Internat. J. Math. & Math. Sci.
VOL. 14 NO. 4 (1991) 797-802

STRONG CONSISTENCIES OF THE BOOTSTRAP MOMENTS

TIEN-CHUNG HU

Department of Mathematics
National Tsing Hua University
Hsinchu, Taiwan 30043, R.O.C.

and

Department of Statistics
University of North Carolina
Chapel Hill, NC 27599 U.S.A.

(Received November 14, 1990 and in revised form January 23, 1991)

ABSTRACT. Let X be a real valued random variable with $E|X|^{r+\delta} < \infty$ for some positive integer r and real number, δ, $0 < \delta \leq r$, and let $\{X, X_1, X_2, \ldots\}$ be a sequence of independent, identically distributed random variables. In this note, we prove that, for almost all $w \in \Omega$, $\mu^*_{r:n}(w) \to \mu_r$ with probability 1, if $\liminf_{n\to\infty} m(n)n^{-\beta} > 0$ for some $\beta > \frac{r-\delta}{r+\delta}$, where $\mu^*_{r:n}$ is the bootstrap r^{th} sample moment of the bootstrap sample with sample size m(n) from the data set $\{X_1, X_2, \ldots, X_n\}$ and μ_r is the r^{th} moment of X. The results obtained here not only improve on those of Athreya [3] but also the proof is more elementary.

KEY WORDS AND PHRASES. Bootstrap sample size, Sample moment,, Convergence with probability 1.

1980 AMS SUBJECT CLASSIFICATION CODES. Primary 60F15; secondary 62G05.

1. INTRODUCTION

Let X be a real valued random variable with $E|X|^{r+\delta} < \infty$ for some positive real number $\delta \leq r$, and let $\{X, X_1, X_2, \ldots\}$ be a sequence of independent, identically distributed random variables. Let

$$F_n(x;w) = \frac{1}{n}\sum_{i=1}^{n} I[X_i(w) \leq x], \quad n = 1,2,\ldots, \quad w \in \Omega, \tag{1.1}$$

be the empirical distribution functions associated with the sequence $\{X_1(w), X_2(w), X_3(w), \ldots\}$. For every positive integer n and $w \in \Omega$, let $\{X_{n1}(w), X_{n2}(w), \ldots, X_{nm(n)}(w)\}$ be independent, identically distributed random variables with distribution function $F_n(x;w)$ defined as in (1.1). We call $\{X_{n1}(w), X_{n2}(w), \ldots, X_{nm(n)}(w)\}$ the bootstrap sample set with bootstrap sample size m(n); it is required that $m(n) \to \infty$ as $n \to \infty$. Denote by $\mu_{n;r}(w)$ and $\mu^*_{n;r}(w)$ the r^{th} sample moment of $\{X_1(w), X_2(w), \ldots, X_n(w)\}$ and the bootstrap r^{th} sample moment of $\{X_{n1}(x), X_{n2}(w), \ldots, X_{nm(n)}(w)\}$ respectively and denote by μ_r the r^{th} moment of X. (When r=1, we use $\mu_n(w)$ and $\mu^*_n(w)$ instead of $\mu_{n;1}(w)$ and $\mu^*_{n;1}(w)$; further $\mu_n(w)$ and $\mu^*_n(w)$ are called sample mean and bootstrap sample mean respectively.) A problem, from the bootstrap theory of Efron [1], is to find conditions such that, for almost all w, the bootstrap sample mean converges to the population mean (when it exists). That is, for almost all w,

$$\mu^*_{n;r}(w) \to \mu_r \quad \text{as} \quad n \to \infty \tag{1.2}$$

with probability 1. By using the abstract "Vasserstein's metric" among distributions and a Mallow type inequality, Bickel and Freedman [2] showed that if $E|X| < \infty$, then for almost all $w \in \Omega$, (2) holds in probability. Athreya [3] found that if $E|X|^\theta < \infty$ for some $\theta \geq 1$, and $\liminf_{n\to\infty} m(n)n^{-\beta} > 0$ for some $\beta > 0$ such that $\theta\beta > 1$, then for almost all $w \in \Omega$, (1.2) holds with probability 1. To show this he used the difficult and complex inequality of Kurtz [4]. Bickel and Freedman and Athreya used deep mathematics and hard inequalities to prove the consistency of the bootstrap sample mean to the population mean. Their proofs are not easily comprehended. This note, provides an elementary way to obtain the strong consistency, relying on the Markov inequality. Moreover, the consistency property holds under weaker conditions than those presented in Athreya [3].

2. RESULTS AND PROOFS

THEOREM 2.1. *Let $\{X, X_1, X_2, \ldots\}$ be a sequence of independent, identically distributed random variables with $E|X|^{r+\delta} < \infty$ for some integer r and real number $\delta \leq r$. Then, for almost all $w \in \Omega$, (1.2) holds with probability 1, if $\liminf_{n\to\infty} m(n)n^{-\beta} > 0$ for some real number $\beta > 0$ such that $\beta > \frac{r-\delta}{r+\delta}$.*

First, a lemma is needed in proving the theorem. The lemma is known in the literature. For the sake of completeness, a proof for the lemma is given.

LEMMA 2.2. Let $\{X, X_1, X_2, \ldots\}$ be a sequence of independent, identically distributed random variables. Then, for any $0 < p < 1$, $E|X|^p < \infty$ implies that

$$\sum_{n=1}^{\infty} \frac{|X_n|}{n^{1/p}} < \infty \quad \text{with probability 1.}$$

PROOF. Let $Y_n = \frac{|X_n|}{n^{1/p}} I\left[\frac{|X_n|}{n^{1/p}} \le 1 \right]$. Thus

$$\sum_{n=1}^{\infty} P\left[\frac{|X_n|}{n^{1/p}} \neq Y_n \right] = \sum_{n=1}^{\infty} P(|X_n| > n^{1/p})$$

$$= \sum_{n=1}^{\infty} P(|X|^p > n) < \infty$$

since $E|X|^p < \infty$. Defining $A_j = \{(j-1)^{1/p} < |X| \le (j)^{1/p}\}$, we have, for $\alpha \ge 1$,

$$\sum_{n=1}^{\infty} E(|Y_n|^\alpha) = \sum_{n=1}^{\infty} \sum_{j=1}^{n} \int_{A_j} \frac{1}{n^{\alpha/p}} |X|^\alpha dP = \sum_{j=1}^{\infty} \left[\sum_{n=j}^{\infty} \frac{1}{n^{\alpha/p}} \right] \int_{A_j} |X|^\alpha \, dP$$

$$\le \sum_{j=1}^{\infty} \left[\frac{C}{j^{(\alpha/p)-1}} \right] \int_{A_j} |X|^\alpha dP \le \sum_{j=1}^{\infty} C \int_{A_j} |X|^p = C\, E|X|^p < \infty$$

where the constant C depends only on α and p. Choosing $\alpha = 1$ and 2, we have that

$$\sum_{n=1}^{\infty} E|Y_n| < \infty$$

and

$$\sum_{n=1}^{\infty} \operatorname{Var} Y_n < \infty.$$

Thus, by the "three series Theorem" of Kolmogorov the lemma is established.

We are now in a position to prove the main result, which provides the strong consistency of the bootstrap sample moments.

PROOF OF THEOREM 2.1. It suffices to prove the result for the case r=1. The other cases can be proved in a similar way with minor changes. Recall that, for each n and $w \in \Omega$, $\{X_{n1}(w), X_{n2}(w), \ldots, X_{nm(n)}(w)\}$ are the independent, identically distributed random variables with distribution function defined in (1.1). By the strong law of large numbers, we have for almost all $w \in \Omega$,

$$\mu_n(w) \to \mu \quad \text{as} \quad n \to \infty. \tag{2.1}$$

Thus, it suffices to show that, for almost all $w \in \Omega$,

$$|\mu_n^*(w) - \mu_n(w)| \to 0 \quad \text{as} \quad n \to \infty$$

with probability 1. From the Borel-Cantelli lemma, we only need to prove for almost all $w \in \Omega$ and for every $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P\left\{\left[\left|\frac{1}{m(n)} \sum_{i=1}^{m(n)} X_{ni}(w) - \mu_n(w)\right| > \epsilon\right] \,|X_1(w),\ X_2(w), \dots,\ X_n(w)\right\} < \infty. \quad (2.2)$$

For the case of presentation, we suppress all the symbol w in $X_i(w)$, $X_{ni}(w)$ and $\mu_n(w)$ and the symbol n in m(n). Let q be an integer such that $\frac{1}{q} < \beta - \frac{1-\delta}{1+\delta}$ and from the Markov inequality, we have

$$P\left[\left|\frac{1}{m}\sum_{i=1}^{m} X_{ni}-\mu_n\right| > \epsilon|X_1,X_2,\dots,X_n\right] \le \left(\frac{1}{m\epsilon}\right)^{2q} E\left[\left(\sum_{i=1}^{m}(X_{ni}-\mu_n)\right)^{2q}|X_1,X_2,\dots,X_n\right] \quad (2.3)$$

Now we write

$$B_{2q} = E\left[\left(\sum_{i=1}^{n} (X_{ni}-\mu_n)\right)^{2q}|X_1,X_2,\dots,X_n\right]$$

$$= \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \cdots \sum_{i_{2q}=1}^{m} E\left[(X_{ni_1}-\mu_n)\cdots(X_{ni_{2q}}-\mu_n)|X_1,X_2,\dots,X_n\right]$$

$$= \sum_{t=1}^{2q} \sum_{\substack{q_1+\dots+q_t=2q \\ q_i\ge 1, i=1,\dots,t}} \frac{(2q)!}{q_1!\dots q_t!} \binom{m}{t} E\left[(X_{n1}-\mu_n)^{q_1}\dots(X_{nt}-\mu_n)^{q_t}|X_1,X_2,\dots,X_n\right]$$

$$= \sum_{t=1}^{q} \sum_{\substack{q_1+\dots+q_t=2q \\ q_i\ge 2, i=1,\dots,t}} \frac{(2q)!}{q_1!\dots q_t!} \binom{m}{t} E\left[(X_{n1}-\mu_n)^{q_1}|X_1,X_2,\dots,X_n\right]$$

$$\dots E\left[(X_{nt}-\mu_n)^{q_t}|\ X_1,X_2,\dots,X_n\right] \quad (2.4)$$

where the third equality in (2.4) holds since X_{n1}, $X_{n2}, \dots,$ X_{nm} are identically distributed. Further the last equality in (2.4) is justified since X_{n1}, $X_{n2}, \dots,$ X_{nm} are independent and $E(X_{n1}-\mu_n) = 0$ implies that there is no contribution for those terms which contain at least one of $q_i=1$. In the sequel, we use the shorthand notation $a_n \sim b_n$ for $a_n = O(b_n)$ as $n \to \infty$. Thus, from (2.4) we have

$$B_{2q} \sim \sum_{t=1}^{q} m^t\ E\left[(X_{n1}-\mu_n)^{q_1}|X_1,X_2,\dots,X_n\right] \cdots E\left[(X_{nt}-\mu_n)^{q_t}|\ X_1,X_2,\dots,X_n\right]$$

$$= \sum_{t=1}^{q} m^t\left[\frac{1}{n}\sum_{i=1}^{n} (X_i-\mu_n)^{q_1}\right] \cdots \left[\frac{1}{n}\sum_{i=1}^{n} (X_i-\mu_n)^{q_t}\right]$$

$$= \sum_{t=1}^{q} \left(\frac{m}{n}\right)^t \left[\sum_{i=1}^{n} (X_i-\mu_n)^{q_1}\right] \cdots \left[\sum_{i=1}^{n} (X_i-\mu_n)^{q_t}\right]$$

$$\leq \sum_{t=1}^{q} (\tfrac{m}{n})^t \left[\sum_{i=1}^{n} 2^{q_1}(|X_i|^{q_1} + |\mu_n|^{q_1})\right] \cdots \left[\sum_{i=1}^{n} 2^{q_t}(|X_i|^{q_t}+|\mu_n|^{q_t})\right]$$

$$\leq 2^{2q} \sum_{t=1}^{q} (\tfrac{m}{n})^t (n)^{\Sigma_{j=1}^{t} \frac{q_j}{q_j-1+q}} \left[\sum_{i=1}^{n} (\tfrac{1}{i})^{\frac{q_1}{q_1-1+\delta}} (|X_i|^{q_1}+|\mu_n|^{q_1})\right]$$

$$\cdots \left[\sum_{i=1}^{n} (\tfrac{1}{i})^{\frac{q_t}{q_t-1+\delta}} (|X_i|^{q_t} + |\mu_n|^{q_t})\right] \qquad (2.5)$$

since $q_j \geq 2$, $q_1+q_2+\ldots+q_t = 2q$ and where the first inequality in (2.5) is obtained for fact that $|a+b|^s \leq 2^s(|a|^s+|b|^s)$ for a, b and s real numbers. Since μ is finite it follows from (2.1) that, for almost all w, there exists a constant C such that $\mu_n < C$ for every n. Further note that $\delta < 1$ and $q_j \geq 2$ for $j = 1,2,\ldots,q$ which imply $\frac{q_j}{q_j-1+\delta} > 1$. Thus, for almost all $w \in \Omega$,

$$\sum_{i=1}^{\infty} (\tfrac{1}{i})^{\frac{q_j}{q_j-1+\delta}} |\mu_n|^{q_j} < \infty, \quad j = 1,2,\ldots,q. \qquad (2.6)$$

Now applying our lemma to $|X_i|^{q_j}$ and choosing $p = \frac{q_j-1+\delta}{q_j} < 1$, we have, for almost all $w \in \Omega$,

$$\sum_{i=1}^{\infty} (\tfrac{1}{i})^{\frac{q_j}{q_j-1+\delta}} |X_i|^{q_j} < \infty, \quad j = 1,2,\ldots,q. \qquad (2.7)$$

Without loss of generality we put $m(n) = n^{\beta}$ where β is some real number such that $\beta > \frac{1-\delta}{1+\delta}$, then from (2.5), (2.6) and (2.7), we have

$$(\tfrac{1}{m})^{2q} B_{2q} \sim \sum_{t=1}^{q} \left[\tfrac{1}{n}\right]^{2q\beta+(1-\beta)t\ \Sigma_{j=1}^{t} \frac{q_j}{q_j-1+\delta}} \qquad (2.8)$$

Denote the exponent of $(\frac{1}{n})$ in (2.8) as $\phi(t)$ for $t = 1,2,\ldots,q$. Note that

$$\phi(t) = 2q\beta + (1-\beta)t - \sum_{j=1}^{t} \frac{q_j}{q_j-1+\delta}$$

$$\geq 2q\beta - t(\beta-1 + \tfrac{2}{1+\delta}) \qquad (2.9)$$

$$\begin{aligned} &\geq 2q\beta - q(\beta + \frac{1-\delta}{1+\delta}) \\ &= q(\beta - \frac{1-\delta}{1+\delta}) \end{aligned} \qquad (2.10)$$

for $t = 1,2,\ldots,q$, where (2.9) holds since $\frac{q_j}{q_j - 1+\delta} < \frac{2}{1+\delta}$ for $j = 1,2,\ldots,t$. By the appropriate choice of q, (2.2) follows from (2.3), (2.8) and (2.10). This completes the proof.

Theorem 1 of Athreya, stated in Corollary 2.3, is an immediate consequence of our Theorem 2.1.

COROLLARY 2.3. If $E|X| < \infty$ and $m(n) = \rho^n$ for some $\rho > 1$, then (2.2) holds, for $r = 1$.

COROLLARY 2.4. If $E\, X^2 < \infty$ and $m(n) = n^\beta$ for some $\beta > 0$, then (2.2) holds for $r = 1$.

PROOF. By letting $\delta = 1-\alpha$ for some $\alpha < \frac{2\beta}{1+\beta}$.

REMARKS. Let m(n) growth with an algebraic rate; that is, $m(n) = n^\beta$ for some $\beta > 0$. First note that $1+\delta$ in our notation plays the role of θ in Athreya's. In case $1 < \theta < 2$ we have $\beta > \frac{1}{1+\delta} > \frac{1-\delta}{1+\delta}$. Hence, in this case, our condition is strictly weaker than the one is posed by Athreya. In case $\theta \geq 2$, we only require $\beta > 0$ which is even much weaker than $\beta > \frac{1}{\theta}$ the requirement in the Theorem 2 of Athreya. However, if $\theta = 1$, then both of our theorem and Athreya's theorem require $\beta > 1$.

Recently, the author learned that Professor Csorgo also improved the result of Athreya using a different approaching.

ACKNOWLEGEMENT. The author is most grateful to the Department of Statistics, University of North Carolina, Chapel Hill, for their hospitality. The author also wishes to thank the editor and a referee for helpful suggestions.

REFERENCES

1. Efron, B. Bootstrap methods: Another look at the jacknife, *Annals of Statistics* 7 (1979), 1-26.

2. Bickel, P.J. and Freedman, D.A. Some asymptotic theory for the bootstrap. *Annals of Statistics* 9 (1981), 1196-1217.

3. Athreya, K.B. Strong law for bootstrap, *Statistics and Probability Letters* 1 (1983), 147-150.

4. Kurtz, T.G. Inequalities for laws of large numbers, *Annals of Mathematical Statistics* 43 (1972), 1874-1883.

5. Chung, K.L. *A Course in Probability Theory*, Harcourt, Brace and World, New York, 1968.