Cambridge University Press
978-0-521-76542-8 - Stochastic Processes for Physicists: Understanding Noisy Systems
Kurt Jacobs
Excerpt
More information

1

A review of probability theory

In this book we will study dynamical systems driven by noise. Noise is something that changes randomly with time, and quantities that do this are called *stochastic processes*. When a dynamical system is driven by a stochastic process, its motion too has a random component, and the variables that describe it are therefore also stochastic processes. To describe noisy systems requires combining differential equations with probability theory. We begin, therefore, by reviewing what we will need to know about probability.

1.1 Random variables and mutually exclusive events

Probability theory is used to describe a situation in which we do not know the precise value of a variable, but may have an idea of the relative likelihood that it will have one of a number of possible values. Let us call the unknown quantity X. This quantity is referred to as a *random variable*. If X is the value that we will get when we roll a six-sided die, then the possible values of X are $1, 2, \ldots, 6$. We describe the likelihood that X will have one of these values, say 3, by a number between 0 and 1, called the *probability*. If the probability that $X = 3$ is unity, then this means we will *always* get 3 when we roll the die. If this probability is zero, then we will never get the value 3. If the probability is $2/3$ that the die comes up 3, then it means that we expect to get the number 3 about two thirds of the time, if we roll the die many times.

The various values of X, and of any random variable, are an example of *mutually exclusive* events. That is, whenever we throw the die, X can have only one of the values between 1 and 6, no more and no less. Rather obviously, if the probability for X to be 3 is $1/8$, and for X to be 6 is $2/8$, then the probability for X to be *either* 3 *or* 6 is $1/8 + 2/8 = 3/8$. That is, the total probability that one of two or more mutually exclusive events occurs is the *sum* of the probabilities for each event. One usually states this by saying that “mutually exclusive probabilities sum”. Thus, if

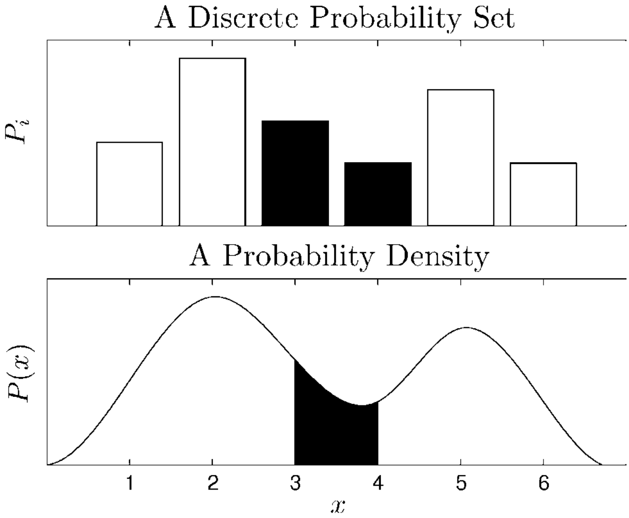

Figure 1.1. An illustation of summing the probabilities of mutually exclusive events, both for discrete and continuous random variables.

we want to know the probability for X to be in the range from 4 to 6, we sum all the probabilities for the values from 4 to 6. This is illustrated in Figure 1.1. Since X *always* takes a value between 1 and 6, the probability for it to take a value in this range must be unity. Thus, the sum of the probabilities for all the mutually exclusive possible values must always be unity. If the die is *fair*, then all the possible values are equally likely, and each is therefore equal to $1/6$.

Note: in mathematics texts it is customary to denote the unknown quantity using a capital letter, say X, and a variable that specifies one of the possible values that X may have as the equivalent lower-case letter, x. We will use this convention in this chapter, but in the following chapters we will use a lower-case letter for both the unknown quantity and the values it can take, since it causes no confusion.

In the above example, X is a *discrete random variable*, since it takes the discrete set of values $1, \ldots, 6$. If instead the value of X can be any real number, then we say that X is a *continuous* random variable. Once again we assign a number to each of these values to describe their relative likelihoods. This number is now a function of x (where x ranges over the values that X can take), called the *probability density*, and is usually denoted by $P_X(x)$ (or just $P(x)$). The probability for X to be in the range from $x = a$ to $x = b$ is now the area under $P(x)$ from $x = a$ to $x = b$. That is

$$\text{Prob}(a < X < b) = \int_a^b P(x)dx. \tag{1.1}$$

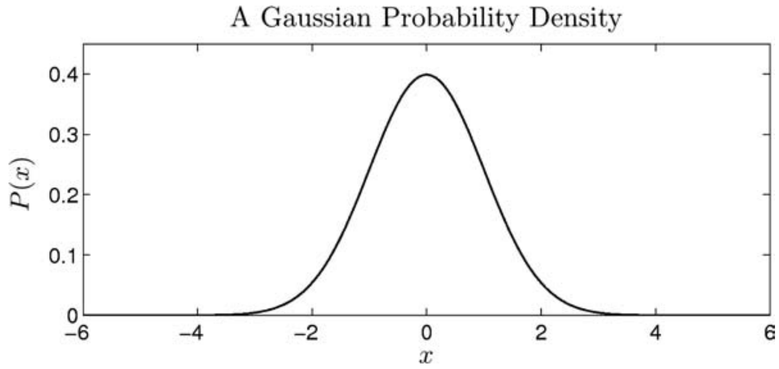

Figure 1.2. A Gaussian probability density with variance $V = 1$, and mean $\langle X \rangle = 0$.

This is illustrated in Figure 1.1. Thus the integral of $P(x)$ over the whole real line (from $-\infty$ to ∞) must be 1, since X must take one of these values:

$$\int_{-\infty}^{\infty} P(x)dx = 1. \tag{1.2}$$

The average of X, also known as the *mean*, or *expectation value*, of X is defined by

$$\langle X \rangle \equiv \int_{-\infty}^{\infty} P(x)x\, dx. \tag{1.3}$$

If $P(x)$ is symmetric about $x = 0$, then it is not difficult to see that the mean of X is zero, which is also the center of the density. If the density is symmetric about any other point, say $x = a$, then the mean is also a. This is clear if one considers a density that is symmetric about $x = 0$, and then shifts it along the x-axis so that it is symmetric about $x = a$: shifting the density shifts the mean by the same amount.

The *variance* of X is defined as

$$V_X \equiv \int_{-\infty}^{\infty} P(x)(x - \langle X \rangle)^2\, dx = \int_{-\infty}^{\infty} P(x)x^2\, dx - \langle X \rangle^2 = \langle X^2 \rangle - \langle X \rangle^2. \tag{1.4}$$

The *standard deviation* of X, denoted by σ_X and defined as the square root of the variance, is a measure of how broad the probability density for X is – that is, how much we can expect X to deviate from its mean value.

An important example of a probability density is the Gaussian, given by

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{1.5}$$

The mean of this Gaussian probability density is $\langle X \rangle = \mu$ and the variance is $V(x) = \sigma^2$. A plot of this probability density in given in Figure 1.2.

1.2 Independence

Two random variables are referred to as being *independent* if neither of their probability densities depends on the value of the other variable. For example, if we rolled our six-sided die two times, and called the outcome of the first roll X, and the outcome of the second roll Y, then these two random variables would be independent. Further, we speak of the event $X = 3$ (when the first die roll comes up as 3) and the event $Y = 6$ as being independent. When two events are independent, the probability that both of them occur (that $X = 3$ *and* $Y = 6$) is the *product* of the probabilities that each occurs. One often states this by saying that "independent probabilities multiply". The reason for this is fairly clear if we consider first making the die roll to obtain X. Only if $X = 3$ do we then make the second roll, and only if that comes up 6 do we get the result $X = 3$ and $Y = 6$. If the first roll only comes up 3 one eighth of the time, and the second comes up 6 one sixth of the time, then we will only get both of them $1/8 \times 1/6 = 1/48$ of the time.

Once again this is just as true for independent random variables that take a continuum of values. In this case we speak of the "joint probability density", $P(x, y)$, that X is equal to x and Y is equal to y. This joint probability density is the product of the probability densities for each of the two independent random variables, and we write this as $P(x, y) = P_X(x)P_Y(y)$. The probability that X falls within the interval $[a, b]$ *and* Y falls in the interval $[c, d]$ is then

$$\begin{aligned}\text{Prob}(X \in [a, b] \text{ and } Y \in [c, d]) &= \int_a^b \int_c^d P(x, y)dydx \\ = \int_a^b \int_c^d P_X(x)P_Y(y)dydx &= \left(\int_a^b P_X(x)dx\right)\left(\int_c^d P_Y(y)dy\right) \\ &= \text{Prob}(X \in [a, b]) \times \text{Prob}(Y \in [c, d]).\end{aligned}$$

In general, if we have a joint probability density, $P(x_1, \ldots, x_N)$, for the N variables $X_1, \ldots, X_N$, then the expectation value of a function of the variables, $f(X_1, \ldots, X_N)$, is given by integrating the joint probability density over all the variables:

$$\langle f(X_1, \ldots, X_N)\rangle = \int_{-\infty}^{\infty} f(x_1, \ldots, x_N)P(x_1, \ldots, x_N)\, dx_1 \ldots dx_N. \quad (1.6)$$

It is also worth noting that when two variables are independent, then the expectation value of their product is simply the product of their individual expectation values. That is

$$\langle XY\rangle = \langle X\rangle\langle Y\rangle. \quad (1.7)$$

1.3 Dependent random variables

Random variables, X and Y, are said to be *dependent* if their joint probability density, $P(x, y)$, does not factor into the product of their respective probability densities.

To obtain the probability density for one of the variables alone (say X), we integrate the joint probability density over all values of the other variable (in this case Y). This is because, for each value of X, we want to know the total probability summed over all the mutually exclusive values that Y can take. In this context, the probability densities for the single variables are referred to as the *marginals* of the joint density.

If we know nothing about the value of Y, then our probability density for X is just the marginal

$$P_X(x) = \int_{-\infty}^{\infty} P(x, y)dy. \tag{1.8}$$

If X and Y are dependent, and we learn the value of Y, then in general this will change our probability density for X (and vice versa). The probability density for X *given* that we know that $Y = y$, is written $P(x|y)$, and is referred to as the *conditional* probability density for X *given* Y.

To see how to calculate this conditional probability, we note first that $P(x, y)$ with $y = a$ gives the *relative* probability for different values of x given that $Y = a$. To obtain the conditional probability density for X given that $Y = a$, all we have to do is divide $P(x, a)$ by its integral over all values of x. This ensures that the integral of the conditional probability is 1. Since this is true for any value of y, we have

$$P(x|y) = \frac{P(x, y)}{\int_{-\infty}^{\infty} P(x, y)dx}. \tag{1.9}$$

Note also that since

$$P_Y(y) = \int_{-\infty}^{\infty} P(x, y)dx, \tag{1.10}$$

if we substitute this into the equation for the conditional probability above (Eq. (1.9)) we have

$$P(x|y) = \frac{P(x, y)}{P_Y(y)}, \tag{1.11}$$

and further that $P(x, y) = P(x|y)P_Y(y)$.

As an example of a conditional probability density consider a joint probability density for X and Y, where the probability density for Y is a Gaussian with zero mean, and that for X is a Gaussian whose mean is given by the value of Y. In this case X and Y are not independent, and we have

$$P(x, y) = P(x|y)P(y) = \frac{e^{-(1/2)(x-y)^2}}{\sqrt{2\pi}} \times \frac{e^{-(1/2)y^2}}{\sqrt{2\pi}} = \frac{e^{-(1/2)(x-y)^2-(1/2)y^2}}{2\pi}, \tag{1.12}$$

where we have chosen the variance of Y, and of X given Y to be unity. Generally, when two random variables are dependent, $\langle XY\rangle \neq \langle X\rangle\langle Y\rangle$.

1.4 Correlations and correlation coefficients

The expectation value of the product of two random variables is called the *correlation* of the two variables. The reason that we call this quantity a *correlation* is that, if the two random variables have zero mean and fixed variance, then the larger the value of the correlation, the more the variables tend to fluctuate *together* rather than independently; that is, if one is positive, then it is more likely that the other is positive. The value of the correlation therefore indicates how *correlated* the two variables are.

Of course, if we increase the variance of either of the two variables then the correlation will also increase. We can remove this dependence, and obtain a quantity that is a clearer indicator of the mutual dependence between the two variables by dividing the correlation by $\sqrt{V(X)V(Y)}$. This new quantity is called the correlation *coefficient* of X and Y, and is denoted by C_{XY}:

$$C_{XY} \equiv \frac{\langle XY\rangle}{\sqrt{V(X)V(Y)}}. \tag{1.13}$$

If the means of X and Y are not zero, then we can remove these when we calculate the correlation coefficient, so as to preserve its properties. Thus, in general, the correlation coefficient is defined as

$$C_{XY} \equiv \frac{\langle (X-\langle X\rangle)(Y-\langle Y\rangle)\rangle}{\sqrt{V(X)V(Y)}} = \frac{\langle XY\rangle - \langle X\rangle\langle Y\rangle}{\sqrt{V(X)V(Y)}}. \tag{1.14}$$

The quantity on the top line, $\langle XY\rangle - \langle X\rangle\langle Y\rangle$ is called the *covariance* of X and Y, and is zero if X and Y are independent. The correlation coefficient is therefore zero if X and Y are independent (completely uncorrelated), and is unity if $X = cY$, for some positive constant c (perfect correlation). If $X = -cY$, then the correlation coefficient is -1, and we say that the two variables are perfectly *anti-correlated*. The correlation coefficient provides a rough measure of the mutual dependence of two random variables, and one which is relatively easy to calculate.

1.5 Adding random variables together

When we have two continuous random variables, X and Y, with probability densities P_X and P_Y, it is often useful to be able to calculate the probability density of the random variable whose value is the sum of them: $Z = X + Y$. It turns out that the probability density for Z is given by

$$P_Z(z) = \int_{-\infty}^{\infty} P_X(s - z)P_Y(s)ds \equiv P_X * P_Y, \tag{1.15}$$

which is called the *convolution* of P_X and P_Y [1]. Note that the convolution of two functions, denoted by "$*$", is another function. It is, in fact, quite easy to see directly why the above expression for $P_Z(z)$ is true. For Z to equal z, then if $Y = y$, X must be equal to $z - y$. The probability (density) for that to occur is $P_Y(y)P_X(z - y)$. To obtain the total probability (density) that $Z = z$, we need to sum this product over all possible values of Y, and this gives the expression for $P_Z(z)$ above.

It will be useful to know the mean and variance of a random variable that is the sum of two or more random variables. It turns out that if $X = X_1 + X_2$, then the mean of X is

$$\langle X\rangle = \langle X_1\rangle + \langle X_2\rangle, \tag{1.16}$$

and if X_1 and X_2 are independent, then

$$V_X = V_{X_1} + V_{X_2}. \tag{1.17}$$

That is, when we add independent random variables both the means and variances add together to give the mean and variance of the new random variable. It follows that this remains true when we add any number of independent random variables together, so that, for example, $\langle \sum_{n=1}^{N} X_n\rangle = \sum_{n=1}^{N}\langle X_n\rangle$.

If you have ever taken an undergraduate physics lab, then you will be familiar with the notion that averaging the results of a number of independent measurements produces a more accurate result. This is because the variances of the different measurement results add together. If all the measurements are made using the same method, we can assume the results of all the measurements have the same mean, μ, and variance, V. If we average the results, X_n, of N of these independent measurements, then the mean of the average is

$$\mu_{\text{av}} = \left\langle \sum_{n=1}^{N} \frac{X_n}{N} \right\rangle = \sum_{n=1}^{N} \frac{\mu}{N} = \mu. \tag{1.18}$$

But because we are dividing each of the variables by N, the variance of each goes down by $1/N^2$. Because it is the variances that add together, the variance of the sum is

$$V_{\text{av}} = V\left[\sum_{n=1}^{N} \frac{X_n}{N}\right] = \sum_{n=1}^{N} \frac{V}{N^2} = \frac{V}{N}. \tag{1.19}$$

Thus the variance gets smaller as we add more results together. Of course, it is not the variance that quantifies the uncertainty in the final value, but the standard deviation. The standard deviation of each measurement result is $\sigma = \sqrt{V}$, and hence the standard deviation of the average is

$$\sigma_{\text{av}} = \sqrt{\frac{V}{N}} = \frac{\sigma}{\sqrt{N}}. \tag{1.20}$$

The accuracy of the average therefore increases as the square root of the number of measurements.

1.6 Transformations of a random variable

If we know the probability density for a random variable X, then it can be useful to know how calculate the probability density for a random variable, Y, that is some function of X. This is referred to as a *transformation* of a random variable because we can think of the function as transforming X into a new variable Y. Let us begin with a particularly simple example, in which Y is a linear function of X. This means that $Y = aX + b$ for some constants a and b. In this case it is not that hard to see the answer directly. Since we have multiplied X by a, the probability density will be stretched by a factor of a. Then adding b will shift the density by b. The result is that the density for Y is $Q(Y) = P(y/a - b/a)/a$.

To calculate the probability density for $Y = aX + b$ in a more systematic way (which we can then use for much more general transformations of a random variable) we use the fact that the probability density for Y determines the average value of a function of Y, $f(Y)$, through the relation

$$\langle f(Y)\rangle = \int_{-\infty}^{\infty} P(y)f(y)dy. \tag{1.21}$$

Now, since we know that $Y = g(X) = aX + b$, we also know that

$$\langle f(Y)\rangle = \int_{-\infty}^{\infty} P(x)f(y)dx = \int_{-\infty}^{\infty} P(x)f(ax+b)dx. \tag{1.22}$$

Changing variables in the integral from x to y we have

$$\langle f(Y)\rangle = \int_{-\infty}^{\infty} P(x)f(ax+b)dx = \frac{1}{a}\int_{-\infty}^{\infty} P(y/a - b/a)f(y)dy. \tag{1.23}$$

Thus the density for Y is

$$Q(y) = \frac{1}{a} P(y/a - b/a). \tag{1.24}$$

In addition, it is simple to verify that $\langle Y \rangle = a \langle X \rangle + b$ and $V_Y = a^2 V_X$.

More generally, we can derive an expression for the probability density of Y when Y is an arbitrary function of a random variable. If $Y = g(X)$, then we determine the probability density for Y by changing variables in the same way as above. We begin by writing the expectation value of a function of Y, $f(Y)$, in terms of $P(x)$. This gives

$$\langle f(Y) \rangle = \int_{x=a}^{x=b} P(x) f(g(x)) dx, \tag{1.25}$$

where a and b are, respectively, the lower and upper limits on the values that X can take. Now we transform this to an integral over the values of Y. Denoting the inverse of the function g as g^{-1}, so that $X = g^{-1}(Y)$, we have

$$\begin{aligned} \langle f(Y) \rangle = \int_{x=a}^{x=b} P(x) f(g(x)) dx &= \int_{y=g(a)}^{y=g(b)} P(g^{-1}(y)) \left(\frac{dx}{dy} \right) f(y) dy \\ = \int_{y=g(a)}^{y=g(b)} \frac{P(g^{-1}(y))}{g'(x)} f(y) dy &= \int_{y=g(a)}^{y=g(b)} \frac{P(g^{-1}(y))}{g'(g^{-1}(y))} f(y) dy. \end{aligned} \tag{1.26}$$

We now identify the function that multiplies $f(y)$ inside the integral over y as the probability density for Y. But in doing so we have to be a little bit careful. If the lower limit for y, $g(a)$, is *greater* than the upper limit for y, then the probability density we get will be negative to compensate for this inversion of the integral limits. So the probability density is actually the absolute value of the function inside the integral. The probability density for y is therefore

$$Q(y) = \frac{P(g^{-1}(y))}{|g'(g^{-1}(y))|}. \tag{1.27}$$

One must realize also that this expression for $Q(y)$ only works for functions that map a single value of x to a single value of y (invertible functions), because in the change of variables in the integral we assumed that g was invertible. For non-invertible functions, for example $y = x^2$, one needs to do the transformation of the integral on a case-by-case basis to work out $Q(y)$.

1.7 The distribution function

The *probability distribution function*, which we will call $D(x)$, of a random variable X is defined as the probability that X is less than or equal to x. Thus

$$D(x) = \text{Prob}(X \le x) = \int_{-\infty}^{x} P(z)\, dz. \tag{1.28}$$

In addition, the fundamental theorem of calculus tells us that

$$P(x) = \frac{dD(x)}{dx}. \tag{1.29}$$

1.8 The characteristic function

Another very useful definition is that of the *characteristic* function, $\chi(s)$. The characteristic function is defined as the *Fourier transform* of the probability density. Thus before we discuss the characteristic function, we need to explain what the Fourier transform is. The Fourier transform of a function $P(x)$ is another function given by

$$\chi(s) = \int_{-\infty}^{\infty} P(x)e^{isx} dx. \tag{1.30}$$

The Fourier transform has many useful properties. One of them is the fact that it has a simple inverse, allowing one to perform a transformation on $\chi(s)$ to get back $P(x)$. This inverse transform is

$$P(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \chi(s)e^{-isx} ds. \tag{1.31}$$

Another very useful property is the following. If we have two functions $F(x)$ and $G(x)$, then the Fourier transform of their convolution is simply the *product* of their respective Fourier transforms! This can be very useful because a product is always easy to calculate, but a convolution is not. Because the density for the sum of two random variables in the convolution of their respect densities, we now have an alternate way to find the probability density of the sum of two random variables: we can either convolve their two densities, or we can calculate the characteristic functions for each, multiply these together, and then take the inverse Fourier transform.

Showing that the Fourier transform of the convolution of two densities is the product of their respective Fourier transforms is not difficult, but we do need to use the Dirac δ-function, denoted by $\delta(x)$. The Dirac δ-function is zero everywhere except at $t = 0$, where it is infinite. It is defined in such a way that it integrates to