Math 639: Lecture 3

The law of large numbers

Bob Hough

January 31, 2017

Convergence in probability

Definition

A sequence of random variables $\{Y_n\}$ converges to Y *in probability* if for all $\epsilon > 0$,

$$\mathrm{Prob}(|Y_n - Y| > \epsilon) \to 0$$

as $n \to \infty$.

Uncorrelated variables

Recall that random variables X_1, X_2 are said to be uncorrelated if $\mathrm{E}\left[X_1^2\right] < \infty$, $\mathrm{E}\left[X_2^2\right] < \infty$ and $\mathrm{E}[X_1 X_2] = \mathrm{E}[X_1]\,\mathrm{E}[X_2]$.

Theorem

Let $X_1, ..., X_n$ be uncorrelated random variables satisfying $\mathrm{E}\left[X_i^2\right] < \infty$. Then

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

Uncorrelated variables

Proof.

- We may assume that each variable is mean 0, since both sides of the equation are unchanged under translation.
- We have

$$\mathrm{E}\left[(X_1 + \cdots + X_n)^2\right] = \mathrm{E}\left[X_1^2\right] + \cdots + \mathrm{E}\left[X_n^2\right],$$

since the cross-terms vanish.

□

Convergence in probability

Lemma

If $p > 0$ and $\mathrm{E}\left[|Z_n|^p\right] \to 0$ as $n \to \infty$ then $Z_n \to 0$ in probability.

Proof.

By Markov's inequality, for each $\epsilon > 0$, $\mathrm{Prob}(|Z_n| \geq \epsilon) \leq \epsilon^{-p}\,\mathrm{E}\left[|Z_n|^p\right]$, which gives the claim. □

L^2 weak law

Theorem (L^2 weak law)

Let $X_1, X_2, \ldots$ be uncorrelated random variables satisfying $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) \leq C < \infty$. If $S_n = X_1 + \ldots + X_n$ then as $n \to \infty$, $\frac{S_n}{n} \to \mu$ in L^2 and in probability.

L^2 weak law

Proof.

Observe

$$\mathrm{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\left(\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)\right) \leq \frac{Cn}{n^2} \to 0.$$

This proves convergence in L^2. Convergence in probability follows from the previous lemma. □

Independent and identically distributed

Definition

A sequence of random variables $X_1, X_2, X_3, \ldots$ which have the same distribution and are independent are called *independent and identically distributed* or *i.i.d.*.

The L^2 weak law applies to i.i.d. variables of finite variance.

Weierstrass approximation theorem

Example

Let f be a continuous function on $[0, 1]$. The *Bernstein polynomial* of degree n associated to f is

$$f_n(x) = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} f\left(\frac{m}{n}\right).$$

As a consequence of the weak law, we show that $f_n(x) \to f(x)$ uniformly as $n \to \infty$.

Weierstrass approximation theorem

Proof.

- Let S_n be the sum of n i.i.d. random variables satisfying $\mathrm{Prob}(X_i = 1) = p$, $\mathrm{Prob}(X_i = 0) = 1 - p$. Thus $\mathrm{E}[X_i] = p$, $\mathrm{Var}[X_i] = p - p^2$.
- Note
$$\mathrm{Prob}(S_n = m) = \binom{n}{m} p^m (1-p)^{n-m},$$
thus $\mathrm{E}\left[f\left(\frac{S_n}{n}\right)\right] = f_n(p)$.
- Given $\delta > 0$, by Chebyshev's inequality,
$$\mathrm{Prob}\left(\left|\frac{S_n}{n} - p\right| > \delta\right) \leq \frac{\mathrm{Var}\left(\frac{S_n}{n}\right)}{\delta^2} = \frac{p(1-p)}{n\delta^2} \leq \frac{1}{4n\delta^2}.$$

□

Weierstrass approximation theorem

Proof.

- Let $\delta > 0$ be such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$, and let $M = \sup_{x \in [0,1]} |f(x)|$.
- We have

$$\begin{aligned} \left| \mathrm{E}\left[f\left(\frac{S_n}{n} \right) \right] - f(p) \right| &\leq \mathrm{E}\left[\left| f\left(\frac{S_n}{n} \right) - f(p) \right| \right] \\ &\leq \epsilon + 2M \,\mathrm{Prob}\left(\left| \frac{S_n}{n} - p \right| > \delta \right) \\ &\leq \epsilon + \frac{M}{2n\delta^2}. \end{aligned}$$

The claim follows. □

Concentration of the 2-norm

Example

- Let $X_1, ..., X_n$ be independent and uniformly distributed on $(-1, 1)$. Their joint distribution is uniform measure on the cube $(-1, 1)^n$.
- Let $Y_i = X_i^2$. These variables are independent and satisfy $\mathrm{E}[Y_i] = \frac{1}{3}$ and $\mathrm{Var}[Y_i] \leq \mathrm{E}\left[Y_i^2\right] \leq 1$.
- The weak law implies $\frac{1}{n}\left(X_1^2 + \cdots + X_n^2\right) \to \frac{1}{3}$ in probability, as $n \to \infty$.
- Given $0 < \epsilon < 1$, let

$$A_{n,\epsilon} = \left\{x \in \mathbb{R}^n : (1-\epsilon)\sqrt{\frac{n}{3}} < \|x\|_2 < (1+\epsilon)\sqrt{\frac{n}{3}}\right\}.$$

- By the weak law, $\frac{|A_{n,\epsilon} \cap (-1,1)^n|}{2^n} \to 1$.

L^2 weak law, again

A slightly stronger variant of the L^2 weak law is as follows.

Theorem (L^2 weak law)

Let $X_1, X_2, ..., X_n$ be random variables satisfying $\mathrm{E}\left[X_i^2\right] < \infty$, and let $S_n = X_1 + \cdots + X_n$. Let $\mu_n = \mathrm{E}[S_n]$ and $\sigma_n^2 = \mathrm{Var}(S_n)$. Let $\{b_n\}$ be a sequence of non-zero numbers such that $\frac{\sigma_n^2}{b_n^2} \to 0$. Then $\frac{S_n - \mu_n}{b_n} \to 0$ in probability.

Proof.

Since $\mathrm{E}\left[\left(\frac{S_n-\mu_n}{b_n}\right)^2\right] = \frac{\mathrm{Var}[S_n]}{b_n^2} \to 0$, the conclusion follows from Chebyshev's inequality. □

Coupon collector's problem

- Let $X_1, X_2, \ldots$ be i.i.d. on $\{1, 2, \ldots, n\}$
- Let $\tau_k^n = \inf\{m : |\{X_1, \ldots, X_m\}| = k\}$ be the waiting time until collecting the kth distinct coupon. Set $\tau_0^n = 0$.
- We are interested in $T_n = \tau_n^n$, the waiting time until collecting a complete set of coupons.

Coupon collector's problem

- Let $Y_{n,k} = \tau_k^n - \tau_{k-1}^n$ be the incremental waiting time to collect the kth coupon. $Y_{n,k}$ has a geometric distribution with parameter $1 - \frac{k-1}{n}$.
- A geometric distribution with parameter p has mean $\frac{1}{p}$ and variance $\leq \frac{1}{p^2}$.
- Hence $T_n = \sum_{k=1}^n Y_{n,k}$ satisfies

$$\mathrm{E}[T_n] = \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right)^{-1} = n \sum_{k=1}^n \frac{1}{k} \sim n \log n$$

$$\mathrm{Var}[T_n] \leq \sum_{k=1}^n \left(1 - \frac{k-1}{n}\right)^2 < n^2 \sum_{m=1}^\infty \frac{1}{m^2}.$$

- Taking $b_n = n \log n$ in the previous theorem proves $\frac{T_n - n\sum_{m=1}^n \frac{1}{m}}{n \log n} \to 0$ in probability, or $\frac{T_n}{n \log n} \to 1$ in probability.

Random permutations

The cycle representation of a permutation π on $\{1, 2, ..., n\}$ is found by writing

$$(1, \pi(1), \pi^2(1), ..., \pi^{k-1}(1))$$

where k is the least positive integer such that $\pi^k(1) = 1$, then repeating this process starting with the least number not contained in $1, \pi(1), ..., \pi^{k-1}(1)$, and iterating. For example, the permutation

i	1	2	3	4	5	6	7	8	9
$\pi(i)$	3	9	6	8	2	1	5	4	7

has cycle structure $(1, 3, 6)(2, 9, 7, 5)(4, 8)$.

Random permutations

- Let π be chosen at uniform from the symmetric group $\mathfrak{S}_n$ on n letters.
- Let $X_{n,k}$ indicate the event that the kth letter in the cycle structure of π closes a cycle, and let $S_n = \sum_{k=1}^n X_{n,k}$ denote the number of cycles.

Random permutations

Lemma

The events $X_{n,1}, \ldots, X_{n,n}$ *are independent, and* $\text{Prob}(X_{n,j} = 1) = \frac{1}{n-j+1}$.

Proof.

- Build the cycle structure at random left to right, starting from 1, by assigning $\pi(i)$ only once i is reached in the cycle structure.
- The number of choices for $\pi(i)$ is $n - k + 1$ where k is the position of i in the cycle structure, and exactly one choice leads to completing a cycle.

□

Random permutations

By the previous lemma,

$$\mathrm{E}[S_n] = \frac{1}{n} + \frac{1}{n-1} + \ldots + 1$$

$$\mathrm{Var}[S_n] = \sum_{k=1}^{n} \mathrm{Var}[X_{n,k}] \le \sum_{k=1}^{n} \mathrm{E}\left[X_{n,k}^2\right] \le \mathrm{E}[S_n].$$

It follows that for $\epsilon > 0$,

$$\frac{S_n - \sum_{m=1}^{n} \frac{1}{m}}{(\log n)^{\frac{1}{2}+\epsilon}} \to 0$$

in probability.

Occupancy

- Suppose that r balls are dropped independently at random in n boxes, so that each of n^r assignments is equally likely.
- Let A_i be the event that box i is empty, and $N = \sum_i A_i$ the number of empty boxes.
- We have

$$\mathrm{Prob}[A_i] = \left(1 - \frac{1}{n}\right)^r, \qquad \mathrm{E}[N] = n\left(1 - \frac{1}{n}\right)^r.$$

- If $r/n \to c$ then $\frac{1}{n}\mathrm{E}[N] \to e^{-c}$.

Occupancy

- Calculate

$$
\begin{aligned}
\mathrm{E}\left[N^2\right] &= \mathrm{E}\left[\left(\sum_{m=1}^{n} \mathbf{1}_{A_m}\right)^2\right] = \sum_{1\le k,m\le n} \mathrm{Prob}(A_k \cap A_m)\\
\mathrm{Var}[N] &= \mathrm{E}\left[N^2\right] - \mathrm{E}[N]^2\\
&= \sum_{1\le k,m\le n} \mathrm{Prob}(A_k \cap A_m) - \mathrm{Prob}(A_k)\,\mathrm{Prob}(A_m)\\
&= n(n-1)\left[\left(1-\frac{2}{n}\right)^r - \left(1-\frac{1}{n}\right)^{2r}\right] + O(n)\\
&= O(n).
\end{aligned}
$$

- It follows that $\frac{N}{n} \to e^{-c}$ in probability.

Triangular arrays

Definition

A *triangular array* of random variables is a collection $\{X_{n,k}\}_{1\leq k\leq n}$. Many classical limit theorems of probability theory apply to the row sums

$$S_n = \sum_{1\leq k\leq n} X_{n,k}.$$

Truncation

Definition

Let $M > 0$. The *truncation* at height M of random variable X is

$$\overline{X} = X\mathbf{1}_{(|X| \leq M)} = \begin{cases} X & |X| \leq M \\ 0 & |X| > M \end{cases}.$$

Weak law for triangular arrays

Theorem (Weak law for triangular arrays)

For each n let $X_{n,k}$, $1 \leq k \leq n$, be independent. Let $b_n > 0$ with $b_n \to \infty$, and let $\overline{X}_{n,k} = X_{n,k}\mathbf{1}_{(|X_{n,k}| \leq b_n)}$. Suppose that as $n \to \infty$,

- $\sum_{k=1}^n \mathrm{Prob}(|X_{n,k}| > b_n) \to 0$, *and*
- $b_n^{-2} \sum_{k=1}^n \mathrm{E}\left[\overline{X}_{n,k}^2\right] \to 0.$

Set $S_n = X_{n,1} + X_{n,2} + ... + X_{n,n}$ and $a_n = \sum_{k=1}^n \mathrm{E}[\overline{X}_{n,k}]$. Then $\frac{S_n - a_n}{b_n} \to 0$ in probability.

Weak law for triangular arrays

Proof.

- Let $\overline{S}_n = \overline{X}_{n,1} + \cdots + \overline{X}_{n,n}$. Bound

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

- Use a union bound to estimate

$$\begin{aligned}\mathrm{Prob}(S_n \neq \overline{S}_n) &\leq \mathrm{Prob}\left(\bigcup_{k=1}^{n}\{X_{n,k} \neq \overline{X}_{n,k}\}\right)\\ &\leq \sum_{k=1}^{n}\mathrm{Prob}(|X_{n,k}| > b_n) \to 0.\end{aligned}$$

□

Weak law for triangular arrays

Proof.

- The second term is bounded by

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

□

Moments

Lemma

If $Y \geq 0$ *and* $p > 0$ *then* $\mathrm{E}[Y^p] = \int_0^\infty p y^{p-1} \mathrm{Prob}[Y > y] dy$.

Moments

Proof.

By Fubini's theorem for non-negative random variables,

$$\begin{aligned}\int_0^\infty py^{p-1}\,\mathrm{Prob}[Y>y]dy &= \int_0^\infty \int_\Omega py^{p-1}\mathbf{1}_{(Y>y)}dPdy\\ &= \int_\Omega \int_0^\infty py^{p-1}\mathbf{1}_{(Y>y)}dydP\\ &= \int_\Omega \int_0^Y py^{p-1}dydP = \int_\Omega Y^p = \mathrm{E}\left[Y^p\right].\end{aligned}$$

□

The weak law of large numbers

Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with

$$x \,\mathrm{Prob}(|X_i| > x) \to 0, \qquad x \to \infty.$$

Let $S_n = X_1 + \cdots + X_n$ and let $\mu_n = \mathrm{E}\left[X_1 \mathbf{1}_{(|X_1| \leq n)}\right]$. Then $\frac{S_n}{n} - \mu_n \to 0$ in probability.

The weak law of large numbers

Proof.

We apply the weak law for triangular arrays with $X_{n,k} = X_n$ and with $b_n = n$. There are two conditions to check. The first is satisfied, since

$$\sum_{k=1}^{n} \mathrm{Prob}(|X_{n,k}| > n) = n\,\mathrm{Prob}(|X_i| > n) \to 0.$$

To prove the second condition, it suffices to check that $\frac{1}{n^2}\sum_{k=1}^{n} \mathrm{E}\left[\overline{X}_{n,k}^2\right] = \frac{1}{n}\mathrm{E}\left[\overline{X}_{n,1}^2\right] \to 0$. This follows, since

$$\frac{1}{n}\mathrm{E}\left[\overline{X}_{n,1}^2\right] \leq \frac{1}{n}\int_0^n 2y\,\mathrm{Prob}[|X_1| > y]dy \to 0.$$

□

The weak law of large numbers

Theorem

Let $X_1, X_2, \ldots$ *be i.i.d. with* $\mathrm{E}[|X_i|] < \infty$. *Let* $S_n = X_1 + \cdots + X_n$, *and let* $\mu = \mathrm{E}[X_1]$. *Then* $\frac{S_n}{n} \to \mu$ *in probability.*

Proof.

The condition of the previous weak law is met, since $x \,\mathrm{Prob}(|X_i| > x) \leq \mathrm{E}\left[|X_i| \mathbf{1}_{(|X_i|>x)}\right] \to 0$ as $x \to \infty$. The theorem now follows, since $\mu_n \to \mu$ as $n \to \infty$, by dominated convergence. □

The Cauchy distribution

- The *Cauchy distribution* has density $\frac{1}{\pi(1+x^2)}$.
- If $X_1, ..., X_n$ are i.i.d. Cauchy, then $\frac{1}{n}\sum_{i=1}^n X_i$ is again Cauchy of the same distribution. This may be readily checked with characteristic functions, we postpone the proof.
- Thus the Cauchy distribution is a distribution for which a weak law does not hold.

The "St. Petersburg paradox"

Theorem

Let $X_1, X_2, \ldots$ be i.i.d., satisfying

$$\mathrm{Prob}[X_i = 2^j] = 2^{-j}, \qquad j \geq 1.$$

Let $S_n = X_1 + \cdots + X_n$. We have $\frac{S_n}{n \log_2 n} \to 1$ in probability as $n \to \infty$.

The "St. Petersburg paradox"

Proof.

- We apply the weak law for triangular arrays with b_n tending to ∞ faster than n but slower than $n \log n$.
- Since $\text{Prob}[X_1 \geq 2^m] = \sum_{j=m}^{\infty} 2^{-j} \leq 2^{-m+1}$, this condition guarantees that $n\,\text{Prob}[X_1 \geq b_n] \to 0$ as $n \to \infty$.
- To check the second condition, note that $\overline{X}_{n,k} = X_k \mathbf{1}_{(|X_k| \leq b_n)}$ satisfies

$$\mathrm{E}\left[\overline{X}_{n,k}^2\right] = \sum_{j: 2^j \leq b_n} 2^{2j} 2^{-j} \leq 2b_n.$$

In particular, $\frac{1}{b_n^2} \sum_{k=1}^{n} \mathrm{E}\left[\overline{X}_{n,k}^2\right] = O\left(\frac{n}{b_n}\right) \to 0$.

- We have $a_n = \mathrm{E}\left[\overline{X}_{n,k}\right] = \sum_{j: 2^j \leq b_n} 2^j 2^{-j} \sim \log_2 b_j \sim \log_2 n$.
- It follows that $\frac{S_n - n a_n}{b_n} \to 0$ and hence $\frac{S_n}{n \log_2 n} \to 1$ in probability.

□

The Borel-Cantelli Lemmas

Definition

Given A_n a sequence of subsets of Ω, let

$$\limsup A_n = \lim_{m\to\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega : \text{ in infinitely many } A_n\}$$

$$\liminf A_n = \lim_{m\to\infty} \bigcap_{n=m}^{\infty} A_n = \{\omega : \text{ in all but finitely many } A_n\}.$$

First Borel-Cantelli Lemma

Theorem (First Borel-Cantelli lemma)

If $\sum_{n=1}^{\infty} \text{Prob}[A_n] < \infty$ *then* $\text{Prob}\left[A_n \textit{ i.o.}\right] = 0$.

Proof.

Let $N = \sum_k \mathbf{1}_{A_k}$. Since $\text{E}[N] = \sum_k \text{Prob}[A_k] < \infty$, we have $N < \infty$ a.s.. □

Convergence in probability

Theorem

$X_n \to X$ in probability if and only if for every subsequence $X_{n(m)}$ there is a further subsequence $X_{n(m_k)}$ that converges almost surely to X.

Convergence in probability

Proof.

- For the forward direction, for each k there is an $n(m_k) > n(m_{k-1})$ so that $\mathrm{Prob}\left[|X_{n(m_k)} - X| > \frac{1}{k}\right] \leq 2^{-k}$. Since

$$\sum_{k=1}^{\infty} \mathrm{Prob}\left[\left|X_{n(m_k)} - X\right| > \frac{1}{k}\right] < \infty$$

  Thus only finitely many events occur a.s. so $X_{n(m_k)} \to X$ a.s..

- To prove the reverse direction, consider the sequence $y_n = \mathrm{Prob}(|X_n - X| > \delta)$. The conclusion follows from the observation that, in a topological space, if every subsequence of $\{y_n\}$ has a sub-subsequence converging to y, then $y_n \to y$.

□

Convergence of functions

Theorem

If f is continuous and $X_n \to X$ in probability then $f(X_n) \to f(X)$ in probability. If, in addition, f is bounded, then $\mathrm{E}[f(X_n)] \to \mathrm{E}[f(X)]$.

Convergence of functions

Proof.

- Let $X_{n(m)}$ be a subsequence, with sub-subsequence $X_{n(m_k)} \to X$ a.s.
- By continuity, $f(X_{n(m_k)}) \to f(X)$, a.s. which proves the convergence in probability.
- When f is bounded, $\mathrm{E}[f(X_{n(m_k)})] \to \mathrm{E}[f(X)]$, which suffices for the second claim.

□

Strong law of large numbers

Theorem

Let $X_1, X_2, \ldots$ be an i.i.d. sequence satisfying $\mathrm{E}[X_i] = \mu$ and $\mathrm{E}\left[X_i^4\right] < \infty$. If $S_n = X_1 + \cdots + X_n$ then $\frac{S_n}{n} \to \mu$ a.s.

Strong law of large numbers

Proof.

- We can assume $\mu = 0$ by making a translation.
- Expand

$$\mathrm{E}\left[S_n^4\right] = \mathrm{E}\left[\sum_{1 \le i,j,k,l \le n} X_i X_j X_k X_l\right].$$

- Since $\mathrm{E}[X_i] = 0$, the only terms which survive the expectation are of the form X_i^4 or $X_i^2 X_j^2$, $i \ne j$. Thus $\mathrm{E}\left[S_n^4\right] = O(n^2)$.
- It follows that $\mathrm{Prob}\left[|S_n| > n\epsilon\right] = O\left(\frac{1}{n^2 \epsilon^4}\right)$, so only finitely many of these events occur by Borel-Cantelli.

□

The second Borel-Cantelli lemma

Theorem

If events A_n are independent, the $\sum_{n=1}^{\infty} \text{Prob}[A_n] = \infty$ implies $\text{Prob}[A_n \textit{ i.o.}] = 1$.

The second Borel-Cantelli lemma

Proof.

Let $M < N < \infty$. By independence and the inequality $(1-x) \leq e^{-x}$,

$$\mathrm{Prob}\left(\bigcap_{n=M}^{N} A_n^c\right) = \prod_{n=M}^{N}(1-\mathrm{Prob}(A_n)) \leq \exp\left(-\sum_{n=M}^{N}\mathrm{Prob}(A_n)\right) \to 0$$

as $N \to \infty$. Thus $\mathrm{Prob}\left(\bigcup_{n=M}^{\infty} A_n\right) = 1$ for all M. Since $\bigcup_{n=M}^{\infty} A_n \downarrow \limsup A_n$ we obtain $\mathrm{Prob}(\limsup A_n) = 1$. $\square$

Failure of the strong law

Corollary

If $X_1, X_2, \ldots$ are i.i.d. with $\mathrm{E}[|X_i|] = \infty$, *then*
$\mathrm{Prob}\left[\lim_n \frac{S_n}{n} \text{ exists} \in (-\infty, \infty)\right] = 0$.

Proof.

We have

$$\mathrm{E}\left[|X_1|\right] = \int_0^\infty \mathrm{Prob}(|X_1| > x)dx \le \sum_{n=0}^\infty \mathrm{Prob}(|X_1| > n).$$

Thus, by independence and the second Borel-Cantelli lemma, the event $|X_n| > n$ occurs infinitely often with probability 1, which is sufficient to guarantee the non-convergence. □

Stronger Borel-Cantelli

Theorem

If $A_1, A_2, \ldots$ are pairwise independent and $\sum_{n=1}^{\infty} \mathrm{Prob}(A_n) = \infty$, then as $n \to \infty$

$$\sum_{m=1}^{n} \mathbf{1}_{A_m} \Bigg/ \sum_{m=1}^{n} \mathrm{Prob}(A_m) \to 1 \text{ a.s.}$$

Stronger Borel-Cantelli

Proof.

- Let $X_m = \mathbf{1}_{A_m}$ and $S_n = X_1 + ... + X_n$.
- By pairwise independence, $\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + ... + \mathrm{Var}(X_n)$. Since each X_n is an indicator variable, $\mathrm{Var}(S_n) \leq \mathrm{E}[S_n]$. Thus,

$$\mathrm{Prob}\left(|S_n - \mathrm{E}[S_n]| > \delta\, \mathrm{E}[S_n]\right) \leq \frac{1}{\delta^2\, \mathrm{E}[S_n]}.$$

- Let $n_k = \inf\{n : \mathrm{E}[S_n] > k^2\}$ and let $T_k = S_{n_k}$. By summability of $\sum_k \frac{1}{\mathrm{E}[T_k]}$ we find that $T_k / \mathrm{E}[T_k] \to 1$ a.s.
- To conclude the theorem in general, note that for $n_k < n < n_{k+1}$, use

$$\frac{T_k}{\mathrm{E}[T_{k+1}]} \leq \frac{S_n}{\mathrm{E}[S_n]} \leq \frac{T_{k+1}}{\mathrm{E}[T_k]}$$

and $\frac{\mathrm{E}[T_{k+1}]}{\mathrm{E}[T_k]} \to 1$.

□

Record values

Example (Record values)

- Let $X_1, X_2, \ldots$ be i.i.d. random variables having a continuous distribution.
- Let $A_k = \{X_k > \sup_{j<k} X_j\}$ be the event of a record at index k.
- Since the distributions are continuous, $X_i \neq X_j$ a.s.. The ordering of $X_1, X_2, \ldots, X_k$ induces the uniform measure on permutations in $\mathfrak{S}_k$, since for any permutation σ, $(X_1, \ldots, X_k)$ and $(X_{\sigma(1)}, \ldots, X_{\sigma(k)})$ are equal in distribution.
- Hence $A_1, A_2, \ldots$ are independent and $\mathrm{Prob}(A_k) = \frac{1}{k}$.
- By the strong law of large numbers, $R_n = \sum_{m=1}^n \mathbf{1}_{A_m}$ satisfies as $n \to \infty$

$$\frac{R_n}{\log n} \to 1, \qquad a.s.$$

Head runs

Example (Head runs)

- Let X_n, $n \in \mathbb{Z}$ be i.i.d. with $\mathrm{Prob}(X_n = 1) = \mathrm{Prob}(X_n = -1) = \frac{1}{2}$.
- Let $\ell_n = \max\{m : X_{n-m+1} = \cdots = X_n = 1\}$ be the length of the run of 1's at time n, and let $L_n = \max_{1 \leq m \leq n} \ell_m$. We show $\frac{L_n}{\log_2 n} \to 1$, a.s.
- Since $\mathrm{Prob}(\ell_n \geq (1+\epsilon)\log_2 n) \leq n^{-(1+\epsilon)}$ is summable, this event happens finitely often with probability 1, by Borel-Cantelli.

Head runs

Example (Head runs)

- To prove the lower bound, let $n = 2^k$ and split the block between $[n/2, n)$ into pieces of length $[(1-\epsilon)\log_2 n] + 1$.
- Each of these is entirely 1 with probability $\gg n^{-1+\epsilon}$, and the events are independent.
- There are $\gg \frac{n}{\log n}$ events in the block, so that, summed over the block their probabilities sum to $\gg n^{\epsilon/2}$.
- Summing in blocks, infinitely many of the events occur with probability 1, by Borel-Cantelli.

Strong law of large numbers

Theorem (Strong law of large numbers)

Let $X_1, X_2, \ldots$ *be pairwise independent identically distributed random variables with* $\mathrm{E}[|X_i|] < \infty$. *Let* $\mathrm{E}[X_i] = \mu$ *and* $S_n = X_1 + \cdots + X_n$. *Then* $\frac{S_n}{n} \to \mu$ *a.s.*

Strong law of large numbers

Lemma

Let $Y_k = X_k \mathbf{1}_{(|X_k| \leq k)}$ and $T_n = Y_1 + \cdots + Y_n$. It is sufficient to prove that $T_n/n \to \mu$ a.s.

Proof.

Observe $\sum_k \mathrm{Prob}(|X_k| > k) \leq \int_0^\infty \mathrm{Prob}(|X_1| > t)dt = \mathrm{E}[|X_1|] < \infty$. Thus $\mathrm{Prob}(Y_k \neq X_k \text{ i.o.}) = 0$. It follows that

$$\sup_n |T_n(\omega) - S_n(\omega)| < \infty, \; a.s.,$$

which suffices for the claim. □

Strong law of large numbers

Lemma

We have $\sum_k \frac{\mathrm{Var}(Y_k)}{k^2} < \infty$.

Proof.

Write

$$\sum_k \frac{\mathrm{E}\left[Y_k^2\right]}{k^2} \leq \sum_{k=1}^{\infty} k^{-2} \int_0^{\infty} \mathbf{1}_{(y<k)} 2y \,\mathrm{Prob}(|X_1| > y) dy$$
$$= \int_0^{\infty} \left\{ \sum_{k=1}^{\infty} k^{-2} \mathbf{1}_{(y<k)} \right\} 2y \,\mathrm{Prob}(|X_1| > y) dy$$
$$\ll \int_0^{\infty} \mathrm{Prob}(|X_1| > y) dy = \mathrm{E}[|X_1|] < \infty.$$

□

Proof of the strong law.

It suffices to prove the theorem for $X_n \geq 0$, since the general case may be separated into positive and negative parts.

- Let $\alpha > 1$ and set $k(n) = [\alpha^n]$. Recall $T_n = Y_1 + ... + Y_n$.
- By Chebyshev's inequality, for $\epsilon > 0$,

$$\sum_{n=1}^{\infty} \mathrm{Prob}\left(|T_{k(n)} - \mathrm{E}[T_{k(n)}]| > \epsilon k(n)\right) \leq \epsilon^{-2} \sum_{n=1}^{\infty} \frac{\mathrm{Var}(T_{k(n)})}{k(n)^2}$$

$$= \epsilon^{-2} \sum_{n=1}^{\infty} k(n)^{-2} \sum_{m=1}^{k(n)} \mathrm{Var}(Y_m) = \epsilon^{-2} \sum_{m=1}^{\infty} \mathrm{Var}(Y_m) \sum_{n: k(n) \geq m} k(n)^{-2}$$

$$\ll \epsilon^{-2} \sum_{m=1}^{\infty} \frac{\mathrm{Var}(Y_m)}{m^2} < \infty.$$

□

Strong law of large numbers

Proof of the strong law.

- It follows that $(T_{k(n)} - \mathrm{E}[T_{k(n)}])/k(n) \to 0$ a.s. Meanwhile, $\frac{\mathrm{E}[T_{k(n)}]}{k(n)} \to \mathrm{E}[X_1]$ by dominated convergence.
- For $k(n) \leq m < k(n+1)$

$$\frac{T_{k(n)}}{k(n+1)} \leq \frac{T_m}{m} \leq \frac{T_{k(n+1)}}{k(n)}.$$

- Since $\frac{k(n+1)}{k(n)} \to \alpha$, we have a.s.

$$\frac{1}{\alpha}\mathrm{E}[X_1] \leq \liminf_{n\to\infty} \frac{T_m}{m} \leq \limsup_{m\to\infty} \frac{T_m}{m} \leq \alpha\,\mathrm{E}[X_1].$$

- Since $\alpha > 1$ was arbitrary, the limit follows.

□

Strong law of large numbers

Theorem

Let $X_1, X_2, \ldots$ *be i.i.d. with* $\mathrm{E}[X_i] = \infty$ *and* $\mathrm{E}[|X_i^-|] < \infty$. *Let* $S_n = X_1 + \cdots + X_n$. *Then* $\frac{S_n}{n} \to \infty$ *a.s.*

Strong law of large numbers

Proof.

Let $M > 0$ and set $X_i^M = \min(X_i, M)$. By the strong law, $\frac{1}{n}\sum_{i=1}^n X_i^M \to \mathrm{E}[X_i^M]$ a.s. as $n \to \infty$. Hence

$$\liminf_{n\to\infty} \frac{S_n}{n} \geq \mathrm{E}\left[X_i^M\right].$$

Since $\mathrm{E}\left[X_i^M\right] \to \infty$ as $M \to \infty$, the claim follows. □

Renewal theory

Let $X_1, X_2, \ldots$ be i.i.d. with $0 < X_i < \infty$. Let $T_n = X_1 + \cdots + X_n$ and

$$N_t = \sup\{n : T_n \leq t\}.$$

Given a sequence of events which happen in succession with waiting time X_n to the nth event, we think of N_t as the number of events which have happened up to time t.

Theorem

If $\mathrm{E}[X_1] = \mu \leq \infty$, *then as* $t \to \infty$,

$$\frac{N_t}{t} \to \frac{1}{\mu} \text{ a.s.}.$$

Proof.

Since $T(N_t) \leq t < T(N_t + 1)$, dividing through by N_t gives

$$\frac{T(N_t)}{N_t} \leq \frac{t}{N_t} \leq \frac{T(N_t + 1)}{N_t + 1} \frac{N_t + 1}{N_t}.$$

We have $N_t \to \infty$ a.s.. Hence, by the strong law,

$$\frac{T_{N_t}}{N_t} \to \mu, \qquad \frac{N_t + 1}{N_t} \to 1.$$

□

Empirical distribution functions

Let $X_1, X_2, \ldots$ be i.i.d. with distribution F and let

$$F_n(x) = \frac{1}{n} \sum_{m=1}^{n} \mathbf{1}_{(X_m \le x)}.$$

Theorem (Glivenko-Cantelli Theorem)

As $n \to \infty$,

$$\sup_x |F_n(x) - F(x)| \to 0 \ a.s..$$

Empirical distribution functions

Proof.

Note that F is increasing, but can have jumps.

- For $k = 1, 2, \ldots$, and $1 \leq j \leq k-1$, define $x_{j,k} = \inf\{x : F(x) \geq \frac{j}{k}\}$. Set $x_{0,k} = -\infty$, $x_{k,k} = \infty$.
- Write $F(x-) = \lim_{y \uparrow x} F(y)$.
- Since each of $F_n(x_{j,k}-)$ and $F_n(x_{j,k})$ converges by the strong law, and $F_n(x_{j,k}-) - F_n(x_{j-1,k}) \leq \frac{1}{k}$, the uniform convergence follows.

□

Entropy

- Let $X_1, X_2, \ldots$ be i.i.d., taking values in $\{1, 2, \ldots, r\}$ with all possibilities of positive probability. Set $\mathrm{Prob}(X_i = k) = p(k) > 0$.
- Let $\pi_n(\omega) = p(X_1(\omega))p(X_2(\omega))\ldots p(X_n(\omega))$. By the strong law, a.s.

$$-\frac{1}{n}\log \pi_n \to H \equiv -\sum_{k=1}^{r} p(k)\log p(k).$$

The constant H is called the *entropy*.