Important concepts from today: probability measure; measurable function; random variable.

1. Measure

Definition 1 (Measure). A measure on a set $S$ is a map $m$ assigning to some subsets $A \subseteq S$ nonnegative real numbers, or infinity: $m(A) \in \mathbb{R}_+ \cup \{\infty\}$. The collection $\mathcal{F}$ of subsets of $S$ on which $m$ is defined is called the $\sigma$-field (or $\sigma$-algebra) of measurable sets. A $\sigma$-field $\mathcal{F}$ is a collection of sets satisfying the following properties:

1. The empty set $\emptyset \in \mathcal{F}$.
2. If $A \in \mathcal{F}$, then its complement $A^c \in \mathcal{F}$. The complement is $S \setminus A = \{x : x \notin A\}$.
3. If countably many sets $A_1, A_2, \ldots \in \mathcal{F}$, then also their union $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ and intersection $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

A measure is a map $m : \mathcal{F} \to \mathbb{R}$ satisfying the following properties:

1. The empty set $\emptyset$ has measure 0: $m(\emptyset) = 0$.
2. If $A_i \in \mathcal{F}$ are disjoint, that is, $A_i \cap A_j = \emptyset$ whenever $i \neq j$, then

$$m(\bigcup A_i) = \sum A_i.$$

This property is called $\sigma$-additivity.

If $m(S) = 1$, we say that $m$ is a probability measure, and we say that the triple $(S, \mathcal{F}, m)$ is a probability space.

We will often use the letters $\mu, \nu$ for measures. For probability spaces we will often use $\Omega$; for probability measures we will often use $\mathbb{P}$ or sometimes $Q$.

An interlude on countable sets:

Definition 2. A set is countable if there is a one-to-one map from it to the natural numbers. That is, I can assign different natural indices to the elements of the set.

- Any finite set is countable.
- The natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$ are countable.
- Any subset of a countable set is countable.
- Any nonempty open interval on the real line is uncountable.
- Any set that contains an uncountable set is itself uncountable. In particular, the real numbers $\mathbb{R}$, the $n$-dimensional space $\mathbb{R}^n$, the complex numbers $\mathbb{C}$ are all uncountable.
Exercise 1 (In-class exercise). Which of the following are countable:

1. \( \{1, 2, \ldots, 10\} \)
2. \{all sheep\}
3. \( \{3, 5\} \)
4. \( \{f : \mathbb{R} \to \mathbb{R}, \ f \text{ continuous}\} \)
5. \( \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \)

(Answer: the first two are countable, the others aren’t.)

Exercise 2. Prove these statements, or at least look up their proofs.

Definition 3. The \( \sigma \)-field \( \text{generated} \) by a collection of sets \( S \) is the smallest \( \sigma \)-field that contains those sets. We can try to build it by starting with \( S \) and adding in all countable unions and intersections of elements of \( S \); then all countable unions and intersections of elements of \( S \) and the stuff we just added; and so on. Beware that the process doesn’t have to end after a fixed number of steps if the set you’re working with is uncountable!

To define a measure, it’s enough to define it on the generators of a \( \sigma \)-field.

Example 1. (1) \( S = \{1, 2, \ldots, 6\} \), with \( \mathcal{F} = 2^S = \text{all subsets of } S \). Because \( S \) is finite, \( \mathcal{F} \) is generated by the one-element sets. So to define \( m \), it is enough to define \( m(\{i\}) \) for each \( i \in S \). We can put different measures on \( (S, \mathcal{F}) \):

(a) We may take \( m(\{i\}) = 1/6 \) for all \( i \) (fair die),
(b) or \( m(\{6\}) = 0.5 \) and \( m(\{i\}) = 0.1 \) for all other \( i \) (heavily loaded die that often comes up 6).
(c) We can also take \( m(\{i\}) = 1 \); then for any set \( A \subseteq S \), \( m(A) \) is the number of elements in \( A \).

2. \( S = \{1, 2, \ldots, 6\} \), but only the sets \( \emptyset, \{2, 3, 5\}, \{1, 4, 6\}, \) \( S \) are measurable \( (\mathcal{F} \neq 2^S) \). If \( m \) is a probability measure, it will be enough to define \( m(\{2, 3, 5\}) \). Note that \( m(\{2\}) \) is not defined.

3. \( S = \mathbb{N} \), with \( \mathcal{F} = 2^S \). Because \( S \) is countable, it’s enough to define the measure of every point, e.g. \( m(\{n\}) = 1/n \).

4. The Lebesgue measure on \( \mathbb{R} \): \( S = \mathbb{R} \), \( m([a, b]) = m([a, b]) = b - a \). The \( \sigma \)-algebra of measurable sets is generated by the open intervals, and is called the Borel \( \sigma \)-algebra. Here, \( \mathcal{F} \neq 2^\mathbb{R} \): there exist non-measurable sets (although it’s hard to give an explicit example). The measure of a set is its “length”. Observe that \( m(\{x\}) = m([x, x]) = 0 \) for any \( x \in S \), but there are sets (consisting of uncountably many points) with non-zero measure.

By \( \sigma \)-additivity, \( m(A) = 0 \) for any countable set \( A \): so \( m(\mathbb{Z}) = m(\mathbb{Q}) = 0 \).

5. The uniform probability measure on an interval, or on a subset of \( \mathbb{R}^n \): it’s proportional to the Lebesgue measure (length, area, volume) but the total measure of the set is 1. For example, if the measure \( \mu \) is uniform on the interval \( (a, b) \), then \( \mu(A) = m(A)/(b-a) \).

6. If \( (S, \mathcal{F}, \mu) \) and \( (T, \mathcal{G}, \nu) \) are measure spaces, then \( S \times T \) is a measure space. The measurable sets are generated by \( \mathcal{F} \times \mathcal{G} \); the measure is given by \( m(A \times B) = \mu(A)\nu(B) \) for one of the basic sets, and then whatever it has to be for the remaining measurable sets.

(a) \( \mathbb{R}^2 = \mathbb{R} \times \mathbb{R} \) is a measure space with the product Lebesgue measure. \( \mathcal{F} \times \mathcal{G} \) will consist of rectangles, but a circle is also measurable (you can fill it up with countably many rectangles). The measure of a set is its 2-dimensional area, which is length times width for a rectangle.

Aside: The Banach-Tarski paradox tells you that there are non-measurable sets in \( \mathbb{R}^3 \).
(b) $S = \{\text{sequences of four } U\text{'s or } D\text{'s}\} = \{U, D\}^4$. If $\mu(\{U\}) = p$ and $\mu(\{D\}) = 1 - p$, then $m(\{UUDD\}) = p^2(1-p)$, and

$m(\{UUDD, UDDU, UDUD, DUUD, DUUU\}) = 6p^2(1-p)^2$.

(c) $S = \{\text{infinite sequences of } U\text{'s and } D\text{'s}\} = \{U, D\}^\infty$. In the case of an infinite product, the measurable sets are generated by sets of the form $A_1 \times \ldots \times A_k \times (\text{entire space in all the remaining components})$. For example,

$m(\{\text{sequences that start with 2 } D\text{'s and 2 } U\text{'s}\}) = 6p^2(1-p)^2$.

We can represent a particular infinite sequence as a countable intersection of its initial segments, so we can find its measure; usually it’ll be 0.

(d) $S = \{f : \mathbb{R} \to \mathbb{R}\} = \mathbb{R}^\mathbb{R}$, all functions from $\mathbb{R}$ to itself. Measurable sets are generated by saying something about the function at finitely many points; when we look at countable intersections, we can say something at countably many points. The set of functions that are continuous everywhere is nonmeasurable, because “everywhere” is uncountable.

A probability measure is a set of weights: when you integrate a function against it, you’re computing that function’s weighted average value, or mean. If the measure is discrete, the “integral” is a sum (with weights). If the measure does not integrate to 1, then integrating against it is like computing a weighted sum rather than a weighted average.

**Exercise 3.** Let $S = \mathbb{N}$, and let $m$ be given by $m(\{n\}) = 1/n$ for $n \geq 0$. What is the measure of the set of even integers $A = 2\mathbb{N}$?

**Solution:**

$m(\{2, 4, 6, \ldots \}) = m(\{2\}) + m(\{4\}) + m(\{6\}) + \ldots = \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \ldots = \infty$.

**Exercise 4.** Let $S = \mathbb{R}$, and let $m$ be given by $m([a, b]) = \int_a^b 1_{[0, 1]}(x)dx$. What is the measure of the set $A = \{x : x^2 < \frac{1}{2}\}$?

**Solution:** The indicator function $1_{[0, 1]}(x)$ is equal to 1 if $x \in [0, 1]$ and to 0 otherwise.

The set $A$ takes the form of an open interval, $A = (-1/\sqrt{2}, 1/\sqrt{2})$. Just as for the Lebesgue measure, it doesn’t matter whether or not we include the endpoints of $A$, because $m(\{a\}) = 0$ for any single point $a$. We compute

$m(A) = \int_{-1/\sqrt{2}}^{1/\sqrt{2}} 1_{[0, 1]}(x)dx = \int_0^{1/\sqrt{2}} 1dx = \frac{1}{\sqrt{2}}$.

2. Random variable

**Definition 4 (Measurable function).** Let $(S, \mathcal{F}, \mu)$ and $(T, \mathcal{G}, \nu)$ be two measure spaces. A function $f : S \to T$ is called **measurable** if the preimages of $\nu$-measurable sets are $\mu$-measurable. That is, if $B \in \mathcal{G}$, then $f^{-1}(B) \in \mathcal{F}$.

If $T = \mathbb{R}$, we almost always consider the Borel $\sigma$-algebra, so we require $f^{-1}(B)$ to be measurable for every Borel set $B$. It’s enough to require that $f^{-1}(B)$ is measurable for every open set $B$. In particular, if $f : S \to \mathbb{R}$ is measurable, then for every $y \in \mathbb{R}$, we must have $f^{-1}(y) \in \mathcal{F}$.

Why is this a sensible definition? You should think of the $\sigma$-field as a way of structuring information: when you look at an element of $S$ or $T$, you don’t get to see the element itself, you get to see only what the $\sigma$-field tells you about it. We are saying that “information
about $x \in S$ lets me determine information about $f(x) \in T$. If the preimages of points aren’t elements of $\mathcal{F}$, there might be different $x \in S$ that are indistinguishable based on the $\sigma$-field of information, but which map to different real numbers.

(“Measurable” is really a misnomer here, since we don’t care about the measures $\mu$ and $\nu$ yet, only about their $\sigma$-fields.)

**Example 2.** (1) When all subsets of $S$ are measurable ($\mathcal{F} = 2^S$), then all functions $f : S \to T$ are measurable. (The preimage of anything is a subset of $S$, and therefore measurable.)

(2) Suppose $S = \{1, 2, \ldots, 6\}$, and $\mathcal{F} = \{\emptyset, \{2, 3, 5\}, \{1, 4, 6\}, S\}$. Then any measurable function $f : S \to \mathbb{R}$ must be constant on $\{2, 3, 5\}$ and on $\{1, 4, 6\}$. This is because I can’t distinguish elements of these sets based on the information in $\mathcal{F}$. From the definition, if $x = f(2)$, then $2 \in f^{-1}\{x\}$, so $\{2, 3, 5\} \subseteq f^{-1}\{x\}$.

(3) Note that I can have different measures and $\sigma$-algebras on the same underlying set $S$, so the same function $f : S \to T$ may or may not be measurable depending on $\mathcal{F}$.

(4) If $S = \mathbb{R}$ and $m$ is the Lebesgue measure, then any continuous function is measurable; but many measurable functions are not continuous. For example, $f(x) = \lfloor x \rfloor$ is measurable: the preimage of any set is a subset of $\mathbb{R}$. It is quite hard to write down a function $f : \mathbb{R} \to \mathbb{R}$ that is not measurable.

**Definition 5 (Random variable).** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A real-valued random variable is a measurable map $X : \Omega \to \mathbb{R}$. For any set $S$ with $\sigma$-algebra $\mathcal{G}$, an $S$-valued random variable is a measurable map $X : \Omega \to S$.

For $A \subset \mathbb{R}$ (or $A \subset S$), when we write $\mathbb{P}(X \in A)$, we mean $\mathbb{P}\{\omega \in \Omega : X(\omega) \in A\}$.

Notice that $\mathbb{P}$ is a measure on $\Omega$, so we have to feed a subset of $\Omega$ to it!

In real life, we can think of $\omega \in \Omega$ as the “state of the universe”, and $X(\omega)$ is some observable quantity. For example, in rolling a die, $\omega$ is the set of conditions (how hard I threw the die, how I held it, what is the die made of, what’s the desk made of, what are the air currents in the room, ...) and if I knew $\omega$ exactly, I should be able to determine how the die is going to land. The probability measure on $\Omega$ quantifies our beliefs about which states of the world are more likely than others.

On the other hand, we can also think of $\Omega$ for this example as the set of numbers $\{1, \ldots, 6\}$, which are about equally likely.

**Definition 6 ($\sigma(X)$ and induced probability).** Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ be a random variable. We get a measure on $\mathbb{R}$ from this:

$$\mu(A) = \mathbb{P}(X \in A) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\}.$$  

This is called the induced measure. It’s not equal to the Lebesgue measure, since $m(\mathbb{R}) = \infty$ but $\mu(\mathbb{R}) = 1$.

We also get a $\sigma$-field $\sigma(X) = \{X^{-1}(B)\}$. This encapsulates the information I can tell about events based only on information about the value of $X$. This usually tells me less information than the original $\sigma$-field.

**Example 3.** (1) Let $S = \{U, D\}^4$, and let $X : S \to \mathbb{R}$ count the number of $U$’s in the sequence. (The range of $X$ is thus the discrete set $\{0, \ldots, 4\}$.) Suppose the measure on $S$ is the product measure, with $\mu(\{U\}) = p$ and $\mu(\{D\}) = 1 - p$. 


The induced measure – now on the set of numbers \{0, \ldots, 4\} and not on the set of sequences of Us and Ds – is

\[ m(\{0\}) = (1 - p)^4, \quad m(\{1\}) = 4p(1 - p)^3, \quad m(\{2\}) = 6p^2(1 - p)^2, \]

\[ m(\{3\}) = 4p^3(1 - p), \quad m(\{4\}) = p^4. \]

If \( p = 0.5 \), the CDF of \( m \) looks like this:

\[
\begin{array}{c}
1 \quad 2 \quad 3 \quad 4 \\
\bullet \quad \circ \quad \bullet \quad \circ \\
\end{array}
\]

The algebra \( \sigma(X) \) has 5 sets: \{DDDD\}, \{DDDU, DUDU, UDDD, UDDB\}, \{sequences with 2 U’s\}, etc. If \( Y : S \to \mathbb{R} \) has \( Y(UUDU) \neq Y(DUUU) \) then \( Y \) is not \( \sigma(X) \)-measurable.

(2) If \( S = \{U, D\}^\mathbb{N} \) and \( X : S \to \{U, D\}^2 \) takes the first two elements of the sequence, then \( \sigma(X) \) has 4 sets. If \( Y(UUDUUDD\ldots) \neq Y(UUDUDD\ldots) \) then \( Y \) is not \( \sigma(X) \)-measurable.

**Definition 7** (Cumulative distribution function). For a probability measure \( m \) on \( \mathbb{R} \), its cumulative distribution function (CDF) is the increasing function

\[ F_m(x) = m((\infty, x]). \]

For a random variable \( X : \Omega \to \mathbb{R} \), the cumulative distribution function of (the law of) \( X \) is

\[ F_X(x) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in (\infty, x]\}) \]

\[ = \mathbb{P}(X \leq x). \]

We often call this the CDF of \( X \) itself, omitting the “law of”.

Notice that \( F_m \) is always increasing (or at least nondecreasing), since it is looking at the measure of larger and larger sets.