Exercise sheet 4

Exercise 4 - 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:

LAPLACE.

Exercise 4 - 2

Imagine you wanted to store a probability distribution $P(x)$ over two events $x \in \{0, 1\}$ on a computer with very limited memory and precision. Of course it is enough to store only $P(0)$ since $P(1)$ can then be calculated through normalization, but you still have to round the numbers up to machine precision. Let X be the set of all numbers the computer can represent. Furthermore, let

$$q_{\text{low}} = \max\{q | q \in X \wedge q \leq P(0)\}$$
$$\text{and } q_{\text{high}} = \min\{q | q \in X \wedge q \geq P(0)\}$$

i.e., q_{low} the highest number in X that is still lower than $P(0)$ and q_{high} the lowest number in X that is still higher than $P(0)$.

a) Derive a decision rule for when to round to q_{low} or q_{high} for general $P(0)$, q_{low} and q_{high} based on the rule that you want to lose the least amount of information from the original distribution $P(x)$. (2 points)

b) Using the decision rule derived in a), determine wether it is better to round

- $P(1) = 0.146$ to 0.1 or 0.2
- $P(1) = 0.01$ to 0 or 0.5?

(1 point)

Exercise 4 - 3

Assume that for a PDF $P(x)$ only a set of constraints —typically moment constraints— are known, like

$$\begin{aligned} Q(x) = P(x \mid I) &= \text{const}, \\ P(x) &> 0, \\ \int P(x)\,\mathrm{d}x &= 1, \\ \int P(x) f_i(x)\,\mathrm{d}x &= \alpha_i. \end{aligned}$$

From the principle of maximum entropy one derives that

$$P(x) = \exp\left(\lambda_0 + \sum_{i=1}^{n} \lambda_i f_i(x)\right)$$

with λ_i such that the constraints are satisfied.

Assume a PDF $P(x|\alpha, \mu)$ with $x \in \mathbb{R}$, $\alpha > 0$, $\mu \in \mathbb{R}$ and $\langle |x - \mu| \rangle_{(x|\alpha,\mu)} = \alpha$. Calculate the PDF of maximum entropy. (3 points)

Exercise 4 - 4

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 $s \in \{0, 1, \ldots, n-1\}$ 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)

This exercise sheet will be discussed during the exercises.
Group 01, Wednesday 18:00 - 20:00, Theresienstr. 37, A 449,
https://wwwmpa.mpa-garching.mpg.de/ ensslin/lectures/lectures.html