Intelligence and Field Theory
Summer term 2025

PD Dr. Torsten Enßlin
Vishal Johnson, Andreas Popp

Exercise sheet 5

Exercise 5 - 1

It is a lazy Friday morning and you are slumped on your couch watching the “Bayes Chronicles” with a hot cup of Kolmogorov brew. Suddenly, you are startled by a loud knock on your front door. You open the door to find a mysterious package — the wrapper is as black as it is shiny and has an ethereal feel, it seems unreal as if it isn’t from your world. Looking around you find no one there. A bit alarmed, you take the package into your home. Unwrapping the package reveals a box containing N buttons; it contains no other details except the following instruction at its bottom:

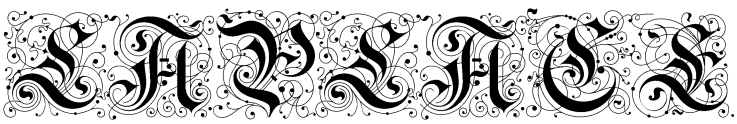
“Press the correct button and the Djinn of the box would grant you a wish. Press the wrong button and the box shall disappear into a puff of smoke and you would suffer a month of bad luck!”

Perplexed, you think of contacting professor Alice. Professor Alice is a peculiar character, once the most famous mathematician in all the world she is now ousted from the mathematical society because of her finally believing in the existence of arbitrary path integrals. Nonetheless, for a situation as obscure as the one that confronts you you cannot think of a single better person to contact. However, you must address her in the peculiar way of her choosing — she only communicates through written letter.

It has been 17 days now. Professor Alice had sent a follow up message two weeks ago that she found the box most baffling and that she would need two weeks to do some research on the box and get back to you. Meanwhile you have stored the box in a levitating field as instructed by professor Alice — you do not want to risk the box interacting too much with the outside world.

Finally, you hear the unmistakable tinkle of the post van. There is the singular letter you were expecting from professor Alice. You anxiously open up the letter and wait with bated breath for the message to be revealed; and there it is: a completely empty letter. What is the probability distribution you assign to the buttons given this message? Use both, the Laplace principle and MaxEnt principle to decode the empty message received! (2 points)

Epilogue: Alas, you click on the wrong button and find that the box disappears into a cloud of smoke. However, as you carefully observe, the smoke forms a vague pattern:



Exercise 5 - 2

In this exercise we calculate the achieved information gain for the Binomial distribution. We first consider a Bernoulli trial with two outcomes $s \in \{0, 1\}$. We want to calculate the achieved information gain with respect to Alice I_A of the update of Bob’s state from I_0 to I_B. Here are the information contents: for $X \in \{A, B, 0\}$ referring to Alice, Bob final, and Bob initial respectively, $p_X = P(s = 0|I_X)$.

a) What is achieved information gain for a single Bernoulli trial? (1 point)

b) Now, let us consider the Binomial distribution. We are looking at the random variable $s \in \{0, 1, \dots, n-1\}$ for n Bernoulli trials referring to the number of times outcome 0 turned out to be the case for the n Bernoulli trials. Guess the achieved information gain without actually calculaing. (1 point)

c) Calculate the actual information gain. (3 points)

Exercise 5 - 3

In this exercise we trace Shannon's argument for the entropy function. Consider a random variable that can take on the discrete set of values $(x_1, \ldots, x_N)$ with probabilities $(p_1, \ldots, p_N)$. We request an entropy function $H(p_1, \ldots, p_N)$ with some desirable properties.

First, we would like H to have the composition property. Let us partition the set of events $\{x_i\}$ into sets $\{y_j\}$ $(j \in \{1, \ldots, M\})$ with $y_j = \{x_{j_1} \ldots, x_{j_{N_j}}\}$ — of course $\sum_j N_j = N$. This means distributing the events $\{x_i\}$ into non-intersecting subsets y_j whose union is the entire set $\{x_i\}$ — $y_j \cap y'_j = \phi$ for $j \neq j'$ and $\cup_j y_j = \{x_i\}$. Note also that the probabilities of events y_j is $q_j = \sum_k p_{j_k}$.

a) The composition property states that the entropy associated with choosing the partitions and then the individual events within the partition equals the entropy of choosing the events themselves.. Fill in the blanks for the below equation and explain your reasoning:

$$H(p_1, \ldots, p_N) = \underline{\hspace{3cm}} + \sum_j q_j \underline{\hspace{3.5cm}} . \tag{1}$$

(3 points)

b) We now consider the functions $A(n) = H(1/n, \ldots, 1/n)$ (with n arguments for this particular H).

The second assumption we use for H is that the A function constructed from it is a monotonically increasing function of n. This just means that a larger entropy is associated with systems with larger number of elements. Consider dividing the events N into M sets of n equiprobable events so that $N = Mn$. What does eq. (1) now become? What is a function that solves it? (2 points)

c) The A functions are only defined above for integer values. Requesting an anlytic extension to real numbers fixes the A function and therefore also H. This would be the third assumption about H — that it is a continuous function of all its variables. Work out the expression for H using this assumption by first restricting to rational probabilities $p_i = n_i/N$ where $N = \sum_{i=1}^{M} n_i$ and then analytically extending. (3 points)

Exercise 5 - 4

Your knowledge I about a quantity $\mathbf{x} = (x_1, \ldots x_n)^{\mathrm{T}} \in \mathbb{R}^n$ is characterized by being separable, i.e. $\mathcal{P}(\mathbf{x}|I) = \prod_{i=1}^{n} \mathcal{P}(x_i|I)$ and by being isotropic, i.e. $\mathcal{P}(\mathbf{x}|I) = \mathcal{P}(\mathbf{O}\,\mathbf{x}|I)$ for all orthonormal transformations with $\mathbf{O}^\dagger = \mathbf{O}^{-1}$.

Using the definition of the information Hamiltonian:

$$\mathcal{H}(\mathbf{x}|I) = \sum_{i=1}^{n} \mathcal{H}(x_i|I) \equiv \sum_{i=1}^{n} h_i(x_i) \ , \tag{2}$$

we may write

$$\begin{aligned}
\prod_{i=1}^{n} \mathcal{P}(x_i|I) &= \prod_{i=1}^{n} \mathcal{P}((\mathbf{O}\,\mathbf{x})_i|I) && (3)\\
\sum_{i=1}^{n} \mathcal{H}(x_i|I) &= \sum_{i=1}^{n} \mathcal{H}((\mathbf{O}\,\mathbf{x})_i|I) && (4)\\
\sum_{i=1}^{n} h_i(x_i) &= \sum_{i=1}^{n} h_i((\mathbf{O}\,\mathbf{x})_i) \ . && (5)
\end{aligned}$$

a) Let $\mathbf{x} = r\, e^{(i)}$ be parallel to the i-th unit vector of $\mathbb{R}^n$, $e^{(i)}_j = \delta_{ij}$, and $\mathbf{O}$ such that $\mathbf{O}e^{(i)} = e^{(j)}$ with $i \neq j$. Show that $h_i(r) = h_j(r) + \mathrm{const(i,j)}$. (3 points)

b) Use the result of a) to show that $\mathcal{H}(\mathbf{x}|I) = \sum_i h(x_i) + const.$ with $h(x_i) \equiv h_0(x_i)$. (1 point)

c) Given a general $\mathbf{x} \in \mathbb{R}^n$ with length $r = |\mathbf{x}| = \sqrt{\sum_i x_i^2}$, we can choose $\mathbf{O}$ such that $\mathbf{Ox} = (r, 0, \ldots, 0)$. Show that in this case

$$\frac{h'(x_j)}{x_j} = \frac{h'(r)}{r} \quad \forall j \ , \tag{6}$$

where h' denotes the derivative of h. (3 points)

d) Given the result of c), derive the general functional form of h. (2 points)

e) Finally, derive an expression for $\mathcal{P}(\mathbf{x}|I)$. (2 points)

This exercise sheet will be discussed during the exercises.
Group 01, Wednesday 18:00 - 20:00, Theresienstr. 37, A 449,
Group 02, Thursday, 10:00 - 12:00, Theresienstr. 37, A 249,

https://wwwmpa.mpa-garching.mpg.de/ ensslin/lectures/lectures.html