Ch.4 Information Theory

From 4.2. ~~to 4.3.~~

Zi-Seok LEE

July 12, 2023

Summer Brain Study

Histogram Equalization

Our objective is to find the probability density function $p[r]$ which maximizes

$$-\int_0^{r_{max}} dr p[r] \log_2 p[r] \tag{1}$$

subject to the constraint

$$\int_0^{r_{max}} dr p[r] = 1 \tag{2}$$

Lagrange Multipliers

How do we use Lagrange multipliers here?

$$\mathcal{L}(p, \lambda) = -\int_0^{r_{max}} dr p[r] \log_2 p[r] + \lambda \cdot \left(\int_0^{r_{max}} dr p[r] - 1\right) \tag{3}$$

Note that the function $\mathcal{L}$ defined in (0.3) is a functional (a function on the function space)

$$\partial_p \mathcal{L}(p, \lambda) = 0 \Leftrightarrow \lim_{\epsilon \to 0} \frac{\mathcal{L}(p + \epsilon h, \lambda) - \mathcal{L}(p, \lambda)}{\epsilon} = 0 \tag{4}$$

$$\partial_\lambda \mathcal{L}(p, \lambda) = 0 \Leftrightarrow \int_0^{r_{max}} dr p[r] = 1 \tag{5}$$

where h is a function defined on $[0, r_{max}]$. (Calculus of variations)

Continued

Then,

$$-\int_0^{r_{max}} dr \lim_{\epsilon \to 0} \frac{(p[r] + \epsilon h[r]) \log_2(p[r] + \epsilon h[r]) - p[r] \log_2 p[r]}{\epsilon} \tag{6}$$

$$+\lambda \int_0^{r_{max}} dr \lim_{\epsilon \to 0} \frac{p[r] + \epsilon h[r] - p[r]}{\epsilon} = 0 \tag{7}$$

$$-\int_0^{r_{max}} drh[r](\log_2 p[r] + \frac{1}{\ln 2} + \lambda) = 0 \tag{8}$$

Since this should hold for any function $h[r]$, the maximizer $p[r] = constant$.

Since the uniform pdf also satisfies (0.5),

$$p[r] = \frac{1}{r_{max}} \tag{9}$$

Entropy Maximization

Suppose the stimulus s invokes response $r = f(s)$. For small Δs, we have

$$p[s]\Delta s = \frac{|f(s + \Delta s) - f(s)|}{r_{max}} \tag{10}$$

The solution to (0.9) for a monotonically increasing response is

$$f(s) = r_{max} \int_{s_{min}}^{s} ds' p[s'] \tag{11}$$

This is the maximum-entropy encoding for the constraint $r \leq r_{max}$

Other Possible Constraints

If the average firing rate is constrained to a fixed value, the maximizing probability density function is exponential. If the variance is also constrained, the maximizing pdf becomes a Gaussian.

We want to maximize

$$H = -\int_0^\infty drp[r] \log_2 p[r] \tag{12}$$

subject to

$$\int_0^\infty dr\ rp[r] = r_{avg} \tag{13}$$

$$\int_0^\infty drp[r] = 1 \tag{14}$$

Continued

Using similar methods,

$$\mathcal{L}(p, \lambda_1, \lambda_2) = -\int_0^\infty drp[r] \log_2 p[r] \tag{15}$$

$$+ \lambda_1(\int_0^\infty dr\ rp[r] - r_{avg}) \tag{16}$$

$$+ \lambda_2(\int_0^\infty drp[r] - 1) \tag{17}$$

Then, for any function h,

$$-\int_0^\infty drh[r](\log_2 p[r] + \frac{1}{\ln 2} + \lambda_1 r + \lambda_2) = 0 \tag{18}$$

Hence, $p[r] = e^{Ar+B}$ for some constants A, B that satisfy the other two constraints.

Variance Constraint gives a Gaussian

If a third constraint is imposed,

$$\int_0^{\infty} dr \, (r - r_{avg})^2 p[r] = r_{var} \tag{19}$$

we get a Gaussian.

Quiz

Problem 1. We want to maximize

$$H = -\int_0^\infty drp[r] \log_2 p[r]$$

subject to

$$\int_0^\infty drp[r] = 1$$

Show that if the average firing rate is constrained to a fixed value r_{avg}, the maximizing pdf is exponential.

Problem 2. In **Problem 1,** show that if the variance is also fixed to r_{var}, the maximizing probability density function becomes a Gaussian. Find the probability density function $p[r]$.

Application to Retinal Ganglion Cell Receptive Fields

We estimate the firing rate of a neuron in response to a particular image by,

$$r_{est} = r_0 + \int_0^{\infty} d\tau D(\tau) s(t-\tau) = r_0 + L(t) \tag{20}$$

The integral term is the linear estimate L, which represents a weighted sum of the stimulus accros time and space.

$$L(t) = \int_0^{\infty} d\tau \int d\overrightarrow{x} D(\overrightarrow{x}, \tau) s(\overrightarrow{x}, t-\tau) \tag{21}$$

Assuming the space-time receptive field $D(\overrightarrow{x}, t)$ is separable, we can rewrite $L(t) = L_s L_t(t)$ where

$$L_s = \int d\overrightarrow{x} D_s(\overrightarrow{x}) s_s(\overrightarrow{x}) \tag{22}$$

$$L_t(t) = \int_0^{\infty} d\tau D_t(\tau) s_t(t-\tau) \tag{23}$$

Spatial Receptive Fields of Retinal Ganglion Cells

We consider an array of retinal ganglion cells. We assume that the statistics of the **input** are translation-invariant. This means that input from a ball on my left is equivalent to when ball is moved to my right. Hence, the kernel D_s is only affected by $\overrightarrow{a}$. $\overrightarrow{a}$ denotes the center point of the receptive field of a retinal ganglion cell.

$$L_s(\overrightarrow{a}) = \int d\overrightarrow{x} D_s(\overrightarrow{x} - \overrightarrow{a}) s_s(\overrightarrow{x}) \tag{24}$$

When we consider many neurons, the array of vectors $\overrightarrow{a_i}$ may fill the receptive field quite densely. In this case, we can treat $\overrightarrow{a}$ as a continuous variable and integrate $L(\overrightarrow{a})$ over $\overrightarrow{a}$.

Making the Problem Easier (Solvable)

$$\boldsymbol{r} = (r_1, r_2, \ldots, r_N)$$

When maximizing entropy, exact factorization and probability equalization are difficult to achieve. Instead we require a weaker condition:

$$< r_i >=< r > \text{ and } < (r_i - < r >)^2 >= \sigma_r^2 \tag{25}$$

(25) is a necessary condition for individual response pdfs being equal. Furthermore, assume (26) for independence.

$$Q_{ij} = \int d\boldsymbol{r} p[\boldsymbol{r}](r_i - < r >)(r_j - < r >) = \sigma_r^2 \delta_{ij} \tag{26}$$

Finding a distribution that satisfies (25), (26) is usually tactible. We apply this method to find the whitening filter.

The Whitening Filter

The continuous analog of (26) gives

$$Q_{LL}(\overrightarrow{a}, \overrightarrow{b}) = \sigma_L^2 \delta(\overrightarrow{a} - \overrightarrow{b}) \tag{27}$$

Where the correlation function is defined as

$$Q_{LL}(\overrightarrow{a}, \overrightarrow{b}) :=< L_s(\overrightarrow{a}), L_s(\overrightarrow{b}) >= \int d\overrightarrow{x} d\overrightarrow{y} D_s(\overrightarrow{x} - \overrightarrow{a}) D_s(\overrightarrow{y} - \overrightarrow{a}) < s(\overrightarrow{x}), s(\overrightarrow{y}) > \tag{28}$$

Note that in (28), the term $< s(\overrightarrow{x}), s(\overrightarrow{y}) >$ is actually a function of $\overrightarrow{x} - \overrightarrow{y}$ (Why?) We write it as

$$Q_{ss}(\overrightarrow{x} - \overrightarrow{y}) =< s(\overrightarrow{x}), s(\overrightarrow{y}) > \tag{29}$$

Finding the Optimal Filter D_s

We can solve (28) for D_s by expressing D_s, Q_{ss} in terms of their inverse Fourier transforms $\tilde{D}_s, \tilde{Q}_{ss}$. If the inverse Fourier transform of f is $\tilde{f}$, we can recover the original function by the applying the Fourier transform

$$f(\xi) = \int_{-\infty}^{\infty} \tilde{f}(x) \cdot e^{-2\pi i \xi x} dx \tag{30}$$

Since D_s, Q_{ss} are functions on $\mathbb{R}^2$, we apply (30) on both dimensions. From (27), we find

$$|\tilde{D}_s(\vec{\kappa})| = \frac{\sigma_L}{\sqrt{\tilde{Q}_{ss}(\vec{\kappa})}} \tag{31}$$

Therefore, the output of the optimal filter has a power spectrum (power as a function of frequency) $\tilde{Q}_{ss}(\vec{\kappa})|\tilde{D}_s(\vec{\kappa})|^2$ independent of the spatial frequency $\vec{\kappa}$. This is like white noise. That is why (31) is called a whitening filter. Also, the filter is not unique - it just needs to satisfy (31).

Power Spectrum

Energy spectral density describes how the energy of a signal is distributed with frequency. Here, energy E of a signal $x(t)$ is defined as

$$E := \int_{-\infty}^{\infty} |x(t)|^2 dt$$

From Parseval's Theorem,

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |\hat{x}(\omega)|^2 d\omega$$

where $\hat{x}(\omega)$ is the Fourier transform of $x(t)$. (In general, no physical power is actually involved) The power spectrum is given by

$$P = \lim_{T\to\infty} \frac{1}{T} \int_{-\infty}^{\infty} |x(t)\mathbb{k}_{[-T/2,T/2]}|^2 dt$$

Optical Modulation Transfer Function

Measurements reflect that $\tilde{Q}_{ss}(\vec{\kappa}) \propto 1/|\vec{\kappa}|^2$. Also, an additional factor $\exp(-\alpha|\vec{\kappa}|)$ must be included to account for the filtering introduced by the optics of the eye (the optical modulation transfer function). As a result,

$$\tilde{Q}_{ss}(\vec{\kappa}) \propto \frac{\exp(-\alpha|\vec{\kappa}|)}{|\vec{\kappa}|^2 + \kappa_0^2}$$

Then, substituting in (31), we obtain the result that $|\tilde{D}_s(\vec{\kappa})|$ is predicted to grow exponentially for large $|\vec{\kappa}|$. What does this imply? Whitening filters maximize entropy by equalizing the distribution of response power over the entire spatial frequency range. High spatial frequency components of images are relatively rare in natural scenes and, even if they occur, are greatly attenuated by the eye. The whitening filter compensates for this by boosting the responses to high spatial frequencies.

The Problem with this Strategy

Real inputs to retinal ganglion cells involve a mixture of **true signal and noise** coming from biophysical sources in the retina. At high spatial frequencies, for which the true signal is weak, inputs to retinal ganglion cells are likely to be **dominated by noise**, especially in low-light conditions. Boosting the amplitude of this noise-dominated input and transmitting it to the brain is **not an efficient visual encoding strategy**.

Why did this problem occur? And how do we resolve it?

Because no distinction has been made between the entropy coming from true signals and that coming from noise. To correct this problem, we should maximize the information transmitted by the retinal ganglion cells about natural scenes, rather than maximize the entropy. We will follow an approximate procedure that **prefilters the input** to eliminate as much noise as possible, and then uses the results of this section to **maximize the entropy of a linear filter** acting on the prefiltered input signal.

Filtering Input Noise

Suppose the visual stimulus is the sum of true stimulus and a noise term (a distrotion). We express the Fourier transform of the linear kernel $\tilde{D}$ as a product of a noise-eliminating filter, $\tilde{D}_\eta$, and the whitening filter from (31), $\tilde{D}_w$.

$$\tilde{D}_s(\vec{\kappa}) = \tilde{D}_\eta(\vec{\kappa})\tilde{D}_w(\vec{\kappa}) \tag{32}$$

How do we determine the form of the noise filter? We choose the optimal kernel that makes the total input as close to the true signal as possible. The solution to this problem utilizes functional derivatives. The answer: the Fourier transform of the optimal filter is the Fourier transform of the cross-correlation between the quantity being filtered and the quantity being approximated divided by the Fourier transform of the autocorrelation of the quantity being filtered.

We assume that the signal and noise terms are uncorrelated, so $< s_s(\vec{x})\eta(\vec{y}) >= 0$. The relevant cross-correlation for this problem is

$$< (s_s(\vec{x}) + \eta(\vec{x})) \cdot s_s(\vec{y}) >= Q_{ss}(\vec{x} - \vec{y}) \tag{33}$$

and the autocorrelation is

$$< (s_s(\vec{x}) + \eta(\vec{x}))(s_s(\vec{y}) + \eta(\vec{y})) >= Q_{ss}(\vec{x} - \vec{y}) + Q_{\eta\eta}(\vec{x} - \vec{y}) \tag{34}$$

Here, $Q_{ss}, Q_{\eta\eta}$ are the stimulus and noise autocorrelation functions, respectively. These results imply that the optimal noise filter is real and given by

$$\tilde{D}_\eta(\vec{\kappa}) = \frac{\tilde{Q}_{ss}(\vec{\kappa})}{\tilde{Q}_{ss}(\vec{\kappa}) + \tilde{Q}_{\eta\eta}(\vec{\kappa})} \tag{35}$$

The Optimal Kernel

By design, we may approximate $\tilde{D}_w$ using (31) to get

$$|\tilde{D}_s(\vec{\kappa})| \propto \frac{\sigma_L \sqrt{\tilde{Q}_{ss}(\vec{\kappa})}}{\tilde{Q}_{ss}(\vec{\kappa}) + \tilde{Q}_{\eta\eta}(\vec{\kappa})} \tag{36}$$

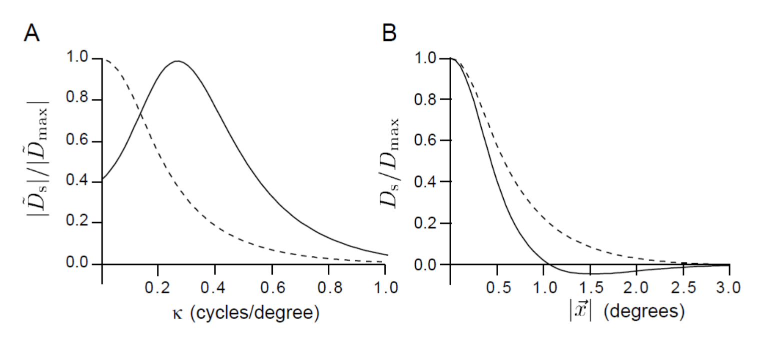
(Computation) We want to minimize an expression of the form,

$$Error = \frac{1}{T} \int_0^T dt (r_0 + \int_0^\infty d\tau D(\tau) s(t-\tau) - r(t))^2$$

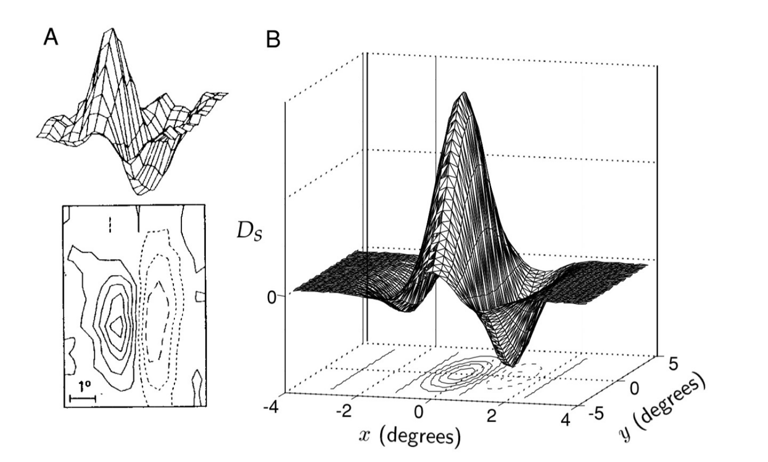
Although I have not tried the calculations yet, this is another potential quiz problem.

Receptive Field Properties Predicted by our Caclulations

Calculations, we mean by entropy maximization and noise suppression of responses to natural images.



The resulting function $D_s(\vec{x})$ is radially symmetric, so it only depends on $|\vec{x}|$. Under low noise conditions (solid lines in figure 4.3), the linear kernel has a bandpass character and the predicted receptive field has a center-surround structure, which matches the retinal ganglion receptive fields shown in chapter 2.


A
1°
B
D_S
0
-4
-2
0
2
4
x (degrees)
5
0
-5
y (degrees)

Continued

This structure eliminates one major source of redundancy in natural scenes: the strong **similarity of neighboring inputs** owing to the predominance of low spatial frequencies in images.

When the noise level is high (dashed lines in figure 4.3), the structure of the optimal receptive field is different. In spatial frequency terms, the filter is now **low-pass, and the receptive field loses its surround**. This structure averages over neighboring pixels to extract the true signal obscured by the uncorrelated noise. In the retina, we expect the signal-to-noise ratio to be controlled by the level of ambient light, with low levels of illumination corresponding to the high-noise case. The predicted change in the receptive fields at **low illumination (high noise)** matches what actually happens in the retina. At low light levels, circuitry changes within the retina **remove the opposing surrounds** from retinal ganglion cell receptive fields.

Temporal Processing

Natural images tend to change relatively slowly over time. This means that there is substantial redundancy in the succession of natural images, suggesting an opportunity for efficient temporal filtering to complement efficient spatial filtering.

Finding the optimal filter takes us through analogous, though lower dimensional, steps:

$$< L_t(t)L_t(t') >= \sigma_L^2 \delta(t - t') \tag{37}$$

(37) is the equation for decorrelation and variance equalization. We take the same steps by expressing the filter $D_t(\tau)$ by the Fourier transform of $\tilde{D}_t(\omega)$. Then, analogous to (36), we get

$$|\tilde{D}_t(\omega)| \propto \frac{\sigma_L \sqrt{\tilde{Q}_{ss}(\omega)}}{\tilde{Q}_{ss}(\omega) + \tilde{Q}_{\eta\eta}(\omega)} \tag{38}$$

Optimal Coding

Dong and Atick (1995) determined that the temporal power spectrum has the form,

$$\tilde{Q}_{ss}(\omega) \propto \frac{1}{\omega^2 + \omega_0^2} \tag{39}$$

To determine the temporal kernel, we require it to be causal ($D_t(\tau) = 0$ for $\tau < 0$) and impose a technical condition known as minimum phase, which assures that the output changes as rapidly as possible when the stimulus varies.