Thomas Kesselheim
Alexander Braun

June 17, 2020
Due: June 24, 2020 at noon

Algorithms and Uncertainty

Summer Term 2020

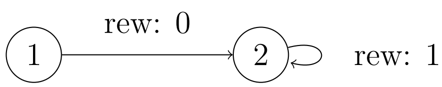
Exercise Set 7

Exercise 1: (3+2+2+2 Points)

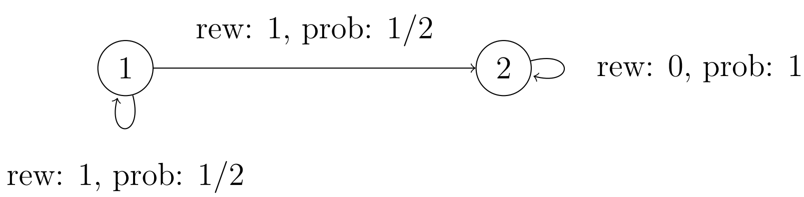
For the following single-armed bandits, give the fair charges of all states. Unless states otherwise, the transitions are deterministic. Justify your statements if necessary. For part (a), consider $\gamma = \frac{1}{2}$; for the remaining parts an arbitrary $\gamma \in (0, 1)$.

(a)

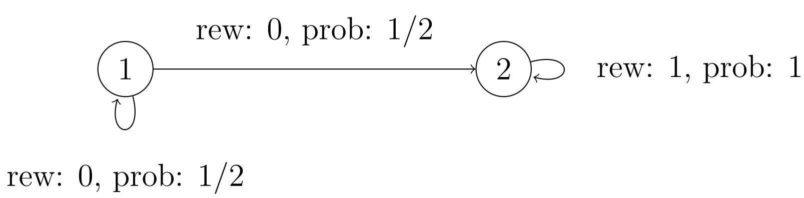
(b)



(c)



(d)



Exercise 2: (4 Points)

Consider the following explore-exploit algorithm. In the first $\frac{T}{2}$ steps (so $k = \frac{T}{2n}$), we explore. Afterwards, we exploit the most promising arm. Use the approach from the lecture to derive an upper-bound for the expected regret of this algorithm.

Exercise 3: (8 Points)

Use the one-sided version of Hoeffding's inequality to show a regret bound for UCB1 of $\sum_{i \neq i^*} \frac{4 \ln T}{\Delta_i} + 2\Delta_i$. The one-sided version of Hoeffding's inequality is as follows: Let $Z_1, \ldots, Z_N$ be independent random variables such that $a_i \leq Z_i \leq b_i$ with probability 1. Let $\bar{Z} = \frac{1}{N} \sum_{i=1}^{N} Z_i$ be their average. Then for all $\gamma \geq 0$

$$\Pr\left[\bar{Z} - \mathbf{E}[\bar{Z}] \geq \gamma\right] \leq \exp\left(-\frac{2N^2\gamma^2}{\sum_{i=1}^{N}(b_i - a_i)^2}\right) .$$

Hint: Note that the one-sided version of Hoeffding's inequality also implies a bound on $\Pr\left[\bar{Z} \leq \mathbf{E}[\bar{Z}] - \gamma\right]$.