

Mathematics for Understanding Probability Theory

Probability theory is a mathematical discipline, and it is only natural that it has been developed in terms of the specialized notation and techniques of mathematics. The important thing to keep in mind is that all of the rebarbative symbolism can be translated back into the basic language that philosophers are more familiar with, though often the translation takes a great deal more space than the original statement and sometimes, if the notation is well chosen, the translation will obscure some important features that are brought out strikingly in the abbreviated form.

The following notes are designed to help philosophers as they are reading probability theory. I have tried to presuppose only a rusty recollection of highschool math.

* Fractions

- * Multiplying fractions is thrillingly simple: $a/b \times c/d = ac/bd$. This equality can be exploited in both directions. In probability, for example, we will sometimes write:

$$\frac{P(H|E)}{P(\sim H|E)} = \frac{P(H)}{P(\sim H)} \times \frac{P(E|H)}{P(E|\sim H)}$$

This equation separates out two ratios on the right-hand side, which are called the prior odds and the likelihood ratio, respectively. We arrive at it by taking the simple form of Bayes's Theorem for $P(H|E)$ and $P(\sim H|E)$:

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

$$P(\sim H|E) = \frac{P(\sim H)P(E|\sim H)}{P(E)}$$

We then divide $P(H|E)$ by $P(\sim H|E)$ and also do the corresponding division on the right. The term $P(E)$ cancels out, leaving us with

$$\frac{P(H|E)}{P(\sim H|E)} = \frac{P(H)P(E|H)}{P(\sim H)P(E|\sim H)}$$

In the final step, we separate out the respective ratios on the right-hand side.

- * Adding fractions requires that we find a common denominator. Note that in general

$$a/b + c/d \neq ac/bd,$$

a common error known as "freshman addition." The correct form is

$$a/b + c/d = (ad + bc)/bd.$$

Nevertheless, for some cases of combining evidence we may make use of something like freshman addition.

- * Sometimes we need to transpose terms in fractions. Provided that we are careful not to divide by zero, we may find the following transformations handy:

$$a/b = c \quad \Leftrightarrow \quad a/c = b \quad (\text{multiplying both sides by } b/c)$$

$$a/b = c/d \quad \Leftrightarrow \quad d/b = c/a \quad (\text{multiplying both sides by } d/a)$$

In any event, again remembering that we must restrict ourselves to cases where there is no division by zero, the following identity is important and worth remembering:

$$a/b = c/d \quad \Leftrightarrow \quad ad = bc \quad (\text{multiplying both sides by } bd)$$

- * Comparing fractions can be tricky. Although $1/2$ and $5/10$ have the same *value*, they are distinct fractions. In some contexts we will not want to reduce the latter to the former, because the magnitude of the numerator and of the denominator conveys information above and beyond the value of their ratio. For example, if we are flipping a thumbtack to see how frequently it lands point up, we might want to represent the number of successes s in a number of trials t by the fraction s/t . If we have flipped it twice and it has landed heads up once, then s/t will be $1/2$. Flipping it a third time with another point-up result, we will update this to $(s+1)/(t+1)$, which is $2/3$. On the other hand, if we have flipped it ten times with a point-up result five times, the ratio s/t will be $5/10$; and if we have a success on the next trial, the new number $(s+1)/(t+1)$ will be $6/11$, which is *not* equal to $2/3$. Context will help us to determine whether it is appropriate to reduce fractions or not.

* Intervals

- * Quite frequently it is useful to set boundaries within which a given variable must lie. Probabilities, for example, may be as low as 0 or as high as 1, but they may not exceed these bounds. There are several useful notations that help us to indicate this. We might write the foregoing fact about probabilities in either of these ways:

$$r \in [0, 1]$$

$$0 \leq r \leq 1$$

The first expression says that r is an *element* of the interval $[0, 1]$, the set of values in the number line from 0 to 1, including the end points as possible values. The second says that r is a number greater than or equal to 0 and less than or equal to 1. Obviously, these come to the same thing.

It is a natural question why we use the weak inequalities (\leq) here rather than the strong ones ($<$). The answer is that we can use the strong ones when we don't want to include the endpoints. If both zero and 1 are (in some particular case) inadmissible values, we will write:

$$r \in (0, 1)$$

$$0 < r < 1$$

The use of the rounded parentheses rather than the squared-off brackets is a convention that indicates the end points are *not* included; this is called an *open* interval, while the interval with endpoints included, written with square brackets, is called *closed*. If for some reason we wanted to include one endpoint but not the other, we could write

$$r \in [0, 1)$$

or

$$r \in (0, 1]$$

These are called *half-open* intervals. One useful thing to note about half-open intervals is that we can use them to *tile* the real number line without missing a single point, e.g. $[0, 1), [1, 2), \dots [n, n+1), \dots$

In general, we can do the same thing with real numbers a and b instead of 0 and 1, writing, for example, $[a, b]$, provided that $a \leq b$. In the limiting case where the two numbers are the same, the interval reduces to $[a, a]$, in which case we identify the interval with the real number a (or, if we are being picky about it, with $\{a\}$, the set containing a as its only member).

- * Intervals provide us with the means of indicating some important sets that have infinitely many members without having to use ellipses. We cannot enumerate all of the natural numbers, though we can *start* the enumeration and use a notation that indicates how we intend it to go on by writing $\{0, 1, 2, \dots, n, n+1, \dots\}$. But when it comes to the points in a line segment, we cannot even start the enumeration. Start with zero – what is the *very next point*? The question is misleading: there is no next point. Still, with the compact notation $[0, 6]$ we can indicate precisely which set of points we have in mind, even though they defy an orderly enumeration.

- * Special properties of 0, 1, and ∞

You should recall from elementary mathematics that adding 0 (zero) to any number leaves it unchanged, just as multiplying any number by 1 leaves it unchanged. Multiplying any number by 0 yields 0. Division by 0 is undefined. The sideways eight, called the lemniscate, is the symbol for infinity. This is not a natural number and one cannot divide or multiply by it. In general, we will see ∞ coming up in probability only as the *limit* of a fraction or a function when some pieces of the function take extreme values, not as the value of the function in any real case.

* Σ notation

The Greek capital letter sigma (Σ) denotes a *sum* or the operation of *summation*, that is, of adding things up. It is a useful abbreviation when the things to be added together are related in such a fashion that they can be compactly described by a single formula. Often the context indicates clearly enough what we are summing over, but sometimes we will attach a subscript to it (Σ_i) indicating that we are to put successive values of i (such as $i=1, i=2, \dots, i=n$) into the function that follows and sum up the results; if necessary, we might also add a superscript denoting the upper value n . Thus, the sum of the first n powers of 2 might be written $\Sigma_i 2^i$. We can think of this as a set of instructions: “Take the powers of 2, beginning with 2^1 , and sum them up.” Of course, if we don’t know when to stop, we may be baffled. And sometimes we will be told to take an *infinite* sum, usually called an *infinite series* – one with infinitely many terms to be added together. It comes as a shock to many people that such infinite sums do not always take infinite values. For example, for $i=1$ to ∞ , $\Sigma_i 1/2^i$ comes to 1. Sometimes the range of values of the “index” (which is often denoted by i but also sometimes by other letters) is explicitly noted, often by writing the lower and upper limits below and above the Σ . We might write the previous infinite series as

$$\sum_{i=1}^{\infty} \frac{1}{2^i}$$

If we wanted only the first n terms of this series, we would write:

$$\sum_{i=1}^n \frac{1}{2^i}$$

One way of writing Bayes’s Theorem looks like this:

$$P(H_1|E) = \frac{P(H_1) P(E|H_1)}{\sum_i P(H_i) P(E|H_i)}$$

Written out the long way with three hypotheses that form a *partition* – a mutually exclusive and jointly exhaustive set – it would look like this:

$$P(H_1|E) = \frac{P(H_1) P(E|H_1)}{P(H_1) P(E|H_1) + P(H_2) P(E|H_2) + P(H_3) P(E|H_3)}$$

Even for three terms, it is obvious that Σ notation saves us a great deal of space. When there are more terms involved, the notation becomes indispensable.

* Π notation

It is occasionally convenient to have a notation for products parallel to the notation for sums, and for this purpose, we use the Greek capital letter pi (Π). The Π notation permits us to represent the product of a set of terms that can be described by a single formula. Thus, in Bovens and Hartmann we encounter a formula something like this:

$$m_s(S) = \frac{P(R_1, \dots, R_m)}{\prod P(R_i)}$$

This is the probability of a conjunction divided by the product of the probabilities of the conjuncts. Written out the long way for three terms ($m=3$), the right-hand side would look like this:

$$\frac{P(R_1 \& R_2 \& R_3)}{P(R_1) \times P(R_2) \times P(R_3)}$$

* order of operations

To avoid confusion, mathematicians have imposed a definite order of operations: parentheses first, then exponents, multiplication, division, addition and subtraction. You can remember this hierarchy of operations by the silly sentence “Please Excuse My Dear Aunt Sally.”

* logarithms

A logarithm is a measure of the size of a number. You can think of it as an exponent to which a certain base is raised in order to get the number. Thus, if our base is 10, the logarithm of 1,000 is 3 – since $10^3 = 1,000$; the logarithm of 10,000,000 to the base 10 is 7. In base 10, there is a very close connection between the log of a number and the number of digits it has. If our base is 2, the logarithm of 64 is 6, since $2^6 = 64$.

One must generally specify the base when giving logarithms unless context makes this plain. The most common bases are 10 (for obvious reasons), 2 (useful for computer science applications), and Euler’s constant e , which is approximately 2.718281828. The use of Euler’s constant simplifies many calculations in higher mathematics.

Note that for any of these bases b , by definition, $b^0 = 1$.

That’s all for now. A future edition of this handout may also include discussions of:

* limits

* integrals and integration

* set theory

* measure theory

* functions and function notation