Bull. Aust. Math. Soc. **97** (2018), 513–517
doi:10.1017/S0004972718000059

A SIMPLE PROOF OF THE STRONG LAW OF LARGE NUMBERS WITH RATES

NUNO LUZIA

(Received 6 November 2017; accepted 10 January 2018; first published online 28 March 2018)

Abstract

We give a simple proof of the strong law of large numbers with rates, assuming only finite variance. This note also serves as an elementary introduction to the theory of large deviations, assuming only finite variance, even when the random variables are not necessarily independent.

2010 *Mathematics subject classification*: primary 60F15; secondary 60F10.

Keywords and phrases: strong law of large numbers, rates of convergence, large deviations.

1. Introduction and statements

Suppose that $X_1, X_2, \ldots$ are independent, identically distributed random variables with mean μ and $\tau := \mathbb{E}|X_i - \mu| < \infty$. Let $S_n = X_1 + \cdots + X_n$.

The *strong law of large numbers* (SLLN) is a fundamental theorem in probability and statistics. It says that

$$\frac{S_n}{n} \to \mu \quad \text{a.s.},$$

where a.s. is an abbreviation for almost surely (see (1.2) for an equivalent statement). There are already, in the literature, elementary treatments of SLLN (for example, Etamadi's famous proof [2]), but they do not take into account the *rate of convergence*. In applications it is important to quantify the rate of convergence, so that, for a given n, we can estimate the error between S_n/n and μ. For that purpose, it is necessary to assume the existence of higher moments, say $\sigma^2 := \mathbb{E}(X_i - \mu)^2 < \infty$. In this note we will give an elementary proof of the SLLN with rates, when $\sigma^2 < \infty$. In subsection 1.2, we will also consider a version for not necessarily independent random variables.

It is well known that, when $\sigma^2 < \infty$, Chebyshev's inequality gives the weak law of large numbers *with rates*: for every $\varepsilon > 0$ and $n \geq 1$,

$$\mathbb{P}\left\{\left|\frac{1}{n}S_n - \mu\right| > \varepsilon\right\} \leq \frac{\sigma^2}{\varepsilon^2 n}. \tag{1.1}$$

The SLLN is equivalent to saying that, for each $\varepsilon > 0$,

$$\mathbb{P}\Big\{\sup_{k\geq n}\Big|\frac{1}{k}S_k-\mu\Big|>\varepsilon\Big\}\to 0 \quad \text{as } n\to\infty. \tag{1.2}$$

Note that, even when $\sigma^2 < \infty$, the SLLN does not follow from (1.1) because $\sum 1/n$ is a divergent series. We will prove the following SLLN *with rates*.

Theorem 1.1. *Assume that $\sigma^2 < \infty$. Let $\beta > 1$ and $0 < \varepsilon \leq 1$. There are constants C_β, $D_\beta > 0$ (depending only on β) such that, for every $n \geq n(\varepsilon,\beta)$,*

$$\mathbb{P}\Big\{\sup_{k\geq n}\Big|\frac{1}{k}S_k-\mu\Big|>\tau\varepsilon\Big\}\leq\frac{\sigma^2}{\tau^2\varepsilon^2\, n}(C_\beta+D_\beta(\log n)^{\beta-1}), \tag{1.3}$$

where

$$n(\varepsilon,\beta) := \max\{6\varepsilon^{-1},\exp((9\beta^{-1}\varepsilon^{-1})^{\beta^{-1}/(1-\beta^{-1})})\}.$$

Remark 1.2. From the proof of Theorem 1.1, we may take

$$C_\beta = 72+72\beta\lfloor\beta\rfloor! \quad \text{and} \quad D_\beta = 72\beta+72(e-1)\beta\lfloor\beta\rfloor!.$$

See how the right-hand side of (1.3) is almost as good as the right-hand side of (1.1), but it gives a much stronger kind of convergence: *almost sure convergence* instead of *convergence in probability*. We have introduced the parameter $\beta > 1$ for practical reasons. When β is big, the right-hand side of (1.3) gets worse but $n(\varepsilon,\beta)$ gets better, so we may choose the best β according to what is most favourable for a particular application.

In the specialised literature, there are much better estimates than (1.3) under the much stronger assumption that the *moment-generating function* $\varphi(t) := \mathbb{E}(e^{tX_i}) < \infty$ for some $t > 0$. Such estimates were initiated by Cramér, and fall within the theory known today as *large deviations* (see [1] for an introduction). We also note that, in this case, the large-deviation estimates depend heavily on the distribution of X_i.

Our estimates are *universal*, meaning that they depend only on σ^2, not on the specific distribution of X_i. This could be of practical importance when we are dealing with *fat-tailed* distributions, as long as they have finite variance.

1.1. Higher orders of convergence. A simple adaptation to the proof of Theorem 1.1 allows us to give rates of convergence of

$$\frac{S_n-n\mu}{n^\alpha}\to 0 \quad \text{a.s.}$$

for $2/3 < \alpha \leq 1$.

Theorem 1.3. *Assume that $\mu = 0$ and $\sigma^2 < \infty$. Let $2/3 < \alpha \leq 1$. Take any β with $\beta > (2\alpha-1)^{-1}$, $\beta(1-\alpha) < 1$ and $0 < \varepsilon \leq 1$. Then, for every $n \geq (9\beta^2\varepsilon^{-1})^{\beta/(1-\beta(1-\alpha))}$,*

$$\mathbb{P}\Big\{\sup_{k\geq n}\Big|\frac{S_k}{k^\alpha}\Big|>\tau\varepsilon\Big\}\leq\frac{72\sigma^2}{\tau^2\varepsilon^2}\Big(\frac{1}{n^{2\alpha-1}}+\frac{(\beta(2\alpha-1)-1)^{-1}}{n^{2\alpha-1-\beta^{-1}}}\Big). \tag{1.4}$$

Again, there are much better estimates than (1.4), whenever $\mathbb{E}(e^{tX_i}) < \infty$ for some $t > 0$, which depend heavily on the distribution of X_i (see [3, Section XVI.7] and [4]). Such estimates (for $1/2 < \alpha < 1$) are known as *moderate deviations*.

https://doi.org/10.1017/S0004972718000059 Published online by Cambridge University Press

1.2. Dependent random variables. In this subsection, the identically distributed random variables X_i are not necessarily independent.

Introduce the notation $X_i^+ = \max\{0, X_i\}$, $X_i^- = \max\{0, -X_i\}$, $S_n^{(+)} = X_1^+ + \cdots + X_n^+$ and $S_n^{(-)} = X_1^- + \cdots + X_n^-$. For example, if

$$\operatorname{var}(S_n^{(+)}) \le C\sigma^2 n \quad \text{and} \quad \operatorname{var}(S_n^{(-)}) \le C\sigma^2 n, \tag{1.5}$$

for some constant $C > 0$ and every $n \ge 1$, then Theorems 1.1 and 1.3 still hold (the proofs are exactly the same, we just have to multiply the right-hand sides of (1.3) and (1.4) by C). Condition (1.5) is satisfied if, for example,

$$|\mathbb{E}(X_i^\diamond X_{i+n}^\diamond) - (\mu^\diamond)^2| \le \frac{D}{n(\log n)^\gamma} \quad \text{for } \diamond = + \text{ and } \diamond = -,$$

for some constants $D > 0$, $\gamma > 1$ and every $i, n \ge 1$, where μ^+, μ^- are the means of X_i^+, X_i^-, respectively. Moreover, a straightforward modification in the proof of Theorem 1.3 gives the following result.

THEOREM 1.4. *Assume that* $\mu = 0$ *and* $\sigma^2 < \infty$. *Let* $1 < \theta < 2$ *and assume that*

$$\operatorname{var}(S_n^{(+)}) \le Cn^\theta, \quad \operatorname{var}(S_n^{(-)}) \le Cn^\theta, \tag{1.6}$$

for some $C > 0$ *and every* $n \ge 1$. *Suppose that* $(1/3)(1 + \theta) < \alpha \le 1$. *Take any* β *with* $\beta > (2\alpha - \theta)^{-1}$, $\beta(1 - \alpha) < 1$ *and* $0 < \varepsilon \le 1$. *Then, for every* $n \ge (9\beta^2\varepsilon^{-1})^{\beta/(1-\beta(1-\alpha))}$,

$$\mathbb{P}\Big\{\sup_{k\ge n}\Big|\frac{S_k}{k^\alpha}\Big| > \tau\varepsilon\Big\} \le \frac{72C}{\tau^2\varepsilon^2}\Big(\frac{1}{n^{2\alpha-\theta}} + \frac{(\beta(2\alpha-\theta)-1)^{-1}}{n^{2\alpha-\theta-\beta^{-1}}}\Big).$$

Condition (1.6) is satisfied if, for example,

$$|\mathbb{E}(X_i^\diamond X_{i+n}^\diamond) - (\mu^\diamond)^2| \le Dn^{\theta-2} \quad \text{for } \diamond = + \text{ and } \diamond = -,$$

for some constant $D > 0$ and every $i, n \ge 1$.

2. Proofs

2.1. Proof of Theorem 1.1. The proof follows the spirit of the proof of Theorem 1 of [5].

Since $X_i = X_i^+ - X_i^-$ and $\max\{X_i^+, X_i^-\} \le X_i^+ + X_i^- = |X_i|$, we can assume without loss of generality that $X_i \ge 0$ and $\mu > 0$ (we will give the details at the end of the proof).

Let $\beta > 1$ and $f(x) = e^{x^{\beta^{-1}}}$ for $x \ge 1$. Consider the subsequence $n_k = \lceil f(k)\rceil$, $k = 1, 2, \ldots$ (where, as usual, $\lceil x\rceil$ is the least integer $\ge x$). We will see that $\rho_k := n_{k+1}/n_k \to 1$ as $k \to \infty$.

By (1.1),

$$\mathbb{P}\Big\{\sup_{k\ge m}\Big|\frac{1}{\mu n_k}S_{n_k} - 1\Big| > \varepsilon\Big\} \le \frac{\sigma^2}{\mu^2\varepsilon^2}\sum_{k=m}^{\infty}\frac{1}{f(k)}.$$

https://doi.org/10.1017/S0004972718000059 Published online by Cambridge University Press

Using the integral test (and the change of variable $y = f(x)$), we can estimate this series from above by

$$\frac{1}{f(m)} + \beta \int_{f(m)}^{\infty} \frac{(\log y)^{\beta-1}}{y^2}\, dy.$$

Using integration by parts ($\lceil\beta\rceil - 1$ times), we see that the expression above is less than

$$\frac{1}{f(m)}(1 + \beta\lfloor\beta\rfloor! + (\beta + (e-1)\beta\lfloor\beta\rfloor!)(\log f(m))^{\beta-1}).$$

Now take any n and let k be such that $n_k \le n < n_{k+1}$. Since we are assuming that $X_i \ge 0$,

$$\frac{n_k}{n_{k+1}}\frac{1}{n_k}S_{n_k} \le \frac{1}{n}S_n \le \frac{n_{k+1}}{n_k}\frac{1}{n_{k+1}}S_{n_{k+1}} \tag{2.1}$$

and so

$$\left|\frac{1}{\mu n}S_n - 1\right| \le 3\varepsilon$$

if

$$\left|\frac{1}{\mu n_l}S_{n_l} - 1\right| \le \varepsilon, \quad \text{for } l = k, k+1, \quad \text{and} \quad \rho_k \le 1+\varepsilon,\ \rho_k^{-1} \ge 1-\varepsilon.$$

Let $h(x) = x^{\beta^{-1}}$. Then $h'(x) = \beta^{-1}x^{\beta^{-1}-1}$, which implies that

$$f(k+1)/f(k) \le e^{\beta^{-1}k^{\beta^{-1}-1}} \le 1 + \tfrac{3}{2}\beta^{-1}k^{\beta^{-1}-1} \le 1 + \tfrac{1}{2}\varepsilon$$

if $k \ge (3\beta^{-1}\varepsilon^{-1})^{(1-\beta^{-1})^{-1}}$. This implies that $\rho_k \le 1+\varepsilon$, and $\rho_k^{-1} \ge 1-\varepsilon$ if also $n_k \ge 2\varepsilon^{-1}$.

Putting all this together (and replacing 3ε by ε) yields

$$\mathbb{P}\Big\{\sup_{j\ge n}\left|\frac{1}{\mu j}S_j - 1\right| > \varepsilon\Big\} \le \frac{9\sigma^2}{\mu^2\varepsilon^2 n}(1 + \beta\lfloor\beta\rfloor! + (\beta + (e-1)\beta\lfloor\beta\rfloor!)(\log n)^{\beta-1})$$

for $n \ge n(\varepsilon,\beta)$.

Let us now explain in more detail why we could assume that $X_i \ge 0$. For general X_i, it is clear that we can assume, without loss of generality, that $\mu = 0$ (by working with $X_i - \mu$). Then we see that the means of X_i^+ and X_i^- satisfy $\mu^+ = \mu^- = \frac{1}{2}\tau$. Of course we may assume that $\mu^+ > 0$ (otherwise the random variables X_i are trivial). Then we apply the above proof twice, once for $S_n^{(+)} := X_1^+ + \cdots + X_n^+$ and then for $S_n^{(-)} := X_1^- + \cdots + X_n^-$. The result then follows easily for $S_n = S_n^+ - S_n^-$.

2.2. Proof of Theorem 1.3. As in the proof of Theorem 1.1, we can assume that $X_i \ge 0$ and $\mu > 0$.

By (1.1),

$$\mathbb{P}\Big\{n^{1-\alpha}\left|\frac{1}{\mu n}S_n - 1\right| > \varepsilon\Big\} \le \frac{\sigma^2}{\mu^2\varepsilon^2\, n^{2\alpha-1}}. \tag{2.2}$$

Let $f(x) = x^\beta$, $x \ge 1$, and $n_k = \lceil f(k)\rceil$, $k = 1, 2, \ldots$. Then we easily see (by using the integral test) that

$$\mathbb{P}\Big\{\sup_{k\ge m} n_k^{1-\alpha}\left|\frac{1}{\mu n_k}S_{n_k} - 1\right| > \varepsilon\Big\} \le \frac{\sigma^2}{\mu^2\varepsilon^2}\Big(\frac{1}{f(m)^{2\alpha-1}} + \frac{(\beta(2\alpha-1)-1)^{-1}}{f(m)^{2\alpha-1-\beta^{-1}}}\Big).$$

https://doi.org/10.1017/S0004972718000059 Published online by Cambridge University Press

Now if $n_k \le n < n_{k+1}$, by (2.1) we see that

$$n^{1-\alpha}\left|\frac{1}{\mu n}S_n - 1\right| \le 3\varepsilon \quad \text{if} \quad n_l^{1-\alpha}\left|\frac{1}{\mu n_l}S_{n_l} - 1\right| \le \varepsilon, \quad \text{for } l = k, k+1,$$

and $\rho_k \le 1 + \varepsilon n_k^{\alpha-1}$, $\rho_k^{-1} \ge 1 - \varepsilon n_k^{\alpha-1}$. We see that this last hypothesis is satisfied whenever $k \ge (3\beta^2\varepsilon^{-1})^{(1-\beta(1-\alpha))^{-1}}$. Consequently,

$$\mathbb{P}\left\{\sup_{j\ge n} j^{1-\alpha}\left|\frac{1}{\mu j}S_j - 1\right| > \varepsilon\right\} \le \frac{9\sigma^2}{\mu^2\varepsilon^2}\left(\frac{1}{n^{2\alpha-1}} + \frac{(\beta(2\alpha-1)-1)^{-1}}{n^{2\alpha-1-\beta^{-1}}}\right)$$

whenever $n \ge (9\beta^2\varepsilon^{-1})^{\beta/(1-\beta(1-\alpha))}$.

The general result follows by decomposing $X_i = X_i^+ - X_i^-$ and arguing as we did at the end of the proof of Theorem 1.1.

2.3. Proof of Theorem 1.4. This is a straightforward modification of the proof of Theorem 1.3. For example, instead of (2.2), we have now

$$\mathbb{P}\left\{n^{1-\alpha}\left|\frac{1}{\mu n}S_n - 1\right| > \varepsilon\right\} \le \frac{C}{\mu^2\varepsilon^2\, n^{2\alpha-\theta}}.$$

We leave the details as an exercise.

References

[1] R. Durrett, *Probability: Theory and Examples*, 2nd edn (Duxbury Press, Belmont, CA, 2004).

[2] N. Etamadi, 'An elementary proof of the strong law of large numbers', *Z. Wahrscheinlicheitstheorie verw. Geb.* **55** (1981), 119–122.

[3] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd edn, Vol. 2 (Wiley, New York, 1971).

[4] J. Fill, 'Convergence rates related to the strong law of large numbers', *Ann. Probab.* **11** (1983), 123–142.

[5] N. Luzia, 'A Borel–Cantelli lemma and its applications', *Trans. Amer. Math. Soc.* **366** (2014), 547–560.

NUNO LUZIA, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-909, Brazil
e-mail: nuno@im.ufrj.br

https://doi.org/10.1017/S0004972718000059 Published online by Cambridge University Press