t) - (u_n - At)^\top S^{-1}(u_n - At)\| = o_P(1) \tag{3.2.4}$$

and therefore

$$\sup_{\|t\|\le C} \| \log \frac{\mathscr{R}_n(t)}{\mathscr{R}_n(0)} - t^\top A^\top S^{-1} u_n + \frac{1}{2} t^\top A^\top S^{-1} A t\| = o_P(1). \tag{3.2.5}$$

3.3 Proof of Theorem 3.1.1

With the help of the Lemma 3.2.1 – Lemma 3.2.4 we now prove the main result.

Proof of Theorem 3.1.1. We verify the assumptions of Lemma 3.2.3 with $\mathbb{T}_{nj}(t) = V_{nj}(\theta + n^{-1/2}t)$, $S = W$, $u_n = n^{-1/2}\sum_{j=1}^{n} m\tilde{h}_1(Z_j;\theta)$, and $A = -E[m\dot{\tilde{h}}_1(Z;\theta_0)]$. Since $E\|m\tilde{h}_1(Z;\theta_0)\|$ is finite, we obtain

$$\max_{1\le j\le n} \|m\tilde{h}_1(Z_j;\theta_0)\| = o_P(n^{1/2}). \tag{3.3.1}$$

For a fixed C, note that we have the bound

$$\sup_{\|t\|\le C} \max_{1\le j\le n} \|V_{nj}(\theta + n^{-1/2}t)\| \le \max_{1\le j\le n} \|m\tilde{h}_1(Z_j;\theta)\| + n^{1/2}\hat{D}_n^{1/2}$$

where $\hat{D}_n$ is given in (3.2.1). It then follows from Lemma 3.2.1 that $\hat{D}_n = o_p(1)$ hence

$$\sup_{\|t\|\le C} \max_{1\le j\le n} \|V_{nj}(\theta + n^{-1/2}t)\| = o_P(n^{1/2}).$$

This implies (B1). From Lemma 3.2.2 and the central limit theorem it follows that (B3) holds. We are now left to verify (B2). To this end, set

$$\bar{W}_n = \frac{1}{n}\sum_{j=1}^{n} \tilde{h}_1(Z;\theta)\tilde{h}_1^\top(Z;\theta).$$

By (A2),

$$\|\bar{W}_n - W\| = o_P(1). \tag{3.3.2}$$

We conclude (B2) from Lemma 3.2.1, (3.3.2) and the bound,

$$\begin{aligned}\|a^\top(\mathbb{S}_n(t) - \bar{W}_n)a\| &= \|\frac{1}{n}\sum_{j=1}^{n}(a^\top V_{nj}(\theta + n^{-1/2}t))^2 - \frac{1}{n}\sum_{j=1}^{n}(a^\top\tilde{h}_1(Z_j;\theta))^2\| \\ &\le \hat{D}_n + 2\Big(\frac{1}{n}\sum_{j=1}^{n}\|\tilde{h}_1(Z_j;\theta)\|^2\hat{D}_n\Big)^{1/2},\end{aligned}$$

valid for every unit vector a in $\mathbb{R}^k$, every t with $\|t\| \le C$. We now apply Lemma 3.2.3 to complete the proof of Theorem 3.1.1. ■

4. EASY MAXIMUM EMPIRICAL LIKELIHOOD ESTIMATION

In this chapter, we discuss the existence of maximum empirical likelihood estimate (MELE) $\hat{\theta}$. We derive the estimating equations for the MELE's and obtain their asymptotic behaviors. In the end, we give a class of easy MELE's and establish their asymptotic distributions.

4.1 MELE's and semiparametric effeciency

If h_1 were a known function, we would work with the empirical likelihood ratio,

$$\mathscr{R}_n(\theta) = \sup\Big\{\prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^{n} \pi_j m\tilde{h}_1(Z_j;\theta) = 0\Big\}, \quad \theta \in \Theta. \tag{4.1.1}$$

It follows from Owen (1988) that if $\mathrm{Var}(mh_1)$ is finite and positive definite then (4.1.1) reaches its maximum when

$$\pi_j = \frac{1}{n}\frac{1}{1+\xi^\top m\tilde{h}_1(Z_j;\theta)}, \qquad j = 1, ..., n,$$

where $\xi = (\xi_1, ..., \xi_d)^\top$ are Lagrange multipliers, which is a $d \times 1$ vector and satisfies

$$\sum_{j=1}^{n} \frac{m\tilde{h}_1(Z_j;\theta)}{1+\xi^\top m\tilde{h}_1(Z_j;\theta)} = 0.$$

Moreover, $\xi \to 0$ as $n \to \infty$.

In addition to satisfying (A1)-(A3), we further assume $EG^3 < \infty$. It then follows from Lemma 1 of Qin and Lawless (1994) that as $n \to \infty$, with probability 1 the EL ratio function in (4.1.1) attains its maximum value at some $\tilde{\theta}$ in the interior of the ball $\|\theta - \theta_0\| \le n^{-1/3}$, and $\tilde{\theta}$ and $\tilde{\xi} = \xi(\tilde{\theta})$ satisfy

$$A_{1n}(\tilde{\theta}, \tilde{\xi}) = 0, \qquad A_{2n}(\tilde{\theta}, \tilde{\xi}) = 0, \tag{4.1.2}$$

where

$$A_{1n}(\theta, \xi) = \frac{1}{n}\sum_{j=1}^{n} \frac{m\tilde{h}_1(Z_j; \theta)}{1 + \xi^\top m\tilde{h}_1(Z_j; \theta)}, \tag{4.1.3}$$

$$A_{2n}(\theta, \xi) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{1 + \xi^\top m\tilde{h}_1(Z_j; \theta)} \frac{\partial m\tilde{h}_1(Z_j; \theta)^\top}{\partial \theta}\xi. \tag{4.1.4}$$

The equations in (4.1.2) are theoretically useful and can't be used to find the MELE's defined by the JEL ratio (2.2.1) because h_1 is unknown. Instead, we find the MELE's by solving the estimating equations in (3.1.10) – (3.1.11). The next lemma states that both sets of equations give the same solutions as the sample size tends to infinity. To this end, let $N_0(\theta_0, 0) = \{\theta : \|\theta - \theta_0\| \leq n^{-1/3}\} \times \{\xi : \|\xi\| \leq n^{-1/3}\}$ denote a neighborhood of $(\theta_0, 0)$. We have the following lemma with the proof delayed to Chapter 7.

Lemma 4.1.1 *Assume (A1)-(A3) hold. Then*

$$\sup_{(\theta,\xi)\in N_0(\theta_0,0)} \|B_{1n}(\theta, \xi) - A_{1n}(\theta, \xi)\| = o_P(1), \tag{4.1.5}$$

and

$$\sup_{(\theta,\xi)\in N_0(\theta_0,0)} \|B_{2n}(\theta, \xi) - A_{2n}(\theta, \xi)\| = o_P(1). \tag{4.1.6}$$

Consequently, from Lemma 4.1.1 it follows that for large n, there exists some point $\hat{\theta}$ in a shrinking neighborhood of θ_0, such that $\hat{\theta}$ and $\hat{\xi} = \xi(\hat{\theta})$ satisfy (3.1.9) and the JEL (2.2.1) reaches its maximum value at $(\hat{\theta}, \hat{\xi})$. These statement hold on an event Ω with $P(\Omega) = 1$ at least for sufficiently large n. In general, on its complement Ω^c where $(\hat{\theta}, \hat{\xi})$ are not defined we define them to be arbitrary numbers. The below theorem gives the asymptotic normality and the proof can be found in Chapter 7.

Theorem 4.1.1 *Assume (A1)-(A3) hold with the dominating function G satisfying $E(G^3) < \infty$. Then, as n tends to infinity, with probability one $\hat{\theta}(\theta)$ attains its maximum at some $\hat{\theta}$ in a shrinking neighborhood of θ_0, and $\hat{\theta}$ and $\hat{\xi} = \xi(\hat{\theta})$ solves (3.1.9), and satisfy*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathscr{N}(0, V), \quad \sqrt{n}(\hat{\xi} - 0) \to \mathscr{N}(0, U), \tag{4.1.7}$$

where

$$V = \big(E(\dot{\tilde{h}}_1)^\top \big(E(\tilde{h}_1^{\otimes 2})\big)^{-1} E(\dot{\tilde{h}}_1)\big)^{-1},$$

$$U = \frac{1}{m^2}\big(E(\tilde{h}_1^{\otimes 2})\big)^{-1}\big(I_d - E(\dot{\tilde{h}}_1) V E(\dot{\tilde{h}}_1)^\top \big(E(\tilde{h}_1^{\otimes 2})\big)^{-1}\big),$$

and $\hat{\theta}$ *and* $\hat{\xi}$ *are asymptotically uncorrelated.*

As a result, the weights π_j, $j = 1, \ldots, n$ can be estimated by

$$\hat{\pi}_j = \frac{1}{n}\frac{1}{1 + \hat{\xi}^\top m \tilde{h}_1(Z_j; \hat{\theta})}, \qquad j = 1, ..., n. \tag{4.1.8}$$

This, in turn, yields an efficient estimate of the DF F as follows:

$$\hat{F}_n(z) = \sum_{j=1}^{n} \hat{\pi}_j \mathbf{1}(Z_j \leq z). \tag{4.1.9}$$

Remark 4.1.1 The asymptotic variance matrix V can be consistently estimated by

$$\hat{V} = \Big(\big(\sum_{j=1}^{n} \hat{\pi}_j \dot{\tilde{h}}_1(Z_j; \hat{\theta})\big)^\top \big(\sum_{j=1}^{n} \hat{\pi}_j \tilde{h}_1^{\otimes 2}(Z_j; \hat{\theta})\big)^{-1} \big(\sum_{j=1}^{n} \hat{\pi}_j \dot{\tilde{h}}_1(Z_j; \hat{\theta})\big)\Big)^{-1}. \tag{4.1.10}$$

Theorem 4.1.1 also can be used to get approximate confidence limits for θ or F.

By the U-statistics theory, the asymptotic distribution of a U-statistic $U_{mn}(h)$ of order m with kernel h is dictated by h_1. Thus we apply Corollary 1 of Qin and Lawless (1994) with their estimating function $g = h_1$ to obtain the below result.

Corollary 4.1.1 *Assume that the assumptions of Theorem 4.1.1 hold. Suppose* $d > k$*. Then the asymptotic covariance-variance matrix* $V = V_d$ *of* $\sqrt{n}(\hat{\theta} - \theta_0)$ *does not increase (in the sense of positive definiteness of positive definite matrices) as the number of estimating equations increases.*

Using the same argument as above and applying Theorem 3 of Qin and Lawless (1994), we have the following.

Theorem 4.1.2 *Under the assumptions of Theorem 4.1.1, the MELE* $\hat{\theta}$ *is fully efficient in the sense of Van der Vaart (1988) and Bickel, Klaassen, Ritov and Wellner (1993).*

The efficiency criteria used are that of a least dispersed regular estimator or that of a locally asymptotic minimax (LAM) estimator. These criteria are based on the convolution theorems and on the lower bounds on the local asymptotic risk in LAN (locally asymptotically normal) and LAM families, see the above references and additional references therein.

4.2 Easy maximum empirical likelihood estimation

In this section, we study a special case of the U-statistics based estimating equations, that is, some of the equations do not involve parameters. At a first glance, it is seemingly restrictive for use. But actually it is quite useful as we shall demonstrate below. As to this special case, we derive the estimating equations for the MELE's which are computationally faster than the usual MELE's – easy MELE's – as the solutions of the estimating equations given before.

Consider a kernel functions of the form,

$$h(Z_1,\dots,Z_m;\theta) = (u(Z_1,\dots,Z_m;\theta)^\top, v(Z_1,\dots,Z_m)^\top)^\top,$$

where $u : \mathbb{R}^m \times \Theta \mapsto \mathbb{R}^p$ and $v : \mathbb{R}^m \mapsto \mathbb{R}^q$ are measurable functions and Θ is an open subset of $\mathbb{R}^k$. Suppose u and v satisfy

$$E(u(Z_1,\dots,Z_m;\theta)) = 0, \quad \theta \in \Theta, \tag{4.2.1}$$

and

$$E(v(Z_1,\dots,Z_m)) = 0. \tag{4.2.2}$$

While (4.2.1) serves as a criterion equation for the parameter θ, (4.2.2) describes side information about the underlying distribution. The parameter θ is usually estimated by the M-estimate, the solution to the sample version of (4.2.1), that is, the UGEE with kernel $u(;\theta)$,

$$U_{nm}(u(;\theta)) = \binom{n}{m}^{-1} \sum_{1\le i_1<\cdots<i_m\le n} u(Z_{i_1},\dots,Z_{i_m};\theta) = 0, \quad \theta \in \Theta, \tag{4.2.3}$$

where we assume, without loss of generality, that u is argument symmetric about its m variables (otherwise we symmetrize it). The M-estimate is not efficient in general as the side information given by (4.2.2) is not utilized. Often we assume that the number p of equations in (4.2.1) is equal to the dimension k of the parameter θ (otherwise we can eliminate those redundant equations). Throughout this section we assume $p = k$.

We now work with the JEL,

$$\hat{\mathscr{R}}_n(\theta) = \sup\Big\{\prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^{n} \pi_j \mathrm{vec}(V_{nj}^u(\theta), V_{nj}^v) = 0\Big\}, \quad \theta \in \Theta, \tag{4.2.4}$$

where V_{nj}^u, V_{nj}^v are the jackknife pseudo values based on the U-statistics with kernel function u and v respectively.

In this case, $V = m^2 E\big((\tilde{u}_1(Z;\theta_0)^\top, \tilde{v}_1(Z)^\top)^\top(\tilde{u}_1(Z;\theta_0)^\top, \tilde{v}_1(Z)^\top)\big)$, so that

$$V = m^2 \begin{pmatrix} E(\tilde{u}_1^{\otimes 2}) & E(\tilde{u}_1\tilde{v}_1^\top) \\ E(\tilde{v}_1\tilde{u}_1^\top) & E(\tilde{v}_1^{\otimes 2}) \end{pmatrix}. \tag{4.2.5}$$

Introduce the following assumption.

(A4) Suppose u satisfies (A1), u and v satisfy (A2), and V satisfies (A3).

Under (A4), we can apply Lemma 3.2.1 and Lemma 3.2.2 to u, so that it holds for every finite C,

$$\hat{D}_n(C) = \sup_{\|t\|\le C} \frac{1}{n}\sum_{j=1}^{n} \|V_{nj}^u(\theta_0 + n^{-1/2}t) - m\tilde{u}_1(Z_j;\theta_0)\|^2 = O_P(\frac{1}{n}). \tag{4.2.6}$$

$$\sup_{\|t\|\le C} \|n^{-1/2}\sum_{j=1}^{n}(V_{nj}^u(\theta_0 + n^{-1/2}t) - m\tilde{u}_1(Z_j;\theta_0)) + A_u t\| = o_P(1), \tag{4.2.7}$$

where $A_u = -E[m\dot{\tilde{u}}_1(Z;\theta_0)]$ of full rank. Let $A_v = 0_{q\times k}$ and $A = (A_u^\top, A_v^\top)^\top$. It follows from Theorem 3.1.1 holds for the JEL (4.2.4). Let $\xi = (\xi_{u_{(p\times 1)}}^\top, \xi_{v_{(q\times 1)}}^\top)$ be the Lagrange multipliers. Then we have

$$\pi_j = \frac{1}{n}\frac{1}{1 + \xi_u^\top V_{nj}^u(\theta) + \xi_v^\top V_{nj}^v}, \quad j = 1, \dots, n,$$

where ξ satisfies

$$\sum_{j=1}^{n} \frac{\mathrm{vec}(V_{nj}^{u}(\theta), V_{nj}^{v})}{1+\xi_u^\top V_{nj}^{u}(\theta)+\xi_v^\top V_{nj}^{v}} = 0.$$

Thus, similar to the discussion in the previous section, as $n \to \infty$, (4.2.4) attains its maximum value in probability at some $\hat\theta$ in a shrinking neighborhood of θ_0, and $\hat\theta$ and $\hat\xi = (\hat\xi_u^\top, \hat\xi_v^\top)^\top = (\xi_u(\hat\theta)^\top, \hat\xi_v^\top)^\top$ satisfy

$$C_{1n}(\hat\theta, \hat\xi) = 0, \qquad C_{2n}(\hat\theta, \hat\xi) = 0, \tag{4.2.8}$$

where

$$C_{1n}(\theta, \xi) = \frac{1}{n}\sum_{j=1}^{n} \frac{\mathrm{vec}(V_{nj}^{u}(\theta), V_{nj}^{v})}{1+\xi_u^\top V_{nj}^{u}(\theta)+\xi_v^\top V_{nj}^{v}}, \tag{4.2.9}$$

$$C_{2n}(\theta, \xi) = \frac{1}{n}\sum_{j=1}^{n} \frac{\xi_u^\top \dot V_{nj}^{u}(\theta)}{1+\xi_u^\top V_{nj}^{u}(\theta)+\xi_v^\top V_{nj}^{v}}. \tag{4.2.10}$$

Therefore, as a corollary of Theorem 4.1.1, the asymptotic properties of the MELE's $\hat\theta$ of (4.2.4) can be obtained as stated below. As a convention, we drop the argument at the true value of parameter so that $u(Z;\theta_0) = u(Z)$, $E(u(Z;\theta_0)^{\otimes 2}) = E(u^{\otimes 2}(Z))$ and of course $E(u^{\otimes 2}(Z)) = E(u^{\otimes 2})$, etc. Under (A4), V is invertible. Let the inverse of the block matrix V be $V^{-1} = (V^{ij})_{i,j=1,2}$. By the inverse formulas for a block matrix, we have

$$V^{-1} = \begin{pmatrix} V_{11\cdot 2}^{-1} & -V_{11\cdot 2}^{-1} E(\tilde u_1 \tilde v_1^\top)(E(\tilde v_1^{\otimes 2}))^{-1} \\ -V_{22\cdot 1}^{-1} E(\tilde v_1 \tilde u_1^\top)(E(\tilde u_1^{\otimes 2}))^{-1} & V_{22\cdot 1}^{-1} \end{pmatrix}, \tag{4.2.11}$$

where

$$V_{11\cdot 2} = E(\tilde u_1^{\otimes 2}) - E(\tilde u_1 \tilde v_1^\top)\big(E(\tilde v_1^{\otimes 2})\big)^{-1} E(\tilde v_1 \tilde u_1^\top),$$
$$V_{22\cdot 1} = E(\tilde v_1^{\otimes 2}) - E(\tilde v_1 \tilde u_1^\top)\big(E(\tilde u_1^{\otimes 2})\big)^{-1} E(\tilde u_1 \tilde v_1^\top).$$

Theorem 4.2.1 *Suppose (A4) holds. Suppose also $p = k$. Then, as n tends to infinity, with probability one $\hat\theta(\theta)$ attains its maximum at some $\hat\theta$ in a shrinking neighborhood of θ_0, $\hat\xi = (0, \hat\xi_v^\top)^\top$ solves (4.2.8), and satisfy*

$$\sqrt{n}(\hat\theta - \theta_0) \to \mathscr{N}(0, \bar V), \quad \sqrt{n}(\hat\xi_v - 0) \to \mathscr{N}(0, \bar U),$$

where

$$\bar{V} = \big(E(\dot{\tilde{u}}_1)\big)^{-1} V_{11\cdot 2} \big(E(\dot{\tilde{u}}_1)\big)^{-\top}, \quad \bar{U} = V_{22\cdot 1} - V^{21} V_{11\cdot 2}^{-1} V^{12}.$$

and $\hat{\theta}$ *and* $\hat{\xi}$ *are asymptotically uncorrelated.*

Remark 4.2.1 As the asymptotic variance of the estimate of the UGEE (4.2.3) is $\Sigma = \big(E(\dot{\tilde{u}}_1)\big)^{-1} E(\tilde{u}_1^{\otimes 2}) \big(E(\dot{\tilde{u}}_1)\big)^{-\top}$, the asymptotic variance of the MELE's satisfies $\bar{V} \leq \Sigma$.

Now let us expand the constraints in (4.2.4) as

$$\begin{cases} \sum_{j=1}^{n} \pi_j V_{nj}^u(\theta) = 0, \quad \vartheta \in \Theta, & \text{(4.2.12a)} \\ \sum_{j=1}^{n} \pi_j V_{nj}^v = 0. & \text{(4.2.12b)} \end{cases}$$

Based on (4.2.12b), we look at the JEL,

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^{n} \pi_j V_{nj}^v = 0 \Big\}. \tag{4.2.13}$$

This has the solution

$$\hat{\pi}_j = \frac{1}{n} \frac{1}{1 + \hat{\xi}_v^\top V_{nj}^v}, \quad j = 1, \ldots, n,$$

where $\hat{\xi}_v$ satisfies

$$\sum_{j=1}^{n} \frac{V_{nj}^v}{1 + \xi_v^\top V_{nj}^v} = 0. \tag{4.2.14}$$

Naturally, we substitute $\hat{\pi}_j$ in (4.2.12a) to get

$$\frac{1}{n} \sum_{j=1}^{n} \frac{V_{nj}^u(\theta)}{1 + \hat{\xi}_v^\top V_{nj}^v} = 0. \tag{4.2.15}$$

Therefore, we find $\xi_u = 0$, and $\hat{\xi} = (0_{p\times 1}, \hat{\xi}_v)$. It is worth to note that (4.2.14) – (4.2.15) are identical to (4.2.9) – (4.2.10). Consequently, we can find the MELE's by solving (4.2.14) – (4.2.15). We refer the MELE's as the solutions to (4.2.15) to as the *easy empirical likelihood estimates* as they have less computational burden than those usual MELE's as the solutions to (4.2.9) – (4.2.10) (or (3.1.9)) as pointed by Peng (2015).

5. APPLICATIONS: EXAMPLES

In this chapter, we give several examples.

5.1 Estimating the expected values with side information

We are interested in estimation of the expected value $\theta = E(\psi(Z_1, \ldots, Z_m))$ for some known function $\psi : \mathbb{R}^m \mapsto \mathbb{R}^d$ in the presence of side information given by $ET(Z_1, \ldots, Z_m) = a$ for some measurable function $T : \mathbb{R}^m \mapsto \mathbb{R}$ and constant a. Without loss of generality, we assume both ψ and T are argument-symmetric. Hence our kernel functions are given by

$$\begin{cases} u(Z_1, \ldots, Z_m; \theta) = \psi(Z_1, \ldots, Z_m) - \theta, & (5.1.1a) \\ v(Z_1, \ldots, Z_m) = T(Z_1, \ldots, Z_m) - a. & (5.1.1b) \end{cases}$$

We shall apply Theorem 4.2.1 to derive the asymptotic behaviors.

Let us first mention that U-statistics are quite general. Heffernan (1997) showed that a statistical functional $\theta = \theta(Q)$ of a distribution Q admits an unbiased estimator iff there is a function ψ of m variables such that $\theta(Q) = \int \cdots \int \psi \, dQ^m$, and derived the U-statistic as the unique MVUE of a central moment. Moment based distribution characteristics (e.g. Pearson's correlation) are functions of central moments, so that the sample versions as test statistics can be expressed as functions of U-statistics.

Example 1 Estimating the mean difference in the presence of known covariance. Let $Z = (X, Y)$ be a bivariate random vector with finite second moments. We are interested in estimating the mean difference $\theta = E(X - Y)$ when there is available the side information $\mathrm{Cov}(X, Y) = a$ for some known a. Let $Z_i = (X_i, Y_i)$, $i = 1, \ldots, n$ be a random sample of Z. Let

$$u(Z_1, Z_2; \theta) = \frac{1}{2}(X_1 + X_2 - Y_1 - Y_2) - \theta$$

be the symmetrized kernel function. The side information can be expressed by the U-statistic based equation $E(U_n(v)) = 0$ with the kernel equal to

$$v(Z_1, Z_2) = \frac{1}{2}(X_1Y_1 + X_2Y_2 - X_1Y_2 - X_2Y_1) - a.$$

We shall apply Theorem 4.2.1 to derive the asymptotic behavior of the MELE $\hat{\theta}$.

Let $E(X) = \mu_1$, $E(Y) = \mu_2$, $\mathrm{Var}(X) = \sigma_1^2$, and $\mathrm{Var}(Y) = \sigma_2^2$. Then

$$\begin{aligned} u_1(z_1;\theta) &= E(u(Z_1, Z_2;\theta)|Z_1 = z_1) \\ &= E(\frac{1}{2}(x_1 - y_1) + \frac{1}{2}(X_2 - Y_2) - \theta) \\ &= \frac{1}{2}(x_1 - y_1) + \frac{1}{2}(\mu_1 - \mu_2) - \theta, \\ &= \frac{1}{2}(x_1 - y_1) - \frac{1}{2}\theta, \\ v_1(z_1) &= E(v(Z_1, Z_2)|Z_1 = z_1) \\ &= \frac{1}{2}(x_1y_1 + E(X_2Y_2) - x_1\mu_2 - y_1\mu_1) - a \\ &= \frac{1}{2}(x_1y_1 + \mu_1\mu_2 + a - x_1\mu_2 - y_1\mu_1) - a, \\ &= \frac{1}{2}(x_1 - \mu_1)(y_1 - \mu_2) - \frac{1}{2}a, \end{aligned}$$

and

$$\dot{u}_1(z;\theta) = -\frac{1}{2}.$$

Let $\hat{\theta}$ be the MELE. Then it follows from Theorem 4.2.1 that

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathscr{N}(0, \bar{V}),$$

where

$$\begin{aligned} \bar{V} =& E((X-Y) - \theta_0)^2 - \frac{\left(E[\big((X-Y)-\theta_0\big)\big((X-\mu_1)(Y-\mu_2)-a\big)]\right)^2}{E\big((X-\mu_1)(Y-\mu_2)-a\big)^2} \\ =& \sigma_1^2 + \sigma_2^2 - 2a - \frac{\Big(E\big((X-\mu_1)^2(Y-\mu_2)\big) - E\big((X-\mu_1)(Y-\mu_2)^2\big)\Big)^2}{E\big((X-\mu_1)^2(Y-\mu_2)^2\big) + a^2}. \end{aligned}$$

Example 2 Estimating the DF in the presence of known CV. By (4.1.9), we can construct an improved distribution function (DF) F of a random variable Z in when side information is available. Let $Z_1, \ldots, Z_n$ be i.i.d. copies of Z. Let

$$\psi(Z_1, Z_2; t) = \frac{1}{2}(\mathbf{1}[Z_1 \le t] + \mathbf{1}[Z_2 \le t]), \quad t \in \mathbb{R}.$$

Then

$$\theta_t = E(\psi(Z_1, Z_2; t)) = \frac{1}{2}(P(Z_1 \le t) + P(Z_2 \le t)) = F(t).$$

Often we have some side information about F. Here we assume the side information is given by $\sigma/\mu = c_0$, that is, the coefficient of variation of Z equals to a constant c_0. This side information can be expressed as a U-statistic equation by taking

$$T(Z_1, Z_2) = \frac{1}{2}(Z_1^2 + Z_2^2) - (1 + c_0^2)Z_1 Z_2.$$

Therefore, with the kernel functions equal to

$$\begin{aligned} u(Z_1, Z_2; \theta_t) &= \frac{1}{2}(\mathbf{1}[Z_1 \le t] + \mathbf{1}[Z_2 \le t]) - \theta_t, \\ v(Z_1, Z_2) &= \frac{1}{2}(Z_1^2 + Z_2^2) - (1 + c_0^2)Z_1 Z_2, \end{aligned}$$

the jackknife pseudo values V_{nj}^u, V_{nj}^v can be computed, and the estimates $(\hat{\theta}_t, \hat{\xi})$ can be obtained as the solutions to the estimating equations (3.1.9). Alternatively, we can apply (4.2.14) and (4.2.15) to this example and obtain computationally faster estimates as the solutions to the below equations: Find $\hat{\xi}$ as the solution to

$$\sum_{j=1}^{n} \frac{V_{nj}^v}{1 + \xi_v^\top V_{nj}^v} = 0,$$

while $\hat{\theta}_t$ is the solution to

$$\frac{1}{n}\sum_{j=1}^{n} \frac{V_{nj}^u(\theta_t)}{1 + \hat{\xi}_v^\top V_{nj}^v} = 0.$$

As

$$\hat{\pi}_j = \frac{1}{n}\frac{1}{1 + \hat{\xi}_v^\top V_{nj}^v}, \qquad j = 1, \ldots, n.$$

by (4.1.9), an efficient estimate of the DF $F(z)$ is given by

$$\hat{F}_n(z) = \sum_{j=1}^{n} \hat{\pi}_j \mathbf{1}[Z_j \le z].$$

Here we suppress the dependence of $\hat{\pi}_j$ on the fixed t.

Example 3 Estimating the convolution with side information. Let Z_1, $\ldots$, Z_n be i.i.d. copies with a random variable Z on $\mathbb{R}$ with finite $\mu = E(Z)$ and

$\sigma^2 = \mathrm{Var}(Z_1)$. Let $\psi(z_1, z_2) = \mathbf{1}[z_1 + z_2 \le t], z_1, z_2 \in \mathbb{R}$ for known $t \in \mathbb{R}$. Hence $\theta_t = P(Z_1 + Z_2 \le t)$ is the convolution of Z_1 and Z_2. Suppose that there is available the side information that the coefficient of variation is known: $\sigma/\mu = c_0$ for some known constant c_0. The kernel functions can be constructed as

$$u(Z_1, Z_2; \theta_t) = \mathbf{1}(Z_1 + Z_2 \le t) - \theta_t,$$
$$v(Z_1, Z_2) = \frac{1}{2}(Z_1^2 + Z_2^2) - (1 + c_0^2) Z_1 Z_2.$$

Let F be the DF of Z. It follows

$$u_1(z_1; \theta) = E(u(Z_1, Z_2; \theta) | Z_1 = z_1) = F(t - z_1) - \theta_t,$$
$$v_1(z_1) = E(v(Z_1, Z_2) | Z_1 = z_1) = \frac{1}{2}(z_1^2 + (1 + c_0^2)\mu^2) - (1 + c_0^2)\mu z_1;$$

and

$$\dot{u}_1(z_1; \theta_t) = -1.$$

By Theorem 4.2.1, the MELE $\hat{\theta}_t$ is asymptotically normally:

$$\sqrt{n}(\hat{\theta}_t - \theta_t) \to \mathscr{N}(0, \bar{V}_t),$$

where

$$\bar{V}_t = E(F(t - Z_1) - \theta_t)^2 - \frac{\Big(E(F(t - Z_1) - \theta_t)(\frac{1}{2}(Z_1^2 + \sigma^2 + \mu^2) - (1 + c_0^2)\mu Z_1)\Big)^2}{E(\frac{1}{2}(Z_1^2 + (1 + c_0^2)\mu^2) - (1 + c_0^2)\mu Z_1)^2}.$$

Example 4 Estimating Gini's mean difference with side information. Let $Z_1, \ldots, Z_n$ be i.i.d. with r.v Z. Gini's mean difference of Z_1 and Z_2 is defined as $E|Z_1 - Z_2|$, which is an alternative index of variability. It is estimated by the U-statistic

$$U_{n2} = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} |Z_i - Z_j|.$$

Suppose there is available side information that the inter-quartile range is known,

$$P(q_1 \le Z \le q_3) = 0.5,$$

for some known constants q_1, q_3. Clearly the U-statistic as estimate does not utilize this side information. Here we use the JEL method to incorporate the side information. To this end, let $\theta = E|Z_1 - Z_2|$. The kernel is then given by $u(Z_1, Z_2; \theta) = |Z_1 - Z_2| - \theta$. The side information can be expressed by the U-statistic with the kernel equal to

$$v(z_1, z_2) = (\mathbf{1}[q_1 \le z_1 \le q_3] + \mathbf{1}[q_1 \le z_2 \le q_3])/2 - 0.5.$$

It is easy to calculate

$$u_1(z_1; \theta) = E|z_1 - Z_2| - \theta,$$

$$v_1(z_1) = \mathbf{1}[q_1 \le z_1 \le q_3]/2 - 0.25,$$

and $\dot{u}_1(z_1; \theta) = -1$. By Theorem 4.2.1, the MELE $\hat{\theta}$ has asymptotic normal distribution with the variance equal to

$$E(u_1^2) - (E(u_1 v_1))^2 / E(v_1^2) = E(u_1^2) - 16(E(u_1 v_1))^2.$$

Example 5 ESTIMATING THE OVERDISPERSION PARAMETER WITH SIDE INFORMATION. Overdispersion is common in count data. This can be modeled by overdispersion parameter ϕ as

$$\mathrm{Var}(Z) = \phi E(Z), \qquad \phi > 1. \tag{5.1.2}$$

Let $Z_1, \ldots, Z_n$ be count data (frequency data). Suppose there is available side information that

$$P(Z = 0) = P_0.$$

We are interested in estimating the overdispersion parameter ϕ. Clearly (5.1.2) can be written as

$$E(Z^2) = \phi E(Z) + (E(Z))^2,$$

which is equivalent to

$$E(Z_1^2 - \phi Z_1 - Z_1 Z_2) = 0.$$

Thus the corresponding U-statistic based equation with the kernel equal to the symmetrized function:

$$u(Z_1, Z_2; \phi) = \frac{1}{2}(Z_1^2 + Z_2^2) - \frac{\phi}{2}(Z_1 + Z_2) - Z_1 Z_2. \tag{5.1.3}$$

The jackknife pseudo value of (5.1.3) is one constraint for estimating ϕ. Moreover, the kernel function of the side information can be constructed as

$$v(Z_1, Z_2) = \frac{1}{2}(\mathbf{1}[Z_1 = 0] + \mathbf{1}[Z_2 = 0]) - P_0. \tag{5.1.4}$$

Now we apply Theorem 4.2.1 to (5.1.3) and (5.1.4) to estimate ϕ. Let $E(Z) = \mu$. We have

$$\begin{aligned}
u_1(z_1; \phi) =& E(u(Z_1, Z_2; \phi)|Z_1 = z_1) \\
=& E\big(\frac{1}{2}(z_1^2 + Z_2^2) - \frac{\phi}{2}(z_1 + Z_2) - z_1 Z_2\big) \\
=& \frac{1}{2}z_1^2 - \frac{\phi}{2}z_1 + \frac{1}{2}\mu^2 - z_1\mu; \\
v_1(z_1) =& E(v(Z_1, Z_2)|Z_1 = z_1) \\
=& \frac{1}{2}(\mathbf{1}[z_1 = 0] + P(Z_2 = 0)) - P_0 \\
=& \frac{1}{2}(\mathbf{1}[z_1 = 0] - P_0),
\end{aligned}$$

and

$$\dot{u}_1(z; \theta) = \frac{z}{2}.$$

Hence, by Theorem 4.2.1,

$$\sqrt{n}(\hat{\phi} - \phi_0) \to \mathscr{N}(0, \bar{V}),$$

where

$$\bar{V} = \frac{4}{\mu^2}\Big(E(u_1^2) - \frac{\big(E(u_1 v_1)\big)^2}{E(v_1^2)}\Big) = \frac{4}{\mu^2}\Big(E(u_1^2) - 4\frac{\big(E(u_1 v_1)\big)^2}{P_0(1 - P_0)}\Big).$$

Example 6 THE SIMPLICIAL DEPTH FUNCTION. Let $Z_1, \dots, Z_n$ be i.i.d. with a distribution Q on $\mathbb{R}^m$. The *simplicial depth function* $D(z)$ of a point $z \in \mathbb{R}^m$ with respect to distribution Q is defined as follows:

$$D(z) = P\big(z \in \Delta(Z_1, \dots, Z_{m+1})\big), \quad z \in \mathbb{R}^m,$$

where $\Delta(Z_1, \dots, Z_{m+1})$ denotes the random simplex with vertices $Z_1, \dots, Z_{m+1}$, i.e., the closed simplex with vertices $Z_1, \dots, Z_{m+1}$. For a point $z \in \mathbb{R}^m$, $D(z)$ is the value of the population simplicial depth at point z. The usual estimate of the depth function is the sample simplicial depth $D_n(z)$ given by the U-statistic,

$$D_n(z) = \binom{n}{m+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{m+1} \le n} \mathbf{1}[z \in \Delta(Z_{i_1}, \ldots, Z_{i_{m+1}})], \quad z \in \mathbb{R}^m.$$

The depth function can be used to define the multivariate medians and possess robustness property. When additional information is available about the underlying distribution Q, our JEL approach is capable to employ it into the estimation. In this example, we assume the marginal medians of Z are known as the side information. Let $Z = (Z^{(1)}, \ldots, Z^{(m)})^\top$, and $M = (a^{(1)}, \ldots, a^{(m)})^\top$ where $a^{(l)} = \mathrm{med}(Z^{(l)})$, $l = 1, \ldots, m$. Let $Z_j = (Z_j^{(1)}, \ldots, Z_j^{(m)})^\top$, $j = 1, \ldots, n$. Fix $z \in \mathbb{R}^m$ and $D = D(z)$. The kernel functions can be constructed as

$$u(Z_1, \ldots, Z_{m+1}; D) = \mathbf{1}[z \in \Delta(Z_1, \ldots, Z_{m+1})] - D;$$

and

$$v^{(l)}(Z_1, \ldots, Z_{m+1}) = \frac{1}{m+1} \sum_{j=1}^{m+1} \big(\mathbf{1}[Z_j^{(l)} \le a^{(l)}] - \frac{1}{2}\big), \quad l = 1, \ldots, m.$$

Thus

$$\begin{aligned} u_1(z_1; D) =& P\big(z \in \Delta(z_1, Z_2, \ldots, Z_{m+1})\big) - D; \\ v_1^{(l)}(z_1) =& \frac{1}{m+1}\big(\mathbf{1}[z_1^{(l)} \le a^{(l)}] - \frac{1}{2}\big) + \frac{1}{m+1} \sum_{j=2}^{m+1} \big(P(Z_j^{(l)} \le a^{(l)}) - \frac{1}{2}\big) \\ =& \frac{1}{m+1}\big(\mathbf{1}[z_1^{(l)} \le a^{(l)}] - \frac{1}{2}\big), \qquad l = 1, \ldots, m, \end{aligned}$$

and

$$\dot{u}_1(z; D) = -1.$$

Let $v_1 = (v_1^{(1)}, \ldots, v_1^{(m)})^\top$. Hence, by Theorem 4.2.1, the MELE $\hat{D}$ satisfies

$$\sqrt{n}(\hat{D} - D) \to \mathscr{N}(0, \bar{V}),$$

where

$$\bar{V} = E(u_1^2) - E(u_1 v_1^\top)\big(E(v_1 v_1^\top)\big)^{-1} E(v_1 u_1).$$

5.2 Smoothed U-quantiles

The theory of U-quantile provides a unified treatment of several commonly used statistics, see Arcones (1993). Let $\kappa : \mathbb{R}^m \mapsto \mathbb{R}$ be a measurable argument-symmetric function. Associated with κ there induces a distribution function $H(t) = P(\kappa(Z_1, \dots, Z_m) \le t)$, $t \in \mathbb{R}$. The minimum variance unbiased estimate (MVUE) of $H(t)$ is the U-statistic of order m given by

$$U_n(t) = U_{nm}(t) =: \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} \mathbf{1}[\kappa(Z_{i_1}, \dots, Z_{i_m}) \le t], \quad t \in \mathbb{R}, \qquad (5.2.1)$$

and κ shall be referred to as the kernel (of the U-quantile). As $H(t)$ is a distribution function, its p-th quantile t_p is well defined by $t_p = \inf\{t : H(t) \ge p\}$ for $p \in [0,1]$. The U-quantiles include the Hodges-Lehmann median estimator, Gini's mean difference, Theil's estimator of the slope in a simple linear model, and Kendall's tau. They correspond to the U-quantiles with $p_0 = 1/2$ and the kernels $\kappa(z_1, z_2) = 2^{-1}(z_1 + z_2)$, $|z_1 - z_2|$, $(y_1 - y_2)/(x_1 - x_2)$, and $(x_1 - x_2)(y_1 - y_2)$ respectively.

As the U-quantiles are discontinuous, our theory does not apply here. We now consider smoothed U-quantiles. Let F be a continuous DF. A continuous estimator of $H(t)$ is the smoothed U-quantile,

$$H_{nm}(t) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} F_b((t - \kappa(Z_{i_1}, \dots, Z_{i_m})), \quad t \in \mathbb{R}, \qquad (5.2.2)$$

where $F_b(t) = F(t/b)$ with b a bandwidth. This is the smoothed version of the U-statistic in (5.2.1). Given $p \in [0,1]$, the p-th U-quantile solves $H(t_p) = p$. The smoothed sample p-th U-quantile $\hat{t}_p$ is a solution of $H_{nm}(t_p) = p$.

Let us take the Theil-Sen estimator for an illustration. The Theil-Sen estimator is a robust estimator of the slope in a simple linear model. Suppose that $Z_i = (X_i, Y_i)^\top, i = 1, \dots, n$ are independent and satisfy

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where α is the intercept and β is the slope, and $\varepsilon_i, i = 1, \ldots, n$ are i.i.d. The Theil-Sen estimator of the slope is the median of the slopes $(Y_i - Y_j)/(X_i - X_j)$:

$$\beta = \operatorname{med}\left\{(Y_i - Y_j)/(X_i - X_j) : \ 1 \le i < j \le n\right\},$$

where $X_1, \ldots, X_n$ are assumed to be distinct for the sake of convenience. In this case, $\kappa(z_1, z_2) = (y_1 - y_2)/(x_1 - x_2)$ and

$$H(t) = P((Y_1 - Y_2)/(X_1 - X_2) \le t).$$

While the corresponding U-statistic is

$$U_{n2}(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbf{1}[(Y_i - Y_j)/(X_i - X_j) - \beta < 0],$$

the smoothed version is

$$H_{n2}(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} F_b\big(\beta - (Y_i - Y_j)/(X_i - X_j)\big). \tag{5.2.3}$$

Hence, to estimate β, one constraint can be chosen as

$$H_{n2}(\beta) - \frac{1}{2} = 0. \tag{5.2.4}$$

In a simple linear model, we assume errors ε_i, $i = 1, \ldots, n$ are i.i.d. with the normal with zero mean. Therefore, we have

$$P(0 \le \varepsilon_j) = \frac{1}{2}. \tag{5.2.5}$$

Here we relax the normality assumption to (5.2.5), the assumption of zero median of the error and use it to improve the efficiency of the Theil-Sen estimator. The U-statistic based equation of (5.2.5) is

$$\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbf{1}[0 < \frac{\varepsilon_i + \varepsilon_j}{2}] - \frac{1}{2} = 0. \tag{5.2.6}$$

Thus, we obtain the smoothed version of (5.2.6)

$$\binom{n}{2}^{-1} \sum_{1 \le i < j \le n} F_b(\frac{\varepsilon_i + \varepsilon_j}{2}) - \frac{1}{2} = 0. \tag{5.2.7}$$

This gives us another constraint.

Now let us apply Theorem 4.1.1 to get the asymptotic distribution of the improved Theil-Sen estimator β. From (5.2.4) and (5.2.7), the kernel is $h(Z_1, Z_2; \beta) = (h^{(1)}(Z_1, Z_2; \beta), h^{(2)}(Z_1, Z_2; \beta))^\top$, where

$$h^{(1)}(Z_1, Z_2; \beta) = F_b\big(\beta - (Y_1 - Y_2)/(X_1 - X_2)\big) - 1/2,$$
$$h^{(2)}(Z_1, Z_2; \beta) = F_b\big((Y_1 + Y_2 - \beta(X_1 + X_2))/2\big) - 1/2.$$

Therefore,

$$h_1^{(1)}(z_1; \beta) = E[F_b(\beta - (y_1 - Y_2)/(x_1 - X_2))] - 1/2,$$
$$h_1^{(2)}(z_1; \beta) = E[F_b((y_1 + Y_2 - \beta(x_1 + X_2))/2)] - 1/2.$$

Let f be the pdf of F and $f_b(t) = f(\frac{t}{b})$. Since F is a cdf, it follows

$$\dot{h}_1^{(1)}(z_1; \beta) = E\big(b^{-1} f_b(\beta - (y_1 - Y_2)/(x_1 - X_2))\big),$$
$$\dot{h}_1^{(2)}(z_1; \beta) = E\big(-(x_1 + X_2)/(2b) f_b((y_1 + Y_2 - \beta(x_1 + X_2))/2)\big).$$

Using these we can obtain the asymptotic normal distribution of the MELE by Theorem 4.1.1.

6. AN EL APPROACH TO GOODNESS-OF-FIT TESTING FOR MULTIVARIATE AND HIGH DIMENSIONAL SYMMETRIES

In this chapter, we develop empirical likelihood tests to various multivariate and high dimensional symmetries based on the characterizations of the symmetries. We report some simulation results.

6.1 Testing multivariate symmetries

6.1.1 Spherical symmetry

A random vector X in $\mathbb{R}^d$ is spherically symmetric about a point $\theta \in \mathbb{R}^d$ if

$$X - \theta \stackrel{d}{=} \Gamma(X - \theta),$$

for every orthogonal $d \times d$ matrix Γ, where $\stackrel{d}{=}$ denotes both sides of the equality have an identical distribution. Spherical symmetry is equivalent to the assertion that the radius $V = \|X - \theta\|$ is *independent* of the spatial unit vector $U = (X - \theta)/\|X - \theta\|$, which is uniformly distributed on the unit sphere $\mathcal{S}^{d-1}$ in $\mathbb{R}^d$, i.e. $U \sim \mathscr{U}(\mathcal{S}^{d-1})$. Independence of V and U of course implies

$$E[a_j(V)b_k(U)] = 0, \quad a_j \in L_{2,0}(F_V),\, b_k \in L_{2,0}(F_U), j, k = 1, 2, \ldots, \tag{6.1.1}$$

where $L_2(F) = \left\{h : \int h^2\, dF < \infty\right\}$ and $L_{2,0}(F) = \left\{h \in L_2(F) : \int h\, dF = 0\right\}$ for a distribution F, and F_V and F_U are the distribution functions of V and U respectively. There are numerous choices for a and b, for example, one can choose the sign function $a_1(v) = \text{sign}(v)$, Huber's function $a_2(v) = v\mathbf{1}[\|v\| \le 1.4] + 1.4\text{sign}(v)\mathbf{1}[\|v\| > 1.4]$ and the coordinatewise projection functions $b_k(U) = U_k, k = 1, \ldots, d$ of U.

As U is uniformly distributed over $\mathscr{U}(\mathcal{S}^{d-1})$, we can use it as side information. To this end, we shall resort to the Jackknife empirical likelihood for vector U-statistics developed by Tan, *et al.* (2015). Let $(V_i, U_i), i = 1, \ldots, n$ be a random sample of (V, U), and let $R = U_1 + U_2$ and $R^0 = R/\|R\|$. It is well known that U being uniformly distributed over the unit sphere $\mathscr{U}(\mathcal{S}^{d-1})$ is equivalent to the assertion that $\|R\|$ and R^0 are *independent*. Independence implies

$$E(c_l(\|R\|)R^0) = 0, \quad c_l \in L_{2,0}(G), l = 1, 2, \ldots,$$

where G denotes the distribution function of $\|R\|$. Note that G is known and computable. We mention the formula below in the usual case.

Remark 6.1.1 For $d = 3$, the distribution function G of $\|R\|$ is given by

$$G(r) = \frac{1}{16\pi^2} \int_{\|x\|=1, \|y\|=1, \|x+y\| \leq r} \frac{dx_1 dx_2 dy_1 dy_2}{\sqrt{(1 - x_1^2 - x_2^2)(1 - y_1^2 - y_2^2)}},$$

where $x = (x_1, x_2, x_3), y = (y_1, y_2, y_3)$ and $0 \leq r \leq 2$.

We can choose c_l as for a_j. As G is known, a systematic way of choosing $c_l, l = 1, 2, \ldots$ is to take them to be basis functions of $L_{2,0}(G)$, for example, $c_l = \varphi_l \circ G$, where $\varphi_l(t) = \sqrt{2}\cos(l\pi t), t \in (0, 1)$ is the usual orthonormal trigonometric basis. Denote $c_L = (c_1, \ldots, c_L)^\top$ for some positive integer L. Let $\kappa(U_1, U_2) = c_L(\|R\|) \otimes R^0$, where $\otimes$ denotes the Kronecker product. Then it is an argument-symmetric vector kernel and satisfies $E(\kappa(U_1, U_2)) = 0$ by the preceding independence. The vector U-statistic with κ as the kernel is now given by

$$U_n(c_L) = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \kappa(U_i, U_j).$$

The Jackknife pseudo values of the vector U-statistics are calculated by

$$V_{nj} = nU_n(c_L) - (n-1)U_{n-1}^{(-j)}(c_L), \quad j = 1, \ldots, n,$$

where $U_{n-1}^{(-j)}$ is the vector U-statistic based on the $n-1$ observations with the deletion of the jth. The preceding discussion motivates us to use the first few equations in

(6.1.1) as constraints to construct the Jackknife empirical likelihood ratio with side information as follows:

$$\mathscr{R}_n^{ssu} = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j \text{vec}\big(a_J(V_j) \otimes b_K(U_j), V_{nj}(c_L)\big) = 0\Big\},$$

where $a_J = (a^{(1)}, \dots, a^{(J)})^\top$, $b_K = (b^{(1)}, \dots, b^{(K)})^\top$, and $\text{vec}(X, Y)$ denotes the column vector consisting of stacking X, Y. Let $\mathscr{R}_n^{ss}$ be the empirical likelihood ratio when the Jackknife pseudo values are not included. It follows from Corollary 3.1 of Tan, *et al.* (2015) that the following holds. Denote $\kappa_1(u) = E(\kappa(U_1, U_2)|U_1 = u)$.

Theorem 6.1.1 *Suppose the covariance* $\text{Cov}(\text{vec}(a_J(V), b_K(U), \kappa_1(U))$ *has full rank* $JK + Ld$. *Then* $-2\log \mathscr{R}_n^{ssu}$ *has an asymptotic chisquare distribution with* $JK + Ld$ *degrees of freedom, that is,*

$$-2\log \mathscr{R}_n^{ssu} \Rightarrow \chi^2_{(JK+Ld)}.$$

If $E(a_J^{\otimes 2}(V)E(b_K^{\otimes 2}(U))$ *has full rank* JK, *then* $-2\log \mathscr{R}_n^{ss} \Rightarrow \chi^2_{JK}$.

A systematic way of choosing $a^{(j)}$ is $a^{(j)} = \varphi_j \circ F_V, j = 1, 2, \dots,$ which is an orthonormal basis of $L_{2,0}(F_V)$. But it is not computable as F_V is unknown. One can estimate it by the empirical distribution function $\hat{F}_V$ and obtain computable $\hat{a}^{(j)} = \varphi_j \circ \hat{F}_V$. With estimated constraints, we now work with the empirical likelihood ratio:

$$\hat{\mathscr{R}}_n^{ss} = \sup\Big\{\prod_{i=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j \hat{a}_J(V_j) \otimes U_j = 0\Big\}.$$

We will allow J, d to depend on the sample size n, $J = J_n, d = d_n$, and grow to infinity slowly. The following is the asymptotic result with the proof delayed to the last section.

Theorem 6.1.2 *Suppose* $J_n d_n \to \infty$ *but* $J_n^3 d_n^5 + J_n^4 d_n^3 + J_n^5 d_n^3 = o(n)$. *Then*

$$\frac{-2\log \hat{\mathscr{R}}_n^{ss} - J_n d_n}{\sqrt{J_n d_n}} \Rightarrow \mathscr{N}(0, 1). \tag{6.1.2}$$

This shows $-2\log\hat{\mathscr{R}}_n^{ss}$ is approximately chisquare distributed with $J_n d_n$ degrees of freedom. Thus for $0 < \alpha < 1$,

$$P\big(-2\log\hat{\mathscr{R}}_n^{ss} > \chi^2_{J_n d_n}(1-\alpha)\big) \to \alpha,$$

where $\chi^2_d(1-\alpha)$ is the $(1-\alpha)$-th percentile of the chisquare distribution with d degrees of freedom. Accordingly the test $\mathbf{1}[-2\log\hat{\mathscr{R}}_n^{ss} > \chi^2_{J_n d_n}(1-\alpha)]$ has an asymptotic size α. If $J_n = J$ for all n, then this is the case of the sphere with infinity dimension. If $d_n = d$ for all n, then this is the case of a d-dimensional sphere.

6.1.2 Rotational symmetry

A random vector $X \in \mathcal{S}^d$ is rotationally symmetric about a fixed direction θ if

$$X - \theta \overset{d}{=} \mathbb{O}(X-\theta),$$

for every $d \times d$ rotation matrix $\mathbb{O}$ about the direction θ in $\mathbb{R}^d$. Rotational symmetry is equivalent to the assertion that the projection $T = \theta^\top X$ of X onto the direction θ is *independent* of the unit tangent ξ at θ to $\mathcal{S}^{d-1}$, which is uniformly distributed on $\mathcal{S}^{d-2}$, i.e. $\xi \sim \mathscr{U}(\mathcal{S}^{d-2}(\theta))$, where $\mathcal{S}^{d-2}(\theta) = \big\{x \in \mathbb{R}^d : \|x\| = 1, x^\top\theta = 0\big\}$. For more details, see page 179 of Madia and Jupp (2000). Notice that X satisfies the tangent-normal equation,

$$X = T\theta + \sqrt{1-T^2}\,\xi. \tag{6.1.3}$$

Rotationally symmetric distributions include von Mises-Fisher-type distributions with densities of the form $f(\theta^\top x), x \in \mathcal{S}^d$, Waston-type distributions with densities of the form $g\big(\kappa(\theta^\top x)^2\big), x \in \mathcal{S}^d$, and Bingham-type distributions of densities of the form $h(x^\top \mathbb{K} x), x \in \mathcal{S}^d$, where f, g, h are nonnegative functions and $\mathbb{K}$ is a positive definite matrix.

In modeling directional and axial data using parametric distributions, one often wishes to test the null hypothesis that the underlying distribution is rotationally symmetric about $\theta = \theta_0$. By exploiting the preceding independence, a nonparametric test is the empirical likelihood ratio elaborated below. Let $(T_i, \xi_i), i = 1, \ldots, n$ be a random sample of (T, ξ), where $T = \theta_0^\top X$ and $\xi = (X - T\theta_0)/\sqrt{1 - T^2}$. As in the spherical symmetry, the uniform distribution of ξ can be used as side information. To this end, let $R = \xi_1 + \xi_2$ and $R^0 = R/\|R\|$. Again the statement that $\xi \sim \mathscr{U}(\mathcal{S}^{d-2})$ is equivalent to the assertion that $\|R\|$ and R^0 are independent. Let G be the distribution function of $\|R\|$. Let $d^{(m)} = \varphi_m \circ G$ and $d_M = (d_{(1)}, \ldots, d^{(M)})^\top$. This is computable as G is known. Let $\kappa(\xi_1, \xi_2) = d_M(\|R\|) \otimes R^0$. Then it is argument-symmetric and satisfies $E(\kappa(\xi_1, \xi_2)) = 0$ by the preceding independence. The vector U-statistic with κ as kernel is now given by

$$U_n(d_M) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \kappa(\xi_i, \xi_j).$$

The Jackknife pseudo values of the vector U-statistics are calculated by

$$V_{nj}(d_M) = nU_n(d_M) - (n-1)U_{n-1}^{(-j)}(d_M), \quad j = 1, \ldots, n,$$

Analogously we construct the Jackknife empirical likelihood ratio with side information as follows:

$$\mathscr{R}_n^{rsu} = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j \mathrm{vec}(a_J(T_j) \otimes b_K(\xi_j), V_{nj}(d_M)\big) = 0 \Big\},$$

for some choices of $a_J = (a^{(1)}, \ldots, a^{(J)})^\top$ and $b_K = (b^{(1)}, \ldots, b^{(K)})^\top$. Let $\mathscr{R}_n^{rs}$ be the empirical likelihood ratio when the jackknife pseudo values are not included. It follows from Corollary 3.1 of Tan, *et al.* (2015) that the following holds. Let $\xi_1(x) = E(\kappa(\xi_1, \xi_2)|\xi_2 = x)$.

Theorem 6.1.3 *Suppose the matrix* $\mathrm{Cov}(\mathrm{vec}(a_J(T)), b_K(\xi), \kappa_1(\xi))$ *has full rank* $JK + M(d-1)$. *Then* $-2\log \mathscr{R}_n^{rsu} \Rightarrow \chi^2_{(JK+M(d-1))}$. *If* $E(a_J^{\otimes 2}(T)E(b_K^{\otimes 2}(\xi))$ *has full rank* JK, *then* $-2\log \mathscr{R}_n^{rs} \Rightarrow \chi^2_{JK}$.

As the distribution F_T of T is unknown, we estimate it by the empirical distribution function $\hat{F}_T$. Let J_n, d_n be positive integers and tend to infinity and take

$\hat{a}_{J_n} = (a^{(1)}, \ldots, a^{(J_n)})^\top = (\varphi_1, \ldots, \varphi_{J_n})^\top \circ \hat{F}_T$. With the estimated constraints, the empirical likelihood ratio as test statistic is now given by

$$\hat{\mathscr{R}}_n^{rs} = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j \hat{a}^{(J_n)}(T_j) \otimes \xi_j = 0\Big\}.$$

We have the following.

Theorem 6.1.4 *Suppose $J_n d_n \to \infty$ but $J_n^5 d_n^5 = o(n)$. Then*

$$\frac{-2\log \hat{\mathscr{R}}_n^{rs} - J_n(d_n-1)}{\sqrt{J_n(d_n-1)}} \Rightarrow \mathscr{N}(0,1). \tag{6.1.4}$$

Thus $-2\log \hat{\mathscr{R}}_n^{rs}$ is approximately chisquare distributed with $J_n(d_n-1)$ degrees of freedom and the test $\mathbf{1}[-2\log \hat{\mathscr{R}}_n^{rs} > \chi^2_{J_n(d_n-1)}(1-\alpha)]$ has an asymptotic size $\alpha \in (0,1)$.

6.1.3 Antipodally symmetric distributions

Let X have the continuous generalized Scheiddegger-Watson distribution, i.e., the density is of the form $g(\|x_v\|), x \in \mathcal{S}^{d-1}$, where g is some known function, and x_v is the part of x in an s-dimensional subspace $\mathcal{V}$. Then the tangent-norm equation of X is given by

$$X = T\eta + (1-T^2)^{1/2}\xi, \quad T = \|x_v\|,\ \eta \in \mathcal{V},\ \xi \in \mathcal{V}^\perp.$$

where $\|\eta\| = 1$, $\eta \in \mathcal{V}$, $\|\xi\| = 1$, $\xi \in \mathcal{V}^\perp$. Here we take $\eta = X_v/\|X_v\|$, $\xi = (X - T\eta)/(1-T^2)^{1/2}$. A relationship similar to the rational symmetry is that T, η and ξ are independent, and η and ξ are uniformly distributed on unit spheres in $\mathcal{V}$ and $\mathcal{V}^\perp$. Analogous to the preceding discussions, one can construct an empirical likelihood ratio test and we shall omit the details.

6.1.4 Coordinatewise symmetry

A random vector X in $\mathbb{R}^d$ has a distribution coordinatewise symmetric about θ if

$$(X_1-\theta_1, \ldots, X_d-\theta_d) \stackrel{d}{=} (s_1(X_1-\theta_1), \ldots, s_d(X_d-\theta_d)), \quad s_j = \pm 1,\ j = 1, \ldots, d.$$

Coordinatewise symmetry of X about θ is equivalent to the assertion that the coordinatewise radius vector $V = (V_1, \ldots, V_d)^\top$ is *independent* of the coordinatewise sign vector $U = (U_1, \ldots, U_d)^\top$, where $V_j = \|X_j - \theta_j\|$ and $U_j = \text{sign}(X_j - \theta_j)$. Independence implies

$$E[a^{(j)}(V)b^{(k)}(U)] = 0, \quad a^{(j)} \in L_{2,0}(F_V), b^{(k)} \in L_{2,0}(F_U), j, k = 1, 2, \ldots.$$

Analogous to the preceding discussions, we choose several $a^{(j)} \in L_{2,0}(F_V), j = 1, \ldots, J$ and $b^{(k)} \in L_{2,0}(F_U), k = 1, \ldots, K$ and construct the empirical likelihood ratio:

$$\mathscr{R}_n^{cs} = \sup\Big\{ \prod_{i=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j a_J(V_i) \otimes b_K(U_i) = 0 \Big\},$$

By Owen's theorem, we have the following.

Theorem 6.1.5 *Suppose the matrix $E(a_J a_J^\top(V)) \otimes E(b_K b_K^\top(U))$ has full rank JK. Then*

$$-2 \log \mathscr{R}_n^{cs} \Rightarrow \chi^2_{JK}.$$

6.1.5 Exchangeability

A random vector X in $\mathbb{R}^d$ is exchangeable if

$$(X_1, \ldots, X_d) \stackrel{d}{=} (X_{\pi_1}, \ldots, X_{\pi_d})$$

for every permutation $\pi_1, \ldots, \pi_d$ of $1, \ldots, d$. Then $O = (X_{d,1}, \ldots, X_{d,d})^\top$, where $X_{d,1} \leq \ldots \leq X_{d,d}$ are the order statistics of X, is *independent* of $R = (R_1, \ldots, R_d)^\top$, where $R_j = \sum_{i=1}^d \mathbf{1}[X_i \leq X_j]$ are the rank statistics. Independence implies

$$E[a^{(j)}(O)b^{(k)}(R)] = 0, \quad a^{(j)} \in L_{2,0}(F_O), b^{(k)} \in L_{2,0}(F_R), \quad j, k = 1, 2, \ldots.$$

In the same fashion, we choose several $a^{(j)} \in L_{2,0}(F_O), j = 1, \ldots, J$ and $b^{(k)} \in L_{2,0}(F_R), k = 1, \ldots, K$ and construct the empirical likelihood ratio:

$$\mathscr{R}_n^{es} = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \sum_{j=1}^n \pi_j a_J(O_j) \otimes b_K(R_j) = 0 \Big\},$$

where $(O_j, R_j), j = 1, \ldots, n$ is a random sample of (O, R).

By Owen's theorem, we have the following.

Theorem 6.1.6 *Suppose the matrix* $E(a_J a_J^\top(O)) \otimes E(b_K b_K^\top(R))$ *has full rank* JK*. Then*

$$-2\log \mathscr{R}_n^{es} \Rightarrow \chi^2_{JK}.$$

6.2 Simulation study

6.2.1 Testing the center of spherical symmetry

Consider the null hypothesis that the center of symmetry of a distribution is the origin of $\mathbb{R}^d$ versus the alternative hypothesis that the center is θ_1 different from the origin,

$$H_0 : \theta = \mathbf{0} \quad \text{vs} \quad H_1 : \theta = \theta_1.$$

We were interested in the power performance of the proposed tests at the nominal level of significance $\alpha = .05$ in different cases. Specifically, we looked at the following cases. Case 1. How the power of the tests decreases as the dimension d grows with respect to different sample sizes n. Case 2. How the power of the tests increases as θ_1 moves away from the origin for a fixed sample size n with respect to different dimensions d and different number of constraints J, K. Case 3. How the side information increases the power of the tests. We examined the test $\mathbf{1}[-2\log \hat{\mathscr{R}}_n^{ss} > \chi^2_{J_n d_n}(1-\alpha)]$ in Cases 1 and 2 and the test $\mathbf{1}[-2\log \mathscr{R}_n^{ssu} > \chi^2_{JK+Ld}(1-\alpha)]$ in Case 3. In the latter situation, we choose $a(v) = (\text{sign}(v), v\mathbf{1}[\|v\| \le 1.4] + 1.4\text{sign}(v)\mathbf{1}[\|v\| > 1.4])^\top$ and $b_K(U) = U$. As $E(U) = 0$, these choices still satisfy the equalities in (6.1.1) by the independence of V and U even though the components $a^{(1)}, a^{(2)}$ of a are not in $L_{2,0}(F_V)$. Examining Cases 2 and 3, one observes that there is significant power increase and sample size reduction with the use of the side information

Table 6.1.
Case 1: The simulated $\alpha = .05$ level of significance of the EL test about the center of spherical symmetry $H_0 : \theta = \mathbf{0}$ with $J = r$, $d = \text{dim}$, $n = 400$ and $m = 2000$ repetitions. Data generated from multivariate normal.

	r=1	r=2	r=3	r=4	r=5
dim=1	0.0395	0.0520	0.0470	0.0585	0.0465
dim=2	0.0475	0.0610	0.0480	0.0460	0.0560
dim=3	0.0570	0.0605	0.0520	0.0615	0.0595
dim=4	0.0555	0.049	0.05450	0.06500	0.0705

Table 6.2.
Case 1: The simulated power of the test about the center of spherical symmetry $H_0 : \theta = \mathbf{0}$ vs $H_1 : \theta = \text{rep}(0.4, \text{dim})$ with $J = r$, $d = \text{dim}$, and $m = 2000$ repetitions at $\alpha = 0.05$ level of significance. Data generated from multivariate normal.

	n	r=1	r=2	r=3	r=4	r=5
dim=1	50	0.3170	0.2425	0.2005	0.183	0.1835
	100	0.5460	0.4070	0.3810	0.3530	0.3175
	150	0.7190	0.6260	0.5880	0.5105	0.4500
	200	0.8350	0.7500	0.7285	0.6495	0.6190
	250	0.9035	0.8590	0.8200	0.7870	0.7285
	300	0.9425	0.9165	0.8900	0.8550	0.8225
	350	0.9745	0.9530	0.9330	0.9050	0.8870
	400	0.9875	0.9780	0.9680	0.9565	0.9390
	450	0.9955	0.9900	0.9845	0.9715	0.9685
	500	0.9975	0.9910	0.9880	0.9815	0.9820
	600	1	0.9985	0.9975	0.9980	0.9930
	700	1	1	0.9995	0.9990	0.9985
	800	1	1	1	1	0.9995
dim=2	50	0.2180	0.1690	0.1720	0.2035	0.2565
	100	0.4290	0.3190	0.2480	0.2555	0.2335
	150	0.6125	0.4870	0.3805	0.3675	0.3200
	200	0.7525	0.6245	0.5580	0.4795	0.4495
	250	0.8660	0.7440	0.6735	0.6200	0.5745
	300	0.9065	0.8485	0.7715	0.7165	0.6745
	350	0.9620	0.8970	0.8615	0.8130	0.7615
	400	0.9860	0.9425	0.9140	0.8595	0.8210
	450	0.9905	0.9625	0.9480	0.9185	0.8965
	500	0.9955	0.9815	0.9675	0.9555	0.9280
	600	0.9990	0.9940	0.9915	0.9885	0.9815
	700	1	0.9980	0.9970	0.991	0.991
	800	1	1	1	0.998	0.9985
dim=3	50	0.1745	0.1660	0.2390	0.3365	0.4720
	100	0.3355	0.2325	0.2305	0.2110	0.2440
	150	0.4845	0.3285	0.2980	0.2770	0.2705
	200	0.6430	0.4755	0.4165	0.3395	0.3175
	250	0.7685	0.6040	0.5250	0.4370	0.3945
	300	0.8630	0.7050	0.6370	0.5565	0.5060
	350	0.9100	0.8095	0.7355	0.6710	0.6040
	400	0.9475	0.8680	0.8070	0.7285	0.6840
	450	0.9700	0.9105	0.8720	0.7980	0.7525
	500	0.9865	0.9405	0.9125	0.8645	0.8360
	600	0.9945	0.9790	0.9600	0.9350	0.9230
	700	0.9995	0.9955	0.9890	0.9695	0.9580
	800	1	0.9975	0.9960	0.9920	0.9860
dim=4	50	0.1580	0.1995	0.353	0.5375	0.7570
	100	0.2510	0.2120	0.2125	0.2580	0.3345
	150	0.3830	0.2680	0.2255	0.2250	0.2525
	200	0.5045	0.3350	0.3195	0.2640	0.2845
	250	0.6465	0.4580	0.4050	0.3515	0.3335
	300	0.7465	0.5655	0.5050	0.4315	0.3600
	350	0.8420	0.6590	0.5925	0.5150	0.4345
	400	0.8825	0.7665	0.6560	0.5825	0.4920
	450	0.9260	0.8150	0.7440	0.6710	0.5965
	500	0.9665	0.8670	0.8260	0.7335	0.6605
	600	0.9855	0.9470	0.9075	0.8430	0.8100
	700	0.9980	0.9780	0.9605	0.9085	0.8885
	800	0.9995	0.9905	0.9840	0.9590	0.9370

Table 6.3.
Case 2: The simulated power of the test about the center of spherical symmetry $H_0 : \theta = \mathbf{0}$ vs $H_1 : \theta = \theta_1$ with $J = r$, $d =$ dim, $n = 400$ and $m = 2000$ repetitions at $\alpha = 0.05$ level of significance. Data generated from multivariate normal.

	θ_1	r=1	r=2	r=3	r=4	r=5
dim=1	0.05	0.0920	0.0820	0.0735	0.0695	0.0685
	0.10	0.1875	0.1595	0.1460	0.1345	0.1170
	0.15	0.3615	0.2860	0.2535	0.2450	0.2305
	0.20	0.5980	0.4990	0.4540	0.4090	0.4170
	0.25	0.7875	0.6995	0.6440	0.6080	0.5685
	0.30	0.9020	0.8415	0.7985	0.7630	0.7500
	0.35	0.9580	0.9265	0.9085	0.8675	0.8595
	0.40	0.9910	0.9775	0.9670	0.9490	0.9365
dim=2	(0.05,0.05)	0.0845	0.0755	0.0770	0.0670	0.0630
	(0.10,0.10)	0.1830	0.1405	0.1290	0.1165	0.1020
	(0.15,0.15)	0.3615	0.2670	0.2350	0.2005	0.2020
	(0.20,0.20)	0.5405	0.4370	0.4125	0.3400	0.3230
	(0.25,0.25)	0.7510	0.6275	0.5830	0.4840	0.4620
	(0.30,0.30)	0.8680	0.7995	0.7515	0.6675	0.6310
	(0.35,0.35)	0.9480	0.8920	0.8425	0.7995	0.7590
	(0.40,0.40)	0.9770	0.9410	0.9075	0.8820	0.8500
dim=3	(0.05,0.05,0.05)	0.0865	0.0665	0.0680	0.0745	0.0690
	(0.10,0.10,0.10)	0.1695	0.1235	0.1210	0.1180	0.1040
	(0.15,0.15,0.15)	0.3065	0.2415	0.2030	0.1905	0.1655
	(0.20,0.20,0.20)	0.5020	0.3765	0.3420	0.3010	0.2410
	(0.25,0.25,0.25)	0.6805	0.5585	0.4730	0.4155	0.3825
	(0.30,0.30,0.30)	0.8145	0.6690	0.6265	0.5530	0.5045
	(0.35,0.35,0.35)	0.8995	0.7935	0.7295	0.6390	0.5995
	(0.40,0.40,0.40)	0.9460	0.8550	0.8040	0.7400	0.6790
dim=4	(0.05,0.05,0.05,0.05)	0.0605	0.0610	0.0695	0.0690	0.0805
	(0.10,0.10,0.10,0.10)	0.1320	0.1165	0.1160	0.1120	0.1060
	(0.15,0.15,0.15,0.15)	0.2710	0.1965	0.1730	0.1480	0.1660
	(0.20,0.20,0.20,0.20)	0.4315	0.3045	0.2870	0.2440	0.2425
	(0.25,0.25,0.25,0.25)	0.6165	0.4485	0.4160	0.3560	0.3455
	(0.30,0.30,0.30,0.30)	0.7440	0.5690	0.5240	0.4490	0.392
	(0.35,0.35,0.35,0.35)	0.8310	0.6935	0.5895	0.5325	0.4885
	(0.40,0.40,0.40,0.40)	0.8960	0.7435	0.6640	0.5750	0.5225

Table 6.4.
Case 3: The simulated $\alpha = .05$ level of significance of the test about the center of spherical symmetry $H_0 : \theta = \mathbf{0}$ with $J = K = r$, $d = L =$ dim, $n = 50$, and $m = 100$ repetitions. Data generated from multivariate normal.

	r=1	r=2	r=3	r=4	r=5
dim=2	0	0	0.08	0.06	0.08
dim=3	0.02	0.03	0.08	0.10	0.14
dim=4	0.03	0.02	0.10	0.26	0.41

Table 6.5.
Case 3: The simulated power of the test about the center of spherical symmetry $H_0 : \theta = \mathbf{0}$ vs $H_1 : \theta = \theta_1$ with $J = K = r$, $d = L =$ dim, $n = 50$, and $m = 100$ repetitions at $\alpha = 0.05$ level of significance. Data generated from multivariate normal.

	θ_1	r=1	r=2	r=3	r=4	r=5
dim=2	(0.1,0.1)	0.11	0.05	0.09	0.15	0.13
	(0.2,0.2)	0.17	0.23	0.15	0.20	0.25
	(0.3,0.3)	0.59	0.53	0.39	0.47	0.46
	(0.4,0.4)	0.82	0.79	0.76	0.74	0.79
	(0.5,0.5)	0.98	0.91	0.95	0.90	0.93
dim=3	(0.1,0.1,0.1)	0.06	0.06	0.12	0.14	0.21
	(0.2,0.2,0.2)	0.31	0.20	0.31	0.34	0.5
	(0.3,0.3,0.3)	0.69	0.55	0.59	0.66	0.84
	(0.4,0.4,0.4)	0.91	0.92	0.90	0.91	0.95
	(0.5,0.5,0.5)	0.99	0.99	0.97	1	1

6.2.2 Testing rotational symmetry

Consider the null hypothesis that the direction of rotational symmetry of a distribution is $(0,0,1)$ versus the alternative hypothesis that the direction is θ_1 different from $(0,0,1)$,

$$H_0 : \theta = (0,0,1) \quad \text{vs} \quad H_1 : \theta = \theta_1.$$

We looked at the same Cases 1 and 2 as in Subsection (6.2.1). Specifically, we studied the test $\mathbf{1}[-2\log\hat{\mathscr{R}}_n^{rs} > \chi^2_{J_n(d_n-1)}(1-\alpha)]$ and chose the same a_J as in Subsection (6.2.1). Here the data were generated from the von Mises-Fisher distribution. This is a probability distribution defined on the sphere $\mathcal{S}^d$ with the pdf given by

$$f_d(\theta^\top x;\kappa) = C_d(\kappa)\exp(\kappa\theta^\top x), \quad x \in \mathcal{S}^d,$$

where $\kappa \geq 0$, $\|\theta\| = 1$ and the normalization constant $C_d(\kappa)$ is given by

$$C_d(\kappa) = \frac{\kappa^{d/2-1}}{(2\pi)^{d/2} I_{d/2-1}(\kappa)},$$

where I_v denotes the modified Bessel function of the first kind with order v. When $d = 3$, $C_3(\kappa)$ reduces to

$$C_3(\kappa) = \frac{\kappa}{4\pi\sinh\kappa} = \frac{\kappa}{2\pi(\exp(\kappa) - \exp(-\kappa))}.$$

The parameter θ is the mean direction and the parameter κ the concentration parameter. The distribution is more concentrated around the mean direction θ with higher κ. When $\kappa = 0$, the distribution is uniform on $\mathcal{S}^d$. Obviously, the von Mises-Fisher distribution is rotationally symmetric about the mean direction θ.

6.3 Proof of Theorem 6.1.2

We shall apply Theorem 7.4 of Peng and Schick (2013) to prove the result. For self-containedness, we quote their result below. Assume that $(\mathcal{Z}, \mathscr{S})$ is a measurable space, that $Z_1, \ldots, Z_n$ are independent copies of the $\mathcal{Z}$-valued random variable Z with distribution Q, and that m_n is a positive integer that tends to infinity with n. Let w_n

Table 6.6.
Case 1: The simulated $\alpha = 0.05$ level of significance of the test about the direction of rotational symmetry $H_0 : \theta = (0,0,1)^\top$ with $d = 3$, $J = r$, and $m = 2000$ repetitions. Data generated from the von Mises-Fisher distribution.

	r=1	r=3	r=5
n=50	0.0505	0.0990	0.2275
n=100	0.0530	0.0685	0.0865

Table 6.7.
Case 2: The simulated power of the test about rotational symmetry $H_0 : \theta = (0,0,1)^\top \quad vs \quad H_1 : \theta = (0.14, 0.14, 0.98)^\top$ with $d = 3$, $J = r$, and $m = 2000$ repetitions at $\alpha = 0.05$. Data generated from the von Mises-Fisher distribution.

	r=1	r=3	r=5
n=50	0.949	0.9995	0.9955
n=100	0.9995	1	1

denote a measurable function from $\mathcal{Z}$ to $\mathbb{R}^{m_n}$ such that $\int w_n \, dQ = 0$ and $\int \|w_n\|^2 \, dQ$ is finite. Let $\hat{w}_n$ be an estimator of w_n. Consider

$$\hat{\mathscr{R}}_n = \sup\Big\{ \prod_{j=1}^n n\pi_j : \pi \in \mathscr{P}_n, \frac{1}{n}\sum_{j=1}^n \pi_j \hat{w}_n(Z_j) = 0 \Big\}.$$

Let $\|\mathbb{M}\|_o$ be the spectral norm (the largest eigenvalue) of matrix $\mathbb{M}$ and set

$$\mathbb{W}_n = \int w_n w_n^\top \, dQ, \quad \text{and} \quad \hat{\mathbb{W}}_n = \frac{1}{n}\sum_{j=1}^n \hat{w}_n \hat{w}_n^\top(Z_j).$$

Lemma 6.3.1 *Suppose the $m_n \times m_n$ dispersion matrices $\mathbb{W}_n$ is* regular *in the sense that*

$$0 < \inf_n \inf_{\|u\|=1} u^\top \mathbb{W}_n u \le \sup_n \sup_{\|u\|=1} u^\top \mathbb{W}_n u < \infty.$$

Assume

$$m_n \max_{1\le j\le n} \|\hat{w}_n(Z_j)\| = o_p(n^{1/2}), \tag{6.3.1}$$

$$\|\hat{\mathbb{W}}_n - \mathbb{W}_n\|_o = o_p(m_n^{-1/2}) \tag{6.3.2}$$

and

$$\frac{1}{n}\sum_{j=1}^n \hat{w}_n(Z_j) = \frac{1}{n}\sum_{j=1}^n v_n(Z_j) + o_p(n^{-1/2}) \tag{6.3.3}$$

for some measurable function v_n from $\mathcal{S}$ into $\mathbb{R}^{m_n}$ such that $\int v_n \, dQ = 0$ and $\|v_n\|$ is Lindeberg, that is, for every $\epsilon > 0$, $\int \|v_n\|^2 \mathbf{1}[\|v_n\| > \epsilon\sqrt{n}] \, dQ \to 0$. Suppose the dispersion matrix $\mathbb{U}_n = \mathbb{W}_n^{-1/2} \int v_n v_n^\top \, dQ \mathbb{W}_n^{-1/2}$ of $\mathbb{W}_n^{-1/2} v_n(Z)$ satisfies $\|\mathbb{U}_n\|_o = O(1)$ and $m_n/\mathrm{trace}(\mathbb{U}_n^2)$ is bounded. Then, as m_n tends to infinity with n,

$$\frac{-2\log \hat{\mathscr{R}}_n - \mathrm{trace}(\mathbb{U}_n)}{\sqrt{2\mathrm{trace}(\mathbb{U}_n^2)}} \implies \mathscr{N}(0,1).$$

Proof of Theorem 6.1.2. We shall prove this by applying Lemma 6.3.1 with

$$Z = (V, U^\top)^\top, \quad w_n(Z) = \sqrt{d_n} a_{J_n}(V) \otimes U \quad \text{and} \quad \hat{w}_n(Z) = \sqrt{d_n} \hat{a}_{J_n}(V) \otimes U.$$

Then $m_n = J_n d_n$ and $\mathbb{W}_n = d_n E(a_{J_n}^{\otimes 2}(V)) \otimes E(U^{\otimes 2}) = \mathbb{I}_{J_n} \otimes \mathbb{I}_{d_n}$ as $E(U^{\otimes 2}) = \mathbb{I}_{d_n}/d_n$. Thus $\mathbb{W}_n$ is regular. Clearly (6.3.1) is met as $\|\hat{w}(Z_j)\| \le \sqrt{2} J_n^{1/2} d_n$ and $J_n^3 d_n^4 = o(n)$. For (6.3.2), by the triangle inequality,

$$\|\hat{\mathbb{W}}_n - \mathbb{W}_n\|_o \le \|\hat{\mathbb{W}}_n - \bar{\mathbb{W}}_n\|_o + \|\bar{\mathbb{W}}_n - \mathbb{W}_n\|_o. \tag{6.3.4}$$

Since

$$nE\|\bar{\mathbb{W}}_n - \mathbb{W}_n\|^2 \le d_n^2 E\|a_J(V) \otimes U\|^4 \le 4J_n^2 d_n^4,$$

it follows $\|\bar{\mathbb{W}}_n - \mathbb{W}_n\|_o \le 2J_n d_n^2/\sqrt{n} = o(m_n^{-1/2})$ as $J_n^3 d_n^5 = o(n)$. Note

$$\|\hat{\mathbb{W}}_n - \bar{\mathbb{W}}_n\|_o \le D_n + 2\|\bar{\mathbb{W}}_n\|_o^{1/2} D_n^{1/2},$$

where $D_n = d_n \frac{1}{n}\sum_{j=1}^n \|[\hat{a}_{J_n}(V_j) - a_{J_n}(V_j)] \otimes U_j\|^2$. Thus (6.3.2) is implied by $D_n = o_p(m_n^{-1})$ in view of (6.3.4). To this end, let $\varphi_n(t) = (\varphi^{(1)}, \ldots, \varphi^{(J_n)})^\top$ so $a_{J_n} = \varphi_n \circ F_V$ and $\hat{a}_{J_n} = \varphi_n \circ \hat{F}_V$. We need the following properties of the trigonometric basis: for $t \in [0, 1]$,

$$\|\varphi_n(t)\| \le (2J_n)^{1/2}, \quad \|\varphi_n'(t)\| \le \sqrt{2}\pi J_n^{3/2}, \quad \|\varphi_n''(t)\| \le \sqrt{2}\pi^2 J_n^{5/2}, \tag{6.3.5}$$

where φ_n' and φ_n'' denote the first and second order derivatives of φ. Using the second inequality in (6.3.5), we derive

$$\frac{1}{n}\sum_{j=1}^n \|\varphi_n(\hat{F}_V(V_j)) - \varphi_n(F_V(V_j))\|^2 \le 2\pi^2 J_n^3 \sup_{t\in\mathbb{R}} \|\hat{F}_V(t) - F_V(t)\| = O_P(J_n^3/n).$$

Hence the desired $D_n = O_P(J_n^3 d_n^2/n) = o_P(m_n^{-1})$ as $J_n^4 d_n^3 = o(n)$. We now show (6.3.3) holds with $v_n(Z) = \sqrt{d_n} a_{J_n}(V) \otimes U = \sqrt{d_n}\varphi_n \circ F_V(V) \otimes U$. Clearly $E(v_n(Z)) = 0$ as $E(U) = 0$ and V are independent of U, and v_n is Lindeberg as $\|v_n(Z)\| \le \sqrt{2} J_n^{1/2} d_n = o(\sqrt{n})$. Moreover, since $\int v_n v_n^\top\, dQ = \mathbb{W}_n$ it follows that $\mathbb{U}_n = \mathbb{I}_{J_n} \otimes \mathbb{I}_{d_n}$, hence $\mathbb{U}_n = O_p(1)$ and $m_n/\mathrm{trace}(\mathbb{U}_n^2) = 1$. Now using Taylor expansion, we write

$$\frac{1}{n}\sum_{j=1}^n \sqrt{d_n}\big(\varphi_n(\hat{F}_V(V_j)) - \varphi_n(F_V(V_j))\big) \otimes U_j = L_n + M_n, \quad \text{say},$$

where

$$L_n = \frac{1}{n}\sum_{j=1}^n \sqrt{d_n}\varphi_n'(F_V(V_j))\big(\hat{F}_V(V_j) - F_V(V_j)\big) \otimes U_j,$$

$$M_n = \frac{1}{n}\sum_{j=1}^n \sqrt{d_n}\varphi_n''(F_{nj}^*)\big(\hat{F}_V(V_j) - F_V(V_j)\big)^2 \otimes U_j,$$

where F^*_{nj} lies in between $\hat{F}_V(V_j)$ and $F_V(V_j)$. Using the second inequality in (6.3.5), we get

$$\begin{aligned}
E\big(\|L_n\|^2\big) &= \operatorname{trace}\big(E(L_n^{\otimes 2})\big) \\
&= \frac{d_n}{n}\operatorname{trace}\Big(E\big[\varphi_n'(V_1)^{\otimes 2}\big(\hat{F}_V(V_1) - F_V(V_1)\big)^2\big] \otimes E(U_1^{\otimes 2})\Big) \\
&\le \frac{d_n}{n} E\Big(\|\varphi_n'(F_V(V_1))\|^2\big(\hat{F}_V(V_1) - F_V(V_1)\big)^2\Big) \\
&\le 2\pi^2 \frac{J_n^3 d_n}{n} E\big(\sup_{t\in\mathbb{R}}\big(\hat{F}_V(t) - F_V(t)\big)^2\big) \\
&= O_P(J_n^3 d_n/n^2) = o_P(n^{-1})
\end{aligned}$$

as $J_n^3 d_n = o(n)$. This shows $L_n = o_P(n^{-1/2})$. Using the third inequality in (6.3.5), we find

$$\|M_n\| \le \sqrt{2}\pi^2 J_n^{5/2} d_n^{3/2} \sup_{t\in\mathbb{R}} \|\hat{F}_V(t) - F_V(t)\|^2 = O_P(J_n^{5/2} d_n^{3/2}/n) = o_P(n^{-1/2})$$

as $J_n^5 d_n^3/n = o(1)$. This yields $M_n = o_P(n^{-1/2})$ and hence the desired (6.3.3). We now apply Lemma 6.3.1 to complete the proof. ■

7. TECHNICAL DETAILS

In this chapter, we prove the lemmas and theorems.

7.1 Proofs for Lemmas in Chapter 3

Let us recall some expressions and properties of U-statistics and their jackknife pseudo values. Given the definitions in the first section, by the Hoeffding decomposition, we have

$$V_{nj}(\theta) = m\tilde{h}_1(Z_j;\theta) + R_{nj}(\theta), \quad j = 1, \ldots, n. \tag{7.1.1}$$

where R_{nj} is the remainder given by

$$R_{nj}(\theta) = \sum_{c=2}^{m} \binom{m}{c} \left(nU_{nc}(h_c^*(\theta)) - (n-1)U_{(n-1)c}^{(-j)}(h_c^*(\theta)) \right), \; j = 1, \ldots, n. \tag{7.1.2}$$

After rearranging equation (7.1.2), we have for $j = 1, \ldots, n$ that

$$R_{nj}(\theta) = \sum_{c=2}^{m}\binom{m}{c}\left(cU_{(n-1)(c-1)}(h_{(c-1)j}^*(\theta)) - (c-1)U_{(n-1)c}^{(-j)}(h_c^*(\theta))\right), \tag{7.1.3}$$

where

$$h_{(c-1)j}^* = h_c^*(Z_j; z_1, \ldots, z_{c-1}). \tag{7.1.4}$$

Proof of Lemma 3.2.1. Let $\theta_{nt} = \theta_0 + n^{-1/2}t$. By (7.1.1) and Cauchy inequality,

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n} \|V_{nj}(\theta_{nt}) - m\tilde{h}_1(Z_j;\theta_0)\|^2 \\
=&\frac{1}{n}\sum_{j=1}^{n} \|m\tilde{h}_1(Z_j;\theta_{nt}) + R_{nj}(\theta_{nt}) - m\tilde{h}_1(Z_j;\theta_0)\|^2 \\
=&\frac{1}{n}\sum_{j=1}^{n} \|m[\tilde{h}_1(Z_j;\theta_{nt}) - \tilde{h}_1(Z_j;\theta_0)] + R_{nj}(\theta_{nt})\|^2 \\
\leq&\frac{1}{n}\sum_{j=1}^{n} 2\left(\|m[\tilde{h}_1(Z_j;\theta_{nt}) - \tilde{h}_1(Z_j;\theta_0)]\|^2 + \|R_{nj}(\theta_{nt})\|^2\right) \\
=&\frac{2m^2}{n}\sum_{j=1}^{n} \|\tilde{h}_1(Z_j;\theta_{nt}) - \tilde{h}_1(Z_j;\theta_0)\|^2 + \frac{2}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt})\|^2 \\
=&\frac{2m^2}{n}\sum_{j=1}^{n} \|h_1(Z_j;\theta_{nt}) - h_1(Z_j;\theta_0)\|^2 + \frac{2}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt})\|^2 \\
:=&2m^2 A_n(t) + 2B_n(t).
\end{aligned}$$

It follows that

$$\hat{D}_n(C) \leq 2m^2 \sup_{\|t\|\leq C} A_n(t) + 2\sup_{\|t\|\leq C} B_n(t).$$

Let $\tilde{Z}_1, \ldots, \tilde{Z}_m$ be i.i.d copies of Z_1. Then

$$h_1(Z_j;\theta) = E(h(\tilde{Z}_1, \ldots, \tilde{Z}_m;\theta)|\tilde{Z}_1 = Z_j).$$

For large n and $\|t\| \leq C$, $\theta_{nt} \in N(\theta_0)$, so that by (A1) we get

$$\begin{aligned}
A_n(t) &= \frac{1}{n}\sum_{j=1}^{n} \|E\left(h(\tilde{Z}_1, \ldots, \tilde{Z}_m;\theta_{nt}) - h(\tilde{Z}_1, \ldots, \tilde{Z}_m;\theta_0)|\tilde{Z}_1 = Z_j\right)\|^2 \\
&\leq \frac{1}{n}\sum_{j=1}^{n} E\left(\|\frac{\partial h}{\partial \theta}(\tilde{Z}_1, \ldots, \tilde{Z}_m;\theta^*_{jt})\|^2|\tilde{Z}_1 = Z_j\right)\|n^{-1/2}t\|^2 \\
&\leq C^2\frac{1}{n}\left(\frac{1}{n}\sum_{j=1}^{n} E(G^2(\tilde{Z}_1, \ldots, \tilde{Z}_m)|\tilde{Z}_1 = Z_j)\right)
\end{aligned}$$

where θ^*_{jt} lies between θ_0 and θ_{nt}. By the law of large umbers, for $M > 0$,

$$P\left(\frac{1}{n}\sum_{j=1}^{n} E(G^2(\tilde{Z}_1, \ldots, \tilde{Z}_m)|\tilde{Z}_1 = Z_j) > M\right) \leq \frac{E(G^2)}{M} = o(1)$$

since G is square-integrable. This shows

$$A_n(t) = O_P(\frac{1}{n}) \quad \text{uniformly in} \quad \|t\| \leq C.$$

Next we show this also holds for $B_n(t)$.

Without loss of generality, we prove the case of $\theta \in \mathbb{R}^2$. Let's denote the coordinates of t as (t_1, t_2). Select C_0 as $-C < C_0 < C$, then point $t_0 = (C_0, C_0)$ is located inside the circle $\{\|t\| < C\}$.

$$\sup_{\|t\| \leq C} B_n(t) \leq B_n(t_0) + \sup_{\|t\| \leq C} |B_n(t) - B_n(t_0)|. \tag{7.1.5}$$

For any $\epsilon > 0$,

$$\begin{aligned} &P(\sup_{\|t\| \leq C} B_n(t) > \epsilon) \\ \leq &P(B_n(t_0) > \epsilon/2) + P(\sup_{\|t\| \leq C} |B_n(t) - B_n(t_0)| > \epsilon/2) \\ \leq &\frac{EB_n(t_0)}{\epsilon/2} + P(\sup_{\|t\| \leq C} |B_n(t) - B_n(t_0)| > \epsilon/2) \\ := &P_{1n} + P_{2n}. \end{aligned}$$

By (7.1.3) and the Cauchy inequality,

$$\begin{aligned} EB_n(t_0) = &\frac{1}{n} \sum_{j=1}^{n} E\|R_{nj}(\theta_{nt_0})\|^2 = E\|R_{n1}(\theta_{nt_0})\|^2 \\ \leq &2m \sum_{c=2}^{m} \binom{m}{c}^2 \big\{c^2 Var(U_{(n-1)(c-1)}(h^*_{(c-1)j}(\theta_{nt_0}))) \\ &+ (c-1)^2 Var(U^{(-j)}_{(n-1)c}(h^*_c(\theta_{nt_0})))\big\}. \end{aligned}$$

In the proof of Theorem 7.1 of Peng and Tan (2016), it suffices to show $Var(U_n(\theta_{nt_0})) = O(\frac{1}{n})$ holds. Therefore we get

$$E\|U_n(\theta_{nt_0})\|^2 = O(\frac{1}{n}),$$

and this proves $P_{1n} = O(\frac{1}{n})$. Let us now show $P_{2n} = O(\frac{1}{n})$. To this end, let

$$r_{nj}(t) = R_{nj}(\theta_{nt}) - R_{nj}(\theta_{nt_0}), \quad \|t\| \leq C.$$

Using (7.1.3) again, we have

$$\begin{aligned}r_{nj}(t) =& \sum_{c=2}^{m}\binom{m}{c}c[U_{(n-1)(c-1)}(h^*_{(c-1)j}(\theta_{nt})) - U_{(n-1)(c-1)}(h^*_{(c-1)j}(\theta_{nt_0}))]\\ &- \sum_{c=2}^{m}\binom{m}{c}(c-1)[U^{(-j)}_{(n-1)c}(h^*_c(\theta_{nt})) - U^{(-j)}_{(n-1)c}(h^*_c(\theta_{nt_0}))].\end{aligned} \tag{7.1.6}$$

By the definition of U-statistics, (A1) and using the mean value theorem, there is t^* satifying $\|t^*\| \le C$ such that

$$\begin{aligned}\|h(\tilde{Z}_1, ..., \tilde{Z}_m; \theta_{nt}) - h(\tilde{Z}_1, ..., \tilde{Z}_m; \theta_{nt_0})\| &\le n^{-1/2}2C\|\dot{h}(\tilde{Z}_1, ..., \tilde{Z}_m; \theta_{nt^*})\|\\ &\le n^{-1/2}2CG(\tilde{Z}_1, ..., \tilde{Z}_m).\end{aligned}$$

Therefore the difference in the first sum on the right side of (7.1.6) satisfies

$$\begin{aligned}&\|U_{(n-1)(c-1)}(h^*_{(c-1)j}(\theta_{nt})) - U_{(n-1)(c-1)}(h^*_{(c-1)j}(\theta_{nt_0}))\|\\ =&\|\binom{n-1}{c-1}^{-1}\sum_{i_1<...<i_{c-1}}(h^*(Z_j; Z_{i_1}, ..., Z_{i_{c-1}}; \theta_{nt}) - h^*(Z_j; Z_{i_1}, ..., Z_{i_{c-1}}; \theta_{nt_0}))\|\\ \le&\binom{n-1}{c-1}^{-1}\sum_{i_1<...<i_{c-1}}(\delta_{Z_j} + P)(\delta_{Z_{i_1}} + P)\dots(\delta_{Z_{i_{c-1}}} + P)P^{m-c}\\ &\|h(\tilde{Z}_1, ..., \tilde{Z}_m; \theta_{nt}) - h(\tilde{Z}_1, ..., \tilde{Z}_m; \theta_{nt_0})\|\\ \le&n^{-1/2}2C\binom{n-1}{c-1}^{-1}\sum_{i_1<...<i_{c-1}}(\delta_{Z_j} + P)(\delta_{Z_{i_1}} + P)\dots(\delta_{Z_{i_{c-1}}} + P)P^{m-c}G(Z_1, ..., Z_m),\end{aligned}$$

where $\sum_{i_1<...<i_{c-1}}$ denotes the sum over all permutations of $1, ..., j-1, j+1, ..., n$. Similar inequalities can be derived for the difference in the second sum on the right side of (7.1.6). Let

$$\begin{aligned}g^*(Z_{i_1}, ..., Z_{i_c}) &= (\delta_{Z_{i_1}} + P)\dots(\delta_{Z_{i_c}} + P)P^{m-c}G,\\ g^*_j(Z_{i_1}, ..., Z_{i_c}) &= g^*(Z_j; Z_{i_1}, ..., Z_{i_{c-1}}).\end{aligned} \tag{7.1.7}$$

Note that g^* is argument-symmetric. Combining the inequalities above yields

$$\frac{n^{1/2}}{2C}\|r_{nj}(t)\| \le \sum_{c=2}^{m}\binom{m}{c}cU_{(n-1)(c-1)}(g^*_j) + \sum_{c=2}^{m}\binom{m}{c}(c-1)U^{(-j)}_{(n-1)c}(g^*), \tag{7.1.8}$$

uniformly in $\|t\| \le C$. Recall

$$B_n(t) = \frac{1}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt})\|^2,$$

and after the partition,

$$\sup_{\|t\|\le C} B_n(t) \le B_n(t_0) + \sup_{\|t\|\le C} |B_n(t) - B_n(t_0)|;$$

also, recall the definition of P_{2n} as

$$P_{2n} = P(\sup_{\|t\|\le C} |B_n(t) - B_n(t_0)| > \epsilon/2).$$

By the Cauchy inequality, we derive

$$\begin{aligned}
P_{2n} =& P\Big(\sup_{\|t\|\le C} \frac{1}{n}\sum_{j=1}^{n} |\|R_{nj}(\theta_{nt})\|^2 - \|R_{nj}(\theta_{nt_0})\||^2 > \epsilon/2\Big) \\
\le& P\Big(\sup_{\|t\|\le C} \frac{1}{n}\sum_{j=1}^{n} \|R_{nj}(\theta_{nt}) - R_{nj}(\theta_{nt_0})\|(\|R_{nj}(\theta_{nt})\| + \|R_{nj}(\theta_{nt_0})\|) > \epsilon/2\Big) \\
\le& P\Big(\sup_{\|t\|\le C} \frac{1}{n}\sum_{j=1}^{n} \|r_{nj}(t)\|(\|r_{nj}(t)\| + 2\|R_{nj}(\theta_{nt_0})\|) > \epsilon/2\Big) \\
\le& P\Big(\sup_{\|t\|\le C} \frac{1}{n}\sum_{j=1}^{n} \big(\|r_{nj}(t)\|^2 + 2\|r_{nj}(t)\|\|R_{nj}(\theta_{nt_0})\|\big) > \epsilon/2\Big). \\
\le& P\Big(\sup_{\|t\|\le C} \frac{1}{n}\sum_{j=1}^{n} \|r_{nj}(t)\|^2 > \epsilon/4\Big) \\
&+ P\Big(\sup_{\|t\|\le C} \big(\frac{1}{n}\sum_{j=1}^{n} \|r_{nj}(t)\|^2\big)\big(\frac{1}{n}\sum_{j=1}^{n} \|R_{nj}(\theta_{nt_0})\|^2\big) > \epsilon^2/64\Big).
\end{aligned}$$

$$\frac{1}{n}\sum_{j=1}^{n} \|R_{nj}(\theta_{nt_0})\|^2 = O_P(1). \tag{7.1.9}$$

For arbitrary fixed $M > 0$, when n is large, there is $\epsilon' > 0$ such that

$$P(\frac{1}{n}\sum_{j=1}^{n} \|R_{nj}(\theta_{nt_0})\|^2 > M) = \epsilon',$$

and

$$P(\frac{1}{n}\sum_{j=1}^{n} \|R_{nj}(\theta_{nt_0})\|^2 \le M) = 1 - \epsilon'.$$

Let $C_n(t) = \frac{1}{n}\sum_{j=1}^{n} \|r_{nj}(t)\|^2$, we have

$$
\begin{aligned}
&P\Big(\sup_{\|t\|\le C} C_n(t)\big(\frac{1}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt_0})\|^2\big) > \epsilon^2/64\Big)\\
=&P\Big(\{\sup_{\|t\|\le C} C_n(t)\big(\frac{1}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt_0})\|^2\big) > \epsilon^2/64\}\bigcap\{\frac{1}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt_0})\|^2 > M\}\Big)\\
&+P\Big(\{\sup_{\|t\|\le C} C_n(t)\big(\frac{1}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt_0})\|^2\big) > \epsilon^2/64\}\bigcap\{\frac{1}{n}\sum_{j=1}^{n}\|R_{nj}(\theta_{nt_0})\|^2 \le M\}\Big)\\
\le&\epsilon' + P\Big(\sup_{\|t\|\le C} C_n(t) > \frac{\epsilon^2}{64M}\Big).
\end{aligned}
$$

when $\epsilon' \to 0$. Using this and (7.1.8), we derive for $0 < \epsilon < 1$ and M that

$$
\begin{aligned}
P_{2n} \le& P\Big(\sup_{\|t\|\le C} C_n(t) > \frac{\epsilon}{4}\Big) + P\Big(\sup_{\|t\|\le C} C_n(t) > \frac{\epsilon^2}{64M}\Big) + \epsilon'\\
\le& 2P\Big(\sup_{\|t\|\le C} C_n(t) > \frac{\epsilon}{4}\Big) + \epsilon'\\
\le& 2P\Big(\frac{1}{n}\sum_{j=1}^{n}\Big\{\sum_{c=2}^{m}\binom{m}{c} cU_{(n-1)(c-1)}(g_j^*)\Big\}^2 > \frac{\epsilon}{8}4C^2 n\Big) + \epsilon'\\
&+2P\Big(\frac{1}{n}\sum_{j=1}^{n}\Big\{\sum_{c=2}^{m}\binom{m}{c}(c-1)U^{(-j)}_{(n-1)c}(g^*)\Big\}^2 > \frac{\epsilon}{8}4C^2 n\Big) + \epsilon'\\
\le& \frac{2}{\frac{\epsilon}{8}n4C^2}E\left(\sum_{c=2}^{m}\binom{m}{c} cU_{(n-1)(c-1)}(g_1^*)\right)^2\\
&+\frac{2}{\frac{\epsilon}{8}n4C^2}E\left(\sum_{c=2}^{m}\binom{m}{c}(c-1)U^{(-1)}_{(n-1)c}(g^*)\right)^2 + \epsilon'\\
\le& \frac{4}{\epsilon nC^2}\sum_{c=2}^{m}\binom{m}{c}^2 c^2 Var(U_{(n-1)(c-1)}(g_1^*))\\
&+\frac{4}{\epsilon nC^2}\sum_{c=2}^{m}\binom{m}{c}^2 (c-1)^2 Var(U^{(-1)}_{(n-1)c}(g^*)) + \epsilon'\\
=& O(\frac{1}{n}) + \epsilon'.
\end{aligned}
$$

Take $M \to \infty$, then $\epsilon' \to 0$, hence this completes the proof. ■

Proof of Lemma 3.2.2. Let $\Delta_n(t)$ be the expression inside the supremum in (3.2.2). We now bound it by $\|\Delta_n(t)\| \le \|\Delta_{1n}(t)\| + \|\Delta_{2n}(t)\|$, where

$$\Delta_{1n}(t) = n^{-1/2}\sum_{j=1}^{n}(V_{nj}(\theta_{nt}) - m\tilde{h}_1(Z_j;\theta_{nt})),$$
$$\Delta_{2n}(t) = n^{-1/2}\sum_{j=1}^{n}(m\tilde{h}_1(Z_j;\theta_{nt}) - m\tilde{h}_1(Z_j;\theta_0)) + At.$$

By the Mean Value Theorem, for each $j = 1, ..., n$, there is some θ^*_{jt} lying between θ_0 and θ_{nt}, such that

$$\begin{aligned}
\Delta_{2n}(t) &= n^{-1/2}\sum_{j=1}^{n}(m(\tilde{h}_1(Z_j;\theta_{nt}) - \tilde{h}_1(Z_j;\theta_0))) + At \\
&= n^{-1/2}\sum_{j=1}^{n} mE[(h(\tilde{Z}_1, ..., \tilde{Z}_m;\theta_{nt}) - h(\tilde{Z}_1, ..., \tilde{Z}_m;\theta_0))|\tilde{Z}_1 = Z_j] + At \\
&= n^{-1/2}\sum_{j=1}^{n} mE[\frac{\partial h}{\partial\theta}(\tilde{Z}_1, ..., \tilde{Z}_m;\theta^*_{jt})(n^{-1/2}t)|\tilde{Z}_1 = Z_j] + At \\
&= \frac{t}{n}\sum_{j=1}^{n} mE[\frac{\partial h}{\partial\theta}(\tilde{Z}_1, ..., \tilde{Z}_m;\theta^*_{jt})|\tilde{Z}_1 = Z_j] + At \\
&= \frac{t}{n}\sum_{j=1}^{n} \frac{\partial}{\partial\theta}\Big(mE(h(\tilde{Z}_1, ..., \tilde{Z}_m;\theta^*_{jt})|\tilde{Z}_1 = Z_j)\Big) + At \\
&= \frac{t}{n}\sum_{j=1}^{n} \frac{\partial}{\partial\theta}(m\tilde{h}_1(Z_j;\theta^*_{jt})) + At \\
&= \frac{t}{n}\sum_{j=1}^{n} \Big(m\dot{\tilde{h}}_1(Z_j;\theta^*_{jt}) - m\dot{\tilde{h}}_1(Z_j;\theta_0)\Big) + \Big(\frac{1}{n}\sum_{j=1}^{n} m\dot{\tilde{h}}_1(Z_j;\theta_0) + A\Big)t \\
&:= A_n(t) + B_n(t),
\end{aligned} \tag{7.1.10}$$

where the equality from the fifth step to the sixth step in (7.1.10) is obtained by the Dominated Convergence Theorem, see Remark 7.1.1 below. For the second term of the last line, we apply the law of large numbers to get $\sup_{\|t\|\le C}\|B_n(t)\| = o_P(1)$. Let $G_1(z) = E(G(Z_1,...,Z_m)|Z_1 = z)$. We now apply (A1) to bound the first term by

$$\begin{aligned} A_n(t) \le & m\frac{\|t\|}{n}\sum_{j=1}^{n} G_1(Z_j)\|\theta^*_{jt} - \theta_0\| \\ \le & m\frac{C}{n}\sum_{j=1}^{n} G_1(Z_j)n^{-1/2}C \\ = & mC^2 n^{-1/2}(\frac{1}{n}\sum_{j=1}^{n} G_1(Z_j)) \\ = & O_P(n^{-1/2}) \end{aligned}$$

uniformly in $\|t\| \le C$. Combining these two, we obtain $\sup_{\|t\|\le C}\|\Delta_{2n}(t)\| = o_P(1)$.

To deal with Δ_{1n}, we now introduce $\hat{\mathbf{U}}_n(\theta)$, the projection of a vector U-statistic $\mathbf{U}_n(\theta)$ of order m with kernel $\mathbf{h}$ onto some sum space, which enables a U-statistic to be approximated within a sufficient degree of accuracy by a sum of i.i.d. random variables (for the details see Section 5.3.1, Serfling (1980)). Specifically, the projection $\hat{\mathbf{U}}_n(\theta)$ of a U-statistic $\mathbf{U}_n(\theta)$ is defined as

$$\hat{\mathbf{U}}_n(\theta) = \sum_{j=1}^{n} E(\mathbf{U}_n(\theta)|Z_j) - (n-1)E(\mathbf{U}_n(\theta)). \tag{7.1.11}$$

This is a sum of i.i.d. random variables, and satisfies

$$\hat{\mathbf{U}}_n(\theta) - E(\mathbf{U}_n(\theta)) = \frac{m}{n}\sum_{j=1}^{n}\tilde{\mathbf{h}}_1(Z_j;\theta), \tag{7.1.12}$$

where $\tilde{\mathbf{h}}_1$ is defined as before. The proof of this can be found in Remark 7.1.2. It is useful to express the difference $\mathbf{U}_n - \hat{\mathbf{U}}_n$ as a U-statistic,

$$\mathbf{U}_n(\theta) - \hat{\mathbf{U}}_n(\theta) = \binom{n}{m}^{-1}\sum_{1\le i_1<\ldots<i_m\le n}\mathbf{H}(Z_{i_1},\ldots,Z_{i_m};\theta), \tag{7.1.13}$$

based on the symmetric kernel

$$\mathbf{H}(z_1,\ldots,z_m;\theta) = \mathbf{h}(z_1,\ldots,z_m;\theta) - \tilde{\mathbf{h}}_1(z_1;\theta) - \ldots - \tilde{\mathbf{h}}_1(z_m;\theta). \tag{7.1.14}$$

Assume $E\|\mathbf{h}_\theta\|^2 < \infty$ uniformly in $\theta \in N(\theta_0)$. Then it is shown in Remark 7.1.3 that uniformly in $\theta \in N(\theta_0)$,

$$E\|\mathbf{U}_n(\theta) - \hat{\mathbf{U}}_n(\theta)\|^2 = O(n^{-2}). \tag{7.1.15}$$

We now express Δ_{1n} as

$$n^{-1/2}\Delta_{1n}(t) = \frac{1}{n}\sum_{j=1}^{n} V_{nj}(\theta_{nt}) - \frac{1}{n}\sum_{j=1}^{n} m\tilde{h}_1(Z_j;\theta_{nt}) \\ = U_n(\theta_{nt}) - \frac{m}{n}\sum_{j=1}^{n}\tilde{h}_1(Z_j;\theta_{nt}). \tag{7.1.16}$$

Using the projection $\hat{\mathbf{U}}_n(\theta)$ in (7.1.12) and noting (2.1.1), we further rewrite (7.1.16) as

$$n^{-1/2}\Delta_{1n}(t) = U_n(\theta_{nt}) - \hat{U}_n(\theta_{nt}). \tag{7.1.17}$$

Thus it is left to show

$$\sup_{\|t\|\le C} \|U_n(\theta_{nt}) - \hat{U}_n(\theta_{nt})\| = o_P(n^{-1/2}). \tag{7.1.18}$$

where C is an arbitrary positive constant.

Denote $U_{nt} = U_n(\theta_{nt}) - \hat{U}_n(\theta_{nt})$. It can then be expressed as a U-statistic,

$$U_{nt} = \binom{n}{m}^{-1} \sum_{1\le i_1<\ldots<i_m\le n} H(Z_{i_1},\ldots,Z_{i_m};\theta_{nt}), \tag{7.1.19}$$

where $H = \mathbf{H}$ is given in (7.1.14) with $\mathbf{h} = h$. To prove (7.1.18), it suffices to show $P(\sup_{\|t\|\le C}\|U_{nt}\| > n^{-1/2}\epsilon) \to 0$ as $n\to\infty$. Using the same technique as in the proof of Lemma 3.2.1, without loss of generality, we prove the case of two dimensional $t\in\mathbb{R}^2$ and denote $t = (t_1,t_2)^\top$. Equally partition $[-C, C]$ as $-C = C_0 < C_1 < C_2 < \ldots < C_L = C$ and obtain L^2 rectangles as

$$a_{ll'} : \{t = (t_1,t_2) : C_{l-1} < t_1 \le C_l,\, C'_{l'-1} < t_2 \le C'_{l'}\}, \quad 1\le l,l'\le L.$$

We take $t_{ll'} = (C_l, C_{l'})$ from each rectangle to get

$$\begin{aligned} &P\big(\sup_{\|t\|\le C}\|U_{nt}\| > n^{-1/2}\epsilon\big) \\ &\le \sum_{l,l'=1}^{L} P\big(\|U_{nt_{ll'}}\| > n^{-1/2}\frac{\epsilon}{2}\big) + \sum_{l,l'=1}^{L} P\big(\sup_{t\in a_{ll'}}\|U_{nt} - U_{nt_{ll'}}\| > n^{-1/2}\frac{\epsilon}{2}\big) \\ &:= \Omega_{1n} + \Omega_{2n}. \end{aligned}$$

Applying (7.1.15) with $\mathbf{U} = U$, we get

$$E\|U_n(\theta_{nt}) - \hat{U}_n(\theta_{nt})\| \leq \left(E\|U_n(\theta_{nt}) - \hat{U}_n(\theta_{nt})\|^2\right)^{\frac{1}{2}} = O(\frac{1}{n})$$

holds uniformly in $t = t_{ll'}, l, l' = 1, \ldots, L$. Consequently, there exists $M > 0$ such that

$$\Omega_{1n} \leq \frac{\sum_{l,l'=1}^{L} E\|U_{nt_{ll'}}\|n^{1/2}}{\epsilon/2} \leq \frac{L^2 M n^{-1} n^{1/2}}{\epsilon/2} \to 0$$

for $L = L_n = \log n$. This shows $\Omega_{1n} = o_P(1)$.

Finally, we are now left to show $\Omega_{2n} = o_P(1)$. To ease notation, set $\mathbb{Z}_c = (Z_{i_1}, \ldots, Z_{i_m})$ and write $\sum_c h(\mathbb{Z}_c)$ the sum of all the permutations $c = (i_1, ..., i_m)$ with $1 \leq i_1 < i_2 < ... < i_m \leq n$. With the aid of (7.1.14) and by (A1), we now use the mean value theorem to get

$$\begin{aligned}
&\|U_{nt} - U_{nt_{ll'}}\| \\
=&\|\binom{n}{m}^{-1} \sum_c \left(H(\mathbb{Z}_c; \theta_{nt}) - H(\mathbb{Z}_c; \theta_{nt_{ll'}})\right) \| \\
\leq&\binom{n}{m}^{-1} \sum_c \|(h(\mathbb{Z}_c; \theta_{nt}) - h(\mathbb{Z}_c; \theta_{nt_{ll'}})) - \sum_{q=1}^{m}(\tilde{h}_1(Z_{i_q}; \theta_{nt}) - \tilde{h}_1(Z_{i_q}; \theta_{nt_{ll'}}))\| \\
\leq&\binom{n}{m}^{-1} \sum_c \left(\|h(\mathbb{Z}_c; \theta_{nt}) - h(\mathbb{Z}_c; \theta_{nt_{ll'}})\| + \sum_{q=1}^{m} \|\tilde{h}_1(Z_{i_q}; \theta_{nt}) - \tilde{h}_1(Z_{i_q}; \theta_{nt_{ll'}})\|\right) \\
=&\binom{n}{m}^{-1} \sum_c \left(\|h(\mathbb{Z}_c; \theta_{nt}) - h(\mathbb{Z}_c; \theta_{nt_{ll'}})\| + \sum_{q=1}^{m} \|E(h(\mathbb{Z}_c; \theta_{nt}) - h(\mathbb{Z}_c; \theta_{nt_{ll'}})|Z_{i_q})\|\right) \\
\leq&\binom{n}{m}^{-1} \sum_c \left(\|\frac{\partial h}{\partial \theta}(\mathbb{Z}_c; \theta_{nt^*_{ll'}})\| \frac{1}{\sqrt{n}} \frac{2\sqrt{2}C}{L} + \sum_{q=1}^{m} \|E\left(\frac{\partial h}{\partial \theta}(\mathbb{Z}_c; \theta_{nt^*_{ll'_q}})|Z_{i_q}\right)\| \frac{1}{\sqrt{n}} \frac{2\sqrt{2}C}{L}\right) \\
\leq&\binom{n}{m}^{-1} \sum_c \left(\|\frac{\partial h}{\partial \theta}(\mathbb{Z}_c; \theta_{nt^*_{ll'}})\| \frac{1}{\sqrt{n}} \frac{2\sqrt{2}C}{L} + \sum_{q=1}^{m} E\left(\|\frac{\partial h}{\partial \theta}(\mathbb{Z}_c; \theta_{nt^*_{ll'_q}})\||Z_{i_q}\right) \frac{1}{\sqrt{n}} \frac{2\sqrt{2}C}{L}\right) \\
\leq&\frac{2\sqrt{2}C}{L\sqrt{n}} \binom{n}{m}^{-1} \sum_c \left(G(\mathbb{Z}_c) + \sum_{q=1}^{m} E(G(\mathbb{Z}_c)|Z_{i_q})\right),
\end{aligned}$$

where $t^*_{ll'}, t^*_{ll'_q} \in a_{ll'}$. Let $\kappa(\mathbb{Z}_c) = G(\mathbb{Z}_c) + \sum_{q=1}^m E(G(\mathbb{Z}_c)|Z_{i_q})$. The average in the last line is actually a U-statistic of order m with kernel κ. Since G is square-integrable, it follows $E(\kappa(\mathbb{Z}_c)) < \infty$. Hence

$$\begin{aligned}
\Omega_{2n} &\leq L^2 P\big(\|U_{nm}(\kappa - E\kappa)\| + E\kappa > \frac{\epsilon}{2}\frac{L}{2\sqrt{2}C}\big) \\
&\leq L^2 P\big(\|U_{nm}(\tilde{\kappa})\| > \frac{\epsilon}{4}\frac{L}{2\sqrt{2}C}\big) + L^2\mathbf{1}\big[E\kappa > \frac{\epsilon}{4}\frac{L}{2\sqrt{2}C}\big] \\
&\leq L^2 \frac{\mathrm{Var}(U_{nm}(\tilde{\kappa}))}{\frac{\epsilon^2}{16}\frac{L^2}{8C^2}} + L^2\mathbf{1}\big[E\kappa > \frac{\epsilon}{4}\frac{L}{2\sqrt{2}C}\big] \\
&\leq O(1/n) + L^2\mathbf{1}\big[E\kappa > \frac{\epsilon}{4}\frac{L}{2\sqrt{2}C}\big] \\
&= o_P(1)
\end{aligned}$$

as $L = L_n = \log n \to \infty$ while $Eg < \infty$. This completes the proof. ■

Remark 7.1.1 Assume (A1) is met. Let $\{\theta_n\} \in N(\theta_0)$ be a sequence such that $\theta_n \to \theta_0$ as $n \to \infty$. Since h_θ is differentiable with respect to θ, $\lim_{n\to\infty} \frac{h_{\theta_n} - h_{\theta_0}}{\theta_n - \theta_0} = \dot{h}_{\theta_0}$ holds. Under (A1), $\|\frac{\partial}{\partial\theta}h_\theta\|$ can be bound by a square-integrable function G for $\theta \in N(\theta_0)$. Thus it follows from the Dominated Convergence Theorem that $\frac{\partial}{\partial\theta}Eh_\theta = E(\frac{\partial}{\partial\theta}h_\theta)$.

Remark 7.1.2 The proof of (7.1.12) can be obtained by expanding $\hat{\mathbf{U}}_n(\theta)$ and applying the definition of $\mathbf{h}_1(Z_j;\theta)$.

$$\begin{aligned}
\hat{\mathbf{U}}_n(\theta) - E(\mathbf{U}_n(\theta)) &= \sum_{j=1}^{n} E(\mathbf{U}_n(\theta)|Z_j) - nE\mathbf{h} \\
&= \sum_{j=1}^{n} \binom{n}{m}^{-1} \sum_{1\le i_1<\ldots<i_m\le n} E(\mathbf{h}(Z_{i_1},\ldots,Z_{i_m};\theta)|Z_j) - nE\mathbf{h} \\
&= \sum_{j=1}^{n} \binom{n}{m}^{-1} \Big(\binom{n-1}{m-1}\mathbf{h}_1(Z_j;\theta) + \binom{n-1}{m}E\mathbf{h}\Big) - nE\mathbf{h} \\
&= \sum_{j=1}^{n} \Big(\frac{m}{n}\mathbf{h}_1(Z_j;\theta) + (1-\frac{m}{n})E\mathbf{h}\Big) - nE\mathbf{h} \\
&= \sum_{j=1}^{n} \Big(\frac{m}{n}\mathbf{h}_1(Z_j;\theta) + (1-\frac{m}{n})E\mathbf{h} - E\mathbf{h}\Big) \\
&= \frac{m}{n}\sum_{j=1}^{n}(\mathbf{h}_1(Z_j;\theta) - E\mathbf{h}) \\
&= \frac{m}{n}\sum_{j=1}^{n}\tilde{\mathbf{h}}_1(Z_j;\theta).
\end{aligned}$$

Remark 7.1.3 The proof of (7.1.15) is similar to the proof of Lemma 5.2.2B, Serfling (1980).

Proof Let us prove that (7.1.15) holds for the case of $d = 1$ for ease of notation. For $d > 1$, it can be obtained by stacking the coordinates. Define $\zeta_0 = 0$ and, for $1 \le c \le m$,

$$\zeta_c(\theta) = \mathrm{Var}[\mathbf{h}_c(Z_1,\ldots,Z_c;\theta)] = E[\tilde{\mathbf{h}}_c^2(Z_1,\ldots,Z_c;\theta)].$$

We have

$$0 = \zeta_0 \le \zeta_1(\theta) \le \cdots \le \zeta_m(\theta) = \mathrm{Var}(h_\theta) < \infty, \quad \theta \in N(\theta_0).$$

Consider two sets $\{a_1,\ldots,a_m\}$ and $\{b_1,\ldots,b_m\}$ of m distinct integers from $\{1,\ldots,n\}$ and let c be the number of integers common to the two sets. It follows from symmetry of $\tilde{\mathbf{h}}$ and independence of $\{Z_1,\ldots,Z_m\}$ that

$$E[\tilde{\mathbf{h}}(Z_{a_1},\ldots,Z_{a_m};\theta)\tilde{\mathbf{h}}(Z_{b_1},\ldots,Z_{b_m};\theta)] = \zeta_c(\theta).$$

Recall the difference (7.1.13) of $\mathbf{U}_n$ and its projection $\hat{\mathbf{U}}_n$ with the kernel $\mathbf{H}$ given in (7.1.14). Clearly,

$$\tilde{\mathbf{H}}_1(Z_1;\theta) = \tilde{\mathbf{h}}_1(Z_1;\theta) - \tilde{\mathbf{h}}_1(Z_1;\theta) = 0.$$

Hence,

$$\zeta_0 = \zeta_1(\theta) = 0. \tag{7.1.20}$$

Write

$$E(\mathbf{U}_n(\theta) - \hat{\mathbf{U}}_n(\theta))^2 = \binom{n}{m}^{-2} \sum E \prod_{j=1}^{2} \mathbf{H}(Z_{i_{j1}}, \ldots, Z_{i_{jm}};\theta), \tag{7.1.21}$$

where $\{i_{j1}, \ldots, i_{jm}\}$, $j = 1, 2$ are two sets of permutations among $\{1, \ldots, n\}$, and $\sum$ denotes summation over all $\binom{n}{m}^2$ of indicated terms. Consider a typical term of the product. For the jth factor, let p_j denote the number of indices repeated in the other factor. Since (7.1.20), if $p_j \leq 1$, then the product has zero expectation. Thus a term in (7.1.21) can have nonzero expectation only if each factor in the product contains at least 2 indices which appear in the other factor in the product. Note that each nonzero expectation is bounded by $E(G^{*2})$ which is independent of the parameter θ, where G^* is the square-integrability function given in Remark 3.1.1. Let q denote the number of distinct elements among the repeated indices in the two factors of a given product. Then

$$2q \leq \sum_{j=1}^{2} p_j. \tag{7.1.22}$$

For fixed values of q, p_1 and p_2, the number of ways to select the indices in the two factors of a product is of order

$$O(n^{q+(m-p_1)+(m-p_2)}), \tag{7.1.23}$$

where the implicit constants depend upon m, but not upon n. Moreover, by (7.1.22),

$$q \leq \lceil \frac{1}{2} \sum_{j=1}^{2} p_j \rceil,$$

where $\lceil \cdot \rceil$ denotes integer part. Thus

$$q + \sum_{j=1}^{2}(m - p_j) \leq 2m + \lceil \frac{1}{2} \sum_{j=1}^{2} p_j \rceil - \sum_{j=1}^{2} p_j = 2m - \lceil \frac{1}{2}(\sum_{j=1}^{2} p_j + 1) \rceil,$$

since, for any integer x, $x - \lceil \frac{1}{2}x \rceil = \lceil \frac{1}{2}(x+1) \rceil$. Note that $p_1, p_2 \geq 2$, we have $\sum_{j=1}^{2} p_j \geq 4$, so that

$$q + \sum_{j=1}^{2}(m - p_j) \leq 2m - \lceil \frac{1}{2}(4+1) \rceil = 2m - 2. \tag{7.1.24}$$

Thus, by (7.1.23) and (7.1.24), it follows that the number of terms in the sum in (7.1.21) for which the expectation is possibly nonzero is of order

$$O(n^{2m-2}).$$

The sum of such possibly nonzero terms is bounded by $O(n^{2m-2})E(G^2)$. Since $\binom{n}{m}^{-1} = O(n^{-m})$, it follows that (7.1.15) is proved. ■

Proof of Lemma 3.2.3. Let $\lambda_n(t)$ and $\Lambda_n(t)$ denote the smallest and largest eigen values of $\mathbb{S}_n(t)$. It follows from (B2) that there are constants $0 < \eta < K < \infty$ such that

$$P(\sup_{\|t\| \leq C} \Lambda_n(t) > K) \to 0 \quad \text{and} \quad P(\inf_{\|t\| \leq C} \lambda_n(t) > \eta) \to 0. \tag{7.1.25}$$

If follows from (B3) that

$$\sup_{\|t\| \leq C} \|\bar{\mathbb{T}}_n(t)\| = O_P(n^{-1/2}).$$

This and (B1) yield

$$\sup_{\|t\| \leq C} \mathbb{T}_n^*(t)\|\bar{\mathbb{T}}_n(t)\| = o_P(1) \tag{7.1.26}$$

and

$$\sup_{\|t\| \leq C} n\mathbb{T}_n^*(t)\|\bar{\mathbb{T}}_n(t)\|^3 = o_P(1). \tag{7.1.27}$$

From (3.2.3), (7.1.25) - (7.1.27) it follows that

$$\sup_{\|t\| \leq C} \| -2\log \mathscr{R}_n(t) - n\bar{T}_n(t)^\top \mathbb{S}_n(t)^{-1}\bar{T}_n(t)\| = o_P(1). \tag{7.1.28}$$

From (B2) we derive

$$\sup_{\|t\| \leq C} \|\mathbb{S}_n(t)^{-1} - S^{-1}\| = o_P(1)$$

and thus obtain the expansion

$$\sup_{\|t\|\leq C} \|n\bar{T}_n(t)^\top \mathbb{S}_n(t)^{-1}\bar{T}_n(t) - n\bar{T}_n(t)^\top S^{-1}\bar{T}_n(t)\| = o_P(1). \tag{7.1.29}$$

The first conclusion (3.2.4) in the lemma follows from (7.1.28), (7.1.29) and (B3). The second conclusion (3.2.5) is a simple consequence of (3.2.4). ■

7.2 Proofs for Lemmas and Theorems in Chapter 4

Proof of Lemma 4.1.1. Taking partial derivatives of equations (4.1.3) – (3.1.11) w.r.t. θ and ξ, we get

$$\frac{\partial A_{1n}}{\partial \theta^\top}(\theta, 0) = \frac{1}{n}\sum_{j=1}^{n} \frac{\partial m\tilde{h}_1}{\partial \theta^\top}(Z_j;\theta), \quad \frac{\partial A_{1n}}{\partial \xi^\top}(\theta, 0) = -\frac{1}{n}\sum_{j=1}^{n} m^2\tilde{h}_1(Z_j;\theta)^{\otimes 2},$$

$$\frac{\partial A_{2n}}{\partial \theta^\top}(\theta, 0) = 0, \qquad \frac{\partial A_{2n}}{\partial \xi^\top}(\theta, 0) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial m\tilde{h}_1}{\partial \theta}(Z_j;\theta)\right)^\top,$$

and

$$\frac{\partial B_{1n}}{\partial \theta^\top}(\theta, 0) = \frac{1}{n}\sum_{j=1}^{n} \frac{\partial V_{nj}}{\partial \theta^\top}(\theta), \quad \frac{\partial B_{1n}}{\partial \xi^\top}(\theta, 0) = -\frac{1}{n}\sum_{j=1}^{n} V_{nj}(\theta)^{\otimes 2},$$

$$\frac{\partial B_{2n}}{\partial \theta^\top}(\theta, 0) = 0, \qquad \frac{\partial B_{2n}}{\partial \xi^\top}(\theta, 0) = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial V_{nj}}{\partial \theta^\top}(\theta)\right)^\top.$$

Thus, expanding $A_{1n}(\theta,\xi)$ and $A_{2n}(\theta,\xi)$ at $(\theta_0, 0)$, we have

$$\begin{aligned}
&A_{1n}(\theta,\xi)\\
=&A_{1n}(\theta_0,0) + \frac{\partial A_{1n}}{\partial \theta^\top}(\theta_0,0)(\theta-\theta_0) + \frac{\partial A_{1n}}{\partial \xi^\top}(\theta_0,0)(\xi-0) + o_P(\delta_n)\\
=&\frac{1}{n}\sum_{j=1}^{n} m\tilde{h}_1(Z_j;\theta_0) + \frac{1}{n}\sum_{j=1}^{n}\frac{\partial m\tilde{h}_1}{\partial \theta^\top}(Z_j;\theta_0)(\theta-\theta_0)\\
&-\frac{1}{n}\sum_{j=1}^{n} m^2\tilde{h}_1(Z_j;\theta_0)^{\otimes 2}\xi + o_P(\delta_n),
\end{aligned} \tag{7.2.1}$$

and

$$\begin{aligned}
A_{2n}(\theta,\xi) =&A_{2n}(\theta_0,0) + \frac{\partial A_{2n}}{\partial \theta^\top}(\theta_0,0)(\theta-\theta_0) + \frac{\partial A_{2n}}{\partial \xi^\top}(\theta_0,0)(\xi-0) + o_P(\delta_n)\\
=&\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\partial m\tilde{h}_1}{\partial \theta^\top}(Z_j;\theta)\right)^\top \xi + o_P(\delta_n),
\end{aligned} \tag{7.2.2}$$

where $\delta_n = \|\theta - \theta_0\| + \|\xi\|$, $(\theta, \xi) \in N_0(\theta_0, 0)$. These expansions follow from the usual mean value theorem and the bounded assumptions of the kernel functions.

Similarly, expanding $B_{1n}(\theta, \xi)$ and $B_{2n}(\theta, \xi)$ at $(\theta_0, 0)$, we get

$$\begin{aligned} B_{1n}(\theta, \xi) = \frac{1}{n}\sum_{j=1}^{n} V_{nj}(\theta_0) + \frac{1}{n}\sum_{j=1}^{n} \frac{\partial V_{nj}}{\partial \theta^\top}(\theta_0)(\theta - \theta_0) \\ - \frac{1}{n}\sum_{j=1}^{n} V_{nj}(\theta_0)^{\otimes 2}\xi + o_P(\delta_n), \end{aligned} \tag{7.2.3}$$

and

$$B_{2n}(\theta, \xi) = \frac{1}{n}\sum_{j=1}^{n} \left(\frac{\partial V_{nj}}{\partial \theta^\top}(\theta)\right)^\top \xi + o_P(\delta_n). \tag{7.2.4}$$

Hence it follows from (7.2.1)-(7.2.4) that

$$\begin{aligned} &\sup_{(\theta,\xi)\in N_0(\theta_0,0)} \|B_{1n}(\theta, \xi) - A_{1n}(\theta, \xi)\| \\ \leq &\frac{1}{n}\sum_{j=1}^{n} \|R_{nj}(\theta_0)\| + \frac{1}{n}\sum_{j=1}^{n} \|\dot{R}_{nj}(\theta_0)\| |n^{-1/2}| \\ &+ \frac{1}{n}\sum_{j=1}^{n} \|V_{nj}(\theta_0)^{\otimes 2} - m^2 \tilde{h}_1(Z_j; \theta_0)^{\otimes 2}\| |n^{-1/2}| + o_P(\delta_n), \end{aligned}$$

and

$$\sup_{(\theta,\xi)\in N_0(\theta_0,0)} \|B_{2n}(\theta, \xi) - A_{2n}(\theta, \xi)\| \leq \frac{1}{n}\sum_{j=1}^{n} \|\dot{R}_{nj}(\theta_0)\| |n^{-1/2}| + o_P(\delta_n).$$

By (7.1.3),

$$\dot{R}_{nj}(\theta) = \sum_{c=2}^{m} \binom{m}{c} \left(c U_{(n-1)(c-1)}(\dot{h}^*_{(c-1)j}(\theta)) - (c-1) U^{(-j)}_{(n-1)c}(\dot{h}^*_c(\theta)) \right).$$

Hence

$$\|\dot{R}_{nj}(\theta_0)\| \leq \sum_{c=2}^{m} \binom{m}{c} \left(c U_{(n-1)(c-1)}(\|\dot{h}^*_{(c-1)j}(\theta_0)\|) + (c-1) U^{(-j)}_{(n-1)c}(\|\dot{h}^*_c(\theta_0)\|) \right).$$

Expanding the U-statistics in the above inequality and using (7.1.7), we obtain

$$\|\dot{R}_{nj}(\theta_0)\| \leq \sum_{c=2}^{m} \binom{m}{c} c U_{(n-1)(c-1)}(g^*_j) + \sum_{c=2}^{m} \binom{m}{c} (c-1) U^{(-j)}_{(n-1)c}(g^*), \quad j = 1, \ldots, n.$$

Thus, for $M > 0$,

$$
\begin{aligned}
&P(\frac{1}{n}\sum_{j=1}^{n}\|\dot{R}_{nj}(\theta_0)\| > M) \\
\leq &P(\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{c=2}^{m}\binom{m}{c}cU_{(n-1)(c-1)}(g_j^*)\right) > \frac{M}{2}) \\
&+ P(\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{c=2}^{m}\binom{m}{c}(c-1)U_{(n-1)c}^{(-j)}(g^*)\right) > \frac{M}{2}) \\
\leq &\frac{2}{M}E\left(\sum_{c=2}^{m}\binom{m}{c}cU_{(n-1)(c-1)}(g_1^*)\right) + \frac{2}{M}E\left(\sum_{c=2}^{m}\binom{m}{c}(c-1)U_{(n-1)c}^{(-1)}(g^*)\right) \\
= &\frac{2}{M}\sum_{c=2}^{m}\binom{m}{c}cE(U_{(n-1)(c-1)}(g_1^*)) + \frac{2}{M}\sum_{c=2}^{m}\binom{m}{c}(c-1)E(U_{(n-1)c}^{(-1)}(g^*)) \\
= &O_P(1)
\end{aligned}
$$

as $M \to \infty$. Therefore, we arrive at

$$
\frac{1}{n}\sum_{j=1}^{n}\|\dot{R}_{nj}(\theta_0)\| = o_P(1)
$$

Now we are left to show

$$
\frac{1}{n}\sum_{j=1}^{n}\|V_{nj}(\theta_0)^{\otimes 2} - m^2\tilde{h}_1(Z_j;\theta_0)^{\otimes 2}\| = o_P(n^{1/2}).
$$

In the proof of Lemma 3.2.1 (the proof about $B_n(t)$), we have shown

$$
E\|R_{nj}(\theta_0)\|^2 = O(\frac{1}{n}), \quad j = 1,\ldots,n.
$$

Moreover,

$$
\sum_{j=1}^{n}\|R_{nj}(\theta_0)\|^2 = \sum_{j=1}^{n}\|V_{nj}(\theta_0) - m\tilde{h}_1(Z_j;\theta_0)\|^2 = O_P(1), \tag{7.2.5}
$$

as the expected value of the above sum is $O(1)$. Thus by Markov's inequality, we derive for any $\epsilon > 0$,

$$
\begin{aligned}
P\Big(\max_{1\leq j\leq n}\|V_{nj}(\theta_0)\| > n^{1/2}\epsilon\Big) &\leq \sum_{j=1}^{n}P\left(\|V_{nj}(\theta_0)\| > n^{1/2}\epsilon\right) \\
&\leq \epsilon^{-2}E\left(\|V_{n1}(\theta_0)\|^2\mathbf{1}[\|V_{n1}\| > n^{1/2}\epsilon]\right) \to 0, \quad n \to \infty.
\end{aligned}
$$

Thus

$$\max_{1\le j\le n} \|V_{nj}(\theta_0)\| = o_P(n^{1/2}). \tag{7.2.6}$$

It follows that

$$\max_{1\le j\le n} \|m\tilde{h}_1(Z_j;\theta_0)\| = o_P(n^{1/2}). \tag{7.2.7}$$

Now let us write the components of $h(\cdot;\theta_0)$ as $(h^{(1)}(\cdot;\theta_0),\ldots,h^{(d)}(\cdot;\theta_0))$. By (7.2.5) and the Cauchy inequality, we have

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{j=1}^{n}\Big(V_{nj}(h^{(l)})V_{nj}(h^{(l')}) - m\tilde{h}_1^{(l)}(Z_j;\theta_0)m\tilde{h}_1^{(l')}(Z_j;\theta_0)\Big)\Big|^2\\
\le&2\frac{1}{n}\sum_{j=1}^{n}\Big(V_{nj}(h^{(l)}) - m\tilde{h}_1^{(l)}(Z_j;\theta_0)\Big)^2\frac{1}{n}\sum_{j=1}^{n}V_{nj}(h^{(l')})^2\\
&+2\frac{1}{n}\sum_{j=1}^{n}\big(m\tilde{h}_1^{(l)}(Z_j;\theta_0)\big)^2\frac{1}{n}\sum_{j=1}^{n}\Big(V_{nj}(h^{(l')}) - m\tilde{h}_1^{(l')}(Z_j;\theta_0)\Big)^2\\
=&O_p(n^{-1}), \quad l,l' = 1,\ldots,d.
\end{aligned} \tag{7.2.8}$$

And this completes the proof. ■

Proof of Theorem 4.1.1. Under (A1) – (A3), the JEL $\hat{\mathscr{R}}_n(\theta)$ in (3.1.8) is continuously differentiable and its maximizer must satisfy (3.1.9). By Lemma 4.1.1, the euclidean norm of the difference of the solutions of (4.1.2) and (3.1.9) tends to zero as $n\to\infty$. Consequently, we prove the first part of the theorem by applying Lemma 1 of Qin and Lawless (1994). We now apply their Theorem 1 to prove the remaining (4.1.7). By Lemma 4.1.1 and Taylor's expansion of B_{1n} and B_{2n}, we have

$$\begin{aligned}
0 =&B_{1n}(\hat\theta,\hat\xi)\\
=&B_{1n}(\theta_0,0) + \frac{\partial B_{1n}}{\partial\theta^\top}(\theta_0,0)(\hat\theta-\theta_0) + \frac{\partial B_{1n}}{\partial\xi^\top}(\theta_0,0)(\hat\xi-0) + o_P(\delta_n),\\
0 =&B_{2n}(\hat\theta,\hat\xi)\\
=&B_{2n}(\theta_0,0) + \frac{\partial B_{2n}}{\partial\theta^\top}(\theta_0,0)(\hat\theta-\theta_0) + \frac{\partial B_{2n}}{\partial\xi^\top}(\theta_0,0)(\hat\xi-0) + o_P(\delta_n),
\end{aligned}$$

where $\delta_n = \|\hat\theta-\theta_0\| + \|\hat\xi\|$. We have

$$\begin{pmatrix}\hat\xi\\ \hat\theta-\theta_0\end{pmatrix} = S_n^{-1}\begin{pmatrix}B_{1n}(\theta_0,0)+o_P(\delta_n)\\ o_P(\delta_n)\end{pmatrix},$$

where

$$S_n = \begin{pmatrix} \frac{\partial B_{1n}}{\partial \xi^\top} & \frac{\partial B_{1n}}{\partial \theta} \\ \frac{\partial B_{2n}}{\partial \xi^\top} & 0 \end{pmatrix}_{(\theta_0, 0)} = \begin{pmatrix} -\frac{1}{n}\sum_{j=1}^n V_{nj}^{\otimes 2} & \frac{1}{n}\sum_{j=1}^n \dot{V}_{nj} \\ \frac{1}{n}\sum_{j=1}^n \dot{V}_{nj}^\top & 0 \end{pmatrix}$$

Thus

$$S_n \to_p \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & 0 \end{pmatrix} = \begin{pmatrix} -E(m^2 \tilde{h}_1^{\otimes 2}) & E(m\dot{\tilde{h}}_1) \\ E(m\dot{\tilde{h}}_1)^\top & 0 \end{pmatrix}.$$

The convergence to S_{11} can be obtained by the law of large numbers to (7.2.8), and the convergence to S_{12} (S_{21}) from the proof of Lemma 4.1.1. From this and $B_{1n}(\theta_0, 0) = \frac{1}{n}\sum_{j=1}^n V_{nj}(\theta_0) = O_P(n^{-1/2})$, we derive $\delta_n = O_P(n^{-1/2})$. Consequently, we arrive at

$$\sqrt{n}(\hat{\theta} - \theta_0) = S_{22.1}^{-1} S_{21} S_{11}^{-1} \sqrt{n} B_{1n}(\theta_0, 0) + o_P(1) \to \mathscr{N}(0, V).$$

From this, Lemma 4.1.1 and Theorem 3.1.1 it follows the desired result. ■

REFERENCES

REFERENCES

[1] Arcones M.A. (1996). The Bahadur-Kiefer Representation for U-Quantiles. *Ann. Statist.* **24**: 1400 – 1422.

[2] Arcones, M.A. and Giné, E. (1993). Limit Theorems for U-Processes. *Ann. Statist.* **21**: 1494 – 1542.

[3] Arvesen, J.N. (1969). Jackknifing U-statistics. *Ann. Statist.* **40**: 2076 – 2100.

[4] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation in Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore.

[5] Chen J, Variyath A., Abraham B. (2008). Adjusted Empirical Likelihood and its Properties. *J. Comp. Graph. Stat.* **17(2)**, 426-443.

[6] Chen, S.X., Peng, L. and Qin, Y.-L. (2009). Effects on data dimension on empirical likelihood. *Biometrika* **96**: 711–722.

[7] Chen, S.X. and Qin, Y.S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* **87**, 946–953.

[8] Chen, S.X. and Van Keilegom, I. (2009). A goodness-of-fit test for parametric and semiparametric models in multiresponse regression. *Bernoulli* **15**: 955–976.

[9] DiCiccio, T.J., Hall, P. and Romano, J.P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**: 1053–1061.

[10] Emerson, S. and Owen, A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* **3**, 1161–1192.

[11] Fang, F., Hong, Q. and Shao, J. (2009). A pseudo empirical likelihood approach for stratified samples with nonresponse. *Ann. Statist.* **37**: 371–393.

[12] Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Review*, **58**: 109–127.

[13] Heffernan, P. M. (1997). Unbiased Estimation of Central Moments by using U-statistics. *J. R. Statist. Soc. B* **59**: 861 – 863.

[14] Helland, I. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9**, 79–94.

[15] Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**: 1079–1111.

[16] Jing, B.-Y., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *J. Amer. Statist. Assoc.* **104**: 1224–1232.

[17] Kolaczyk, E.D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4**, 199–218.

[18] Kowalski, J. and Tu, X.M. (2008). *Modern Applied U-Statistics*. John Wiley & Sons, Inc.

[19] Lee, A. J. (1990). *U-Statistics: Theory and Practice*. Marcel Dekker, Inc.

[20] Liu, Y. and Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *Ann. Statist.* **38**, 1341-1362.

[21] Nordman, D.J. and Lahiri, S.N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34**: 3019–3050.

[22] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**: 237–249.

[23] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18**: 90–120.

[24] Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725–1747.

[25] Owen, A. (2001). *Empirical Likelihood*. Chapman & Hall/CRC, London.

[26] Peng, H. (2015). Efficient estimation in models with side information: an easy empirical likelihood approach. The *Manuscript* can be found at the URL: `https://math.iupui.edu/~hpeng/preprints_hp.html`.

[27] Peng, H. and Schick, A. (2004). Estimation of linear functionals of bivariate distributions with parametric marginals. *Statist. Decisions* **22**, 61–77.

[28] Peng, H. and Schick, A. (2005). Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown marginals: the least-squares approach. *J. Multivariate Anal.* **95**, 385–409.

[29] Peng, H. and Schick, A. (2013a). Maximum empirical likelihood estimation and related topics. The *Manuscript* can be found at the URL: `http://www.math.binghamton.edu/anton/preprint.html`.

[30] Peng, H. and Schick, A. (2013b). Maximum empirical likelihood estimation and related topics. The *Manuscript* can be found at the URL: `https://math.iupui.edu/~hpeng/preprints_hp.html`.

[31] Peng, H. and Schick, A. (2013c). An empirical likelihood approach of goodness of fit testing. *Bernoulli* **19**: 954–981

[32] Peng, H. and Tan, F. (2016). Jackknife Empirical Likelihood Goodness-Of-Fit Tests For U-Statistics Based General Estimating Equations. To appear in *Bernoulli*.

[33] Portnoy, S. (1988). Behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16**, 356–366.

[34] QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**: 30–325.

[35] SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc.

[36] SHI, J. and LAU, T.S. (2000). Empirical likelihood for partially linear models. *J. Multivariate Anal.* **72**, 132–148.

[37] SHI, X. (1984). Approximate Independence of Jackknife Pseudo Values and the Bootstrap Method. *Journal of Wuhan Institute of Hydra-Electric Engineering* (Chinese) **2**, 83 – 90.

[38] TAN, F., LIN, Q., ZHENG, W. and PENG, H. (2013). A Jackknife empirical likelihood approach to goodness of fit U-Statistics testing with side information. The *Manuscript* can be found at the URL: `http://math.iupui.edu/~ftan.`

[39] TANG, C. Y. and LENG, C. (2012). An empirical likelihood approach to quantile regression with auxiliary information. *Statist. & Probabil. Lett.* **82**, 29–36.

[40] VAART, VAN DER A.W. (2000). *Asymptotic Statistics.* Cambridge University Press.

[41] WANG, Q.-H. and JING, B.-Y. (2003). Empirical likelihood for partially linear models. *Ann. Inst. Statist. Meth.* **55**: 585–595.

[42] YUAN, A., HE, W., WANG, B. and QIN, G. (2012). U-statistics with side information. *J. Multiv. Analy.* **111**: 20 – 38.

[43] ZHANG, B. (1995). M-estimation and quantile estimation in the presence of auxiliary information. *J. Statist. Plann. Infer.* **44**, 77 – 94.

[44] ZHANG, B. (1997). Quantile processes in the presence of auxiliary information. *Ann. Inst. Statist. Math.* **49**, 35 – 55.

[45] ZUO, Y. and SERFING, Y. (1997). General notions of statistical depth function. *Ann. Statist.* **28**, 461 – 482.

VITA

VITA

My name is Lingnan Li. I obtained my Bachelor's degree in Science from the University of Science and Technology of China in 2009. After that, I came to the United States and studied in the Ph.D. program in the Department of Mathematical Sciences at Indiana University Purdue University Indianapolis. Since 2010, I have been working on Statistics as a Ph.D. student.