Some prominent problem-solving patterns in probability proofs

Alex Fu

2023-01-09

There are elements of creativity, intuition, and technique to proofs, but quite often, especially in "routine" proofs and arguments, there is also an element of simple pattern matching or recognition. Despite the title of this note, our discussion will focus on some common patterns appearing in these more "mundane" checks and verifications, rather than the more variable approaches to proofs of deeper results. These patterns are not entirely uninteresting, however: they can still make appearances as intermediate steps in more complex proofs. Their value lies in being an important part of fluency, just like mechanical skills such as breathing or reading, which once learned, you can wield naturally as part of your skillset without needing to consciously apply.

We have already seen some patterns in the proof of the Strong Law of Large Numbers, paradigms like managing complexity, reduction to simpler cases, and transforming qualitative statements into finer quantitative statements. Here, we will aim to discuss a few more. An overarching pattern that we want to highlight first is to

> Follow the definition; work from the definition.

If you have previously done exercises in axiomatic systems, e.g. set theory, abstract algebra, or point-set topology, then this should be a familiar story. Less creativity is required than systematically applying the 2–4 properties you are given as part of the definition. In probability in particular, many constructions like the Lebesgue σ-algebra or the class of all integrable functions are too complicated to be described explicitly. Instead, they are defined via a multistage development process, or *story*, of increasing complexity. Each intermediate step will be straightforward, e.g. taking finite unions or taking monotone limits; the complexity arises from having multiple steps.

Very often, proofs of properties of such objects involve *tracing* the same story as the definition, one of the most prominent examples being Lebesgue integration, which we will soon expand upon.

> Use the few properties that you are given.

1. **Showing two σ-algebras are equal**.

   This seems fairly standard: $\Sigma_1 = \Sigma_2$ iff $\Sigma_1 \subseteq \Sigma_2$ and $\Sigma_2 \subseteq \Sigma_1$. But, as remarked, σ-algebras are difficult to specify explicitly. How can you show $A \in \Sigma_2$ for all $A \in \Sigma_1$ if you don't know what "all $A \in \Sigma_1$" means? In practice, most σ-algebras of interest will have a **generating set** $\mathcal{C}$, a "smaller" collection of sets that *can* be described, such as the collection of open intervals.

   Here is the key: the σ-algebra $\sigma(\mathcal{C})$ generated by $\mathcal{C}$ is the **minimal** σ-algebra containing $\mathcal{C}$.

Let $\mathcal{C}_1, \mathcal{C}_2$ be generating sets of Σ_1, Σ_2 respectively. We can now prove $\Sigma_1 = \Sigma_2$ by showing that $\mathcal{C}_1 = \mathcal{C}_2$, i.e. $\mathcal{C}_1 \subseteq \mathcal{C}_2$ and $\mathcal{C}_2 \subseteq \mathcal{C}_1$, but there is an even less intensive method: showing that $\mathcal{C}_1 \subseteq \Sigma_2$ and $\mathcal{C}_2 \subseteq \Sigma_1$. Why? Minimality. Σ_2 is a σ-algebra containing $\mathcal{C}_1$, but $\Sigma_1 = \sigma(\mathcal{C}_1)$ is the *minimal* σ-algebra containing $\mathcal{C}_1$, so $\Sigma_1 \subseteq \Sigma_2$. Symmetrically, $\mathcal{C}_2 \subseteq \Sigma_1$ implies $\Sigma_2 \subseteq \Sigma_1$.

The same applies to more general structures such as λ-systems or topologies, not just σ-algebras.

> $\Sigma_1 = \Sigma_2$ iff $\mathcal{C}_1 \subseteq \Sigma_2$ and $\mathcal{C}_2 \subseteq \Sigma_1$.

As a side note, interesting set-theoretic issues arise for σ-algebras. For one, not even the set of all countable unions of countable intersections of countable unions of ... sets in $\mathcal{C}$ encompasses all of $\sigma(\mathcal{C})$. To generate a σ-algebra "bottom-up," i.e. iterating set operations, requires *transfinite induction*. This is why we define $\sigma(\mathcal{C})$ "top-down," as the intersection of all σ-algebras containing $\mathcal{C}$, or the minimal σ-algebra containing $\mathcal{C}$. There also exists a "cardinality gap" not present for algebras: σ-algebras are either finite or uncountable.

2. **Showing that every measurable set / every set in a σ-algebra has some property P.**

Minimality is again key. Suppose Σ is generated by $\mathcal{C}$, and the property P holds for every $A \in \mathcal{C}$. Then it suffices to show that $\mathcal{P} = \{A \in \Sigma : A \text{ has property } P\}$ is closed under complements and countable unions! In this case, $\mathcal{P}$ is a σ-algebra, and $\mathcal{P}$ contains $\mathcal{C}$, which means that $\mathcal{P} \supseteq \Sigma$, by which we are done.

> Every measurable set has property P if P holds on a generating set of Σ, such as the collection of open intervals; $P(A) \implies P(A^c)$; and $P(A_1), P(A_2), \ldots \implies P(\bigcup_{i=1}^{\infty} A_i)$.

Naturally, we can generalize this pattern to show that every element in some structure has a given property. In many ways, structures are determined by their generating sets and properties closed under their operations.

3. **Showing that a given set A is measurable.**

This is surprisingly nontrivial for general A. Of course, we are done if we can reduce A to a countable union of known measurable sets, such as generating sets of Σ, but not all $A \subseteq \Omega$ are so nice. One approach is to leverage **countability**. Consider the set of points on which a sequence of measurable functions converges. This requires a statement of the form "for all $\varepsilon > 0, \ldots$," which translates to an uncountable intersection over $\mathbb{R}^+$, but we can make it countable by taking just $\varepsilon = \frac{1}{k}$ for $k \in \mathbb{Z}^+$, e.g. $\bigcap_{k \in \mathbb{Z}^+} \left\{x : |f_n(x) - f(x)| < \frac{1}{k}\right\}$.

Or, the denseness of the rationals in the reals is commonly used to translate into countability. For example, for $r \in \mathbb{R}$, $\{x : f(x) + g(x) > r\}$ is a countable union over $q \in \mathbb{Q}$ of $\{x : f(x) > q\} \cap \{x : g(x) > r - q\}$! Other techniques like diagonalization may be useful for introducing countability as well.

Alternatively, like proving identities for probability using properties of expectation, we can introduce a more complicated object — measurable functions. If A is the preimage of a known measurable set under some f which we show is measurable, then A itself is measurable! Which brings us to the following:

4. **Showing that a given function f is measurable.**

> It suffices to check measurability on a generating set, i.e. $f^{-1}(B) \in \Sigma$ for all $B \in \mathcal{C}'$.

The preimage works very nicely with the boolean set operations of union, intersection, and complement; in fact, the preimage f^{-1} commutes with the set operations. For continuous functions, which are *very* similar to measurable functions in definition, it suffices to check that preimages of open sets in a generating base or subbase are open. The same reasoning applies here: $\{B \in \Sigma' : f^{-1}(B) \in \Sigma\}$ is closed under (arbitrary) union and complements as noted, which makes it a σ-algebra. If it contains $\mathcal{C}'$ generating Σ', then it really equals Σ', which is precisely the definition of f being measurable.

5. **Showing that given σ-algebras $\{\Sigma_i\}_{i\in I}$ are independent**.

   It suffices to check the definition of independence for generating π-systems $\{\Pi_i\}_{i\in I}$. The proof of this fact first reduces to the case where I is finite without loss of generality, then leverages the principle of *successive approximations* or *successive upgrades*, which also finds itself in proofs like those of Dynkin's π-λ theorem or the Lindeberg–Feller Central Limit Theorem.

   Informally, to prove a statement about $(\Sigma_1, \ldots, \Sigma_n)$ given something about $(\Pi_1, \ldots, \Pi_n)$, our first step is to upgrade the known result to $(\Sigma_1, \Pi_2, \ldots, \Pi_n)$, then work step-by-step to $(\Sigma_1, \ldots, \Sigma_n)$, by induction for example. We thus break a tough implication into smaller, more manageable "rungs" on a ladder, and climb up one step at a time.

6. **Showing that something is true for all integrable functions**.

   Use the "approximation" or "monotone class argument," the same steps through which the Lebesgue integral and the class of integrable functions were initially defined. For instance, show that property P holds for all indicator functions, is closed under taking linear combinations, pointwise limits, and differences of unsigned parts in order to show that $\{f \text{ integrable } : f \text{ has property } P\}$ contains all integrable functions.

   You usually don't have to start all the way down at the level of indicators; the starting point could be simple functions, bounded measurable functions, or even nonnegative measurable functions. These "generate" the class of all integrable functions just as well as indicators.

   Lastly, a few miscellanous patterns that boil down to the following principle.

   Reduce your problems to match the form of known results. Invoke what you know.

7. **Showing weak convergence, especially to a Gaussian**. Manipulate your problem into a statement about triangular arrays satisfying the Lindeberg conditions, or reuse the method of characteristic functions in the general CLT by relating via approximation or otherwise, e.g. by the Lindeberg exchange method.

8. **Given convergence in distribution**. Invoke Skorohod's representation theorem. Upgrading weak distributional convergence to strong or almost sure convergence is a powerful move because it gives you access to many of the convergence theorems that you might want, like a higher vantage point. And, the extra height will not hurt, because you always "jump down" at the end using a.s. $\implies$ i.d., or taking the distributions.

9. **Showing something (elementary) about martingales**. Formulate the statement in terms of the optional stopping theorem (OST) or any of its variations often does the trick. Or, construct a new random process and show that it is a martingale, then apply known properties.

■