## **Notes on Probability and Probability Densities**

As we begin thinking about statistical physics it will be very useful to make sure you are comfortable with some basic concepts from probability. These concepts are involved in answering questions like these:

1. What is actually meant by "probability"?

2. How do you calculate averages, when you know the probabilities?

3. What do you do to extend the idea of probability to the situation in which the quantity being measured can actually have any one of a continuous range of possibilities, such as the x-position of a particle lying between  $x = X_1$  and  $x = X_2$ ?

4. How do you estimate how accurately a given measurement is likely to result in a value close to the average?

These notes will provide a brief description of how one goes about answering such questions. First, we need to be clear about what we mean by "probability". Imagine an experiment that can have only certain discrete outcomes. For example, flipping a balanced coin that either gives "heads" or "tails", throwing a well-made die, which results in the upper face showing a number of dots equal to either 1, 2, 3, 4, 5 or 6. We all know the chance of getting a head on any one flip is 1/2; the chance of a 3 on the die is 1/6, but what exactly is meant by this?

What we usually mean is something like this:

If we throw the coin many times, or equivalently, many people throw identical coins, the ratio of the number of heads observed n(H), divided by the total number of coin tosses N, tends to come out near 0.500 as N becomes very large. We can consider defining the probability of observing heads on any one toss as  $P(H) = \frac{\lim_{N \to \infty} n(H)}{N \to \infty}$ .

There are problems with such a definition though, if you think about it. For a well-defined limit of 0.500 to exist, means that if I pick a small number, say 0.001, you can find a total number of flips N, such that if we flip it that many times, you can guarantee me the ratio will be within  $\pm 0.001$  of the

limit 0.500. Of fact, you can't make any such guarantee, because in any one experiment of flipping the coin N times, it is possible to come out with far more heads than tails. Of course, in practice this doesn't happen very often at all. Mathematicians have actually wrestled with this problem for quite some time, and several schools of thought have held sway at various times.

For basic understanding without mathematical rigor, we will just stick with the idea that if you flip a coin  $10^6$  times, or do some equivalent experiment than can be done much more quickly, and repeat that entire process a great many times, you will find that the ratio comes out close to 0.500, mostly lying in the range from 0.499 to 0.501, i.e., with numbers as far from 0.5 as  $0.5\pm0.01$  being very rare indeed. If you flip the coin  $10^8$  times as an different experiment, you find most of the times you repeat this experiment, i.e. flip it another  $10^8$  times, the ratio lies very close to 0.5, typically in the range  $0.5\pm0.0001$ , with results lying outside the range  $0.5\pm0.001$  being quite rare. So the picture that emerges is that if you flip it enough times, it becomes very unlikely that you get a ratio very far at all from 0.5. Of course, this is rather circular reasoning, but it will have to do for us!

Another useful way to think about things like flipping a balanced coin is to state from the outset that the coin is fair and is thus exactly as likely to come up heads as tails in any one flip. If so, we can simply list all the possible outcomes of doing the N flips. To be specific, let's look as such a list for 3 flips,

 $\{\{H,H,H\},\{H,H,T\},\{H,T,H\},\{T,H,H\},\{H,T,T\},\{T,H,T\},\{T,T,H\},\{T,T,T\}\},\$ which shows all of the 8 possible outcomes. We can easily see that there is only one outcome with 3 heads, so if all outcomes are equally likely we can say the probability of getting 3 heads is 1/8. Similarly, the probability of getting 2 heads, P(2 heads) = 3/8, P(1H)=3/8, and P(0H) = 1/8. Note the important fact that P(3H)+P(2H)+P(1H)+P(0H)=1 exactly, because we are certain of getting either 0,1,2, or 3 heads; no other outcomes are possible.

If we know the probabilities for all the possible outcomes of an experiment, then we are said to know the underlying probability distribution, and with that we can make many predictions. For example, when throwing a die, each number is equally likely so we know P(n) = 1/6, n = 1,2...6, which is the underlying distribution.

Suppose we wanted to know what the average value is for the number we expect to get when throwing a single die. We can do the experiment of throwing the die many times, call it N times. We know that the average value for the number we get is defined by

 $\langle n \rangle = [1 * n(1) + 2 * n(2) + 3 * n(3) + 4 * n(4) + 5 * n(5) + 6 * n(6)] / N$ , and this can easily be rewritten as:

$$\langle n \rangle = 1 * P(1) + 2 * P(2) + 3 * P(3) + 4 * P(4) + 5 * P(5) + 6 * P(6)$$

or more simply,  $\langle n \rangle = \sum_{n=1}^{\infty} n P(n)$ .

This is a fundamental result, relating average values to the underlying probability distribution P(n). For the die we get 21/6 = 3.5, a result we will never see, but the average nevertheless! *Make sure you understand and remember this well enough to use it when you need it.* 

What if we don't just want to know the average value for n? What if we want to know the average we would get for some function of n, e.g.  $n^2$ ? This is also quite easy if we again recall the definition of average:

$$\langle n^2 \rangle = 1^2 * P(1) + 2^2 * P(2) + 3^2 * P(3) + 4^2 * P(4) + 5^2 * P(5) + 6^2 * P(6)$$

or again more simply,  $\langle n^2 \rangle = \sum_{n=1}^{6} n^2 P(n)$ .

Similarly, for any function of n, we have:

 $\langle f(n) \rangle = \sum_{n=1}^{5} f(n) P(n)$ , which is quite useful and important because, if we

know the probabilities of all possible outcomes, i.e. the underlying probability distribution P(n), we can easily calculate the average of any function of the outcomes. (Of course you should remember and understand this most basic result!) These ideas and definitions apply to all situations in which the outcomes of an experiment are distributed randomly over some set of possibilities, either because we don't know enough about the actual experiment (such as throwing the die) to actually predict the outcome, or because the outcome is by its very nature random, as is often the case for quantum systems.

Now let's tackle the problem of an experiment that can produce any one of a continuous range of results. For example we might have a particle bouncing around on the *x*-axis between x = 0 and x = 2 m, and our experiment is to measure its *x*-position. Let's suppose we have no idea of where we will find it,

i.e. all locations are equally probable. If you think about this for a while it isn't very nice. What do you think the chance of finding it to be at x = 1/2 m might be? Well there are an infinite number of real numbers between 0 and 2, so the chance of finding it at exactly any particular place is clearly zero!

We have to be careful how we phrase our question. The way out of the problem is to divide the *x*-axis into small steps of size  $\delta x$ , which can be as small as you like but are still finite. So for example, we could divide our two-meter interval into two million small pieces, each 1 micron long. If that isn't good enough for you, then use 10 million, or 100 million! OK, lets use N pieces, i.e.  $\delta x = 1/N$  meters. Now we ask a more intelligent question: what is the probability of finding the particle between some position *x* and  $x + \delta x$ , in the range where we know it to be located? Now that one we can answer; clearly it is exactly 1/N because it is equally likely to be found in any one of our N pieces.

What is the chance of locating it between 0.230 meters and 0.653 meters? Clearly the answer is the sum of all the probabilities of finding it in each of the little intervals lying in that range. This brings us to the idea of a continuous probability density function p(x), having the property that the probability  $P(x_1, x_2)$  that the particle will be found in the range between  $x_1$  and  $x_2$  is just given by  $P(x_1, x_2) = \int_{x_1}^{x_2} p(x) dx$ . Clearly in this case, we must have p(x) = 1/2, so that the probability of finding the particle between x = 0 and x= 2 m, will be 1.000 as it must be. *This idea of a probability density function to calculate probabilities for continuous variables is essential in physics and is used all the time, so be sure you think this bit through before going on.* 

For example, we might have a situation in which the probability density is a Gaussian, given by  $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-x_o)^2/2\sigma^2}$ . In this case the particle is most likely to be found where p(x) is largest, i.e. near  $x = x_o$ . Clearly the parameter  $\sigma$  describes the width of the region of relatively high probability. If you go out a distance  $2\sigma$  from the peak at  $x_o$ , the probability density is already down by a factor of  $e^2$ .

Where did the  $1/\sqrt{2\pi\sigma^2}$  come from? Well, the thing must be somewhere, so we must impose the condition that the probability of finding it somewhere on

the x-axis is exactly 1.000, i.e.  $1 \equiv \int_{-\infty}^{\infty} p(x) dx$  and to have this hold requires the factor of  $1/\sqrt{2\pi\sigma^2}$ .

How would you find the probability that the particle was located between  $x - \sigma$  and  $x + \sigma$ ? Answer: just integrate p(x) over that range,

$$P(x_{o} - \sigma \le x \le x_{o} + \sigma) = \int_{x_{o} - \sigma}^{x_{o} + \sigma} p(x) dx = \frac{1}{\sqrt{2\pi\sigma^{2}}} \int_{x_{o} - \sigma}^{x_{o} + \sigma} e^{-(x - x_{o})^{2}/2\sigma^{2}} dx$$

What if you wanted to find the average position? Just multiply p(x) by x and integrate, i.e.  $\langle x \rangle = \int x p(x) dx$ . What should the limits be for the integral? Answer, whatever is required to cover the entire range where the probability density is not zero. In this case, from negative to positive infinity.

What if you wanted the average value to expect for some function of position, f(x) = g.  $f(x) = x^2$ ? Easy, just use the same idea we already worked out for discrete distributions like coin flipping or throwing a die, and you get  $\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x) p(x) dx$ . This assumes that your probability density function is already properly normalized so that  $1 = \int_{-\infty}^{\infty} p(x) dx$ .