1 From probabilistic problems to probability spaces

1a Two approaches

1a1 Problem. Two friends come between 6:00 and 7:00 at random. Each one waits for the other at most 20 min. Find the probability of their meeting.

Here is a dialogue between two persons, D (prefers discrete to continuous) and C (prefers continuous to discrete).

D: One friend comes at $6:X$, the other at $6:Y$, where $X, Y \in \{0, 1, \dots, 59\}$. There are $60 \cdot 60 = 3600$ possibilities. The following possibilities are successful:

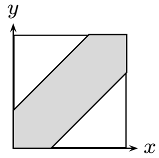
$$\begin{array}{lll} X = 0, & Y \in \{0, \dots, 20\}; & = 21 \\ X = 1, & Y \in \{0, \dots, 21\}; & = 22 \\ \dots & \dots & \dots \\ X = 19, & Y \in \{0, \dots, 39\}; & = 40 \\ X = 20, & Y \in \{0, \dots, 40\}; & = 41 \\ X = 21, & Y \in \{1, \dots, 41\}; & = 41 \\ \dots & \dots & \dots \\ X = 39, & Y \in \{19, \dots, 59\}; & = 41 \\ X = 40, & Y \in \{20, \dots, 59\}; & = 40 \\ \dots & \dots & \dots \\ X = 58, & Y \in \{38, \dots, 59\}; & = 22 \\ X = 59, & Y \in \{39, \dots, 59\}; & = 21 \end{array}$$

In all, $2 \cdot (21 + 22 + \dots + 40) + 20 \cdot 41 = 2040$. The probability:

$$p = \frac{2040}{3600} = \frac{17}{30} = 0.566\dots$$

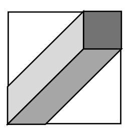
C: Why use minutes? Maybe, seconds? Rather, I use hours and all their parts: $X, Y \in (0, 1)$. All points (x, y) of the square $[0, 1) \times [0, 1)$ are possibilities. Successes satisfy

$$|x - y| \le 1/3\,.$$



The area is equal to

$$2 \cdot \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} = \frac{4}{9} + \frac{1}{9} = \frac{5}{9}\,.$$



The probability:

$$p = \frac{5/9}{1} = \frac{5}{9} = 0.555\dots \tag{1a2}$$

D: You compare infinite sets via their areas. I doubt that it is correct.

C: You replace the continuum with a finite set (of 60 points). It is surely incorrect, since your result (17/30) depends on the parameter (60) chosen at will.

D: Well, I can take $3n$ points (not just 60). Then I get $9n^2$ possibilities; among them, the following number of successes:

$$\underbrace{2\big((n+1)+\dots+(n+n)\big)}_{n(3n+1)} + n(2n+1) = n(5n+2)\,.$$

The discrete probability is

$$p_n = \frac{n(5n+2)}{9n^2} = \frac{5}{9} + \frac{2}{9n}\,.$$

The continuous probability is the limit,

$$p = \lim_{n\to\infty} p_n = \frac{5}{9}\,. \tag{1a3}$$

C: But I got the same without limits.

D: And I got it without areas.

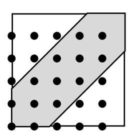
1b Area as a limit

Relations between the two approaches of Sect. 1a may be understood as follows. We have a square $\Omega = [0,1) \times [0,1)$ and its subset $A \subset \Omega$, $A = \{(x,y) \in \Omega : |x-y| \le 1/3\}$. For an arbitrary $n \in \{1,2,\dots\}$ we consider the lattice[1]

$$\frac{1}{n}\mathbb{Z}^2 = \{(\tfrac{k}{n}, \tfrac{l}{n}) : k,l \in \mathbb{Z}\}\,,$$

the number of lattice points in A,

$$\#\big(A \cap \tfrac{1}{n}\mathbb{Z}^2\big)\,,$$



the fraction

$$P_n(A) = \frac{\#\big(A \cap \frac{1}{n}\mathbb{Z}^2\big)}{\#\big(\Omega \cap \frac{1}{n}\mathbb{Z}^2\big)}$$

and its limit

$$P(A) = \lim_{n\to\infty} P_n(A)\,,$$

[1] $\mathbb{Z} = \{\dots,-2,-1,0,1,2,\dots\}$.

if it exists.

If A is a polygon (or the union of several polygons) then the limit $P(A)$ exists and is equal to the area of A (as defined by elementary geometry).

In general, geometry defines the *outer area* of A as the infimum of areas of all polygons containing A; the *inner area* of A as the supremum of areas of all polygons contained in A; if the inner and outer areas coincide, their common value is called the (Jordan) *area* of A, and A is called *Jordan measurable.* Otherwise, the inner area is strictly smaller than the outer area, A is not Jordan measurable, and Jordan area of A is not defined.

Every domain with a piecewise smooth boundary is Jordan measurable.

Jordan measurable sets are an algebra of sets (as defined in 1e).

For every Jordan measurable set A, the limit $P(A)$ exists and is equal to the area of A.

Three-dimensional Jordan measurable sets $A \subset \mathbb{R}^3$ are introduced similarly; the volume is used rather than the area. More generally, $A \subset \mathbb{R}^d$ for any $d = 1, 2, 3, 4, \dots$ may be considered.

1c Possible and impossible

Recall that discrete probability calls an event A impossible, if its probability vanishes, $P(A) = 0$.

A singleton[2] is Jordan measurable, and its area is 0. However, the union of all singletons is the whole Ω. We do not want to say that a singleton is an impossible event, since one of such events must occur. Instead we say that

$$\left.\begin{array}{l} A \text{ is negligible,} \\ A \text{ occurs almost never,} \\ \text{almost surely, } A \text{ does not occur} \end{array}\right\} \text{ when } P(A) = 0\,;$$

$$\left.\begin{array}{l} A \text{ is almost certain,} \\ A \text{ occurs almost always,} \\ \text{almost surely, } A \text{ occurs} \end{array}\right\} \text{ when } P(A) = 1\,.$$

1c1 Problem. Two friends come between 6:00 and 7:00 at random. Can it happen that they come simultaneously?

C: No. The set $\{(x, y) : x = y\}$ is of zero area;[3] a random point has no chance to hit it.

D: Indeed, only a mathematician can give such an answer, rigorous and irrelevant! My answer is "yes". Let $A = \{(x, y) \in [0, 1) \times [0, 1) : x = y\}$, then

$$\#\bigl(A \cap \tfrac{1}{n}\mathbb{Z}^2\bigr) = \#\{(0,0), (\tfrac{1}{n}, \tfrac{1}{n}), (\tfrac{2}{n}, \tfrac{2}{n}), \dots, (\tfrac{n-1}{n}, \tfrac{n-1}{n})\} = n\,;$$

$$P_n(A) = \frac{n}{n^2} = \frac{1}{n}\,.$$

[2] That is, a set of one point.

[3] You see, it is a line.

The probability is positive, therefore, the event is possible.

C: However, $\lim_{n\to\infty} P_n(A) = 0$.

D: So what?

C: You did not say, what is the probability of the event. You cannot say just "$1/n$", since there is no "n"in the formulation of the problem. The parameter n is your trick. You must take the limit for $n \to \infty$.

D: In real life, there is always a finite resolution. You cannot say that friends do not come simultaneously since one friend comes 1 microsecond after the other. Therefore you cannot say that $P(A)$ is less than $\frac{1}{1\,000\,000} \cdot \frac{1}{60} \cdot \frac{1}{60}$.

C: If you apply mathematics to real life, you need some idealization. A resolution is not given in the formulation, therefore I use continuum.

D: Therefore you say "no" while the right answer is "yes". You may neglect the difference between $\frac{1}{1\,000\,000} \cdot \frac{1}{60} \cdot \frac{1}{60}$ and 0, but you may not neglect the distinction between "yes" and "no'.

C: And you may not neglect the distinction between reality and mathematics.

MATHEMATICAL PROBLEMS ARE FORMAL. I may ask you questions about coins tossed, friends coming etc., but never forget: I am a mathematician; my questions are mathematical (unless otherwise stated); translate all these coins and friends into mathematical language, that is, USE A MATHEMATICAL MODEL, then answer my questions.

Do not be confused by two different approaches; one is AN INFINITE SEQUENCE OF FINITE MODELS, the other A SINGLE INFINITE MODEL. Both are very useful; confront but do not confuse one with the other. An example was given in Sect. 1a; the same result was reached by C via a single infinite model (recall (1a2)), and by D via an infinite sequence of finite models (recall (1a3)).

The present course is mostly about CONTINUOUS PROBABILITY. (Discrete probability was the matter of "Introduction to probability" course.) Thus, I mean a SINGLE INFINITE (CONTINUOUS) MODEL, unless otherwise stated.

Do not say that the probability of a single point *tends to* 0; leave it to the other approach. Say that THE PROBABILITY OF A SINGLE POINT IS EQUAL TO 0; this is the typical situation for continuous probability.

CAN WE OBSERVE AN EVENT OF ZERO PROBABILITY? In real life we cannot, but in the theory we can. Why? Since points of zero size, infinite collections of coins etc. exist in the theory but not in the real life.

HOW TO OBSERVE AN EVENT OF ZERO PROBABILITY? It is easy. Choose at random a point within the square.[4] Ask yourself, what was the chance of this specific point to be chosen. It was equal to zero.

D: So, the continuous probability is absurd.

C: Why?

D: The only bridge between probability theory and its real-life applications is the claim that events of very small probability do not occur in practice. For example: if I toss a fair coin 1000 times, I am pretty sure that I'll not get 1000 "heads"; it is too improbable.

[4]Or the interval $(0, 1)$, if you like.

Moreover, the number of "heads" will not escape the interval $[400, 600]$. In contrast, you say that even an event of zero probability occurs quite easily. Maybe your mathematics is correct, but your philosophy is inconsistent.

C: Just a moment. You say, the sequence H^{1000} of 1000 "heads" is improbable. What about the sequence, say, $(HT)^{500} = HTHT \dots HT$ of 500 pairs "head, tail"?

D: The same. Still, the probability is $2^{-1000} \approx 10^{-300}$; more than improbable.

C: Give me an example of a more probable sequence.

D: There is no such example. Every single sequence (of length 1000) is of probability 2^{-1000}.

C: So, every single sequence is improbable. Now, can you toss a coin 1000 times?

D: No problem.

C: Do it, and you'll get an improbable sequence! Continuous probability is not at all responsible for the absurdity.

D: What is responsible?

C: Two words missing in your phrase "events of very small probability do not occur in practice". Rather, you should say: A SINGLE PREDICTED EVENT OF VERY SMALL PROBABILITY DOES NOT OCCUR IN PRACTICE.

D: I see. And the theoretical counterpart is, A PREDICTED EVENT OF ZERO PROBABILITY DOES NOT OCCUR.

C: I agree.

A SET OF ZERO PROBABILITY MAY BE IGNORED when calculating probabilities. In particular, uniform distributions on $(0, 1)$, $[0, 1]$, $[0, 1)$ and $(0, 1]$ are all the same. Also, in 1a1 it does not matter whether we write $|x - y| < 1/3$ or $|x - y| \le 1/3$.

1d Uniform and non-uniform

It was implicitly assumed in Sect. 1a that all points of the square are equiprobable.[5] Is it a realistic assumption? I think, it is not. If a friend comes between 6:00 and 7:00, the central values (around 6:30) could be more probable than peripheral values (close to 6:00 or 7:00). Also, the two friends need not be independent. Say, a rain could influence both. The uniform distribution on the square is a simplification. During the course we'll deal with various distributions, generally non-uniform.

Discrete probability defined "uniform" by $p(\omega_1) = p(\omega_2)$ for all $\omega_1, \omega_2 \in \Omega$; all points are equiprobable. Equivalently, $p(\omega) = 1/n$ for all $\omega \in \Omega$; here $n = \#(\Omega)$.

Now, in the continuous case, the equality $p(\omega_1) = p(\omega_2)$ turns into $0 = 0$ and becomes void. It does not express uniformity. (After all, $p(\omega_1) = 2p(\omega_2)$ is also true.) Here is a correct formulation of uniformity:

$$P(A) = P(B) \qquad \text{whenever sets } A, B \subset \Omega \text{ are of equal area.}$$

Equivalent formulations will appear during the course. For now I emphasize the following. In the continuous case, PROBABILITIES OF SETS ARE RELEVANT; PROBABILITIES OF POINTS

[5]That is, of equal probability.

ARE IRRELEVANT. This is why the following section is difficult for a beginner. You'd like a function of a point; however, probability is rather a function of a set. Sorry...

1e Probability space, the definition

1e1 Definition. A *probability space* $(\Omega, \mathcal{F}, P)$ consists of:

$$\left.\begin{array}{l} \Omega \ne \emptyset \\ \mathcal{F} \subset 2^\Omega \\ \left.\left.\begin{array}{l} \emptyset, \Omega \in \mathcal{F} \\ A \in \mathcal{F} \implies \Omega \setminus A \in \mathcal{F} \\ A, B \in \mathcal{F} \implies A \cap B, A \cup B \in \mathcal{F} \end{array}\right\} \begin{array}{c}\mathcal{F} \text{ is a field} \\ \text{(algebra)}\end{array} \\ A_1, A_2, \dots \in \mathcal{F} \implies (A_1 \cup A_2 \cup \dots) \in \mathcal{F} \end{array}\right\} \begin{array}{c}\mathcal{F} \text{ is a } \sigma\text{-field} \\ (\sigma\text{-algebra})\end{array} \\ \left.\left.\begin{array}{l} P : \mathcal{F} \to [0,1] \\ P(\Omega) = 1 \\ P(A \uplus B) = P(A) + P(B) \end{array}\right\} \begin{array}{c}P \text{ is a finitely ad-} \\ \text{ditive probability}\end{array} \\ P(A_1 \uplus A_2 \uplus \dots) = P(A_1) + P(A_2) + \dots \end{array}\right\} \begin{array}{c}P \text{ is a proba-} \\ \text{bility measure}\end{array} \end{array}\right\} \begin{array}{c}(\Omega, \mathcal{F}, P) \text{ is a} \\ \text{probability space}\end{array}$$

Ω	—	sample space,
$\mathcal{F}$	—	σ-field of events,
P	—	probability measure.

1f Probability space, first examples

1f1 Example. A single toss of a fair coin is described by the following probability space:

$$\Omega = \{0,1\}\,; \qquad \mathcal{F} = 2^\Omega = \{\emptyset, \{0\}, \{1\}, \Omega\}\,, \qquad \begin{aligned} P : \emptyset &\mapsto 0\,, \\ \{0\} &\mapsto 1/2\,, \\ \{1\} &\mapsto 1/2\,, \\ \Omega &\mapsto 1\,. \end{aligned}$$

When Ω is finite or countable, all subsets of Ω may be treated as events; that is, we may take $\mathcal{F} = 2^\Omega$ and define the probability of any event $A \subset \Omega$ as the sum of probabilities of its points,

$$P(A) = \sum_{\omega \in A} p(\omega) \qquad \text{(1f2)}$$

(a finite sum or the sum of a series). That is, we start with a function $p : \Omega \to [0,1]$ satisfying $\sum_{\omega\in\Omega} p(\omega) = 1$ and construct $P : \mathcal{F} \to [0,1]$ by (1f2). It is always a probability measure, which is elementary in the finite case; in the countable case, some calculus is needed for the proof.

1f3 Example. An experiment with 10 equiprobable outcomes $0, 1, \dots, 9$ is repeated until 0 encounters.

Here Ω consists of all finite sequences of digits $1, \dots, 9$ (including the empty sequence), and

$$p(\omega) = 10^{-(1+\mathrm{length}(\omega))} .$$

$\omega \in \Omega$	$p(\omega)$	
empty	0.1	0.1
1	0.01	0.09
…	…	
9	0.01	
11	0.001	0.081
…	…	
99	0.001	
111	0.0001	
…	…	…

Note that $0.1+0.09+0.081+\dots = \frac{1}{10}\bigl(1+\frac{9}{10}+(\frac{9}{10})^2+\dots\bigr) = \frac{1}{10}\cdot\frac{1}{1-\frac{9}{10}} = 1$; the corresponding probability measure P is defined by (1f2).

Here is an example of an event: $A_1 =$ "the first outcome is 1"; A_1 is an infinite subset of Ω;

$\omega \in A_1$	$p(\omega)$	
1	0.01	0.01
11	0.001	0.009
…	…	
19	0.001	
111	0.0001	0.0081
…	…	
199	0.0001	
…	…	…

thus, $P(A_1) = 0.01 + 0.009 + 0.0081 + \dots = \frac{1}{100}\bigl(1+\frac{9}{10}+(\frac{9}{10})^2+\dots\bigr) = \frac{1}{10}$, as it should be.

Instead of sequences of digits we may use numbers:

$\omega \in \Omega$	$p(\omega)$
0	0.1
0.1	0.01
…	…
0.9	0.01
0.11	0.001
…	…
0.99	0.001
0.111	0.0001
…	…

You see a very special discrete probability distribution on $[0, 1)$.

1f4 Example. An experiment with 10 equiprobable outcomes $0, 1, \dots, 9$ is repeated endlessly.

Here Ω consists of all *infinite* sequences of digits $0, \dots, 9$. Each sequence is of zero probability; we cannot use $p(\omega)$ for constructing $P(A)$.

Instead of sequences we may use numbers;[6] say,

$$(3, 1, 4, 1, 5, 9, 2, 6, 5, \dots) \mapsto 0.314159265\dots$$

What is the corresponding probability distribution on $[0, 1)$?

The event $A_0 =$ "the first outcome is 0" is of probability 1/10, thus $P\big([0, 0.1)\big) = 0.1$. Similarly, $P\big([0.1, 0.2)\big) = 0.1, \dots, P\big([0.9, 1)\big) = 0.1$. It does not mean yet that the distribution is uniform. (After all, these 10 equalities hold also for the previous example.) However, the event $A_{00} =$ "the first two outcomes are 00" is of probability 1/100, thus $P\big([0, 0.01)\big) = 0.01$. Similarly, $P\big([0.01, 0.02)\big) = 0.01, \dots, P\big([0.99, 1)\big) = 0.01$. And so on; say,

$$P\big([0.314\,159, 0.314\,16)\big) = 0.000\,001\,.$$

It follows (by adding) that

$$P\big([a, b)\big) = b - a \tag{1f5}$$

whenever $0 \le a < b \le 1$ and a, b are of the form $k/2^n$. Such numbers are dense; it follows that[7] (1f5) holds for *all* a, b such that $0 \le a < b \le 1$.

So, the probability of any interval is equal to its length. Does it mean that the distribution is uniform? It is natural to *define* the uniform distribution by (1f5). However, what about events more complicated than intervals? Is every $P(A)$ uniquely determined by (1f5)?

Now the σ-field $\mathcal{F}$ becomes relevant. If you take $\mathcal{F} = 2^\Omega$, $\Omega = [0, 1)$, then of course $\mathcal{F}$ is a σ-field, however, $P(\cdot)$ is not uniquely determined by (1f5). The proof is far beyond our course, but the intuitive reason is simple: a very complicated set A cannot be reached from intervals by taking complements, unions and intersections.[8] For such A we have no idea what is $P(A)$.

Do we need bizarre sets? It depends on our goal. Such problems as 1a1 do not need complicated sets; Jordan measurable sets are enough. However, properties of infinite sequences lead to more complicated sets. For example, the Strong Law of Large Numbers[9] states (in particular) that the event

$$\Big\{ (x_1, x_2, \dots) \in \Omega : \lim_{n\to\infty} \frac{x_1 + \dots + x_n}{n} = 4.5 \Big\} \tag{1f6}$$

is of probability 1 (under conditions of 1f4). The corresponding subset of $[0, 1)$,

$$\Big\{ (0.x_1 x_2 \dots)_{10} : \lim_{n\to\infty} \frac{x_1 + \dots + x_n}{n} = 4.5 \Big\}\,, \tag{1f7}$$

[6] A tail of digits 9 (say, $0.24999\dots$) makes some troubles; however, such a case is of zero probability and may be ignored.

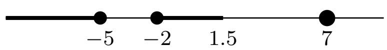
[7] Try to deduce it now, or wait for next sections.

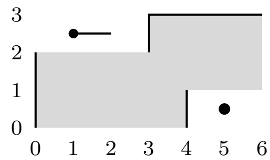
[8] The union (or intersection) of a *sequence* of sets is meant.

[9] We'll return to SLLN later.

is dense in $[0,1)$, and its complement is also dense. (Thus, it cannot be Jordan measurable.) A simpler example is the set of rational numbers (the corresponding event is periodicity of a sequence).

How to measure bizarre sets? We start with an algebra $\mathcal{A}$ of sets that can be measured easily. In $\mathbb{R}^1$ it may be the algebra generated by intervals. In $\mathbb{R}^2$ we may do the same (after introducing two-dimensional intervals). Examples: $A = (-\infty,-5] \cup [-2,1.5) \cup \{7\}$ (one-dim); $A = [0,4] \times (0,2) \cup [3,6) \times (1,3] \cup [1,2) \times \{2.5\} \cup \{5\} \times \{0.5\}$ (two-dim).

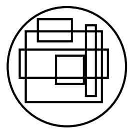




Alternatively, we may consider the algebra generated by polygons. Or, say, polygons and disks. Or, all domains with a piecewise smooth boundary.

Having such an algebra $\mathcal{A}$, we define

(1f8)
$$\mathcal{A}_\sigma = \{ A_1 \cup A_2 \cup \dots : A_1, A_2, \dots \in \mathcal{A} \} \, ,$$
$$\mathcal{A}_\delta = \{ A_1 \cap A_2 \cap \dots : A_1, A_2, \dots \in \mathcal{A} \} \, .$$



(In fact, all open sets belong to $\mathcal{A}_\sigma$, and all closed sets belong to $\mathcal{A}_\delta$.) Equivalently, an $\mathcal{A}_\sigma$-set is the limit of an increasing sequence, $A = \lim A_k$, of sets $A_k \in \mathcal{A}$ (think, why). For $\mathcal{A}_\delta$, the sequence is decreasing; anyway, we have $A = \lim A_k$ for a *monotone* sequence of sets $A_k \in \mathcal{A}$, and we define the measure of A to be the limit of measures of A_k. It is crucial that the limit does not depend on the choice of A_k (for given A) (which is proven by measure theory).

Thus, all open sets and all closed sets have their measures. Such sets can be somewhat bizarre. Say, there exists an open set of measure < 0.01 dense in the whole plane $\mathbb{R}^2$ (think, why). However, the set (1f7) is neither $\mathcal{A}_\sigma$ nor $\mathcal{A}_\delta$. Rather, it is of the type $\mathcal{A}_{\delta\sigma\delta}$,

$$A = \underbrace{\bigcap_\varepsilon \underbrace{\bigcup_m \underbrace{\bigcap_{n>m} \underbrace{\Big\{ (0.x_1 x_2 \dots)_{10} : \Big| \frac{x_1 + \dots + x_n}{n} - 4.5 \Big| < \varepsilon \Big\}}_{A_{\varepsilon,n} \in \mathcal{A}}}_{\in \mathcal{A}_\delta}}_{\in \mathcal{A}_{\delta\sigma}}}_{\in \mathcal{A}_{\delta\sigma\delta}}$$

and its probability is defined accordingly,

$$\mathbb{P}\big(A \big) = \lim_{\varepsilon \to 0} \lim_{m \to \infty} \lim_{n \to \infty} \mathbb{P}\big(A_{\varepsilon,m+1} \cap \dots \cap A_{\varepsilon,n} \big) \, .$$

Well, we should not deepen into measure theory. If you want to know more, try additional literature. Basically, the question 'how to measure sets' is answered as follows.

- Discrete case
  A set gets its measure from its points.
- Continuous case
  - A Jordan measurable set gets its measure from its lattice points via a limiting procedure. Or alternatively — from polygons.
  - A Borel measurable set gets its measure from simpler sets (say, an $\mathcal{A}_{\delta\sigma\delta}$-set from $\mathcal{A}_{\delta\sigma}$-sets, these from $\mathcal{A}_{\delta}$-sets, these from $\mathcal{A}$-sets).
  - A Lebesgue measurable set gets its measure from Borel measurable sets.
  - Otherwise, if a set is not Lebesgue measurable, its measure remains undefined.

WHAT IS A BOREL SET? Informally, a Borel measurable set (or 'a Borel set') is any set that can be constructed from a countable number of simple sets (say, intervals) using the set operations.

Formally, the Borel σ-field is defined as the smallest σ-field containing intervals. A Borel set is any set that belongs to the Borel σ-field.

The word 'intervals' may be interpreted here as 'open intervals', 'closed intervals', etc., without changing the Borel σ-field. Moreover, the word 'intervals' may be replaced with '$\mathcal{A}$-sets'; here $\mathcal{A}$ is any one of the algebras listed before; they all lead to the same Borel σ-field.

WHAT IS THE LEBESGUE MEASURE? Informally, it is the natural extension of length (area, volume,...) from Jordan measurable sets to all Borel sets.

Formally, the one-dimensional Lebesgue measure, denote it 'mes', is a set function,

$$\mathcal{B} \ni A \mapsto \operatorname{mes}(A) \in [0, +\infty]$$

(here $\mathcal{B}$ is the Borel σ-field) such that

$$\begin{aligned} A = (a,b) \quad &\Longrightarrow \quad \operatorname{mes}(A) = b - a \qquad \text{for } -\infty < a < b < +\infty\,, \\ \operatorname{mes}(A_1 \uplus A_2 \uplus \dots) &= \operatorname{mes}(A_1) + \operatorname{mes}(A_2) + \dots \qquad \text{for disjoint } A_1, A_2, \dots \in \mathcal{B}\,. \end{aligned}$$

(Existence and uniqueness of Lebesgue measure is proved by measure theory.) The same holds for dimension 2,

$$A = (a,b) \times (c,d) \quad \Longrightarrow \quad \operatorname{mes}_2(A) = (d-c)(b-a)\,,$$

and any dimension $d = 1, 2, 3, \dots$

WHAT IS A LEBESGUE MEASURABLE SET? It is any set $A \subset \mathbb{R}$ such that $B_1 \subset A \subset B_2$ and $\operatorname{mes}(B_2 \setminus B_1) = 0$ for some Borel sets B_1, B_2 ('sandwich'). Lebesgue measurable sets are a σ-field larger than the Borel σ-field. Naturally, one defines $\operatorname{mes}(A) = \operatorname{mes}(B_1)$ (or $\operatorname{mes}(B_2)$, which is the same), thus extending the Lebesgue measure. The same holds in any dimension $d = 1, 2, 3, \dots$

Note that all subsets of (say) a straight segment on $\mathbb{R}^2$ are Jordan measurable and (therefore) Lebesgue measurable (think, why), but not Borel measurable, in general. Note also that the notion of a Borel set depends only on the topology on $\mathbb{R}$ (or $\mathbb{R}^d$), while notions of Jordan measurable set and Lebesgue measurable set depend also on the measure.

Here is a summary of what will be used in our course.

1f9. For each dimension $d \in \{1, 2, 3, \dots\}$, the class $\mathcal{B}_d$ of all Borel subsets of $\mathbb{R}^d$ is a σ-field.

1f10. In $\mathbb{R}$, every interval[10] is a Borel set. In $\mathbb{R}^2$, every polygon is a Borel set. More generally, every domain with a piecewise smooth boundary is a Borel set. Still more generally, all open sets and all closed sets in $\mathbb{R}^d$ are Borel sets. Also, every finite or countable subset of $\mathbb{R}^d$ is a Borel set.

1f11. If two probability measures on $\mathcal{B}_1$ coincide on all intervals then they coincide on the whole $\mathcal{B}_1$. If two probability measures on $\mathcal{B}_2$ coincide on all polygons then they coincide on the whole $\mathcal{B}_2$.

1f12. Every Borel set $B \in \mathcal{B}_d$ has its d-dimensional *Lebesgue measure* $\operatorname{mes}_d(B) \in [0, +\infty]$. If $d = 1$ and B is an interval then $\operatorname{mes}(B)$ is equal to the length of the interval B.[11] If $d = 2$ and a Borel set B is Jordan measurable then $\operatorname{mes}_2(B)$ is equal to the area of B.

1f13. Let $\Omega \in \mathcal{B}_d$, $\operatorname{mes}_d(\Omega) \in (0, \infty)$. Define $\mathcal{F}$ as the class of all Borel subsets of Ω; define $P : \mathcal{F} \to [0, 1]$ by

$$P(A) = \frac{\operatorname{mes}_d(A)}{\operatorname{mes}_d(\Omega)} \quad \text{for all } A \in \mathcal{F}\,; \tag{1f14}$$

then $(\Omega, \mathcal{F}, P)$ is a probability space.

The latter gives us a lot of examples of continuous probability spaces.

[10] Be it (a, b), $[a, b)$, $(-\infty, a]$ etc.

[11] Of course, $\operatorname{mes}(B)$ is the same as $\operatorname{mes}_1(B)$.