Strong Law of Large Numbers

Will Perkins

February 12, 2013

The Theorem

Theorem (Strong Law of Large Numbers)

Let $X_1, X_2, \ldots$ be iid random variables with a finite first moment, $\mathbb{E}X_i = \mu$. Then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$$

almost surely as $n \to \infty$.

The word 'Strong' refers to the type of convergence, almost sure.

We'll see the proof today, working our way up from easier theorems.

Using Chebyshev's Inequality, we saw a proof of the Weak Law of Large Numbers, under the additional assumption that X_i has a finite variance.

Under an even stronger assumption we can prove the Strong Law.

Theorem (Take 1)

Let $X_1, \ldots$ be iid, and assume $\mathbb{E}X_i = \mu$ and $\mathbb{E}X_i^4 = m_4 < \infty$. Then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$$

almost surely as $n \to \infty$.

Proof with a 4th moment

Proof: Since we have a finite 4th moment, we can try a 4th moment version of Chebyshev:

$$\Pr[|Z - \mathbb{E}Z| > \epsilon] \leq \frac{\mathbb{E}|Z - \mathbb{E}Z|^4}{\epsilon^4}$$

First to simplify, we can assume $\mathbb{E}X_i = 0$ just by subtracting μ from each.
Now let $U_n = \frac{X_1+X_2+\cdots+X_n}{n}$. $\mathbb{E}U_n = 0$.
Then calculate

$$\mathbb{E}U_n^4 = \frac{1}{n^4}\mathbb{E}[\sum_i X_i^4 + 4\sum_{i\neq j} X_i X_j^3 + 3\sum_{i\neq j} X_i^2 X_j^2 + 6\sum_{i,j,k} X_i X_j X_k^2 + \sum_{i,j,k,l} X_i X_j X$$

Proof with a 4th moment

Now all the terms with an X_i to the first power are 0 in expectation. [Why?]
Which leaves:

$$\mathbb{E}U_n^4 = \frac{1}{n^4}\left[n\mathbb{E}X_i^4 + 3n(n-1)\mathbb{E}X_i^2X_j^2\right]$$

$$\leq \frac{m_4}{n^3} + \frac{3\sigma^4}{n^2}$$

Now applying the 4th moment Markov's Inequality:

$$\Pr[|U_n - \mathbb{E}U_n| > \epsilon] \leq \frac{\frac{m_4}{n^3} + \frac{3\sigma^4}{n^2}}{\epsilon^4}$$

Proof with a 4th moment

But for ϵ fixed, we can sum the RHS from $n = 1$ to ∞ and get a finite sum. ($1/n^2$ is summable).

Now apply Borel-Cantelli: fix $\epsilon > 0$, and let A_n^ϵ be the event that $|U_n| > \epsilon$. We've shown that

$$\sum_{n=1}^{\infty} \Pr(A_n^\epsilon) < \infty$$

and so by the Borel-Cantelli Lemma, with probability 1, only finitely many of the A_n^ϵ's occur.

This is precisely what it means for $U_n \to 0$ almost surely.

Removing Higher Moment Conditions

What remains is to remove the conditions for X_i to have finite higher moments.

Strong Law with 2nd Moment

Theorem (Take 2)

Let $X_1, \ldots$ be iid with mean μ and variance σ^2. Then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu$$

almost surely as $n \to \infty$.

Two tricks:

1. Assume X_i's are non-negative
2. First prove for a subsequence

Non-negativity

Let $X_i = X_i^+ - X_i^-$ where $X_i^+ = \max\{0, X_i\}$, $X_i^- = -\min\{0, X_i\}$
X_i^+ and X_i^- are both non-negative, with finite expectation and variance, so if we prove the SLLN holds for non-negative RV's, we can apply spearately to the two parts and recombine.

Subsequence

We will find a subsequence of natural numbers so that the empirical averages along the subsequence converge alsmost surely. The subsequence will be explicit: $1, 4, 9, \dots n^2, \dots$

Let

$$A^{\epsilon}_{n^2} = \left\{ \left| \frac{X_1 + \cdots + X_{n^2}}{n^2} - \mu \right| > \epsilon \right\}$$

We bound with Chebyshev

$$\Pr(A^{\epsilon}_{n^2}) \leq \frac{\mathrm{var}\left(\frac{X_1 + \cdots + X_{n^2}}{n^2}\right)}{\epsilon^2}$$

Subsequence

$$\text{var}\left(\frac{X_1 + \cdots + X_{n^2}}{n^2}\right) = \frac{1}{n^4} n^2 \sigma^2 = \frac{\sigma^2}{n^2}$$

So

$$\sum_n \Pr(A^{\epsilon}_{n^2}) \leq \sum \frac{\sigma^2}{\epsilon^2 n^2} < \infty$$

Applying the Borel-Cantelli Lemma shows that along the subsequence $\{n^2\}$, the empirical averages converge to μ almost surely.

From Subsequence to Full Sequence

We want to show that for every $\epsilon > 0$ with probability 1 there is N large enough so that

$$\left| \frac{X_1 + \cdots + X_n}{N} - \mu \right| < \epsilon$$

We know this holds for large enough $N = n^2$. And here is where we will use non-negativity.
Start by picking n large enough so that

$$\left| \frac{X_1 + \cdots + X_{n^2}}{n^2} - \mu \right| < \epsilon/3$$

and

$$\left| \frac{X_1 + \cdots + X_{(n+1)^2}}{(n+1)^2} - \mu \right| < \epsilon/3$$

From Subsequence to Full Sequence

For $n^2 \leq N \leq (n+1)^2$,

$$\frac{X_1 + \cdots + X_{n^2}}{(n+1)^2} \leq \frac{X_1 + \cdots + X_n}{N^2} \leq \frac{X_1 + \cdots + X_{(n+1)^2}}{n^2}$$

and

$$\left(\mu - \frac{\epsilon}{3}\right) \frac{n^2}{(n+1)^2} \leq \frac{X_1 + \cdots + X_{n^2}}{(n+1)^2}$$

and

$$\frac{X_1 + \cdots + X_{(n+1)^2}}{n^2} \leq \left(\mu + \frac{\epsilon}{3}\right) \frac{(n+1)^2}{n^2}$$

If n is large enough so that $\frac{n^2}{(n+1)^2}$ is close to 1, then we are done.

Removing the finite variance condition

To get the full theorem under the fewest conditions we need one more trick: truncation.

Again assume that $X_i \geq 0$, with $\mathbb{E}X_i = \mu < \infty$.
Let $Y_n = \min\{X_n, n\}$.

Fact: $X_n - Y_n \to 0$ almost surely.
Proof:

$$\sum_n \Pr[X_n \neq Y_n] = \sum_n \Pr[X_1 > n] \leq \mathbb{E}X_1 < \infty$$

and apply Borel-Cantelli.
In particular, it's enough to prove the strong law for the Y_n's.

Removing the finite variance condition

Now we apply the same methods we've used before.

This time we will use an even sparser subsequence, $1, c, c^2, c^3, \ldots$ for some $c > 1$ which will depend on ϵ.
The main estimate we need to apply Borel-Cantelli is:

$$\sum_{j=1}^{\infty} \frac{1}{c^j} \min\{X_i, c^j\}^2 = O(X_j)$$

and so

$$\sum_{j=1}^{\infty} \frac{1}{c^j} \mathbb{E}[Y_{c^j}]^2 < \infty$$

Removing the finite variance condition

Now we use Chebyshev:
Let

$$A^{\epsilon}_{c^j} = \left\{ \left| \frac{Y_1 + \cdots + Y_{c^j}}{c^j} - \mu \right| > \epsilon \right\}$$

and

$$\Pr(A^{\epsilon}_{c^j}) \leq \frac{\text{var}\left(\frac{Y_1+\cdots+Y_{c^j}}{c^j}\right)}{\epsilon^2}$$

$$\leq \frac{1}{\epsilon^2 c^j} \mathbb{E}[Y_{c^j}]^2$$

Finishing Up

From above,

$$\sum_{j=1}^{\infty} \frac{1}{\epsilon^2 c^j} \mathbb{E}[Y_{c^j}]^2 < \infty$$

and so Borel-Cantelli says that along the subsequence c^j, the empirical averages converge almost surely.

Again we can use the fact that the Y_i's are non-negative to go from the sparse sequence to the full sequence.