**Intelligence and Field Theory**
Summer term 2025

*PD Dr. Torsten Enßlin*
*Vishal Johnson, Andreas Popp*

# Exercise sheet 4

## Exercise 4 - 1

Consider the following coin toss experiment:

- A number $n_A$ respectively $n_B$ coin tosses are performed with coin $A$ and coin $B$. The results are stored in data vectors $d_A^{(n)} = (d_1, \dots, d_n) \in \{0,1\}^n$ and $d_B^{(n)}$, where 1 and 0 represent the possible outcomes head and tail, respectively.
- Individual tosses are independent from each other.
- All tosses are done with either coin $A$ or coin $B$ with each having an unknown bias $f_A, f_B \in [0,1]$ denoting the probability of head $P(\text{head}|f_i) = f_i \ \forall i \in \{A, B\}$.

Coin $A$ and $B$ are used in a game in which you are paid $G_A = 1$ € if you get a 1 with coin $A$ and $G_B = 4$ € if you get a 1 with coin $B$. If the coin shows 0, you are not paid anything. You are now given the following data $d_A^{(2)} = (1,1)$, $d_B^{(2)} = (0,0)$.

Hint: For your convenience, we note that for $a, b \in \mathbb{N}$ it holds that $\int_0^1 dx\, x^a (1-x)^b = \frac{a!b!}{(a+b+1)!}$.

**a)** Derive the expected financial gain for each coin and decide which coin to toss next. (5 points)

**b)** When is the optimal point in time to switch to the other coin assuming your chosen coin fails in subsequent tosses. (3 points)

**c)** How much information did you gain on $f_i$ for your chosen coin $i \in \{A, B\}$ by these additional (loosing) tosses in part b)? Please also state the units of the information gain in your answer. (3 points)

## 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

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

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

**a)** Derive a decision rule for when to round to $q_{\text{low}}$ or $q_{\text{high}}$ for general $P(0)$, $q_{\text{low}}$ and $q_{\text{high}}$ 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

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

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 $\lambda_i$ 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

A bacterial species lives in an environment that sustains exactly $N_\mathrm{i} = \mathcal{O}(10^6)$ individuals. Per generation, each bacterium first splits into two, the original and a copy, and then half of the population dies according to natural selection. Each genome $g$ of the bacteria is a sequence of $N_\mathrm{b} = \mathcal{O}(10^6)$ bases (letters), where each is drawn out of four possible molecules (the genetic alphabet).

**a)** What is the probability that a randomly synthesized genome of length $N_\mathrm{b}$ is identical to that of a specific bacterium? Provide also a rough decimal number for the case $N_\mathrm{b} = 10^6$! You can use $\log_{10} 4 \approx 0.6$. (1 point)

**b)** How many bits of information is stored in a genome? (1 point)

**c)** We identify the distribution function

$$f(g) \equiv \frac{N(g)}{N_\mathrm{i}} \equiv \frac{\#\text{ bacteria with genom } g}{\#\text{ bacteria}}$$

with the probability $P(g)$ that an individual bacterium has the genome $g$, $P(g) = f(g)$. What is the maximal and minimal amount of information the population can represent with respect to the flat distribution in the space of all sequences as measured in bits? (2 points)

**d)** **Deterministic evolution:** Assume that the replication is free of mutations and also that the natural selection is deterministic. Each possible genome $g$ has a different fitness value $\mathcal{F}(g) \in \mathbb{R}$. For simplicity, we assume that the natural selection eliminates in each generation all bacteria with fitness below the population median. How many bits of information can maximally be generated per generation (relative information between generations)? After how many generations is the fittest genotype present in the initial population guaranteed to be the sole survivor? Provide the number of generation for the case $N_i = 10^6$! You can use $\log_2 10 = 3.3$. (4 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*