Intelligence and Field Theory
Summer term 2025

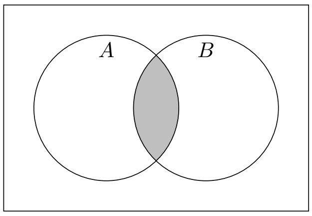
PD Dr. Torsten Enßlin
Vishal Johnson, Andreas Popp

Exercise sheet 3

Exercise 3 - 1

As in the lecture, we consider a set $\mathcal{S}$ of states of a physical system S (for example, the microstates of a gas in a box). We ask questions about this physical system that correspond to subsets of the set $\mathcal{S}$ (for example, the microstates of a gas in a box corresponding to a certain total energy E).

There is, thus, a correspondence between subsets of $\mathcal{S}$ and statements one can make about the system S. For example, the negation of a statement corresponds to the complement of a subset. Consider the statements "A. Property a holds for the system S." and "B. Property b holds for the system S." and the subsets A and B of all microstates for which these statements are respectively true. Then, "A and B. Properties a and b hold for the system S." corresponds to the intersection $A \cap B$:



.

a) Consider two statements A and B. What are the set theory equivalents of the natural language statements:

1. not A,
2. A or B,
3. not both A and B,
4. neither A nor B.

Draw their Venn diagrams. (3 points)

b) What are the set theoretic equivalents of the following:

1. either A or B,
2. A if B,
3. A if and only if B,
4. A but not B.

Draw their Venn diagrams. (3 points)

c) What is the Venn digram for the statement A implies B. Does this Venn diagram match the intuition $A \subseteq B$? If not, point out the inconsistency. (3 points)

Exercise 3 - 2

Attention!
Consider a scenario in which Alice wants to communicate some information to Bob. The information concerns a continuous quantity s and resulting outcomes. Alice's knowledge state is $\mathrm{I_A}$ which gives rise to the probability distribution $p(s) = \mathcal{P}(s|\mathrm{I_A})$. Similarly, Bob's knowledge state is $\mathrm{I_B}$ which gives rise to the probability distribution $q(s) = \mathcal{P}(s|\mathrm{I_B})$. The aim is to find an appropriate loss function to quantify "distances" between the information contents $\mathrm{I_A}$ and $\mathrm{I_B}$ (and therefore probability distributions $p(s)$ and $q(s)$).

We wish the loss function to satisfy certain properties:

1. We assume that the loss function obeys *locality* - that the loss corresponding to outcome s_* must only depend on the probability $q(s_*)$ induced in Bob about event s_*. Therefore, the loss function takes the form $\mathcal{L}(s, q(s))$.
2. Furthermore, we assume that the loss function obeys *properness* - that the loss function takes on its minimal value when $q(s) = p(s)$.

Alice calculates the average loss L by averaging the above loss function over her probability distribution:

$$\mathrm{L} = \langle \mathcal{L}(s, q(s)) \rangle_{p(s)}. \tag{1}$$

a) Calculate the sensitivity of the loss L to the probability $q(s_*)$ that Bob's knowledge state assigns to signal value s_*. In order to do this, calculate the functional derivate $\delta \mathrm{L}/\delta q(s_*)$. (2 points)
Hint: Try to solve the problem for a discrete signal $\{s_i\}$ first.

b) For a proper loss function $\mathcal{L}$, the functional derivate evaluated at $q = p$ must be minimal. Therefore $\delta \mathrm{L}/\delta q(s_*)|_{q=p} = 0$. Calculate the functional relation satisfied by $\mathcal{L}$ under this condition. Remember that $q(s)$ must be normalised. (2 points)

c) Find out the loss function which satisfies the above functional relation. (1 point)

d) There are scenarios in which certain signal values s deserve more attention than other values of the signal. Using a weighting scheme $w(s)$ (with $w(s) > 0$) to quantify the attention deserved by different signal values one can define an attention function

$$\mathcal{A}(s) = \frac{w(s)\mathcal{P}(s)}{\int w(s)\mathcal{P}(s) ds} \tag{2}$$

corresponding to a probability distribution $\mathcal{P}(s)$.

Consider $\mathcal{A}_q(s)$ and $\mathcal{A}_p(s)$, the attention functions corresponding to the probability distributions $q(s)$ and $p(s)$. Alice wants to ensure that her communication influences Bob to place maximal attention $\mathcal{A}_q(s_*)$ on what actually happens (s_*) in the end.

First, show that properness in probability and properness in attention are equivalent. (2 points)

e) Repeat the above analysis for attention functions $\mathcal{A}_q(s)$ and $\mathcal{A}_p(s)$. Note that expectation averages are still to be performed over probabilities. (4 points)

For a more detailed discussion please check out the paper onlinelibrary.wiley.com/doi/10.1002/andp.202300334.

Exercise 3 - 3

A sequence of n coin tosses is performed and stored in a data vector $d^{(n)} = (d_1, \dots d_n) \in \{0, 1\}^n$. The coin produced a head (denoted by a 1 in the data vector) with constant, but unknown frequency $f = P(d_i = 1|f) \in [0, 1]$.

a) How many bits of extra information on f are provided by the data vector $d^{(n)} = (1, \dots 1)$ of only ones?

Hint: The extra information contained in a probability distribution $p(x)$ compared to a probability distribution $q(x)$ (in bits) is given by $\int \mathrm{d}x\, p(x) \log_2 \left(\frac{p(x)}{q(x)}\right)$ ($\mathcal{D}_\mathcal{S}(p(x), p(x), q(x))$). Furthermore you may use the following integral formulas:
$\int_0^1 dx\, x^n (1-x)^m = \frac{n!\, m!}{(n+m+1)!}$ for $n, m \in \mathbb{N}$, $\int_0^1 dx\, x^n \ln x = -\frac{1}{(n+1)^2}$ (2 points)

b) After how many such sequential heads did one obtain 10 bits of information on f? An accuracy of 10% is sufficient.

Hint: If $n > 10$ you can use $\frac{n}{n+1} \approx 1$. Use $2^{1/\ln 2} = 2^{\ln e/\ln 2} = 2^{\log_2 e} = e \approx 2.7$. (1 point)

c) How many bits on the outcome of the next toss is provided by a sequence of n heads? Provide also the asymptotic for $n \to \infty$!

Hint: It is helpful to guess the maximal amount of obtainable information before the detailed calculation is done. (2 points)