# 1 Introduction

## 1.1 What is Statistical Mechanics?

We perceive the world at a macroscopic level, dealing with substances composed of atoms, yet wholly ignorant of the state of those individual atoms. As physicists, there are many situations in which we may wish to describe the macroscopic behaviour of some physical system, even though we do not have full knowledge of the state of all of the microscopic degrees of freedom of that system. For instance, suppose we want to know the physical properties of a container of some gas, whose atoms barely interact with each other, such as argon. The most brute-force approach to trying to understand the physical properties of the gas would be to take the Schrödinger equation for the $\sim 10^{23}$ atoms in the container, and try to solve it to find the eigenvalues and eigenstates of any observable that we might be interested in. Even if we were able to do this (which we aren't), we would still need to know the initial state of the gas in order to be able to determine the state of the gas at any particular time. Moreover, the subsequent dynamics will be chaotic, so that we can never know the initial conditions with sufficient precision to be able to calculate the state of the gas at arbitrary times. Clearly, such an approach is impractical, even with the fastest computers currently available (or likely to be available in the future).

One of the triumphs of eighteenth- and nineteenth-century physics was the development of thermodynamics, which allowed the description of the physical properties of a system such as a container of gas to be reduced to a single equation of state, the ideal gas law

$$PV = Nk_BT, \tag{1.1}$$

which relates thermodynamic quantities pressure $P$, volume $V$ and temperature $T$ without any reference to the microscopic details of the gas.

However, thermodynamics does not explain why, even though the ideal gas law does a good job of describing dilute gases of both argon and oxygen (or any other simple gas), argon has a heat capacity $C_V = \frac{3}{2}k_B$, whilst oxygen has a heat capacity $C_V = \frac{5}{2}k_B$. The distinction between these two heat capacities depends on the microscopic degrees of freedom that are available to atoms or molecules in each gas. As diatomic molecules, oxygen molecules have extra degrees of freedom than argon atoms, which leads to more possible ways for energy to be deposited in such a gas. For instance, oxygen molecules can have rotational kinetic energy in addition to the translational kinetic energy that argon atoms can have. Statistical mechanics is a framework to calculate such macroscopic properties, from a knowledge of microscopic details of a system. Researchers

such as Maxwell, Boltzmann and Gibbs developed these ideas in a classical context, but they have proven to be equally applicable to quantum systems.

While we study a number of problems in this book that can be solved exactly, most problems in statistical physics cannot be solved explicitly. Nevertheless, the concepts that we develop here can be applied either in numerical contexts, such as molecular dynamics or Monte Carlo simulations, or in approximate analytic approaches. Statistical mechanics also allows us to understand why we can describe the properties of a complex system with many different possible arrangements of its microscopic constituents, for example $\sim 10^{23}$ atoms, in terms of just a few quantities, such as temperature, pressure and chemical potential, and why thermal equilibrium exists. As such, it provides a microscopic foundation to the thermodynamic concepts used in physics and other fields such as chemistry, engineering and biology.

An important part of thermodynamics is the concept of entropy, which is often described in somewhat unsatisfying terms as "the amount of disorder in a system" or some similar phrase. In the context of statistical mechanics, entropy can be given a precise meaning, as a measure of our ignorance about a system. Such a view is in accord with the idea that we describe the statistical properties of a system rather than trying to solve the equations of motion for its constituents (e.g. Newton's laws or the Schrödinger equation). More broadly, the idea of entropy also plays an important role in information theory. In Chapter 2 we will see how entropy is related to irreversibility, as embodied in the second law of thermodynamics, even though the relevant equations of motion are invariant under time reversal.

Examples of systems and situations that we will encounter and/or can be described in a statistical mechanical framework include: ideal gases, spins in magnetic materials, photons in the Sun, white dwarf stars, molecular motors, DNA, Bose–Einstein condensation, impurities in crystals, polymer folding, stock market fluctuations, electrical noise in circuits, magnetic domains, superfluidity and phase transitions.

## 1.2 Probabilistic Behaviour

Having established that it is futile to try to follow the trajectories of every single molecule in a gas or to solve the corresponding equations of motion for the particles, we turn to a statistical description that will allow us to obtain the collective behaviour of the system without knowing what any individual particle is doing. This is in accord with how we describe physical systems – we generally are interested in macroscopic properties, such as pressure or temperature in a gas, that arise from the average over the motion of individual molecules. We know that certain configurations of the molecules are very unlikely, e.g. all the molecules being found in one corner of a container. To quantify this intuitive understanding we will assign probabilities to different configurations. Averages over these probabilities will give us the average values of physical quantities, which we will be able to relate to likely outcomes of macroscopic measurements when the number of constituents of a system becomes large. To give such a description, we need to start with a careful discussion of probabilities and probability distributions.

There are a number of different ways to think about probability. An empirical approach is the frequentist point of view: if we do $N$ trials of an experiment and measure $X$ for some quantity in $N_X$ of those trials, then the frequentist view is that the probability of measuring $X$ is

$$P(X) = \lim_{N\to\infty} \frac{N_X}{N}. \tag{1.2}$$

A problem with this approach is that $N \to \infty$ is an unattainable limit, it is impossible to make an infinite number of measurements – it also ties a probability to an actual sequence of events. Another way to think about probability is via a propensity approach, in which the probability is viewed as intrinsic to a particular physical situation and is then reflected in subsequent measurements. In this view, $P(X)$ exists for the physical situation we are considering and then if we make a sequence of measurements, those measurements will be determined by $P(X)$. A third approach is the Bayesian point of view, where the probabilities assigned to different possible outcomes, e.g. $X$ or $Y$, depend on prior knowledge. We will tend to take a propensity approach to probabilities, but from a mathematical point of view, provided a probability satisfies the axioms, then it is acceptable.

### 1.2.1 Axioms of Probability

We introduce $P(A)$ as the probability that some outcome $A$ occurs given some initial conditions, e.g. this could be the probability that a fair coin comes up heads when tossed. In order to be a probability, $P(A)$ must satisfy the following four axioms, which express basic properties that we require of any reasonable probability:

**(1)** $P(A) \geq 0$ – this guarantees there are no negative probabilities.

**(2)** $\sum_i P(A_i) = 1$ – this expresses the idea that if we consider all possible outcomes $A_i$ for the given initial conditions then one of these must happen and so the sum of all the probabilities is 1.

**(3)** The probability when combining independent outcomes for the same event is additive (e.g. this could be the probability of rolling either a 1 or a 4 on a die, in which case the probability is equal to the sum of the probability of rolling 1 with the probability of rolling 4). Hence for independent outcomes $A$ and $B$

$$P(A \vee B) = P(A) + P(B), \tag{1.3}$$

where $\vee$ is logical *or*, or equivalently

$$P(A) = 1 - P(\bar{A}), \tag{1.4}$$

where $\bar{A}$ is logical *not* $A$. This is also equivalent to

$$P(A) + P(\bar{A}) = 1, \tag{1.5}$$

which reflects that it is certain that we get one of the outcomes $A$ or $\bar{A}$.

**(4)** The probability of independent events is multiplicative (e.g. we might want to know the probability that we get a head when we flip a coin and a 6 when we roll a die):

$$P(A \wedge B) = P(A)P(B), \tag{1.6}$$

where $\wedge$ is logical *and*.

### 1.2.2 Example: Coin Toss Experiment

A simple example which illustrates the ideas expressed in the axioms in a concrete way, and which also introduces the binomial distribution, is a coin toss experiment. We can flip a coin many times and we will get some sequence of heads and tails, e.g.

HTTHHHTHTTTTTHTHTHTHTHHTHTTHTHTHTHTHTHTTTHHTH . . .

We can't predict the outcome of any one coin toss, but for a fair coin, the odds of having a head or a tail when we toss the coin are $p = \frac{1}{2}$ and $q = \frac{1}{2}$, respectively.

For $N$ tosses of the coin, we can determine the probability of there being $N_H$ heads and $N_T$ tails by drawing a tree diagram as illustrated in Fig. 1.1 and counting possibilities, for instance, or more systematically by using the binomial theorem:

$$\begin{aligned} 1 = (p+q)^N &= p^N + Np^{N-1}q + \frac{1}{2}N(N-1)p^{N-2}q^2 + \cdots + q^N \\ &= \sum_{M=0}^{N} \binom{N}{M} p^{N-M} q^M \\ &= \sum_{M=0}^{N} P(N, N-M), \end{aligned} \tag{1.7}$$

where

$$\binom{N}{M} = \frac{N!}{M!\,(N-M)!}, \tag{1.8}$$

and $P(N, N-M)$ is the probability of having $N-M$ heads and $M$ tails. If we label the number of heads as $N_H$ and the number of tails as $N_T$, then the probability of having $N_H$ heads and $N_T$ tails after a total of $N = N_H + N_T$ coin tosses will be given by

$$P(N, N-M) = P(N, N_H) = \binom{N}{M} p^{N-M} q^M = \frac{N!}{N_H!\,N_T!} p^{N_H} q^{N_T}. \tag{1.9}$$

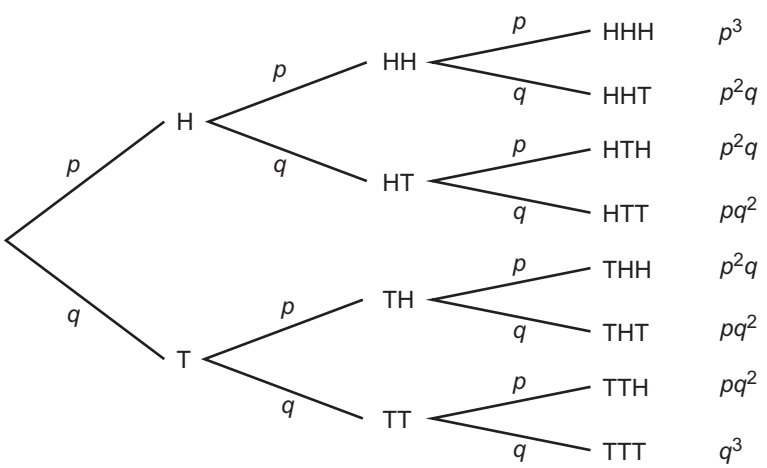


Fig. 1.1 Probability tree diagram for three coin tosses. The probability of obtaining a head (H) on any coin toss is $p$ and the probability of obtaining a tail (T) is $q$.

Now, suppose we pick $N = 100$, then our general expectation based on experience is that $\Delta N = N_H - N_T$ should not be too large (i.e. much less than 100). We can be quantitative and we see that the probability that we get $\Delta N = \pm 100$ is $p^{100} = 1/2^{100} \simeq 8 \times 10^{-31}$. In contrast, we might guess that it is quite likely that we will get $\Delta N = 0$, which is given by the probability of 50 heads:

$$P(100, 50) = \binom{100}{50} \frac{1}{2^{100}} \simeq 8.0\%. \tag{1.10}$$

However, this is not very different from the probability of $\Delta N = \pm 2$, corresponding to 49 or 51 heads:

$$P(100, 51) = P(100, 49) = \binom{100}{49} \frac{1}{2^{100}} \simeq 7.8\%, \tag{1.11}$$

but considerably more than $\Delta N = \pm 20$, which corresponds to 40 or 60 heads:

$$P(100, 40) = P(100, 60) = \binom{100}{40} \frac{1}{2^{100}} \simeq 1.1\%. \tag{1.12}$$

If we were to perform $N = 10$ coin flips, for instance, then looking at $\Delta N$ as we have done here would not be very useful to compare with what we find when $N = 100$. A quantity that is more useful is

$$\frac{\Delta N}{N} = \frac{N_H - N_T}{N},$$

which gives us the relative deviation from the behaviour expected of a fair coin (i.e. $N_H = N_T$). We might expect that the probability that this quantity is close to zero gets larger as $N$ increases. For instance, when $N = 100$, the probability that $-0.1 \leq \Delta N/N < 0.1$ is $\simeq 73\%$, whereas for $N = 1000$, the probability is $\simeq 99.9\%$.

### 1.2.3 Probability Distributions

The coin toss example suggests that rather than focusing on individual outcomes of events, which are random, we should look at quantities that are robust properties of the system, such as average values or the distribution of allowed values – we may not know the exact result of any particular event, but would like to know the range of highly probable outcomes.

The objects that will be central to this discussion are probability distributions. A probability distribution $P(X)$ for some quantity $X$ gives the probability of the different allowed values of $X$. It can be either discrete, in which case it will take the form $p(x_i)$, which measures the probability that $X$ takes the value $x_i$, or continuous, in which case we can write the probability as $p(x)dx$, which corresponds to the probability that $X$ takes a value between $x$ and $x + dx$. In this second case, $p(x)$ is a probability density, as illustrated in Fig. 1.2.

Given a probability distribution $p(x_i)$, we can calculate the mean value of some function $f$ of $X$ by weighting the values of $f$ by the probability distribution:

$$\langle f(x) \rangle = \overline{f(x)} = \sum_i f(x_i)\, p(x_i). \tag{1.13}$$

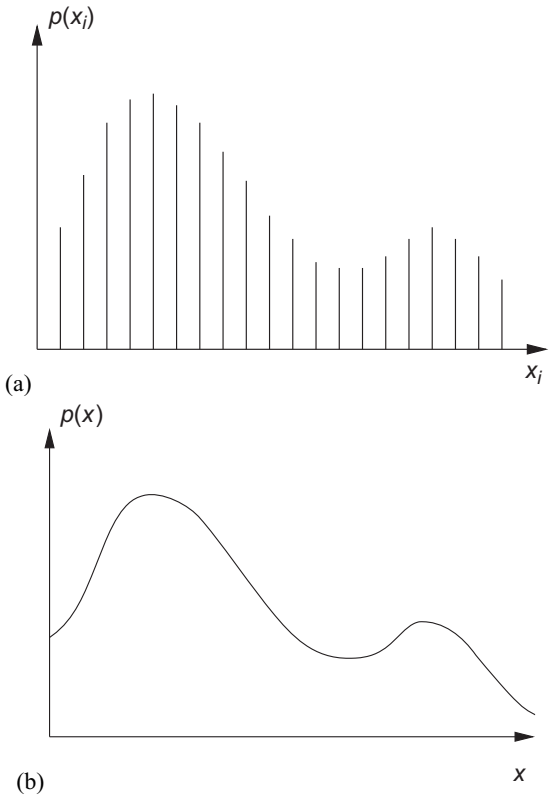


Fig. 1.2 Examples of (a) a discrete and (b) a continuous probability distribution.

For a continuous probability distribution we should replace the sum with an integral:

$$\langle f(x)\rangle = \int dx\, f(x)\, p(x). \tag{1.14}$$

In the following, results written with summations should be understood as applying to both discrete and continuous probability distributions. We will be particularly interested in moments of probability distributions as these will allow us to characterize the shape of unknown probability distributions. For instance, for the $m^{\text{th}}$ moment:

$$\langle x^m\rangle = \overline{x^m} = \sum_i x_i^m\, p(x_i). \tag{1.15}$$

The main moments of distributions that we will be concerned with are:

(i) the zeroth moment

$$\left\langle x^0\right\rangle = \sum_i p(x_i) = 1 \tag{1.16}$$

(as we would expect for a probability);

(ii) the first moment (the mean)

$$\langle x \rangle = \bar{x} = \sum_i x_i \, p(x_i), \tag{1.17}$$

and;

(iii) the second moment

$$\left\langle x^2 \right\rangle = \sum_i x_i^2 \, p(x_i). \tag{1.18}$$

With knowledge of these quantities we can also determine the central moments of a distribution, i.e. the moments with respect to the mean. Define

$$\Delta x = x - \langle x \rangle ,$$

then we can see immediately that

$$\begin{aligned} \langle \Delta x \rangle &= \langle x - \langle x \rangle \rangle \\ &= \sum_i x_i \, p(x_i) - \sum_i \langle x \rangle \, p(x_i) \\ &= \langle x \rangle - \langle x \rangle \, 1 \\ &= 0, \end{aligned} \tag{1.19}$$

and that the variance is

$$\begin{aligned} \left\langle (\Delta x)^2 \right\rangle &= \left\langle (x - \langle x \rangle)^2 \right\rangle \\ &= \left\langle x^2 \right\rangle - 2 \langle x \langle x \rangle \rangle + \langle x \rangle^2 \\ &= \left\langle x^2 \right\rangle - 2 \langle x \rangle^2 + \langle x \rangle^2 \\ &= \left\langle x^2 \right\rangle - \langle x \rangle^2 , \end{aligned} \tag{1.20}$$

which is closely related to the standard deviation

$$\sigma = \sqrt{\left\langle (\Delta x)^2 \right\rangle} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}. \tag{1.21}$$

We note in passing that the third and fourth central moments give the skew and the kurtosis, respectively, which can be used to characterize the shape of a probability distribution. These are important quantities in the theory of probability distributions but we will not use them in our discussion of statistical physics.

### 1.2.4 Example: Random Walk

Random walks describe many processes that can take place in nature. A very short list of examples includes the folding of polymers, Brownian motion of small particles, photons diffusing in the Sun, electrons diffusing in a wire, molecular motors, chemotaxis, genetic drift and stock prices. A schematic example of a random walk in two dimensions is illustrated in Fig. 1.3.

We will initially consider the simplest version of a random walk. Imagine a walker placed on a one-dimensional line. The walker takes steps of equal length randomly either to the left or right. Without knowing the exact sequence of steps that the walker takes, we

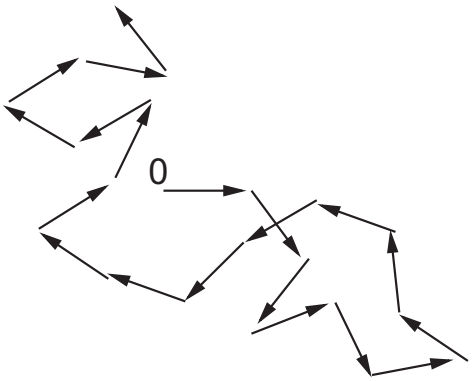


Fig. 1.3 A random walk in two dimensions with fixed step length but random direction for each step starting from the origin 0.

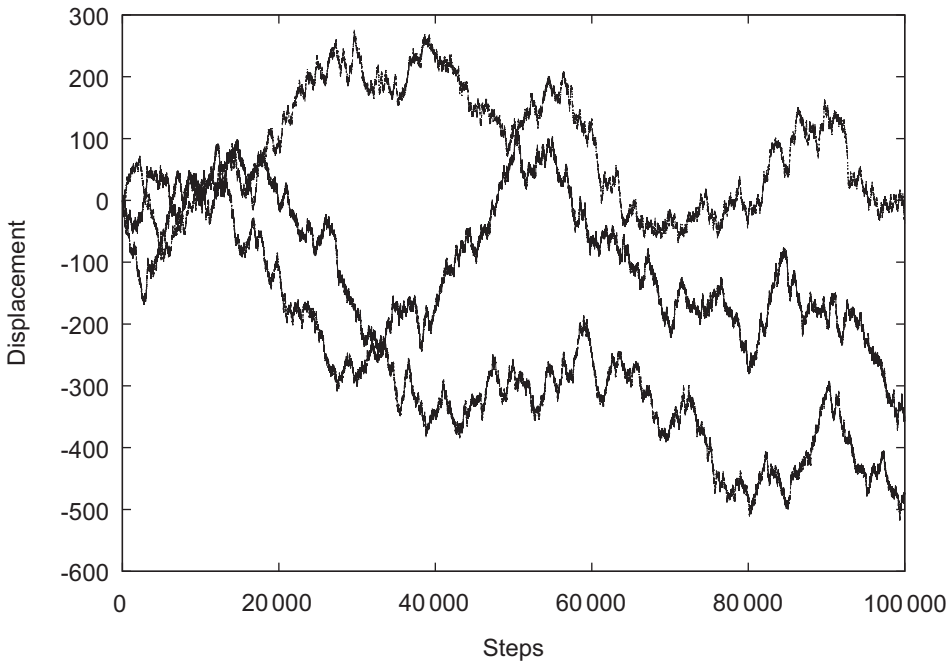


Fig. 1.4 Three random walks in one dimension, each with 100 000 steps.

cannot determine exactly where they will end up (see Fig. 1.4). However, we can say very precisely what the probabilities for finding the walker at any position along the line are after the walker has taken $N$ steps. Note that we could regard our coin-flipping example as a random walk, with heads corresponding to a move to the right and tails to a move to the left, so we are also obtaining the probability distribution for coin flips.

To simplify the problem, suppose the walker starts at the origin, and takes steps of length $a$. After they have taken $N$ steps, $N_L$ will have been to the left and $N_R$ will have been to the right, and let us call their position on the line $Ra$. Then

$$N = N_R + N_L, \tag{1.22}$$
$$R = N_R - N_L, \tag{1.23}$$

and so

$$N_R = \frac{1}{2}(N + R), \tag{1.24}$$
$$N_L = \frac{1}{2}(N - R). \tag{1.25}$$

There are two possibilities at each step, a move to the right with probability $p$ and a move to the left with probability $q$. In this sense the problem is just like the coin flip problem, with $N_R$ equivalent to $N_H$ and $N_L$ equivalent to $N_T$. Hence, for a walk with displacement from the origin of $Ra$, we can determine from the binomial theorem that the probability of any path with $N$ steps of which $N_R$ are to the right and $N_L$ are to the left is

$$P(N,R) = \frac{N!}{N_R!\,N_L!}p^{N_R}q^{N_L} = \frac{N!}{\left[\frac{1}{2}(N+R)\right]!\left[\frac{1}{2}(N-R)\right]!}p^{\left[\frac{1}{2}(N+R)\right]}q^{\left[\frac{1}{2}(N-R)\right]}. \tag{1.26}$$

The probability of each individual path is $p^{N_R}q^{N_L}$ (where the combinatorial factor $N!/N_R!\,N_L!$ takes care of the fact that there are multiple paths that end up at the same endpoint).

If we want to calculate the mean displacement for a walk with $N$ steps, then we can do so as follows (recalling that $N_R - N_L = N_R - [N - N_R]$):

$$\begin{aligned}
\langle R\rangle &= \sum_{R=-N}^{N} R\,P(N,R) \\
&= \sum_{N_R=0}^{N}(N_R - N_L)\frac{N!}{N_R!\,(N-N_R)!}p^{N_R}q^{(N-N_R)} \\
&= p\frac{\partial}{\partial p}\left[\sum_{N_R=0}^{N}\frac{N!}{N_R!\,(N-N_R)!}p^{N_R}q^{(N-N_R)}\right]\Bigg|_{p+q=1} \\
&\quad - q\frac{\partial}{\partial q}\left[\sum_{N_R=0}^{N}\frac{N!}{N_R!\,(N-N_R)!}p^{N_R}q^{(N-N_R)}\right]\Bigg|_{p+q=1} \\
&= p\frac{\partial}{\partial p}(p+q)^N\Big|_{p+q=1} - q\frac{\partial}{\partial q}(p+q)^N\Big|_{p+q=1} \\
&= Np\,(p+q)^{N-1}\Big|_{p+q=1} - Nq\,(p+q)^{N-1}\Big|_{p+q=1} \\
&= N(p-q).
\end{aligned} \tag{1.27}$$

The trick we used in order to evaluate this sum was to write the sum in terms of derivatives of the binomial expansion

$$(p+q)^N = \sum_{M=0}^{N}\binom{N}{M}p^{N-M}q^M, \tag{1.28}$$

where we note that $p$ and $q$ are continuous variables, and wait until the end of the calculation to set $p + q = 1$. If $p = q$ then we can see that $\langle R \rangle = 0$. We should expect this on general grounds without doing any calculations, because there is no asymmetry between right and left. This is because for any path that leads to a positive displacement, there will be an equally probable path that leads to the same negative displacement and the average of these displacements will be zero.

An alternative physically motivated argument that reaches the same value for $\langle R \rangle$ is that if the probability of a step to the right is $p$, then after $N$ steps we should expect that the number of steps to the right will be $\langle N_R \rangle = pN$. Similiarly, we should expect that the number of steps to the left will be $\langle N_L \rangle = qN$. Given these two estimates we would expect that $\langle R \rangle = \langle N_R \rangle - \langle N_L \rangle = (p-q)N$, which is exactly what we found in a more complicated calculation in deriving Eq. (1.27).

We can use a similar approach to the one we used in Eq. (1.27) to calculate the mean square displacement:

$$
\begin{aligned}
\left\langle R^2 \right\rangle &= \sum_{R=-N}^{N} R^2 P(N,R) \\
&= \sum_{N_R=0}^{N} (N_R - N_L)^2 \frac{N!}{N_R!\,(N-N_R)!} p^{N_R} q^{(N-N_R)} \\
&= \left( p \frac{\partial}{\partial p} \right)^2 \left[ \sum_{N_R=0}^{N} \frac{N!}{N_R!\,(N-N_R)!} p^{N_R} q^{(N-N_R)} \right] \Bigg|_{p+q=1} \\
&\quad - 2pq \frac{\partial^2}{\partial p \partial q} \left[ \sum_{N_R=0}^{N} \frac{N!}{N_R!\,(N-N_R)!} p^{N_R} q^{(N-N_R)} \right] \Bigg|_{p+q=1} \\
&\quad + \left( q \frac{\partial}{\partial q} \right)^2 \left[ \sum_{N_R=0}^{N} \frac{N!}{N_R!\,(N-N_R)!} p^{N_R} q^{(N-N_R)} \right] \Bigg|_{p+q=1} \\
&= \left\{ \left( p \frac{\partial}{\partial p} \right)^2 (p+q)^N - 2pq \frac{\partial^2}{\partial p \partial q} (p+q)^N + \left( q \frac{\partial}{\partial q} \right)^2 (p+q)^N \right\} \Bigg|_{p+q=1} \\
&= \left\{ p \frac{\partial}{\partial p} pN(p+q)^{N-1} - 2pqN(N-1)(p+q)^{N-2} + q \frac{\partial}{\partial q} qN(p+q)^{N-1} \right\} \Bigg|_{p+q=1} \\
&= \Big[ Np(p+q)^{N-1} + p^2 N(N-1)(p+q)^{N-2} - 2pqN(N-1)(p+q)^{N-2} \\
&\quad + Nq(p+q)^{N-1} + q^2 N(N-1)(p+q)^{N-2} \Big] \Big|_{p+q=1} \\
&= (p-q)^2 N(N-1) + N \\
&= N(p+q)^2 + (p-q)^2 N(N-1) \\
&= N^2 (p-q)^2 + 4Npq \\
&= \langle R \rangle^2 + 4Npq.
\end{aligned}
\tag{1.29}
$$