Notes on basic measure theory for computer scientists

Mathias Ruggaard Pedersen
mrp@cs.aau.dk

November 30, 2016

These notes are aimed at people learning measure theory for use in computer science. Much of the material in these notes is taken from [4] and [3]. However, I have rearranged some of the material, added and fleshed out some proofs, added a few more examples, and other small changes. The only prerequisites for these notes are some basic set theory and some analysis. However, measure theory (like most other parts of mathematics) has interesting connections to many other mathematical disciplines, in particular to topology. Hence, I have added pointers, indicated by e.g. ☞[**Topology**], throughout the text to show where measure theory connects with other subjects. If you are not familiar with the subject which is being referred to, then these pointers can easily be skipped without loss of continuity. If you find any errors in the text, please let me know.

Contents

1 Measurable spaces

Just as topology generalises the notion of "distance", so measure theory is intended to generalise the notion of "size" by "measuring" the size of sets. It turns out, however, that we can not measure the size of any sets we like, but we have to restrict ourselves to what we will call measurable sets, which is the analogue of open sets in topology.

Before we define measurable sets, we will introduce a few related kinds of collections of sets. In what follows, let X be an arbitrary set.

Definition 1.1 (π-system)**.** A non-empty collection of subsets $\Pi \subseteq 2^X$ is called a **π-system** if

(A1) $A, B \in \Pi$ implies $A \cap B \in \Pi$ (closed under finite intersection).

▲

Definition 1.2 (λ-system)**.** A non-empty collection of subsets $\Lambda \subseteq 2^X$ is called a **λ-system** if

(B1) $X \in \Lambda$,

(B2) $A \in \Lambda$ implies $A^c \in \Lambda$ (closed under complement),

(B3) $A_1, A_2, A_3, \dots \in \Lambda$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ implies $\bigcup_{n=1}^{\infty} A_n \in \Lambda$ (closed under countable union of pairwise disjoint sets).

▲

Definition 1.3 (Semiring)**.** A non-empty collection of subsets $\mathcal{S} \subseteq 2^X$ is called a **semiring** if

(C1) $\emptyset \in \mathcal{S}$,

(C2) $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$ (closed under finite intersection),

(C3) $A, B \in \mathcal{S}$ implies that there exists finitely many mutually disjoint sets $C_i \in \mathcal{S}$ such that $A \setminus B = \bigcup_i C_i$.

▲

Definition 1.4 (Field of sets)**.** A non-empty collection of subsets $\mathcal{F} \subseteq 2^X$ is called a **field of sets** if

(D1) $X \in \mathcal{F}$,

(D2) $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$ (closed under complement),

(D3) $A, B \in \mathcal{F}$ implies $A \cup B \in \mathcal{F}$ (closed under finite union),

(D4) $A, B \in \mathcal{F}$ implies $A \cap B \in \mathcal{F}$ (closed under finite intersection).

▲

Note that since we have closure under complement, (D1) could be replaced by $\emptyset \in \mathcal{F}$ (this is also true for λ-systems), and one of (D3) and (D4) could be removed, since they both imply each other by de Morgan's laws. Note also that any field of sets is a semiring.

Definition 1.5 (σ-algebra)**.** A non-empty collection of subsets $\Sigma \subseteq 2^X$ is called a **σ-algebra** (on X) if

(E1) $X \in \Sigma$,

(E2) $A \in \Sigma$ implies $A^c \in \Sigma$ (closed under complement),

(E3) $A_1, A_2, A_3, \dots \in \Sigma$ implies $\bigcup_{n=1}^{\infty} A_n \in \Sigma$ (closed under countable union).

▲

As before, (E1) could be replaced by $\emptyset \in \Sigma$. Notice also that (E2) and (E3) together imply closure under countable intersection, by de Morgan's laws.

☞ [**Boolean algebra**] Both fields of sets and σ-algebras are special cases of Boolean algebras. Fields of sets in particular are very important in the study of Boolean algebras. Also, you can informally translate the requirements of Definition 1.4 and Definition 1.5 into the language of logic in the following way. Complement becomes negation, intersection becomes conjunction, union becomes disjunction, and the empty set becomes the statement $\perp$ that is always false. This translation can be made precise in the study of Boolean algebra.

Let us remark a bit on the connections between these collections. Clearly, any σ-algebra is a field of sets. Any σ-algebra is also both a π-system and a λ-system. The converse is also true.

Proposition 1.6. *If $A \subseteq 2^X$ is both a π-system and a λ-system, then A is a σ-algebra.*

Proof. That (E1) and (E2) hold is immediate from Definition 1.2, since A is a λ-system. So we need to show that (E3) holds, i.e. that A is closed under countable union. Let A_n be a countable collection of sets in A and define

$$B_n = A_n \cap \left(\bigcap_{i=1}^{n-1} (A_i)^c \right).$$

We know by (B2) that $(A_i)^c \in A$ for all i, since A is a λ-system, and we know by (A1) that $B_n \in A$, since A is a π-system, and all the intersections are finite.

Now suppose that $x \in \bigcup_{i=1}^{\infty} A_i$. Let j be the smallest index such that $x \in A_j$. Then $x \in B_j$ but $x \notin B_i$ for $i \neq j$. In other words, all the B_n are pairwise disjoint. Hence it follows from (B3), since A is a λ-system, that $\bigcup_{i=1}^{\infty} B_i \in A$, and since

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i,$$

it follows that $\bigcup_{i=1}^{\infty} A_i \in A$. ■

In topology, the basic notion is that of an open set, which is given by the topology τ, which then defines the topological space (X, τ). Similarly, in measure theory the basic notion is that of a measurable set. For any set X, we can construct a σ-algebra Σ, and we then call (X, Σ) a measurable space. The measurable sets are those that are in Σ. In other words, the measurable sets are those that are given by the σ-algebra for the measurable space.

Definition 1.7 (Measurable set)**.** Given any set X, the pair (X, Σ) is called a **measurable space** if Σ is a σ-algebra on X. If $A \in \Sigma$, then we say that A is a **measurable set**. ▲

Example 1.8. The set $\{\emptyset, X\}$ is a σ-algebra on X. This is known as the **trivial** or **indiscrete** σ-algebra on X. ⧫

Example 1.9. Given any set X, the power set 2^X is a σ-algebra on X, known as the **discrete** σ-algebra. This is the σ-algebra that we often work with when dealing with discrete probability distributions. ⧫

Proposition 1.10. *The intersection Σ of an arbitrary collection (possibly uncountable) of σ-algebras on X is a σ-algebra on X.*

Proof. Let Σ_i, $i \in \mathcal{I}$ be a collection of σ-algebras on X, and let $\Sigma = \bigcap_{i \in \mathcal{I}} \Sigma_i$. We need to verify properties (E1)-(E3).

(E1) Since $\emptyset \in \Sigma_i$ for all $i \in \mathcal{I}$, we have $\emptyset \in \Sigma$.

(E2) If $A \in \Sigma$, then $A \in \Sigma_i$ for all $i \in \mathcal{I}$ and hence $A^c \in \Sigma_i$ for all $i \in \mathcal{I}$, so $A^c \in \Sigma$.

(E3) If A_j is a countable collection, then $A_j \in \Sigma_i$ for all $i \in \mathcal{I}$ and j. This means that $\bigcup_j A_j \in \Sigma_i$ for all $i \in \mathcal{I}$, so $\bigcup_j A_j \in \Sigma$.

■

It is not true in general that a union of σ-algebras is again a σ-algebra, not even for finite unions.

Example 1.11. Let $\Sigma_1 = \{\emptyset, A, A^c, X\}$ and $\Sigma_2 = \{\emptyset, B, B^c, X\}$ where $A \neq B$ and $A, B \in 2^X \setminus \{\emptyset, X\}$. Clearly, both Σ_1 and Σ_2 are σ-algebras on X, but $A \cup B \notin \Sigma_1 \cup \Sigma_2$, so the union of Σ_1 and Σ_2 violates (E3), and hence it is not a σ-algebra. ⧫

Corollary 1.12. *Given a set X and $Y \subseteq 2^X$, the intersection Σ of all σ-algebras on X that contain Y is the smallest σ-algebra containing Y.*

Proof. By Proposition 1.10, the intersection Σ of all σ-algebras on X that contain Y is a σ-algebra, so we only need to prove that it is the smallest, i.e. that for any σ-algebra Σ' on X such that $Y \subseteq \Sigma'$, $\Sigma \subseteq \Sigma'$.

We have that $Y \subseteq \Sigma$ by the properties of intersection. Now take any other σ-algebra Σ' such that $Y \subseteq \Sigma'$. Then we must have $\Sigma' = \Sigma_i$ for some $i \in \mathcal{I}$ from the construction of Σ. From this it follows that $\Sigma \subseteq \Sigma'$. ■

From Corollary 1.12 it follows that the following definition is well-defined.

Definition 1.13. Let X be a set and $Y \subseteq 2^X$. We define **the σ-algebra generated by** Y, written $\sigma(Y)$, as the smallest σ-algebra containing Y, i.e. $Y \subseteq \sigma(Y)$. More precisely,

$$\sigma(Y) \stackrel{\text{def}}{=} \bigcap \{\Sigma \mid \Sigma \text{ is a } \sigma\text{-algebra on } X \text{ and } Y \subseteq \Sigma\}.$$

▲

☞[**Topology**] The following example gives a canonical way of constructing a σ-algebra for an arbitrary topological space. Given a topological space (X, τ), the **Borel σ-algebra** (or simply Borel algebra) $\mathbb{B}(X, \tau)$ on X is the σ-algebra generated by the open sets (or, equivalently, the closed sets) of (X, τ). The elements of $\mathbb{B}(X, \tau)$ are called **Borel sets**. Note that in a measurable space $(X, \mathbb{B}(X, \tau))$ with the Borel σ-algebra, all open sets are measurable, but not all measurable sets are open. In fact, since σ-algebras are closed under complement, and the complement of an open set is a closed set, all closed sets are also measurable.

Now that we have the concept of a generated σ-algebra, we can describe some more relations between the kinds of sets that we introduced in the beginning of this section.

For a nested collection of sets satisfying

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$$

and $\bigcup_n A_n = A$, we write $A_n \uparrow A$.

Similarly, for a nested collection of sets satisfying

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$$

and $\bigcap_n A_n = A$, we write $A_n \downarrow A$.

Definition 1.14 (Monotone class)**.** A collection of sets $\mathcal{M}$ is called a **monotone class** if

1) whenever $A_n \uparrow A$ with all $A_n \in \mathcal{M}$, then $A \in \mathcal{M}$,
2) whenever $A_n \downarrow A$ with all $A_n \in \mathcal{M}$, then $A \in \mathcal{M}$.

▲

Proposition 1.15. *Any σ-algebra is a monotone class and if a monotone class is also a field of sets, then it is a σ-algebra.*

Proof. First suppose that Σ is a σ-algebra on X. If $A_n \uparrow A$ with all $A_i \in \Sigma$, then $A = \bigcup_n A_n$, so $A \in \Sigma$. Likewise, if $A_n \downarrow A$ with all $A_i \in \Sigma$, then $A = \bigcap_n A_n$, so $A \in \Sigma$. Hence Σ is a monotone class.

Next suppose that $\mathcal{M}$ is a monotone class which is also a field of sets. Since it is a field of sets, we have $X \in \mathcal{M}$ and $\mathcal{M}$ is closed under complement, so we only need to show that $\mathcal{M}$ is closed under countable union. Let $\{A_j \mid j \in \mathbb{N}\}$ be a countable collection of sets in $\mathcal{M}$ and define $B_i = \bigcup_{j=1}^{i} A_j$. Since $\mathcal{M}$ is a field of sets, it is closed under finite union, and hence $B_i = \bigcup_{j=1}^{i} A_j \in \mathcal{M}$. We have that $B_i \uparrow \left(\bigcup_j A_j\right)$, so $\bigcup_j A_j \in \mathcal{M}$, since $\mathcal{M}$ is a monotone class. ■

What Proposition 1.15 tells us is that instead of thinking of a σ-algebra as closed under complement and countable union, we can also think of it as closed under complement and closed under limits of increasing sequences. Sometimes this latter characterisation is more natural to think of.

Theorem 1.16 (Monotone class theorem)**.** *Let $\mathcal{F}$ be a field of sets, and let $m(\mathcal{F})$ be the smallest monotone class containing $\mathcal{F}$. Then $m(\mathcal{F}) = \sigma(\mathcal{F})$.*

Proof. By Proposition 1.15, it is enough to show that $m(\mathcal{F})$ is a field of sets, because this implies that $m(\mathcal{F})$ is a σ-algebra. Since $\sigma(\mathcal{F})$ is the smallest σ-algebra containing $\mathcal{F}$, this implies $\sigma(\mathcal{F}) \subseteq m(\mathcal{F})$, and since any σ-algebra is a monotone class and $m(\mathcal{F})$ is the smallest monotone class, we have $m(\mathcal{F}) \subseteq \sigma(\mathcal{F})$, so that $m(\mathcal{F}) = \sigma(\mathcal{F})$.

Now we proceed to show that $m(\mathcal{F})$ is a field of sets. Clearly, since $\mathcal{F} \subseteq m(\mathcal{F})$, we have $X \in m(\mathcal{F})$.

Next we show that $m(\mathcal{F})$ is closed under complement. Define $\mathcal{N} \overset{\text{def}}{=} \{E \in m(\mathcal{F}) \mid E^c \in m(\mathcal{F})\}$. Notice that $\mathcal{F} \subseteq \mathcal{N}$, because $\mathcal{F}$ is a field of sets and hence closed under complement, and $m(\mathcal{F})$ contains $\mathcal{F}$. We want to argue that $\mathcal{N}$ is a monotone class, so suppose that $A_i \uparrow A$ with all $A_i \in \mathcal{N}$. Then all $(A_i)^c$ are in $m(\mathcal{F})$, and $(A_i)^c \downarrow A^c$. Because $m(\mathcal{F})$ is a monotone class, this means that $A^c \in m(\mathcal{F})$, which implies $A \in \mathcal{N}$. A similar argument shows that if $A_i \downarrow A$ with all $A_i \in \mathcal{N}$, then $A \in \mathcal{N}$, which shows that $\mathcal{N}$ is a monotone class. Since $m(\mathcal{F})$ is the smallest monotone class containing $\mathcal{F}$, we have $m(\mathcal{F}) \subseteq \mathcal{N}$, and by the definition of $\mathcal{N}$, we have $\mathcal{N} \subseteq m(\mathcal{F})$. This implies that $\mathcal{N} = m(\mathcal{F})$, which shows that $m(\mathcal{F})$ is closed under complement.

Lastly we need to show that $m(\mathcal{F})$ is closed under finite intersection. To this end, define $M_A \overset{\text{def}}{=} \{E \mid E \cap A \in m(\mathcal{F})\}$. Like before, we wish to show that M_A is a monotone class. Suppose that $B_i \uparrow B$ with all $B_i \in M_A$. Then $(B_i \cap A) \uparrow (B \cap A)$, and since $(B_i \cap A) \in m(\mathcal{F})$, we also have $(B \cap A) \in m(\mathcal{F})$, which implies that $B \in M_A$. In a similar manner, we can show that $B_i \downarrow B$, then $B \in M_A$. Thus, M_A is a monotone class. Now, note that $A \in M_B$ if and only if $B \in M_A$. If $A \in \mathcal{F}$, then M_A is a monotone class which contains $\mathcal{F}$, which implies that $m(\mathcal{F}) \subseteq M_A$. Furthermore, if $A \in \mathcal{F}$ and $E \in m(\mathcal{F})$, then $E \in M_A$, and hence $A \in M_E$, which means that M_E is a monotone class which contains $\mathcal{F}$ and $m(\mathcal{F}) \subseteq M_E$. Putting all this together, if we have $D, E \in m(\mathcal{F})$, this means that $D \in M_E$, so $D \cap E \in m(\mathcal{F})$. This shows that $m(\mathcal{F})$ is closed under finite intersection, so $m(\mathcal{F})$ is a field of sets. ∎

Theorem 1.17 (Dynkin's π-λ theorem)**.** *If $P \subseteq 2^X$ is a π-system and $L \subseteq 2^X$ is a λ-system, then $P \subseteq L$ implies that $\sigma(P) \subseteq L$.*

Proof. Let $\lambda(P)$ be the smallest λ-system containing P. We want to show that $\lambda(P)$ is a π-system, and hence, by Proposition 1.6, a σ-algebra, because this will imply that

$$P \subset \sigma(P) \subseteq \lambda(P) \subseteq L.$$

Let $L_A \overset{\text{def}}{=} \{B \subseteq X \mid A \cap B \in \lambda(P)\}$. We now show that if $A \in \lambda(P)$, then L_A is a λ-system.

(B1) $A \cap X = A \in \lambda(P)$, so $X \in L_A$.

(B2) Assume $B \in L_A$, which means that $A \cap B \in \lambda(P)$. We have assumed that $A \in \lambda(P)$. Since $\lambda(P)$ is a λ-system, $A^c \in \lambda(P)$, and since A^c and $A \cap B$ are disjoint, $A^c \cup (A \cap B) \in \lambda(P)$. Taking the complement of this, we get

$$(A^c \cup (A \cap B))^c = A \cap (A^c \cup B^c) = (A \cap A^c) \cup (A \cap B^c) = A \cap B^c,$$

so $A \cap B^c \in \lambda(P)$, which means that $B^c \in L_A$.

(B3) If B_n is a pairwise disjoint collection in L_A, then $A \cap B_n$ is a pairwise disjoint collection in $\lambda(P)$. Because $\lambda(P)$ is a λ-system, this means that $\bigcup_{n=1}^{\infty}(A \cap B_n) = A \cap (\bigcup_{n=1}^{\infty} B_n) \in \lambda(P)$, which further implies that $\bigcup_{n=1}^{\infty} B_n \in L_A$.

We have thus verified that L_A is a λ-system.

Next we prove that $\lambda(P)$ is a π-system. Assume $A, B \in P$. Since P is a π-system, we know that $A \cap B \in P \subseteq \lambda(P)$, and hence $B \in L_A$. This means that $P \subseteq L_A$, and since we proved that L_A is a λ-system, we get $\lambda(P) \subseteq L_A$. This means that if $A \in P$ and $B \in \lambda(P)$, then we also have $B \in L_A$, so $A \cap B \in \lambda(P)$, and hence $A \in L_B$. This shows that $P \subseteq L_B$, and again $\lambda(P) \subseteq L_B$ because L_B is a λ-system. Finally, assume that $A, B \in \lambda(P)$. We know that $\lambda(P) \subseteq L_B$, so $A \in L_B$, which means that $A \cap B \in \lambda(P)$. This proves that $\lambda(P)$ is a π-system. ■

2 Measures

Recall that the idea of measure theory is to define a notion of size on sets. So far we have talked about measurable sets, which are the sets that we wish to measure the size of, but we have not spoken about how to actually do this measurement.

In this section, we will introduce the notion of measure, which assigns to each measurable set a non-negative value which we interpret as its size.

Definition 2.1 (Additivity)**.** Let (X, Σ) be a measurable space, and let $\{A_i \mid i \in \mathcal{I}\} \subseteq \Sigma$ be a countable collection of measurable sets. A function $\mu\colon \Sigma \to \mathbb{R}_{\geq 0}$ is

- **finitely additive** if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A, B \in \Sigma$ and $A \cap B = \emptyset$,
- **countably subadditive** if
$$\mu\left(\bigcup_{i \in \mathcal{I}} A_i\right) \leq \sum_{i \in \mathcal{I}} \mu(A_i),$$
- **countably additive** if all the A_i are pairwise disjoint and
$$\mu\left(\bigcup_{i \in \mathcal{I}} A_i\right) = \sum_{i \in \mathcal{I}} \mu(A_i).$$

▲

Definition 2.2 (Measure)**.** A **measure** μ on a measurable space (X, Σ) is a function $\mu\colon \Sigma \to \mathbb{R}_{\geq 0}$ such that

(F1) $\mu(\emptyset) = 0$,

(F2) μ is countably additive.

A **subprobability measure** is a measure μ such that $\mu\colon \Sigma \to [0, 1]$. A **probability measure** is a subprobability measure that furthermore satisfies $\mu(X) = 1$. ▲

(Sub-)probability measures are also often called **(sub-)probability distributions**. A measurable space with a measure is called a **measure space**. If μ is a probability measure, then the measure space (X, Σ, μ) is often called a **probability space**.

☞ **[Boolean algebra]** We have defined measures to be functions on σ-algebras, as is usually done in measure theory. However, measures can be defined in general for any Boolean algebra, though some results for measures only hold in the case of σ-algebras.

In computer science, it can be convenient to work with subprobability measures instead of probability measures, since, when we define probabilities on transition systems, a subprobability measure allows for the possibility of no transition, i.e. deadlock. However, note that subprobability measures are not strictly necessary to do this, as we can simply add a "dead" state with only a self-loop where we direct all of the remaining probability mass.

The following proposition gives some general properties of measures.

Proposition 2.3. *Let (X, Σ, μ) be a measure space. Then the following hold for all $A_n, A, B \in \Sigma$.*

1) If $A \subseteq B$, then $\mu(A) \leq \mu(B)$ (monotonicity).

2) If $A \subseteq B$ and $\mu(B)$ is finite, then $\mu(B \setminus A) = \mu(B) - \mu(A)$.

3) If $A_n \uparrow A$, then $\lim_{i\to\infty} \mu(A_i) = \mu(A)$ (continuity with respect to increasing sequences).

4) If $A_n \downarrow A$ and $\mu(A_1)$ is finite, then $\lim_{i\to\infty} \mu(A_i) = \mu(A)$ (continuity with respect to decreasing sequences).

Proof. 1) Observe that if $A \subseteq B$, then $B = A \cup (B \setminus A)$, and hence $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$ by countable additivity.

2) As in the proof of 1), we get $\mu(B) = \mu(A) + \mu(B \setminus A)$, so $\mu(B \setminus A) = \mu(B) - \mu(A)$ because $\mu(B)$ was assumed to be finite.

3) We first define a countable collection of pairwise disjoint sets, so that we can make use of countable addivitity. Let $B_1 = A_1$ and $B_{i+1} = A_{i+1} \setminus A_i$. Then all the B_i are pairwise disjoint and we have that for every n, $\bigcup_{i=1}^n B_i = A_n$ and $\bigcup_{i=1}^\infty B_i = A$. Hence, by countable addivitity, we get $\mu(A_n) = \sum_{i=1}^n \mu(B_i)$ and $\mu(A) = \sum_{i=1}^\infty B_i$. This means that

$$\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \sum_{i=1}^{\infty} \mu(B_i) = \mu(A).$$

4) We construct an increasing sequence B_i given by $B_i = A_1 \setminus A_i$. Now

observe that

$$
\begin{aligned}
A_1 \setminus \left(\bigcap_{i=1}^{\infty} A_i \right) &= A_1 \cap \left(\bigcap_{i=1}^{\infty} A_i \right)^c \\
&= A_1 \cap \left(\bigcup_{i=1}^{\infty} (A_i)^c \right) = \bigcup_{i=1}^{\infty} (A_1 \cap (A_n)^c) \\
&= \bigcup_{i=1}^{\infty} (A_1 \setminus A_i) = \bigcup_{i=1}^{\infty} B_i.
\end{aligned}
$$

Because of this, we get that

$$
\mu(A_1) - \mu\left(\bigcap_{i=1}^{\infty} A_i \right) = \mu\left(A_1 \setminus \left(\bigcap_{i=1}^{\infty} A_i \right) \right) = \mu\left(\bigcup_{i=1}^{\infty} B_i \right).
$$

Since the B_i are an increasing sequence, we get by 3) that

$$
\mu\left(\bigcup_{i=1}^{\infty} B_i \right) = \lim_{i\to\infty} \mu(B_i) = \lim_{i\to\infty} \mu\left(A_1 \setminus \left(\bigcap_{i=1}^{\infty} A_i \right) \right) = \mu(A_1) - \lim_{i\to\infty} \mu(A_i).
$$

Hence

$$
\mu(A_1) - \mu\left(\bigcap_{i=1}^{\infty} A_i \right) = \mu(A_1) - \lim_{i\to\infty} \mu(A_i),
$$

and the result now follows because we have assumed $\mu(A_1)$ to be finite. ■

Next we give a number of examples of measures.

Example 2.4. For any measurable space (X, Σ), the **null measure** ω is given by $\omega(A) = 0$ for any $A \in \Sigma$. ♦

Example 2.5. For any $x \in X$ and $A \in \Sigma$, we define the **Dirac delta measure** δ (or simply **Dirac measure**) as

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

Furthermore, if we fix the set A, we get the **characteristic function** χ_A (also known as the **indicator function** $\mathbb{1}_A$) of A, defined by

$$
\chi_A(x) = \mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}
$$

Note that the characteristic (indicator) function is **not** a measure, since it has X, not Σ, as its domain. ♦

Example 2.6. Given two measure spaces (X, Σ_X, μ) and (Y, Σ_Y, ν), the **product space** $(X, \Sigma_X, \mu) \otimes (Y, \Sigma_Y, \nu)$ is the measure space $(X \times Y, \Sigma_X \otimes \Sigma_Y, \mu \times \nu)$, where $\Sigma_X \otimes \Sigma_Y$ is the σ-algebra generated by the rectangles $A \times B$ for $A \in \Sigma_X$ and $B \in \Sigma_Y$, and $\mu \times \nu$ is the **unique** product measure defined by $(\mu \times \nu)(A \times B) = \mu(A) \cdot \nu(B)$ for all $(A, B) \in \Sigma_X \times \Sigma_B$. ♦

Example 2.7. Given a measure space (X, Σ_X, μ), a measurable space (Y, Σ_Y), and a measurable function $f\colon X \to Y$, we can define a measure ν on (Y, Σ_Y) as follows. For any $B \in \Sigma_Y$, let $\nu(B) \overset{\text{def}}{=} \mu(f^{-1}(B))$. This is called the **image measure** or **push forward** of μ under f. The measurability of f ensures that this is a well-defined measure. ♦

Example 2.8. In computer science, we often speak about the **trace** of some process, which is just a sequence of symbols over some alphabet. For the set of all traces, we can construct a σ-algebra as follows. Intuitively, the measurable sets are those that have some fixed, finite prefix, and then an arbitrary, infinite tail. That is, we characterise the measurable sets by a finite prefix. More precisely, a measurable set is of the form

$$\mathcal{C}(\alpha) \overset{def}{=} \{\beta \mid \alpha \sqsubseteq \beta\},$$

where $\alpha \sqsubseteq \beta$ means that α is a prefix of β. These sets of traces are called **cylinder sets** or **cones**.

For a concrete example, consider the case of Markov chains. The following construction is adapted from [1], but see also [2] and [5]. Let $M = (S, \tau, \ell)$ be a (discrete-time) Markov chain, where S is the state space, $\tau\colon S \to \mathcal{D}(S)$ is a transition probability function that for each state s returns a probability distribution on all states, representing the probability that s transitions to another given state, and $\ell\colon S \to 2^A$ is a labelling function that assigns a set of atomic propositions from the fixed (countable) set A to each state. Note that the labels are not of importance here, and can be omitted if necessary.

We define a **trace** on a Markov chain $M = (S, \tau, \ell)$ as an infinite sequence $\rho = s_0, s_1, s_2, \ldots$ such that $s_i \in S$ and $\tau(s_i)(s_{i+1}) > 0$ for all $i \in \mathbb{N}$. Let $\mathcal{T}(M)$ denote the set of all traces on M. For any $\rho \in \mathcal{T}(M)$, we further define $\rho|^n = s_0, s_1, s_2, \ldots, s_{n-1}, s_n$. Then, for any $\rho' = s_0, s_1, \ldots, s_{n-1}, s_n, s_{n+1}, \ldots$ we define the cylinder set of rank n as

$$\mathcal{C}_n(\rho') \overset{\text{def}}{=} \mathcal{C}(s_0, s_1, \ldots, s_{n-1}, s_n) = \{\rho \in \mathcal{T}(M) \mid \rho|^n = s_0, s_1, \ldots, s_{n-1}, s_n\}.$$

Let $\mathcal{C}_M$ denote the set of all cylinder sets for all ranks $n \in \mathbb{N}$, and let $\mathcal{C}_M^\sigma = \sigma(\mathcal{C}_M)$. Now, for any $s \in S$ we can define the **unique** probability measure $\mathbb{P}_s$ on $\mathcal{C}_M^\sigma$ as

$$\mathbb{P}_s(\mathcal{C}(s_0, s_1, \ldots, s_{n-1}, s_n)) = \mathbb{1}_{\{s\}}(s_0) \cdot \prod_{i=0}^{n-1} \tau(s_i)(s_{i+1})$$

for all $s_i \in S$. Intuitively, this means that the probability assigned to each cylinder set is the product of all the probabilities given by the transitions of the prefix for that cylinder set.

In Figure 1, part of a Markov chain is given. We can compute the probabilities of some of its cylinder sets as follows.

$$\begin{aligned}
\mathbb{P}_{s_1}(\mathcal{C}(s_1, s_2, s_4)) &= 0.7 \cdot 1 = 0.7 \\
\mathbb{P}_{s_1}(\mathcal{C}(s_1, s_2, s_3)) &= 0.7 \cdot 0 = 0 \\
\mathbb{P}_{s_1}(\mathcal{C}(s_1, s_3, s_4)) &= 0.3 \cdot 0.4 = 0.12 \\
\mathbb{P}_{s_1}(\mathcal{C}(s_1, s_3, s_5)) &= 0.3 \cdot 0.6 = 0.18 \\
\mathbb{P}_{s_2}(\mathcal{C}(s_1, s_3, s_5)) &= 0 \cdot 0.3 \cdot 0.6 = 0
\end{aligned}$$

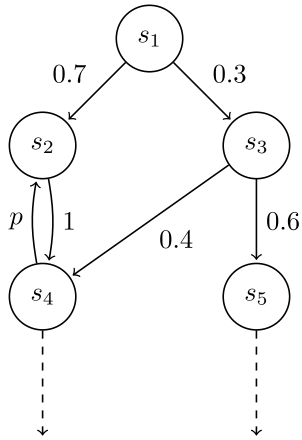

Figure 1: Part of a Markov chain. The dashed arrows indicate that the states may have additional transitions to states that are not pictured.

Note also that since s_4 has a possible transition back to s_2, we have that

$$\mathcal{C}(s_1, s_2, s_4, s_2, s_4) \subseteq \mathcal{C}(s_1, s_2, s_4),$$

since by choosing the p-transition back to s_2, we lose some possible traces, so $\mathcal{C}(s_1, s_2, s_4)$ contains more traces. Under what conditions are they equal? Clearly, if $p = 1$, then it is the only transition from s_4, and then

$$\mathcal{C}(s_1, s_2, s_4, s_2, s_4) = \mathcal{C}(s_1, s_2, s_4).$$

But since $p = 1$, we have

$$\mathbb{P}_{s_1}(\mathcal{C}(s_1, s_2, s_4, s_2, s_4)) = \mathbb{P}_{s_1}(\mathcal{C}(s_1, s_2, s_4)) = 0.7,$$

so the probability is still well-defined. In other words, this means that when the same cylinder set can be described by two different prefixes, these prefixes must have some determinism such that the resulting probabilities are still the same. ♦

For the familiar example of the real number line $\mathbb{R}$, things turn out to be more complicated than one might expect. Intuitively, we would like to define a measure μ on $\mathbb{R}$, such that

(a) μ is translation-invariant, i.e. for any set $A \subseteq \mathbb{R}$ and $r \in \mathbb{R}$, the sets A and $A + r \stackrel{def}{=} \{a + r \mid a \in A\}$ have the same measure,

(b) μ assigns a measure to all sets $A \subseteq \mathbb{R}$,

(c) μ assigns the measure $b - a$ to any interval $[a, b]$.

However, this turns out to be impossible. This result is sometimes known as the **Vitali theorem**.

Theorem 2.9 (Vitali theorem)**.** *Any measure on the real numbers $\mathbb{R}$ which satisfy (a) and (c) can not satisfy (b).*

Proof. Let μ be a measure on $\mathbb{R}$ which satisfies (a) and (c). Construct a relation $\mathcal{R}$ on $\mathbb{R}$ by

$$x\mathcal{R}y \text{ iff } x - y \text{ is rational.}$$

This relation is reflexive because $x - x = 0 \in \mathbb{Q}$ for all $x \in \mathbb{R}$, it is symmetric because $x - y \in \mathbb{Q}$ implies $-(x - y) = y - x \in \mathbb{Q}$, and it is transitive because $x - y \in \mathbb{Q}$ and $y - z \in \mathbb{Q}$ implies that $x - y + y - z = x - z \in \mathbb{Q}$. Hence, $\mathcal{R}$ is an equivalence relation. It is clear that for every $x \in \mathbb{R}$ there exists a $y \in (0, 1)$ such that $x\mathcal{R}y$. Now use the axiom of choice to pick from every equivalence class $[x]_{\mathcal{R}}$ one such $y \in (0, 1)$, and let E be the set of those chosen elements.

Define $E + r = \{x + r \mid x \in E\}$ for any $r \in \mathbb{Q}$, and note that if $r \neq r'$, then $E + r \cap E + r' = \emptyset$, i.e. the two sets are disjoint. For any $x \in (0, 1)$, there is a rational number $r \in (-1, 1) \cap \mathbb{Q}$ such that $x \in E + r$. To see this, note that if $x \in (0, 1)$, then there is an element $y \in E$ such that $x\mathcal{R}y$. Now let $r = x - y \in \mathbb{Q}$. Then $x = y + r \in E + r$, and since both $x, y \in (0, 1)$, we must have $r \in (-1, 1) \cap \mathbb{Q}$.

Now construct the countable union of disjoint sets

$$S = \bigcup_{r \in (-1,1) \cap \mathbb{Q}} E + r.$$

Assume towards a contradiction that S is measurable. Clearly $(0, 1) \subseteq S \subseteq (-1, 2)$, and therefore $1 = \mu((0, 1)) \leq \mu(S) \leq \mu((-1, 2)) \leq 3$ by (c) and monotonicity. However, if $\mu(E) = 0$, then

$$\mu(S) = \sum_{r \in (-1,1) \cap \mathbb{Q}} \mu(E + r) = \sum_{r \in (-1,1) \cap \mathbb{Q}} \mu(E) = \sum_{r \in (-1,1) \cap \mathbb{Q}} 0 = 0$$

by countable additivity and (a). Likewise, if $\mu(E) = \alpha \neq 0$, then

$$\mu(S) = \sum_{r \in (-1,1) \cap \mathbb{Q}} \mu(E) = \sum_{r \in (-1,1) \cap \mathbb{Q}} \alpha = \infty.$$

In either case, we obtain a contradiction, and hence S can not be measurable. ■

The canonical way to define a measure on the real numbers is known as the Lebesgue measure, but because of Vitali's theorem, there must be some subsets of $\mathbb{R}$ that are (Lebesgue) **non-measurable**. Thus, for the Lebesgue measure, we have to give up condition (b). Note that the existence of these non-measurable sets depend on the assumption of the axiom of choice. Because of these complications, we start with defining a so-called outer measure on all subsets of $\mathbb{R}$, and then restrict this to the σ-algebra on $\mathbb{R}$.

Definition 2.10 (Outer measure)**.** An **outer measure** on a set X is a function $\mu^* \colon 2^X \to [0, \infty]$ such that

(G1) $\mu^*(\emptyset) = 0$,

(G2) $A \subseteq B$ implies $\mu^*(A) \leq \mu^*(B)$ (monotonicity),

(G3) μ^* is countably subadditive.

▲

From an outer measure, we can construct a σ-algebra and a measure defined on it. Even though the outer measure is defined on all subsets of X, the σ-algebra we will construct, will in general not consist of all subsets of X, because we only want those subsets that behave nicely with respect to the outer measure and complements (cf. Definition 2.2). The construction is given by the following theorem.

Theorem 2.11. *Let X be a set and let μ^* be an outer measure defined on X. Denote by Σ the collection of all subsets $A \subseteq X$ which satisfy*

$$\mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)$$

for every $E \subseteq X$. Define $\mu(A) = \mu^*(A)$ for all $A \in \Sigma$. Then (X, Σ, μ) is a measure space.

Proof. We first wish to show that Σ as defined is indeed a σ-algebra. Our strategy for doing this will be to show that Σ is both a π-system and a λ-system. By Proposition 1.6, this will imply that Σ is a σ-algebra.

First we show that Σ is a π-system, so let $A, B \in \Sigma$, and let E be any subset of X. First we recall some set-theoretic identities.

$$A \cap B^c = (A \cap B)^c \cap A,$$

so

$$(A \cap B^c) \cap E = ((A \cap B)^c \cap A) \cap E = ((A \cap B)^c \cap E) \cap A.$$

Furthermore,

$$A^c \cap E = ((A \cap B)^c \cap A^c) \cap E = ((A \cap B)^c \cap E) \cap A^c.$$

Now, since $A \in \Sigma$, we know that for any subset E of X that

$$\begin{aligned}
\mu^*(E) &= \mu^*(A \cap E) + \mu^*(A^c \cap E) \\
&= \mu^*(A \cap B \cap E) + \mu^*(A \cap B^c \cap E) + \mu^*(A^c \cap E) \\
&= \mu^*(A \cap B \cap E) + \mu^*(((A \cap B)^c \cap E) \cap A) + \mu^*(((A \cap B)^c \cap E) \cap A^c) \\
&= \mu^*(A \cap B \cap E) + \mu^*((A \cap B)^c \cap E).
\end{aligned}$$

Hence, $A \cap B \in \Sigma$, so Σ is a π-system.

Next we show that Σ is a λ-system.

(B1) Clearly, $X \in \Sigma$ since

$$\mu^*(E) = \mu^*(X \cap E) + \mu^*(X^c \cap E) = \mu^*(X \cap E) + 0 = \mu^*(E)$$

for any $E \subseteq X$.

(B2) Σ is closed under complement because if $A \in \Sigma$ then

$$\mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E) = \mu^*(A^c \cap E) + \mu^*(A \cap E)$$

for all $E \subseteq X$, so $A^c \in \Sigma$.

(B3) Note first that if $A, B \in \Sigma$ and $A \cap B = \emptyset$, then

$$\begin{aligned} \mu^*((A \cup B) \cap E) &= \mu^*(A \cap (A \cup B) \cap E) + \mu^*(A^c \cap (A \cup B) \cap E) \qquad (1) \\ &= \mu^*(A \cap E) + \mu^*(B \cap E). \end{aligned}$$

Now assume that $\{A_i \mid i \in \mathbb{N}\}$ is a countable collection of pairwise disjoint sets in Σ. We wish to show that $\bigcup_{i=1}^{\infty} A_i \in \Sigma$. Since we have shown that Σ is closed under finite intersection and complement, it is also closed under finite union. Hence, for all $n \in \mathbb{N}$ we have

$$\mu^*(E) = \mu^*\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap E\right) + \mu^*\left(\left(\bigcup_{i=1}^{n} A_i\right)^c \cap E\right).$$

By (1), this means that

$$\mu^*(E) = \sum_{i=1}^{n} \mu^*(A_i \cap E) + \mu^*\left(\left(\bigcup_{i=1}^{n} A_i\right)^c \cap E\right).$$

Now note that $\mu^*((\bigcup_{i=1}^{n} A_i)^c \cap E) \supseteq \mu^*((\bigcup_{i=1}^{\infty} A_i)^c \cap E)$. Since μ^* is an outer measure and hence monotone, this means that

$$\mu^*(E) \geq \sum_{i=1}^{n} \mu^*(A_i \cap E) + \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c \cap E\right).$$

Since, as noted, this holds for any $n \in \mathbb{N}$, we get

$$\mu^*(E) \geq \sum_{i=1}^{\infty} \mu^*(A_i \cap E) + \mu^*\left(\left(\bigcup_{i=1}^{\infty}\right)^c \cap E\right). \qquad (2)$$

μ^* is also countably subadditive, since it is an outer measure, so

$$\begin{aligned} \mu^*(E) &\geq \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap E\right) + \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c \cap E\right) \\ &\geq \mu^*\left(\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap E\right) \cup \left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c \cap E\right)\right) \\ &= \mu^*(E). \end{aligned}$$

Putting all this together, we see that

$$\mu^*(E) \geq \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap E\right) + \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c \cap E\right) \geq \mu^*(E),$$

so we conclude that

$$\mu^*(E) = \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap E\right) + \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c \cap E\right),$$

which means that $\bigcup_{i=1}^{\infty} A_i \in \Sigma$.

Thus Σ is both a π-system and a λ-system, so it is a σ-algebra.

Finally, we need to show that μ^* is a measure when we restrict it to only sets in Σ. Let μ denote μ^* restricted to Σ.

(F1) $\mu(\emptyset) = 0$, since $\mu^*(\emptyset) = 0$ because μ^* is an outer measure.

(F2) Consider (2). The A_is in that equation constitute a countable collection of pairwise disjoint sets, and we have shown that the left-hand side and right-hand side are in fact equal. Hence, for any countable collection of pairwise disjoint sets we have

$$\mu^*(E) = \sum_{i=1}^{\infty} \mu^*(A_i \cap E) + \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c \cap E\right)$$

for any $E \subseteq X$. Letting $E = \bigcup_{i=1}^{\infty} A_i$, we get

$$\begin{aligned}
\mu^*(E) &= \sum_{i=1}^{\infty} \mu^*\left(A_i \cap \left(\bigcup_{i=1}^{\infty} A_i\right)\right) + \mu^*\left(\left(\bigcup_{i=1}^{\infty} A_i\right)^c \cap \left(\bigcup_{i=1}^{\infty} A_i\right)^c\right) \\
&= \sum_{i=1}^{\infty} \mu^*(A_i) + \mu^*(\emptyset) \\
&= \sum_{i=1}^{\infty} \mu^*(A_i).
\end{aligned}$$

This proves that μ is countably additive.

■

With these results, we can define the Lebesgue measure on $\mathbb{R}$. We first start by defining

$$m((a,b)) = m((a,b]) = m([a,b)) = m([a,b]) = b - a,$$

for any interval with $a, b \in \mathbb{R}$, thus satisfying (c). Next we define m for intervals. Let $\{I_k \mid k \in \mathcal{K}\}$ be a countable collection of pairwise disjoint intervals. Then we define

$$m\left(\bigcup_{k\in\mathcal{K}} I_k\right) = \sum_{k\in\mathcal{K}} m(I_k),$$

as required by the countable additivity condition. Note that for any collection, we can turn it into a pairwise disjoint collection as long as the collection is a semiring. Now we can define μ^* for arbitrary subsets of $\mathbb{R}$ as

$$\mu^*(A) = \inf\left\{ m\left(\bigcup_{k\in\mathcal{K}} I_k\right) \,\middle|\, I_k \text{ countable collection of intervals, } A \subseteq \bigcup_{k\in\mathcal{K}} I_k \right\}.$$

Then μ^* is an outer measure of all subsets of $\mathbb{R}$. Intuitively, $\mu^*(A)$ picks from all the possible interval coverings of A, and chooses the smallest one, i.e. the one that gives the smallest measure. Now, by Theorem 2.11, we can define Σ as all the subsets $A \subseteq \mathbb{R}$ such that

$$\mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E)$$

for every $E \subseteq \mathbb{R}$, and we can define $\mu(A) = \mu^*(A)$ for all $A \in \Sigma$. Then $(\mathbb{R}, \Sigma, \mu)$ is a measure space and μ is called the Lebesgue measure.

Example 2.12. Let $(\mathbb{R}, \Sigma, \mu)$ be the real numbers equipped with the Lebesgue measure. Then $\mu(\mathbb{Q}) = 0$. To see this, note that the rational numbers $\mathbb{Q}$ are countable. Hence, we can enumerate all the rational numbers q_i, and get $\mathbb{Q}$ as a countable union of all these

$$\mathbb{Q} = \bigcup_{i=0}^{\infty} \{q_i\}.$$

Notice also that the measure of any single number is zero, since a single number can be seen as an interval with the same left and right end-point

$$\mu(\{a\}) = \mu([a, a]) = a - a = 0.$$

Hence, it follows that

$$\mu(\mathbb{Q}) = \mu\left(\bigcup_{i=0}^{\infty} \{q_i\}\right) = \sum_{i=1}^{\infty} \mu(\{q_i\}) = \sum_{i=1}^{\infty} 0 = 0.$$

♦

The construction of the Lebesgue measure can be generalised to work for an arbitrary field of sets.

3 Measurable functions

If $f : X \to Y$ is a function and $B \subseteq Y$, then we denote by $f^{-1}(B)$ the **preimage** of B under f, i.e.

$$f^{-1}(B) \stackrel{\text{def}}{=} \{x \in X \mid f(x) \in B\}.$$

Preimages have many nice properties. In particular, they preserve complement and arbitrary union and intersection, i.e.

- $(f^{-1}(B))^c = f^{-1}(B^c)$,
- $\bigcup_{i \in I} f^{-1}(B_i) = f^{-1}(\bigcup_{i \in I} B_i)$, and
- $\bigcap_{i \in I} f^{-1}(B_i) = f^{-1}(\bigcap_{i \in I} B_i)$.

Definition 3.1 (Measurable function)**.** Given two measurable spaces (X, Σ_X) and (Y, Σ_Y), a function $f \colon X \to Y$ is said to be **measurable** if $f^{-1}(B) \in \Sigma_X$ for all $B \in \Sigma_Y$. ▲

☞ **[Topology]** Notice the similarity between the definitions of measurable function and continuous function.

☞ **[Category theory]** The category of measurable spaces **Meas** has measurable spaces as objects and measurable functions as morphisms.

☞ **[Topology]** If we take an arbitrary topological space (X, τ), and consider the measurable space $(X, \mathbb{B}(X, \tau))$ with the Borel σ-algebra, then clearly any continuous function is also measurable, but the converse is not true in general.

Example 3.2. If (X, Σ_X) and (Y, Σ_Y) are measurable spaces where $\Sigma_Y = \{\emptyset, Y\}$ is the trivial σ-algebra, then any function $f : X \to Y$ is measurable, regardless of how f and Σ_X are defined. To see this, note that it will always be the case that $f^{-1}(\emptyset) = \emptyset$. Because preimages preserve complement, this also means that $f^{-1}(Y) = f^{-1}(\emptyset^c) = \emptyset^c = X$. Hence $f^{-1}(\emptyset) = \emptyset \in \Sigma_X$ and $f^{-1}(Y) = X \in \Sigma_X$, so f is measurable. ♦

The following proposition gives a way to generate a σ-algebra on any set by considering a function (not necessarily measurable) from that set to a measurable space.

Proposition 3.3. *Let X be a set and (Y, Σ_Y) be a measurable space and let $f\colon X \to Y$ be a function. Then the set*

$$f^{-1}(\Sigma_Y) \stackrel{def}{=} \{f^{-1}(B) \mid B \in \Sigma_Y\}$$

is a σ-algebra on X.

Proof. We need to verify properties (E1)-(E3).

(E1) Since f is a function, we have $f^{-1}(Y) = X$, and since Σ_Y is a σ-algebra, $Y \in \Sigma_Y$, so $X \in f^{-1}(\Sigma_Y)$.

(E2) Take an arbitrary $A \in f^{-1}(\Sigma_Y)$. Then there must exist some $B \in \Sigma_Y$ such that $A = f^{-1}(B)$. Now,

$$A^c = X \setminus A = f^{-1}(Y) \setminus f^{-1}(B) = f^{-1}(Y \setminus B) = f^{-1}(B^c),$$

and since Σ_Y is a σ-algebra, it follows that $B^c \in \Sigma_Y$, and hence $f^{-1}(B^c) = A^c \in f^{-1}(\Sigma_Y)$.

(E3) Let $A_1, A_2, A_3, \dots \in f^{-1}(\Sigma_Y)$ be a countable collection of sets. Then, for every A_i, there must exist some $B_i \in \Sigma_Y$ such that $A_i = f^{-1}(B_i)$. Hence

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} f^{-1}(B_i) = f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right),$$

and since Σ_Y is a σ-algebra, $\bigcup_{i=1}^{\infty} B_i \in \Sigma_Y$, so $\bigcup_{i=1}^{\infty} A_i \in f^{-1}(\Sigma_Y)$.

■

Proposition 3.4. *Let X and Y be sets and let $f\colon X \to Y$ be a function. Suppose $\mathcal{G}$ is a collection of subsets of Y. Define*

$$f^{-1}(\mathcal{G}) \stackrel{def}{=} \{f^{-1}(A) \mid A \in \mathcal{G}\}.$$

Then

$$f^{-1}(\sigma(\mathcal{G})) = \sigma(f^{-1}(\mathcal{G})).$$

Proof. First we show that $f^{-1}(\sigma(\mathcal{G})) \supseteq \sigma(f^{-1}(\mathcal{G}))$. To do this, we show that $f^{-1}(\sigma(\mathcal{G}))$ is a σ-algebra and that it contains $f^{-1}(\mathcal{G})$. To show that it contains $f^{-1}(\mathcal{G})$, observe that $\mathcal{G} \subseteq \sigma(\mathcal{G})$, which implies $f^{-1}(\mathcal{G}) \subseteq f^{-1}(\sigma(\mathcal{G}))$. To show that it is a σ-algebra, we verify properties (E1)-(E3).

(E1) Since $\emptyset = f^{-1}(\emptyset)$ and $\emptyset \in \sigma(\mathcal{G})$, we have that $\emptyset \in f^{-1}(\sigma(\mathcal{G}))$.

(E2) Assume $A \in f^{-1}(\sigma(\mathcal{G}))$. Then there must exist some $G \in \sigma(\mathcal{G})$ such that $A = f^{-1}(G)$. Since $\sigma(\mathcal{G})$ is closed under complement, we have $G^c \in \sigma(\mathcal{G})$, implying that $f^{-1}(G^c) \in f^{-1}(\sigma(\mathcal{G}))$. Hence

$$f^{-1}(G^c) = f^{-1}(Y \setminus G) = f^{-1}(Y) \setminus f^{-1}(G) = X \setminus A = A^c \in f^{-1}(\sigma(\mathcal{G})).$$

(E3) Assume $A_1, A_2, \dots \in f^{-1}(\sigma(\mathcal{G}))$. Then for every A_i there must exist some $G_i \in \sigma(\mathcal{G})$ such that $A_i = f^{-1}(G_i)$. Since $\sigma(\mathcal{G})$ is closed under countable union, we have $\bigcup_i G_i \in \sigma(\mathcal{G})$, implying that $f^{-1}(\bigcup_i G_i) \in f^{-1}(\sigma(\mathcal{G}))$. Thus we get

$$f^{-1}(\bigcup_i G_i) = \bigcup_i f^{-1}(G_i) = \bigcup_i A_i \in f^{-1}(\sigma(\mathcal{G})).$$

Next we show that $f^{-1}(\sigma(\mathcal{G})) \subseteq \sigma(f^{-1}(\mathcal{G}))$. First we define

$$D \stackrel{\texttt{def}}{=} \{G \subseteq Y \mid f^{-1}(G) \in \sigma(f^{-1}(\mathcal{G}))\}.$$

Observe that what we want to show will follow if we prove that D is a σ-algebra and that $\mathcal{G} \subseteq D$. This is because it will imply that $\sigma(\mathcal{G}) \subseteq D$, which, by the definition of D, will imply that if $G \in \sigma(\mathcal{G})$, then $f^{-1}(G) \in \sigma(f^{-1}(\mathcal{G}))$. This further implies that

$$f^{-1}(\sigma(\mathcal{G})) = \{f^{-1}(G) \mid G \in \sigma(\mathcal{G})\} \subseteq \sigma(f^{-1}(\mathcal{G})).$$

Hence, we now show that D is a σ-algebra and that $\mathcal{G} \subseteq D$. First observe that $f^{-1}(\mathcal{G}) \subseteq \sigma(f^{-1}(\mathcal{G}))$. This implies that if $G \in \mathcal{G}$ then $f^{-1}(G) \in \sigma(f^{-1}(\mathcal{G}))$. By the definition of D, this further implies that $\mathcal{G} \subseteq D$. To show that D is a σ-algebra, we verify properties (E1)-(E3).

(E1) Since $\emptyset = f^{-1}(\emptyset)$ and $\emptyset \in \sigma(f^{-1}(\mathcal{G}))$, we have $\emptyset \in D$.

(E2) Assume $G \in D$. Then $f^{-1}(G) \in \sigma(f^{-1}(\mathcal{G}))$. Since $\sigma(f^{-1}(\mathcal{G}))$ is closed under complement, we get

$$X \setminus f^{-1}(G) = f^{-1}(Y) \setminus f^{-1}(G) = f^{-1}(Y \setminus G) = f^{-1}(G^c) \in \sigma(f^{-1}(\mathcal{G})),$$

which implies that $G^c \in D$.

(E3) Assume $G_1, G_2, \dots \in D$. Then for all G_i we have $f^{-1}(G_1) \in \sigma(f^{-1}(\mathcal{G}))$. Since $\sigma(f^{-1}(\mathcal{G}))$ is closed under countable union, we get

$$\bigcup_i f^{-1}(G_i) = f^{-1}(\bigcup_i G_i) \in \sigma(f^{-1}(\mathcal{G})),$$

which implies that $\bigcup_i G_i \in D$.

■

References

[1] Gorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On the total variation distance of semi-markov chains. **Foundations of Software Science and Computation Structures (FoSSaCS)**, 2015.

[2] Christel Baier and Holger Hermanns. Weak bisimulation for fully probabilistic processes. Technical Report TR-CTIT-99-12, University of Twente, Centre for Telematics and Information Technology (CTIT), Enschede, the Netherlands, 1999. `http://doc.utwente.nl/18158/`.

[3] Radu Mardare. **Logical Foundations of Metric Behavioural Theory for Markov Processes**. Aalborg University Press, 2015.

[4] Prakash Panangaden. **Labelled Markov Processes**. Imperial College Press, 2009.

[5] Roberto Segala. Probability and nondeterminism in operational models of concurrency. In **CONCUR 2006 - Concurrency Theory, 17th International Conference, CONCUR 2006, Bonn, Germany, August 27-30, 2006, Proceedings**, pages 64–78, 2006. `http://dx.doi.org/10.1007/11817949_5`.