8.581 Entropy, information and the brain

Problems for Lectures 3 & 4.

---

1. Imagine that we count discrete events (spikes, synaptic vesicles, cars, ...) and that on average we count $\bar{n}$. Prove that the distribution of counts $P(n)$ that has maximum entropy given this mean count is an exponential distribution, and find the maximum entropy $S_{\max}(\bar{n})$ as a function of $\bar{n}$.

2. Suppose that we are counting spikes, so the mean count $\bar{n} = \bar{r}\Delta\tau$, where $\bar{r}$ is the mean spike rate and $\Delta\tau$ is the size of the time bins in which we count. Clearly the maximum entropy for the whole spike train is obtained if each bin is independent of every other. Thus you can compute the maximum entropy per spike as $S_{\max}(\bar{n})/\bar{n}$. Find the limiting behavior for small and large values of time time resolution $\Delta\tau$. For a cell that generates $\bar{r} = 30\,\text{spike/s}$, plot the maximum entropy per spike as function of $\Delta\tau$.

3. The major problem in computing the entropy and information from real spike trains is that one requires large amounts of data to give a good sampling of the relevant probability distributions. Suppose that we observe a total of N neural responses (words, as discussed in the lecture). Of these responses, n_i are the word labelled i, and we try to estimate $\tilde{p}_i = n_i/N$ and $\tilde{S} = -\sum_i \tilde{p}_i \log_2 \tilde{p}_i$. Assuming that the variance in the count n_i will be equal to the mean, find the systematic and random errors in this naive estimate of the entropy.

4. If we choose two responses at random, the probability that they are the same is $P_c = \sum_i p_i^2$. Prove that this probability of a coincidence can be used to give a lower bound on the entropy, $S \geq -\log_2 P_c$. Explain why this bound is much less sensitive to sampling errors than the naive estimate given above. You might try to consider the case where K possible words are equally likely (although we don't know K), and then ask if you have N examples, how accurately can you estimate P_c ... .

5. Imagine that we observe a variable y and try to estimate some underlying signal s. Prove that the optimal estimator in the least squares sense is the conditional mean. That is, if we consider all possible functions $s_{\rm est}(y)$ that we can use as estimators, show that we will minimize the mean square error,
$$\chi^2 = \langle [s_{\rm est}(y) - s]^2 \rangle,$$
if we choose the conditional mean
$$s_{\rm est}(y) = \int ds P(s|y) s.$$

6. Consider the case where $y = s + \eta$ with η a Gaussian noise source. Show that if s is also chosen from a Gaussian distribution then the optimal estimator is linear, $s_{\rm est}(y) = Ky$, and find the gain K. You might think that the best estimate is linear because the relation between y and s is linear. Show that this is wrong by studying the case where s is chosen from an exponential distribution,

$$P(s) = \frac{1}{2s_0} \exp(-|s|/s_0);$$

what is the form of the optimal estimator in this case? You might start by trying to find the most likely value of s given y, before doing the integrals.

7. Consider, as in the lectures, a neuron that generates spikes according to a Poisson process. Let the rate of this process be $r(t) = \bar{r}x^2(t)$, where $x(t)$ is a Gaussian variable with unit variance and a correlation time τ; the probability distribution for $x(t)$ is then

$$P[x(t)] \propto \exp\left[-\frac{\tau}{4}\int dt \left(\frac{dx(t)}{dt}\right)^2 - \frac{1}{4\tau}\int dt x^2(t)\right].$$

Find a formal expression for the optimal estimator of $s(t) = x^2(t)$ given the spike arrival times $\{t_i\}$, as in the lecture. You should be able to evaluate this expression exactly, and explain what approximations are required to make linear decoding a valid scheme.