Nick Ouellette  
Physics 114  

**Maximum Entropy Methods**

Jaynes (57), in an attempt to remove the dependencies on microscopic dynamics from statistical mechanics, proposes an interesting derivation of the canonical probability distribution using information theory as its base. This derivation never needs to appeal to postulates such as equal *a priori* probabilities or ergodicity, and instead takes the maximization of entropy to be its fundamental principle. In some ways, entropy becomes the most important characterization of a system rather than energy. Jaynes also makes the claim that this new formulation does away with the need to consider calculations on ensembles and instead allows prediction of the dynamics of single systems.

We begin with a derivation of the canonical probability distribution. Consider the Shannon entropy, defined by

\[ H = -K \sum_i P_i \ln P_i. \]  

(1)

To follow the maximum entropy hypothesis, we must maximize this entropy, subject to the definition of the expectation value,

\[ \langle f(x) \rangle = \sum_i P_i f(x_i), \]  

(2)

and the certainty condition,

\[ \sum_i P_i = 1. \]  

(3)

We can perform this maximization using the method of Lagrange multipliers, which gives us the equation

\[ \frac{\partial H}{\partial P_i} = \mu \frac{\partial}{\partial P_i} \left( \sum_i P_i f(x_i) \right) + \lambda \frac{\partial}{\partial P_i} \left( \sum_i P_i \right), \]

where \( \mu \) and \( \lambda \) are our multipliers. Setting \( K \) in equation (1) to 1 and carrying out these derivatives, we have

\[ \ln P_i + 1 = \mu f(x_i) + \lambda. \]

Solving for \( P_i \), we get

\[ P_i = e^{-\mu f(x_i) - (\lambda + 1)}. \]

We now make the definition

\[ \sum_i e^{-\mu f(x_i)} = Z(\mu), \]

and we call \( Z \) the partition function, noticing that the exponential looks very similar to a Boltzmann factor. Now, using the certainty condition, we have

\[ 1 = \sum_i P_i = \sum_i e^{-(\lambda + 1)} e^{-\mu f(x_i)} = e^{-(\lambda + 1)} Z(\mu), \]

which says that

\[ \lambda + 1 = \ln Z(\mu). \]
Now we work with equation (2), which becomes

$$\langle f(x_i) \rangle = e^{-(\lambda+1) \sum_i f(x_i)} e^{-\mu f(x_i)}.$$  

Noticing that the multiplicative factor is simply one over the partition function and the sum can be expressed as a derivative of the partition function, we have

$$\langle f(x_i) \rangle = -\frac{1}{Z(\mu)} \frac{\partial}{\partial \mu} Z(\mu) = -\frac{\partial}{\partial \mu} \ln Z(\mu),$$

which is what we would expect for the canonical ensemble. Indeed, if $\mu = \beta$, where $\beta = 1/kT$, and $f(x) = E$, it is identical. We also note that the $P_i$ have become

$$P_i = \frac{1}{Z(\mu)} e^{-\mu f(x_i)},$$

which is the canonical probability distribution. We have invoked no knowledge of microscopic dynamics in this derivation, and have used only the maximization of the entropy as a fundamental postulate.