CHAPTER 3

# Basic Concepts in Probability and Statistics, Part 2

*The Relationship of Probability to Other Magnitudes*
*The Concept of Chance*
*The Philosophers' Dispute About The Concept Of Probability*
*The Relationship Of Probability To The Concept Of Resampling*
*Conclusion*

This chapter may be skipped by those anxious to reach the actual machinery of estimating probabilities.

## The relationship of probability to other magnitudes

An important argument in favor of approaching the concept of probability with the concept of the operational definition is that an estimate of a probability often (though not always) is the opposite side of the coin from an estimate of a physical quantity such as time or space.

For example, uncertainty about the probability that one will finish a task within 9 minutes is another way of labeling the uncertainty that the time required to finish the task will be less than 9 minutes. Hence, if an operational definition is appropriate for time in this case, it should be equally appropriate for probability. The same is true for the probability that the quantity of radios sold will be between 200 and 250 units.

Hence the concept of probability, and its estimation in any particular case, should be no more puzzling than is the "dual" concept of time or distance or quantities of radios. That is, lack of certainty about the probability that an event will occur is not different in nature from lack of certainty about the amount of time or distance in the event. There is no essential difference between whether a part 2 inches in length will be the next to emerge from the machine, or what the length of the next part will be, or the length of the part that just emerged (if it has not yet been measured).

The information available for the measurement of (say) the length of a car or the location of a star is exactly the same information that is available with respect to the concept of probability in those situations. That is, one may have ten disparate observations of an auto's length which then constitute a probability distribution, and the same for the altitude of a star in the heavens. All the more reason to see the parallel between Einstein's concept of *time* and *length* as being what you measure on a clock and on a meter stick, respectively—or better, that time and length are *equivalent to* the *measurements* that one makes on a clock or meter stick—and the notion that *probability* should be defined by the measurements made on a clock or a meter stick. Seen this way, all the discussions of logical and empirical notions of probability may be seen as being made obsolete by the Einsteinian invention of the operational definition, just as discussions of absolute space and time were made obsolete by it.

Or: Consider having four different measurements of the length of a model auto. Which number should we call the length? It is standard practice to compute the mean. But the mean could be seen as a weighted average of each observation by its probability. That is:

(.25 * 20 inches + .25 * 22 inches...) = mean model length, instead of (20 + 22 + ...) / 4 = mean model length

This again makes clear that the decimal weights we call "probabilities" have no extraordinary properties when discussing frequency series; they are just weights we put on some other values.

It should be noted that the view outlined above has absolutely no negative implications for the formal mathematical theory of probability.

In a book of puzzles about probability (Mosteller, 1965/1987, #42), this problem appears: "If a stick is broken in two at random, what is the average length of the smaller piece?" This particular puzzle does not even mention probability explicitly, and no one would feel the need to write a scholarly treatise on the meaning of the word "length" here, any more than one would one do so if the question were about an astronomer's average observation of the angle of a star at a given time or place, or the average height of boards cut by a carpenter, or the average size of a basketball team. Nor would one write a treatise about the "meaning" of "time" if a similar puzzle involved the average time between two bird calls. Yet a rephrasing of the problem reveals its tie to the concept of probability, to wit: What is the probability that the smaller piece will be (say) more than half the length of the larger piece? Or, what is the probability distribution of the sizes of the shorter piece?

The duality of the concepts of probability and physical entities also emerges in Whitworth's discussion (1897/1965) of fair betting odds:

> …What sum ought you fairly give or take now, while the event is undetermined, in exchange for the assurance that you shall receive a stated sum (say $1,000) if the favourable event occur? The chance of receiving $1,000 is worth something. It is not as good as the certainty of receiving $1,000, and therefore it is worth less than $1,000. But the prospect or expectation or chance, however slight, is a commodity which may be bought and sold. It must have its price somewhere between zero and $1,000. (p. xix.)
>
> …And the ratio of the expectation to the full sum to be received is what is called the chance of the favourable event. For instance, if we say that the chance is 1/5, it is equivalent to saying that $200 is the fair price of the contingent $1,000. (p. xx.)…
>
> The fair price can sometimes be calculated mathematically from *a priori* considerations: sometimes it can be deduced from statistics, that is, from the recorded results of observation and experiment. Sometimes it can only be estimated generally, the estimate being founded on a limited knowledge or experience. If your expectation depends on the drawing of a ticket in a raffle, the fair price can be calculated from abstract considerations: if it depend upon your outliving another person, the fair price can be inferred from recorded statistics: if it depend upon a benefactor not revoking his will, the fair price depends upon the character of your benefactor, his habit of changing his mind, and other circumstances upon the knowledge of which you base your estimate. But if in any of these cases you determine that $300 is the sum which you ought fairly to accept for your prospect, this is equivalent to saying that your chance, whether calculated or estimated, is 3/10... (p. xx.)

It is indubitable that along with frequency data, a wide variety of other information will affect the odds at which a reasonable person will bet. If the two concepts of probability stand on a similar footing here, why should they not be on a similar footing in *all* discussion of probability? Why should both kinds of information not be employed in an operational definition of probability? I can think of no reason that they should not be so treated.

Scholars write about the "discovery" of the concept of probability in one century or another. But is it not likely that even in pre-history, when a fisherperson was asked how long the big fish was, s/he sometimes extended her/his arms and said, "About this long, but I'm not exactly sure," and when a scout was asked how many of the enemy there were, s/he answered, "I don't know for sure...probably about fifty." The uncertainty implicit in these statements is the functional equivalent of probability statements. There simply is no need to make such heavy work of the probability concept as the philosophers and mathematicians and historians have done.

## The concept of chance

The study of probability focuses on randomly generated events—that is, events about which there is uncertainty whether or not they will occur. And the uncertainty refers to your knowledge rather than to the event itself. For example, consider this lecture illustration with a salad spoon.

I spin the salad spoon like a baton twirler. If I hold it at the handle and attempt to flip it so that it turns only half a revolution, I can be almost sure that I will correctly get the spoon end and not the handle. And if I attempt to flip it a full revolution, again I can almost surely get the handle successfully. It is not a random event whether I catch the handle or the head (here ignoring those throws when I catch neither end) when doing only half a revolution or one revolution. The result is quite predictable in both these simple maneuvers so far.

When I say the result is "predictable," I mean that you would not bet with me about whether this time I'll get the spoon or the handle end. So we say that the outcome of my flip aiming at half a revolution is not "random."

When I twirl the spoon so little, I *control* (almost completely) whether it comes down the handle or the spoon end; this is the same as saying that the outcome does not occur by chance.

The terms "random" and "chance" implicitly mean that you believe that I cannot control or cannot know in advance what will happen.

Whether this twirl will be the rare time I miss, however, *should* be considered chance. Though you would not bet at even odds on my catching the handle versus the spoon end if there is to be only a half or one full revolution, you might bet—at (say) odds of 50 to 1—whether I'll make a mistake and get it wrong, or drop it. So the very same flip can be seen as random or determined depending on what aspect of it we are looking at.

Of course you would not bet *against* me about my *not* making a mistake, because the bet might *cause* me to make a mistake purposely. This "moral hazard" is a problem that emerges when a person buys life insurance and may commit suicide, or when a boxer may lose a fight purposely. The people who stake money on those events say that such an outcome is "fixed" (a very appropriate word) and not random.

Now I attempt more difficult maneuvers with the ladle. I can do 1-1\2 flips pretty well, and two full revolutions with some success—maybe even 2-1/2 flips on a good day. But when I get much beyond that, I cannot determine very well whether I'll get handle or spoon. The outcome gradually becomes less and less predictable—that is, more and more random.

If I flip the spoon so that it revolves three or more times, I can hardly control the process at all, and hence I cannot predict well whether I'll get the handle or the head. With 5 revolutions I have absolutely no control over the outcome; I cannot predict the outcome better than 50-50. At that point, getting the handle or the spoon end has become a very random event for our purposes, just like flipping a coin high in the air. So at that point we say that "chance" controls the outcome, though that word is just a synonym for my lack of ability to control and predict the outcome. "Chance" can be thought to stand for the myriad small factors that influence the outcome.

We see the same gradual increase in randomness with increasing numbers of shuffles of cards. After one shuffle, a skilled magician can know where every card is, and after two shuffles there is still much order that s/he can work with. But after (say) five shuffles, the magician no longer has any power to predict and control, and the outcome of any draw can then be thought of as random chance.

At what point do we say that the outcome is "random" or "pure chance" as to whether my hand will grasp the spoon end, the handle, or at some other spot? *There is no sharp boundary to this transition.* Rather, the transition is gradual; this is the crucial idea, and one that I have not seen stated before.

Whether or not we refer to the outcome as random depends upon the twirler's skill, which influences how predictable the event is. A baton twirler or juggler might be able to do ten flips with a non-random outcome; if the twirler is an expert and the outcome is highly predictable, we say it is not random but rather is determined.

Again, this shows that the randomness is not a property of the physical event, but rather of a person's knowledge and skill.

## What Do We Mean by "Chance"?

We have defined "chance" as the absence of predictive power and/or explanation and/or control. Here we should not confuse the concepts of determinacy-indeterminacy and predictable-unpredictable. What matters for *decision purposes* is whether you can predict. Whether the process is "really" determinate is largely a matter of definition and labeling, an unnecessary philosophical controversy for our operational purposes (and perhaps for any other purpose). Much more discussion of this general topic may be found in my forthcoming book *The Philosophy and Practice of Resampling Statistics.*

The ladle in the previous demonstration *becomes* unpredictable—that is, random—even though it still is subject to similar physical processes as when it is predictable. I do not deny *in principle* that these processes can be "understood," or that one could produce a machine that would—like a baton twirler—make the course of the ladle predictable for many turns. But in practice we cannot make the predictions—and it is the practical reality, rather than the principle, that matters here.

When I flip the ladle half a turn or one turn, I control (almost completely) whether it comes down at the handle end or the spoon end, so we do not say that the outcome is chance. Much the same can be said about what happens to the predictability of drawing a given card as one increases the number of times one shuffles a deck of cards.

Consider, too, a set of fake dice that I roll. Before you know they are fake, you assume that the probabilities of various outcomes is a matter of chance. But after you know that the dice are loaded, you no longer assume that the outcome is chance. This illustrates how the probabilities you work with are influenced by your knowledge of the facts of the situation.

Admittedly, this way of thinking about probability takes some getting used to. For example, suppose a magician does a simple trick with dice such as this one:

> The magician turns his back while a spectator throws three dice on the table. He is instructed to add the faces. He then picks up any *one* die, adding the number on the *bottom* to the previous total. This same die is rolled again. The number it now shows is also added to the total. The magician turns around. He calls attention to the fact that he has no way of knowing which of the three cubes was used for the second roll. He picks up the dice, shakes them in his hand a moment, then correctly announces the final sum.
>
> **Method:** Before the magician picks up the dice he totals their faces. Seven [the opposite sides of the dice always add to seven] added to this number gives the total obtained by the spectator. (Gardner, 1956, pp. 42-44).

Can the dice's sum really be random if the magician knows exactly what it is—as you also could, if you knew the trick? Forget about "really," I suggest, and accept that this simply is a useful way of thinking.

Later we will talk about the famous draft lottery where the balls were not well mixed and the outcomes did not all have the same probabilities, but where we still consider the lottery "fair."

Later on we shall also consider the distributions of heights of various groups of living things (including people). When we consider all living things taken together, the shape of the overall distribution—many individuals at the tiny end where the viruses are found, and very few individuals at the tall end where the giraffes are—is determined mostly by the distribution of species that have different mean heights. Hence we can explain the shape of that distribution, and we do not say that is determined by "chance." But with a homogenous cohort of a single species—say, all 25-year-old human females in the U.S.—our best description of the shape of the distribution is "chance." With situations in between, the shape is partly due to identifiable factors—e.g. age—and partly due to "chance."

Or consider the case of a basketball shooter: What causes her or him to make (or not make) a basket this shot, after a string of successes? Only chance, because the "hot hand" does not exist. But what causes a given shooter to be very good or very poor relative to other players? For that explanation we can point to such factors as the amount of practice or natural talent.

Again, all this has nothing to do with whether the mechanism is "really" chance, unlike the arguments that have been raging in physics for a century. That is the point of the ladle demonstration. Our knowledge and our power to predict the outcome gradually transits from non-chance (that is, "determined") to chance ("not determined") in a gradual way even though the same sort of physical mechanism produces each throw of the ladle.

Earlier I mentioned that when we say that chance controls the outcome of the spoon flip after (say) five revolutions, we mean that there are many small forces that affect the outcome. The effect of each force is not known, and each is independent of the other. None of these forces is large enough for me (as the spoon twirler) to deal with, or else I would deal with it and be able to improve my control and my ability to predict the outcome. This concept of many small influences—"small" meaning in practice those influences whose effects cannot be identified and allowed for—which affect the outcome and whose effects are not knowable and which are independent of each other is fundamental in statistical inference. This concept is the basis of the Theory of Errors and the Central Limit Theorem, which enable us to predict how the mean of a distribution will behave in repeated sampling from the distribution, as will be discussed later.

That is, the assumptions of the Central Limit Theorem and of the Normal distribution are the conditions that produce an event that we say is chance-like.

It is interesting to consider the relationship of this concept to the quincunx: Therein, any one ball's fate seems chance-like, but the overall distribution is determined.

## The philosophers' dispute about the concept of probability

Those who call themselves "objectivists" or "frequentists" and those who call themselves "personalists" or "Bayesians" have been arguing for hundreds or even thousands of years about the "nature" of probability. The objectivists insist (correctly) that any estimation not based on a series of observations is subject to potential bias, from which they conclude (incorrectly) that we should never think of probability that way. They are worried about the perversion of science, the substitution of arbitrary assessments for value-free data-gathering. The personalists argue (correctly) that in many situations it is not possible to obtain sufficient data to avoid considerable judgment. Indeed, if a probability is about the future, some judgment is *always* required—about which observations will be relevant, and so on. They sometimes conclude (incorrectly) that the objectivists' worries are unimportant.

As is so often the case, the various sides in the argument have different sorts of situations in mind. As we have seen, the arguments disappear if one thinks *operationally* with respect to the *purpose of the work*, rather than in terms of *properties*, as mentioned earlier. (Much more about this in my book, *The Philosophy and Practice of Statistics and Resampling).*

Here is an example of the difficulty of focusing on the supposed properties of the mechanism or situation: The mathematical theorist asserts that the probability of a die falling with the "5" side up is 1/6, on the basis of the physics of equally-weighted sides. But if one rolls a particular die a million times, and it turns up "5" less than 1/6 of the time, one surely would use the observed proportion as the practical estimate. The probabilities of various outcomes with cheap dice may depend upon the number of pips drilled out on a side. In 20,000 throws of a red die and 20,000 throws of a white die, the proportions of 3's and 4's were, respectively, .159 and .146, .145 and .142 – all far below the expected proportions of .167. That is, 3's and 4's occurred about 11 percent less often that if the dice had been perfectly formed, a difference that could make a big difference in a gambling game (Bulmer, 1979, p. 18).

It is reasonable to think of both the *engineering* method (the theoretical approach) and the *empirical* method (experimentation and data collection) as two alternative ways to estimate a probability. The two methods use different processes and different proxies for the probability you wish to estimate. One must adduce additional knowledge to decide which method to use in any given situation. It is sensible to use the empirical method when data are available. (But use both together whenever possible.)

In view of the inevitably subjective nature of probability estimates, you may prefer to talk about "degrees of belief" instead of probabilities. That's fine, just as long as it is understood that we operate with degrees of belief in exactly the same way as we operate with probabilities. The two terms are working synonyms.

Most important: One cannot sensibly talk about probabilities in the abstract, without reference to some set of facts. The topic then loses its meaning, and invites confusion and argument. This also is a reason why a general formalization of the probability concept does not make sense.

## The relationship of probability to the concept of resampling

There is no all-agreed definition of the concept of the resampling method in statistics. Unlike some other writers, I prefer to apply the term to problems in *both* pure probability and statistics. This set of examples may illustrate:

**1.** Consider asking about the number of hits one would expect from a .250 (25 percent) batter in a 400 at-bat season. One would call this a problem in "probability." The sampling distribution of the batter's results can be calculated by formula or produced by Monte Carlo simulation.

**2.** Now consider examining the number of hits in a given batter's season, and asking how likely that number (or fewer) is to occur by chance if the batter's long-run batting average is .250. One would call this a problem in "statistics." But just as in example (1) above, the answer can be calculated by formula or produced by Monte Carlo simulation. And the calculation or simulation is exactly the same as used in (1).

Here the term "resampling" might be applied to the simulation with considerable agreement among people familiar with the term, but perhaps not by all such persons.

**3.** Next consider an observed distribution of distances that a batter's hits travel in a season with 100 hits, with an observed mean of 150 feet per hit. One might ask how likely it is that a sample of 10 hits drawn with replacement from the observed distribution of hit lengths (with a mean of 150 feet) would have a mean greater than 160 feet, and one could easily produce an answer with repeated Monte Carlo samples. Traditionally this would be called a problem in probability.

**4.** Next consider that a batter gets 10 hits with a mean of 160 feet, and one wishes to estimate the probability that the sample would be produced by a distribution as specified in (3). This is a problem in statistics, and by 1996, it is common statistical practice to treat it with a resampling method. The actual simulation would, however, be identical to the work described in (3).

Because the work in (4) and (2) differ only in question (4) involving measured data and question (2) involving counted data, there seems no reason to discriminate between the two cases with respect to the term "resampling." With respect to the pairs of cases (1) and (2), and (3) and (4), there is no difference in the actual work performed, though there is a difference in the way the question is framed. I would therefore urge that the label "resampling" be applied to (1) and (3) as well as to (2) and (4), to bring out the important fact that the procedure is the same as in resampling questions in statistics.

One could easily produce examples like (1) and (2) for cases that are similar except that the drawing is without replacement, as in the sampling version of Fisher's permutation test—for example, a tea taster. And one could adduce the example of prices in different state liquor control systems (see Chapter 8) which is similar to cases (3) and (4) except that sampling without replacement seems appropriate. Again, the analogs to cases (2) and (4) would generally be called "resampling."

The concept of resampling is defined in a more precise way in Chapter 10. Fuller discussion may be found in my *The Philosophy and Practice of Statistics and Resampling* Conclusion.

## Conclusion

We define "chance" as the absence of predictive power and/or explanation and/or control.

When the spoon rotates more than three or four turns I cannot control the outcome—whether spoon or ladle end—with any accuracy. That is to say, I cannot predict much better than 50-50 with more than four rotations. So we then say that the outcome is determined by "chance."

As to those persons who wish to inquire into what the situation "really" is: I hope they agree that we do not need to do so to proceed with our work. I hope all will agree that the outcome of flipping the spoon gradually *becomes* unpredictable (random) though still subject to similar physical processes as when predictable. I do not deny *in principle* that these processes can be "understood," certainly one can develop a machine (or a baton twirler) that will make the outcome predictable for many turns. But this has nothing to do with whether the mechanism is "really" something one wants to say is influenced by "chance." This is the point of the cooking-spoon demonstration. The outcome traverses from non-chance (determined) to chance (not determined) in a smooth way even though the physical mechanism that produces the revolutions remains much the same over the traverse.

## Endnote

1. The idea that our aim is to advance our work in improving our knowledge and our decisions, rather than to answer "ultimate" questions about what is "really" true is in the same spirit as some writing about quantum theory. In 1930 Ruarck and Urey wrote: "The reader who feels disappointed that the information sought in solving a dynamic problem on the quantum theory is [only] statistical…should console himself with the thought that we seldom need any information other than that which is given by the quantum theory" (quoted by Cartright, 1987, p. 420).

2. This does not mean that I think that people should confine their learning to what they need in their daily work. Having a deeper philosophical knowledge than you ordinarily need can help you deal with extraordinary problems when they arise.