

# STATS 310A - Lecture 2

Persi Diaconis  
Scribed by Michael Howes

09/23/21

## 1 Strong law of large numbers

Recall that we have  $\Omega = (0, 1]$  and for  $\omega \in \Omega$  we write  $\omega = 0.d_1(\omega)d_2(\omega)\dots$  where  $d_i(\omega)$  is the  $i^{th}$  binary digit of  $\Omega$ . If  $\omega$  has two binary expansions, we pick the expansion that ends in all 1's. We also defined  $r_i = 2d_i - 1 \in \{-1, 1\}$  and  $S_n = \sum_{i=1}^n r_i$ . Let  $B \subseteq \Omega$  be the subset

$$B = \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} |S_n(\omega)/n| = 0 \right\}.$$

Recall also that a subset  $A \subseteq \Omega$  is said to be negligible if for every  $\varepsilon > 0$ , there exists a countable collection of intervals  $\{(a_i, b_i]\}_{i=1}^\infty$  such that  $A \subseteq \cup_i (a_i, b_i]$  and  $\sum_i b_i - a_i = \sum_i p((a_i, b_i]) < \varepsilon$ . The strong law of large numbers (slln) states that  $B^c$  is negligible.

*Proof.* Fix  $\delta > 0$  and note that  $\{|S_n/n| > \delta\} = \{S_n^4 > \delta^4 n^4\}$ . By Markov's inequality we can thus conclude that

$$p(\{|S_n/n| > \delta\}) \leq \frac{1}{\delta^4 n^4} \int_0^1 S_n(\omega)^4 d\omega.$$

Since  $S_n = \sum_{i=1}^n r_i$  we know that  $(S_n)^4 = \sum_{i,j,k,l=1}^n r_i r_j r_k r_l$ . Note that there are five possibilities for the term  $r_i r_j r_k r_l$ . These are

- (a) The case  $r_i^4$ . This case occurs  $n$  times and when it occurs  $\int_0^1 r_i(\omega) r_j(\omega) r_k(\omega) r_l(\omega) d\omega = \int_0^1 1 d\omega = 1$ .
- (b) The case  $r_i^2 r_j r_k$  when  $i, j, k$  are distinct. In this case  $\int_0^1 r_i r_j r_k r_l d\omega = \int_0^1 r_j r_k d\omega = 0$ .
- (c) The case  $r_i^2 r_j^2$  when  $i \neq j$ . This case occurs  $3n(n-1)$  times and in this case  $\int_0^1 r_i r_j r_k r_l d\omega = \int_0^1 1 d\omega = 1$ .
- (d) The case  $r_i^3 r_j$  where  $i \neq j$ . In this case  $r_i r_j r_k r_l = r_i r_j$  and thus  $\int_0^1 r_i r_j r_k r_l d\omega = 0$ .
- (e) The case  $r_i r_j r_k r_l$  and  $i, j, k, l$  are all distinct. In this case  $\int_0^1 r_i r_j r_k r_l d\omega = 0$ .

Combining all of the above we have

$$\int_0^1 S_n(\omega)^4 d\omega = \sum_{i,j,k,l=1}^n \int_0^1 r_i(\omega) r_j(\omega) r_k(\omega) r_l(\omega) d\omega = n + 3n(n-1) \leq 3n^2.$$

Thus we have  $p(\{|S_n/n| > \delta\}) \leq \frac{3}{\delta^4 n^2}$ . Now set  $\delta_n = \frac{1}{n^{1/8}}$  so that  $\delta_n \rightarrow 0$  and  $\sum_{n=1}^\infty \frac{3}{\delta_n^4 n^2} < \infty$ . Note that for all  $m \in \mathbb{N}$ , we have

$$\bigcap_{n=m}^\infty \{\omega : |S_n(\omega)/n| \leq \delta_n\} \subseteq B.$$

By taking complements we see

$$B^C \subseteq \bigcup_{n=m}^{\infty} \{\omega : |S_n(\omega)/n| > \delta_n\}.$$

The set  $\{\omega : |S_n(\omega)/n| > \delta_n\}$  can be written as a finite union of disjoint intervals  $\cup_{i=1}^{k_n} I_{n,i}$  such that  $\sum_{i=1}^{k_n} p(I_{n,i}) < \frac{3}{\delta_n^4 n^2}$ . Thus we have

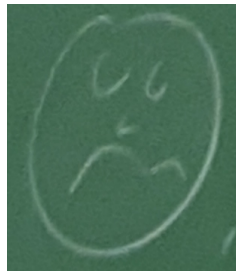
$$B^C \subseteq \bigcup_{n=m}^{\infty} \bigcup_{i=1}^{k_n} I_{n,i},$$

for all  $m \in \mathbb{N}$ . Thus given  $\varepsilon > 0$  we can choose  $m$  such that  $\sum_{n=m}^{\infty} \frac{3}{\delta_n^4 n^2} < \varepsilon$  and note that

$$\sum_{n=m}^{\infty} \sum_{i=1}^{k_n} p(I_{n,i}) \leq \sum_{n=m}^{\infty} \frac{3}{\delta_n^4 n^2} < \varepsilon.$$

Showing that  $B^C$  is negligible, □

Thus we can say that  $\lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} = 0$  for *almost all*  $\omega \in \Omega$  or simply that  $\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0$  *almost surely*. But!



This statement does not have any qualitative bounds to it. Also how well does our model actually reflect coin flipping? A true model would have a lot of physics and observations of how people flip coins. Persi has written papers about such things. One is “Dynamic bias in the coin toss” written with collaborators.

## 2 Assigning Probabilities

See “Ten Great Ideas About Chance” by Persi.

**Definition 1.** Let  $\Omega$  be a set. A collection  $\mathcal{F}_0$  of subset of  $\Omega$  is a *field* is

- (a) The set  $\Omega$  is in  $\mathcal{F}_0$ .
- (b) If  $A \in \mathcal{F}_0$ , then  $A^c \in \mathcal{F}_0$ .
- (c) If  $A, B \in \mathcal{F}_0$ , then  $A \cup B \in \mathcal{F}_0$ .

We say that  $\mathcal{F}_0$  is closed under compliments and finite unions.

**Definition 2.** Let  $\Omega$  be a set and  $\mathcal{F}$  a collection of subsets of  $\Omega$ . The collection  $\mathcal{F}$  is a  $\sigma$ -field if it is a field and closed under countable unions. That is if  $A_1, A_2, \dots \in \mathcal{F}$ , then  $\cup_i A_i \in \mathcal{F}$ .

**Examples 1.** (a)  $\Omega = (0, 1]$  and  $\mathcal{F}_0 = \{\text{finite unions of disjoint intervals of the form } (a, b]\}$ . This is a field since  $(a, b]^c = (0, a] \cup (b, 1]$ . It is not a  $\sigma$ -field since  $(0, 1/2) = \cup_{i=1}^{\infty} (0, 1/2 - 1/i] \notin \mathcal{F}_0$ .

- (b) The collection of all subsets of  $\Omega$  is a  $\sigma$ -field.
- (c)  $\{\Omega, \emptyset\}$  is a  $\sigma$ -field.
- (d) If  $\mathcal{F}_i$  are  $\sigma$ -fields on  $\Omega$  for all  $i \in I$ , then

$$\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i,$$

is a  $\sigma$ -field.

- (e) If  $\mathcal{C}$  is any collection of subsets of  $\Omega$ , then define

$$\sigma(\mathcal{C}) := \bigcap_i \mathcal{F}_i,$$

where the intersection is of all  $\sigma$ -fields on  $\Omega$  that contain  $\mathcal{C}$ . This is a  $\sigma$ -field and is called the  $\sigma$ -field generated by  $\mathcal{C}$ .

- (f) The Borel set (in  $(0, 1]$ ) is the  $\sigma$ -field generated by all the intervals in  $(0, 1]$ . For example our set

$$B = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{|S_n/n| < 1/k\},$$

is a Borel set.

**Definition 3.** Let  $\Omega$  be a set and  $\mathcal{F}$  a field. The pair  $(\Omega, \mathcal{F})$  is called a measurable space. A probability on  $(\Omega, \mathcal{F})$  is a function

$$P : \mathcal{F} \rightarrow [0, 1], \quad A \mapsto P(A),$$

such that

- (a)  $P(\Omega) = 1, P(\emptyset) = 0$ .
- (b)  $P(A^C) = 1 - P(A)$ .
- (c) If  $\{A_i\}_{i=1}^{\infty}$  are pairwise disjoint and  $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$ , then  $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ .

**Remark 1.** It is often “easy” to assign probabilities to a field  $\mathcal{F}_0$  of sets and there is a standard way to uniquely extend these probabilities to the  $\sigma$ -field generated by  $\mathcal{F}_0$ .

**Definition 4.** Let  $P$  be a probability on a field  $\mathcal{F}_0$ . For all  $A \subseteq \Omega$ , define

$$P^*[A] := \inf \left\{ \sum_{i=1}^{\infty} P(B_i) : A \subseteq \bigcup_{i=1}^{\infty} B_i \text{ and } B_i \in \mathcal{F}_0 \right\}.$$

The function  $P^*$  is called the *outer measure associated to*  $(\Omega, \mathcal{F}, P)$ .

The function  $P^*$  has the following properties

- (a)  $P^*(\Omega) = 1, P^*(\emptyset) = 0$ .
- (b)  $P^*$  is countable subadditive. That is if  $A = \cup_{i=1}^{\infty} A_i$ , then  $P^*[A] \leq \sum_{i=1}^{\infty} P^*[A_i]$ .

*Proof.* Fix  $\varepsilon > 0$  and let  $(B_{i,j})_{j=1}^{\infty}$  be a countable cover of  $A_i$  by intervals such that

$$\sum_{j=1}^{\infty} P[B_{i,j}] \leq P^*[A_i] + \varepsilon 2^{-i}$$

. Thus  $\{B_{i,j}\}_{i,j=1}^{\infty}$  is a countable cover of  $A$  by elements of  $\mathcal{F}_0$ . Since  $P[B_{i,j}] \geq 0$ , we can rearrange the infinite series  $\sum_{i,j=1}^{\infty} P[B_{i,j}]$  and thus conclude the following

$$\begin{aligned} P^*[A] &\leq \sum_{i,j=1}^{\infty} P[B_{i,j}] \\ &= \sum_i \sum_{j=1}^{\infty} P[B_{i,j}] \\ &\leq \sum_{i=1}^{\infty} P^*[A_i] + \varepsilon \sum_{i=1}^{\infty} 2^{-i} \\ &= \sum_{i=1}^{\infty} P^*[A_i] + \varepsilon \\ &= \sum_{i=1}^{\infty} P^*[A_i] + \varepsilon. \end{aligned}$$

Since  $\varepsilon > 0$  was arbitrary, we can conclude that  $P^*[A] \leq \sum_{i=1}^{\infty} P^*[A_i]$ .  $\square$

**Definition 5.** [Caratheodory] Let  $P$  be a probability measure on a field  $\mathcal{F}_0$ . A set  $A \in \Omega$  is *measurable* if for all  $E \subseteq \Omega$ ,

$$P^*[E] = P^*[E \cap A] + P^*[E \cap A^c].$$

We will use  $\mathcal{M}$  to denote the set of all measurable sets.

We will prove:

- $\mathcal{M}$  is a  $\sigma$ -field that contains  $\mathcal{F}_0$ .
- $P^*$  restricted to  $\mathcal{M}$  is a probability (in particular  $P^*$  is countable additive on measurable sets).
- $P^*$  restricted to  $\mathcal{F}_0$  equals  $P$ .
- $P^*$  is unique.

This goal will preoccupy us for the next lecture but we will start working on it today. With the notation as above we will show that  $\mathcal{M}$  is a field.

*Proof.* The one main trick we will use is that to show  $A \in \mathcal{M}$ , it is enough to show

$$P^*(E \cap A) + P^*(E \cap A^c) \leq P^*(E),$$

for all  $E \subseteq \Omega$ . This is because we always have  $P^*(E) \leq P^*(E \cap A) + P^*(E \cap A^c)$  by subadditivity.

We can see immediately that  $\Omega \in \mathcal{M}$  and that  $\mathcal{M}$  is closed under complements. Thus suppose that  $A, B \in \mathcal{M}$  and  $E \subseteq \Omega$ , then

$$\begin{aligned} P^*(E) &= P^*(E \cap A) + P^*(E \cap A^c) \\ &= P^*(E \cap A \cap B) + P^*(E \cap A \cap B^c) + P^*(E \cap A^c \cap B) + P^*(E \cap A^c \cap B^c) \\ &\geq P^*(E \cap ([A \cap B] \cup [A \cap B^c] \cup [A^c \cap B])) + P^*(E \cap [A^c \cap B^c]) \\ &= P^*(E \cap [A \cup B]) + P^*(E \cap [A \cup B]^c). \end{aligned}$$

Thus, by the main trick,  $A \cup B \in \mathcal{M}$ .  $\square$

### 3 Conclusion

One might ask why are we doing all this just to talk about probabilities? There are several reasons

- People want to work with infinite sequence spaces, random curves, Brownian motion and the set of all probabilities measures on  $[0, 1]$ . These are complicated spaces and it can be hard to assign probabilities on them by hand.
- We simply cannot assign a consistent notion of length to all subsets of  $[0, 1]$ .
- Keep an eye out for a halloween talk on non-measurable set.

A remark on finite vs countable additivity. Let  $\mathbb{N} = \{1, 2, \dots\}$ . One would like to say that  $j$  chosen “at random” from  $\mathbb{N}$  has a 50% chance of being even. We can make sense of this by defining

$$P_n(A) = \frac{|A \cap [n]|}{n},$$

where  $|B|$  is the number of elements in  $B$  and  $[n] = \{1, 2, \dots, n\}$ . If  $\lim_{n \rightarrow \infty} P_n(A) = l$  exists we say that  $A$  has density  $D(A) = l$ . One can show

- $D(\text{multiples of } j) = \frac{1}{j}$ ,
- $D(\{\text{primes}\}) = 0$ , and
- $D(\{\text{square free numbers}\}) = \frac{6}{\pi^2}$ .

Not every set has a density. For example if

$$A = \{1, 10, 11, 100, 101, \dots, 199, 1000, 10001, \dots, 1999, 10000\}.$$

That is,  $A$  is the set of numbers that start with a 1 when written in decimal. Then  $P_n(A)$  moves up and down between  $1/9$  and  $5/9$  infinitely often. The answer to this problem is to use the Hahn-Banach theorem to extend  $D$  to all subsets of  $\mathbb{N}$ . This gives us a measure that is finitely additive but not countably additive. This example is highly non-constructive and non unique since there are many possible Hahn Banach extensions and none of them are “natural.”