Statistics

David W.H. Swenson

1 Probability Distribution Functions

A probability distribution function gives a mathematical way of showing how probabilities vary for a given system. As a mathematical tool, they allow one to easily obtain various averages both for systems which only allow certain values (discrete systems) and for systems in which any value is possible (continuous systems).

1.1 Discrete PDFs

The best way to describe PDFs may be to give an example. Let's assume that we have a die with 6 sides. One side is marked 1, two sides are marked 2, and three sides are marked 3. Since there are six total sides, this means that the probability of rolling each number is as shown below:

Number Rolled	Probability
1	$1/6 \approx 0.17$
2	$2/6 = 1/3 \approx 0.33$
3	$3/6 = 1/2 = 0.50$

Table 1: Table of a Discrete PDF

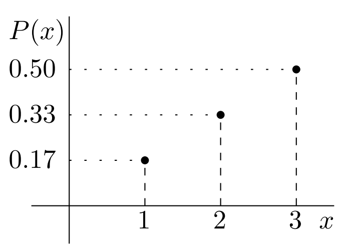

Figure 1: Graph of a Discrete PDF

Table 1 defines a probability distribution function. Since only certain values can be rolled, it is a "discrete" probability function. It can also be represented as a graph of values versus probabilities, as shown in figure 1.

Exercise 1. If you roll this die 25 times, about how many times will you expect to get each value (1, 2, and 3)?

1.2 Continuous PDFs

Although only certain values can come from rolling a die, for other systems a continuum of values (or a near continuum) can be found. For example, imagine we measured, to perfect accuracy, the height of every student at Berkeley. Normally we speak of height in discrete terms. When asked our height, we give a certain number of feet and inches. Of course, that isn't exact. Some people will then subdivide those inches into half-inches, or even quarter-inches. But even that isn't exact. In reality, no two people are exactly the same height. If you take enough decimal places, there is a difference. Therefore, there is a continuous distribution of heights.

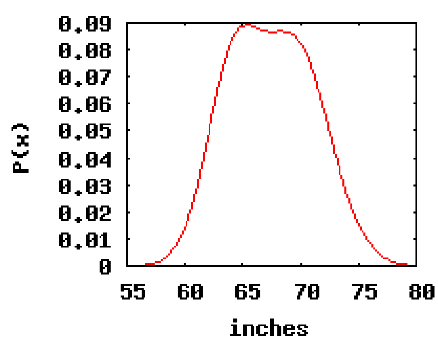

Figure 2: Continuous Probability Distribution Function for Heights

When we think of the distribution of heights as based only on the number of (whole) inches, we're thinking of it as a discrete probability function. We take all the heights in a certain range, and group them together in a "bin" that covers a certain range. This process is called "binning." It takes a continuous probability distribution function and makes a discrete probability distribution function out of it.

In science, we frequently talk about the "continuum limit." This is when we take something which is (in principle) discrete, but we assume that there are so many discrete possibilities which are so close to one another that we can treat them as an integral. Essentially, this is the opposite of the "binning" process. Mathematically, this ends up being the same as making the transformation

$$\sum_i \to \int_I \mathrm{d}x$$

where the discrete values were denoted x_i and the domain of the function is I. When dealing with probability distribution functions, the change from a discrete PDF to a continuous PDF is essentially the same process. If we could imagine some near-infinite number of experiments, which fit a certain (continuous) distribution function no matter how small we make the "buckets," then that function is the continuous PDF for that process.

One important difference between continuous PDFs and discrete PDFs is that the probabilty of having a *specific* value is always negligible. What you can determine is the probability of finding a value within a certain range. In order to find the probability between values a and b for a continuous probability distribution function $P(x)$, you use the formula

$$\text{Prob} = \int_a^b \mathrm{d}x\, P(x) \tag{1}$$

Exercise 2. The height distribution function shown above is generated from the function

$$P(x) = \frac{1}{944.998}\left(x\left(e^{-0.095(x-64)^2} + 1.1\, e^{-0.055(x-69.4)^2}\right)\right) \tag{2}$$

What is the probability that someone is shorter than 60 inches? Taller than 76 inches? Between 64 inches and 69 inches? [Hint : you'll probably want to use a calculator or computer program to do the numeric integration for you.]

Exercise 3. What would be wrong with using $P(x) = x^2$ as a probability distribution function over the domain of all real numbers? [Hint : what is the probability for a value between 0 and 2?]

Exercise 4. What would be wrong with using $P(x) = x^3$ as a probability distribution function if the domain was restricted to $[-1, 1]$? [Hint : what is the probablity for a value between -1 and 0?]

1.3 Normalization of a PDF

A probability distribution function is said to be "normalized" if the sum of all its possible results is equal to one. Physically, you can think of this as saying "we've listed every possible result, so the probability of one of them happening has to be 100%!"

For the discrete probability example in section 1.1, we implicitly normalized the probabilities. If you take all the probabilities, you will see that they add to 1. In general, you get the probability of a discrete value by doing (often implicitly) a thought experiment. Imagine a very large number (we'll call it N) of identical trials. With M possible results, we can define n_j for $j \leq M$ such that n_j is the number of trials that give the jth possible result. In this case, it is clear that

$$\sum_{j=1}^{M} n_j = N \tag{3}$$

since every trial is included in one of the n_j. This also gives us a probability for the jth possible result (p_j) of

$$p_j = \frac{n_j}{N} \tag{4}$$

Exercise 5. From equations 3 and 4, show that

$$\sum_{j=1}^{M} p_j = 1 \tag{5}$$

That is, show that the sum of all the probabilities p_j is equal to one.

For continuous PDFs, this procedure is easily generalized. Essentially we take the continuum limit of this process. So instead of setting the sum over discrete probabilities p_j (where j is determined by the value of x), we take the integral of the probability distribution function over its domain. In other words, the sum turns into an integral, the bounds of the integral come from the range over which the PDF is valid (as is true for the bounds of the sum), and the discrete probability p_j becomes the normalized PDF $P(x)$. So we have the equation

$$\int_I \mathrm{d}x\, P(x) = 1$$

where I is the interval over which the PDF is defined. Frequently, this interval is $\mathbb{R}$, the set of all real numbers.

If we're given an unnormalized PDF, we just need to divide the unnormalized PDF $p(x)$ by a normalization constant A given by

$$A = \int_I \mathrm{d}x\, p(x)$$

This gives us an important formula. Using the notation from above,

$$P(x) = \frac{p(x)}{A} \tag{6}$$

Exercise 6. Show that equation 6 implies that $\int_I \mathrm{d}x\, P(x) = 1$.

Exercise 7. What is the normalization constant A for the Gaussian PDF $p(x) = e^{-ax^2}$, defined over all space? Give your answer in terms of a. [Hint : you'll probably need to look up the definite integral.]

Exercise 8 (Advanced)**.** Show that the PDF from exercise 2 is normalized. Note that this is more an exercise in integration than in statistics. You can cheat by doing it approximately numerically, but it would good practice to actually try it by doing the integrals. By using the approximations $\int_{-\infty}^{-b} \mathrm{d}x\, e^{-ax^2} \approx 0$ and $\int_{-\infty}^{-b} \mathrm{d}x\, x\, e^{-ax^2}$ for $b > 3.5/a$ (with a and b positive), you can solve the integrals without using a computer. You'll need to remember that $\int_a^b \mathrm{d}x\, f(x) + \int_b^c \mathrm{d}x\, f(x) = \int_a^c \mathrm{d}x\, f(x)$ and $\int_a^b \mathrm{d}x\,(f(x)+g(x)) = \int_a^b \mathrm{d}x\, f(x) + \int_a^b \mathrm{d}x\, g(x)$. Remember that the bounds of normalization are $[0, \infty]$ (height can't be negative).

2 Expectation Values

2.1 Expectation Values for Discrete PDFs

An expectation value is essentially an average. Given a certain probability distribution function, it is the average value of a given function. For example, the expectation value of x is the same as the mean of a normalized probability distribution function $P(x)$. We'll start with the mean, and then generalize to the expectation value of other functions.

First, we start with the familiar definition of the mean (average) of a set of N numbers labeled $x_1, \ldots, x_N$:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$$

However, since we're looking at a discrete PDF, only certain values of x_i are possible. Let's call J the set of accepted values (in our loaded-die example from section 1.1, $J = \{1, 2, 3\}$). Let's call M the number of possible values, χ_j the jth element of J (in this case, $\chi_j = j$, and n_j the number of times χ_j occurs in the set of x_i. Note that $\sum_{j=1}^{M} n_j = N$. Now we have

$$\langle x \rangle = \frac{1}{N} \sum_{j=1}^{M} n_j \chi_j$$

Exercise 9. Imagine that in the loaded-die example from section 1.1 we rolled the die 30 times, and 5 times we got the number 1, 10 times we got the number 2, and 15 times we got the number 3. In the two formulae above, what is N? What is M? n_1? n_2? n_3? Is there any meaning to n_4? What is χ_j for $j = 2$? Do we know what x_i is for $i = 2$?

Now we just do a little bit of mathematical manipulation:

$$\begin{aligned} \langle x \rangle &= \sum_{j=1}^{M} \frac{n_j}{N} \chi_j \\ &= \sum_{j=1}^{M} p_j \chi_j \end{aligned}$$

where $p_j = n_j/N$ is the probability of having the jth result (number of times the result occurs divided by the total number of results). So of course p_j is just given by our discrete probability distribution function!

Exercise 10. For some generic function $f(x)$, show that given a discrete probability distribution function p_j, we have:

$$\langle f(x) \rangle = \sum_{j=1}^{M} p_j f(\chi_j)$$

Exercise 11. In summation notation, write out $\langle f(x) \rangle$ and $f(\langle x \rangle)$. In general, does $\langle f(x) \rangle - f(\langle x \rangle) = 0$?

2.2 Expectation Values for Continuous PDFs

Finding the expectation values for continuous PDFs just involves taking the "continuum limit" of the discrete version, *i.e.*:

$$\sum_j p_j \to \int_I \mathrm{d}x \, P(x)$$

So the expectation value of some function $f(x)$ for a continuous PDF $P(x)$ normalized over the interval I is

$$\langle f(x) \rangle = \int_I \mathrm{d}x \, P(x) f(x)$$

Exercise 12. What is the expectation value of x for the normalized Gaussian distribution function $P(x) = \sqrt{\frac{a}{\pi}} e^{-ax^2}$, defined over all space? [Hint: A Gaussian is a bell curve (this one centered at 0). So you can do the first part in your head.] Using that, what is $\langle x \rangle^2$? What about the expectation value $\left\langle x^2 \right\rangle$?

3 Variance

The variance is defined as

$$\sigma^2 = \left\langle (x - \langle x \rangle)^2 \right\rangle$$

It basically measures how far the average point is from the mean value. The square root of the variance is the standard deviation, which as you may already know, is also a measure of how "spread out" a distribution is.

Exercise 13. What is the variance of a normalized (over all space) Gaussian, $\sqrt{\frac{a}{\pi}} e^{-ax^2}$? If the Gaussian was not normalized, would that change its variance?

Exercise 14. Show that $\sigma^2 = \left\langle x^2 \right\rangle - \langle x \rangle^2$. [Hint: $\langle f(x) + g(x) \rangle = \langle f(x) \rangle + \langle g(x) \rangle$ for any functions f and g.]