On Thompson Sampling for Smoother-than-Lipschitz Bandits

James A. Grant and David S. Leslie

Lancaster University

AISTATS 2020

Lancaster University

Abstract

- Extend understanding of **Thompson Sampling** for stochastic bandits.
- Bound on the Bayesian regret of Thompson Sampling for continuum-armed bandits with **nonparametric, smooth reward functions**, and **sub-exponential** noise.
- Achieved by analysis based on the **eluder dimension** (a smoothness measure) of the reward function class.

Problem Setting

- Continuum armed bandit specified by a tuple $(\mathcal{A}, f, p)$
  - $\mathcal{A} \subset \mathbb{R}^d$ is the action set,
  - $f: \mathcal{A} \to \mathbb{R}$ is the reward function, lying in a function class $\mathcal{F}$,
  - p on $\mathbb{R}$ is the reward noise distribution.

- A learner who knows $\mathcal{A}$ (but not f) iterates, for $t = 1,2, \ldots T$,
  - Select an action $a_t \in \mathcal{A}$
  - Observe a reward $R(a_t) = f(a_t) + \eta_t$, where $\eta_t \sim p$.

- The learner's objective is to minimise Bayesian regret,

$$\min_{a_1,\ldots,a_T} E_{\pi_0}\left(\sum_{t=1}^{T}\left(\max_{a\in\mathcal{A}} f(a) - f(a_t)\right)\right).$$

Smoother-than-Lipschitz Functions

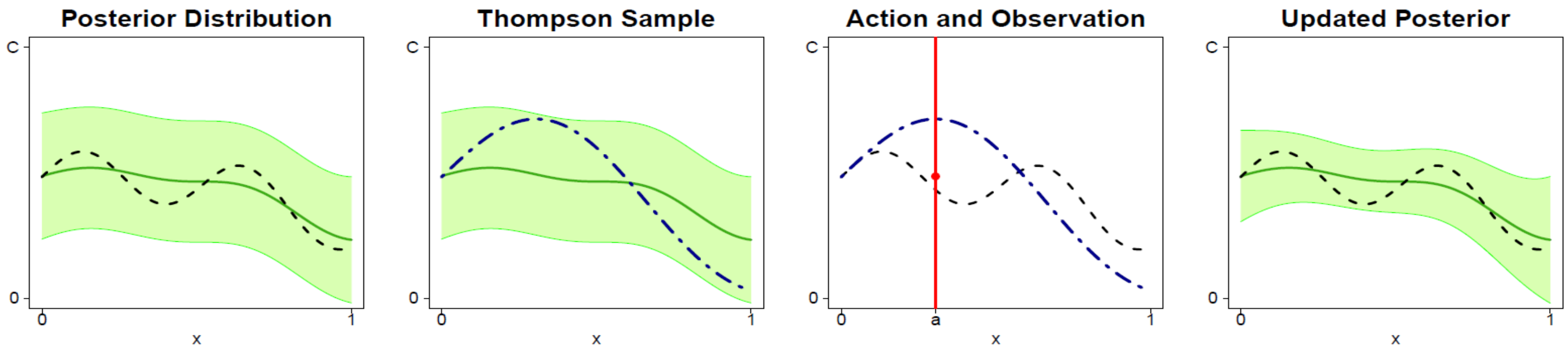
- The achievable scaling of regret depends on the smoothness of f
- Some known results,
  - For f Lipschitz: Optimal regret $\Omega\left(T^{2/3}\right)$ - [K05]
  - For f drawn from a Gaussian Process: Optimal regret $\Omega\left(\sqrt{T}\right)$ - [SKKS12]

- We focus on f having $M \in \mathbb{N}$ Lipschitz derivatives,

$$f \in \mathcal{F}_{C,M,L} = \left\{g \colon \mathcal{A} \to [0, C] \;\; \text{s.t.} \;\; \left|g^{(m)}(a) - g^{(m)}(a')\right| \leq L|a - a'|, m \leq M\right\}$$

Thompson Sampling

- Thompson Sampling is a Bayesian approach to choosing $a_t \in \mathcal{A}$ in each round.
- Initialised by a prior distribution π_0 on $\mathcal{F}$, at each $t = 1, \dots, T$, do,
  - Draw a function $\tilde{f}_t \sim \pi_{t-1}$
  - Choose an action $a_t \in \operatorname{argmax}_{a \in \mathcal{A}} \tilde{f}_t(a)$
  - Observe $R(a_t)$ and compute π_t as posterior on f.



Main Result

Theorem *The Bayesian regret of Thompson Sampling with prior distribution p_0 on $\mathcal{F}_{C,M,L}$ applied a the continuum armed bandit problem with reward function f_0 drawn from p_0, and sub-exponentially distributed noise satisfies*

$$BR(T) = O\left(T^{(2M^2+11M+10)/(4M^2+14M+12)}\right).$$

- Recall that M is the number of Lipschitz derivatives.
- For $M = 0$, the bound is $O\left(T^{5/6}\right)$, and for $M = 1$, the bound is $O\left(T^{23/30}\right)$.
- As $M \to \infty$ the bound approaches $O\left(\sqrt{T}\right)$.

Proof Sketch

- For parametric problems, where $f = f_\theta, \theta \in \mathbb{R}^d$, and confidence sets
$$\{\Theta_t\}_{t=1}^T \text{ with } P(\theta \in \Theta_{t+1} | a_{1:t}, R_{1:t}) \geq 1 - \delta,$$
- [RVR14] show,
$$BR(T, \pi^{TS}) \leq T\delta + \mathbb{E}\left(\sum_{t=1}^{T} \sup_{\theta \in \Theta_t} f_\theta(a_t) - \inf_{\theta \in \Theta_t} f_\theta(a_t)\right)$$
- They derive sets $\widehat{\Theta}_t$ centred on the least squares estimator, whose width may be expressed in terms of properties of $\mathcal{F}$ - the class of potential reward functions.

Proof Sketch

First step is an analogue of $\widehat{\Theta}_t$ for non-parametric settings.

Lemma (informal) *For sets,*

$$\mathcal{F}_t = \left\{ f \in \mathcal{F}\colon \sum_{i=1}^{t} \left(\hat{f}_{LS,t}(a_i) - f(a_i) \right)^2 \leq \beta(\delta, \alpha(t)) \right\}$$

where $\beta(\delta, \alpha(t)) \propto N(\alpha(t), \mathcal{F}, \|\cdot\|_\infty)$*, and*

$$\hat{f}_{LS,t} \in \operatorname*{argmin}_{f \in \mathcal{F}} \sum_i \left(f(a_i) - R(a_i) \right)^2 ,$$

we have $P\left(f_0 \in \cap_{i=1}^{t} \mathcal{F}_t \right) \geq 1 - 2\delta.$

Proof Sketch

Second step is to bound the sum of diameters of $\mathcal{F}_t$ sets.

Lemma (informal) *For sets $\mathcal{F}_t \subset \mathcal{F}$ as defined previously, and all non-increasing functions $\kappa : \mathbb{N} \to \mathbb{R}$ we have that the sum of diameters $\sum_{t=1}^{T} \sup_{f \in \mathcal{F}_t} f(a_t) - \inf_{f \in \mathcal{F}_t} f(a_t)$ is bounded by,*

$$T\kappa(T) + d_E\big(\mathcal{F}, \kappa(T)\big) + \sqrt{d_E\big(\mathcal{F}, \kappa(T)\big)\beta(\delta, \alpha(T))T}.$$

- $d_E(\mathcal{F}, \kappa(T))$ is the eluder dimension of $\mathcal{F}$ - defined momentarily.
- Everything to this point is for general $\mathcal{F}$ and doesn't depend on Lipschitzness.
- We choose $\kappa(T)$ and $\alpha(T)$ to optimise a bound.

Proof Sketch

Finally, specialise to the smoother-than-Lipschitz setting by bounding covering number and eluder dimension.

1. Covering number bound available from classic theory [KT1961],

$$\log N\left(\alpha, \mathcal{F}_{C,M,L}, \|\cdot\|_\infty\right) = \Theta\left(\alpha^{-\frac{1}{M+1}}\right)$$

2. $d_E(\mathcal{F}, \kappa(T))$ is the length, D, of longest sequence $a_1, \dots, a_D \in \mathcal{A}$, such that for every $i \in \{1, \dots, D\}$ there exist $f, f' \in \mathcal{F}$ such that

$$f(a_i) - f'(a_i) > \kappa(T)$$

and

$$\sqrt{\sum_{j=1}^{i} \left(f(a_j) - f'(a_j)\right)^2} \leq \kappa(T)$$

Proof Sketch

Lemma *The eluder dimension of* $\mathcal{F}_{C,M,L}$ *can be bounded as*

$$d_E\left(\mathcal{F}_{C,M,L}, \kappa\right) = o\left(\left(\frac{\kappa}{L}\right)^{-1/(M+1)}\right).$$

- The proof considers functions $h = f - f'$ where $f, f' \in \mathcal{F}_{C,M,L}$.
- In particular, look at if $h(a) > \kappa$, how small can δ be such that $h(a \pm \delta) \ll \kappa$.
- The more smooth derivatives, the larger δ, and the smaller the eluder dimension.

Lower Bound

Theorem *For any algorithm for continuum armed bandit problems of the form* $([0,1], f, p)$ *where* $f \in \mathcal{F}_{C,M,L}$ *and* p *is sub-exponential, there exists a problem instance such that the regret incurred satisfies*

$$Reg(T) = \Omega\left(T^{(M+2)/(2M+3)}\right).$$

- Recall the upper bound is $O\left(T^{(2M^2+11M+10)/(4M^2+14M+12)}\right)$.
- There is a gap of order $T^{(3M+2)/(4M^2+14M+12)}$, which vanishes as $M \to \infty$.
- **Open Question:** Is this gap a feature of TS or of the eluder dimension based analysis?

Summary

- Presented Bayesian regret bound for non-parametric Thompson Sampling on smooth continuum-armed bandit problems.

- Proof via eluder dimension analysis, leads to result which matches lower bound in case of infinitely many smooth derivatives.

- Open questions around gap for finitely many smooth derivatives, and the extension to higher dimensional action spaces.

Thank you for watching

-

j.grant@lancaster.ac.uk

[K05] – Kleinberg (2005). Nearly Tight Bounds for the Continuum-Armed Bandit Problem. *NeurIPS.*

[KT1961] – Kolmogorov and Tikhomirov (1961). Ɛ-entropy and Ɛ-capacity of Sets in Function Spaces. *Translations of the American Mathematical Society.*

[RVR14] – Russo and Van Roy (2014). Learning to Optimise via Posterior Sampling. *Mathematics of Operations Research.*

[SKKS12] – Srinivas, Krause, Kakade and Seeger (2012). Information Theoretic Regret Bounds for Gaussian Process Optimisation in the Bandit Setting. *IEEE Transactions on Information Theory.*