The Weak Law of Large Numbers

Robin Liu

November 30th, 2020

1 Introduction

In this short note, we prove the weak law of large numbers. Section 2 provides a review of probability and measure theory and section 3 illustrates the proof of the theorem.

2 Probability and Measure Theory

The setting of probability takes place in a measure space $(\Omega, \mathcal{F}, P)$ where $P(\Omega) = 1$. This measure space is known as a **probability space**. Probabilists call Ω the **sample space** and P the **probability measure**. We define **random variables** as real-valued functions X on Ω such that $X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathcal{F}$ for every Borel set B. In other words, $P(X^{-1}(B))$ is defined. If $X \geq 0$, then we call $EX = \int X\, dP$ the **expected value** of X. It is clear that the expected value of X is the Lebesgue integral of X with respect to the probability measure. We also write EX as μ and call it the **mean** of X. The **variance** of X is defined as $Var(X) = E(X - \mu)^2$. Two useful equations for variance that should be obvious to anyone with previous experience in probability are

$$Var(X) = E(X - \mu)^2 = EX^2 - 2\mu EX + \mu^2 = EX^2 - \mu^2$$

and

$$Var(aX + b) = E(aX + b - E(aX + b))^2 = a^2 E(X - EX)^2 = a^2 Var(X).$$

We now provide definitions of concepts that are unique to probability and not measure theory. We call the random variables $X_1, X_2, ..., X_n$ **independent** if for every Borel set B_i it is the case that

$$P(\cap_{i=1}^{n} \{X_i \in B_i\}) = \prod_{i=1}^{n} P(X_i \in B_i)$$

for $i = 1, 2, ..., n$. We say $X_1, X_2, ..., X_n$ are **independent and identically distributed**, or **i.i.d.**, if they are independent and have the same distribution, in other words, if $P(X_1 \leq x) = ... = P(X_n \leq x)$. Furthermore, we call random variables X_i where $EX_i^2 < \infty$ **uncorrelated** if $E(X_i X_j) = EX_i EX_j$ for $i \neq j$. A key result that we will rely on in the proof of the weak law is that *independent random variables are uncorrelated*. Another useful result taught in introductory probability courses is addressed in the following theorem.

Theorem 2.1. *Let $X_1, X_2, ...X_n$ be uncorrelated random variables such that $EX_i^2 < \infty$. Then*

$$Var(X_1 + ... + X_n) = Var(X_1) + ... + Var(X_n).$$

Three major theorems from measure theory are required for proving the weak law of large numbers: Chebyshev's Inequality, Fubini's Theorem, and Lebesgue's Dominated Convergence Theorem. These theorems are stated in succession.

Theorem 2.2 (Chebyshev's Inequality). *Suppose $f : \mathbb{R} \to \mathbb{R}$ such that $f \geq 0$, $A \in \mathcal{B}$ where $\mathcal{B}$ is the Borel algebra, and $i_A = \inf\{f(y) : y \in A\}$. Then it follows that*

$$i_A P(X \in A) \leq E(f(X); X \in A) \leq Ef(X).$$

Theorem 2.3 (Fubini's Theorem). *If $f \geq 0$ or $\int |f|\, d\mu < \infty$, then*

$$\int_X \int_Y f(x,y)\mu_2(dy)\mu_1(dx) = \int_Y \int_X f(x,y)\mu_1(dx)\mu_2(dy).$$

Theorem 2.4 (Lebesgue's Dominated Convergence Theorem). *If $X_n \to X$ almost everywhere and $|X_n| \leq Y$ for all n and for some Y such that $EY < \infty$, then $EX_n \to EX$.*

We end this section with a note on some notation. Probabilists typically refer to the concept of convergence in measure as **convergence in probability**. In other words, a random variable Y_n converges in probability to Y if for every $\epsilon > 0$ we have $P(|Y_n - Y| > \epsilon) \to 0$ as $n \to \infty$.

3 The Weak Law of Large Numbers

Having established the key results in probability and measure theory that will be required for this project, we begin our proof of the weak law of large numbers. The ultimate goal of this section is to prove the following theorem.

Theorem 3.1 (Weak Law of Large Numbers). *Let $X_1, X_2, ...$ be i.i.d. such that $E|X_i| < \infty$, $S_n = X_1 + ... + X_n$, and $\mu = EX_1$. Then $S_n/n \to \mu$ in probability.*

Before we can prove this theorem, we need to prove a few more related theorems and lemmas. The first theorem we need is a form of the weak law for **triangular arrays**, which are arrays of $X_{n,k}$, $1 \leq k \leq n$.

Theorem 3.2 (Weak Law for Triangular Arrays). *Fixing n, let $X_{n,k}$, $1 \leq k \leq n$ be independent. Furthermore, let $b_n > 0$ such that $b_n \to \infty$ and $\bar{X}_{n,k} = X_{n,k}1_{|X_{n,k}| \leq b_n}$ where 1 is the indicator function. As $n \to \infty$, suppose that (1): $\sum_{k=1}^n P(|X_{n,k}| > b_n) \to 0$ and (2): $b_n^{-2} \sum_{k=1}^n E\bar{X}_{n,k}^2 \to 0$. If $S_n = X_{n,1} + ...X_{n,n}$ and $a_n = \sum_{k=1}^n E\bar{X}_{n,k}$, then $(S_n - a_n)/b_n \to 0$ in probability.*

Proof. Define $\bar{S_n} = \bar{X}_{n,1} + ... + \bar{X}_{n,n}$. Suppose $\bar{S_n} = S_n$. Then $\{|\frac{S_n - a_n}{b_n}| > \epsilon\} = \{|\frac{\bar{S}_n - a_n}{b_n}| > \epsilon\}$. It follows that $\{|\frac{S_n - a_n}{b_n}| > \epsilon\} \subset \{|\frac{\bar{S}_n - a_n}{b_n}| > \epsilon\} \cup \{S_n \neq \bar{S}_n\}$. Then monotoncity implies that

$$P\left(\left|\frac{S_n - a_n}{b_n}\right| > \epsilon\right) \leq P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon\right) + P(S_n \neq \bar{S}_n).$$

For it to be the case that $S_n \neq \bar{S}_n$, there must exist a $X_{n,k}$ such that $\bar{X}_{n,k} \neq X_{n,k}$. Then $\{S_n \neq \bar{S}_n\} \subset \cup_{k=1}^n \{\bar{X}_{n,k} \neq X_{n,k}\}$. By assumption (1), it follows that

$$P(S_n \neq \bar{S}_n) \leq P(\cup_{k=1}^n \{\bar{X}_{n,k} \neq X_{n,k}\}) \leq \sum_{k=1}^n P(|X_{n,k}| > b_n) \to 0.$$

Notice that $a_n = E\bar{S}_n$. By Chebyshev's Inequality, Theorem 2.1, and the fact that $Var(X) = EX^2 - \mu^2$, and assumption (2), we have

$$P\left(\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon\right) \leq \epsilon^{-2} E\left|\frac{\bar{S}_n - a_n}{b_n}\right|^2 = \epsilon^{-2} b_n^{-2} Var(\bar{S}_n) \leq (\epsilon b_n)^{-2} \sum_{k=1}^n E(\bar{X}_{n,k})^2 \to 0.$$

Therefore, we have

$$P\left(\left|\frac{S_n - a_n}{b_n}\right| > \epsilon\right) \to 0,$$

which is what we wanted to show. □

We also need the following lemma.

Lemma 3.3. *Let $Y \geq 0$ and $p \geq 0$. Then $E(Y^p) = \int_0^\infty py^{p-1}P(Y > y)\, dy$.*

Proof. By Fubini's Theorem, we have

$$\begin{aligned}\int_0^\infty py^{p-1}P(Y>y)\,dy &= \int_0^\infty \int_\Omega py^{p-1}1_{Y>y}\,dP\,dy \\ &= \int_\Omega \int_0^\infty py^{p-1}1_{Y>y}\,dy\,dP \\ &= \int_\Omega \int_0^Y py^{p-1}\,dy\,dP \\ &= \int_\Omega Y^p\,dP = EY^p.\end{aligned}$$

□

The last theorem that we need is another generalized form of the weak law.

Theorem 3.4. *Let $X_1, X_2, \dots$ be i.i.d with $xP(|X_i| > x) \to 0$ as $x \to \infty$. Let $S_n = X_1 + \dots + X_n$ and $\mu_n = E(X_1 1_{|X_1|\le n})$. Then $S_n/n - \mu_n \to 0$ in probability.*

Proof. We will apply the Weak Law for Triangular Arrays to this problem by setting $X_{n,k} = X_k$ and $b_n = n$. Notice that

$$\sum_{k=1}^n P(|X_k| > n) = nP(|X_i| > n) \to 0.$$

Therefore, assumption (1) holds. Notice that $P(|\bar{X}_{n,1}| > y) = 0$ for $y \ge n$ and $P(|\bar{X}_{n,1}| > y) = P(|X_1| > y) - P(|X_1| > n)$ for $y \le n$. By Lemma 3.3, we have

$$E(\bar{X}_{n,1}^2) = \int_0^\infty 2yP(|\bar{X}_{n,1}| > y)\,dy \le \int_0^n 2yP(|X_1| > y)\,dy.$$

Set $g(y) = 2yP(|X_1| > y)$. Notice that $0 \le g(y) \le 2y$. By assumption, we know $g(y) \to 0$ as $y \to \infty$. Therefore, $g(y)$ is bounded, so $M = \sup g(y) < \infty$. Let $\epsilon_k = \sup\{g(y) : y > K\}$. Considering the integrals of $[0, K]$ and $[K, n]$, we get

$$\int_0^n 2yP(|X_1| > y)\,dy \le KM + (n-K)\epsilon_K.$$

Dividing both sides by n and taking the limit w.r.t. n gives

$$\lim_{n\to\infty} \frac{1}{n}\int_0^n 2yP(|X_1| > y)\,dy \le \epsilon_K.$$

Taking the limit w.r.t. K gives

$$\lim_{n\to\infty} \frac{1}{n}\int_0^n 2yP(|X_1| > y)\,dy = 0.$$

Then $E(\bar{X}_{n,1}^2)/n \to 0$. Therefore, assumption (2) is satisfied and we conclude that $S_n/n - \mu_n \to 0$ in probability. □

Finally, we are ready to take on the weak law of large numbers.

Proof of the Weak Law of Large Numbers. By Chebyshev's Inequality, we have $xP(|X_1| > x) \le E(|X_1|1_{|X_1|>x})$. Notice that $1_{|X_1|>x} \to 0$ as $x \to \infty$. By Lebesgue's Dominated Convergence Theorem, it follows that $E(|X_1|1_{|X_1|>x}) \to 0$ as $x \to \infty$. By Theorem 3.4, we have for every $\epsilon > 0$ that $P(|S_n/n - \mu_n| > \epsilon/2) \to 0$. Lebesgue's Dominated Convergence also implies that

$$\mu_n = E(X_1 1_{|X_1|\le n}) \to EX_1 = \mu$$

as $n \to \infty$. Since $\mu_n \to \mu$, it follows that μ will be arbitrarily close enough to μ_n for large n such that $\{|S_n/n - \mu| > \epsilon\} \subset \{|S_n/n - \mu_n| > \epsilon/2\}$. Then $P(|S_n/n - \mu| > \epsilon) \to 0$. Therefore, $S_n/n \to \mu$ in probability. □