Smoothness-Adaptive Contextual Bandits

Yonatan Gur, Ahmadreza Momeni, and Stefan Wager

Stanford University, Stanford, CA 94305

{ygur, amomenis, swager}@stanford.edu

Abstract

We study a non-parametric multi-armed bandit problem with stochastic covariates, where a key complexity driver is the smoothness of payoff functions with respect to covariates. First, we establish that adapting to unknown smoothness of payoff functions is, in general, impossible. However, under a self-similarity condition (which does not reduce the minimax complexity of the problem at hand), we establish that adapting to unknown smoothness is possible, and further devise a general policy for achieving smoothness-adaptive performance.

Formualtion: Non-parametric Contextual Bandits

- T steps, $K=2$ arms, context space $[0,1]^d$
- In each decision period, agent selects an arm
- Before selection, context $X_t \overset{\text{i.i.d.}}{\sim} P_X$ observed
- Rewards $Y_{k,t} \in [0,1]$ s.t. $\mathbb{E}[Y_{k,t}] = f_k(X_t)$
- f_k payoff function of arm k
- Performance of a policy π measured by regret w.r.t. a *dynamic* oracle:

$$\mathcal{R}^{\pi}(\mathsf{P};T) = \mathbb{E}^{\pi}\left[\Sigma_{t=1}^{T} \max_k f_k(X_t) - f_{\pi_t}(X_t)\right]$$

- Complexity depends on payoff structure
- In non-parametric model, **smoothness** of payoffs drives complexity
- Existing work assumes prior knowledge about smoothness (not a practical assumption)

General Model Assumptions

- **Payoff smoothness:** payoff functions are (β, L)-Hölder; $\exists\, \beta \in [\underline{\beta}, \bar{\beta}]$, $L>0$ s.t. $\forall x, x'$

$$\Big| f_k(x) - \underbrace{\mathrm{TE}(f_k, \lfloor \beta \rfloor, x'; x)}_{\text{Taylor expansion of degree } \lfloor\beta\rfloor} \Big| \le L \|x - x'\|_\infty^\beta$$

  - larger $\beta \Rightarrow$ smoother payoffs $\Rightarrow$ easier problem
- **Context distribution:**

$$\exists\, 0 < \underline{\rho} \le \bar{\rho} : \ \underline{\rho} \le p_X(x) \le \bar{\rho} \qquad \forall x$$

  - e.g., uniform distribution
- **Margin:**

$$P_X\{0 < |f_1(X) - f_2(X)| \le \delta\} \le C_0 \delta^\alpha \ \forall \delta > 0$$

  - captures mass of contexts near decision boundary
  - larger $\alpha \Rightarrow$ less context mass near decision boundary $\Rightarrow$ easier problem

Minimax Regret Rate with the Knowledge of Smoothness

It has been shown (see Rigollet and Zeevi (2010), Perchet and Rigollet (2013), Hu et al. (2019)) that

$$\inf_{\pi \in \Pi} \sup_{\mathsf{P} \in \mathcal{P}(\beta,\alpha,d)} \mathcal{R}^{\pi}(\mathsf{P};T) = \Theta\left(T^{\zeta(\beta,\alpha,d)}\right), \quad \text{where } \zeta(\beta,\alpha,d) = 1 - \frac{\beta(1+\alpha)}{2\beta+d}.$$

- Minimax regret rate depends on smoothness parameter β
- Previous policy design based on knowledge of smoothness parameter β

Impossibility of Costless Adaptation to Smoothness

Theorem (Impossibility of adapting to smoothness)

Fix two Hölder exponents $\beta < \gamma$. Assume policy π is rate-optimal over γ-smooth problems.

1. (At most Lipschitz-smooth) For $0 < \beta < \gamma \le 1$: $\sup_{\mathsf{P} \in \mathcal{P}(\beta,\alpha,d)} \mathcal{R}^{\pi}(\mathsf{P};T) \ge \underline{C} T^{1 - \frac{d}{\alpha(2\beta+d-\alpha\beta)}} \left[T^{\zeta(\gamma,\alpha,d)}\right]^{-\frac{d}{2\beta+d-\alpha\beta}}$;
2. (At least Lipschitz-smooth) For $\beta = 1 < \gamma$: $\sup_{\mathsf{P} \in \mathcal{P}(1,\alpha,d)} \mathcal{R}^{\pi}(\mathsf{P};T) \ge \underline{C} T^{1 - \frac{1}{2\alpha}} \left[T^{\zeta(\gamma,\alpha,d)}\right]^{-\frac{1}{2}}$.

- Lower bound on performance over rough problems as a function of performance over smooth problem
- Implies existance of pairs (β, γ) s.t. optimality for both impossible (see the following examplea)
- Smoothness-adaptivity impossible, without additional requirements

Example (Impossibility of adapting to smoothness)

1. (At most Lipschitz-smooth) $\gamma = \frac{15}{100}$, $\beta = \frac{\gamma}{2}$, $\alpha = \frac{99}{100\gamma}$, and $d=1$: Optimal rate over β-smooth problems for policies that are rate-optimal over γ-smooth problems without knowledge of β is $\Omega\left(T^{0.58}\right)$ while with knowledge of β, the optimal rate is $\mathcal{O}(T^{0.504348})$.
2. (At least Lipschitz-smooth) $\gamma > 1$, $\beta = 1$, $\alpha = 1$ and $d = 1$: Optimal rate over β-smooth problems for policies that are rate-optimal over γ-smooth problems without knowledge of β is $\Omega\left(T^{\frac{\gamma}{2\gamma+1}}\right)$ while with knowledge of β, the optimal rate is $\mathcal{O}(T^{\frac{1}{3}})$.

A Sufficient Condition for Aapting to Smoothness

Definition (Self-similarity): A set of payoff functions $\{f_k\}_{k\in\mathcal{K}}$ is self-similar if for some $\beta \in [\underline{\beta}, \bar{\beta}]$

- all payoffs are β-Hölder;
- $\exists b > 0, l_0 > 0$ s.t. $\forall l \ge l_0 : \max_{\mathsf{B}\in\mathcal{B}_l} \max_{k\in\mathcal{K}} \sup_{x\in\mathsf{B}} |\Gamma_l^p f_k(x;\mathsf{B}) - f_k(x)| \ge b 2^{-l\beta} \quad \forall p \in \{\lfloor \underline{\beta} \rfloor, \ldots, \lfloor \bar{\beta} \rfloor\}$
  - $\Gamma_l^p f_k(x;\mathsf{B})$: projection of f_k to polynomials of degree p over B
  - $\mathcal{B}_l = \{\mathsf{B}_m,\ m = 1,\ldots,2^{ld}\}$ is a collection of the hypercubes: $\mathsf{B}_m = \mathsf{B}_{\mathsf{m}} = \{x \in [0,1]^d : \frac{\mathsf{m}_i - 1}{2^l} \le x_i \le \frac{\mathsf{m}_i}{2^l},\ i \in \{1,\ldots,d\}\}$

- Effectively implies a global lower bound on the estimation bias
- Can be viewed as complementing Hölder smoothness, which implies upper bound on bias
- Does not reduce regret complexity
- **Example:** $f_1(x) = \frac{1}{2}$, $f_2(x) = x^\beta$ for some $\beta \le 1 = \bar{\beta}$

Smoothness-Adaptive Contextual Bandits (SACB) Policy

Algorithm 1 `SACB`

Require: Set of non-adaptive policies $\{\pi_0(\beta_0)\}_{\beta_0 \in [\underline{\beta}, \bar{\beta}]}$, horizon length T, minimum and maximum smoothness exponents $\underline{\beta}$ and $\bar{\beta}$, and a tuning parameter γ

1: **Initialize:** $g \leftarrow \left\lceil \frac{(\underline{\beta}+d-1)\log_2 T}{(2\bar{\beta}+d)^2} \right\rceil$; $\bar{r} \leftarrow \lceil 2l\bar{\beta} + (\frac{2d}{\underline{\beta}} + 4)\log_2 \log T \rceil$

2: Partition the context space $[0,1]^d$ into equal sized hypercubes with side-length 2^{-g}; Denote the set of hypercubes by $\mathcal{B}_g$

3: **Sampling:** Collect samples in each hypercube $\mathsf{B} \in \mathcal{B}_g$ by alternating between the arms for $r \in \{1, \ldots, \bar{r}\}$ rounds; in each round collect 2^r samples for each arm

4: **Estimation:** At the end of each sampling round r in each hypercube $\mathsf{B} \in \mathcal{B}_g$, form two separate estimates of each payoff function using polynomial regression of degree $\lfloor \bar{\beta} \rfloor$ and badnwidths 2^{-j} for $j = j_1 = g$, and $j = j_2 = g + \lceil \frac{1}{\underline{\beta}} \log_2 \log T \rceil$; Denote the estimates by $\hat{f}_k^{(\mathsf{B},r)}(x;j)$

5: **Hypothesis test:** At the end of each sampling round r in each hypercube $\mathsf{B} \in \mathcal{B}_g$, check whether the difference between the estimation using the two bandwidths exponents j_1 and j_2 exceeds a pre-determined threshold for some tuning param. γ:

$$\sup_{k\in\mathcal{K}, x\in\mathsf{B}} \left| \hat{f}_k^{(\mathsf{B},r)}(x; j_1^{(\mathsf{B})}) - \hat{f}_k^{(\mathsf{B},r)}(x; j_2^{(\mathsf{B})}) \right| \ge \frac{\gamma (\log T)^{\frac{d}{2\underline{\beta}} + \frac{1}{2}}}{2^{r/2}}.$$

Denote by $r_{\text{last}}^{(\mathsf{B})}$ the smallest round index for which the above inequality holds in hypercube B

6: **Smoothness estimation:** After sampling finished, set

$$\hat{\beta}_{\texttt{SACB}} = \frac{1}{2g}\left[\min_{\mathsf{B}\in\mathcal{B}_g} r_{\text{last}}^{(\mathsf{B})} - \left(\frac{2d}{\underline{\beta}} + 4\right)\log_2 \log T\right]$$

7: **Model selection:** Choose the corresponding non-adaptive policy $\pi_0 \leftarrow \pi_0(\min[\max[\underline{\beta}, \hat{\beta}_{\texttt{SACB}}], \bar{\beta}])$ and run it for the remaining time steps

- Estimates smoothness by comparing estimation bias and variance
- Key idea of estimation: estimation bias of self-similar and Hölder-smooth payoffs is bounded from above and below
- Adaptively integrates a smoothness estimation sub-routine with some collection of non-adaptive rate-optimal policies $\{\pi_0(\beta_0)\}_{\beta_0 \in [\underline{\beta}, \bar{\beta}]}$
- Achieves smothness-adaptivity when paired with rate-optimal off-the-shelf policies $\{\pi_0(\beta_0)\}_{\beta_0 \in [\underline{\beta}, \bar{\beta}]}$