# Assignment 7
# Theoretical Neuroscience

Maneesh Sahani, Liam Paninski, Peter Dayan

19 November 2004

1. **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.

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?

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 maximum.

* 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 $\alpha^n$ where $\alpha$ is a measure of synaptic depression and $n$ is the number of spikes in the symbol. (This is a terrible model of synaptic behaviour). 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.

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

Optional: Improve on the model of synaptic transmission and repeat the optimisation. Do you get a qualitatively similar result?

2. **Entropy Estimation**

   (a) Define $\hat{H}_{\text{MLE}}$ (as in class) to be the entropy of the normalised histogram generated by $N$ samples from a discrete distribution over $M$ values, with probability mass function $p(m)$. We are interested in its properties as an estimator for the entropy of the underlying distribution $H(p)$. Prove each of the following statements made in class:

   - $\hat{H}_{\text{MLE}}$ is non-positively biased for all $p$. *
   - A rough estimate of the bias is: $-(M-1)/2N$. **
   - The variance of $\hat{H}_{\text{MLE}}$ is much smaller: $\approx (\log M)^2/N$. **
   - No unbiased estimator for the entropy exists (although asymptotically unbiased estimators do exist). ***

   * Hint: try Jensen. The bias isn't actually negative for all $p$; there are some $p$ for which the bias is zero. Which $p$ are these? What does the bias in $\hat{H}_{\text{MLE}}$ imply about the bias of $\hat{I}_{\text{MLE}}$?

   ** Hint: expand $H(p)$ about the asymptotic limit of the MLE for $p$. (This general technique of expanding around a limit is called the "delta method" in statistics, if you ever happen to run into that term.) Explain why our rough bias estimate is independent of $p$. Does this make sense, or is there something wrong here?

   *** Hint: try writing out the bias of any arbitrary estimator $\hat{H}(x)$, where $\hat{H}(x)$ is a function of the observed data $x = \{x_1, x_2, \ldots x_N\},\ x_i \in \{1, 2, \ldots m\}$. Notice any problems setting this to zero for all possible underlying $p$?

   (b) Yet more fly data: compute the Gaussian lower bound on the information rate in the spike train about the velocity signal, given the estimators you constructed in the last homework. Do your estimators show any notable differences in terms of the Fourier energy they succeed in capturing? Discuss how tight you think the lower bound is in this case. How can you a) ensure that the bound is tight; and b) if the bound isn't sufficiently tight, make it tighter?

3. **Population Coding**

   Shadlen and collaborators have claimed that if the activities of neurons in population codes are corrupted by *correlated* noise, then there is a sharp limit to the useful number of neurons in the population. *Prima facie* this is wrong – the stronger the correlations, the lower the entropy of the noise, and therefore the stronger the signal.

   Resolve this issue for the case of additive and multiplicative noise by considering the following three models for the noisy activities $r_1$ and $r_2$ of two neurons which form a population code for a real-valued quantity $x$:

$$\text{a)} \quad \begin{cases} r_1^a = x + \epsilon_1 \\ r_2^a = x + \epsilon_2 \end{cases} \tag{1}$$

$$\text{b)} \quad \begin{cases} r_1^b = x(1-\delta) + \epsilon_1 \\ r_2^b = x(1+\delta) + \epsilon_2 \end{cases} \tag{2}$$

$$\text{c)} \quad \begin{cases} r_1^c = x(1-\delta)(1+\eta_1) \\ r_2^c = x(1+\delta)(1+\eta_2) \end{cases} \tag{3}$$

   where $\delta \neq 0$ is known, and, $\epsilon$ and $\eta$ are Gaussian, with mean 0 and covariance matrices:

$$\Sigma = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$$

(a) What is the maximum likelihood (ML) estimator for $x$ on the basis of $r_1$ and $r_2$ in each case?

(b) What is the appropriate measure of the expected accuracy of the ML estimate, and why?

(c) How does the expected accuracy in each case depend on the degree of correlation $c$?

(d) What conclusions would you draw about the clash between Shadlen and common sense?