6 The Basic Rules of Probability

This chapter summarizes the rules you have been using for adding and multiplying probabilities, and for using conditional probability. It also gives a pictorial way to understand the rules.

The rules that follow are informal versions of standard axioms for elementary probability theory.

ASSUMPTIONS

The rules stated here take some things for granted:

- The rules are for finite groups of propositions (or events).
- If A and B are propositions (or events), then so are AvB, A&B, and ~A.
- Elementary deductive logic (or elementary set theory) is taken for granted.
- If A and B are *logically equivalent*, then Pr(A) = Pr(B). [Or, in set theory, if A and B are events which are provably the same sets of events, Pr(A) = Pr(B).]

NORMALITY

The probability of any proposition or event A lies between 0 and 1.

(1) $0 \leq \Pr(A) \leq 1$

Why the name "normality"? A measure is said to be *normalized* if it is put on a scale between 0 and 1.

CERTAINTY

An event that is sure to happen has probability 1. A proposition that is certainly true has probability 1.

(2) Pr(certain proposition) = 1
Pr(sure event) = 1

Often the Greek letter Ω is used to represent certainty: $\Pr(\Omega) = 1$.

ADDITIVITY

If two events or propositions A and B are mutually exclusive (disjoint, incompatible), the probability that one or the other happens (or is true) is the sum of their probabilities.

(3) If A and B are mutually exclusive, then
$\Pr(A \vee B) = \Pr(A) + \Pr(B)$.

OVERLAP

When A and B are not mutually exclusive, we have to subtract the probability of their overlap. In a moment we will *deduce* this from rules (1)–(3).

(4) $\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \& B)$

CONDITIONAL PROBABILITY

The only basic rules are (1)–(3). Now comes a *definition*.

(5) If $\Pr(B) > 0$, then $\Pr(A/B) = \dfrac{\Pr(A \& B)}{\Pr(B)}$

MULTIPLICATION

The definition of conditional probability implies that:

(6) If $\Pr(B) > 0$, $\Pr(A \& B) = \Pr(A/B)\Pr(B)$.

TOTAL PROBABILITY

Another consequence of the definition of conditional probability:

(7) If $0 < \Pr(B) < 1$, $\Pr(A) = \Pr(B)\Pr(A/B) + \Pr(\sim B)\Pr(A/\sim B)$.

In practice this is a very useful rule. What is the probability that you will get a grade of A in this course? Maybe there are just two possibilities: you study hard, or you do not study hard. Then:

Pr(A) = Pr(study hard)Pr(A/study hard) + Pr(don't study)Pr(A/don't study).

Try putting in some numbers that describe yourself.

LOGICAL CONSEQUENCE

When B logically entails A, then

$$Pr(B) \leq Pr(A).$$

This is because, when B entails A, B is logically equivalent to A&B. Since

$$Pr(A) = Pr(A\&B) + Pr(A\&{\sim}B) = Pr(B) + Pr(A\&{\sim}B),$$

$Pr(A)$ will be bigger than $Pr(B)$ except when $Pr(A\&{\sim}B) = 0$.

STATISTICAL INDEPENDENCE

Thus far we have been very informal when talking about independence. Now we state a *definition* of one concept, often called statistical independence.

(8) If $0 < Pr(A)$ and $0 < Pr(B)$, then,
A and B are statistically independent if and only if:
$Pr(A/B) = Pr(A)$.

PROOF OF THE RULE FOR OVERLAP

(4) $Pr(AvB) = Pr(A) + Pr(B) - Pr(A\&B)$.

This rule follows from rules (1)–(3), and the logical assumption on page 58, that logically equivalent propositions have the same probability.

AvB is logically equivalent to: $(A\&B)\ v\ (A\&{\sim}B)\ v\ ({\sim}A\&B)$ (*)

Why? Those familiar with "truth tables" can check it out. But you can see it directly. A is logically equivalent to $(A\&B)\ v\ (A\&{\sim}B)$. B is logically equivalent to $(A\&B)\ v\ ({\sim}A\&B)$.

Now the three components $(A\&B)$, $(A\&{\sim}B)$, and $({\sim}A\&B)$ are mutually exclusive. (Why?) Hence we can add their probabilities, using (*).

$Pr(AvB) = Pr(A\&B) + Pr(A\&{\sim}B) + Pr({\sim}A\&B)$ (**)
A is logically equivalent to $[(A\&B)v(A\&{\sim}B)]$, and
B is logically equivalent to $[(A\&B)v({\sim}A\&B)]$.

So,

$$Pr(A) = Pr(A\&B) + Pr(A\&{\sim}B).$$
$$Pr(B) - Pr(A\&B) + Pr({\sim}A\&B).$$

Since it makes no difference to add and then subtract something in (**):

$$Pr(AvB) = Pr(A\&B) + Pr(A\&{\sim}B) + Pr({\sim}A\&B) + Pr(A\&B) - Pr(A\&B)$$

Hence,

$$Pr(AvB) = Pr(A) + Pr(B) - Pr(A\&B).$$

CONDITIONALIZING THE RULES

It is easy to check that the basic rules (1)–(3), and (5), the definition of conditional probability, all hold in conditional form. That is, the rules hold if we replace $Pr(A)$, $Pr(B)$, $Pr(A/B)$, and so on, by $Pr(A/E)$, $Pr(B/E)$, $P(A/B\&E)$, and so on.

Normality

(1C) $0 \leq Pr(A/E) \leq 1$

Certainty

We need to check that for E, such that $Pr(E) > 0$,

(2C) $Pr([\text{sure event}]/E) = 1$.

Now E is logically equivalent to the occurrence of E with something that is sure to happen. Hence,

$$Pr([\text{sure event}]\ \&\ E) = Pr(E).$$
$$Pr([\text{sure event}/E]) = [Pr(E)]/[Pr(E)] = 1.$$

Additivity

Let $Pr(E) > 0$. If A and B are mutually exclusive, then

$$Pr[(AvB)/E] = Pr[(AvB)\&E]/Pr(E) = Pr(A\&E)/Pr(E) + Pr(B\&E)/Pr(E).$$

(3C) $Pr[(AvB)/E] = Pr(A/E) + Pr(B/E)$.

Conditional probability

This is the only case you should examine carefully. The conditionalized form of (5) is:

(5C) If $Pr(E) > 0$ and $Pr(B/E) > 0$, then

$$Pr[A/(B\&E)] = \frac{Pr[(A\&B)/E]}{Pr(B/E)}.$$

We prove this starting from (5),

$$Pr[A/(B\&E)] = \frac{Pr(A\&B\&E)}{Pr(B\&E)}.$$

The numerator (on top of the fraction) is $Pr(A\&B\&E) = Pr[(A\&B)/E] \times Pr(E)$.
The denominator (bottom of the fraction) is $Pr(B\&E) = Pr(B/E) \times Pr(E)$.
Dividing the numerator by the denominator, we get (5C).

Many philosophers and inductive logicians take conditional probability, rather than categorical probability, as the primitive idea. Their basic rules are, then, versions of (1C), (2C), (3C), and (5C). Formally, the end results are in all essential respects identical to our approach that begins with categorical probability and then defines conditional probability. But when we start to ask about various meanings of these rules, we find that a conditional probability approach sometimes makes more sense.

STATISTICAL INDEPENDENCE AGAIN

Our first intuitive explanation of independence (page 25) said that trials on a chance setup are independent if and only if the probabilities of the outcomes of a trial are not influenced by the outcomes of previous trials. But this left open what "influenced" really means. We also spoke of *randomness*, of trials having no *memory*, and of *the impossibility of a gambling system*. These are all valuable metaphors.

The idea of conditional probability makes one exact definition possible. The probability of A should be no different from the probability of A given B, Pr(A/B).

Naturally, independence should be a symmetric relation: A is independent of B if and only if B is independent of A.

In other words, when $0 < \Pr(A)$ and $0 < \Pr(B)$, we expect that:

If Pr(A/B) = Pr(A), then Pr(B/A) = Pr(B) (and vice versa).

This is proved from definition (8) on page 60.

Suppose that Pr(A/B) = Pr(A).
By (5), Pr(A) = [Pr(A&B)]/[Pr(B)].
And so Pr(B) = [Pr(A&B)]/[Pr(A)].
So, since A&B is logically equivalent to B&A,
Pr(B) = Pr(B&A)/Pr(A) = Pr(B/A).

MULTIPLE INDEPENDENCE

Definition (8) defines the statistical independence of a pair of propositions. That is called "pairwise" independence. But a whole group of events or propositions could be mutually independent. This idea is easily defined.

It follows from (6) and (8) that when A and B are statistically independent:

Pr(A&B) = Pr(A)Pr(B)

(See exercise 3.) This can be generalized to the statistical independence of any number of events. For example A, B, and C are statistically independent if and only if A, B, and C are pairwise independent, and

Pr(A&B&C) = Pr(A)Pr(B)Pr(C).

VENN DIAGRAMS

John Venn (1824–1923) was an English logician who in 1866 published the first systematic theory of probabilities explained in terms of relative frequencies. Most people remember him only for "Venn diagrams" in deductive logic. Venn diagrams are used to represent *deductive* arguments involving the quantifiers *all*, *some*, and *no*.

You can also use Venn diagrams to represent probability relations. These drawings help some people who think spatially or pictorially.

Imagine that you have eight musicians:

Four of them are singers, with no other musical abilities.
Three of them can whistle but cannot sing.
One can both whistle and sing.

A Venn diagram can picture this group, using a set of circles. One circle is used for each class. Circles overlap when the classes overlap. Our diagram looks like this:

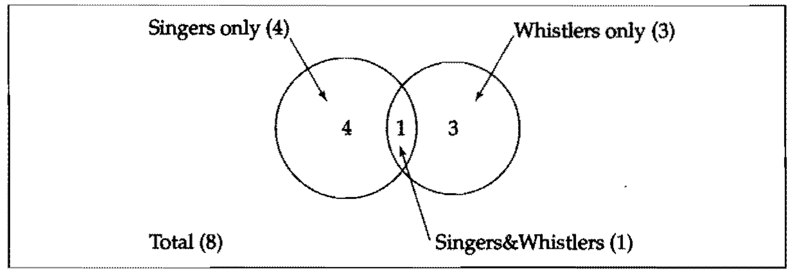


FIGURE 6.1

The circle representing the singers contains five units (four singers plus one singer&whistler), while the circle representing the whistlers has four units (three whistlers plus one singer&whistler). The overlapping region has an area of one unit, since only one of the eight people fits into both categories. We will think of the area of each segment as proportional to the number of people in that segment.

Now say we are interested in the probability of selecting, at random, a singer from the group of eight people. Since there are five singers in the group of eight people, the answer is 5/8.

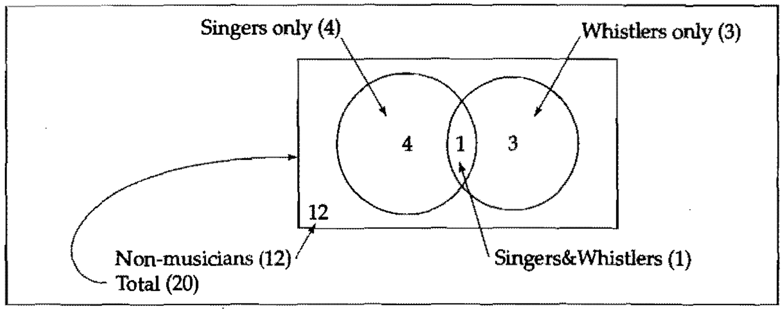
What is the probability that a singer is chosen, on condition that the person chosen is also a whistler? Since you know the person selected is a whistler, this limits the group to the whistlers' circle. It contains four people. Only one of the four is in the singers' circle. Hence only one of the four possible choices is a singer. Hence, the probability that a singer is chosen, given that the singer is also a whistler, is 1/4.

Now let us generalize the example. Put our 8 musicians in a room with 12 nonmusical people, resulting in a group of 20 people. Imagine we were interested in these two events:

Event A = a singer is selected at random from the whole group.
Event B = a whistler is selected at random from the whole group.

Here is a Venn diagram of the situation, where the entire box represents the room full of twenty people.

FIGURE 6.2



Notice the major change from the previous diagram: Figure 6.2 now has its circles enclosed in a rectangle. By convention, the area of the rectangle is set to 1. The areas of each of the circles correspond to the probability of occurrence of an event of the type that it represents: the area of circle A is 5/20, or 0.25, since there are 5 singers among 20 people. Likewise, the area of circle B is 4/20, or 0.2. The area of the region of overlap between A & B is 1/20, or 0.05.

These drawings can be used to illustrate the basic rules of probability.

(1) *Normality*: $0 \leq \Pr(A) \leq 1$.

This corresponds to the rectangle having an area of 1 unit: since all circles must lie within the rectangle, no circle, and hence no event can have a probability of greater than 1.

(2) *Certainty*: Pr(sure event) = 1. Pr(certain proposition) = 1.

With Venn diagrams, an event that is sure to happen, or a proposition that is certain, corresponds to a "circle" that fills the entire rectangle, which by convention has unit area 1.

(3) *Additivity*: If A and B are mutually exclusive, then:

$$\Pr(A \vee B) = \Pr(A) + \Pr(B).$$

If two groups are mutually exclusive they do not overlap, and the area covering members of either group is just the sum of the areas of each.

(4) *Overlap*:

To calculate the probability of AvB, determine how much of the rectangle is covered by circles A and B. This will be all the area in A, plus the area that

appears *only* in B. The area only in B is the areas in B, less the area of overlap with A.

$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \& B)$$

(5) *Conditional*:

Given that event B has happened, what is the probability that event A will also happen? Look at Figure 6.2. If B has happened, you know that the person selected is a whistler. So we want the proportion of the area of B, that includes A. That is, the area of A&B divided by the area of B.

$$\Pr(A/B) = \Pr(A \& B) \div \Pr(B), \text{ so long as } \Pr(B) > 0.$$

So, in our numerical example, $\Pr(A/B) = 1/4$.
Conversely, $\Pr(B/A) = \Pr(A \& B)/\Pr(A) = 1/5 = 0.2$.

ODD QUESTION 2

Recall the Odd Question about Pia:

2. Pia is thirty-one years old, single, outspoken, and smart. She was a philosophy major. When a student, she was an ardent supporter of Native American rights, and she picketed a department store that had no facilities for nursing mothers. Rank the following statements in order of probability from 1 (most probable) to 6 (least probable). (Ties are allowed.)

_______(a) Pia is an active feminist.
_______(b) Pia is a bank teller.
_______(c) Pia works in a small bookstore.
_______(d) Pia is a bank teller and an active feminist.
_______(e) Pia is a bank teller and an active feminist who takes yoga classes.
_______(f) Pia works in a small bookstore and is an active feminist who takes yoga classes.

This is a famous example, first studied empirically by the psychologists Amos Tversky and Daniel Kahneman. They found that very many people think that, given the whole story:

> The most probable description is (f) Pia works in a small bookstore and is an active feminist who takes yoga classes.

In fact, they rank the possibilities something like this, from most probable to least probable:

(f), (e), (d), (a), (c), (b).

But just look at the logical consequence rule on page 60. Since, for example, (f) logically entails (a) and (b), (a) and (b) must be more probable than (f).

In general:

$\Pr(A\&B) \le \Pr(B)$.

It follows that the probability rankings given by many people, with (f) most probable, are completely wrong. There are many ways of ranking (a)–(f), but any ranking should obey these inequalities:

$\Pr(a) \ge \Pr(d) \ge \Pr(e)$.
$\Pr(b) \ge \Pr(d) \ge \Pr(e)$.
$\Pr(a) \ge \Pr(f)$.
$\Pr(c) \ge \Pr(f)$.

ARE PEOPLE STUPID?

Some readers of Tversky and Kahneman conclude that we human beings are irrational, because so many of us come up with the wrong probability orderings. But perhaps people are merely *careless*!

Perhaps most of us do not attend closely to the exact wording of the question, "Which of statements (a)–(f) are more probable, that is have the highest probability."

Instead we think, "Which is the most useful, instructive, and likely to be true thing to say about Pia?"

When we are asked a question, most of us want to be informative, useful, or interesting. We don't necessarily want simply to say what is most probable, in the strict sense of having the highest probability.

For example, suppose I ask you whether you think the rate of inflation next year will be (a) less than 3%, (b) between 3% and 4%, or (c) greater than 4%.

You could reply, (a)-or-(b)-or-(c). You would certainly be right! That would be the answer with the highest probability. But it would be totally uninformative.

You could reply, (b)-or-(c). That is more probable than simply (b), or simply (c), assuming that both are possible (thanks to additivity). But that is a less interesting and less useful answer than (c), or (b), by itself.

Perhaps what many people do, when they look at Odd Question 2, is to form a character analysis of Pia, and then make an interesting guess about what she is doing nowadays.

If that is what is happening, then people who said it was most probable that Pia works in a small bookstore and is an active feminist who takes yoga classes, are not irrational.

They are just answering the wrong question—but maybe answering a more useful question than the one that was asked.

AXIOMS: HUYGENS

Probability can be axiomatized in many ways. The first axioms, or basic rules, were published in 1657 by the Dutch physicist Christiaan Huygens (1629–1695), famous for his wave theory of light. Strictly speaking, Huygens did not use the idea of probability at all. Instead, he used the idea of the fair price of something like a lottery ticket, or what we today would call the expected value of an event or proposition. We can still do that today. In fact, almost all approaches take probability as the idea to be axiomatized. But a few authors still take expected value as the primitive idea, in terms of which they define probability.

AXIOMS: KOLMOGOROV

The definitive axioms for probability theory were published in 1933 by the immensely influential Russian mathematician A. N. Kolmogorov (1903–1987). This theory is much more developed than our basic rules, for it applies to infinite sets and employs the full differential and integral calculus, as part of what is called measure theory.

EXERCISES

1 *Venn Diagrams.*

Let L: A person contracts a lung disease.
Let S: That person smokes.

Write each of the following probabilities using the Pr notation, and then explain it using a Venn diagram.

(a) The probability that a person either smokes or contracts lung disease (or both).
(b) The probability that a person contracts lung disease, given that he or she smokes.
(c) The probability that a person smokes, given that she or he contracts lung disease.

2 *Total probability.* Prove from the basic rules that $\Pr(A) + \Pr(\sim A) = 1$.

3 *Multiplying.* Prove from the definition of statistical independence that if $0 < \Pr(A)$, and $0 < \Pr(B)$, and A and B are statistically independent,

$\Pr(A\&B) = \Pr(A)\Pr(B)$.

4 *Conventions.* In Chapter 4, page 40, we said that the rules for normality and certainty are just conventions. Can you think of any other plausible conventions for representing probability by numbers?

5 *Terrorists.* This is a story about a philosopher, the late Max Black.

One of Black's students was to go overseas to do some research on Kant. She was afraid that a terrorist would put a bomb on the plane. Black could not convince her that the risk was negligible. So he argued as follows:

BLACK: Well, at least you agree that it is almost impossible that *two* people should take bombs on your plane?

STUDENT: Sure.

BLACK: Then you should take a bomb on the plane. The risk that there would be another bomb on your plane is negligible.

What's the joke?

KEY WORDS FOR REVIEW

Normality	Conditional probability	Total probability
Certainty	Venn diagrams	Logical consequence
Additivity	Multiplication	Statistical independence

7 Bayes' Rule

One of the most useful consequences of the basic rules helps us understand how to make use of new evidence. Bayes' Rule is one key to "learning from experience."

Chapter 5 ended with several examples of the same form: urns, shock absorbers, weightlifters. The numbers were changed a bit, but the problems in each case were identical.

For example, on page 51 there were two urns A and B, each containing a known proportion of red and green balls. An urn was picked at random. So we knew:

Pr(A) and Pr(B).

Then there was another event R, such as drawing a red ball from an urn. The probability of getting red from urn A was 0.8. The probability of getting red from urn B was 0.4. So we knew:

Pr(R/A) and Pr(R/B).

Then we asked, what is the probability that the urn drawn was A, *conditional* on drawing a red ball? We asked for:

Pr(A/R) = ? Pr(B/R) = ?

Chapter 5 solved these problems directly from the definition of conditional probability. There is an easy rule for solving problems like that. It is called *Bayes' Rule.*

In the urn problem we ask which of two *hypotheses* is true: Urn A is selected, or Urn B is selected. In general we will represent hypotheses by the letter H.