Physical Interpretation of KL Divergence

Shi Feng

January 11, 2022

1

The relative information (or KL divergence or discrimitive information) is defined as:

$$\mathbb{D}_{KL}(p,q) = \sum_i p_i \log\left(\frac{p_i}{q_i}\right) \tag{1.1}$$

Let us assume p_i is a probability distribution at thermodynamic equilibrium, such that

$$p_i = \frac{1}{Z}\exp(-E_i/k_BT) \tag{1.2}$$

thermal dynamics tells us that the free energy is $F = U - TS$, which can be written in a more information-theorectic way:

$$F(p) = \sum_i p_iE_i + k_BT\sum_i p_i \log p_i \tag{1.3}$$

where we used $S = -\sum_i p_i \log p_i$. Now, suppose a system is out of equilibrium, which is featured by a probability distribution q_i that does not obey Boltzmann distribution. We would like to know how much its free energy is different from the equalibrium free energy $F(p)$. Remarkably, their difference is exactly proportional to $\mathbb{D}_{KL}$. The calculation is straightforward:

$$\begin{aligned}
F(q) - F(p) &= \sum_i q_iE_i + k_BT\sum_i q_i\log q_i - \sum_i p_iE_i - k_BT\sum_i p_i\log p_i \\
&= k_BT\sum_i q_i\frac{E_i}{k_BT} + k_BT\sum_i q_i\log q_i - \sum_i p_iE_i + k_BT\sum_i p_i\left(\frac{E_i}{k_BT} + \log Z\right) \\
&= k_BT\sum_i q_i\frac{E_i}{k_BT} + k_BT\sum_i q_i\log q_i - \sum_i p_iE_i + \sum_i p_iE_i + k_BT\underbrace{\sum_i p_i\log Z}_{=\log Z = \sum_i q_i\log Z} \\
&= -k_BT\sum_i q_i\log p_i + k_BT\sum_i q_i\log q_i \\
&= k_BT\sum_i q_i\log\left(\frac{q_i}{p_i}\right) \\
&= k_BT\mathbb{D}_{KL}
\end{aligned} \tag{1.4}$$

that means, the the non-equalibrium free energy differ from the equalibrium free energy by their relative information.