

# 1 Measures

Let  $E$  be any set. A collection  $\mathcal{E}$  of subsets of  $E$  is called a  $\sigma$ -algebra if the following holds:

1.  $\emptyset \in \mathcal{E}$ .
2. If  $A \in \mathcal{E}$ , then  $A^c = E \setminus A \in \mathcal{E}$ .
3. If  $(A_n : n \in \mathbb{N})$ ,  $A_n \in \mathcal{E}$ , then  $\bigcup_n A_n \in \mathcal{E}$ .

**Examples.**

- $\mathcal{E} = \{\emptyset, E\}$
- $\mathcal{E} = \mathcal{P}(E)$ , the set of all subsets of  $E$ .

Note that  $\bigcap_n A_n = (\bigcup_n A_n^c)^c$ , any  $\sigma$ -algebra  $\mathcal{E}$  is also closed under countable intersection of its elements. Also  $B \setminus A = B \cap A^c \in \mathcal{E}$  whenever  $A, B \in \mathcal{E}$ .

Any set  $E$  with a choice of  $\sigma$ -algebra  $\mathcal{E}$  is called a *measurable space*, and the elements of  $\mathcal{E}$  are called *measurable sets*.

A *measure*  $\mu$  is a set-function  $\mu : \mathcal{E} \rightarrow [0, \infty]$  such that  $\mu(\emptyset) = 0$ , and for any  $(A_n : n \in \mathbb{N})$ ,  $A_n \in \mathcal{E}$  pairwise disjoint ( $A_n \cap A_m = \emptyset$  for all  $n \neq m$ ) then

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n) \quad (\text{countable additivity of } \mu)$$

If  $\mathcal{E}$  is countable, then for any  $A \in \mathcal{P}(E)$  and a measure  $\mu$

$$\mu(A) = \mu\left(\bigcup_{x \in A} \{x\}\right) = \sum_{x \in A} \mu(\{x\})$$

hence there is a one-to-one correspondence between measures and mass functions on  $E$ .

For any collection  $\mathcal{A}$  of subsets of  $E$ , we define the  $\sigma$ -algebra  $\sigma(\mathcal{A})$  generated by  $\mathcal{A}$  as

$$\sigma(\mathcal{A}) = \{A \subseteq E : A \in \mathcal{E} \text{ } \forall \sigma\text{-algebras } \mathcal{E} \supseteq \mathcal{A}\}$$

which equals (Example sheet)

$$\sigma(\mathcal{A}) = \bigcap_{\sigma\text{-algebras } \mathcal{E} \supseteq \mathcal{A}} \mathcal{E}$$

To construct good ‘generators’ we define

1.  $\mathcal{A}$  is called a *ring over  $E$*  if  $\emptyset \in \mathcal{A}$ ; if  $A, B \in \mathcal{A}$  then  $B \setminus A \in \mathcal{A}$  and  $A \cup B \in \mathcal{A}$ .

2.  $\mathcal{A}$  is called an *algebra over  $E$*  if  $\emptyset \in \mathcal{A}$ ; if  $A, B \in \mathcal{A}$  then  $A^c \in \mathcal{A}$ ,  $A \cup B \in \mathcal{A}$ .

Notice that in a ring  $A \Delta B = (B \setminus A) \cup (A \setminus B) \in \mathcal{A}$  and  $A \cap B = (A \cup B) \setminus (A \Delta B) \in \mathcal{A}$ . Also,  $B \setminus A = B \cap A^c = (B^c \cup A)^c \in \mathcal{A}$ , so an algebra is a ring.

**Fact:** If  $\bigcup_n A_n$ ,  $A_n \in \mathcal{E}$ ,  $\mathcal{E}$  some  $\sigma$ -algebra (or a ring if the union is finite) - then we can find  $B_n \in \mathcal{E}$  disjoint such that  $\bigcup_n A_n = \bigcup_n B_n$ . Indeed, define  $\tilde{A}_n = \bigcup_{j \leq n} A_j$ , and set  $B_n = \tilde{A}_n \setminus \tilde{A}_{n-1}$ , then the fact follows. [“disjointification of countable unions”]

**Definition.** A *set function* on any collection  $\mathcal{A}$  of subsets of  $E$  (where  $\emptyset \in \mathcal{A}$ ) is a map  $\mu : \mathcal{A} \rightarrow [0, \infty]$  such that  $\mu(\emptyset) = 0$ . We say  $\mu$  is

1. *increasing* if  $\mu(A) \leq \mu(B)$  whenever  $A \subseteq B$ ;  $A, B \in \mathcal{A}$
2. *additive* if  $\mu(A \cup B) = \mu(A) + \mu(B)$  whenever  $A, B \in \mathcal{A}$ ;  $A \cup B \in \mathcal{A}$  and  $A \cap B = \emptyset$ .
3. *countably additive* if  $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$  for any  $(A_n : n \in \mathbb{N})$  where  $A_n \in \mathcal{A}$  disjoint and  $\bigcup_n A_n \in \mathcal{A}$ .
4. *countably sub-additive* if  $\mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$  for all  $(A_n : n \in \mathbb{N})$  such that  $\bigcup_n A_n \in \mathcal{A}$

**Remark:** one can show that a measure  $\mu$  on a  $\sigma$ -algebra satisfies 1-4 above.

**Theorem** (Caratheodory). *Let  $\mu$  be a countably additive set function on a ring  $\mathcal{A}$  of subsets of  $E$ . Then there exists a measure  $\mu^*$  on  $\sigma(\mathcal{A})$  such that  $\mu^*|_{\mathcal{A}} = \mu$ .*

*Proof.* For  $B \subseteq E$  define the *outer measure*  $\mu^*$  as

$$\mu^*(B) = \inf \left\{ \sum_{n \in \mathbb{N}} \mu(A_n) : A_n \in \mathcal{A}, B \subseteq \bigcup_n A_n \right\}$$

and set  $\mu^*(B) = \infty$  if the set within the infimum is empty.

Define

$$\mathcal{M} = \{A \subseteq E : \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) \forall B \subseteq E\}$$

the “ $\mu^*$ -measurable” sets.

Step 1:  $\mu^*$  is countably sub-additive on  $\mathcal{P}(E)$ . For any  $B \subseteq E$  and  $B_n \subseteq E$  such that  $B \subseteq \bigcup_n B_n$  we have

$$\mu^*(B) \leq \sum_n \mu^*(B_n) \quad (\dagger)$$

WLOG we assume  $\mu^*(B_n) < \infty$  for all  $n$  so for all  $\varepsilon > 0$ , there exists  $A_{nm}$  such that  $B_n \subseteq \bigcup_m A_{nm}$  and

$$\mu^*(B_n) + \frac{\varepsilon}{2^n} \geq \sum_m \mu(A_{nm})$$

Now since  $\mu^*$  and since  $B \subseteq \bigcup_n B_n \subseteq \bigcup_{n,m} A_{nm}$ , hence

$$\mu^*(B) \leq \mu^*\left(\bigcup_{n,m} A_{nm}\right) \leq \sum_{n,m} \mu(A_{nm}) \leq \sum_n \mu^*(B_n) + \underbrace{\sum_n \frac{\varepsilon}{2^n}}_{=\varepsilon}$$

so  $(\dagger)$  follows since  $\varepsilon$  was arbitrary.

Step 2:  $\mu^*$  extends  $\mu$ . Let  $A \in \mathcal{A}$ . Clearly  $A = A \cup \emptyset \cup \dots \cup \emptyset$ , so by definition of  $\mu^*$ ,  $\mu^*(A) \leq \mu(A) + 0 + \dots + 0$ . So we need to prove  $\mu(A) \leq \mu^*(A)$ . Again, assume  $\mu^*(A) < \infty$  WLOG, and let  $A_n \in \mathcal{A}$  be such that  $A \subseteq \bigcup_n A_n$ . Then  $A = \bigcup_n \underbrace{(A \cap A_n)}_{\in \mathcal{A}}$ , and since  $\mu$  is countably sub-additive on  $\mathcal{A}$ , we have

$$\mu(A) = \mu\left(\bigcup_n (A \cap A_n)\right) \leq \sum_n \mu(\underbrace{A \cap A_n}_{\subseteq A_n}) \leq \sum_n \mu(A_n)$$

so since the  $(A_n)$  were arbitrary, by taking infima, we have  $\mu(A) \leq \mu^*(A)$ .

Step 3:  $\mathcal{M} \supseteq \mathcal{A}$ . Let  $A \in \mathcal{A}$ , then  $B \subseteq (B \cap A) \cup (B \cap A^c) \cup \emptyset \dots \cup \emptyset = \bigcup_n B_n$  so by  $(\dagger)$  we have

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c) + 0 \dots + 0$$

so we need to prove  $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$ . Again, WLOG assume  $\mu^*(B) < \infty$ , and so for all  $\varepsilon > 0$  there exist  $A_n \in \mathcal{A}$  such that  $B \subseteq \bigcup_n A_n$  and

$$\mu^*(B) + \varepsilon \geq \sum_n \mu(A_n) \quad (\circ)$$

now  $B \cap A \subseteq \bigcup_n \underbrace{(A_n \cap A)}_{\in \mathcal{A}}$  and  $B \cap A^c \subseteq \bigcup_n \underbrace{(A_n \cap A^c)}_{A_n \setminus A \in \mathcal{A}}$ . Therefore by definition of inf in  $\mu^*$  and additivity of  $\mu$

$$\begin{aligned} \mu^*(B \cap A) + \mu^*(B \cap A^c) &\leq \sum_n (\mu(A_n \cap A) + \mu(A_n \cap A^c)) \\ &= \sum_n \mu(A_n) \\ &\underbrace{\leq}_{\circ} \mu^*(B) + \varepsilon \end{aligned}$$

since  $\varepsilon$  was arbitrary,  $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$ , so  $A \in \mathcal{M}$ .

Step 4:  $\mathcal{M}$  is an algebra. Clearly  $\emptyset \in \mathcal{M}$ , and by the definition of  $\mathcal{M}$  it's obvious that  $A^c \in \mathcal{M}$  whenever  $A \in \mathcal{M}$ . So let  $A_1, A_2 \in \mathcal{M}$

$$\mu^*(B) = \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c), \text{ since } A_1 \in \mathcal{M}$$

$$= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c), \text{ since } A_2 \in \mathcal{M}$$

Clearly  $A_1 \cap A_2^c = (A_1 \cap A_2^c) \cap A_1$  and  $A_1^c = (A_1 \cap A_2)^c \cap A_1^c$  so

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap (A_1 \cap A_2)^c), \text{ since } A_1 \in \mathcal{M} \end{aligned}$$

so  $A_1 \cap A_2 \in \mathcal{M}$ , and  $\mathcal{M}$  is an algebra.

Step 5: Let  $A = \bigcup_n A_n$ ,  $A_n \in \mathcal{M}$ , WLOG  $A_n$  disjoint (disjointification). Want  $A \in \mathcal{M}$  and  $\mu^*(A) = \sum_n \mu^*(A_n)$ . By  $(\dagger)$  we clearly have

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c) + 0 \dots + 0$$

and

$$\mu^*(A) \leq \sum_n \mu^*(A_n)$$

so we only need two converse inequalities. Similar to before

$$\begin{aligned}
 \mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
 &= \mu^*(B \cap A_1) + \mu^*(B \cap \underbrace{A_1^c \cap A_2}_{=A_2 \text{ as disjoint}}) + \mu^*(B \cap A_1^c \cap A_2^c) \\
 &= \sum_{n \leq N} \mu^*(B \cap A_n) + \mu^*(B \cap A_1^c \cap \dots \cap A_N^c)
 \end{aligned}$$

since  $\bigcup_{n \leq N} A_n \subseteq A$  so  $\bigcap_{n \leq N} A_n^c \supseteq A^c$ , taking limits

$$\mu^*(B) \geq \sum_{n=1}^{\infty} \mu^*(B \cap A_n) + \mu^*(B \cap A^c)$$

and by (†)

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

so  $A \in \mathcal{M}$ . Applying the previous with  $B = A$ , we see

$$\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A \cap A_n) + \mu^*(A \cap A^c) = \sum_n \mu^*(A_n)$$

□

**Definition.** A collection  $\mathcal{A}$  of subsets of  $E$  is called a  $\pi$ -system if  $\emptyset \in \mathcal{A}$  and if  $A, B \in \mathcal{A}$ , then  $A \cap B \in \mathcal{A}$ .

**Definition.**  $\mathcal{A}$  is called a  $d$ -system if  $E \in \mathcal{A}$ , and if  $B_1, B_2 \in \mathcal{A}$  such that  $B_1 \subseteq B_2$ , then  $B_2 \setminus B_1 \in \mathcal{A}$ , and if  $A_n \in \mathcal{A}$ ,  $A_n \uparrow \bigcup_n A_n = A$ , then  $A \in \mathcal{A}$ .

One shows (Example sheet) that a  $d$ -system which is also a  $\pi$ -system is a  $\sigma$ -algebra.

**Lemma** (Dynkin). *Let  $\mathcal{A}$  be a  $\pi$ -system. Then any  $d$ -system that contains  $\mathcal{A}$  also contains  $\sigma(\mathcal{A})$ .*

*Proof.* Define

$$\mathcal{D} = \bigcap_{\mathcal{D}' \supseteq \mathcal{A} \text{ a } d\text{-system}} \mathcal{D}'$$

which is again a  $d$ -system (Example sheet). We show that  $\mathcal{D}$  is a  $\pi$ -system, hence a  $\sigma$ -algebra containing  $\mathcal{A}$ . Define

$$\mathcal{D}' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \forall A \in \mathcal{A}\}$$

which contains  $\mathcal{A}$  as  $\mathcal{A}$  is a  $\pi$ -system. Next we show  $\mathcal{D}'$  is a  $d$ -system. Clearly  $E \cap A = A \in \mathcal{A} \subseteq \mathcal{D}$ , so  $E \in \mathcal{D}'$ . Next let  $B_1, B_2 \in \mathcal{D}'$  such that  $B_1 \subseteq B_2$  then  $(B_2 \setminus B_1) \cap A = \underbrace{(B_2 \cap A)}_{\in \mathcal{D}} \setminus \underbrace{(B_1 \cap A)}_{\in \mathcal{D}} \in \mathcal{D}$  and so  $B_2 \setminus B_1 \in \mathcal{D}'$ .

Next take  $B_n \uparrow B$ ,  $B_n \in \mathcal{D}'$  then  $\underbrace{B_n \cap A}_{\in \mathcal{D}} \uparrow B \cap A \in \mathcal{D}$  so  $B \in \mathcal{D}'$ .

Hence  $\mathcal{D}'$  is a  $d$ -system containing  $\mathcal{A}$ , so by minimality of  $\mathcal{D}$ ,  $\mathcal{D} \subseteq \mathcal{D}'$ . Conversely, by construction  $\mathcal{D}' \subseteq \mathcal{D}$ , so  $\mathcal{D}' = \mathcal{D}$ .

Next define

$$\mathcal{D}'' = \{B \in \mathcal{D} : B \cap A \in \mathcal{D} \forall A \in \mathcal{D}\}$$

which by the preceding step ( $\mathcal{D}' = \mathcal{D}$ ) contains  $\mathcal{A}$ . Just as before, one shows that  $\mathcal{D}'' = \mathcal{D}$  and so  $\mathcal{D}$  is a  $\pi$ -system (as  $\mathcal{D}''$  is by construction).  $\square$

**Theorem** (Uniqueness of extension). *Let  $\mu_1, \mu_2$  be measures on  $(E, \mathcal{E})$  such that  $\mu_1(E) = \mu_2(E) < \infty$ , and suppose  $\mu_1 = \mu_2$  on a  $\pi$ -system  $\mathcal{A}$  such that  $\mathcal{E} \subseteq \sigma(\mathcal{A})$ . Then  $\mu_1 = \mu_2$  on  $\mathcal{E}$ .*

*Proof.* Define

$$\mathcal{D} = \{A : \mu_1(A) = \mu_2(A)\}$$

which contains  $\mathcal{A}$  by hypothesis. We show that  $\mathcal{D}$  is a  $d$ -system, and hence by Dynkin's Lemma, contains  $\sigma(\mathcal{A})$ , so the theorem follows.

To see this, note first that  $E \in \mathcal{D}$  by hypothesis. Next, by additivity and finiteness of  $\mu_1, \mu_2$ , for  $B_1 \subseteq B_2$ ,  $B_1, B_2 \in \mathcal{D}$ .

$$\mu_1(B_2 \setminus B_1) = \mu_1(B_2) - \mu_1(B_1) = \mu_2(B_2) - \mu_2(B_1) = \mu_2(B_2 \setminus B_1)$$

so  $B_2 \setminus B_1 \in \mathcal{D}$ . Finally take  $B_n \uparrow B$ ,  $B_n \in \mathcal{D}$ . This implies  $B \setminus B_n \downarrow \emptyset$  and (by Example sheet)  $\mu_i(B \setminus B_n) \rightarrow \mu_i(\emptyset) = 0$  for  $i = 1, 2$ . This implies for  $\mu_i(B) < \infty$  that  $\mu_i(B_n) \rightarrow \mu_i(B)$  as  $n \rightarrow \infty$  for both  $i = 1, 2$ . But then

$$\mu_1(B) = \lim_{n \rightarrow \infty} \mu_1(B_n) = \lim_{n \rightarrow \infty} \mu_2(B_n) = \mu_2(B)$$

and so  $B \in \mathcal{D}$ , and thus  $\mathcal{D}$  is a d-system.  $\square$

**Remark:** the above theorem applies to finite measures  $\mu$  such that  $\mu(E) < \infty$ . The above theorem extends (as we will see) to  $\sigma$ -finite measures  $\mu$  for which  $E = \bigcup_{n \in \mathbb{N}} E_n$  such that  $\mu(E_n) < \infty$ .

## Borel- $\sigma$ -algebras

**Definition.** Let  $E$  be a topological space (Hausdorff, or metric space). The  $\sigma$ -algebra generated by  $\mathcal{A} = \{B \subseteq E : B \text{ is open}\}$  is called the *Borel- $\sigma$ -algebra*, denoted by  $\mathcal{B}(E)$ , or just  $\mathcal{B}$  when  $E = \mathbb{R}$ . Elements of  $\mathcal{B}(E)$  are the Borel subsets of  $E$ . A measure  $\mu$  on  $(E, \mathcal{B}(E))$  is called a *Borel measure on  $E$* . A *Radon* measure  $\mu$  is a Borel measure such that  $\mu(K) < \infty$  for all  $K \subseteq E$  compact (closed in Hausdorff spaces, hence measurable).

## Construction of Lebesgue measure

We will (eventually) construct a unique Borel measure  $\mu$  on  $\mathbb{R}^d$  such that

$$\mu\left(\prod_{i=1}^d [a_i, b_i]\right) = \prod_{i=1}^d |b_i - a_i|, \quad a_i < b_i, \quad i = 1, \dots, d$$

We will do  $d = 1$  first.

**Theorem.** *There exists a unique Borel measure (called the Lebesgue measure)  $\mu$  on  $\mathbb{R}$  such that*

$$\mu((a, b]) = b - a, \quad \forall a < b \quad (\dagger)$$

*Proof.* Consider the collection  $\mathcal{A}$  of subsets of  $\mathbb{R}$  of the form

$$A = (a_1, b_1] \cup \dots \cup (a_n, b_n]$$

(intervals pairwise disjoint) which form a ring ( $\emptyset = ((a, a])$ , unions and differences are clear), which generates (Example sheet) generates the same  $\sigma$ -algebra on the open such intervals, and open intervals with rational endpoints generate  $\mathcal{B}$ , so  $\sigma(\mathcal{A}) \supseteq \mathcal{B}$ .

Define a set function  $\mu$  on  $\mathcal{A}$  by

$$\mu(A) = \sum_{i=1}^n (b_i - a_i)$$

$\mu$  is clearly additive, and well-defined since if  $A = \bigcup_j C_j$  and  $A = \bigcup_k D_k$  for distinct disjoint unions, then  $C_j = \bigcup_k (C_j \cap D_k)$  and  $D_k = \bigcup_j (D_k \cap C_j)$ , so

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_j C_j\right) = \sum_j \mu(C_j) = \sum_j \mu\left(\bigcup_k (C_j \cap D_k)\right) \\ &= \sum_{j,k} \mu(C_j \cap D_k) = \dots = \mu\left(\bigcup_k D_k\right) = \mu(A) \end{aligned}$$

by additivity of  $\mu$ . Now to prove existence of  $\mu$ , we apply Caratheodory's theorem and need to check that  $\mu$  is countably additive on  $\mathcal{A}$ . By the Example sheet, it suffices to show that for all  $A_n \in \mathcal{A}$  such that  $A_n \downarrow \emptyset$  we have  $\mu(A_n) \rightarrow 0$ .

Assume this is not the case, so there exists some  $\varepsilon > 0$  and  $B_n \in \mathcal{A}$  such that  $B_n \downarrow \emptyset$  but  $\mu(B_n) \geq 2\varepsilon$  for all  $n$ . We can approximate  $B_n$  from within by  $C_n = \bigcup_{i=1}^{N_n} \left(a_{n_i} + \frac{\varepsilon 2^{-n}}{N_n}, b_{n_i}\right] \in \mathcal{A}$  such that  $\mu(B_n \setminus C_n) = \varepsilon 2^{-n} \cdot \frac{N_n}{N_n} = \varepsilon 2^{-n}$ .

Now since  $B_n \downarrow$ , we have  $B_N = \bigcap_{n \leq N} B_n$  and

$$B_N \setminus (C_1 \cap \dots \cap C_N) = B_N \cap \left(\bigcup_{n \leq N} C_n^c\right) = \bigcup_{n \leq N} B_N \setminus C_n \subseteq \bigcup_{n \leq N} B_n \setminus C_n$$

Hence since  $\mu$  is increasing

$$\mu(B_N \setminus (C_1 \cap \dots \cap C_N)) \leq \mu\left(\bigcup_{n \leq N} B_n \setminus C_n\right) \leq \sum_{n \leq N} \mu(B_n \setminus C_n) \leq \varepsilon$$

Hence the “length” of what was removed  $(C_1 \cap \dots \cap C_N)$  must be at least  $\varepsilon$ , i.e

$$\mu(C_1 \cap \dots \cap C_N) \geq \varepsilon > 0$$



This means that  $C_1 \cap \dots \cap C_N$  is non-empty for all  $N$ , and so is

$$K_N = \overline{C_1} \cap \dots \cap \overline{C_N}$$

( $\overline{C_i}$  denotes the closure of  $C_i$ ) Thus  $K_N$  is a nested sequence of non-empty closed intervals, so  $\emptyset \neq \bigcap_N K_N$ . But  $K_N \subseteq \overline{C_N} \subseteq B_N$ , so  $\emptyset \neq \bigcap_N K_N \subseteq \bigcap_N B_n = \emptyset$ , a contradiction. So a measure  $\mu$  satisfying  $(\dagger)$  must exist.

For uniqueness, suppose  $\mu, \lambda$  measures such that  $(\dagger)$  holds, and define  $\mu_n(A) = \mu(A \cap (n, n+1])$ ,  $\lambda(A) = \lambda(A \cap (n, n+1])$  for  $n \in \mathbb{Z}$ , which are finite measures such that  $\mu_n(E) = 1 = \lambda_n(E)$  and  $\mu_n = \lambda_n$  on the  $\pi$ -system  $\mathcal{A}$ . So by the uniqueness theorem, we must have  $\mu_n = \lambda_n$  on  $\mathcal{B}$ , and

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_n A \cap (n, n+1]\right) = \sum_n \mu(A \cap (n, n+1]) = \sum_n \mu_n(A) \\ &= \sum_n \lambda_n(A) = \dots = \lambda(A) \end{aligned}$$

so  $\lambda = \mu$ . □

**Remarks:**

1. a set  $B \in \mathcal{B}$  is called a Lebesgue null set if  $\mu(B) = 0$ . Can write  $\{x\} = \bigcap_n (x - \frac{1}{n}, x]$  and so  $\mu(\{x\}) = \lim_n \frac{1}{n} = 0$ . In particular  $\mu((a, b)) = \mu((a, b]) = \mu([a, b])$ , and any countable set  $Q$  satisfies  $\mu(Q) = \mu\left(\bigcup_{q \in Q} \{q\}\right) = \sum_{q \in Q} \mu(\{q\}) = 0$ . But there exist  $C$  uncountable (and measurable) in  $\mathcal{B}$  such that  $\mu(C) = 0$  [Cantor set].
2. Translation invariance of  $\mu$ : let  $x \in \mathbb{R}$ , then  $B + x = \{b + x : b \in B\}$  is in  $\mathcal{B}$  whenever  $B \in \mathcal{B}$  and we can define

$$\mu_x(B) = \mu(B + x)$$

and by uniqueness in the preceding theorem

$$\mu_x((a, b]) = \mu((a + x, b + x]) = (b + x) - (a + x) = b - a$$

so  $\mu_x = \mu$ .

3. Lebesgue-measurable sets: in the extension theorem,  $\mu$  was assigned on the class  $\mathcal{M}$ , which can be shown (Example sheet) to equal

$$\mathcal{M} = \{M = A \cup N : A \in \mathcal{B}, N \subseteq B \in \mathcal{B} \text{ s.t. } \mu(B) = 0\}$$

## Existence of non-measurable sets

Consider  $E = (0, 1]$  with addition “+” modulo 1, and Lebesgue measure  $\mu$  is still translation invariant modulo 1.

Consider the subgroup  $Q = E \cap \mathbb{Q}$  of  $E$  and declare  $x \sim y$  if  $x - y \in Q$ . This gives equivalence classes  $[x] = \{y \in E : x \sim y\}$  on  $E$ . Assuming the axiom of choice, we can select a representative of  $[x]$ , and denote by  $S$  the set of selections running over all equivalence classes. Then we can partition  $E$  into the union of its cosets,

$$E = \bigcup_{q \in Q} (S + q)$$

a disjoint union.

Assume  $S$  is a Borel set (in  $\mathcal{B}(E)$ ), then  $S + q$  is also a Borel set for all  $q \in Q$ , and we can write (by countable additivity and translation invariance)

$$1 = \mu(E) = \mu\left(\bigcup_{q \in Q} (S + q)\right) = \sum_{q \in Q} \mu(S + q) = \sum_{q \in Q} \mu(S)$$

which is a contradiction. So  $S \notin \mathcal{B}(E)$ .

One can further show that  $\mu$  cannot extend to  $\mathcal{P}(E)$ ,

**Theorem** (Banach, Kuretwski). *Assuming the continuum hypothesis, there exists no measure on  $([0, 1])$  such that  $\mu((0, 1]) = 1$  and  $\mu(\{x\}) = 0$  for all  $x \in (0, 1]$ .*

*Proof.* Not given [see Dudley, 2002]. □

## Probability Spaces

If  $(E, \mathcal{E}, \mu)$  (a measure space) is such that  $\mu(E) = 1$ , we often call it a *probability space* and write  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\Omega$  is the set of outcomes/the sample space;  $\mathcal{F}$  is the set of events and  $\mathbb{P}$  is the probability measure.

The axioms of probability theory (Kolmogorov, 1933) are

1.  $\mathbb{P}(\Omega) = 1$
2.  $0 \leq \mathbb{P}(E) \leq 1, \forall E \in \mathcal{F}$
3. If  $(A_n : n \in \mathbb{N})$  are disjoint,  $A_n \in \mathcal{F}$ , then  $\mathbb{P}(\bigcup_n A_n) = \sum_n \mathbb{P}(A_n)$  [so  $\mathbb{P}$  is a measure on a  $\sigma$ -algebra]

We further say that  $(A_i : i \in I)$  are *independent* if for all  $J \subseteq I$  finite, we have

$$\mathbb{P} \left( \bigcap_{j \in J} A_j \right) = \prod_{j \in J} \mathbb{P}(A_j)$$

We further say  $\sigma$ -algebras  $(\mathcal{A}_i : i \in I)$  are *independent* if for any  $A_j \in \mathcal{A}_j$ ,  $j \in J$ ,  $J \subseteq I$  finite, the  $A_j$ 's are independent.

**Proposition.** Let  $\mathcal{A}_1, \mathcal{A}_2$  be  $\pi$ -systems of sets in  $\mathcal{F}$ , and suppose  $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$  for all  $A_1 \in \mathcal{A}_1$ ,  $A_2 \in \mathcal{A}_2$ . Then the  $\sigma$ -algebras  $\sigma(\mathcal{A}_1), \sigma(\mathcal{A}_2)$  are independent.

*Proof.* Exercise. □

### The Borel-Cantelli Lemmas

For a sequence  $(A_n : n \in \mathbb{N})$ ,  $A_n \in \mathcal{F}$ , define

$$\limsup_n A_n = \bigcap_n \bigcup_{m \geq n} A_m = \{A_n \text{ infinitely often "i.o."}\}$$

$$\liminf_n A_n = \bigcup_n \bigcap_{m \geq n} A_m = \{A_n \text{ eventually}\}$$

**Lemma** (1st Borel-Cantelli Lemma). *If  $A_n \in \mathcal{F}$  are such that  $\sum_n \mathbb{P}(A_n) < \infty$  then  $\mathbb{P}(A_n \text{ i.o.}) = 0$*

*Proof.*

$$\mathbb{P} \left( \bigcap_n \bigcup_{m \geq n} A_m \right) \leq \mathbb{P} \left( \bigcup_{m \geq n} A_m \right) \leq \sum_{m \geq n} \mathbb{P}(A_m) \rightarrow 0$$

□

**Remark:** the proof actually works for any measure  $\mu$ .

**Lemma** (2nd Borel-Cantelli Lemma). *Suppose  $A_n \in \mathcal{F}$  are independent and  $\sum_n \mathbb{P}(A_n) = \infty$ . Then  $\mathbb{P}(A_n \text{ i.o.}) = 1$ .*

*Proof.* By independence, for any  $N \geq n$  and using  $1 - a \leq e^{-a}$ ,

$$\mathbb{P} \left( \bigcap_{m=n}^N A_m^c \right) = \prod_{m=n}^N (1 - \mathbb{P}(A_m)) \leq \exp \left( - \sum_{m=n}^N \mathbb{P}(A_m) \right) \rightarrow 0 \text{ as } N \rightarrow \infty$$

Since  $\bigcap_{m=n}^N A_m^c \downarrow \bigcap_{m \geq n} A_m^c$ , by countable additivity we have

$$\mathbb{P} \left( \bigcap_{m \geq n} A_m^c \right) = 0$$

But then

$$\begin{aligned}\mathbb{P}(A_n \text{ i.o.}) &= \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m\right) = 1 - \mathbb{P}\left(\bigcup_n \bigcap_{m \geq n} A_m^c\right) \\ &\geq 1 - \sum_n \underbrace{\mathbb{P}\left(\bigcap_{m \geq n} A_m^c\right)}_{=0} = 1\end{aligned}$$

□

## 2 Measurable functions

Let  $(E, \mathcal{E})$ ,  $(G, \mathcal{G})$  be measurable spaces and let  $f : E \rightarrow G$ . We say that  $f$  is  $\mathcal{E}$ - $\mathcal{G}$ -measurable if  $f^{-1}(A) \in \mathcal{E}$  for all  $A \in \mathcal{G}$ . If  $G = \mathbb{R}$  with  $\mathcal{G} = \mathcal{B}(\mathbb{R})$ , we just say  $f : (E, \mathcal{E}) \rightarrow \mathbb{R}$  is *measurable*.

Moreover, if  $E$  is a topological space and  $\mathcal{E} = \mathcal{B}(E)$ , we say  $f$  is *Borel measurable*.

Preimages preserve set operations:  $f^{-1}(\bigcup_i A_i) = \bigcup_i f^{-1}(A_i)$  and  $f^{-1}(G \setminus A) = E \setminus f^{-1}(A)$ , which implies that  $\{f^{-1}(A) : A \in \mathcal{G}\}$  is a  $\sigma$ -algebra over  $E$ , and likewise  $\{A : f^{-1}(A) \in \mathcal{E}\}$  is also a  $\sigma$ -algebra over  $G$ .

This implies that if  $\mathcal{A}$  is a collection of subsets of  $G$  generating  $\mathcal{G}$  and such that  $f^{-1}(A) \in \mathcal{E}$  for all  $A \in \mathcal{A}$ , then  $\{A : f^{-1}(A) \in \mathcal{E}\}$  is a  $\sigma$ -algebra containing  $\mathcal{A}$ , and hence  $\mathcal{G}$ . In particular, it suffices to check  $f^{-1}(A) \in \mathcal{E}$ ,  $\forall A \in \mathcal{A}$  to conclude that  $f$  is measurable.

If  $f$  takes real values, then

$$\mathcal{A} = \{(-\infty, y] : y \in \mathbb{R}\}$$

generates  $\mathcal{B}(\mathbb{R})$  (Example sheet), and so  $f$  will be measurable whenever  $f^{-1}((-\infty, y]) = \{x \in E : f(x) \leq y\} \in \mathcal{E}$  for all  $y \in \mathbb{R}$ . Moreover, if  $E$  is a topological space with  $\mathcal{E} = \mathcal{B}(E)$ , then if  $f : E \rightarrow \mathbb{R}$  is continuous, it is Borel measurable.

The indicator function

$$1_A(x) = \begin{cases} 1 & \text{when } x \in A \\ 0 & \text{when } x \notin A \end{cases}$$

is measurable if and only if  $A \in \mathcal{E}$ .

One shows that compositions of measurable maps are measurable, and so are  $f_1 + f_2$ ,  $f_1 \cdot f_2$ ,  $\inf_n f_n$ ,  $\liminf_n f_n$ ,  $\limsup_n f_n$  whenever the  $f_n$  are.

Moreover, given a collection of maps  $\{f_i : E \rightarrow (G, \mathcal{G}), i \in I\}$  we can make them all measurable for

$$\sigma(f_i^{-1}(A) : A \in \mathcal{G}, i \in I)$$

**Theorem** (Monotone class theorem). *Let  $\mathcal{A}$  be a  $\pi$ -system generating the  $\sigma$ -algebra  $\mathcal{E}$  over  $E$ . Let further  $\mathcal{V}$  be a vector space of bounded maps from  $E$  to  $\mathbb{R}$  such that*

1.  $1_E \in \mathcal{V}$ ,  $1_A \in \mathcal{V}$ ,  $\forall A \in \mathcal{A}$ .
2. If  $f$  is bounded and  $f_n \in \mathcal{V}$  is such that  $0 \leq f_n \uparrow f$  pointwise on  $E$ , then  $f \in \mathcal{V}$ .

Then  $\mathcal{V}$  contains all bounded measurable  $f : E \rightarrow \mathbb{R}$ .

*Proof.* Define  $\mathcal{D} = \{A \in \mathcal{E} : 1_A \in \mathcal{V}\}$ . By hypothesis,  $\mathcal{D}$  contains the  $\pi$ -system  $\mathcal{A}$  and we now show it is also a  $d$ -system, so by Dynkin's lemma,  $\mathcal{E} = \mathcal{D}$ . Indeed,  $E \in \mathcal{D}$  since  $1_E \in \mathcal{V}$  by hypothesis. Also if  $A \subseteq B$ ,  $A, B \in \mathcal{D}$ , then  $1_{B \setminus A} = 1_B - 1_A \in \mathcal{V}$  as  $\mathcal{V}$  is a vector space. Finally, if  $A_n \in \mathcal{D}$  and  $A_n \uparrow A$ , then  $1_{A_n} \uparrow 1_A$  pointwise and so  $1_A \in \mathcal{V}$  by hypothesis, so  $A \in \mathcal{D}$ . In particular  $A \in \mathcal{V}$  for all  $A \in \mathcal{E}$ .

Let now  $f : E \rightarrow \mathbb{R}$  be bounded, non-negative and measurable. Define

$$f_n = \sum_{j=0}^{n2^n} \frac{j}{2^n} 1_{A_{n_j}}$$

where  $A_{n_j} = \{x \in E : \frac{j}{2^n} < f(x) \leq \frac{j+1}{2^n}\} = f^{-1}((\frac{j}{2^n}, \frac{j+1}{2^n}]) \in \mathcal{E}$  for  $j = 0, \dots, n2^n - 1$ , and  $A_{n_{n2^n}} = \{x \in E : f(x) > n\} = f^{-1}((n, \infty)) \in \mathcal{E}$ .

Clearly since  $f$  is bounded, for  $n > \|f\|_\infty$ , we see

$$f_n \leq f \leq f_n + 2^{-n}$$

so  $|f_n - f| \leq 2^{-n} \rightarrow 0$ . So by hypothesis  $f \in \mathcal{V}$ . For general  $f$  bounded and measurable, we can decompose  $f = f^+ - f^-$  where  $f^\pm \geq 0$ , and repeat the argument above.  $\square$

## Image Measures

If  $f : (E, \mathcal{E}) \rightarrow (G, \mathcal{G})$  is  $\mathcal{E}$ - $\mathcal{G}$  measurable, and  $\mu$  is a measure on  $\mathcal{E}$ , then the image measure  $\nu = \mu \circ f^{-1}$  is obtained from

$$\nu(A) = \mu(f^{-1}(A)), \quad \forall A \in \mathcal{G}$$

which is indeed a measure on  $\mathcal{G}$  (Example sheet).

**Lemma.** Let  $g : \mathbb{R} \rightarrow \mathbb{R}$  be a right-continuous, monotone increasing function, and set  $g(\pm\infty) = \lim_{z \rightarrow \pm\infty} g(z)$ . On  $I = (g(-\infty), g(\infty))$  define

$$f(x) = \inf\{y \in \mathbb{R} : x \leq g(y)\}, \quad x \in I$$

Then  $f$  is monotone increasing, left-continuous and

$$f(y) \leq y \iff x \leq g(y) \quad \forall x, y$$

*Proof.* Define  $J_x = \{y \in \mathbb{R} : x \leq g(y)\}$ . Since  $x > g(-\infty)$ ,  $J_x$  is non-empty and bounded below, so  $f(x) \in \mathbb{R}$ . Now if  $y \in J_x$  then  $y' \geq y$  implies  $y' \in J_x$  as well since  $g \uparrow$ . Moreover if  $y_n \downarrow y$ ,  $y_n \in J_x$ , then we can take limits in  $x \leq g(y_n)$  to see  $x \leq \lim_n g(y_n) = g(y)$  as  $g$  is right-continuous, so  $y \in J_x$ . We conclude that  $J_x = [f(x), \infty)$ , which shows the equivalence.

Moreover, if  $x \leq x'$ , then  $J_x \supseteq J_{x'}$  since  $g \uparrow$ . So by properties of the infimum  $f(x) \leq f(x')$ . Likewise if  $x_n \uparrow x$ , then  $J_x = \bigcap_n J_{x_n}$  so  $f(x_n) \rightarrow f(x)$  as  $x_n \rightarrow x$ .  $\square$

We call  $f$  the *generalised inverse of  $g$* .

**Theorem.** Let  $g$  be as in the above lemma. Then there exists a unique Radon measure  $\mu_g$  on  $\mathbb{R}$  such that  $\mu_g((a, b]) = g(b) - g(a)$  for all  $a < b$ . Every Radon measure on  $\mathbb{R}$  can be obtained in this way.

*Proof.* For  $f$  as defined in the previous lemma, note that for all  $z \in \mathbb{R}$

$$f^{-1}((-\infty, z]) = \{x : f(x) \leq z\} = \{x : x \leq g(y)\} = (g(-\infty), g(z)] \in \mathcal{B}(I)$$

Where the 2nd equality follows again from the lemma. So  $f$  is  $\mathcal{B}\text{-}\mathcal{B}(I)$  measurable, and the image measure  $\mu \circ f^{-1} = \mu_g$ , where  $\mu$  is the Lebesgue measure on  $I$ , exists.

Then for  $-\infty < a < b < \infty$  we have

$$\mu_g((a, b]) = \mu(f^{-1}((a, b])) = \mu(x \in I : a < f(x) \leq b) = \mu((g(a), g(b)]) = g(b) - g(a)$$

Which uniquely determines  $\mu_g$  by the same arguments as for the Lebesgue measure on  $\mathbb{R}$ . (Since  $g$  maps into  $\mathbb{R}$ ,  $\mu_g$  is a Radon measure).

Conversely, let  $\nu$  be any Radon measure on  $\mathbb{R}$ , define

$$g(y) = \begin{cases} \nu((0, y]) & y \geq 0 \\ -\nu((y, 0]) & y < 0 \end{cases}$$

Which is clearly increasing in  $y$  (since  $\nu$  is increasing). If  $y_n \downarrow y$ , then  $(0, y_n] \downarrow (0, y]$  so  $g(y_n) \rightarrow g(y)$  since  $\nu$  is countably additive, so  $g$  is right-continuous. Finally (assuming  $a < 0 < b$ , the other cases are similar),

$$\nu((a, b]) = \nu((a, 0]) + \nu((0, b]) = -g(a) + g(b) = g(b) - g(a)$$

And by uniqueness as before, the result follows.  $\square$

**Remark:** The  $\mu_g$  are called Lebesgue-Stieltjes measures, with Stieltjes distribution  $g$ .

For example, the Dirac measure  $\delta_x$  at  $x \in \mathbb{R}$ , defined by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Which has Stieltjes distribution  $g = 1_{[x, \infty)}$ .

## Random Variables

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, and  $(E, \mathcal{E})$  a measurable space.

**Definition.** An  $E$ -valued random variable  $X$  is any  $\mathcal{F}$ - $\mathcal{E}$  measurable map

$$X : \Omega \rightarrow E$$

When  $E = \mathbb{R}, \mathbb{R}^d$  (with Borel  $\sigma$ -algebras) we call  $X$  a *random variable*, or random vector. The *law* or *distribution*  $\mu_X$  of a random variable is given by  $\mu_X = \mathbb{P} \circ X^{-1}$  (the image measure) with, for  $E = \mathbb{R}$  distribution function

$$F_X(z) = \mu_X((-\infty, z]) = \mathbb{P}(X^{-1}((-\infty, z])) = \mathbb{P}(\omega \in \Omega : X(\omega) \leq z) = \mathbb{P}(X \leq z)$$

which uniquely determines  $\mu_X$ .

Using properties of measures one shows that any distribution function satisfies

1.  $F_X \uparrow$
2.  $F_X$  is right-continuous
3.  $\lim_{z \rightarrow -\infty} F_X(z) = \mu_X(\emptyset) = 0$  and  $\lim_{z \rightarrow \infty} F_X(z) = \mu_X(\mathbb{R}) = \mathbb{P}(\Omega) = 1$

Given any distribution function  $F_X$  satisfying 1,2 & 3, we can on  $(\Omega, \mathcal{F}, \mathbb{P}) = ((0, 1), \mathcal{B}(0, 1), \mu)$ , where  $\mu$  is the Lebesgue measure obtain a random variable  $X : \Omega \rightarrow \mathbb{R}$  by

$$X(\omega) = \inf\{x : \omega \leq F_X(x)\}$$

with distribution function  $F_X$ .