Notes on Probability Theory

David Kinney
Foundations and Applications of Humanities Analytics

1 Introduction

Probability is one of the most important mathematical concepts for humanities and cultural analytics. The goal of these notes is to provide a reference on probability theory that complements the lecture material. Our aim is to present probability theory in a way that is both *rigorous* (i.e., it leaves as little as possible imprecise and cuts as few corners as it can), and *accessible* (i.e., it does not assume prior knowledge of any mathematics beyond arithmetic).

Probability theory is a mathematical language that allows us to speak with precision about how likely it is that a given process results in a given outcome. The term 'process' can be understood very broadly. Flipping a coin is a process, as is flipping 1,000 coins, as is measuring the temperature of a glass of water, as is writing a novel. For our purposes here, we'll say that a process is anything that begins with some initial conditions and ends with an outcome, where we as the inquirers can divide things into initial conditions and outcomes however we want. So in the examples above, the initial conditions might be broadly described as:

1. The material composition of a coin and the force with which it is flipped.
2. The material composition of 1,000 coins and the force with which they are flipped.
3. Some description of the room and container in which a gas of water are stored, and the properties of the thermometer used to measure its temperature.
4. The psychological state of an author and the cultural context in which they live when they begin writing a novel.

The corresponding outcomes of interest might be:

1. Whether the coin lands heads or tails.
2. The sequence of heads/tails outcomes of the 1,000 coins.
3. The reading of a thermometer placed in the glass of water.
4. The contents of the novel that the author eventually publishes.

These examples are not exhaustive; probability theory has a myriad of applications. But they should give you an idea of the sorts of uses that probability theory can and does have, including in the context of quantitative humanities research. In what follows, we'll provide the necessary background to understand and develop some useful applications of probability theory.

2 Set Theory

The language of probability theory is actually a special application of the more general language that a lot of mathematical theories are written in: the language of set theory. One can spend their whole life studying set theory. Thankfully, we won't. Instead, we'll introduce only the minimal set-theoretic language needed to do some useful probability theory. Specifically, we'll define:

- The concept of a *set*.
- The *subset* relation between sets.
- The *power set* of a given set.

- The set-theoretic operations of *union* and *intersection*.
- A *function* between two sets.

Intuition can be a good guide to understanding some of these concepts. In the case of *subset*, *union*, and *intersection*, the terms mean more or less what you might think they mean from an understanding of ordinary English. Nevertheless, for the sake of rigor, we'll define each more precisely.

2.1 Sets

We'll begin by defining a set very generally.

Definition 1. A **set** is a collection of elements.

On its own, this doesn't tell us much, because it raises an obvious question: what are elements? The simple answer is that elements are whatever we want them to be. They might be numbers, they might be letters or words, they might be shapes or symbols, and so on. For the sake of doing probability theory, we'll often want to think of the elements of a set as symbols that denote possible outcomes of processes. How exactly symbols denote possible events in the world is a question that has kept philosophers of language from Bertrand Russell to Jacques Derrida very busy, but let's ignore that for now and just accept that sets are collections of elements, where those elements can be whatever we want them to be. This flexibility is a virtue of set theory; because the elements of sets can be anything, we can use set theory to talk about a wide variety of subjects.

Let us now introduce some standard set-theoretic notation, i.e., some symbols that allow us write efficiently in the language of set theory. This notation is purely conventional; there's nothing about set theory that *requires* us to use this particular notation. But these symbols are so common that it's worth being familiar with them. We'll sometimes use italicized capital letters, like S, to refer to sets. We'll also typically list the elements of a set in the brackets $\{ \quad \}$ to indicate that those elements are all contained within a single set. So for instance, if the set S contains all and only the numbers 1, 2, 3, and 4, we can write this as $S = \{1, 2, 3, 4\}$. We might also use Greek letters, like Ω or Σ, to refer to sets. So a set Ω containing all and only the letters h and t, which might stand for a heads or tails outcome of a coin toss, can be written as $\Omega = \{h, t\}$. Finally, we might want to remain very agnostic about the nature of the elements of a set. In that case, we'll use a numbered, lower-case letter to refer to the elements of a set. So we might have the set $S = \{s_1, s_2, \ldots, s_n\}$. Each element s_i, where i is any number from 1 to n, denotes some element of the set, where that element can be anything at all. If we switch to Greek letters, we might do the same thing using the notation $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$. Note that the choice to use Roman or Greek letters is also just a matter of convention; in different contexts it is typical to use one or the other, but there's no real rhyme or reason behind it other than the cultural quirks of a particular application of set theory. We denote claims about set-membership using the symbols $\in$ and $\notin$. We read $s_i \in S$ as 's_i is an element of S', and $s_i \notin S$ as 's_i is not an element of S'.

The ordering of elements in a set does not matter. To illustrate, the set $\{a, b, c\}$ is identical to the set $\{c, b, a\}$. This would be just as true if these sets contained numbers, shapes, or anything else instead of letters. Similarly, an element is only "counted" as a belonging to a set once. There's no distinction, in basic set theory, between the set $\{3, 4, 4\}$ and the set $\{3, 4\}$. This would be just as true if we replaced 3 and 4 with letters, shapes, words, etc.

Importantly, *sets can be elements of sets.* For example, the set $S = \{\{1, 2\}, \{3, 4\}\}$ has as its elements the sets $\{1, 2\}$ and $\{3, 4\}$. Crucially, this does *not* mean that 1, 2, 3, or 4 are elements of S. Indeed, they're not. Rather, *only* the sets $\{1, 2\}$ and $\{3, 4\}$ are elements of S. The numbers 1 and 2 are elements of $\{1, 2\}$, but not S, and 3 and 4 are elements of $\{3, 4\}$, but not S. More generally, we say that *set membership is not transitive.* That is, if $s_i \in S$ and $S \in \Sigma$, *it does not follow* that $s_i \in \Sigma$. This is an admittedly counter-intuitive part of set theory, so it may be worth reading this paragraph a few times, and then trying the following exercise:

Exercise 1. Let C be a set containing all the clubs in a standard deck of cards, and let D be a set containing all the diamonds in a standard deck of cards. Let S be a set defined so that $S = \{C, D\}$. Now answer the following questions (correct answers in footnote):

1. Is C an element of S?

2. Is D an element of S?
3. Is the king of clubs an element of C?
4. Is the six of diamonds an element of C?
5. Is the seven of diamonds an element of D?
6. Is the three of clubs an element of D?
7. Is the king of clubs an element of S?
8. Is the king of diamonds an element of S?[1]

We recommend not moving on until you understand how to answer these questions correctly. You'll need this understanding to grasp the concept of a *power set*, which in turn is crucial to doing some useful probability theory.

Before moving on, we'll introduce one final piece of notation. The symbol $\emptyset$ denotes the 'empty set', or the set with no elements. It may seem bizarre that we'll even need to talk about this set, but just as the number zero plays an important role in mathematics, it turns out that the empty set plays a crucial role in many applications of set theory, including probability theory.

2.2 The Subset Relation

We'll often want to say that one element is a subset of another, and indeed we'll need to in order to speak the language of probability theory. We can define the subset relation between two sets formally as follows:

Definition 2. A set A is a **subset** of a set B if and only if every element of A is an element of B.[2]

In symbols, we write $A \subseteq B$ to denote that A is a subset of B. To illustrate by way of some simple examples, the set of English nouns is a subset of the set of all English words, the set of odd integers is a subset of the set of all integers, and the set of oranges is a subset of a set of all fruits.

This is all we'll really need to say about the subset relation here, but we'll close with two important remarks.

Remark 1. Any set is a subset of itself. To see this, note that for any set A it is trivially the case that any element of A is an element of A, and so A is a subset of A.

Remark 2. The empty set $\emptyset$ is a subset of any set. To see this, note that for any set A it is trivially the case that any element of $\emptyset$ is an element of A, just because $\emptyset$ has no elements in the first place. Thus, $\emptyset$ is a subset of A.

2.3 Power Sets

As promised, we come now to the all-important notion of a power set. Equipped with the concepts we've already defined, we can define a power set very simply.

Definition 3. For any set S, the power set of S is the set of all subsets of S.

We'll use the somewhat unusual symbol $\wp$ to denote the power set. That is, we can read $\wp(S)$ as 'the power set of the set S'. To illustrate, if $S = \{1, 2, 3, 4\}$, then

$$\wp(S) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}, \{1,2,3\}, \{1,2,4\}, \{1,3,4\}, \{2,3,4\}, \{1,2,3,4\}\}.$$

Notice that $S = \{1, 2, 3, 4\}$ is itself an element of $\wp(S)$. More generally, we can make the following remark:

[1] Answers: 1) Yes, 2) Yes, 3) Yes, 4) No, 5) Yes, 6) No, 7) No, 8) No.

[2] Often in defining relations between sets, we'll say that A stands in that relation to B if and only if some other condition holds. This just means that if A stands in the relation B, then the condition holds, and that if the condition holds, then A stands in the relation to B. For instance, we might say that Alice is older than Bob if and only if Alice was born before Bob. This just means that if Alice is older than Bob then Alice was born before Bob, and if Alice was born before Bob, then Alice is older than Bob.

Remark 3. Any set is an element of its own power set. This follows from the fact that the power set of any set A is the set of all subsets of A, and the fact, demonstrated in Remark 1 above, that any set is a subset of itself.

Notice further that while 1, 2, 3, and 4 are all elements of S, *none* of them are elements of $\wp(S)$. Sets *containing* these integers are elements of $\wp(S)$, but this is not the same thing as those integers being elements of S, because set membership is not transitive. This would be just as true if S did not contain integers but instead contained letters, shapes, words, or anything else. If this point does not make sense (and indeed, this part of set theory has the potential to be confusing), then we suggest returning to Exercise 1. Once you have mastered that, the idea that any element of S need not be an element of the power set of S should be clear.

Notice as well that the empty set $\emptyset$ is an element of $\wp(S)$. Here too, we can make a more general remark:

Remark 4. The empty set is an element of any power set. This follows from the fact that the power set of any set A is the set of all subsets of A, and the fact, demonstrated in Remark 2, that the empty set is a subset of any set.

2.4 Union and Intersection

For the sake of doing probability theory, we'll often be interested in defining a set that contains either all the elements of both A and B, or all the elements shared by A and B, where A and B are themselves sets. This can be made more precise by defining the concept of the *union* and *intersection* of two sets.

We begin by defining the union of any two sets:

Definition 4. The union of any two sets A and B is the set containing all and only those elements that are in either A or B.

Using symbols, we write $A \cup B$ for the union of A and B. To illustrate, the union of the sets of words $A = \{\text{democracy}, \text{capitalism}, \text{oligarchy}\}$ and $B = \{\text{democracy}, \text{religion}, \text{materialism}\}$ is:

$$A \cup B = \{\text{democracy}, \text{capitalism}, \text{oligarchy}, \text{religion}, \text{materialism}\}.$$

Note that for any sets A or B, $A \subseteq A \cup B$ and $B \subseteq A \cup B$. This is because any element of a set A is also an element of the union of B with any other set.

Next, we define the intersection of any two sets:

Definition 5. The intersection of any two sets A and B is the set containing all and only those elements that are in both A and B.

Using symbols, we write $A \cap B$ for the intersection of A and B. To illustrate, the intersection of the sets of words $A = \{\text{democracy}, \text{capitalism}, \text{oligarchy}\}$ and $B = \{\text{democracy}, \text{religion}, \text{materialism}\}$ is $A \cap B = \{\text{democracy}\}$. Note that for any sets A or B, $A \cap B \subseteq A$ and $A \cap B \subseteq B$. This is because any element of the intersection $A \cap B$ is also an element of both A and B. If A and B share no elements, then the only set containing all and only those elements in both A and B is the set containing no elements, i.e., the empty set. In this case, we write $A \cap B = \emptyset$.

2.5 Functions

A function between two sets is a mathematical object that takes the elements from one set and matches them to elements of a second set. This can be defined more formally as follows:

Definition 6. A function f from set A to set B is a relation that associates each $a_i \in A$ with an element $f(a_i)$ of B.

More rigorous definitions of a function are possible, but this will suffice for our purposes. Symbolically, we use lower-case letters to represent functions, and write $f : A \to B$ to denote that f is a function from A to B. Importantly, if f is a function from A to B, then $f(a_i)$ must be an element of B for every $a_i \in A$. However, it is not the case that every element of B must have an associated element of A that is mapped to it. To illustrate, if we let $A = \{x, y, z\}$ and $B = \{\alpha, \beta, \gamma\}$, then we can define a function $f : A \to B$ such that $f(x) = \alpha$, $f(y) = \alpha$, and $f(z) = \gamma$. Alternatively, we might define a function $g : A \to B$ such that $g(x) = \alpha$, $g(y) = \beta$, and $g(z) = \gamma$. This shows that functions are not generally symmetric; a function from A to B does *not* necessarily define a function from B to A. You may recall from other coursework in mathematics that we sometimes summarize functions using equations. For instance, we can define a function f from the integers to the integers such that each integer x is mapped to its square. This is summarized by the equation $f(x) = x^2$. However, in what follows, we'll define *probability* functions, which usually cannot be summarized in this way.

3 Probability Theory

Now for the fun part. Recall from the introduction that probability theory is mathematical language that allows us to speak precisely about the likelihood of any given outcome of some process. In what follows, we'll introduce the crucial notion of a *probability space*, and then move on to the concept of conditional probability.

3.1 The Probability Space

By the send of this subsection, we'll define the all-important concept of a probability space. To get there, we'll have to define some other notions first.

Let Ω be a set containing all the possible outcomes of some process. We'll assume for now that Ω has finitely many elements (i.e., that there is some integer n that is equal to the number of elements in Ω). While there are many instances in which one might want to consider processes that have an infinite set of possible outcomes, doing so makes probability theory much more complicated, so we'll leave such cases aside for now. To illustrate, if the process we are modeling is a single roll of a die, then Ω might be the set $\{1, 2, 3, 4, 5, 6\}$, where each number denotes the side of the die that shows after the roll.

Next, consider the power set $\wp(\Omega)$. This is the set of all subsets of all possible outcomes of the process being modelled. In the case of the set Ω representing possible outcomes $\{1, 2, 3, 4, 5, 6\}$ of the die roll, the power set $\wp(\Omega)$ is:

$$\begin{aligned}\wp(\Omega) = \{&\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{2,3\}, \{2,4\}, \{2,5\},\\ &\{2,6\}, \{3,4\}, \{3,5\}, \{3,6\}, \{4,5\}, \{4,6\}, \{5,6\}, \{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,2,6\}, \{1,3,4\},\\ &\{1,3,5\}, \{1,3,6\}, \{1,4,5\}, \{1,4,6\}, \{1,5,6\}, \{2,3,4\}, \{2,3,5\}, \{2,3,6\}, \{2,4,5\}, \{2,4,6\},\\ &\{2,5,6\}, \{3,4,5\}, \{3,4,6\}, \{3,5,6\}, \{4,5,6\}, \{1,2,3,4\}, \{1,2,3,5\}, \{1,2,3,6\}, \{1,2,4,5\},\\ &\{1,2,4,6\}, \{1,2,5,6\}, \{1,3,4,5\}, \{1,3,4,6\}, \{1,3,5,6\}, \{1,4,5,6\}, \{2,3,4,5\}, \{2,3,4,6\},\\ &\{2,3,5,6\}, \{3,4,5,6\}, \{1,2,3,4,5\}, \{1,2,3,4,6\}, \{1,2,3,5,6\}, \{1,2,4,5,6\}, \{1,3,4,5,6\},\\ &\{2,3,4,5,6\}, \{1,2,3,4,5,6\}\}.\end{aligned}$$

Don't worry about reading each element of this power set; we write it out in full here just to give you a sense of what its elements are.

Taking stock, we now have a set Ω whose elements are the possible outcomes of a process, and a power set $\wp(\Omega)$ whose elements are sets of possible outcomes of that process. This puts us in a position to define a probability function:

Definition 7. A **probability function** $P : \wp(\Omega) \to [0, 1]$ is a function from the set of sets of possible outcomes $\wp(\Omega)$ into the set of all real numbers between $[0, 1]$, where p has the following properties:

1. $P(\Omega) = 1$.
2. $P(\emptyset) = 0$.
3. For any $A \in \wp(\Omega)$ and $B \in \wp(\Omega)$ such that $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

This is a somewhat more involved definition than we've had so far, so we'll break it down piece-by-piece, with examples. Hopefully, it will be satisfying to do so, because it will bring together most of the concepts we've defined so far.

First, there is the definition of a probability function p as a function from $\wp(\Omega)$ into $[0, 1]$. To get a sense for what this means, consider first any S that is a subset of the set of possible outcomes

Ω. We know that S is an element of $\wp(\Omega)$, since $\wp(\Omega)$ is the set of *all* subsets of Ω. So we thereby know that the probability function p assigns S a number $P(S) = x$, where x is between 0 and 1, inclusive. This number is then *interpreted by us* as the likelihood that the actual outcome of the process will be in the subset S. This interpretive move is crucial, so we'll give it a name:

Definition 8. For any process with the set of possible outcomes Ω, any probability function $P : \wp(\Omega) \to [0, 1]$, any set of possible outcomes $S \in \wp(\Omega)$, and any x between 0 and 1 inclusive, the **standard probabilistic interpretation** of the formalism $P(S) = x$ is to read $P(S) = x$ as saying 'the likelihood that the actual outcome of the process is in S is x'.

It is important to remember that that nothing in any of the set theory we've defined so far *requires* us to interpret probabilities in this way. Rather, the standard probabilistic interpretation is an imputation of meaning onto the set theory that *we*, as inquirers, create. Arguably, the set theory that we've defined here *becomes* probability theory once we impose this meaning upon it. What's truly incredible is that the results of imputing this meaning onto the set-theoretic language have proven so useful in such a wide range of domains.

To illustrate, if Ω is the set of possible outcomes of die rolls, then the probability function p assigns a number between 0 and 1 to any subset of the set of all possible outcomes of the die roll. We then interpret these numbers as the likelihood of the particular die roll in question resulting in one of the outcomes in the set in question. So, for instance, if $P(\{1, 2, 3\}) = .5$, then we read this as saying that the likelihood that the die roll results in the outcome 1, 2, or 3 is .5, or 50%.

Next, consider the properties of a probability function listed in Definition 7. Recall from Remark 3 that any set is an element of its power set. That is, for any set of possible outcomes Ω, the full set of possibilities Ω is in $\wp(\Omega)$. Since the probability function p assigns a probability to all elements of $\wp(\Omega)$, it assigns a probability to Ω. The first property of a probability function is that $P(\Omega) = 1$. Under the standard probabilistic interpretation, this means that the likelihood of the process resulting in any of its possible outcomes is one, which is the maximum allowable likelihood. In other words, we are committed to the idea that *something* has to result from our process, and it has to be one of the things in our set of possibilities Ω. In the case of the roll of a die, this amounts to a commitment to the idea that the outcome of the die roll will be either 1, 2, 3, 4, 5, or 6.

As for the second property of a probability function, recall from Remark 4 that the empty set is a member of any power set. Thus, for any set of possible outcomes Ω, the empty set $\emptyset$ is in $\wp(\Omega)$. Since the probability function p assigns a probability to all elements of $\wp(\Omega)$, it assigns a probability to $\emptyset$. The second property of a probability function is that $P(\emptyset) = 0$. Under the standard probabilistic interpretation, this means that the likelihood of the process resulting in none of its possible outcomes is zero, which is the minimum allowable likelihood. This is the flipside of our commitment, discussed in the previous paragraph, to the idea that *something* has to result from our process.

Finally, consider the third property of a probability function, which is that if two sets of possible outcomes A and B share no common elements (so that $A \cap B = \emptyset$), then the probability of an outcome in either A or B (i.e., the probability of the union $A \cup B$) must be the sum of the probability of an outcome in A or an outcome in B. So in the die roll example, if the probability of the outcome being in the set $\{1, 2\}$ is $\frac{1}{3}$, and the probability of the outcome being in the set $\{3, 4\}$ is also $\frac{1}{3}$, then the probability of the outcome being in the set $\{1, 2, 3, 4\}$ is $\frac{2}{3}$. In fact, if we know the values of the probabilities $P(\{1\})$, $P(\{2\})$, $P(\{3\})$, $P(\{4\})$, $P(\{5\})$, and $P(\{6\})$, then the third property of a probability function allows us to calculate the probability of any other set of possible outcomes for the die roll. To test your understanding, assign each of the probabilities listed above the value $\frac{1}{6}$, and then use the third property on a probability function to calculate the value of any other set of possible outcomes of a die roll.

We can put all this together in a concise statement of what it means to model a process probabilistically, by defining a probability space as follows:

Definition 9. A **probability space** is comprised of three objects: a set of outcomes Ω, the power set $\wp(\Omega)$, and the probability function $P : \wp(\Omega) \to [0, 1]$.

Equipped with the standard interpretation of the values of a probability function, it should now be clear how we can assign likelihoods to the possible outcomes of processes using a probability space. It is crucial to note, however, that the probabilistic formalism on its own *does not tell us the exact probabilities to assign to all sets of outcomes.* It only provides some constraints on how we assign said probabilities. For instance, nothing in probability theory itself says that we should assign all outcomes of a die roll equal probability; this is an assumption we'll have to justify by other means, if we can justify it at all. It may be that, for some reason, a die is weighted so that even outcomes are more likely than odd outcomes. If this were true, it might change the probability function that we'd want to use in a model of the die roll process. All that probability theory itself tells us is that, whatever probabilities we do assign to sets of possible outcomes, the probability function must satisfy the three properties listed in Definition 7.

3.1.1 A Caveat

At this stage, we have to make a confession. We have not told you the whole truth here. So now we will. In many contexts, one *can* use a probability space in which probabilities are not defined on the full power set of outcomes. In fact, when the set of possible outcomes is infinite, we may be *required* to define our probability space differently. So, if you choose to go on and study more advanced aspects of probability theory, then you'll have to prepare to deviate somewhat from what we've presented above. At the same time, *a lot* of useful applications of probability theory can be done using just what we've presented so far, including everything that you will encounter in this course. Specifically, as long as the process that you're studying has finitely many possible outcomes (even if there's ten trillion of them), then you'll be compliant with all mathematical rules if you just stick to the techniques we've presented here.

3.2 Conditional Probability

Often, we'll want to change the probability that we assign to a given set of outcomes of a process once we learn some information about the actual outcome of that process. Returning to our example of a die roll, we might initially believe that the likelihood of the die roll resulting in an outcome where the die shows a 1 is $\frac{1}{6}$. However, if we learn that the outcome of the die roll was such that the die showed an odd number, then we may wish to revise this belief, and instead say that the likelihood of the die showing a 1 is $\frac{1}{3}$ (since we now now that the actual outcome must be either 1, 3, or 5).

It turns out that the language of probability theory offers a very precise way of talking about this practice of changing or updating our beliefs. Suppose that we have a probability space consisting of a set of possible outcomes Ω, its power set $\wp(\Omega)$, and a probability function $P : \wp(\Omega) \to [0, 1]$. Let A and B be any two elements of the power set $\wp(\Omega)$, i.e., any two subsets of Ω. The **conditional probability** $P(A|B)$ can be read as 'the probability that the outcome of the process is an element of A, given that it is an element of B'. For instance, in the example given above, the claim 'the probability that the outcome of the die roll is in $\{1\}$, given that it is in the set odd outcomes $\{1, 3, 5\}$, is $\frac{1}{3}$' can be written as $P(\{1\}|\{1, 3, 5\}) = \frac{1}{3}$.

In fact, if we already have a well-defined probability function over all the elements of a power set $\wp(\Omega)$, then we can calculate the value of any conditional probability. Letting A and B be any elements of $\wp(\Omega)$, the value of the conditional probability $P(A|B)$ can be calculated using the formula

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

That is, the value of the conditional probability $P(A|B)$ is the ratio between the probability assigned to the intersection $A \cap B$ and the probability assigned to the set of possible outcomes B. Thus, the equation above is often referred to as the "ratio formula" for prior probability.

To illustrate using the example above, if we assume that all outcomes of a die roll are equally likely, then the probability of an outcome in the set $\{1\} \cap \{1, 3, 5\} = \{1\}$ is $\frac{1}{6}$, whereas the probability that the outcome of the die roll is in the set $\{1, 3, 5\}$ is $\frac{3}{6}$, or .5. Using the ratio formula, we can calculate the value of the conditional probability $P(\{1\}|\{1, 3, 5\})$ as follows:

$$P(\{1\}|\{1, 3, 5\}) = \frac{P(\{1\} \cap \{1, 3, 5\})}{P(\{1, 3, 5\})} = \frac{P(\{1\})}{P(\{1, 3, 5\})} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

Thus, the probability of the outcome of the die roll being 1, given that we know the outcome will be an odd number, is $\frac{1}{3}$. Note that, for any set of possible outcomes B, when $P(B) = 0$, the conditional probability $P(A|B)$ is undefined, since division is not defined when the denominator is zero. In more advanced applications of probability theory, one can explore alternative ways of defining conditional probability which avoid this issue, but for our purposes here, we don't need to worry about this, since we won't assign probability zero to any events which might occur.

4 Application in Humanities Analytics

In this final section, we'll show how one can use formal probability theory to set up a context of inquiry for a toy example of project in humanities analytics. The example we use will be *very* simplistic, so as to make it very clear how these concepts apply. Indeed, it may not be the case that an actual project in humanities analytics or digital humanities would ever be quite this simple.

Suppose that we're interested in the use of floral imagery and animal imagery in nineteenth-century British novels. To study this probabilistically, we'll think of a generic nineteenth-century British novelist as a *process* that can result in one of four outcomes:

- A novel with floral imagery and animal imagery.
- A novel with floral imagery but not animal imagery.
- A novel with animal imagery but not floral imagery.
- A novel with neither floral imagery nor animal imagery.

Let Ω be the set containing these four outcomes, and let $\wp(\Omega)$ be the power set of Ω, containing each of its subsets. Our goal is to define a probability function p over Ω. This, it must be noted, is the hard part. The simplest way to do it would be to go through every nineteenth-century British novel and count whether it contains either variety of imagery. This would enable us to calculate probabilities for the sets containing one and only one element of Ω using the following equations:

$$P(\{\text{Novel with floral and animal imagery}\}) = \frac{\#\text{ of novels with floral and animal imagery}}{\text{Total }\#\text{ of novels}}$$

$$P(\{\text{Novel with floral but not animal imagery}\}) = \frac{\#\text{ of novels with floral but not animal imagery}}{\text{Total }\#\text{ of novels}}$$

$$P(\{\text{Novel with animal but not floral imagery}\}) = \frac{\#\text{ of novels with animal but not floral imagery}}{\text{Total }\#\text{ of novels}}$$

$$P(\{\text{Novel with neither animal nor floral imagery}\}) = \frac{\#\text{ of novels with neither animal nor floral imagery}}{\text{Total }\#\text{ of novels}}$$

As a practical matter, it will be impossible to calculate these four ratios with perfect historical accuracy. But through a careful mix of archival knowledge, text-processing technology, and (most importantly) expertise in literary study and history, it may be possible to arrive at decent estimates.

Once we have probabilities assigned to these four sets, each of which contains one and only one element of Ω, we note that all other elements of the power set $\wp(\Omega)$ can be formed by taking the union of some combination of the four sets assigned probabilities above. So, via the third property of a probability function, we can use the four probabilities calculated above, along with simple addition, to calculate probabilities for every other element of the power set $\wp(\Omega)$.

This allows us, in turn, to calculate conditional probabilities. For instance, suppose that we wanted to know the probability the a novel contains animal imagery, given that it contains floral imagery. Let A be the set of outcomes in which the novel produced has animal imagery, which is defined as follows:

$$A = \{\text{Novel with floral and animal imagery}, \text{Novel with animal but not floral imagery}\}.$$

Let F be the set of outcomes in which the novel produced has floral imagery, which is defined as follows:

$$F = \{\text{Novel with floral and animal imagery}, \text{Novel with floral but not animal imagery}\}.$$

Note that the intersection $A \cap F$ is defined as follows:

$$A \cap F = \{\text{Novel with floral and animal imagery}\}.$$

If, using the techniques above, we are able to calculate that $P(F) = .7$ and $P(A \cap F) = .2$, then we can calculate $P(A|F)$ using the ratio formula as follows:

$$P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{.2}{.7} \approx .286$$

This probability allows us to quantify, to some extent, the relationship between animal and floral imagery in our domain of interest (i.e., nineteenth-century British novels).

Beyond all of the statistical guesses and fudges that may be necessary in order to estimate the probabilities used in this calculation, there's a deeper idealization at work here as well. Namely, to get this whole context of inquiry started, we had to think of a process whereby a generic nineteenth-century British novelist produces a novel. Clearly, there is no such thing as a generic nineteenth-century British novelist. There is also no such thing as a generic author of a Brazilian women's slave narrative, a generic immigrant poet, a generic Turkish diarist, etc. However, this kind of idealization, in which a process that is always realized in a unique way at the local level is studied in a more generic form at the statistical level, is at the heart of much quantitative research in the sciences as well as the humanities. For example, no two COVID-19 infections are identical, and yet epidemiological models must nevertheless distinguish between the infected and the uninfected while ignoring certain nuances of each infection. There's an unavoidable trade-off between our ability to make informative statistical generalizations and the kind of understanding that comes from detailed study of a single case. This is why we view statistical approached in the humanities as complementing, rather than replacing, more traditional humanities scholarship that provides the kind of irreplaceable understanding of specific bodies of work that cannot quite be captured statistically.

5 Conclusion

We conclude by noting that, when one reads a quantitative journal article in any field, one frequently sees probabilistic language, but rarely sees probabilities presented in the precise language of probability theory. Specifically, authors rarely present the underlying probability space in which probabilities are defined. Sometimes, this is because such explication would be pedantic, and the underlying probability space can be inferred from context by someone familiar with the article's background theory. However, it is also the case that sometimes scientists and other quantitative researchers play fast and loose with probabilistic language without giving careful thought to the nature of the underlying probability space that they are using to model a given process. This can occur in even the most "rigorous" of scientific contexts. So we conclude with a plea to you, the reader, to produce scholarship in humanities analytics and beyond that can set an example of good probabilistic hygiene for the entire scientific community to emulate.