

# STATS310A - Lecture 5

Persi Diaconis  
Scribed by Michael Howes

10/05/21

## Contents

<b>1 Applying Borel Cantelli</b>	<b>1</b>
<b>2 Homework</b>	<b>2</b>
<b>3 Measure Theory</b>	<b>2</b>
<b>4 Distribution functions</b>	<b>4</b>

## 1 Applying Borel Cantelli

Recall,  $l_n(\omega)$  = length of the head run starting at  $n$ . That is  $l_n(\omega) = k$  if and only if  $d_n(\omega) = d_{n+1}(\omega) = \dots = d_{n+k-1}(\omega) = 1$  and  $d_{n+k} = 0$ . Last time we saw

$$P(\omega : l_n(\omega) > (1 + \varepsilon) \log_2(n) \text{ i.o.}) = 0.$$

We will now see

$$P(w : l_n(w) > \log_2(n) \text{ i.o.}) = 1.$$

The problem is that the events  $\{l_n \geq r_n\}$  are not independent. The trick is to use subsequences.

**Proposition 1.** *Let  $r_n$  be a weakly increasing subsequence such that  $r_n > 0$  and  $\sum_{n=1}^{\infty} \frac{2^{-r_n}}{r_n} = \infty$  (for example  $r_n = \log_2(n)$ ). Then  $P(l_n \geq r_n \text{ i.o.}) = 1$ .*

*Proof.* Define a sequence  $n_k$  by  $n_1 = 1$  and  $n_{k+1} = n_k + r_{n_k}$  so that  $n_{k+1} - n_k = r_{n_k}$ . Define  $A_k = \{l_{n_k} \geq r_{n_k}\}$ . Then  $A_k = \{\omega : d_i(\omega) = 1 \text{ for } n_k \leq i \leq n_{k+1} - 1\}$  since  $n_k + r_{n_k} = n_{k+1}$ . Thus the events  $\{A_k\}_{k=1}^{\infty}$  are independent. Note also that  $P(A_k) = \frac{1}{2^{r_{n_k}}}$ . The second Borel Cantelli theorem tells us that if

$$\sum_{k=1}^{\infty} P(A_k) = \infty,$$

then  $P(A_k \text{ i.o.}) = 1$ . Note that

$$\begin{aligned}
 \sum_{k=1}^{\infty} P(A_k) &= \sum_{k=1}^{\infty} 2^{-r_{n_k}} \\
 &= \sum_{k=1}^{\infty} 2^{-r_{n_k}} \frac{n_{k+1} - n_k}{r_{n_k}} \\
 &= \sum_{k=1}^{\infty} \sum_{n=n_k}^{n_{k+1}-1} \frac{2^{-r_{n_k}}}{r_{n_k}} \\
 &\geq \sum_{k=1}^{\infty} \sum_{n=n_k}^{n_{k+1}-1} \frac{2^{-r_n}}{r_n} \\
 &= \sum_{n=1}^{\infty} \frac{2^{-r_n}}{r_n} \\
 &= \infty.
 \end{aligned}$$

□

## 2 Homework

Read sections 10, 11 and 12. Do problems 10.3, 10.4, 11.2, 14.5 and two more.

## 3 Measure Theory

Let  $\Omega$  be a set and  $\mathcal{F}$  an algebra of subsets of  $\Omega$ .

**Definition 1.** A *measure* on  $(\Omega, \mathcal{F})$  is a function  $\mu : \mathcal{F} \rightarrow [0, \infty]$  such that

- (a) (Non-trivial)  $\mu(\emptyset) = 0$ .
- (b) (Monotone) If  $A, B \in \mathcal{F}$  and  $A \subseteq B$ , then  $\mu(A) \leq \mu(B)$ .
- (c) (Countable additivity) If  $\{A_i\}_{i=1}^{\infty}$  is a countable collection of disjoint subsets in  $\mathcal{F}$  and  $\bigcup_i A_i \in \mathcal{F}$ , then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

For example with  $\Omega = \mathbb{R}$  and  $\mathcal{F}$  equal to the collection of all subsets, then  $\mu(\emptyset) = 0$  and  $\mu(A) = \infty$  otherwise is a measure. Likewise  $\mu(A) = \#$  of points in  $A$ , is also a measure of  $(\Omega, \mathcal{F})$ .

One might ask why we are doing this?

- (a) We want to define probabilities via densities, that is we want to write

$$P(A) = \int_A f(\omega) d\mu(\omega),$$

where  $\mu$  is a measure that is not necessarily a probability measure.

- (b) It's free from the hard work we've already done.

A word of caution! For probabilities we had if  $A \subseteq B$ , then  $P(B \setminus A) = P(B) - P(A)$ . This is not true in general. For measures we have  $P(B) = P(A) = \infty$ . For example if  $\mu$  is Lebesgue measure,  $B = (0, \infty)$  and  $A = (x, \infty)$  for some  $x \geq 0$ , then  $\mu(B \setminus A) = \mu((0, x]) = x$  but  $\mu(B) = \mu(A) = \infty$ .

**Definition 2.** A measure  $\mu$  on  $(\Omega, \mathcal{F})$  is  $\sigma$ -finite if there exists a countable collection  $B_i \in \mathcal{F}$  s.t.  $\bigcup_i B_i = \Omega$  and  $\mu(B_i) < \infty$  for all  $i$ .

For example Lebesgue measure is  $\sigma$ -finite since we can set  $B_i = (-i, i)$ . The measure  $\mu(A) = \#$  of points in  $A$  is not  $\sigma$ -finite on  $\mathbb{R}$ .

**Theorem 1.** Suppose that we have two measure  $\mu_1, \mu_2$  on  $\sigma(\mathcal{P})$  where  $\mathcal{P}$  is a  $\pi$ -system. If  $\mu_1$  and  $\mu_2$  agree on  $\mathcal{P}$  and there exists a countable collection  $B_i \in \mathcal{P}$  such that  $\mu_1(B_i) = \mu_2(B_i) < \infty$  and  $\Omega = \bigcup_i B_i$ , then  $\mu_1$  and  $\mu_2$  agree on  $\sigma(\mathcal{P})$ .

*Proof.* For  $B \in \mathcal{P}$  with  $\mu_1(B) < \infty$ , define  $L_B = \{A \in \sigma(\mathcal{P}) : \mu_1(A \cap B) = \mu_2(A \cap B)\}$ . Then  $L_B$  is a  $\lambda$ -system and it contains  $\mathcal{P}$  and thus by the  $\pi$ - $\lambda$  theorem  $\sigma(\mathcal{P}) = L_B$ .

By assumption we have  $\Omega = \bigcup_i B_i$  where  $B_i \in \mathcal{P}$  and  $\mu_1(B_i) < \infty$ . Fix  $A \in \sigma(\mathcal{P})$  and for  $i = 1, 2$ , note that, by the inclusion exclusion formula

$$\mu_i \left( \bigcup_{j=1}^n A \cap B_j \right) = \sum_{j=1}^n \mu_i(A \cap B_j) - \sum_{1 \leq j < k \leq n} \mu_i(A \cap B_j \cap B_k) + \dots$$

Furthermore, finite intersections of  $\{B_i\}_{i=1}^\infty$  are in  $\mathcal{P}$  and have finite measure. Thus since  $L_B = \sigma(\mathcal{P})$  for all  $B \in \mathcal{P}$  with finite measure we have  $\mu_1(A \cap B) = \mu_2(A \cap B)$  where  $B$  is any finite intersection of  $\{B_i\}_{i=1}^\infty$ . Thus

$$\mu_1 \left( \bigcup_{j=1}^n A \cap B_i \right) = \mu_2 \left( \bigcup_{j=1}^n A \cap B_i \right),$$

for all  $n$ . Letting  $n \rightarrow \infty$  we see that  $\mu_1(A) = \mu_2(A)$ .  $\square$

Note that  $\sigma$ -finiteness is needed. If  $\mu_1(A) = \#$  of points in  $A$  and  $\mu_2(A) = \infty$  if  $A \neq \emptyset$ , then  $\mu_1 = \mu_2$  on all intervals but  $\mu_1 \neq \mu_2$  on the Borel  $\sigma$ -algebra.

**Definition 3.** Let  $\Omega$  be a set. An outer measure  $\mu^*$  is a function defined on all subsets of  $\Omega$  such that

- (a)  $\mu^*(A) \in [0, \infty]$
- (b)  $\mu^*(\emptyset) = 0$ ,
- (c)  $\mu^*(A) \leq \mu^*(B)$  if  $A \subseteq B$ , and
- (d)  $\mu^*(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mu^*(A_i)$ .

For example let  $\Omega$  be any set and let  $\mathcal{A}$  be any collection of subsets of  $\Omega$  and let  $\rho : \mathcal{A} \rightarrow [0, \infty]$  be any function such that  $\rho(\emptyset) = 0$ . Then

$$\mu^*(A) = \inf \left\{ \sum_{i=1}^\infty \rho(A_i) : A \subseteq \bigcup_{i=1}^\infty A_i, A_i \in \mathcal{A} \right\},$$

where  $\inf \emptyset = \infty$  is an outer measure. This can be seen by using the usual  $\varepsilon 2^{-i}$  trick to prove countable subadditivity.

A special case of this is Hausdorff measure when  $M \subseteq \mathbb{R}^n$  is a crinkly manifold or a fractal or something similar. We define  $\mathcal{A} = \{B_\varepsilon(x) : x \in \mathbb{R}^n, \varepsilon > 0\}$  where  $B_\varepsilon(x)$  is a ball of radius  $\varepsilon$  centred at  $x$ . We then define  $\rho_{n,\gamma}(B_\varepsilon(x)) = (2\varepsilon)^\gamma$  and call the resulting  $\mu_{n,\gamma}^*$  the  $\gamma$ -Hausdorff measure on  $\mathbb{R}^n$ .

**Theorem 2.** Let  $\mu^*$  be an outer measure on all subsets of  $\Omega$  and define

$$\mathcal{M} = \{A \subseteq \Omega : \mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E) \text{ for all } E \subseteq \Omega\},$$

then  $\mathcal{M}$  is a  $\sigma$ -algebra and  $\mu$  restricted to  $\mathcal{M}$  is a measure.

*Proof.* Symbol for symbol, letter for letter, the same as the proof of the Caratheodory extension theorem.  $\square$

## 4 Distribution functions

“People at microsoft research” were putting probabilities on  $\mathbb{R}^n$  using distribution functions in a funny way.

If  $\mathbb{P}$  is a probability measure on  $\mathbb{R}^n$  we can define a function  $F : \mathbb{R}^n \rightarrow [0, 1]$  by  $F(x) = \mathbb{P}(A_x)$  where  $A_x = \{y : y_i \leq x_i, \text{ for } i = 1, \dots, n\}$ . We call  $F(x)$  the distribution function (DF) generated by  $\mathbb{P}$ . A distribution function satisfies the following

- $\lim_{x \rightarrow +\infty} F(x) = 1.$
- $\lim_{x \rightarrow -\infty} F(x) = 0.$
- $F$  is monotone in each coordinate.
- $F$  is right continuous. If  $x_n \searrow x$ , then  $F(x_n) \searrow F(x).$

But not every  $F$  that satisfies the above is a DF. At microsoft they had the idea to take define on  $\mathbb{R}^3$

$$MS(x, y, z) = F(x, y, z)G(x, y)H(x),$$

where  $F, G$  and  $H$  were all distribution functions. One can ask is  $MS$  a distribution function? A function  $F : \mathbb{R}^2 \rightarrow [0, 1]$  is a distribution function if and only if

- (a)  $\lim_{x \rightarrow +\infty} F(x) = 1.$
- (b)  $\lim_{x \rightarrow -\infty} F(x) = 0.$
- (c)  $F$  is monotone in each coordinate.
- (d)  $F$  is right continuous. If  $x_n \searrow x$ , then  $F(x_n) \searrow F(x).$
- (e) For all  $(x_1, y_1), (x_2, y_2)$  such that  $x_1 \leq x_2$  and  $y_1 \leq y_2$ , we have

$$F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0.$$

This property requires that the probability of every rectangle is non-negative.

Furthermore, if  $F$  satisfies (a)-(e), then there is a unique probability measure  $\mathbb{P}$  on  $\mathbb{R}^2$  such that  $\mathbb{P}(A_{(x,y)}) = F(x, y).$