MSDM 5058 Information Science
Assignment 3 (due 6th April, 2024)

Submit your assignment solution on canvas. You may discuss with others or seek help from your TA, but should not directly copy from others. Otherwise, it will be considered as plagiarism.

(1) **Conditioning reduces Entropy**

In the lecture, it is stated that "Conditioning reduces entropy", i.e., $H[X|Y] \leq H[X]$, with equality iff X is independent of Y. However, it can happen that $H[X|Y{=}y] < H[X]$, or $H[X|Y{=}y] > H[X]$. We here study an example to see if this is the case.

$P(x,y)$ y \ x	1	2	3	4	$P(y)$
1	1/8	1/16	1/32	1/32	
2	1/16	1/8	1/32	1/32	
3	1/16	1/16	1/16	1/16	
4	1/4	0	0	0	
$P(x)$					

(a) Fill in the values of $P(x)$ and $P(y)$ in the table.
(b) What is the joint entropy $H[X,Y]$?
(c) What are the marginal entropies $H[X]$ and $H[Y]$?
(d) For each value of y, what is the conditional entropy $H[X|Y{=}y]$?
(e) For what values of y can one have $H[X|Y{=}y] < H[X]$? Similarly, $H[X|Y{=}y] > H[X]$.
(f) What are the conditional entropies $H[X|Y]$ and $H[Y|X]$?
(g) Compute the mutual information $I[X{:}Y]$ and verify that

$$I[X{:}Y] = H[X] - H[X|Y] = H[Y] - H[Y|X]$$

(2) **Principle of Maximum Entropy**

(a) Start with a given distribution for an "unfair" die with distribution {1/12, 1/12 , 1/6, 1/6, 1/4, 1/4}. Calculate the best guess of the distribution for the cases (i) of no information and (ii) the case of knowing only the average $\sum_{i=1}^{6} i p_i = \frac{25}{6}$.

(b) Let $p_1, p_2, \ldots, p_n$ by the probabilities of a particle having energy level $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$, respectively, where n is the number of energy levels, and let the mean value of energy be $\bar{\varepsilon}$. By maximizing the Shannon entropy,

$$-\sum_{i=1}^{n} p_i \log p_i$$

Subject to

$$\sum_{i=1}^{n} p_i = 1 \text{ and } \sum_{i=1}^{n} p_i \varepsilon_i = \bar{\varepsilon}.$$

Obtain (i) the Maxwell-Boltzmann distribution,

$$p_i = \frac{e^{-\frac{\varepsilon_i}{kT}}}{\sum_{i=1}^{n} e^{-\frac{\varepsilon_i}{kT}}}$$

and (ii) hence show that the mean energy is given by,

$$\frac{\sum_{i=1}^{n} \varepsilon_i e^{-\frac{\varepsilon_i}{kT}}}{\sum_{i=1}^{n} e^{-\frac{\varepsilon_i}{kT}}} = \bar{\varepsilon}$$

(*Hint: Start by writing down the Lagrangian similar to Lecture 9, p.74 and identify one of the Lagrange multiplier λ to be equal to -1/kT, where k is the Boltzmann constant and T is the absolute temperature.*)

(3) **Does entropy increase?**

In the lecture, you learn that entropy is a state function, which only depends on the start and end of its path, as well as macroscopic quantities such as temperature and volume. Furthermore, the **Second Law of Thermodynamics** states that entropy always increases. However, one can show that in an isolated system, no matter what non-equilibrium initial state it is, entropy computed at the microscopic level indeed *stays constant* in time. Let us see how this happens.

In classical statistical and Hamiltonian mechanics, there is the **Liouville's Theorem**, which states that the probability density of states in an ensemble of many identical states with different initial conditions is constant along every trajectory in phase space. In other words, the total time derivative of the probability is zero, i.e., $\frac{d\rho}{dt} = 0$. The equilibrium states have probability densities that only depend on energy and particle number. Furthermore, the velocity field has *zero divergence*, i.e., $\nabla.\mathbf{V} = 0$.

Puzzle: If the probability density starts in a non-equilibrium initial state, how can it evolve into an equilibrium state with largest entropy?

Denote f to be any function that depends on ρ (= $\rho(q_\alpha, p_\alpha, t)$ is the probability density distribution in phase space). $\mathbf{V} = (\dot{\mathbf{P}}, \dot{\mathbf{Q}}) = (\dot{p_\alpha}, \dot{q_\alpha})$ is the 6N-dimensional velocity in phase space, **P** and **Q** are the momentum p_α and position q_α variables. The total time derivative is given by, $\frac{d}{dt} = \frac{\partial}{\partial t} + \sum_\alpha \left(\dot{p_\alpha}\frac{\partial}{\partial p_\alpha} + \dot{q_\alpha}\frac{\partial}{\partial q_\alpha}\right)$

(a) Show that

$$\frac{\partial f(\rho)}{\partial t} = -\nabla.[f(\rho)\mathbf{V}] = -\sum_\alpha \left[\frac{\partial}{\partial p_\alpha}(f(\rho)\dot{p_\alpha}) + \frac{\partial}{\partial q_\alpha}(f(\rho)\dot{q_\alpha})\right].$$

(b) Hence, show that $\int \frac{\partial f(\rho)}{\partial t} d\mathbf{P}d\mathbf{Q} = 0$, assuming that the probability density vanishes at large momenta and positions, and $f(0) = 0$.

(c) Thus, show that the entropy $S = -k_B \int \rho \ln \rho \, d\mathbf{P}d\mathbf{Q}$ is constant in time. What is your conclusion? (*Hint: Let* $f(\rho) = \rho \ln \rho$ *and use the results in* (b))

(4) **Relative Entropy**

In the lecture, we have shown that the relative entropy is non-negative from probabilistic arguments. One can also prove this divergence inequality in a straightforward way. Let us begin by considering the following function

$$f(a) = \log a - a + 1$$

a) Show that for $a > 0$, $f(a) \leqq 0$, with equality if and only if $a = 1$.

b) Hence, show that the relative entropy,

$$D(p||q) = \sum_{x\in\chi} p(x)\log\left(\frac{p(x)}{q(x)}\right)$$

is non-negative by using the result in a).

c) We know that relative entropy is asymmetric with respect to p and q, i.e. $D(p||q) \neq D(q||p)$. Now consider again a binary communication channel where the probability of sending out a 0 and a 1 are p and $(1-p)$, while the probability of receiving a 0 and a 1 are q and $(1-q)$, respectively.

(i) Assume that $p = 0.3$ while $q = 0.7$, compute both $D(p||q)$ and $D(q||p)$. What do you get?

(ii) Hence, what is your conclusion when $D(p||q) = D(q||p)$ in this case?