Chapter 2

Probability Distributions

The main purpose of this chapter is to introduce Kolmogorov's probability axioms. These are the first three core rules of Bayesianism. They represent constraints that an agent's unconditional credence distribution at a given time must satisfy in order to be rational.

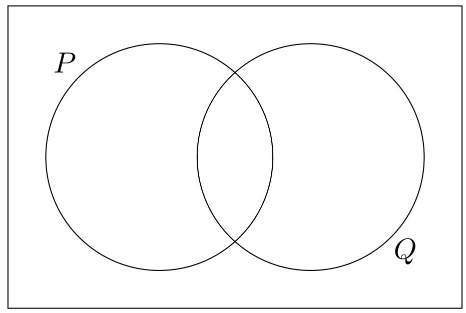
The chapter begins with a quick overview of propositional and predicate logic. The goal is to remind readers of logical notation and terminology we will need later; if this material is new to you, you can learn it from any introductory logic text. Next I introduce the notion of a numerical distribution over a propositional language, the tool Bayesians use to represent an agent's degrees of belief. Then I present the probability axioms, which are mathematical constraints on such distributions.

Once the probability axioms are on the table, I point out some of their more intuitive consequences. The probability calculus is then used to analyze the Lottery Paradox scenario from Chapter 1, and Tversky and Kahneman's Conjunction Fallacy example.

Kolmogorov's axioms are the canonical way of defining a probability distribution, and are useful for doing probability proofs. Yet there are other, equivalent mathematical structures that Bayesians often use to illustrate points and solve problems. After presenting the axioms, this chapter describes how to work with probability distributions in two alternate forms: Venn diagrams and stochastic truth-tables.

I end the chapter by explaining what I think are the most distinctive elements of probabilism, and how probability distributions go beyond what one obtains from a comparative confidence ordering.

Figure 2.1: The space of possible worlds



2.1 Propositions and propositional logic

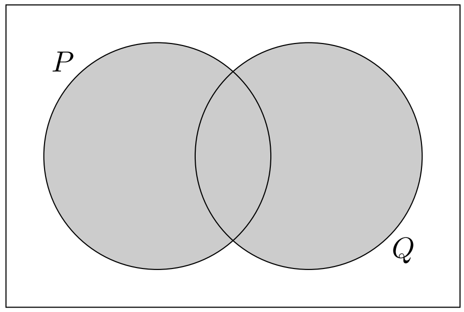
While other approaches are sometimes used, we will assume that degrees of belief are assigned to propositions.[1] In any particular application we will be interested in the degrees of belief an agent assigns to the propositions in some language $\mathcal{L}$. $\mathcal{L}$ will contain a finite number of **atomic propositions**, which we will usually represent with capital letters (P, Q, R, etc.).

The rest of the propositions in $\mathcal{L}$ are constructed in standard fashion from atomic propositions using five **propositional connectives**: $\sim$, &, $\vee$, $\supset$, and $\equiv$. $\sim P$ is true just in case P is false. $P \,\&\, Q$ is true just in case both P and Q are. "$\vee$" represents inclusive "or"; $P \vee Q$ is false just in case P and Q are both false. "$\supset$" represents the material conditional; $P \supset Q$ is false just in case P is true and Q is false. $P \equiv Q$ is true just in case P and Q are both true or P and Q are both false.

Philosophers sometimes think about propositional connectives using sets of **possible worlds**. Possible worlds are somewhat like the alternate universes to which characters travel in science-fiction stories—events occur in a possible world, but they may be different events than occur in the **actual world** (the possible world in which *we* live). Possible worlds are maximally specified, such that for any event and any possible world that event either does or does not occur in that world. And the possible worlds are plentiful enough such that for any combination of events that *could* happen, there is a possible world in which that combination of events *does* happen.

We can associate with each proposition the set of possible worlds in which that proposition is true. Imagine that in the **Venn diagram** of Figure 2.1, the possible worlds are represented as points inside the rectangle. Proposition P might be true in some of those worlds, false in others. We

Figure 2.2: The set of worlds associated with $P \vee Q$



can draw a circle around all the worlds in which P is true, label it P, and then associate proposition P with the set of all possible worlds in that circle (and similarly for proposition Q).

The propositional connectives can also be thought of in terms of possible worlds. $\sim P$ is associated with the set of all worlds lying outside the P-circle. $P \,\&\, Q$ is associated with the set of worlds in the overlap of the P-circle and the Q-circle. $P \vee Q$ is associated with the set of worlds lying in either the P-circle or the Q-circle. (The set of worlds associated with $P \vee Q$ has been shaded in Figure 2.2 for illustration.) $P \supset Q$ is associated with the set containing all the worlds except those that lie both inside the P-circle and outside the Q-circle. $P \equiv Q$ is associated with the set of worlds that are either in both the P-circle and the Q-circle or in neither one.[2]

Warning: I keep saying that a proposition can be "associated" with the set of possible worlds in which that proposition is true. It's tempting to think that the proposition just *is* that set of possible worlds, but we will avoid that temptation. Here's why: The way we've set things up, any two logically equivalent propositions (such as P and $\sim P \supset P$) are associated with the same set of possible worlds. So if propositions just *were* their associated sets of possible worlds, P and $\sim P \supset P$ would be the same proposition. Since we're taking credences to be assigned to propositions, that would mean that *of necessity* every agent assigns P and $\sim P \supset P$ the same credence. Eventually we're going to suggest that if an agent assigns P and $\sim P \supset P$ different credences she's making a rational mistake. But we want our formalism to suggest it's a *rational requirement* that agents assign the same credence to logical equivalents, not a *necessary truth.* It's useful to think about propositions in terms of their associated sets of possible worlds, so we will continue to do so. But to keep logically equivalent propositions separate entities we will not say that a proposition just is a set of possible worlds.

Before we discuss logical relations among propositions, a word about notation. I said we will use capital letters as atomic propositions. We will also use capital letters as metavariables ranging over propositions. I might say, "If P entails Q, then…". Clearly the atomic proposition P doesn't entail the atomic proposition Q. So what I'm trying to say in such a sentence is "Suppose we have one proposition (which we'll call 'P' for the time being) that entails another proposition (which we'll call 'Q'). Then…". At first it may be confusing sorting atomic proposition letters from metavariables, but context will hopefully make my usage clear. (Look especially for such phrases as: "For any propositions P and Q…".)[3]

2.1.1 Relations among propositions

Propositions P and Q are **equivalent** just in case they are associated with the same set of possible worlds—in each possible world, P is true just in case Q is. In that case I will write "$P \mathrel{=}\!\mathrel{|}\!\mathrel{\models} Q$". P **entails** Q ("$P \vDash Q$") just in case there is no possible world in which P is true but Q is not. On a Venn diagram, P entails Q when the P-circle is entirely contained within the Q-circle. (Keep in mind that one way for the P-circle to be entirely contained in the Q-circle is for them to be the same circle! When P is equivalent to Q, P entails Q and Q entails P.) P **refutes** Q just in case $P \vDash \sim Q$. When P refutes Q, every world that makes P true makes Q false.[4]

For example, suppose I roll a six-sided die. The proposition that the die came up six entails the proposition that it came up even. The proposition that the die came up six refutes the proposition that it came up odd. The proposition that the die came up even is equivalent to the proposition that it did not come up odd—and each of those propositions entails the other.

P is a **tautology** just in case it is true in every possible world. In that case we write "$\vDash P$". I will sometimes use the symbol "T" to stand for a tautology. A **contradiction** is false in every possible world. I will sometimes use "F" to stand for a contradiction. A **contingent** proposition is neither a contradiction nor a tautology.

Finally, we have properties of proposition *sets* of arbitrary size. The propositions in a set are **consistent** if there is at least one possible world in which all those propositions are true. The propositions in a set are **inconsistent** if no world makes them *all* true.

The propositions in a set are **mutually exclusive** if no possible world makes *more than one* of them true. Put another way, any two propositions in a mutually exclusive set are inconsistent with each other. (For any propositions P and Q in the set, $P \vDash {\sim}Q$.) The propositions in a set are jointly **exhaustive** if each possible world makes at least one of the propositions in the set true. In other words, the disjunction of all the propositions in the set is a tautology.

We will often work with proposition sets whose members are both mutually exclusive and jointly exhaustive. A mutually exclusive, jointly exhaustive set of propositions is called a **partition**. Intuitively, a partition is a way of dividing up the available possibilities. For example, in our die-rolling example the proposition that the die came up odd and the proposition that the die came up even form a partition. When you have a partition, each possible world makes *exactly* one of the propositions in the partition true. On a Venn diagram, the regions representing the propositions combine to fill the entire rectangle without overlapping at any point.

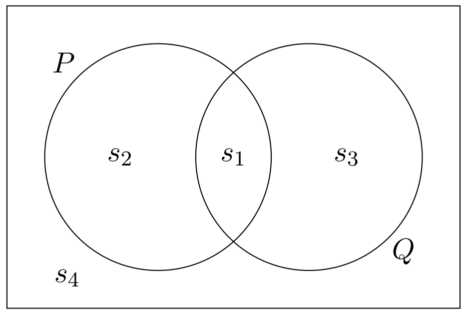
2.1.2 State-descriptions

Suppose we are working with a language that has just two atomic propositions, P and Q. Looking back at Figure 2.1, we can see that these propositions divide the space of possible worlds into four mutually exclusive, jointly exhaustive regions. Figure 2.3 labels those regions s_1, s_2, s_3, and s_4. Each of the regions corresponds to one of the lines in the following truth-table:

	P	Q	state-description
s_1	T	T	$P \,\&\, Q$
s_2	T	F	$P \,\&\, {\sim}Q$
s_3	F	T	${\sim}P \,\&\, Q$
s_4	F	F	${\sim}P \,\&\, {\sim}Q$

Each line on the truth-table can also be described by a kind of proposition called a **state-description**. A state-description in language $\mathcal{L}$ is a conjunction in which (1) each conjunct is either an atomic proposition of $\mathcal{L}$ or its negation; and (2) each atomic proposition of $\mathcal{L}$ appears exactly once. For example, $P \,\&\, Q$ and ${\sim}P \,\&\, Q$ are each state-descriptions. A state-description succinctly describes the possible worlds associated with a line on the truth-table. For example, the possible worlds in region s_3 are

Figure 2.3: Four mutually exclusive, jointly exhaustive regions



just those in which P is false and Q is true; in other words, they are just those in which the state-description $\sim P \,\&\, Q$ is true. Given any language, its state-descriptions will form a partition.

Notice that the state descriptions available for use are dependent on the language we are working with. If instead of language $\mathcal{L}$ we are working with a language $\mathcal{L}'$ with three atomic propositions (P, Q, and R), we will have eight state-descriptions available instead of $\mathcal{L}$'s four. (You'll work with these eight state-descriptions in Exercise 2.1. For now we'll go back to working with language $\mathcal{L}$ and its paltry four.)

Each proposition in a language (except for contradictory propositions) has an equivalent that is a disjunction of state-descriptions. We call this disjunction the proposition's **disjunctive normal form**. For example, the proposition $P \vee Q$ is true in regions s_1, s_2, and s_3. Thus

$$P \vee Q \;=\!\mid\models\; (P \,\&\, Q) \vee (P \,\&\, \sim Q) \vee (\sim P \,\&\, Q) \tag{2.1}$$

The proposition on the righthand side is the disjunctive normal form equivalent of $P \vee Q$. To find the disjunctive normal form of a non-contradictory proposition, figure out which lines of the truth-table it's true on, then make a disjunction of the state-descriptions associated with each such line.[5]

2.1.3 Predicate logic

Sometimes we will want to work with languages that represent objects and properties. To do so, we will first identify a **universe of discourse**, the total set of objects under discussion. Each object in the universe of discourse will be represented by a **constant**, which will usually be a lower-case letter

$(a, b, c, \ldots)$. Properties of those objects and relations among them will be represented by **predicates**, which will be capital letters.

Relations among propositions in such a language are exactly as described in the previous sections, except that we have two new kinds of propositions. First, our atomic propositions are now generated by applying a predicate to a constant, as in "Fa". Second, we can generate quantified sentences, as in "$(\forall x)(Fx \supset \sim Fx)$". Since we will be using predicate logic rarely, I won't work through the details here; a thorough treatment can be found in any introductory logic text.

I do want to emphasize, though, that as long as we restrict our attention to finite universes of discourse, all the predicate logic we need can be handled by the propositional machinery discussed above. If, say, our only two constants are a and b and our only predicate is F, then the only atomic propositions in $\mathcal{L}$ will be Fa and Fb, for which we can build a standard truth-table:

Fa	Fb	state-description
T	T	$Fa \,\&\, Fb$
T	F	$Fa \,\&\, \sim Fb$
F	T	$\sim Fa \,\&\, Fb$
F	F	$\sim Fa \,\&\, \sim Fb$

For any proposition containing a quantifier, we can find an equivalent composed entirely of atomic propositions and propositional connectives. A universally-quantified sentence will be equivalent to a *conjunction* of its substitution instances, while an existentially-quantified sentence will be equivalent to a *disjunction* of its substitution instances. For example, when our only two constants are a and b we have:

$$(\exists x)Fx \mathrel{=\!\!\mid\!\models} Fa \vee Fb \tag{2.2}$$

$$(\forall x)(Fx \supset \sim Fx) \mathrel{=\!\!\mid\!\models} (Fa \supset \sim Fa) \,\&\, (Fb \supset \sim Fb) \tag{2.3}$$

As long as we stick to finite universes of discourse, every proposition will have an equivalent that uses only propositional connectives. So even when we work in predicate logic, every non-contradictory proposition will have an equivalent in disjunctive normal form.

2.2 Probability distributions

A **distribution** over language $\mathcal{L}$ assigns a real number to each proposition in the language.[6] Bayesians represent an agent's degrees of belief as a distribution over a language; I will use "cr" to symbolize an agent's credence distribution. For example, if an agent is 70% confident that it will rain tomorrow, we will write

$$\mathrm{cr}(R) = 0.7 \tag{2.4}$$

where R is the proposition that it will rain tomorrow. Another way to put this is that the agent's **unconditional credence** in rain tomorrow is 0.7. (*Unconditional* credences contrast with *conditional* credences, which we will discuss in Chapter 3.)

Bayesians hold that a *rational* credence distribution satisfies certain rules. Among these are our first three core rules, **Kolmogorov's axioms**:

Non-Negativity: For any proposition P in $\mathcal{L}$, $\mathrm{cr}(P) \geqslant 0$.

Normality: For any tautology T in $\mathcal{L}$, $\mathrm{cr}(\mathsf{T}) = 1$.

Finite Additivity: For any mutually exclusive propositions P and Q in $\mathcal{L}$, $\mathrm{cr}(P \vee Q) = \mathrm{cr}(P) + \mathrm{cr}(Q)$

Kolmogorov's axioms are often referred to as "the probability axioms". Mathematicians call any distribution that satisfies these axioms a **probability distribution**. Kolmogorov (1950) was the first to articulate these axioms as the foundation of mathematical probability theory.[7]

Warning: To a mathematician, these axioms *define* what it is for a distribution to be a probability distribution. This is distinct from the way we use the word "probability" in everyday life. For one thing, the word "probability" in English may not mean the same thing in every use. And even if it does, it would be a substantive philosophical thesis that "probabilities" can be represented by a numerical distribution satisfying Kolmogorov's axioms. Going in the other direction, there are numerical distributions satisfying these axioms that don't count as "probabilistic" in any ordinary sense. For example, we could invent a distribution "tv" that assigns 1 to every true proposition and 0 to every false proposition. To a mathematician, the fact that tv satisfies Kolmogorov's axioms makes it a probability distribution. But a proposition's tv-value might not match its probability in the everyday sense. Improbable propositions can turn out to be true (I just rolled snake-eyes!), and propositions with high probabilities can turn out to be false.

Probabilism is the philosophical view that rationality requires an agent's credences to form a probability distribution (that is, to satisfy Kolmogorov's axioms). Probabilism is attractive in part because it has intuitively appealing consequences. For example, from the probability axioms we can prove:

Negation: For any proposition P in $\mathcal{L}$, $\mathrm{cr}(\sim P) = 1 - \mathrm{cr}(P)$.

According to Negation, rationality requires an agent with $\mathrm{cr}(R) = 0.7$ to have $\mathrm{cr}(\sim R) = 0.3$. Among other things, Negation embodies the sensible thought that if you're highly confident that a proposition is true, you should be unconfident that its negation is.

Usually I'll leave it as an exercise to prove that a particular consequence follows from the probability axioms, but in this case I'll lay out a proof to show how it might be done.

Negation Proof:

(1)	P and $\sim P$ are mutually exclusive.	logic
(2)	$\mathrm{cr}(P \vee \sim P) = \mathrm{cr}(P) + \mathrm{cr}(\sim P)$	(1), Finite Additivity
(3)	$P \vee \sim P$ is a tautology.	logic
(4)	$\mathrm{cr}(P \vee \sim P) = 1$	(3), Normality
(5)	$1 = \mathrm{cr}(P) + \mathrm{cr}(\sim P)$	(2), (4)
(6)	$\mathrm{cr}(\sim P) = 1 - \mathrm{cr}(P)$	(5), algebra

2.2.1 Consequences of the probability axioms

Below are a number of further consequences of the probability axioms. Again, these consequences are listed in part to demonstrate the intuitive things that follow from the probability axioms. But I'm also listing them because they'll be useful in future proofs.

Maximality: For any proposition P in $\mathcal{L}$, $\mathrm{cr}(P) \leqslant 1$.

Contradiction: For any contradiction F in $\mathcal{L}$, $\mathrm{cr}(\mathsf{F}) = 0$.

Entailment: For any propositions P and Q in $\mathcal{L}$, if $P \vDash Q$ then $\mathrm{cr}(P) \leqslant \mathrm{cr}(Q)$.

Equivalence: For any propositions P and Q in $\mathcal{L}$, if $P \dashv\vDash Q$ then $\mathrm{cr}(P) = \mathrm{cr}(Q)$.

General Additivity: For any propositions P and Q in $\mathcal{L}$, $\mathrm{cr}(P \vee Q) = \mathrm{cr}(P) + \mathrm{cr}(Q) - \mathrm{cr}(P \,\&\, Q)$.

Decomposition: For any propositions P and Q in $\mathcal{L}$,
$\mathrm{cr}(P) = \mathrm{cr}(P \,\&\, Q) + \mathrm{cr}(P \,\&\, {\sim}Q)$.

Partition: For any finite partition of propositions in $\mathcal{L}$, the sum of their unconditional cr-values is 1.

Together, Non-Negativity and Maximality establish the bounds of our credence scale. Rational credences will always fall between 0 and 1 (inclusive). Working within these bounds, Bayesians represent certainty that a proposition is true as a credence of 1 and certainty that a proposition is false as credence 0. The upper bound is arbitrary—we could have set it at whatever positive number we wanted. But using 0 and 1 lines up nicely with everyday talk of being 0% confident or 100% confident in particular propositions, and also with various considerations of frequency and chance discussed later in this book.

Entailment is motivated just as we motivated Comparative Confidence in Chapter 1; we've simply moved from an expression in terms of confidence orderings to one using numerical credences. Understanding equivalence as mutual entailment, Entailment entails Equivalence. General Additivity is a generalization of Finite Additivity that allows us to calculate an agent's credence in any disjunction, not just a disjunction of mutually exclusive disjuncts. (When the disjuncts *are* mutually exclusive, their conjunction is a contradiction, the $\mathrm{cr}(P \,\&\, Q)$ term equals 0, and General Additivity takes us back to Finite Additivity.) The Decomposition and Partition rules naturally go together. In Partition, you have a set of mutually exclusive propositions with a tautological disjunction, so their unconditional credences add up to the tautology's credence of 1. In Decomposition you have two mutually exclusive propositions whose disjunction is equivalent to P, so their unconditional credences add up to proposition P's.

Finally, here's a trick that involves multiple applications of Finite Additivity. Suppose we have a finite set of propositions $P, Q, R, S, \ldots$ that are mutually exclusive. By Finite Additivity,

$$\mathrm{cr}(P \vee Q) = \mathrm{cr}(P) + \mathrm{cr}(Q) \tag{2.5}$$

Logically, since P and Q are each mutually exclusive with R, $P \vee Q$ is also mutually exclusive with R. So Finite Additivity yields

$$\mathrm{cr}([P \vee Q] \vee R) = \mathrm{cr}(P \vee Q) + \mathrm{cr}(R) \tag{2.6}$$

Combining Equations (2.5) and (2.6) then gives us

$$\mathrm{cr}(P \vee Q \vee R) = \mathrm{cr}(P) + \mathrm{cr}(Q) + \mathrm{cr}(R) \tag{2.7}$$

Next we would invoke the fact that $P \vee Q \vee R$ is mutually exclusive with S to derive

$$\mathrm{cr}(P \vee Q \vee R \vee S) = \mathrm{cr}(P) + \mathrm{cr}(Q) + \mathrm{cr}(R) + \mathrm{cr}(S) \tag{2.8}$$

and repeating this process for each element of the set, we'd eventually have

$$\mathrm{cr}(P \vee Q \vee R \vee S \vee \ldots) = \mathrm{cr}(P) + \mathrm{cr}(Q) + \mathrm{cr}(R) + \mathrm{cr}(S) + \ldots \tag{2.9}$$

The idea here is that once you have Finite Additivity for proposition sets of size 2, you have it for propositions sets of any larger finite size as well. When the propositions in a finite set are mutually exclusive, the probability of their disjunction equals the sum of the probabilities of the disjuncts.

2.2.2 A Bayesian approach to the Lottery

In upcoming sections I'll explain two alternative ways of thinking about the probability calculus. But first let's use it to *do* something: a Bayesian analysis of the situation in the Lottery Paradox. Recall the scenario from Chapter 1: A fair lottery has one million tickets.[8] An agent is skeptical of each ticket that it will win, but takes it that some ticket will win. In Chapter 1 we saw that it's difficult to articulate norms on binary belief that depict this agent as believing rationally. But once we move to degrees of belief, the analysis is straightforward.

We'll use a language in which the constants $a, b, c, \ldots$ stand for the various tickets in the lottery, and the predicate W says that a particular ticket wins. A reasonable credence distribution over the resulting language sets

$$\mathrm{cr}(Wa) = \mathrm{cr}(Wb) = \mathrm{cr}(Wc) = \ldots = 1/1{,}000{,}000 \tag{2.10}$$

Negation then gives us

$$\mathrm{cr}({\sim}Wa) = \mathrm{cr}({\sim}Wb) = \mathrm{cr}({\sim}Wc) = 1 - 1/1{,}000{,}000 = 0.999999 \tag{2.11}$$

reflecting the agent's high confidence for each ticket that that ticket won't win.

What about the disjunction saying that some ticket will win? Since the $Wa, Wb, Wc, \ldots$ propositions are mutually exclusive,[9] we can use multiple applications of Finite Additivity and the trick discussed at the end of the previous section to derive

$$\begin{aligned} &\mathrm{cr}(Wa \vee Wb \vee Wc \vee Wd \vee \ldots) = \\ &\qquad \mathrm{cr}(Wa) + \mathrm{cr}(Wb) + \mathrm{cr}(Wc) + \mathrm{cr}(Wd) + \ldots \end{aligned} \tag{2.12}$$

On the righthand side of Equation (2.12) we have one million disjuncts, each of which has a value of 1/1,000,000. Thus the credence on the lefthand side is 1.

We now have a model of the Lottery situation in which the agent is both highly confident that some ticket will win and highly confident of each ticket that it will not. (Constructing a similar model of the Preface is left as an exercise for the reader.) There is no tension with the rules of rational confidence represented in Kolmogorov's axioms. The Bayesian model not only accommodates but *predicts* that if an agent has a small confidence in each proposition of the form Wx, is certain that no two of those propositions can be true at once, and yet has a high enough number of Wx propositions lying around, that agent will be certain (or close to certain) that at least one of the Wx is true.

This analysis also reveals why it's difficult to simlutaneously maintain both the Lockean thesis and the Belief Consistency norm from Chapter 1. The Lockean thesis implies that a rational agent believes a proposition just in case her credence in that proposition is above some numerical threshold (where the threshold is greater than 0.5 but less than 1). For any such threshold we pick, it's possible to generate a Lottery-type scenario in which the agent's credence that at least one ticket will win clears the threshold, but her credence for any given ticket that that ticket will lose also clears the threshold. Given the Lockean thesis, a rational agent will therefore believe that at least one ticket will win but also believe of each ticket that it will lose. This violates Belief Consistency, which says that every rational belief set is logically consistent.

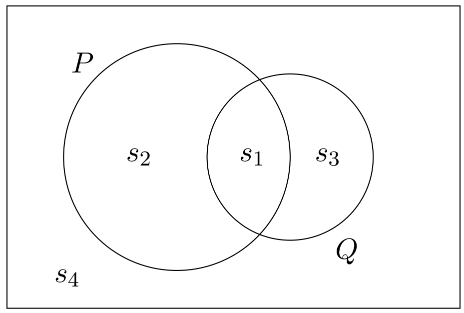
2.2.3 Probabilities are weird! The Conjunction Fallacy

As you work with credences it's important to remember that probabilistic relations can function very differently from the relations among categorical concepts that inform many of our intuitions. In the Lottery situation it's perfectly rational for an agent to be highly confident of a disjunction while having low confidence in each of its disjuncts. That may seem strange.

Tversky and Kahneman (1983) offer another probabilistic example that runs counter to most people's intuitions. In a famous study, they presented subjects with the following prompt:

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Figure 2.4: Areas equal to unconditional credences



The subjects were then asked to rank the probabilities of the following propositions (among others):

- Linda is active in the feminist movement.
- Linda is a bank teller.
- Linda is a bank teller and is active in the feminist movement.

The "great majority" of Tversky and Kahneman's subjects ranked the conjunction as more probable than the bank teller proposition. But this violates the probability axioms! A conjuction will always entail each of its conjuncts. By our Entailment rule—which follows from the probability axioms—the conjunct must be at least as probable as the conjunction. Being more confident in a conjunction than its conjunct is known as the **Conjunction Fallacy**.

2.3 Probability and Venn diagrams

Earlier we used Venn diagrams to visualize propositions and the relations among them. We can also use Venn diagrams to picture probability distributions. All we have to do is attach significance to something that was unimportant before: the *size* of regions in the diagram. We stipulate that the area of the entire rectangle is 1. The area of a region inside the rectangle equals the agent's unconditional credence in any proposition associated with that region. (Note that this visualization technique works only for credence functions that satisfy the probability axioms.)[10]

For example, consider Figure 2.4. There we've depicted a probabilistic credence distribution in which the agent is more confident of proposition P than she is of proposition Q, as indicated by the P-circle's being larger than the Q-circle. What about $\text{cr}(Q \,\&\, P)$ versus $\text{cr}(Q \,\&\, {\sim}P)$? On the diagram the region labeled s_3 is slightly bigger than the region labeled s_1, so the agent is slightly more confident of $Q \,\&\, {\sim}P$ than $Q \,\&\, P$. (When you construct your own Venn diagrams you need not include state-description labels like "s_3"; I've added them for later reference.)

> Warning: It is tempting to think that the size of a region in a Venn diagram represents the *number* of possible worlds in that region—the number of worlds that make the associated proposition true. But this would be a mistake. Just because an agent is more confident of one proposition than another does not necessarily mean she associates more possible worlds with the former than the latter. For example, if I tell you I have a weighted die that is more likely to come up 6 than any other number, your increased confidence in 6 does not necessarily mean that you think there are disproportionately many *worlds* in which the die lands 6. The area of a region in a Venn diagram is a useful visual representation of an agent's confidence in its associated proposition. We should not read too much out of it about the distribution of possible worlds.[11]

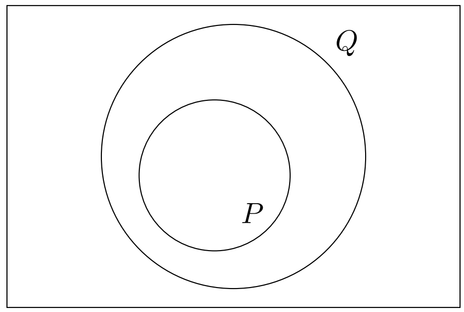
Venn diagrams make it easy to see why certain probabilistic relations hold. For example, take the General Additivity rule from Section 2.2.1. In Figure 2.4, the $P \vee Q$ region contains every point that is in the P-circle, in the Q-circle, or in both. We could calculate the area of that region by adding up the area of the P-circle and the area of the Q-circle, but in doing so we'd be counting the $P \,\&\, Q$ region (labeled s_1) twice. We adjust for the double-counting as follows:

$$\text{cr}(P \vee Q) = \text{cr}(P) + \text{cr}(Q) - \text{cr}(P \,\&\, Q) \tag{2.13}$$

That's General Additivity.

Figure 2.5 depicts a situation in which proposition P entails proposition Q. As discussed earlier, this requires the P-circle to be wholly contained within the Q-circle. But since areas now represent unconditional credences, the diagram makes it obvious that the cr-value of proposition Q must be at least as great as the cr-value of proposition P. That's exactly what our

Figure 2.5: $P \vDash Q$



Entailment rule requires. (It also shows why the Conjunction Fallacy is a mistake—imagine Q is the proposition that Linda is a bank teller and P is the proposition that Linda is a feminist bank teller.)

Venn diagrams can be a useful way of visualizing probabilistic relationships. Bayesians often clarify a complex situation by sketching a quick Venn diagram of the agent's credence distribution. There are limits to this technique; with more than 3 circles it becomes difficult to get all the overlapping regions one needs and to make areas proportional to credences. But there are also cases in which it's much easier to see on a diagram why a particular theorem holds than it is to prove that theorem from the axioms.

2.4 Stochastic truth-tables

Suppose I want to describe an agent's entire unconditional credence distribution over a particular language $\mathcal{L}$. There are infinitely many propositions in $\mathcal{L}$, so do I have to specify infinitely many values? Here the Equivalence rule helps. If the agent is rational, she will assign the same credence to equivalent propositions. So if I tell you her unconditional credence in one proposition, I've also told you her credence in its infinitely-many equivalents. All I really need to do is figure out how many *different* (non-equivalent) propositions are expressible in $\mathcal{L}$, and tell you the agent's credence in each of those.

The trouble is, that can still be a lot. If $\mathcal{L}$ has n atomic propositions, it will contain 2^{2^n} non-equivalent propositions. For 2 atomics that's only 16 credence values to specify, but by the time we reach 4 atomics it's up to 65,536 distinct values. Luckily, another shortcut is available.

Looking back at Figure 2.4, it's clear that we really need to specify only

4 values to determine the areas of all the regions in the figure. Suppose I give you the following **stochastic truth-table**:

	P	Q	cr
s_1	T	T	0.1
s_2	T	F	0.3
s_3	F	T	0.2
s_4	F	F	0.4

This truth-table tells you immediately what the agent's unconditional credences are in our four state-descriptions. But it can also be used to determine the agent's credences in other propositions. For example, since the Venn diagram region associated with $P \vee Q$ contains s_1, s_2, and s_3, we find $\mathrm{cr}(P \vee Q)$ by adding up the cr-values on the first three rows. In this case the result is 0.6.

A stochastic truth-table for language $\mathcal{L}$ is a standard truth-table for $\mathcal{L}$ to which one column has been added: a column specifying the agent's unconditional credence in each state-description. There are two rules for filling in the values in the final column:

1. Each value must be non-negative.

2. The values in the column must sum to 1.

The first rule is required because each of the values in the final column is the agent's unconditional credence in some state-description. By Non-Negativity, it can't be a negative number. The second rule is required because the state-descriptions form a partition, so by our Partition rule their unconditional credences must sum to 1.

As long as we're working with an agent whose credences satisfy the probability axioms, specifying unconditional credences for the state-descriptions in this way suffices to specify the rest of the agent's credence distribution as well. For any non-contradictory proposition in $\mathcal{L}$, we can find the agent's unconditional credence in that proposition by summing the values of the rows on which the proposition is true. (Contradictions automatically receive credence 0.)

Unconditional credences can be calculated this way because each row on which the proposition is true represents one disjunct in its disjunctive normal form. Since the disjuncts are mutually exclusive (being state-descriptions), we can find the credence of the whole disjunction by summing the credences of its parts. For example, we've already seen that

$$P \vee Q \;=\!\mid\models\; (P \,\&\, Q) \vee (P \,\&\, {\sim}Q) \vee ({\sim}P \,\&\, Q) \tag{2.14}$$

reflecting the fact that $P \vee Q$ is true on the first, second, and third rows of the truth-table. By Equivalence,

$$\mathrm{cr}(P \vee Q) = \mathrm{cr}[(P \,\&\, Q) \vee (P \,\&\, {\sim}Q) \vee ({\sim}P \,\&\, Q)] \tag{2.15}$$

Since the state-descriptions on the right are mutually exclusive, multiple applications of Finite Additivity yield

$$\mathrm{cr}(P \vee Q) = \mathrm{cr}(P \,\&\, Q) + \mathrm{cr}(P \,\&\, {\sim}Q) + \mathrm{cr}({\sim}P \,\&\, Q) \tag{2.16}$$

So we find the unconditional credence in $P \vee Q$ by summing the values on the first, second, and third rows of the table.

Stochastic truth-tables describe an entire credence distribution in an efficient manner; instead of specifying a credence value for each non-equivalent proposition in the language, we need only specify values for its state-descriptions. Credences in state-descriptions can then be used to calculate credences in other propositions.[12]

Stochastic truth-tables can also be used to prove theorems and solve problems. To do so, we simply replace the credence values with variables:

	P	Q	cr
s_1	T	T	a
s_2	T	F	b
s_3	F	T	c
s_4	F	F	d

This stochastic truth-table for an $\mathcal{L}$ with two atomic propositions makes no assumptions about the agent's specific credence values. It is therefore fully general, and can be used to prove general theorems about probability distributions. For example, on this table

$$\mathrm{cr}(P) = a + b \tag{2.17}$$

But a is just $\mathrm{cr}(P \,\&\, Q)$, and b is $\mathrm{cr}(P \,\&\, {\sim}Q)$. This gives us a very quick proof of the Decomposition rule from Section 2.2.1.

As for problem-solving, suppose I tell you that my credence distribution satisfies the probability axioms and also has the following features: I am certain of $P \vee Q$, and I am equally confident in Q and ${\sim}Q$. I then ask you to tell me my credence in $P \supset Q$.

You might be able to solve this problem by drawing a careful Venn diagram—perhaps you can even solve it in your head! If not, the stochastic truth-table provides a purely algebraic solution method. We start by expressing the constraints on my distribution as equations using the variables from the table. Given the second rule for filling out stochastic truth-tables above, we know that:

$$a + b + c + d = 1 \tag{2.18}$$

(Sometimes it also helps to invoke the first rule, writing inequalities specifying that a, b, c, and d are each greater than or equal to 0. In this particular case that wouldn't be useful.) Next we represent the fact that I am equally confident in Q and $\sim Q$:

$$\mathrm{cr}(Q) = \mathrm{cr}(\sim Q) \tag{2.19}$$

$$a + c = b + d \tag{2.20}$$

Finally, we represent the fact that I am certain of $P \vee Q$. The only line of the truth-table on which $P \vee Q$ is false is line s_4; if I'm certain of $P \vee Q$, I must assign this state-description a credence of 0. So

$$d = 0 \tag{2.21}$$

Now what value are we looking for? I've asked you for my credence in $P \supset Q$; that proposition is true on lines s_1, s_3, and s_4; so you need to find $a + c + d$. Applying a bit of algebra to Equations (2.18), (2.20), and (2.21), you should be able to determine that $a + c + d = 1/2$.

2.4.1 Working with alternate partitions

It's very common for a Bayesian epistemologist (or a statistician, or an economist) to say that when an agent rolls a fair six-sided die, her credence in each of the six possible outcomes should be 1/6. By repeated applications of Finite Additivity, the agent's credence in the disjunction that the die will land with one of its six faces up is 1. And by Negation, the agent's credence that something *else* will happen is 0. But couldn't the die spontaneously combust mid-roll? Couldn't it be destroyed by lasers? Couldn't it stop perfectly balanced on one corner? Should a rational agent assign these possibilities *no* credence whatsoever?

We will return to this question in our discussion of the Regularity Principle in Chapters 4 and 5. For the moment I will set it aside, and note that it's often useful methodologically to allow the ruling out of genuine logical possibilities without worrying about this maneuver's rationality. Pollsters calculating confidence intervals for their latest sampling data don't factor in the possibility that the United States will be overthrown before the next presidential election.

Section 2.1 defined various relations among propositions in terms of possible worlds. In that context, the appropriate set of possible worlds to consider was the full set of *logically* possible worlds. But following the methodological point just mentioned, it's often useful to simplify our models by narrowing our focus to an agent's **doxastically possible worlds**—the subset of logically possible worlds she hasn't ruled out of consideration.[13] For example, when we analyzed the Lottery scenario in Section 2.2.2, we effectively ignored possible worlds in which no tickets win the lottery or in which more than one ticket wins. Such worlds are logically possible, but for our purposes it's simpler to treat the agent as ruling them out of consideration.

All of our earlier definitions work just as well for the set of doxastically possible worlds as they do for the full set of logically possible worlds. So in our model of the lottery we treated the proposition that ticket a will win as mutually exclusive with the proposition that ticket b will win, allowing us to apply Finite Additivity to the disjunction of those propositions. If we were working with the full space of logically possible worlds we would have a world in which both those propositions are true, so they wouldn't count as mutally exclusive. But relative to the set of possible worlds we've supposed the agent entertains, they are.

Once we've confined our attention to doxastically possible worlds, we can still build a stochastic truth-table to describe the agent's credence distribution. But sometimes it will be more convenient to work with a partition of doxastic space other than our language's state-descriptions. It turns out that we can construct something like a stochastic truth-table for any partition.

For example, suppose I tell you I'm going to roll a loaded die that comes up 6 on half its rolls (with the remaining rolls distributed equally among the other numbers). We can represent your credence distribution in light of this information using a language with six atomic propositions (the die comes up 1, the die comes up 2, etc.). Six atomic propositions would give us a stochastic truth-table of 64 rows. Given the certainties we're supposing you have about how a die works, most of those rows—any row on which more than one number comes up, the row on which no number comes up— receive a credence of 0. But that's still an unwieldy truth-table to work with.

An alternate approach: For your space of doxastic possibilities, the atomic propositions that the die comes up 1, that the die comes up 2, etc. form a partition. You're certain that exactly one of these propositions is true. So we can build a table using just those six propositions:

proposition	cr
Die comes up 1.	1/10
Die comes up 2.	1/10
Die comes up 3.	1/10
Die comes up 4.	1/10
Die comes up 5.	1/10
Die comes up 6.	1/2

As in a stochastic truth-table, this table assigns each element of the partition an unconditional credence. Credences in other propositions can then be calculated just as before. If I ask how confident you are in the proposition that the die comes up odd, you add up the values on the rows on which that proposition is true. In this case that's the first, third, and fifth rows, so your credence in an odd roll is $1/10 + 1/10 + 1/10 = 3/10$.

2.5 What the probability calculus adds

In Chapter 1 we moved from describing agents' doxastic attitudes in terms of binary (categorical) beliefs and confidence comparisons to working with numerical degrees of belief. As we saw there, credences' added fineness of grain confers both advantages and disadvantages. On the one hand, credences allow us to say *how much more* confident an agent is of one proposition than another. On the other hand, assigning credences over a set of propositions immediately makes them all commensurable with respect to the agent's confidence, which may be an unrealistic result.

Chapter 1 also offered a norm for comparative confidence orderings:

Comparative Entailment: For any pair of propositions such that the first entails the second, rationality requires an agent to be at least as confident of the second as the first.

I now want to explore how Kolmogorov's probability axioms go beyond what this constraint requires.

Comparative Entailment can easily be derived from the probability axioms—we've already seen that by the Entailment rule, if $P \vDash Q$ then rationality requires $\mathrm{cr}(P) \leqslant \mathrm{cr}(Q)$. What's more interesting is how much of the probability calculus can be recreated simply by assuming that Comparative Entailment holds. We saw in Chapter 1 that if Comparative Entailment holds, a rational agent will assign equal, maximal confidence to all tautologies and equal, minimal confidence to all contradictions. This doesn't give specific *numerical confidence values* to contradictions and tautologies, because

Comparative Entailment doesn't work with numbers. But the probability axioms' 0-to-1 scale for credence values is fairly stipulative and arbitrary anyway. The real essence of Normality, Contradiction, Non-Negativity, and Maximality can be obtained from Comparative Entailment.

That leaves one axiom unaccounted for. To me the key insight of probabilism—and the element most responsible for Bayesianism's distinctive contributions to epistemology—is Finite Additivity. Finite Additivity places demands on rational credence that don't follow from any of the other norms we've seen. To see how, consider the following two credence distributions over a language with one atomic proposition:

Mr. Prob:	$\mathrm{cr}(\mathsf{F}) = 0$	$\mathrm{cr}(P) = 1/6$	$\mathrm{cr}(\sim P) = 5/6$	$\mathrm{cr}(\mathsf{T}) = 1$
Mr. Weak:	$\mathrm{cr}(\mathsf{F}) = 0$	$\mathrm{cr}(P) = 1/36$	$\mathrm{cr}(\sim P) = 25/36$	$\mathrm{cr}(\mathsf{T}) = 1$

From a confidence ordering point of view, Mr. Prob and Mr. Weak are identical; they each rank $\sim P$ above P and both those propositions between the tautology and the contradiction. Both agents satisfy Comparative Entailment. Both agents also satisfy the Non-Negativity and Normality probability axioms. But only Mr. Prob satisfies Finite Additivity. His credence in the tautologous disjunction $P \vee \sim P$ is the sum of his credences in its mutually exclusive disjuncts. Mr. Weak's credences, on the other hand, are **superadditive**: he assigns *more* credence to the disjunction than the sum of his credences in its mutually exclusive disjuncts. $(1 > 1/36 + 25/36)$

Probabilism goes beyond Comparative Entailment by exalting Mr. Prob over Mr. Weak. By endorsing Finite Additivity, the probabilist holds that Mr. Weak's credences have an *irrational* feature not present in Mr. Prob's. When we apply Bayesianism in later chapters, we'll see that Finite Additivity—a kind of linearity constraint—gives rise to some of the theory's most interesting and useful results.

Of course, the fan of comparative confidence orderings need not restrict herself to the Comparative Entailment norm. Chapter **??** will explore further comparative constraints that have been proposed. We will ask whether those non-numerical norms can replicate all the desirable results secured by Finite Additivity for the Bayesian credal regime. This will be an especially pressing question because the impressive numerical credence results come with a price. When we examine explicit philosophical arguments for the probability axioms in Part IV of this book, we'll find that while Normality and Non-Negativity can be straightforwardly argued for, Finite Additivity is the most difficult part of Bayesian Epistemology to defend successfully.

2.6 Exercises

Problem 2.1. (a) List all eight state-descriptions available in a language with the three atomic sentences P, Q, and R.

(b) Give the disjunctive normal form of $(P \vee Q) \supset R$.

Problem 2.2. Here's a fact: For any propositions P and Q, $P \vDash Q$ if and only if every disjunct in the disjunctive normal form equivalent of P is also a disjunct of the disjunctive normal form equivalent of Q.

(a) Use this fact to show that $(P \vee Q) \,\&\, R \vDash (P \vee Q) \supset R$.

(b) Explain why the fact is true. (Be sure to explain both the "if" direction and the "only if" direction!)

Problem 2.3. Explain why a language $\mathcal{L}$ with n atomic propositions can express exactly 2^{2^n} non-equivalent propositions. (Hint: Think about the number of state-descriptions available, and the number of distinct disjunctive normal forms.)

Problem 2.4. Suppose your universe of discourse contains only two objects, named by the constants "a" and "b".

(a) Find a quantifier-free equivalent of the proposition $(\forall x)[Fx \supset (\exists y)Gy]$.

(b) Find the disjunctive normal form of your quantifier-free proposition from part (a).

Problem 2.5. Consider the probabilistic credence distribution specified by this stochastic truth-table:

P	Q	R	cr
T	T	T	0.1
T	T	F	0.2
T	F	T	0
T	F	F	0.3
F	T	T	0.1
F	T	F	0.2
F	F	T	0
F	F	F	0.1

Calculate each of the following values on this distribution:

(a) $\mathrm{cr}(P \equiv Q)$

(b) $\mathrm{cr}(R \supset Q)$

(c) $\mathrm{cr}(P \,\&\, R) - \mathrm{cr}(\sim P \,\&\, R)$

(d) $\mathrm{cr}(P \,\&\, Q \,\&\, R)/\mathrm{cr}(R)$

Problem 2.6. Can a probabilistic credence distribution assign $\mathrm{cr}(P) = 0.5$, $\mathrm{cr}(Q) = 0.5$, and $\mathrm{cr}(\sim P \,\&\, \sim Q) = 0.8$? Explain why or why not.*

Problem 2.7. Can an agent have a probabilistic cr-distribution meeting all of the following constraints?

1. The agent is certain of $A \supset (B \equiv C)$.
2. The agent is equally confident of B and $\sim B$.
3. The agent is twice as confident of C as $C \,\&\, A$.
4. $\mathrm{cr}(B \,\&\, C \,\&\, \sim A) = 1/5$.

If not, prove that it's impossible. If so, provide a stochastic truth-table and demonstrate that the resulting distribution satisfies each of the four constraints.

Problem 2.8. Starting with only the probability axioms and Negation, prove all of the probability rules listed in Section 2.2.1. Your proofs must be straight from the axioms—no using Venn diagrams or stochastic truth-tables! Once you prove a rule you may use it in further proofs. (Hint: You may want to prove them in an order different from that in which they're listed.)

Problem 2.9. Tversky and Kahneman's finding that ordinary subjects commit the Conjunction Fallacy has held up to a great deal of experimental scrutiny. Kolmogorov's axioms make it clear that the propositions involved cannot range from most probable to least probable in the way subjects consistently rank them. Do you have any suggestions for *why* subjects might consistently make this mistake? Is there any way to read what the subjects are doing as rationally acceptable?

Problem 2.10. Recall Mr. Prob and Mr. Weak from Section 2.5. Mr. Weak assigns lower credences to each contingent proposition than does Mr. Prob. While Mr. Weak's distribution satisfies Non-Negativity and Normality, it violates Finite Additivity by being superadditive: it contains a disjunction

*I owe this problem to Julia Staffel.

whose credence is *greater* than the sum of the credences of its mutually exclusive disjuncts.

Construct a credence distribution for Mr. Bold over language $\mathcal{L}$ with single atomic proposition P. Mr. Bold should rank every proposition in the same order as Mr. Prob and Mr. Weak. Mr. Bold should also satisfy Non-Negativity and Normality. But Mr. Bold's distribution should be **subadditive**: it should contain a disjunction whose credence is *less* than the sum of the credences of its mutually exclusive disjuncts.

2.7 Further reading

Introductions and Overviews

Merrie Bergmann, James Moor, and Jack Nelson (2013). *The Logic Book*. 6th edition. New York: McGraw Hill

One of many available texts that thoroughly covers the logical material assumed in this book.

Ian Hacking (2001). *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press

Brian Skyrms (2000). *Choice & Chance: An Introduction to Inductive Logic*. 4th. Stamford, CT: Wadsworth

Each of these books contains a Chapter 6 offering an entry-level, intuitive discussion of the probability rules—though neither explicitly uses Kolmogorov's axioms. Hacking has especially nice applications of probabilistic reasoning, along with many counter-intuitive examples like the Conjunction Fallacy from our Section 2.2.3.

Classic Texts

A. N. Kolmogorov (1950). *Foundations of the Theory of Probability*. Translation edited by Nathan Morrison. New York: Chelsea Publishing Company

Text in which Kolmogorov laid out his famous axiomatization of probability theory.

Extended Discussion

J. Robert G. Williams (2015). Probability and Non-Classical Logic. In: *Oxford Handbook of Probability and Philosophy*. Ed. by Alan Hájek and Christopher R. Hitchcock. Oxford University Press

Covers probability distributions in non-classical logics, such as logics with non-classical entailment rules and logics with more than one truth-value. Also briefly discusses probability distributions in logics with extra connectives and operators, such as modal logics.

Branden Fitelson (2008). A Decision Procedure for Probability Calculus with Applications. *The Review of Symbolic Logic* 1, pp. 111–125

Fills in the technical details of solving probability problems algebraically using stochastic truth-tables, including the relevant meta-theory. Also describes a Mathematica package that will solve probability problems and evaluate probabilistic conjectures for you, downloadable for free at http://fitelson.org/PrSAT/.

Notes

[1]Among various alternatives, some authors assign degrees of belief to sentences, statements, or sets of events. Some views of propositions make them identical to one of these alternatives. I will not assume much about what propositions are, except that they are capable of having truth-values, are expressible by declarative sentences, and have enough internal structure to contain logical operators. This last assumption could be lifted with a bit of work.

[2]Bayesians sometimes define degrees of belief over a **sigma algebra**. A sigma algebra is a set of sets that is closed under union, intersection, and complementation. Given a language $\mathcal{L}$, the sets of possible worlds associated with the propositions in that language form a sigma algebra. The algebra is closed under union, instersection, and complementation because the propositions in $\mathcal{L}$ are closed under disjunction, conjunction, and negation (respectively).

(Strictly speaking, a sigma algebra is closed under *countably many* applications of set operations, so some sigma algebras are representable in a language of propositions only if we allow infinitely many atomic propositions and propositions of infinite length. We will ignore these complications here.)

[3]I'm also going to be fairly cavalier about the use-mention distinction, corner-quotes, and the like.

[4]Throughout this book we will be assuming a classical logic, in which each proposition has exactly one of two available truth-values and entailment obeys the inference rules taught in standard introductory logic classes. For information about probability in non-classical logics, see the Further Readings at the end of this chapter.

[5]Strictly, in order to get the result that the state-descriptions in a language form a partition and the result that each non-contradictory proposition has a *unique* disjunctive normal form, we need to further regiment our definitions. To our definition of a state-description we add that the atomic propositions must appear in alphabetical order. We then introduce a canonical ordering of the state-descriptions in a language (say, the order in which they appear in a standardly-ordered truth-table), and require disjunctive normal form propositions to contain their disjuncts in canonical order with no repeats.

[6]In the statistics community probability distributions are often assigned over random variables. Propositions are then thought of as dichotomous random variables capable of taking only the values "true" and "false" (or 1 and 0). Only rarely in this book will we look past distributions over propositions to more general random variables.

[7]Some authors also include Countable Additivity—which we'll discuss in Chapter 5—among "Kolmogorov's axioms". I'll use this phrase to pick out only Non-Negativity, Normality, and Finite Additivity.

[8]This analysis could easily be generalized to any large number of tickets other than one million.

[9]You may be concerned that $Wa, Wb, Wc, \dots$ are not strictly speaking mutually exclusive—for instance, there are *logically* possible worlds in which both Wa and Wb are true—so Finite Additivity does not apply. We'll address this concern in Section 2.4.1.

[10]A probability distribution over sets of possible worlds is an example of what mathematicians call a "measure". The function that takes any region of a two-dimensional space and outputs its area is also a measure. That's what makes probabilities representable by areas in a rectangle.

[11]To avoid the confusion discussed here, some authors use "muddy" Venn diagrams in which all atomic propositions have regions of the same size and probability weights are indicated by piling up more or less "mud" on top of particular regions. Muddy Venn diagrams are difficult to depict on two-dimensional paper, so I've stuck with representing increased confidence as increased region size.

[12]We have argued *from* the assumption that an agent's credences satisfy the probability axioms *to* the conclusion that her unconditional credence in any non-contradictory proposition is the sum of her credences in the disjuncts of its disjunctive normal form. One can also argue successfully in the other direction. Suppose I stipulate an agent's credence distribution over language $\mathcal{L}$ as follows: (1) I stipulate unconditional credences for $\mathcal{L}$'s state-descriptions that are non-negative and sum to 1; (2) I stipulate that for every other non-contradictory proposition in $\mathcal{L}$, the agent's credence in that proposition is the sum of her credences in the disjuncts of that proposition's disjunctive normal form; and (3) I stipulate that the agent's credence in each contradiction in 0. We can then *prove* that the credence distribution I've just stipulated satisfies Kolmogorov's three probability axioms. I'll leave the (somewhat challenging) proof as an exercise for the reader.

[13]Philosophers sometimes describe the worlds an agent hasn't ruled out of consideration as her "epistemically possible worlds". Yet that term also carries the more specific meaning of worlds not ruled out by what the agent *knows*. So I'll discuss doxastically possible worlds, which concern what an agent *takes* to be ruled out rather than what she *knows* to be ruled out. (Note that ruling a possibility out in this sense is more than just *believing* the possibility does not obtain. One can believe a proposition is false without *ruling out* the possibility that it's true.)