YOKOHAMA MATHEMATICAL
JOURNAL VOL. 42, 1994

TRIMMED L-STATISTICS BASED ON WEAKLY DEPENDENT RANDOM VARIABLES

By

K. YOSHIHARA and S. KANAGAWA

(Received May 6, 1994)

Abstract. A relation between the trimming size of trimmed L-statistics and their limit distributions is considered for strictly stationary sequence of random variables satisfying the ϕ-mixing condition or the strong mixing condition.

1. Introduction and results

Let $\xi_{n:1}\leqq\xi_{n:2}\leqq\cdots\leqq\xi_{n:n}$ be the order statistics corresponding to some stationary sequence of random variables $\{\xi_1, \xi_2, \cdots, \xi_n\}$ with the uniform distribution over (0, 1). An L-statistic is defined by

$$T:=\frac{1}{n}\sum_{i=1}^{n}c_{ni}g(\xi_{n:i}),$$

where, $\{c_{ni}, 1\leqq i\leqq n\}$ is a triangular array of positive constants and g is a left continuous and nondecreasing function on (0, 1). L-statistics play an important role in the theory of statistical inference. (See Chapter 7 in [5] for more details.) Besides it is well known that the rate of convergence in the central limit theorem can be improved by trimming extreme random variables (see e.g. [1] and [4]). Furthermore the trimming method gives some advantages to the central limit theorem in addition to the rate of convergence. However the limit distribution differs from the original normal distribution or the central limit theorem does not hold while the trimming size is so large. In this paper, according to [3], we consider a weak convergence of some trimmed L-statistics for weakly dependent random variables, here three types of trimming size are treated; trimming fixed numbers, trimming fixed fractions and trimming vanishing fractions.

Let $\{\xi_j, j\geqq 1\}$ be a strictly stationary sequence of random variables with uniform distribution over (0, 1) satisfying some mixing condition as follows. Denote $\mathcal{M}_a^b$ the σ-algebra generated by $\xi_a, \cdots, \xi_b$, $1\leqq a\leqq b\leqq\infty$. We say that

1991 AMS Subject Classification: Primary 62G30, Secondary 60F05.
Key words and phrase: L-statistics, trimming, ϕ-mixing, strong mixing.

$\{\xi_j\}$ is ϕ-mixing if as $n\to\infty$

$$(1)\qquad \phi(n):=\sup_{k\geq 1}\sup_{A\in\mathcal{M}_1^k,\,B\in\mathcal{M}_{k+n}^\infty}\left|\frac{\boldsymbol{P}(AB)-\boldsymbol{P}(A)\boldsymbol{P}(B)}{\boldsymbol{P}(A)}\right|\downarrow 0.$$

Furthermore $\{\xi_j\}$ is said to be strong mixing if as $n\to\infty$

$$(2)\qquad \alpha(n):=\sup_{k\geq 1}\sup_{A\in\mathcal{M}_1^k,\,B\in\mathcal{M}_{k+n}^\infty}|\boldsymbol{P}(AB)-\boldsymbol{P}(A)\boldsymbol{P}(B)|\downarrow 0.$$

It is easy to see that $\alpha(n)\leqq\phi(n)$ for each n (e.g. [6]).

Consider the trimmed L-statistic defined by

$$T_n:=\frac{1}{n}\sum_{i=k_n+1}^{m_n}c_{ni}g(\xi_{n:i}),$$

where k_n and m_n are nonnegative integers with $0\leqq k_n<m_n\leqq n$. As for $\{c_{ni}\}$ we suppose that there exists a nonnegative and continuous function J on (0, 1) such that for each $1\leqq i\leqq n$

$$c_{ni}=n\int_{(i-1)/n}^{i/n}J(t)dt.$$

Specifically we assume that J is Lipschitz continuous on any interval $[\delta, 1-\delta]$ with $0<\delta\leqq 1/2$ and for some real ρ_0, ρ_1

$$J(t)=t^{\rho_0}\nu_0(t)\quad\text{and}\quad J(1-t)=t^{\rho_1}\nu_1(t)\qquad\text{for}\quad 0<t\leqq\delta,$$

where ν_0 and ν_1 are slowly varying at $t=0$ and

$$\nu_0'(t)=t^{-1}\nu_0(t)\varepsilon_0(t)\quad\text{and}\quad \nu_1'(t)=t^{-1}\nu_1(t)\varepsilon_1(t)$$

for some continuous functions ε_0 and ε_1 with $\varepsilon_0(t)$, $\varepsilon_1(t)\to 0$ as $t\to 0$. Moreover define a key function K on (0, 1) by

$$K(t):=\begin{cases}0 & \text{for}\quad 0<t<c\\ \int_0^t J(t)dg(t) & \text{for}\quad c\leqq t<1,\end{cases}$$

here c is a fixed continuity point of g with $0<c<1$. We also assume that K is not the trivial zero function. Since K is nondecreasing and $K(\varepsilon)=0$ for some $\varepsilon>0$, K is the left continuous inverse of some distribution function H. For some uniformly distributed random variables $\{\xi_j\}$ define $Y_{ab}(\xi_j)$ by

$$Y_{ab}(\xi_j):=K_{ab}(\xi_j)-\boldsymbol{E}(K_{ab}(\xi_j))=-\int_a^b\zeta_j(t)dK(t),$$

where

$$\zeta_j(t):=I\{\xi_j<t\}-t,\qquad 0<t<1,$$

$$K_{ab}(t):=\begin{cases}K(a) & \text{for}\quad 0<t\leqq a\\ K(t) & \text{for}\quad a<t<b.\\ K(b) & \text{for}\quad b\leqq t<1\end{cases}$$

We investigate the asymptotic normality of T_n under the next three conditions for trimming sizes k_n and m_n.

Condition I. (Trimming fixed numbers).

$k_n=k$ and $m_n=m$ for all n, where k and m are fixed numbers.

Condition II. (Trimming fixed fractions).

For $0<a<b<1$, $\lim_{n\to\infty}\sqrt{n}\left(\frac{k_n}{n}-a\right)=0$ and $\lim_{n\to\infty}\sqrt{n}\left(\frac{m_n}{n}-b\right)=0$.

Condition III. (Trimming vanishing fractions).

$\lim_{n\to\infty}k_k=\infty$ and $\lim_{n\to\infty}\frac{k_n}{n}=0$, $\lim_{n\to\infty}(n-m_n)=\infty$ and $\lim_{n\to\infty}\frac{m_n}{n}=1$.

In the case where $\{\xi_j\}$ is a sequence of i.i.d. random variables [1] succeeded to characterize the possible limiting behavior of trimmed sum T_n. Furthermore [3] obtained necessary and sufficient conditions of asymptotic normality of T_n. In this note we shall extend their results to the case where $\{\xi_j\}$ are weakly dependent random variables satisfying mixing conditions.

Theorem 1. (Trimming a fixed number). *Suppose that Condition I is satisfied. Furthermore assume that $\{\xi_j\}$ is a strictly stationary sequence of random variables with zero mean and $\boldsymbol{E}(|g(\xi_1)|^{2+\delta})<\infty$ for some $\delta>0$. Assume that $\{\xi_j\}$ satisfies either the ϕ-mixing condition* (1) *with*

$$\sum_{k=1}^{\infty} k^{2p-2}\{\phi(k)\}^{1/2}<\infty \quad \text{and} \quad \phi(k)=O(k^{-3(1-q)/4q}), \tag{3}$$

or the strong mixing condition (2) *with*

$$\sum_{k=1}^{\infty} k^{2p-2}\{\alpha(k)\}^{1/2}<\infty \quad \text{and} \quad \alpha(k)=O(k^{-2(1+r)/3r}), \tag{4}$$

here $p\in\boldsymbol{Z}^+$ and q, $r<1/4$ are some positive constants. Then

$$\sigma_{01}^2:=\lim_{n\to\infty}\frac{1}{n}\boldsymbol{E}\Big(\sum_{j=1}^{n}Y_{01}(\xi_j)\Big)^2$$

exists. If $\sigma_{01}^2>0$, then we can redefine $\{T_n\}$ and a sequence of independent copies of Gaussian process $\{B_n(t),\ 0<t<1\}$ with covariance

$$\Gamma(s,t)=\boldsymbol{E}(\zeta_1(s)\zeta_1(t))+\sum_{k=1}^{\infty}\boldsymbol{E}(\zeta_1(t)\zeta_{k+1}(s))+\sum_{k=1}^{\infty}\boldsymbol{E}(\zeta_1(s)\zeta_{k+1}(t)) \tag{5}$$

on a common probability space such that

$$\lim_{n\to\infty}\boldsymbol{P}\left\{\left|\sqrt{n}(T_n-\mu_n)+\int_0^1 B_n(t)dK(t)\right|>\varepsilon\right\}=0, \tag{6}$$

for any $\varepsilon>0$, where $\mu_n:=\int_{k_n/n}^{m_n/n}J(t)g(t)dt$. Furthermore

(7) $$\sqrt{n}(T_n-\mu_n)/\sigma_{01} \xrightarrow{d} N(0, 1)$$

as $n\to\infty$, *where* "$\xrightarrow{d}$" *means convergence in law.*

Theorem 2. (Trimming a fixed fraction). *Assume that Condition II is satisfied. Then, for any* $0<a<b<1$,

$$\sigma_{ab}^2 := \lim_{n\to\infty}\frac{1}{n}\boldsymbol{E}\Big(\sum_{j=1}^{n} Y_{ab}(\xi_j)\Big)^2$$

exists. Suppose that J *is Lipschitz continuous on a neighborhood containing* $[a, b]$ *with* $J(a+t)=t^{\rho_0}\nu_0(t)$ *and* $J(b-t)=t^{\rho_1}\nu_1(t)$ *for* $0<t\leqq\delta$ *and* $\sigma_{ab}^2>0$. *Under the same assumptions as in Theorem* 1, *if*

$$K(a+)-K(a-)=K(b+)-K(b-)=0,$$

then we can redefine $\{T_n\}$ *and* $\{B_n(t)\}$ *on a common probability space such that*

(8) $$\lim_{n\to\infty}\boldsymbol{P}\Big\{\Big|\sqrt{n}(T_n-\mu_n)+\int_a^b B_n(t)dK(t)\Big|>\varepsilon\Big\}=0$$

for any $\varepsilon>0$. *Furthermore, as* $n\to\infty$

(9) $$\sqrt{n}(T_n-\mu_n)/\sigma_{ab} \xrightarrow{d} N(0, 1).$$

Theorem 3. (Trimming a vanishing fraction). *Assume that Conditions III is satisfied. Under the same assumptions in Theorem* 1, *if* $\Psi_{in}(c)\to 0$ *as* $n\to\infty$ *holds for all real* c *and* $i=0, 1$, *then we can redefine* $\{T_n\}$ *and* $\{B_n(t)\}$ *on a common probability space such that* (6) *and* (7) *hold, where* Ψ_{in} *is some nondecreasing and left continuous function with* $\Psi_{in}(0)=0$ *for* $i=0, 1$ *defined by*

$$\Psi_{0n}(x):=\begin{cases}\sqrt{k_n/n}\{K((3k_n)/(2n))-K(k_n|n)\}, & x>\sqrt{k_n}/2\\ \sqrt{k_n/n}\{K((k_n/n)+x\sqrt{k_n}/n)-K(k_n/n)\}, & |x|\leqq\sqrt{k_n}/2\\ \sqrt{k_n/n}\{K(n_n/(2n))-K(k_n/n)\}, & x<-\sqrt{k_n}/2\end{cases}$$

$$\Psi_{1n}(x):=\begin{cases}\sqrt{(n-m_n)/n}\{K((m_n/n)+(n-m_n)/(2n))-K(m_n/n)\}, & x>\sqrt{n-m_n}/2\\ \sqrt{(n-m_n)/n}\{K((m_n/n)+x\sqrt{n-m_n}/n)-K(m_n/n)\}, & |x|\leqq\sqrt{n-m_n}/2.\\ \sqrt{(n-m_n)/n}\{K((m_n/n)-(n-m_n)/(2n))-K(m_n/n)\}, & x<-\sqrt{n-m_n}/2\end{cases}$$

2. Proof of Theorem 2

We first treat the ϕ-mixing case with (3). The strong mixing case can be considered similarly. Before proving we prepare the following lemma due to [2].

Lemma 1. *Put* $U_n(t):=n^{-1/2}\sum_{i=1}^{n}\zeta_i(t)$ *for* $0<t<1$. *Under the assumptions in*

Theorem 1, *on a common probability space, we can construct* $\{U_n(t)\}$ *and independent copies of Gaussian process* $\{B_n(t)\}$ *with the covariance* $\Gamma(s, t)$ *defined by* (5) *such that*

$$P\{\sup_{t\in(0,1)} |U_n(t)-B_n(t)| \geqq Cn^{-\beta}(\log n)^{1/2}\} \leqq Cn^{-\beta}(\log n)^{1/2}, \tag{10}$$

here β *is some positive constant dependent on* $\{\alpha(k)\}$ *or* $\{\phi(k)\}$.

Now we show that $\sigma_{ab}^2 := \lim_{n\to\infty}(1/n)E\Big(\sum_{j=1}^{n} Y_{ab}(\xi_j)\Big)^2$ exists for any $0<a<b<1$. Using Theorem 1.2.2 in [6] and (3),

$$\begin{aligned} E\Big(\sum_{j=1}^{n} Y_{ab}(\xi_j)\Big)^2 &= \sum_{1\le i,j\le n} E\Big(\int_a^b \zeta_i(t)dK(t)\cdot\int_a^b \zeta_j(t)dK(t)\Big) \\ &= \sum_{1\le i,j\le n} \int_a^b\int_a^b [E(I\{\xi_i<s\}I\{\xi_j<t\})-st]dK(s)dK(t) \\ &= \sum_{1\le i,t\le n} 2\phi(|i-j|)\int_a^b\int_a^b dK(s)dK(t) = O(n), \end{aligned} \tag{11}$$

as $n\to\infty$. Hence the limit exists clearly.

We next prove (8). According to [3], p. 127, we have

$$\begin{aligned} &\sqrt{n}(T_n-\mu_n)+\int_a^b B_n(t)dK(t) \\ &= \sqrt{n}\Big\{\int_{(\xi_{n:k},\,\xi_{n:m_n}]} g(t)d\Gamma(G_n^*(t)) - \int_{k_n/n}^{m_n/n} g(t)d\Gamma(t)\Big\} + \int_a^b B_n(t)dK(t) \\ &= -\int_{\xi_{n:k_n}}^{m_n/n} n^{-1/2}\int_t^{G_n^*(t)} J(s)dsdg(t) - \int_{\xi_{n:k_n}}^{k_n/n} n^{-1/2}\int_{k_n/n}^{G_n^*(t)} J(s)dsdg(t) \\ &\quad + \int_{\xi_{n:m_n}}^{m_n/n} n^{-1/2}\int_{m_n/n}^{G_n^*(t)} J(s)dsdg(t) + \int_a^b B_n(t)dK(t) \\ &= \int_a^{k_n/n} B_n(t)dK(t) + \int_{m_n/n}^b B_n(t)dK(t) - \int_{k_n/n}^{m_n/n}(U_n(t)-B_n(t))dK(t) \\ &\quad - \int_{k_n/n}^{m_n/n}(U_n^*(t)-U_n(t))dK(t) - \frac{1}{2}n^{-1/2}\int_{k_n/n}^{m_n/n} U_n^*(t)^2 J'(t_n^*)dg(t) \\ &\quad - \int_{\xi_{n:k_n}}^{k_n/n} n^{-1/2}\int_{k_n/n}^{G_n^*(t)} J(s)dsdg(t) + \int_{\xi_{n:m_n}}^{m_n/n} n^{-1/2}\int_{k_n/n}^{G_n^*(t)} J(s)dsdg(t) \\ &=: I_1+I_2+I_3+I_4+I_5+I_6+I_7, \end{aligned} \tag{12}$$

where

$$d\Gamma(t) := J(t)dt,\ G_n(t) := \frac{1}{n}\sum_{i=1}^{n} I(\xi_i \leqq t),\ U_n^*(t) := \sqrt{n}\,\{G_n^*(t)-t\},$$

$$G_n^*(t):=\begin{cases} 1/n, & 0<t<\xi_{n:1} \\ G_n(t), & \xi_{n:1}\leqq t<\xi_{n:n}, \\ 1-1/n, & \xi_{n:n}\leqq t<1 \end{cases}$$

for $t\in(0, 1)$ and t_n^* lies between $G_n^*(t)$ and t. Since $k_n/n\to a$, $m_n/n\to b$ and $K(a+)-K(a-)=K(b+)-K(b-)=0$, it is easy to see that

$$\lim_{n\to\infty}\boldsymbol{P}\{|I_1/\sigma_{ab}|>\varepsilon\}=\lim_{n\to\infty}\boldsymbol{P}\{|I_2/\sigma_{ab}|>\varepsilon\}=0. \tag{13}$$

We next show that I_3 converges to zero in probability. Put

$$A:=\{\sup_{t\in(0,1)}|U_n(t)-B_n(t)|\geqq Cn^{-\beta}(\log n)^{1/2}\}.$$

Using Lemma 1, we can construct $\{U_n(t)\}$ and $\{B_n(t)\}$ on a common probability space such that for any $\varepsilon>0$

$$\begin{aligned}(14)\quad & \boldsymbol{P}\{|I_3/\sigma_{ab}|>\varepsilon\}\leqq\boldsymbol{P}\{|I_3/\sigma_{ab}|>\varepsilon,\ A^c\}+\boldsymbol{P}\{A\} \\ & \leqq\boldsymbol{P}\Big\{\int_{k_n/n}^{m_n/n}|U_n(t)-B_n(t)|\,dK(t)/\sigma_{ab}>\varepsilon,\ \sup_{t\in(0,1)}|U_n(t)-B_n(t)|<Cn^{-\beta}(\log n)^{1/2}\Big\} \\ & \quad+\boldsymbol{P}\{A\} \\ & \leqq\boldsymbol{P}\Big\{Cn^{-\beta}(\log n)^{1/2}\int_{k_n/n}^{m_n/n}dK(t)/\sigma_{ab}>\varepsilon,\ \sup_{t\in(0,1)}|U_n(t)-B_n(t)|<Cn^{-\beta}(\log n)^{1/2}\Big\} \\ & \quad+\boldsymbol{P}\{A\} \\ & \leqq\boldsymbol{P}\Big\{Cn^{-\beta}(\log n)^{1/2}\int_{k_n/n}^{m_n/n}dK(t)/\sigma_{ab}>\varepsilon\Big\}+Cn^{-\beta}(\log n)^{1/2}\longrightarrow 0 \quad\text{as}\quad n\to\infty.\end{aligned}$$

Furthermore we can also treat $I_4\sim I_7$ similarly and conclude the proof of (8) from (12)~(14). Finally we see easily that $\int_a^b B_n(t)dK(t)/\sigma_{ab}$ obeys the standard normal distribution which implies (9) from (8).

2. Proofs of Theorems 1 and 3

According to [3] we can show these theorems similarly to Theorem 2.

References

[1] S. Csörgö, E. Haeusler and D. Mason, The asymptotic distribution of trimmed sums, Ann. Probab., **16** (1988), 672-699.

[2] P. Douhkan and F. Portal, Principle d'invariance principle faible pour la fonction de repartition empirique dans un vadre multidimensionnel et melangeant, Prob. Math. Statist., **8** (1987), 117-132.

[3] D. Mason and G. Shorack, Necessary and sufficient conditions for asymptotic normality of trimmed L-statistics, J. Statist. Planning Inference, **25** (1990), 111-139.

[4] T. Mori, On the limit distributions of lightly trimmed sums, Camb. Phil. Soc., **96**

(1984), 507-516.
[5] P.K. Sen, *Sequential Nonparametrics,* John Wiley & Sons, New York, 1981.
[6] K. Yoshihara, *Weakly Dependent Stochastic Sequences and Their Applications, Vol. 3, Asymptotic Statistics based on Weakly Dependent Data,* Sanseido Co. Ltd., Tokyo, 1993.

Department of Mathematics
Faculty of Engineering
Yokohama National University
156, Tokiwadai, Hodogaya
Yokohama 240, JAPAN

Department of Mathematics
Faculty of Liberal Arts & Education
Yamanashi University
4-4-37, Takeda
Kôfu 400, JAPAN