1

Introduction to Combinatorics

How we count things turns out to have a powerful significance in physical problems! One of the oldest problems stems from undercounting and over-counting the number of possible configurations a particular system can have – mathematically, this is usually due to the fact that objects are mistakenly assumed to be indistinguishable when they are not, and vice versa. However, one of the great surprises of physics is that identical particles are fundamentally indistinguishable. In this chapter, we will introduce some of the basic mathematical objects that occur in physical problems, and give their enumeration. Statistical mechanics is one of the key sources of ideas, so we spend some time on the basic concepts here, especially as partition functions are clear examples of generating functions that we will encounter later on. We will recall some of the basic mathematical concepts in enumeration, leading on to the role of generating functions. At the end, we make extensive use of generating functions, exploiting the methods for dealing with partition functions in statistical mechanics, but for specific combinatorial families such as permutations and partitions.

We start, however, with the touchstone for all combinatorial problems: how to distribute balls in urns.

1.1 Counting: Balls and Urns

Proposition 1.1.1 *There are K^N different ways to distribute the N distinguishable balls among K distinguishable urns.*

The proof is based on the simple observation that there are K choices of urn for each of the N balls. Suppose next that we have more urns than balls.

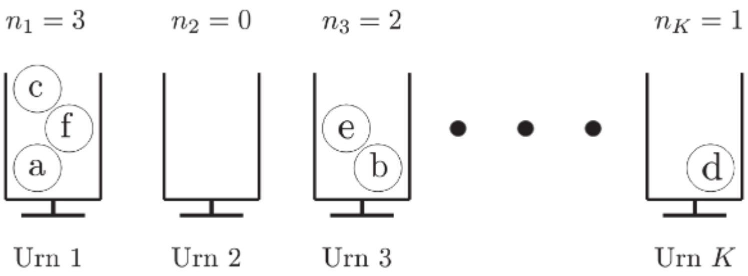

Figure 1.1 Occupation numbers of distinguishable balls in distinguishable urns.

Proposition 1.1.2 *The total number of ways to distribute the N distinguishable balls among K distinguishable urns so that no urn ends up with more than one ball is (later we will call this a falling factorial)*

$$K^{\underline{N}} \triangleq K(K-1)\cdots(K-N+1)\,.$$

The argument here is simple enough: if we have already distributed M balls among the urns, with no urn having more than one ball inside, so if we now want to place in an extra ball we have these $K - M$ empty urns remaining to choose from; therefore, $K^{\underline{M+1}} = K^{\underline{M}}(K-M)$ with $K^{\underline{1}} = K$.

Let n_k denote the number of balls in urn k; we call this the **occupation number** of the urn. See Figure 1.1. We now give the number of possibilities leading to a given set of occupation numbers, subject to the constraint of a fixed total number of balls, $\sum_{k=1}^{K} n_k = N$.

Proposition 1.1.3 *The number of ways to distribute N distinguishable balls among N distinguishable urns so that we have a prescribed number n_k balls are in the kth urn, for each $k = 1, \ldots, K$, is the multinomial coefficient*

$$\binom{N}{n_1 \ldots n_K} = \frac{N!}{n_1!\, n_2! \ldots n_K!}.$$

The proof here is based on the observation that there are $\binom{N}{n_1}$ ways to choose the n_1 balls to go into the first urn, then $\binom{N-n_1}{n_2}$ ways to choose next n_2 balls to go into the second urn, and so on, leading to

$$\binom{N}{n_1}\binom{N-n_1}{n_2}\cdots\binom{N-n_1-\cdots-n_{K-1}}{n_K} \equiv \frac{N!}{n_1!\, n_2! \ldots n_K!}.$$

Suppose however that the balls are in fact indistinguishable as in Figure 1.2! Then we do not distinguish between distributions having the same occupation numbers for the urns.

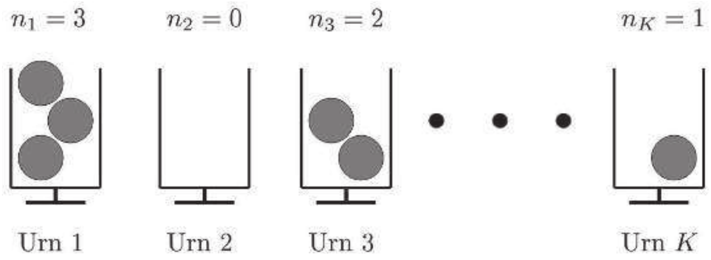

Figure 1.2 Occupation numbers of indistinguishable balls in distinguishable urns.

Proposition 1.1.4 *There are $\binom{N+K-1}{N}$ ways to distribute N indistinguishable balls among K distinguishable urns.*

Proof Take, for example, $K = 6$ urns and $N = 8$ balls and consider the distribution represented by occupation sequence $(1, 2, 0, 4, 1, 0)$, then encode this as follows:

$$\bullet|\bullet\bullet||\bullet\bullet\bullet\bullet|\bullet|$$

which means one ball in urn 1, two balls in urn 2, no balls in urn 3, and so on. In this encoding, we have $N + K - 1$ symbols (balls and sticks), N of which are balls and $K - 1$ of which are sticks (separations between the urns). In any such distribution, we must choose which N of the $N + K - 1$ symbols are to be the balls, and there are $\binom{N+K-1}{N}$ different ways to do this. Each way of selecting these symbols corresponds to a unique distribution, and vice versa. □

Proposition 1.1.5 *The number of ways to distribute N indistinguishable balls among K distinguishable urns, if we only allow at most one ball per urn, is $\binom{K}{N}$.*

That is, we must choose N out of the K urns to have a ball inside.

These enumerations turn out to be of immediate relevance to sampling theory in statistics. We note that if we have a set of K items and we draw a sample of size N, then if we make no replacement there will be $\binom{K}{N}$ such samples – imagine placing a ball into urn j if the jth element is selected! If replacement is allowed, then the number of samples is $\binom{N+K-1}{N}$.

1.2 Statistical Physics

Counting problems surfaced early on in the theory of statistical mechanics, and we recall the basic setting next.

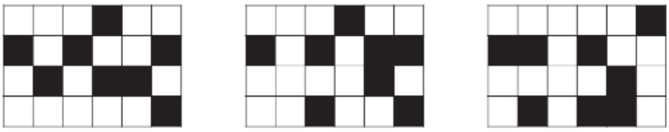

Figure 1.3 Three microstates, each corresponding to $N = 24$ particles, with 8 "on" (the black boxes).

1.2.1 The Microcanonical Ensemble

Two-State Model

Ludwig Boltzmann pioneered the microscopic derivation of laws of thermodynamics. To understand his ideas, we consider a very simple model of a solid material consisting of N particles, where each particle can be in either one of two states: an "off" state of energy 0 and an "on" state of energy ε. The total energy is therefore $U = \varepsilon M$, where M is the number of particles in the "on" state. In Figure 1.3, we have three typical examples where $N = 24$ and $U = 8\varepsilon$, that is, in each of these we have 8 "on" states of of a total of 24. Each of these configurations is referred to as a **microstate**, and we say that they are consistent with the **macrostate** $(U = 8\varepsilon, N = 24)$.

Boltzmann's idea was that if the system was isolated, so that the energy U was fixed, the system's internal dynamics would make it jump from one microstate to another with only microstates consistent with the fixed macrostate (U, N) allowed. (In other words, the number of particles, N, and their energy U, are to be constants on the motion – whatever that happens to be.) Here the total number of microstates consistent with macrostate $(U = \varepsilon M, N)$ is then

$$W(U, M) = \binom{N}{M}.$$

He then made the **ergodic hypothesis**: *over a long enough period of time, each of these microstates was equally likely*: that is, the system may be found to be in a given microstate with frequency $1/W$. Therefore, long time averages would equate to averages over all the microstates consistent with the macrostate, with each microstate having equal probabilistic weight $1/W$. The latter probability system is known as the **microcanonical ensemble**. We note that the set of all microstates consistent with $(U = 8\varepsilon, N = 24)$ also includes some less-than-random-looking configurations such as the ones shown in Figure 1.4. But, nevertheless, they each get equal weight: here $W = \binom{24}{8} = 735, 471$.

At this resolution, we would expect to see the system run through the possible microstates, so that the picture over time would appear something like static on a TV screen. Configurations with a discernable pattern, as in Figure 1.4,

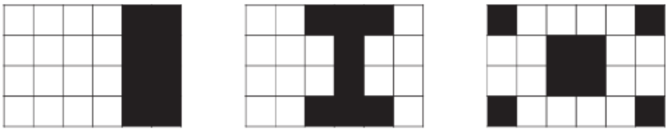

Figure 1.4 Another three microstates consistent with the $(U = 8\varepsilon, N = 24)$. Despite their apparent structure, each has the same 1/735,471 chance to occur as the more random ones in Figure 1.3.

may flash up briefly from time to time, but most of the time we are looking at fairly random-looking configurations such as in Figure 1.3. If N is large, and the particles are small, then we would expect to be looking most of the time at a uniform gray – the shade of gray determined by the ration M/N.

Boltzmann's remarkable proposal was that the entropy associated with a macrostate (U, N) was the logarithm of the number of consistent microstates

$$S(U, N) = k \ln W(U, N)$$

where k is a scale factor fixing our eventual definition of temperature scale. In the present case, we have $W(U = \varepsilon M, N) = \frac{N!}{(N-M)!M!}$. If we take N large with $U = Nu$ for some fixed ratio u, then using Stirling's approximation, $\ln N! = N \ln N - N + O(\ln N)$, we find that the entropy per particle in the bulk limit $(N \to \infty)$ is

$$s(u) = \lim_{N\to\infty} \frac{S(U = Nu, N)}{N} = -kp_0 \ln p_0 - kp_1 \ln p_1$$

where $p_1 = \frac{M}{N} \equiv \frac{u}{\varepsilon}$ and $p_0 = 1 - p_1$. (Note p_1 is the proportion of particles that are "on" in each of these microstates, with p_0 the proportion "off".) Alternatively, we may write this as

$$s(u) = -k \left\{ \frac{u}{\varepsilon} \ln \frac{u}{\varepsilon} + \left(1 - \frac{u}{\varepsilon}\right) \ln \left(1 - \frac{u}{\varepsilon}\right) \right\}.$$

From thermodynamics, one should identify the temperature T via the relation $1/T = \frac{\partial s}{\partial u} \equiv -\frac{k}{\varepsilon} \ln \left(\frac{\varepsilon}{u} - 1\right)$, and so in this model we have

$$u = \frac{\varepsilon}{e^{\varepsilon/kT} + 1}.$$

Somewhat surprisingly, this artificial model actually shows very good qualitative behavior for small values of u – see, for instance, Callen (1985, chapter 15). (Note that for $0 \le u < \varepsilon/2$, the temperature will be positive, but becomes negative for higher values $\varepsilon/2 < u \le \varepsilon$. Negative temperatures do actually

0	0	0	0	1	0
0	0	2	0	0	0
0	3	0	0	0	1
0	0	0	1	0	0

Figure 1.5 A microstate consistent with the $(U = 8\varepsilon, N = 24)$ in the Einstein model: one of 7,888,725.

makes physical sense, however, and are encountered in the related model of a two-state ferromagnet.)

Einstein's Model

A related model is the Einstein model for a crystalline solid. The difference is that each particle can have allowed energies $0, \varepsilon, 2\varepsilon, 3\varepsilon, \ldots$. This time we may depict a microstate as in Figure 1.5, where the number n in each box tells us that the corresponding particle has energy $n\varepsilon$ (or, perhaps more physically, that there are n quanta in the box!). We see that the number of microstates consistent with a macrostate $(U = \varepsilon M, N)$ will be

$$W = \binom{N + M - 1}{M}.$$

In other words, the number of ways of distributing M is indistinguishable quanta among N distinguishable boxes. This time, again using Stirling's identity, one may show that the entropy per particle in the bulk limit is

$$s\left(u\right) = k\left\{\ln\left(1 + \frac{u}{\varepsilon}\right) + \frac{u}{\varepsilon}\ln\left(1 + \frac{\varepsilon}{u}\right)\right\},$$

and that $u = \frac{\varepsilon}{e^{\varepsilon/kT} - 1}$.

1.2.2 The Canonical Ensemble

Boltzmann's ergodic hypothesis marks the introduction of probability theory into physics. So far we have not used probability explicitly; however, this changes if we consider the situation depicted in Figure 1.6. For simplicity, we work with the two-state model.

We fix the total energy U_{tot} of the $N_{\text{tot}} = N + N'$ particles, and then Boltzmann's principle tells us that each microstate of the total system is equally likely. However, the total number of "on" particles in total is constant, the number that are inside the system will vary with a probability distribution determined by the ergodic hypothesis. Let $M_{\text{tot}} = U_{\text{tot}}/\varepsilon$, and suppose that we have x particles in the "on" state in the system – then there are $\binom{N}{x}$ microstates

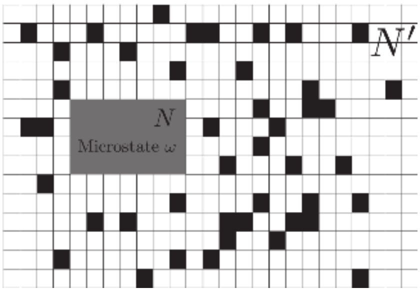

Figure 1.6 A system with N particles and microstate ω forms a subsystem of a larger system of $N+N'$ particles.

of the system consistent with this, and $\binom{N'}{M_{\text{tot}}-x}$ microstates possible for the complement. Allowing for all possible x leads to

$$\binom{N+N'}{M_{\text{tot}}} = \sum_{x=0}^{M_{\text{tot}}} \binom{N}{x}\binom{N'}{M_{\text{tot}}-x}.$$

If N, N' and M_{tot} are large of the same order, then it turns out that the largest term in the sum comes from $x \approx \rho N$, where $\rho = M_{\text{tot}}/N_{\text{tot}}$. In fact, this one term alone dominates to the extent that one may ignore all the other terms. This is related to the large deviation principle, which we discuss in Chapter 7.

The energy of the subsystem is now a random variable, and Boltzmann's hypothesis tells us its distribution. We now consider the situation where the number N of particles in our system is large but fixed, but we take $N' \to \infty$. We do this in such a way that the average number $\bar{u} = \varepsilon\, M_{\text{tot}}/N_{\text{tot}}$ is a constant. Suppose we have a fixed microstate ω of our subsystem with energy $E(\omega)$, i.e. the number of "on" particles in the subsystem. Then the probability of the particular microstate, ω, occurring is

$$\begin{aligned} p_{N'}(\omega) &= W(U', N')/W(U_{\text{tot}}, N_{\text{tot}}) \\ &= \binom{N'}{\frac{\bar{u}}{\varepsilon}N' + \frac{\bar{u}-u}{\varepsilon}N} \Bigg/ \binom{N+N'}{\frac{\bar{u}}{\varepsilon}(N+N')} \end{aligned}$$

where $u = E(\omega)/N$ is the energy density of the subsystem. Now making the approximation that $W(uN, N) \approx e^{Ns(u)/k}$, we find

$$\begin{aligned} k \ln p_{N'}(\omega) &= N'\, s\left(\bar{u} + (\bar{u}-u)\frac{N}{N'}\right) - (N+N')\, s(\bar{u}) \\ &= -N\, s(\bar{u}) + kN'\left[s\left(\bar{u} + (\bar{u}-u)\frac{N}{N'}\right) - s(\bar{u})\right] \\ &\to -N\, s(\bar{u}) + kNs'(\bar{u})(\bar{u}-u), \end{aligned}$$

since s is differentiable. Specifying the average energy per particle in the bulk limit $N' \to \infty$ to be $\bar{u}$ is equivalent to fixing the temperature T via the relation $1/T = s'(\bar{u})$, from which we see that $\ln p_{N'}(\omega) \to \frac{1}{kT}\big(F - uN\big)$, where $F = N\big(\bar{u} - Ts(\bar{u})\big)$. (Note that the relation between $\bar{u}$ and T is one-to-one. The variable $F = U - TS$ is the *Helmholtz free-energy* in thermodynamics.) That is, we obtain the probability

$$p_{\text{can.}}(\omega) = \frac{1}{Z} e^{-E(\omega)/kT},$$

where the normalization is given by the canonical partition function

$$Z_N = e^{-F/kT} = \sum_{\omega} e^{-E(\omega)/kT},$$

where the sum is over all microstates consistent with having a fixed number N of particles.

The probability distribution that we obtain in this way is called the **canonical ensemble** and is interpreted as saying that our subsystem is in thermal equilibrium with a heat bath at temperature T.

The derivation presented in the preceding is actually very general. We relied on the relation $W(uN, N) \approx e^{Ns(u)/k}$, but not the specifics of the entropy per particle, $s(u)$. So the same argument will go through so long as $s(u)$ exists and defines a monotone increasing, strictly concave function of u.

1.2.3 The Grand Canonical Ensemble

We now describe the situation in statistical mechanics where we have a gas consisting of a number of particles, N, each of which can have one of K distinguishable states with energy values $\varepsilon_1 \le \varepsilon_2 \le \cdots \le \varepsilon_K$; see Callen (1985). We allow both the energy and the number of particles to vary: we allow microstates ω, which have $N(\omega)$ particles and energy $E(\omega)$, and introduce the **grand canonical ensemble**

$$p_{\text{g.c.}}(\omega) = \frac{1}{\Xi} e^{-\big(E(\omega) - \mu N(\omega)\big)/kT},$$

where the normalization is given by the **grand canonical partition function**

$$\Xi = \sum_{\omega} e^{-\big(E(\omega) - \mu N(\omega)\big)/kT},$$

where the sum is now over all microstates – unrestricted in both number and energy. We introduce the standard notation of the *inverse temperature* $\beta = 1/kT$, and the parameter μ is known as the chemical potential. The alternate parameter $z = e^{\beta\mu}$ is called the *fugacity*.

We will make extensive use of the following lemma.

Lemma 1.2.1 (The $\sum\prod \longleftrightarrow \prod\sum$ Lemma) *Let* $\mathbb{M}$ *and* $\mathbb{K}$ *be countable sets, and let* $\mathbb{M}^{\mathbb{K}}$ *denote the sequences* $\mathbf{m} = (m_k)_{k\in\mathbb{K}}$ *where* $m_k \in \mathbb{M}$, *and let* $f : \mathbb{K} \times \mathbb{M} \to \mathbb{C}$. *We have the formal series relation*

$$\sum_{\mathbf{m}\in\mathbb{M}^{\mathbb{K}}} \prod_{k\in\mathbb{K}} f(k, m_k) = \prod_{k\in\mathbb{K}} \left\{ \sum_{m\in\mathbb{M}} f(k, m) \right\}.$$

Proof If we expand out the right-hand side, we find that we get a sum over terms of the form $\prod_{k\in\mathbb{K}} f(k, m_k)$, where all possible values $m_k \in \mathbb{M}$ will occur. Written in terms of the $\mathbf{m} = (m_k)_{k\in\mathbb{K}}$ gives the left-hand side. □

In many cases, the expression will be convergent.

1.2.4 Maxwell–Boltzmann Statistics

Here we assume that the particles are all distinguishable. Suppose that the jth particle has energy $\varepsilon_{k(j)}$, then the sequence of numbers $\mathbf{k} = (k(1), \ldots, k(N))$ determine the state of the gas. In particular, the set of all possible configurations is

$$\Omega_{N,K} = \{1, \ldots, K\}^N,$$

and we have $\#\Omega_{N,K} = K^N$. We give a Boltzmann weight to a state $\mathbf{k} \in \Omega_{N,K}$ of $e^{-\beta E(\mathbf{k})}$, where the total energy is $E(\mathbf{k}) = \sum_{j=1}^N \varepsilon_{k(j)}$. We shall be interested in the canonical partition function

$$Z_N(\beta) = \sum_{\mathbf{k}\in\Omega_{N,K}} e^{-\beta E(\mathbf{k})} = \sum_{\mathbf{k}\in\Omega_{N,K}} \prod_{j=1}^N e^{-\beta\varepsilon_{k(j)}} \equiv \left(\sum_{k=1}^K e^{-\beta\varepsilon_k} \right)^N,$$

where we used the $\sum\prod \longleftrightarrow \prod\sum$ Lemma for the last part. The associated grand canonical partition function is

$$\Xi(\beta, z) = \sum_{N=0}^\infty z^N Z_N(\beta) = \frac{1}{1 - z\sum_{k=1}^K e^{-\beta\varepsilon_k}}.$$

This was recognized as leading to an unphysical answer as some of the thermodynamic potentials (in particular, the entropy) ended up being nonextensive. This was known as the Gibbs paradox, and resolution was to apply a correction factor $1/N!$ to each $Z_N(\beta)$, nominally to account for indistinguishability of the gas particles and to crudely correct for overcounting of possibilities. This now leads to the physically acceptable form

$$\Xi\left(\beta,z\right)=\sum_{N=0}^{\infty}\frac{1}{N!}z^{N}Z_{N}\left(\beta\right)=\exp\left\{z\sum_{k=1}^{K}e^{-\beta\varepsilon_{k}}\right\}.$$

1.2.5 Bose–Einstein Statistics

Bosons are fundamentally indistinguishable particles. We are able to say, for instance, that n_k particles have the kth energy value ε_k, for each k, but physically there is no more detailed description to give – the particles have no identities of their own beyond that. The set of all possible configurations is therefore

$$\Omega^{+}_{N,K}=\left\{(n_1,\dots,n_K)\in(\mathbb{N}_+)^K:\sum_{k=1}^{K}n_k=N\right\}$$

where $\mathbb{N}_+\triangleq\{0,1,2,\dots\}$. In particular, we have $\#\Omega^{+}_{N,K}=\binom{N+K-1}{N}$. The energy associated with a state $\mathbf{n}=(n_1,\dots,n_K)$ is then $E\left(\mathbf{n}\right)=\sum_{k=1}^{K}\varepsilon_k n_k$, and we are led to the Boson canonical partition function

$$Z^{+}_{N}\left(\beta\right)=\sum_{\mathbf{n}\in\Omega^{+}_{N,K}}e^{-\beta E(\mathbf{n})}=\sum_{\mathbf{n}\in\Omega^{+}_{N,K}}\prod_{k=1}^{K}e^{-\beta\varepsilon_j n_k}.$$

This time the associated grand canonical partition function is

$$\begin{aligned}\Xi^{+}\left(\beta,z\right)&=\sum_{N=0}^{\infty}z^{N}Z^{+}_{N}\left(\beta\right)\\&=\sum_{(n_1,\dots,n_K)\in(\mathbb{N}_+)^K}\prod_{k=1}^{K}\left(ze^{-\beta\varepsilon_k}\right)^{n_k}\\&=\prod_{k=1}^{K}\sum_{n=0}^{\infty}\left(ze^{-\beta\varepsilon_k}\right)^{n}\\&=\prod_{k=1}^{K}\frac{1}{1-ze^{-\beta\varepsilon_k}},\end{aligned}\qquad(1.1)$$

where again we use the $\sum\prod\longleftrightarrow\prod\sum$ Lemma at the last stage. The thermodynamic potentials have the correct scaling properties and we do not have to resort to any ad hoc corrections of the type needed for Maxwell–Boltzmann statistics.