Motivation

- preserve user privacy in online recommender systems.
- conceal individual choices about sensitive behaviors and beliefs.

Example: Randomized response method (RR) [Warner (1965)]

We introduce generalized corruption functions.

Problem setting

- $K$ arms with means $\mu_1, \ldots, \mu_K$ w.l.o.g.
- $\mu_1 > \mu_2, \ldots, \mu_K$
- Learner pulls an arm $A_t$ at time $t = 1, \ldots, T$
- receives reward $\sim$ Bernoulli distribution with mean $\mu_{A_t}$
- observes feedback $\sim$ Bernoulli distribution with mean $\lambda_{A_t}$
- A known corruption function $g_a: \mu_a \mapsto \lambda_a$
- Assumption: $g_a$ is monotonic and continuous.
- Goal: Minimize $\text{Regret}_{T} = \sum_{t=1}^{T} \Delta_a \mathbb{E}[N_a(T)]$ where $N_a(T) = \sum_{t=1}^{T} 1_{\{A_t = a\}}$, and $\Delta_a = \mu_1 - \mu_a$.

Randomized response

- Corruption function $g_a: \lambda_a = p_{\theta_0}(a) + (p_{\theta_1}(a) - p_{\theta_0}(a))\mu_a$
- $p(\text{feedback} = x \mid \text{reward} = y) = M_a(x, y) = \begin{cases} 0 & \text{if } p_{\theta_0}(a) = p_{\theta_1}(a) \\ 0.5 & \text{if } p_{\theta_0}(a) \neq p_{\theta_1}(a) \end{cases}$
- Initialization: Pull each arm once.
- Pull arm $A_{t+1} = \arg \max_a \text{Index}_a(t)$.

Lower bound on regret

**Definition 1.** An uniformly efficient algorithm for the corrupt bandit problem is an algorithm which, for any bandit model, has $\text{Regret}_{T} = o(T^\alpha)$ for all $\alpha \in [0, 1]$.

**Theorem 1.** Fix the corruption functions $(g_a)_{a=1}^K$. Any uniformly efficient algorithm, for a corrupt bandit problem, satisfies

$$\liminf_{T \to \infty} \frac{\text{Regret}_{T}}{\log(T)} \geq \sum_{a=2}^{K} \frac{\Delta_a}{d(\lambda_a, g_a(\mu_1))},$$

where $d(x, y) = \text{KL}(B(x), B(y))$.

**KLUCB-CF**

1. Input: A bandit model having $K$ arms
2. Parameters: $(g_a)_{a=1}^K$, a non-decreasing (exploration) function $f: N \to \mathbb{R}$, $d(x, y) = \text{KL}(B(x), B(y))$.
3. Initialization: Pull each arm once.
4. At time $t \geq K + 1$, do:
   5. Compute for each arm $a$, one of the following quantities:
      - $\text{Index}_a(t) = \begin{cases} g_a^{-1}(\ell_a(t)) & \text{if } g_a \text{ is decreasing} \\ g_a^{-1}(a_u(t)) & \text{if } g_a \text{ is increasing} \end{cases}$
      - where $\ell_a(t) = \min\{q : N_a(t) \cdot d(\lambda_a(t), q) \leq f(t)\}$
      - $a_u(t) = \max\{q : N_a(t) \cdot d(\lambda_a(t), q) \leq f(t)\}$
   6. Pull arm $A_{t+1} = \arg \max_a \text{Index}_a(t)$.
   7. Observe feedback $A_{t+1}$.

**Theorem 2.** The expected regret of KLUCB-CF using $f(t) = \log(t) + 3\log(\log(t))$ on a $K$-armed corrupted bandit with corruption functions $(g_a)_{a=1}^K$ is upper bounded by

$$\text{Regret}_T \leq \sum_{a=2}^{K} \frac{\Delta_a \log(T)}{d(\lambda_a, g_a(\mu_1))} + O(\sqrt{\log(T)}).$$

**Corrupted feedback to enforce differential Privacy**

**Definition 2.** A corrupt feedback corruption scheme $g$ is $(\epsilon, \delta)$-differentially private if for all reward sequences $R_{t1}, \ldots, R_{t2}$ and $R_{t1}^\prime, \ldots, R_{t2}^\prime$ that differ in at most one reward, and for all $S \subseteq \text{Range}(g)$,

$$\Pr[g(R_{t1}, \ldots, R_{t2}) \in S] \leq e^\epsilon \cdot \Pr[g(R_{t1}^\prime, \ldots, R_{t2}^\prime) \in S] + \delta$$

- Privacy preserving input
- Differential privacy requires that $\max_{a \in K} \left( \frac{p_{\theta_0}(a)}{p_{\theta_1}(a)}, \frac{p_{\theta_1}(a)}{p_{\theta_0}(a)} \right) \leq e^\epsilon + \delta$
- To achieve $(\epsilon, \delta)$-differential privacy with randomized response,

$$M_a = \begin{cases} 0 & \text{if } p_{\theta_0}(a) = p_{\theta_1}(a) \\ 1 & \text{if } p_{\theta_0}(a) \neq p_{\theta_1}(a) \end{cases}$$

**Experiments**

- Randomized response as corruption function.
- Scenario 1: Two arms with mean rewards 0.9 and 0.6
- Figure 1(a) shows average regret for $p_{\theta_0}(1) = p_{\theta_1}(1) = 0.6$ and $p_{\theta_0}(2) = p_{\theta_1}(2) = 0.9$
- Figure 1(b) shows the performance for varying values of $p = p_{\theta_0}(1) = p_{\theta_1}(1) = p_{\theta_0}(2) = p_{\theta_1}(2)$ with $T = 10^4$

**Conclusion**

- UCB-CF, KLUCB-CF, and Thompson Sampling-CF provide suitable solutions. KLUCB-CF is the best solution as it is asymptotically optimal and outperforms others in experiments.
- We provide appropriate corruption matrices that achieve a desired level of differential privacy.

**Key references**

[Auer Peter, Cesa-Bianchi Nicolò, and Fischer Paul (2002)] Finite-time Analysis of the Multiarmed Bandit Problem.