# CLASSICAL PROBABILITY 2008
# 3. THE LAW OF LARGE NUMBERS

JOHN MORIARTY

At school, the first experiment in probability might be to flip a fair coin a large number of times and calculate the ratio of heads to tails. The student might be tempted to take a short cut and invent the data - because they believe strongly that the ratio should approximate 1. How would we prove that the student is right?

Let $X_1, X_2, X_3, ...$ be a sequence of independent random variables, and let

$$S_n = \sum_{k=1}^{n} X_k,$$

which we call the $n$th partial sum. A law of large numbers provides conditions under which there exists a constant $\mu$ such that

$$\frac{S_n}{n} \to \mu \tag{1}$$

as $n \to \infty$. From the previous section, we know that statement (1) isn't precise - we must also specify the mode of convergence. If we have convergence in probability, then we call it a <u>weak law of large numbers</u>, whereas a <u>strong law of large numbers</u> means convergence almost surely. We see from Theorem 2.2 that if the strong law holds then the weak law holds - which helps to explain these names.

## 1. A weak law of large numbers

It is fairly easy to prove a weak law of large numbers using Chebychev's inequality.

**Theorem 1.1. (A weak law of large numbers)** *Assume that $X_1, X_2, ...$ are independent (but not necessarily identically distributed). Also assume that $E(X_n^2) \leq M$ for some constant $M > 0$. Then*

$$\frac{S_n}{n} - \frac{E(S_n)}{n} \to 0$$

*in probability as $n \to \infty$.*

*Proof.* Let $Y_n := X_n - E(X_n)$ and let $S'_n = \sum_{k=1}^n Y_k$. By Chebychev's inequality,

$$\begin{aligned} P\left(\frac{|S'_n|}{n} \geq \epsilon\right) &\leq \frac{E(S'_n)^2}{n^2\epsilon^2} = \frac{E(Y_1 + ... + Y_n)^2}{n^2\epsilon^2} \\ &= \frac{\sum_{k=1}^n E(Y_k^2)}{n^2\epsilon^2} \leq \frac{M}{n\epsilon^2} \to 0 \end{aligned}$$

as $n \to \infty$. Since

$$\frac{S_n}{n} - \frac{E(S_n)}{n} = \frac{S'_n}{n},$$

this finishes the proof. □

We can also give a short proof for a strong law of large numbers, for iid sequences, if we assume that the fourth moment is finite. Later, in Section 3, we will see a version with less restrictive assumptions—the tradeoff is that its proof will be longer.

**Theorem 1.2. (A strong law of large numbers)** *Assume $X_1, X_2, ...$ are iid with $E(X_1^4) < \infty$. Then $S_n/n \to \mu$ almost surely as $n \to \infty$, where $\mu = E(X_1)$.*

*Proof.* Assume without loss of generality that $E(X_n) = 0$. Recall that, to prove $S_n/n \to \mu$ almost surely, it suffices to show that

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \infty$$

for all $\epsilon > 0$ (Consequence 2.7 in the second part of the lecture notes). From Chebychev's inequality we have

$$P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) \leq \frac{E(S_n^4)}{n^4\epsilon^4}.$$

Moreover,

$$\begin{aligned} E(S_n^4) &= E(X_1 + ... + X_n)^4 \\ &= nE(X_1^4) + \binom{4}{2}\binom{n}{2}(E(X_1^2))^2 \leq C(n^2+1) \end{aligned}$$

for some constant $C$ large enough (the remaining terms are 0 due to independence and $E(X_n) = 0$). Consequently,

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_n}{n}\right| > \epsilon\right) \leq \sum_{n=1}^{\infty} \frac{C(n^2+1)}{\epsilon^4 n^4} < \infty,$$

which finishes the proof. □

We will now concentrate on proving the less restrictive version of the strong law of large numbers, mentioned above. The version we prove next only requires a first moment.

## 2. Preparations for the general version of the strong law of large numbers

In this section we provide a few results which will be needed in the next section. However, especially the dominated convergence result Theorem 2.3 and Theorem 2.4 are also of independent interest.

We begin with two lemmas from analysis, the first of which is very intuitively appealing.

**Lemma 2.1. (The Cesaro lemma)** *Let $a_1, a_2, \ldots$ be a sequence of real numbers. If $a_n \to a$ as $n \to \infty$, then also*

$$\frac{\sum_{k=1}^n a_k}{n} \to a$$

*as $n \to \infty$.*

*Proof.* Let $\epsilon > 0$ and let $N_0$ be a large number so that

$$|a_n - a| \leq \epsilon/2 \text{ for all } n \geq N_0.$$

With $N_0$ fixed, we may choose $N_1 > N_0$ so that

$$\frac{\sum_{k=1}^{N_0} |a_k - a|}{N_1} \leq \epsilon/2.$$

For $n > N_1$ we then have

$$\begin{aligned}
\left| \frac{\sum_{k=1}^n a_k}{n} - a \right| &\leq \frac{\sum_{k=1}^n |a_k - a|}{n} \\
&= \frac{\sum_{k=1}^{N_0} |a_k - a|}{n} + \frac{\sum_{k=N_0}^n |a_k - a|}{n} \\
&\leq \frac{\sum_{k=1}^{N_0} |a_k - a|}{N_1} + \frac{\epsilon(n - N_0)}{2n} \leq \epsilon,
\end{aligned}$$

which finishes the proof. □

**Lemma 2.2. (The Kronecker lemma)** *Let $a_1, a_2, \ldots$ be a sequence of real numbers. If*

$$\sum_{n=1}^{\infty} \frac{a_n}{n}$$

*converges, then*

$$\frac{\sum_{k=1}^n a_k}{n} \to 0$$

*as $n \to \infty$.*

*Proof.* Let

$$S_n = \sum_{k=1}^n \frac{a_k}{k}.$$

Then $S_n$ converges by assumption. Moreover, if we define $S_0 = 0$, then

$$\sum_{k=1}^{n} a_k = \sum_{k=1}^{n} k(S_k - S_{k-1}) = nS_n - \sum_{k=1}^{n} S_{k-1},$$

so

$$\frac{\sum_{k=1}^{n} a_k}{n} = S_n - \frac{\sum_{k=1}^{n} S_{k-1}}{n} \to 0$$

as $n \to \infty$ by the Cesaro lemma. □

**Theorem 2.3. (Dominated convergence)** *Assume $X_n \to X$ almost surely. Also assume there exists a random variable $Z$ such that $|X_n| \leq Z$ almost surely for all $n$, and $EZ < \infty$. Then*

$$EX_n \to EX$$

*as $n \to \infty$.*

*Proof.* Let $Y_n = |X - X_n|$, and let $\epsilon > 0$ and $M > 0$. Then

$$\begin{aligned} EY_n &= E\left(Y_n 1_{\{Y_n \leq \epsilon\}}\right) + E\left(Y_n 1_{\{Y_n > \epsilon\}}\right) \\ &\leq \epsilon + 2E\left(Z 1_{\{Y_n > \epsilon\}}\right) \\ &= \epsilon + 2E\left(Z 1_{\{Y_n > \epsilon\}} 1_{\{Z \leq M\}}\right) + 2E\left(Z 1_{\{Y_n > \epsilon\}} 1_{\{Z > M\}}\right) \\ &\leq \epsilon + 2MP(Y_n > \epsilon) + 2E\left(Z 1_{\{Z > M\}}\right). \end{aligned}$$

Letting $M \to \infty$, the last term can be made arbitrary small since$^*$ $EZ < \infty$, and then the middle term can be made arbitrary small by letting $n \to \infty$ since $X_n \to X$ in probability. Thus we have that$^*$

$$|EX_n - EX| \leq E|X_n - X| = EY_n \to 0$$

as $n \to \infty$, which finishes the proof.

*: these steps may seems reasonable, but actually we cannot justify them yet from what we have done. I will fix this if there is time. □

**Question.** Suppose you know the distribution function $F$ of a random variable $X$. What can you say about the mean, $E(X)$?

The next result gives a direct way to estimate $E(X)$ from $F$, without any intermediate steps.

**Theorem 2.4.** *Assume $X$ is non-negative. Then*

$$\sum_{n=1}^{\infty} P(X \geq n) \leq EX \leq 1 + \sum_{n=1}^{\infty} P(X \geq n).$$

*In particular, $EX < \infty$ if and only if $\sum_{n=1}^{\infty} P(X \geq n) < \infty$.*

*Proof.*

$$\begin{aligned}
\sum_{n=1}^{\infty} P(X \geq n) &= \sum_{n=1}^{\infty} \sum_{k \geq n} P(k \leq X < k+1) \\
&= \sum_{k=1}^{\infty} \sum_{n=1}^{k} P(k \leq X < k+1) \\
&= \sum_{k=0}^{\infty} kP(k \leq X < k+1) \\
&\leq \sum_{k=0}^{\infty} E(X 1_{\{k \leq X < k+1\}}) = E(X)
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(X) &\leq \sum_{k=0}^{\infty} (k+1) P(k \leq X < k+1) \\
&= \sum_{n=1}^{\infty} P(X \geq n) + \sum_{k=0}^{\infty} P(k \leq X < k+1) \\
&= \sum_{n=1}^{\infty} P(X \geq n) + 1.
\end{aligned}$$

□

Recall that Chebyshev's inequality from Lecture Notes 2 gives a bound on the probability that a random variable $X$ is large. Recall also the simulated random walks shown in lectures, where the position of the walker at time $n$ was the $n$th partial sum of her steps. The next result combines these two ideas, giving a bound on the probability that the random walk is large at any time up to time $n$.

**Lemma 2.5. (Kolmogorov's inequality)** *Assume $X_1, X_2, ..., X_n$ are independent, that $E(X_k) = 0$ and that $E(X_k^2) < \infty$ for all $k$. Then, for all $\epsilon > 0$ we have*

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \epsilon\right) \leq \frac{E(S_n^2)}{\epsilon^2}.$$

*Proof.* We keep track of the first time that $S_k$ gets large; this gives a set of disjoint events, making the calculations easier. Let

$$A_k = \{|S_i| < \epsilon \text{ for } i = 1, 2, ..., k-1 \text{ and } |S_k| \geq \epsilon\},$$

and set

$$A = \bigcup_{k=1}^{n} A_k = \{\max_{1 \leq k \leq n} |S_k| \geq \epsilon\}.$$

Then

$$E(S_n^2) \geq E(S_n^2 1_A) = \sum_{k=1}^{n} E(S_n^2 I_{A_k}),$$

where

$$\begin{aligned} E(S_n^2 I_{A_k}) &= E\left((S_k + X_{k+1} + ... + X_n)^2 1_{A_k}\right) \\ &= E(S_k^2 1_{A_k}) + 2E\left(S_k(X_{k+1} + ... + X_n) 1_{A_k}\right) \\ &\quad + E\left((X_{k+1} + ... + X_n)^2 1_{A_k}\right) \\ &\geq E(S_k^2 1_{A_k}) \geq \epsilon^2 P(A_k), \end{aligned}$$

where we have used the fact that $X_1, \dots, X_k$ are independent of $X_{k+1}, \dots, X_n$. **(Question:** Can you see why $E(S_k X_{k+1} 1_{A_k}) = 0$? (Hint: think about independence.) Can you see where this is used in the proof?)

Consequently,

$$E(S_n^2) \geq \epsilon^2 \sum_{k=1}^{n} P(A_k) = \epsilon^2 P(A),$$

which finishes the proof. □

When studying sequences and series, you met various ways to test whether a series of numbers converges—ratio test, integral test, . . . But what about a series of random numbers? Intuitively, you might think about a random walk - when does the walker's position converge?

The next result gives sufficient conditions.

**Lemma 2.6.** *Assume $X_1, X_2, ...$ are independent with $E(X_n) = 0$. Also assume that $\sum_{n=1}^{\infty} E(X_n^2) < \infty$. Then*

$$S_n = \sum_{k=1}^{n} X_k$$

*converges almost surely as $n \to \infty$.*

*Proof.*

$$\begin{aligned} P\left(\sup_{k \geq 0} |S_{n+k} - S_n| \geq \epsilon\right) &= \lim_{N \to \infty} P\left(\sup_{0 \leq k \leq N} |S_{n+k} - S_n| \geq \epsilon\right) \\ &\leq \lim_{N \to \infty} \frac{E(X_{n+1} + ... + X_{n+N})^2}{\epsilon^2} \\ &= \frac{\sum_{k=n+1}^{\infty} E(X_k^2)}{\epsilon^2} \to 0 \end{aligned}$$

as $n \to \infty$ (note that we used Kolmogorov's inequality above). It follows (why?) that $S_n$ converges almost surely as $n \to \infty$. (Hint: think about Cauchy sequences, and note that the events $A_n = \{\sup_{k,l \geq n} |S_k - S_l| \geq 2\epsilon\}$ are decreasing as $n \to \infty$) □

**Example** Let $X_1, X_2, ...$ be independent, and assume

$$X_n = \begin{cases} n^{1-\epsilon} & \text{with probability } \frac{1}{2n\log n} \\ -n^{1-\epsilon} & \text{with probability } \frac{1}{2n\log n} \\ 0 & \text{with probability } 1 - \frac{1}{n\log n} \end{cases}$$

for some $\epsilon > 0$. Then since

$$\sum_{n=1}^{\infty} \frac{V(X_n)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^{1+2\epsilon}\log n} < \infty,$$

Lemma 2.6 shows that

$$P\left(\sum_{n=1}^{\infty} \frac{X_n}{n} \text{ converges}\right) = 1$$

so by the Kronecker lemma,

$$P\left(\frac{X_1 + ... + X_n}{n} \to 0 \text{ as } n \to \infty\right) = 1.$$

Compare this with Problem 3 on Exercise Sheet 5 which shows that $S_n/n$ does not converge almost surely in the example above if $\epsilon = 0$.

## 3. The strong law of large numbers

Now we are ready to prove a general version of the strong law of large numbers.

**Theorem 3.1. Strong law of large numbers** *Let $X_1, X_2, ...$ be a sequence of iid random variables such that $E|X_1| < \infty$. Then*

$$\frac{S_n}{n} \to E(X_1)$$

*almost surely as $n \to \infty$.*

*Proof.* Without loss of generality (why?) we may assume that $E(X_n) = 0$. Since $E|X_1| < \infty$ we have

$$\sum_n P(|X_n| \geq n) = \sum_n P(|X_1| \geq n) < \infty,$$

so the first Borel-Cantelli lemma implies that

$$P\left(|X_n| < n \text{ for all but finitely many } n\right) = 1.$$

Define the sequence $\tilde{X}_1, \tilde{X}_2, ...$ by

$$\tilde{X}_n = \begin{cases} X_n & \text{if } |X_n| < n \\ 0 & \text{if } |X_n| \geq n, \end{cases}$$

and note that $S_n/n \to 0$ almost surely if and only if $(\tilde{X}_1 + ... + \tilde{X}_n)/n \to 0$ almost surely. Moreover,

$$E(\tilde{X}_n) = E(X_1 1_{\{|X_1|<n\}}) \to E(X_1) = 0$$

by dominated convergence. Thus

$$\frac{\sum_{k=1}^{n} E(\tilde{X}_k)}{n} \to 0$$

by the Cesaro lemma, so

$$\frac{S_n}{n} \to 0 \text{ a.s.} \iff \frac{\sum_{k=1}^{n} Y_k}{n} \to 0 \text{ a.s.},$$

where

$$Y_k := \tilde{X}_k - E(\tilde{X}_k).$$

Thus, by the Kronecker Lemma it suffices to show that $\sum_{k=1}^{\infty} \frac{Y_k}{k}$ converges almost surely. According to Lemma 2.6 above, it therefore suffices to show that

$$\sum_{n=1}^{\infty} \frac{V(Y_n)}{n^2} < \infty.$$

But we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{V(Y_n)}{n^2} &\leq \sum_{n=1}^{\infty} \frac{E(\tilde{X}_n^2)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} E\left(X_1^2 1_{\{|X_1|<n\}}\right) \\
&= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} E\left(X_1^2 1_{\{k-1\leq|X_1|<k\}}\right) \\
&= \sum_{k=1}^{\infty} E\left(X_1^2 1_{\{k-1\leq|X_1|<k\}}\right) \sum_{n=k}^{\infty} \frac{1}{n^2} \\
&\leq \sum_{k=1}^{\infty} \frac{2}{k} E\left(X_1^2 1_{\{k-1\leq|X_1|<k\}}\right) \\
&\leq 2\sum_{k=1}^{\infty} E\left(|X_1| 1_{\{k-1\leq|X_1|<k\}}\right) = 2E|X_1| < \infty,
\end{aligned}$$

which thus finishes the proof. □

We finish this section with a converse to the strong law of large numbers. It shows that the assumption about a finite moment is necessary. Thus there is no hope of proving a more general version than the one above.

**Theorem 3.2.** *Assume $X_1, X_2, \ldots$ are iid and that*

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

*almost surely as $n \to \infty$ for some $\mu$. Then $E|X_n| < \infty$, and $\mu = E(X_n)$.*

*Proof.* First note that, by assumption,

$$\frac{X_n}{n} = \frac{S_n}{n} - \frac{n-1}{n} \frac{S_{n-1}}{n-1} \to 0$$

almost surely as $n \to \infty$. Therefore

$$P(|X_n| \geq n \text{ i.o.}) = 0.$$

By the second Borel-Cantelli lemma,

$$\sum_{n=1}^{\infty} P(|X_1| \geq n) = \sum_{n=1}^{\infty} P(|X_n| \geq n) < \infty.$$

It follows from Theorem 2.4 above that $E|X_1| < \infty$. Moreover, from the strong law of large numbers, $\mu = E(X_n)$. $\square$