Big Probability Theorems Intuitively Explained

Xingjian Wang

January 6, 2026

I will attempt to organize and show several theorems from measure-theoretic probability in the most intuitive manner. This does not mean that the proofs are concise; rather, I expect the exact opposite. The advantage of this approach is that the exposition remains readable and that each step of the argument follows one's normal intuition; our proof ideas are explicitly motivated and traceable, rather than relying on abrupt manipulations.

We expect the reader to bear the burden of summarizing and condensing the material, as we only provide the most stream-of-consciousness proofs. The goal here is that the proof ideas are completely dissected and understood upon the first close reading and leaves no fuzzy understanding.

First, we attempt to show the Riesz-Weyl subsequence theorem. We can do this by first revisiting the set theoretic expressions of convergence and using Borel-Cantelli, before relying on a diagonal trick to find a subsequence.

Prop 1. For a sequence of functions (f_n) sending $(\Omega, \mathcal{F}, \mu)$ to $\mathbb{R}$ (matter of fact, any metric space works), we denote the set of points x on which f_n does not converge to f pointwise as $\{x : \exists \epsilon > 0 \text{ s.t. } |f_n(x) - f(x)| \geq \epsilon \text{ i.o.}\}$, or $\{x : \exists \epsilon > 0 \text{ s.t. } \forall N \in \mathbb{N}, \exists n \geq N \text{ s.t. } |f_n(x) - f(x)| \geq \epsilon\}$, or

$$\bigcup_{\epsilon>0} \left\{ \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} \{x : |f_n - f|(x) \geq \epsilon\} \right\}$$

where by definition, inside the braces we have

$$\limsup_n \{x : |f_n - f|(x) \geq \epsilon\}.$$

Or, we can represent the entire thing as

$$\bigcup_{m=1}^{\infty} \left\{ \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} \{x : |f_n - f|(x) \geq \epsilon_m\} \right\}$$

if we take $\epsilon_m \downarrow 0$. These are equal, as one direction is simply inclusion, and the other reduces to the statement "for all $\epsilon > 0$, we can find $\epsilon_m < \epsilon$" from where an inclusion is established by mapping ϵ to some smaller ϵ_m.

Taking complement and flipping everything set theoretically gives points that do converge. Think in the language of $\liminf$.

Lemma 2. (Borel-Cantelli) For a sequence of events (E_n), suppose that $\sum_n \mu(E_n) < \infty$, then $\mu(\limsup_n E_n) = 0$.

Proof. Assume that the summation identity holds, then $\sum_{n \geq N} \mu(E_n) \to 0$ as $N \to \infty$. Since also $\mu\left(\bigcup_{n \geq N} E_n\right) \leq \sum_{n \geq N} \mu(E_n)$, we have

$$\mu\left(\lim_{N\to\infty} \sup_{n \geq N} E_n\right) = \lim_{N\to\infty} \mu\left(\bigcup_{n \geq N} E_n\right) \leq \lim_{N\to\infty} \sum_{n \geq N} \mu(E_n) = 0$$

where the first equality holds by continuity from above, and second equality holds by our previous statement.

$\square$

Remark. In the language of probability, we have that if $\sum_n \mathbb{P}[E_n] < \infty$ then $\mathbb{P}[\limsup_n E_n] = 0$.

These two results allow us to bridge convergence in measure (in probability) and convergence almost everywhere (almost surely).

Theorem 3. (Riesz-Weyl, Weak Version) For any $f_n \xrightarrow{\text{m}} f$, we may find some subsequence $(n_k)_{k=1}^{\infty}$ such that $f_{n_k} \xrightarrow{\text{a.e.}} f$.

Proof. We want to find a subsequence f_{n_k}. We want for all $\epsilon > 0$ that $\mu(\{x : |f_{n_k}(x) - f(x)| \geq \epsilon \text{ i.o.}\}) = 0$ holds.

Since $\{x : |f_{n_k}(x) - f(x)| \geq \epsilon \text{ i.o.}\} = \limsup_k \{x : |f_{n_k}(x) - f(x)| \geq \epsilon\}$, we label this as $\limsup_k E^{\epsilon}_{n_k}$ and try to show $\mu\left(\limsup_k E^{\epsilon}_{n_k}\right) = 0$ by showing $\sum_k \mu(E^{\epsilon}_{n_k}) < \infty$ and using Borel-Cantelli.

We run into a problem; for a fixed ϵ, we can pick n_k accordingly so that $\mu(E^{\epsilon}_{n_k}) < 2^{-k}$ (by the definition of convergence in measure) so the sum is finite. However, for some other $\epsilon' < \epsilon$, this subsequence does not work.

We patch that by the standard diagonal argument where we find $\eta_k \downarrow 0$ and each n_k satisfying $\mu(E^{\eta_k}_{n_k}) < 2^{-k}$. The interpretation is that at n_k, only on a 2^{-k}-sized domain does f_{n_k} get η_k-far from f. Then, for any $\epsilon > 0$, find K such that $\eta_k < \epsilon$ if $k \geq K$, so that the first $\mu(E^{\eta_1}_{n_k})$ to $\mu(E^{\eta_K}_{n_k})$ terms sum to a finite measure, and all $k > K$ has $\mu(E^{\epsilon}_{n_k}) < \mu(E^{\eta_k}_{n_k})$, since suppose

$\eta_k < \epsilon$ then more x enables f_{n_k} to get η_k-far than ϵ-far. This yields the inequality below. Specifically, we have

$$\begin{aligned}\sum_{k=1}^{\infty} \mu\left(E_{n_k}^{\epsilon}\right) &= \sum_{k=1}^{K} \mu\left(E_{n_k}^{\epsilon}\right) + \sum_{k=K+1}^{\infty} \mu\left(E_{n_k}^{\epsilon}\right) \\ &\leq \sum_{k=1}^{K} \mu\left(E_{n_k}^{\epsilon}\right) + \sum_{k=K+1}^{\infty} \mu\left(E_{n_k}^{\eta_k}\right)\end{aligned}$$

and the first part is finite while the second part has each term bounded by 2^{-k}. Therefore, $\sum_{k=1}^{\infty} \mu\left(E_{n_k}^{\epsilon}\right)$ is finite, moreover $\mu\left(\limsup_k E_{n_k}^{\epsilon}\right) = 0$ by Borel-Cantelli. Since $\limsup_k E_{n_k}^{\epsilon} = \{x : |f_{n_k}(x) - f(x)| \geq \epsilon \text{ i.o.}\}$, we have for any ϵ the sequence (f_{n_k}) gets ϵ-close almost everywhere. Let $\epsilon_m \downarrow 0$ so $\mu(\{x : |f_{n_k}(x) - f(x)| \geq \epsilon_m \text{ i.o.}\}) \uparrow \mu(\{x : \exists \epsilon \text{ s.t. } |f_{n_k}(x) - f(x)| > \epsilon \text{ i.o.}\})$, by continuity from below. This gives us precisely the nonconvergent domain, where $0 \uparrow 0$, and we are done.

To summarize, let $\eta_k \downarrow 0$ and for each η_k pick n_k so that f_{n_k} is η_k-close to f barring a 2^{-k}-measure set. For each ϵ, "$|f_{n_k} - f|$ exceeding ϵ i.o." is null because $\sum_k \{|f_{n_k} - f| \text{ exceeding } \epsilon\} < \infty$, relying on some $\eta_k < \epsilon$, then Borel-Cantelli. Intersect over ϵ to get $\mu(\text{nonconvergent domain}) = 0$.

□

Remark. Notice that ϵ_m exists for countability purposes. The statement $\{x : |f_{n_k}(x) - f(x)| \geq \epsilon_m \text{ i.o.}\} \uparrow \{x : \exists \epsilon \text{ s.t. } |f_{n_k}(x) - f(x)| > \epsilon \text{ i.o.}\}$ simply captures $\bigcup_{\epsilon_k} \{\cdot\}$ vs $\bigcup_{\epsilon > 0} \{\cdot\}$'s equality, which we showed in prop 1.

Next we will visit some standard theorems that rely on similar techniques. The ones shown above are quite universal in proving convergence. One must be familiar with the language of i.o. vs ult., $\limsup$ vs $\liminf$, and $\bigcup_N \bigcap_{n \geq N}$ vs $\bigcap_N \bigcup_{n \geq N}$ before proceeding.

Theorem 4. (Egorov) Assume $\mu(X) < \infty$. Then $f_n \xrightarrow{\text{a.e.}} f$ gives $f_n \xrightarrow{\text{a.u.}} f$.

Remark. Without any restrictions, convergent almost uniformly implies convergent almost everywhere. Adding $\mu(X) < \infty$ establishes that as N grows, the measure of "unconquered x at N" remains finite and goes to zero. This move establishes the reverse implication.

Proof. For any $\eta > 0$, we need to find E such that $\mu(E) < \eta$ and that f_n converges uniformly to f on $X \setminus E$.

Let's examine pointwise convergence more carefully and see what stops us from getting uniform convergence. Fix $\epsilon > 0$, assuming convergence a.e., then for each x in a full set, exists $N = N(x, \epsilon)$ such that $|f_n(x) - f(x)| < \epsilon$ for all $n \geq N$ (definition). Hence, for each N there is a set $X_N^\epsilon \subset X$ on which $|f_n(x) - f(x)|$ stays $< \epsilon$ starting at N. In other words, some x are "conquered" at N and makes $f_n(x)$ remain ϵ-close to $f(x)$ starting at N. Taking union over all N gives a full set in X, as each x must pick an X_N^ϵ.

Let's look at x that remains "unconquered" after N, i.e. points that still allow a ϵ-deviation somewhere after time N. Suppose E_n^ϵ denotes the event $\{x : |f_n(x) - f(x)| > \epsilon\}$. Then $X \setminus X_N^\epsilon = \bigcup_{n \geq N} E_n^\epsilon$, with the interpretation that x "remains unconquered at N" equals "still allows an ϵ-deviation at some $n \geq N$". This is a decreasing sequence that converge to a null set.

Notice the interpretation $\bigcup_{n \geq N} E_n^\epsilon = \sup_{n \geq N} E_n^\epsilon \downarrow \limsup_n E_n^\epsilon$, denoting points on which (f_n) never get fully ϵ-close. This has measure 0 by definition. We try to think about its relationship with the aforementioned "nonconvergent set", which is also represented with $\limsup$. The finiteness of $\mu(X)$ allows for continuity from above, where $\mu(\bigcup_{n \geq N} E_n^\epsilon) < \infty$ works for any N. This is the only use of requiring $\mu(X) < \infty$.

Suppose we want to build E, then for all ϵ, it suffices to pick some N_ϵ, build X_N^ϵ (the conquered set), and throw the remaining unconquered x into E. The only task is to bound $\mu(E) < \eta$. We have two tools; one is the freedom to vary N_ϵ, while the other is countability trick, where instead of including all ϵ, we only include all ϵ_k, where any $(\epsilon_k) \downarrow 0$ works. This way for any ϵ, find $\epsilon_k < \epsilon$ so that given the N_k we pick, the unconquered set $X \setminus X_{N_k}^{\epsilon_k}$ includes $X \setminus X_{N_k}^\epsilon$. The interpretation is that x that allows ϵ-deviation after N_k must also allow ϵ_k-deviation after N_k. Same idea as Riesz-Weyl.

From this construction, if for each ϵ_k, E includes its unconquered set (each built with a corresponding N_k), then every ϵ can rely on some ϵ_k, to borrow its N_k so that if $X \setminus X_{N_k}^{\epsilon_k}$ is in E, then $X \setminus X_{N_k}^\epsilon$ is also in E.

Then our problem comes down to picking each N_k for ϵ_k. Since for each $\epsilon > 0$, its unconquered set $X \setminus X_N^\epsilon = \sup_{n \geq N} E_n^\epsilon$ converges to a null set (by a.e. convergence) as $N \to \infty$, we can find for each ϵ_k some N_k so that $\mu(X \setminus X_{N_k}^{\epsilon_k}) < 2^{-k}\eta$. Letting $E = \bigcup_k (X \setminus X_{N_k}^{\epsilon_k})$, summing over ϵ_k gives a bound of $\mu(E) < \eta$. We are done.

To summarize, given η, let $\epsilon_k \downarrow 0$ and for each k pick N_k so the "un-ϵ_k-conquered x after N_k", $\mu(\{x : \exists n \geq N_k \text{ s.t. } |f_n - f|(x) > \epsilon_k\}) < 2^{-k}\eta$, which is possible as "$..._k$ after N" decreases to a null set as $N \to \infty$, using continuity from above. Make its union over all k precisely E, then for any ϵ pick $\epsilon_k < \epsilon$ and use N_k; we get uniformly $\|f_n - f\|_\infty < \epsilon$ on $X \setminus E$.

$\square$

We employ all the techniques we have used to tackle the strong version of Riesz-Weyl.

Theorem 5. (Riesz-Weyl, Strong Version) If f_n is measure-Cauchy, then it has an a.u. (hence a.e.) convergent subsequence (f_{n_k}). Moreover, this would imply that (f_n) itself is measure convergent.

Proof. At this point, all the tools were already used. Extending to this result is not a huge jump from the previous two results. It would be a good exercise to attempt the proof freehand.

Find an a.e. Cauchy (hence convergent) subsequence using the same mechanism as the weak Riesz-Weyl, and use the second half of Egorov to establish that since "unconquered set at n_k" goes to null as $k \to \infty$, find N_k so that the tail sum (bounds the measure of E) is controlled.

□

Theorem 6. (Dominated Convergence)

Theorem 7. (Monotone Convergence)

Lemma 8. (Fatou) For a sequence of measurable functions $(f_n) : (\Omega, \mathcal{F}, \mu) \to [0, \infty]$, we have

$$\int \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n.$$

Proof. □