Scratchwork: The Law of Large Numbers

Let's collect as many proofs as we can of the **Law of Large Numbers**.

Thm Suppose that $X_1, X_2, \ldots$ are independent random variables, and that for some constant K in $[0, \infty)$,
$$\mathbb{E}[X_k] = 0 \text{ and } \mathbb{E}[X_k^4] \leq K$$
The odds that the average of these numbers tends to zero is 1:
$$\mathbf{P}\left[\frac{1}{n}\sum_{i=0}^{n} X_i \to 0\right] = 1$$

Clarification, the event is $\{\omega : \frac{1}{n}\sum X_i(\omega) \to 0\}$, if we constructed the probability space very carefully. We could look at rare events such as, despite the average usually tending to zero, we could considtion that the empirircal average $S_n \to \frac{1}{2}$ or something small but definitively separate from zero, $S_n \to \frac{1}{100}$.

More clarification. What if the X_i are not independent, maybe they are only "loosely correlated" so that $\mathbb{E}[X_i X_j] \ll 1$. Or maybe these are pseudorandom, X_n is an orbit of a group action in number theory.

Proof Using the Binomial Formula, and linearity of expectation :
$$\mathbb{E}[S_n^4] = \mathbb{E}[(X_1 + \cdots + X_n)^4] = \mathbb{E}[\sum_k X_k^4 + 6\sum\sum_{i<j} X_i^2 X_j^2]$$

The mixed terms, go away. Here $\binom{4}{2,2} = \frac{4}{2!2!} = 6$ In fact, teh next line is:
$$\mathbb{E}[X_i X_j^3] = \mathbb{E}[X_i X_j^2 X_k] = \mathbb{E}[X_i X_j X_k X_l] = 0$$
Here we use "independence" and "centering" that $\mathbb{E}[X_i] = 0$.

In number theory problems we choose a specific sequence to study, usually elements of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$. However, there is usually also a group element such as a quadratic form Q or prime numbers $p \in \mathbb{Z}$.

- $[\mathbb{E}[(X_i)^2] \leq \mathbb{E}[X_i^4] \leq K$ by $\ldots$ Cauchy-Schwarz inequality ?
- $\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] \leq K$

We've had to count the number of terms in our sequence:
$$\mathbb{E}[S_n^4] \leq nK + 3n(n-1)K \leq 3Kn^2$$

The next step involved the **monotone convergence theorem** we'd like to actually see some of the weirdness that could be implied here.[1]

$$\left[\mathbb{E}\sum (S_n/n)^4 \overset{?}{\leq} 3K\sum n^{-2} < \infty\right] \to \left[\mathbf{P}\left[\sum (S_n/n)^4 < \infty\right] = 1\right] \to \left[\mathbf{P}[S_n/n \to 0] = 1\right]$$

---

[1] Monotone convegence theorem is used show convergence of some integers, $\lim \int f_n = \int \lim f_n$ which is *false* for Riemann integrals. What does it mean to have a result this extreme and never use it? This seems wrong.

Lots of normal grown-up human decision making involves us deciding that certain events are rare or will never occur. Is that the only philosophy here?

What does this mean that are sum is finite almost surely. What are these probability zero points? Remember that Number Theory studies functions over $\mathbb{Q}$ and $\mu(\mathbb{Q}) = \mathbb{R}$ as a measure zero subset of the real numbers. These exceptions could be everywhere.

In the process of our proof of the Law of Large numbers, we computed an $\mathcal{L}^4$ norm. We could talk about random or pseudo-random sequences of numbers.[2] Without any specific example in mind, this proof is totally correct without any further questions. The Law of Large numbers is a theoretical way of checking that loosely correlated observations about the world will tend to the average.[3]

Tutorial Monotone Convergence Theorem

Tutorial $\mathcal{L}^2$ spaces

Our use of the Cauchy-Schwartz inequality is not free. Covariance and variance, are tightly related to Pythagoras' theorem and projection operators on Hilbert spaces.

The only Hilbert space I have any chance at understanding is $L^2([0,1], dx)$ spanned by the $\sin$ and $\cos$ function. There is also $L^2(S^1, d\theta)$ spanned by $[\theta \mapsto e^{in\theta}]$. Here they study the geometry of the Hilbert space $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$. Our probability space (and . . . our Hilbert space) depends on:

- the space of outcomes Ω

- $\mathbf{P}$ the probability measure

- also depends on the collection of "measurable" events $\mathcal{F}$

This jargon could make sense if we saw a random process – they're everywere – the randomness seperates that is constantly separating us from a "perfect" description of the world. Things that we're uncertain about.

(MON) if $0 \le X_n \uparrow X$, then $\mathbf{E}(X_n) \uparrow \mathbf{E}(X) \le \infty$.

(MON) if f_n is a sequence of elements of $(m\Sigma)^+$ such that $f_n \uparrow f$ then $\mu(f_n) \uparrow \mu(f) \le \infty$, or $\int f_n d\mu \to \int f d\mu$.

These are limits of unions and intersections of measurable sets. The integrals that we're thinking of could go seriously go astray when taking limits. As a consolation, these examples are ubiquitous in physics, engineering, biology, economics and social sciences. Think about where the measurable sets are in Sociology and what could be doing there?

Example: level sets are measurable. f is "integrable" (e.g. continuous) and $f^{-1}[a,b]$ is measurable. We need to able to say that $\{\omega : 0 < n^{-1} \sum X_i < \epsilon\}$ is measurable. . .

Thm If X and Y are in $\mathcal{L}^2$ (square-integrable) then $XY \in \mathcal{L}^1$ (integrable) and the CS inequality is obviousy true: $|\mathbf{E}(XY)| \le \mathbf{E}(|XY|) \le ||X||_2 \, ||Y||_2$.

Thm If X and Y are in $\mathcal{L}^2$ then so is $X + Y$ and we have the triangle inequality, $||X + Y||_2 \le ||X||_2 + ||Y||_2$.

---

[2]Pseudorandom numbers usually involve a generator or "seed" with lots of information either stored in the initial data itself or in the transition operator. Group orbits can be quite complicated and this proves a great place to have shape and complexity.

[3]If they have more correlation (such as the weather) we now have **Ergodic theory**. The mixing and diffusion of air, and the relation between temperature and pressure, are great examples from Chemistry. Elecrical Engineering could also have examples of random processes.

Triangle inequalit not-so-innocent here. And the Law of Large numbers requires an $\mathcal{L}^4$ norm, where the geometry is even less clear.

There seems to be a lot of uncertainty as to whether $\int X\,d\mu$ or $\mathbf{P}[X]$ is even a number in the first place. The answer could depend on how we approxmate X and μ. Arrangements like these could appear in Combinatorics or Number Theory (e.g. pigeon-hole principle). So this entire discussion is missing an input, just an interesting problem that we care about.

We need Monotone Convergence Theorem to prove the CS inequality.

- We could only consider $|X|$ and $|Y|$ and focus on $X \geq 0$ and $Y \geq 0$.

- Define "cutoff" random variables $X_n = X \wedge n$ and $Y_n = Y \wedge n$ ("bounded").

$$0 \leq \mathbb{E}[(aX_n + bY_n)^2] = a^2\mathbb{E}[X_n^2] + 2ab\mathbb{E}[X_nY_n] + b^2\mathbb{E}[Y_n^2]$$

This is a quadratic form in a and b (over the real numbers, which are themselves equivalences of classes of sequences of numbers or observations):

$$\left[2\mathbb{E}(X_nY_n)\right]^2 \leq 4\mathbb{E}[X_n^2]\mathbb{E}[Y_n^2] \leq 4\mathbb{E}[X^2]\mathbb{E}[Y^2]$$

In court, how do we compare the "lightness" of two measurments? How do we measure the lightness or heaviness of quantitites or events or distributions?

Let $n \to \infty$, and our proof is complete by the Monotone Convergence Theorem.

References

[1] $\cdots$