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

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

Solution:

(a)

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

(b) The joint entropy is:

$$H[X,Y] = -\sum_{x,y} P(x,y)\log_2 P(x,y) \;= \frac{1}{4}\log_2 4 + \frac{1}{4}\log_2 8 + \frac{3}{8}\log_2 16 + \frac{1}{8}\log_2 32$$

$$= \frac{27}{8}\text{ bits}$$

(c) $H[X] = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{2}{8}\log_2 8 = \frac{7}{4}\text{bits}$;

$H[Y] = 4 \times \frac{1}{4}\log_2 4 = 2\text{ bits}$

(d) $H[X|Y=1] = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{2}{8}\log_2 8 = \frac{7}{4}\text{bits}$

$H[X|Y=2] = \frac{1}{4}\log_2 4 + \frac{1}{2}\log_2 2 + \frac{2}{8}\log_2 8 = \frac{7}{4}\text{bits}$

$H[X|Y=3] = \frac{4}{4}\log_2 4 = 2\text{bits}$

$H[X|Y=4] = 1\log_2 1 = 0$

(e) Compare the results in c) and d), we get $H[X|Y{=}4] < H[X]$ and $H[X|Y{=}3] > H[X]$.

(f) $H[X|Y] = -\sum_i P(Y=i)H[X|Y=i]$

$$= \frac{1}{4}(H[X|Y=1] + H[X|Y=2] + H[X|Y=3] + H[X|Y=4]) = \frac{11}{8}\text{bits}$$

Similarly, one can get $H[Y|X] = 13/8$ bits

(g) $I[X{:}Y] = H[X] + H[Y] - H[X,Y] = 7/4 + 2 - 27/8 = 3/8$ bits

$H[X] - H[X|Y] = 7/4 - 11/8 = 3/8$ bits

$H[Y] - H[Y|X] = 2 - 13/8 = 3/8$ bits , all of them give the same value.

(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 $\sum_{i=1}^{6} ip_i = \frac{25}{6}$.

(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 (a) the Maxwell-Boltzmann distribution,

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

and (b) 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.*)

Solution:

(a) Refer to Lecture 9, p.74 – 78. With no information, the best guess is a uniform distribution with each $p_i = \frac{1}{6}$. By knowing the average, we have one more constraint, and the Lagrangian becomes

$$J(p) = -\sum_{i=1}^{n} p_i \log p_i + \lambda_0 \left(\sum_{i=1}^{n} p_i - 1 \right) + \lambda_1 \left(\sum_{i=1}^{n} i p_i - \frac{25}{6} \right)$$

Take derivative with respect to p_i: $-1 - \log p_i + \lambda_0 + \lambda_1 i$ and set the derivatives to zero for the condition of maximum.

$$p_i = e^{-1+\lambda_0+\lambda_1 i}$$

Therefore,

$$1 = \sum p_i = e^{-1+\lambda_0} \sum e^{\lambda_1 i} \,; \frac{25}{6} = \sum i p_i = \frac{\sum i e^{\lambda_1 i}}{\sum e^{\lambda_1 i}} = \frac{\sum i x_i}{\sum x_i}$$

where $x_i = e^{\lambda_1 i}$. One then obtains, $\lambda_0 = -1.69905$; $\lambda_1 = 0.23634$; $p_i = \{0.085204, 0.10792, 0.136692, 0.173134, 0.219293, 0.277757\}$.

(b) The Lagrangian for the system is given by

$$J(p) = -\sum_{i=1}^{n} p_i \log p_i + \lambda_0\left(\sum_{i=1}^{n} p_i - 1\right) + \lambda_1\left(\sum_{i=1}^{n} p_i\varepsilon_i - \bar{\varepsilon}\right)$$

Take derivative with respect to p_i: $-1 - \log p_i + \lambda_0 + \lambda_1\varepsilon_i = 0$

$$p_i = \frac{e^{\lambda_1\varepsilon_i}}{e^{1-\lambda_0}}$$

$$p_i = \frac{e^{\lambda_1\varepsilon_i}}{\sum_{i=1}^{n} e^{\lambda_1\varepsilon_i}}$$

$$\frac{\sum_{i=1}^{n} \varepsilon_i e^{\lambda_1\varepsilon_i}}{\sum_{i=1}^{n} e^{\lambda_1\varepsilon_i}} = \bar{\varepsilon}$$

Substituting $\lambda_1 = -\frac{1}{kT}$, where k is the Boltzmann constant and T is the absolute temperature, we get (i) the Maxwell-Boltzmann distribution,

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

and (ii) by substituting into $\sum_{i=1}^{n} p_i\varepsilon_i = \bar{\varepsilon}.$, the mean energy is given by,

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

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

Solution:

(a) For any function f that satisfies the assumptions in the above,

$$\frac{df(\rho)}{\partial t} = \frac{\partial}{\partial t} f\big(\rho(q_\alpha, p_\alpha, t)\big) = \frac{\partial f(\rho)}{\partial \rho}\frac{\partial}{\partial t}\rho(q_\alpha, p_\alpha, t)\,.$$

Liouville's Theorem implies that along a particular trajectory,

$$\frac{d}{dt}\rho(q_\alpha(t), p_\alpha(t), t) = \frac{\partial}{\partial t}\rho(q_\alpha, p_\alpha, t) + (\mathbf{V}.\nabla)\rho(q_\alpha, p_\alpha, t) = 0\,.$$

We therefore have,

$$\frac{\partial f(\rho)}{\partial t} = \frac{\partial f(\rho)}{\partial \rho}\frac{\partial}{\partial t}\rho(q_\alpha, p_\alpha, t) = -\mathbf{V}.\left(\frac{\partial f(\rho)}{\partial \rho}\nabla\rho\right) = -\mathbf{V}.(\nabla f) = -\nabla(f\mathbf{V}),$$

since $\nabla.\mathbf{V} = 0$

(b) Using the result in (a) and by Gauss Law, we obtain

$$\int \frac{\partial f(\rho)}{\partial t} d\mathbf{P}d\mathbf{Q} = -\int \nabla(f\mathbf{V})d\mathbf{P}d\mathbf{Q} = \oint_{surface\ at\ R\to\infty} f(\rho)\mathbf{V}.d\vec{x} = 0\,.$$

The integral on the right-hand side is integrated over the whole phase space. The divergence over the whole phase space becomes a surface integral by employing the Gauss Law, and thus gives zero when the integrand falls to zero on the surface.

(c) The rate of change of the entropy is

$$\frac{dS}{dt} = -k_B \int \frac{\partial}{\partial t}(\rho \ln \rho) d\mathbf{P} d\mathbf{Q} = -\int \nabla\big((\rho \ln \rho)\mathbf{V}\big) d\mathbf{P} d\mathbf{Q} = 0$$

By taking $f(\rho) = \rho \ln \rho$, one can see that $f(\rho) \to 0$ as $\rho \to 0$. Notice that since entropy is a state function, it does not depend on the momentum and position variables. Therefore, the entropy is a constant in time with a complete microscopic treatment.

** Note that since entropy increases at the macroscopic level, this implies the increase of entropy is indeed an emergent property.

(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?

Solution:

(a) Take first and second derivative of $f(a)$, we get $f'(a) = 1/a - 1$, and $f''(a) = -1/a^2$. The first derivative has a zero when a = 1, and since the second derivative is always smaller than zero, this means it is a maximum, and $f(1) = 0$.

(b)

$$D(p||q) = \sum_{x\in\chi} p(x)\log\left(\frac{p(x)}{q(x)}\right) = -\sum_{x\in\chi} p(x)\log\left(\frac{q(x)}{p(x)}\right) \geq \sum_{x\in\chi} p(x)\left(1 - \frac{q(x)}{p(x)}\right)$$

Since

$$\sum_{x\in\chi} p(x)\left(1 - \frac{q(x)}{p(x)}\right) = \sum_{x\in\chi} p(x) - \sum_{x\in\chi} q(x) = 0 \rightarrow \ D(p||q) \geq 0$$

(c) (i) $D(p||q) = 0.3\log\left(\frac{0.3}{0.7}\right) + 0.7\log\left(\frac{0.7}{0.3}\right)$, and $D(q||p) = 0.7\log\left(\frac{0.7}{0.3}\right) +$ $0.3\log\left(\frac{0.3}{0.7}\right)$. Therefore, $D(p||q) = D(q||p)$.

(ii) For $D(p||q) = D(q||p)$, we should have

$$p\log\frac{p}{q} + (1-p)\log\frac{(1-p)}{(1-q)} = q\log\frac{q}{p} + (1-q)\log\frac{(1-q)}{(1-p)} \ .$$

Therefore, when $q = p$, or $q = 1 - p$, one will get $D(p||q) = D(q||p)$.