Introduction to analytic proofs for students going into likelihood theory

Patrick Breheny

September 1, 2022

The purpose of this document is provide an introduction (or refresher) to the kinds of proofs that one sees in a course on real analysis, as these sorts of proofs come up a lot in theoretical statistics. If you've never had a course in real analysis, or it's been a while and you've forgotten, I think this document will be useful if you're thinking about taking Likelihood Theory.

First, a note on style. Most textbooks and papers provide proofs in an unstructured paragraph style. For the purposes of learning how to prove things, however, I recommend a more structured approach based on making consecutive explicit statements with explicit justifications. There are several reasons for this:

1. When you finish a structured proof, it is very clear exactly which conditions were required, why they were required, and what supporting theorems or results were used. When you're learning in a course, this is extremely valuable as it will be more clear to you how everything is connected.
2. The other major reason is that it's much harder to make a mistake in a structured proof. This doesn't make the proof easier, it just means that in an unstructured proof, one can easily skip steps without realizing it. We'll see some examples of this later.
3. It's also very beneficial from a grading and feedback perspective, as it makes it much clearer to the person grading the proof whether you understand all the pieces or not.

Definitions, "there exists", and "for all"

Many proofs in theoretical statistics involve convergence; loosely speaking, as we collect more and more data, we can be increasingly sure that certain things will happen. We'll discuss probabilistic convergence in the course, but before that, it will help to be familiar with ordinary, deterministic convergence.

Definition. *A sequence $x_1, x_2, \ldots$ of real numbers is said to* converge *if there exists a real number x such that for all $\epsilon > 0$, there is an integer N such that $|x_n - x| < \epsilon$ for all $n > N$. In this case, we say that x_n converges to x, or that x is the limit of x_n, and write $x_n \to x$ or $\lim x_n = x$ (these mean the same thing). If no such number exists, the sequence is said to diverge.*

To make this definition concrete, let's consider a specific sequence: $e^{-1}, e^{-2}, \ldots$. Intuitively, these numbers are getting closer to zero, so it seems right to say the sequence converges to zero. But does it satisfy the definition? Notice that there are two specific clauses that come up in the above definition, and they occur constantly in mathematical proofs and theorems:

- "there exists": There is at least one number satisfying this condition. Denoted $\exists$.
- "for all": This has to be true in every possible case, no exceptions. Denoted $\forall$.

So, when we say that "there exists a real number x", we're only requiring these conditions to be satisfied for the limit point; in the case of our example $x_n = e^{-n}$, we only need to show that it happens at $x = 0$. Maybe this condition is met at other points, maybe it isn't.

On the other hand, when we say "for all $\epsilon > 0$", the rest of the condition has to true for *every single positive real number*. Maybe we can find ϵ values for which the condition is met, but it if doesn't hold for all of them, not good enough.

For example, suppose we consider the point $x = 0.001$: does x_n converge to 0.001? Well, if $\epsilon = 0.002$, then yes, we can find a number N such that $|x_n - x| < \epsilon$ for all $n > N$. For example, $N = 6$ works:

$$\begin{aligned} \left|e^{-6} - 0.001\right| &= 0.0015 < 0.002 \\ \left|e^{-7} - 0.001\right| &= 0.000088 < 0.002 \\ &\ldots \end{aligned}$$

However, this isn't enough – we need to be able to find a suitable N for *all* positive ϵ. If $\epsilon = 0.0005$, for example, we can never find an N that will work.

One final important point to note here is that the order of these clauses matters. For every ϵ, there exists a suitable N. We choose ϵ first, and can then find an N that works for that ϵ – we don't need to find an N that works for all ϵ (and in fact, there typically isn't such a N). If $\epsilon = 0.001$, then $N = 7$ works. If $\epsilon = 0.0000000000000000000000001$, we are going to have to choose a much larger N (but we can *always* find such an N).

Finding N is fairly simple in this case; we can simply take the log of both sides:

$$\begin{aligned} e^{-n} &< \epsilon \\ -n &< \log \epsilon \\ n &> -\log \epsilon. \end{aligned}$$

So, $N = \lceil -\log \epsilon \rceil$, where $\lceil \cdot \rceil$ denotes the ceiling function (i.e., round up the number inside to the next integer). To make things even more concrete, one could think of $N(\epsilon)$ as a function:

```
n <- function(eps) {ceiling(-log(eps))}
n(0.001)
## 7
n(0.0000001)
## 17
n(0.00000000000000001)
## 40
```

Structured proofs

That was a lot of preliminary stuff, but it's very important to understand definitions before you move on to proofs. Now, let's prove that $e^{-n} \to 0$. What, specifically, do we need to prove? That for all $\epsilon > 0$, there is an integer N such that $|e^{-n}| < \epsilon$ for all $n > N$. So, we need to start with "Let $\epsilon > 0$" (this is how **a lot** of proofs start):

Proof. Let $\epsilon > 0$, and let $N = \lceil -\log \epsilon \rceil$. Then for all $n > N$,

$$\begin{aligned}
\left|e^{-n}\right| &= e^{-n} && e^{-n} \text{ is always positive} \\
&< e^{-N} && e^{-n} \text{ is strictly decreasing} \\
&= \exp\{-\lceil -\log \epsilon \rceil\} && \text{Definition of } N \\
&\le e^{\log \epsilon} && e^{-n} \text{ decreasing} \\
&= \epsilon && \square
\end{aligned}$$

The left column is a series of statements or claims; this is the main logic of what's happening in the proof. Chaining all the lines together, we have $|e^{-n}| < \epsilon$. Thus, we've done what we needed to do: given any positive ϵ, I can find always find an N that meets the requirement. Thus, $e^{-n} \to 0$ by the definition of convergence.

The right column provides the justification for each step. For example, in the fourth line we claimed that $\exp\{-\lceil -\log \epsilon \rceil\} \le e^{\log \epsilon}$. How do we know that this is true? Well, $\lceil x \rceil \ge x$ since the operation involves rounding up, and e^{-x} is a decreasing function of x. So by replacing the argument to e^{-x} with something smaller, the result must be larger (or equal, since $-\log \epsilon$ could be an integer already).

How far to go with these justification depends on what the reader of the proof would likely consider obvious, and is therefore a judgment call. For example, we could include a proof of the fact that e^{-x} is a decreasing function of x, but in my judgment this is going off on a bit of a tangent that distracts from the main point of the proof. In a similar fashion, I didn't even provide a justification for the final step, but to be really thorough, I could have added that exp() and log() are inverse functions. In general, for the purposes of a course, you should err on the side of being very thorough and justifying every step. In a paper or thesis, however, going into this level of detail for every proof is probably unnecessary as the audience (other people with PhDs in statistics or biostatistics) will probably find many of the steps obvious and not requiring justification.

One final comment: as the above proof should hopefully make clear, a final proof does not describe one's thought process in terms of how you arrived at the result. Clearly I didn't know what I should set N equal to until I'd worked through the math to solve for n in the previous section. If you're reading the proof for the first time, the line "let $N = \lceil -\log \epsilon \rceil$" is going to seem mysterious; where did *that* come from? Keep in mind that we're writing a proof, not a novel of self-discovery. The point is to construct an iron-clad, rigorous argument, not to communicate our thoughts and feelings and realizations. It might not be immediately obvious *why* I am letting $N = \lceil -\log \epsilon \rceil$ ", but it should be completely obvious that I *can* set $N = \lceil -\log \epsilon \rceil$". One can certainly write about the thought process and intuition behind the proof, but this should be done outside the formal proof – we don't want to mix formal logic with informal intuition.

Convergence

Let's go through a few more proofs. For example, you might find the "there exists a real number x" clause in the definition of convergence unsatisfying, as it leaves open the possibility that a many such numbers exist and a sequence might converge to lots of different things. However, this is in fact not possible.

Theorem. *If $x_n \to a$ and $x_n \to b$, then a=b.*

Proof. Let $\epsilon > 0$.

①	$\exists N_a : n > N_a \implies \lvert x_n - a \rvert < \frac{\epsilon}{2}$	$x_n \to a$
②	$\exists N_b : n > N_b \implies \lvert x_n - b \rvert < \frac{\epsilon}{2}$	$x_n \to b$

Thus, for all $n > N = \max(N_a, N_b)$, we have

$$\begin{aligned} |a-b| &= |a - x_n + x_n - b| \\ &\leq |x_n - a| + |x_n - b| && \text{Triangle inequality} \\ &< \epsilon \end{aligned}$$

$\therefore a = b$ ∎

In the above proof, the symbol $\implies$ means "implies" and the symbol $\therefore$ means "therefore" (writing out the words is fine too, I'm just introducing the symbols because I occasionally use them in class). This proof introduces a standard technique that comes up often: since $x_n \to a$, I can get x_n as close to a as I want. I could find an n such that $|x_n - a| < \epsilon$, but why stop there? I can keep going and get $|x_n - a| < \epsilon/2$, which ends up making the rest of the proof more clear. I could keep going even further and get $|x_n - a| < \epsilon/1000000$ if I wanted, but this isn't necessary for what I'm trying to show.

Summarizing the above proof, if $x_n \to a$ and $x_n \to b$, then for any positive number ϵ, it must be the case that $|a - b| < \epsilon$. In other words, they have to be equal $(a - b = 0)$. Here, the triangle inequality states that

$$|x + y| \leq |x| + |y| .$$

Applied to line three, this gives

$$\begin{gathered} |a - x_n + x_n - b| \leq |a - x_n| + |x_n - b| \\ |a - b| < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2}. \end{gathered}$$

Hopefully the logic here is clear and all the steps of the proof make sense (if not, feel free to come by my office). You may very well be thinking "I understand the proof, but I could never come up with that!" This is normal; don't worry about it. Any technique that you've never seen before is going to seem incredible and clever, but the more exposure to proofs you have, the more you will recognize all of the above steps as fairly standard and you will definitely be able to recognize when you need them in the future.

Theorem. *If $x_n \to x$, then the sequence x_n is bounded.*

Proof.

$\exists N : n > N \implies \lvert x_n - x \rvert < 1$	$x_n \to x$
$\exists r = \max\{1, \lvert x_1 - x \rvert, \ldots, \lvert x_N - x \rvert\}$	Maximum of finite set exists
$\therefore \{x_n\}_{n=1}^{\infty}$ is bounded	$r + 1 + \lvert x \rvert$

□

This is a fairly simple proof, but it illustrates a few important points. First, the choice of "1" in the first line is completely arbitrary; I could have chosen any number. Second, and more importantly, the heart of this proof is the second line, where we establish that there is a number r that bounds $|x_n - x|$ and therefore also bounds x_n. However, it is important to recognize that this is a claim (of existence), and it requires a justification. It is very easy in an unstructured proof to just say "let r by the maximum of $\{1, |x_1 - x|, \ldots, |x_N - x|\}$". But how do you know that this maximum exists? If the set were infinite, we wouldn't know this. For example, we saw earlier that $e^{-n} \to 0$, so by our proof, $e^{-1}, e^{-2}, \ldots$ is bounded. However, consider the set $\{\ldots, e^{-(-2)}, e^{-(-1)}, e^{-0}, e^{-1}, e^{-2}, \ldots\}$. This sequence converges as $n \to \infty$, but because there is an infinite collection of numbers leading up to x_N, our proof above doesn't work – we don't know that this set has a maximum (and indeed, it doesn't have a maximum, and the set isn't bounded).

This is a common way in which it is easy to skip steps in an unstructured proof. Of course, one could still write "Let $r = \max\{1, |x_1 - x|, \ldots, |x_N - x|\}$" in a structured proof, but it would be immediately clear that the right hand column is empty and that this statement has not been justified. Now, sometimes it's fine to say "let" without a justification: writing "Let $\epsilon > 0$" and justifying it with "positive numbers exist" is pedantic. One could argue that letting $N = \lceil -\log \epsilon \rceil$ should be justified in the sense that this only exists if $\epsilon > 0$ (which it is), and furthermore we should be more careful in the definition: $N = \max\{1, \lceil -\log \epsilon \rceil\}$ is more technically sound (if ϵ is very large, N could be negative according to our original definition). It never hurts to think about these things, but at the same time, it's something of a judgment call and I felt that going into these justifications was a distraction from the main idea. On the other hand, the "maximum of a finite set" justification is quite important because it's really the main idea of the proof.

Now would be a good time to try proving things on your own. Here are three theorems to start with (solutions will be provided).

Theorem. *Suppose $x_n \to x$ and $y_n \to y$. Then $x_n + y_n \to x + y$.*

Theorem. *Suppose $x_n \to x$ and $y_n \to y$. Then $x_n y_n \to xy$.*

Theorem. *Suppose $x_n \to x$, with $x_n \neq 0$ for all n and $x \neq 0$. Then $1/x_n \to 1/x$.*

For the third theorem, is $x_n \neq 0$ actually required? If we know that $x \neq 0$, can we still have $x_n = 0$?

Continuity

Another major concept in analysis is that of continuity. Suppose we have a function f and a convergent sequence $x_n \to x$. Do we know that as $f(x_n) \to f(x)$? The answer is that no, this doesn't always happen. Only some functions have this property, and those functions are said to be "continuous". Below is the formal definition.

Definition. *A function f is said to be* continuous *at the point x_0 if for every $\epsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(x_0)| < \epsilon$ for all $x : |x - x_0| < \delta$. If f is continuous at every point in its domain, then the entire function f is said to be* continuous.

This is similar to the concept of convergent sequences, except now instead of a countable sequence of points $x_1, x_2, \ldots$, we are concerned with all the points in the "neighborhood" $\{x : |x - x_0| < \delta\}$; that is, the points near x_0. Many of the techniques that we encountered earlier with convergence are very similar to the techniques one uses with continuity. These techniques are often referred to as "delta-epsilon" techniques. For example, the techniques used in the following proof should look fairly familiar by now.

Theorem. *Suppose $x_n \to x_0$ and f is continuous at x_0. Then $f(x_n) \to f(x_0)$.*

Proof. Let $\epsilon > 0$.

$$\begin{aligned} &\exists \delta : |x - x_0| < \delta \implies |f(x) - f(x_0)| < \epsilon && f \text{ continuous at } x_0 \\ &\exists N : n > N \implies |x_n - x_0| < \delta && x_n \to x_0 \\ &\therefore n > N \implies |f(x_n) - f(x_0)| < \epsilon \end{aligned}$$

□

One important thing to note here is that the order of these steps is important. Students often switch the order of the first two lines in this proof, but this makes no sense. The claim "$\exists N : n > N \implies |x_n - x_0| < \delta$" is meaningless if δ hasn't been defined yet. This isn't just semantics: if you were trying to determine how large n had to be in order to ensure that $f(x_n)$ is within a certain tolerance of $f(x_0)$, you couldn't start by finding N. Without using continuity first, you'd have no idea how close x_n must be to x_0 in order to ensure that $f(x_n)$ is within ϵ of $f(x_0)$.

It is worth noting that we can actually make the above theorem into an "if and only if" statement, and thus, an equivalent definition of continuity, but we would have to add the condition that $f(x_n) \to f(x)$ for all sequences $x_n \to x$. For example, a function could satisfy $f(x_n) \to f(x_0)$ for increasing sequences $x_n \nearrow x_0$ but not for decreasing sequences $x_n \searrow x_0$; such functions are not continuous at x_0.

Here are some additional proof exercises related to continuity for you to practice with. Note that the sum and product proofs are very similar to the corresponding proofs for sequences; however, they are still useful exercises if you've never done delta-epsilon proofs before.

Theorem. *Let the functions f and g be continuous at x_0. Then $h = f + g$ is continuous at x_0.*

Theorem. *Let the functions f and g be continuous at x_0. Then $h = f \cdot g$ is continuous at x_0.*

Theorem. *Let the function f be continuous at x_0 and the function g be continuous at $f(x_0)$. Then $h(x) = g(f(x))$ is continuous at x_0.*

Exercise: Write an R function `n(eps)` that returns the smallest N for which $n > N \implies |f(x_n) - f(x_0)| < \epsilon$ for $x_n = 2^{1/n}$ and $f(x) = e^x$.