Convergence of sums of random variables in contrast to the weak and strong laws dealing with averaged sums in probability theory

Paper ID: IJIFR/ V2/ E2/ 006
Page No: 290-297
Research Area: Mathematics

Key Words: Random Variables, Convergence Notions, Weak & Strong Laws, Probability Theory

Dr. Nidhi Sharma
Assistant Professor, Department of Mathematics
Pt. C.L.S. Govt. College, Karnal (Haryana) India

Abstract
Probability theory is a fundamental pillar of modern mathematics with relations to other mathematical areas like algebra, topology, analysis, geometry or dynamical systems. The language of probability fits well into the classical theory of dynamical systems. For example, the ergodic theorem of Birkhoff for measure preserving transformations has as a special case the law of large numbers which describes the average of partial sums of random variables $\frac{1}{n} \sum_{k=1}^{n} X_k$ there are different versions of the law of large numbers. "Weak laws" make statements about convergence in probability; "strong laws" make statements about almost everywhere convergence. There are versions of the law of large numbers for which the random variables do not need to have a common distribution and which go beyond Birkhoff's theorem. Another important theorem is the central limit theorem which shows that $S_n = X_1 + X_2 + \cdots + X_n$ normalized to have zero mean and variance 1 converges in law to the normal distribution or the law of the iterated logarithm which says that for centred independent and identically distributed $X_k$, the scaled sum $S_n/n$ has accumulation points in the interval $[-\sigma, \sigma]$ if $\Delta_n = \sqrt{2n \log \log n}$ and $\sigma$ is the standard deviation of $X_k$. While stating the weak and strong law of large numbers and the central limit theorem, different convergence notions for random variables appear: almost sure convergence is the strongest; it implies convergence in probability and the later implies convergence in law. There is also $L^1$ convergence which is stronger than convergence in probability.

1 Introduction

1.1 The weak law of large numbers
Consider a sequence $X_1, X_2, \ldots$ of random variables on a probability space $(\Omega, \mathcal{A}, P)$. We are interested in the asymptotic behavior of the sums $S_n = X_1 + X_2 + \cdots + X_n$ for $n \to \infty$ and especially in the convergence of the averages $S_n/n$. The limiting behavior is described by "laws of large numbers". Depending on the definition of convergence, one speaks of "weak" and "strong" laws of large numbers.
We first prove the weak law of large numbers. There exist different versions of this theorem since more assumptions on $X_n$ can allow stronger statements.

**Definition:** A sequence of random variables $Y_n$ converges in probability to a random variable $Y$, if for all $\epsilon > 0$,

$$\lim_{n \to \infty} P[|Y_n - Y| \geq \epsilon] = 0.$$ 

One calls convergence in probability also stochastic convergence.

**Theorem 1:** (Weak law of large numbers for uncorrelated random variables). Assume $X_i \in L^2$ have common expectation $E[X_i] = m$ and satisfy $\sup_n \frac{1}{n} \sum_{i=1}^{n} \text{Var}[X_i] < \infty$. If $X_n$ are pairwise uncorrelated, then

$$\frac{S_n}{n} \to m$$

in probability. 

**Proof:** Since $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \cdot \text{Cov}[X,Y]$ and $X_n$ are pairwise uncorrelated, we get $\text{Var}[X_n + X_m] = \text{Var}[X_n] + \text{Var}[X_m]$ and by induction $\text{Var}[S_n] = \sum_{i=1}^{n} \text{Var}[X_i]$. Using linearity, we obtain $E[S_n/n] = m$ and

$$\text{Var}\left[\frac{S_n}{n}\right] = \frac{E\left[S_n^2\right]}{n^2} - \frac{E[S_n]^2}{n^2} = \frac{\text{Var}[S_n]}{n^2} = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}[X_i].$$

The right hand side converges to zero for $n \to \infty$. With Chebychev's inequality, we obtain

$$P\left[\left|\frac{S_n}{n} - m\right| \geq \epsilon\right] \leq \frac{\text{Var}\left[\frac{S_n}{n}\right]}{\epsilon^2}.$$

**Theorem 2:** (Weak law of large numbers for independent $L^4$ random variables). Assume $X_i \in L^4$ have common expectation $E[X_i] = m$ and satisfy $M = \sup_n \|X_i\| < \infty$. If $X_i$ are independent, then

$$\frac{S_n}{n} \to m$$

in probability. Even converges for all $\epsilon > 0$.

**Proof:** We can assume without loss of generality that $m = 0$. Because the $X_i$ are independent, we get

$$E[S_n^4] = \sum_{i_1, i_2, i_3, i_4 = 1}^{n} E[X_{i_1}X_{i_2}X_{i_3}X_{i_4}]$$

Again by independence, a sum $m$ and $E[X_{i_1}X_{i_2}X_{i_3}X_{i_4}]$ is zero if an index $i = i_k$ occurs alone, is $E[X_i^4]$ if all indices are the same and $E[X_i^2]E[X_i^2]$, if there are two pairwise equal indices. Since by Jensen's inequality and $E[X_i^2] \leq E[X_i^4] \leq M$, we get

$$E[S_n^4] \leq nM + n(n+1)M$$

By using Chebychev-Markov inequality, with $h(x) = x^4$ to get

$$P\left[\left|\frac{S_n}{n}\right| \geq \epsilon\right] \leq \frac{E[(S_n/n)^4]}{\epsilon^4} \leq \frac{Mn + n^2}{\epsilon^4n^4} \leq 2M \frac{1}{\epsilon^4n^2}.$$  

### 1.2 The probability distribution function

**Definition:** The law of a random variable $X$ is the probability measure $\mu$ on $\mathbb{R}$ defined by
\[
\mu B = P\{X^{-1}(B)\}\] for all \(B\) in the Borel \(\sigma\)-algebra of \(\mathbb{R}\). The measure \(\mu\) is also called the push-forward measure under the measurable map \(X: \Omega \rightarrow \mathbb{R}\).

**Definition:** The distribution function of a random variable \(X\) is defined as

\[
F_X(s) = \mu((-\infty, s]) = P[X \leq s]
\]

The distribution function is sometimes also called cumulative density function (CDF) but we do not use this name here in order not to confuse it with the probability density function (PDF)

\[
F_X(s) = F_Y(s)
\]

for continuous random variables.

**Remark:** Remark. The distribution function \(F\) is very useful. For example, if \(X\) is a continuous random variable with distribution function \(F\), then \(Y = F(X)\) has the uniform distribution on \([0,1]\). We can reverse this. If we want to produce random variables with a distribution function \(F\), just take a random variable \(Y\) with uniform distribution on \([0,1]\) and define \(X = F^{-1}(Y)\). This random variable has the distribution function \(F\) because

\[
\{X \in [a, b]\} = \{F^{-1}(Y) \in [a, b]\} \subseteq \{Y \in F([a, b])\} = \{Y \in [F(a), F(b)]\} = F(b) - F(a)
\]

We see that we need only to have a random number generator which produces uniformly distributed random variables in \([0, 1]\) to produce random variables with a given continuous distribution.

# 2 Convergences of Random Variables

In order to formulate the strong law of large numbers, we need some other notions of convergence.

**Definition:** A sequence of random variables \(X_n\) converges in probability to a random variable \(X\), if

\[
P[|X_n - X| \geq \epsilon] \rightarrow 0
\]

for all \(\epsilon > 0\).

**Definition:** A sequence of random variables \(X_n\) converges almost everywhere or almost surely to a random variable \(X\), if \(P[X_n \rightarrow X] = 1\).

**Definition:** A sequence of \(L^p\) random variables \(X_n\) converges in \(L^p\) to a random variable \(X\), if

\[
||X_n - X||_p \rightarrow 0 \quad \text{for} \ n \rightarrow \infty.
\]

**Definition:** A sequence of random variables \(X_n\) converges fast in probability, or completely if

\[
\sum_n P[|X_n - X| \geq \epsilon] < \infty
\]

for all \(\epsilon > 0\).

We have so four notions of convergence of random variables \(X_n \rightarrow X\), if the random variables are defined on the same probability space \((\Omega, A, P)\). We have seen two equivalent but weaker notions convergence in distribution and weak convergence, which not necessarily assume \(X_n\) and \(X\) to be defined on the same probability space.

# 3 The strong law of large numbers

The weak law of large numbers makes a statement about the stochastic convergence of sums
of random variables $X_n$. The strong laws of large numbers make analog statements about almost everywhere convergence. The first version of the strong law does not assume the random variables have the same distribution. They are assumed to have the same expectation and have to be bounded in $L^4$.

**Theorem 3.1**: (Strong law for independent $L^1$ random variables). Assume $X_n$ are independent random variables in $L^4$ with common expectation $E[X_n] = m$ and for which $M = \sup_n \|X_n\|_4 < \infty$.

Then $S_n/n \to m$ almost everywhere.

*Proof*: From previous theorems, we have derived that

$$ P[|S_n/n - m| \geq \epsilon] \leq 2M \frac{1}{\epsilon^4 n^2} $$

This means that $S_n/n \to m$ converges completely. By proposition (2) we have almost everywhere convergence.

**Definition**: A real number $x \in [0,1]$ is called normal to the base 10, if its decimal expansion $x = x_1x_2...$ has the property that each digit appears with the same frequency $1/10$.

*Proof*: Define the random variables $X_n(x) = x_n$, where $x_n$ is the $n^{th}$ decimal digit. We have only to verify that $X_n$ are IID random variables. The strong law of large numbers will assure that almost all $x$ are normal. Let $\Omega = \{0, 1, ..., 9\}^\infty$ be the space of all infinite sequences $(\omega_1, \omega_2, \omega_3,...)$. Define on $\Omega$ the product $\sigma$-algebra $A$ and the product probability measure $P$. Define the measurable map $S(\omega) = \sum_{n=1}^{\infty} \omega_k/10^k = x$ from $\Omega$ to $[0,1]$. The integers $\omega_k$ are just the decimal digits of $x$. The map $S$ is measure preserving and can be inverted on a set of measure 1 because almost all real numbers have a unique decimal expansion.

Because $X_n(x) = X_n(S(\omega)) = Y_n(\omega) = \omega_n$, if $S(\omega) = x$

We see that $X_n$ are the same random variables than $Y_n$. The later are by construction IID with uniform distribution on $[0, 1, ...., 9]$.

**Theorem 3.2**: (Strong law for pairwise independent $L^1$ random variables). Assume $X_n \in L^1$ are pairwise independent and identically distributed random variables. Then $S_n/n \to E[X]$ almost everywhere.

*Proof*. We can assume without loss of generality that $X_n \geq 0$ (because we can split $X_n = X_n^+ + X_n^-$ into its positive $X_n^+ = X_n \vee 0 = \max(0, X_n)$, and negative part $X^- = -X_n 0 = \max(-X, 0)$). Knowing the result for $X_n^\pm$ implies the result for $X_n$.

Define $f_R(t) = 1_{-R, R}$, the random variables $X_n^{(R)} = f_R(X_n)$ and $Y_n = X_n^{(n)}$ as well as

$$ S_n = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad T_n = \frac{1}{n} \sum_{i=1}^{n} Y_i. $$

(i) It is enough to show that $T_n - E[T_n] \to 0$.

*Proof*. Since $E[Y_n] \to E[X] = m$, we get $E[T_n] \to m$. Because
we get by the first Borel-Cantelli lemma that \( P[Y_n \neq X_n, \text{infinitely often}] = 0 \). This means 

\[ T_n - S_n \to 0 \text{ almost everywhere, proving } E[S_n] \to m. \]

(ii) Fix a real number \( a > 1 \) and define an exponentially growing subsequence \( k^n = [a^n] \) which is the integer part of \( a^n \). Denote by \( \mu \) the law of the random variables \( X_n \). For every \( \varepsilon > 0 \), we get using Chebychev inequality, pairwise independence for \( k^n = [a^n] \) & constant which vary from line to line.

\[
\sum_{n=1}^{\infty} P[|T_{k_n} - E[T_{k_n}]| \geq \varepsilon] \leq \sum_{n=1}^{\infty} \frac{\text{Var}[T_{k_n}]}{\varepsilon^2 k_n^2}
\]

\[
= \frac{1}{\varepsilon^2} \sum_{m=1}^{k_n} \text{Var}[Y_m]
\]

\[
= \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \text{Var}[Y_m] \sum_{n: k_n \geq m} \frac{1}{k_n^2}
\]

\[
\leq (1) \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \text{Var}[Y_m] \frac{C}{m^2}
\]

\[
\leq C \sum_{m=1}^{\infty} \frac{1}{m^2} E[Y_m^2]
\]

In (1) we used that with \( \kappa_n = [a^n] \) one has \( \sum_{n: k_n \geq m} \kappa_n \geq C \cdot m^{-2} \).

\[
\sum_{n=1}^{\infty} P[|T_{k_n} - E[T_{k_n}]| \geq \varepsilon] \leq C \sum_{m=1}^{\infty} \frac{1}{m^2} E[Y_m^2]
\]

\[
\leq C \sum_{m=1}^{\infty} \frac{1}{m^2} \sum_{l=0}^{m-1} \int_{l}^{l+1} x^2 \, d\mu(x)
\]

\[
= C \sum_{l=0}^{\infty} \sum_{m=l+1}^{\infty} \frac{1}{m^2} \int_{l}^{l+1} x^2 \, d\mu(x)
\]

\[
\leq C \sum_{l=0}^{\infty} \sum_{m=l+1}^{\infty} \frac{(l+1)}{m^2} \int_{l}^{l+1} x \, d\mu(x)
\]
\[
\leq (2) \quad C \sum_{l=0}^{\infty} \int_{l}^{l+1} x \, d\mu(x) \\
\leq C \cdot \mathbb{E}[X_1] < \infty.
\]

In (2) we used that \(\sum_{m=l+1}^{n} m^{-2} \leq C \cdot (l + 1)^{-1}\).

We have now proved complete (= fast stochastic) convergence. The almost everywhere convergence of \(T_{k_n} = \mathbb{E}[T_{k_n}] \to 0\).

(iii) So far, the convergence has only been verified along a subsequence \(\{k_n\}\). Because we assumed \(X_n \geq 0\), the sequence \(U_n = \sum_{l=1}^{n} Y_l = \tau n\) is tonically increasing. For \(k \in [k_m, k_{m+1}]\), we get therefore

\[
\frac{k_m}{k_{m+1}} \frac{U_{k_m}}{k_m} = \frac{U_{k_m}}{k_m} \leq \frac{U_{n}}{n} \leq \frac{U_{k_{m+1}}}{k_m} = \frac{k_{m+1}}{k_m} \frac{U_{k_{m+1}}}{k_{m+1}}.
\]

and from \(\lim_{n \to \infty} T_n = \mathbb{E}[X_1]\) almost everywhere,

\[
\frac{1}{\alpha} \mathbb{E}[X_1] \leq \liminf_{n} T_n \leq \limsup_{n} T_n \leq \alpha \mathbb{E}[X_1].
\]

Remark. The strong law of large numbers can be interpreted as a statement about the growth of the sequence \(\sum_{k=1}^{n} X_n\). For \(\mathbb{E}[X_1] = 0\), the convergence \(\frac{1}{n} \sum_{k=1}^{n} X_n \to 0\) means that for all \(\epsilon > 0\) there exists \(m\) such that for \(n > m\)

\[
|\sum_{k=1}^{n} X_n| \leq \epsilon n.
\]

This means that the trajectory \(\sum_{k=1}^{n} X_n\) is finally contained in any arbitrary small cone. In other words, it grows slower than linear. The exact description for the growth of \(\sum_{k=1}^{n} X_n\) is given by the law of the iterated logarithm of Khinchin which says that a sequence of IID random variables \(X_n\) with \(\mathbb{E}[X_n] = m\) and \(\sigma(X_n) = \sigma \neq 0\) satisfies

\[
\lim_{n \to \infty} \sup_{n} \frac{S_n}{\Lambda_n} = +1, \quad \lim_{n \to \infty} \inf_{n} \frac{S_n}{\Lambda_n} = -1,
\]

with \(\Lambda_n = \sqrt{2\sigma^2 n \log \log n}\).

Theorem: Given a sequence \(X_n \in \mathcal{L}\) which is independent. Choose \(a_{ik} \in \text{med}(S_l - S_k)\). Then, for all \(n \in \mathbb{N}\) and all \(\epsilon > 0\).

\[
P[\max_{1 \leq k \leq n} |S_n + a_{n,k}| \geq \epsilon] \leq 2P[S_n \geq \epsilon]
\]

Proof. Fix \(n \in \mathbb{N}\) and all \(\epsilon > 0\). The sets

\[
A_1 = \{S_1 + a_{n,1} \geq \epsilon\}, \quad A_{k+1} = \{\max_{1 \leq l \leq k} (S_n + a_{n,l}) < \epsilon, S_{k+1} + a_{n,k+1} \geq \epsilon\}
\]
for $1 \leq k \leq n$ are disjoint and $\bigcup_{k=1}^{n} A_k = \{ \max_{1 \leq k \leq n} (S_k + \alpha_{n,k}) \geq \epsilon \}$. Because $\{ S_n \geq \epsilon \}$ contains all the sets $A_k$ as well as $\{ S_n - S_k \geq \alpha_{n,k} \}$ for $1 \leq k \leq n$, we have using the independence of $\sigma(A_k)$ and $\sigma(S_n - S_k)$

$$
P[S_n \geq \epsilon] \geq \sum_{k=1}^{n} P[S_n - S_k \geq \alpha_{n,k} \cap A_k]$$

$$= \sum_{k=1}^{n} P[S_n - S_k \geq \alpha_{n,k}] P[A_k]$$

$$\geq \frac{1}{2} \sum_{k=1}^{n} P[A_k]$$

$$= \frac{1}{2} P[\bigcup_{k=1}^{n} A_k]$$

$$= \frac{1}{2} P[\max_{1 \leq k \leq n} (S_n + \alpha_{n,k}) \geq \epsilon].$$

Applying this inequality to $-X_n$, we get also $P[-S_n - \alpha_{n,m} \geq -\epsilon] \geq 2P[-S_n \geq -\epsilon]$ and so

$$P[\max_{1 \leq k \leq n} |S_n + \alpha_{n,k}| \geq \epsilon] \leq 2P[S_n \geq \epsilon].$$

4 Weak Convergence

**Definition:** Denote by $C_b(R)$ the vector space of bounded continuous functions on $\mathbb{R}$. This means that $\|f\|_{\infty} = \sum_{x \in \mathbb{R}} |f(x)| < \infty$ for every $f \in C_b(R)$. A sequence of Borel probability measures $\mu_n$ on $\mathbb{R}$ converges weakly to a probability measure $\mu$ on $\mathbb{R}$ if for every $f \in C_b(\mathbb{R})$ one has $\int f d\mu_n \to \int f d\mu$ in the limit $n \to \infty$.

For weak convergence, it is enough to test

**Remark:** We need to show that any sequence $\int f d\mu_n$ converges to $\int f d\mu$ for a dense set in $C_b(\mathbb{R})$. This dense set can consist of the space $P(\mathbb{R})$ of polynomials or the space $C_b(\mathbb{R})$ of bounded smooth functions. An important fact is that a sequence of random variables $X_n$ converges in distribution to $X$ if and only if $E[h(X_n)] = E[h(X)]$ for all smooth functions $h$ on the real line. This will be used the proof of the central limit theorem.

Weak convergence defines a topology on the set $M_1(\mathbb{R})$ of all Borel probability measures on $\mathbb{R}$. Similarly, one has a topology for $M_1([a, b])$.

**Lemma:** The set $M_1(I)$ of all probability measures on an interval $I = [a, b]$ is a compact topological space.

**Proof:** We need to show that any sequence $\mu_n$ of probability measures on $I$ has an accumulation point. The set of functions $f_k(x) = x^k$ on $[a, b]$ span all polynomials and so a dense set in $C_b([a, b])$. The sequence $\mu_n$ converges to $\mu$ and only if all the moments $\int_a^b x^k d\mu_n$ converge for $n \to \infty$ and for all $k \in \mathbb{N}$. In other words, the compactness of $M_1([a, b])$
is equivalent to the compactness of the product space $\prod$ with the product topology, which is Tychnov's Theorem.

Theorem: (Weak convergence = convergence in distribution). A sequence $X_n$ of random variables converges in law to a random variable $X$ if and only if $X_n$ converges in distribution to $X$.

Proof. (i) Assume we have convergence in law. We want to show that we have convergence in distribution. Given $s \in \text{Cont}(f)$ and $\delta > 0$. Define a continuous function $1_{(-\infty,s]} \leq f \leq 1_{(-\infty,s+\delta]}$. Then

$$F_n(s) = \int_{\mathbb{R}} 1_{(-\infty,s]} \ d\mu_n \leq \int_{\mathbb{R}} f \ d\mu_n \leq \int_{\mathbb{R}} 1_{(-\infty,s+\delta]} \ d\mu_n = F_n(s + \delta).$$

This gives

$$\limsup_{n \to \infty} F_n(s) \leq \lim_{n \to \infty} \int f \ d\mu_n = \int f \ d\mu \leq F(x + \delta).$$

Similarly, we obtain with a function $1_{(-\infty,s-\delta]} \leq f \leq 1_{(-\infty,s]}$

$$\liminf_{n \to \infty} F_n(s) \geq \lim_{n \to \infty} \int f \ d\mu_n = \int f \ d\mu \geq F(s - \delta).$$

Since $F$ is continuous at $x$ we have for $\delta \to 0$:

$$F'(s) = \lim_{\delta \to 0} F'(s - \delta) \leq \liminf_{n \to \infty} F_n(s) \leq \limsup_{n \to \infty} F_n(s) \leq F(s).$$

That is we have established convergence in distribution.

(ii) Assume now we have no convergence in law. There exists then a continuous function $f$ so that $\int f \ d\mu_n$ to $\int f \ d\mu$ fails. That is, there is a subsequence and $\varepsilon > 0$ such that $\int f \ d\mu_{n_k} - \int f \ d\mu \geq \varepsilon > 0$. There exists a compact interval $I$ such that $\int f \ d\mu_{n_k} - \int f \ d\mu \geq \varepsilon/2 > 0$ and we can assume that $\mu_{n_k}$ and $\mu$ have support on $I$. The set of all probability measures on $I$ is compact in the weak topology. Therefore, a subsequence of $d\mu_{n_k}$ converges weakly to a measure $\nu$ and $|\nu(f) - \mu(f)| \geq \varepsilon/2$. Define the $\pi$-system $\mathcal{I}$ of all intervals $\{(-\infty,s] \mid s \text{ continuity point of } F\}$. We have $\mu_n((-\infty,s]) = F_{X_n}(s) = F_X(s) = \mu((-\infty,s])$. Using (i) we see $\mu_{n_k}((-\infty,s]) \to \nu((-\infty,s])$ also, so that $\mu$ and $\nu$ agree on the $\pi$-system $\mathcal{I}$. If $\mu$ and $\nu$ agree on $\mathcal{I}$, they agree on the $\pi$-system of all intervals $\{(-\infty,s]\}$. By lemma (2.1.4), we know that $\mu = \nu$ on the Borel $\sigma$-algebra and so $\mu = \nu$. This contradicts $|\nu(f) - \mu(f)| \geq \varepsilon/2$. So, the initial assumption of having no convergence in law was wrong.

References: