Information Theory Problem Set

Dan Butts, dab@umd.edu

1 Calculation of information quantities in simple contexts

Consider a disease that afflicts a fraction f of the population. There are two tests for the disease. Test #1 gives no false positives, but only has a 50% chance of detecting the disease if it is there. Test #2 has 100% chance of detecting the disease if it is there, but 50% chance of a false positive if it is not there.

A. Write down the joint probability distribution $p(\text{disease}, \text{test result})$ for each test, in terms of the fraction f of the population that has it. Use the following shorthand:

Y = has disease, N = no disease, + = test is positive, - = test is negative

B. Suppose the incidence of the disease in the population is 10%: $f = 0.1$. How much information does each test provide? Which test provides more information, and, intuitively, why?

C. Specific information and stimulus-specific information. For test #2, what is the specific information of each response? What is the stimulus-specific information of the disease state and normal state?

Recall that the specific information of a response is given by $i_{sp}(r) = H[S] - H[S|r]$, and the stimulus-specific information of a stimulus is given by $i_{ssi}(s) = \sum_r p(r|s) i_{sp}(r)$.

Advanced: Write a general expression for the information of test #2 as a function of f. What fraction is it maximized at?

2 The information of spike trains

Consider an experiment consisting of one-second-long trials (over which a stimulus ensemble is presented), where the neuron recorded fires once and only once on every repeat of the stimulus.

A. Divide the experiment into time bins of width Δt. Considering each bin independent, what is the channel capacity $H[R]$ of that neuron. As a result, what is the maximum information rate that the neuron can provide? What happens to the information rate as $\Delta t \rightarrow 0$? Why? *(note the set handed out in class said channel capacity, but this was a typo)*

B. Assume that the PSTH of the neuron is a Gaussian of width σ, centered at 0.5 sec. For a bin width Δt, what is the mutual information rate ($= I/\Delta t$) between the stimulus (during the experiment) and the response of a single bin? What happens as $\Delta t \rightarrow 0$?

C. The single spike information is given by:

$$I_1 = \langle r(t) \log_2[r(t)/\bar{r}] \rangle_t \tag{1}$$

where $\bar{r}$ is the average firing rate, and the angle-brackets represent the average over time. What is the single-spike information for the example in B?

Advanced: Is the single spike information an over-estimate or under-estimate of the true information of the spike train? Why?

3 The information in tuning curves

[*To be done in Matlab*] Consider the model of a single cricket cercal system neuron given in the last experiment. There, the neurons tuning curve, given the *average* firing rate of the neuron in response to a particular wind direction θ is given by:

$$f(\theta) = \cos(\theta - \theta_i)]_+ \tag{2}$$

where $[]_+$ indicates half-wave rectification (turning negative values to zero). Note that here, we are normalizing firing rates so the maximum firing rate at the tuning curve peak is 1. While the tuning curve gives the average firing rate, there is "noise", such that the firing rate is not the same on every trial. For simplification, assume that noise is constant, such that for a given stimulus θ, the distribution of possible responses is given by a Gaussian centered at the average firing rate:

$$p(r|\theta) = A\exp\{[r - f(\theta)]^2)/2\sigma_n^2\} \quad (3)$$

and the width is given by a constant noise term $\sigma_n = 0.2$. Here, A is a constant that is adjusted to make sure that $\sum_r p(r|\theta) = 1$.

A. Assume a stimulus bin width of 5 degrees (so there are 72 bins), and a response bin width of 0.05 (make 30 bins so there is a maximum firing rate of 1.5). What is the mutual information between the stimulus and response?

To do this, first construct the joint probability distribution. The prior $p(\theta)$ is uniform, and the conditional distributions are given by Gaussian distributions shown above. This involves making a two dimensional matrix of the correct size:

$$\text{jpd} = \text{zeros(72,31)}; \quad (4)$$

Then, for each response distribution, use the Gaussian described above. Assume the tuning curve is centered at 180 degrees, so the nth response distribution is

$$\text{rs} = 0.05\text{*}(0{:}30); \quad (5)$$

$$\text{jpd(n,:)} = \text{exp(-(rs-f(n)).}^2)/(2 * 0.2^2)); \quad (6)$$

Note: you need to have f (the tuning curve) calculated first (as an array with 72 elements for this to work. The distribution would then need to be normalized

$$\text{jpd(n,:)} = \text{jpd(n,:) / sum(jpd(n,:))}; \quad (7)$$

and once this is done, be sure to make sure that $\sum_r \sum_\theta p(r,\theta) = 1$. Then calculate the mutual information given the joint probability distribution $p(r,\theta)$.

B. Calculate the specific information of the response $i_{sp}(r)$, defined in problem #1, and plot.

C. *OPTIONAL*: Calculate the stimulus-specific information $i_{ssi}(\theta)$, also defined above, and plot.

D. *OPTIONAL*: Multiply the noise by a factor of 4, and recalculate $i_{sp}(r)$ and $i_{ssi}(\theta)$, and plot each.