PHYS 273, Winter 2016, Homework 4

Due date: Thursday, February 18th, 2016

1. *Information capacity of genetic regulatory elements.* A gene regulatory module can be viewed as a device that computes an output, given an input that depends on the internal or external environment of the cell. For instance, the input could be the physical concentration c of a transcription factor (TF) that regulates the expression $g(c)$ of a gene. The intrinsic noise in gene expression imposes fundamental limits on the number of distinct environmental states that the cell can distinguish. In this problem, we will compute the information capacity of a regulatory module by treating it as a noisy, memoryless communication channel.

Experiments yield the noisy input-output relation $p(g|c)$, typically in the form of mean gene expression levels $\bar{g}(c) = \int gp(g|c)dg$ and the noise $\sigma_g^2(c) = \int (g-\bar{g})^2 p(g|c)dg$. The distribution of input concentrations $p_{\mathrm{TF}}(c)$ is unknown; to compute the capacity, we require the optimal input distribution $p^*_{\mathrm{TF}}(c)$ that maximizes the mutual information:

$$I(c;g) = \int dc\, p_{\mathrm{TF}}(c) \underbrace{\int dg\, p(g|c) \log_2 p(g|c)}_{=\mathrm{I}} - \int dc\, p_{\mathrm{TF}}(c) \underbrace{\int dg\, p(g|c) \log_2 p_{\mathrm{exp}}(g)}_{=\mathrm{II}}, \tag{1}$$

where $p_{\mathrm{exp}}(g) = \int dg\, p(g|c) p_{\mathrm{TF}}(c)$. We will assume that the input-output relation is a Gaussian distribution,

$$p(g|c) = \frac{1}{\sqrt{2\pi\sigma_g^2(c)}} \exp\left(- \frac{[g-\bar{g}(c)]^2}{2\sigma_g^2(c)} \right), \tag{2}$$

and calculate the capacity in the small-noise approximation $\sigma_g(c) \ll \bar{g}(c)$. You may proceed as follows:

a. Calculate the integral I explicity. To compute integral II, the $\log_2 p_{\mathrm{exp}}(g)$ term can be approximated for small noise by expanding it about the mean $\bar{g}$ and retaining terms to the zeroth order i.e., show that $\langle \log_2 p_{\mathrm{exp}}(g) \rangle_{p(g|c)} \approx \log_2 p_{\mathrm{exp}}(\bar{g})$. (**Optional:** show that the first order term is zero and the second order term is $\frac{1}{2}\sigma_g^2 f''(\bar{g})$, where $f(g) = \log_2 p_{\mathrm{exp}}(g)$.)

b. The integral over the input distribution $p_{\mathrm{TF}}(c)$ can be rewritten in terms of the distribution of mean expression levels $\hat{p}_{\mathrm{exp}}(\bar{g})$ using $p_{\mathrm{TF}}(c)dc = \hat{p}_{\mathrm{exp}}(\bar{g})d\bar{g}$. Assume that for small noise, $p_{\mathrm{exp}}(\bar{g}) = \hat{p}_{\mathrm{exp}}(\bar{g})$, and verify that you are left with

$$I(c;g) = -\int d\bar{g}\, \hat{p}_{\mathrm{exp}}(\bar{g}) \log_2[\sqrt{2\pi e}\sigma_g(c)\hat{p}_{\mathrm{exp}}(\bar{g})] \tag{3}$$

c. Set up a variational principle to maximize $I(c; g)$ with respect to $\hat{p}_{\text{exp}}(\bar{g})$ while ensuring its normalization. Compute the optimal solution $\hat{p}^*_{\text{exp}}(\bar{g})$ and get an explicit result for the capacity.

2. *Information rate of a grid cell.* Grid cells in the rat medial entorhinal cortex fire rapidly whenever the rat is at particular locations in its environment. In typical laboratory settings, the firing field of each grid cell is a collection of localized regions forming a grid-like structure. We will calculate the information that the firing pattern of a grid cell yields about its location in a very simplified, idealized setting.

We shall assume a simple Poisson process as a model for the spiking activity of the grid cells. Let the firing rate of a grid cell when the rat is inside and outside the firing field be r_f and r_0 respectively. We will ignore the grid-like pattern of the firing field and simply assume that the firing field of the grid cell covers a fraction p of the full environment. Thus, the probability that the rat is on the firing field of a particular grid cell at any moment is p. Calculate the rate of information gain $\dot{I}(x; s)$ that the spiking activity s yields about the rat's location x. In the limit of $r_f \gg r_0$, show that the information gain is maximized when $p \approx 1/e$. (**Hint:** To calculate the rate of information gain, calculate the mutual information between x and s in a small time interval dt and divide this quantity by dt).

3. *Second law of thermodynamics.* In statistical mechanics, we define the thermodynamic entropy as the logarithm of the number of microstates, corresponding exactly to the case of a uniform distribution for Shannon entropy. Is there an analog for the second law of thermodynamics in information theory?

Consider an isolated system modeled as a Markov chain with transitions obeying the physical laws that govern the system. We will show that the entropy of any distribution on the states of the Markov chain increases monotonically if the stationary distribution of the process is uniform. Let $p(x_n)$ and $q(x_n)$ be any two distributions on the state space of the Markov chain and $T(x_{n+1}|x_n)$ be the probability transition function. We have $p(x_n, x_{n+1}) = p(x_n)T(x_{n+1}|x_n)$ and $q(x_n, x_{n+1}) = q(x_n)T(x_{n+1}|x_n)$. (a) Show that

$$\begin{aligned} D(p(x_n, x_{n+1})||q(x_n, x_{n+1})) &= D(p(x_n)||q(x_n)) + D(p(x_{n+1}|x_n)||q(x_{n+1}|x_n)) && (4) \\ &= D(p(x_{n+1})||q(x_{n+1})) + D(p(x_n|x_{n+1})||q(x_n|x_{n+1})) && (5) \end{aligned}$$

where $D(p||q)$ is the Kullback-Leibler divergence or relative entropy between distributions p and q. Recall that the relative entropy is a measure of the distance between two distributions. (b) Using $p(x_{n+1}|x_n) = q(x_{n+1}|x_n) = T(x_{n+1}|x_n)$ and the non-negativity of the relative entropy function conclude that

$$D(p(x_n)||q(x_n)) \geq D(p(x_{n+1})||q(x_{n+1})) \tag{6}$$

The above inequality shows that the distance between any two distributions is non-increasing with time for any Markov chain. Now suppose that the stationary distribution $\mu(x)$ of the Markov chain (if it exists) is uniform. (c) Set $q(x_n) = \mu(x)$ and use

$D(p(x_n)||\mu) = \log \Omega - H(p(x_n))$ (where Ω is the total number of states) to show that the entropy is always a non-decreasing quantity with time for any initial distribution.