Throughout, $\sigma(z) = 1/(1+e^{-z})$ is the logistic function. We admit Hoeffding's inequality in the following form: if $Z_1, \dots, Z_m$ are independent random variables with values in an interval of length c, then for any $\varepsilon > 0$,

$$\mathbb{P}\Big(\frac{1}{m}\sum_{i=1}^{m}\big(Z_i - \mathbb{E}[Z_i]\big) \geq \varepsilon\Big) \leq \exp\Big(-\frac{2m\varepsilon^2}{c^2}\Big) \ . \tag{1}$$

Exercise 1 (Sequential halving)**.** A bandit has K arms with Bernoulli rewards. Sequential halving with budget n runs $L = \lceil \log_2 K \rceil$ phases. In each phase, each remaining arm is pulled $\lfloor n/(|S|L) \rfloor$ times, where $|S|$ is the number of remaining arms; the $\lceil |S|/2 \rceil$ arms with the largest empirical means over this phase are kept, ties going to the smallest index. The last remaining arm is recommended.

1. Let $K = 4$ and $n = 16$. Give the number of phases, the number of pulls of each arm in each phase, and the total number of pulls.
2. In phase 1, the rewards are $(1, 0)$ for arm 1, $(1, 1)$ for arm 2, $(0, 0)$ for arm 3 and $(1, 0)$ for arm 4. In phase 2, the rewards of the remaining arms are $(1, 1, 0, 1)$ for the one of smallest index and $(1, 0, 0, 1)$ for the other one. Give the arms kept after phase 1 and the recommended arm.
3. Let a and b be two arms with means μ_a and $\mu_b = \mu_a - \Delta$, where $\Delta > 0$, each pulled m times independently, with empirical means $\hat{\mu}_a$ and $\hat{\mu}_b$. Show that $\mathbb{P}(\hat{\mu}_b \geq \hat{\mu}_a) \leq \exp(-m\Delta^2/2)$.
   Hint: apply (1) to $Z_i = Y_i - X_i$, where X_i and Y_i are the i-th rewards of a and b.
4. For $\Delta = 0.1$, give the smallest m for which this bound is at most 0.05. One may use $\log 20 \approx 3.00$.
5. Let $K = 8$ and $n = 4800$. Give the number of pulls of each remaining arm in each phase of sequential halving. Assume that arms 1 and 2, with a gap $\Delta = 0.1$, both reach the last phase. Bound the probability that arm 2 beats arm 1 in this phase, and compare with uniform allocation, which pulls each of the 8 arms 600 times. One may use $e^{-4} \approx 0.018$ and $e^{-3} \approx 0.050$.

Exercise 2 (Bradley-Terry model and Elo ratings)**.** There are M models with scores $s \in \mathbb{R}^M$. In the Bradley-Terry model, model i beats model j with probability $\sigma(s_i - s_j)$, independently from one duel to the next. After a series of duels, W_{ij} is the number of wins of i against j, and $N_{ij} = W_{ij} + W_{ji}$. The log-likelihood of the scores is

$$\ell(s) = \sum_{i \neq j} W_{ij} \log \sigma(s_i - s_j) \ . \tag{2}$$

1. Show that $\sigma(-z) = 1 - \sigma(z)$ and $\sigma'(z) = \sigma(z)(1-\sigma(z))$ for any $z \in \mathbb{R}$, and that $\ell(s + c\mathbf{1}) = \ell(s)$ for any $c \in \mathbb{R}$, where $\mathbf{1}$ is the vector of ones.
2. Show that $\dfrac{\partial \ell}{\partial s_i}(s) = \sum_{j \neq i} \big(W_{ij} - N_{ij}\,\sigma(s_i - s_j)\big)$.
3. Let $M = 2$, $W_{12} = 6$ and $W_{21} = 2$. Show that the maximizers of ℓ are the scores with $s_1 - s_2 = \log 3$, and give the estimated probability that model 1 beats model 2.
4. The Elo update after a duel between i and j, with $y = 1$ if i wins and $y = 0$ otherwise, is

$$s_i \leftarrow s_i + k\big(y - \sigma(s_i - s_j)\big) \ , \qquad s_j \leftarrow s_j - k\big(y - \sigma(s_i - s_j)\big) \ ,$$

   with $k > 0$. Show that it is a gradient ascent step with step size k on the log-likelihood of this single duel, $y \log \sigma(s_i - s_j) + (1-y) \log \sigma(s_j - s_i)$.
5. Three players A, B and C start with ratings 0, and $k = 1$. A beats B, then B beats C, then A beats C. Compute the ratings after each duel. One may use $\sigma(-0.5) \approx 0.378$ and $\sigma(1.122) \approx 0.754$.