8.581 Entropy, information and the brain

Problems for Lecture 1.

---

1. **Maximum entropy property of Gaussians.** Consider a random variable x, drawn from some distribution $P(x)$. Find the distribution that maximizes the entropy, $S = -\int dx P(x) \log P(x)$, given that you know the average value of x and the variance of x. It is easiest to formulate this maximization problem by introducing Lagrange multipliers to constrain the values of
$$\langle x \rangle = \int dx P(x) x \quad \text{and} \quad \langle x^2 \rangle = \int dx P(x) x^2,$$
and remember that you also need to preserve normalization, $\int dx P(x) = 1$. The answer should be the Gaussian distribution.

2. **Information with Gaussian signals and noise.** Consider a system in which we observe a detector with output y, driven by a signal s with added noise η; for simplicity let's set the 'gain' of the detector to unity, so that
$$y = s + \eta.$$
Assume that both the signal and the noise are drawn from Gaussian distributions, and find the information that y provides about s in terms of the signal and noise variances; that is, compute the integral
$$I(y \to s) = \int ds \int dy P(y, s) \log \left(\frac{P(y, s)}{P(s)P(y)} \right).$$
You should find that $I = (1/2) \log(1 + SNR)$, where $SNR = \langle s^2 \rangle / \langle \eta^2 \rangle$, assuming that all variable have zero mean. More generally these second moments should be the variances; you should show that the mean values are irrelevant for information transmission.

3. **Multicomponent signals.** Generalize the results in the previous problem to the case where we observe a set of outputs $y_1, y_2, ...$,
$$y_i = s_i + \eta_i.$$
The signals and noise may each have correlations among their components, characterized by covariance matrices
$$\langle s_i s_j \rangle = S_{ij} \quad \text{and} \quad \langle \eta_i \eta_j \rangle = N_{ij},$$

but we assume that the distributions are still Gaussian. Show that the information y provides about s is given by

$$I(\vec{y} \to \vec{s}) = \frac{1}{2}\mathrm{Tr}\log(1 + N^{-1}S),$$

where N^{-1} is the inverse of the matrix N_{ij}, Tr is the trace of the matrix, and here 1 is to be interpreted as the unit matrix.

4. **Time dependent signals.** Consider now the case where signals depend on time, so we have

$$y(t) = s(t) + \eta(t).$$

The signals and noise are still Gaussian, and now in place of the covariance matrix we have the correlation function or its Fourier transform, the power spectrum, as described in the lecture. To compute the rate of information transmission, imagine observing signals in a finite time window of size T; then use a discrete Fourier series to decompose the functions of time and compute the information from each Fourier component, and add up contributions from all the components. Finally, you let the window T become large, and sums over discrete frequencies become integrals, just as when we some over modes in statistical mechanics. Go through all these steps! Things to be careful about: positive and negative frequencies (remember the functions here are real; try using sine and cosine components?), the justification for independence of the different components (which is correct only in the Gaussian case). You should find the rate of information transmission

$$R_{\rm info} = \lim_{T\to\infty} \frac{1}{T} I[y(t) \to s(t)] = \frac{1}{2}\int \frac{d\omega}{2\pi} \log[1 + S(\omega)/N(\omega)],$$

where $S(\omega)$ and $N(\omega)$ are the power spectra of signal and noise, respectively.

5. **Optimal Spectra.** Given the noise power spectrum $N(\omega)$, find the signal power spectrum $S(\omega)$ that maximizes information transmission while constraining the dynamic range of the signal,

$$\langle s^2 \rangle = \int \frac{d\omega}{2\pi} S(\omega).$$

This problem is discussed by Shannon, and the results are used by de Ruyter van Steveninck and Laughlin in the analysis of information transmission by the first stages of the fly visual system; both references are in the reader.