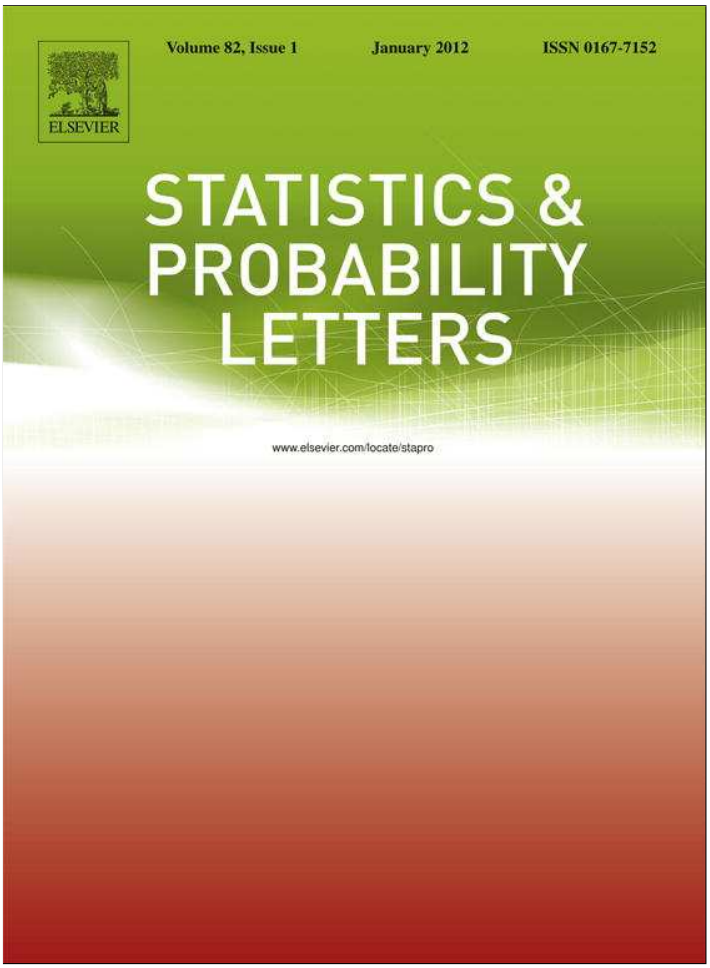
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Statistics and Probability Letters 82 (2012) 965–971

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A law of the single logarithm for weighted sums of arrays applied to bootstrap model selection in regression

Pierre Lafaye de Micheaux*, Christian Léger

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128 Succursale Centre-ville, Montréal, QC, H3C 3J7, Canada

ARTICLE INFO

Article history:
Received 23 August 2011
Received in revised form 18 January 2012
Accepted 18 January 2012
Available online 2 February 2012

MSC:
60B12
60F15
60G50

Keywords:
BIC
Linear regression
Variable selection
Rowwise independent
Triangular arrays

ABSTRACT

We generalize a law of the single logarithm obtained by Qi (1994) and Li et al. (1995) to the case of weighted sums of triangular arrays of random variables. We apply this result to bootstrapping the all-subsets model selection problem in regression, where we show that the popular Bayesian Information Criterion of Schwarz (1978) is no longer asymptotically consistent.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

According to van der Vaart (1998), "The law of the iterated logarithm is an intriguing result but appears to be of less interest to statisticians". Whether or not this is true, statisticians sometimes need to know bounds on the almost sure size of means or linear combinations of random variables to establish certain statistical results, and results concerning the law of the iterated (or single) logarithm are then needed. In studying model selection in regression in a bootstrap context, we needed such a result, but to our surprise none of the large number of existing results applied to our problem. We first begin by giving some background on the statistical application. Then we proceed with the probability result which is of independent interest, and adds to the large literature on the subject. In particular, we are looking at a triangular array result for the almost sure size of $S_n = \sum_{i=1}^{n} a_{n,i}X_{n,i}$ where $X_{n,1}, \ldots, X_{n,n}$ are independent and identically distributed (i.i.d.) random variables from distribution F_n with mean 0 and finite variance and some extra conditions on the moments of order $4 + \delta$ with $\delta > 0$ and where the triangular array of constants $\{a_{n,i}\}$ satisfies conditions typical of a regression context. This is a generalization of a result obtained simultaneously and independently by Qi (1994) and Li et al. (1995), and it turns out that the size of $|S_n|$ is $(n \log n)^{1/2}$. Depending whether the sum S_n is weighted or unweighted ($a_{n,i} \equiv 1$), whether or not there is a single sequence $X_1, \ldots, X_n$ or a triangular array of independent random variables, and depending on conditions on the scalars $a_{n,i}$ and on the moments of the distribution of the random variables, sometimes the rate is a single logarithm $(n \log n)^{1/2}$, other times an iterated logarithm $(n \log \log n)^{1/2}$. See, for instance, Sung (2009), Li et al. (2009), Ahmed et al. (2001), Hu and Weber (1992), and Lai and Wei (1982). The single logarithm rate that we have here, as opposed to the iterated logarithm rate that holds if there is a single sequence of random variables $X_1, X_2, \ldots, X_n$ i.i.d. from F, has important statistical implications: whereas

* Corresponding author. Tel.: +1 5143436607.
E-mail address: lafaye@dms.umontreal.ca (P. Lafaye de Micheaux).

0167-7152/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2012.01.018

Author's personal copy

the Bayesian Information Criterion (BIC) of Schwarz (1978) is a consistent method of model selection in regression, it is no longer the case for bootstrap data. Consequently, statisticians do need to pay attention to the law of the iterated (or single) logarithm and associated results!

2. Consistent model selection and the bootstrap

Before considering the bootstrap context, we begin with model selection in the usual multiple linear regression context. Consider the sequence of embedded models

$$Y_n = X_{(k),n}\beta_{(k)} + \epsilon_n \tag{1}$$

where ϵ_n is a sample of i.i.d. random variables from distribution F with mean 0 and finite variance (extra conditions will be necessary later), $X_{(k),n} = (x_{1,n}, \ldots, x_{k,n})$ is a matrix of size $n \times k$, $\beta_{(k)}$ is a k-vector, and $k = 1, 2, \ldots, p$. For simplicity, we will often drop the subscript n, e.g., using $X_{(k)}$. We assume that k_0 identifies the true model. That is, we assume without loss of generality (see Rao and Wu (1989)) that the independent variables are so ordered that if one considers the full model (1) with $k = p$, the model that gave rise to the data is such that the first k_0 components of $\beta_{(p)}$ are non-zero whereas the last $p - k_0$ are zero. The search for the true model therefore consists of choosing the order k.

Rao and Wu (1989) look at the problem of consistently choosing the regression model by minimizing a criterion computed for each possible model. Many authors have introduced criteria and most are (at least asymptotically) equivalent to

$$D_n(k) = S_{(k)} + k\hat{\sigma}^2_{(p)}C_n, \tag{2}$$

where $S_{(k)} = Y_n'(I - P_{(k)})Y_n$ is the sum of squared residuals of model k, $\hat{\sigma}^2_{(p)}$ is the (strongly consistent and unbiased) estimator of the error variance from the full model, and C_n is a constant depending on n which, through the multiplier k, penalizes larger models. Here, $P_{(k)} = X_{(k)}(X_{(k)}'X_{(k)})^{-1}X_{(k)}'$ is the projection matrix on the column space of $X_{(k)}$. A large value of C_n favors smaller models which will therefore ensure that unnecessary variables will be deleted whereas a small value of C_n favors larger models ensuring that all important variables will be included. We say that a model selection method is consistent if the selected model converges to the true model with probability 1. To have an asymptotically consistent model selection, $D_n(k)$ must be larger than $D_n(k_0)$ for all $k \neq k_0$ and all large values of n. It turns out that the size of $X_{(j)}'\epsilon_n$ plays a key role in the difference $D_n(k) - D_n(k_0)$.

For simplicity, we assume the classical condition: $n^{-1}X_{(p)}'X_{(p)}$ converges to a fixed $p \times p$ positive definite matrix G; weaker conditions can be imposed; see Rao and Wu (1989). This condition implies a number of linear algebra results, including $X_{(k_0)}'(I - P_{(k_0-1)})X_{(k_0)} \geq a_1 n$ for a positive constant a_1. Let d_n be a sequence of constants that will be used in defining a rate of convergence. As shown in Rao and Wu (1989), if

$$X_{(j)}'\epsilon_n = O(nd_n)^{1/2} \quad \text{a.s.} \tag{3}$$

then linear algebra and Cauchy–Schwarz's inequality imply that for $j = 1, \ldots, p$,

$$\epsilon_n'P_{(j)}\epsilon_n = O(d_n) \quad \text{a.s.} \quad \text{and} \quad \epsilon_n'P_{(k_0-1)}X_{(k_0)} = O(nd_n)^{1/2} \quad \text{a.s.} \tag{4}$$

To find the conditions on C_n which will ensure consistent model selection, consider first the comparison of the criterion for a model of order $k < k_0$ which leaves out important variables. Then it can be shown that

$$D_n(k) - D_n(k_0) \geq \beta^2_{(k_0)}X_{(k_0)}'(I - P_{(k_0-1)})X_{(k_0)} + 2\beta_{(k_0)}\epsilon_n'(I - P_{(k_0-1)})X_{(k_0)} - (k_0 - k)C_n\hat{\sigma}^2_{(p)} = T_1 + T_2 + T_3. \tag{5}$$

As argued above, $T_1 \geq \beta^2_{(k_0)}a_1 n > 0$. The negative term T_3 will be dominated by the positive term T_1 provided that $C_n = o(n)$ which is the upper bound rate on C_n. Assuming that (3) holds, using (4) $T_2 = O(nd_n)^{1/2}$ a.s. so that $D_n(k) - D_n(k_0) \geq 0$ a.s. for n large, provided that $d_n = o(n)$. Indeed, through a law of the iterated logarithm type result, Rao and Wu (1989) show that (3) holds for $d_n = \log\log n$, i.e., $T_2 = O(n\log\log n)^{1/2}$, leading to $D_n(k) - D_n(k_0) \geq 0$ a.s. for n large. Hence, asymptotically $\hat{k}_n \geq k_0$ a.s.

We must now consider what happens when we are looking at a model which contains unnecessary variables, i.e., $k > k_0$. Then it can be shown that

$$D_n(k) - D_n(k_0) = (k - k_0)C_n\hat{\sigma}^2_{(p)} - \sum_{j=k_0+1}^{k} \epsilon_n'(P_{(j)} - P_{(j-1)})\epsilon_n = T_4 - T_5. \tag{6}$$

As discussed above, provided that (3) holds, $\epsilon_n'P_{(j)}\epsilon_n = O(d_n)$ a.s., so that $T_5 = O(d_n)$ a.s. Since T_4 is positive, C_n diverging faster than d_n will guarantee that asymptotically, $D_n(k_0)$ will be smaller than $D_n(k)$, leading to a consistent choice of the model. And as mentioned above, $d_n = \log\log n$. So, as Rao and Wu (1989) have shown, provided that C_n satisfies

$$(\log\log n)^{-1}C_n \to \infty \quad \text{and} \quad n^{-1}C_n \to 0, \tag{7}$$

minimizing $D_n(k)$ leads to consistent model selection. Note that $C_n = \log n$, asymptotically equivalent to the BIC rule of Schwarz (1978), satisfies conditions (7).

Author's personal copy

The probability result that we will be presenting became necessary when we investigated the behavior of bootstrap model selection to study the distribution of the regression estimator when the model is chosen from the data — that work will be presented elsewhere. More precisely, we were looking at conditions on C_n to guarantee an (conditionally on the observed data) almost sure model selector for bootstrap data. Here we use resampling of the errors as opposed to resampling of the pairs; see Efron and Tibshirani (1993). To apply the bootstrap in this context, one constructs bootstrap observations Y_n^* by using the model in (1) where the matrix of regressors is $X_{(\hat{k}_n)}$ with $\hat{k}_n$ the value minimizing $D_n(k)$ for the original regression data, the vector of regression coefficients is $\beta^*_{(\hat{k}_n)} = (X'_{(\hat{k}_n)}X_{(\hat{k}_n)})^{-1}X'_{(\hat{k}_n)}Y_n$ and the regression errors ϵ_n^* are i.i.d. from the empirical distribution of (centered) residuals from the chosen model $\hat{k}_n$. Note that if conditions (7) on C_n are satisfied, $\hat{k}_n$ computed on the original data will converge a.s. to k_0 and so the regressors of the bootstrap observations will asymptotically be exactly those of the true model. To work out the theory, we consider a triangular array of i.i.d. random variables $\epsilon_{n;i}$ from distribution F_n, $i = 1, \ldots, n$ where some conditions on F_n are imposed; see Theorem 1. Writing $D_n^*(k) = S^*_{(k)} + k(\hat{\sigma}^*_{(p)})^2C_n^*$ where $S^*_{(k)} = Y_n^{*\prime}(I - P_{(k)})Y_n^*$ and $(\hat{\sigma}^*_{(p)})^2$ is the unbiased estimator of variance in the (bootstrap) full model, the bootstrap choice of model is defined by minimizing criterion $D_n^*(k)$. Note that we put a star on C_n^* to indicate that the constant at the bootstrap level could be different from the constant C_n used with the original observations.

The bootstrap version of Eqs. (5) and (6) hold. As was previously the case, for $k < k_0$, $D_n^*(k) > D_n^*(k_0)$ provided that $X'_{(j)}\epsilon_n^* = O(nd_n)^{1/2}$ a.s. where $d_n = o(n)$, as long as $n^{-1}C_n^* \to 0$ whereas for $k > k_0$ it will also be the case provided that C_n^* diverges faster than d_n, therefore defining the lower bound on C_n^*. As we will see in Theorem 1, because of the triangular nature of the bootstrap random variables involved, $d_n = \log n$ instead of $\log\log n$, i.e., to ensure an almost sure convergence of the bootstrap selected model where the bootstrap data is generated from a consistent choice of the (original) data, we need that at the bootstrap level C_n^* satisfies

$$(\log n)^{-1}C_n^* \to \infty \quad \text{and} \quad n^{-1}C_n^* \to 0. \tag{8}$$

The single logarithm in (8) as opposed to the iterated logarithm in (7) has important statistical implications as the popular BIC criterion with $C_n^* = \log n$ no longer satisfies these conditions and so BIC would not consistently choose the model at the bootstrap level. Interestingly, if one is willing to live with the weaker condition that $\hat{k}_n^* \to k_0$ in probability instead of a.s., then it is possible to show that $(\log n \log n)^{-1}C_n^* \to \infty$ is sufficient.

3. Main result

Now we state and prove our main result. This is an extension of the result of Qi (1994) and Li et al. (1995). Our method of proof follows that of Qi (1994). Note that in the following theorem, the fixed regressors $X_{(j)}$ of our example will be scalars $a_{n,i}$ whereas the random variables $\epsilon_{n,i}$ become $X_{n,i}$.

Theorem 1. *Let $X_{n,1}, \ldots, X_{n,n}$ be i.i.d. random variables with distribution function F_n. We suppose that $\mathbb{E}(X_{n,1}) = 0$,*

$$\mathbb{E}({X_{n,1}}^2) = \sigma_n^2 \to \sigma^2 \in (0, \infty) \quad \textit{when } n \to \infty, \tag{9}$$

and

$$\exists \delta > 0, \quad \sup_{n\geq 1}\mathbb{E}|X_{n,1}|^{4+\delta} < \infty. \tag{10}$$

Let $a_{n,i}$ be a triangular array and let $S_n = \sum_{i=1}^n a_{n,i}X_{n,i}$. We define

$$v_n = \mathbb{V}\text{ar}(S_n) = \sigma_n^2 \sum_{i=1}^n a_{n,i}^2. \tag{11}$$

If

$$\sup_{n\geq 1} \max_{1\leq i\leq n} |a_{n,i}| < \infty \tag{12}$$

and there exist two positive constants b_1 and b_2 such that

$$b_1 n \leq \sum_{i=1}^n a_{n,i}^2 \leq b_2 n, \quad n \geq 1, \tag{13}$$

then

$$\frac{|S_n|}{\sqrt{2v_n \log v_n}} = O(1), \quad \textit{a.s.} \tag{14}$$

Remark 1. In the regression example, if $n^{-1}X'_{(p)}X_{(p)}$ converges to a fixed $p \times p$ positive definite matrix G, then the columns of $X_{(p)}$ automatically satisfy conditions (12) and (13). Note also that condition (12) does *not* imply the existence of a smaller bound in (13), although it does imply an upper bound.

Author's personal copy

Remark 2. Qi (1994) obtains the exact constant in the right hand side of (14) when $F_n = F$, $a_{n,i} = 1$, for all i and n and under the necessary and sufficient condition that $\mathbb{E}[|X_{n,1}|^4/\log^2(\max|X_{n,1}|, e)] < \infty$. When the distribution F_n changes with n, his argument cannot be adapted and we have had to assume a bound on moments of order $4+\delta$ (condition (10) is equivalent to the existence of a $\gamma > 0$ such that $\sup_{n\geq 1} \mathbb{E}[|X_{n,1}|^{4+\gamma}/\log^2(\max(|X_{n,1}|, e))] < \infty$, but the former is more natural).

Proof. Noting that $|x|^4 \leq 1 + |x|^{4+\delta}$, it is easy to see that (10) implies that there exists a finite constant C_1 such that

$$\mathbb{E}|X_{n,1}|^4 < C_1 \quad \forall n, \tag{15}$$

which in turn implies that there exists a finite constant C_2 such that

$$\mathbb{E}|X_{n,1}|^3 < C_2 \quad \forall n. \tag{16}$$

Also (12) implies that there exists a finite constant C_a such that

$$\max_{i,n} |a_{n,i}| < C_a \quad \forall n, i. \tag{17}$$

In order to prove the theorem, we will show that for all ϵ positive and small enough,

$$\mathrm{P}(S_n > (1+\epsilon)\sqrt{2v_n \log v_n}, \text{ infinitely often}) = 0,$$

and a similar argument will be used to conclude that

$$\mathrm{P}(S_n < -(1+\epsilon)\sqrt{2v_n \log v_n}, \text{ infinitely often}) = 0.$$

We will use the Borel–Cantelli lemma after showing that

$$\sum_{n=1}^{\infty} \mathrm{P}(S_n > (1+\epsilon)\sqrt{2v_n \log v_n}) < \infty. \tag{18}$$

For a constant θ that will be determined later, let $Y_{n,i} = X_{n,i}\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n \log v_n})$, $\tilde{X}_{n,i} = Y_{n,i} - \mathbb{E}(Y_{n,i})$ and $\tilde{S}_n = \sum_{i=1}^n a_{n,i}Y_{n,i}$, where $\mathbb{1}(\cdot)$ is the indicator function. Then

$$\begin{aligned}
\mathrm{P}(S_n > (1+\epsilon)\sqrt{2v_n \log v_n}) &= \mathrm{P}\left(\{S_n > (1+\epsilon)\sqrt{2v_n \log v_n}\} \cap \left\{\max_{1\leq i\leq n} |X_{n,i}| > \theta\sqrt{v_n \log v_n}\right\}\right) \\
&\quad + \mathrm{P}\left(\{S_n > (1+\epsilon)\sqrt{2v_n \log v_n}\} \cap \left\{\max_{1\leq i\leq n} |X_{n,i}| \leq \theta\sqrt{v_n \log v_n}\right\}\right) \\
&\leq \mathrm{P}\left(\max_{1\leq i\leq n} |X_{n,i}| > \theta\sqrt{v_n \log v_n}\right) + \mathrm{P}\left(\{S_n > (1+\epsilon)\sqrt{2v_n \log v_n}\}\right. \\
&\quad \left. \cap \left\{\max_{1\leq i\leq n} |X_{n,i}| \leq \theta\sqrt{v_n \log v_n}\right\}\right) \\
&\leq \mathrm{P}\left(\max_{1\leq i\leq n} |X_{n,i}| > \theta\sqrt{v_n \log v_n}\right) + \mathrm{P}(\tilde{S}_n > (1+\epsilon)\sqrt{2v_n \log v_n}).
\end{aligned}$$

Thus, (18) will be proved if we can show that

$$\sum_{n=1}^{\infty} \mathrm{P}(\tilde{S}_n > (1+\epsilon)\sqrt{2v_n \log v_n}) < \infty \tag{19}$$

and that

$$\sum_{n=1}^{\infty} \mathrm{P}\left(\max_{1\leq i\leq n} |X_{n,i}| > \theta\sqrt{v_n \log v_n}\right) < \infty. \tag{20}$$

First, we consider (19). Since $\mathbb{E}(X_{n,i}) = 0$, then

$$\begin{aligned}
|\mathbb{E}(a_{n,i}Y_{n,i})| &= |\mathbb{E}[a_{n,i}X_{n,i}\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n \log v_n})]| = |\mathbb{E}[a_{n,i}X_{n,i}\mathbb{1}(|X_{n,i}| > \theta\sqrt{v_n \log v_n})]| \\
&\leq |a_{n,i}|\mathbb{E}|X_{n,i}\mathbb{1}(|X_{n,i}| > \theta\sqrt{v_n \log v_n})| \\
&\leq |a_{n,i}|(\mathbb{E}|X_{n,i}|^3)^{1/3}(\mathbb{E}[\mathbb{1}^{3/2}(|X_{n,i}| > \theta\sqrt{v_n \log v_n})])^{2/3} \quad \text{by Hölder's inequality} \\
&\leq C_2^{1/3}|a_{n,i}|\{\mathrm{P}(|X_{n,i}| > \theta\sqrt{v_n \log v_n})\}^{2/3} \quad \text{by (16)}
\end{aligned}$$

Author's personal copy

$$\leq C_2^{1/3}|a_{n,i}|\left(\frac{\mathbb{E}|X_{n,i}|^3}{\theta^3(v_n\log v_n)^{3/2}}\right)^{2/3} \quad \text{by Markov's inequality}$$
$$\leq \frac{C_2|a_{n,i}|}{\theta^2 v_n\log v_n} \quad \text{by (16)}.$$

Thus

$$|\mathbb{E}(\tilde{S}_n)| = \left|\sum_{i=1}^n \mathbb{E}(a_{n,i}Y_{n,i})\right| \leq \sum_{i=1}^n |\mathbb{E}(a_{n,i}Y_{n,i})| \leq \frac{C_2\sum_{i=1}^n |a_{n,i}|}{\theta^2 v_n\log v_n} \leq \frac{C_2 n C_a}{\theta^2 v_n\log v_n} \quad \text{by (17)}.$$

For n large enough, this last term is bounded since $v_n \geq b_1 n\sigma^2/2$ by (9), (11) and (13) and thus

$$\begin{aligned}\mathrm{P}(\tilde{S}_n > (1+\epsilon)\sqrt{2v_n\log v_n}) &= \mathrm{P}(\tilde{S}_n - \epsilon/2\sqrt{2v_n\log v_n} > (1+\epsilon/2)\sqrt{2v_n\log v_n})\\ &\leq \mathrm{P}(\tilde{S}_n - \mathbb{E}(\tilde{S}_n) > (1+\epsilon/2)\sqrt{2v_n\log v_n}).\end{aligned}$$

So proving

$$\sum_{n=1}^{\infty}\mathrm{P}(\tilde{S}_n - \mathbb{E}(\tilde{S}_n) > (1+\epsilon)\sqrt{2v_n\log v_n}) < \infty$$

for all ϵ small enough will be sufficient to conclude that (19) is true.

Let $t = \sqrt{2\log v_n/v_n}$ and $0 < \epsilon < 1$. Using inequality (11) from Qi (1994):

$$\exp(ax) \leq 1 + ax + \frac{1+\epsilon}{2}a^2x^2 + \frac{a^4}{\epsilon^4}x^4 + \frac{|a|^5|x|^5}{5!}\exp(|ax|),$$

we obtain

$$\begin{aligned}\mathbb{E}(\exp(ta_{n,i}\tilde{X}_{n,i})) &\leq 1 + ta_{n,i}\mathbb{E}(\tilde{X}_{n,i}) + \frac{1+\epsilon}{2}t^2a_{n,i}^2\mathbb{E}(\tilde{X}_{n,i}^2) + \frac{t^4a_{n,i}^4\mathbb{E}(\tilde{X}_{n,i}^4)}{\epsilon^4} + \frac{t^5|a_{n,i}|^5}{5!}\mathbb{E}(|\tilde{X}_{n,i}|^5\exp(|ta_{n,i}\tilde{X}_{n,i}|))\\ &\leq 1 + \frac{1+\epsilon}{2}t^2a_{n,i}^2\sigma_n^2 + \frac{t^4a_{n,i}^4\mathbb{E}(\tilde{X}_{n,i}^4)}{\epsilon^4} + \frac{t^5|a_{n,i}|^5}{5!}\mathbb{E}(|\tilde{X}_{n,i}|^5\exp(|ta_{n,i}\tilde{X}_{n,i}|))\\ &= 1 + \frac{1+\epsilon}{2}t^2a_{n,i}^2\sigma_n^2 + T_{n,i} + U_{n,i}, \qquad (21)\end{aligned}$$

since $\mathbb{E}(\tilde{X}_{n,i}) = 0$, $\mathbb{E}(\tilde{X}_{n,i}^2) = \mathbb{V}\mathrm{ar}(Y_{n,i}) \leq \mathbb{E}(Y_{n,i}^2) \leq \mathbb{E}(X_{n,i}^2) = \sigma_n^2$ and where we define

$$T_{n,i} = \frac{t^4a_{n,i}^4\mathbb{E}(\tilde{X}_{n,i}^4)}{\epsilon^4}$$

and

$$U_{n,i} = \frac{t^5|a_{n,i}|^5}{5!}\mathbb{E}(|\tilde{X}_{n,i}|^5\exp(|ta_{n,i}\tilde{X}_{n,i}|)).$$

Now, let us bound $\mathbb{E}(\tilde{X}_{n,i}^4)$ in terms of $\mathbb{E}(X_{n,i}^4)$. First, using the C_r inequality (Shorack, 2000, p. 47), we get $(a-b)^4 \leq 8(a^4+b^4)$. Thus,

$$\begin{aligned}\mathbb{E}(\tilde{X}_{n,i}^4) &= \mathbb{E}\{X_{n,i}\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n\log v_n}) - \mathbb{E}[X_{n,i}\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n\log v_n})]\}^4\\ &\leq 8\mathbb{E}\{X_{n,i}^4\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n\log v_n}) + \mathbb{E}^4[X_{n,i}\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n\log v_n})]\}\\ &\leq 8\mathbb{E}\{X_{n,i}^4\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n\log v_n}) + \mathbb{E}[X_{n,i}^4\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n\log v_n})]\} \quad \text{using Jensen's inequality}\\ &= 16\mathbb{E}[X_{n,i}^4\mathbb{1}(|X_{n,i}| \leq \theta\sqrt{v_n\log v_n})] \leq 16\mathbb{E}[X_{n,i}^4].\end{aligned}$$

So,

$$T_{n,i} = \frac{t^4a_{n,i}^4\mathbb{E}(\tilde{X}_{n,i}^4)}{\epsilon^4} = \frac{4(\log v_n)^2a_{n,i}^4}{\epsilon^4v_n^2}\mathbb{E}(\tilde{X}_{n,i}^4) \leq \frac{64(\log v_n)^2a_{n,i}^4}{\epsilon^4v_n^2}\mathbb{E}[X_{n,i}^4] \leq \frac{64C_1C_a^4(\log v_n)^2}{\epsilon^4v_n^2},$$

using (15) and (17).

Author's personal copy

Let us consider the term $U_{n,i}$. We have $|\tilde{X}_{n,i}| = |Y_{n,i} - \mathbb{E}(Y_{n,i})| \leq |Y_{n,i}| + |\mathbb{E}(Y_{n,i})| \leq 2\theta\sqrt{v_n \log v_n}$. This implies that

$$\begin{aligned}
U_{n,i} &= (5!)^{-1}\mathbb{E}(|ta_{n,i}\tilde{X}_{n,i}|^5 \exp(|ta_{n,i}\tilde{X}_{n,i}|)) = (5!)^{-1}\mathbb{E}(|ta_{n,i}\tilde{X}_{n,i}| \exp(|ta_{n,i}\tilde{X}_{n,i}|)|ta_{n,i}\tilde{X}_{n,i}|^4) \\
&\leq (5!)^{-1}2\theta t\sqrt{v_n \log v_n}|a_{n,i}| \exp(2\theta t|a_{n,i}|\sqrt{v_n \log v_n})t^4 a_{n,i}^4 \mathbb{E}(\tilde{X}_{n,i}^4) \\
&\leq (5!)^{-1}2\theta t^5\sqrt{v_n \log v_n}|a_{n,i}|^5 \exp(2\theta t|a_{n,i}|\sqrt{v_n \log v_n})16C_1 \\
&= \frac{32\theta C_1}{5!}t^5|a_{n,i}|^5\sqrt{v_n \log v_n} \exp(2\theta t|a_{n,i}|\sqrt{v_n \log v_n}).
\end{aligned}$$

Using the value of t and choosing $\theta = 1/(4\sqrt{2}C_a)$, we get

$$\begin{aligned}
\exp(2\theta t|a_{n,i}|\sqrt{v_n \log v_n}) &= \exp\{2(4\sqrt{2}C_a)^{-1}\sqrt{2\log v_n/v_n}|a_{n,i}|\sqrt{v_n \log v_n}\} \\
&= \exp\{(1/(2C_a))|a_{n,i}| \log v_n\} = v_n^{|a_{n,i}|/(2C_a)}.
\end{aligned}$$

Thus

$$\begin{aligned}
U_{n,i} &\leq \frac{32C_1 2^{5/2}}{5!4\sqrt{2}C_a}|a_{n,i}|^5\frac{(\log v_n)^3}{v_n^2}v_n^{|a_{n,i}|/(2C_a)} \leq \frac{32C_1 2^{5/2}}{5!4\sqrt{2}C_a}|a_{n,i}|^5\frac{(\log v_n)^3}{v_n^{3/2}} \\
&\leq \frac{4}{15}C_1 C_a^4\frac{(\log v_n)^3}{v_n^{3/2}}.
\end{aligned}$$

Here we used again (15) and (17).

Thus, from (21) and using the inequality $1 + x \leq \exp(x)$, we obtain

$$\begin{aligned}
\mathbb{E}(\exp(ta_{n,i}\tilde{X}_{n,i})) &\leq \exp\left\{\frac{1+\epsilon}{2}t^2 a_{n,i}^2\sigma_n^2 + \frac{64C_1C_a^4(\log v_n)^2}{\epsilon^4 v_n^2} + \frac{4}{15}C_1C_a^4\frac{(\log v_n)^3}{v_n^{3/2}}\right\} \\
&= \exp\left\{\frac{1+\epsilon}{2}t^2 a_{n,i}^2\sigma_n^2\right\}\exp\left\{\frac{64C_1C_a^4(\log v_n)^2}{\epsilon^4 v_n^2} + \frac{4}{15}C_1C_a^4\frac{(\log v_n)^3}{v_n^{3/2}}\right\}.
\end{aligned}$$

Now, noting that $\tilde{S}_n - \mathbb{E}(\tilde{S}_n) = \sum_{i=1}^n a_{n,i}\tilde{X}_{n,i}$ and by independence of the $\tilde{X}_{n,i}$, we have

$$\begin{aligned}
&\mathrm{P}(\tilde{S}_n - \mathbb{E}(\tilde{S}_n) > (1+\epsilon)\sqrt{2v_n \log v_n}) \\
&= \mathrm{P}[\tilde{S}_n - \mathbb{E}(\tilde{S}_n) > (1+\epsilon)tv_n] \\
&= \mathrm{P}[\exp\{t(\tilde{S}_n - \mathbb{E}(\tilde{S}_n))\} > \exp\{(1+\epsilon)t^2 v_n\}] \\
&\leq \exp\{-(1+\epsilon)t^2 v_n\}\mathbb{E}[\exp\{t(\tilde{S}_n - \mathbb{E}(\tilde{S}_n))\}] \text{ by Markov's inequality} \\
&\leq \exp\{-(1+\epsilon)2\log v_n\}\exp\left(\frac{1+\epsilon}{2}t^2\sigma_n^2\sum_{i=1}^n a_{n,i}^2\right)\exp\left(\frac{64C_1C_a^4 n(\log v_n)^2}{\epsilon^4 v_n^2} + \frac{4}{15}C_{1C_a}{}^4\frac{n(\log v_n)^3}{v_n^{3/2}}\right) \\
&= \exp\{-(1+\epsilon)2\log v_n\}\exp\left(\frac{1+\epsilon}{2}2\frac{\log v_n}{v_n}v_n\right)\exp\left(\frac{64C_1C_a^4 n(\log v_n)^2}{\epsilon^4 v_n^2} + \frac{4}{15}C_1C_a^4\frac{n(\log v_n)^3}{v_n^{3/2}}\right) \\
&= \exp\{-(1+\epsilon)\log v_n\}\exp\left(\frac{64C_1C_a^4 n(\log v_n)^2}{\epsilon^4 v_n^2} + \frac{4}{15}C_1C_a^4\frac{n(\log v_n)^3}{v_n^{3/2}}\right) \leq 2v_n^{-(1+\epsilon)}
\end{aligned}$$

for n large enough and considering (9), (11) and (13).

We thus conclude that $\sum_{n=1}^\infty \mathrm{P}(\tilde{S}_n - \mathbb{E}(\tilde{S}_n) > (1+\epsilon)\sqrt{2v_n \log v_n}) < \infty$.

Now we consider (20). We try to prove that

$$\sum_{n=1}^\infty \mathrm{P}\left(\max_{1\leq i\leq n}|X_{n,i}| > \theta\sqrt{v_n \log v_n}\right) < \infty.$$

But we have

$$\begin{aligned}
\mathrm{P}\left[\max_{1\leq i\leq n}|X_{n,i}| > \theta\sqrt{v_n \log v_n}\right] &= 1 - \mathrm{P}\left[\max_{1\leq i\leq n}|X_{n,i}| \leq \theta\sqrt{v_n \log v_n}\right] = 1 - \prod_{i=1}^n \mathrm{P}[|X_{n,i}| \leq \theta\sqrt{v_n \log v_n}] \\
&= 1 - \mathrm{P}[|X_{n,1}| \leq \theta\sqrt{v_n \log v_n}]^n = 1 - (1 - \mathrm{P}[|X_{n,1}| > \theta\sqrt{v_n \log v_n}])^n \\
&= 1 - \exp(\log(1 - \mathrm{P}[|X_{n,1}| > \theta\sqrt{v_n \log v_n}])^n) \\
&= 1 - \exp(n\log(1 - \mathrm{P}[|X_{n,1}| > \theta\sqrt{v_n \log v_n}])) \\
&\approx 1 - \exp(-n\mathrm{P}[|X_{n,1}| > \theta\sqrt{v_n \log v_n}]) \approx n\mathrm{P}[|X_{n,1}| > \theta\sqrt{v_n \log v_n}]
\end{aligned}$$

Author's personal copy

(where $a_n \approx b_n$ means $a_n/b_n \to 1$ when $n \to \infty$) because, using Chebyshev, $n\mathrm{P}[|X_{n,1}| > \theta\sqrt{v_n \log v_n}] \leq n\frac{\sigma_n^2}{\theta^2 v_n \log v_n} \to 0$ when $n \to \infty$.

Thus it suffices to show that (see Spivak, 2006, p. 468)

$$\sum_{n=1}^{\infty} n\mathrm{P}[|X_{n,1}| > \theta\sqrt{v_n \log v_n}] < \infty.$$

Now, since $v_n \geq b_1\sigma_n^2 n$, it suffices to show that

$$\sum_{n=1}^{\infty} n\mathrm{P}[|X_{n,1}| > C_3(n \log n)^{1/2}] < \infty,$$

where C_3 is a constant that depends on b_1, σ^2 and θ. Let $s(x) = x^{2+\delta/2}$ which is a strictly increasing function on $[0, +\infty)$, and where δ is the strictly positive real number hypothesized in the theorem.

So, $n\mathrm{P}(|X_{n,i}| > C_3\sqrt{n \log n}) = n\mathrm{P}(s(|X_{n,i}|) > s(C_3\sqrt{n \log n})) \leq n\mathbb{E}[s^2(|X_{n,i}|)]/s^2(C_3\sqrt{n \log n})$, using Markov. But

$$\frac{n\mathbb{E}[s^2(|X_{n,i}|)]}{s^2(C_3\sqrt{n \log n})} = \frac{n\mathbb{E}[s^2(|X_{n,i}|)]}{(C_3^2 n \log n)^{2+\delta/2}} = \frac{\mathbb{E}[s^2(|X_{n,i}|)]}{(n^{1+\delta/2}(C_3^2 \log n)^{2+\delta/2})} \leq \frac{1}{C_3^{4+\delta}} \frac{\mathbb{E}[s^2(|X_{n,i}|)]}{n^{1+\delta/2}}$$

(for n large enough) whose series is convergent since $\mathbb{E}[s^2(|X_{n,i}|)] = \mathbb{E}|X_{n,1}|^{4+\delta}$ is bounded by (10). □

Acknowledgment

We would like to thank the anonymous referee for his suggestions that improved the presentation of the main theorem.

References

Ahmed, S.E., Li, D., Rosalsky, A., Volodin, A.I., 2001. Almost sure lim sup behavior of bootstrapped means with applications to pairwise i.i.d. sequences and stationary ergodic sequences. J. Statist. Plann. Inference 98 (1–2), 1–14.
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall/CRC, Boca Raton, Florida.
Hu, T.C., Weber, N.C., 1992. On the rate of convergence in the strong law of large numbers for arrays. Bull. Austral. Math. Soc. 45 (3), 479–482.
Lai, T.L., Wei, C.Z., 1982. A law of the iterated logarithm for double arrays of independent random variables with applications to regression and time series models. Ann. Probab. 10 (2), 320–335.
Li, D., Qi, Y., Rosalsky, A., 2009. Iterated logarithm type behavior for weighted sums of i.i.d. random variables. Statist. Probab. Lett. 79 (5), 643–651.
Li, D.L., Rao, M.B., Tomkins, R.J., 1995. A strong law for *B*-valued arrays. Proc. Amer. Math. Soc. 123 (10), 3205–3212.
Qi, Y.C., 1994. On strong convergence of arrays. Bull. Austral. Math. Soc. 50 (2), 219–223.
Rao, C.R., Wu, Y., 1989. A strongly consistent procedure for model selection in a regression problem. Biometrika 76 (2), 369–374.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.
Shorack, G.R., 2000. Probability for Statisticians. Springer-Verlag, New York.
Spivak, M., 2006. Calculus, third ed. Cambridge University Press.
Sung, S.H., 2009. A law of the single logarithm for weighted sums of i.i.d. random elements. Statist. Probab. Lett. 79 (10), 1351–1357.
van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge.