A New Proof of Lebesgue Density Theorem

Xiang Huang

July 12, 2018

Abstract

I ran into Lebesgue Density Theorem in the very early stage of my research (while I was reading the first Chapter of *Algorithmic Randomness and Complexity* by Downey et al.). However, due to some typos, it is really hard to understand some critical facts about the proof. Plus, the original proof is non-constructive. Here I developed a constructive one.

The Original Proof

A measurable set $\mathcal{A} \subseteq 2^\omega$ has *density d at X* if

$$\lim_{n\to\infty} \mu(\mathcal{A} \mid [X \upharpoonright n])2^n = d.$$

Define $\phi(\mathcal{A}) = \{X \in 2^\omega : \mathcal{A} \text{ has density 1 at } X\}$.

The following proof of Lebesgue Density Theorem can be found in [1].

Lebesgue Density Theorem. *If $\mathcal{A}$ is measurable then so is $\phi(\mathcal{A})$, and $\mu(\mathcal{A}\Delta\phi(\mathcal{A})) =$ 0.*

Proof. It suffices to show that $\mathcal{A} - \phi((A))$ is a null set since $\phi(\mathcal{A}) - \mathcal{A} \subseteq \overline{\mathcal{A}} - \phi(\overline{\mathcal{A}})$ and $\overline{\mathcal{A}}$ is measurable. Define for every positive rational ε

$$\mathcal{B}_\varepsilon = \{X \in \mathcal{A} : \liminf_{n\to\infty} \mu(\mathcal{A} \cap [X \upharpoonright n])2^n < 1 - \varepsilon\}.$$

Then $\mathcal{A} - \phi(\mathcal{A}) = \bigcup_\varepsilon \mathcal{B}_\varepsilon$, hence it suffices o prove that every $\mathcal{B}_\varepsilon$ is a null set. Suppose for a contradiction that for $\mathcal{B} = \mathcal{B}_\varepsilon$ we have that the outer measure

$$\mu^*(\mathcal{B}) := inf\{\mu(\mathcal{U}) : \mathcal{B} \subseteq \mathcal{U} \wedge \mathcal{U} \text{ open}\} > 0.$$

Then there exists $\mathcal{G} \supseteq \mathcal{B}$ open with $\mu(\mathcal{G})(1 - \varepsilon) < \mu^*(\mathcal{B})$. Define

$$I = \{x \in 2^{<\omega} : [x] \subseteq \mathcal{G} \wedge \mu(\mathcal{A} \cap [x]) < (1 - \varepsilon)2^{-|x|}\}.$$

Then

(i) for any $X \in \mathcal{B}$, I contains $X \upharpoonright n$ for some n, and

(ii) if $\{x_i\}_{i\in\omega}$ is a prefix-free set of elements of I then $\mu^*(\mathcal{B} - \bigcup_i [x_i]) > 0$.

The first statement holds since $\mathcal{G}$ is open, and for any $X \in \mathcal{B}$, there are infinite many n such that $\mu(\mathcal{A} \cap [X \upharpoonright n])2^n < 1 - \varepsilon$. Hence there must be a long enough prefix of X satisfies both of the conditions.

The second statement holds because

$$\mu^*(\mathcal{B} \cap \bigcup_i [x_i]) \leq \sum_i \mu(\mathcal{A} \cap [x_i]) < \sum_i (1-\varepsilon)2^{-|x_i|} \leq (1-\varepsilon)\mu(\mathcal{G}) < \mu^*(\mathcal{B}).$$

Construct a sequence $\{x_i\}_{i\in\omega}$ as follows. Let x_0 in I be arbitrary, and if x_i, $i \leq n$ are defined such that $\{x_0, \cdots, x_n\}$ is prefix-free, define $x_{n+1} \in I$ of minimal length such that the set $\{x_0, \cdots, x_{n+1}\}$ is again prefix-free.

Now let $X \in \mathcal{B} - \bigcup_i [x_i]$. X exists by (ii). By (i), let $x \in I$ be such that $X \in [x]$. Let k be the smallest number with $[x] \cap [x_k] \neq \emptyset$. Note that k exists because there are only finitely many $y \in I$ of length shorter than x so either $x = x_k$ or $x \sqsubset x$ for some k. In both cases $X \in [x] \subseteq [x_k]$, contradicting that X was not in $\bigcup_i [x_i]$. □

The above is a proof by contradiction. But I don't really get the point of this proof.

Comment. (1) The open set $\mathcal{G}$ gives us some extra room to pick cylinders.

(2) The selection of the prefix set is some sort of maximal prefix set one can get, and this prefix set should cover everything (infinite sequence contained) in the set I.

(3) But according to the second statement above, no matter what kind of prefix we are going to pick, we always miss something. There are always some left-over element that we can not cover.

(4) One key fact is: if $\mu^*(\mathcal{B})$ is not 0, then there is an open cover that has a measure smaller than $\mu^*(\mathcal{B})/(1-\varepsilon)$. This is not true for null set.

(5) However, I can not see **why the target set is small** , other than having this contradiction that I do not understand why it matters.

A Constructive Proof

We now show that the set $\mathcal{A}\Delta\phi(\mathcal{A})$ actually has constructive measure zero. Ideas of the proof: We use the fact that a measurable set can be approximated by open set. That is, given a measurable set $\mathcal{A}$, there is always some open cover of $\mathcal{A}$, $\mathcal{G}$, such that $\mu(\mathcal{G}) < (1+\delta)\mu(\mathcal{A})$, for any fixed $\delta > 0$.

Then imagine that we are going to take an infinite sequence from $\mathcal{G}$. The average "chance" that we get something from $\mathcal{A}$ is, obviously, $\frac{\mu(\mathcal{A})}{\mu(\mathcal{G})} > \frac{1}{1+\delta}$. However, the set of cylinders we care about have density smaller than $1-\varepsilon$. Observer that $\frac{1}{1+\delta}$ can be very close to 1, if we push so hard by picking δ **very** small. Comparatively, $1-\varepsilon$ seems some far off its "average". Then we can see why the set we care about is small by applying some "Markov inequality" argument.

Here comes the proof.

Proof. If $\mu(\mathcal{A})=0$, the result follows directly. Hence we assume $\mu(\mathcal{A})>0$. Let $\delta>0$ to be some fixed number that we decide its value later. Let $\mathcal{G}$ be an open cover of A such that $\mu(\mathcal{G})\le(1+\delta)\mu(\mathcal{A})$.

Define

$$I=\{x\in 2^{<\omega}:[x]\subseteq\mathcal{G}\wedge\mu(\mathcal{A}\cap[x])<(1-\varepsilon)2^{-|x|}\}.$$

As we pointed out in the original proof, we can see for $X\in\mathcal{B}$, there is always a prefix of X in I.

Now lets consider the following fact: Let $\mathcal{G},\mathcal{A}$ be and $\mathcal{G}=\bigcup_n[x_n]$, where $x_n\in 2^{<\omega}$ form a partition of $\mathcal{G}$. Define a random variable $d(X)$, such that

$$d(X)=\begin{cases}\frac{\mu(A\cap[x_n])}{\mu([x_n])}, & \text{if } X\in\mathcal{G} \text{ and } X\in[x_n],\\ 0. \quad X\notin\mathcal{G}.\end{cases}$$

Then $E[1-d(X)\mid\mathcal{G}]\le\frac{\delta}{1+\delta}$. To see this, we note that:

$$\begin{aligned}E[d(X)\mid\mathcal{G}]&=\frac{\int_{\mathcal{G}}d(X)d\mu}{\mu(\mathcal{G})}\\&=\frac{\sum_n\int_{[x_n]}d(X)d\mu}{\mu(\mathcal{G})}=\frac{\sum_n\mu([x_n])\frac{\mu(\mathcal{A}\cap[x_n])}{\mu([x_n])}}{\mu(\mathcal{G})}\\&=\frac{\sum_n\mu(\mathcal{A}\cap[x_n])}{\mu(\mathcal{G})}=\frac{\mu(\mathcal{A}\bigcup_n[x_n])}{\mu(\mathcal{G})}\\&=\frac{\mu(\mathcal{A}\cap\mathcal{G})}{\mu(\mathcal{G})}=\frac{\mu(\mathcal{A})}{\mu(\mathcal{G})}\ge\frac{1}{1+\delta}\end{aligned}$$

By linearity of expectation, we have $E[1-d(X)\mid\mathcal{G}]\le\frac{\delta}{1+\delta}$.

We now come back to the proof. Let $U=\{x_1,\cdots,x_n,\cdots\}$ be picked from I as the original proof. Then since $\mathcal{G}-\bigcup_n[x_n]$ is still an open set, we can write it as a union of cylinders. Let $V=\{y_m\}_{m\in\omega}$ be such a union of cylinders.

Then $U\cup V$ is a partition of $\mathcal{G}$. We define $d(X)$ on this partition of $\mathcal{G}$ as above. We now measure the size of U. We have

$$\begin{aligned}\mu(U)&=\mu(\bigcup_n[x_n])\\&=\mu(\{X\in\mathcal{G}\mid 1-d(X)>\varepsilon\})\\&<E[(1-d(X))\mid\mathcal{G}]/\varepsilon\\&\le\frac{\delta}{1+\delta}\frac{1}{\varepsilon}\\&<\delta\cdot\frac{1}{\varepsilon}\end{aligned}$$

Let $\delta = \varepsilon \cdot \frac{1}{2^n}$. We can create a sequence of $\{U_n\}$ that is a witness of constructive measure zero. $\square$

Application of Lebesgue Density Theorem

Next we prove a classical theorem from computability theory: Sacks' theorem on the measure of upper cones. The proof is taken from [1].

Theorem. *(Sacks [2]) For every noncomputable set $A \in 2^\omega$ the upper cone*

$$A^{\leq T} = \{B : A \leq_T B\}$$

has measure zero.

Proof. Let A be an element of 2^ω such that $\mu(\{B : A \leq_T B\}) \neq 0$. Note that this set is measurable hence must have positive measure. We will show that A is computable. For two class $\mathcal{C}$ and $\mathcal{D}$ define $\mu(\mathcal{C} \mid \mathcal{D}) = \mu(\mathcal{C} \cap \mathcal{D})/\mu(\mathcal{D})$. and for a formula P write $\mu(P(B))$ for $\mu(\{B \in 2^\omega : P(B)\})$. Since $\{B : A \leq_T B\} = \bigcup_\varepsilon \{B : A = \{e\}^B\}$ has positive measure there exists $e \in \omega$ such that $\mu(\{B : A = \{e\}^B\}) > 0$. It follows from Lebesgue Density Theorem that there is a point $X \in 2^\omega$ such that $\mathcal{A}$ has density 1 at X. From this follows the existence of a $\sigma \in 2^{<\omega}$ such that $\mu(A = \{e\}^B \mid [\sigma]) \geq \frac{3}{4}$. Using σ we can compute A as follows. For any $x \in \omega$ the set $T_n = \{X \sqsupset \sigma : \{e\}^X \downarrow = n\}$ are uniformly c.e., so we can enumerate the $T_n' s$, $T_n = \bigcup_s T_{n,s}$, until we find n with $\mu(T_{n,s}) \geq \frac{3}{4}$. Then $A(x) = n$.

$\square$

Comment. (1)

The last sentence seems to be very confusing. We can read $A(x) = n$ as x-th bit of A is $n \in \{0, 1\}$. Actually, n can be something from Baire space and hence we can interpret this broadly.

2. The proof idea is since around $[\sigma]$, with high probability, every infinite sequence under $[\sigma]$ is going to be correct. We just need to take a vote and vote for the majority.

3. This is an existential proof. We only know that there exist σ, and we can hard-code σ to implement a oracle TM to compute A, but we do not know exactly which σ we are going to use. We can surely search for it, but we do not know which one it is. We might end up a σ that will give us a wrong vote.

4. also note that the process $\{e\}^X(x) \downarrow = n$ will eventually ends, we can have an output. The number of inquires to the oracle machine is finite. So we can enumerate all such X. But how can we compute the measure of the set of X we get? Well, when we enumerate X's, we actually do not enumerate them one by one; we enumerate them cylinder by cylinder. Say for some X we only make 100

queries and halt. The there must be a fastest point we make our queries. Then the whole cylinder of sequence with the same prefix up to the fastest point above will have the same effect.

Definition 1. (Tail set) $\mathscr{E} \subseteq 2^\omega$ is a *tail set* if $\mathcal{A}$ is closed under finite variances, i.e., if $v \in 2^{<\omega}$ and $X \in 2^\omega$ are such that $vX \in \mathscr{E}$ for every string ω of length $|v|$.

Theorem. *(Kolmogorov's 0-1 law) If $\mathcal{A} \subseteq 2^\omega$ is a measurable tail set then either $\mu(\mathcal{A}) = 0$ or $\mu(\mathcal{A}) = 1$.*

Proof. Suppose $\mu(\mathcal{A}) > 0$. By Lebesgue Density Lemma, choose $X \in \mathcal{A}$ such that $\mathcal{A}$ has density 1 at X. Let $\varepsilon \in (0, 1)$ be arbitrary. Choose n large enough such that $\mu(\mathcal{A} \cap [X \upharpoonright n])2^n > 1 - \varepsilon$ for *any* ω of length n. So $\mu(\mathcal{A}) > 1 - \varepsilon$. Since ε was arbitrary it follows that $\mu(\mathcal{A}) = 1$. $\square$

Comment. (1) Very nice proof. But can we do it in another way? Say, if the set is not with measure 1, how can we prove that I has measure zero?

The Proof in Analysis Setting

Theorem. *(**Lebesgue Density Theorem**) For any measurable set $E \subset \mathbb{R}$, $m(E\delta\phi(E)) = 0$.*

Proof. It is sufficient to show that $E - \phi(E)$ is a nullset, since $\phi(E) - E \subset E' - \phi(E')$ and E' is measurable. We may also assume that E is bounded. Furthermore, $E - \phi(E) = \bigcup_{\varepsilon > 0} A_\varepsilon$, where

$$A_\varepsilon = \{x \in E \mid \liminf_{h \to 0} \frac{m(E \cap [x - h, x + h])}{2h} < 1 - \varepsilon\}$$

Hence it is sufficient to show that A_ε is a nullset for every $\varepsilon > 0$. Putting $A = A_\varepsilon$, we shall obtain a contradiction from the supposition that $m^*(A) > 0$.

If $m^*(A) > 0$ there exist a bounded open set G containing A such that $m(G) < m^*(A)/(1 - \varepsilon)$. Let $\mathscr{E}$ denote the class of all closed intervals I such that $I \subset G$ and $m(E \cap I) \leq (1 - \varepsilon)|I|$. Observe that

(i) $\mathscr{E}$ includes arbitrarily short intervals about each point of A, and

(ii) for any disjoint sequence $\{I_n\}$ of members of $\mathscr{E}$, we have $m^*(A - \bigcup I_n) > 0$. Property (ii) follows from the fact that

$$m^*(A \cap \bigcup I_n) \leq \sum m(E \cap I_n) \leq (1 - \varepsilon) \sum |I_n| \leq (1 - \varepsilon)m(G) < m^*(A). \quad (1)$$

We construct inductively a disjoint sequence I_n of members of $\mathscr{E}$ as follows. Choose I_1 arbitrarily from $\mathscr{E}$. Having chosen $I_1, \cdots, I_n$, let $\mathscr{E}$ be the set of members of $\mathscr{E}$ that are disjoint to $I_1, \cdots, I_n$. Properties (i) and (ii) imply that $\mathscr{E}_n$ is non-empty. Let d_n

be the least upper bound of the lengths of members of $\mathscr{E}_n$, and choose $I_{n+1} \in \mathscr{E}$ such that $|I_{n+1}| > \frac{d}{2}$. Put $B = A - \bigcup_1^\infty I_n$. By (ii), we have $m^*(B) > 0$. Hence there exists a positive integer N such that

$$\sum_{N+1}^{\infty} |I_n| < m^*(B)/3.$$

For each $n > N$ let J_n denote the interval concentric with I_n with $|J_n| = 3|I_n|$. The inequality (1) implies that the intervals $J_n (n > N)$ do not cover B, Hence there exist a point $x \in B - \bigcup_{N+1}^\infty J_n$. Since $x \in A - \bigcup_1^N I_n$, it follows from (i) that there exists an interval $I \in \mathscr{E}_N$ with center x. I must meet some interval I_n with $n > N$. (Otherwise $|I| \leq d_n \leq 2|I_{n+1}|$) for all n, contrary to $x \notin \bigcup_{N+1}^\infty J_n$. $\square$

Comment. 1. Compare to the Cantor space proof. Here when we want to pick the "largest interval", we may not be able to do so. Since there are infinite many intervals, there may not be an interval actually assumes the sup of sizes. Hence we compromise and instead pick something large enough.

But in Cantor space, we can always pick the largest cylinder by picking the one with shortest length. However, since there are infinitely many strings. To pick the shortest one out of them, the process can not be done by a Turing machine.

2. One big difference in my proof and this proof and its Cantor space version is, they are trying to approximate $\mathcal{A} - \phi(\mathcal{A})$, but I am trying to approximate $\mathcal{A}$ itself. I try to approx $\mathcal{A}$, and get an open cover of $\mathcal{B}$. They try approx $\mathcal{A} - \phi(\mathcal{A})$, and by definition, the inequality (1) hence holds.

3. $d_n \to 0$ as $n \to \infty$.

4. for any d_n, there are only finitely many intervals have measures greater than d_n.

5. I can develop a proof for LDT using similar idea to what I have for the Cantor space version.

6. Pick $> d_n/2$ interval and expand the things we pick is a very good way to bound the measure of a set of intervals.

7. The classic proof is basically the same as the Vitali Covering Lemma in page 109 of *real analysis* by Royden and Fitzpatrick. The only difference is here our intervals center at x, and in the textbook the intervals contains x. So $3I$ is enough here, while in the textbook we have to take $5I$.

I spent too much time on this single problem. I have to stop here now.

References

[1] Terwijn, S. A. (2003). Complexity and randomness. Department of Computer Science, The University of Auckland, New Zealand.

[2] G. E. Sacks, Degrees of unsolvability, Annals of Mathematics Studies 55, Princeton University Press, 1963.