

On the law of large numbers for the bootstrap mean

Sándor Csörgő *

Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1027, USA

Received July 1990

Dedicated to my brother, Miklós Csörgő, for his sixtieth birthday.

Abstract: Direct and elementary proofs are given for weak and strong laws of large numbers for bootstrap sample means under minimal moment conditions. Concerning the required rate of divergence of the bootstrap sample size, the strong laws obtained improve on those of Athreya (1983).

AMS 1980 Subject Classifications: Primary 60F15; Secondary 62G05.

Keywords: Bootstrap, sample mean, law of large numbers.

1. Results and discussion

Let X_1, X_2, \dots be an infinite sequence of independent random variables with a common distribution function $F(x) = P\{X \leq x\}$, $x \in \mathbb{R}$, defined on the same probability space, and for each $n = 1, 2, \dots$, let $X_1^{(n)}, \dots, X_m^{(n)}$ be the ordinary Efron bootstrap sample from X_1, \dots, X_n with bootstrap sample size $m = m(n)$, where $\{m(n)\}$ is a sequence of positive integers. The variables $X_1^{(n)}, \dots, X_m^{(n)}$ result from sampling m times the sequence X_1, \dots, X_n with replacement such that at each stage any one element has probability $1/n$ to be picked.

There is an extensive literature on the asymptotic distribution of the bootstrap sample mean $m^{-1} \sum_{j=1}^m X_j^{(n)}$. (See Arcones and Giné (1989) for the newest results and some earlier references.) Strong laws of large numbers have been proved by Athreya (1983) for this mean together with the weak law. The aim of the present note is to improve on Athreya's strong laws by using a simple direct and elementary proof, an easy modification of which is also applicable to give the weak law.

The validity of the strong or the weak law of large numbers, i.e.

$$\lim_{n \rightarrow \infty} \frac{1}{m(n)} \sum_{j=1}^{m(n)} X_j^{(n)} = E(X) \quad \text{almost surely} \quad (1.1)$$

or

$$\frac{1}{m(n)} \sum_{j=1}^{m(n)} X_j^{(n)} \rightarrow E(X) \quad \text{in probability as } n \rightarrow \infty, \quad (1.2)$$

* Professor at large from the Bolyai Institute, University of Szeged, Hungary. Work done while a Visiting Professor at the Department of Statistics, University of North Carolina at Chapel Hill.

under the minimal moment condition $E(|X|) < \infty$, is one of the most natural theoretical problems and, in view of the statistical importance of the bootstrap, the results are likely to find numerous applications in large sample theory. In fact, the technical necessity of a law, preferably a strong law, has become apparent in work with Murray Burke and Edit Gombay on applying the nonparametric bootstrap in composite goodness-of-fit bootstrap tests when unknown parameters are estimated using maximum likelihood based on the bootstrap sample, and thus one wishes to substantiate condition (A6) in Burke and Gombay (1988) in the maximum likelihood setting. I also thank Murray Burke for calling my attention to Athreya (1983) after the first version of this note was completed.

Introducing the sequences

$$l_1^\delta(n) = (\log n)^{1+\delta}, \quad l_r^\delta(n) = (\log n)(\log \log n) \cdots (\log_{r-1} n)(\log_r n)^{1+\delta},$$

$r = 2, 3, \dots$, $\delta > 0$, where \log_r stands for the r times iterated logarithm and n is always taken as large as needed to make these definitions meaningful, our main result is the following.

Theorem. *If for some $0 \leq \delta < 1$,*

$$E(|X|^{1+\delta}) < \infty \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{n^{1-\delta}}{m(n)} < \infty, \tag{1.3}$$

then we have (1.1). If for some $r = 1, 2, \dots$ and $\delta > 0$,

$$E(X^2) < \infty \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{l_r^\delta(n)}{m(n)} < \infty, \tag{1.4}$$

then again we have (1.1). Furthermore, if

$$E(|X|) < \infty \quad \text{and} \quad m(n) \rightarrow \infty \quad \text{as } n \rightarrow \infty, \tag{1.5}$$

then we have (1.2).

The weak law in the third statement has been first proved by Bickel and Freedman (1981). Another proof of it was given by Athreya (1983). Since we give below what we believe to be a simplest possible proof, we included this statement for the sake of completeness.

Athreya (1983) formulates his strong laws, as first results of their kind, in two theorems. His Theorem 1, stating (1.1) under $E(|X|) < \infty$ and requiring an exponential growth rate for $m(n)$, is redundant in the sense that the special case $\theta = 1$ of his Theorem 2 provides a better result in demanding only an algebraic power growth. This Theorem 2 of Athreya establishes (1.1) under the condition that

$$E(|X|^\theta) < \infty \quad \text{and} \quad \limsup_{n \rightarrow \infty} \frac{n^\beta}{m(n)} < \infty \quad \text{for some } \theta \geq 1 \text{ and } \beta > 0 \text{ such that } \theta\beta > 1.$$

While requiring an algebraic growth of $m(n)$, our condition in (1.3) is strictly weaker in the range $1 \leq \theta = 1 + \delta < 2$. As long as $\theta \geq 2$, condition (1.4) asks only for a logarithmic growth of $m(n)$. (See also the finer form (1.10) of this condition below.) Athreya's proof is indirect and not elementary in the sense that it arrives at (1.1) through a conditional strong law and it is based on a rather sophisticated inequality of Thomas Kurtz. We aim at equivalent forms of (1.1) and (1.2) directly and use only the Markov and Chebyshev inequalities.

To expose the nature of our approach and an interpretation of the meaning of (1.1) and (1.2), a few more remarks are in order before the proofs in Section 2. Clearly, $\sum_{j=1}^m X_j^{(n)} = \sum_{j=1}^n f_j X_j$, where $f_j \geq 0$ is the frequency of X_j , $1 \leq j \leq n$, in $X_1^{(n)}, \dots, X_m^{(n)}$ and the Multinomial($m, 1/n, \dots, 1/n$) frequency vector

(f_1, \dots, f_n) is independent of (X_1, \dots, X_n) . To achieve an explicit representation, suppose without loss of generality that the original probability space is rich enough to carry an infinite sequence U_1, U_2, \dots of independent random variables uniformly distributed on the interval $(0, 1)$ such that the two sequences $\{X_j\}$ and $\{U_j\}$ are independent, and let $G_m(t) = m^{-1} \#\{1 \leq j \leq m: U_j \leq t\}$, $t \in \mathbb{R}$, denote the m th uniform empirical distribution function. Then we have the distributional equality

$$\left\{ \frac{1}{m(n)} \sum_{j=1}^{m(n)} X_j^{(n)} \right\}_{n=1}^{\infty} \stackrel{\mathcal{D}}{=} \left\{ \sum_{j=1}^n \left(G_{m(n)}\left(\frac{j}{n}\right) - G_{m(n)}\left(\frac{j-1}{n}\right) \right) X_j \right\}_{n=1}^{\infty}.$$

Such a representation is implicit already in Bickel and Freedman (1981). (See also Burke and Gombay (1988), for example.) Setting

$$w_n(j) = G_{m(n)}\left(\frac{j}{n}\right) - G_{m(n)}\left(\frac{j-1}{n}\right), \quad j = 1, \dots, n, \tag{1.6}$$

it follows that the statement (1.1) of the theorem is equivalent to

$$\mu_n := \sum_{j=1}^n w_n(j) X_j \rightarrow E(X) \quad \text{a.s. as } n \rightarrow \infty \tag{1.7}$$

and the statement (1.2) is equivalent to

$$\mu_n \rightarrow E(X) \quad \text{in probability as } n \rightarrow \infty. \tag{1.8}$$

These are laws for random linear combinations of the original variables X_1, X_2, \dots . In this form, the intuitive meaning of the law is that the linear combination μ_n behaves asymptotically as if all the random weights $w_n(1), \dots, w_n(n) \geq 0$ were replaced by their common expectation

$$E(w_n(j)) = \frac{1}{n}, \quad j = 1, \dots, n. \tag{1.9}$$

Note that, even though the sequence of weights $(w_n(1), \dots, w_n(n))$ and the sequence of independent weighted variables (X_1, \dots, X_n) are independent, the summands in (1.7) are not because the weights themselves are dependent: we have $w_n(1) + \dots + w_n(n) = G_{m(n)}(1) = 1$ almost surely. Hence the ordinary strong law of large numbers (cf. Etemadi, 1981; or Billingsley, 1986, pp. 290–292, for a recent best proof; and Stout, 1974, Section 3.2, for example, for old proofs) or triangular-array versions of the Etemadi-type strong laws for non-identically distributed variables in Csörgő, Tandori and Totik (1983) are not applicable, nor are results from the theory of strong laws for deterministic linear combinations (Stout 1974, Chapter 4). Nevertheless, it is completely natural to follow the basic outline of Etemadi’s best proof. This is what we do, though because of some new subtleties certain details have to be refined and part of this outline has to be essentially reversed.

Of course, the intuitive idea of substituting in the expectations (1.9) does not work by itself. Assuming only $E(|X|) < \infty$, which is necessary for the ordinary law, the resampling size $m(n)$ must go to infinity, as $n \rightarrow \infty$, to have (1.1). When we have $E(|X|^{1+\delta}) < \infty$ for some $0 \leq \delta < 1$, the condition in (1.3) requires $m(n)$ to diverge to infinity at least at the rate of $n^{1-\delta}$ and thus compensates for possibly heavier tails in a clear fashion. I do not have a proof to show that these conditions on $m(n)$ in (1.3) are in fact necessary but conjecture that they are close to being so. Note that in the most important special case when $\delta = 0$, the condition requires $m(n)$ to grow at least as the sample size n , and this is satisfied in most practical situations, as for the naive bootstrap when $m(n) \equiv n$.

When $E(X^2) < \infty$, the tentative (very mild) growth rates required in (1.4) are only typical but not optimal. This condition, while keeping the same proof given below, can be replaced by the more

complicated but mathematically weaker condition that there exists a sequence of positive numbers $l(n)$, slowly varying at infinity, such that

$$\limsup_{n \rightarrow \infty} \frac{l(n)}{m(n)} < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{nl(n)} < \infty. \tag{1.10}$$

This condition cannot be far from being necessary either.

It is of some interest to underline generally that here we talk about conditions for (1.1) that require the resampling size $m(n)$ to grow fast enough. Contrary to this, the slower $m(n)$ grows, the better chance we have for a reasonable asymptotic distribution statement when only moments of order less than 2 exist. This is clear from the results reviewed or proved by Arcones and Giné (1989).

Finally we point out that our proof may in principle be applicable to bootstrap means resulting from resampling schemes different from the uniform scheme treated here, which lead to weights $w_n(j)$ different from those in (1.6). It always works whenever

$$\min_{1 \leq j \leq n} \left\{ E(w_n(j)) - \frac{1}{n} \right\} \rightarrow 0 \quad \text{and} \quad \max_{1 \leq j \leq n} \left\{ E(w_n(j)) - \frac{1}{n} \right\} \rightarrow 0$$

as $n \rightarrow \infty$, and an inequality of the type of (2.3) below holds both for the truncated and the original variables Y_j and X_j . The corresponding extended versions of the Theorem are easily formulated for linear combinations of the form in (1.7). Not surprisingly, the trivial special case $w_n(j) = 1/n, j = 1, \dots, n$, of the case $\delta = 0$ of any one of such extended versions gives Etemadi's (1981) strong law for (pairwise) independent, identically distributed random variables.

2. Proof

Following Etemadi (1981), we assume without loss of generality that $X \geq 0$. This can be done here, too, since the weights are non-negative.

First we prove (1.7) under the conditions in (1.3). Set $k_n = k_n(\alpha) = [\alpha^n]$ for some $\alpha > 1$, where the brackets denote integer part, and introduce

$$Y_j = X_j I(X_j \leq j), \quad j = 1, 2, \dots,$$

using the Khinchin-Kolmogorov truncation scheme, where $I(\cdot)$ is the indicator function. Etemadi (1981), and in fact already Khinchin (1929) proved that under $E(X) < \infty$,

$$\sum_{j=1}^{\infty} P\{X_j \neq Y_j\} < \infty. \tag{2.1}$$

Extending somewhat the argument of Etemadi (1981), who establishes the case when $\delta = 0$, under our present moment condition $E(X^{1+\delta}) < \infty, 0 \leq \delta < 1$, it is straightforward to show (once the original argument is given) that

$$\sum_{n=1}^{\infty} \frac{1}{k_n^{2-\delta}} \sum_{j=1}^{k_n} E(Y_j^2) < \infty. \tag{2.2}$$

Using the fact that for $0 \leq s \leq t \leq 1$,

$$E(G_{m(n)}(s)G_{m(n)}(t)) = \frac{s}{m(n)} + \frac{m(n) - 1}{m(n)} st,$$

by elementary calculation we obtain

$$E(w_n^2(j)) = \frac{1}{nm(n)} + \frac{m(n) - 1}{m(n)} \frac{1}{n^2} \leq \frac{1}{nm(n)} + \frac{1}{n^2}, \quad 1 \leq j \leq n,$$

and

$$\text{Cov}(w_n(j), w_n(k)) = -\frac{1}{n^2 m(n)}, \quad 1 \leq j < k \leq n.$$

Hence, if we define $w_n(j) = 0$ for $j \geq n + 1$, by non-negativity,

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^k w_n(j)Y_j\right) &= \sum_{j=1}^k \text{Var}(w_n(j)Y_j) + 2 \sum_{1 \leq i < j \leq k} \text{Cov}(w_n(i)Y_i, w_n(j)Y_j) \\ &\leq \sum_{j=1}^k E(w_n^2(j))E(Y_j^2) + 2 \sum_{1 \leq i < j \leq k} \text{Cov}(w_n(i), w_n(j))E(Y_i)E(Y_j) \\ &\leq \left(\frac{1}{nm(n)} + \frac{1}{n^2}\right) \sum_{j=1}^k E(Y_j^2) \end{aligned} \tag{2.3}$$

for every $k \geq 1$ and $n \geq 1$, and if $C(\delta)$ denotes the finite supremum of the sequence $\{n^{1-\delta}/m(n)\}$ in (1.3),

$$\sum_{n=1}^{\infty} P\{|w_n(j)| > \varepsilon\} \leq \frac{1}{\varepsilon^2} \left\{ C(\delta) \sum_{n=1}^{\infty} \frac{1}{n^{2-\delta}} + \sum_{n=1}^{\infty} \frac{1}{n^2} \right\} < \infty \tag{2.4}$$

for every fixed $j \geq 1$ and $\varepsilon > 0$.

After these preliminaries, for each $n = 1, 2, \dots$, let $i(n)$ be the integer such that

$$k_{i(n)} < n \leq k_{i(n)+1}. \tag{2.5}$$

Clearly, $i(n) \rightarrow \infty$ as $n \rightarrow \infty$ and by non-negativity,

$$L(n) := \sum_{j=1}^{k_{i(n)}} w_n(j)X_j \leq \mu_n \leq \sum_{j=1}^{k_{i(n)+1}} w_n(j)X_j =: U(n). \tag{2.6}$$

Consider now the truncated versions of the lower and upper bounds $L(n)$ and $U(n)$ defined by

$$L^*(n) = \sum_{j=1}^{k_{i(n)}} w_n(j)Y_j \quad \text{and} \quad U^*(n) = \sum_{j=1}^{k_{i(n)+1}} w_n(j)Y_j.$$

It follows from (2.4) and the Borel–Cantelli lemma that

$$\max_{1 \leq j \leq k} w_n(j) \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty$$

and hence, for every fixed $k \geq 1$,

$$\sum_{j=1}^k w_n(j)Y_j \rightarrow 0 \quad \text{and} \quad \sum_{j=1}^k w_n(j)X_j \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

These relations, combined with (2.1), give

$$\lim_{n \rightarrow \infty} (L(n) - L^*(n)) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} (U(n) - U^*(n)) = 0 \quad \text{a.s.,}$$

and hence by (2.6), almost surely,

$$\liminf_{n \rightarrow \infty} L^*(n) \leq \liminf_{n \rightarrow \infty} \mu_n \leq \limsup_{n \rightarrow \infty} \mu_n \leq \limsup_{n \rightarrow \infty} U^*(n). \tag{2.7}$$

Next, using the inequality (2.3) with $k = k_{i(n)+1}$ and the fact that

$$\limsup_{n \rightarrow \infty} \left\{ \frac{k_{i(n)+1}^{2-\delta}}{nm(n)} + \frac{k_{i(n)+1}^{2-\delta}}{n^2} \right\} \leq \limsup_{n \rightarrow \infty} \left\{ \left(\frac{k_{i(n)+1}}{k_{i(n)}} \right)^{2-\delta} \left(\frac{n^{1-\delta}}{m(n)} + 1 \right) \right\} \leq \alpha^{2-\delta} (C(\delta) + 1),$$

where we used (2.5), by (1.3) and (2.2) we obtain

$$\sum_{n=1}^{\infty} P\{|U^*(n) - E(U^*(n))| > \varepsilon\} \leq \frac{\alpha^{2-\delta} (C(\delta) + 1)}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k_{i(n)+1}^{2-\delta}} \sum_{j=1}^{k_{i(n)+1}} E(Y_j^2) < \infty$$

for every $\varepsilon > 0$. Thus, again by the Borel–Cantelli lemma,

$$\lim_{n \rightarrow \infty} (U^*(n) - E(U^*(n))) = 0 \quad \text{a.s.},$$

and by the same argument,

$$\lim_{n \rightarrow \infty} (L^*(n) - E(L^*(n))) = 0 \quad \text{a.s.}$$

Therefore, (2.7) yields the almost sure inequalities

$$\liminf_{n \rightarrow \infty} E(L^*(n)) \leq \liminf_{n \rightarrow \infty} \mu_n \leq \limsup_{n \rightarrow \infty} \mu_n \leq \limsup_{n \rightarrow \infty} E(U^*(n)). \tag{2.8}$$

Finally, using (1.9), (2.5) and an easy argument involving Cesàro summation,

$$\begin{aligned} \liminf_{n \rightarrow \infty} E(L^*(n)) &= \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{k_{i(n)}} E(Y_j) \\ &\geq \liminf_{n \rightarrow \infty} \frac{k_{i(n)}}{k_{i(n)+1}} \frac{1}{k_{i(n)}} \sum_{j=1}^{k_{i(n)}} E(Y_j) = \frac{1}{\alpha} E(X), \end{aligned}$$

and similarly,

$$\limsup_{n \rightarrow \infty} E(U^*(n)) \leq \alpha E(X).$$

Since $\alpha > 1$ is as close to 1 as we wish, these two inequalities and those in (2.8) give (1.7) and hence the first statement of the theorem. \square

Even though the variance is finite, when proving (1.7) under (1.4) we still use the same truncation. Elaborating further on Etemadi’s (1981) argument one can easily show that under $E(X^2) < \infty$,

$$\sum_{n=1}^{\infty} \frac{1}{k_n l_r^\delta(k_n)} \sum_{j=1}^{k_n} E(Y_j^2) < \infty$$

for every fixed $r = 1, 2, \dots$ and $\delta > 0$, and if $C_r(\delta)$ denotes the finite supremum of the sequence $\{l_r^\delta(n)/m(n)\}$ in (1.4) then by (2.3) again,

$$\sum_{n=1}^{\infty} P\{|w_n(j)| > \varepsilon\} \leq \frac{1}{\varepsilon^2} \left\{ C_r(\delta) \sum_{n=1}^{\infty} \frac{1}{n l_r^\delta(n)} + \sum_{n=1}^{\infty} \frac{1}{n^2} \right\} < \infty$$

for every fixed $j \geq 1$, $r \geq 1$, $\delta > 0$, and $\varepsilon > 0$. In fact, in both relations one can replace the sequences $l_r^\delta(\cdot)$ by any slowly varying sequence $l(\cdot)$ satisfying (1.10). Using these two inequalities in place of (2.2) and (2.4), or the corresponding two inequalities with the general $l(\cdot)$ function, respectively, the same proof as above gives (1.7) under (1.4) and hence the second statement of the theorem. \square

Finally, to prove the third statement of the theorem, we prove (1.8) under (1.5). We can still assume without loss of generality that $X \geq 0$, but modify the truncation by introducing the triangular array

$$Y_{j,n} = X_j I(X_j \leq n \wedge m(n)), \quad j = 1, \dots, n, \quad n = 1, 2, \dots,$$

where $n \wedge m = \min(n, m)$. We have

$$E\left(\sum_{j=1}^n w_n(j) Y_{j,n}\right) = \int_0^{n \wedge m(n)} x \, dF(x) \rightarrow E(X) \quad \text{as } n \rightarrow \infty. \quad (2.9)$$

Thus, on the one hand, by non-negativity and the Markov inequality,

$$0 \leq \sum_{j=1}^n w_n(j) X_j - \sum_{j=1}^n w_n(j) Y_{j,n} \rightarrow 0 \quad \text{in probability as } n \rightarrow \infty,$$

and hence it is enough to show that

$$\sum_{j=1}^n w_n(j) Y_{j,n} \rightarrow E(X) \quad \text{in probability as } n \rightarrow \infty.$$

To show this via the Chebyshev inequality, (2.9) implies, on the other hand, that it is enough to see that the variance of the latter sequence converges to zero as $n \rightarrow \infty$. But by (2.3), applied with $k = n$ and $Y_{1,n}, \dots, Y_{n,n}$ replacing Y_1, \dots, Y_n , we have

$$\text{Var}\left(\sum_{j=1}^n w_n(j) Y_{j,n}\right) \leq \frac{1}{n} \frac{2}{n \wedge m(n)} \sum_{j=1}^n E(Y_{j,n}^2) = 2 \int_0^{n \wedge m(n)} \frac{x}{n \wedge m(n)} x \, dF(x),$$

and this upper bound goes to zero by the dominated convergence theorem of Lebesgue. \square

References

- Arcones, M.A. and E. Giné (1989), The bootstrap of the mean with arbitrary bootstrap sample size, *Ann. Inst. H. Poincaré* **25**, 457–481.
- Athreya, K.B. (1983), Strong law for the bootstrap, *Statist. Probab. Lett.* **1**, 147–150.
- Bickel, P.J. and D.A. Freedman (1981), Some asymptotic theory for the bootstrap, *Ann. Statist.* **9**, 1196–1217.
- Billingsley, P. (1986), *Probability and Measure* (Wiley, New York, 2nd ed.).
- Burke, M.D. and E. Gombay (1988), On goodness-of-fit and the bootstrap, *Statist. Probab. Lett.* **6**, 287–293.
- Csörgő, S., K. Tandori and V. Totik (1983), On the strong law of large numbers for pairwise independent random variables, *Acta. Math. Hungar.* **42**, 319–330.
- Etemadi, N. (1981), An elementary proof of the strong law of large numbers, *Z. Wahrsch. Verw. Gebiete* **55**, 119–122.
- Khinchin, A. (1929), Sur la loi des grandes nombres, *Comptes Rendus Acad. Sci. Paris* **188**, 477–479.
- Stout, W.F. (1974), *Almost Sure Convergence* (Academic Press, New York).