Assignment 4
Theoretical Neuroscience

Maneesh Sahani

27 October 2005

1. **The expected autocorrelation function of a renewal process.**

   In class, we analysed the autocorrelation function of a point process in terms of its intensity function $\lambda(t, \ldots)$. For a self-exciting point process, λ depends on the past history of spiking, and so computing the expected value of the correlation in this way can be quite difficult. Fortunately, for the special case of a renewal process (i.e. a point process with iid inter-event intervals), there is an alternative way to compute the autocorrelation function.

   Consider a neuron whose firing can be described by a renewal process with inter-spike interval probability density function $p(\tau)$.

   (a) Given an event at time t, the probability that the next spike arrives in the interval $I_\tau = [t+\tau, t+\tau+d\tau)$ is $p(\tau)d\tau$. What is the probability that the *second* spike after the one at t arrives in I_τ instead? The third spike?

   (b) What is the probability that, given a spike at t, there is a spike in I_τ, regardless of the number of intervening spikes?

   (c) Your answer to the previous question has given you the right half of the autocorrelation function. What does the left half look like? What happens at $\tau = 0$?

   (d) Show that for a Gamma process with ISI density

   $$p(\tau) = \beta^2 \tau e^{-\beta\tau},$$

   the Laplace transform of (the right half of) the expected autocorrelation function is

   $$\mathcal{L}[Q(\tau)](s) = \frac{\beta^2}{(\beta+s)^2 - \beta^2}.$$

   [Hint: Recall that $\mathcal{L}[f](s) = \int_0^\infty dx\ f(x)e^{-sx}$. Apply the Laplace convolution theorem, after setting $p(\tau) = 0$ for $\tau < 0$. Finally, use the fact that for $|x| < 1$, $(1-x)^{-1} = 1 + x + x^2 + x^3 + ...$]

   Find the expected power spectrum (i.e. the Fourier transform of the expected autocorrelation function) for this process.

2. **Information theory**

   (a) Use Jensen's inequality to prove that the KL divergence $\mathsf{KL}[p\|q]$ is convex in the pair (p, q); i.e., if (p_1, q_1) and (p_2, q_2) are two pairs of distributions and $a \in [0, 1]$, then

   $$\mathsf{KL}[ap_1 + (1-a)p_2 \| aq_1 + (1-a)q_2] \leq a\mathsf{KL}[p_1\|q_1] + (1-a)\mathsf{KL}[p_2\|q_2].$$

(b) Differential entropy. What is the differental entropy of the continuous distribution with density $u_a(x)$ that is uniform on the interval $[-a/2, a/2]$ and 0 elsewhere? When is this entropy equal to 0? Compare to the case of the entropy of a discrete distribution being 0. What happens when $a = 0.5$? What happens as $a \to \infty$? Do these results correspond to your intuitive notion of uncertainty in a distribution?

(c) Is the KL divergence defined on continuous distributions "well-behaved"? Calculate $KL[u_1 \| u_2]$ as well as $KL[u_2 \| u_1]$ (u_a as defined above). Interpret these results in terms of the "coding penalty" discussed in class.

3. **Stochastic processes and entropy rates**

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

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

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?]

4. **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 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 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 and repeat the optimisation. Do you get a qualitatively similar result?