Statistical Mechanics Lecture 2 **DRAFT**

Information Theory

Stephen Whitcomb

whitcomb.s@northeastern.edu

Preface

Given this is the first lecture, I will briefly stop to discuss the direction and goal of these notes. As a student of physics myself, I am familiar with many of the pitfalls and confusions new students may have when approaching statistical mechanics. My hope is that by pooling insights and explanations from various resources as well as clarifying confusions I have struggled with, these notes may be helpful for other students. My expectation is that these notes should be somewhere between the level of an undergraduate course and a introductory graduate course, and will hopefully grow to include all of the material covered in a one-semester statistical mechanics course. **SOURCES THUS FAR: patrascioiu, a complicated problem in mathematics and physics; birkhoff, dynamical systems pp 154,**

Contents

1 Whats the Big Idea? **1**

2 What Physical Axioms does Statistical Mechanics Use? **2**

2.1 The Ergodic Hypothesis 2

2.2 Equal *a priori* Probabilities 4

2.3 How are these two postulates the same 4

2.4 The Failure of these Postulates in QSM 4

2.5 Entropy Maximization at Equilibrium 4

3 What is Information Theory? **4**

1 Whats the Big Idea?

Statistical mechanics is fundamentally a formalism by which we can characterize the dynamics of large systems. Whether this system is an ideal gas, a harmonic oscillator, or the famous ising model, we need to accept several nontrivial assumptions in order to use the standard machinery of statistical mechanics. Chief among these axioms are the postulate of equal *a priori* probability and the ergodic hypothesis. Although these axioms produce results that agree pretty well with experiments, the ergodic hypothesis has only been proven for very few systems (and has non-trivial examples where it doesn't hold), and it turns out that the equal a priori assumption is simply a restatement of the ergodic hypothesis for physical (Hamiltonian) systems. Many physicists rest on the laurels on statistical mechanics and don't devote much thought to the limitations of their validity. As we shall see, however, the ergodic hypothesis does not hold in all Hamiltonian systems. More importantly, as we shall see, examples where classical statistical mechanics break down (e.g. the heat capacity of a diatomic gas) are also situations where the ergodic assumption is dubious at best. **rwokaowpkda**.

2 What Physical Axioms does Statistical Mechanics Use?

2.1. The Ergodic Hypothesis

What is the ergodic hypothesis? Fortunately, there is a very intuitive physical answer,

The Ergodic Hypothesis (Informal)

The average of some observable over time is equivalent to the average of the observable over all possible configurations of the system

This hypothesis is indispensable for equilbrium statistical mechanics, and makes it easy to test the validity of statistical mechanics by comparing quantities we can measure experimentally (the time average of an observable) with the average over configurations of the observable (which is usually more feasible to calculate). As an obvious example of this, consider the energy of a system averaged over different configurations in the canonical ensemble as given by

$$\langle E\rangle_\mu = \frac{\sum_\nu E_\nu e^{-\beta E_\nu}}{\sum_\nu e^{-\beta E_\nu}} = -\frac{\partial \ln Z}{\partial \beta} \tag{1}$$

It is only by the ergodic hypothesis that we can equate the time average ($\langle E\rangle_t$) of the energy with this ensemble average ($\langle E\rangle_\mu$). This time average is what we can actually observe in experiments, so this equivalence makes it easy to translate theory into experimental predictions.

We can easily restate this hypothesis in a more mathematical language. Note, however, that this is still a very active research field in mathematics (known as ergodic theory). Accordingly, we need to present technical arguments and several terms that must be used without introduction. With that being said, this section can be skipped over on a first read. We refer the curious reader to any standard textbook on dynamical systems (ex. Birkhoff) for a more rigorous and thorough study of dynamical systems.

Throughout this treatment, we will be working with a system with N degrees of freedom. Let $\{q_1, q_2, ..., q_n\}$ denote the set of positions and $\{p_1, p_2, ..., p_n\}$ denote their corresponding canonical conjugate momenta. Clearly, the time evolution of this system will trace out some path through phase space (see fig. **fig**).

Thus, we can write

$$\langle O(\{q_i\}, \{p_i\})\rangle_t = \frac{1}{T}\int_0^T O(\{q(t)\}, \{p(t)\})dt \tag{2}$$

where $\{q_i\} = \{q(t)\}$ and $\{p_i\} = \{p(t)\}$ at t=0. This definition requires we restrict our attention to observables that are bounded along the trajectory. This is not prohibitively restrictive, but it is necessary if we want $\lim_{t->\infty}\langle O\rangle_t$ to exist.

We must now propose a mathematical definition of 'equilibrium'. In order to do this, we can think of a simple system (ex. SHM). To each point in our configuration space, we can assign a frequency with which the system visits it. Assuming there is no driving force and energy is conserved, we expect that this frequency assigned to each point will eventually converge to some value $\mu(\{q\}, \{p\})$. This is equivalent to the standard statistical mechanical definition of equilibrium (that probabilities of each state are time-independent), and will prove very useful for our development.

A theorem by Birkhoff guarantees the existence of such limits under quite general circumstances. Unfortunately, to properly describe this result would take many pages, so we encourage the reader to see Birkhoff p. 154. The probability that the system is in some neighborhood A of a phase point $(\{q_n\}, \{p_n\})$ is

$$P_A(\{q_n\}, \{p_n\}) = \int_A \mu(\{q_i\}, \{p_i\}) \cdot \prod dq_i dp_i \tag{3}$$

In keeping with standard notation, we expect that μ is normalized. That is,

$$\int_S \mu(\{q_i\}, \{p_i\}) \cdot \prod dq_i dp_i = 1 \tag{4}$$

We can now define the ensemble average of an observable. We write

$$\langle O\rangle_\mu = \int_S O(\{q_i\}, \{p_i\}) \cdot \mu(\{q_i\}, \{p_i\}) \cdot \prod dq_i dp_i \tag{5}$$

Furthermore, the theorem by Birkhoff states that, for many initial conditions, the following relation holds

$$\lim_{t\to\infty} \langle O(\{q_i\}, \{p_i\})\rangle_t = \langle O\rangle_\mu \tag{6}$$

Note, Birkhoff's theorem only implies this relation holds true for many initial conditions. If this relation holds for *almost all* initial conditions (that is, the set of initial conditions for which this relation does not hold is of measure zero), then we say the flow is *fully ergodic.* In equilibrium statistical mechanics, this assumption that we can replace time averages (experimentally obtainable) with configuration averages (generally more computationally tractable) is aptly named the *ergodic hypothesis.*

Now that we've explicitly described what ergodicity is, the next question is, "are all systems ergodic?". The answer to this is a resounding *no.* Imagine you flip a coin at t=0, record the result, flip the coin at t=1, and so on. This system **would** be ergodic— that is, the time average of the coin flip (0 if heads is +1 and tails is -1) would be the same as the average over configurations. Now imagine the coin flip is not repeated after t=0, but the state is still sampled every second. Then the time average for any given starting configuration is still either +1 or -1, whereas the configuration average is still zero. Thus, the ergodic hypothesis is not obeyed in this system.

This coin flip example seems highly pathological and perhaps of no worry to a physicist. Consider instead a 2d harmonic oscillator with rational frequencies $\omega_1 = \frac{a}{b}, \omega_2 = \frac{c}{d}$. Such a system is clearly non-ergodic, as the initial conditions obviously determine which portions of the phase space are accessible before the system forms a loop in phase space. There is <u>much</u> more that can be said about this and about mathematically demonstrating ergodicity, but it is widely beyond the purview of these notes.

2.2. Equal *a priori* Probabilities

The principle of *equal a priori probabilities* (PEAPP) is a simple postulate with far-reaching consequences. To be explicit, the

2.3. How are these two postulates the same

2.4. The Failure of these Postulates in QSM

2.5. Entropy Maximization at Equilibrium

However, none of these approaches directly address why this entropy maximization assumption is physically realistic. Of course

Let us now present an alternative argument from an information theoretic perspective.

$$S_{Boltzmann} = -k_B \sum_{\nu} P_\nu \ln P_\nu \tag{7}$$

However, we will

3 What is Information Theory?

Information theory, as the name suggests, is the study of **information**. Let (x_i, p_i) be a probability distribution. If f is some arbitrary function and we're told given the value of

$$\langle f(x) \rangle = \sum_i x_i \cdot p_i \tag{8}$$

can one then determine what $\langle g \rangle$ is for some arbitrary g? Generally, no. Here, two constraints have been provided to solve for arbitrarily many probabilities. In fact, if there are N discrete values of x, we unsurprisingly need (N-2) additional constraints to determine the value of $\langle g \rangle$.

This class of problem, where one tries to specify probabilities when little information is known, can be traced back hundreds of years. The natural solution, adopted by Laplace in his "Principle of Insufficient Reason" is to assume the probability of every outcome is uniform. However, this choice is just as arbitrary as any other! In fact a uniform distribution can easily be changed to a non-uniform distribution by a smooth change of variables (see fig. 1), so such a prescription is not even well defined depending how one defines coordinates. Clearly, a different approach is needed.

Instead, the modern solution is generally to pick the probability distribution that minimizes the likelihood of being wrong. More precisely, we should guess a probability distribution (given certain constraints) which maximizes uncertainty in the underlying probability distribution. Really, any other choice is an inappropriate assumption about additional constraints and should be avoided. Fortunately, Shannon discovered a unique quantity which is positive, increases with increasing uncertainty, and is additive for independent sources of uncertainty. We can write

$$S(p_1, p_2, \ldots, p_n) = -K \sum_i p_i \log(p_i) \tag{9}$$

or, for a probability distribution,

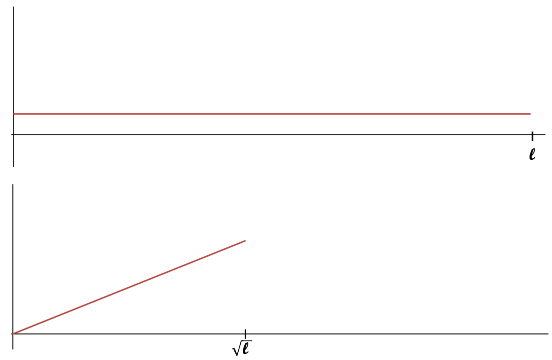

Figure 1: The uniform distribution from 0 to l changes qualitatively under the coordinate transformation $x \to \sqrt{x}$

$$S(p(x)) = -K \int x \log(x) dx \tag{10}$$

where $K \in \mathbb{R}$ is some constant.

Fortunately, identifying some p(x) that maximizes this quantity given constraints is not, in general, difficult. All one must do is write

$$a \tag{11}$$