Assignment 6
Theoretical Neuroscience

Maneesh Sahani

16 November 2012

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 positive half of the autocorrelation function. What does the negative 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 $|x| < 1$, $(1-x)^{-1} = 1 + x + x^2 + x^3 + ...$]

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

2. **Estimation Theory**

   (a) We derived the Fisher information $J(\theta)$ as the expected value of the second derivate (curvature) of the log-likeihood in the lecture.

      i. Repeat the derivation for a *vector* parameter (or stimulus in our setting) $\boldsymbol{\theta}$, showing that the Fisher information in this case is given by a matrix.

As mentioned in the lecture, there is an alternate definition in terms of the first derivative. For vector parameters this is:

$$J(\boldsymbol{\theta}_0) = \mathrm{Cov}_{\boldsymbol{\theta}_0}\left(\nabla \log p(n|\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0}\right).$$

where $\mathrm{Cov}_{\boldsymbol{\theta}_0}$ means the covariance evaluated under $p(n|\boldsymbol{\theta}_0)$.

(b) Demonstrate that these two definitions are the same (or more precisely, give conditions under which these two definitions are the same).

(c) Consider an LNP model:

$$p(n|\mathbf{x}) = \mathsf{Poiss}(g(\mathbf{w}\cdot\mathbf{x}))$$

i. What is $J(\mathbf{x})$ (the Fisher Information about the stimulus value available to the rest of the brain)? How does it depend on $\mathbf{w}$? Working in two dimensions (recall the picture from lecture) show how $J(\mathbf{x})$ varies around the vector linear projection vector $\mathbf{w}$.

ii. What is $J(\mathbf{w})$ (the Fisher Information about the weight vector available to an experimenter — consider the case of multiple measurements n_i, each in response to a different stimulus $\mathbf{x}_i$)? How does it depend on the distribution of $\mathbf{x}$? What would be a good distribution with which to probe the cell if we knew (say) the orthant of stimulus space in which $\mathbf{w}$ lay?

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, ϵ and η are Gaussian, with mean 0 and covariance matrices:

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

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

(b) How does the expected accuracy in each case depend on the degree of correlation c? [Hint: begin by showing that the Fisher Information for a Gaussian distribution with mean $\mu(\theta)$ and variance $\Sigma(\theta)$ both dependent on a scalar parameter θ is:

$$J(\theta) = \nabla\mu^{\mathsf{T}}\Sigma^{-1}\nabla\mu + \frac{1}{2}\mathsf{Tr}\left[\Sigma^{-1}(\nabla\Sigma)\Sigma^{-1}(\nabla\Sigma)\right]$$

where the matrix "gradient" is the matrix of elementwise derivatives.]

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