Assignment 7
Theoretical Neuroscience

Maneesh Sahani and Peter Dayan

Due 5 December, 2008

1. **Stochastic processes and entropy rates**

   (a) Prove that the two definitions of the entropy rate given in class:

   $$\lim_{n\to\infty} \frac{1}{n} H(X_1, \dots X_n) \qquad \text{and} \qquad \lim_{n\to\infty} H(X_n \mid X_{n-1} \dots X_1),$$

   are equivalent. [Hint: If $a_n \to a$ as $n \to \infty$, what can be said about the running averages $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$?]

   (b) Consider a point process $\mathcal{P}_\lambda$with a constant mean rate constrained to be λ. We are interested in the form of the maximum entropy process consistent with the constraint.

      i. First, consider the stochastic process defined by taking successive inter-event intervals generated by $\mathcal{P}_\lambda$. How does the constraint on $\mathcal{P}_\lambda$'s rate constrain the ISI process? What is the maximum entropy ISI process? What does this imply about $\mathcal{P}_\lambda$?

      ii. Now consider the stochastic process defined by counting events from $\mathcal{P}_\lambda$ that fall in successive intervals of length Δ. How is the mean rate constraint reflected in this counting process? What is the maximum entropy counting process under this constraint? What does this imply about $\mathcal{P}_\lambda$?

      iii. Suppose we were to expect spike trains in the brain to achieve maximum entropy with constrained spike rate. Which of the two preceding approaches to the obtaining the maximum entropy distribution is likely to be the more relevant to the brain. [Hint: how does the process obtained in the second case depend on Δ?]

2. **Communication through a probabilistic synapse**

   (a) The Blahut-Arimoto algorithm.

   In this part of the question, we derive an algorithm to find an input distribution that achieves the capacity of an arbitrary discrete channel.

      i. Given a channel characterised by the conditional distribution $P(R|S)$, we wish to find a source distribution $P(S)$ that maximises the mutual information $I(R;S)$. Show that

      $$I(R;S) \geq \sum_{s,r} P(s)P(r|s) \log \frac{Q(s|r)}{P(s)}$$

      for any conditional distribution $Q(S|R)$. When is equality achieved?

      ii. Use this result to derive (in closed form) an iterative algorithm to find the optimal $P(S)$.* This is called the Blahut-Arimoto algorithm. Prove that the algorithm converges to a unique maximal value of $I(R;S)$.

* Hint: by analogy to EM, alternate maximisations of the bound on the right hand side with respect to Q and to $P(S)$.

(b) Synaptic failure.

Many synapses in the brain appear to be unreliable; that is, they release neurotransmitter stochastically in response to incoming spikes. Here, we will build an extremely crude model of communication under these conditions.

Assume that the input to the synapse is represented by the number of spikes arriving in a 10 ms interval, while the output is the number of times a vesicle is released in the same period. Let the minimum inter-spike interval be 1 ms (taking into account both the length of the spike and the refractory period), and assume that at most 1 vesicle is released per spike. Thus, both input and output symbols on this channel are integers between 0 and 10 inclusive.

Let the probability of vesicle release be independent for each spike in the input symbol, and be given by α^n where α is a measure of synaptic depression and n is the number of spikes in the symbol. (We are neglecting order-dependent effects within each 10ms symbol, and any interactions between successive symbols. This is a terrible model of synaptic behaviour).

i. Generate (in MATLAB) the conditional distribution of output given input for this synapse. Take $\alpha = 0.9$. Use Blahut-Arimoto to derive the capacity-achieving input distribution and plot it.

ii. Try to interpret your result intuitively. Might this have anything to do with the short "bursts" of action potentials found in many spike trains?

iii. `OPTIONAL`: Improve on the model of synaptic transmission. Consider 5 ms input and output symbols, each being a 5-bit binary number where a 1 indicates a spike or a vesicle release. The probability of transmission for each spike in the symbol is again α^n but n is now the number of vesicles released *so far* for this symbol. Construct a new conditional distribution table and repeat the optimisation. Do you get a qualitatively similar result?

3. Show that the averaged form of the single-trial Oja rule in equation 8.16 is given by

$$\tau_w \frac{d\mathbf{w}}{dt} = \mathbf{Q} \cdot \mathbf{w} - \alpha(\mathbf{w} \cdot \mathbf{Q} \cdot \mathbf{w})\mathbf{w}.$$

Prove that if it converges, the averaged learning rule produces a set of weights proportional to an eigenvector of the correlation matrix $\mathbf{Q}$, normalized so that $|\mathbf{w}|^2 = 1/\alpha$

4. Simulate the ocular dominance model of figure 8.7 using a subtractively normalized version of equation 8.31 (i.e. equation 8.14) with saturation limits at 0 and 1, and cortical interactions generated as in figure 8.8 from

$$\mathbf{K}_{aa'} = exp(-\frac{(a-a')^2}{2\sigma^2}) - \frac{1}{9}exp(-\frac{(a-a')^2}{18\sigma^2})$$

where $\sigma = 0.066$ mm. Use 512 cortical cells with locations a spread evenly over a nominal 10 mm of cortex, and periodic boundary conditions (this means that you can use Fourier transforms to calculate the effect of the cortical interactions). Also use the discrete form of equation 8.31

$$\mathbf{W} \rightarrow \mathbf{W} + \epsilon \mathbf{K} \cdot \mathbf{W} \cdot \mathbf{Q}$$

with a learning rate of $\epsilon = 0.01$. Plot $\mathbf{w}_-$ as it evolves from near **0** to the final form of ocular dominance. Calculate the magnitude of the discrete Fourier transform of $\mathbf{w}_-$. Repeat this around 100 times, work out the average of the magnitudes of the Fourier transforms, and compare this to the Fourier transform of $\mathbf{K}$.