

Chapter 1

Introduction

Probability theory has been used in physics since the work of Maxwell, Boltzmann and Gibbs. However, many-body theory was not very successful until Planck introduced the quantum of action. Then many puzzles about black-body radiation, specific heats, and the like, were resolved. Looking today at Planck's argument we see that he used classical probability theory, even though his energy was quantised, to get the famous Planck black-body law. The same can be said of Einstein's seminal paper on stimulated emission. Even Bose's paper on the derivation of Planck's law from radiation theory is a correct use of classical probability, if we admit that we are discussing the statistics of field configurations, rather than configurations of particles. It is not clear that Bose realised the difference. It was Einstein who noticed that what Bose had done was to use a new counting rule for identical particles (assuming that photons were particles). At first, the new quantum mechanics was regarded as a modification of *mechanics*, not probability theory. Indeed, it was not until **1933** that Kolmogorov formally defined what a probability theory is. The upshot of Bell's notable work is that the Copenhagen interpretation of quantum theory, formulated between **1925** and **1930**, is not actually a consistent set of rules within classical probability, but needs a generalisation, which we call *quantum probability*. This was formalised by von Neumann in his book [1] before Kolmogorov set up the foundations of the classical theory. There were two reasons for the lack of success of the statistical mechanics of classical particles; the first is the continuous nature of the phase space, with the result that the statistical entropy of a distribution is infinite. In statistical mechanics, we hope to identify the statistical entropy with the experimental entropy, which is finite. We may introduce a coarse-graining by replacing phase-space by a discrete set of points, such as the set of labels for phase-cells; but then the entropy depends on the choice of coarse-graining, and increases as we make the divisions finer and finer. The second trouble with the statistical mechanics of Newtonian particles

comes from their distinguishability. Thus, if there are two particles in three-dimensional space, phase-space is $\Gamma = \mathbf{R}^{12}$; the configuration $(\mathbf{x}_1, \mathbf{p}_1; \mathbf{x}_2, \mathbf{p}_2)$ is taken to be a different point in Γ from $(\mathbf{x}_2, \mathbf{p}_2; \mathbf{x}_1, \mathbf{p}_1)$ (unless the points 1 and 2 coincide). A consequence of this was noticed by Gibbs in the early days, and is known as the Gibbs paradox. The problem arises when we try to relate the physical entropy of a gas (say) to the statistical entropy. To define the latter, we need a sample space Ω with sample points $\omega \in \Omega$, and a distribution $p(\omega)$, that is, a probability measure on Ω . The Shannon entropy, for a discrete space, is

$$S(p) = - \sum_{\omega} p(\omega) \log(\omega).$$

The entropy is measured in bits if the logarithm is taken to base two, and could be infinite; it is always non-negative. If the sample space has a finite number of elements, N , then the probability giving rise to the largest entropy is the uniform distribution

$$p(\omega) = 1/N.$$

Then the entropy is $\log N$, which should be compared with Boltzmann's formula for the physical entropy of a gas in equilibrium,

$$S = k, \log N$$

where here N is the number of configurations of a given number of the molecules with a given total energy, and k , is Boltzmann's constant. For a system in the usual continuum \mathbf{R}^3 the number of configurations is infinite, but if we divide space into cells (and do the same in momentum space), and restrict to a finite volume, then the entropy can be calculated. Gibbs' paradox arises when we compare the number of Configurations of n particles in m cells in phase space with the number of configurations of $2n$ particles in $2m$ cells, corresponding to twice the volume in r-space and the same volume in momentum space. According to thermodynamics, which is confirmed by experiment, the physical entropy of the second system is twice that of the first; we say that entropy is an *extensive* variable. Looking at Boltzmann's formula, we see that for this to be **true**, the number of configurations in the second system must be N^2 , the product of the numbers in the two halves. This is just not true for Newtonian particles. Not only can each of the N configurations in the first half be accompanied by each of the N configurations in the second half (already giving us the N^2 configurations), but a particle in the first half of space can be in any cell in the second half, and vice versa; there are far more than N^2 configurations, and the entropy of twice the volume is more than twice that of its parts (other things being equal). Gibbs called this extra entropy *the entropy of mixing*. He noted that its presence is not observed experimentally when two gases of identical molecules are mixed. Gibbs' paradox has been clearly discussed by [2]. It is not so much

a paradox within probability theory as a demonstration that molecules do not obey classical mechanical laws, given that we have faith in the use of probability theory. Gibbs' proposed solution to his paradox is to divide the number of configurations by suitable factorials. In modern parlance, this amounts to altering the configuration space of n particles, by collecting together the $n!$ configurations that differ only by a permutation of the particles into an equivalence class; one then takes the whole equivalence class as a single point in a new sample space. The Gibbsian sample space has $N/n!$ points; it is the quotient of the Newtonian phase space on n particles by the permutation group $S(n)$, which acts on it. It was not appreciated at the time, but this is tantamount to introducing Bose statistics for the particles, and using the field description of the configuration. For gases at room temperature and pressure, n is about 10^{23} for each cubic centimetre, so the over-counting of states in Newtonian physics is by a factor of $(10^{23})!$, one of the biggest errors in any theory still taught to students. No wonder that classical statistical mechanics **was** not able to explain even the main features of microscopic many-body systems. Statistical methods applied to Newtonian particles do give good answers for the behaviour of large numbers of ball bearings or tossed coins, which are consequently said to obey Boltzmann statistics. This is true even if the samples are so accurately similar that no difference between them can be discerned. The failure of molecules to obey Boltzmann statistics is not related to the difficulty, in practice, of following such a small individual along its path, but is due to the need for a new formulation, which we call the "field" point of view. We describe this in the next section, and contrast it with the "particle" point of view of Newton. This will resolve Gibbs' paradox; but the first difficulty, the lack of a natural unit of size for a cell in phase space, will not be resolved without quantum theory.

The first half of the book is devoted to giving the theory of classical statistical dynamics, based on a discrete sample space and using the field point of view. The objective is to provide a theory of nonequilibrium thermodynamics which obeys the first and second laws. The formalism is constructed in a way that makes clear that the classical theory is a special case, rather than a limiting case, of the quantum version, which is the subject of Part II.

We start with Chapter 2, with some remarks on probability theory. It is emphasised that the sample space must be clearly defined; one's choice of sample space, denoted Ω , is just as much part of the physics as is the choice of dynamics, and indeed must come first. The number of points in R , for example, determines the entropy of the uniform state. We also mention the concept of *local structure*. There should be a set A , usually finite, a point of which represents a point or small region in physical space; to each $x \in A$ is assigned the local sample space Ω_x . A configuration of the system, a point $\omega \in R$, is then a *field*, that is, a map ω from A such that $\omega(x) \in \Omega_x$. It is explained how this formalism avoids

Gibbs' paradox. Indeed, we prove that entropy is extensive. To make the models tractable, we choose a discrete, but possibly infinite, sample space at each $\mathbf{x} \in \Lambda$. In this context, marginal probabilities and conditional expectations are entirely elementary, and can be introduced gently. The idea of a random variable, often disguised as a mystery in books on Physics, is simply a real function on Ω . To get a model, we must specify a random variable, \mathbf{E} , the energy; this divides Ω into energy-shells Ω_E ; this is the set of all sample points with energy E . We shall always assume that Ω_E is finite; this is needed for thermodynamic stability, in that the microcanonical state, known to physicists as an "ensemble", then exists. The assumption of finiteness also enables us to avoid measure theory and the many niceties of functional analysis.

The bounded random variables form an algebra, always denoted by \mathbf{d} ; its dual, the set of linear functionals on this algebra, contains the positive cone. The normalised part of this is the set of possible states of the system. We shall limit our attention to the set of *normal* states, each given by a (countably additive) probability measure. In the finite case we can forget this refinement, and the duality between states and algebras is mutual. The characteristic of Kolmogorov's probability is that the algebra of observables is *abelian*, that is, commutative. For this reason we call this part the *classical* theory, even though there are no Newtonian particles present. The discrete sample space leads to a discrete energy-spectrum, so we are able to incorporate this aspect of quantum theory without which a classical theory would give very poor results.

The main mathematical tool is convexity, which plays a large role in the entropy-estimates, the paradigm of which is Kullback's inequality.

In Chapter 3 we define reversible dynamics as a group $\{\tau_t\}$ of invertible maps on \mathbf{R} , labelled by discrete time $t \in \mathbf{Z}$. Here \mathbf{Z} denotes the group of all integers. We prefer the term reversible to "conservative dynamics", since all our dynamics will obey the first law, conservation of energy. This is called "linear" not because the equations of motion are given by linear equations, as in a noninteracting system, but because τ_t defines a linear operator on \mathbf{d} ; more, it gives rise to a group of automorphisms of \mathcal{A} . The dynamics will be limited to energy-conserving maps. Thus τ_t maps Ω_E bijectively onto itself. This allows us to extend the duality theory (valid if $|\Omega| < \infty$) to all cases of interest to us, even if $|\Omega| = \infty$. Here we get our first success: we can construct models in which the mean energy is independent of time (the first law). We introduce random dynamics by forming the convex sum of reversible dynamics. This is a generalisation of the construction of ergodic means and leads to the class of bistochastic Markov chains with discrete time. These have the Markov property: the state at time $t + 1$ depends on the state at time t , but not on the history, that is, the state at all times before t . Not only does the discrete time lead to a readily computable theory and avoids some hard questions on the existence and

uniqueness of solutions to non-linear equations, but the time-step mitigates the Markov assumption. It may be a poor approximation to real systems to assume the Markov property if time is continuous, but if the time-step is chosen to be large on the time-scale of the fast variables, it should give a good account of the dynamics of the slow variables. This division into fast and slow variables is well explained in [3]. The dynamics given by a mixture of reversible maps is entropy-increasing, and so we get models obeying the second law. Here we meet the first “ensemble”, the microcanonical state, and show that it is the limit of any ergodic bistochastic process if the initial state has a sharp energy. This uses Lyapunov’s direct method, which is explained simply. We illustrate with models of discrete diffusion and of reacting particles.

In Chapter 4 we introduce the Boltzmann map. This is an abstract version of the Stosszahlansatz of Boltzmann. Given any algebra $\mathbf{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ (a tensor product), we can replace any state on \mathbf{A} by the product of its marginal states. This is the state of greatest entropy having the same marginal distributions as the given state. It follows that the map increases the entropy (or leaves it the same if it was already a product). This was the step in Boltzmann’s derivation of his equation that caused most controversy. We do not attempt to prove that validity, or approximate validity, of the step; rather, the step is added to the dynamics to model randomness which increases entropy but does not alter the mean energy, and which is not accounted for by the random reversible dynamics. It models the increase in entropy caused by the fast variables which break the correlations between the factors involved in the map. The resulting dynamics is non-linear on the space of states, which causes worry to some people. In particular, the powerful and popular theory of Markov chains, briefly described in Chapter 3, does not apply. But, as with Boltzmann’s original equation, the non-linearity is necessary if the isolated system is to converge to equilibrium at its own temperature. It is the Boltzmann map that mixes up the different energy-shells, and leads to the canonical state at large times. A further simplifying map, called the LTE-map, standing for *local thermodynamic equilibrium* is introduced. This increases the entropy still further, by replacing the state after one time-step by the product of states which, on each local algebra, are in a canonical state relative to the local energy. After the LTE-map the state is independent at each site \mathbf{x} . This map conserves energy only if there is no energy of interaction between sites. In that case it leads to very tractable equations of motion, in which the state is parametrised by a field of intensive variables, such as the beta and chemical potential, far simpler than the detail needed to specify a general state. The highly non-linear equations then converge to equilibrium. This is illustrated with simple models of chemical kinetics, and a free gas in a gravitational field. This last model allows us to give a natural definition of the pressure.

Then we introduce an idea possibly even more controversial than nonlinearity: the heat-particle. This is a convenient way to incorporate the first law of thermodynamics, which requires that the total energy, including heat, must be conserved by the dynamics. In statistical mechanics, heat is most often identified with random kinetic energy. More precisely, it is any form of energy that has been randomised. By introducing a new degree of freedom explicitly to carry the heat, we mitigate the omission of all velocity variables, such as would occur in the particle picture. Then we can formulate a stochastic law that conserves total energy. It is one of the tenets of statistical dynamics that in irreversible dynamics it is *information* that is dissipated as time progresses, not energy; energy simply moves between the different forms in which it can appear. In the simplest models there is just one heat-particle, which represents the shared, thermalised kinetic energy of the fields in the system. In more detailed models, the heat-particle is located on the bond between two molecules, and represents the energy of vibration of the bond, and the shared electromagnetic energy contained in the photons moving between them, and the thermalised phonons. The chapter ends with a general theory of chemical reactions. Chemical kinetics has been an active field since 1850, when the law of mass action was formulated in the form we learn at school: the rate of a reaction is proportional to the product of the concentrations of the reactants. Later, it was recognised that this could not be completely correct; the problem is that if the forward reaction rate is in balance with the backward rate, so that we are in equilibrium, then we get an equation between products of concentrations, and this equation is not satisfied by the canonical or grand canonical state. So the law of mass action was rewritten, so that the rate was proportional to the product of the *activities* of the reactants, rather than their concentrations. This fitted the data even better than the original form of the law, and, in addition, the grand canonical states are among its fixed points. The original law became known as the “density-led” form, and the new form was known as the “activity-led” version. A large part of the Chapter is used to show that the activity-led theory can be fitted into statistical dynamics by a suitable choice of sample space, energy-function, and bistochastic map. The good properties of the dynamics (energy-conservation, increasing entropy, convergence to equilibrium) then follow from our general theory.

In Chapter 5 we introduce the concept of isothermal dynamics. This arises from the isolated dynamics by adjusting the state after each time-step so that the heat-particle is in thermal equilibrium at a specified temperature. The motion then reduces to a linear dynamics through the states of the system, and so the theory can again be described by a linear stochastic process. The adjustment to the state after each time-step models the heat-flows that must occur to keep the temperature constant. We show that this is related to the Legendre

transform. The theory is applied to chemical reactions, and we quickly obtain the famous results of chemical kinetics: the principle of detailed balance, which says that for closed systems of chemical reactions at equilibrium each chemical reaction is separately in equilibrium; moreover the *law of Arrhenius*, which says that the forward and backward rates of a reaction are related by a Boltzmann factor $e^{-\beta E}$, where E is the energy of reaction, is a simple consequence of the dynamics. The main theorem is that the free energy decreases along the orbit, and **so** is a Lyapunov function. This contrasts with the isolated dynamics, in which the entropy increases. The difference between these is the heat that must be put in or taken out to maintain constant beta, and indeed we derive the axiom of thermodynamics, that entropy changes by $\geq \beta dQ$ when heat dQ is added at beta β . We use the free-energy theorem to prove the “fundamental theorem of chemical kinetics”, which states that isothermal chemical reactions converge to thermal and chemical equilibrium at large times.

We end Chapter 5 with a convergence theorem and a structure theorem. The convergence of systems to equilibrium is shown to occur in the very strong sense defined by a large norm. We also state the law of “equivalence of ensembles”. This says that the state of a subsystem of fixed size is close to a canonical state if the full system is large and is in a microcanonical state. In our version, the “closeness” is measured by the norm. This should not be construed as saying that the isothermal and isolated dynamics are nearly the same; as every chemist **knows**, isothermal dynamics needs carefully controlled heat-flows through the boundary. The distinction between the microcanonical and the canonical state disappears in the thermodynamic limit: it is one of the casualties of the idealisation implicit in the infinite-volume theory. This makes the boundary conditions difficult to specify, and some ambiguity in the dynamics seems inevitable. Ingarden [4] says “At present, many physicists think that this assumption [that the volume is infinite] is the only possible one for thermodynamical and statistical theories. Such an extreme point of view seems to be, however, absurd.”

The structure theorem says that every Markov chain with positive fixed point, and small time step, can be regarded as the isothermal dynamics of the system coupled to its heat-particle. This shows that the theory, and in particular the introduction of the heat-particle, leads by passing to its isothermal version to the most general Markov chain with the canonical state as its fixed point.

In Chapter 6 we consider driven systems; these are open in the sense of chemistry, in that particles can flow in and out, in order to keep the chemical potentials of certain chemicals fixed. These are called the driving chemicals. In driven systems the fundamental theorem of chemical kinetics does not apply. This is illustrated by heat conduction through a lattice gas. There is a brief introduction to linear stability theory and the Lyapunov indices. It is applied to a simple version of the Brussellator of Prigogine, both with and without noise.

There is also a numerical study of this model. The noise in the model is *thermal* and we show how to include it in the equations. It is shown by example that noise can destroy a limit cycle. This ends the classical part; we return to some harder problems in Chapter 13.