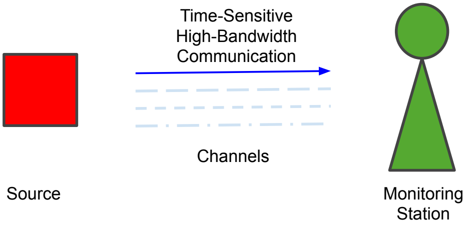
Correlated Age-of-Information Bandits

Ishank Juneja - 16D070012

Advisor: Prof. Sharayu Moharir
Deptartment of Electrical Engineering

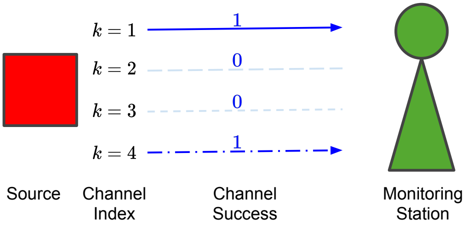
November 23, 2020

The Problem Setting



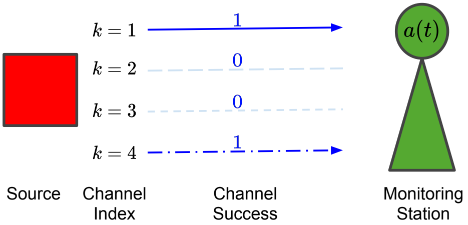
- System: A sensor node (the source) and a monitoring station
- Aim: Communicate **time-sensitive** and **high-bandwidth** information between the sensor and central monitoring station
- Problem: Find a channel selection **policy** such that cumulative performance is maximized

The Channel Selection Problem



- The available bandwidth is partitioned into K frequency channels
- Schedule channel k_t among K available channels at every time step t
- Channel either successful - 1 or unsuccessful - 0
- Assume stationary channel statistics across T trials

Scheduling Problem Formulation



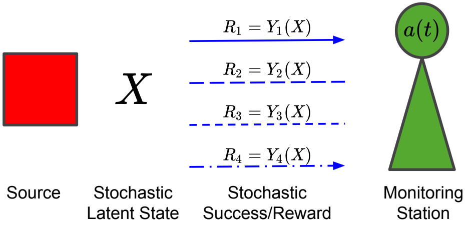
- Schedule in a manner that minimizes cumulative Age-of-Information
- AoI - $a(t)$ is the time elapsed since the most recent successful update
- Multi-Armed Bandit (MAB) framework applicable
- MAB policies applied and analysed in the work of [2]

High Bandwidth - 5G Uses Shorter Waves

Link: Image Source

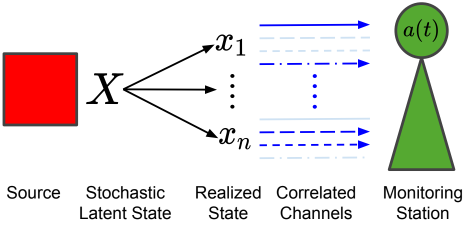
- Next generation: Higher data rates, move to higher frequency band
- 5G will use the 30GHz - 300GHz (mmWave/EHF) band
- New challenges arise: line-of-sight paths, attenuation
- Line-of-sight affects all channels, attenuation is frequency selective [3]

Performance of Channels is Correlated



- Stochastic success (reward) of arm k is $Y_k(X)$
- $Y_k(X)$ is a known deterministic function of state X
- X is a latent stochastic state with unknown distribution
- Correlation model introduced in the work of [4]

The Correlation Model



- Realizations of X lie in alphabet $\{x_1, x_2, \ldots, x_n\}$
- Realization dictates which channels would be successful if used
- Successes across channels at a given time are correlated depending on the functions $Y_1, Y_2, \ldots, Y_K$

Contributions of Work

- Variants of the UCB and Thompson Sampling policies that account for correlation analysed for the AoI regret metric
- Lower bound on AoI regret of $\Omega(\log T)$ for certain problem instances
- An upper bound on AoI regret for Correlated-UCB (CUCB) and Correlated-Thompson Sampling
- Simulations to compare the performance of UCB, Thompson Sampling and their correlated variants

Correlated Bandit Model Definitions

- Construct a Correlated Bandit instance with K arms
- Sample arm k - Obtain reward $Y_k(X)$, mean reward $\mu_k = \mathbb{E}_X[Y_k(X)]$
- Sub-optimality gap: $\Delta_k = \mu^* - \mu_k$

Definition (Expected pseudo reward and pseudo gap)

Pseudo reward for arm ℓ with respect to arm k is given by,

$$s_{\ell,k}(r) = \sup_{x:Y_k(x)=r} Y_\ell(x).$$

Expected pseudo reward in turn is defined as,

$$\phi_{\ell,k} = \mathbb{E}_X[\, s_{\ell,k}(Y_k(X))\,].$$

The pseudo gap is defined as $\tilde{\Delta}_{\ell,k^*} = \mu^* - \phi_{\ell,k^*}$.

Correlated Bandit Model Definitions

- If $\tilde{\Delta}_{\ell,k^*} > 0$, then arm ℓ is non-competitive
- Arm ℓ is *competitive* if $\tilde{\Delta}_{\ell,k^*} \leq 0$ and *strictly competitive* if the inequality is strict
- C denotes the number of competitive arms excluding arm k^*.

Definition (Expected pseudo reward and pseudo gap)

Pseudo reward for arm ℓ with respect to arm k is given by,

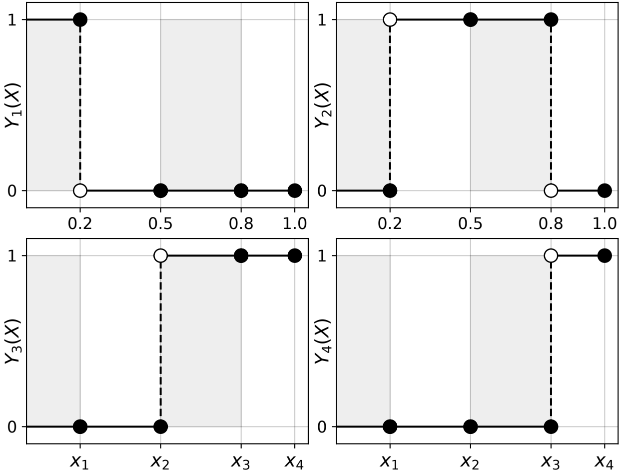
$$s_{\ell,k}(r) = \sup_{x : Y_k(x) = r} Y_\ell(x).$$

Expected pseudo reward in turn is defined as,

$$\phi_{\ell,k} = \mathbb{E}_X[\, s_{\ell,k}(Y_k(X))\,].$$

The pseudo gap is defined as $\tilde{\Delta}_{\ell,k^*} = \mu^* - \phi_{\ell,k^*}$.

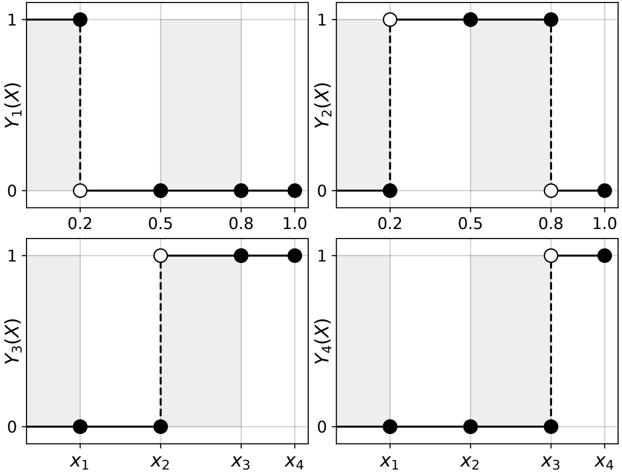
Example 1: Correlated Bandit Model



$$\mu_1 = 1 \times 0.2 = 0.2$$
$$\mu_2 = 1 \times 0.3 + 1 \times 0.3 = 0.6$$
$$\mu_3 = 1 \times 0.3 + 1 \times 0.2 = 0.5$$
$$\mu_4 = 1 \times 0.2 = 0.2$$

- Arm 2 is optimal
- Optimal arm $k^* = 2$.
- $\mu^* = \mu_2 = 0.6$

Example 1: Correlated Bandit Model

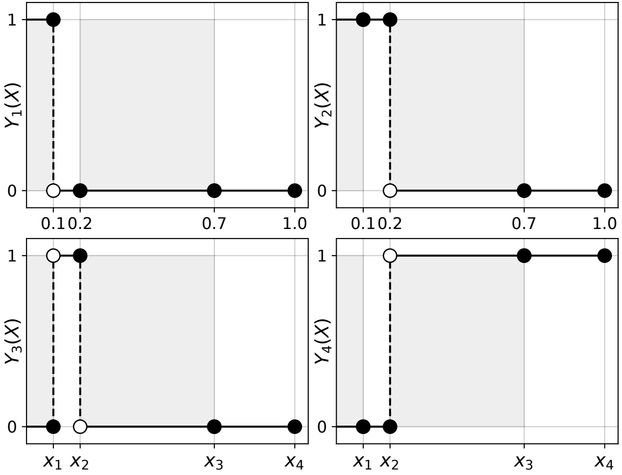


Using Definition 1,

$$\phi_{1,2} = 1 \times 0.4 = 0.4$$
$$\phi_{2,2} = \mu_2 = 0.6$$
$$\phi_{3,2} = 1 \times 0.4 + 1 \times 0.6 = 1.0$$
$$\phi_{4,2} = 1 \times 0.4 = 0.4$$

- Only competitive sub-optimal is arm 3
- For this example $C = 1$.

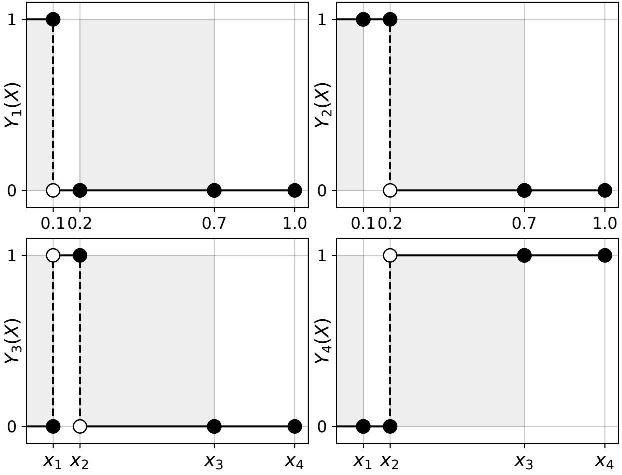
Example 2: Correlated Bandit Model



$$\mu_1 = 1 \times 0.1 = 0.1$$
$$\mu_2 = 1 \times 0.1 + 1 \times 0.1 = 0.2$$
$$\mu_3 = 1 \times 0.1 = 0.1$$
$$\mu_4 = 1 \times 0.5 + 1 \times 0.3 = 0.8$$

- Arm 4 is optimal
- Optimal arm $k^* = 4$.
- $\mu^* = \mu_4 = 0.8$

Example 2: Correlated Bandit Model



Using Definition 1,

$$\phi_{1,4} = 1 \times 0.2 = 0.2$$
$$\phi_{2,4} = 1 \times 0.2 = 0.2$$
$$\phi_{3,4} = 1 \times 0.2 = 0.2$$
$$\phi_{4,4} = \mu_4 = 0.8$$

- No competitive sub-optimal arms
- For this example $C = 0$.

Age-of-Information (AoI) Definition

The current AoI is the time elapsed since the last successful update.
More formally,

Definition (Age-of-Information (AoI))

At the start of time slot t, let $a(t)$ denote the AoI at the central monitoring station and let $u(t)$ denote the time index at which the recent most successful update was received by the monitoring station. Then, $a(t) = t - u(t)$. Alternatively,

$$a(t) = \begin{cases} 1 & \text{if the update in } t-1 \text{ succeeds} \\ a(t-1)+1 & \text{otherwise.} \end{cases}$$

AoI Regret Metric

Definition (Age-of-Information Regret (AoI Regret))

AoI regret for a policy ρ, over T slots is given by,

$$R_\rho(T) = \sum_{t=1}^{T} \mathbb{E}[a_\rho(t) - a^*(t)] = \sum_{t=1}^{T} \mathbb{E}[a_\rho(t)] - \frac{T}{\mu^*}, \tag{1}$$

where (1) follows from the expectation of a geometric random variable with parameter μ^*.

- $a_\rho(t)$ denotes the AoI in time slot t under policy ρ.
- $a^*(t)$ is the AoI under the optimal policy.

AoI Regret Lower Bound - Policy Family

Lower Bound on AoI regret is derived for a certain α-consistent family of policies.

Definition (α-consistent policies [6])

Let k_s denote the index of the channel scheduled in time-slot s. The index k^* denotes the index of the optimal channel. A scheduling policy is called α-consistent, for a constant $\alpha \in (0,1)$, if there exists an instance dependent constant M such that,

$$\mathbb{E}\left[\sum_{s=1}^{t} \mathbb{1}\{k_s = k\}\right] \leq M t^{\alpha}, \; \forall \, k \neq k^*. \tag{2}$$

AoI Regret Lower Bound

Theorem (Lower bound on AoI regret)

If a bandit instance I *has at least one strictly competitive arm* k *with* $\tilde{\Delta}_{k,k^*} < 0$*, then for any* α*-consistent policy* ρ*, we have,*

$$R_\rho(T) \geq \max_{k \in \mathcal{C}'} \frac{\Delta_k}{\mathrm{D}(P_k, P'_k)} \frac{(1-\alpha)\log T - \log(4M)}{\mu^*}.$$

Otherwise, if $\tilde{\Delta}_{k,k^*} \geq 0 \; \forall \, k \in [K]$*,* $R_\rho(T) \geq 0$*.*

- $\mathrm{D}(P_k, P'_k)$ is the KL divergence between the reward distribution of arm k and a suitably chosen perturbed reward distribution
- $\mathcal{C}'$ is the set of strictly competitive arms
- M is an instance dependent constant as in Definition 4

Definitions Associated With CUCB

Let C denote the number of competitive sub-optimal arms and $\mathcal{C}$ denote the set of competitive arms inclusive of k^*.

$$t_o = \inf\left\{\tau \geq 2 : \Delta_{\min}, \tilde{\Delta}_{k,k^*} \geq 4\sqrt{\frac{2K\log\tau}{\tau}}\right\},$$

$$U^{(nc)}_{k,\text{CUCB}} = Kt_0 + K^3\sum_{t=Kt_0}^{T} 2\left(\frac{t}{K}\right)^{-2} + \sum_{t=1}^{T} 3t^{-3},$$

$$U^{(c)}_{k,\text{CUCB}} = 8\frac{\log(T)}{\Delta_k^2} + \left(1 + \frac{\pi^2}{3}\right) + \sum_{t=1}^{T} 2Kt\exp\left(-\frac{t\Delta_{\min}^2}{2K}\right).$$

Where, $\Delta_{\min} = \min_{k\neq k^*}\mu^* - \mu_k$.

Upper bound on AoI Regret Under CUCB

Theorem (Upper bound on AoI regret under CUCB)

Let $\mu_{\min} = \min_k \mu_k$, *then for* $T > t_0$,

$$\mathbb{E}[R_{\text{CUCB}}(T)] \leq \frac{1-\mu^*}{\mu^*\mu_{\min}} + \Big(\frac{1}{\mu_{\min}} - \frac{1}{\mu^*}\Big)\Bigg(\sum_{k' \in [K]\setminus\mathcal{C}} \Delta_{k'} U^{(nc)}_{k,\text{CUCB}} + \sum_{k \in \mathcal{C}\setminus\{k^*\}} \Delta_k U^{(c)}_{k,\text{CUCB}}\Bigg)$$

$$= O(1) + O(C \log T),$$

and for $T \leq t_0$,

$$\mathbb{E}[R_{\text{CUCB}}(T)] \leq \Big(\frac{1}{\mu_{\min}} - \frac{1}{\mu^*}\Big)T.$$

Definitions Associated With C - Thompson Sampling

Let C denote the number of competitive sub-optimal arms and $\mathcal{C}$ denote the set of competitive arms inclusive of k^*.

$$t_b = \inf \left\{ \tau \geq \exp\left(11\beta\right) : \Delta_{\min}, \tilde{\Delta}_{k,k^*} \geq 6\sqrt{\frac{2K\beta\log\tau}{\tau}} \right\},$$

$$U_{k,\mathrm{CTS}}^{(nc)} = Kt_b + \sum_{t=1}^{T} 3t^{-3} + K^2 \sum_{t=Kt_b}^{T} \left((2K+3)\left(\frac{t}{K}\right)^{-2} + \left(\frac{t}{K}\right)^{1-2\beta} \right),$$

$$U_{k,\mathrm{CTS}}^{(c)} = 18\frac{\log(T\Delta_k^2)}{\Delta_k^2} + \exp\left(11\beta\right) + \frac{9}{\Delta_k^2} + \sum_{t=1}^{T} 2Kt\exp\left(-\frac{t\Delta_{\min}^2}{2K} \right).$$

Where, $\Delta_{\min} = \min_{k \neq k^*} \mu^* - \mu_k$ and $\beta > 1$ is a parameter of the Thompson Sampling with Gaussian priors algorithm.

Upper Bound on AoI Regret Under C - Thompson Sampling

Theorem (Upper bound on AoI regret under CTS)

Then, for any choice of $\beta > 1$ *and for* $T > t_b$,

$$\mathbb{E}[R_{\text{CTS}}(T)] \leq \frac{1-\mu^*}{\mu^* \mu_{\min}} + \Big(\frac{1}{\mu_{\min}} - \frac{1}{\mu^*}\Big)\Bigg(\sum_{k' \in [K] \setminus \mathcal{C}} \Delta_{k'} U^{(nc)}_{k,\text{CTS}} + \sum_{k \in \mathcal{C} \setminus \{k^*\}} \Delta_k U^{(c)}_{k,\text{CTS}}\Bigg)$$

$$= O(1) + O(C \log T),$$

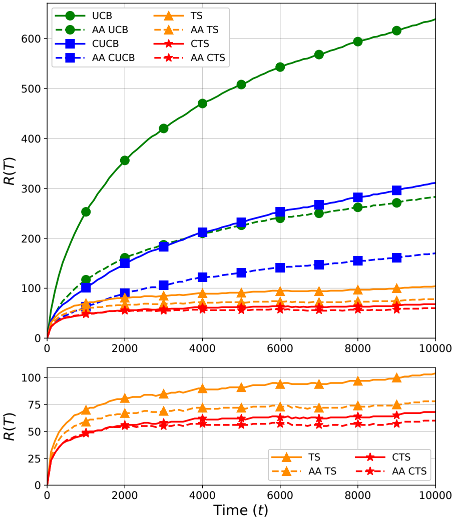
and for $T \leq t_b$,

$$\mathbb{E}[R_{\text{CTS}}(T)] \leq \big(\frac{1}{\mu_{\min}} - \frac{1}{\mu^*}\big)T.$$

Discussion About Regret Bounds

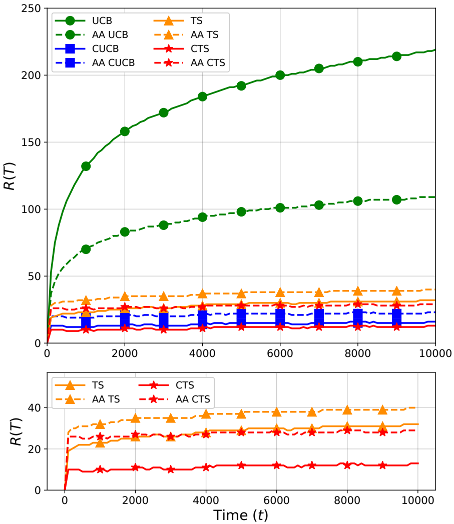
- If $C > 0$ for a Correlated Bandit instance, and if at least one arm is strictly competitive, then the lower bound and upper bound on AoI regret are both $\mathrm{O}(\log T)$
- When there are no competitive arms ($C = 0$), there is no meaningful lower bound on the expected AoI regret
- The $C = 0$ case agrees with the fact that the set $\mathcal{C}\setminus\{k^*\}$ being empty results in an $\mathrm{O}(1)$ upper bound on AoI regret
- For these cases of Correlated Bandit instances AoI regret bounds are order-optimal

Simulation Results for Example 1



- An AoI-aware policy follows the policy or is greedy based on a Threshold [2]
- CTS and its AoI-aware variant perform the best on AoI regret
- CUCB and CTS have significantly lower AoI regret compared to UCB and TS

Simulation Results for Example 2



- Bandit instance in Example 2 had no competitive sub-optimal arms, i.e. $C = 0$
- As predicted by the bounds in the preceding Theorems both CUCB and CTS have constant expected AoI regret
- The AoI-aware variant of a policy need not perform better than its parent policy as is the case for CUCB, CTS and TS in this example

Key Takeaways and Conclusion

- 5G frequency band means the problem of line-of-sight occlusion becomes more significant
- Exploit correlation to identify certain channels as sub-optimal within a few time steps
- Assumption: Deterministic reward functions
- Strength: Once determined, reward functions can be applied in other communication systems with a similar configuration but a different and unknown distribution of X
- Distribution agnostic model and algorithms analysed in this work would be highly beneficial in such scenarios

Directions Planned to be Pursued

- Dropping the assumption of channel reward distribution being stationary across time
- Theoretical analysis of AoI-aware policies
- Alternate correlation models to exploit 0-1 Binary rewards
- AoI regret upper bound for C - Thompson Sampling with Beta priors

References I

[1] I. Juneja, S. Fatale, and S. Moharir, "Correlated age-of-information bandits," *arXiv preprint arXiv:2011.05032*, 2020.

[2] S. Fatale, K. Bhandari, U. Narula, S. Moharir, and M. K. Hanawal, "Regret of age-of-information bandits," *arXiv preprint arXiv:2001.09317*, 2020.

[3] K. C. Huang and Z. Wang, *Millimeter Wave Characteristics.* John Wiley & Sons, Ltd, 2011, ch. 1, pp. 1–31.

[4] S. Gupta, S. Chaudhari, G. Joshi, and O. Yağan, "Multi-armed bandits with correlated arms," *arXiv preprint arXiv:1911.03959*, 2019.

[5] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*, March 2012, pp. 2731–2735.

[6] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[7] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285–294, 1933.

References II

[8] L. Zhou, "A survey on contextual multi-armed bandits," 2015.

[9] V. Dani, T. Hayes, and S. Kakade, "Stochastic linear optimization under bandit feedback." 01 2008, pp. 355–366.

[10] A. Kosta, N. Pappas, V. Angelakis *et al.*, "Age of information: A new concept, metric, and tool," *Foundations and Trends® in Networking*, vol. 12, no. 3, pp. 162–259, 2017.

[11] V. Tripathi and S. Moharir, "Age of Information in Multi-Source Systems," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[12] P. R. Jhunjhunwala and S. Moharir, "Age-of-Information aware scheduling," *SPCOM*, 2018.

[13] I. Kadota, A. Sinha, and E. Modiano, "Optimizing age of information in wireless networks with throughput constraints," in *Proc. INFOCOM*, 2018.

[14] B. Sombabu and S. Moharir, "Age-of-Information Aware Scheduling for Heterogeneous Sources," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '18. New York, NY, USA: ACM, 2018, pp. 696–698.

References III

[15] A. B. Tsybakov, *Introduction to Nonparametric Estimation*, ser. Springer series in statistics. Springer, 2009.

[16] T. Lattimore and C. Szepesvári, *Bandit Algorithms*. Cambridge University Press, 2020.

End of Slides

Thank You!

Appendix: CUCB Decision Making Algorithm [4] - I

Input: Pseudo-rewards $s_{\ell,k}(r)$
Initialize: Set $\hat{\mu}_k$, $\hat{\phi}_{\ell,k}$ and n_k as $0 \; \forall \; k \in [K]$.
while $1 \leq t \leq K$ **do**
  Schedule update on Channel $k_t = t$
  Receive reward r_t drawn from Ber(μ_{k_t})
  $\hat{\mu}_{k_t} = r_t$
  $n_{k_t}(t) = 1$
  $t = t + 1$
end while
while $t \geq K + 1$ **do**
  Find $\mathcal{S}_t = \{k : n_k(t-1) \geq \frac{t-1}{K}\}$, the set of arms pulled a significant number of times till $t - 1$. Define $k^{\text{emp}}(t) = \arg\max_{k \in \mathcal{S}_t} \hat{\mu}_{k_t}$
  Initialize the empirically competitive set $\mathcal{A}_t$ as $\{\}$
  for $k \in [K]$ **do**

Appendix: CUCB Decision Making Algorithm [4] - II

if $\min_{\ell \in S_t} \hat{\phi}_{k,\ell}(t) \geq \hat{\mu}_{k^{\text{emp}}}(t)$ **then**

Add empirically competitive arms k to the set: $\mathcal{A}_t = \mathcal{A}_t \cup \{k\}$

end if

end for

Schedule update on Channel k_t such that, $k_t = \arg\max_{k \in \mathcal{A}_t \cup \{k^{\text{emp}}(t)\}} \hat{\mu}_{k_t} + \sqrt{\frac{2 \log t}{n_k(t-1)}}$

Receive reward r_t drawn from Ber(μ_{k_t})

$\hat{\mu}_{k_t} = (\hat{\mu}_{k_t} \cdot n_{k_t}(t-1) + r_t)/(n_{k_t}(t-1) + 1)$

$n_{k_t}(t) = n_{k_t}(t-1) + 1$

$\hat{\phi}_{k,k_t} = \sum_{\tau : k_\tau = k_t} s_{k,k_\tau}(r_\tau)/n_{k_t}(t) \; \forall \, k \neq k_t$

$t = t + 1$

end while

Appendix: C-TS Decision Making Algorithm [4] - I

Input: Pseudo-rewards $s_{\ell,k}(r)$

Initialize: Set the number of successes, $S_k(t)$, failures, $F_k(t)$, $\hat{\mu}_k$, $\hat{\phi}_{\ell,k}$ and n_k as 0 $\forall$ $k \in [K]$.

while $t \geq 1$ **do**

  Find $\mathcal{S}_t = \{k : n_k(t-1) \geq \frac{t-1}{K}\}$, the set of arms pulled a significant number of times till $t-1$. Define $k^{\text{emp}}(t) = \arg\max_{k \in \mathcal{S}_t} \hat{\mu}_{k_t}$

  Initialize the empirically competitive set $\mathcal{A}_t$ as $\{\}$

  for $k \in [K]$ **do**

    if $\min_{\ell \in \mathcal{S}_t} \hat{\phi}_{k,\ell}(t) \geq \hat{\mu}_{k^{\text{emp}}}(t)$ **then**

      Add empirically competitive arms k to the set: $\mathcal{A}_t = \mathcal{A}_t \cup \{k\}$

    end if

  end for

  For each k in $[K]$, draw a sample $\theta_k(t)$, where, $\theta_k(t) \sim \text{Beta}(S_k(t-1)+1, F_k(t-1)+1)$

Appendix: C-TS Decision Making Algorithm [4] - II

Schedule update on Channel k_t such that $k_t = \arg\max_{k \in \mathcal{A}_t \cup \{k^{\text{emp}}(t)\}} \theta_k(t)$

Receive reward r_t drawn from Ber(μ_{k_t})

$S_{k_t}(t) = S_{k_t}(t-1) + r_t$

$F_{k_t}(t) = F_{k_t}(t-1) + (1 - r_t)$

$\hat{\mu}_{k_t} = (\hat{\mu}_{k_t} \cdot n_{k_t}(t-1) + r_t)/(n_{k_t}(t-1) + 1)$

$n_{k_t}(t) = n_{k_t}(t-1) + 1$

$\hat{\phi}_{k,k_t} = \sum_{\tau : k_\tau = k_t} s_{k,k_\tau}(r_\tau)/n_{k_t}(t) \, \forall \, k \neq k_t$

$t = t + 1$

end while