Multi-User Information Theory 2	Oct 31, 2013

Statistical Mechanics and Information Theory

Lecturer: Dror Vinkler   *Scribe: Dror Vinkler*

I. INTRODUCTION TO STATISTICAL MECHANICS

In order to see the need for a statistical approach to mechanics, let's look at a system of N particles, each having a property called "a spin", which can be $\pm m$. Each of those particles contribute to the total magnetic moment of the system, such that the total moment is given by

$$M = \sum_{i=1}^{N} m_i. \tag{1}$$

Definition 1 *(Spin Excess)* For a system with N particles with spins $\pm m$, the spin excess is given by how many particles more than $N/2$ have spin $+m$, i.e. $S = N_\uparrow - \frac{N}{2}$ where $N_\uparrow$ is the number of particles with spin $+m$.

From Definition 1 and Equation (1) it follows that the total magnetic moment is given by $M = 2mS$.

Theoretically, one could calculate the system's magnetic moment at some point in time by using Newton's laws for each particle (given the forces that act on each particle). However, as there are about 10^{23} particles in a gram of matter, this is not very practical. A different approach would be to use statistical tools to find the *probability* of having a certain magnetic moment at some point in time. Since the total magnetic moment is a deterministic function of the spin excess, this is equivalent to finding the probability of a certain spin excess.

Definition 2 *(Multiplicity Function)* The multiplicity function states how many microscopic states (a specific set of spins) result in the same macroscopic state (a certain magnetic moment). This function is marked by $g(S, N)$.

Lemma 1 The multiplicity function is given by

$$g(S, N) = \binom{N}{N/2 + S} = \frac{N!}{(N/2 + S)!(N/2 - S)!}. \tag{2}$$

Looking at this system of spins as a sequence of bits, a microscopic state is equivalent to a specific sequence x^n and a macroscopic state is equivalent to some type P. The multiplicity function then becomes the size of the type class defined by S. I.e. $P = (\frac{1}{2} + \frac{S}{N}, \frac{1}{2} - \frac{S}{N})$ and $g(S, N) = |T(P)|$.

Lemma 2 Under the assumption that the probability of each microscopic state is the same, and noticing that there are a total of 2^N possible microscopic states, the probability of finding the system with a given spin excess is

$$P^N(S) = \frac{g(S, N)}{2^N}. \tag{3}$$

Lemma 3 For $S \ll N$, $P^N(S)$ is a normal distribution with a mean of 0 and variance of $N/4$.

Proof: For $S \ll N$ the Stirling approximation applies and it follows that

$$\ln \frac{P^N(S)}{P^N(0)} = \ln \frac{(N/2)!(N/2)!}{(N/2 + S)!(N/2 - S)!}$$

$$= 2\left(\frac{N}{2}\ln\frac{N}{2} - \frac{N}{2}\right) - \left(\frac{N}{2} + S\right)\ln\left(\frac{N}{2} + S\right)$$

$$+ \left(\frac{N}{2} + S\right) - \left(\frac{N}{2} - S\right)\ln\left(\frac{N}{2} - S\right) + \left(\frac{N}{2} - S\right)$$

$$= N\ln\frac{N}{2} - \left(\frac{N}{2} + S\right)\ln\left(\frac{N}{2} + S\right) - \left(\frac{N}{2} - S\right)\ln\left(\frac{N}{2} - S\right)$$

$$= -\left(\frac{N}{2} + S\right)\ln\left(1 + \frac{2S}{N}\right) - \left(\frac{N}{2} - S\right)\ln\left(1 - \frac{2S}{N}\right)$$

$$\overset{(a)}{\approx} -\left(\frac{N}{2} + S\right)\left(\frac{2S}{N} - \frac{2S^2}{N^2}\right) + \left(\frac{N}{2} - S\right)\left(\frac{2S}{N} + \frac{2S^2}{N^2}\right)$$

$$= -\frac{2S^2}{N}, \tag{4}$$

where (a) follows from the Taylor expansion near $\frac{2S}{N} = 0$. Thus

$$P^N(S) = P^N(0)e^{-\frac{2S^2}{N}}. \tag{5}$$

From Lemma 3 follows that the system will most likely have a zero magnetic moment, and that there's a very low chance of finding the system with a high excess spin relative to the number of particles.

We will now use the insights we have gained so far to analyze systems which are a bit more complex. The analysis is done by setting certain parameters of the system as constant, and analyzing the behaviour of the system under these conditions. Depending on what parameters are constant, the system is said to be in a different "Thermodynamic Ensemble".

II. THE MICROCANONICAL ENSEMBLE

In the microcanonical ensemble, the constant parameters are the system's total energy U, volume V, and number of particles N. In order to visualize such a system, let's consider a rigid box completely isolated from its environment. This box is divided into two parts by a divider which allows only energy to flow from one part to the other (see Fig. 1). In other words, N_1, N_2, V_1, V_2 are all constant, and only U_1 can change over time. Notice that, since U is constant, we can write $U_2 = U - U_1$ and use only one variable to describe our system over time.

$$\boxed{\begin{array}{c|c} N_1, V_1, U_1 & N_2, V_2, U - U_1 \end{array}}$$
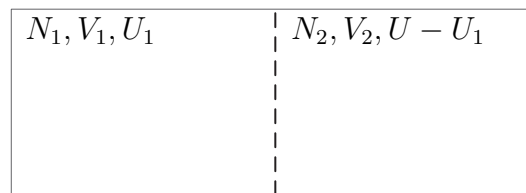
Fig. 1. The microcanonical ensemble model system - a system isolated from its environment divided into two subsystems, where only energy can flow from one subsystem to another.

Suppose this box contains spin particles as discussed in the previous section, and there's an external magnetic field B which affects the system's energy according to the formula $U = -MB$. Thus, since U is constant, M must also be constant, which means the macroscopic state of the box is constant. However, the macroscopic state of each side of the box can change, with the only limitation being that $M_1 + M_2 = M = const$.

Lemma 4 The probability of finding the left part of the box with some energy U_1 is

given by

$$P(U_1) = \frac{g(U_1, N_1)g(U - U_1, N_2)}{g(U, N)}. \tag{6}$$

Proof: The left part of the box has an energy of U_1, which means it a the spin excess $S_1 = -U_1/2mB$ and that the right part of the box has an energy of $U - U_1$. According to Equation (3), the probability of this is given by the number microscopic states that yield this macroscopic result, divided by the number of all possible microscopic states.

As this is a multiplication of Gaussians, the expected value will be the same as the most probable value. The most probable value of U_1 is given by the point where the derivative of $P(U_1)$ is zero, but for ease of calculation the derivative of $\ln P(U_1)$ will be considered. Setting

$$\frac{\partial}{\partial U_1} \ln[g(U_1, N_1)g(U - U_1, N_2)] = 0 \tag{7}$$

yields the following equality

$$\frac{\partial}{\partial U_1} \ln g(U_1, N_1) = \frac{\partial}{\partial(U - U_1)} \ln g(U - U_1, N_2). \tag{8}$$

Definition 3 *(Entropy)* The entropy of a system is marked σ and defined as the logarithm of the multiplicity function. I.e.

$$\sigma = \ln g(U, N) \tag{9}$$

Definition 4 *(Temperature)* The temperature of a system is marked τ and defined through the derivative of the system's entropy by the system's energy, in the following manner

$$\frac{1}{\tau} = \frac{\partial\sigma}{\partial U} \tag{10}$$

Theorem 1 When two systems are isolated from the environment and attached to one another such that only energy can flow between them, the expected state is the one where the systems have the same temperature.

Proof: Follows directly from Equation (8) and the definition of temperature.

Looking at our system again as a sequence of equiprobable bits, and remembering that the multiplicity function is now the size of the type class, we see the definition of entropy given here agrees with what we know about the size of type classes, namely that $|T(P)| = 2^{nH(X)}$. Also, this microcanonical setting can be seen as taking a sequence

of bits and dividing it randomly to two subsets of constant lengths N_1, N_2. Substituting $g(\cdot)$ in Equation (8) with the Gaussian we found in Lemma 3, and replacing U with its equivalent S, one finds that the expected state is that both sequences will have the same type.

Theorem 2 *(The First Law of Thermodynamics)* In a microcanonical system where the entropy depends on the volume V and the energy U, the full differential of the energy is

$$dU = \tau d\sigma - pdv, \tag{11}$$

where p is pressure.

Theorem 3 *(The Second Law of Thermodynamics)* In an isolated system, over a long enough period of time, the entropy always rises.

III. The Canonical Ensemble

In the canonical ensemble, the constant parameters are the system's temperature τ, volume V and number of particles N. The way to keep the system, marked S, in a constant temperature is to put it in contact with a much larger system, called "a heat bath" and marked R (see Fig. 2).
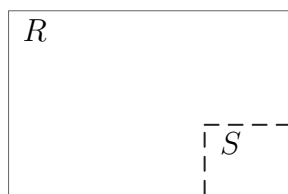

Fig. 2. The canonical ensemble model system - an isolated system divided in two: a very big heat bath, R, and a small system, S.

The system *and* the heat bath are a microcanonical ensemble, with constant energy U_0. Marking the system's energy $\mathcal{E}$, and noting $\mathcal{E} \ll U_0 - \mathcal{E}$, it is easy to see that the system's energy has negligible affect on the heat bath's energy, meaning its temperature is constant. According to Theorem 1, this means the system's temperature will be constant as well, thus realizing the canonical ensemble.

Theorem 4 The probability of the system to be in a certain microscopic state m is given by $P_S(m) = \frac{1}{Z} e^{-\frac{\mathcal{E}(m)}{\tau}}$, where Z is a normalization factor and $\mathcal{E}(m)$ is the energy of the system when its microscopical state is m.

Proof: For a given microscopic state of the system, the number of possible states overall is given only by the number of possible states of the heat bath. Thus,

$$
\begin{aligned}
P_S(m) &= \frac{1}{Z} g_R(U_0 - \mathcal{E}) \\
&= \frac{1}{Z} e^{\ln g_R(U_0 - \mathcal{E})} \\
&= \frac{1}{Z} e^{\sigma_R(U_0 - \mathcal{E})} \\
&\stackrel{(a)}{=} \frac{1}{Z} \exp\left(\sigma_R\Big|_{U_0} - \left(\frac{\partial \sigma_R}{\partial \mathcal{E}}\right)\Big|_{U_0} \cdot \mathcal{E} \right) \\
&= \frac{1}{Z} \exp\left(\sigma_R\Big|_{U_0} - \frac{\mathcal{E}}{\tau} \right) \\
&= \frac{1}{Z} e^{-\frac{\mathcal{E}}{\tau}},
\end{aligned}
\tag{12}
$$

where (a) follows from $\mathcal{E} \ll U_0$ and the Taylor expansion of σ_R.

Definition 5 *(Partition Function)* The normalization factor Z is called the partition function of the system, and is given by

$$
Z = \sum_m e^{-\frac{\mathcal{E}(m)}{\tau}}.
\tag{13}
$$

Lemma 5 Much like Shannon's entropy, the statistical mechanics entropy can be calculated using the formula

$$
\sigma = -\sum_m P(\mathcal{E}) \ln P(\mathcal{E})
\tag{14}
$$

Proof:

$$
\begin{aligned}
Z &= \sum_m e^{-\frac{\mathcal{E}(m)}{\tau}} \\
&= \sum_{\mathcal{E}} g(\mathcal{E}) e^{-\frac{\mathcal{E}}{\tau}} \\
&\stackrel{(a)}{\approx} g(\langle \mathcal{E} \rangle) e^{-\frac{\langle \mathcal{E} \rangle}{\tau}} \sqrt{N},
\end{aligned}
\tag{15}
$$

where $\langle\mathcal{E}\rangle$ is the expected value of $\mathcal{E}$ and (a) follows from the fact that $g(\mathcal{E})$ is a Gaussian of width $\sqrt{N}$. Taking the logarithm on both sides yields

$$
\ln Z = \ln g(\langle\mathcal{E}\rangle) - \frac{\langle\mathcal{E}\rangle}{\tau} + \frac{1}{2}\ln N
$$
$$
\stackrel{(a)}{\approx} \sigma - \frac{\langle\mathcal{E}\rangle}{\tau},
\tag{16}
$$

where (a) follows from the fact that $\ln N \ll \sigma$. Thus,

$$
\begin{aligned}
\sigma &= \ln Z + \frac{\langle\mathcal{E}\rangle}{\tau} \\
&= \frac{1}{Z}\sum_m e^{-\frac{\mathcal{E}}{\tau}}\ln Z + \frac{1}{\tau}\frac{1}{Z}\sum_m \mathcal{E}e^{-\frac{\mathcal{E}}{\tau}} \\
&= \frac{1}{Z}\sum_m \left(\ln Z - \ln e^{-\frac{\mathcal{E}}{\tau}}\right)e^{-\frac{\mathcal{E}}{\tau}} \\
&= -\sum_m \frac{1}{Z}e^{-\frac{\mathcal{E}}{\tau}}\ln\frac{e^{-\frac{\mathcal{E}}{\tau}}}{Z} \\
&= -\sum_m P\ln P.
\end{aligned}
\tag{17}
$$