A. Proof of Theorem 3

For $p, q \in (0,1)$ let $d(p,q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$ be the relative entropy between Bernoulli distributions with biases p and q respectively. For $\theta \in [0,1]^K$ let $\mathbb{E}_\theta$ denote the expectation when the algorithm interacts with the Bernoulli bandit determined by $\theta \in [0,1]^K$. Let $\theta = (1/2 + \Delta, 1/2, \ldots, 1/2)$ where $\Delta \in (0, 1/4)$ is some parameter to be tuned subsequently. Then let

$$i = \operatorname*{arg\,min}_{k>1} \mathbb{E}_\theta[N_k(T)]\,.$$

By the pigeonhole principle it follows that $\mathbb{E}_\theta[N_i(T)] \le T/(K-1)$. Then define $\phi \in [0,1]^K$ so that $\phi_j = \theta_j$ for all $j \neq i$ and $\phi_i = 1/2 + 2\Delta$. By the definitions of θ and ϕ we have

$$R_\theta(T) \ge \Delta(T - \mathbb{E}_\theta[N_1(T)]) \quad \text{and} \quad R_\phi(T) \ge \Delta\mathbb{E}_\phi[N_1(T)]\,,$$

which means that

$$R_\theta(T) \ge \frac{T\Delta}{2}\mathbb{P}_\theta(N_1(T) \le T/2) \quad \text{and} \quad R_\phi(T) \ge \frac{T\Delta}{2}\mathbb{P}_\phi(N_1(T) > T/2)\,.$$

Summing the two regrets and applying the Bretagnolle-Huber inequality shows that

$$\begin{aligned} R_\theta(T) + R_\phi(T) &\ge \frac{T\Delta}{2}\left(\mathbb{P}_\theta(N_1(T) \le T/2) + \mathbb{P}_\phi(N_1(T) > T/2)\right) \\ &\ge \frac{T\Delta}{4}\exp\left(-KL(\mathbb{P}_\theta, \mathbb{P}_\phi)\right)\,. \end{aligned}$$

The next step is to calculate the relative entropy between $\mathbb{P}_\theta$ and $\mathbb{P}_\phi$. Both bandits behave identically on all arms except action i. When action i is played the learner effectively observes a reward with bias either $\tau_m/2$ or $\tau_m(1/2+2\Delta)$. Therefore

$$KL(\mathbb{P}_\theta, \mathbb{P}_\phi) = \mathbb{E}_\theta\left[N_i(T)\right] d(\tau_m/2, \tau_m(1/2+2\Delta))\,.$$

Upper bounding the relative entropy by the χ-squared distance shows that

$$d(\tau_m/2, \tau_m(1/2+2\Delta)) \le \frac{2\left(\tau_m/2 - \tau_m(1/2+2\Delta)\right)^2}{\tau_m(1/2-2\Delta)} \le 32\tau_m\Delta^2\,,$$

where we used the assumption that $2\Delta \le 1/4$. Therefore

$$KL(\mathbb{P}_\theta, \mathbb{P}_\phi) \le 32\tau_m\Delta^2\mathbb{E}_\theta[N_i(T)] \le \frac{32\tau_m\Delta^2 T}{K-1}\,.$$

Finally we conclude that

$$R_\theta(T) + R_\phi(T) \ge \frac{T\Delta}{4}\exp\left(-\frac{32\tau_m\Delta^2 T}{K-1}\right)\,.$$

The result follows by tuning Δ.