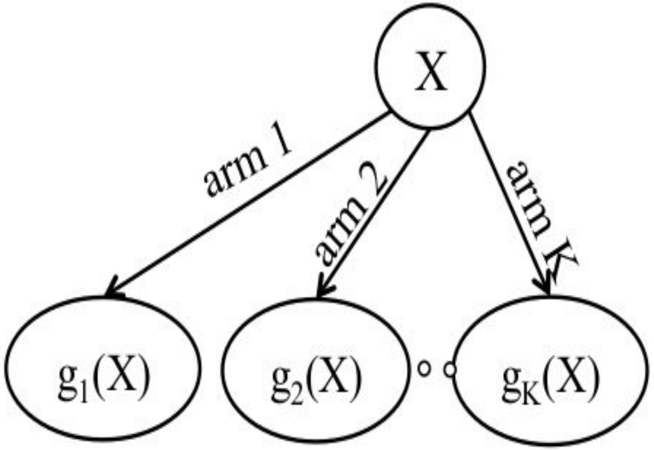
Correlated Bandits

GROUP 11
ISHANK JUNEJA 16D070012
DHRUV VARSHNEY 16D070033
DEVANSHU SINGH GAHARWAR 16D070042

Introduction : What are correlated bandits ?

- Standard MAB Setup
- Independence assumption between arms relaxed
- Correlation between arms can be exploited if present
    - Skip pulling some arms based on correlation

Problem Formulation



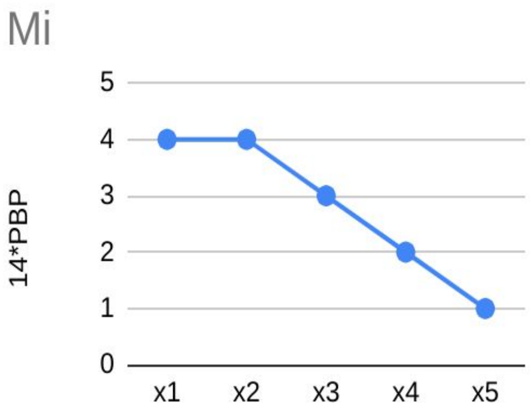
- X is the hidden random variable
- $g_1(X)$, $g_2(X)$, … , $g_k(X)$ are the dependent reward functions
- g_1 ,g_2 ….g_k are known functions.
  - Assumption Valid ???

Motivating Example

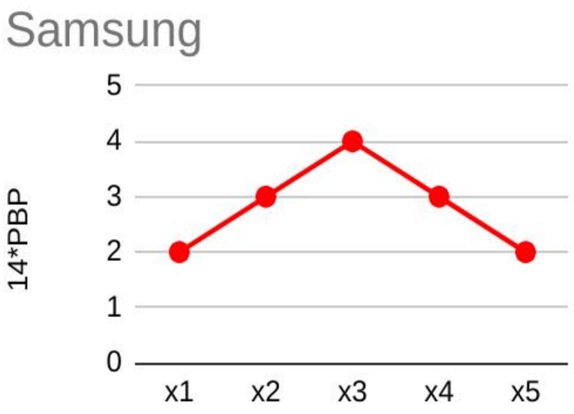
- Consider the case of Amazon expanding to a new country
- k arms ≡ k mobile companies
  - $g_1, g_2 \ldots g_k$ = product buying probability (PBP)
- Random variable ≡ Discrete Income levels
- $g_1, g_2 \ldots g_k \rightarrow$ found using paid surveys.

Fact : Amazon to start operating in Bangladesh in 2020. Refer [this]

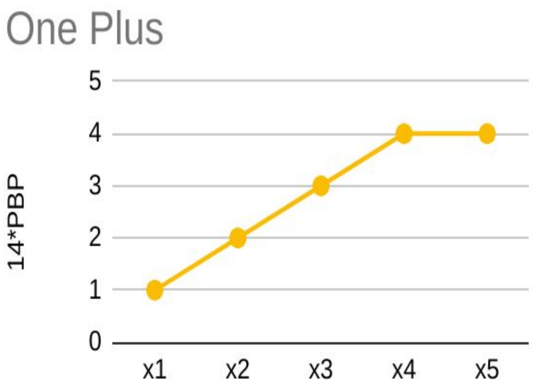
Motivating Example (Continued)



Income level -->



Income level -->



Income level -->

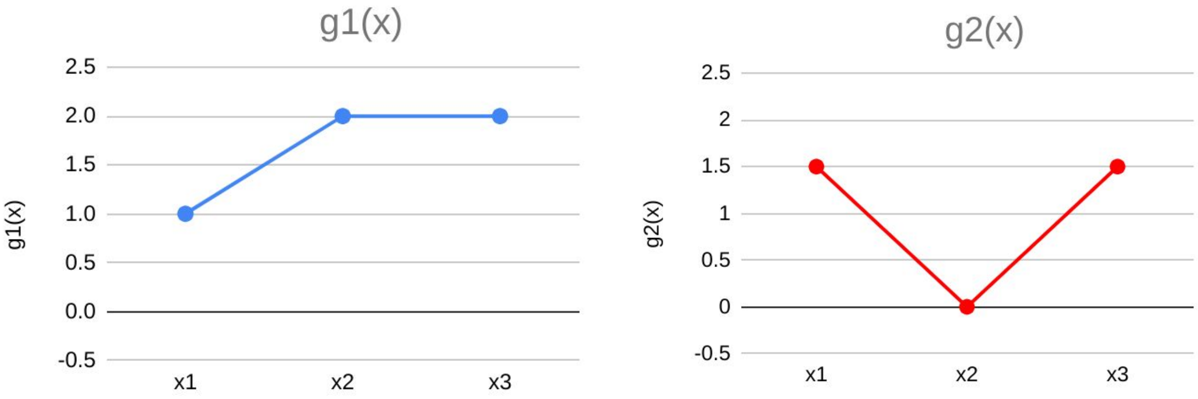
Approach 1.0

- C-UCB algorithm proposed
- Need to classify arms as Competitive and Non-Competitive
- How to decide whether an arm is competitive ?
  - Pseudo Reward of arm ℓ wrt k : $s_{\ell,k}(r) \triangleq \max_{x:g_k(x)=r} g_l(x)$
  - Expected Pseudo Gap of arm ℓ wrt k : $\Delta_{\ell,k} \triangleq \mu_k - E[s_{\ell,k}(g_k(X))]$
  - Arm ℓ is **non-competitive** wrt k if pseudo gap is positive.

Approach 1.0 (Continued)

- Compute the expected quantities empirically (Law of large numbers)
  - Empirical reward of arm k : $\mu_k = \sum_{\tau} \frac{1_{k(\tau)=k}\, g_{k(\tau)}(X(\tau))}{n_k(t)}$
  - Empirical Pseudo Reward : $\Phi_{l,k}(t) \triangleq \sum_{\tau} \frac{1_{k(\tau)=k}\, s_{l,k}(r_{\tau}))}{n_k(t)}$
  - $\mu_k > \Phi_{l,k}(t) \Rightarrow$ Arm l is non-competitive wrt arm k

Example :

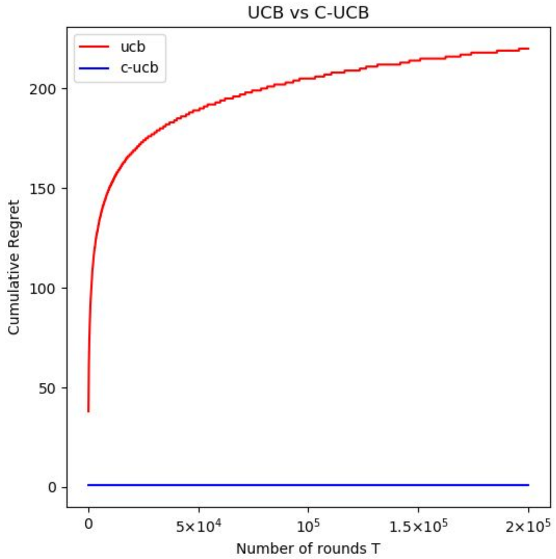


- Arm 1 pulled 10 times out of which reward 1 (3 times) and 2 (7 times)
  - $\mu_1 = (1*3 + 2*7)/10 = 1.7$
  - $s_{21}(1) = s_{21}(2) = 1.5 \Rightarrow \Phi_{21}(t) = 1.5$
  - Clearly $\Delta_{21} = \mu_1 - \Phi_{21}(t) > 0 \Rightarrow$ Arm 2 is non - competitive

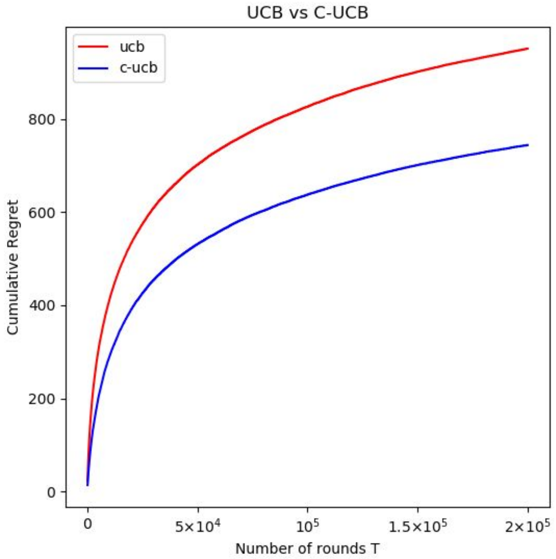
C-UCB ALGORITHM

- Initialise using standard UCB method ($n_k = 0$, $I_k = \infty$)
- For every iteration t do :
  1. Choose reference arm.
  2. Find the empirically competitive set wrt reference arm.
  3. Apply UCB over the set of competitive arms to get optimal arm k_t
  4. For all arms $k \neq k_t$ update the empirical pseudo rewards.
  5. Update the standard UCB parametres (n_k, I_k)
  6. Update the empirical reward for arm k_t

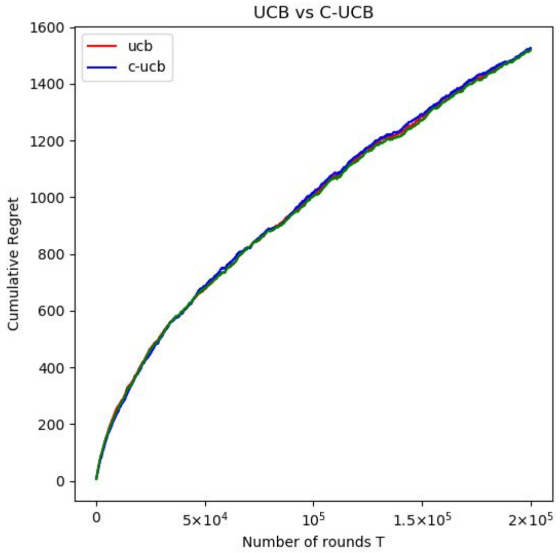
Simulation Results



No Competitive arm

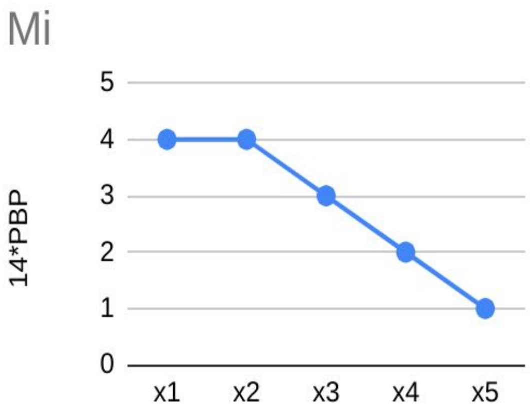


1 Competitive arm

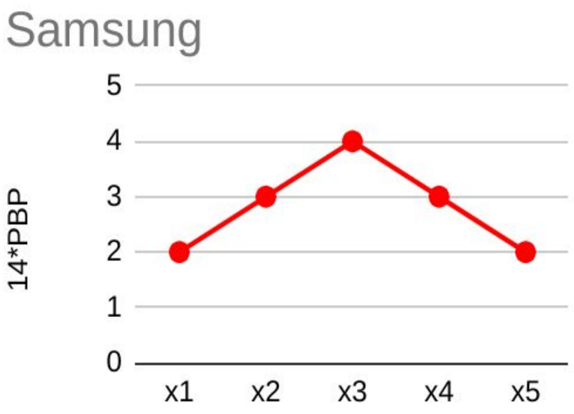


All arms are competitive

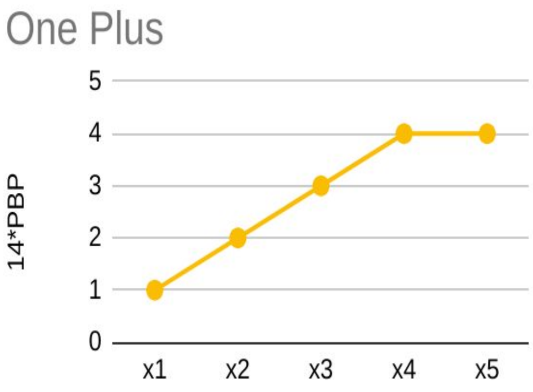
Reward Functions for Simulation Results



Income level -->



Income level -->



Income level -->

Our Contribution
A New Algorithm : Uni-C-UCB

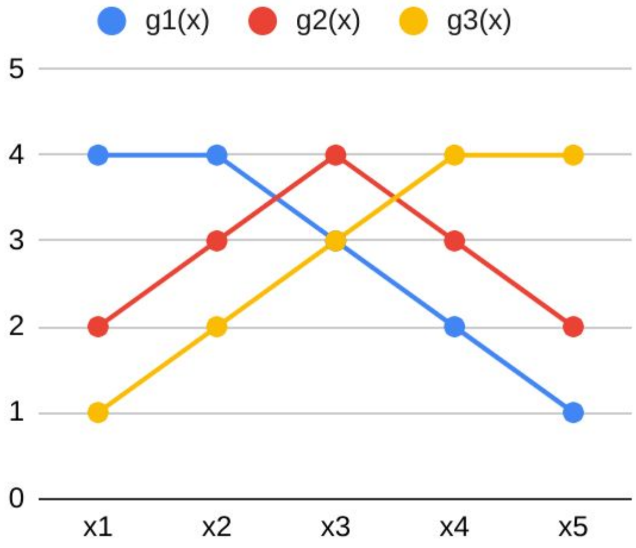
Approach 2.0

- Underlying distribution of X which can be learnt is used.
- How to decide whether an arm is competitive ?
  - **Pseudo distribution** - Empirically estimate the pmf(s)
  - **Confidence Set** (C*) :
    - First sort the pmf in decreasing order of probability.
    - C* = {1,2 ...j} where j is the minimum k s.t. $\sum_{i=1}^{k} p(x_i) > 1 - \epsilon$
  - Arm k is **Non-Competitive** if $g_k(x) < g_j(x)\ \forall\ x \in C^*$ and some arm j.

Example :

Consider the following 3 cases :

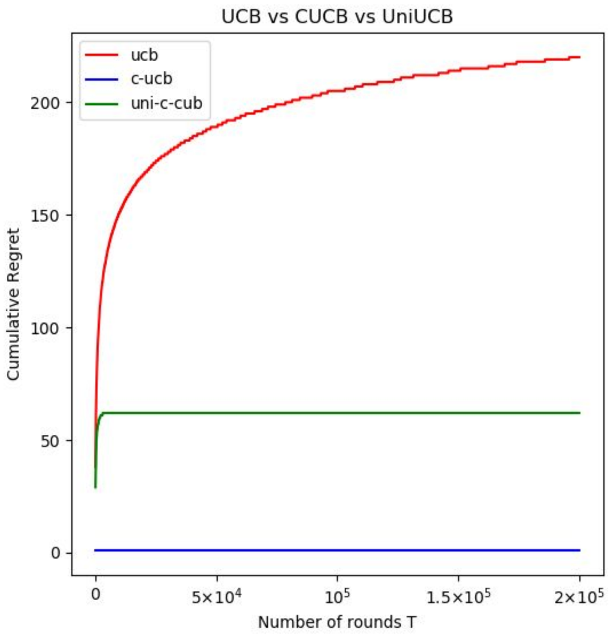
- If $\{x_1,x_2\} \in C^* \Rightarrow$
  Arms 1 is competitive
- If $\{x_2,x_3,x_4\} \in C^* \Rightarrow$
  All arms are competitive
- If $\{x_1,x_2,x_3\} \in C^* \Rightarrow$
  Arms 1,2 are competitive



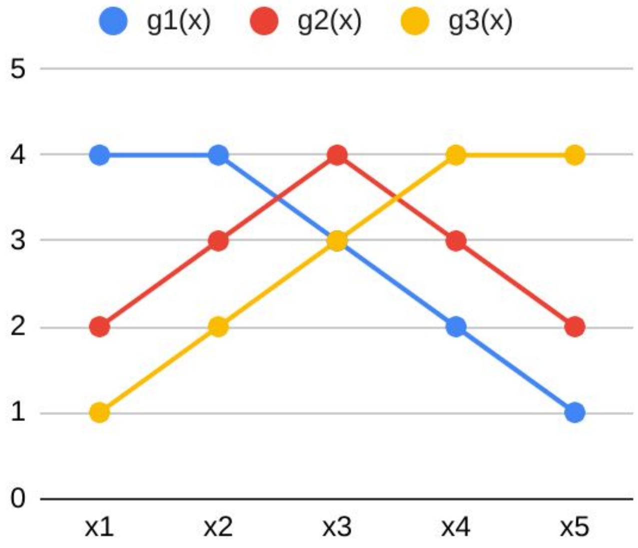
Uniform C-UCB ALGORITHM

- Initialise using standard UCB method (n_k = 0, I_k = ∞)
- Initialise C* = support of r.v. X and ∈ = 0.1 (tuneable)
- For every iteration t do :
  1. Find the competitive set using C*
  2. Apply UCB over the set of competitive arms to get optimal arm k_t
  3. Update the pseudo-distribution using Bayesian updates
  4. Update the Confidence Set (C*)
  5. Update the standard UCB parametres (n_k, I_k)

Simulation Results

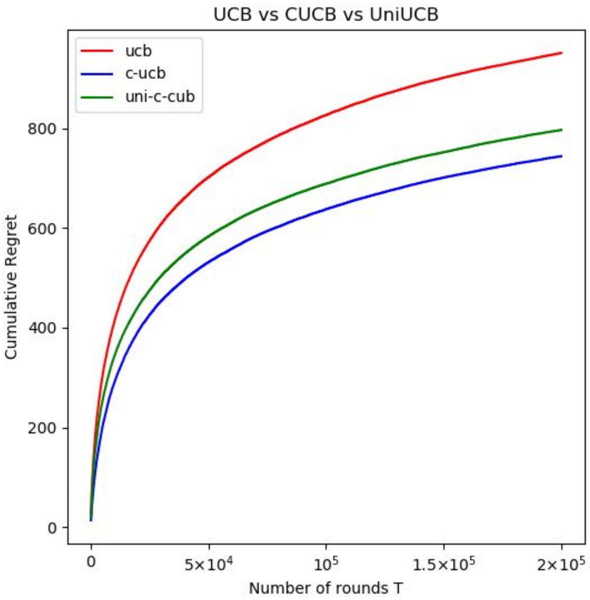


No Competitive arm

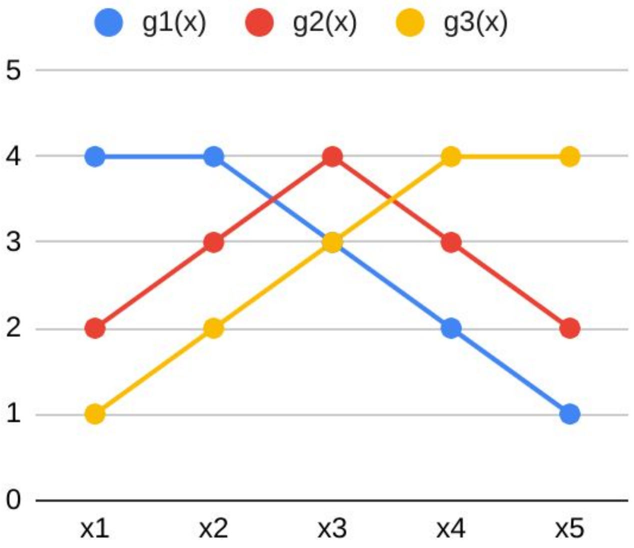


0.8 0.1 0.04 0.03 0.03

pmf

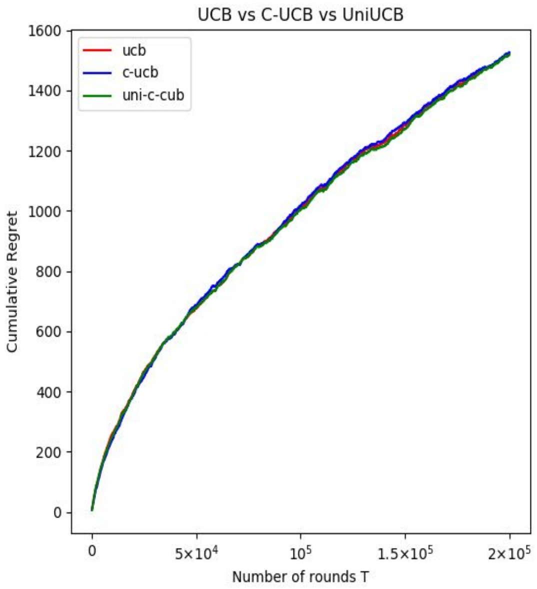


1 Competitive arm

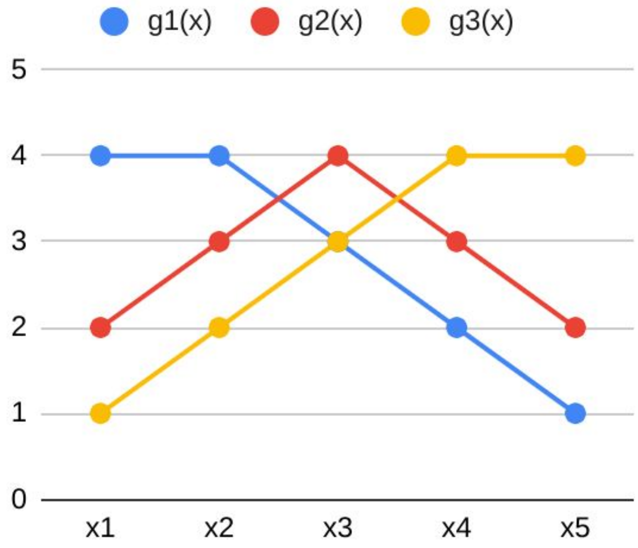


0.03 0.07 0.41 0.245 0.245

pmf

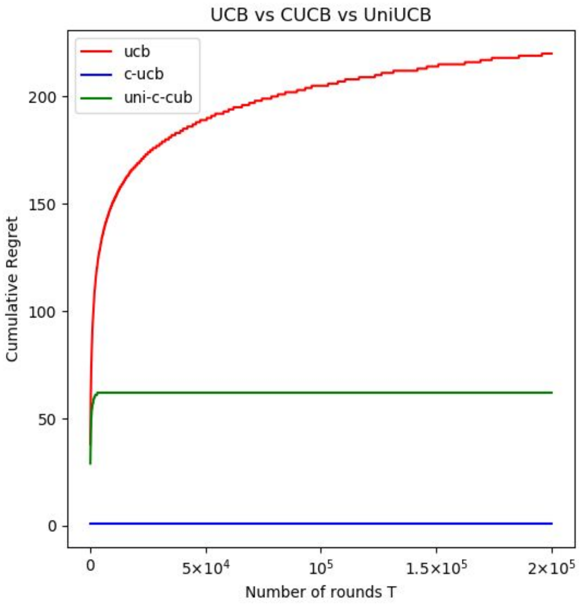


All arms are competitive

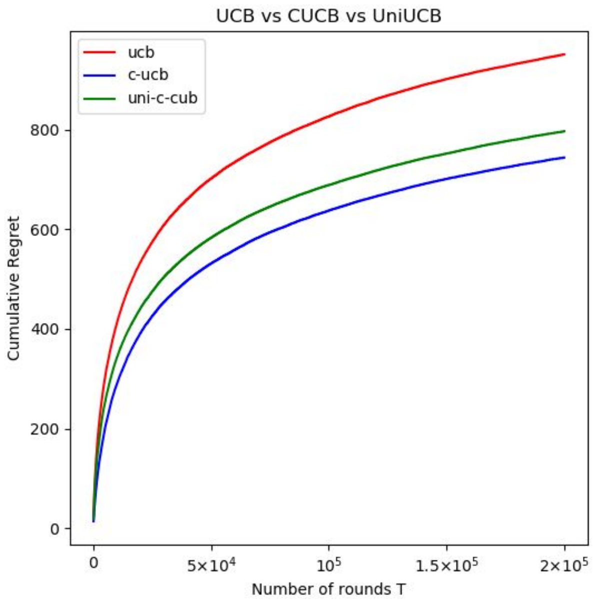


0.13 0.3 0.24 0.2 0.13

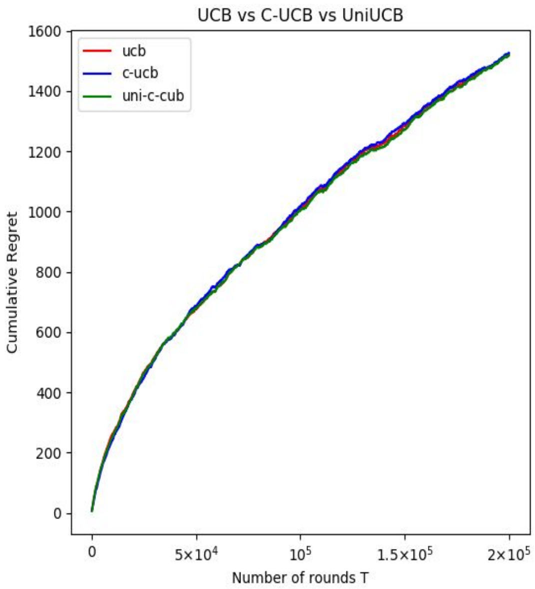
pmf



No Competitive arm



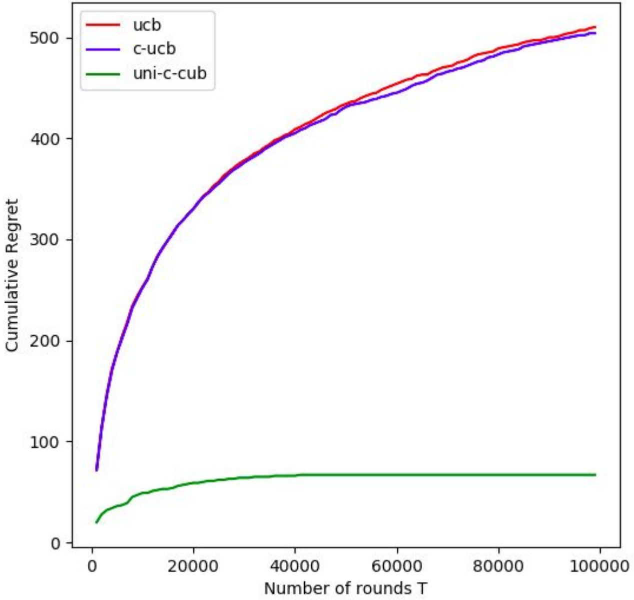
1 Arm Competitive

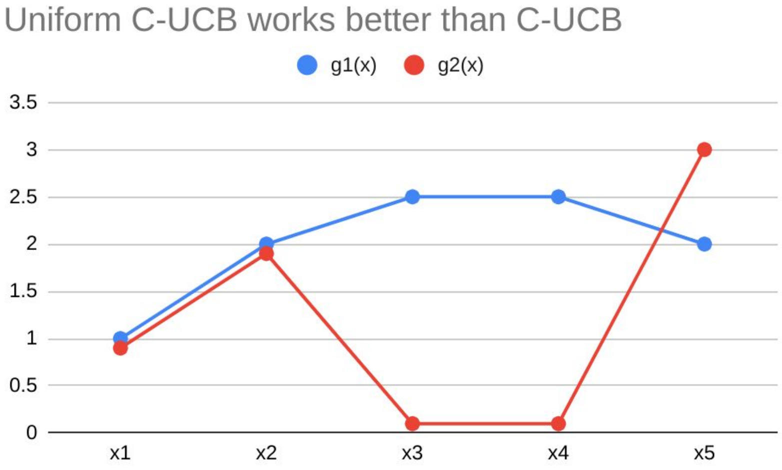


All Arms Competitive

Can Uni-C-UCB outperform C-UCB ?

Yes, Indeed !





0.45 0.45 0.045 0.045 0.01

pmf

Concluding Remarks

- Two approaches presented
- Both perform better than UCB by exploiting correlation
- Which approach is better ?
  - Depends on the exact nature of functions
- In the report, we would include our attempt on the regret analysis.

Thank You

Questions ?

Algorithm 1 C-UCB Correlated UCB Algorithm

Input: Reward Functions $\{g_1, g_2 \dots g_K\}$
Initialize: $n_k = 0, I_k = \infty$ for all $k \in \{1, 2, \dots K\}$
for each round t **do**
  Find $k^{\max} = \arg\max_k n_k(t-1)$, the arm that has been pulled most times until round $t-1$
  Initialize the empirically competitive set $\mathcal{A} = \{1, 2, \dots, K\} \setminus \{k^{\max}\}$.
  for $k \neq k^{\max}$ **do**
    if $\hat{\mu}_{k^{\max}} > \hat{\phi}_{k,k^{\max}}$ **then**
      Remove arm k from the empirically competitive set: $\mathcal{A} = \mathcal{A} \setminus \{k\}$
    end if
  end for
  Apply UCB1 over arms in $\mathcal{A} \cup \{k^{\max}\}$ by pulling arm $k_t = \arg\max_{k \in \mathcal{A} \cup \{k^{\max}\}} I_k(t-1)$
  Receive reward r_t, and update $n_{k_t} = n_{k_t} + 1$
  Update Empirical reward: $\hat{\mu}_{k_t}(t) = \frac{\hat{\mu}_{k_t}(t-1)(n_{k_t}(t)-1)+r_t}{n_{k_t}(t)}$
  Update the UCB Index: $I_{k_t}(t) = \hat{\mu}_{k_t} + B\sqrt{\frac{2\log t}{n_{k_t}}}$
  Compute pseudo-rewards for all arms $k \neq k_t$: $s_{k,k_t}(r_t) = \max_{x: g_{k_t}(x) = r_t} g_k(x)$.
  Update empirical pseudo-rewards for all $k \neq k_t$: $\hat{\phi}_{k,k_t}(t) = \sum_{\tau: k_\tau = k_t} s_{k,k_\tau}(r_\tau)/n_{k_t}$
end for