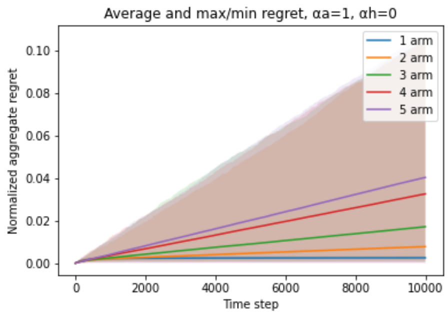
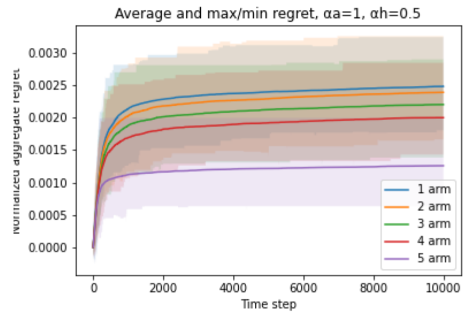
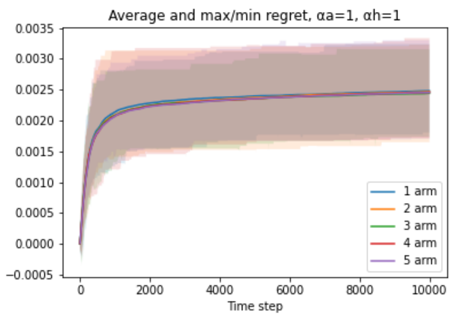


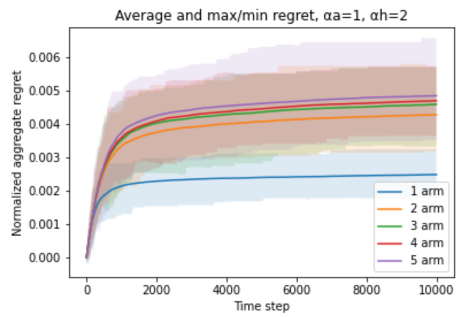
(a) $\alpha_h = 0$



(b) $\alpha_h = 0.5$



(c) $\alpha_h = 1$



(d) $\alpha_h = 2$

Figure 1: Plots show regret for (line shows average regret, shaded region is max and min regret, for 100 simulations, each with 10,000 time steps). Each simulation has 5 arms total, where the single best arm has reward drawn from $\mathcal{N}(\mu = 0.5, \sigma = 0.1)$ and the other arms have reward $\mathcal{N}(\mu = 0.45, \sigma = 0.1)$.

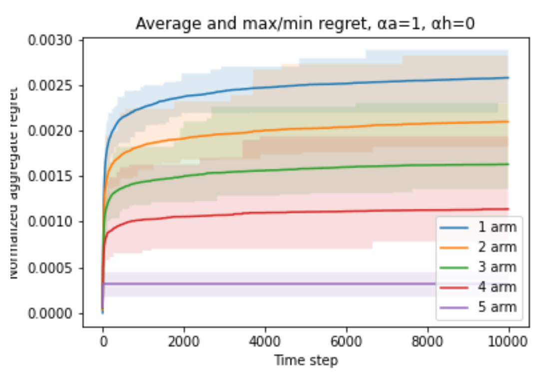
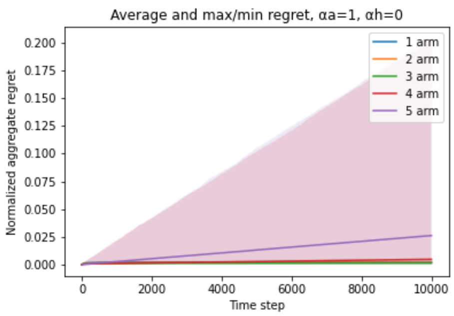
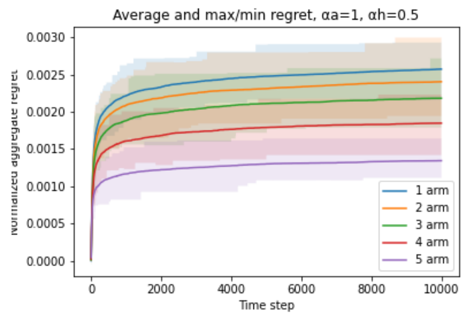


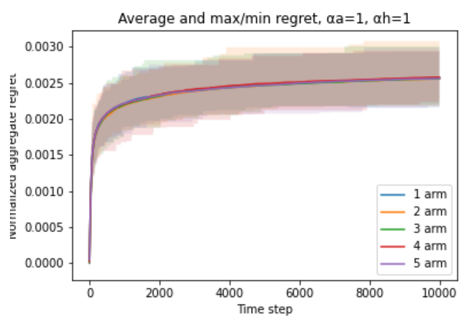
Figure 2: Simulation identical to Figure 5a 1, except that, while the best arm still has reward drawn from $\mathcal{N}(\mu = 0.5, \sigma = 0.1)$, the other arms now have reward drawn from $\mathcal{N}(\mu = 0.1, \sigma = 0.1)$. The larger gap in rewards means that $p_{i,\epsilon}$ is smaller, so the linear dependence on T does not dominate overall regret.



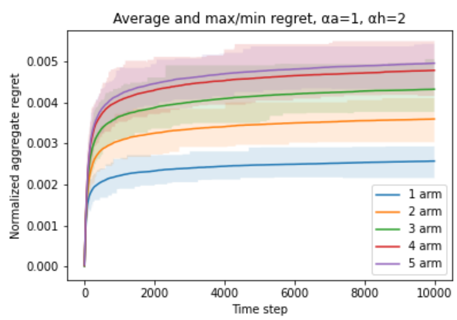
(a) $\alpha_h = 0$



(b) $\alpha_h = 0.5$

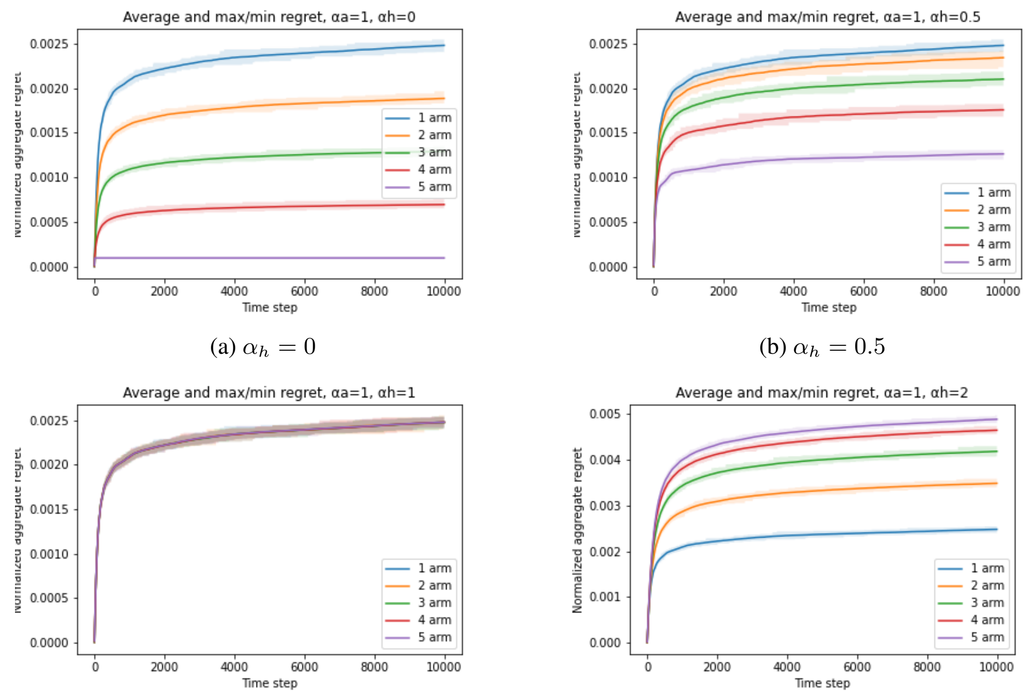


(c) $\alpha_h = 1$



(d) $\alpha_h = 2$

Figure 3: Plots show regret for (line shows average regret, shaded region is max and min regret, for 100 simulations, each with 10,000 time steps). Each simulation has 5 arms total, where arm i has reward drawn from $\mathcal{N}(\mu = \mu_i, \sigma = 0.1)$ for $\mu_i \in \{0.5, 0.4, 0.3, 0.2, 0.1\}$. As compared to Figure 1, the average regret curves seem similarly spaced apart, but the shaded region seems narrower.



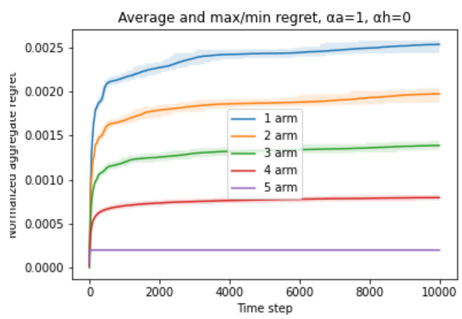
(a) $\alpha_h = 0$

(b) $\alpha_h = 0.5$

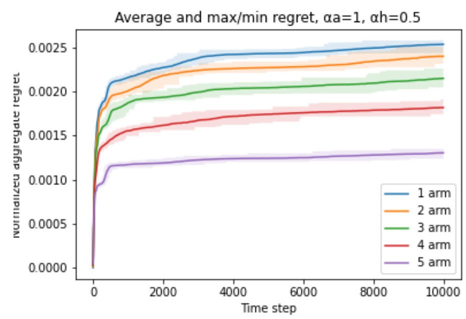
(c) $\alpha_h = 1$

(d) $\alpha_h = 2$

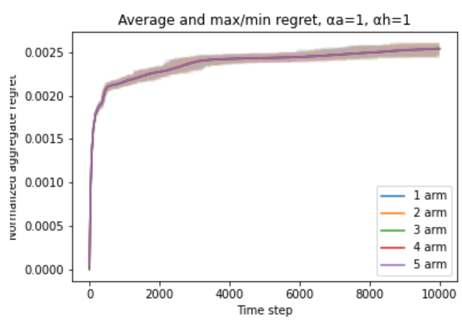
Figure 4: Plots show regret for (line shows average regret, shaded region is max and min regret, for 100 simulations, each with 10,000 time steps). Each simulation has 5 arms total, where arm i has reward drawn from $\mathcal{N}(\mu = \mu_i, \sigma = 0.01)$ for $\mu_i \in \{0.5, 0.4, 0.3, 0.2, 0.1\}$. As compared to Figure 3, the average regret curves again seem similarly spaced apart, but the shaded region seems much narrower - the maximum and minimum regret seems to have no overlap. Figure 4a shows sub-linear regret, likely because $p_{i,\epsilon}$ is smaller because of the smaller σ value.



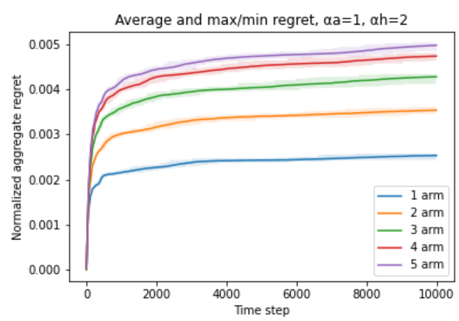
(a) $\alpha_h = 0$



(b) $\alpha_h = 0.5$



(c) $\alpha_h = 1$



(d) $\alpha_h = 2$

Figure 5: Plots show regret for (line shows average regret, shaded region is max and min regret, for 100 simulations, each with 10,000 time steps). Each simulation has 5 arms total, where arm i has reward drawn from $\mathcal{N}(\mu = \mu_i, \sigma = 0.01)$ for $\mu_i \in \{0.5, 0.45, 0.4, 0.35, 0.3\}$