

THE STATISTICS ANALYSIS OF RANDOMLY TRIMMED MEANS AND THEIR BOOTSTRAP

LUO Kui¹, MA Xue-min², MA Zhi-wei³, ZHOU Xuan¹

(1. Industrial Training Centre, Shenzhen Polytechnic, Shenzhen 518055, China)

(2. School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China)

(3. Center for Chinese Entrepreneur Studies, Tsinghua University, Beijing 100084, China)

Abstract: In this paper, we consider studentized randomly trimmed means and their bootstrap under relaxed conditions. By using empirical process approach of CG, we obtain the studentized central limit theorem of randomly trimmed means whose asymptotic properties do not depend on the underlying density under relaxed conditions, which extend the results of randomly trimmed means by Chen and Gine [1].

Keywords: trimmed means; central limit theorems; bootstrap

2010 MR Subject Classification: 60F05; 62E20

Document code: A **Article ID:** 0255-7797(2015)02-0237-15

1 Introduction

It is well known that the sample mean is very sensitive to outliers in the data. As an alternative approach, randomly trimmed means have been studied widely in statistics. We can construct a randomly trimmed mean as follows: choose two levels $a_n < b_n$ depending on the sample and calculate the average of the data points between a_n and b_n . Above levels can be obtained by evaluating empirical distribution F_n of functions $a(F)$ and $b(F)$ where F is the probability distribution of X . If $a(F) = F^{-1}(\alpha/2)$ and $b(F) = F^{-1}(1 - \alpha/2)$, we can have the classical trimming. Let $a(F) = F^{-1}(\alpha/2)$ and $b(F) = F^{-1}(1 - \alpha/2)$, X is trimmed at level α . Let $a(F) = \mu(F) - cs(F)$ and $b(F) = \mu(F) + cs(F)$ where μ denotes median, s denotes median absolute deviation (MAD), and c is a constant of choice, we can have trimmed mean of Hampel [2].

Trimmed means are robust and can be bootstrapped. Hall and Padmanahban [3] study the bootstrap for the studentized classical trimmed mean. Shorack [4] provides a systematic study of bootstrapping to various L-statistics. In a recent paper, Chen and Gine [1] (hence CG, 2004) present a unified, empirical process based approach to the central limit

* Received date: 2014-04-13

Accepted date: 2014-06-04

Foundation item: Supported in part by the National Natural Science Foundation of China (31100958) and the Natural Science Foundation of Shenzhen Polytechnic (2213K3170021).

Biography: Luo Kui(1981-), female, born at Jingshan, Hubei, PH. D., major in financial mathematics.

Corresponding author: Ma Xuemin.

theorem and the bootstrap central limit theorem for the general trimmed mean under mild assumptions on the levels $a(F)$ and $b(F)$. Their basic assumption is that $\sqrt{n}(a_n - a)$ can be asymptotically linearized. It is asymptotically equivalent to $\sum_{i=1}^n (h_1(X_i) - Eh_1(X))/\sqrt{n}$ for some square integrable function h_1 for F . Similar conditions also hold for $\sqrt{n}(b_n - b)$.

Despite their generality, the central limit theorems of CG include a complicated expression for the asymptotic variance of trimmed means that depends on the underlying density of X . Following the empirical process approach of CG, we study studentized randomly trimmed means and their bootstrap under relaxed conditions. We can avoid the complex expression of variance in our central limit theorems. The results in this thesis may be used to obtain asymptotic confidence intervals and tests for trimmed means and also goes to the confidence intervals and tests to the usual means. We do not pursue that here.

This thesis is organized as follows. In Section 2, we introduce some definitions and major results of CG. Section 3 studies the asymptotic properties of studentized trimmed means. Section 4 verifies the validity of bootstrap. A simple example is introduced in Section 5.

2 Definitions and Existing Results

Before describing our findings, we will introduce some basic definitions in this section and will review some existing results.

Let $X, X_1, \dots, X_n, \dots$ be independent identically distributed real random variables with common probability law P and, for each $n \in N$, let $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ be the empirical measure corresponding to the first n observations X_1, \dots, X_n . F and F_n denote the cumulative distribution functions associated to P and $P_n(\omega)$ for all $n \in N$ and $\omega \in \Omega$. In addition, we assume $a < b$ and s.t. $F(b(P)) - F(a(P)-) \neq 0$ and $a \leq b$. The trimmed mean of P based on a and b is defined as

$$\theta = \theta(P) := \frac{\int_{a(P)}^{b(P)} x dF(x)}{F(b(P)) - F(a(P)-)} = E(X|X \in [a(P), b(P)]). \quad (2.1)$$

For convenience, we denote $a(P)$ and $b(P)$ as a and b in the following. And a_n and b_n are for $a(P_n)$ and $b(P_n)$. Similarly as (2.1), we can also define the empirical trimmed mean based on a and b as follows:

$$\theta_n = \theta(P_n) = \frac{\sum_{i=1}^n X_i I_{[a_n, b_n]}(X_i)}{\sum_{i=1}^n I_{[a_n, b_n]}(X_i)} = \frac{\int_{a_n}^{b_n} x dF_n(x)}{F_n(b_n) - F_n(a_n-)}. \quad (2.2)$$

Given $X, X_i, i \in N$ with c.d.f. F and density f . Let $P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ be the empirical measure and $\nu_n = \sqrt{n}(P_n - P)$ the empirical process. Let $-\infty < a < b < \infty$ and let a_n, b_n be random variables such that $-\infty < a_n \leq b_n < \infty$ a.s.. Following CG, we assume the following conditions:

(D.1) The c.d.f. F has a derivative f on an open set containing a and b and f is continuous there, hence, f is uniformly continuous on a compact set K whose interior contains a and b .

(D.2) $f(a) + f(b) \neq 0$.

(L) There exist measurable, P -square integrable functions h_1 and h_2 such that

$$\sqrt{n}(a_n - a) = \nu_n(h_1) + o_p(1) \quad \text{and} \quad \sqrt{n}(b_n - b) = \nu_n(h_2) + o_p(1). \quad (2.3)$$

Based on the above framework, CG proves the following lemma and theorem:

Lemma 1 Assume (D.1) and (L). Define, for $x \in R$,

$$g_1(x) = I_{[a,b]}(x) + f(b)h_2(x) - f(a)h_1(x), \quad (2.4)$$

$$g_2(x) = xI_{[a,b]}(x) + bf(b)h_2(x) - af(a)h_1(x), \quad (2.5)$$

Then

$$\sqrt{n} \left[\int_{a_n}^{b_n} dF_n(x) - \int_a^b dF(x) \right] = \nu_n(g_1) + o_p(1), \quad (2.6)$$

$$\sqrt{n} \left[\int_{a_n}^{b_n} x dF_n - \int_a^b x dF(x) \right] = \nu_n^b(g_2) + o_p(1). \quad (2.7)$$

Theorem 1 Assume (D.1), (D.2) and (L), and set

$$g(x) := \frac{1}{\int_a^b dF(t)} g_2(x) - \frac{\int_a^b t dF(t)}{(\int_a^b dF(t))^2} g_1(x), \quad x \in R$$

with g_1 and g_2 as defined in Lemma 1. Let θ_n be the trimmed mean based on a_n and b_n and let θ be its population counterpart, as defined in above. Then

$$\sqrt{n}(\theta_n - \theta) \rightarrow \sqrt{\text{Var}_F(g)}Z,$$

in distribution, where Z is standard normal.

Theorem 1 is a very general result. It includes many cases studied in previous literature like Hampel's means (Hampel [2, 5]), Symmetrically trimmed means (Huber [6]) and Kim's metrically trimmed means (Kim [7]). However, the variance of randomly trimmed means, $\text{Var}_F(g)$, in Theorem 1 has a very complicated form. Especially, it depends on the underlying density $f(a)$ and $f(b)$. With these constraints, Theorem 1 may be difficult to implement in some practice, for example, to construct tests of sample means. To overcome the above deficiency, we prove a studentized central limit theorem of randomly trimmed means whose asymptotic properties do not depend on the underlying density under relaxed conditions. Furthermore, the bootstrap is also valid in our case.

3 Asymptotics of Studentized Trimmed Means

In the similar spirit as Lemma 1, we prove the following lemma. It plays an important role in our proof of studentized central limit theorem in the next.

Lemma 2 Assume (D.1) and (L) hold. Define, for $x \in R$,

$$g_3(x) = x^2 I_{[a,b]}(x) + b^2 f(b) h_2(x) - a^2 f(a) h_1(x).$$

Then

$$\sqrt{n} \left(\int_{a_n}^{b_n} x^2 dF_n - \int_a^b x^2 dF \right) = \nu_n(g_3) + o_p(1). \quad (3.1)$$

Proof We can write the left side of (3.1) as follows:

$$\sqrt{n} \left[\int_{a_n}^{b_n} x^2 dF_n - \int_a^b x^2 dF \right] = \sqrt{n} \int_{a_n}^{b_n} x^2 d(F_n - F) + \sqrt{n} \int_{a_n}^a x^2 dF + \sqrt{n} \int_b^{b_n} x^2 dF.$$

Since x is on the compact set K , x^2 is bounded. Then we have that

$$\|x^2 I_{[c,d]}(x)\| \leq \|x^2\|_K := \sup_{x \in K} x^2$$

is finite. Also, there is $\delta_0 > 0$ such that $[a - \delta_0, a + \delta_0] \cup [b - \delta_0, b + \delta_0] \in K$. Therefore, by the asymptotic equicontinuity of the empirical process, see Dudley [8], we have

$$\lim_{\delta \rightarrow 0} \limsup_n \Pr \left\{ \sup_{|c-a| \leq \delta, |d-b| \leq \delta} |\nu_n(x^2 I_{[c,d]}(x)) - \nu_n(x^2 I_{[a,b]}(x))| \geq \varepsilon \right\} = 0$$

for all $\varepsilon > 0$. Since

$$\begin{aligned} & \Pr \left\{ \left| \sqrt{n} \int_{a_n}^{b_n} x^2 d(F_n - F) - \nu_n(x^2 I_{[a,b]}(x)) \right| \geq \varepsilon \right\} \\ & \leq \Pr \left\{ \sup_{|c-a| \leq \delta, |d-b| \leq \delta} |\nu_n(x^2 I_{[c,d]}(x)) - \nu_n(x^2 I_{[a,b]}(x))| \geq \varepsilon \right\} \\ & \quad + \Pr \{ |b_n - b| > \delta \} + \Pr \{ |a_n - a| > \delta \}, \end{aligned}$$

the equicontinuity condition and condition (L) give, upon taking limits first as $n \rightarrow \infty$ and then as $\delta \rightarrow 0$,

$$\sqrt{n} \int_{a_n}^{b_n} x^2 d(F_n - F) = \nu_n(x^2 I_{[a,b]}(x)) + o_p(1).$$

Given $0 < \delta \leq \delta_0$ and $M_n, n \in N$, such that $M_n \rightarrow \infty$ and $M_n/\sqrt{n} \rightarrow 0$, set

$$T_n := \sup \{ |f(b) - f(c)| : |b - c| \leq (\delta + M_n)/\sqrt{n} \},$$

which tends to zero by uniform continuity of f on K . Then, on the event where $|\nu_n(h_2) -$

$\sqrt{n}(b_n - b) \leq \delta$ and $|\nu_n(h_2)| \leq M_n$, we have

$$\begin{aligned}
& \left| \sqrt{n} \int_b^{b_n} x^2 dF - b^2 f(b) \nu_n(h_2) \right| \\
= & \left| \sqrt{n} \int_b^{b_n} x^2 (f(x) - f(b) + f(b)) dx - b^2 f(b) \nu_n(h_2) \right| \\
\leq & \left| \sqrt{n} \int_b^{b_n} x^2 |f(x) - f(b)| dx + \left| \sqrt{n} \int_b^{b_n} x^2 f(b) dx - b^2 f(b) \nu_n(h_2) \right| \right| \\
= & \sqrt{n} \int_b^{b_n} x^2 |f(x) - f(b)| dx + \left| \sqrt{n} \int_b^{b_n} x^2 dx - b^2 \nu_n(h_2) \right| f(b) \\
\leq & \sqrt{n} |b_n - b| \left(|b| + \frac{\delta + M_n}{\sqrt{n}} \right)^2 T_n + \left| \sqrt{n} \int_b^{b_n} x^2 dx - \sqrt{n} b^2 (b_n - b) \right. \\
& \left. + b^2 [\sqrt{n}(b_n - b) - \nu_n(h_2)] \right| f(b) \\
\leq & \sqrt{n} |b_n - b| \left(|b| + \frac{\delta + M_n}{\sqrt{n}} \right)^2 T_n + \left(\left| \sqrt{n} \int_b^{b_n} (x^2 - b^2) dx \right| \right. \\
& \left. + b^2 \left| \sqrt{n}(b_n - b) - \nu_n(h_2) \right| \right) f(b) \\
= & \sqrt{n} |b_n - b| \left(|b| + \frac{\delta + M_n}{\sqrt{n}} \right)^2 T_n + \left(\sqrt{n} \left| \frac{b_n^3 - b^3}{3} - b^2 (b_n - b) \right| \right. \\
& \left. + b^2 \left| \sqrt{n}(b_n - b) - \nu_n(h_2) \right| \right) f(b) \\
\leq & \sqrt{n} |b_n - b| \left(|b| + \frac{\delta + M_n}{\sqrt{n}} \right)^2 T_n + \left(\sqrt{n} \left| (b_n - b)(b_n^2 - b_n b + b^2) - b^2 (b_n - b) \right| \right. \\
& \left. + b^2 \left| \sqrt{n}(b_n - b) - \nu_n(h_2) \right| \right) f(b) \\
\leq & \sqrt{n} |b_n - b| \left(|b| + \frac{\delta + M_n}{\sqrt{n}} \right)^2 T_n + \left(\sqrt{n} \left| b_n - b \right| \left| b_n^2 + b_n b \right| \right. \\
& \left. + b^2 \left| \sqrt{n}(b_n - b) - \nu_n(h_2) \right| \right) f(b) \\
\leq & \sqrt{n} |b_n - b| \left(|b| + \frac{\delta + M_n}{\sqrt{n}} \right)^2 T_n + \left(\sqrt{n} \left| b_n - b \right| \left(\frac{\delta + M_n}{\sqrt{n}} \right)^2 + |b| \left(\frac{\delta + M_n}{\sqrt{n}} \right) \right) \\
& \left. + b^2 \left| \sqrt{n}(b_n - b) - \nu_n(h_2) \right| \right) f(b) \\
= & \sqrt{n} |b_n - b| b^2 T_n + \left| \sqrt{n}(b_n - b) - \nu_n(h_2) \right| b^2 f(b) + o_p(1).
\end{aligned}$$

We get the last equation above since $\frac{\delta + M_n}{\sqrt{n}} \rightarrow 0$. Based on the above results, we have the following equation, for any $\varepsilon > 0$:

$$\begin{aligned}
& \Pr \left\{ \left| \sqrt{n} \int_b^{b_n} x^2 dF - b^2 f(b) \nu_n(h_2) \right| > \varepsilon \right\} \\
\leq & \Pr \left\{ \sqrt{n} |b_n - b| b^2 T_n + \left| \sqrt{n}(b_n - b) - \nu_n(h_2) \right| b^2 f(b) > \varepsilon \right\} \\
& + \Pr \left\{ \left| \nu_n(h_2) - \sqrt{n}(b_n - b) \right| > \delta \right\} + \Pr \left\{ \left| \nu_n(h_2) \right| > M_n \right\} \rightarrow 0.
\end{aligned}$$

Then we have

$$\sqrt{n} \int_b^{b_n} x^2 dF = b^2 f(b) \nu_n(h_2) + o_p(1).$$

Similarly, we can prove

$$\sqrt{n} \int_{a_n}^a x^2 dF = -a^2 f(a) \nu_n(h_1) + o_p(1).$$

Then the lemma is proven.

Based on above lemma, we can prove the following result:

Lemma 3 Assume conditions (D.1), (D.2) and (L) hold. Denote

$$S_n^2 = \frac{\int_{a_n}^{b_n} (x - \theta_n)^2 dF_n}{\int_{a_n}^{b_n} dF_n} = \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_{a_n}^{b_n} dF_n} - (\theta_n)^2,$$

where $\theta_n = \frac{\int_{a_n}^{b_n} x dF_n}{\int_{a_n}^{b_n} dF_n}$. Then we have $S_n^2 = S^2 + o_p(1)$, where

$$S^2 = \frac{\int_a^b (t - \theta)^2 dF}{\int_a^b dF} = \frac{\int_a^b t^2 dF}{\int_a^b dF} - \frac{(\int_a^b t dF)^2}{(\int_a^b dF)^2}.$$

Proof

$$\begin{aligned} S_n^2 &= \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_{a_n}^{b_n} dF_n} - \frac{\int_{a_n}^{b_n} x dF_n}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n}^{b_n} x dF_n}{\int_{a_n}^{b_n} dF_n} \\ &= \frac{\int_a^b dF}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_a^b dF} - \frac{\int_a^b dF}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \frac{\int_a^b dF}{\int_{a_n}^{b_n} dF_n}. \end{aligned}$$

Since $\sqrt{n}[\int_{a_n}^{b_n} dF_n - \int_a^b dF] = \nu_n(g_1) + o_p(1)$,

$$\left| \frac{\int_{a_n}^{b_n} dF_n}{\int_a^b dF} - 1 \right| = \frac{1}{\int_a^b dF} \left| \int_{a_n}^{b_n} dF_n - \int_a^b dF \right| \rightarrow 0 \text{ in Pr.}$$

We can write $\frac{\int_a^b dF}{\int_{a_n}^{b_n} dF_n} = 1 + \tau_n$ where $\tau_n \rightarrow 0$ in probability. Then we have

$$\begin{aligned} S_n^2 &= \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_a^b dF} (1 + \tau_n) - \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} (1 + \tau_n)^2 \\ &= \left(\frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_a^b dF} - \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \right) + \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_a^b dF} \tau_n \\ &\quad - \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} 2\tau_n + \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \tau_n^2 \\ &= \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_a^b dF} - \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} \frac{\int_{a_n}^{b_n} x dF_n}{\int_a^b dF} + o_p(1). \end{aligned}$$

Based on Lemma 1 and Lemma 2, we have

$$\begin{aligned}
 S_n^2 &= \frac{\int_a^{b_n} x^2 dF_n}{\int_a^b dF} - \frac{\int_a^{b_n} x dF_n}{\int_a^b dF} \frac{\int_a^{b_n} x dF_n}{\int_a^b dF} + o_p(1) \\
 &= \frac{1}{\sqrt{n}} \frac{\sqrt{n} \int_a^b t^2 dF + \nu_n(g_3)}{\int_a^b dF} - \frac{1}{n} \frac{\sqrt{n} \int_a^b t dF + \nu_n(g_2)}{\int_a^b dF} \frac{\sqrt{n} \int_a^b t dF + \nu_n(g_2)}{\int_a^b dF} + o_p(1) \\
 &= \frac{\int_a^b t^2 dF}{\int_a^b dF} + \frac{1}{\sqrt{n}} \frac{\nu_n(g_3)}{\int_a^b dF} - \frac{1}{n} \left[\frac{n(\int_a^b t dF)^2 + 2\sqrt{n} \int_a^b t dF \cdot \nu_n(g_2) + \nu_n^2(g_2)}{(\int_a^b dF)^2} \right] + o_p(1) \\
 &= \frac{\int_a^b t^2 dF}{\int_a^b dF} - \frac{(\int_a^b t dF)^2}{(\int_a^b dF)^2} + \frac{1}{\sqrt{n}} \frac{\nu_n(g_3)}{\int_a^b dF} - \frac{2}{\sqrt{n}} \frac{\int_a^b t dF}{(\int_a^b dF)^2} \nu_n(g_2) - \frac{1}{n} \frac{\nu_n^2(g_2)}{(\int_a^b dF)^2} + o_p(1) \\
 &= \frac{\int_a^b t^2 dF}{\int_a^b dF} - \frac{(\int_a^b t dF)^2}{(\int_a^b dF)^2} + o_p(1) = S^2 + o_p(1).
 \end{aligned}$$

The lemma is proven.

Based on above lemmas, we can obtain the following theorem.

Theorem 2 Assume conditions (D.1), (D.2) and (LS) hold:

(LS) $\sqrt{n}(a_n - a) = o_p(1)$ and $\sqrt{n}(b_n - b) = o_p(1)$.

Define $\sigma_n^2 = \frac{S_n^2}{\int_a^{b_n} dF_n} = \frac{\int_a^{b_n} (x - \theta_n)^2 dF_n}{(\int_a^{b_n} dF_n)^2}$, we have the studentized trimmed mean $\frac{\sqrt{n}(\theta_n - \theta)}{\sigma_n} \rightarrow$

$N(0, 1)$ in distribution, where θ is the population randomly trimmed mean $\theta = \frac{\int_a^b t dF}{\int_a^b dF}$.

Proof According to conditions (L) and (LS), $h_1 = h_2 = 0$. We have

$$g_1(x) = I_{[a,b]}(x) + f(b)h_2(x) - f(a)h_1(x) = I_{[a,b]}(x)$$

and

$$g_2(x) = xI_{[a,b]}(x) + bf(b)h_2(x) - af(a)h_1(x) = xI_{[a,b]}(x).$$

Then

$$g(x) := \frac{1}{\int_a^b dF(t)} g_2(x) - \frac{\int_a^b t dF(t)}{(\int_a^b dF(t))^2} g_1(x) = \frac{xI_{[a,b]}(x)}{\int_a^b dF(t)} - \frac{\int_a^b t dF(t)}{(\int_a^b dF(t))^2} I_{[a,b]}(x). \tag{3.2}$$

Since $E_F(g) = \frac{\int_a^b t dF(t)}{\int_a^b dF(t)} - \frac{\int_a^b t dF(t)}{\int_a^b dF(t)} = 0$, we have

$$\begin{aligned}
 \text{Var}_F(g) &= E_F(g^2) - (E_F(g))^2 \\
 &= \frac{\int_a^b t^2 dF(t)}{(\int_a^b dF(t))^2} - 2 \frac{\int_a^b t dF(t)}{(\int_a^b dF(t))^3} \int_a^b t dF(t) + \frac{(\int_a^b t dF(t))^2}{(\int_a^b dF(t))^4} \int_a^b dF(t) \\
 &= \frac{\int_a^b t^2 dF(t)}{(\int_a^b dF(t))^2} - \frac{(\int_a^b t dF(t))^2}{(\int_a^b dF(t))^3} \\
 &= \frac{S^2}{\int_a^b dF(t)}.
 \end{aligned}$$

Since $S_n^2 = S^2 + o_p(1)$,

$$\begin{aligned}\text{Var}_F(g) &= \frac{S_n^2}{\int_a^b dF(t)} + o_p(1) = \frac{S_n^2}{\int_{a_n}^{b_n} dF_n(t)} \frac{\int_{a_n}^{b_n} dF_n(t)}{\int_a^b dF(t)} + o_p(1) \\ &= \sigma_n^2(1 + o_p(1)) + o_p(1) = \sigma_n^2 + o_p(1).\end{aligned}$$

According to Theorem 1, we have $\frac{\sqrt{n}(\theta_n - \theta)}{\sigma_n} \rightarrow N(0, 1)$ in distribution.

4 Bootstrap of Studentized Trimmed Means

In the above section, we study the asymptotic properties of studentized trimmed means. But with a small sample data, the estimators and tests based on those asymptotic properties may not be accurate. In this case, bootstrap may provide some efficient improvement. In the case of the mean, Hall [9] proved that if $Ex^3 < \infty$ then the bootstrap approximation is better than the normal. Although we do not know if this is the case for the trimmed mean, it is interesting to know that the studentized CLT in the trimmed mean can be bootstrapped.

With the sample X_1, \dots, X_n , we take n samples from it with replacement. We denote $X_{n,1}^b, \dots, X_{n,n}^b$ as the n th bootstrap sample. Following the notion of CG, we write F_n^b, P_n^b and ν_n^b respectively the empirical c.d.f, the empirical measure and the empirical process based on this sample: $F_n^b(A) = \frac{1}{n} \sum_{i=1}^n \delta_{X_{n,i}^b}(A)$, $F_n^b(x) = P_n^b(-\infty, x]$ and $\nu_n^b = \sqrt{n}(p_n^b - p_n)$. In addition, we denote by $\text{Pr}_b = \text{Pr}_b(\omega)$ the conditional probability given the sample X_1, \dots, X_n (its dependence on ω will not be displayed for convenience). Also, L^b will denote conditional law given the sample. The symbol $o_{p_b}(1)$ means the following: $V_n(X_{n,1}^b, X_{n,n}^b, X_1, \dots, X_n)$ is $o_{p_b}(1)$ a.s if a.s. $\text{Pr}_b\{|V_n| > \varepsilon\} \rightarrow 0$ for every $\varepsilon > 0$. This is also equivalent to: in almost every ω , every subsequence of $V_n(\omega, \omega')$ has a subsequence that converges $\omega' - \text{a.s.}$

(D.3) P has a density f on R , the set $B_f = \{f > 0\}$ is open and f is continuous on B_f .

CG proves a central limit theorem which can be considered as the bootstrap version of Theorem 1. Under the similar conditions, the bootstrap is also valid for the studentized trimmed means CLT in Theorem 3. To show this result, we need prove the following lemma at first.

Lemma 4 Assume (D.2), (D.3), (L), $a, b \in B_f$, $a_n \rightarrow a$ a.s. and $b_n \rightarrow b$ a.s.. Assume also that a_n^b and b_n^b , defined respectively as $a_n^b = a(F_n^b)$ and $a_n^b = a(F_n^b)$, satisfy

$$\sqrt{n}(a_n^b - a_n) = \nu_n^b(h_1) + o_{P_b}(1) \quad \text{and} \quad \sqrt{n}(b_n^b - b_n) = \nu_n^b(h_2) + o_{P_b}(1) \quad \text{a.s..}$$

Define, for $x \in R$,

$$\begin{aligned}g_1(x) &= I_{[a,b]}(x) + f(b)h_2(x) - f(a)h_1(x), \\ g_2(x) &= xI_{[a,b]}(x) + bf(b)h_2(x) - af(a)h_1(x), \\ g_3(x) &= x^2I_{[a,b]}(x) + b^2f(b)h_2(x) - a^2f(a)h_1(x).\end{aligned}$$

We have

$$\sqrt{n} \left(\int_{a_n^b}^{b_n^b} dF_n^b - \int_{a_n}^{b_n} dF_n \right) = \nu_n^b(g_1) + o_{p^b}(1) \text{ a.s.}, \tag{4.1}$$

$$\sqrt{n} \left(\int_{a_n^b}^{b_n^b} x dF_n^b - \int_{a_n}^{b_n} x dF_n \right) = \nu_n^b(g_2) + o_{p^b}(1) \text{ a.s.}, \tag{4.2}$$

$$\sqrt{n} \left(\int_{a_n^b}^{b_n^b} x^2 dF_n^b - \int_{a_n}^{b_n} x^2 dF_n \right) = \nu_n^b(g_3) + o_{p^b}(1) \text{ a.s.} \tag{4.3}$$

Proof We provide the complete proof for (4.1) in the following. (4.2) and (4.3) can be proven in the similar spirit. Let us denote

$$\sqrt{n} \left(\int_{a_n^b}^{b_n^b} dF_n^b - \int_{a_n}^{b_n} dF_n \right) = \sqrt{n} \left(\int_{a_n}^{b_n} d(F_n^b - F_n) + \int_{a_n^b}^{a_n} dF_n^b + \int_{b_n}^{b_n^b} dF_n^b \right). \tag{4.4}$$

According to Gine [10], the class of functions

$$\mathcal{F} = \{I_{[a,b]}(x), xI_{[a,b]}(x), x^2I_{[a,b]}(x) : -\infty < c_1 \leq a < b \leq c_2 < \infty\}$$

is uniform P-Donsker. We have

$$\{\nu_n^b(f) : f \in \mathcal{F}\} \rightarrow_{L^b} \{G_p(f) : f \in \mathcal{F}\} \text{ a.s.},$$

where \rightarrow_{L^b} denotes convergence in law conditionally on the sample and G_p is a centered Gaussian process. Then $\nu_n^b(I_{[a,b]}(x)) \rightarrow_{L^b} G_p(I_{[a,b]}(x))$.

Since

$$\sqrt{n} \int_{a_n}^{b_n} d(F_n^b - F_n) = \nu_n^b(I_{[a_n, b_n]}(x)) = \nu_n^b(I_{[a_n, b_n]}(x)) - \nu_n^b(I_{[a, b]}(x)) + \nu_n^b(I_{[a, b]}(x)).$$

To prove $\sqrt{n} \int_{a_n}^{b_n} d(F_n^b - F_n) \rightarrow_{L^b} G_p(I_{[a, b]}(x))$ a.s., we only need show

$$\nu_n^b(I_{[a_n, b_n]}(x)) - \nu_n^b(I_{[a, b]}(x)) \rightarrow 0 \text{ in } p^b \text{ a.s.} \tag{4.5}$$

According to Gine [10],

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} Pr_b \left\{ \sup_{g, h \in \mathcal{F}, E_p(g-h)^2 \leq \delta} |\nu_n^b(g) - \nu_n^b(h)| > \varepsilon \right\} = 0 \quad \forall \varepsilon > 0 \text{ a.s.}, \tag{4.6}$$

where the class of functions \mathcal{F} is uniform P-Donsker. In addition,

$$\lim_{n \rightarrow \infty} E_p(I_{[a_n, b_n]}(x) - I_{[a, b]}(x))^2 = 0. \tag{4.7}$$

Based on (4.6) and (4.7), we have

$$\lim_{n \rightarrow \infty} Pr_b \left\{ |\nu_n^b(I_{[a_n, b_n]}(x)) - \nu_n^b(I_{[a, b]}(x))| > \varepsilon \right\} = 0 \quad \forall \varepsilon > 0 \text{ a.s.}$$

Then (4.5) is proven.

Next, we will prove

$$\sqrt{n} \int_{b_n}^{b_n^b} dF_n^b - f(b)\nu_n^b(h_2) = o_{p^b}(1) \text{ a.s..} \quad (4.8)$$

We can write

$$\begin{aligned} & \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n^b - f(b)\nu_n^b(h_2) \right| \\ = & \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n^b - \sqrt{n} \int_{b_n}^{b_n^b} dF_n + \sqrt{n} \int_{b_n}^{b_n^b} dF_n - \sqrt{n} \int_{b_n}^{b_n^b} dF + \sqrt{n} \int_{b_n}^{b_n^b} dF - f(b)\nu_n^b(h_2) \right| \\ \leq & \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n^b - \sqrt{n} \int_{b_n}^{b_n^b} dF_n \right| + \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n - \sqrt{n} \int_{b_n}^{b_n^b} dF \right| \\ & + \left| \sqrt{n} \int_{b_n}^{b_n^b} dF - f(b)\nu_n^b(h_2) \right| \\ = & \gamma_1 + \gamma_2 + \gamma_3. \end{aligned}$$

In the following, we will prove the convergence of above three terms separately. At first, we can write

$$\begin{aligned} \gamma_1 &= \left| \nu_n^b(-\infty, b_n^b) - \nu_n^b(-\infty, b_n) \right| \\ &\leq \sup_{|b_n - \lambda| < \delta} \left| \nu_n^b(-\infty, \lambda) - \nu_n^b(-\infty, b_n) \right| \text{ if } |b_n^b - b_n| < \delta. \end{aligned}$$

Then given ε ,

$$\begin{aligned} \Pr_b \left\{ \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n^b - \sqrt{n} \int_{b_n}^{b_n^b} dF_n \right| > \varepsilon \right\} &\leq \Pr_b \left\{ \sup_{|s-t| < \delta} \left| \nu_n^b(-\infty, s) - \nu_n^b(-\infty, t) \right| > \varepsilon \right\} \\ &+ \Pr_b \{ |b_n^b - b_n| > \delta \} \text{ for all } \delta. \end{aligned} \quad (4.9)$$

According to conditions in the lemma, we know that $\sqrt{n}(b_n^b - b_n) - v_n^b(h_2) \rightarrow 0$ in \Pr_b a.s. and we also know $\nu_n^b(h_2) \rightarrow_{d^b} N(0, \text{Var}_p(h_2))$. So, we have

$$\Pr_b(\sqrt{n}|b_n^b - b_n| > M_n) \rightarrow 0 \text{ a.s. for any } M_n \rightarrow \infty.$$

Let $M_n = \frac{\sqrt{n}}{(\lg \lg n)^2} \rightarrow \infty$, $\Pr_b \left(\sqrt{n}|b_n^b - b_n| > \frac{\sqrt{n}}{(\lg \lg n)^2} \right) \rightarrow 0$ a.s.. Then

$$\Pr_b \left(|b_n^b - b_n| > \frac{1}{(\lg \lg n)^2} \right) \rightarrow 0 \text{ a.s..}$$

We have $b_n^b - b_n \rightarrow 0$ in p^b a.s.. Similarly, we can prove $a_n^b - a_n \rightarrow 0$ in p^b a.s.. Then we have $a_n \rightarrow a$ and $b_n \rightarrow b$ in p^b a.s.. According to above equations, $\lim_{n \rightarrow \infty} \Pr_b \{ |b_n^b - b_n| > \delta \} \rightarrow 0$

a.s. for any $\delta > 0$. Since $E_p(I_{(-\infty,s)} - I_{(-\infty,t)})^2 \leq |s - t| \sup_{x \in K} f(x)$, we have the following equation according to (4.6),

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \Pr\left\{ \sup_b \sup_{|s-t| < \delta} |\nu_n^b(-\infty, s] - \nu_n^b(-\infty, t]| > \varepsilon \right\} \rightarrow 0.$$

Then taking limits first as $n \rightarrow \infty$ and then as $\delta \rightarrow 0$, we can obtain the following result based on (4.9):

$$\gamma_1 = \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n^b - \sqrt{n} \int_{b_n}^{b_n^b} dF_n \right| = o_{P^b}(1) \text{ a.s..} \tag{4.10}$$

Next, we focus on γ_2 .

$$\begin{aligned} & \Pr_b \left\{ \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n - \sqrt{n} \int_{b_n}^{b_n^b} dF \right| > \varepsilon \right\} \\ &= \Pr_b \left\{ \left| \nu_n(-\infty, b_n] - \nu_n(-\infty, b_n^b) \right| > \varepsilon \right\} \\ &\leq \Pr_b \left\{ \sup_{|b_n - \lambda| < \frac{1}{(\log \log n)^2}} \left| \nu_n(-\infty, b_n] - \nu_n(-\infty, \lambda) \right| > \varepsilon \right\} + Pr_b \left\{ |b_n - b_n^b| > \frac{1}{(\log \log n)^2} \right\} \\ &= \xi_1 + \xi_2. \end{aligned} \tag{4.11}$$

$\xi_2 \rightarrow 0$ in P^b has been proven before. $\xi_1 \rightarrow 0$ in P^b can be validated by the following theorem, see Gine [10].

Since the functions

$$\mathcal{F} = \{I[a, b](x), I[-\infty, a](x), xI_{[a, b]}(x) : -\infty < c_1 \leq a < b \leq c_2 < \infty\}$$

is uniform P -Donsker, we have

$$\lim_{n \rightarrow \infty} \sup_{g, h \in \mathcal{F}; E_p(g-h)^2 < 1/(\log \log n)^2} |\nu_n(g) - \nu_n(h)| = 0 \text{ a.s..}$$

Then we have

$$\gamma_2 = \left| \sqrt{n} \int_{b_n}^{b_n^b} dF_n - \sqrt{n} \int_{b_n}^{b_n^b} dF \right| = o_{P^b}(1) \text{ a.s..} \tag{4.12}$$

At last, we focus on γ_3 :

$$\begin{aligned} \gamma_3 &= \left| \sqrt{n} \left(\int_{b_n}^{b_n^b} dF \right) - f(b) \nu_n^b(h_2) \right| \\ &= \left| \sqrt{n}(F(b_n^b) - F(b_n)) - f(b) \nu_n^b(h_2) \right| \\ &= \left| \sqrt{n}(F(b_n^b) - F(b_n)) - f(b) \sqrt{n}(b_n^b - b_n) + f(b) \cdot o_{P^b}(1) \right|. \end{aligned}$$

Since $f(b)$ is bounded, $\gamma_3 = \left| \sqrt{n}(F(b_n^b) - F(b_n)) - f(b) \sqrt{n}(b_n^b - b_n) \right| + o_{P^b}(1)$. Recall the set we defined before. b_n^b will locate in K as n becomes big enough with $Pr_b \rightarrow 1$ a.s.. Then there exists a point denoted η_n between b_n^b and b_n such that

$$\begin{aligned} \gamma_3 &= \left| \sqrt{n}(b_n^b - b_n) f(\eta_n) - f(b) \sqrt{n}(b_n^b - b_n) \right| + o_{P^b}(1) \\ &= \left| \sqrt{n}(b_n^b - b_n) (f(\eta_n) - f(b)) \right| + o_{P^b}(1) \\ &= (\nu_n^b(h_2) + o_{P^b}(1)) |f(\eta_n) - f(b)| + o_{P^b}(1). \end{aligned}$$

Since f is continuous, $|f(\eta_n) - f(b)| \rightarrow 0$ in P^b a.s.. And $\nu_n^b(h_2)$ is $O_{P^b}(1)$ a.s.. We have

$$\gamma_3 = o_{P^b}(1) \quad \text{a.s..} \tag{4.13}$$

Based on (4.9), (4.12) and (4.13), we have

$$\sqrt{n} \int_{b_n}^{b_n^b} dF_n^b - f(b)\nu_n^b(h_2) = o_{P^b}(1) \quad \text{a.s..} \tag{4.14}$$

Similarly, we can prove

$$\sqrt{n} \int_{a_n^b}^{a_n} dF_n^b + f(a)\nu_n^b(h_1) = o_{P^b}(1) \quad \text{a.s..} \tag{4.15}$$

Combine above results, we can prove (4.1). In the same spirit, we can also prove (4.2) and (4.3). Finally, Lemma 4 is proven.

Based on above lemma, we can prove that following theorem.

Theorem 3 Assume conditions (D.2), (D.3) and (LS) hold. Suppose $a, b \in B_f$, $a_n \rightarrow a$ a.s. and $b_n \rightarrow b$ a.s.. Assume also that a_n^b and b_n^b , defined respectively as $a_n^b = a(F_n^b)$ and $b_n^b = b(F_n^b)$, satisfy

$$(Lb) \quad \sqrt{n}(a_n^b - a_n) = o_{P^b}(1), \quad \sqrt{n}(b_n^b - b_n) = o_{P^b}(1) \quad \text{a.s..}$$

Define θ_n^b as the bootstrap trimmed mean, $\theta_n^b = \frac{\sum_{i=1}^n X_{n,i}^b I_{[a_n^b, b_n^b]}(X_{n,i}^b)}{\sum_{i=1}^n I_{[a_n^b, b_n^b]}(X_{n,i}^b)}$, $n \in N$, where θ_n

and θ are as in Theorem 3. Define

$$(S_n^b)^2 = \frac{\sum_{i=1}^n (x_{n,i}^b - \theta_n^b)^2 I_{[a_n^b, b_n^b]}(x_{n,i}^b)}{\sum_{i=1}^n I_{[a_n^b, b_n^b]}(x_{n,i}^b)}.$$

Denote $(\sigma_n^b)^2 = \frac{(S_n^b)^2}{\int_{a_n^b}^{b_n^b} dF_n^b}$. We have

$$\frac{\sqrt{n}(\theta_n^b - \theta_n)}{\sigma_n^b} \rightarrow_{L^b} N(0, 1) \quad \text{a.s..} \tag{4.16}$$

Proof In the proof in above section, we have shown $\text{Var}_F(g) = \frac{S^2}{\int_a^b dF(t)}$ where $S^2 = \frac{\int_a^b (x-\theta)^2 dF}{\int_a^b dF}$ and $\theta = \frac{\int_a^b x dF}{\int_a^b dF}$. Denote $\sigma^2 = \frac{S^2}{\int_a^b dF(t)}$. We have

$$\frac{\sqrt{n}(\theta_n^b - \theta_n)}{\sigma_n^b} = \frac{\sqrt{n}(\theta_n^b - \theta_n)}{\sqrt{\text{Var}_F(g)}} \times \frac{\sqrt{\text{Var}_F(g)}}{\sigma} \times \frac{\sigma}{\sigma_n^b} = \frac{\sqrt{n}(\theta_n^b - \theta_n)}{\sqrt{\text{Var}_F(g)}} \times \frac{\sigma}{\sigma_n^b}. \tag{4.17}$$

Under more genl conditions, CG has proven

$$\sqrt{n}(\theta_n^b - \theta_n) \rightarrow_{L^b} \sqrt{\text{Var}_F(g)}Z. \tag{4.18}$$

To prove (4.16), we only need show $\sigma_n^b \rightarrow_{Pr^b} \sigma$ a.s.. Based on Lemma 4, we have

$$\begin{aligned} (S_n^b)^2 &= \frac{\int_{a_n^b}^{b_n^b} x^2 dF_n^b}{\int_{a_n^b}^{b_n^b} dF_n^b} - \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n^b}^{b_n^b} dF_n^b} \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n^b}^{b_n^b} dF_n^b} \\ &= \frac{\int_{a_n}^{b_n} dF_n}{\int_{a_n^b}^{b_n^b} dF_n^b} \frac{\int_{a_n}^{b_n} x^2 dF_n^b}{\int_{a_n}^{b_n} dF_n} - \frac{\int_{a_n}^{b_n} dF_n}{\int_{a_n^b}^{b_n^b} dF_n^b} \frac{\int_{a_n}^{b_n} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n}^{b_n} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n}^{b_n} dF_n}{\int_{a_n^b}^{b_n^b} dF_n^b}. \end{aligned} \quad (4.19)$$

According to (4.1), $\int_{a_n^b}^{b_n^b} dF_n^b - \int_{a_n}^{b_n} dF_n = \frac{1}{\sqrt{n}}(\nu_n^b(g_1)) + o_{P^b}(1) = o_{P^b}(1)$ a.s., where $\nu_n^b(g_1)$ is bounded in P^b . We can write

$$\frac{\int_{a_n}^{b_n} dF_n}{\int_{a_n^b}^{b_n^b} dF_n^b} = 1 + \tau_n, \quad (4.20)$$

where $\tau_n \rightarrow 0$ in P^b a.s.. Substitute (4.20) to (4.19):

$$\begin{aligned} (S_n^b)^2 &= \frac{\int_{a_n^b}^{b_n^b} x^2 dF_n^b}{\int_{a_n}^{b_n} dF_n} (1 + \tau_n) - \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} (1 + \tau_n)^2 \\ &= \left(\frac{\int_{a_n^b}^{b_n^b} x^2 dF_n^b}{\int_{a_n}^{b_n} dF_n} - \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \right) + \frac{\int_{a_n^b}^{b_n^b} x^2 dF_n^b}{\int_{a_n}^{b_n} dF_n} \tau_n \\ &\quad - \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} 2\tau_n + \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \tau_n^2 \\ &= \frac{\int_{a_n^b}^{b_n^b} x^2 dF_n^b}{\int_{a_n}^{b_n} dF_n} - \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} + o_{P^b}(1) \text{ a.s..} \end{aligned}$$

Based on Lemma 4, we have

$$\begin{aligned} (S_n^b)^2 &= \frac{\int_{a_n^b}^{b_n^b} x^2 dF_n^b}{\int_{a_n}^{b_n} dF_n} - \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n^b}^{b_n^b} x dF_n^b}{\int_{a_n}^{b_n} dF_n} + o_{P^b}(1) \text{ a.s.} \\ &= \frac{1}{\sqrt{n}} \frac{\int_{a_n}^{b_n} x^2 dF_n + \nu_n^b(g_3)}{\int_{a_n}^{b_n} dF_n} - \frac{1}{n} \frac{(\sqrt{n} \int_{a_n}^{b_n} x dF_n + \nu_n^b(g_2))^2}{\int_{a_n}^{b_n} dF_n \int_{a_n}^{b_n} dF_n} + o_{P^b}(1) \text{ a.s.} \\ &= \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_{a_n}^{b_n} dF_n} - \left(\frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_{a_n}^{b_n} dF_n} \right)^2 + \frac{1}{\sqrt{n}} \frac{\nu_n^b(g_3)}{\int_{a_n}^{b_n} dF_n} - \frac{2}{\sqrt{n}} \frac{\int_{a_n}^{b_n} x dF_n}{(\int_{a_n}^{b_n} dF_n)^2} \nu_n^b(g_2) \\ &\quad - \frac{1}{n} \frac{(\nu_n^b(g_2))^2}{(\int_{a_n}^{b_n} dF_n)^2} + o_{P^b}(1) \text{ a.s.} \\ &= \frac{\int_{a_n}^{b_n} x^2 dF_n}{\int_{a_n}^{b_n} dF_n} - \left(\frac{\int_{a_n}^{b_n} x dF_n}{\int_{a_n}^{b_n} dF_n} \right)^2 + o_{P^b}(1) \text{ a.s.} \\ &= S_n^2 + o_{P^b}(1) \text{ a.s..} \end{aligned}$$

Under the assumption $a_n \rightarrow a$ a.s. and $b_n \rightarrow b$ a.s., we can prove $S_n^2 \rightarrow S^2$ a.s. in the similar spirit as Lemma 3. Then we have $(S_n^b)^2 = S^2 + o_{P^b}(1)$ a.s.. We can also prove $\int_{a_n}^{b_n} dF_n \rightarrow \int_a^b dF$ a.s..

Based on the definition of σ_n^b , we have

$$(\sigma_n^b)^2 = \frac{(S_n^b)^2}{\int_{a_n}^{b_n} dF_n^b} = \frac{S^2}{\int_a^b dF} \frac{\int_a^b dF}{\int_{a_n}^{b_n} dF_n} \frac{\int_{a_n}^{b_n} dF_n}{\int_{a_n}^{b_n} dF_n^b} = \sigma^2 + o_{P^b}(1) \quad \text{a.s..}$$

Then the theorem is proven.

5 An Example

To examine the application of above theorems, we consider the following simple example in order statistics. Let $X, X_1, \dots, X_n, \dots$ be independent identically distributed real random variables on support $[a, b]$. Suppose f is their density function, with $f(X) \neq 0, X \in [a, b]$. Let $a_n = X_{(j_n)}, b_n = X_{(n-j_n)}$. Then, if $j_n = o(\sqrt{n})$, we have

$$\sqrt{n}(a_n - a) \rightarrow 0, \quad \sqrt{n}(b_n - b) \rightarrow 0. \quad (5.1)$$

To prove above equations, we may assume, without loss of generality that $a = 0$. $F(X)$ is uniform on $[0, 1]$ and $(F(X))_{(j_n)} = F(X_{(j_n)})$ by monotonicity. Also,

$$P(X_{(j_n)} > \frac{\varepsilon}{\sqrt{n}}) = P(F(X_{(j_n)}) > F(\frac{\varepsilon}{\sqrt{n}}))$$

and $F(\frac{\varepsilon}{\sqrt{n}}) \approx f(0)\frac{\varepsilon}{\sqrt{n}}$. Hence we may also assume that the law of X is uniform on $[0, 1]$. It is well known that the law of $X_{(j_n)}$ in this case is $L(X_{(j_n)}) = L(\frac{w_1 + \dots + w_{j_n}}{w_1 + \dots + w_n})$, where w_i are i.i.d exponential with $\lambda = 1$ (e.g. Breiman [11]). Then

$$P\left(X_{(j_n)} > \frac{\varepsilon}{\sqrt{n}}\right) = P\left(\frac{w_1 + \dots + w_{j_n} - j_n}{\sqrt{j_n}} > \frac{\varepsilon}{\sqrt{n}j_n}(w_1 + \dots + w_n - n) + \frac{\varepsilon\sqrt{n}}{\sqrt{j_n}} - \sqrt{j_n}\right).$$

and assuming $j_n \rightarrow \infty$, by the central limit theorem, this probability tends to zero if and only if $\frac{\varepsilon\sqrt{n}}{\sqrt{j_n}} - \sqrt{j_n} \rightarrow \infty$, which happens in all $\varepsilon > 0$ if and only if $j_n = o(\sqrt{n})$.

In this case, we can define the trimmed mean

$$\theta_n = \frac{\sum_{i=1}^n X_i I(X_{(j_n+1)} \leq X_i \leq X_{(n-j_n)})}{\sum_{i=1}^n I(X_{(j_n+1)} \leq X_i \leq X_{(n-j_n)})} = \frac{\sum_{j=j_n+1}^{n-j_n} X_{(j)}}{n - 2j_n},$$

which is the average of the data with the smallest j_n and the largest j_n data removed (outliers). Let $\theta(P) = E(X)$. According to Theorem 3, we have

$$\frac{\sqrt{n}(\theta_n - E(X))}{\sigma_n} \rightarrow_d N(0, 1), \quad \text{where} \quad \sigma_n^2 = \frac{\sum_{j=j_n+1}^{n-j_n} (X_{(j)} - \theta_n)^2}{(n - 2j_n)^2}.$$

References

- [1] Chen Z, Giné. Another approach to asymptotics and bootstrap of randomly trimmed means[J]. *Evarist Annals Institute of Stat. Math.*, 2004, 56(4): 771–790.
- [2] Hampel F. A general qualitative definition of robustness[J]. *Annals Math. Stat.*, 1971, 42(6): 1887–1896.
- [3] Hall P, Padmanabhan A R. On the bootstrap and the trimmed mean[J]. *J. Multi. Anal.*, 1992, 41(1): 132–153.
- [4] Shorack G R. Uniform CLT, WLLN, LIL and bootstrapping in a data analytic approach to trimmed L-statistics[J]. *J. Stat. Plan. Infer.*, 1997, 60(1): 1–44.
- [5] Hampel F. The breakdown points of the mena combined with some rejection rules[J]. *Technometrics*, 1985, 27(2): 95–107.
- [6] Seong Ju Kim. The metrically trimmed mean as a measure of location[J]. *Annals Stat*, 1992, 20(3): 1534–1547.
- [7] Huber P. *Robust statistics*[M]. New York: Wiley, 1981.
- [8] Dudley R. *Uuiform central limit theorems*[M]. New York: Cambridge University Press, 1999.
- [9] Hall P. Rate of convergence in bootstrap approximations[J]. *The Annals of Probability*, 1988, 16(4): 1665–1684.
- [10] Gine E. *Lectures on some aspects of the bootstrap*[J]. *Lecture Notes in MathEmatics*, 1996: 37–151.
- [11] Breiman L. *Probablity*[M]. New Jersey: Addison-Wesley, 1968.

随机切尾均值及其自举的统计分析

罗 葵¹, 马学敏², 马志伟³, 周 旋¹

(1. 深圳职业技术学院工业中心, 广东 深圳 518055)

(2. 武汉大学数学与统计学院, 湖北 武汉 430072)

(3. 清华大学中国企业研究中心, 北京 100084)

摘要: 本文研究了松弛条件下学习随机切尾均值及其自举的统计性质. 利用CG的经验过程方法, 本文证明了在松弛条件下, 学习随机切尾均值的中心极限定理, 其渐进性质不依赖密度函数. 进一步, 得到了该性质对于学习切尾均值的自举依然成立, 推广了Chen和Gine^[1]随机切尾均值的相关研究结果.

关键词: 切尾均值; 中心极限定理; 自举

MR(2010)主题分类号: 60F05; 62E20 中图分类号: O211.4