Chapter 5. Basic Concepts of Probability
Part I

"Not chaos-like, together crush'd and bruis'd,
 But, as the world, harmoniously confused
 Where order in variety we see,
 And where, tho' all things differ, all agree."
   —Pope, *Windsor Forest*

Imagine you are sitting in your statistics class one day when a man walks in and proclaims:

> "I have developed the power of mind over matter. If one of you will stand in one corner of this room and toss 100 pennies, I will stand in the opposite corner and, through the sheer power of thought, cause those penny tosses to come up heads. Now mind you, I don't claim that each and every toss will come up heads—I haven't yet perfected this skill to the level of 100 percent. But what I will do is produce an *impressive* number of heads, to the point where you will at least be willing to take my claim seriously."

And so your class decides to put him to the test. One student stands in the front left corner of the room to toss the pennies one-by-one; another is positioned nearby to record the head or tail outcome of each toss; and the man who has made the claim stands in the right rear corner of the room, attempting to exert the sheer power of his thought upon the outcomes. The rest of the class watches closely to make sure that there is no hanky-panky or collusion.

I have concocted this little scenario to show that you already have some good, solid intuitions about the generalities of probability, even though you might not yet know anything at all about its technical details. You can begin eliciting these intuitions by asking yourself the question: How many heads would have to turn up in the 100 penny tosses before I would find the results "impressive" enough to take the man's claim seriously? When I ask the students in my own statistics classes to reflect on this question, an occasional ardent skeptic will proclaim that he or she would not be impressed even if 100% of the tosses turned up as heads. Most, however, say that their threshold for being sufficiently impressed falls somewhere between 70% and 80%, that is, between 70 and 80 heads out of 100 tosses. Never in all my years before the blackboard have I found a student who would be impressed with only 50% heads, nor with anything less than 50% heads.

But of course not, you will say—obviously! And indeed it is obvious—but do pause for a moment to consider why you find it so. I expect it comes down to something like this. Somewhere within you is the idea that a standard minted coin, on any particular toss, has a 50% chance of coming up heads and a 50% chance of coming up tails. And then there is a process of inference. For some students it might take place step-by-step as a deliberate conscious process; for others it is perhaps a kind of

global intuitive leap. Either way, here is the underlying logic of it. If any particular toss of a coin has a 50% chance of coming up heads and a 50% chance of coming up tails, then in any multiplicity of coin tosses we would expect the total numbers of heads and tails outcomes to be about evenly divided, about half-and-half, 50/50. Thus you would not be impressed if our mind-over-matter claimant came up with only 50 heads out of 100 tosses, because that is just about what you would expect of any set of coin tosses, irrespective of whether anyone was there trying to zap the tosses with the sheer power of thought.

I expect you would also not be impressed by 51 heads out of 100, nor 52, nor 53. But what about 63, or 73, or 83? On the other hand, all but the most fervent skeptic would be impressed with 100 heads out of 100. But what about 95, or 85, or 75? In brief, where do you draw the line? Somewhere between 50 heads and 100 heads is a point where a rational and open-minded person could reasonably say "I am impressed." Intuitively, you know the line is there, and you probably have an intuitive hunch of its general vicinity. Let us suppose that after reflecting on this question you decide that the line falls somewhere in the vicinity of 70%. Anything less than 70 heads out of 100 will not impress you; an even 70 heads out of 100 will begin to impress you; 71 out of 100 will impress you a bit more; 72 out of 100 still more; and so on.

Now for another question, also aimed at eliciting some intuitions that you already have about the subject. Suppose that the test of the man's claim had involved only 10 penny tosses. The question is, would you in this case still draw the line at 70%? That is, would you be anywhere near as impressed with 7 heads out of 10 tosses as you would be with 70 heads out of 100 tosses? I expect your answer to this question will be an immediate "no," based on a strong underlying intuition to the effect that a 70% line could be much more easily reached or exceeded with only 10 tosses than it could be with as many as 100 tosses. Later in this chapter we will examine inferential procedures by which you can determine that the respective likelihoods for these two outcomes occurring by mere chance are

OUTCOME	LIKELIHOOD
$\geq$70% heads in 10 tosses:	17%
$\geq$70% heads in 100 tosses:	0.005%

The first of these would fall far short of the standard 5% criterion of statistical significance introduced in Chapter 4 (i.e., equal to or less than 5%), while the second would far surpass it. So if our mind-over-matter claimant actually were to get as many as 70% heads in 100 tosses, you could allow yourself to be very impressed indeed—assuming of course that the pennies were not somehow mechanically biased in favor of heads, that there was no collusion, that the outcomes of the tosses were accurately recorded, and so on. For it is very highly unlikely (5/1000ths of 1%) that he or anyone else would get as many as 70% heads in 100 tosses by mere chance coincidence.

But more of such details later. First we must lay some foundations. A couple of centuries ago the great mathematician Laplace observed that the theory of probability is "at bottom only common sense reduced to calculation." I think a better way of putting it would be to say that the theory of probability is common sense expanded by calculation—for in fact there are many aspects of the subject that go well beyond common sense, and some that are even flatly contradictory of common sense. Still, the conceptual and computational apparatus of probability does have its roots in common sense, and for that reason you will find you have a substantial head start in the task of studying it. As we begin developing the subject, please do not be lulled if you find some of the illustrative examples—coin tosses and the like—to be rather trivial. Keep your eye not on the examples, but on the general concepts that lie behind them. And bear in mind that these concepts are the foundation of all statistical inference, and thus of virtually everything else in this text that will follow.

¶Elementary Probabilities: *"common sense reduced to calculation"*

The basic concept is an idea of utter simplicity. Imagine you have four small balls, similar in every respect except that one is red and the other three are blue. Place the balls in a box, close the lid, and shake the box so as to jumble the order of its contents. Now reach into the box and blindly withdraw one ball—but before you do so, place a bet on whether the ball you draw will be red or blue. Assuming that you would want to place your bet rationally rather than just arbitrarily, the task is to determine which of the two colors has the greater chance of being drawn. And that is tantamount to asking which color has the greater probability of being drawn. The common sense of it is this. If you are blindly drawing 1 ball out of 4, and if these 4 balls include 3 blue and 1 red, then you have 3 chances out of 4 of drawing a blue ball, but only 1 chance out of 4 of drawing a red one. Thus the rational choice would be to bet on blue, for while you would still have 1 chance out of 4 of losing your bet, you would have 3 chances out of 4 of winning it. A bet on red, on the other hand, would have only 1 chance out of 4 of winning and 3 chances out of 4 of losing.

Although the calculations to which this common sense is "reduced" can sometimes grow quite complex, the basic operation is just elementary arithmetic. It amounts to taking common-sense concepts such as "3 chances out of 4" and converting them into a meaningful and useful numerical form. In general, for any common-sense concept of the form "this particular event has x chances out of y of occurring," the probability of that event can be defined numerically as the ratio of x to y. Thus the probability, **P**, of drawing a blue ball (with 3 chances out of 4) is

$\mathbf{P}_{(blue)} = 3/4 = .75$

and the probability of drawing a red ball (with 1 chance out of 4) is

$\mathbf{P}_{(red)} = 1/4 = .25$

So far, so good. Calculating that the respective probabilities are $\mathbf{P}_{(blue)} = .75$ and $\mathbf{P}_{(red)} = .25$, you rationally place your bet on drawing a blue ball. The bet is down, you blindly reach into the box and draw a ball—and it is red! The moral of this scenario is that a probability value such as $\mathbf{P}_{(blue)} = .75$ or $\mathbf{P}_{(red)} = .25$ is not a reliable predictor of the outcome of a singular event. It is a reliable predictor only in reference to a multiplicity of events; and the greater the number of such events, the more reliable the prediction. For the present example, the practical predictive meaning of the two probability values is this: If you were to perform the ball-drawing operation an indefinitely large number of times (returning the drawn ball to the box and shaking the box prior to each new draw), you would draw one or another of the blue balls in 75% of the cases and the red ball in 25% of the cases. If you were to do it only a few times, your observed percentages of blue and red would perhaps differ markedly from these theoretical values of 75% and 25%; however, the more you repeated the operation, the closer they would come.

To take another example, imagine a statistics class containing 30 students, of whom 7 are freshmen, 12 are sophomores, 10 are juniors, and one is a senior. If you were to select one student at random from this class, what is the probability that the student you select will be a freshman? As there is a total of 30 students in the class, of whom exactly 7 are freshmen, that probability is

$$\mathbf{P}_{(freshman)} = 7/30 = .2333$$

By the same reasoning, the probability of selecting a sophomore is

$$\mathbf{P}_{(sophomore)} = 12/30 = .40$$

and so on for the categories junior and senior:

$$\mathbf{P}_{(junior)} = 10/30 = .3333$$

$$\mathbf{P}_{(senior)} = 1/30 = .0333$$

These examples illustrate what is known as the **relative frequency** concept of probability, so named because it defines the probability of an event in terms of the number (or frequency) of possibilities favorable to the occurrence of that event, relative to the total number of possibilities. Thus, in its general conceptual structure, the probability of the occurrence of a certain event, x, is framed as

$$\mathbf{P}_{(x)} = \frac{\text{number of possibilities favorable to the occurrence of x}}{\text{total number of pertinent possibilities}}$$

So the probability of randomly drawing a blue ball from a box is

$$\mathbf{P}_{(blue)} = \frac{\text{the total number of blue balls in the box}}{\text{the total number of balls in the box}}$$

and the probability of randomly selecting a sophomore from a statistics class is

$$\mathbf{P}_{(sophomore)} = \frac{\text{the number of sophomores in in the class}}{\text{the total number of students in the class}}$$

Similarly, if we were to select at random one person from a room full of persons, the probability of selecting a woman would be

$$\mathbf{P}_{(woman)} = \frac{\text{the number of women in in the room}}{\text{the total number of persons in the room}}$$

It will be fairly evident that probability values determined in this fashion will always fall somewhere between $\mathbf{P}=0$ and $\mathbf{P}=1.0$, inclusive; for the operation always involves the division of one number by another number, and the divisor is in every case equal to or larger than the number it divides. Abstract though this point might seem, it makes direct contact with common sense. If there are no women in a room of 100 persons, then the probability of randomly selecting a woman from that room is nil. Hence

$$\mathbf{P}_{(woman)} = 0/100 = 0$$

Conversely, if the room with 100 persons contains *only* women, the probability of selecting a woman is what common sense would call one-hundred percent:

$$\mathbf{P}_{(woman)} = 100/100 = 1.0$$

If the number of women in the room is greater than zero but less than 100, then the probability of selecting a woman will fall somewhere in between $\mathbf{P}=0$ and $\mathbf{P}=1.0$; and the greater the proportion of women, the greater the probability. That, indeed, is exactly what a probability value is—a statement of proportion. Multiply any probability value by 100, and you convert it into a common-sense statement of percentage.

There are certain kinds of situations where the particular values for "number of possibilities favorable to the occurrence of x" and "total number of pertinent possibilities" can be precisely known in advance, either by physically counting all the relevant possibilities, or else by enumerating them through logical analysis. In a case of this sort, the resulting probability value is said to be determined **a priori**. With our

example of balls in a box, for instance, we know in advance that the box contains exactly 4 balls, of which exactly 3 are blue. Nothing else is needed. Given these prior facts, we can then proceed by "pure logic" to conclude that the probability of blindly drawing a blue ball from the box is exactly $\mathbf{P}_{(blue)} = .75$. A variation on the theme of a priori probability reasoning is illustrated by our introductory example of tossing pennies. From everything we know about the physical properties of pennies and the physical principles involved in flipping them, it is reasonable to assume that the two possible outcomes, heads and tails, are equally likely. Thus, $\mathbf{P}_{(head)} = 1/2 = .5$ and $\mathbf{P}_{(tail)} = 1/2 = .5$.

There are many other kinds of situations, however, where the appropriate probability value cannot be known precisely in advance, but rather must be estimated on the basis of observing a large number of actual instances. Here the determination is said to be **a posteriori**. This is what lurks behind the scenes when a meteorologist tells us there is a 30-percent probability of rain today, or when a physician tells a patient there is an 87-percent chance that a certain surgical procedure will be successful. For the meteorologist it is an estimate based on observations of the frequency of rain in the past under similar combinations of temperature, humidity, atmospheric pressure, and other relevant factors:

$$\mathbf{P}_{(rain)} = \frac{\text{number of previously observed similar days that produced rain}}{\text{total number of previously observed similar days}}$$

while for the physician it is an estimate based on the previously observed success rate of this particular surgical procedure for patients of this particular age, gender, physical condition, and so forth:

$$\mathbf{P}_{(success)} = \frac{\text{number of similar patients for whom the surgery proved successful}}{\text{total number of similar patients on whom the surgery was performed}}$$

I expect it is intuitively obvious to you that the confidence one can have in such a probability estimate increases in proportion to the number of observations on which it is based. We will make more of this point later in the present chapter and in other chapters as well.

End of Chapter 5, Part I.

[Return to Top of Chapter 5, Part I](#)
[Go to Chapter 5, Part 2](#)

Home	Click this link only if the present page does not appear in a frameset headed by the logo **Concepts and Applications of Inferential Statistics**