2
Probability Distributions

This chapter introduces Kolmogorov's probability axioms, the first three core normative rules of Bayesian epistemology. 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* what it is to be a probability distribution, and they 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 three alternative forms: Venn diagrams, probability tables, and odds.

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 ranking.

2.1 Propositions and propositional logic

Following the discussion in Chapter 1, we will assume that degrees of belief are propositional attitudes—they are attitudes agents assign 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

Fundamentals of Bayesian Epistemology 1: Introducing Credences. Michael G. Titelbaum, Oxford University Press.
 DOI: 10.1093/oso/9780198707608.003.0002

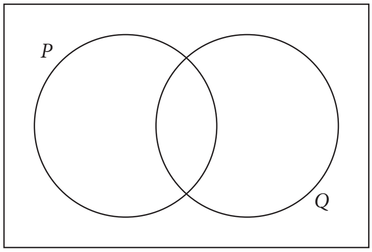

Figure 2.1 The space of possible worlds

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 $\equiv$. A **negation** $\sim P$ is true just in case P is false. A **conjunction** P & Q is true just in case its **conjuncts** P and Q are both true. "$\vee$" represents inclusive "or"; a **disjunction** $P \vee Q$ is false just in case its **disjuncts** P and Q are both false. "$\supset$" represents the **material conditional**; $P \supset Q$ is false just in case its **antecedent** P is true and its **consequent** Q is false. A **material biconditional** $P \equiv Q$ 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 (named after a logical technique developed by John Venn), the possible worlds are represented as points inside the rectangle. Proposition P might be true in some of those worlds, false in others. We 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

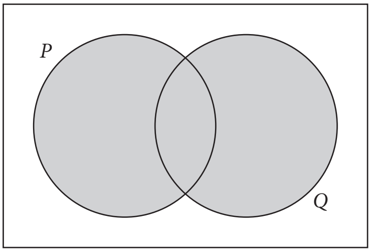

Figure 2.2 The set of worlds associated with $P \vee Q$

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 deem it 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 relations among propositions, a word about notation. I said we will use capital letters to represent specific 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 saying 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 out especially for phrases like: "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 "$P \dashv\vDash Q$". 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 have rolled 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 of 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, each proposition in the set refutes each of the others. (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

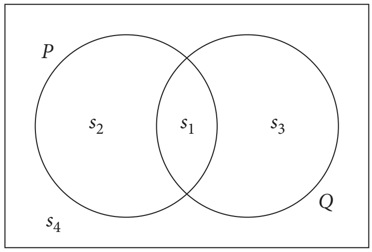

Figure 2.3 Four mutually exclusive, jointly exhaustive regions

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 together 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 in a partition 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 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.[5]

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 $\mathcal{L}'$ containing 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.)

Every non-contradictory proposition in a language 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

$$P \vee Q \Dashv\vDash (P \,\&\, Q) \vee (P \,\&\, {\sim}Q) \vee ({\sim}P \,\&\, Q) \tag{2.1}$$

The proposition on the right-hand 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.[6]

2.1.3 Predicate logic

Sometimes we will want to work with languages that represent objects and properties. To do so, we 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, ...). 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 rarely be using predicate logic, 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 logical relations 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 in this language containing a quantifier, we can find an equivalent composed entirely of atomic propositions and propositional connectives. To do this we need the notion of a **substitution instance**: a substitution instance of a quantified sentence is produced by removing the quantifier and replacing its variable throughout what remains with the same constant. (So, for example, $Fa \supset \sim Fa$ is a substitution instance of $(\forall x)(Fx \supset \sim Fx)$.) A universally quantified sentence is equivalent to a *conjunction* of all its substitution instances for constants in $\mathcal{L}$, while an existentially quantified sentence is equivalent to a *disjunction* of its substitution instances. For example, when our only two constants are a and b we have:

$$(\forall x)(Fx \supset \sim Fx) \;\dashv\vDash\; (Fa \supset \sim Fa) \,\&\, (Fb \supset \sim Fb) \tag{2.2}$$

$$(\exists x)Fx \;\dashv\vDash\; Fa \vee Fb \tag{2.3}$$

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 The probability axioms

A **distribution** over language $\mathcal{L}$ assigns a real number to each proposition in the language.[7] 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.) The higher the numerical value of an agent's unconditional credence in a proposition, the more confident the agent is that that proposition is true.

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 (1933/1950) was the first to articulate these axioms as the foundation of mathematical probability theory.[8]

Warning

Kolmogorov's work inaugurated a mathematical field of probability theory distinct from the philosophical study of what probability is. While this was an important advance, it gave the word "probability" a special meaning in mathematical circles that can generate confusion elsewhere.

For a twenty-first-century mathematician, Kolmogorov's axioms *define* what it is for a distribution to be a "probability distribution". This is distinct from the way people use "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 (in the everyday sense) can be represented by a numerical distribution satisfying Kolmogorov's axioms. Going in the other direction, there are numerical distributions satisfying the 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 (the Titanic should've made it to port).

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 dubious that its negation is.

Usually I'll leave it as an exercise to prove that a particular consequence follows from the probability axioms, but here I will prove Negation as an example for the reader.

Negation Proof:

	Claim	Justification
(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 illustrate intuitive things that follow from the axioms. But I'm also listing them because they'll be useful in future proofs.

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

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

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

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

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

Finite Additivity (Extended): For any finite set of mutually exclusive propositions $\{P_1, P_2, \ldots, P_n\}$,
$\text{cr}(P_1 \vee P_2 \vee \ldots \vee P_n) = \text{cr}(P_1) + \text{cr}(P_2) + \ldots + \text{cr}(P_n)$.

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

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). Given these bounds, Bayesians represent absolute certainty that a proposition is true as a credence of 1 and absolute certainty that a proposition is false as credence 0. The upper bound is arbitrary—we could have set it at whatever positive real 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 plausible for all the same reasons Comparative Entailment was plausible in Chapter 1; we've simply moved from an expression in terms of confidence comparisons 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, whether the disjuncts are mutually exclusive or not. (When the disjuncts *are* mutually exclusive, their conjunction is a contradiction, the $\text{cr}(P \,\&\, Q)$ term equals 0, and General Additivity takes us back to Finite Additivity.)

Finite Additivity (Extended) can be derived by repeatedly applying Finite Additivity. Begin with any finite set of mutually exclusive propositions $\{P_1, P_2, \ldots, P_n\}$. By Finite Additivity,

$$\mathrm{cr}(P_1 \vee P_2) = \mathrm{cr}(P_1) + \mathrm{cr}(P_2) \tag{2.5}$$

Logically, since P_1 and P_2 are each mutually exclusive with P_3, $P_1 \vee P_2$ is also mutually exclusive with P_3. So Finite Additivity yields

$$\mathrm{cr}([P_1 \vee P_2] \vee P_3) = \mathrm{cr}(P_1 \vee P_2) + \mathrm{cr}(P_3) \tag{2.6}$$

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

$$\mathrm{cr}(P_1 \vee P_2 \vee P_3) = \mathrm{cr}(P_1) + \mathrm{cr}(P_2) + \mathrm{cr}(P_3) \tag{2.7}$$

Next we would invoke the fact that $P_1 \vee P_2 \vee P_3$ is mutually exclusive with P_4 to derive

$$\mathrm{cr}(P_1 \vee P_2 \vee P_3 \vee P_4) = \mathrm{cr}(P_1) + \mathrm{cr}(P_2) + \mathrm{cr}(P_3) + \mathrm{cr}(P_4) \tag{2.8}$$

Clearly this process iterates as many times as we need to reach

$$\mathrm{cr}(P_1 \vee P_2 \vee \ldots \vee P_n) = \mathrm{cr}(P_1) + \mathrm{cr}(P_2) + \ldots + \mathrm{cr}(P_n) \tag{2.9}$$

The idea here is that once you have Finite Additivity for proposition sets of size two, you have it for proposition 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.

Combining Finite Additivity and Equivalence yields Decomposition. For any P and Q, P is equivalent to the disjunction of the mutually exclusive propositions $P \& Q$ and $P \& {\sim}Q$, so $\mathrm{cr}(P)$ must equal the sum of the cr-values of those two. Partition then takes a finite set of mutually exclusive propositions whose disjunction is a tautology. By Finite Additivity (Extended) the cr-values of the propositions in the partition must sum to the cr-value of the tautology, which by Normality must be 1.

2.2.2 A Bayesian approach to the Lottery scenario

In future sections I'll explain some alternative ways of thinking about the probability calculus. But first let's use probabilities 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.[9] 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 plausible 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

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

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, Finite Additivity (Extended) yields

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

On the right-hand side of Equation (2.12) we have one million terms, each of which has a value of 1/1,000,000. Thus the credence on the left-hand side equals 1.

The Lottery *Paradox* is a problem for particular norms on binary belief. We haven't done anything to resolve that paradox here. Instead, we've shown that the lottery situation giving rise to the paradox can be easily modeled by Bayesian means. We've build 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 for 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 a rational agent has a small confidence in each of a set of mutually exclusive propositions, yet has a large enough number of those propositions available, that agent will be certain (or close to certain) that at least one of the propositions is true.

This analysis also reveals why it's difficult to simultaneously 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. For any such threshold we pick (less than 1), 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 Doxastic possibilities

In the previous section we considered propositions of the form Wx, each of which says of some particular ticket that it will win the lottery. To perform various calculations involving these W propositions, we assumed that they form a partition—that is, that they are mutually exclusive and jointly exhaustive. But you may worry that this isn't right: what about worlds in which ticket a and ticket b both win the lottery due to a clerical error, or worlds in which no ticket wins the lottery, or worlds in which the lottery never takes place, or worlds in which humans never evolve? These worlds are **logically possible**—the laws of logic alone don't rule them out. Yet the credence distribution we crafted for our agent assigns these worlds degree of belief 0. Could it ever be rational for an agent to assign a logical possibility *no* credence whatsoever?

We will refer to the possible worlds an agent entertains as her **doxastically possible worlds**.[10] Perhaps a fully rational agent never rules out any logically possible world; if so, then a rational agent's set of doxastic possibilities is always the full set of logical possibilities, and includes worlds like the one in which every ticket wins, the one in which no ticket wins, the one in which no humans exist, etc. We will discuss this position when we turn to the Regularity Principle in Chapters 4 and 5. For the time being I want to note that even if a rational agent should never entirely rule out a logically possible world, it might be convenient in particular contexts for her to *temporarily* ignore certain worlds as live possibilities. Pollsters calculating confidence intervals for their latest sampling data don't factor in the possibility that our sun will explode before the next presidential election.

How is the probability calculus affected when an agent restricts her doxastically possible worlds to a proper subset of the logically possible worlds? 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 we can reinterpret those definitions as quantified over an agent's doxastically possible worlds. In our analysis of the Lottery scenario above, we effectively ignored possible worlds in which no tickets win the lottery or in which more than one ticket wins. For our purposes it was simpler to suppose that the agent rules them out of consideration. So our Bayesian model treated each *Wx* proposition as mutually exclusive with all the others, allowing us to apply Finite Additivity to generate equations like (2.12). If we were working with the full space of logically possible worlds we would have worlds in which more than one *Wx* proposition was true, so those propositions wouldn't count as mutally exclusive. But relative to the set of possible worlds we've supposed the agent entertains, they are.

2.2.4 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.

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.

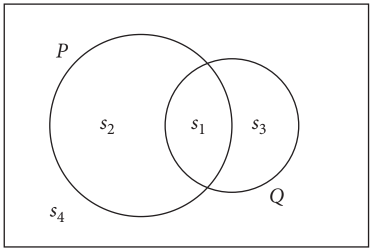

Figure 2.4 Areas equal to unconditional credences

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 Alternative representations of probability

2.3.1 Probabilities in 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 *sizes* 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 distributions that satisfy the probability axioms.)[11]

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 has slightly more area 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 reference.)

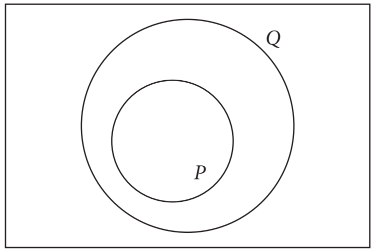

Figure 2.5 $P \vDash Q$

Warning

It's 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 six than any other number, your increased confidence in six does not necessarily mean that you think there are disproportionately many *worlds* in which the die lands six. 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 into it about the distribution of possible worlds.[12]

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 this double-counting as follows:

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

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 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 probability 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; when our languages grow beyond three or so atomic propositions 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 understand why a particular theorem holds by looking at a diagram than by working with the axioms.

2.3.2 Probability tables

Besides being represented visually in a Venn diagram, a probability distribution can be represented precisely and efficiently in a **probability table**. To build a probability table, we begin with a set of propositions forming a partition of the agent's doxastic possibilities. For example, suppose an agent is going to roll a loaded six-sided die that comes up six on half of its rolls (with the remaining rolls distributed equally among the other numbers). A natural partition of the agent's doxastic space uses the propositions that the die comes up one, the die comes up two, the die comes up three, etc. The resulting probability table looks like this:

proposition	cr
Die comes up one.	1/10
Die comes up two.	1/10
Die comes up three.	1/10
Die comes up four.	1/10
Die comes up five.	1/10
Die comes up six.	1/2

The probability table first lists the propositions in the partition. Then for each proposition it lists the agent's unconditional credence in that proposition. If the agent's credences satisfy the probability axioms, the credence values in the table will satisfy two important constraints:

1. Each value is non-negative.
2. The values in the column sum to 1.

The first rule follows from Non-Negativity, while the second follows from our Partition theorem.

Once we know the credences of partition members, we can calculate the agent's unconditional credence in any other proposition expressible in terms of that partition. First, any contradiction receives credence 0. Then for any other proposition, we figure out which rows of the table it's true on, and calculate its credence by summing the values on those rows. For example, we might be interested in the agent's credence that the die roll comes up even. The proposition that the roll comes up even is true on the second, fourth, and sixth rows of the table. So the agent's credence in that proposition is $1/10 + 1/10 + 1/2 = 7/10$.

We can calculate the agent's credence in this way because

$$E \mathrel{=\!\!\mid\!\models} 2 \vee 4 \vee 6 \tag{2.14}$$

where E is the proposition that the die came up even, "2" represents its coming up two, etc. By Equivalence,

$$\mathrm{cr}(E) = \mathrm{cr}(2 \vee 4 \vee 6) \tag{2.15}$$

Since the propositions on the right are members of a partition, they are mutually exclusive, so Finite Additivity (Extended) yields

$$\mathrm{cr}(E) = \mathrm{cr}(2) + \mathrm{cr}(4) + \mathrm{cr}(6) \tag{2.16}$$

The agent's unconditional credence in E can be found by summing the values on the second, fourth, and sixth rows of the table.

Given a propositional language $\mathcal{L}$, it's often useful to build a probability table using the partition containing $\mathcal{L}$'s state-descriptions. For example, for a language with two atomic propositions P and Q, I might give you the following probability 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

The state-descriptions in this table are fully specified by the Ts and Fs appearing under P and Q in each row, but I've also provided labels $(s_1, s_2, \ldots)$ for each state-description to show how they correspond to regions in Figure 2.4.

Suppose a probabilistic agent has the unconditional credences specified in this table. What credence does she assign to $P \vee Q$? From the Venn diagram we can see that $P \vee Q$ is true on state-descriptions s_1, s_2, and s_3. So we find $\text{cr}(P \vee Q)$ by adding up the cr-values on the first three rows of our table. In this case $\text{cr}(P \vee Q) = 0.6$.

A probability table over state-descriptions is a particularly efficient way of specifying an agent's unconditional credence distribution over an entire propositional language.[13] A language $\mathcal{L}$ closed under the standard connectives contains infinitely many propositions, so a distribution over that language contains infinitely many values. If the agent's credences satisfy the probability axioms, the Equivalence rule tells us that equivalent propositions must all receive the same credence. So we can specify the entire distribution just by specifying its values over a maximal set of non-equivalent propositions in the language.

But that can still be a lot of propositions! If $\mathcal{L}$ has n atomic propositions, it will contain 2^{2^n} non-equivalent propositions (see Exercise 2.3). 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.

On the other hand, a language with n atomics will contain only 2^n state-descriptions. And once we provide unconditional credences for these propositions in our probability table, all the remaining values in the distribution follow. Every contradictory proposition receives credence 0, while each non-contradictory proposition is equivalent to a disjunction of state-descriptions (its disjunctive normal form). By Finite Additivity (Extended), the credence in a disjunction of state-descriptions is just the sum of the credences assigned to those state-descriptions. So the probability table contains all the information we need to specify the full distribution.[14]

2.3.3 Using probability tables

Probability 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.

But probability tables can also be used to prove theorems and solve problems. To do so, we replace the numerical credence values in the table 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 probability 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

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

But a is just cr(P & Q), and b is cr(P & $\sim Q$). This gives us a very quick proof of the Decomposition rule from Section 2.2.1. It's often much easier to prove a general probability result using a probability table built on state-descriptions than it is to prove the same result from Kolmogorov's axioms.

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 probability table provides a purely algebraic solution method. We start by expressing the constraints on my distribution as equations using the variables from the table. From our second constraint on probability tables we have:

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

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

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

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

Finally, we represent the fact that I am certain of $P \vee Q$. The only line of the 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 $a + c + d = 1/2$.

2.3.4 Odds

Agents sometimes report their levels of confidence using odds rather than probabilities. If an agent's unconditional credence in P is $\mathrm{cr}(P)$, her **odds for** P are $\mathrm{cr}(P) : \mathrm{cr}(\sim P)$, and her **odds against** P are $\mathrm{cr}(\sim P) : \mathrm{cr}(P)$.

For example, there are thirty-seven pockets on a European roulette wheel. (American wheels have more.) Eighteen of those pockets are black. Suppose an agent's credences obey the probability axioms, and she assigns equal credence to the roulette ball's landing in any of the thirty-seven pockets. Then her credence that the ball will land in a black pocket is 18/37, and her credence that it won't is 19/37. Her odds for black are therefore

$$18/37 : 19/37, \text{ or } 18 : 19 \tag{2.22}$$

(Since the agent assigns equal credence to each of the pockets, these odds are easily found by comparing the number of pockets that make the proposition true to the number of pockets that make it false.) Yet in gambling contexts we usually report odds *against* a proposition. So in a casino someone might say that the odds against the ball's landing in the single green pocket are "36 to 1". The odds against an event are tightly connected to the stakes at which it would be fair to gamble on that event, which we will discuss in Chapter 7.

Warning

Instead of using a colon or the word "to", people sometimes quote odds as fractions. So someone might say that the odds for the roulette ball's landing in a black pocket are "18/19".[15] It's important not to mistake this fraction for a probability value. If your odds for black are 18 : 19, you take the ball's landing on black to a bit less likely to happen than not. But if your unconditional credence in black were 18/19, you would always bet on black!

It can be useful to think in terms of odds not only for calculating betting stakes but also because odds highlight differences that may be obscured by probability values. Suppose you hold a single ticket in a lottery that you take to be fair. Initially you think that the lottery contains only two tickets, of which yours is one. But then someone tells you there are 100 tickets in the lottery. This is a significant blow to your chances, witnessed by the fact that your assessment of the odds against winning has gone from 1 : 1 to 99 : 1. The significance of this change can also be seen in your unconditional credence that you will lose, which has jumped from 50% to 99%.

But now it turns out that your informant was misled, and there are actually 10,000 tickets in the lottery! This is another significant blow to your chances, intuitively at least as bad as the first jump in size. And indeed, your odds against winning go from 99 : 1 to 9,999 : 1. Yet your credence that you'll lose moves only from 99% to 99.99%. Probabilities work on an additive scale; from that perspective a move from 0.5 to 0.99 looks important while a move from 0.99 to 0.9999 looks like a rounding error. But odds use ratios, which highlight multiplicative effects more obviously.

2.4 What the probability calculus adds

In Chapter 1 we moved from thinking of agents' doxastic attitudes in terms of binary (categorical) beliefs and confidence comparisons to working with numerical degrees of belief. At a first pass, this is a purely *descriptive* maneuver, yielding descriptions of an agent's attitudes at a higher fineness of grain. As we saw in Chapter 1, this added level of descriptive detail 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 numerical credences over a set of propositions introduces a complete ranking, making all the propositions commensurable with respect to the agent's confidences. This may be an unrealistic result.

Chapter 1 also offered a *norm* for comparative confidence rankings:

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.

We have now introduced Kolmogorov's probability axioms as a set of norms on credences. Besides the descriptive changes that happen when we move from comparative confidences to numerical credences, how do the probability axioms go beyond Comparative Entailment? What *more* do we demand of an agent when we require that her credences be probabilistic?

Comparative Entailment can be derived from the probability axioms—we've already seen that by the Entailment rule, if $P \vDash Q$ then rationality requires $\mathrm{cr}(P) \leq \mathrm{cr}(Q)$. But 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 assign 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 comparative 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:	$\mathrm{cr}(\mathsf{F}) = 0$	$\mathrm{cr}(P) = 1/36$	$\mathrm{cr}(\sim P) = 25/36$	$\mathrm{cr}(\mathsf{T}) = 1$

With respect to their confidence comparisons, Mr. Prob and Mr. Weak are identical; they each rank $\sim P$ above P and both those propositions between a tautology and a 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. In 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 gives rise to some of the theory's most interesting and useful results. It does so by demanding that rational credences be *linear*, in the sense that a disjunction's credence is a linear combination[16] of the credences in its mutually exclusive disjuncts.

Of course, the fan of confidence comparisons need not restrict herself to the Comparative Entailment norm. Chapter 14 will explore further comparative constraints that have been proposed, some of which are capable of discriminating between Mr. Prob and Mr. Weak. We will ask whether those non-numerical norms can replicate all the desirable results secured by Finite Additivity for the Bayesian credence regime. This will be an especially pressing question because the impressive Bayesian numerical 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.

2.5 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 non-contradictory propositions X and Y, $X \vDash Y$ if and only if every disjunct in the disjunctive normal form equivalent of X is also a disjunct of the disjunctive normal form equivalent of Y.

(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. Can a probabilistic credence distribution assign $\text{cr}(P) = 0.5$, $\text{cr}(Q) = 0.5$, and $\text{cr}(\sim P \,\&\, \sim Q) = 0.8$? Explain why or why not.[17]

Problem 2.6. Starting with only the probability axioms and Negation, write out proofs for all of the probability rules listed in Section 2.2.1. Your proofs must be straight from the axioms—no using Venn diagrams or probability 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 the one in which they're listed. And I did Finite Additivity (Extended) for you.)

Problem 2.7. Prove that for any propositions P and Q, if $\text{cr}(P \equiv Q) = 1$ then $\text{cr}(P) = \text{cr}(Q)$.

Problem 2.8. In *The Empire Strikes Back*, C-3PO tells Han Solo that the odds against successfully navigating an asteroid field are 3,720 to 1.

(a) What is C-3PO's unconditional credence that they will successfully navigate the asteroid field? (Express your answer as a fraction.)

(b) Suppose C-3PO is certain that they will survive if they either successfully navigate the asteroid field, or fail to successfully navigate it but hide in a cave. He is also certain that those are the only two ways they can survive, and his odds against the conjunction of failing to successfully navigate and hiding in a cave are 59 to 2. Assuming C-3PO's credences obey the probability axioms, what are his odds against their surviving?

(c) In the movie, how does Han respond to 3PO's odds declaration? (Apparently Han prefers his probabilities quoted as percentages.)

Problem 2.9. Consider the probabilistic credence distribution specified by this probability 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 2.10. 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 probability table and demonstrate that the resulting distribution satisfies each of the four constraints. (Hint: Start by building a probability table; then figure out what each of the constraints says about the credence values in the table; then figure out if it's possible to meet all of the constraints at once.)

Problem 2.11. Tversky and Kahneman's finding that ordinary subjects commit the Conjunction Fallacy has held up to a great deal of experimental replication. 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 2.12. Recall Mr. Prob and Mr. Weak from Section 2.4. 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 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.

2.6 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 and Chance: An Introduction to Inductive Logic*. 4th edition. 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 counterintuitive examples like the Conjunction Fallacy from our Section 2.2.4.

CLASSIC TEXTS

A. N. Kolmogorov (1933/1950). *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 (2016). Probability and Non-Classical Logic. In: *Oxford Handbook of Probability and Philosophy*. Ed. by Alan Hájek and Christopher R. Hitchcock. Oxford: 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 probability tables (which Fitelson calls "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. Other authors describe degrees of belief as assigned to sentences, statements, or sets of events. Also, propositions are sometimes taken to be identical to one of these alternatives. As mentioned in Chapter 1, I will not assume much about what propositions are, except that: they are capable of having truth-values (that is, capable of being true or false); they are expressible by declarative sentences; and they 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 (countable) union, (countable) 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, intersection, and complementation because the propositions in $\mathcal{L}$ are closed under disjunction, conjunction, and negation (respectively).
3. I'm also going to be fairly cavalier about corner-quotes, the use-mention distinction, etc.

4. Throughout this book we will be assuming a classical logic, in which each proposition has exactly one of two available truth-values (true/false) and entailment obeys the inference rules taught in standard introductory logic classes. For information about probability in non-classical logics, see the Further Reading at the end of this chapter.
5. The cognoscenti will note that in order for the state-descriptions of $\mathcal{L}$ to form a partition, the atomic propositions of $\mathcal{L}$ must be (logically) independent. We will assume throughout this book that every propositional language employed contains logically independent atomic propositions, unless explicitly noted otherwise.
6. 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 repetition.
7. In the statistics community, probability distributions are often assigned over the possible values of sets of random variables. Propositions are then thought of as dichotomous random variables capable of taking only the values 1 and 0 (for "true" and "false", respectively). Only rarely in this book will we look past distributions over propositions to distributions over more general random variables.
8. The axioms I've presented are not precisely identical to Kolmogorov's, but the differences are insignificant for our purposes. Some authors also include Countable Additivity—which we'll discuss in Chapter 5—among "Kolmogorov's axioms", but I'll use the phrase to pick out only Non-Negativity, Normality, and Finite Additivity.

   Galavotti (2005, pp. 54–5) notes that authors such as Mazurkiewicz (1932) and Popper (1938) also provided axioms for probability around the time Kolmogorov was working. She recommends Roeper and Leblanc (1999) for an extensive survey of the axiomatizations available.
9. This analysis could easily be generalized to any large, finite number of tickets.
10. Philosophers sometimes describe the worlds an agent entertains as her "epistemically possible worlds". Yet that term also carries a connotation of being determined by what the agent *knows*. So I'll discuss doxastically possible worlds, which are determined by what an agent *takes* to be possible rather than what she *knows*.
11. 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 Euclidean two-dimensional space and outputs its area is also a measure. That makes probabilities representable by areas in a rectangle.
12. To avoid the confusion discussed here, some authors use "muddy" Venn diagrams in which all atomic propositions are associated with 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 higher confidence as greater area.
13. Truth-tables famously come to us from Wittgenstein's *Tractatus Logico-Philosophicus* (Wittgenstein 1921/1961), in which Wittgenstein also proposed a theory of probability

assigning equal value to each state-description. But to my knowledge the first person to characterize probability distributions in *general* by the values they assign to state-descriptions was Carnap, as in his (1945, Sect. 3).

14. 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 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 is 0. We can prove that any credence distribution stipulated in this fashion will satisfy Kolmogorov's three probability axioms. I'll leave the (somewhat challenging) proof as an exercise for the reader.
15. Odds against a proposition, quoted with a slash like a fraction, are known as "British odds", for their popularity among British and Irish bookies.
16. Given two variables x and y and two constants a and b, we call $z = ax + by$ a **linear combination** of x and y. Finite Additivity makes $\text{cr}(X \vee Y)$ a linear combination of $\text{cr}(X)$ and $\text{cr}(Y)$, with the constants a and b each set to 1.
17. I owe this problem to Julia Staffel.