12 Theories about Probability

There have been two fundamentally different approaches to probability. One emphasizes the frequency idea. The other emphasizes the belief idea.

Some theorists say that only one of those two ideas really matters. We will call them *dogmatists*. In this book we are *eclectic*. Here are two definitions taken from a dictionary:

- Eclectic. Adjective. 1. (in art, philosophy, etc.) Selecting what seems best from various styles, doctrines, ideas, methods, etc.
- Dogmatic. Adjective. 1. Characterized by making authoritative or arrogant assertions or opinions, etc.

FREQUENCY DOGMATISTS

Some experts believe that all inductive reasoning should be analyzed in terms of frequency-type probabilities. This is a dogmatic philosophy, saying that inductive reasoning should rely on exactly one use of probability. Belief dogmatists often say that frequency-type probabilities "have no role in science."

BELIEF DOGMATISTS

Some experts believe that all inductive reasoning should be analyzed in terms of belief-type probabilities. This is a dogmatic philosophy, saying that inductive reasoning should rely on exactly one use of probability. Belief dogmatists often say that frequency-type probability "doesn't even make sense."

OUR ECLECTIC APPROACH

Chapters 16–19 and 22 use the frequency idea. Chapters 13–15 and 21 use the belief idea.

Luckily, most (but not all) data and arguments that a frequency dogmatist can analyze, a belief dogmatist can analyze too. And vice versa. Only in rather specialized situations do the two schools of thought draw really different inferences from the same data. They differ more about how to get data in the first place. They disagree about how to design experiments in order to get the most useful information.

Our eclectic approach values both families of ideas.

THE WEATHER

Weather forecasts come with probabilities:

(a) The probability of precipitation tomorrow is 30%.

What does that mean? Since the statement is about rain (or snow or hail) *tomorrow*, it is about a single case. Either it rains, or it does not. There is only one tomorrow. So this must be a belief-type probability.

If you ask the weather office what it means, you will hear something like this:

(b) We have compared the weather conditions for the past week with a very large data bank, and have selected all those weeks which in the relevant ways are like the week just ended. Of those weeks, 30% were followed by rain. The probability of rain the next day, after a week like last week, is 30%.

This is stated as a frequency-type probability. If the weather office means what it says, then it must be using the frequency principle to go from (b) to (a). Some meteorologists say that when their office issues forecast (a), it just *means* (b)—or better, it is a short form of (b). They say that they always mean frequency-type probabilities. They are *frequency dogmatists*.

Belief dogmatists suggest that the weather people don't know what they are talking about. Meteorologists ought never to mean frequencies. For they don't rely only on past frequencies of precipitation. They use all sorts of information, including (b), and make some judgment calls. But statement (a) is exactly what it seems. It is used to state the extent to which the weather people expect rain.

Luckily, most people don't know about such dogmatic infighting. No one, in daily life, is worried about what forecast (a) means.

FOUR THEORIES ABOUT PROBABILITY

Both kinds of dogmatist try to make their ideas as clear as possible. Hence there are quite a lot of what philosophers call "theories of probability." We will briefly describe two frequency-type theories, and two belief-type theories.

PERSONAL PROBABILITY

Recall the dinosaurs. Someone says, "The reign of the dinosaurs was brought to an end when a giant asteroid hit the Earth." We can say something quite impersonal about this claim:

It is probable that the reign of the dinosaurs was brought to an end when a giant asteroid hit the Earth.

That suggests that "anyone" who is well informed ought to make the same judgment. This *interpersonal* notion leads to anonymous assertions like:

One can be very confident that the dinosaurs were extinguished after a giant asteroid hit the Earth.

Then there are *personal* statements, which express the speaker's own personal confidence or degree of belief:

I am very confident that the dinosaurs were extinguished after a giant asteroid hit the Earth.

This leads to the idea of personal probability. It is a kind of belief-type probability. You might think that personal probability is totally "subjective" and of no value for inductive logic. In fact, it is a very powerful idea, developed in Chapters 13–15. People who want to be consistent must (we show) have personal probabilities that obey the basic rules of probability.

LOGICAL PROBABILITY

Personal probability is one rather extreme kind of belief-type probability. At the opposite extreme is logical probability. Here is a long probability statement about dinosaurs:

Relative to recent evidence about a layer of iridium deposits in many parts of the Earth, geologically identified as contemporary with the extinction of the dinosaurs, the probability is 90% that the reign of the dinosaurs was brought to an end when a giant asteroid hit the Earth.

One could also say:

Relative to the *available evidence*, it is *reasonable* to have a high *degree of belief* in the dinosaur/asteroid hypothesis.

This is a statement of a conditional probability. We called it *interpersonal/evidential*.

Let H be the hypothesis about dinosaurs, and let E be the evidence. Then this statement is of the form,

$Pr(H/E) = 0.9$.

The theory of logical probability says statements like that express a *logical* relation between H and E, analogous to logical deducibility.

Logical probabilities are supposed to be logical relations between evidence and an hypothesis. On this theory, it makes no sense to talk about the probability of H all by itself.

According to the theory of logical probability, probability is always *relative to evidence*. When we don't actually mention evidence, we are implicitly referring to some body of evidence.

Frequency dogmatists cannot abide this conclusion. One of the first explicit frequency theories about probability was due to John Venn (1834–1923), better known as the inventor of Venn diagrams (page 63). He said that

The probability of an event is no more relative to something else than the area of a field is relative to something else.

THE PRINCIPLE OF INSUFFICIENT REASON

What if there is no relevant evidence at all? From the earliest days of numerical probabilities, people have been inclined to say that if there is no reason to choose among several alternatives, they should be treated as equally probable.

Example: Peter on a road in the Netherlands (page 114) does not know if the sign,

[redacted]TERDAM →

indicates the way to Amsterdam or to Rotterdam. He sees no reason to choose among the alternatives. He does not know of any other major city ending in "terdam," so he says that the probability of each is 0.5.

Example: Mario Baguette is to meet his friend Sonia at the Metropolitan Airport. She is flying in from New York. She has told Mario that she is arriving between 3:00 and 4:00 in the afternoon. Five different airlines have flights from New York arriving during that hour—Alpha Air, Beta Blaster, Gamma Goways, Delta, and Eatemup Airlines. Mario can think of no reason to favor one over the others, so he says that the probability of each is 0.2. Alpha and Beta arrive at terminal 1, Gamma and Delta at terminal 2, and Eatemup arrives at terminal 3. Hence his expectation for Eatemup and terminal 3 is lowest, and he flips a coin to see whether he goes to terminal 1 or 2.

The advocate of logical probability says that when there is no evidence favoring one of n mutually exclusive and jointly exhaustive hypotheses, then one should assign them each probability $1/n$. This is called *the principle of insufficient reason*, or *the principle of indifference*.

TOO EASY BY HALF

The principle of insufficient reason sounds fine. But there are real problems. Would you say it is equally probable that a car is red or not red? Or that it is red or blue or green or some other color? The principle quickly leads to a lot of paradoxes.

BAYESIANS

We will see in Chapter 15 how both personal and logical probability theories make heavy use of Bayes' Rule. Belief dogmatists who make exclusive use of belief-type probabilities and place great emphasis on Bayes' Rule are often called Bayesians.

KEYNES AND RAMSEY

The idea of logical probability—which is always "relative to the evidence"—goes back at least 200 years, but the first systematic presentation of logical probability, as a logical relation, was by John Maynard Keynes (1883–1946). Keynes is the famous economist whose theories are sometimes credited with saving capitalism during the crisis of the Great Depression of 1929–1936. His *Treatise on Probability* (1921) was his first major work, submitted for a prize fellowship (something like a post-doc, but with no Ph.D.) at Cambridge University.

In 1926, a second young man at Cambridge, Frank Plumpton Ramsey (1903–1930), presented the first modern theory of personal probability (see page 000). He died during an operation when he was twenty-six, after having made fundamental contributions to economics, mathematical logic, probability theory, and philosophy.

Logicians, and especially Rudolf Carnap, have been very attracted to Keynes's logical probability. Nevertheless, a few simple words by Ramsey are impressive.

> The most "fundamental criticism of Mr. Keynes' views . . . is the obvious one that there really do not seem to be any such things as the probability relations he describes."

DE FINETTI AND SAVAGE

Two great figures in the history of personal probability are the Italian mathematician Bruno de Finetti and the American L. J. Savage. You will find out more about Savage on page 184. De Finetti pioneered his ideas, in Italian, at about the same time as Ramsey did in English. But Ramsey thought there was room, in his mental space, for a frequency-type concept of probability, especially in quantum mechanics. De Finetti was convinced that only a personal belief-type concept made sense. So was Savage. Both were belief dogmatists, who taught us an enormous amount about the idea of personal probability and what is now called Bayesian statistics.

LIMITING FREQUENCY

Belief-type probabilities split into logical probabilities at one extreme and personal probabilities at another. Something similar happens with frequency-type probabilities.

Thus far we have mentioned a family of ideas: frequency, long run, geometrical symmetry, physical property, tendency, propensity. How can these be made precise? One way is to emphasize relative frequency in the long run. What is a "long run"? Mathematicians immediately think of an idealization: an *infinite* sequence of trials.

The next idea that occurs to a mathematician is that of a *convergence* and *mathematical limit*. That leads to the idealization of an infinite sequence of trials in which the relative frequency of an outcome S *converges* to a *limit*.

But there is another requirement. Think of a sequence of tosses of a coin, in which heads and tails alternate:

H T H T H T. . . .

The relative frequency certainly converges to ½, but we do not think of this as a probability, because the sequence is fully determinate. So we add a further clause to this definition of probability. It was mentioned earlier on page 28: *The impossibility of a gambling system.*

This theory about probability was first stated clearly by John Venn in 1866. It was systematically developed in 1928 by the Austrian-born applied mathematician (aerodynamics) and philosopher Richard von Mises (1883–1953), who became a professor at Harvard.

The idea of "the impossibility of a gambling system" has turned out to be very fruitful. It has led to fundamental new ideas about computational complexity, pioneered by A. N. Kolmogorov (see page 67). Roughly speaking, a sequence is now said to be random, relative to a class of ideal computer programs, if any program sufficient to generate the sequence is at least as long as the sequence itself.

PROPENSITY

The limiting frequency idea is one extreme frequency-type probability. It idealizes a series of actual results of trials on a chance setup. It emphasizes what can be seen to happen, rather than underlying causes or structures.

Stable relative frequencies exist only if a setup has some underlying physical or geometrical properties. And we can have that underlying structure even if in fact no actual trials are ever made.

A second extreme frequency-type idea emphasizes the tendency, disposition, or *propensity* of the chance setup. This seems particularly natural when we think of stochastic processes found in nature, such as radioactive decay. We are concerned less with the distribution of clicks on a Geiger counter, than with the basic physical processes.

The propensity account of probability was developed by the Austrian-born philosopher Karl Popper (1902–1994), who became a professor at the London School of Economics. Popper may have been the most influential philosopher of science of the twentieth century. We discuss his evasion of the problem of induction in Chapter 20. Popper believed that his propensity approach was important for understanding fundamental issues of quantum mechanics.

EXPECTED VALUES

We have spent so much space on kinds of probability, that we may be in danger of ignoring expected values and fair prices.

When expected values are explained in terms of probability, every ambiguity in the word "probability" shows up in "expected value." Thus there are personal expected values, logical expected values (conditional on evidence), propensity expected values, and so forth. And there are the usual debates between different theories. This anecdote will make the point.

THE OLD COKE MACHINE

Once upon a time a bottle of Coca-Cola cost only five cents. Coke machines would dispense a bottle for a nickel. Then the price went up to six cents. *Unhappy Sailor* remembered this to his dying day. (He told me this story himself, when I visited him at the U.S. Bureau of Standards, where he had become chief statistician.)

He visited a Pacific naval base where the machine had been fixed to allow for inflation. Each time you put a nickel in, there was a one-in-six chance that you got nothing. *Sailor* was ashore for one day only, and he wanted just one bottle of Coke. He lost the first time, and he had to pay another nickel to get his drink. "That's not fair!" said he.

The machine was equivalent to the following lottery. There are six tickets, five of which have a prize of a six-cent bottle of Coke, and one of which has no prize. The fair price is then five cents.

Up in philosophers' paradise, the debate is still going on.

GHOSTS OF VENN, VON MISES, POPPER, AND OTHER FREQUENCY DOGMATISTS: And that, of course, was the fair price, because on average, sailors on the base paid six cents for each bottle of Coke.

GHOST OF *Unhappy Sailor*: I didn't care about the long run. I was only on the island for a day, and I had to pay a dime for my six-cent bottle of Coke. It may have been fair for the people stationed permanently at the base, but it was not fair for me.

GHOSTS OF KEYNES AND OTHER BELIEF DOGMATISTS: Exactly so. Talk of averages in the long run is completely irrelevant. The machine was fair. Putting a nickel in the machine was equivalent to having a ticket in a symmetric zero-sum lottery with six tickets. Five tickets give a prize worth six cents, one ticket gives nothing. So the total prizes added up to 30 cents. There are six players, and each should put in a nickel. The unfairness is not in the lottery. The unfairness is in forcing *Unhappy Sailor* to play this lottery. There are many kinds of unfairness, and being forced to gamble for a soft drink is one of them.

WHY DO WE HAVE DIFFERENT KINDS OF PROBABILITY?

Here is a simple answer. Numerical probabilities began to be used in earnest a long time ago, around 1650. Naturally, people needed simple examples. Gambling was very common, and so were lotteries. A lot of gambling relied on artificial randomizers such as dice, coins, and packs of cards. Lotteries used an artificial randomizer. Moreover, many of the randomizers of choice were supposed to be symmetric. Hence it really did not make much difference what you meant by probability. Let's take a lottery with 1,000 numbered balls placed in an urn, and see why it makes little difference.

Relative frequency talk: On repeated draws with replacement from this lottery box, each ball is drawn as often as every other. So the probability of drawing ball #434 is the relative frequency of drawing that ball, namely 1/1000.

Propensity talk: The urn, balls, and drawing procedure are arranged so that the propensity for one ball to be drawn is the same as for any other. Hence the probability of drawing ball #434 is 1/1000.

Personal talk: So far as I'm concerned, every ball is equally likely, so for me, the probability of drawing ball #434 is 1/1000.

Interpersonal talk: No reasonable person would give greater probability to one ball over another, so the probability of drawing ball #434 is 1/1000.

Logical talk: The logical relation between any h_j (the hypothesis that ball #*j* is drawn) and the available evidence, and that between any other h_k and the available evidence, is just the same. So the probabilities are equal, and the probability of drawing ball #434 is 1/1000.

Principle of insufficient reason talk: There is no evidence favoring one of the 1000 mutually exclusive and jointly exhaustive hypotheses over any other. Hence one should assign them each probability 1/1000.

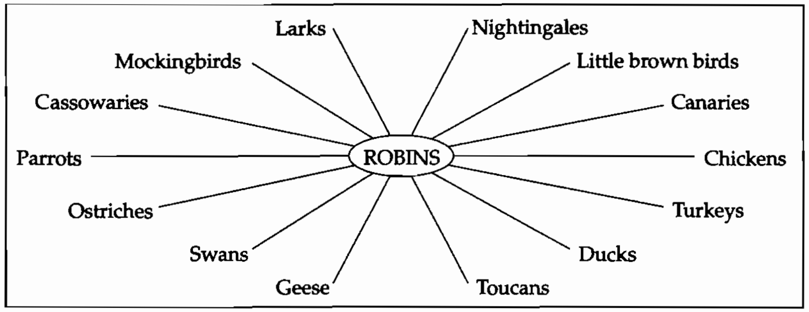
MODELING

The next step in the development of probabilities was to model more complex situations on core examples such as urns, lotteries, dice, cards. It still did not matter much how one understood the core examples. Only when we get quite sophisticated modeling does it matter very much how we think of probabilities. Yes, our shock absorber example is naturally understood in terms of frequencies, and our strep throat example is naturally understood in terms of degrees of belief. But until one starts thinking very carefully about statistical inference, only pedantic philosophers care much about the meaning of probability.

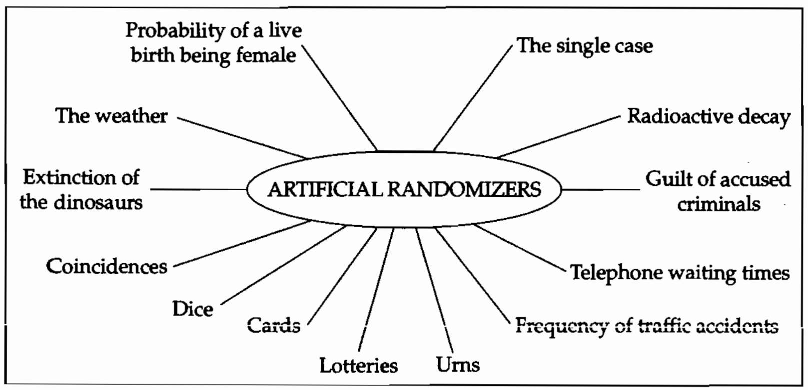
MEANINGS

"Define your terms!" How often have you heard that. But definition is not always what we want. Many words are best explained, in the beginning, by examples. Suppose you wanted to explain the meaning of the word "bird" to an alien. You would probably point to small or middle-sized birds that are common where you live. Famous experiments in California showed that most students thought a robin was the prototypical bird. In England you can get the same result, except that the English robin (redbreast) is a completely different kind of bird from the North American robin! Myself, I would not pick a robin but some of those little brown birds that congregate in the tree outside my window. But except for smart alecks who know about this experiment, no one in North America chooses an ostrich as the prototypical bird, let alone a cassowary. Few choose ducks, fewer still swans. Few choose eagles or owls.

Who can define a bird? Maybe some ornithologists can, but most of us get on just fine talking about birds without defining "bird." We think of birds as being more or less like our prototype of a bird. Ostriches are in some ways like my little brown birds. So are eagles, but ostriches and eagles are not much alike. We could think of birds as radiating out from a center, where in the center we have our prototype for birds, and the birds on the periphery have some resemblance to the birds in the center. According to the experiments, those students in northern California have a picture of birds rather like this:



Much the same is true of probability. Our prototypical examples are artificial randomizers. But as we start to think hard about more real-life examples, we get further and further away from the core examples. Then our examples tend to cluster into belief-type examples, and frequency-type examples, and in the end we develop ideas of two different kinds of probability:



THE LAST WORD

From now on we will be careful about kinds of probability. In previous chapters we tended to speak, wherever possible, in a frequency-type way. It is only fair that we end the first part of the book by letting a belief dogmatist have the last word.

John Maynard Keynes, advocate of the logical theory of belief-type probability, had no use for frequency-type probabilities. He jeered at the idea of relative frequency in the long run. He asked what could be meant by "the long run."

"The long run" is just a metaphor. Infinite long runs may be a nice idealization, but have nothing to do with real life. In another context Keynes remarked:

"In the long run we are all dead."

EXERCISES

1 *Indifference.* Look back at Mario Baguette's problem about meeting Sonia (page 143). If Mario suspects that Sonia will take the budget airline, Eatemup, he may not be indifferent among the five choices. Now he has a problem in personal probabilities; does he think it more likely she would come on Alpha *or* Beta, or on the cheap Eatemup? What would you do?

The remaining exercises are based on an imaginary symposium. Once upon a time four dogmatists were put on a stage and asked to explain their views about probability. We'll call these men

Venn, the frequency theorist.
Popper, the propensity theorist.
De Finetti, the personal probability theorist.
Keynes, the theorist of logical probability.

The master of ceremonies read out several statements and asked each dogmatist to say what it meant, if anything. What did each of the four men say about the italicized statements below?

2 *Happy Harry's Preparation Kit for the Graduate Record Examination (GRE). "Students who use Happy Harry's Kit have a 90% probability of success in the GRE!"*

3 *The informed source.* An informed source who did not wish to be identified was asked about the recent spate of bombings and retaliations. She said that *there is little probability of a lasting Middle East peace settlement in the next two years.*

4 *Conditional Happy Harry.* "If your probability of success on the GRE is only 50–50 before you use Happy Harry's kit, it will be 75% after you use it." The emcee told the dogmatists to take this as a compound statement involving two probability statements: "If *your probability of success is 0.5, on condition that you do not use Harry's kit*, then *your probability of success is 0.75, on condition that you do use Harry's kit.*"

5 *The Great One.* Wayne Gretzky, who scored more goals than anyone in the history of ice hockey, developed arthritis at age thirty-eight and started doing paid TV ads for Tylenol. How could a healthy man have arthritis in his thirties? Dr. Clarfield, a hockey physician, said, "He played enough hockey for three careers, so *the trauma in his joints was probably more than it was for most players.*"

6 *Landing on Mars.* A NASA spokeswoman said that *the first person to set foot on Mars would in all probability be an engineer.*

7 *Lunatic coin tossers.* A fair coin is being tossed. We say that after N tosses, heads is in the lead if heads has come up more times than tails. Conversely, tails is in the lead if tails has come up more often. (With an even number N, they could be tied.) We say that the lead changes at toss N if heads was in the lead at N−1, but tails is in the lead at N+1. *The probability that in 10,000 tossings of a fair coin, the lead never changes, is about .0085.*

8 *An amazing answer.* Here you are asked about the meaning of probability in a question. The answer to this question is given in the answer section, but you are not expected to know the answer here. But make a guess. *What is the probability that in 10,000 tosses, one side is in the lead for more than 9,930 tosses, while the other side is in the lead for fewer than 70 tosses?*

KEY WORDS FOR REVIEW

Frequency theory	Personal theory
Propensity theory	Logical theory

13 Personal Probabilities

How personal degrees of belief can be represented numerically by using imaginary gambles.

Chapters 1–10 were often *deliberately ambiguous* about different kinds of probability. That was because the basic ideas usually applied, across the board, to most kinds of probability.

Now we develop ideas that matter a lot for belief-type probabilities. They do not matter so much from the frequency point of view.

THE PROGRAM

There are three distinct steps in the argument, and each deserves a separate chapter.

- *This chapter* shows how you might use numbers to represent your degrees of belief.
- *Chapter 14* shows why these numbers should satisfy the basic rules of probability. (And hence they should obey Bayes' Rule.)
- *Chapter 15* shows how to use Bayes' Rule to *revise* or *update* personal probabilities in the light of new evidence. This is the fundamental motivation for the group of chapters, 13–15.

In these chapters we are concerned with a person's degrees of belief. We are talking about personal probabilities. But this approach can be used for other versions of belief-type probability, such as the logical perspective of Keynes and Carnap.

Because Bayes' Rule is so fundamental, this approach is often called *Bayesian.* "Belief dogmatists" are often simply called Bayesians because the use of Bayes'