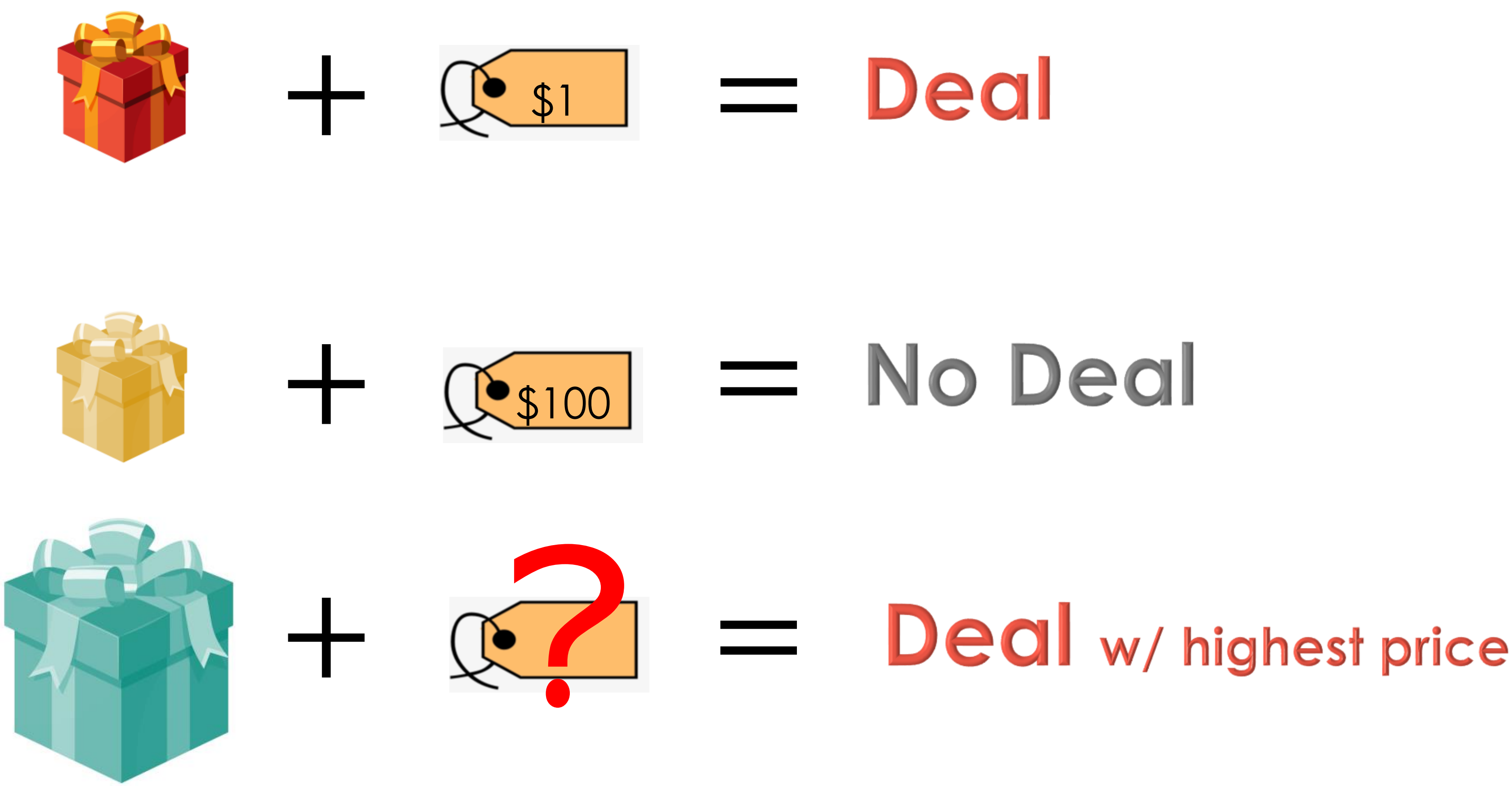
11/10/2020

Dynamic Pricing in High Dimensions

Adel Javanmard, Hamid Nazerzadeh

Reading group led by: Jianyu Xu

What's dynamic pricing?



Problem Setting

For $t = 1, 2, \ldots, T$:

- The nature chooses a feature x_t;

• The firm observes x_t;	• The customer observes x_t;
• The firm *proposes* a price $p_t = \pi_t(x_t)$;	• The customer valuates the product as $v_t(x_t)$
	• The customer compares p_t and v_t, and decides $y_t = \mathbb{I}(v_t \geq p_t)$;
• The firm receives a reward $r_t = p_t \cdot y_t$	

Recap: Contextual Bandit

For $t = 1, 2, \ldots, T$:

0. Nature draws $(x_t, \boldsymbol{r}_t)$ from dist. $\mathcal{D}$ over $\mathcal{X} \times [0,1]^{\mathcal{A}}$.
1. Observe context $x_t \in \mathcal{X}$. [e.g., user profile, search query]
2. Choose action $a_t \in \mathcal{A}$. [e.g., ad to display]
3. Collect reward $r_t(a_t) \in [0,1]$. [e.g., 1 if click, 0 otherwise]

Task: choose a_t's that yield high expected reward (w.r.t. $\mathcal{D}$).

Contextual: use features x_t to choose good actions a_t.

Bandit: $r_t(a)$ for $a \neq a_t$ is not observed.

https://www.cs.columbia.edu/~djhsu/papers/ilovetoconbandits-slides.pdf

Basic Assumptions

I.I.D.

- $x_t \sim \mathbb{P}_X \subset \mathbb{R}^d$ independently and identically;
  - $\mathbb{P}_X$ is unknown to us
  - $\mathbb{P}_X$ is supported by a bounded set $\mathcal{X}$.

Linear

- $v_t(x) = \alpha_0 + \theta_0^T x + z_t$, or $v(x) = \mu_0^T \tilde{x} + z_t$
  - here z_t are marketing shocks (noises)
  - z_t drawn i.i.d. from a distribution with 0-mean and CDF F
  - F is known to us

Achieve-
-able

- $\mu_0 \in \Omega = \{\mu \in \mathbb{R}^{d+1}: ||\mu||_0 \leq s_0, ||\mu||_1 \leq W\}$
  - s_0 is a sparsity factor, and $s_0 = d + 1$ in a dense case

Technical Assumptions

- Assumption 2.1: The noise CDF $F(v)$ is:
  - Known!
  - strictly increasing;
  - $F(v)$ and $\left(1 - F(v)\right)$ are both log-concave w.r.t. v.
  - E.g.: normal, uniform, Laplace, exponential, logistic,…
- Assumption 2.2: The distribution $\mathbb{P}_X$ satisfies:
  - $\mathbb{E}_{x_t \sim \mathbb{P}_X}[x_t] = 0, \forall\, t = 1,2, \dots, T, \dots;$
    - Normalized by α_0
  - $\Sigma = \mathbb{E}_{x_t \sim \mathbb{P}_X}[x_t x_t^T]$ with any singular value $C_{\min} \leq \sigma_i \leq C_{\max}$
    - Here $C_{\max} \geq 1 \geq C_{\min} > 0$ are constants.

Recap: settings

For $t = 1,2, \ldots, T$:

- Samples $x_t \sim \mathbb{P}_X$, i.i.d.

The firm	The customer
• The firm observes x_t;	• The customer observes x_t;
• Proposes a price $p_t = \pi_t(x_t)$;	• Valuates the product as $v_t(x_t) = x_t^T \mu_0 + z_t$
	• Compares p_t and v_t, and decides $y_t = \mathbb{I}(v_t \geq p_t)$; • $\mathbb{P}(y_t = 1) = 1 - F(p_t - x_t^T \mu_0)$
• Receives a reward $r_t = p_t \cdot y_t$ • $\mathbb{E}[r_t \mid x_t, p_t] = p_t\left(1 - F(p_t - x_t^T \mu_0)\right)$	

Greedy function $g(v)$

- Reward is random, so we maximize its expectation.
- $\mathbb{E}[r_t(p)] = p \cdot \left(1 - F(p - \mu_0^T \tilde{x}_t)\right)$
- Define $g(v) \triangleq argmax_p\, p \cdot \left(1 - F(p - v)\right)$.
  - A greedy pricing function.
- Therefore, $p_t^* = g(\mu_0^T \tilde{x}_t)$.

Expected Regret

- We define the expected regret as:

$$\mathrm{Regret}_\pi(T) \equiv \max_{\substack{\mu_0 \in \Omega \\ \mathbb{P}_X \in Q(\mathcal{X})}} \mathbb{E}\left[\sum_{t=1}^{T} \left(p_t^* \mathbb{I}(v_t \geq p_t^*) - p_t \mathbb{I}(v_t \geq p_t) \right)\right]$$

- This is a worst-case regret w.r.t. μ_0 and $\mathbb{P}_X$

Main Results

- An algorithm of $O(s_0 \log d \cdot \log T)$ regret.
  - Under Assumption 2.1 and 2.2
- A lower bound of $\Omega\big(s_0(\log d + \log T)\big)$.
- They are almost matching.

Recap: notations

Notation	Definition
$x_t, \tilde{x}_t$	Feature vector; x_t padding an "1"
p, p_t, v_t	Price; price proposed at time t; valuation at time t
θ_0, μ_0	Valuation parameter; θ_0 padding an α_0
z_t	Noise
r_t, y_t	Reward ($r_t = p_t \cdot y_t$); decision (buy: 1; not buy: 0)
F, f	Noise CDF; noise PDF
s_0	Sparsity
$\Sigma, C_{\max}, C_{\min}$	$\Sigma \triangleq \mathbb{E}[xx^T]$, with $C_{\max} I_d \succcurlyeq \Sigma \succcurlyeq C_{\min} > 0$
$g(v)$	$g(v) \triangleq argmax_p \, p \cdot \big(1 - F(p - v)\big)$
p_t^*	$p_t^* \triangleq g(\mu_0^T \tilde{x}_t)$
$\mathbb{P}_X, \mathcal{X}$	Distribution of x; support of $\mathbb{P}_X$
Ω, W	Parameter domain, ℓ_1-norm-bound of any parameter

Idea of Algorithm Designing

- Max likelihood estimator (MLE)
  - A well-parameterized model
- Greedy policy
  - Make best use of estimators
- Doubling Episodes
  - Fewer parameter updates
  - Easier analysis …
- Regularization parameter
  - Promote sparsity structure in the estimated parameter

Maximum Likelihood Estimation

- The negative log-likelihood function:

$$\mathcal{L}(\mu) = -\frac{1}{n}\sum_{t=1}^{n} \mathbb{I}(y_t = 1)\log\Big(1 - F(p_t - \tilde{x}_t^T\mu)\Big)$$
$$+\mathbb{I}(y_t = 0)\log\Big(F(p_t - \tilde{x}_t^T\mu)\Big)$$

- Strongly convex with high probability.
    - (See Proposition A.2.)

Greedy Policy

- Define $g(v) \triangleq argmax_p \mathbb{E}[r_t(p)]$
  $= argmax_p \, p \cdot \big(1 - F(p - v)\big).$

- $p_t^* = g(\mu_0^T \tilde{x}_t)$.

- We assume $p_t = g(\tilde{x}_t^T \mu_t)$ for some μ_t, without losing generality.
  - If μ_t is approaching μ_0 then the regret will be small.

Doubling Episodes

- The first episode: $\tau_1 = 0$, no periods.
  - Initialize all parameters to 0.
- For $k = 2,3, \ldots,$ let $\tau_k = 2^{k-1}$.
- Within each episode, we adopt the same μ_k.
  - Remember our pricing policy: $p_t = g(\tilde{x}_t^T \mu_t)$ for some μ_t

Algorithm 1: RMLP (Regularized Maximum Likelihood Pricing)

Double episodes: $\tau_k = 2^{k-1}$

Update parameter:

$$\hat{\mu}^k = \arg\min_{\|\mu\|_1 \le W} \{\mathcal{L}(\mu) + \lambda_k \|\mu\|_1\}$$

where $\mathcal{L}(\mu)$ being negative log-likelihood

Price greedily:

$$p_t = g\left(\tilde{x}_t^T \hat{\mu}^k\right)$$

$t = 1,2, \dots, \tau_{k-1}$

$t = \tau_k$

Algorithm 1: RMLP

Input: (at time 0**)** function g, regularizations λ_k, W (bound on $\|\mu_0\|_1$),
Input: (arrives over time) covariate vectors $\{\tilde{x}_t\}_{t\in\mathbb{N}}$
Output: prices $\{p_t\}_{t\in\mathbb{N}}$

1: $\tau_1 \leftarrow 1$, $p_1 \leftarrow 0$, $\widehat{\mu}^1 \leftarrow 0$
2: **for** each episode $k = 2, 3, \ldots$ **do**
3: Set the length of k-th episode: $\tau_k \leftarrow 2^{k-1}$.
4: Update the model parameter estimate $\widehat{\mu}^k$ using the regularized ML estimator obtained from observations in the previous episode:

$$\widehat{\mu}^k = \underset{\|\mu\|_1 \le W}{\arg\min} \ \{\mathcal{L}(\mu) + \lambda_k \|\mu\|_1\} \tag{8}$$

with

$$\mathcal{L}(\mu) = -\frac{1}{\tau_{k-1}} \sum_{t=\tau_{k-1}}^{\tau_k - 1} \Big\{ \mathbb{I}(y_t = 1)\log(1 - F(p_t - \mu\cdot\tilde{x}_t)) + \mathbb{I}(y_t = -1)\log(F(p_t - \mu\cdot\tilde{x}_t)) \Big\} \tag{9}$$

5: For each period t during the k-th episode, set

$$p_t \leftarrow g(\widehat{\mu}^k \cdot \tilde{x}_t) \tag{10}$$

Algorithm 1: RMLP policy for dynamic pricing

Parameter λ_k Chosen

- λ_k constraints the ℓ_1 –norm of the estimator $\hat{\mu}^k$
- We select $\lambda_k = 4u_W\sqrt{\frac{\log d}{\tau_{k-1}}}$, where
$$u_W = \max\{\log' F(-2W), -\log'\left(1 - F(2W)\right)\}.$$

Remember that $||\mu||_1 \leq W$.

Remarks on RMLP

- Deterministic
  - Exploitation by greedy policy $g(\mu^T \tilde{x})$
  - Exploration naturally through random x_t and z_t
- Oblivious
  - Only rely on data from previous episode
    - Remark: previous episode is half as large as the whole
  - Suitable for perishable data
- Efficient
  - Fewer updates of estimators
- Low regret

Regret Analysis – Main Idea

- Step 1: bound estimation error $||\hat{\mu} - \mu_0||_2$
  - Tool 1: $\mathcal{L}(\hat{\mu}_t) + \lambda||\hat{\mu}_t||_1$ is **optimal**;
  - Tool 2: $\mathcal{L}(\mu)$ is **concentrated** (Azuma-Hoeffding);
  - Tool 3: $\mathcal{L}(\mu)$ is **strongly convex** ($\mathbb{E}[x_t x_t^T] \succcurlyeq C_{\min} \cdot I \succ 0$);
- Step 2: bound pricing difference $|p_t^* - p_t|$
  - Tool 4: g is **Lipschitz;**
  - $p_t^* - p_t = g(\tilde{x}_t^T \mu_0) - g(\tilde{x}_t^T \hat{\mu}) \leq |\tilde{x}_t^T (\hat{\mu} - \mu_0)|$.
- Step 3: bound reward difference $r_t(p_t^*) - r_t(p_t)$
  - Tool 5: $r_t(p) = p(1 - F(p - \tilde{x}_t^T \mu_0))$ is **strongly convex;**
  - $r_t(p_t^*) - r_t(p_t) = O\big((p_t^* - p_t)^2\big)$.

Bound estimation error $\|\hat{\mu} - \mu_0\|_2$

Proposition 8.1 (Estimation Error). *Consider linear model (1) with $\mu_0 = (\theta_0, \alpha_0) \in \Omega$, under Assumptions 2.1 and 2.2. Let $\widehat{\mu}$ be the solution of optimization problem (33) with $\lambda \geq 4u_W\sqrt{(\log d)/n}$. Then, there exist positive constants c_0 and C such that, for $n \geq c_0 s_0 \log(d)$, the following inequality holds with probability at least $1 - 1/d - 2e^{-n/(c_0 s_0)}$:*

$$\|\widehat{\mu} - \mu_0\|_2^2 \leq \frac{16 s_0 \lambda^2}{\ell_W^2 C_{\min}^2}\,. \tag{35}$$

Proposition 8.3. *Under assumptions of Proposition 8.1, there exist constants $c, c_1 > 0$, such that for $n \geq c_1 d$, the following holds true:*

$$\mathbb{E}(\|\widehat{\mu} - \mu_0\|_2^2) \leq \frac{16(s_0 + 1)\lambda^2}{\ell_W^2 C_{\min}^2} + 4W^2 e^{-cn^2}\,. \tag{37}$$

Therefore, we may divide n into 3 cases:

- $1 \leq n < c_0 s_0 \log(d)$ (where n is small)
- $c_0 s_0 \log(d) \leq n < c_1 d$ (where n trades off with δ)
- $n \geq c_1 d$ (where n is large but δ is small)

Main idea of Proposition 8.1 & 8.3

- Second-order Taylor expansion of $\mathcal{L}(\mu)$:

$$\mathcal{L}(\mu_0) - \mathcal{L}(\widehat{\mu}) = -\langle \nabla \mathcal{L}(\mu_0), \widehat{\mu} - \mu_0 \rangle - \frac{1}{2}\langle \widehat{\mu} - \mu_0, \nabla^2 \mathcal{L}(\bar{\mu})(\widehat{\mu} - \mu_0) \rangle$$

- Red circle is bounded by

$$\mathcal{L}(\widehat{\mu}) + \lambda \|\widehat{\mu}\|_1 \le \mathcal{L}(\mu_0) + \lambda \|\mu_0\|_1$$

  - and then by triangular inequalities (parameterized by s_0).
- Blue circle is bounded by concentration inequality

$$\nabla \mathcal{L}(\mu) = \frac{1}{n} \sum_{t=1}^{n} \xi_t(\mu) \tilde{x}_t$$

- Black circle **bounds** the quadratic error:
  - $\nabla^2 \mathcal{L}(\mu) \succcurlyeq \frac{\ell_W}{n} \tilde{X}^T \tilde{X}$
  - $\mathbb{E}[x x^T] \succcurlyeq C_{\min} I$

Proof details

- See liveboard…

Recap- RMLP

Assumption

- I.I.D. features
- Parameterized linear model
  - Feasible domain
- Known F
- Upper & Lower bounds on $\mathbb{E}[xx^T]$

Design

- Episodes
- MLE
- Regularizor
- Greedy

Proof

- Bound estimation error
  - 2nd –order Taylor expansion
  - Hoeffding Concentration
  - Strong convexity
  - Support set
  - Triangular inequality
- Bound price difference
  - In the same order of estimation error
  -
- Bound regret
  - Quadratic to the pricing difference

Lower bound on regret $\Omega(s_0 \log T)$

- Suppose we can observe v_t in each round…
  - For each time: $\mathrm{v_t} = \tilde{x}_t^T \mu_0 + z_t$
  - A linear regression!

- What is the lower bound of online linear regression?

Lower bound on regret $\Omega(\log T)$

- Assume $z_t \sim \mathcal{N}(0, \sigma^2)$, and we have:

$$\min_{\pi \in \Pi} \text{Regret}_\pi(T) \geq C' \left\{ s_0 \log\left(\frac{T}{s_0}\right) + \min\left[\frac{T}{s_0}, s_0 \log\left(\frac{d}{s_0}\right)\right] \right\}.$$

(Theorem 5.1)

- Key theorem towards the lower bound.

Lower bound on regret $\Omega(\log T)$

General idea: Reductions

- Minimax regret → minimax estimation error

- Estimation error → distinguish in a $\boldsymbol{\delta}$**-packing** parameter set
  - Far enough to enlarge regret
  - Close enough to hardly distinguish
  - Le Cam's method

- Lower bound the error probability of distinguish
  - Fano's Inequality

Fano's Inequality

- Fano's Lemma:

Lemma 11 (Fano) *Let* $X_1, \ldots, X_n \sim P$ *where* $P \in \{P_1, \ldots, P_N\}$. *Let* ψ *be any function of* $X_1, \ldots, X_n$ *taking values in* $\{1, \ldots, N\}$. *Let* $\beta = \max_{j \neq k} \mathsf{KL}(P_j, P_k)$. *Then*

$$\frac{1}{N} \sum_{j=1}^{N} P_j(\psi \neq j) \geq \left(1 - \frac{n\beta + \log 2}{\log N}\right).$$

Intuition:

- LHS: Probability of incorrectly distinguishing (estimating) the distribution
- RHS: a high probability
  - Increases as N goes larger
  - Decreases as KL-divergence goes larger

Discussion: Why not $\Omega(\sqrt{T})$?

- An $\Omega(\sqrt{T})$ regret is necessary for generic stochastic/adverserial contextual bandit problems.
- A "separability assumption" will reduce it to $\Omega(\log T)$:
  - Constant reward gap between the best and the second best actions.
  - Pricing is continuous and thus NOT separable!
- Which assumption(s) leads to this logarithmic regret?

Which assumption qualifies $\Omega(\sqrt{T})$?

- Linearity?
- Parametrization?
- Known noise distribution?
- Stochastic feature?
- $C_{\min} > 0$?

Which assumption qualifies $\Omega(\sqrt{T})$?

Linearity? ------ NO.

- A nonlinear model:

$$v(x_t) = \psi(\theta_0 \cdot \phi(x_t) + \alpha_0 + z_t)\,,$$

- Theorem 6.3:

Theorem 6.3. *Let ψ be log-concave and strictly increasing. Suppose that Assumptions 2.1 and 6.1 (or its alternative, Assumption 6.2) hold. Then, regret of the RMLP policy described as Algorithm 2 is of $O(s_0 \log d \cdot \log T)$.*

- Assumption 6.1: define $\Sigma_\phi \coloneqq \mathbb{E}[\phi(x)\phi(x)^T]$, and then $C_{\max} \cdot I \succcurlyeq \Sigma_\phi \succcurlyeq C_{\min} \cdot I \succ 0$.

Which assumption qualifies $\Omega(\sqrt{T})$?

Parametrization? ------Yes.

- For totally non-parametrized model, it is at least as hard as contextual bandits.
  - $\Omega(\sqrt{T})$ regret is necessary
  - maybe not sufficient, due to an infinite action set.

Which assumption qualifies $\Omega(\sqrt{T})$?

Noise Distribution? ------Yes.

- A theorem in [BR12]:

Theorem 3.1 (General Regret Lower Bound). *Define a problem class* $\mathcal{C}_{\mathsf{GenLB}} = (\mathcal{P}, \mathcal{Z}, d)$ *by letting* $\mathcal{P} = [3/4, 5/4]$, $\mathcal{Z} = [1/3, 1]$, *and* $d(p; z) = 1/2 + z - zp$. *Then for any policy* ψ *setting prices in* $\mathcal{P}$, *and any* $T \geq 2$, *there exists a parameter* $z \in \mathcal{Z}$ *such that*

$$\mathrm{Regret}(z, \mathcal{C}_{\mathsf{GenLB}}, T, \psi) \geq \frac{\sqrt{T}}{48^3}\ .$$

- Is $O(\sqrt{T})$ achievable in our linear setting?

Broder, J., & Rusmevichientong, P. (2012). Dynamic pricing under a general parametric choice model. *Operations Research*, *60*(4), 965-980.

Which assumption qualifies $\Omega(\sqrt{T})$?

- Theorem 7.1

Theorem 7.1. *Consider the valuation model (1), where noises z_t are generated from a distribution $F_{m,\sigma}$, with unknown mean m and variance σ^2. Under Assumption 2.2 and assuming that distribution $F_{m,\sigma}$ satisfies Assumption 2.1, the regret of RMLP-2 policy is of $O\big(s_0(\log d)\sqrt{T}\big)$. Further, regret of any pricing policy in this case is $\Omega(\sqrt{T})$.*

Algorithm: RMLP-2

Input: Pricing function g (corresponding to $F_{0,1}$), regularizations λ_k, W (bound on $\|\mu_0\|_1$)
Input: (arrives over time) covariate vectors $\{\tilde{x}_t = (x_t, 1)\}_{t\in\mathbb{N}}$
Output: prices $\{p_t\}_{t\in\mathbb{N}}$

1: **for** each episode $k = 1, 2, \ldots$ **do**
2: For the first period of the episode, offer the price uniformly at random from $[0, 1]$.
3: Denote by $\mathcal{A}_k$ the set of first periods in episodes $1, \ldots, k$.
4: Update the model parameter estimate $\widehat{\mu}^k$ using the regularized ML estimator:

$$(\widehat{\mu}^k, \widehat{\beta}^k) = \underset{\|(\mu/\beta, \beta)\|_1 \le W}{\arg\min} \{\mathcal{L}(\beta, \mu) + \lambda_k \|\mu\|_1\} \tag{25}$$

with

$$\mathcal{L}(\mu, \beta) = -\frac{1}{k} \sum_{t\in\mathcal{A}_k} \Big\{\mathbb{I}(y_t = 1) \log(1 - F(\beta p_t - \mu \cdot \tilde{x}_t)) + \mathbb{I}(y_t = -1) \log(F(\beta p_t - \mu \cdot \tilde{x}_t))\Big\} \tag{26}$$

5: For each period t during the k-th episode, set

$$p_t \leftarrow \frac{1}{\widehat{\beta}^k} g(\widehat{\mu}^k \cdot \tilde{x}_t) \tag{27}$$

Algorithm 3: RMLP-2 policy for dynamic pricing

Which assumption qualifies $\Omega(\sqrt{T})$?

Stochastic feature? ------ Not sure.

- In [CLPL16], the features are adversarial.
- They achieves $O(T^{\frac{2}{3}})$ regret, based on EXP4.
  - This seems suboptimal.

Cohen, M. C., Lobel, I., & Paes Leme, R. (2020). Feature-based dynamic pricing. *Management Science*.

Which assumption qualifies $\Omega(\sqrt{T})$?

$C_{\min} > 0$? ------ No.

- In this paper a $C_{\min}$ helps prove $O(\log T)$ regret.
  - Without $C_{\min}$, an $O(\sqrt{T})$ regret bound is guaranteed, but not necessary.

Theorem 4.2. *Suppose that product feature vectors are generated independently from a probability distribution $\mathbb{P}_X$ with a bounded support $\mathcal{X} \in \mathbb{R}^d$. Under Assumption 2.1, the regret of RMLP policy is of $O(\sqrt{(\log d)T})$.*

- In our new works, we proved an $O(\log T)$ regret without $C_{\min}$.

Recap: which qualifies $\Omega(\sqrt{T})$?

- ~~Linearity?~~
- Parametrization?
- Known noise distribution?
- *Stochastic feature?*
- $\cancel{C_{\min} > 0}$~~?~~

Conclusion

Main Results

- Problem: dynamic pricing in high-dimensional features.
  - Linear valuation
  - Random feature
  - Known and fixed noise distribution
- Algorithm: RMLP
  - Max likelihood estimator with ℓ_1-regularizor
  - Episode-based greedy policy
    - Computationally efficient
    - Easy to analysis (avoiding martingale concentrations)
  - Upper regret bound: $O(s_0 \log d \cdot \log T)$
- Lower regret bound: $\tilde{\Omega}(s_0 \log T)$
- Nonlinear cases: $O(s_0 \log d \cdot \log T)$
- Unknown noise distribution
  - Parametrized: $O(\sqrt{T})$ and $\Omega(\sqrt{T})$
  - Unparametrized: $O(\delta T)$

Next Steps

- Proposed by the authors:
  - A tighter bound of upper & lower regret
  - μ_0 (or θ_0) is not sparse but close to a sparse vector.
  - Multiple-product sales at a time.
- Proposed by ourselves:
  - *Dynamic regret*, and adaptive regret
  - Adversarial features
    - Is it still $O(logT)$?
  - Totally unparametrized model
    - Is it $\Omega(\sqrt{T})$ or $\Omega\left(T^{\frac{2}{3}}\right)$?
    - Is it harder than contextual bandits?

UC SANTA BARBARA