2 Probability and statistics

Topics: Random variables, probability distributions, mean, variance, counting, Gaussian distribution, delta function

2.1 Motivation and approach

The field of statistical mechanics, which connects the microscopic description of a system to macroscopic observables, is intricately tied with probability and statistics. When describing a large number (think Avogadro's number) of particles, it's not important (and too difficult) to know the details of each particle, so we often think of them as a statistical ensemble, or a probability distribution, over all possible states of the system (*i.e.* "microstates"). The mean and variance of this distribution correspond to observables like total energy, volume, temperature, or pressure.

Quantum mechanics also takes on a probabilistic view – predictions made by quantum mechanics are statistical by nature. Without getting too philosophical, observed quantities in quantum mechanics can be interpreted as the result of measuring an ensemble of infinite, identically-prepared systems.

Finally, statistics is important in analyzing and understanding any experimental data. The world around us is made up of random variables, and any time we measure these random variables or make predictions about how they will behave in the future, we are making use of probability and statistics.

This section describes random variables and probability distributions, providing formulae for computing means and variances, which are measures of probability distributions. It also introduces two distributions, the Gaussian distribution and delta function, which are commonly encountered in physics, and basic concepts of counting and combinatorics.

2.2 Random variables and probability distributions

Conceptual video: Random variables by Khan Academy (4:31)

Conceptual video: Density curves by Khan Academy (9:33)

A **random variable** is a variable whose values depend on the outcome of random events. The **probability distribution** of a random variable is the function that gives the probabilities that possible outcomes actually occur. Random variables can either be discrete (*i.e.* the result of a coin toss) or continuous (*i.e.* the amount of time you end up waiting for the bus).

If $p(x)$ is the **probability density function** of a continuous random variable x, then $p(x) \geq 0$ at all points x, as we know there cannot be negative probabilities! The probability that x is between the values a and b is given by

$$p(a < x < b) = \int_a^b p(x)dx\,, \tag{2.1}$$

and this probability should always be between 0 and 1.

Finally, all probability distributions should be **normalized**,

$$\int_{-\infty}^{\infty} p(x)dx = 1\,, \tag{2.2}$$

since we know that the probability of x being any value between negative and positive infinity should be 100%.

For a discrete random variable with **probability mass function** P, the probability that x is equal to a is given by $P(x = a)$ and the normalization condition is given by $\sum_i P(x_i) = 1$.

2.2.1 Mean and variance

Conceptual and technical video: Deriving the mean and variance of a continuous probability distribution by JB Statistics (7:21)

The **mean** or **expected value** of a random variable X is a value that captures the center of the random variable's distribution. If you perform many random events that come from the same probability distribution, the average value of all the outcomes is represented by the mean.

The expectation value $E[X]$ is often denoted as μ and can be computed as

$$\mu = E[X] = \int xp(x)dx \tag{2.3}$$

$$\mu = E[X] = \sum_i x_i P(x_i) \tag{2.4}$$

for continuous and discrete random variables, respectively.

The expectation is a linear operator, so

$$E[aX + b] = aE[X] + b\,. \tag{2.5}$$

The expected value of a constant is itself (*i.e.* $E[b] = b$) since the constant is not random and its value does not vary.

The **variance** measures the spread of a distribution, or the magnitude of fluctuations that occur around the mean. The variance is often denoted as σ^2 and can be computed as

$$\sigma^2 = \mathrm{Var}[X] = E\big[(X - \mu)^2\big] = E[X^2] - \mu^2\,. \tag{2.6}$$

For continuous and discrete random variables, this becomes

$$\sigma^2 = \int x^2 p(x)dx - \left(\int xp(x)dx \right)^2 \tag{2.7}$$

$$\sigma^2 = \sum_i x_i^2 P(x_i) - \left(\sum_i x_i P(x_i) \right)^2 . \tag{2.8}$$

Note that, unlike the expectation, the variance is not a linear operator. It does, however, follow the general rule:

$$\mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X] . \tag{2.9}$$

The square root of the variance is called the **standard deviation** and is often denoted as σ.

Example 2.1: Compute the mean and variance of a random variable X that is distributed according to the Bernoulli distribution, which is a discrete probability distribution with

$$P(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases} , \tag{2.10}$$

where p is a parameter.

2.2.2 Multivariate distributions

A **multivariate distribution** or **joint distribution** is a distribution of multiple random variables. The measure of **correlation** between two random variables X and Y is given by

$$\sigma_{XY} = \langle XY \rangle - \langle X \rangle \langle Y \rangle . \tag{2.11}$$

Two random variables are said to be uncorrelated if $\langle XY \rangle = \langle X \rangle \langle Y \rangle$ so that $\sigma_{XY} = 0$. In this case, the joint probability distribution is simply a product of each probability distribution:

$$p(x, y) = p(x)p(y) . \tag{2.12}$$

In physics, we see that two things separated by shorter distances or timescales tend to be more correlated than things that are far away from each other in space or time.

To obtain the **marginal distribution** from the joint distribution, simply integrate over one of the random variables:

$$p(x) = \int p(x, y)dy . \tag{2.13}$$

2.3 Special distributions

2.3.1 Gaussian distribution

Conceptual video: Sampling distribution of the sample mean by Khan Academy (10:51)

Technical video: Standard error of the mean by Khan Academy (15:14)

Conceptual and technical video: Central limit theorem by Khan Academy (9:48)

The **Gaussian distribution** (*i.e.* **normal distribution** or **bell curve**) is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{2.14}$$

Why does the Gaussian distribution come up literally everywhere? The answer is the **central limit theorem**.

Often, we have a random variable for which we don't know the probability distribution, but we want to use data to estimate its mean and variance.

For example, we can conduct an experiment n times, the results of which are independent and identically distributed random variables: $X_1, X_2, \ldots, X_n$. From there, we can compute the **sample mean** as:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i\,. \tag{2.15}$$

The sample mean $\bar{X}_n$ is itself a random variable that has $E(\bar{X}_n) = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, where μ and σ^2 are the mean and variance of the underlying distribution from which the data X_i were drawn.

When plotting data corresponding to experiments we've repeated multiple times under the same conditions, we tend to use a point (the sample mean) and errorbars (the sample variance), which highlight both the result of our experiments and the uncertainty in our measurements.

According to the **law of large numbers**, the sample mean $\bar{X}_n$ converges to the true mean μ as $n \to \infty$. What's more, the central limit theorem says that for large n, the distribution of $\bar{X}_n$ approaches a Gaussian distribution with μ and variance σ^2/n, regardless of the underlying distribution for each of the X_i.

Conceptual and technical video: Why does pi show up here? The Gaussian integral, explained by vcubingx (5:45)

The Gaussian distribution can also be very helpful for integration. If you have a function of the form e^{x^2}, you know that it can be manipulated to match the exponential in Eq. (2.14) and then use the property of normalization to solve the integral.

The definite integral of any arbitrary Gaussian function is

$$\int_{-} \infty^{\infty} e^{-a(x+b)^2}\,dx = \sqrt{\frac{\pi}{a}}\,. \tag{2.16}$$

Example 2.2: Evaluate the following integral:

$$I = \int_0^{\infty} e^{-x^2/\sigma^2}\,dx\,. \tag{2.17}$$

Wikipedia has a very useful list of Gaussian integrals that is good to keep handy.

2.3.2 Delta function

Conceptual video: Dirac delta function by Khan Academy (17:47)

The delta function is a weird "function." Consider first the **Kronecker delta**:

$$\delta_{nm} = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{if } n \neq m \end{cases} . \tag{2.18}$$

This is normalized (*i.e.* $\sum_m \delta_{nm} = 1$), so it can be thought of as a discrete probability distribution that picks out a particular value.

The continuous version of this distribution is the **Dirac delta function**:

$$\delta(x - x') = \begin{cases} \infty & \text{if } x = x' \\ 0 & \text{otherwise} \end{cases} . \tag{2.19}$$

The Dirac delta function is also normalized ($\int_{-\infty}^{\infty} \delta(x - x')dx = 1$), and it can also be used to pick out particular values:

$$\int f(x)\delta(x - x')dx = f(x') . \tag{2.20}$$

For this reason, the Dirac delta function is commonly used in physics to impose certain constraints, like the position of a particle or the total energy of a system.

The Dirac delta function comes up in other contexts, such as a Gaussian function with the variance tending to 0, in the Fourier transform of constants, and as the derivative of the Heaviside function.

2.4 Counting

Conceptual video: Introduction to combinations by Khan Academy (6:17)

For an experiment in which all outcomes are equally likely, the probability of an event occurring is given by

$$p = \frac{\text{number of outcomes satisfying event}}{\text{number of all possible outcomes}} . \tag{2.21}$$

Example 2.3: When considering a fair six-sided die, what is the probability of rolling an even number?

Sometimes, determining the numerator and denominator in the above probability expression can be challenging.

Often, we will need to figure out how many ways there are to order n objects. There are n ways to choose the first object, $(n - 1)$ ways to choose the second object, . . . , and 1 way to choose the last object. We can multiply all of these ways to obtain $n \cdot (n-1) \cdot \dots \cdot 1 = n!$ ("n factorial") ways.

Other times, we will need to figure out how many ways there are to pick k objects from n total objects. If order matters, there are n ways to choose the first object, $(n-1)$ ways to choose the second object, …, and $n-(k-1)$ ways to choose the k^{th} and final object. This means there are $n \cdot (n-1) \cdot \dots \cdot n-(k-1) = n!/(n-k)!$ ways.

If order doesn't matter, then we can divide the above expression by the number of ways to choose those k objects ($k!$ ways) so we have $\binom{n}{k} = \binom{n}{n-k} = \frac{n!}{(n-k)!k!}$ ("n choose k") ways.

Example 2.4: How many ways can n distinguishable particles be placed in k distinguishable boxes? Assume there are no restrictions on the number of particles that can be in each box.

2.5 Connections to physical chemistry

As noted in the Introduction and Motivation, both the fields of statistical mechanics and quantum mechanics are probabilistic in nature.

Counting and combinatorics can be helpful in determining the number of microstates of a system, *i.e.* the number of configurations of an ensemble of particles.

The probability that a macroscopic system will be in a certain microstate is given by a probability distribution. The specific probability distribution depends on the physical circumstances of the problem, but one of the most common is the Boltzmann probability distribution, for which the probability that the system is in microstate i with energy ε_i is given by $p_i \propto e^{-\beta\varepsilon_i}$, where β is related to the temperature of the system. This probability distribution can be used to calculate expectation values (the mean) and fluctuations (the variance) of observables like total energy, volume, temperature, or pressure.

In quantum mechanics, probability distributions appear frequently. One common example is that the square modulus of a particle's normalized wavefunction $|\psi(x)|^2$ gives the probability distribution of position measurements of that particle.

2.6 Additional text resources

Seeing Theory: A visual introduction to probability and statistics by Kunin, Guo, Devlin, and Xiang: Chapters 1 and 3

2.7 Example problem solutions

Example 2.1: *Compute the mean and variance of a random variable X that is distributed according to the Bernoulli distribution, which is a discrete probability distribution with*

$$P(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}, \tag{2.22}$$

where p is a parameter.

We can compute the mean directly via Eq. (2.4):

$$\mu = \sum_i x_i P(x_i) \tag{2.23}$$
$$= 0 \cdot P(0) + 1 \cdot P(1) \tag{2.24}$$
$$= p\,. \tag{2.25}$$

Similarly, we can use this result along with Eq. (2.8) to compute the variance:

$$\sigma^2 = E[X^2] - \mu^2 \tag{2.26}$$
$$= \sum_i x_i^2 P(x_i) - \mu^2 \tag{2.27}$$
$$= 0^2 \cdot P(0) + 1^2 \cdot P(1) - \mu^2 \tag{2.28}$$
$$= p - p^2\,. \tag{2.29}$$

Example 2.2: *Evaluate the following integral:*

$$I = \int_0^\infty e^{-x^2/\sigma^2} dx\,. \tag{2.30}$$

First, it's important to notice that the integrand has the form of a Gaussian function centered at 0. Next, we notice that the limits of integration are from 0 to ∞. Since this Gaussian function is symmetric about its center at $x = 0$, this integral is half of that from $-\infty$ to ∞. Finally, we can use the result from Eq. (2.16) and divide by 2 to obtain

$$I = \frac{1}{2}\sqrt{\sigma^2 \pi}\,. \tag{2.31}$$

Example 2.3: *When considering a fair six-sided die, what is the probability of rolling an even number?*

Since all outcomes are equally likely, the probability of rolling an even number is the ratio of the number of outcomes resulting in an even roll to the number of all possible outcomes: $p = 3/6 = 1/2$.

Example 2.4: *How many ways can n distinguishable particles be placed in k distinguishable boxes? Assume there are no restrictions on the number of particles that can be in each box.*

First, we must ask ourselves if order matters. Since the particles are distinguishable, order does matter – we can tell which particles are in which boxes! The first particle can be arranged in k ways. Since there are no restrictions on the number of particles in each box, the second particle can also be arranged in k ways, and so on, for each of the n particles. Since each particle's box assignment is independent of all the others', we can multiply these n values to obtain k^n ways.

To consider further... Think about how this solution would be different if the particles were indistinguishable. In classical mechanics, identical particles are distinguishable, while they are indistinguishable in quantum mechanics. This difference has an implication in probabilities of events in classical versus quantum mechanics. Also, consider how this solution would be different if we put restrictions on the numbers of particles allowed in each box.