A Possible-Worlds Semantics for Kolmogorov's Axiomatization of Probability Theory

Francisco Corella

Revised August 4, 2020

Abstract

The nature of probability has been a philosophical open question for three centuries. In his 1933 treatise Kolmogorov proposed a formalization of probability theory that has very successfully supported the development of the theory and its use in a large variety of applications and a growing number of fields of science and engineering. But a frequentist interpretation of probability theory proposed in the same treatise is inconsistent in several respects with Kolmogorov's own formalization. Later authors have moved away from Kolmogorov's interpretation but have not been able to provide a satisfactory explanation of the nature of the elements of the first component Ω of what is called today a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We propose to view Ω as a set of possible worlds, borrowing a philosophical concept that goes back to Leibniz, but staying away from modal and probabilistics logics and keeping probabilistic reasoning inside ordinary mathematics, where a probability space is a structure defined by Kolmogorov's axioms, within the context of first-order logic and Zermelo-Frankel set theory. We also propose an intensional definition of the notion of event as a condition on the values of one or more random variables, the less intuitive traditional definition as a subset of Ω being then more easily understood as the extensional counterpart of the intensional definition.

Contents

1 Introduction

The formalization and axiomatization of probability theory in Kolmogorov's classic treatise [1], dated 1933, is universally accepted by practitioners as the foundation of modern probability theory. But while most, if not all, philosophers agree with practitioners on the adequacy of Kolmogorov's axioms for reasoning mathematically about probabilities, there is no agreement as to what probabilities mean. The meaning of probabilities is a philosophical open question still being debated today, as can be seen in the entry on *Interpretation of Probability* in the Standford Encyclopedia of Philosophy [2].

Yet Kolmogorov did not propose his axioms as an unmotivated, purely mathematical theory. An interpretation is included along with the axioms in [1]. Given the broad acceptance of the axioms, that interpretation could have settled the question of the meaning of probability once and for all. But that did not happen. Kolmogorov's interpretation of his own axioms is not even mentioned in [2]. What went wrong?

As we shall see below, Kolmogorov's interpretation is couched in terms of frequencies of outcomes of a repeated experiment. It can therefore be objected to on general grounds as a *frequentist interpretation* [2, §3.4]. But it can also be objected to on specific grounds as being inconsistent in several respects with the formalization of probability theory that it is trying to interpret. This paper analyzes these inconsistencies and proposes an alternative interpretation based on possible-worlds semantics. It also proposes an intensional definition of the notion of event as a condition on the values of one or more random variables, the less intuitive traditional definition as a subset of Ω being then more easily understood as the extensional counterpart of the intensional definition.

2 Inconsistencies of Kolmogorov's interpretation with his formalization

2.1 Kolmogorov's formalization

Kolmogorov's formalization defines a *field of probability* consisting of a set E, a set $\mathfrak{F}$ of subsets of E called *random events* or simply *events*, and a function P that assigns to each event $A \in \mathfrak{F}$ a non-negative real number $\mathrm{P}(A)$ called the *probability* of A. He refers to the elements of E as *elementary events* and uses the variable ξ to refer to them. Today E, $\mathfrak{F}$ and P are often called Ω, $\mathcal{F}$ and $\mathbf{P}$, the structure $(E, \mathfrak{F}, \mathbf{P})$ is called a *probability triple* or *probability space*, and the variable ω is used to refer to elements Ω.

The formalization is introduced in three stages. In Chapter 1, $\mathfrak{F}$ is a *field of sets*, or *algebra*, closed under finite unions, intersections and differences, and containing E. The probability function P satisfies the axioms $P(E) = 1$ and $P(A + B) = P(A) + P(B)$, where $+$ denotes union, for disjoint A and B. Chapter II Section 1 adds an Axiom of Continuity from which countable additivity follows. And in Chapter II Section 2 the event space is required to be a *Borel field*, today called a *σ-algebra*, closed under countable unions and intersections.

Random variables are not introduced until Chapter III, and only real-valued variables are defined. The general concept of random variable, not necessarily numeric, had to wait until 1948, when it was introduced by Fréchet [3].

2.2 Kolmogorov's interpretation

Kolmogorov provides his interpretation in Section 2 of Chapter I, entitled *The Relation to Experimental Data*, saying in a footnote that he has used in large measure the work of R. v. Mises at [4, pp. 21-27].

To "*apply the theory of probability to the actual world of experiments*" he considers a "*complex of conditions* $\mathfrak{S}$", i.e. an experiment, "*which allows of any number of repetitions*", and gives as an example the tossing of a coin two times, each tossing resulting in either a head (H) or a tail (T) coming up. In the example the outcome of the experiment has four possible "*variants*", HH, HT, TH and TT, which play the roles of the *elementary events*, i.e. of the elements of E, in the formalization. An event is then a set of elementary events. For example the event that the same side of the coin comes up in the two tossings, either head or tail, is the set $A = \{\mathrm{HH}, \mathrm{TT}\}$.

He then considers multiple repetitions of the experiment, and refers to the event A as having *multiple occurrences*, one for each repetition. This allows him to interpret the probability $\mathrm{P}(A)$ of event A as follows (quoting literally from [1, page 4]):

> *Under certain conditions, which we shall not discuss here, we may assume that to an event A which may or may not occur under conditions $\mathfrak{S}$, is assigned a real number $\mathrm{P}(A)$ which has the following characteristics:*
>
> *(a) One can be practically certain that if the complex of conditions $\mathfrak{S}$ is repeated a large number of times, n, then if m be the number of occurrences of event A, the ratio m/n will differ very slightly from $\mathrm{P}(A)$.*
>
> *(b) If $\mathrm{P}(A)$ is very small, one can be practically certain that, when conditions $\mathfrak{S}$ are realized only once, the event A would not occur at all.*

2.3 A first inconsistency with the formalization

This interpretation is inconsistent with how a sequence of experiments is treated in Kolmogorov's own formalization of probability theory as it is used in the rest of [1] and as it has subsequently been used as the foundation of modern probability theory. There is no such thing in modern probability theory as an event that has "multiple occurrences" in a sequence of "repetitions" of the same experiment. Multiple tosses of a coin are modeled as a sequence of distinct experiments. Kolmogorov himself considers sequences of experiments in Sections 5 and 6 of Chapter I of [1].

In the example used in his interpretation, Kolmogorov defines the event A in a context where a single experiment takes place, consisting of a pair of tosses of a coin. In that context, he says that E consists of four elementary events HH, HT, TH and TT and he defines A as $\{\mathrm{HH}, \mathrm{TT}\}$.

Then he introduces a sequence of "repetitions of the experiment". In modern probability theory as formalized by Kolmogorov, the proper treatment of such "repetitions" is as a sequence $(Y_i)_{1\leq i\leq n}$ of random variables, each ranging over a four-element set $\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$. (While Kolmogorov does not have the concept of general, not necessarily real-valued random variables, an equivalent sequence of real-valued random variables could be defined by encoding HH, HT, TH and TT as any four real numbers.) That four-element set is the common range of the random variables, not the set E of the formalization, which must have many more than 4 elements in the multi-experiment context. To say that the event A of the single-experiment context occurs in one of the experiments of the multi-experiment context is an informal, but formally incorrect, way of saying that the corresponding random variable takes either the value HH or the value TT. The event that this happens in, say, the second experiment, is properly formalized as the preimage

$$Y_2^{-1}[\{\mathrm{HH}, \mathrm{TT}\}] = \{\xi \in E \mid Y_2(\xi) \in \{\mathrm{HH}, \mathrm{TT}\}\},$$

which may be more concisely referred to as "the event $Y_2 \in \{\mathrm{HH}, \mathrm{TT}\}$", omitting the argument ξ of the random variable Y_2 (which is a function with domain E). And the event that it happens in the fifth experiment as the preimage

$$Y_5^{-1}[\{\mathrm{HH}, \mathrm{TT}\}] = \{\xi \in E \mid Y_5(\xi) \in \{\mathrm{HH}, \mathrm{TT}\}\},$$

or "the event $Y_5 \in \{\mathrm{HH}, \mathrm{TT}\}$". These are different events in the multi-experiment context, not two occurrences in the multi-experiment context of the event A defined in the single-experiment context.

It is worth pointing out that the single experiment context can be modeled in the same manner as the multi-experiment context, using a single random variable Y that ranges over $\{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$; and it would be best to do so. The event A would then be defined as $Y^{-1}[\{\mathrm{HH}, \mathrm{TT}\}] = \{\xi \in E \mid Y(\xi) \in \{\mathrm{HH}, \mathrm{TT}\}$, or "the event $Y \in \{\mathrm{HH}, \mathrm{TT}\}$". There would be no need to specify the elements of the set E, and it would not matter if E (or Ω with current notations) had four elements or some other finite, or countably or uncountably infinite number of elements; it would just have to be a Borel field in the terminology of [1], or a σ-algebra in current terminology.

A random variable maps an underlying probability space to a target probability space called the probability distribution of the variable. In applications of probability theory there should never be a need to specify the underlying probability space of a random variable. The probability distribution should be sufficient for all probabilistic reasoning involving the variable. In an application involving multiple random variables, all those variables are sourced from the same underlying probability space, and their joint probability distribution should be sufficient for all probabilistic reasoning in the context of that application without having to specify the common underlying probability space.

2.4 The law of large numbers

What is Kolmogorov referring to when he mentions "*certain conditions, which we shall not discuss here*"? He must be referring to the hypotheses of the law of large numbers. A modern statement of the strong law of large numbers, as found, e.g., in [5, Theorem 6.1, p. 70] or [6, Theorem 5.4.4, p. 62] asserts that

$$\mathbf{P}\left(\lim_{n\to\infty}\frac{1}{n}(X_1+X_2+\ldots+X_n)=m\right)=1 \tag{1}$$

if $X_1, X_2, \ldots$ are independent identically distributed random variables with mean m. An equivalent statement can be found in Kolmogorov's treatise, with a footnote saying that a proof of the statement has not yet been published [1, footnote 9, page 67].

Given the random variables $(Y_i)_{i=1,2,\ldots}$ of the above multi-experiment context, we can define random variables $(X_i)_{i=1,2,\ldots}$ such that each X_i is the characteristic function of the event that $Y_i \in \{\mathrm{HH}, \mathrm{TT}\}$, i.e. the function that maps $\xi \in E$ to 1 if $Y_i(\xi) \in \{\mathrm{HH}, \mathrm{TT}\}$ or to 0 otherwise. The expected value, or mean, of X_i is then the probability of the event $Y_i \in \{\mathrm{HH}, \mathrm{TT}\}$. If we assume that the X_i are independent and identically distributed random variables whose common mean is equal to the probability of the event A of the single-experiment context, and if we take what Kolmogorov calls "*the number of occurrences of event A*" (and denotes with the variable m, not to be confused with the variable m that denotes the common mean of the X_i in the above statement of the law of large numbers), to mean the number of variables X_i, $1 \le i \le n$ that take the value HH or the value TT in a run of the multiple experiments, then the above quoted paragraph (a) is an informal statement of the conclusion of the strong law of large numbers when applied to the sequence $(X_i)_{1\le i\le n}$.

Thus the frequentist interpretation of the probability function in paragraph (a) is in fact a feature of the probability function that follows from the axioms. This is puzzling. Since the axioms are not concerned with frequencies, it should be possible to provide a non-frequentist interpretation, and indeed we provide one below. The adequacy of the axioms would then be highlighted by the fact that they can be used to prove the law of large numbers even though they can be interpreted without any reference to frequencies. Why did Kolmogorov need to use frequencies in his interpretation?

2.5 A second inconsistency

The answer can be found in the paragraph that follows the above quotation, entitled "*The Empirical Deduction of the Axioms*". There, Kolmogorov uses the frequentist interpretation of the probability function to justify Axiom V of the formalization of Section 1, the axiom of finite additivity. The finite additivity of probabilities of incompatible events is justified by the finite additivity of their frequencies.

But why is this justification necessary? Events are formalized as sets of elements of E, called elementary events, and incompatible events as disjoint sets of elementary events. Doesn't it follow that the probability of the union of two incompatible events is the sum of their probabilities?

That follows if the elements of E are assigned probabilities and the probability of an event is the sum of the probabilities of the elements of E that it contains. The axioms do not assign probabilities to the elements of E, but an interpretation can do so, and Kolmogorov defines such interpretations when he constructs fields of probabilities with finite E in footnote 7 page 9, and with countable E in Chapter II, Section 3, paragraph I.

But it is not possible to assign probabilities to the elements of E that add up to event probabilities in a context that involves one or more continuous random variables. This is because a continuous random variable takes an uncountable number of distinct values, which requires an uncountable number of distinct event probabilities. If an event probability is to be the sum of probabilities of elements of E, that requires an uncountable number of elements of E with non-zero probabilities, which is impossible because the sum of their probabilities would have to be $\mathrm{P}(E) = 1$.

This means that interpreting the elements of E as *elementary events* is inconsistent with the formalization, because such terminology strongly suggests that they are of a nature similar to that of the *random events* that are the elements of $\mathcal{F}$, and that both the elements of E and the elements of $\mathcal{F}$ have probabilities. The similar nature of the elements of E and the elements of $\mathcal{F}$ in Kolmogorov's interpretation is emphasized by the fact that Kolmogorov refers to the elements of E as random events rather than elementary events in the definition of "the probability function of u" in Chapter III, Section 1; but perhaps this is just an inadvertent error.

3 The nature of the elements of E (or Ω)

Later authors have avoided this inconsistency by not referring to the $\xi \in E$, or the $\omega \in \Omega$ in modern terminology, as events of any kind. Rosenthal [6] avoids using any generic name for them; others refer to them as "*outcomes*".

For example, Billingsley [5, §2, p. 15] interprets Ω as consisting of "*all the possible results or outcomes of an experiment or observation*", and provides many examples:[1]

> *For observing the number of heads in n tosses of a coin the space Ω is $[0, 1, \ldots, n]$; for describing the complete history of the n tosses Ω is the space of all 2^n n-long sequences of H's and T's; for an infinite sequence of tosses Ω can be taken Ω can be taken as the unit interval as in the preceding section; for the number of α-particles emitted by a substance over a unit interval of time of the number of telephone calls arriving at an exchange Ω is $[0, 1, 2, \ldots]$; for the position of a particle Ω is three-dimensional Euclidean space; for describing the motion of the particle Ω is an appropriate space of functions; and so on.*

There is something to be noticed in these examples: in each of them, Ω is the range of values of a random variable. This clashes with the fact that a random variable is a function with domain Ω, for two reasons: first, if we interpret Ω both as the domain of a random variable and its range, we must be interpreting the random variable as an identity function; second, all the random variables in a particular context cannot be required to have the same range.

It also goes counter the fundamental intuition of modern probability theory, which all practitioners consciously or subconsciously share, that the role of Ω is only to model the fact

[1] Billingsley uses square brackets instead of curly brackets to denote sets.

that a random variable can take any of a number of possible values, and that there is never an operational need to specify Ω or discuss the nature of its elements. In a probabilistic context with multiple variables, each $\omega \in \Omega$ models a slate of possible outcomes of all the variables. In a context with a temporal dimension and a notion of a present time, an $\omega \in \Omega$ models both the uncertainty about the values that random variables may take in the future, and the fact that they could have taken alternative values in the past.

So, if they are not events of any kind, nor outcomes of experiments or observations, what is the nature of the elements of Ω? The philosophical concept of *possible worlds* provides an answer.

4 An alternative interpretation

4.1 The concept of possible worlds

The concept of possible worlds goes back to Leibniz, who viewed them as alternative worlds that God could have created, the actual world having been chosen by God as the best of all possible worlds. This notion was ridiculed by Voltaire and other philosophers after a devastating earthquake followed by a tsunami hit Lisbon in 1755 causing tens of thousands of deaths and tremendous suffering [7, 8, 9]. A good explanation of the modern version of the concept can be found in the first two paragraphs of the entry on *Possible Worlds* in the Stanford Encyclopedia of Philosophy [10].

The elements of Ω can be viewed as *possible worlds* because Ω is the domain of the random variables and each random variable can take a range of *possible values*. The fact that the same Ω is the domain of all the random variables captures the intuition that only some combinations of values of the random variables are possible. Each possible combination corresponds to an element $\omega \in \Omega$, and is the result of applying each random variable, viewed as a function, to ω. If we view ω as a possible world, we can say that the possible combination of values corresponding to ω consists of the values that the random variables take in that possible world.

This interpretation views probability theory as a way of assigning *probabilities* to *sets of possibilities*. The elements of Ω are possibilities concerning the values of random variables, organized as possible worlds. The elements of $\mathcal{F}$ are sets of possible worlds that are assigned probabilities.

4.2 An intensional concept of event

When we view the elements of Ω as possible worlds, a natural definition of the concept of *event* is as a condition on the values that one or more random variables take when applied, as functions, to an element $\omega \in \Omega$. The event is said to happen in the possible world ω if the condition is satisfied for that ω. The set S of possible worlds $\omega \in \Omega$ that satisfy the condition must be an element of the σ-algebra $\mathcal{F}$, and the probability of the event is then $\mathbf{P}(S)$.

When a formula is used to specify the condition, the argument ω of the random variables is omitted so that the formula reads as if the variables were not random. For example, in the conclusion (1) of the strong law of large numbers shown in Section 2.4, the formula

$$\lim_{n\to\infty} \frac{1}{n}(X_1 + X_2 + \ldots + X_n) = m \tag{2}$$

denotes an event that happens in a possible world ω if

$$\lim_{n\to\infty} \frac{1}{n}(X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega)) = m.$$

If we identify the event with the set

$$S = \{\omega \in \Omega \mid \lim_{n\to\infty} \frac{1}{n}(X_1(\omega) + X_2(\omega) + \ldots + X_n(\omega)) = m\}, \tag{3}$$

of possible worlds in which it happens (which is an element of $\mathcal{F}$ as shown, for example, in the proof of [5, Theorem 6.1, p. 70]), we get the traditional definition of the event as an element of the σ-algebra $\mathcal{F}$. Condition 2 is an intensional definition of the event, while equation 3 is a corresponding extensional definition.

The dual intensional and extensional definitions of the concept of event bring probability theory into agreement with the informal notion of probability and suggest a better way of introducing the formal concepts of the theory.

Today events are introduced before random variables. An event is defined as a set of elements of Ω, in disagreement with the ordinary notion of event. How is the event that a coin toss comes up heads any kind of set? Saying that the elements of Ω are outcomes, such as the possible outcomes of the toss of a coin or a die, adds to the confusion when Ω is then said to be the common domain of all the random variables, rather the range of any one random variable.

The better way is to define random variables before events. A random variable is a variable that can take a range of possible values. A random variable X that models the tossing of a coin can take the value H or T. These two possibilities are mathematically modeled by saying that X is a function whose domain is a set Ω of possible worlds, which takes the value H in some possible worlds $\omega \in \Omega$, and T in others. The fact that the toss comes up heads is an event, intensionally defined by the condition $X(\omega) = \mathrm{H}$, abbreviated $X = \mathrm{H}$ when it is understood that X is a random variable rather than an ordinary variable. The event is said to happen in those possible worlds ω where the condition is true, and the set E of those possible worlds is the extensionally defined event. Having said all this, the fact that E is a set is no longer unintuitive.

When an element of $\mathcal{F}$ is defined without reference to any conditions on random variables, it is unintuitive to refer to it as an event. It would be better to refer to it instead as an element of $\mathcal{F}$, or as an "*$\mathcal{F}$-set*", a terminology that Billingsley uses, e.g., when he restates the conclusion of the strong law of large numbers [5, Theorem 6.1] as saying that $n^{-1}\sum_{i=1}^{n}(X_i - m) \to 0$ "*except on an $\mathcal{F}$-set of probability 0*".

5 Logic, mathematics, and probabilities

The modern concept of possible worlds is primarily used in modal logic [11], where it serves as the semantics for an extension of propositional logic with two operators, "it is necessary that" ($\Box$) and its dual "it is possible that" ($\Diamond$). If p is a propositional variable, $\Box p$ is interpreted as "p is true in all possible worlds", and $\Diamond p$ as "p is true in some possible world".

In first approximation this agrees with the interpretation of elements of Ω as possible worlds: an event is possible if it has a probability greater than 0, which implies that it happens in at least one possible world, and necessary if it has probability 1, which is the case if it happens in all possible worlds. But there are differences. First, in probability theory there are non-empty $\mathcal{F}$-sets with probability 0, and $\mathcal{F}$-sets other than Ω with probability 1. Second, the **S1**–**S5** systems of modal logic [12, §1.1] have a syntactic Uniform Substitution rule that allows any formula to be substituted for a propositional variable; this results in formulas with nested modal operators that require interpretations where there is a relation of "accessibility" between worlds [12, §3.2]. Third, and most significantly, probability theory is a part of mathematics, rather than an ad-hoc formal logic with syntax and semantics designed to express the concepts of necessity and possibility.

Yet another entry in the Stanford Encyclopedia of Philosophy, the entry on *Logic and Probability* [13], discusses a large collection of logics specifically designed to support probabilistic reasoning. But probability theory as formalized by Kolmogorov is being very successfully used in a large variety of applications and a growing number of fields of science and engineering without recourse to probabilistic logics. There seems to be no need of ad-hoc logical formalisms to support probabilistic reasoning.

Probability theory as formalized by Kolmogorov is an ordinary part of mathematics. A probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is a structure in the sense of Bourbaki [14, 15], just like a group, a vector space, or a topological space. Reasoning about probabilities in mathematics, like most other mathematical reasoning, can be formalized in the context of first-order logic augmented with a system of axioms of set theory such as Zermelo-Frankel. The possible-world semantics proposed here is just a way of thinking about the nature of the elements of Ω, which hopefully makes probability theory more intuitive and easier to understand and teach.

6 Conclusion

We have analyzed Kolmogorov's frequentist interpretation of probability theory and shown that it is inconsistent in several respects with his own formalization. Later authors have moved away from Kolmogorov's interpretation but have not been able to provide a satisfactory explanation of the nature of the elements of the first component Ω of what is called today a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We have proposed to view Ω as a set of possible worlds, borrowing a philosophical concept that goes back to Leibniz, but staying away from modal and probabilistic logics and keeping probabilistic reasoning inside ordinary mathematics, where a probability space is a structure defined by Kolmogorov's axioms, within the context of first-order logic and Zermelo-Frankel set theory. We have also proposed an intensional definition of the notion of event as a condition on the values of one or more random variables, the less intuitive traditional definition as a subset of Ω being then more easily understood as the extensional counterpart of the intensional definition.

Acknowledgements

The author is grateful to Michel and Catherine Baudin for providing comments and catching an error.

References

[1] A. N. Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Pub. Co. Second English edition, 1956. https://books.google.com/books?id=EmetAAAAMAAJ.

[2] Alan Hájek. Interpretations of probability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2019 edition. https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/.

[3] Maurice René Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, 10(4):215–310, 1948.

[4] Richard von Mises. *Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und theoretischen Physik*. 1931. Published in 1945 by Rosenberg, NY under authority of the United States Alien Property Custodian. Available from AbeBooks.com.

[5] P. Billingsley and Karreman Mathematics Research Collection. *Probability and Measure*. Wiley, 1979.

[6] J.S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, 2006.

[7] Wikipedia. Poème sur le désastre de Lisbonne. https://en.wikipedia.org/wiki/Po%C3%A8me_sur_le_d%C3%A9sastre_de_Lisbonne. Accessed July 18, 2020.

[8] Voltaire. Poème sur le désastre de Lisbonne, 1956. Gallica, Les Essentiels, Littérature. https://gallica.bnf.fr/essentiels/anthologie/poeme-desastre-lisbonne. Retrieved July 19, 2020.

[9] F.M. Arouet (Voltaire). *Candide*. Dover thrift editions. Dover Publications, 2012.

[10] Christopher Menzel. Possible worlds. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition. https://plato.stanford.edu/archives/win2017/entries/possible-worlds/.

[11] James Garson. Modal logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2018 edition. https://plato.stanford.edu/archives/fall2018/entries/logic-modal/.

[12] Roberta Ballarin. Modern origins of modal logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition. https://plato.stanford.edu/archives/sum2017/entries/logic-modal-origins/.

[13] Lorenz Demey, Barteld Kooi, and Joshua Sack. Logic and probability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition. https://plato.stanford.edu/archives/sum2019/entries/logic-probability/.

[14] Leo Corry. Nicolas bourbaki and the concept of mathematical structure. *Synthese*, 92(3):315–348, 1992.

[15] Nicolas Bourbaki. *Théory des Ensembles*. Diffusion CCLS, 1970.