1 Fundamentals of Statistical Physics

"I know nothing ... nothing" - John Banner

1.1 Ignorance, Entropy and the Ergodic Theorem

Consider a large number of systems $N_s \to \infty$, each of which can be in some state specific quantum state. Let n_i be the number of systems that are in the state i. We will define the *ignorance* I as a measure of the number of ways to arrange the systems given $n_0, n_1 \cdots$.

$$I = \frac{N_s!}{n_0! n_1! \cdots}, \tag{1.1}$$

with the constraint that $n_0 + n_1 + \cdots = N_s$. Our immediate goal is to find n_i that maximizes ignorance while satisfying the constraint. However, before doing so, we will define S as:

$$S \equiv \frac{1}{N_s} \ln(I), \tag{1.2}$$

which will be maximized when I is maximized, but by defining it as the log of the ignorance, the entropy will have some convenient properties which we will see below. The quantity S is the *entropy*, the most fundamental quantity of statistical mechanics. It is divided by the number of systems so that one can speak of the entropy in an individual system. Using Stirling's expansion,

$$\lim_{N\to\infty} \ln N! = N \ln N - N + (1/2)\ln N + (1/2)\ln(2\pi) + 1/(12N) + \cdots, \tag{1.3}$$

we keep the first two terms to see that

$$\begin{aligned} S &= \frac{1}{N_s}\left(N_s \ln N_s - \sum_i n_i \ln n_i - N_s + \sum_i n_i + \cdots\right) \\ &= -\sum_i p_i \ln p_i \quad \text{as } N_s \to \infty, \end{aligned} \tag{1.4}$$

where $p_i \equiv n_i/N_s$ is the probability a given system is in state i. As $N_s \to \infty$, all terms beyond the four expressed in the first line of Eq. (1.4) above vanish. Note that if all the probability is confined to one state, the entropy will be zero. Furthermore, since for each probability, $0 < p_i \leq 1$, the entropy is always positive.

Our goal is to maximize S. Maximizing a multi-dimensional function (in this case a function of $n_0, n_1 \cdots$) with a constraint is often done with Lagrange multipliers. In that case, one maximizes the quantity, $S - \lambda C(\vec{n})$, with respect to all variables and with respect to λ. Here, the constraint C must be some function of the variables constrained to zero, in our case $C = \sum_i p_i - 1$. The coefficient λ is called the Lagrange multiplier. Stating the minimization,

$$\begin{aligned} \frac{\partial}{\partial p_i}\left(-\sum_j p_j \ln p_j - \lambda\left[\sum_j p_j - 1\right]\right) &= 0, \\ \frac{\partial}{\partial \lambda}\left(-\sum_j p_j \ln p_j - \lambda\left[\sum_j p_j - 1\right]\right) &= 0. \end{aligned} \tag{1.5}$$

The second expression leads directly to the constraint $\sum_j p_j = 1$, while the first expression leads to the following value for p_i,

$$\ln p_i = -\lambda - 1, \quad \text{or} \quad p_i = e^{-\lambda-1}. \tag{1.6}$$

The parameter λ is then chosen to normalize the distribution, $e^{-\lambda-1}$ multiplied by the number of states is unity. The important result here is that all states are equally probable. This is the result of stating that you know nothing about which states are populated, i.e., maximizing ignorance is equivalent to stating that all states are equally populated. This can be considered as a fundamental principle – *Disorder (or entropy) is maximized.* **All statistical mechanics derives from this principle**.

ASIDE: REVIEW OF LAGRANGE MULTIPLIERS

Imagine a function $F(x_1 \cdots x_n)$ which one minimizes w.r.t. a constraint $C(x_1 \cdots x_n) = 0$. The gradient of F projected along the hyper-surface of the constraint must vanish, or equivalently, ∇F must be parallel to ∇C. The constant of proportionality is the Lagrange multiplier λ,

$$\nabla F = \lambda \nabla C.$$

The two gradients are parallel if,

$$\nabla (F - \lambda C) = 0.$$

However, this condition on its own merely enforces that $C(x_1 \cdots x_n)$ is equal to some constant, not necessarily zero. If one fixes λ to an arbitrary value, then solves for $\vec{x}$ by solving the parallel-gradients constraint, one will find a solution to the minimization constraint with $C(\vec{x})$ = some constant, but not zero. Fixing $C = 0$ can be accomplished by additionally requiring the condition,

$$\frac{\partial}{\partial \lambda}(F - \lambda C) = 0.$$

Thus, the n-dimensional minimization problem with a constraint is translated into an $(n+1)$-dimensional minimization problem with no constraint, where λ plays the role as the extra dimension.

The principle of maximizing entropy is related to the Ergodic theorem, which provides the way to understand why all states are equally populated from the perspective of dynamics. The Ergodic theorem assumes the symmetry of time-reversal, i.e., the rate at which one changes from state i to state j is the same as the rate at which one changes from state j to state i. If a state is particularly difficult to enter, it is equivalently difficult to exit. Thus, a time average of a given system will cycle through all states and, if one waits long enough, the system will spend equal amounts of net time in each state.

Satisfaction of time reversal is sometimes rather subtle. As an example, consider two large identical rooms, a left room and a right room, separated by a door manned by a security guard. If the rooms are populated by 1000 randomly oscillating patrons, and if the security guard grants and denies access with equal probability when going right-to-left vs. left-to-right, the population of the two rooms will, on average, be equal. However, if the security guard denies access to the left room while granting exit of the left room, the population will ultimately skew towards the right room. This explicit violation of the principle of maximized entropy derives from the fact that moving left-to-right and right-to-left, i.e. the time reversed motions, are not treated equivalently.

The same security guard could, in principle, police the traversal of gas molecules between two partitions of a box. Such paradoxes were discussed by Maxwell, and the security guard is referred to as Maxwell's demon. As described by Maxwell,

> ... if we conceive of a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are as essentially finite as our own, would be able to do what is impossible to us. For we have seen that molecules in a vessel full of air at uniform temperature are moving with velocities by no means uniform, though the mean velocity of any great number of them, arbitrarily selected, is almost exactly uniform. Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower molecules to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction to the second law of thermodynamics.

This apparent violation of the second law of thermodynamics was explained by Leó Szilárd in 1929, who showed that the demon would have to expend energy to measure the speed of the molecules, and thus increase entropy somewhere, perhaps in his brain, thus ensuring that the entropy of the entire system (gas + demon) increased. Check out http://en.wikipedia.org/wiki/Maxwell's_demon.

We have defined the entropy with logarithms in such a way that it is additive for two uncorrelated systems. For instance, we consider a set of N_a systems of type a which can be arranged I_a ways, and a second independent set of N_b systems of type b which can be arranged I_b ways. The combined systems can be arranged $I = I_a I_b$ number of ways, and the entropy of the combined systems is $S = \ln I = S_a + S_b$.

1.2 Statistical Ensembles

The previous section discussed the manifestations of maximizing ignorance, or equivalently entropy, without regard to any constraints aside from the normalization constraint. In this section, we discuss the effects of fixing energy and/or particle number. These other constraints can be easily incorporated by applying additional Lagrange multipliers. For instance, conserving the average energy can be enforced by adding an extra Lagrange multiplier β related to fixing the average energy per system. Minimizing the entropy per system with respect to the probability p_i for being in state i,

$$\frac{\partial}{\partial p_i}\left(-\sum_j p_j \ln p_j - \lambda[\sum_j p_j - 1] - \beta[\sum_j p_j \epsilon_j - \bar{E}]\right) = 0, \qquad (1.7)$$

gives

$$p_i = \exp(-1 - \lambda - \beta\epsilon_i). \qquad (1.8)$$

Thus, the states are populated proportional to the factor $e^{-\beta\epsilon_i}$, which is the Boltzmann distribution, with β being identified as the inverse temperature. Again, the parameter λ is chosen to normalize the probability. However, again the Lagrange multipliers for a given β only enforce the constraint that the average energy is some constant, not the particular energy one might wish. Thus, one must adjust β to find the desired energy, a sometimes time-consuming process.