

PROBABILITY

Probability theory is central to several areas of physics, most notably quantum mechanics and statistical mechanics, so this post will give some of the basics of the theory for those who haven't encountered it before.

Central to the theory of probability is the concept of randomness, and this is a notoriously difficult notion to define. How can we tell if a number or other event is in some sense 'random'? Is 64 a random number? Is it truly a random event whether toast falls buttered side down or is Murphy's law true?

The only realistic way to cope with the definition of randomness is to define a set of probabilities for the occurrence of the events in question. We can illustrate this by considering the question above of whether 64 is a random number. Suppose we are choosing integers from the inclusive set [1...100] and we would like all numbers in this set to occur equally often. If the first number we pick turns out to be 64, this tells us nothing. If we pick 100 numbers (perhaps by picking one of a group of tiles, each of which has one of the numbers from 1 to 100 written on it, out of a bag, noting the number so picked and then replacing the tile) and 64 turns up exactly once, we might say that's what we expect, but suppose we get no 64s, or two or three of them. Does that mean that 64 isn't being selected randomly? Not necessarily. The only way we can demonstrate experimentally that 64 is occurring randomly in a way that is consistent with our hypothesis that all numbers between 1 and 100 are equally probable is to pick a very large (ideally infinite, but if we're doing the simulation on a computer, we would probably stop after several million) number of numbers and then check to see if 64 occurred around 1/100th of the time. If it did, then we would be fairly confident in stating that, yes, 64 is indeed a random number in this context.

But suppose we were dealing with a different situation in which powers of 2 were meant to be twice as likely to occur as other numbers (but we're still restricting our attention to numbers between 1 and 100). Thus the seven numbers 1, 2, 4, 8, 16, 32 and 64 should show up twice as often as the other numbers. In this case (as we'll see in a moment) we would expect 64 to occur a fraction of 2/107 of the time as opposed to a non-power of 2 which would occur 1/107th of the time. So if we got the same results as in the first experiment, then we would say, no, 64 is not occurring randomly, or at least it is evidence that the probability of its occurrence is not what we predicted.

In other words, whether or not an event is 'random' depends on the context of the problem. Before we can say whether something is random, we must first specify its probability.

So what exactly is probability? Mathematically, to define probability we must first define a set of possible events. Thus in the case of selecting a number out of a bag, we must specify which numbers are possible outcomes of each selection. Suppose we define this set to be the numbers from 1 to 100.

Having specified which events can happen, we need to then define a probability for each event in this set, but we need to do this so that it satisfies an important constraint: the sum of probabilities for all possible mutually exclusive events in the set must add up to 1. Since a probability of 1 indicates certainty, this condition is the mathematical way of saying that the event *must* be one of those specified in the original set. This is a valuable check when you are doing calculations involving probability: if the sum of probabilities for all the events in a set is *not* 1, you've done something wrong, so go back and check the definition of the set and the probability assigned to each event in the set.

Thus in the case of choosing numbers from the set [1..100] there are 100 possible events. In our first example above, where selection of all numbers is equally likely, the probabilities of choosing each number must be the same, and they must add up to 1. If we denote the probability of choosing number n as $P(n)$, then we have the conditions

$$P(n) = k \tag{1}$$

$$\sum_{n=1}^{100} P(n) = 1 \tag{2}$$

where k is a constant, independent of n . From the second equation, since all probabilities are equal, we get $100k = 1$ or $k = 0.01$.

In the second example, the powers of two are to be twice as likely as the other numbers, so we get

$$P(n) = \begin{cases} 2k & \text{if } n = \text{power of 2} \\ k & \text{otherwise} \end{cases} \tag{3}$$

$$\sum_{n \neq 2^m} k + \sum_{n=2^m} 2k = 1 \quad (4)$$

$$93k + 7(2k) = 1 \quad (5)$$

$$k = \frac{1}{107} \quad (6)$$

In the second line, there are 93 non-powers of 2 and 7 powers of 2. So the probability of getting a power of 2 is $2/107$ as stated above.

Now we have a look at a few of the basic terms in probability.

The *mean* or *average* of a set of numbers is defined as

$$\langle n \rangle \equiv \sum nP(n) \quad (7)$$

where the sum is taken over all values in the set. That is, the average value is a weighted average, where each value n is multiplied by its probability and then the whole lot added up. In the simple case where all values have the same probability, this is the same as adding up all the numbers and dividing by the size of the set. For example, in the case of the numbers from 1 to 100, if all numbers are equally likely:

$$\langle n \rangle = \sum_{n=1}^{100} nP(n) \quad (8)$$

$$= \frac{1}{100} \sum_{n=1}^{100} n \quad (9)$$

$$= \frac{1}{100} \frac{1}{2} (100)(101) \quad (10)$$

$$= 50.5 \quad (11)$$

where in the third line we've used the formula for summing up the numbers from 1 to N :

$$\sum_{n=1}^N n = \frac{1}{2} n(n+1) \quad (12)$$

In the second case, where powers of 2 are twice as likely, we get

$$\langle n \rangle = \sum_{n=1}^{100} nP(n) \quad (13)$$

$$= \frac{1}{107} \sum_{n \neq 2^m} n + \frac{2}{107} \sum_{n=2^m} n \quad (14)$$

$$= \frac{1}{107} \left(\frac{1}{2}(100)(101) - (2^7 - 1) \right) + \frac{2}{107}(2^7 - 1) \quad (15)$$

$$= \frac{1}{107} (5050 - 127 + 2 \times 127) \quad (16)$$

$$= \frac{5177}{107} \quad (17)$$

$$\approx 48.38 \quad (18)$$

where in the third line we used the formula for the first N powers of 2 (a special case of the formula for a geometric series):

$$\sum_{m=0}^N 2^m = 2^{N+1} - 1 \quad (19)$$

The mean in this case is lower than in the first case, since there are more powers of 2 less than 50 than greater than 50, thus the lower half of the number set is weighted more.

Using similar logic, we can calculate the mean of any function of n . For example, the *mean square* is the mean of the squares of n :

$$\langle n^2 \rangle = \sum n^2 P(n) \quad (20)$$

A common measure of the spread of values around the mean is given by the *variance*, defined as follows:

$$\text{var}(n) \equiv \langle n - \langle n \rangle \rangle^2 \quad (21)$$

$$= \sum (n - \langle n \rangle)^2 P(n) \quad (22)$$

$$= \sum n^2 P(n) - 2\langle n \rangle \sum nP(n) + \langle n \rangle^2 \quad (23)$$

$$= \langle n^2 \rangle - \langle n \rangle^2 \quad (24)$$

where we have used $\sum P(n) = 1$ and $\sum nP(n) = \langle n \rangle$ in the derivation.

The *standard deviation* is the (positive) square root of the variance, and is commonly denoted with a lower-case sigma σ :

$$\sigma \equiv \sqrt{\langle n^2 \rangle - \langle n \rangle^2} \quad (25)$$

For the examples above, we have, using the formula for the sum of squares: $\sum_{n=1}^N n^2 = \frac{N(N+1)(2N+1)}{6}$

$$\text{var}(n) = \frac{1}{100} 338350 - 50.5^2 \quad (26)$$

$$= 833.25 \quad (27)$$

$$\sigma = 28.87 \quad (28)$$

for the equi-probable set, and, for the set with powers of 2 weighted at twice the probability:

$$\text{var}(n) = \frac{1}{107} (338350 - 5461) + \frac{2}{107} 5461 - \left(\frac{5177}{107} \right)^2 \quad (29)$$

$$= 872.26 \quad (30)$$

$$\sigma = 29.53 \quad (31)$$

The larger variance in the second case is due to the fact that some of the numbers far from the mean (the lower powers of 2) have a higher weight, so the distribution is spread out more than in the equiprobable case.

This should give you the main ideas that you need to use probabilities of discrete events in physics. Frequently, however, we need to consider probabilities of continuous variables, such as in quantum mechanics where we consider the probability of find a particle at a given point in space. This requires the concept of a probability density function, and is discussed in the context of the wave function in quantum mechanics here.

PINGBACKS

Pingback: Buffon's needle: estimating pi