中山大学学报（自然科学版）
第39卷 第5期 ACTA SCIENTIARUM NATURALIUM Vol. 39 No. 5
2000年 9月 UNIVERSITATIS SUNYATSENI Sep. 2000

文章编号：0529-6579（2000）05-0001-05

Bayesian Bootstrap for Empirical Processes Indexed by a Class of Functions*

ZHANG Di-xin[1]， LIN Hai-ming[2]
（1. Department of Statistics，Zhongshan University， Guangzhou 510275，China
2. Department of Mathematics，Guizhou Normal University，Guizhou 550001）

Abstract：Let $\mathscr{F}$ be a class of bounded or unbounded functions. It is shown that, for almost all sample sequences, Bayesian bootstrapped empirical processes（BBEP）indexed by a class of functions conditionally converge in distribution to a P-bridge process，i. e. BBEP and empirical processes over the same index class have the same limit distribution almost surely. Therefore, one can use BBEP to simulate empirical processes.

Keywords：Bayesian bootstrap； empirical process； conditional central limit theorem

CLC number：O212.7；O211.4 **Document Code**：A

1 Introduction

Let $\{X_n\}$ be a sequence of i. i. d. d-dimensional random vectors with a common probability measure P. Efron[1] introduces the “bootstrap”，a resampling method for approximating the distribution of statistics $H_n(X_1, \cdots, X_n, P)(n \geqslant 1)$. The bootstrap of Efron[1] has a Bayesian counterpart called the Bayesian bootstrap，which is due to Rubin[2]. Rubin's Bayesian bootstrap is based on simulation rather than resampling. This suggestive method has been validated with limit theorems for many particular statistics by Lo[3,4] and others. A great advantage of Rubin's Bayesian bootstrap approach is that the Monte Carlo method can be applied to simulate the approximated posterior distribution. The steps of the simulation are given by（1.1）~（1.3）in [3]. Lo[3] shows that，for empirical processes indexed by a class of indicator functions in $\mathbf{R}^1$， the Bayesian bootstrap[2] and the bootstrap[1] are first-order asymptotically equivalent for almost all sample sequences，they have almost surely the same limiting conditional distribution. Zhang et al[5] extend Lo[3]'s result to multivariate cases. Zhang et al[6] give the Bayesian bootstrap approximation for empirical processes indexed by a bounded class of functions in $\mathbf{R}^d$，$d \geqslant 1$. It is natural to ask whether Zhang et al[6]'s results hold for empirical processes indexed by an unbounded class of functions in $\mathbf{R}^d(d \geqslant 1)$. In this paper，we discuss this problem. The answer is affirmative.

Our theory relies on some important results of Pollard[7] on empirical processes. Our approach is parallel to the approach on empirical processes in Chapter Ⅱ and Ⅶ of Pollard[7]. However，there are significant differences between the two theories. Empirical processes are constructed from sums of independent random variables，whereas the Bayesian bootstrapped empirical processes（BBEP）do not have the forms. Another difference is that：for empirical processes，we are interested in non-conditional central limit theorem；for BBEP， we are interested in conditional central limit theorems on the given

* 基金项目：国家自然科学基金资助项目（19661001）；广东省高等教育厅自然科学基金资助项目（199989）
收稿日期：1999-10-07； 作者简介：张涤新（1954~），男，教授，博士，现广东商学院工作.

sample. The paper is divided into three parts. In Section 2, we obtain that, for almost all sample sequences, BBEP indexed by a class of functions conditionally converge in distribution to a P-bridge process, i. e. BBEP and empirical processes over the same index class have the same limit distribution almost surely. Therefore, one can use the BBEP to simulate the empirical processes. In Section 3, the idea of the symmetrization method introduced by Pollard[7] is developed. Some lemmas and the main results given in Section 2 are proven.

2 Main results

Assume that $X_1, \cdots, X_n, \cdots$ are i. i. d. random samples valued in $\mathbf{R}^d\ (d \geqslant 1)$ on a probability space $(\Omega_1, \mathscr{F}_1, P_1)$ with law P. Let $\{y_n\}$ be a sequences of independent standard exponential random variables on a probability space $(\Omega_2, \mathscr{F}_2, P_2)$ with common probability density function $f(x) = \exp\{-x\}$ if $x \geqslant 0$ and $f(x) = 0$ if $x < 0$. Furthermore, suppose that $\{X_n\}$, $\{y_n\}$ are mutually independent random sequences. Let $\Omega = \Omega_1 \times \Omega_2$, $P = P_1 \times P_2$ and $\mathscr{F} = \mathscr{F}_1 \times \mathscr{F}_2$. Thus, we obtain a probability space $(\Omega, \mathscr{F}, P)$. Denote $V_i = y_i / \sum_{i=1}^n y_i$, $i = 1, \cdots, n$, $n \geqslant 1$, then $(V_1, \cdots, V_{n-1})$ obeys Dirichlet distribution $D(1, \cdots, 1)$. Without loss of generality, we assume $V_i \geqslant 0$, $i \geqslant 1$. Let P_n be the empirical probability measure of $X_1, \cdots, X_n$ and $P_{n,v}$ be a probability measure that assigns the mass of V_i to X_i, $1 \leqslant i \leqslant n$. Write $\mathscr{F}$ for a class of Borel measurable real functions on $\mathbf{R}^d$. Let $Qf = \int f\,\mathrm{d}Q$, $L^i(Q) = \{f: \int |f|^i \mathrm{d}Q < \infty\}$ for any probability measure Q, $i = 1, 2$. Thus,

$$P_n f = 1/n \sum_{i=1}^n f(X_i), \qquad P_{n,v} f = \sum_{i=1}^n V_i f(X_i).$$

Definition 1 Let $D_{n,v} f = \sqrt{n}\,(P_{n,v} f - P_n f)$ for $f \in \mathscr{F}$. $\{D_{n,v} f: f \in \mathscr{F}\}$, $n \geqslant 1$, are called Bayesian bootstrapped empirical processes indexed by $\mathscr{F}$.

Our notation conforms in general with that of Pollard[7]. For instance, we say that a function F is an envelope of $\mathscr{F}$ if $\sup_{\mathscr{F}} |f(\circ)| \leqslant F(\circ)$ and so on.

Condition 1 $\mathscr{F}$ is permissible class of functions with an envelope $F > 0$ and $PF^2 < \infty$ (see p. 196 in [7] for "permissible").

Condition 2 The class of graphs of functions in $\mathscr{F}$ has polynomial discrimination (for "graph" and "polynomial discrimination" see p. 17~27 in [7]).

Suppose that $\mathscr{X}$ is the class of real functions on $\mathscr{F}$. Equip $\mathscr{X}$ with the metric generated by the uniform norm $\|\circ\|$, where $\|U\| = \sup_{f \in \mathscr{F}} |U(f)|$ to each $U \in \mathscr{X}$. Let $C(\mathscr{F}, P) = \{U \in \mathscr{X}: \|U\| < \infty$, for any $r > 0$, there exists $\delta > 0$, if $f, g \in \mathscr{F}$ and $(P(f-g)^2)^{1/2} < \delta$, then $|U(f) - U(g)| < r\}$. Throughout the paper, we suppose that B_P is a Gaussian random element of $\mathscr{X}$ with $\mathrm{Cov}(f, g) = P(fg) - PfPg$ for any $f, g \in \mathscr{F}$, whose sample paths all belong to $C(\mathscr{F}, P)$ and that B_P is a P-bridge process over $\mathscr{F}$ and is tight, i. e. the probability measure induced by B_P is tight (cf. p. 149~157 in [7]). Let $\boldsymbol{X} = (X_1, X_2, \cdots)$.

Theorem 1 Let $\mathscr{F}$ be a pointwise bounded, totally bounded, permissible class of functions. If for any $\varepsilon > 0$, $r > 0$, there exists a $\delta > 0$ such that, for almost all samples $\boldsymbol{X} = (X_1, X_2, \cdots)$,

$$\limsup_{n \to \infty} P\left(\sup_{(f,g) \in [\delta]} |D_{n,v}(f-g)| > \eta \,\middle|\, \boldsymbol{X}\right) \leqslant \varepsilon,$$

then, for almost all samples $\boldsymbol{X}$, $D_{n,v} \xrightarrow{\mathrm{d}} B_P \quad [P(\circ \mid \boldsymbol{X})]$.

Let "$\xrightarrow{\mathrm{d}}$" denote convergence in distribution.

Theorem 2　If $\mathscr{F}$ satisfies Conditions 1 and 2, then, for almost all samples $\boldsymbol{X}$,

$$1)\ D_{n,v} \xrightarrow{\mathrm{d}} B_P \quad [P(\circ \mid \boldsymbol{X})];\quad 2)\ \sup_{\mathscr{F}} |D_{n,v} f| \xrightarrow{\mathrm{d}} \sup_{\mathscr{F}} |B_P f| [P(\circ \mid \boldsymbol{X})].$$

3　Proofs of the main results and lemmas

In order to prove the main results, we require the following supporting lemmas. Let

$$P_{n,v^2} f = n \sum_{i=1}^{n} (V_i - 1/n)^2 f(X_i),\qquad \mathscr{F}_1 = \{(f-g)^2 : f, g \in \mathscr{F}\}.$$

Lemma 1　If $\mathscr{F}$ satisfies Conditions 1 and 2, then, for almost all samples $\boldsymbol{X}$,

$$\sup_{\mathscr{F}_1} |P_{n,v^2} h - P_n h| \to 0 \quad \text{a.s.}\ [P(\circ \mid \boldsymbol{X})] \tag{1}$$

Proof　Let $\boldsymbol{Y} = (y_1, y_2, \cdots)$ and $f(\boldsymbol{X}, \boldsymbol{Y}) = 1 - I_{A(\boldsymbol{X},\boldsymbol{Y})}$, where $A(\boldsymbol{X}, \boldsymbol{Y})$ denotes the set $\{\lim_{n\to\infty} \sup_{h\in\mathscr{F}_1} |P_{n,v^2} h - P_n h| = 0\}$ and $\{y_i\}$ is defined in Section 2. By Lemma 2[8], $\sup_{\mathscr{F}_1} |P_{n,v^2} h - P_n h| \to 0$ a.s. We therefore obtain that $0 = \mathrm{E}(f(\boldsymbol{X}, \boldsymbol{Y})) = \mathrm{E}(\mathrm{E}(f(\boldsymbol{X}, \boldsymbol{Y}) \mid \boldsymbol{X}))$. Condition 1 implies that $f(\boldsymbol{X}, \boldsymbol{Y})$ is a non-negative measurable function on the probability space $(\Omega, \mathscr{T}, P)$. See Appendix C[7] for more details. Since $\boldsymbol{X} = (X_1, X_2, \cdots)$ and $\boldsymbol{Y}$ are mutually independent, Fubini's theorem implies that, for almost all samples $\boldsymbol{X}$, $\mathrm{E}(f(\boldsymbol{X}, \boldsymbol{Y}), \boldsymbol{X}) = 0$. This implies that Lemma 1 holds.

Given a pseudo metric space $(\mathscr{F}, d)$, the covering number (see p. 143 in [7]) is defined by $N(\varepsilon, d, \mathscr{F}) = \min\{m : \text{there are } f_1, \cdots, f_m \in \mathscr{F} \text{ such that } \sup_{f\in\mathscr{F}} \min_{1\leqslant j\leqslant m} d(f, f_j) \leqslant \varepsilon\}$.

Let $d_2(f, g, P_{n,v^2}) = (P_{n,v^2}(f-g)^2)^{1/2}$ and $N_2(u, P_{n,v^2}, \mathscr{F})$ denote $N(u, d_2(\circ, \circ, P_{n,v^2}), \mathscr{F}))$. Define

$$J_2(\delta, P_{n,v^2}, \mathscr{F}) = \int_0^{\delta} [2\ln(N_2(u, P_{n,v^2}, \mathscr{F})^2/u)]^{1/2} \mathrm{d}u,\quad \delta > 0.$$

Lemma 2　If $\mathscr{F}$ satisfies the conditions 1 and 2, then, for any $\varepsilon > 0$ and $\eta > 0$, there exists a constant $\delta > 0$ such that, for almost all samples $\boldsymbol{X}$,

$$\limsup_{n\to\infty} P\left(\sup_{(f,g)\in[\delta]} |D_{n,v}(f-g)| > \eta \mid \boldsymbol{X}\right) \leqslant \varepsilon \tag{2}$$

where $[\delta] = \{(f, g) : f, g \in \mathscr{F}, d_2(f, g, P) := (P(f-g)^2)^{1/2} \leqslant \delta\},\ \delta > 0$.

Proof　Let $D_n = \sup_{(f,g)\in[\delta]} P_n(f-g)^2$. For any $\varepsilon \in (0, 1)$ and $\eta > 0$, choose δ satisfying $0 < \delta \leqslant \eta/6$. From Theorems Ⅱ.24 and Ⅱ.36 in [7],

$$\sup_{\mathscr{F}_1} |P_n h - Ph| \to 0 \qquad \text{a.s.} \quad [P] \tag{3}$$

where $\mathscr{F}_1 = \{(f-g)^2 : f, g \in \mathscr{F}\}$. Hence, for almost all samples $\boldsymbol{X}$, there exists $N_{\boldsymbol{X}} > 0$ such that, for any $n \geqslant N_{\boldsymbol{X}}$, $D_n \leqslant 2\delta^2$ and

$$\sup_{[\delta]} P\left(|D_{n,v}(f-g)| > \eta/2 \mid \boldsymbol{X}\right) \leqslant 4\eta^{-2} \sup_{[\delta]}\{1/n \sum_{i=1}^{n} (f(X_i) - g(X_i))^2 +$$
$$1/(n(n+1)) \sum_{i\neq j}^{n} |(f(X_i) - g(X_i))(f(X_j) - g(X_j))|\} \leqslant 8\eta^{-2} D_n \leqslant 1/2$$

Generate a sequence of i. i. d. sign random variables $\{\sigma_i\}$ with $P(\sigma_i = +1) = P(\sigma_i = -1) = 1/2$ which is independent of the sequences $\{X_i\}$ and $\{y_i\}$. Let $[\delta]_n = \{(f, g): f, g \in \mathscr{F}, d_2(f, g, P_{n,v^2}) \leqslant \delta\}$. From Lemma Ⅱ. 8[7] and the Second Symmetrization Method (cf. p. 14 in [7]), if $n \geqslant N_X$, then

$$P\left(\sup_{(f,g)\in[\delta]} |D_{n,v}(f-g)| > \eta \mid \boldsymbol{X}\right) \leqslant 4P\left(\sup_{[\delta]} |D^0_{n,v}(f-g)| > \eta/4 \mid \boldsymbol{X}\right) \tag{4}$$

where $D^0_{n,v}(f-g) = \sqrt{n}\sum_{i=1}^{n}\sigma_i(V_i - 1/n)(f(X_i) - g(X_i))$. Obviously,

$$P\left(\sup_{[\delta]} |D^0_{n,v}(f-g)| > \eta/4 \mid \boldsymbol{X}\right) \leqslant$$

$$P\left(\sup^{[2\delta]_n} |D^0_{n,v}(f-g)| > \eta/4 \mid \boldsymbol{X}\right) + P([\delta] - [2\delta]_n \mid \boldsymbol{X}) \tag{5}$$

By Lemma 1 and (3), as $n \to \infty$,

$$P([\delta] - [2\delta]_n \mid \boldsymbol{X}) \leqslant P\left(\sup_{[\delta]} |P_{n,v^2}(f-g)^2 - P(f-g)^2| > \delta^2/2 \mid \boldsymbol{X}\right) \to 0 \tag{6}$$

Let $V(n) = (V_1, \cdots, V_n)$. By Hoeffding inequality (cf. Appendix C[7]), for any $r > 0$ and $f, g \in \mathscr{F}$, if $\boldsymbol{X}$ and $V(n)$ are given,

$$P(|D^0_{n,v}(f-g)| > r d_2(f, g, P_{n,v^2}) \mid \boldsymbol{X}, V(n)) \leqslant 2\exp\{-r^2/2\} \tag{7}$$

Notice that $n\sum_{i=1}^{n}(V_i - 1/n)^2 \to 1$ a. s. $[P_2]$. For any $\varepsilon_1 \in (0, \varepsilon/8)$, there exists $N > 0$ and $S_2 \in \mathscr{F}_2$ not depending on the samples with $P_2(S_2) \geqslant 1 - \varepsilon_1$, if $n \geqslant N$,

$$0 < n\sum_{i=1}^{n}(V_i - 1/n)^2 \leqslant 2 \text{ on } S_2 \text{ uniformly} \tag{8}$$

Let $S = \Omega_1 \times S_2$. When (8) holds, we define a probability measure P^*_{n,v^2}, which puts the mass of $(V_i - 1/n)^2/\left[\sum_{i=1}^{n}(V_i - 1/n)^2\right]$ on X_i, $i = 1, \cdots, n$. Thus,

$$P^*_{n,v^2}(f-g)^2 = \frac{P_{n,v^2}(f-g)^2}{n\sum_{i=1}^{n}(V_i - 1/n)^2} \quad \text{for any } f, g \in \mathscr{F}$$

From Lemma Ⅱ. 36[7], there exist constants $A > 0$ and $W > 0$ such that, for any $n \geqslant N$ and each $u \in (0, 1)$,

$$N_2(2u, P_{n,v^2}, \mathscr{F}) \leqslant N_2(u, P^*_{n,v^2}, \mathscr{F}) \leqslant Au^{-W} \quad \text{uniformly on } S$$

This implies $J_2(\delta, P_{n,v^2}, \mathscr{F}) \to 0$ as $\delta \to 0$ on S uniformly. Hence, there exists a constant $\delta_0 \in (0, \min\{\varepsilon/32, \eta/6\})$, if $\delta \in (0, \delta_0]$, then $J_2(2\delta, P_{n,v^2}, \mathscr{F}) \leqslant \eta/208$ on S uniformly. Let $P_S(\circ) = P(\circ \cap S)/P(S)$ and let $n \geqslant N$. (7) and Chaining Lemma Ⅶ 9[7] imply

$$P_S\left(\sup_{[2\delta]_n} |D^0_{n,v}(f-g)| > 26J_2(2\delta, P_{n,v^2}, \mathscr{F}) \mid \boldsymbol{X}, V(n)\right) \leqslant 4\delta$$

By taking expectation over $V(n)$ for the last inequality, we can obtain that

$$P_S\left(\sup_{[2\delta]_n} |D^0_{n,v}(f-g)| > \eta/4 \mid \boldsymbol{X}\right) \leqslant 4\delta, \text{ and}$$

$$P\left(\sup_{[2\delta]_n} |D^0_{n,v}(f-g)| > \eta/4 \mid \boldsymbol{X}\right) \leqslant P_S\left(\sup_{[2\delta]_n} |D^0_{n,v}(f-g)| > \eta/4 \mid \boldsymbol{X}\right) + \varepsilon_1 \leqslant \varepsilon/4 \tag{9}$$

We deduce the result of the lemma by (4), (5), (6) and (9).

Lemma 3 Suppose that $\mathscr{F} \subset L^2(P)$ and that $f_1, \cdots, f_m$ belong to $\mathscr{F}$, where $m \geqslant 1$ is any positive integer. Then, for almost all samples $\boldsymbol{X}$,

$$(D_{n,v}f_1, \cdots, D_{n,v}f_m) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Sigma) \quad [P(\circ \mid \boldsymbol{X})]$$

where $\Sigma = (\sigma_{i,j})_{m\times m}$, $\sigma_{i,j} = P(f_if_j) - Pf_iPf_j$, $1 \leqslant i, j \leqslant m$.

Proof　For any $f \in \mathscr{F}$, Lo[3] shows that, for almost all samples $\boldsymbol{X}$,

$$D_{n,v}f \xrightarrow{\mathrm{d}} \mathrm{N}(0, Pf^2 - (Pf)^2) \quad [P(\circ \mid \boldsymbol{X})] \tag{10}$$

Notice that $D_{n,v}$ is a linear operator. The result of Lemma 3 follows from (10) and the properties of characteristic function.

Proof of Theorem 1　Replace E_n and P in Theorem Ⅶ21[7] with $D_{n,v}$ and $P(\circ \mid \boldsymbol{X})$ respectively. By Lemma 3, the rest of the proof is the same as that of Theorem Ⅶ21[7].

Proof of Theorem 2　Result 1) follows from Lemmas 2 and 3 and Theorem 1. Result 2) is obvious from result 1) and Theorem Ⅳ. 12[7].

References:

[1] EFRON B. Bootstrap methods: another look at the jackknife [J]. Ann Statist, 1979, 7: 1~26.

[2] RUBIN D B. The Bayesian bootstrap [J]. Ann Statist, 1981, 9: 130~134.

[3] LO A Y. A large sample study for Bayesian bootstrap [J]. Ann Statist, 1987, 15: 360~375.

[4] LO A Y. Bayesian bootstrap for censored data [J]. Ann Statist, 1993, 21: 100~123.

[5] ZHANG D, CHENG P. The goodness-of-fit test by using multivariate Random weighting method [J]. SystemsScience and Mathematical Sciences, 1995, 8: 355~363.

[6] ZHANG D, CHENG P. Random weighting approximation for empirical processes [J]. Chinese Science Bulletin, 1995, 40 (17): 1415~1418.

[7] POLLARD D. Convergence of Stochastic processes [M]. New York: Springer-Verlag, 1984.

[8] ZHANG D. Asymptotic properties for Dirichlet processes indexed by a class of functions [J]. Statistics & Probability Letters, 1997, 32: 25~33.

函数指标集上经验过程的贝叶斯自助方法

张涤新*，林海明

摘　要：设$\mathscr{F}$是一个有界或无界的函数指标集．对几乎所有的样本序列，函数指标集上贝叶斯自助经验过程（BBEP）条件地依分布收敛到一个 P-桥过程，即：在同一指标集上，BBEP 和经验过程几乎肯定地有相同的极限分布．因此能够使用 BBEP 去模拟经验过程．

关键词：贝叶斯自助方法；经验过程；条件中心极限定理

* 中山大学统计学系，广东 广州 510275