A New Proof for a Strong Law of Large Numbers of Kolmogorov’s Type via Weak Convergence

Yu-Lin Chou*

Abstract

In terms of the Dirac representation of sample mean and the weak convergence of empirical distributions that holds almost surely, we construct a new proof for a strong law of large numbers of Kolmogorov’s type with i.i.d. random variables \(X_1, X_2, \ldots\) such that \(\lim_{c \to \infty} \sup_{n \in \mathbb{N}} n^{-1} \sum_{i=1}^{n} |X_i| \cdot I_{[c, +\infty]} \circ |X_i| = 0\) almost surely. That each random variable \(X_i\) is \(L^1\) is also a conclusion. Our proof is independent of both Kolmogorov’s strong law and its known proof(s), and potentially furnishes a new way to obtain a short proof of Kolmogorov’s strong law.

Keywords: classical Glivenko-Cantelli theorem for right closed rays; convergence of moments; convergence of probability measures; Dirac representation of sample mean; law of large numbers; uniform integrability

AMS MSC 2020: 60F15; 60B10; 62F99

1 Introduction

By Kolmogorov’s strong law of large numbers (KSLLN) we mean the statement that the sample mean of independent identically distributed (i.i.d.) random variables with finite mean converges almost surely to the common mean. The typical proof (for example, Section 3 of Chapter 4 in Shiryaev [2]) depends on several (important) results within the scope of elementary analysis; the length of the typical proof is then inevitably great.

It would be desirable to obtain a short proof of KSLLN that possibly and hopefully also gains new insight into the nature of the “problem”. Toward a new proof of KSLLN having the desired properties, we propose a new proof of a strong law with the original \(L^1\) assumption replaced by exactly one mild assumption on the tail-asymptotic behavior of the random variables; the original \(L^1\) assumption then becomes a conclusion.

On the basis of the Dirac representation of sample mean, our proof depends mostly on the classical Glivenko-Cantelli theorem (for right closed rays), which can in fact be obtained without first utilizing KSLLN, and a result connecting the weak convergence of probability measures with the convergence of the corresponding first moments. It turns out that this approach in principle leads to a short\(^1\) proof independent of KSLLN and its known proof(s). Moreover, apart from showing a new connection between strong laws of large numbers and the weak convergence theory, the proof in general points out another possibility for a weak convergence to imply an almost everywhere convergence; an existing well-known possibility is the Skorokhod’s representation theorem.\(^2\)

Our proof scheme is as follows. There is a version of strong law of large numbers, given in Theorem 1 in Section 3 of Chapter 4 in Shiryaev [2] with a short, elementary proof, which imposes a stronger moment condition than KSLLN — the uniform boundedness of the fourth moment. Since the empirical

*Author for correspondence: Yu-Lin Chou, Institute of Statistics, National Tsing Hua University, Hsinchu 30013, Taiwan; Email: y.l.chou@gapp.nthu.edu.tw The author would like to express gratitude for the reviewing comments received for the previous version.

\(^1\)Although one may argue that there is a trade-off between length and depth.

\(^2\)Theorem 6.7 in Billingsley [1] furnishes a compact statement of the Skorokhod’s representation theorem with a proof.
distributions are involved with and only with the indicator functions, it follows that the classical Glivenko-
Cantelli theorem can be obtained without using KSLLN. We then establish by the classical Glivenko-
Cantelli theorem (viewed as obtained independently of KSLLN) the weak convergence of the empirical
distribution functions, and we prove the convergence of the corresponding first moments.

2 Result

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Without employing KSLLN and its known proof(s), we prove

Theorem. If $X_1, X_2, \ldots$ are i.i.d. random variables on $\Omega$, and if $\lim_{c \to \infty} \sup_{n \in \mathbb{N}} n^{-1} \sum_{i=1}^n |X_i(\omega)| \cdot 1_{[c, +\infty]}(|X_i(\omega)|) = 0$ for $P$-almost all $\omega \in \Omega$, then $X_1 \in L^1(P)$ and

$$n^{-1} \sum_{i=1}^n X_i \to \text{a.s. } \mathbb{E}X_1.$$ 

Proof. Let $P$ be the probability measure induced by $X_1$ on the Borel sigma-algebra $\mathcal{B}_\mathbb{R}$ over $\mathbb{R}$. For all $n \in \mathbb{N}$ and all $\omega \in \Omega$, let $\hat{P}_n(B; \omega) := n^{-1} \sum_{i=1}^n 1_B \circ X_i(\omega)$ for all $B \in \mathcal{B}_\mathbb{R}$, so that $\hat{P}_n(\cdot; \omega)$ is the
empirical distribution constructed from $X_1(\omega), \ldots, X_n(\omega)$.

Since the classical Glivenko-Cantelli theorem, viewed as obtained from Theorem 1 in Section 3 of Chapter 4 in Shiryaev [2] and hence independently of KSLLN, implies that the distribution functions $\hat{F}_n(\cdot; \omega)$ of $\hat{P}_n(\cdot; \omega)$ converge uniformly to the distribution function $F$ of $P$ for $P$-almost all $\omega \in \Omega$, in particular there is some $A \in \mathcal{F}$ of $P$-measure 1 such that $\hat{F}_n(c; \cdot) \to F(c)$ on $A$ and for all continuity
points $c$ of $F$; from the portmanteau theorem we then have the weak convergence

$$\hat{P}_n(\cdot; \omega) \rightharpoonup P$$

for all $\omega \in A$.

Write

$$n^{-1} \sum_{i=1}^n X_i = \int y \, d\hat{P}_n(y; \omega)$$

for all $n \in \mathbb{N}$; we claim that the weak convergence implies the convergence $\int y \, d\hat{P}_n(y; \omega) \to \int x \, dP(x)$ for $P$-almost all $\omega \in \Omega$. Indeed, let $\mathbb{D}^X_i$ denote the Dirac measure at $X_i$ for each $i \in \mathbb{N}$. Given any $\omega \in A$, define for each $n \in \mathbb{N} \cup \{0\}$ the function $\xi_n(\cdot; \omega)$ as the natural projection of the product $X_{n e^{n \cup \{0\}}}^n H_n$
on $H_n$, where $H_0$ is the probability space $(\mathbb{R}, \mathcal{B}_\mathbb{R}, P)$ and $H_n$ is for each $n \in \mathbb{N}$ the probability space obtained simply from probabilizing $\mathbb{R}$ with respect to $\mathcal{B}_\mathbb{R}$ by assigning probability $1/n$ to each of
$X_1(\omega), \ldots, X_n(\omega)$. With respect to the apparent product sigma-algebra, the existence of the (product)
probability measure $P$ over the product $\bigotimes_{n \in \mathbb{N} \cup \{0\}}^n H_n$ is well-known; thus $\xi_0(\cdot; \omega), \xi_1(\cdot; \omega), \ldots$ form a sequence of random variables defined on the probability space $\bigotimes_{n \in \mathbb{N} \cup \{0\}}^n H_n$ for every $\omega \in A$. Then,
for every $\omega \in A$, the random variable $\xi_0(\cdot; \omega)$ has the same distribution $P$ as $X_1$, and each $\xi_n(\cdot; \omega)$
with $n \in \mathbb{N}$ has $n^{-1} \sum_{i=1}^n \mathbb{D}X_i(\omega) = \hat{P}_n(\cdot; \omega)$ as its distribution (concentrated on $\{X_1(\omega), \ldots, X_n(\omega)\}$).

It follows that

$$\int y \, d\hat{P}_n(y; \omega) = \mathbb{E}\xi_n(\cdot; \omega)$$

for all $n \in \mathbb{N}$ and all $\omega \in A$; moreover, we have

$$\xi_n(\cdot; \omega) \rightharpoonup X_1$$

---

3 A short, simple proof can be found under Theorem 19.1 in van der Vaart [3].

4 The use of different symbols for the bound variable appearing in the integrals is simply in order to minimize any possible distraction.
for all $\omega \in A$. Here the expectation of each $\xi_n(\cdot; \omega)$ is taken with respect to $\mathcal{P}$, and the relation $\rightsquigarrow$ between the random variables $\xi_n(\cdot; \omega)$ and $X_1$ means the weak convergence relation between their distributions.

Now

$$\int_{\{|\xi_n(\cdot; \omega)| \geq c \}} |\xi_n(\cdot; \omega)| \, d\mathcal{P} = n^{-1} \sum_{i=1}^{n} |X_i(\omega)| \cdot 1_{[c, +\infty)}(|X_i(\omega)|)$$

for all $n \in \mathbb{N}$, all $\omega \in A$, and all $c \in \mathbb{R}$. Since there is by assumption some $\mathcal{A} \in \mathcal{F}$ with $P(\mathcal{A}) = 1$ such that the collection $\{\xi_n(\cdot; \omega)\}_{n \in \mathbb{N}}$ is uniformly $\mathcal{P}$-integrable for all $\omega \in A \cap \mathcal{A}$ since the weak convergence $\xi_n(\cdot; \omega) \rightsquigarrow X_1$ holds for all $\omega \in A \cap \mathcal{A}$, and since $P(A \cap \mathcal{A}) = 1$, the desired implication then follows from Theorem 3.5 in Billingsley [1].

But Theorem 3.5 in Billingsley [1] also ensures that $X_1 \in L^1(P)$; the proof is complete. □

References

[1] P. Billingsley, *Convergence of Probability Measures*, second edition. John Wiley & Sons (1999).

[2] A. N. Shiryaev, *Probability*, second edition, translated by R. P. Boas. Springer (1996).

[3] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press (1998).

[4] A. M. Zapala, *Unbounded mappings and weak convergence of measures*, Statistics and Probability Letters 78, 2008, pp. 698–706.

---

5 Results neighboring the cited one exist, and may be more general in some respects. Examples of such a result are enumerated here for possible reference to potential unexpected applications. Zapala [4] gives some generalizations of the classical portmanteau theorem for unbounded maps on completely regular topological spaces. Theorem 2.20 in van der Vaart 3 serves as a special case of Theorem 2 in Zapala 4.