Chapter 2

Statistical Descriptions of Systems

2.1 Quantitative Descriptions of Distributions

The mean of a series of N measurements of a variable x (*i.e.*, $x_1, x_2, \cdots, x_N$) is determined using the following equation:

$$\overline{x} = \left(\frac{1}{N}\right) \sum_{n=1}^{N} x_n$$

The mean of x can also be expressed in terms of the probability distribution for x, denoted as $p(x_n)$.

$$\overline{x} = \frac{\sum_{n=1}^{N} p(x_n) x_n}{\sum_{n=1}^{N} p(x_n)} \tag{2.1}$$

This expression can be made more general by writing the summations as integrals with the change of $p(x_n) \rightarrow p(x)$.

$$\overline{x} = \frac{\int x\, p(x)\, dx}{\int p(x)\, dx} \tag{2.2}$$

We can then generalize this discussion further by recognizing that $\overline{x}$ can be expressed as a moment of the probability distribution $p(x)$. Indeed, the moment of x around c is determined by the following equation

$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n\, p(x)\, dx$$

Thus, if $p(x)$ is normalized (*i.e.*, $\mu_0 = 1$), then $\overline{x} = \mu_1$ around $c = 0$. In classical mechanics, the zeroth moment, first moment (around $c = 0$), and second moment of a distribution of mass $p(x)$ are related to the total mass (M_{total}), the center of mass (x_{CM}), and the moment of inertia (I), respectively.

$$M_{total} = \int_{-\infty}^{\infty} p(x)dx \quad \rightarrow \quad M_{total} = \mu_0$$

$$x_{CM} = \int_{-\infty}^{\infty} x\, p(x)dx \quad \rightarrow \quad x_{CM} = \frac{\mu_1}{M_{total}}$$

$$I = \int_{-\infty}^{\infty} (x-c)^2\, p(x)dx \quad \rightarrow \quad I = \mu_2$$

The variable c in the equation for μ_2 denotes the location of the axis of rotation. Let's now build on this connection to distributions further by calculating the variance, $\overline{(\Delta x)^2}$, of the distribution; note that the variance is sometimes also referred to as the *dispersion*[1].

$$\overline{(\Delta x)^2} = \overline{(x-\overline{x})^2} = \overline{x^2 - 2\overline{x}x + \overline{x}^2} = \overline{x^2} - 2\overline{x}^2 + \overline{x}^2$$

$$\overline{(\Delta x)^2} = \overline{x^2} - \overline{x}^2 \tag{2.3}$$

The starting point of a lot of physics is the characterization of distributions so we should try our best to get comfortable with these analyses.

2.2 Random Walk

Let's consider an object that can move along a one-dimensional coordinate axis. Let's denote the probability that the object moves in one direction along this coordinate axis (*e.g.*, to the right) as p and the probability that the object moves in the other direction as q. In both cases, we will assume that the object moves *1 unit* or *step* along the coordinate axis. Finally, we assume that each subsequent step taken by object is statistically independent of any previous step. In other words, all successive steps are statistically independent. With all of this in mind, the probability that the object will move n steps to the right out of a total of N steps is

$$p_n = \frac{N!}{n!\,(N-n)!} p^n q^{N-n}$$

We can then use this probability to calculate the mean number of steps to the right using Equation 2.1. The smallest number of steps to the right would be zero and the largest number of steps to the right would be N.

[1] Other names include mean squared displacement or mean square fluctuation.

$$\overline{n} = \frac{\sum_{n=0}^{N} n p_n}{\sum_{n=0}^{N} p_n} \quad \rightarrow \quad \overline{n} = \frac{\sum_{n=0}^{N} n \frac{N!}{n!(N-n)!} p^n q^{N-n}}{\sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n q^{N-n}}$$

We can simplify this distribution using the following identity

$$n \frac{N!}{n!\,(N-n)!} p^n q^{N-n} = p \frac{\partial}{\partial p} \left[\frac{N!}{n!\,(N-n)!} p^n q^{N-n} \right]$$

We can also take advantage of the equation for the binomial distribution

$$\sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!} p^n q^{N-n} = (p+q)^N$$

Putting it all together then gives us

$$\overline{n} = \frac{p \frac{\partial}{\partial p} \left[(p+q)^N \right]}{(p+q)^N} \quad \rightarrow \quad \overline{n} = \frac{pN\,(p+q)^{N-1}}{(p+q)^N} \quad \rightarrow \quad \overline{n} = \frac{pN}{p+q}$$

We can further simplify this if the probabilities are normalized so that the sum p and q is one.

$$p + q = 1 \quad \rightarrow \quad \overline{n} = pN$$

We can interpret this result as the product of N steps and the probability p of taking a single step to the right. This is the general relation for a distribution consisting of N steps. Namely, the parameters of the distribution are found by multiplying the number of steps in the distribution N by the parameters of a single step[2].

We can go through a similar calculation to obtain the following expression for the dispersion of the distribution of the number of steps to the right using Equation 2.3. To begin, let's calculate $\overline{n^2}$.

$$\overline{n^2} = \frac{\sum_{n=0}^{N} n^2 p_n}{\sum_{n=0}^{N} p_n} \quad \rightarrow \quad \overline{n^2} = \frac{\sum_{n=0}^{N} n^2 \frac{N!}{n!(N-n)!} p^n q^{N-n}}{\sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n q^{N-n}}$$

$$n^2 p^n = \left(p \frac{\partial}{\partial p} \right)^2 p^n \quad \rightarrow \quad \overline{n^2} = \frac{\sum_{n=0}^{N} \left(p \frac{\partial}{\partial p} \right)^2 \frac{N!}{n!(N-n)!} p^n q^{N-n}}{\sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n q^{N-n}}$$

[2]We'll come back to this in Chapter 10 when we discuss transport properties.

$$\overline{n^2} = \frac{\left(p\frac{\partial}{\partial p}\right)^2 \left[(p+q)^n\right]}{(p+q)^n}$$

Simplifying further than gives us

$$\overline{n^2} = \frac{\left(p\frac{\partial}{\partial p}\right) \left[pN\,(p+q)^{N-1}\right]}{(p+q)^n}$$

$$\overline{n^2} = \frac{p\left[N\,(p+q)^{N-1} + pN(N-1)\,(p+q)^{N-2}\right]}{(p+q)^n}$$

Let's now assume that $p+q=1$.

$$\overline{n^2} = p\left[N + pN(N-1)\right] \quad \rightarrow \quad \overline{n^2} = pN\left[1 + pN - p\right]$$

$$\overline{n^2} = pN\left[pN + (1-p)\right] \quad \rightarrow \quad \overline{n^2} = (pN)^2 + pNq$$

Let's substitute this result into Equation 2.3

$$\overline{(\Delta n)^2} = \overline{n^2} - \overline{n}^2 \quad \rightarrow \quad \overline{(\Delta n)^2} = (pN)^2 + pqN - (pN)^2$$

$$\overline{(\Delta n)^2} = pqN$$

This dispersion can be thought of as a a measure of the deviation of the position of the random walker with respect to the mean of the position distribution; this mean is thus the reference point for measurements of the position of the random walker. It can also be considered to be a measurement of the extent of the system sampled or explored by the random walker. As expected, the extent of the space explored by the random walk increases with the number of steps taken by the random walker.

The square root of the dispersion[3] is a linear measure of the width of the distribution of n. A good measure of the relative width of the distribution is thus

$$\frac{\left(\overline{(\Delta n)^2}\right)^{\frac{1}{2}}}{\overline{n}} = \frac{\sqrt{pqN}}{pN} = \sqrt{\frac{p}{q}}\frac{1}{\sqrt{N}}$$

As N increases, the relative width decreases. As shown in Figure 2.1, as the number of steps N increases, the probability distribution of n becomes tighter packed around its mean value. We can interpret this result as indicating that as N becomes large, it becomes increasingly unlikely that the system will be found at any value of n other than the mean value of n. In other words, ***for large systems, we can approximate the instantaneous value of a distribution with the mean value of that distribution without introducing***

[3] In other words, the root mean square of the number of steps to the right.

significant error as there is a low probability of finding the system at any other configuration[4]. This will be the starting point for the discussion of entropy in the next chapter.

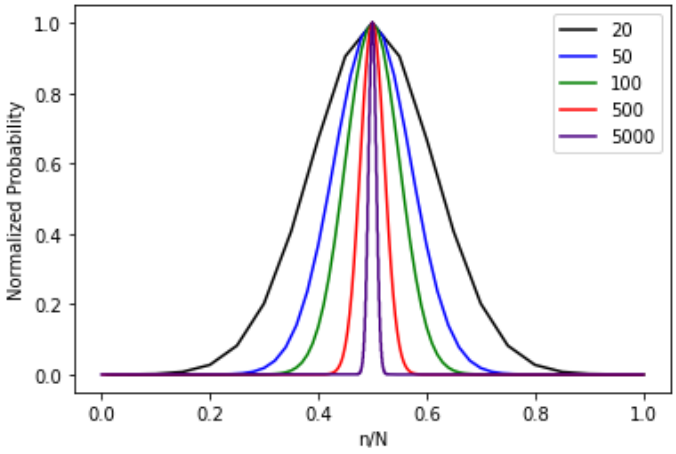

Figure 2.1: Normalized probability of observing a random walk system in a configuration associated with n steps to the right out of a total of N steps for different values of N; the probability of taking a step to the right is 0.5. As N increases, the probability of observing any outcome other than the mean outcome becomes vanishing small.

Approximation by Normal Distribution

When the number of steps in the random walk is large, the binomial distribution describing it can be approximated by a normal distribution.

$$P(n)dn = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}} dn$$

Let's now calculate the probability that a particular outcome (*i.e.*, a particular microstate) is within α of the mean of this distribution.

$$P = \int_{1-\alpha Np}^{1+\alpha Np} \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}} dn \quad \rightarrow \quad P = 2 \int_{0}^{1+\alpha Np} \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(n-Np)^2}{2Npq}} dn$$

Let's now define a new variable.

[4] Another interpretation of this same distribution is that of coin flipping. In this case p can denote the probability of getting *heads* and q the probability of getting *tails*. Then, for a fair coin, as the number of coin flips increases, the probability of obtaining any result other than half heads and half tails decreases.

N	α	$1-\alpha$
10	0.89	0.11
10^2	0.28	0.62
10^3	0.09	0.91
10^4	0.03	0.97
10^5	0.009	0.991

Table 2.1: Caption

$$t^2 = \frac{(n-Np)^2}{2Npq} \quad \rightarrow \quad dt = \frac{dn}{\sqrt{2Npq}}$$

Substitution gives us

$$P = \frac{2}{\sqrt{\pi}} \int_0^{\alpha\sqrt{\frac{Np}{2q}}} e^{-t^2} dt \quad \rightarrow \quad P = \text{erf}\left(\alpha\sqrt{\frac{Np}{2q}}\right)$$

This is the amusingly named error function. It is approximately equal to one when its argument is greater than 2. We can use this fact to explore the relationship between α and N.

$$\alpha\sqrt{\frac{Np}{2q}} \geq 2 \quad \rightarrow \quad \alpha^2 \geq \frac{8q}{Np}$$

If $p = q = \frac{1}{2}$, $\alpha \geq \sqrt{\frac{8}{N}}$. As shown in Table 2.1, the value of α decreases with increasing N, indicating that the width around the mean enclosing effectively all of the probability of the distribution narrows as N increases. When N is on the order of 10^5, the probability function is effectively contained within 1% of the mean of the function.

Displacement

Next, let's write the equation for this distribution in terms of the variable m, which denotes the difference between the number of steps taken to the right and the number of steps taken to the left. In other words, m denotes the net displacement (in steps) from the starting point of the random walk. When N is large, the probability of obtaining a displacement m out of a total of N steps can be approximated by

$$P(m) = \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(m-N(p-q))^2}{8Npq}}$$

When the probabilities of taking steps to the right and left are equal (*i.e.*, when $p = q$), this distribution is centered on $m = 0$, as expected.

If the steps have size l, the displacement can be described by $x = ml$. When N is large, we can consider x to be a continuous variable and determine the

probability that random walker is found in a range between x and $x+dx$. Since m can have only integer values separated by $\Delta m = \pm 2$, the range dx contains $\frac{dx}{2l}$ possible values of m, all of which occur with nearly the same probability $p(m)$. Thus, the probability of finding the random walker between $x + dx$ is

$$P(x)dx = P(m)\frac{dx}{2l}$$

$$P(x)dx = \frac{1}{2l\sqrt{2\pi Npq}} e^{-\frac{(x-Nl(p-q))^2}{8Npq}} dx$$

We see that the standard deviation of this distribution is

$$\sigma = 2l\sqrt{Npq}$$

Thus, the standard deviation of the distribution is proportional to the square root of the dispersion. To put this another way, the variance of the distribution is proportional to the dispersion, which sorta kinda makes sense.

Timescales

If we assume that each step of the random walk requires a time Δt, then

$$\overline{(\Delta n)^2} \approx N\Delta t \quad \rightarrow \quad \left(\overline{(\Delta n)^2}\right)^{\frac{1}{2}} \approx (N\Delta t)^{\frac{1}{2}}$$

The linear spread of the distribution increases as the square root of the total time of the random walk.

2.3 Statistical Approach to Work and Heat

Building on what we learned in Section 3.1.1, we can now express the mean energy of a system in terms of the energy states E_n accessible to that system and the probabilities p_n that those states are occupied. Applying Equation 2.1 gives us

$$\overline{E} = \frac{\sum\limits_n E_n p_n}{\sum\limits_n p_n}$$

Let's assume that the probability distribution p_n is normalized.

$$\sum_n p_n = 1 \quad \rightarrow \quad \overline{E} = \sum_n E_n p_n$$

It follows that the mean energy of the system can change if the accessible energy states change or if the probability those states are occupied changes.

$$d\overline{E} = \sum_n \left(p_n dE_n + E_n dp_n\right)$$

$$d\overline{E} = \sum_n p_n dE_n + \sum_n E_n dp_n \tag{2.4}$$

Let's consider the first term on the right-side of Equation 2.4. We can express dE_n in terms of its partial derivatives with respect to the generalized position coordinates of our system, denoted as q_k.

$$dE_n = \sum_k \frac{\partial E_n}{\partial q_k} dq_k$$

Thus,

$$\sum_n p_n dE_n = \sum_n p_n \sum_k \left(\frac{\partial E_n}{\partial x_k} \right) dx_k = \sum_k \sum_n p_n \left(\frac{\partial E_n}{\partial x_k} \right) dx_k$$

The derivative of energy with regards to position in this expression is the generalized force for the k^{th} generalized coordinate.

$$F_k = -\left(\frac{\partial E_n}{\partial x_k} \right) \quad \rightarrow \quad \sum_n p_n dE_n = \sum_k \sum_n p_n \left(-F_k \right) dx_k$$

The summation of probability and generalized force in this expression is the average generalized force

$$\sum_n p_n F_k = \overline{F}_k$$

Thus,

$$\sum_n p_n dE_n = -\sum_k \overline{F}_k dx_k$$

The product of generalized force and associated infinitesimal displacement in this equation is the corresponding infinitesimal work done by that force[5].

$$\overline{F}_k dx_k = đW_k \quad \rightarrow \quad \sum_n p_n dE_n = -\sum_k đW_k$$

$$\sum_n p_n dE_n = -đW$$

Substitution of this result into Equation 2.4 then gives us the following expression

$$d\overline{E} = -đW + \sum_n E_n dp_n$$

Equating $d\overline{E}$ in this expression with dE in Equation 1.2 then gives us

[5]We have already discussed in Section 1.5 that work is the product of a generalized force and a generalized displacement

$$-đW + \sum_n E_n dp_n = đQ - đW \quad \rightarrow \quad đQ = \sum_n E_n dp_n$$

$$đW = -\sum_n p_n dE_n \qquad đQ = \sum_n E_n dp_n \tag{2.5}$$

Work changes the energy of each state by changing the external parameters of the system but does not change the probability of occupying each state. Heat changes the probability of occupying any state while leaving the energy of each state unaffected.

The Infinite Square Well

Let's take a moment to remember some quantum mechanics. Specifically, let's recall the solution to the Schrödinger equation for a particle in a 1-dimensional infinite square well potential described by

$$V(x) = \begin{cases} 0 & -a < x < a \\ \infty & |x| > a \end{cases}$$

Since the potential energy is infinite at $x = \pm a$, the probability of finding the particle outside of the well is zero and thus the wave function $\psi(x)$ must therefore vanish for $|x| > a$. This leads to a quantization of energy states accessible to the particle.

$$E_n = \frac{\hbar^2 \pi^2}{2m} \frac{n^2}{(2a)^2} \quad n = 1, 2, 3, 4, 5, \ldots$$

We see that when the width of the potential well changes (*i.e.*, when a changes), the energy states change. Changing a, exerting a force to expand the length of the well, for example, would require work as it would be associated with a force being exerted over a displacement. This is a trivial example of the relationship between $đW$ and dE_n.

2.4 Boltzmann's H-Theorem

In a publication in 1872, Ludwig Boltzmann defined a quantity H as the average of the natural logarithm of the probability of the accessible microstates of a system.

$$H = \overline{\ln p} = \sum_n p_n \ln(p_n) \tag{2.6}$$

Let's now differentiate H with respect to time.

$$\frac{dH}{dt} = \sum_n \left(\frac{dp_n}{dt} \ln(p_n) + \frac{dp_n}{dt} \right) \quad \rightarrow \quad \frac{dH}{dt} = \sum_n \frac{dp_n}{dt} \left(\ln(p_n) + 1 \right)$$

Time-dependent changes in p_n can be written in terms of transitions into the state n from a different state, denotes as s, and transitions from state n into state s. We can denote the probability of these transitions as p_{sn} and p_{ns}, respectively.

$$\frac{dp_n}{dt} = \sum_s p_s p_{sn} - \sum_s p_n p_{ns} \quad \rightarrow \quad \frac{dp_n}{dt} = \sum_s \left(p_s p_{sn} - p_n p_{ns}\right)$$

Let's now assume that the Hamiltonian associated with the transition between states is Hermitian. In that case, $p_{sn} = p_{ns}$.

$$p_{sn} = p_{ns} \quad \rightarrow \quad \frac{dp_n}{dt} = \sum_s p_{ns} \left(p_s - p_n\right)$$

Substituting this expression into our equation for $\frac{dH}{dt}$ then yields

$$\frac{dH}{dt} = \sum_n \sum_s p_{ns} \left(p_s - p_n\right) \left(\ln(p_n) + 1\right) \tag{2.7}$$

Of course, since our designation of the states s and n is completely arbitrary, we could have equally validly written this double sum as

$$\frac{dH}{dt} = \sum_s \sum_n p_{sn} \left(p_n - p_s\right) \left(\ln(p_s) + 1\right) \tag{2.8}$$

When we add Equation 2.7 and Equation 2.8 together (while recognizing that $p_{sn} = p_{ns}$) we obtain

$$2\frac{dH}{dt} = \sum_n \sum_s p_{ns} \left(p_s - p_n\right) \left(\ln(p_n) - \ln(p_s)\right)$$

Now let's switch the order of the probabilities in the first parenthetical term in this expression.

$$2\frac{dH}{dt} = -\sum_n \sum_s p_{ns} \left(p_n - p_s\right) \left(\ln(p_n) - \ln(p_s)\right)$$

We can rewrite this expression as

$$\frac{dH}{dt} = -\frac{1}{2} \sum_n \sum_s p_{ns} \left(p_n - p_s\right) \left(\ln(p_n) - \ln(p_s)\right) \tag{2.9}$$

If $p_n > p_s$ then $\ln(p_n) > \ln(p_s)$. Similarly, if $p_n < p_s$ then $\ln(p_n) < \ln(p_s)$. It is then follows from Equation 2.9 that

$$\frac{dH}{dt} \leq 0 \tag{2.10}$$

The equality in Equation 2.10 is valid when $p_n = p_s$. Thus, we can interpret Equation 2.10, which is referred to as Boltzmann's H-theorem, as fact that the quantity H always tends to decrease with time, reaching a minimum at equilibrium.

Hermitian Operator Review

An operator H is defined as Hermitian if its off-diagonal matrix elements are complex conjugates of one another.

$$H = H^*$$

Recall that H^* denotes the conjugate transpose of H (*i.e.*, the transpose of the complex conjugate of H). A Hermitian operator can always be moved over the the other side of an inner product.

$$\langle b| H |a\rangle = \langle b|Ha\rangle = \langle Hb|a\rangle$$

But perhaps more important for us is

$$\langle b|Ha\rangle = \langle a|Hb\rangle^*$$

because we want to talk about transitions between states. If the operator H couples ψ_a and ψ_b, thus allowing for transitions between these states, then the probability of a transition from state a to state b.

$$p_{ba} \propto |\langle b| H |a\rangle|^2$$

If H is Hermitian, then

$$\langle b| H |a\rangle = \langle Hb|a\rangle = \langle a|Hb\rangle^* = \langle a| H |b\rangle^*$$

So,

$$p_{ba} \propto |\langle b| H |a\rangle|^2 \quad \rightarrow \quad p_{ba} \propto \left|\langle a| H |b\rangle^*\right|^2 \quad \rightarrow \quad p_{ba} \propto |\langle a| H |b\rangle|^2$$

$$p_{ba} = p_{ab}$$

If the Hamiltonian governing the transition between states a and b is Hermitian, the rate of the transition from state a to state b is the same as the rate of the transition from state b to state a.

A few things about Boltzmann's H-Theorem

People got quite upset back in 1872 when Boltzmann published his work on H-theorem. First, we have what is referred to as Loschmidt's reversibility paradox. The argument is that since the processes underlying the motion of the molecules in a system are governed by Newtonian mechanics, which is time reversible, a macroscopic thermodynamic description of those molecules must also be time reversible. Therefore, Boltzmann's argument that H always decreases cannot be true. Boltzmann responded to this criticism by stating that his H-theorem should be interpreted as a probability-based discussion of equilibrium. Indeed, one could interpret the statement that H must decrease with time as

saying that as time goes by systems will tend to be found in their most probable configuration/state. Much later, Lanford showed that the time-reversal invariant Hamiltonian equations of motion for the movement and collision of gas molecules can lead directly to the irreversible behavior of an ideal gas encapsulated by Boltzmann's H-theorem.

Boltzmann next addressed Zermelo's argument that H-theorem violated the Poincairé recurrence theorem. What's the Poincairé recurrence theorem you ask? Well, it states that dynamical systems will return to a state arbitrary close to their initial state after a sufficiently long but finite time. This follows from Liouville's theorem (see Appendix), which tells us that all Hamiltonian described systems are *volume-preserving* in phase space. Eventually, all generalized coordinates in the phase space of a closed system[6] must be intersected by the flow of the system. Boltzmann responded to this criticism by saying that while systems must return to their initial state in a finite amount of time, that amount of time is nevertheless so large that measurements of systems returning to their initial conditions are extremely unlikely. This is part of a larger discussion of probability and measurement. For example, there is a non-zero probability that all of the air in the room will collapse to one corner of the room, but this probability is so low that the time required for us to observe it happen is very, very large, perhaps larger than the age of the universe, but nevertheless finite.

2.5 Summary

- Work changes the energy of each state by changing the external parameters of the system but does not change the probability of occupying each state. Heat changes the probability of occupying any state while leaving the energy of each state unaffected.

$$đW = -\sum_n p_n dE_n \qquad đQ = \sum_n E_n dp_n$$

- The mean of x in terms of the probability distribution for x, denoted as $p(x_n)$, can be expressed as

$$\overline{x} = \frac{\sum_{n=1}^{N} p(x_n) x_n}{\sum_{n=1}^{N} p(x_n)} \qquad \overline{x} = \frac{\int x\, p(x)\, dx}{\int p(x)\, dx}$$

- The variance of a distribution is also known as the dispersion of the distribution.

$$\overline{(\Delta x)^2} = \overline{x^2} - \overline{x}^2$$

[6] A closed system is another way of saying a system where the phase space is bounded