Appendix A

Consider the Thompson samples $\theta_{i,t}$, $i = 1, \ldots, N$ for any round indexed by t. The samples for stationary optimal arm 1 and stationary suboptimal arm i are denoted by $\theta_{1,t}$ and $\theta_{i,t}$ respectively. Let $M_i(t)$ denote the event

$$M_i(t) : \begin{cases} \theta_{j,t} \leq \kappa_j,\ j \notin \{1, 2, i\}\,, \text{ and} \\ \kappa_2 < \theta_{2,t} \leq \mu_2, \end{cases} \tag{1}$$

where we choose thresholds such that $\kappa_2 > \eta$. Consider the probability of the event where suboptimal arm i is selected under the filtration $\mathcal{F}_{t-1}$ and the Thompson sample $\theta_{i,t}$ such that E_i^θ is true, i.e., $\Pr(i(t) = i | E_i^\theta(t), \mathcal{F}_{t-1})$. We have that arm i is a part of the optimal solution only if arm 1 is below the threshold κ_1, all other stationary arms are below their respective thresholds κ_j, $j \notin \{1, 2, i\}$, and the stationary optimal arm 2 is above κ_2 (in which case at least some of the optimal solutions are supported by arms i and 2). Hence, we have

$$\begin{aligned} &\Pr(i(t) = i | E_i^\theta(t), \mathcal{F}_{t-1}) \\ &\quad\leq \Pr(i(t) = i, \theta_1(t) < \kappa_1, M_i(t) | E_i^\theta(t), \mathcal{F}_{t-1}) \\ &\quad= (1 - p_{i,t}) \Pr(M_i(t) | E_i^\theta(t), \mathcal{F}_{t-1}), \end{aligned} \tag{2}$$

where the second step follows from the independence of events conditional on the filtration $\mathcal{F}_{t-1}$.

Next, we bound the probability of selecting arm 1. We observe that, conditioned on $M_i(t)$ and E_i^θ, arm 1 forms a part of the optimal solution at time t along with arm 2. Further, the probability mass assigned to arm 1 is $(\theta_{2,t} - \eta)/(\theta_{2,t} - \theta_{1,t})$. For any Thompson samples such that $\theta_2 > \kappa_2$ and $\theta_1 > \kappa_1$, the probability mass assigned to arm 1 is at least $(\kappa_2 - \eta)/(\kappa_2 - \kappa_1) = \epsilon_{1,i}$. Consequently, we have

$$\begin{aligned} &\Pr\big(i(t) = 1 | E_i^\theta(t), \mathcal{F}_{t-1}\big) \\ &\geq \Pr\big(i(t) = 1, M_i(t) | E_i^\theta(t), \mathcal{F}_{t-1}\big) \\ &= \Pr\big(M_i(t) | E_i^\theta(t), \mathcal{F}_{t-1}\big) \cdot \Pr\big(i(t) = 1 | M_i(t), E_i^\theta(t), \mathcal{F}_{t-1}\big) \\ &\geq \epsilon_{1,i} \cdot p_{i,t} \cdot \Pr\big(M_i(t) | E_i^\theta(t), \mathcal{F}_{t-1}\big), \end{aligned} \tag{3}$$

Combining (2) and (3) we get the desired result.

Appendix B

Similar to the approach in [Agrawal and Goyal, 2013], we bound the number of plays of any suboptimal arm in the following manner:

$$\begin{aligned} \mathbb{E}\big[k_i(T)\big] &= \sum_{t=1}^{T} \Pr(i(t) = i) \\ &= \sum_{t=1}^{T} \Pr(i(t) = i, E_i^\mu(t), E_i^\theta(t)) \\ &\quad + \sum_{t=1}^{T} \Pr(i(t) = i, E_i^\mu(t), \overline{E_i^\theta(t)}) \\ &\quad + \sum_{t=1}^{T} \Pr(i(t) = i, \overline{E_i^\mu(t)}) \end{aligned}$$

The last two terms of this expression are upper bounded by (17) and (16) respectively. Then, following the approach in [Agrawal and Goyal, 2013], we bound the first term of the expression above using Lemma 1, where we exploit the fact that the number of plays of arm i are a linear function of the number of playes of arm 1,

$$\begin{aligned} &\sum_{t=1}^{T} \Pr(i(t) = i, E_i^\mu(t), E_i^\theta(t)) \\ &\quad= \sum_{t=1}^{T} \mathbb{E}\Big[\Pr(i(t) = i, E_i^\mu(t), E_i^\theta(t) \big| \mathcal{F}_{t-1})\Big] \\ &\quad\leq \sum_{t=1}^{T} \mathbb{E}\left[\frac{1 - p_{i,t}}{\epsilon_{1,i} \cdot p_{i,t}} \Pr(i(t) = 1, E_i^\mu(t), E_i^\theta(t) \big| \mathcal{F}_{t-1})\right] \\ &\quad= \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}\left[\frac{1 - p_{i,t}}{\epsilon_{1,i} \cdot p_{i,t}} I(i(t) = 1, E_i^\mu(t), E_i^\theta(t))\right]\right] \\ &\quad\leq \sum_{k=0}^{T-1} \mathbb{E}\left[\left(\frac{1}{\epsilon_{1,i} \cdot p_{i,\tau_k+1}} - 1\right) \sum_{t=\tau_k+1}^{\tau_{k+1}} I(i(t) = 1)\right] \\ &\quad= \sum_{k=0}^{T-1} \mathbb{E}\left[\frac{1}{\epsilon_{1,i} \cdot p_{i,\tau_k+1}} - 1\right], \end{aligned}$$

where I is the indicator function, and we have used the fact that $\epsilon_{1,i}$ is independent of the history of plays. From (18), we have an upper bound on $\mathbb{E}(\frac{1}{p_{i,\tau_j+1}})$. By collecting the upper bounds from (16), (17), and (18), we directly obtain Lemma 2.

Appendix C

We present the pseudocode for the LinCon-KL-UCB algorithm in this section.

Algorithm 1 LinCon-KL-UCB

Input: Reward Values $r_{\{1,\ldots,N\}}$, Constraint η, c
Initialize: $k_{\{1,\ldots,N\},0} = 0, s_{\{1,\ldots,N\},0} = 0$.
for Time index $t = 1$ **to** T **do**
  if $t < N$ **then**
    $i(t) = t$
  else
    for Arm index $i = 1$ **to** N **do**
      $\tilde{\mu}_{i,t} = \max\left\{q \in \Theta : k_i d\left(\frac{k_i}{s_i}, q\right) \log(t) + c \log\log(t)\right\}$
    end for
    Solve, if feasible, the linear program:

$$LP(\tilde{\mu}_t) : \text{maximize} \sum_i x_{i,t} \tilde{\mu}_{i,t} r_i$$

$$\text{subject to } \begin{cases} \sum_i x_{i,t} \tilde{\mu}_{i,t} \geq \eta \\ \sum_i x_{i,t} = 1 \\ x_{i,t} \geq 0 \quad \forall\, i \in \{1, \ldots, N\} \end{cases}, \tag{4}$$

    if a (feasible) optimal solution existed **then**
      Sample $i(t) \sim [x_{1,t}, \ldots, x_{N,t}]$
    else
      Sample $i(t)$ uniformly from $\{1, \ldots, N\}$.
    end if
  end if
  Observe: Reward event $c_{i(t)} \in \{0, 1\}$.
  Update:
    $k_{i(t)} = k_{i(t)} + 1$
    $s_{i(t)} = s_{i(t)} + c_{i(t)}$.
end for

References

[Agrawal and Goyal, 2013] Agrawal, S. and Goyal, N. (2013). Further optimal regret bounds for thompson sampling. In *Artificial Intelligence and Statistics*, pages 99–107.