EMPIRICAL LIKELIHOOD RATIO

YAN LIU

1. Reference

Owen (1991), AS.
Chen and Hall (1993), AS
Chen and Wong (2006), Sinica
Owen (2010), book.

2. Notations

1. K	an rth-order kernel function
2. M	the block length
3. L	the gap between the beginnings of two adjacent blocks
4. r	some integer $r \geq 2$
5. $\theta_q = F^{-1}(q)$	the qth quantile
6. $f = F'$	the first derivative of F
7. G and G_h	proper distribution functions ($G_h(x) = G(x/h)$)
8. h	bandwidth
9. c	satisfying $P(\chi_1^2 \leq c) = \alpha$.
10. μ_j	$= E[G\{(\theta_q - X_i)/h\} - q]^j$
11. $\hat{\theta}_q$	the usual estimate of θ_q
12. $\hat{\mu}_j$	$= n^{-1}\sum_{i=1}^n [G\{(\hat{\theta}_q - X_i)/h\} - q]^j$
13. $\beta = 1/6(3\mu_2^{-2}\mu_4 - 2\mu_2^{-3}\mu_3^2)$	
14. γ	either β or $\hat{\beta}$
15. $d(c,\gamma) = c(1 + n^{-1}\gamma)$	for the Bartlett-corrected confidence region $I_{h,d(c,\beta)}$

3. Concepts and definitions

3.1. **rth-order kernel function.** for some integer $r \geq 2$ and constant $\kappa \neq 0$, K is a function satisfying

(3.1) eqch:3.1
$$\int u^j K(u)du = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \leq j \leq r-1, \\ \kappa, & \text{if } j = r. \end{cases}$$

Date: December 10, 2014.

3.2. **Empirical log likelihood.** Some functions are

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

and

$$L_h(\theta) = \sup_{p:\hat{F}_{p,h}(\theta)=q} \prod_{i=1}^{n}(np_i).$$

ELL is defined by

$$l_h(\theta) = -2\log L_h(\theta).$$

3.2.1. *Blockwise empirical likelihood.*

(1) some functions

$$g_h(X_i, \theta_q) = G_h(\theta_q - X_i) - q$$

(2) the i-th block average

$$T_i(\theta_q) = \frac{1}{M}\sum_{j=1}^{M} g_h(X_{(i-1)L+j}, \theta_q)$$

(3) The block empirical likelihood for θ_q

$$L_h(\theta_q) = \sup \prod_{i=1}^{Q} p_i,$$

subject to

$$\sum_{i=1}^{Q} p_i = 1, \quad \sum_{i=1}^{Q} p_i T_i(\theta_q) = 0.$$

4. Assumptions

4.1. **Assumption A.**

(A1) K satisfies (3.1) and is bounded and compactly supported.

(A2) f and $f^{(r-1)}$ exist in a neighborhood of θ_q and are continuous at θ_q.

(A3) $f(\theta_q) > 0$.

(A4) for some $t > 0$, $nh^t \to 0$.

4.2. **Assumption A'.** In addition to Assumption A,

$$nh^{2r} \to 0 \quad \text{and} \quad nh/\log n \to \infty.$$

4.3. **Assumption C.**

(C1) $\{X_i\}_{i=1}^n$ is a strictly stationary α-mixing sequence.

(C2) $\sum_{k=1}^{\infty} k\alpha^{1/p}(k) < \infty$ for some $p > 1$.

(C3) the spectral density $\phi(t)$ of $\{\mathbb{1}\{X_k < \theta_q\}_{k=1}^n$ satisfies $\phi(0) > 0$.

(C4) (A1)

(C5) $nh^{2r} \to 0$ and $nh \to \infty$.

(C6) (A2) and (A3).

(C7) The block length $M \to \infty$ and $M = o(n^{1/2})$. The gap L satisfies $kL \leq M$ and $(k+1)L > M$ for some $k > 0$.

5. Formulae

For the solution λ of the equation

$$\begin{aligned}\sum_{i=1}^n w_i(1+\lambda w_i)^{-1} &= 0.\\ \sum_{i=1}^n \lambda w_i(1+\lambda w_i)^{-1} &= 0.\\ \sum_{i=1}^n \frac{1+\lambda w_i}{1+\lambda w_i} &= n.\\ \sum_{i=1}^n \frac{1}{1+\lambda w_i} &= n.\end{aligned}$$

Note that

$$-\frac{w_i}{1+\lambda w_i} = \frac{\lambda w_i^2}{1+\lambda w_i} - \frac{(1+\lambda w_i)w_i}{1+\lambda w_i},$$

also

$$\sum_{i=1}^n \Big\{\frac{\lambda w_i^2}{1+\lambda w_i} - w_i\Big\} = 0.$$

6. Results

Theorem 6.1. *Suppose Assumption A. If $nh^{2r} \to 0$, then*

$$l_h(\theta_q) \xrightarrow{\mathcal{L}} \chi_1^2.$$

Proof. **Step 1.** Define $w_i = w_i(\theta_q) = G_h(\theta_q - X_i) - q$ with the solution $\lambda = \lambda(\theta_q)$ of the equation

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

Note that $p_i = n^{-1}(1+\lambda w_i)^{-1} \geq 0$, we have

$$|1+\lambda w_i|^{-1} \geq (1+|\lambda| \max|w_i|)^{-1}.$$

□

Theorem 6.2. *Suppose Assumption A'. If nh^r is bounded, then*

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

The right-hand side cannot be rendered equal to $\alpha + o(n^{-1})$ by appropriately choosing h.

Theorem 6.3. *If n^3h^{2r} is bounded, for either $\gamma = \beta$ or $\gamma = \hat{\beta}$,*

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

Suppose $\beta_0 = 1/6q^{-1}(1-q)^{-1}(1-q+q^2)$. Since $\beta = \beta_0 + O(h)$,

$$P(\theta_q \in I_{h,d(c,\beta_0)}) = \alpha + O(n^{-1}h).$$

7. Introduction

(1) If $F_n \xrightarrow{\mathcal{L}} F_0$, then

$$T(F_n) \xrightarrow{\mathcal{L}} T(F_0) \quad !$$

8. Setting

(1) The empirical distribution

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i};$$

(2) The common distribution function F_0;
(3) The likelihood function

$$L(F) = \prod_{i=1}^{n} F\{x_i\} \quad \text{under } F;$$

(4) The empirical likelihood ratio function

$$R(F) = L(F)/L(F_n);$$

(5) Confidence Region C for $T(F_0)$

$$C = \{T(F)|R(F) \geq r\}.$$

9. Results

⟨owen1990:thm1⟩ **Theorem 9.1.** *Let* X, X_1, ... *be i.i.d. random vectors in* $\mathbb{R}^p$, *with* $E(X) = \mu_0$ *and* $V(X) = \Sigma$ *of rank* $q > 0$. *For positive* $r < 1$, *let* $C_{r,n} = \{\int X\, dF | F \ll F_n, R(F) \geq r\}$. *Then* $C_{r,n}$ *is a convex set and*

$$\lim_{n\to\infty} P(\mu_0 \in C_{r,n}) = P(\chi^2_{(q)} \leq -2\log r).$$

Moreover if $E|X|^4 < \infty$, *then*

$$|P(\mu \in C_{r,n}) - P(\chi^2_{(q)} \leq -2\log r)| = O(n^{-1/2}).$$

Theorem 9.2 (Empirical likelihood for triangular arrays)**.** *Let* $Z_{in} \in \mathbb{R}^p$ *for* $1 \leq i \leq n$ *and* $p \leq n < \infty$, *be a collection of random vectors, with* $Z_{1n}, \ldots, Z_{nn}$ *independent for each* n. *Suppose that* $E(Z_{in}) = m_n$, $V(Z_{in}) = V_{in}$ *and let* $V_n = (1/n)\sum_{i=1}^n V_{in}$, $\sigma_{1n} = \max \operatorname{eig}(V_n)$ *and* $\sigma_{pn} = \min \operatorname{eig}(V_n)$. *Assume that as* $n \to \infty$,

(9.1) owen91:eq3.3a
$$P(m_n \in \operatorname{ch}(\{Z_{1n}, \ldots, Z_{nn}\})) \to 1$$

and

(9.2) owen91:eq3.3b
$$n^{-2}\sum_{i=1}^n E(\|Z_{in} - m_n\|^4 \sigma_{1n}^{-2}) \to 0$$

and that for some $c > 0$ *and all* $n \geq p$,

(9.3) owen91:eq3.3c
$$\sigma_{pn}/\sigma_{1n} \geq c.$$

Then $-2\log R(m_n) \to \chi^2_{(p)}$ *in distribution as* $n \to \infty$, *where*

$$R(m) = sup\{\prod n\omega_i | \omega_i \geq 0, \sum \omega_i = 1, \sum \omega_i Z_{in} = m\}.$$

Remark 9.3. The error in Theorem 9.1 is shown to be
- $O(n^{-1})$ if the assumptions justifying Edgeworth expansions are met;
- $O(n^{-2})$ if there is a Bartlett factor;
- $O(n^{-1/2})$ still obtains for one sided problems.

10. Some lemmas

10.1. **Setting.**

(1) Let $w = (w_1, \ldots, w_n)$ be defined as

(10.1) owen1990:eq2.1
$$w_i \geq 0, \qquad \sum_{j\colon X_j = X_i} w_j = F(X_i)$$

for $1 \leq i, j \leq n$.

(2) Define

$$\tilde{R}(F, w) = \prod_{i=1}^n n w_i.$$

10.2. **Lemmas.**

Lemma 10.1. *For any* $r \in [0,1]$,

$$\{F|R(F) \geq r\} = \{F|\tilde{R}(F,w) \quad \textit{some } w \textit{ satisfying } (10.1)\}$$

Lemma 10.2. *Let* F_0 *be a distribution on* $\mathbb{R}^p$ *with mean* μ_0 *and finite covariance matrix* Σ *of full rank* p*. Let* Ω *be the set of unit vectors in* $\mathbb{R}^p$*. Then for* $X \sim F_0$*,*

$$\inf_{\theta\in\Omega} P((X-\mu_0)'\theta > 0) > 0.$$

Sketch of proof.

(1) Suppose $\mu_0 = 0$;
(2) By compactness of Ω, we have $\theta_n^* \to \theta_0 \in \Omega$;

□

11. Regression model

In the regression model, the data are of the form (x_i, y_i) for $1 \leq i \leq n$.

$$Y_i = x_i\beta_0 + \epsilon_i,$$

where $\beta_0 \in \mathbb{R}^p$ is a column vector of coefficients and ϵ_i is a random variable with mean 0 and variance $\sigma^2(x_i) < \infty$.

On the condition (9.1)
Let $P = \{x_i|Y_i - x_i\beta_0 > 0\}$ and $N = \{x_i|Y_i - x_i\beta_0 < 0\}$. If

$$\mathrm{ch}(N) \cap \mathrm{ch}(P) \neq \emptyset, \tag{11.1}$$

(11.1) owen91:eq5.1

then 0 is in the convex hull of the Z_{in}.

Proof. Using contraposition. □

On the condition (9.2)
Introduce

$$\mu_4(x) = \int (Y-\mu(x))^4\, dF_x.$$

If $\min\mathrm{eig}(V_n) > a > 0$ for all sufficiently large n, a sufficient condition for (9.2) is

$$n^{-2}\sum_{i=1}^{n}\|x_i\|^4\mu_4(x_i) \to 0. \tag{11.2}$$

(11.2) owen91:eq5.2

On the condition (9.3)
Note that

$$V_n = \frac{1}{n}\sum x_i' x_i \sigma^2(x_i).$$

(9.3) follows if $(1/n)\sum\|x_i\|^{\alpha+2} < \infty$ and both $\min\mathrm{eig}(1/n)X'X$ and $\sigma^2(x_i)$ are bounded away from 0.

Corollary 11.1. Let $n_0 \geq p$, $\alpha \geq 0$ and $a, b > 0$. Assume that (11.2) holds and as $n \to \infty$, (11.1) holds with probability tending to 1. Suppose $a < \sigma^2(x_i) < b\|x_i\|^\alpha$ for all i and that for all $n \geq n_0$, $a < \min \mathrm{eig}(X'X)/n$ and $(1/n)\sum\|x_i\|^{\alpha+2} < b$. Then

$$-2\log R(\beta_0) \to_d \chi^2_{(p)} \quad \text{as } n \to \infty.$$

12. Linear Algebra

12.1. **Variance.** Suppose that $Z_i = Z_i(\beta) = x_i'(Y_i - x_i\beta)$, where x_i now the x_i are not random. If $\beta = \beta_0$ in the regression model, then

$$V(Z_i) = V(x_i'\epsilon_i) = E(x_i'\epsilon\,(x_i'\epsilon)') = x_i'E(\epsilon\epsilon')x_i = x_i'x_i\sigma^2(x_i).$$

13. Simplex

A k-simplex is a k-dimensional polytope which is the convex hull of its $k+1$ vertices. Suppose the $k+1$ points $u_0, \ldots, u_k \in \mathbb{R}^n$ are affinely independent, which means $u_1 - u_0, \ldots, u_k - u_0$ are linearly dependent. Then the simplex determined by them is the set of points

$$C = \{\theta_0 u_0 + \cdots + \theta_k u_k | \theta_i \geq 0, 0 \leq i \leq k, \sum_{i=0}^{k} \theta_i = 1\}.$$

Note that

$$\theta_0 u_0 + \cdots + \theta_k u_k = u_0 + \sum_{i=1}^{k} \theta_i(u_i - u_0),$$

and $u_1 - u_0, \ldots, u_k - u_0$ are linearly dependent. We can see that u_0 causes "affine" and other points cause the figure.

Table 2. Simplex

0-simplex	a single point
1-simplex	a line
2-simplex	triangle
3-simplex	tetrahedron
4-simplex	5-cell

?⟨simplex⟩?

13.1. **face.** The convex hull of any nonempty subset of the $n+1$ points that define an n-simplex is called a face of the simplex. In particular, the convex hull of a subset of size $m+1$ is an m-simplex, called an m-face of the n-simplex.

Remark 13.1. The number of m-faces is equal to the binomial coefficient $\binom{n+1}{m+1}$.

Example 1.
The vertices of tetrahedron are ${}_4C_1 = 4$;
The edges of tetrahedron are ${}_4C_2 = 6$;
The faces of tetrahedron are ${}_4C_3 = 4$.

14. Applications

EL method is applied to the problem of setting a confidence interval of

(1) means,
(2) regression models,
(3) generalized linear models,
(4) estimating equations,
(5) space curves estimated by kernel smooths.

Furthermore, they can be applied to

(6) quantiles

15. Advantages

15.1. ELR.
The vantages are that

- it does not require us to specify a family of distributions for the data; (cf. bootstrap, jackknife.)
- it makes an automatic determination of the shape of confidence regions; (cf. parametric likelihood methods.)
- it enables the shape of a region

Empirical likelihood can be thought of as a bootstrap that does not resample, and as a likelihood without parametric assumptions.

15.2. smoothed ELR.

- it is not necessary to accurately determine an "optimal" value of the parameter.
- it is Bartlett-correctable.

Table 3. Comparison of ELR and simple confidence interval

ELR	Order	A.B.C. order	CI	Order	normal-approx
regular case	$O(n^{-1})$	$O(n^{-2})$			
quantile	$O(n^{-1/2})$		(two-sided) sign-test CI (or binomial-method)	$O(n^{-1/2})$	$O(n^{-1})$
smoothed	$O(n^{-1})$	$O(n^{-2})$	Interpolation method	$O(n^{-1})$	not B.C.

?⟨default⟩?

16. Foundations

16.1. General version of Wilks' theorem for quantiles.
The error in the chi-squared approximation is given by

TABLE 4. with smoothing parameter

error	A.B.C. order
$O(n^{-1})$	$O(n^{-2})$
simple B.C. case	$o(n^{-1})$

?⟨default⟩?

16.2. Questions.

(1) Ka.Plan-Meier estimator ?
(2) bootstrap ?
(3) Vapnik-Cervonenkis theory ? (exponentially fast ?)

16.3. Summary.

ELTs (empirical likelihood theorems) $\neq$ Wilk's theorem !!

EL method combines the reliability of the nonparametric method with the flexibility and effectiveness of the likelihood approach.

The empirical likelihood is distinguished more by being a likelihood than by being empirical.

16.4. On the parametric method.

The vantages of the parametric method are that

(1) the effective estimators can be found;
(2) tests with good power properties can be constructed;
(3) its flexibility.

However, a problem with parametric inferences is that we might not know which parametric family to use and then

- the inefficiency is caused by such misspecification;
- the corresponding confidence intervals and tests can fail completely.

17. WORDS

(1) fascinating うっとりさせる
(2) engagingly 人を引きつける
(3) vantage 強み
(4) inherent 本来備わっている
(5) facilitate を容易にする
(6) in-depth 詳細な
(7) myriad 無数の
(8) indulge を満足させる
(9) instill in 徐々に教え込む
(10) at a time 一度に
(11) offset 相殺する
(12) hindsight あと知恵
(13) rudimentary 初歩的な
(14) encompass 含む
(15) commendably 称賛に値するように

18. FURTHER READING

(1) http://www.stanford.edu/ owen/empirical
(2) Jing Qin
(3) Jerry Lawless

(4) Yuichi Kitamura
(5) Mykland P. A. (1995) "Dual likelihood"
(6) Mykland P. A. (1999) "Bartelett identities and large deviations in likelihood theory"
(7) Hall and La Scala (1990)