1. Problem sheet for **Sequential Decision Making**

Exercise 1 Formally show that the regret defined by

$$\mathcal{R}(\nu, T) = T\mu^*(\nu) - \mathbb{E}\left[\sum_{t=1}^{T} R_t\right] \tag{1}$$

can also be written as

$$\mathcal{R}(\nu, T) = \sum_{a \in \mathcal{A}} \Delta_a \mathbb{E}[N_a(T)]$$

using the following definitions:

- Action gap: $\Delta_a(\nu) = \mu^* - \mu_a(\nu)$
- Number of times action a was chosen by the learner:

$$N_a(t) = \sum_{s=1}^{t} \mathbb{I}\{A_s = a\} \tag{2}$$

Exercise 2 Implement a multi-armed bandit in a Jupyter notebook with $K = 6$ Bernoulli arms (rewards 1 or 0) with $\mu_1 = 0.3$, $\mu_2 = 0.5$, $\mu_3 = 0.4$, $\mu_4 = 0.45$, $\mu_5 = 0.3$, and $\mu_6 = 0.35$.

1. Draw random samples from all arms for $T = 1000$ rounds and store them in a matrix.
2. Compute the empirical mean for each arm after 10 rounds, after 100 rounds, and after 1000 rounds.
3. Compare these empirical values to the true mean values, and compute the deviations from the mean for the different sample sizes.

Exercise 3 Use the multi-armed bandit from Exercise 2, this time using random means for the arms and restricting to two arms. Implement the following three algorithms:

1. Uniform Exploration
2. Follow The Leader
3. Explore-Then-Commit

in a Jupyter notebook. Compute the empirical regret averaged for each of the algorithms over 50 different runs (i.e., this means to resample all the arms 50 times with a different seed).

Exercise 4 Plot the expected regret calculated in the lecture for two Bernoulli arms for the Explore-Then-Commit algorithm for different action gaps and different m.

Exercise 5 Suppose that X is σ-subgaussian and X_1 and X_2 are independent and σ_1 and σ_2-subgaussian, respectively. Then:

- $\mathbb{E}[X] = 0$ and $\mathbb{V}[X] \leq \sigma^2$.
- cX is $|c|\sigma$-subgaussian for all $c \in \mathbb{R}$.