Statistical Information Theory

Problem sheet 1

Problems for Lecture 2: Entropy and typicality

1. **Discrete distributions**. Calculate the entropy of the following discrete distributions:

$$\boldsymbol{p}_1 = \{1, 0, 0, 0, 0, 0, 0\}$$

$$\boldsymbol{p}_2 = \{\frac{1}{2}, 0, 0, 0, 0, 0, \frac{1}{2}\}$$

$$\boldsymbol{p}_3 = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\}$$

$$\boldsymbol{p}_4 = \{\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\}$$

2. **The blurry microscope**. A cell can be in one of K discrete positions along a 1D space, with probabilities $\{p_1, p_2, ..., p_K\}$. We want to find the position of the cell by looking through a microscope. However, the microscope is blurry around the edge, so it cannot distinguish the cases when the cell is in position 1 or in position 2. In other words, whenever the cell is in positions 1 or 2, our microscope yields the same measurement of the cell being in the edge (denoted as state E).

   (a) What is the probability distribution of our microscope measurements? Show this is a valid probability distribution.

   (b) Prove that the entropy of our microscope measurements is lower than the underlying entropy of the cell's position. What does this mean?

3. **The poisoned wine**. You own a restaurant in Mayfair. The Prime Minister and a cohort of foreign diplomats are coming for dinner. You have received a new shipment of 1,000 bottles of wine from a new supplier. A whistleblower has told you one of the bottles is poisoned. You possess testing kits that allow you to test for poison in a wine glass, but you only have 10 kits.

   Question: Can you design a sequence of tests that will find the poisoned bottle with high probability? (Whether you throw it away or give it to the Prime Minister is up to you.)

   Hint: You can mix wine from multiple bottles in a single glass.

4. **Entropy bounds**. Let X be a discrete random variable with a finite alphabet $\mathcal{X}$. Let's prove some important bounds for entropy:

$$0 \leq H(X) \leq \log |\mathcal{X}|$$

You can use the fact that both $\log x$ and $H(X)$ are a concave-down ($\bigcap$) functions of x and $p(x)$ respectively.

(a) Prove that $0 \leq H(X)$. Show the equality is achieved by the Kronecker delta distribution, i.e. $p(x_i) = \delta_{ij}$ for a given j.

(b) Prove that $H(X) \leq \log |\mathcal{X}|$. (*Hint*: Use Jensen's inequality.) Show the equality is achieved by the uniform distribution, i.e. $p(x_i) = 1/|\mathcal{X}|$.

5. **Exponential distribution**. Calculate the differential entropy of an exponential distribution for any parameter $\lambda \in \mathbb{R}^+$:

$$p(x) = \lambda \exp(-\lambda x) \tag{1}$$

6. **Change of units**. Differential entropy is indeed a rather quirky measure. One of its quirks is that it depends on the units one uses to measure x.

   Consider an experiment where the voltage V across a neuron's membrane is measured. Measurements reveal a normal distribution of voltages with mean $\mu = -65\,\mathrm{mV}$ and standard deviation $\sigma = 5\,\mathrm{mV}$.

   (a) Calculate the entropy of the voltage measurement V.

   (b) We change the units of our voltmeter and measure the neuron's voltate in volts. What is the entropy of the resulting quantity?

   (c) We change the reference value of our voltmeter so that the new measurement is $V' = V + 65\,\mathrm{mV}$. Without using the formula for the Gaussian entropy, argue that a shift in the random variable does not cause a change in entropy. (You may use the formula of the differential entropy for any continuous PDF $p(x)$.)

7. **In high-dimensional space, no one can hear anyone scream**. Consider a sample of M points from a zero-mean, spherical n-dimensional Gaussian. Prove that as n grows all samples are equally far from each other (in terms of squared norm) and that this distance grows with n.

   Hint: The difference between two Gaussian RVs, $\boldsymbol{D} = \boldsymbol{X} - \boldsymbol{Y}$, where $\boldsymbol{X}, \boldsymbol{Y} \sim \mathcal{N}(\mathbf{0}, \sigma \boldsymbol{I}_n)$, is also Gaussian. Begin by deriving the mean and variance of $\boldsymbol{D}$.

Programming exercises

The following exercises are designed to give you an intuition for high-dimensional spaces, independent random variables, and typicality. Take some time to experiment with different values and distributions, and visualise the results in different forms.

1. **High-dimensional Gaussians**.

(a) Take k random samples from an n-dimensional spherical Gaussian with unit variance. Calculate the (L2) distance between each sample and the origin, and plot these distances in a histogram. How does this histogram change as n increases? (The values $k = 1,000$ and $1 \leq n \leq 500$ are a good start.)

(b) Evaluate the likelihood (i.e. probability density), under the same Gaussian distribution, of each sample and of the mode of the distribution. How do these two compare? What intuitions can we get from this?

2. **AEP and typical sets**. Let's try to get some intuition on the AEP and typical sets with some numerical examples.

   i. Pick any discrete PDF you like that allows you to draw samples, evaluate likelihoods, and calculate entropies. (*Hint*: just pick any distribution from the `scipy.stats` library.)

   ii. Sample k strings of length n and calculate their likelihood p_i of each sample $i = 1, ..., k$ (you can use the `pmf()` method of the `scipy.stats` distribution; try $k = 5,000$ and $10 \leq n \leq 1,000$). Calculate the values $-n^{-1} \log p_i$ and plot them in a histogram.

   iii. Calculate the entropy H of the distribution (you can use the `entropy()` method). Pick a positive number ε (for example, $\varepsilon = 0.1$), and add X-ticks at $H - \varepsilon$, H, and $H + \varepsilon$.

   What do you observe as n grows? What is the relationship between the average log-likelihood of the strings and the entropy of the original distribution?