Egorov's Theorem is about Speed

Hello!

This blogpost is about Egorov's theorem, a standard result of any typical undergraduate measure theory class. It's an interesting theorem, but I feel that it often goes overlooked by students, partly because it seems like such a technical result, partly because it gets buried in a pile of results about all the different ways sequences of measurable functions can converge and how they relate to each other.

What I want to do is show that there's a very neat way to think about it, that makes its proof very natural. Ideally, after you read this article, you'll feel like you could have come up with it.

Pointwise and uniform convergence

It's a well-known fact that pointwise convergence of a sequence of functions does not, in general, imply uniform convergence.

For instance, the sequence of functions $\{f_n\}$ defined by

$$f_n = \mathbb{1}_{[n,+\infty)}$$

converges pointwise but not uniformly to 0.

The issue, broadly speaking, is that although for every point x, $f_n(x)$ converges to $f(x)$, the *speed* at which this convergence happens is arbitrary, i.e. the convergence can happen arbitrarily slowly. This is a problem because the sequence $\{\sup_x |f_n(x) - f(x)|\}$ converges, by definition, slower than any individual $\{f_n(x)\}$.

To make an analogy that is more "mathematical", it's like having a set $A \subset \mathbb{R}$ that only contains positive numbers, but contains arbitrarily small ones, so that $\inf A = 0$ is not positive.[1]

Despite not being rigorous, thinking about convergence in this way helps us identify under which conditions pointwise convergence implies uniform convergence: we need a way to give a "positive" lower bound to the speed of convergence of $\{f_n(x)\}$. A trivial example of this is the following result:

Proposition 1: If X is a finite set, and $f_n : X \to \mathbb{R}$ is a sequence of functions that converges pointwise to $f : X \to \mathbb{R}$, then the convergence is uniform.

[1] To clarify, in this analogy, the positive numbers represent the "speeds" of the individual $\{f_n(x)\}$, and we imagine that a speed of zero means that the sequence doesn't actually converge.

Obviously, if there are only finitely many sequences $f_n(x)$, the worst of the speeds is the lower bound that we're looking for.

A more sophisticated result of this kind is Dini's theorem, where instead compactness is used to provide a positive lower bound.

Egorov's Theorem

Egorov's theorem is another example of a result that tries to leverage pointwise convergence into uniform convergence, and in this section we'll see in detail how we can reframe it in terms of "speed", and how this reframing makes the usual proof of the theorem very natural.

The statement is as follows:

Theorem 2: Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f_n, n \geq 1$ and f be measurable[2] functions $X \to \mathbb{R}$. Assume that $\{f_n\}$ converges μ-almost everywhere to f, and that μ is finite. Then, for any $\varepsilon > 0$, there exists a measurable set A with $\mu(X \setminus A) < \varepsilon$ and

$$\sup_{x \in A} |f_n(x) - f(x)| \underset{n \to \infty}{\longrightarrow} 0.$$

In other words, $\{f_n\}$ converges uniformly to f on A.

Here, we're talking about almost-everywhere convergence instead of pointwise convergence, but this doesn't matter for the intuition. The trick is that Egorov's theorem allows us a little bit of leeway: we only care about what happens on A.

So now we're in the following position: we have a collection of sequences that (almost-)all converge (the sequences $\{f_n(x)\}$, for any x), but at arbitrary speeds. We want to find a positive lower bound for the speeds of convergence, and we're allowed to remove from our consideration a part of the sequences.

I think it's useful to pause here and think about how this reframing of the theorem helps us. Because it makes the following idea very obvious: we need to put aside all the values of x such that $f_n(x)$ converges "too slowly", and consider the rest.

At this point the fact that the theorem is true should hopefully be clear. Now we need to prove it.

Proof: Let $\varepsilon > 0$.

[2] By measurable, we mean $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$-measurable

We're looking for a set A that contains all the values of x such that $\{f_n(x)\}$ converges "fast enough". What "fast enough" means depends on ε. The smaller the value of ε, the more values of x we need in our set, so the less strict we'll be on the minimum required speed.

Unfortunately, we can't represent the speed of convergence of a sequence with a single number, we need something more complicated.

Take a nondecreasing sequence $\{N_k\}_{k\geq 1}$ of integers, and consider the set A_k defined by

$$\begin{aligned} A_k &:= \left\{ x \in X, \forall n \geq N_k, |f_n(x) - f(x)| \leq \frac{1}{k} \right\} \\ &= \left\{ x \in X, \sup_{n \geq N_k} |f_n(x) - f(x)| \leq \frac{1}{k} \right\}. \end{aligned}$$

and let

$$A := \bigcap_{k \geq 1} A_k.$$

The way I think of this definition is that all of our sequences $\{f_n(x)\}$ are participating in a long-distance race where the goal is to converge to your limit, and we're placing little checkpoints along the way to monitor their progress. Then, we can think of N_k as some sort of "maximum time" to get to the kth checkpoint, otherwise you're eliminated for being too slow. When we say "$x \in A_k$", we're saying "by the time N_k, the sequence $\{f_n(x)\}$ has passed the k-th checkpoint (i.e. is within $\frac{1}{k}$ of its limit.)" And if $x \in A$, it means that $\{f_n(x)\}$ got to every checkpoint in time.

In other words, the sequence $\{N_k\}$ is the mathematical object that we use to represent a speed, and A is the set of x values such that the sequences $\{f_n(x)\}$ are "faster than $\{N_k\}$".

From this definition, what is relatively easy to show is that $\{f_n\}$ converges uniformly to f on A. Indeed, as n grows to $+\infty$, it passes every N_k, and if $n \geq N_k$, then

$$\sup_{x \in A} |f_n(x) - f(x)| \leq \frac{1}{k}.$$

Now, we're going to prove that for the right choice of $\{N_k\}_k$, this set A is the one we're looking for. Essentially, we need to find $\{N_k\}$ that grows slowly enough that $X \setminus A$, the set we eliminate for being too slow, has measure at most ε.

We'll use the following bound on $\mu(X \setminus A)$:

$$\mu(X \setminus A) = \mu\left(\bigcup_{k\geq 1}\left\{x \in X : \sup_{n\geq N_k} |f_n(x) - f(x)| > \frac{1}{k}\right\}\right)$$
$$\leq \sum_{k=1}^{\infty} \mu\left(\left\{x \in X : \sup_{n\geq N_k} |f_n(x) - f(x)| > \frac{1}{k}\right\}\right).$$

The trick here is that we can consider

$$g_N(x) := \sup_{n\geq N} |f_n(x) - f(x)|$$

as a single function. Given that $\{f_n\}$ converges μ-almost-everywhere to f, g_N is finite and converges to 0 μ-almost everywhere. Furthermore, for any x, the sequence $\{g_N(x)\}$ is nonincreasing, which means that for any fixed k, the sets

$$B_N = \left\{x : g_N(x) > \frac{1}{k}\right\}$$

form a decreasing family of sets.

We're going to apply the continuity from above of μ. We can always do that because μ is finite. We have

$$\lim_{N\to+\infty} \mu(B_N) = \mu\left(\bigcap_{N\geq 1} B_N\right) = 0.$$

The equality on the right above stems from the fact that $\{g_N\}$ converges μ-almost everywhere to 0, and if x is in every B_N, then $g_N(x)$ never goes below $\frac{1}{k}$, and hence does not converge to 0.

Intuitively, since we've fixed k, this fact says that if we stay at a particular checkpoint, the set of x's that we haven't seen cross in front of us gets smaller and smaller.[3]

The consequence of this is that for any k, we can find N_k such that

$$\mu\left(B_{N_k}\right) < \frac{\varepsilon}{2^k}.$$

With this,

[3] in the sense that its measure goes to 0.

$$\begin{aligned}\mu(X \setminus A) &= \mu\left(\bigcup_{k\geq 1} B_{N_k}\right)\\ &\leq \sum_{k\geq 1} \mu\Big(B_{N_k}\Big)\\ &< \sum_{k\geq 1} \frac{\varepsilon}{2^k}\\ &< \varepsilon.\end{aligned}$$

This concludes the proof.

Going further

There are many extensions of Egorov's theorem, and I invite you to look at its Wikipedia page to learn more. What I'm interested in is seeing where the intuition leads us. Because right now, to me, the fact that we required the measure μ to be finite seems more like a technical constraint than a fundamental one. By this, I mean that even if we consider a measure space $(X, \mathcal{A}, \mu)$ where μ is not necessarily finite, we can still define this notion of speed, and consider sequences that grow slower and slower. By doing so, we should get larger and larger subsets of X that eventually cover μ-almost all of X.

In essence, it seems the following statement is true:

Theorem 3: Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f_n, n \geq 1$ and f be measurable functions $X \to \mathbb{R}$. Assume that $\{f_n\}$ converges μ-almost everywhere to f. Then, there exists measurable sets H and $\{B_n, n \geq 1\}$ such that $\mu(H) = 0$, $B_n \subset B_{n+1}$ for all $n \geq 1$,

$$X = H \cup \bigcup_n B_n$$

and $\{f_n\}$ converges uniformly to f on each set B_n.

To be a little clearer on what the notation means, we'd expect H to be the set where $\{f_n(x)\}$ doesn't actually converge to $f(x)$, and each B_n to be the set of x values such that $\{f_n(x)\}$ converges to $f(x)$ faster than some "positive" speed $\{N_k\}$, such that the speeds $\{N_k\}$ "go to zero"[4] as the B_n grow.

[4]Note that there is one sequence $\{N_k\}$ per set B_k. And to clarify, we're not expecting each individual sequence $\{N_k\}$ to go to 0, as they are nondecreasing sequences of integers. But since each $\{N_k\}$ represents a speed, we expect this represented speed to "decrease" towards a value that represents a sequence that does not converge.

It is not hard to see that this result is true when the measure μ is finite. In that case, we just have to pick B_n to be the set we get when we apply Egorov's theorem with $\varepsilon = \frac{1}{n}$.[5]

It happens that this statement is also true when we only require $(X, \mathcal{A}, \mu)$ to be σ-finite. The proof of this fact is relatively easy, and is available on the Wikipedia page I linked earlier.

Unfortunately, and dare I say counter-intuitively, the statement of Theorem 3 is not true in general.

This means we need to dig a little deeper. The crux of the issue is something I alluded to above. I keep talking about this "speed" like it's a number, but it's not. In some ways, it does behave like a number – it's helped us before – but remember that rigorously, we defined it to be a sequence. And while it's easy to have a sequence of numbers that goes to 0, sequences of sequences are more complicated.

If this makes you think of diagonal arguments, then you're right! This is how we are going to construct a counterexample to Theorem 3.

More specifically, we need to find a measure space $(X, \mathcal{A}, \mu)$ and a sequence of functions $f_n : X \to \mathbb{R}$ for $n \geq 1$ that converges μ-almost everywhere to 0, but the result of Theorem 3 does not hold.

To make this proof a little easier to digest, I'm going to introduce some notation. Let S be our set of "speeds":

$$S := \{\{N_k\} \in \mathbb{N}^{\mathbb{N}} : \forall k, N_{k+1} \geq N_k\}.$$

We are going to give a partial order to S, such that $\{N_k\} \leq \{M_k\}$ means $\{N_k\}$ represents a "slower speed" than $\{M_k\}$. Now remember that since we defined our speeds in terms of "time it takes to reach checkpoints", the definition of $\leq$ on S is going to be backwards:

$$\{N_k\} \leq \{M_k\} \quad \text{iff} \quad \forall k, N_k \geq M_k.$$

$\{N_k\}$ is slower than $\{M_k\}$ iff, for any k the time it takes $\{N_k\}$ to reach the kth checkpoint is larger than that of $\{M_k\}$.

We're going to use S as our space X, and we're going to define our functions $f_n : S \to \mathbb{R}$ in the following way: for each speed $s \in S$, we'll define the sequence $\{f_n(s)\}$ such that its speed is s.

[5] Actually, to guarantee that the B_n form a nondecreasing sequence of sets, we take $B_n = C_1 \cup \cdots \cup C_n$ where C_n is the set we get from Egorov's theorem with $\varepsilon = \frac{1}{n}$.

To guarantee that the functions f_n defined in this way are measurable, we just take $\mathcal{A} = \mathcal{P}(S)$ as our σ-algebra. For our measure, we take the cardinal measure: for $A \subset S$, $\mu(A)$ is the number of elements of A if A is finite, and $+\infty$ if A is infinite.[6]

I'm also going to properly define this "speed" function that I've been talking about now: let V be the set of sequences of real numbers that converge to 0. Define the function $\beta : V \to S$ by setting

$$\beta(\{u_n\})_k = \inf\left\{N \geq 1 : \forall n \geq N, |u_n| \leq \frac{1}{k}\right\}.$$

Consider the two following lemmas:

Lemma 4: For any element $\{N_k\} \in S$, there exists a sequence $\{u_n\}$ in V such that $\beta(\{u_n\}) = \{N_k\}$.

Lemma 5: Let $U \subset V$. If we define $\{v_n\}$ by

$$v_n = \sup_{\{u_n\} \in U} |u_n|,$$

then $\{v_n\}$ converges to 0 if and only if there exists $\{N_k\} \in S$ such that for any $\{u_n\} \in U$,

$$\{N_k\} \leq \beta(\{u_n\}).$$

In the interest of streamlining this article, I'm not going to spend time on proving either of these lemmas. The proofs are relatively simple, and you are welcome to attempt them on your own.

Intuitively, the first lemma says that there exist sequences of every speed, and the second is a restatement of the fact that for a set of sequences to converge uniformly (i.e. at the same time), their speeds need to be bounded below.

We're now ready to define the functions f_n: for any $s \in S$, use Lemma 4 to pick the sequence $\{f_n(s)\} \in V$ such that $\beta(\{f_n(s)\}) = s$. By definition, $\{f_n\}$ converges pointwise to 0.

Assume that there exist $B_n, n \geq 1$ as in the statement of Theorem 3[7]. Then, applying Lemma 5 to each B_n, we find $s_n \in S$ a speed such that for any $s \in B_n$,

$$s_n \leq \beta(\{f_n(s)\}) = s.$$

Perhaps now you see where this is going: define $\{N_k\} \in S$ such that

[6]Note that with this measure, the only set of measure 0 is the empty set, and converging μ-almost everywhere is the same thing as converging pointwise.

[7]Since the only set of measure 0 is the empty set, we can omit H.

$$N_k = (s_n)_k + 1.$$

If s_n is the lower bound for the speeds of elements in B_n, we've defined an element of S that is slower than every s_n. This means that $\{N_k\}$ belongs to none of the sets B_n, for any n, which is a contradiction, as we've assumed that

$$S = \bigcup_n B_n.$$

Conclusion

I like Egorov's theorem because it highlights what makes measure theory so useful: you have a way to provide results that are more nuanced than the black-and-white ones you see in real analysis. It makes concrete things like "this theorem is almost true but for these few pesky counterexamples…". In a similar vein, Lusin's theorem states that measurable functions are "almost" continuous. Perhaps unsurprisingly, its proof is a straightforward application of Egorov's.

I was a little disappointed that I wasn't able to find a proof of it that was meaningfully different from the one I presented here, so please send me an email if you know of one.

I hope this was an enjoyable read, and that I was able to effectively communicate how I think of the notions I covered here.

Until next time.