

BASIC PROBABILITY THEORY 2024

JUHAN ARU

1

¹Version of 2024. All kinds of feedback, including smaller or bigger typos, is appreciated - juhan.aru@epfl.ch. This is a third version of the notes. In writing previous version of these notes I have consulted notes of I. Manolescu (Fribourg), Y. Velenik (Geneva), A. Eberle (Bonn) (all on their websites) and the book by R. Dalang & D. Conus published by EPFL press.

SECTION 0

Introduction

This course is about probability theory: the mathematical framework for formalising our questions about random phenomena, and their mathematical study.

When we want to describe a random phenomena in the real world, we build a mathematical model. This is itself an interesting process and a good model involves lots of well-chosen simplifications and righteous choices - e.g. to model a coin toss, we usually discard the possibility of it landing on the edge, or without further knowledge we consider the heads and tails equiprobable, although that may not be the case for example already because of different weight distributions. But this all is not the topic of this course.

In this course we will study the general mathematical framework and formulation of such models and then discuss the mathematical tools necessary and useful to study such models. Hopefully we also have some time to discuss some interesting models.

SECTION 1

Basic notions

In this chapter we discuss some basic but important notions of probability theory:

- Probability space
- Random variables
- Independence

1.1 Probability space

Our first aim is to motivate the notion of a probability space or a probabilistic model. To do this let us consider two examples:

- (1) A random number with values in $\{1, 2, \dots, 12\}$ e.g. something that comes from a lottery.
- (2) Describing the weather in Lausanne the day after.

In describing these two random phenomena we will still use everyday vocabulary / intuitions. Thereafter we will give the mathematical definitions that will fix the vocabulary for the rest of the course.

(1) Random number. To describe a random number mathematically, we basically need three inputs:

- The set of all possible outcomes: in this case $\Omega = \{1, 2, 3, \dots, 12\}$
- The collection of yes / no questions that we can answer about the actual outcome, i.e. this random number. For example:
 - Is this number equal to 3?
 - Is this number even?
 - Is this number smaller than 4?

To each of these questions we put in correspondence the subset of outcomes that corresponds to the answer yes: $\{3\}$, $\{2, 4, 6, 8, 10, 12\}$ or $\{1, 2, 3\}$ respectively. We call each such subset an event.

- Finally, to each event $E \subseteq \Omega$ we want to assign a numerical value $\mathbb{P}(E) \in [0, 1]$ that we call the probability. This should correspond to the fraction of times an event happens if the random number is given to us many times, e.g. if the lottery is played many times. ²

Here the set of possible outcomes was easy and directly given by the problem. Also it is natural to assume that each subset $E \subseteq \Omega$ is an event - or in other words that for each E we can ask the question: is the number in E ? This means that we can take the collection of events to correspond to all subsets of Ω .

Determining the probability really depends on what we want to model - e.g. if we are trying to model the lottery, we may assume that all numbers are equally likely and then we

²In fact, one uses probabilistic models also to model phenomena that only happens once. In that case probability measures somehow our degree of belief.

rediscover the model from high-school: we set $\mathbb{P}(E) = |E|/|\Omega|$. However, if we wanted to describe the sum of two dice, we would need to choose the numbers $\mathbb{P}(E)$ very differently! ³

Now, if we want our model to correspond to the intuitive notion of probability and to predict the fraction of repeated experiments, then these choices are not quite free - we need to add some constraints. E.g. we cannot put in an arbitrary function $\mathbb{P}$: indeed, if we have two events $E_1 \subseteq E_2$ then we should have $\mathbb{P}(E_1) \leq \mathbb{P}(E_2)$ as every time E_1 happens, also E_2 happens. We should also have $\mathbb{P}(\Omega) = 1$ as something always happens and $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$ if E and F are disjoint (why?). Of course not all these constraints are distinct - some might imply others and when giving the definition of a probability space below we will purify and choose only some conditions that will then mathematically imply all the others.

(2) Weather in Lausanne the day after. We would again want to make the three decisions, but here the task is already harder at the very first step. What should be the state space? A natural state space could probably be all possible microscopic states of the atmosphere up to 20km of height over Lausanne...but here we of course have many arbitrary choices - why 20 km, how wide should we look over Leman etc? And in any case, any natural state would be impossibly complicated!

Luckily, we do not actually need to worry about it - we only have to assign probabilities to all the events in the collection of events. And we have some freedom in choosing this collection events - it could be determined by our possibility to measure the states, e.g. we are able to measure the temperature up to some precision, or the density of CO_2 or water molecules to some precision and this determines some subsets of the state space. We could of course just choose our state space to be exactly so small that we can distinguish each state with our measurement devices, but allowing the state space to be larger and just restricting the set of yes and no questions / events has an advantage - if we get better at observing and measuring we can keep the same sample space and just change the collection of events. In particular, if we have fixed $\mathbb{P}$, we would need to only extend it and not rebuild the whole model.

However, as with the probability function, also for the collection of events there are some natural consistency conditions: we would assume that if one can observe if event E happened, we should be also able to measure if its complement E^c happened. Or if we are able to say if E happened or if F happened, we should be able to say if one of the two happened - i.e. $E \cup F$ should also be an event. And in fact it comes out that this is all we need!

Naturally, setting up probabilities for this model is also horribly complicated - there are no natural symmetry assumptions like the one we used for the uniform distribution. Also, even the best physicist in the world will not be able to describe the natural probability distribution of all microscopic states of the atmosphere, especially as it will heavily depend on what is happening just before! Thus, our only choice basically is to try to somehow use the combination of our knowledge about atmospheric processes together with our observations from history to set up some estimates for the model; and then naturally we will try to improve it with every next day. Luckily, this difficult task is not up to us but rather the office of meteo and the statisticians!

³See Exercise sheet 1.

Remark 1.1. Finally, before giving the mathematical definitions, let us stress again that all three components of the model - the sample space, the set of events and their probabilities - are inputs that we choose to build our model. When trying to model a real world phenomena we usually make simplifications for each of these choices. For example, for the coin toss we use only two outcomes: heads and tails, although theoretically edge is also possible. Also, we usually set probabilities to be a half, although that is not exactly true either.

1.2 Mathematical definition of a probability space

We are now ready to use our mathematical filter and give a mathematical definition of a probability space. In fact, we first use the mathematical purifier to come up with a definition in the restricted setting where Ω is a finite set, and then generalize it further.

Indeed, the discussions above lead us directly to:

Definition 1.2 (Elementary probability space, Kolmogorov 1933). *An elementary probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where*

- Ω is a finite set, called the state or sample space or the universe.
 - $\mathcal{F}$ is a set of subsets of Ω , satisfying:
 - $\emptyset \in \mathcal{F}$;
 - if $A \in \mathcal{F}$, then also $A^c \in \mathcal{F}$;
 - If $A_1, A_2 \in \mathcal{F}$, then also $A_1 \cup A_2 \in \mathcal{F}$.
- $\mathcal{F}$ is called the collection of events and any $A \in \mathcal{F}$ is called an event.
- And finally, we have a function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfying $\mathbb{P}(\Omega) = 1$ and additivity for disjoint sets: if $A_1, A_2 \in \mathcal{F}$ are pairwise disjoint, then

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

This function $\mathbb{P}$ is called the probability

Notice that some properties discussed above, like the fact that for events $E_1 \subseteq E_2$, we have $\mathbb{P}(E_1) \leq \mathbb{P}(E_2)$, follow directly from the definition.⁴

Now, most phenomena in the real world can be described by finite sets just because we are able to measure things only to a finite level of precision. However, like the notion of a continuous or differentiable function helps to simplify our mathematical descriptions of reality and thus improve our understanding, continuous probability spaces also make the mathematical descriptions neater, simpler and thereby also make it easier to understand and study the underlying random phenomena.

Some natural examples where infinite sample spaces come in: an uniform point on a line segment e.g. stemming from breaking a stick into several pieces; the position on the street where the first raindrop of the day falls; or the space of all infinite sequences of coin tosses. In all these cases the mathematically natural state space is even uncountable. Countably infinite state spaces can also come up: for example if we want to model the first moment that a repeated coin toss comes up heads, the value might be 1, 2, 3 or with very very small probability also 10^{10} , so a natural state space would contain all natural numbers.

So let us state the general definition:

Definition 1.3 (Probability space, Kolmogorov 1933). *A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where*

⁴See Exercise sheet 1.

- Ω is a set, called the state or sample space or the universe.
- $\mathcal{F}$ is a set of subsets of Ω , satisfying:
 - $\emptyset \in \mathcal{F}$;
 - if $A \in \mathcal{F}$, then also $A^c \in \mathcal{F}$;
 - If $A_1, A_2, \dots \in \mathcal{F}$, then also $\bigcup_{n \geq 1} A_n \in \mathcal{F}$. $\mathcal{F}$ is called the collection of events or a σ -algebra and any $A \in \mathcal{F}$ is called an event.
- And finally, we have a function $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfying $\mathbb{P}(\Omega) = 1$ and additivity for disjoint sets: if $A_1, A_2, \dots \in \mathcal{F}$ are pairwise disjoint,

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n).$$

This function $\mathbb{P}$ is called the probability

Notice the only differences are 1) we do not assume Ω to be finite 2) we assume that the set of events is stable under countable unions 3) we assume also the additivity of the probability under countable unions.

Exercise 1.1. Show that each elementary probability space is a probability space.

In fact probability spaces are an example of a general notion of measure spaces - probability spaces are just measure spaces with total mass equal to 1.

Definition 1.4 (Measure space, Borel 1898, Lebesgue 1901-1903). A measure space is a triple $(\Omega, \mathcal{F}, \mu)$, where

- Ω is a set, called the sample space or the universe.
- $\mathcal{F}$ is a set of subsets of Ω , satisfying:
 - $\emptyset \in \mathcal{F}$;
 - if $A \in \mathcal{F}$, then also $A^c \in \mathcal{F}$;
 - If $A_1, A_2, \dots \in \mathcal{F}$, then also $\bigcup_{n \geq 1} A_n \in \mathcal{F}$. $\mathcal{F}$ is called a σ -algebra and any $A \in \mathcal{F}$ is called a measurable set.
- And finally, we have a function $\mu : \mathcal{F} \rightarrow [0, \infty]$ satisfying $\mu(\emptyset) = 0$ and countable additivity for disjoint sets: if $A_1, A_2, \dots \in \mathcal{F}$ are pairwise disjoint,

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n).$$

This function μ is called a measure. If $\mu(\Omega) < \infty$, we call μ a finite measure.

Geometrically we interpret:

- Ω as our space of points
- $\mathcal{F}$ as the collection of subsets for which our notion of volume can be defined
- μ our notion of volume: it gives each measurable set its volume.

It is important to make this link to measure theory as many properties of probability spaces directly come from there. Yet it is also good to keep in mind that probability theory is not just measure theory - as M. Kac has put it well, 'Probability is measure theory with a soul' and we adhere to this philosophical remark.

Remark 1.5. You should compare the definition of a probability space / measure space with the definition of a topological space: there also we use a collection of subsets with certain

properties to attach structure to the set. A question you should ask is: why do we use exactly countable unions and intersections for the events, and not finite or arbitrary?