C

STRONG LAW OF LARGE NUMBERS

In this section we shall state and prove the strong law of large numbers.

Theorem C.0.1. **(Strong Law of Large Numbers)** *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables. Assume that X_1 has finite mean μ and finite variance σ^2. Let $A = \{\lim_{n\to\infty} \overline{X}_n = \mu\}$. Then*

$$P(A) = 1. \tag{C.0.1}$$

As remarked in Chapter 8, the above results states that the convergence of sample mean to μ actually happens with Probability one. This mode of convergence of the sample mean to the true mean is called "convergence with probability 1." We define it precisely below.

Definition C.0.2. *A sequence $X_1, X_2, \ldots$ is said to converge with probability one to a random variable X if $A = \{\lim_{n\to\infty} \overline{X}_n = X\}$.*

$$P(A) = 1. \tag{C.0.2}$$

The following notation

$$X_n \stackrel{w.p.1}{\longrightarrow} X$$

is typically used to convey that the sequence $X_1, X_2, \ldots$ converges with probability one to X.

As alluded earlier that this is a stronger mode of convergence. We prove it in the next proposition.

Proposition C.0.3. *Let $X_1, X_2, \ldots$ be a sequence of random variables on a sample space S. Suppose X_n converges to a random variable X with probability 1 then X_n converges to a random variable X in probability.*

Proof- Let $\epsilon > 0$ and $\delta > 0$ be given. We need to show $\exists N$ such that

$$P(|X_m - X| > \epsilon) < \delta, \,\forall m \geq N. \tag{C.0.3}$$

Let $A = \{\omega \in S : \lim_{n\to\infty} X_n(\omega) = X\}$. We are given that

$$P(A) = 1. \tag{C.0.4}$$

Suppose we denote, for $\eta > 0$ and $n \geq 1$,

$$A_n^\eta = \{\omega \in S : |X_n(\omega) - X(\omega)| \leq \epsilon\}.$$

then

$$A = \cap_{\eta>0} \cup_{k=1}^\infty \cap_{n=k}^\infty A_n^\eta.$$

This can be verified using the fact that $\omega \in A$ if and only if for all $\eta > 0$, there is a $k \equiv k(\omega)$ such that

$$|X_n(\omega) - X(\omega)| \leq \epsilon, \;\forall\, n \geq k.$$

For $m \geq 1$, define $B_m^\epsilon = \cap_{n=m}^\infty A_n^\epsilon$. Note

$$B_m^\epsilon \subset B_{m+1}^\epsilon, \tag{C.0.5}$$

for all $m \geq 1$. So by Exercise 1.1.13, we have

$$\lim_{m\to\infty} P(B_m^\epsilon) \uparrow P(\cup_{m=1}^\infty B_m^\epsilon). \tag{C.0.6}$$

As $A \subset \cup_{m=1}^{\infty} B_m^{\epsilon}$, using (C.0.4) we have $1 = P(A) \leq P(\cup_{m=1}^{\infty} B_m^{\epsilon}) \leq 1$. So

$$P(\cup_{m=1}^{\infty} B_m^{\epsilon}) = 1. \tag{C.0.7}$$

By (C.0.6) and (C.0.7) $\exists N$ such that

$$P(B_m^{\epsilon_0}) > 1 - \delta, \ \forall m \geq N.$$

As $B_m^{\epsilon} \subset A_m^{\epsilon}$,

$$P(A_m^{\epsilon}) > 1 - \delta, \ \forall m \geq N.$$

Therefore by considering the complement of A_m^{ϵ} we obtain (C.0.3). ■

We will need a technical Lemma regarding convergence in probabilty which we state and prove below.

Lemma C.0.4. *Suppose a sequence of random variables X_n is such that*

$$X_n \xrightarrow{p} X \text{ and } X_n \xrightarrow{p} Y$$

for some random variables X, Y then $P(X = Y) = 1$.

Proof- Let $k \geq 1$. Let $A_k = \{| X - Y | \geq \frac{1}{k}\}$. Notice that $A_k \subset A_{k+1}$ and $\cup_{k=1}^{\infty} A_k = \{X \neq Y\}$. Let $k \geq 1, \delta > 0$ be given. As $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{p} Y$, (applying Definition 8.2.3 with $\epsilon = \frac{1}{2k}$), there exists N such that for all $n \geq N$

$$0 \leq P\left(| X_n - X | > \frac{1}{2k}\right) < \frac{\delta}{2} \qquad \text{and} \qquad 0 \leq P\left(| X_n - Y | > \frac{1}{2k}\right) < \frac{\delta}{2}. \tag{C.0.8}$$

Using the triangle inequality we observe that $| X - Y | \leq | X - X_n | + | X_n - Y |$ for all $n \geq 1$. So,

$$A_k \subset \{| X_n - X | > \frac{1}{2k}\} \cup \{| X_n - X | > \frac{1}{2k}\} \tag{C.0.9}$$

for all $n \geq 1$. Combining (C.0.8) and (C.0.9) we have (using any $n \geq N$)

$$0 \leq P(A_k) \leq P\left(| X_n - X | > \frac{1}{2k}\right) + P\left(| X_n - Y | > \frac{1}{2k}\right) \leq \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

As $\delta > 0$ was arbitrary we have $P(A_k) = 0$. Further by Exercise 1.1.13,

$$P(X \neq Y) = \lim_{k \to \infty} P(A_k) = 0.$$

Hence $P(X = Y) = 1$. ■

Proof of Theorem C.0.1**(Special Case)**- We provide a complete proof of Theorem C.0.1 in the special case when the random variables are i.i.d Bernoulli (p) random variables. We will proceed in two steps.

Step 1: $\overline{X}_n$ converges with probability one to a random variable X.

Let $\overline{S} = \limsup_{n \to \infty} \overline{X}_n$ and $\underline{S} = \liminf_{n \to \infty} \overline{X}_n$. Clearly,

$$0 \leq \underline{S} \leq \overline{S} \leq 1.$$

Fix $\epsilon > 0$, then for every k define

$$N_k = \inf\{n \in \mathbb{N} : \frac{X_k + X_{k+1} \ldots + X_{k+n-1}}{n} \geq \overline{S} - \epsilon\}.$$

The random variable N_k, in some sense, measures how close we are to $\overline{S}$ and our main effort will be to control the size N_k. It is easy to see that N_k is finite a.e. and are all identically distributed (because of independence of X_i). Hence we can choose an m such that $P(N_k > m) < \epsilon$ for all k. Define random variables Y_k and N_k^Y by the following mechanism:

$$Y_k = \begin{cases} X_k & \text{if } N_k \leq m \\ 1 & \text{if } N_k > m \end{cases} \tag{C.0.10}$$

$$N_k^Y = \inf\{n \in \mathbb{N} : \frac{Y_k + Y_{k+1} \ldots + Y_{k+n-1}}{n} \geq \overline{S} - \epsilon\}. \tag{C.0.11}$$

Clearly $N_k^Y \leq N_k$ and if k is such that $N_k \geq m$ then $N_k^Y = 1$ (since setting $Y_k = 1$ ensures that we are above $\overline{S} - \epsilon$ immediately). So we have

$$N_k^Y \leq m. \;\; a.e.$$

So for large enough $n \in \mathbb{N}$ we can break up $\sum_{k=1}^n Y_k$ into pieces of lengths atmost M such that the average over each piece is atleast $\overline{S} - \epsilon$. Then finally stop at the n-th term. Then it is clear that,

$$\sum_{k=1}^{n} Y_k \geq (n-m)(\overline{S} - \epsilon). \tag{C.0.12}$$

By our choice of m

$$E(Y_k) = E(X_k 1(N_k \leq m)) + P(N_k > m) < E(X_k) + \epsilon = E(X) + \epsilon,$$

for any k. Take expectations in (C.0.12) and use the above inequality to obtain

$$n(E(X) + \epsilon) \geq (n-m)(E(\overline{S}) - \epsilon).$$

Divide by n and first let $n \to \infty$ followed by $\epsilon \to 0$, to get

$$E(\overline{S}) \leq E(X). \tag{C.0.13}$$

Let $\widetilde{X_k} = 1 - X_k$. Applying the above argument to $\widetilde{X}$ (verify this) we have

$$E(\overline{\widetilde{S}}) \leq E(\widetilde{X}).$$

Since $\underline{S} = -\overline{\widetilde{S}}$ this implies

$$E(\underline{S}) \geq E(X). \tag{C.0.14}$$

Now, $\underline{S} \leq \overline{S}$ a.e. So only way (C.0.14) and (C.0.13) can hold only if $\underline{S} = \overline{S} a.e.$ Therefore $\lim_{n\to\infty} \overline{X}_n$ exists almost everywhere and let us call it X. This completes step 1.

Step 2: We shall now use the Weak Law of Large numbers (Theorem 8.2.1), along with Proposition C.0.3, and Lemma C.0.4 to complete the proof. The weak law implies that

$$\overline{X}_n \xrightarrow{p} \mu \text{ as } n \to \infty.$$

From Step 1, we know that

$$\overline{X}_n \xrightarrow{w.p.1} X \text{ as } n \to \infty.$$

Proposition C.0.3 then implies that

$$\overline{X}_n \xrightarrow{p} X \text{ as } n \to \infty.$$

Finally Lemma C.0.4 implies $P(X = \mu) = 1$. Therefore

$$\overline{X}_n \xrightarrow{w.p.1} \mu \text{ as } n \to \infty.$$

■

Proof of Theorem C.0.1**(General Case)** The essence of the proof is contained in the special case proven above. We provide a sketch of the proof.

Case 1:$(\mathbf{0 \leq X \leq 1})$ An imitation of Step 1 of the proof for Bernoulli p random variables will show that there is a limit. Step 2 of the above proof follows readily.

Case 2: Bounded Case When the random variable X is bounded, i.e. $\mid X \mid \leq M$ for some $M > 0$. One can consider $Y = \frac{X-M}{2M}$ and $Y_i = \frac{X_i - M}{2M}$. As $0 \leq Y \leq 1$ then one can use Case 1 for Y_i to establish that there is a limit. Step 2 of the above proof follows readily.

Case 3: (General Case by Truncation) One fixes $\alpha, \beta > 0$ and defines

$$\overline{S}_{(\alpha)} = \min\{\overline{S}, \alpha\}, X^{(\beta)} = \max\{X, -\beta\} \text{ and } X_k^{(\beta)} = \max\{X_k, -\beta\} \;\forall k \in \mathbb{N}.$$

The above quantities are all bounded. One imitates Step 1 of the above proof and this will result in inequalities depending on α, β. One then allows α, β approach infinity to establish that $\overline{X}_n \xrightarrow{w.p.1} X$ for a random variable X. Step 2 of the above proof follows readily. We refer the reader to [AS09] for the complete proof. ■