Empirical Likelihood Incorporating the Existence Probability

Takuma Bando and Tomonari Sei
Graduate School of Information Science and Technology,
The University of Tokyo

Abstract

Empirical likelihood is an effective nonparametric method for statistical inference, having several good properties from both theoretical and practical aspects. One serious problem of empirical likelihood is that its confidence regions can have poor accuracy, especially when the sample size is small, the parameter is multidimensional, and/or the underlying distribution is skewed. Importantly, the fact that the "probability that the empirical likelihood ratio function exists" (hereinafter, referred to as "existence probability") is always lower than 1 is a strong contributor to such poor accuracy. In this paper, we develop a new approach to improvements in coverage accuracy of empirical likelihood inference by incorporating the existence probability into the inference. It is demonstrated in simulation studies how accurate our methods are; they may work well even in small sample, multidimensional, highly skewed situations.

1 Introduction

Likelihood inference is a powerful device in parametric statistics. Empirical likelihood introduced by [4] brings likelihood inference into the realm of nonparametric or semiparametric statistical inference, employing a nonparametric version of likelihood functions referred to as empirical likelihood ratio function. It is not necessary to specify a family of distributions of the data; nevertheless, the empirical likelihood holds several advantages of parametric likelihood inference. Among them, it is appealing that the empirical likelihood does not require the construction of a pivotal statistic, does not generally require variance estimation, produces confidence regions with the natural shape, and has very desirable asymptotic power. In the past three decades, the empirical likelihood has found applications in many areas of statistics and is still under active research.

Here, our interest is in the "practical" difficulties constructing accurate empirical likelihood confidence regions; even though Bartlett corrected empirical likelihood introduced by [1, 2] has the coverage error of $O(n^{-2})$ with sample size n, it still has poor accuracy when the sample size is small, the data is multidimensional, and/or the underlying distribution is highly skewed; see, for example, [6]. These difficulties are partially attributed to the *existence probability*, the probability that the empirical likelihood ratio function exists.

We develop two calibration methods for improvements in coverage accuracy of empirical likelihood confidence regions; these incorporate the information of the existence probability into empirical likelihood inference each in a different way but are based on a common idea, the *quasi-bootstrap*. The quasi-bootstrap is a statistical resampling method for estimating the existence probability, in which the data is resampled from what we call the quasi-bootstrap

distribution. We put a detailed discussion of our methods in Section 3, including their asymptotic properties. On top of that, in Section 5, simulation studies show that our methods have better performance than conventional counterparts.

We only discuss the mean as a parameter of interest in this paper; nevertheless it may be possible to apply our methods to other parameters.

2 Preliminaries

Here, we provide some notations and previous studies concerning empirical likelihood necessary for later sections.

2.1 Empirical likelihood

To set up notation, let $X_1, ..., X_n \in \mathbb{R}^d$ be i.i.d. random variables having mean μ_0 and variance-covariance matrix V_0. The empirical likelihood ratio function for the mean is defined by

$$R(\mu) = \sup\left\{\prod_{i=1}^{n} n\omega_i \mid \sum_{i=1}^{n} \omega_i(X_i - \mu) = 0, \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1\right\};$$

we also define the empirical log-likelihood ratio function $\ell(\mu) = -2\log R(\mu)$. Here we introduce two conditions:

Condition 2.1. V_0 is finite and has rank $q > 0$,

Condition 2.2. $\limsup_{\|t\|\to\infty} |E[\exp(\mathrm{it}^\top \mathrm{X_i})]| < 1$.

One noteworthy property of the empirical likelihood is that, under Condition 2.1,

$$\ell(\mu_0) = -2\log R(\mu_0) \rightsquigarrow \chi^2_{(q)}, \tag{1}$$

which provides an asymptotic justification for statistical hypothesis testing that reject μ_0 at some nominal level. Then, given the nominal level $1-\alpha$, we typically choose the threshold $\gamma_{1-\alpha}$ satisfying $P(\chi^2_{(q)} \leq \gamma_{1-\alpha}) = 1-\alpha$; the $100(1-\alpha)\%$ empirical likelihood confidence region for the mean may be written as

$$C^{n,\alpha} = \{\mu \mid \ell(\mu) \leq \gamma_{1-\alpha}\} = \left\{\sum_{i=1}^{n} \omega_i X_i \mid -2\log\prod_{i=1}^{n} n\omega_i \leq \gamma_{1-\alpha}, \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1\right\}.$$

This is the basic way to calibrate empirical likelihood, what is called *chi-squared calibration*. [1] showed that the coverage error of $C^{n,\alpha}$ is of order n^{-1} under Conditions 2.1 and 2.2:

$$P(\mu_0 \in C^{n,\alpha}) = P(\ell(\mu_0) \leq \gamma_{1-\alpha}) = P(\chi^2_{(q)} \leq \gamma_{1-\alpha}) + O(n^{-1}). \tag{2}$$

In the same time, they established another key property that, under Conditions 2.1 and 2.2, the empirical likelihood is Bartlett correctable, reducing the coverage error to $O(n^{-2})$: for a constant a chosen in an appropriate manner,

$$P\left(\ell(\mu_0)\left(1 - \frac{a}{n}\right) \leq \gamma_{1-\alpha}\right) = P(\chi^2_{(q)} \leq \gamma_{1-\alpha}) + O(n^{-2}). \tag{3}$$

Then the corresponding confidence region is defined by

$$C_{\mathrm{Bart}}^{n,\alpha} = \left\{ \mu \mid \ell(\mu) \left(1 - \frac{a}{n}\right) \leq \gamma_{1-\alpha} \right\}.$$

We refer to [1, 2] for the details.

Another effective calibration method is the *bootstrap calibration*. Define the bootstrap empirical likelihood ratio function for $\bar{X}_n$ by

$$R^*(\bar{X}_n) = \sup \left\{ \prod_{i=1}^{n} n\omega_i \mid \sum_{i=1}^{n} \omega_i (X_i^* - \bar{X}_n) = 0, \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1 \right\}, \tag{4}$$

where $X_1^*, ..., X_n^*$ are independently sampled from the empirical distribution of $X_1, ..., X_n$. Let $\hat{\zeta}_{n,\alpha}$ be the $1-\alpha$ quantile of $-2 \log R^*(\bar{X}_n)$. Then the confidence region of the bootstrap calibrated empirical likelihood is written as

$$C_{\mathrm{B}}^{n,\alpha} = \{ \mu \mid \ell(\mu) \leq \hat{\zeta}_{n,\alpha} \}. \tag{5}$$

Since it is almost impossible to obtain the exact values of $\hat{\zeta}_{n,\alpha}$ in practice, the simulation procedure is used for approximating them: for a sufficiently large B, calculating the quantiles of $-2 \log R^{b*}(\bar{X}_n),\ b = 1, ..., B$.

2.2 Existence Probability

By the definition of the empirical likelihood ratio function for the mean, It is obvious that $R(\mu_0)$ and $\ell(\mu_0)$ is defined if and only if

$$\mu_0 \in \{ \mu \mid \ell(\mu) < \infty \} = \{ \mu \mid \mu \in \mathrm{Conv}(X_1, ..., X_n) \},$$

where $\mathrm{Conv}(X_1, ..., X_n)$ is the convex hull of X_i; otherwise, $\ell(\mu_0) = \infty$. Then what we call the existence probability is written as

$$p = P(\ell(\mu) < \infty) = P(\mu_0 \in \mathrm{Conv}(X_1, ..., X_n)).$$

The important thing is that p itself may be in part responsible for the poor accuracy of empirical likelihood inference. We illustrate how p acts on the the chi-squared calibration and the bootstrap calibration.

As for the chi-squared calibration, $1-p$ may be regarded as negligible, for, in the proof of (1) (see [5]), it is assumed that sample size n is large enough that $1 - p = P(\mu_0 \notin \mathrm{Conv}(X_1, ..., X_n))$ is negligible and the mean μ_0 is inside the convex hull of X_i; (2) and (3) are more precisely expressed as

$$\begin{aligned} P(\ell(\mu_0) \leq \gamma_{1-\alpha} \mid \ell(\mu_0) < \infty) &= \frac{P(\ell(\mu_0) \leq \gamma_{1-\alpha})}{p} \\ &= 1 - \alpha + O(n^{-1}) \end{aligned}$$

and likewise

$$P\left(\ell(\mu_0)\left(1 - \frac{a}{n}\right) \leq \gamma_{1-\alpha} \mid \ell(\mu_0) < \infty\right) = 1 - \alpha + O(n^{-2}). \tag{6}$$

This implies that when p is small, the coverage probability of chi-squared calibrated empirical likelihood may tend to be lower than the nominal level $1-\alpha$:

$$P(\ell(\mu_0) \le \gamma_{1-\alpha}) = (1-\alpha)p + O(n^{-1}), \tag{7}$$

$$P\left(\ell(\mu_0)\left(1-\frac{a}{n}\right) \le \gamma_{1-\alpha}\right) = (1-\alpha)p + O(n^{-2}). \tag{8}$$

There is no doubt that both (2) and (3) are correct since, by taking $n \to \infty$, we have $p \to 1$ with higher order than $O(n^{-2})$ as we shall see in Section 6.

In the second place, the bootstrap calibrated empirical likelihood is subject to influence by the existence probability; it is possible that its coverage probabilities are much greater than their nominal levels in contrast with the chi-squared calibration in which the coverage probabilities are subject to underestimation. Let us see how such overestimation happens. For a sufficiently large B, let $\ell^{1*}(\bar{X}_n), ..., \ell^{B*}(\bar{X}_n)$ be bootstrap replicants of the empirical likelihood likelihood ratio function for the empirical mean $\bar{X}_n$, given $X_1, ..., X_n$ having mean μ_0. Let $\ell^1(\mu_0), ..., \ell^B(\mu_0)$ be empirical log-likelihood ratio functions for μ_0 obtained by a simulation procedure. Note that some of the bootstrap replicants and the simulated functions are infinite; $\ell^i(\mu_0) = \infty$ with the probability $1-p = P(\ell(\mu_0) = \infty)$ and $\ell^{i*}(\bar{X}_n) = \infty$ with $P(\ell^*(\bar{X}_n) = \infty \mid \mathbb{F}_n)$, in which the notation $P(\cdot \mid \mathbb{F}_n)$ indicates that the distribution of the observations is $\mathbb{F}_n$ given $X_1, ..., X_n$. What is matter here is that the number of $\ell^{i*}(\bar{X}_n)$ being infinite is prone to be larger than the counterpart of $\ell^i(\mu_0)$. In other words, $P(\ell^*(\bar{X}_n) = \infty \mid \mathbb{F}_n)$ overestimates $1-p$. There is an implication that the coverage probabilities of bootstrap calibrated empirical likelihoods may still be above their nominal levels.

What we want to highlight here is that it needs to incorporate the information of the existence probability into empirical likelihood inference for practical use.

3 Proposed Methods

In this section, we propose two calibration methods incorporating existence probabilities into empirical likelihood inference. They are based on a common idea, the quasi-bootstrap, which is a new resampling scheme designed to estimate the existence probability p well. There are two ways to apply the quasi-bootstrap to empirical likelihood inference, corresponding to our two methods. The *quasi-bootstrap calibration*, our first method, employs the quasi-bootstrap for resampling empirical likelihood ratio functions. On the other hand, the *modifying chi-squared calibration*, our second method, uses it for estimation of the existence probability.

3.1 Quasi-Bootstrap

The quasi-bootstrap is a resampling method, constructed so as to take the existence probability into account in empirical likelihood inference. It is basically much the same as the bootstrap introduced by [3]. The only difference between the quasi-bootstrap and the bootstrap is their distributions used for resampling. Let $X_1, ..., X_n \in \mathbb{R}^d$ be i.i.d. random variables with mean μ_0, let $\mathbb{F}_n$ be the empirical distribution obtained from $X_1, ..., X_n$, let $\varepsilon = d/n$, and let G be some *projection symmetric* distribution having mean $\bar{X}_n = \sum_{i=1}^n X_i/n$ and finite variance-covariance matrix. Here, the distribution of Y is projection symmetric if the distribution of the projection of Y on the unit sphere is symmetric with respect to its mean; for example, Gaussian distributions and uniform distributions on the surface of spheres. A

quasi-bootstrap sample $X_1^*, ..., X_n^*$ is drawn from

$$T_n = (1-\varepsilon)\mathbb{F}_n + \varepsilon G; \tag{9}$$

meanwhile the bootstrap relies on $\mathbb{F}_n$ only. Hence, its distribution function is written as

$$T_n(x) = (1-\varepsilon)\mathbb{F}_n(x) + \varepsilon G(x) = (1-\varepsilon)\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}} + \varepsilon G(x).$$

This formulation may seem strange and to reduce estimation accuracy. In fact, T_n is intended to estimate the existence probability well:

$$P(\ell^*(\bar{X}_n) < \infty \mid T_n) \simeq P(\ell(\mu_0) < \infty),$$

where

$$\ell^*(\bar{X}_n) = -2\log\left[\sup\left\{\prod_{i=1}^{n} n\omega_i \mid \sum_{i=1}^{n}\omega_i(X_i^* - \bar{X}_n) = 0,\ \omega_i \ge 0,\ \sum_{i=1}^{n}\omega_i = 1\right\}\right]. \tag{10}$$

The notation $P(\cdot \mid T_n)$ denotes that the observations are sampled from T_n given $X_1, ..., X_n$. Indeed, $P(\ell^*(\bar{X}_n) < \infty)$ may be good estimator of $P(\ell(\mu_0) < \infty)$, whereas the bootstrap tends to underestimate.

3.2 Quasi-Bootstrap Calibration

The quasi-bootstrap calibration is almost the same as the bootstrap calibration except for their distributions. Let $X_1, ..., X_n \in \mathbb{R}^d$ be i.i.d. random variables with mean μ_0. Denote the quasi-bootstrap distribution by T_n as in (9). For $b = 1, ..., B,\ i = 1, ..., n$, let X_i^{*b} be i.i.d. random variables according to quasi-bootstrap distribution T_n. Now, we can derive $\ell^{*b}(\bar{X}_n)$ in accordance with (10). Then the quantiles of $\ell^{*i}(\bar{X}_n)$ under T_n are used as an estimator for a quantile of $\ell(\mu_0)$; the $1-\alpha$ quantile is provided as the smallest value $x = \hat{\xi}_{n,\alpha}$ that satisfies

$$\frac{\#\{b \mid \ell^{*b}(\bar{X}_n) \le x\}}{B} \ge 1-\alpha. \tag{11}$$

Hence, the quasi-bootstrap calibrated $(1-\alpha)100\%$ confidence region $C_{\mathrm{QB}}^{n,\alpha}$ is written as

$$C_{\mathrm{QB}}^{n,\alpha} = \{\mu \mid \ell(\mu) \le \hat{\xi}_{n,\alpha}\}.$$

3.3 Modifying Chi-Squared Calibration

The modifying chi-squared calibration is different from the usual chi-squared calibration in that the nominal level $1-\alpha$ may be modified. To be more precise, we use the $(1-\alpha)\hat{p}$ quantile of chi-squared distribution as the threshold instead of $1-\alpha$, where $\hat{p}$ is the estimator of the existence probability.

We assume Condition 2.1 so that $\ell(\mu_0) \rightsquigarrow \chi^2_{(q)}$ as n goes to infinity. As mentioned previously, the quasi-bootstrap is aimed to estimate the existence probability $p = P(\ell(\mu_0) < \infty)$; the estimator $\hat{p}$ is defined by

$$\hat{p} = P(\ell^*(\bar{X}_n) < \infty \mid T_n) \simeq \frac{\#\{\mu_0 \in \mathrm{Conv}(X_1^{*b}, ..., X_n^{*b})\}}{B}.$$

In practice, the right-hand side is typically used as $\hat{p}$ for a sufficiently large B. Here, plugging in $p = \hat{p}$ for (7) yeilds

$$P(\ell(\mu_0) \leq \gamma_{1-\alpha}) \simeq (1-\alpha)\hat{p}. \tag{12}$$

Then, it follows immediately that

$$P(\ell(\mu_0) \leq \gamma_{(1-\alpha)/\hat{p}}) \simeq 1-\alpha,$$

where $\gamma_{(1-\alpha)/\hat{p}}$ is the $(1-\alpha)/\hat{p}$ quantile of $\chi^2_{(q)}$. It is implied that the $(1-\alpha)100\%$ confidence region based on the modifying chi-squared calibration is

$$C^{n,\alpha}_{\mathrm{MC}} \equiv \{\mu \mid \ell(\mu) \leq \gamma_{(1-\alpha)/\hat{p}}\}.$$

What matters is that, under Conditions 2.1 and 2.2, the modifying chi-squared calibrated empirical likelihood can have second order accuracy by employing the Bartlett correction, that is, for an appropriately chosen constant,

$$P\left(\ell(\mu_0)\left(1-\frac{a}{n}\right) \leq \gamma_{(1-\alpha)/\hat{p}}\right) = 1-\alpha+O(n^{-2}).$$

The corresponding confidence region is expressed as

$$C^{n,\alpha}_{\mathrm{MC}_2} \equiv \left\{\mu \mid \ell(\mu)\left(1-\frac{a}{n}\right) \leq \gamma_{(1-\alpha)/\hat{p}}\right\}.$$

These properties will be seen once again through Theorem 4.3 in Section 4.

4 Asymptotic Properties

In this section, we introduce some asymptotic properties of our methods. The consistency of the quasi-bootstrap calibration is illustrated first, which is broken down into two parts, Proposition 4.1 and Theorem 4.2. Then we introduce Theorem 4.3 that summarizes the asymptotic properties of the modifying chi-squared calibration. We will prove them in Section 6.

4.1 Consistency of the Quasi-Bootstrap Calibration

The consistency of the quasi-bootstrap calibration can be regarded as the consistency of the quasi-bootstrap estimator relative to the Kolmogorov–Smirnov distance; see Chapter 23 in [7]. That is, we call that the quasi-bootstrap calibration is consistent if

$$\sup_{\gamma}\left|P(\ell^*(\bar{X}_n) \leq \gamma \mid T_n) - P(\ell(\mu_0) \leq \gamma)\right| \xrightarrow{P} 0.$$

The following lemma indicates that it suffices to show the consistency of the quasi-bootstrap calibration that, for every γ,

$$P(\ell(\mu_0) \leq \gamma) \to P(\chi^2_{(q)} \leq \gamma), \quad P(\ell^*(\bar{X}_n) \leq \gamma \mid T_n) \xrightarrow{P} P(\chi^2_{(q)} \leq \gamma).$$

Proposition 4.1. *Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean μ_0, and let $\bar{X}_n = \sum_{i=1}^n X_i/n$. For some $q > 0$, suppose that $\ell(\mu_0) \rightsquigarrow \chi^2_{(q)}$, and that $\ell^*(\bar{X}_n) \rightsquigarrow \chi^2_{(q)}$, given $X_1, X_2, \ldots$, in probability, where $\ell^*(\bar{X}_n)$ is defined as (10). Then, the quasi-bootstrap calibration is consistent.*

Since Owen (1990) showed that, if i.i.d. random sample X_i has mean μ_0 and finite variance-covariance matrix of rank $q > 0$, then $\ell(\mu_0) \rightsquigarrow \chi^2_{(q)}$, what remains to be done for our objective is to show the assumption left in Proposition 4.1 is satisfied, that is, conditionally on $X_1, X_2, ..., \ell^*(\bar{X}_n) \rightsquigarrow \chi^2_{(q)}$ in probabiulity. In fact, a strong result,

$$P(\ell^*(\bar{X}_n) \leq \gamma \mid T_n) \to P(\chi^2_{(q)} \leq \gamma) \quad \text{a.s.}$$

is obtained.

Theorem 4.2. *Let $X_1, X_2, ... \in \mathbb{R}^d$ be i.i.d. random variables having mean μ_0 and variance-covariance matrix V_0, and let $X_1^*, ..., X_n^*$ be i.i.d. random variables according to T_n, given $X_1, X_2, ...$, as in (9). The corresponding quasi-bootstrap empirical likelihood ratio function is denoted by $\ell^*(\bar{X}_n)$ as (10). Then under Condition 2.1, for almost every sequence $X_1, X_2, ...$,*

$$\ell^*(\bar{X}_n) \rightsquigarrow \chi^2_{(q)},$$

given $X_1, X_2, ...$

Combining Proposition 4.1 and Theorem 4.2, we can establish the consistency of the quasi-bootstrap calibration. We refer the reader to Section 6 for their proofs; the similar processes can also be applied just as well to the bootstrap calibration, which indicates that the bootstrap calibration is asymptotically consistent.

4.2 Asymptotic Properties of the Modifying Chi-Squared Calibration

The modifying chi-squared calibration is not only asymptotically valid, but Bartlett correctable and then can achieve the second order accuracy. These properties are put together by the next theorem. As for its proof, see Section 6.

Theorem 4.3. *Let $X_1, X_2, ... \in \mathbb{R}^d$ be i.i.d. random variables having mean μ_0 and variance-covariance matrix V_0 satisfying Condition 2.1. Then, for $\alpha \in (0, 1)$,*

$$\hat{p} \to 1 \quad \text{a.s.}, \quad P(\ell(\mu_0) \leq \gamma_{(1-\alpha)/\hat{p}}) \to 1 - \alpha,$$

and under Condition 2.2,

$$P\left(\ell(\mu_0)\left(1 - \frac{a}{n}\right) \leq \gamma_{(1-\alpha)/\hat{p}}\right) = 1 - \alpha + O(n^{-2}),$$

where a is an appropriately chosen constant.

Consequently, both of our methods are asymptotically valid. The reason behind is $\hat{p} \to 1$ a.s., implying that the quasi-bootstrap calibration is asymptotically equivalent to the bootstrap calibration, and the same goes for the modifying chi-squared calibration and the chi-squared calibration. In other words, the conventional chi-squared calibration can be regarded as the modifying chi-squared calibration with $\hat{p} = 1$. Most important thing is, as repeatedly said yet, $p = P(\mu_0 \in \text{Conv}(X_1, ..., X_n))$ itself indeed produces superfluous bias and makes the accuracy of empirical likelihood inference poor in practice; it is necessary to incorporate the information of p into the conventional empirical likelihood inferences.

Table 1: Simulated coverage probabilities with the underlying distributions $N(0, I)$ with sample size n and dimension d; these are based on 1000 samples and, as for B-EL and QB-EL, 2000 bootstrap replicants. The columns of MC-EL and QB-EL are shown in bold for highlighting our methods.

Dist	$\boldsymbol{n}$	$\boldsymbol{d}$	**level**	**EL**	**Bart-EL**	**MC-EL**	**B-EL**	**QB-EL**
$N(0, I)$	10	1	0.90	0.853	0.872	**0.877**	0.896	**0.896**
			0.95	0.898	0.917	**0.930**	0.954	**0.950**
			0.99	0.956	0.966	**0.977**	0.986	**0.989**
		2	0.90	0.747	0.795	**0.843**	0.917	**0.911**
			0.95	0.804	0.850	**0.910**	0.965	**0.958**
			0.99	0.882	0.905	**0.974**	0.974	**0.974**
	20	3	0.90	0.802	0.839	**0.840**	0.904	**0.905**
			0.95	0.867	0.898	**0.899**	0.949	**0.948**
			0.99	0.936	0.951	**0.953**	0.991	**0.991**
		5	0.90	0.713	0.789	**0.813**	0.950	**0.930**
			0.95	0.782	0.844	**0.875**	0.990	**0.970**
			0.99	0.871	0.902	**0.993**	0.993	**0.993**
	30	3	0.90	0.845	0.876	**0.876**	0.900	**0.898**
			0.95	0.906	0.930	**0.930**	0.960	**0.958**
			0.99	0.975	0.984	**0.986**	0.993	**0.993**
		7	0.90	0.682	0.773	**0.778**	0.926	**0.916**
			0.95	0.765	0.841	**0.852**	0.970	**0.956**
			0.99	0.874	0.906	**0.925**	0.998	**0.995**
	50	10	0.90	0.707	0.780	**0.780**	0.900	**0.898**
			0.95	0.785	0.841	**0.841**	0.966	**0.960**
			0.99	0.887	0.929	**0.929**	0.997	**0.995**
		15	0.90	0.496	0.631	**0.647**	0.976	**0.933**
			0.95	0.583	0.711	**0.732**	0.993	**0.979**
			0.99	0.718	0.816	**0.976**	0.997	**0.997**

5 Simulation Studies

We put some of simulation studies together in this chapter, which compare our methods and conventional ones. We denote throughout this section the chi-squared calibrated empirical likelihood by EL, the Bartlett corrected empirical likelihood by Bart-EL, the second order modifying chi-squared calibrated empirical likelihood by MC-EL, the bootstrap calibrated empirical likelihood confidence region by B-EL, and the quasi-bootstrap calibrated empirical likelihood confidence region by QB-EL. Namely, EL, Bart-EL, MC-EL, B-EL, and QB-EL have the corresponding coverage probabilities, $P(\ell(\mu_0) \le \gamma_{1-\alpha})$, $P(\ell(\mu_0)(1 - a/n) \le \gamma_{1-\alpha})$, $P(\ell(\mu_0)(1 - a/n) \le \gamma_{(1-\alpha)/\hat{p}})$, $P(\ell(\mu_0) \le \hat{\zeta}_{n,\alpha})$, and $P(\ell(\mu_0) \le \hat{\xi}_{n,\alpha})$, respectively.

Table 1, 2 report some of simulated coverage probabilities of confidence regions for the means of $N(0, I)$ and $\chi^2_{(1)}$, respectively. Every result in both tables are based on 1000 samples; each simulated observations $X_1, ..., X_n \in \mathbb{R}^d$ yields an empirical likelihood ratio function for the mean. At the same time, 2000 bootstrap and quasi-bootstrap samples per simulation are drawn from each of $\mathbb{F}_n$ and T_n, they are needed for B-EL, QB-EL, and estimating existence

Table 2: Simulated coverage probabilities with the underlying distributions $\chi^2_{(1)}$ with sample size n and dimension d; for example, when $d = 2$, we use $(X_1, X_2) \in \mathbb{R}^2$ with $X_1, X_2 \sim \chi^2_{(1)}$. The columns of MC-EL and QB-EL are shown in bold for highlighting our methods.

Dist	***n***	***d***	**level**	**EL**	**Bart-EL**	**MC-EL**	**B-EL**	**QB-EL**
$\chi^2_{(1)}$	10	1	0.90	0.773	0.876	**0.888**	0.861	**0.859**
			0.95	0.832	0.908	**0.923**	0.901	**0.903**
			0.99	0.897	0.935	**0.951**	0.956	**0.959**
		2	0.90	0.616	0.817	**0.858**	0.862	**0.832**
			0.95	0.685	0.842	**0.888**	0.893	**0.886**
			0.99	0.766	0.870	**0.897**	0.897	**0.897**
	20	3	0.90	0.697	0.861	**0.866**	0.896	**0.870**
			0.95	0.771	0.901	**0.909**	0.943	**0.932**
			0.99	0.861	0.935	**0.966**	0.987	**0.983**
		5	0.90	0.567	0.850	**0.871**	0.933	**0.877**
			0.95	0.632	0.870	**0.925**	0.943	**0.935**
			0.99	0.730	0.894	**0.943**	0.943	**0.943**
	30	3	0.90	0.809	0.902	**0.902**	0.912	**0.904**
			0.95	0.870	0.936	**0.937**	0.957	**0.948**
			0.99	0.937	0.974	**0.976**	0.996	**0.992**
		7	0.90	0.566	0.872	**0.880**	0.938	**0.878**
			0.95	0.650	0.905	**0.916**	0.980	**0.936**
			0.99	0.751	0.934	**0.982**	0.984	**0.984**
	50	10	0.90	0.595	0.902	**0.902**	0.927	**0.887**
			0.95	0.674	0.940	**0.941**	0.974	**0.948**
			0.99	0.803	0.974	**0.976**	0.998	**0.991**
		15	0.90	0.369	0.877	**0.885**	0.972	**0.871**
			0.95	0.436	0.902	**0.915**	0.973	**0.931**
			0.99	0.558	0.926	**0.973**	0.973	**0.973**

probabilities in MC-EL. Here, as for the quasi-bootstrap distribution T_n, we use the Gaussian distribution having the empirical mean and variance-covariance matrix as G. Then it is obvious that T_n meets the requirements of the quasi-bootstrap distribution because G is projection symmetric and has a finite variance-covariance matrix. The columns of MC-EL and QB-EL are shown in bold for highlighting our methods.

Throughout these tables, B-EL and QB-EL stand out conspicuously than the others. What is most interesting is that QB-EL remains accurate when B-EL overestimates the coverage level. Another interesting thing is that MC-EL dominates Bart-EL in every case. However, MC-EL obtains high accuracy only in low-dimensional settings, whereas QB-EL does in almost every case. One apparent reason for this is the vulnerability of Bart-EL to multidimensional applications. By and large, QB-EL may be the best for real applications of empirical likelihood inference.

6 Proofs

We require the following lemma for the proof of Proposition 4.1.

Lemma 6.1. *(Van der Vaart (1998), Lemma 21.2) For any sequence of distribution functions, $F_n^{-1} \rightsquigarrow F^{-1}$ if and only if $F_n \rightsquigarrow F$, in which F^{-1} and F_n^{-1} are left-continuous inverses defined by $F^{-1}(p) = \inf\{x \mid F(x) \geq p\}$ and $F_n^{-1}(p) = \inf\{x \mid F_n(x) \geq p\}$.*

Proof. See [7]. □

Proof of Proposition 4.1. We can assume throughout that

$$P(\ell^*(\bar{X}_n) \leq \gamma \mid T_n) \to P(\chi^2_{(q)} \leq \gamma) \quad \text{a.s.},$$

without loss of generality since every subsequence of $P(\ell^*(\bar{X}_n) \leq \gamma \mid T_n)$ has a further subsequence which converges almost surely.

Let H be the distribution of $\chi^2_{(q)}$ and let $\hat{H}_n$ be the distribution of $\ell^*(\bar{X}_n)$. Denote the $1-\alpha$ quantile of $\chi^2_{(q)}$ by $\gamma_{1-\alpha}$. Note that H^{-1} is continuous at every point on $(0,1)$. Using Lemma 6.1 and Glivenko–Cantelli's theorem, we have

$$\hat{H}_n^{-1}(1-\alpha) \to H^{-1}(1-\alpha) \quad \text{a.s.}.$$

Moreover, by Slutsky's lemma

$$\ell(\mu_0) - \hat{H}_n^{-1}(1-\alpha) \rightsquigarrow \chi^2_{(q)} - H^{-1}(1-\alpha).$$

Thus

$$P\big(\ell(\mu_0) \leq \hat{H}_n^{-1}(1-\alpha)\big) \to P\left(\chi^2_{(q)} \leq H^{-1}(1-\alpha)\right) = 1-\alpha,$$

from which the consistency of the quasi-bootstrap calibration follows. □

Before getting down to the proof of Theorem 4.2, three necessary lemmas are introduced. Lemma 6.2 below is a quasi-bootstrap version of Theorem 23.4 in [7], establishing, given the original observations, the asymptotic normality of the sequence $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)$ in which $\bar{X}_n^*$ is the conditional mean under T_n.

Lemma 6.2. *Let $X_1, X_2, ...$ be i.i.d. random variables having mean μ and finite variance-covariance matrix Σ and let $X_1^*, ..., X_n^*$ be i.i.d. random variables with the quasi-bootstrap distribution $T_n = (1-\varepsilon)\mathbb{F}_n + \varepsilon G$ conditionally on $X_1, ..., X_n$. Define $\bar{X}_n = \sum_{i=1}^n X_i/n$ and also $\bar{X}_n^* = \sum_{i=1}^n X_i^*/n$. Then, given $X_1, X_2, ...$, for almost every sequence $X_1, X_2, ...$*

$$\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \rightsquigarrow N(0, \Sigma).$$

Proof. Let Y be a random variable drawn from G. Given $X_1, X_2, ...$, the mean and the variance-covariance matrix of X_i^* are

$$E[X_i^* \mid T_n] = \frac{1-\varepsilon}{n}\sum_{i=1}^n X_i + \varepsilon E[Y] = (1-\varepsilon)\bar{X}_n + \varepsilon\bar{X}_n = \bar{X}_n$$

$$\mathrm{Var}(X_i^* \mid T_n) = \frac{1-\varepsilon}{n}\sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^\top + \varepsilon E[(Y - \bar{X}_n)^2 \mid X_1, X_2, ...].$$

By the strong law of large numbers and Slutsky's Lemma,

$$\mathrm{Var}(X_i^*|T_n) \to \Sigma \quad \text{a.s.},$$

remembering that $\varepsilon \to 0$.

Next, we will show that triangular arrays of X_i^* satisfy the Lindeberg condition: for every $\eta > 0$,

$$E\left[\|X_i^*\|^2 \mathbb{1}_{\{\|X_i^*\| > \eta\sqrt{n}\}} \mid T_n\right] = \frac{1}{n}\sum_{i=1}^{n} \|X_i\|^2 \mathbb{1}_{\{\|X_i\| > \eta\sqrt{n}\}} \to 0 \quad \text{a.s.}.$$

For $M \leq \eta\sqrt{n}$, by the strong law of large numbers,

$$\begin{aligned} E\left[\|X_i^*\|^2 \mathbb{1}_{\{\|X_i^*\| > \eta\sqrt{n}\}}\right] &\leq \frac{1}{n}\sum_{i=1}^{n} \|X_i\|^2 \mathbb{1}_{\{\|X_i\| > M\}} \\ &\to E\left[\|X_i\|^2 \mathbb{1}_{\{\|X_i\|^2 > M\}}\right] \quad \text{a.s.}. \end{aligned}$$

This implies that the Lindeberg condition holds for almost every sequence $X_1, X_2, ...$ because the last term is arbitrarily small for arbitrarily large M.

Applying the Lindeberg central limit theorem to the quasi-bootstrap replications yields the lemma. □

The next lemma guarantees that the existence probability has its limit 1:

$$P(\ell(\mu) < \infty) = P(\mu \in \mathrm{Conv}(X_1, ..., X_n)) \to 1.$$

Lemma 6.3. *Let $X_1, ..., X_n \in \mathbb{R}^d$ be i.i.d. random variables having mean μ and let Θ be the set of unit vectors in $\mathbb{R}^d$. Define*

$$p_{\mathrm{m}} = \sup_{\theta \in \Theta} P\left(\theta^\top (X_i - \mu) > 0\right).$$

Then $0 < p_{\mathrm{m}} < 1$ and

$$P\left(\mu \notin \mathrm{Conv}(X_1, ..., X_n)\right) \leq 2\binom{n}{d-1} p_{\mathrm{m}}^{n-d+1}.$$

Proof. By definition of hyperplane, $\mathrm{Conv}(X_1, ..., X_n)$ does not contain μ if, for some split of $X_1, ..., X_n$ into some $d-1$ points $X_{(1)}, ..., X_{(d-1)}$ and $n-d+1$ points left $X_{(d)}, ..., X_{(n)}$ are on one side of the hyperplane through $X_{(1)}, ..., X_{(d-1)}$ and μ_0.

For $p \in (0,1), k = 2, 3, ...$, $p^k + (1-p)^k$ is minimized by $p = 1/2$ and monotonically increasing as p grows on $[1/2, 1)$.

From these results, we have

$$P\left(\mu \notin \mathrm{Conv}(X_1, ..., X_n)\right) \leq \binom{n}{d-1}(p_{\mathrm{m}}^{n-d+1} + (1-p_{\mathrm{m}})^{n-d+1}).$$

Then the desired inequality follows. □

From now, we use the notation $\|\cdot\|$ be Euclidean norm. The following lemma is that Lemma 3 in [5] is adjusted for the quasi-bootstrap, providing some order bounds of quasi-bootstrap replicants.

Lemma 6.4. *Let $X_1, X_2, ... \in \mathbb{R}^d$ be i.i.d. random variables with $E[\|X_i\|^2] < \infty$. Denote i.i.d. quasi-bootstrap replicants by $X_1^*, X_2^*, ... \sim T_n$, given $X_1, X_2, ...$ Then, for almost every sequence $X_1, X_2, ...,$*

$$\max_{1\le i\le n} \|X_i^*\| = o(n^{1/2}),$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \|X_i^*\|^3 = o(n^{1/2}).$$

Proof. First, we show $\max_{1\le i\le n} \|X_i\| = o(n^{1/2})$ in the same way as the proof of Lemma 3 in [5]. By Markov's inequality, $\sum_{i=1}^{n} P(\|X_i\| > \sqrt{n}) \le E[\|X_i\|^2] < \infty$. Then, by Borel–Cantelli's lemma,

$$P\left(\limsup_{n\to\infty} \left\{\|X_n\| > n^{1/2}\right\}\right) = 0.$$

Likewise, for any $A > 0$,

$$P\left(\limsup_{n\to\infty} \left\{\max_{1\le i\le n} \|X_i\| n^{-1/2} \le A\right\}\right) = 1.$$

It implies that $\max_{1\le i\le n} \|X_i\| = o(n^{1/2})$.

What is more, an analogous result holds for quasi-bootstrap replicants in the same time. Let Y_i be i.i.d. random variables drawn from G. From the above result,

$$\max_{1\le i\le n} \|X_i^*\| \le \max_{1\le i\le n} \|X_i\| + \max_{1\le i\le n} \|Y_i\| = o(n^{1/2}). \tag{13}$$

Finally, by the strong law of large numbers and (13), for almost every sequence $X_1, X_2, ...,$

$$\frac{1}{n}\sum_{i=1}^{n} \|X_i^*\|^3 \le \max_{1\le i\le n} \|X_i^*\| \frac{1}{n}\sum_{i=1}^{n} \|X_i^*\|^2 = o(n^{1/2}).$$

□

We will establish Theorem 4.2 below following the flow of Proof of Theorem 3.2. in [6].

Proof of Theorem 4.2. Without loss of generality, we may assume $d = q$; otherwise consider q-dimensional linear transformation. It is also assumed that, for a sufficiently large n and almost every sequence $X_1, X_2, ...,$ $\mathrm{Var}(X_i^* \mid T_n)$ is full rank.

By Lemma 6.2, for almost every sequence $X_1, X_2, ..,$ the probability that $\bar{X}_n$ is inside the convex hull of X_i^* has the limit 1. Therefore it suffices to show that, for X_i^* the convex hull of which contains $\bar{X}_n$ and almost every sequence $X_1, X_2, ...,$ $\ell^*(\bar{X}_n) \rightsquigarrow \chi^2_{(d)}$.

When $\bar{X}_n$ is inside of the convex hull of X_i^*, there exists a unique set of weights $\omega_i > 0$ maximizing $\prod_{i=1}^n n\omega_i$ under the constraints $\sum_{i=1}^n \omega_i = 1$ and $\sum_{i=1}^n \omega_i(X_i^* - \bar{X}_n) = 0$. Through use of the method of Lagrange multipliers, we can write

$$\omega_i = \frac{1}{n}\frac{1}{1 + \lambda^\top (X_i^* - \bar{X}_n)}, \tag{14}$$

where the multiplier $\lambda = \lambda(\bar{X}_n) \in \mathbb{R}^d$ satisfies

$$g(\lambda) \equiv \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^* - \bar{X}_n}{1 + \lambda^\top (X_i^* - \bar{X}_n)} = 0.$$

In the second place, we will derive the bounds of $\|\lambda\|$. For simplicity, we may use the abbreviations "a.e.s." to denote "given $X_1, X_2, ...$, for almost every sequence" in the following. Write $\lambda = \|\lambda\|\theta$ in which $\theta \in \mathbb{R}^d$, $Y_{n,i} = \lambda^\top(X_i^* - \bar{X}_n)$, and $Z_n = \max_{1\le i\le n}\|X_i^* - \bar{X}_n\|$. Then

$$\begin{aligned}\bar{X}_n^* - \bar{X}_n &= \frac{1}{n}\sum_{i=1}^n \frac{(X_i^* - \bar{X}_n)\big(1 + Y_{n,i}\big)}{1 + Y_{n,i}} \\ &= \frac{1}{n}\sum_{i=1}^n \frac{X_i^* - \bar{X}_n}{1 + Y_{n,i}} + \frac{1}{n}\sum_{i=1}^n \frac{(X_i^* - \bar{X}_n)Y_{n,i}}{1 + Y_{n,i}} \\ &= \frac{1}{n}\sum_{i=1}^n \frac{(X_i^* - \bar{X}_n)(X_i^* - \bar{X}_n)^\top}{1 + Y_{n,i}}\lambda.\end{aligned}$$

Thus

$$\theta^\top(\bar{X}_n^* - \bar{X}_n) = \|\lambda\|\theta^\top \frac{1}{n}\sum_{i=1}^n \frac{(X_i^* - \bar{X}_n)(X_i^* - \bar{X}_n)^\top}{1 + Y_{n,i}}\theta. \tag{15}$$

Denote the bootstrap sample variance-covariance matrix by S:

$$S = \frac{1}{n}\sum_{i=1}^n (X_i^* - \bar{X}_n)(X_i^* - \bar{X}_n)^\top.$$

Remember $\omega_i > 0$ so $1 + Y_{n,i} > 0$. According to $\theta^\top g(\lambda) = 0$ and (15),

$$\begin{aligned}\|\lambda\|\theta^\top S\theta &\le \|\lambda\|\theta^\top \frac{1}{n}\sum_{i=1}^n \frac{(X_i^* - \bar{X}_n)(X_i^* - \bar{X}_n)^\top}{1 + Y_{n,i}}\theta\left(1 + \max_i Y_{n,i}\right) \\ &\le \|\lambda\|\theta^\top \frac{1}{n}\sum_{i=1}^n \frac{(X_i^* - \bar{X}_n)(X_i^* - \bar{X}_n)^\top}{1 + Y_{n,i}}\theta(1 + \|\lambda\|Z_n) \\ &= \theta^\top(\bar{X}_n^* - \bar{X}_n)(1 + \|\lambda\|Z_n),\end{aligned}$$

indicating

$$\|\lambda\|\big(\theta^\top S\theta - Z_n\theta^\top(\bar{X}_n^* - \bar{X}_n)\big) \le \theta^\top(\bar{X}_n^* - \bar{X}_n).$$

By Lemma 6.2 and 6.4,

$$\|\lambda\|(\theta^\top S\theta + o_p(1)) = O_p(n^{-1/2}) \quad \text{a.e.s..} \tag{16}$$

Here, let σ_1, σ_d be, respectively, the largest and smallest eigenvalues of V_0:

$$\sigma_1 = \sup_{\theta\in\Theta}\theta^\top V_0\theta, \quad \sigma_d = \inf_{\theta\in\Theta}\theta^\top V_0\theta.$$

To avoid confusion, component of vectors is expressed as $X_i = (X_{i1}, ..., X_{id})$, and entries of the matrices are denoted by $S = (s_{kl})_{1\le k,l\le d}$, $V_0 = (v_{kl})_{1\le k,l\le d}$. Let $X^\dagger = (X_1^\dagger, ..., X_d^\dagger)$ be a random variable drawn from G. By the strong law of large numbers, we have

$$\begin{aligned}E[s_{kl} \mid T_n] =& E[(X_{ik}^* - \bar{X}_{nk})(X_{il}^* - \bar{X}_{nl}) \mid T_n] \\ =& \frac{1-\varepsilon}{n}\sum_{i=1}^n (X_{ik} - \bar{X}_{nk})(X_{il} - \bar{X}_{nl}) \\ &+ \varepsilon E[(X_k^\dagger - \bar{X}_{nk})(X_l^\dagger - \bar{X}_{nl}) \mid X_1, X_2, ...] \\ \to& v_{kl} \quad \text{a.s.},\end{aligned}$$

and thus

$$s_{kl} = \frac{1}{n}\sum_{i=1}^{n}(X_{ik}^* - \bar{X}_{nk})(X_{il}^* - \bar{X}_{nl}) \to v_{kl} \quad \text{a.e.s.},$$

from which we obtain bounds about σ_1 and σ_d. Let $\theta_0 = \arg\max_{\theta\in\Theta} \theta^\top S\theta$. Then

$$\begin{aligned}\sup_{\theta\in\Theta} \theta^\top S\theta - \sup_{\theta\in\Theta} \theta^\top V_0\theta &\le \sum_{k,l=1}^{d} \theta_{0k}\theta_{0l}s_{kl} - \sum_{k,l=1}^{d} \theta_{0k}\theta_{0l}v_{kl} \\ &= \sum_{k,l=1}^{d} \theta_{0k}\theta_{0l}s_{kl} - \sum_{k,l=1}^{d} \theta_{0k}\theta_{0l}(s_{kl} + o_p(1)) \quad \text{a.e.s.} \\ &= o_p(1) \quad \text{a.e.s.},\end{aligned}$$

and likewise

$$\inf_{\theta\in\Theta} \theta^\top S\theta - \inf_{\theta\in\Theta} \theta^\top V_0\theta \ge o_p(1) \quad \text{a.e.s.}.$$

Therefore we have

$$\sigma_1 + o_p(1) \ge \theta^\top S\theta \ge \sigma_d + o_p(1) \quad \text{a.e.s.}. \tag{17}$$

Combining (16) and (17), we establish the bounds of $\|\lambda\|$:

$$\|\lambda\| = O_p(n^{-1/2}) \quad \text{a.e.s.}.$$

Now by Lemma 6.4, we have

$$\max_{1\le i\le n} |Y_{n,i}| = O_p(n^{-1/2})o(n^{1/2}) = o_p(1) \quad \text{a.e.s.}, \tag{18}$$

and hence

$$\begin{aligned}\left\| \frac{1}{n}\sum_{i=1}^{n} \frac{(X_i^* - \bar{X}_n)Y_{n,i}^2}{1+Y_{n,i}} \right\| &\le \frac{1}{n}\sum_{i=1}^{n} \|X_i^* - \bar{X}_n\|^3 \|\lambda\|^2 |1+Y_{n,i}|^{-1} \\ &= o(n^{1/2})O_p(n^{-1})O_p(1) \\ &= o_p(n^{-1/2}) \quad \text{a.e.s.}.\end{aligned}$$

It is implied that

$$\lambda = S^{-1}(\bar{X}_n^* - \bar{X}_n) + \beta \quad \text{a.e.s.}, \tag{19}$$

where $\beta = o_p(n^{-1/2})$, since we can expand

$$\begin{aligned}0 &= g(\lambda) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i^* - \bar{X}_n)\left(1 - Y_{n,i} + \frac{Y_{n,i}^2}{1+Y_{n,i}}\right) \\ &= \bar{X}_n^* - \bar{X}_n - S\lambda + \frac{1}{n}\sum_{i=1}^{n} \frac{(X_i^* - \bar{X}_n)Y_{n,i}^2}{1+Y_{n,i}}.\end{aligned}$$

What is more, by (18) and use of the Taylor expansion,

$$\log(1+Y_{n,i}) = Y_{n,i} - \frac{1}{2}Y_{n,i}^2 + \eta_i, \tag{20}$$

where, for some $0 < B < \infty$,

$$P(|\eta_i| \le B|Y_{n,i}|^3, 1 \le i \le n \mid T_n) \to 1 \quad \text{a.s..}$$

Substituting (19) and (20) into (14), we can write

$$\begin{aligned}
\ell^*(\bar{X}_n) &= -2\sum_{i=1}^{n} \log(n\omega_i) \\
&= 2\sum_{i=1}^{n} \log(1 + Y_{n,i}) \\
&= 2\sum_{i=1}^{n} Y_{n,i} - \sum_{i=1}^{n} Y_{n,i}^2 + 2\sum_{i=1}^{n} \eta_i \\
&= 2n\lambda^\top(\bar{X}_n^* - \bar{X}_n) - n\lambda^\top S\lambda + 2\sum_{i=1}^{n} \eta_i \\
&= n(\bar{X}_n^* - \bar{X}_n)^\top S^{-1}(\bar{X}_n^* - \bar{X}_n) - n\beta^\top S\beta + 2\sum_{i=1}^{n} \eta_i. \qquad (21)
\end{aligned}$$

The quasi-bootstrap precision matrix S^{-1} can be eigendecomposed as $S^{-1} = U^\top \Sigma U$ such that Σ is a diagonal matrix having a decomposition of the form $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$. Using Lemma 6.2, we have

$$\sqrt{n}\Sigma^{1/2}U(\bar{X}_n^* - \bar{X}_n) \rightsquigarrow N(0, I) \quad \text{a.e.s.,}$$

from which, by the continuous mapping theorem,

$$n(\bar{X}_n^* - \bar{X}_n)^\top S^{-1}(\bar{X}_n^* - \bar{X}_n) \rightsquigarrow \chi^2_{(d)} \quad \text{a.e.s..}$$

Furthermore, we have

$$n\beta^\top S\beta = no_p(n^{-1/2})O_p(1)o_p(n^{-1/2}) = o_p(1) \quad \text{a.e.s.,}$$

and also

$$\left|\sum_{i=1}^{n} \eta_i\right| \le B\|\lambda\|^3 \sum_{i=1}^{n} \|X_i^* - \bar{X}_n\|^3 = O_p(n^{-3/2})o_p(n^{3/2}) = o_p(1) \quad \text{a.e.s..}$$

Hence, the final conclusion follows from (21) by taking $n \to \infty$. □

Lastly, Theorem 4.3 is proved.

Proof of Theorem 4.3. By Lemma 6.3, for some $p_m \in [0.5, 1)$,

$$P(\ell^*(\bar{X}_n) < \infty \mid T_n) = 1 + O\left(n^{d-1}p_m^{n-d+1}\right) \quad \text{a.s..} \qquad (22)$$

Then the first part of the theorem immediately follows from (22) under Condition 2.1, the continuity of $\chi^2_{(q)}$ distribution function and Theorem 1 in [5].

The second part essentially follows from [1, 2]; they showed that, under Condition 2.2, for a fixed constant a,

$$P\left(\left(\ell(\mu_0)+O_p(n^{-3/2})\right)\left(1-\frac{a}{n}+O_p(n^{-3/2})\right)\le\gamma_{1-\alpha}\right)=1-\alpha+O(n^{-2}). \tag{23}$$

Denote $\ell(\mu_0)(1-a/n)$ by $\ell'_n(\mu_0)$ and $O(n^{d-1}p_m^{n-d+1})$ by $\varepsilon_n>0$ for simplicity. Combining and (22) and (23), we have

$$\begin{aligned}\left|P\left(\ell'_n(\mu_0)\le\gamma_{(1-\alpha)/\hat{p}}\right)-1+\alpha\right|\le&\left|P\left(\ell'_n(\mu_0)\le\gamma_{(1-\alpha)/\hat{p}}\right)-P\left(X_n\le\gamma_{1-\alpha}\right)\right|\\&+\left|P\left(\ell'_n(\mu_0)\le\gamma_{1-\alpha}\right)-1+\alpha\right|\\\le&P\left(\ell'_n(\mu_0)\le\gamma_{1-\alpha}+\varepsilon_n\right)+O(n^{-2})\\=&O(n^{-2})\end{aligned}$$

□

References

[1] DiCiccio, T., Hall, P. and Romano, J. (1988). Bartlett adjustment for empirical likelihood. Technical Report 298, Dept. Statistics, Stanford Univ., Stanford, CA.

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

[3] Efron, B. (1979). Bootstrap Methods: Another Look at the Jacknife. *Ann. Statist.* **7** 1–26.

[4] Owen, A. B. (1988). Empirical Likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249.

[5] Owen, A. B. (1990). Empirical Likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120.

[6] Owen, A. B. (2001). *Empirical Likelihood*. Chapman and Hall/CRC, London, 2001.

[7] Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge, U.K.