

# Probability and Measure Solutions

WispyAbyss

December 13, 2024

## Contents

<b>Forward</b>	<b>1</b>
<b>Chapter 1.1 - Borel's Normal Number Theorem</b>	<b>1</b>
Notes	1
The Unit Interval	2
The Weak Law of Large Numbers	3
The Strong Law of Large Numbers	4
The Measure Theory of Diophantine Approximation	7
Problems	9
1.1	9

## Forward

This document will contain notes and solutions corresponding to Probability and Measure, Third Edition, by Patrick Billingsley [amazon].

## Chapter 1.1 - Borel's Normal Number Theorem

### Notes

For a complete understanding of probability, you need to understand an infinite number of events as well as a finite number of events. We try and present why that must be so here.

## The Unit Interval

We take the length of an interval  $I = (a, b] = b - a$ . Note, for  $A$  a disjoint set of intervals in  $(0, 1]$ , we have that  $P(A)$  is well defined. If  $B$  is a similar disjoint set, and is disjoint from  $A$ ,  $P(A + B) = P(A) + P(B)$  is well defined as well. Note - we haven't defined anything for intersections yet. These definitions can also directly stem from the Riemann integral of step functions.

The unit interval can give the probability that a single particle is emitted in a unit interval of time. Or a single phone call comes in. However, it can also model an infinite coin toss. This is done as follows - for  $\omega \in (0, 1]$ , define:

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n}$$

Where  $d_n(\omega)$  is 0 or 1, and comes from the binary expansion of  $\omega$ . We take  $\omega$  as the non terminating representation. Note, we were particular when we defined intervals as half inclusive. Examine the set of  $\omega$  for which  $d_i(\omega) = u_i$  for  $i = 1, \dots, n$ ,  $u_i \in \{0, 1\}$ . We have that:

$$\sum_{i=1}^n \frac{u_i}{2^i} < \omega \leq \sum_{i=1}^n \frac{u_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i}$$

We cannot have the lower extreme value, as this would imply  $\omega$  takes on its terminating binomial representation, which is what we said we would not do. This is our first taste, I guess, of measure 0 sets, we we still have:

$$\mathbb{P}[\omega : d_i(\omega) = u_i, i = 1, \dots, n] = \frac{1}{2^n}$$

Note, probabilities of various familiar events can be written down immediately. Ultimately, note, however, each probability is the sum of disjoint dyadic intervals of various ranks  $k$ . Ie, all the events are still well defined by our probability definition above. We have:

$$\mathbb{P}\left[\omega : \sum_{i=1}^n d_i(\omega) = k\right] = \binom{n}{k} \frac{1}{2^n}$$

All these results have been for finitely many components of  $d_i(\omega)$ . What we are interested in, however, is properties of the entire sequence of  $\omega = (d_1(\omega), d_2(\omega), \dots)$ .

## The Weak Law of Large Numbers

What I like about this chapter, is to me - it *emphasizes* the connection between the *structure of real numbers*, and probability. At the end of the day - probability can be seen as just extracting properties of *frequency* over the real numbers, to be understood as probabilistic statements. However, with just our basic real numbers - we can't really prove a lot of properties about infinite things. That is when measure theory comes in later. However, for now, we look at what we can prove - and that starts with the weak law of large numbers. We have:

**Theorem 1.1 - The Weak Law of Large Numbers** For each  $\epsilon$ :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \omega : \left| \frac{1}{n} \sum_{i=1}^n d_i(\omega) - \frac{1}{2} \right| \geq \epsilon \right] = 0$$

Probabilistically - this is saying that if  $n$  is large, then there is a small probability that the fraction/relative *frequency* of heads in  $n$  tosses will deviate much from  $1/2$ . Think about it as a statement over the real numbers as well - it is also interesting. Ultimately, the intervals containing  $\omega$  that do not satisfy the above are getting smaller and smaller and smaller. We formalize this with the following concept:

As  $d_i(\omega)$  are constant over each dyadic interval of rank  $n$  if  $i \leq n$ , the sums  $\sum_{i=1}^n d_i(\omega)$  are also constant over rank  $n$ . Thus, the set in the theorem is just a disjoint union of dyadic intervals of rank  $n$ . Note - the theorem is saying, that the total weight given to those intervals gets smaller and smaller as  $n$  goes to infinity.

Now, we go over how to prove the theorem. It relies on rademacher variables:

$$r_n(\omega) = 2d_n(\omega) - 1$$

These are  $\pm 1$  when  $d_n = 1/0$ . Note, these have the same "being constant on dyadic intervals" properties as  $d_n(\omega)$ . We define:

$$s_n(\omega) = \sum_{i=1}^n r_i(\omega)$$

And so, our theorem is equivalent to proving:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \omega : \left| \frac{1}{n} s_n(\omega) \right| \geq \epsilon \right] = 0$$

Note, rademacher functions also have interpretations, probabilistically, of random walks and such. With these variables, we can ultimately find properties, going all the way to:

$$\int_0^1 s_n^2(\omega) d\omega = 1$$

However, what interests me is the following: Chebyshev's Lemma, but as a property of the real numbers. We have:

**Lemma - Chebyshev's Inequality** If  $f$  is a nonnegative step function, then  $[\omega : f(\omega) \geq \alpha]$  is for  $\alpha > 0$  a finite union of intervals, and:

$$\mathbb{P}[\omega : f(\omega) \geq \alpha] \leq \frac{1}{\alpha} \int_0^1 f(\omega) d\omega$$

Proof: Note, it is all just properties of step functions. Let  $c_j$  correspond to the step intervals  $(x_{j-1}, x_j]$ , and let  $\sum'$  be the sum over  $c_j \geq \alpha$ . Then, we have quite easily:

$$\int_0^1 f(\omega) d\omega = \sum c_j(x_j - x_{j-1}) \geq \sum' c_j(x_j - x_{j-1}) \geq \sum' \alpha(x_j - x_{j-1}) = \alpha \mathbb{P}[\omega : f(\omega) \geq \alpha]$$

Thus, we have Chebyshev's inequality, and with it, we can easily prove the Weak Law of Large Numbers. However - it is important to note - these are *properties over the real numbers*, as much as they are probabilistic properties.

## The Strong Law of Large Numbers

Just to first formalize some terms - the frequency of 1 in  $\omega$  is  $\sum_{i=1}^n d_i(\omega)$ , the relative frequency is that number normalized, ie  $\frac{1}{n} \sum_{i=1}^n d_i(\omega)$ , and the asymptotic relative frequency is the limit. We can derive, with some technical tools outside of discrete probability theory, results on the set:

$$N = \left[ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n d_i(\omega) = 1/2 \right]$$

We call this the set of normal numbers  $N$ . The tools themselves are the concepts of negligibility. A set  $A$  is negligible if for every  $\epsilon > 0$ , there is a countable number of intervals (not necessarily disjoint) such that:

$$A \subset \bigcup_k I_k \quad \sum_k I_k = \sum_k b_k - a_k < \epsilon$$

For one - I like to note here interpretations. Essentially - if  $A$  is negligible, it is a practical impossibility that  $\omega$  randomly drawn will lie within  $A$ . And if  $A^c$  is negligible, it is a practical certainty that  $\omega$  randomly drawn will lie within  $A$ . These are just how they should be understood - and these understandings are reasonable, as the total "length" that  $A$  takes up can be understood to be incredibly incredibly small.

Some properties of negligibility - note, these are the standard properties, stemming from infinite sums  $(1/2^k)$  summing to values less than  $\epsilon$ . Individual points are negligible, and so to thus are countable sets. So to are countable unions of countable sets.

With these properties - we understand that the property of our model not including  $\omega$  with a terminating sequence (all 0 ending) is not a short coming. These  $\omega$  form a countable set - and so, they can be considered negligible.

**Theorem 1.2** The set of normal numbers  $N$  has negligible complement.

**Proof** As an aside - we note that this proof is stronger than just the negligibility properties we noted above. This is because  $N^c$  is not countable. The set of  $d_i(\omega) = 1$  unless  $i$  is a multiple of 3 clearly belongs to  $N$  - as for each  $n$ ,  $n^{-1} \sum_{i=1}^n d_i(\omega) \geq 2/3$ . However, note this set is uncountable (diagonalization argument).

Note, the proof relies on equivalently defining  $N$  as:

$$N = \left[ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} s_n(\omega) = 0 \right]$$

Then, we can again make use of Chebyshev's Inequality (step function version) to find that:

$$\mathbb{P}[\omega : |s_n(\omega)| \geq n\epsilon] \leq \frac{1}{n^4 \epsilon^4} \int_0^1 s_n^4(\omega) d\omega = \frac{n + 3n(n-1)}{n^4 \epsilon^4} \leq \frac{3}{n^2 \epsilon^4}$$

Where the last step is just via an in depth (but simple) investigation of the integrals of multiplications of rademacher variables. With this property, we can find that if  $A_n = [\omega : |n^{-1} s_n(\omega)| \geq \epsilon_n]$ , then we have a sequence of  $\epsilon_n$  such that  $P(A_n) \leq 3\epsilon_n^{-4} n^{-2}$ , and we can find such a sequence such that:

$$\sum_n \mathbb{P}[A_n] < \infty$$

The final step to proving the theorem is noting that:

$$\bigcap_{n=m}^{\infty} A_n^c \subset N \implies N^c \subset \bigcup_{n=m}^{\infty} A_n$$

Which will ultimately prove the theorem. Note - a lot of details are left out, but I do not consider them important. You should be able to fill in. These are just the major strokes, outlining the proof. It essentially hinges on our integral value, and the relationship between  $A_n$  and the set of normal number  $N$ . qed.

So, we have  $N^c$  is negligible. But, can we have that  $N$  itself is negligible? Well, we could say no - using our "practically impossible" notions, and noting that for  $\omega \in [0, 1]$  randomly drawn, it must be in  $[0, 1]$ , and  $N^c \cup N = [0, 1]$ . But, that is not rigorous. And so, the following theorem will give us our initial basis of *measure*, and also help us note that  $N$  is not negligible.

### Theorem 1.3 - Lebesgue Measure Starting Point

1. If  $\bigcup_k I_k \subset I$ , and the  $I_k$  are disjoint, then  $\sum_k |I_k| \leq |I|$
2. If  $I \subset \bigcup I_k$  (the  $I_k$  need not be disjoint), then  $|I| \leq \sum_k |I_k|$
3. If  $\bigcup I_k = I$ , and the  $I_k$  are disjoint, then  $|I| = \sum_k |I_k|$

Note, this Theorem is true for countably infinite intervals as well. **Proof:** Note that the third part follows directly from (1) and (2). We start with the finite cases. For (1), we can prove by induction on the number of intervals  $n$ . It is clearly true for  $n = 1$ , and it is a fairly simple induction hypothesis to prove in general. We similarly have the same for (2).

The difficult part comes when going to infinite intervals. For (1), it is a simple limit, ie:

$$\sum_k |I_k| = \lim_{n \rightarrow \infty} \sum_{k=1}^n |I_k|$$

Note, each sum is less than  $|I|$ , as the finite case to 1 applies for each finite sum. And so, the inequality can be expanded to the limit. However - we can't do that for (2). Ultimately, the difference between the two cases is the inclusion of unions. We note:

$$\bigcup_k I_k \subset I \implies \bigcup_{k=1}^n I_k \subset I$$

Ie, the inclusion is true for every subset. However, we *do not* necessarily have:

$$I \subset \bigcup_k I_k \implies I \subset \bigcup_{k=1}^n I_k$$

Note that in the following way:  $I = (a, b]$ . We have that  $I_i = (a + 1/i, b]$ . We do indeed have that:

$$I \subset \bigcup_k I_k$$

As if you take  $x \in I$ ,  $a < x \leq b$ , and so we must have for  $i$  large enough,  $a + 1/i < x \leq b$ , and so  $x \in I_i$ . However, note that the inclusion is not actually true for a specific finite subunion. So, we need to take a different strategy to prove the infinite case. This comes from dealing with *open covers of compact spaces*, and relying on the Heine-Borel theorem, which says that intervals  $[a, b]$  are indeed compact. In this case, we are able to bridge between infinite unions and finite unions - as we can take a finite sub cover of an open cover on compact spaces. We prove the theorem essentially for  $[a + \epsilon, b]$ , that:

$$|I| - \epsilon = b - (a + \epsilon) \leq \sum_k |I_k| + \epsilon$$

However, as the  $\epsilon$  is arbitrary, we can conclude the fact for the infinite case as well. qed.

Note - this implies that  $N$  is not negligible. As it it was,  $[0, 1]$  would be negligible, but that is incorrect by the above, as any open covering must have total sum at least 1, ie, the total sum is not smaller than arbitrary  $\epsilon$ .

## The Measure Theory of Diophantine Approximation

This section is just additional, so my notes here are sparse. However, I do read through it, and record the theorems, plus some notes I have on them.

**Theorem 1.4** If  $x$  is irrational, there are infinitely many irreducible fractions  $p/q$  such that:

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^2}$$

Honestly, this proof is so good. I like it a lot - it is pretty clever. However, I don't just want to copy it down here - it is in the book. I'm not sure if there is any broad message I can glean from it - just that, it is a property of the real numbers. It just hinges on the following fact (which itself is pretty

difficult to prove), that for every  $Q$  positive integer, there is an integer  $q < Q$  and corresponding  $p$  such that:

$$\left| x - \frac{p}{q} \right| < \frac{1}{qQ} \leq \frac{1}{q^2}$$

Note, this is true for  $x$  rational or irrational. However, we have an infinite number of such irreducible fractions for the irrational case, and the contradiction derived in the book is nice as well. Anyway - read the book for this. qed.

Anyway, the above essentially means that, apart from a negligible set of  $x$ , each real number has an infinite set of irreducible rationals such that the bounds in Theorem 1.4 are true. We now consider a generalization - when can we tighten the inequality in Theorem 1.4 - Consider:

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^2 \varphi(q)}$$

Let  $A_\varphi$  consist of the real  $x$  for which the above has infinitely many irreducible solutions. Under what conditions on  $\varphi$  will  $A_\varphi$  have negligible complement? Note that if  $\varphi(q) < 1$ , then the condition is weaker than Theorem 1.4, and so  $A_\varphi$  has negligible complement immediately. It becomes interesting if  $\varphi(q) > 1$ . We will later prove the theorem:

**Theorem 1.5** Suppose that  $\varphi$  is positive and nondecreasing. If:

$$\sum_q \frac{1}{q\varphi(q)} = \infty$$

Then  $A_\varphi$  has negligible complement. We will prove this later, but we can now prove:

**Theorem 1.6** Suppose that  $\varphi$  is positive. If

$$\sum_q \frac{1}{q\varphi(q)} < \infty$$

Then  $A_\varphi$  is negligible.

We will go over the proof soon for this theorem. However - just note what the theorems are saying. Note that in the second -  $\varphi(q)$  must be growing quite quickly. We need the denominator to be quite large, so that the infinite sum



is ultimately finite. However, in theorem 1.5, we don't want the  $\varphi(q)$  to be too large, lest the sum actually does become finite. Ultimately - both theorems are conditions on how  $\varphi(q)$  grows. Which, ultimately does make sense. If  $\varphi(q)$  grows to large - it becomes unreasonable to expect our condition to hold infinitely many times. If I ever encounter such situations, where I might want to examine the growth of a function  $\varphi$  - I think examining whether the infinite sum of  $1/\varphi$  equals infinity or not is often a good property that is related to the growth of a function.

**Proof of Theorem 1.6** I'll give the full proof here, as it is interesting to me, and rather short. We want to show that  $A_\varphi$  is negligible. Well, given that the sum is finite, there is a  $q_0$  large enough such that the tail sum  $\sum_{q \geq q_0} \frac{1}{q\varphi(q)} < \epsilon/4$ . If  $x \in A_\varphi$ , then our definition holds for some  $q \geq q_0$ , and as  $0 < x < 1$ , we have that the corresponding  $p$  lies in  $0 \leq p \leq q$ . Thus, we have that:

$$A_\varphi \subset \bigcup_{q \geq q_0} \bigcup_{p=0}^q \left( \frac{p}{q} - \frac{1}{q^2\varphi(q)}, \frac{p}{q} + \frac{1}{q^2\varphi(q)} \right]$$

Which stems from every  $x \in A_\varphi$  being in the right expression, given that  $x$  is within one of the intervals on the right by the property we just described. Now, we have a covering interval - we just need to find the length of it. Note, by assumption, we have all of our  $q$  must satisfy  $q \geq 1$  (or, we can add that in). And so, we have the sum of the intervals is:

$$\sum_{q \geq q_0} \sum_{p=0}^q \frac{2}{q^2\varphi(q)} = \sum_{q \geq q_0} \frac{2(q+1)}{q^2\varphi(q)} \leq \sum_{q \geq q_0} \frac{2(q+q)}{q^2\varphi(q)} \leq \sum_{q \geq q_0} \frac{4}{q\varphi(q)} < \epsilon$$

And thus,  $A_\varphi$  is negligible. qed.

## Problems

### 1.1

1. As for why the existence of an infinite sequence of independent events each with probability  $1/2$  in a discrete probability space would make the section superfluous - I think this is because, in the section, we *rely* on the uncountability of the real numbers. This allows us to make notions like negligible, which helps us make Borel's Number Theorem (the Strong Law of Large numbers). We could then just handle infinite cases with a countable, discrete space - which would make the section unnecessarily in depth ("superfluous").

As for why a discrete space cannot have an infinite sequence of independent events. Note, we can partition the space into sets  $A_1 \cap A_2$ ,  $A_1 \cap A_2^c$ ,  $A_1^c \cap A_2$ , and  $A_1^c \cap A_2^c$ . Note that each has probability  $2^{-2}$ , by independence. And so, each countable point  $\omega$  belongs to one of these sets, and  $P(\omega) \leq 2^{-2}$ . We can continue on for arbitrary  $2^k$  partitions, each of probability at most  $2^{-k}$ . Thus, we find that  $P(\omega) = 0$  for each point. This is a contradiction, as:

$$\sum P(\omega) = 1 \neq 0 = \sum_{\omega} 0 = \sum P(\omega)$$

2. This portion draws the same contradiction as above, namely,  $P(\omega) = 0$  for all  $\omega$ . Each  $\omega$  belongs to a sequence of  $A_1, A_2, A_3^c$ , something like that. Let  $t_i = p_i/1 - p_i$  that corresponds to  $\omega \in A_i$  or  $\omega \in A_i^c$ . We find:

$$P(\omega) \leq \prod_{i=1}^n t_i \leq \exp \left( - \sum_{i=1}^n (1 - t_i) \right)$$

Where the second step notes a property for  $t_i \in [0, 1]$ . Note, it is clear that the above is bounded by:

$$\leq \exp \left( - \sum_{i=1}^n \alpha_i \right)$$

If  $\sum_n \alpha_n$  diverges, the above goes to 0, and we can conclude:

$$P(\omega) = 0$$

Which again, draws out our contradiction. qed.